**Exploratory Data Analysis**

Jessie Le, Leon Zhou, Calvin Ochi

# **Introduction:**

Life expectancy at birth, defined by the United Nations Human Development Report as “the years a newborn infant would live if prevailing patterns of age-specific mortality rates at the time of birth were to stay the same throughout the infant’s life” (Human Development Report, 2019), is a common indicator of national development and quality of life for the average citizen. Life expectancy has grown progressively since the 1990’s, with a drastic increase during the Millennium Development Goals (MDGs) era of 2000-2015 (WHO, 2015). Despite this, variability among high- and low-income countries persists. Learning more about what factors influence life expectancy may help inform policy decision-making and to effectively reallocate resources.

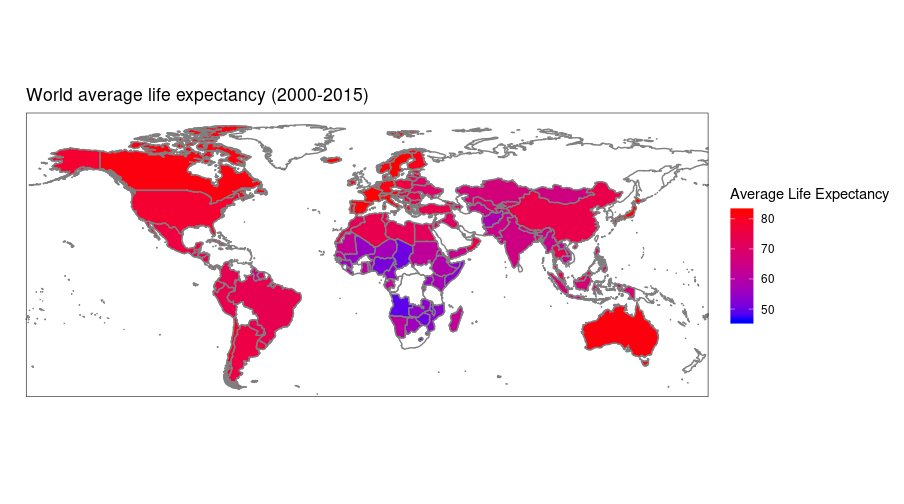
There has been extensive research on the effects of socioeconomic status on life expectancy, but little research has conducted combined analysis of demographic, socioeconomic, and healthcare-related factors across all countries in recent decades. Our investigation attempts to integrate significant factors from previous studies to build explanatory model that would describe important factors and their contributions to one country’s life expectancy. Previous studies show that socioeconomic determinants include Gross Domestic Product (GDP) per capita, percent expenditure on health, and education, represented by years of schooling (Lin et al 2012 and Kabir 2008). Moreover, life expectancy can be attributed to health factors including alcohol consumption, percent immunization, BMI, and HIV/AIDS prevalence (Mondal and Shinta 2014 and Khan et al 2016). Regional factors also contribute to the differences in life expectancy (Kabir 2008 and Kpolovie 2016). This present study hopes to determine the significance of the above variables after controlling multiple socioeconomic and health factors.

# **Data management**

# There are ten countries (Table 1) that are included in the original data set but do not have data on life expectancy or other variables. Thus, we removed these countries from our analysis, and 183 countries remain in our dataset. We did not have data indicating each country with its respective continents. So we added a continent variable, which includes 5 continents Asia, Africa, Europe, Oceana, and the Americas.

1. **Exploratory Data Analysis**

***Baseline variables EDA :***

Accounting for all observations between 2000-2015, the mean life expectancy is 69.2 years with a standard deviation of 9.5 years (n = 2938). For life expectancy of each country from 2000 to 2015, Japan has the highest average life expectancy (82.5). Meanwhile, Sierra Leone has the lowest average life expectancy (46.1). When considering all the available observations, years have a positive relationship with life expectancy: as year increases from 2000 to 2015, the life expectancy increases as well. However, this correlation does not seem to be strong (cor=0.17). When taking the average life expectancy of each country (193 observations), the relationship between years and average life expectancy seems to be stronger (cor=0.995). For country status, developed countries have a higher average life expectancy (mean=79.2) and less variation (sd=3.9), and developing countries have lower average life expectancy (mean=67.1) and more variation (sd=9.0). Based on PlotA6, TableA10, and Map 1, when looking at continental differences in life expectancy, Africa has the lowest life expectancy (mean=58.6) and the largest variation (sd=7.3) while Europe has the highest life expectancy (mean=77.4). Based on this, continental differences may be strong potential predictors for life expectancy. 

## 

## ***Healthcare factors:***

Among healthcare-related variables, there is a universal separation between developed and developing countries. Moreover, the effects of each variable differs between developed and developing countries, which indicates that country status is an important controlling variable and interacts with healthcare factors. When comparing healthcare effects with life expectancy between 2000 and 2015, there seems to be no significant change in the effects of each healthcare-related variable on life expectancy. Furthermore, there is a moderate correlation between life expectancy with alcohol consumption (0.403), BMI (0.542), and deaths per 1,000 HIV/AIDS live births (-0.592). The other variables do not display a strong correlation with life expectancy (correlation coefficients are less than 0.35). On the other hand, percent Hepatitis B, Polio, and Diphtheria immunizations among 1-year-olds show relatively strong correlation between one another (Table B2).

Although there are differences in baselines of percent Hepatitis B immunization, and measles cases between developed and developing countries, Hepatitis B and measles do not show any obvious effect on life expectancy, after controlling for country status.

Percent Polio and Diphtheria immunization among 1-year-olds show moderate positive relationships with life expectancy. Alcohol consumption, BMI, and number of deaths per 1,000 HIV/AIDS live births (0-4 years) show evident relationships with life expectancy, after adjusting for country status. Surprisingly, there is a positive trend between alcohol consumption and life expectancy: as alcohol consumption increases, life expectancy increases. While among developing countries, there is a positive relationship between BMI and life expectancy, the trend is slightly negative among developed countries.

Over the years, the global average number of deaths of 1,000 HIV/AIDS live births has drastically decreased. Nonetheless, deaths per 1,000 HIV/AIDS live births show a similar exponential trend with life expectancy, after adjusting for year. Logarithm transformation of deaths per 1,000 HIV/AIDS live births shows a negative relationship with life expectancy.

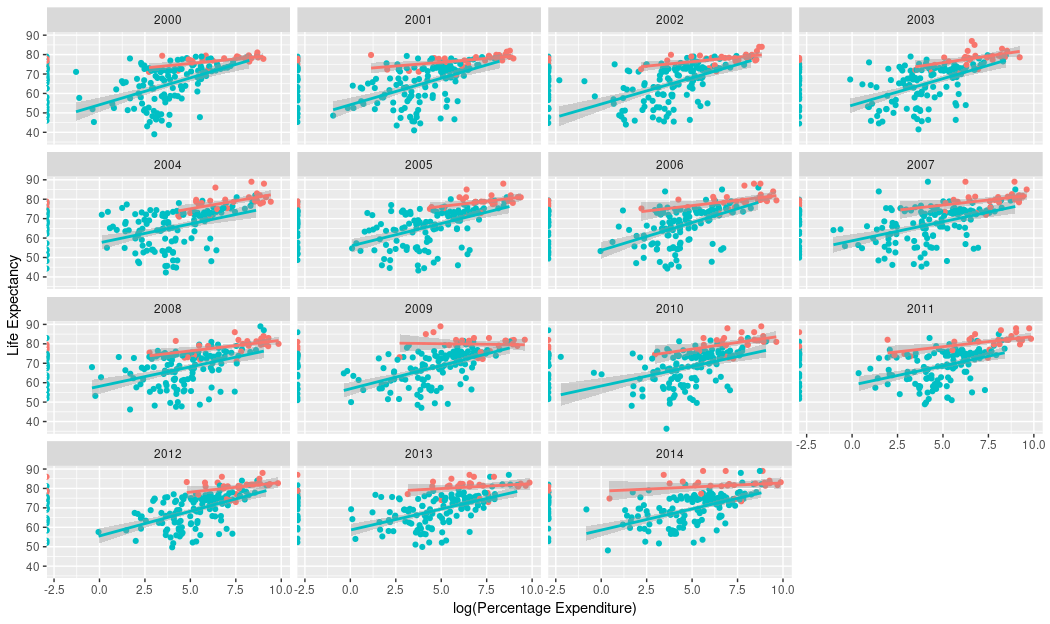
***Socioeconomic factors:***

Among developed countries, between 2000-2015, there is an overall increase in percentage of expenditure on health, but no obvious increase among developing countries. While total health expenditure shows a weak correlation with life expectancy (0.17), the percentage of expenditure on health displays a strong correlation with life expectancy (0.41). There is a positive relationship between the logarithm of percentage expenditure and life expectancy.

The population of a country does not show any evident trend with life expectancy. There is no strong correlation between a country’s population and any other socioeconomic variables.

Developed countries have an overall increase in mean GDP, while GDP in developing countries does not change dramatically. There is greater variability in GDP among developing countries than developed countries. There are moderate correlations between life expectancy and GDP. A country’s GDP has a strong positive correlation with percent expenditure in health (0.959), a moderate correlation with years of schooling (0.468), and income composition of resources (0.447). There is a consistent positive relationship between log(percent expenditure) and life expectancy across different years, with a potential interaction between country status and log(percent expenditure) (**Fig. 1**).

There is a strong correlation between years of schooling and income composition of resources (0.785). Both of these explanatory variables display a noticeable positive relationship with life expectancy, and their effects do not change over the years.



**Figure 1.** Life expectancy versus log(Percentage of Expenditure on Health) across years from 2000-2014.

1. **Modeling preview**

Because of the skewness of the distributions, logarithm transformation is needed for the following explanatory variables: Hepatitis B, Polio, and Diphtheria percent immunization among one-year-olds, number of measles cases, and number of deaths per 1,000 HIV/AIDS live births.

After Tukey Honest Significant Differences test, Africa’s average life expectancy is significantly lower than that of other continents (p<0.05 for all paris). Thus, we would like to relevel continents into two categories: African and non-African continents for modeling.

We would like to begin modeling with preliminary linear models. We expect some assumptions for linear regression to be violated such as independence: each of our observation (country) is not totally independent since there are multiple records for each country for each year.

Since our data set follows a multilevel data structure, we would like to implement multilevel models for our analysis. For level one variables, we have: year, healthcare factors including alcohol, number of measles cases, BMI, percents hepatitis B, polio, diphtheria immunization among one-year-olds, number of deaths per 1,000 HIV/AIDS live births, and socioeconomic factors including expenditure percentage on healthcare, total expenditure, GDP, population, income composition of resources, and number of schooling years. For our level two variables, we have continent and country status. In addition, since one of our level one variable is years (2000-2015), we could potentially use a longitudinal model (repeated measures analysis) approach for modeling life expectancy.

1. **References**

Human Development Report (2019). Human Development Indices and Indicators: 2018 Statistical Update. 15-110.

Kabir, M. (2008). Determinants of Life Expectancy in Developing Countries. *The Journal of Developing Areas, 41(2)*, 185-204.

Khan, A., Khan S., &Khan, M. (2016, November). Factors effecting life expectancy in developed and developing countries of the world (An approach to available literature). *International Journal of Yoga, Physiotherapy and Physical Education, 1(1),* 04-06.

Kpolovie, P. J. , Oshodi, P. O., and Iwuchukwu, H. (December 2016). Continental Inequalities in Life Expectancy. *European Journal of Biology and Medical Science Research, 4(6),* 30-47.

Lin, R., Chen, Y., Chien L., and Chan C. (2012). Political and social determinants of life expectancy in less developed countries: a longitudinal study. *BMC Public Health, 12(85).*

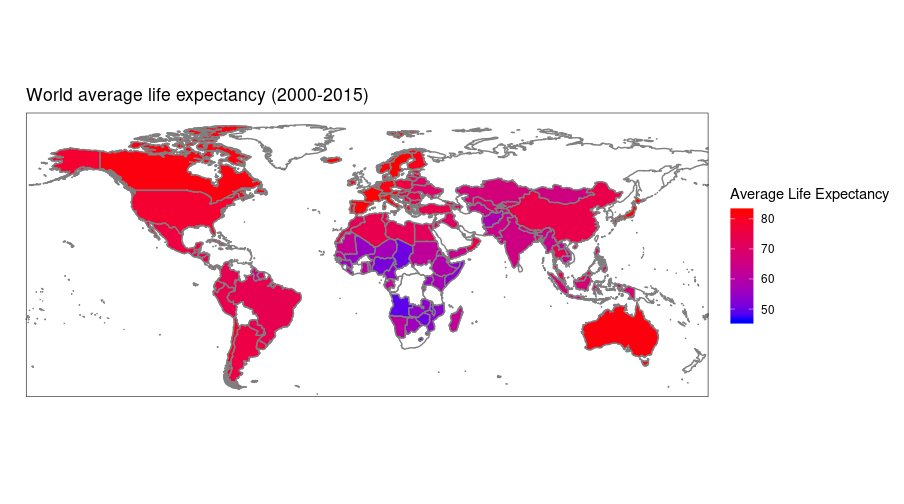
Mondal, M. N. I, and Shitan, M. (2014). Relative Importance of Demographic, Socioeconomic and Health Factors on Life Expectancy in Low- and Lower-Middle-Income Countries. *Journal of Epidemiology, 24(2*), 117-124.

World Health Organization. (2015)‎. Health in 2015: from MDGs, Millennium Development Goals to SDGs, Sustainable Development Goals.

