CT5100 Data Visualisation: Assignment 2

Kevin Derrane, 12409118.

Robbie Deegan, 17232577.

Jamie O'Halloran, 12458152.

16/02/2018

Introduction

Matt Ridley's hypothesis states that "fertility rates everywhere will converge to 2.1 in a few decades, and the world population will stabilise at 9 billion people" This is known as the replacement rate. When this rate is hit the population will begin to shrink.

Our project uses compiled datasets of fertility rates, education and population growth using data collected from the World Bank and United Nations data sources. Our collective objective is to illustrates through appropriate visualisations the evidence that might support Matt Ridley's hypothesis. Our four main visualtions are as follows:

1. Percentage increase in world population from 1950 till present.
2. Changing fertility rates from 1950 to the present for selected countries of the world.
3. Regional view of the above visualisation
4. The observed relationship between fertility & per capita income & education

To execute these visualizations, we used a programming language known as R. To create the best visualizations, we utilized many different r packages such as; scales; ggplot2; reshape; dplyr and many more. After exploring and creating the visualizations we then uploaded them to a HTML page rendered from a RMarkdown script.

1. Percentage increase in world population.

The dataset below was obtained from the UN data source:

1. World Population http://data.un.org/Data.aspx?q=world+population&d=PopDiv&f=variableID%3a12%3bcrID%3a900.

```
head(world_pop)
```

```
##   Country.or.Area Year.s.   Variant   Value
## 1           World    2018 No change 7637740
## 2           World    2017 No change 7553284
## 3           World    2016 No change 7468369
## 4           World    2015 No change 7383009
## 5           World    2014 No change 7298453
## 6           World    2013 No change 7213426
```

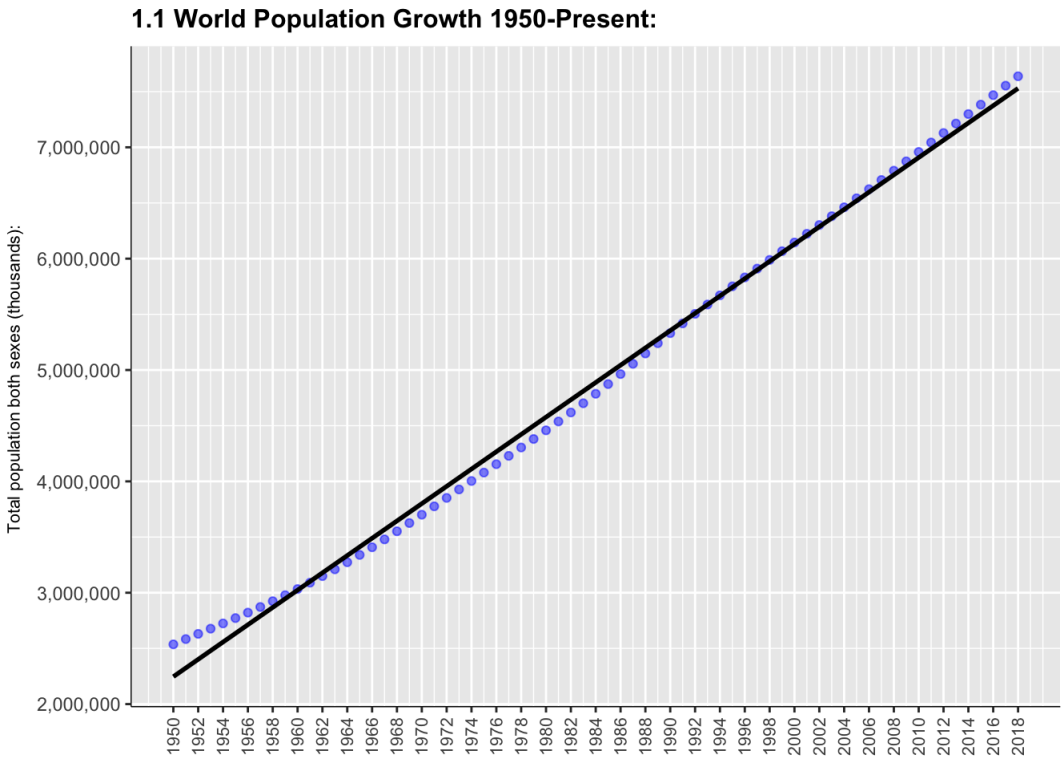
Plots:

Plot 1.1 below illustrates the world population growth since 1950. The graph shows the graudal inrease in the worlds population.

