

**FROM ALCOHOL TO HEALTH: CONSUMPTION, CONSEQUENCES, CONTROL**

**Link to Mercury website**

**https://mercury.swin.edu.au/cos30045/s104169617/index.html**

Due Date: Nov 24

Word count: 2731 words

Tutor:Dr Hoang Nhu Vinh

Tutorial session:

Friday, 13:00 – 17:00

Students:

* Nguyen Thi Thanh Minh

(104169617)

* Nguyen Sy Thanh Trung

(104224529)

SWINBURNE UNIVERSITY OF TECHNOLOGY

**COS30045 – Data Visualisation**

*Semester September, 2024*

Process Book

Table of Contents

[I. Introduction 2](#_Toc183549095)

[Background and Motivation 2](#_Toc183549096)

[Key Areas for Analysis 2](#_Toc183549097)

[Objectives 3](#_Toc183549098)

[Target Audience 3](#_Toc183549099)

[Visualization Purpose 4](#_Toc183549100)

[2.1 What are the global patterns of alcohol consumption by type (wine, beer, spirits)? 4](#_Toc183549101)

[2.2 How do alcohol consumption patterns differ by demographic (gender, age) across countries? 5](#_Toc183549102)

[2.3 How does alcohol consumption correlate with health and social consequences (e.g., deaths, diseases, disorders, traffic accidents) by country and gender? 5](#_Toc183549103)

[How have the negative impacts of alcohol (e.g., deaths, disorders, diseases, traffic accidents) changed over time? 6](#_Toc183549104)

[What alcohol-related policies do countries have in place (e.g., taxes, prices, restrictions), and how do they vary globally? 8](#_Toc183549105)

[II. Data 10](#_Toc183549106)

[1. Data Sources 10](#_Toc183549107)

[2. Data Processing 10](#_Toc183549108)

[III. Visualisation Design 12](#_Toc183549109)

[IV. Validation 15](#_Toc183549110)

[1. Evaluation Objectives 15](#_Toc183549111)

[2. Evaluation Methodology 16](#_Toc183549112)

[2.1 Participants 16](#_Toc183549113)

[2.2 Data Overview 16](#_Toc183549114)

[2.3 Validation Steps 16](#_Toc183549115)

[2.4 Recommendations 17](#_Toc183549116)

[V. Conclusion 17](#_Toc183549117)

[VI. References 17](#_Toc183549118)

# Introduction

## **Background and Motivation**

### **Key Areas for Analysis**

1. **Alcohol Consumption Patterns**

* *By Type:* Analyze preferences for wine, beer, and spirits across different countries.
* *By Demographics:* Break down consumption by gender and age groups to uncover patterns in different population segments.

1. **Consequences of Alcohol Consumption**

* *Health Outcomes:* Highlight alcohol-related health issues like liver disease, cardiovascular problems, and addiction disorders.
* *Social and Economic Impact:* Include data on alcohol-related traffic accidents, crime rates, and economic costs, emphasizing the broader impact on society.

1. **Policies and Controls**

* *Taxation and Pricing Policies:* Show how alcohol taxes affect consumption rates, comparing countries with higher vs. lower taxes.
* *Regulations:* Include any bans, minimum age laws, or restricted sales hours, and analyze their effectiveness in reducing consumption or harmful outcomes.

### O**bjec**tives

**1. Analyze Alcohol Consumption Patterns**

* Provide an overview of global and demographic patterns in alcohol consumption by breaking down data by country, type of alcohol, gender, and age. This will highlight where consumption is highest, what types of alcohol are most popular, and which demographics are more inclined to consume alcohol.

**2. Investigate Health and Social Consequences**

* Examine the impact of alcohol consumption on health outcomes and social issues, such as liver disease, traffic accidents, and mental health disorders. This will help illustrate the potential risks associated with different levels and patterns of alcohol use.

**3. Evaluate the Effectiveness of Alcohol Control Policies**

* Show how various countries’ policies (e.g., taxes, sales restrictions, advertising bans) align with consumption rates and health outcomes. By visualizing these relationships, the project can identify potential policy measures that correlate with reduced consumption and better health outcomes.

**4. Track Changes Over Time**

* Analyze how alcohol consumption trends have evolved over recent decades and assess the impact of major events (e.g., the COVID-19 pandemic) on these trends. This helps in understanding the dynamic nature of alcohol consumption and its changing impact on health and society.

**5. Provide an Interactive, User-Friendly Exploration Tool**

* **Objective:** Develop a website that hosts these visualizations interactively, allowing users to explore specific questions or trends of interest by filtering by country, type of alcohol, demographic group, or year. The goal is to make data accessible and engaging, enhancing users' understanding of alcohol as a lifestyle factor.

### Target Audience

* **Public Health Officials and Policymakers:** Understanding the scope of alcohol-related health issues and how policy measures might reduce negative outcomes. The visualizations offer data-driven insights into effective policies and highlight areas for intervention.
* **Researchers and Academics in Public Health and Social Sciences:** Analyzing cross-country and demographic comparisons to study the impact of alcohol consumption on public health. This project provides a rich dataset and an accessible way to explore associations between alcohol use and health outcomes.
* **Nonprofit Organizations and Health Advocates:** Supporting advocacy for healthier lifestyle choices and policies that reduce harmful alcohol consumption. The project data can help these organizations create targeted campaigns or policy recommendations.
* **General Public and Health-Conscious Individuals:** Gaining awareness of how alcohol consumption patterns and risks vary across regions and demographics. Interactive charts offer a digestible way to learn about the health risks associated with alcohol and explore data by country or demographic.
* **Media and Data Journalists:** Accessing clear and interactive visuals to report on alcohol consumption trends, health impacts, and policies. The data can support stories on lifestyle trends, health risks, or policy effectiveness.

