Life Expectancy Analysis

Nikhil Gupta

11/15/2021

# **INTRODUCTION** :

Life Expectancy of human life has been one of the central idea behind various researches and increasing it has been a constant effort of humanity.Thus, understanding the available data about it as well as various parameters that effect the life expectancy becomes increasingly important. In light of this notion, this study is trying to fundamentally analyse the available LIfe expectancy data as well as make suggestions on how we can try to improve it by influencing various primary and secondary parameters associated with it. This study will focus on immunization factors, mortality factors, economic factors, social factors and other health related factors as well. Since the observations this dataset are based on different countries, it will be easier for a country to determine the predicting factor which is contributing to lower value of life expectancy. This will help in suggesting a country which area should be given importance in order to efficiently improve the life expectancy of its population.

# **DATA SOURCE** :

This given data has been sourced from the data repository of *Kaggle*.

url : [*https://www.kaggle.com/kumarajarshi/life-expectancy-who*](https://www.kaggle.com/kumarajarshi/life-expectancy-who)

This dataset related to life expectancy, health factors for 193 countries has been collected from the The Global Health Observatory (GHO) data repository under World Health Organization (WHO) and its corresponding economic data was collected from United Nation website. Among all categories of health-related factors only those critical factors were chosen which are more representative. It has been observed that in the 15 years period ranging from 2000-2015 , there has been a huge development in health sector resulting in improvement of human mortality rates especially in the developing nations in comparison to the past 30 years. Therefore, in this project we have considered data from year 2000-2015 for 193 countries for further analysis. The individual data files have been merged together into a single dataset. All predicting variables was then divided into several broad categories:​Immunization related factors, Mortality factors, Economical factors and Social factors.

# **QUESTIONS OF INTEREST** :

The data-set aims to answer the following key questions: -Compare the life expectancy trends of Developed world and Developing world.

-Does various predicting factors which has been chosen initially really affect the Life expectancy? What are the predicting variables actually affecting the life expectancy?

-Does Life Expectancy has positive or negative correlation with eating habits, lifestyle, exercise, smoking, drinking alcohol etc.

-How does mortality factors like Infant and Adult mortality rates affect life expectancy?

-What is the impact of Immunization coverage on life Expectancy?

-What is the impact of increased health spending on the lifespan of humans?Analyse the relationship of various economic factors vis-a-vis Life Expectancy.

-Find the mean life expectancy for all the people across globe, averaged over the given time period ?

-Top 10 countries with highest and Least Life expectancy across globe ?

-Probability density function of global Avg. Life Expectancies to find most likely L.E. for world countries.

-Country-wise comparison of top performers : in terms of LIfe Expectancy

-Linear Model to predict Life expectancy against the parameter of schooling

# **DETAILED STATISTICAL ANALYSIS** :

## Importing relevant *libraries*

library(readr)  
library(httr)  
library(tidyverse)

## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5 v dplyr 1.0.7  
## v tibble 3.1.4 v stringr 1.4.0  
## v tidyr 1.1.3 v forcats 0.5.1  
## v purrr 0.3.4

## -- Conflicts ------------------------------------------ tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

library(lubridate)

##   
## Attaching package: 'lubridate'

## The following objects are masked from 'package:base':  
##   
## date, intersect, setdiff, union

library(dplyr)  
library(ggplot2)  
library(readr)

## Importing the *data* for Analysis

data<- read\_csv( "Life Expectancy Data.csv",show\_col\_types = FALSE)  
data

## # A tibble: 2,938 x 22  
## Country Year Status `Life expectanc~ `Adult Mortalit~ `infant deaths`  
## <chr> <dbl> <chr> <dbl> <dbl> <dbl>  
## 1 Afghanistan 2015 Developing 65 263 62  
## 2 Afghanistan 2014 Developing 59.9 271 64  
## 3 Afghanistan 2013 Developing 59.9 268 66  
## 4 Afghanistan 2012 Developing 59.5 272 69  
## 5 Afghanistan 2011 Developing 59.2 275 71  
## 6 Afghanistan 2010 Developing 58.8 279 74  
## 7 Afghanistan 2009 Developing 58.6 281 77  
## 8 Afghanistan 2008 Developing 58.1 287 80  
## 9 Afghanistan 2007 Developing 57.5 295 82  
## 10 Afghanistan 2006 Developing 57.3 295 84  
## # ... with 2,928 more rows, and 16 more variables: Alcohol <dbl>,  
## # percentage expenditure <dbl>, Hepatitis B <dbl>, Measles <dbl>, BMI <dbl>,  
## # under-five deaths <dbl>, Polio <dbl>, Total expenditure <dbl>,  
## # Diphtheria <dbl>, HIV/AIDS <dbl>, GDP <dbl>, Population <dbl>,  
## # thinness 1-19 years <dbl>, thinness 5-9 years <dbl>,  
## # Income composition of resources <dbl>, Schooling <dbl>

## Understanding the Data-set

We are looking at the tibble and the summary statistics to get more insight about the data such as size of data, quantum of data, nature of data (clean/unclean). The above results indicates that some of the data entries are missing or N/A. The most of the missing data was for population, Hepatitis B and GDP. The missing data were from less known countries like Vanuatu, Tonga, Togo,Cabo Verde etc.The final merged file(final dataset) consists of 22 Columns and 2938 rows which meant 20 predicting variables.

summary(data)

