UiT The Arctic University of Norway

DEPARTMENT OF ARCTIC & MARINE BIOLOGY

HUMAN RESPONSES TO YEAR-ROUND LIGHTING

USING GOOGLE TRENDS TO INVESTIGATE SEASONALITY IN ACTIVITY PATTERNS AND LIBIDO

HEDDA LÆRUM STARCK

MASTER'S THESIS IN BIOLOGY BIO-3950, NOVEMBER 2023

Acknowledgements

I want to express my sincere gratitude to several individuals who played a crucial role in the completion of this Master's Thesis. Firstly, I extend my heartfelt thanks to Dr Shona Wood for allowing me to focus my attention on the area of biology that captivates me the most: human behaviour and physiology. This opportunity made it a joy to work on the thesis, and for that, I will be forever grateful.

I am deeply appreciative of Dr Cody FitzGerald for his guidance throughout the entire process. His expertise and insights have been invaluable in shaping the technical aspects of this thesis. Thank you also for always encouraging me through obstacles and successes.

I sincerely appreciate my husband, Hector, whose unwavering support has been immeasurable. Thank you for brainstorming ideas with me around the dinner table and for always helping me when I ran into coding issues. I could not have asked for a better partner.

Lastly, I am thankful to Gunnstein for always being there to accompany me on a much-needed break and to my supportive friends and family. Your contributions made this thesis possible.

Thank you all from the bottom of my heart.

Hedda Lærum Starck

Tromsø, 8.11.23

Abstract

Seasonal rhythms in humans are fascinating but studying them is inherently challenging due to the need for data that covers multiple years. Due to the experimental obstacles, it remains debated whether humans exhibit seasonal fluctuations in aspects such as activity levels and libido. To approach these topics through indirect digital measures, this thesis used multi-year data from Google Trends to investigate seasonal variations in daily search activity patterns and pornography interest. We demonstrate a new method for modelling aggregate search data and calculating daily search activity durations. Our analyses detected seasonal differences in daily search activity duration in areas in the Northern and Southern Hemispheres (Norway, Sweden, Finland, New Zealand, and Victoria (Australia)). There was an approximately 20-minute average difference in search activity duration between spring and winter in the Northern Hemisphere and between summer and winter in the Southern Hemisphere. We also detected a clear inverse relationship between the Northern and Southern Hemispheres' seasonal fluctuations in search activity length. Similarly, we detected seasonal differences in search interest for pornography. Both the Northern and Southern Hemispheres showed a larger pornography search interest during the summer. However, significant surges in searches during the Christmas season and smaller spikes around Easter indicate that leisure time during holidays plays a crucial role in driving search interest for pornography. In brief, this study indicates seasonality in two important aspects of human lives: activity levels and libido but does not attempt to identify the underlying cause of the observed patterns.

Table of Contents

List of Tables

List of Figures

1 Preface

Our solar system was created 4.6 billion years ago when an immense cloud of gas and dust collapsed under its own gravity. A swirling motion was initiated by the collapse and maintained by the forming planets. The movement never stopped because of a lack of substantial resistance in the universe. We see the remainder of these immense forces today as planet Earth rotates daily on its axis and revolves yearly around the Sun (2).

When Earth was still young, it was hit by another planet, Theia, in a massive collision that ended up both tilting our world and producing the moon (3). The resulting tilt of 23.4° positions the two hemispheres closer or further away from the sun, depending on the Earth's location in its orbit. This produces seasons at all latitudes except the equator (4).

Some million years after this cataclysmic event, the first microbes appeared. Since then, life on Earth has evolved into the myriad of creatures that exist today. Many of these lineages have been exposed to yearly changes in temperature and light conditions and have evolved to optimise their life history traits in accordance with the seasons. The results are a range of circannual (approximately yearly) adaptations in physiology and behaviour exhibited in both prokaryotic and eukaryotic species (5). For animals, these adaptations include moulting of fur and feathers, yearly migrations, fluctuations in sex drive, and annual changes in activity levels (6).

Humans are increasingly disconnected from the harshness of the outer world. We stay warm during winter, illuminate our houses when it gets dark, turn on the air conditioner during the hot days of summer, and have constant access to foods from all over the world regardless of the season. In recent years, we have also become increasingly more *logged on* to social stimuli, even in times of solitude.

Have we disconnected entirely from the seasonal laws that govern other animals' behaviour? Or is it possible that nature has a larger impact on us than we realise, despite our attempts to distance ourselves from it? These are questions I have been investigating in this Master's Thesis, and I will share my findings on the following pages.

2 Glossary

Biological Clock	An organism's innate, physiological mechanism of timekeeping.
Circadian rhythm	An endogenous, approximately 24-hour cycle that can be entrained by a zeitgeber with a period in the circadian range of approximately 19 to 28 hours.
Circannual	An endogenous, approximately one-year cycle that can be entrained by a zeitgeber with a period in the circannual range of approximately 8 to 16 months.
Diel	Corresponding to the duration of the 24-hour day.
Entrainment	The synchronisation of a self-sustaining rhythm (i.e., a circadian or circannual rhythm) by a zeitgeber. Under steady entrainment, there is a stable phase relationship between the self-sustaining rhythm and the rhythm of the zeitgeber.
Free-run	The state of a self-sustaining rhythm in the absence of effective zeitgebers.
Photoperiod	The hours of daylight during the 24-hour day.
Photoperiodism	The response of organisms to changes in photoperiod, which allows them to adapt to the environmental challenges of forthcoming seasonal change.
Seasonal	A seasonal organism responds to the yearly fluctuations in environmental conditions but does not necessarily have intrinsic timekeeping mechanisms.
Social Clock	The local clock time which allows us to be in sync with societal expectations and institutions such as school, work, and public transport.
Social jet lag	The chronic discrepancy between an individual's intrinsic Biological Clock and the Social Clock.
Zeitgeber	Any external or environmental cue that entrains or synchronises an organism's biological rhythms.

3 Introduction

3.1 Photoperiod, Photoperiodism and Circannual Rhythms

Photoperiod – the hours of daylight exposure during the 24-hour day – fluctuates seasonally in most of the world (4, 5). As the Earth orbits the Sun, it exhibits an axial tilt of 23.4°, as illustrated in Figure 1. This tilt causes the Northern Hemisphere to be closer to the Sun during the summer solstice and further away from the Sun during the winter solstice. The opposite is true for the Southern Hemisphere (1). Thus, the hemispheres experience "opposite" seasons at any given time.

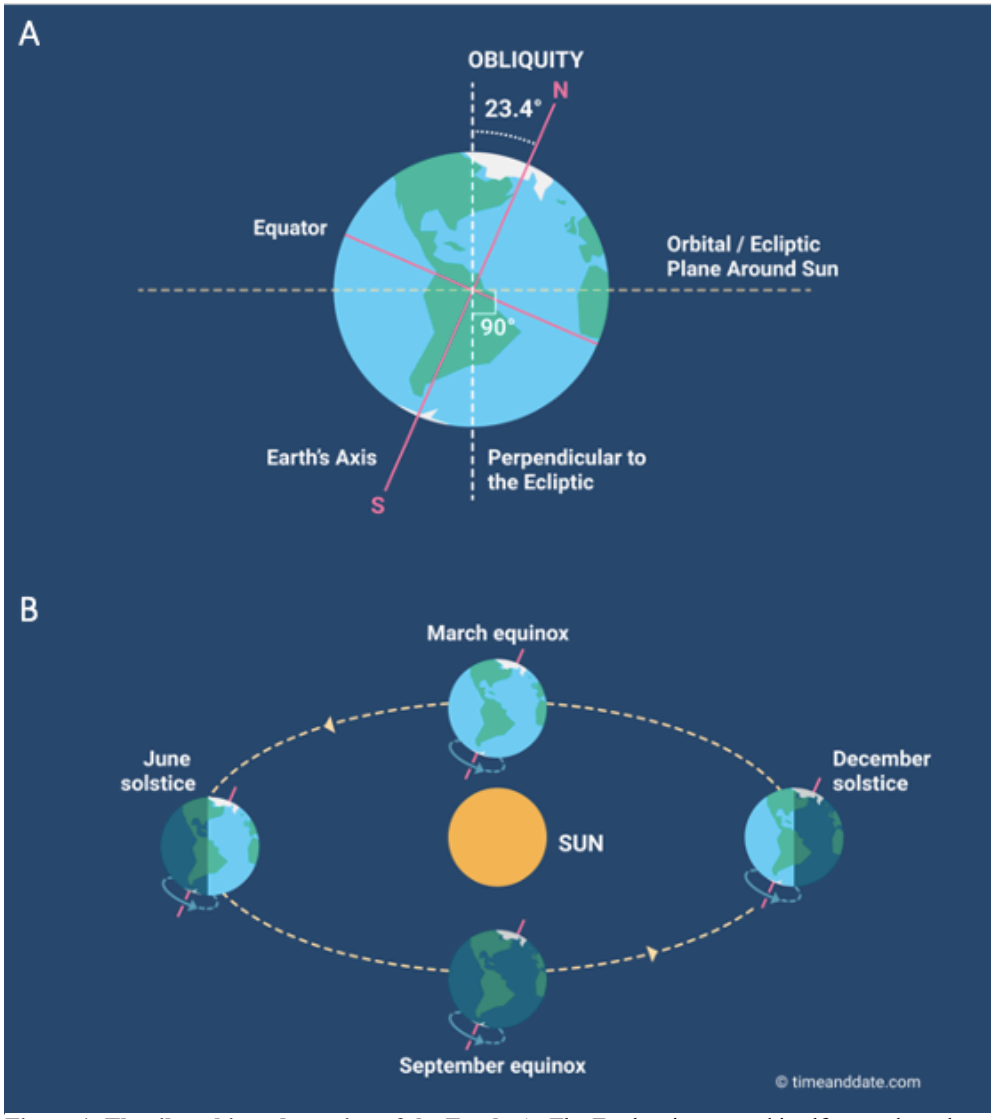


Figure 1: **The tilt, orbit and rotation of the Earth. A:** The Earth spins around itself to produce the 24-hour day/night cycle. This occurs with an axial tilt of 23.4° relative to the Sun. **B**: The Earth orbits the Sun in an ellipse, but this is not the leading cause of the seasons. The axial tilt causes different hemispheres of the Earth to point toward or away from the Sun at different times of the year. The hemisphere closest to the Sun experiences summer, and the hemisphere furthest away from the Sun experiences winter. Only the equator has near-constant yearly light conditions. Figure (slightly modified) from timeanddate.com (1).

During the equinoxes, days and nights are roughly equal in duration across all latitudes. In comparison, nights are shorter, and days are longer during summers in both the Northern and Southern Hemispheres. During winter, nights are longer, and days are shorter. Only at the Equator, located at 0° latitude where the impact of Earth's axial tilt is minimal, there are no noteworthy seasonal changes in day length (6, 7). The further you move away from the Equator, the larger the seasonal photoperiodic changes become. Tropical regions located between the Tropics of Cancer and Capricorn at 23° 27′ north and south, respectively, experience only minor photoperiodic changes, although they can have pronounced dry and wet seasons (8-10). Between the tropical belt and the Arctic Circles are temperate regions characterised by warm summers, cold winters, and intermediate springs and autumns (10, 11). Beyond the Arctic Circles at 66° 33' south and north, photoperiods are the most extreme. These regions experience *polar nights* during winter and *midnight sun* during summer: periods when the sun is below and above the horizon throughout the 24-hour day (12). Solar radiation intensity also varies throughout the year and is strongest in summer and weakest in winter (13). The yearly fluctuations in solar energy input produce seasonal climate patterns such as changes in temperature, humidity, winds, and ocean currents (14). Unlike temperature and weather conditions, however, photoperiod is a mathematically reliable signal of the time of year (7).

For animals in temperate and arctic regions, executing certain life history events at the right time of year is essential. In an evolutionary context, it is easy to see why. An animal that produces offspring during winter, when food is scarce, will probably not create a long lineage of descendants. Similarly, an animal that enters hibernation when food is plentiful or during the mating season will also suffer reduced fitness.

Many organisms have evolved timing mechanisms that allow them to anticipate the future seasonal environment and perform activities such as reproduction, growth, moult, migration, and hibernation at the appropriate time of year (6). Timekeeping comprises three elements, as illustrated in Figure 2. Firstly, sensory systems receive cues from the environment, particularly light cues, which serve as synchronizing signals or *zeitgebers*. Secondly, an innate clock maintains an intrinsic, self-sustaining rhythm in the absence of zeitgebers but can also be entrained to match its rhythm to the environment. Thirdly, output pathways link the internal clock to physiological processes and behavioural responses.

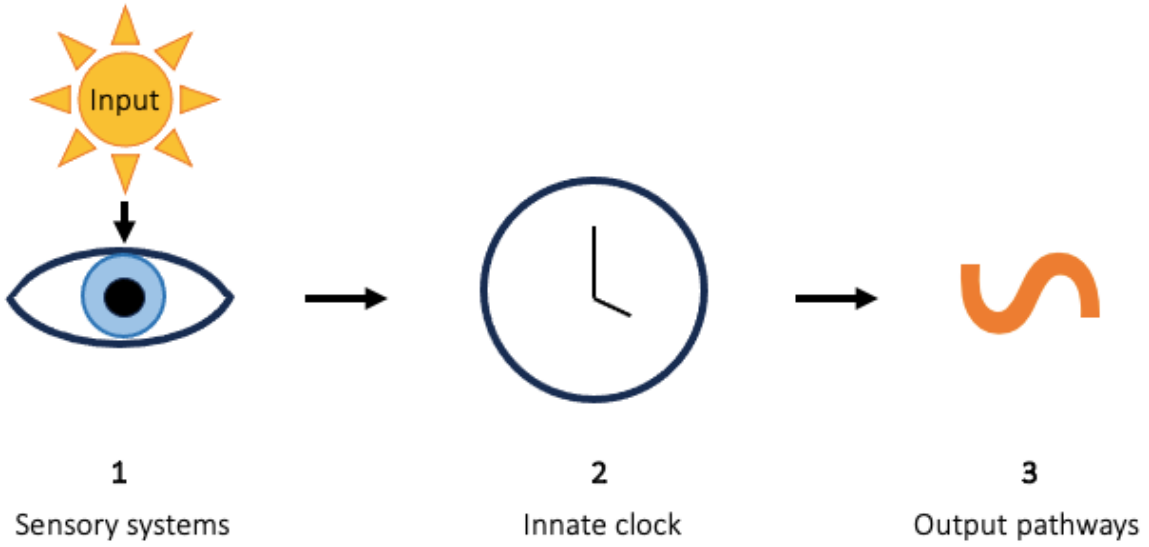


Figure 2: Eskinogram of the three basic components for timekeeping. A clock pathway has three components: a perceived input, an innate clock, and an output. The sensory system perceives the input and relays it to the oscillator (the innate clock), which produces physiological and behavioural responses.

In the context of seasonal timing, the innate clock is called the circannual (*circa* – approximately, *annual* – yearly) clock. Photoperiod is used as a seasonal zeitgeber that entrains the circannual clock, which in turn drives a range of physiological changes. Animals use the absolute day length and the directions of daylight changes to track the time of year and anticipate coming seasons (15). Photoperiodic animals use photoperiod to initiate seasonal programmes of physiological change that can also be stimulated in a lab setting with artificial changes in photoperiod. This is called *photoperiodism* (16). Zeitgebers such as temperature, social stimuli and seasonal food availability appear less important than photoperiod for the entrainment of circannual rhythms (6, Ch. 4).

It is worth emphasising that the term *circannual* should only be used to describe an organism that maintains a self-sustained, approximately yearly rhythm under constant conditions. If no such self-sustained rhythm has been proven, it cannot be known if the organism is indeed circannual or if it is simply *seasonal.* A seasonal organism responds to the yearly fluctuations in environmental conditions but does not necessarily exhibit intrinsic timekeeping.

Similarly, the term *circadian* should only be used to describe an organism that maintains a self-sustained, approximately 24-hour rhythm under constant conditions. When approximately 24-hour rhythms are seen in variable conditions, these rhythms should be

described as *diel.* Circadian rhythms will not be explained thoroughly here, but the interested reader is encouraged to read Chapter 3, Fundamental Properties of Circadian Rhythms, in the book Chronobiology by Dunlap et al. (2004). In short, circadian rhythms are endogenous rhythms with a period of approximately 24 hours (7, Ch. G-7). Although circadian rhythms are self-sustained under constant laboratory conditions, they are entrained by the rhythmic changes in light intensity between night and day. Without an external zeitgeber, circadian periods are usually not precisely 24 hours long. However, the Sun (and for humans, electric light sources) entrains the endogenous daily rhythm to the exact 24-hour day (6, Ch. 1). The circadian clock has been implicated in photoperiodism, as discussed further in section 3.2.

In terms of innate timing, there are two classifications of biological circannual rhythms: Type I and Type II. Type I species have an internal circannual clock that requires external cues to persist for more than one cycle. The cues must be appropriate to their respective season and occur during both short and long days. As an example, many wild rodents use long photoperiods as a signal to reproduce. However, if the long photoperiod continues – instead of progressing into autumn and winter photoperiods – the reproductive state also continues (6, Ch. 4). Also, many Type I animals will eventually become refractory to long-term housing on short photoperiods, and spontaneously exhibit their summer reproductive phenotype. To break this state of refractoriness, the animal needs to be re-exposed to a long photoperiod (15). Type II species have self-sustaining rhythms that persist under constant environmental conditions. They are driven by an internal circannual clock and entrained by changing photoperiod so that the endogenous timekeeping is synchronised with the seasonal cues of the outer world. Type II species have free-running periods of approximately 10-12 months that have been documented to persist as much as seven cycles in mammals. Examples of animals with Type II circannual rhythms include longer-lived species such as squirrels, sheep, deer, and starlings (6, Ch. 4).

3.2 Encoding of Photoperiod by the SCN, the Melatonin Signal and Photoperiodism

The anatomical structures behind the circannual clock have not been conclusively identified in any species (15, 17). However, as mentioned above, there exists some evidence connecting photoperiodic time measurement to the circadian clock, and in recent years, promising suggestions have been forwarded about the anatomical location of the circannual clock.

In mammals, light stimuli are relayed from the eyes via the retinohypothalamic tract to the SCN in the ventral periventricular zone of the anterior hypothalamus (18). The SCN is a paired structure, where each half consists of approximately 10,000 neurons that give rise to circadian rhythms through specific, rhythmic gene expression and by the rate at which they fire action potentials (19).

Electrical activity in the SCN is high during the day and low at night for both diurnal and nocturnal mammals. The light intensities that occur during dawn and dusk, at approximately 50% of the maximum level, trigger, respectively, behavioural onset and offset for diurnal species and the opposite for nocturnal species. Consequently, as dawn and dusk times change with the seasons, so do behavioural activity lengths, as discussed in section 3.3 (20).

The electrical activity patterns of the SCN have been shown to differ between different photoperiods (18). VanderLeest et al. (2007) demonstrated this by keeping freely moving mice at long (LD 16:8) and short (LD: 8:16) photoperiods and recording the multiunit electrical activity of the SCN. They found that the multiunit electrical pattern differed between the different photoperiods because single neurons exhibited alterations in their activity phases relative to each other (see Figure 3) (21). The SCN can thus encode short winter days and long summer days by manipulating the phase relationship among its constituent neurons (18). Notably, there have not been reported any differences in single-neuron activity between long and short days. The seasonal waveform patterns of SCN electrical activity, that result from alterations in the phase relationship between the individual units, is sometimes referred to as *plasticity* of the SCN (5)

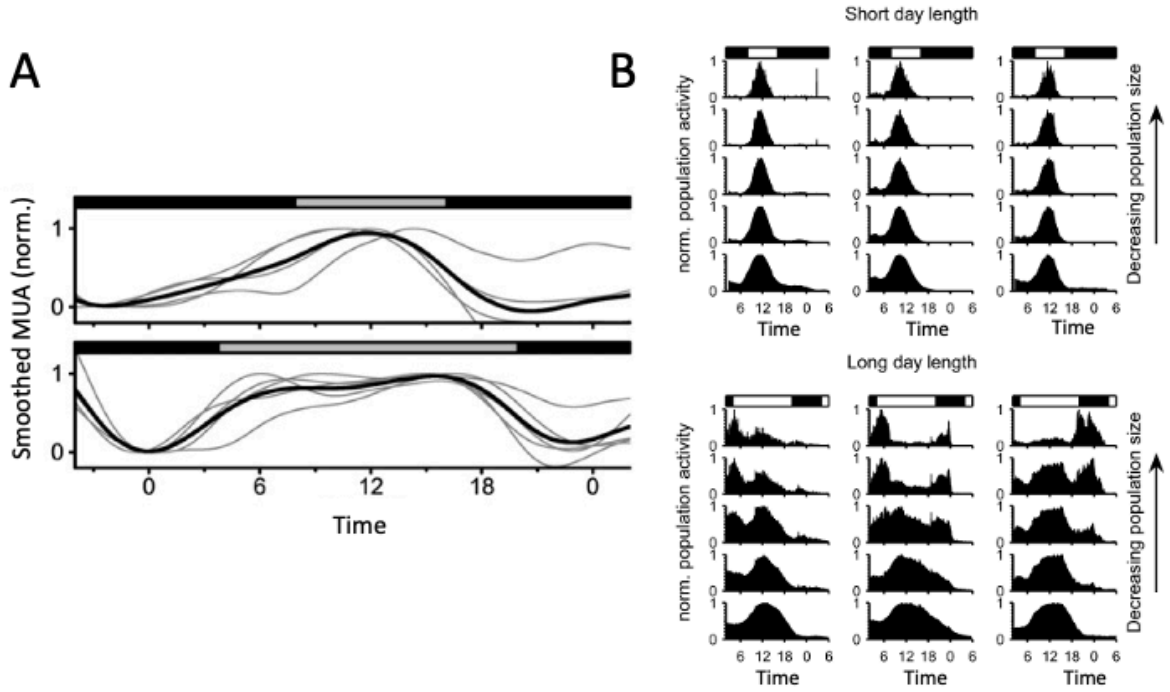


Figure 3: SCN multiunit electrical patterns relative to previous photoperiod exposure. A: Smoothed waveforms of the multiunit electrical activity (MUA) in the SCN of freely moving mice after exposure to a short photoperiod (above) and a long photoperiod (below). Time (x-axis) is in hours. The recordings were performed in constant darkness to assess the properties of the neuronal activity in the absence of light. The top bars indicate the light-dark cycle the animals were exposed to before the experimental recording. The data are normalised from 0 to 1. Black lines are the averaged waveforms, and grey lines are individual recordings. **B**: Normalised cumulative electrical activity for the SCN of mice housed under LD 8:16 (above) and LD 16:8 (below). The bars on the top of each graph represent the light-dark cycle to which the mice were entrained. The spike frequency of action potentials is normalised from 0 to 1. The population size is near single-neuron at the top and increasingly large towards the bottom. The graphs demonstrate that the electrical activity pattern of the SCN is composed of out-of-phase activity exhibited by individual neurons. The phase distribution was larger on long days than on short days. Figures (slightly modified) from VanderLeest et al. (2007) (21).

Several neurotransmitters, such as VIP (Vasoactive Intestinal Peptide), GRP (Gastrin-Releasing Peptide), AVP (Arginine Vasopressin) and GABA (Gamma-Aminobutyric Acid), are involved in the synchronisation between SCN neurons. Therefore, each of these transmitters may play a role in the photoperiodic functions of the SCN (18). For example, VIP-expressing cells in the ventral SCN receive photic information from the retina and project to the dorsal SCN (18). Lucassen et al. (2012) investigated the circadian and photoperiodic functions of VIP-knockout mice by recording the multiunit neural activity of the SCN *in vivo*. Circadian rhythms were largely unaltered, but photoperiodic adaptations of the SCN were abolished. They concluded that VIP must be indispensable for photoperiodic encoding (22).

In mammals, nightly melatonin secretion from the pineal gland is regulated by the SCN. Melatonin levels in blood and saliva, or the metabolite 6-sulfatoxymelatoninin in urine, are

frequently used as a marker for the phase of the circadian clock (23). Interestingly, levels are high during the dark phase for both diurnal and nocturnal mammals (17, 18).

In the context of photoperiodism, the duration of melatonin production reflects the length of the night for many species of mammals (24). For these animals, photoperiodic information is encoded in the daily melatonin signal (see schematic illustration in Figure 4), and pinealectomy abolishes seasonal physiological responses to photoperiod (18, 25). Melatonin has also been implicated in the entrainment of circannual rhythms. Woodfill et al. (1994) gave melatonin infusions to pinealectomised sheep and found that 90 consecutive days were sufficient to entrain their circannual reproductive cycle (26). Therefore, identifying the site of action of melatonin may indicate where seasonal physiological responses are controlled and coordinated.

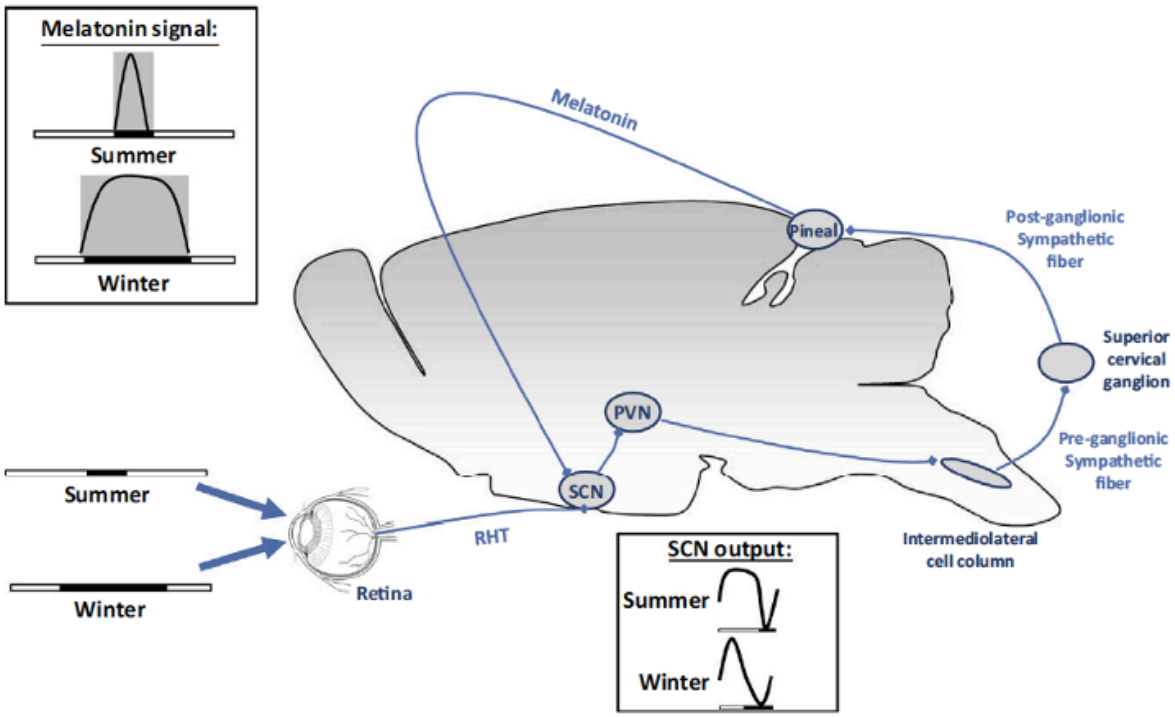


Figure 4: The pathway of seasonal melatonin secretion in a mammalian brain. Photic information is received by the retina and conveyed to the SCN via the retinohypothalamic tract (RHT). The SCN displays a seasonal pattern of electrical activity depending on the duration of light stimuli, which varies with the seasons. This photoperiodic information is transmitted from the SCN to the pineal gland through a polysynaptic pathway that goes through the paraventricular nucleus (PVN), the intermediolateral cell column of the thoracic spinal cord, the superior cervical ganglion and the postganglionic adrenergic fibres that innervate the pineal gland. During summer nights, the nocturnal melatonin signal is temporally compressed, and during winter nights, it is prolonged. Figure (slightly modified) from Coomans et al. (2014) (18).

The pars tuberalis of the pituitary gland has the highest melatonin receptor density of all tissues; specifically, MT1, one of two subtypes of melatonin receptors, is strongly expressed in the pars tuberalis (17, 18). There is a well-established photoperiodic circuit from the retina through the SCN to the pineal gland (see Figure 4), where melatonin production occurs, and to the pars tuberalis. The pars tuberalis uses the melatonin signal from the pineal gland to measure photoperiod duration. Since the extent of melatonin secretion depends on the duration of the night, photoperiod indirectly serves as a seasonal cue for the pars tuberalis. Consequently, the thyroid hormone (TH) metabolism in the hypothalamus is altered in a day-length-dependent manner. This is achieved by altering the TSH (thyroid stimulating hormone) secretion from the pars tuberalis. Specifically, long photoperiods activate TSH through a pathway involving EYA3, an enzyme that stimulates TSH production by regulating the expression of the TSHβ subunit. EYA3 is itself regulated by circadian clock genes (activated by BMAL2 and suppressed by DEC1) (27, 28). TSH is a glycoprotein hormone that consists of the two subunits αGSU (glycoprotein subunit α) and TSHβ (29). Under long photoperiods, TSHβ is strongly co-expressed with αGSU-expressing cells in the pars tuberalis but not under short photoperiods. Thus, on long photoperiods, TSHβ and αGSU are combined to produce TSH, which is secreted in a retrograde manner back up into the hypothalamus (30). Specialised hypothalamic glial cells lining the 3rd ventricle, called tanycytes, have TSH receptors whose binding stimulates the expression of deiodinase enzyme 2 (DIO2). DIO2 converts locally available thyroxine (T4) to the active form of thyroid hormone, triiodothyronine (T3), and this change in thyroid hormone levels drives the seasonal metabolic and reproductive changes in physiology (17). The pars tuberalis has also been suggested to be the site of the circannual clock on the basis that it integrates the photoperiod signal and spontaneously reverts in expression profile when held on constant photoperiods (30). Also, when experiments to disconnect the pars tuberalis from the hypothalamus were conducted under constant light conditions, the seasonal cycles of prolactin, secreted from the downstream anterior pituitary, continued (31).

3.3 Seasonal Changes in Diel Activity of Temperate and Arctic Animals

Seasonality refers to a seasonally fluctuating pattern that may not depend on an innate timer. Here I will discuss seasonal changes in diel activity which may be related to the SCN (see

section 3.2). Since the main zeitgeber, day length, varies with season and latitude, so do the activity patterns of many animals. Daan et al. (1974) investigated seasonality in diel activity levels in three species of mammals and five species of birds kept either at the Arctic Circle or in Germany. The article does not state whether all the animals were captured at the same latitude; however, two of the mammalian species, Golden Hamster (*Mesocricetus auratus*) and Southern Flying Squirrel (*Glaucomys volans*), are not native to Europe and presumably had *one* geographical origin. All the species demonstrated clear seasonal patterns of daily activity levels, with less activity in winter than in summer, although birds had the most marked seasonal shifts (see Figure 5). For the flying squirrel, which is nocturnal, activity bouts were longer during winter (Figure 5B). Also, daily activity onset and offset times appeared to vary with latitude, presumably because photoperiod varies with latitude (Figure 5C). This indicated that seasonal and latitudinal differences in sunrise and sunset influenced the activity length in a roughly mirror-image manner for these animals (32).

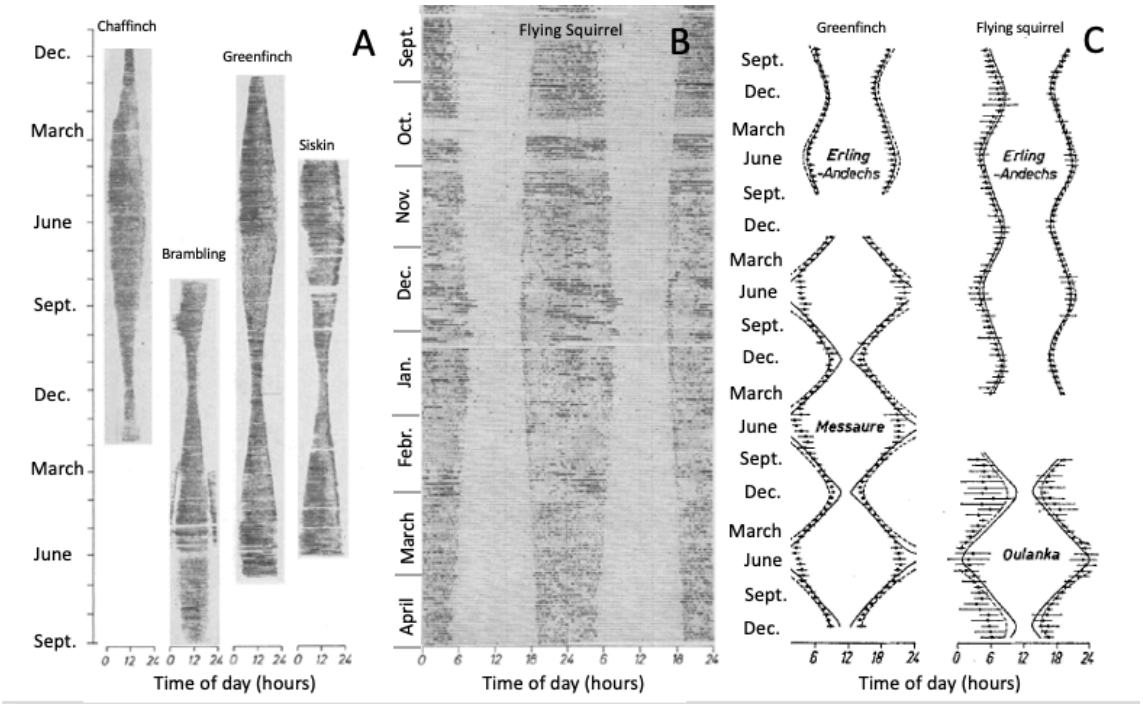


Figure 5: **Seasonal and latitudinal differences in activity patterns of birds and mammals. A**: Seasonal fluctuations in activity for four species of birds (Chaffinch, Brambling, Greenfinch and Siskin) and **B:** Southern Flying Squirrel. The birds were kept in Miessaure, Sweden, and the flying squirrel was held at Erling-Andechs, Germany. **C**: Daily onsets and offsets of activity in a greenfinch (left) and a flying squirrel (right) in Germany (top) and at the Arctic Circle (bottom). The solid lines represent sunrise and sunset, and the horizontal lines represent activity start and cease with two standard deviations on each side of the mean. Figure (slightly modified) from Daan et al. (1974) (33).

Midnight sun and polar night constitute strange deviations from the normal, daily light-dark cycle that governs most of the world most of the time. Animals that live in arctic regions must deal with the relatively continuous signals of light or darkness. However, even during the summer and winter solstices in arctic regions, variations in solar elevation relative to the horizon produce daily cycles of light intensity that might be used by some animals to keep track of time (16). Diel behavioural patterns can also become essentially continuous during polar summer and winter. For example, the diurnal feeding patterns of Svalbard ptarmigan (*Lagopus muta hyperborea*) are intermittently continuous throughout the 24-hour day around the solstices in their natural arctic habitat (34). It has been proposed that the Svalbard ptarmigan possesses a dampened circadian clock that enables non-diel rhythms to take precedence around the solstices (16). Svalbard reindeer also have a breakdown of diel rhythms around the solstices (35). For reindeer, however, the molecular circadian clock appears to be weakened or completely absent, although these animals have strong circannual rhythms (36, 37).

3.4 Photoperiodism, Circannual Clocks and Seasonality in Humans

Bronson (2004) speculated that photoperiodism (specifically, seasonal regulation of reproduction) probably evolved in a common ancestor of all modern mammals (38). Dunlap (2004) hypothesised that photoperiodic time measurement is an ancient property of all vertebrates (6, Ch. 4). Moreover, seasonal breeding is common in non-human primates living at the higher latitudes of the tropics or in the lower temperate zone. Thus, it seems likely that modern humans would possess the genetic potential for photoperiodism and a circannual clock (38).

If humans are photoperiodic, there could potentially also be genetic differences in the trait. Photoperiodism is strongly responsive to selection, as demonstrated by Heideman et al. (1999) on white-footed mice (*Peromyscus leucopus*) (see Figure 6). In an artificial selection experiment, they produced one line of highly photoresponsive mice and another line of remarkably unresponsive mice in only three generations of breeding. At the end of the experiment, 80% of the mice in the photoresponsive-selected group responded strongly to photoperiod, and only 16% in the unresponsive-selected group did the same (39). Based on this, Bronson (2004) forwarded the hypothesis that some humans may be more

photoresponsive than others and that such individual variation may explain the inconsistencies in seasonal research on humans (38). It has been suggested that *Out-of-Africa* migration exposed humans to higher latitudes that might have acted as a selection pressure on genes involved in photoperiodism. Forni et al. (2014) investigated the correlation between yearly changes in photoperiod (Δphotoperiod) and a range of circadian genes from 52 human populations worldwide. There was a significant correlation between Δphotoperiod and single nucleotide polymorphisms (SNPs) in the investigated genes of people living at the respective latitudes, and thus signatures of latitude-driven selection on circadian genes (40). Several species exhibit varying degrees of seasonal responses depending on the latitude at which they live. For some species, such differences between populations may be the result of genetic adaptation to distinct environments. For example, Lynch et al. (1981) studied white-footed mice (*Peromyscus leucopus*) from three latitudinally distinct populations in the United States under identical laboratory conditions. They found that only mice from the southernmost population were unresponsive to a short-day photoperiod (41).

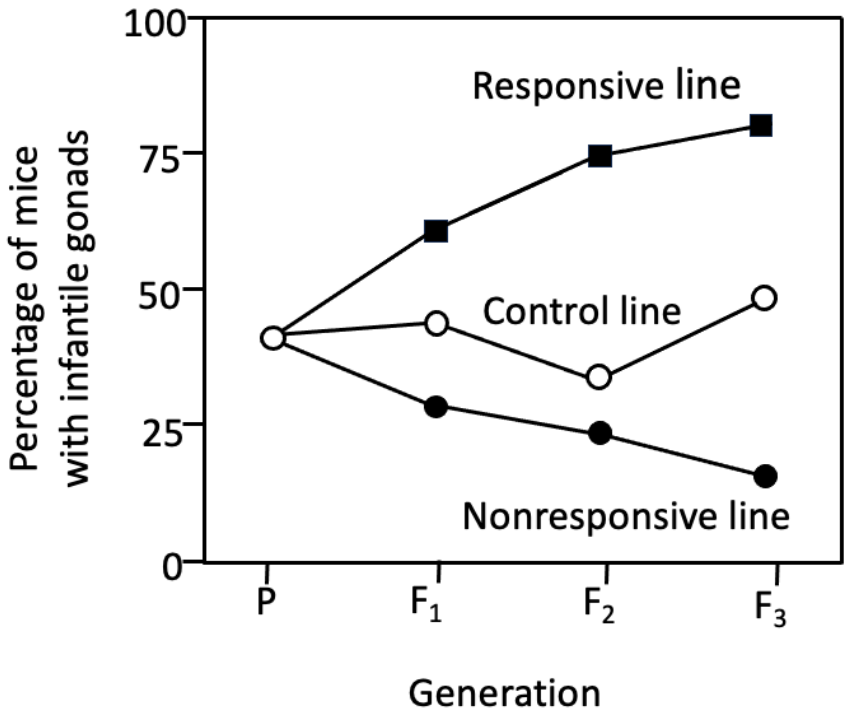


Figure 6: Selection experiment on photoperiodism. The percentage of white-footed mice strongly responsive to photoperiod in each generation (P, F_1, F_2, F_3), measured as having regressed gonads during a short photoperiod. Squares represent the responsive-selected line, solid circles represent the unresponsive-selected line, and open circles are the control line. Figure redrawn from Heideman et al. (1999).

However, it is difficult to say whether we will ever know if humans are photoperiodic or have a circannual clock, primarily due to the inherent challenges in conducting long-term, controlled studies on human subjects (section 3.4.3). Therefore, we are left with evaluating the evidence for seasonality in humans, which remains a subject of debate. Some researchers attribute human seasonality to fluctuations in temperature rather than photoperiod, while other studies do not detect any seasonality at all (42, 43). Modern humans are largely protected from seasonal fluctuations in both temperature and photoperiod, making the topic even more difficult to investigate. Despite somewhat inconsistent evidence, seasons appear to have a notable influence on human behaviour and physiology. However, the mechanisms behind human seasonality are not certain – we could have an endogenous timer or simply respond to seasonal changes (44).

