DATA100 Group Project – 2019 Fall

Ronak Arora, Tirth Desai, Avery Dowling, Zofia Wajda

28/11/2019

List of Contents:

- Introduction
- Political Situation
- Health Care
- Economics
- Corelations between different factors
- Conclusions


```
library(tidyverse,warn.conflicts = FALSE, quietly = TRUE, verbose = FALSE)
```

```
## -- Attaching packages ---------------------------------------------------------
------------------------------------------------------------- tidyverse 1.2.1 --
```

```
## <U+221A> ggplot2 3.2.1     <U+221A> purrr   0.3.2
## <U+221A> tibble  2.1.3     <U+221A> dplyr   0.8.3
## <U+221A> tidyr   1.0.0     <U+221A> stringr 1.4.0
## <U+221A> readr   1.3.1     <U+221A> forcats 0.4.0
```

```
## -- Conflicts ------------------------------------------------------------------
---------------------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(modelr)
library(broom)
```

```
##
## Attaching package: 'broom'
```

```
## The following object is masked from 'package:modelr':
##
##     bootstrap
```

```
library(tidyr)
```

Analysis of World Happiness among different

countries - Introduction

The World Happiness Report is a landmark survey of the state of global happiness that ranks 156 countries (“Country” - character value) by how happy their citizens perceive themselves to be (“Happiness score” - double variable). Data set includes 156 observations.

```
happy2018 <- read_tsv("WorldHappinessReport2018-Score.csv",
                          col_types = cols(
                            .default = col_double(),
                            Country = col_character()
                          ))
glimpse(happy2018)
```

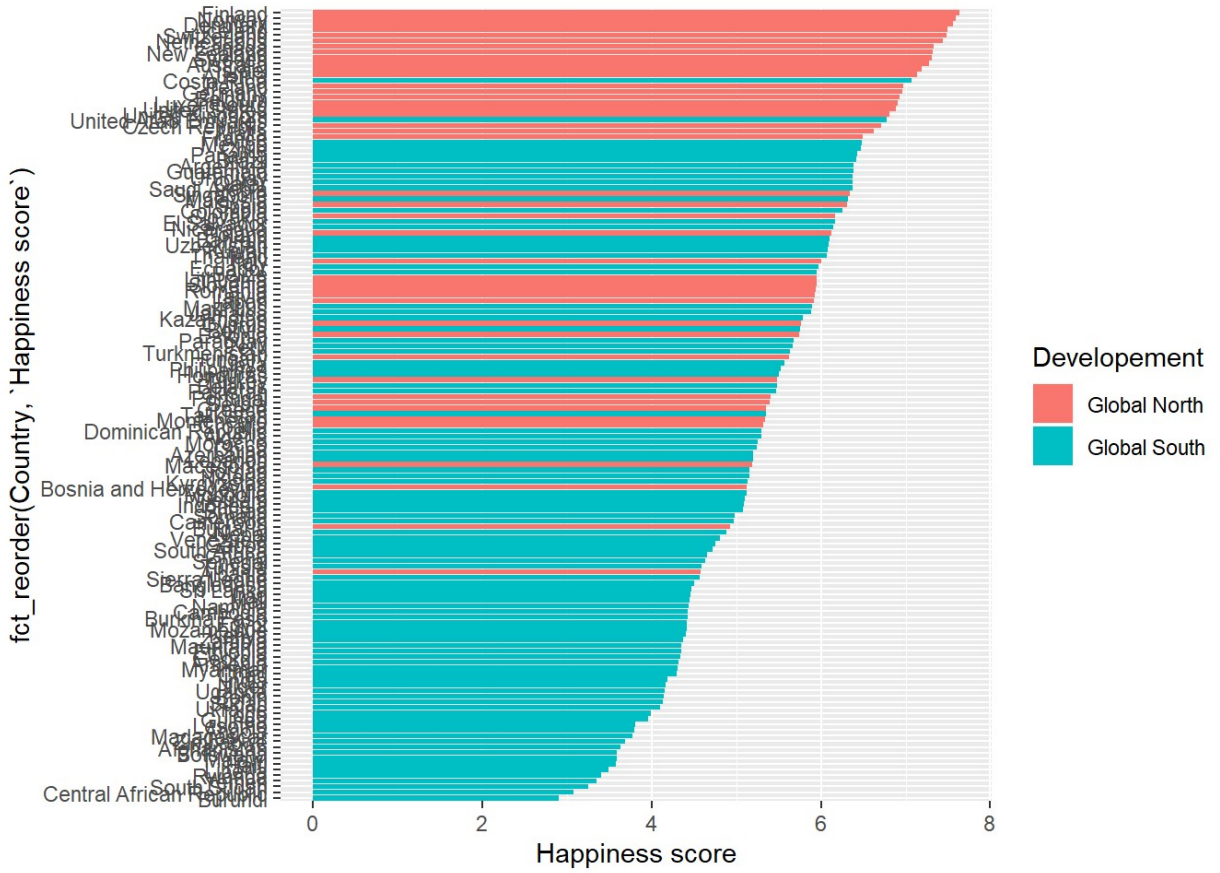
```
## Observations: 156
## Variables: 2
## $ Country           <chr> "Finland", "Norway", "Denmark", "Iceland", "...
## $ `Happiness score` <dbl> 7.632, 7.594, 7.555, 7.495, 7.487, 7.441, 7....
```

The map shows that regions with the happiest citizens are Australia, North America and Western and Northern Europe. The lowest score on the other hand is obtained by Africa and Middle East. This allows us to form a thesis that happiness is related to developement - people in developed countries are happier than those in developing ones. “WorldRegions.csv” data set contains the list of countries along with region and information whether the country is developed or not. Variables: Country(character), Region(character), Global South(“Global South”=developing, “Global North”=developed)(character). There are 258 observations.

```
regionclassification <- read_tsv("WorldRegions.csv",col_types = cols(Region=col_fac
tor()))
regionclassification<-regionclassification%>%
  mutate(Region=fct_recode(Region,"Asia & Pacific"="Asic & Pacific"))
names(regionclassification)[3]<-"Developement"
regionclassification<-semi_join(regionclassification,happy2018,by="Country")%>%
  filter(!is.na(Developement))
glimpse(regionclassification)
```

```
## Observations: 140
## Variables: 3
## $ Country      <chr> "United Arab Emirates", "Afghanistan", "Albania",...
## $ Region       <fct> Arab States, Asia & Pacific, Europe, Europe, Afri...
## $ Developement <chr> "Global South", "Global South", "Global North", "...
```

```
ggplot(inner_join(happy2018, regionclassification, by="Country"))+
  geom_col(mapping=aes(x=fct_reorder(Country, `Happiness score`), y=`Happiness scor
e`, fill=Developement))+
  coord_flip()
```



In fact turs out to be true that Global North (developed countries) tend to score higher than Global South (developing countries). It is not a clear distinction, but there is some connection between those two factors.

In the further parts of the project we will investigate deeper which characteristics of the countries affect their happiness score the most.

I Political situation

Predictions: One of the factors that I suspect affects people's happiness is political situation and level of democratisation in the country. As far as I can guess for now the more democratic the country is the happier are its citizens, because democracy often goes in the line with personal freedom, human rights, safety and good care over citizens.

To investigate that I will be using DEMOCRACYINDEX data set, that contains score in five different categories and a score that is a resultant of them (Score - double). Based on that score the countries are assigned to their regime type (full democracy, flawed democracy, hibrid regime and authoritarian). There are 167 observations.

```
democracyindex <- read_tsv("DEMOCRACYINDEX.csv")
```

```
## Parsed with column specification:
## cols(
##   Rank = col_double(),
##   Country = col_character(),
##   Score = col_double(),
##   `Electoral process and pluralism` = col_double(),
##   `Functioning of government` = col_double(),
##   `Political participation` = col_double(),
##   `Political culture` = col_double(),
##   `Civil liberties` = col_double(),
##   `Regime type` = col_character(),
##   Continent = col_character()
## )
```

```
#I'm leaving only those countries, that also appear in WorldHappinessReport2018-Score.csv
democracyindex<-inner_join(democracyindex,happy2018,by="Country")
glimpse(democracyindex)
```

```
## Observations: 144
## Variables: 11
## $ Rank                            <dbl> 1, 2, 3, 4, 5, 6, 6, 8, 9, 1...
## $ Country                         <chr> "Norway", "Iceland", "Sweden...
## $ Score                           <dbl> 9.87, 9.58, 9.39, 9.26, 9.22...
## $ `Electoral process and pluralism` <dbl> 10.00, 10.00, 9.58, 10.00, 1...
## $ `Functioning of government`     <dbl> 9.64, 9.29, 9.64, 9.29, 9.29...
## $ `Political participation`       <dbl> 10.00, 8.89, 8.33, 8.89, 8.3...
## $ `Political culture`             <dbl> 10.00, 10.00, 10.00, 8.13, 9...
## $ `Civil liberties`               <dbl> 9.71, 9.71, 9.41, 10.00, 9.1...
## $ `Regime type`                   <chr> "Full democracy", "Full demo...
## $ Continent                       <chr> "Europe", "Europe", "Europe"...
## $ `Happiness score`               <dbl> 7.594, 7.495, 7.314, 7.324, ...
```

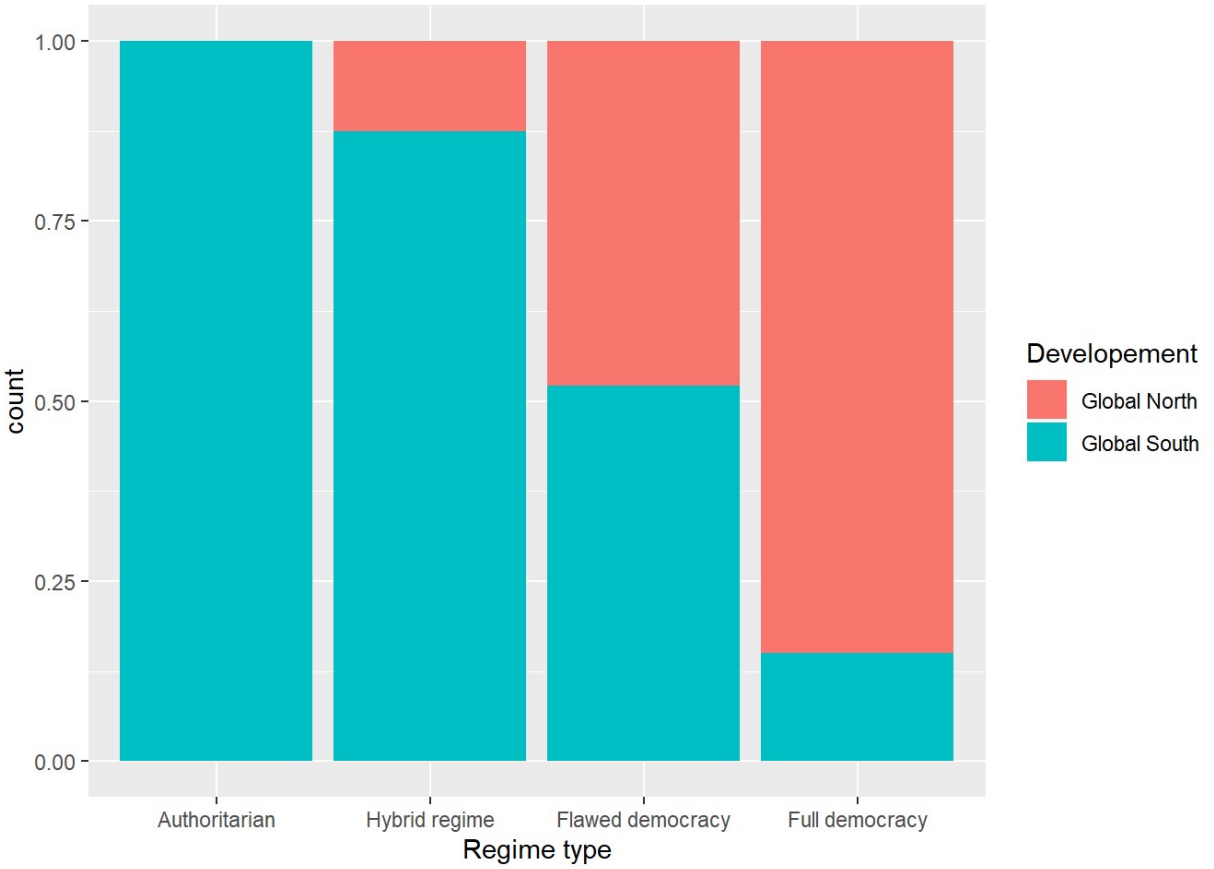
```
summary(democracyindex)
```

```
##       Rank          Country              Score
##  Min.   :  1.00   Length:144         Min.   :1.430
##  1st Qu.: 38.75   Class :character   1st Qu.:3.728
##  Median : 82.00   Mode  :character   Median :5.720
##  Mean   : 80.65                      Mean   :5.631
##  3rd Qu.:120.50                      3rd Qu.:7.312
##  Max.   :166.00                      Max.   :9.870
##  Electoral process and pluralism Functioning of government
##  Min.   : 0.000                  Min.   :0.000
##  1st Qu.: 3.170                  1st Qu.:3.140
##  Median : 7.250                  Median :5.215
##  Mean   : 6.157                  Mean   :5.012
##  3rd Qu.: 9.170                  3rd Qu.:6.790
##  Max.   :10.000                  Max.   :9.640
##  Political participation Political culture Civil liberties
##  Min.   : 1.110          Min.   : 1.880    Min.   : 0.000
##  1st Qu.: 3.890          1st Qu.: 4.380    1st Qu.: 3.748
##  Median : 5.560          Median : 5.630    Median : 6.180
##  Mean   : 5.374          Mean   : 5.697    Mean   : 5.924
##  3rd Qu.: 6.670          3rd Qu.: 6.407    3rd Qu.: 8.240
##  Max.   :10.000          Max.   :10.000    Max.   :10.000
##  Regime type         Continent         Happiness score
##  Length:144         Length:144         Min.   :2.905
##  Class :character   Class :character   1st Qu.:4.446
##  Mode  :character   Mode  :character   Median :5.378
##                                        Mean   :5.390
##                                        3rd Qu.:6.195
##                                        Max.   :7.632
```

In all categories data is only very slightly skewed - the median is a little lower than median. The only exception is political culture. After a short research I found out, that this category is defined by the International Encyclopedia of the Social Sciences as the “set of attitudes, beliefs and sentiments that give order and meaning to a political process and which provide the underlying assumptions and rules that govern behavior in the political system”. That means it is subjective and depends not on political and democratical state of the country, but rather its history, norms and peoples attitude. For this reason it may behave strangely and is unpredictable.

```
democracyindex$`Regime type`<-fct_reorder(democracyindex$`Regime type`,democracyind
ex$Score)
ggplot(inner_join(democracyindex,regionclassification))+
  geom_bar(mapping=aes(`Regime type`,fill=`Developement`),position="fill")
```

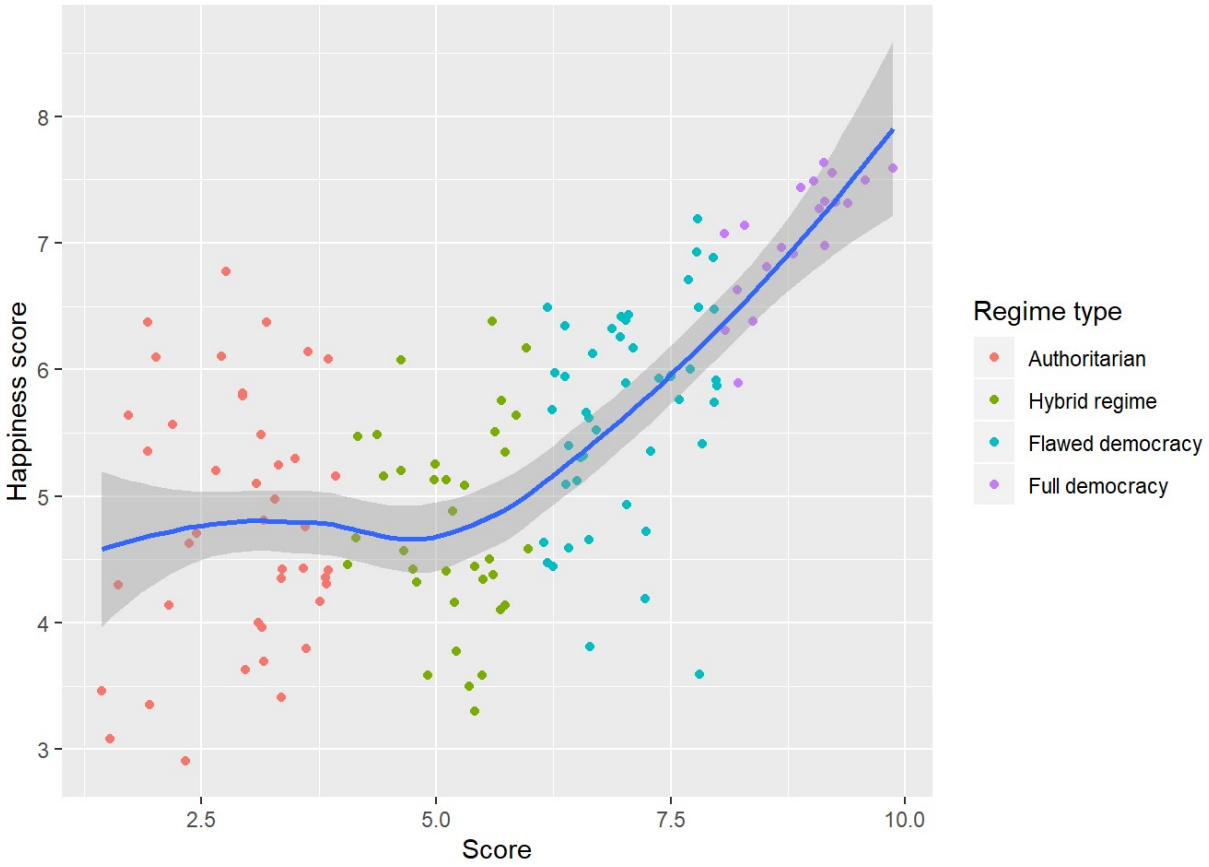
```
## Joining, by = "Country"
```



There is a correlation between level of developement of the country and its regime type - countries in global north tend to be more democratic than in global south. If, as I suspect, democracy has positive affect on happiness then this is one of the reasons one global north is happier than global south.

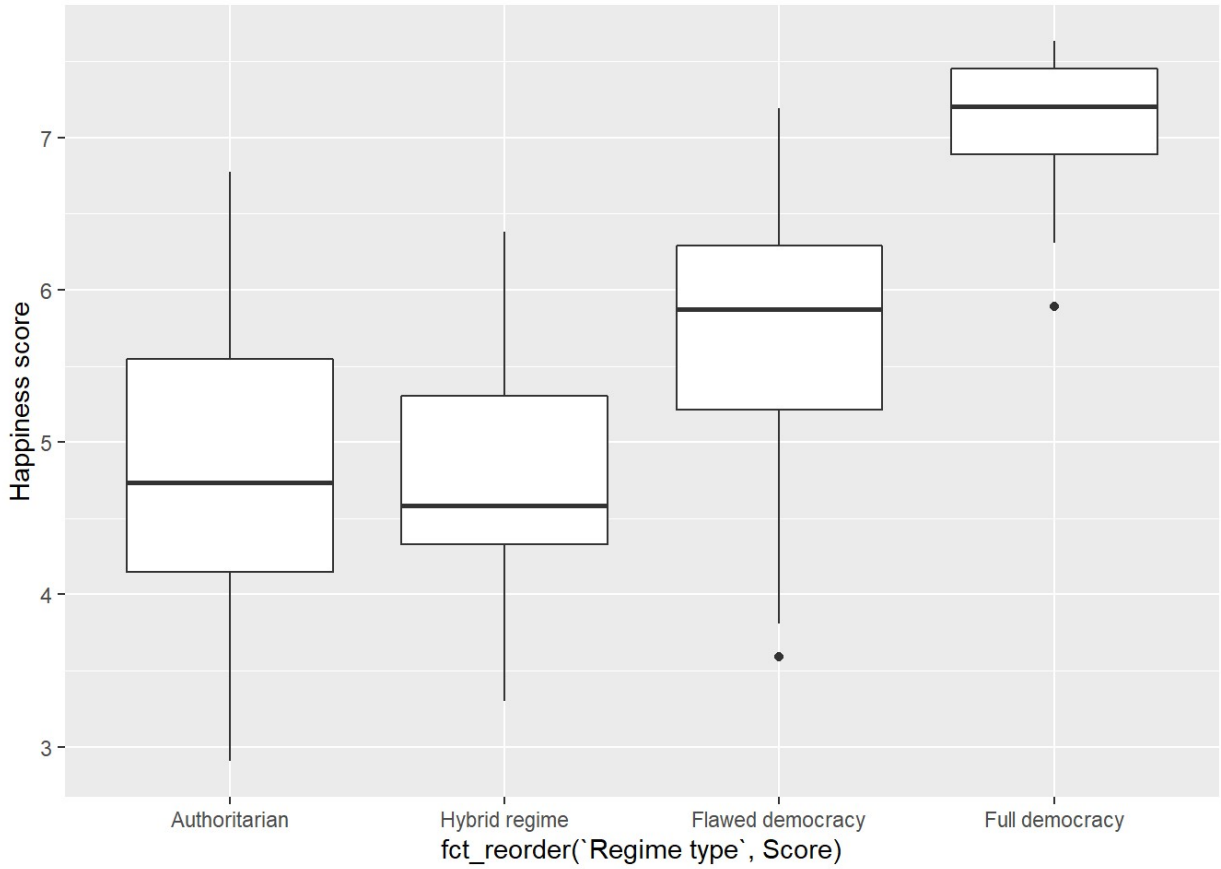
```
plot1<-ggplot(democracyindex,mapping=aes(x=Score, y=`Happiness score`))+
  geom_point(aes(color=`Regime type`))+
  geom_smooth()
plot1
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



The plot suggests that there is in fact a positive corelation between democracy score and happiness in the country, but it is only true for the countries with regime types full democracy, flawed democracy and hybrid regimes with higher score. In less democratic countries there seems to be no corelation between those two factors - the happines does not vary, when level of democracy changes. The insight into different factors does not reveal much either - all the factors behave similarly to the general score of democracy.

