CT5100 Data Visualisation : Assignment 3

Surya Balakrishnan Ramakrishnan (18231072) and Sai Krishna Lakshminarayanan (18230229)

4/3/2019

Analysing Ridley's Hypothesis, GDP, population growth and fertilty rate.

Work Split-Up

Member 1

Surya Balakrishnan Ramakrishnan (18231072)

Question 1 (a), plot, code, observation (plot 1)

Question 3 plot, code, observation (Plot 3)

Question 5 plot, code, observation (plot 5)

Report

Member 2

Sai Krishna Lakshminarayanan (18230229)

Question 1 (b), plot, code, observation (plot 2)

Question 4 plot, code, observation (Plot 4)

Question 5 plot, code, observation (plot 6)

Report

Introduction:

Matt Ridley in his book titled *The Rational Optimist: How Prosperity Evolves* claims to have observed the following major trends

- Fertility rates in richer countries are low.
- Fertility rates will converge to 2.1 globally in a few decades.
- Global population would stagnate at around the 9 Billion count.
- Developing counties are showing a decrease in average fertility.

Ridley also suggested that due to the above mentioned trents if the fertility rates dips past the 2.1 mark then the population will sink immensely over time in the region.

Before analysing the fertility rate lets analyse the trends with GDP and average life expectancy rate across different regions.

Part 1 Plots indicating growth in GDP and life expectancy.

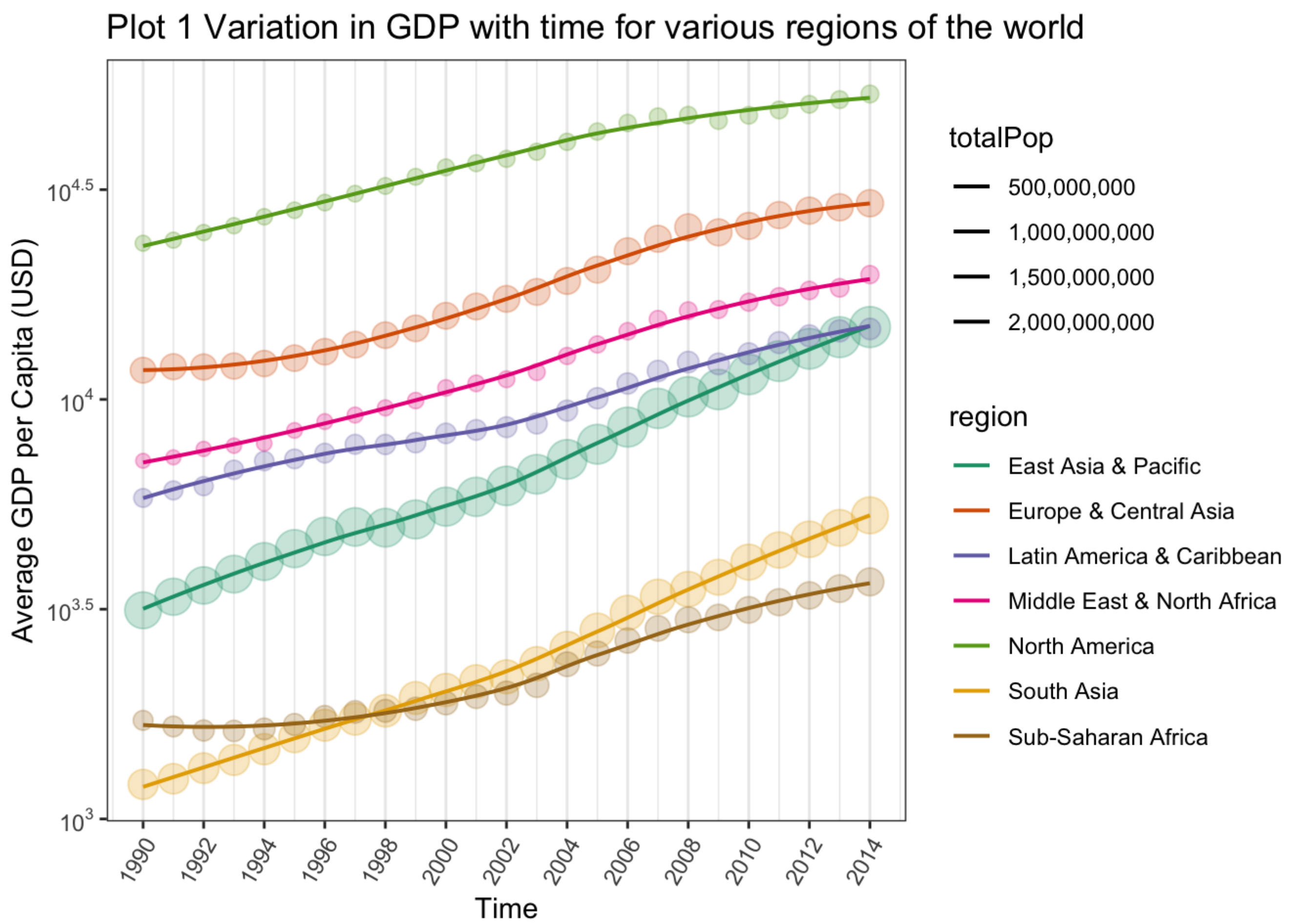
```
#Loading the dataset
NationsData <- read.csv("nations.csv")
head(NationsData)
```

```
##   iso2c iso3c country year gdp_percap life_expect population birth_rate
## 1    AD   AND Andorra 1996         NA          NA      64291       10.9
## 2    AD   AND Andorra 1994         NA          NA      62707       10.9
## 3    AD   AND Andorra 2003         NA          NA      74783       10.3
## 4    AD   AND Andorra 1990         NA          NA      54511       11.9
## 5    AD   AND Andorra 2009         NA          NA      85474        9.9
## 6    AD   AND Andorra 2011         NA          NA      82326         NA
##   neonat_mortal_rate                region      income
## 1                2.8 Europe & Central Asia High income
## 2                3.2 Europe & Central Asia High income
## 3                2.0 Europe & Central Asia High income
## 4                4.3 Europe & Central Asia High income
## 5                1.7 Europe & Central Asia High income
## 6                1.6 Europe & Central Asia High income
```

