M1 EE
2023/2024

Introduction to R

Perceived usefulness of work by employees: The European case

Dan Sellem
&
Quentin Le Boulch

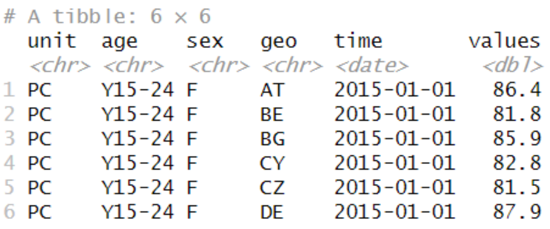
Objectif

From a general Database to a reduced database maximizing information

The main goal of this article is to demonstrate to our subscribers how the open-source software R works. This program enables the processing, analysis and extraction of relevant information from a large dataset to empirically represent the functioning of our economies. Starting with a database obtained from Eurostat website, we will gradually reduce the superfluous information it contains to ultimately construct a two-variable database that maximizes information quality.

The story line of our application is job meaningful or uesful work. More specially, we will focus on potential variation in job usefulness perception among European countries (UK include) by employees. The sample consists of individual workers aged between 25 and 64, and we focus on 2015.

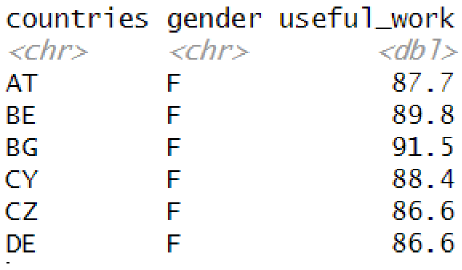
To carry out the database cleanup, we will work in three steps: firstly we'll describe the initial database to understand its major trends and highlight important points. Secondly, we will identify statistically significant information using inference tools. Finally, we will construct our reduced data base and provide some commentary.



```
# A tibble: 6 × 6
  unit  age     sex   geo   time       values
  <chr> <chr>   <chr> <chr> <date>      <dbl>
1 PC    Y15-24 F     AT    2015-01-01   86.4
2 PC    Y15-24 F     BE    2015-01-01   81.8
3 PC    Y15-24 F     BG    2015-01-01   85.9
4 PC    Y15-24 F     CY    2015-01-01   82.8
5 PC    Y15-24 F     CZ    2015-01-01   81.5
6 PC    Y15-24 F     DE    2015-01-01   87.9
```

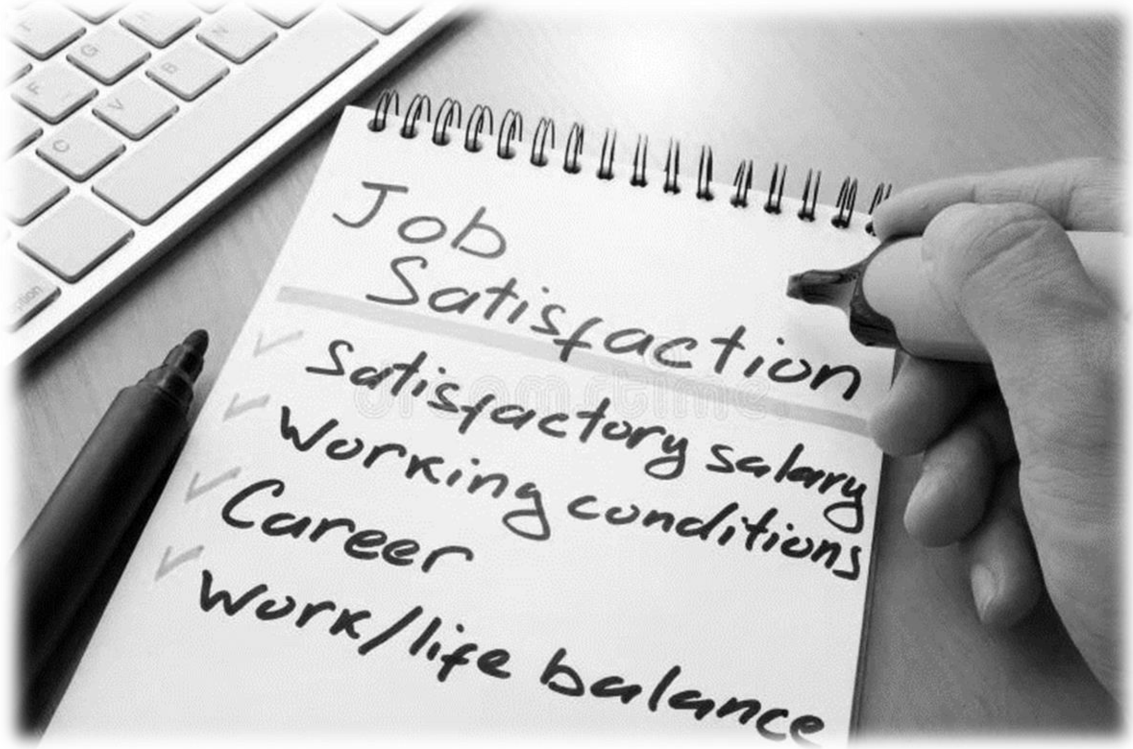
Document 1 : This screen picture represents the firsts 6 lines of the initial database extracted from the Eurostat website.

