ANALYSIS OF GROSS DOMESTIC PRODUCT PER CAPTIA(GDPPC) & SOCIAL ECONOMIC STATUS(SES) USING TABLEAU

By

Radha Rani Kommineni

PS ID: 2156574

Introduction

The data used in this analysis was downloaded from Kaggle and is freely available. It originally contained 1000 data samples and 5 columns with four of them being numerical and the rest categorical. The data is clean and has no missing values and thus missing values treatment was not carried out in this analysis. This GDP data frame contains data like country, gdppc, SES, popshare and year. Data visualization been a wide topic have used here to analyze and visualize my data based on Tufte's and Colin ware's theories using line and bar graphs.

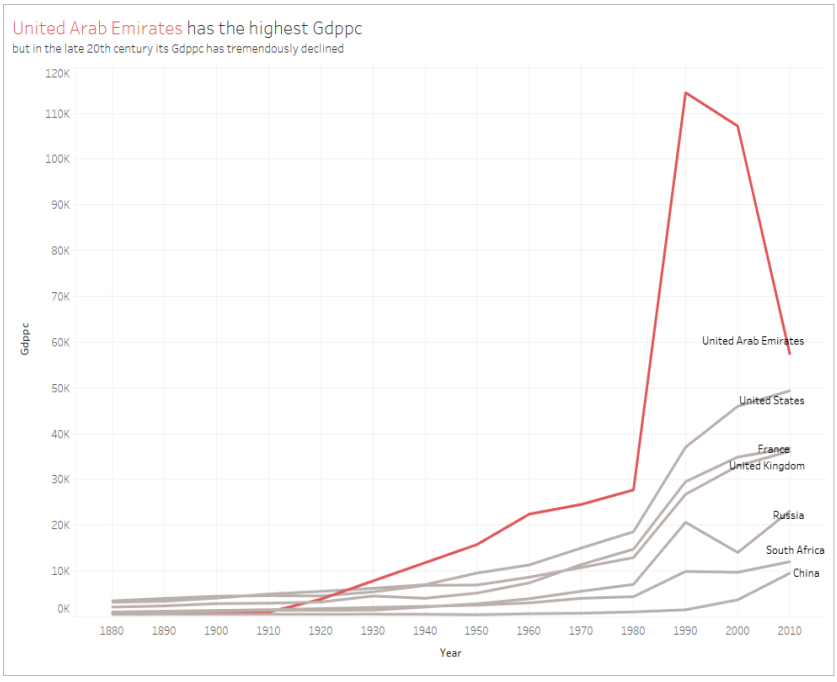
Discussion

The data was explored and two theories in data visualisations were employed, Tufte's and Colin Ware's theories. The principles of visualisation [Edward Tufte 2001] are:

- Show – and emphasise – the data
- Induce the viewer to thinking about the substance of the information as opposed to the design of the graphic
- Avoid distorting what the data has to say
- Make large data sets coherent
- Encourage the eye to compare different pieces of data
- Reveal data at several levels of detail, from a summarised view to fine structure
- Serve a reasonable and clear purpose: description, exploration, tabulation – or decoration if that's the objective.

Tufte suggests that the less the non-data ink in a visualisation, it distils it to its essence and aids in better comprehension.

