STA 141A Group Project: World GDP Growth and Population Growth Analysis

I. Group Member

Name	Email	ID
Jingqi Chen	cjqen@ucdavis.edu	998841054
Yili Wang	ylwang@ucdavis.edu	999316533
Hao Wu	hawu@ucdavis.edu	914446556
Monireangsey Taing	tmktaing@ucdavis.edu	912979232

II. Role of the group members

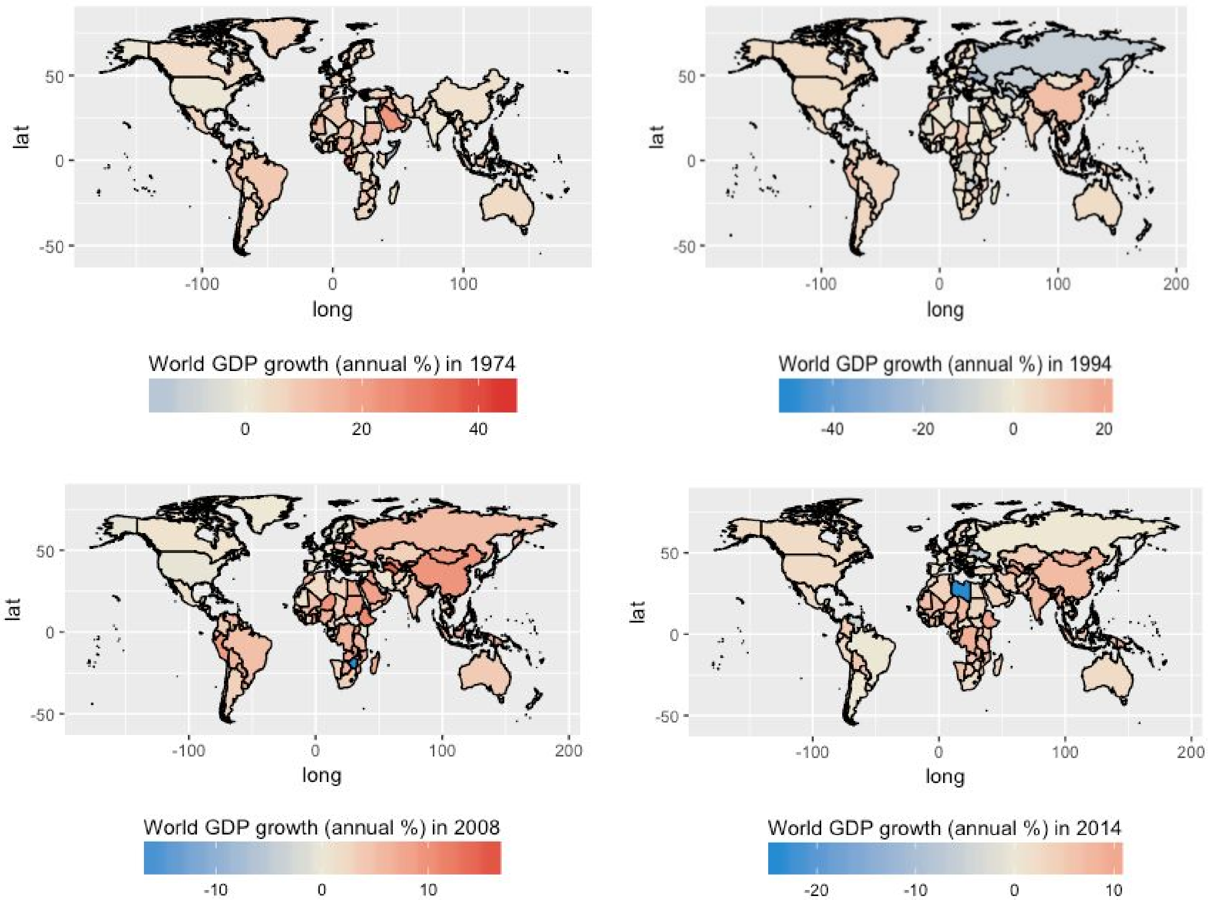
Hao Wu: Use ggplot2 and maps to display world annual GDP growth and world total population in 1974, 1994, 2008 and 2014.
Jingqi Chen & Yili Wang: Examine relationship between Population Growth Rate and its relevant indicators, particularly Birth Rate and various Mortality rates.
Monireangsey Taing: Use time series analysis to predict the population and the GDP for the next 10 years with confidence band for USA

Introduction

In the data of the World Development Indicators from the World Bank, it contains over a thousand annual indicators of economic development from hundreds of countries around the world. In the project, we explore the world GDP growth, the population, and other indicators from selected years which reflecting four phrase of world development: 1974, 1994, 2008, 2014. The purposes of the project are to observe the changes of GDP and population from countries over the world and to predict the these variables from a selected country, which is the United States. The methodology to explore the data is using function ggplot for plotting the world map and other plots for showing the data, and employing the time series analysis to predict the value.

World GDP Growth Rate

We examine the GDP growth rate distribution all over the world.