```
ggplot(democracyindex,mapping=aes(x=fct_reorder(`Regime type`,Score), y=`Happiness score`))+
  geom_boxplot()
```



In authoritarian countries 1st quantile, median and 3rd quantile are significantly futher away from one another that in other regime types - that means that in this regime type happiness varies much more - so people are less affected by democratic level of the country. One reason may be that in global south people have less access to information so they do not know about the situation they live in so it cannot affect their lifes that much. The data set I will use is INTERNET.cvs, that contains the information about internet users in the countries (226 observations). I'm also using POPULATION.csv (238 observations)

```
internetuser <- read_tsv("INTERNET.csv")
```

```
## Parsed with column specification:
## cols(
##   Rank = col_double(),
##   Country = col_character(),
##   `Internet users` = col_double()
## )
```

```
population <- read_tsv("POPULATION.csv")
```

```
## Parsed with column specification:
## cols(
##   Rank = col_double(),
##   Country = col_character(),
##   Population = col_double(),
##   `Date of Information` = col_character()
## )
```

```
internetuser<-inner_join(internetuser,population,by="Country")%>%
  mutate("Internet users percentage"=round(`Internet users`/Population*100,2))%>%
  select(Country,`Internet users`,`Internet users percentage`,Population)
#For the sake of general assesement, I will create variable that states whether peo
ple in the country have high(higher than average) or low (lower than average) acces
s to internet:
mean_percentage_of_internet_users<-mean(internetuser[["Internet users percentag
e"]], na.rm = FALSE)
internetuser<-internetuser%>%
  mutate("Access to internet"=ifelse(`Internet users percentage`>mean_percentage_of
_internet_users,"High","Low"))
glimpse(internetuser)
```

```
## Observations: 225
## Variables: 5
## $ Country                    <chr> "China", "India", "United States",...
## $ `Internet users`           <dbl> 730723960, 374328160, 246809221, 1...
## $ `Internet users percentage` <dbl> 52.77, 28.86, 74.96, 58.82, 92.39,...
## $ Population                 <dbl> 1384688986, 1296834042, 329256465,...
## $ `Access to internet`       <chr> "High", "Low", "High", "High", "Hi...
```

```
inner_join(democracyindex,internetuser,by="Country")
```

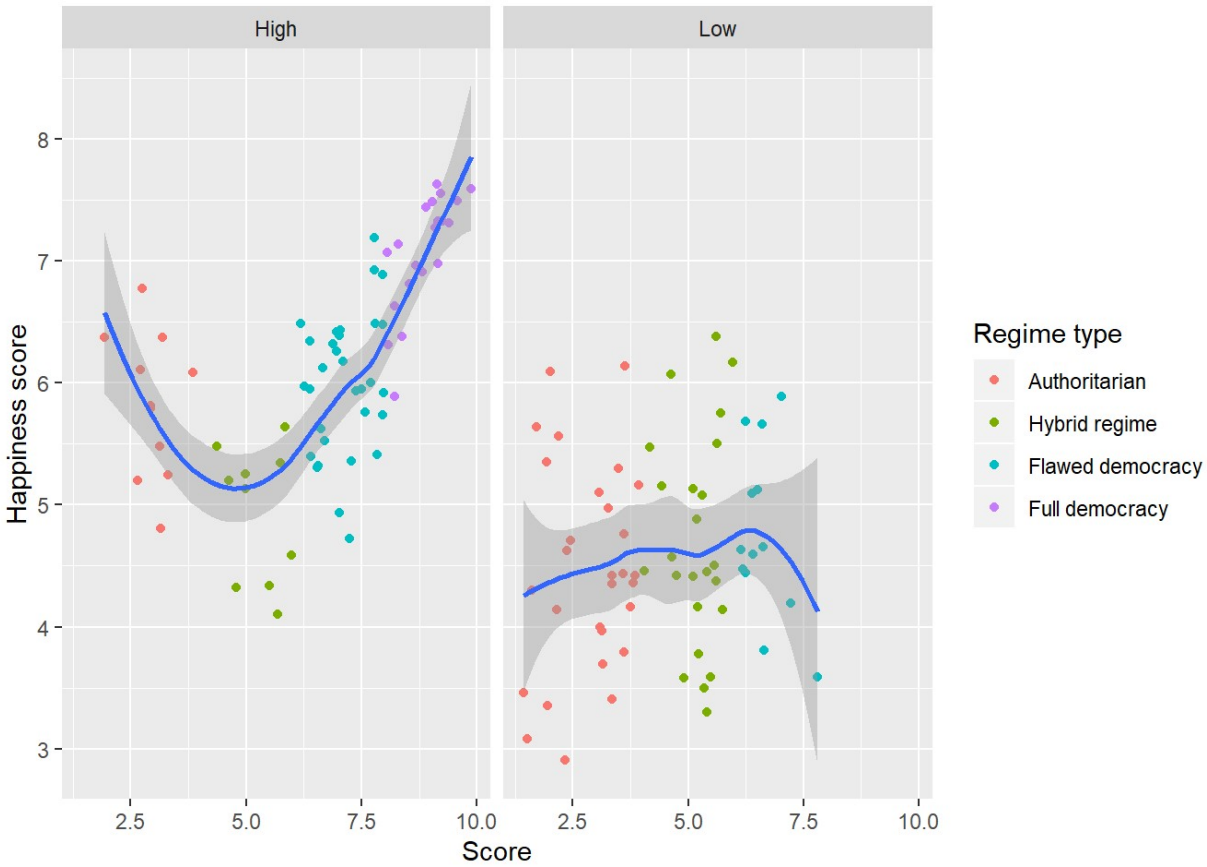
R…	Country	Sc…	Electoral process and pluralism
<dbl>	<chr>	<dbl>	<dbl>
1	Norway	9.87	10.00
2	Iceland	9.58	10.00
3	Sweden	9.39	9.58
4	New Zealand	9.26	10.00
5	Denmark	9.22	10.00
6	Ireland	9.15	9.58
6	Canada	9.15	9.58
8	Finland	9.14	10.00
9	Australia	9.09	10.00

R…	Country	Sc…	Electoral process and pluralism	Fu
<dbl>	<chr>	<dbl>	<dbl>	
10	Switzerland	9.03	9.58	

1-10 of 139 rows | 1-5 of 15 columns Previous **1** 2 3 4 5 6 … 14 Next

```
ggplot(inner_join(democracyindex,internetuser,by="Country"),mapping=aes(x=Score, y=
`Happiness score`))+
  geom_point(aes(color=`Regime type`))+
  geom_smooth()+
  facet_wrap(vars(`Access to internet`))
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



As I can conclude from the graph the level of access to internet (more or less equal to access to independent information in general) impacts not only happiness in general but also the connection between democracy and happiness. Among countries with low access this corelation is basically nonexistent. On the other hand among countries with high access it is stronger: turns out that happiness falls as democracy rises for authoritarian countries, for the rest the ralation is positive.

```
table<-inner_join(democracyindex,internetuser,by="Country")%>%
  group_by(`Regime type`,`Access to internet`)%>%
  count()%>%
  spread(key=`Access to internet`, value=n)
glimpse(table)
```

```
## Observations: 4
## Variables: 3
## Groups: Regime type [4]
## $ `Regime type` <fct> Authoritarian, Hybrid regime, Flawed democracy, ...
## $ High          <int> 11, 10, 32, 20
## $ Low           <int> 29, 24, 13, NA
```

Among authoritarian countries and those with autoritarian regime there are significantly more countries with low acces to internet.

To sum up the whole part, there is in fact positive connection between political situation of the country and its happiness, but only in countries with higher democratic score. In authoritarian countries people do not seem to be affected by the political situation that much. If they are the ralationship is, curiously enough, negative. I have an idea, how to explain both of those phenomena. First of all authoritarian countries (often global south) have less access to internet that the rest of the world. Therefore people do not know what is happening around them and further abroad hence there are not affected by that. The reason why those few authoritarian countries with high access to internet tend to be less happy as level of democracy rises id as follows: one of the characteristics of those countries is governments control over media. So the less democratic the county, the more manipulated people are to believe, that they in fact live in happy country.

II Health Care

I will be using the HEALTHEXP.csv and DEATHRATE.csv files to show that health influences happiness.

Predictions: I think that health and happiness have a strong positive correlation. I predict that as health expenditures increase, the happiness score will increase as well. I think this is true because countries who spend more money on their healthcare system should have less citizens with poor health which would make them upset. I think that death rates and happiness have a strong negative correlation. I predict that as death rates increase, the happiness score will decrease. I think this is true because there are many reasons that death rates would be high in a country and would make people unhappy (war, epidemic, poor healthcare system, etc.).

In the HEALTHEXP.csv file, there are 3 variables; Country, Current Health Expenditure, and Year. I will work with Country and Current Health Expenditure. In the DEATHRATE.csv file, there are 4 variables; Rank, Country, (deaths/1,000 population), and Date of Information. I will work with Country and (deaths/1,000 population).

```
deathrate <- read_tsv("DEATHRATE.csv")
```

```
## Parsed with column specification:
## cols(
##   Rank = col_double(),
##   Country = col_character(),
##   `(deaths/1,000 population)` = col_double(),
##   `Date of Information` = col_character()
## )
```

```
deathrate %>% head()
```

R…	Country	(deaths/1,000 population)	Date of Information
<dbl>	<chr>	<dbl>	<chr>
1	South Sudan	19.3	2018 est.
2	Lesotho	15.1	2018 est.
3	Lithuania	14.8	2018 est.
4	Bulgaria	14.5	2018 est.
5	Latvia	14.5	2018 est.
6	Ukraine	14.3	2018 est.

6 rows

```
healthexp <- read_tsv("HEALTHEXP.csv")
```

```
## Parsed with column specification:
## cols(
##   Country = col_character(),
##   `Current Health Expenditure` = col_character(),
##   Year = col_double()
## )
```

```
healthexp %>% head()
```

Country	Current Health Expenditure	Year
<chr>	<chr>	<dbl>
Afghanistan	10.3%	2015
Albania	6.8%	2015
Algeria	7.1%	2015
Andorra	12%	2015
Angola	2.9%	2015
Antigua and Barbuda	4.8%	2015

6 rows

```
#In order to be able to use the variable `Current Health Expenditure` as a continou
s one I will parse it into double.
healthexp$`Current Health Expenditure` <- parse_number(healthexp$`Current Health Ex
penditure`)
```

```
summary(healthexp)
```

```
##    Country          Current Health Expenditure      Year
##  Length:190         Min.   :  2.000            Min.   :2015
##  Class :character   1st Qu.:  4.800            1st Qu.:2015
##  Mode  :character   Median :  6.300            Median :2015
##                     Mean   :  8.913            Mean   :2015
##                     3rd Qu.:  8.200            3rd Qu.:2015
##                     Max.   :403.000            Max.   :2015
##                     NA's   :1
```

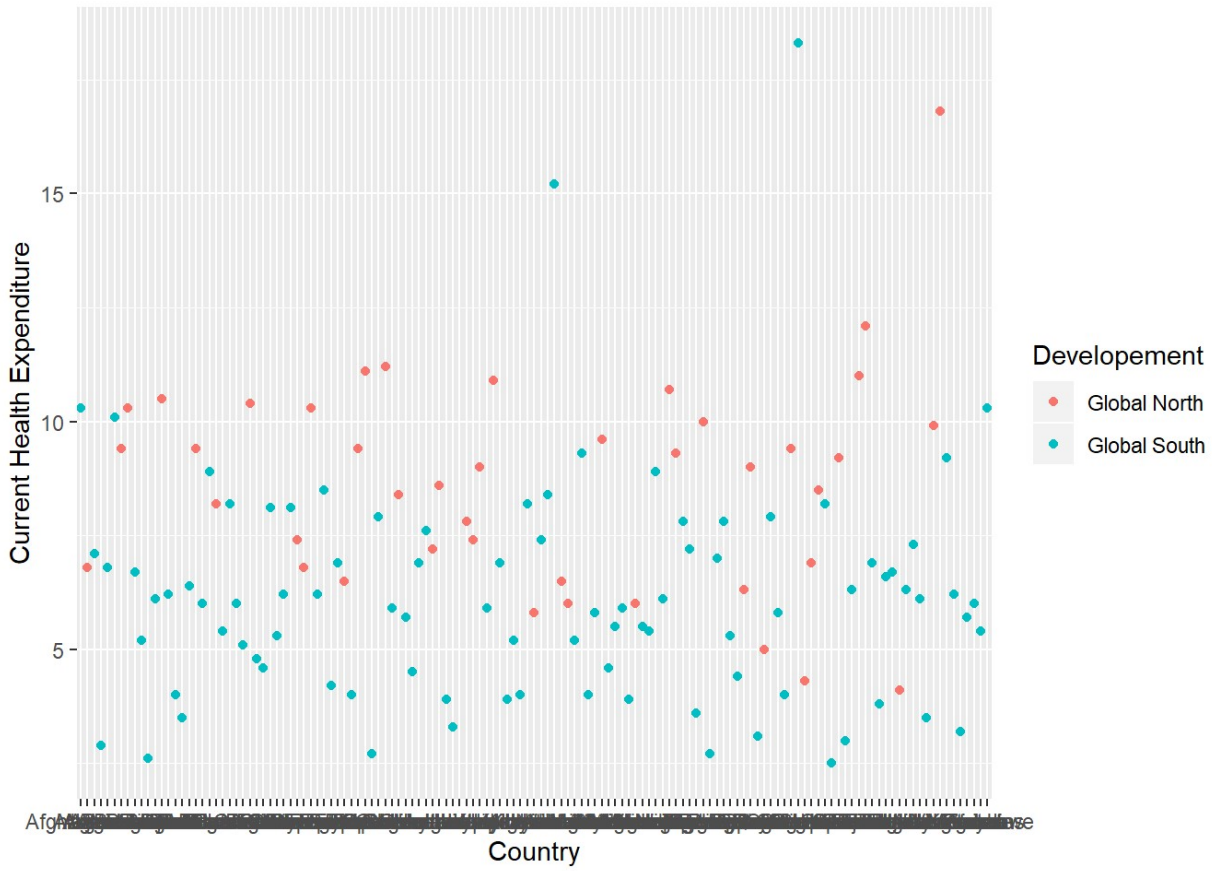
```
summary(deathrate)
```

```
##       Rank          Country          (deaths/1,000 population)
##  Min.   :  1.00   Length:226         Min.   : 1.600
##  1st Qu.: 57.25   Class :character   1st Qu.: 5.900
##  Median :113.50   Mode  :character   Median : 7.400
##  Mean   :113.50                      Mean   : 7.650
##  3rd Qu.:169.75                      3rd Qu.: 9.275
##  Max.   :226.00                      Max.   :19.300
##  Date of Information
##  Length:226
##  Class :character
##  Mode  :character
##
##
##
```

```
ggplot(data=inner_join(healthexp,regionclassification)) +
  geom_point(mapping=aes(x = Country, y = `Current Health Expenditure`, color = `De
velopement`))
```

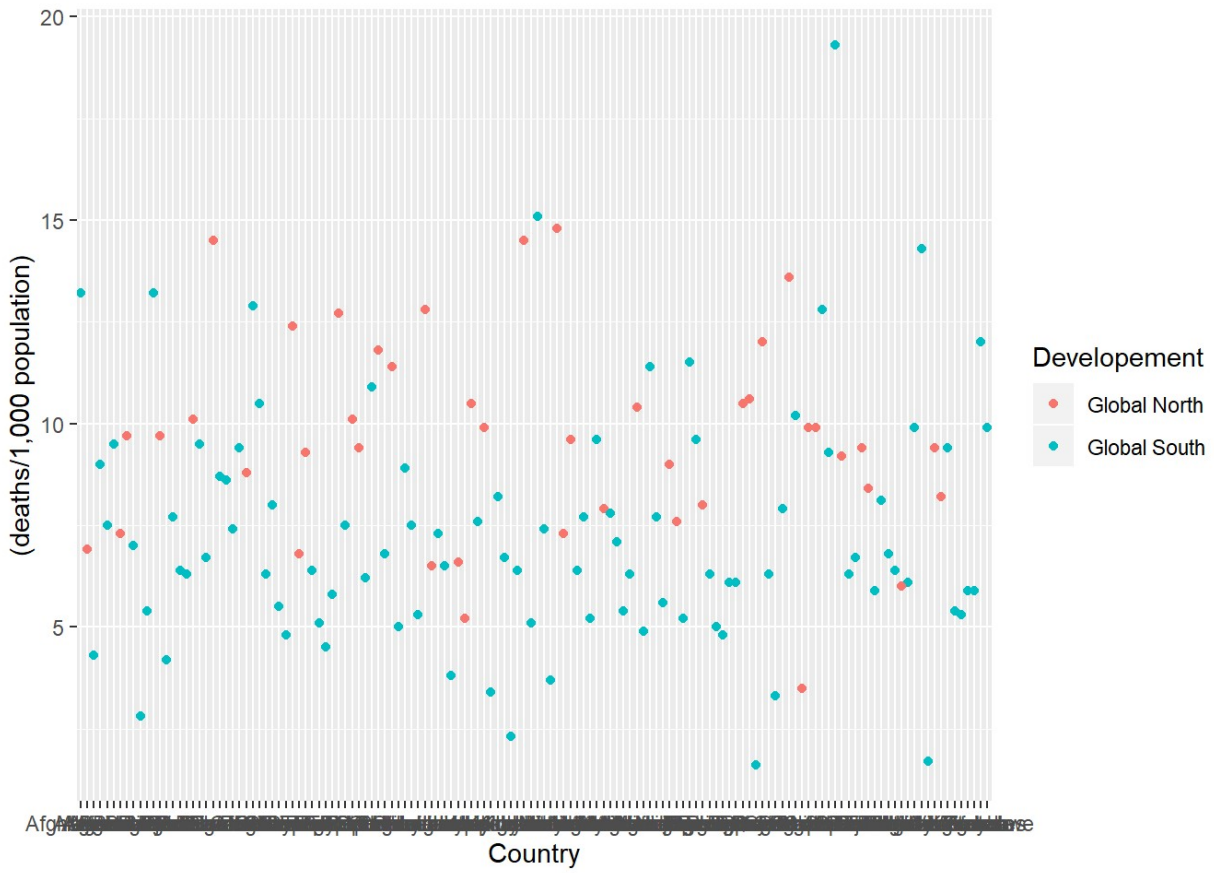
```
## Joining, by = "Country"
```

```
## Warning: Removed 1 rows containing missing values (geom_point).
```




```
ggplot(data=inner_join(deathrate,regionclassification)) +
  geom_point(mapping=aes(x = Country, y = `(deaths/1,000 population)`, color = `Dev
elopement`))
```

```
## Joining, by = "Country"
```

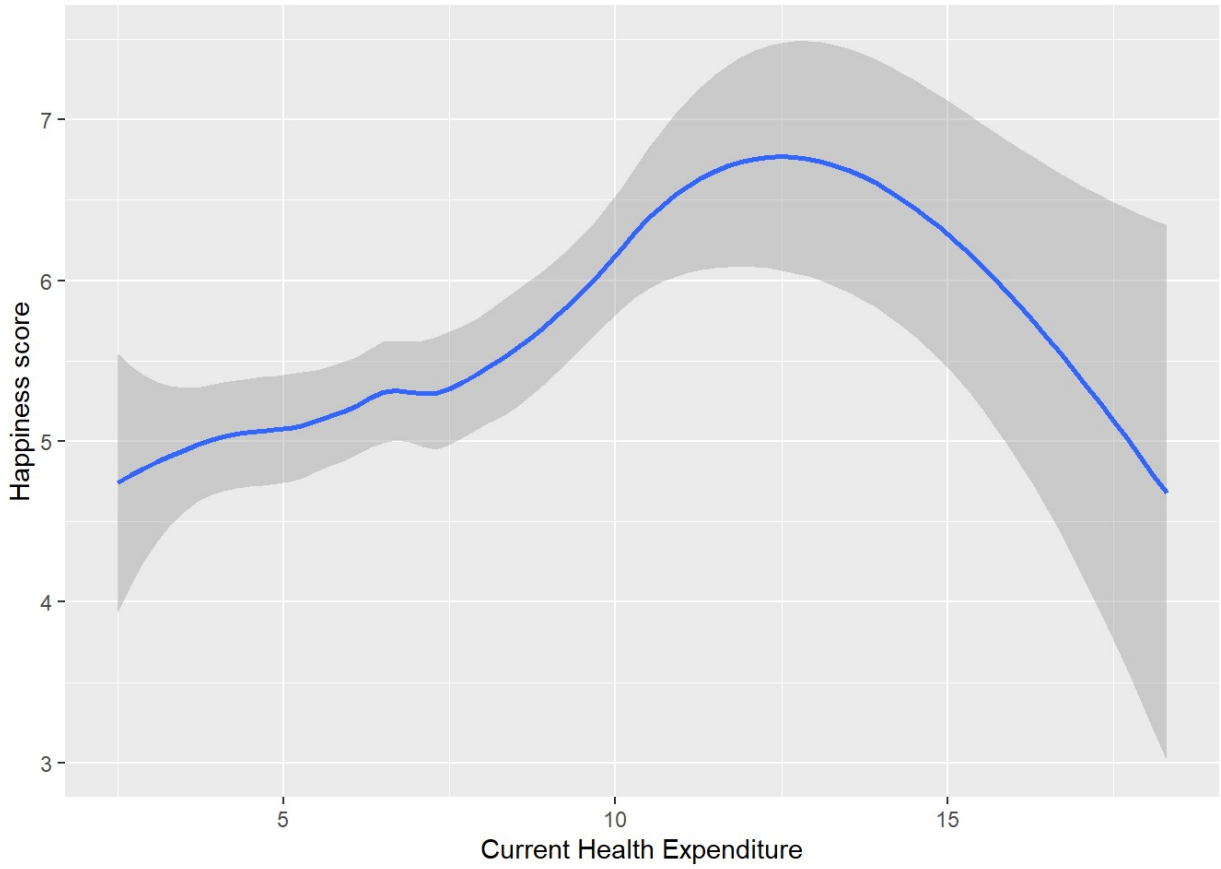


The diagram reveals that the data varies much across the countries, but there exists some tendency: health expenditure is higher in global north, but the death rate is also higher. The reason for that may be deseases connected to developement or sitting style of living.

```
ggplot(data=inner_join(healthexp,happy2018,by="Country")) +
  geom_smooth(mapping=aes(x = `Current Health Expenditure`, y = `Happiness score`))
```

