COMP4037 Research Methods Spring

Coursework 2: Data Visualization

Number of Credits: 25% of module
Recommended hours: 20-25 hours (for distinction level)
Submission Deadline: 25 April by 16:00
~~**Reassessment Deadline**: 21 Aug by 17:00~~
Late submissions will incur a penalty of 10% per day including weekends and bank holidays. Submit a **PDF** file on Moodle-a link to a submission form will be provided a few days before the deadline.

Questions: If you have any questions about this coursework, please ask them: 1) in class, 2) after class, or 3) in the Research Methods Teams chat. Individual questions received via email will be redirected towards these other options due to the class size. *See:*

Research Methods Questions Management: Asking and Answering Questions

Page Limit: Your submission is limited to 2 pages. Any graphics you use must be on page 1. The description may extend to page 2. Your supplementary video is limited to 2 minutes.

The **aim** of this coursework is to:

- Help you to build an understanding of the research methods in your area of interest
- Develop and assess your ability to present research in a concise visual form
- Give you practice in the use of data as evidence to support your work

This coursework supports the following **Learning Objectives** of the module:

- To increase understanding of the research process and the application of appropriate research methods to given problems.
- To develop the ability to critically appraise and choose research methods, and justify their application to appropriate research problems.
- To enhance skills to present results as part of implementing a project.

Task Description and Context

Professor C was approached by a team of researchers from the University of Oxford studying the impact that diet and food consumption has on the environment. They have published a fascinating research paper on their research (Scarborough et al,2023), however, there is a major limitation with their work. There is a startling absence of any graphics or visual representations of their research data and findings. They need your help with this challenge. They have been collecting a massive amount of diet and food consumption data from over 55,000 consumers linked together with food production from 38,000 farms across 119 countries..

Each row in the data set from the project contains the following information:

1. Participant number

2. Type of Diet
3. Gender
4. Age Group
5. Mean GHG emissions (GreenHouse Gas) measured in kg
6. Mean Agricultural Land Use in square meters
7. Mean Water Scarcity
8. Mean Eutrophication[1] Potential– measured in g of PO_4e, gPO_4e
9. Mean GHG from CH_4 emissions (Methane) from livestock management measured in kg
10. Mean GHG from N_2O (Nitrous Oxide) emissions associated with fertilizer use[2]
11. Mean Biodiversity Impact–species extinction per day
12. Mean Agricultural Water Usage in cubic meters (1 m^3 - 1,000 liters)
13. Mean Acidification Potential

It is a lot of data. In fact, it feels like data chaos. The researchers would like to know if visualization can be used to gain insight into their data.

Understanding the Data: As with real-world projects, acronyms and special terminology are used. Thus, you may need to conduct some research, like Googling various terms, in order to gain an understanding of the data and its associated terminology. *We recommend you start by reading the research paper about the dataset:*
Vegans, vegetarians, fish-eaters and meat-eaters in the UK show discrepant environmental impacts | Nature Food

Download the Data: The data can be downloaded from:
Dataset for 'Vegans, vegetarians, fish-eaters and meat-eaters in the UK show discrepant environmental impacts' - ORA - Oxford University Research Archive

Optional Link to Scripts that Generate the Figures
Supplemental Data for 'Estimating the environmental impacts of 57,000 food products' - ORA - Oxford University Research Archive

Better Than Professor C: Professor C would like to obtain and convey the insight contained in the data visually. The goal is to create images that maximize our understanding of the data. These are some examples of questions that can be considered in the coursework.
Some obvious factual information, such as:

- Which diet is most beneficial for the environment?
- What is the disparity between the best and worst diets?
- Which age group has the best diet?

---

[1] What is Eutrophication?
https://oceanservice.noaa.gov/facts/eutrophication.html

[2] The Green House Gas No One Is Talking About
https://civileats.com/2019/09/19/the-greenhouse-gas-no-ones-talking-about-nitrous-oxide-on-farms-explained/

- Is there a difference in diet according to gender?

Professor C would also like to find out some less obvious information, such as:

- Are there any interesting patterns or trends in the data?
- Are there any outliers?
- What insight can we gain from multivariate visualizations of the data?
- Can we compare the diets and their environmental impacts across all categories?

Professor C made some attempts to visualize the data using old-fashioned pie charts, bar charts, bubble charts, and line graphs. He was disappointed by the results, which are not very insightful, not visually uninformative, nor aesthetically impressive. *Can you do better than Professor C?*

A Real-World Challenge

Selecting a tool: You are required to select an appropriate visualization tool for depicting the dataset concerned. You may have a look at the visualization resources paper (Liu, et al 2022) for a description of available tools. See the Section called "*Free, Off-the-Shelf Software Collections for Data Visualization Practitioners*" in the survey paper by Liu et al (2022) in the References. You may also use Python, Java, R, Matlab or some other appropriate programming library.

