HORIZON 2020 Analysis.

Ana Radoselovics
Programa en Inteligencia de Negocio y Big Data
Business Analytics

Associate professor: Gustavo Pino Moure
Coordinating professor: Joan Melià Seguí / Maria Pujol Jover

Delivery date: June 2019

HORIZON 2020 Analysis

TFM Business Analytics Ana Radoselovics

DATA SHEET

Title:	*HORIZON 2020 – CORDIS - Analysis*
Author:	*Ana Radoselovics Almagro*
Associate professor:	*Gustavo Pino Moure*
Coordinating professors:	*Joan Melià Seguí / Maria Pujol Jover*
Delivery Date: (mm/aaaa):	06/2019
Program:	Business Intelligence and Big Data
Project area:	*TFM Business Analytics*
Language:	*English*
Key Words:	*Horizon Dashboard, CORDIS for the economy, CORDIS for human development*

Abstract:

Purpose of this work is to analyse Open Data published in the context of EU H2020 CORDIS Project as well as macroeconomic data of the involved countries.

The objective is to know how the EU funds are assigned by topic, year and location as well as to analyse how are the projects impacting in the economic and human development of the european society.

Horizon 2020 is the biggest EU research and innovation program ever. Purpose is to improve the quality of life as well as the development level of each country.

Horizon 2020 focuses on three key areas: Excellence Science, Industrial Leadership and Societal Challenges.

The analysis has three parts. Two correlation analysis and a clustering analysis done with R. A Dashboard developed with Power BI.

The correlation analysis relates H2020 Assigned Funds as well as local Country investments in Education, Health and R&D with macroeconomic parameters.

Conclusions are that the EU Assigned Funds as well as the investments of each participant country are directly related with the economic development and the quality of life.

The clustering analysis does a segmentation of the participant states depending how much money was invested in every key area and how big is the GDP per capita. The result gives three groups of countries. The bigger are the assignments the higher is the GDP.

The Dashboard allows the user to navigate through four sheets which show in detail the relevant information about CORDIS as well as the correlations between assignments, economic and social parameters.

Index

List of tables and graphics

1. Introduction

1.1 Project context and justification

This work is based in the analysis of the Open Data published by the European Union about Research and Innovation Projects in the context of Horizon 2020. [1]. Horizon 2020 is the biggest EU Research and Innovation Programme ever.

It will lead to more breakthroughs, discoveries and world-firsts by taking great ideas from the lab to the market. Almost €80 billion of funding is available over 7 years (2014 to 2020) – in addition to the private and national public investment that this money will attract.

Excellent science, competitive industry and tackling societat challenges are at the heart of Horizon 2020. Files with information about the diferent projects, organisations and researchers are periodically published in the EU Open Data Portal in Excel or csv format.

Justification of this work is by one side to create a Dashboard which can enhance the information uploaded by the EU in the Open Data Portal. This Dashboard will allow the users to navigate through data related to CORDIS project as well as macroeconomic indicators. Purpose is to find correlations between EU assignments and economic as well as social parameters.

By other side deeper analytic questions can be raised with a statistical approach using R. For that purpose two correlation analysis and a clustering analysis will be done.

The personal motivations to do this project are by one side to learn more about an interesting and current topic; innovation and research in Europa. My scientifc background has been useful to understand the data.

Another motivation is to apply my knowledge and experience with R; correlation and clustering algorithms to analyse and relate the data.

And the last, to go deeper in my knowledge with visualization tools.

1.2 Project objectives definition.

Like previously described, H2020 project is created to support basic and applied investigation in order to improve the industry, economy, sustainable development and social benefit in Europa, as well as other countries.

DataSet for the analysis will be taken from the EU Open Data Portal. Excel and csv files with information about more than 8000 projects.

Those files contain information about the topic, a brief summary of every project, coordinator country as well as money invested on that. Three main investment areas are supported: Excellence Science, Industrial benefits and Societal Challenges.

The DataSet has to be enhanced with Macroeconomic and demographic data also downloaded from Open Portals. It's important to remark that we refer to EU Investments as the Assigned Funds of CORDIS projects and local Investments as the money assigned by each country in Education, Health and R&D respectively.

The first analysis is done with R. Two objectives are defined:

Correlation analysis to measure the success of the program.
Following analytic questions are formulated for each country:

1) How is the CORDIS investment related with the PIB?.
2) How is the investment done by the respective governments related with the same parameter?.
3) Which correlation is bigger?.
4) Same questions related to the average salary.
5) How are correlated the assigned funds by the EU with the respective investments of each country in education, Health and R&D?.
6) How is the live expectancy related with the EU investments?.
7) How is the live expectancy related with the country public investments in Health?.
8) Which correlation is bigger?.
9) Are there other factors which also can impact in the live expectancy?. For example level of CO2?. Is the live expectancy also related to environmental aspects?.
10) Same questions for the innovation ratio. How is it correlated to the EU investments?.
11) How is the innovation ratio correlated to the local funds assigned to R&D?.
12) Which correlation is bigger?.
13) Same analysis for the unemployment Rate Percentage.
14) Same analysis for the Gender Gap Ranking.

After the discussion the analysis is in the position to evaluate the impact or CORDIS projects in the economy, quality of life and development level of the involved states. Moreover, it can be discussed if the CORDIS strategies are aligned with the local investment strategies.

Second objective is to do a cluster analysis to segment the participant countries. Four parameters are taken into account: assigned funds per capita in the three main investigation areas: Excellent Science, Industrial Leadership, Societal Challenges. The fourth parameter is the GDP per capita.

Following questions are raised after the clustering:

1) How are the countries classified ?.
2) How is the trend of each group of countries?. Which branch receives more funds?.
3) How are related the funds invested in each cluster with the GDP?.
4) How can the states be grouped for future collaborations?.

After the clustering we are in the position to go deeper and find some new correlations based on specific human indicators related to the investments in Societal Challenges projects.

R analysis has given a more statistical approach which can be useful to understand how the investments can impact in the economy as well as the quality of life.

The second analysis with Power BI Desktop (Dashboard) will give a visual approach that can help the user to understand how the data are structured and related. The user will have the opportunity to navigate through the data and see the impact that the investments can have in their lives.

Those are some of the reports which will be designed:

- Summary of Cordis H2020: list of all the projects with the most relevant information.
- Pie Report total invest amount by investigation area and sub-area.
- Bar Plot with the EU Investments vs GDP and Average Salary in each country.
- Visualization of the temporal evolution of EU Investments, GDP and Average Salary for each country.
- Gender Gap vs EU Investments.
- Treemap showing EU Investments in Societal Challenges projecs vs HDI, Global Peace Ranking and Immigration Rate by country
- Map showing the life expectancy vs local investments in health and CO2 emissions.
- EU Investments in Industrial Leadership vs local investments in R&D.
- Debt per capita and Trade Balance %GDP.

The pages can be filtered by year and country respectively.

1.3 Project scope and methodology

As described in the previous section, this Project will consist of two parts: correlation and clustering analysis with R and a Dashboard done with Power BI Desktop.

As a source DataSet for the analysis will be used an enhanced Set combining data of the EU Open Data Portal with macroeconomic parameters.

More detailed information about Source DataSets will be developed in the second Chapter as well as Data Cleansing and Data Modelling Methodology.

For Management Methodology we will use an Agile approach [2]. We will consider each PEC as a Sprint and the multiple discussions with the assigned Professor as the periodic team meetings.

For the Risk Management also an Agile approach has been used. Some risks were identified since the very beginning but others came up after the periodic alignments.

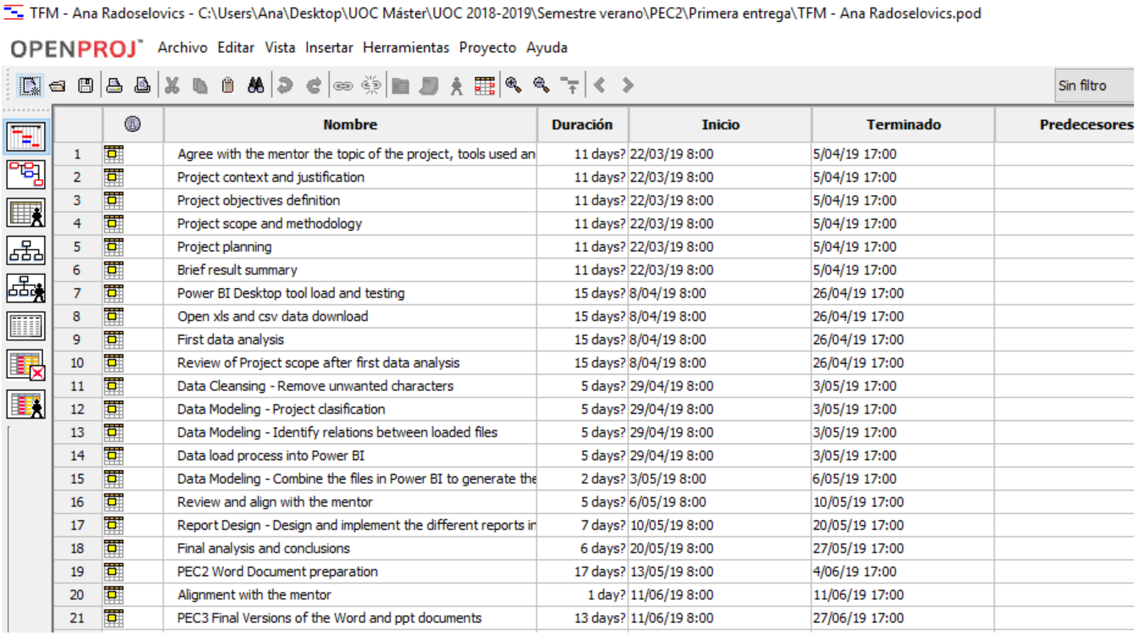
The next Section (Project Planning) will go more in detail with these topics. Next table summarizes the identified risks of the Project and the actions performed to mitigate them:

Decision	Associated risks	Actions to mitigate the risk
Tool Selection: initially Power BI	Less expertise. Analytic capabilities not enough to go deeper in correlation and clustering	Start the analysis with a well known tool: R.
Initial DataSet Selection: Data published by the BCE	Limited functional background to analyse the data	Chose other DataSet.
Derived from the previous risk	Objectives of PEC1 were not achieved.	Another DataSet which allows to go deeper in the analytical topics.
Second DataSet: Projects and reports of the EU Open Data Portal (H2020)	Bad data quality: the text contains many strange characters. Parameters are not enough to perform the analysis	Cleansing process. Enhance the DataSet with macroeconomic data.
Derived from the previous risks	Delay in the plan.	High dedication. Frequent discussions.

Table 1: Project Risks

1.4 Project planning.

This graphic represents the initial suggested planification

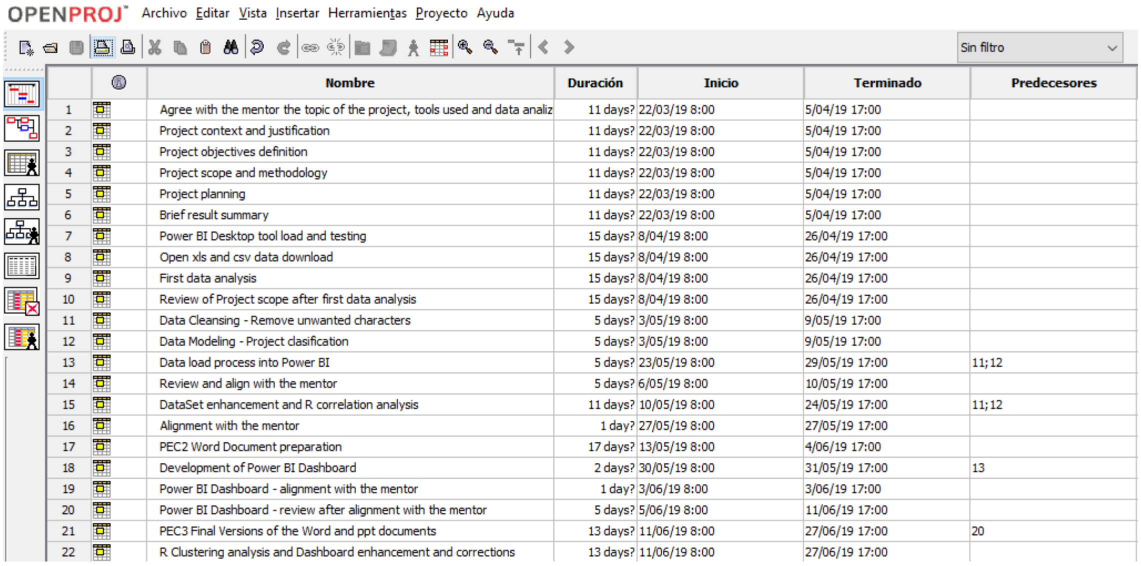


TFM - Ana Radoselovics - C:\Users\Ana\Desktop\UOC Máster\UOC 2018-2019\Semestre verano\PEC2\Primera entrega\TFM - Ana Radoselovics.pod
OPENPROJ
Archivo Editar Vista Insertar Herramientas Proyecto Ayuda
Sin filtro