**Appendix:**

| Cook Islands | Dominica | Marshall Islands | Monaco | Nauru | Niue | Palau | St. Kitts and Nevis | San Marino | Tuvalu |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |

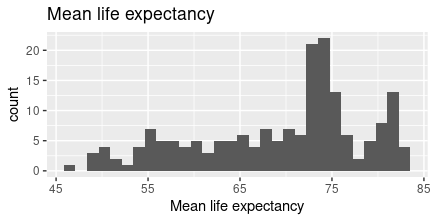
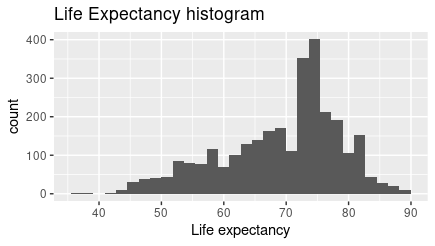
**Table 1**: 10 countries do not have any life expectancy data and are removed from the analysis data.

**Map 1:** World map with average life expectancy from 2000-2015. White area indicates we do not have data on that country’s life expectancy. 

# 

**A - Response Variables/Baseline Variables:**

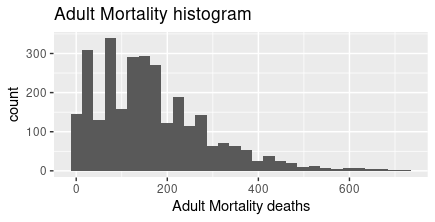
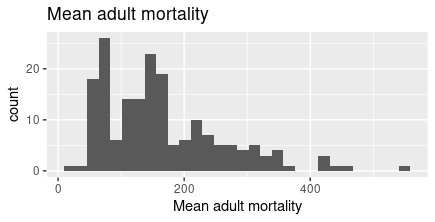
**Life Expectancy:**

**Figure A1.** histograms of the distribution of life expectancy. Left: n=2983. Right: n=193

|  | **Mean** | **Standard Deviation** | **Minimum** | **Maximum** |
| --- | --- | --- | --- | --- |
| **Life expectancy (years)**  **(n=2938)** | 69.2 | 9.5 | 36.3 | 89 |
| **Average life expectancy (years)**  **(n=193)** | 69.2 | 9.2 | 46.1 | 82.5 |

**Table A1.** Summary statistics for life expectancy

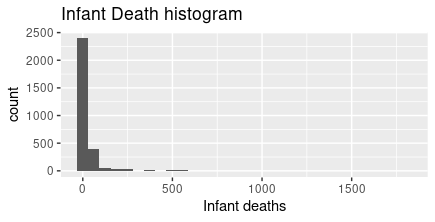
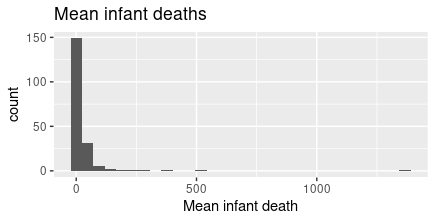
**Adult Mortality:**

**Figure A2.** Histogram of the distribution of adult mortality. Left: n=2983. Right: n=193

|  | **Mean** | **Standard Deviation** | **Minimum** | **Maximum** |
| --- | --- | --- | --- | --- |
| **Adult Mortality**  **(n=2938)** | 164.8 | 124.3 | 1.0 | 723 |
| **Average adult mortality**  **(n=193)** | 164.8 | 96.2 | 18.8 | 550.1 |

**Table A2.** Summary statistics for adult mortality

**Infant Death:**

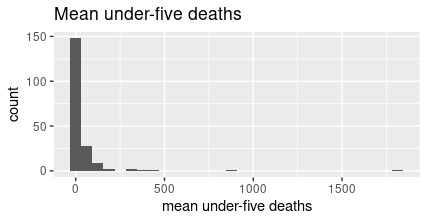
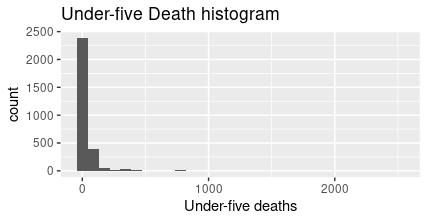


**Figure A3.** Histograms of distribution of infant deaths. Left: n=2983. Right: n=193

|  | **Mean** | **Standard Deviation** | **Minimum** | **Maximum** |
| --- | --- | --- | --- | --- |
| **Infant deaths**  **(n=2938)** | 30.3 | 117.9 | 0 | 1800 |
| **Average infant deaths**  **(n=193)** | 28.8 | 113.1 | 0 | 1366.7 |

**Table A3.** Summary statistics for infant deaths

**Under-five deaths:**

**Figure A4.** histograms of the distribution of under-five deaths. Left: n=2983. Right: n=193

|  | **Mean** | **Standard Deviation** | **Minimum** | **Maximum** |
| --- | --- | --- | --- | --- |
| **Under-five deaths**  **(n=2938)** | 42.0 | 160.4 | 0 | 2500 |
| **Average under-five deaths**  **(n=193)** | 39.9 | 153.6 | 0 | 1812.5 |

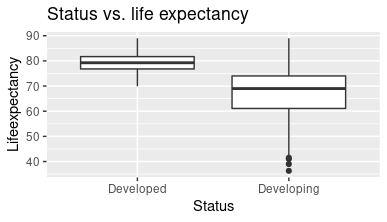
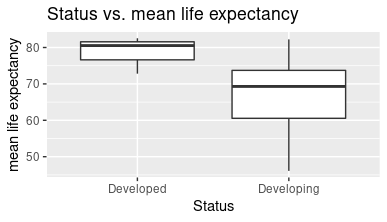
**Table A4.** Summary statistics for infant deaths

**Baseline Variables:**

**Country Status:**

| **Status** | **Developed** | **Developing** | **Total** |
| --- | --- | --- | --- |
| **Counts** | 32 (16.6%) | 161 (83.4%) | 193 |

**Table A5.** Count table for country status with proportion of the total countries.

**Figure A5.** Boxplots of life expectancy between developed and developing countries. Left: n=2983. Right: n=193

**Status vs. life expectancy table (n=2983)**

| Status | **Mean** | **Standard deviation** | **Minimum** | **Maximum** |
| --- | --- | --- | --- | --- |
| **Developed** | 79.2 | 3.9 | 69.9 | 89 |
| **Developing** | 67.1 | 9.0 | 36.3 | 89 |

**Table A6.** Summary statistics of life expectancy between developed and developing countries.

**Status vs. mean life expectancy table (n=193)**

| Status | **Mean** | **Standard deviation** | **Minimum** | **Maximum** |
| --- | --- | --- | --- | --- |
| **Developed** | 79.2 | 3.2 | 72.8 | 82.5 |
| **Developing** | 67.1 | 8.6 | 46.1 | 82.2 |

**Table A7.** Summary statistics of average life expectancy between developed and developing countries.

**Country vs. life expectancy:**

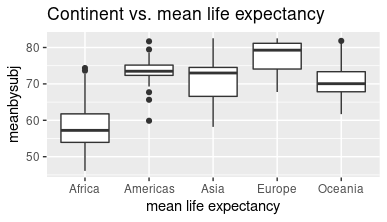
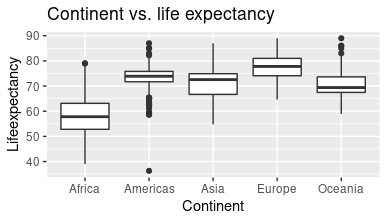
|  | **Country** | **Ave years life expectancy** |
| --- | --- | --- |
| **Max** | Japan | 82.5 |
| **Min** | Sierra Leone | 46.1 |

**Table A8.** Countries with highest life expectancy and lowest life expectancy.

**Continents:**

| **Continent** | **Africa** | **Americas** | **Asia** | **Europe** | **Oceania** | **Total** |
| --- | --- | --- | --- | --- | --- | --- |
| **Counts** | 54 (28.2%) | 35 (18%) | 47 (24.3%) | 41 (21.2%) | 16 (8.3%) | 193 |

**Table A9.** Count table of 5 main continents with proportions.

**Figure A6.** Boxplots of life expectancy between continents. Left: n=2983. Right: n=193. 

Continent vs. life expectancy table (n=2983)

| Continent | **Mean** | **Standard deviation** | **Minimum** | **Maximum** |
| --- | --- | --- | --- | --- |
| **Africa** | 58.6 | 8.0 | 39 | 79 |
| **Americas** | 73.5 | 4.4 | 36.3 | 87 |
| **Asia** | 71.2 | 5.9 | 54.8 | 87 |
| **Europe** | 64.6 | 4.9 | 64.6 | 89 |
| **Oceania** | 71.2 | 6.4 | 58.9 | 89 |

**Table A10.** summary statistics of life expectancy between different continents.

Continent vs. mean life expectancy table (n=193)

| Continent | **Mean** | **Standard Deviation** | **Minimum** | **Maximum** |
| --- | --- | --- | --- | --- |
| **Africa** | 58.6 | 7.2 | 46.1 | 74.4 |
| **Americas** | 73.5 | 4.0 | 59.9 | 81.7 |
| **Asia** | 71.2 | 5.6 | 58.2 | 82.5 |
| **Europe** | 77.4 | 4.3 | 67.8 | 82.5 |
| **Oceania** | 71.2 | 6.5 | 61.7 | 81.8 |

**Table A11.** summary statistics of life expectancy between different continents.

**Year**

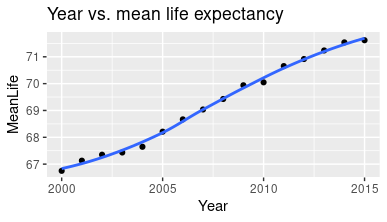
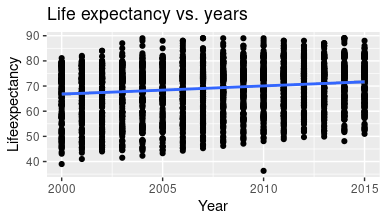
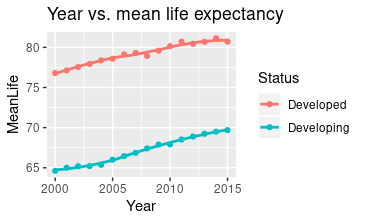
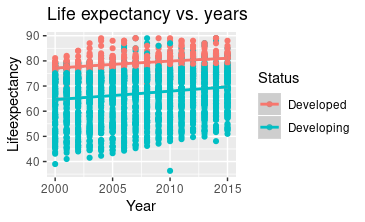
**Figure A7.** Scatterplots of year and life expectancy. Left: n=2983. Right: n=193.

Table of correlation between life expectancy and year

|  | **Year** |
| --- | --- |
| **Life expectancy (n=2983)** | 0.17 |
| **Average life expectancy (n=193)** | 0.995 |

**Table A12.** Table of correlation between years and life expectancy.

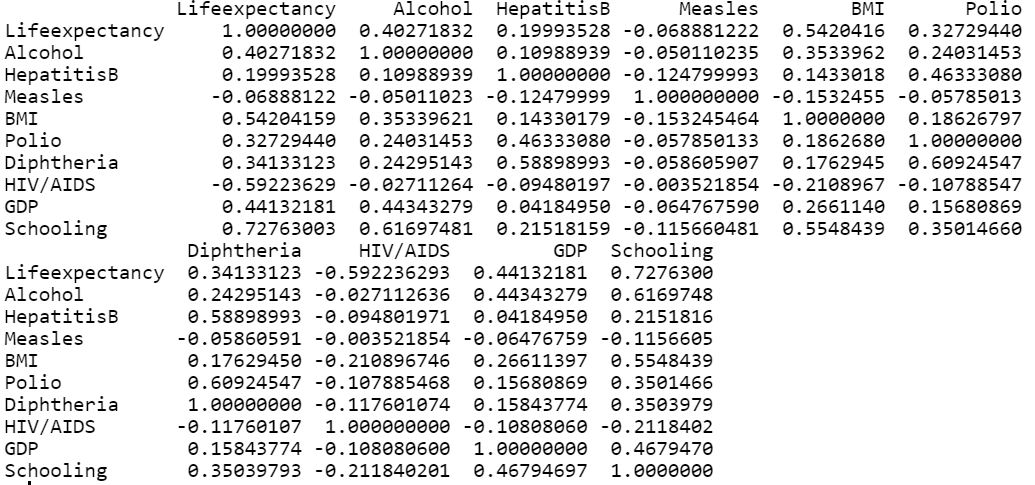
**Figure A8.** Scatterplots of year and life expectancy between developed and developing countries. Left: n=2983. Right: n=193

**B - Health-related Variables:**

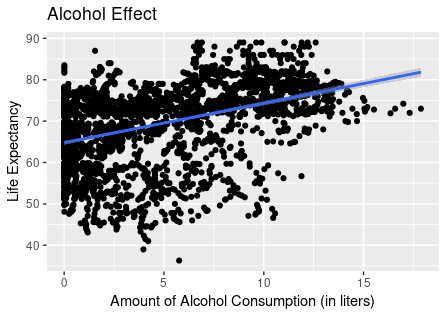
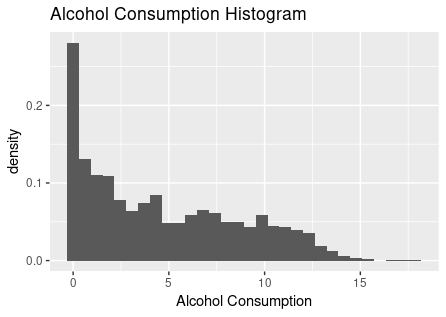
**Table B1.** Statistics Summary Table

| **Variables (unit)** | **Mean** | **Standard Deviation** | **Minimum** | **Maximum** |
| --- | --- | --- | --- | --- |
| Alcohol (liters) | 4.6 | 4.05 | 0.01 | 17.87 |
| Hepatitis B (% immunization among 1-year-old) | 80.9 | 25 | 1 | 99 |
| Measles (cases/1000 population) | 2420 | 11467 | 0 | 212183 |
| BMI | 38.3 | 20.0 | 1 | 87.3 |
| Polio (% immunization among 1-year-olds) | 82.55 | 23.4 | 3 | 99 |
| Diphtheria (% immunization among 1-year-old) | 82.32 | 23.7 | 2 | 99 |
| HIV/AIDS (deaths per 1,000 live births HIV/AIDS (0-4 years)) | 1.74 | 5.1 | 0.1 | 50.6 |

**Correlation coefficients healthcare-related variables**

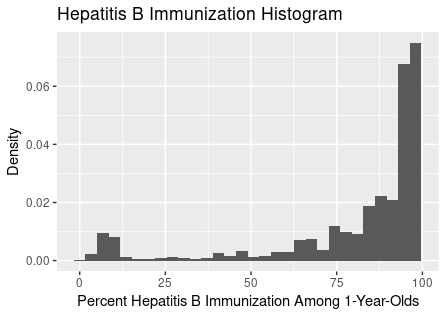
**Table B2**. Correlation coefficients among health variables and GDP and number of years of schooling.

