Name: Ayano Yamamoto

Module: Data Visualisation

Table of Contents

Selected Datasets

Data from eight files below were imported into a Python file, cleaned, and merged into two CSV files, `producers.csv` and `consumers.csv`.

- Historical Data on the Global Coffee Trade from International Coffee Organization

  Five Excel files below were used in this analysis. Since the original files did not include ISO 3166-1 alpha-3 codes, they were assigned using a function to fuzzy-search country names in `pycountry` and return alpha-3 codes. Country names that could not be matched using the function were assigned the alpha-3 codes manually.

  - Exports - Calendar Year
  - Imports - Calendar Year
  - Prices to Growers - Annual Averages
  - Retail Prices - Annual Averages
  - Disappearance (consumption) - End of Year

  "Prices to Growers - Annual Averages" includes prices per type of coffee beans per country per year (Figure 1). If a country had prices for multiple types of coffee beans, their average was calculated per year. For all five files, the rows that did not have any yearly country data were dropped, and their values were converted into a list for each row and then exploded vertically so that one row was assigned to one year per country (Figure 2).

	Calendar years	1990	1991	1992
0	NaN	NaN	NaN	NaN
1	Colombian Milds	NaN	NaN	NaN
2	Colombia	69.5230	67.1251	54.5709
3	Kenya	60.5721	50.6778	28.9495

Figure 1: Original data frame structure

	alpha3	country	price	year
0	AGO	Angola	85.6569	1990
0	AGO	Angola	91.1153	1991
0	AGO	Angola	44.9183	1992
0	AGO	Angola	11.8708	1993
0	AGO	Angola	NaN	1994

Figure 2: Dataframe after cleaning and restructuring

- Coffee Bean Production from Our World in Data

  Rows without ISO 3166-1 alpha-3 codes were dropped since they were referring to continents or groups of countries. Years were filtered only to include data from 1990 to 2019.

- Population, total and GDP (current US$) from World Bank Data

  The original datasets were structured with one year of data per country per column (Figure 3). The columns were filtered to only include data from 1990 to 2019, and their values were converted into a list for each row and then exploded vertically so that one row was assigned to one year per country (Figure 4).

	Country Name	Country Code	1990	1991	1992
0	Aruba	ABW	65712.0	67864.0	70192.0
1	Africa Eastern and Southern	AFE	309890664.0	318544083.0	326933522.0
2	Afghanistan	AFG	10694796.0	10745167.0	12057433.0

Figure 3: Original data frame structure

	country	alpha3	population	year
0	Aruba	ABW	65712.0	1990
0	Aruba	ABW	67864.0	1991
0	Aruba	ABW	70192.0	1992
0	Aruba	ABW	72360.0	1993
0	Aruba	ABW	74710.0	1994

Figure 4: Dataframe after restructuring

The final CSV files imported into Tableau are as below.

Features in `producers.csv`

- alpha3: ISO 3166-1 alpha-3 codes from “Exports - Calendar Year”
- country: Country names
- year: Years from 1990 to 2019
- production: Values from Coffee Bean Production
- export: Values from “Exports - Calendar Year”
- price: Values from “Prices to Growers - Annual Averages”
- population: Values from Population, total
- gdp: Values from GDP (current US$)

Features in `consumers.csv`

- alpha3: ISO 3166-1 alpha-3 codes from “Imports - Calendar Year”
- country: Country names
- year: Years from 1990 to 2019
- consumption: Values from “Disappearance (consumption) - End of Year”
- import: Values from “Imports - Calendar Year”
- retail: Values from “Retail Prices - Annual Averages”
- population: Values from Population, total
- gdp: Values from GDP (current US$)

Intended Audience

The audience for exploration: A writer for a web magazine about the coffee industry

Images: Flaticon.com

> As a writer for a coffee blog, I would like to analyze the historical data on the global coffee trade so that I can write about interesting insights for my readers.

A web magazine features an overview of the coffee market as well as the most recent global coffee news and leader profiles. They are publishing an article based on historical data on the global coffee trade. Questions to be answered are:

- What is the overall trend of coffee imports and exports from 1990 to 2019?
- Which are the top 5 countries in coffee production and consumption? How have they changed in the last three decades?
- What is the overall trend of the prices paid to the growers? What is the overall trend of the retail prices of coffee?

The audience for insights: A reader of the web magazine and a cafe owner

Images: Flaticon.com

> As a reader of a coffee blog, I would like to read about key insights from historical data on the global coffee trade so that I can take this information into account for my business.

