CSC4730/6730, CMIS 4200   
Data Visualization

Summer 2019

Project #2

**Due date:** 11:59 pm,07/03/2019 (Wednesday)

**Requirements**

In this homework, you will learn to create statistical charts using data visualization libraries Plotly and Seaborn.

1. You will create 10 charts. Each chart is worth 10 points.
2. Load HollywoodsMostProfitableStories.csv and use **Plotly** to create the following charts. Every figure must include a title. Each axis must be labelled.
   1. A bar chart showing the profitability of the film. The X axis is the film. The Y axis is the profitability. The bars should be sorted from the most to the least profitable
   2. A histogram showing the number of films for each Genre. The X axis is the Genre. The Y axis is the number of films in the spreadsheet from each Genre.
   3. A figure with two bar plots: Worldwide Gross and Audience. The X axis is the film index. When the mouse cursor hovers over a bar, the film’s title should be displayed in the tooltip.
   4. A line chart showing the profitability of the films over the years. The X axis is the Year. The Y axis is the profitability.
   5. A dot plot showing the Rotten Tomato %. The X axis is the Rotten Tomato %. The Y axis is the film title.
   6. You must use Plotly API to create the charts. **Do not use Plotly Chart Studio.**
3. Load Housing\_price.csv and use Plotly to create the following charts. Every figure must include a title. Each axis must be labelled.
   1. A figure that contains three boxplots: price2014, squarefeet, and acre.
   2. A histogram showing the number of houses per Zip code. The X axis is the Zip codes. Create an annotation pointing to the Zip code with the most houses.
   3. A line chart with four lines: price1998, price2007, price2011, and price2014. The X axis is the house num. The Y axis is the value from the four columns listed above. When the mouse cursor hovers over a marker, the street address of the house should be displayed in the tooltip.
   4. You must use Plotly API to create the charts. **Do not use Plotly Chart Studio.**

1. Load wimbledons\_champions.csv and use **Seaborn** to create the following charts.
   1. A scatterplot showing when a player from different countries won the championship. The X axis is the country. The Y axis is the year. Each circle/dot indicates a player from certain country won the championship in a certain year. The circles should be filled with green color.
   2. Create a grid with four cells. In the first row, show two charts: a histogram showing the number of men’s champions for different countries and histogram showing the number of women’s champions for different countries.   
      In the second row, show two charts: a histogram showing the number of men’s runners-up for different countries and histogram showing the number of women’s runners-up for different countries.
2. Write your code in Jupyter Notebook. Submit both your Jupyter Notebook file (or Python code) and a PDF file.
   1. You can download your Jupyter Notebook as a PDF.
   2. The PDF file must contain all the data visualizations. If you submit files without any data visualization, even if your Python code is correct, you will lose half of the credit.
   3. Your Jupyter Notebook and the PDF file should be of the following format:
      1. Markdown cell: question number (e.g. Question 1.a) and the requirements
      2. Code cell: your Python code.
      3. Output: the data visualization
      4. Repeat the above pattern for each requirement.
   4. . If Jupyter Notebook doesn’t work, you may use Spyder. But you still need to create a PDF file that follow the above pattern: question number/requirements, Python code, data visualization, …

* **Late submissions policy.**
  + **There will be a 10% late submission penalty for each day that is late. For example, if you submit the project two days late, there will be a 20% penalty.**

**Deliverables**:

* Put your Jupyter Notebook and the PDF file in a Zip file. Submit the Zip file to iCollege under folder Project 2.
* Do not submit by email.