DATA300_LAB4_HanhPhan

Part 1: Tidying the Data

1. Subset the data by only keeping the year 2018's data
2. Convert the variable "trade" to numeric, and change the missing values to NA
3. Change the negative "democ" values to NA

```
mydata <- read.csv("world_bank.csv")

df <- subset(mydata,mydata$Time=="2018")
df$trade <- as.numeric(df$trade)

## Warning: NAs introduced by coercion

df$trade <- ifelse(df$trade==".." ,NA,df$trade)
df$democ <- ifelse(df$democ<0,NA,df$democ)
```

Part 2: Trade and GDP

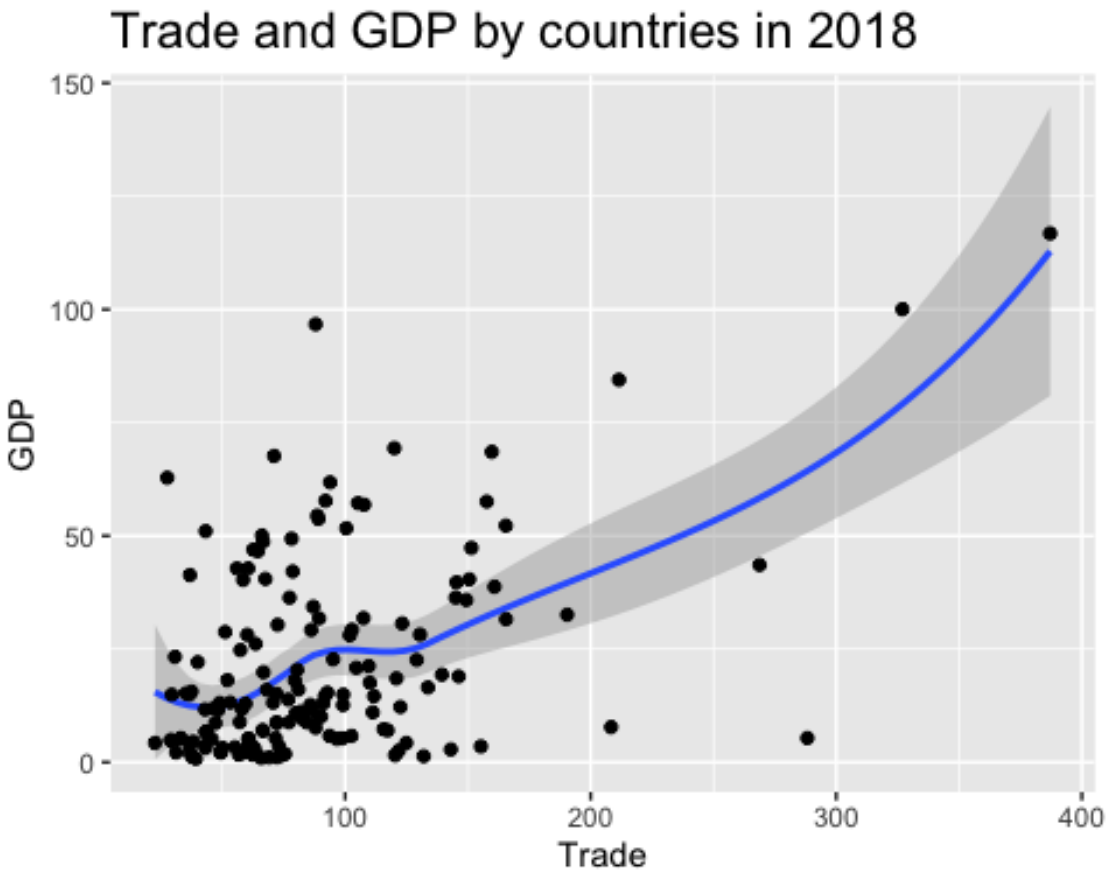
The trend line is going up as trade and GDP get bigger. Therefore, we can conclude that countries that trade more are likely to have higher GDP.

```
library(ggplot2)
G <- ggplot(data=df,mapping=aes(x=trade,y=gdp)) + geom_smooth() +
geom_point() + labs(title = "Trade and GDP by countries in 2018", x="Trade",
y = "GDP") + theme(plot.title = element_text(size=16))
G

## `geom_smooth()` using method = 'loess' and formula 'y ~ x'

## Warning: Removed 21 rows containing non-finite values (stat_smooth).

## Warning: Removed 21 rows containing missing values (geom_point).
```



Part 3:

The smooth loess curve and the straight-line linear fit both show the trend and relationship between Democracy and GDP of countries. The smooth loess curve fits the data better because the data point does not appear to be following a really straight line. For example, many non-democratic countries have high GDP and many democratic countries have low GDP.