A large portion of the readers of this web magazine are people who work in the coffee industry, such as cafe owners, roasters, bean sourcers, and baristas. They are looking for interesting industry facts and knowledge that will be valuable for their businesses. Questions to be answered are:

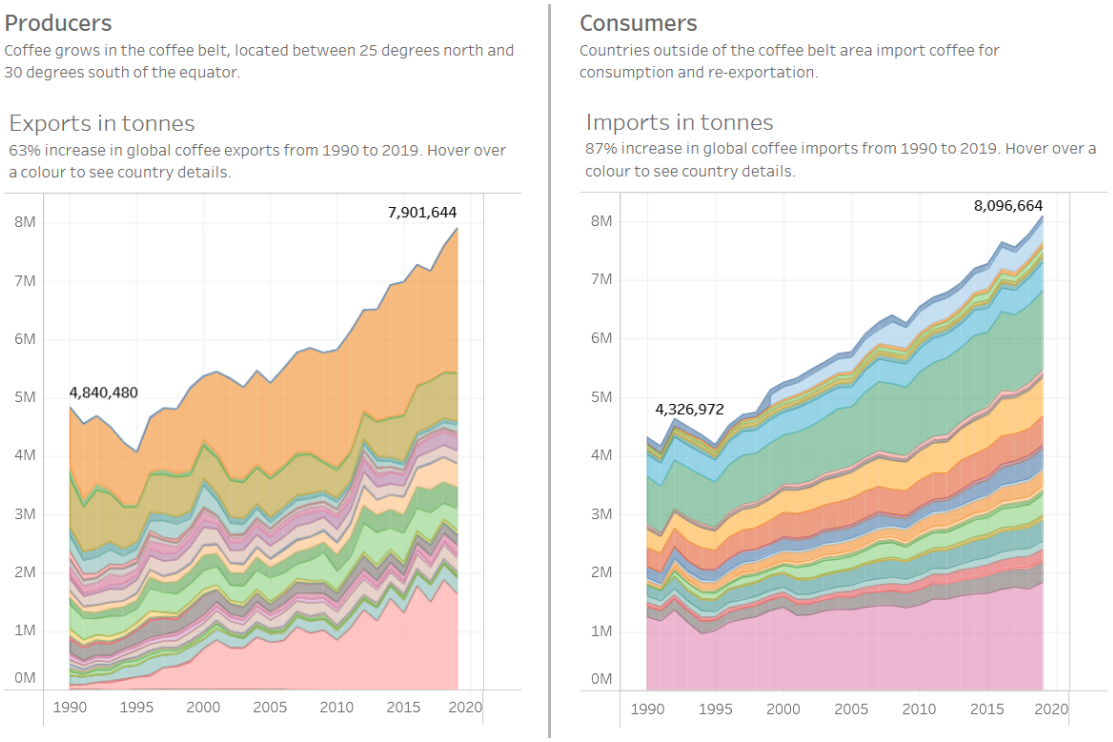
- How has the per capita consumption changed in Europe over the last three decades?
- How do the production countries view coffee as part of their economy? Does coffee make up significant portions of their GDPs, and have there been any changes since 1990?
- Is there any trend in the retail prices of coffee that stands out?

Data Exploration

Link to Tableau Public:
https://public.tableau.com/app/profile/ayano.yamamoto/viz/Assignment1_16778865208690/Exploration

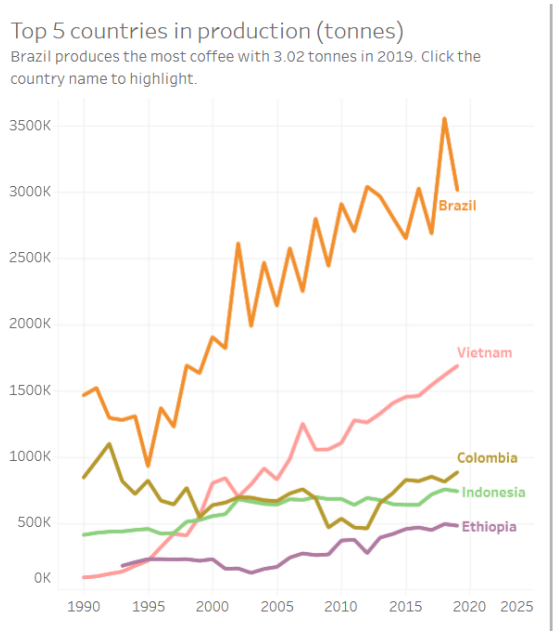
- What is the overall trend of coffee imports and exports from 1990 to 2019?

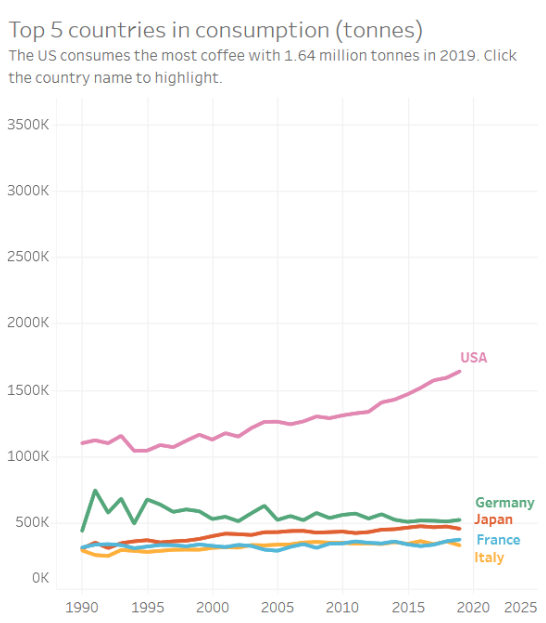
  Two area charts show export and import trends over time, as well as details from each country in different colours. Export and import data were originally recorded in the unit of thousand 60kg bags, which were transformed into tonnes in calculated fields.