Although the cause remains uncertain, seasonality has been demonstrated in various aspects of human behaviour and physiology, such as disease susceptibility and immunological responses, mental health, various hormone levels, and body weight (42, 44-47). For example, *seasonal affective disorder* (SAD) is a well-documented phenomenon, with recurrent episodes of depression or bipolar disorder that only occur during a specific time of year – normally during winter (48). Another example is the seasonal susceptibility to infectious diseases, such as influenza in the winter or chickenpox in the spring (49). The mechanisms behind this phenomenon are not fully understood; however, suggestive explanations include seasonal variations in environmental conditions that impact the host's immune response and seasonal fluctuations in contact rates (e.g., school terms). Another suggestion is that endogenous circannual restructurings of immunity, metabolism or body condition could play a role, but this remains highly speculative due to a lack of evidence (49). The following sections will focus on two of the most studied seasonally fluctuating aspects of human biology: reproduction and the sleep-wake cycle.

3.4.1 Evidence for Seasonal Changes in Diel Activity, Sleep and Melatonin Secretion in Humans

The studies that found seasonal differences in sleep duration typically had two distinguishing features:

1. The studies were conducted at extreme latitudes with large annual fluctuations in daylight.
2. Or the study removed voluntary control over the modern light environment. For example, in a lab or camping.

However, even under extreme photoperiods, seasonal fluctuations in the sleep-wake cycle are not always detected. Antarctic studies often occur at large research stations with set working hours and mealtimes. Only rarely are the workers at antarctic research stations encouraged to sleep and wake up according to their preferences, but this was the case in a study by Kennaway et al. (1991), where four workers at a small antarctic research station showed free-running rhythms of sleep, melatonin production, cortisol levels and electrolyte excretion during winter (see Figure 7) (50). Similarly, Usui et al. (2000) found that eight men on an antarctic research expedition struggled to entrain to the 24-hour period during winter (51). Contrary to this, Yoneyama et al. (1999) saw no changes in the sleep-wake cycle but did detect phase shifts in melatonin and temperature rhythms, with a clear delay occurring during winter (52). Arendt et al. (2017) also found no change in the duration of melatonin release but a marked phase delay in winter compared to summer (53).

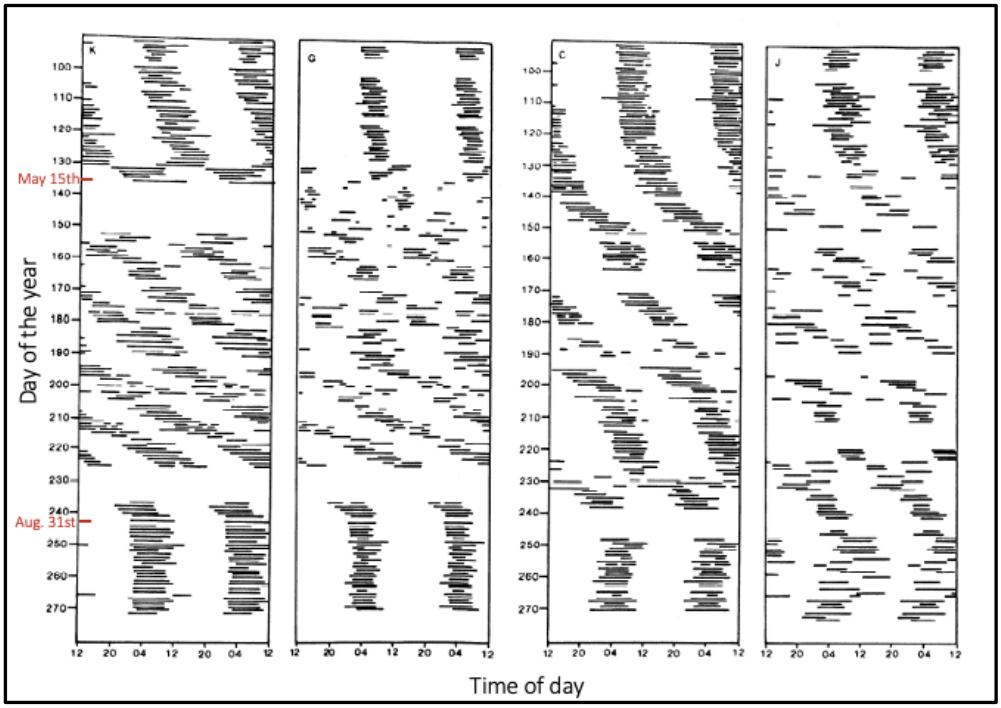


Figure 7: **Free-running sleep-wake rhythms of humans in Antarctica**. The four multi-plotted graphs represent the sleep-wake rhythms of four human subjects on an Antarctic research station in 1987. Sleep times, registered by sleep diaries, are plotted as black lines. The participants were free to sleep, wake up and work at their preferred times. During the Antarctic winter, the sleep-wake rhythms of all participants free-ran. The gaps in the records occurred during field trips when the participants did not keep a sleep diary. Figure (slightly modified) from Kennaway et al. (1991) (50).

There have been limited long-lasting investigations into the human sleep-wake cycle in experimental settings, with a few interesting exceptions. Wehr et al. (1991) exposed eight healthy people to a summer photoperiod of LD 16:8 for one week and a winter photoperiod of LD 10:14 for 4 weeks under laboratory conditions. The subjects slept longer ("winter": 11.0 ± 0.8, "summer": 7.7 ± 0.2) and showed longer nocturnal melatonin secretion ("winter": 12.5 ± 1.8, "summer": 10.3 ± 0.8 h) under the short photoperiod (54). Stothard et al. (2016) demonstrated that total sleep duration was longer during winter camping (9.9 ± 0.4 hours) compared to summer camping (8.0 ± 1.1) in Colorado (USA). They also showed that the average sleep duration was ~2.3 hours longer during winter camping than in participants' everyday lives and that melatonin onset occurred ~2.6 hours earlier, although there were no differences in the midpoint and offset of the melatonin rhythm (55, incl. S3). However, although the results of these two studies might be suggestive of seasonal changes in the sleep-wake patterns of humans, multi-year monitoring would be necessary to conclusively demonstrate such patterns.

There is no consensus on whether human melatonin secretion varies between the seasons, as the experimental evidence is mixed (24, 54). Many studies on humans appear to encounter a phase delay in the melatonin rhythm during winter, even in countries that do not use daylight saving time, such as Japan (13).

Other studies have compared natural light exposure to the everyday electrical environments of their subjects. During a German reality TV program with participants living under Stone Age conditions for two months during summer, sleep-wake rhythms were recorded before, after and during filming. The participants slept approximately 1.5 hours longer when living under Stone Age conditions than at home (56). Studies that compare natural light exposure to modern electrical lighting are interesting in the context of seasonal rhythms because our modern homes, in many aspects, function as a "constant" summer photoperiod. The duration of the biological night is similar between natural summer photoperiods and modern indoor environments, but there is a phase delay in the latter (55).

3.4.1.1 Chronotypes

People's phase of entrainment to the same light-dark cycle varies – a characteristic called *chronotype*. Colloquially, some extreme chronotype variants may be termed *larks* and *owls*,

but there are many intermediates between these two (57). Importantly, chronotypes appear to be approximately normally distributed in the population (58).

Genetic differences in the protein components of the circadian clock might be the reason for this inter-individual variation. Such genetic differences could lead to variations in how people´s clocks respond to light and darkness and in how long their subjective days are. If a circadian clock generates daily rhythms that are somewhat shorter than 24 hours, the internal day needs to be either delayed or extended to match the environmental day. Similarly, if a circadian clock produces internal days slightly longer than 24 hours, then it must be advanced or shortened to remain entrained to the external day.

Chronotype assessment (for example, with the Munich Chronotype Questionnaire, MCTQ) is usually done during work-free days to minimise the disagreement between the *biological* and the *social clock* (i.e. local time) (57). A person's chronotype is more of a state than a stable trait because it reflects the phase of entrainment, which can change with varying zeitgebers (57). Thus, in the Stone Age reality show, the participants changed their chronotypes between home and Stone Age conditions. Chronotypes also change with age – teenagers have the latest chronotypes – and vary with sex – men are generally later chronotypes than women before the age of 40 and earlier after (58).

Longitude has also been demonstrated to influence chronotype. Randler (2008) showed that pupils in Western Germany are later chronotypes than pupils in Eastern Germany. This can be explained by equal *social clocks* but different *sun clocks*: School start times are similar all throughout Germany, but sunrise is earlier in Eastern Germany. Thus, East German pupils had earlier rise and bedtimes, and Western German pupils were at a higher risk of sleep deprivation and poor learning outcomes (59).

Roenneberg and colleagues investigated people’s sleep logs and saw that many sleep-wake cycles shifted during the weekends as if the subjects flew several time zones to the West. This shift was not caused by travel, however, but by a mismatch between the *social clock* and the *biological clock.* The phenomenon was named *social jet lag.* The researchers proposed that electric lighting has weakened the natural photoperiodic zeitgeber and that humans now live under roughly constant light conditions except during sleep. This has led to delayed chronotypes for almost all people, and a chronic discrepancy has arisen between the *biological* and *social clock,* which is compensated for during weekends (57). Different

chronotypes and *social jet lag* should be considered when investigating activity patterns in humans since they could potentially obscure their detection.

3.4.2 Seasonality in Human Reproduction

Seasonal reproduction is seen in almost all animals that do not live in tropical areas, and the reproduction time is usually set to deliver offspring between spring and early summer (60). Humans also have clear seasonal patterns in birth rates, although they vary between countries, continents, and latitudes (44, 61). In Europe, conceptions most often occur in spring or summer, depending on the latitude (38, 62). In Sweden, August has historically been the month with the most conceptions, resulting in more births occurring in April (see Figure 8) (62, 63). Although the amplitude of this pattern has declined in recent decades, there is still a peak of births in spring and a trough in the final quarter of the year (63). The declining trend of seasonal births has been observed in many countries, for example, Spain, as illustrated in Figure 9. Interestingly, after a Spanish industrialisation campaign in the 1960s, the rhythm of annual births exhibited a phase shift, and the peak moved from spring to autumn/winter (64).

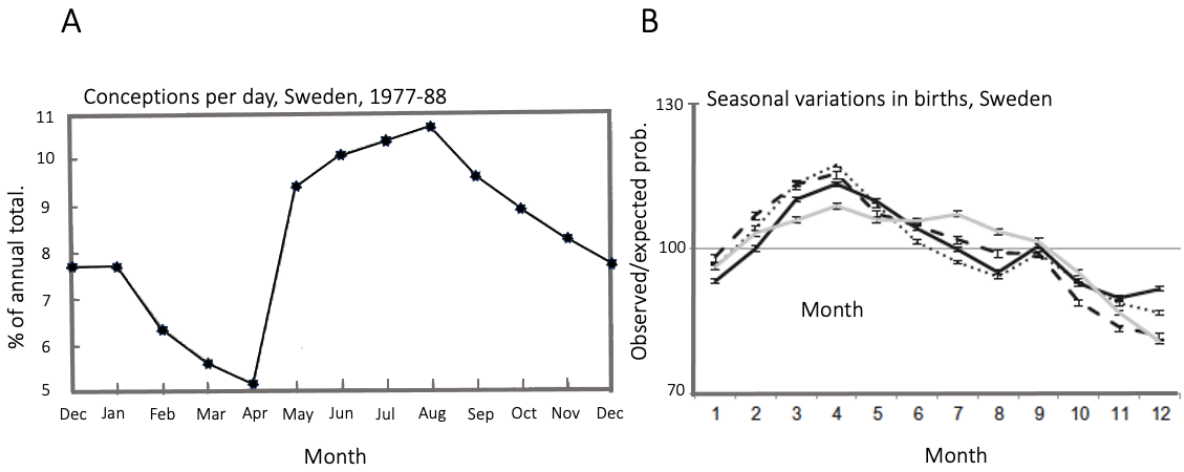


Figure 8: Seasonality in conceptions and births in Sweden. A: The number of conceptions in Sweden between 1977 and 1988, stated as the percentage of the annual total. **B:** Observed vs. expected number of births in Sweden for each month. Solid line: 1940-1959, small-dotted line: 1960-1979, dotted line: 1980-1999, grey line: 2000-2012. Error bars indicate the 95% confidence interval. Figure A is from Moos et al. (1994), B is from Dahlberg et al. (2018). Both graphs have been slightly modified.

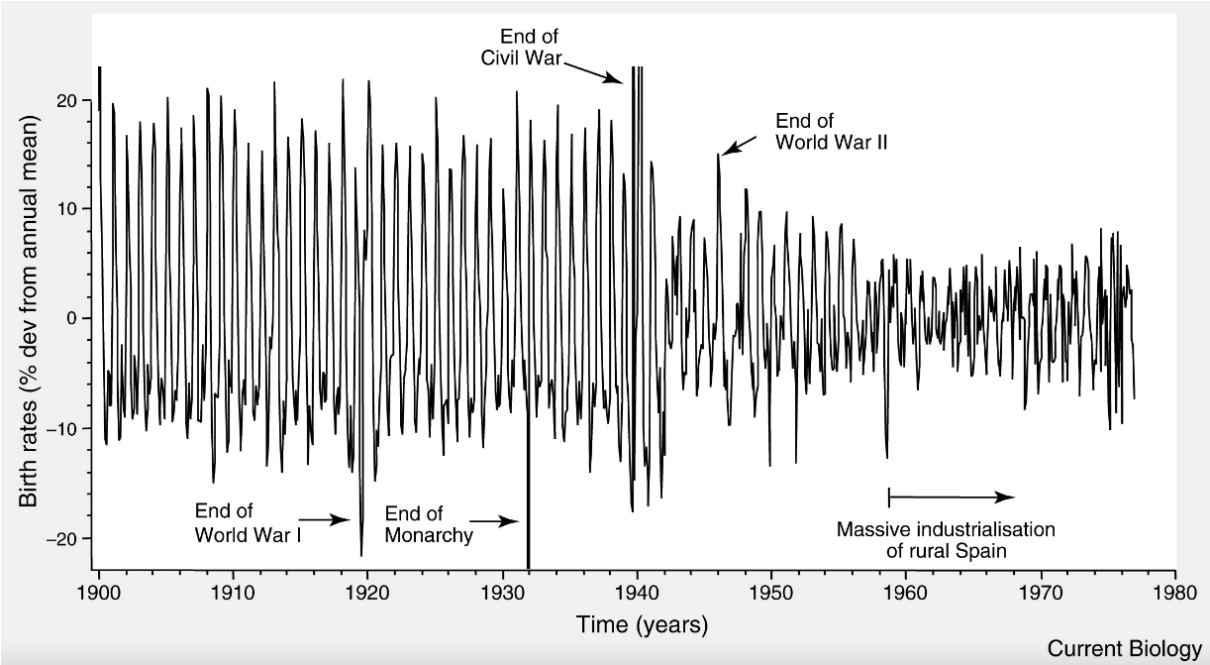


Figure 9: Monthly birth rates in Spain from 1900 to 1978. Before 1940, the rhythm was highly regular, with peak conception levels occurring in spring and annual differences of about 30% between the peaks and troughs. During and after World War 2, the amplitude of the rhythm decreased. After the 1960s, when a massive industrialisation campaign was launched by Francisco Franco, the phase of conception peaks shifted to autumn/winter. Figure from Roenneberg (2004) (64).

Although there is a seasonal pattern in birth rates, there is no doubt that humans are born all year round in all parts of the world. Foster and Roenneberg (2008) suggested this is due to the long-lasting nurturing of human foetuses and babies. Both pregnancy and lactation are highly energy-demanding processes, and in humans, they are spread over such a lengthy period that there may not be any *one* optimal season to give birth (44). Bronson (2004), on the other hand, stated that the most energetically demanding stages of reproduction are late pregnancy, birth, and lactation. Thus, most births have historically occurred in spring and summer to align this energy-depleting phase with the highest food availability (38).

Some researchers have investigated whether seasonality in birth rates is a consequence of seasonality in libido, but the results are mixed, and the studies are few. Demir et al. (2016) evaluated testosterone, luteinising hormone (LH), follicle-stimulating hormone (FSH) and prolactin levels during winter and summer for 80 Turkish men. They also provided the men with a questionnaire to determine the frequency of sexual thoughts and ejaculations during the two seasons. They found a significant difference in FSH and testosterone levels between seasons but no significant changes in LH and prolactin. Also, the men had more sexual thoughts and ejaculations during summer (65).

Contrary to this, Caniklioğlu et al. (2021) investigated hormone levels and morning erections of 221 Turkish men and found no differences between summer and winter. However, the definitions of the seasons were very liberal, with everything between September 24th and March 21st counting as *winter* and all the remaining year as *summer (43).* Svartberg et al. (2003) investigated total and free testosterone levels in 1548 men living in Northern Norway, an area with large seasonal variations in temperature and photoperiod. Total testosterone had a prominent peak in October and November, while free testosterone peaked in December (see Figure 10). The variations in free testosterone were large, with a 31% difference between the lowest and highest mean monthly levels (66). I have not succeeded in finding any studies on seasonality in female libido, perhaps because the female ovulatory cycle makes seasonal patterns harder to detect.

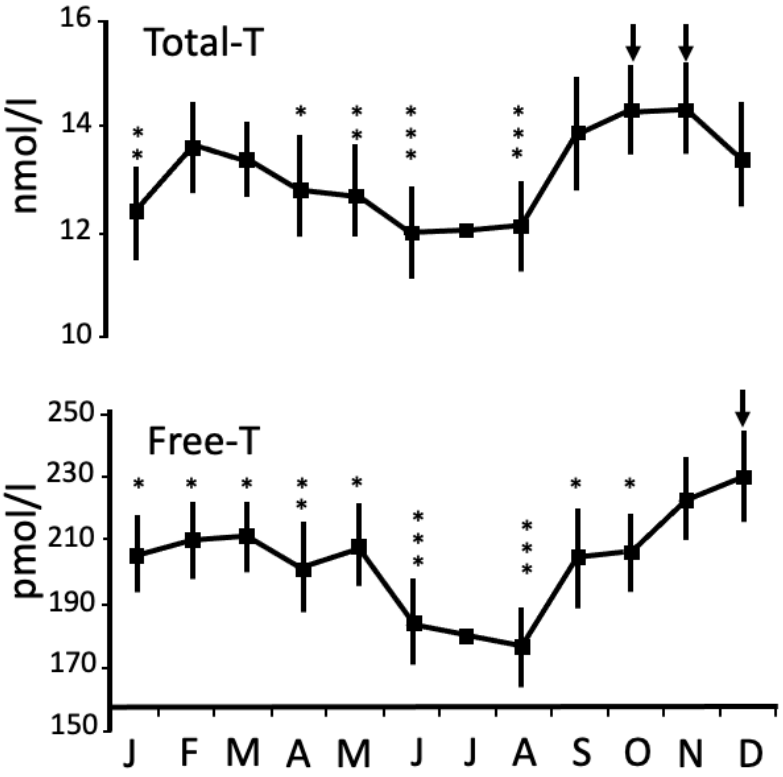


Figure 10: Seasonal variations in total and free testosterone in men from Northern Norway. Mean serum concentration of total testosterone (Total-T) (nmol/litre) and free testosterone (Free-T), including 95% confidence intervals. Arrows indicate peak values, and stars indicate significant differences from the peak (* is P< 0.05, ** is P < 0.01, *** is P < 0.001). Figure redrawn from Svartberg et al. (2003) (66).

3.4.3 Difficulties and Limitations of Studying Seasonality in Human Activity Levels

Studies on seasonality in humans would ideally be carried out over several years in constant laboratory conditions, but for obvious ethical and practical reasons, this is not really an option. Thus, researchers opt for alternative approaches, such as the ones described in section 3.4.1, but the downside to these methods is that they are temporary and can only provide suggestive evidence for seasonal rhythms. Multi-year observational studies can be executed, but there are several potential caveats, such as self-reporting, which is unreliable. The experimental participants can forget to monitor their behaviour, and they can remember wrong when filling out questionnaires.

Several other potentially confounding factors have already been mentioned, such as our modern lifestyles with little exposure to natural light or any other seasonally fluctuating factors, such as temperature and food availability. Human lifestyles are also highly variable, with, for example, different mealtimes and levels of screen exposure. Add to this the commonly found *social jet lag* and the variations in chronotype, and it is easy to see why observational studies can come short in investigating such complex systems.

3.5 Exploring Human Behaviour through Phone or Internet Usage

Indirect measures of human activity through modern technological tools might be the most reliable measure of human activity in their natural environments. Such methods avoid the risks associated with self-reporting.

3.5.1 Tappigraphy

Many studies have used actigraphy (monitoring of data generated by movement) to estimate the activity levels of various animals. However, such analyses typically overestimate rest periods in humans since we can remain cognitively active while performing few body movements. Borger et al. (2019) found that *tappigraphy* from smartphone interactions was better able to capture the total range of human activity levels, also during wakeful rest, than traditional actigraphy. Touchscreen interactions were suggested as a new approach to measure

sleep since most participants used their phones around sleep onset and offset (67). Similarly, a study by Massar et al. (2021) found a strong agreement in the estimated bed and wake times between tappigraphy and a sleep tracker (Oura Ring) (68).

3.5.2 Google Trends

Research of online queries can provide valuable insight into human behaviour. Google Trends is an online tool for investigating Google searches around the world. It allows you to select a location, a search term (or several), and a timeline and provides a graph showing interest over time for your specific query. The data used to make the graph is based on a representative selection of Google searches matching your requirements. There has been a surge of studies using tools such as Google Trends in the last two decades. A search for *Google Trends* on *Web of Science* renders over 8,000 articles, demonstrating the popularity of the resource. The research topics of the published papers include computer science, ecology, engineering, and public health, among others.

A review article on Google Trends health-related research by Mavragani et al. (2018) revealed that most studies used time series analyses and that seasonality was examined in 23.1% (69). One such study was conducted by Ayers et al. (2013), who elegantly demonstrated that mental health-related queries (ADHD, anxiety, bipolar, depression, anorexia, bulimia, OCD, schizophrenia, and suicide) oscillated annually with regular winter peaks and summer troughs between 2006 and 2010 (see Figure 11). This study also used data from locations on opposite hemispheres to reveal an antiphase pattern in the seasonal search activity (46). Such an inversion of the yearly rhythm between hemispheres is commonly interpreted as a sign of seasonality since the seasons are opposite at any time above and below the Equator. Similarly, Bakker et al. (2016) demonstrated that seasonal peaks in Google searches for “chicken pox” varied with latitude and occurred at opposite times of year in the Northern and Southern Hemispheres (70).

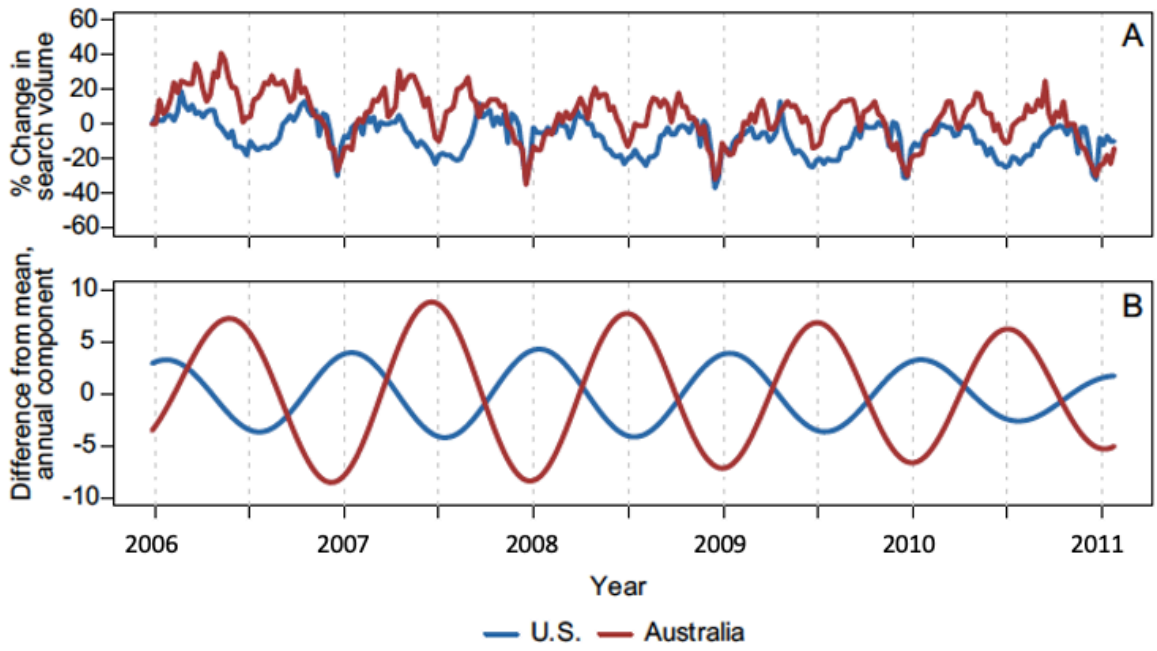


Figure 11: Seasonal trends in mental health searches on Google. Mental health searches (ADHD, anxiety, bipolar, depression, anorexia, bulimia, OCD, schizophrenia, and suicide) oscillated annually with winter peaks and summer troughs between 2006 and 2010. Figure from Ayers et al. (2013).

Google Trends has also been used to investigate libido. For instance, Zattoni et al. (2020) looked at pornography habits during the Covid lockdown (71).

Weekly and annual sleep patterns were investigated by Leypunskiy et al. (2018), who used Twitter activity records to investigate *social jet lag*. Although this study did not use Google Trends, the Twitter data had many similarities with the former. Among other findings, they reported more social jetlag during February-March and September-October than in July-August (see Figure 12). According to their analyses, the timing of Twitter use was more affected by social cues than by day length, although there were some tendencies of dawn-tracking in winter and dusk-tracking in the summer (72).

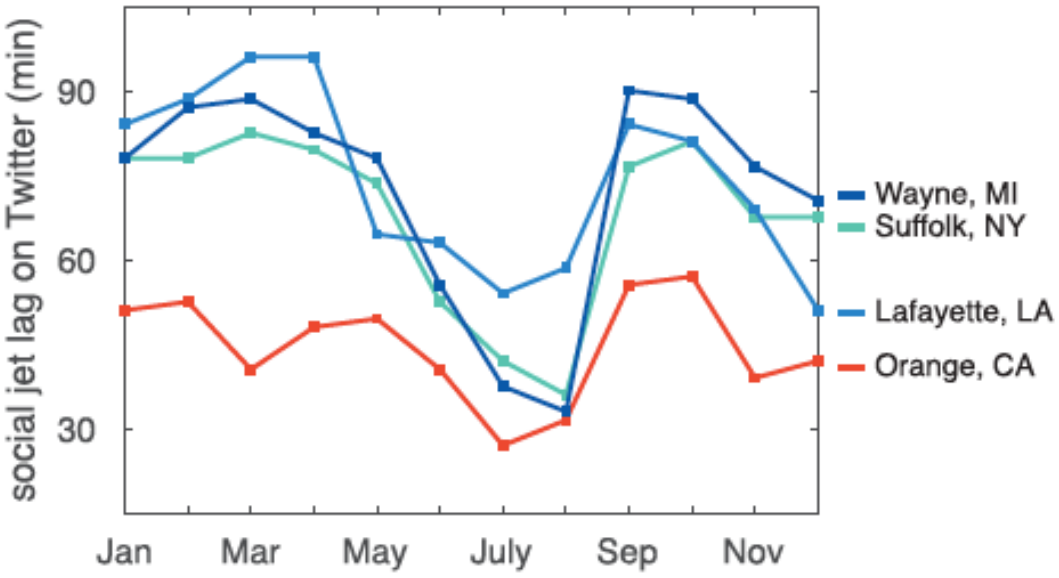


Figure 12: Social jet lag investigated through Twitter use. Seasonal changes in Twitter social jet lag in 2012–2013 in four US counties. Figure from Leypunskiy et al. (2018).

3.6 The Weibull Distribution

As seen in Figure 11 (top part), data based on many people's searches can be somewhat noisy, which can make it challenging to uncover the underlying patterns. Curve-fitting can be useful to smoothen out noisy curves, but it is not necessarily obvious which model to utilise.

User behaviour on websites tends to exhibit *negative ageing,* where the chance of the user leaving the website increases with time. When people browse the web, they rarely dive deeply into every webpage. Instead, they quickly look at a page to decide if it is worth their time. This initial look is like a quick "screening" process, and if they find something interesting, they will examine the page in more detail. This behavioural pattern is called "screen-and-glean," and Liu et al. (2010) recommended the Weibull Distribution for modelling this behaviour (73).

The Weibull Distribution is named after the Swedish mathematician Ernst Hjalmar Waloddi Weibull (1887-1979) and is often used for modelling survival and error analyses, such as the duration of marriage and product lifespans. The probability density function f(x) of the Weibull function is:

$$f(x) = \begin{cases} \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-\left(\frac{x}{\lambda}\right)^{k}} & if\ x \geq 0 \\ 0 & if\ x < 0 \end{cases}$$

Where k > 0 is a shape parameter and λ is a scale parameter. x is the time before an error occurs, and the error rate decreases with time for k < 1 (see Figure 13) (74).

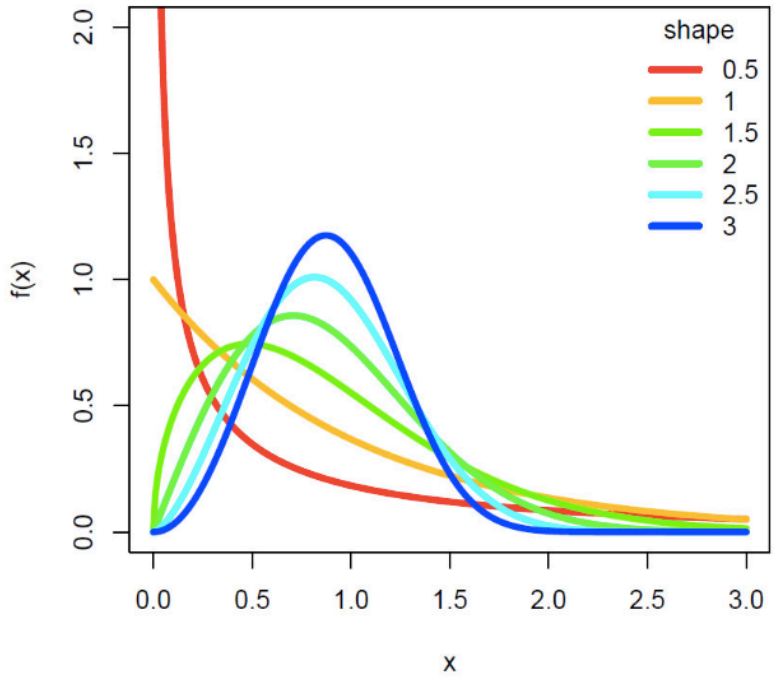


Figure 13: Illustration of the Weibull distribution. The probability density function for the Weibull Distribution with scale parameter $\lambda = 1$ and shape parameter (k) from 0.5 to 3. Figure from the Department of Biosciences (UiO) (74).

3.7 Objectives

The aim was to use Google Trends data to analyse human seasonal rhythms in activity levels and libido. For this purpose, Google searches for news would function as a proxy for activity levels, and pornography searches would be a proxy for sexual drive.

We wanted to address the following study questions regarding the potential seasonality in human activity-rest patterns and libido:

1. Do activity-rest periods vary between seasons?

2. Does libido vary between seasons?

3. Are there hemispherical differences in activity and/or libido patterns? A seasonal anti-phase in the search patterns of the Northern and Southern Hemispheres (as seen in Figure 11) was expected if a seasonal effect was present.

4 Methods

This Methods section should really be named the “Process” section in the case of this Master’s Thesis. There were no established methods for the analyses we wished to perform, so there was a lot of trial and error involved in producing the results. The following paragraphs will explain the process, including planning, programming, obstacles, and the production of the final method. For a gross overview of the methods only, skip to section 4.6 - Summarised Methods.

4.1 Planning

Between 2018 and 2019, Shona Wood downloaded datasets from Google Trends with the intention of investigating seasonal rhythms in human search patterns. The datasets she downloaded had an hourly resolution, contained the keywords *porn, Facebook* or the name of a local news outlet (e.g., *Diario Rio Negro*, *nrk* or *svt*) and were from approximately 120 places around the world (e.g., *Buenos Aires*, *Telemark* and *Uppsala*). Her final data set consisted of 386 files. Each file was specific to one location and contained hourly data for approximately one year for one of the three keywords, as illustrated in Table 1.

Table 1: The original dataset. Illustrative schematisation of a small part of Shona Woods's original dataset of 386 files. Each file contained a unique combination of a keyword and location, as shown below. All the files had an hourly resolution and contained data for approximately one year.

Keyword	Location	Resolution	Time frame
porn	Buenos Aires Telemark Uppsala …	Hourly	~ April 2018 – April 2019
Facebook	Buenos Aires Telemark Uppsala …		
Diario Rio Negro nrk svt …	Buenos Aires Telemark Uppsala …		

Despite the effort behind downloading all these files, it was decided not to use them in this thesis. This decision was, first and foremost, based on the noisiness of the data. The files were based on single-keyword queries; however, another method for downloading Google Trends

data is to use categories. Categories are mainly used to specify the nature of your query. For instance, if you query Google Trends for the keyword *jaguar,* all Google searches that include this word will be used to produce a graph – it does not matter if people have searched for the animal or the car. However, if you filter your search by a category, for example, *cars and vehicles,* then all searches for *Panthera onca* will be excluded (75). In Computer Science, an asterisk (*) is often used as a wildcard character representing any other string or character (76). When an asterisk was combined with a category query on the Google Trends webpage, all searches within the category were used to produce the resulting graph, as illustrated in Figure 14. This resulted in a smoother and less noisy graph. It is worth noting that conducting a search in *all categories* with an asterisk (*) yielded no results. However, this approach would have been even better if it were possible.

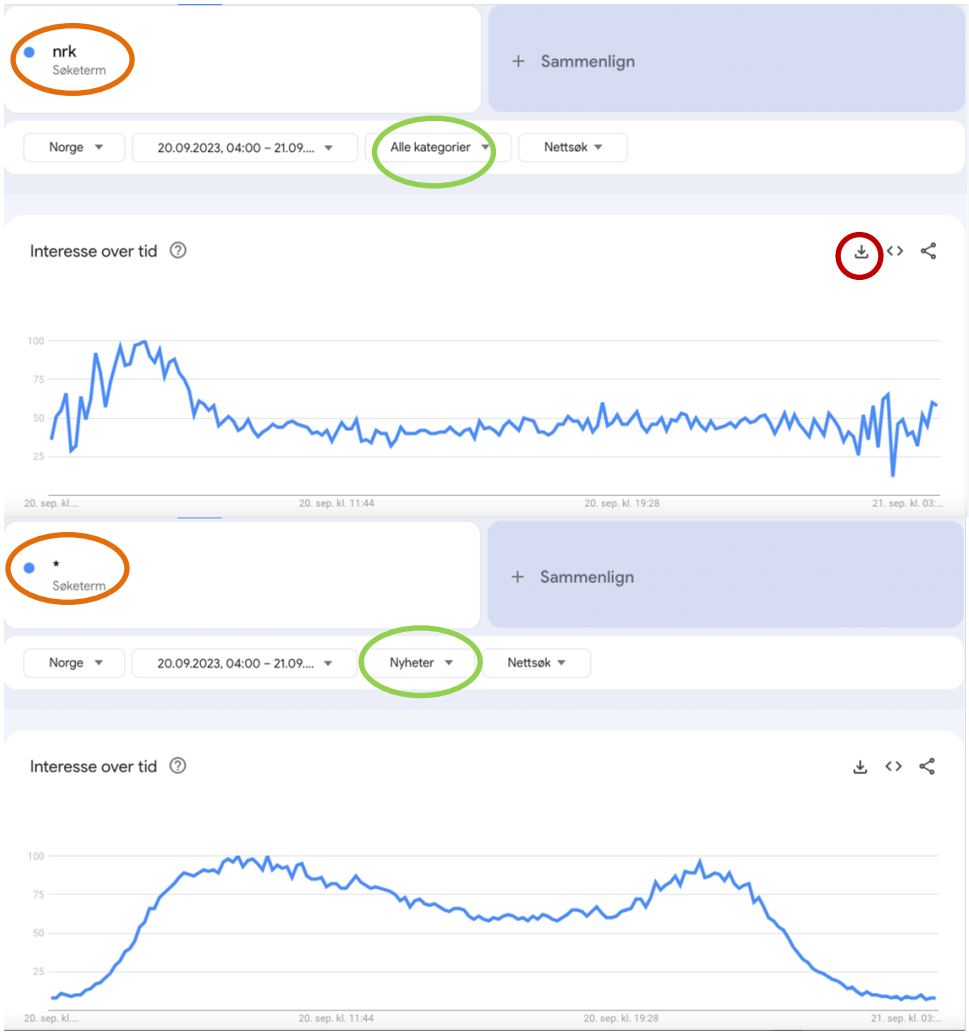


Figure 14: Example of a Google Trends keyword and category search. An unfiltered search for nrk (top) vs an *-search in the *News* category (bottom). Both queries are for the same period (20.09.23, 04:00 – 21.09.23, 04:00) and location (Norway). The bottom graph has a smoother curve with a clearer daily pattern because it is based on more search data. The red circle shows where to click to download the CSV file. The orange circles show the site of the keyword, and the green circles show the site of the category specification.

There is no Google Trends category for pornography, but it is possible to combine up to five keywords. Zattoni et al. (2020) had already used Google Trends to determine the five most popular search words for pornography: Porn, XNXX, PornHub, xVideos and xHamster. Thus, we could use these five words, separated by plus signs (+), to create a more extensive query, as illustrated in Figure 15.

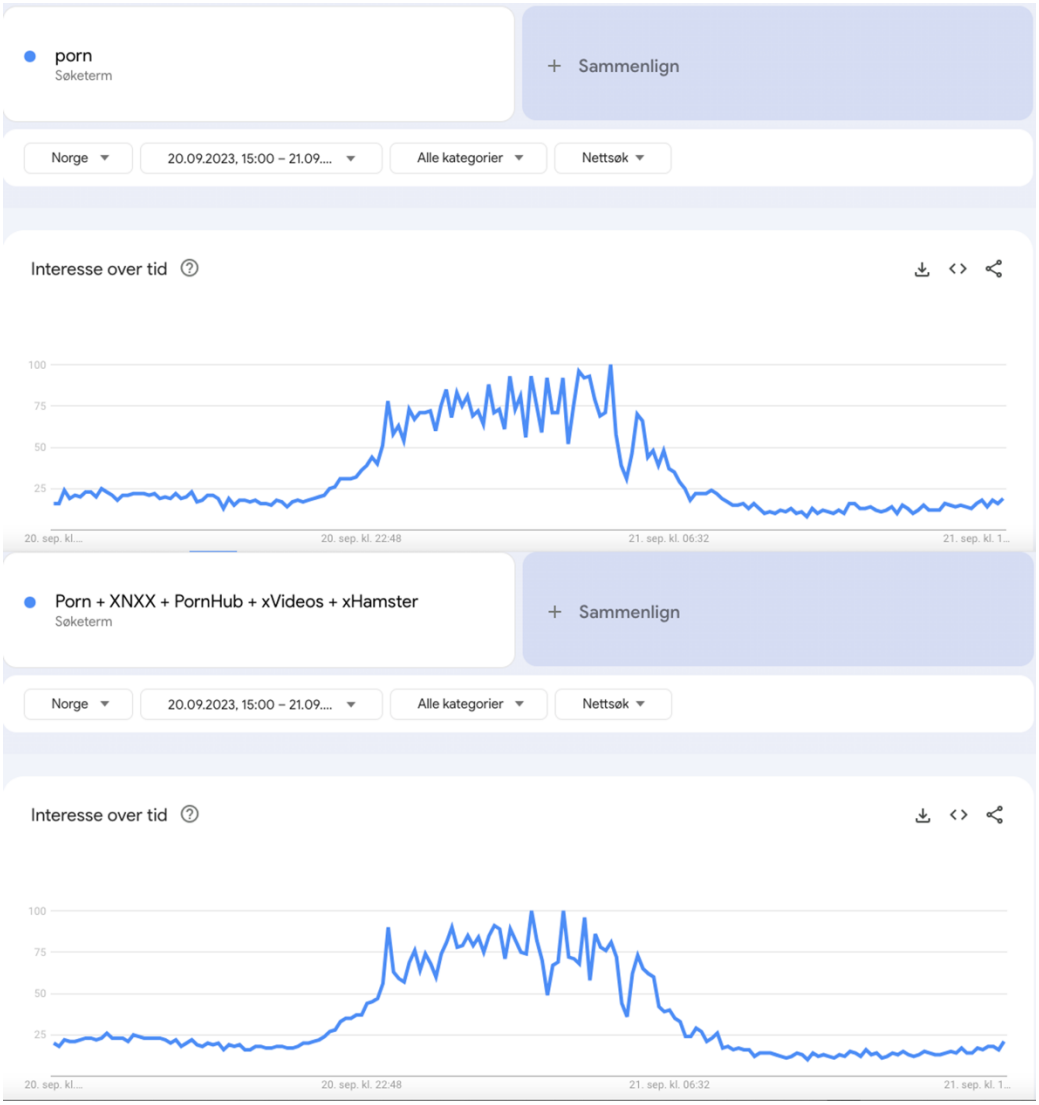


Figure 15: Google Trends pornography queries with one or five keywords. A query for *Porn* (top) renders a slightly noisier graph than for *Porn* + *XNXX* + *PornHub* + *xVideos* + *xHamster* (bottom). Otherwise, both queries are equal in the time frame (20.09.23, 15:00 – 21.09.23, 15:00) and location (Norway) and have no specified category.