**Alcohol Consumption**

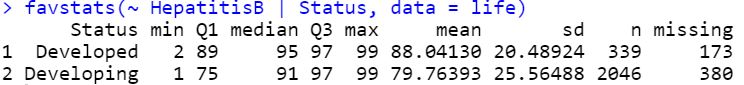


**Figure B1**. Alcohol consumption histogram (a) and scatter plot against life expectancy (b).

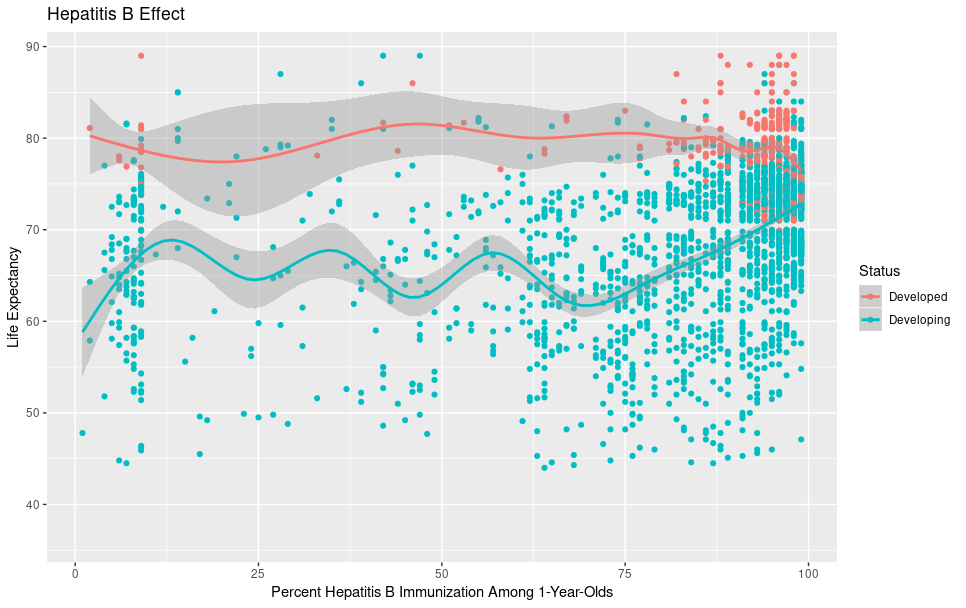
**Hepatitis B Immunization**

****

**Figure B2.** Distribution of percent Hepatitis B immunization among one-year-olds

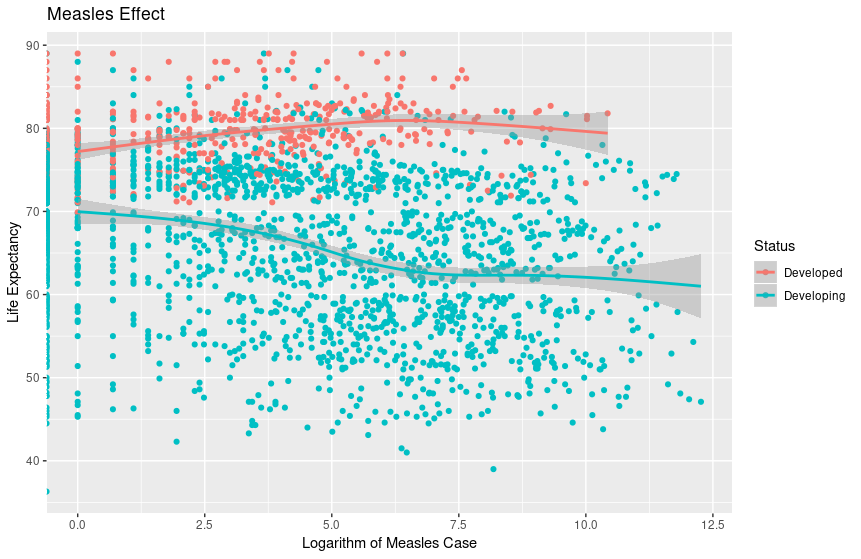
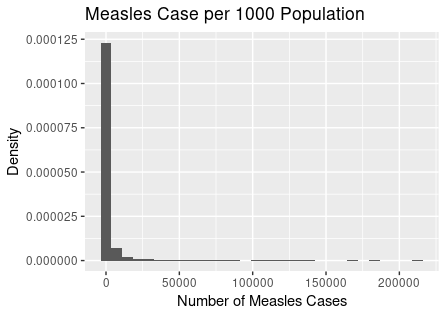
****

There is a large difference between percent Hepatitis B immunization between developed and developing countries.

****

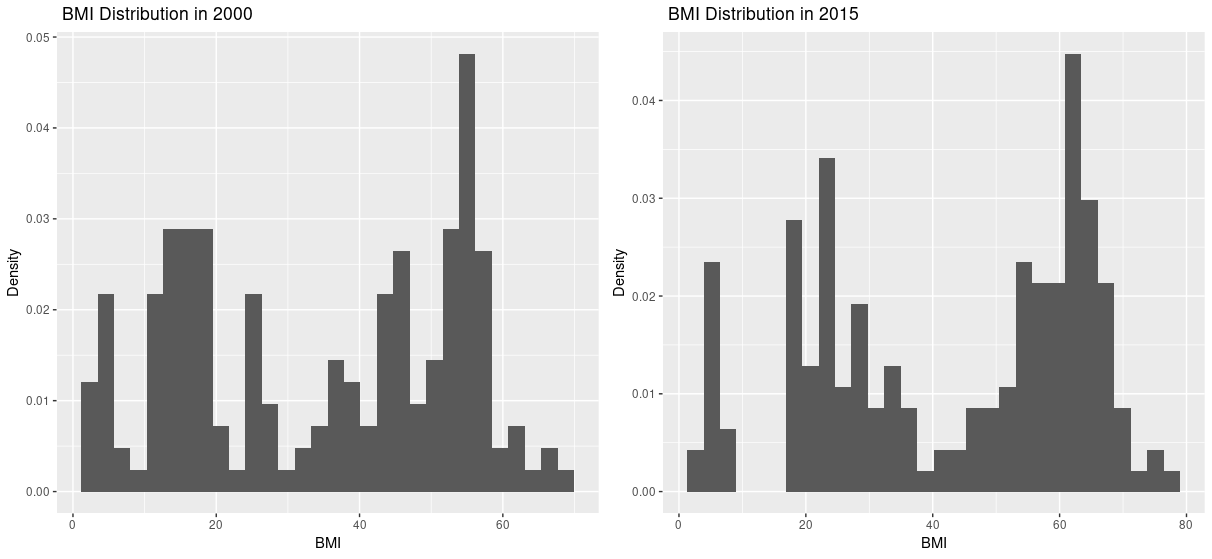
**Figure B3.** Scatter plot of percent Hepatitis B immunization against life expectancy, controlled by country status.

**Number of Measles Cases**

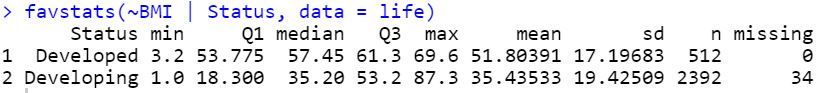
****

**Figure B4.** Distribution of number of measles cases (a) and scatter plot of log(measles cases) against life expectancy.

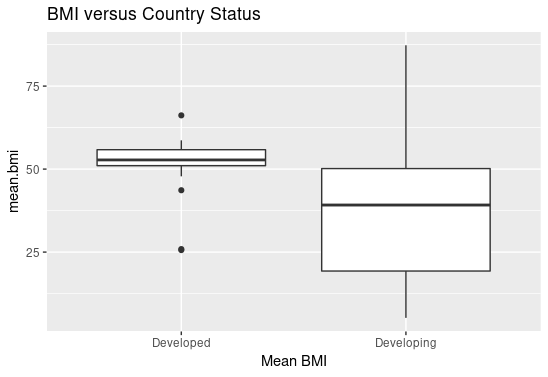
**BMI**

****

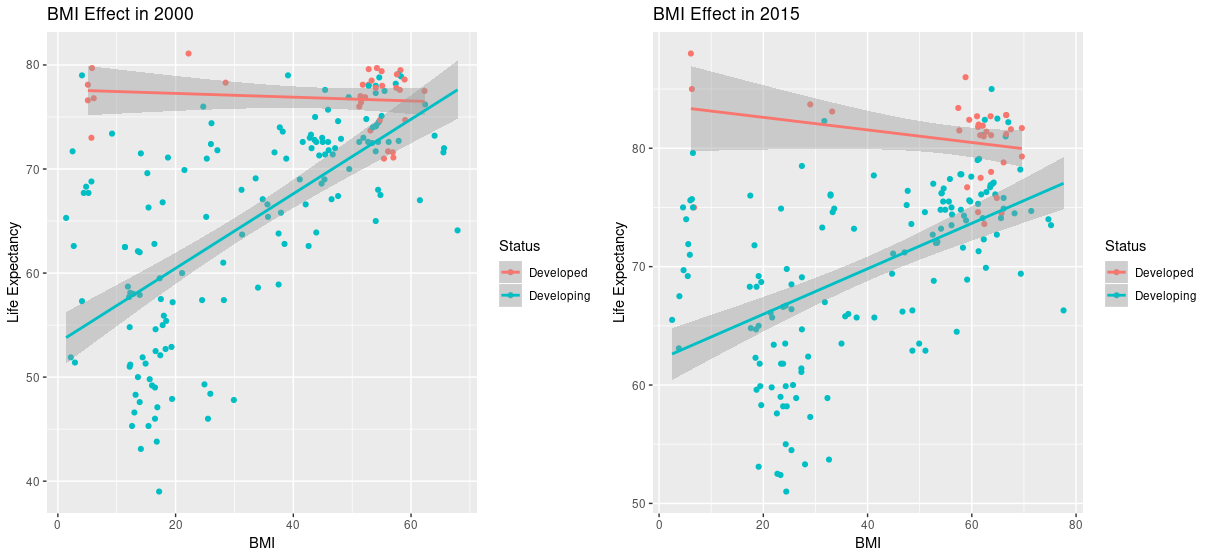
**Figure B5.** BMI distributions in 2000 (a) and 2015 (b) show that BMI shifted higher over year.

****

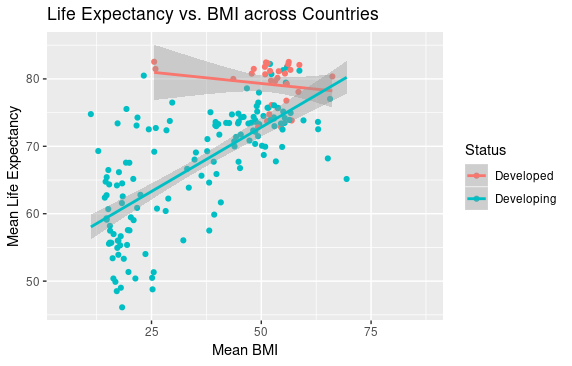
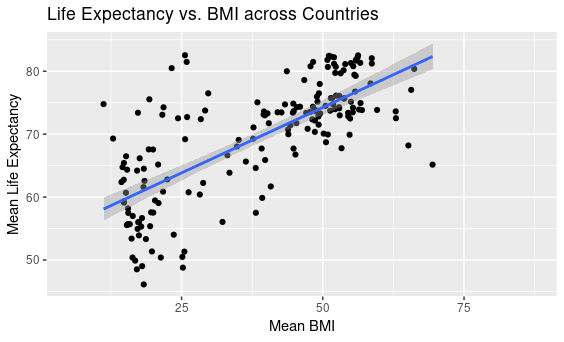
There is a large difference in BMI between developed and developing countries.

****

**Figure B6.** Boxplot of mean BMI between developed and developing countries. There are wide variability within developing countries. Developed countries have noticeably higher BMI.

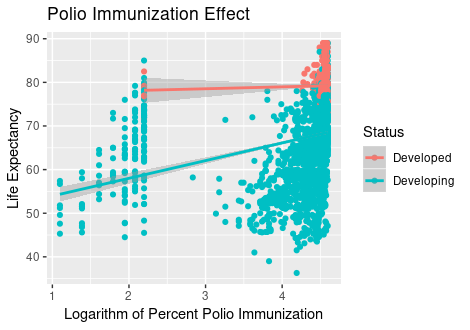
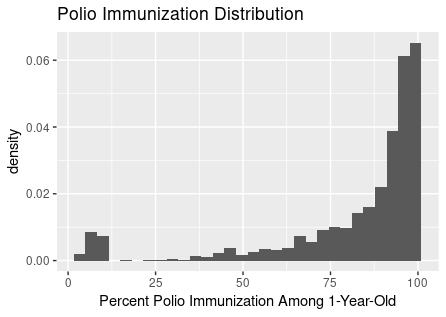
****

**Figure B7.** Scatter plots of BMI against life expectancy controlled by country status in 2000 (a) and 2015 (b). There is a difference in BMI effect between the two groups.

