Commodities and Yields of the World

GitHub: https://github.com/BUCKERS99/C7083-Data-Visualisation

17239400

10/03/2022

Contents

1 Introduction

The data sets included in this portfolio come from the github for R4DS and contain data collected from Our world in data. Four data sets are used throughout to create visualisations concerning crops and their countries from 1885 to 2019; however, there is limited data for years preceding 1960. The data sets contain observations for a mix of over 200 countries and continents. Each data set is explained below in more detail.

1.1 Key crop yields

A dataset recording the yield of certain commodities with a range of year from 1961 to 2018 and over 13000 observations. 14 variables are described below:

Variable	Class	Description
Entity	character	Country or Region Name
Code	character	Country Code (NA is for regions)
Year	double	Year
Wheat	double	Wheat Yield
Rice	double	Rice Yield
Maize	double	Maize Yield
Soybeans	double	Soybean Yield
Potatoes	double	Potato Yield
Beans	double	Beans Yield
Peas	double	Peas Yield
Cassava	double	Cassava (yuca) Yield
Barley	double	Barley Yield
Cocoa	double	Cocoa Yield
Bananas	double	Banana Yield

1.2 Arable land

A dataset recording the area of land used for arable agriculture with a range of year from 1961 to 2014 and over 11000 observations. 4 variables are described below:

Variable	Class	Description
Entity	Character	Country or Region Name
Code	Character	Country Code (NA is for regions)
Year	Double	Year
land requirement	Double	Arable land normalised to 1961

1.3 Land use

A dataset recording the change in land use for arable agriculture with a range of year from 10000 BCE to 2019 and over 49000 observations. This data contains lots of missing values due to population data having been added from Gapminder. 6 variables are described below:

Variable	Class	Description
Entity	Character	Country or Region Name
Code	Character	Country Code (NA is for regions)
Year	Double	Year
Cereal yield index	Double	Cereal Yield Index
land change	Double	Change in land area use for cereal production relative to 1961
Total population (Gapminder)	Double	Total Population (Gapminder)

1.4 Tractors

A dataset recording yield of crops and the number of tractors per 100 square kilometer of arable land with a range of year from 10000 BCE to 2019 AD and over 49000 observations. This data contains lots of missing values due to population data having been added from Gapminder. 6 variables are described below:

Variable	Class	Description
Entity	Character	Country or Region Name
Code	Character	Country Code (NA is for regions)
Year	Double	Year
tractor	Double	Number of tractors per 100 sq km of arable land
Yield	Double	Cereal Yield in kg per hectare
Total population (Gapminder)	Double	Total Population (Gapminder)

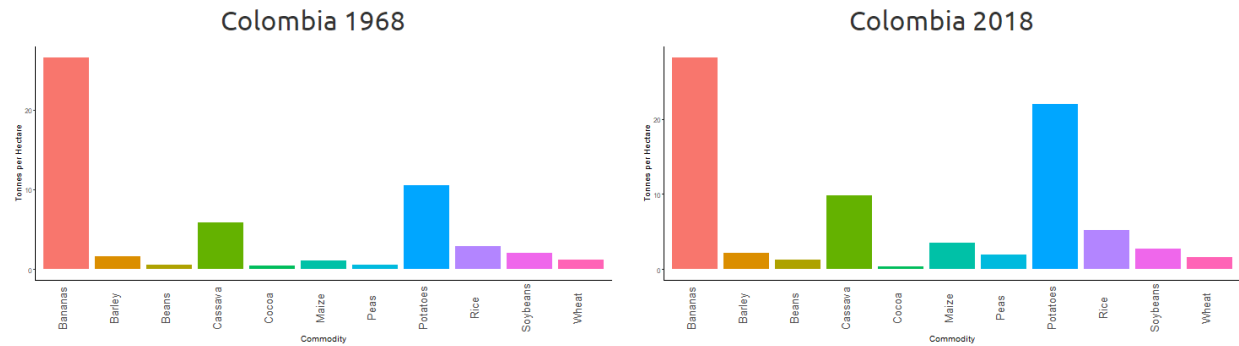## 2 Comparitive app between countries and years

Figure 1 - Commodities and their yields for each country.

