# Biostatistics Assignment 1: Report

Name: Deep Pooja
Roll  No.: 17074

## Section 1

There are a total 187 countries and I wanted to have a big picture, so I decided  to analyse the data on a global scale. I have made a small animation using the data which gives us a clear visualization of how the countries in the world are doing in terms of various factors in the data such as life expectancy, child mortality, children per women, GDP per capita and total population on a world map. It also helps in forming the hypothesis and capturing imperative patterns in the data.

### Life expectancy

At the beginning, no country in the world had a life expectancy longer than 40 years. Every country is shown in purple. Almost everyone in the world lived in extreme poverty, we had very little medical knowledge, and in all countries our ancestors had to prepare for an early death. Life expectancy has increased rapidly since the Age of Enlightenment. In the early 19th century, life expectancy started to increase in the early industrialized countries while it stayed low in the rest of the world.

Over the next 150 years *some* parts of the world achieved substantial health improvements. A global divide opened (global inequality). In 1950 the life expectancy for newborns was already over 60 years in Europe, North America, Oceania, Japan and parts of South America. But elsewhere a newborn could only expect to live around 30 years.

If you look closely then you realize that there is a sudden black out in 1918 in regions Asia, Africa and Europe, I suspect this is due to Spanish flu. Global inequality in terms of health is maximum in late nineties and started to decrease from the beginning of twenties and It is projected in data that at the end of twenties we will have global equality in terms of life expectancy which implies improvement in global health conditions. Also I would like to point out that life expectancy has increased at all ages since decrease in child mortality(apparent from child mortality animation) could be thought as the reason to increase in life expectancy but this is not the case world is confident of the health of people at all ages.

| Color code | life expectancy range |
|---|---|
| Black | 0-20 |
| Purple | 20-40 |
| Pink | 40-60 |
| Mustard | 60-80 |
| Light yellow | 80-100 |
| White | Missing |


```
################### BEGIN OF CODE ###########################################
#Step 1: Convert Life expectancy in year data to long format
getwd()
setwd("C:/Users/hp/Downloads/Biostatistics/assignment1")  # setting the working directory
lifeExpectancyInYears <- read.csv("life_expectancy_years.csv", header = T, check.names =
FALSE)  # reading the life_expectancy_years CSV file
# The data is in wide format and I am used to work with long format data so I will use tidyr
#library to convert this wide format data into long format data
library(tidyr)  # importing tidyr library to wide
 long_format <- lifeExpectancyInYears.long <- pivot_longer(lifeExpectancyInYears, cols=2:302,
names_to = "year", values_to = "life_expec")
long_format  # life expectancy data in long format
country          year    life_expec
Afghanistan     1800    28.20
Afghanistan     1801    28.20
#Step 2: Get dataframe for geometries of counties in world (map dataframe) and merge the map
#dataframe and life expectancy dataframe
library(rnaturalearth)
library(dplyr)
library(rgeos)
library(sf)
world = ne_countries(returnclass = "sf")%>%
  select(iso_a2, name_long, continent)
World   # world map dataframe contains countries and their geomtries
merged_df <- merge(world, long_format, by.x = "name_long", by.y = "country", all.x = TRUE)
head(merged_df)  #  merging using merge() function
    name_long iso_a2  continent   year   life_expec                      geometry
1 Afghanistan     AF      Asia       2001          54.8         MULTIPOLYGON (((61.21082 35...
2 Afghanistan     AF      Asia       1997           53.7         MULTIPOLYGON (((61.21082 35...
3 Afghanistan     AF      Asia       1996           53.8         MULTIPOLYGON (((61.21082 35...
```