```
#Preprocessing the data by removing na, and selecting only yhe column required.
#Then using summarise we take the total population and take weighted mean of gdp, and
grouping them by year and region.
plot1 <- na.omit(NationsData)
plot1 <- plot1 %>%
  select(region,year, gdp_percap,population) %>%
  group_by(year, region) %>%
  summarize(totalPop = sum(population,na.rm = TRUE), ave_gdp_percap = weighted.mean(g
dp_percap, population))

head(plot1)
```

```
## # A tibble: 6 x 4
## # Groups:   year [1]
##    year region                       totalPop ave_gdp_percap
##   <int> <fct>                           <dbl>          <dbl>
## 1  1990 East Asia & Pacific        1721688790          3143.
## 2  1990 Europe & Central Asia       797531121         11738.
## 3  1990 Latin America & Caribbean   402141882          5823.
## 4  1990 Middle East & North Africa  233167927          7140.
## 5  1990 North America               277414000         23567.
## 6  1990 South Asia                 1120547744          1209.
```

```
#Plotting the data
ggplot(plot1, aes(x = year, y = ave_gdp_percap, size=totalPop, color = region))+
  ggtitle("Plot 1 Variation in GDP with time for various regions of the world") +
  xlab("Time") +
  ylab("Average GDP per Capita (USD)") +
  geom_point(alpha=0.25, show.legend =FALSE)+
  scale_x_continuous(breaks = seq(1990, 2015, 2))+ # Scaling for x axis
  scale_y_log10(labels = trans_format("log10", math_format(10^.x)), breaks = scales::
trans_breaks("log10", function(x) 10^x)) +
  scale_size_area(max_size = 7, labels  = comma) +
  scale_color_brewer(palette='Dark2') +
  geom_line(size = 0.7  , stat="smooth",method = "loess",se=FALSE, show.legend =TRUE)
+
  theme_bw()+
  theme(
    panel.grid.major.y =element_blank(),
    panel.grid.minor.y = element_blank(),
    axis.text.x = element_text(angle = 60, hjust=1) # angle the axis text at 60 degre
es so we can fit the names
  )
```