Figure 1 was created with the intention of the user comparing countries against one another to gain information on the types of commodity that the country produce. Creating the shiny dashboard was complicated and there were features that I would have liked to include, such as data labels on the bars, that would have enhanced the visualisations usability. The drop down boxes require the year and country to be typed as there were too many countries to display them all. The bar plots used show a clear distinction between commodities along with the colour choices.
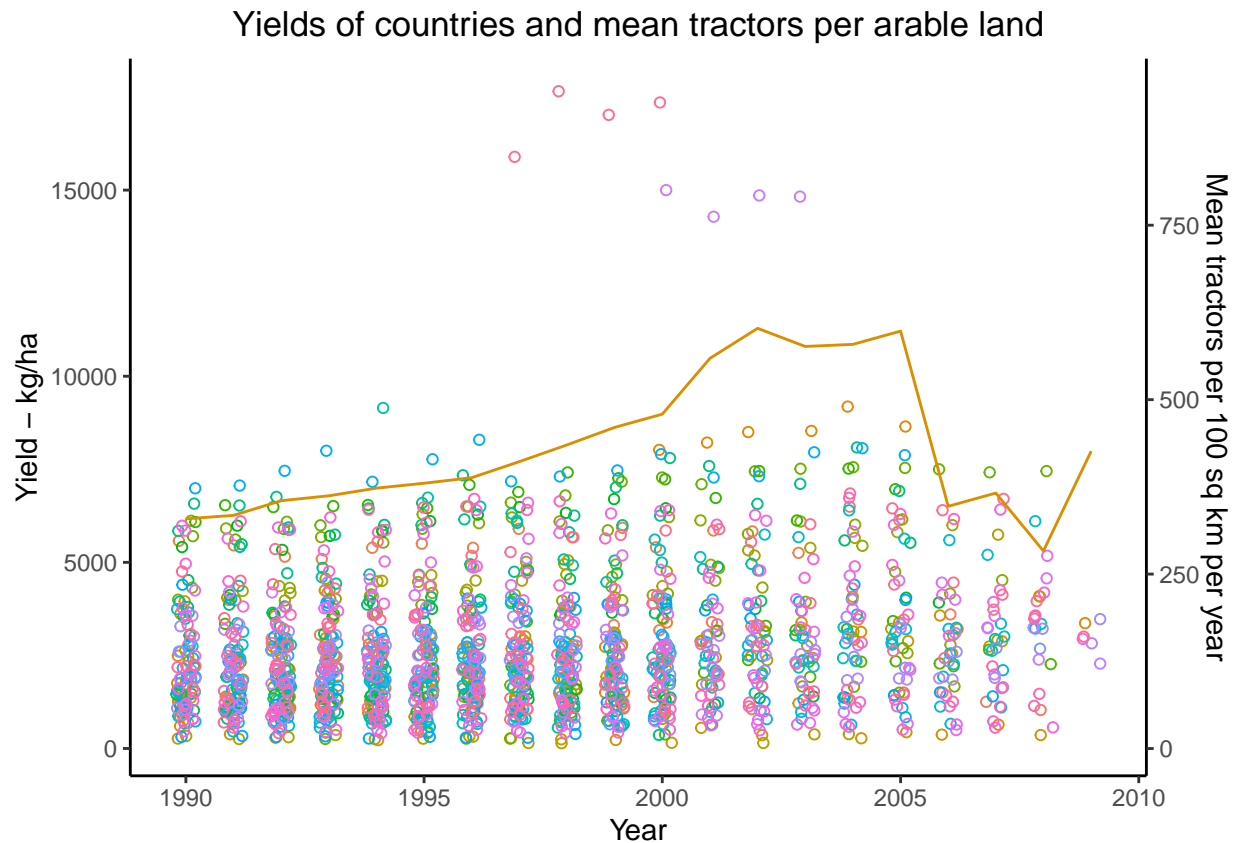
3 Cereal yield of countries from 1990 to 2009

Yields of countries and mean tractors per arable land

Figure 2 - Cereal yield of countries and the average number of tractors per 100 sq kn of arable land.

