Aashima Parvathi Poudhar
Heriot-Watt ID: H00423612
F21DV Coursework 1
Link to video- F21_DV_CW2_Video

Course code and name:	F21DV Data Visualisation & Analytics
Type of assessment:	**Individual** *(delete as appropriate)*
Coursework Title:	F21DV Data Visualisation & Analytics Coursework 2
Student Name:	Aashima Parvathi Poudhar
Student ID Number:	H00423612

Declaration of authorship. By signing this form:

- **I declare** that the work I have submitted for individual assessment OR the work I have contributed to a group assessment, is entirely my own. I have NOT taken the ideas, writings or inventions of another person and used these as if they were my own. My submission or my contribution to a group submission is expressed in my own words. Any uses made within this work of the ideas, writings or inventions of others, or of any existing sources of information (books, journals, websites, etc.) are properly acknowledged and listed in the references and/or acknowledgements section.

- I confirm that I have read, understood and followed the University's Regulations on plagiarism as published on the University's website, and that I am aware of the penalties that I will face should I not adhere to the University Regulations.

- I confirm that I have read, understood and avoided the different types of plagiarism explained in the University guidance on Academic Integrity and Plagiarism

Student Signature *(type your name):* *Aashima Parvathi Poudhar*

Date: *12/04/2023*

Table of Contents

F21DV - Data Visualisation & Analytics: Coursework 2

1 Overview

Coffee is one of the top 5 most popular beverages in the world and it is the second or the most popular hot beverage according to various statistical publications. While this is a generally popular and accepted piece of information, this coursework attempts to delve deeper to understand how the popularity of coffee varies across countries with particular focus on the Nordic Countries, namely, Denmark, Finland, Iceland, Norway, Sweden, the potential reasons behind this coffee culture, and the impact of it on people and the environment.

1.1 Story

Theme: **Coffee Consumption Trends and Impact.**
The data collected for this theme was analysed during the exploratory phase and the key insights uncovered are exposed through **explanatory visualizations** (Knaflic, 2014).

The story first unveils the countries with an unparalleled coffee addiction and moves on to show the user where in the world the popularity is higher. i.e., whether or not the love for caffeine has a geographical association. Along with this, the visualization focuses on other key parameters which are likely to influence higher levels of coffee consumption namely, average leisure hours of people, the land temperature, and average sunshine duration for each country. Finally, the story goes on to show the positive and negative effects of coffee intake.

2 Visualizations

2.1 Section 1 – The Top Coffee Consumers

Swarm Plot

To first draw the attention of the user towards highest ranked countries based on their per capita coffee consumption, a creative variant of a swarm plot is used (Heinz, 2020) (Shander, 2022).

- The data points are centred around the amount consumed (in kg).
- Instead of using standard circles for the swarm points, **coffee bean images** are used to make the visualization more bound to the theme.
- Content is strategically greyed out to shift the focus towards Finland, Norway, and other Nordic countries which are in the top 10.
- The text annotations are animated to appear gradually giving the user sufficient time to grasp the content.
- The colours of the title and sub-title are used in the annotations for the highest and second highest coffee-consuming countries so that the user immediately associates Finland as the “Caffeine Champion” and Norway as one of the “most saturated coffee drinkers of the world.”
- The text annotations state some interesting facts surrounding the theme and the country to add more context to the story.

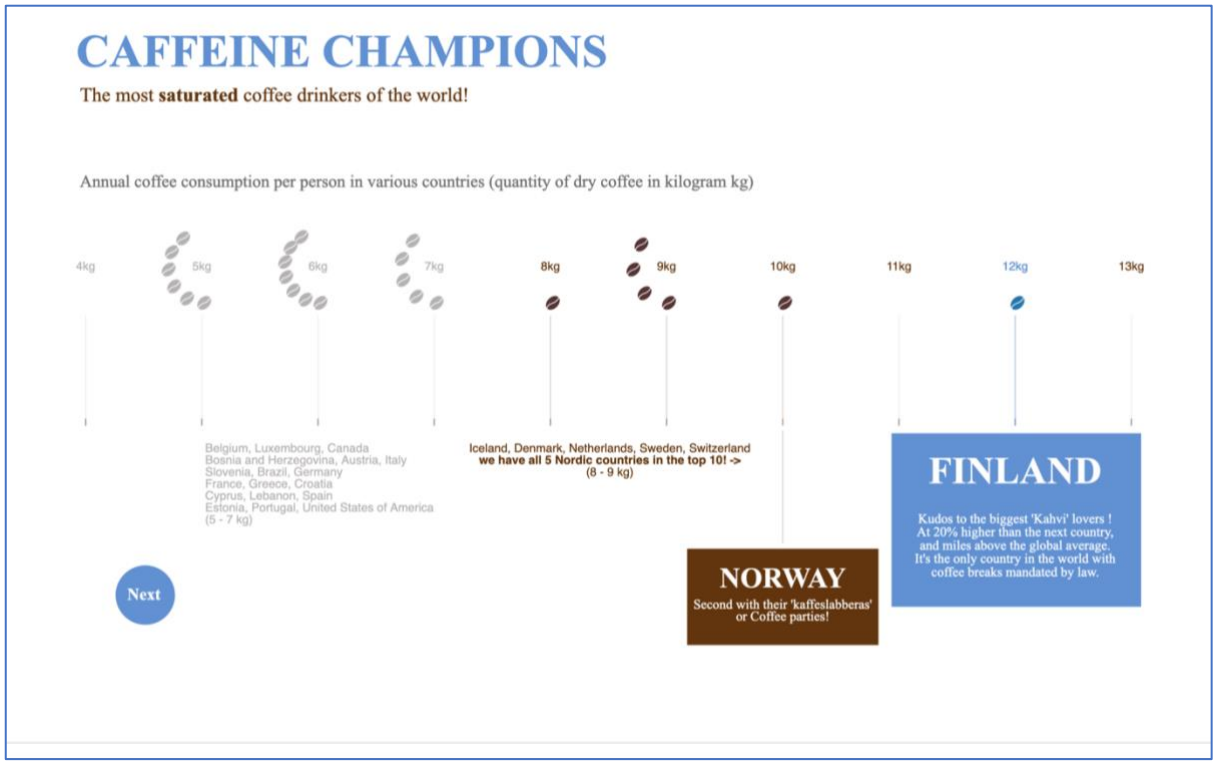


Figure 2.1: Top Per-Capita Coffee Consumers

2.2 Section 2 – Influencing Parameters

This visualization is divided into two main sections to show the trends and the reasons behind them.