	ⓘ	Nombre	Duración	Inicio	Terminado	Predecesores
1	▦	Agree with the mentor the topic of the project, tools used an	11 days?	22/03/19 8:00	5/04/19 17:00	
2	▦	Project context and justification	11 days?	22/03/19 8:00	5/04/19 17:00	
3	▦	Project objectives definition	11 days?	22/03/19 8:00	5/04/19 17:00	
4	▦	Project scope and methodology	11 days?	22/03/19 8:00	5/04/19 17:00	
5	▦	Project planning	11 days?	22/03/19 8:00	5/04/19 17:00	
6	▦	Brief result summary	11 days?	22/03/19 8:00	5/04/19 17:00	
7	▦	Power BI Desktop tool load and testing	15 days?	8/04/19 8:00	26/04/19 17:00	
8	▦	Open xls and csv data download	15 days?	8/04/19 8:00	26/04/19 17:00	
9	▦	First data analysis	15 days?	8/04/19 8:00	26/04/19 17:00	
10	▦	Review of Project scope after first data analysis	15 days?	8/04/19 8:00	26/04/19 17:00	
11	▦	Data Cleansing - Remove unwanted characters	5 days?	29/04/19 8:00	3/05/19 17:00	
12	▦	Data Modeling - Project clasification	5 days?	29/04/19 8:00	3/05/19 17:00	
13	▦	Data Modeling - Identify relations between loaded files	5 days?	29/04/19 8:00	3/05/19 17:00	
14	▦	Data load process into Power BI	5 days?	29/04/19 8:00	3/05/19 17:00	
15	▦	Data Modeling - Combine the files in Power BI to generate the	2 days?	3/05/19 8:00	6/05/19 17:00	
16	▦	Review and align with the mentor	5 days?	6/05/19 8:00	10/05/19 17:00	
17	▦	Report Design - Design and implement the different reports ir	7 days?	10/05/19 8:00	20/05/19 17:00	
18	▦	Final analysis and conclusions	6 days?	20/05/19 8:00	27/05/19 17:00	
19	▦	PEC2 Word Document preparation	17 days?	13/05/19 8:00	4/06/19 17:00	
20	▦	Alignment with the mentor	1 day?	11/06/19 8:00	11/06/19 17:00	
21	▦	PEC3 Final Versions of the Word and ppt documents	13 days?	11/06/19 8:00	27/06/19 17:00	

After the PEC1 (first Sprint) the 5th of April, it was concluded that the project objectives and scope as well as the methodology were not completely defined. Particularly the analytical questions we pretended to answer. Risks 1, 2 and 3 of the previous section were detected.

After that a new Project DataSet was loaded. Project Scope was reviewed. Risk 4 was detected. The quality of the new DataSet was not optimal. Loaded files contained a high number of strange characters. To mitigate the risk it was planned a Data Cleansing task between the 29th of April and the 3rd of May. In parallel the Data Modelling task was also performed. It included a semi-manual task to classify the project area or key.
The data were loaded in Power BI and the conclusions were aligned with the Professor.
After the alignment on May the 10th it was concluded that the model was simple. More parameters were needed to go deeper into the analytical questions. Risks 5 and 6 were detected.
To mitigate them the DataSet needed to be enhanced. A new planification was done.



		Nombre	Duración	Inicio	Terminado	Predecesores
1		Agree with the mentor the topic of the project, tools used and data analiz	11 days?	22/03/19 8:00	5/04/19 17:00	
2		Project context and justification	11 days?	22/03/19 8:00	5/04/19 17:00	
3		Project objectives definition	11 days?	22/03/19 8:00	5/04/19 17:00	
4		Project scope and methodology	11 days?	22/03/19 8:00	5/04/19 17:00	
5		Project planning	11 days?	22/03/19 8:00	5/04/19 17:00	
6		Brief result summary	11 days?	22/03/19 8:00	5/04/19 17:00	
7		Power BI Desktop tool load and testing	15 days?	8/04/19 8:00	26/04/19 17:00	
8		Open xls and csv data download	15 days?	8/04/19 8:00	26/04/19 17:00	
9		First data analysis	15 days?	8/04/19 8:00	26/04/19 17:00	
10		Review of Project scope after first data analysis	15 days?	8/04/19 8:00	26/04/19 17:00	
11		Data Cleansing - Remove unwanted characters	5 days?	3/05/19 8:00	9/05/19 17:00	
12		Data Modeling - Project clasification	5 days?	3/05/19 8:00	9/05/19 17:00	
13		Data load process into Power BI	5 days?	23/05/19 8:00	29/05/19 17:00	11;12
14		Review and align with the mentor	5 days?	6/05/19 8:00	10/05/19 17:00	
15		DataSet enhancement and R correlation analysis	11 days?	10/05/19 8:00	24/05/19 17:00	11;12
16		Alignment with the mentor	1 day?	27/05/19 8:00	27/05/19 17:00	
17		PEC2 Word Document preparation	17 days?	13/05/19 8:00	4/06/19 17:00	
18		Development of Power BI Dashboard	2 days?	30/05/19 8:00	31/05/19 17:00	13
19		Power BI Dashboard - alignment with the mentor	1 day?	3/06/19 8:00	3/06/19 17:00	
20		Power BI Dashboard - review after alignment with the mentor	5 days?	5/06/19 8:00	11/06/19 17:00	
21		PEC3 Final Versions of the Word and ppt documents	13 days?	11/06/19 8:00	27/06/19 17:00	20
22		R Clustering analysis and Dashboard enhancement and corrections	13 days?	11/06/19 8:00	27/06/19 17:00	

Between the 10th and 24th of May more data were loaded to improve the DataSet. This part was complex, how to compare assigned data (EU H2020 project) with macroeconomic indicators. Details will be clarified in the second chapter. The correlation analysis with R was also performed in the same period.

Preparation and documentation of the PEC2 Word is also scheduled in parallel.

During the PEC3 the R analysis has been completed with a segmentation analysis and a deeper correlation related to Projects in Societal Challenges area.

1.5 Brief summary of the obtained results

The deliveries of this project consist in two parts:

Analysis with R about correlations between assigned funds by the EU as well as the respective countries with macroeconomic parameters like PIB, average salary, unemployment rate, innovation ratio or life expectancy.
The questions that the analysis answers are how the investments are related with indicators that represent the quality of life or development level of a State.

Segmentation analysis of the countries using four parameters:

EU Investments in project type Excellence Science
EU Investments in project type Industrial Leadership
EU Investments in project type Societal Challenges
GDP per capita

The result gives three main groups. Conclusions will be detailed in Chapter 3.

Deeper correlation analysis between human and social parameters and EU Investments in Societal Challenges projects.

Dashboard performed with Power BI Desktop. This application will allow the user to navigate through diferent visualizations of the data. The goal is to have a view how the assignments of the EU are distributed by Country, Project type and year. The questions that the application will answer to the user are:

1) How are the funds distributed by country?.
2) How are the funds distributed by year?. Which is the trend?.
3) How are the funds distributed by project type?.
4) How can impact the EU Investments in the economical development?
5) How can impact the EU Investments in the human development?
6) How are the EU Investments related with the Local Investments?

More details about the Dashboard design and structure will be given in

Chapter 2. [3]

2. Project development

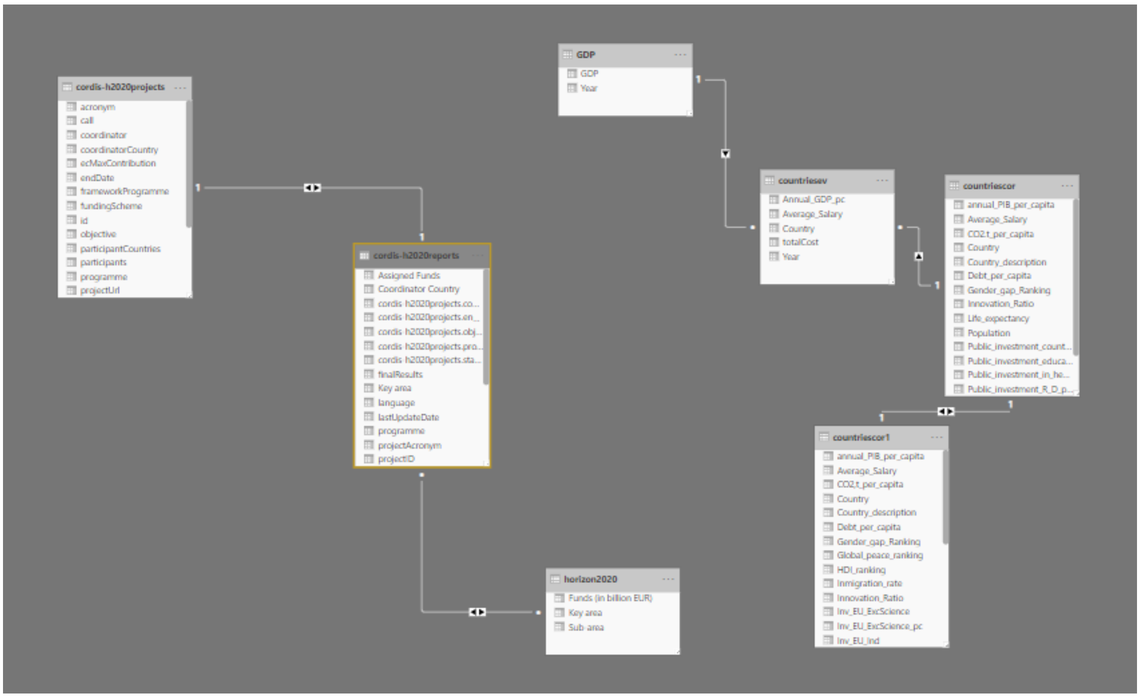
2.1 Data Cleansing.

For the Power BI analysis files were downloaded from the EU Open Data Portal Ref [4]: H2020 Projects and H2020 Report summaries.
It was detected that the files contained many strange characters due to the translation of languages like german or polish.

A Data Cleansing effort was added to the planification. For that purpose a semi-manual process using Excel find&replace function was performed.

2.2 Data Modelling.

Power BI Data Model: this is how the Data Model for the Dashboard results. Annex[1].



To have a more clear view how the Data Sets have been created, following table shows the Data Source, the Relevant fields and the Data Modelling Process.

Data Set	Source	Relevant Fields	Data Modelling process
cordis-h2020projects	Ref[4]	Start Date, Total Cost, Coordinator Country	Power BI
cordis-h2020reports	Ref[4]	Title, Key Area	Power BI
horizon2020	Ref[1]	Key area, Sub-area, Funds (in billion EUR)	Manual
GDP	Ref[5]	GDP, year	Manual
countriesev	Ref[5]	Evolution of GDP, Average Salary and total EU investments per year	Annex[4]
countriescor	Ref[5];Ref[6]	Economic parameters	Annex[2]
countriescor1	Ref[5]	Human indicators	Annex[3]

Table 2: Power BI Data Model

Column Key Area of **cordis-h2020reports** file has been semimanually calculated to distinguish between projects related to Excellence Science, Industrial leadership or Societal Challenges.

Cordis-h2020 reports table in Power BI has been enhanced with following columns of the **cordis-h2020 projects** file:*Start date*, *End date*, *Project url*, *Objective*, *Total cost*, *Coordinator and coordinator country.*

countriesev, countriescor and countriescor1 have been created using R programs and also used as input files for the correlation and segmentation analysis.

countriescor is focused on economic indicators as GDP, Average Salary, Unemployment Rate Percentage, Innovation Ratio, Gender Gap Ranking, EU Investments as well as Local investments in Education, Health and R&D. It is used as input file for the first correlation analysis which will be detailed in the next Chapter.

countriescor1 is an enhancement of countriescor with the calculated EU Investments per Project Type (Excellence Science, Industrial leadership or Societal Challenges). New indicators have been added like HDI (Human development index), Inmigration Rate and Global Peace Ranking.

This file is used as input of the clustering analysis and the second correlation analysis performed with R.

To generate the files countriesev, countriescor and countriescor1 in the Data Modelling process performed with R, files detailed in Annex [5], [6], [7] and [8] were used. Data were taken from Ref [4], [5] and [6].