Figure 2 shows the yield of each country over the space of 10 years. A strip chart was chosen as there was a lot of data to represent in a small area with the use of the jitter function allowing it to be viewed clearly. The graph shows that there is little change in the range of yield other than a few outliers. These outliers are the UAE and St Vincent and the Grenadines, for the former the yield of certain commodities were investigated over these years to see if conclusions could be made on the reason for the increase in yield.
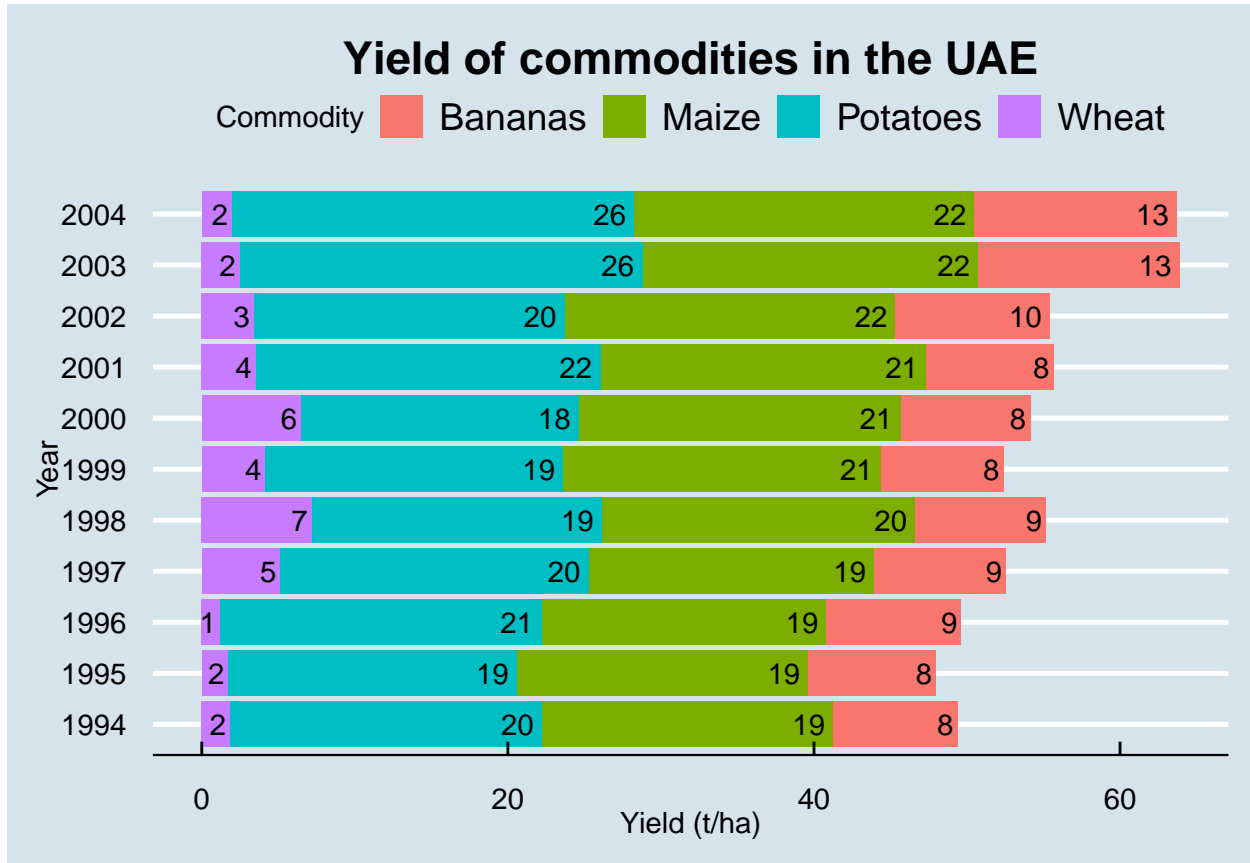
4 UAE Commodities

Figure 3 - UAE Commodities and Yields from 1994 to 2004.