Similarly, there is no "social media" category on Google Trends. Observations of the graphs made me suspect that most social media users do not use Google to enter these platforms but rather an app on their phones, which would not be observable through Google Trends (see

Figure 16). Thus, a decision was made to focus the analyses on data from the *News* category (and potentially other categories) and the combined pornography keywords.

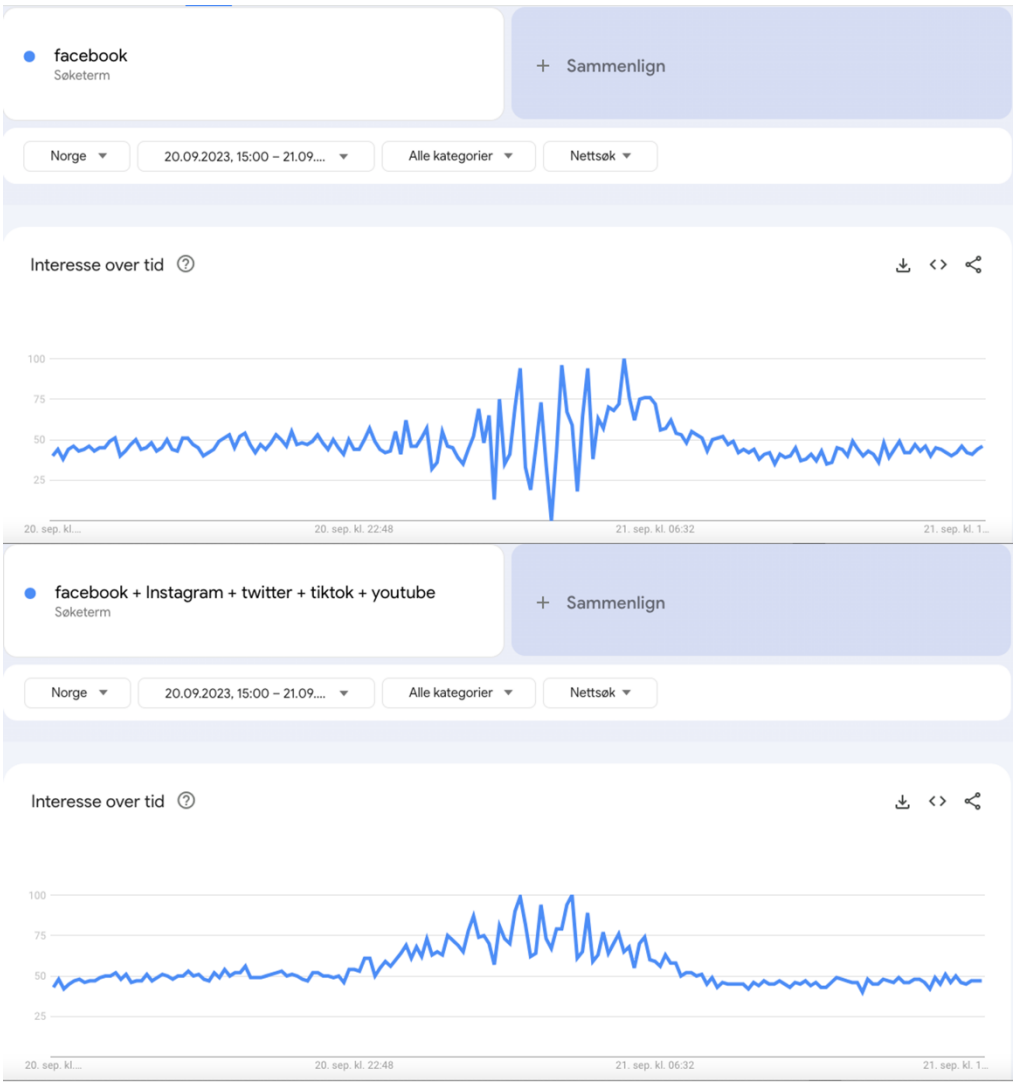


Figure 16: Google Trends social media queries with one or five keywords. A query for *Facebook* (top) renders a noisier graph than a query for *Facebook* + *Instagram* + *Twitter* + *TikTok* + *YouTube* (bottom). However, both graphs are inferior to the *News* category *-query because they are noisier and have less defined slopes in the early morning and late evening. These queries were thus discarded from the analyses. Both queries are equal in time frame (20.09.23, 15:00 – 21.09.23, 15:00) and location (Norway) and have no specified category.

The data files behind Google Trends graphs can be downloaded manually by clicking the downward-pointing arrow in the upper right corner, as shown in Figure 14. However, this method is tedious if the intent is to obtain several hundred or thousands of data files, and an unofficial API (Application Programming Interface) called *Pytrends* could be used to automatise the process (77). Implementing *Pytrends* would make it possible to download a large number of files from regions worldwide and obtain a new data set with the preferred

keyword and/or category combinations. The attempted implementation of *Pytrends* will be explained in section 4.2.

Another step in the planning phase was determining how to analyse the data. Two methods were made: one for analysing seasonality in daily search activity levels and another for investigating seasonality in pornography interest. These approaches are explained in the following paragraphs (sections 4.1.1 and 4.1.2).

4.1.1 Method for Analysing Seasonality in Activity Patterns

The method developed to analyse seasonality in daily search activity levels was based on the slopes that could be seen in the early morning and late evening for the *News*-category *-queries, as shown in Figure 17.

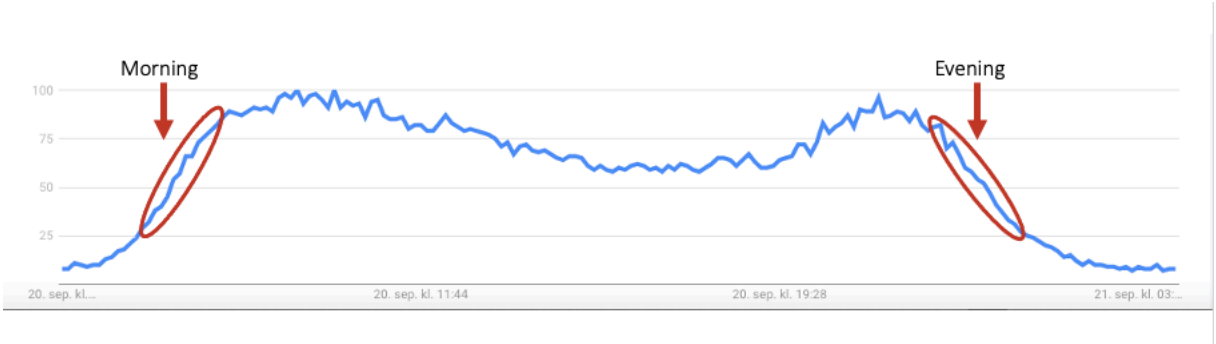


Figure 17: Morning and evening slopes on a daily Google Trends graph. The regions of interest for the planned analyses were the morning and evening slopes. These regions were interpreted as potentially representing when most people wake up and go to sleep on a given day, based on the assumption that many people use their phones right after waking up and right before going to sleep. The graph is a *-query in the *News* category, period 20.09.2023, 04:00 – 21.09.2023, 04:00, location Norway.

Two critical assumptions were made in this regard:

1. Many people use their phones right after waking up and before sleep.
2. Based on the previous statement, the first and last daily slopes roughly represent when people wake up and fall asleep.

A prerequisite for detecting any possible seasonal trends in the data would be to reduce the noise of the data as much as possible. If the data behind several daily graphs could be

averaged to create smooth graphs, this would be a good starting point for further analyses. The planned process is illustrated in Figure 18 and consisted of the following steps:

1. Average several daily datasets from Google Trends (across weeks, regions, or both) to obtain smooth curves.
2. Find the points with the steepest slopes in the morning and evening based on the first derivative.
3. Calculate the distance between the maximum slopes in the morning and evening to estimate activity length.
4. Compare the activity lengths for different seasons with an ANOVA and an appropriate post-hoc test if required.

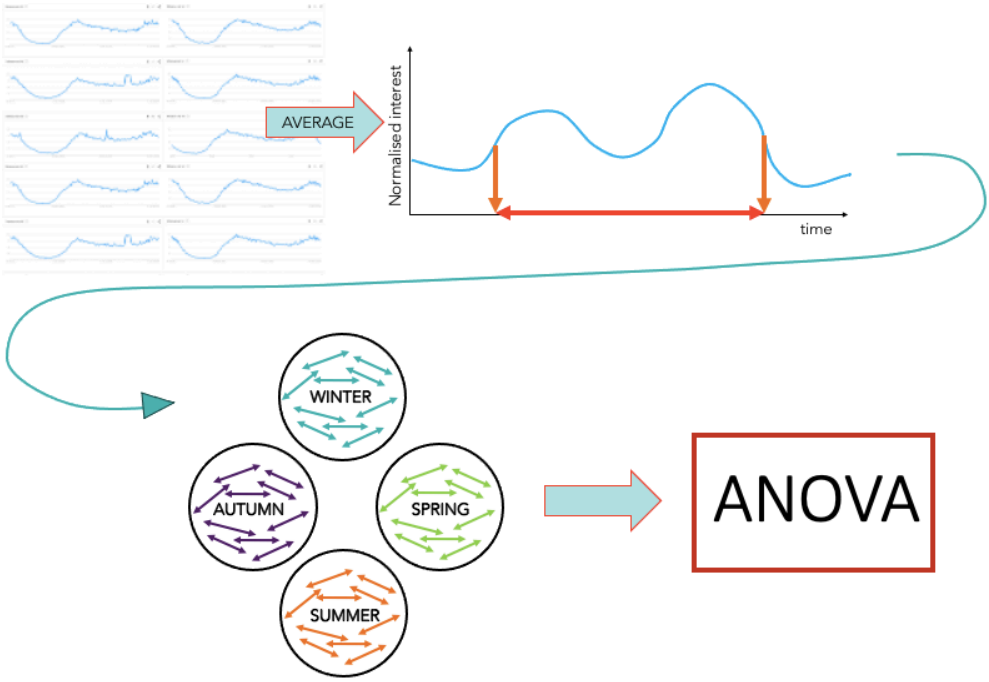


Figure 18: Work-flow diagram as imagined during the planning stage. Daily datasets for a particular time of year would be averaged to provide one smooth graph. Such averaged graphs would be used to find the points of the steepest slope in the morning and evening, based on the maximum 1st derivative. The distance between the times of the steepest morning and evening slopes would represent the activity length and would be the unit of interest in further analyses. The activity lengths would be grouped by seasons, and an ANOVA (and a Tukey post hoc test) would be used to look for any seasonal differences in activity length.

4.1.2 Method for Analysing Seasonality in Libido

The method that was developed to analyse seasonality in libido was based on yearly datasets from Google Trends (as illustrated in Figure 19). Yearly datasets have a weekly resolution,

which means that each data point represents the average search interest for a particular week (e.g., February 6th – 12th, July 17th – 23rd).

The plan consisted of the following steps:

1. Download yearly datasets based on the keywords *Porn* + *XNXX* + *PornHub* + *xVideos* + *xHamster*. The query would be performed without any specified category. Since the intention was to compare pornography use relative to light intensity, the seasons were centred around the solstices and equinoxes. Thus, Winter was November – January, Spring was February – April, Summer was May – June, and Autumn was August – October.
2. Compare weekly normalised search volumes grouped into four seasons with an ANOVA and potentially an appropriate post-hoc test.

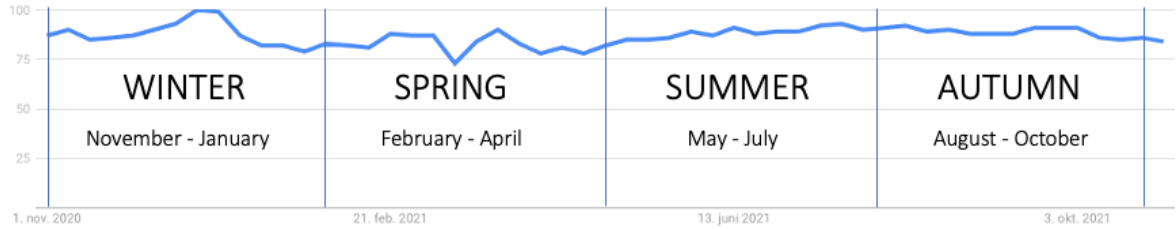


Figure 19: Yearly pornography query grouped by seasons. Seasonality in pornography searches would be analysed by downloading yearly datasets (as above) and splitting them into seasons. Next, the corresponding seasons from different years would be grouped and used in an ANOVA to look for seasonal differences in pornography search activity. The query in this illustration was for *Porn* + *XNXX* + *PornHub* + *xVideos* + *xHamster* for the period 01.11.2021 – 31-10.2022 within "All categories", location Norway.

4.2 Pytrends

To closely match seasonal changes in search activity and pornography interest with day length changes, it was deemed a good idea to download regional data to obtain high latitudinal detail. One goal was to create a gradient plot showing any latitudinal differences within Norway, Europe, or the World. The dataset would need to be huge to obtain this goal, but this could be solved (or so I thought) with *Pytrends*.

As mentioned, *Pytrends* is an unofficial API for the automatic download of data files from Google Trends (77), and it seemed like the perfect tool for this thesis. With the aid of

Pytrends, it would be possible to obtain a much larger dataset than could ever be downloaded manually through the Google Trends webpage.

The *Pytrends* webpage (https://pypi.org/project/pytrends/) was used as a reference as the code was developed. The aim was to build a pipeline to download historical search data from any region worldwide for the desired categories and/or keywords. Next, this data would be sorted and aggregated by latitude (Geopandas was implemented for this purpose (78)) and used in further analyses. A challenge was that all Google Trends data used UTC (Coordinated Universal Time), and this needed to be converted to the local time for each site to correctly identify the morning and evening slopes in an automated process. In this regard, knowing which countries used daylight saving time (DST), and when DST was implemented in each country was yet another obstacle that was worked on. However, this process was never completed because of issues that arose with *Pytrends*.

While working on the pipeline, it became clear that Google detected *Pytrends* queries as different from website visits since the message `USER_TYPE_SCRAPER` was often returned. When this happened, the returned data was either faulty or completely lacking. Faulty data contained chunks of missing data (0's) that were not seen in the graphs or datasets on the Google Trends website. Thus, Google Trends detected *Pytrends* as a robot downloader and appeared to sabotage such requests.

The subsequent step involved deploying a more advanced web scraping solution through a proxy service (OxyLabs) to conceal the IP address and obtain clean data while using Pytrends and browser automation tools. For a short time, this seemed to work, and some regional data from the Nordic countries (Norway, Sweden, Finland) was downloaded and tentatively analysed (see Appendix A1). However, later it was discovered that also this dataset contained occasional chunks of missing data and the data was later deleted by mistake.

The plan was to have the entire script in order prior to downloading the final datasets since we were paying per GB for using the proxy. For this reason, finishing the script and all methodological plans was prioritised over downloading data. However, in the early summer of 2023, Google Trends went through an update that changed both the website and the protocols behind it. For anyone using *Pytrends,* the result was a complete failure in retrieving data (see Figure 20). The previous messages of `USER_TYPE_SCRAPER` were gone, but all returned datasets were faulty or missing after the update. In a last effort, a computer scientist

investigated the problem to see if it was feasible to retrieve the necessary data to proceed with the analyses. However, Google's recognition of scrapers had become very advanced using anti-bot protection features, and it was unrealistic to omit this problem within the given time. Thus, downloading data with *Pytrends* could not be done. To this day (07.11.23), the issue has still not been solved, as seen on the Pytrends GitHub page (79).

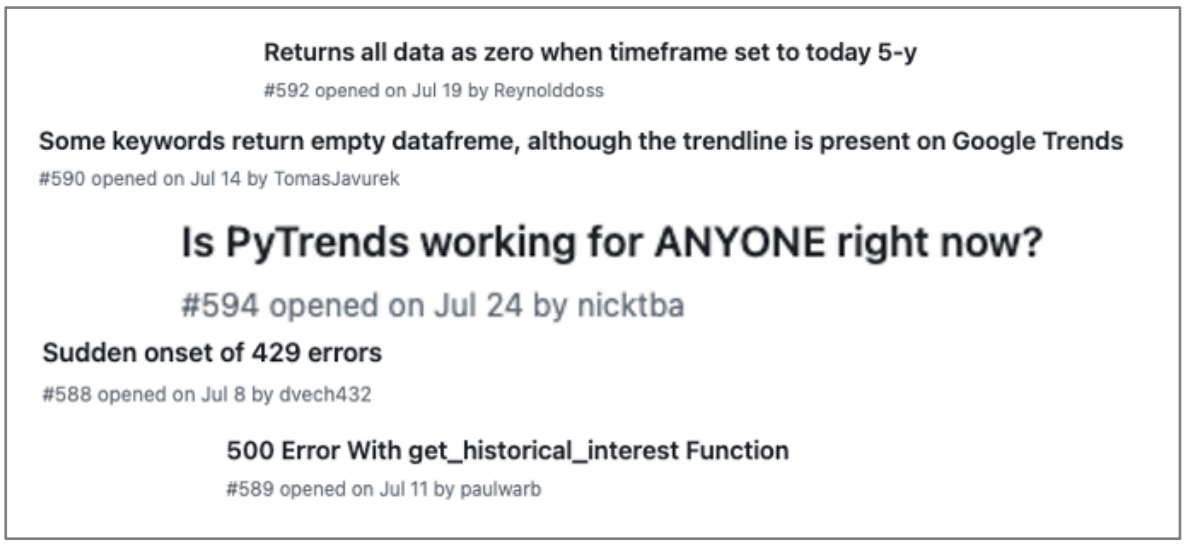


Figure 20: Desperate *Pytrends* users. A subset of the comments that appeared on the *Pytrends* GitHub Issues page after Google Trends updated their systems and *Pytrends* stopped working. The comment threads can be read in their full length at *https://github.com/GeneralMills/pytrends/issues.*

4.3 Manual Download

Around August 10th, 2023, it had become clear that Pytrends would not be able to provide the necessary data to complete the analyses. It seemed a daunting task at first, but a solution was to download sufficient data manually using the Google Trends webpage and the *download CSV* button in the upper right corner of the page (as shown in Figure 14). Downloading data manually would be time-consuming, and it would not be possible to obtain a dataset of the previously imagined size. The regions and time frames had to be narrowed down to make it possible to download them while also leaving enough time to complete the analyses. Thus, five regions were chosen: Norway, Sweden, Finland, New Zealand, and Victoria (the southernmost state of Australia). These areas were selected because they are located at relatively extreme latitudes above and below the equator and they were assumed to be similar regarding smartphone and/or screen access (i.e., they are all wealthy, industrialised countries). The chosen time frame was Saturdays between January 2016 and August 2023. The cut-off in

2016 was based on the observation that data from 2015 had drastically worse quality, as shown in Figure 21. Saturdays were chosen because downloading every day of the week would be too time-consuming and because it might be the day of the week that is the least influenced by work schedules (people who do not work during the weekend can sleep as long as they want on Saturdays and do not have to wake up early on Sundays).

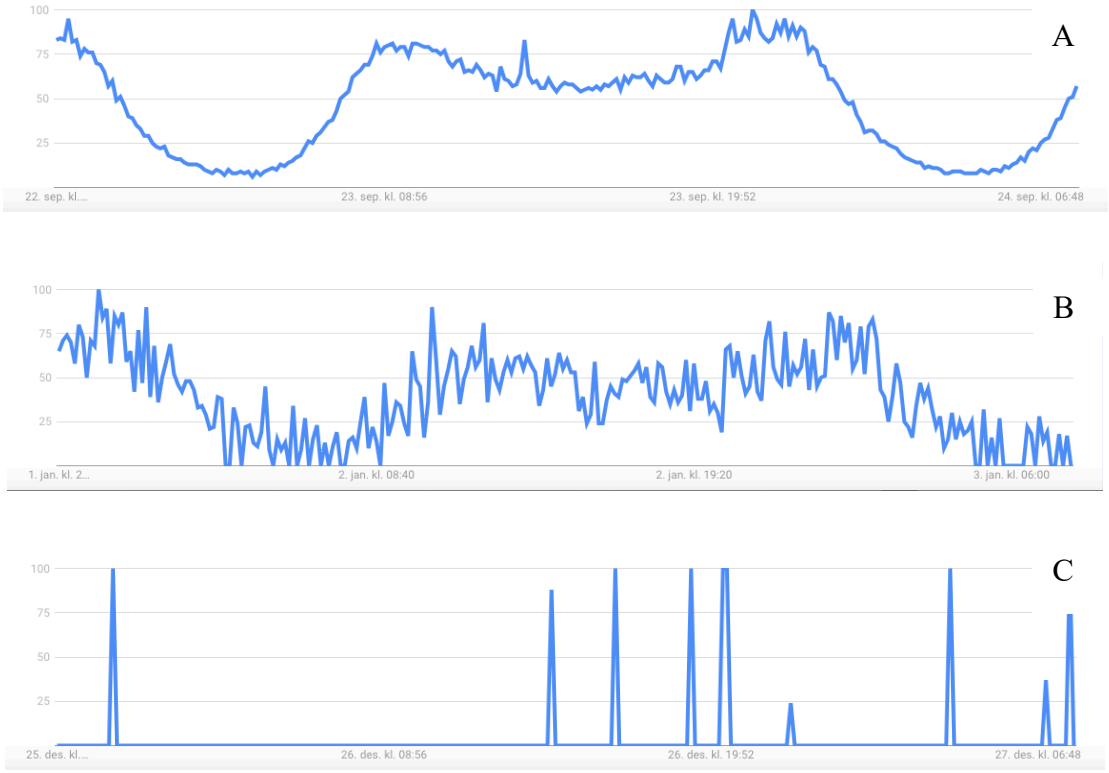


Figure 21: Deterioration of *-queries in the *News* category over time. A: Any dataset from the previous two weeks has a very high quality. This graph represents 22.09.23 – 24.09.23 and was produced 25.09.23. **B:** After two weeks, the datasets lose some of their original smoothness, but it is still possible to see patterns in the data. This graph represents 01.01.16 – 03.01.16. **C:** Prior to 2016, data is very sparse, which makes it impossible to detect daily oscillations. This graph represents 25.12.15 – 27.12.15.

The downloading process consisted of cross-checking a calendar (http://www.uke-kalender.com/) and changing the dates in the search field in the web browser at the sites indicated in Figure 22. This was done to download every Saturday between January 2016 to August 2023 for all five regions.

By visual evaluation, it was decided that UTC time 20:00 (Friday) to 06:00 (Sunday) included the morning and evening slopes of the desired Saturdays with generous margins in Norway. This time interval was at the boundaries of 8-minute resolution datasets and extending it any further would yield results on an hourly basis. This time span corresponded to the local times

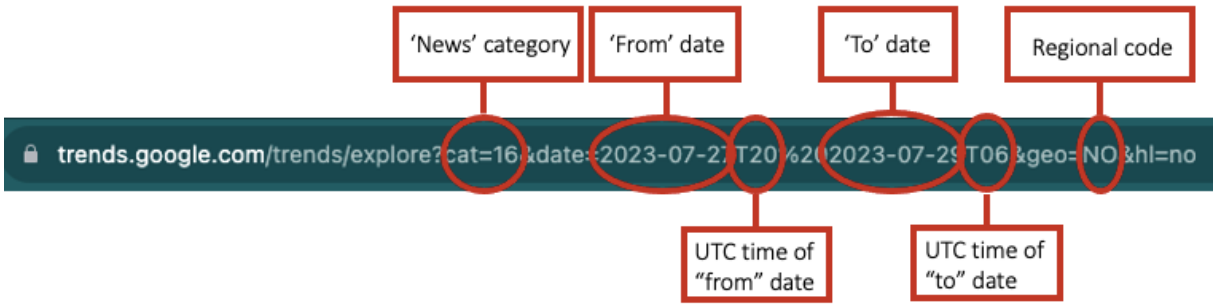


Figure 22: Sites of manipulation in the link to download daily data manually from Google Trends. Category 16 (cat=16) represents the *News* category. Data was downloaded between Friday–Sunday in the Nordic countries and between Friday-Saturday in the Oceanic regions – this difference was due to discrepancies between UTC and local time. The regional codes also vary between different sites, geo=NO represents Norway. The link was manually changes for each query.

Friday 21:00 to Sunday 07:00 under standard time and Friday 22:00 to Sunday 08:00 under DST. The UTC times in the links were adjusted to get the same local times in the other regions (see Figures 22 and 23). However, a calculation error led to a shift of one hour for the downloaded data from New Zealand and Victoria. These regions were downloaded with local times from Friday 22:00 to Sunday 08:00 under standard time and Friday 23:00 to Sunday 09:00 under DST. Also, the data from Finland was downloaded with a one-hour advance relative to the data from Norway and Sweden. These shifts were, however, unimportant to the analyses since they were within the generous margins around the Saturday of interest.

In the Northern countries, each data file spanned three days in UTC: from Friday to Sunday. In the Southern Hemisphere, each data file spanned two days in UTC: from Friday to Saturday. This was an artefact caused by the different time zones, and as shown in Figure 23, the selected time frames resulted in similar graphs for the two hemispheres. Importantly, all graphs/datasets included the Saturday morning and evening slopes. The UTC corrections that were performed to obtain the desired local times in the downloaded Google Trends datasets are summarised in Table 2.

Figure 23: Different time frames, similar graphs. *-queries in the *News* category spanning three days in Sweden (09.08.23 – 11.08.23) and two days in New Zealand (09.08.23-10.08.23) rendered similar graphs centred around Saturday.

Table 2: Region-specific time conversions for data download. These conversions were done to obtain an appropriate local time frame that would encompass all of Saturday in addition to a slight overlap with Friday evening and Sunday morning. Example of how to read the table: from the top left, Norway, with an example link, uses UTC + 1 as standard time and UTC+2 as daylight saving time (DST). To obtain a local standard time from 21:00 (Friday) to 07:00 (Sunday), the UTC time needed to be 20 (Friday) to 06 (Sunday) in the Google Trends link. This UTC time corresponds to 22:00 (Friday) to 08:00 (Sunday) during DST. All data was downloaded with 8-minute intervals between each data point. Note that the UTCs of Finland, New Zealand and Victoria all were subjected to calculation errors which rendered a ± 1-hour shift in these datasets relative to Norway and Sweden (see the text for further explanations).

Region	Example link	UTC → local	UTC time	Local time	Resolution
Norway	https://trends.google.com/trends/explore?cat=16&date=2023-07-28T20%202023-07-30T06&geo=NO&hl=no	UTC + 1 UTC + 2 (DST)	From: 20 (Fri) To: 06 (Sun)	Fri: 21 (22 DST) Sun: 07 (08 DST)	8 min
Sweden	https://trends.google.com/trends/explore?cat=16&date=2023-08-18T20%202023-08-20T06&geo=SE&hl=no				
Finland	https://trends.google.com/trends/explore?cat=16&date=2023-08-11T19%202023-08-13T05&geo=FI&hl=no	UTC + 2 UTC + 3 (DST)	From: 19 (Fri) To: 05 (Sun)		
New Zealand	https://trends.google.com/trends/explore?cat=16&date=2023-09-15T10%202023-09-16T20&geo=NZ&hl=no	UTC + 12 UTC + 13 (DST)	From: 10 (Fri) To: 20 (Sat)	Fri: 22 (23 DST) Sun: 08 (09 DST)	
Victoria	https://trends.google.com/trends/explore?cat=16&date=2023-08-04T12%202023-08-05T22&geo=AU-VIC&hl=no	UTC + 10 UTC + 11 (DST)	From: 12 (Fri) To: 22 (Sat)		

Pornography datasets were downloaded for the same five regions using the five previously selected keywords (*Porn* + *XNXX* + *PornHub* + *xVideos* + *xHamster*). These datasets, however, were yearly, starting in early November one year and ending in late October of the subsequent year. It was not possible to make the datasets start exactly on November 1st and end on October 31st because queries greater than nine months provide weekly resolution of the data, which clusters days. It was important to have all the seasons (centred around solstices and equinoxes) included in *one* dataset because of Google's normalisation procedure. It would, for example, not be possible to join January of 2017 to an already existing file for 2016, because the values of January 2017 would be scaled by the dataset to which they originally belonged and would thus not be independent.

Pornography data was downloaded from 2016 – 2021 in the following intervals:

- ~1 Nov. 2016 – ~31 Oct. 2017
- ~1 Nov. 2017 – ~31 Oct. 2018
- ~1 Nov. 2018 – ~31 Oct. 2019
- ~1 Nov. 2019 – ~31 Oct. 2020
- ~1 Nov. 2020 – ~31 Oct. 2021

Pornography data from ~1 November 2021 to ~31 October 2022 was not downloaded because of a change is Google's data collection system that occurred around January 1st, 2022, which distorted the normalised data for this year (see Figure 24). Data for ~1 November 2022 – ~31 October 2023 was not included because October 31st, 2023, had not yet occurred when the analyses were executed.

Lastly, a *-query in the *Arts and Entertainment* category was done for Norway, collecting data from every Saturday between 2016-2023, similar to what was done for the *News* category. This was done to verify that any potential findings in the *News* category were not only artefacts of that particular category. The *Arts and Entertainment* category was written as *cat=3* in the link shown in Figure 22.

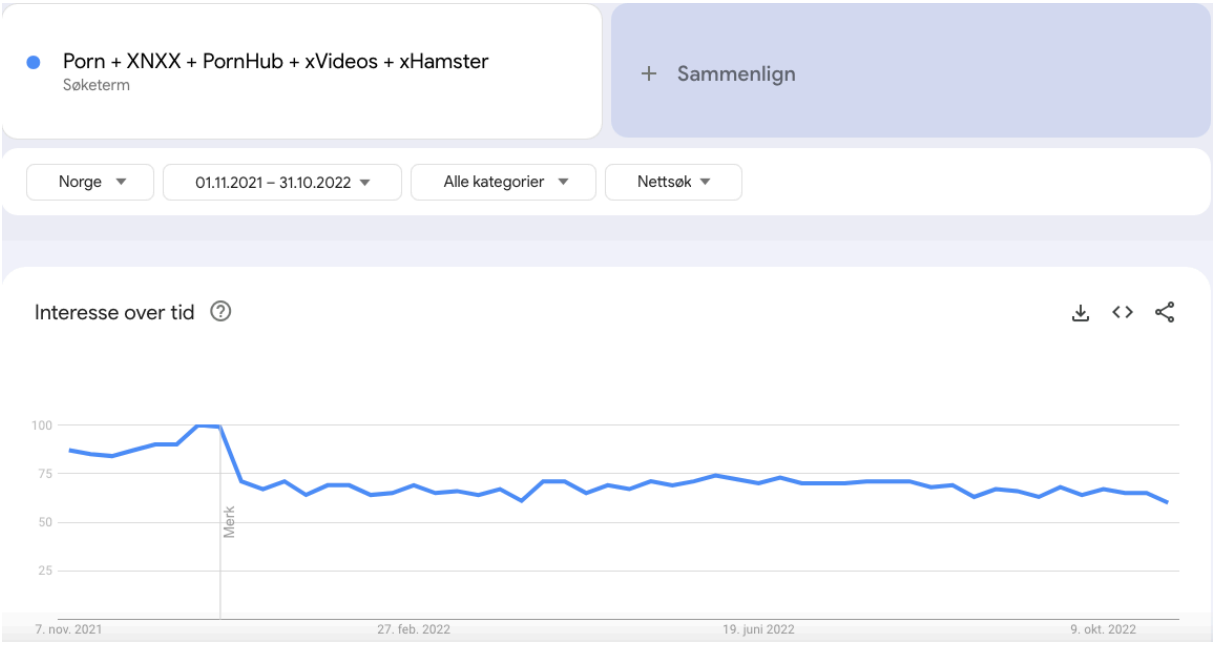


Figure 24: Updates in Google's data collection systems can distort datasets. Pornography data from November 2021 – October 2022 was not downloaded because of a change is Google's data collection system that occurred around January 1st, 2022, which distorted the normalised data for this time frame.

When all data had been downloaded, there were a total of 1980 datasets for the *-query in the *News* category, 25 datasets for the combined pornography keywords, and 399 datasets for the *-query in the *Arts and Entertainment* category. The total number of datasets was 2404. Each dataset consisted of a certain number of data points: 256 datapoints per dataset for the *News* or *Arts and Entertainment* category queries, and 52 datapoints per dataset for pornography keyword queries. The total number of data points was ~506,880 for *News* category queries, ~1300 for pornography keyword queries, and ~101,745 for *Arts and Entertainment* category queries. These numbers are summarised in Table 3.

Table 3: Total number of datasets (files) and data points (data) for each region and query.

Query	Region	Start	End	Files/region	Tot. files	Data/file	Total data points
News	Norway Sweden Finland NZ Victoria	2016	2023	396	1980	256	~ 506 880
Pornography	Norway Sweden Finland NZ Victoria		2021	5	25	52	~1300
Arts and Entertainment	Norway		2023	399	399	256	~101745
Total					**2404**		**609 925**

4.4 Curve Fitting

During the initial tests with *Pytrends*, I had downloaded regional data from every county in Norway, Sweden, and Finland across ten years. After averaging such an abundance of data across the corresponding weeks of different years and between regions, the results were very smooth curves compared to the raw graphs. However, the morning and evening slopes still contained some irregularities, and this would make it impossible to produce a script that used the averaged data directly to find the steepest morning and evening gradients. Initial tests showed that this could be solved by interpolating each averaged dataset.

After *Pytrends* was discarded and data had been downloaded manually, the data was no longer regional, but national. Also, the data was primarily downloaded from *one* Google Trends category (*News*), not several as was intended with *Pytrends.* The result was that there was now much less data to average, and this made the graphs/datasets much noisier than previously. Thus, interpolation no longer provided a reliable way to systematically avoid spurious minima and maxima, making it difficult to find the point of maximum derivative by programming. The solution was to fit a curve of some sort to the data. A fitted curve would entail sleek morning and evening slopes that could be used to find a reliable estimate of the maximum derivatives for any graph in an automated script.

First, a polynomial curve was fitted to the Nordic data, and by looking at the trending number of peaks it was decided that a 5^{th} order polynomial made the most sense (see Appendix A2). Yet, despite the fitted curve's resemblance to the dataset's shape, the fit was poor at the extremes, with the model frequently intersecting the dataset midway down the slope.

Also, the polynomial curve overshot the peaks in the original data from the Northern Hemisphere. On the Southern Hemisphere graphs, the fit became even worse, because they did not have only two distinct peaks as did the graphs from the North. Rather, these datasets had two distinct curves (morning and evening) and a varying number of smaller peaks in the middle of the day. Thus, the 5th-order polynomial could not be used as a general solution, and other options had to be explored.

The similarities between the northern and southern datasets were the morning and evening slopes and the morning and evening peaks. The difference was that in the South, there was often not the same mid-day trough as in the North. Based on this difference, it seemed like the

graphs were mixtures of different distributions where three or more bell-shaped curves were combined into one daily curve with a varying number of clear peaks. From this emerged the idea of trying to sum different distributions to produce one well-fitted curve.

First, the Gaussian distribution was tested for this purpose. However, the sum of Gaussian distributions was not a good fit because the peaks of the original data were too sharp for the Gaussian curve, and this shifted the important morning and evening slopes. Then Liu et al. (2010) was discovered, which stated that user browsing behaviour can follow a Weibull distribution (73). Summing different Weibull distributions produced a good fit for the morning and evening slopes, which were the important parts of the planned analyses.

The consistent observation of morning and evening slopes and peaks made it natural to fit one Weibull at each of these sites. Although the mid-day distribution was often not evident in the Northern datasets, and often underfitted *several* scattered and smaller distributions in the midpart of the Southern datasets, one Weibull served to summarise (and somewhat ignore) what was happening in the middle sections for the purposes of this thesis. Also, as more curves were attempted to be fitted in this middle section, the optimisation process (non-linear least squares) became increasingly unstable. The more peaks were fitted the higher the risk became of convergence issues like parameters reaching bounds, Weibull curves disappearing, unbalanced overlapping of distributions, unbalanced bias and curves taking the places of other curves. Also, these issues were not systematic and happened randomly, highly depending on the dataset and its noise. In the end, three Weibull distributions were combined for each dataset since this served to place the morning and evening distributions in suitable positions on the x-axis.

The mixture of three Weibull distributions was used for both the North and the South. The three Weibull distributions were forced to share the k (shape) and λ (scale) parameters to facilitate convergence. I called this mixture of Weibull distributions TRIWEI. Visual verification ensured that all TRIWEI plots had obtained a good fit with the data and that the bound limits of the parameters were not met (the best values were found within the provided bounding limits, see further explanations below).

Specifically, the *translated* Weibull distribution was used:

$$WEI(x; k, \lambda, \mu) = \frac{k}{\lambda}\left(\frac{x-\mu}{\lambda}\right)^{k-1} e^{-\left(\frac{x-\mu}{\lambda}\right)^{k}} \quad for \quad x \geq \mu \qquad \text{Eq. 2}$$

Where $k > 0$, $\lambda > 0$, and μ is the location parameter.

The three *translated Weibull distributions* were combined in the following manner:

$$TRIWEI(x; k, \lambda, \mu_1, P_1, \mu_2, P_2, \mu_3, P_3, B) = P_1\, WEI(x; k, \lambda, \mu_1) + P_2\, WEI(x; k, \lambda, \mu_2) + P_3\, WEI(x; k, \lambda, \mu_3) + B$$

Eq. 3

Where P_1, P_2, P_3 are multiplicative factors to scale the individual Weibull distributions and B is a bias that slightly translates all curves vertically to optimise the fit.

Each of the three *translated Weibull distributions* had to be scaled and translated vertically and horizontally. This was done by changing μ, P and B with the `scipy.optimize.curve_fit` function from the SciPy library in Python. This function used the TRIWEI equation (Eq. 3), the Google Trends data points and initial guesses of the bounds of $k, \lambda, \mu_1, P_1, \mu_2, P_2, \mu_3, P_3$ and B. To make these guesses, the x-axis consisting of time stamps was converted to integer units with the `np.linspace` function.

A trial-and-error process was initiated that roughly adhered to the following approach. The process began with initial "guess" and "bounds" as null values. While the model fit was deemed as bad (for example, if bounds were reached) through visual inspection, an iterative process was performed. In this process, models were fitted to the Google Trends graphs with the `scipy.optimize.curve_fit` function. These models had parameter values that were subsequently averaged to produce a new guess value. If a bound value was met, the following bound guess was expanded, and if the bound value was not met, the next bound guess was shrunk. This cycle was repeated until all plots were visually deemed as satisfactory within the provided bounds.