Map

This section depicts how the coffee culture is spread out geographically with the help of a **map visualization**. The idea behind using a map here is to convey to the user which specific region of the world makes up a larger portion of the top consumers.

- The geoMercator projection is used for the map even though it does not provide an accurate indication of the sizes of countries, especially towards the poles because people are more familiar with this projection of the world map, and it would be more convenient for the user.
- Circles are used to represent countries where a larger radius indicates higher per capita intake of coffee.
- The colour scheme is used to indicate which continent the countries belong to.
- The animated caption over the cluster on the map calls out the continent, Europe as well as the region, Northwest which includes the Nordic countries.
- A **zoom feature** is built into the map that helps the user see better which countries are part of the severely overlapping European cluster. The **labels for countries** become visible when the user zooms in.
- This visualization becomes the foundation for the parallel plot because the parameters explored there such as temperature and sunshine duration are closely related to the geographical position of a country.

Lollipop Chart

A tooltip is added over the country circles which show the total coffee consumption trend of each country over time from 2000 to 2019 with the help of a **lollipop chart** (Project).

Parallel Coordinates Chart

Now, the second half of the visualization is a parallel plot or a parallel coordinates chart (Holtz, 2018b) which lets the user explore and understand some of the reasons behind the coffee trend.

- In keeping with the theme of the visualizations, the parallel plot lines for the Nordic countries are highlighted and they convey to the user that generally, these top coffee-consuming countries have:
    - higher leisure hours
    - lower average temperatures, and
    - lower annual sunshine hours
- The first parallel coordinate axis shows per capita consumption, and it is placed there purposefully even though it is a repetition so that the user can easily correlate it with the other parameters.
- Interactivity is added between the map and the parallel coordinates chart. When a user hovers over a circle on the map, if there's associated data available for that country, only that is retained on the parallel plot so that the user can explore and understand that influencing parameters individually for each country.

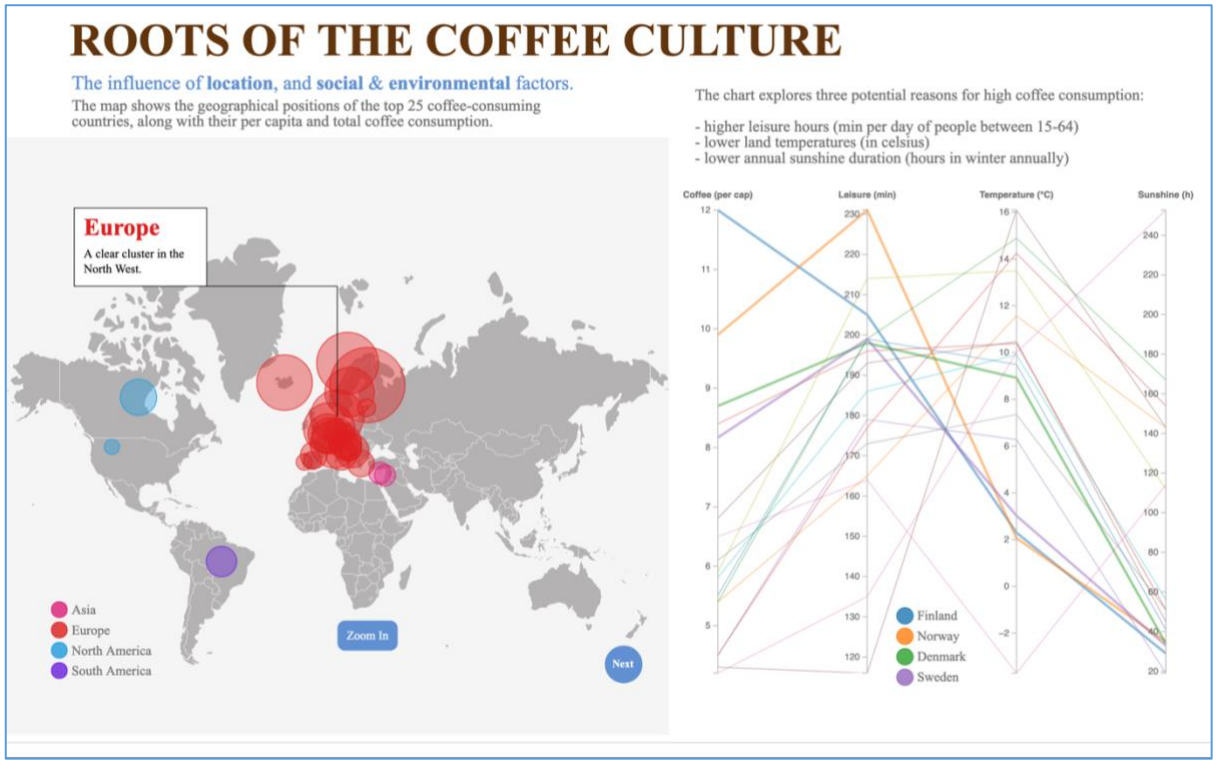


Figure 2.2: Influencing Parameters

2.3 Section 3 - The Positive and Negative Aspects of Coffee Consumption

This visualization is also divided into two main sections to show positive and negative aspects of coffee consumption.

For the positive aspect, a social perspective is chosen and for the negative, an environmental perspective.