Process the data: You are allowed and encouraged to further abstract the data, e.g., aggregate data records into groups or add new data or meta-data. In fact, you can process or supplement the data in many ways and you are expected to make changes, e.g., perhaps linking some spreadsheets together or creating new spreadsheets.

Your task is to produce an advanced visual design which can convey some meaningful and hopefully interesting insight into the data. You are required to supply a unique observation about the data. For your final image, fill out a description template as below. See Figure 1 for an example.

Visual Design Description Template:

Provide the following information for the image you create:

- **Image:** The visualization itself as an image
- **Visual Design Type:** The name/type of the visual design
- **Name of Tool:** The tool that was used to generate the image
- **Diet Groups**: The diet groups shown
- **Variables:** the other data attributes shown and *why you chose them*
- **Visual Mappings:** Each of the visual design mappings. Include the data mapping information about **color**, **position** (x,y axes), **shape**, **size**, **hierarchy**, and any other visual mappings.
- **Unique Observation:** Things we can learn from the visualization, e.g, from this image we can see this pattern… **Make sure you describe *where and how* in the imagery**

your unique observation can be seen. Is it shown in the accompanying video? If so, at what time in the video? Note that outliers are normally very interesting. They indicate that something unusual is happening.
- **Data Preparation:** Any modifications to the original data that had to be performed to generate your beautiful image.
- (**Optional**) **URL** to **screen-capture demo** showing any animation or user-interaction (any form of user input like mouse interaction, user-options, keyboard options, filtering, selection, brushing). You can also use the video to emphasize your unique observation and highlight where in the image this can be seen. You can include a link to a YouTube or Vimeo video to demonstrate any interaction or animation (similar to CW1). I recommend **two minutes maximum time**. Due to the large volume of submissions, we will not have time to look at and interact with Tableau public software web pages. Therefore, we require a video.
- **(Optional) URL to source code:** if you use a programming language like R or Python you can provide a link to GItLab or GitHub where your code is stored.

A good observation requires interpretation of the resulting image that you generate. Also, if your visual design features interaction or animation, you may upload a screen-capture demonstration to YouTube and provide the link in the PDF you submit. Note that all user-interface components should be shown in English.

Advanced Visual Designs: You are to do better than Professor C! That means you are required to use more advanced visual layouts than typical line graphs, bar charts, pie charts, and bubble charts. That means the following are not allowed: bar charts, pie charts, line charts, and bubble charts, however, you can use these simple charts to support more advanced visual designs, in other words, as complementary charts to support an interesting observation.

What is Better? Create a visualization that:
- conveys information and knowledge
- enables discovery of: patterns, deviation, hierarchy, relationships and association
- identifies relationships between data attributes
- depicts data at different scales
- separates noise from the signal
- is intuitive to laypersons and easy to learn
- is aesthetically pleasing

Describe the insight that your visualizations provide. What can we learn from your visualization? How is it better than a standard line, pie, or bar chart? You are not required to answer all of the above questions. They merely serve as examples of what you could discuss when providing a unique observation about the data through the lens of your imagery.

Examples: Some examples of more advanced visual designs are treemaps, parallel coordinates, matrix charts, radar charts, scatterplot matrices, and digital maps.

Very Common Mistake with Treemaps: A common mistake that students make is using a treemap to show nonhierarchical data. Treemaps are a hierarchical data visualization technique. Without a hierarchy, they may be even less effective than a bar chart. However, you can create a hierarchy from your data to visualize. You may look at the following video tutorial(s) for extra help on this topic:
A Short Introduction to Treemaps
Treemaps for Complete Beginners
Visualization Lecture 5 1. Treemap Construction
You will be penalized by -20 points for this error.

Color: For a color map, you are encouraged to use this one: **https://paletton.com/**

Digital Maps: You are encouraged to explore the use of digital maps in your visualizations. Using digital maps will yield bonus points. For a good resource on finding digital maps, please refer to the survey paper on visualization resources by Liu et al 2022 in the References. You may certainly use the digital, geo-spatial maps from the Tableau tutorials in your submission. However, you will not receive credit for the bar charts and line charts demonstrated in the tutorial. Trying to introduce another dataset that links together diet, ingredients, and location is very difficult. It would require reading the research paper for CW 2 and the research paper from Workshop 5 on data visualization.