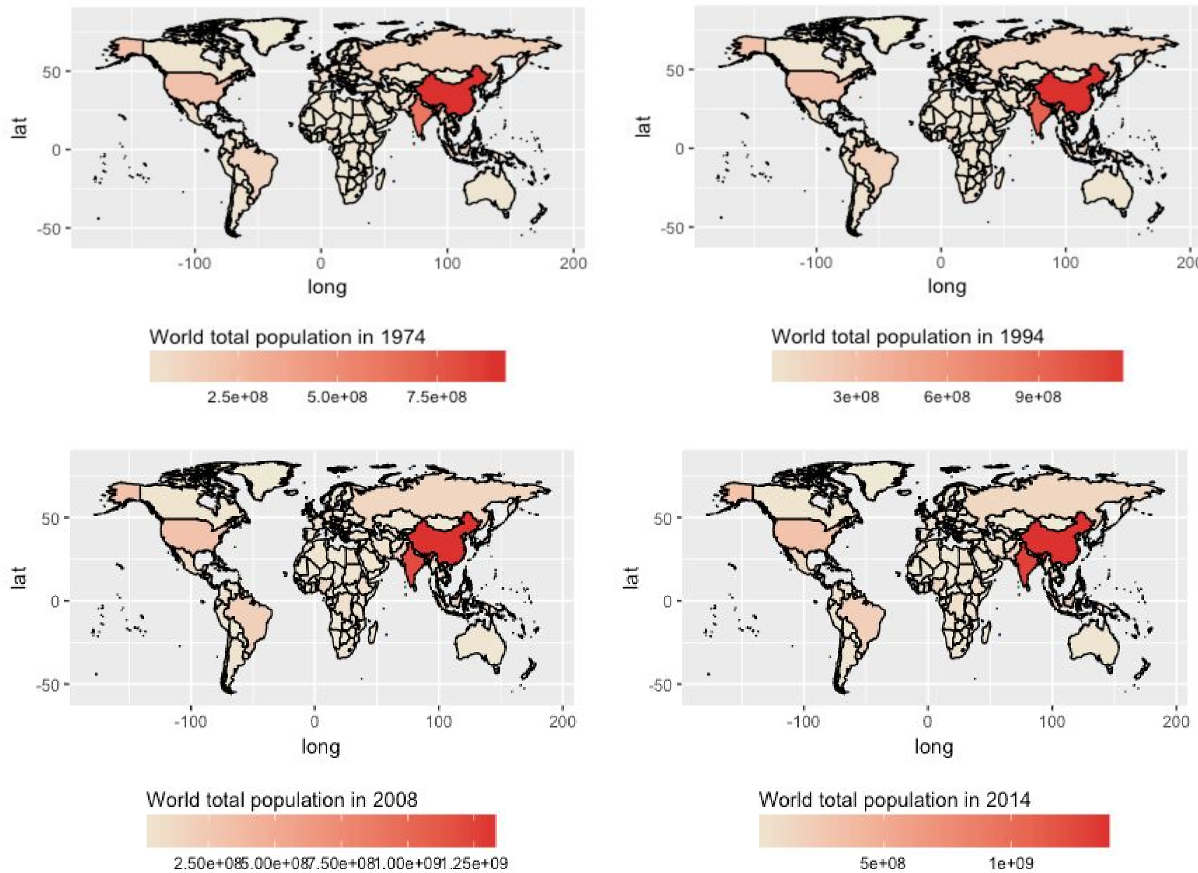


(Graph1: World GDP Growth Rate in 1974,1994,2008, and 2014)

Based on Graph 1, we can observe comprehensive political changes and economics development in different regions in the world. In 1974, Middle East was the region with highest GDP growth rate and continent of America had relatively high GDP growth. In 1994, we can observe that the area of Russia had negative GDP growth rate and it might be the aftermath of Cold War. Meanwhile, China was apparently having high GDP growth rate, compared to other countries. As time goes on, in 2008, the whole continent of East Asia, some regions in Africa, and Southern Americas had rapidly growing GDP. In 2014, the redness of regions like China and Africa was gradually reducing. Most noticeably, Libya had turned to completely negative GDP growth rate and it was possibly resulted from warfares.

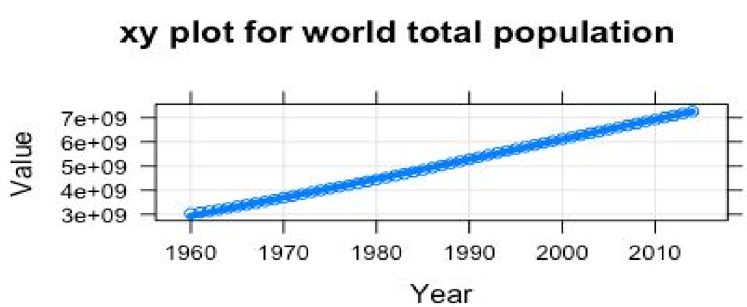
World Population Growth Rate

Next, we focus on the population development happened all around world in the last forty years.



(Graph2: World Total Population in 1974, 1994, 2008, and 2014)

According to Graph2, across the four time phrase, China and India were having highest total population. Countries like America, Russia and Brazil also had relatively high total population. The number of population in each region were reducing gradually. By the Graph2, China held the highest population from 1974 to 2014, whereas the countries in Africa continent have the smallest population among all the continent in the world.