Lollipop Chart

The first half or the left half uses a lollipop chart to show the ranking of countries based on 2 factors that are generally associated with coffee consumption:

1. Happiness / life satisfaction
2. Work productivity

- The green colour scheme is chosen to pre-attentively convey to the user that we are talking about a positive aspect.
- Strategic opacity levels directly show the user what the caption of the visualization effectively conveys – i.e., the top coffee consuming countries rank higher in terms of overall life satisfaction and productivity of its people. This again includes the Nordic countries.
- There is bidirectional interactivity between the 2 lollipop charts where hovering over one circle also highlights the lollipop associated with the same country in the other chart and fades out all the others. This way, the user can see how the same country ranks in terms of both the positive parameters.

Stacked Column Chart

The second half uses a stacked column chart (*Visual Vocabulary*) (Holtz, 2018a) and brings the user's attention to the negative environmental aspects of coffee intake and how increased consumption can take a toll on the environment through:

1. Greenhouse gas (GHG) emissions
2. Agricultural land exploitation

- The rectangles are animated to add a little life to the visualization.
- Text annotations provide an indication of how coffee production and supply chain become the third largest contributor to GHG and sixth largest contributor to land footprint.
- The red colour palette is used to indirectly convey to the user that we are exploring a negative aspect.
- Only 6 of the top contributors are retained for the chart to not clutter the visualization with multiple categories. Moreover, the 7th category "Others" which takes into account the contribution of all remaining 26 food products is greyed out so that the user's focus goes automatically to the relevant section.

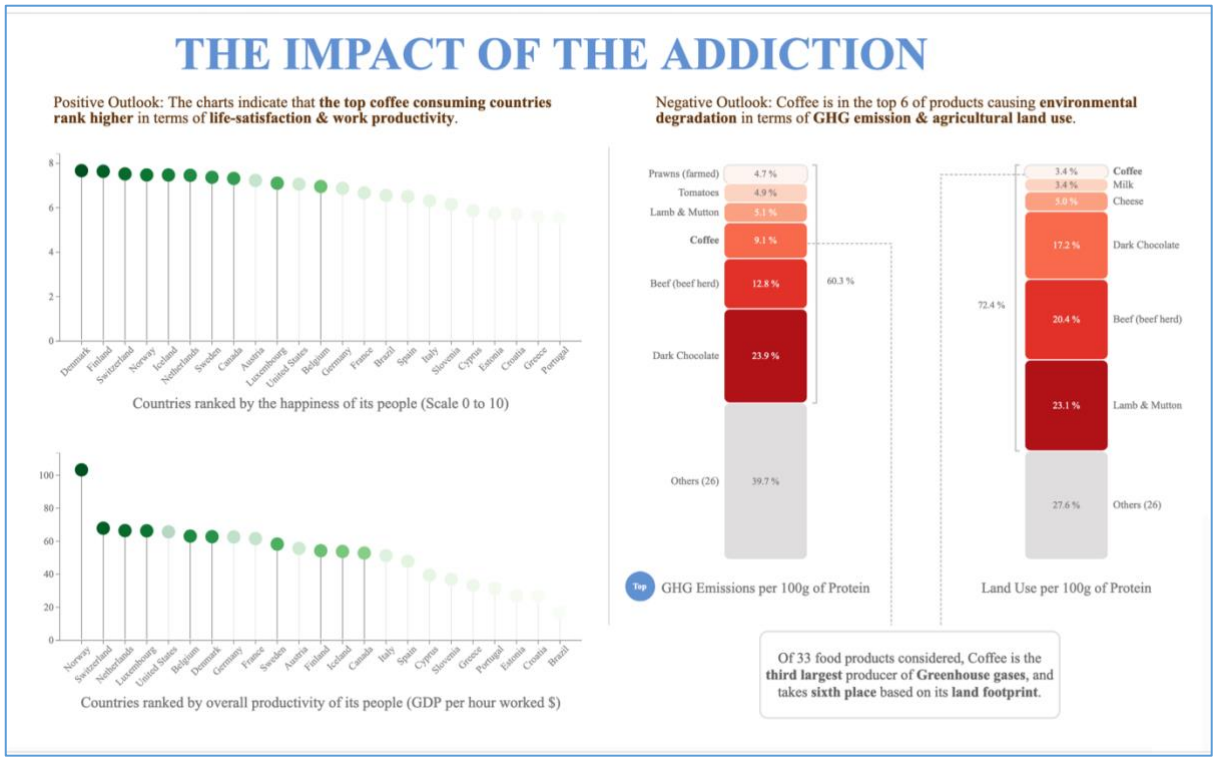


Figure 2.3: Positive and Negative Aspects

3 Application Requirements and Core Requirements

3.1 Application Requirements

A1. Clear theme

The visualizations are centred around **coffee consumption, its patterns and potential causes, and some positive and negative outcomes of large-scale coffee intake**. Care has been taken to ensure that the colour theme also remains uniform. The central colour in all the visualizations is – **this**. In addition, a second colour that goes well with the coffee theme, i.e., the colour of **coffee beans,** also has a prominent presence in all the visualizations. The colour theme is introduced and emphasized in the very first visualization to set the tone for all the remaining. Figure 2.1: Top Per-Capita Coffee Consumers clearly illustrates this application requirement.

Also, as a creative addition to the visualization, **image of a coffee bean** is used in place of the usual circles that are used to represent points on a regular swarm plot.

