**Colour Guide:**

**Red : Important information**

**Grey: slightly important / Do not forget**

**Bold: Important key words**

**Box: Notes, theory or tips.**

**Keats information:**

**7CCSMSDV - Simulation and Data Visualization Assignment Description**

The module assignment component is an individual project.

The overall assignment expected time is about 15 days of work.

You can start working on the assignment as soon as it becomes available and submit by the deadline.

I suggest you pace yourself by addressing the parts that you can already complete.

The assignment will comprise of 3 parts: Analytics, Design and Implementation.

You will be provided with a Dataset, research question and visualization examples.

You are allowed to access offline (e.g. books, articles, notes etc.) and online resources, including example code.

Final Submission will include:

* a 2-3 pages report (either a 1 page A0 poster or 2-3 pages short paper) in .pdf format. The page report shall follow the VGTC conference two-column format in line with the [IEEE VIS Posters’ formatting guidelines](https://tc.computer.org/vgtc/publications/conference), using the poster format option. Note: You shall have maximum 3 pages of text, and as many pages as you need for references and images. If you choose to submit a 1 page poster then it will have to be: A0 size. Either submission shall be in .pdf format;
* source code of the developed visualization;
* a 2-3 min (maximum) video demo presentation of your implemented visualization in action.

Note: If the submitted report is more than 3 pages long I will stop marking at page 3.

Likewise if the submitted video is more than 3 min in length I will stop watching at minute 3.

Assignment script and submission link will appear in this section.

**Submission Link and Instructions**

Submission instructions.   
You shall submit a zipped file named "EmailID.zip", where EmailID shall be replaced by the first part of your KCL e\_mail address

Example:

if e\_mail = john.smith@kcl.ac.uk

then zipped\_file\_name = john.smith.zip

The zip folder shall contain the following:

1. The answers to Part 1 and Part 2 in a single report in .pdf format. Please label each section of the report clearly with Part 1, Part 2 headings.
2. The answer to Part 3 shall be submitted as a folder containing the code of your webapp implementation (accepted formats: a WebStorm project, or a self-contained .html page).  If you created a live version of your webapp accessible online you can also add the direct link in the report. Please mark it visible under a section or subsection titled "WebApp Live".
3. A 2-3 min (maximum) video presentation describing the design principle behind your webappp, and a demo of  how the implemented visualization support analysis of the data and answer your research question(s).

King’s College London

This paper is part of an examination of the College counting towards the award of a degree. Examinations are governed by the College Regulations under the authority of the Academic Board.

Degree Programmes MSc

Module Code 7CCSMSDV

Module Title Introduction to Data Visualization Examination Period Summer 2023 (Period 2)

Time Allowed Submission Deadline April 21, 4PM (BST)

Rubric Part 1 carries TWENTY marks. Part 2 carries TWENTY marks. Part 3 carries THIRTY marks.

Time: From Dr Rita Borgo recommendation she says it takes 15 days of work to complete the coursework. Carlota the TA took a total of 3 weeks as she was working on the coursework and studying for exams at the same time. In the end it is your time and you know how you work best, simultaneously with other work or concurrently. Just make sure you do not rush the work as creativity needs time and creativity is rewarded in this project.

Remember: The deadline is several weeks after the Big Final Year Project deadline, so you could work on the coursework once you give in that project. However, you will not get as much feedback from Tas and possibly no questions answered from Tas and the Lecturer.

Also questions will not be answered in the closer dates to the deadline. You can get feedback and questions answered during LGTs and SGTs. This coursework counts as 70% of your final grade.

Tip: Split the work in sections and give yourself personal deadlines so you stay on top and not have to cram a lot of work in the last minute.

© 2023 King’s College London

Individual Project Assignment

Introduction

This assignment is divided into **3 parts: Analytics, Design and Prototyping**, Implementation. Please **read** carefully the instructions related to each part. Data Scientists are called to put their skills to action to help mine the large and diverse amount of data available to support knowledge discovery in front of diverse .

To this aim the theme of this assignment will be focusing on analysis of three main topics: cost of living, social media, sport events.

