How not To Put A Nation to Sleep

An In-Depth Analysis of Sleep

Blake Whittington (19200321)
Gabriele Spiridaviciute(19303821)

A report submitted in part fulfilment of COMP 30780

Data Science in Practice

Demonstrator: William Blanzeisky

UCD School of Computer Science
University College Dublin

May 15, 2022

Table of Contents

Abstract

Sleep is one of the most important biological functions necessary to human survival. Acting as nature's very own charging station, sleep allows our bodies and minds to recuperate from the previous day's activities and prepares us for the day ahead. A healthy sleeping pattern paves the way for a healthy mind; without a sufficient amount of sleep, our brains simply cannot function properly.

Sleeping requirements vary from person to person, and the duration of sleep we need is greatly dependant on age. The National Sleep Foundation suggests that most adults need approximately 7-9 hours of sleep per night in order to function at their best.[1]. Our main goal for this study is to investigate the sleeping habits of people around the world, in order to gain a better understanding of what Americans can do to optimise their time, and to analyse the various substances that influence their sleep and overall health.

Chapter 1: **Introduction**

We, as humans, spend roughly 1/3 of our lives sleeping. Many studies have shown that sleep and health are strongly related. This means that we should try to prioritise getting an optimal amount of sleep (7-9 hours) as often as we can.

Our goal for this report is to give the reader an insight into where America's sleep ranks on a global scale, to understand the effects of economic health on the nation's sleeping habits and to investigate whether or not America's citizens have an accurate perception of their personal health and sleep when looking at the substances they use.

We will accomplish this by utilizing sleep survey data, economic performance metrics, and documented standards relating to tobacco, caffeine, and alcohol usage.

Chapter 2: Project Objectives

2.1 Objectives & Motivations

Knowing and understanding how different substances contribute to the quality of our sleep is vital in achieving a healthy lifestyle. Many of us are guilty of abusing caffeine, alcohol and tobacco, failing to acknowledge the impact they can have on our health and sleep. Our primary objective for this project is to understand whether or not adults, more specifically American adults, are aware of their own substance habits and the effects these habits have on their sleep and health at an individual level.

In the next chapter we will outline some of the key research questions we explored and the data that we used in the process. This will be followed by a brief summary of some 'Related Work' that has been conducted on sleep analysis throughout history. Chapter 4 will describe the datasets we used whilst carrying out our research. In Chapter 5 we will describe how we have answered each of the key research questions and present our findings in detail. Chapter 6 concludes with a summary of the key findings, a discussion of the limitations of the work as presented as well as several possibilities for future work.

2.2 Research Questions

1. RQ1 – Looking Around the World - What Can America Learn About Sleep? There is always more than one 'right' way to do something - sleep included. No two countries are the same, each with their own unique cultures and traditions. This poses the question - how do sleeping habits differ from country to country, and what effect do certain lifestyle choices have on the quality of sleep of any individual society? Throughout this paper we will focus primarily on the sleeping habits of the United States, and what Americans can learn from other countries about time allocation, in order to optimise their sleep.
2. RQ2: How Does Economic Health Affect People's Sleep? While observing economic trends, what is the relationship that links economic health to the quantity of America's sleep? Is there a correlation between certain economic health indications, and how the population sleeps?
3. RQ3 – Are People a Good Judge of Their Own Sleep and Health? We want to look at how participants respond to various questions such as "How would you rate your overall sleep?" and "How would you rate your overall health?". Then use the participant's answers to compare to external sources (standards recommendations) to see if they indicate good/bad health.

Chapter 3: **Related Work**

As the field of medicine evolved over time, so too did the area of sleep.[2] Throughout history, sleep has always been a topic of interest, however, most cultures focused on interpreting dreams, and did not pay much attention to the role of sleep in one's health.[3] In fact, the correlation between sleep and health wasn't acknowledged until 1913, when French psychologist Henri Piéron published a book titled "Le Probleme Physiologique du Sommeil" ("The Physiological Problem of Sleep"). This marked the first time sleep was studied from a physiological perspective.[3]

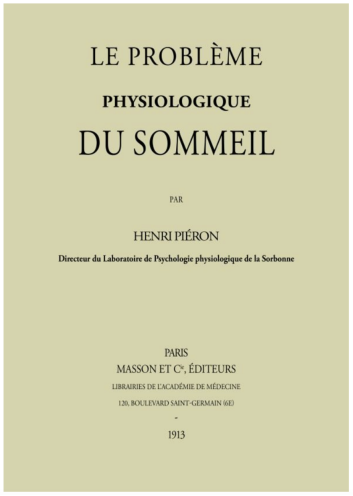

Figure 3.1: Piéron's "Le Probleme Physiologique du Sommeil"; from https://www.books.fr/pieron-le-pionnier/

In 1925, Nathaniel Kleitman, regarded by many as the "father of American sleep research", [3] joined the University of Chicago and established the world's first sleep laboratory which was filled with measuring devices designed and built by himself and his students.[2] In 1953, he and one of his students, Dr. Eugene Aserinsky, discovered R.E.M. (Rapid Eye Movement) and made the link between sleeping and dreaming.[4]

Figure 3.2: Kleitman and Aserinsky, Cave Dwellers, 1938; from https://www.the-scientist.com/foundations/cave-dwellers-1938-33966

Following the work of Kleitman and his students, there was rapid advancement within the field of sleep research. In more recent times, a large amount of attention has been placed on sleep disorders[5] and the importance of sleep for mental health.[6]. In this report, we would like to take a different approach, by addressing how various substances can affect sleeping patterns and in turn affect one's personal perception of health and sleep.

Chapter 4: Data Considerations

In this report, we will be utilising 3 main sources of data to gather information on the sleeping habits of American adults, and adults worldwide.

- Time Use from Organisation for Economic Co-operation and Development (OECD).[7]
- Time Americans Spent Sleeping from Bureau of Labour and Statistics.[8]
- 2013 Sleep and Health Survey from Sleep Foundation.[9]

We will also be utilising 3 smaller datasets to explore trends in various economic metrics.

- GDP Ratings from 2003 - 2019 from Macrotrends[10]
- Unemployment Ratings from 2003 - 2019 from Macrotrends[11]
- Inflation Ratings from 2003 - 2019 from Macrotrends[12]

Finally we will be utilising a custom dataset that we created, containing various limits and recommended standards for:

- Caffeine [13]
- Tobacco [14]
- Alcohol [15]

4.1 Data Collection

In terms of data collection, we gathered multiple datasets from a wide range of sources. We collected our OECD dataset from 'Statista' and our survey data from the 'Sleep Health Journal' website. When dealing with economic data, we found American sleep data and economic trend data through the BLS(Bureau of Labour and Statistics).

4.2 Data Preparation

Cleaning the datasets in the 'Sleep Foundation' surveys was quite the labour-intensive task, while the clean-up in the 'OECD' & 'TASS' datasets was much more tame in comparison.

When we began cleaning the sleep survey data, we quickly found that the raw data was far messier than we had originally thought. Each file came from a different survey, with different placements

of nondescript column names. This meant that we had to manually reference each individual questionnaire, and accurately replace usable column names, while dropping all other irrelevant columns.

Many of the columns were found to have values that were nondescript, labelled with basic numerical values. We swapped out these values for their descript questionnaire counterparts, in order to better manipulate the data frame and more accurately describe the analysis to come. (e.g. changing a value from '1' to 'Male', and '2' to 'Female' etc)

While BLS (TASS and economic datasets) were far easier to clean with a simple loading procedure, OECD simply required splitting the original '.xslx' file into 3 separate files. These files contained the total sleep for an entire nation, the total of men's sleep in a nation, as well as the total of women's sleep in a nation. After this was completed we were able to make an immediate start on the analysis.

Chapter 5: **Results**

5.1 RQ1: Looking Around The World – What Can Americans Learn About Sleep?

If there's one thing everyone in the world has in common, it's sleep. We all reach a certain point in the day where we close our eyes, relax our minds and recharge our bodies.

With that in mind, we live in a world full of diverse countries and cultures that often influence how we spend our time. We want to explore the similarities and differences in regard to time use and sleep across the globe. In this chapter we will explore the similarities and differences between the sleeping habits of people across the globe.

For this global analysis we will be using the OECD (Organisation for Economic Co-operation and Development) dataset. This dataset contains time use information on 33 different countries worldwide.

5.1.1 What Does Sleep Look Like Around The World?

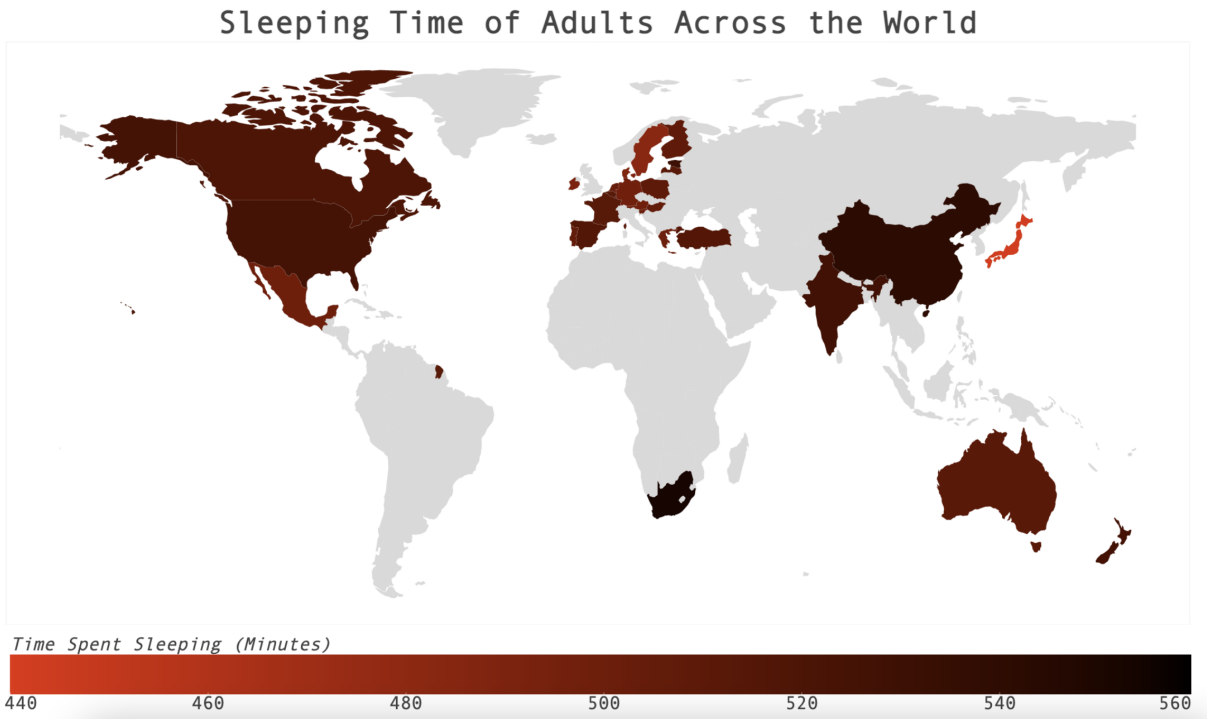

Figure 5.1: Sleeping Time Across the World

The results are presented in Figure 5.1 as a Geopandas choropleth map. We merged a shape file containing all the countries in the world [16] with the OECD dataset. This allowed us to determine the geographical location (longitude and latitude) of each of the 33 countries in the OECD dataset and map them accordingly.

Looking at how long each country spends sleeping, we can note that there is no apparent consistency between different countries and their sleeping habits. Additionally, there appears to be no uniformity between the sleep times of countries in any one given continent. In Asia, for example, we see that China averages out at around 540 minutes (9 hours) of sleep per night, while on the opposite end of the scale in Japan, citizens are getting on average 440 minutes, or just over 7 hours of sleep each night.

