**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

Thursday, January 9, 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.

1. **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.
* **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 calculating 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**

* 1. Line Plots

**Alcohol Consumption Over Time**

**Ein Bild, das Text, Screenshot, Reihe, Diagramm enthält.

Automatisch generierte Beschreibung**

* **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 high level of consumption.
  + India exhibits a relatively low and stable trend.
  + The United States and Japan show fluctuating consumption patterns.
* **Relevance:** Highlights cultural and policy differences affecting alcohol consumption.

**Daily Calorie Intake Over Time**

**Ein Bild, das Text, Screenshot, Reihe, Diagramm enthält.

Automatisch generierte Beschreibung**

* **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.
  1. Bar Charts

**Gini Coefficient (2019)**

**Ein Bild, das Text, Screenshot, Reihe, parallel enthält.

Automatisch generierte Beschreibung**

* **Data:** Gini coefficients for income inequality across 11 countries.
* **Insights:** Scandinavian countries (e.g., Norway, Sweden) exhibit lower Gini coefficients, indicating better income equality. Conversely, countries like the United States and Croatia have higher values.
* **Relevance:** Demonstrates disparities in income distribution and economic policy effectiveness.

**GDP Per Capita (2023)**

**Ein Bild, das Text, Reihe, Diagramm, Screenshot enthält.

Automatisch generierte Beschreibung**

* **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 India and South Africa show significantly lower values.
* **Relevance:** Highlights the economic divide and varying levels of development.
  1. Scatter Plots

**Alcohol Consumption vs Gini Coefficient (2019)**

**Ein Bild, das Text, Screenshot, Diagramm, Zahl enthält.

Automatisch generierte Beschreibung**

* **Data:** Alcohol consumption (liters/person) and Gini coefficients for income inequality.
* **Insights:**
  + Weak correlation (Pearson's r ≈ [calculated value]) indicates no significant relationship.
  + Variations suggest that alcohol consumption is not directly influenced by income inequality.
* **Relevance:** Suggests the role of other factors, such as culture or regulations.

**Calorie Intake vs GDP Per Capita (2019)**

**Ein Bild, das Text, Screenshot, Diagramm, Reihe enthält.

Automatisch generierte Beschreibung**

* **Data:** Daily calorie intake and GDP per capita for 26 countries.
* **Insights:**
  + Positive correlation (Pearson's r ≈ [calculated value]) shows that higher GDP per capita often aligns with higher calorie intake.
  + Outliers, such as Japan, emphasize the influence of cultural and dietary habits.
* **Relevance:** Highlights the link between economic prosperity and food availability/nutrition.

**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 significant patterns, such as the relationship between GDP and calorie intake and variations in alcohol consumption across countries. Future work could include deeper analyses, such as regression modeling or time-series forecasting, to gain further insights.