1. Overview of project

The object to of my project is to find out whether there is any relationship between birth rates of various countries in this world and percentage of internet users in that country

2. Data gathered, source and variables it contains

I have gathered major part of my data from superdatascience.com and I have modified this raw data accordingly to suit my needs.

	A	B	C	D	E
1	Country.Name	Birth.rate	Internet.users	Income.Group	region
2	Afghanistan	35.253	5.9	Low income	Asia
3	Albania	12.877	57.2	Upper middle income	Europe
4	Algeria	24.738	16.5	Upper middle income	Africa
5	Angola	45.985	19.1	Upper middle income	Africa
6	Antigua and Ba	16.447	63.4	High income	North America
7	Argentina	17.716	59.9	High income	South America
8	Armenia	13.308	41.9	Lower middle income	Asia
9	Aruba	10.244	78.9	High income	North America
10	Australia	13.2	83	High income	Oceania
11	Austria	9.4	80.6188	High income	Europe
12	Azerbaijan	18.3	58.7	Upper middle income	Asia
13	Bahamas, The	15.339	72	High income	North America
14	Bahrain	15.04	90.0000397	High income	Middle East
15	Bangladesh	20.142	6.63	Lower middle income	Asia
16	Barbados	12.188	73	High income	North America
17	Belarus	12.5	54.17	Upper middle income	Europe
18	Belgium	11.2	82.1702	High income	Europe
19	Belize	23.092	33.6	Upper middle income	North America
20	Benin	36.44	4.9	Low income	Africa

Screenshot of sample data

My data contains five variables, those are country name, birth rate, internet users, income group, region (continent)

Definition of these variables and how they are calculated:

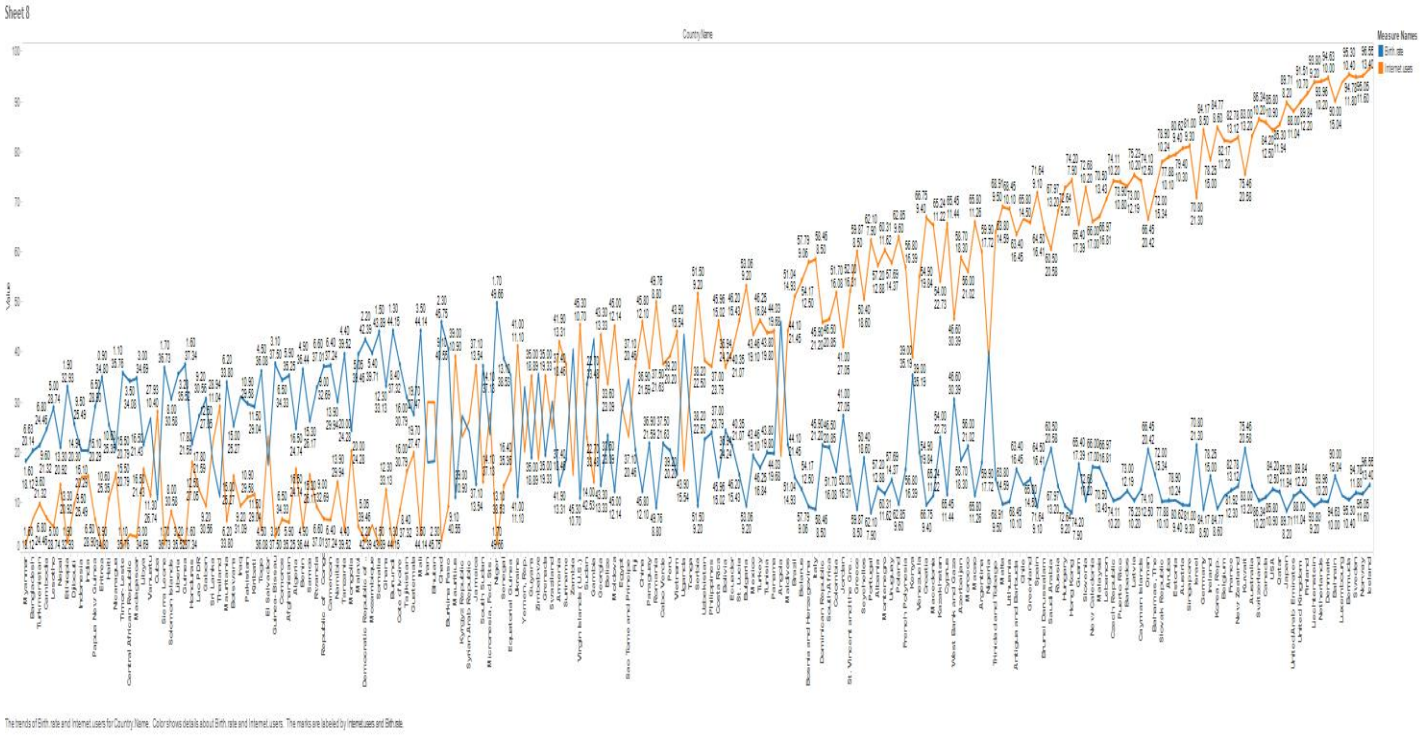
1. Birth rate: The birth rate (births/population rate) is the total number of live births per 1,000 of a population in a year.