The assignment has therefore three main themes (or tracks). You are required to **choose ONE**.

Remember: Use the links provided to brainstorm and start, they will help you with the coursework. But in the end you are allowed to use your own data and research. Just remember to reference it.

Proposed Tracks:

1. Cost of Living data analysis;
2. Social Media data analysis;
3. Sport data analysis.

Data. As a starting point we are providing you with:

* Three starting datasets reporting on: (i) Cost of living increase in London (ii) Social media influence on shopping (iii) Formula 1 World Championship from 1950 to 2023.

At least one of the three provided datasets should be used for the assignment. [Datasets available on module Keats page]

* A series of links to relevant data repositories and projects. These are given as a starting point, it is not mandatory to use them and you are invited to look for any other datasets available on the web that you may find interesting. [List provided on module Keats page]

Visualizations. We are providing you with links to existing visualizations, again these are given as a starting point and for inspiration, it is not mandatory to follow them and you are invited to look for other visualizations that are available on the web and that you may find inspirational. [List provided on module Keats page]

Research questions. You are provided with an initial research question for each track:

*Q1 - Track 1: “Analyze the development of cost-of-living over time. Are there any detectable trends?”*

*Q1 - Track 2: “Analyze the effects of social media influence on shopping. Are there any detectable trends?”*

*Q1 - Track 3: “Analyze performances of Formula 1 teams. Are there any detectable trends?”*

Tips:

When creating and analysing the research questions break them down in to keywords and analyse them in detail to be able to answer later. Look for keywords that frame the data: Time, Quantity, Quality, Location etc.

Stuck on which topic to chose?

Tips:

* Read into the datasets, which stands out the most to you?
* Look and search for visualisations in each topic, which seems more interesting?
* Research the topics to learn a little more about them, which sparks more research questions.

Remember: Do not worry about your decision as it will not affect your grade. All range of grades can be achieved in all the topics.

**Part 1. Analytics [20 marks total]**

Based on the track you chose and the respective data provided in the Data section and/or other datasets you may have encountered or may find online, you are required to:

1. Propose **two or more** exploratory research questions (non-trivial questions) beyond Q1, label them as Q2, Q3, etc. [Note: You can propose questions that might require access to datasets beyond the one provided, these can be questions complementary to Q1, investigating other aspects than the ones investigated by Q1, **or** refinements/expansions of Q1, you are free to decide.]  
   [10 Marks]

Tips: To come up with research questions first try to dissecting the topic, what is it about? Are there articles about them that could give you ideas for more questions? Why is it important? What aspects do you want to learn about?

Diagram, schematic

Description automatically generated

1. Explain **what type of data** could be used to answer Q1 and each one of the questions you proposed in a.. Assess the appropriateness of each dataset(s) (the ones we provided and other resources you may have found, please provide links to the latter), you would potentially be using to answer the questions.

[5 marks]

Theory: Types of data: Continuous, quality, quantity, discrete, ordinal, table, interval, integer etc.

What makes data appropriate?

Appropriate: For purposes of data quality, data are appropriate if they are suitable for the intended purpose of an analysis and relevant to the system or process being analysed. So how does the data do this?

c. Explain **if** and **how** datasets, you described in b., are or could be correlated.

[5 marks]

Theory:

Correlated: To have mutual relationship or connection in which one thing affects or depends on another.

Remember: Correlation does not mean causation [(fun examples)](https://www.wnycstudios.org/podcasts/otm/articles/spurious-correlations)

Also YOU MUST link and reference the resources you end up using either on the report, your code in the comments, or in the interactive website.