Titre : Employed persons thinking that they do useful work by sex and age (source: Eurofound)
Code : qoe_ewcs_7b3



```
countries gender useful_work
<chr>     <chr>        <dbl>
AT        F             87.7
BE        F             89.8
BG        F             91.5
CY        F             88.4
CZ        F             86.6
DE        F             86.6
```

Document 2 : The first 6 rows of the database after arrangement and reorganization.



The **useful work rate or percentage** is a subjective value represented as a ratio ranging from 0% to 100 %. Employees are surveyed in their workplace and respond to a set of questions about their feeling on their jobs, if they think that the job they do is useful. It's all about perceived usefulness of work by employees. These responses are then synthesized with **a meaningful work rate/percentage**, **usefulness work rate/percentage, perceived usefulness of work** . Both are the same it is just to be clear about the terminology. We talk about the SAME idea.

Database description

Perceived usefulness of work by country and gender

The initial database (wich we will refer to as "base"), contains 1080 individuals/observations. The first variable « unit » informs us about the nature of the "value" variable : in our case, it represents a percentage. This variable does not contain any significatant information and will be removed at a later stage. The second variable "age" provvides information abouut the age of the surveyed people. In a practical approach and considering that the labor rights differs from one country to another, we will select only people aged between 25 and 64 years old. The "sexe" variable is an indicator variable with the character "F" (female), "M" (male)) and "T" (total): since we aim to understand the differentiated effect of useful work % by gender, we will exclude "T". The « geo » variable indicates the country code where the survey was conducted: we will later transform these codes into excat names of the countries using R function. The 4th study variable « Time » has has three options to choose from: 2005, 2010, or 2015. Our report focuses only on the year 2015. Finally, the last variable, "values"' indicates the useful work rate as a percentage (Document 1).

The database consists of 1080 observations: for the 30 country entities, the useful work rate (reminder we talk about perceived usefulness of work by employees) is provided for 3 periods, 4 age categories, and 3 gender categories. By multiplying the entities by the number of categories and options, we obtain our 1080 observations.

The country entities include "UE27" and "EU28"' which correspond to the average useful work % in European Union countries. Because we are interested in significant differences between countries and to avoid biasing our results, we intentionally remove these two categories. It is important to note that our study focuses on the year 2015, at a time when the United Kingdom was still a member of the European Union.

After grouping the useful work rates by country and selecting the data for the year 2015, we obtain a database with 3 variables, as represented above **(Document 2).**

This transformed database contains 3 variables for 29 countries. For better presentation, we have intentionally renamed the variable names. The variables "Unit", "Age" and "Time" are not included because we know that we are only interested in the useful work rate for EU member countries in the year 2015 and for workers aged 25 to 64. Lastly, it should be noted that the "Gender" variable now has only 2 categories: "F" or "M." This transformed database consists of 60 observations, which means 2 observations per country for the 2 gender categories.