## Visualization Purpose

### 2.1 What are the global patterns of alcohol consumption by type (wine, beer, spirits)?

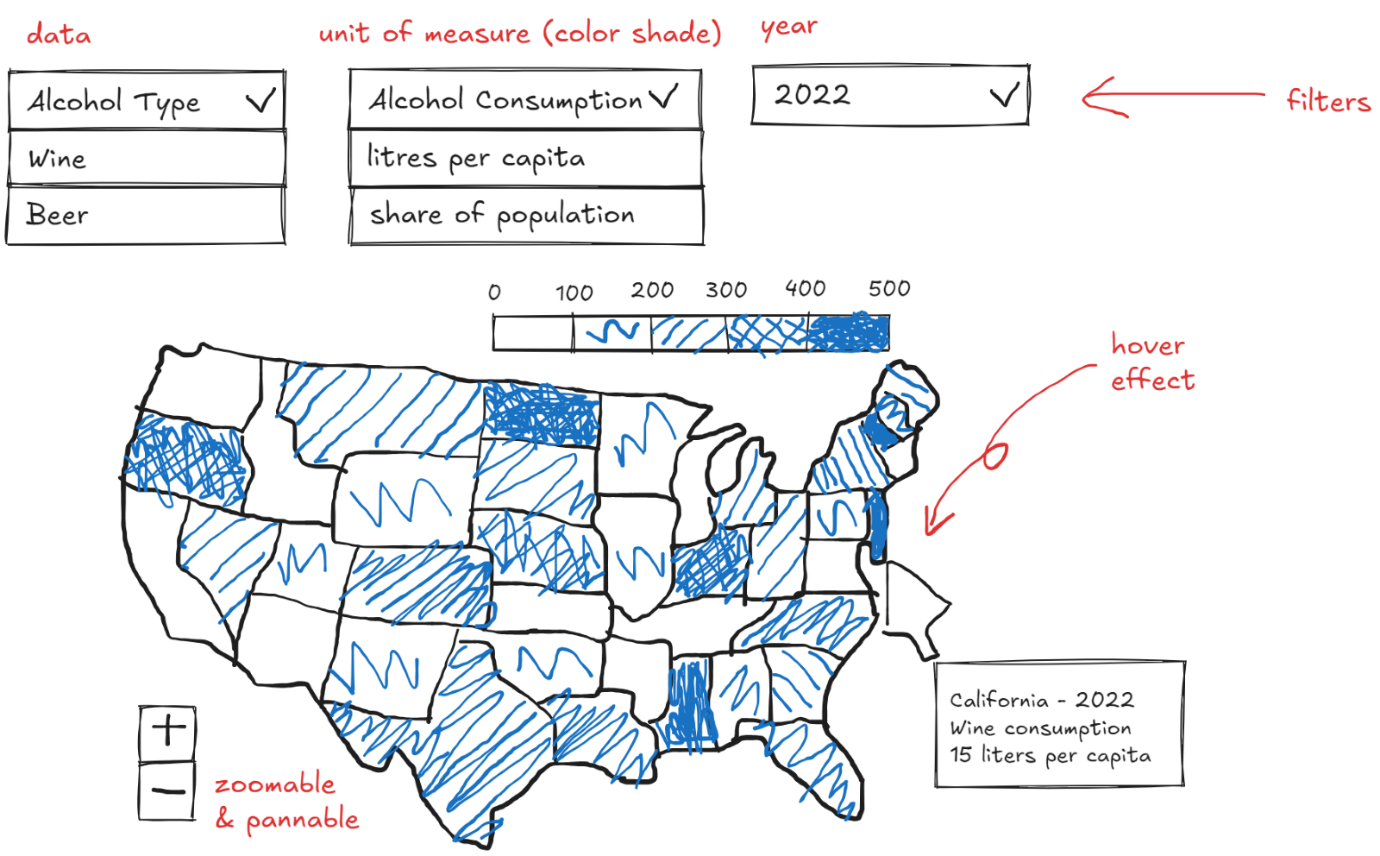


Figure 1. Choropleth Map

* **Chart Type:** **Interactive Choropleth Map**
* **Purpose:** This map will offer a visual comparison of alcohol consumption rates across countries by type. Users will see which regions favor certain types of alcohol, helping reveal cultural preferences and consumption patterns.
* **Design and Interactivity:**
* **Filter Options:** Filters for year and type of alcohol allow users to switch between wine, beer, and spirits and view how each is distributed globally over time.
* **Tooltip:** When hovering over a country, a tooltip shows specific consumption data (e.g., “Wine Consumption: 15 liters per capita”).
* **Color Scale:** A gradient color scale (e.g., from light to dark) visually represents the consumption rate, highlighting countries with high and low alcohol use.
* **Insights:** This chart reveals which countries have high or low alcohol consumption rates for different types, helping users identify patterns across continents and pinpoint regions with specific preferences.

### 2.2 How do alcohol consumption patterns differ by demographic (gender, age) across countries?

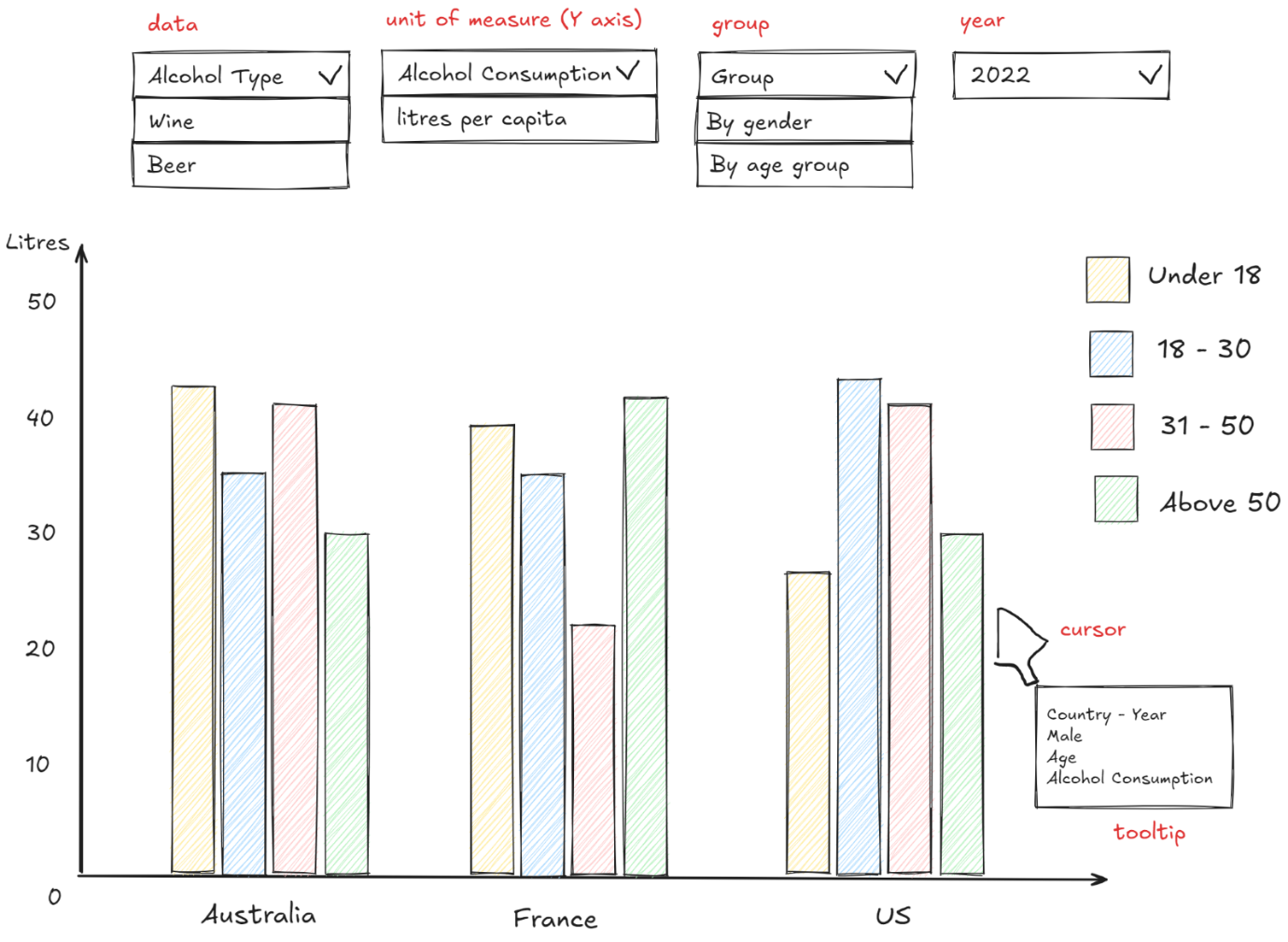


Figure 2. Stacked Bar Chart

* **Chart Type:** **Grouped / Stacked Bar Chart**
* **Purpose:** The stacked bar chart will break down alcohol consumption by age and gender within each country, enabling comparisons across demographics.
* **Design and Interactivity:**
* **Filter Options:** Users can filter by year and type of alcohol, comparing demographic data within and across countries.
* **Group Options:** Users can switch between types of grouping, by age group or by gender.
* **Tooltip:** Hovering over each bar segment shows specific values, e.g., “Female, 18-25, Wine Consumption: 5 liters per capita.”
* **Color Coding:** Different colors represent demographic groups, making it easy to distinguish age and gender distributions.
* **Chart Type:** Users can choose between a grouped bar chart or stacked bar chart, with smooth transitions between the 2 types.
* **Insights:** This chart helps users understand which demographics consume more alcohol, showing if certain age groups or genders tend to favor specific types of alcohol.

### How does alcohol consumption correlate with health and social consequences (e.g., deaths, diseases, disorders, traffic accidents) by country and gender?