- Which are the top 5 countries in coffee production and consumption? How have they changed in the last three decades?

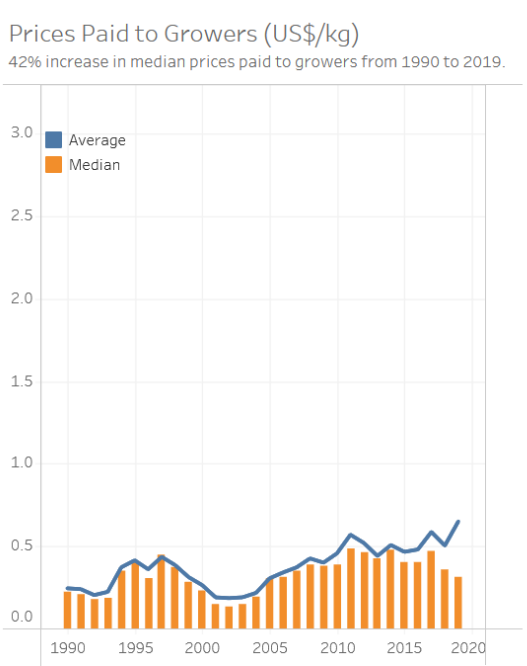
  Two line graphs show the top 5 countries in coffee production and consumption, and how they performed over time. Consumptions were originally recorded in the unit of thousand 60kg bags, which were transformed into tonnes in calculated fields. Country names are set to match the line colours for clarity, especially when there are lines close to each other as seen in the consumption graph.

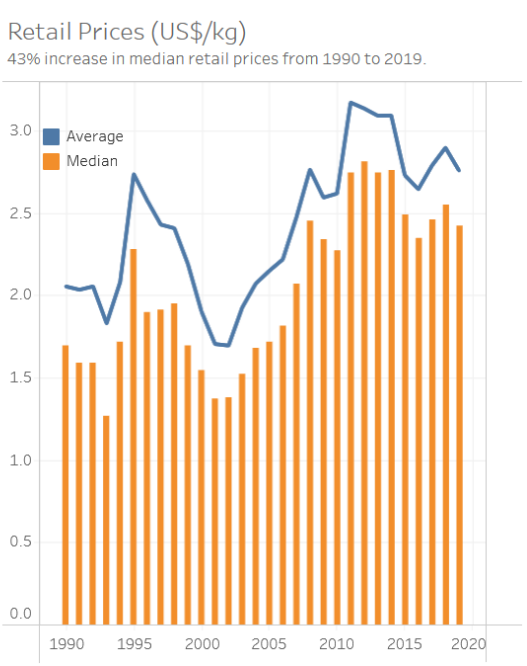




- What is the overall trend of the prices paid to the growers? What is the overall trend of the retail prices of coffee?

  Two graphs show the average and median prices paid to the growers and the retail prices. The prices paid to the growers were originally recorded in US cents/lb, and the retail prices were recorded in US$/lb. They were both converted to US$/kg in calculated fields.





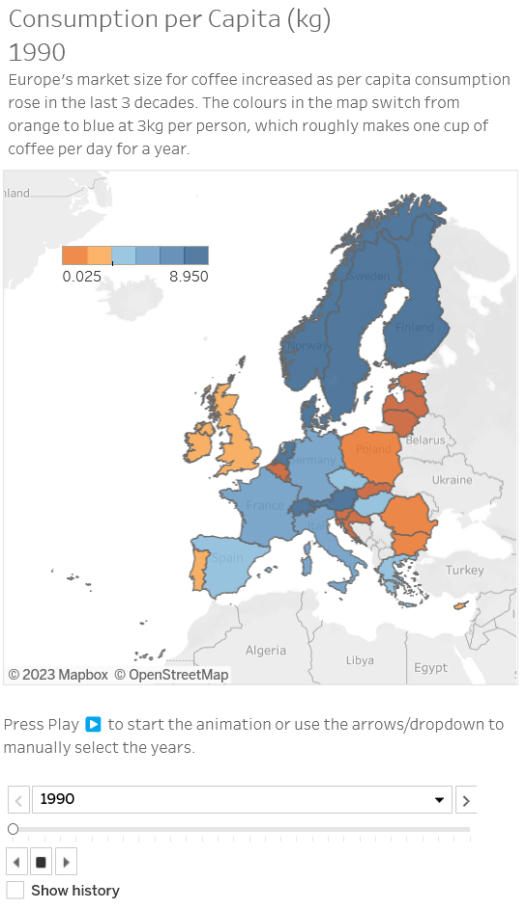
3 Insights from the Data

Link to Tableau Public:
https://public.tableau.com/app/profile/ayano.yamamoto/viz/Assignment1_16778865208690/Insights

- How has the per capita consumption changed in Europe over the last three decades?