## Country Year Status Life expectancy  
## Length:2938 Min. :2000 Length:2938 Min. :36.30   
## Class :character 1st Qu.:2004 Class :character 1st Qu.:63.10   
## Mode :character Median :2008 Mode :character Median :72.10   
## Mean :2008 Mean :69.22   
## 3rd Qu.:2012 3rd Qu.:75.70   
## Max. :2015 Max. :89.00   
## NA's :10   
## Adult Mortality infant deaths Alcohol percentage expenditure  
## Min. : 1.0 Min. : 0.0 Min. : 0.0100 Min. : 0.000   
## 1st Qu.: 74.0 1st Qu.: 0.0 1st Qu.: 0.8775 1st Qu.: 4.685   
## Median :144.0 Median : 3.0 Median : 3.7550 Median : 64.913   
## Mean :164.8 Mean : 30.3 Mean : 4.6029 Mean : 738.251   
## 3rd Qu.:228.0 3rd Qu.: 22.0 3rd Qu.: 7.7025 3rd Qu.: 441.534   
## Max. :723.0 Max. :1800.0 Max. :17.8700 Max. :19479.912   
## NA's :10 NA's :194   
## Hepatitis B Measles BMI under-five deaths  
## Min. : 1.00 Min. : 0.0 Min. : 1.00 Min. : 0.00   
## 1st Qu.:77.00 1st Qu.: 0.0 1st Qu.:19.30 1st Qu.: 0.00   
## Median :92.00 Median : 17.0 Median :43.50 Median : 4.00   
## Mean :80.94 Mean : 2419.6 Mean :38.32 Mean : 42.04   
## 3rd Qu.:97.00 3rd Qu.: 360.2 3rd Qu.:56.20 3rd Qu.: 28.00   
## Max. :99.00 Max. :212183.0 Max. :87.30 Max. :2500.00   
## NA's :553 NA's :34   
## Polio Total expenditure Diphtheria HIV/AIDS   
## Min. : 3.00 Min. : 0.370 Min. : 2.00 Min. : 0.100   
## 1st Qu.:78.00 1st Qu.: 4.260 1st Qu.:78.00 1st Qu.: 0.100   
## Median :93.00 Median : 5.755 Median :93.00 Median : 0.100   
## Mean :82.55 Mean : 5.938 Mean :82.32 Mean : 1.742   
## 3rd Qu.:97.00 3rd Qu.: 7.492 3rd Qu.:97.00 3rd Qu.: 0.800   
## Max. :99.00 Max. :17.600 Max. :99.00 Max. :50.600   
## NA's :19 NA's :226 NA's :19   
## GDP Population thinness 1-19 years  
## Min. : 1.68 Min. :3.400e+01 Min. : 0.10   
## 1st Qu.: 463.94 1st Qu.:1.958e+05 1st Qu.: 1.60   
## Median : 1766.95 Median :1.387e+06 Median : 3.30   
## Mean : 7483.16 Mean :1.275e+07 Mean : 4.84   
## 3rd Qu.: 5910.81 3rd Qu.:7.420e+06 3rd Qu.: 7.20   
## Max. :119172.74 Max. :1.294e+09 Max. :27.70   
## NA's :448 NA's :652 NA's :34   
## thinness 5-9 years Income composition of resources Schooling   
## Min. : 0.10 Min. :0.0000 Min. : 0.00   
## 1st Qu.: 1.50 1st Qu.:0.4930 1st Qu.:10.10   
## Median : 3.30 Median :0.6770 Median :12.30   
## Mean : 4.87 Mean :0.6276 Mean :11.99   
## 3rd Qu.: 7.20 3rd Qu.:0.7790 3rd Qu.:14.30   
## Max. :28.60 Max. :0.9480 Max. :20.70   
## NA's :34 NA's :167 NA's :163

## **Cleaning** the data set

looking at the *summary of the data* to see if there are any problems. Summary statistics show that there are many rows which contain NA values.Hence, cleaning the data by deleting the rows which contain NA’s.

data\_clean <-na.omit(data)  
print(data\_clean)

## # A tibble: 1,649 x 22  
## Country Year Status `Life expectanc~ `Adult Mortalit~ `infant deaths`  
## <chr> <dbl> <chr> <dbl> <dbl> <dbl>  
## 1 Afghanistan 2015 Developing 65 263 62  
## 2 Afghanistan 2014 Developing 59.9 271 64  
## 3 Afghanistan 2013 Developing 59.9 268 66  
## 4 Afghanistan 2012 Developing 59.5 272 69  
## 5 Afghanistan 2011 Developing 59.2 275 71  
## 6 Afghanistan 2010 Developing 58.8 279 74  
## 7 Afghanistan 2009 Developing 58.6 281 77  
## 8 Afghanistan 2008 Developing 58.1 287 80  
## 9 Afghanistan 2007 Developing 57.5 295 82  
## 10 Afghanistan 2006 Developing 57.3 295 84  
## # ... with 1,639 more rows, and 16 more variables: Alcohol <dbl>,  
## # percentage expenditure <dbl>, Hepatitis B <dbl>, Measles <dbl>, BMI <dbl>,  
## # under-five deaths <dbl>, Polio <dbl>, Total expenditure <dbl>,  
## # Diphtheria <dbl>, HIV/AIDS <dbl>, GDP <dbl>, Population <dbl>,  
## # thinness 1-19 years <dbl>, thinness 5-9 years <dbl>,  
## # Income composition of resources <dbl>, Schooling <dbl>

## Understanding the **Clean Data**

Clearly, the summary of the clean data shows that there are no more missing data fields.Though the minimum value for certain attributes are still 0, but 0 is a possible minimum value for those columns. Hence assuming zero to be the part of data and retaining them in the clean dataset. After cleaning the datasets, the final cleaned tibble contains 1649 rows with 22 columns. To get some more insight into the clean data, looking at the summary statistics.

summary(data\_clean)

## Country Year Status Life expectancy  
## Length:1649 Min. :2000 Length:1649 Min. :44.0   
## Class :character 1st Qu.:2005 Class :character 1st Qu.:64.4   
## Mode :character Median :2008 Mode :character Median :71.7   
## Mean :2008 Mean :69.3   
## 3rd Qu.:2011 3rd Qu.:75.0   
## Max. :2015 Max. :89.0   
## Adult Mortality infant deaths Alcohol percentage expenditure  
## Min. : 1.0 Min. : 0.00 Min. : 0.010 Min. : 0.00   
## 1st Qu.: 77.0 1st Qu.: 1.00 1st Qu.: 0.810 1st Qu.: 37.44   
## Median :148.0 Median : 3.00 Median : 3.790 Median : 145.10   
## Mean :168.2 Mean : 32.55 Mean : 4.533 Mean : 698.97   
## 3rd Qu.:227.0 3rd Qu.: 22.00 3rd Qu.: 7.340 3rd Qu.: 509.39   
## Max. :723.0 Max. :1600.00 Max. :17.870 Max. :18961.35   
## Hepatitis B Measles BMI under-five deaths  
## Min. : 2.00 Min. : 0 Min. : 2.00 Min. : 0.00   
## 1st Qu.:74.00 1st Qu.: 0 1st Qu.:19.50 1st Qu.: 1.00   
## Median :89.00 Median : 15 Median :43.70 Median : 4.00   
## Mean :79.22 Mean : 2224 Mean :38.13 Mean : 44.22   
## 3rd Qu.:96.00 3rd Qu.: 373 3rd Qu.:55.80 3rd Qu.: 29.00   
## Max. :99.00 Max. :131441 Max. :77.10 Max. :2100.00   
## Polio Total expenditure Diphtheria HIV/AIDS   
## Min. : 3.00 Min. : 0.740 Min. : 2.00 Min. : 0.100   
## 1st Qu.:81.00 1st Qu.: 4.410 1st Qu.:82.00 1st Qu.: 0.100   
## Median :93.00 Median : 5.840 Median :92.00 Median : 0.100   
## Mean :83.56 Mean : 5.956 Mean :84.16 Mean : 1.984   
## 3rd Qu.:97.00 3rd Qu.: 7.470 3rd Qu.:97.00 3rd Qu.: 0.700   
## Max. :99.00 Max. :14.390 Max. :99.00 Max. :50.600   
## GDP Population thinness 1-19 years  
## Min. : 1.68 Min. :3.400e+01 Min. : 0.100   
## 1st Qu.: 462.15 1st Qu.:1.919e+05 1st Qu.: 1.600   
## Median : 1592.57 Median :1.420e+06 Median : 3.000   
## Mean : 5566.03 Mean :1.465e+07 Mean : 4.851   
## 3rd Qu.: 4718.51 3rd Qu.:7.659e+06 3rd Qu.: 7.100   
## Max. :119172.74 Max. :1.294e+09 Max. :27.200   
## thinness 5-9 years Income composition of resources Schooling   
## Min. : 0.100 Min. :0.0000 Min. : 4.20   
## 1st Qu.: 1.700 1st Qu.:0.5090 1st Qu.:10.30   
## Median : 3.200 Median :0.6730 Median :12.30   
## Mean : 4.908 Mean :0.6316 Mean :12.12   
## 3rd Qu.: 7.100 3rd Qu.:0.7510 3rd Qu.:14.00   
## Max. :28.200 Max. :0.9360 Max. :20.70

## \*\* Data Visualization and Analysis\*\*

In order to answer the above mentioned questions of interest,we will be going through various visualizations along with their subsequent analysis in a orderly manner.