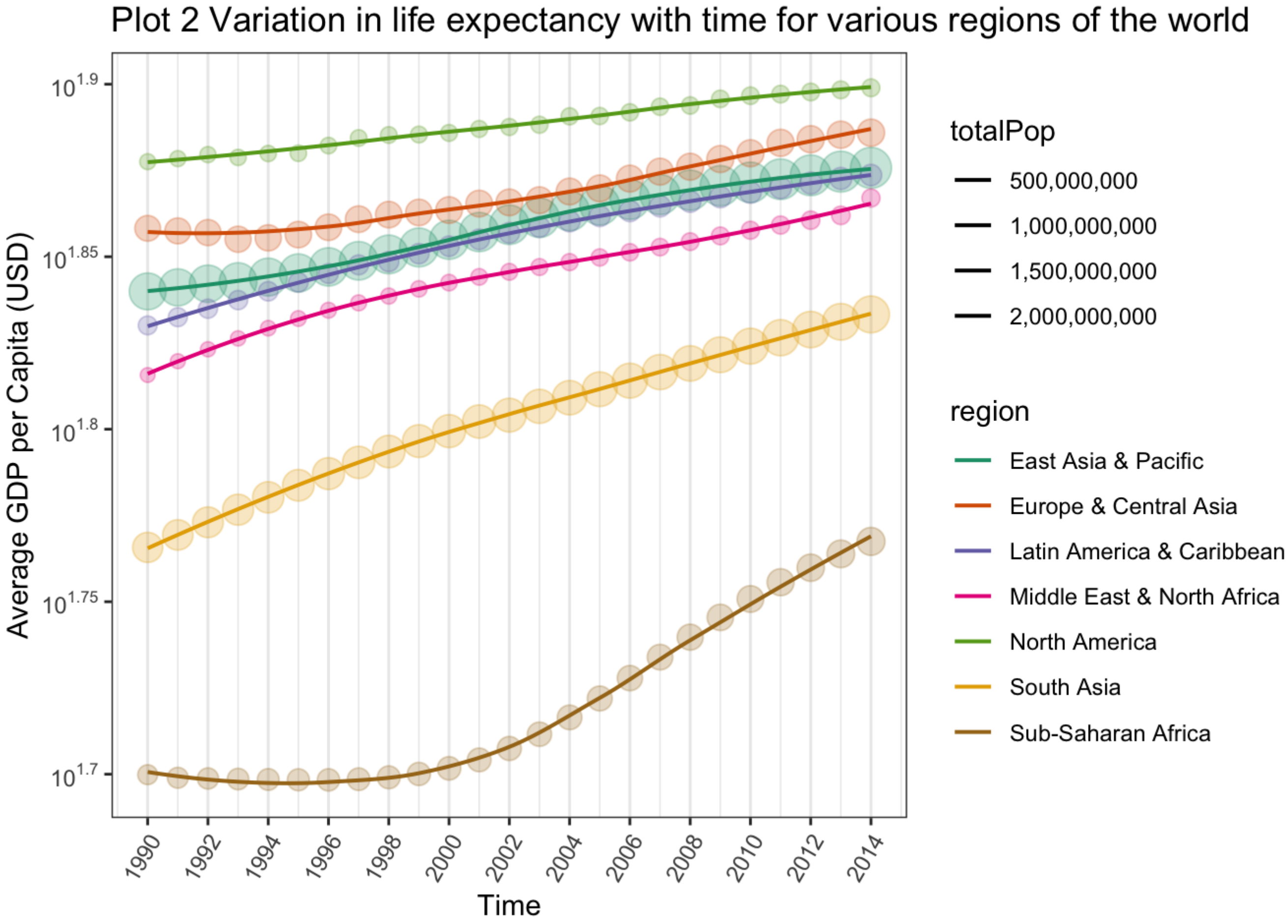


From the plots indicating growth in GDP between 1990 and 2014 for different regions of the world titled Plot 1, where we consider average gdp measured by taking population as a weight parameter. We observe that there is an increasing trend which suggests that every region of the world is heading towards the right path.

```
#Preprocessing the data by removing na, transforming the data by grouping the data by
year and region and taking the sum of population and weighted mean of gdp.
plot2 <- na.omit(NationsData)
plot2 <- plot2 %>%
  select(region,year, life_expect,population) %>%
  group_by(year, region) %>%
  summarize(totalPop = sum(population,na.rm = TRUE), ave_life_expect = weighted.mean
(life_expect, population))
head(plot2)
```

```
## # A tibble: 6 x 4
## # Groups:   year [1]
##    year region                       totalPop ave_life_expect
##   <int> <fct>                           <dbl>           <dbl>
## 1  1990 East Asia & Pacific        1721688790            69.2
## 2  1990 Europe & Central Asia       797531121            72.1
## 3  1990 Latin America & Caribbean   402141882            67.6
## 4  1990 Middle East & North Africa  233167927            65.4
## 5  1990 North America               277414000            75.4
## 6  1990 South Asia                 1120547744            58.3
```

```
#Plotting the data
ggplot(plot2, aes(x = year, y = ave_life_expect, size=totalPop, color = region))+
  ggtitle("Plot 2 Variation in life expectancy with time for various regions of the w
orld") +
  xlab("Time") +
  ylab("Average GDP per Capita (USD)") +
  geom_point(alpha=0.25, show.legend =FALSE)+
  scale_x_continuous(breaks = seq(1990, 2015, 2))+ # Scaling for x axis
  scale_y_log10(labels = trans_format("log10", math_format(10^.x)), breaks = scales::
trans_breaks("log10", function(x) 10^x)) +
  scale_size_area(max_size = 7, labels  = comma) +
  scale_color_brewer(palette='Dark2') +
  geom_line(size = 0.7  , stat="smooth",method = "loess",se=FALSE, show.legend =TRUE)
+
  theme_bw() +
  theme(
    panel.grid.major.y =element_blank(),
    panel.grid.minor.y = element_blank(),
    axis.text.x = element_text(angle = 60, hjust=1) # angle the axis text at 60 degre
es so we can fit the names
  )
```



Similarly from the plots indicating the growth in average life expectancy rate between 1990 and 2014 accross all the regions of the world titled plot 2, we observe that even though in the sub saharan region there is a slight negative trend between 1990 and 2000, but overall all the regions of the world are in the right path progressing towards achieving high life expectancy rates.

Part 3 Plots indicating percentage change in population between 1950 and 2017.

