



COS30045 Individual Assignment

NAME: Ng Kang Jen
STUDENT ID: 105002263

Table of Contents

1. Introduction	2
1.1 Overview of the Assignment	
1.2 Objectives of the Report	
2. Visualization 1: Risk Factors for Health	3
2.1 Dataset Overview	
2.2 Analysis of the Visualization	
2.3 Recommendations for Improvement	
3. Visualization 2: Actual Salary of Teachers and School Heads	4
3.1 Dataset Overview	
3.2 Analysis of the Visualization	
3.3 Recommendations for Improvement	
4. Visualization 3: Total OECD Development Assistance for Education by Level	5
4.1 Dataset Overview	
4.2 Analysis of the Visualization	
4.3 Recommendations for Improvement	
5. Conclusion	6
5.1 Summary of Findings	
5.2 Importance of Data Visualization Integrity	
5.3 Lessons Learned	
6. References	7
6.1 Sources Used for Data and Visualizations	

Introduction

For this assignment, I will be analysing 3 datasets taken from OECD website. The objective of this assignment is to improve our skills at judging datasets more effectively. I hope that I can improve on understanding and presenting data more clearly after this assignment. When doing the assignment, I will study each dataset in depth, what is the context and what kind of data it contains? I will give critiques to the visualisations of the datasets according to concepts of good data visualisation and suggest anything that can improve it. Before I analyse, I will use Tufte's Integrity Principles as guidelines for myself. The principles focus is making sure that numbers are accurate according to the dataset, labels are clear and visible, and data is presented with context without any distortion. If I follow these principles, this will ensure that the analysis is thorough and meaningful.

The formal list of principles is:

- Make sure the labels are explicit, thorough, and all-encompassing to avoid confusion or misunderstanding stemming from visual distortions. Include explanations on the chart for better comprehension and clearly mark important data points that everyone wants to see.
- Emphasize displaying differences in the data instead of differences in design features. The focus should be on presenting the data accurately for better comprehension, rather than on choosing visually appealing designs which is really not necessary.
- It is usually better to utilize deflated and standardized units of measurement rather than nominal units when presenting monetary data over time. This method considers inflation and gives a more precise depiction of shifts in buying power throughout time.
- Do not use more dimensions in the visualization than those present in the data. Focus on illustrating only the necessary factors to avoid unnecessary complexity and disorientation.
- Make sure that visuals correctly depict the data's context and avoid misrepresentation through isolated data quotes. Be honest and open when displaying information to uphold trustworthiness and honesty between the researcher and the public.

