Master Degree Program in

**Data Science and Advanced Analytics**

# PORTRAIT OF YOUNG ADULTS IN THE EUROPEAN UNION IN 2019

## Data Visualization


Álvaro Reis, Number: 20220679

David Martins, Number: 20221006

Diogo Martins, Number: 20221361

Marta Correia, Number: 20220709

**NOVA Information Management School**

**Instituto Superior de Estatística e Gestão de Informação**

Universidade Nova de Lisboa


April, 2023

# 1. INTRODUCTION

Young adults are a crucial part of the European Union's population, yet they face significant challenges in various areas of their lives. According to the Eurofound report on younth in Europe in 2019 [1], many young adults in the EU struggle with education, employment and housing issues. Also, according to the European Health Interview Survey (2019) [2], a significant proportion of young adults in the EU reported engaging in risky health behaviours (such as smoking or insufficient physical activity). In this context, the project aims to provide a comprehensive portrait of young adults in the EU in 2019, using reliable data sources and visually engaging graphics.

In the project, we have presented two different views of the situation of young adults (aged 15-24) in the European Union: The first view focuses on health factors, while the second view highlights various socioeconomic factors. In this way, we aimed to provide a comprehensive picture of the circumstances of young adults in the EU in 2019, highlighting key areas of concern and offering insights into the factors that shape their lives. By using data visualization techniques, we sought to make this information more accessible and engaging, helping to promote greater understanding and awareness of the challenges and opportunities facing young adults in the EU.

The data visualization project will provide various interactive options, including hovering over data points for additional information, clicking on data points for navigation or detailed data display, and zooming in/out. Users can filter and sort data based on specific criteria and compare data points side-by-side using correlation plots. These interactive features create a more engaging and immersive experience, enabling users to analyze and understand the data better.

# 2. DATA AND METHODS

Our dataset includes a collection of 11 datasets, 6 of which are focused on sociodemographic factors and 5 on health factors. Each dataset includes data for all European Union countries for the year 2019. After collecting the data, we created two main datasets, one for sociodemographic factors and the other for health factors. These datasets were used to create our visualizations. Our data collection process involved downloading datasets from Eurostat. We ensured that our data was reliable and up-to-date, making it suitable for our data visualization project. Overall, our dataset provides a comprehensive view of young adults in the European Union in 2019, highlighting key sociodemographic and health factors that shape their lives. In *Table 1* it is possible to observe the variables we used and their description.

In this project, all the data visualizations were implemented using Tableau. This allowed us to create interactive and visually appealing graphics. Throughout the process, we imported our data sources into Tableau and made use of a range of different Tableau features and functions in order to ensure that our final product was both informative and visually engaging. It is possible to access the project via this **link**.

Our data visualization project was developed with a keen focus on the three pillars of storytelling for data visualization: visual design, message, and interactivity [3]. In terms of visual design, we have adopted an open exploration, reader-driven narrative approach. This means that instead of presenting a fixed storyline, we have aimed to create an interactive and collaborative experience for the reader.

Our intention with this visualization is to encourage the reader to actively engage with the data and to explore different aspects of young adults' lives in the European Union. In this way, we hope to create a dynamic and interactive visualization as our goal is to make this project and engaging and informative experience for the readers, while also encouraging critical thinking and active participation. Regarding messaging, we used headlines, captions and annotations that effectively highlighted the most important information in our data visualization. Finally, with regards to interactivity, we aimed to create a highly engaging experience through our data visualization project. To achieve this, we incorporated interactive features into each of the graphs, allowing users to explore the data and interact with the visualization on their terms.

A health view was created to focus on health factors affecting young adults in the European Union in 2019. The aim is to provide insights into their health behaviors and outcomes, highlighting areas of concern. We will now discuss the visualization and interaction choices for each specific visualization.

The first visualization featured prominently in our dashboard shows the distribution of individual health factors, including "Overweight", "Smoking", "Alcohol", "Exercise", and "Fruits and Veggies", across each country (*Figure 1*). Users can easily interact with this visualization by selecting a specific country of interest from the drop-down menu located at the top right corner of the histogram. This feature provides users with a dynamic and customizable viewing experience, enabling them to tailor their analysis and explore different health trends across countries. By using a histogram, we can effectively communicate the distribution of each health factor within each country, making it easy for users to interpret the data and identify potential patterns and trends. Moreover, the histogram's visual appeal and ease of use make it a compelling tool for users to engage with and explore the data.