Instead of designing this graph with only one line we decided to opt for the actual population growth followed by a fitted linear regression line to emphasize the upwards trend in world population. The solid black line represents the upwards population trend, while the actual population growth is displayed by the spotted blue line. With simplicity in mind we decided to separate our X-axis into two year break ticks instead of individual years. We felt that this left our graph less cluttered, while still allowing you to visuale as much information as possible.

```
p1.1 <- ggplot(world_pop , aes(x=Year.s., y=Value))
p1.1 +  geom_point(na.rm = TRUE, alpha=0.5, color = "blue") +
  ggtitle("1.1 World Population Growth 1950-Present:") +
  geom_smooth(method=lm,se=FALSE, color = "black")  +
  scale_x_log10(labels = trans_format("log10", math_format(10^.x)), breaks = scales::trans_breaks("log10", f
unction(x) 10^x)) +
  scale_size_area(max_size = 10, labels  = comma) +
  scale_x_continuous(lim=c(1950,2018), breaks=seq(1950,2018,2)) +
  scale_y_continuous(name="Total population both sexes (thousands): ", labels = comma) +
  theme(axis.line = element_line(colour = "black", size = 0.25), plot.title = element_text(size = 12, face
= "bold"), axis.title.x=element_blank(), axis.text.x = element_text(size=8, angle = 90, hjust=1, vjust = .5
), title = element_text(size=8), legend.key = element_rect(fill = NA, colour = NA, size = 0.25))
```

```
## Scale for 'x' is already present. Adding another scale for 'x', which
## will replace the existing scale.
```

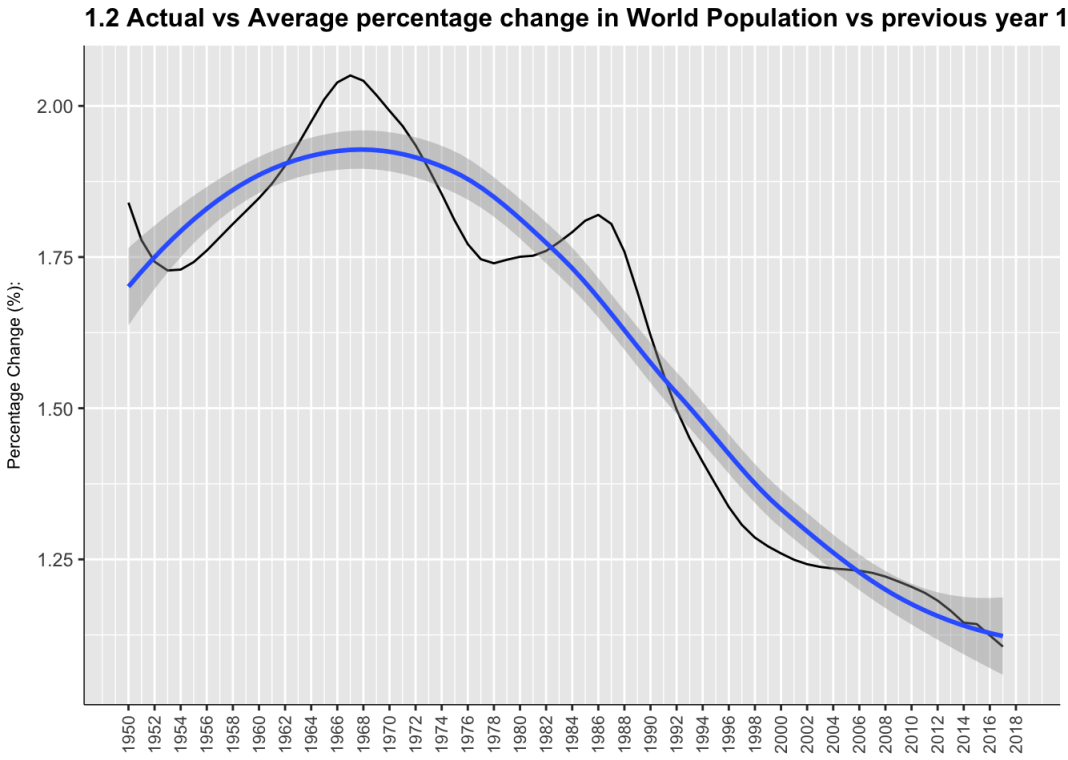


Plot 1.2 below illustrates the world population percentage growth versus the previous year since 1950. The graph shows the graudal downward trend in the percentage increase compared to the previous year.

