Report

Data Analysis & Visualization

Assignment 2: Working with Real Data

Course Informatik 1 - WS 2024/25

Group BG

Fabian Müller, 01410690

Daniel Schipfer, 11723440

Thomas Walchhofer, 51800280

Monday, January 13, 2025

1. Introduction

This report provides insights derived from OECD datasets using various visualizations. The objective was to explore and analyze data related to alcohol consumption, Gini coefficients, GDP per capita, calorie intake, and their relationships across multiple countries over time. The datasets were chosen for their relevance to social, economic, and health-related trends and their potential to uncover interesting patterns.

2. Code Design

The code was designed with modularity and reusability in mind. The following functions were implemented to streamline data analysis and visualization:

- **Data Loading Functions:** Separate functions (get_alcohol_data, get_gini_data, get_calorie_data, get_gdp_data) load and preprocess individual datasets. The reason separate functions were used is that the filtering for every dataset was different.
- **Visualization Functions:**
    - line_plot: For plotting time-series data of variables across multiple countries.
    - bar_plot: For comparing a variable across countries in a specific year.
    - scatter_plot: For visualizing the relationship between two variables and visualizing their correlation coefficient.
- **Utility Functions:** The merge_df function combines datasets for scatter plots based on common attributes (country, year).

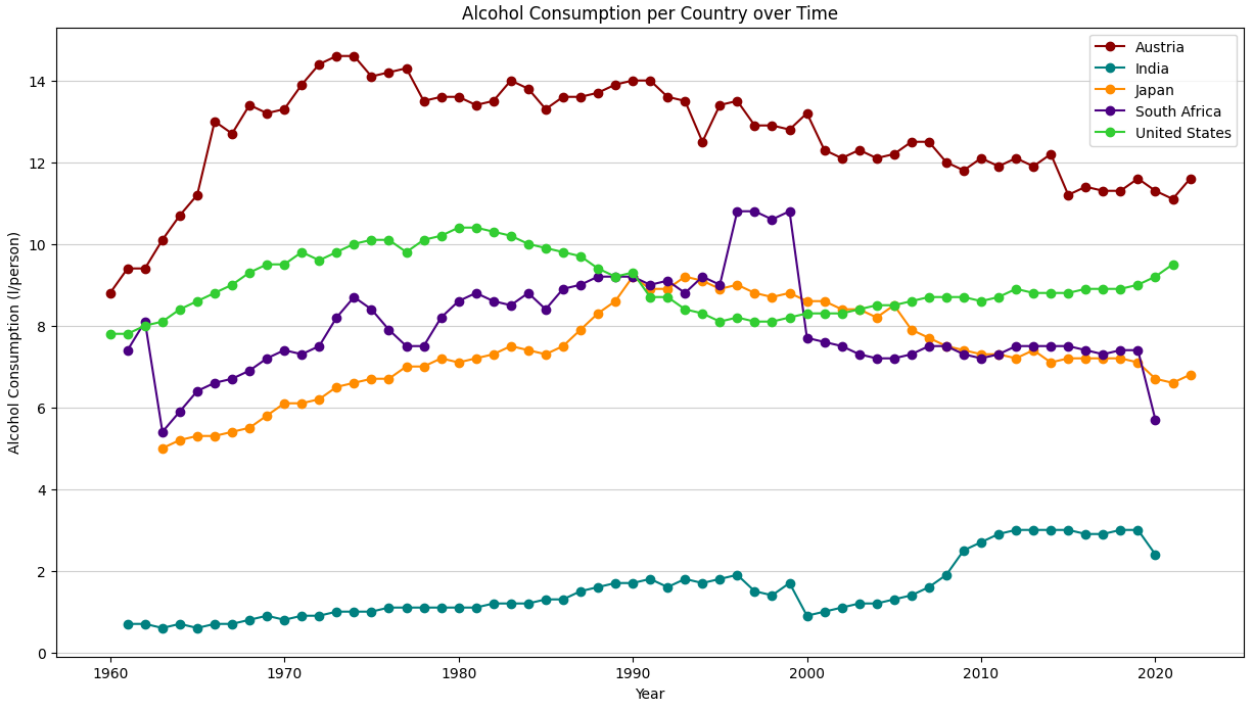
Function Design Rationale

1. **Reusability:** Functions accept parameters for customization, enabling flexibility in plotting.
2. **Modularity:** Each function performs a specific task, making the code easier to understand and maintain.
3. **Scalability:** The design allows for easy addition of new datasets or modifications to visualizations.

