



Homework 1 - Part 3

Who is CS7295

2. Data Type

City - categorical
Country - categorical
State - categorical
Native language - categorical
Languages known other than English - categorical
Preferred OS - categorical
Mobile OS - categorical
Favorite language - categorical
Chocolate or Vanilla - categorical
Preferred OS - categorical
Color - categorical
Favorite number - categorical
Temperate - quantitative
Amount of coffee - quantitative
Wake up - quantitative
To bed - quantitative
Visual encoding - quantitative

4. Data Clean-Up

To clean up the data, the data was exported to a new excel spreadsheet. Categorical data such as the Country which varied in naming convention (i.e.: USA vs United States) was manually corrected to be one category. In the category "what other languages do you speak besides English", I filled in blank values with None and also made multi-lingual students have comma separated values for languages. State abbreviations were spelled out as well or no values were changes to 'NA'. Some preprocessing was done around lengthy answers related to Mobile OS preference and removing words such as 'degrees'. Exclamation points were removed, numbers that were spelled out were many quantitative, and any temperature listed in something other than Fahrenheit was converted.

6. Data Gathering

From Piazza I compared the information hints as to what text blurb matched with the data in the csv. I simply inserted two rows for department (CCIS or other) and what degree program the individual is pursuing. There wasn't a great deal of cleaning the data here.

7. Additional questions

1. Is there a correlation between preferred programming language and preferred OS?
2. What percentage of the class speaks another language if their native language is not English?
3. What are peoples favorite color?

9. Visualizations

@TODO - PUT IN ANOTHER PDF / DOC