Instead of designing this graph with only one line we decided to opt for the actual percentage values followed by a regression line to emphasize the downwards trend in percentage increase compared to the previous year. The solid black line represents the actual percentage values, while the blue line is a regression line. Continuing with the above design in mind we decided to separate our X-axis into two year break ticks instead of individual years. We felt that this left our graph less cluttered, while still allowing you to visuale as much information as possible.

```
# Percentage increase for each year
Year <- world_pop $Year.s.
Population <- world_pop $Value
df <- data_frame(Year, Population)

df %>% mutate(Previous_Year = lag(Population, 1), Change = Previous_Year - Population, Percent_Change = Chan
ge/Previous_Year*100) %>%
  ggplot(aes(x = Year, y = Percent_Change)) +
  geom_line() +
  geom_smooth() +
  xlab("Year") +
  ylab("Percentage Change (%):") +
  ggtitle("1.2 Actual vs Average percentage change in World Population vs previous year 1950-Present:") +
  scale_x_continuous(lim=c(1950,2018), breaks=seq(1950,2018,2)) +
  theme(axis.line = element_line(colour = "black", size = 0.25), plot.title = element_text(size = 12, face
= "bold"), axis.title.x=element_blank(), axis.text.x = element_text(size=8, angle = 90, hjust=1, vjust = .5
), title = element_text(size=8), legend.key = element_rect(fill = NA, colour = NA, size = 0.25))
```

```
## `geom_smooth()` using method = 'loess'
```



Plots 1.1 & 1.2 show the increase in the worlds population and also the decrease in the percentage increase compared to the previous year. This shows that as the world is getting closer to 9 billion people, it seems to be stabilising. This confirms part of Matt Ridley's hypothesis.

2. Changing fertility rates from 1950 to the present for selected countries of the world.

The datasets below were obtained from the World Bank & UN data source:

1. GDP http://data.un.org/Data.aspx?q=world+population&d=PopDiv&f=variableID%3a12%3bcrID%3a900.
2. Fertility Rates http://data.un.org/Data.aspx?q=fertility%20rates&d=PopDiv&f=variableID%3A54.

These datasets were then edited online prior to downloading.

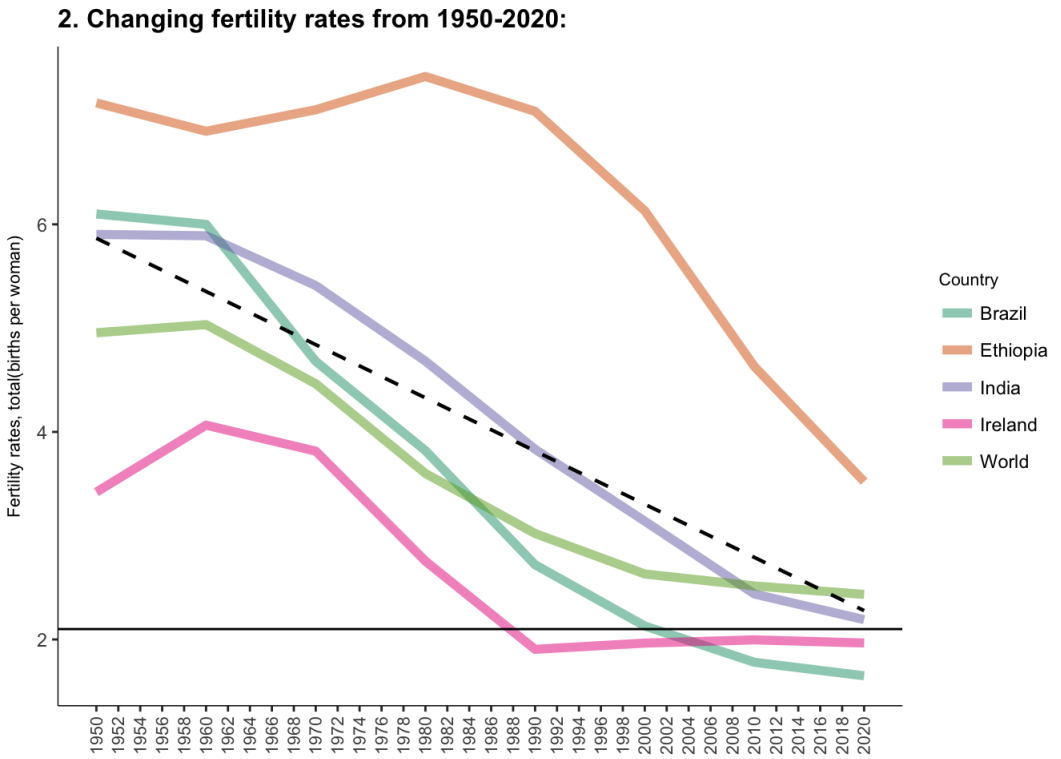
```
head(countries)
```

```
##   Country Year End_Year Variant Value
## 1  Brazil 2020     2025  Medium  1.65
## 2  Brazil 2010     2015  Medium  1.78
## 3  Brazil 2000     2005  Medium  2.13
## 4  Brazil 1990     1995  Medium  2.72
## 5  Brazil 1980     1985  Medium  3.82
## 6  Brazil 1970     1975  Medium  4.68
```