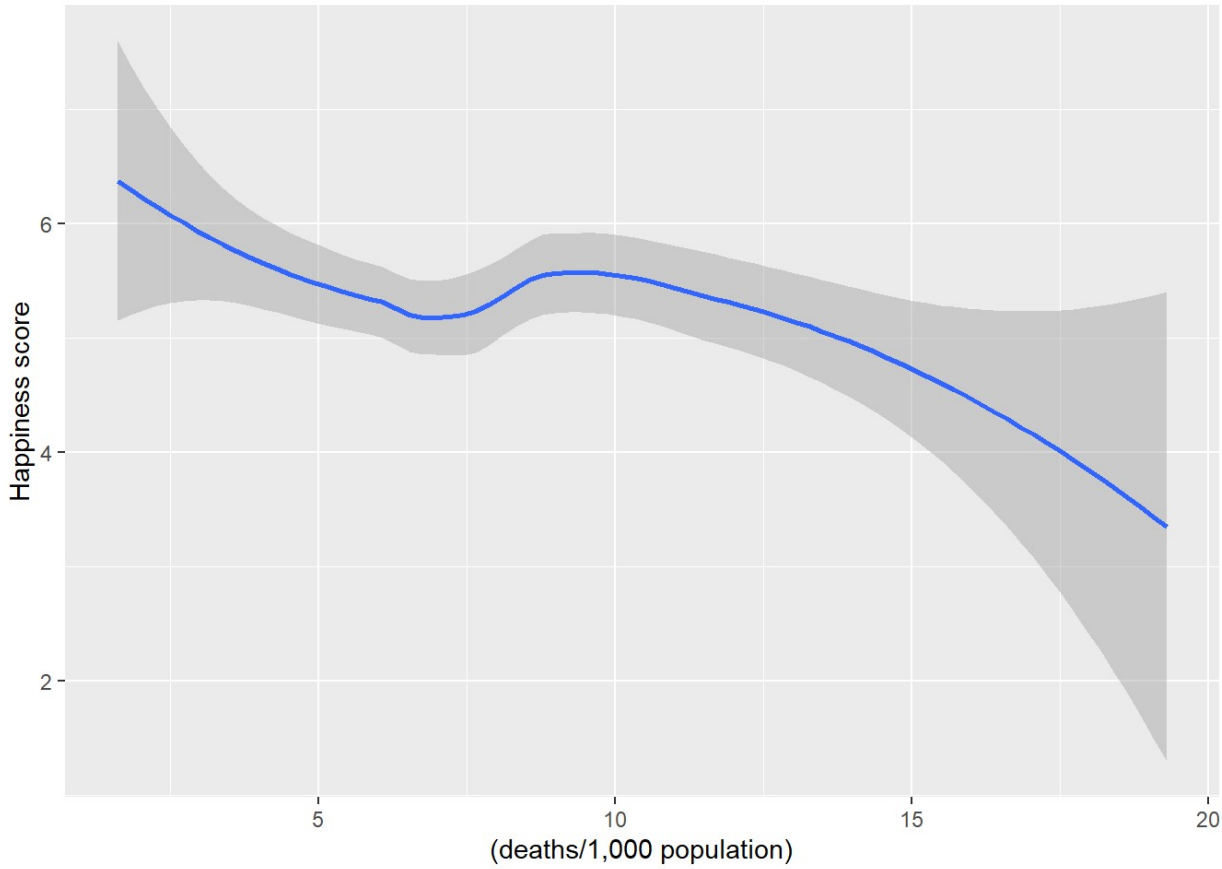
```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

```
## Warning: Removed 1 rows containing non-finite values (stat_smooth).
```




```
ggplot(data=inner_join(deathrate,happy2018,by="Country")) +
  geom_smooth(mapping=aes(x = `(deaths/1,000 population)`, y = `Happiness score`))
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



You can see from the trendlines in these graphs that my original predictions were pretty accurate. Although the relations are not perfectly clear, it is visible that generally, countries with higher happiness scores have higher health expenditures and lower death rates in their country. In conclusion, the level of healthcare in a country is related to the happiness of the country. The only weird observation is that within higher health expenditures happiness stops to rise and starts to fall. Probably in those countried death rate and spread of diseases is so high that the huge amount of money is spent not to help people live healthier but to stop this appearance.

III Economics

Rproject by Ronak Arora for group, Data set imported will be GINI, GDPCOMPOSITION, GDPPP.

```
urbanization <- read_tsv("URBANIZATION.csv")
```

```
## Parsed with column specification:
## cols(
##   Country = col_character(),
##   `Urban v.s. total` = col_character(),
##   Year = col_double(),
##   `Rate of urbanization` = col_character(),
##   `Year range` = col_character(),
##   Notes = col_character()
## )
```

```
urbanization%>%head()
```

Country <chr>	Urban v.s. total <chr>	Y…	Rate of urbanization <chr>	Year range <chr>	N <c
Afghanistan	25.8%	2019	3.37%	2015-20 est.	*N*
Albania	61.2%	2019	1.69%	2015-20 est.	*N*
Algeria	73.2%	2019	2.46%	2015-20 est.	*N*
American Samoa	87.1%	2019	0.07%	2015-20 est.	*N*
Andorra	88%	2019	-0.31%	2015-20 est.	*N*
Angola	66.2%	2019	4.32%	2015-20 est.	*N*

6 rows

```
gdppp <- read_tsv("GDPPP.csv")
```

```
## Parsed with column specification:
## cols(
##   Rank = col_double(),
##   Country = col_character(),
##   `GDP - PER CAPITA (PPP)` = col_character(),
##   `Date of Information` = col_character()
## )
```

```
gdppp%>%head()
```

R…	Country <chr>	GDP - PER CAPITA (PPP) <chr>	Date of Information <chr>
1	Liechtenstein	$139,100	2009 est.
2	Qatar	$124,500	2017 est.
3	Monaco	$115,700	2015 est.
4	Macau	$111,600	2017 est.
5	Luxembourg	$106,300	2017 est.
6	Bermuda	$99,400	2016 est.

6 rows

```
gdpcomp <- read_tsv("GDPCOMPOSITION.csv")
```

```
## Parsed with column specification:
## cols(
##   Country = col_character(),
##   algriculture = col_character(),
##   industry = col_character(),
##   services = col_character(),
##   year = col_character(),
##   notes = col_character()
## )
```

```
gdpcomp %>% head()
```

Country <chr>	algriculture <chr>	industry <chr>	services <chr>	year <chr>	notes <chr>
Afghanistan	23%	21.1%	55.9%	2016 est.	data exclude opium pro
Albania	21.7%	24.2%	54.1%	2017 est.	*NA*
Algeria	13.3%	39.3%	47.4%	2017 est.	*NA*
American Samoa	27.4%	12.4%	60.2%	2012	*NA*
Andorra	11.9%	33.6%	54.5%	2015 est.	*NA*
Angola	10.2%	61.4%	28.4%	2011 est.	*NA*

6 rows

```
gini <- read_tsv("GINI.csv")
```

```
## Parsed with column specification:
## cols(
##   Rank = col_double(),
##   Country = col_character(),
##   `Distribution of family income - Gini index` = col_double(),
##   `Date of Information` = col_character()
## )
```

```
gini %>% head()
```

R…	Country <dbl><chr>	Distribution of family incor
1	Lesotho	
2	South Africa	
3	Micronesia, Federated States of	

R…	Country	Distribution of family income
<dbl>	<chr>	
4	Haiti	
5	Botswana	
6	Namibia	

6 rows | 1-3 of 4 columns

```
gdpcomp<-gdpcomp%>%mutate(services=parse_number(`services`))
gdpcomp<-gdpcomp%>%mutate(industry=parse_number(`industry`))
gdpcomp<-gdpcomp%>%mutate(algriculture=parse_number(`algriculture`))
gdppp<-gdppp%>%mutate(`GDP - PER CAPITA (PPP)`=parse_number(`GDP - PER CAPITA (PPP)
`))
urbanization<-urbanization%>%mutate(`Urban v.s. total`=parse_number(`Urban v.s. tot
al`))
urbanization<-urbanization%>%mutate(`Rate of urbanization`=parse_number(`Rate of ur
banization`))
```

Now we analyse each file to see what it contains and what can be used as a key.

```
summary(gdppp)
```

```
##       Rank       Country          GDP - PER CAPITA (PPP)
##  Min.   :  1   Length:229         Min.   :   700
##  1st Qu.: 58   Class :character   1st Qu.:  5400
##  Median :115   Mode  :character   Median : 14900
##  Mean   :115                      Mean   : 23405
##  3rd Qu.:172                      3rd Qu.: 34400
##  Max.   :229                      Max.   :139100
##  Date of Information
##  Length:229
##  Class :character
##  Mode  :character
##
##
##
```

```
summary(gdpcomp)
```

```
##    Country           algriculture      industry        services
##  Length:231         Min.   : 0.00   Min.   : 0.00   Min.   : 28.40
##  Class :character   1st Qu.: 2.30   1st Qu.:17.43   1st Qu.: 51.77
##  Mode  :character   Median : 6.75   Median :24.65   Median : 61.20
##                     Mean   :11.23   Mean   :26.03   Mean   : 62.54
##                     3rd Qu.:16.85   3rd Qu.:33.00   3rd Qu.: 73.40
##                     Max.   :60.70   Max.   :61.40   Max.   :100.00
##                     NA's   :5       NA's   :5       NA's   :3
##      year              notes
##  Length:231         Length:231
##  Class :character   Class :character
##  Mode  :character   Mode  :character
##
##
##
##
```

```
summary(gini)
```

```
##       Rank          Country
##  Min.   :  1.00   Length:158
##  1st Qu.: 40.25   Class :character
##  Median : 79.50   Mode  :character
##  Mean   : 79.50
##  3rd Qu.:118.75
##  Max.   :158.00
##  Distribution of family income - Gini index Date of Information
##  Min.   : 0.30                              Length:158
##  1st Qu.:32.10                              Class :character
##  Median :37.90                              Mode  :character
##  Mean   :38.67
##  3rd Qu.:45.23
##  Max.   :63.20
```

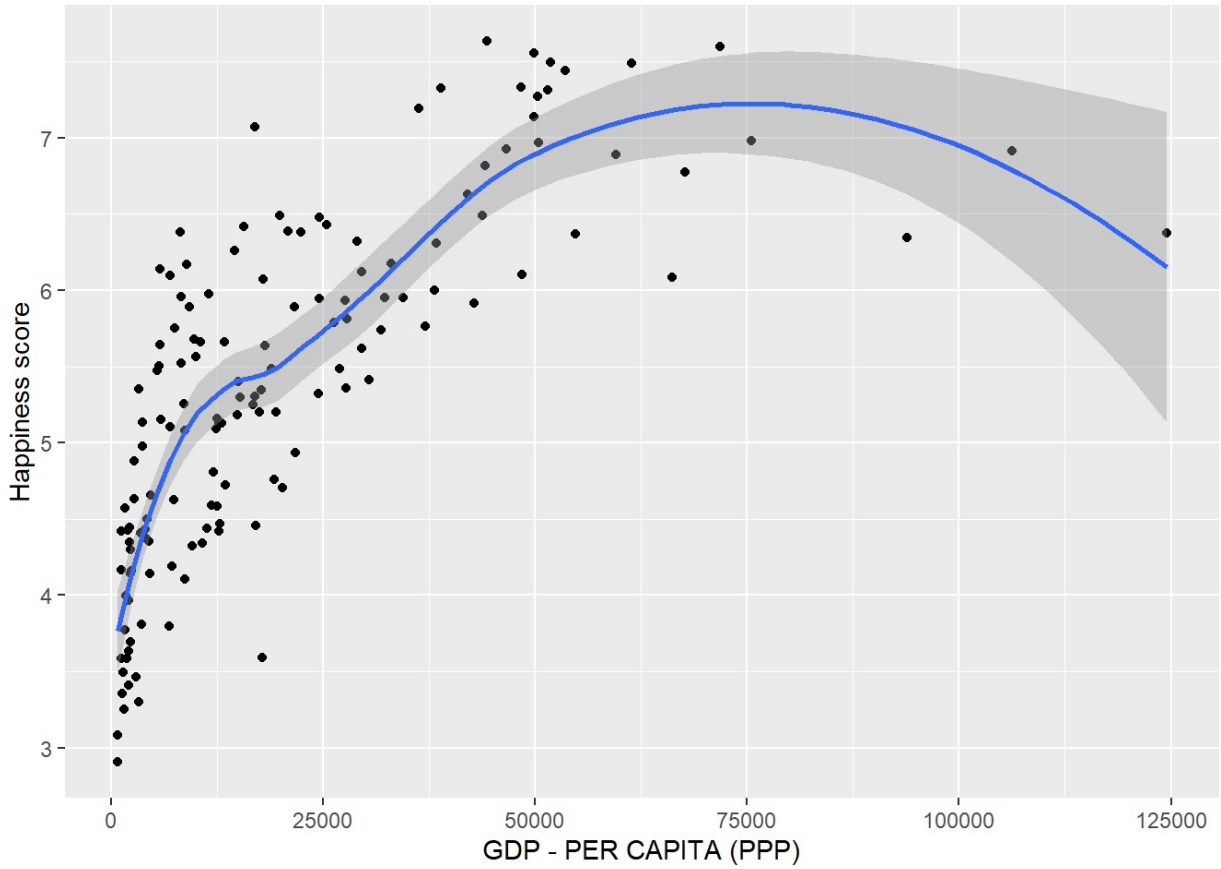
```
summary(urbanization)
```

```
##    Country          Urban v.s. total      Year      Rate of urbanization
##  Length:231         Min.   :  0.00   Min.   :2012   Min.   :-0.930
##  Class :character   1st Qu.: 42.00   1st Qu.:2019   1st Qu.: 0.620
##  Mode  :character   Median : 61.90   Median :2019   Median : 1.555
##                     Mean   : 60.63   Mean   :2019   Mean   : 1.755
##                     3rd Qu.: 80.65   3rd Qu.:2019   3rd Qu.: 2.810
##                     Max.   :100.00   Max.   :2019   Max.   : 5.700
##                                                     NA's   :1
##   Year range           Notes
##  Length:231         Length:231
##  Class :character   Class :character
##  Mode  :character   Mode  :character
##
##
##
##
```

Here we join the tables by the country and then we plot out the point graph to see the how GDP per capita impacts happiness and see if the monthly income has any relation with the happiness score.

```
gdppp<-inner_join(gdppp,happy2018,by="Country")
ggplot(gdppp,aes(`GDP - PER CAPITA (PPP)`,`Happiness score`))+geom_point()+geom_smo
oth(aes(x=`GDP - PER CAPITA (PPP)`,y=`Happiness score`))
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

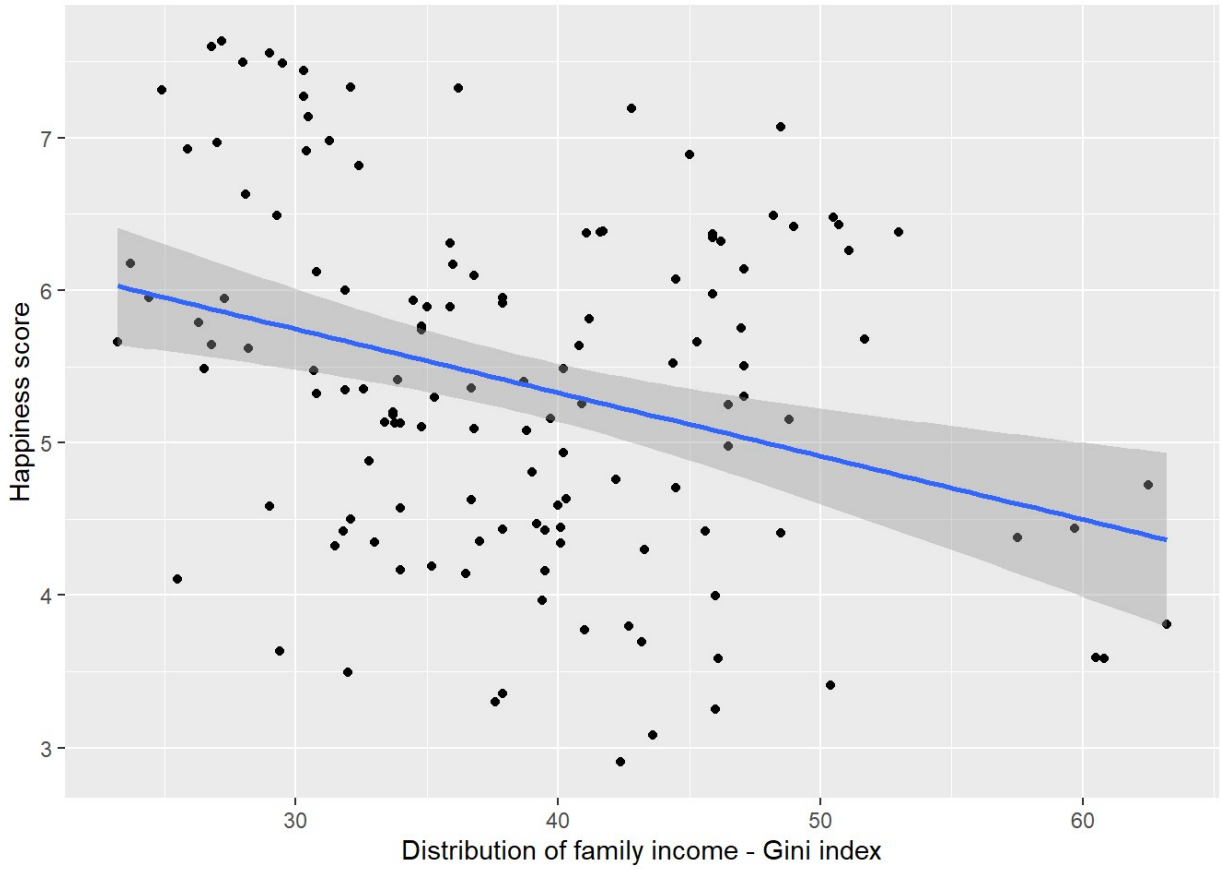


```
cor(gdppp$`GDP - PER CAPITA (PPP)`,gdppp$`Happiness score`)
```

```
## [1] 0.7162236
```

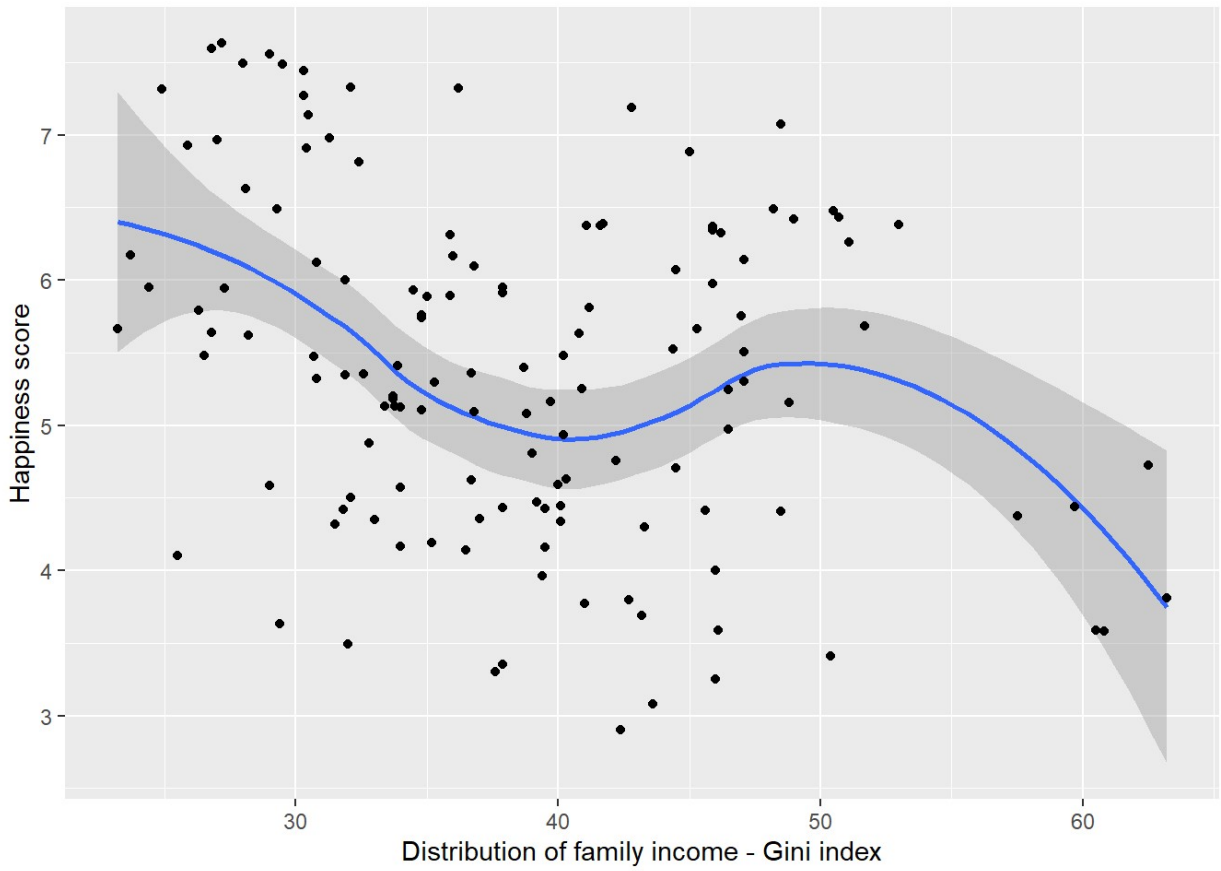
MONEY MATTERS: Per capita gross domestic product (GDP) is a metric that breaks down a country's GDP per person.It is calculated by dividing GDP over a country's population (STRONG POSITIVE CORRELATION HERE). After looking at this data closely we come to conclusions that the money is strongly correlated to happiness and as the per capita income increases the happiness increases that might be because as the average income of people increases, their quality of life improves and their standard of living do too, hence people are able to enjoy their life more and access more amenties.

