Data Representation Samples

Vedant Shah

December 24, 2020

Graphs Using R

These are a set of graphs that are considered to be good which I had to make in R programming. One of the major philosophy of the professor was to not have labels of the X-axis and the Y-axis (For figure 1 and 2). The heading of the graph should be self sufficient. My opinion: This might be true in scientific/research papers as the graph has to be majorly self explanatory with less nomenclatures on it, in business and presentations having legends and labels makes it easier to understand the graph.

Multi-Factorial Graph

Figure 1:The data is a multi-factorial dataset making putting two variables on the x-axis with multiple conditions in each variable. Each condition is represented with different shapes and the label is provided.

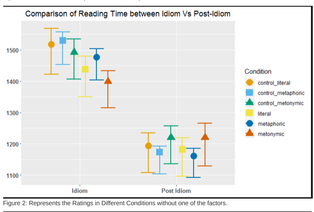
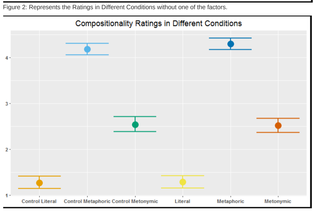


Figure 2: Represents the Ratings in Different Conditions without one of the factors.



Graphs Using Python (Matplotlib and Bokeh)

Figure 3: This graph can be taken an example of the data distribution given different factor.

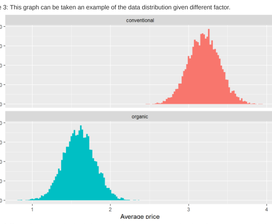
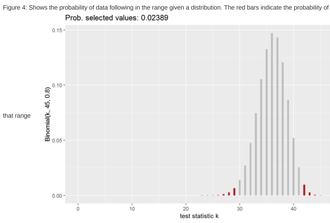


Figure 4: Shows the probability of data following in the range given a distribution. The red bars indicate the probability of data(0.02389) following in that range



Graphs on a hand tracking dataset

Figure 5: Shows a simple regression of of correct or non correct answers (0 representing as incorrect). Its an interaction between the reaction time of the participant and the Maximum absolute deviation. The graph manages to show the distribution of the data and the regression for both condition and its difference

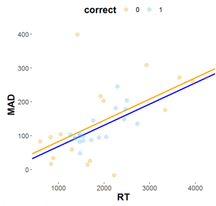
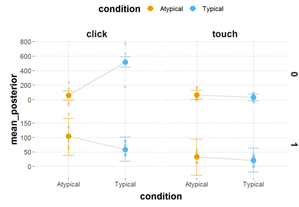


Figure 6: Performed a Bayesian analysis on the data in hand and calculated the mean of the data with 3 factor in consideration. The condition (typical and atypical), method (click and touch) and answer (Correct or non Correct). The graph also represents the concentration of data and the confidence interval and the mean. The line connecting the conditions is the regression line between two conditions.



Neural Network Performace Analysis Graphs

Figure 7: Graphs shows the loss, which is prediction error, in the neural network learning of each epoch. The graphs shows a decrease in loss in the training and the test set indicating a good learning pattern.

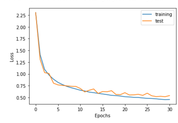
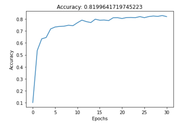


Figure 8: Shows the Accuracy of the neural network over each epoch.



Vector Dimension Comparion of Image Data

Figure 9: Original 2 dimensional vector Representation of the images which were categorized into 10 categories. The dataset shows the distribution and the color shows each category. This graph could use a label for each colour which is missing.

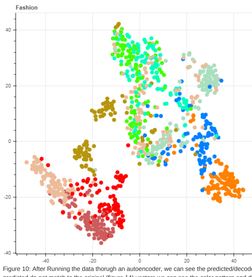
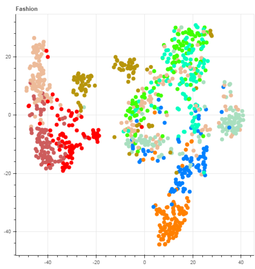


Figure 10: After Running the data thorugh an autoencoder, we can see the predicted/decoded vectors of the images. Now the vectors which are predicted do not match to the original (figure 14) vectors we can see the color pattern and distribution predictions are similiar. The vectors are not matched for accuracy, the predication of the category is compared. So even though the network did not get the vectors right it got a good accuracy in predicting the categories and the colors and the graph pattern indicates.



Extra Comments

1. I have also worked on Pie Charts, Word Clouds and Network Graphs
2. The above are examples of few of the graphs as a student in my masters
3. The data visualization software I have used before is Kibana. The packages I have used in R is Ggplot and for python is matplotlib and Bokeh. I am currently working on MNE package in python for neuroscientific data.
4. There is a course I did in my masters on basic visualizations where I had to sign a data protection form so the data in figure 1 and 2 are not actual data.