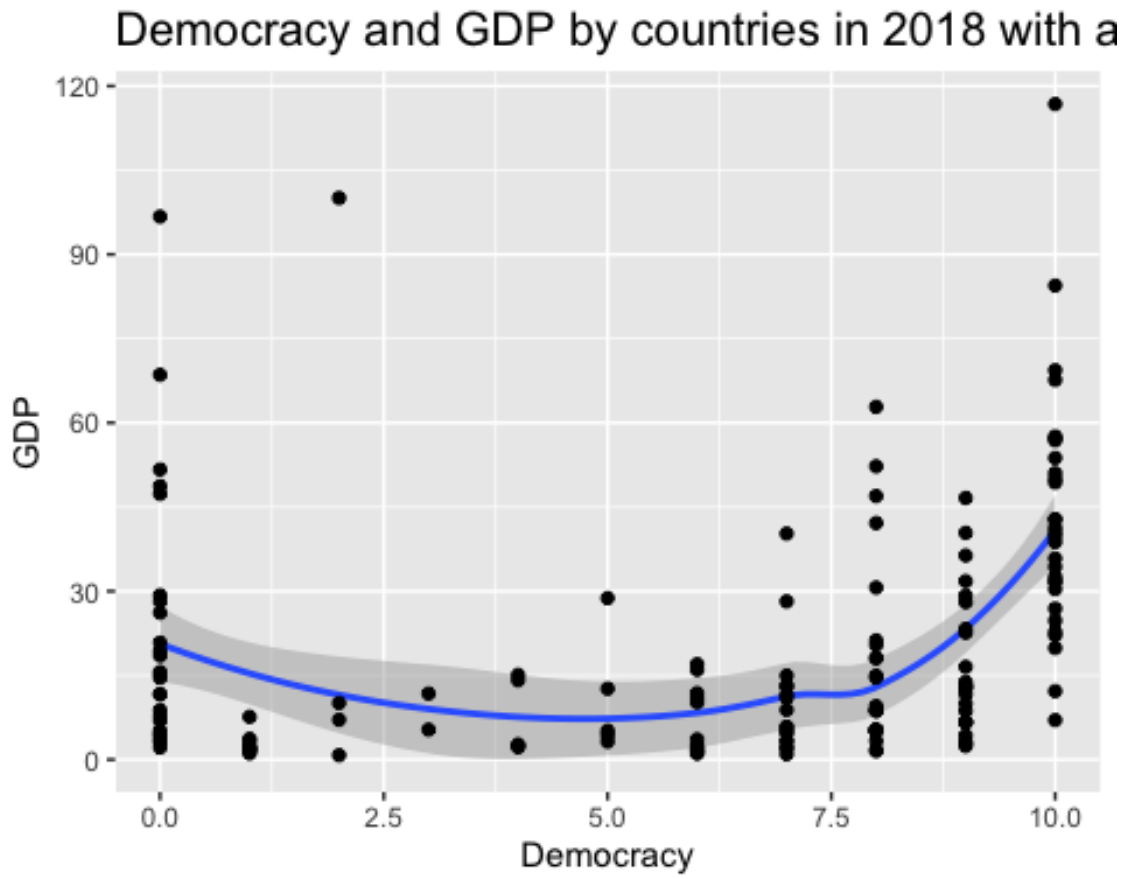
There are more countries that are democratic than non-democratic. Also, there are many democratic countries with low GDP per capita and many non-democratic one with high GDP capita. Therefore, democratic countries are not likely to have higher GDP.

```
G2 <- ggplot(data=df,mapping=aes(x=democ,y=gdp)) + geom_smooth()
+geom_point()+ labs(title = "Democracy and GDP by countries in 2018 with a
smooth loess curve", x="Democracy", y = "GDP")+ theme(plot.title =
element_text(size=15))
G2

## `geom_smooth()` using method = 'loess' and formula 'y ~ x'

## Warning: Removed 28 rows containing non-finite values (stat_smooth).

## Warning: Removed 28 rows containing missing values (geom_point).
```