### Comparing the life expectancy of Developed world and Developing world

data\_global\_trend <- data\_clean %>% group\_by(Year,Status)%>% select(-c(Country))%>% summarise\_each(funs( mean( .,na.rm = TRUE)))%>%ungroup()

## Warning: `summarise\_each\_()` was deprecated in dplyr 0.7.0.  
## Please use `across()` instead.

## Warning: `funs()` was deprecated in dplyr 0.8.0.  
## Please use a list of either functions or lambdas:   
##   
## # Simple named list:   
## list(mean = mean, median = median)  
##   
## # Auto named with `tibble::lst()`:   
## tibble::lst(mean, median)  
##   
## # Using lambdas  
## list(~ mean(., trim = .2), ~ median(., na.rm = TRUE))

ungroup(data\_clean)

## # A tibble: 1,649 x 22  
## Country Year Status `Life expectanc~ `Adult Mortalit~ `infant deaths`  
## <chr> <dbl> <chr> <dbl> <dbl> <dbl>  
## 1 Afghanistan 2015 Developing 65 263 62  
## 2 Afghanistan 2014 Developing 59.9 271 64  
## 3 Afghanistan 2013 Developing 59.9 268 66  
## 4 Afghanistan 2012 Developing 59.5 272 69  
## 5 Afghanistan 2011 Developing 59.2 275 71  
## 6 Afghanistan 2010 Developing 58.8 279 74  
## 7 Afghanistan 2009 Developing 58.6 281 77  
## 8 Afghanistan 2008 Developing 58.1 287 80  
## 9 Afghanistan 2007 Developing 57.5 295 82  
## 10 Afghanistan 2006 Developing 57.3 295 84  
## # ... with 1,639 more rows, and 16 more variables: Alcohol <dbl>,  
## # percentage expenditure <dbl>, Hepatitis B <dbl>, Measles <dbl>, BMI <dbl>,  
## # under-five deaths <dbl>, Polio <dbl>, Total expenditure <dbl>,  
## # Diphtheria <dbl>, HIV/AIDS <dbl>, GDP <dbl>, Population <dbl>,  
## # thinness 1-19 years <dbl>, thinness 5-9 years <dbl>,  
## # Income composition of resources <dbl>, Schooling <dbl>

print(data\_global\_trend)

## # A tibble: 31 x 21  
## Year Status `Life expectancy` `Adult Mortality` `infant deaths` Alcohol  
## <dbl> <chr> <dbl> <dbl> <dbl> <dbl>  
## 1 2000 Developed 76.1 91.7 1.38 10.6   
## 2 2000 Developing 68.6 197. 26.7 3.76  
## 3 2001 Developed 76.6 84.9 1.29 10.4   
## 4 2001 Developing 68.6 181. 25.8 3.85  
## 5 2002 Developed 76.9 94.9 1.29 10.7   
## 6 2002 Developing 66.9 188. 25.7 3.81  
## 7 2003 Developed 77.2 78.9 1.07 10.6   
## 8 2003 Developing 66.7 191. 30.2 3.59  
## 9 2004 Developed 77.9 87.8 1.07 10.9   
## 10 2004 Developing 66.8 201. 45.7 3.67  
## # ... with 21 more rows, and 15 more variables: percentage expenditure <dbl>,  
## # Hepatitis B <dbl>, Measles <dbl>, BMI <dbl>, under-five deaths <dbl>,  
## # Polio <dbl>, Total expenditure <dbl>, Diphtheria <dbl>, HIV/AIDS <dbl>,  
## # GDP <dbl>, Population <dbl>, thinness 1-19 years <dbl>,  
## # thinness 5-9 years <dbl>, Income composition of resources <dbl>,  
## # Schooling <dbl>

#### Plotting the avg. Life expectancy trend for Developed Vs Developing.

Below is the plot of Avg. Life Expectancy of developed nations Vs developing nations, averaged over the past few years (2000-2015) period .

data\_global\_trend %>%  
 ggplot(aes(x = Year, y = `Life expectancy`)) +  
 geom\_line(aes(color = data\_global\_trend$`Life expectancy`, shape = data\_global\_trend$Status)) +  
 geom\_point(aes(color = data\_global\_trend$`Life expectancy`, shape = data\_global\_trend$Status)) +  
 labs(shape = "Type", color = NULL)+  
 ylim(60,90)+  
 theme(legend.position="bottom",  
 axis.text.x = element\_text(angle = 90)) +  
 labs(title = "Avg Life Expectancy Trend", y= "Life Expectancy (in years)" )

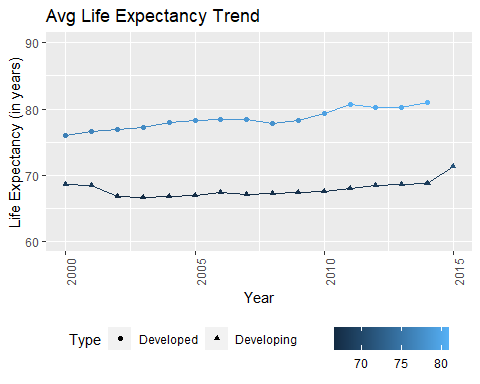
## Warning: Ignoring unknown aesthetics: shape

## Warning: Use of `data\_global\_trend$`Life expectancy`` is discouraged. Use `Life  
## expectancy` instead.

## Warning: Use of `data\_global\_trend$Status` is discouraged. Use `Status` instead.

## Warning: Use of `data\_global\_trend$`Life expectancy`` is discouraged. Use `Life  
## expectancy` instead.

## Warning: Use of `data\_global\_trend$Status` is discouraged. Use `Status` instead.

 #### Analyzing the above plots.

As can be seen from the Plots, following observations are made : - Life expectancy of developed countries is significantly higher than developing countries (almost 10yr higher). - Though both developed and developing countries have shown an increase in Life expectancy but the increase is higher for developed countries as compared to developing countries. - Even the maximun Life expectancy for developing countries is lower than the least Life expectancy of developed countries.