This comes as no surprise as Japan has always had a poor reputation for its sleeping habits [16]. However, Japan takes pride in its citizens working themselves to exhaustion, even considering it a subtle sign of diligence.[17]. Both Japan's government and its companies are guilty of enforcing intense work schedules that often lead its citizens to burnout. A common phrase you'll hear in Japan is 'karoshi', which roughly translates to "death by overwork". This is a legal term recognized by Japanese government as a cause of death.[18] To combat this exhaustion, Japan has introduced short sleeping breaks for workers; the 'hirune' which translates to "lunchtime sleep" and 'inemuri' which translates to "being present while asleep".

5.1.2 Is There A Difference In The Way Men and Women Sleep Across The Globe?

Even though there is not uniformity of countries in a geographic sense, we can still attempt to see if there is a difference in how men and women sleep in these countries.

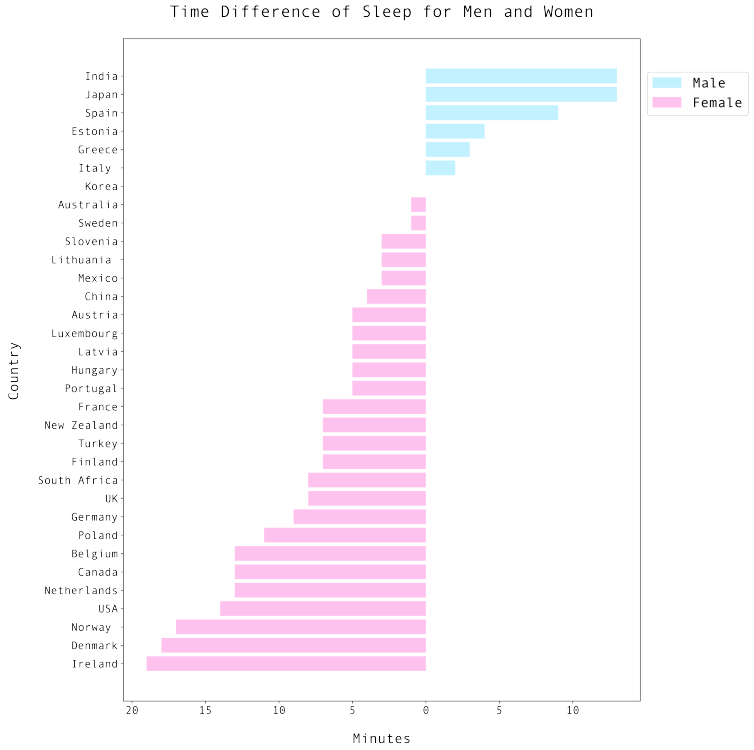

Figure 5.2: Time Difference Between Genders In Terms of Sleep

In Figure 5.2, we see that Japan and India lead in the net positive male difference section with 13 minutes of extra sleep when compared against women in the same countries. Ireland leads in the net positive women difference section with women sleeping 19 minutes longer than men on average.

Since, overall, we are talking about hours of sleep in these countries, it can be said that there is no significant difference in the way women and men sleep. We are all human beings with a largely identical need for sleep in order to function properly.

5.1.3 How Do Countries Around the World Spend Their Day?

We can see that sleep times vary throughout the world, with some countries having a 2-hour discrepancy in average rest time per night. Despite the 2-hour disparity, it is clear from Figure 5.3 that each country makes a similar effort to ensure their citizens are well rested, despite allocating their time differently.

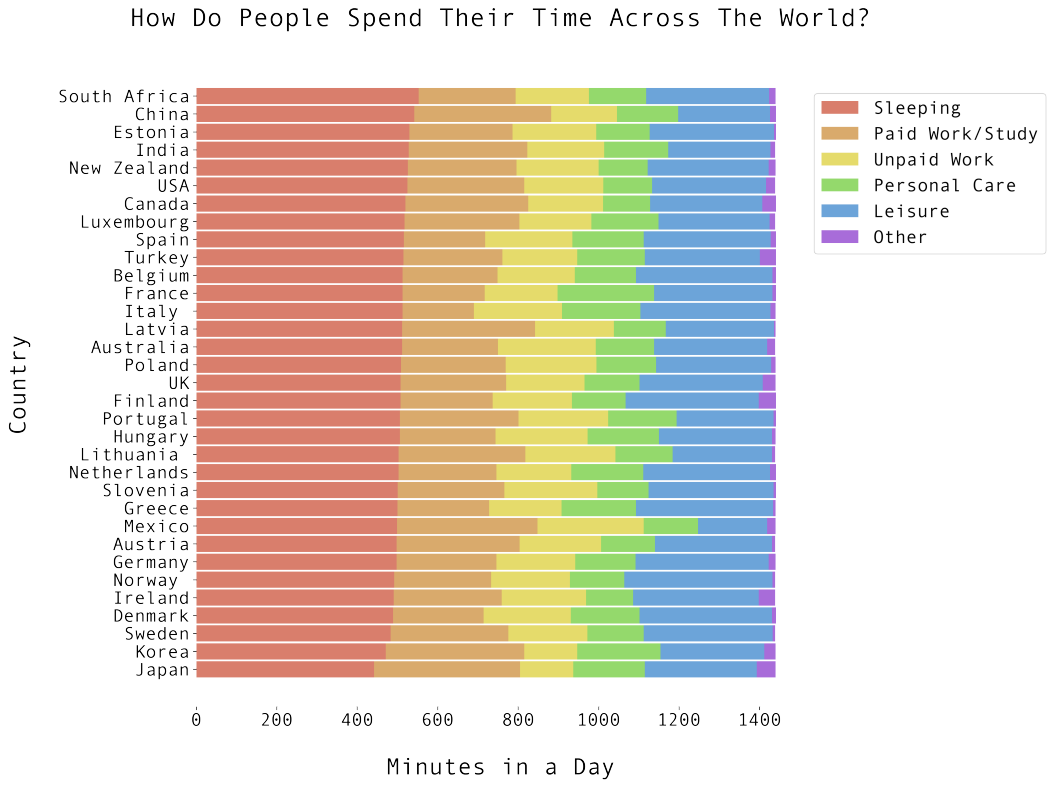

Figure 5.3: Time Usage Across the World

5.1.4 How to the Most/Least Well Rested Countries Divide Up Their Time?

One thing we want to take a look at is how the 3 Longest Sleeping Countries and the 3 Shortest Sleeping Countries divide their time, and if there are any common behaviours and patterns that lead them to good or bad sleep.

The top 3 sleeping countries are - South Africa, China and Estonia.

For these countries, we found that their average distribution of time is quite rigid. The only instance where we see a change is in China, where paid work takes up more time than leisurely activities. This is an understandable switch because China is a very work-oriented country[19] and this change might be caused by companies enforcing the 996 rule whereby workers work from 9am-9pm for 6 days a week.[20]

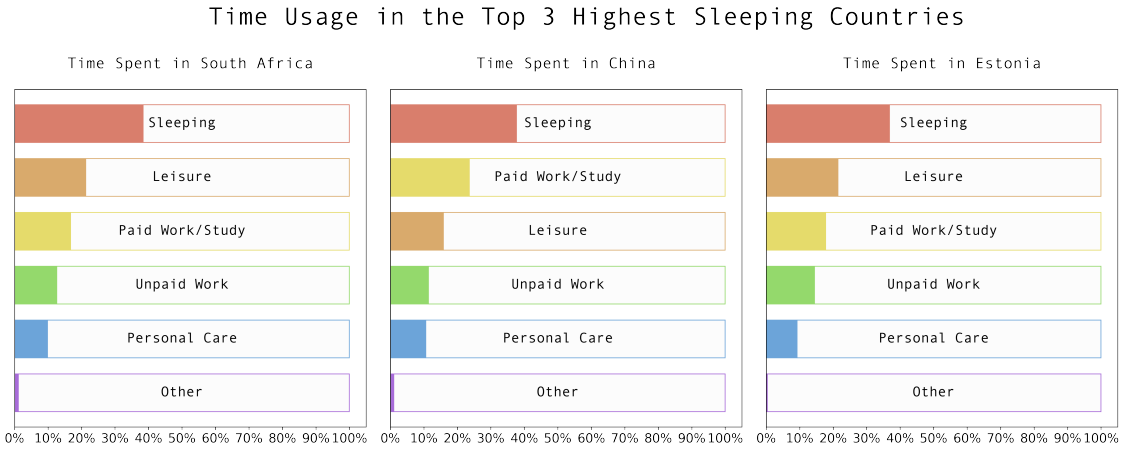

Figure 5.4: Top 3 Sleeping Countries

Looking at the shortest sleeping countries (Sweden, Korea and Japan) we can see that switching time use category priorities is more common. It is then assumed that for the countries that are sleeping poorly, the priorities of time use are a case by case basis.

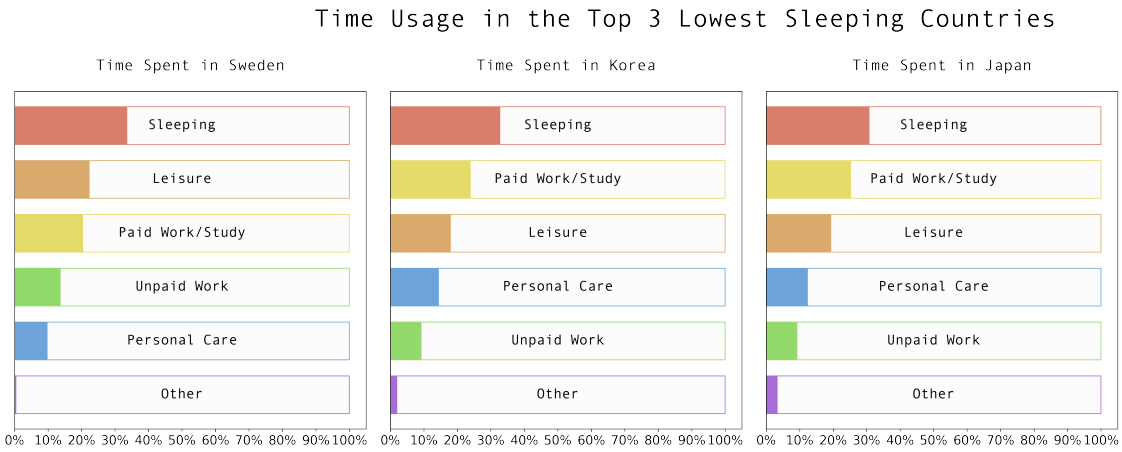

Figure 5.5: Bottom 3 Sleeping Countries

While the top 3 stay fairly steady we can see that the bottom 3 shift around more.

