# "DOWN WITH THE BLUES"

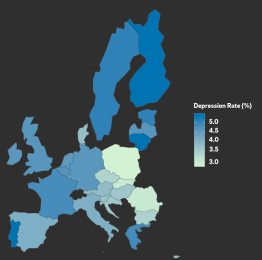
## Understanding the influence of social, environmental and economic variables on depression


**Authors:**

Filipa Sá da Costa | 20180051
Guilherme Martins | 20180056
João Fernandes | 20180061
Liah Rosenfeld | 20180044


## 01 OUR PROBLEM

Considering previously established theoretical (and practical) connections between economical [1,4], social [1,3,4] and environmental [2] factors and depression, our goal in this project was to understand the influence of these variables on this mental illness. Also, we wanted to see if a time dimension (from 2000 to 2016) would prove to be significant in the explanation of the depression rate.




### Data Description

- 16 variables
- 28 european countries

### Variables

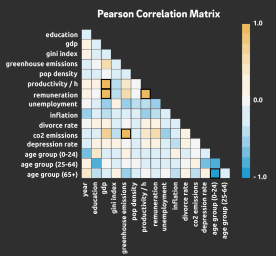
- **Environmental** (e.g. CO2 Emissions)
- **Social** (e.g. Population Density)
- **Economic** (e.g. GDP per capita)

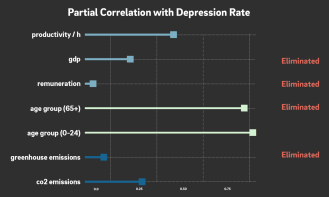
### Techniques

- Linear Regresssion
- Decision Tree
- Clustering
- ANOVA | Kruscal Wallis
- Post Hoc

## 02 PREPARING FOR THE MODEL

As we can see in the correlation matrix, we have multicollinearity issues meaning that some variables (more specifically the ones highlighted in the matrix) have a high correlation between them. We computed partial correlation between these "problematic" variables and our dependent variable and kept the ones with the higher partial correlation.




## 03 UNDERSTANDING THE INFLUENCE

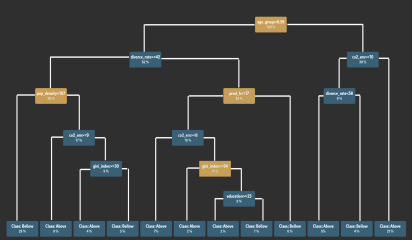
Now that we have our final variables, we followed two approaches in order to understand which ones influence the depression rate. First, a linear regression was computed and conclusions were drawn from the standardised coefficients and their significance. Secondly, a decision tree was made in order to predict wether a country would have a depression rate above average (in each specific year). The results show that Gini Index, Population Density, Productivity per Hour and proportion of adults between the ages 25 to 64 appear to have influence on the depression rate in both approaches (using $\alpha = 0.05$ for the regression).

### Linear Regression Results

| | std coefficient | std error | p-value |
|---|---|---|---|
| intercept | 0.222 | 0.146 | 0.130 |
| year | 0.059 | 0.075 | 0.437 |
| pop density | 0.299 | 0.096 | 0.002 |
| divorce rate | 0.000 | 0.046 | 0.994 |
| age group (25-64) | 0.274 | 0.073 | 0.000 |
| productivity / h | - 0.280 | 0.114 | 0.014 |
| co2 emissions | - 0.097 | 0.072 | 0.178 |
| education | 0.001 | 0.047 | 0.983 |
| unemployment | 0.007 | 0.041 | 0.871 |
| gini index | 0.112 | 0.040 | 0.006 |
| inflation | - 0.157 | 0.097 | 0.106 |
| age group (0-24) | 0.142 | 0.120 | 0.237 |

Adjusted $R^2$: 0.4 %

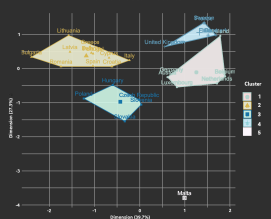
### Decision Tree



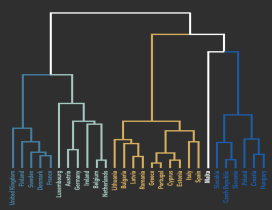
## 04 CLUSTERING

Our second objective consisted in characterizing the euro-28 countries according to the significant influencing variables on depression. We performed cluster analysis using the variables that consistently appeared as significant in the previous step. We did this analysis with data from the most recent year in our dataset (2016). The Dendrogram presented below results from an agglomerative hierarchical clustering with Ward method (which yielded the best results). Finally, the presented k-means results below were obtained with an initialization of the centroids based on the centroids obtained in the hierarchical clustering