****

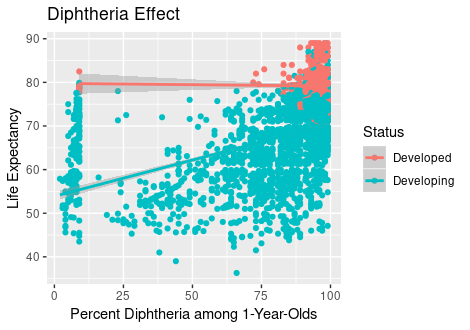
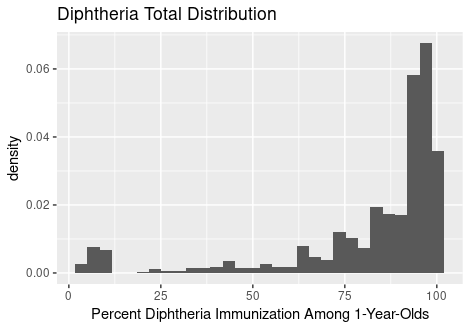
**Figure B8.** Scatter plots of mean BMI from each country against life expectancy. Comparing between the two plots, we can see that country status is a controlling factor.

**Polio Immunization**

****

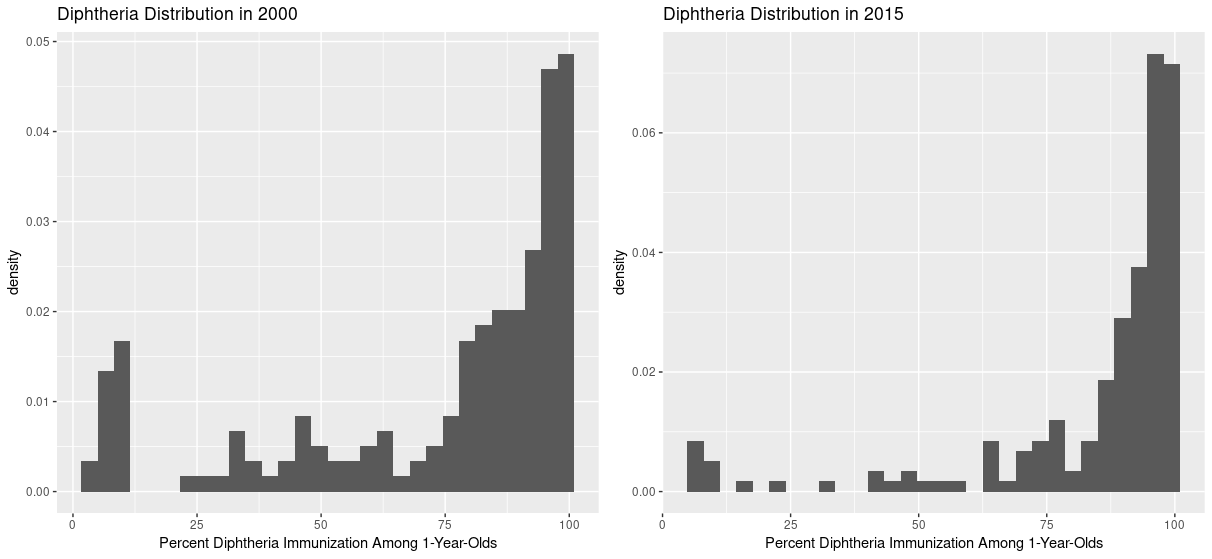
**Figure B9.** Histogram of percent Polio immunization among one-year-olds (a) is very left-skewed. Scatter plot of logarithm of percent Polio immunization against life expectancy controlled for country status (b) displays a slightly positive trend.

**Diphtheria Immunization Effect**

****

**Figure B10.** Histogram of percent Diphtheria immunization among one-year-olds (a) and scatter plot against life expectancy shows similarity with percent Polio immunization. There is likelihood of high correlation between the two variables.

**(a) (b)**

****

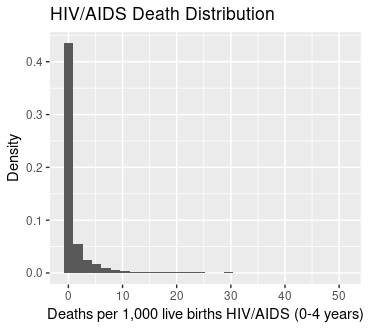
**Figure B11.** Distributions of percent Diphtheria immunization in 2000 (a) and 2015 (b) show that percent immunization shifted right over time.

**(a) (b)**

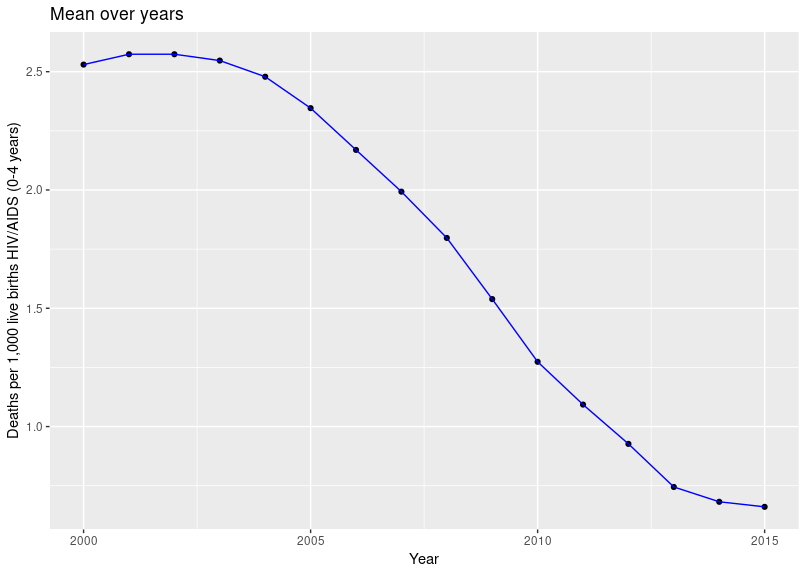
****

**Figure B12.** Scatter plots of percent Diphtheria among one-year-olds against life expectancy controlled for country status in 2000 (a) and 2015 (b) indicates there is a similar trend over years among developing countries.

**HIV/AIDS Effect**

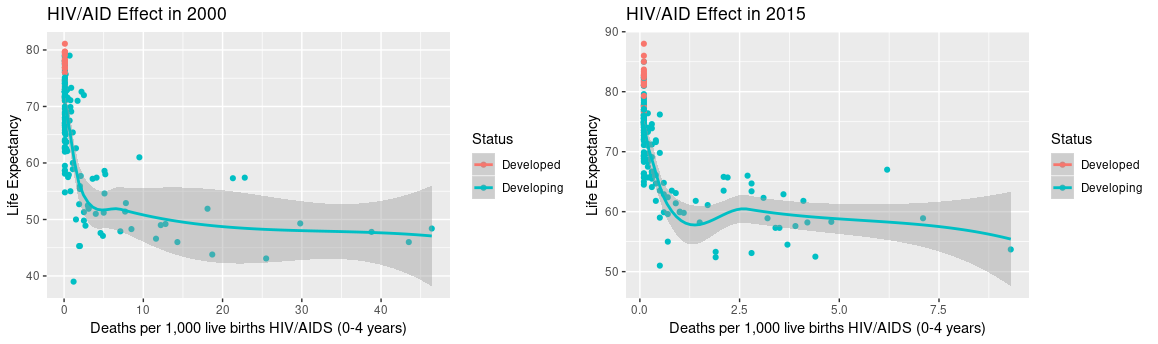


**Figure B13.** Histogram of deaths per 1,000 HIV/AIDS live births.

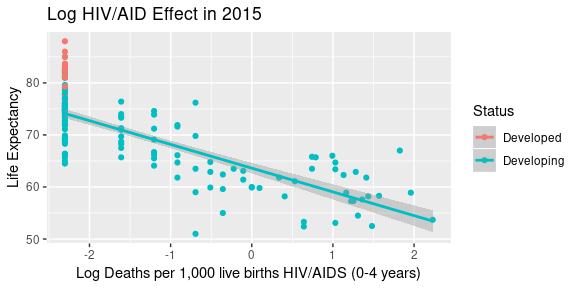


**Figure B14.** Scatter plot shows global mean number of deaths per 1,000 HIV/AIDS live births decreases over time.

**(a) (b)**



**Figure B15.** Scatter plots of number of deaths per 1,000 HIV/AIDS live births against life expectancy in 2000 (a) and 2015 (b) show a similar trend.



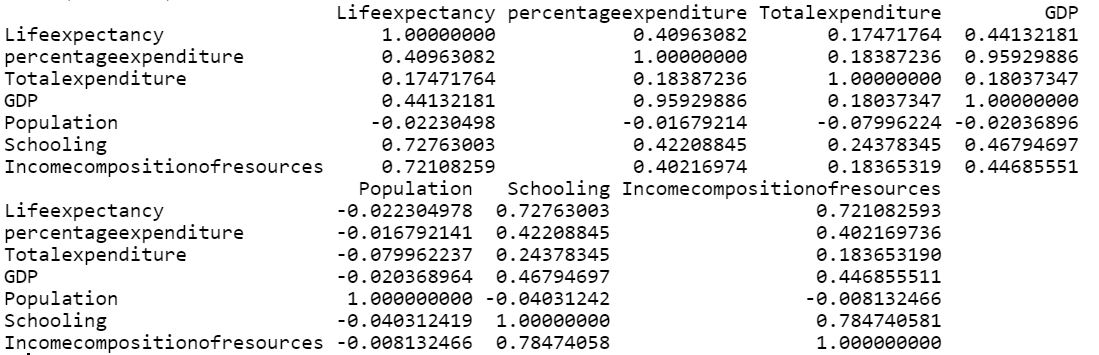
**Figure B16.** Scatter plot of log(deaths rate HIV/AIDS births) against life expectancy shows a linear relationship.

**C - Socioeconomic related variables:**

**Table C1.** Statistics Summary Table for all socioeconomic explanatory variables

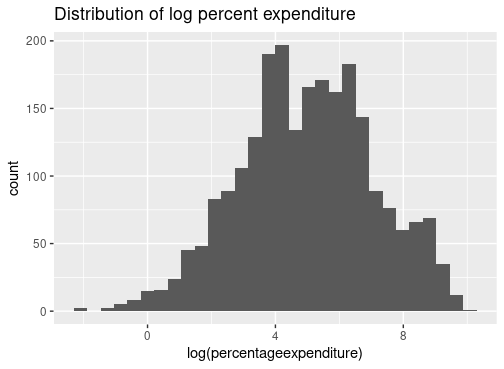
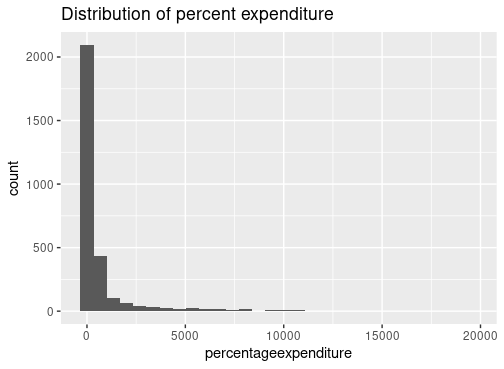
| **Variables (unit)** | **Mean** | **Standard Deviation** | **Minimum** | **Maximum** |
| --- | --- | --- | --- | --- |
| Percent expenditure | 738.3 | 1987.9 | 0 | 19479.9 |
| Total expenditure | 5.9 | 2.5 | 0.37 | 17.6 |
| GDP | 7483.2 | 14270.2 | 1.7 | 119172.7 |
| Population | 12753375 | 61012097 | 34 | 1293859294 |
| Income Composition of resources | 0.6 | 0.2 | 0 | 0.948 |
| Schooling | 11.9 | 3.4 | 0 | 20.7 |

**Table C2.** Correlation coefficients among socioeconomic variables.

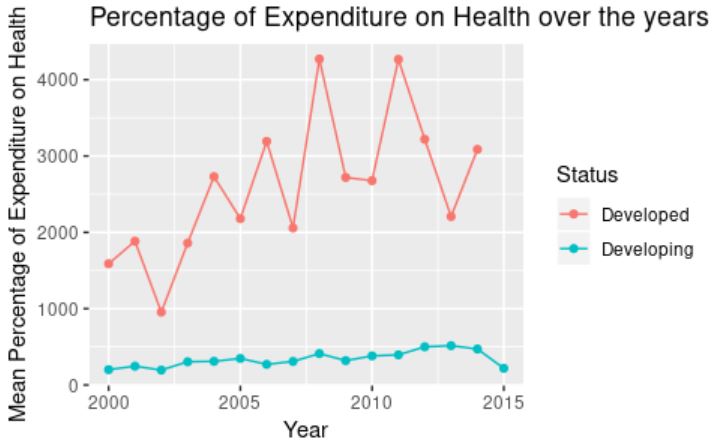
****

**Percent Expenditure on Health**

**a) b)**

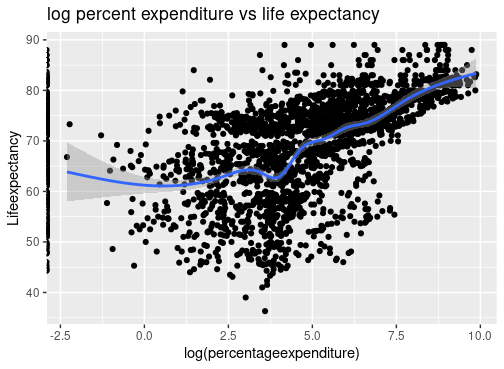
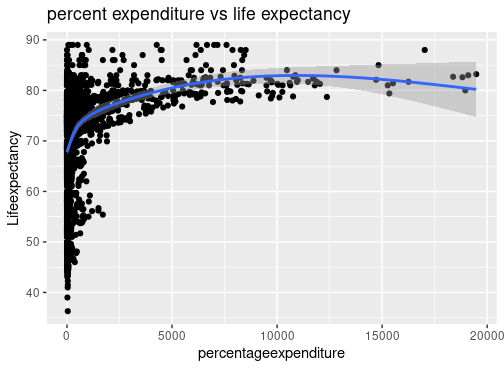
****

**Figure C1.** Histogram of percentage of expenditure on health (a) is very right skewed, while histogram log transformation (b) more resembles normal distribution.