  The map shows the per capita coffee consumption of European countries over time. The values are calculated using data from the World Bank on population and consumption in kilograms.

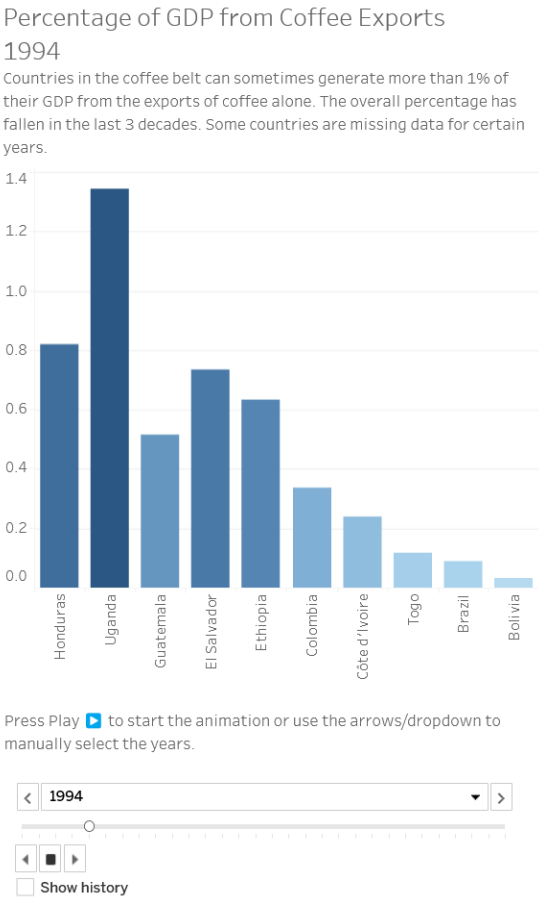
  The user can interact with the visualisation using the year slider, and see the overall trend of coffee consumption per capita increasing from 1990 to 2019. The colour scales are designed to show countries in orange if their per capita consumptions are less than 3kg per year, which approximately makes one cup of coffee every day for a year. It is also apparent that Nordic countries such as Norway and Finland consumed high amounts of coffee per capita throughout the three decades.



- How do the production countries view coffee as part of their economy? Does coffee make up significant portions of their GDPs, and have there been any changes since 1990?

  The bar chart shows the percentage of GDPs from coffee exports for 10 countries in the coffee belt. The values are calculated using data from the World Bank on GDP, exports in kilograms, and prices paid to the growers in US$/kg.

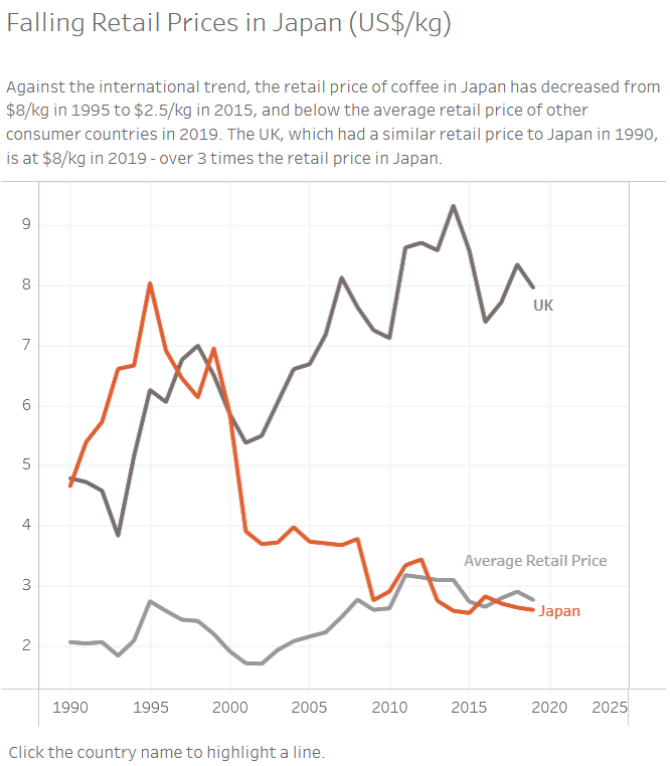
  The user can interact with the visualisation using the year slider, and see the overall percentages of GDP from coffee exports decreasing from 1990 to 2019. The colours are designed to accentuate higher values. In particular years, coffee exports alone can make up over 1% of the GDPs of countries such as Honduras and Uganda.



- Is there any trend in the retail prices of coffee that stands out?

  The line graph shows the retail prices of Japan and the UK over time compared to the average retail price of consumer countries. The average retail price was created as a calculated field using `FIXED [Year]`.

  Country names are set to match the line colours for clarity. The users can click a country name (or “Average Retail Price”) to highlight a specific line. Compared to the average, it is apparent that the retail price of coffee in Japan has decreased over the three decades. Values for the UK are also shown for comparison since the two countries had similar retail prices in 1990. By 2019 the price in the UK has become over 3 times the price in Japan.