2. Internet users: calculated based on what percentage of population has access to internet

3. Income group: categorized based on per capita income of the country.

4. Region: categorized based on the continent a country belongs to.

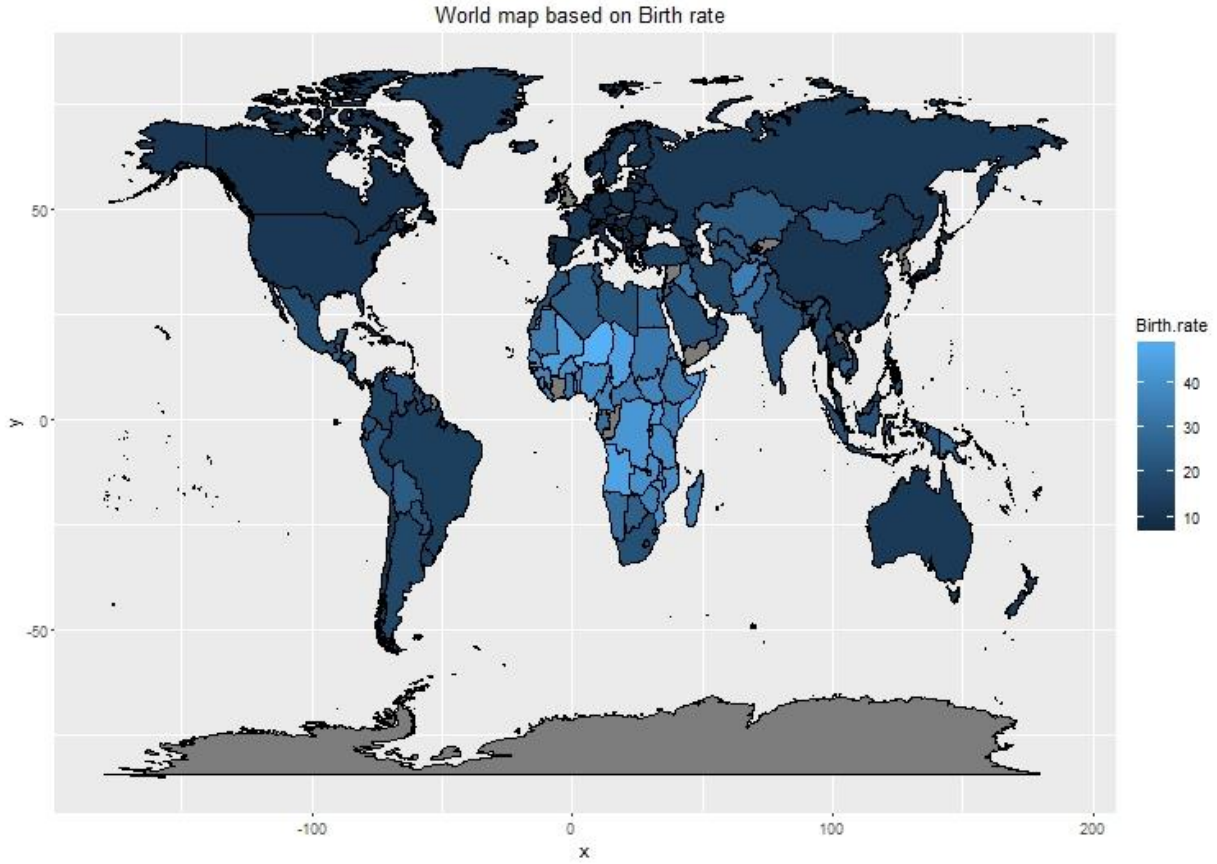
3. Plots



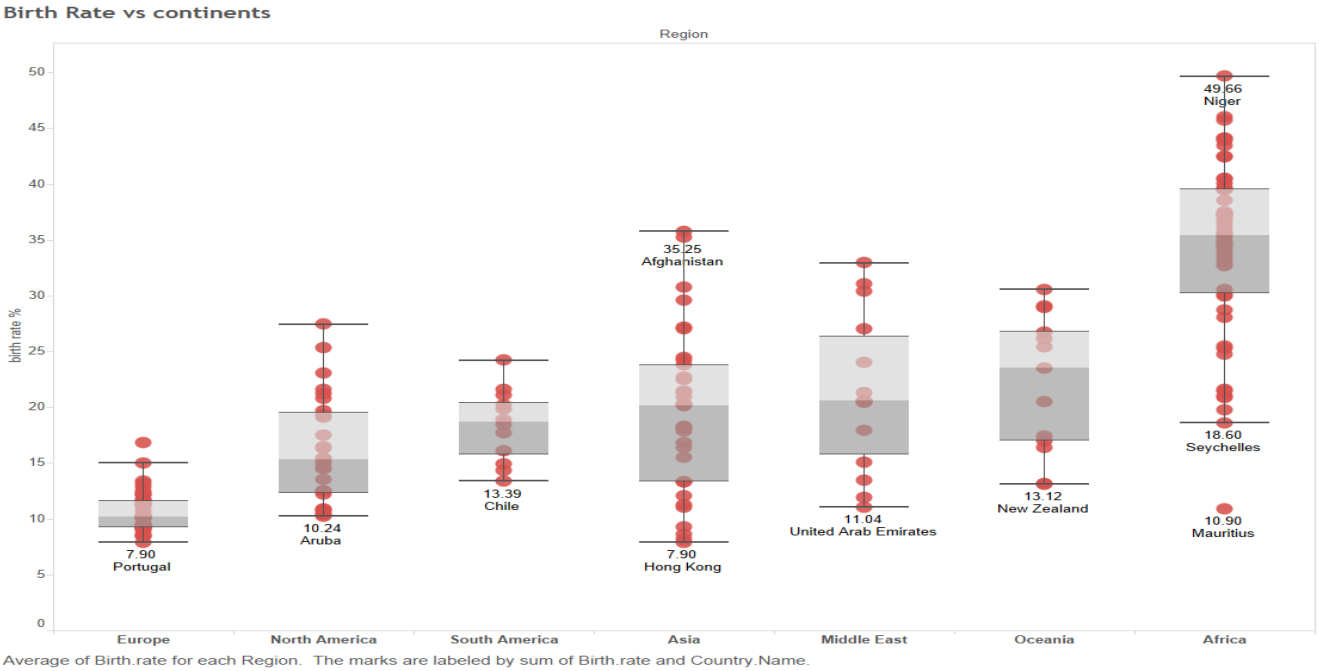
plot 1: this is plot that represents birth rate and internet users in country.

On X axis we have nearly 195 country names and blue line indicate birth rate and orange line indicate internet users

The plot is unclear as we represented 195 countries in a single plot. This plot is created to give us a glimpse that that there might be a relationship between internet users and birth rate.