The second visualization available is a map that provides a quick overview of the health status of young adults in the EU (*Figure 2*), allowing users to easily identify countries with higher or lower rates of the chosen health behavior. By using the dropdown menu, it is possible to switch between different health behaviors (smoking rate, alcohol consumption, physical activity levels, and fruit and vegetable consumption and overweight levels). The map visualization was chosen as it provides a clear and intuitive way to present geographical data and allows for easy identification of regional patterns and variations. The dropdown menu and color-coded legend allow users to interact with the map and customize their view based on specific health factors of interest.

The treemap visualization was chosen for displaying the causes of death, providing an intuitive way to present these data (*Figure 3*). Each rectangle in the tree map represents a category of causes of death, with the size of the rectangle proportional to the number of deaths attributed to that category. The hierarchical structure of the tree map also allows users to drill down into specific causes of death by clicking on smaller rectangles within larger ones. This makes it easier for users to explore the data and identify patterns or trends in the causes of death across different countries.

Finally, there is a scatterplot where users can select two health variables of interest to display on the X and Y axis (*Figure 4*). The scatterplot shows the correlation between two variables that can be chosen by the user and provides an additional layer of information for the user to consider when analyzing the correlation between the two health variables. This graphic was considered as a useful tool for exploring relationships between the different health factors. Also, by allowing users to select two variables to visualize, the graphic can help identify potential correlations or patterns in the data. The graphic allows also to see the correlation coefficient (users can hover over the line to see the exact

correlation coefficient value) and the line of best fit for the two indicators. Finally, the scatterplot allows users to see the position of each EU country relative to the line of best fit, enabling them to identify outliers and explore any interesting trends or patterns in the data.

A sociodemographic view was created to provide insights into the demographic and social characteristics of young adults in the European Union in 2019. The aim is to understand the diversity of the population and the challenges faced by different subgroups. We will now discuss the visualization and interaction choices for each specific visualization.

The histogram visualization in the sociodemographic view displays the distribution of a chosen sociodemographic measure for a selected number of top and bottom EU countries. Users can interact with the histogram by selecting different measures and country numbers to compare and understand the distribution of different sociodemographic data. The histogram provides an effective way to analyze and compare sociodemographic information in a simple and clear manner (*Figure 5*).

The bubble chart allows users to choose between emigrants and immigrants and displays the corresponding data points for each selected category as bubbles (*Figure 6*). The size of each bubble represents the total number of emigrants or immigrants for the selected country, with larger bubbles indicating a larger number of emigrants or immigrants. Users can interact with the visualization by hovering over individual bubbles to see the specific values for each country, providing a detailed understanding of the emigration and immigration patterns of different EU countries. Overall, the bubble chart provides an intuitive and easy-to-understand way of visualizing complex data related to emigration and immigration in the EU.

The boxplots available provide a visual representation of the distribution of four sociodemographic variables (*Figure 7*). It is possible to hover over the individual boxes to display more detailed information (such as the minimum and maximum values, median and quartiles). The choice of using boxplots as a visualization technique allows for a clear and concise representation of the distribution of each variable and easy comparison between variables.

In the sociodemographic view, there is a scatterplot that is similar to the one in the health view (*Figure 8*). All justification of visualization and interaction choices for this graph are the same as presented above in health view.

# 3. RESULTS AND DISCUSSION

## 3.1. DATA ENCODING

The histogram (*Figure 1*) primarily uses the horizontal position of the bars as a channel to encode the rate of each health factor within each country. The y-axis represents the different health factors being plotted, while the x-axis represents the rate of each health factor within each country. The data encoding can be further customized using the drop-down menu, which allows users to select a specific country of interest and view the distribution of health factors for that country.

The map visualization (*Figure 2*) uses color as a channel to represent different health indicators for each country, with darker colors indicating higher rates and lighter colors representing lower rates.

The drop-down menu at the top of the visualization allows users to select which health indicator they want to visualize, adding an additional layer of interactivity to the visualization.

