

Proposal Project

Giovanny Caluña - Ecuador

7/20/2021

For this project, I propose to develop an analysis and find relationships among the worked time by people through the years with the quality life level and happiness in different countries. Also, the work's purpose is to create models to predict the GDP values according to factors that conform to "happiness" for example freedom. To perform the project, I obtained different datasets which contain information about hours of work per week in different years and happiness score per country (with more interesting features).

As a small proof of this interesting project, I plotted 3 graphics to show the available data.

Data Sources:

<https://www.kaggle.com/orhankaramancode/city-quality-of-life-dataset?select=uaScoresDataFrame.csv>

<https://www.kaggle.com/unsdsn/world-happiness?select=2015.csv>

<https://data.oecd.org/emp/hours-worked.htm>

```
dataProject <-readr::read_csv('horasTrabajadasEuropa.csv')
```

```
##
## -- Column specification -----
## cols(
##   LOCATION = col_character(),
##   INDICATOR = col_character(),
##   SUBJECT = col_character(),
##   MEASURE = col_character(),
##   FREQUENCY = col_character(),
##   TIME = col_double(),
##   Value = col_double(),
##   'Flag Codes' = col_logical()
## )
```

```
dataProject2 <-readr::read_csv('uaScoresDataFrame.csv')
```

```
## Warning: Missing column names filled in: 'X1' [1]
```

```
##
## -- Column specification -----
## cols(
##   .default = col_double(),
##   UA_Name = col_character(),
##   UA_Country = col_character(),
##   UA_Continent = col_character()
## )
## i Use 'spec()' for the full column specifications.
```

```
dataProject3 <-readr::read_csv('world-happiness-report.csv')
```

```
##
## -- Column specification -----
## cols(
##   'Country name' = col_character(),
##   year = col_double(),
##   'Life Ladder' = col_double(),
##   'Log GDP per capita' = col_double(),
##   'Social support' = col_double(),
##   'Healthy life expectancy at birth' = col_double(),
##   'Freedom to make life choices' = col_double(),
##   Generosity = col_double(),
##   'Perceptions of corruption' = col_double(),
##   'Positive affect' = col_double(),
##   'Negative affect' = col_double()
## )
```

```
dataProject4 <-readr::read_csv('world-happiness-report-2021.csv')
```

```
##
## -- Column specification -----
## cols(
##   .default = col_double(),
##   'Country name' = col_character(),
##   'Regional indicator' = col_character()
## )
## i Use 'spec()' for the full column specifications.
```

```
dataProject5 <-readr::read_csv('2015.csv')
```

```
##
## -- Column specification -----
## cols(
##   Country = col_character(),
##   Region = col_character(),
##   'Happiness Rank' = col_double(),
##   'Happiness Score' = col_double(),
##   'Standard Error' = col_double(),
##   'Economy (GDP per Capita)' = col_double(),
##   Family = col_double(),
##   'Health (Life Expectancy)' = col_double(),
##   Freedom = col_double(),
##   'Trust (Government Corruption)' = col_double(),
##   Generosity = col_double(),
##   'Dystopia Residual' = col_double()
## )
```

```
dataProject6 <-readr::read_csv('2016.csv')
```

```
##
```

```
## -- Column specification -----
## cols(
##   Country = col_character(),
##   Region = col_character(),
##   'Happiness Rank' = col_double(),
##   'Happiness Score' = col_double(),
##   'Lower Confidence Interval' = col_double(),
##   'Upper Confidence Interval' = col_double(),
##   'Economy (GDP per Capita)' = col_double(),
##   Family = col_double(),
##   'Health (Life Expectancy)' = col_double(),
##   Freedom = col_double(),
##   'Trust (Government Corruption)' = col_double(),
##   Generosity = col_double(),
##   'Dystopia Residual' = col_double()
## )
```

```
dataProject7 <-readr::read_csv('2017.csv')
```

```
##
## -- Column specification -----
## cols(
##   Country = col_character(),
##   Happiness.Rank = col_double(),
##   Happiness.Score = col_double(),
##   Whisker.high = col_double(),
##   Whisker.low = col_double(),
##   Economy..GDP.per.Capita. = col_double(),
##   Family = col_double(),
##   Health..Life.Expectancy. = col_double(),
##   Freedom = col_double(),
##   Generosity = col_double(),
##   Trust..Government.Corruption. = col_double(),
##   Dystopia.Residual = col_double()
## )
```

```
dataProject8 <-readr::read_csv('2018.csv')
```

```
##
## -- Column specification -----
## cols(
##   'Overall rank' = col_double(),
##   'Country or region' = col_character(),
##   Score = col_double(),
##   'GDP per capita' = col_double(),
##   'Social support' = col_double(),
##   'Healthy life expectancy' = col_double(),
##   'Freedom to make life choices' = col_double(),
##   Generosity = col_double(),
##   'Perceptions of corruption' = col_character()
## )
```

```
dataProject9 <-readr::read_csv('2019.csv')
```

```
##  
## -- Column specification -----  
## cols(  
##   'Overall rank' = col_double(),  
##   'Country or region' = col_character(),  
##   Score = col_double(),  
##   'GDP per capita' = col_double(),  
##   'Social support' = col_double(),  
##   'Healthy life expectancy' = col_double(),  
##   'Freedom to make life choices' = col_double(),  
##   Generosity = col_double(),  
##   'Perceptions of corruption' = col_double()  
## )
```