Plots:

Graph 2 below is a comparison visualization of four different countries rates versus the world average fertility rate since 1950. Each country represents either a 'rich' (Ireland), 'poor' (Ethiopia), 'newly developed' (Brazil) or 'developing' (India) country. To best differentiate the countries, we used the 'Dark2' palette. We felt that this clearly distinguished each country from one another. The plot includes a black line, which Ridley's hypothesis of a 2.1 fertility rate. This was executed by using the 'geom_hline' function in the ggplot2 package. As there are several times our data overlaps we decided to add some transparency to the graph. We did this by setting the alpha function to 0.5, which resulted in a much clearer plot. The black dashed line is a downward sloping regression line, which allows the viewer to visualize the trajectory of fertility rates.

```
countries$Year <- as.integer(countries$Year)
p2 <- ggplot(countries, aes(x=Year, y=Value, color=Country))
p2 + geom_line(alpha = 0.5, size = 2 ) +
  stat_smooth(method=lm, se= FALSE, colour = "black", size = 0.8, linetype="dashed")+
  scale_x_continuous(lim=c(1950,2020), breaks=seq(1950,2020,2)) +
  ylab("Fertility rates, total(births per woman)")+
  ggtitle("2. Changing fertility rates from 1950-2020:")+
  geom_hline(yintercept=2.1) +
  theme( panel.background = element_blank()) +
  scale_color_brewer(palette='Dark2') +
  theme(panel.grid.major = element_blank(), panel.background = element_blank(), axis.line = element_line(col
our = "black", size = 0.25), plot.title = element_text(size = 12, face = "bold"),
        axis.title.x=element_blank(), axis.text.x = element_text(size=8, angle = 90, hjust=1, vjust = .5),
title = element_text(size=8), legend.key = element_rect(fill = NA, colour = NA, size = 0.25))
```



Originally, Brazil had a higher fertility rate than India, but from 1960 onwards Brazil's fertility rate was less than India's. This is important because Brazil is seen as a newly developed country, while India is seen as a developing country.

This graph highlights the downward trend of wordlwide fertility rates. It shows this rate getting closer to the 2.1 replacement rate.

3. Regional view of the above visualisation

The datasets below were obtained from the World Bank & UN data source:

1. GDP http://data.un.org/Data.aspx?q=world+population&d=PopDiv&f=variableID%3a12%3bcrID%3a900.
2. Fertility Rates http://data.un.org/Data.aspx?q=fertility%20rates&d=PopDiv&f=variableID%3A54.

These datasets were then edited online prior to downloading.

```
head(region)
```

