

# INFO 1998 Project A Deliverable

## Data Analysis and Visualization

### Guideline and Rubric

**Release Date:** February 21th

**Due Date:** March 6th at 11:59 pm

**Submit Through:** CMS

### Overview

These first two lectures have been about data analysis, manipulation, and visualization tools. Now, we will look at the practical implementation of these skills with the train data from this [housing dataset](#) in a group of 3-4 people. This is a fairly open ended assignment; there are some basic requirements, but within those requirements you should approach the problem however you best see fit.

There is both a data analysis and visualization component for this project. First, for data analysis, you must write a summary of the data using summary statistics. Then, you must fill in missing values in the data (don't just delete rows or columns with missing data) and write a brief explanation justifying your decisions for data imputation. Second, for the data visualizations, you must produce two sets of visualizations using different column(s) in the dataset and all different types of graphs. The choice of features is up to you to decide. These should be proper graphs with a title and labels. Below the graphs, write a short description of what the visualization revealed to you about the data, specifically mentioning how the type of graph influenced your observations. Note that the visualization you choose should not be done just for the sake of graphing, but should give insight into the dataset.

Every group will produce **two sets** of graphs with **two graphs per set**. Each set will focus on the same column(s) in the dataset. You should produce set one with two graphs and set two on a different set of features with two more graphs. **Every visualization must be a different type of graph**. There will be a **maximum of half the points** awarded for the second set if the same features are used as the first. Additionally, **max of half the points** will be given for a graph if the type of graph is duplicated.

Criteria	Points
<b>Data Analysis</b>	
Description of the dataset using summary statistics	3
Proper data imputation	6
Explanation of filled in data	3
<b>Data Visualization</b>	
<i>Set One</i>	
<i>Graph One</i>	
Reasonable attempt	2
Title and labels	1
Easy Interpretation	2
Explanation of insight gained	2
<i>Graph Two</i>	
Reasonable attempt	2
Title and labels	1
Easy Interpretation	2
Explanation of insight gained	2
<i>Set Two (Max half points if same features used)</i>	
<i>Graph Three</i>	
Reasonable attempt	2
Title and labels	1
Easy Interpretation	2
Explanation of insight gained	2
<i>Graph Four</i>	
Reasonable attempt	2
Title and labels	1
Easy Interpretation	2
Explanation of insight gained	2
<b>Total</b>	<b>40</b>