3. Data Visualization and Analysis

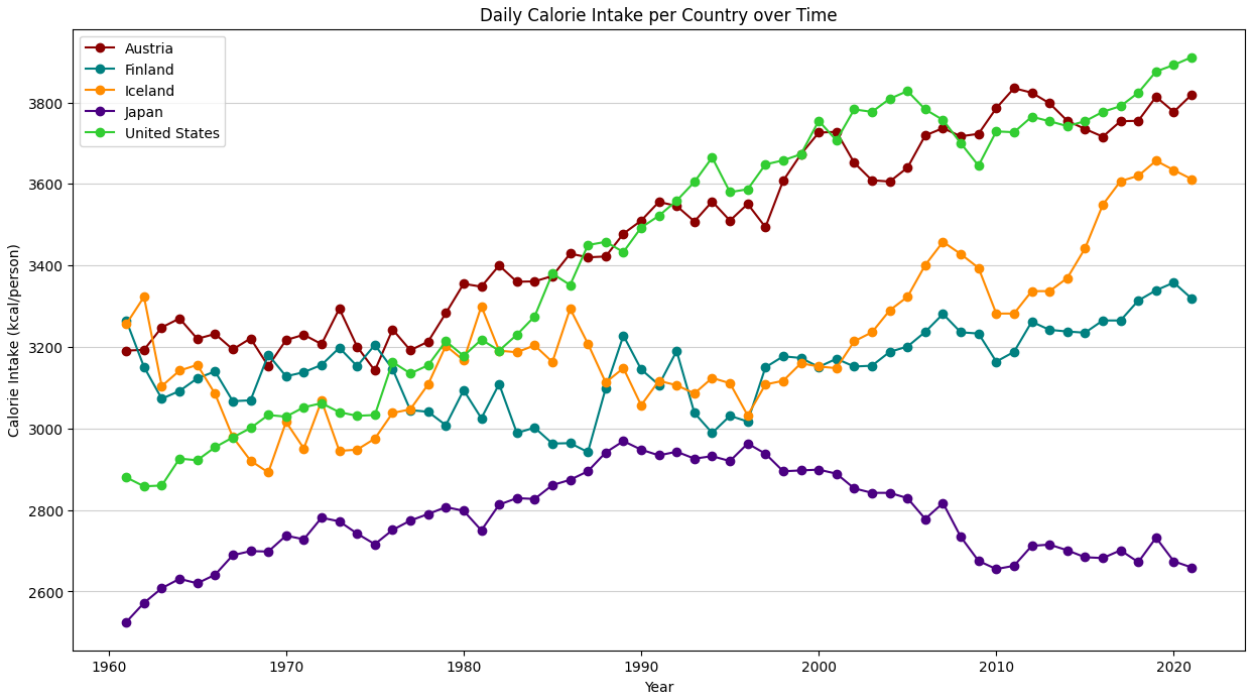
3.1 Line Plots

Alcohol Consumption Over Time



- **Data:** Alcohol consumption (liters per person) across Austria, United States, Japan, South Africa, and India from 1960 to 2024.
- **Insights:** Significant differences in alcohol consumption trends are visible:
    - Austria shows a consistently higher level of consumption than the other countries.
    - India exhibits the lowest consumption. However, it has been increasing since the year 2000.
    - The United States and Japan show fluctuating consumption patterns.
    - Africa experienced a sharp decline in consumption between 1999-2000.
    - The recent decline in consumption in the year 2020 in South Africa might be due to an alcohol ban in July 2020 [1].
- **Relevance:** Highlights cultural and policy differences affecting alcohol consumption.