### K-Means Cluster Results



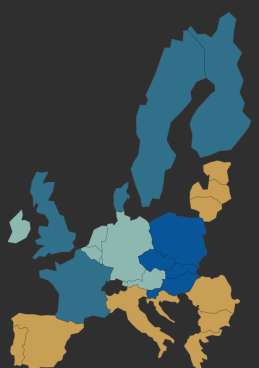

### Hierarchical Clustering Results



Clustering based on pop density, age group 25-64, productivity/h and gini index.

## 05 CONCLUSIONS

We conducted a one-way Anova to see if the clusters differ significantly when it comes to the depression rate. Since the assumptions were not held, we computed a Kruskal-wallis test. We concluded that there is a statistically significant difference in the average depression rate between the clusters (Kruswal-Wallis(4)=10.53,p<0.05). Games-Howell Post hoc tests allowed us to conclude that only cluster number 3 (that presents the lowest average depression rate) differs significantly from the remaining ones (p<0.05).



**Cluster 1**

This cluster includes countries like Germany, Luxemburg and the Netherlands. It's distinguished for its high productivity per hour.

**Cluster 2**

This cluster is composed by countries like Portugal, Romania and Greece. It's characterized by having an elevated gini index and,thus, a high social inequality state.

**Cluster 3**

This cluster consists in the countries with the lowest amount of social inequality (i.e., with the lowerst gini index). It includes countries like Hungary, Slovakia and Poland.

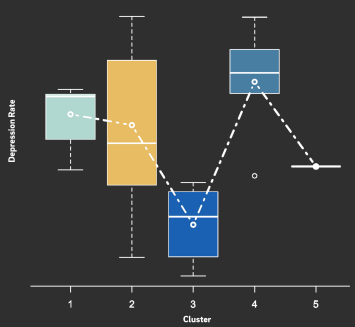
**Cluster 4**

Containing countries with an elevated productivity per hour, a small proportion of population in the active ages (between 25 and 64) and a relatively high population density, this cluster includes countries like Denmark, Sweden and UK.

**Cluster 5**

Situated in a country by iteself we have Malta. This country is distinguished by an astonishing elevated population density.

### Depression Rate Variation by Cluster




Our results show the importance of social, economical and socio-economical variable on depression.

- We concluded that the more individuals between the ages of 25 and 64, the higher the depression rate in that country tends to be. Taking in consideration that, as stated in [4], the risk group for this illness consists in individuals between the ages of 18 and 29, our results may shed some light on the fact that the risk group is more bound to the upper boundary of that interval.

- We also concluded that the cluster with the countries that have the lowest depression rate, is also the one with the lowest Gini-index, shedding some light on the influence of that variable on depression.

- Furthermore, our results indicated that the more productive the country the less it's propense to have a high depression rate. This makes sense considering the importance of economic variables presented in [2], being that this variable can be an indicative of wealth.

- Finally, we concluded that the more the population density, the higher tends to be the depression rate of a country. This could happen because this variable could be indicative of a high urbanisation of the country, which can lead to rising income inequalities [5] reflecting in the also proved significant Gini-index.


References:
[1]Freeman, A., Tyrovolas, S., Koyanagi, A., Chatterji, S., Leonardi, M., Ayuso-Mateos, J. L., ... & Haro, J. M. (2016). The role of socio-economic status in depression: results from the COURAGE (aging survey in Europe). BMC public health, 16(1), 1098.
[2]Heikkinen, K., & Poivios, L. (2015). Environmental and Economic Factors Associated with Mental Illness.
[3] Allen, J., Balfour, R., Bell, R., & Marmot, M. (2014). Social determinants of mental health. International review of psychiatry, 26(4), 392-407.
[4] American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (DSM-5®). American Psychiatric Pub.
[5] McGranahan, G., & Satterthwaite, D. (2014). Urbanisation: concepts and trends. London: IIED.


Data Source:
- [-] PORDATA
- [-] Our World in Data