VISUALIZATION 1

Risk factors for health ⓘ

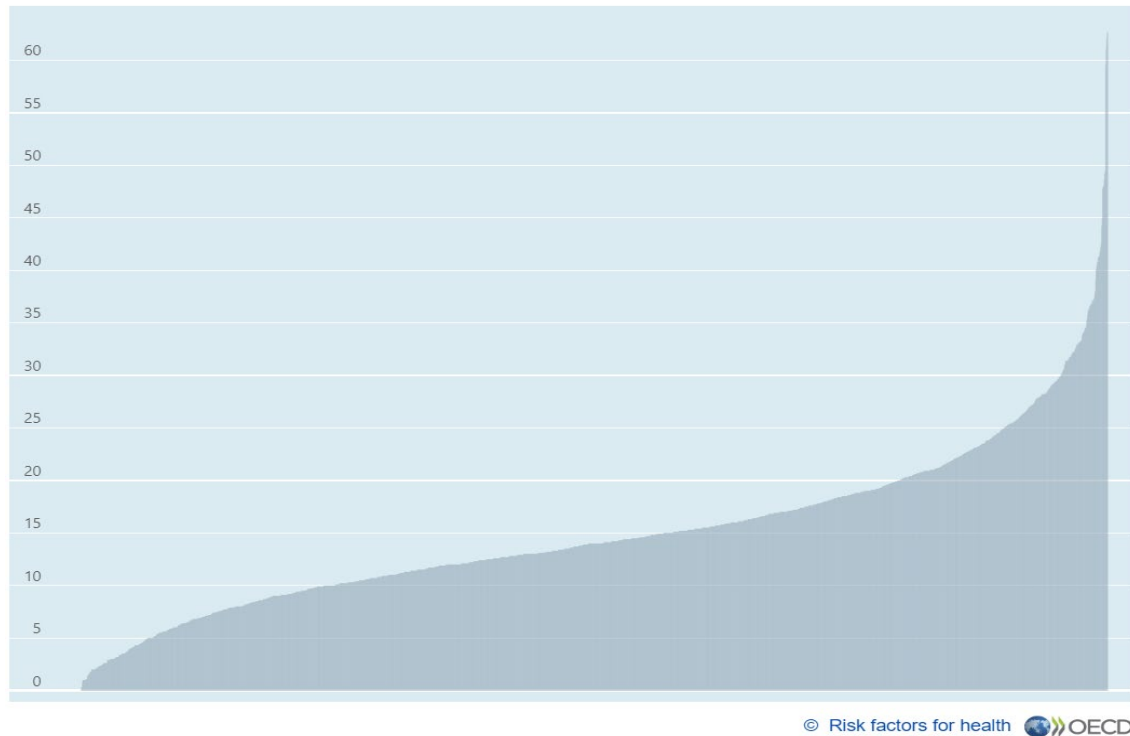


Figure 1: Risk Factors for Health

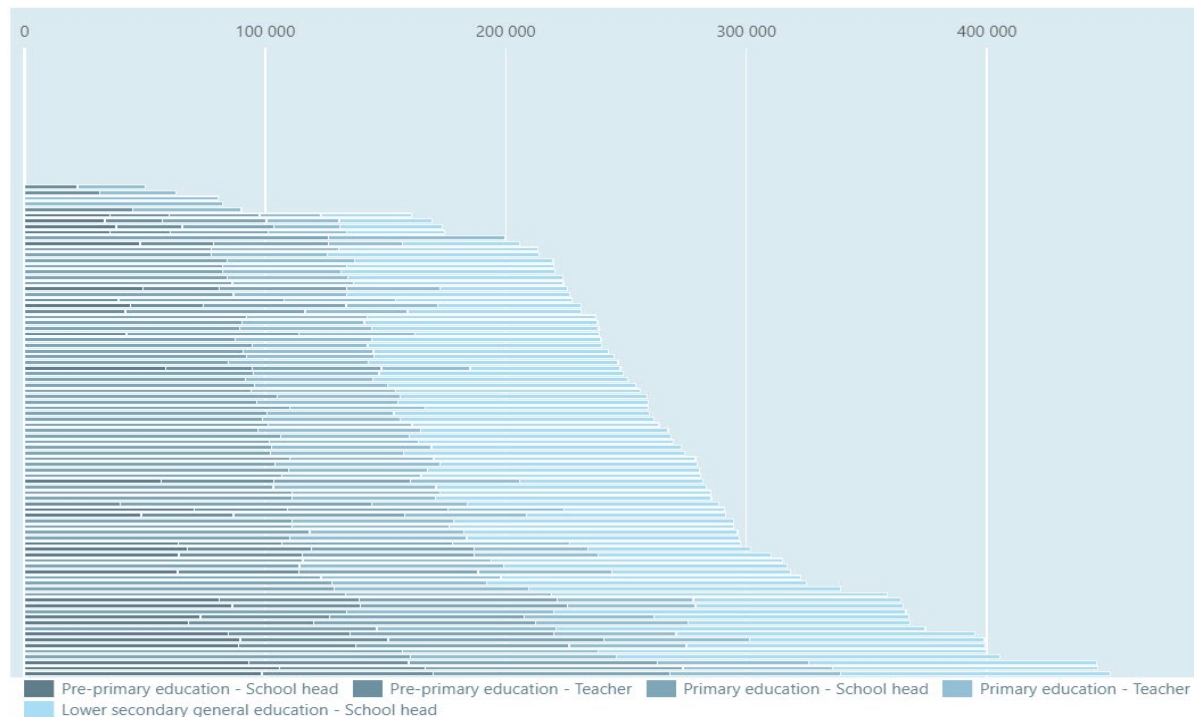
Figure 1 shows Risk Factors for Health which I sourced from OECD Data explorer website. This dataset shows the updated data for several risk factors for health such as tobacco consumption, vaping, and many more. It is also by several countries such as United States, United Kingdom, and many more. The Y-axis is based on the percentage of the risk factors such as percentage of population that are daily smokers (i.e. 15 years and over – Female), total cigarettes per smoker per day, and percentage of population that are obese (self-report). This chart here has 3106 data points. Referring Tufte's Integrity Principle, there isn't any label which makes this chart incredibly hard to read. The data density is very high as well which includes from the year 2010 to now, which makes it hard for us to find what we want in this dataset. Even though, there is a customisation feature to allow us to find what we want, for most of us normal people, we would want to find our data easily. The chart should be displayed in a visible manner once we searched for it. The improved chart should include labels for countries and different colours for them as well. Considering the huge amount of data point and the context behind the chart, we can also consider changing the type of chart to use for this to like several smaller bar charts, grouped bar chart, or scatter plot. This ensures viewers can easily tell the difference when accessing the chart. Other than separate charts, we can bin or aggregate data into broader categories to reduce visual clutter. For example, by region instead of country, or by year range rather individual years. This would reduce the number of data points and makes it easier to recognize something at a glance.