2.3 First correlation analysis results and conclusions.

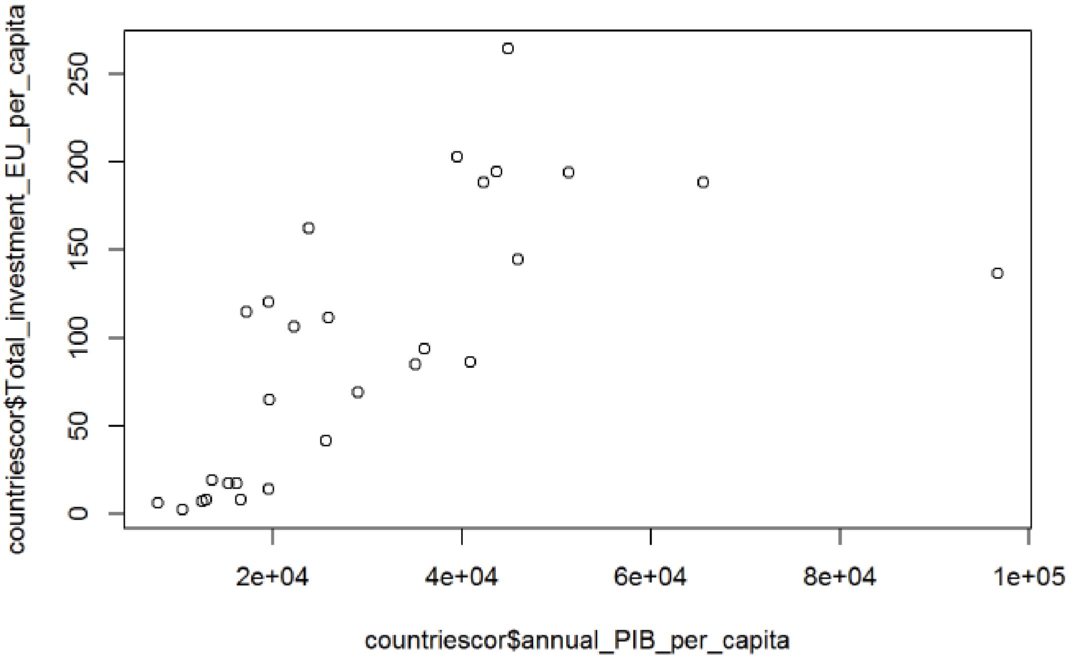
Annex [9] shows in html format the result of the execution with R of the first correlation analysis. Assignments of the EU have been correlated with macroeconomic parameters using the cor function.

Cor function between two variables calculates the correlation coefficient which adjusts these parameters to a linear function.

A correlation coefficient near to 1 means that the relationship between the variables is a linear approach (linear regression).

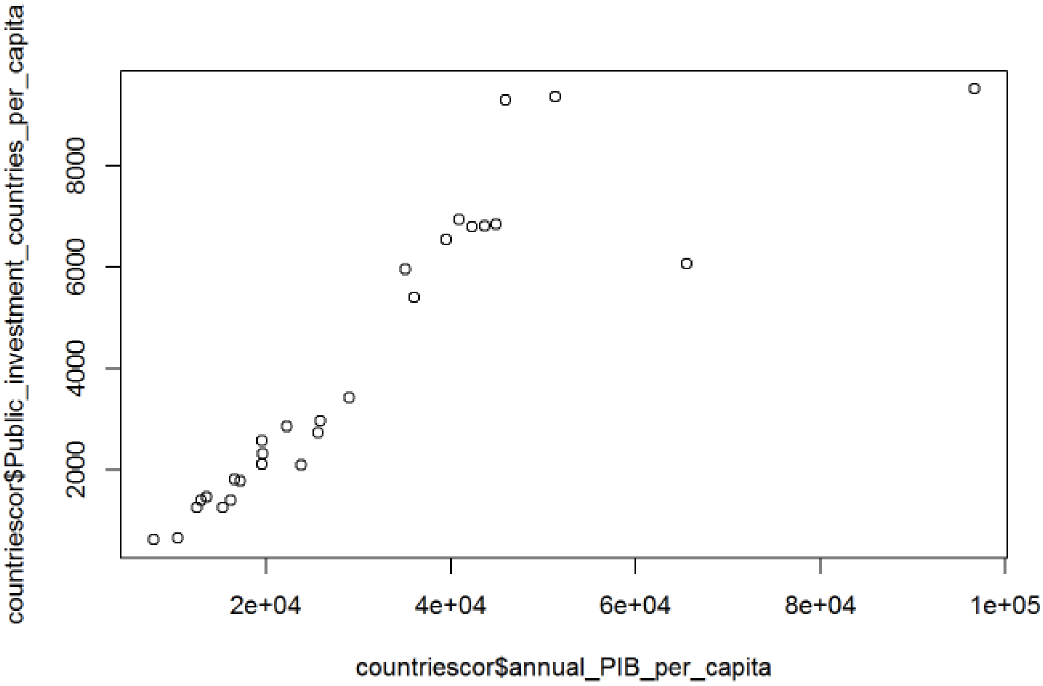
Let´s see the results and conclusions of the analysis:

Plot 1: GDP vs EU investments pc



This graphic shows that actually the investments of the EU are related to the GDP with a correlation factor of 0.6688031.

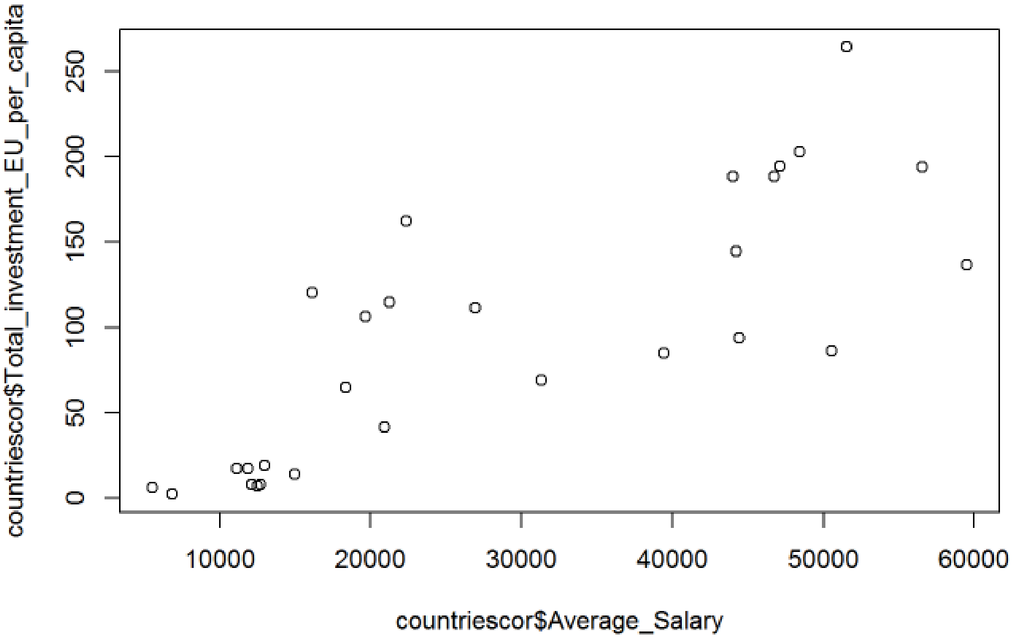
Plot 2: GDP pc vs Local investments pc



This correlation factor is still bigger if we compare with the respective investments of each country in education, Health and R&D respectively: 0.8824068.

Same analysis for the Average Salary:

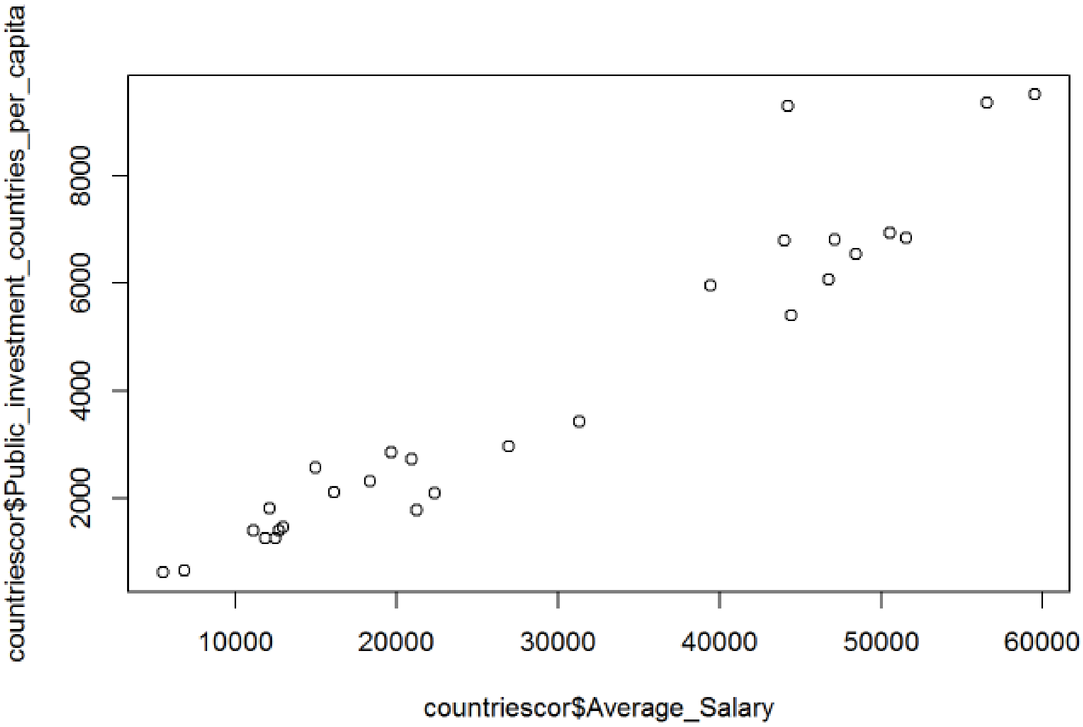
Plot 3: Average Salary vs EU investments



The correlation factor in this case is 0.805711. The points fit very good a linear approach.

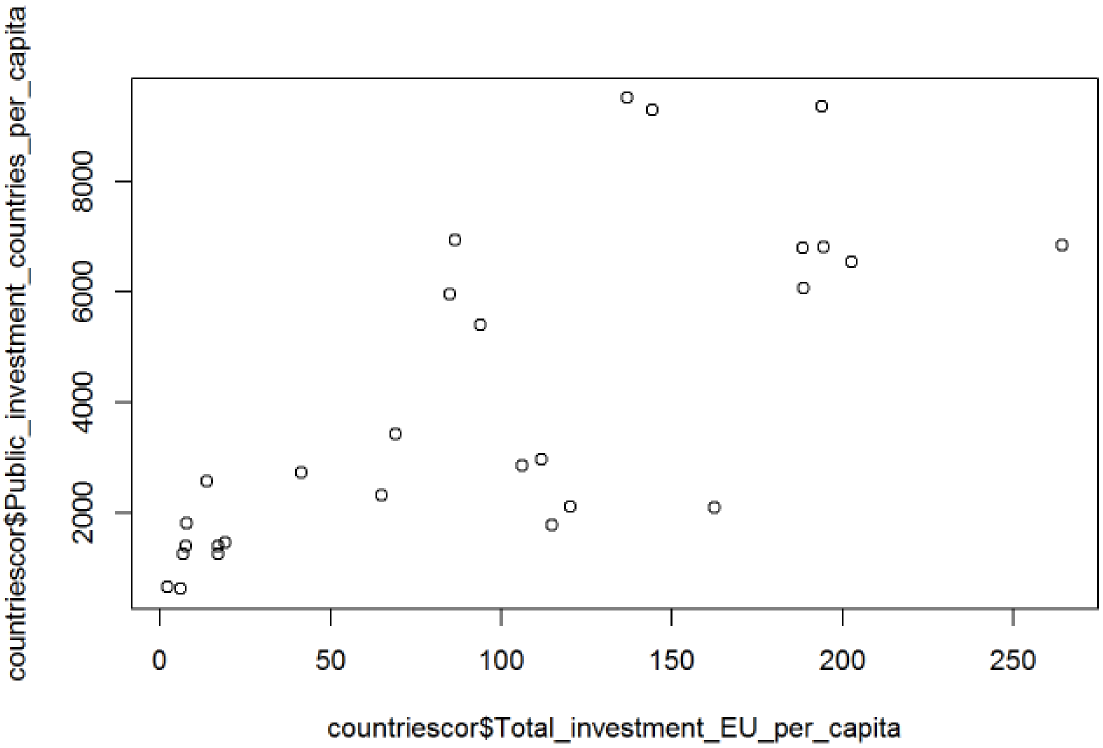
The linear trend is still bigger if we compare Average Salary with the local investments of each country. Coefficient: 0.9605182

Plot 4: Average Salary vs Local inv.



The following question is. Which relation exists between the EU investment funds and the respective local assignations for each country?.

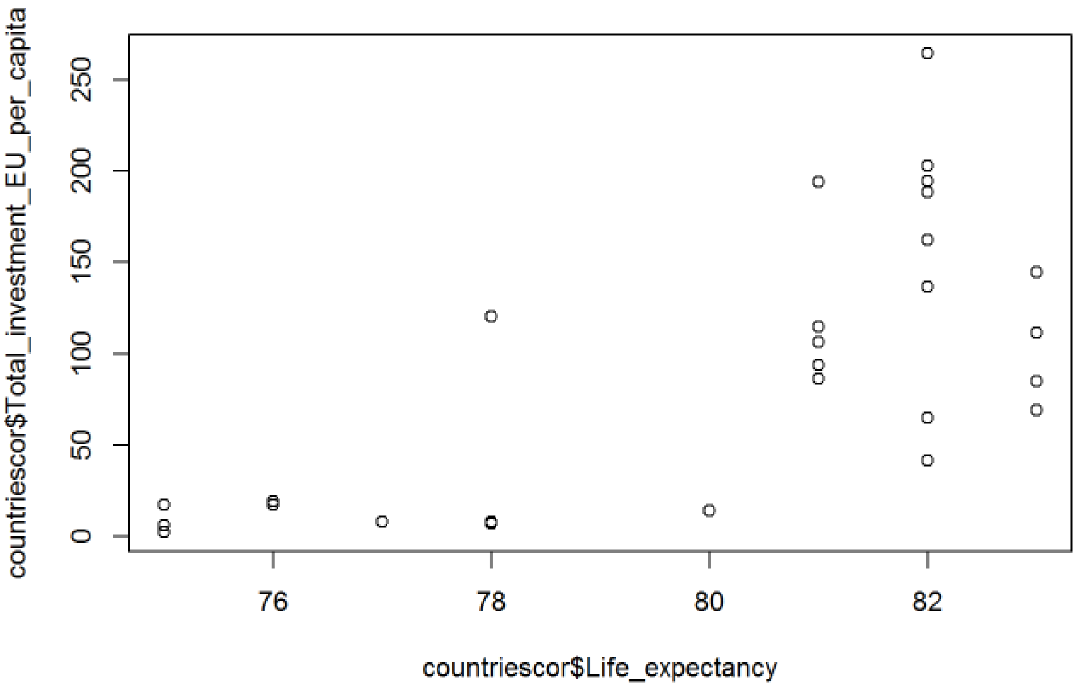
Plot 5: EU investments pc vs Local inv



Actually, the assigned funds by the EU are proportional to the Money assigned by each european State in education, Health and R&D. The correlation factor is 0.7411864.

