

Week 13 Diary Entry

Yuha Kim

2023-11-24

Week 9 Diary Entry

Part 1

Q: What is the topic that you have finalized?

A: The topic I have finalized for is about the volcanoes in Japan. The question I want to investigate is: How are volcanoes distributed across Japan?

Part 2

Q: What are the data sources that you have curated so far?

A: The data source I have curated so far is a visualization of volcanoes in Japan from Tableau Public. Here is the link: https://public.tableau.com/app/profile/david.pires/viz/volcanoes_15907677233490/LandofVolcanoes (https://public.tableau.com/app/profile/david.pires/viz/volcanoes_15907677233490/LandofVolcanoes)

Week 10 Diary Entry

(I changed topic from Week 9!)

Part 1

Q: What is the question that you are going to answer?

A: How does the amount of physical activity affect the quality of sleep for people in different occupations?

Part 2

Q: Why is this an important question?

A: According to the National Institutes of Health, sleep plays a vital role in maintaining physical and mental health. It is widely known that physical activity improves sleep, as proved by numerous academic studies over the years. The specific question I am asking is important because some jobs demand a high degree of stress that may negatively affect sleep despite the high level of physical activity.

Part 3

Q: Which rows and columns of the dataset will be used to answer this question?

A: I will use the columns "Occupation", "Quality of Sleep", "Physical Activity" and all of 374 rows.

Additionally, I may use the columns "Sleep Duration" and "Sleep Disorder" when analyzing the quality of sleep factor.

Challenges and errors I faced

I faced a challenge when I was reviewing the data source for my previous topic on volcanoes in Japan. The csv file had a good chunk of null values for the columns I needed to use. Therefore, I searched for a different data set with a better usability. This new data set also helped me formulate a more interesting research question because it displays the occupation of each subject.

```
library(tidyverse)
```

```
## — Attaching core tidyverse packages — tidyverse 2.0.0 —
## ✓ dplyr      1.1.3      ✓ readr      2.1.4
## ✓ forcats    1.0.0      ✓ stringr    1.5.0
## ✓ ggplot2    3.4.4      ✓ tibble     3.2.1
## ✓ lubridate  1.9.3      ✓ tidyr      1.3.0
## ✓ purrr      1.0.2
## — Conflicts — tidyverse_conflicts() —
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
read_csv("sleep.csv")
```

```
## Rows: 374 Columns: 13
## — Column specification —
## Delimiter: ","
## chr (5): Gender, Occupation, BMI Category, Blood Pressure, Sleep Disorder
## dbl (8): Person ID, Age, Sleep Duration, Quality of Sleep, Physical Activity...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
## # A tibble: 374 × 13
##   `Person ID` Gender Age Occupation `Sleep Duration` `Quality of Sleep`
##   <dbl> <chr> <dbl> <chr> <dbl> <dbl>
## 1 1 Male 27 Software Engine... 6.1 6
## 2 2 Male 28 Doctor 6.2 6
## 3 3 Male 28 Doctor 6.2 6
## 4 4 Male 28 Sales Represent... 5.9 4
## 5 5 Male 28 Sales Represent... 5.9 4
## 6 6 Male 28 Software Engine... 5.9 4
## 7 7 Male 29 Teacher 6.3 6
## 8 8 Male 29 Doctor 7.8 7
## 9 9 Male 29 Doctor 7.8 7
## 10 10 Male 29 Doctor 7.8 7
## # i 364 more rows
## # i 7 more variables: `Physical Activity Level` <dbl>, `Stress Level` <dbl>,
## # `BMI Category` <chr>, `Blood Pressure` <chr>, `Heart Rate` <dbl>,
## # `Daily Steps` <dbl>, `Sleep Disorder` <chr>
```

Week 11 Diary Entry

Part 1

Q: List the visualizations that you are going to use in your project (Answer: What are the variables that you are going to plot? How will it answer your larger question?)

A: I will plot the variable “Physical Activity” in the x-axis and the variable “Sleep Duration” in the y-axis. I will map occupation to the colour of each point in order to compare how the relationship between physical activity and quality of sleep differs by jobs.

Part 2

Q: How do you plan to make it interactive? (Answer: features of ggplot2/shiny/markdown do you plan to use to make the story interactive)

I plan to make it interactive by using the library(ggplotly) command to display the content of each point in the graph when you hover a cursor above it.

Part 3

Q: What concepts incorporated in your project were taught in the course and which ones were self-learnt?