5.1.5 What Can America Learn from the distribution of time use?

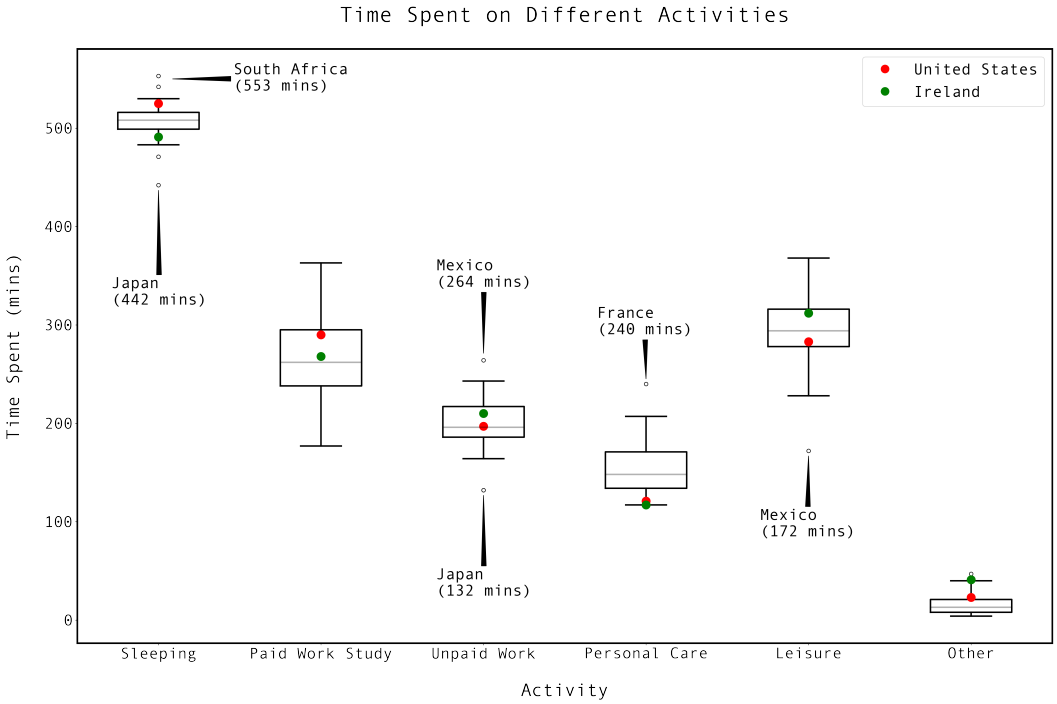

Figure 5.6: Time Distribution For Various Activities

As we can see in Figure 5.6, most countries across the world tend to sleep within a strict window. While some countries such as South Africa and Japan deviate quite strongly, America places quite well in terms of sleep - it's at the max without becoming an outlier. With that in mind, there are aspects that Americans can improve on such as personal care. Seeing as America ranks so well in terms of its sleep, some Americans might be tempted to sacrifice this in order to improve their personal care. However they should instead consider drawing from paid work and study since the interquartile range is so high and the US is very close to the 3rd Quartile. Something the US government should probably consider is increasing the minimum wage so people don't have to work as much as they do, allowing them more time to focus on personal care which may lead to an overall improvement of life quality.

5.1.6 Summary

The goal of this research question is to investigate whether or not sleeping habits vary around the world and if so, if there is much disparity in sleep times between different countries.

Using our worldwide sleep data and an external shape file, we were able to look at how sleeping times change across the globe, in the form of a choropleth. We saw some considerable differences, some varying by up to two hours.

We saw in Figure 5.2 that in most countries, women tend to sleep more than men, albeit the difference quite minor at about 20 minutes more.

Figure 5.3 showed us that most countries make sleep a priority, however, the way they allocate the rest of their time changes.

We wanted to see if the way countries allocate their time has an impact on how their sleep ranks worldwide. The Top 3 sleeping countries displayed less disparity than the Bottom 3 sleeping countries; as well as that, the top 3 sleeping countries seemed to spend more time on leisurely activities than doing paid work or study. We could say that that those who play hard tend to get more rest than those who work hard.

After seeing how the top and bottom 3 sleeping countries spent their time, we wanted to see how America, specifically, allocated its time. Generally speaking, we saw that America placed quite well in terms of its sleep, however, there were areas such as personal care that needed improvement. In order to improve this, we suggest that America draws time from its work and paid study portion as they work quite long hours when compared to the rest of the world. We think the US government should raise minimum wages to reduce the amount of time Americans spend working - leaving them with more time to focus on personal care.

5.2 RQ2: How Does Economic Health Affect People's Sleep?

Various economic factors such as inflation, unemployment and Gross Domestic Product (GDP) can affect the quality of our everyday lives. A poor economic status leads to significant economic damage[21] that impacts a country's population as a whole. We want to investigate if sudden changes in different economic factors affect the sleep patterns of US citizens and which ones are the most influential in dictating the sleep of a nation.

5.2.1 How Do Changes In Inflation Affect Sleeping Patterns?

First we want to look at how sleep changes over time with regards to changes in inflation rates.

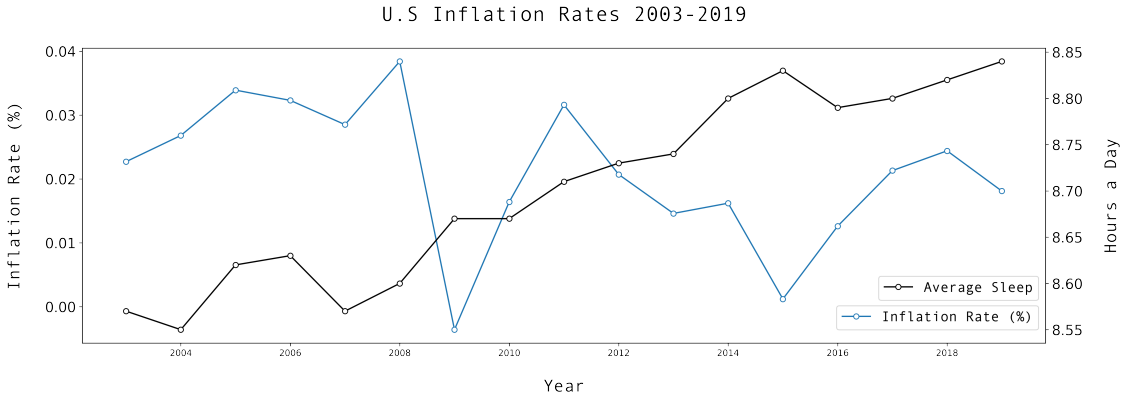

Figure 5.7: US Inflation Rate and Sleep Time 2003-2019

We can see in Fig 5.7 that inflation increased steadily from 2003, until it peaked in 2005 and subsequently began to decrease. Within this time frame, we can see that sleep time increased and decreased accordingly with inflation. In 2008, annual inflation in the US rose to 3.8%, the highest we see between 2003 - 2019. In 2009, we see a massive drop in inflation rates. This deflationary period occurred during the Great Recession, a time that was very difficult for American citizens[22].

Even though the drop in inflation rates does not seem to drastically alter how long people spend sleeping, we can see that when the sudden deflation occurs both in 2009 and 2015, the period of time people spend sleeping tends to increase at a higher rate.

5.2.2 How Do Changes In Unemployment Affect Sleeping Patterns?

Now let's move onto unemployment and investigate if there is a clearer link between that and the time Americans spend sleeping.

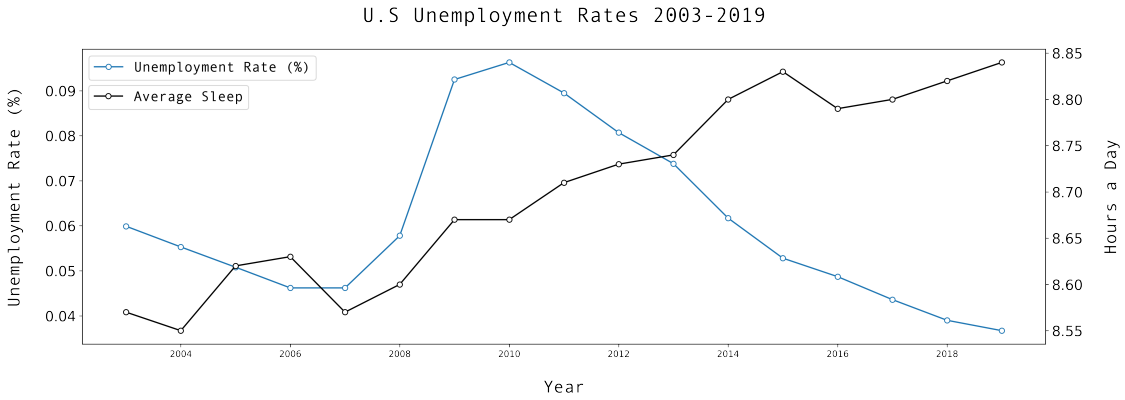

Figure 5.8: US Unemployment Rate and Sleep Time 2003-2019

The relationship between unemployment and personal well-being is widespread knowledge. Many studies the have documented the connection between joblessness and various health factors such as obesity[23][24], heart disease[25][26] and mental illness[27][28][29] . However, the relationship between unemployment and sleep is less researched.[30]

We can see that as the Great Depression begins in 2007, America's unemployment rates start to soar. In only 3 years, the rate of unemployment goes from 4.6% up to 9.6%. Even though this sudden increase in unemployment is joined by an increase in the amount of time Americans spend sleeping, it is unclear whether or not they are correlated as sleep time continues to rise beyond 2010, when unemployment rates start going down.

5.2.3 How Do Changes In Global Domestic Product Affect Sleeping Patterns?

Another economic factor that we would like to look at is Gross Domestic Product (GDP).

GDP measures the total value of goods produced in a certain country during a certain period of time[31]

A great deal of research has been done on how sleep deprivation can lead to large economic expenses, in terms of poor labour productivity and GDP loss.[32] However, the opposite has not been looked at in depth. We want to see if a change in GDP affects how long Americans sleep.

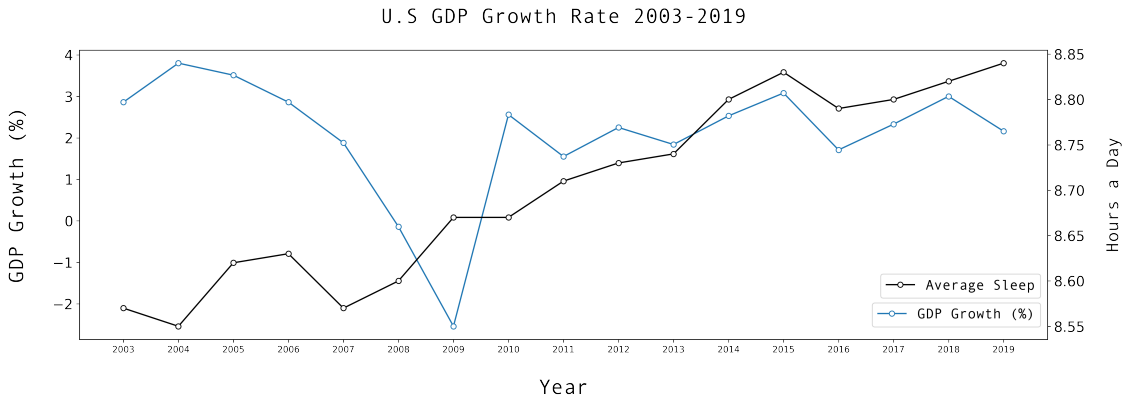

Figure 5.9: US GDP Rate and Sleep Time 2003-2019

Looking at Figure 5.9, we can see that there is no clear trend between the rate of GDP and the time Americans spend sleeping. Even with the sudden dip in GDP between 2004 and 2009, sleep time was going up. In fact, while the rate of GDP jumped up and down between 2003 and 2019, sleep increased steadily over time without being influenced by GDP rates. This is in not surprising for us as we have been looking at various countries around the world sleeping better than others with a higher GDP. If South America(GDP: $300 Billion) has been proven to, on average, sleep better than Korea(GDP: $1.6 Trillion) then it follows that changes in GDP in a singular country like the USA should not affect sleep patterns.

Now that we've seen how different economic factors and sleep time in America changed over time, let's have a look at which economic factors had the highest correlation to sleep.

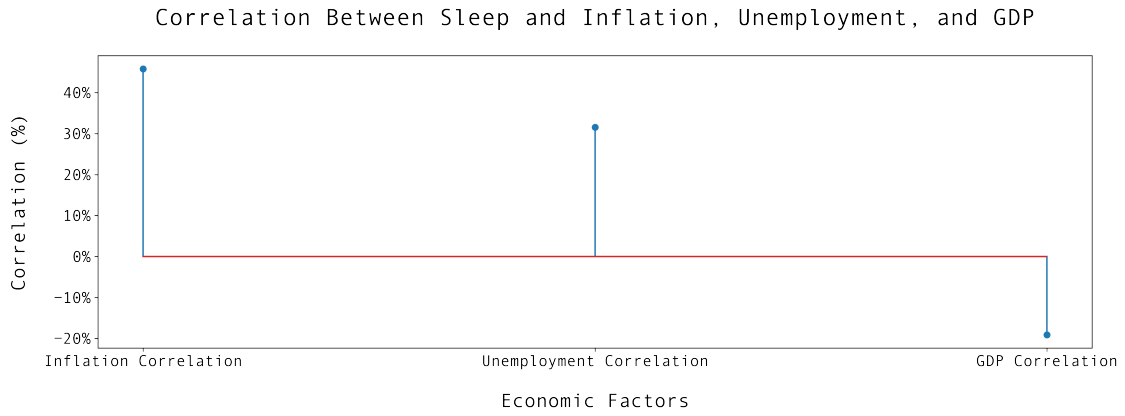

Figure 5.10: Correlation Between Various Economic Factors and Sleep

As we can see in Figure 5.10, inflation has the highest correlation to sleep. We can say that inflation is the most influential economic factor when it comes to influences how long Americans spend sleeping, followed by unemployment then followed by GDP which shows a negative relationship.