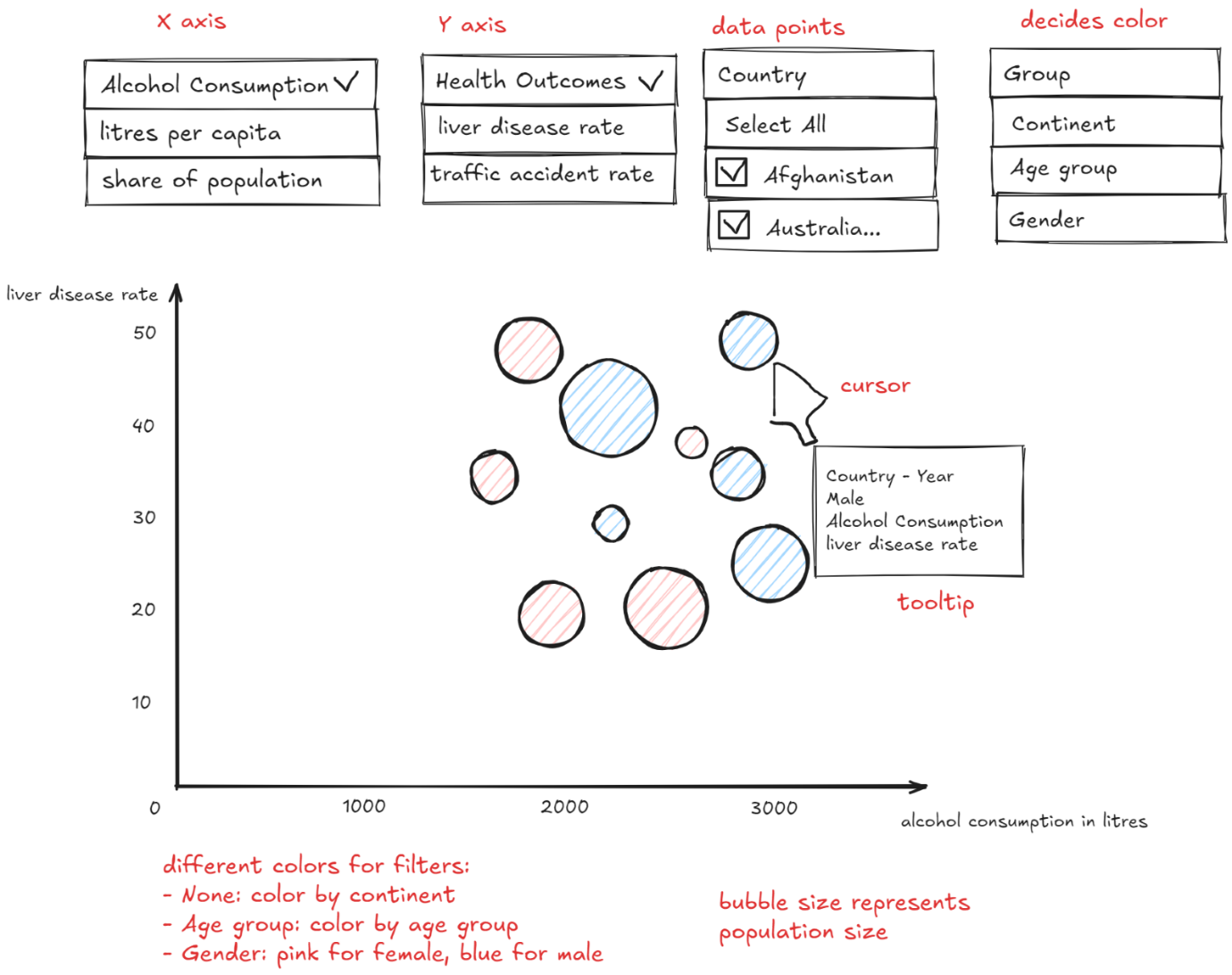


Figure 3. Bubble Chart

* **Chart Type:** **Interactive Bubble Chart**
* **Purpose:** This bubble chart will show the relationship between alcohol consumption and various health outcomes across countries and genders, allowing for a nuanced look at health impacts.
* **Design and Interactivity:**
* **Axes Options:** Users can select different health outcomes (e.g., liver disease rate or traffic accident rates) on the y-axis and alcohol consumption on the x-axis.
* **Filters for Time and Gender:** Choose to view data for a particular year or by gender to examine specific trends.
* **Bubble Size and Color:** Bubble size represents population size, and colors represent gender, making it easy to compare countries and genders.
* **Insights:** This visualization reveals correlations between alcohol consumption and health consequences. Users can identify patterns, such as whether high alcohol consumption coincides with high disease or accident rates, and how these differ by gender and country.