Lets cross-check the observations with the summary statistics.

developed<- data\_global\_trend %>% filter (Status == "Developed")%>% select(Year,`Life expectancy`)  
developing<-data\_global\_trend %>% filter (Status == "Developing")%>% select(Year, `Life expectancy`)  
summary(developed)

## Year Life expectancy  
## Min. :2000 Min. :76.08   
## 1st Qu.:2004 1st Qu.:77.55   
## Median :2007 Median :78.34   
## Mean :2007 Mean :78.52   
## 3rd Qu.:2010 3rd Qu.:79.83   
## Max. :2014 Max. :81.03

summary(developing)

## Year Life expectancy  
## Min. :2000 Min. :66.68   
## 1st Qu.:2004 1st Qu.:67.11   
## Median :2008 Median :67.51   
## Mean :2008 Mean :67.92   
## 3rd Qu.:2011 3rd Qu.:68.57   
## Max. :2015 Max. :71.40

The summary statistics are also in tune with the observations of the Plots, with min. and Max. values of Life Expectancy as follows:

Developed : 76.08 to 81.03 Developing : 66.68 to 71.40 Gap b/w the two : approx: 10yrs

### Finding the relationship between Avg Life Expectancy and various predicting variables.

Preparing copies of data in relevant format. - global: containing data for both developed and developing nations - developed\_stats : containing data for developed nations - developing\_stats : containing data for developing nations

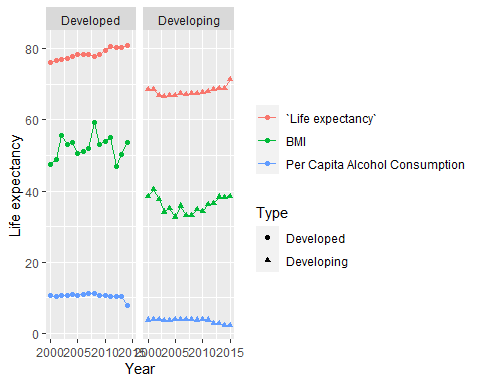
global <- data\_global\_trend  
developed\_stats<- data\_global\_trend %>% filter (Status == "Developed")  
developing\_stats<- data\_global\_trend %>% filter (Status == "Developing")

#### Finding the relationship between Avg. Life Expectancy and Social factors like BMI and Drinking Alcohol.

Plotting the required visualization.

global %>%  
 ggplot(aes(x = Year, y =`Life expectancy` )) +  
 geom\_line(aes(color = "`Life expectancy`")) +  
 geom\_point(aes(color = "`Life expectancy`", shape = global$Status)) +  
 geom\_line(aes(y = Alcohol, color = "Per Capita Alcohol Consumption")) +  
 geom\_point(aes(y = Alcohol, color = "Per Capita Alcohol Consumption", shape= global$Status)) +  
 geom\_line(aes(y = BMI, color = "BMI")) +  
 geom\_point(aes(y = BMI, color = "BMI", shape = global$Status)) +  
 labs(shape = "Type", color = NULL) +  
 facet\_wrap(~Status)

## Warning: Use of `global$Status` is discouraged. Use `Status` instead.  
  
## Warning: Use of `global$Status` is discouraged. Use `Status` instead.  
  
## Warning: Use of `global$Status` is discouraged. Use `Status` instead.



theme(legend.position="bottom",  
 axis.text.x = element\_text(angle = 90)) +  
 labs(title = "Life Expectancy Vs Per Capita Alcohol consumption (in litres)", y= NULL)

## List of 4  
## $ axis.text.x :List of 11  
## ..$ family : NULL  
## ..$ face : NULL  
## ..$ colour : NULL  
## ..$ size : NULL  
## ..$ hjust : NULL  
## ..$ vjust : NULL  
## ..$ angle : num 90  
## ..$ lineheight : NULL  
## ..$ margin : NULL  
## ..$ debug : NULL  
## ..$ inherit.blank: logi FALSE  
## ..- attr(\*, "class")= chr [1:2] "element\_text" "element"  
## $ legend.position: chr "bottom"  
## $ y : NULL  
## $ title : chr "Life Expectancy Vs Per Capita Alcohol consumption (in litres)"  
## - attr(\*, "class")= chr [1:2] "theme" "gg"  
## - attr(\*, "complete")= logi FALSE  
## - attr(\*, "validate")= logi TRUE

**Analyzing the above Plots** :

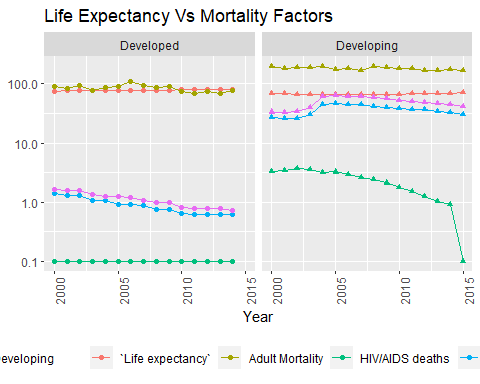
The plots depict the trend of Life Expectancy against lifestyle factors of BMI and alcohol consumption. As per the Plots, - Developed countries have a more unhealthy Lifestyle as indicated by higher BMI and Higher per capita alcohol consumption.This may be partially true because of higher per capita income of developed countries thus more expenditure on drinks and food. But, Despite leading a more unhealthy lifestyle, Developed countries have a higher Life Expectancy as compared to developing countries. This suggest that lifestyle factors like alcohol consumption , BMI are not playing a major role in influencing the Life Expectancy. This is a little counter-intuitive and needs more detailed analysis from diverse datasets. - Another good observation is that with decrease in per capital alcohol consumption, there seems to be a general increase in the Avg. Life Expectancy for both developed and developing clusters. - The BMI data is very haphazard and any direct correlation with Life Expectancy seems to be a haste.

#### Life Expectancy Vs Mortality Factors : Adult Mortality, Infant deaths, under 5 deaths and deaths due to HIV/AIDS

Plotting the required visualization.

global %>%   
 ggplot(aes(x = Year, y =`Life expectancy` )) +  
 geom\_line(aes(color = "`Life expectancy`" )) +  
 geom\_point(aes(color = "`Life expectancy`", shape = global$Status)) +  
 geom\_line(aes(y = `Adult Mortality`, color = "Adult Mortality")) +  
 geom\_point(aes(y = `Adult Mortality`, color = "Adult Mortality", shape= global$Status)) +  
 geom\_line(aes(y = `infant deaths`, color = "infant deaths")) +  
 geom\_point(aes(y = `infant deaths`, color = "infant deaths", shape = global$Status)) +  
 geom\_line(aes(y = `under-five deaths`, color = "under-five deaths")) +  
 geom\_point(aes(y = `under-five deaths`, color = "under-five deaths", shape = global$Status)) +  
 geom\_line(aes(y = `HIV/AIDS`, color = "HIV/AIDS deaths")) +  
 geom\_point(aes(y = `HIV/AIDS`, color = "HIV/AIDS deaths", shape = global$Status)) +  
 scale\_y\_log10() +  
 facet\_wrap(~Status) +  
 labs(shape = "Type", color = NULL) +  
 theme(legend.position="bottom",  
 axis.text.x = element\_text(angle = 90)) +  
 labs(title = "Life Expectancy Vs Mortality Factors", y= NULL)

## Warning: Use of `global$Status` is discouraged. Use `Status` instead.  
  
## Warning: Use of `global$Status` is discouraged. Use `Status` instead.  
  
## Warning: Use of `global$Status` is discouraged. Use `Status` instead.  
  
## Warning: Use of `global$Status` is discouraged. Use `Status` instead.  
  
## Warning: Use of `global$Status` is discouraged. Use `Status` instead.



**Analysing the above Plots**

-As can be seen from the plots, Developing nations have much higher mortality rates as compared to developed countries. Hence, all three mortality rates(Adult, infant and under-five) have a negative correlation with Life Expectancy. - The infant mortality rates and under-5-mortality rates are almost parallel curves suggesting similar factors like poor nutrition,level of vaccination, neonatal care etc. are at play. -The deaths due to HIV/AIDS are also significantly higher in developing countries resulting into higher overall mortality rates in them. But the Deaths due to HIV/AIDS are showing a decreasing trend with timein developing nations which can be one of the reasons for increasing Life expectancy with time.