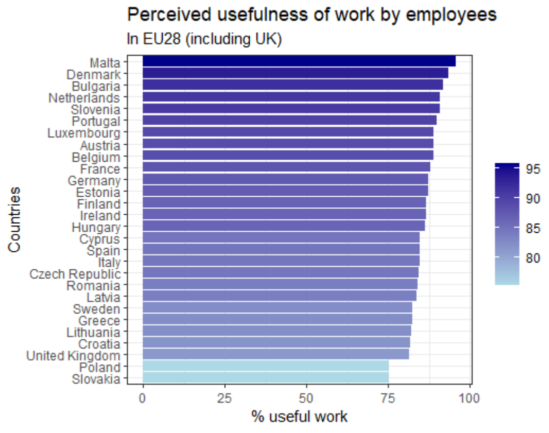
Now, we can outline the initial trends of this new database. Overall, the average **useful work rate** (perceived usefulness of work by employees) – for all countries and both genders combined – stands at 85.95%, which appears to be a good ratio. The median is close to the mean: as many people attribute a usefulness for their jobs level lower as they do higher than 86.45%. The intercountry dispersion is measured by the corrected standard deviation, which is 5.2.

Let's now compare these overall results between countries and gender:

- Firstly, we calculate the average useful work for all EU member countries by gender: this information is contained in the 'mean_gender' dataframe. In 2015, the average perceived usefulness of work for women was 86.8%, compared to 58% for men. We will test the significance of the equality of means in the second part, but we can already observe that there is no significant difference between the genders.
- Then, we calculate the average perceived usefulness of work by country - regardless of gender. This information is available through the 'mean_countries' dataframe, and **Document 3** represents the rates for each country.

Document 3 allows for the separation of countries into three more or less distinct groups: the first group (sky blue) consists of 2 countries with an average useful work rate of 75%, a second, more heterogeneous group (light blue) composed of around twenty countries – including France, Germany, and the United Kingdom – with useful work rates around 80%. Finally, the podium, composed of Malta and Denmark, forms the last group (dark blue) with useful work rate exceeding 90%.

Document 3: useful work rate by country



In more detail, the country with the lowest useful work rate is Slovakia (75.3%), while Malta has the highest with 95.75%.

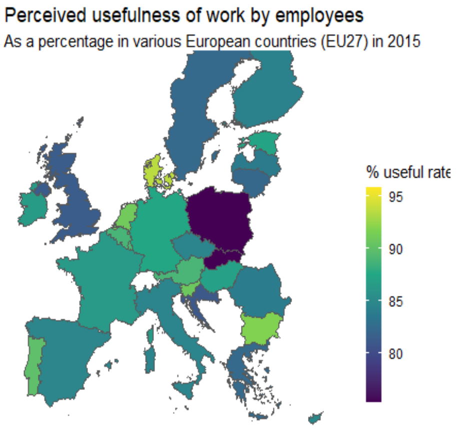
Trend-wise, we observe that it is not necessarily the countries with more developed economies that occupy the top positions. Other institutional factors (labor market composition, labor laws, benefits, and personalized support) as well as natural factors (climate, quality of life, green spaces) can work in favor or against a country.

A specific point of detail should be given to Denmark, which has adopted a 'flexicurity' employment policy, as have many Scandinavian countries. The ease of layoffs that is characteristic of these systems does not appear to be unfavorable to perceived usefulness of work. In fact, other aspects such as support in returning to employment or significant benefits (the 'security' part) offset the negative aspects of labor market flexibility.