From the previous Figure it was clear that between 1997 and 2000 the UAE recorded some extraordinary yields. Figure 3 was therefore created to explain this dramatic increase in yield by breaking down yield by each commodity. There was scope to include fertiliser application figures but there was no data for the UAE in this regard. The visualisation made in Figure 3 does not shed light upon the dramatic increase in yield for the UAE in 1997, it does show that there was a small increase from 1 to 5 tonnes per hectare of wheat. Due to proportion being hard for the eye to calculate the decision was taken to include data labels for ease of interpretation of this graph.
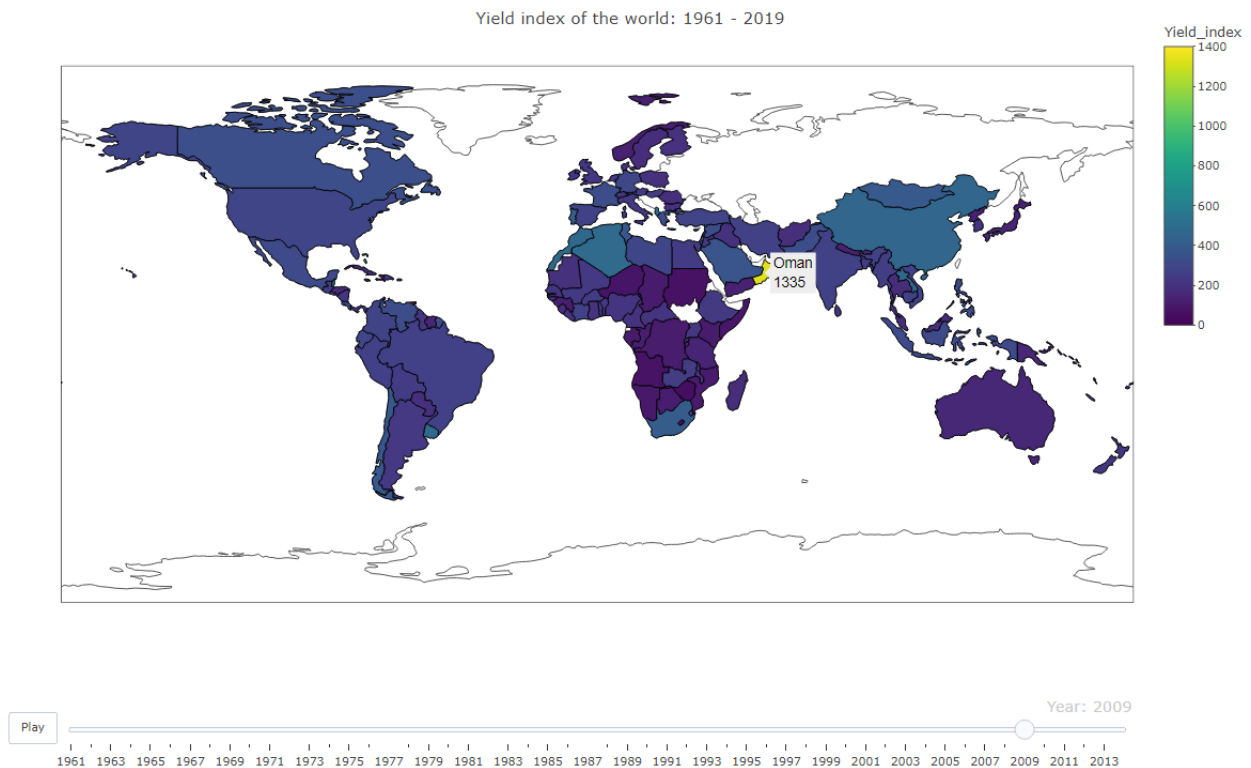
5 Cereal Yield Index of the World

Figure 4 - Interactive map of world yield index.

