

# CS 503 – Data Visualization

## Assignment 2: Exploratory Data Analysis

### 1. Introduction

In this assignment, you will design two visualizations techniques for a small dataset and provide a rigorous rationale for your design choices.

### 2. Tasks

The dataset contains some important statistics from a large sample of movies. The data includes the movie budget and revenue from different sources as well as ratings from [RottenTomatoes](#), [The Numbers](#) and [IMDB](#).

**Step 1.** Pose an initial question that you would like to answer.

For example: Is there a relationship between columns? Are the columns IMDB rating and Production budget correlated? Is there any relationship between the movie budget and revenue?

**Step 2.** Assess the fitness of the data for answering your question.

Inspect the data--it is invariably helpful to first look at the raw values. Does the data seem appropriate for answering your question? If not, you may need to start the process over. If so, does the data need to be reformatted or cleaned prior to analysis? Perform any steps necessary to get the data into shape prior to visual analysis.

**Step 3.** Design two visualization techniques that you believe effectively communicates the data and provide a short write-up (no more than 4 paragraphs) describing your design. This design is must be implemented in Python. The Python codes provided on [Python data Visualization: demos](#), [Multi-Dimensional Data Visualization: Demos code](#), and [Visualization Tools: Python Libraries for Data Visualization](#) will give you an idea how to use different visualization techniques in Python. While you must use the data set given, note that you are free to transform the data as you see fit. Your chart image should be

interpretable without recourse to your short write-up. Do not forget to include title, axis labels or legends as needed!

As different visualizations can emphasize different aspects of a data set, you should document what aspects of the data you are attempting to communicate effectively. In short, what story (or stories) are you trying to tell? Just as important, also note which aspects of the data might be obscured or down-played due to your visualization design. In your write-up, you should provide a rigorous rationale for your design decisions. Document the visual encodings you used and why they are appropriate for the data. These decisions include the choice of visualization type, size, color, scale, and other visual elements, as well as the use of sorting or other data transformations.

### **3. Submission**

You must submit the *pdf* file of your report. In fact, it must contain the illustration of your visualization technique in term of images (three images that illustrates three visualization techniques) and a rigorous explanation. Indeed, your explanation must justify which one of the two visualization techniques will help the reader to give a relevant interpretation. The use of a multidimensional visualization (3D, 4D, etc.) technique will be an asset. In addition, you have to submit your *.py* or *.ipynb* if you prefer using *Jupyter Notebook*. **Please, do not submit your assignment in .zip or .rar files.** If you do a transformation to the dataset, you will have to mention it on your report and to submit it.