DATA 211 Winter 2023 Individual Project

It may become necessary to amend this specification at some point. If that is the case you will be notified via D2L.

Due Date: April 12, 2023 @ 23:59
Late submission: -10 marks/day (minimum unit is *day*); no submission will be accepted over two days late.

Overview

This project helps you strengthen various core programming concepts throughout the course. You need to download a dataset, write a Python 3 program to read the file, and create an infographic using the *SimpleGraphics* library (you can download the library from D2L). You need to do this project **individually** (i.e., you cannot partner with anyone).

Policy

Violation of any of the following points is considered as academic misconduct.

- You must not submit or look at solutions to the assignment problems that are not your own.
- You must not share your solution code with other students, nor ask others to share their solutions with you.
- You must indicate on your submission any assistance received.

Submission

Submit the following files to the **dropbox on D2L**:

- your Python 3 program — **.py file**, and
- your dataset(s) — plain text file(s).
    - the original dataset that you downloaded.
    - if you perform data cleaning on your dataset or modified the dataset in any way, you must submit the cleaned dataset as well, so that we can use it to run your program. If you use the downloaded dataset as is, you only need to submit the one dataset.
- any other files (for example, images) that are required to run your program.

If you have any questions regarding the specifications for this project or how to submit, please ask your TA because **it is your TA who will mark your project**. The TA who will mark your project is **Mahnaz Booshehrian (mahnaz.booshehrian@ucalgary.ca)**

Deliverables [Total 150 marks]

A. Dataset [20 marks]

Decide on a topic that interests you. **Download** a dataset related to the topic and use it for this project. Your dataset must:

- be a plain text file
- have at least 10 rows and 5 columns
- be downloaded from the Internet (i.e., not obtained by conducting a survey)
- be different from the dataset you used in your DATA 201 assignments or projects.

You must:

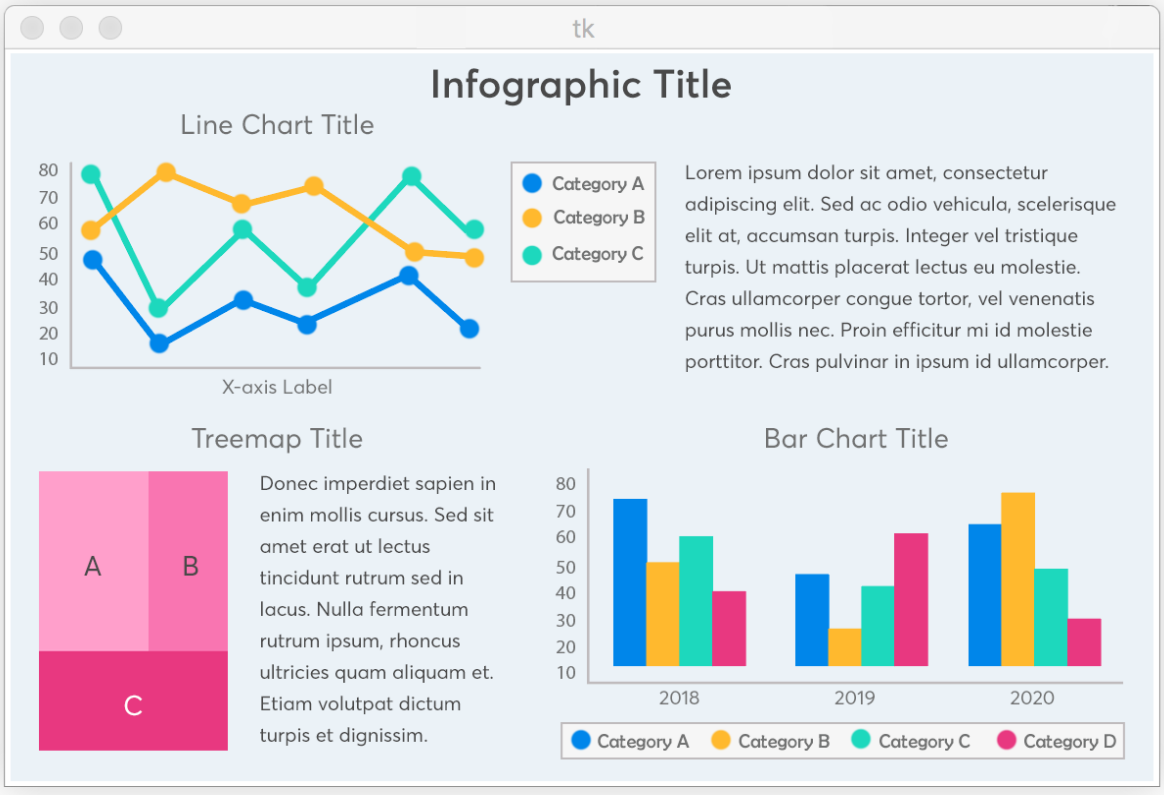
- provide the link to obtain the dataset so that we can verify your work. Put the link as a comment near the top of your program.
- submit the downloaded dataset in its original form.
- submit the cleaned or modified dataset if you changed the dataset in any way.
    - **Note**: You do not need to perform any data cleaning, but you can do it if you want. No marks will be given for cleaning the data.
- use meaningful file names for the dataset file(s).
    - For example, *data.csv* and *data_cleaned.csv*.