After some trial and error of fitting the curves, the following parameter guesses were chosen:
$k = 2.7, \quad \lambda = 81.33, \quad \mu_1 = 21.84, \quad P_1 = 55.54, \quad \mu_2 = 85.70, \quad P_2 = 54.06,$

$$\mu_3 = 181.26, \qquad P_3 = 56.02, \qquad B = 0.05$$

Based on these estimates, the following bounds were selected after several rounds of testing:

$2.6 \leq k \leq 6\,, \quad 80 \leq \lambda \leq 120, \quad -10 \leq \mu_1 \leq 30, \quad 30 \leq P_1 \leq 100,$

$$80 \leq \mu_2 = 130, \quad 0 \leq P_2 \leq 100, \quad 160 \leq \mu_3 \leq 210, \quad 30 \leq P_3 \leq 100, \quad 0 \leq B \leq 0.2$$

These bounds were consistent with the hypothesis that the data followed a mixture of Weibull distributions since they provided well-fitting curves for the datasets. Within the bounds, the fitting algorithm found optimal parameters for all the datasets.

TRIWEI functions were used to fit curves to data of four different types (as summarised in Table 4):

1: **Raw Google Trends data.**

2**: Data averaged within hemispheres.** For each Saturday between 2016 to 2023, Google Trends datasets were averaged for the Northern countries (Norway, Sweden, Finland) and for the Southern Hemisphere regions (New Zealand and Victoria) separately. This provided one averaged dataset for each hemisphere for each Saturday from 2016 to 2023.

3: **Data averaged across years.** All datasets from week number X of different year were averaged for each region separately to provide 52 averaged datasets for each site.

4: **Data averaged across years and within hemispheres**. First, the same week numbers of the different years were averaged for each site. Then, the corresponding weeks within each hemisphere were averaged to produce 52 datasets for each hemisphere.

Some years have 53 weeks, although most years have 52. For this reason, there are 53 TRIWEIs per region for *Data averaged across years* and per hemisphere for *Data averaged across years and within hemispheres.* However, the 53rd weeks were excluded from all analyses because their presence in the datasets was inconsistent.

Table 4: Four data types fitted with TRIWEI functions. Fitting of TRIWEI functions was done to four different types of *News* data, as explained above. This resulted in a varying number (#) of TRIWEIs in total and per hemisphere (North, South) or region (Norway, Sweden, Finland, New Zealand, Victoria).

Data types	# files before average	# TRIWEI	# TRIWEI per region
Raw data	1980	1980	North: 1188 South: 792
Data averaged within hemispheres		792	North: 396 South: 396
Data averaged across years		265	53 per region
Data averaged across years and within hemispheres		106	North: 53 South: 53

Since data had been downloaded from Friday to Sunday and the region of interest was Saturday, the datasets were transformed to local time and cut between Saturday 03:00 AM and Sunday 03:00 AM. This cutting disrupted the ratios of the datasets since they had been normalised by Google within the requested time frame. Before fitting the TRIWEI curves, data was normalised between zero and one. This was done to re-establish similar Y-axis values to make the guessed TRIWEI bounds narrower. A narrower bound range was desired since it provided smaller chances of errors.

4.5 Activity Lengths, Statistical Analyses, and Plotting

4.5.1 Activity Length Calculation

For each fitted TRIWEI, morning and evening slope derivatives were used to calculate the activity length, as explained in section 4.1.1. Activity lengths were found by subtracting the time of the morning derivative from the time of the evening derivative:

$$\text{Activity Length} = \text{Derivative time 2} - \text{Derivative time 1} \qquad \text{Eq. 4}$$

This procedure was also carried out using the morning and evening peaks to investigate if this measure could function as an alternative to the derivatives (see Appendices A3 and A4):

$$\text{Alternative Activity Length} = \text{Peak time 2} - \text{Peak time 1} \qquad \text{Eq. 5}$$

Activity lengths were grouped by seasons that were centred around the solstices and equinoxes. The weeks corresponding to the Northern and Southern hemisphere seasons are schematised in Table 5.

Table 5: Seasonal grouping of weeks. The weeks of the year belonging to winter, spring, summer, and autumn in the Northern and Southern Hemispheres when seasons are centred around solstices and equinoxes.

Northern season	Southern season	Weeks
Winter	Summer	46–52 and 1–6
Spring	Autumn	7 – 19
Summer	Winter	20 – 32
Autumn	Spring	33 – 45

4.5.2 *News* Category ANOVAs and Post Hoc Tests

The seasonal groups (see Table 5) were used directly to perform ANOVAs and post hoc tests for the activity lengths of each data type. Two-way ANOVAs and Tukey post hoc tests were done using the `bioinfokit.analys` library and its `anova_stat` and `tukey_hsd` functions. These statistical tests were decided on early in the planning process and were a fundamental part of the study design. The ANOVA was chosen because it detects whether the means of different groups are equal or not (80). The Tukey post hoc test was used since it was recommended in the course STA-3300 for when you want to do all pair-wise comparisons.

For the activity length analyses, the response (dependent) variable was activity length, and the predictor (independent) variables were seasons, hemispheres and the interaction between season and hemisphere.

The model can be expressed as follows:

$$Activity\ length = \mu + Season + Hemisphere + Season \cdot Hemisphere + \varepsilon$$

Where:

- *Activity Length* is the dependent variable representing the duration of daily search activity.
- μ is the overall population mean.

- *Season* is a categorical, independent variable with four levels (*Winter*, *Spring*, *Summer*, and *Autumn*).
- *Hemisphere* is another categorical, independent variable with two levels (*North*, *South*).
- The interaction term *Season · Hemisphere* represents how the combination of *Season* and *Hemisphere* affects *Activity length*.
- ε represents the error term, which accounts for random variability and factors not considered in the model.

The hypotheses related to the effects of *Season*, *Hemisphere*, and the interaction *Season · Hemisphere* can be expressed as follows:

- Null hypothesis: The group means are equal: H_0: $\mu_1=\mu_2=...=\mu_8$
- Alternative hypothesis: H_1: Not all μ are equal.

Statistical tests for seasonal differences were not performed within each separate country/region. All statistical tests used the standard significance level of $\alpha= 0.05$.

To find the seasonal average times of the morning and evening gradients, `pandas.Series.dt.tz_convert` was used to convert the downloaded data to the local times of the different regions.

The ANOVA model assumes the data are independent, stem from normally distributed populations and have equal variance (81). The last two of these assumptions were tested as shown in Appendix A5.

Since the *Raw data* violated the assumption of homoscedasticity, Welch's ANOVA and Games-Howell post hoc test were carried out as an alternative. StatisticsHowTo.com and several other academic websites recommend Welch's ANOVA for normal, balanced data with heteroscedasticity (82, 83). However, Welch's ANOVA is an alternative to the one-way ANOVA and only takes one predictor variable. Thus, *season* was the only predictor variable in these analyses and *activity length* remained the response variable. Since Welch's ANOVA is sensitive to unequal sample sizes, all activity lengths from 2023 (January-August) were removed from these analyses.

4.5.3 *News* Category Polar Plots

Polar plots were made with the `bar_polar` function from the `plotly.express` library. The polar plots had 52 weeks on the polar axis. The activity lengths from the data types *Data averaged across years* and *Data averaged across years and within hemispheres* could be plotted directly. The activity lengths from the data types *Raw data* and *Data averaged within hemispheres* had to be averaged across the corresponding weeks of the different years before the polar plots could be made. *Data averaged across years* was also used to make hemispherical polar plots; in these, the activity lengths for each region within a particular hemisphere were averaged.

4.5.4 *Arts and Entertainment* ANOVA, Post Hoc and Polar Plots

ANOVAs and Tukey post hoc tests were also performed for the data from the *Arts and Entertainment* category, but since this data was only downloaded for Norway, these analyses only had one predictor variable, *Season,* and the test was thus a one-way ANOVA. The response variable was activity length, as in the analyses of the *News* category. The analyses were done with data of two types, *Raw data* and *Data averaged across years* (as described in section 4.4).

The model can be expressed as follows:

$$\text{Activity length} = \mu + \text{Season} + \varepsilon$$

Where:

- Activity Length is the dependent variable representing the duration of daily search activity.
- μ is the overall population mean.
- Season is a categorical, independent variable with four levels (Winter, Spring, Summer, and Autumn).
- ε represents the error term, which accounts for random variability and factors not considered in the model.

The hypotheses related to the effects of *Season* was:

- Null hypothesis: The group means are equal: H_0: $\mu_1=\mu_2=\mu_3=\mu_4$
- Alternative hypothesis: H_1: Not all μ are equal.

The ANOVA assumptions (see section 4.5.2) were tested as shown in Appendix A6. Since the *Raw data* did not meet the assumption of homoscedasticity, Welch's ANOVA and Games-Howell post hoc test were carried out as an alternative. Since Welch's ANOVA is sensitive to unequal sample sizes, all activity lengths from 2023 (January-August) were removed from these analyses.

Polar plots for the *Arts and Entertainment* category were made as described in section 4.5.3. The activity lengths from the data type *Data averaged across* could be plotted directly. The activity lengths from the data type *Raw data* had to be averaged across the corresponding weeks of the different years before the polar plots could be made.

4.5.5 Pornography ANOVA, Post Hoc and Polar Plots

Since the pornography data were heteroscedastic (see model assumption tests in Appendix A7), Welch's ANOVA and Games-Howell post hoc tests were used on these data. The data were split between the two hemispheres and separate analyses were carried out for the North and the South. The predictor variable was *Season,* and the response variable was *Normalised search interest.*

The pornography data was also plotted as polar plots (as explained in section 4.5.3). First, the corresponding weeks of different years were averaged. For each week of a particular year, there was only one number/data point, since Google Trends only provided weekly resolution on datasets > 9 months. To produce pornography polar plots for the Northern and Southern Hemispheres, the corresponding, averaged weekly data for the different regions within the Northern and Southern Hemispheres were also averaged.

4.5.6 Seasonal Wave Plots

For the data types that were not averaged across years (*Raw data* and *Data averaged within hemispheres*), seasonal wave plots were made with the activity lengths of all Saturdays between 2016-2023. This was done for data from the *News* and *Arts and Entertainment* categories as well as for the pornography data.

4.5.7 Photoperiod Graphs

Some of the data types (see bullet points below) were also plotted against the photoperiod of a northern and southern point of the respective geographical region. These graphs were made in Excel. The photoperiods were gotten from https://www.timeanddate.no/astronomi/sol, where you can search for sunrise, sunset and daylength for any place in the world. The daylengths were from Saturdays of 2020, and the activity length datasets that were used for this purpose were:

- *Data averaged across years* for Norway and the Northern and Southern Hemispheres.
- Weekly data averaged across years for the pornography queries in Norway, the Northern Hemisphere, and the Southern Hemisphere.

The places used for day lengths in the photoperiod graphs are summarised in Table 6:

Table 6: The northern and southern locations that were used to represent the most extreme photoperiods within a particular region. Nordkapp and Lindesnes are the northernmost and southernmost parts of Norway, respectively. Smygehuk is the southernmost point of Sweden and of the joint area of Norway, Sweden, and Finland. Lindsay point is the northernmost point of Victoria (Australia), and Stewart Island is an island in the south of New Zealand. The southernmost point of New Zealand was not used because it is a remote island with few inhabitants which is located very far south compared to the rest of the country.

Region	Northern photoperiod	Southern photoperiod
Norway	Nordkapp (Finnmark, Norway)	Lindesnes (Agder, Norway)
Northern hemisphere	Nordkapp (Finnmark, Norway)	Smygehuk (Skåne, Sweden)
Southern hemisphere	Lindsay point (Victoria, Australia)	Stewart Island (Raikura, New Zealand)

4.5.8 Code and Data Access

The code and data that were used to perform the analyses described above are available through https://github.com/gtthesis/code_thesis.

4.6 Summarised Methods

This section does not provide complete explanations about the methodology but serves to summarise the final approach.

Activity analyses:

1. Data from the *News* category in Google Trends was downloaded for each Saturday between January 2016 and August 2023 for the following regions: Norway, Sweden, Finland, New Zealand, and Victoria. Later it was transformed to local time, cut to a specific time frame, and normalised.
2. A TRIWEI (mix of three Weibull functions) was fitted to Google Trends datasets that were 1) raw datasets, 2) averaged within hemispheres, 3) averaged across years, 4) averaged across years and within hemispheres. See section 4.4 for further details.
3. The first derivative was used to find the steepest point on the morning and evening slopes on the TRIWEI graphs.
4. The distance between the points of maximum morning and minimum evening slopes was used as a measure of the activity length of search activity on Saturdays. As an alternative metric the distance between the first and last peak was also considered.
5. Activity lengths were grouped by seasons centred around the solstices and equinoxes. Plots were made, and statistical tests (ANOVA + Tukey post hoc and Welch's ANOVA + Games-Howell post hoc) were performed.
6. Data from the *Arts and Entertainment* category was also downloaded and analysed as described in the steps above. However, this was only done for Norway.

Libido analyses:

1. The keywords *Porn* + *XNXX* + *PornHub* + *xVideos* + *xHamster* were used to perform yearly Google Trends queries between 2016 and 2021 for the following regions: Norway, Sweden, Finland, New Zealand, and Victoria. All datasets were downloaded from approximately November 1st of one year to October 31st of the consecutive year.
2. The weeks of the year were grouped by seasons centred around the solstices and equinoxes. Statistical tests (Welch's ANOVA + Games-Howell post hoc) were performed to investigate seasonal differences in search interest.

3. The corresponding weeks of different years were averaged to produce polar plots for each region. The weeks were also averaged for the regions within each hemisphere to produce plots for the Northern and Southern hemispheres.

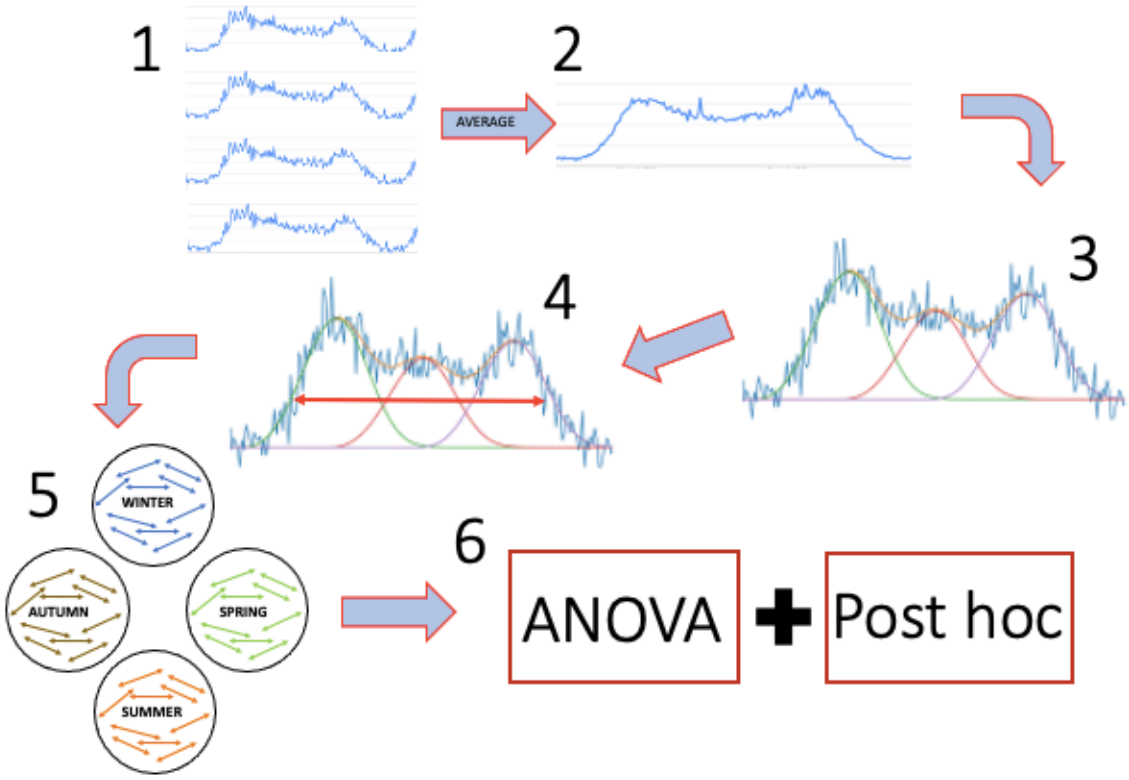


Figure 25: Final workflow diagram, activity length analyses. 1: Google trends daily datasets were **2:** averaged. **3:** Averaged datasets were fitted with TRIWEIs. **4:** the activity lengths were measured as the length between the steepest morning and evening slopes. **5:** Activity lengths were grouped by seasons and **6:** statistical tests were performed.

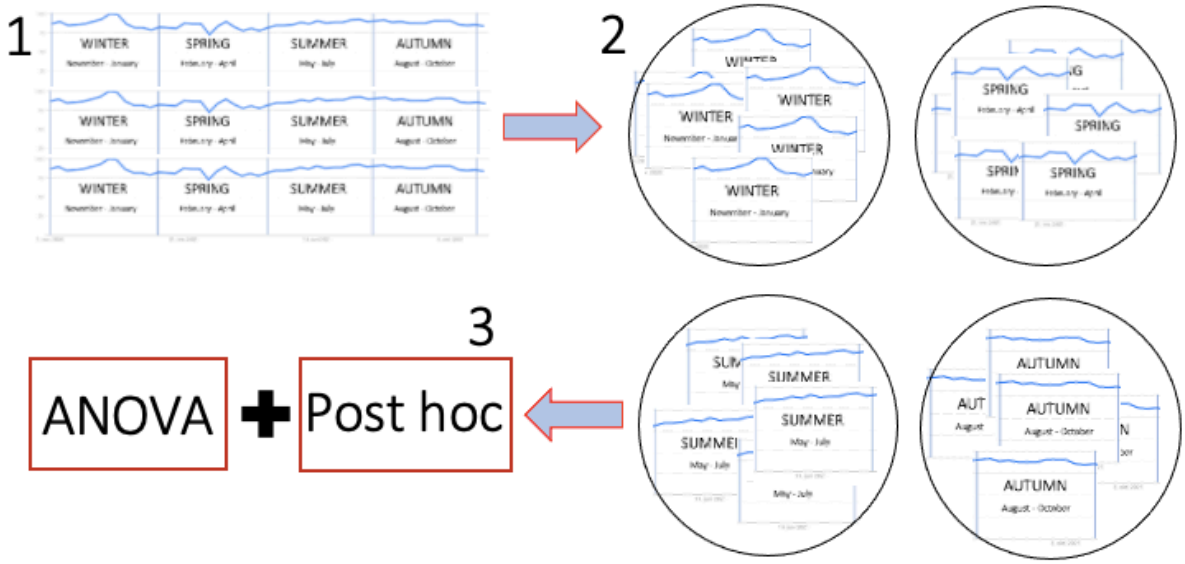


Figure 26: Final workflow diagram, pornography interest analyses. 1: Google trends yearly pornography queries were **2:** grouped by season (weekly resolution). **3:** Statistical tests were performed to investigate seasonal differences in weekly search interest.

5 Results

5.1 TRIWEI Curve Fitting Performed Well with Daily Google Trends Datasets

Some examples of fitted TRIWEIs for the different data types are shown in Figures 27, 28, 29 and 30. The top text of the graphs contains the parameters of each Weibull, where G is k/ λ, L is λ, M is μ. Based on an evaluation of the TRIWEI fits, the number of datasets for each data type (as listed in Table 4), the number of relevant predictor variables and the model assumption tests (see Appendices A5 and A6), it was decided to only present the statistical analyses of the *Data averaged across years* data in this Results section. The analyses of the other data types can be found in Appendices A8 - A11.

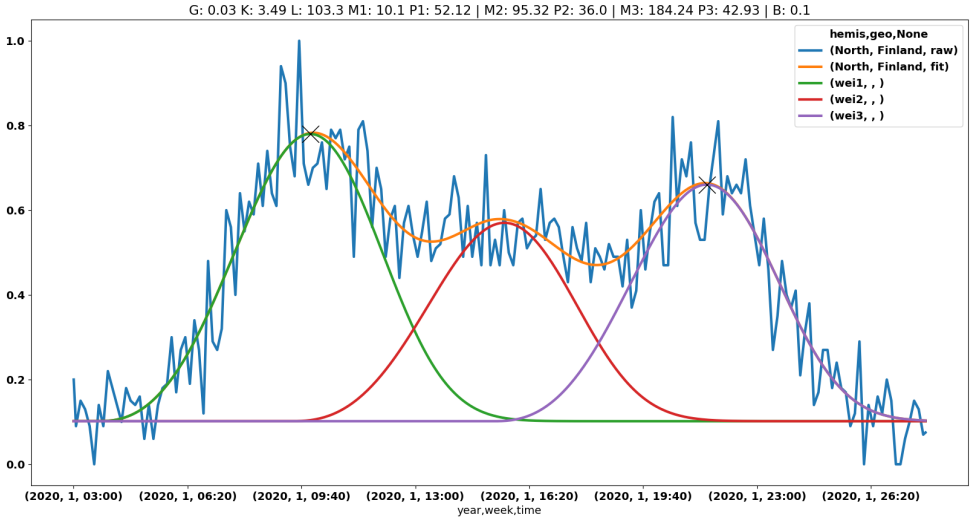


Figure 27: Example graph of TRIWEIs fitted to Raw *News* data from Finland, week 1, 2020. Top text: G is k/ λ, L is λ, M is μ.

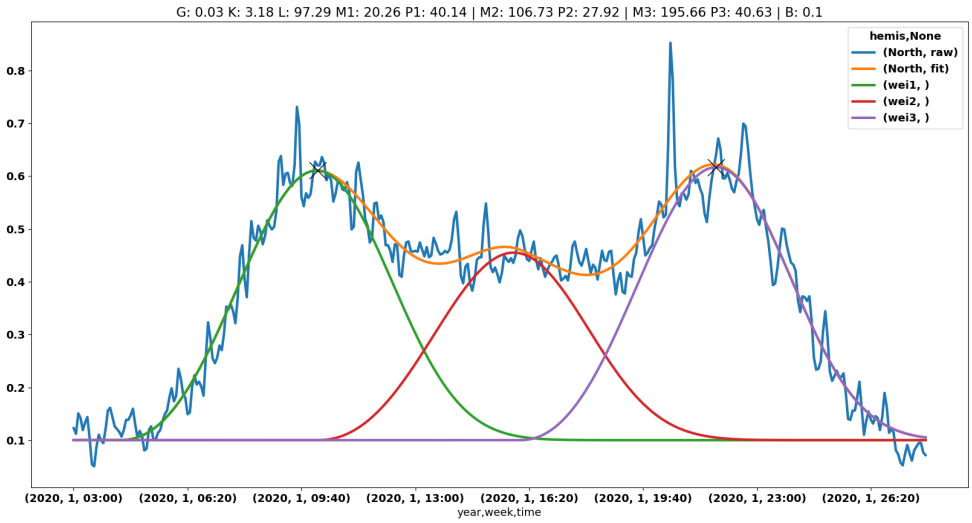


Figure 28: Example graph of TRIWEI fitted to *News* Data averaged within hemispheres, North, week 1, 2020. Top text: G is k/ λ, L is λ, M is μ.

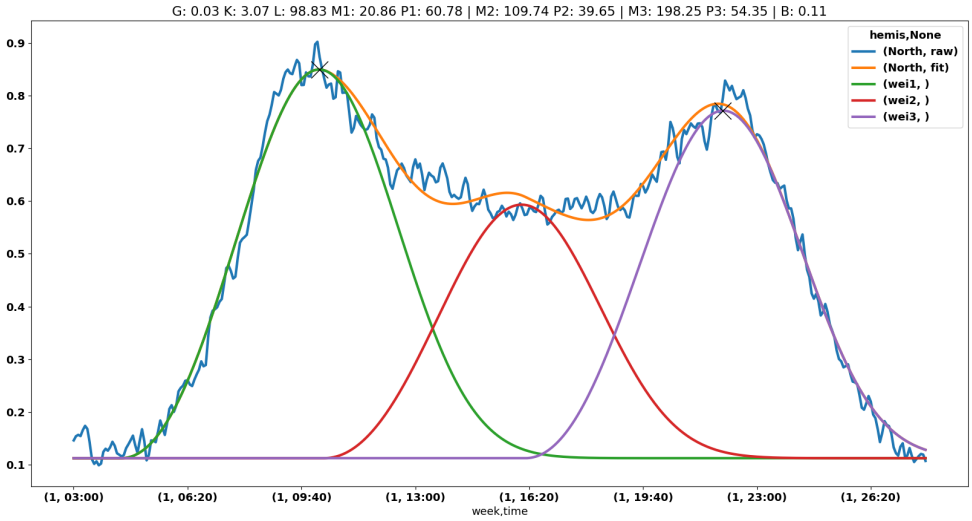


Figure 29: Example graph of TRIWEI fitted to *News* Data averaged across years and within hemispheres, North, weeks 1. Top text: G is k/ λ, L is λ, M is μ.

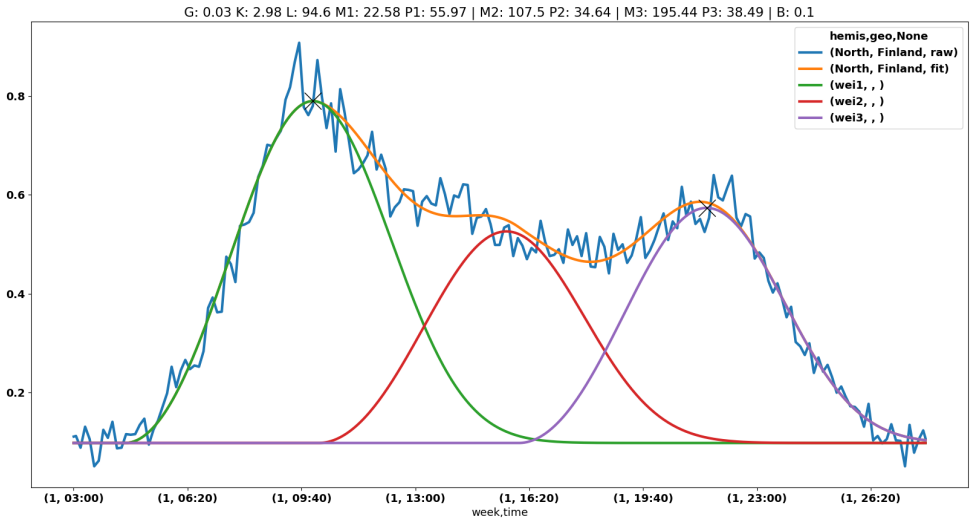


Figure 30: Example graph of TRIWEI fitted to *News* Data averaged across years, Finland, weeks 1 averaged. Top text: G is k/ λ, L is λ, M is μ.

5.2 Significant Seasonal Differences in Activity Lengths Derived from the *News* Category.

The two-way ANOVA of the 265 TRIWEI gradient-based activity lengths derived from *Data averaged across years* showed a significant effect of season, hemisphere and the interaction between season and hemisphere, as shown in Table 7.

The following Tukey post hoc test for the Northern Hemisphere revealed statistically significant differences between winter and spring (mean difference: 18.2 min, 95% CI: 8.1– 28.2, p = 0.001), winter and summer (mean difference: 14.8 min, 95% CI: 4.7– 28.8, p = 0.001) and spring and autumn (mean difference: 11.1 min, 95% CI: 0.8– 21.3, p = 0.028). All other pairwise comparisons in the Northern Hemisphere were nonsignificant (p > 0.2) (see Table 8). The Tukey post hoc test for the Southern Hemisphere revealed statistically significant differences between summer and autumn (mean difference: 11.2 min, 95% CI: 1.5– 21.0, p = 0.017), summer and winter (mean difference: 18.0 min, 95% CI: 8.2 – 27.8, p = 0.001), and winter and spring (mean difference: 12.2 min, 95% CI: 2.2– 22.1, p = 0.010). All other pairwise comparisons in the Southern Hemisphere were nonsignificant (p > 0.2) (see Table 9).

Table 7: Two-way ANOVA based on TRIWEIs from *News*/ *Data averaged across years*.

ANOVA	**df**	**F**	**Pr (>F)**
Season	3	15.6	$2.3 \cdot 10^{-9}$
Hemisphere	1	709.5	$7.0 \cdot 10^{-76}$
Season : Hemis	3	0.9	$4.4 \cdot 10^{-1}$

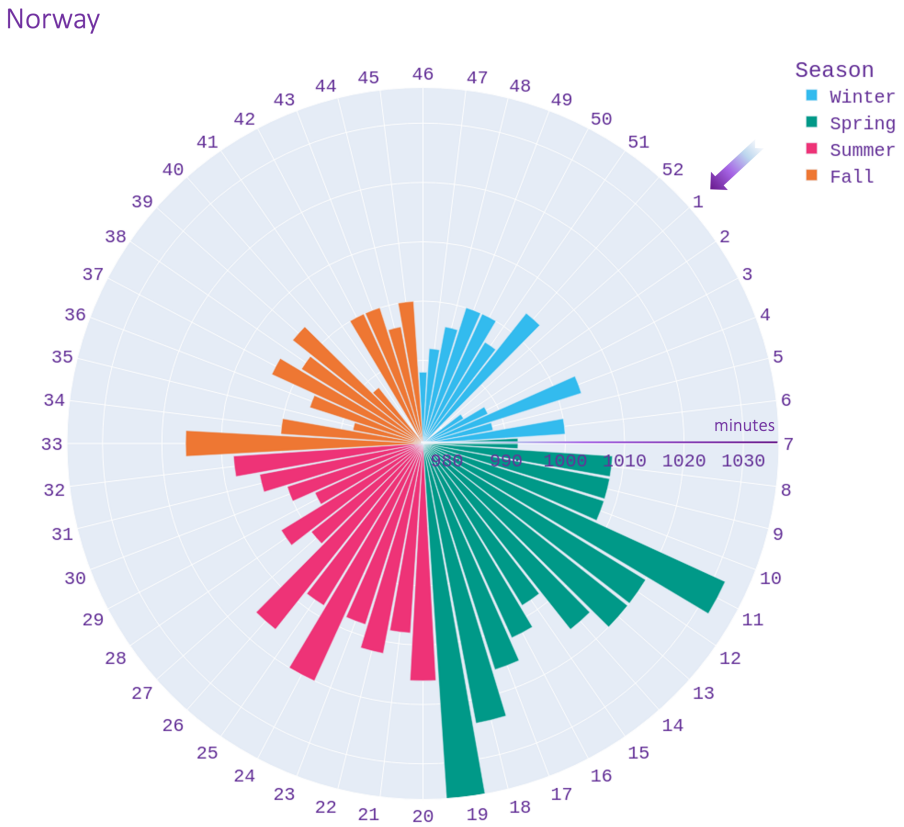
Table 8*:* **Tukey post hoc, Northern Hemisphere, based on TRIWEIs from *News*/ *Data averaged across years*.**

Season 1	**Season 2**	**Difference**	**Lower CI**	**Upper CI**	**p-value**
Winter	Spring	18.2	8.1	28.2	0.001
Winter	Summer	14.8	4.7	24.8	0.001
Winter	Autumn	7.1	-3.0	17.1	0.262
Spring	Summer	3.4	-6.8	13.6	0.803
Spring	Autumn	11.1	0.8	21.3	0.028
Summer	Autumn	7.7	-2.5	17.9	0.210

Table 9: Tukey post hoc, Southern Hemisphere, based on TRIWEIs from *News*/ *Data averaged across years*.

Season 1	**Season 2**	**Difference**	**Lower CI**	**Upper CI**	**p-value**
Summer	Autumn	11.2	1.5	21.0	0.017
Summer	Winter	18.0	8.2	27.8	0.001
Summer	Spring	5.8	-3.9	15.6	0.406
Autumn	Winter	6.8	-3.2	16.7	0.291
Autumn	Spring	5.4	-4.6	15.3	0.494
Winter	Spring	12.2	2.2	22.1	0.010

The polar plots based on activity lengths from TRIWEI-fitted *Data averaged across years* are presented in Figures 31, 32 and 33. Figure 31 shows that the northern countries appear to have larger activity lengths in spring and summer than in winter. In particular, Norway and Sweden appear to have the longest activity lengths in spring and Finland in summer. Figure 32 shows that in the southern hemisphere regions, summers and springs have the longest activity lengths. In Victoria, early autumn has similar activity lengths as late spring. Winter is the season with the shortest activity lengths in both southern regions. Figure 33, which represents the Northern and Southern Hemispheres, shows the longest activity lengths occurring in spring in the Northern Hemisphere and in summer in the Southern Hemisphere. Both hemispheres appear to have the shortest activity lengths during winter. Note that the x-axis (activity length in minutes) is scaled differently between the different plots.



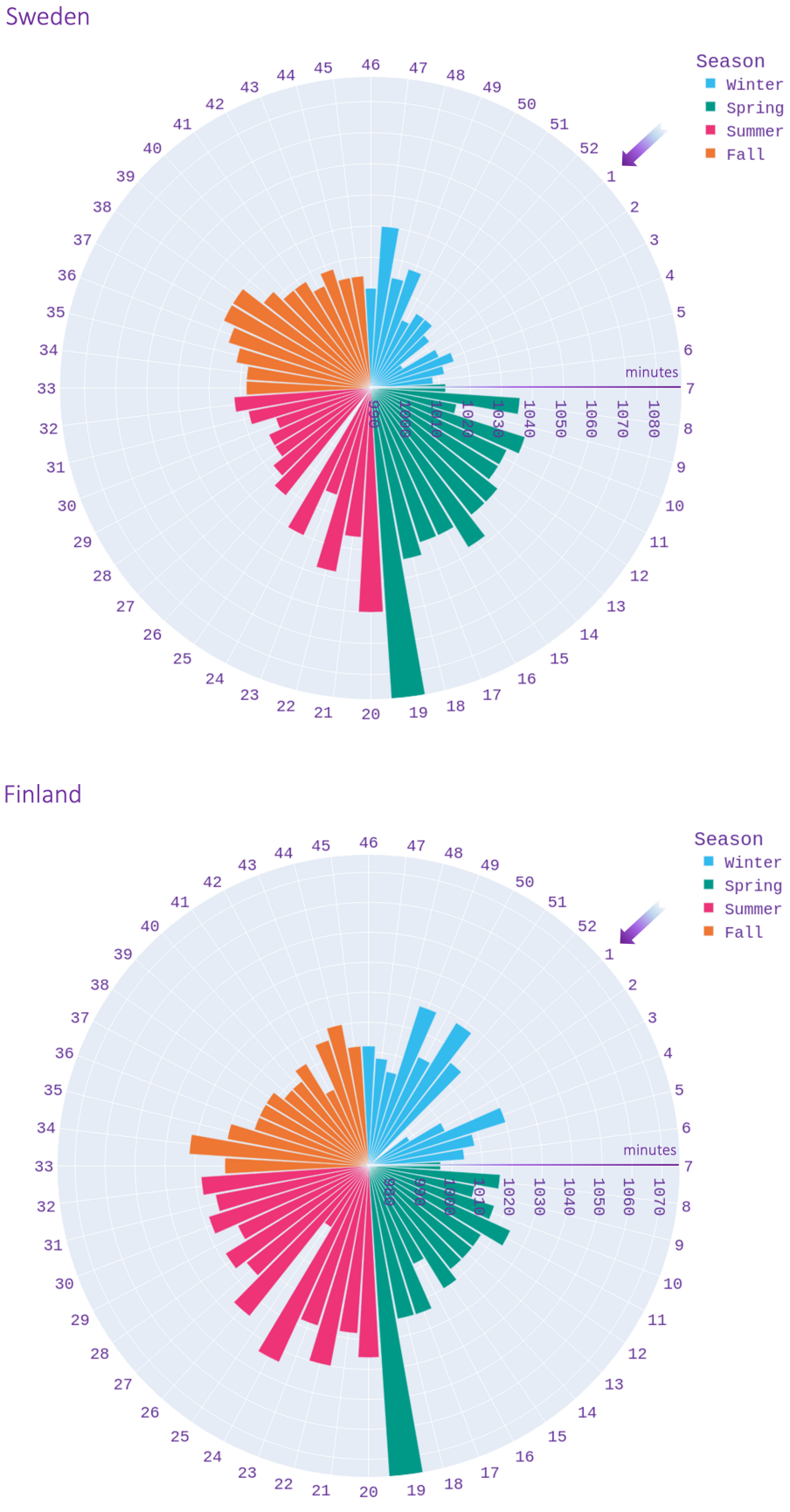
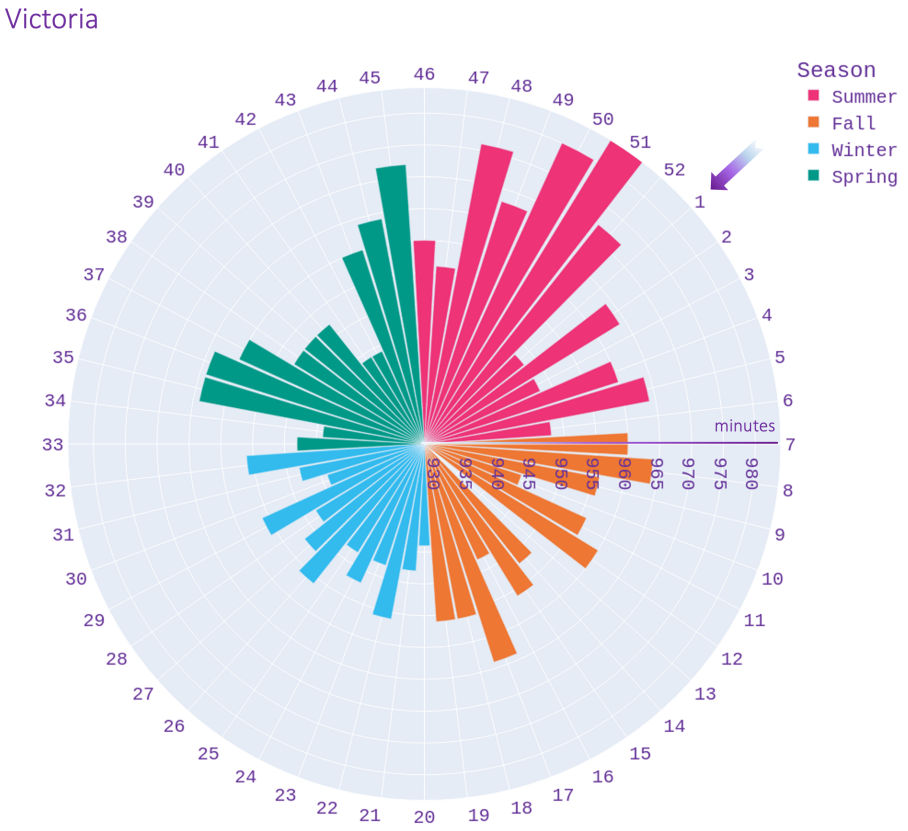


Figure 31: *News* polar plots, Norway, Sweden, Finland. Seasonal activity lengths in Norway, Sweden, and Finland based on TRIWEIs from *Data averaged across years* derived from Google Trends *News* category *-queries. The polar axis represents the weeks of the year, and the x-axis is activity length in minutes. Note that the x-axis (activity length in minutes) is scaled differently between the different plots.