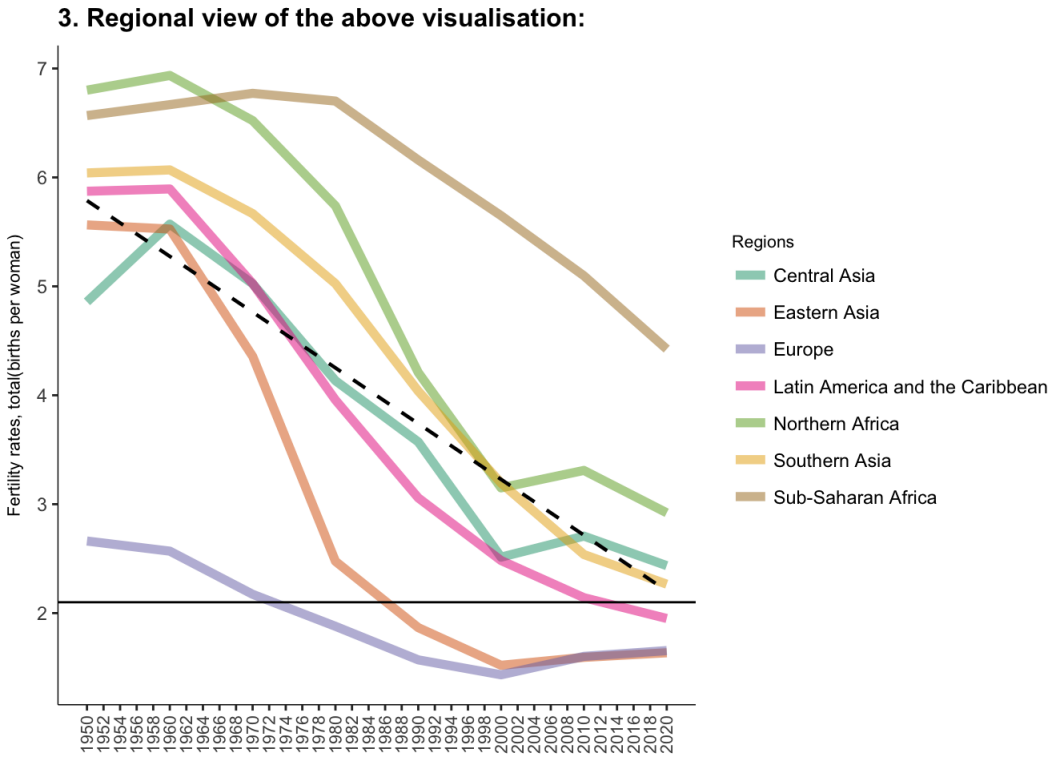
```
##        Regions Year End_Year Variant Value
## 1 Central Asia 2020     2025  Medium 2.436
## 2 Central Asia 2010     2015  Medium 2.707
## 3 Central Asia 2000     2005  Medium 2.514
## 4 Central Asia 1990     1995  Medium 3.572
## 5 Central Asia 1980     1985  Medium 4.134
## 6 Central Asia 1970     1975  Medium 5.024
```

Plots:

Graph 3 below is compiled from the same data from the regions designated in this week's dataset. The seven different regions are clearly

distinguished by the 'Dark2' palette. We incorporated a regression to illustrate which direction fertility rates are headed, which is negatively sloped. Since the dataset goes back all the way to 1950 we felt that it was better to separate the X-axis by every two years. As there were many lines in this visualization transparency within the plot was crucial, so we set the alpha to 0.5

```
region$Year <- as.integer(region$Year)

p3 <- ggplot(region, aes(x=Year, y=Value, color=Regions))
p3 + geom_line( alpha = 0.5, size = 2) +
  stat_smooth(method=lm, se= FALSE, colour = "black", size = 0.8, linetype="dashed")+
  scale_x_continuous(lim=c(1950,2020), breaks=seq(1950,2020,2)) +
  ylab("Fertility rates, total(births per woman)") +
  ggtitle("3. Regional view of the above visualisation:")+
  geom_hline(yintercept=2.1) +
  theme( panel.background = element_blank()) +
  scale_color_brewer(palette='Dark2') +
  theme(panel.grid.major = element_blank(), panel.background = element_blank(), axis.line = element_line(col
our = "black", size = 0.25), plot.title = element_text(size = 12, face = "bold"),
        axis.title.x=element_blank(), axis.text.x = element_text(size=8, angle = 90, hjust=1, vjust = .5),
title = element_text(size=8), legend.key = element_rect(fill = NA, colour = NA, size = 0.25))
```



This graph highlights the downward trend of regional fertility rates. It shows this rate getting closer to the 2.1 replacement rate.

4. Observed relationship between fertility & per capita income & education

The datasets below were obtained from the World Bank data source:

1. GDP https://data.worldbank.org/indicator/NY.GDP.PCAP.CD?end=2016&start=1970&view=chart.
2. Literacy Rates https://data.worldbank.org/indicator/SE.ADT.LITR.ZS.
3. Fertility Rates https://data.worldbank.org/indicator/SP.DYN.TFRT.IN?end=2016&start=1970.