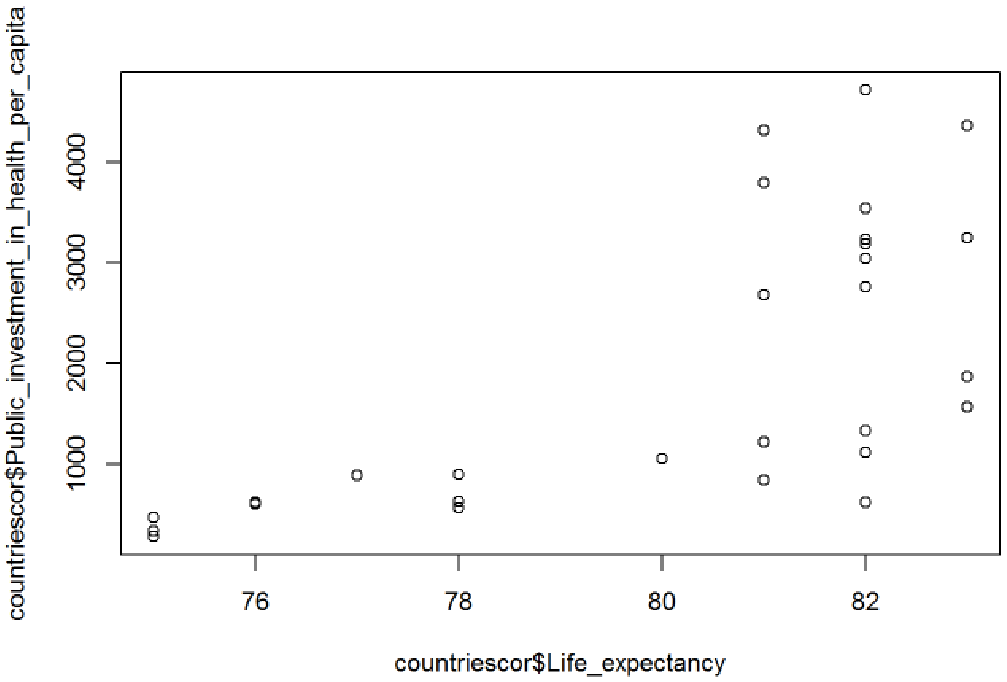
How is the Life Expectancy related with the H2020 assignments?. And how is it related with the local assignments in Health?.
Are there more factors which can impact in the Quality of Life and therefore in the Life Expectancy?.
Those are the questions we try to answer with the following plots:

Plot 6: Life expectancy vs EU investments



The correlation coefficient for this linear approach is 0.6830489. Sligthly bigger than the relation with the local assignments in Health: 0.6621424.

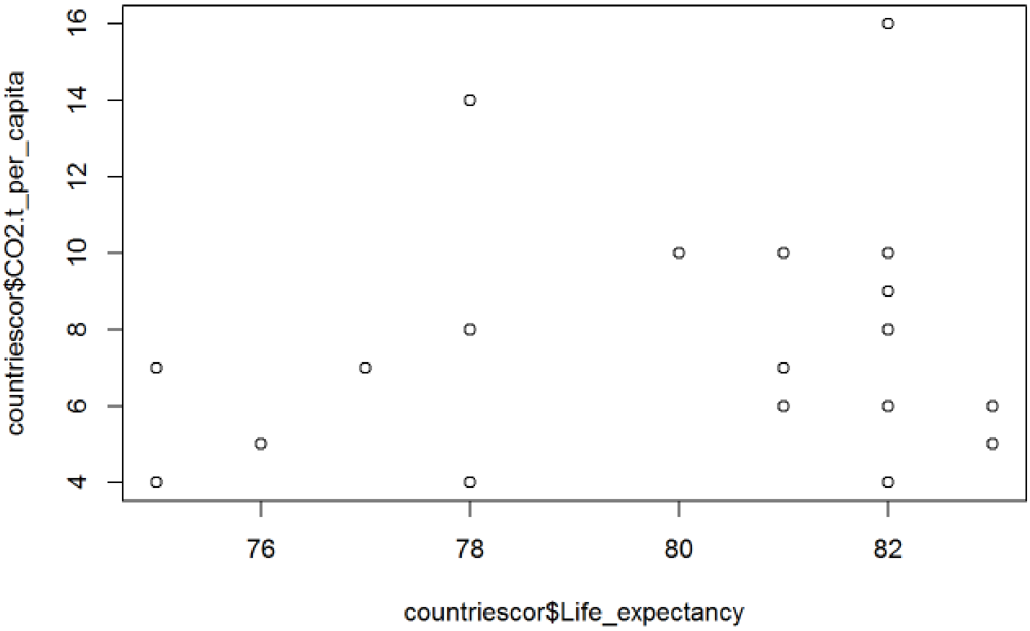
Plot 7: Life expectancy vs Local inv



In this case we can conclude that there are more parameters which can influence the Life Expectancy. Let´s see if the assignments in environmental projects can improve the Live Expectancy of the citizens.

Let´s relate this indicator with the emissions of CO2 per capita.
The approach is not linear but we observe that the countries with higher Expectancy have lower levels of CO2 in the atmosphere.

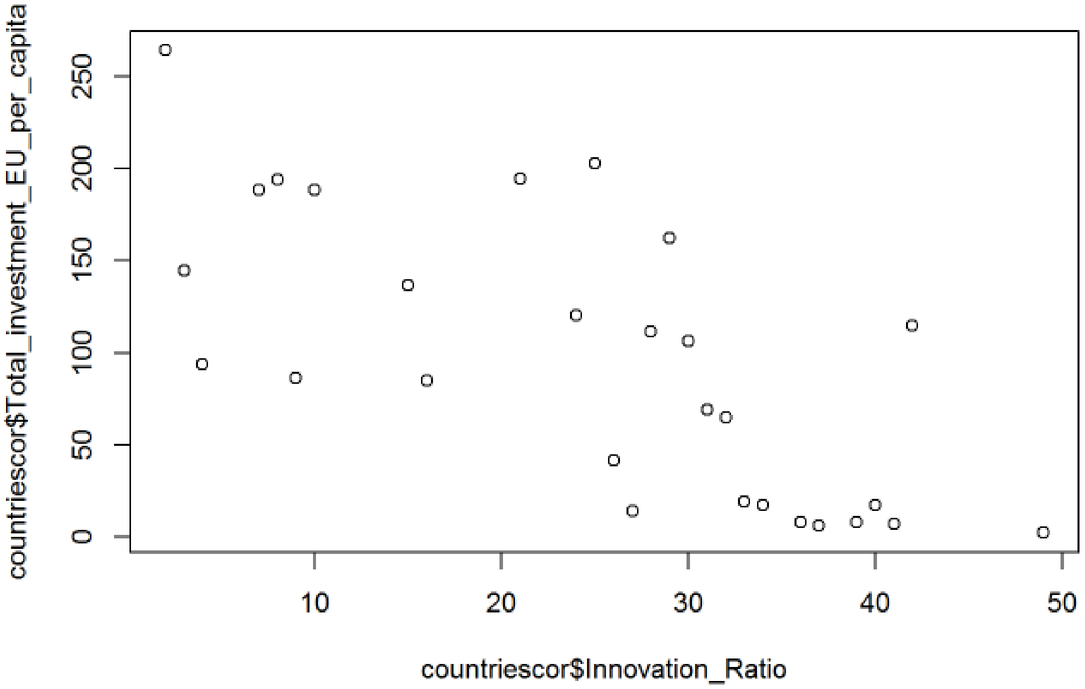
Plot 8: Life expectancy vs CO2.t pc



Industrial Leadership and Excellence Science are two of the three focus area of H2020 project. How can we measure the development level of a Country in this sense?. Innovation Ranking is a good indicator for that purpose. In this case correlation factor has to be negative. It means a Country which dedicates much Money in investigation should be in the top level (first positions in the Ranking).

Let´s relate the Innovation Ranking with the EU assignments and the local assignments in Research and Development.

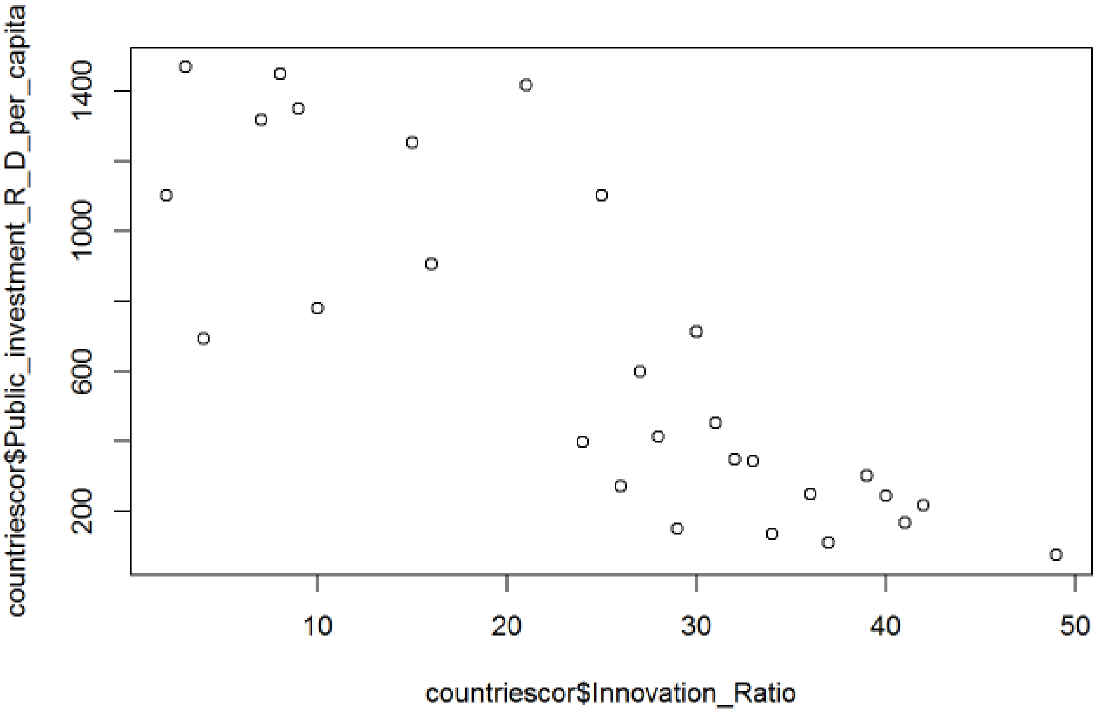
Plot 9: Innovation Ratio vs EU inv



The Relationship fits to a linear approach with correlation coefficient of -0.711801.

The correlation is still stronger with the local assignments in Research and Development.

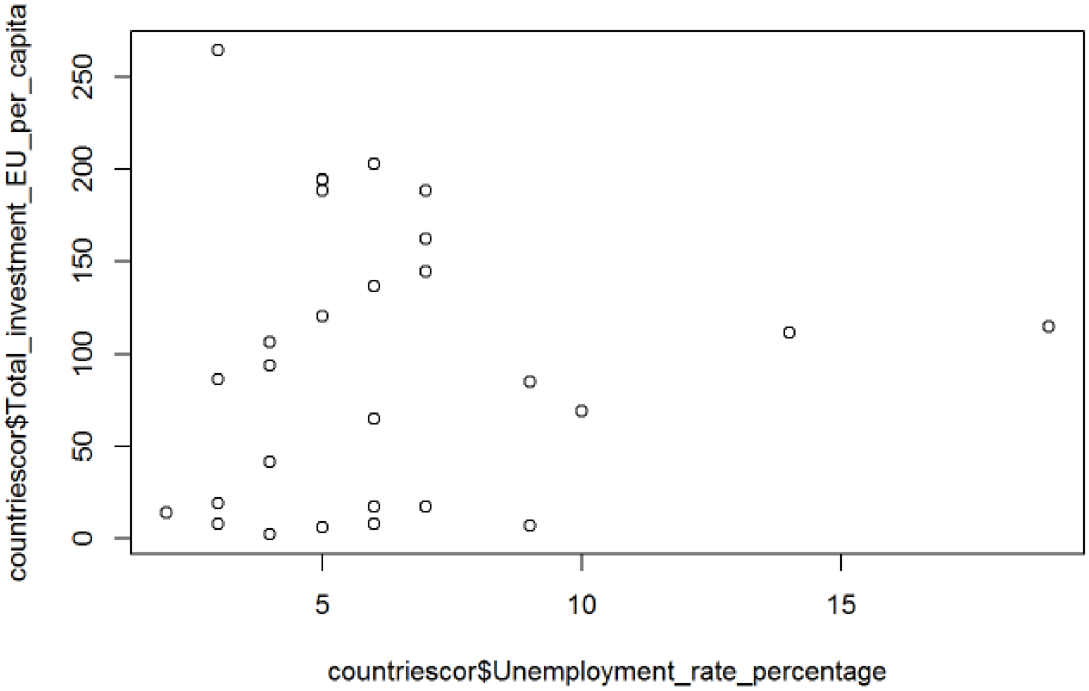
Plot 10: Innovation Ratio vs Local inv



Unemployment rate percentage is an indicator which also measures quality of life and economic development level of a State.