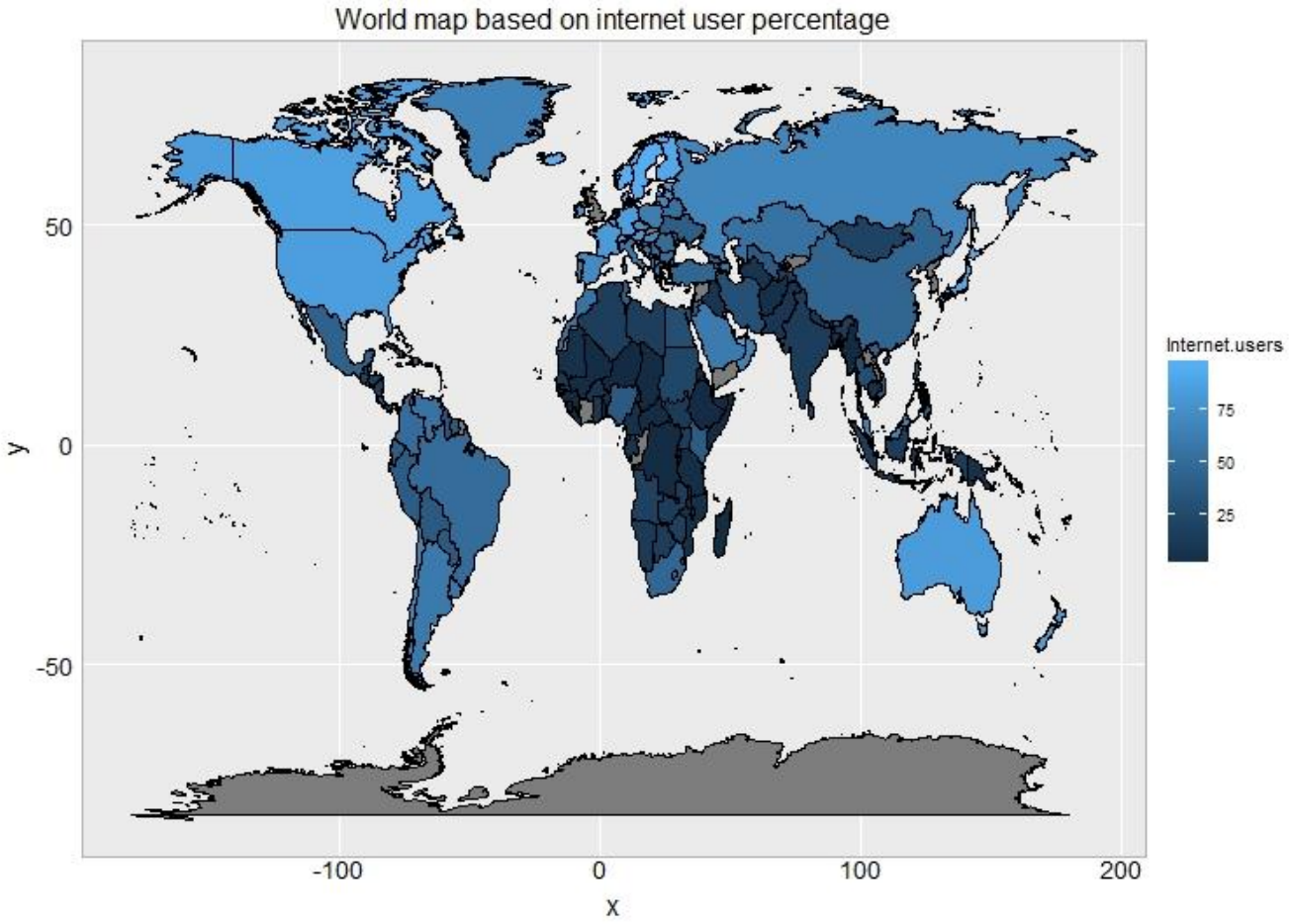


Plot 2: countries based on birth rate



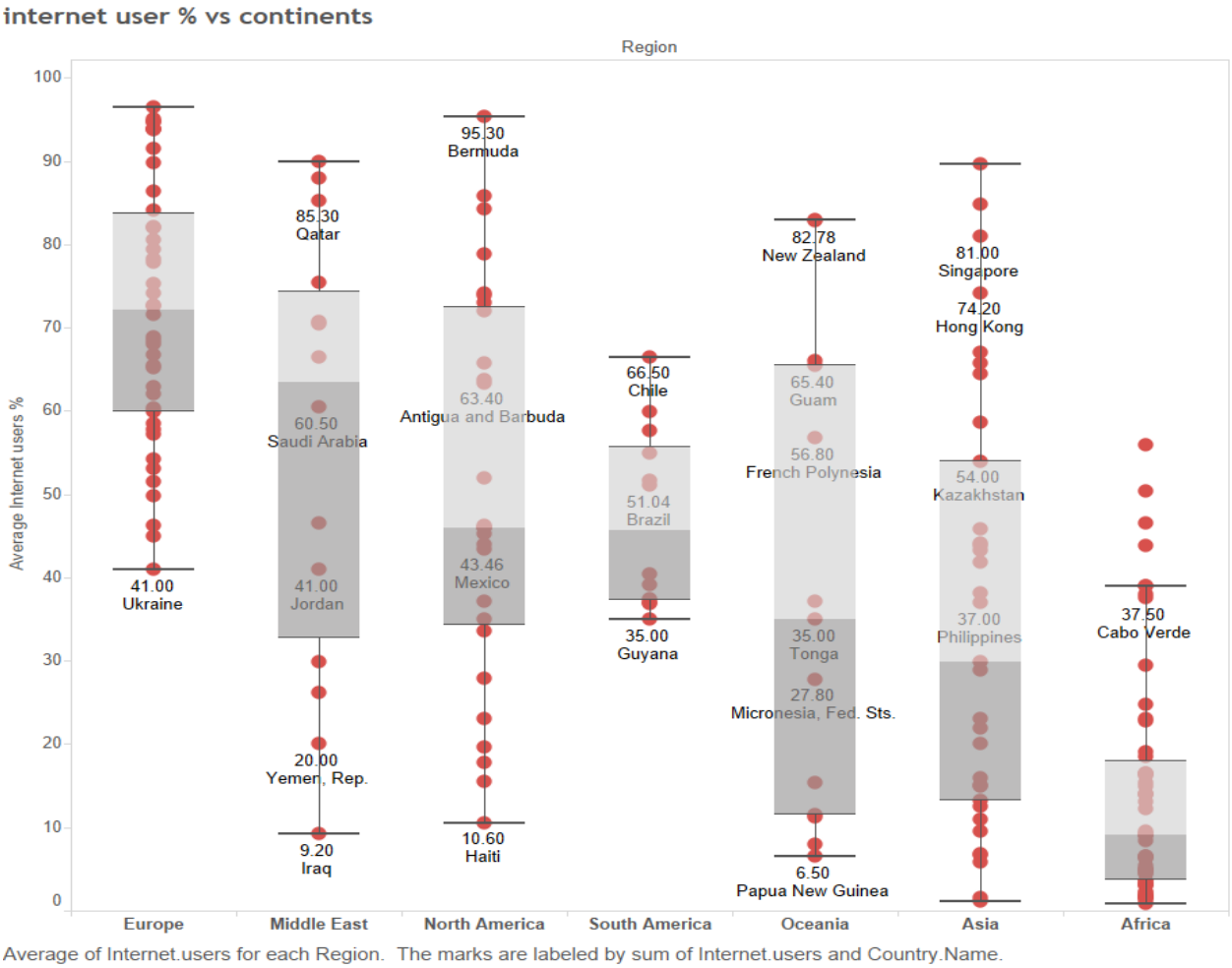
Plot 3: countries based on birth rate categorized based on regions