This is an understandable order as inflation can strongly affect people's personal lives. Necessities such as food, fuel, and utilities soar in price leaving people with less money for savings or discretionary spending. In general the lack of financial freedom causes strain on people leading them to alter their sleeping patterns.

5.2.4 Summary

In this research question, we wanted to see if changes in various economic factors played a big role in how long Americans spend sleeping. Generally speaking, there did not seem to be any clear trends in how inflation, unemployment and Gross Domestic Product (GDP) affected sleeping patterns. We looked a little bit closer to determine which of these factors had the strongest relation to sleep time, and we found that inflation had a positive correlation of 45%, followed quite closely by unemployment with 32%. GDP, however, had a negative correlation with sleep time with -20%.

5.3 RQ3: Are People A Good Judge Of Their Own Sleep And Health?

Within the realm of sleep analysis, it's hard to talk about how the world effects the individual without talking about the individuals habits and daily activities. It is precisely because of this that we want to analyse the surveyors of the 2013 sleep poll "Sleep and Health" conducted by sleepfoundation.org in order to understand if they are accurately assessing their own sleep and health.

We will be evaluating this question on the basis of how the surveyors answer the following 2 questions asked in the survey:

- How would you rate your overall health?
- During the past two weeks, how would you rate your overall sleep quality?

Both of these questions set the foundation for the current mentality of each surveyor. Combined with other, more finite and specific questions asked, we set the foundation for being able to directly compare the statistical to the qualitative. In specific, we want to look at how each surveyor recorded their usage of 3 main substances that are of detriment to their health and sleep.

Questions we will be looking at in close detail are:

Alcohol Related:

- In the last two weeks, how many days have you had an alcoholic beverage?
- On days you have a drink, how many alcoholic beverages do you normally consume?

Caffeine Related:

- Thinking about the last two weeks, how many 12 ounce servings of caffeinated beverages, such as soda, soft drinks, coffee, tea, and energy drinks do you drink on an average weekday or workday ?

Tobacco Related:

- Do you now or have you ever smoked cigarettes, cigars or a pipe?

In order to answer this question to the best quality and reach possibly, it was decided to split this question into 2 distinct approaches. A more general analysis to answer this question on the basis of the population of surveyors taken as a whole, and a clustering approach to split the surveyors into meaningful groups to see how various clusters feel about their health and sleep patterns.

5.3.1 General Analysis Approach

How do Americans Rate Their Health and Sleep?

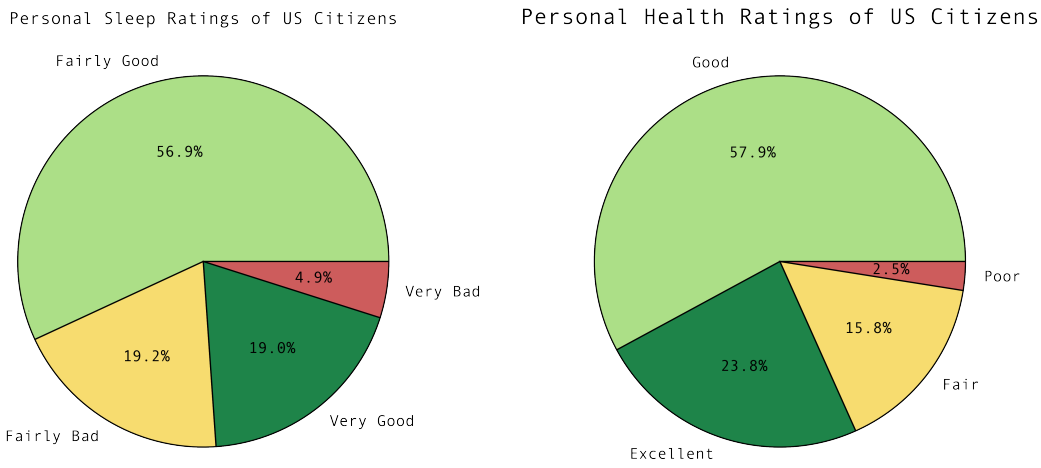

Figure 5.11: Personal Sleep and Health Ratings of US Citizens

It is evident from Figure 5.11, that most of America's citizens would rate their sleep as "Fairly Good", and rate their health as "Good". Now equipped with the understanding that Americans tend to rate their health and sleep as predominantly "Good", the truthfulness of these perceptions must be called into question. How can a population become healthier and sleep more, if they believe they don't need improvement to begin with? In this analysis, the use of caffeine, tobacco, and alcohol are our primary focus. Caffeine has been proven to worsen sleep, tobacco is an obvious detriment to one's health, while alcohol has been shown to cause harm to the quality of both sleep and health.

In order to gain a grand overarching view of troubles facing everyone in the 2013 sleep survey, a factual statistical basis for standards is needed to compare against the health and sleep perceptions that are presented. The analysis from hereon will involve a dataset built from the ground up, which contains recommended limits for each of these substances. The recommended limits describe the moment you start putting your health and sleep at risk through the consumption of each substance.

gender	alcohol_(drink)	tobacco	caffeine_(mg)
men_limit_day	14	0	400
women_limit_day	7	0	400
men_limit_week	98	0	2800
women_limit_week	49	0	2800
men_limit_2_week	196	0	5600
women_limit_2_week	98	0	5600

Figure 5.12: External Substance Dataset

Do People Recognize Caffeine's Affect on Sleep?

Keeping these substance metrics in mind, the first of our analysis of health perception starts with caffeine.

The survey question asked regarding caffeine is as follows : "Thinking about the last two weeks, how many 12 ounce servings of caffeinated beverages such as soda, soft drinks, coffee, tea, and energy drinks do you drink on an average weekday or workday?". Since the specific beverages that each surveyor is drinking when consuming 12 ounce servings is not explicitly stated, the average caffeine in a 12 ounce serving of soda, coffee, tea, and energy drinks were all combined to create an average 12 ounce caffeine serving. Looking back at the format the surveyors were given to answer the question, they were given 3 different sections:

a. "Between 5:00 AM and noon?"

b. "Between noon and 5:00 PM?"

c. "Between 5:00 PM and 5:00 AM the next morning?"

A sample of the survey answers in our dataset looks like this:

	caffeine_between_05:00-11:59	caffeine_between_12:00-16:59	caffeine_between_17:00-04:59
0	1	0	0
1	2	2	0
2	0	2	0
3	2	2	0
4	2	0	1

The problem we ran into here is that we can't determine if someone is above or below the daily caffeine threshold since we only have data on the caffeine intake for different time-frames of the day. So what we want to do is combine these to get how much caffeine they consumed through out the day.

	caffeine_servings_a_day
0	1
1	4
2	2
3	4
4	3

We still don't have the milligrams of caffeine that they consume a day. To get a figure for this, we calculated the average milligrams of caffeine in a 12oz serving of these drinks.

Since we're dealing with 4 different caffeinated beverages (soda, coffee, tea, and energy drinks) we want to find what the average amount of caffeine is in an average caffeinated beverage. After calculating it we get 77.81 mg of caffeine in average caffeinated beverage

Now we can see who falls above the caffeine threshold of 400mg of caffeine.

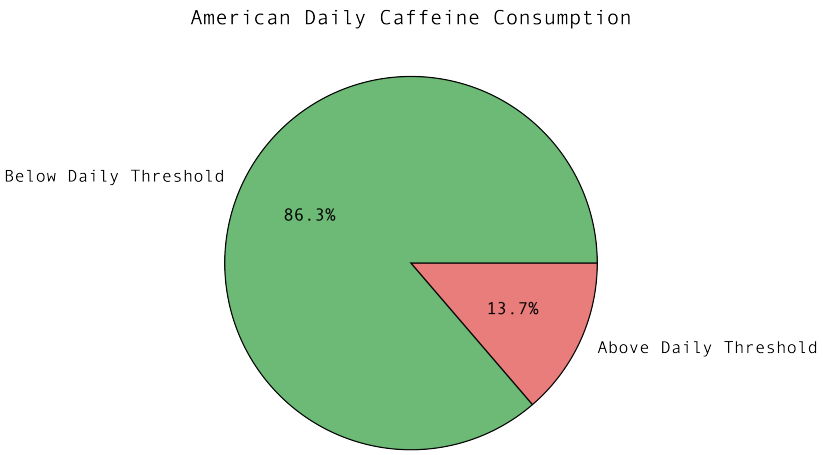

Figure 5.13: Percentage of US Citizens Above and Below the Caffeine Threshold

Now we want to delve deeper, to see if the time Americans consume caffeine has an effect on how they would rate their sleep and health.

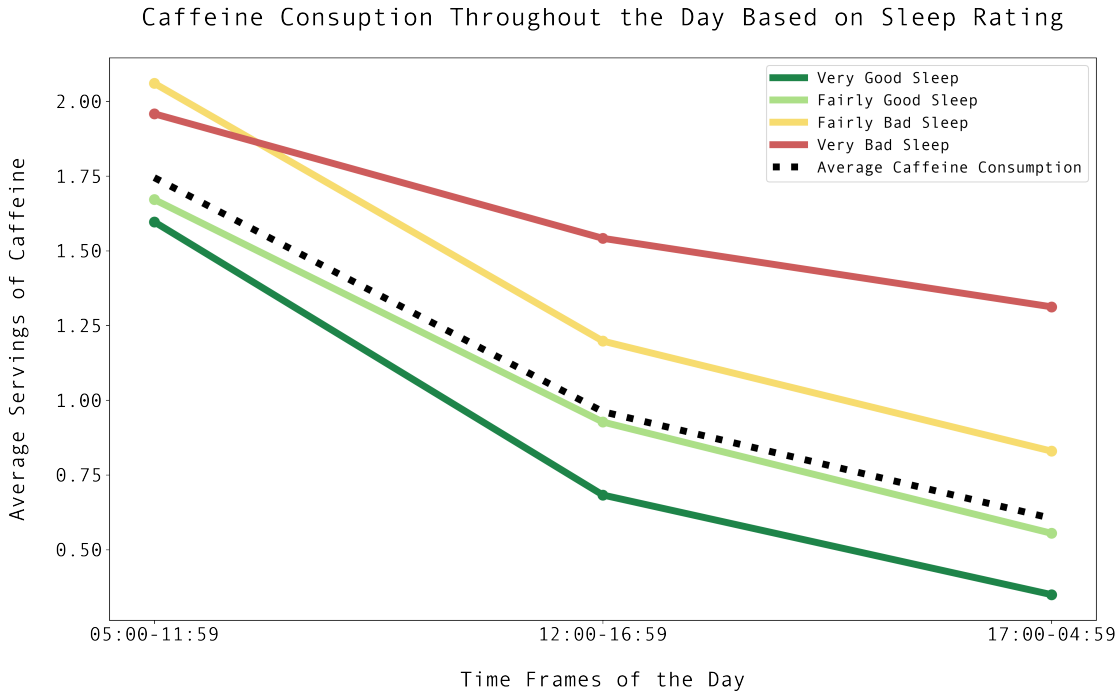

Figure 5.14: Caffeine Intake Throughout the Day

It's common knowledge that consuming caffeine later in the day worsens one's sleep. We can see in Figure 5.14 that generally, people tend to drink less caffeine as the day progresses.