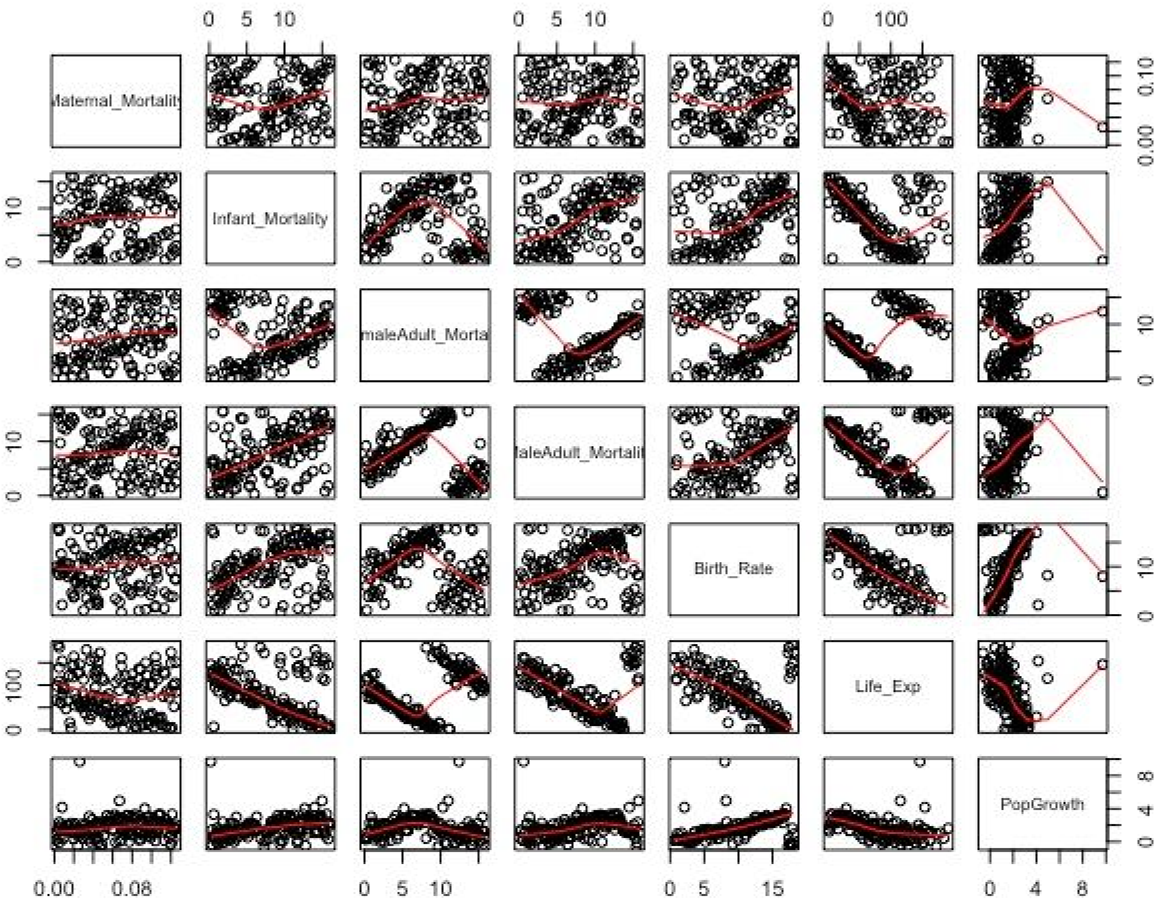


(Graph3: Total Population vs. Time)

From Graph3, we can conclude a self-evident fact that from 1960 to 2010, the world total population has been continuously increasing.

Population Growth Rate and Potential Related Factors

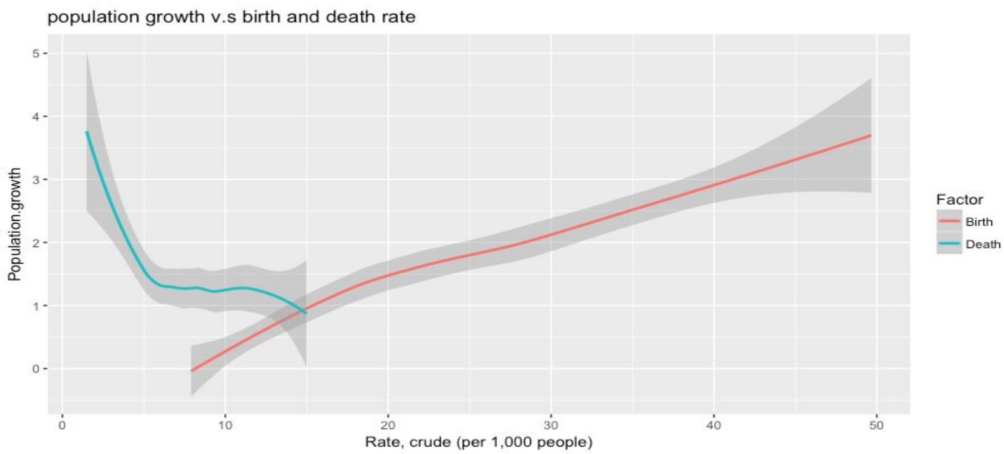
After having a general understanding towards the world population, we are going to explore the factors that potentially related to population growth rate. We estimate the following factors' relationship: Maternal mortality, Infant mortality, Male adult mortality, Female adult mortality, Birth rate, Life expectancy, and Population growth.



(Graph 4: Linear relationship between population growth and its related variables)

In Graph 4, we can take a close look at the pair relationship between these factors and we can find out easily that Female adult mortality and Infant mortality are linearly related. Also, Birth rate and Life expectancy are linearly related. This can be explained by the fact that both birth rate and life expectancy can reflect a country's health care levels and medical technologies, since new-born babies and the elderly are the major target groups of health care plans and medical treatment.

In addition, we want to figure out some elements that are related to population growth. Below in Graph 5, the red line represents population growth rate v.s birth rate, crude (per 1,000 people), the blue line represents population growth rate v.s death rate, crude (per 1,000 people). We can see that there is a linear relationship between birth rate and population growth rate. However, there is no obvious linear relationship between death rate and population growth rate. From the graph5, we can tell that as the death rate increases, population growth rate decreases. Contrarily, as the birth rate increases, the population growth rate is always increases simultaneously.



(Graph 5: Population growth vs Birth rate and Death rate)

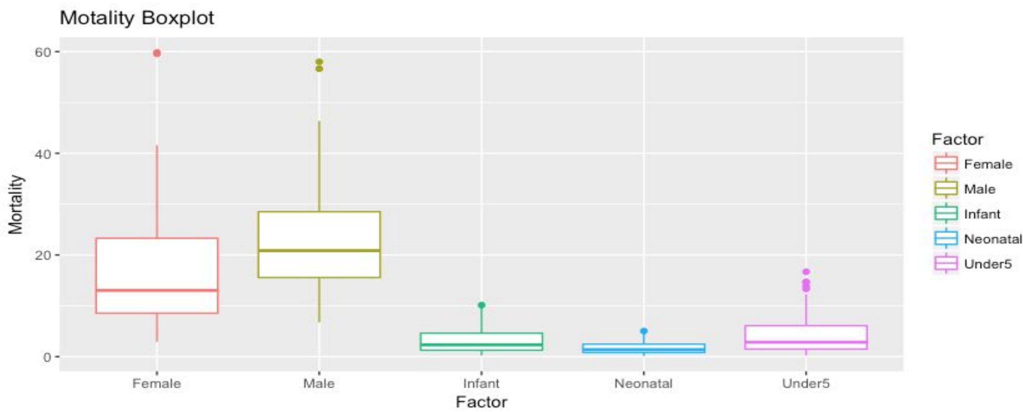
We want to fit a model to see if it can birth rate can be a good estimate of population growth rate

```
> model1 = lm(Population.growth~Birth.rate,birth_death_population)
> summary(model1)

Call:
lm(formula = Population.growth ~ Birth.rate, data = birth_death_population)

Residuals:
    Min      1Q  Median      3Q     Max 
-1.7657 -0.3455 -0.0186  0.2167  8.4001 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) -0.452836   0.140455  -3.224  0.00148 ** 
Birth.rate   0.086599   0.005926  14.613  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8829 on 199 degrees of freedom
Multiple R-squared:  0.5176,	Adjusted R-squared:  0.5152 
F-statistic: 213.5 on 1 and 199 DF,  p-value: < 2.2e-16
```

The p-value is less than 0.05, therefore it is not a good estimate for population growth rate.