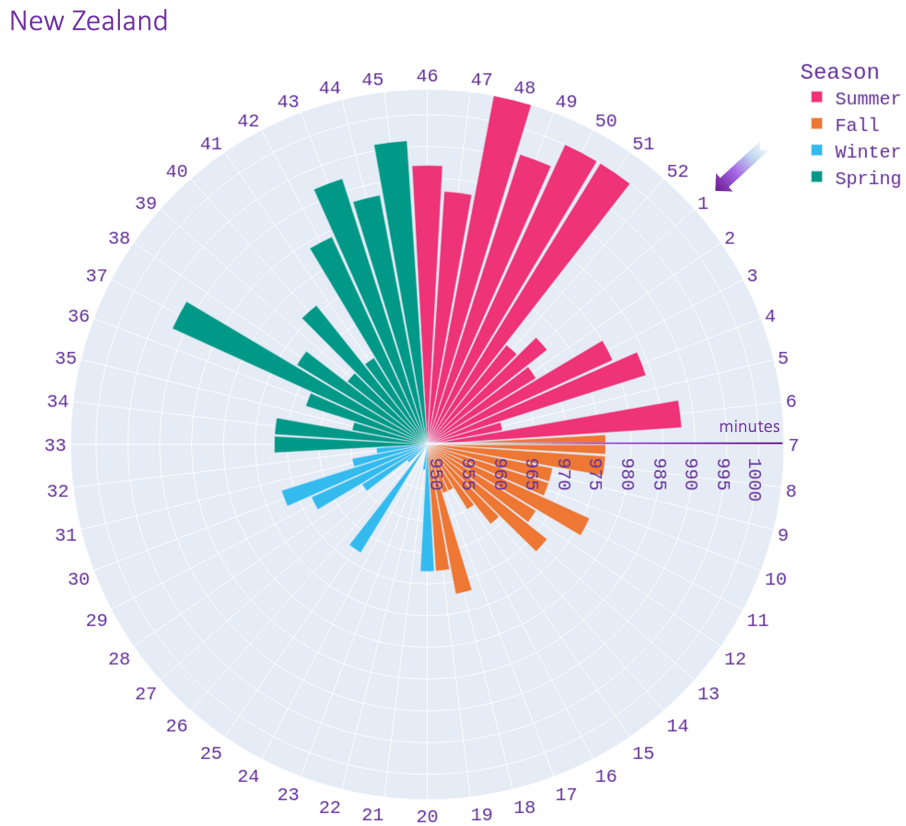
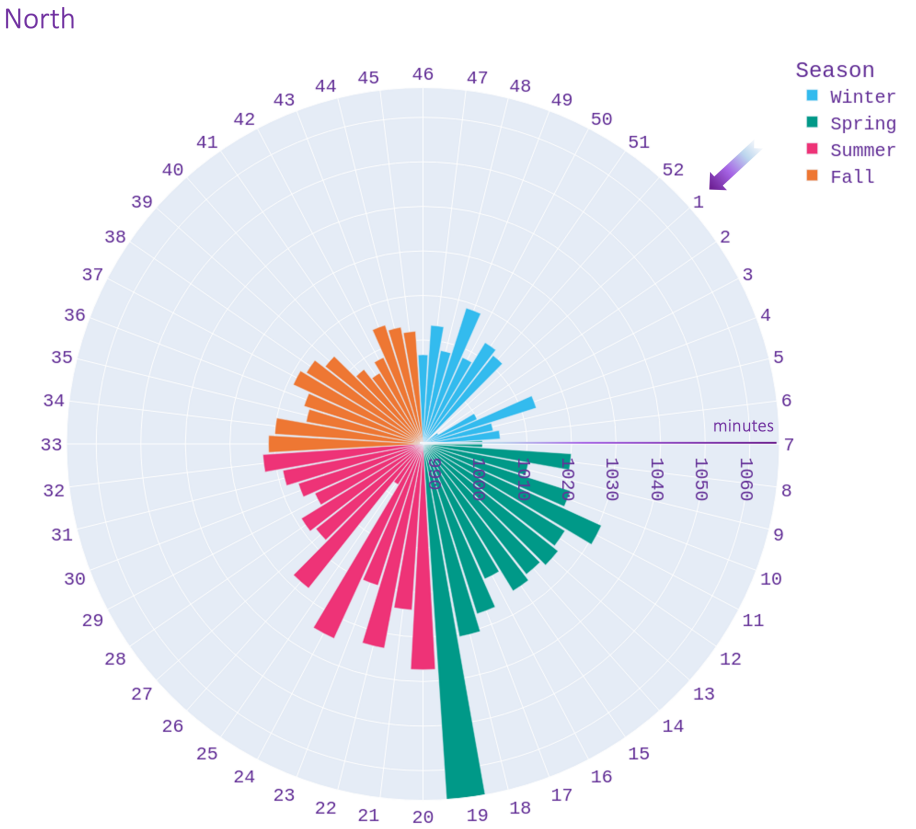


Figure 32: ***News* polar plots, Victoria, New Zealand.** Seasonal activity lengths in New Zealand, and Victoria based on TRIWEIs from *Data averaged across years* derived from Google Trends *News* category *-queries. The polar axis represents the weeks of the year, and the x-axis is activity length in minutes. Note that the x-axis (activity length in minutes) is scaled differently between the different plots.



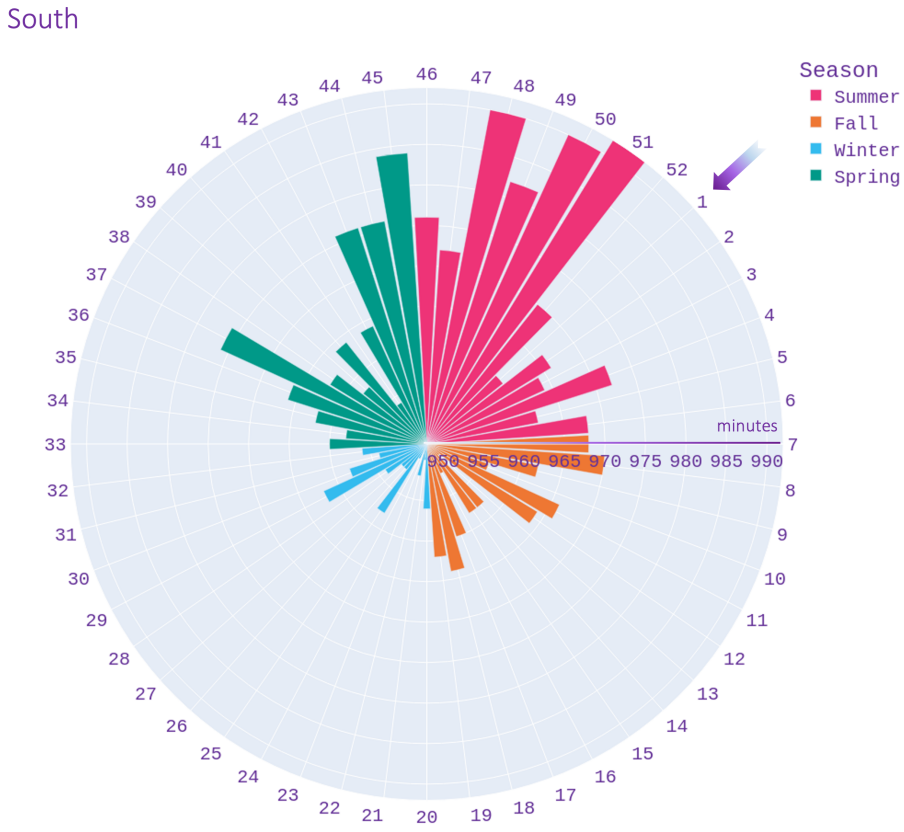


Figure 33: *News* polar plots, North, South. Seasonal activity lengths in the Northern and Southern Hemispheres based on TRIWEIs from *Data averaged across years* derived from Google Trends *News* category *-queries. The polar axis represents the weeks of the year and x-axis is the activity length in minutes. Note that the x-axis (activity length in minutes) is scaled differently between the different plots.

The average activity length and the average times of the steepest morning and evening slopes are summarised for the Northern Hemisphere in Table 10 and for the Southern Hemisphere in Table 11. Based on *Data averaged across years* in the Northern Hemisphere, the average activity length was longest in spring (17 h 5 min) and shortest in winter (16 h 46 min). Summer and autumn had intermediate activity lengths of 17 h 2 min and 16 h 54 min, respectively. The time of the maximum morning derivative was earliest in spring (07:02 AM) and latest in Winter (07:28 AM), and intermediate in summer (07:15 AM) and autumn (07:08 AM). The time of the minimum evening derivative was earliest in autumn (00:02 AM), latest in summer (00:16 AM), and intermediate in winter (00:14 AM) and spring (00:07 AM) (see Table 10).

In the Southern Hemisphere, the average activity length was longest in summer (16 h 14 min) and shortest in winter (15 h 54 min). Spring and autumn had intermediate activity lengths of 16 h 6 min and 16 h 1 min, respectively. The time of the maximum morning derivative was earliest in spring (07:00 AM) and latest in winter (07:14 AM), and intermediate in summer (07:04 AM) and autumn (07:09 AM). The time of the minimum evening derivative was earliest in spring (23:07 PM), latest in summer (23:19 PM), and intermediate in winter (23:08 PM) and spring (23:07 PM) (see Table 11).

Table 10: Northern Hemisphere activity lengths and the times of the steepest morning and evening derivatives. Northern hemisphere activity lengths in hours (based on *Data averaged across years*), time of morning gradient (maximum derivative) and time of evening gradient (minimum derivative).

	Spring	**Summer**	**Autumn**	**Winter**
Activity length	17:05	17:02	16:54	16:46
Morning gradient	07:02	07:15	07:08	07:28
Evening gradient	00:07	00:16	00:02	00:14

Table 11: Southern Hemisphere activity lengths and the times of the steepest morning and evening derivatives. Southern hemisphere activity lengths in hours (based on *Data averaged across years*), time of morning gradient (maximum derivative) and time of evening gradient (minimum derivative).

	Spring	Summer	Autumn	Winter
Activity length	16:06	16:14	16:01	15:54
Morning gradient	07:00	07:04	07:09	07:14
Evening gradient	23:07	23:19	23:10	23:08

5.3 Significant Seasonal Differences in Activity Lengths Derived from the *Arts and Entrainment* Category.

Data averaged across years from *-queries in the *Arts and Entertainment* category of Google Trends (location Norway) were fitted with TRIWEIs to find the activity lengths based on gradients, as explained in section 4.5. One dataset, from the Saturday of week 11, 2020, was excluded from the analyses because of a faulty Google Trends graph (see Appendix A12). The remaining activity lengths were used to produce the polar plot shown in Figure 34 and to perform a one-way ANOVA (see Table 12) that had a significant effect of season (the only predictor in this analysis). The following Tukey post hoc test revealed significant differences between winter and spring (mean difference: 12.9 min. 95% CI: 2.6– 23.1. $p = 0.009$) and between spring and autumn (mean difference: 13.2 min. 95% CI: 2.8– 23.7. $p = 0.008$) (see Table 13). All other pairwise comparisons were insignificant ($p > 0.1$).

Table 12: One-way ANOVA based on TRIWEIs from *Arts and Entertainment/ Data averaged across years.*

ANOVA	**df**	**F**	**Pr (>F)**
Season	3	4.8	0.005

Table 13: Tukey post hoc, Norway, based on TRIWEIs from *Arts and Entertainment/ Data averaged across years.*

Season 1	Season 2	Difference	Lower CI	Upper CI	p-value
Winter	Spring	12.9	2.6	23.1	0.009
Winter	Summer	7.9	-2.3	18.2	0.182
Winter	Autumn	0.4	-9.9	10.6	0.900
Spring	Summer	4.9	-5.5	15.4	0.586
Spring	Autumn	13.2	2.8	23.7	0.008
Summer	Autumn	8.3	-2.1	18.8	0.163

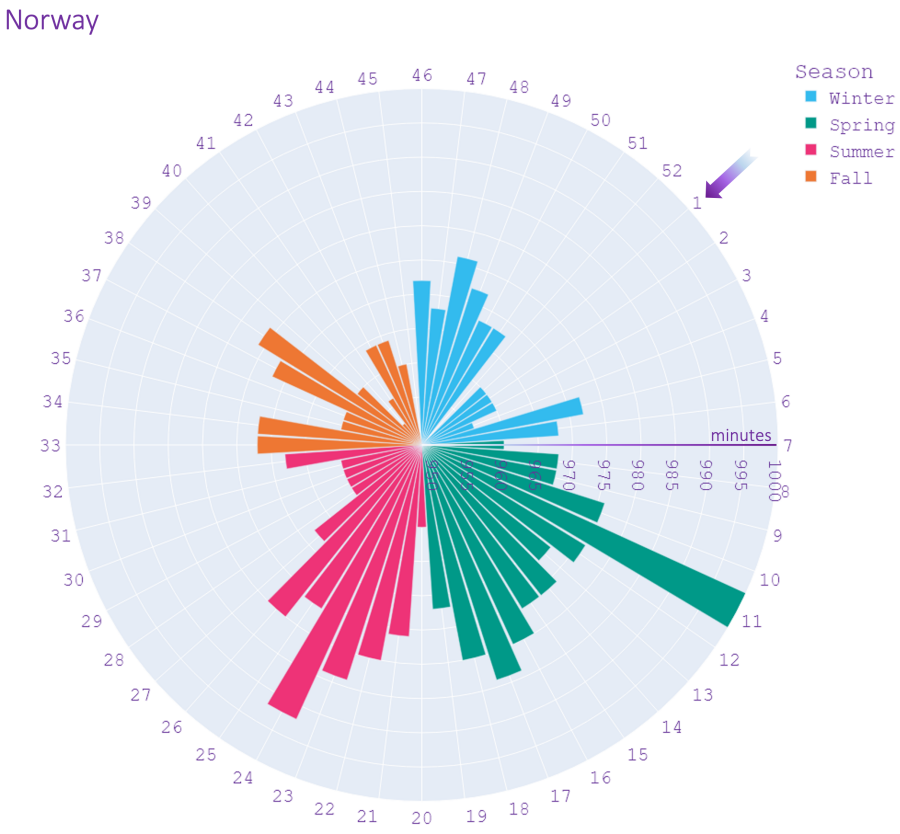


Figure 34: *Arts and Entertainment* polar plot, Norway. Polar plot based on *Data averaged across years* TRIWEI activity lengths.

5.4 Inverse Relationship between Hemispheres for Activity Lengths from *Raw Data.*

Seasonal wave plots are presented for *Data averaged within hemispheres* from the *News* category (see Figure 35), for the *Raw data* from the *Arts and Entertainment* category from Norway (see Figure 36), and for the raw weekly data for the pornography keywords (Figure 37).

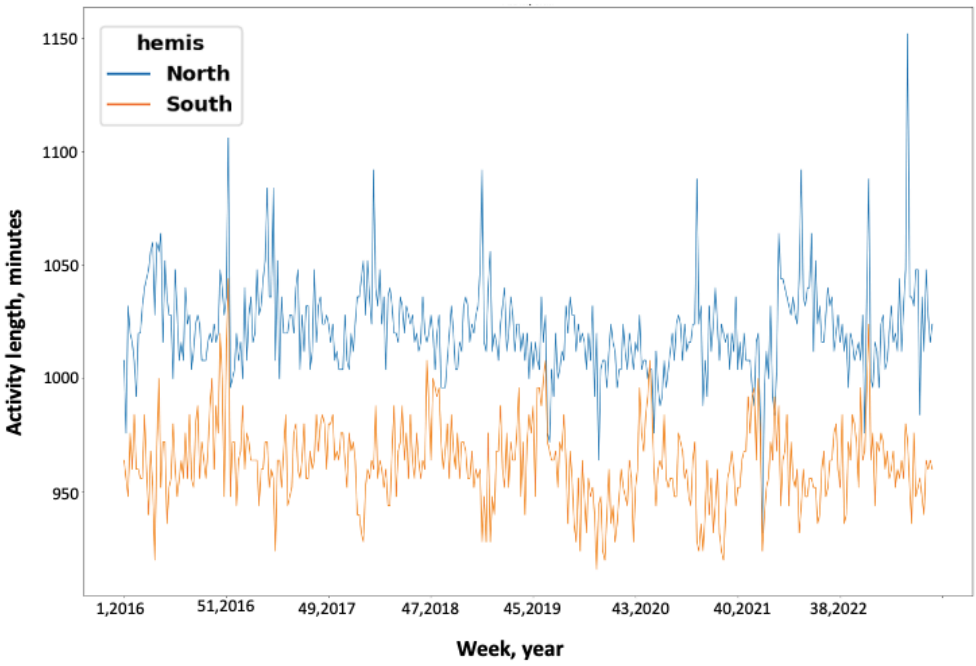


Figure 35: Wave plot, *News*. Seasonal wave plot of activity lengths derived from *Data averaged within hemispheres* from the *News* category. Activity lengths are plotted from January 2016 to August 2023.

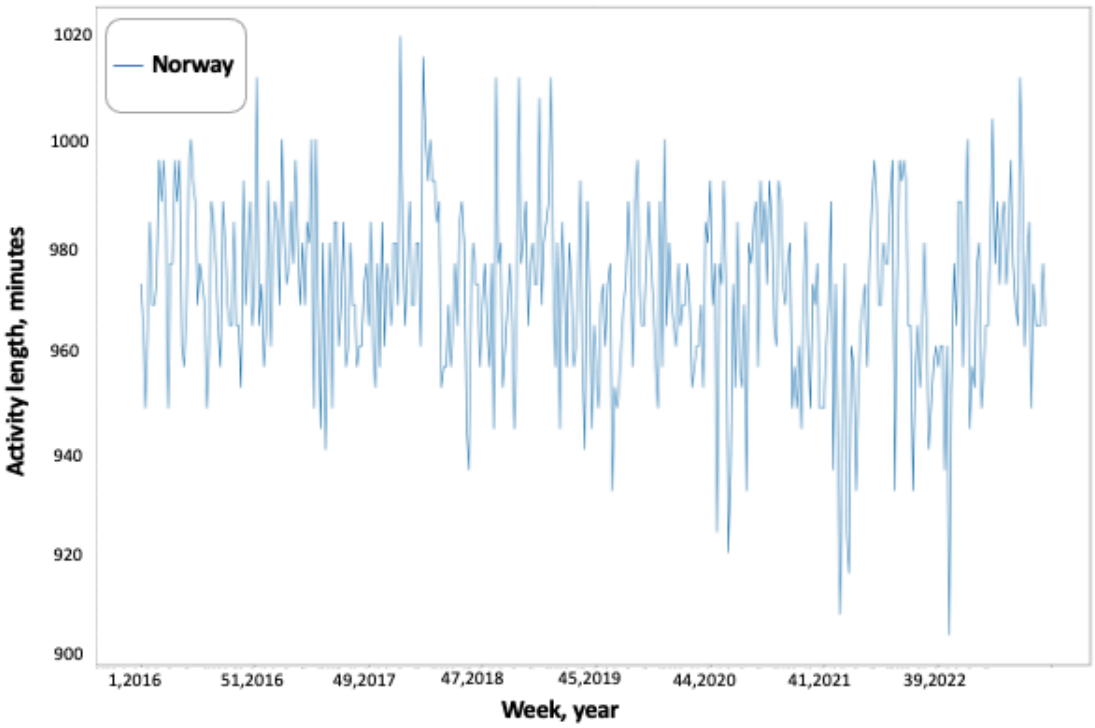


Figure 36: ***Wave plot, Arts and Entertainment.*** Seasonal wave plot of activity lengths derived from the *Arts and Entertainment* category. Activity lengths are plotted from January 2016 to August 2023.

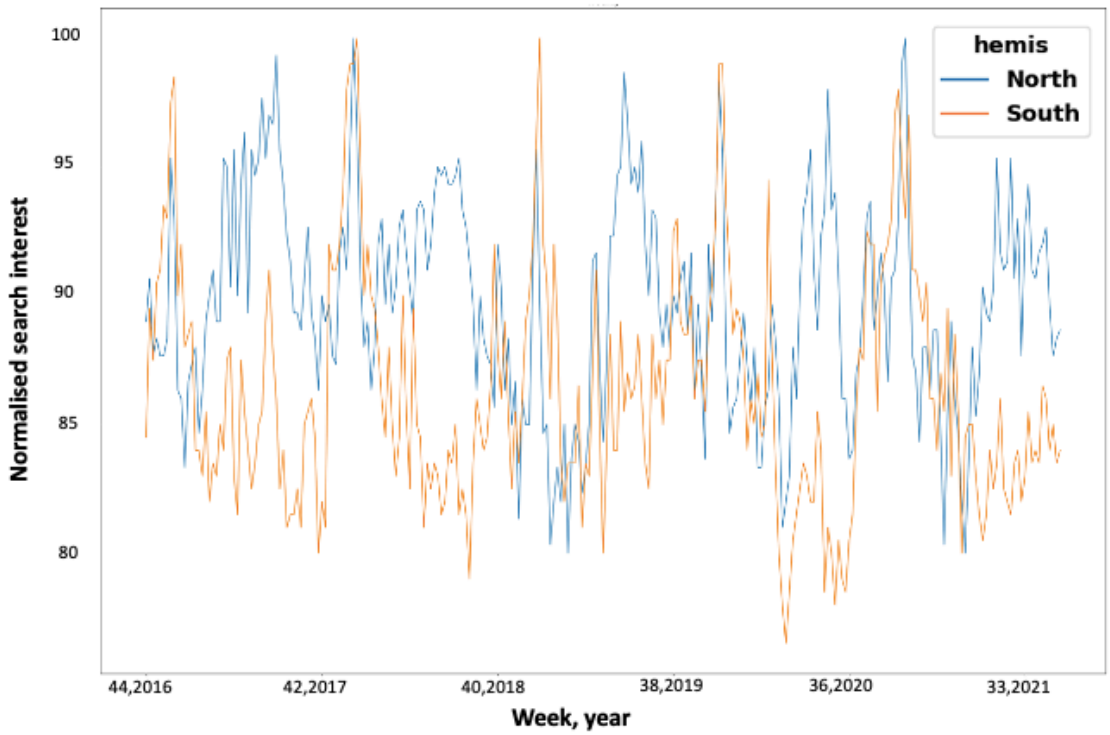


Figure 37: Wave plot, pornography keywords. Seasonal wave plot of normalised search interest for the pornography keywords *Porn, XNXX, PornHub, xVideos and xHamster*, plotted from January 2016 to August 2023.

5.5 Seasonal and Hemisphere Differences in Pornography Searches

The Welch's ANOVA of the Pornography data from the Northern Hemisphere showed a significant effect of season, as shown in Table 14. The following Games-Howell post hoc test for the Northern Hemisphere revealed statistically significant differences between autumn and spring (mean difference: 1.6, $p = 0.002$), autumn and summer (mean difference: 4.0, $p < 0.000$), spring and summer (mean difference: 5.6, $p < 0.000$) and between summer and winter (mean difference: 4.3, $p < 0.000$) (see Table 15). The other pairwise comparisons in the Northern Hemisphere were non-significant ($p>0.05$).

Table 14: Welch's ANOVA based on seasonal groups of raw, yearly pornography datasets from the Northern Hemisphere.

Welch's ANOVA	df	F	Pr (>F)
Season	3	64.5	$1.9 \cdot 10^{-34}$

Table 15*:* Games-Howell post hoc of seasonal groups from raw, yearly pornography data from the Northern Hemisphere.

Season 1	Season 2	Mean S1	Mean S2	Diff	p-value
Autumn	Spring	89.3	87.7	1.6	0.002
Autumn	Summer	89.3	93.3	4.0	< 0.000
Autumn	Winter	89.3	88.9	0.3	0.885
Spring	Summer	87.7	93.3	5.6	< 0.000
Spring	Winter	87.7	88.9	1.3	0.058
Summer	Winter	93.3	88.9	4.3	< 0.000

The Welch's ANOVA of the Pornography data from the Southern Hemisphere showed a significant effect of season, as shown in Table 16. The following Games-Howell post hoc test for the Southern Hemisphere revealed statistically significant differences between autumn and

summer (mean difference: 7.0, $p < 0.000$), spring and summer (mean difference: 5.8, $p < 0.000$), spring and winter (mean difference: 2.1, $p < 0.000$) and between summer and winter (mean difference: 8.0, $p < 0.000$) (see Table 17). The other pairwise comparisons in the Southern Hemisphere were non-significant ($p>0.07$).

Table 16: Welch's ANOVA based on seasonal groups of raw, yearly pornography datasets from the Southern Hemisphere.

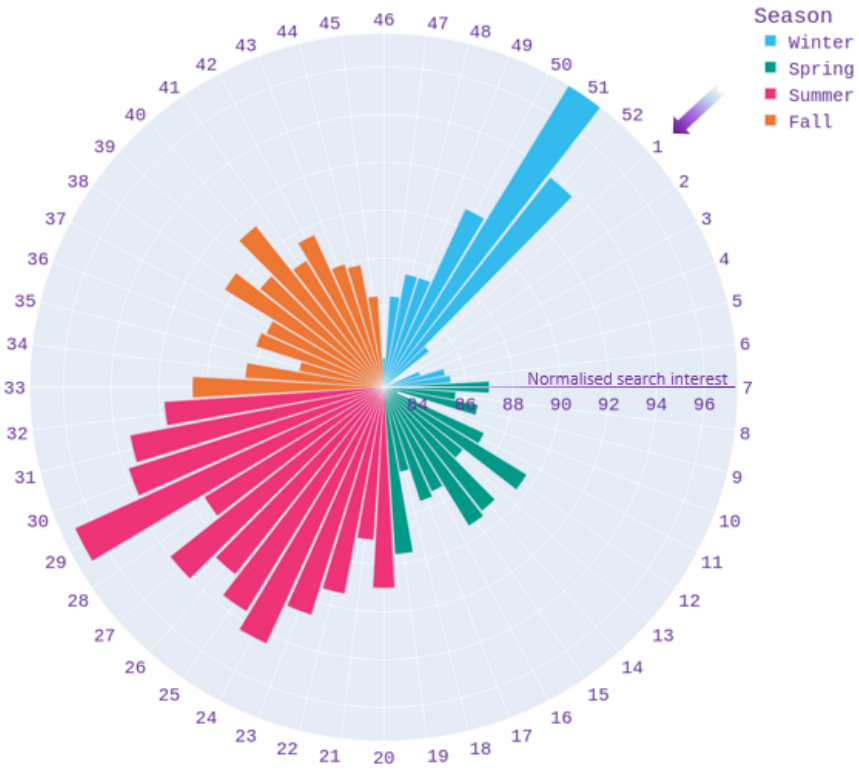
Welch's ANOVA	**df**	**F**	**Pr (>F)**
Season	3	82.6	$1.4 \cdot 10^{-38}$

Table 17: Games-Howell post hoc of seasonal groups from raw, yearly pornography data from the Southern Hemisphere.

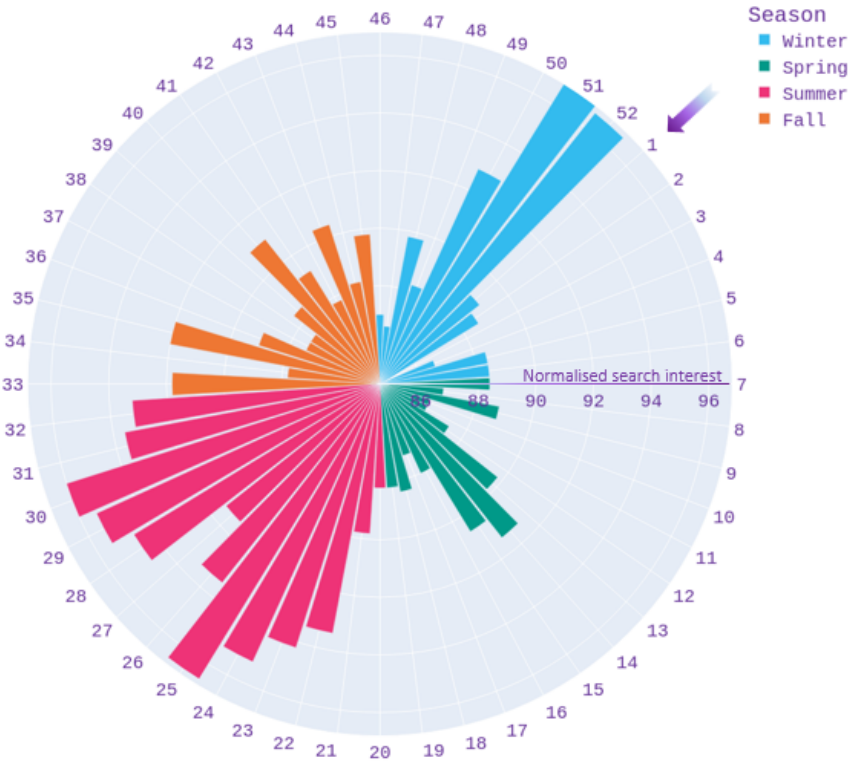
Season 1	**Season 2**	**Mean S1**	**Mean S2**	**Diff**	**p-value**
Autumn	Spring	84.6	85.8	1.2	0.074
Autumn	Summer	84.6	91.6	7.0	< 0.000
Autumn	Winter	84.6	83.6	0.9	0.217
Spring	Summer	85.8	91.6	5.8	< 0.000
Spring	Winter	85.8	83.6	2.1	< 0.000
Summer	Winter	91.6	83.6	8.0	< 0.000

The polar plots based on pornography data are presented in Figures 38, 39 and 40. Figure 38 shows that the northern countries appear to have a higher interest in pornography during summer and around the Christmas holiday. Figure 39 shows that in the southern hemisphere regions, summers have the highest interest in pornography. Figure 40, which represents the Northern and Southern Hemispheres, shows the same trends as described above. The northern hemisphere countries appear to be the least interested in pornography during spring. In the Southern Hemisphere, the season with the least interest is winter. Note that the x-axis (normalised interest) is scaled differently between the different plots.

Norway



Sweden



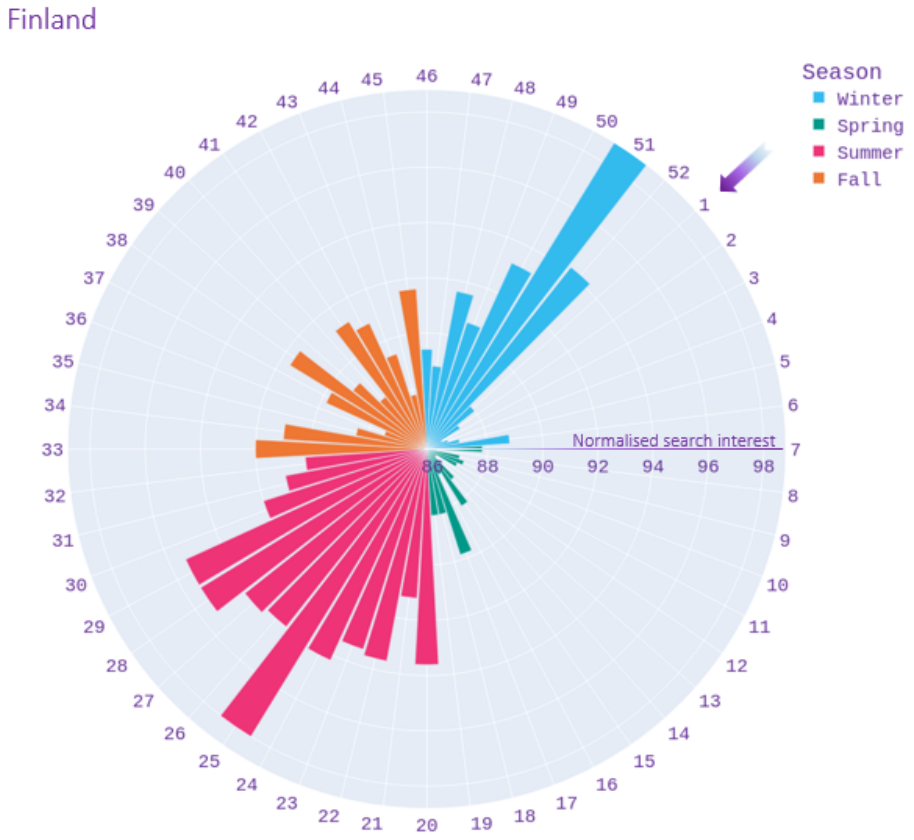
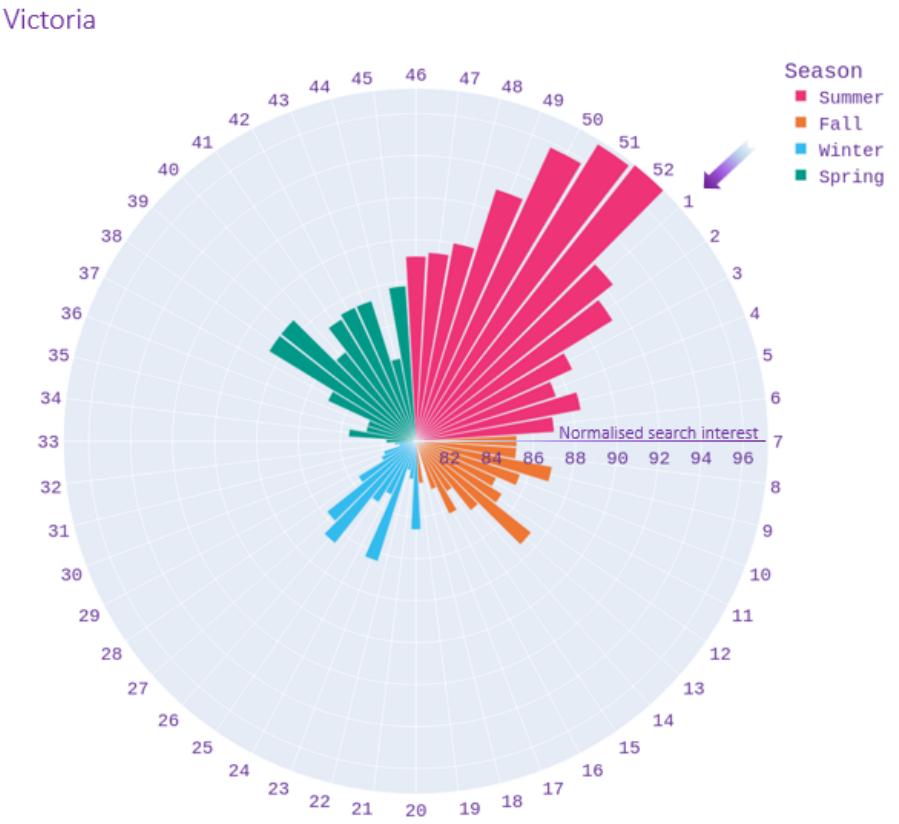


Figure 38: Pornography polar plots, Norway, Sweden, Finland. Averaged normalised search interest for pornography keywords between 2016-2021 in Nordic countries. Note that the x-axis (normalised search interest) is scaled differently between the different plots.



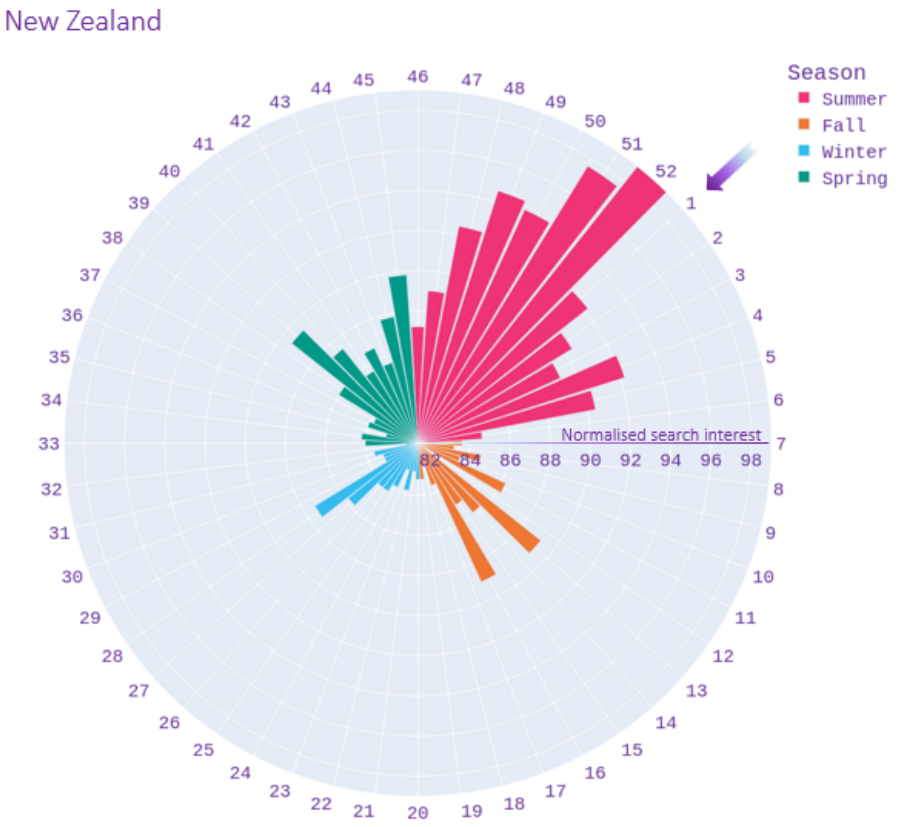


Figure 39: Pornography polar plots, Victoria, New Zealand. Averaged normalised search interest for pornography keywords between 2016-2021 in Southern regions. Note that the x-axis (normalised search interest) is scaled differently between the different plots.

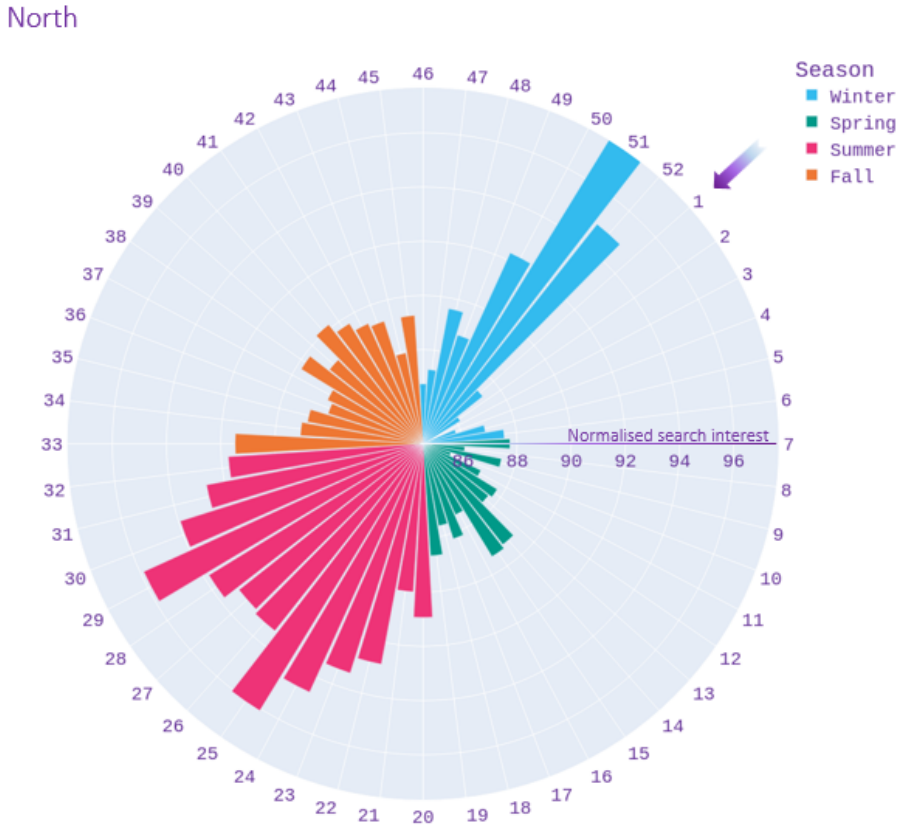


Figure 40: Pornography polar plots, North, South. Averaged normalised search interest for pornography keywords between 2016-2021 in the Northern and Southern Hemisphere. Note that the x-axis (normalised search interest) is scaled differently between the different plots.

5.6 No Clear Relationship between Activity Lengths and Photoperiod or between Pornography Interest and Photoperiod in Norway and Regionally Averaged Data

Figures 41, 42 and 43 show activity lengths based on *News* category queries plotted against some extreme photoperiods in the region of interest, as described in section 4.5.7. Figure 41 shows *Data averaged across years* for Norway plotted against the photoperiod of each Saturday in Nordkapp (beige) and Lindesnes (yellow) in 2020. Figure 42 shows the averaged *Data averaged across years* for the Northern hemisphere plotted against the photoperiod of each Saturday in Nordkapp (beige) and Smygehuk (yellow) in 2020. Figure 43 shows the averaged *Data averaged across years* for the Southern hemisphere plotted against the photoperiod of each Saturday in Lindsay point (beige) and Stewart Island (yellow) in 2020.

