

Week 10 Diary Entry

Yuha Kim

2023-10-23

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    3.4.3    ✓ tibble     3.2.1
## ✓ lubridate  1.9.2    ✓ 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 1 Male 27 Software Engine... 6.1 6
## 2 2 2 Male 28 Doctor 6.2 6
## 3 3 3 Male 28 Doctor 6.2 6
## 4 4 4 Male 28 Sales Represent... 5.9 4
## 5 5 5 Male 28 Sales Represent... 5.9 4
## 6 6 6 Male 28 Software Engine... 5.9 4
## 7 7 7 Male 29 Teacher 6.3 6
## 8 8 8 Male 29 Doctor 7.8 7
## 9 9 9 Male 29 Doctor 7.8 7
## 10 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>
```