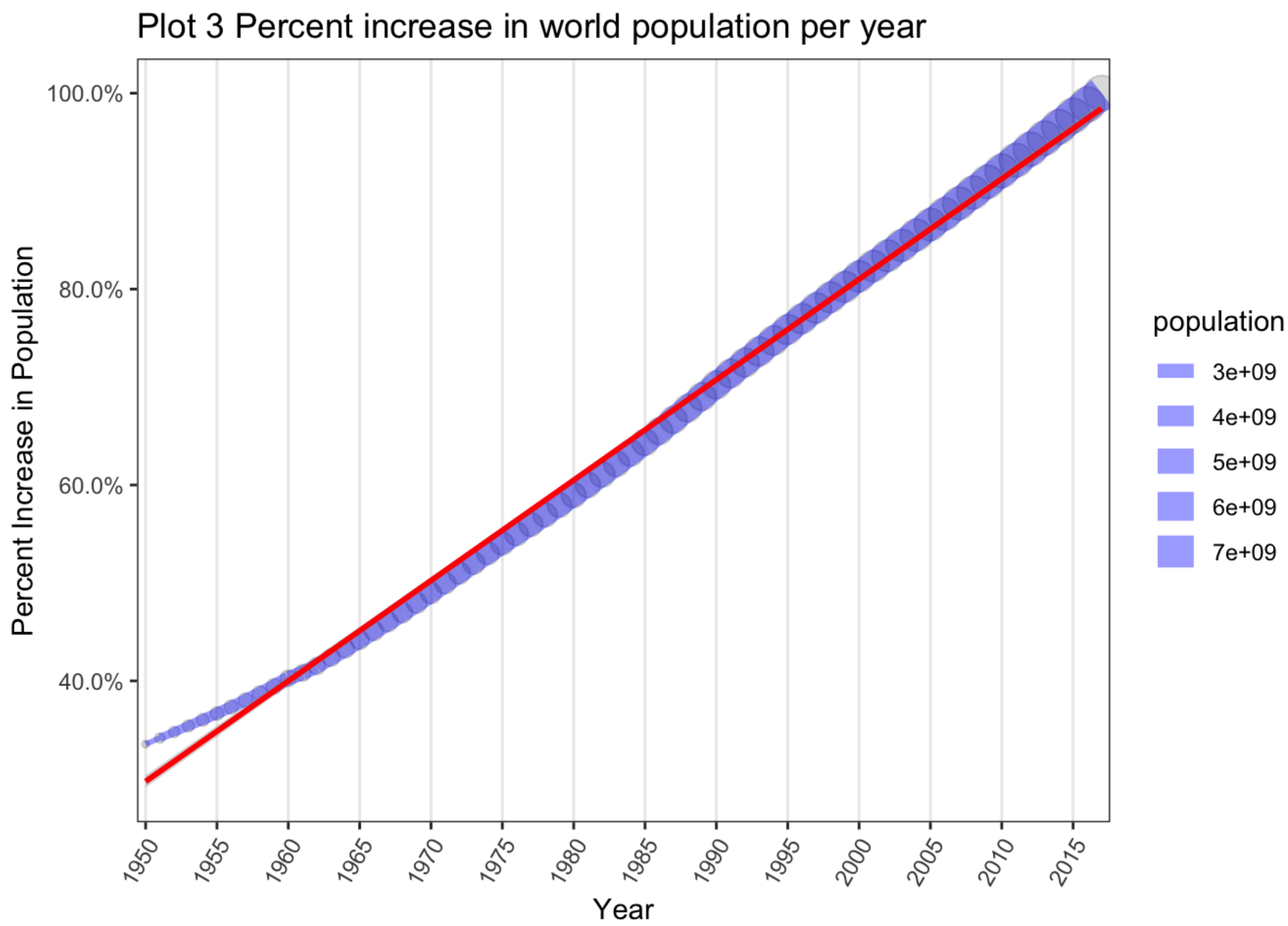
```
#Loading the dataset
PopulationDataset<- as.data.frame(read.csv("population.csv", header = FALSE))


WorldPopulation1 <- as.data.frame(read.csv("worldpopulation_percentage.csv", header =
FALSE))
```

```
#Preprocessing the data by giving appropreate column name, removing na and transformi
ng the data.
# Using loop to allocate column names and then for columns without column names we ma
nually give names
i <- 0
while (i<2) {
  PopulationDataset <- PopulationDataset[-1, ]
  i <- i+1
}

colnames(PopulationDataset) <- PopulationDataset[1,]
colnames(PopulationDataset)[1] <- "RegionName"
colnames(PopulationDataset)[2] <- "CountryCode"
colnames(PopulationDataset)[2] <- "IndicatorName"
colnames(PopulationDataset)[3] <- "IndicatorCode"
PopulationDataset <- PopulationDataset[-1, ]
WorldPopulation <- PopulationDataset[258,]
WorldPopulation <- WorldPopulation[,-64]
WorldPopulation <- WorldPopulation[,-63]
colnames(WorldPopulation1) <- c("year", "population", "average annual", "average annu
al")
WorldPopulation1 <- WorldPopulation1[-1,]

second <- as.data.frame(WorldPopulation1$year)
second$population <- WorldPopulation1$population
second$population <- as.numeric(gsub(",","",second$population))
second <- second[-11:-142,]
```

```
#Combining two datasets to get the required range of data
WorldPopulation <- na.omit(WorldPopulation)
second <- na.omit(second)
plot3 <- gather(WorldPopulation, key = Year, value = Population, `1960`:`2017`)
plot3 <- select(plot3, Year, Population)
plot3 <- aggregate(plot3$Population, by = list(plot3$Year), FUN = sum)
plot3 <- select(plot3, Group.1, x)
names(plot3) <- c("year","population")
names(second) <- c("year","population")
plot3 <- rbind(second,plot3)
plot3$Percentage <- (plot3$population/max(plot3$population))
head(plot3)
```

```
##   year population Percentage
## 1 1950 2525778669  0.3354127
## 2 1951 2572850917  0.3416637
## 3 1952 2619292068  0.3478309
## 4 1953 2665865392  0.3540157
## 5 1954 2713172027  0.3602978
## 6 1955 2761650981  0.3667356
```

```
#Plotting the data
ggplot(plot3, aes(year, Percentage, size = population, group = 1)) +
  ggtitle("Plot 3 Percent increase in world population per year") +
  geom_point(alpha=0.15, show.legend =FALSE)+
  geom_line( alpha=0.4, color="blue") +
  geom_smooth(method = lm, colour = "red", show.legend =FALSE) +
  scale_x_discrete(breaks = seq(1950, 2017, 5))+ # Scaling for x axis
  scale_y_continuous(labels=scales::percent) +
  xlab("Year")+
  ylab("Percent Increase in Population")+
  theme_bw()+
  theme(
    panel.grid.major.y =element_blank(),
    panel.grid.minor.y = element_blank(),
    axis.text.x = element_text(angle = 60, hjust=1) # angle the axis text at 60 degre
es so we can fit the names
  )
```



From plot 3 we can observe that in the past 67 years between 1950 and 2017 we can see that in the year 1950 we onnly had around 35% of the population as compared to 2017. Ever since then we can observe a steady increase in the population increase. We also observe that population curve is steeper **between 2005 and 2017** which suggests that the population growth is increasing in the recent years. Several factors including GDP increase and improved life expectancy rate could have played a role in the observed outcome.