Lots of Extra Help

- **Supplementary Questions and Answers:**
  Research Methods: Questions and Answers on Courseworks 2 & 3
- **More Questions and Answers on Courseworks 2 & 3**
  https://youtu.be/6DJYgh8B1uE
- **Research Methods Lecture on Visualization**
- **Supplementary Visualization Lectures:** You will see many examples in the supplementary data visualization lectures. If you'd like to get extra help, you may watch the course lectures 1-6 on YouTube: Data Visualization Classes 2019 - YouTube
- **Hands-On Tutorials:** For extra help, you may also watch either one of the following live, hands-on, video tutorials of how to create a visual design including a geo-spatial map:
- Visualization in Practice: A Live Introduction and Hands-on Demonstration of Tableau
  Tableau for Absolute Beginners: A Live Introduction and Hands-on Demonstration
- Applied Visualization using Tableau: A Hands-On Tutorial and Demonstration
  Applied Visualization using Tableau: A Hands-On Tutorial and Demonstration
- **Help from Visualization Experts:** You may also ask any visualization questions on the Vis Guides Website: http://VisGuides.org

- **Help from Diet Data Experts:** Or you may contact any member of the Oxford University's Research Team: https://www.nature.com/articles/s43016-023-00795-w#citeas

Both carefully prepared tutorials start from the very beginning with a description of a sample data set and proceed step-by-step creating a selection of visual designs.

Assessment Criteria

The following will be examined to assess the coursework

The **quality** of the image supplied
The appropriateness of the **visual design**
The **visual design** type
The appropriateness of the **tool** used
The completeness of the description of the **attributes/columns**
The description of the **attributes/columns** conveyed in the image
The completeness and clarity of the **Visual Mappings**
The quality of the **Unique Observation**.
The quality of the **Data Preparation** description.
Adherence to the report **template** provided.
The quality of the supplementary **video**.

See the detailed assessment criteria in the appendix after the references.

Academic Integrity: This is an individual assessment that should consist of your own unaided work. You are permitted to use any material (e.g., diagrams and quotations) from the research paper that your video is based on, but you must make it clear when you are quoting from the paper. The University has detailed advice about academic integrity and submissions that demonstrate a lack of that integrity will be treated under appropriate disciplinary procedures.

References

(Liu et al, 2022) Xiaoxiao Liu, Mohammed Alharbi, Jian Chen, Alexandra Diehl, Elif E Firat, Dylan Rees, Qiru Wang, and Robert S Laramee, **Visualization Resources: A Survey**, *Information Visualization*, Volume 22, Issue 1, pages 3-30 (PDF file, web page, https://doi.org/10.1177/14738716221126992)

(Scarborough et al 2023) Scarborough, P., Clark, M., Cobiac, L. *et al.* **Vegans, Vegetarians, Fish-eaters and Meat-eaters in the UK Show Discrepant Environmental Impacts**. *Nature Food* 4, 565–574 (2023). https://doi.org/10.1038/s43016-023-00795-w

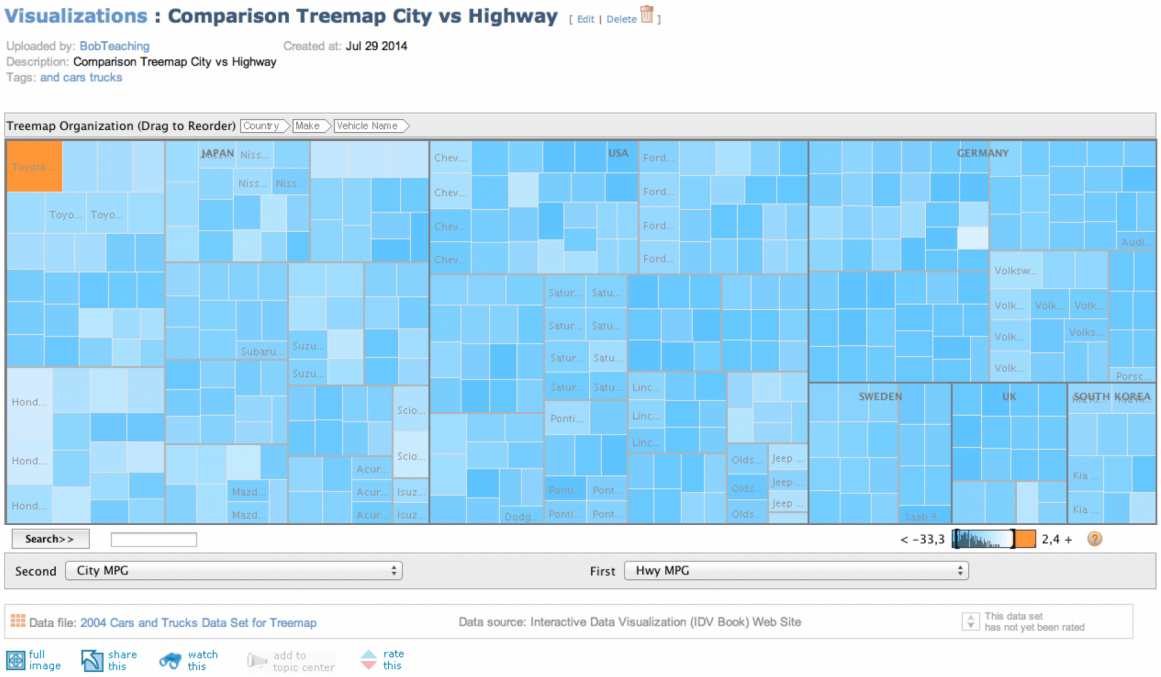
Carbon Independent (n.d.). **Emissions from food.** [online] www.carbonindependent.org. Available at: Emissions from food.