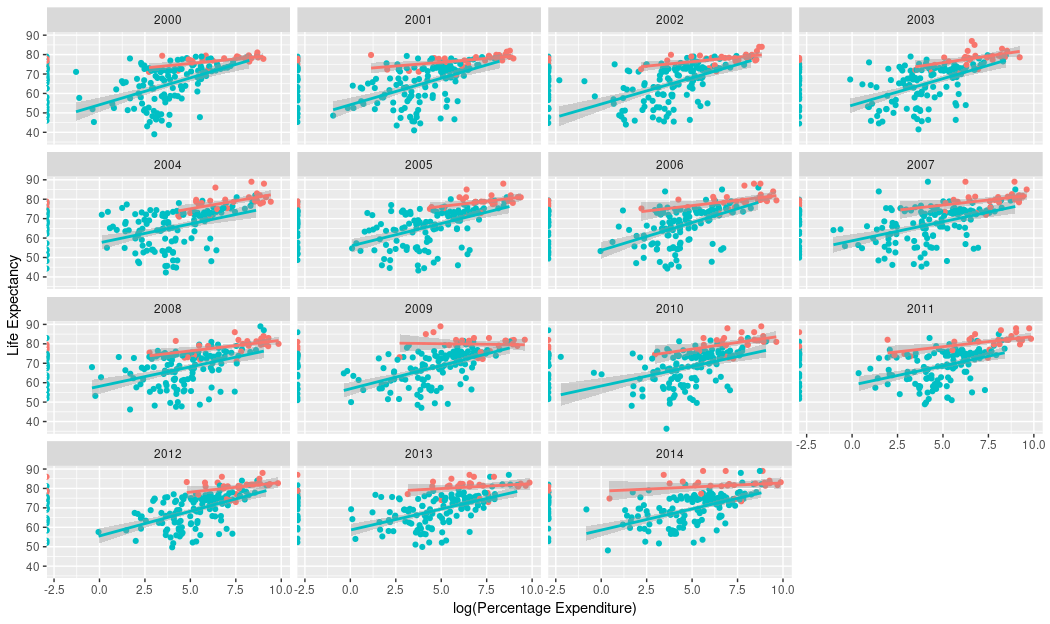


**Figure C2.** Percentage of expenditure on health over the years.

**a) b)**



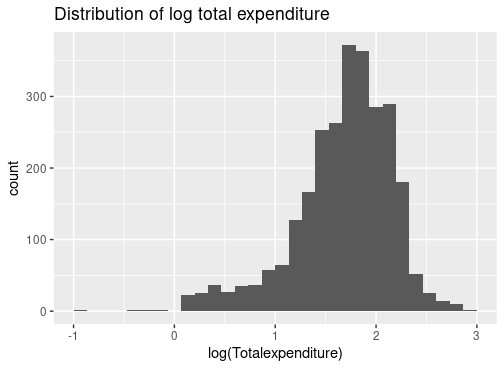
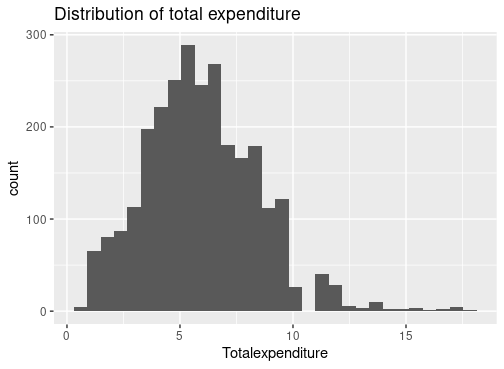
**Figure C3.** Scatter plots of percentage of expenditure on health versus life expectancy (a) and log(percent expenditure) versus life expectancy (b).



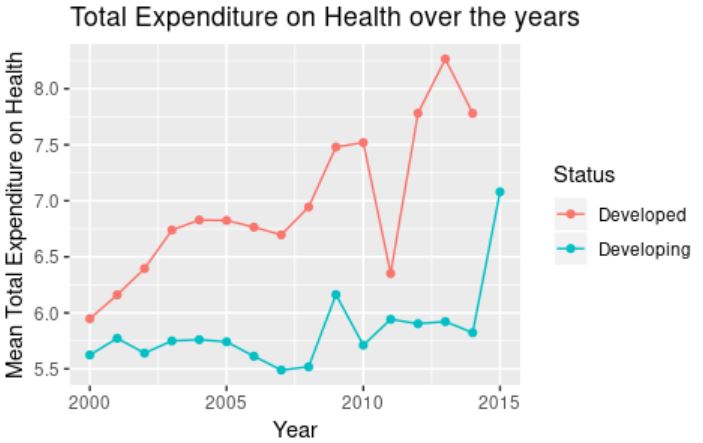
**Figure C4.** Scatter plots of log(percent expenditure) versus life expectancy in spanning from 2000-2014

**Total Expenditure:**

**a) b)**

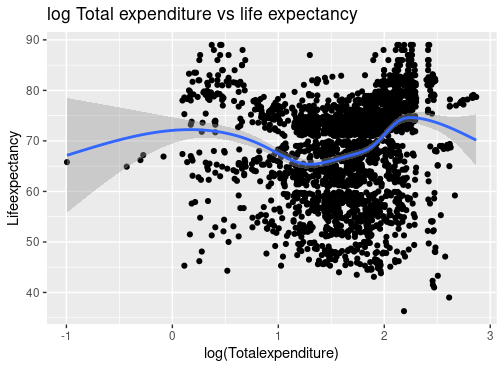
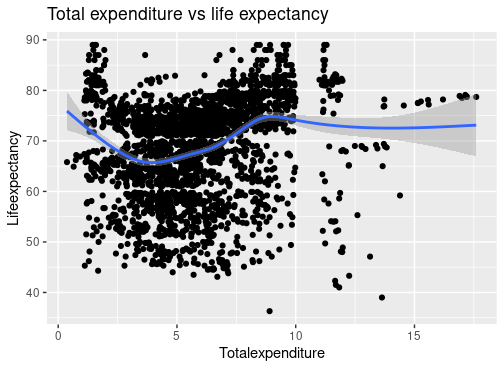
****

**Figure C5.** Histogram of total expenditure (a) and distribution of log(total expenditure) (b)

****

**Figure C6.** Mean total expenditure over the years. There is a more drastic increase among developed countries, compared with developing countries.

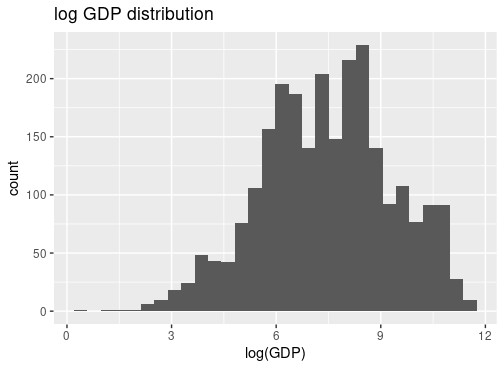
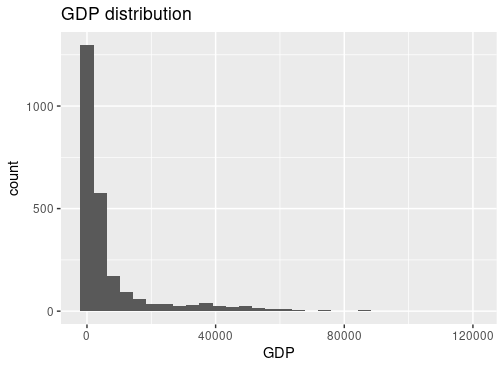
**a) b)**

****

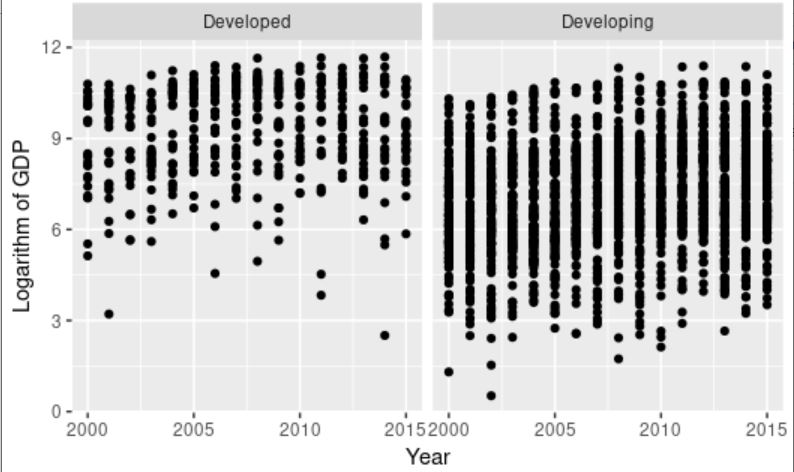
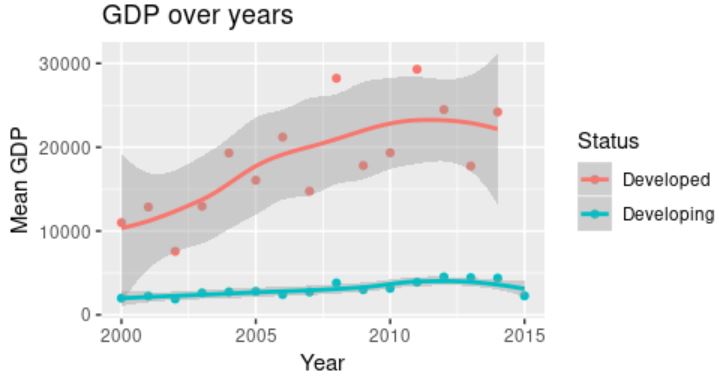
**Figure C7.** Scatter plots of total expenditure on health against life expectancy (a) and log(total expenditure) against life expectancy (b). None shows any evident relationship with life expectancy.

**GDP:**

**a) b)**

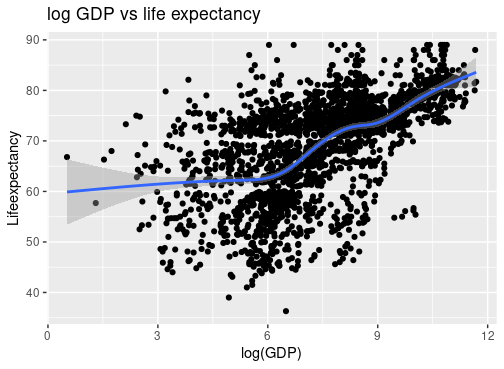
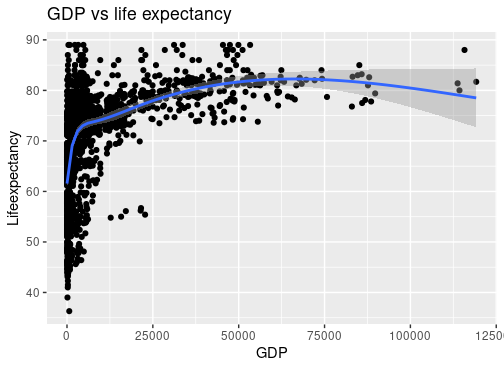
****

**Figure C8.** Distribution of GDP (a) is very right skewed, while distribution of log(GDP) is more normally distributed.

**a) b)  
**

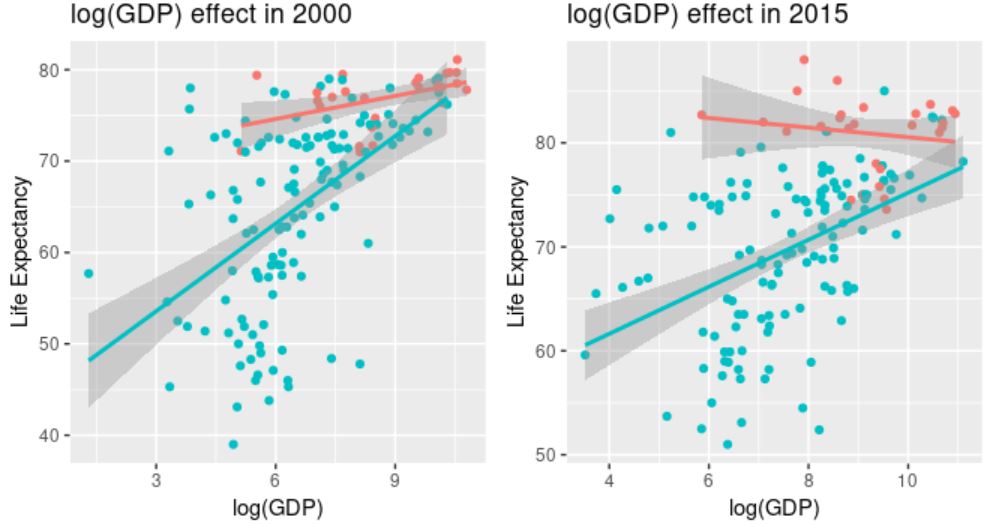
**Figure C9.** Scatter plots of mean GDP over time (a) and log(GDP) over time (b).