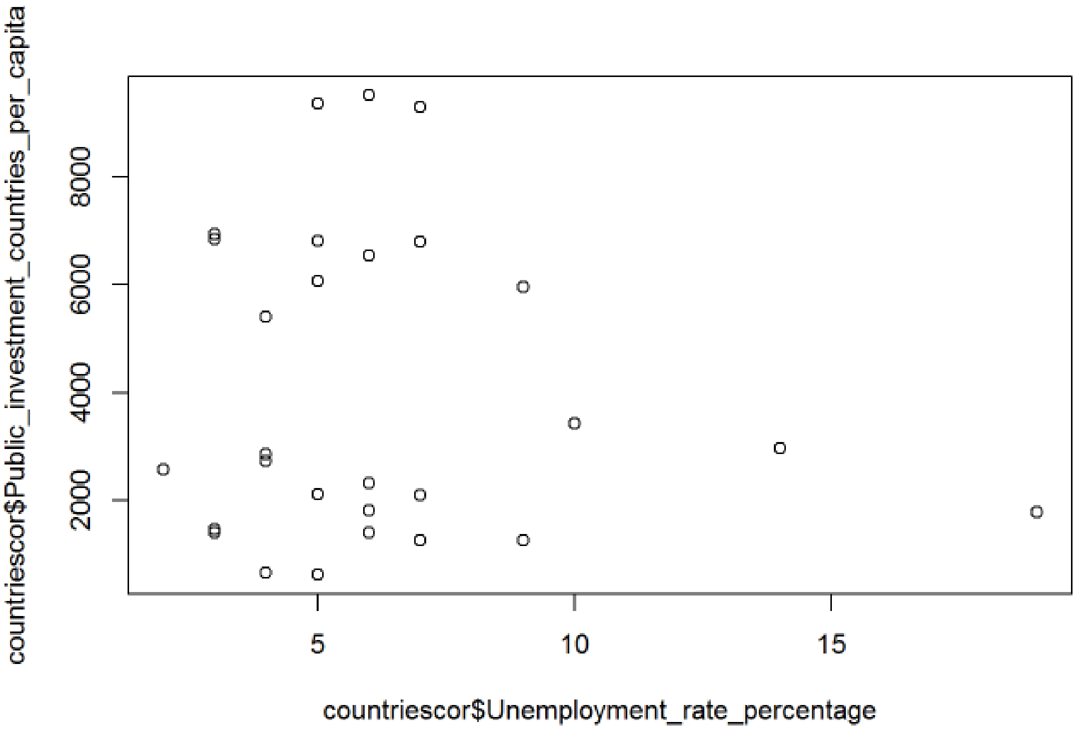
How is the unemployment rate percentage related with the CORDIS assignments?. And with the local assignments?.

Plot 11: Unemployment Rate Percentage vs total EU inv



Although the relationship is not a linear approach it can be observed that the bigger are the EU assignments the lower is the unemployment rate.

Plot 12: Unemployment Rate Percentage vs local inv



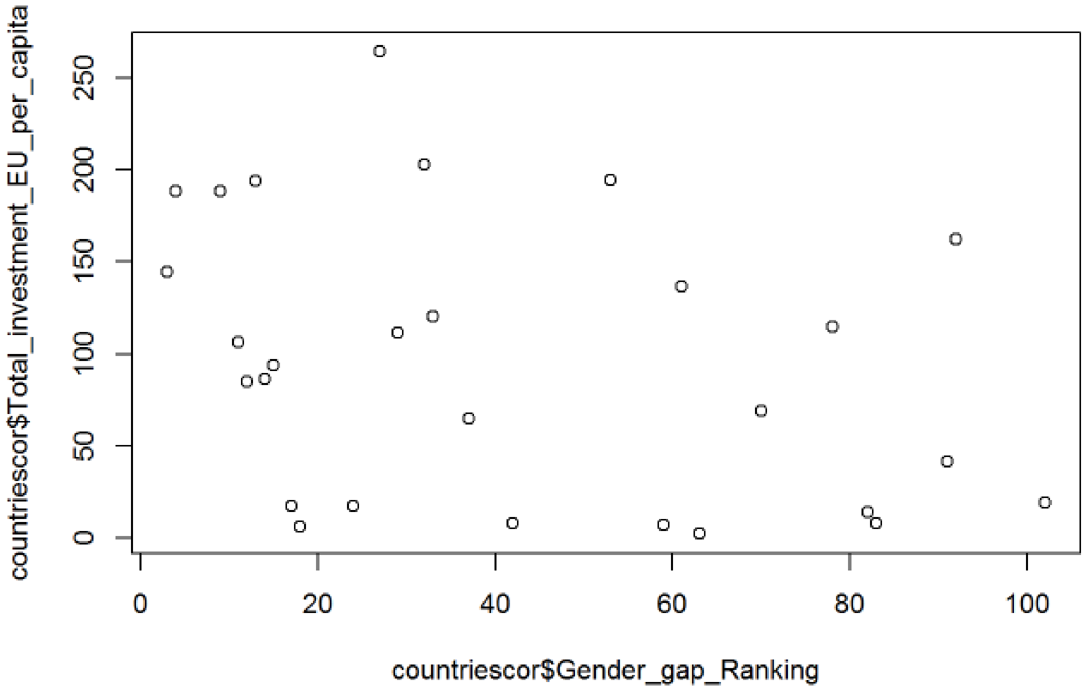
The same trend is observed regarding public local investments. Correlation is slightly stronger: -0.1126793.

To finish the analysis let´s study other social indicator which gives an approach how developed is a society: the gender gap.

Societal Challenges is the third focus of CORDIS projects. Not only Societal Challeges projects but also Excellence Science and Industrial Leadership are topics which influence the development level of a society and therefore the gender gap ranking.

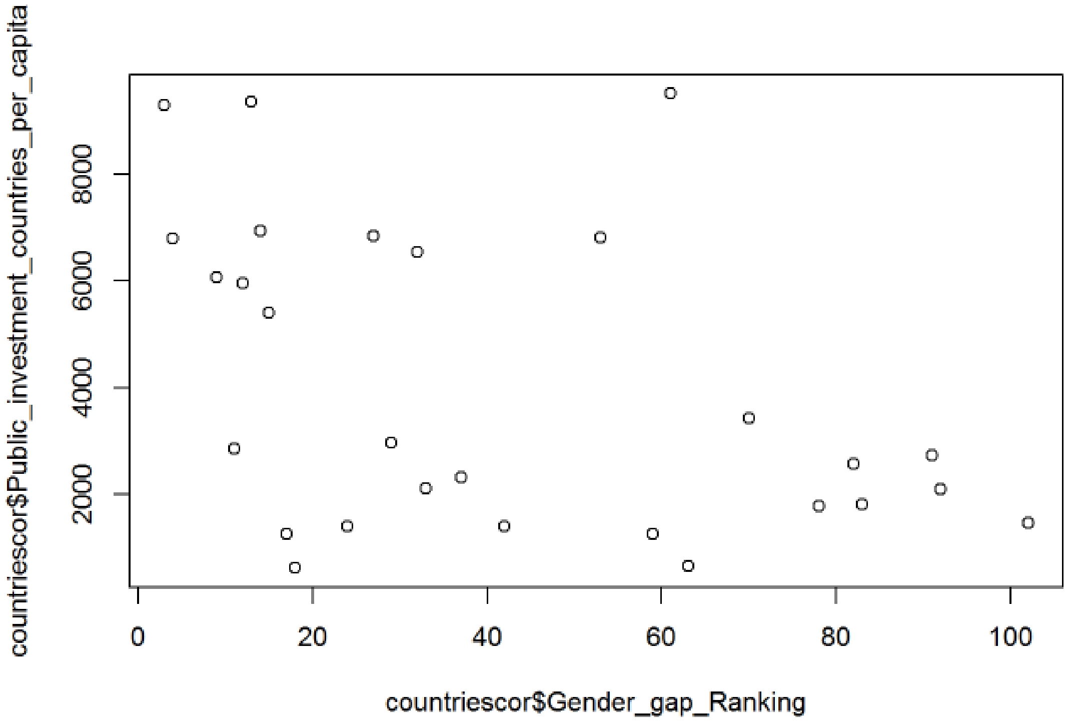
In this case the relationship should also be inverse. The more developed is a country the highest positions in the ranking.

Plot 13: Gender Gap vs EU investments



The trend is as expected although the points don´t fit a linear approach (correlation coefficient: -0.346117). Again the correlation with the local assignments is bigger -0.4418252

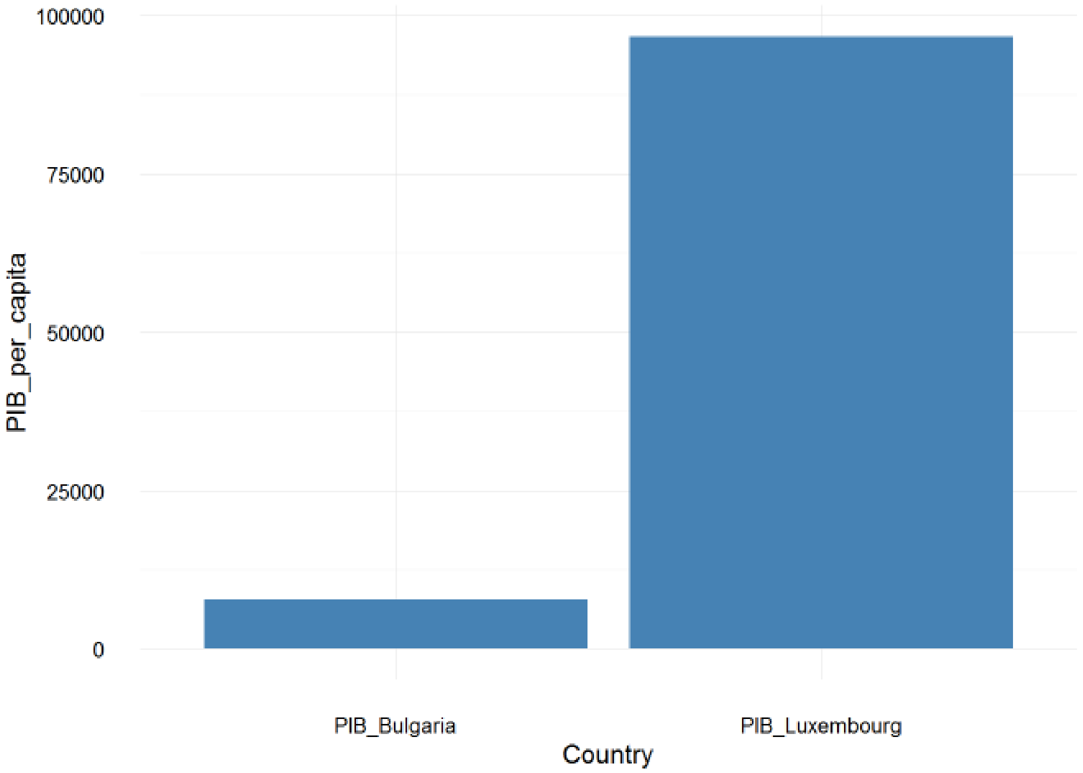
Plot 14: Gender Gap vs Local investments



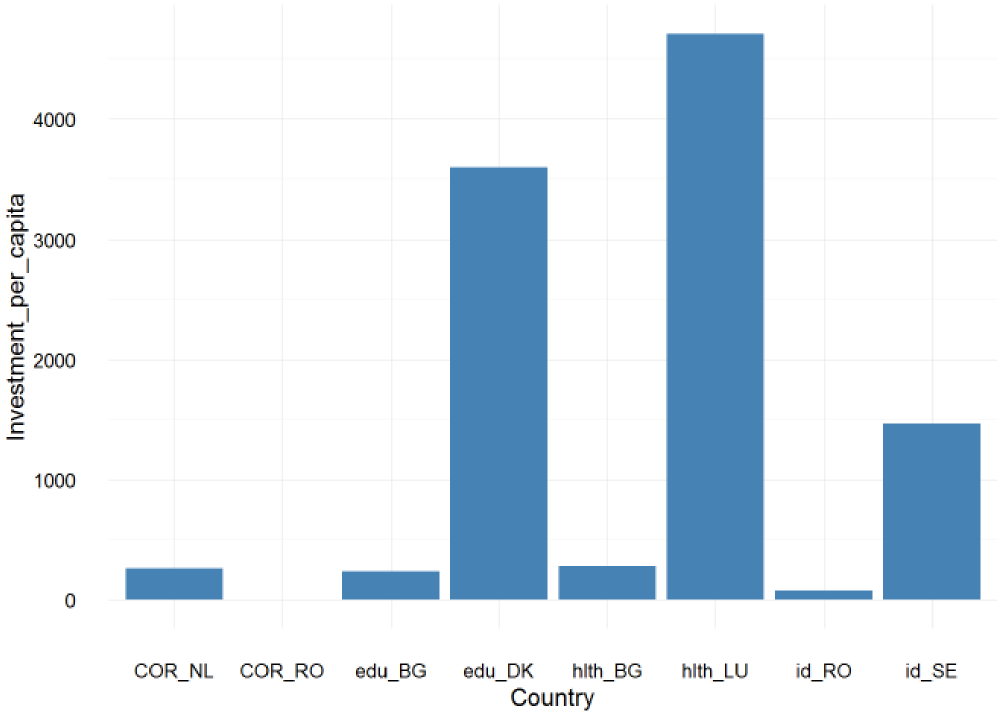
This section is concluded with some statistical analysis of the input file (countriescor.csv). The summary function in R gives as result between others the maximum and the minimum value for every column.

