

DATA VISUALIZATION IN TABLEAU:

Create an **explanatory** data visualization from a data set that communicates a clear finding or that highlights relationships or patterns in a data set. Your work should be a reflection of the theory and practice of data visualization, and your final deliverable will be a write up along with a Tableau Public workbook.

SUMMARY:

For this project I have chosen to work with the PISA data set.

The data set itself was very huge, the size was over 2GB. So for the purpose of this project I trimmed the data first using Python and then imported the new file to Tableau for Visualization purpose. My Tableau story contains a total of 7 dashboards containing various charts and mapping data. My visualizations mainly give an idea about how each country ranks up in the PISA test, the study approach of the top performing countries vs the bottom ones and wealth information of each nation.

DESIGN:

Data Wrangling:

The data set for PISA test information was over 2.5 GB, so it was impossible to load it in notepad++. It would hang up the PC while doing this. So for purpose of the project I referred the PISADICT2012.csv file which had key value pairs showing what each field meant. Referring this I used the key value pairs I needed for my analysis. The file "**PISA_dict final mappings.xlsx**" contains those columns I have used. So first I trimmed the data using Pandas and after that I changed the headers to the values given in the dictionary csv file. More details are given in the jupyter notebook attached "**PISA_Dataset.ipynb**".

Data Visualization:

Next I imported the new csv file to Tableau. Now since the data set had information about countries and measures which gave performance of the countries. So I used multiple chart types. **First a mapping chart had to be there, a bar chart was a must. But as I progressed in the story making process an Area curve, Scatter plot, Pie chart and simple tabular data format followed.** Now since the data set had country data so I thought of displaying the data in a map, this was the first criteria. Now each country has different average scores for the PISA test. So it is a good idea to show the scores in a bar chart for each country since bar chart is easy to understand. Separate color for each of these charts has been used. I have used Red and Green color in most of the charts when highlighting a nation's performance. For showing OECD and Non-OECD countries I have used bar chart again. Since there is only one dimension – the number of countries. Next there is a data containing class size and learning time. So I thought why not use scatter plot to check correlation between these variables and scores of each country. By this way I can get an indicator to prove that these parameters actually have an effect on the test performance. While examining wealth index of each country I saw it had negative values also. So a line chart was enough for this, but to make the visualization more appealing I made an area curve. Now when seeing the data for parents I had a hard time to

think how to display it since for each country there was 4 parameters to check, so I have kept it in a tabular form so that the user can use the scroll bar and check the data for the countries. For my next chart I have used Pie chart for comparing different parameters for a single nation like Do they initiate industrial tours or do they host quiz sessions on practical applications of the things they have learnt. So we can see the percentage of people in a nation who undergoes through all this and who does not. Pie chart is most suitable for this kind of comparison between parameters. I also used Pie chart for visualizing some student learning parameters, but after getting the review I changed them to bar chart as it was easy to understand.

Green means – The nation has scored above average in the PISA test

Red means – The nation has scored below average in the PISA test

I have also labelled the marks for helping out the user to understand each mark. There were a lot of null data for some of the countries so I renamed them to “Did not participate” since that must be the only reason the data set does not have feedback about some of the reviews.

Link to the Data visualization in Tableau public server:

https://public.tableau.com/profile/anirban7826#!/vizhome/PISA_Observations1/PISADataAnalysis?publish=yes

FEEDBACK:

I gathered feedback from 2 people and those are as follows

Feedback #1

One very important rule that I have learned from Udacity and other places. You should hardly use pi charts. Pi charts are said to be poor visualisation technique. A human being can understand height and width much better than area and radius. Bar chart is hence always better.

Feedback #2

- The text at the top is helpful, but may be a bit more than preferred, since we want to make visualization the focus. So consider using more succinct language and limit to one to two sentences.
- I'm not exactly sure how the linear fit line works in Tableau, but the p-value (when you mouse over) is not significant, therefore, it'll be farfetched to claim a strong linear relationship here. On the other hand, it's also not founded to claim causal relationship ("affect"), without experiment, we're dealing with observational studies which only support association.

Feedback #3

It's good overall. I found the pie chart a bit confusing, as the percentages do not add up to 100%. And same for the following slide. Consider either using a different graph type, and/or present the numbers in a different way (maybe do some initial calculation?).

DESIGN POST FEEDBACK:

After taking the feedback these were the things I changed.

There were many null values in the data set and that's why the percentages in pie chart was not adding upto 100 because I hid the null values. Now I have renamed them to "Did not Participate" hopefully it makes sense.

For one of the dashboard which contained only pie charts, I changed them to bar graph as per the suggestion.

In the stories section I edited the wordings which said that "something had caused something" since we did not perform any experiment this type of prediction would be dangerous. Since there are no proof to suggest if it is totally true. These are just data from different nations.

Link to the data visualization post feedback:

https://public.tableau.com/profile/anirban7826#!/vizhome/PISA_Observations_Final/PISADataAnalysis?publish=yes

RESOURCES:

1. http://onlinehelp.tableau.com/current/pro/desktop/en-us/help.htm#dashboards_sheet_selector.html
2. <https://public.tableau.com/en-us/s/blog/2015/06/rough-guide-dashboard-actions>
3. <https://discussions.udacity.com/>
4. <http://www.oecd.org/pisa/>
5. [PISA Data Analysis Manual - SPSS, Second Edition](#)
6. Stack overflow
7. <https://www.youtube.com/watch?v=boJcT-lerFQ>
8. <https://www.youtube.com/watch?v=IDyMMPiNVGw>