**a) b)**

****

**Figure C10.** Scatter plot of GDP against life expectancy (a) and log(GDP) against life expectancy (b)

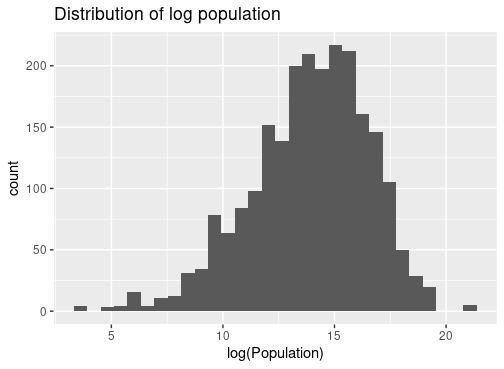
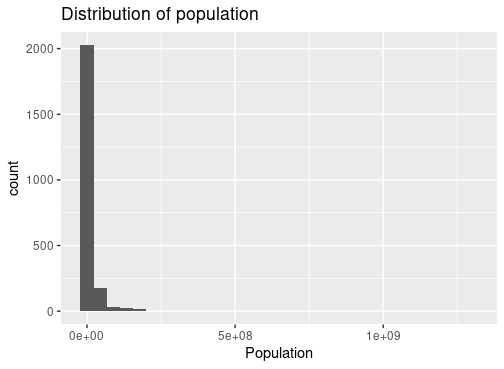
**a) b)**

****

**Figure C11.** Scatter plot of life expectancy versus log(GDP) in 2000 (a) and 2015 (b)

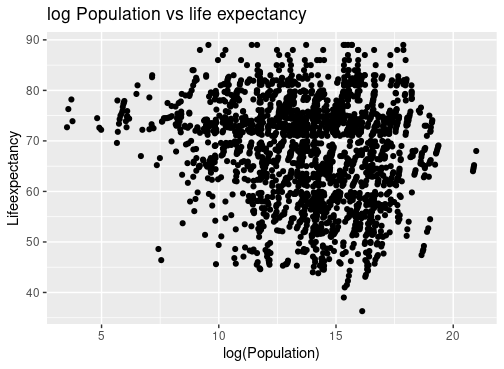
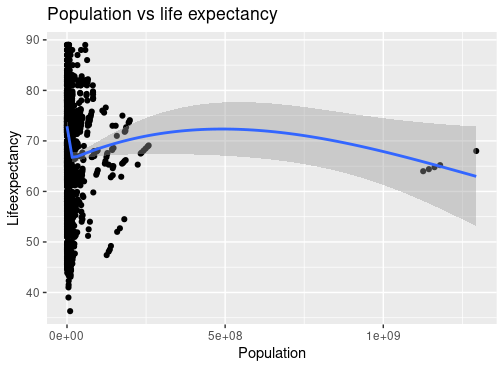
**Population:**

**a) b)**

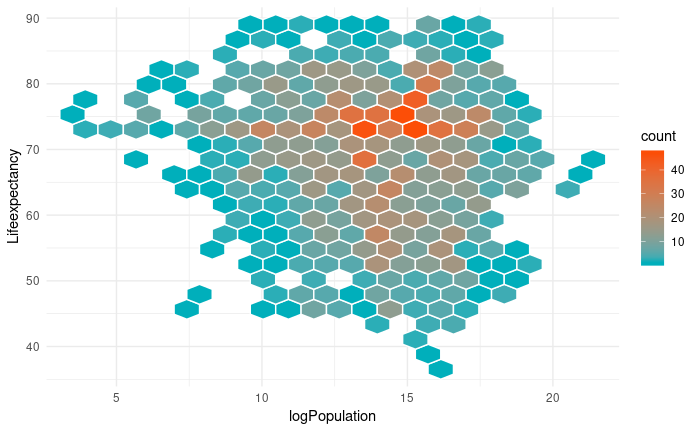
****

**Figure C12.** Histogram of population (a) and log(population) (b)

**a) b)**

****

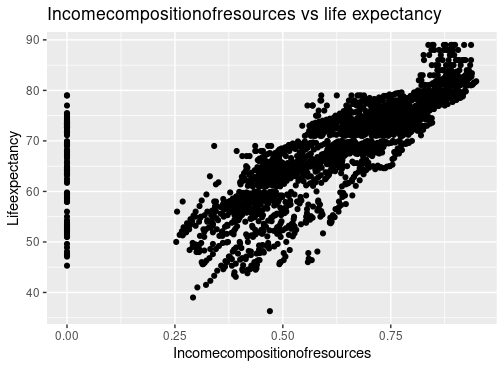
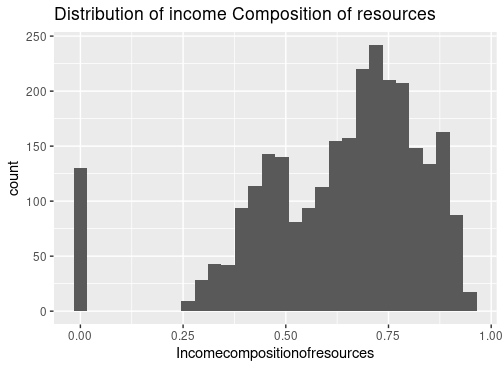
**Figure C13.** Scatter plot of life expectancy versus population (a) and log(population) (b)

****

**Figure C14.** Hexagon intensity plot displaying the distribution in life expectancy versus log population

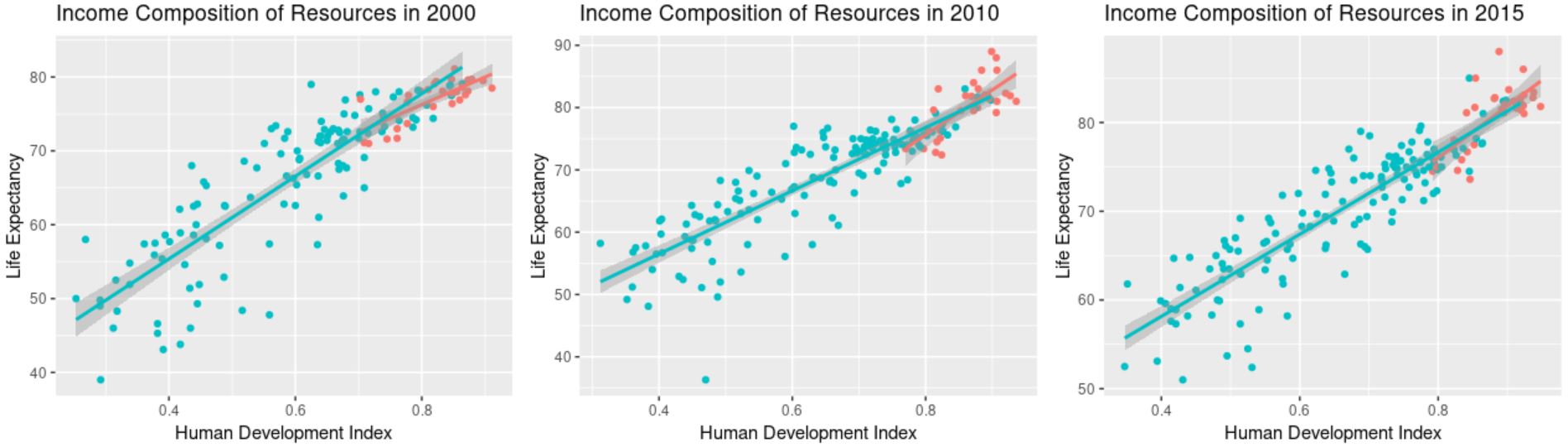
**Income Composition of Resources**

**a) b)**

****

**Figure C15.** Distribution of income composition of resources (a) and scatter plot against life expectancy.

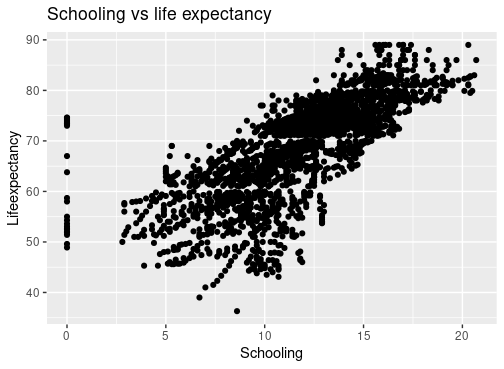
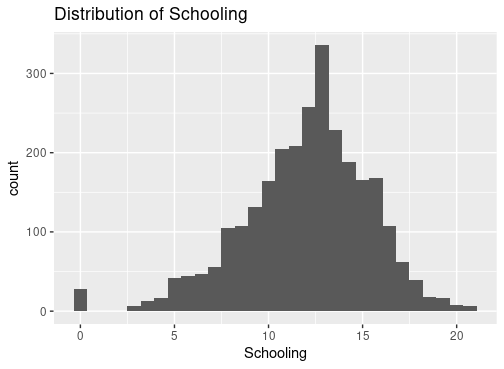
**a) b) c)**

****

**Figure C16.** Scatter plot of life expectancy income composition of resources in human development index in 2000 (a), 2010 (b), and 2015 (c).

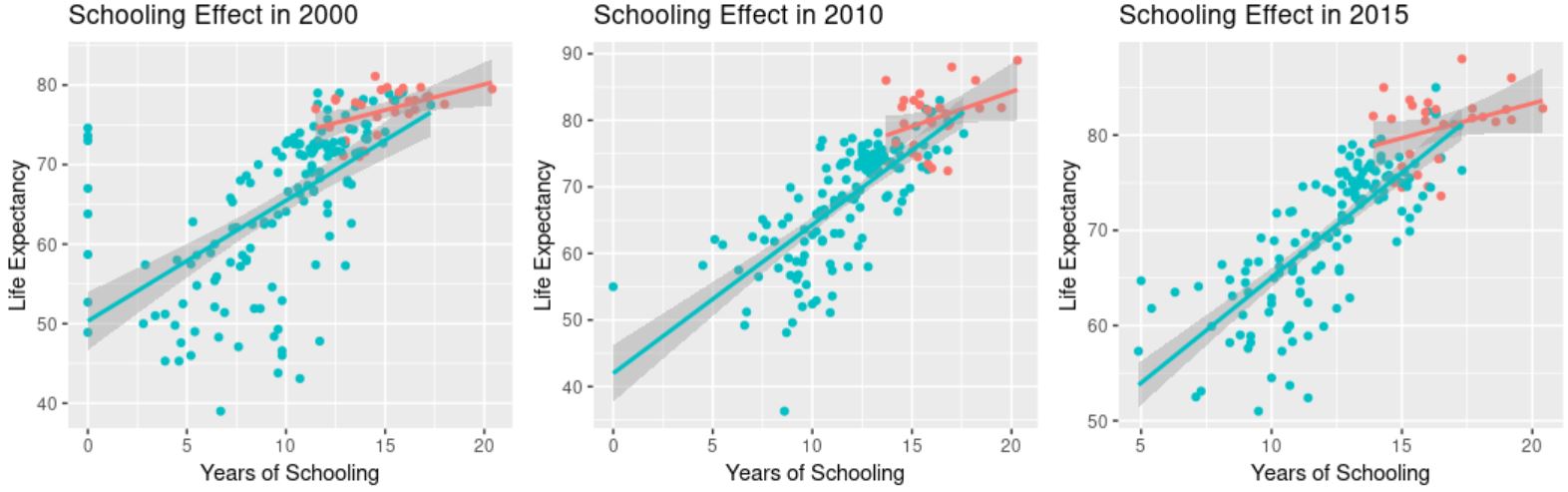
**Number of Years of Schooling:**

**a) b)**

****

**Figure C17.** Distribution of years of schooling (a) and scatter plot against life expectancy (b)

**a) b) c)**

****

**Figure C18.** Scatter plots of life expectancy versus years of schooling in 2000 (a), 2010 (b), and 2015 (c)

**Codes:**

Data manipulation:

>library(readr)

>library(countrycode)

>library(dplyr)

>le.df <- read\_csv("~/Stats 316 F19/Project/Leon-Jessie-Calvin/Data/life-expectancy-original.csv")

#add continents based on country names

>le.df$Continent <- countrycode(sourcevar = le.df$Country, origin ="country.name",

destination = "continent")

#countrycode function could not recognize following countries, add their continents manually

>le.df <- le.df %>% mutate(Continent = ifelse(Country == "CentralAfricanRepublic", "Africa",Continent)) %>%

mutate(Continent = ifelse(Country == "DominicanRepublic", "Americas",Continent)) %>%

mutate(Continent = ifelse(Country == "LaoPeople'sDemocraticRepublic", "Asia",Continent)) %>% mutate(Continent = ifelse(Country == "RepublicofKorea", "Asia",Continent)) %>%

mutate(Continent = ifelse(Country == "SaintLucia", "Americas",Continent)) %>%

mutate(Continent = ifelse(Country == "SouthAfrica", "Africa",Continent)) %>%

mutate(Continent = ifelse(Country == "UnitedStatesofAmerica", "Americas",Continent))

#check

>sum(is.na(le.df$Continent))

Baseline/Response variables:

>library(readr)

>library(countrycode)

>library(dplyr)

>library(mosaic)

>life <- read\_csv("~/Stats 316 F19/Project/Leon-Jessie-Calvin/Leon-EDA/Life-Analysis.csv")

*#convert character cols to factor*

>life <- mutate\_if(life, is.character, as.factor)

>summary(life) ##Overall summary

>keydata <- life %>% ##create a new dataset with only 193 observations.

dplyr::select(Country, Year, Status, Continent,Lifeexpectancy, infantdeaths, AdultMortality,`under-fivedeaths`)

>country.lev2 <- keydata %>%

group\_by(Country) %>%

filter(row\_number() == 1) %>%

select(Country, Status, Continent)

*# Add average across all performances for each subject for EDA plots*

>meanbysubj <- life %>% group\_by(Country) %>%

> summarise(meanbysubj = mean(Lifeexpectancy, na.rm = TRUE),

meanAdult = mean(AdultMortality, na.rm=T),

meanUnder = mean(`under-fivedeaths`, na.rm=T),

meanInf = mean(infantdeaths))

>country.lev2 <- country.lev2 %>%

left\_join(meanbysubj, by = "Country")

>cat("Min year:", min(life$Year))

>cat("Max year:", max(life$Year))