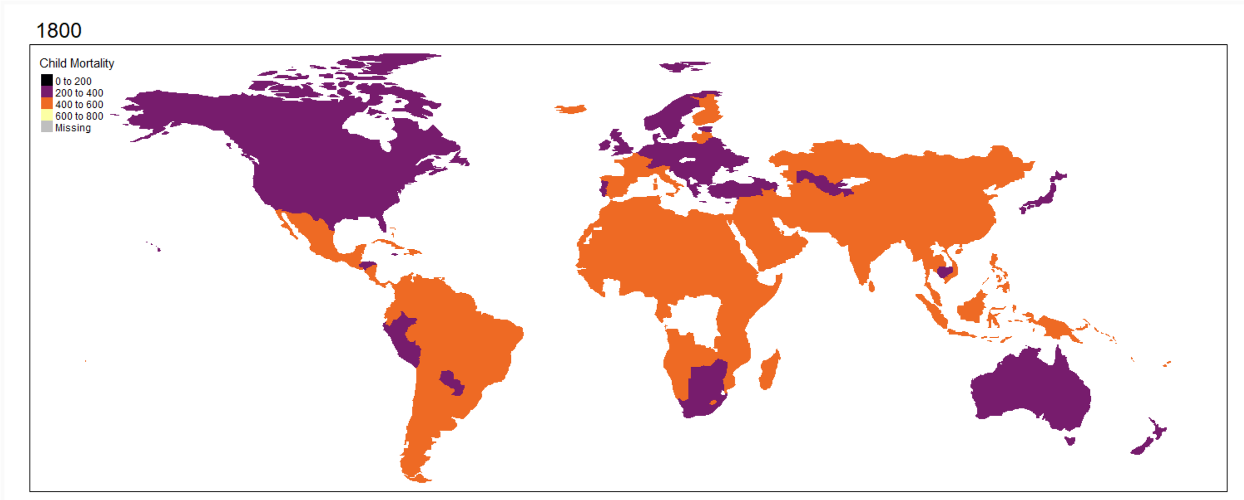
```
#Note: column year got disordered while merging to fix it, I have used order() function to sort the
#dataframe
attach(merged_df)
newdata <- merged_df[order(year),]
detach(merged_df)
newdata
 name_long   iso_a2    continent   year    life_expec                    geometry
 Afghanistan     AF         Asia        1800       28.2         MULTIPOLYGON (((61.21082 35...
Albania            AL        Europe      1800       35.4         MULTIPOLYGON (((20.59025 41...
Algeria            DZ        Africa        1800       28.8         MULTIPOLYGON (((11.99951 23…
# Hooray! Years are ordered now
# Step 3: Making the Animation using tmap library :)
library(tmap)
library(sf)
m <- "+proj=longlat +datum=WGS84 +no_defs +ellps=WGS84 +towgs84=0,0,0"  # CRS of the
#newdata needed for projection
my_ani <- tm_shape(newdata, projection = m ) +
  tm_fill("life_expec", title = "Life expectancy",
          palette = "inferno") +
  tm_facets( along = "year" ) +tm_layout(frame = TRUE)

tmap_animation(my_ani, filename = "Animation.gif",
             width = 1500, height = 600, delay = 60)
##################### END OF CODE #########################################
```

## Child Mortality

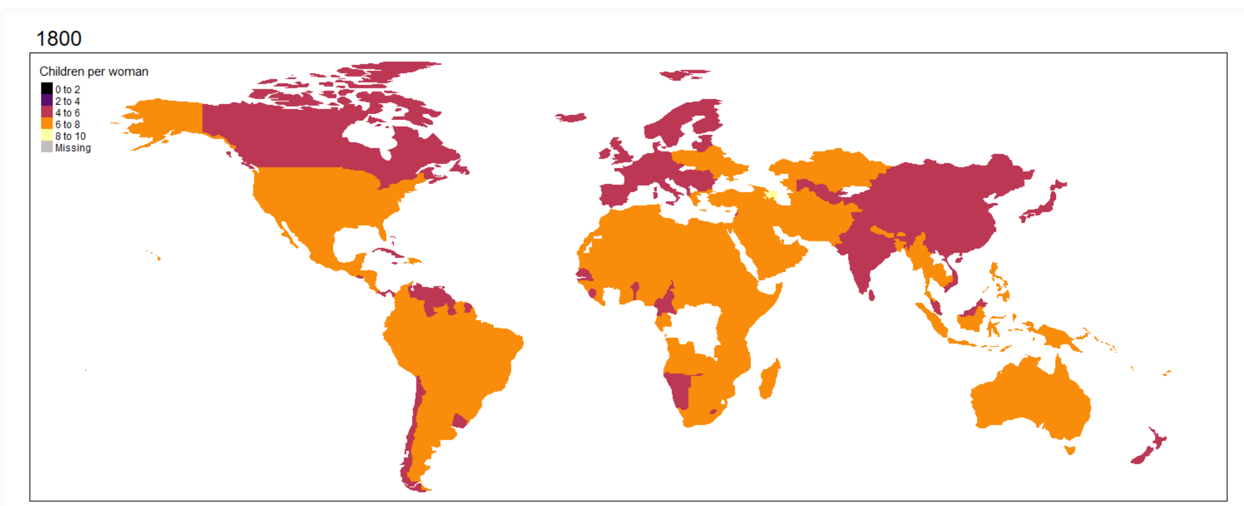
Similar trend is observed in child mortality animation. At the beginning, child mortality was high in the world especially in Asia and Africa, western world was doing a little better. Over the 150 years child mortality rate has decreased dramatically across all regions in the world and in the late nineties the world looked black ( equality in terms of child mortality) but if you see the life expectancy animation there was still a global divide. Therefore child mortality is not a direct measure of overall health of the world's population. I hypothesise life expectancy is a much better measure of world’s health.

