# Final Essay – Methods of Empirical Social Research

## Outlined requirements Q4

Read and evaluate the
paper https://journals.sagepub.com/doi/full/10.1177/05694345221129022

Based on the publication create a data driven design yourself that operationalizes a different variable than income inequality. Choose a different variable to analyze (e.g., education level, health outcomes, employment rates).

State the main variables you are analyzing and justify your choice of visualization method (e.g., bar chart, scatter plot, heatmap).

Find a scientific publication that supports the choice of your main variables and the approach you used for visualization.

Cite this publication and explain how it backs your design and analysis.

## Introduction

In this paper, I will provide an in-depth evaluation of "Teaching Income Inequality with Data-Driven Visualization" by Sang T. Truong and Humberto Barretto. Following this analysis, I will present my own approach to data-driven visualization, focusing on employment rates as the primary variable. This analysis will compare employment rates to geographical location, aiming to uncover insights into how employment rate disparities correlate with regional factors. I will focus on the NUTS2 regions in my visualization. The whole project is interactive so you can compare any of the 272 regions with each other.

## NUTS2 Regions

NUTS2 regions are part of the Nomenclature of Territorial Units for Statistics (NUTS), which is a hierarchical system for dividing up the economic territory of the European Union (EU) and the United Kingdom (UK) for statistical purposes. The NUTS classification was created and is maintained by Eurostat, the statistical office of the EU. (Eurostat, 2022)

## Evaluation of "Teaching Income Inequality with Data-Driven Visualization"

This paper presents an innovative approach to visualizing income distribution and inequality in the United States using U.S. household income microdata from the Current Population Survey. By leveraging modern visualization techniques, the authors aim to make the concept of income inequality more accessible to non-expert audiences. The 3D dynamic visualizations highlight changes in income distribution over time, adjusted for inflation, regional price differences, and household size, offering a more nuanced understanding than traditional measures like the Gini coefficient. The interactive nature of the visualization promotes engagement and further exploration of economic disparities, providing valuable resources for educators and

policymakers. Despite some limitations due to data collection variability, this methodology significantly enhances the communication of complex economic data. (Truong & Barreto, 2022)

## GitHub repository

Link: https://github.com/rangeraider/emprs_finalproject

## Data Set

I am using two datasets from Eurostat to create my visualizations.

### Dataset 1:

https://ec.europa.eu/eurostat/databrowser/view/lfst_r_lfsd2pwn/default/table?lang=en

Full data set download as a csv file. I provided the downloaded version which was available at 14.07.2025 at 16:53. This dataset holds population data from NUTS2 regions across Europe and their employment status as well as age group, sex, and working status. This dataset can only be download by the official link since it is too big for GitHub.

### Dataset 2:

Is created by myself by using an excel file where all labels of the NUTS2 regions are stored. I had to do this since the file wasn't readable by any python extensions. I again provided a copy in the data directory of my project in my GitHub repository.

The original xml file comes from: https://ec.europa.eu/eurostat/web/nuts/ which was accessed at 14.07.2024 at 16:58.

## Data Preprocessing

The data preprocessing involved several key steps: loading data from CSV and Excel files into Data Frames, storing these into an SQLite database, and creating new tables while dropping irrelevant columns. Data cleaning included removing rows with unwanted values in the `geo`, `citizen`, and `age` columns and updating region names using a reference table. Further, employment and population data were filtered, merged, and used to calculate employment rates. The processed data was then visualized through a Tkinter GUI, allowing users to select regions and view employment rate trends by sex over time.

## Used Variables and Explanation

- unit: The unit of measurement, consistently "THS_PER" in both data frames which means thousand persons.
- wstatus: The working status, with distinct values including "ACT", "EMP", "INAC", "POP", "UNE", "UNK" in df_emprate_nuts2 and df_emprate_country.
    - ACT actively looking for work
    - EMP employment
    - INAC inactive
    - POP whole working population
    - UNE unemployment
    - UNK unknown
- citizen: The citizenship status, consistently "TOTAL" in both data frames.
- sex: The gender, with distinct values "F", "M", "T" in both data frames.
- age: The age group, consistently "Y15-64" in both data frames.