Figure 4 details the cereal yield index for all countries that had supplied data from 1961 to 2019. Cereal yield index is the average yield of all cereals grown that year, interestingly the country with the highest yield index is Oman with 1335. This animated map is also interactive; the user is able to zoom in to areas of interest and when the mouse is held over the country it will show information. The scroll bar at the bottom also allows for a specific year to be chosen for comparison, or playing across the range of years. A set scale was used to see the distinction between early and later years with colors from the viridis package, which are colourblind friendly.
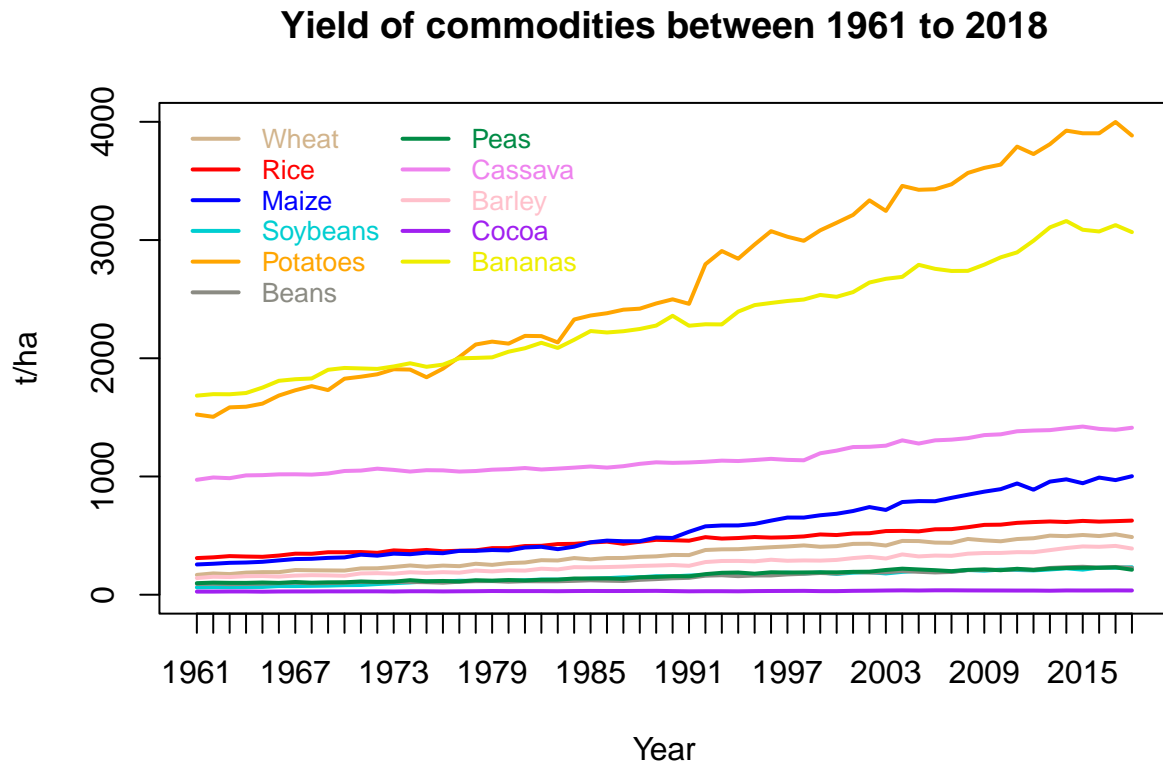
6   Commodities

Yield of commodities between 1961 to 2018

Figure 5 - Total commodity yield between 1961 and 2018.

Figure 5 is a base R visualisation constructed to detail the yield of commodities over a span of 57 years. In general it shows that the yield has increased for all commodities other than cocoa. The yield for cocoa should not be seen as reliable however, as the countries that grow this commodity do not have accurate recording methods and yields are not necessarily correct. Colour was used to distinguish between commodities and the cumulative total is calculated and plotted in order to easily see trends over a time series.
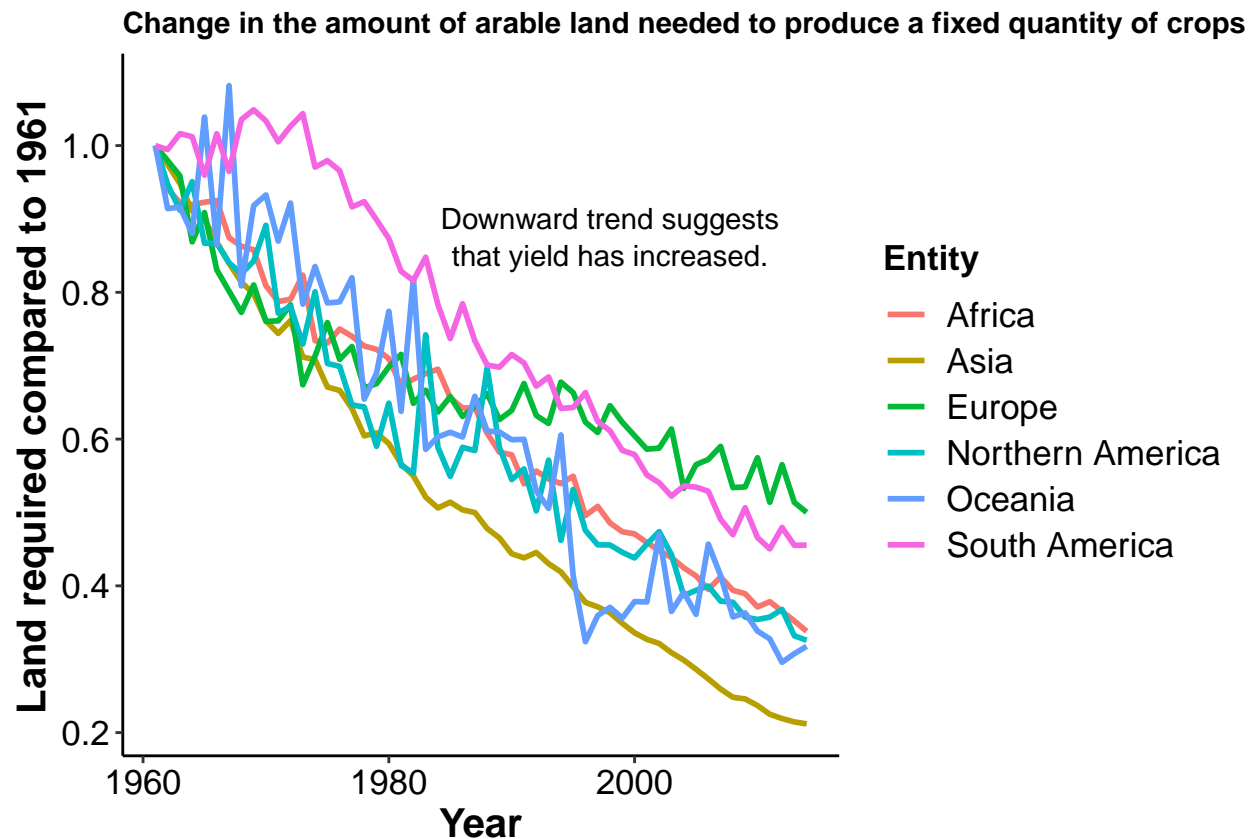
7 Land usage

