

Starting with Tableau

Taten Knight
tatenhknight@lewisu.edu
DATA-53000, Summer 2022
Data Visualization
Lewis University

I. INTRODUCTION

The goal of this week's project is to get familiar with Tableau by visualizing a dataset of our choice, and providing visuals that illustrate our findings within the dataset. We also needed to take into account lessons from this week on data visualization. I chose a dataset from the UN, covering population, surface area, and density data for different regions worldwide [1].

II. DATA DESCRIPTION

The UN dataset is fairly simple. It has various columns covering population breakdowns, surface area, and population density. A full schema is given below in Table 1.

TABLE I. DATA ATTRIBUTES

Attribute	Type	Example Value(s)	Description
Region	Categorical	China, Western Europe	Geographic Region
Year	Date	2011	Year
Series	Categorical	Population mid-year estimates (millions)	Description of Data
Value	Numeric	100	Value of Series
Footnotes	~	~	Notes
Source	~	~	Data source

III. METHODOLOGY AND RESULTS

I chose a simple dataset in order to get more familiar with Tableau, but it was difficult to find one that provided a wide enough range of well formatted data to create quality visualizations. Because of this I explored a few datasets before settling on this one. I first looked through the data in a csv and added appropriate column headers as needed so that Tableau was able to correctly interpret the data. Then I looked through the series values and decided on things that may be interesting to see. The first visualization I chose is the 'Treemap' of population density shown in Fig 1.