```
gini<-inner_join(gini,happy2018,by="Country")
ggplot(data=gini)+geom_point(mapping=aes(x=`Distribution of family income - Gini in
dex`,y=`Happiness score`))+
geom_smooth(mapping=aes(x=`Distribution of family income - Gini index`,y=`Happiness
score`),method="lm")
```




```
ggplot(data=gini)+geom_smooth(mapping=aes(x=`Distribution of family income - Gini i
ndex`,y=`Happiness score`))+geom_point(mapping=aes(x=`Distribution of family income
- Gini index`,y=`Happiness score`))
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

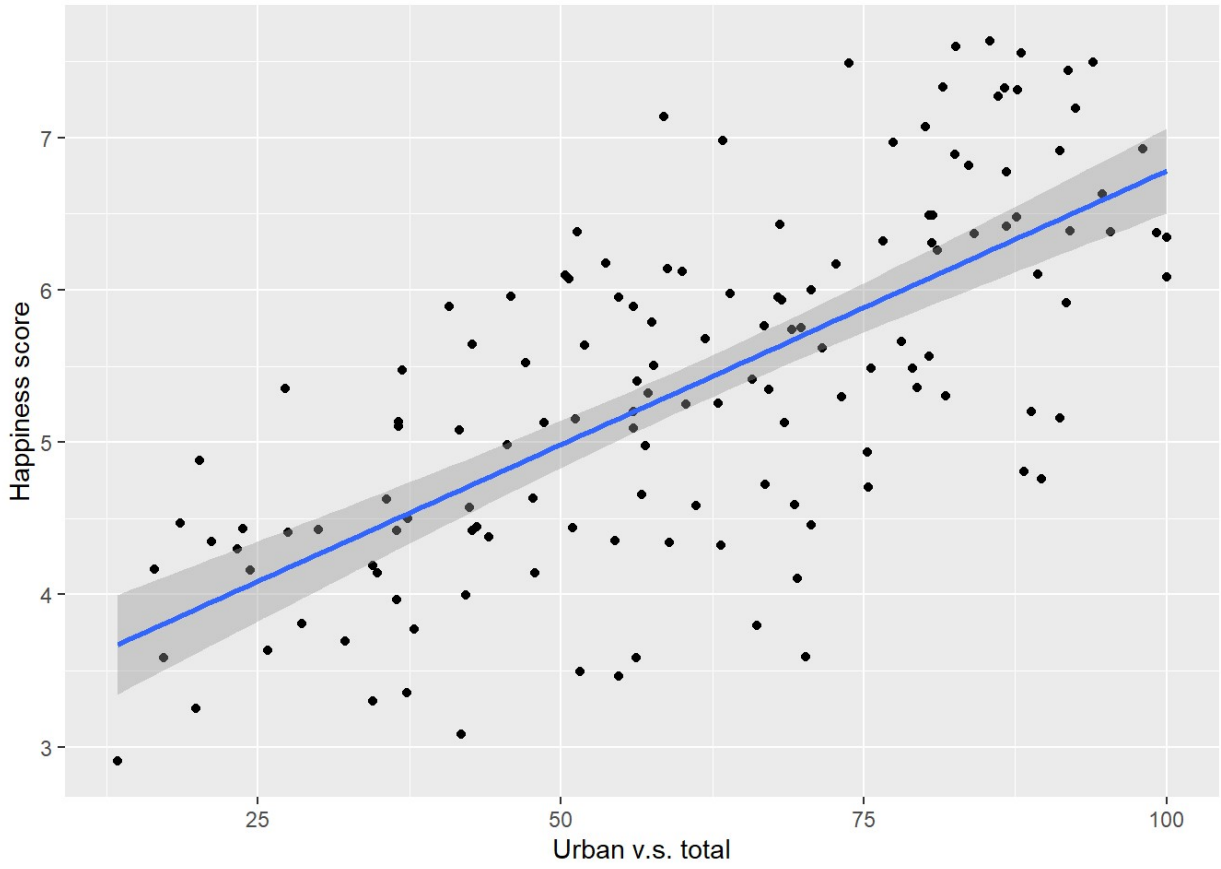



```
cor(gini$`Distribution of family income - Gini index`,gini$`Happiness score`)
```

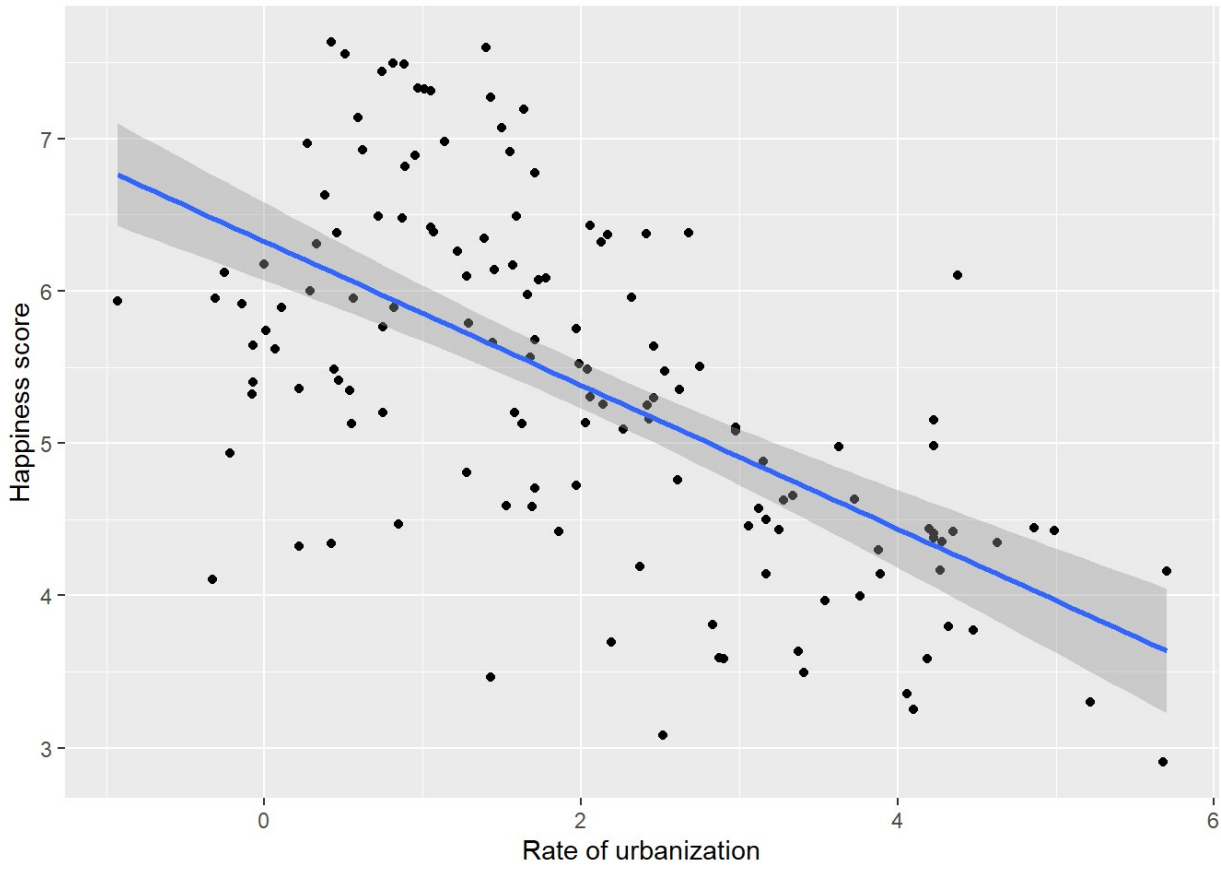
```
## [1] -0.3113098
```

Here we find a relation between the GINI index and Happiness score. GINI INDEX= Difference in household income of rich and poor people, (BECAUSE THE CORRELATION IS negative and it's closer to zero so we can say that the factors are inversly related and are not also not strongly correlated to each other). BUT CORRELATION STILL EXISTS. Let's filter out parts and look more closely at other factors.

```
urbanization<-inner_join(urbanization,happy2018,by="Country")
ggplot(urbanization,aes(`Urban v.s. total`,`Happiness score`))+geom_point()+geom_sm
ooth(method="lm")
```




```
ggplot(urbanization,aes(`Rate of urbanization`,`Happiness score`))+geom_point()+geo
m_smooth(method="lm")
```



```
cor(urbanization$`Urban v.s. total`,urbanization$`Happiness score`)
```

```
## [1] 0.6961388
```

```
cor(urbanization$`Rate of urbanization`,urbanization$`Happiness score`)
```

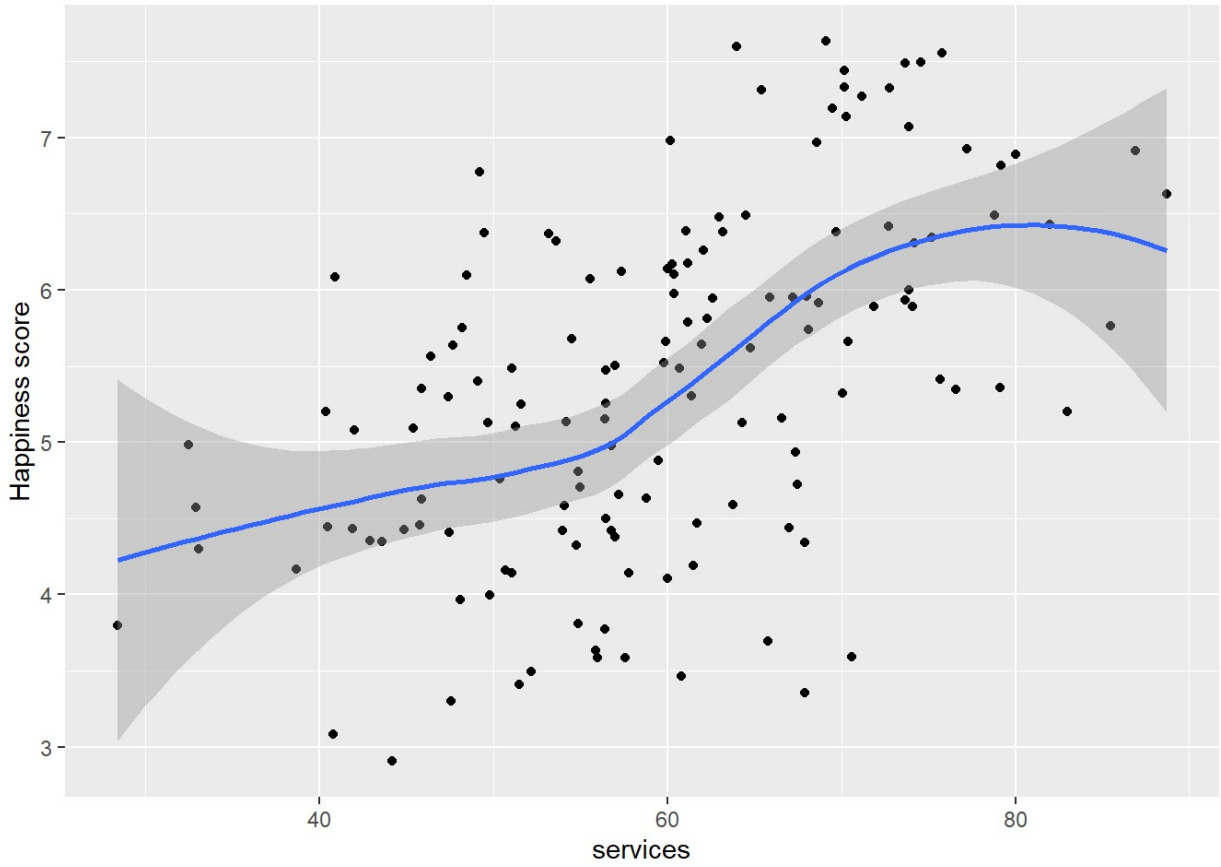
```
## [1] -0.6115063
```

In the plot 1 we can clearly notice a general increase in the trend of the happiness of the people and a strong correlation (almost 0.7) which is represting a strong correlation and showing us that clearly that people prefer to live in more urbanized cities and metropolitian areas and due to the amenties offered due to their greater development gives the living area a better night life of itself and increases the recreational activites and that could lead to greater happiness among people.

In the plot 2 we can clearly notice a general downward trend in relation between rate of urbanization and the and the happines and the correlation shows us this exact value which is (almost -0.7) this shows that people are not happy with the increasing rate of construccction because of the fact that the greater the rate of urbansation the greater is the area under construction greater is the polution and environmental depletion and this deteriorates the quality of life and also the happiness.

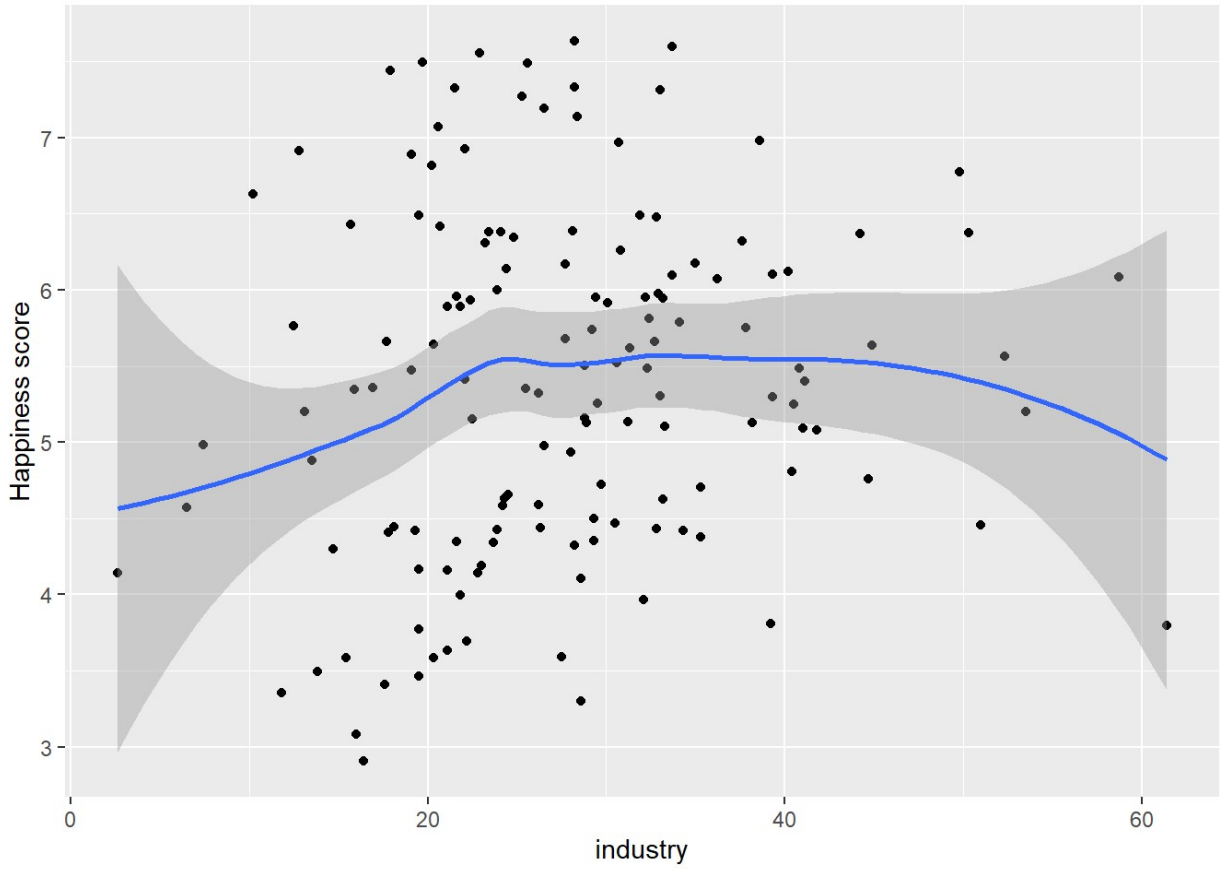
```
gdpcomp<-inner_join(gdpcomp,happy2018,by="Country")
ggplot(data=gdpcomp)+geom_point(mapping=aes(x=services,y=`Happiness score`))+geom_s
mooth(mapping=aes(x=services,y=`Happiness score`))
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



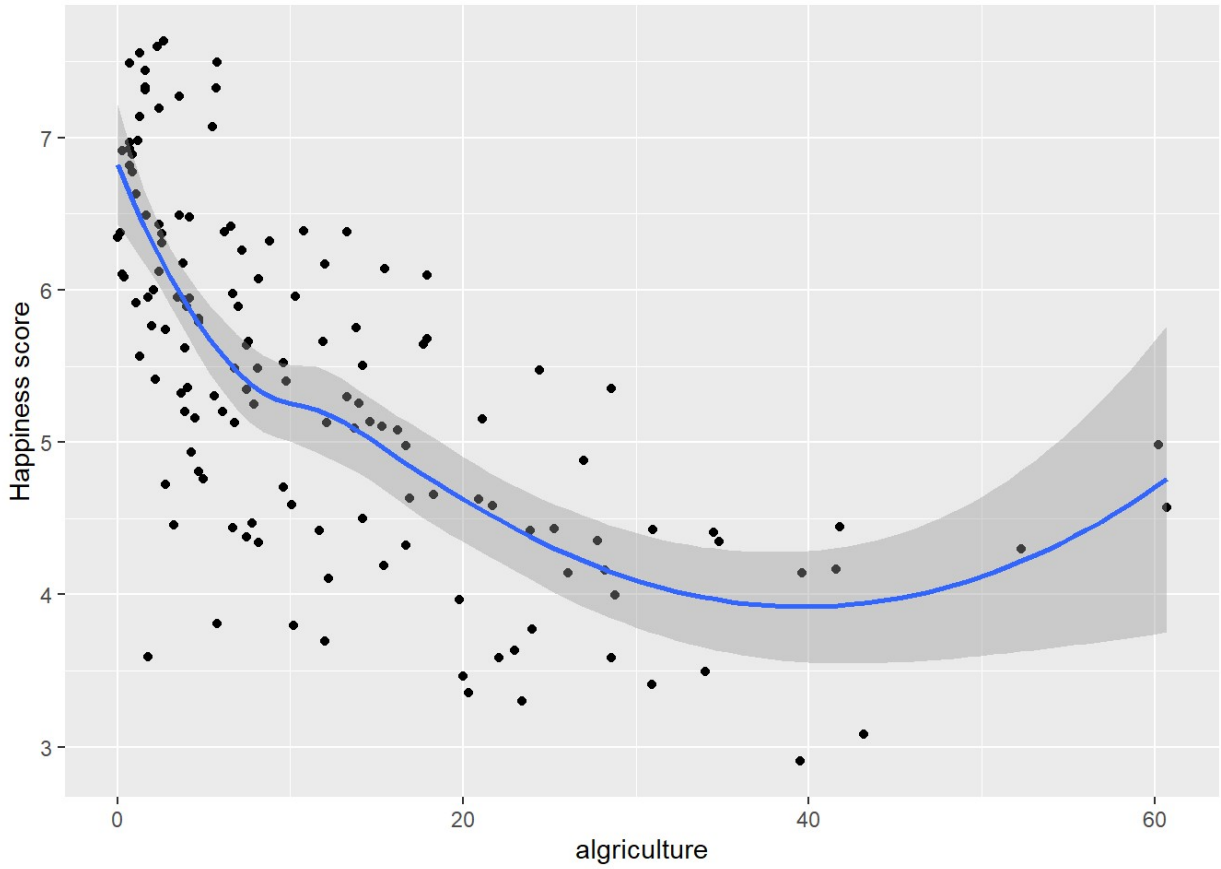

```
ggplot(data=gdpcomp)+geom_point(mapping=aes(x=industry,y=`Happiness score`))+geom_s
mooth(mapping=aes(x=industry,y=`Happiness score`))
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```




```
ggplot(data=gdpcomp)+geom_point(mapping=aes(x=algriculture,y=`Happiness score`))+ge
om_smooth(mapping=aes(x=algriculture,y=`Happiness score`))
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```




```
cor(gdpcomp$services,gdpcomp$`Happiness score`)
```

```
## [1] 0.5322104
```

```
cor(gdpcomp$industry,gdpcomp$`Happiness score`)
```

```
## [1] 0.1283523
```

```
cor(gdpcomp$algriculture,gdpcomp$`Happiness score`)
```

```
## [1] -0.6287781
```

Here in the plot 1 we can find a clear postive growth however when we look at it's corelation with the factors we observe that the relation is quite weak and this could be rergarding the fact that teritiary sector is a new and an upcoming service sector in the industry but people aren't extremly bothered by the fact on how much the governmnet invests in their industry.

Here in plot 2 we can not find a general trend which is intresting, but it's hard to make any insights. Probably the factor of the people actually taking part in these industry is as a miniscule factor in their day to day lives. However when we try to find the correlation of the data it's positive but it's very close to zero so the correlation is very small.

Here in plot 3 we can see a general trend of decrease in values and it forms as the values are inversly proportional to happiness showing us that data. We presume that this could be because of the fact that the countries which intensly invest in agriclture do not take part in other industries or are not that developed and the fact that people in this country would be aware with under development and obviously not be happy with it (ALSO a negative correaltion gives us the same insight).

IV Corelations between different factors

```
unemployment <- read_tsv("UNEMP.csv")
```

```
## Parsed with column specification:
## cols(
##   Rank = col_double(),
##   Country = col_character(),
##   `(%)` = col_double(),
##   `Date of Information` = col_character()
## )
```

```
unemployment %>% head()
```

Rank <dbl>	Country <chr>	(%) <dbl>	Date of Information <chr>
1	Cocos (Keeling) Islands	0.1	2011
2	Cambodia	0.3	2017 est.
3	Niger	0.3	2017 est.
4	Laos	0.7	2017 est.
5	Thailand	0.7	2017 est.
6	Belarus	0.8	2017 est.

6 rows

```
laborforce <- read_tsv("LABORFORCE.csv")
```

```
## Parsed with column specification:
## cols(
##   Rank = col_double(),
##   Country = col_character(),
##   `Labor force` = col_number(),
##   `Date of Information` = col_character()
## )
```

```
laborforce %>% head()
```

Rank <dbl>	Country <chr>	Labor force <dbl>	Date of Information <chr>
1	China	806700000	2017 est.
2	India	521900000	2017 est.
3	United States	160400000	2017 est.
4	Indonesia	125000000	2016 est.
5	Brazil	104200000	2017
6	Russia	76530000	2017 est.

6 rows