Previous Iterations

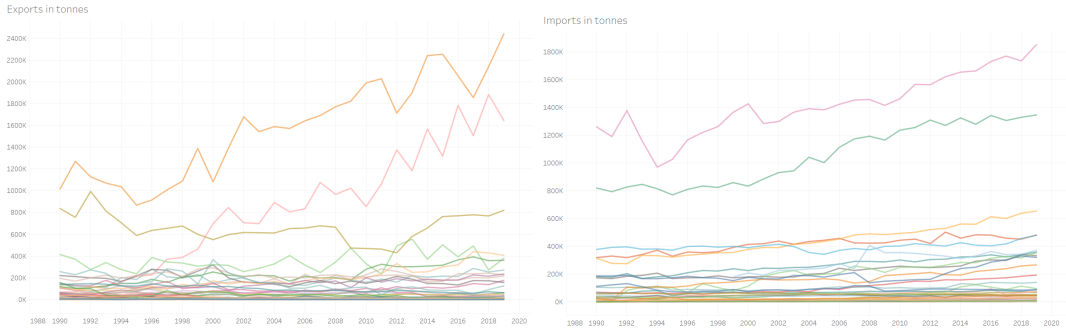
- Imports and exports: from line graphs to area charts

  The two visualisations showing imports and exports were originally line graphs with colours for each country. The issues in these graphs were:

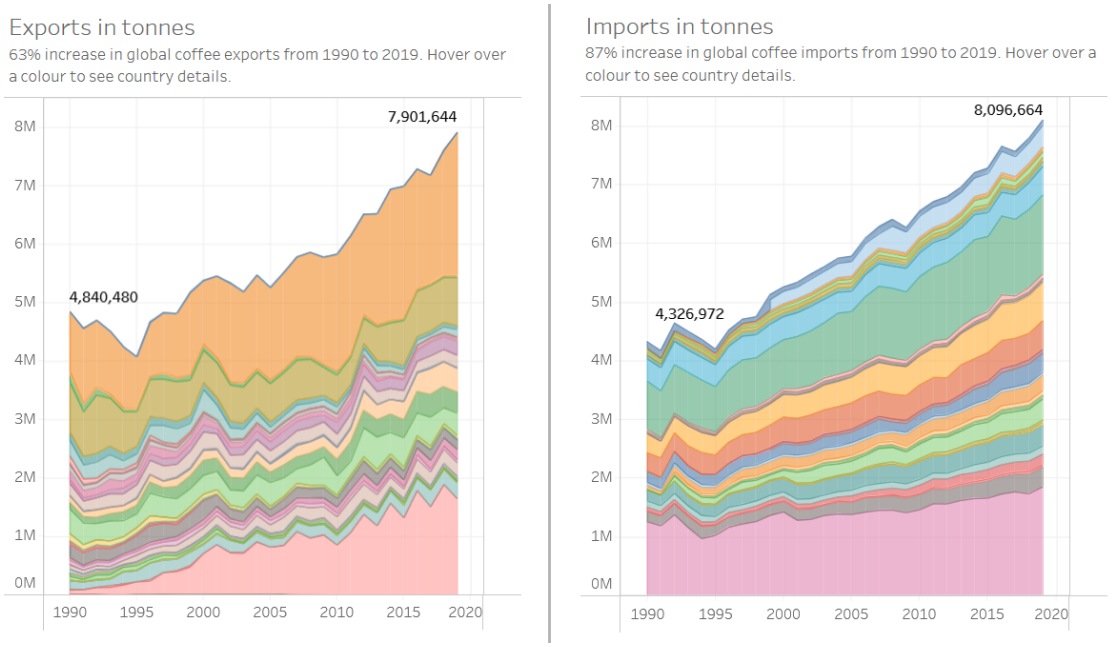
  - No context on the overall size of the trade
  - Different ranges for the y-axis
  - Hard to read in areas where many countries had similar values

  They were converted into area charts with matching axis ranges, with labels showing the total size of the trades at the beginning and end of the timeline to represent the upward trends.

  Before



  After



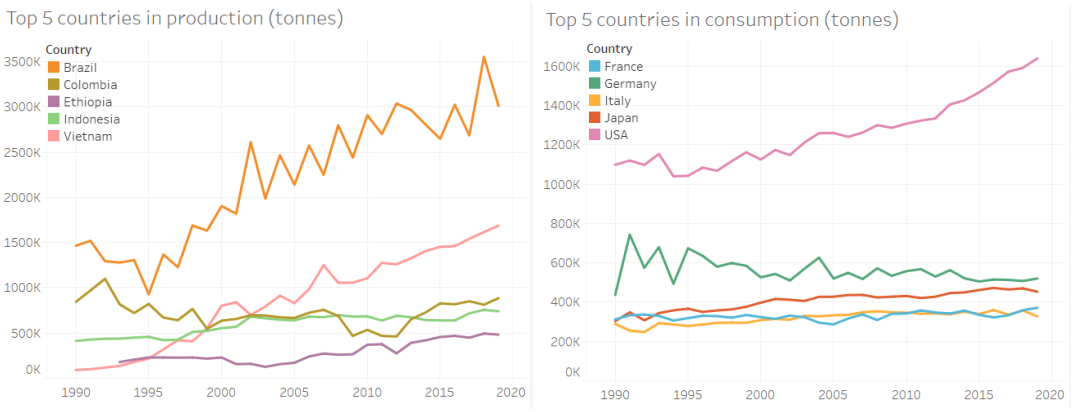
- Production and consumption: legends as labels