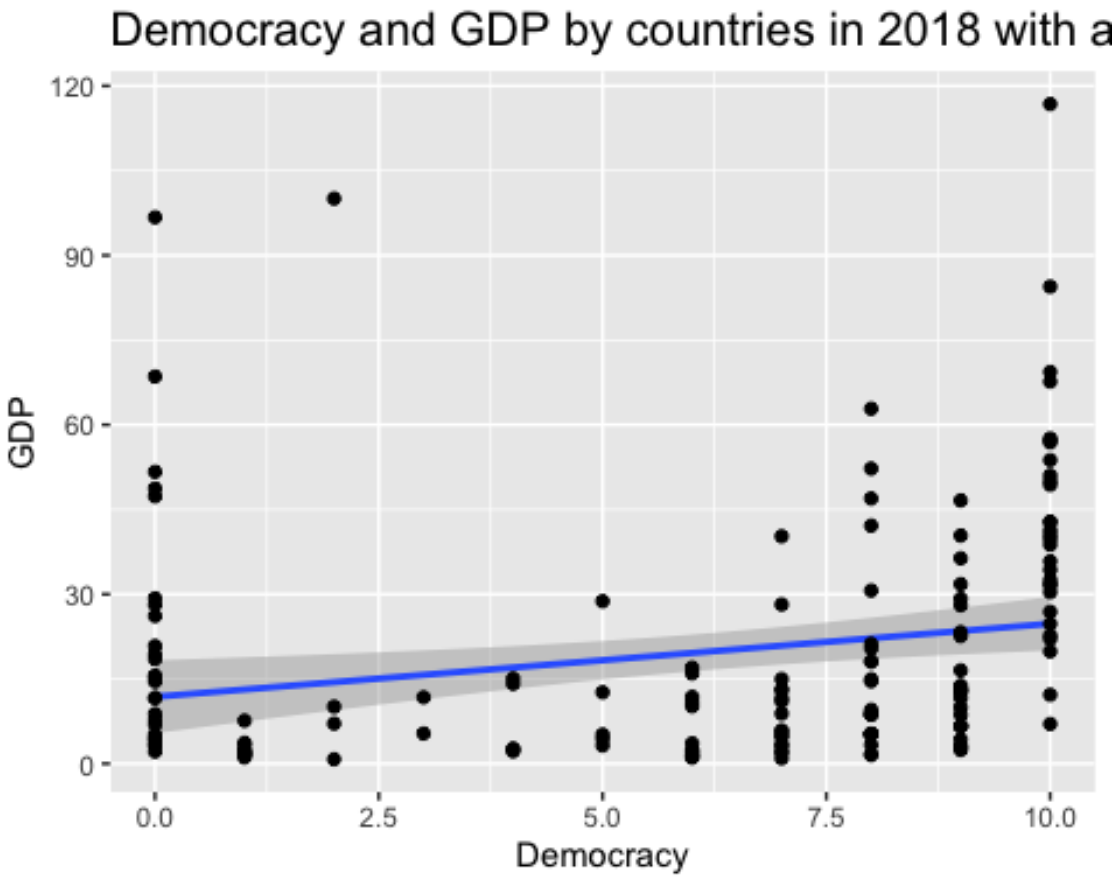



```
G3 <- ggplot(data=df,mapping=aes(x=democ,y=gdp)) + geom_smooth(method='lm')
+geom_point()+ labs(title = "Democracy and GDP by countries in 2018 with a
straight-line linear fit", x="Democracy", y = "GDP") + theme(plot.title =
element_text(size=15))
G3
```

```
## `geom_smooth()` using formula 'y ~ x'

## Warning: Removed 28 rows containing non-finite values (stat_smooth).

## Warning: Removed 28 rows containing missing values (geom_point).
```


Democracy and GDP by countries in 2018 with a
120
90
60
30
0
GDP
0.0
2.5
5.0
7.5
10.0
Democracy