**Part 2. Design and Discussion [20 marks total]**

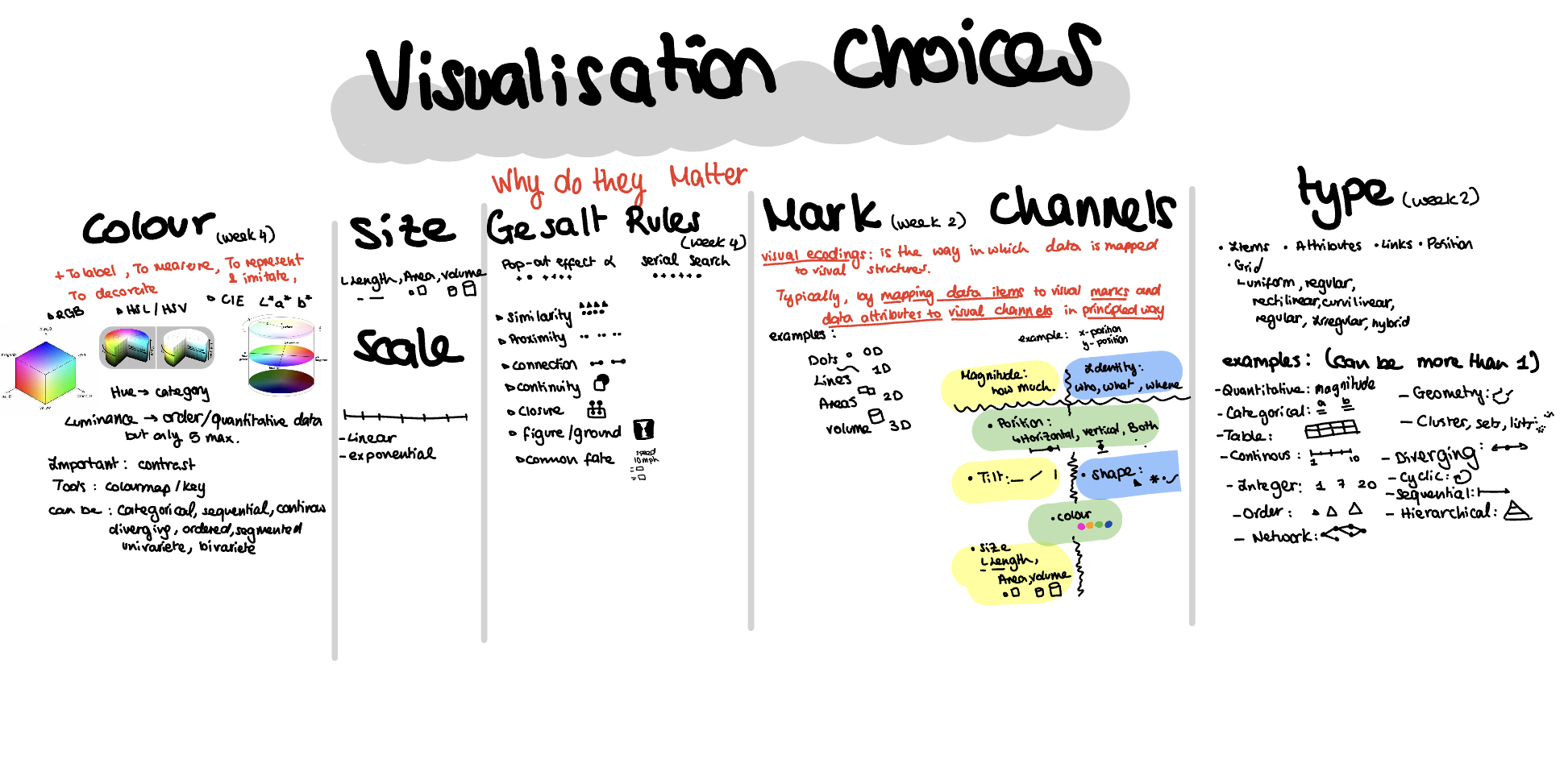
Based on visualization approaches surveyed in class and in recommended readings:

1. Propose and design a **minimum of 3** visualizations that would answer the research questions proposed in Part 1 (creativity will be rewarded). By design we mean drawing/sketching a prototype. Design can be hand drawn on paper **or** using a tool of your choice e.g. PowerPoint, Sketch, Illustrator, D3, Tableau, etc.  
   [9 marks]

Remember: If you have more than three research questions you can create more than three visualizations. Have fun and try different things. Do not limit yourself to things you think you might be able to implement be as creative as you please. However, there is no need to reinvent the wheel simple can be effective too. Just make sure to justify your decisions.