  The line graphs showing the top 5 countries in production and consumption initially had their legends in the top left corners. The issues in these graphs were:

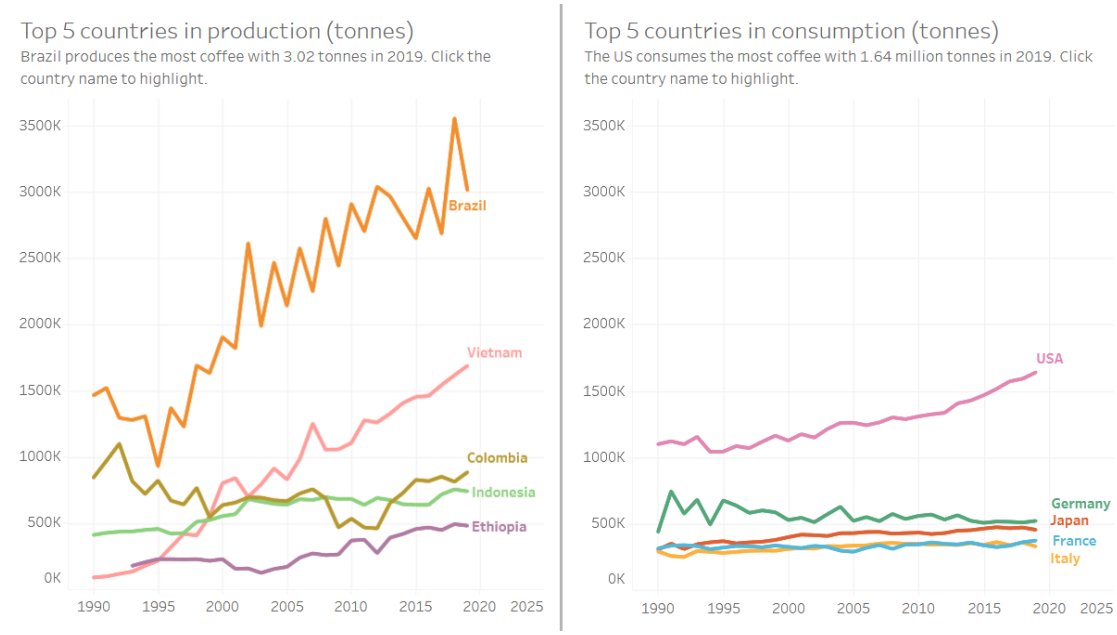
  - Requires the readers' effort to match the legend and line colours
  - Different ranges for the y-axis

  The legends were moved to labels set to match the line colours for clarity. The axis ranges were also standardised for comparison.

Before



After



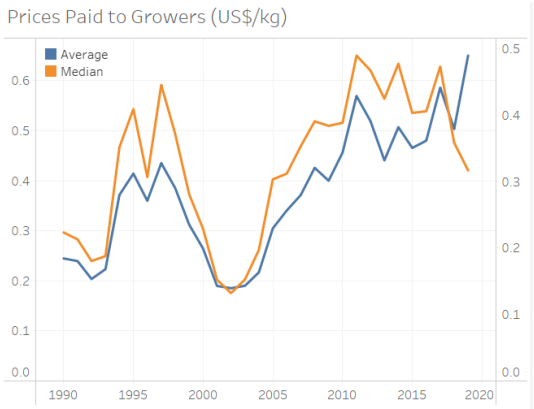
- Prices: combining multiple graph types

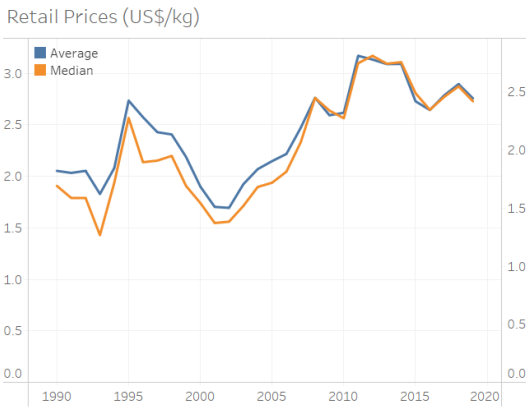
  Prices paid to growers and retail prices had their averages and medians as line graphs with a different axis for each line. The issues in these graphs were:

  - Hard to differentiate between average and median
  - Different ranges for the y-axis