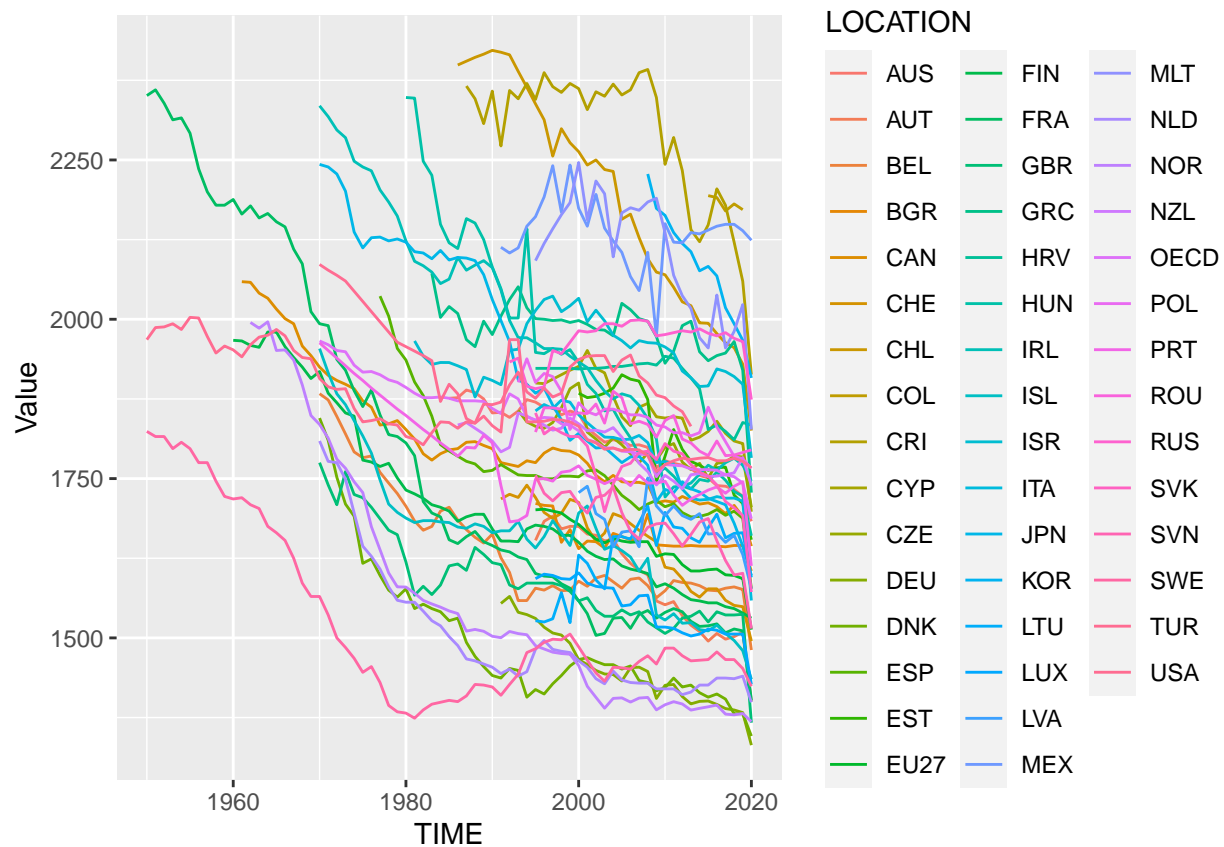
```
library(ggplot2)  
library(dplyr)
```

```
##  
## Attaching package: 'dplyr'
```

