HW #1

1.2, 1.6, 1.11, 1.12, 1.16, 1.18, 1.26, 1.27

Hayden Atchley

5 September 2022

1.2

The sex discrimination study is an observational study rather than a randomized study. Because of this, it is impossible to know what other factors may have contributed to the results other than the sex of the employees. Even if the researchers were to control for variables such as education, experience, age, and/or any other variables they could think of, there may still be confounding variables that aren't known and can't be accounted for. It is therefore difficult to prove sex discrimination from this study alone.

1.6

This study is an observational study, as it wasn't randomized which subjects used marijuana and which didn't. As such, it is difficult to conclude that marijuana use affects short-term memory because any number of other factors might be at play. The selection of the subjects was also not random, and so the results can't be generalized to a larger group.

Many other factors could potentially be confounding variables in this study. These could include education, genetics, factors from how they were raised, or potentially even indirect factors such as household income.

1.11

While at first glance it seems that the Vitamin C should have the same effect whether or not the participants know they're taking it, much research has shown that the "placebo effect", where the mere belief that a treatment will have a certain effect can cause that effect, is significant. The extra confidence of the participants who knew they weren't in the control group may have contributed to the results. Additionally, those participants may have (consciously or not) modified their behaviors in a way that lowered their chances of getting a cold. Therefore, it is no longer possible to conclude that the difference in cold rates is due solely to the Vitamin C.

1.12

The results, though they can't be generalized to a population due to the non-random sample, are still important because the treatment was randomized. The results therefore indicate that the treatment did have a significant effect on the subjects, and they indicate a need for future studies to determine if this effect can be generalized for a larger population.

1.16

I obtained the data for this problem from [https://www.cia.gov/the-world-factbook/about/archives/download/factbook-2004.zip] in `./rankorder/2004rank.txt`. This data does not seem to match Display 1.11, but it appears to be the data the exercise is pointing to. I am not sure what other data to use, so I will use this.

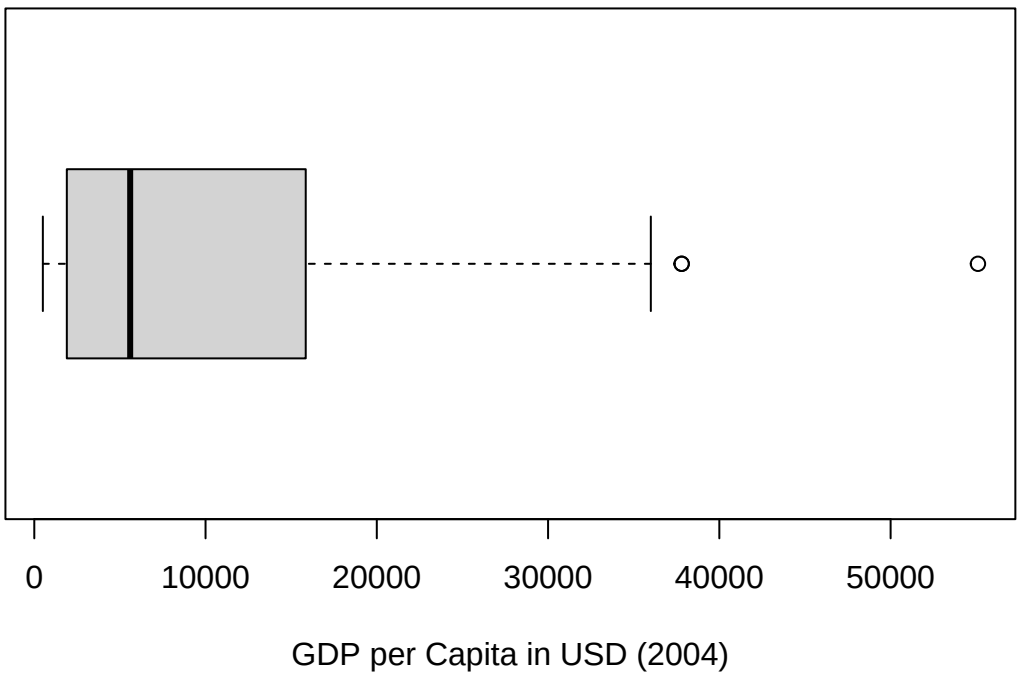
```
rank <- read_tsv("HW/HW1/2004rank.txt", skip = 1, n_max = 232) %>%
    mutate(`GDP - per capita` = `GDP - per capita` %>%
                     str_replace_all("[^\\d]", "") %>%
                     as.integer())

head(rank, n = 4)
```

```
## # A tibble: 4 x 4
##    Rank Country       `GDP - per capita` `Date of Information`
##   <dbl> <chr>                      <int> <chr>
## 1     1 Luxembourg                 55100 2003 est.
## 2     2 Norway                     37800 2003 est.
## 3     3 United States              37800 2003 est.
## 4     4 Bermuda                    36000 2003 est.
```