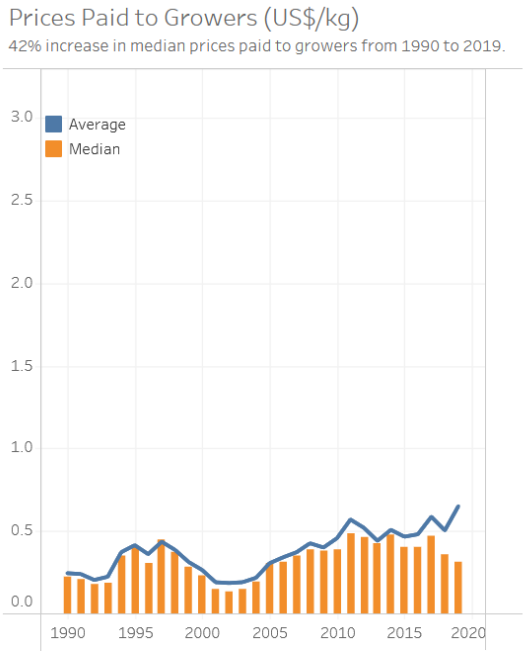
  Medians were transformed into bar graphs, with the axis ranges standardised for comparison.

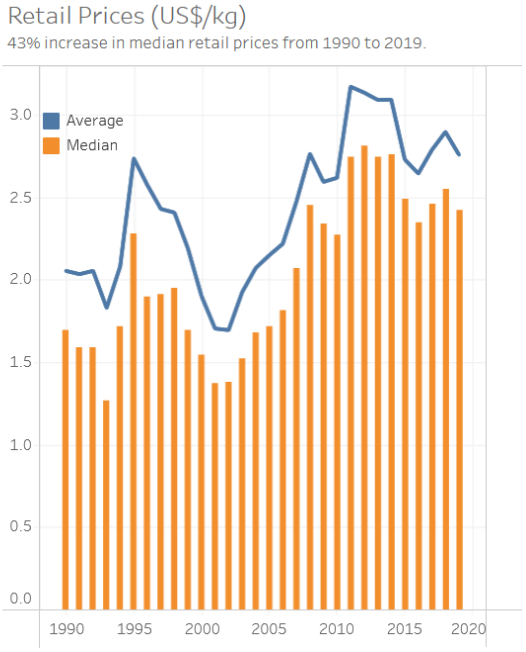
  Before





  After





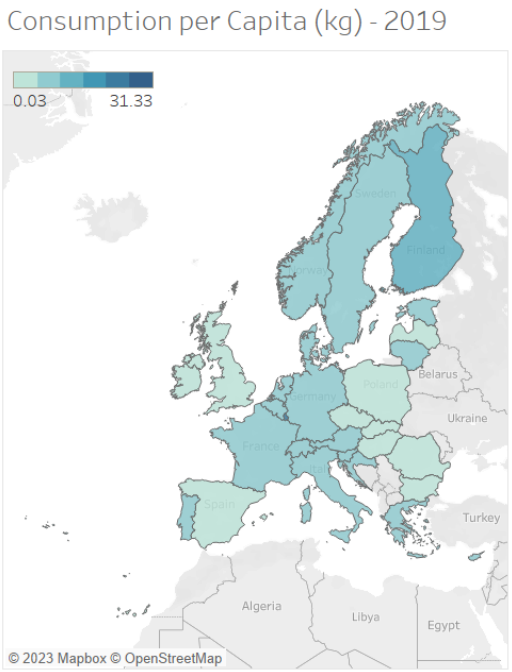
- Consumption per capita map: colour adjustments

  Since the consumption per capita data had a high outlier (Luxembourg), the colour gradient in the map looked very similar for the rest of the countries. The issues in this visualisation were:

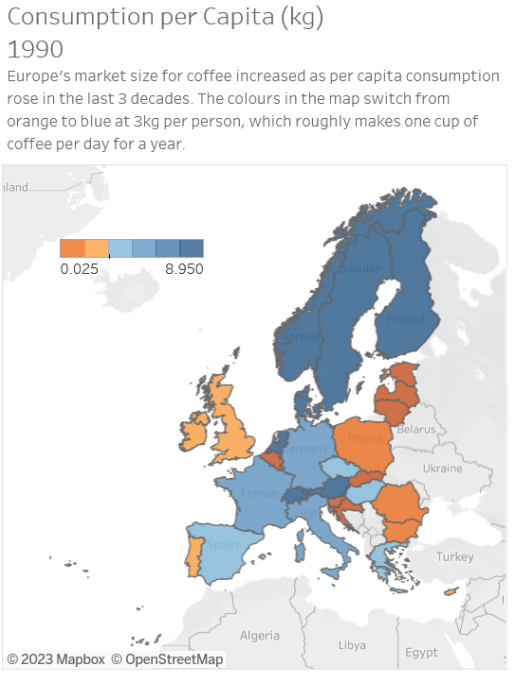
  - Hard to see the difference between countries
  - Hard to observe the change over time when animated

  It was adjusted to have two colours in the gradient, with 3kg set as the centre point so it visualises a point where the consumption per capita goes reaches one cup per day if a country's colour changes from orange to blue.