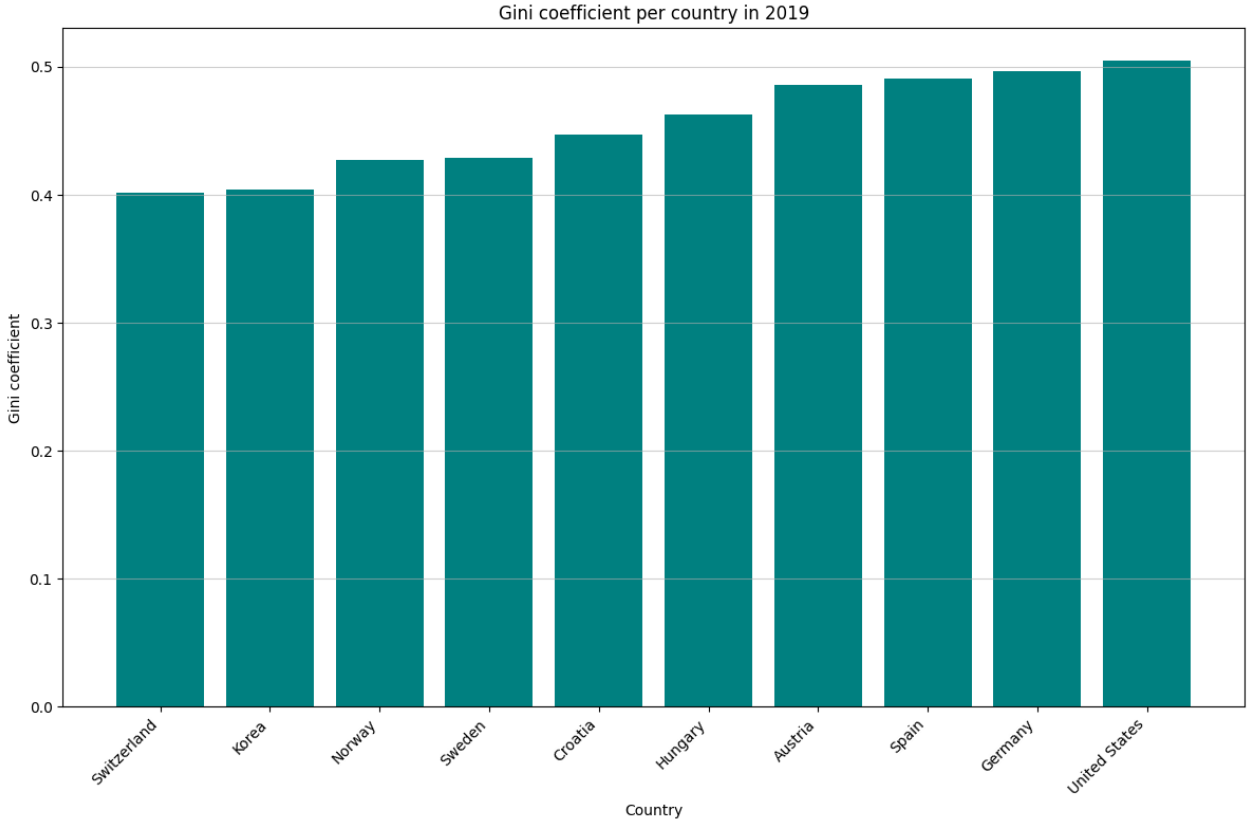
Daily Calorie Intake Over Time



- **Data:** Average daily calorie intake across Austria, Iceland, Finland, United States, and Japan.
- **Insights:**
    - While most countries exhibit a gradual increase in calorie intake over time, Japan maintains a relatively stable trend, reflecting dietary and cultural factors.
    - The daily calorie intake for Austria, Finland, Iceland and the United States is well above the recommended 2600-2800 kcal/day for men and 2000-2200 kcal/day for women [2]. This could explain the rise in obesity rates in these countries.
- **Relevance:** Highlights the trend in daily calorie consumption across countries.