From the result above we want to get a closer look at how the different mortalities rate affect population growth rate. For children (infant, neonatal, and under 5), their mortality rates are generally less than those of the adults. From graph 6, there is no linear relationship between population growth rate and each group's mortality rate. From graph 7, Male seems to have higher mortality rate than female.

(Graph 6: Population Growth vs. Different Mortality Groups)

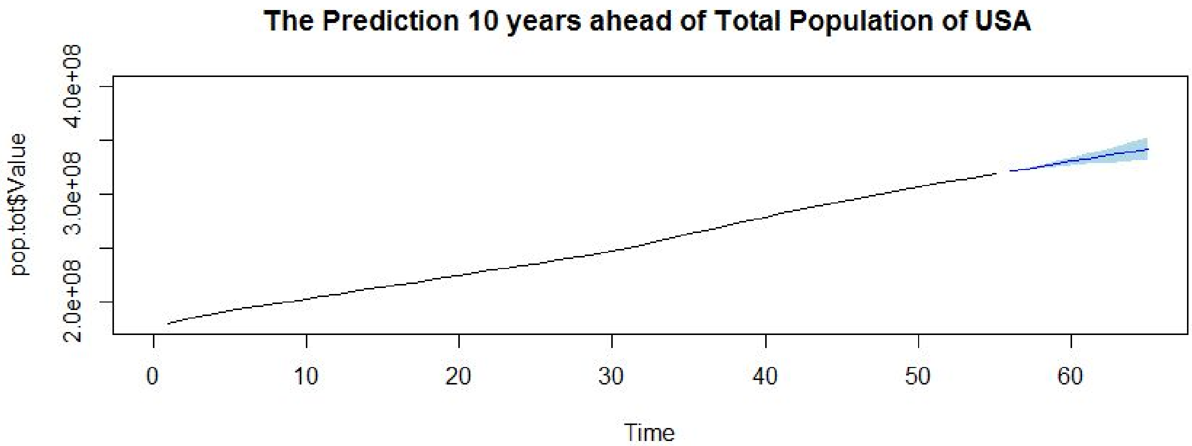


(Graph 7: Mortality Boxplot)

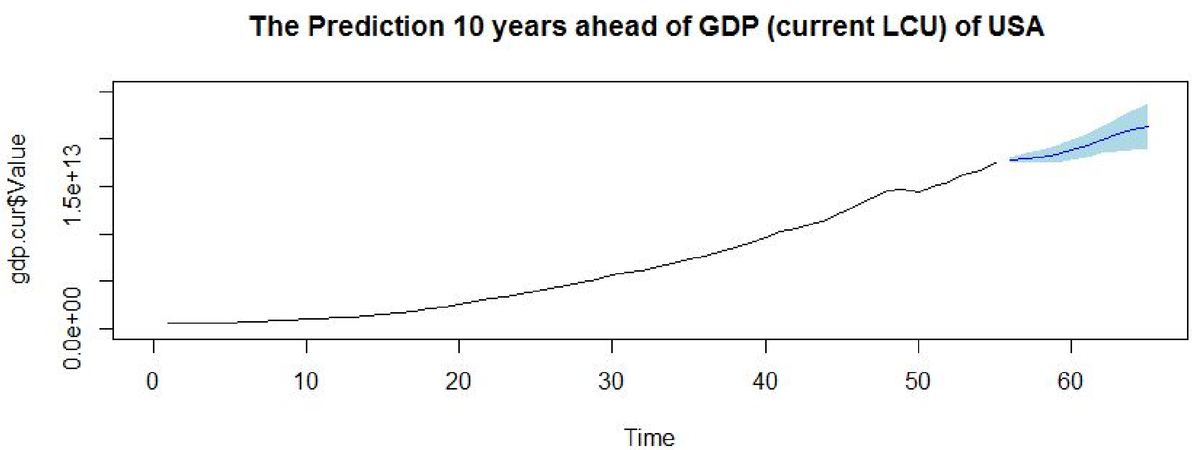
Time Series Analysis of Population and GDP of the United States

Now, we plot the time series of the total population and the GDP (current LCU) of USA from 1960 to 2014, which are indicated from 0 to 55. Before we predict the population and GDP, we would like to have a Box-Ljung test to see whether the residual of population and GDP is identically and independently distributed. By the Box-Ljung test with lag of 10, the p-value of population and GDP is 0.9538 and 0.7699, respectively, which is greater than any value of significant level of alpha. Thus, we fail to reject the null hypothesis that all autocorrelation of a time series is zero, which means the residual is identically and independently distributed.

By using function auto.arima(), we obtain the autoregressive integrated moving average (ARIMA) of the 1st order Auto-Regressive and the 2nd order Moving Average model for population and ARIMA of 2nd order Auto-Regressive and the 2nd order Moving Average model for GDP. Through using function predict(), we obtain the prediction in the future 10 years for the population and GDP. From Graph 8 and Graph 9, we can tell that both population and the GDP is growing constantly. Therefore, we expect the population and the GDP is increasing in the next 10 years



(Graph 8: The Prediction 10 years ahead of Total Population of USA)



(Graph 9: The Prediction 10 years ahead of GDP of USA)

Conclusion

In this report, we first illustrate World GDP Growth Rate and World Population Growth Rate. Then we explore the factors that potentially related to population growth rate, especially focused on Maternal mortality, Infant mortality, Male adult mortality, Female adult mortality, Birth rate, Life expectancy. In the end, we perform a time series analysis of population and GDP of the United States. From the result, the overall world population is increasing constantly although some countries have a negative growth rate in population due to the war, disease, and other factors. With a further understanding in world population, insteading of using the world population variable, we use the growth rate, comparing with different growth stage. We find that infant has the lowest death rate with the highest growth rate, while adult has the highest death rate with the lowest growth rate.