| Color code | Child Mortality |
|---|---|
| Black | 0-200 |
| Purple | 200-400 |
| Orange | 400-600 |
| Lemon yellow | 600-800 |
| White | Missing |




## Children per Women

| Color code | Children Per Woman |
|---|---|
| Black | 0-2 |
| Purple | 2-4 |
| Pink | 4-6 |
| Orange | 6-8 |
| Lemon yellow | 8-10 |
| White | Missing |




Similar trends are observed, At the beginning, every woman had on an average six children in the world and the number started to decrease dramatically in early twenties (except Asia) this could be due to the health improvements of infants (as evident from decrease in child mortality during this time).
So I hypothesise, In pre-modern world women had more children because there was a risk of child death (high child mortality) because of poor health facilities for infants, once there was improvement in health facilities across the world especially infants health women chose to have fewer children as risk of losing

the child decreased. Also as the number decreases leads to increase in life expectancy as we see in later animations

## Income per Person

I really suspect the trend here, since almost till late nineties income per person in every country was lying in one range i.e 1 to 50 thousand. It started to increase in oil rich countries (which is expected because of the rise in demand for oil all across the world). It also increases later in western world.

| Color code | Income Per Person(k = thousand) |
|---|---|
| Black | (1-50)k |
| Purple | (50k-100)k |
| Orange | (100k-150)k |
| Lemon yellow | (150k-200)k |
| White | Missing |


## Total Population

The population of every county almost remains constant in certain ranges but in the early twenties there was a boom in population of china and India and it is projected that chinese population will remain on an average same but India's population will further increase.

| Color code | Total Population(mln = million) |
|---|---|
| Black | (0-500 )mln |
| Purple | (500-1000)mln |
| Orange | (1000-1500)mln |
| Lemon yellow | (1500-2000)mln |
| White | Missing |


## Section 2

In this section I have analyzed how factors like child mortality, children per woman, total population and GDP per capita have affected the life expectancy of all countries over the years because I hypothesize that life expectancy is a much better measure of health improvements all across the world.

```
########################### Begin of Code #########################################
#Step 1: Reading each csv file and converting it to long format
getwd()
setwd("C:/Users/hp/Downloads/Biostatistics/assignment1")
population_total <- read.csv("population_total.csv", header = T, check.names = FALSE)
child_mortality <- read.csv("child_mortality_0_5_year_olds_dying_per_1000_born.csv", header
= T, check.names = FALSE)
child_per_woman <- read.csv("children_per_woman_total_fertility.csv", header = T,
check.names = FALSE)
lifeExpectancyInYears <- read.csv("life_expectancy_years.csv", header = T, check.names =
FALSE)
income_per_person <-
read.csv("income_per_person_gdppercapita_ppp_inflation_adjusted.csv", header = T,
check.names = FALSE)

library(tidyr)
long_population <- population_total.long <- pivot_longer(population_total, cols=2:302, names_to
= "year", values_to = "population_total")

long_child_mortality <- child_mortality.long <- pivot_longer(child_mortality, cols=2:302,
names_to = "year", values_to = "child_mortality")
```

