

INFO I590: Data Visualization

Visualization Critique – 01

Author: Sanjana Pukalay (spukalay)

Source: <http://www.r2d3.us/visual-intro-to-machine-learning-part-1/>

1. Overview (What is the visualization about?)

As the title says, this visualization is a picturesque tutorial about the various machine learning models. An example about homes from New York and San Francisco is taken which walks us through how to develop a machine learning model which essentially distinguishes between homes from different areas. Multiple visualization techniques are used as part of this tutorial which makes it quite interesting and easy to interpret.

2. Explanation of Data

The dataset used for this visual tutorial is about the homes from New York and San Francisco areas and contains various attributes like elevation, dimension, cost etc. to describe them and which can make them unique which is basically what is taken advantage of to make the distinction. A classification model is developed using this data.

3. Explanation of Visualization Techniques

The attributes of the homes are described in brief. Some concepts regarding feature engineering is explained using a scatterplot which distinguishes homes based on their elevation. A scatterplot matrix is used to show the relationship between various attributes used to develop the model. This is what we call an exploratory data analysis which helps give an idea as to how the data looks like and what appropriate method should we use to handle this kind of data. As we move further into the tutorial, the scatterplot is very interestingly converted into a histogram which helps to see better how frequently homes appear at each elevation. Pie charts are used to describe the evaluation metrics like accuracy, type-I,II errors etc. Ultimately, a combination of the above mentioned visualizations paired with the concept of decision trees is used to give an overall picture of how each step works and how to determine what is the best methodology to use.

4. Effectiveness of the Visualization

The visualization techniques used in the tutorial seem pretty effective and overall the visualization looks very clean and presentable. It was certainly able to achieve the objective which was 'Introduction'. However, the concepts were not explained in depth and more visual elements could have been used to describe it further. A lot of other alternative visualizations can be explored and experimented with.

5. Integrity of the Visualization

As mentioned, pie charts used to describe the evaluation metrics. Many a times, pie chart is not a very effective way to visualize data. Bar graphs or histograms do a

better job at giving a clearer picture. There doesn't seem to be any significant biases in the visualization.

6. Design

The colours used for this visualization are very pleasant to the eyes. Most importantly, there are limited number of colours used avoiding much of the ambiguity. However, a more contrasted colour combination would have been a better way to distinguish between the labels.