>status.df <- life %>%

subset(Year==2000) %>% *#status did not change from this period*

select(Country, Status) %>%

distinct(Country, .keep\_all = TRUE)

summary(status.df$Status)

>continent.df <- life %>%

subset(Year==2015) %>%

select(Country, Continent) %>%

distinct(Country, .keep\_all = TRUE)

>summary(continent.df$Continent)

>favstats(life$Lifeexpectancy)

>favstats(country.lev2$meanbysubj)

>ggplot(data=life, aes(x=Lifeexpectancy)) +

geom\_histogram() +

labs(title = "Life Expectancy histogram") +

xlab("Life expectancy")

>ggplot(country.lev2, aes(x=meanbysubj)) +

geom\_histogram() +

xlab("Mean life expectancy") +

labs(title = "Mean life expectancy")

>favstats(life$AdultMortality)

favstats(country.lev2$meanAdult)

>ggplot(data=life, aes(x=AdultMortality)) +

geom\_histogram() +

labs(title = "Adult Mortality histogram") +

xlab("Adult Mortality deaths")

>ggplot(data=country.lev2,aes(x=meanAdult)) +

geom\_histogram() +

xlab("Mean adult mortality") +

labs(title = "Mean adult mortality")

>favstats(life$infantdeaths)

>favstats(country.lev2$meanInf)

>sum(country.lev2$meanInf==0)

>ggplot(data=life, aes(x=infantdeaths)) +

geom\_histogram() +

labs(title = "Infant Death histogram") +

xlab("Infant deaths")

>ggplot(data=country.lev2, aes(x=meanInf)) +

geom\_histogram() +

labs(title = "Mean infant deaths") +

xlab("Mean infant death")

>favstats(life$`under-fivedeaths`)

>sum(country.lev2$meanUnder==0)

>ggplot(data=life, aes(x=`under-fivedeaths`)) +

geom\_histogram() +

labs(title = "Under-five Death histogram") +

xlab("Under-five deaths")

>ggplot(data=country.lev2,aes(x=meanUnder)) +

geom\_histogram() +

labs(title = "Mean under-five deaths") +

xlab("mean under-five deaths")

>favstats(Lifeexpectancy ~ as.factor(Country), data=life)

>country.df <- life %>%

select(Country, Lifeexpectancy) %>%

filter(!is.na(Lifeexpectancy)) %>%

group\_by(Country) %>%

summarise(MeanLife = mean(Lifeexpectancy))

head(country.df)

>country.df[which.max(country.df$MeanLife),]

>country.df[which.min(country.df$MeanLife),]

>life0 <- life %>%

filter(!is.na(Lifeexpectancy))

cor(life0$Lifeexpectancy, life0$Year)

>ggplot(life, aes(x=Year, y=Lifeexpectancy)) +

geom\_point() +

geom\_smooth() +

labs(title="Life expectancy vs. years")

>ggplot(life, aes(x=Year, y=Lifeexpectancy, col=Status)) +

geom\_point() +

geom\_smooth() +

labs(title="Life expectancy vs. years")

>li1 <- life %>%

filter(!is.na(Lifeexpectancy)) %>%

select(Country, Year, Lifeexpectancy, Status) %>%

group\_by(Year) %>%

summarise(MeanLife = mean(Lifeexpectancy))

>cor(li1$Year, li1$MeanLife)

>life %>%

filter(!is.na(Lifeexpectancy)) %>%

select(Country, Year, Lifeexpectancy, Status) %>%

group\_by(Year) %>%

summarise(MeanLife = mean(Lifeexpectancy)) %>%

ggplot(aes(x=Year,y=MeanLife)) +

geom\_point() +

geom\_smooth(se=F) +

labs(title = "Year vs. mean life expectancy")

>life %>%

filter(!is.na(Lifeexpectancy)) %>%

select(Country, Year, Lifeexpectancy, Status) %>%

group\_by(Year,Status) %>%

summarise(MeanLife = mean(Lifeexpectancy)) %>%

ggplot(aes(x=Year,y=MeanLife, col=Status)) +

geom\_point() +

geom\_smooth(se=F) +

labs(title = "Year vs. mean life expectancy")

>ggplot(data=life,aes(x=Status, y=Lifeexpectancy)) +

geom\_boxplot() +

labs(title="Status vs. life expectancy")

>ggplot(data=country.lev2, aes(x=Status, y=meanbysubj)) +

geom\_boxplot() +

labs(title="Status vs. mean life expectancy") +

ylab("mean life expectancy")

>ggplot(life, aes(x=Continent, y=Lifeexpectancy)) +

geom\_boxplot() +

labs(title="Continent vs. life expectancy")

>ggplot(country.lev2, aes(x=Continent, y=meanbysubj)) +

geom\_boxplot() +

labs(title="Continent vs. mean life expectancy") +

xlab("mean life expectancy")

>favstats(meanbysubj ~ Continent, data=country.lev2)

>a1 <- aov(country.lev2$meanbysubj~ country.lev2$Continent)

>posthoc <- TukeyHSD(x=a1, conf.level=0.95)

>print(posthoc)

*#map*

>WorldData <- map\_data('world') %>% filter(region != "Antarctica") %>% fortify

>df <- country.lev2

>p <- ggplot() +

geom\_map(data = WorldData, map = WorldData,

aes(x = long, y = lat, group = group, map\_id=region),

fill = "white", colour = "#7f7f7f", size=0.5) +

geom\_map(data = df, map=WorldData,

aes(fill=meanbysubj, map\_id=Country),

colour="#7f7f7f", size=0.5) +

coord\_map("rectangular", lat0=0, xlim=c(-180,180), ylim=c(-60, 90)) +

scale\_fill\_continuous(low="blue", high="red")+

scale\_y\_continuous(breaks=c()) +

scale\_x\_continuous(breaks=c()) +

labs(fill="Average Life Expectancy", title="World average life expectancy (2000-2015)", x="", y="") +

theme\_bw()

**Healthcare related variables:**

>new.life <- na.omit(life)

df.corr <- new.life %>% select(Lifeexpectancy, Alcohol, HepatitisB, Measles, BMI, Polio, Diphtheria, `HIV/AIDS`, GDP, Schooling)

>cor(df.corr)

>summary(life$Lifeexpectancy)

>ggplot(data = life, aes(x=Lifeexpectancy,y=..density..)) +

geom\_histogram() +

labs(x = "Life expectancy", title = "Life Expectancy Distribution across Countries", y = "Density")

>le2015 <- life %>% filter(Year == 2015) %>%

ggplot(aes(x=Lifeexpectancy,y=..density..)) +

geom\_histogram() +

labs(x = "Life expectancy", title = "Life Expectancy in 2015", y = "Density") +

theme(plot.title=element\_text(hjust=.9,face="italic",size=12))

>le2000 <- life %>% filter(Year == 2000) %>%

ggplot(aes(x=Lifeexpectancy,y=..density..)) +

geom\_histogram() +

labs(x = "Life expectancy", title = "Life Expectancy in 2000", y = "Density") +

theme(plot.title=element\_text(hjust=.9,face="italic",size=12))

>grid.arrange(le2000, le2015, ncol = 2)

>life %>%

group\_by(Status, Year) %>%

summarise(meanle = mean(Lifeexpectancy)) %>%

ggplot(aes(x=Year, y=meanle, color = Status)) +

geom\_point() +

geom\_smooth(size = 0.5, method = "lm") +

labs(y = "Mean Life Expectancy", title = "Life Expectancy Rate")

>summary(life$Alcohol)

>ggplot(data = life, aes (x = Alcohol, y=..density..)) +

geom\_histogram() +

labs(x = "Alcohol Consumption", title = "Alcohol Consumption Histogram")

>ggplot(data = life,aes(x = Alcohol, y = Lifeexpectancy)) +

geom\_point(na.rm = T) +

geom\_smooth(method = "lm") +

labs(title = "Alcohol Effect", x = "Amount of Alcohol Consumption (in liters)", y = "Life Expectancy")

>summary(life$HepatitisB)

>ggplot(data=life, aes(x=HepatitisB, y = ..density..)) +

geom\_histogram() +

labs(x="Percent Hepatitis B Immunization Among 1-Year-Olds", y="Density", title = "Hepatitis B Immunization Histogram")

>favstats(~ HepatitisB | Status, data = life)

>ggplot(data = life, aes(x=HepatitisB, y = Lifeexpectancy, color = Status)) +

geom\_point() +

geom\_smooth() +

labs(x = "Percent Hepatitis B Immunization Among 1-Year-Olds", y = "Life Expectancy", title = "Hepatitis B Effect")

>summary(life$Measles)

>ggplot(data = life[], aes(x=Measles, y =..density..)) +

geom\_histogram() +

labs(title = "Measles Case per 1000 Population", x = "Number of Measles Cases", y = "Density")

>ggplot(data = life, aes(x=log(Measles), y = Lifeexpectancy, color = Status)) +

geom\_point() +

geom\_smooth() +

labs(x = "Logarithm of Measles Case", y = "Life Expectancy", title = "Measles Effect")

>life2015 <- life %>% filter(Year == 2015)

>summary(life$BMI)

>favstats(~BMI | Status, data = life2015)

>ggplot(data = life, aes(x=BMI, y=..density..)) +

geom\_histogram() +

labs(title =" BMI Distribution", y = "Density")

>bmi2000 <- life %>% filter(Year == 2000) %>%

ggplot(aes(x=BMI, y=..density..)) +

geom\_histogram() +

labs(title =" BMI Distribution in 2000", y = "Density")

>bmi2015 <- life %>% filter(Year == 2015) %>%

ggplot(aes(x=BMI, y=..density..)) +

geom\_histogram() +

labs(title =" BMI Distribution in 2015", y = "Density")

>grid.arrange(bmi2000, bmi2015, ncol=2)

ggplot(data = life, aes(x=BMI, y=Lifeexpectancy, color = Status)) +

geom\_point() +

geom\_smooth(method = "lm") +

labs(title = "BMI Effect")

>bmi2000plot <- life %>% filter(Year == 2000) %>%

ggplot(aes(x=BMI, y=Lifeexpectancy, color = Status)) +

geom\_point() +

geom\_smooth(method = "lm") +

labs(title = "BMI Effect in 2000", y = "Life Expectancy")

>bmi2015plot <- life %>% filter(Year == 2015) %>%

ggplot(aes(x=BMI, y=Lifeexpectancy, color = Status)) +

geom\_point() +

geom\_smooth(method = "lm") +

labs(title = "BMI Effect in 2015", y = "Life Expectancy")

>grid.arrange(bmi2000plot,bmi2015plot, ncol = 2)

>mean.bmi <- life %>%

group\_by(Country) %>%

summarise(mean.bmi = mean(BMI),

mean.le = mean(Lifeexpectancy))

>mean.bmi2 <- life %>%

group\_by(Country) %>%

filter(row\_number() == 1) %>%

left\_join(mean.bmi, by ="Country")

>ggplot(data = mean.bmi2,aes(x=Status,y=mean.bmi)) +

geom\_boxplot() +

labs(title = "BMI versus Country Status", x = "Mean BMI")

>ggplot(data = mean.bmi2, aes(x=mean.bmi, y=mean.le, color = Status)) +

geom\_point() +

geom\_smooth(method = "lm") +

labs(x = "Mean BMI", y="Mean Life Expectancy", title = "Life Expectancy vs. BMI across Countries")

>summary(life$Polio)

>ggplot(data = life, aes(x=Polio,y=..density..)) +

geom\_histogram() +

labs(x = "Percent Polio Immunization Among 1-Year-Old", title = "Polio Immunization Distribution")

>ggplot(data = life, aes(x=log(Polio), y=Lifeexpectancy, color = Status)) +

geom\_point() +

geom\_smooth(method = "lm") +

labs(x = "Logarithm of Percent Polio Immunization", y = "Life Expectancy", title = "Polio Immunization Effect")

>summary(life$Diphtheria)

>favstats(~ Diphtheria | Status, data = life)

>ggplot(data = life, aes(x=Diphtheria,y=..density..)) +

geom\_histogram() +

labs(title ="Diphtheria Total Distribution", x = "Percent Diphtheria Immunization Among 1-Year-Olds")

>diph2000 <- life %>% filter(Year == 2000) %>%

ggplot(aes(x=Diphtheria,y=..density..)) +

geom\_histogram() +

labs(title ="Diphtheria Distribution in 2000", x = "Percent Diphtheria Immunization Among 1-Year-Olds")

>diph2015 <- life %>% filter(Year == 2015) %>%

ggplot(aes(x=Diphtheria,y=..density..)) +

geom\_histogram() +

labs(title ="Diphtheria Distribution in 2015", x = "Percent Diphtheria Immunization Among 1-Year-Olds")

>grid.arrange(diph2000,diph2015, ncol = 2)

>ggplot(data = life, aes(x=Diphtheria, y=Lifeexpectancy, color = Status)) +

geom\_point() +

geom\_smooth(method = "lm") +

labs(x = "Percent Diphtheria among 1-Year-Olds", y="Life Expectancy", title ="Diphtheria Effect")

>diph2000plot <- life %>% filter(Year == 2000) %>%

ggplot(aes(x=Diphtheria, y=Lifeexpectancy, color = Status)) +

geom\_point() +

geom\_smooth(method = "lm") +

labs(x = "Percent Diphtheria among 1-Year-Olds", y="Life Expectancy", title ="Diphtheria Effect in 2000")

>diph2015plot <- life %>% filter(Year == 2015) %>%

ggplot(aes(x=Diphtheria, y=Lifeexpectancy, color = Status)) +

geom\_point() +

geom\_smooth(method = "lm") +

labs(x = "Percent Diphtheria among 1-Year-Olds", y="Life Expectancy", title ="Diphtheria Effect in 2015")

>grid.arrange(diph2000plot,diph2015plot, ncol =2)

>summary(life$`HIV/AIDS`)