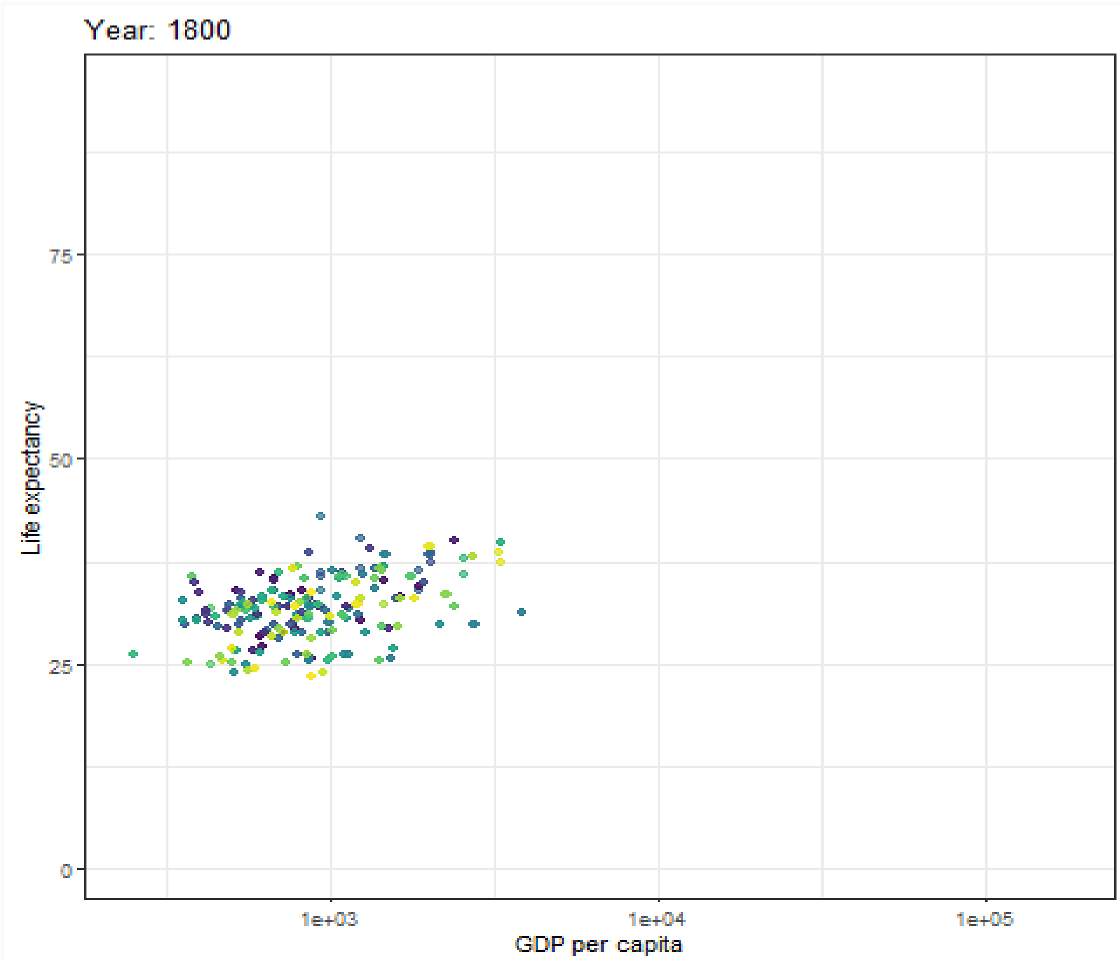
```
long_income <- income_per_person.long <- pivot_longer(income_per_person, cols=2:242,
names_to = "year", values_to = "income_per_person")
long_child_per_woman <- child_per_woman.long <- pivot_longer(child_per_woman, cols=2:302,
names_to = "year", values_to = "child_per_woman")
long_life_expectancy <- lifeExpectancyInYears.long <- pivot_longer(lifeExpectancyInYears,
cols=2:302, names_to = "year", values_to = "life_expec")
#Step 2: Merging all long format dataframes  to one dataframe
library(dplyr)
merged <- Reduce(function(x, y) merge(x, y, all=TRUE),
list(long_population,long_child_mortality, long_income,  long_child_per_woman,
long_life_expectancy))

merged$year <- as.numeric(as.character(merged$year))
merged
# Step 3: Plotting :)
library(ggplot2)
library(gganimate)
theme_set(theme_bw())

p <- ggplot(
  merged,
  aes(x =income_per_person , y=life_expec, colour = country)
  ) +
  geom_point(show.legend = FALSE, alpha = 0.7) +
  scale_color_viridis_d() +
  scale_size(range = c(2, 12)) +
  scale_x_log10() +
  labs(x = "GDP per capita", y = "Life expectancy")
p
p + transition_time(year) +
  labs(title = "Year: {frame_time}") +
shadow_wake(wake_length = 0.1, alpha = FALSE)
#################################### End Of Code ##############################
```