Population Density

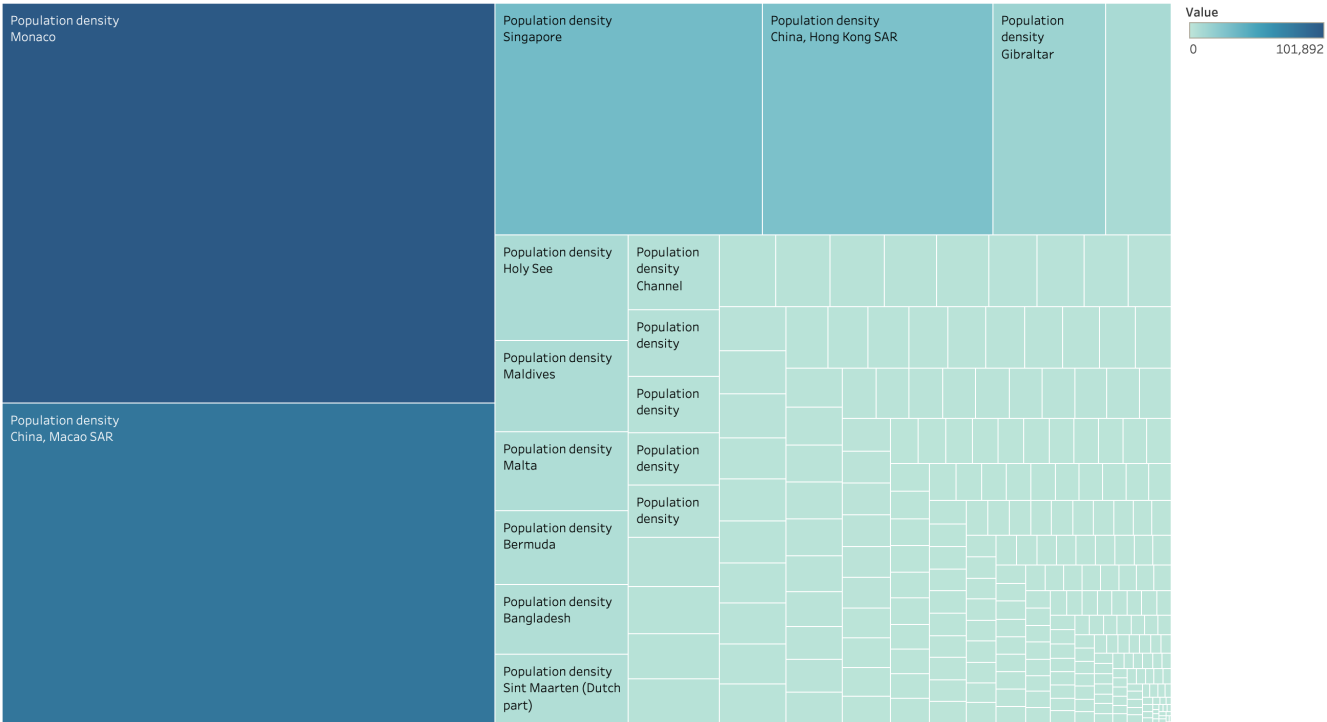


Fig. 1 Population Density by Region

I chose this image because it a nice way to visualize the differences in population density among the most and least dense countries. This visual would be best in an interactive setting, where one can hover over the smaller boxes to see what country they represent. Fig 2. uses size again to compare the surface areas of the regions.

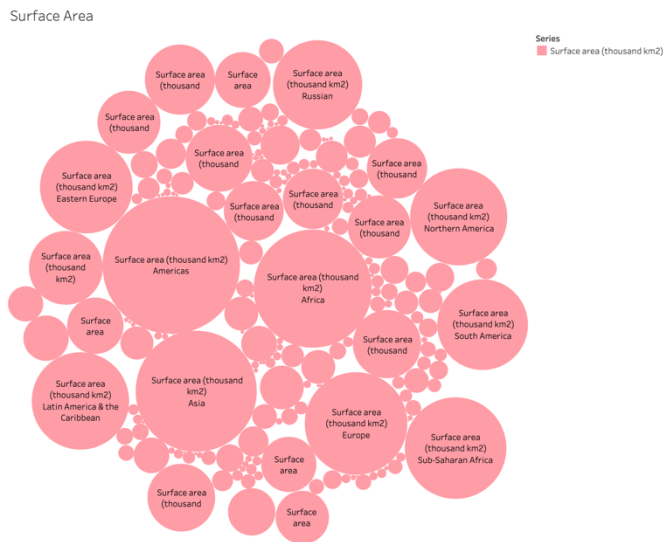


Fig. 2 Surface Area by Region

The first two visuals take advantage of comparative size to give the viewer an idea of what the values are relative to one another. This is very useful if any one value rings a bell. For example, if we know the visual size of the United States on a map, and we have the packed bubble plot, we can gain a decent understanding of the surface area of other regions by using the USA as a reference. Next we have the sex ratio (men per 100 women), shown via bar graph in Fig 3.

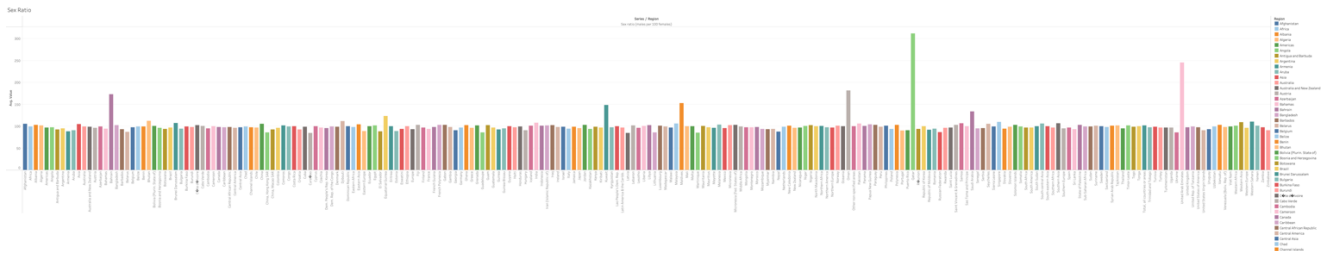


Fig. 3 Sex Ratio by Region

This figure is terrible at this scale and if the goal here was to communicate the data effectively within the paper, then I would count this as a failure. I included the figure to point out its usefulness in detecting anomalous or extraordinary values. We see in the graph above a few little spikes that indicate that the number of men relative to the number of women is extremely high in these regions. Specifically, in Bahrain, Oman, the United Arab Emirates, and Qatar. The bar graph is a great way to immediately recognize large discrepancies in height among the values. Fig 4. takes advantage of visual ratios to show the percent of the population that is under 14 years old with the percent of the population that is above 60.