The tree map visualization (*Figure 3*) encodes data using size and color as channels to represent different information. The size of each rectangle corresponds to the number of deaths attributed to a specific cause, while the color indicates the proportion of deaths relative to the total number of deaths in a particular country. For example, if a rectangle is dark blue, it represents a cause of death that accounts for a large percentage of total deaths in that country.

The scatterplot (*Figure 4*) uses two continuous variables to encode data, with each point (mark) on the plot representing a single country. The position of each point on the vertical and horizontal axes corresponds to the values of the two selected health indicators, allowing users to visualize the correlation between them. Additionally, a trendline is included in the scatterplot to help users identify the general trend in the data. Hovering over individual data points displays the country name and the exact values of the two selected indicators for that country, enabling users to explore the data in more detail.

The histogram visualization in the sociodemographic view (*Figure 5*) uses a basic data encoding approach where the x-axis represents the range of values for the selected sociodemographic measure, and the y-axis represents the frequency of countries within each range. The height of each bar represents the number of countries with a value falling within that range. The use of a simple data encoding approach in the histogram allows for a clear and concise representation of the distribution of the selected sociodemographic measure across different EU countries.

In the bubble chart (*Figure 6*), the data encoding involves the use of three variables: the country, the number of immigrants, and the number of emigrants. The size of each bubble represents the total number of immigrants or emigrants. This allows users to compare the number of immigrants and emigrants across different countries and to see which countries have the largest inflow and outflow of people.

The four boxplots in the sociodemographic view (*Figure 7*) encode a continuous variable on the y-axis and a categorical variable (country) on the x-axis. The boxplots allow for easy identification of median, interquartile range or outliers for each variable in a visually appealing way.

All justification of data encoding for the scatterplot of the sociodemographic view (*Figure 8*) are the same as presented above relatively to the scatterplot of the health view.

### 3.2. DATA FILTERING

The histogram visualization in the dashboard (*Figure 1*) provides a way for users to interactively filter the data by selecting a specific country of interest from the drop-down menu located at the top right corner of the histogram.

The map visualization (*Figure 2*) offers an interactive filtering option that corresponds to the drop-down meu. The drop-down menu allows the user to select a range of health indicators.

The tree map visualization (*Figure 3*) allows for interactive data filtering through the use of a dropdown menu that enables users to select a specific country of interest. Once a country is selected,

the tree map will display the causes of death for that country. Additionally, Users can also hover over individual squares to view the number of deaths associated with each cause.

The scatterplot (*Figure 4*) offers interactive data filtering by allowing users to choose two health factors to compare. This enables users to focus on specific variables and investigate the relationships between them. Moreover, the scatterplot allows users to hover individual data points to see the location of each European Union country on the plot. This feature helps users to explore the data in more detail and compare different countries. It is also possible for users to hover over the line and see the value of the correlation coefficient, which provides users with additional information about the strength and direction of the relationship between the two variables being compared.

Data filtering is a key feature of the histogram visualization in the sociodemographic view (*Figure 5*). Users can choose the number of top and bottom countries they want to display in the histogram, as well as the specific sociodemographic measure they want to analyze. This allows users to filter the data and focus on specific aspects of the sociodemographic data that are relevant to their analysis.

The bubble chart in the sociodemographic view (*Figure 6*) features a toggle button that enables users to switch between viewing data on emigrants or immigrants. The size of the bubbles corresponds to the number of emigrants or immigrants for each EU country, with larger bubbles indicating countries with higher numbers of emigrants or immigrants. This allows users to explore the data and compare different countries in terms of their emigration and immigration patterns.

In the four boxplots (*Figure 7*), it is possible to hover the over the boxes to display the specific values for each country. This allows users to easily compare the values for each country and identify any outliers or interesting patterns in the data.

All justification of data filtering for the scatterplot of the sociodemographic (*Figure 8*) view are the same as presented above relatively to the scatterplot of the health view.

## 4. CONCLUSION

In conclusion, our data visualization project has provided a comprehensive portrait of young adults in the European Union in 2019, highlighting key sociodemographic and health factors that shape their lives. Through the use of reliable and up-to-date data sources, visually engaging graphics and interactive options, we aimed to promote greater understanding and awareness of the challenges and opportunities facing young adults in the EU. In future work, we could explore additional factors that affect young adults' lives such as youth participation in society or youth and the digital world. Additionally, we could broaden our visualizations to encompass all continents and incorporate data from multiple years to provide a more comprehensive view of the trends and changes over time.