These datasets were then combined and edited within Microsoft Excel.

```
head(rates)
```

```
##   Country Year Fertility.Rate      GDP Literacy.Rate
## 1   World 1975       4.154705 1449.619      68.76034
## 2   World 1976       4.040733 1549.510      68.89088
## 3   World 1977       3.939673 1722.394      69.01872
## 4   World 1978       3.848233 1992.436      69.42679
## 5   World 1979       3.780500 2274.798      69.82894
## 6   World 1980       3.722039 2516.286      70.04151
```

Plots:

Plot 4.1, 4.2 and 4.3 below illustrates the relationship between fertility & per capita income & education from 1975-2015. The graph shows the downward and upward trends of the relationships between the variables.

4.1 Graph 4.1 shows the average worldwide fertility rates. It shows the downward trend of the decrease in fertility rates worldwide. A regression line was fitted to show and emphasise this trend. As we can see the fertility rates have dramatically dropped since 1975. In this graph a 2.1 replacement line was fitted to show Matt Ridleys fertility hypothesis. As we can see from the graph and the fitted regression line we are no far off of the 2.1 replacement rate for fertility rates.

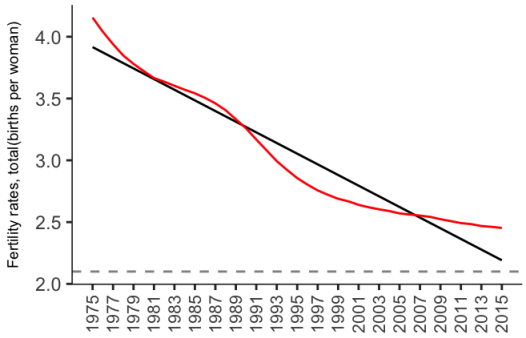
4.2 Graph 4.2 shows the average worldwide GDP per capita in USD from 1975 to 2015. It shows the upward trend of the increase in acergae worldwide GDP per capita. A regression line was also fitted to show and emphasise this trend. As we can see the average GDP per capita has increased by more than seven times since 1975.

4.3 Graph 4.3 shows the average worldwide adult literacy rate from 1965 to 2015. We decided to use literacy rates as a means to measure the level of education of a country. Datasets relation to literacy rates were also easier to find. A regression line was fitted to show the upward trend in literacy rates. Literacy rates have increased by almost 20% since 1975. Literacy rates also seem to be rising steadily compared to the other two graphs.

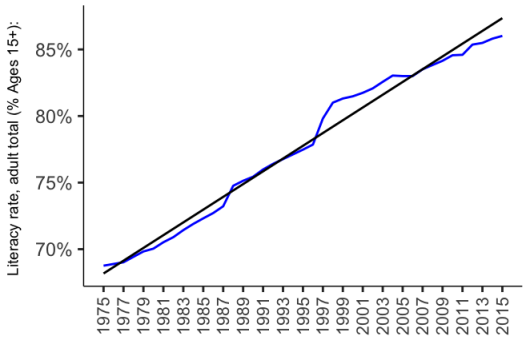
Continuing with the above design in mind we decided to separate our X-axis into two year break ticks instead of individual years. We felt that this left our graph less cluttered, while still allowing you to visuale as much information as possible.

```
p4.1 <- ggplot(rates, aes(x = Year, y = Fertility.Rate)) +
  stat_smooth(method = lm, se = FALSE, color = "black", size = 0.5) +
  geom_line(color = "red") +
  geom_hline(yintercept=2.1, alpha = 0.5, linetype="dashed") +
  scale_x_continuous(lim=c(1975,2015), breaks=seq(1975,2015,2)) +
  ggtitle("4.1 Worldwide Fertility rates 1975-2015:") +
  ylab("Fertility rates, total(births per woman)") +
  theme(panel.grid.major = element_blank(), panel.background = element_blank(), axis.line = element_line(col
our = "black", size = 0.25), plot.title = element_text(size = 9, face = "bold"),
        axis.title.x=element_blank(), axis.text.x = element_text(size=8, angle = 90, hjust=1, vjust = .5),
title = element_text(size=7), legend.key = element_rect(fill = NA, colour = NA, size = 0.25))


p4.2 <- ggplot(rates, aes(x=Year, y=GDP)) +
  stat_smooth(method = lm, se = FALSE, color = "black", size = 0.5) +
  geom_line(color = "green") +
  ggtitle("4.2 Avg Worldwide GDP per capita USD 1975-2015:") +
  ylab("GDP per capita (current USD)") +
  scale_x_continuous(lim=c(1975,2015), breaks=seq(1975,2015,2)) +
  theme(panel.grid.major = element_blank(), panel.background = element_blank(), axis.line = element_line(col
our = "black", size = 0.25), plot.title = element_text(size = 9, face = "bold"),
        axis.title.x=element_blank(), axis.text.x = element_text(size=8, angle = 90, hjust=1, vjust = .5),
title = element_text(size=7), legend.key = element_rect(fill = NA, colour = NA, size = 0.25))


p4.3 <- ggplot(rates, aes(x=Year, y=Literacy.Rate)) +
  geom_line(color = "blue") +
  stat_smooth(method = lm, se = FALSE, color = "black" , size = 0.5) +
  scale_x_continuous(lim=c(1975,2015), breaks=seq(1975,2015,2)) +
  scale_y_continuous(name="Literacy rate, adult total (% Ages 15+): ", labels = function(x){ paste0(x, "%")
}) +
  ggtitle("4.3 Avg Worldwide adult literacy rates 1975-2015:") +
  theme(panel.grid.major = element_blank(), panel.background = element_blank(), axis.line = element_line(col
our = "black", size = 0.25), plot.title = element_text(size = 8, face = "bold"),
        axis.title.x=element_blank(), axis.text.x = element_text(size=8, angle = 90, hjust=1, vjust = .5),
title = element_text(size=7), legend.key = element_rect(fill = NA, colour = NA, size = 0.25))

# plot the graphs together using the multiplot function.
multiplot(p4.1, p4.2, p4.3, cols=2)
```