### How have the negative impacts of alcohol (e.g., deaths, disorders, diseases, traffic accidents) changed over time?

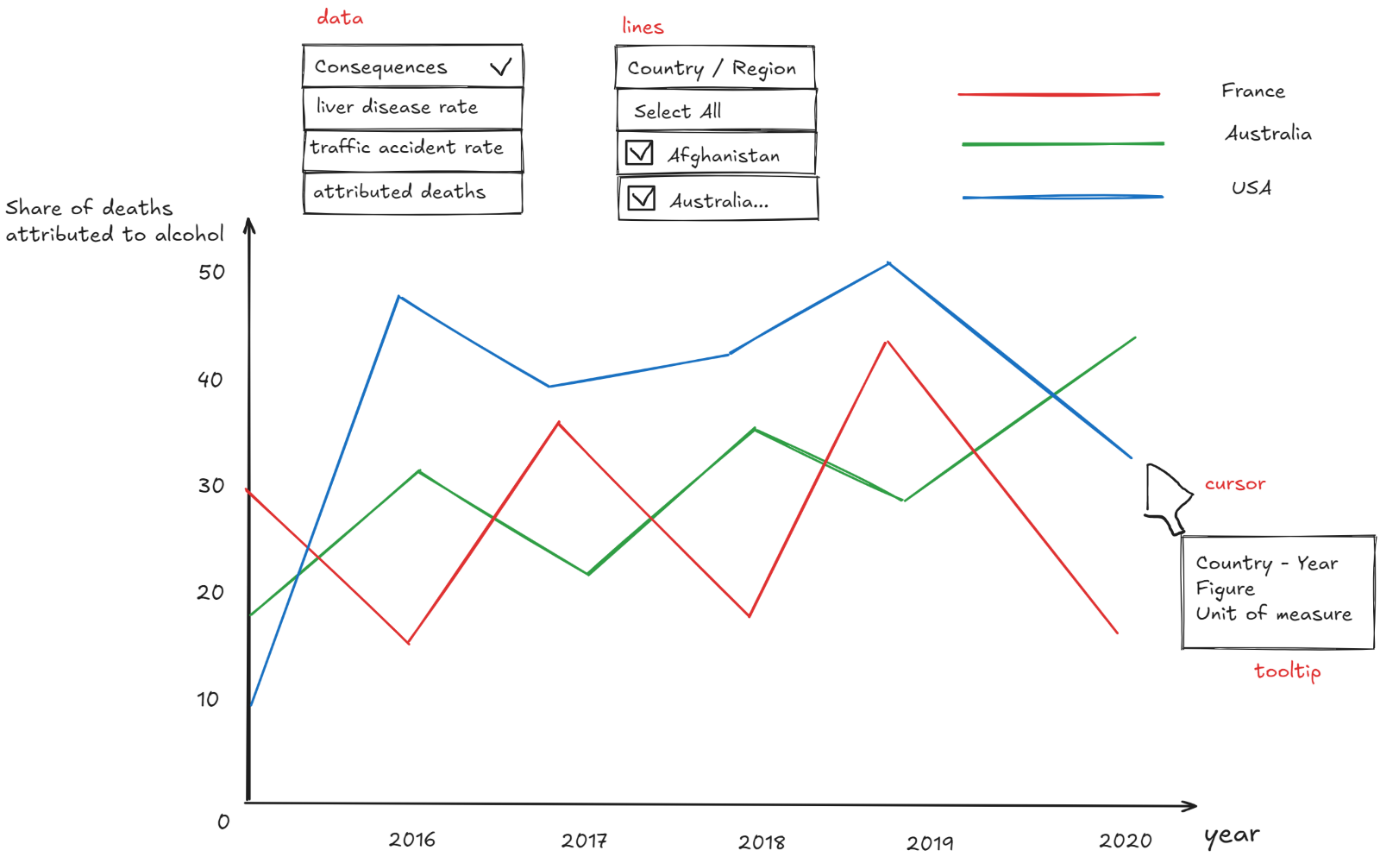


Figure 4. Line Chart

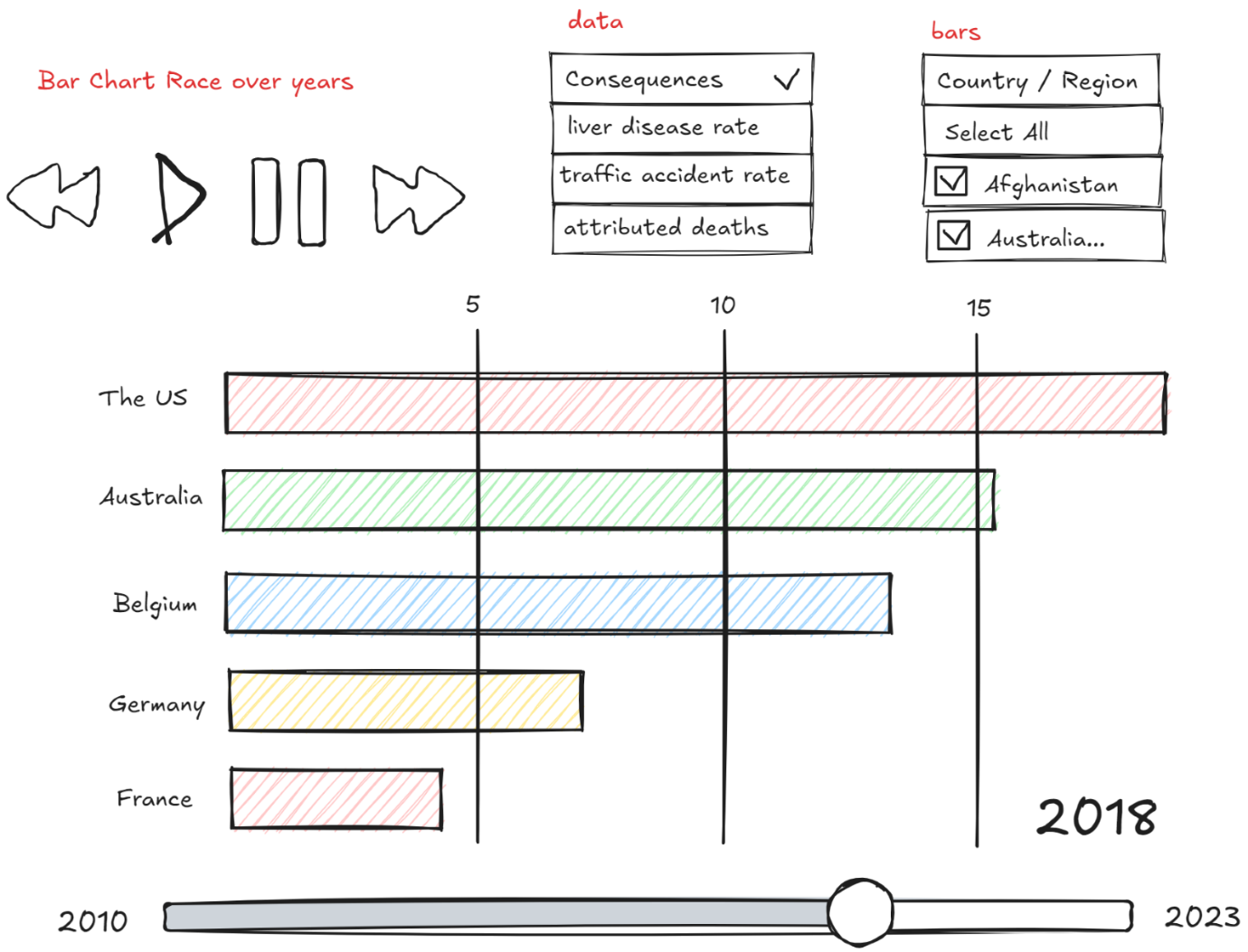


Figure 5. Bar Chart Race

* **Chart Type:** **Line Chart / Bar Chart Race**
* **Purpose:** This chart will show the trends in health and social consequences linked to alcohol use over the years, illustrating if these impacts are worsening, improving, or stabilizing.
* **Design and Interactivity:**
* **Filter Options:** Allow users to choose specific impacts (e.g., alcohol-related deaths, traffic accidents) or view all in one chart for broader trends. Filters for country and year can further refine the view.
* **Hover Tooltips:** Display exact data values for each year, giving users detailed insight into the intensity of impacts over time.
* **Color-Coded Lines:** Each impact type has a unique color, making it easy to distinguish trends.
* **Bar Chart Race Design:** Allow users to switch between views.
* Country Selection: Consider an option for viewers to highlight or focus on specific countries, so they can track selected nations’ positions over time.
* Transition Speed Control: Provide a speed control to adjust the animation speed (pause/resume, faster, slower...), letting users analyze trends at their preferred pace.
* **Insights:** This visualization highlights which impacts are most severe, reveals if alcohol-related issues are on the rise, and shows how outcomes have shifted, possibly in response to changing consumption rates or policies.

### What alcohol-related policies do countries have in place (e.g., taxes, prices, restrictions), and how do they vary globally?

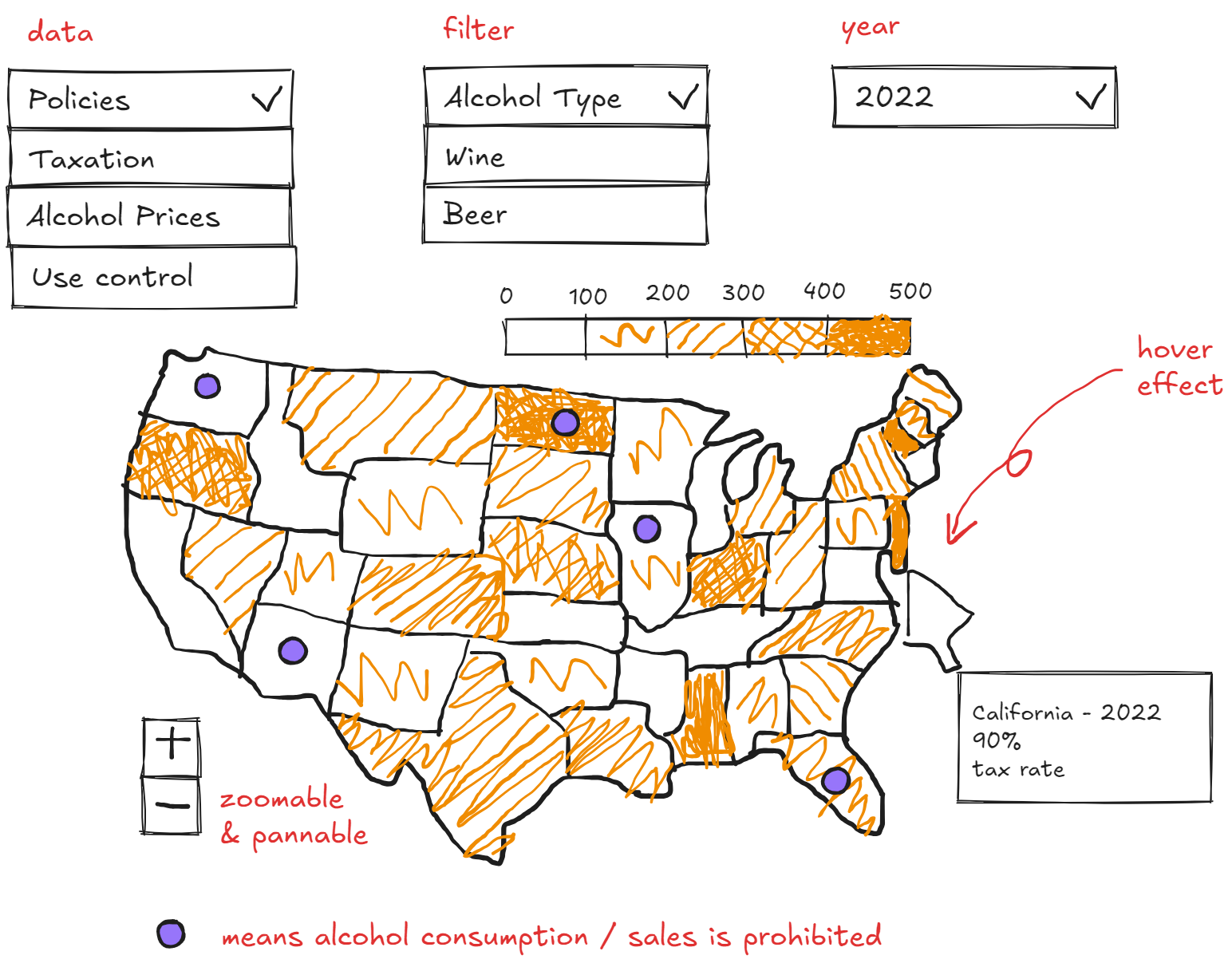


Figure 6. Geo Map (Choropleth or Bubble Overlay)

* **Chart Type:** **Geo Map (Choropleth or Bubble Overlay)**
* **Purpose:** The geo chart will provide a comparative view of alcohol-related policies across countries, showing which countries have higher taxes, bans, or restrictions on advertising.
* **Design and Interactivity:**
* **Filter Options:** Filters for different types of policies (tax rates, prices, restriction levels) let users explore each policy type separately.
* **Tooltip:** When hovering over a country, the tooltip shows detailed policy information (e.g., “Tax Rate: 15%, Ad Restriction: Moderate”).
* **Color Coding or Bubble Size:** For example, darker shades can represent higher taxes, or bubble size could indicate price levels.
* **Insights:** This chart helps users quickly see how policies vary worldwide, understand which regions have stricter or more lenient approaches, and explore potential links between policy strictness and consumption or health outcomes.

# Data

## Data Sources

* [Non-medical determinants of health](https://www.oecd-ilibrary.org/social-issues-migration-health/data/oecd-health-statistics/oecd-health-data-non-medical-determinants-of-health_data-00546-en?parentId=http%3A%2F%2Finstance.metastore.ingenta.com%2Fcontent%2Fcollection%2Fhealth-data-en) (OECD Health Statistics, including 2023)
* [Alcohol consumption](https://data-explorer.oecd.org/vis?fs%5b0%5d=Topic%2C1%7CHealth%23HEA%23%7CRisk%20factors%20for%20health%23HEA_RIS%23&pg=0&fc=Topic&bp=true&snb=7&df%5bds%5d=dsDisseminateFinalDMZ&df%5bid%5d=DSD_HEALTH_LVNG%40DF_HEALTH_LVNG_AC&df%5bag%5d=OECD.ELS.HD&df%5bvs%5d=1.0&dq=.A.....&pd=2010%2C&to%5bTIME_PERIOD%5d=false&vw=tb) (OECD Health, including 2023)
* [Prevalence of weekly drinking by age and sex](https://www.oecd-ilibrary.org/sites/6e4b4ffb-en/1/3/2/index.html?itemId=/content/publication/6e4b4ffb-en&_csp_=2c8abebfcd351c6cf1626284896818ce&itemIGO=oecd&itemContentType=book#figure-d1e6539) (OECD analysis)
* [Prevalence of monthly binge drinking by age and sex](https://www.oecd-ilibrary.org/sites/6e4b4ffb-en/1/3/2/index.html?itemId=/content/publication/6e4b4ffb-en&_csp_=2c8abebfcd351c6cf1626284896818ce&itemIGO=oecd&itemContentType=book#figure-d1e6575) (OECD analysis)
* [Global Burden of Disease 2021](https://vizhub.healthdata.org/gbd-results/) (IHME, for consequences)
* [Global Information System on Alcohol and Health](https://apps.who.int/gho/data/node.main.GISAH?lang=en) (WHO, old version)
* [Global Information System on Alcohol and Health](https://www.who.int/data/gho/data/themes/global-information-system-on-alcohol-and-health) (WHO, new version)
* [Global prices and taxes on alcoholic beverages](https://www.who.int/data/gho/data/themes/topics/taxes-on-alcoholic-beverages) (WHO)
* [World Development Indicators](https://datacatalog.worldbank.org/search/dataset/0037712/World-Development-Indicators) (World Bank)