Which are the countries which have the maximum and the minimum GDP per capita respectively?. Let´s see it graphically.

Plot 15: Max _ Min GDP pc



Plot 16: Investments per capita



And to finish, plot 15 shows maximum and minimum values of CORDIS investments per capita, country investment in education per capita, country investment in Health per capita, country investment in Research and Development per capita (i+d).

Country codes: NL: Netherlands (the), RO: Romania, BG: Bulgaria,
DK: Denmark, LU: Luxembourg, SE: Sweden.

Files in Annex [6] and [7] have been used for the plots. They have been created with the results of the summary function applied to the source file countriescor.csv.

2.4 Clustering and second correlation analysis.

This second part has also been developed with R. Annex [10] and [11] show code executions.

As input file has been used countriescor1.csv which contains data about EU Investments splitted by Project Type: Excellence Science, Industrial Leadership and Societal Challenges.

It has also been enhanced with social parameters like HDI (Human development index), Inmigration Rate and Global Peace Ranking.

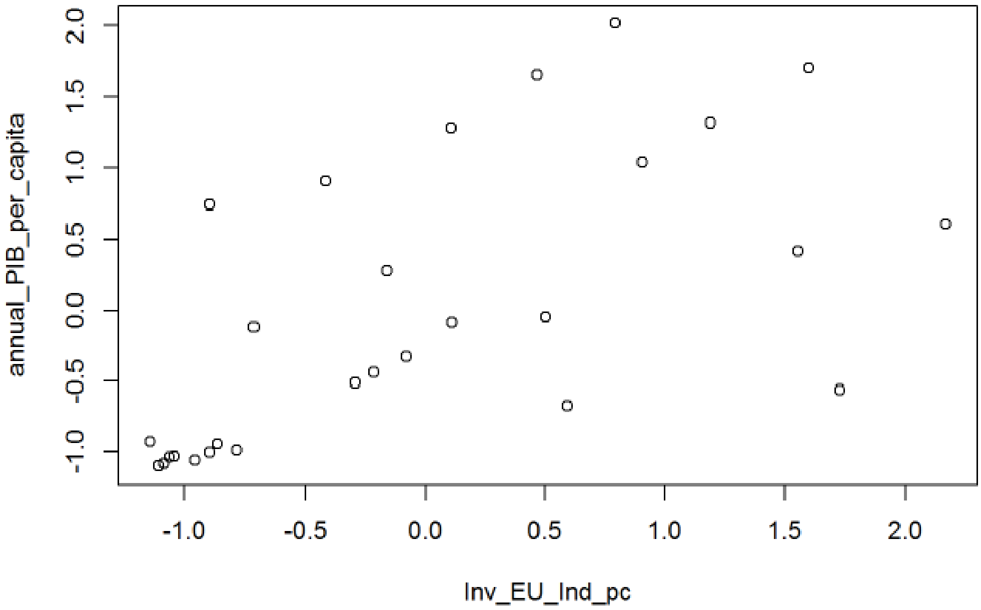
Objectives of this second analysis are:

Segmentation of the participant countries taken into account four parameters:
EU investments in Excellence Science pc, EU investments in Industrial Leadership pc, EU investments in Societal Challenges pc, GDP per capita.
Purpose is to know how the countries are grouped in order to plan future investments and collaborations.
For the clustering it has been used kmeans() function: Ref[7].

Deeper correlation analysis between EU investments in Societal Challenges pc and social parameters like Gender Gap Ranking, Global Peace Ranking, Inmigration Rate and HDI Ranking.
Purpose is to know how the specific investments in Societal Challenges projects can impact in the human development in comparison with the total EU investments.
As the first correlation it is done with the function cor: Ref[8]. And the correlation matrix has been calculated using the corrplot function: Ref[9].

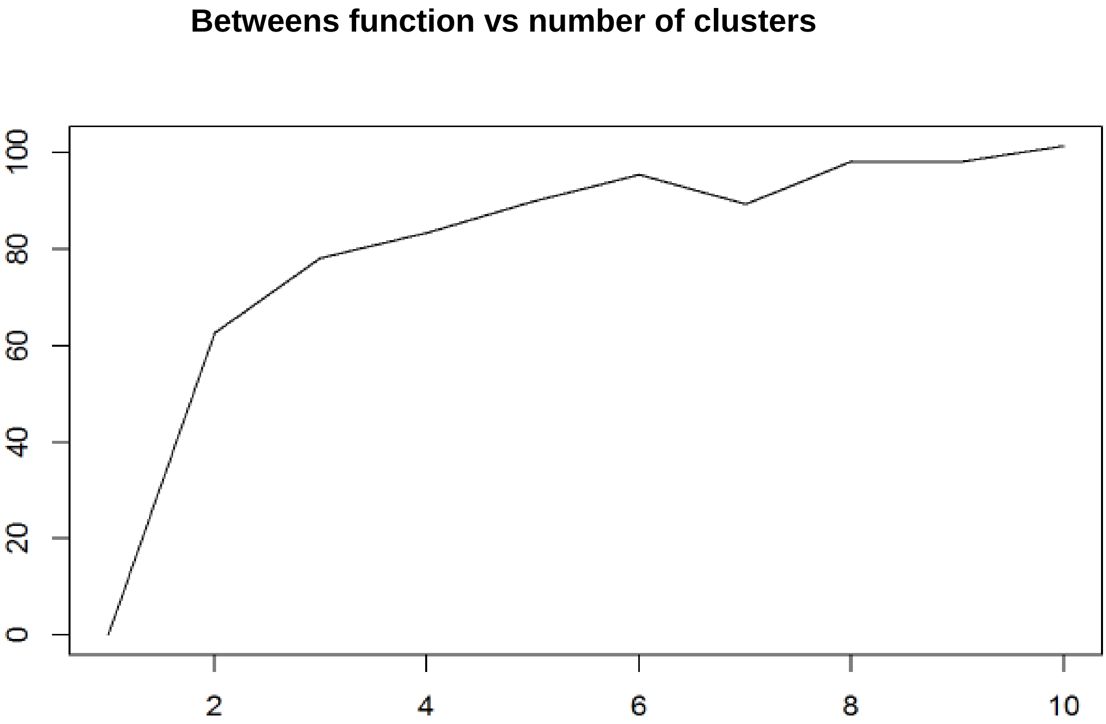
Before the clustering it is important to scale the data. This is how the normalized data look like.

Plot 17: Normalized data for clustering

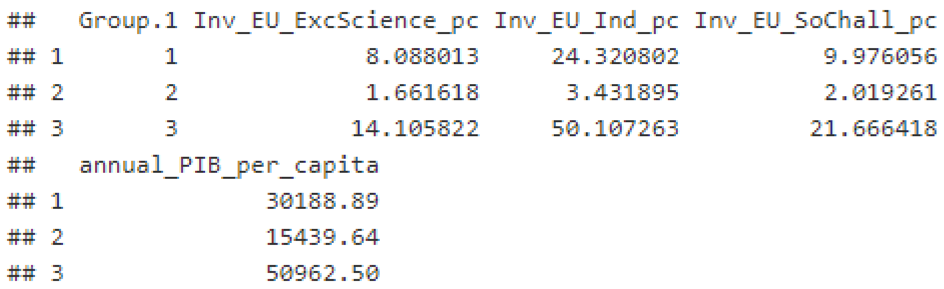


Result of the clustering gives as optimal distribution three clusters. Up from this point the betweens function becames stable.

Plot 18: Betweens function vs clusters



And those are the denormalized values for the centers:



```
##   Group.1 Inv_EU_ExcScience_pc Inv_EU_Ind_pc Inv_EU_SoChall_pc
## 1       1             8.088013     24.320802          9.976056
## 2       2             1.661618      3.431895          2.019261
## 3       3            14.105822     50.107263         21.666418
##   annual_PIB_per_capita
## 1              30188.89
## 2              15439.64
## 3              50962.50
```

The clustering fits to a Gauss Distribution. What does it mean?:

There´s a central group (Group 1) with 9 Countries which have intermediate values of all parameters.

There´s a top group (Group 3) with 8 Countries with the highest values.

And there´s a group (Group 2) of 11 Countries which presents the lowest values.

Conclusions are:

1. The reacher is the Country the higher the investments, the better is the quality of life.

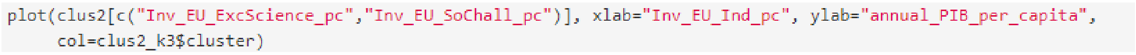
2. The highest investments are always in Industrial Leadership projects followed by Societal Challenges projects. Excellence Science receives less money than the others.

This result can arise following questions for next phase of CORDIS project:

1. Could it be possible to invest more money in the poorest countries in order to reduce the differences?.

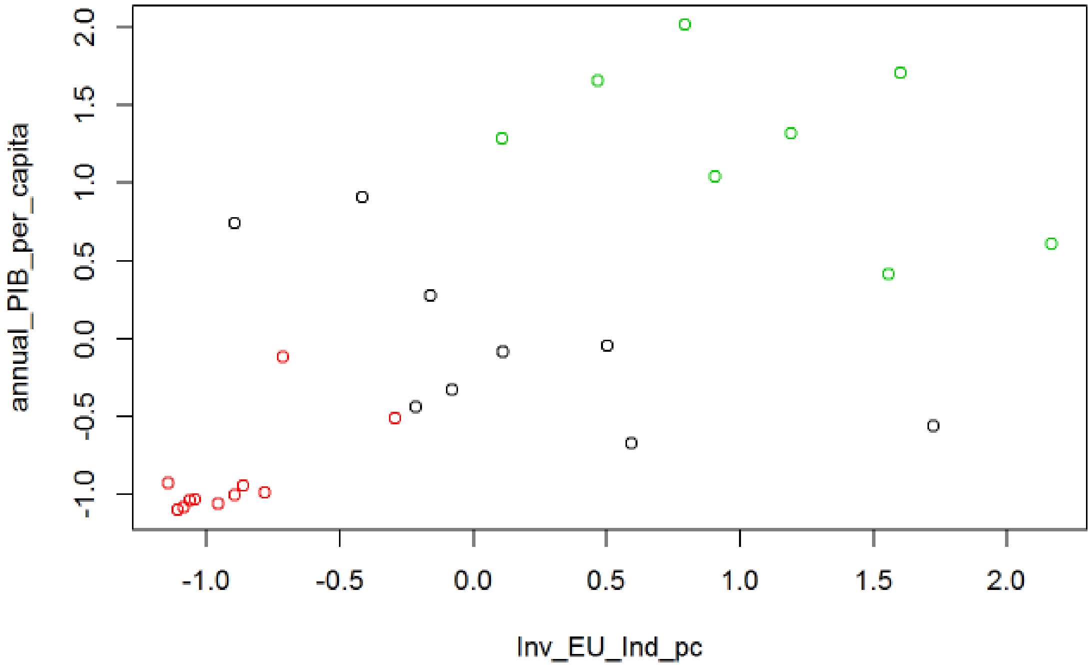
2. Could it be possible to invest more money in Excellence Science projects?.

Following plot shows the three Clusters by colour.



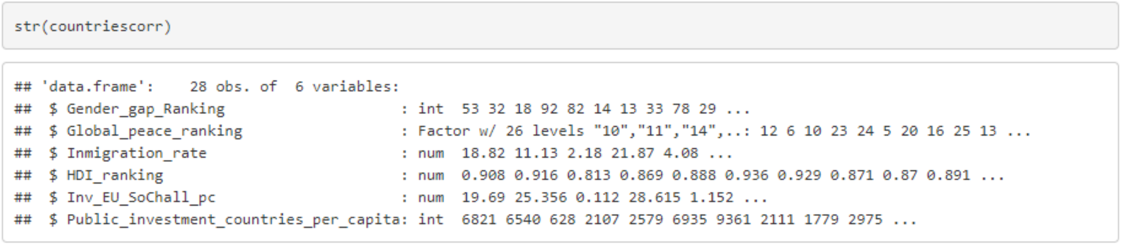
```
plot(clus2[c("Inv_EU_ExcScience_pc","Inv_EU_SoChall_pc")], xlab="Inv_EU_Ind_pc", ylab="annual_PIB_per_capita",
     col=clus2_k3$cluster)
```

Plot 19: Result of Clustering Analysis



The second conclusion of the segmentation suggests a deeper correlation analysis: how are the EU investments in Societal Challenges related with social parameters like Gender Gap Ranking, Global Peace Ranking, Inmigration Rate or HDI Ranking.

