Week 11 Diary Entry

Yuha Kim

2023-10-30

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 Tablaeu 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 chuck 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 conflic
ts 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)

```
library(tidyverse)
```

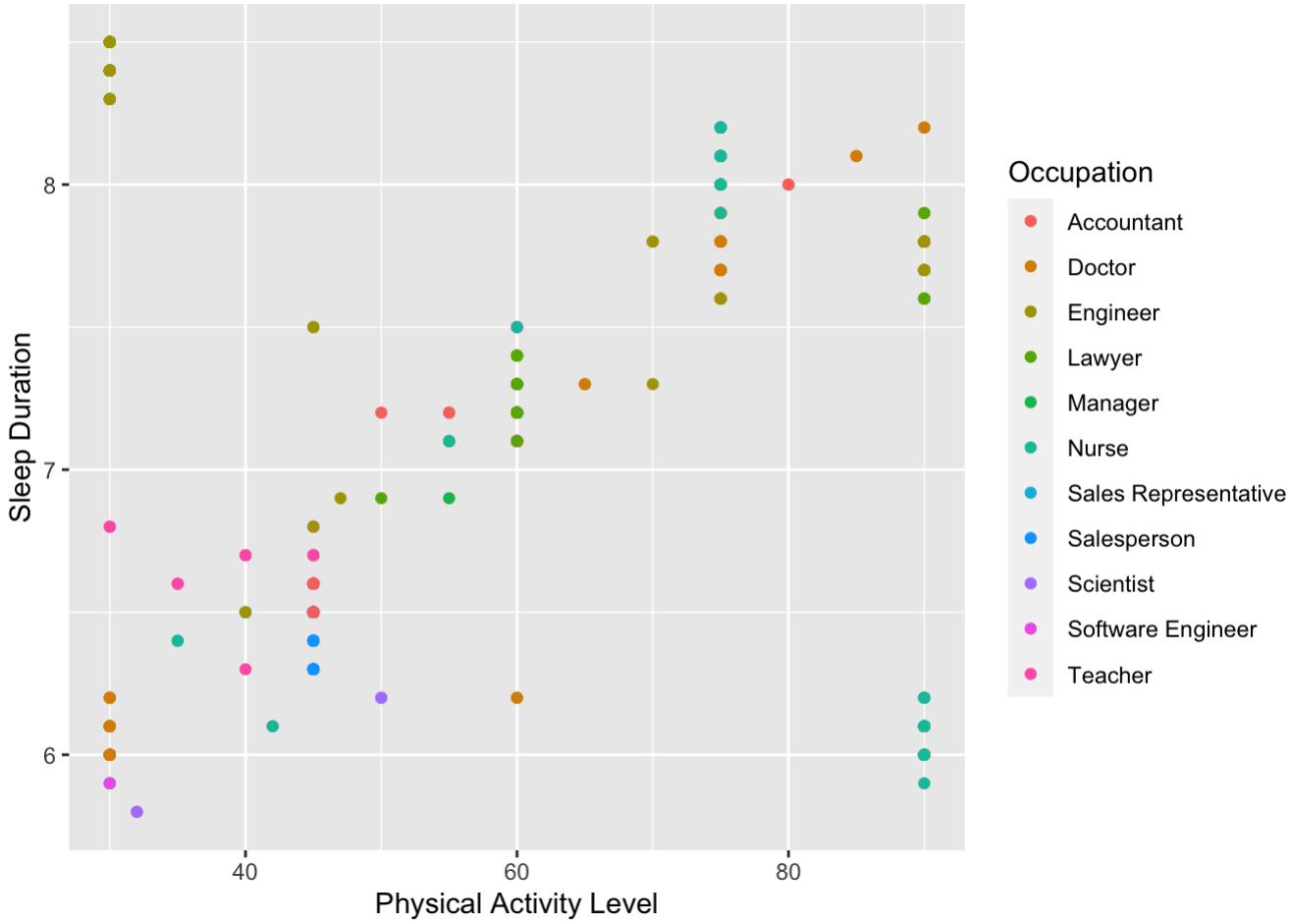
```
sleep <- read_csv("sleep.csv")
```

```
## 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.
```

```
glimpse(sleep)
```

```
## Rows: 374
## Columns: 13
## $ `Person ID`               <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 1…
## $ Gender                    <chr> "Male", "Male", "Male", "Male", "Male", "Mal…
## $ Age                       <dbl> 27, 28, 28, 28, 28, 28, 29, 29, 29, 29, 29, …
## $ Occupation                <chr> "Software Engineer", "Doctor", "Doctor", "Sa…
## $ `Sleep Duration`          <dbl> 6.1, 6.2, 6.2, 5.9, 5.9, 5.9, 6.3, 7.8, 7.8,…
## $ `Quality of Sleep`        <dbl> 6, 6, 6, 4, 4, 4, 6, 7, 7, 7, 6, 7, 6, 6, 6,…
## $ `Physical Activity Level` <dbl> 42, 60, 60, 30, 30, 30, 40, 75, 75, 75, 30, …
## $ `Stress Level`            <dbl> 6, 8, 8, 8, 8, 8, 7, 6, 6, 6, 8, 6, 8, 8, 8,…
## $ `BMI Category`            <chr> "Overweight", "Normal", "Normal", "Obese", "…
## $ `Blood Pressure`          <chr> "126/83", "125/80", "125/80", "140/90", "140…
## $ `Heart Rate`              <dbl> 77, 75, 75, 85, 85, 85, 82, 70, 70, 70, 70, …
## $ `Daily Steps`             <dbl> 4200, 10000, 10000, 3000, 3000, 3000, 3500, …
## $ `Sleep Disorder`          <chr> "None", "None", "None", "Sleep Apnea", "Slee…
```

```
ggplot(data = sleep,
       mapping = aes(x = `Physical Activity Level`, y = `Sleep Duration`, colour = Oc
cupation)) + geom_point()
```



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.