PROJECT

You will conduct exploratory on a dataset that you choose.

# Data Set Selection

Pick a dataset on a topic that you can (or willing to) raise interesting questions. It is better if you pick a topic that you are familiar with, or willing to spend some time on understanding it.

Pick a dataset that is suitable for predictive analysis. That is, the dataset should allow you to predict the values of a column based on the other variables in the dataset.

The prediction process should be meaningful. To explain it with an example, predicting potential ice-cream consumption based on temperature makes sense and useful especially for ice-cream sellers. Predicting temperature based on ice cream consumption is not very meaningful since the invention of thermometer.

Kaggle datasets are ok but you need to cite the project. Also, you can’t replicate existing analyses.

Here are two other data sources:

* <https://github.com/awesomedata/awesome-public-datasets>
* Data.gov

# Project 1: Exploratory Analysis

You will conduct exploratory analysis as discussed in the class. You will submit a report that includes the following sections. It can be in MSWord or HTML. It will be 3 to 5 pages.

In addition, you will upload your data file and the .rmd file that shows the calculations and creation of graphs.

## **Report Sections:**

Your report is expected to have the following sections:

## Introduction

Brief description about the topic and the project.

* How does your project help those who are interested in this topic?
* Why data analysis is needed for this?
* How do you expect your analysis to improve decision making in this area?

## Data

Describe your data;

* What is the source?
* What are your variables? If there are too many, focus on the ones that you will explore
* Describe the data cleaning process if you needed
* Provide descriptive statistics. Try to write a paragraph that gives descriptive statistics instead of just giving numbers in a table

## Exploratory Analysis

You will answer at least 4 or 5 questions (that you come up with). These questions will focus on whether

* a variable in your dataset have an interesting aspect
* a variable has any relationship with the variable you are trying to predict in the next step.

Each question will be answered with a graph.

## Results and Conclusion

You will finalize your report by summarizing your findings.

## **Project 1 Evaluation Criteria:**

You will be evaluated based on the following criteria:

**Problem selection:** Demonstration of choosing a topic/problem that is meaningful and applicable to data analysis. An example for a bad problem selection would be “understanding temperature changes by studying ice cream conception”. A better one is “understanding ice-cream consumption changes in relation to weather conditions”.

**Data:** Demonstration of good data selection. Your dataset should appropriate for the problem selected and it should be rich enough to conduct a thorough analysis. A rich dataset will have enough variables and observations.

**Analysis:** Demonstration of using techniques that you learned in the class. Your analysis should show some versatility. Try to use different graphs for the each of the questions that you are answering in your exploratory analysis.

**Results:** Demonstration of good interpretation of your exploratory analysis. After you answered the 4 of 5 questions in the analysis, you need bring it all together and give the summary (big picture) of your analysis. You will also explain briefly what your plan is for your predictive analysis.

**Report:** Demonstration of professional technical report writing. Your hard work will only be as good as you can communicate them. Since we don’t have any presentation for first project, your report will have to ‘speak’ for you. Make sure your report is concise but also consist of full sentences. Your assumption should be that you are writing this report for someone who is interested in your topic but hasn’t taken any data analysis training (i.e., your future manager). So ,to paraphrase Albert Einstein, “your writing style should be as non-technical as possible, but not a bit less”.

**Presentation:** You will be presenting your findings to the class with a formal presentation. There will be some discussions about the expectations. But in general, a well-rehearsed, engaging presentation with professional looking slides is expected.