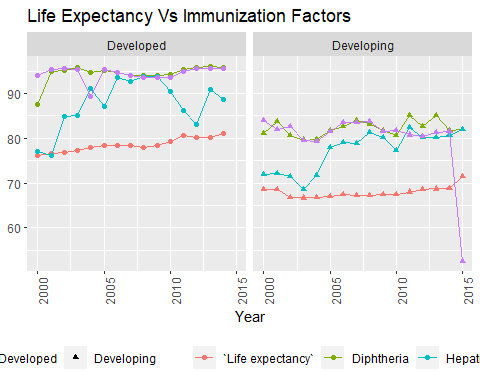
#### Average Life Expectancy Vs Immunization Factors.

Considering immunization of 1year old child as the benchmark and vaccinations of Hepatitis B, Polio and Diphtheria as index.

Plotting the required visualization.

global %>%   
 ggplot(aes(x = Year, y =`Life expectancy` )) +  
 geom\_line(aes(color = "`Life expectancy`" )) +  
 geom\_point(aes(color = "`Life expectancy`", shape = global$Status)) +  
 geom\_line(aes(y = `Diphtheria`, color = "Diphtheria")) +  
 geom\_point(aes(y = `Diphtheria`, color = "Diphtheria", shape= global$Status)) +  
 geom\_line(aes(y = `Hepatitis B`, color = "Hepatitis B")) +  
 geom\_point(aes(y = `Hepatitis B`, color = "Hepatitis B", shape = global$Status)) +  
 geom\_line(aes(y = `Polio`, color = "Polio")) +  
 geom\_point(aes(y = `Polio`, color = "Polio", shape = global$Status)) +  
 facet\_wrap(~Status) +  
 labs(shape = "Type", color = NULL) +  
 theme(legend.position="bottom",  
 axis.text.x = element\_text(angle = 90)) +  
 labs(title = "Life Expectancy Vs Immunization Factors", y= NULL)

## Warning: Use of `global$Status` is discouraged. Use `Status` instead.  
  
## Warning: Use of `global$Status` is discouraged. Use `Status` instead.  
  
## Warning: Use of `global$Status` is discouraged. Use `Status` instead.  
  
## Warning: Use of `global$Status` is discouraged. Use `Status` instead.

 **Analysis of the above plots** .

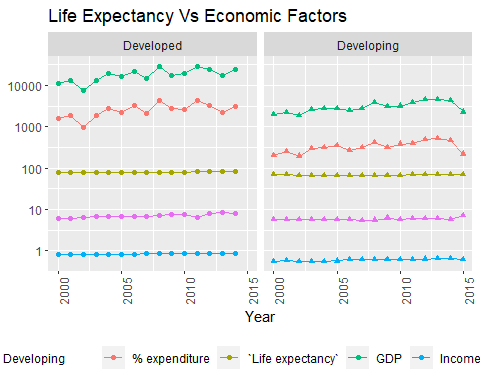
* The above curve does not show any definite trend, specially in the data of Hepatitis-B vaccination. Hence no direct corelation with the LIfe expectancy is clear.
* But one good observation is that overall there is higher level of immunization in the developed countries as compared to developing one, which can be a factor for higher Life expectancy of developed world.

#### Average Life Expectancy Vs Economic Factors.

Looking at the various economic factors like % expenditure, Total Expenditure, GDP of the nation and income composition

global %>%   
 ggplot(aes(x = Year, y =`Life expectancy` )) +  
 geom\_line(aes(color = "`Life expectancy`" )) +  
 geom\_point(aes(color = "`Life expectancy`", shape = global$Status)) +  
 geom\_line(aes(y = `percentage expenditure`, color = "% expenditure")) +  
 geom\_point(aes(y = `percentage expenditure`, color = "% expenditure", shape= global$Status)) +  
 geom\_line(aes(y = `Total expenditure`, color = "Total expenditure")) +  
 geom\_point(aes(y = `Total expenditure`, color = "Total expenditure", shape = global$Status)) +  
 geom\_line(aes(y = `Income composition of resources`, color = "Income composition")) +  
 geom\_point(aes(y = `Income composition of resources`, color = "Income composition", shape = global$Status)) +  
 geom\_line(aes(y = GDP, color = "GDP")) +  
 geom\_point(aes(y = GDP, color = "GDP", shape= global$Status)) +  
 facet\_wrap(~Status) +  
 scale\_y\_log10() +  
 labs(shape = "Type", color = NULL) +  
 theme(legend.position="bottom",  
 axis.text.x = element\_text(angle = 90)) +  
 labs(title = "Life Expectancy Vs Economic Factors", y= NULL)

## Warning: Use of `global$Status` is discouraged. Use `Status` instead.  
  
## Warning: Use of `global$Status` is discouraged. Use `Status` instead.  
  
## Warning: Use of `global$Status` is discouraged. Use `Status` instead.  
  
## Warning: Use of `global$Status` is discouraged. Use `Status` instead.  
  
## Warning: Use of `global$Status` is discouraged. Use `Status` instead.



**Analysing the above Plots.**

As can be seen from the graphs :

* All the economic factors have a positive correlation with the Life expectancy. This suggests better the fiscal health of the nation –> better is the expenditure on health infrastructure and services –> better is the Life expectancy.
* Also, as expected developed nations have stronger fiscal health as compared to developing nations.