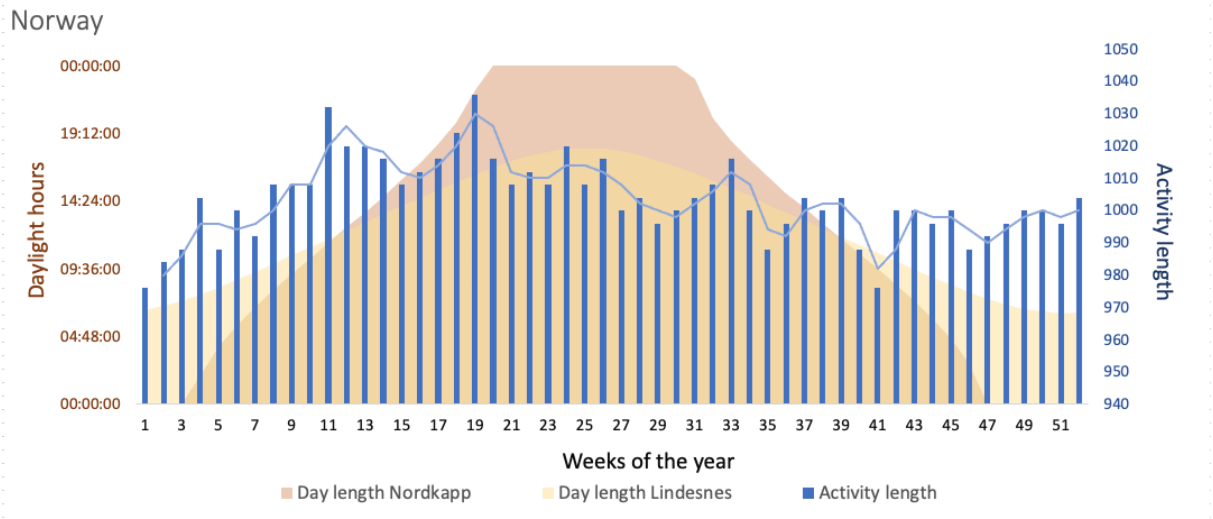


Figure 41: Activity lengths and photoperiod, Norway. Activity length (in minutes) in Norway averaged across years (blue bars) plotted against photoperiod in Nordkapp (beige) and Lindesnes (yellow).

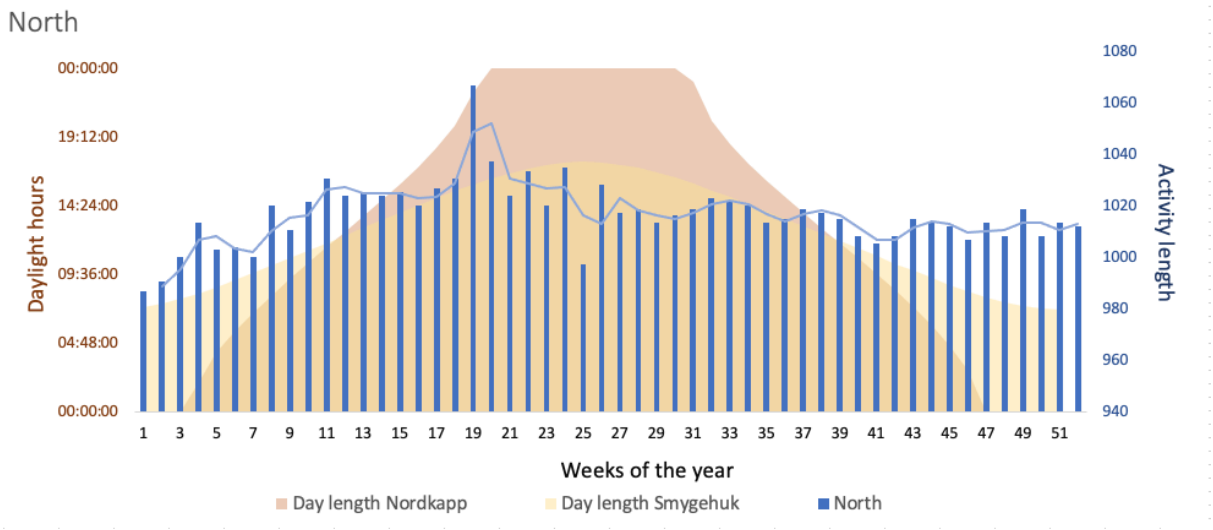


Figure 42: Activity lengths and photoperiod, North. Activity lengths (in minutes) in the Northern hemisphere (Norway. Sweden. Finland) averaged across years (blue bars) plotted against photoperiod in Nordkapp (beige) and Smygehuk (yellow).

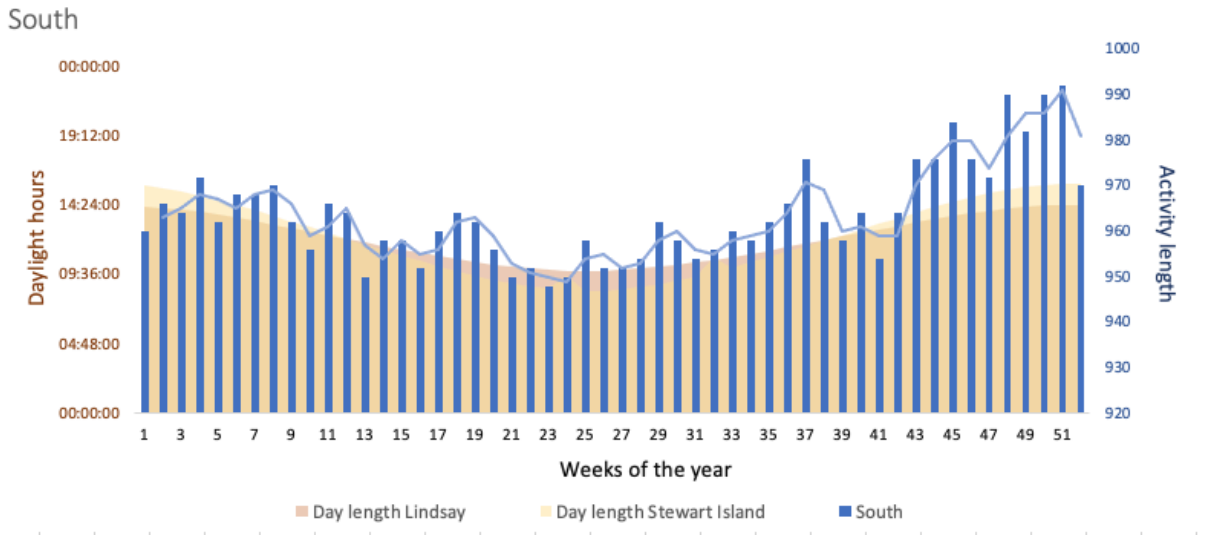


Figure 43: Activity lengths and photoperiod, South. Activity lengths (in minutes) in the Southern hemisphere (New Zealand. Victoria) averaged across years (blue bars) plotted against photoperiod in Lindsay point (beige) and Stewart Island (yellow).

Figures 44, 45 and 46 show normalised pornography interest, averaged over weeks between 2016 and 2021 and plotted against some extreme photoperiods in the region of interest. Figure 44 shows pornography interest in Norway, plotted against the photoperiod of each Saturday in Nordkapp (beige) and Lindesnes (yellow) in 2020. Figure 45 shows pornography interest from 2016-2021 for the Northern Hemisphere plotted against the photoperiod of each Saturday in Nordkapp (beige) and Smygehuk (yellow) in 2020. Figure 46 shows pornography interest from 2016-2021 for the Southern Hemisphere plotted against the photoperiod of each

Saturday in Lindsay point (beige) and Stewart Island (yellow) in 2020. The graphs for the Northern and Southern Hemispheres contain data on normalised pornography interest that has been firstly averaged between weeks for each region and secondly averaged between regions.

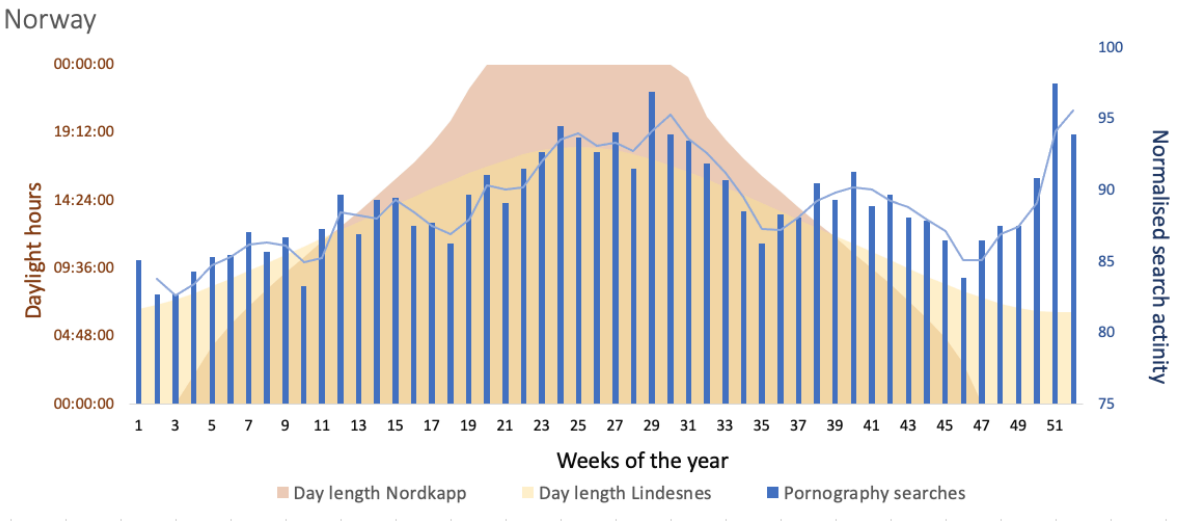


Figure 44: Pornography interest and photoperiod, Norway. Normalised pornography interest in Norway averaged across 2016-2021 (blue bars) plotted against photoperiod in Nordkapp (beige) and Lindesnes (yellow).

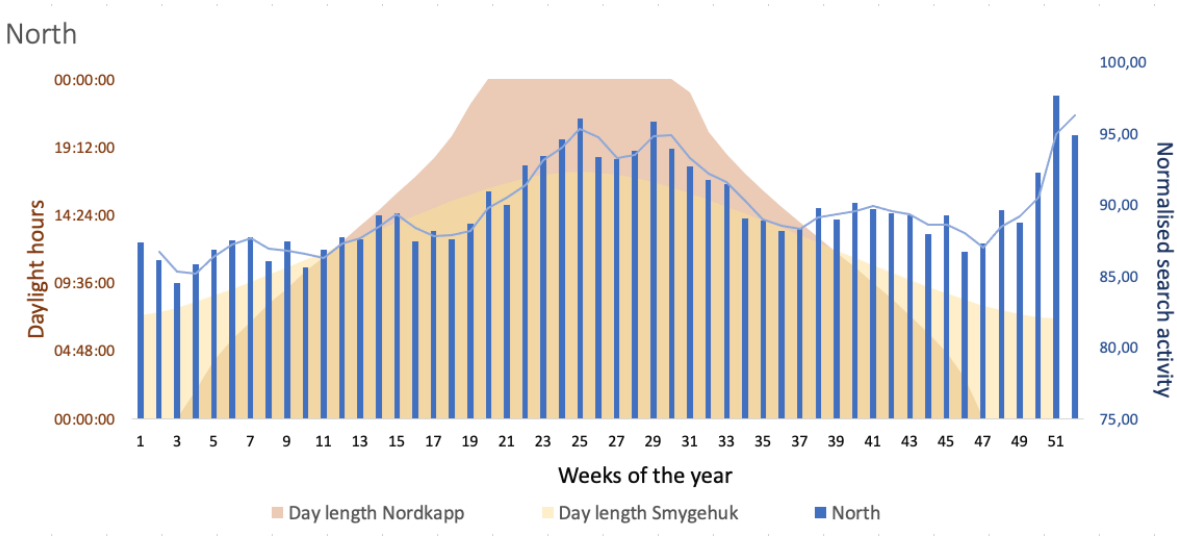


Figure 45: Pornography interest and photoperiod, North. Normalised pornography interest in the Northern hemisphere (Norway. Sweden. Finland) averaged across 2016-2021(blue bars) plotted against photoperiod in Nordkapp (beige) and Smygehuk (yellow).

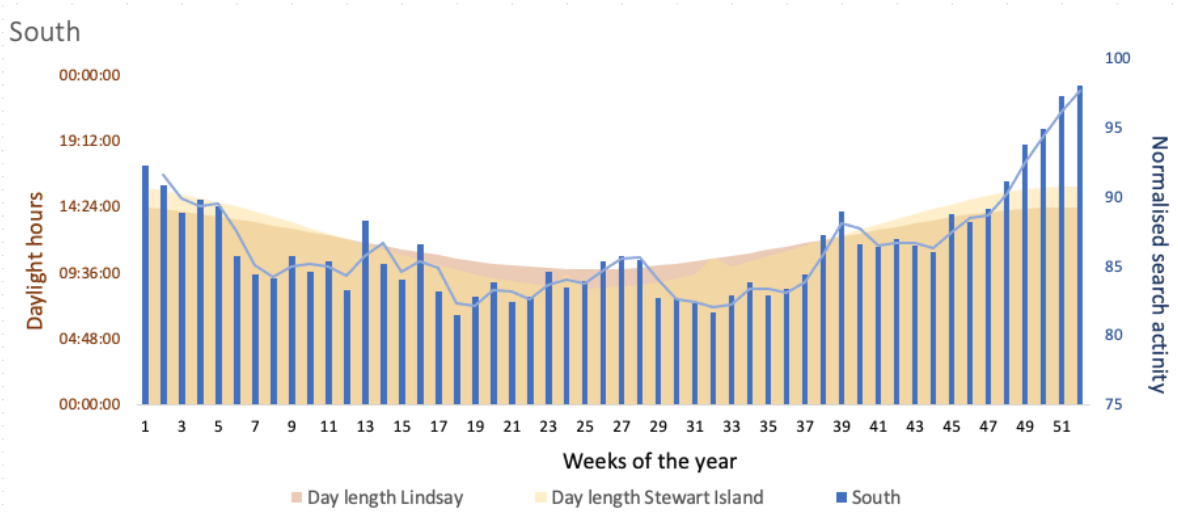


Figure 46: Pornography interest and photoperiod, South. Normalised pornography interest in the Southern hemisphere (New Zealand. Victoria) averaged across 2016-2021(blue bars) plotted against photoperiod in Lindsay point (beige) and Stewart Island (yellow).

6 Discussion

The objectives of this thesis, as stated in section 3.7, were to investigate seasonal differences in activity-rest periods and libido as well as any potential hemisphere differences. The results indicated that daily search activity and pornography search interest vary between seasons. In addition, the results indicated that there were seasonally inverse patterns in daily search activity and pornography search interest between the Northern and Southern Hemispheres. Since *activity lengths* from daily Google Trends datasets were used as a proxy for activity-rest periods, and since pornography search interest was used as a proxy for libido, the findings might be indicative of seasonal fluctuations in daily activity-rest behaviour and libido.

6.1 Activity Analyses

The activity analyses for the *News* category (*Data averaged across years*) were similar for the Northern and Southern Hemispheres. They both revealed a longer daily activity span in *summer* compared to *winter* and in *spring* compared to *winter* ($p \leq 0.01$). The statistically significant average differences in activity length were ~18 minutes at the most (between *winter* and *spring* in the Northern Hemisphere and between *summer* and *winter* in the Southern Hemisphere) and ~11 minutes at the least (between *spring* and *autumn* in the Northern Hemisphere and between *summer* and *autumn* in the Southern Hemisphere). The shorter seasonal differences of ~11 minutes are similar to that found between summer and winter in a study by Suzuki et al. (2019), who investigated seasonal changes in sleep duration in 1,388 Japanese citizens (84). Similarly, Johnsen et al. (2013) found that subjects in Tromsø slept ~12 minutes longer on free days during winter compared to summer (90).

An important question that arises is whether these are *meaningful* differences. If we first assume that the measurements did indeed represent daily activity lengths (this assumption is further discussed in section 6.1.1), then my answer would be *yes*. A seasonal fluctuation in daily activity length of 11-18 minutes (not considering confidence intervals that went below and above this range) is not extreme compared to other seasonal animals. For example, Daan et al. (1974) showed that the daily activity lengths of various bird species (Eurasian chaffinch, European greenfinch, and great spotted woodpecker) were ~6 hours in winter and ~18 hours in summer, which comprises a staggering 12-hour seasonal difference (32). Although mammals have smaller seasonal variations in daily activity lengths than birds, they can still be

pronounced. Erkert et al. (2004) demonstrated that the primate species Verreaux's sifaka had motor activities centred around a 15-hour interval during long days and a nine-hour interval during short days in Madagascar (85).

However, humans live very different lives from all other animals (except pets and domesticated animals), and so the observed activity lengths of humans cannot necessarily be directly compared to those of other species. As discussed in section 3.4.1, the artificial lighting environments of humans significantly reduce total sleep time. Since artificial light exposure might imitate a *constant* summer photoperiod (55), yearly changes in light exposure should be smaller for humans than for animals that are only exposed to natural light. Thus, if seasonal fluctuations in daily activity levels are caused by yearly changes in light exposure, activity fluctuations should be smaller in humans than in other animals. In addition, the search data from Google Trends was likely based principally on people whose lives are largely removed from natural cycles of temperature, food access or other commonly acknowledged zeitgebers. Add to this that all search activity (the unit of measure in our analyses) is done on electronic screens, which are linked to shorter sleep durations, later sleep onset and increased sleep deficiency (86). Therefore, it was remarkable that this study was still able to detect significant seasonal patterns.

It seems unlikely that the seasonal differences in activity length were only artefacts of the *News* category since the ANOVA and Tukey post hoc analyses of *Data averaged across years* from the *Arts and Entertainment* category in Norway detected similar results (see Tables 12 and 13). It is worth noting that the *Arts and Entertainment*-based *Data averaged across years* analyses were based on fewer data than the corresponding *News* activity lengths. The *Raw data* analyses were based on more data and detected more seasonal differences (as shown in Appendices A8 and A11). The considerations regarding which analyses to present are discussed in section 6.1.6.

Although the results from the different categories were relatively concordant, differences in search patterns between different Google Trends categories cannot be completely ruled out. Indeed, the *News* category was chosen as the main category for our analyses since news searches were assumed to be more stable throughout the year than searches within other categories. *Entertainment* might, for example, be more popular during certain times of the year due to weather conditions or holidays. The potential differences in the reliability of categories are difficult to estimate since Google Trends provides no information about what

searches constitute each category (this issue is discussed further in section 6.2). One way to mitigate any categorical differences would have been to average several categories (as planned while using *Pytrends*), but some categories have very low-quality data where any daily patterns are essentially impossible to spot. The best solution would have been if Google Trends offered an "All searches" category that was based on all search activity regardless of keyword. This option was, unfortunately, not available as of October 2023.

6.1.1 Assumption of Activity Measurement

A critical assumption in the analyses of *News* and *Arts and Entertainment* datasets was that the distance between the steepest morning and evening slopes of each fitted TRIWEI was representative of the general daily activity of Google users. The assumption entailed that each activity length roughly stretched between the average wake and bedtimes in the population. Between seasons, activity lengths were approximately 17 hours in the Northern Hemisphere and 16 hours in the Southern Hemisphere (this hemisphere discrepancy is discussed further in section 6.1.2). These activity lengths would correspond to daily *inactivity lengths* of approximately 7 and 8 hours, respectively. Without taking into consideration any sleep debt from the weekdays and the consequent social jet lag, 7-8 hours is within the commonly recommended sleep durations for adults (7-9 hours) (87). Therefore, it seems our assumption may be valid.

Whether people follow the sleep duration recommendations is arguable. A study on 16-19-year-old Norwegians showed that self-reported sleep duration on weekends was ~8 hours and 32 minutes between February and May 2012 (88). However, average sleep duration might have declined after 2013, potentially in relation to screen use, and the percentage of Americans who owned a smartphone surged from 35% in 2011 to 77% in 2016 and had reached 97% by 2021 (89, 90). Although the estimated *inactivity lengths* seem to agree with observed and recommended *sleep durations,* without corresponding individual sleep duration data we cannot be certain. Nevertheless, this similarity is an interesting observation and might allude to a correlation between Google searches and the general activity onsets and offsets in the population.

6.1.2 Hemisphere Differences in Activity Length

Perhaps the largest conundrum about the presented results is the general difference in activity lengths, as shown in Figure 35. Regardless of the season, the measured activity lengths were ~1 hour shorter in the Southern Hemisphere compared to the Northern Hemisphere (Tables 10 and 11). I can only speculate as to what caused this difference.

The northernmost point in the Oceanic regions used in the analyses was Lindsay point, which is located at about 34ºS. This latitude approximately corresponds to Rabat (Morocco) or Los Angeles (USA) in the Northern Hemisphere. On the other hand, the southernmost point in the Nordic region was Smygehuk, located at about 55° N. On the day of their respective summer solstices, Smygehuk has a photoperiod of 17 h 27 min, whereas Lindsay Point has a photoperiod of 14 h 26 min. On their respective winter solstices, the photoperiods are 7 h 5 min in Smygehuk and 9 h 52 min in Lindsay Point. The difference in the experienced photoperiods is, in other words, noteworthy and could perhaps be a part of the explanation for the observed hemispherical differences in general activity lengths.

We could, for example, imagine that the Nordic countries rely more on electrical lighting than the Oceanic regions during winters and that this could make the general activity length in winter longer in the Nordic regions. Similarly, we could imagine that the longer natural photoperiod in summer makes the summer activity lengths longer in the Nordic regions. Long winter and summer photoperiods (natural or electrical) and their consequent influence on human activity could then be thought to influence the spring and autumn behavioural patterns in the Northern Hemisphere due to, for example, the establishment of culturally accepted activity lengths. Another suggestion could be that some other cultural factor caused the difference between the Nordic countries and the Oceanic regions. This could, for example, be a higher expectation to be productive and efficient in the Nordic regions, leading to longer-lasting daily activity. There could also be a generally higher use of screens around bedtime in the Northern Hemisphere. Again, all suggestions remain highly speculative, and I have no clear opinion of what caused this hemispherical difference in activity lengths.

6.1.3 Assumptions about Activity Data Representativity

Given that Google has indeed provided a representative sample of searches (see section 6.3 for further discussion), another question that arises is whether the representative search

sample represents the general population. Unfortunately, this question is difficult to answer since there is no metadata associated with the Google Trends datasets, such as the age or sex distribution.

We can, however, speculate about potential issues regarding representativity. The fundamental question in this respect is: *who* uses the internet in a way that creates the observed morning and evening slopes on the daily Google Trends graphs? We could, for example, imagine that older people are more likely to start their days *googling* for news and that younger people would be more likely to get their news from other sources, such as social media apps. We could also imagine a difference between chronotypes. For example, maybe earlier chronotypes are less likely to spend time on their phones in the morning because they are eager to get their day started. Or, on the contrary, maybe later chronotypes feel more pressure to get the day started since they wake up later and choose not to spend time on the internet for that reason. These suggestions illustrate an important shortage of this study: we do not know on whose searches the data are based.

The data also represent a limited time frame in the searcher's lives, since all data were from Saturdays. Saturdays were chosen based on Roenneberg and colleagues' assertion that human behaviour on free days better reflects their overall circadian phase (57). This is supported by Johnsen et al. (2013) who detected seasonal differences in sleep duration for employed participants on free days, but not on workdays (91). Our data were presumably derived from a highly variable population, with an unspecified portion working during weekends. Thus, the use of Saturdays to approximate "free days" may be flawed in the current study.

6.1.4 Causation behind Seasonal Activity Lengths

Although the antiphase pattern between the Northern and Southern Hemispheres (best seen in the wave plot in Figure 35) suggests that some seasonal factor was contributory to the observed trends, there is no way of identifying which factor drives these patterns. It could be photoperiod, temperature, humidity, or any other environmental or societal factor that fluctuates seasonally.

Plots were generated to display the relationship between photoperiod lengths and activity duration (Figures 41-43). There was potentially a very weak relationship but as the activity

data were from a wide range of photoperiods, this was a flawed approach. Latitudinally split Google Trends data combined with latitudinally specific photoperiod data could be combined for improved figures and a cross-correlation analysis in future research. It is worth mentioning that the first experiments with latitudinally split *Pytrends* data did *not* indicate any clear latitudinal differences in activity length patterns (see graph in Appendix A1).

There are a few potential obstacles to analyses of natural photoperiod correlations with activity. The larger cities, where most people live, are often heavily light-polluted reducing natural light exposure (92). Thus, Google Trends-derived activity lengths from the larger cities could potentially poorly reflect the impact of natural photoperiod changes. Additionally, people who live in large cities could be more affected by social entrainment simply because there are more people around them. On the other hand, activity lengths based on search activity in more rural places with less light pollution could be unreliable and noisy simply because they are based on fewer searches. Thus, it is not certain that such an analysis could be performed at all with the methods used in this thesis because the TRIWEI distribution could not be fitted properly. In fact, a test query revealed that local daily datasets from Troms and Oslo were both highly noisy compared to a national daily dataset from all of Norway (see Appendix A13 for example figures). The higher noisiness of local data might have been an obstacle in detecting any latitudinal differences in search activity, as shown in Appendix A1.

It is possible that Google does, in fact, possess local data of sufficient quality to perform local TRIWEI-based activity length analyses, but that these data are not available through the Google Trends website. Perhaps future studies would benefit from contacting Google and suggesting a collaboration to obtain higher-quality data.

Cultural events might have influenced the activity lengths of some Saturdays. For example, there is a large spike in the *News*-based polar plot from the Northern Hemisphere in week 19. The weekend of week 19 sometimes, but not always, coincides with a series of holidays in the Nordic countries (i.e. Ascension Day, Pentecost, and the Norwegian national day). When some or all of these holidays fall closely together, it is customary to take an extended break from work and school. My suspicion is that this “unexpected” free time is thoroughly exploited, perhaps resulting in later bedtimes. Another small spike occurs in week 11, which in some years coincides with Easter (e.g. 2016). Interestingly, Saturday of week 11 can also coincide with (and is always close to) the spring equinox (e.g. 2021). Perhaps the longer photoperiod has really started to affect people in the Nordic regions by week 11, and from this

point, the daily activity lengths continue to increase throughout the summer. I am not as familiar with the holidays in the Southern Hemisphere. However, Figure 33 demonstrates that the longest activity length occurred on the Saturdays closest to Christmas. Another spike in the polar plots from the Southern occurred in week 37, which is only a few days away from the spring equinox in the Southern Hemisphere.

The spiking activity lengths around the spring equinoxes in both the Northern and Southern Hemisphere are complemented by the observation of generally longer activity lengths in spring than in autumn. This observation was only statistically significant for the Northern Hemisphere in the Tukey post hoc test (and also in the alternative *Raw data*-based analyses using Welch's ANOVA and Games-Howell post hoc, as seen in Appendices A8 and A11), although the polar plot from the Southern Hemisphere seems suggestive of a potential difference. In this context, it is interesting to note that birds often have longer diel activity spans before the summer solstice than after, even though the photoperiods are equal (32). Although the potential link to activity length is unknown to me, it is interesting to note that some brain areas have seasonal fluctuations in size. The paraventricular nucleus (PVN), which has been associated with relaying photoperiodic information from the brain to the rest of the body, was shown to be larger in spring than during any other time of the year in humans (18, 93). Perhaps activity lengths in humans, birds, and other vertebrates could have evolved to be longer in spring. During late spring and summer, animals in their natural environments typically experience higher food abundance. Animals could benefit from being more active at this time of year to gain large fat reserves that can be depleted during late autumn and winter. In autumn, on the other hand, the body might prepare for the season with the lowest food availability and start decreasing its energy expenditure by shortening the diel activity period.

A link between season, activity levels and metabolism could have large implications for human health. If some innate, biological mechanism leads to lower energy expenditure in late autumn and winter compared to spring and summer, this could have ramifications for the obesity epidemic in the Western World. Modern humans in industrialised countries have constant access to food all year round, but that is likely not how we evolved. In our evolutionary past, humans could have benefitted from having lower energy expenditure in winter to survive on what little food they had. However, such a reduction in energy expenditure would be maladaptive in most modern lives. A possible *benefit* from seasonal variations in food access is intriguing, especially considering the fasting trends that have

gained attention in recent years, such as intermittent fasting and the 5:2 diet. Perhaps the most beneficial fasting pattern is *seasonal?* This thesis has not addressed these potential associations. However, the observed seasonal differences in activity lengths in spring/summer compared to autumn/winter are suggestive of seasonal changes in energy requirements. The potential link between season, activity levels and metabolic expenditure should be investigated in future studies.

6.1.5 Data Handling

6.1.5.1 The TRIWEI Model

Visual inspection indicated that the TRIWEIs fit well to the daily Google Trends datasets. Although the TRIWEIs were based on a least squares method, it is possible that the fit could have been improved with other methods that were not explored in this thesis, especially in the middle part of the graphs. Any model fit might entail small deviations from the underlying structure of the data. However, from visual evaluation, I would say that the TRIWEIs did a good job of representing the morning and evening slopes of the daily datasets. This method, created for the purpose of this thesis, could be used in future analyses to model daily search patterns.

As seen from the results in Appendices A3 and A4, using the morning and evening peaks rendered similar statistical results as the derivatives. The only exception was that with the peak analyses, there was an additional statistically significant difference in the Southern Hemisphere between summer and spring. Presumably, search intensity increases in the morning (and decreases in the evening) because people start (and stop) using Google for the day. The steepest point on the slope should then be when most people start (or cease) their daily search activity. If we assume that the onset and offset of search activity occur near the onset and offset of daily activity, then we have an approximate, indirect measure of the latter. The morning peak, by this logic, should represent the midpoint between when most people have started and ended their first daily search bout. This is supported by the activity length measurements between peaks, which were approximately 12 hours in the Northern Hemisphere and 11 hours in the Southern Hemisphere (see Appendix A3). These time intervals are too short to represent daily activity, whereas the 16–17-hour activity lengths between gradients seem like a reasonable estimate (as discussed in section 6.1.1). Using the

peaks instead of the gradients might be beneficial in some contexts, but for this thesis, the interpretation of the slopes was more intuitive.

6.1.5.2 Re-normalisation of Normalised Data

As explained in section 4.4 the downloaded datasets were cut between Saturday 03:00 AM and Sunday 03:00 AM, and re-normalised. The re-normalisation process ensured that all the datasets, that were to be fitted with TRIWEIs, were similarly scaled. The second normalisation did not change the relative ordering or relationships between the data. In other words, larger values in the normalised dataset were still larger after re-normalisation, and smaller values remained smaller. However, re-normalisation could exaggerate the relative differences between data points. This would depend on the range of values in the cut dataset, which could be closer to 0 or 1, depending on the relative ratios in that part of the original, normalised dataset. The alteration in ratios among data points had no impact on the inherent relationships between the data, and consequently, the steepest slopes on the curve should maintain the same position on the x-axis for both the normalised and re-normalised data. Since these were the only points of interest for our analyses, I see no concerns regarding the re-normalisation procedure.

However, future studies of a similar kind could avoid this step by directly downloading Google Trends data within the desired time frame. This was not done here due to some methodological uncertainty at the time of the download. Also, one would have to be very thorough regarding the time and date calculations to get the same time frame regardless of the time zone for the individual regions. If the data has no margins, it is not possible to correct the time frame after the download.

6.1.5.3 Data Types, Averaging Procedures and Lack of Variation Measures

Data of four different types were tentatively investigated as explained in section 4.4 and Table 4. The "*Raw data"* data type was the only one that was directly fitted with TRIWEIs. In the following statistical tests, the activity lengths of the *Raw data* TRIWEIs were grouped by season and hemisphere and analysed directly. However, the ANOVA assumptions were not met for this data type since the Bartlett's test (see Appendices A5 and A6) revealed

heterogeneity of variance. When the variance is heterogenous, there is a greater chance of falsely rejecting the null hypothesis. I am not sufficiently versed in this topic to evaluate whether the *Raw data* analyses with regular ANOVAS had any value when this assumption was violated. However, since several websites suggested that Welch's ANOVA can be used when the assumption of homoscedasticity is violated, this approach was used for the *Raw data* (as presented in Appendices A8 and A11).

All the other data types met the ANOVA assumptions. However, these data types were averaged in one way or another, and there are issues associated with this approach. Specifically, in the averaging process of Google Trends datasets, I did not account for the associated variation/standard deviations, and this information was thus lost in the consequent analyses.

The loss of information in the averaging procedures seems especially worrisome regarding the ANOVAs and less so for the polar plots (although they were sometimes averaged more extensively) because the latter mostly served to illustrate what had already been detected in the statistical tests.

One issue in terms of averaging the Google Trends datasets (before fitting a TRIWEI) is that averaged Saturdays with higher variability will have the same weight as those with lower variability in the following tests. This could perhaps have been mitigated with some sort of weighted averages, where Saturdays with smaller standard deviations could have a higher weight than those with larger standard deviations. This was not tested.

Before performing the analyses, the *Raw data* were thought of as "clean" data that could support the validity of the results from the averaged data types. It was assumed that the *Raw data*-based activity lengths would be more variable than the activity lengths of the other data types. This was simply because individual daily Google Trends datasets were expected to be more variable than datasets averaged over eight years. However, this plan partly failed when the *Raw data* activity lengths were too variable to pass the Bartlett's test. Thus, the *Raw data* could not be used to verify the findings from the other data types with the same use of statistical tests. This issue was omitted by using Welch's ANOVA and Games-Howell post hoc tests, but since these tests required equal sample sizes, the data from 2023 were not included. Although the alternative tests supported the findings from the regular ANOVAs and Tukey post hoc tests (and in fact detected more seasonal differences than the latter, as shown

in Appendices A8 and A11), the tests are not directly comparable due to the inclusion/exclusion of 2023.

It is likely that the high variability of the *Raw data* was caused by the spikes that can be seen in the seasonal wave plot in Figure 35. These spikes were often caused by extended activity lengths in weeks 19, 20 and 52 (which coincide with holidays as discussed in section 6.1.4). Also, a large drop in search activity length occurred on January 1st 2021, which happened to be a Saturday. Celebrations on New Year's Eve likely made people wake up later than usual on this day in both hemispheres. It is possible that these data points (activity lengths) could be considered outliers, and that by removing them, the Bartletts test would have been insignificant so that the null hypothesis of homoscedasticity would not have been rejected. If so, the ANOVA approach could have been used for the *Raw data* to support the findings of the averaged data types. However, information can also become lost by removing "outliers" that do in fact constitute important trends in the data. Unfortunately, there was no time to investigate this further within the given time.

In summary, averaging without considering variability could lead to misleading results. However, we should only expect this to occur if seasonal patterns differed significantly between years and between regions. This does not seem to be the case, as indicated by the *Raw data* in Figure 35. Also, the *Raw data* analyses performed with Welch's ANOVA and Games-Howell post hoc tests did support the findings of the averaged data types.

6.1.5.4 Unequal Sample Sizes

The Google Trends datasets for analysing activity lengths were downloaded from January 1st, 2016, to August 5th, 2023. The cut-off in August 2023 occurred since this was the time that the data was downloaded. All the data (apart from faulty data, see Appendix 12) were used in regular ANOVAs and Tukey post hoc tests. For the analyses of *Raw data* and *Data averaged across hemispheres*, this resulted in a slight unevenness in sample size when using these tests. The period between August 6th and December 31st (corresponding to weeks 32-52, mostly representing the Autumn seasonal group) only had seven activity lengths per week, whereas the rest of the weeks of the year had eight activity lengths. This should not be a big issue regarding the ANOVA, since it is fairly robust to slight differences in sample size (94).

However, the Tukey post hoc test assumes that the sample sizes are equal (95), further supporting the choice of alternative statistical tests for the *Raw data.*

For the other two data types, *Data averaged across years* and *Data averaged across years and within hemispheres,* eight data sets were averaged across years between weeks 1-31 and seven data sets were averaged across years for weeks 32 – 52. This can have influenced the variability of the averages, but this was not investigated.

6.1.6 Choices Regarding which Results to Present

There were pros and cons to both kinds of statistical analyses, as discussed in the previous sections. Although I chose to present regular ANOVAs for *Data averaged across years*-based activity lengths in the Results section, this does not necessarily mean these analyses were better. The choice was mainly made based on the output I could get from the different tests. For example, the regular ANOVA provided confidence intervals for the estimated seasonal differences, and Welch's ANOVA did not. Also, the two-way ANOVA allowed several predictor variables and their interactions to be included in the model. This way, *hemisphere* could be accounted for in the model, whereas *season* was the only predictor variable in Welch's ANOVA. I wanted to compare the results of the two activity length analyses (based on *News* and *Arts and Entertainment*), and therefore needed the methodology for these investigations to be as similar as possible. Thus, it did not seem like a good option to present, for example, the regular ANOVA analyses for *News* and Welch's ANOVA analyses for *Arts and Entertainment*.

An argument could be made for presenting only *Raw data* analyses with Welch's ANOVA and Games-Howell post hoc tests. These tests were based on larger sample sizes and detected even more seasonal differences than the regular ANOVAs. For example, the Games-Howell post hoc test (based on *Raw data* from the *News* category) detected additional seasonal differences between summer and autumn (mean difference 8.4 minutes, $p = 0.001$) and between fall and winter (mean difference 6.5 minutes, $p = 0.02$) in the Northern Hemisphere. For the Southern Hemisphere, the Games-Howell post hoc test (based on *Raw data* from the *News* category) detected additional seasonal differences between autumn and winter (mean difference 6.2 minutes, $p = 0.04$) and between spring and summer (mean difference 7.5 minutes, $p = 0.03$). Perhaps due to the larger sample sizes for these analyses, the *Raw data*

Games-Howell post hoc tests generally had lower p-values for the statistically significant comparisons than did the Tukey post hoc tests.

Zimmerman (2004) recommended against using preliminary tests (such as Bartlett's test), and to rather aim for separate-variances analyses (such as Welch's ANOVA) unconditionally, especially in the presence of unequal sample sizes (96). With more time, this topic could have been investigated further to decide on the optimal approach. A solution could have been to use Welch's ANOVA, irrespective of the validity of the normal ANOVAs, although this would require separate analyses for each hemisphere.

6.2 Pornography Analyses

The analyses of search interest in pornography keywords revealed several seasonal differences for both hemispheres. The tests revealed that all seasonal comparisons were statistically significant in the Northern Hemisphere, except for *autumn-winter* and *spring-winter*. In the Southern Hemisphere, only *spring-autumn* and *autumn-winter* had insignificant seasonal comparisons. There was a relatively high search interest for pornography in summer in both the Northern and Southern Hemispheres (as shown in Figure 40). However, there was a pronounced difference between the Northern and Southern seasonal trends during winter. In particular, the Northern Hemisphere had a massive spike in pornography search interest during the last two weeks of the year, whereas the Southern Hemisphere had relatively low search interest during the Southern Hemisphere winter. The last two weeks of the year were nevertheless the weeks with the highest search interest in the Southern Hemisphere, but they coincided with Southern Hemisphere summer. The regional polar plots showed the same trends as the hemispherical ones.

The large Christmas spike in the Northern Hemisphere and the highest measured interest around Christmas in the Southern Hemisphere makes it obvious that free time and holidays must play an influential role in general pornography interest. Interestingly, Ayers et al. seem to have demonstrated an opposite Christmas effect in mental health queries, although they did not comment on this (46). Search activity for mental health-related queries was low around Christmas in both the Northern and Southern Hemispheres, as shown in Figure 11.

The occurrence of Easter peaks around weeks 12-15 in both hemispheres further support the significance of holidays in relation to pornography interest. These peaks were smaller, probably because Easter falls on different weeks depending on the year. This makes the interpretation of the summer interest difficult. Summer holidays are common in both Oceania and the Nordic countries, and this prolonged break could thus be responsible for the higher pornography interest at this time of year. Unfortunately, there is no way to detect the underlying cause for fluctuations in pornography interest through the analyses that were performed in this thesis.

Nevertheless, the potential link to photoperiod should be considered. The summer period of prolonged and heightened pornography interest coincides with the longest photoperiod, as shown in Figures 44 - 46. In fact, one of the three weeks with the highest pornography search interest in the Northern Hemisphere coincides with the week of the summer solstice. This can be said also for the Southern Hemisphere, although the longest day of the year coincides with the Christmas holiday. Additionally, there seems to be a steady increase in pornography interest as the photoperiod approaches its maximum length, and this is seen in both hemispheres.