Appendix:

```
library(ggmap)
library(maps)
library(ggplot2)
library(plyr)
library(dplyr)
library(utils)
library(igraph)
library(lattice)

data=read.csv("~/Desktop/rdata/world-development-indicators2/Indicators.csv",stringsAsFactors=FALSE)
data_2014 = subset(data,Year==2014)
data_2008 = subset(data,Year==2008)
data_1994 = subset(data,Year==1994)
data_1974 = subset(data,Year==1974)

##world total gdp growth
#select row after WLD
world =  map_data("world")
gdpgrowth_2014 = subset(data_2014, IndicatorCode=="NY.GDP.MKTP.KD.ZG")
gdpgrowth_2008 = subset(data_2008, IndicatorCode=="NY.GDP.MKTP.KD.ZG")
gdpgrowth_1994 = subset(data_1994, IndicatorCode=="NY.GDP.MKTP.KD.ZG")
gdpgrowth_1974 = subset(data_1974, IndicatorCode=="NY.GDP.MKTP.KD.ZG")
#correction of the name from kaggle post:
#https://www.kaggle.com/benhamner/d/worldbank/world-development-indicators/indicators-in-data/comments
correction = c("Antigua and Barbuda"="Antigua", "Bahamas, The"="Bahamas",
        "Brunei Darussalam"="Brunei", "Cabo Verde"="Cape Verde",
        "Congo, Dem. Rep."="Democratic Republic of the Congo",
        "Congo, Rep."="Republic of Congo", "Cote d'Ivoire"="Ivory Coast",
        "Egypt, Arab Rep."="Egypt", "Faeroe Islands"="Faroe Islands",
        "Gambia, The"="Gambia", "Iran, Islamic Rep."="Iran",
        "Korea, Dem. Rep."="North Korea", "Korea, Rep."="South Korea",
        "Kyrgyz Republic"="Kyrgyzstan", "Lao PDR"="Laos",
        "Macedonia, FYR"="Macedonia", "Micronesia, Fed. Sts."="Micronesia",
        "Russian Federation"="Russia", "Slovak Republic"="Slovakia",
        "St. Lucia"="Saint Lucia", "St. Martin (French part)"="Saint Martin",
        "St. Vincent and the Grenadines"="Saint Vincent",
        "Syrian Arab Republic"="Syria", "Trinidad and Tobago"="Trinidad",
        "United Kingdom"="UK", "United States"="USA",
        "Venezuela, RB"="Venezuela", "Virgin Islands (U.S.)"="Virgin Islands",
        "Yemen, Rep."="Yemen")
for (c in names(correction)) {
  gdpgrowth_2014$CountryName[which(gdpgrowth_2014$CountryName==c)] = correction[c]
  gdpgrowth_2008$CountryName[which(gdpgrowth_2008$CountryName==c)] = correction[c]
  gdpgrowth_1994$CountryName[which(gdpgrowth_1994$CountryName==c)] = correction[c]
  gdpgrowth_1974$CountryName[which(gdpgrowth_1974$CountryName==c)] = correction[c]
}
```

```
names(gdpgrowth_2014)[1] = "region"
names(gdpgrowth_2008)[1] = "region"
names(gdpgrowth_1994)[1] = "region"
names(gdpgrowth_1974)[1] = "region"

#match region
worldgdpgrowth_2014 = inner_join(world, gdpgrowth_2014, by = "region")
worldgdpgrowth_2008 = inner_join(world, gdpgrowth_2008, by = "region")
worldgdpgrowth_1994 = inner_join(world, gdpgrowth_1994, by = "region")
worldgdpgrowth_1974 = inner_join(world, gdpgrowth_1974, by = "region")

ggplot() +
  geom_polygon(data = worldgdpgrowth_1974, aes(x=long, y = lat, group = group,fill=Value), color="black") +
  coord_fixed(1.3) +
  theme(legend.position="bottom") +
  guides(fill = guide_colorbar(barwidth=13, title.position = "top", direction = "horizontal")) +
  scale_fill_gradient2(midpoint = 0.5, mid="#eee8d5", high="#dc322f", low="#268bd2",
                   "World GDP growth (annual %) in 1974")
ggplot() +
  geom_polygon(data = worldgdpgrowth_1994, aes(x=long, y = lat, group = group,fill=Value), color="black") +
  coord_fixed(1.3) +
  theme(legend.position="bottom") +
  guides(fill = guide_colorbar(barwidth=13, title.position = "top", direction = "horizontal")) +
  scale_fill_gradient2(midpoint = 0.5, mid="#eee8d5", high="#dc322f", low="#268bd2",
                   "World GDP growth (annual %) in 1994")
ggplot() +
  geom_polygon(data = worldgdpgrowth_2008, aes(x=long, y = lat, group = group,fill=Value), color="black") +
  coord_fixed(1.3) +
  theme(legend.position="bottom") +
  guides(fill = guide_colorbar(barwidth=13, title.position = "top", direction = "horizontal")) +
  scale_fill_gradient2(midpoint = 0.5, mid="#eee8d5", high="#dc322f", low="#268bd2",
                   "World GDP growth (annual %) in 2008")
ggplot() +
  geom_polygon(data = worldgdpgrowth_2014, aes(x=long, y = lat, group = group,fill=Value), color="black") +
  coord_fixed(1.3) +
  theme(legend.position="bottom") +
  guides(fill = guide_colorbar(barwidth=13, title.position = "top", direction = "horizontal")) +
  scale_fill_gradient2(midpoint = 0.5, mid="#eee8d5", high="#dc322f", low="#268bd2",
                   "World GDP growth (annual %) in 2014")

##world total population
pop_2014 = subset(data_2014, IndicatorCode=="SP.POP.TOTL")
pop_2008 = subset(data_2008, IndicatorCode=="SP.POP.TOTL")
pop_1994 = subset(data_1994, IndicatorCode=="SP.POP.TOTL")
pop_1974 = subset(data_1974, IndicatorCode=="SP.POP.TOTL")
for (c in names(correction)) {
  pop_2014$CountryName[which(pop_2014$CountryName==c)] = correction[c]
```

```
  pop_2008$CountryName[which(pop_2008$CountryName==c)] = correction[c]
  pop_1994$CountryName[which(pop_1994$CountryName==c)] = correction[c]
  pop_1974$CountryName[which(pop_1974$CountryName==c)] = correction[c]
}
names(pop_2014)[1] = "region"
names(pop_2008)[1] = "region"
names(pop_1994)[1] = "region"
names(pop_1974)[1] = "region"

#match region
worldpop_2014 = inner_join(world, pop_2014, by = "region")
worldpop_2008 = inner_join(world, pop_2008, by = "region")
worldpop_1994 = inner_join(world, pop_1994, by = "region")