An interesting note to make is that those who rate their sleep as 'Very Good' or 'Fairly Good', tend to stay below the average servings line, while those who rate it as 'Fairly Bad' or 'Very Bad' are above it. This shows us that people can understand the effect that caffeine has on sleep, and that they rate their sleep quality accordingly. It also gives us a very interesting understanding into the human psyche. Since there isn't a single group that is drastically differentiating themselves from the rest between the hours of 5AM to 12PM, it is clear, everyone knows that it doesn't matter how much caffeine is being consumed in the morning, since every one group drinks over 1.5 servings

of caffeine within this time frame. The one group that exemplifies this very well is those who rate their sleep as 'Fairly Bad'. This group starts their day consuming more caffeine that those who rate their health as 'Very Bad', yet by mid-day, the surveyors who have a 'Fairly Bad' health rating are drinking less than their 'Very Bad' counterparts. This shows recognition of the importance of drinking less caffeine as they day goes on, even though they don't get good sleep due to other factors.

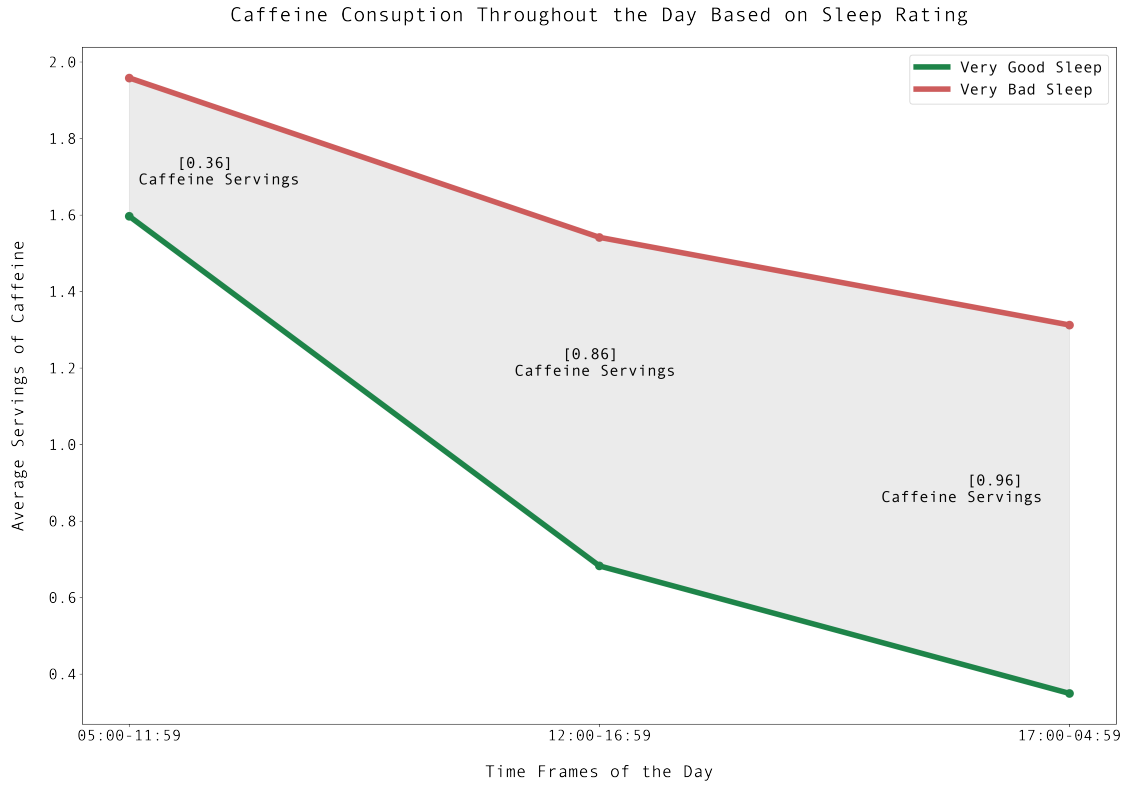

Figure 5.15: Caffeine Intake Throughout the Day - Very Good Health vs Very Bad Health

Looking at the two opposite ends of the scale in Figure 5.15, we can see that those who rate their sleep as 'Very Good' tend to drink less caffeine than those who rate their sleep as 'Very Bad' - most notably, a difference of 0.96, almost a whole serving between 5PM and 5AM - the time when you should really be limiting your caffeine intake. It's clear that people can recognise the negative effect that caffeine has on sleep when dealing with the two extremes of sleep perception.

How does Smoking Influence How People Perceive Their Own Health?

Moving on from caffeine, we want to take a look at tobacco usage to see if people take their tobacco intake into account when rating their health.

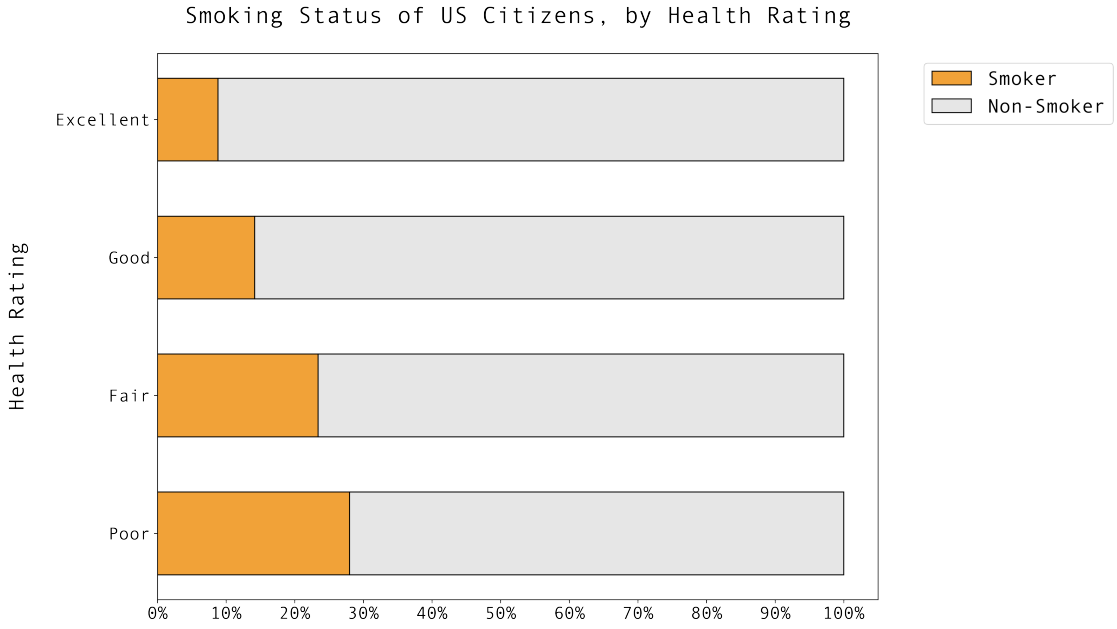

Figure 5.16: Health Status According to Tobacco

In Figure 5.16, we can see that as health quality worsens, the percentage of smokers increase. This could lead us to believe that people can acknowledge the effect of tobacco on their health, however, an interesting thing to note is that of those who rate their health as 'Excellent', 9% smoke, and of those who rate their health as 'Good', 14% smoke. Clearly people are not perfect when it comes to assessing their health.

How Much Alcohol is Too Much For Men and Women?

Now looking at alcohol, we want to see how many people are guilty of drinking too much. But how much at is too much? Men can drink up to 28 standard drinks in 2 weeks and it would still be considered safe. Anything above 28 standard drinks is dangerous. For women the limit is 14 standard drinks for 2 weeks. The yellow sections in both Figure 5.17 and 5.18 contain people whose drinking range is non-determinable i.e. we cannot accurately state whether they are above or below the threshold due to their ranges. Because of this we will focusing on the red and green sections.

Drinking Range for Males (2 Weeks)

	0 Days	1-3 Days	4-6 Days	7-10 Days	10+ Days
0 Drinks	0 Drinks	0 Drinks	0 Drinks	0 Drinks	0 Drinks
1-2 Drinks	0 Drinks	1-6 Drinks	4-12 Drinks	7-20 Drinks	10+ Drinks
3-5 Drinks	0 Drinks	3-15 Drinks	12-30 Drinks	21-50 Drinks	30+ Drinks
6-9 Drinks	0 Drinks	6-27 Drinks	20-54 Drinks	42-90 Drinks	60+ Drinks
10+ Drinks	0 Drinks	10-30+ Drinks	40-60+ Drinks	70-100+ Drinks	100+ Drinks

Figure 5.17: Drinking Ranges for Men (2 Weeks)

Drinking Range for Females (2 Weeks)

	0 Days	1-3 Days	4-6 Days	7-10 Days	10+ Days
0 Drinks	0 Drinks	0 Drinks	0 Drinks	0 Drinks	0 Drinks
1-2 Drinks	0 Drinks	1-6 Drinks	4-12 Drinks	7-20 Drinks	10+ Drinks
3-5 Drinks	0 Drinks	3-15 Drinks	12-30 Drinks	21-50 Drinks	30+ Drinks
6-9 Drinks	0 Drinks	6-27 Drinks	20-54 Drinks	42-90 Drinks	60+ Drinks
10+ Drinks	0 Drinks	10-30+ Drinks	40-60+ Drinks	70-100+ Drinks	100+ Drinks

Figure 5.18: Drinking Ranges for Women (2 Weeks)

Does Alcohol Influence How People Perceive Their Own Health?

Now that we know how much alcohol men and women can consume in two weeks, let's examine how many people are above and below the alcohol limit when looking at different health ratings.

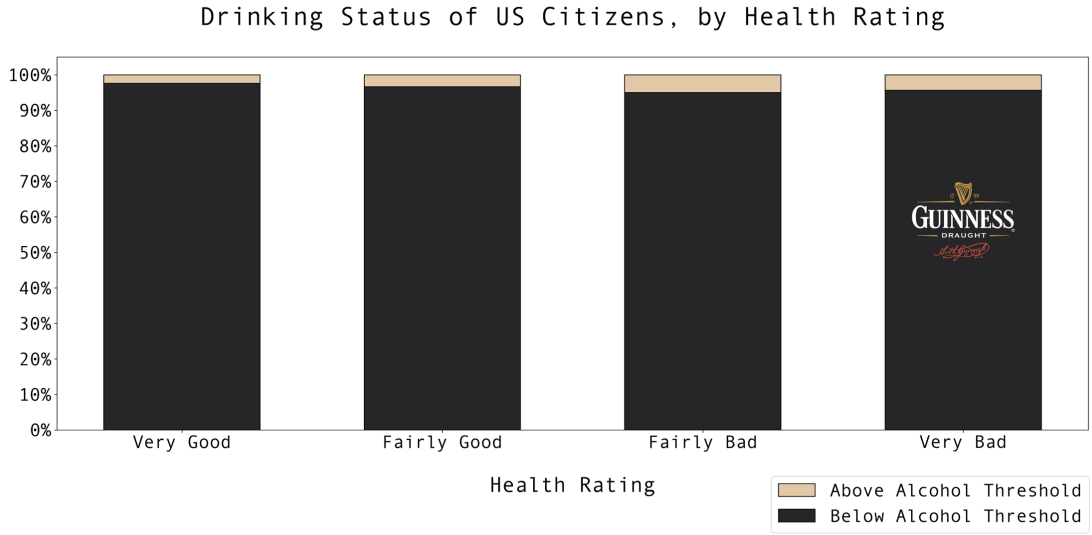

Figure 5.19: Health Status According to Alcohol Thresholds

Looking at Figure 5.19, we can see that most Americans are below the alcohol intake level. In a similar way to the tobacco graph5.16, we can see that as health quality gets worse, the percentage of people who are above the alcohol limit gets higher. Something to note is that those who rate their health as 'Fairly Bad', contain more people above the alcohol threshold than those who would rate their health quality as Very Bad'. While the surveyors that rate their health as 'Very Bad' and 'Fairly Bad' are correct in saying that they are on the lower end of the scale, they are not nearly as accurate in determining just how bad their health really is, when compared to those who rate their health as 'Very Good' and 'Fairly Good'.