This plot was created based on the data mapped to various counties in the world. From this data we can observe that countries in Africa has most birth rate with Nigeria having 49.66 as birth rate. In North America majority of counties have very low birth rate with countries having as low as 10%. Europe we have countries with least birth rate. Asia is mix where we can see countries with very high birth rate and low birth rate. Middle east too has above average birthrate. Oceania which consist of Australia has got 2nd highest average birth rate.



Plot 4: countries based on Internet Users.

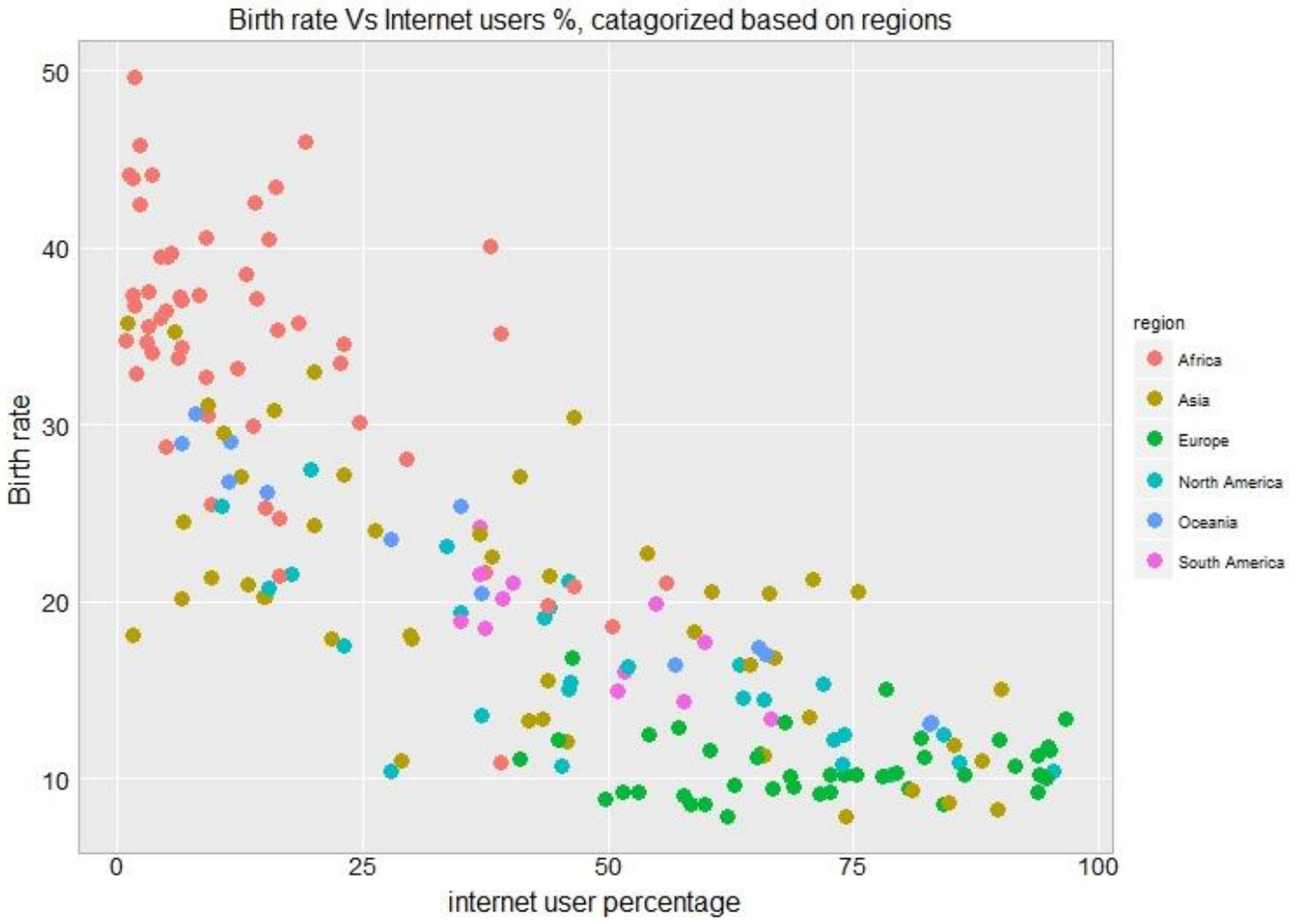
This plot was created based on the data mapped to various counties in the world. From the box plot and world map we can see that European counties has most % of internet users and Africa being the least. They are countries in Asia where the internet users are as low those living in Africa.



Plot 5 : box plot based on regions vs internet users

From this box plot we can see that Europe has average internet users more than 70%. Middle East has around 65% which is quite surprising. North America as countries which got internet users to be around 10.6 % which is very low. In Asia we have maximum variance not so surprising as it has very technologically advanced countries and very poor countries at same time.