The generally higher pornography interest in summer is also interesting in relation to birth rates, as discussed in section 3.4.2. If we assume that pornography interest represents general libido, then that could lead to higher conception rates in summer and, consequently, more births in spring. This fits nicely with previous observations in birth rates, as reported by Moos et al. (1994) and Dahlberg et al. (2018), among others (62, 63). It also corresponds with the study by Demir et al. (2016) who detected more sexual thoughts and ejaculations in summer, accompanied by higher testosterone and FSH levels, compared to winter (65).

6.2.1 Validity of Test Results

As shown by the Bartletts test in Appendix A7, the assumption of homogeneity of variance was violated for the pornography search interest data. This can easily be understood by looking at the polar plots in Figure 40: the variance in data points during winter must have been a lot larger than in any other season due to the large Christmas spike.

Whether the violation of the homoscedasticity assumption invalidates the findings of the ANOVA and Tukey post hoc tests has not been easy to understand. Several websites and articles state that heteroscedasticity can alter the rates of Type 1 errors (false positives, i.e. rejecting H_0 when it is true), but that this is especially likely when sample sizes are unequal. When a large variance is combined with a large sample size, the probability of Type 1 errors can fall below the significance level (usually α=0.05). On the contrary, when a smaller sample size is linked to a larger variance, the likelihood of Type 1 errors increases, occasionally exceeding the significance level by a significant margin (96, 97). In the pornography analyses, sample sizes were equal and consisted of 13 data points for each season and year, making a total of 78 data points per season across the six years that were included in the analysis (2016-2021). Whether this sample size is large enough to dodge the issues caused by heteroscedasticity remains unknown to me, but to err on the side of caution, Welch's ANOVA and Games-Howell post hoc tests were presented for these data.

Another question arises beyond the selection of analyses: how to interpret any observed differences? The Google Trends datasets were normalised by Google in an untransparent process (discussed further in section 6.3). The downloaded datasets contained normalised search interest values that generally varied between 80-100 on a scale from 0-100. It is not clear what other data affected the scale of the datasets in this manner. In contrast, the daily graphs that were used for the activity analyses were always scaled from 0 – 100. Although the prior processing of the Google Trends data is unclear, I suppose the relatively high numbers in the pornography datasets are reflective of generally high search interest. Thus, even if the statistical tests picked up on true seasonal differences, these should be expected to be modest in size. In fact, the generally high values for normalised pornography interest are most likely indicative of high year-round interest with rather small, though detectable, seasonal fluctuations.

6.2.2 Assumption of Libido Measurement

Another important consideration is whether pornography search interest can reasonably be used as a proxy for libido. It seems obvious that people who search for pornography, in most cases, seek sexual stimulation, but I have not found any research papers that evaluate this connection.

Two questions arise:

1. How many people use Google to search for pornography instead of other sources?
2. Who uses pornography at all?

I have not found any research papers that discussed the first question. However, PornHub reported that Google's Chrome web browser made up 48% of mobile traffic and 53% of desktop traffic in 2022. Chrome searches might be relevant to this thesis, as discussed further in section 6.3. Thus, Google Trends could potentially contain information about ~50% of all global PornHub visits that are made through Chrome (98). That estimate might vary between countries, but I do not possess specific information about the countries used in this thesis. Semrush.com reports that ~ 21% of xVideos and xHamster users, ~23% of PornHub users, and ~28% of XNXX users visit the sites through Google (99). These numbers, however, do not include "direct" traffic through the Google Chrome web browser. Thus, as much as ~70% of all PornHub visits might be included in the Google Trends database. This is a substantial amount of data, since PornHub is the 4th most visited website globally, with over 12.5 billion visits in September 2023 (100).

The second question – *who uses pornography at all?* – is also challenging to answer. I have not found any single study that breaks down pornography use by a range of factors, such as sex, age, religion, education level, etc. However, several narrower studies together make it obvious that pornography consumers do not represent the general population. First and foremost, women use pornography significantly less than men – 16% of women compared to 46% of men between 18 to 39 years of age viewed pornography in a given week of 2014 (101). In accordance with this estimate, PornHub reported that only 36% of users were female in 2022 (98). We can thus expect our "libido" estimate to be highly male-skewed. Also, a study conducted by Ballester-Arnal et al. (2023) found that individuals in the 41-60 age group consumed slightly less pornography content than adults below 40 (102). This finding might support our use of pornography consumption as a proxy for libido since human libido typically decreases with age, particularly after the age of 50 (103)

6.3 Challenges of Using Google Trends

Google has had approximately 90% of the global search engine market share since 2010, with over 5.5 billion daily searches. However, many non-western countries, such as China, Japan,

South Korea, and Russia, mostly use other search engines than Google for political or linguistic reasons (104). Since our analyses only included Western countries, Google should be expected to give a good estimation of total internet search volume.

The validity of the Google Trends data is essential to all conclusions that were drawn in this thesis. Unfortunately, Google Trends, although an interesting and versatile tool, is highly untransparent in various ways. Google seems very protective of its data, as demonstrated by the blocking of Pytrends. My suspicion is that there are economic reasons for keeping the information vague since data and metadata about people's search behaviour can be valuable to advertisers, business owners, competing search engines and others.

One of the most important questions is whether Chrome searches are included in the Google Trends datasets. Should Chrome not be included in the Google Trends datasets, then all direct pathways to any website are also not included. Since any previously visited website is normally autocompleted in the address bar, this would exclude a lot of relevant searches for both news and pornography. Unfortunately, it remains unknown whether Chrome searches are included in the Google Trends statistics.

Similarly, we lack the knowledge about which search keywords are included in any given category. If we knew how many and which keywords constituted certain categories, it would be easier to evaluate whether the graphs were indeed valid representations of general interest and whether noisiness in the data stems from relatively few keywords being included in each category.

Another question related to the quality of the data is why datasets lose quality over time (as shown in Figure 21) and how this degradation works. Also, how might Google's occasional updates of data processing methods (see Figure 24) impact our analyses?

Finally, a key characteristic of Google Trends data is their normalisation, which has its pros and cons. For the *activity length* analyses, normalisation aided the investigations. For the *pornography interest* analyses, on the other hand, non-normalised data could have made it easier to say something about the true magnitude of the seasonal fluctuations in search patterns.

6.3.1 Critiques of Google Trends

Although several interesting research questions have been investigated using Google Trends, not all researchers are equally enthusiastic. Franzén (2023) went as far as to advise against using Google Trends in research due to an observed low replicability of results (105). Franzén's issues regarded a single-keyword query within a narrow field of interest (an ex-chief of the Danish security service). I suspect her findings might be more of a problem in individual keyword searches than in category searches or even in combined keyword searches (especially for popular keywords). However, the reproducibility of Google Trends research should be investigated further for all types of studies. To ensure reproducibility, Nuti et al. (2014) provided a guideline to researchers who use Google Trends. They encouraged scientists to provide clear documentation of all modifiable search characteristics, such as location, time period, query category, and search terms, as well as a rationale for the search input (106).

6.4 Improvements and Suggestions for Future Studies

If this study were to be redone, I would have tried to find a statistical analysis that was not sensitive to heteroscedasticity and that could include several predictor variables. In the planning process, I did not expect the variance to differ between groups. However, now we know that certain days of the year, such as holidays, can cause too much variance for the regular ANOVAs (one-way and two-way) to provide trustworthy results. Thus, I would fit TRIWEI curves to *Raw data* to avoid losing information about the variation/standard deviations of the datasets in any averaging processes.

Given that an appropriate statistical approach was found, it would have been interesting to include more predictors in the analysis. An example of a variable that could have been included with the existing data is "holidays". Other factors that would have been interesting to include, if one could get hold of the information, are the sex and age of the searchers.

Also, the developed approach could have included a third site/region near the Equator. If photoperiod plays a role in the observed seasonal fluctuations, we should expect regions closer to the Equator to have less pronounced rhythms. An obstacle in this regard might be to find a region in this part of the world that is comparable to the Nordic countries and Oceania in terms of access to the internet and necessary electronic devices. However, also regions that

are not located exactly at the Equator, but closer to it, could be used to look for dampened seasonal rhythms.

An early goal in the planning stage was to look for latitudinal differences in activity lengths and pornography interest. This was not achieved due to the described issues with data download. However, if future studies were able to access such data, a latitudinal gradient could be very effective in connecting seasonal rhythms to photoperiod.

An important question is whether the developed method is a good way to investigate seasonal rhythms in activity levels and libido of humans. To this, my answer is yes. There are many different approaches to studying seasonal rhythms in humans. Many might say lab experiments would be more informative, and as discussed, they would be necessary to determine any potential circannual rhythms in humans. However, the Google Trends data also has its advantages. Firstly, it is based on a much larger sample than could ever be investigated in a lab. Secondly, it is anonymous, which can make investigations of certain topics, such as sexuality, easier. Thirdly, it is free, whereas other human studies could be costly. Lastly, this method might be more representative of people's actual lives than controlled lab experiments. The people who generated the Google Trends data were not aware of their role in this study. They were simply living their lives as normal, demonstrating to us how modern humans behave on the internet.

Finally, despite our modern lifestyles featuring electric lights, heating, screens, and various technological devices that distance us from the natural world, and despite our persistent states of comfort, humans seem to display seasonal fluctuations in behaviour. This finding is relevant, not just theoretically but also practically, to all of us.

7 Conclusion

The results suggest that there are indeed seasonal differences in search activity patterns on Google. These seasonal differences have been found in the *News* and *Arts and Entertainment* categories as well as for combined pornography keyword queries. The hypothesis that seasons contribute to the observed patterns is supported by inverse yearly oscillations in searches between the Northern and Southern Hemispheres. Based on these analyses, it is not possible to estimate to which degree the observed seasonal changes in search patterns represent real-life fluctuations in human activity levels and libido. The analyses used an indirect measure to approximate a field of study which is difficult to investigate directly. Since the study is observational, it cannot be used to make any conclusions about causation. Also, there is little transparency surrounding the units of investigation in this study since we do not know how Google Trends processes its data. Nevertheless, the observed seasonal changes in search patterns are by no means trivial and are suggestive of seasonal rhythms in two important aspects of human life: activity levels and libido.

Works Cited:

1. Kher A. What Causes Seasons on Earth? timeanddate.com [updated 30.08.23. Available from: https://www.timeanddate.com/astronomy/seasons-causes.html.
2. CoolCosmos. Why does Earh spin? coolsosmos.com: IPAC; [updated 14.08.23. Available from: https://coolcosmos.ipac.caltech.edu/ask/59-Why-does-Earth-spin-.
3. NASA. What Causes the Seasons? spaceplace: NASA; [updated 14.08.23. 22.07.21:[Available from: https://spaceplace.nasa.gov/seasons/en/.
4. Greshko M. Planet Earth, Explained Nationalgeographic.com: National Geographic; [updated 14.08.23. Available from: https://www.nationalgeographic.com/science/article/earth.
5. Meijer JH, Michel S, VanderLeest HT, Rohling JHT. Daily and seasonal adaptation of the circadian clock requires plasticity of the SCN neuronal network. Eur J Neurosci. 2010;32(12):2143-51.
6. Dunlap JC, Loros JJ, DeCoursey PJ. Chronobiology : biological timekeeping. Sunderland, Mass: Sinauer Associates; 2004.
7. Hill RW, Wyse GA, Anderson M. Animal physiology. Fourth edition. ed. Sunderland, Massachusetts: Sinauer Associates, Inc. Publishers; 2016.
8. Lilleøren K. Tropene snl.no2022 [updated 30.08.23. 22.03.22:[Available from: https://snl.no/tropene.
9. Mamen J. Klima snl.no2023 [updated 30.08.23. 05.04.23:[Available from: https://snl.no/klima.
10. Dahle H. Vendekretser snl.no2021 [updated 30.08.23. 27.04.21:[Available from: https://snl.no/vendekretser.
11. Temperate Zone dictionary.com [Available from: https://www.dictionary.com/browse/temperate-zone.
12. Dahle H. Polarsirkel snl.no2020 [updated 29.08.2023. 15.06.2020:[Available from: https://snl.no/polarsirkel.
13. Honma K, Honma S, Kohsaka M, Fukuda N. Seasonal variation in the human circadian rhythm: dissociation between sleep and temperature rhythm. American Journal of Physiology - Regulatory, Integrative and Comparative Physiology. 1992;262(5):885-91.
14. Author, Urry LA, Campbell NA, Cain ML, Minorsky PV, Wasserman SA. Biology: a global approach: Pearson Education; 2017.
15. Goldman BD. Mammalian Photoperiodic System: Formal Properties and Neuroendocrine Mechanisms of Photoperiodic Time Measurement. J Biol Rhythms. 2001;16(4):283-301.
16. Hazlerigg D, et.al. Biological timekeeping in polar environments: Lessons from terrestrial vertebrates. Journal of experimental biology (in press). 2023.
17. Wood S, Loudon A. The pars tuberalis: The site of the circannual clock in mammals? Gen Comp Endocrinol. 2018;258:222-35.
18. Coomans CP, Ramkisoensing A, Meijer JH. The suprachiasmatic nuclei as a seasonal clock. Front Neuroendocrinol. 2015;37:29-42.
19. Touitou Y, Reinberg A, Touitou D. Association between light at night, melatonin secretion, sleep deprivation, and the internal clock: Health impacts and mechanisms of circadian disruption. Life Sci. 2017;173:94-106.
20. Houben T, Deboer T, van Oosterhout F, Meijer JH. Correlation with Behavioral Activity and Rest Implies Circadian Regulation by SCN Neuronal Activity Levels. J Biol Rhythms. 2009;24(6):477-87.
21. VanderLeest HT, Houben T, Michel S, Deboer T, Albus H, Vansteensel MJ, et al. Seasonal Encoding by the Circadian Pacemaker of the SCN. Curr Biol. 2007;17(5):468-73.

22. Lucassen EA, van Diepen HC, Houben T, Michel S, Colwell CS, Meijer JH. Role of vasoactive intestinal peptide in seasonal encoding by the suprachiasmatic nucleus clock. Eur J Neurosci. 2012;35(9):1466-74.
23. Mirick DK, Davis S. Melatonin as a Biomarker of Circadian Dysregulation. Cancer Epidemiol Biomarkers Prev. 2008;17(12):3306-13.
24. Illnerová H, Zvolsky P, Vanecek J. The circadian rhythm in plasma melatonin concentration of the urbanized man: the effect of summer and winter time. Brain Res. 1985;328(1):186-9.
25. Dardente H, Wood S, Ebling F, Sáenz de Miera C. An integrative view of mammalian seasonal neuroendocrinology. J Neuroendocrinol. 2019;31(5):e12729-n/a.
26. Woodfill CJI, Wayne NL, Moenter SM, Karsch FJ. Photoperiodic synchronization of a circannual reproductive rhythm in sheep: identification of season-specific time cues. Biol Reprod. 1994;50(4):965-76.
27. Wood SH, Hindle MM, Mizoro Y, Cheng Y, Saer BRC, Miedzinska K, et al. Circadian clock mechanism driving mammalian photoperiodism. Nat Commun. 2020;11(1):4291-15.
28. EYA3 [Internet]. [cited 25.10.23]. Available from: https://www.proteinatlas.org/ENSG00000158161-EYA3.
29. Physiology, Thyroid Stimulating Hormone [Internet]. National Library of Medicine. 2023 [cited 11.09.23]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK499850/.
30. Wood Shona H, Christian Helen C, Miedzinska K, Saer Ben RC, Johnson M, Paton B, et al. Binary Switching of Calendar Cells in the Pituitary Defines the Phase of the Circannual Cycle in Mammals. Curr Biol. 2015;25(20):2651-62.
31. Lincoln GA, Clarke IJ, Hut RA, Hazlerigg DG. Characterizing a Mammalian Circannual Pacemaker. Science. 2006;314(5807):1941-4.
32. Daan S, Aschoff J. Circadian Rhythms of Locomotor Activity in Captive Birds and Mammals: Their Variations with Season and Latitude. Oecologia. 1975;18(4):269-316.
33. Archer SN, Laing EE, Moeller-Levet CS, van der Veen DR, Bucca G, Lazar AS, et al. Mistimed sleep disrupts circadian regulation of the human transcriptome. Proc Natl Acad Sci U S A. 2014;111(6):E682-E91.
34. Stokkan KA, Mortensen A, Blix AS. Food intake, feeding rhythm, and body mass regulation in Svalbard rock ptarmigan. American journal of physiology Regulatory, integrative and comparative physiology. 1986;251(2):264-R7.
35. Tyler NJC, Gerkema MP, Folkow L, van Oort BEH, Blix AS, Stokkan K-A. Circadian organization in reindeer. Nature. 2005;438(7071):1095-6.
36. Lu W, Meng Q-J, Tyler NJC, Stokkan K-A, Loudon ASI. A Circadian Clock Is Not Required in an Arctic Mammal. Curr Biol. 2010;20(6):533-7.
37. Hazlerigg D, Blix AS, Stokkan K-A. Waiting for the Sun: the circannual programme of reindeer is delayed by the recurrence of rhythmical melatonin secretion after the arctic night. J Exp Biol. 2017;220(Pt 21):3869-72.
38. Bronson FH. Are Humans Seasonally Photoperiodic? J Biol Rhythms. 2004;19(3):180-92.
39. Heideman PD, Bruno TA, Singley JW, Smedley JV. Genetic Variation in Photoperiodism in Peromyscus leucopus: Geographic Variation in an Alternative Life-History Strategy. Journal of mammalogy. 1999;80(4):1232-42.
40. Forni D, Pozzoli U, Cagliani R, Tresoldi C, Menozzi G, Riva S, et al. Genetic adaptation of the human circadian clock to day-length latitudinal variations and relevance for affective disorders. Genome Biol. 2014;15(10):499-.

41. Lynch GR, Heath HW, Johnston CM. Effect of Geographical Origin on the Photoperiodic Control of Reproduction in the White-Footed Mouse, Peromyscus leucopus. Biol Reprod. 1981;25(3):475-80.
42. Kuzmenko NV, Tsyrlin VA, Pliss MG. Seasonal Dynamics of Melatonin, Prolactin, Sex Hormones and Adrenal Hormones in Healthy People: a Meta-Analysis. Journal of evolutionary biochemistry and physiology. 2021;57(3):451-72.
43. Caniklioğlu M, Öztekin Ü, Caniklioğlu A, Selmi V, Sarı S, Işıkay L. Can Annual Daylight Cycles and Seasons Have an Effect on Male Sexual Functions? Curēus (Palo Alto, CA). 2021;13(10):e18879-e.
44. Foster RG, Roenneberg T. Human Responses to the Geophysical Daily, Annual and Lunar Cycles. Curr Biol. 2008;18(17):R784-R94.
45. Dopico XC, Evangelou M, Ferreira RC, Guo H, Pekalski ML, Smyth DJ, et al. Widespread seasonal gene expression reveals annual differences in human immunity and physiology. Nat Commun. 2015;6(1):7000-.
46. Ayers JWPMA, Althouse BMS, Allem J-PMA, Rosenquist JNMDP, Ford DEMDMPH. Seasonality in Seeking Mental Health Information on Google. Am J Prev Med. 2013;44(5):520-5.
47. Ma Y, Olendzki BC, Li W, Hafner AR, Chiriboga D, Hebert JR, et al. Seasonal variation in food intake, physical activity, and body weight in a predominantly overweight population. Eur J Clin Nutr. 2006;60(4):519-28.
48. Partonen T, Lönnqvist J. Seasonal affective disorder. Lancet. 1998;352(9137):1369-74.
49. Martinez ME. The calendar of epidemics: Seasonal cycles of infectious diseases. PLoS Pathog. 2018;14(11):e1007327.
50. Kennaway DJ, Van Dorp CF. Free-running rhythms of melatonin, cortisol, electrolytes, and sleep in humans in Antarctica. American Journal of Physiology - Regulatory, Integrative and Comparative Physiology. 1991;260(6):1137-44.
51. Usui A, Obinata I, Ishizuka Y, Okado T, Fukuzawa H, Kanba S. Seasonal changes in human sleep–wake rhythm in Antarctica and Japan. Psychiatry Clin Neurosci. 2000;54(3):361-2.
52. Yoneyama S, Hashimoto S, Honma K. Seasonal changes of human circadian rhythms in Antarctica. American Journal of Physiology - Regulatory, Integrative and Comparative Physiology. 1999;277(4):1091-7.
53. Arendt J, Middleton B. Human seasonal and circadian studies in Antarctica (Halley, 75°S). Gen Comp Endocrinol. 2017;258:250-8.
54. Wehr T, A. The durations of human melatonin secretion and sleep respond to changes in daylength (photoperiod). Journal of Clinical Endocrinology & Metabolism. 1991;73:1276-80.
55. Stothard ER, McHill AW, Depner CM, Birks BR, Moehlman TM, Ritchie HK, et al. Circadian Entrainment to the Natural Light-Dark Cycle across Seasons and the Weekend. Curr Biol. 2017;27(4):508-13.
56. Piosczyk H, Landmann N, Holz J, Feige B, Riemann D, Nissen C, et al. Prolonged Sleep under Stone Age Conditions. J Clin Sleep Med. 2014;10(7):719-22.
57. Roenneberg T, Pilz LK, Zerbini G, Winnebeck EC. Chronotype and Social Jetlag: A (Self-) Critical Review. Biology (Basel). 2019;8(3):54.
58. Fischer D, Lombardi DA, Marucci-Wellman H, Roenneberg T. Chronotypes in the US - Influence of age and sex. PLoS One. 2017;12(6):e0178782-e.
59. Randler C. Differences in Sleep and Circadian Preference between Eastern and Western German Adolescents. Chronobiol Int. 2008;25(4):565-75.

60. Nishiwaki-Ohkawa T, Yoshimura T. Molecular basis for regulating seasonal reproduction in vertebrates. J Endocrinol. 2016;229(3):R117-R27.
61. Roenneberg T, Aschof J. Annual Rhythm of Human Reproduction: I. Biology, Sociology, or Both? J Biol Rhythms. 1990;5(3):195-216.
62. Moos WS, Randall W. Patterns of human reproduction and geographic latitude. Int J Biometeorol. 1995;38(2):84-8.
63. Dahlberg J, Andersson G. Changing seasonal variation in births by sociodemographic factors: a population-based register study. Hum Reprod Open. 2018;2018(4):hoy015-hoy.
64. Roenneberg T. The Decline in Human Seasonality. J Biol Rhythms. 2004;19(3):193-5.
65. Demir A, Uslu M, Arslan OE. The effect of seasonal variation on sexual behaviors in males and its correlation with hormone levels: a prospective clinical trial. Cent European J Urol. 2016;69(3):285-9.
66. Svartberg J, Jorde R, Sundsfjord J, Bønaa KH, Barrett-Connor E. Seasonal Variation of Testosterone and Waist to Hip Ratio in Men: The Tromsø Study. J Clin Endocrinol Metab. 2003;88(7):3099-104.
67. Borger JN, Huber R, Ghosh A. Capturing sleep-wake cycles by using day-to-day smartphone touchscreen interactions. NPJ Digit Med. 2019;2(1):73-.
68. Massar SAA, Chua XY, Soon CS, Ng ASC, Ong JL, Chee NIYN, et al. Trait-like nocturnal sleep behavior identified by combining wearable, phone-use, and self-report data. NPJ digital medicine. 2021;4(1):90-.
69. Mavragani A, Ochoa G, Tsagarakis KP. Assessing the Methods, Tools, and Statistical Approaches in Google Trends Research: Systematic Review. J Med Internet Res. 2018;20(11):e270.
70. Bakker KM, Martinez-Bakker ME, Helm B, Stevenson TJ. Digital epidemiology reveals global childhood disease seasonality and the effects of immunization. Proc Natl Acad Sci U S A. 2016;113(24):6689-94.
71. Zattoni F, Gul M, Soligo M, Morlacco A, Motterle G, Collavino J, et al. The impact of COVID-19 pandemic on pornography habits: a global analysis of Google Trends. Int J Impot Res. 2020;33(8):824-31.
72. Leypunskiy E, Kıcıman E, Shah M, Walch OJ, Rzhetsky A, Dinner AR, et al. Geographically Resolved Rhythms in Twitter Use Reveal Social Pressures on Daily Activity Patterns. Curr Biol. 2018;28(23):3763-75.e5.
73. Liu C, White R, Dumais S. Understanding web browsing behaviors through Weibull analysis of dwell time. Annual ACM Conference on Research and Development in Information Retrieval [Internet]. 2010:[379-86 pp.].
74. Weibullfordelingen mn.uio.no: Det matematisk-naturvitenskapelige fakultet; 2019 [updated 27.12.19. Available from: https://www.mn.uio.no/ibv/tjenester/kunnskap/plantefys/matematikk/weibullfordelingen.html.
75. Refine Trends results by category support.google.com: Google; [Available from: https://support.google.com/trends/answer/4359597?hl=en.
76. Asterisk Computer Hope: computerhope.com; 2023 [Available from: https://www.computerhope.com/jargon/a/asterisk.htm.
77. pytrends 4.9.2 pypi.org [Available from: https://pypi.org/project/pytrends/.
78. GeoPandas 0.dev+untagged geopandas.org [11.11.23]. Available from: https://geopandas.org/en/stable/index.html.
79. GeneralMills. pytrends github.com [Available from: https://github.com/GeneralMills/pytrends.
80. Løvås GG. Statistikk for universiteter og høyskoler. Oslo: Universitetsforlaget; 2018.
81. Hornæs P. Hypotesetesting for mastergradsstudium i informasjonssikkerhet. In: Gjøvik Hi, editor. 4.11 ed2003.

82. Glen S. Welch's ANOVA: Definition, Assumptions StatisticsHowTo.com: [02.11.23]. Available from: https://www.statisticshowto.com/welchs-anova/.
83. Frost J. Benefits of Welch's ANOVA Compared to the Classic One-Way ANOVA StatisticsByJim.com [07.11.23]. Available from: https://statisticsbyjim.com/anova/welchs-anova-compared-to-classic-one-way-anova/.
84. Suzuki M, Taniguchi T, Furihata R, Yoshita K, Arai Y, Yoshiike N, et al. Seasonal changes in sleep duration and sleep problems: A prospective study in Japanese community residents. PLoS One. 2019;14(4):e0215345-e.
85. Erkert HG, Kappeler PM. Arrived in the Light: Diel and Seasonal Activity Patterns in Wild Verreaux's Sifakas (Propithecus v. verreauxi; Primates: Indriidae). Behavioral ecology and sociobiology. 2004;57(2):174-86.
86. Hysing M, Pallesen S, Stormark KM, Jakobsen R, Lundervold AJ, Sivertsen B. Sleep and use of electronic devices in adolescence: results from a large population-based study. BMJ Open. 2015;5(1):e006748-e.
87. Hirshkowitz MP, Whiton KMHS, Albert SMP, Alessi CMD, Bruni OMD, DonCarlos LP, et al. National Sleep Foundation's sleep time duration recommendations: methodology and results summary. Sleep Health. 2015;1(1):40-3.
88. Hysing M, Pallesen S, Stormark KM, Lundervold AJ, Sivertsen B. Sleep patterns and insomnia among adolescents: a population‐based study. J Sleep Res. 2013;22(5):549-56.
89. Sheehan CM, Frochen SE, Walsemann KM, Ailshire JA. Are US adults reporting less sleep?: Findings from sleep duration trends in the National Health Interview Survey, 2004-2017. Sleep. 2019;42(2):1.
90. Mobile Fact Sheet pewresearch.org: Pew Research Center; 2021 [Available from: https://www.pewresearch.org/internet/fact-sheet/mobile/.
91. Johnsen MT, Wynn R, Allebrandt K, Bratlid T. Lack of major seasonal variations in self reported sleep-wake rhythms and chronotypes among middle aged and older people at 69 degrees North: The Tromsø Study. Sleep Med. 2012;14(2):140-8.
92. Roenneberg T, Kumar CJ, Merrow M. The human circadian clock entrains to sun time. Curr Biol. 2007;17(2):R44-R5.
93. M. A. Hofman, Swaab DF. The human hypothalamus: comparative morphometry

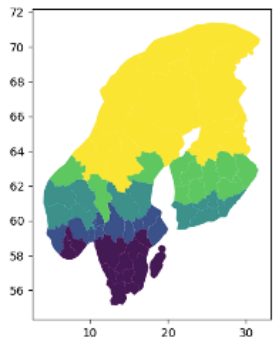
and photoperiodic influences. Progress in Brain Research. 1992;93, Ch. 10.
94. Grace-Martin K. When unequal sample sizes are and are not a problem in anova theanalysisfactor.com [Available from: https://www.theanalysisfactor.com/when-unequal-sample-sizes-are-and-are-not-a-problem-in-anova/.
95. Post hoc Testing uvm.edu2018 [Available from: http://bayes.acs.unt.edu:8083/BayesContent/class/Jon/ISSS_SC/Module009/isss_m91_onewayanova/node7.html.
96. Zimmerman DW. A note on preliminary tests of equality of variances. Br J Math Stat Psychol. 2004;57(1):173-81.
97. Bishop TA, Dudewicz EJ. Heteroscedastic Anova. Sankhyā Series B. 1981;43(1):40-57.
98. PornHub. The 2022 Year in Review pornhub.com2022 [06.11.23]. Available from: https://www.pornhub.com/insights/2022-year-in-review#gender-demographics.
99. SEMRUSH. Free Website Traffic Checker semrush.com2023 [06.11.23]. Available from: semrush.com/website/.
100. pornhub.com, September 2023 Traffic Stats semrush.com [09.11.23]. Available from: https://www.semrush.com/website/pornhub.com/overview/.
101. Regnerus M, Gordon D, Price J. Documenting Pornography Use in America: A Comparative Analysis of Methodological Approaches. J Sex Res. 2016;53(7):873-81.

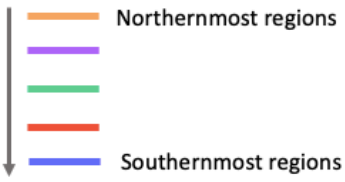
102. Ballester-Arnal R, García-Barba M, Castro-Calvo J, Giménez-García C, Gil-Llario MD. Pornography Consumption in People of Different Age Groups: an Analysis Based on Gender, Contents, and Consequences. Sexuality research & social policy. 2023;20(2):766-79.
103. Segraves RT, Segraves KB. Human Sexuality and Aging. Journal of Sex Education & Therapy. 1995.
104. Jun S-P, Yoo HS, Choi S. Ten years of research change using Google Trends: From the perspective of big data utilizations and applications. Technological forecasting & social change. 2018;130:69-87.
105. Franzen A. Big data, big problems: Why scientists should refrain from using Google Trends. Acta Sociologica. 2023;66(3):343-7.
106. Nuti SV, Wayda B, Ranasinghe I, Wang S, Dreyer RP, Chen SI, et al. The Use of Google Trends in Health Care Research: A Systematic Review. PLoS One. 2014;9(10):e109583-e.
107. Bedre R. How to Perform ANOVA in Python (With Examples) 2023 [updated 07.05.2301.11.23]. Available from: https://www.reneshbedre.com/blog/anova.html.

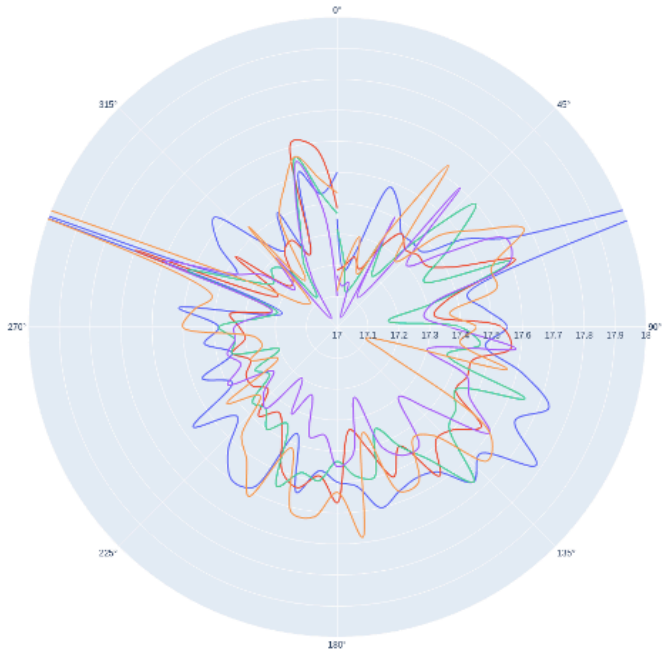
Appendix

A1: Tentative Analyses of Pytrends Data

Daily data was downloaded from the *News* and *Arts and Entertainment* categories of Google Trends through *Pytrends*. The downloads were made separately for a range of regions within the Nordic countries (Norway, Sweden, and Finland), as shown on the map below. Daily datasets for ten years were averaged first across years for each region. The averaged datasets were then interpolated before the maximum and minimum derivatives were used to find the activity lengths. Activity lengths were averaged between regions (colour-coded in the map) and these averaged activity lengths were used to produce the plot below. The plot was never finished because errors in the data (chunks of zeros) were detected, and the data was later lost. Thus, the polar axis is in degrees, although it should have been weeks. The x-axis is 'hours'. Although it is not certain if this would be true with 'clean' data, this initial analysis was not able to detect latitudinal differences between the selected regions.

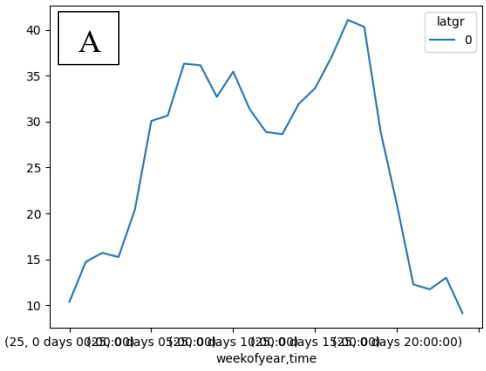


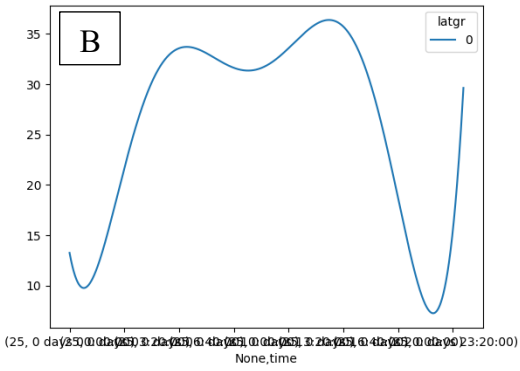


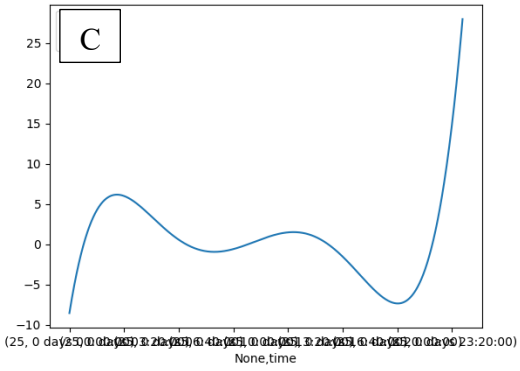


A2: Polynomial Fit of Google Trends Graphs

Example figures of A) google trends data from the Northern Hemisphere, B) 5th order polynomial fit to this data and C) the derivative of B.







A3: Analyses Using Peaks instead of Gradients, *News.*

The following analyses are based on *Data averaged across years*, but instead of using the distance between the derivatives, this method was based on the global maxima of the morning and evening TRIWEI curves. The only difference in the interpretation of the results is that there is an additional significant difference between summer and spring in the Southern Hemisphere when using the peaks instead of the gradients (given a significance level of 0.05).

Anova ActivityPeak

	df	sum_sq	...	F	PR(>F)
C(season)	3.0	5308.313746	...	14.212510	1.343815e-08
C(hemis)	1.0	153023.058029	...	1229.114446	6.384436e-100
C(season):C(hemis)	3.0	884.369923	...	2.367817	7.124301e-02
Residual	257.0	31996.146520	...	NaN	NaN

Tukey ('North', 'ActivityPeak')
{'Fall': 0, 'Spring': 1, 'Summer': 2, 'Winter': 3}

	group1	group2	Diff	Lower	Upper	q-value	p-value
0	3	1	12.439560	5.127827	19.751294	6.249117	0.001000
1	3	2	8.747253	1.435519	16.058986	4.394255	0.011930
2	3	0	3.619048	-3.692686	10.930781	1.818059	0.564966
3	1	2	3.692308	-3.753597	11.138213	1.821438	0.563629
4	1	0	8.820513	1.374608	16.266418	4.351213	0.013081
5	2	0	5.128205	-2.317700	12.574110	2.529775	0.282799

Tukey ('South', 'ActivityPeak')
{'Fall': 0, 'Spring': 1, 'Summer': 2, 'Winter': 3}

	group1	group2	Diff	Lower	Upper	q-value	p-value
0	2	0	9.879121	3.920721	15.837521	6.124675	0.001000
1	2	3	13.109890	7.151490	19.068290	8.127627	0.001000
2	2	1	6.802198	0.843798	12.760597	4.217101	0.018511
3	0	3	3.230769	-2.836968	9.298506	1.966860	0.507030
4	0	1	3.076923	-2.990814	9.144660	1.873200	0.543953
5	3	1	6.307692	0.239955	12.375430	3.840061	0.038364

The distances between the peaks are consistently shorter than the distances between the derivatives. Naturally, morning peaks occur later than morning derivatives, and evening peaks occur earlier than evening derivatives. Although it seems like peaks can be used to detect the same trends as presented in the main part of the thesis, the interpretation of what the peaks might represent is perhaps not as obvious.