```
schooling <- read_tsv("SCHOOLINGEXPECTANCY.csv")
```

```
## Parsed with column specification:
## cols(
##   Country = col_character(),
##   total = col_double(),
##   male = col_double(),
##   female = col_double(),
##   year = col_double(),
##   notes = col_character()
## )
```

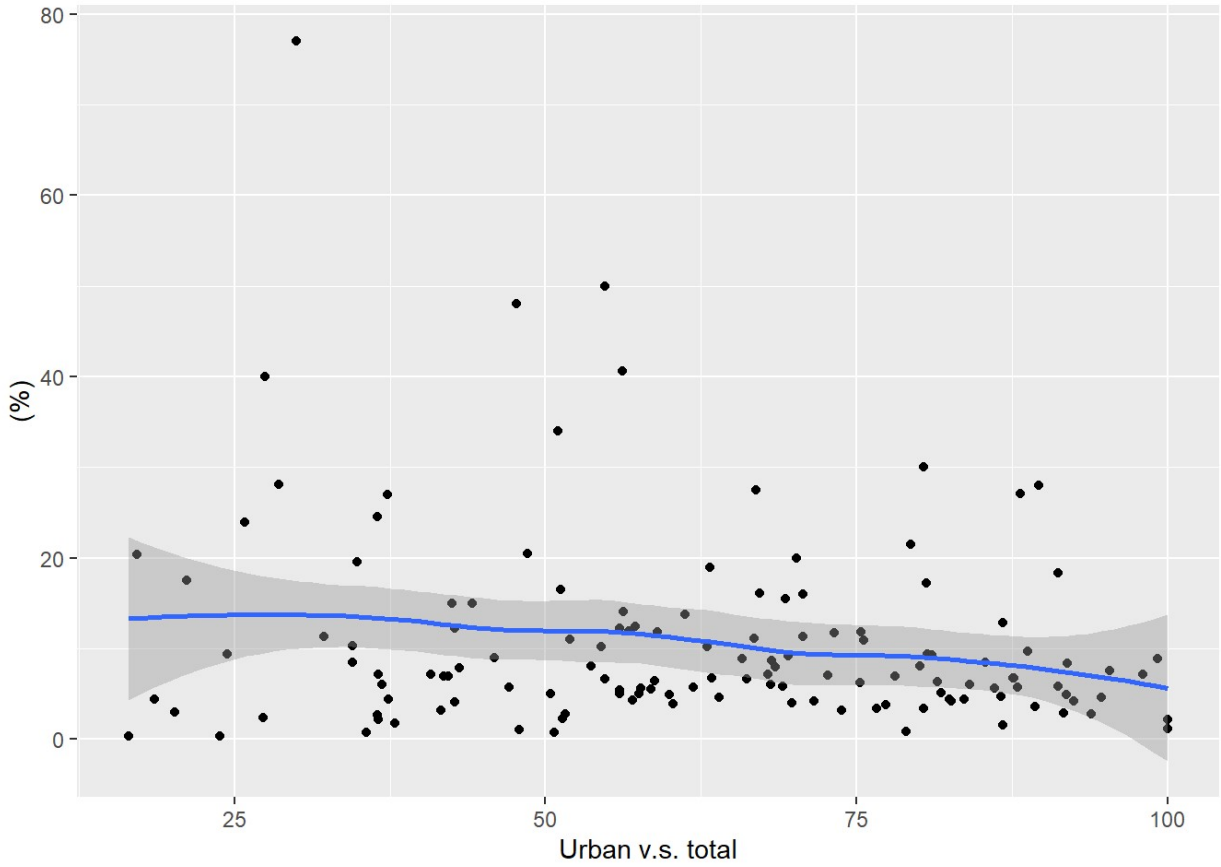
```
schooling%>%head()
```

Country <chr>	total <dbl>	male <dbl>	female <dbl>	year <dbl>	notes <chr>
Afghanistan	10	13	8	2014	*NA*
Albania	15	15	16	2017	*NA*
Algeria	14	14	15	2011	*NA*
Angola	10	13	8	2011	*NA*
Antigua and Barbuda	13	12	13	2012	*NA*
Argentina	18	16	19	2016	*NA*

6 rows

```
x<-inner_join(unemployment,urbanization,by="Country")
y<-arrange(x,`Urban v.s. total`)
ggplot(x)+geom_point(aes(x=`Urban v.s. total`,y=`(%)`))+geom_smooth(aes(x=`Urban v.
s. total`,y=`(%)`))
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

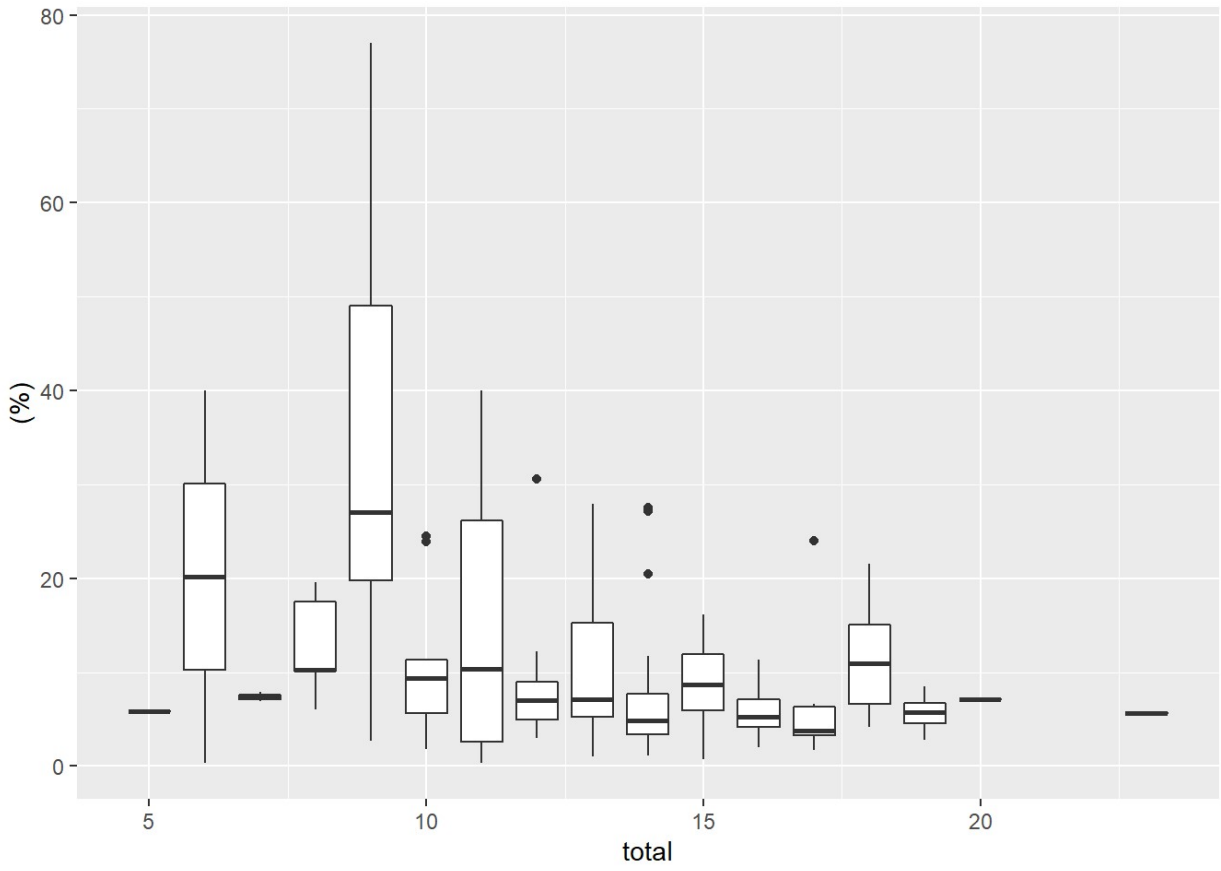



```
cor(y$`Urban v.s. total`,y$`(%)`)
```

```
## [1] -0.1940172
```

We can see that as the urbanization has negligible influence over unemployment according to the graph and corelation. This is leads us to believe probably other factors might influence the unemployment rate.

```
i<-inner_join(unemployment,schooling,by="Country")
j<-arrange(i,desc(`total`))
ggplot(j)+geom_boxplot(aes(group=`total`,x=`total`,y=`(%)`))
```



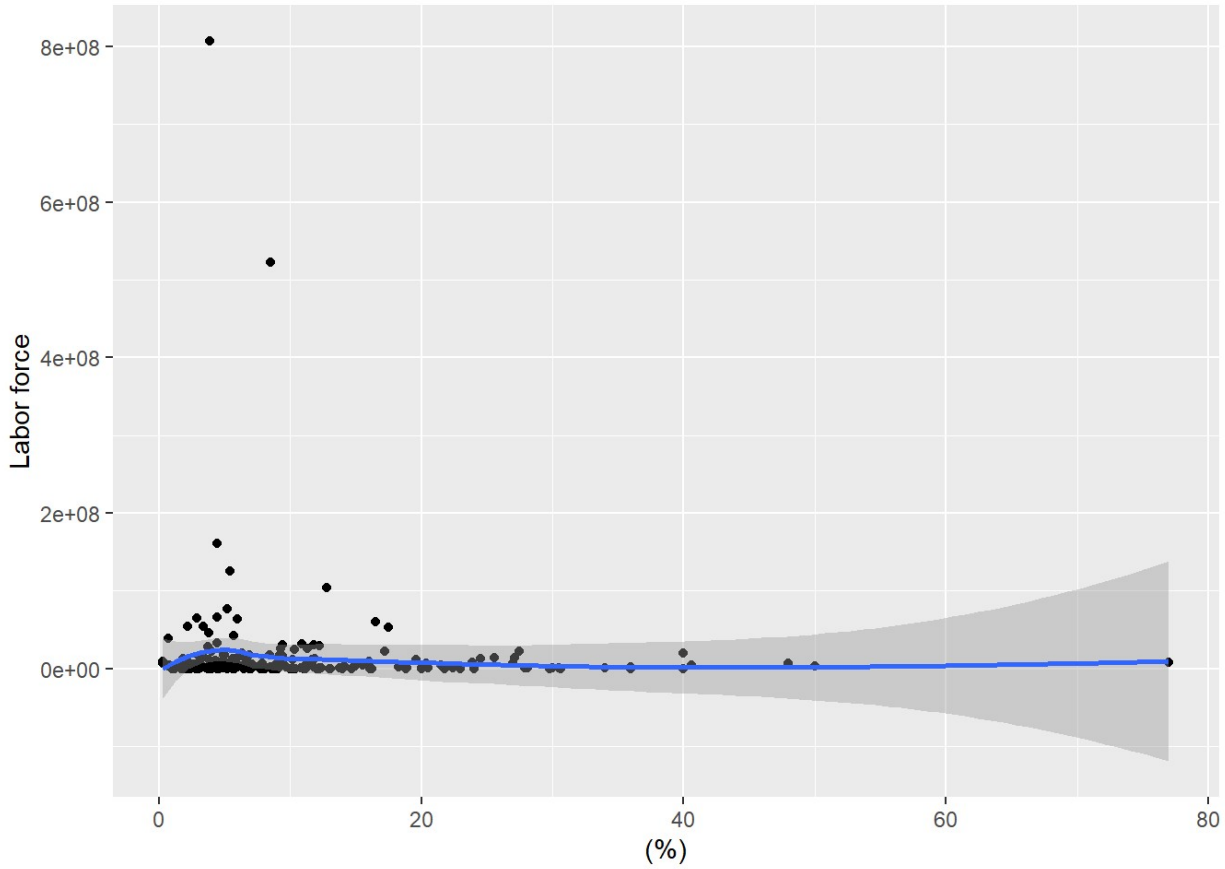

```
cor(j$`total`,j$`(%)`)
```

```
## [1] -0.3188785
```

We can observe that as total schooling life expectancy increases, there is a decrease in unemployment. This is due to the fact, that the population will be well educated and skilled which are the primary factors to getting a job.

```
q<-inner_join(unemployment,laborforce,by="Country")
w<-arrange(q,desc(`Labor force`))
ggplot(w)+geom_point(aes(x=`(%)`,y=`Labor force`))+geom_smooth(aes(x=`(%)`,y=`Labor force`))
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



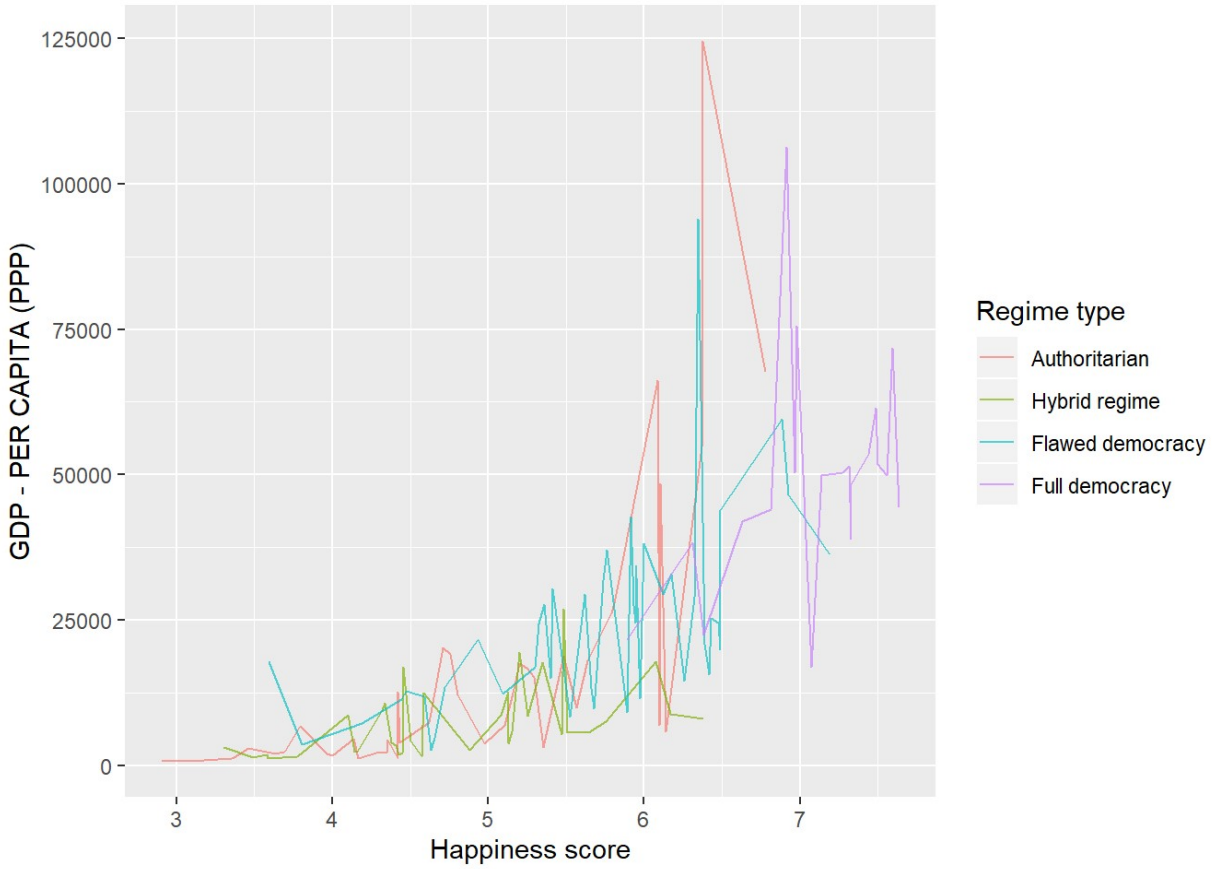

```
cor(w$`Labor force`,w$`(%)`)
```

```
## [1] -0.06818875
```

Unemployment rates seem to have no influence over Labor Force in a country according to the graph and corelation.

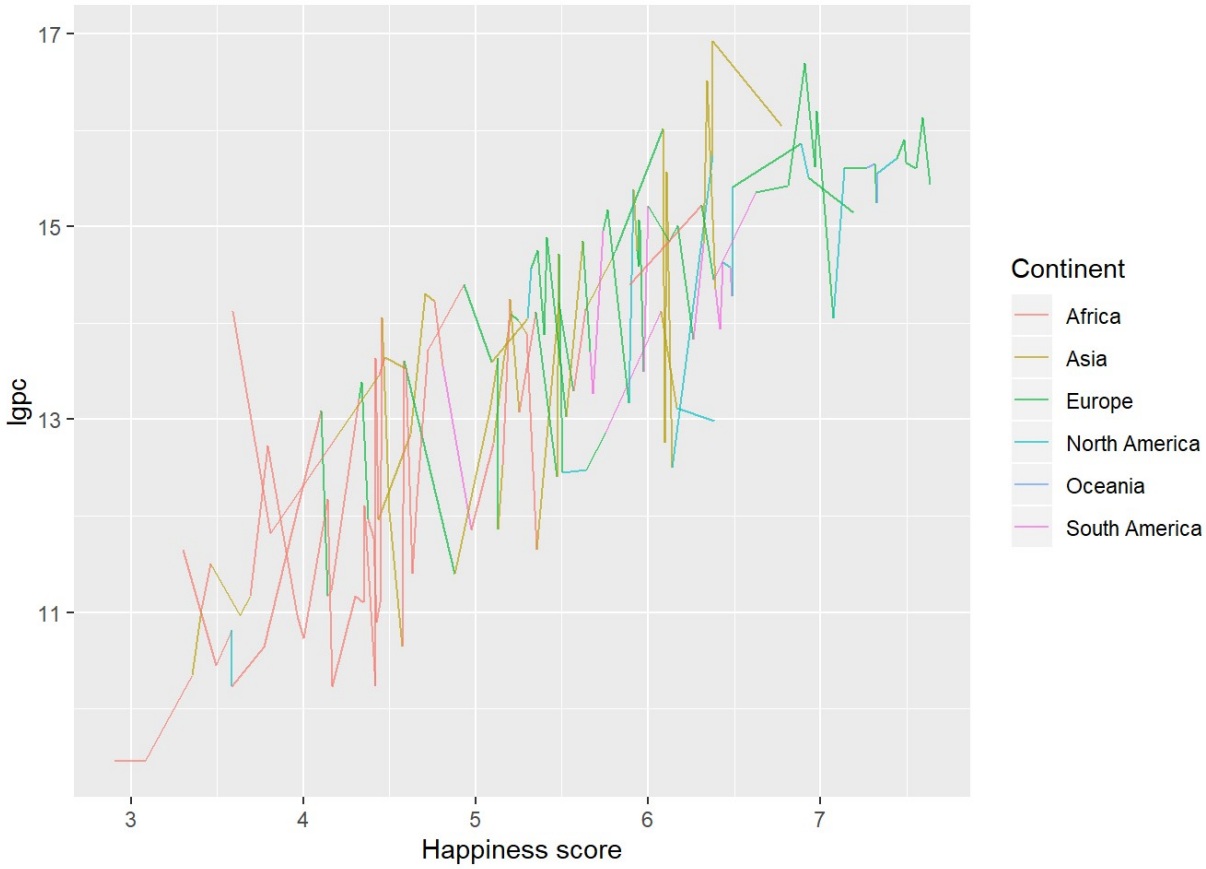
Final Conclusion: To conclude the findings it is safe to say that from the datasets analysed, only Schooling Life Expectancy has a direct influence on the unemployment rate of a country.