4. Insights



Plot 6: birth rate vs internet users categorized based on regions

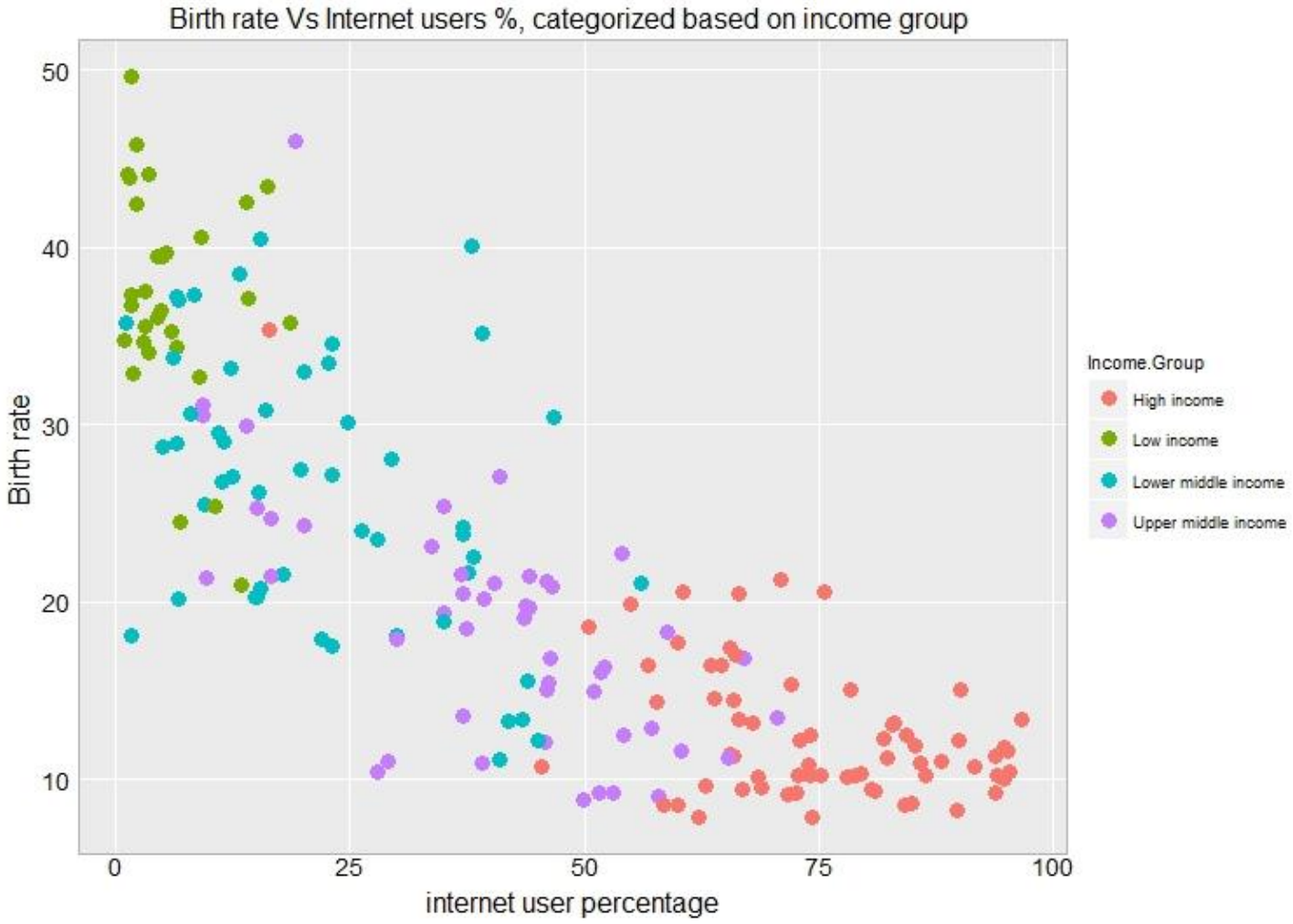
From this plot we can see that countries with low birth rate tend to have high number of internet users and countries with high birth rate have low number of internet users.

Based on regions we can find that African countries have low number of internet users and high birth rate as we more towards right on x- axis we find only European countries in majority.

From the previous plot we feel that percentage of internet users have some form influence of birth rate but it is a very wrong feeling as

Correlation ≠ Causation

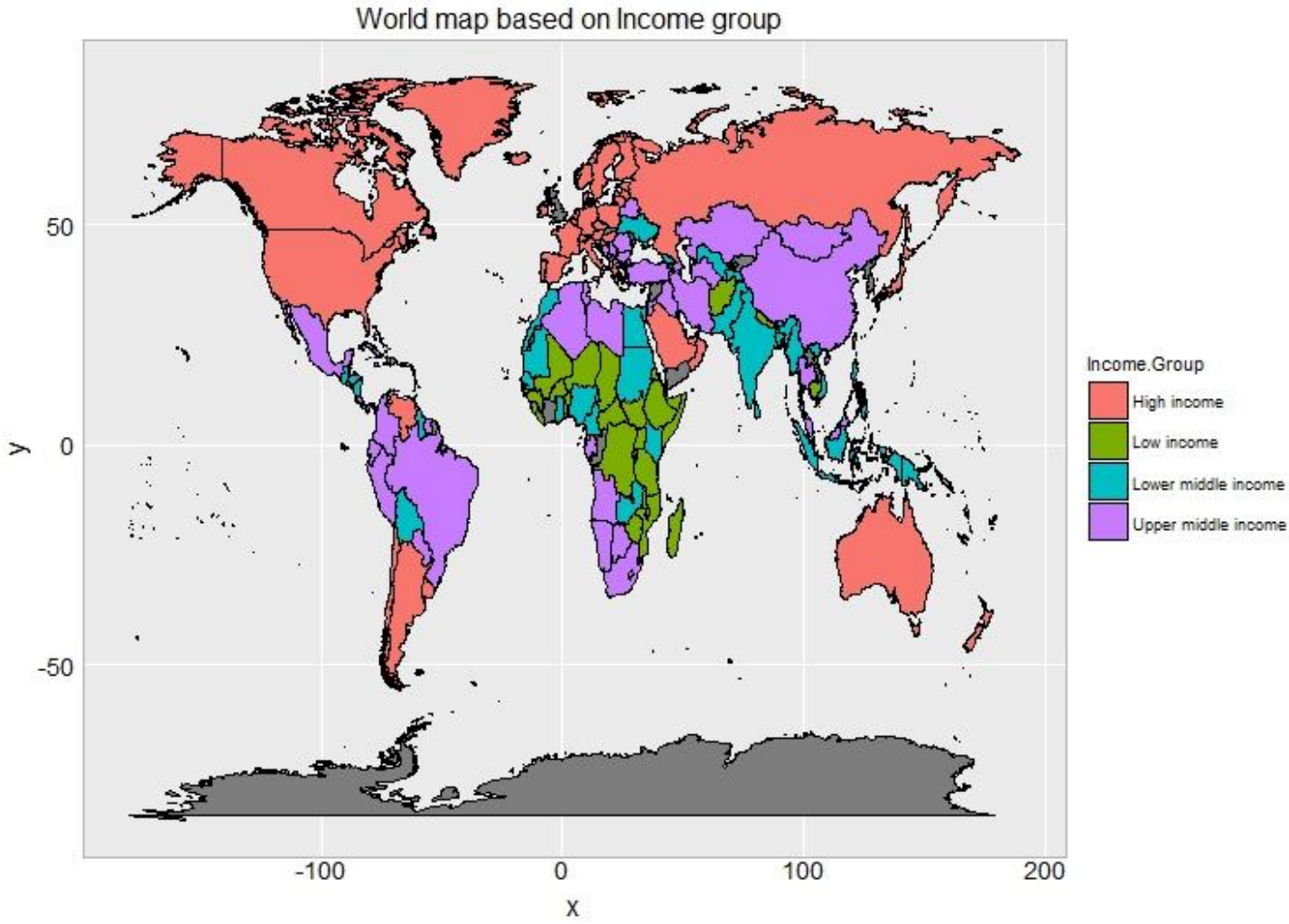
Here is an another plot based on income groups