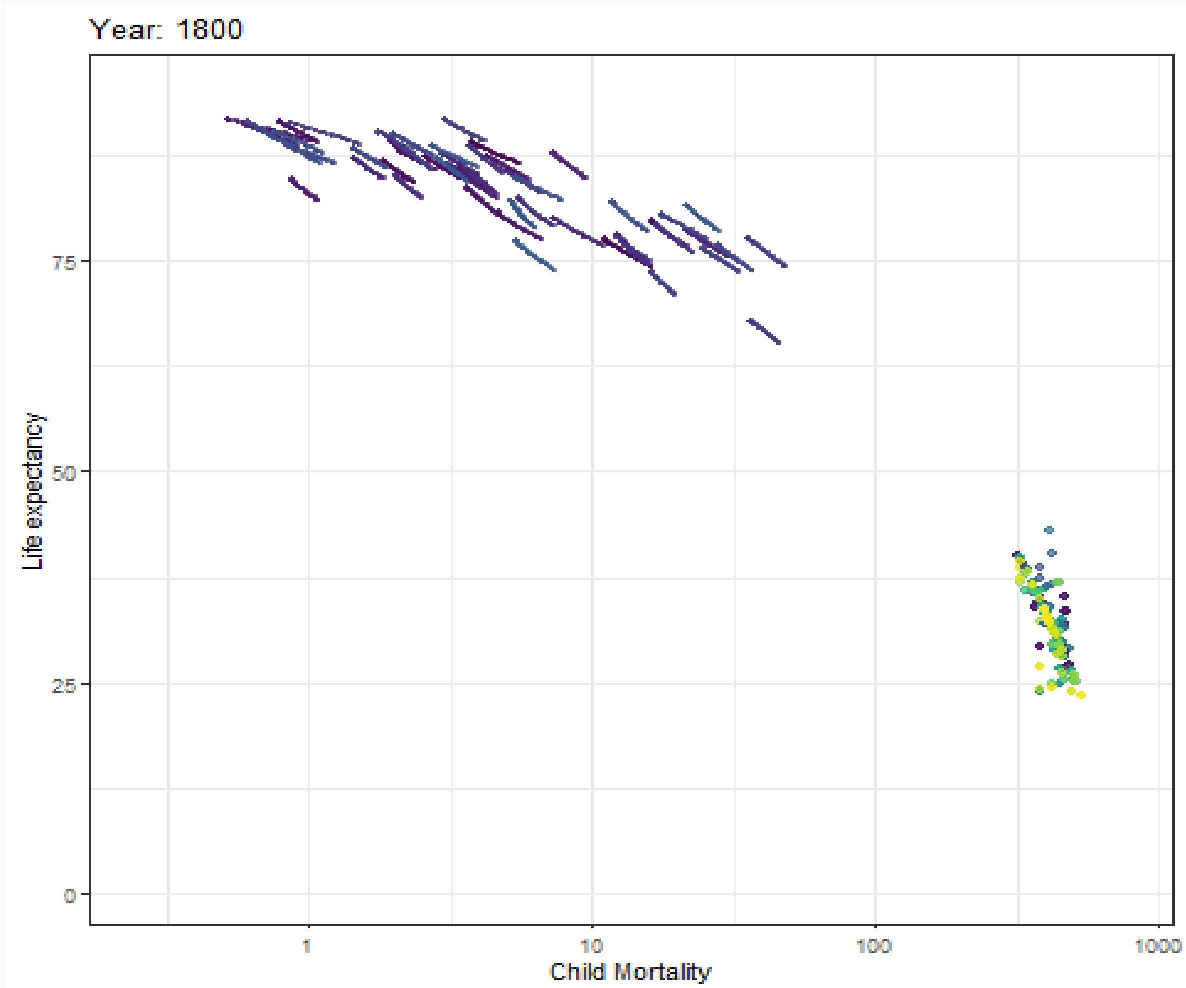
## Life Expectancy VS GDP Per Capita

The small animation below makes it clear that there is a direct correlation between the life expectancy and GDP per capita for almost all countries. Earlier both life expectancy and GDP per capita were low but once the GDP starts to increase life expectancy also increases since people can afford the health facilities.