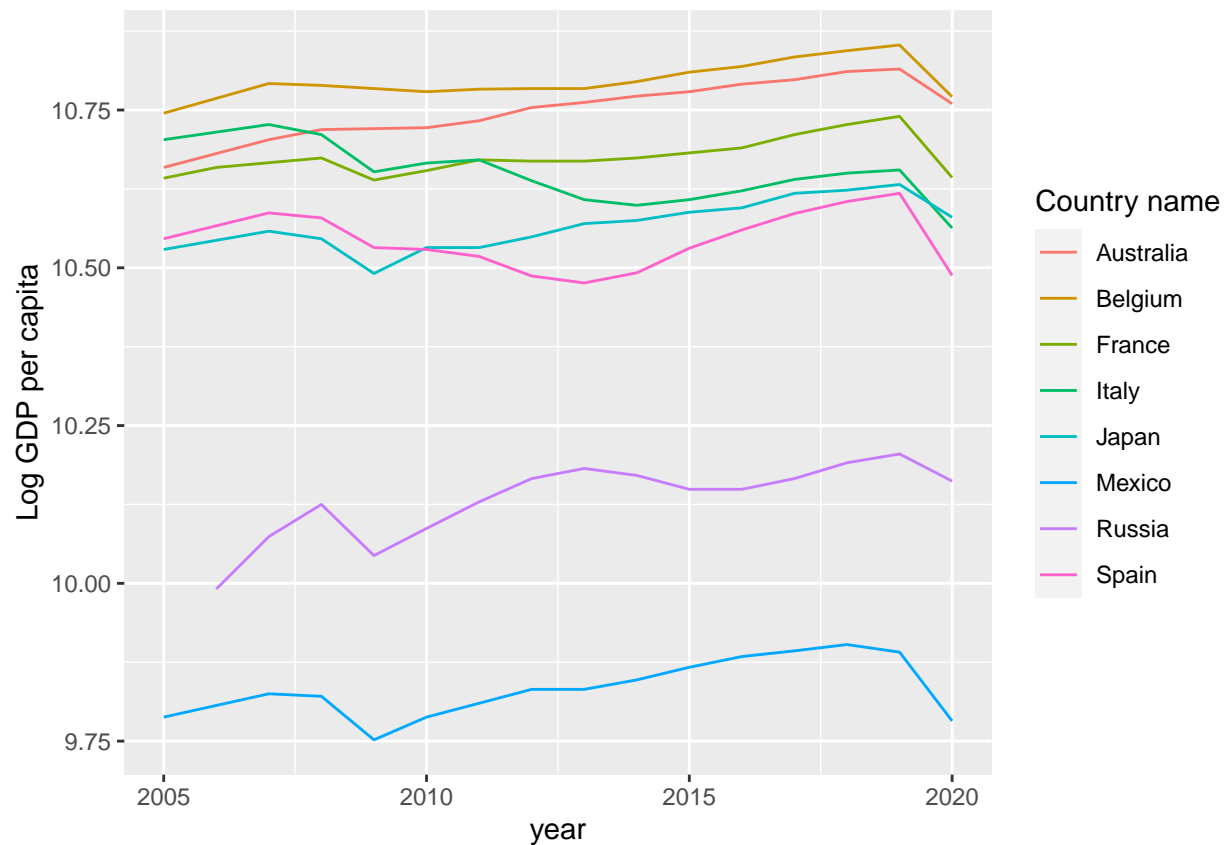
```
## The following objects are masked from 'package:stats':  
##  
##   filter, lag
```

```
## The following objects are masked from 'package:base':  
##  
##   intersect, setdiff, setequal, union
```

```
ggplot(dataProject, aes(x = TIME, y = Value), colour = LOCATION) + geom_line(aes(colour=LOCATION))
```