Plot: 7 birth rate vs internet users categorized based on income group

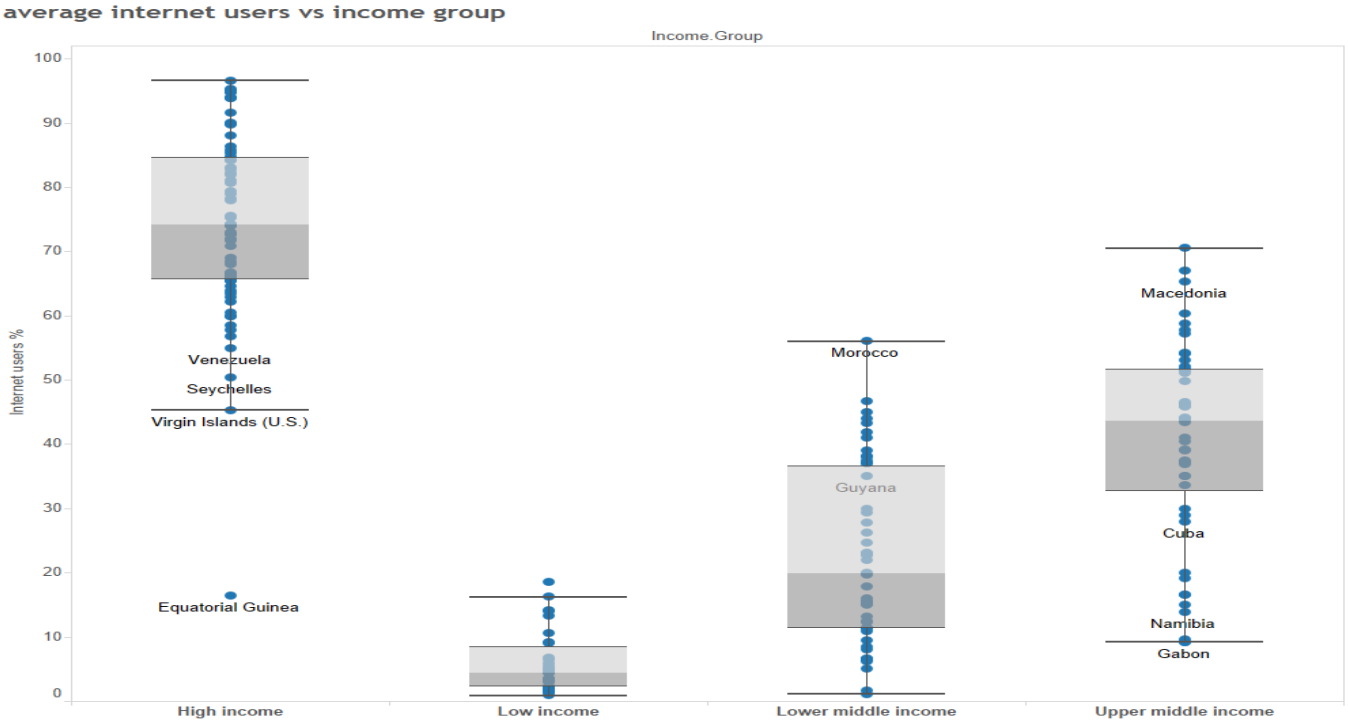
From the previous plot it is quite clear that the income group as important factor behind high percentage of internet users and low birth rate.

It is very clear as the countries belonging to lower income group can't afford internet and neither primary health care services thus they tend to have high birthrate the condition is vice-versa with countries belonging to high income group.