Using Tufte's Theory

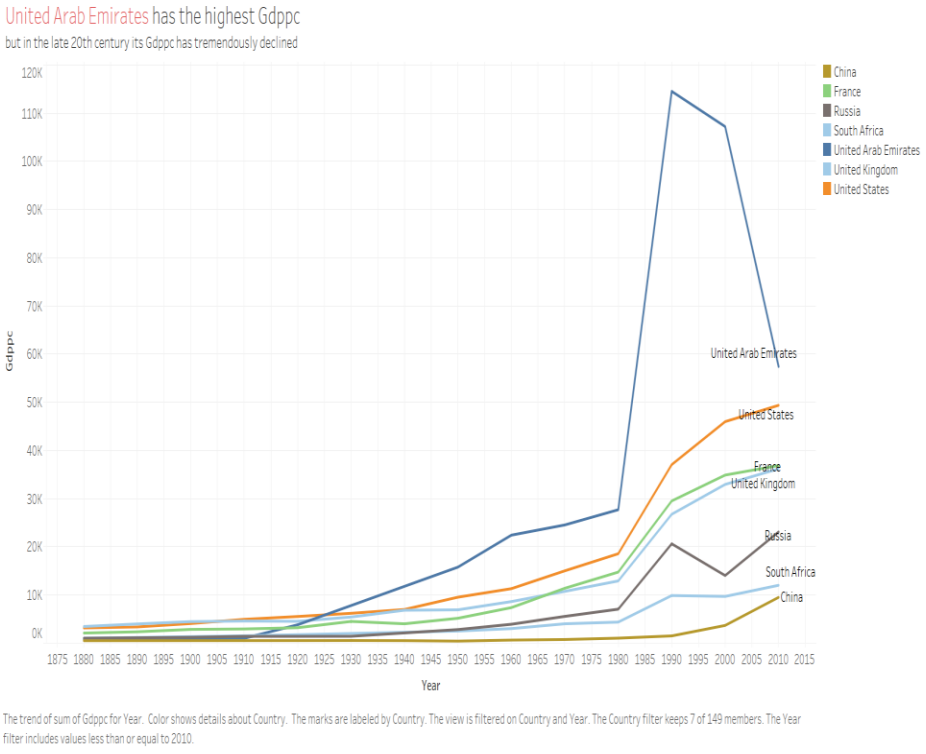


Graph1: gross domestic product per capita (gdppc) against year

Tufte's theory argues that visualization should have less ink print and represent great number of amounts in a small space. To ensure that I stayed in between the theory I used a white background and used light colors so that the lines can be seen with ease. In order to ensure that I did not twist the data I used it as provided by my data frame. I avoided using gridlines and borders to avoid them standing out in my presentation.

In the above chart, compared gross domestic product per capita (gdppc) of seven countries namely China, France, Russia, South Africa, United Arab Emirates, United Kingdom and United States. I erased some of the countries data which I wasn't interested with in this analysis. Colin's work was an advancement of Tufte's work and has been used greatly in the fields of data science and analytics. In his book, [Colin Ware, 2004] Colin provides a comprehensive and quite technical treatment of the theory, mechanics, and practice of visualisation. It delves deep into the theory behind why certain visualisation techniques work better than others. This is mostly discovered by the use of colours in a visualisation. Some colours work better than others in different backgrounds. For example, yellow is more visible in a blue background. In visualisations, having grid lines that are faded makes the chart or graph more visible and less wording makes it difficult to be distracted from the information being conveyed by the visualisation.

Using Tufte's and Colin Ware's theories



Graph2: United Arab Emirates gross domestic product per capita (gdppc) in my pivot table

In this presentation I applied Tufte's and Colin ware's theories which states that ideas should be passed with clarity, efficiency and position. In the graph 1, all secondary element lines are grey colored and a red line is highlighted which shows main information. In graph 2, I had used

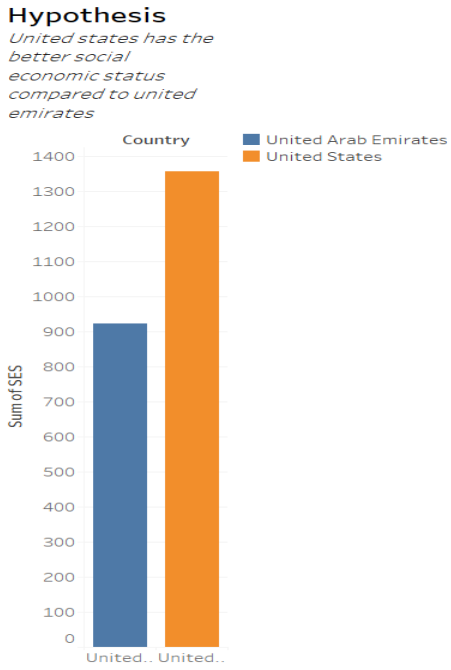
different colors to differentiate and to apply human perception which provides good view or insight to a specific data which in my case was considering the gross domestic product per capita (gdppc) of United Arab Emirates. As you can see the gross domestic product per capita (gdppc) of united emirates experienced a skyrocketing growth in between 1980 and 1990 before it experienced a tremendous decline from then. Title is placed on the top left of the graph so viewer will read it first before looking at the graph. This was to ensure that one will understand what the chart it's all about and I even provided a brief subtitle.