worldpop_1974 = inner_join(world, pop_1974, by = "region")

ggplot() +
  geom_polygon(data = worldpop_1974, aes(x=long, y = lat, group = group,fill=Value), color="black") +
  coord_fixed(1.3) +
  theme(legend.position="bottom") +
  guides(fill = guide_colorbar(barwidth=13, title.position = "top", direction = "horizontal")) +
  scale_fill_gradient2(midpoint = 0.5, mid="#eee8d5", high="#dc322f", low="#268bd2",
                    "World total population in 1974")
ggplot() +
  geom_polygon(data = worldpop_1994, aes(x=long, y = lat, group = group,fill=Value), color="black") +
  coord_fixed(1.3) +
  theme(legend.position="bottom") +
  guides(fill = guide_colorbar(barwidth=13, title.position = "top", direction = "horizontal")) +
  scale_fill_gradient2(midpoint = 0.5, mid="#eee8d5", high="#dc322f", low="#268bd2",
                    "World total population in 1994")
ggplot() +
  geom_polygon(data = worldpop_2008, aes(x=long, y = lat, group = group,fill=Value), color="black") +
  coord_fixed(1.3) +
  theme(legend.position="bottom") +
  guides(fill = guide_colorbar(barwidth=13, title.position = "top", direction = "horizontal")) +
  scale_fill_gradient2(midpoint = 0.5, mid="#eee8d5", high="#dc322f", low="#268bd2",
                    "World total population in 2008")
ggplot() +
  geom_polygon(data = worldpop_2014, aes(x=long, y = lat, group = group,fill=Value), color="black") +
  coord_fixed(1.3) +
  theme(legend.position="bottom") +
  guides(fill = guide_colorbar(barwidth=13, title.position = "top", direction = "horizontal")) +
  scale_fill_gradient2(midpoint = 0.5, mid="#eee8d5", high="#dc322f", low="#268bd2",
                    "World total population in 2014")

world = subset(data,CountryCode=="WLD")
totalpopulation = subset(world,IndicatorCode=="SP.POP.TOTL")
xyplot(Value ~ Year, totalpopulation,
```

```
      grid = TRUE,
      auto.key = TRUE,
      type = c("p", "smooth"), lwd = 4,
      main = "xy plot for world total population")


# Population Growth Rate and Potential Related Factors

# extract 2013 data
data_2013 = data[which(data$Year=="2013"),]
# group
data_2013_group = data_2013[c(1:19541),]
# extract country
data_2013_countries = data_2013[-c(1:19541),]
# total population
data_2013_population_total = data_2013_countries[grepl("Population,
total",data_2013_countries$IndicatorName),]
data_2013_population_total = data.frame(CountryName = data_2013_population_total$CountryName,
                                        Total.population = data_2013_population_total$Value)
# population growth
data_2013_population_growth = data_2013_countries[grepl("Population
growth",data_2013_countries$IndicatorName),]
data_2013_population_growth = data.frame(CountryName = data_2013_population_growth$CountryName,
                                         Population.growth = data_2013_population_growth$Value)

# population data(population growth & total population)
population_2013 = merge.data.frame(data_2013_population_growth,data_2013_population_total)

##Population ages 65 and above (% of total)
data_2013_population_65 = data_2013_countries[which("Population ages 65 and above (% of
total)"==data_2013_countries$IndicatorName),]
data_2013_population_65= data.frame(CountryName = data_2013_population_65$CountryName,
                                    Population.growth = data_2013_population_65$Value)

## Birth rate
data_2013_birth_rate = data_2013_countries[grepl("Birth rate",data_2013_countries$IndicatorName),]
data_2013_birth_rate = data.frame(CountryName = data_2013_birth_rate$CountryName,
                                  Birth.rate = data_2013_birth_rate$Value)
## Death rate
data_2013_death_rate = data_2013_countries[grepl("Death rate",data_2013_countries$IndicatorName),]
data_2013_death_rate = data.frame(CountryName = data_2013_death_rate$CountryName,
                                  Death.rate = data_2013_death_rate$Value)

## Mortality
data_2013_mortality_female = data_2013_countries[grepl("Mortality rate, adult,
female",data_2013_countries$IndicatorName),]
data_2013_mortality_female = c(CountryName = data_2013_mortality_female$CountryName,
```

```
                                    mortality.female = data_2013_mortality_female$Value)
female = data.frame(Mortality = mortality$mortality.female,rep("Female", Factor =
length(mortality$mortality.female)))
data_2013_mortality_male = data_2013_countries[grepl("Mortality rate, adult,
male",data_2013_countries$IndicatorName),]
data_2013_mortality_male = c(CountryName = data_2013_mortality_male$CountryName,
                                    mortality.male = data_2013_mortality_male$Value)
data_2013_mortality_infant = data_2013_countries[grepl("Mortality rate,
infant",data_2013_countries$IndicatorName),]
data_2013_mortality_infant = data.frame(CountryName = data_2013_mortality_infant$CountryName,
                                  mortality.infant = data_2013_mortality_infant$Value)
data_2013_mortality_neonatal = data_2013_countries[grepl("Mortality rate,
neonatal",data_2013_countries$IndicatorName),]
data_2013_mortality_neonatal = data.frame(CountryName = data_2013_mortality_neonatal$CountryName,
                                    mortality.neonatal = data_2013_mortality_neonatal$Value)
data_2013_mortality_under5 = data_2013_countries[grepl("Mortality rate,