```
gdppp2<-inner_join(democracyindex,gdppp,by="Country")
gdppp2<-gdppp2%>%rename(`Happiness score`=`Happiness score.x`)
ggplot(gdppp2,aes(`Happiness score`, `GDP - PER CAPITA (PPP)`,group=`Regime type`,c
olor=`Regime type`
             )) +
    geom_line(alpha = 2/3)
```



Here we look at the GDP_PER_CAPITA of Different countries and compare them with the happiness score. We group our data by the Regime type we can clearly notice that as GDP PER CAPITA increases the happiness increases but along with this as we group by the regime type we come to notice that people with the highest GDP per capita in Authoritarian government are less happy than than the people with lower income(GDP per capita), but living in full democracy and this could mostly be because of the degree of freedom in these countries. Also people in hybrid regime seem to be pretty happy even though their wages are very low which probably means income does not matter to them.

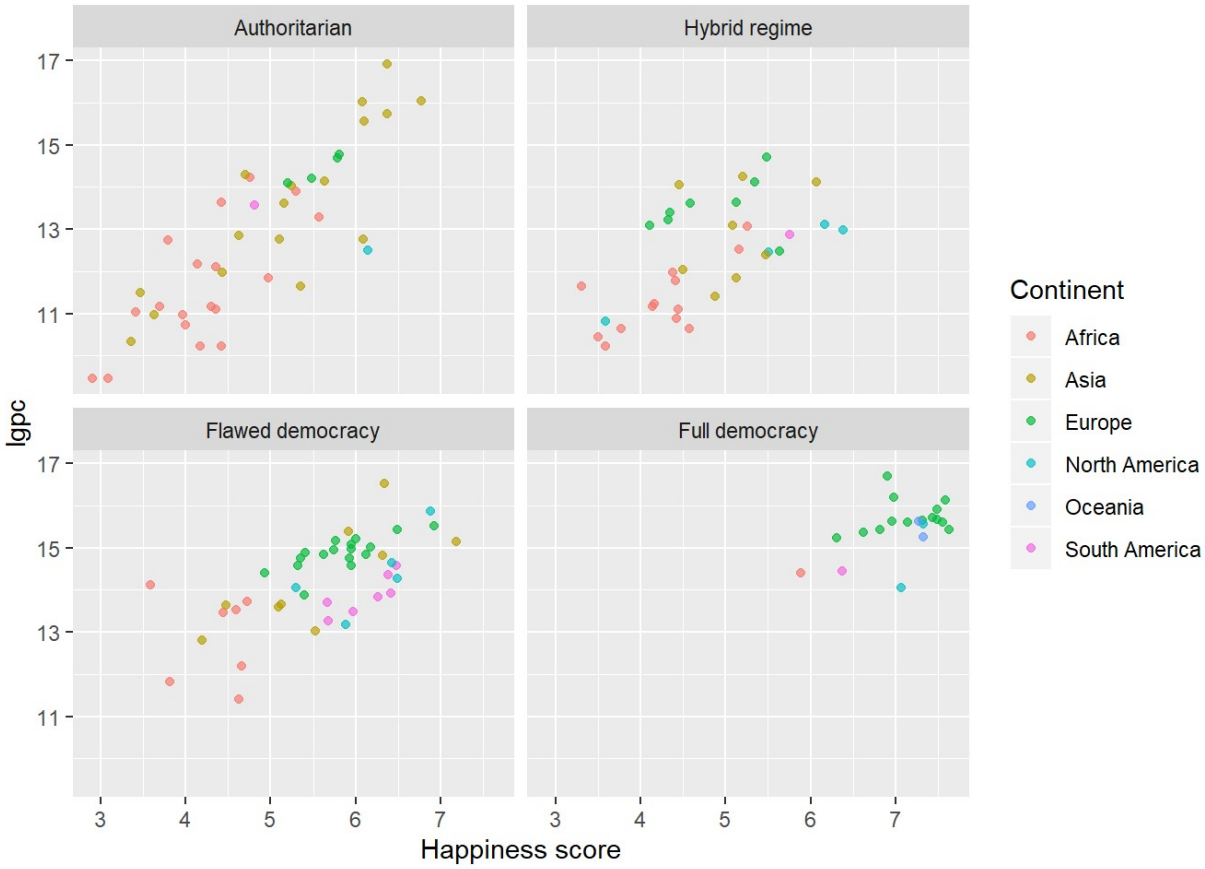
```
lgapminder <- gdppp2 %>%
  mutate(lgpc = log2(`GDP - PER CAPITA (PPP)`))
lgapminder %>%
  ggplot(aes(`Happiness score`, lgpc,group=`Regime type`,color=Continent)) +
    geom_line(alpha = 2/3)
```



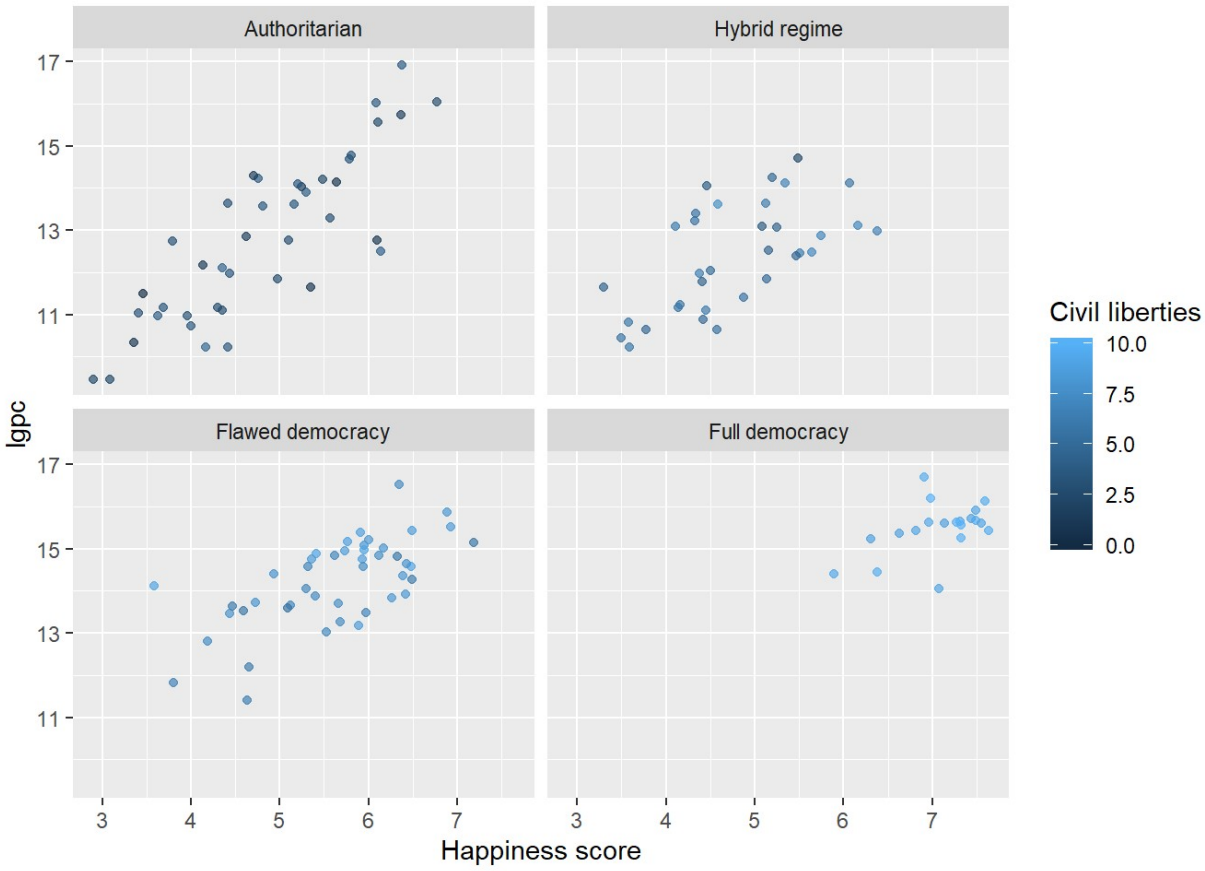

```
cor(lgapminder$lgpc,lgapminder$`Happiness score`)
```

```
## [1] 0.8245981
```

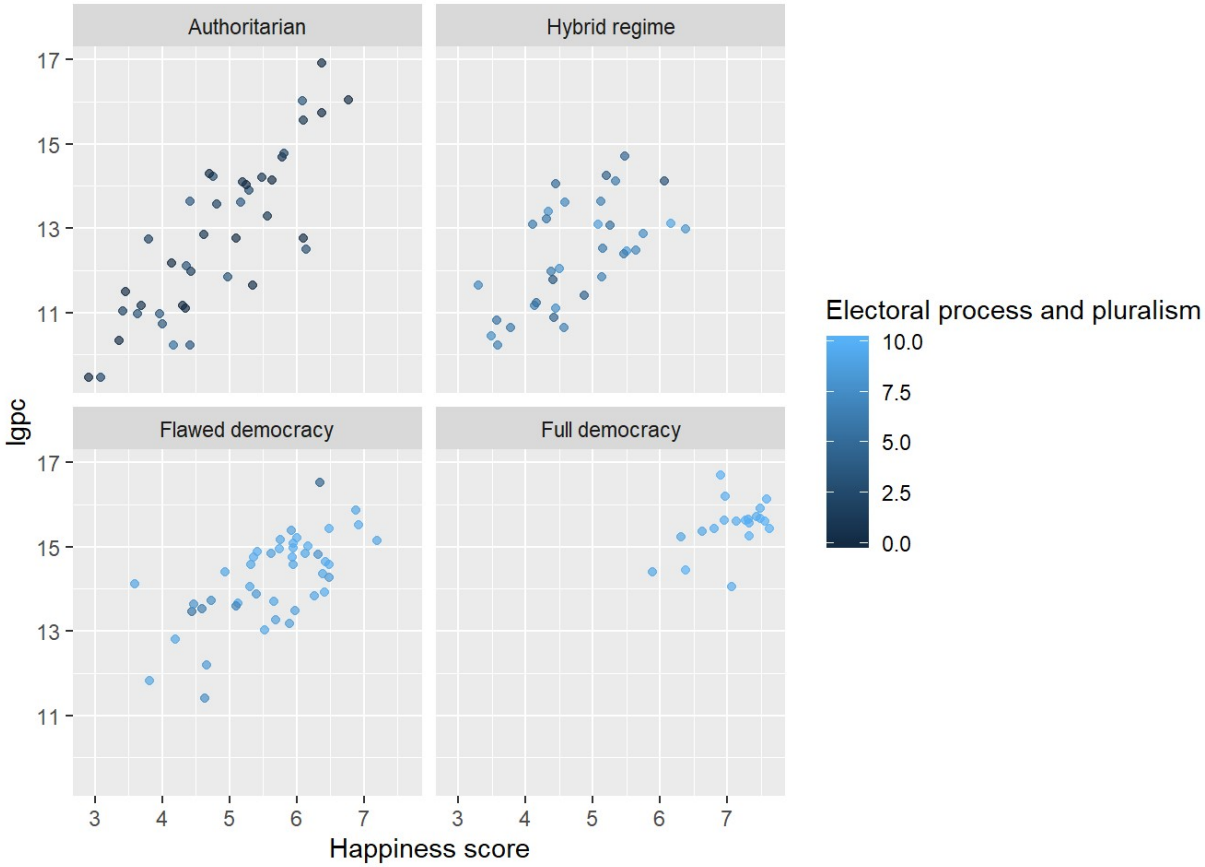
```
ggplot(lgapminder,aes(`Happiness score`, lgpc,color=Continent)) + geom_point(alpha
= 2/3)+facet_wrap(~`Regime type`)
```




```
ggplot(lgapminder,aes(`Happiness score`, lgpc,color=`Civil liberties`)) + geom_poin
t(alpha = 2/3)+facet_wrap(~`Regime type`)
```



```
ggplot(lgapminder,aes(`Happiness score`, lgpc,color=`Electoral process and pluralis
m`)) + geom_point(alpha = 2/3)+facet_wrap(~`Regime type`)
```

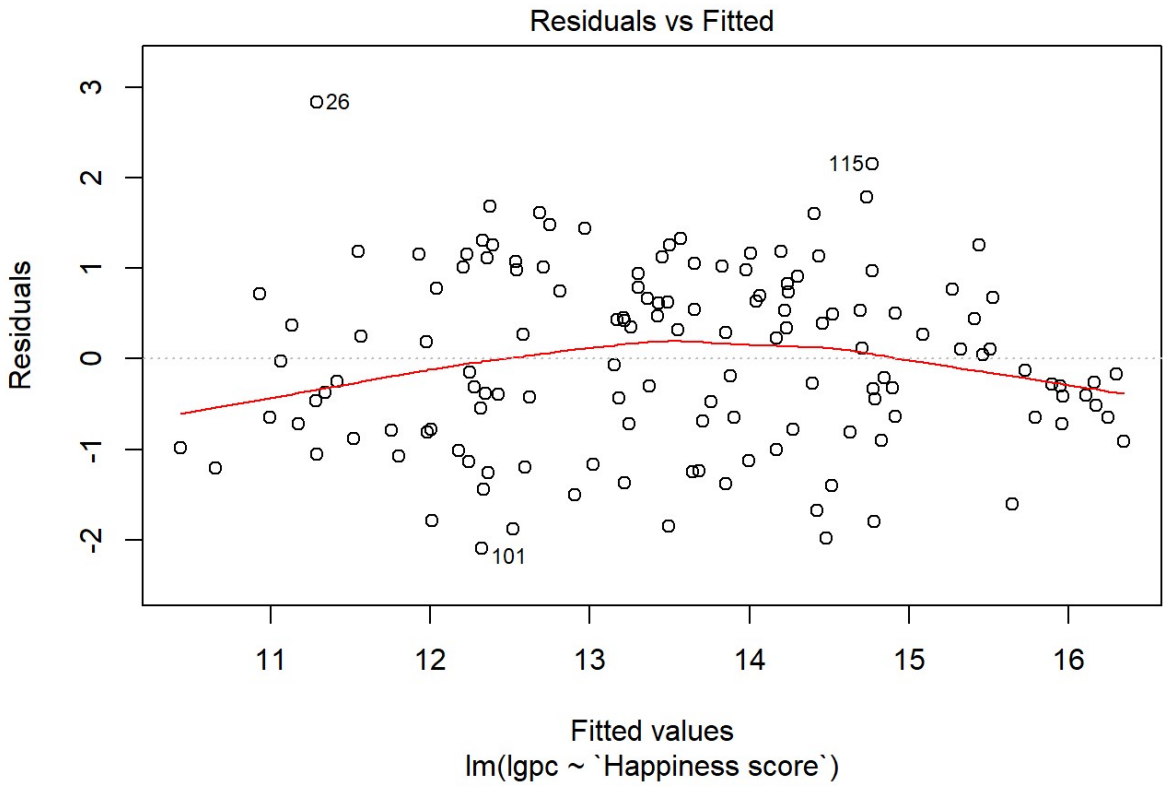


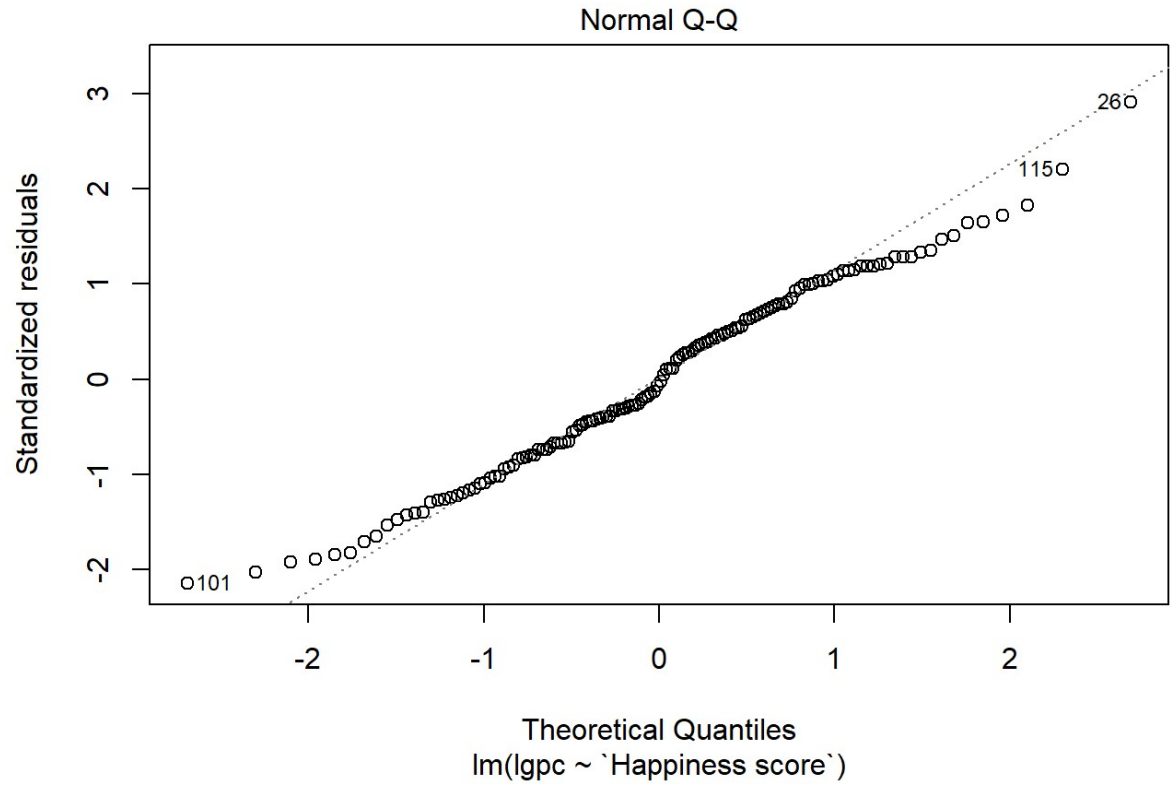
Here we finally notice a very strong correlation as we find the correlation between the log of the GDP PER CAPITA and the Happiness score if the data is gone through a facet wrap by regime type and seperated on the color on the basis of contienent then we come to some pretty intresting and significant plot which clearly indicates the fact that full democrecy make people the happiest and that this practice is most commonly followed in in europeon countries while majority of africa is still under authoritarian government these people will have lower liberty and less freedom hence their overall happinesss levels are lower But as we closely look at the authritarian graph we notice that that there are countires in asia where even though the people are under authoritarian rule they are content with their life these are generally countires like thailand and Japan where people enjoy greater freedom rights even if they are under no democrecy urbanization and other factors also contirbute in their happiness level, Considering how advanced japan is when compared to some african countries the reason seems to be more and more clearer. people in flawed democrecies are still mediocarly happy probably because of freedom as well as the fact that hybrid regime still makes less happy overall (below average also) this is one of our strongest correlations with a staggering 0.825 showing that LGDP is highly correlated to our data set.

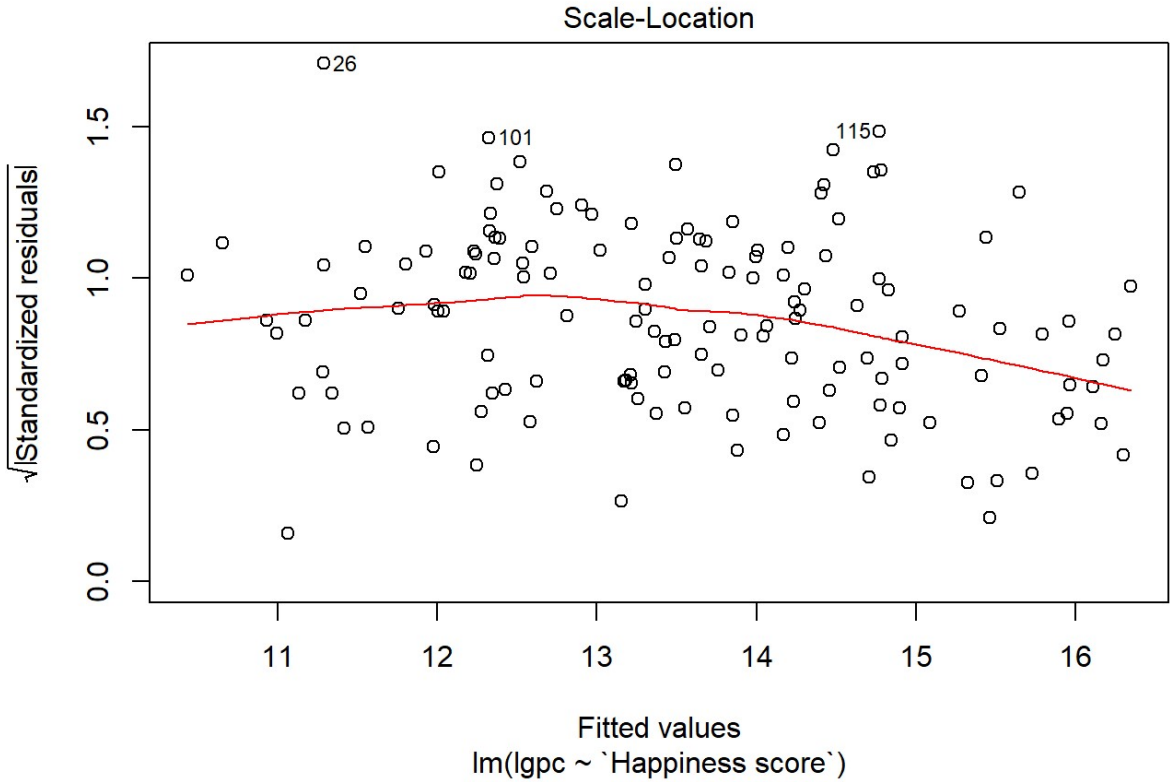
```
lm.fit=lm(lgpc~`Happiness score`,lgapminder)
summary(lm.fit)
```

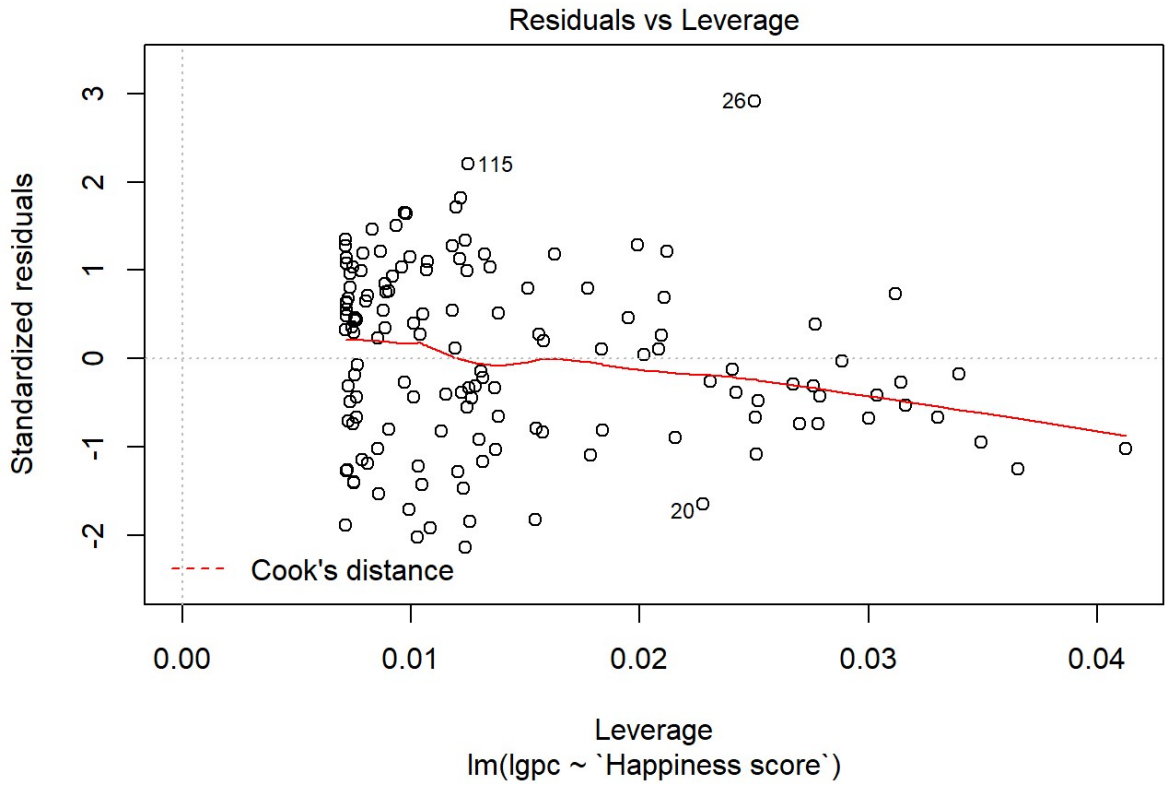
```
##
## Call:
## lm(formula = lgpc ~ `Happiness score`, data = lgapminder)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.09486 -0.71900 -0.04689  0.75729  2.83111
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)         6.7947     0.4027   16.87   <2e-16 ***
## `Happiness score`   1.2517     0.0731   17.12   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9838 on 138 degrees of freedom
## Multiple R-squared:   0.68,  Adjusted R-squared:  0.6776
## F-statistic: 293.2 on 1 and 138 DF,  p-value: < 2.2e-16
```