1. b. Each visualization should be accompanied by a **maximum of 300 words** describing the design rationale, which question(s) your design would help answer and if/how your design may improve upon existing examples.

Remember: If you have more than three research questions you can create more than three visualizations. Have fun and try different things. Do not limit yourself to things you think you might be able to implement be as creative as you please. However, there is no need to reinvent the wheel simple can be effective too. Just make sure to justify your decisions.



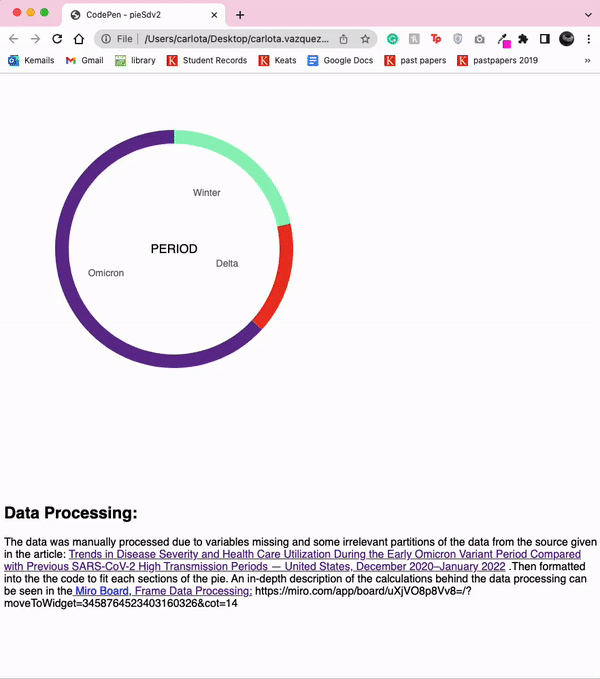
By **design rationale** we mean: the process and principles followed in choosing the specific visualization. You should provide a **rigorous rationale** for your design decisions, e.g. visual encodings used and why they are appropriate for the data. These decisions include the choice of visualization type, size, colour, scale, mark and channels and other visual elements, as well as the use of sorting or other data transformations. Consider **how** these decisions facilitate analysis and/or communication.

[11 marks]

Note: In this part we are only asking to design possible visual layouts not to implement them. If you are hand-drawing your designs take pictures and **add them to your document as figures**. If you are using a tool to develop your designs save them as images and **add them to your document.**

Remember: If you have more than three research questions you can create more than three visualizations. Have fun and try different things. Do not limit yourself to things you think you might be able to implement be as creative as you please. However, there is no need to reinvent the wheel simple can be effective too. Just make sure to justify your decisions.

Diagram

Description automatically generated 

Interaction:

Action carried by the user that changes the visualisation

Examples: Hover, click, slide, pinch, etc.

Data processed can be manual, by you or automatic by calculator online or source.

**Part 3. Implementation [30 marks]**

Out of the visualizations proposed in **Part 2** **implement one** in D3 as a webapp. You shall **use the data** (part or whole of it) provided for your track , you **can** also complement the data with other data sources you foraged yourself.

Your visualization shall support answering one of the research questions, therefore:

* it shall be accompanied by a short description of **how** data are being processed (**and acknowledgement of your data source(s)**). [10 marks]
* it can include either a composition of linked/related simple visual layouts **or** a more sophisticated single visual layout. [10 marks]
* it shall allow some level of **user interaction**. [10 marks]

Note: You are **allowed** to use D3 Example code available on the web as long as it is adapted to **your** data and you explicitly **acknowledge** the source of the original code.

The implementation is to be done using **D3**, visualizations done with other software and embedded in an html page will **not be accepted**.

Starting from scratch can be daunting so use the resources in Keats and some extra links here to help you start with an example: <https://d3-graph-gallery.com/> <https://d3js.org/>

Remember: to link the specific examples you use as inspiration or as template and change the data and certain aspects of the visualisation to make it respond to your research question.