Overall, our data visualization project makes a significant contribution to enhancing the comprehension of the situations facing young adults in the European Union. We have adhered to the fundamental principles of data visualization that we acquired from our classes, resulting in a robust and informative representation of the data.

## 5. REFERENCES

[1] Eurofound. (2020). Living, working and COVID-19: First findings - April 2020. Luxembourg: Publications Office of the European Union.

[2] European Union Agency for Fundamental Rights. (2022). European Health Interview Survey 2019/2020 (EHIS) - Health Status Report. Luxembourg: Publications Office of the European Union.

[3] Segel, E., & Heer, J. (2010). Narrative visualization: Telling stories with data. IEEE transactions on visualization and computer graphics, 16(6), 1139-1148.

## 6. ANNEXES

*Table 1 - Variables and its description*

| Variable | Description |
|---|---|
| Country | Name of the country |
| Overweight | Percentage of youth that is overweight |
| Smoking | Percentage of youth that smokes every week |
| Alcohol | Percentage of youth that consumes alcohol every week |
| Exercise | Percentage of youth that completes more than 150 minutes of weekly aerobic activity |
| Fruits & Veggies | Percentage of youth that does not consume fruits or vegetables daily |
| Lives with Parents | Percentage of youth living with their parents |
| Unemployment | Percentage of unemployed youth as a percentage of the labour force |
| Risk of Poverty | Percentage of youth at risk of poverty rate |
| Tertiary Education | Percentage of youth who have successfully completed tertiary studies |
| Immigrants | Percentage of youth immigrants |
| Emigrants | Percentage of youth emigrants |

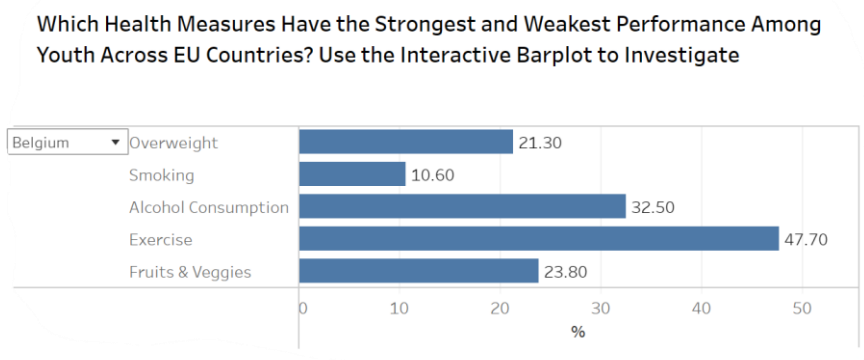


*Figure 1 - Histogram (Health View) - health factors across each country*

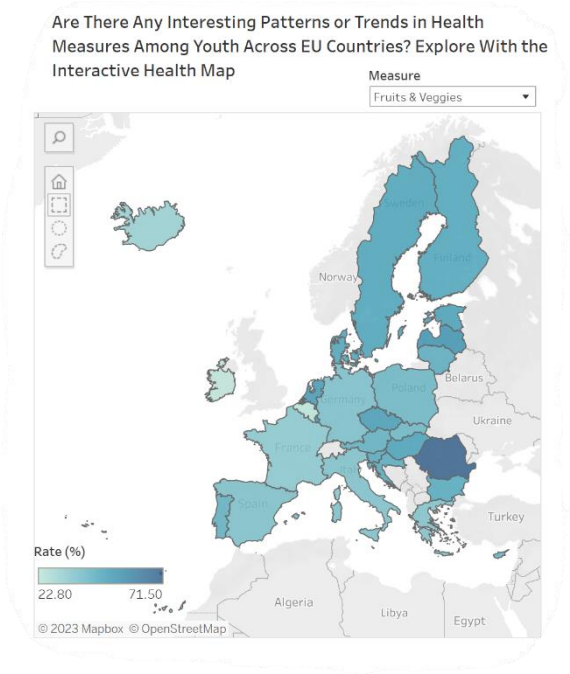


*Figure 2 - Map (Health View) - overview of the health status of young adults in the EU*