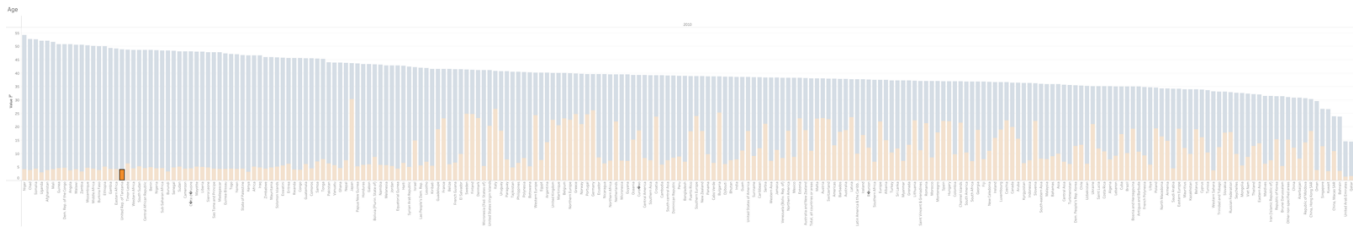


Fig. 4 Age 0-14 versus 60+ in 2010

We have all heard about the problem that Japan has with a very high elderly population relative to the young population. The tallest light orange line that we see is Japan's; 30% of the population is above 60 years old, and 13% is 14 and under, meaning that nearly half of the population is outside of working age. Interestingly, Niger has 50% of the population between 0 and 14, with only 4% over 60. I believe that this is the highest ration of young to older people in the group. It is also important to note that because this graph is of percentages, a lower sum of the blue and orange region indicates a larger 'middle aged' population, meaning that Qatar has the most middle-aged population, and the highest ratio of male to females. Finally, we have a simple line chart in Fig 5. that shows the increase in total population over time.

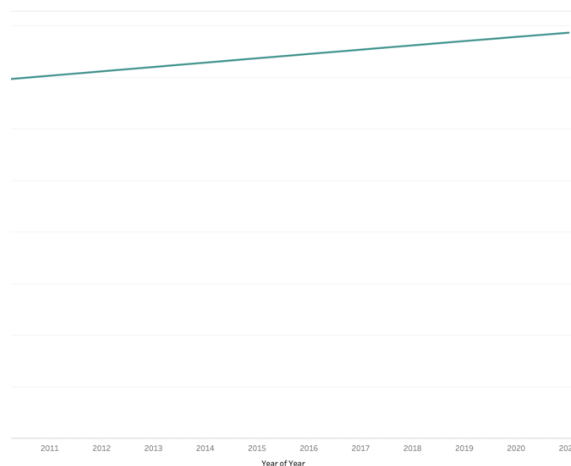


Fig. 5 Total Population from 2010 - 2021

The simple line graph shows that worldwide population is increasing. In 2010 it was just under 7 billion, and in 2021 it was just under 8 billion. We might be able to estimate from this graph that we gain about 1 billion total individuals per decade, putting us at ~10 billion by 2040.

IV. DISCUSSION

The emergent trend that I noticed within my own work was my tendency to prefer charts that were comparative. When viewing something that can only truly be understood or comprehended by comparison, it is much easier and more efficient to compare size. This was very true for population metrics and surface area. For more complex or detailed geographic data, combining color with a map would be a good way to display the data in its actual geographic region.

Working through this report also forced me to reflect on the lack of experience that I have with visualizations; I have only used very basic representations to communicate data in the past, and I do not think that I would have realized that without having taking this course.

V. CONCLUSIONS

In this report we used some very simple visualizations to gain insight into regional population and surface area data. The visualizations were effective, but not necessarily robust in methodology. I am looking forward to understanding and exploring more visualization concepts.

REFERENCES

[1] "Population, Surface Area and Density," UN Data. [Online]. Available: http://data.un.org/_Docs/SYB/CSV/SYB64_1_202110_Population,%20Surface%20Area%20and%20Density.csv.