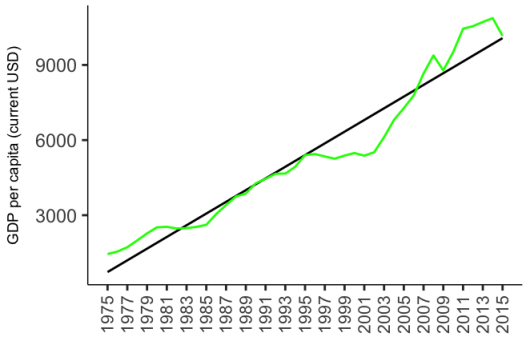
4.1 Worldwide Fertility rates 1975-2015:



4.3 Avg Worldwide adult literacy rates 1975-2015:



4.2 Avg Worldwide GDP per capita USD 1975-2015:



The above graphs show a interesting relationship between fertility & per capita income & education from 1975-2015. As literacy rates and GDP rates rise worldwide fertility rates decrease. You can definetly see a clear correaltion between the 3 graphs. Graph 4.2 and 4.3 are almost mirrors of graph 4.1. You could almost overlay graph 4.2 onto 4.3. This goes with the trend of people being more educated having less kids and also people being richer having access to better health care and better education.

Conclusion:

In conclusion if we take all the above graphs we can see evidence that supports Matt Ridley's hypothesis. In grpahs 2, 3 & 4.1 we can see the worlds fertility rates are nearling Matt's replacement rate of 2.1 births per women. Combine this with graphs 1.1 and 1.1 with the increase in population getting closer to the proposed 9 billion cut off point and followed by the decrease in the percentage increase compared to the previous year we have alot of evidence in favour of the hypothesis. Finally looking at the last three graphs 4.1, 4.2 and 4.3 we see the downward trend in worldwide fertility rates correlationg with the upward trend in GDP per capita and literacy rates worldwide. This suggest the world is getting richer and people are getting more education which again is leading the the replacemnt fertility rate of 2.1. We can safely say that the above visualtions are appropaite in supporting Matt Ridley's hypothesis.

Group member contribution:

Kevin Derrane: I did part four: Observed relationship between fertility & per capita income & education. I did the three graphs for this followed by the write up about them. I also put together this R Markdown file.

Robbie Deegan: I did parts two: Changing fertility rates from 1950 to the present for selected countries of the world & also part three: Regional view of the above visualisation. I created the graphs and I also did my write up on these.

Jamie O'Halloran: I did part one: Percentage increase in world population. I created the graphs and I also did my write up on these.

Group: As a group we all feel we contriubted evenly to this project. We all worked with each other to keep a theme for our graphs and helped each other where necessary. We did the introduction and conclusion together.