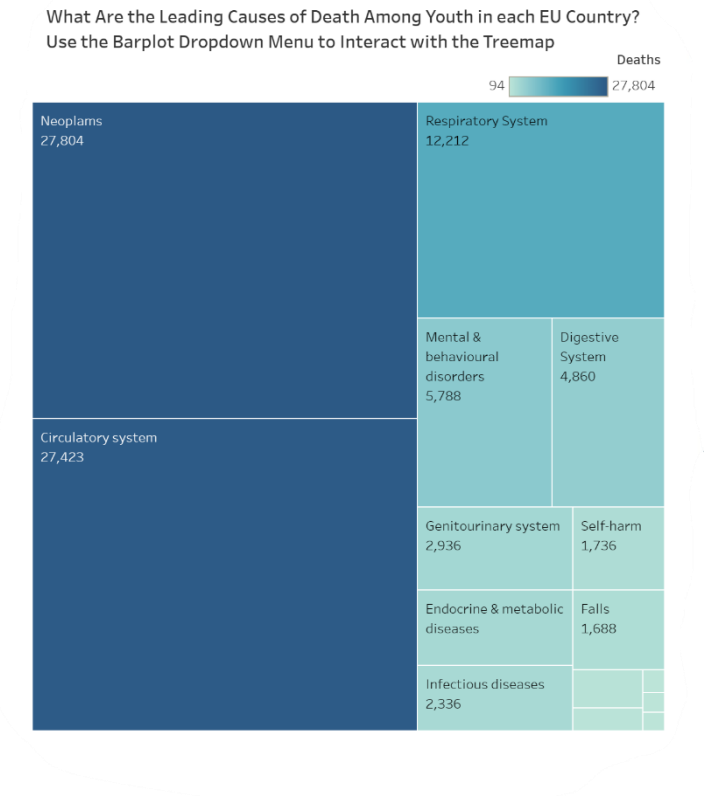


*Figure 3 - Treemap (Health View) - major causes of death accross each country*

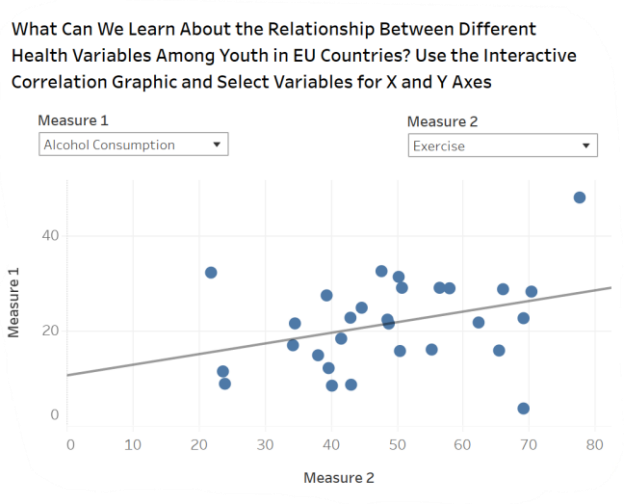


*Figure 4 - Scatterplot (Health View) - correlation between two health-related features*

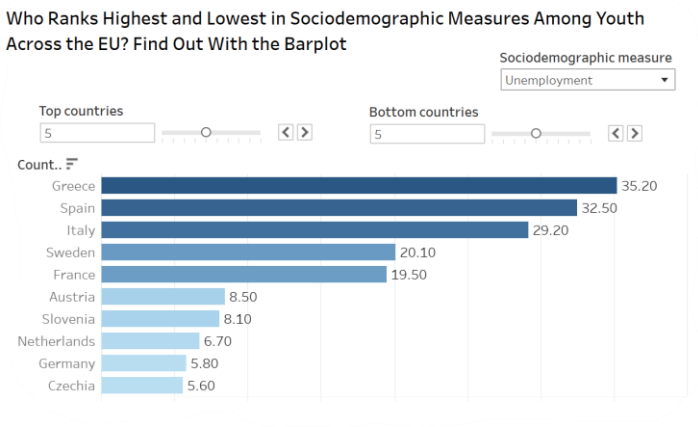


*Figure 5 - Histogram (Sociodemographic View) - top and bottom countries for sociodemographic-related features*