under-5",data_2013_countries$IndicatorName),]
data_2013_mortality_under5 = data.frame(CountryName = data_2013_mortality_under5$CountryName,
                                     mortality.under5 = data_2013_mortality_under5$Value)
# mortality data frame
mortality = merge.data.frame(data_2013_mortality_female,data_2013_mortality_male)
mortality = merge.data.frame(mortality,data_2013_mortality_infant)
mortality = merge.data.frame(mortality,data_2013_mortality_neonatal)
mortality = merge.data.frame(mortality,data_2013_mortality_under5)

# Birth rate, mortality rate & population data frame
birth_mortality_population = merge.data.frame(data_2013_birth_rate,
                                    mortality)
birth_mortality_population = merge.data.frame(birth_mortality_population,
                                         population_2013)

# Birth rate, Death rate & population data frame
birth_death_population = merge.data.frame(data_2013_birth_rate,
                                     data_2013_death_rate)
birth_death_population = merge.data.frame(birth_death_population,
                                         population_2013)
birth = data.frame(Rate = birth_death_population$Birth.rate, Population.growth =
birth_death_population$Population.growth, Factor = c(rep("Birth",length(
birth_death_population$Population.growth))))
death = data.frame(Rate = birth_death_population$Death.rate, Population.growth =
birth_death_population$Population.growth, Factor = c(rep("Death",length(
birth_death_population$Population.growth))))
rate = rbind.data.frame(birth, death)

# population growth v.s birth and death rate
bdp <- ggplot(rate, aes(Rate, Population.growth,color = Factor))
```

```
bdp + geom_smooth() + ggtitle("population growth v.s birth and death rate") + xlab("Rate, crude (per 1,000 people)")

#fit model
model1 = lm(Population.growth~Birth.rate,birth_death_population)
summary(model1)

# population growth v.s different mortality group
female = data.frame(Population.growth = birth_mortality_population$Population.growth,Mortality = (mortality$mortality.female/1000*100),Factor = rep("Female", length(mortality$mortality.female)))
male = data.frame(Population.growth = birth_mortality_population$Population.growth, Mortality = (mortality$mortality.male/1000*100),Factor = rep("Male", length(mortality$mortality.male)))
infant = data.frame(Population.growth = birth_mortality_population$Population.growth, Mortality = (mortality$mortality.infant/1000*100),Factor = rep("Infant", length(mortality$mortality.infant)))
neonatal = data.frame(Population.growth = birth_mortality_population$Population.growth, Mortality = (mortality$mortality.neonatal/1000*100),Factor = rep("Neonatal", length(mortality$mortality.neonatal)))
under5 = data.frame(Population.growth = birth_mortality_population$Population.growth, Mortality = (mortality$mortality.under5/1000*100),Factor = rep("Under5", length(mortality$mortality.under5)))

mor = rbind.data.frame(female,male, infant, neonatal,under5)

mp <- ggplot(mor, aes(Mortality, Population.growth,color = Factor))
mp + geom_smooth(model = lm) + ggtitle("Population Growth v.s Different Mortality Groups")

#mortality boxplot
mpbox <- ggplot(mor, aes(Factor, Mortality, color = Factor))
mpbox + geom_boxplot() + ggtitle("Motality Boxplot")



##time series analysis
data = read.csv("indicators.csv", header = TRUE)
#data filter
library(dplyr)
usa = filter(data, CountryCode == "USA") #only USA
pop.tot = filter(usa, IndicatorCode == "SP.POP.TOTL") #total popular
gdp.cur = filter(usa, IndicatorCode == "NY.GDP.MKTP.CN") #gdp current LCU
library(forecast)
#predicted with confidence interval
mod1 = auto.arima(pop.tot$Value)
res = mod1$res
Box.test(res,lag=10,type="Ljung-Box") #using Box-Ljung test to test IID
n = nrow(pop.tot)
h = 10
fcast = predict(mod1, n.ahead = h) #predict the value
fc = fcast$pred
upper = fc + qnorm(0.975) * fcast$se #upper confidence interval
```

```
lower = fc - qnorm(0.975) * fcast$se #lower confidence interval
plot.ts(pop.tot$Value, xlim = c(0, n+h), ylim = c(180000000, 400000000),
        main = "The Prediction 10 years ahead of Total Population of USA ")
polygon(x = c(n+1:h, n+h:1),y=c(upper,rev(lower)), col = 'lightblue', border=NA)
lines(x = n + (1:h), y=fc, col='blue')

########## Mortality Indicators ############
## Indicator 1: Maternal mortality ratio
data_2013_countries[which(IndicatorName =="Maternal mortality ratio (modeled estimate, per 100,000 live
births)"),] -> Maternal_Mor
data.frame(cbind(Maternal_Mor$CountryName, Maternal_Mor$Value)) -> Maternal_Mor
colnames(Maternal_Mor) <- c("Country","Maternal_Mortality") # rename the columns
# Change the numbers to be measured in Percentage(%)
as.numeric(Maternal_Mor$Maternal_Mortality)/100000 -> Maternal_Mor$Maternal_Mortality
Maternal_Mor$Maternal_Mortality*100 -> Maternal_Mor$Maternal_Mortality

## Indicator 2: Mortality rate, infant (per 1,000 live births)
data_2013_countries[which(IndicatorName =="Mortality rate, infant (per 1,000 live births)"),] -> Infant_Mor
data.frame(cbind(Infant_Mor$CountryName, Infant_Mor$Value)) -> Infant_Mor
colnames(Infant_Mor) <- c("Country","Infant_Mortality") # rename the columns
# Change the numbers to be measured in Percentage(%)
as.numeric(Infant_Mor$Infant_Mortality)/1000 ->Infant_Mor$Infant_Mortality