## Data Processing

* Used Python Jupyter Notebooks with the pandas package to handle data cleaning and aggregation.
* Removed unwanted columns, sorted data, and renamed columns to ensure consistency.
* Standardized country names using ISO 3166 codes, with adjustments for excessively long names (e.g., "United States of America" → "United States").
* Processed data into both wide and long formats for visualization flexibility, depending on the chart requirements (e.g., choropleth maps, stacked bar charts).
* Designed mockups for key visualizations, including:
* Geo chart for alcohol consumption by type with hover effects
* Bubble chart showing alcohol consumption’s relationship with health outcomes, including liver cirrhosis
* Method: Using Python Jupyter Notebooks with the pandas package to handle CSV
* Basic steps:
* Filter out unwanted columns
* Sort the data in desired order
* Change column names into consistent conventions
* Note:
* Country full name (e.g. Croatia, Australia...) columns are omitted, only their codes are kept. This helps reduce file size significantly and keeps data across files consistent (data from different sources may record different country names, but country codes are internationally defined).
* In return, there needs to be a common “dictionary” file to map codes to full names and their continent if needed.  The chosen convention is [ISO 3166](https://github.com/lukes/ISO-3166-Countries-with-Regional-Codes/blob/master/all/all.csv). However, there are some countries with very long full names, so we decided to customize the convention and shorten them. Specifically:

|  |  |
| --- | --- |
| **Full name by ISO 3166** | **Shortened name** |
| United Kingdom of Great Britain and Northern Ireland | United Kingdom |
| United States of America | United States |
| Bolivia, a Plurinational State of | Bolivia |
| Congo, Democratic Republic of the | DR Congo |
| Iran, Islamic Republic of | Iran |
| Korea, Democratic People's Republic of | North Korea |
| Korea, Republic of | South Korea |
| Lao People's Democratic Republic | Laos |
| Micronesia, Federated States of | Micronesia |
| Moldova, Republic of | Moldova |
| Netherlands, Kingdom of the | Netherlands |
| Palestine, State of | Palestine |
| Taiwan, Province of China | Taiwan |
| Tanzania, United Republic of | Tanzania |
| Türkiye | Turkey |
| Venezuela, the Bolivarian Republic of | Venezuela |

* Data aggregation: 2 types
* Wide format: A single row with multiple columns

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Country** | **Year** | **AgeGroup\_Under30** | **AgeGroup\_30\_44** | **AgeGroup\_45\_59** | **AgeGroup\_60+** |
| ABC | 2020 | 10.5 | 20.0 | 15.3 | 25.7 |

* **Visualization Type**: Stacked bar charts, grouped bar charts, or line charts where we need to compare multiple attributes (e.g., age groups) side-by-side or on the same axis. If planning to keep age groups as separate series in your visualization, the wide format is more intuitive.
* **Ease of Access**: When attributes like age groups naturally belong together for a single "entity" (e.g., a country or gender) and need to be processed together.
* Long format: separate rows for each value

|  |  |  |  |
| --- | --- | --- | --- |
| **Country** | **Year** | **Age Group** | **Value** |
| ABC | 2020 | Under30 | 10.5 |
| ABC | 2020 | 30\_44 | 20.0 |
| ABC | 2020 | 45\_59 | 15.3 |
| ABC | 2020 | 60+ | 25.7 |

* **Visualization Type**: Interactive charts like a choropleth, bar chart races, or scatter plots where one axis represents a single attribute (e.g., age group) across entities or time. Allows easier filtering, grouping, and transformations in both D3 and preprocessing tools like Python.
* **Ease of Access**: If we expect to filter by age group, having them as rows is more convenient.
* Full explanation by ChatGPT: [GitHub](https://github.com/minh-swinburne/cos30045-project/issues/12#issuecomment-2481334260)
* Example: [process.ipynb](https://liveswinburneeduau-my.sharepoint.com/:u:/g/personal/104169617_student_swin_edu_au/EaczP-Ggq9JCjEGkvjaIj8YBCPsmk5Zav6OqgHs4ZYQ_Og?e=I8FrrG" \t "_blank) for question 1

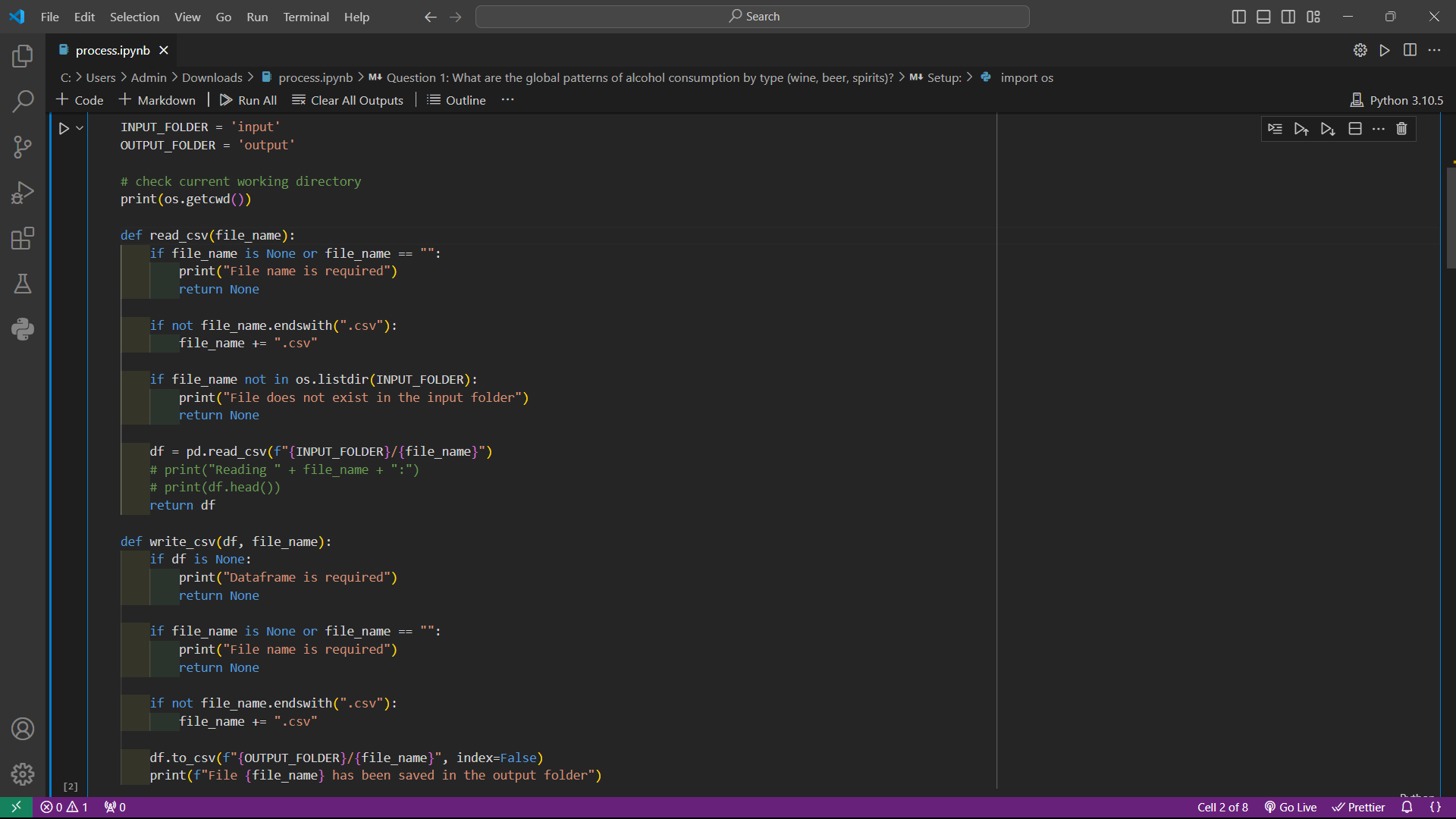


Figure 7 Example data processing for question 1 (Total Alcohol Consumption (OECD, 1960-2023)).