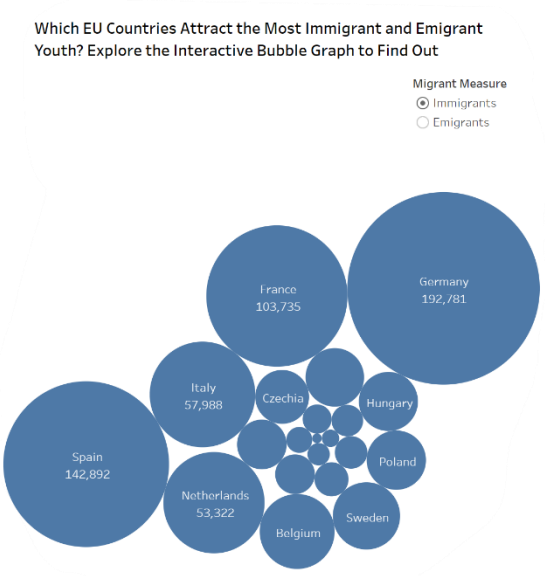


*Figure 6 - Bubble Chart (Sociodemographic View) - Emigrants and Immigrants of European Union countries*

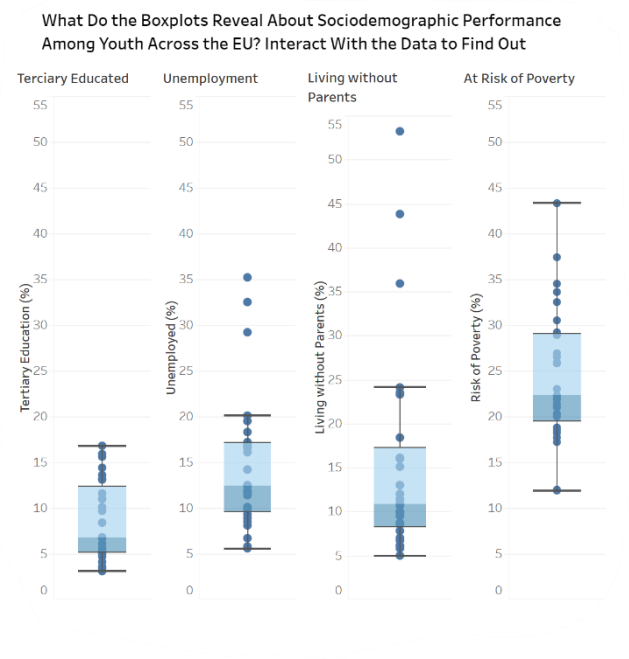


*Figure 7 - Boxplots (Sociodemographic View) - visual representation of the distribution of four sociodemographic-related variables*

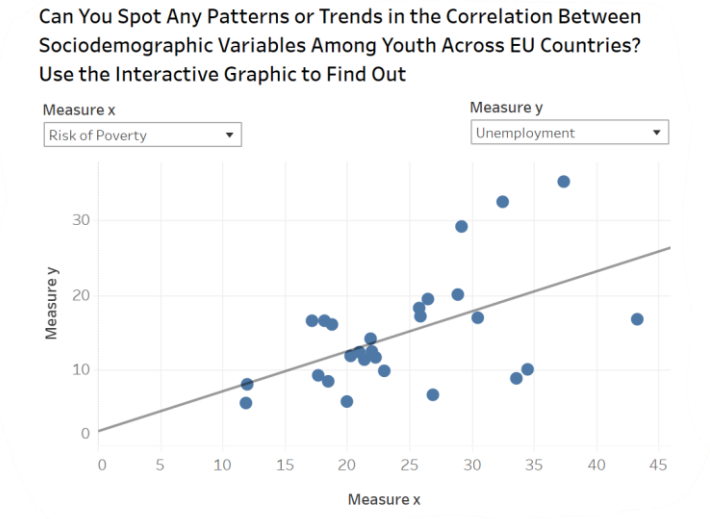


*Figure 8 - Scatterplot (Sociodemographic View) - correlation between two sociodemographic-related features*