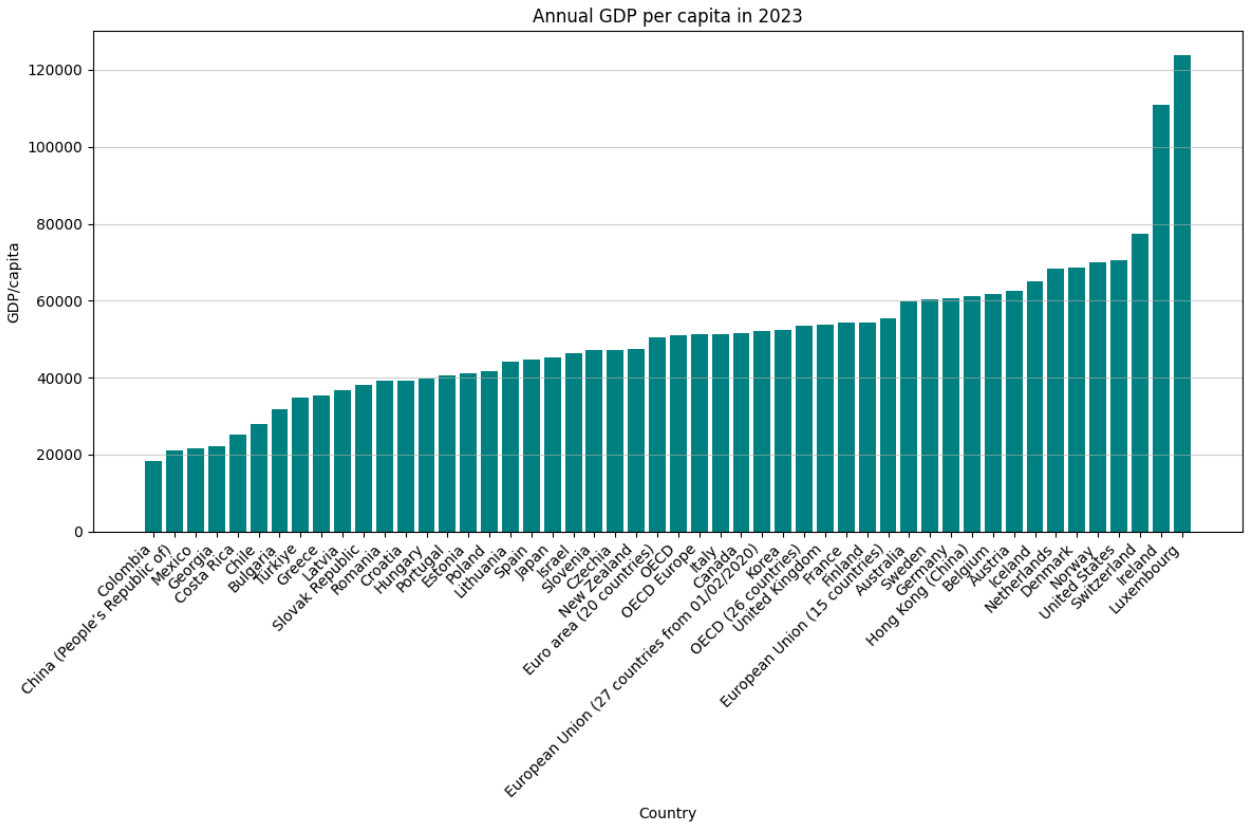
3.2 Bar Charts

Gini Coefficient (2019)



- **Data:** Gini coefficients for income inequality across 11 countries.
- **Insights:**
    - Switzerland, Korea, Norway and Sweden exhibit lower Gini coefficients, indicating better income equality.
    - Conversely, countries like the United States and Germany have higher values indicating a higher income inequality.
- **Relevance:** Demonstrates disparities in income distribution and potentially economic policy effectiveness.