>ggplot(data = life, aes(x=`HIV/AIDS`, y=..density..)) +

geom\_histogram() +

labs(x = "Deaths per 1,000 live births HIV/AIDS (0-4 years)", y = "Density", title = "HIV/AIDS Death Distribution")

>ggplot(data=life, aes(x=`HIV/AIDS`, y = Lifeexpectancy, color = Status)) +

geom\_point() +

geom\_smooth(method = "loess") +

labs(x = "Deaths per 1,000 live births HIV/AIDS (0-4 years)", y = "Life Expectancy", title = "HIV/AID Effect")

>hiv2000 <- life %>% filter(Year == 2000) %>%

ggplot(aes(x=`HIV/AIDS`, y = Lifeexpectancy, color = Status)) +

geom\_point() +

geom\_smooth(method = "loess") +

labs(x = "Deaths per 1,000 live births HIV/AIDS (0-4 years)", y = "Life Expectancy", title = "HIV/AID Effect in 2000")

>hiv2015 <- life %>% filter(Year == 2015) %>%

ggplot(aes(x=`HIV/AIDS`, y = Lifeexpectancy, color = Status)) +

geom\_point() +

geom\_smooth(method = "loess") +

labs(x = "Deaths per 1,000 live births HIV/AIDS (0-4 years)", y = "Life Expectancy", title = "HIV/AID Effect in 2015")

>grid.arrange(hiv2000, hiv2015, ncol=2)

>life %>% filter(Year == 2015) %>%

>ggplot(aes(x= log(`HIV/AIDS`), y = Lifeexpectancy, color = Status)) +

geom\_point() +

geom\_smooth(method = "lm") +

labs(x = "Log Deaths per 1,000 live births HIV/AIDS (0-4 years)", y = "Life Expectancy", title = "Log HIV/AID Effect in 2015")

**Socioeconomic variables:**

life <- read\_csv("~/Stats 316 F19/Project/Leon-Jessie-Calvin/Leon-EDA/Life-Analysis.csv")

*convert character cols to factor*

>life <- mutate\_if(life, is.character, as.factor)

keydata <- life %>%

dplyr::select(Country, Year, Status, Continent,Lifeexpectancy,percentageexpenditure,Totalexpenditure,GDP, Population,Incomecompositionofresources, Schooling)

social.lev2 <- keydata %>%

group\_by(Country) %>%

filter(row\_number() == 1) %>%

select(Country, Status:Schooling)

*# Add average across all performances for each subject for EDA plots*

>meanbysubj <- life %>% group\_by(Country) %>%

summarise(meanbysubj = mean(Lifeexpectancy, na.rm = TRUE))

>social.lev2 <- social.lev2 %>%

left\_join(meanbysubj, by = "Country")

>cor.df <- na.omit(life)

>ggplot(data=life, aes(x=percentageexpenditure)) +

geom\_histogram() +

labs(title="Distribution of percent expenditure")

>ggplot(data=life, aes(x=log(percentageexpenditure)) )+

geom\_histogram() +

labs(title="Distribution of log percent expenditure")

>favstats(life$percentageexpenditure)

>ggplot(data=social.lev2, aes(x=percentageexpenditure)) +

geom\_histogram()

>ggplot(data=life, aes(x=percentageexpenditure, y= Lifeexpectancy)) +

geom\_point() +

geom\_smooth() +

labs(title="percent expenditure vs life expectancy")

>ggplot(data=life, aes(x=log(percentageexpenditure), y= Lifeexpectancy)) +

geom\_point() +

geom\_smooth() +

labs(title="log percent expenditure vs life expectancy")

>ggplot(data=life, aes(x=Totalexpenditure)) +

geom\_histogram() +

labs(title="Distribution of total expenditure")

>ggplot(data=life, aes(x=log(Totalexpenditure))) +

geom\_histogram() +

labs(title="Distribution of log total expenditure")

>ggplot(data=life, aes(x=Totalexpenditure, y= Lifeexpectancy)) +

geom\_point() +

geom\_smooth() +

labs(title="Total expenditure vs life expectancy")

>ggplot(data=life, aes(x=log(Totalexpenditure), y= Lifeexpectancy)) +

geom\_point() +

geom\_smooth() +

labs(title="log Total expenditure vs life expectancy")

>ggplot(data=life, aes(x=GDP)) +

geom\_histogram() +

labs(title="GDP distribution")

>ggplot(data=life, aes(x=log(GDP))) +

geom\_histogram() +

labs(title="log GDP distribution")

>ggplot(data=life, aes(x=GDP, y= Lifeexpectancy)) +

geom\_point() +

geom\_smooth() +

labs(title="GDP vs life expectancy")

>ggplot(data=life, aes(x=log(GDP), y= Lifeexpectancy)) +

geom\_point() +

geom\_smooth() +

labs(title="log GDP vs life expectancy")

>ggplot(data=life, aes(x=Population)) +

geom\_histogram() +

labs(title="Distribution of population")

>ggplot(data=life, aes(x=log(Population))) +

geom\_histogram() +

labs(title="Distribution of log population")

>ggplot(data=life, aes(x=Population, y= Lifeexpectancy)) +

geom\_point() +

geom\_smooth() +

labs(title="Population vs life expectancy")

>ggplot(data=life, aes(x=log(Population), y= Lifeexpectancy)) +

geom\_point() +

labs(title="log Population vs life expectancy")

life\_expectancy2 <- life\_expectancy %>%

mutate(logGDP = log(GDP))

life\_expectancy3 <- life\_expectancy2 %>%

mutate(logpercentageexpenditure = log(percentageexpenditure))

life\_expectancy4 <- life\_expectancy3 %>%

mutate(logTotalexpenditure = log(Totalexpenditure))

life\_expectancy5 <- life\_expectancy4 %>%

mutate(logPopulation = log(Population))

ggplot(life\_expectancy5, aes(logPopulation, Lifeexpectancy)) +

geom\_hex(bins = 20, color = "white")+

scale\_fill\_gradient(low = "#00AFBB", high = "#FC4E07")+

theme\_minimal()

>ggplot(data=life, aes(x=Incomecompositionofresources)) +

geom\_histogram() +

labs(title="Distribution of income Composition of resources")

>ggplot(data=life, aes(x=log(Incomecompositionofresources))) +

geom\_histogram() +

labs(title="Income Composition of resources")

>ggplot(data=life, aes(x=Incomecompositionofresources, y= Lifeexpectancy)) +

geom\_point() +

labs(title="Incomecompositionofresources vs life expectancy")

>ggplot(data=life, aes(x=log(Incomecompositionofresources), y= Lifeexpectancy)) +

geom\_point() +

labs(title="Incomecompositionofresources vs life expectancy")

>ggplot(data=life, aes(x=Schooling)) +

geom\_histogram() +

labs(title="Distribution of Schooling")

>ggplot(data=life, aes(x=Schooling, y= Lifeexpectancy, col=Status)) +

geom\_point() +

geom\_smooth(method = "lm")

>life %>%

group\_by(Year) %>%

summarise(mean.hiv = mean(`HIV/AIDS`)) %>%

ggplot(aes(x=Year,y=mean.hiv)) +

geom\_point() +

labs(y = "Deaths per 1,000 live births HIV/AIDS (0-4 years)", title = "Mean over years") +

geom\_line(color="blue")

>by.year <- life %>%

na.omit() %>%

group\_by(Year, Status) %>%

summarise(mean.life = mean(Lifeexpectancy),

mean.percentexp = mean(percentageexpenditure),

mean.total.exp = mean(Totalexpenditure),

mean.GDP = mean(GDP),

mean.logGDP = log(mean(GDP)),

mean.schooling = mean(Schooling))

>ggplot(data = by.year, aes(x=Year, y=mean.GDP, color = Status)) +

geom\_point() +

geom\_smooth() +

labs(y="Mean GDP", title = "GDP over years")

>ggplot(data = by.year, aes(x=Year, y=mean.percentexp, color = Status)) +

geom\_point() +

geom\_line() +

labs(y = "Mean Percentage of Expenditure on Health", title = "Percentage of Expenditure on Health over the years")

>ggplot(data = by.year, aes(x=Year, y=mean.total.exp, color = Status)) +

geom\_point() +

geom\_line() +

labs(y = "Mean Total Expenditure on Health", title = "Total Expenditure on Health over the years")

>by.status <- life %>%

group\_by(Year, Status) %>%

summarise(mean.life = mean(Lifeexpectancy))

>ggplot(data = by.year, aes(x=Year, y=mean.GDP)) +

geom\_point() +

geom\_smooth(method = "lm") +

labs(y="Mean GDP", title = "Mean GDP over time")

>ggplot(data = by.year, aes(x = Year, y=mean.percentexp)) +

geom\_point() +

geom\_smooth()

>ggplot(data = life, aes(x=Year, y=log(GDP))) +

geom\_point() + facet\_wrap(~Status,ncol=2) +

labs(y="Logarithm of GDP")

>gdp2000 <- life %>% filter(Year == 2000) %>%

ggplot(aes(x=log(GDP),y=Lifeexpectancy, color = Status)) +

geom\_point() +

geom\_smooth(method = "lm") +

theme(legend.position = "none") +

labs(y="Life Expectancy", title = "log(GDP) effect in 2000")

>gdp2015 <- life %>% filter(Year == 2015) %>%

ggplot(aes(x=log(GDP),y=Lifeexpectancy, color = Status)) +

geom\_point() +

geom\_smooth(method = "lm") +

theme(legend.position = "none") +

labs(y = "Life Expectancy", title = "log(GDP) effect in 2015")

>grid.arrange(gdp2000, gdp2015, ncol = 2)

>totalexp2000 <- life %>% filter(Year == 2000) %>%

ggplot(aes(x=log(Totalexpenditure),y=Lifeexpectancy, color = Status)) +

geom\_point() +

geom\_smooth(method = "lm") +

theme(legend.position = "none") +

labs(y="Life Expectancy", title = "log(Total Expenditure) effect in 2000")

>totalexp2014 <- life %>% filter(Year == 2014) %>%

ggplot(aes(x=log(Totalexpenditure),y=Lifeexpectancy, color = Status)) +

geom\_point() +

geom\_smooth(method = "lm") +

theme(legend.position = "none") +

labs(y = "Life Expectancy", title = "log(Total Expenditure) effect in 2014")

>grid.arrange(totalexp2000, totalexp2014, ncol = 2)

>socioeco <- life %>% na.omit() %>%

select(Lifeexpectancy, percentageexpenditure, Totalexpenditure, GDP, Population, Schooling, Incomecompositionofresources)

>cor(socioeco)

>per.exp2000 <- life %>% filter(Year == 2000) %>%

ggplot(aes(x=log(percentageexpenditure),y=Lifeexpectancy, color = Status)) +

geom\_point() +

geom\_smooth(method = "lm") +

theme(legend.position = "none") +

labs(y="Life Expectancy", title = "Percentage Expenditure effect in 2000")

>per.exp2014 <- life %>% filter(Year == 2014) %>%

ggplot(aes(x=log(percentageexpenditure),y=Lifeexpectancy, color = Status)) +

geom\_point() +

geom\_smooth(method = "lm") +

theme(legend.position = "none") +

labs(y = "Life Expectancy", title = "Percentage Expenditure effect in 2014")

>grid.arrange(per.exp2000,per.exp2014, ncol = 2)

>incom2000 <- life %>% filter(Year == 2000, Incomecompositionofresources != 0) %>%

ggplot(aes(x=Incomecompositionofresources,y=Lifeexpectancy, color = Status)) +

geom\_point() +

geom\_smooth(method = "lm") +

theme(legend.position = "none") +

labs(y="Life Expectancy",x = "Human Development Index", title = "Income Composition of Resources in 2000")

>incom2010 <- life %>% filter(Year == 2010, Incomecompositionofresources != 0) %>%

ggplot(aes(x=Incomecompositionofresources,y=Lifeexpectancy, color = Status)) +

geom\_point() +

geom\_smooth(method = "lm") +

theme(legend.position = "none") +

labs(y="Life Expectancy",x = "Human Development Index", title = "Income Composition of Resources in 2010")

>incom2015 <- life %>% filter(Year == 2015, Incomecompositionofresources != 0) %>%

ggplot(aes(x=Incomecompositionofresources,y=Lifeexpectancy, color = Status)) +

geom\_point() +

geom\_smooth(method = "lm") +

theme(legend.position = "none") +

labs(y="Life Expectancy", x = "Human Development Index",title = "Income Composition of Resources in 2015")

>grid.arrange(incom2000, incom2010, incom2015, ncol = 3)

>school2015 <- life %>% filter(Year == 2015) %>%

ggplot(aes(x=Schooling,y=Lifeexpectancy, color = Status)) +

geom\_point() +

geom\_smooth(method = "lm") +

theme(legend.position = "none") +

labs(y="Life Expectancy", x = "Years of Schooling",title = "Schooling Effect in 2015")

>school2010 <- life %>% filter(Year == 2010) %>%

ggplot(aes(x=Schooling,y=Lifeexpectancy, color = Status)) +

geom\_point() +

geom\_smooth(method = "lm") +

theme(legend.position = "none") +

labs(y="Life Expectancy", x = "Years of Schooling",title = "Schooling Effect in 2010")

>school2000 <- life %>% filter(Year == 2000) %>%

ggplot(aes(x=Schooling,y=Lifeexpectancy, color = Status)) +

geom\_point() +

geom\_smooth(method = "lm") +

theme(legend.position = "none") +

labs(y="Life Expectancy", x = "Years of Schooling",title = "Schooling Effect in 2000")

>grid.arrange(school2000, school2010, school2015, ncol = 3)

>life %>% filter(Year != 2015) %>%

ggplot(aes(x=log(percentageexpenditure),y=Lifeexpectancy, color = Status)) + facet\_wrap(~ Year) +

geom\_point() +

geom\_smooth(method = "lm") +

theme(legend.position = "none") +

labs(y="Life Expectancy", x = "log(Percentage Expenditure)")