Before



After



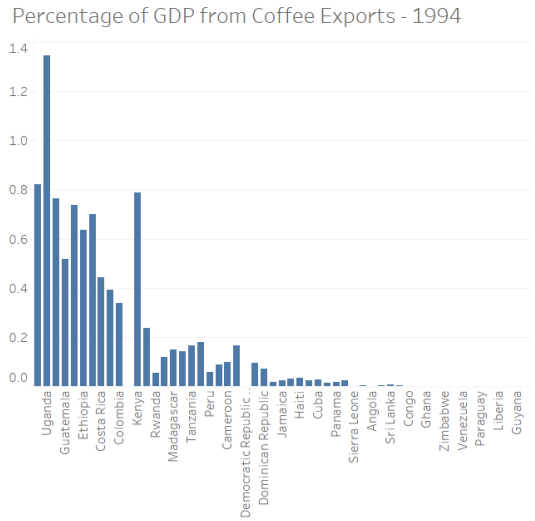
- Percentage of GDP from coffee exports: filtered countries

  Initially there were 55 countries included in the dataset for coffee producers. Their values were plotted on a bar chart in the same colour. The issues in this visualisation were:

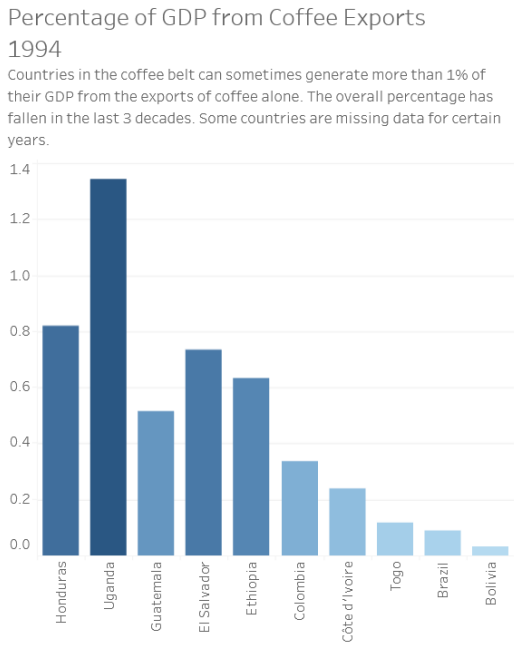
  - Some of the country names are not visible
  - Not clear which values are significant

  The graph was amended to have a filtered selection of countries, with colour gradient to show any values over 1% in the darkest hue.

Before



After



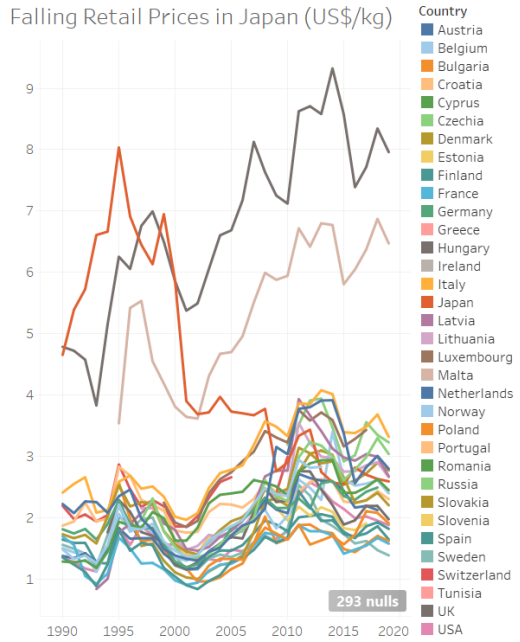
- Retail prices in Japan: summarising values as an average in dual axis

  The line graph showing the retail prices over time originally included values for 34 countries. Although the decline in retail prices in Japan stands out compared to the trend from other countries, it was not well highlighted. The issues in this visualisation were:

  - Hard to see the key insight when there are many other information visible
  - Requires the readers' effort to match the legend and line colours

  An average of retail prices from all consumer countries were calculated to represent the trend to compare with the values from Japan. The UK was also introduced since the two countries had similar retail prices in 1990. The legends were moved to labels set to match the line colours for clarity.

Before



After

References

- International Coffee Organization. (n.d.). *Historical Data on the Global Coffee Trade*. International Coffee Organization. Retrieved March 4, 2023, from https://www.ico.org/new_historical.asp
- Our World in Data. (n.d.). *Coffee Bean Production*. Our World in Data. Retrieved March 4, 2023, from https://ourworldindata.org/grapher/coffee-bean-production?time=latest
- World Bank Group. (n.d.). *GDP (current US$)*. World Bank Data. Retrieved March 4, 2023, from https://data.worldbank.org/indicator/NY.GDP.MKTP.CD
- World Bank Group. (n.d.). *Population, total*. World Bank Data. Retrieved March 4, 2023, from https://data.worldbank.org/indicator/SP.POP.TOTL