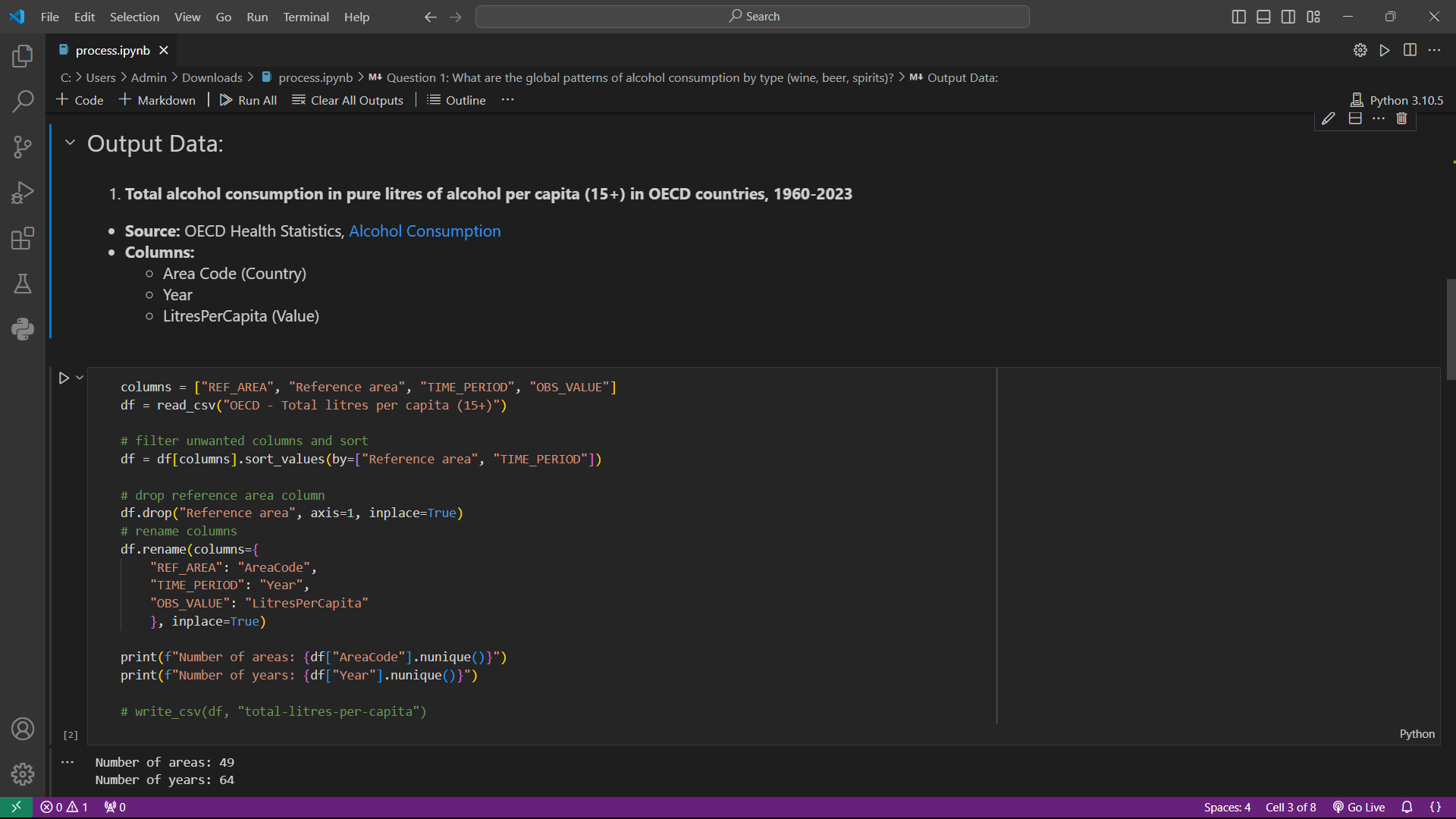


Figure 8. Output data of Total alcohol consumption in pure litres of alcohol per capita (15+) in OECD countries, 1960-2023.

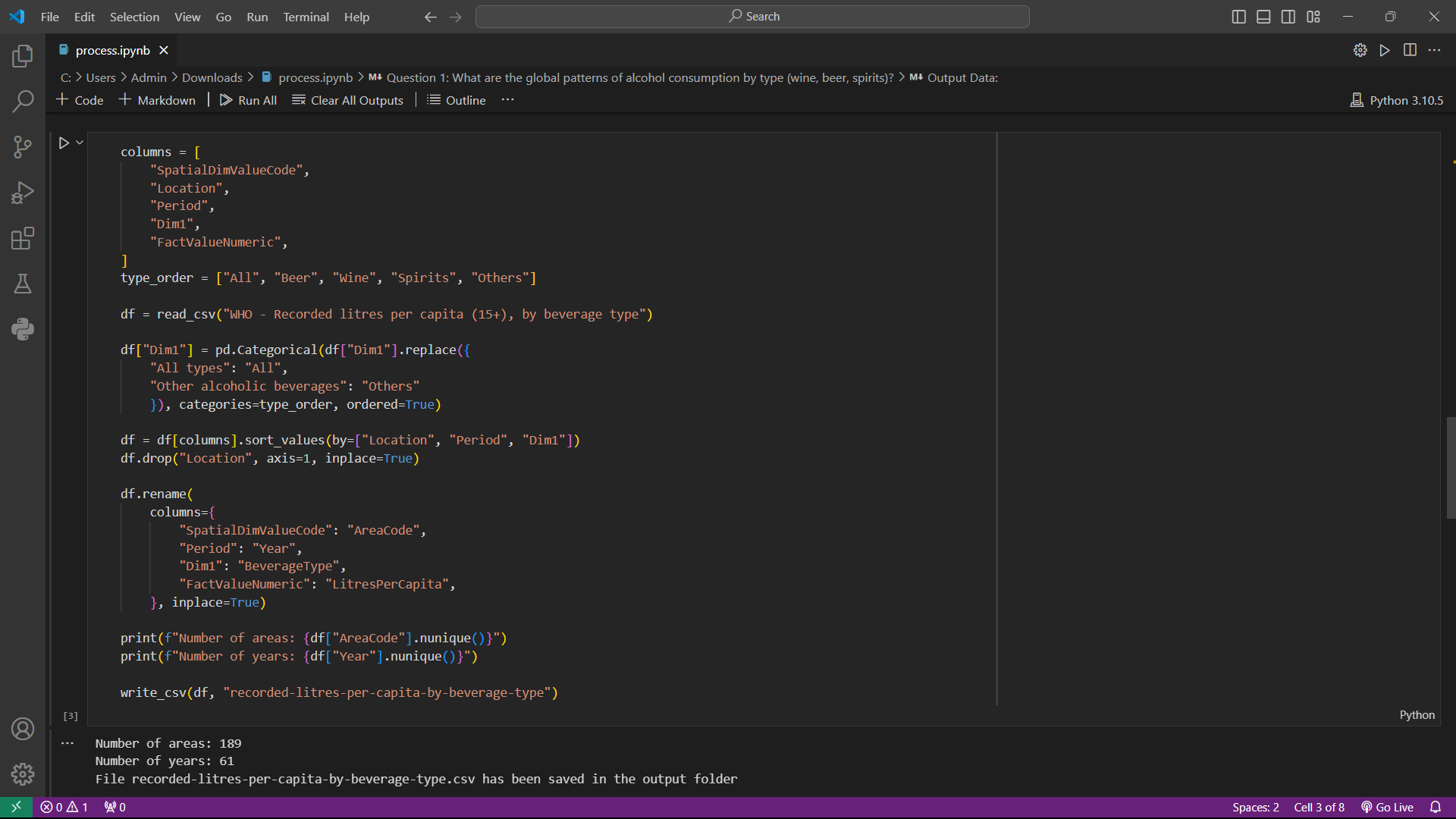


Figure 9. Output data of RECORDED alcohol consumption in pure litres of alcohol per capita (15+), 1961-2020, by beverage type