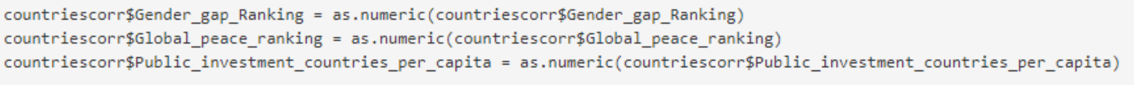
For that purpose the corrplot R Package is used. To apply this function is important that all the parameters of the data set are numeric. This is not the case:



```
str(countriescorr)
```

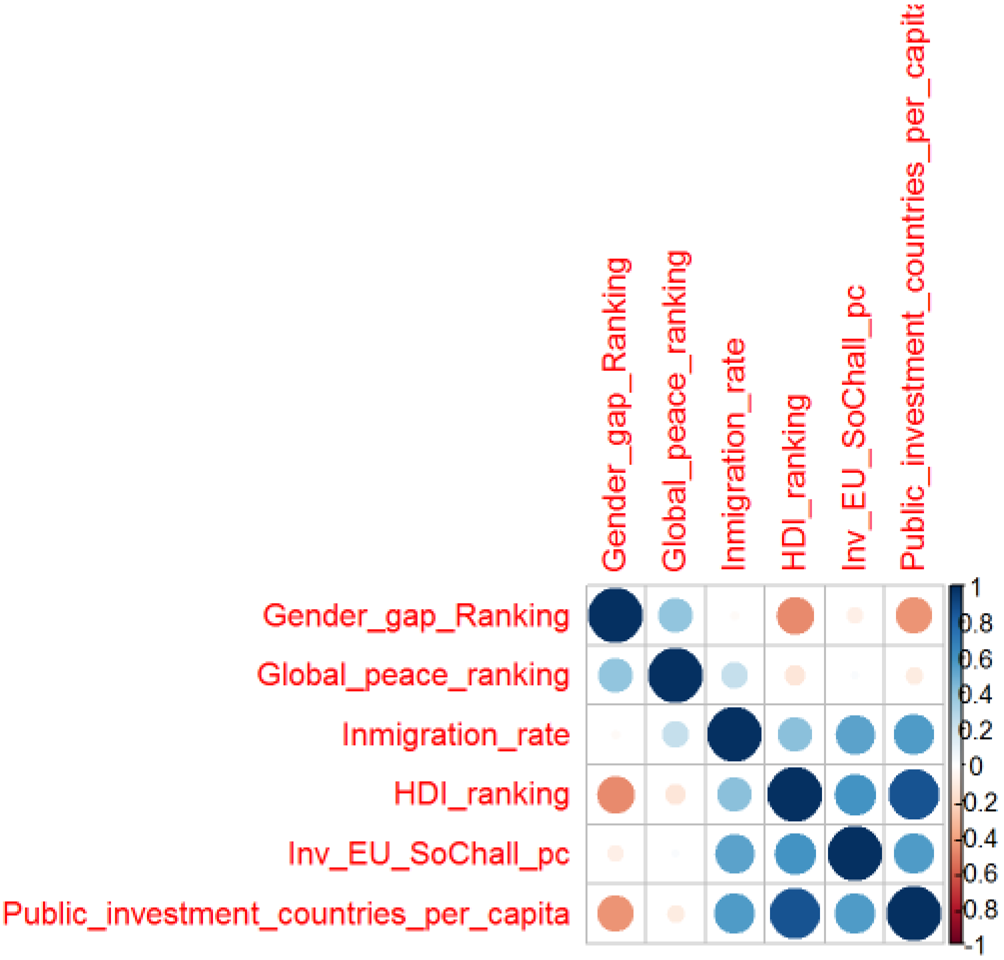
```
## 'data.frame':    28 obs. of  6 variables:
##  $ Gender_gap_Ranking                  : int  53 32 18 92 82 14 13 33 78 29 ...
##  $ Global_peace_ranking                : Factor w/ 26 levels "10","11","14",..: 12 6 10 23 24 5 20 16 25 13 ...
##  $ Inmigration_rate                    : num  18.82 11.13 2.18 21.87 4.08 ...
##  $ HDI_ranking                         : num  0.908 0.916 0.813 0.869 0.888 0.936 0.929 0.871 0.87 0.891 ...
##  $ Inv_EU_SoChall_pc                   : num  19.69 25.356 0.112 28.615 1.152 ...
##  $ Public_investment_countries_per_capita: int  6821 6540 628 2107 2579 6935 9361 2111 1779 2975 ...
```

This has to be corrected with the as.numeric function:



```
countriescorr$Gender_gap_Ranking = as.numeric(countriescorr$Gender_gap_Ranking)
countriescorr$Global_peace_ranking = as.numeric(countriescorr$Global_peace_ranking)
countriescorr$Public_investment_countries_per_capita = as.numeric(countriescorr$Public_investment_countries_per_capita)
```

Now the Dataframe is in the position to be converted in a correlation matrix. This is how the result looks like:



Plot 20: Result of second correlation

The points at the diagonal show the highest correlation coefficient (1). Obviously one variable is always correlated with itself.

Parameters like the Gender Gap Ranking are more correlated with the Local investments than with CORDIS assignments.

Gender Gap Ranking is nevertheless proportional to the Global Peace Ranking and the HDI Index (which goes from 0 to 1).

The Global Peace Ranking is also related to the Inmigration Rate and the HDI Index. It seems again that the Local investments have more impact than the Societal Challenges projects.

It seems that the funds assigned to Societal Challenges have contributed to improve the HDI Index and integrate the inmigration. Also the Local investments.

2.5 Dashboard design and development.

The Dashboard has been developed with Power BI. Annex [13].

Purpose is to create an application to be implemented in an Open Data Portal. The final users are expected to be the citizens that click the site.

Objective is to navigate through four pages which contain information about CORDIS investments, local investments and economic and social indicators.

Clicking and filtering is expected that the final user has a general overview how the funds are distributed by country and year during the project.

As well as how are they impacting in the economic and human development of the society.

This analysis has been done in a more statistical way with R. Purpose of the Dashboard is to analyse in a more visual way the concepts.

Following table summarizes the application design:

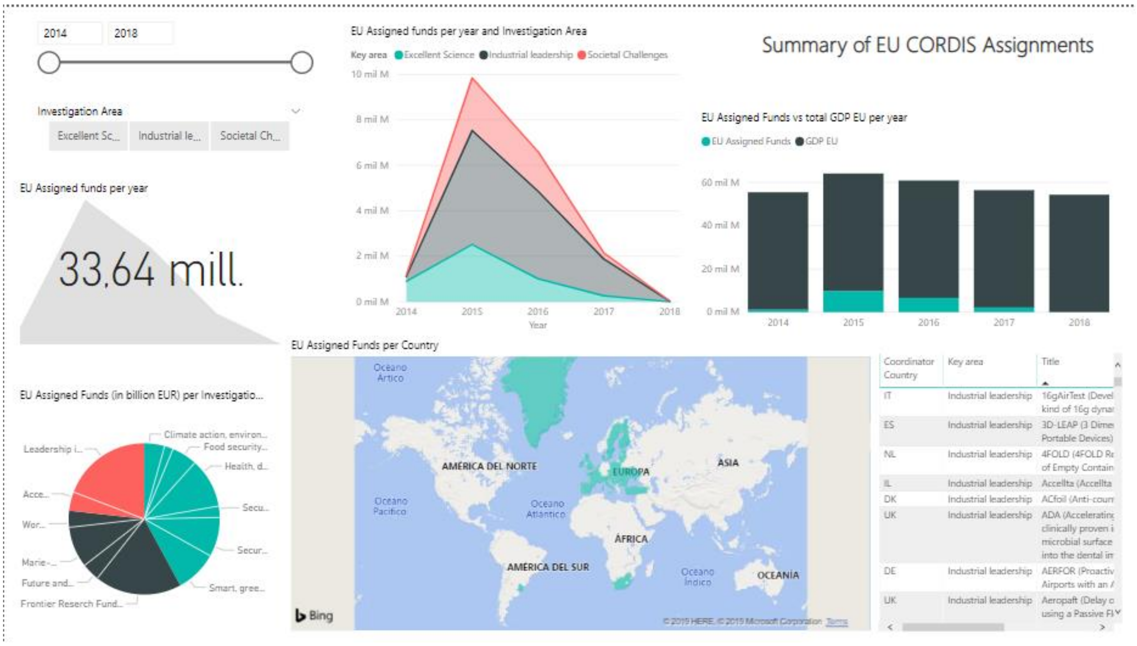
Page name	Report type	Report name	Filter	Source file
Summary	KPI	EU Assigned Funds per year	Year Investigation area	H2020reports
Summary	Stacked Area Chart	EU Assigned Funds per year and Investigation Area	Year Investigation area	H2020reports
Summary	Stacked Column Chart	EU Assigned Funds vs Total GDP EU per year	Year Investigation area	H2020reports GDP
Summary	Pie Chart	EU Assigned Funds per Investigation Area and Sub area	Year Investigation area	Horizon2020
Summary	Map	EU Assigned Funds per Country	Year Investigation area	H2020reports
Summary	Table	Table	Year Investigation area	H2020reports
H2020 Economics	Column Chart with line	Annual GDP per capita vs Average Salary and Total EU Assigned Funds pc	Country Year	Countriesev
H2020 Economics	Bubble Chart	Annual GDP per capita vs Average Salary and Total EU Assigned Funds pc	Country Year	Countriesev

H2020 Social	Treemap	Global peace ranking, HDI index, Inmigration Rate, EU Assigned Funds pc in Societal Challenges	Country	Countriescor1
H2020 Social	Map	Life Expectancy	Country	Countriescor
H2020 Social	Lines	Gender Gap Ranking vs EU Assigned Funds per capita	Country	Countriescor
EU Economics	Radial Gauge Chart	Debt per capita	Country	Countriescor1
EU Economics	Group Bar Chart	TradeBalance %GDP per Country	Country	Countriescor1
EU Economics	Radial Gauge Chart	Annual GDP per capita	Country	Countriescor1
EU Economics	Stacked Area Chart	EU investments Industrial Leadership per capita vs Local investments in R&D	Country	Countriescor1

Table 3: Dashboard design

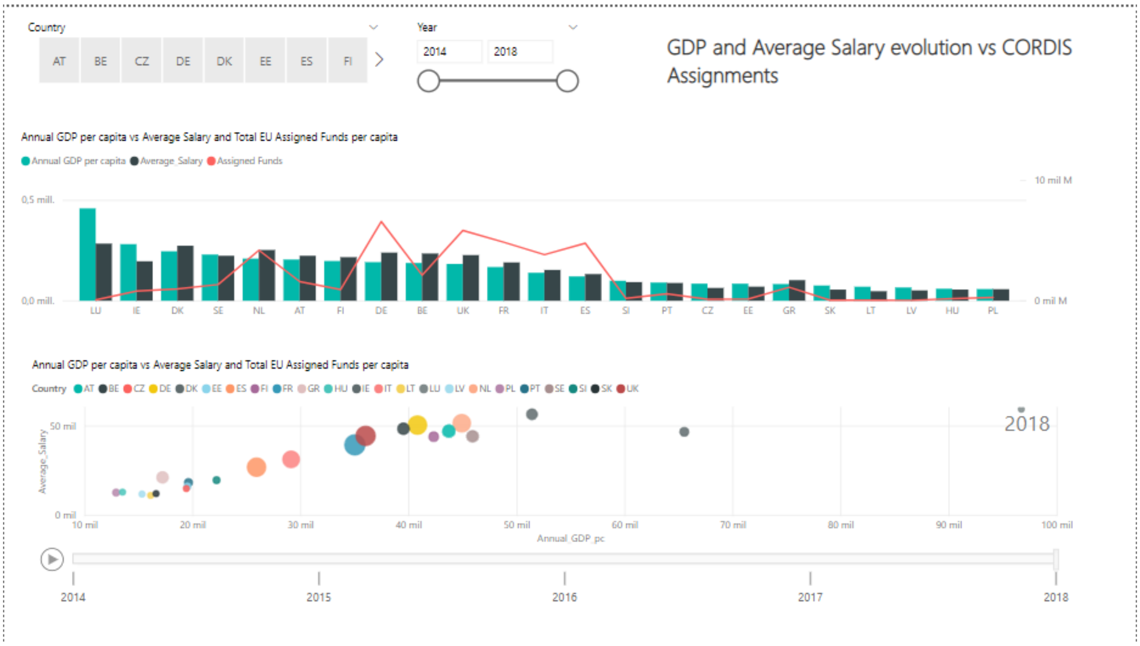
The screenshots below give a picture how each page looks like. More information and details are given in the demo video attached. Annex [14].