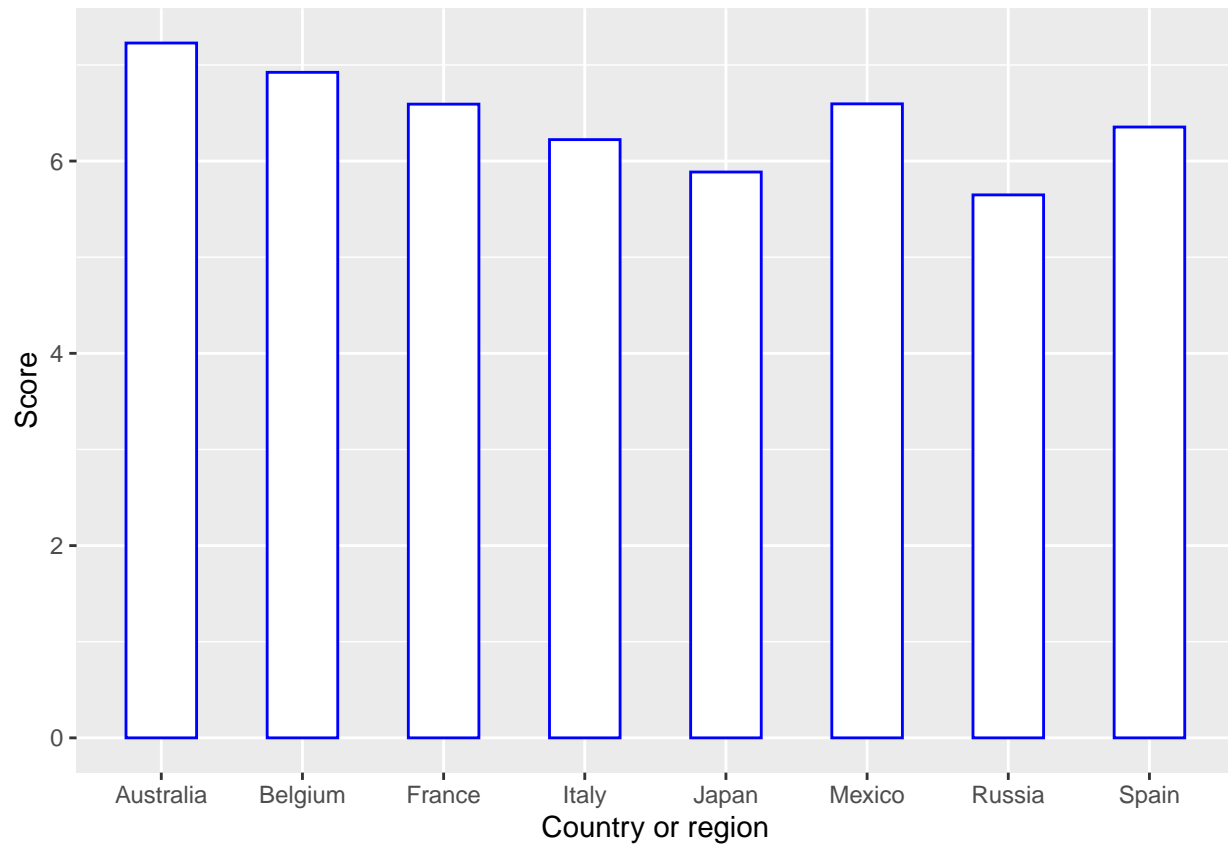


```
dataProject3 <-filter(dataProject3,`Country name` == 'Spain'|`Country name` == 'Australia'|
  `Country name` == 'Belgium'|
  `Country name` == 'Mexico'|
  `Country name` == 'Finlandia'|
  `Country name` == 'France'|
  `Country name` == 'Japan'|
  `Country name` == 'Russia'|
  `Country name` == 'Italy')
ggplot(dataProject3, aes(x = year, y = `Log GDP per capita`))+geom_line(aes(colour=`Country name`))
```



```
dataProject9 <-filter(dataProject9,`Country or region` == 'Spain'|
                        `Country or region` == 'Australia'|
                        `Country or region` == 'Belgium'|
                        `Country or region` == 'Mexico'|
                        `Country or region` == 'Finlandia'|
                        `Country or region` == 'France'|
                        `Country or region` == 'Japan'|
                        `Country or region` == 'Russia'|
                        `Country or region` == 'Italy')

ggplot(dataProject9, aes(x = `Country or region`, y = Score))+ geom_bar(stat="identity", color="blue", fill="blue")
```



All these plots are a small part of the proposed work. An entirely ETL process must be performed in order to normalize the data.

Also, other data sources can be added in order to obtain as much detail as possible.