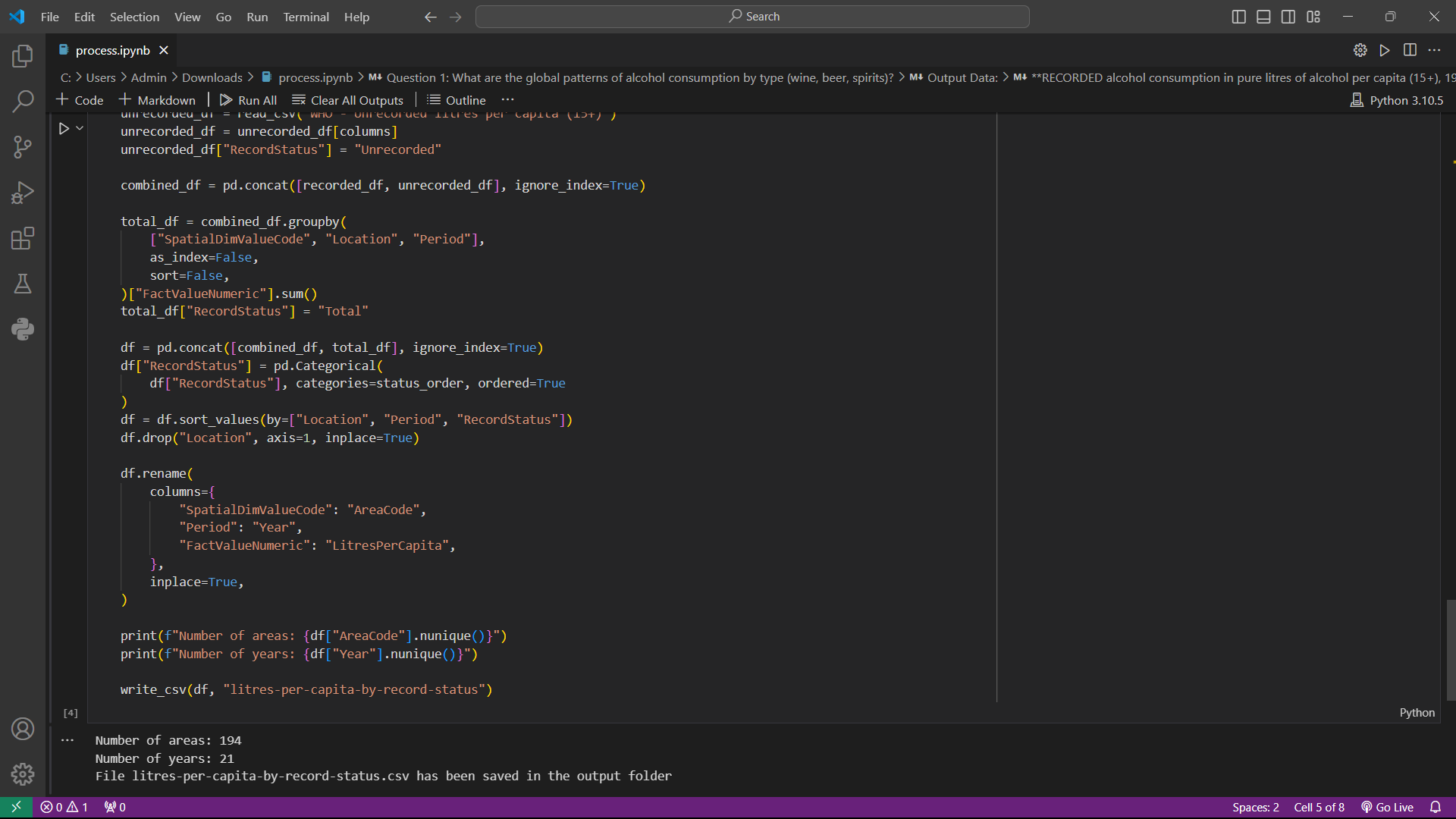


Figure 10. Output data of Alcohol consumption in pure litres of alcohol per capita (15+), 2000-2020, by record status

### 2.1 **Visualization Recommendations**:

* **Dataset 1**: Line chart or area chart to show trends over time in total alcohol consumption for key countries.
* **Dataset 2**: Stacked bar chart to compare beverage types' contributions to alcohol consumption across countries.
* **Dataset 3**: Pie charts or stacked bar charts to highlight recorded vs. unrecorded alcohol proportions by year.

### Storytelling Framework:

1. **Dataset 1: Recorded Litres Per Capita by Beverage Type**
   * Columns: AreaCode, Year, BeverageType, LitresPerCapita.
   * Tracks per capita consumption of various beverage types (Beer, Wine, Spirits, etc.) over time by region.
2. **Dataset 2: Population Consumption, Deaths in 2019 by Cause and Sex**
   * Columns: Cause, AreaCode, Sex, Deaths, LitresPerCapita, Population.
   * Contains data on deaths attributed to causes like liver cirrhosis, broken down by sex and region, with consumption and population statistics.
3. **Dataset 3: Litres Per Capita by Record Status**
   * Columns: AreaCode, Year, LitresPerCapita, RecordStatus.
   * Details total, recorded, and unrecorded beverage consumption by region and year.

### 2.3 Next Steps for Analysis and Storytelling

* **Exploratory Data Analysis (EDA)**:
  + Identify trends in beverage consumption over time across regions.
  + Explore correlations between beverage consumption and health outcomes (e.g., liver cirrhosis).
* **Storytelling Opportunities**:
  + Show how recorded vs. unrecorded consumption affects health outcomes.
  + Compare regional differences in consumption patterns and their impacts on population health.

### **Next Steps**:

* Extract trends specific to alcohol consumption globally and regionally.
* Prepare visualizations based on the above recommendations.

# Visualisation Design

Chart Mockups

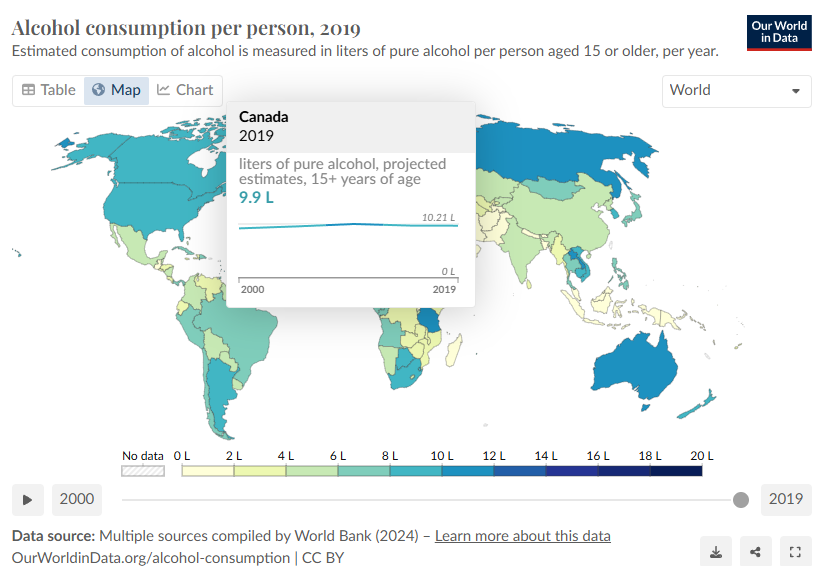


Figure 11. Alcohol consumption geo chart with interactive hover effect by Our World In Data

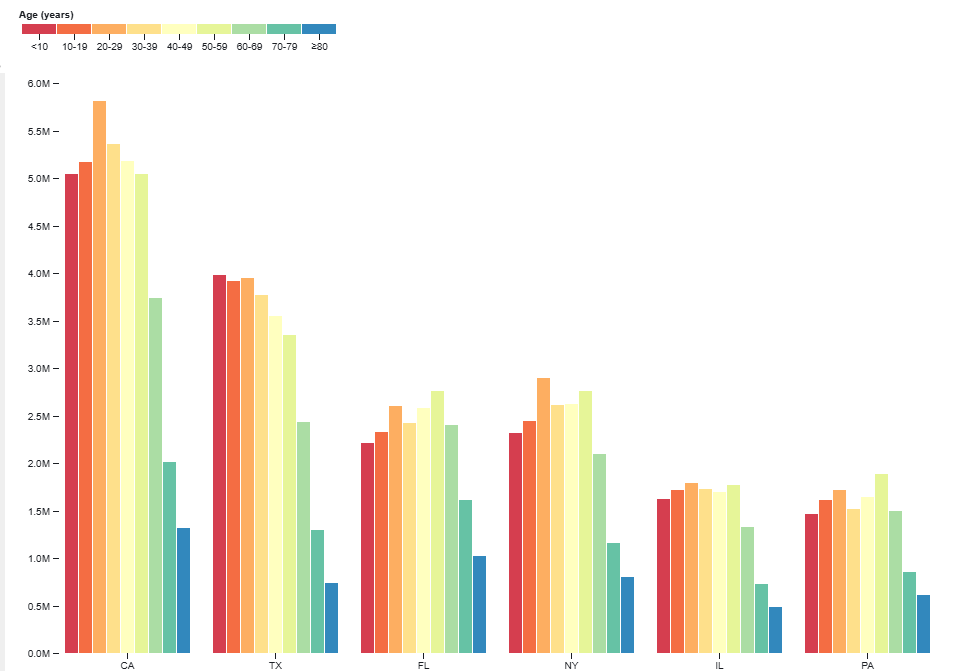


Figure 12. Grouped bar chart example by [D3 Gallery](https://observablehq.com/@d3/grouped-bar-chart/2)