Now we want to look at how each of these substances affect certain groups of our surveyors.

5.3.2 Clustering Approach

As the project progressed, it was clear that clustering was a very real prospect to explore since we had so much data on each surveyor. With 140 usable columns of information on each surveyor, we knew that there were a few key columns that we should consider:

gender	health_rating	sleep_rating	marital_status	age_bin	bmi_bin
Male	Fair	Very Good	single	30-40	20-25
Female	Good	Fairly Bad	married_or_partnered	50-60	30-35
Male	Good	Fairly Good	separated	50-60	25-30
Male	Excellent	Very Good	married_or_partnered	50-60	20-25
Female	Good	Fairly Good	single	30-40	20-25

K-MODES Algorithm

With categorical data, there are no accurately spaced values for an algorithm like K-Means, which requires the use of eigenvalues to reference. This is why K-MODES was used [33]. K-MODES replaces the role of mean in K-Means with dissimilarity(mode) as the main metric for clustering.

$$P(W,Q)=\sum_{l=1}^{k}\sum_{i=1}^{n}w_{il}d_{sim}(x_i,q_l) \quad (1)$$

where, w_{il} is an $N\times K$ matrix where each element belongs to 0 or 1. N is the total number data objects and K is the number of clusters. $d_{sim}(x_i,q_l)$ is the simple dissimilarity measure and it is defined in the following Eq.(2).

$$d_{sim}(x_i,q_i)=\sum_{j=1}^{m}\delta(x_{ij},z_{lj}) \quad (2)$$

where, $\delta(x_{ij}, q_{lj})$ is calculated using the following Eq.(3)

$$\delta(x_{ij},z_{lj})=\begin{cases}1 & if\ x_{ij}=z_{lj}\\ 0 & if\ x_{ij}\neq z_{lj}\end{cases} \quad (3)$$

Figure 5.20: K Modes Algorithm [34]

With the specifics of the algorithm covered, the pre-processing and model parameters can be discussed.

First the surveyors that did not contain a BMI value or valid marital status were dropped from the dataframe. Then age and BMI were binned to create a united front of purely categorical data to cluster.

The scikit-learn pre-processing encoder was then applied to each column in order to get the data into the correct format for the K-MODES algorithm. The notable parameters used were the amount of clusters and the initializing method[35] [36]. The older method created by Huang,Z. in 1997 was used to create a graph using the elbow method. The purpose of the elbow method is to discover the optimal number of clusters to create.

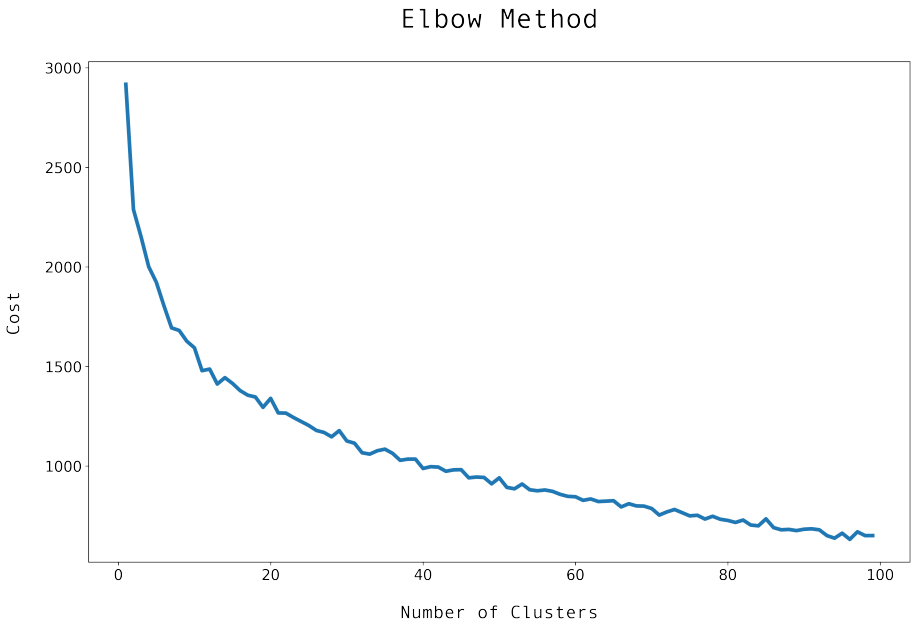

Figure 5.21: Elbow Method

It was then decided that K(the # of clusters) should equal 20. With the elbow method completed the K-MODES algorithm with k=20 is run on the processed dataframe. A column designated for the clusters is created in both a encoded version of the dataframe and the labelled version. The labelled dataframe is saved to be used later in the analysis.

Principal Component Analysis

The encoded dataframe is then used in a principal component analysis in 3 dimensions to visualize the 20 clusters. The explained variation is 67.37% which is high enough to give a decently accurate representation of each surveyor.

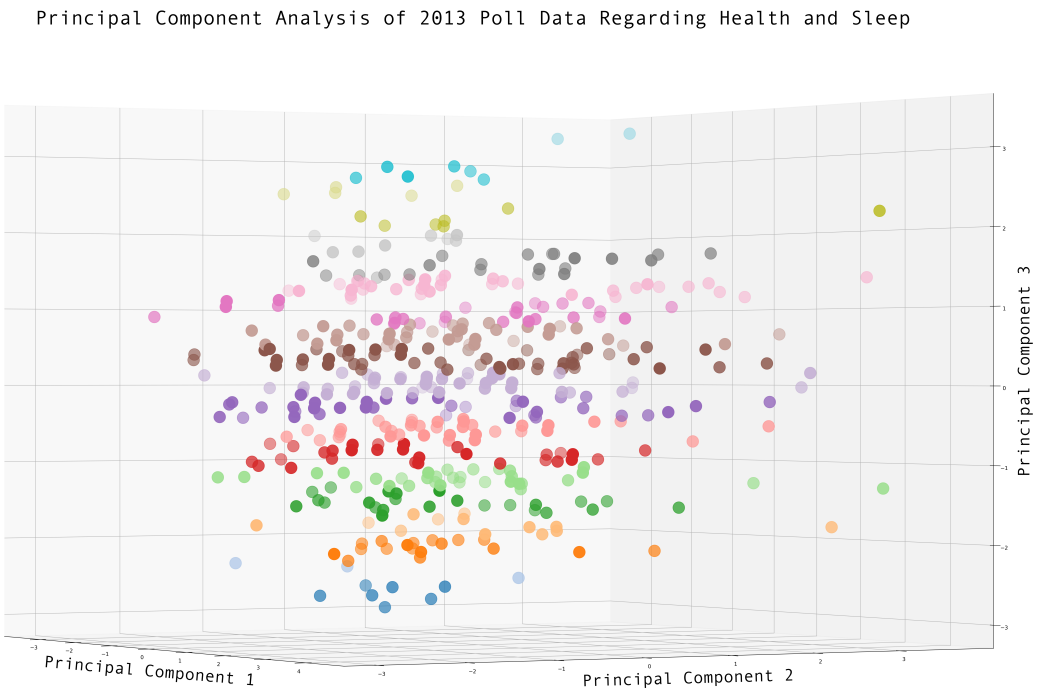

Figure 5.22: Principal Component Analysis of KMODES Clustering

Cluster Population Distribution

With the principal component analysis in Figure 5.22 showing promise that there are significant clusters that can be reviewed, the process of selecting clusters becomes a necessity.

The more straightforward way of approaching this, which is the method utilized in figures 5.23, 5.24, 5.25, 5.26, 5.27, and 5.28 is to observe the distributions of the populations of each clustering response to different columns.

