

## GapMinder\_Analysis

Haneef Meeran  
6/16/2019

---

title: "EDA"  
output: html\_document  
author: "Haneef Meeran"  
date: "6/24/2019"  
output:  
  word\_document:  
    toc: yes  
  pdf\_document:  
    toc: yes  
  html\_document:  
    df\_print: paged  
    highlight: tango  
    theme: cosmo  
    toc: yes

---

### ### Data Description

- Country describes the primary variable describes the location of the life expectancy measured
- Region describes the continent related to the country
- Year - Provided for each country from 1800 to 2015
- Life - Provided for each country by year
- Population - Data available for each starting decade of each country
- Income - Continuous data series for each country by year

##### Income - Gross domestic product per person adjusted for differences in purchasing power (GDP per capita, PPP\$ inflation adjusted)

##### Life - The average number of years a newborn child would live if current mortality pattern were to stay the same

```
##### population - Population for the given country in the given year
```

```
### Set working directory and Import raw data
```

```
`r`  
setwd("~/Desktop/EXPL DATA ANALYSIS/EDA REPORT")
```

```
gapminder <- read.csv("gapminder.csv")
```

```
...
```

```
### Required Libraries
```

```
`r`
```

```
library(dplyr)
```

```
library(ggplot2)
```

```
library(matrixStats)
```

```
library(scales)
```

```
...
```

```
### Summary of data
```

```
`r`{r summary}
```

```
#Summary of Gapminder
```

```
summary(gapminder)
```

```
str(gapminder)
```

```
...
```

```
###Remove blanks
```

```
`r`
```

```
#Number of Unique Countries
```

```
length(unique(gapminder$Country))
```

```
#Cleaning of population field. Removing commas and reloading as numeric in a different  
dataset without NAs.
```

```
gapminder$population1 <- as.numeric(gsub(",", "", gapminder$population))
```

```
gapminder1 <- gapminder %>% filter(!is.na(gapminder$population1))
```

```
...
```

### Questions What is correlation between Income and Population by age groups

### Analysis of Income and Population

```
```{r income study}
```

```
hist(gapminder$income)
```

```
summary(unique(gapminder$income))
```

```
weightedMedian(gapminder$income, w = gapminder$region, na.rm = TRUE)
```

```
weightedMean(gapminder$income, w = gapminder$region, na.rm = TRUE)
```

```
d <- ggplot(gapminder1, aes(x = population1, y = income))
```

```
d + geom_point()+geom_smooth(method = "lm") + scale_x_continuous(labels = scales::comma)
```

```
```
```

Looking at the histogram, a high frequency of income in the range below 50000.

The Histogram shows a very high life expectancy starting at 20, the highest being 20 to 40 and an average income of about 2000 as we observe towards 80.

We can observe that as the population increases, the income hardly increases. If the population is lower, the income increases exponentially.

###Question Compare population between India and China

##### compare the top most populations of the world, Population between India and China

```
```{r population between india and china}
```

```
# Creating subset of India Country
```

```
population_india <- gapminder1 %>% filter(Country=='India')
```

```
# Creating subset of India Country
```

```
population_china <- gapminder1 %>% filter(Country=='China')
```

```
ggplot(data= population_india, mapping = aes(y=Year, x=population1, color="India")) +  
geom_point(data = gapminder1, aes(color="Others")) + geom_jitter()
```

```
ggplot(data= population_china, mapping = aes(y=Year, x=population1, color="China")) +  
geom_point(data = gapminder1, aes(color="Others")) + geom_jitter()
```

```
```
```

The population in India and China has risen exponentially after 1950.

###Question Lets find the highest GDP per Capita

##### Data Analysis of Top Countries with High GDP Per Capita from 1990

```
```{r High GDP, warning = FALSE}
```

```
gdp <- select(gapminder1, Country, Year, income) %>% filter(Year > 1990) %>%  
group_by(Country) %>% summarise(avg = mean(income)) %>% arrange(desc(avg))
```

```
gdp <- gdp[1:8,]
```

```
gdp_top <- filter(gapminder1, Country == gdp$Country) %>% filter(Year > 1990)
```

```
ggplot(gdp_top, aes(Country, income, color = Country)) +  
  geom_boxplot() + labs(title = "Top Countries with High GDP Per Capita", y = "Year", x =  
"Country_Income")  
```
```

###What is the Life expectancy by continent?

##### Data Analysis of Life expectancy by continent

```
```{r Life-expectancy}
```

```
hist(gapminder$life)
```

```
summary(gapminder$life)
```

```
gapminder_2015 %>%
```

```
ggplot(aes(x = life, fill = region)) + facet_wrap(~Year) + # aes = aesthetics  
geom_density(alpha = 0.7) +
```

```
  ggtitle('Life Expectancy by continent') +  
  theme(legend.title = element_text(color = 'black',size = 14, face = 'bold'),  
        legend.background = element_rect(fill = 'gray90', size = 0.5, linetype = 'dashed')) +  
  labs(x='Life Expected (Years)', y='Density')  
```
```

The Histogram shows a very high life expectancy starting at 20, the highest being 20 to 40 and an average income of about 2000 as we observe towards 80.

##### Summary

**\*\*Income\*\***<br />

Income for some of the regions are higher than compared to other regions. From the plots above (population vs income), we can also find that as the population increases, the income hardly increases for most, while is linear for some.

However, if the population is lower always, the income increases exponentially.

**\*\*GDP Per Capita\*\***<br />

We also analyzed the Top Countries with the highest GDP per capita from the year 1990,

**\*\*Life Expectancy\*\***<br />

We find that life expectancy in the given dataset is higher in the range of 25-35 years. The mean of Life Expectancy in 2015 for Sub-Saharan Africa region is lower than all other regions, while Europe & Central Asia is highest.

**\*\*There is a relationship between Life Expectancy and income.**

As the life expectancy grows above 60, the income levels rise higher.

When magnified the analysis to American region, the life Expectancy is almost 70 to 75.

Though the average life expectancy of America improves from below average to above average from 1800 to 2015, the life expectancy of United States has been highest compared to others.