Part 4 Plots indicating change in fertility rate between 1950 and 2017.

```
#Loading the dataset
FertilityDataset <- (read.csv("children-per-woman-UN.csv", header = FALSE))
head(FertilityDataset)
```

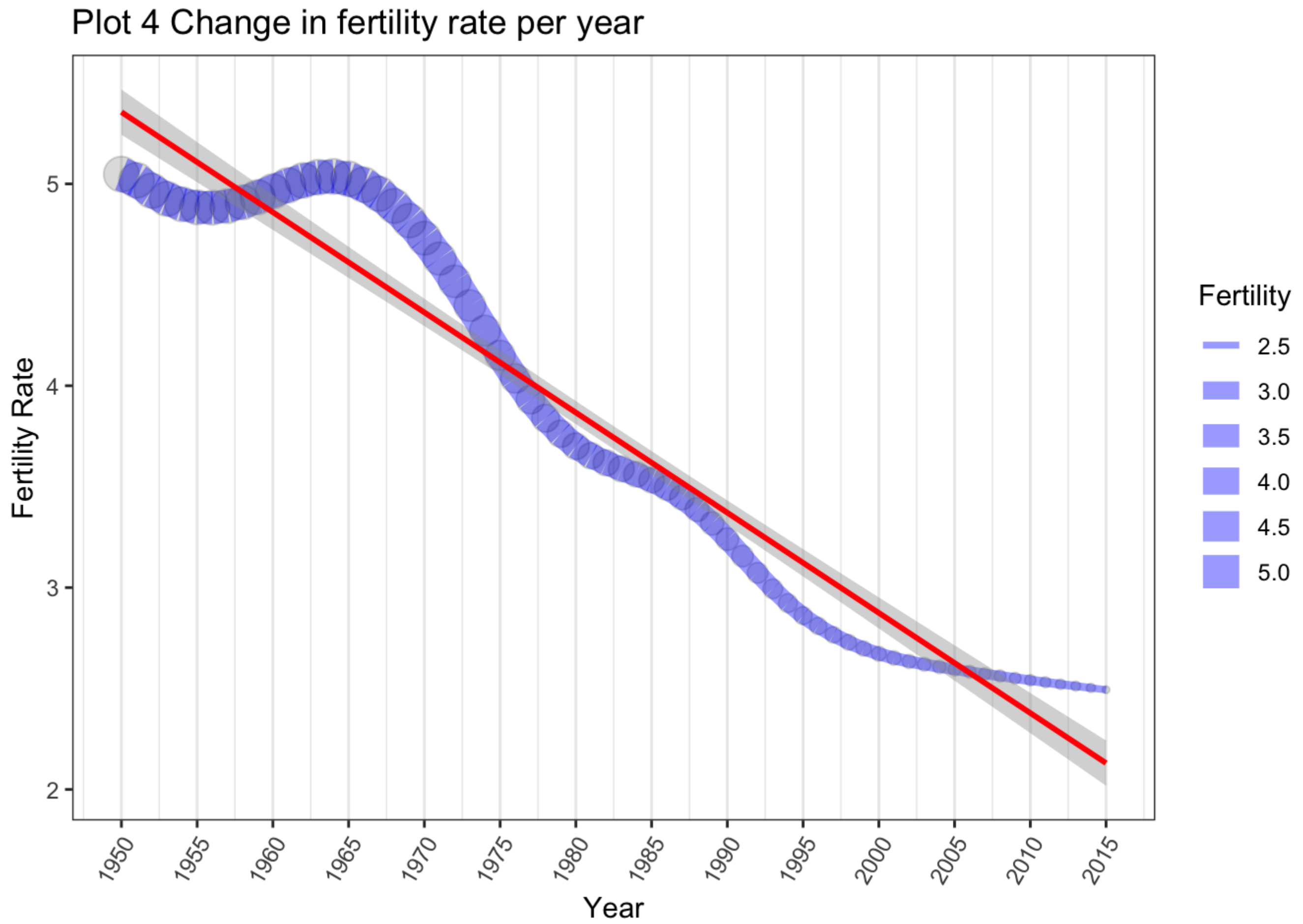
```
##             V1   V2   V3
## 1       Entity Code Year
## 2 Afghanistan  AFG 1950
## 3 Afghanistan  AFG 1951
## 4 Afghanistan  AFG 1952
## 5 Afghanistan  AFG 1953
## 6 Afghanistan  AFG 1954
##
V4
## 1 Estimates, 1950 - 2015: Demographic Indicators - Total fertility (live births pe
r woman) (live births per woman)
## 2
7.45
## 3
7.45
## 4
7.45
## 5
7.45
## 6
7.45
```

```
#Preprocessing the data by converting factors to neumaric data type
colnames(FertilityDataset) <- c("Area","Code","Year","Fertility")
FertilityDataset <- FertilityDataset[-1,]
FertilityDataset$Fertility <- as.numeric(as.character(FertilityDataset$Fertility))
FertilityDataset$Year <- as.numeric(as.character(FertilityDataset$Year))
```

```
#Transforming the data by selecting only the required rows.
FertilityDataset <- na.omit(FertilityDataset)
plot4 <- FertilityDataset
plot4 <- plot4[15643:15708,]
plot4 <- select(plot4, Year, Fertility)
head(plot4)
```

```
##       Year Fertility
## 15644 1950     5.048
## 15645 1951     5.018
## 15646 1952     4.966
## 15647 1953     4.926
## 15648 1954     4.899
## 15649 1955     4.884
```

```
#Plotting the data
ggplot(plot4, aes(Year, Fertility, size = Fertility, group = 1)) +
  ggtitle("Plot 4 Change in fertility rate per year") +
  geom_point(alpha=0.15, show.legend =FALSE)+
  geom_line( alpha=0.4, color="blue") +
  geom_smooth(method = lm, colour = "red", show.legend =FALSE) +
  scale_x_continuous(breaks = seq(1950, 2015, 5))+ # Scaling for x axis
  ylab("Fertility Rate")+
  theme_bw()+
  theme(
    panel.grid.major.y =element_blank(),
    panel.grid.minor.y = element_blank(),
    axis.text.x = element_text(angle = 60, hjust=1) # angle the axis text at 60 degre
es so we can fit the names
  )
```