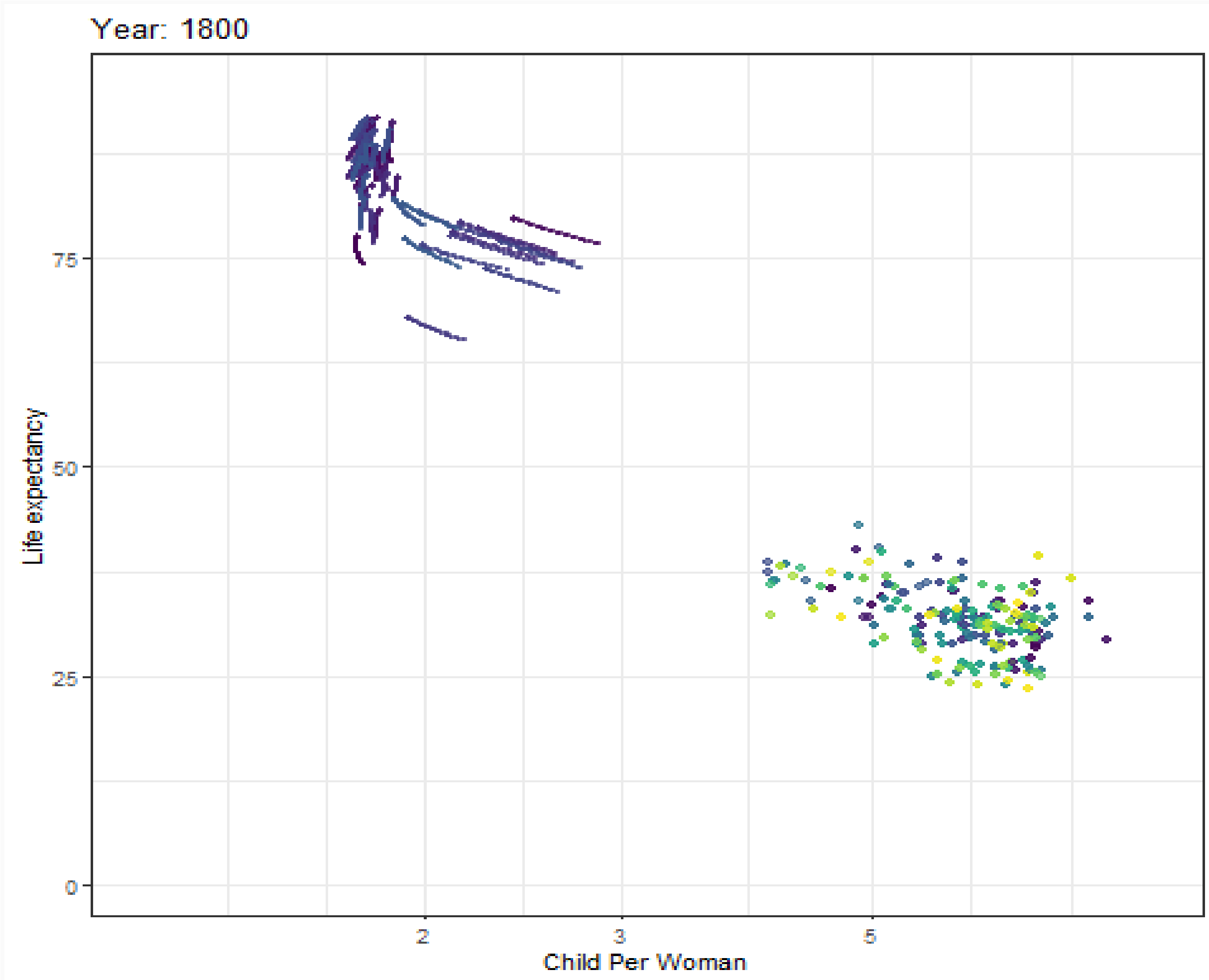

## Life Expectancy VS Child Mortality

This also shows a direct correlation of child mortality rate and life expectancy .Life expectancy has increased at all ages. The average person can expect to live a longer life than in the past, irrespective of what age they are.