Figure 6 - Change in land required to produce a fixed amount of crop (relative to 1961).

Figure 6 suggests in the graphs annotation that the downward trend of the amount of land needed to produce a fixed amount of crop could be attributed to an increase in yield for each commodity. This increase in yield is shown in figure 5 so the visualisations relate to one another well. The graph is plotted as a comparison to the amount of land needed in 1961 to grow a set amount of produce. the graph was constructed using ggplot and is intended to just show an overall trend rather than data values to be read from it.

8 Critique of visualisations

8.1 Bad visualisation

Appendix 1 attempts to show the percentage of the worlds' debt in a choropleth pie-chart. The annotations found throughout the visualisation are informative and helpful with a clear definition in the subtitle of the overall figure. This relates to the principle of detailed captions outlined by Midway (2020), however in the same principle it is outlined that the visuals should be simple. The overall image of the pie chart is challenging to read due to the diverse geometric shapes used for each country. For the reader it creates uncertainty to whether these shapes contain meaning or they are for visual effect. Small annotations that overlap the boundary boxes prevent the reader from being able to gain understanding of what the data relates to. Within these annotations contains the percentage figure which does help the reader with understanding: but requires a calculation to work out the actual information. When working with large figures, such as $69 trillion, it is asking the audience too much to work out what 2.7% of $69 trillion is. The colours chosen for the choropleth scale are difficult to look at and understand their value to the overall graph. This visualisation could be improved through creating an interactive version where the top 10 countries are initially displayed with the option of selecting over countries for comparison. The choice of a pie-chart could also be changes to a bar-graph or map style.