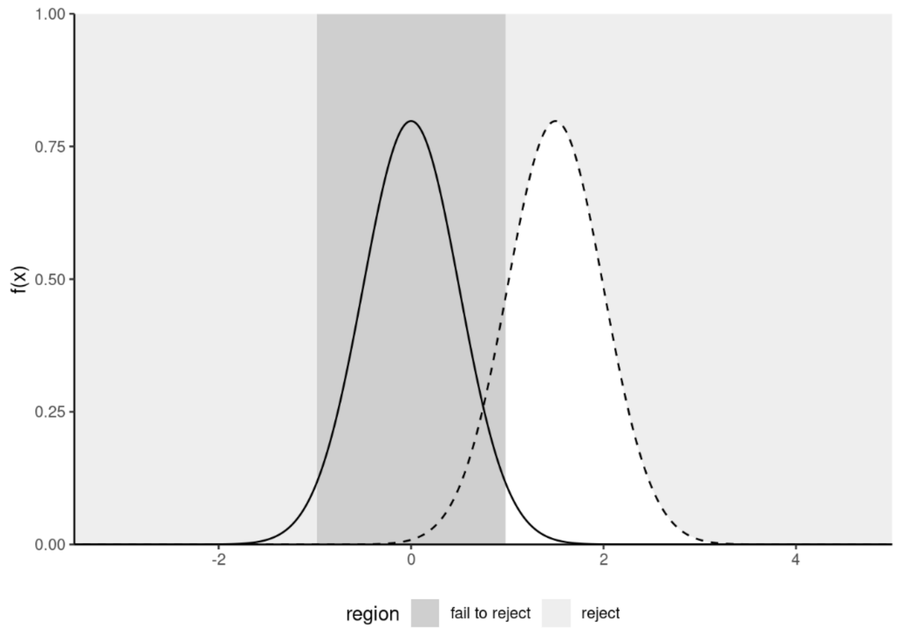
Through this first part, we have transformed our database to retain only variables containing relevant information.

We have demonstrated that there does not appear to be a difference in perception between genders. However, there are three distinct groups of countries within the European Union in terms of job satisfaction. Finally, job satisfaction seems to be based on determinants that are more institutional and structural than cyclical.

Document 4 provides a better representation of the differences in rates between countries. A detailed explanation of the construction of this graph is offered in the 'Technical Point' section on page 3."


Perceived usefulness of work by employees
As a percentage in various European countries (EU27) in 2015
% useful rate
95
90
85
80

Data: Eurostat qoe_ewcs_7b3

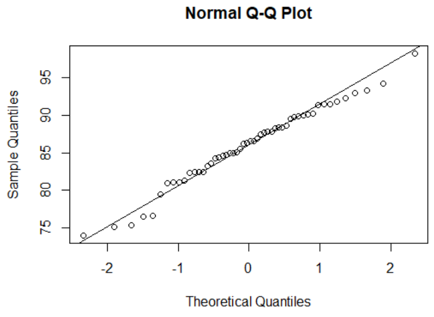


In order to test the significance of useful work % by country and gender, we need to ensure in advance that the "useful_work" variable follows a normal distribution. The first step in this section will be to demonstrate the normality of this variable. In the second step, we will perform a test of equality of useful means between men and women. Finally, we will narrow down our database to a few countries where the significance of the useful work rate is confirmed. For this purpose, we take France as the reference country.

Statistical inference and dimension reduction

Beyond description, let's test the significance of useful work.

In the first section, we showed that the median was relatively close to the mean, suggesting that the distribution of the "useful_work" variable appears to be somewhat normal. The calculated skewness is -0.3, indicating a slight leftward skew. The representation of the variable through its theoretical distribution quantiles (Document on the right) reinforces our idea that the variable is normally distributed.



Except for a few observations that could be considered outliers, the quantiles of the variable of interest converge toward the theoretical quantiles of a normal distribution. We can now use inference tools to test hypotheses

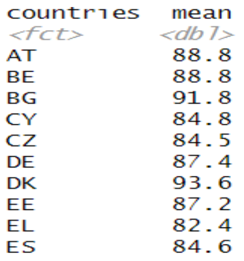
Statistically difference between men and women

Let's start by providing a measure of the relative risk: being a man increases the risk of considering the work as useless 2 times more, or women are 2 times less likely to considering the work as useless (relative risk = 0.52). Note that the p-value associated with this ratio is 0.6, which means that the relative risk cannot be statistically validated.