### Global Life Expectancy Analysis : Country Wise

Preparing copies of data in relevant format. - data\_countries : country-wise avg statistics over the period of 2000 to 2015 - d2 : a copy of data\_countries with a shorter name.

data\_countries <- data\_clean %>%group\_by(Country,Status)%>% select(-c(Year))%>% summarise\_each(funs( mean(.,na.rm = TRUE)))%>%ungroup()  
ungroup(data\_clean)

## # A tibble: 1,649 x 22  
## Country Year Status `Life expectanc~ `Adult Mortalit~ `infant deaths`  
## <chr> <dbl> <chr> <dbl> <dbl> <dbl>  
## 1 Afghanistan 2015 Developing 65 263 62  
## 2 Afghanistan 2014 Developing 59.9 271 64  
## 3 Afghanistan 2013 Developing 59.9 268 66  
## 4 Afghanistan 2012 Developing 59.5 272 69  
## 5 Afghanistan 2011 Developing 59.2 275 71  
## 6 Afghanistan 2010 Developing 58.8 279 74  
## 7 Afghanistan 2009 Developing 58.6 281 77  
## 8 Afghanistan 2008 Developing 58.1 287 80  
## 9 Afghanistan 2007 Developing 57.5 295 82  
## 10 Afghanistan 2006 Developing 57.3 295 84  
## # ... with 1,639 more rows, and 16 more variables: Alcohol <dbl>,  
## # percentage expenditure <dbl>, Hepatitis B <dbl>, Measles <dbl>, BMI <dbl>,  
## # under-five deaths <dbl>, Polio <dbl>, Total expenditure <dbl>,  
## # Diphtheria <dbl>, HIV/AIDS <dbl>, GDP <dbl>, Population <dbl>,  
## # thinness 1-19 years <dbl>, thinness 5-9 years <dbl>,  
## # Income composition of resources <dbl>, Schooling <dbl>

print(data\_countries)

## # A tibble: 133 x 21  
## Country Status `Life expectanc~ `Adult Mortalit~ `infant deaths` Alcohol  
## <chr> <chr> <dbl> <dbl> <dbl> <dbl>  
## 1 Afghanistan Develo~ 58.2 269. 78.2 0.0144  
## 2 Albania Develo~ 75.2 45.1 0.688 4.85   
## 3 Algeria Develo~ 74.2 103. 20.3 0.447   
## 4 Angola Develo~ 50.7 363. 76.6 7.62   
## 5 Argentina Develo~ 75.2 100. 10 8.00   
## 6 Armenia Develo~ 73.3 117. 1 3.70   
## 7 Australia Develo~ 81.9 62.4 1 10.2   
## 8 Austria Develo~ 81.5 65.8 0 12.2   
## 9 Azerbaijan Develo~ 71.1 120. 5.77 1.06   
## 10 Bangladesh Develo~ 70.0 136. 142. 0.01   
## # ... with 123 more rows, and 15 more variables: percentage expenditure <dbl>,  
## # Hepatitis B <dbl>, Measles <dbl>, BMI <dbl>, under-five deaths <dbl>,  
## # Polio <dbl>, Total expenditure <dbl>, Diphtheria <dbl>, HIV/AIDS <dbl>,  
## # GDP <dbl>, Population <dbl>, thinness 1-19 years <dbl>,  
## # thinness 5-9 years <dbl>, Income composition of resources <dbl>,  
## # Schooling <dbl>

#### Global Average Life Expectancy.

Finding the mean life expectancy for all the people across globe, averaged over the period of 2000-2015

The finding is : 68.62 years

d2<- data\_countries  
global\_avg\_LE = mean(d2$`Life expectancy`)   
print(global\_avg\_LE)

## [1] 68.62162

#### Top 10 countries with **highest Life expectancy** across globe, averaged over the period of 2000-2015

d2%>% slice\_max(`Life expectancy`, n = 10)

## # A tibble: 10 x 21  
## Country Status `Life expectanc~ `Adult Mortalit~ `infant deaths` Alcohol  
## <chr> <chr> <dbl> <dbl> <dbl> <dbl>  
## 1 Ireland Develo~ 83.4 55.4 0 11.3   
## 2 Canada Develo~ 82.2 66.8 2 8.16  
## 3 France Develo~ 82.2 72.8 3 12.4   
## 4 Italy Develo~ 82.2 54.1 2.07 8.27  
## 5 Spain Develo~ 82.0 64.1 1.67 10.0   
## 6 Australia Develo~ 81.9 62.4 1 10.2   
## 7 Sweden Develo~ 81.9 56.5 0 7.35  
## 8 Austria Develo~ 81.5 65.8 0 12.2   
## 9 Netherlands Develo~ 81.3 47.2 1 6.68  
## 10 Greece Develo~ 81.2 73.7 0.2 8.80  
## # ... with 15 more variables: percentage expenditure <dbl>, Hepatitis B <dbl>,  
## # Measles <dbl>, BMI <dbl>, under-five deaths <dbl>, Polio <dbl>,  
## # Total expenditure <dbl>, Diphtheria <dbl>, HIV/AIDS <dbl>, GDP <dbl>,  
## # Population <dbl>, thinness 1-19 years <dbl>, thinness 5-9 years <dbl>,  
## # Income composition of resources <dbl>, Schooling <dbl>

As can be seen from the above table of Highest Life expectancy, Ireland tops the chart with avg. Life expectancy of about 83 years.Moreover, all the 10 nations in the list are from the first world nations which are either already well developed or just about to enter the status of being developed. Apart from this , all 10 nations have Avg. Life expectancy over 80yrs, which is way more higher than the global avg. of 68.62 avy across globe over the same time period of 2000 -2015

#### Top 10 countries with the **least Life expectancy** across globe, averaged over the period of 2000-2015

d2%>% slice\_min(`Life expectancy`, n = 10)

## # A tibble: 10 x 21  
## Country Status `Life expectanc~ `Adult Mortalit~ `infant deaths` Alcohol  
## <chr> <chr> <dbl> <dbl> <dbl> <dbl>  
## 1 Sierra Leo~ Develo~ 48.4 335. 26.2 2.42   
## 2 Lesotho Develo~ 48.6 547. 4.42 1.87   
## 3 Zimbabwe Develo~ 49.4 471. 26.9 4.48   
## 4 Malawi Develo~ 50.3 403 36.1 0.882  
## 5 Angola Develo~ 50.7 363. 76.6 7.62   
## 6 Swaziland Develo~ 50.8 337. 2.67 4.46   
## 7 Central Af~ Develo~ 51.4 445. 16.2 0.82   
## 8 Chad Develo~ 52.3 322. 46 0.464  
## 9 Nigeria Develo~ 52.8 340. 524. 8.18   
## 10 Mozambique Develo~ 53.4 230. 74.3 1.27   
## # ... with 15 more variables: percentage expenditure <dbl>, Hepatitis B <dbl>,  
## # Measles <dbl>, BMI <dbl>, under-five deaths <dbl>, Polio <dbl>,  
## # Total expenditure <dbl>, Diphtheria <dbl>, HIV/AIDS <dbl>, GDP <dbl>,  
## # Population <dbl>, thinness 1-19 years <dbl>, thinness 5-9 years <dbl>,  
## # Income composition of resources <dbl>, Schooling <dbl>

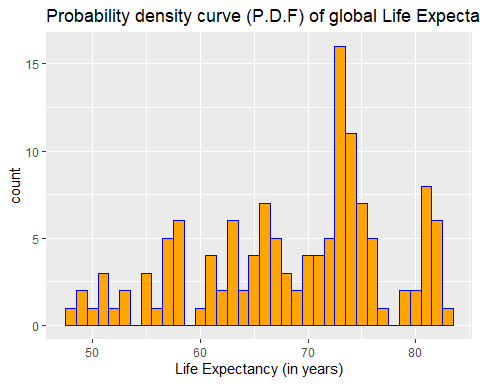
As can be seen from the table above :-

* Sierra Leone tops the chart with Life Expectancy as low as 48yrs, which is almost 20yrs lower than the global avg of 68 yrs.
* All the 10 countries are developing nations and all belong to the continent of Africa.

