CA682i Data Visualisation Assignment

Due: 6th December 2022 before 23:59
Submission: upload via loop, 25% of module grade

Overview: (1) create a data visualisation that may consist of up to 3 graphs, (2) write a short structured report, according to the strict template provided, describing the dataset(s) and the process you used and (3) present the visualisation in a screencast lasting no more than 5 minutes. The visualisation should **illustrate a point, answer a question or otherwise tell a story** so select your datasets accordingly.

You may use any data processing or visualisation tools that you like but you must include a specific section in your report giving reasons for your design choices (graph types, marks, colours, text, style, layout, etc.). Tools like Tableau, PowerBI etc. generally have reasonable default design options so you must explicitly describe why the design used is a good choice.

Dashboards are not appropriate.
The following datasets are **not** permitted: MovieLens (or derivatives); Chicago Crime Dataset; Road traffic or accidents in New York (or similar).

Your submission should contain the following:

(1) Short report in **PDF or DOC** format according to the template provided (end of this document)
(2) Video **file** of the screencast showing your presentation with commentary describing the process of creating and interacting with the visualisation
(3) Source code, project or configuration files (as zip archive if necessary)

Note: A simple chart on a limited amount of static data is a guaranteed fail. You should not present more than 3 charts and this is a limit not a target. A single, sufficiently sophisticated and well-done interactive graph is sufficient. Projects that are a "sampler" of lots of different charts with no connection or story will also do poorly.

What data can you use?
Your data should have some aspect of "big" data. This may mean a dataset of significant size and complexity (volume) or using two datasets from different sources and integrating them (variety) or a dynamic dataset that includes some ongoing updates (velocity). You don't need to graph every point of the dataset but can process it to select a subset to answer your question.
Possible sources: https://data.gov.ie; https://www.data.gov; https://www.gapminder.org/data/; https://github.com/awesomedata/awesome-public-datasets; https://en.wikipedia.org; https://data.europa.eu/euodp/en/home; https://registry.opendata.aws; https://www.google.com/publicdata/directory. There are many others!

Marking Criteria

(1) Dataset: "big" data; showing either data cleaning or transformation or integration.
(2) Visualisation: suitable graph choice; difficulty level; good design/style; use of interactivity or animation.
(3) Report: follows instructions and template; good abstract, critique and reflection.

Declaration on Plagiarism

This form must be filled in and completed by the student submitting an assignment

Name/s:	
Student Number/s:	
Programme:	
Module Code:	CA682i
Assignment Title:	Data Visualisation
Submission Date:	
Module Coordinator:	Dr Suzanne Little

I/We declare that this material, which I/we now submit for assessment, is entirely my own work and has not been taken from the work of others, save and to the extent that such work has been cited and acknowledged within the text of my work. I/We understand that plagiarism, collusion, and copying are grave and serious offences in the university and accept the penalties that would be imposed should I engage in plagiarism, collusion or copying. I/We have read and understood the Assignment Regulations. I/We have identified and included the source of all facts, ideas, opinions, and viewpoints of others in the assignment references. Direct quotations from books, journal articles, internet sources, module text, or any other source whatsoever are acknowledged and the sources cited are identified in the assignment references. This assignment, or any part of it, has not been previously submitted by me/us or any other person for assessment on this or any other course of study.

I/We have read and understood the referencing guidelines found at http://www.dcu.ie/info/regulations/plagiarism.shtml, https://www4.dcu.ie/students/az/plagiarism and/or recommended in the assignment guidelines

Name:_____ Date: _____

Name:_____ Date: _____

TEMPLATE: Project Title or Headline

Abstract (max 200 words)
What is the question you are answering or the story you are trying to tell?
What is the conclusion that you reached?

1. Dataset(s) [½ page]
Where/how did you retrieve your data? Provide a URL if available online.
Describe the data - size (GB or attributes), number of rows, attributes, data types present
What aspects of big data (volume, variety, velocity) are present in your dataset(s)?

2. Data Exploration, Processing, Cleaning and/or Integration [½ page]
What did you need to do to prepare the dataset(s) to create your graph/chart?
How did you choose the attributes and data subset to visualise?

3. Visualisation [1-2 pages]
Screenshot or image of visualisation.
Explain your choice of chart or graph type - what relationship or data type are you showing?
Design choices - justify your use of colour, shapes, marks, layout, structure, font, labels referring to books or articles as necessary.
Comment on any interactivity or animation and how it helps answer your question.
Give a list of tools or libraries used.

4. Conclusion [½-1 page]
Critically analyse the outcome of your visualisation with respect to your question or story.
Were there aspects that you think could be improved upon?
Were there effects or functionality that you were technically unable to achieve that would improve your visualisation?

References
Include any citation of the dataset
Include links to any tutorial or example that contributed significantly to your work
Include any books, articles or web resources supporting your design choices