A:

```
data <- data.frame(
  Topics = c("Install packages", "Invoke the library", "Read csv file", "ggplot2"),
  Weeks = c(2, 2, 3, 7),
  Taught = c("Yes", "Yes", "Yes", "Yes")
)

print(data)
```

```
##           Topics Weeks Taught
## 1  Install packages      2    Yes
## 2 Invoke the library      2    Yes
## 3   Read csv file        3    Yes
## 4         ggplot2        7    Yes
```

Challenges and errors I faced

I learned that I had to use back ticks when creating the ggplot for this particular data set. Initially, I put in the labels without quotations, and it only showed a graph with just one dot in the middle. I was very confused, but when I looked back on the result of glimpse command, I saw that the names of the columns I wanted to use were enclosed in back ticks. I tried again, and it rendered successfully.

Week 12 Diary Entry

Challenges and errors I faced

I struggled with getting rid of the code chunk displayed on the Visualizations Page. I tried setting echo=FALSE and asked for help from TAs, but I still could not find a way to get rid of it. Another challenge I had was in regards to rendering graphs. The visualization did not show up on the actual website, and I learned that it was because I did not remove the code chunk "install.packages" in the qmd file. After deleting it, the graphs rendered successfully on the website.

Week 13 Diary Entry

1. What is the theme of your data story?

The theme of my data story is the relationship between sleep and physical activity. The question that I am investigating is: Is high level of physical activity correlated to better sleep for people in different occupations?

2. Why is it important to address this question?

It is important to address this question because different jobs require different levels of stress. Stress is a critical factor that may hinder getting a good sleep even though a person gets a lot of physical activity. For instance, some jobs are physically demanding, leading to a high score on the level of physical activity, but the stress can have a negative effect on sleep, resulting in less hours of sleep or a bad quality of sleep. Therefore,

it is critical to examine how the relationship between sleep and physical activity differs by occupation so that we can emerge with a more refined understanding of the correlation and hence make better predictions in the future.

3. Why do you think the data sources that you have curated can help you answer the question?

The data source that I have curated is pivotal in helping me answer my question because it provides information on key variables I need for my investigation. These variables include 'Physical Activity Level', 'Sleep Duration', 'Quality of Sleep', and 'Occupation'. The Physical Activity Level refers to the number of minutes the person engages in physical activity daily. Sleep Duration is the number of hours the person sleeps per day. The Quality of Sleep is the subjective rating of the quality of sleep, ranging from 1 to 10. The Occupation refers to the occupation of the person. I first plotted the Physical Activity Level as the x-variable and Sleep Duration and Quality of Sleep as the y-variables. I then assigned a different color for each of the eleven occupations, hence allowing for the comparison of the relationship between x-variable and y-variable for different jobs.

4. What are the insights from the data and how are they depicted in plots?

Visualization #1 depicts the correlation between physical activity levels and sleep duration, revealing a general upward trend. However, certain occupations, like engineers and nurses, deviate from this pattern. Engineers with low physical activity often sleep over 8 hours, while highly active nurses get only about 6 hours, potentially due to the physical demands of their job and increased stress affecting sleep. Visualization #2 introduces sleep quality as a variable, aiming to understand how it differs from sleep duration. Across the 11 graphs for each occupation, a clear trend is challenging to identify. Software engineers and nurses exhibit a positive link between physical activity and sleep quality, while engineers and lawyers lack a discernible trend. This complexity challenges the predictability of sleep quality based on physical activity for engineers and lawyers. Despite nurses getting less sleep, Visualization #2 reveals a positive correlation between their sleep quality and physical activity, highlighting the nuanced nature of the relationship between physical activity, sleep duration, and sleep quality across different occupations.

5. How did you implement this entire project? Were there any new concepts that you learnt to implement some aspects of it?

I implemented the entire project through an iterative process. Initially, I had planned to make only one graph plotting physical activity level and quality of sleep. When I made the graph, I realized that there was no discernable trend, making it difficult to garner any meaningful insights from the visualization. Thus, I substituted sleep duration as the y-variable, and it rendered a more comprehensive graph. Yet I still wanted to include the quality of sleep in my discussion of the topic because quality of sleep is a different measurement than sleep duration. Therefore, I decided to add another visualization that shows each of the 11 occupations in a separate graph.

References

Tharmalingam, L. (2023a, September 18). Sleep health and lifestyle dataset. Kaggle.
<https://www.kaggle.com/datasets/uom190346a/sleep-health-and-lifestyle-dataset>
(<https://www.kaggle.com/datasets/uom190346a/sleep-health-and-lifestyle-dataset>)