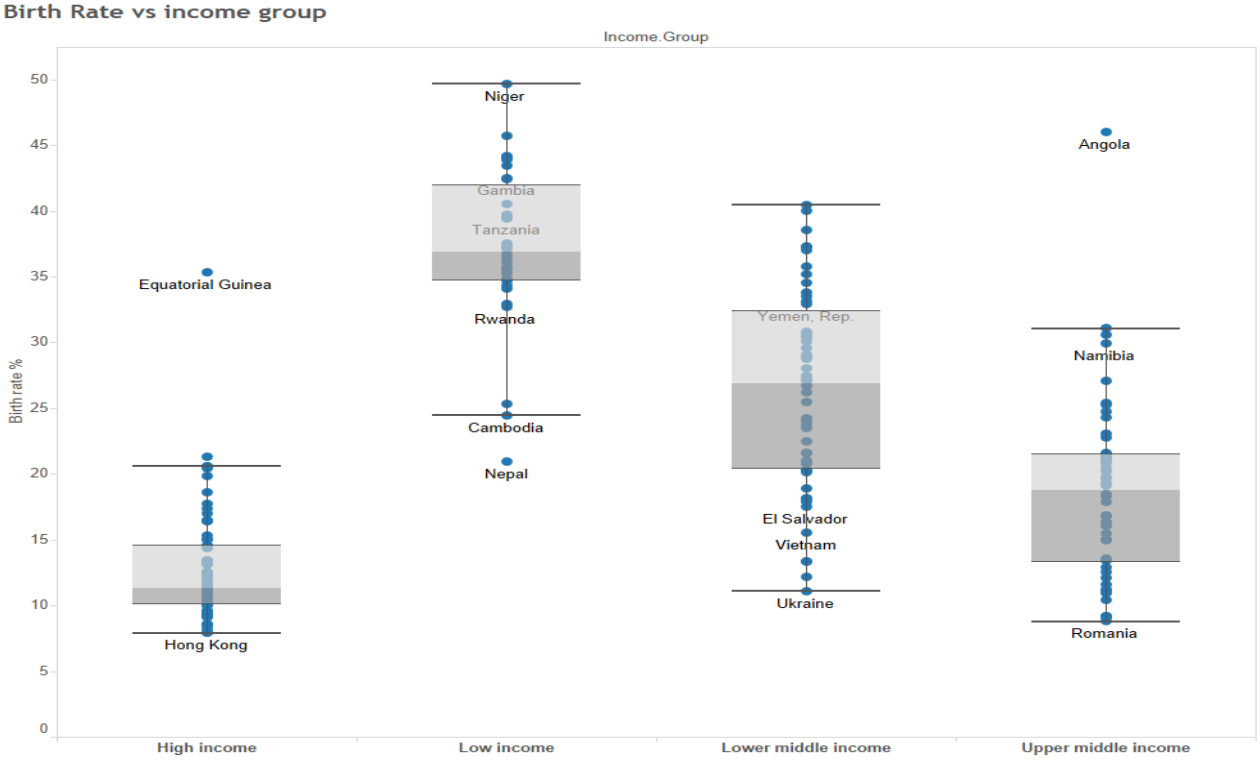


Plot 8 : countries based on income group

From the plot it can be observed that the countries in the high income group tend to be in Europe and as expected majority of low income group counties are in Africa



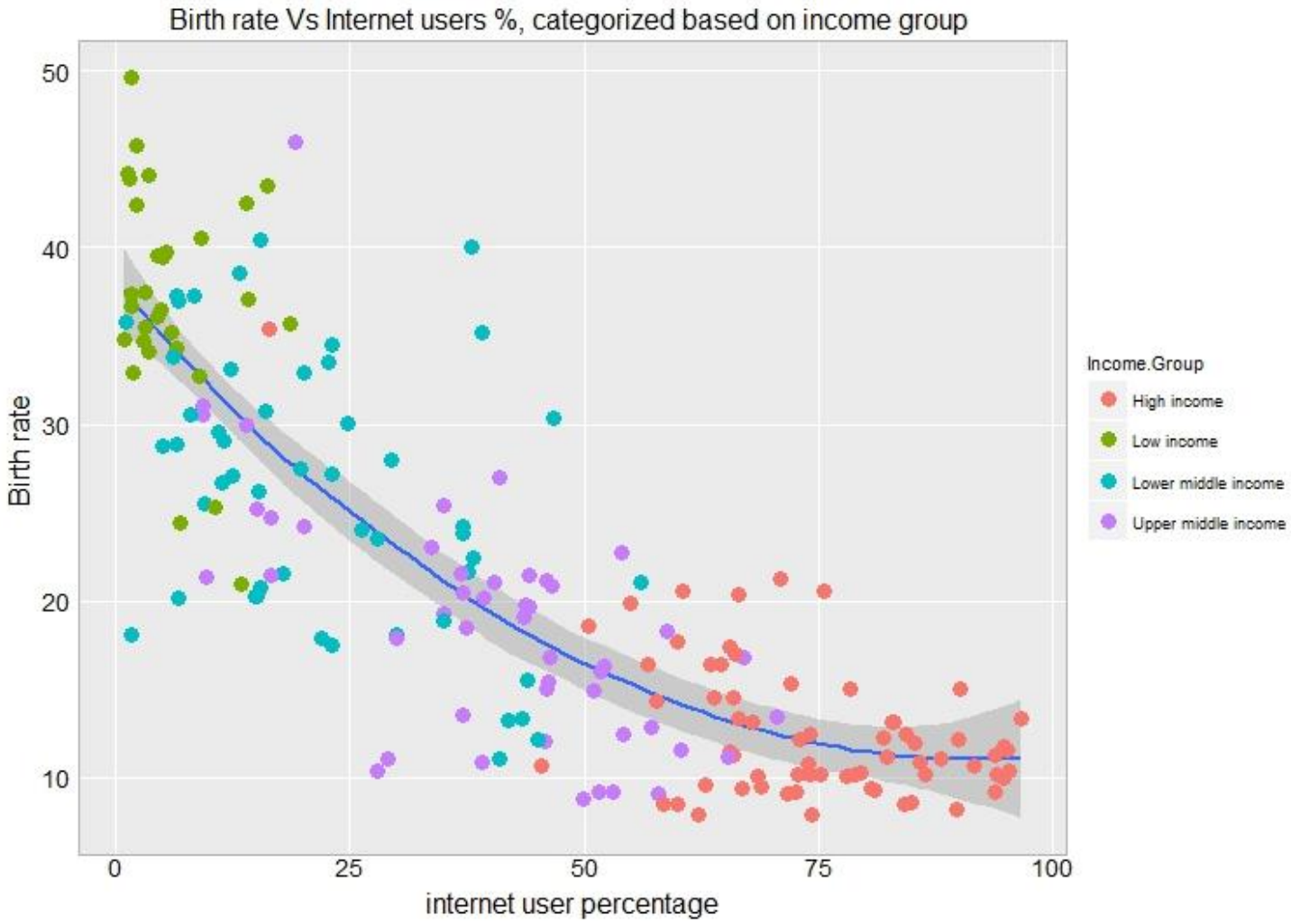
Plot 9 : internet users based on income group



Plot 10: internet users based on income group

From these box plot it can be observed that the income groups with high percentage of internet users tend to have low birth rate and the income groups with high birth rate end to have low percentage of internet users.