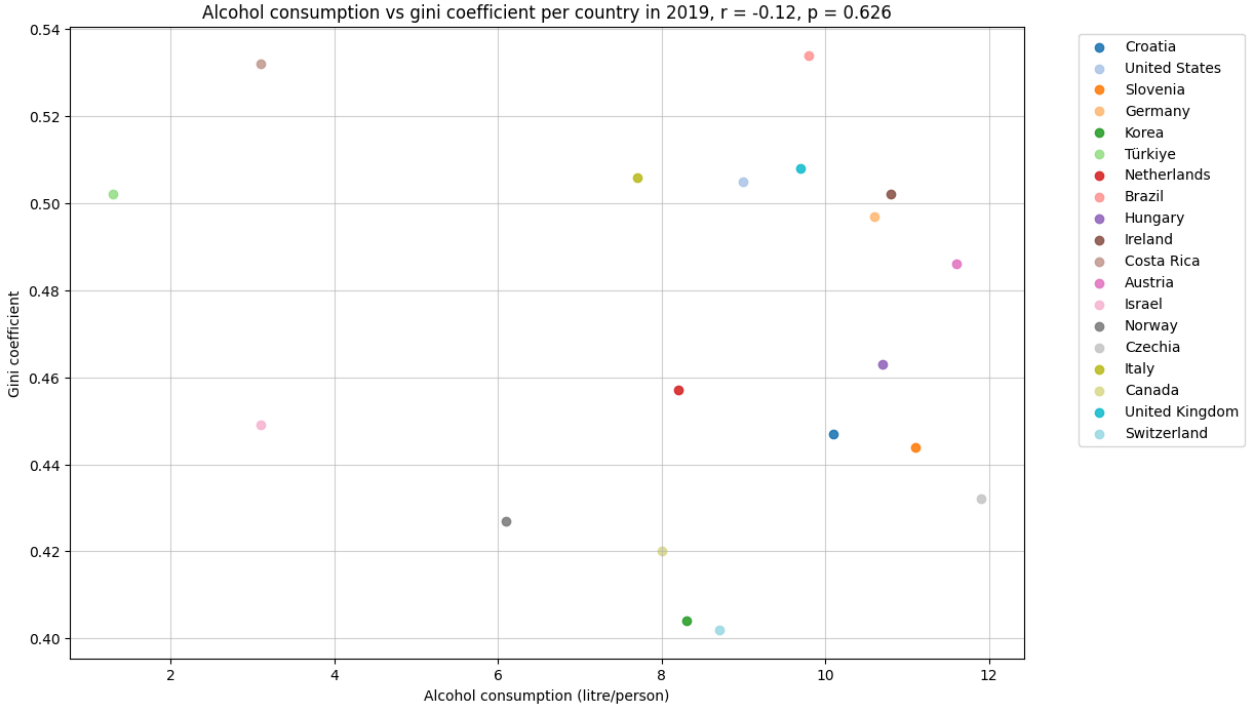
GDP Per Capita (2023)



- **Data:** Annual GDP per capita across a wide range of countries.
- **Insights:**
    - Countries such as Luxembourg and Switzerland stand out with the highest GDP per capita, reflecting strong economies.
    - Developing countries like Colombia and China show significantly lower values.
- **Relevance:** Highlights the economic divide and varying levels of development.