#### **Probability density function** of global Avg. Life Expectancies :

Ploting the pdf curve :

p<-ggplot(d2, aes(x=`Life expectancy`)) +   
 geom\_histogram(binwidth=1, color="Blue", fill="orange") +  
 labs( title = "Probability density curve (P.D.F) of global Life Expectancy" , x = "Life Expectancy (in years)")   
print(p)



According to the Probability density function (pdf) of Avg Life expectancy of different countries over the period of 2000-2015, one can conclude that the most likely average global expected Life expectancy is about **73 year**. Also, the highest Life expectancy across globe is 83yrs which is about 10 years more than the modal value of 73ys.

### Country-wise comparison of top performers : in terms of LIfe Expectancy

Preparing copies of data in relevant format. - d3 : a copy of data\_clean with a shorter name. - d3\_2005 : country wise data for the year 2005 - d3\_2014 : country wise data for the year 2014

d3<-data\_clean  
d3\_2005<-d3 %>% filter(Year == 2005 )  
d3\_2014<-d3 %>% filter(Year == 2014 )  
  
print(d3)

## # A tibble: 1,649 x 22  
## Country Year Status `Life expectanc~ `Adult Mortalit~ `infant deaths`  
## <chr> <dbl> <chr> <dbl> <dbl> <dbl>  
## 1 Afghanistan 2015 Developing 65 263 62  
## 2 Afghanistan 2014 Developing 59.9 271 64  
## 3 Afghanistan 2013 Developing 59.9 268 66  
## 4 Afghanistan 2012 Developing 59.5 272 69  
## 5 Afghanistan 2011 Developing 59.2 275 71  
## 6 Afghanistan 2010 Developing 58.8 279 74  
## 7 Afghanistan 2009 Developing 58.6 281 77  
## 8 Afghanistan 2008 Developing 58.1 287 80  
## 9 Afghanistan 2007 Developing 57.5 295 82  
## 10 Afghanistan 2006 Developing 57.3 295 84  
## # ... with 1,639 more rows, and 16 more variables: Alcohol <dbl>,  
## # percentage expenditure <dbl>, Hepatitis B <dbl>, Measles <dbl>, BMI <dbl>,  
## # under-five deaths <dbl>, Polio <dbl>, Total expenditure <dbl>,  
## # Diphtheria <dbl>, HIV/AIDS <dbl>, GDP <dbl>, Population <dbl>,  
## # thinness 1-19 years <dbl>, thinness 5-9 years <dbl>,  
## # Income composition of resources <dbl>, Schooling <dbl>

print(d3\_2005)

## # A tibble: 110 x 22  
## Country Year Status `Life expectanc~ `Adult Mortalit~ `infant deaths`  
## <chr> <dbl> <chr> <dbl> <dbl> <dbl>  
## 1 Afghanistan 2005 Developing 57.3 291 85  
## 2 Albania 2005 Developing 73.5 15 1  
## 3 Algeria 2005 Developing 72.9 136 19  
## 4 Argentina 2005 Developing 74.9 127 11  
## 5 Armenia 2005 Developing 73 137 1  
## 6 Australia 2005 Developed 81 67 1  
## 7 Austria 2005 Developed 79.4 85 0  
## 8 Azerbaijan 2005 Developing 68.4 162 6  
## 9 Bangladesh 2005 Developing 67.8 155 174  
## 10 Belarus 2005 Developing 68.1 252 1  
## # ... with 100 more rows, and 16 more variables: Alcohol <dbl>,  
## # percentage expenditure <dbl>, Hepatitis B <dbl>, Measles <dbl>, BMI <dbl>,  
## # under-five deaths <dbl>, Polio <dbl>, Total expenditure <dbl>,  
## # Diphtheria <dbl>, HIV/AIDS <dbl>, GDP <dbl>, Population <dbl>,  
## # thinness 1-19 years <dbl>, thinness 5-9 years <dbl>,  
## # Income composition of resources <dbl>, Schooling <dbl>

print(d3\_2014)

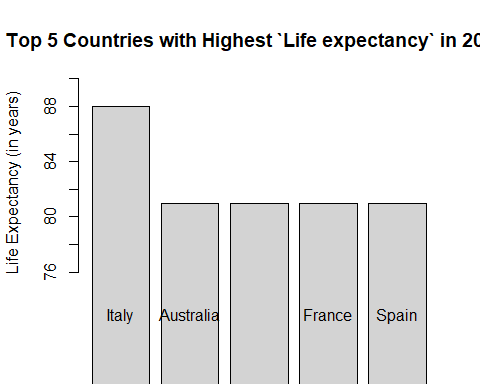
## # A tibble: 131 x 22  
## Country Year Status `Life expectanc~ `Adult Mortalit~ `infant deaths`  
## <chr> <dbl> <chr> <dbl> <dbl> <dbl>  
## 1 Afghanistan 2014 Developing 59.9 271 64  
## 2 Albania 2014 Developing 77.5 8 0  
## 3 Algeria 2014 Developing 75.4 11 21  
## 4 Angola 2014 Developing 51.7 348 67  
## 5 Argentina 2014 Developing 76.2 118 8  
## 6 Armenia 2014 Developing 74.6 12 1  
## 7 Australia 2014 Developed 82.7 6 1  
## 8 Austria 2014 Developed 81.4 66 0  
## 9 Azerbaijan 2014 Developing 72.5 119 5  
## 10 Bangladesh 2014 Developing 71.4 132 98  
## # ... with 121 more rows, and 16 more variables: Alcohol <dbl>,  
## # percentage expenditure <dbl>, Hepatitis B <dbl>, Measles <dbl>, BMI <dbl>,  
## # under-five deaths <dbl>, Polio <dbl>, Total expenditure <dbl>,  
## # Diphtheria <dbl>, HIV/AIDS <dbl>, GDP <dbl>, Population <dbl>,  
## # thinness 1-19 years <dbl>, thinness 5-9 years <dbl>,  
## # Income composition of resources <dbl>, Schooling <dbl>

#### Plotting the best 5 performers of the year 2005

Best\_2005<- d3\_2005 %>% slice\_max(`Life expectancy`, n = 5)  
print(Best\_2005)

## # A tibble: 5 x 22  
## Country Year Status `Life expectanc~ `Adult Mortalit~ `infant deaths` Alcohol  
## <chr> <dbl> <chr> <dbl> <dbl> <dbl> <dbl>  
## 1 Italy 2005 Devel~ 88 66 2 8.65  
## 2 Australia 2005 Devel~ 81 67 1 10.3   
## 3 Canada 2005 Devel~ 81 76 2 8   
## 4 France 2005 Devel~ 81 93 3 12.2   
## 5 Spain 2005 Devel~ 81 77 2 11.9   
## # ... with 15 more variables: percentage expenditure <dbl>, Hepatitis B <dbl>,  
## # Measles <dbl>, BMI <dbl>, under-five deaths <dbl>, Polio <dbl>,  
## # Total expenditure <dbl>, Diphtheria <dbl>, HIV/AIDS <dbl>, GDP <dbl>,  
## # Population <dbl>, thinness 1-19 years <dbl>, thinness 5-9 years <dbl>,  
## # Income composition of resources <dbl>, Schooling <dbl>

barplot(Best\_2005$`Life expectancy`,  
main = "Top 5 Countries with Highest `Life expectancy` in 2005)",  
ylab = "Life Expectancy (in years)",  
ylim = c(75,90),  
names.arg = c("Italy","Australia","Canada","France","Spain"),  
col = "Light Grey",  
  
horiz = F )

