DATA VISUALIZATION PROJECT

DUE, NOVEMBER 04, 2020 @ 11:59PM

There are three main things to consider when completing data visualization: (1) what question do you want to address or what story do you want to illustrate with your data visualization? (2) what is your approach to gathering, understanding, and preparing your data? and (3) how will you choose, apply, and present the visualization of your data of interest? The first two considerations are crucial and in a lot of cases are thought of in tandem. Another important factor will be the type of data you have (i.e., how variables are measured) and how this affects your choice of visualization (of course in relation to your question). Below is a link to a couple of very useful websites that provide you with a collection of charts that can be produced with R. The authors of these sites provide you with descriptions about the chart, common mistakes and the R code.

I encourage you to start with this website: https://www.data-to-viz.com/
Here is another website: https://www.r-graph-gallery.com/index.html

INSTRUCTIONS

1. Select your group members and email the instructor (herbertm@yorku.ca) and TA (madkins@yorku.ca) and include the names and student numbers of these members to us by **Wednesday, September 23, 2020 by 11:59pm**.
2. At the end of this document and attached in eClass, you will find some links to publicly available datasets you may want to consider using. It is important that you review these and select your dataset as early as possible as it can take some time to find a dataset that interests you. Feel free to also look outside the data sources provided to you to find a dataset of interest. For example, if you are working/volunteering in a research lab you may have access to data. Please ensure whatever data you do choose, you have the appropriate permission to use from the owner of the data.
3. Once you have chosen your dataset you will be asked to upload to the "Group Project and Assignment" folder on eClass, a description of your data, your question for visualization, and a description of the variables you will use (i.e., variable name, variable label, variable values or range of values, type of variable - continuous or categorical). This submission is for feedback only and should include the name and student numbers. Submit by **Wednesday, September 30, 2020 by 11:59pm**.
4. **Your final project should include the following:**
    a. A general description of the data.
    b. Your visualization question.
    c. A description of the goals/outcomes of your visualization (i.e., what information do you want to communicate with your data).
    d. Your R code with sufficient comment for anyone to replicate and fully understand your code. These comments can include information about functionality but also any technical aspects you may need to consider.
    e. Discuss any limitations of the data or the visualization you applied.
    f. Please upload to eClass on **Wednesday, November 04, 2020 by 11:59 pm**.
5. Below is a rubric to help guide you as you complete your project.

	Exceeds expectations (4 marks)	**Meets expectations (3 marks)**	**Approaches expectations (2 marks)**	**Fails to meet expectations (0 - 1 mark)**
Conceptual	The project is unique, conceptually compelling, and addresses the question in a creative way. The aesthetic, technical, and conceptual aspects are strongly related and compelling. Data is integrated proficiently and creatively.	The project is creative but could be more unique. The aesthetic, technical, and conceptual aspects are strongly related.	The project is not very creative but addresses the question adequately. The aesthetic, technical, and conceptual aspects are related.	The project is not guided by a clear question. The aesthetic, technical, and conceptual aspects have no relationship. Data is not integrated proficiently or creatively.
Technical	The project utilizes the technical concepts presented in class in a robust way, and includes more concepts not presented in class. The code is commented and is concisely and efficiently written. The project is technically robust with no bugs or glitches.	The project utilizes most of the technical concepts from the class. The code is commented, and well structured, but could be more efficient. The project is sophisticated technically with a few minor bugs or glitches.	The project only uses the most basic technical concepts presented in class. The code is commented, but could be written in a better and more efficient way. The project is technically simplistic and with significant bugs and glitches.	The project doesn't utilize the technical concepts presented in class. The code is written in a confusing or problematic way. Lacks comments. The project is extremely simple technically with serious bugs and glitches.
Design	The overall design is outstanding and professional. Connects compellingly with the technical and conceptual aspects of the project.	The overall design is cohesive and compelling, but could be executed more professionally. Connects with the conceptual aspects of the project.	The overall design reflects effort, but could be stronger and more visually compelling. Connects somewhat with the conceptual aspects of the project.	The overall design is untidy, lacks visual interest, reflects little or no effort. Does not connect with the conceptual aspects of the project.

Some possible datasets

1. https://vincentarelbundock.github.io/Rdatasets/articles/data.html
2. https://data.ontario.ca/en/
3. https://dasl.datadescription.com/datafiles/
4. https://www.causeweb.org/cause/resource-type/dataset
5. http://lib.stat.cmu.edu/datasets/
6. https://www.idvbook.com/teaching-aid/data-sets/the-breakfast-cereal-data-set/
7. https://github.com/awesomedata/awesome-public-datasets
8. http://jse.amstat.org/jse_data_archive.htm
9. https://researchguides.library.yorku.ca/data
10. https://www150.statcan.gc.ca/n1/en/type/data
11. https://www.kaggle.com/datasets
12. https://data.fivethirtyeight.com/
13. https://datasetsearch.research.google.com/
14. https://r-dir.com/reference/datasets.html