```
# Homework 3
install.packages("data.table")
install.packages("arules")
install.packages("arulesViz")
install.packages("lubridate")
install.packages("ggplot2")
install.packages("knitr")
install.packages("plyr")
install.packages("readxl")
install.packages("tidyverse")
install.packages("RColorBrewer")

library(data.table)
library(arules)
library(arulesViz)
library(lubridate)
library(ggplot2)
library(knitr)
library(plyr)
library(readxl)
library(tidyverse)
library(RColorBrewer)

# 1
# A : Importing the Coronavirus dataset

getwd()
setwd("/Users/scottziegler/Desktop")

install.packages("coronavirus")
library(coronavirus)

data(coronavirus)

# B: First 100 rows
head(coronavirus, n = 100)

# Columns
# The first column is the date of the Covid data (when the case occurred), the province is similar to what 'state' in the
# country it occurred. The country shows what country the Covid case occurred, the lat and long is latitude and longitude,
# for a more precise location of the disease. The type refers to whether or not the individual has the disease, died, or
```

```
# recovered. Lastly, the cases column shows the number of cases in that area at that time.

# 2
# A: Showing top 20 countries by cases

coronavirus$country
countries = sort(table(coronavirus$country), decreasing = TRUE, "confirmed" = TRUE)
head(countries, n = 20)

# B: Bar plot of top 5 countries

top_countries = head(countries, n = 5)
barplot(top_countries)

# C: Flip the bar to be a horizontal plot

barplot(top_countries, horiz = TRUE)

# D: Adding a title

barplot(top_countries, horiz = TRUE, main = "Top 5 Countries by Total Cases")
barplot(top_countries, horiz = TRUE, main = "Top 5 Countries by Total Cases", xlab = "Number
of cases", ylab = "Country")

# 3 Recent Cases
# A: Creating the data frame

library(tidyr)

recent_cases = coronavirus %>%
  group_by(type, date) %>%
  summarise(total_cases = sum(cases)) %>%
  pivot_wider(names_from = type, values_from = total_cases) %>%
  arrange(date) %>%
  mutate(active = confirmed - death - recovered) %>%
  mutate(active_total = cumsum(active),
         recovered_total = cumsum(recovered),
         death_total = cumsum(death))

# B: Creating the plot

ggplot(recent_cases, aes(date, confirmed)) + geom_point()
```

Extra credit

#1. Making a blue line on top of the scatter plot

```
ggplot(recent_cases, aes(date, confirmed)) + geom_point() + geom_line( color = "blue")
```

#2. Adding a title to the graph

```
ggplot(recent_cases, aes(date, confirmed)) + geom_point() + geom_line( color = "blue") +
ggtitle("Confirmed Worldwide Cases by Date")
```

#3. Changing the fonts

```
ggplot(recent_cases, aes(date, confirmed)) + geom_point() + geom_line( color = "blue") +
  ggtitle("Confirmed Worldwide Cases by Date") +
  theme(text = element_text(size = 14, family = "Comic Sans MS"))
```

#4. Changing the font color to Red

```
ggplot(recent_cases, aes(date, confirmed)) + geom_point() + geom_line( color = "blue") +
  ggtitle("Confirmed Worldwide Cases by Date") +
  theme(text = element_text(size = 14, family = "Comic Sans MS", color = "red"))
```