VISUALIZATION 2

Actual salary of teachers and school heads - full dataset

Measure: Actual salaries • **Type of educational institution:** Public educational institutions • **Age:** From 25 to 64 years

Combined unit of measure: US dollars, PPP converted, Units



© Actual salary of teachers and school heads - full dataset 

Figure 2: Actual Salary of Teachers and School Heads

Figure 2 shows Actual Salary of Teachers and School Heads which I sourced from OECD Data explorer website. This dataset shows the updated data for teachers and school heads salary up till May 2024. It is also by several states in the United States and other countries as well. The X-axis is based on the value of their salary and how much they cumulated.

This chart here has 864 data points. Referring Tufte's Integrity Principle, there isn't any label to differentiate between states in the US and countries which makes it hard for normal people like us to analyse. The data density is very high as well which includes US states and countries as well, which makes it hard for us to find what we want in this dataset. The colour choices made it worse as we couldn't really differentiate the data because of high density. The categories as well has a very inconsistent grouping like we don't know why specific education levels is grouped together. Even though, there is a customisation feature to allow us to find what we want, for most of us normal people, we would want to find our data easily.

The chart should be displayed in a visible manner once we searched for it. The improved chart should include labels separately for countries and states as well as vibrant colours so we could really tell the difference, and if possible separate teachers and school heads entirely and then compare education levels within each group. This will make it ten times more easier to compare the roles within a specific education level.