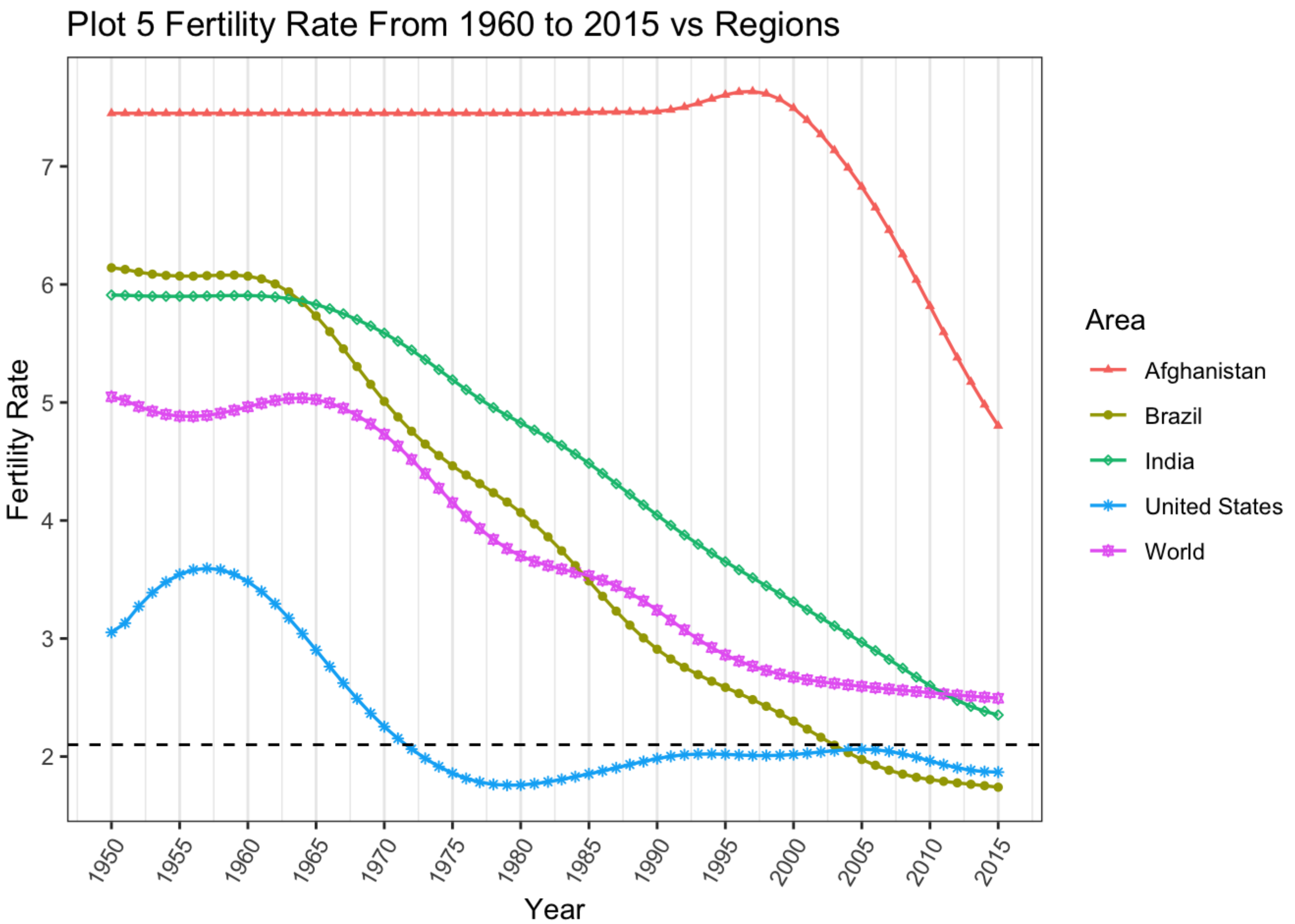


The fertility rate dataset was only available until 2015 so we consider the senario of fertility rate between 1950 and 2015. From the plot 4 we can observe that comparing the years 1950 and 2015 there is a decreasing trend observed in the fertility rate. Since the fertility rate is in the decline it is safe to assume that the birth rate has also taken a hit. Even though the birth rate is in decline but the rate of pupulation is experiencing a steep increase suggesting the increase in life expectancy rate has caused the steep rate of increase in population. Also we observe a period between 1960 and 1975 there is a slight increase in fertility rate similarly between 1987 and 1990.