Infant_Mor$Infant_Mortality*100 -> Infant_Mor$Infant_Mortality

# Indicator 3: Mortality rate, adult, female (per 1,000 female adults)
data_2013_countries[which(IndicatorName =="Mortality rate, adult, female (per 1,000 female adults)"),] ->
FAdult_Mor
data.frame(cbind(FAdult_Mor$CountryName, FAdult_Mor$Value)) -> FAdult_Mor
colnames(FAdult_Mor) <- c("Country","FemaleAdult_Mortality") # rename the columns
# Change the numbers to be measured in Percentage(%)
as.numeric(FAdult_Mor$FemaleAdult_Mortality)/1000 ->FAdult_Mor$FemaleAdult_Mortality
FAdult_Mor$FemaleAdult_Mortality*100 -> FAdult_Mor$FemaleAdult_Mortality

# Indicator 4: Mortality rate, adult, male (per 1,000 male adults)
data_2013_countries[which(IndicatorName =="Mortality rate, adult, male (per 1,000 male adults)"),] ->
MAdult_Mor
data.frame(cbind(MAdult_Mor$CountryName, MAdult_Mor$Value)) -> MAdult_Mor
colnames(MAdult_Mor) <- c("Country","MaleAdult_Mortality") # rename the columns
# Change the numbers to be measured in Percentage(%)
as.numeric(MAdult_Mor$MaleAdult_Mortality)/1000 ->MAdult_Mor$MaleAdult_Mortality
MAdult_Mor$MaleAdult_Mortality*100 -> MAdult_Mor$MaleAdult_Mortality

########## Birth Indicators ############
data_2013_countries[which(IndicatorName =="Birth rate, crude (per 1,000 people)"),] -> BirthRate
data.frame(cbind(BirthRate$CountryName, BirthRate$Value)) -> BirthRate
colnames(BirthRate) <- c("Country","Birth_Rate") # rename the columns
# Change the numbers to be measured in Percentage(%)
```

```
as.numeric(BirthRate$Birth_Rate)/1000 ->BirthRate$Birth_Rate
BirthRate$Birth_Rate*100 -> BirthRate$Birth_Rate

########## Life Expectancy Indicators ############
data_2013_countries[which(IndicatorName =="Life expectancy at birth, total (years)"),] -> LifeExp
data.frame(cbind(LifeExp$CountryName, LifeExp$Value)) -> LifeExp
colnames(LifeExp) <- c("Country","Life_Exp") # rename the columns

########## Population Indicators ############
data_2013_countries[which(IndicatorName =="Population growth (annual %)"),] -> country_PopGrowth_2013
data.frame( Country = country_PopGrowth_2013$CountryName, PopGrowth =
country_PopGrowth_2013$Value) -> country_PopGrowth_2013
colnames(country_PopGrowth_2013) <- c("Country","PopGrowth") # rename the columns
data_2013_countries[which(IndicatorName =="Population density (people per sq. km of land area)"),] ->
country_PopDensity_2013
data.frame(cbind(country_PopDensity_2013$CountryName, country_PopDensity_2013$Value)) ->
country_PopDensity_2013
colnames(country_PopDensity_2013) <- c("Country","PopDensity") # rename the columns
data_2013_countries[which(IndicatorName =="Population, total"),] -> country_PopTotal_2013
data.frame(cbind(country_PopTotal_2013$CountryName, country_PopTotal_2013$Value)) ->
country_PopTotal_2013
colnames(country_PopTotal_2013) <- c("Country","PopTotal") # rename the columns
data_2013_countries[which(IndicatorName =="Population, female (% of total)"),] -> country_PopFemale_2013
data.frame(cbind(country_PopFemale_2013$CountryName, country_PopFemale_2013$Value)) ->
country_PopFemale_2013
colnames(country_PopFemale_2013) <- c("Country","PopFemale") # rename the columns
data_2013_countries[which(IndicatorName =="Urban population growth (annual %)"),] -> country_PopUr_2013
data.frame(cbind(country_PopUr_2013$CountryName, country_PopUr_2013$Value)) -> country_PopUr_2013
colnames(country_PopUr_2013) <- c("Country","PopUrbanGrowth") # rename the columns

# Merge data frames into one data frame
merge.data.frame(Maternal_Mor,Infant_Mor,by = "Country") -> Pop
merge.data.frame(Pop,FAdult_Mor,by = "Country") -> Pop
merge.data.frame(Pop,MAdult_Mor,by = "Country") -> Pop
merge.data.frame(Pop,BirthRate,by = "Country") -> Pop
merge.data.frame(Pop,LifeExp,by = "Country") -> Pop
merge.data.frame(Pop,country_PopGrowth_2013,by = "Country") -> Pop
merge.data.frame(Pop,country_PopDensity_2013,by = "Country") -> Pop
merge.data.frame(Pop,country_PopTotal_2013,by = "Country") -> Pop
merge.data.frame(Pop,country_PopFemale_2013,by = "Country") -> Pop
merge.data.frame(Pop,country_PopUr_2013,by = "Country") -> Pop


attach(Pop)
```

```
#The CIA World Factbook gives the world annual birthrate, mortality rate,
#and growth rate as 1.89%, 0.79%, and 1.096% respectively.

# Divide countries into different categories by Population Growth
length(which(PopGrowth<1.096)) # Low Population Growth Countries : 44
length(which(PopGrowth<2 & PopGrowth>=1.096)) # Medium Population Growth Countries : 44
length(which(PopGrowth>2 & PopGrowth <= 9.7150)) # High Population Growth Countries : 58

# A new column in Pop containing Category of each country
matrix(0, nrow = 147, ncol = 2, byrow = TRUE) -> class
as.character(Pop$Country) -> class[,1]
for(i in 1: 147){ # Create a new column with different categories of countries' population growth
  if(PopGrowth[i]<=1.096){
    "Low"-> class[i,2]
  }
  if((PopGrowth[i]<2 ) & (PopGrowth[i]>1.096)){
    "Medium"-> class[i,2]
  }
  else if (PopGrowth[i]>2){
    "High"-> class[i,2]
  }
}

class
colnames(class) <- c("Country","Category")
merge.data.frame(Pop,class,by = "Country") -> Pop

attach(Pop)

########## Plots #############

pairs(~
Maternal_Mortality+Infant_Mortality+FemaleAdult_Mortality+MaleAdult_Mortality+Birth_Rate+Life_Exp+PopGro
wth,panel = panel.smooth, data = Pop, main = "")
```