```
plot(lm.fit)
```

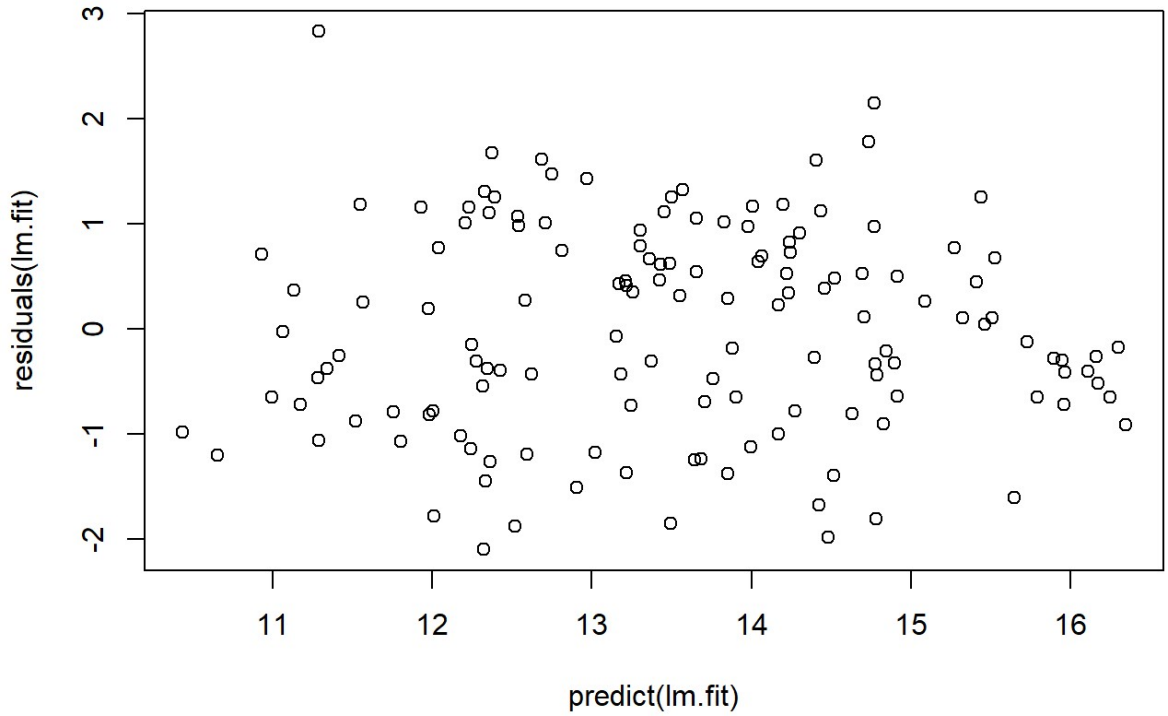



Normal Q-Q
Standardized residuals
Theoretical Quantiles
lm(lgpc ~ `Happiness score`)
26
115
101
-2
-1
0
1
2
3

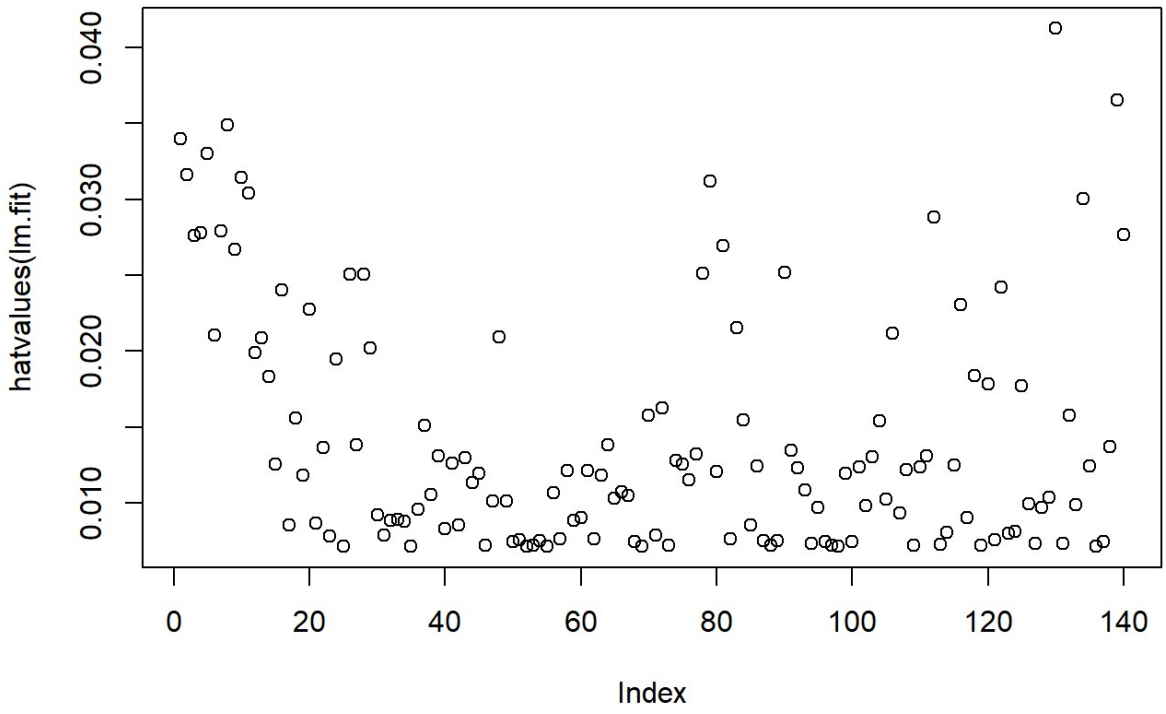

Scale-Location
√|Standardized residuals|
Fitted values
lm(lgpc ~ `Happiness score`)
26
101
115
0.0
0.5
1.0
1.5
11
12
13
14
15
16




```
plot(predict(lm.fit),residuals(lm.fit))
```



```
plot(hatvalues(lm.fit))    #in order to find values of levarage statsistics and whi
ch max to find the highest levarage value
```



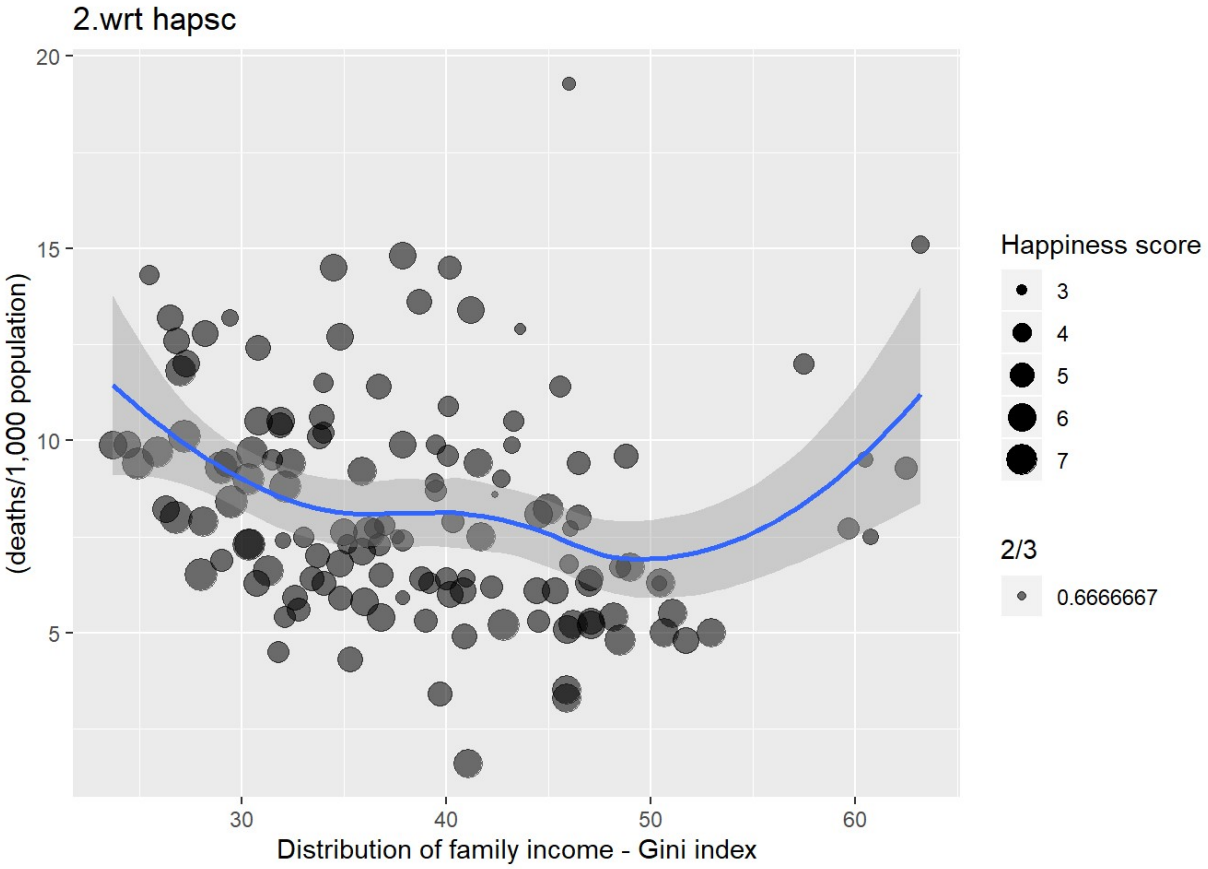

```
which.max(hatvalues(lm.fit))
```

```
## 130
## 130
```

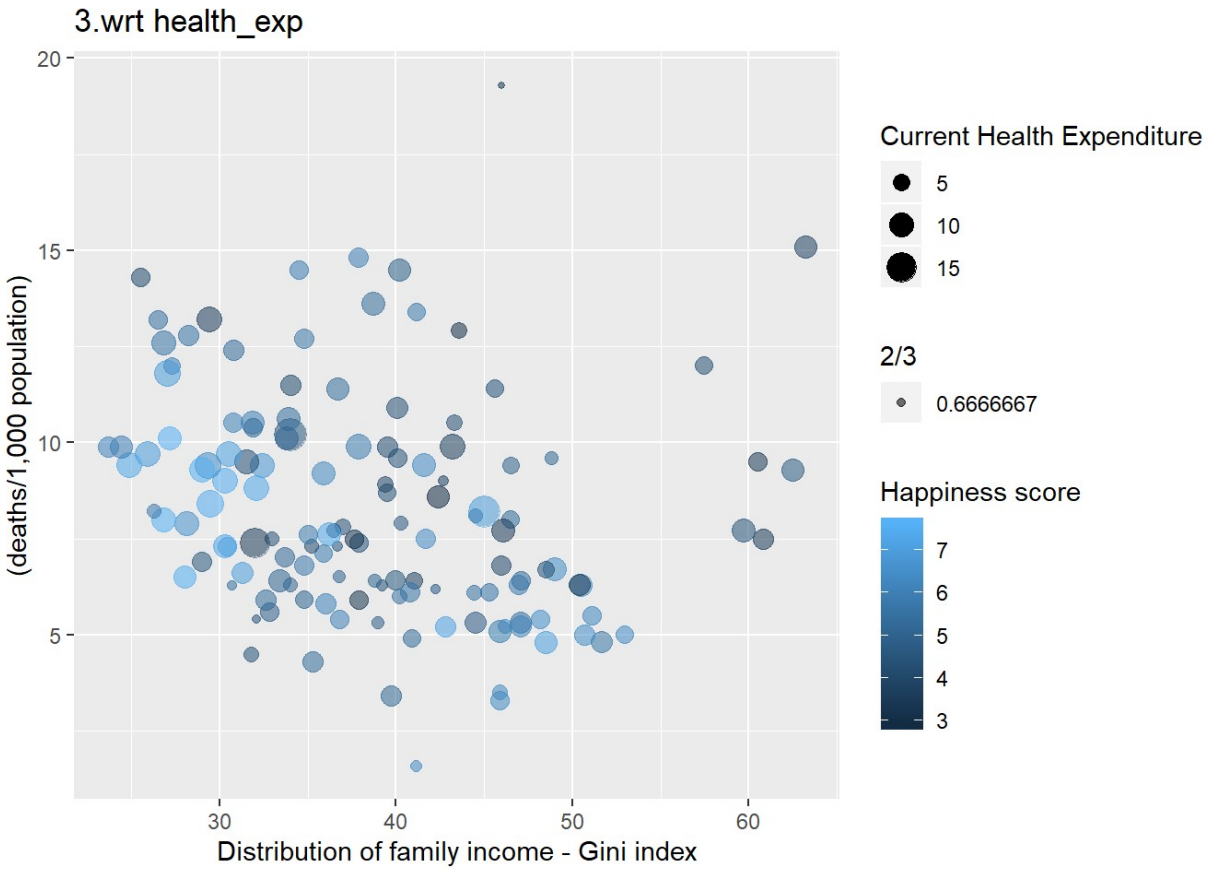
We do summary(lm.fit) in order to know the p values and the R^2 statistic model to check our model's accuracy which seems to be moderatly high 0.67 and and Fstatistic for the model then we plot out the graphs between residuals and the fitted values Then we we fit in the threoretical quantiles which seems to be adequately high following this we have the relation between the sqrt of standerdized residuals and fitted values (it's interesting to look at the deviations of the residuals) Then we look at the residuals vs levarage and try to formulate a relation which seems to be linearly decreasing and then we just look at the levarage values by hat value function and get the largest one.

```
gini2<-inner_join(gini,deathrate,by="Country")
gini2<-inner_join(gini2,healthexp,by="Country")
ggplot(data=gini2)+
  geom_point(mapping=aes(x=`Distribution of family income - Gini index`,y=`(deaths/
1,000 population)`,size=`Happiness score`,alpha=2/3))+
  geom_smooth(mapping=aes(x=`Distribution of family income - Gini index`,y=`(death
s/1,000 population)`))+
  ggtitle("2.wrt hapsc")
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```




```
ggplot(data=gini2)+geom_point(mapping=aes(x=`Distribution of family income - Gini i
ndex`,y=`(deaths/1,000 population)`,size=`Current Health Expenditure`,color=`Happin
ess score`,alpha=2/3))+ggtitle("3.wrt health_exp")
```




```
cor(gini2$`Distribution of family income - Gini index`,gini2$`Current Health Expend
iture`)
```

```
## [1] -0.2003428
```

```
cor(gini2$`Distribution of family income - Gini index`,gini2$`(deaths/1,000 populat
ion)`)
```

```
## [1] -0.1931462
```

```
cor(gini2$`Happiness score`,gini2$`Current Health Expenditure`)
```

```
## [1] 0.3154151
```

```
cor(gini2$`Happiness score`,gini2$`(deaths/1,000 population)`)
```

```
## [1] -0.1657499
```

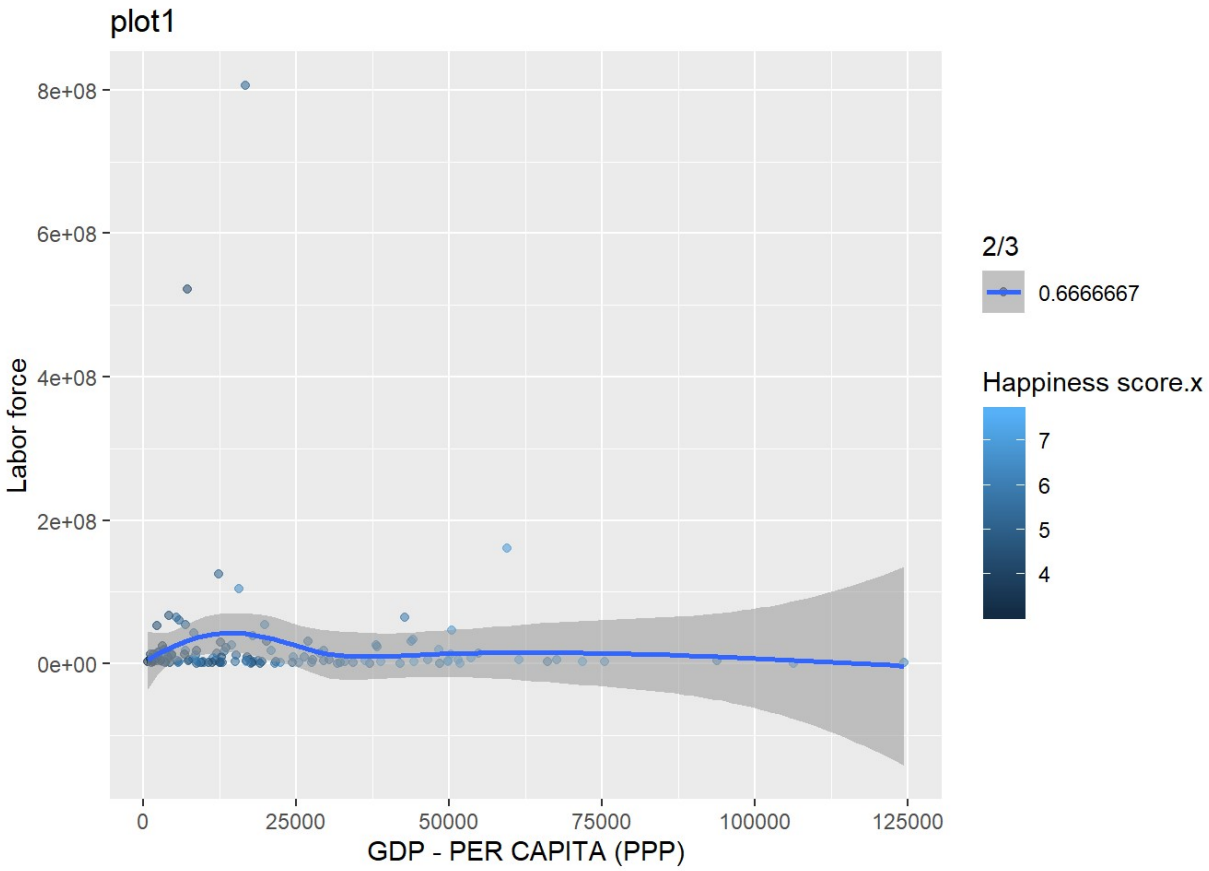
As we can see in the data set provided to us after combinig the tables we come the conclusion that the difference between household income and and Deathrate is not directly correlated or even correlated to health expenditure how we are still able to draft some insights Such as that if we look at the graph (3) we notice that people are some how happy even if the government doesn't have a

lot of health expenditure and even if death rate is high if the wage gap is low the people are happy seems like either one of the factor works for them but over all correlation is very low so we can't really say clearly.

```
gdppp3<-inner_join(gdppp,unemployment,by="Country")
gdppp3<-inner_join(gdppp3,laborforce,by="Country")
gdppp3<-inner_join(gdppp3,internetuser,by="Country")
gdppp3<-inner_join(gdppp3, urbanization, by="Country")
gdppp3<-inner_join(gdppp3, democracyindex, by="Country")
```

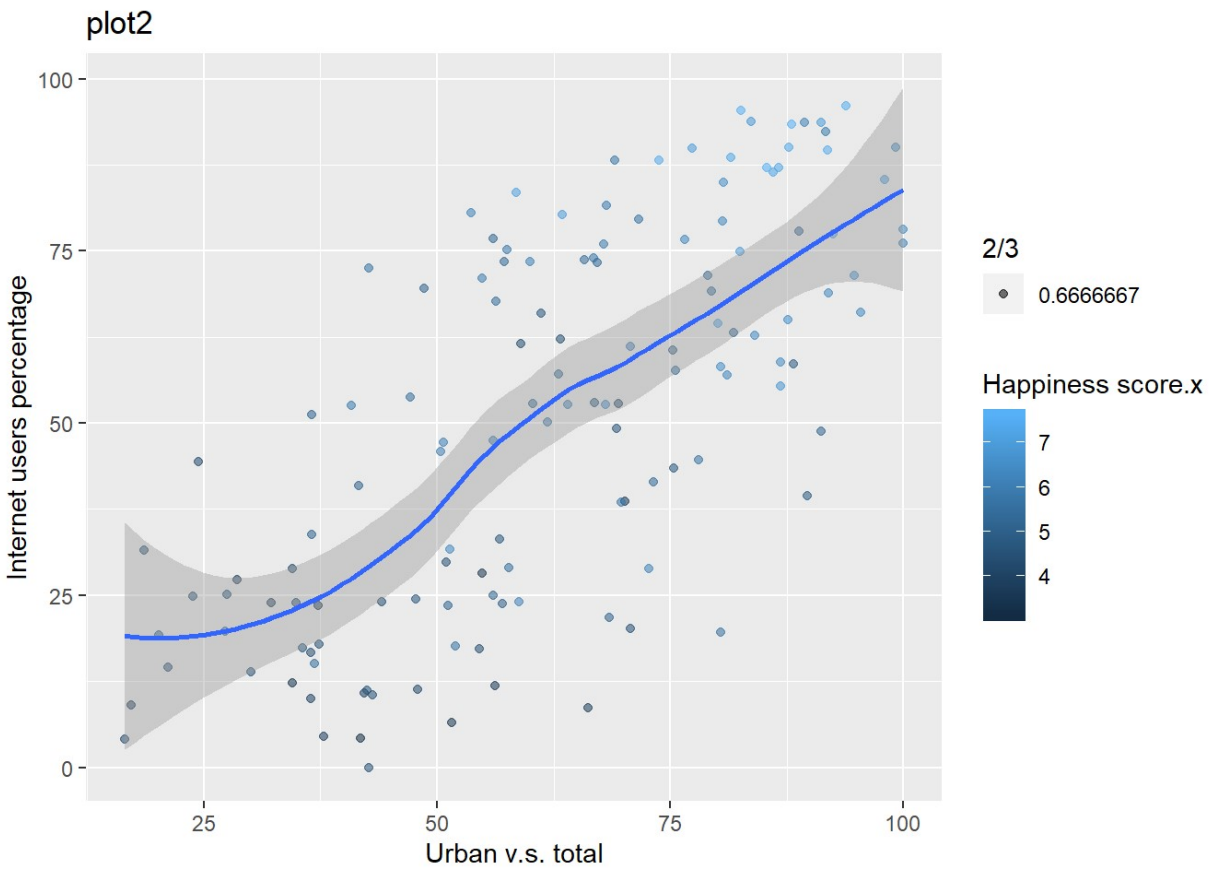
```
ggplot(data=gdppp3)+geom_point(mapping=aes(x=`GDP - PER CAPITA (PPP)`,y=`Labor forc
e`,color=`Happiness score.x`,alpha=2/3))+geom_smooth(mapping=aes(x=`GDP - PER CAPIT
A (PPP)`,y=`Labor force`,color=`Happiness score.x`,alpha=2/3))+ggtitle("plot1 ")
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



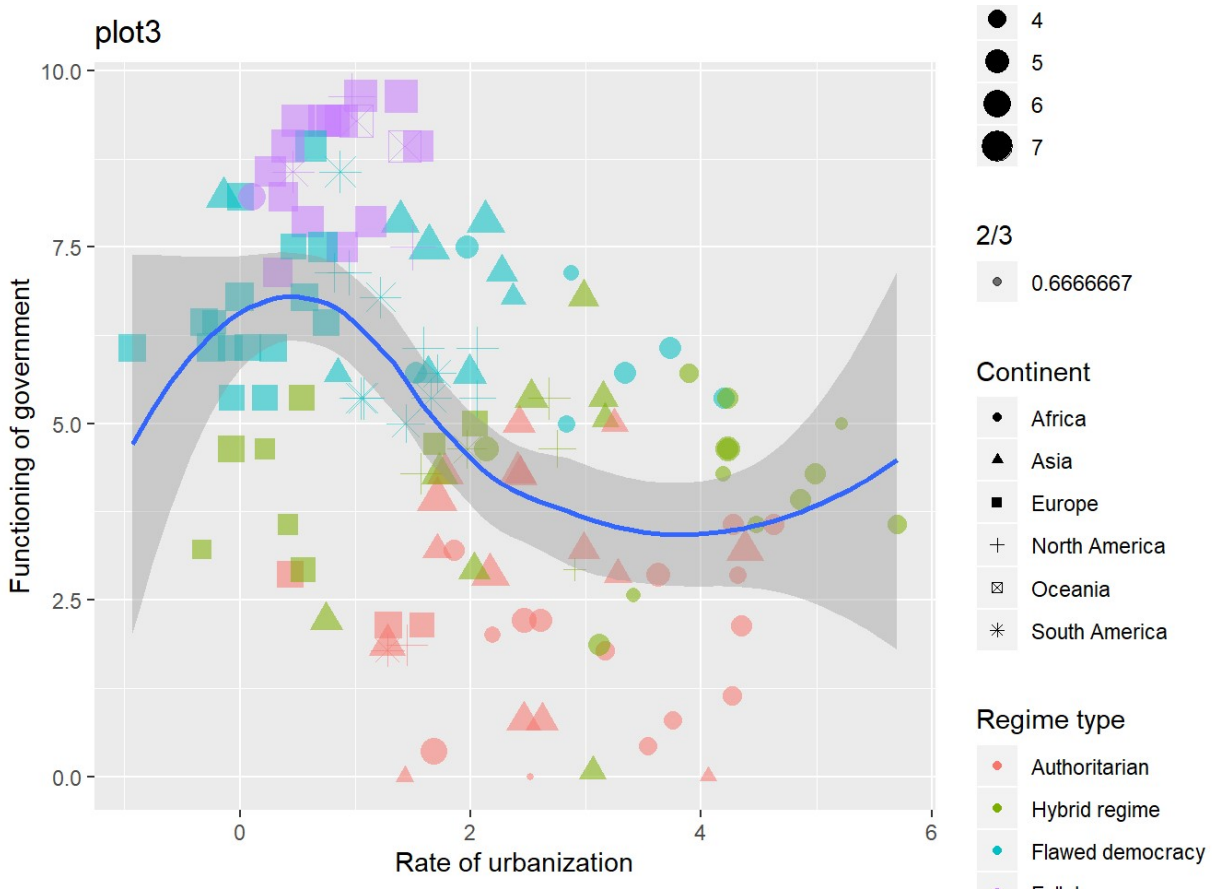