A boxplot of the data is below.

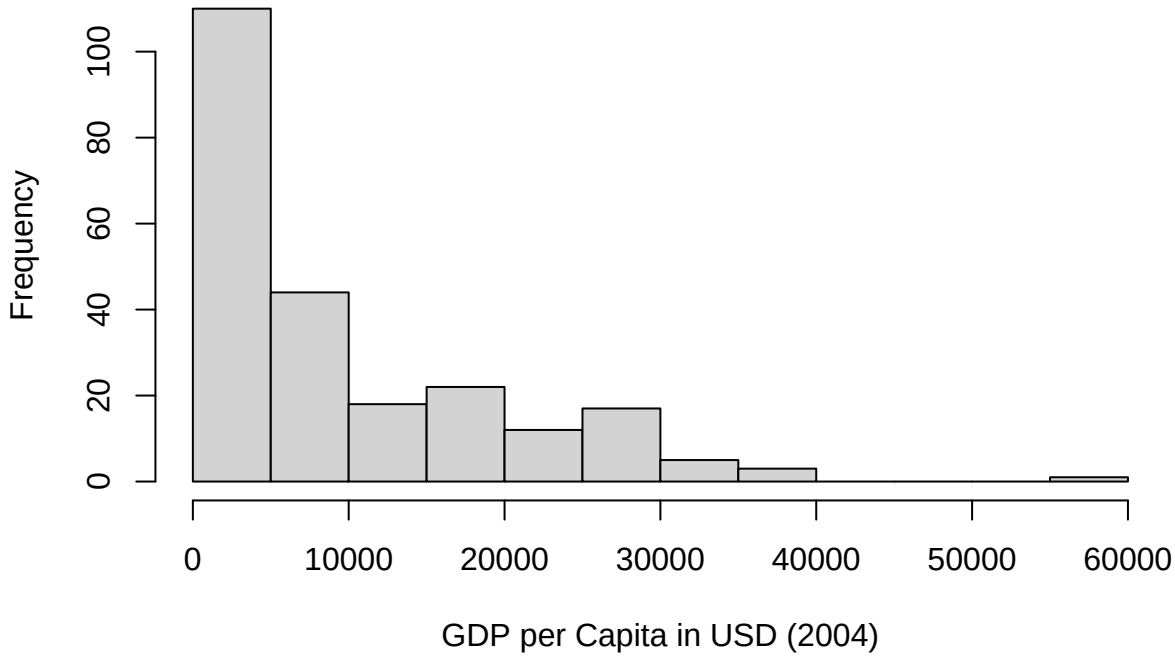
```
rank$`GDP - per capita` %>%
    boxplot(xlab = "GDP per Capita in USD (2004)",
                     horizontal = TRUE)
```



This plot differs from Display 1.11 in several ways. Firstly, as previously mentioned, the data themselves seem to be different. Other than that, there are several presentational differences: Display 1.11 is vertical while this plot is horizontal, the Display names extreme outliers and labels various points of interest in the plot whereas this one does not, and the scale differs by a factor of 1000.

A histogram of the data is below. The bin width defaulted to 5000, and the breaks are listed below the plot.

```
rank$`GDP - per capita` %>%
    hist(xlab = "GDP per Capita in USD (2004)",
             main = NULL) %>%
    {.$breaks}
```




```
##  [1]     0  5000 10000 15000 20000 25000 30000 35000 40000 45000 50000 55000
## [13] 60000
```

The breaks already defaulted to $5000, so no redraw is needed.