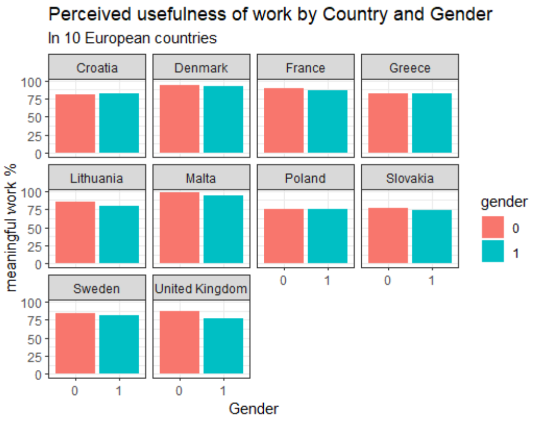
We have shown that there is only a 1.5 percentage point difference between the average useful work % of men and women. To compare these two averages, we need to apply a Student's t-test. It is worth mentioning that this test is statistically valid as long as the intergroup variance is approximately the same. This is indeed the case since the standard deviation of useful work for women is 4.74, compared to 5.24 for men (a ratio of less than 1.5).

Therefore, we set the null hypothesis to be the equality of useful work % means by gender. The p-value is 0.1346, which means there is only a 13.46% difference in significance between the two means. In other words, we cannot conclude a significant difference in perception between men and women.

A first conclusion is evident: since the male-female difference is not statistically significant, the 'gender' variable can be removed in favor of a general analysis without considering gender (**Document 5**). Consequently, our modified database now contains only 2 variables: 'countries' for countries and 'useful_work' for useful work rates/percentages.



```
countries   mean
<fct>      <dbl>
AT          88.8
BE          88.8
BG          91.8
CY          84.8
CZ          84.5
DE          87.4
DK          93.6
EE          87.2
EL          82.4
ES          84.6
```



Document 5: This bar chart clearly illustrates that there is no significant difference in perceived usefulness of work between men and women, regardless of the country of study

Reduction of country number by keeping only those that have a statistically significant correlation with France

Now that our database consists of only 2 variables, as requested by the statement, we will reduce the number of rows in our database, i.e., select the countries for which satisfaction is statistically significant. In the first step, we categorize the 'countries' variable and designate France as the reference country. We could have randomly selected another country, and the results would have been different, but the advantage of selecting France is that its useful work rate (88% for all genders) is close to the median (86.5%). This is equivalent to measuring the significance of rates compared to the median.
We propose a simple linear regression model between work useful rates and different countries, with France as the reference point. The results of this regression are provided in the table below and are available in the script under the name 'model'.

The names of the countries that have been selected (in yellow) are those for which the estimated coefficient associated with them is statistically significant, at least at the 5% significance level.

The model's significance is relevant, as the adjusted coefficient of determination is 0.7382, meaning that 73.52% of the variation in useful work rates/percentage is explained by the differences between the various countries and France.

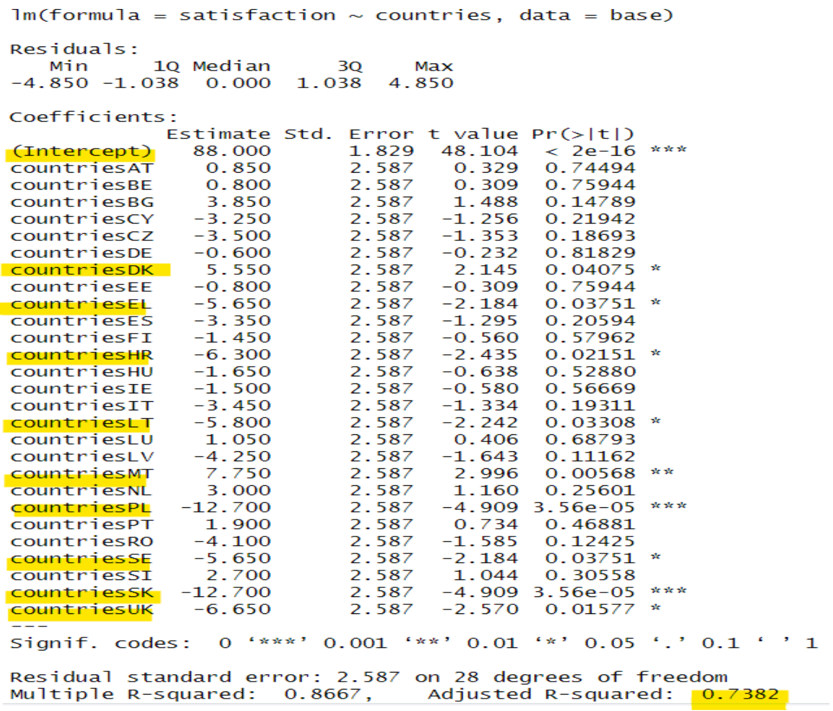
More specifically, the coefficients are interpreted as follows:

- *Intercept* : When we make no inter-country comparisons, the useful work rate is that of France, which is 88%.
- *Countries DK* : In Denmark, useful work rate is higher by *5.55* points compared to France, which means the estimated rate is 88 + *5.55* = 93.55%.
- *Countries PL* : In Poland, the useful work rate is lower by 12.7 points compared to France, which means the estimated rate is 88 - 12.7 = 75.3%.
- *Countries UK* : In the United Kingdom, the useful work rate is lower by 6.65 points compared to France, which means the estimated rate is 88 - 6.65 = 81.35%

It is important to note that the results of the regression are only estimates of the true values of the useful work rate. However, the objective of this regression in our case is to select a few countries for which the rate differences are statistically significant.

To truly assess the relevance of our approach, we can test the significance of the 'countries' variable as a whole concerning the 'useful_work' variable using a Fisher test. The p-value is less than 0.00001, which proves the objectivity of this variable: there is indeed a statistically significant difference in useful work rates based on the country considered.

Through this approach, we can isolate a group of 10 countries, including Denmark, Greece, Croatia, Lithuania, Malta, Poland, Switzerland, Slovakia, the United Kingdom, and of course France. At this stage, it is important to emphasize that there are as many comparison groups as there are countries, but we have chosen to select France. This does not imply that the other countries are less important but rather that the difference with France does not require statistical mention.



```
lm(formula = satisfaction ~ countries, data = base)

Residuals:
   Min     1Q Median     3Q    Max 
-4.850 -1.038  0.000  1.038  4.850 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)     88.000      1.829  48.104  < 2e-16 ***
countriesAT      0.850      2.587   0.329  0.74494    
countriesBE      0.800      2.587   0.309  0.75944    
countriesBG      3.850      2.587   1.488  0.14789    
countriesCY     -3.250      2.587  -1.256  0.21942    
countriesCZ     -3.500      2.587  -1.353  0.18693    
countriesDE     -0.600      2.587  -0.232  0.81829    
countriesDK      5.550      2.587   2.145  0.04075 *  
countriesEE     -0.800      2.587  -0.309  0.75944    
countriesEL     -5.650      2.587  -2.184  0.03751 *  
countriesES     -3.350      2.587  -1.295  0.20594    
countriesFI     -1.450      2.587  -0.560  0.57962    
countriesHR     -6.300      2.587  -2.435  0.02151 *  
countriesHU     -1.650      2.587  -0.638  0.52880    
countriesIE     -1.500      2.587  -0.580  0.56669    
countriesIT     -3.450      2.587  -1.334  0.19311    
countriesLT     -5.800      2.587  -2.242  0.03308 *  
countriesLU      1.050      2.587   0.406  0.68793    
countriesLV     -4.250      2.587  -1.643  0.11162    
countriesMT      7.750      2.587   2.996  0.00568 ** 
countriesNL      3.000      2.587   1.160  0.25601    
countriesPL    -12.700      2.587  -4.909 3.56e-05 ***
countriesPT      1.900      2.587   0.734  0.46881    
countriesRO     -4.100      2.587  -1.585  0.12425    
countriesSE     -5.650      2.587  -2.184  0.03751 *  
countriesSI      2.700      2.587   1.044  0.30558    
countriesSK    -12.700      2.587  -4.909 3.56e-05 ***
countriesUK     -6.650      2.587  -2.570  0.01577 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.587 on 28 degrees of freedom
Multiple R-squared:  0.8667,	Adjusted R-squared:  0.7382
```