```
ggplot(data=gdppp3)+geom_point(mapping=aes(x=`Urban v.s. total`,y=`Internet users p
ercentage`,color=`Happiness score.x`,alpha=2/3))+geom_smooth(mapping=aes(x=`Urban
v.s. total`,y=`Internet users percentage`))+ggtitle("plot2")
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



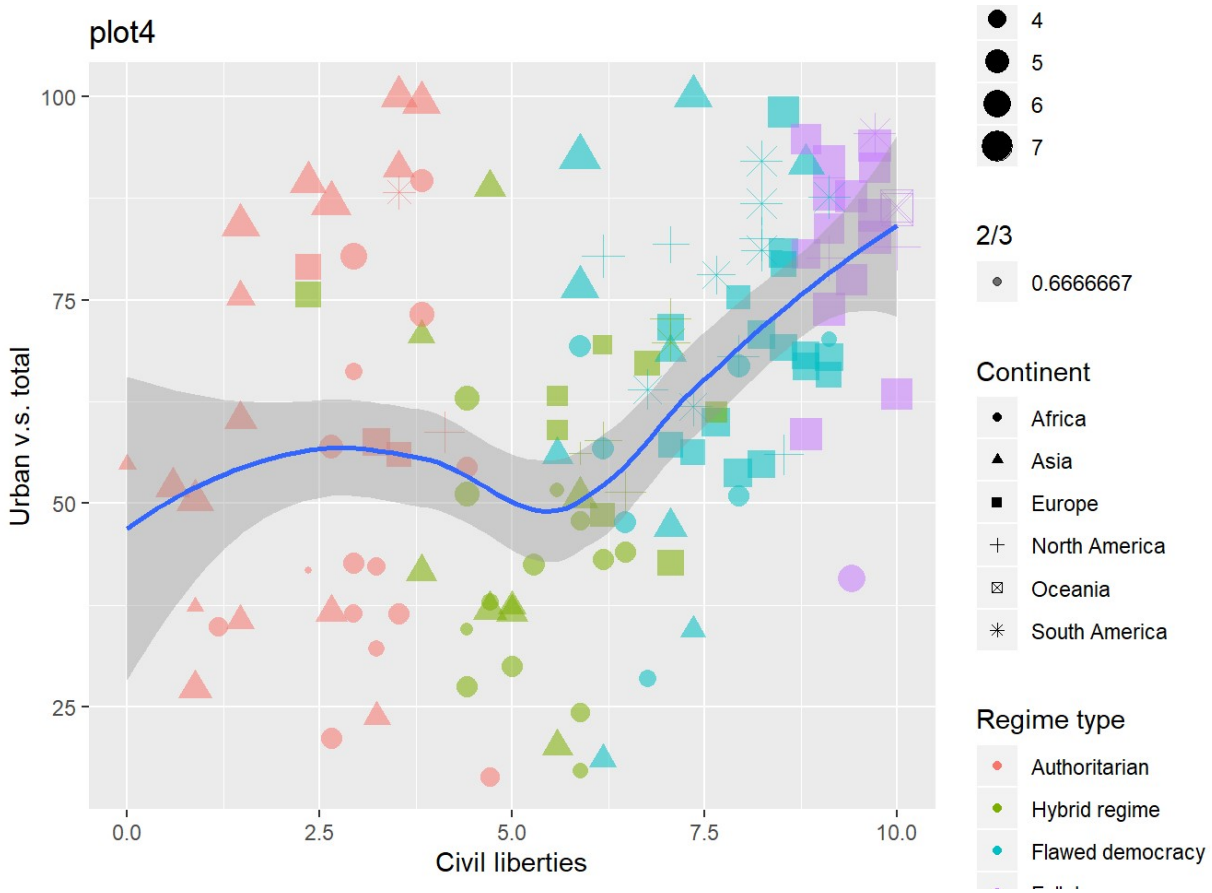

```
ggplot(data=gdppp3)+geom_point(mapping=aes(x=`Rate of urbanization`,y=`Functioning
of government`,size=`Happiness score`,shape=Continent,color=`Regime type`,alpha=2/
3))+geom_smooth(mapping=aes(x=`Rate of urbanization`,y=`Functioning of government
`))+ggtitle("plot3 ")
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



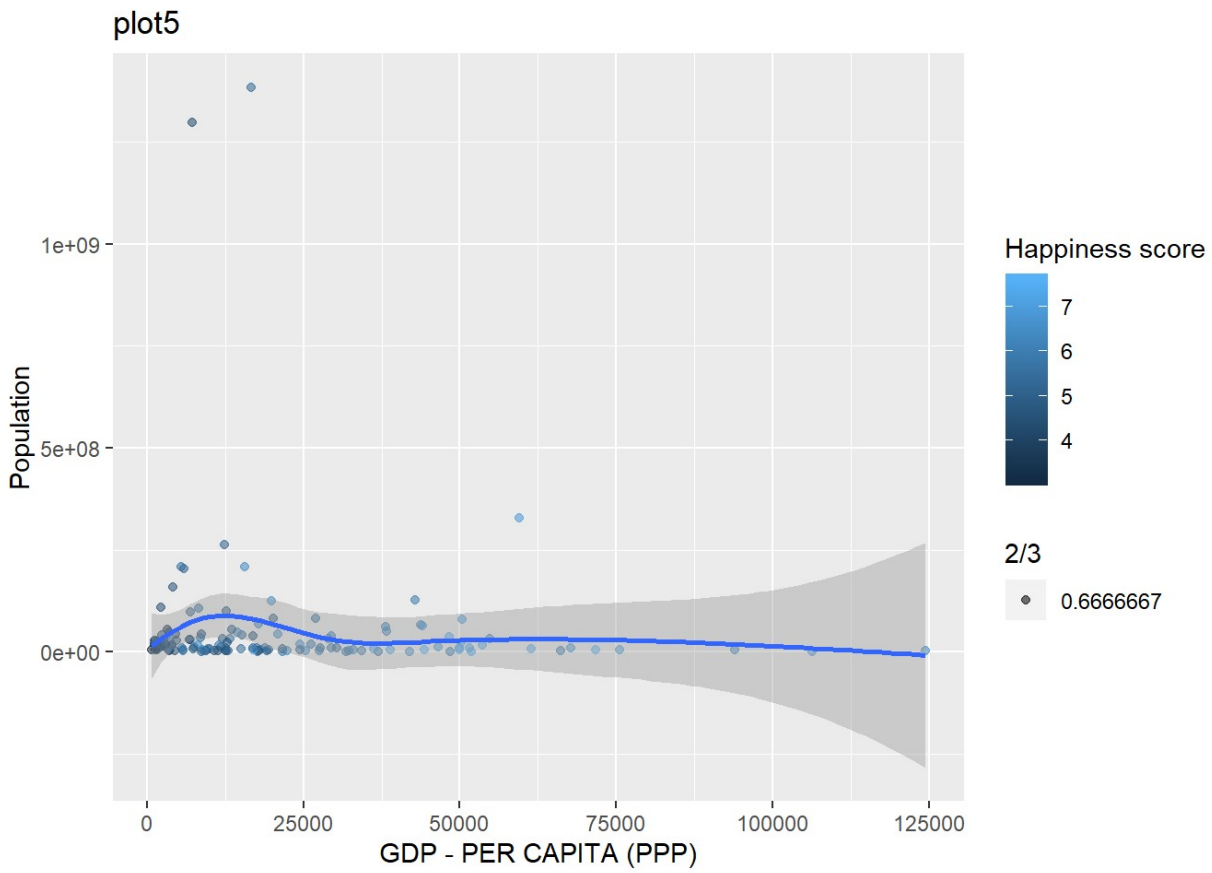

```
ggplot(data=gdppp3)+geom_point(mapping=aes(x=`Civil liberties`,y=`Urban v.s. total
`,size=`Happiness score`,shape=Continent,color=`Regime type`,alpha=2/3))+geom_smoot
h(mapping=aes(x=`Civil liberties`,y=`Urban v.s. total`))+ggtitle("plot4 ")
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



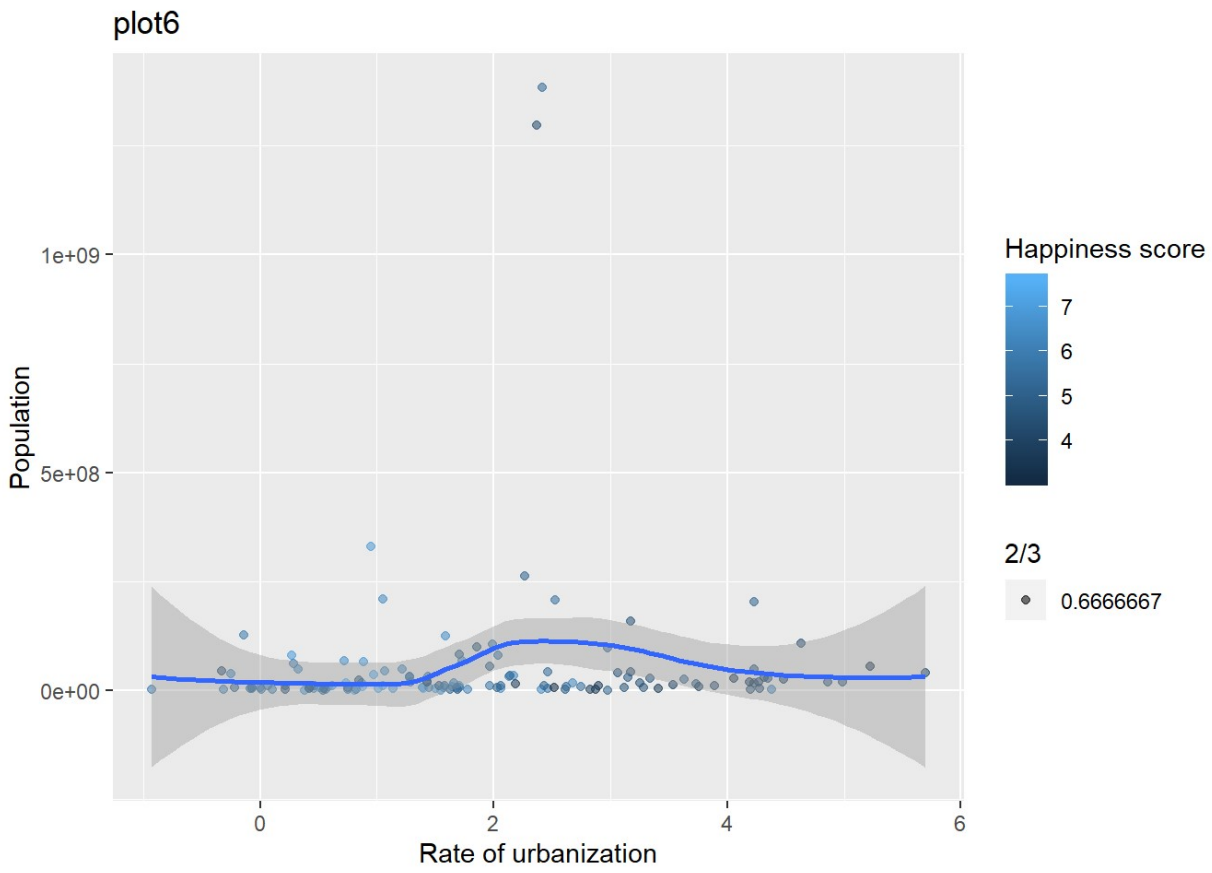

```
ggplot(data=gdppp3)+geom_point(mapping=aes(x=`GDP - PER CAPITA (PPP)`,y=`Population
`,color=`Happiness score`,alpha=2/3))+geom_smooth(mapping=aes(x=`GDP - PER CAPITA
(PPP)`,y=`Population`))+ggtitle("plot5 ")
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



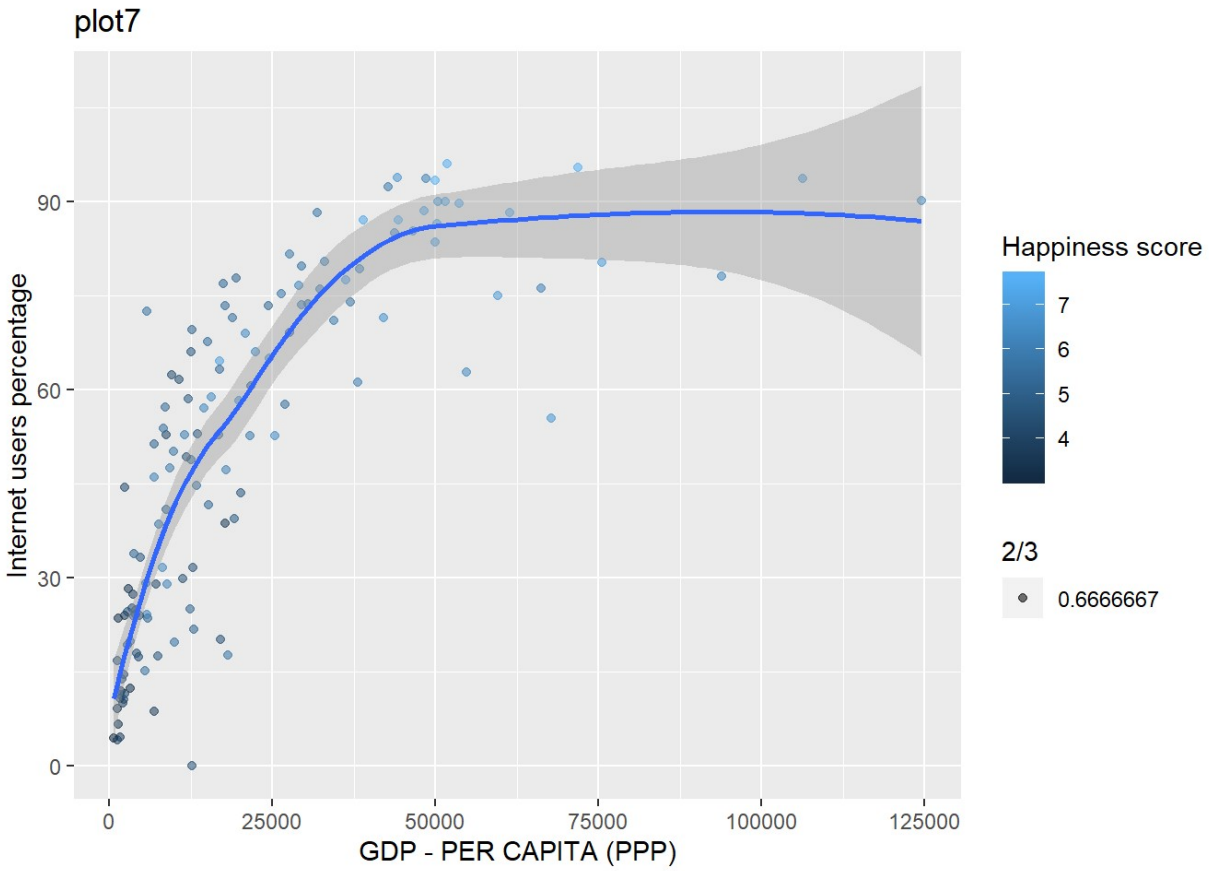

```
ggplot(data=gdppp3)+geom_point(mapping=aes(x=`Rate of urbanization`,y=`Population`,
color=`Happiness score`,alpha=2/3))+geom_smooth(mapping=aes(x=`Rate of urbanization
`,y=`Population`))+ggtitle("plot6 ")
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```




```
ggplot(data=gdppp3)+geom_point(mapping=aes(x=`GDP - PER CAPITA (PPP)`,y=`Internet u
sers percentage`,color=`Happiness score`,alpha=2/3))+geom_smooth(mapping=aes(x=`GDP
- PER CAPITA (PPP)`,y=`Internet users percentage`))+ggtitle("plot7 ")
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



Here we plot graph of other data sets after combining them to make observations as we come notice the graphs for GDP_PER_CAP to labour force is equal to the graph of GDP_PER_CAP and internet users which means we can only use one of these and gate the same results also we notice starkly familiar patterns in GDP_PER_CAP vs POPULATION and GDP_PER_CAP vs LABOUR_FORCE which incites on just using population for further observations and conclusions.Somehow rate of urbanisation provides a small deviation so we can't replace it with the urbanisation vs total here we basically concluded that a lotof these factors are common and as analysed in the upcoming tables by (Zosia) The insights we make on these are going to be very similar HOWEVER plot 3 and plot 4 show us that the people in with the lowest rate of urbanisation and the highest fucntioning government are generally the happiest ones these are prople in the europeon countries mostly and and also they have a low rate of urbanisation i.e they are already develepoed hence their citizens enjoy greater civil liberities.

Conclusions:

- After taking into consideration various factors we can easily form the conclusion that the happiness of a pearson is strongly influenced by the reality that surrounds them. However there is not only one factor that contributes to that. Our coclusoins can be sumed up as follows
- GDP_per capita affects happiness but not strongly
- However political and environmental factors very strongly affect happiness as these factors limit and control the day to day lives of the people
- We also noticed the fact that more developed countries with low rate of urbanisation also keep people happy as people are not facing pollution due to ongoing urbanisation
- We also concluded the fact that when all these factors start combining we finally begin to see happy people:

    - where the government is democratic
    - where gap in household income is low to moderate
    - where gdp per person is high
    - where the rate of urbanisation is low
    - where the percentage of urbanized area is greater
    - where people have more civil right and enjoy a functioning government which listens to them
    - where the electoral process is free and fair(pluralism)

All those factors turn out to be characteristics of more developed countries - which finally proves the thesis we formed in ithe introduction: higher happiness score occures more often in countries of Global North.

Appendix: contribution of each member of the group:

- Ronak Arora: Using ggplot to plot graphs and create insights from it then combining work with the data sets used by my work. Plottiing graphs with trend lines and finding correlation between different variables. Going in depth in a compairison - coputing the p values the relation between the residuals and the standerdized errors and other such facots, also computing the levarage value as well as the R^2 values to check the accuracy of the model on an overall basis which helped landing onto some good conclusions proving that those conclusions were somewhat accurate, performing regression and using trend line in many data sets to see how the line of best fit goes and what's the general trend in the data, plotting out the relation between the residuals and our given variables to get deeper insights. Finding the confidence interval. Relating different parts of the project and factors analysed between one another. Drawing joint conclusions from all parts of the project. Using: ggplot() corr() inner_join() predict() filter() hatvalues() summary() lm() parse_number() which.max() lm.fit() mutate() plot() log2() confirnt()
- Tirth Desai: Finding corelations between some of the data sets (with particular reference to unemployment). Using functions: mutate(),inner_join(), arrange(), ggplot(), geom_point(), geom_smooth(), cor(), geom_boxplot()
- Avery Dowling: Analysing the HEALTHEXP.csv and DEATHRATES.csv files to see if healthcare is related to the happiness of a country. Using: read_tvs(), summary(), ggplot(), geom_point(), geom_smooth() and parse_number.
- Zofia Wajda: Importing data sets and naming them in one manner, so that everyone operated at the same data. Tidying all the data sets (parsing variables inito wanted types). Forming thesis in introduction and comparing every part of the project with it. Analysing how political situation affects happiness with respect of how that connection is affected by access to internet (using data sets DEMOCRACYINDEX.csv, INTERNETUSER.csv and POPULATION.csv). Relating different parts of the project and factors analysed between one another. Drawing joint conclusions from all parts of the project. Using: ggplot(),geom_col(), geom_bar(), geom_point(), geom_smooth(), geom_boxplot(), facet_wrap(), coord_flip(), mutate(), fct_recode(), filter(), fct_reorder(), select(), mean(), ifelse(), group_by(), count(), spread(), read_tvs(), parse_number(), semi_join(), inner_join(), summary(), glimpse().