Part 5 Plots indicating change in fertility rates for different types of countries.

```
#Transforming the data by combining data of different countries
USA <- FertilityDataset[14851:14916,] #RICH COUNTRY
Afganistan <- FertilityDataset[1:66,] #POOR COUNTRY
Brazil <- FertilityDataset[1717:1782,] #NEWLY DEVELOPED
India <- FertilityDataset[6073:6138,] #Developing
World <- plot4 <- FertilityDataset[15643:15708,]
plot5 <- rbind(USA,Afganistan,Brazil,India,World)
```

```
#Plotting the data
ggplot <- ggplot(plot5, aes(x = Year, y = Fertility, color= Area, shape = Area))

ggplot + geom_point(size = 1)+
  geom_line(size=0.6)+
  ggtitle("Plot 5 Fertility Rate From 1960 to 2015 vs Regions") +
  scale_shape_manual(values=c(17, 19, 23, 8, 11))+
  scale_x_continuous(breaks = seq(1950, 2015, 5))+
  scale_y_continuous(breaks = seq(1, 8, 1))+
  geom_hline(yintercept = 2.1, linetype = "dashed")+
  scale_fill_manual(values=c("#FF5C5C", "#E69F00", "#56B4E9", "#009E73", "#FFFF8E",
"#F39BFF")) +
  ylab("Fertility Rate") +
  theme_bw()+
  theme(
    panel.grid.major.y =element_blank(),
    panel.grid.minor.y = element_blank(),
    axis.text.x = element_text(angle = 60, hjust=1) # angle the axis text at 60 degre
es so we can fit the names
  )
```



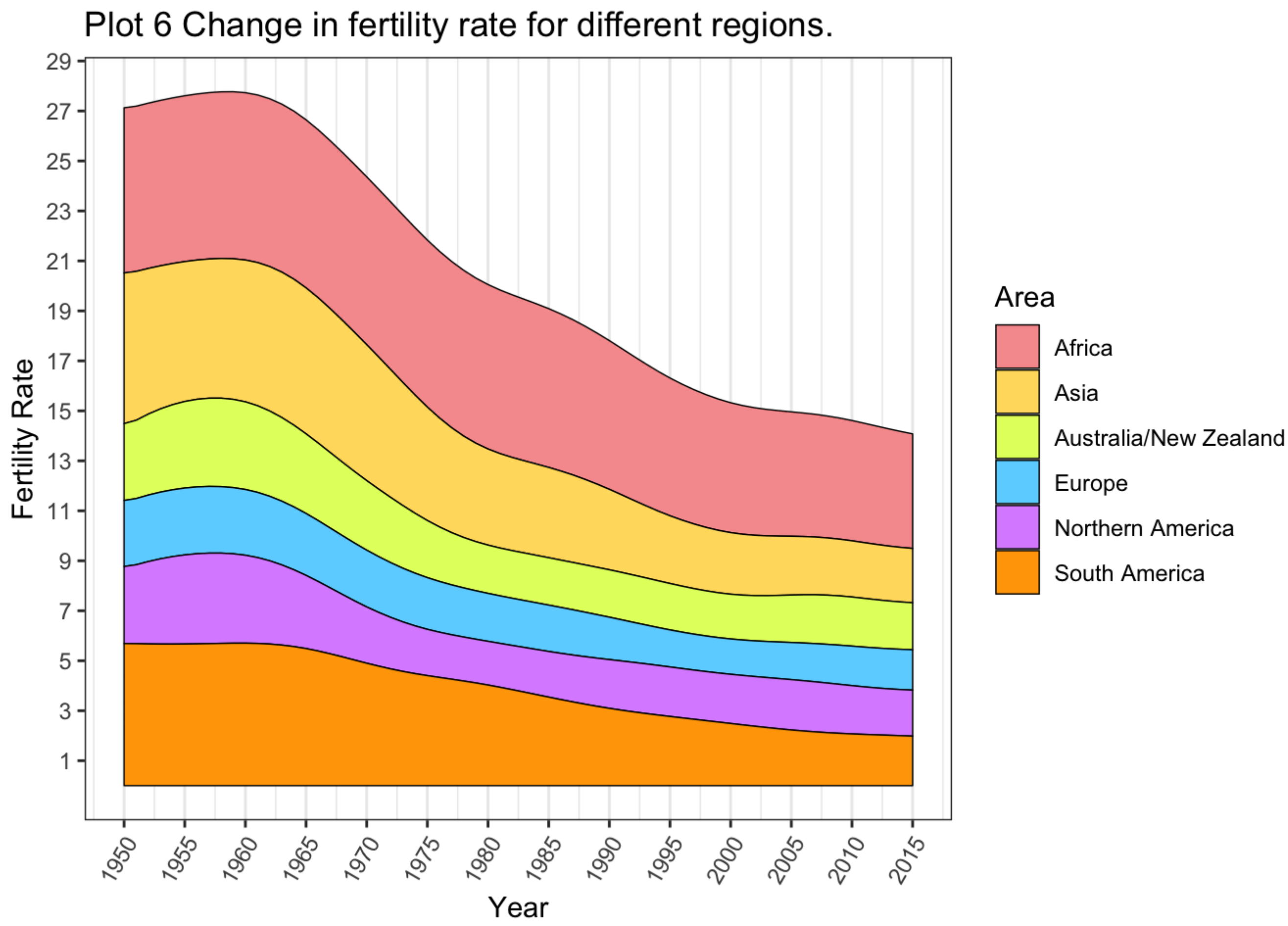
From the plot 5 we can obseve the following:

- Fertility rate in USA which is a rich country is the lowest from 1950 to 2015.
- Fertility rate in newly developed country like Brazil was high initially but in recent times the fertility rate is low.
- Fertility in devepoing cuntry like India was has experienced a steady decrease.
- Countries which go past the 2.1 value of fertility rate experience a significant shrink of population.