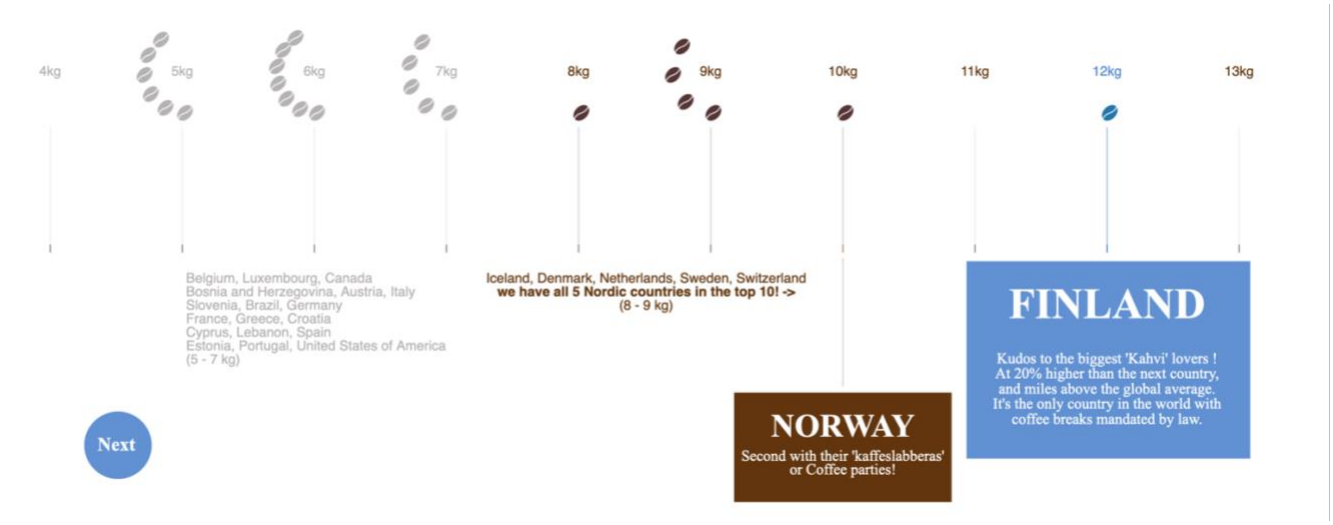


Figure 3.1 A1: Theme of the visualizations

Only visualizations, animations, and strategically positioned text are used to support and convey the theme and story.

A2. Single HTML page

Only one HTML page is used for the coursework, **index.html**. But multiple CSS and JS files are used to organize the code into logical sections.

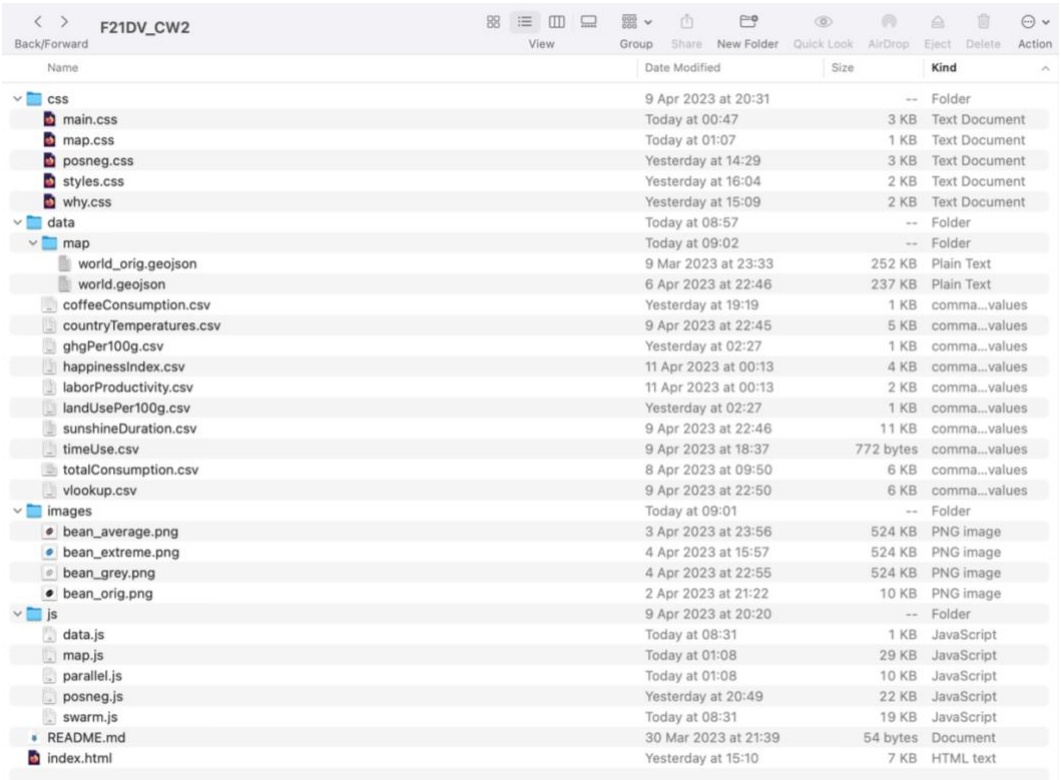


Figure 3.1 A2: Project code structure

A3. All visualizations on single HTML page

All visualizations are loaded on a single HTML page, index.html which is placed at the root of the project folder as evident from Figure 3.1 A2: Project code structure.

A4. Visualization types and Design Considerations

At least 3 visualization types are used for this coursework:

1. **Swarm Plot** – Used to show **distribution** of per capita coffee consumption across countries. Used this chart because it was important to show the value of each data point (Kim, 2022). The visualization is not a standard swarm plot but a creative visualization that is closely related to a swarm.
2. **Map** – To highlight the geographical placement of the top coffee consumers (**spatial visualization**), a map visualization is used. The idea was to show the user that there is a higher concentration in the Northern hemisphere in Northwest Europe. This sets the context for the parallel plot visualization since the parameters – temperature and sunshine duration are tightly bound to the geographical position of a country.
3. **Parallel Coordinates Chart** – Since multiple parameters – leisure hours, land temperature, and annual sunshine hours – are being explored simultaneously to convey to the user how they might influence the coffee culture of country, a parallel coordinate chart was selected. The idea is to convey both **ranking and magnitude along multiple variables**. The first line which once again shows per capita consumption is placed there for ease of **correlation**. For example, it allows the user to follow the lines for Finland and Norway, which are emphasized and deduct that they have higher leisure hours, very low temperatures and very low annual sunshine which may be why they top the list of coffee consumers. The same insight is conveyed to the user with appropriately colour-coded text.
4. **Lollipop Chart** – In the positive facet section, the lollipop charts are used to show **ranking**. The visualization type was chosen because it was important to showcase which countries are the happiest and most productive and if they're the same countries who top the list of caffeine addicts.
5. **Stacked column chart** – Rather than showing the actual values of Greenhouse gas (GHG) emissions and agricultural land use of each food product whose units are not relatable to a user, this visualization shows **composition** or **part-to-whole relationship**. With this, the unit becomes percentage which is a relatable and common unit from which the user can deduct easily that coffee is a large contributor to both GHG emission and land footprint.