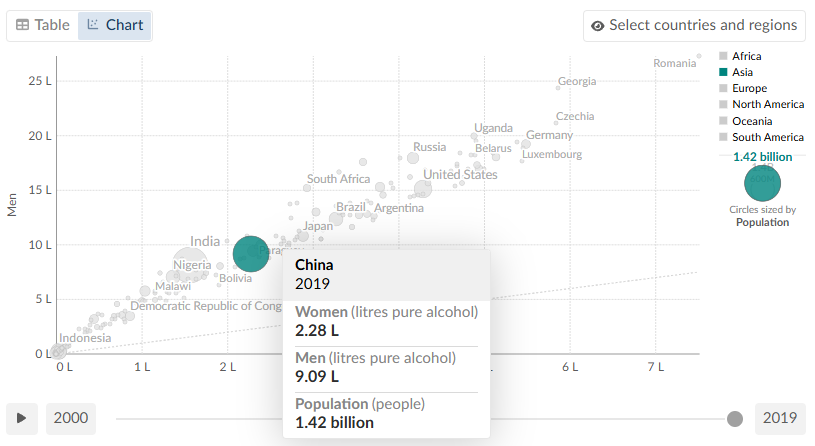


Figure 13. Bubble chart example for question 3

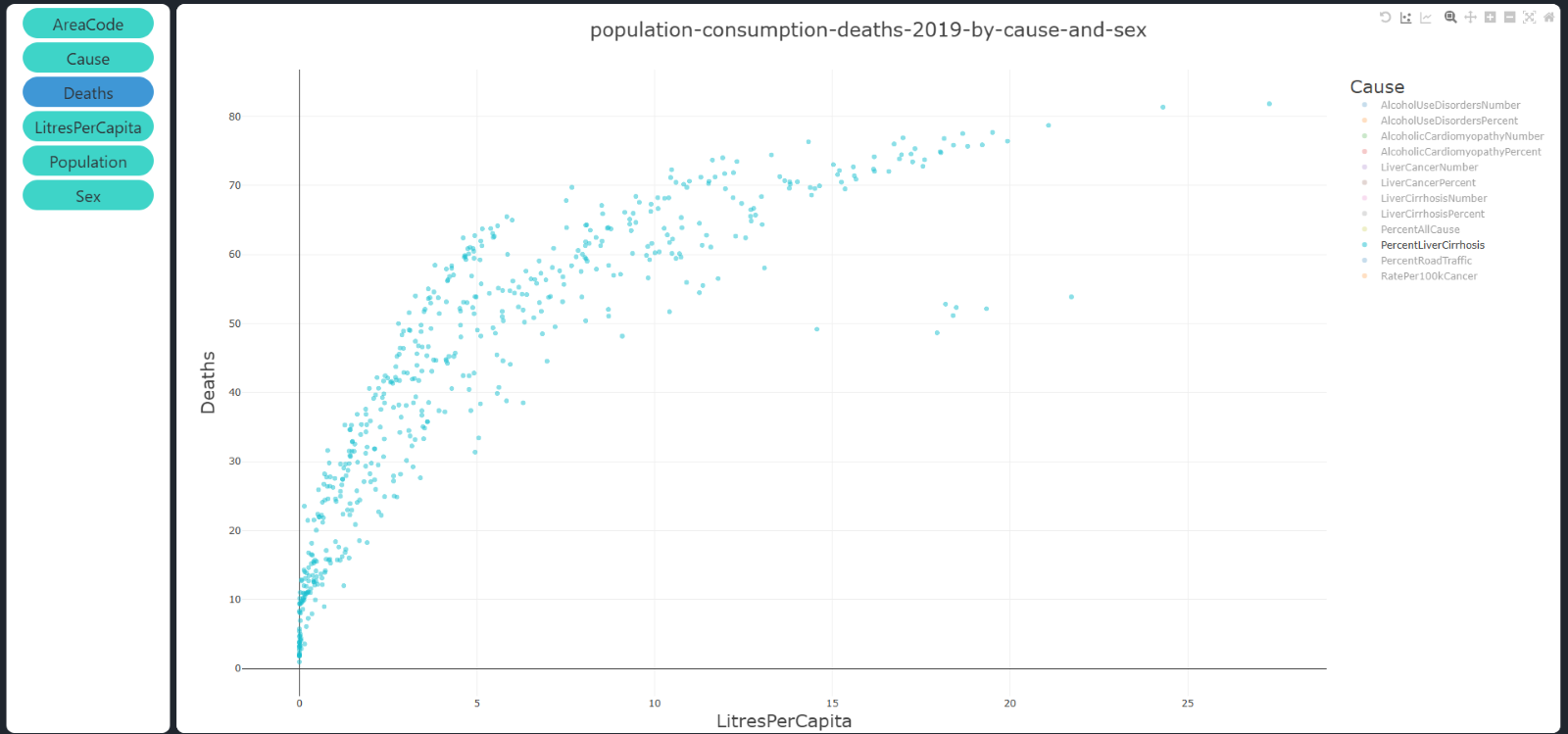


Figure 14. Scatter plot showing the relation between alcohol consumption and liver cirrhosis for question 3, using processed data by csvplot.com

* An additional visualization for comparison of deaths attributed to different risk factors, with those related to alcohol highlighted. Example from [Our World In Data](https://ourworldindata.org/grapher/number-of-deaths-by-risk-factor)

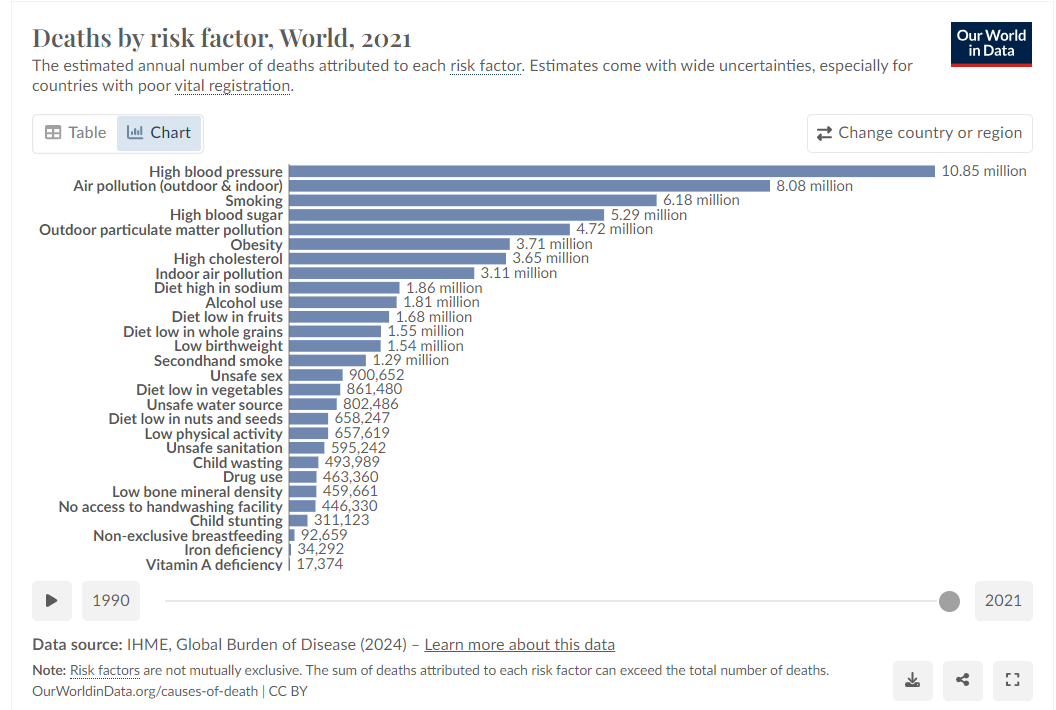


Figure 15. Horizontal bar chart showing deaths by risk factor by Our World in Data

# Validation

## ****Evaluation Objectives****

The objective was to validate whether the visualizations effectively represented and communicated the insights from the **OECD data**. This included assessing:

* The accuracy and clarity of the visualized relationships between alcohol consumption and health outcomes.
* The ability of the visualizations to handle complex, multi-dimensional data.
* The flexibility of the visualizations for analysis across various dimensions, such as countries, age groups, and time.

### ****2.1 Evaluation Methodology****

### 2.2 ****Participants****

We recruited **10 participants** with diverse backgrounds:

* 3 public health professionals
* 2 statisticians/data scientists
* 3 general users with no technical background
* 2 UI/UX designers  
  Participants were chosen to ensure a mix of domain expertise and general usability feedback.

### ****Data Overview****

The evaluation used the following OECD datasets:

* **Non-medical determinants of health (2023)**
* **Alcohol consumption (2023)**
* **Prevalence of weekly drinking by age and sex**
* **Prevalence of monthly binge drinking by age and sex**

Each dataset was preprocessed and formatted into wide and long formats to test the visualizations' compatibility with different data structures.

### ****Validation Steps****

1. **Data Representation Validation**
   * Loaded OECD datasets into the visualizations.
   * Assessed whether the visualizations correctly reflected the key patterns in the data.
   * Verified alignment between raw data values and those represented visually.
2. **Scenario Testing** Specific scenarios were designed to test the visualizations’ ability to address key questions:
   * Identify trends in alcohol consumption across countries over time.
   * Compare the prevalence of binge drinking by age group.
   * Analyze the correlation between alcohol consumption and liver cirrhosis rates.
3. **Flexibility Testing**
   * Tested interactive features such as hover effects, filtering, and zooming for their responsiveness and ease of use.
   * Assessed whether the visualizations could effectively display data from both wide and long formats.
4. **Performance Assessment**
   * Evaluated the visualizations’ performance when handling large datasets.
   * Measured load times and responsiveness during interactions.

### ****2.5 Recommendations****

1. **Tooltip Refinement:** Improve the accuracy of tooltips for aggregated data.
2. **Scalability Enhancements:** Introduce pagination or clustering for charts displaying large numbers of countries.
3. **Data Annotations:** Add annotations for significant outliers or trends to provide additional context.

# Conclusion

This project provides an interactive framework to analyze global alcohol consumption patterns, health impacts, and policy effectiveness using data from OECD, WHO, and the World Bank. The visualizations—geo charts, bar charts, bubble charts, and line charts—effectively present complex data, offering valuable insights for policymakers, researchers, and the public.

The evaluation confirmed the visualizations' accuracy, flexibility, and interactivity, enabling users to explore trends across countries, demographics, and time. While minor improvements like tooltip refinements and scalability enhancements are suggested, the project successfully highlights alcohol's societal impacts and supports data-driven decision-making.

# References

* <https://ourworldindata.org/alcohol-consumption>
* <https://www.healthdata.org/research-analysis/health-risks-issues/alcohol-use>
* <https://www.who.int/data/gho/data/themes/topics/sdg-target-3_5-substance-abuse?lang=en>
* <https://www.who.int/data/gho/data/themes/resources-for-substance-use-disorders>