5. Regression analysis



We can see that the line which is close to majority of dots.

```
> regression_model = lm(data_frame$Birth.rate~ poly(data_frame$Internet.users, 2) )
> summary(regression_model)

Call:
lm(formula = data_frame$Birth.rate ~ poly(data_frame$Internet.users,
    2))

Residuals:
     Min       1Q   Median       3Q      Max
-18.7399  -2.9161  -0.1632   3.1099  19.9940

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                       21.470      0.408  52.628  < 2e-16 ***
poly(data_frame$Internet.users, 2)1 -120.476      5.697 -21.148  < 2e-16 ***
poly(data_frame$Internet.users, 2)2   32.783      5.697   5.755 3.39e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.697 on 192 degrees of freedom
Multiple R-squared:  0.7144,    Adjusted R-squared:  0.7115
F-statistic: 240.2 on 2 and 192 DF,  p-value: < 2.2e-16
```

Here is the output of the quadratic analysis performed in R. birth rate being dependent variable and internet users as regressior

We have P value less than 0.05 these means our findings are statically significant, proving there is clearly a trend.

6. Conclusion

People from countries who belong to high income group can afford internet and the birth rate in these countries is low, may be its because of this they are wealthy enough to afford to visit family planning center and this leads to low birth rate in these countries.

7. Reference

- Wikipedia pages
- Stackoverflow.com forums
- YouTube videos

8. Appendix

- Added continents to my data set on my own
- Code
    - library(ggplot2)
    - library(rworldmap)
    - library(maps)
    - library(MASS)
    - library(grid)
    - hw <- theme_gray()+ theme(
    - strip.text.y = element_blank(),
    - strip.background=element_rect(fill=rgb(.9,.95,1),
    - colour=gray(.5), size=.2),
    -
    - panel.border=element_rect(fill=FALSE,colour=gray(.70)),
    - panel.grid.minor.x = element_blank(),
    - panel.grid.minor.y = element_blank(),
    - panel.margin.x = unit(0.10,"cm"),
    - panel.margin.y = unit(0.05,"cm"),
    -
    - axis.ticks=element_blank(),
    - axis.text=element_text(colour="black"),
    - axis.text.y=element_text(margin=margin(0,3,0,3), size = 14),
    - axis.text.x=element_text(margin=margin(-1,0,3,0), size = 14),
    - axis.title= element_text(size=16),
    - plot.title = element_text(size=16)
    -)
    -
    - data_frame<- read.csv(file.choose())
    -
    -
    -
    - # plot by continents----------
    - ggplot(data_frame, aes(x= Internet.users, y = Birth.rate, color = region))+geom_point(size = 4)+
    - ggtitle("Birth rate Vs Internet users %, catagorized based on regions ") +
    - xlab("internet user percentage")+
    - ylab("Birth rate ")+hw
    -
    -
    - #boxplots
    -
    - #birth rate
    - ggplot(data_frame, aes(x= region, y = Birth.rate)) + geom_boxplot(alpha=.3)
    -

```
#continets

ggplot(data_frame, aes(x= region, y = Internet.users )) + geom_boxplot(
alpha=.3)
#------------------------------------------------------------------------


#plot by income group ----------------------------------------------------


ggplot(data_frame, aes(x= Internet.users, y = Birth.rate, color =
Income.Group))+geom_point(size= 4)+
  ggtitle("Birth rate Vs Internet users %, categorized based on income group ")+
  xlab("internet user percentage")+
  ylab("Birth rate ") +hw

#boxplots

#birth rate
ggplot(data_frame, aes(x= Income.Group, y = Birth.rate )) + geom_boxplot(
alpha=.3)

#continets

ggplot(data_frame, aes(x= Income.Group, y = Internet.users )) + geom_boxplot(
alpha=.3)

#------------------------------------------------------------------------

# regression model----------------------------------------------------------

#PLOT
ggplot(data_frame, aes(x= Internet.users, y = Birth.rate))+ geom_smooth() +
  geom_point(aes(color= Income.Group), size = 4)+
  ggtitle("Birth rate Vs Internet users %, categorized based on income group ")+
  xlab("internet user percentage")+
  ylab("Birth rate ") +hw



regression_model = lm(data_frame$Birth.rate~ poly(data_frame$Internet.users,
2) )

#quadratic annalysis


```

```
summary(regression_model)

#-----------------------------------------------------------------------

#--------- world maps-----------------------------------------------------

#birth rate
world_map <- map_data(map="world")
gg <- ggplot(data_frame)

gg <- gg + geom_map(dat=world_map, map = world_map,
                    aes(map_id=region), fill="gray49", color="black")

gg <- gg + geom_map(map = world_map,
                    aes(map_id = Country.Name, fill = Birth.rate ), colour = "black")


gg <- gg + expand_limits(x = world_map$long, y = world_map$lat) +
ggtitle("World map based on Birth rate ")+hw

gg

#internet users
world_map <- map_data(map="world")
gg <- ggplot(data_frame)

gg <- gg + geom_map(dat=world_map, map = world_map,
                    aes(map_id=region), fill="gray49", color="black")

gg <- gg + geom_map(map = world_map,
                    aes(map_id = Country.Name, fill = Internet.users ), colour =
"black")


gg <- gg + expand_limits(x = world_map$long, y = world_map$lat) +
ggtitle("World map based on internet user percentage")+hw

gg


#income group

world_map <- map_data(map="world")
gg <- ggplot(data_frame)

gg <- gg + geom_map(dat=world_map, map = world_map,
```

```
                aes(map_id=region), fill="gray49", color="black")

gg <- gg + geom_map(map = world_map,
                aes(map_id = Country.Name, fill = Income.Group ), colour =
"black")


gg <- gg + expand_limits(x = world_map$long, y = world_map$lat)
+ggtitle("World map based on Income group ")+hw


gg

#---------------------------------------------------------------------------
```