Clark, D. (2021). *UK: population by country*. [online] Statista. Available at: https://www.statista.com/statistics/294729/uk-population-by-region/.

YouGov (2024). **Dietary choices of Brits (e.g. vegeterian, flexitarian, meat-eater etc)?** [online] yougov.co.uk. Available at: https://yougov.co.uk/topics/society/trackers/dietery-choices-of-brits-eg-vegeterian-flexitarian-meat-eater-etc?crossBreak=london.

Allès, B., Baudry, J., Méjean, C., Touvier, M., Péneau, S., Hercberg, S. and Kesse-Guyot, E. (2017). **Comparison of Sociodemographic and Nutritional Characteristics between Self-Reported Vegetarians, Vegans, and Meat-Eaters**, NutriNet-Santé Study. *Nutrients*, 9(9), p.1023. DOI:https://doi.org/10.3390/nu9091023

Example of What We Are Looking For



Example Template:

- **Visual Design Type**: Treemap.
- **Name of Tool**: IBM ManyEyes
- **Country**: Germany, Japan, South Korea, Sweden, UK, U.S.
- **Disease**: Car efficiency
- **Year**: 2014
- **Visual Mappings**:
  - **color**: color is mapped to the difference between city and highway MPG for each car
  - **shape:** each bottom level rectangle represents a car type
  - **size**: the of each bottom-level rectangle is mapped to the difference in mpg between the city and highway MPG of each car type.
  - **position**: the position of each rectangle / car is determined by decreasing frequency, the more cars manufactured, the further to the left.
  - **hierarchy**: the cars are grouped by country of manufacturer → make → vehicle name
- **Unique Observation**: We can clearly see an outlier, i.e., that the Toyota is the only car with greater city MPG than highway MPG. We can also see that South Korea produces cars. I did not know this previously.
- **Data Preparation**: In order to create the hierarchy, the country of origin for each car has to be included in the spreadsheet as an additional column.

Figure 1: Your task is to produce something like this. ***The dataset used for this example is different from the one used for this assignment.*** Therefore, your report will be different. Also note that this example does not include a link to a supplementary video. A nice supplementary video will help in receiving more points.

Assessment

Grade 80-100. Exceptional insight into the data is provided, similar to that of a good PhD student. The quality of the image and the description provided are at a publishable level. A special effort is invested into aesthetics (nice appearance) and a very clear set of labels, e.g., a labeled hierarchy. This also includes the video. The visual design is advanced. Every aspect of the template is identified and completed at a publishable level.

- The **visual design** is high quality and at a publishable level.
- The name of the **tool** and **data attributes** are provided
- The **attributes/columns** conveyed in the image are provided
- The **Visual Mappings** are completely described where applicable: color legend with labels, axis labels, position, size, shape, hierarchy
- A **Unique Observation** is provided that is unique, non-trivial, and could not be simply guessed using prior knowledge or by looking at the raw data.
- The **Data Preparation** is described at a publishable level
- The accompanying **video** is created at a publishable level
- The URL of any source code is provided, the code follows coding conventions and is commented.

Grade 70-79. Demonstrates a thorough understanding of the data and the image, explanations are very clear for all components in the image template. The visual design is advanced. Very good insight is provided. The writing has a very good structure and the work has been carried out using a professional standard.