A5. Positive and negative facet

A subset of the datasets is taken to create 2 different visualization layouts – a lollipop chart and a stacked column chart to show 2 sides of the subset:

1. The first visualization, i.e., the lollipop chart presents the data in a positive light by showing how coffee might be great for overall life satisfaction and improved productivity.
2. The second, i.e., the stacked column chart presents the data in a negative shade by shedding light on the adverse impact coffee consumption has on the environment.

A6. Data analytics techniques

Analytics Techniques in Code

Most significant data processing by applying appropriate analytics techniques are defined within the **fixData() function** which is included in nearly all the JS files. Most of the

techniques listed below are used in multiple files for different datasets but only including an example each.

1. Parse data / convert value to a number

```
  /* swarm.js */
coffeepercap.forEach(function (d) {
  d.percapitaconsumption = +d.percapitaconsumption;
});
```

2. Sort data

```
/* swarm.js: Sort in descending order of per capita consumption */
coffeepercap.sort(function (a, b) {
  return b.percapitaconsumption - a.percapitaconsumption;
});
```

3. Group data by specific parameters

```
  /* map.js group data by country; group data by isocode */
countries_group = d3.group(coffeepercap1, (d) => d.country);
isocode_group = d3.group(coffeetotal, (d) => d.isocode);
```

4. Mapping data between any 2 parameters

```
  /* map.js map isocode and continent; isocode and country; add unique continents to set */
const continentByCode = new Map();
const perCapByCode = new Map();
const continents = new Set();
  coffeepercap1.map(function (d) {
    continentByCode.set(d.isocode, d.continent);
    countryByCode.set(d.isocode, d.country);
    continents.add(d.continent);
  });
```

5. Filter

```
/* map.js: Filter the world.features array to include only the features whose id matches any of the isocodes
present in the per capita coffee consumption data */
  filteredFeatures = world.features.filter(function (d) {
    return isocodes.includes(d.id);
  });
```

6. Merge data

```
/* parallel.js: Merging to have all parameters in a single dataset*/
 mergedData = d3.merge([
  filteredCoffee.map((d) => ({
    isocode: d.isocode,
    twodecimalplaces: d.value,
  })),
  filteredLeisure.map((d) => ({
    isocode: d.isocode,
    leisuretime: d.value,
  })),
  filteredTemperature.map((d) => ({
    isocode: d.isocode,
    temperature: d.value,
  })),
  filteredSunshine.map((d) => ({
    isocode: d.isocode,
    sunshine: d.value,
  })),
 ]);
```

7. Set

```
 /* posneg.js: Combine the isocodes from filtered arrays, remove duplicates and store in a new set */
 const isocodes = [
   ...new Set([
    ...filteredCoffee.map((d) => d.isocode),
    ...filteredHappiness.map((d) => d.isocode),
    ...filteredProductivity.map((d) => d.isocode),
   ]),
 ];
```

8. Other – d3.max, d3.extent, d3.sum, array.slice

Exploratory Phase – Analytics and Clean-up of Data

1. The visualizations and interactivity between them rely on ISO alpha-3 code as a common element between them. This was chosen as the right property since the geojson file for the world map also contains ISO code. Therefore, the vlookup.csv file which contains countries, continents and their ISO alpha-3 code was used to perform VLOOKUP in Excel for all data that contains information about countries.

2. The leisure hours data set (timeUse.csv) originally contained multiple categories of how people between the age 15-64 spend their time. Since the coursework focuses on leisure hours, the data was filtered for the categories – 'Sports', 'Seeing friends' and 'Other leisure' and the subtotal function was applied to calculate the sum of all 3 categories by grouping the values based on country.
3. The happiness / life satisfaction dataset (happinessIndex.csv) initially contained the happiness cantril ladder score for every year from 2003 to 2021 for all countries. The subtotal function was used to aggregate and calculate the average score for each country. The same operation was applied to the labour productivity dataset (laborProductivity.csv).
4. Percentage contribution was calculated for the GHG Emission and Land Use datasets by applying appropriate formulae to add a new column that shows the percentage share of each food product towards GHG emission and land use.

A7. Complexity, interactivity, and animation

- The swarm plot is animated to gradually show captions that share the insights with the user.
- The map has a zoom feature which allows the user to zoom in on the circles depicting per capita consumption of coffee.
- The labels on the map become visible when the user is sufficiently zoomed in.
- Annotations are used in all visualizations to call out significant insights.
- Tooltips have been added wherever appropriate. The tooltip over the country-circles on the map show the variation in total coffee consumption over a period.
- There is one-sided interactivity between the map and the parallel coordinates chart where hovering over a country retains the line corresponding to that country on the chart and all others are hidden.