Technical note

Creating a shapefile/map using Rstudio

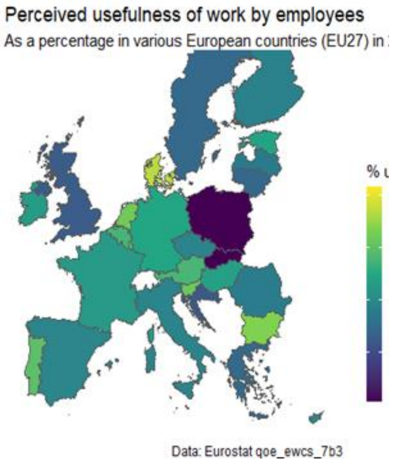
First of all we need the different libraries, including 'tidyverse,' 'eurostat,' 'leaflet,' 'sf,' 'scales,' 'cowplot,' and 'ggthemes'.

Then we integrate a 'shapefile' that provides the map of Europe using the 'get_eurostat_geospatial' function,

After, we create a dataframe that associates a code with each EU country using the 'eu_countries' dataframe from 'eurostat' (we include UK separately)

We creat a 'Map_EU28' dataframe with the 28 countries that you can display on the map using the Europe shapefile data and joining it with the 'EU28' data by country code.

We create a 'mean_countries_map' dataframe that merges the data from 'mean_countries' with 'Map_EU28' using the country code as the joining variable. After that, it's just customization using various functions from the packages: scale_x_continuous and scale_y_continuous to set the plot limits and using theme_void to remove labels and titles, ggplot, geom_sf, scale_fill_continuous etc.


Perceived usefulness of work by employees
As a percentage in various European countries (EU27) in
% u
Data: Eurostat qoe_ewcs_7b3

On this graph, we clearly see that there is no geographical similarity between countries regarding the Usefull Rate.

We can then reject the hypothesis of a country-by-country fragmentation and an inter-border spillover effect: each country has its own labour market and way of thinking.

Conclusion : From a general database to one maximizing data quality

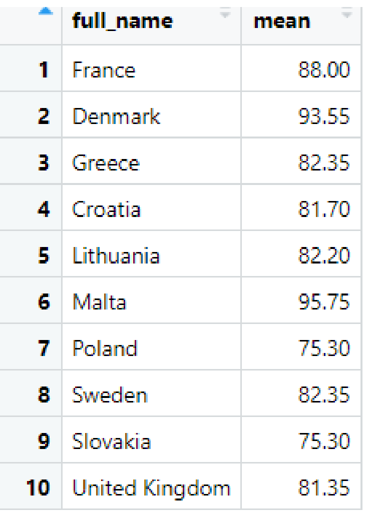
Using tools of descriptive and inferential statistics, our new database is of dimension

Just as a reminder, our initial database was of dimension 1080x6. Gradually, we removed variables without information and redundancy (such as year, unit of measurement, and age). Then, we tested the significance of the 'gender' variable to conclude that no statistically significant differences were observed. Finally, we clustered a set of countries around France that had relevant rate differences.

Document 6 shows the entirety of this database. Although dimension reduction methods like traditional data mining analyses (PCA, CA, clustering) are more appreciated by data analysts, they can still lack objectivity. Throughout this report, we always referred to statistical tests, ensuring that we respected the assumptions of validity (normality of the study variable). The only degree of freedom we took was the choice of the country. Thanks to the R software, our approach was simplified, and we had access to numerous tools to support our argument.

In conclusion, we will briefly comment on our final database and draw the main conclusions.