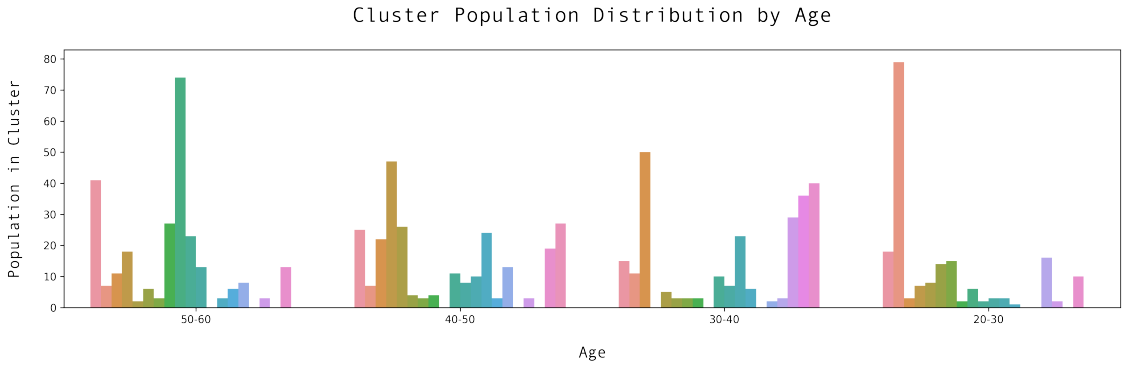

Figure 5.23: Age Distribution of KMODES Clusters

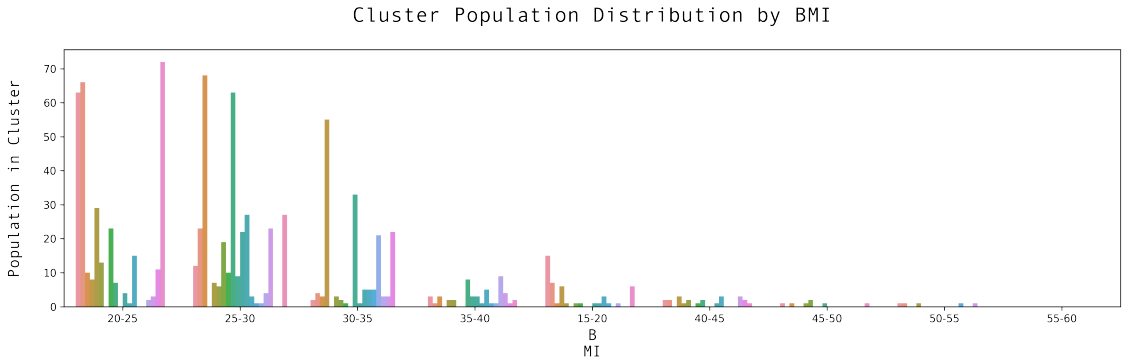

Figure 5.24: BMI Distribution of KMODES Clusters

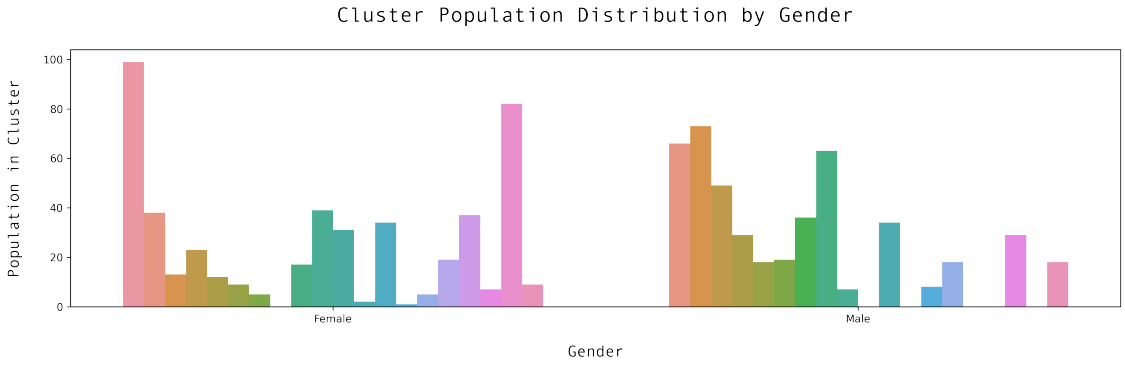

Figure 5.25: Gender Distribution of KMODES Clusters

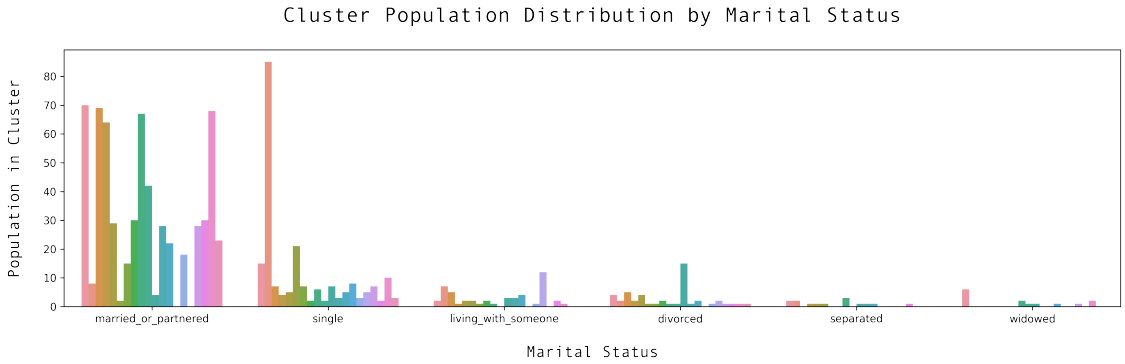

Figure 5.26: Marital Status Distribution of KMODES Clusters

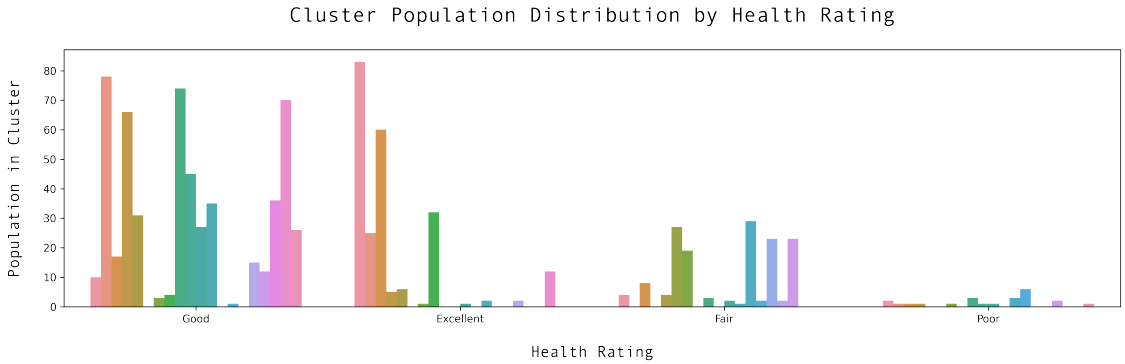

Figure 5.27: Health Rating Distribution of KMODES Clusters

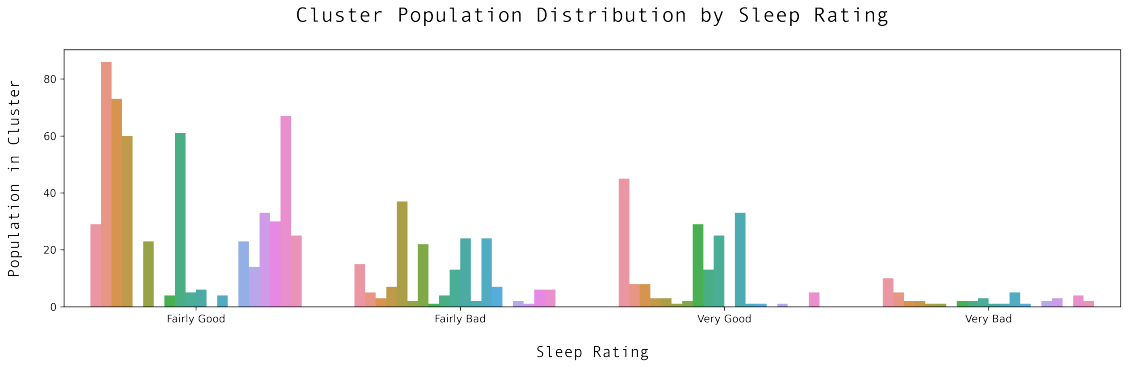

Figure 5.28: Sleep Rating Distribution of KMODES Clusters

It is evident that no cluster shares a common pattern of substance abuse, which shows that different groups of individuals tend to suffer from different problems. In regards to the 4 main clusters we are analysing:

cluster 14 and cluster 6 struggle the most with tobacco usage, while cluster 3 and cluster 2 suffer the most from caffeine problems.

5.3.3 What Are the Main Clusters to Analyze?

Looking back at the previous section of this question titled "General Analysis Approach" at the very end of each of the files related to alcohol, tobacco, and caffeine, newly modified dataframes are all saved and loaded into a new file titled merge_notebooks.ipynb. In this file all 3 modified dataframes are merged into one dataframe which has the content for the metric comparison analysis columns and which surveyors belong to their respective clusters. In this analysis, we focus on 4 of the 20 largest clusters. In order to answer the question of who is a good judge of their own sleep and health, we focus on large clusters filled with sleep and health ratings that are 'Fairly Good' and 'Good' respectively. This lets us directly compare multiple high surveyor count clusters who all rate their sleep and health as 'Good' in some capacity.

The cluster contents that are observed in this report are as follows:

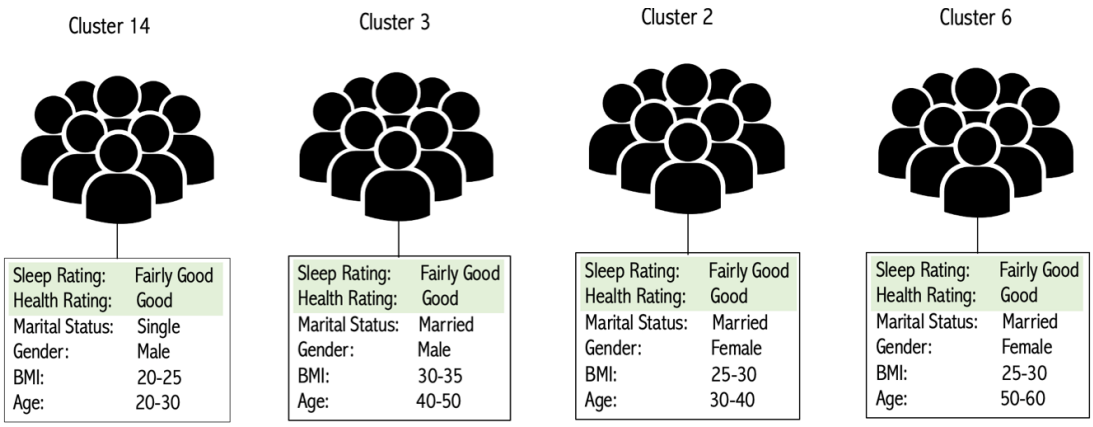

Figure 5.29: Most Important Homogeneously Rated Clusters

5.3.4 Do The Selected Clusters Suffer From Substance Abuse?

Now that the contents of the clusters is clear, and they are observed to be the clusters we want to analyse for the health perceptions of specific surveyors, they discussion of our finding from the clusters can begin.

The first observation we want to make is the percentage of 'healthy' and 'unhealthy' surveyors that are in each cluster. The way we do this is by treating a person as 'unhealthy' if they smoke, drink alcohol in excess, or consume caffeine in excess.

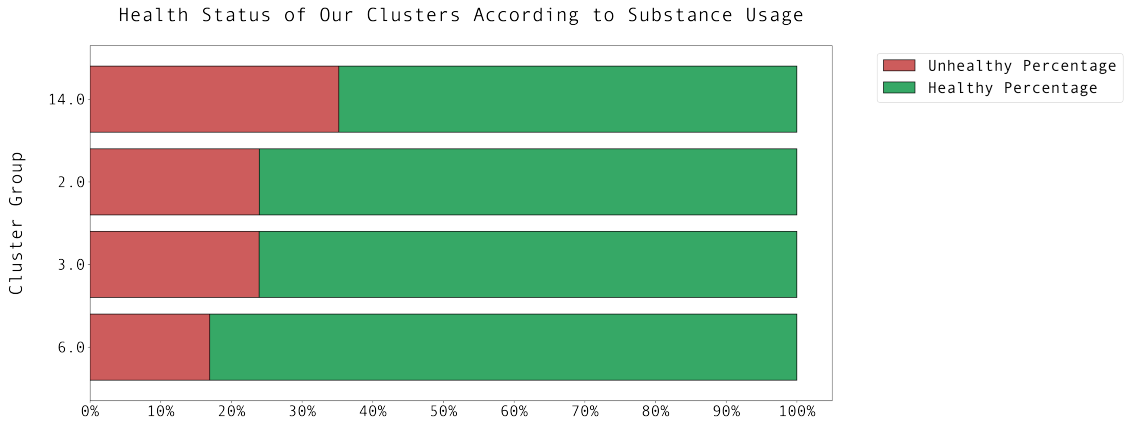

Figure 5.30: Proportion of Healthy vs. Unhealthy Surveyors In Each Cluster

Overall, it is evident that the majority of the clusters consist of people who don't smoke and don't exceed caffeine or alcohol limits. However, there are still a substantial amount of people in each cluster that have a substance issue. Most notably, we can see that 35% of cluster 14 have a problem. This is interesting because it's filled with single men in their 20s-30s who have a healthy BMI. These people rated their health as 'Fairly Good'. One can assume that this rating was strongly based on their BMI and that these people may not have taken into account their personal habits and substance abuse problems. On the other hand, we can see that cluster 6 has the smallest percentage of people with substance problems at 17%. This cluster consists of married women in their 50s-60s, who's BMI classifies them as Overweight. We can see that these people may have omitted their BMI when rating their health.

5.3.5 Which Problems Pose the Biggest Threat?

Now that we have discussed the proportion of unhealthy people in certain clusters, we can look more at which substances each cluster struggle with the most. This graph shows which substance has a dominant hold on each cluster's unhealthy population.

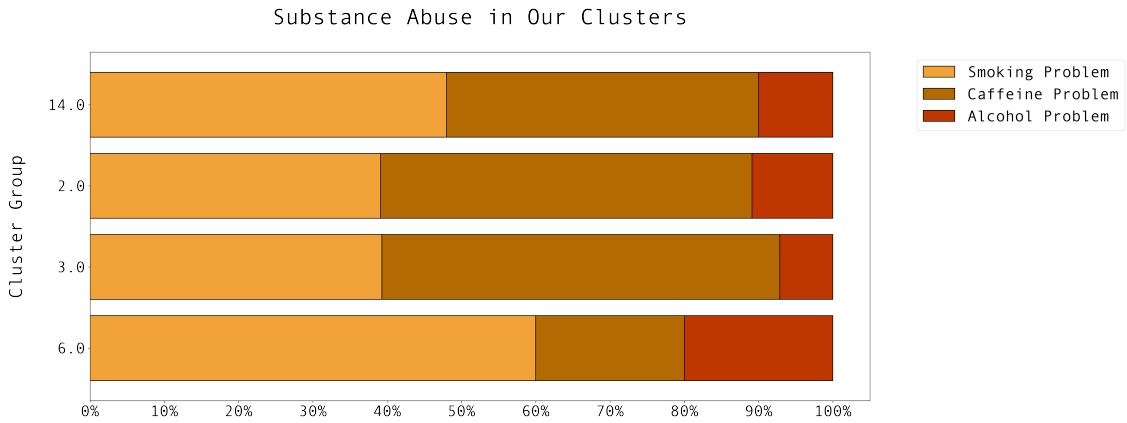

Figure 5.31: Abuse Distribution Among Unhealthy Cluster Populations

It is evident that no cluster shares a common pattern of substance abuse, which shows that different groups of individuals tend to suffer from different problems. In regards to the 4 main clusters we are analysing:

cluster 14 and cluster 6 struggle the most with tobacco usage, while cluster 3 and cluster 2 suffer the most from caffeine problems.