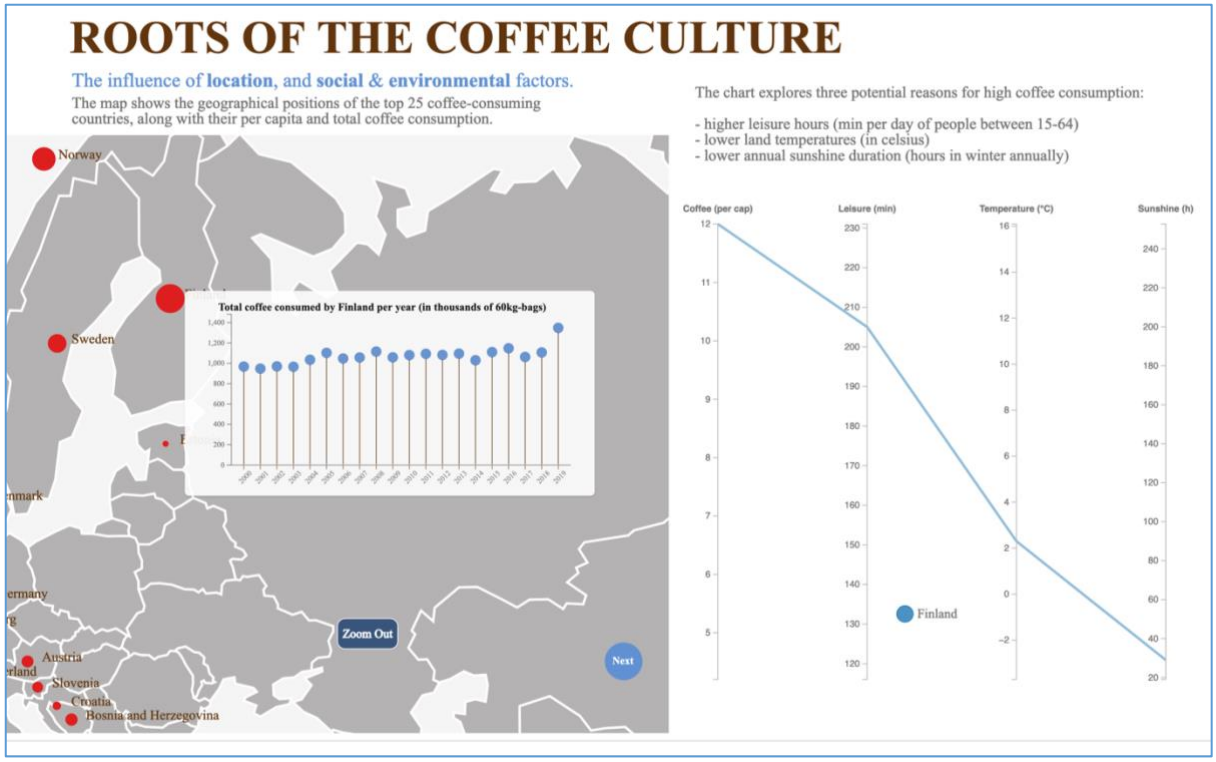


Figure 3.1 A7.1: Zoom, tooltip, interaction between charts

- There is bidirectional interaction between the 2 lollipop charts depicting the ranking of countries based on happiness index and labour productivity where hovering over a lollipop in one chart highlights the lollipop corresponding to that country in both charts and lowers the opacity of all other countries. A guiding tooltip also shows the user the ranks of the country in terms of happiness and productivity.

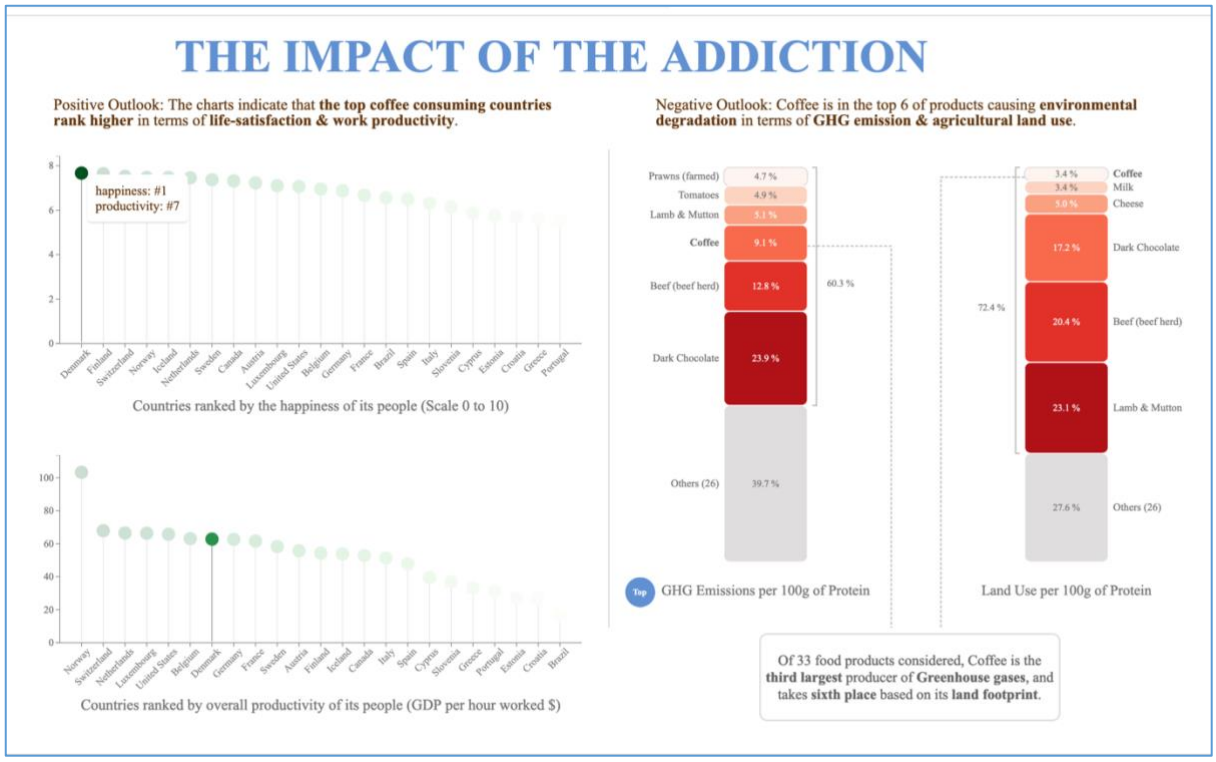


Figure 3.1 A7.2: Bidirectional interaction between charts and tooltip guide

- Bouncing buttons are added with the help of CSS keyframes which help the user navigate through the visualizations without having to scroll.
- CSS media is used to adjust the sizes of various elements based on window width.