Data hypothesis

After the data exploration, a hypothesis was formulated. This hypothesis was formulated from the need to understand which country has high Social Economic Status among US and UAE. This hypothesis is the null hypothesis and the alternative hypothesis, though not stated, is the negation of the null hypothesis.

Created a hypothesis stating *United States has a better social economic status compared to united emirates.*

Research question- Does United States have a better social economic status compared to united emirates.

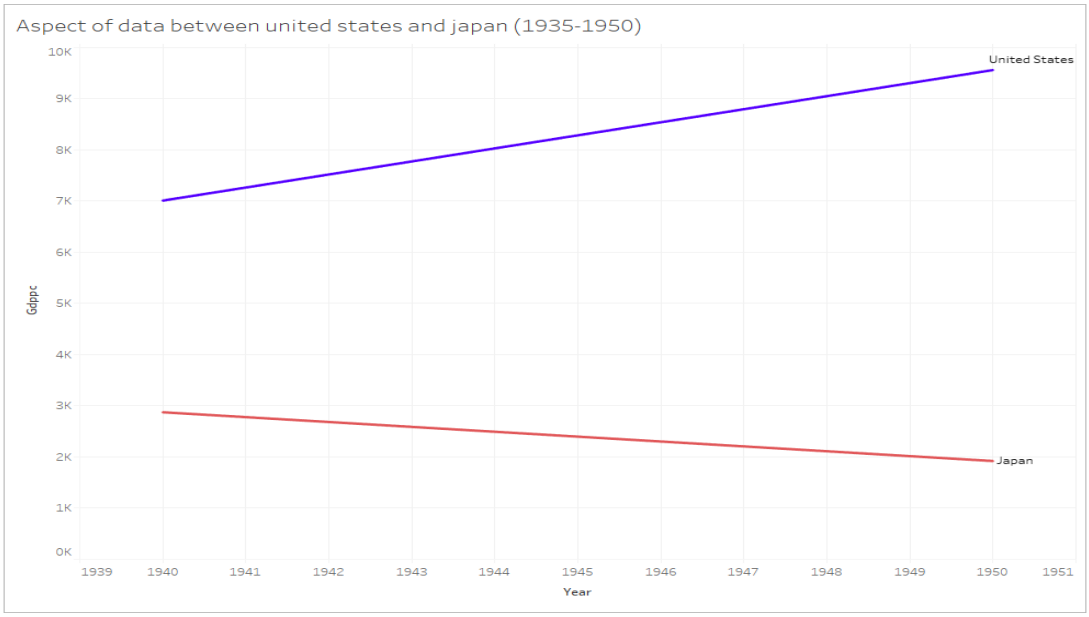


Graph3: hypothesis comparing united Arabs and United States Social Economic Status (SES)

Aspects of the data

Data aspects are used to bring about effects of war on gross domestic product per capita. Therefore, compared the gross domestic product per capita (gdppc) of United States and Japan during the World War 2 which took place between 1939 and 1945. I took a larger range of 1935 to 1950 because war creates its big impact on the economy before and after the war. Prior to the war, the governments redirected most of their resources to fund the production of war machines

and training of the personnel. In our case, Japan involvement in World War II led to stalling gross domestic product per capita (gdppc) since most of its population was recruited in the military and few skilled laborers were left to work in the industries



Graph 4: Data aspects showing effects of World War on Gdppc for Japan and US

United States gross domestic product per capita (gdppc) grew because most of its people were involved in industries for production of machines for the already established military. This meant that the people were earning more and improving their life. After the war, Japan economy was struggling since they had to rebuild all their industries destroyed by the war and fix all its logistics supplies. For United States, its economy had benefited from the war. This data was valid and accurate in proving that a war has a big impact on the gross domestic product per capita (gdppc) of a nation.

CONCLUSION

Data should be visualized in a clear, efficient and simple manner. This ensures that data is reliable and valid. In my hypothesis I failed to reject the null hypothesis that United States has a better social economic status compared to united emirates.

REFERENCES

- *Tufte, Edward R (2001) [1983], the visual Display of Quantitative Information (2nd ed.), Cheshire, CT: Graphics Press, ISBN0-9613921-4-2.*
- *gdp and ses of countries | Kaggle*
- *Information Visualization - Perception for Design, Third Edition (2012) or Fourth Edition (2020). Morgan Kaufmann.*