8.2 Good visualisation

Appendix 2 shows a map type visualisation describing the number of electric car charging stations per 100,000 cars in the United states. The creator has chosen a simple and distinctive colour scheme which allows the reader to differentiate between the high and low values. The states of America also change in size depending on the amount of electric charging stations within the state. The subtitle gives a clear definition of what the graph is about and how to interpret the overall figure. The simplicity of this visualisation allows it to be read and understood without having to perform mental maths, as described in Appendix 1. It is worth highlighting that due to the mapping style the distorted shapes of the states could lead to potential confusion when trying to determine which state is which. The use of the map style also does not make it easily readable to people with little or no American geographic knowledge, and the lack of actual values make the audience guess using the choropleth scale. This visualisation could be improved by including a side-by-side bar chart next with the actual values and state names to solve the potential lack of knowledge in the audience.
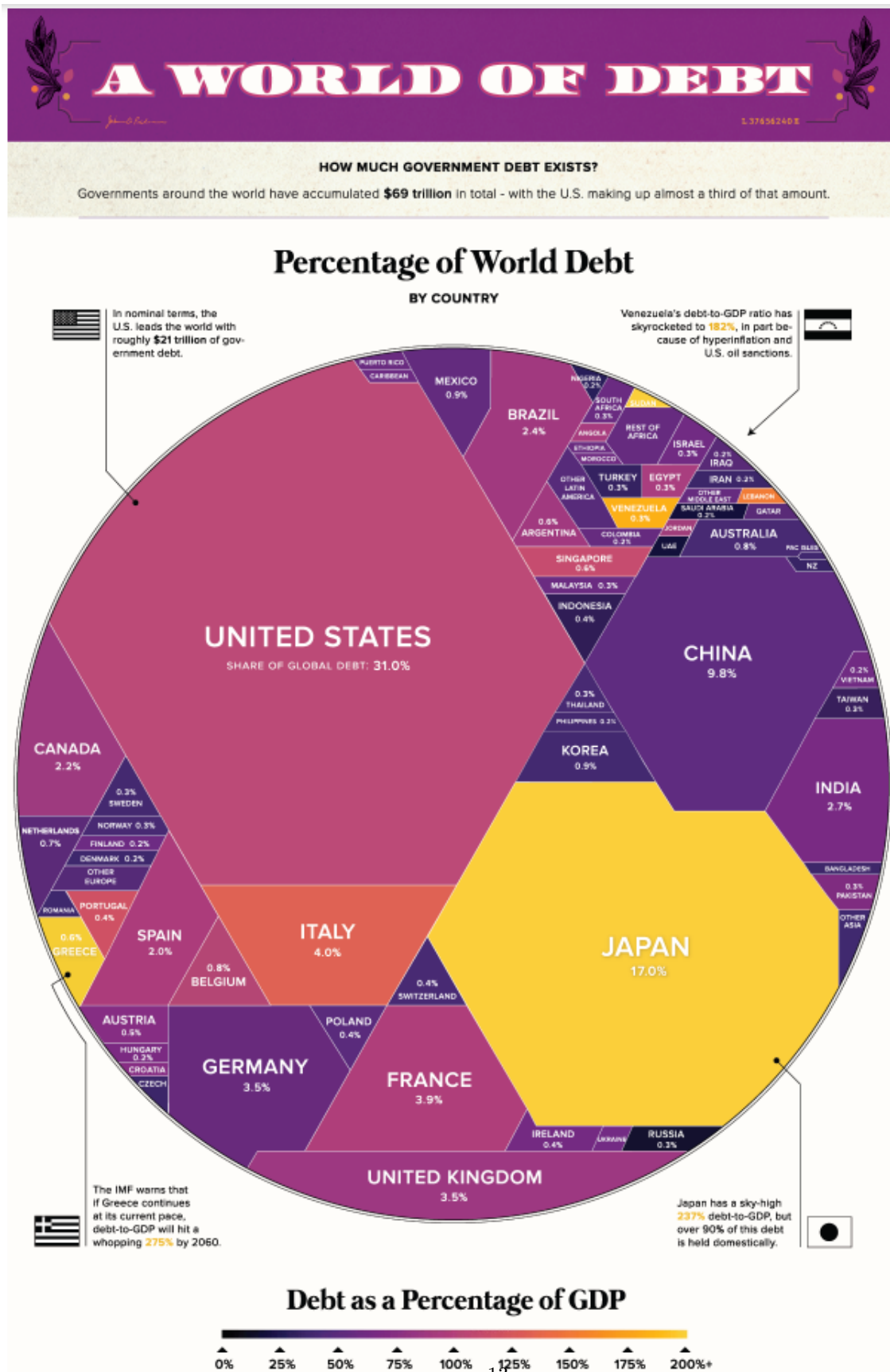
9 References