Plots indicating change in fertility rates for different regions of the world.

```
#Transforming the data by selecting only the required rows and combining data of diff
erent countries.
Asia <- FertilityDataset[595:660,]
Europe <- FertilityDataset[4423:4488,]
North_America <- FertilityDataset[10363:10428,]
South_America <- FertilityDataset[12739:12804,]
Africa <- FertilityDataset[67:132,]
Australia_NewZeland <- FertilityDataset[727:792,]
plot6 <- rbind(Asia,Europe,North_America,South_America,Africa,Australia_NewZeland)
```

```
#Plotting the data
ggplot(plot6, aes(x = Year, y = Fertility, fill = Area)) +
  ggtitle("Plot 6 Change in fertility rate for different regions.") +
  geom_area(stat="identity", colour="black", size = .25) +
  scale_fill_manual(values=c("#F69C9C", "#FFDD6C", "#E2FF6C", "#6CD4FF", "#DB92FF",
"#FFA500")) +
  scale_x_continuous(breaks = seq(1950, 2015, 5))+
  scale_y_continuous(breaks = seq(1, 30, 2))+
  ylab("Fertility Rate") +
  theme_bw()+
  theme(
    panel.grid.major.y =element_blank(),
    panel.grid.minor.y = element_blank(),
    axis.text.x = element_text(angle = 60, hjust=1) # angle the axis text at 60 degre
es so we can fit the names
  )
```



From plot 6 we can observe that the fertility ratio of african countries is the highest whereas the fertility rate of the american countries is at the lowest. We can also observe that across countries in america, europe we have steady fertility rate without much variation on the other hand in asian, australian and african countries we can see the fertility rate is in heavy decline.

Conclusion

From all the above we can come to a conclusion that the hypothesis suggested by ridley seems to be appropriate. There are several factors which support the hypothesis presented which are listed below.

- Fertility rate is in decline among all regions. In rich countries or developed countries the fertility rate has decreased and has stabilised. This means that with more and more countries entering the developing nations list or developed nation status, their fertility rate will stabilise after decreasing.
- If fertility rate stabilises then even birth rate will be stabilised, which will reduce the number of individuals added to the population. This will cause the birth rate to steady down and with life expectancy rate increasing the death rate will also stabilise creating the stabilisation effect on population around the 9 or 10 billion mark.
- Even though the life expectancy rate will increase but after a certain point say in 20 or 30 years eventually mortality rate will increase causing either the population to stabilise or decrease significantly.

All these factors support the hypothesis that claims firtility rate in rich countries is low and developing countries are showing a decrease in average fertility value. Similarly at some point in the near future the population will stabilise at around 9 billion.

Even though these evidances are convincing supporting ridley’s hypothesis, however there are certain factors which can challenge the hypothesis. They are enlisted below.

- Say if the life expectancy rate increases dramatically as suggested from the plot and the death rates will not increase then in the hypothesis might not be true for the near furute.

- If the above point is true the estimate of population count where the world's population is to stabilise might be in accurate as well as the value might be quite high as compared to the suggested figure like say 11-13 Billion
- Also if we remember carefully in plot 4 we observed a period between 1960 and 1975 where there was a slight increase in fertility rate and once gain the same phenominon happened between 1987 and 1990 if we observe such a trend in the future then fertility rates might increase immensely causing the increase in the birth rate which will result in population increase.

References

https://stackoverflow.com/questions/32054368/use-first-row-data-as-column-names-in-r?noredirect=1&lq=1 (https://stackoverflow.com/questions/32054368/use-first-row-data-as-column-names-in-r?noredirect=1&lq=1)

https://www.r-bloggers.com/how-to-aggregate-data-in-r/ (https://www.r-bloggers.com/how-to-aggregate-data-in-r/)

http://garrettgman.github.io/tidying/ (http://garrettgman.github.io/tidying/)

https://data.worldbank.org/indicator/SP.POP.TOTL (https://data.worldbank.org/indicator/SP.POP.TOTL) (World Population Dataset 1)

https://ourworldindata.org/world-population-growth (https://ourworldindata.org/world-population-growth) (World Population dataset 2)

https://ourworldindata.org/fertility-rate#the-global-decline-of-the-fertility-rate-since-1950 (https://ourworldindata.org/fertility-rate#the-global-decline-of-the-fertility-rate-since-1950) (Fertility Dataset)

https://stackoverflow.com/questions/3418128/how-to-convert-a-factor-to-integer-numeric-without-loss-of-information (https://stackoverflow.com/questions/3418128/how-to-convert-a-factor-to-integer-numeric-without-loss-of-information)

http://www.sthda.com/english/wiki/ggplot2-add-straight-lines-to-a-plot-horizontal-vertical-and-regression-lines (http://www.sthda.com/english/wiki/ggplot2-add-straight-lines-to-a-plot-horizontal-vertical-and-regression-lines)

Week 5, 6 and 7 Tutorials