3.2 Core Requirements

C1. D3 version 7

Only d3 version 7 is used for all visualizations.

```
    <!-- Scripts -->
    <script src="https://d3js.org/d3.v7.min.js"></script>
    <script src="http://d3js.org/topojson.v3.min.js"></script>
    <script type="module" src="./js/data.js"></script>
    <script type="module" src="./js/swarm.js"></script>
    <script type="module" src="./js/map.js"></script>
    <script type="module" src="./js/parallel.js"></script>
    <script type="module" src="./js/posneg.js"></script>
  </body>
</html>
```

Figure 3.2 C1: D3 version

C2. Transitions and/or animations

Several transitions and/or animation are used throughout the visualization. The details are provided in section: A7. Complexity, interactivity, and animation

C3. Intuitiveness

- Strategic use of colour makes the visualization intuitive, For instance the same color is used for the title in the swarm plot that says “Caffeine Champions” and the annotation for Finland, the country which tops the list of per capita consumption. In effect, it is clear to the user without even thinking about it that Finland owns the title of “Caffeine Champion”.
- Simple mouse-over interactions have been added which give the user the ability to move around and interact with the visualizations.
- The colours are thoughtfully chosen for the visualizations that present the positive and negative aspects of coffee consumption. Green generally indicates good and red indicates bad and the same principle is applied here.
- Bouncing buttons call the user’s attention to use them to traverse through the visualizations instead of having to scroll.

C4. Accessibility

- **Colour** – The colours chosen for the visualizations included in this coursework have been verified to be visible against the background colour of each visualization which are either white or grey with the help of the Colorable tool (Colorable).

  For example, below are the scores of the main 2 colours which are part of the central theme of the visualizations:
  COLOR1 – 3.23 AA Large
  COLOR2 – 10.32 AAA

- **Font** – Font sizes have been chosen to be sufficiently large and clear. Fallback font family has been set to ensure browser compatibility. Bold font is used for emphasis.

- **Heading elements** – The visualizations are enriched with supporting titles, sub-titles and captions and heading elements are used wherever possible to ensure that screen readers on browsers can be used by those who need it to understand the content.

- **Captions** – The visualizations have been captioned to draw the user’s attention to specific insights. Supports users who use screen readers.

- **Alternative text** – supporting alt-text has been added to the coffee bean images used in the first visualization. The text is sufficiently descriptive to state what it is being used for:

  <image xlink:href="./images/bean_average.png" alt="Image of coffee bean is used to ***represent countries on the chart showing the distribution of per capita coffee consumption of various countries****" x="443.111111111111" y="282" width="20" height="20" class="bean" id="Switzerland-bean"></image>*

C5. Comprehensively documented source code

Source code has been comprehensively documented with comments wherever necessary. A description has been added to every source file that shows what exactly it does and the

functions it uses to accomplish its tasks. There are also individual descriptive comments over every function and other sections of code where a significant task is being carried out.

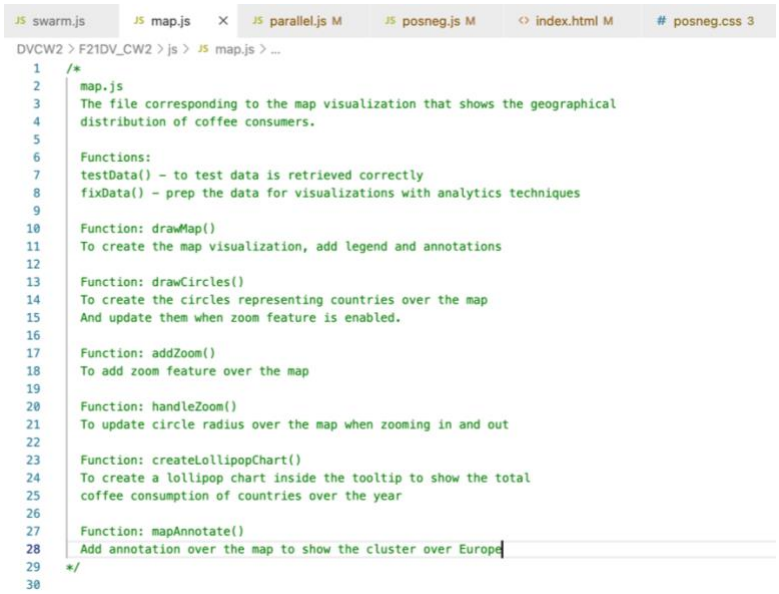


Figure: 3.2 C5: Comprehensive documentation of code

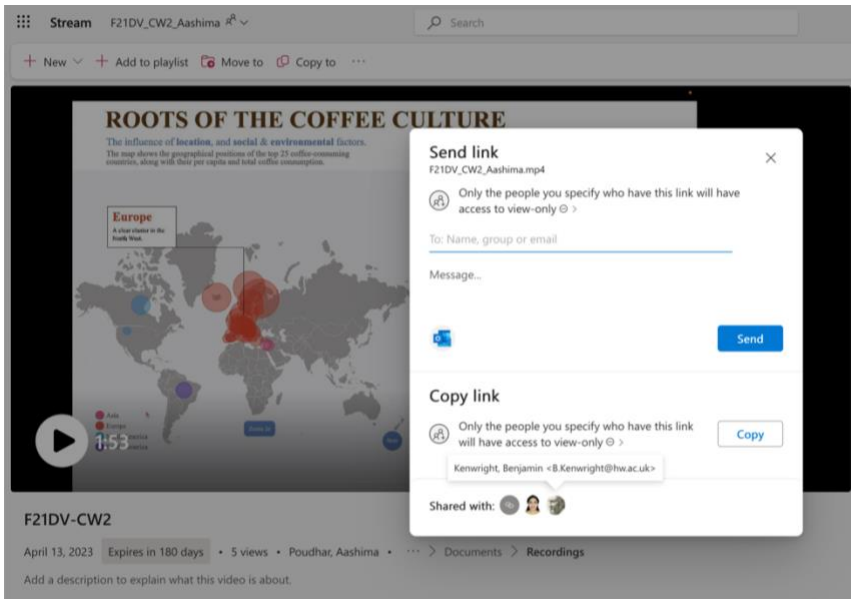
C6. Design and implementation choices

The design, reasoning behind it and the implementation choices are described in detail in section: A4. Visualization types and Design Considerations

C7. Demonstration video

The video demonstrates the application and code. The same has been shared with the course tutor, B. Kenwright.