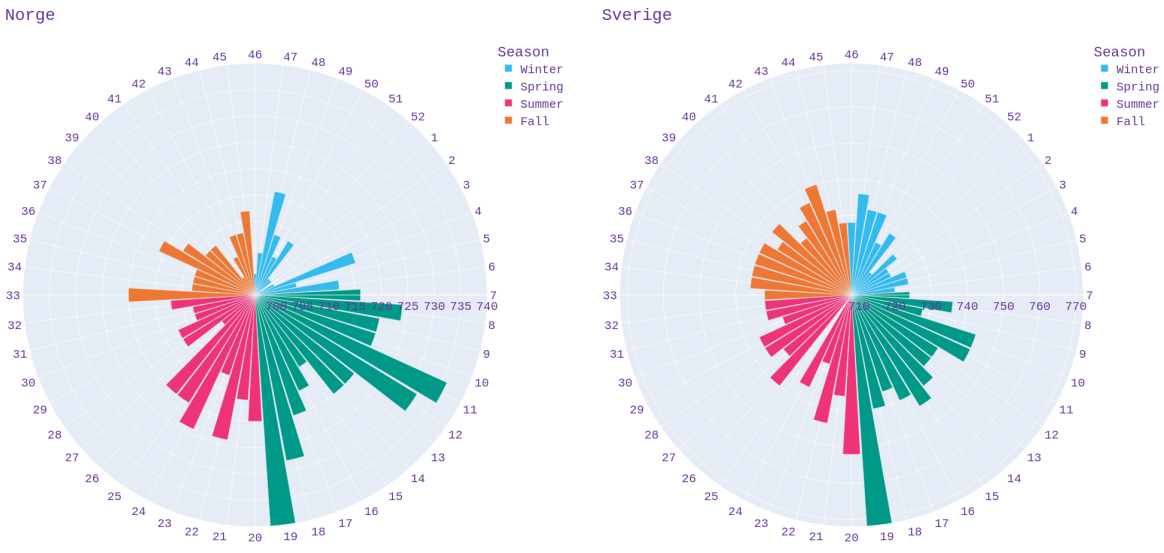
North summary

season	Fall	Spring	Summer	Winter
activity				
ActivityGrad	16:54	17:05	17:02	16:46
ActivityPeak	11:59	12:07	12:04	11:54
EveningGrad	24:02	24:07	24:16	24:14
EveningPeak	21:39	21:43	21:52	21:53
MorningGrad	07:08	07:02	07:15	07:28
MorningPeak	09:40	09:35	09:48	09:59

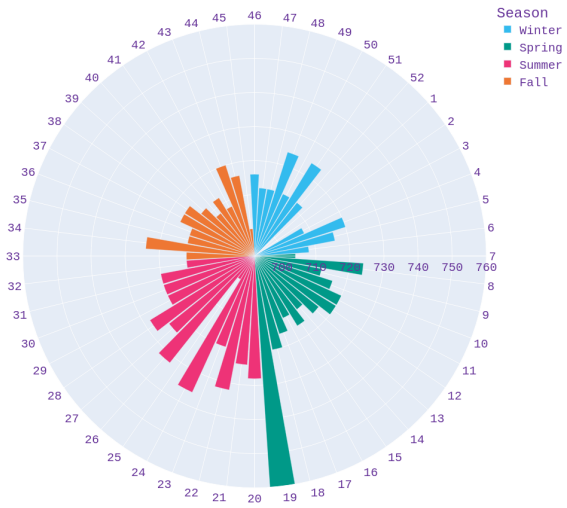
South summary

season	Fall	Spring	Summer	Winter
activity				
ActivityGrad	16:01	16:06	16:14	15:54
ActivityPeak	11:10	11:13	11:20	11:06
EveningGrad	23:10	23:07	23:19	23:08
EveningPeak	20:50	20:45	20:56	20:48
MorningGrad	07:09	07:00	07:04	07:14
MorningPeak	09:40	09:32	09:36	09:42

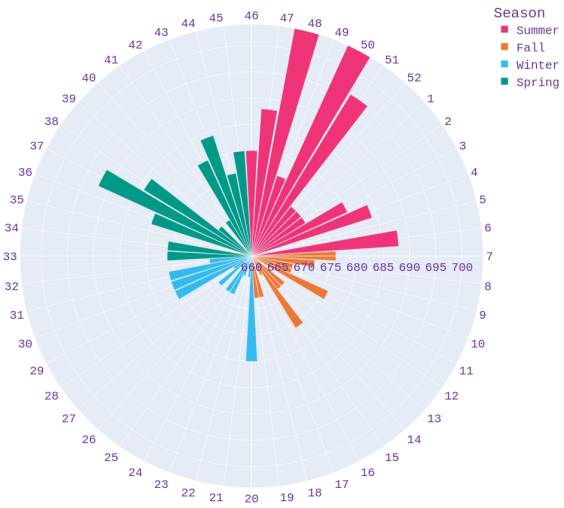
Polar plots using peaks instead of gradients:



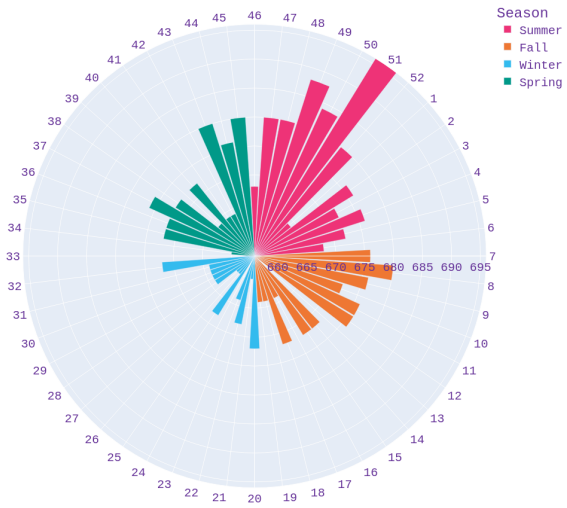
Finland



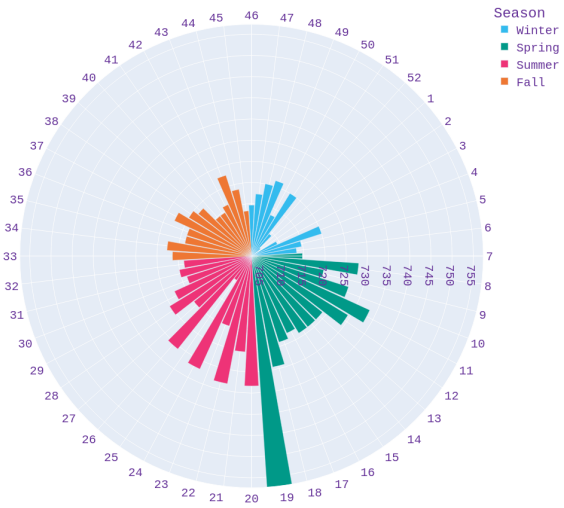
New Zealand



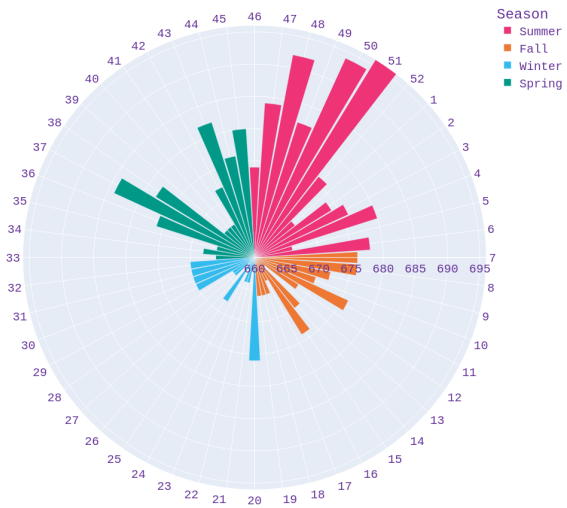
Victoria



North

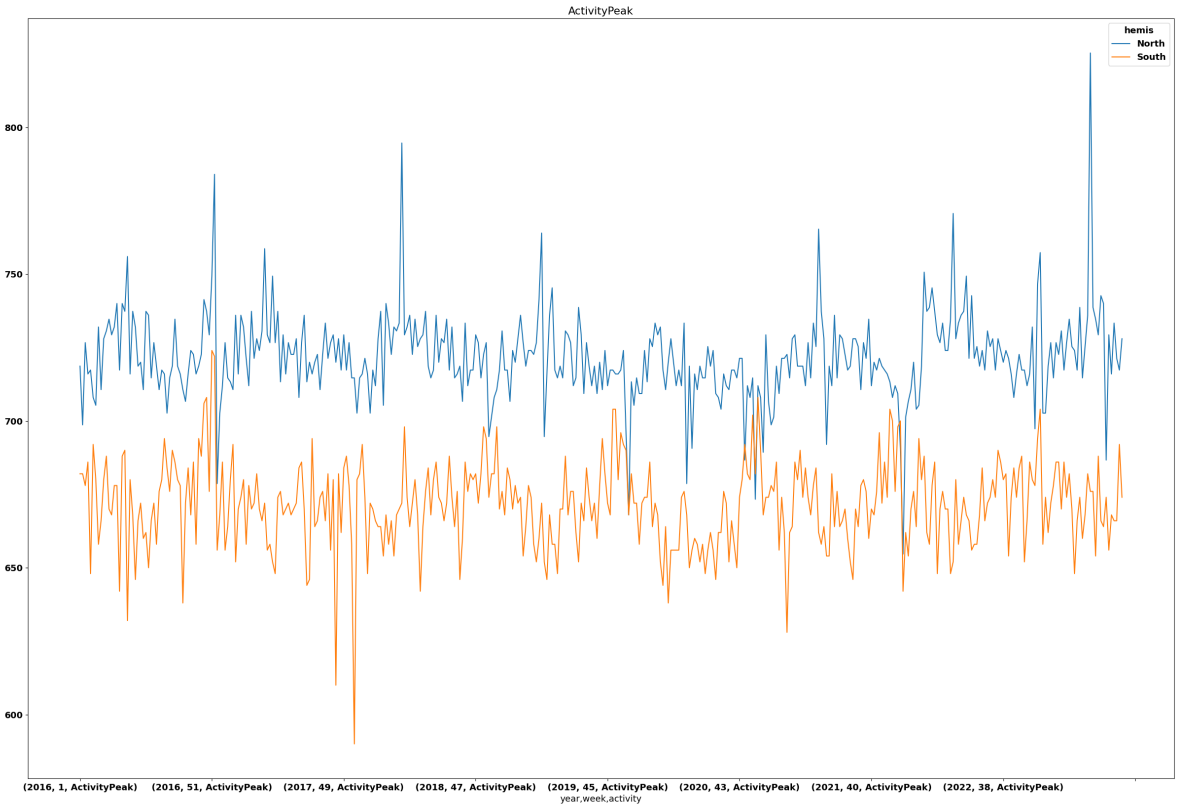


South



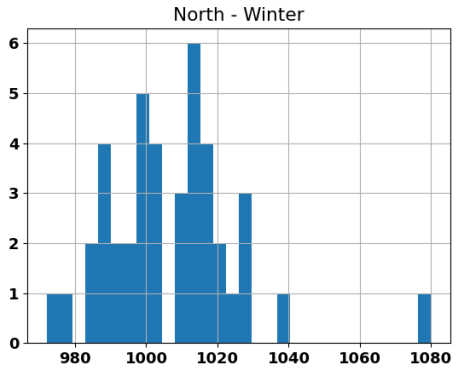
A4: Wave Plot using Peaks instead of Gradients.

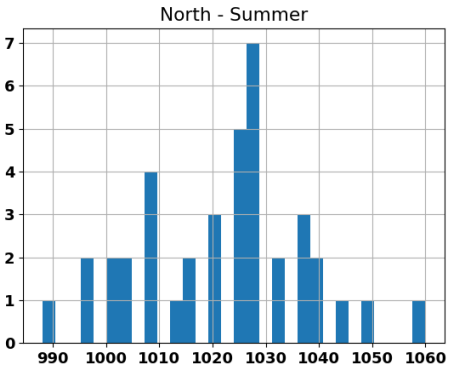
This seasonal wave plot is made with activity lengths (the distance between the global maxima of the first and last peaks on the TRIWEI graphs) based on *News*, *Raw data.*

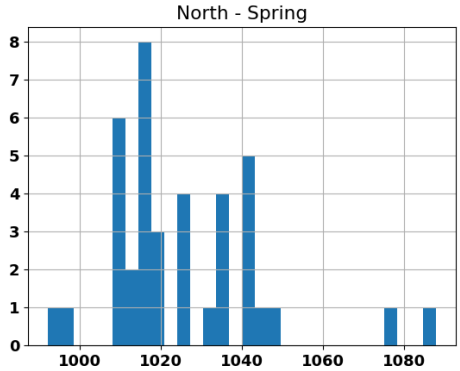


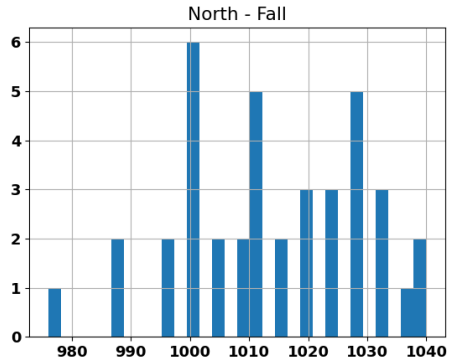
A5: Model Assumptions Tests for *News* Activity Lengths

Normality of data: ANOVAs (one-way and two-way) are fairly robust against violations of the normality assumption. The data are supposed to stem from a population that is normally distributed. One way to approach this assumption is to ask if there is any reason why the data should *not* come from a normally distributed population (81). However, the seasonal groups were plotted as histograms. The example histograms that are included here are from the Northern Hemisphere for *Data averaged across years*-based activity lengths. The samples look as if they could stem from a normally distributed population.





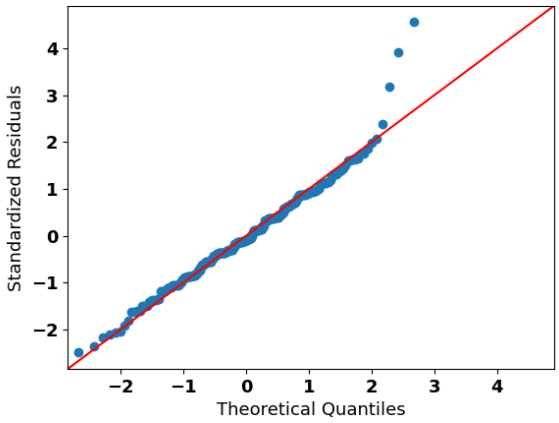


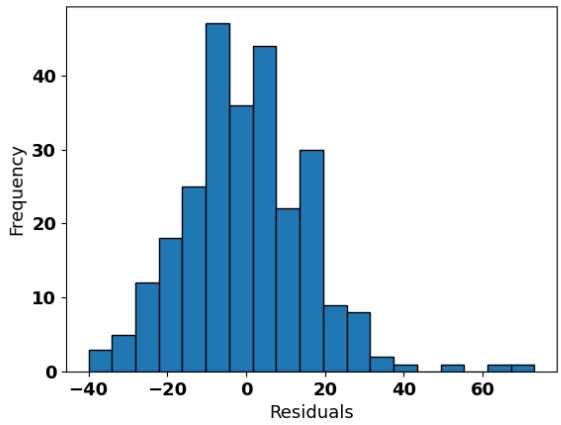


Data averaged across years:

The following analyses and the code to generate them were based on instructions from https://www.reneshbedre.com/blog/anova.html (107).

Q-Q plot and histogram were used to test whether the *News Data averaged across years*-based activity length residuals were normally distributed.





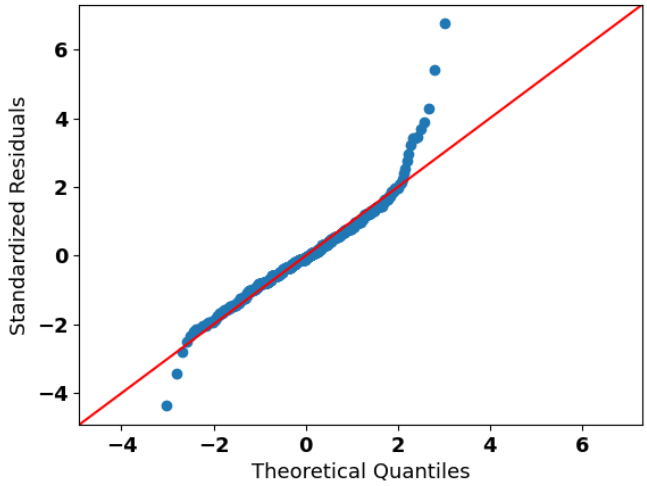
After normality seemed OK (although with some potential outliers), the next step was to test the homogeneity of variances with Bartlett's test. The null hypothesis in this test is that the samples from the different seasons have equal variances. Since the p-value is large (0.64), the null hypothesis cannot be rejected.

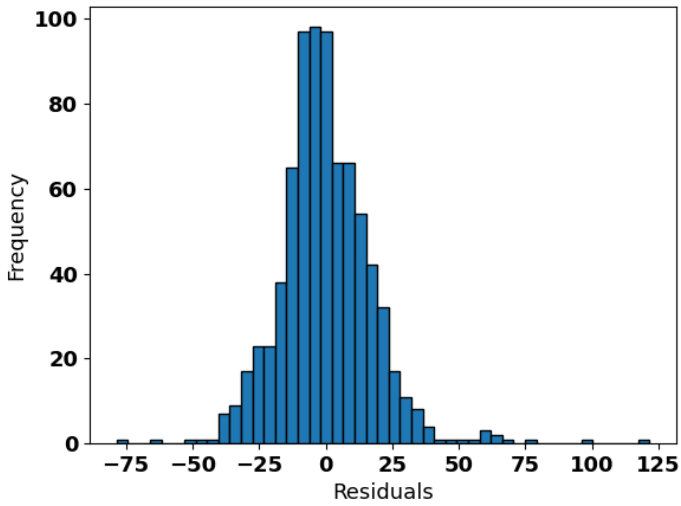
Bartlett

	Parameter	Value
0	Test statistics (T)	1.6643
1	Degrees of freedom (Df)	3.0000
2	p value	0.6449

Data averaged within hemispheres:

The Q-Q plot and histogram suggest that the residuals are fairly normally distributed.





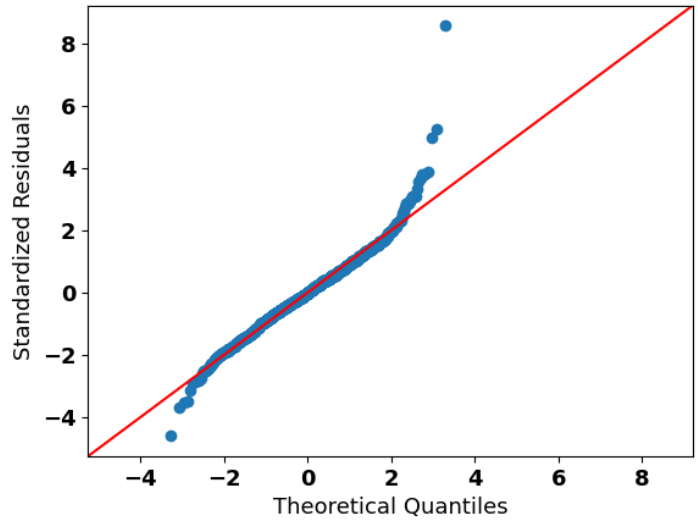
Bartlett's test could not reject the assumption of homogeneity of variances.

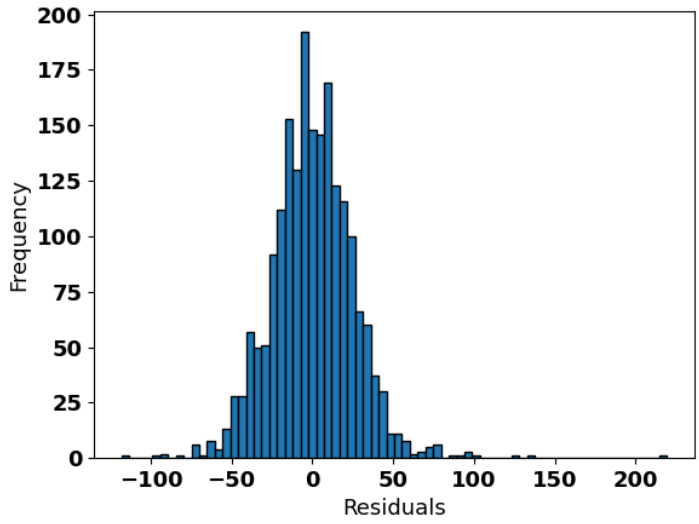
Bartlett

	Parameter	Value
0	Test statistics (T)	7.7158
1	Degrees of freedom (Df)	3.0000
2	p value	0.0523

Raw data:

The Q-Q plot and histogram suggest that the residuals are fairly normally distributed.





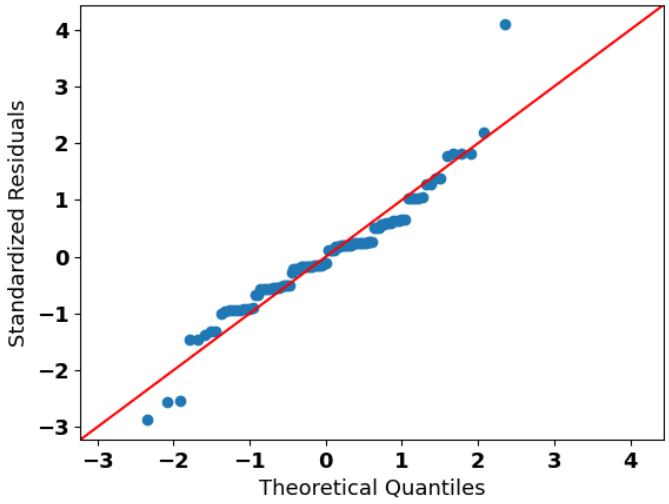
Bartlett's test rejected the assumption of homogeneity of variances.

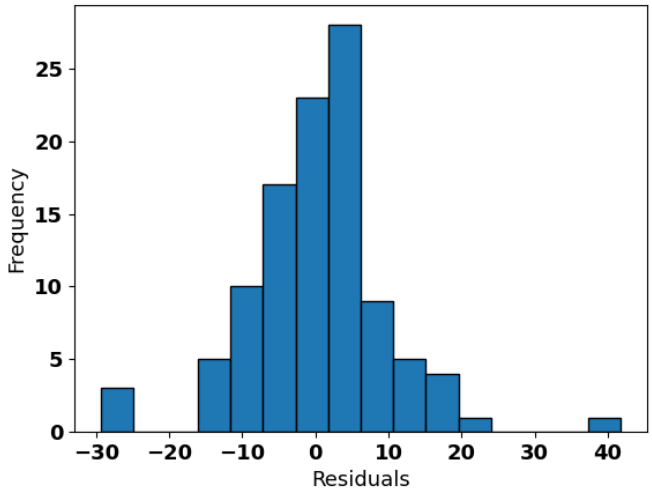
Bartlett

	Parameter	Value
0	Test statistics (T)	13.0012
1	Degrees of freedom (Df)	3.0000
2	p value	0.0046

Data averaged across years and within hemispheres:

The Q-Q plot and histogram suggest that the residuals are fairly normally distributed.





Bartlett's test could not reject the assumption of homogeneity of variances.

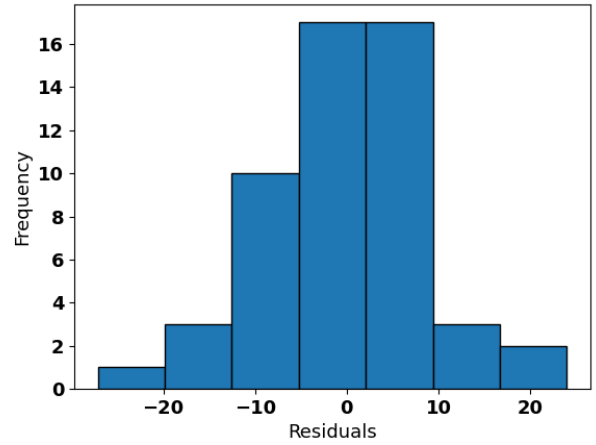
Bartlett

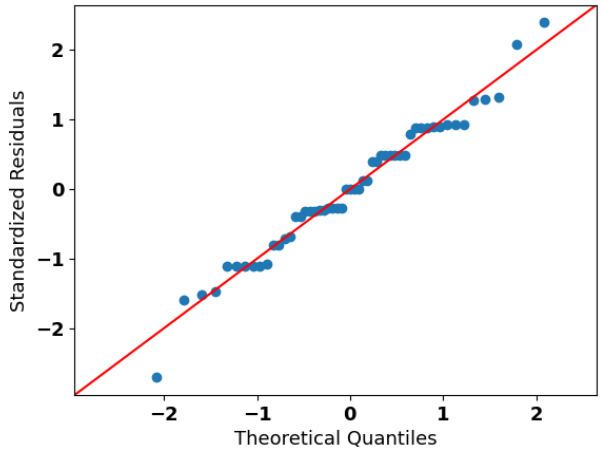
	Parameter	Value
0	Test statistics (T)	1.2433
1	Degrees of freedom (Df)	3.0000
2	p value	0.7426

A6: Model Assumptions Tests for *Arts and Entertainment* Activity Lengths

Data averaged across years:

Q-Q plot and histogram were used to test whether the *Arts and Entertainment Data averaged across years* activity length residuals were normally distributed.





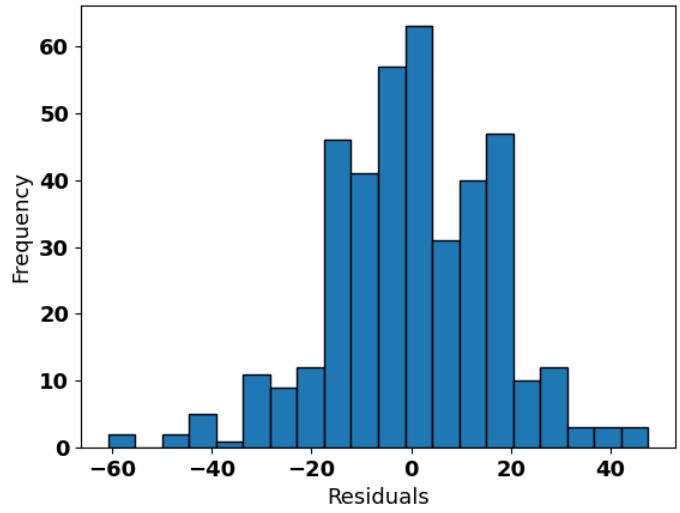
Since the normality assumption seemed OK, the next step was to test the homogeneity of variances with Bartlett's test. The null hypothesis in this test is that the samples from the different seasons have equal variances. Since the p-value is large 0.78), the null hypothesis cannot be rejected.

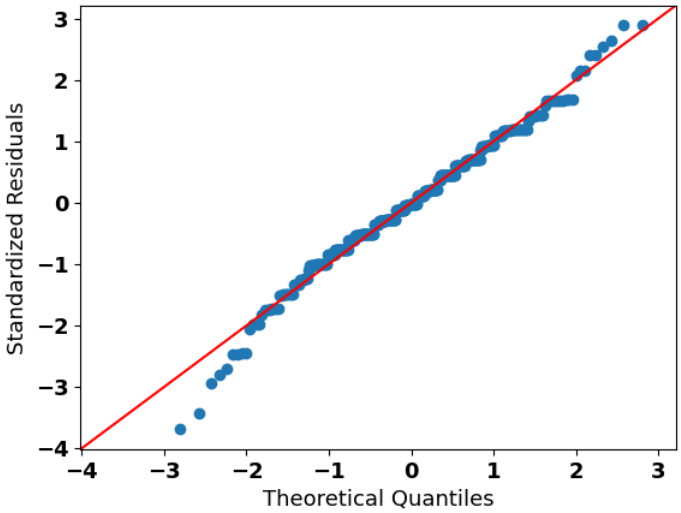
Bartlett

	Parameter	Value
0	Test statistics (T)	1.1004
1	Degrees of freedom (Df)	3.0000
2	p value	0.7770

Raw data:

The Q-Q plot and histogram to test whether the *Arts and Entertainment Raw data* activity length residuals were normally distributed are shown below.





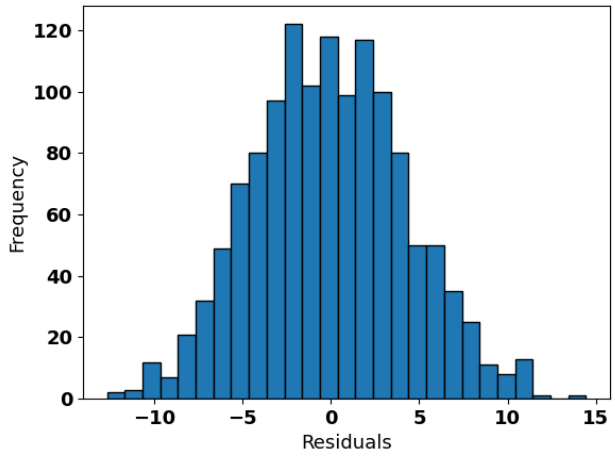
Since the normality assumption seemed OK, the next step was to test the homogeneity of variances with Bartlett's test. The null hypothesis in this test is that the samples from the different seasons have equal variances. Since the p-value is small 0.005), the null hypothesis is rejected.

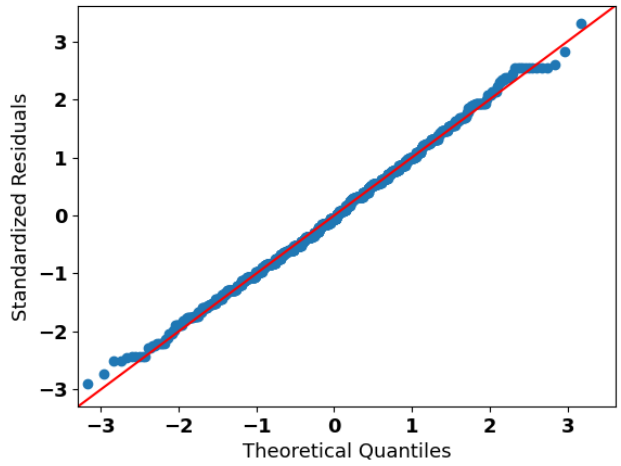
Bartlett

	Parameter	Value
0	Test statistics (T)	12.8891
1	Degrees of freedom (Df)	3.0000
2	p value	0.0049

A7: Model Assumptions Tests for *Pornography* Search Interest

The Q-Q plot and histogram to test whether the pornography residuals were normally distributed are shown below. This assumption seemed OK.





Next, homogeneity of variances was tested with the Bartlett's test. The null hypothesis in this test is that the samples from the different seasons have equal variances. Since the p-value is small 0.0006), the null hypothesis is rejected.

Bartlett

	Parameter	Value
0	Test statistics (T)	17.2931
1	Degrees of freedom (Df)	3.0000
2	p value	0.0006

A8: *Raw Data* Analyses. (*News* Category).

Two-way ANOVA and Tukey post hoc tests (not used due to heteroscedasticity):

```
results\news_category\Activity
Anova ActivityGrad
                       df        sum_sq  ...            F         PR(>F)
C(season)             3.0  8.724279e+04  ...    44.129070   1.344247e-27
C(hemis)              1.0  1.435485e+06  ...  2178.287424  5.587882e-321
C(season):C(hemis)    3.0  6.342899e+03  ...     3.208360   2.224933e-02
Residual           1972.0  1.299542e+06  ...          NaN            NaN
```

```
Tukey ('North', 'ActivityGrad')
{'Fall': 0, 'Spring': 1, 'Summer': 2, 'Winter': 3}
   group1  group2       Diff      Lower      Upper    q-value   p-value
0       3       1  17.946886  12.313083  23.580690  11.590221  0.001000
1       3       2  15.258437   9.611382  20.905493   9.830877  0.001000
2       3       0   6.520147   0.694383  12.345910   4.072009  0.021116
3       1       2   2.688449  -2.874506   8.251405   1.758330  0.587440
4       1       0  11.426740   5.682459  17.171020   7.237540  0.001000
5       2       0   8.738291   2.981012  14.495570   5.522217  0.001000
```

```
Tukey ('South', 'ActivityGrad')
{'Fall': 0, 'Spring': 1, 'Summer': 2, 'Winter': 3}
   group1  group2       Diff      Lower      Upper    q-value   p-value
0       2       0  11.170330   5.113348  17.227312   6.714702  0.001000
1       2       3  17.586685  11.515455  23.657915  10.546889  0.001000
2       2       1   7.120879   0.857518  13.384241   4.139455  0.018451
3       0       3   6.416355   0.435543  12.397168   3.906117  0.029920
4       0       1   4.049451  -2.126307  10.225208   2.387387  0.330721
5       3       1  10.465806   4.276074  16.655538   6.156273  0.001000
```

Welch's ANOVAs and Games-Howell post hoc tests:

```
Anova WELCH NORTH_ACTIVITYGRAD
   Source  ddof1       ddof2          F         p-unc       np2
0  season      3  604.920845  22.856006  5.001600e-14  0.062814

PostHoc Games-Howell NORTH_ACTIVITYGRAD
        A       B    mean(A)    mean(B)  ...       T        df    pval  hedges
0    Fall  Spring  1012.9963  1023.3407  ... -4.8283  533.3342  0.0000 -0.4127
1    Fall  Summer  1012.9963  1021.3480  ... -3.9163  534.4859  0.0006 -0.3347
2    Fall  Winter  1012.9963  1006.5362  ...  2.9709  533.1766  0.0163  0.2530
3  Spring  Summer  1023.3407  1021.3480  ...  0.8741  543.9639  0.8183  0.0747
4  Spring  Winter  1023.3407  1006.5362  ...  7.2467  546.7768  0.0000  0.6176
5  Summer  Winter  1021.3480  1006.5362  ...  6.4127  546.5608  0.0000  0.5465

[6 rows x 10 columns]

Anova WELCH SOUTH_ACTIVITYGRAD
   Source  ddof1       ddof2          F         p-unc       np2
0  season      3  401.737575  15.905906  8.616871e-10  0.069849

PostHoc Games-Howell SOUTH_ACTIVITYGRAD
        A       B   mean(A)   mean(B)  ...       T        df    pval  hedges
0    Fall  Spring  958.8132  963.1648  ... -1.8828  357.7264  0.2372 -0.1970
1    Fall  Summer  958.8132  970.6739  ... -4.6043  337.9712  0.0000 -0.4796
2    Fall  Winter  958.8132  952.5934  ...  2.6499  355.3894  0.0417  0.2772
3  Spring  Summer  963.1648  970.6739  ... -2.7946  353.3322  0.0279 -0.2913
4  Spring  Winter  963.1648  952.5934  ...  4.2826  361.7265  0.0001  0.4480
5  Summer  Winter  970.6739  952.5934  ...  6.6522  356.2992  0.0000  0.6934

[6 rows x 10 columns]
```

A9: Analyses of *Data Averaged within Hemispheres* (*News* Category)

```
results\news_category\Activity from hemisphere averaged
Anova ActivityGrad
                       df         sum_sq  ...            F         PR(>F)
C(season)             3.0   42834.334498  ...    44.084815   2.477632e-26
C(hemis)              1.0  675358.222981  ...  2085.222730  4.487945e-223
C(season):C(hemis)    3.0    4806.152276  ...     4.946460   2.081973e-03
Residual            784.0  253920.523327  ...          NaN            NaN


Tukey ('North', 'ActivityGrad')
{'Fall': 0, 'Spring': 1, 'Summer': 2, 'Winter': 3}
   group1  group2       Diff      Lower      Upper    q-value   p-value
0       3       1  19.772370  12.783724  26.761017  10.323508  0.001000
1       3       2  15.251833   8.246746  22.256919   7.944566  0.001000
2       3       0   7.673469   0.446699  14.900240   3.874442  0.032492
3       1       2   4.520538  -2.380224  11.421299   2.390313  0.330322
4       1       0  12.098901   4.973209  19.224593   6.195559  0.001000
5       2       0   7.578363   0.436547  14.720180   3.871938  0.032650


Tukey ('South', 'ActivityGrad')
{'Fall': 0, 'Spring': 1, 'Summer': 2, 'Winter': 3}
   group1  group2       Diff      Lower      Upper    q-value   p-value
0       2       0  14.624804   8.572901  20.676706   8.817790  0.001000
1       2       3  21.619576  15.553437  27.685714  13.004585  0.001000
2       2       1   9.416013   3.157904  15.674121   5.490167  0.001000
3       0       3   6.994772   1.018975  12.970569   4.271096  0.014266
4       0       1   5.208791  -0.961787  11.379370   3.080156  0.131022
5       3       1  12.203563   6.019022  18.388105   7.200138  0.001000
```

A10: Analyses of *Data Averaged Across Years and Within Hemispheres* (*News* Category)

```
results\news_category\Activity from yearly and hemisphere averaged
Anova ActivityGrad
                      df        sum_sq       mean_sq           F        PR(>F)
C(season)            3.0   5306.869860   1768.956620   16.911319  6.272503e-09
C(hemis)             1.0  88141.025641  88141.025641  842.632892  6.325519e-50
C(season):C(hemis)   3.0    511.575092    170.525031    1.630228  1.873026e-01
Residual            98.0  10250.989011    104.601929         NaN           NaN
```

```
Tukey ('North', 'ActivityGrad')
{'Fall': 0, 'Spring': 1, 'Summer': 2, 'Winter': 3}
   group1  group2       Diff     Lower       Upper   q-value   p-value
0       3       1  19.296703  8.305760  30.287646  6.603588  0.001000
1       3       2  14.989011  3.998068  25.979954  5.129438  0.003696
2       3       0   6.989011 -4.001932  17.979954  2.391732  0.339485
3       1       2   4.307692 -6.884936  15.500321  1.447586  0.713300
4       1       0  12.307692  1.115064  23.500321  4.135960  0.025949
5       2       0   8.000000 -3.192629  19.192629  2.688374  0.241002
```

```
Tukey ('South', 'ActivityGrad')
{'Fall': 0, 'Spring': 1, 'Summer': 2, 'Winter': 3}
   group1  group2       Diff      Lower      Upper   q-value   p-value
0       2       0  11.890110   1.953809  21.826411  4.500833  0.013098
1       2       3  20.197802  10.261501  30.134103  7.645593  0.001000
2       2       1   7.890110  -2.046191  17.826411  2.986690  0.163781
3       0       3   8.307692  -1.810941  18.426326  3.088093  0.142101
4       0       1   4.000000  -6.118633  14.118633  1.486859  0.697927
5       3       1  12.307692   2.189059  22.426326  4.574952  0.011344
```

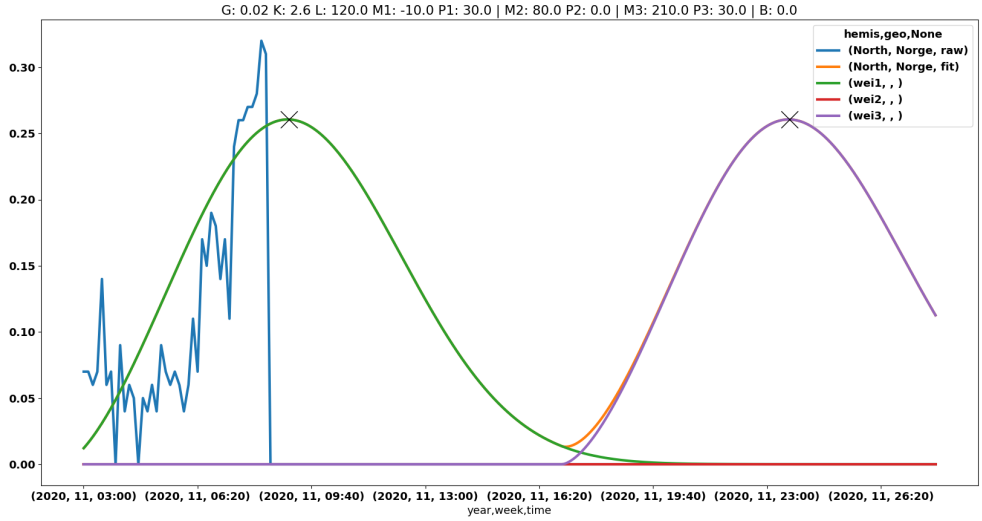
A11: *Raw Data* Analyses. (*Arts and Entertainment* Category)

```
Anova WELCH NORTH_ACTIVITYGRAD
   Source  ddof1       ddof2          F         p-unc       np2
0  season      3  198.695595  16.335768  1.575850e-09  0.101259

PostHoc Games-Howell NORTH_ACTIVITYGRAD
        A       B   mean(A)   mean(B)  ...       T        df    pval  hedges
0    Fall  Spring  964.6593  977.9111  ... -6.1686  177.3831  0.0000 -0.9126
1    Fall  Summer  964.6593  972.5275  ... -3.1415  173.7322  0.0106 -0.4638
2    Fall  Winter  964.6593  964.8261  ... -0.0661  174.2629  0.9999 -0.0097
3  Spring  Summer  977.9111  972.5275  ...  2.2342  165.9920  0.1183  0.3302
4  Spring  Winter  977.9111  964.8261  ...  5.3859  166.3864  0.0000  0.7925
5  Summer  Winter  972.5275  964.8261  ...  2.8009  180.9907  0.0286  0.4123
```

A12: Faulty Google Trends Plot that was Excluded from the Analyses

The blue line shows the faulty Google Trends graph from the *Arts and Entertainment* category (week 11, 2020) that was excluded from the analyses. The TRIWEI fit was performed and produced a very long activity length measurement that showed up as an outlier in the *Arts and Entertainment* wave plot.

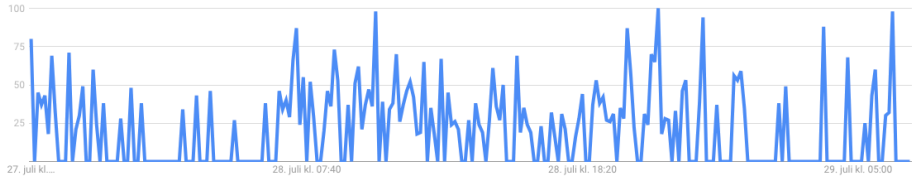


A13: *-Query in the *News* Category, Troms, Oslo, and Norway

The first graph represents search activity in the *News* category in Troms. The time span is from 27.07.23 to 29.07.23. The second and third graphs are equal to the first except that the regions are Oslo (second graph) and Norway (third graph). The daily pattern is hard to detect in the first two graphs.

Oslo | 27.07.2023, 21:00 – 29.07.2... | Nyheter | Nettsøk

Interesse over tid



Norge | 27.07.2023, 21:00 – 29.07.2... | Nyheter | Nettsøk

Interesse over tid