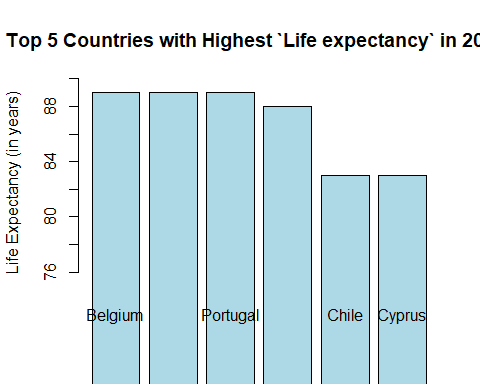


#### Plotting the best 5 performers of the year 2014

Best\_2014<- d3\_2014 %>% slice\_max(`Life expectancy`, n = 5)  
print(Best\_2014)

## # A tibble: 6 x 22  
## Country Year Status `Life expectanc~ `Adult Mortalit~ `infant deaths` Alcohol  
## <chr> <dbl> <chr> <dbl> <dbl> <dbl> <dbl>  
## 1 Belgium 2014 Devel~ 89 76 0 12.6   
## 2 Germany 2014 Devel~ 89 69 2 11.0   
## 3 Portugal 2014 Devel~ 89 78 0 9.88  
## 4 Greece 2014 Devel~ 88 73 0 7.53  
## 5 Chile 2014 Devel~ 83 83 2 7.16  
## 6 Cyprus 2014 Devel~ 83 53 0 0.01  
## # ... with 15 more variables: percentage expenditure <dbl>, Hepatitis B <dbl>,  
## # Measles <dbl>, BMI <dbl>, under-five deaths <dbl>, Polio <dbl>,  
## # Total expenditure <dbl>, Diphtheria <dbl>, HIV/AIDS <dbl>, GDP <dbl>,  
## # Population <dbl>, thinness 1-19 years <dbl>, thinness 5-9 years <dbl>,  
## # Income composition of resources <dbl>, Schooling <dbl>

barplot(Best\_2014$`Life expectancy`,  
main = "Top 5 Countries with Highest `Life expectancy` in 2014)",  
ylab = "Life Expectancy (in years)",  
ylim = c(75,90),  
names.arg = c("Belgium","Germany","Portugal","Greece","Chile", "Cyprus"),  
col = "Light Blue",  
  
horiz = F )



* As can be seen from both the plots, the top performers are still from the first world nations , though the toppers of the past have failed to retain their positions in the chart.
* Also, as can be seen from the graphs, the Life expectancy value for the top performers is higher in 2014 as compared to 2005, which suggest that world as a whole is moving towards longevity of life.

## Linear Model to predict Life expectancy against the parameter of schooling.

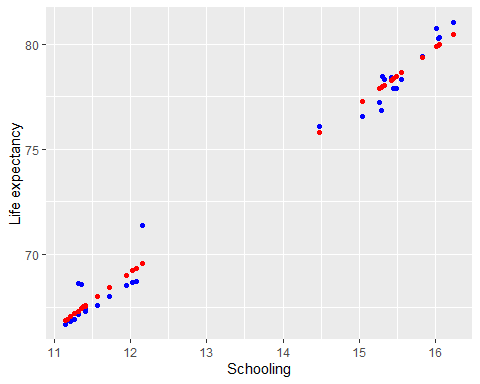
mod <- lm(`Life expectancy` ~ Schooling, data = global)  
summary(mod)

##   
## Call:  
## lm(formula = `Life expectancy` ~ Schooling, data = global)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.0657 -0.4495 -0.1901 0.3199 1.8481   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 37.16285 0.78237 47.50 <2e-16 \*\*\*  
## Schooling 2.66577 0.05747 46.38 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.6511 on 29 degrees of freedom  
## Multiple R-squared: 0.9867, Adjusted R-squared: 0.9862   
## F-statistic: 2152 on 1 and 29 DF, p-value: < 2.2e-16

x\_grid <- seq(1, 151)  
new\_df <- tibble(Schooling = x\_grid)  
global %>% mutate(pred = predict(mod))

## # A tibble: 31 x 22  
## Year Status `Life expectancy` `Adult Mortality` `infant deaths` Alcohol  
## <dbl> <chr> <dbl> <dbl> <dbl> <dbl>  
## 1 2000 Developed 76.1 91.7 1.38 10.6   
## 2 2000 Developing 68.6 197. 26.7 3.76  
## 3 2001 Developed 76.6 84.9 1.29 10.4   
## 4 2001 Developing 68.6 181. 25.8 3.85  
## 5 2002 Developed 76.9 94.9 1.29 10.7   
## 6 2002 Developing 66.9 188. 25.7 3.81  
## 7 2003 Developed 77.2 78.9 1.07 10.6   
## 8 2003 Developing 66.7 191. 30.2 3.59  
## 9 2004 Developed 77.9 87.8 1.07 10.9   
## 10 2004 Developing 66.8 201. 45.7 3.67  
## # ... with 21 more rows, and 16 more variables: percentage expenditure <dbl>,  
## # Hepatitis B <dbl>, Measles <dbl>, BMI <dbl>, under-five deaths <dbl>,  
## # Polio <dbl>, Total expenditure <dbl>, Diphtheria <dbl>, HIV/AIDS <dbl>,  
## # GDP <dbl>, Population <dbl>, thinness 1-19 years <dbl>,  
## # thinness 5-9 years <dbl>, Income composition of resources <dbl>,  
## # Schooling <dbl>, pred <dbl>

LE\_pred <- global %>% mutate(pred = predict(mod))  
LE\_pred %>% ggplot() +  
geom\_point(aes(x = Schooling, y = `Life expectancy`), color = "blue") +  
geom\_point(aes(x = Schooling, y = pred), color = "red")



As can be seen from the above graph : - the schooling has a positive correlation with LIfe Expectancy as the curve has a positive slope. - As both red and blue dots are majorly close, it suggests that linear model to find the correlation and prediction of Life expectancy is quite good.

# **CONCLUSION**

Avg. Life Expectancy of humans is not a naturally fixed constant but a variable being shaped by various anthropogenic factors too such as Lifestyle habits, spending on Health infrastructure and services, better immunization drives as well as other socio-economic and political factors. Hence, in order to lead a better and more longer life, humans need to define their public policy keeping various factors in mind as analysed above. Hence, with better policy and practices on the line of developed world , Humanity can lead a better, healthier and longer life.