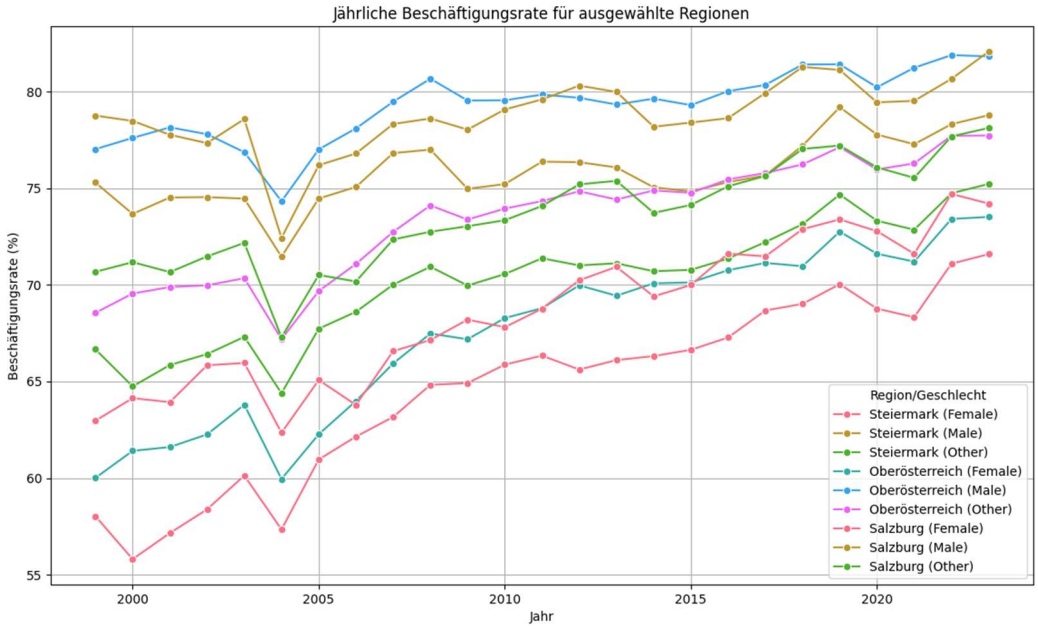
- geo: The geographical region, with NUTS2 codes in df_emprate_nuts2 and country codes in df_emprate_country.
- TIME_PERIOD: The time period, with example years from 1999 to 2022 in df_emprate_nuts2 and from 1999 to 2022 in df_emprate_country.
- OBS_VALUE: The observed values, various numerical values representing employment or population counts.
- nuts2_name: The names of the NUTS2 regions, present only in df_emprate_nuts2.

# Visualizations

## Line Chart

In their paper "Task-Based Effectiveness of Basic Visualizations," the authors explore how various basic visualizations perform across different analytical tasks. Specifically, they found that line charts are highly effective for identifying correlations between data attributes. Users favored line charts for correlation tasks due to their clear depiction of trends and relationships over time, which outperformed other types such as scatterplots. (Saket, Endert, & Demiralp, 2018)

For this project, which aims to visualize employment rates across regions and demographics over time, line charts are an ideal choice. They effectively illustrate trends in employment rates, making it easier to compare changes across different regions and between genders. This aligns perfectly with the project's goal of providing a clear and accessible method for users to analyze and understand employment trends. By utilizing line charts, the project can highlight correlations and trends in the employment data, facilitating a more intuitive and informative analysis.




This just shows one type of visualization which can be created using my application. The full project as well as detailed readme instructions is included in my provided file in my GitHub repository where you can have a look at the developed application.

## Further Expansions

You can develop similar applications for entire countries or other regions worldwide if the necessary data is available. Additionally, you can integrate pie charts to highlight which country has the highest employment percentage. These applications can help visualize employment trends over time and across different regions, making it easier to identify patterns and areas needing attention. With more advanced features, you can also incorporate interactive elements, such as dropdown menus or sliders, to allow users to customize the data view based on their specific interests or requirements.

References:

Truong, S. T., & Barreto, H. (2022). Teaching Income Inequality with Data-Driven Visualization. *the American Economist/the American Economist*, *68*(1), 140–155. https://doi.org/10.1177/05694345221129022

Eurostat. (2022). *NUTS - Nomenclature of territorial units for statistics: Overview*. Publications Office of the European Union. https://doi.org/10.2785/321792

Saket, B., Endert, A., & Demiralp, Ç. (2018). Task-based effectiveness of basic visualizations. *IEEE Transactions on Visualization and Computer Graphics, 25*(7), 2505-2512. https://doi.org/10.1109/TVCG.2018.2825426

Eurostat. (n.d.). *Employment by sex, age and NUTS 2 regions* [Data set]. Retrieved July 14, 2024, from https://ec.europa.eu/eurostat/databrowser/view/lfst_r_lfsd2pwn/default/table?lang=en