Page one. Summary of CORDIS Assignments.

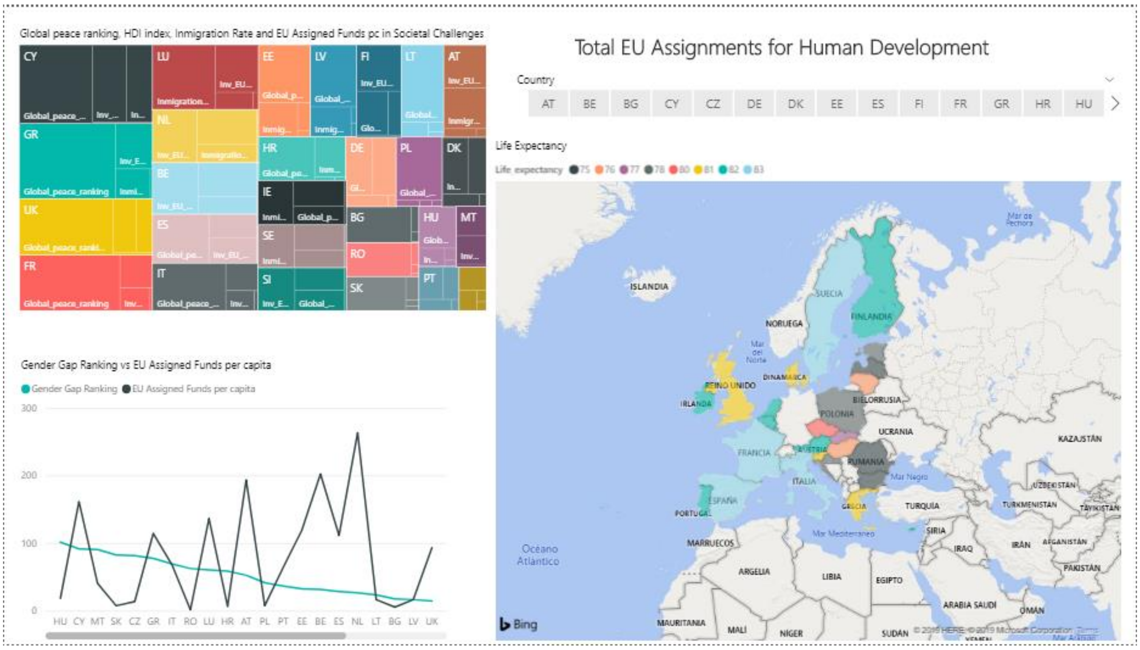


Coordinator Country	Key area	Title
IT	Industrial leadership	16gAirTest (Devel kind of 16g dynar
ES	Industrial leadership	3D-LEAP (3 Dime Portable Devices)
NL	Industrial leadership	4FOLD (4FOLD Re of Empty Contain
IL	Industrial leadership	Accellta (Accellta
DK	Industrial leadership	ACfoil (Anti-coun
UK	Industrial leadership	ADA (Accelerating clinically proven i microbial surface into the dental im
DE	Industrial leadership	AERFOR (Proactiv Airports with an /
UK	Industrial leadership	Aeropaft (Delay o using a Passive Fl

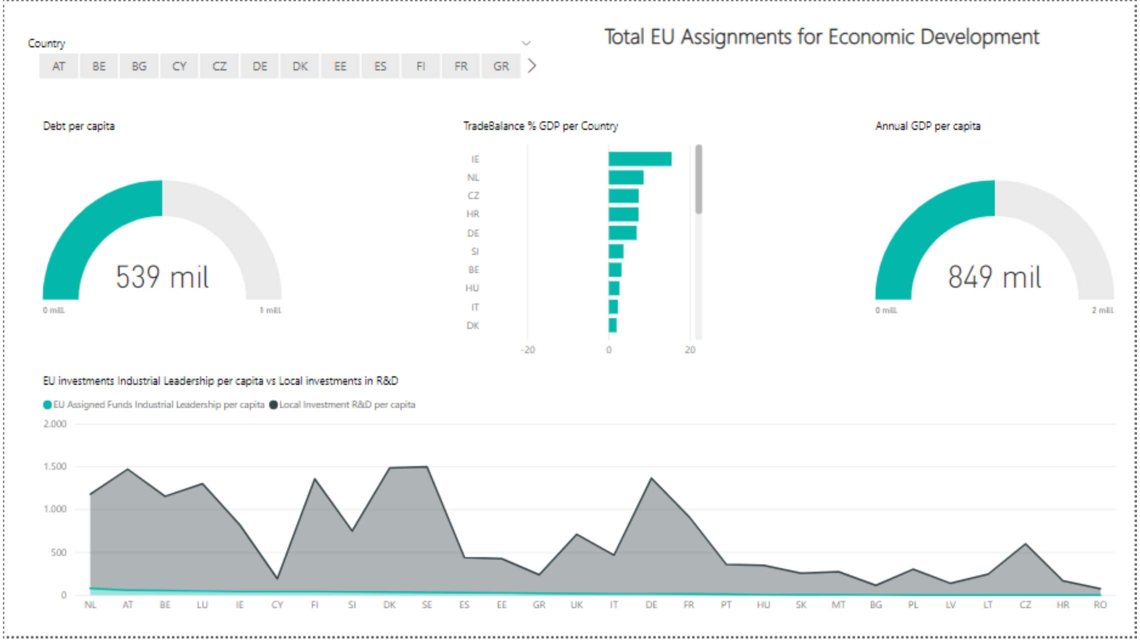
Page two. GDP and Average Salary evolution vs CORDIS Assignments.


Country
AT
BE
CZ
DE
DK
EE
ES
FI
Year
2014
2018
GDP and Average Salary evolution vs CORDIS Assignments
Annual GDP per capita vs Average Salary and Total EU Assigned Funds per capita
● Annual GDP per capita ● Average_Salary ● Assigned Funds
0,5 mill.
0,0 mill.
10 mil M
0 mil M
LU
IE
DK
SE
NL
AT
FI
DE
BE
UK
FR
IT
ES
SI
PT
CZ
EE
GR
SK
LT
LV
HU
PL
Annual GDP per capita vs Average Salary and Total EU Assigned Funds per capita
Country ● AT ● BE ● CZ ● DE ● DK ● EE ● ES ● FI ● FR ● GR ● HU ● IE ● IT ● LT ● LU ● LV ● NL ● PL ● PT ● SE ● SI ● SK ● UK
Average_Salary
50 mil
0 mil
10 mil
20 mil
30 mil
40 mil
50 mil
60 mil
70 mil
80 mil
90 mil
100 mil
Annual_GDP_pc
2018
2014
2015
2016
2017
2018

Page 3. Total EU Assignments for Human Development.



Page 4. Total EU Assignments for Economic Development.


Country
AT
BE
BG
CY
CZ
DE
DK
EE
ES
FI
FR
GR
Total EU Assignments for Economic Development
Debt per capita
539 mil
0 mill.
1 mill.
TradeBalance % GDP per Country
IE
NL
CZ
HR
DE
SI
BE
HU
IT
DK
-20
0
20
Annual GDP per capita
849 mil
0 mill.
2 mill.
EU investments Industrial Leadership per capita vs Local investments in R&D
● EU Assigned Funds Industrial Leadership per capita ● Local Investment R&D per capita
2.000
1.500
1.000
500
0
NL AT BE LU IE CY FI SI DK SE ES EE GR UK IT DE FR PT HU SK MT BG PL LV LT CZ HR RO

3. State of the art.

A Dashboard for the HORIZON project was already developed. Reference Guide can be found in Ref[10] and was published in November 2018 by the European Comission.

The Horizon Dashboard is an intuitive and interactive reporting platform, composed of a set of sheets that allows series of views to discover and filter the Horizon 2020 Data.

The aim of the Horizon Dashboard is to facilitate data sharing, providing public access to real-time programme data in an easy, flexible and user-friendly manner.

This report goes deeper in the granularity of research topics, countries, regions, organisations, sectors and even individual projects and beneficiaries.

Purpose is to allow the participants to prepare future reasearch and innovation projects through an easier identification of possible research topics and partners.

Other similar solutions can be found in the context of EU projects for example European Social Funds in Holland. [11]

Contribution of this TFM is to enhance the scope of the analysis relating the project assignments with macroeconomic and human indicators.

Analytics can provide a deeper view how are they correlated and can help the organizers to understand how are the countries grouped.

The DataSet has been enhanced with macroeconomic and human indicators which can measure the impact of CORDIS investments in the economic and human development of the involved countries.

Moreover, the DataSet contains also the assignments done by the local governments in Education, Science and R&D.

The Dashboard of this TFM is oriented to citizens which can navigate through four pages and see how the funds are distributed by tematic, country and year.

They can compare the funds with economic indicators like GDP, Average Salary, Unemployment Rate Percentage or local assignments.

This work includes an statistical analysis with R which goes deeper into the correlations and classification topics. This analysis can be useful to plan future reasearch projects as well as collaborations and synergies.

4. Conclusions and self evaluation

In this project has been made an analysis comparing investments in investigation projects with macroeconomic parameters.

There have been found relations between the funds invested and economic indicators like GDP, Average Salary, Unemployment rate percentage.

In the same way it has been tested how investments in Societal Challenges, Excellence Science and Industrial Leadership can also impact the human development of a society. We have seen correlations between assigned funds and indicators like Gender Gap, Human Development Index or Global Peace Ranking.

How can contribute the investigation to the economic and human development of a society?.

Or how are the assignments depending on the economic model of a Country?. It has been seen that the Countries with higher positive Trade Commerce Balance invest more money in Industrial Leadership or R&D projects.

Those are some of the lessons lernt:

The relations have been measured in an statistical way with R and have been visualized in a Dashboard. The Dashboard is designed to be implemented in an open portal. It supposes an opportunity for the citizens to see how is the money invested and which are the results which can impact their lifes.

Also a segmentation analysis has been done. EU countries have been grouped by four parameters: EU investments in each investigation area (Excellence Science, Industrial Leadership or Societal Challenges) and GDP per capita.

Results are how expected, the richer is a country the better the level and quality of life. Countries segmentation fit to a Gauss distribution which confirms the intuition.

But the results suggest new lines for future phases of the CORDIS project:

Could it be possible to invest more money in the poorest countries or share resources in order to reduce the differences?.

In this sense the objectives of the planification have been achieved.

A similar analysis can be done in the future with other projects to see relations between them and to have a more general overview.

Moreover, other type of projects can also be analysed. For example investments in culture. How can be related with the human and economic development?.

Self evaluation:

An initial planification was suggested but it was changing along the semester as well as the project scope.

Details are clarified in the Table 1 (Project Risks) and in Section 1.4 Project Planning.

5. Glossary

H2020	Horizon 2020
CORDIS	EU research projects under Horizon 2020
EU	European Union
GDP	Gross domestic product
R&D	Research and development
PC	Per capita
HDI	Human development index
CO2t	CO2 emissions (metric tons per capita)

Table 4: Glossary

6. Bibliography

[1] 06/2019
https://ec.europa.eu/programmes/horizon2020/sites/horizon2020/files/H2020_inBrief_EN_FinalBAT.pdf

[2] 06/2019
https://www.agilealliance.org/wp-content/uploads/2016/01/Agile-Risk-Management-Agile-2012.pdf

[3] Josep Curto Díaz. Materiales UOC fundamentos de Inteligencia de Negocio. Capítulo 6: Diseño de cuadros de mando.

[4] 06/2019
http://data.europa.eu/euodp/es/data/dataset/cordisH2020projects

[5] 06/2019
https://datosmacro.expansion.com/paises/grupos/union-europea

[6] 06/2019
https://es.wikipedia.org/wiki/Anexo:Pa%C3%ADses_por_el_gasto_en_investigaci%C3%B3n_y_desarrollo_I%2BD

[7] 06/2019
https://www.rdocumentation.org/packages/Rcmdr/versions/2.0-4/topics/KMeans

[8] 06/2019
https://www.rdocumentation.org/packages/stats/versions/3.6.0/topics/cor

[9] 06/2019
https://cran.r-project.org/web/packages/corrplot/vignettes/corrplot-intro.html

[10] 06/2019
https://ec.europa.eu/info/funding-tenders/opportunities/docs/manuals/horizon_dashboard_quick_guide.pdf

[11] 06/2019
https://www.cbs.nl/nl-nl/maatwerk/2018/42/europees-sociaal-fonds-in-nederland-2014-2017

7. Annex

[1] Power BI Data Model. WinRAR-ZIP-Archiv.
[2] PEC3_Codigo_R_dm R Archiv.
[3] PEC3_Codigo_R_dm2 R Archiv.
[4] PEC3_Codigo_R_dm3 R Archiv.
[5] countries.csv
[6] countries_sts.csv
[7] countries_sts_1.csv
[8] countries1.csv
[9] PEC3_Codigo_R_cor html Archiv
[10] PEC3_Codigo_R_clus R Archiv
[11] PEC3_Codigo_R_clus html Archiv
[12] PEC3_Codigo_R_cor R Archiv
[13] PowerBI. PBIX file.
[14] Dashboard demo. Video file.

8. DGPR

In this project are used Open Source Tools as well as Open Data source file.

Since DGPR perspective there’s nothing to be taken into account