Final Project EDA

Seth Metcalf

Table of Contents

### Introduction

This is the final project.

##### Importing World Bank Chart

library(ggplot2)  
worldbank = read.csv("C:\\Users\\sethc\\Documents\\2021 UC Berkeley Spring\\STAT20\\Final Project\\worldbank.csv")

*Due 7 May 2021.*

# GDP vs Business Start up Cost

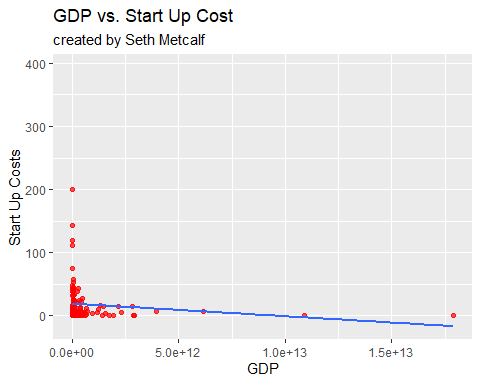
### Code

# Variables  
GDP = worldbank$GDP # Quantitative, gives a number for the Gross Domestic Product of a country  
start\_up\_cost = worldbank$start\_up\_cost # Quantitative, gives a percentage of the GNI for the start up of a business  
  
# Graph  
ggplot(worldbank,aes(x=GDP, y=start\_up\_cost)) +  
 geom\_point(alpha=0.7,color="red") +  
 geom\_smooth(method="lm",se=F) +   
 ggtitle(label = "GDP vs. Start Up Cost", subtitle = "created by Seth Metcalf") +   
 xlab("GDP") +  
 ylab("Start Up Costs")

## `geom\_smooth()` using formula 'y ~ x'

## Warning: Removed 8 rows containing non-finite values (stat\_smooth).

## Warning: Removed 8 rows containing missing values (geom\_point).



### Explanation

I wanted to see if there was a relation between the start up cost of a business and the GDP of the country. Even though that the graph is slightly skewed due to the outliers in GDP, we can see that there is a negative correlation when trying to compare the two. This can be observed to show that countries that have a higher start up caost for businesses tender to have an overall lower GDP compared to countries that don’t have a high start up cost for businesses.

# Amount of Labor vs Amount of Land

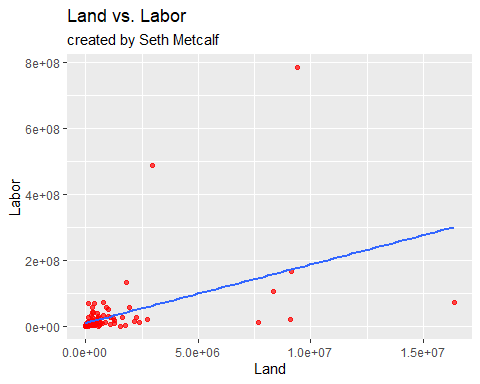
### Code

# Variables  
land = worldbank$land # Quantitative, amount of Land a country has in sq km  
labor = worldbank$labor # Quantitative, amount of Labor a country has (people 15 or older, employed or unemployed seeking a job)  
  
# Graph  
ggplot(worldbank,aes(x=land, y=labor)) +  
 geom\_point(alpha=0.7,color="red") +  
 geom\_smooth(method="lm",se=F) +   
 ggtitle(label = "Land vs. Labor", subtitle = "created by Seth Metcalf") +   
 xlab("Land") +  
 ylab("Labor")

## `geom\_smooth()` using formula 'y ~ x'

## Warning: Removed 2 rows containing non-finite values (stat\_smooth).

## Warning: Removed 2 rows containing missing values (geom\_point).



### Explanation

This graphs the relationship between the amount of labor a country has and the amount of land a country has. The line of best fit that is produced for us is one that shows a positive correlation between the two variables. This means that countries that tend to have more land tend to have a higher population, which makes sense as more land means more resources to expand a nation with. There are a few outliers, but the concentrated group of points in the bottom left of the graph gives us somewhat of an indication that the correlation holds true; it is observed that countries more land tend to have a bigger workforce.