Link - F21DV_CW2_Video



4 Data Sources

All datasets used in this coursework are listed below along with their sources, a brief description of the data and the unit of the key data in each dataset. All datasets are free to use, and the citations have been appropriately added in the report.

2.1 Per Capita Coffee Consumption by Country

Sources: https://www.kaggle.com/datasets/nurielreuven/coffee-consumption-by-country-2022
https://www.ico.org/trade_statistics.asp?section=Statistics
Zippia. "33+ Buzzing Coffee Industry Statistics [2023]: Cafes, Consumption, And Market Trends" Zippia.com. Mar. 19, 2023, https://www.zippia.com/advice/coffee-industry-statistics/

Description: The data shows the per capita consumption of coffee per year for 25 countries.
Unit: Annual per capita coffee consumption in lbs and kilogram.
File name: coffeeConsumption.csv

2.2 Total Coffee Consumption by Country

Sources: https://www.ico.org/historical/1990%20onwards/PDF/4b-disappearance.pdf
Description: Contains the total consumption of coffee by various countries for all years from 1990 to 2019.
Unit: Thousands of 60kg bags
File name: totalConsumption.csv

Map GeoJSON

Source: https://raw.githubusercontent.com/holtzy/D3-graph-gallery/master/DATA/world.geojson

2.3 Influencing parameters

Leisure Hours by Country

Sources: https://ourworldindata.org/time-use-living-conditions#note-3
https://stats.oecd.org/Index.aspx?DataSetCode=TIME_USE
Description: Average leisure time in minutes per day of people between the ages 15 and 64 for various countries.
Unit: Minutes per day
File name: timeUse.csv

Average Temperature by Country

Sources: https://tradingeconomics.com/country-list/temperature
Description: Average temperature by country as of December 2021
Unit: Celsius
File name: countryTemperatures.csv

Annual Sunshine Duration by Country

Sources: https://www.kaggle.com/datasets/thedevastator/the-sunniest-cities-in-the-world
https://data.world/makeovermonday/2019w44
https://en.m.wikipedia.org/wiki/List_of_cities_by_sunshine_duration

Description: This dataset contains the average annual sunshine hours for various cities.
Unit: Hours per year
File name: sunshineDuration.csv

2.4 Positive Impact of Coffee Consumption – Social Impact

Happiness / Life Satisfaction by Country

Sources: https://ourworldindata.org/grapher/happiness-cantril-ladder
https://worldhappiness.report/

Description: Self-reported life satisfaction of people of various countries for 2003 to 2021 on a scale of 1 to 10 where 1 indicates low life satisfaction / low happiness index and 10 indicates extremely high life satisfaction.
Unit: Cantril Ladder Score (scale of 0 to 10)
File name: happinessIndex.csv

Labour Productivity by Country

Sources: https://ourworldindata.org/grapher/labor-productivity-per-hour-PennWorldTable?tab=chart
https://www.rug.nl/ggdc/productivity/pwt/
Description: Productivity in this dataset is measured as the GDP or output per hour of work for different countries.
Unit: Dollar
File name: laborProductivity.csv

2.5 Negative Impact of Coffee Consumption – Environmental Impact

Greenhouse Gas Emissions by Food Product

Sources: https://ourworldindata.org/grapher/ghg-per-protein-poore
Description: This dataset provides the average greenhouse gas emissions per 100 grams of protein for various food products. This combines the emissions during various stages of production including crop production, land use, supply chain and so on.
Unit: This is measured in kilogram where each unit represents a kg of carbon dioxide equivalents per 100 g of protein contained in a food product.
File name: ghgPer100g.csv

Agricultural Land Exploitation by Food Product

Sources: https://ourworldindata.org/grapher/land-use-protein-poore
Description: The dataset contains data for the land required to produce 100 grams of protein for a range of food types. This is sometimes called as land footprint.

Unit: meter squared
File name: landUsePer100g.csv

5 Debugging and Testing

- Used $0__data__ and console.log messages throughout for debugging.
- Tested the visualization with different inputs and combinations.
- Tested and verified to work on Firefox and Chrome.

6 References

Colorable *Colorable*. Available at: https://colorable.jxnblk.com/ (Accessed: 13 April 2023.

Heinz, M. (2020) *Better Data Visualization Using Beeswarm Chart*. Available at: https://martinheinz.dev/blog/27 (Accessed: 13 April 2023.

Holtz, Y. (2018a) *Most basic stacked barplot in d3.js*. Available at: https://d3-graph-gallery.com/graph/barplot_stacked_basicWide.html (Accessed: 13 April 2018.

Holtz, Y. (2018b) *Parallel coordinates chart with color effect*. Available at: https://d3-graph-gallery.com/graph/parallel_custom.html (Accessed: 13 April 2023.

Kim, S.J. (2022) *10 Less-Known Python Visualization Concepts and Hacks*. Available at: https://towardsdatascience.com/10-less-known-python-visualization-concepts-and-hacks-2b5f8ff4a5b7 (Accessed: 13 April 2023.

Knaflic, C.N. (2014) *Exploratory versus explanatory analysis*. Available at: https://www.storytellingwithdata.com/blog/2014/04/exploratory-vs-explanatory-analysis (Accessed: 13 April 2023.

Project, D.V. *Lollipop Chart*. Available at: https://datavizproject.com/data-type/lollipop-chart/ (Accessed: 13 April 2023.

Shander, B. (2022) *Learning Data Visualization*. Available at: https://www.linkedin.com/learning/learning-data-visualization-15572314/creative-innovative-visualizations?autoplay=true&resume=false&u=2374954 (Accessed: 13 April 2023.

Visual Vocabulary. Available at: https://ft-interactive.github.io/visual-vocabulary/ (Accessed: 13 April 2023.