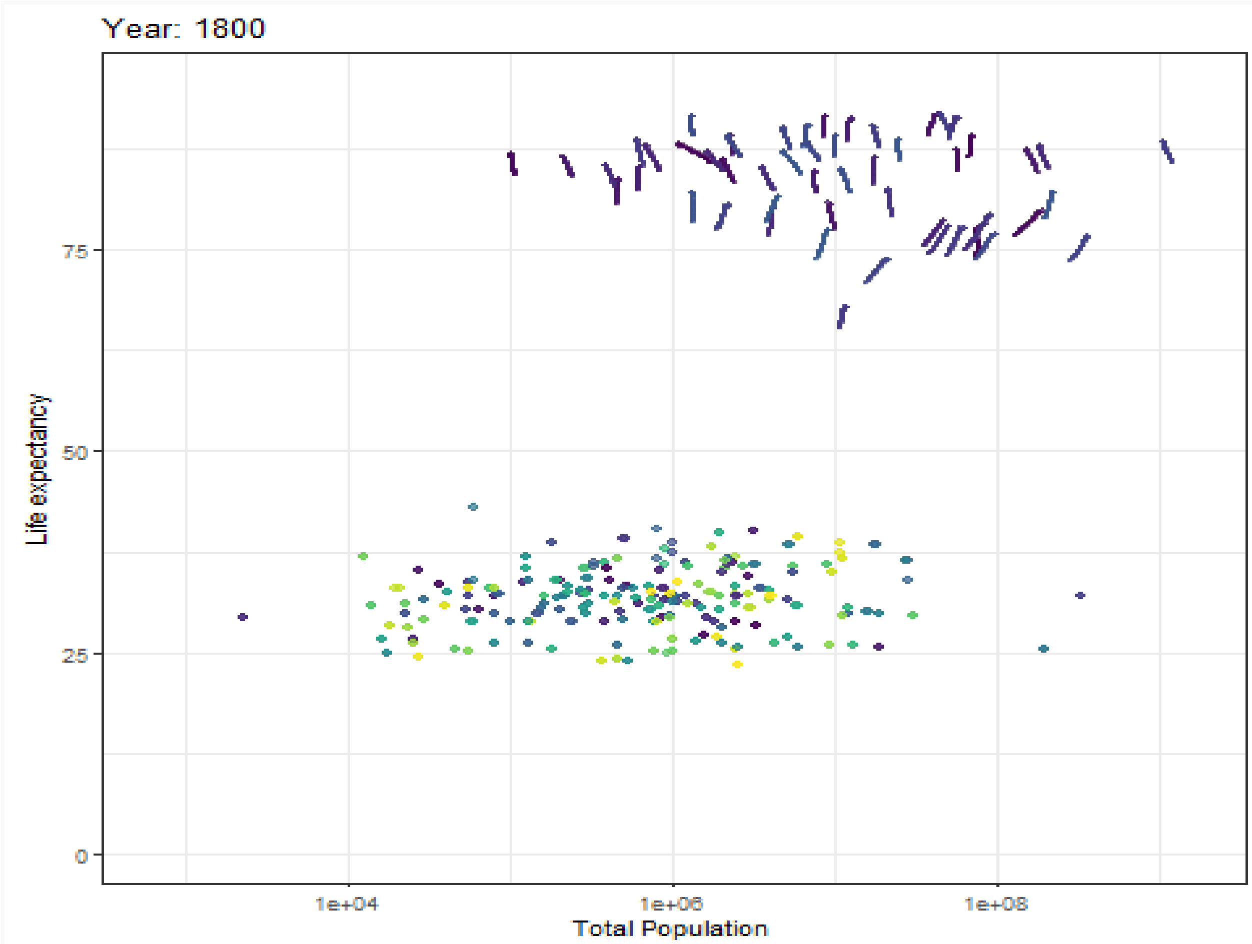

## Life Expectancy VS Child Per Woman

This is interesting as it is projected that the number of children per woman will saturate to two children for all countries . And here also we see direct correlation of children per woman and life expectancy. This is because less the number of children per woman implies better woman's health and hence increases the life expectancy of the population where women count too.




## Life Expectancy VS Total Population

There is an inverse relationship between life expectancy and total population of the country as it seems even an increase in population didn't lead to an increase in life expectancy but as life expectancy increases dramatically total population remains constant and even decreases for some countries.




## Conclusion:

1. Life expectancy is a better indicator of population health and our health and medical capabilities.
2. Life expectancy is directly correlated to child mortality
3. Life expectancy is directly related to children per women
4. Life expectancy is directly related to GDP per capita
5. Life expectancy is inversely related to total population of the country