B. Program [130 marks]

Write a Python 3 program that generates an infographic using the *SimpleGraphics* library based on the dataset from Part A.

Sample Infographic

The following graphic is not based on real data and is simply included as a sample for what a good infographic could look like in terms of layout and colours; your own infographic need not look like this, although you are free to use it as a reference.



This infographic vector was originally created by pikisuperstar on www.freepik.com and was modified for our use.
Source: https://www.freepik.com/free-vector/infographic-element-collection_4173033.htm#page=3&query=infographic&position=43

Infographic Requirements

Your goal is to inform the readers about the information from your dataset via an infographic. Your infographic must show the readers what information they are looking at. This can be accomplished through the proper use of text, images, drawings, and charts. Your infographic

- must have at least **three** charts
    - All charts must have different chart types.
        - Bar charts and pie charts are **different** chart types, but horizontal bar charts, vertical bar charts, and stacked bar charts are all considered to be the **same** chart type.
        - Other example chart types are bar charts, line charts, flow charts, scatter plots, box plots, etc.
    - All charts must have titles, axis labels, and legends when applicable.
    - These charts **must exclusively** make use of the data from the submitted dataset
- should include text, images, and drawings as you see fit.
- must be informative and visually appealing.
    - A **significant portion** of your mark will be determined by how well you can convey the information from your dataset via the infographic **alone**.
        - In other words, by simply looking at the infographic, your TA should be able to understand the data being presented **without** knowing anything about the dataset that was used.

Tips: you should change the window size created by the *SimpleGraphics* library so that the infographic is big enough to clearly show all visual elements.

Marking Information:

Your program must meet the following requirements:

- read the dataset from Part A
    - You should include the filename of the dataset in the code that reads the file
    - **No user input**
        - i.e., when the program is run, an infographic will be displayed without any user input whatsoever
- use the *SimpleGraphics* library to draw your entire infographic
    - While you are allowed to be use images when creating your infographic, **all graphs** should be primarily created **with code** using the *SimpleGraphics* library
        - For example, the bars in a bar chart should be drawn using `rect(...)`, the axes should be drawn using `line(...)`, etc.
        - Images should be used as supplementary material **only**.
- If external images were used in your infographic, you **must** submit them so that your TA can run your program without any errors.
    - If your TA cannot run your code because you did not submit everything necessary to create your infographic, **you will receive zero for this portion.**
- If you decide to change the font for any text, ensure that you are using a **default** (not downloaded), **cross-platform** (available on most operating systems) font such as Calibri - this will ensure that your infographic looks as expected when your TA runs your code.

For full marks, your program **must**:

- have information about you in the form of comments at the top of the program. Include your name, ID number, and lab section. For example,
    - # Name (ID#): Sergio Marquina (1357911)
    - # Lab: 01
- have proper use of functions
- have proper comments
    - comments for functions must include
        - purpose of the function
        - descriptions of all parameters (if no parameters, state no parameters)
        - descriptions of all return values (if no return values, state no return values)
- use meaningful variable names and constants
- make appropriate use of if-statements, loops, and functions
    - A **significant portion** of your mark will be determined by how well you use these fundamental programming concepts.
- proper use of files, including opening and closing
- call the main function to start the program logic
    - the only lines of code that should appear outside of any functions are import statements, constants (if any) and the call to the main function

Each of the following may result in a **30 mark deduction** from your total grade:

- using any libraries that are not covered in labs or lectures
- using *break* or *continue* commands
- using global variables (note that global **constants** are fine)
- define one function inside of another function
- having more than one *return* statement in a function