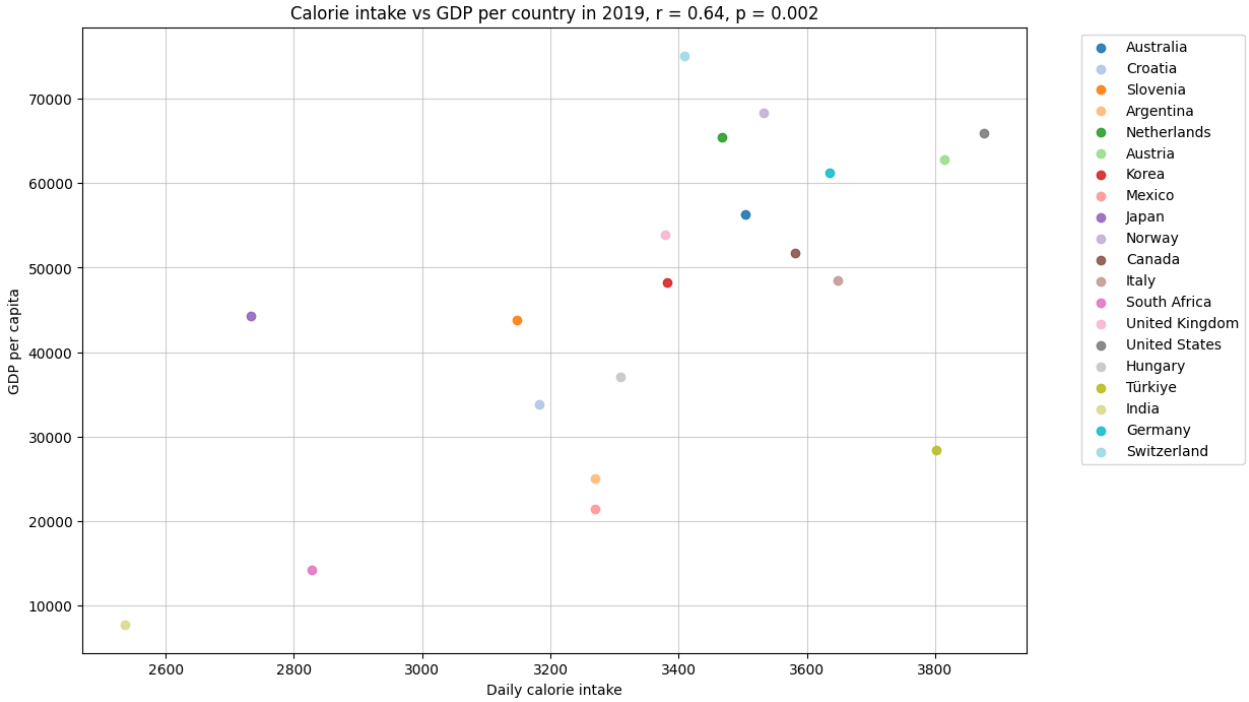
3.3 Scatter Plots

Alcohol Consumption vs Gini Coefficient (2019)



- **Data:** Alcohol consumption (liters/person) and Gini coefficients for income inequality.
- **Insights:**
    - No significant relationship was found between alcohol consumption and the gini-coefficient of a country, $r = -.12$, $p = .626$ (for the calculation pearson correlation was used).
    - Visually this can be seen, because the data points appear randomly distributed across the x and y axis.
    - This suggests that alcohol consumption is not directly influenced by income inequality.
- **Relevance:** This suggests that other factors that weren't analyzed account for the difference between the alcohol consumption.

Calorie Intake vs GDP Per Capita (2019)



- **Data:** Daily calorie intake and GDP per capita for 26 countries.
- **Insights:**
    - Strong positive correlation shows that higher GDP per capita often aligns with higher calorie intake, r = .64, p = .002 (for the calculation pearson correlation was used).
    - This means that countries that do better economically tend to consume more calories.
    - This could be due to the fact that people in economically worse countries might not be able to afford as much food as people from countries with better economic conditions.
    - The relationship appears to be linear. Linear regression could probably be used to model this data and make predictions.
    - Outliers, such as Japan, emphasize the influence of cultural and dietary habits.
- **Relevance:** This highlights the link between economic prosperity and food availability.

4. Conclusion

This analysis provided a comprehensive exploration of social and economic trends through visualizations. The modular and reusable design of the code ensured flexibility and efficiency in handling diverse datasets. The findings reveal some significant patterns, such as the relationship between GDP and calorie intake. Future work could include deeper analyses, such as regression modeling for the relationship between the GDP per capita and daily calorie intake to gain further insights.