VISUALIZATION 3

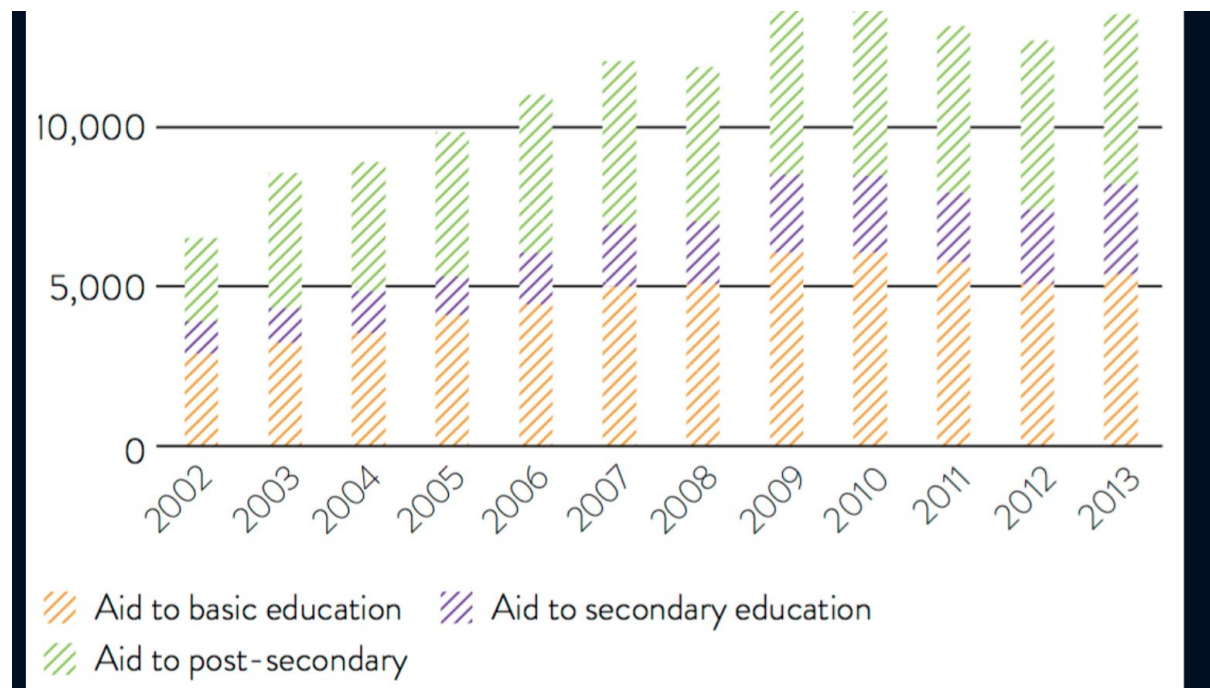


Figure 3: Total OECD development assistance for education by level

Figure 3 shows Total OECD development assistance for education by level which I sourced from ourworldindata.org website. This dataset shows the data for education aid from 2002 to 2013. The education is allocated towards mostly low-income countries that really need the aid. The X-axis is how much aid were allocated for education. Referring Tufte's Integrity Principle, the chart is using stacked bars with similar patterns. The use of diagonal stripes for all categories (aid to basic, secondary, and post-secondary education) makes it hard to tell the difference between them. The colours (orange, purple, and green) have big difference, but the overlapping stripes may still cause confusion, especially for viewers with colour vision deficiencies. Since this is a stacked bar chart, it's essential to ensure that the groupings of aid types are easy to differentiate. However, it's unclear whether the aid amounts add up correctly, and the grouped bars could give a misleading impression of total aid when comparing the years. For example, viewers might find it difficult to tell how much aid is allocated to secondary education vs. primary or post-secondary. Also, the diagonal hatching on top of the colours might add unnecessary complexity. Since colour is already being used as a standout factor, the hatching could be removed for a cleaner and more readable design. To make the chart look better, we can use solid colour instead for the bars so we can tell the differences more clearly. We can also add a label to the X-axis so that when people view it, they don't misunderstand it. Finally, we can separate the data into multiple graphs or using another chart type like grouped bar chart or line chart so that we can analyse the comparison between the aid categories more easily.

CONCLUSION

To conclude this assignment, the analysis of the three visualizations shows the importance of accuracy, simpleness, and thoughtful design in effectively showing data our data to the public. Through these critiques that I did, clearly that even when a visualization is full of mind-boggling stats and data, state of the art looks, the poor design choices, such as high-density layouts, wrong choice of colours, and lack of appropriate labels for the axis, can lead to confusion and misinterpretation between people that view it. This is probably because not all people are familiar with complicated visualizations. Key factors like the choice of colour that we picked, clearness of labels, and structure of the graph are crucial in making sure that the intended message of the stats and data is communicated effectively to the public eyes.

After I analysed each of the visualizations, they have a couple of issues in common: hard to differentiate between the data categories, compact design due to huge amount of information, and not enough context behind the labels, which made it harder to analyse the stats and data. However, the visualizations also demonstrated the potential for improvement through more user-friendly and informative approaches, such as use colours that have better contrast between each other, more designs that are clean, simple and free of unnecessary elements, and labels that clearly show what it is for.

Ultimately, this assignment highlights the importance of smart design decisions in visual portrayal. Even with accurate data, a poorly designed visualization can prevent its purpose of making sure viewers be able to take meaningful stats and data from it. Design improvements that went through careful planning can dramatically enhance how effectively visualizations tell their intended message, making them more accessible and impactful for their audience. I have learnt a lot of things through this assignment including integrity of our visualization, the portrayal of it, and the content of it.

REFERENCES

1. Roser, M., & Ortiz-Ospina, E. (2024, March 12). *Education spending*. Our World in Data. <https://ourworldindata.org/financing-education>
2. Risks factor for health. (n.d.). OECD Data Explorer. [https://data-explorer.oecd.org/vis?df\[ds\]=dsDisseminateFinalDMZ&df\[id\]=DSD_HEALTH_LVNG%40DF_HEALTH_LVNG&df\[ag\]=OECD.ELS.HD&df\[vs\]=1.0&dq=.A.SP_OBS%2BTBC%2BSP_DS...PT_POP%2BCG_SM_D.&pd=2010%2C&to\[TIME_PERIOD\]=false&vw=br](https://data-explorer.oecd.org/vis?df[ds]=dsDisseminateFinalDMZ&df[id]=DSD_HEALTH_LVNG%40DF_HEALTH_LVNG&df[ag]=OECD.ELS.HD&df[vs]=1.0&dq=.A.SP_OBS%2BTBC%2BSP_DS...PT_POP%2BCG_SM_D.&pd=2010%2C&to[TIME_PERIOD]=false&vw=br)
3. Actual salary of teachers and school heads - full dataset. (n.d.). OECD Data Explorer. [https://data-explorer.oecd.org/vis?fs\[0\]=Topic%2C1%7CEducation%20and%20skills%23EDU%23%7CTeachers%23EDU_TEA%23&pg=0&fc=Topic&bp=true&snb=31&vw=sr&df\[ds\]=dsDisseminateFinalDMZ&df\[id\]=DSD_EAG_SAL_ACT%40DF_EAG_SAL_ACT_ALL&df\[ag\]=OECD.EDU.IMEP&df\[vs\]=1.1&dq=..USD_PPP..ISCED11_02%2BISCED11_1%2BISCED11_24%2BISCED11_34.Y25T64.T.&to\[TIME\]=false](https://data-explorer.oecd.org/vis?fs[0]=Topic%2C1%7CEducation%20and%20skills%23EDU%23%7CTeachers%23EDU_TEA%23&pg=0&fc=Topic&bp=true&snb=31&vw=sr&df[ds]=dsDisseminateFinalDMZ&df[id]=DSD_EAG_SAL_ACT%40DF_EAG_SAL_ACT_ALL&df[ag]=OECD.EDU.IMEP&df[vs]=1.1&dq=..USD_PPP..ISCED11_02%2BISCED11_1%2BISCED11_24%2BISCED11_34.Y25T64.T.&to[TIME]=false)