Document 6 : our new database as « general_useful » on the R script



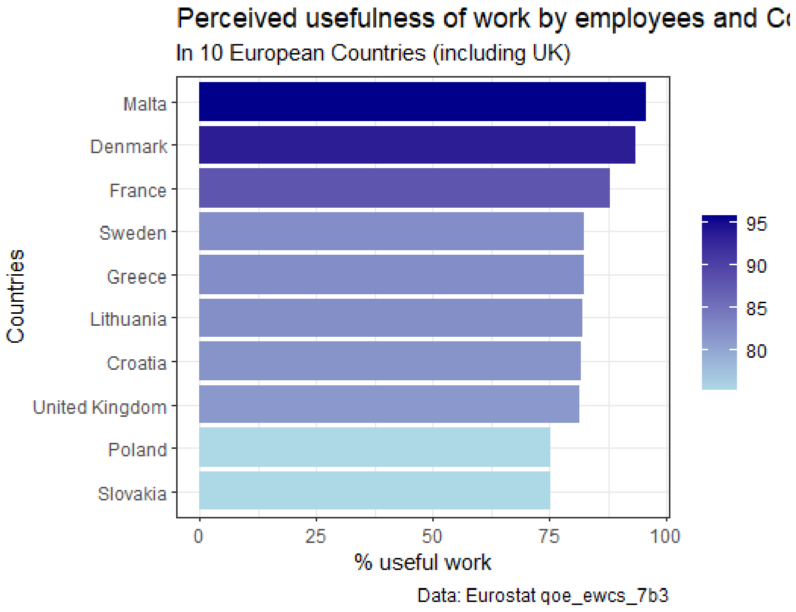
	full_name	mean
1	France	88.00
2	Denmark	93.55
3	Greece	82.35
4	Croatia	81.70
5	Lithuania	82.20
6	Malta	95.75
7	Poland	75.30
8	Sweden	82.35
9	Slovakia	75.30
10	United Kingdom	81.35

In general, the average useful work rate (ie, perceived usefulness of work by employees) is 83.78% compared to 85.95% for the initial database, and the median is 82.28% compared to the previous 86.45%. Despite a difference of more than 1 point, the values remain centered around approximately identical moments.

Moreover, this new database includes the useful rate/percentage for Malta and Poland, which were the two extreme values in the initial database. These similarities allow us to affirm that our new database is not distorted by a selection bias.

Geographically, **Document 7** highlights the existence of three new groups, more or less distinct:

- *Light blue group* : composed of Slovakia and Poland, these two countries have a similar useful work rate of around 75%.
- *Light blue group* : composed of the United Kingdom, Croatia, Lithuania, Greece, and Sweden, these countries have a relatively homogeneous useful work rate around 82%.
- *Dark blue group* : composed of France, Denmark, and Malta, this is the most heterogeneous group with a useful work rate ranging from 90% to 95%.



Document 7 : Cross-Country comparison of useful work (perceived usefulness of work by employees)

This categorization indicates that useful work/ perceived usefulness of work is not indicative of geographic location because the 2nd and 3rd groups consist of relatively diverse countries. Only the first group (sky blue) can be associated with Eastern Europe.

Overall, we observe that there isn't a clear link between a country's economic situation and the level of perceived usefulness of work reported by its workers. If there were such a link, France and the United Kingdom would be at the top of the rankings. Other more internal factors can explain these variations in useful work %, such as work practices, the development of telecommuting, or labor rights in general. Natural external factors, like the favorable climate of Mediterranean countries (Malta), can also play a role. Finally, the mentality and the sense of professional fulfillment are unique factors for each country that can influence the subjective feelings of workers.

Finally, employers attention to their employees working conditions is strongly correlated with the level of business activity. According to DARES (Bulletin 238, August 2020): *"On average, an increase in the level of prevention of one standard deviation is associated with a 23% increase in productivity (value added). Physical constraints, lack of recognition, schedule and working hour constraints, economic insecurity, and organizational changes are consistently associated with lower company performance, averaging between 5% to 12% for an additional standard deviation."*

In conclusion, as mentioned above, meaningful/useful work surveys are of public and economic utility. They allow us to gather insights from employees about their working conditions and the alignment of their positions with their aspirations. Beyond enabling cross-country comparisons, they can provide us with information about the sources of a deceleration in industrial activity.