DATA 606 Project Proposal

Liam Byrne

October 16, 2016

Loading data. Data dictionary available here

```
file_link <-
  "https://raw.githubusercontent.com/Liam-O/Project/master/worldBankProfile.csv"

# Returns a data table
worldbank  <- fread(file_link, header = TRUE, na.strings = c("", ".."), data.table = TRUE, nrows = 1386

worldbank <-worldbank %>%
    gather(year, figure, 6:8) %>%
    select(c(1,3, 6, 7)) %>%
    spread('Series Name', figure)

#Will clean headers later
```

Research question

The project will look at GDP growth for all respective countries provided by The World Country Profiles dataset and possible mitigating factors leading to negative or postie growth

Cases

Every existing country is a case. There are 232 in the set.

Data collection

The data was gathered from the The World Country Profiles dataset

Type of study

The data in `worldbank` deals with historical data for 1990, 2000 and 2015. The study will, thus, be observation.

Data Source

The data was gathered from the The World Country Profiles dataset

Response

The response variable will be GDP in a $US conversion, so numerical.

Explanatory

The explanatory variable has not been decided upon yet. There are 59 possible explanatory variables in the data set. Some could include:

- Credit provided by financial sector
- Fresh water availability
- Urban growth
- Purchasing power
- School enrollment

Relevant summary statistics

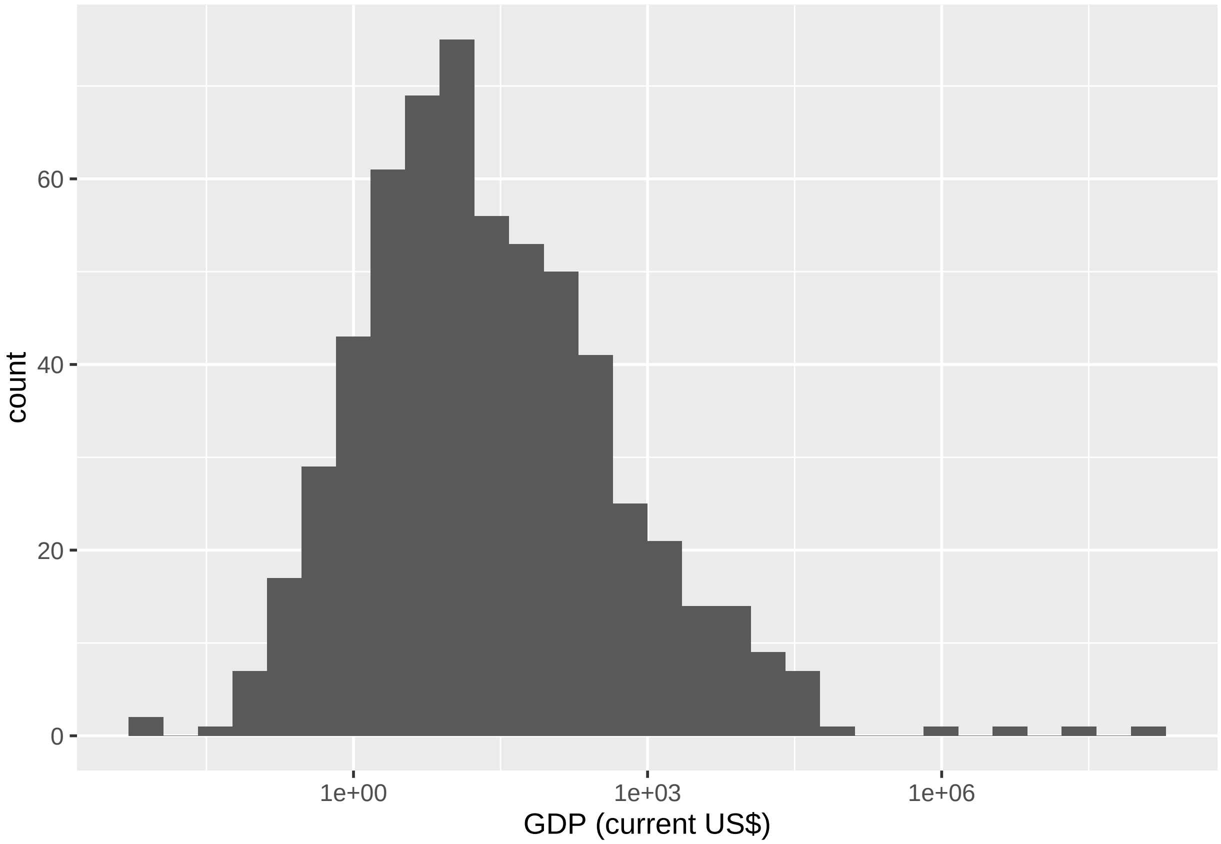
Provide summary statistics relevant to your research question. For example, if you're comparing means across groups provide means, SDs, sample sizes of each group. This step requires the use of R, hence a code chunk is provided below. Insert more code chunks as needed.

```
summary(worldbank$`GDP (current US$)`)
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.      NA's 
##         0         3        15    322500       167 166900000        96
```

```
ggplot(worldbank, aes(`GDP (current US$)`)) + geom_histogram() +  scale_x_log10()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



The GDP data is heavily skewed to the right, but it is generally normalized by a log transform.