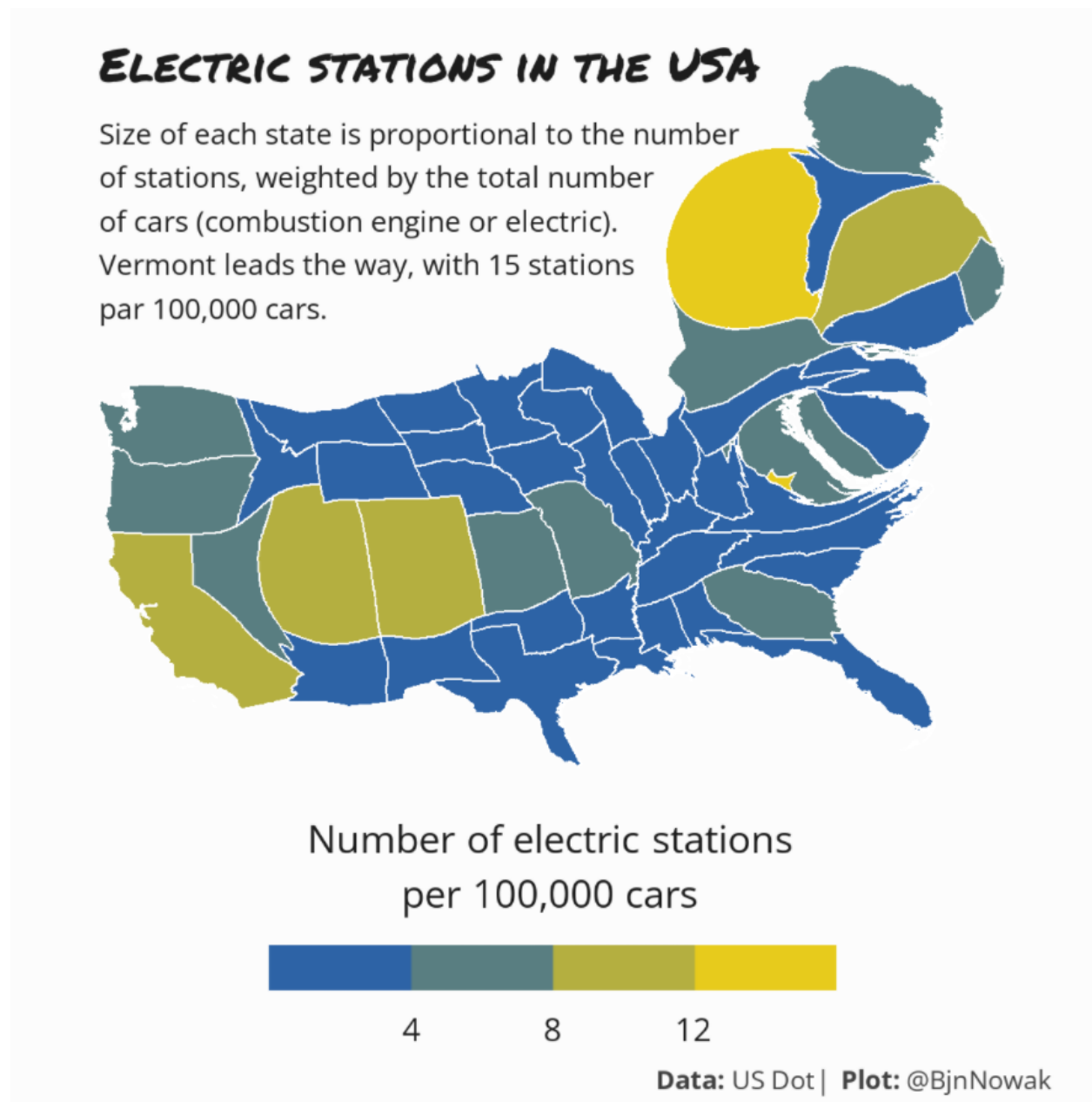Midway, S.R., 2020. Principles of effective data visualization. *Patterns*, 1(9), p.100141.

Twitter. 2022. *Benjamin Nowak* [Online] Available from:https://twitter.com/BjnNowak/status/ 1498658998341537796 [Accessed on 01/03/2022].

VisualCapitalist. 2019. *$60 Trillion of World Debt in One Infographic.* [Online]. Available from:https: //www.visualcapitalist.com/69-trillion-of-world-debt-in-one-infographic/ [Accessed on 01/03/2022].

10 Appendices

Appendix 1 - Example of a bad visualisation Source:(VisualCapitalist, 2019)

ELECTRIC STATIONS IN THE USA

Size of each state is proportional to the number
of stations, weighted by the total number
of cars (combustion engine or electric).
Vermont leads the way, with 15 stations
par 100,000 cars.

Number of electric stations
per 100,000 cars

4   8   12

Data: US Dot | **Plot:** @BjnNowak

Appendix 2 - Example of a good visualisation Source:(Twitter, 2021)