5.3.6 Summary

For this research question, we wanted to explore how Americans rate their overall health and sleep on an individual level. Our main goal was to see if people could recognise the effects of caffeine, tobacco and alcohol on their health and sleep, and if they rate them accordingly.

We saw that for each of the substances, there was a general trend - as sleep and health ratings worsen, the percentage of people who use excessive amounts of these substances grew. This showed us that people had a pretty good understanding of how caffeine, tobacco and alcohol can affect your health and sleep. That being said, there were some people who rated their health on the highest end of the scale, who used more substances like alcohol than their "healthy but not perfect" counterparts. This lead us to conclude that people are not always good judges of their own personal health and sleep.

We analysed specific clusters to see if there were specific groups of people that were worse judges of their health and sleep than others. We found that most of those who rate their sleep as 'Fairly Good' and health as 'Good', fell into the healthy category (not using any substances). That said, there were still a large portion of those people who did use substances. The worst culprits being single men in their 20-30s who have a healthy BMI. These men rated themselves as fairly good and good in their sleep and health ratings respectively, but 40% of them were guilty of using above the recommended substance limit.

We took this one step further to see which substance poses the biggest threat for which group. Single men in their 20-30s who have a healthy BMI and married women in their 50-60s who are overweight, struggle the most with tobacco usage, while married men in their 40-50s who are obese, and married women in their 30-40s who are overweight, suffer the most from caffeine problems.

Chapter 6: **Conclusions**

Sleep is incredibly important to us all and as a society, we are always trying to come up with ways to improve our sleeping habits.

This report describes the ways in which sleeping habits vary around the world, and what Americans can learn from other countries (and each other) about what qualifies as good sleep health. We take an in-depth look at how various substances can affect peoples health and sleep, and which groups of people are more susceptible to believing they have good health when underlying factors such as substance use and BMI are proving them wrong.

We used many different datasets 4.1 and the following key findings were observed based on our analysis.

- Sleep times vary around the world, with some countries sleeping on average, up to two hours more each night than others.
- As a whole, America's quality of sleep is good enough to earn itself a spot in the top 10 sleeping nations. However, there are certain aspects to America's time usage that could be improved, without sacrificing their sleep. Reducing time spent on paid work and study would allow Americans to allocate more time to personal care.
- There did not seem to be any clear trends in how inflation, unemployment and GDP affected sleeping patterns.
- Generally, Americans can recognize the effect of caffeine, tobacco and alcohol on their sleep and health.
- The majority of people in our clusters are considered healthy as they did not exceed recommended thresholds of substances, however, a small percentage of people in each cluster are deemed to be unhealthy, and thus, should take a closer look at their substance intake to see which substance poses the greatest threat to their health.

It is also worth noting some of the limitations that exist within this study.

Our OECD dataset only contained information on 33 countries. While we were able to look at data from many countries, there were many counties left out. This means that in part, we were not able to get a grasp of the full picture. That being said, we were still able to achieve useful analysis the goal things America could learn from the countries we studied.

Our Sleep Foundation survey data proved to be quite the challenge to work with at times. For our alcohol analysis, we had to work with only the ranges we were given in the survey. This meant that we were unable to determine whether or not some people were above or below the alcohol threshold, meaning we had to exclude this percentage from our alcohol analysis. If the survey had given us more specific values, the results for both the general alcohol analysis and the cluster based analysis might have been different.

These limitations should be taken into account when interpreting the results presented, especially when it comes to drawing more detailed conclusions.

With regards to prospects of more detailed analysis, there were many interesting possibilities in each of the research questions that we wanted to explore further, however due to the time limitations

of this course we were not able to complete them. For Research Question 1, "Looking Around the World - What Can Americans Learn About Sleep?, an incredible bit of analysis that could be done is finding more time use surveys from other countries. That way we could look at the world as a whole and understand where America ranks in terms of its time use and sleep patterns. For Research Question 2, "How Does Economic Health Affect People's Sleep?" we hope to gather more data to help us understand the intricacies of various factors of economic health as they relate to sleeping patterns. In this project we dealt with yearly figures, however we think that we could perform a more in-depth analysis on how the economy is functioning and how America is sleeping with monthly or daily figures. Finally research question 3, "Are People A Good Judge of Their Own Sleep and Health?", had so many avenues that we could have chosen from. In the future we hope to develop a more focused analysis on exercise, BMI, and napping. We really wanted to utilize the other Sleep Foundation surveys we cleaned, however we were cautious of our time constraints and focused on 2013's "Sleep and Health" Survey. We know that there are plenty of interesting observations in those surveys that are ready for analysis.

Acknowledgements

A huge appreciation is owed to William, our demonstrator, and Ciara for keeping us motivated and giving extremely insightful help and being a joy to work with. Specifically we owe gratitude in relation to help with correctly configuring the clusters in RQ3. William was also a huge help with setting up the guidelines for the story we wanted to portray onto which this reports "flow" is based off.

A big thank you to Barry Smith and his insightful boot camp which allowed us to perform more in-depth and comprehensive data science than we ever have before.

Finally, thank you to Fareed Idris for making sure our interactions with GitLab and Kreoh run smoothly.

Bibliography

1. Pacheco, D. P. *Why do we need sleep?* Mar. 2022. https://www.sleepfoundation.org/how-sleep-works/why-do-we-need-sleep.
2. Hood, E. & Jenkins, K. P. Evolutionary medicine: A powerful tool for improving human health. *Evolution: Education and Outreach* **1,** 114–120 (2008).
3. Rosenberg, D. C. *The History of Sleep Study: Blog* Nov. 2020. https://www.sleephealthsolutionsohio.com/blog/history-of-sleep-research/.
4. Cheng, V. *Nathaniel Kleitman, sleep expert, dies at 104* Aug. 1999. https://www.nytimes.com/1999/08/19/us/nathaniel-kleitman-sleep-expert-dies-at-104.html.
5. Nutt, D., Wilson, S. & Paterson, L. Sleep disorders as core symptoms of depression. *Dialogues in clinical neuroscience* (2022).
6. Zochil, M. L. & Thorsteinsson, E. B. Exploring poor sleep, mental health, and help-seeking intention in university students. *Australian journal of psychology* **70,** 41–47 (2018).
7. Oecd. https://stats.oecd.org/index.aspx?datasetcode=time_use.
8. *BLS Data viewer* https://beta.bls.gov/dataViewer/view/timeseries/TUU10101AA01000247;jsessionid=F9556C98F34C600F35C8B2B976FAE57D.
9. *2013 Sleep in America® Poll – Exercise and Sleep* https://www.sleephealthjournal.org/content/sleepinamerica.
10. *U.S. GDP growth rate 1961-2022* https://www.macrotrends.net/countries/USA/united-states/gdp-growth-rate.
11. *U.S. unemployment rate 1991-2022* https://www.macrotrends.net/countries/USA/united-states/unemployment-rate.
12. *U.S. inflation rate 1960-2022* https://www.macrotrends.net/countries/USA/united-states/inflation-rate-cpi.
13. hortonj4. *How much caffeine is too much?* Jan. 2022. https://health.clevelandclinic.org/how-much-caffeine-is-too-much/#:~:text=Here%20are%20the%20boundaries.,about%20two%20cans%20of%20cola)..
14. *Tobacco use* Mar. 2022. https://www.cdc.gov/chronicdisease/resources/publications/factsheets/tobacco.htm.
15. *US DRINKING NORMS UPDATED* Oct. 2019. https://arg.org/news/drinking-norms-in-the-us/#:~:text=The%20NIAAA%20weekly%20volume%20guideline,in%20any%20day%20for%20women.).
16. Kelso, N. V. *Natural-earth-vector/ne*$_1 10 m_a dmin_0 countries. SHPatmaster NVKELSO/natural-earth-vector$ May 2018. https://github.com/nvkelso/natural-earth-vector/blob/master/110m_cultural/ne_110m_admin_0_countries.shp.
17. Rousseau, B. *Napping in public? in Japan, that's a sign of diligence* Dec. 2016. https://www.nytimes.com/2016/12/16/world/what-in-the-world/japan-inemuri-public-sleeping.html.
18. Saiidi, U. *Japan has some of the longest working hours in the world. it's trying to change* June 2018. https://www.cnbc.com/2018/06/01/japan-has-some-of-the-longest-working-hours-in-the-world-its-trying-to-change.html.
19. *Work culture in China new 2022: Tips to manage employees* Mar. 2022. https://www.hrone.com/blog/work-culture-china/.

20. Yip, W. *China steps in to regulate brutal '996' work culture* Sept. 2021. https://www.bbc.com/news/world-asia-china-58381538.
21. Pettinger, T., Ray, Yvone & Hicken, J. *Impact of economic recession* Mar. 2020. https://www.economicshelp.org/blog/5618/economics/negative-impact-of-economic-recession/.
22. Team, T. I. *Were there any periods of major deflation in U.S. history?* Feb. 2022. https://www.investopedia.com/ask/answers/040715/were-there-any-periods-major-deflation-us-history.asp.
23. Muenster, E., Rueger, H., Ochsmann, E., Letzel, S. & Toschke, A. M. Association between overweight, obesity and self-perceived job insecurity in German employees. *BMC public health* **11,** 1–7 (2011).
24. Kivimäki, M., Singh-Manoux, A., Nyberg, S., Jokela, M. & Virtanen, M. Job strain and risk of obesity: systematic review and meta-analysis of cohort studies. *International journal of obesity* **39,** 1597–1600 (2015).
25. Schnall, P. L., Landsbergis, P. A. & Baker, D. Job strain and cardiovascular disease. *Annual review of public health* **15,** 381–411 (1994).
26. House, J. S. Occupational stress and coronary heart disease: A review and theoretical integration. *Journal of health and social behavior,* 12–27 (1974).
27. Iacovides, A., Fountoulakis, K. N., Kaprinis, S. & Kaprinis, G. The relationship between job stress, burnout and clinical depression. *Journal of affective disorders* **75,** 209–221 (2003).
28. Mandal, B., Ayyagari, P. & Gallo, W. T. Job loss and depression: The role of subjective expectations. *Social Science & Medicine* **72,** 576–583 (2011).
29. Riumallo-Herl, C., Basu, S., Stuckler, D., Courtin, E. & Avendano, M. Job loss, wealth and depression during the Great Recession in the USA and Europe. *International journal of epidemiology* **43,** 1508–1517 (2014).
30. Staley, O. *Economists have finally measured the connection between unemployment and sleep loss* https://qz.com/1904346/the-connection-between-unemployment-and-sleep/.
31. Fernando, J. *Gross domestic product (GDP) definition* Feb. 2022. https://www.investopedia.com/terms/g/gdp.asp.
32. Hafner, M., Stepanek, M., Taylor, J., Troxel, W. M. & Van Stolk, C. Why sleep matters—the economic costs of insufficient sleep: a cross-country comparative analysis. *Rand health quarterly* **6** (2017).
33. *KModes clustering algorithm for categorical data* June 2021. https://www.analyticsvidhya.com/blog/2021/06/kmodes-clustering-algorithm-for-categorical-data/#:~:text=KModes%20clustering%20is%20one%20of,similar%20our%20data%20points%20are..
34. Jaiswal, S. *K-modesclustering* July 2020. https://medium.com/@shailja.nitp2013/k-modesclustering-ef6d9ef06449.
35. Ashydv. *Bank customer clustering (K-modes clustering)* Mar. 2019. https://www.kaggle.com/code/ashydv/bank-customer-clustering-k-modes-clustering/notebook.
36. teejmahal20. *Clustering categorical data: K-modes (CAT II)* Jan. 2020. https://www.kaggle.com/code/teejmahal20/clustering-categorical-data-k-modes-cat-ii/notebook.