- The **visual design** is high quality.
- The name of the **tool** and **data attributes** are provided
- The **attributes/columns** used in the image are provided
- The **Visual Mappings** are completely described where applicable: color legend with labels, axis labels, position, size, shape, hierarchy
- A Unique **Observation** is provided that is non-trivial, and could not be simply guessed using prior knowledge or by looking at the raw data.
- The **Data Preparation** is described at a very good level
- An accompanying **video** is created at a very good level
- The URL of any source code is provided

Grade 60-69. Demonstrates a decent understanding of most of the aspects of the data and the visual design, explanations are largely clear for most components of the template, the submission has a decent structure, and each part is completed, though perhaps falling short of a thorough, professional writing and image quality. The chosen visual design is advanced.

- The **visual design** is high quality.
- The name of the **tool** and **data attributes** are provided
- The **attributes/columns** used in the image are provided
- The **Visual Mappings** are described where applicable: color legend with labels, axis labels, position, size, shape, hierarchy, although falling short of a thorough and professional level.
- An **Observation** is provided that is unique, non-trivial, and could not be simply guessed using prior knowledge or by looking at the raw data.
- The **Data Preparation** is described at a good level.
- An accompanying **video** is created at a good level
- The URL of any source code is provided

Grade 50-59. Demonstrates some understanding of the data and the visualization, the descriptions and explanations of each component of the template are generally okay but sometimes lack detail, are ambiguous, or contain errors. There is some attempt to structure the submission, and a decent (if perhaps) informal attempt has been made at the writing, image, and the presentation. The insight provided is rather obvious or trivial. The visual design could be improved.

- The **visual design** is of reasonable quality.
- The name of the **tool** and **data attributes** are provided
- The **attributes/columns** used in the image are provided
- The **Visual Mappings** are described where applicable: color legend with labels, axis labels, position, size, shape, hierarchy, at an adequate level.
- A **Unique Observation** is provided but is not unique, trivial, and could be simply guessed using prior knowledge or by looking at the raw data.
- The **Data Preparation** is described.
- An accompanying video is absent.
- The URL of any source code is absent.

Grade 40-49. There is some understanding of the data and visualization, but mostly the descriptions are vague, lack a lot of detail, or have errors. The supplied image is very basic with no special insight given. There is a vague structure to the submission but it isn't easy to follow throughout, some attempt at writing.

- The **visual design** is low quality.
- The name of the **tool** and **data attributes** are not provided.
- The **attributes/columns** used in the image are not provided
- The **Visual Mappings** are described poorly.
- A **Unique Observation** is provided but is not unique, trivial, and could be simply guessed using prior knowledge or by looking at the raw data.
- The **Data Preparation** is poorly described.
- An accompanying video is provided but needs improvement

Grade 30-39. A little bit of understanding and explanation of the data and visual design, but most of the components in the template description are not explained in much detail or contain significant errors. The submission doesn't have much of a structure, there is some attempt at writing. No significant insight is conveyed by the image.

- The **visual design** is low quality.
- The name of the **tool** and **data attributes** are not provided.
- The **attributes/columns** used in the image are not provided
- The **Visual Mappings** are described poorly or erroneously.
- A **Unique Observation** is provided but is not unique, trivial, and could be simply guessed using prior knowledge or by looking at the raw data.
- The **Data Preparation** is poorly or erroneously described.
- An accompanying video is provided but needs improvement

Grade 20-29. A few aspects of the data and visualization have been understood, but overall the submission doesn't show much understanding of the visual design nor the explanation of different components of the template description. The submission is poorly structured and contains little coherent attempt at writing. No effort is made to convey insight into the data.
The visual design is too basic.

- The **visual design** is low quality and too basic.
- The name of the **tool** and **data attributes** are not provided.

- The**attributes/columns** used in the image are not provided
- The **Visual Mappings** are described poorly or erroneously or are not provided.
- No serious attempt at a **Unique Observation**
- No serious attempt describing the **Data Preparation.**
- No attempt at a quality video

Grade 10-19. Shows some awareness of the data and the associated visualization, but doesn't go much beyond mentioning a few keywords and doesn't have any meaningful structure or attempt at writing. The supplied image is low quality or trivial.

- The **visual design** is low quality or contains errors.
- The name of the **tool** and **data attributes** are not provided or contain errors.
- The **attributes/columns** used in the image are not provided or contain errors.
- The **Visual Mappings** are described poorly or erroneously or are not provided.
- No serious attempt at a **Unique Observation**
- No serious attempt describing the **Data Preparation.**

Grade 0-9. No or minimal attempt