**TASK**

# Data Visualisation Approach and Techniques

# JULIANA RIBEIRO

# (C6 Data Science Bootcamp)

# Practical Task 1

Examine the graphs below and use some research and background knowledge to make conclusions based on your observations.

1. The following bar graph shows the gender wage gap in 26 countries based on data collected by the **OECD**. The gender wage gap is calculated by finding the difference between male and female median wages and dividing it by male median wages. It is represented as a percentage in this graph.
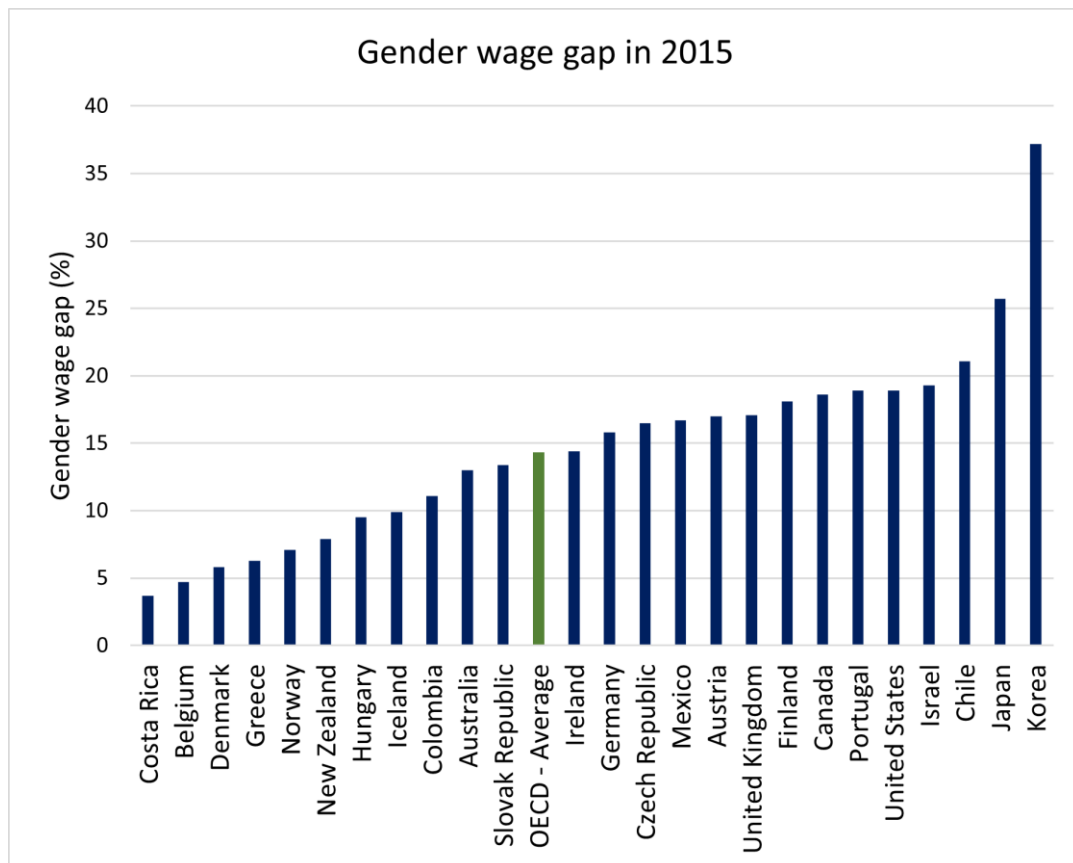


Figure

1: Gender wage gap in 2015 (**Source**)

- **Which three countries have the lowest gender wage gap?**

**Answer:**
Costa Rica, Belgium, and Denmark, respectively, were the first lowest to the third lowest.

- **Which three countries have the highest gender wage gap?**

**Answer:**
Korea is the highest, followed by Japan (2nd), and Chile.

- **Do some research on the country with the lowest gender wage gap and comment on why you think it succeeded in achieving a low gender wage gap in 2015 (max. 150 words).**

**Answer:**

The UNCTAD 2020 IPA Observer report reveals that agencies are increasingly integrating gender equality and women's empowerment in their work, ensuring foreign companies' operations have a positive impact on gender equality (Investment Monitor, 2021).

In this way, Costa Rica has prioritized gender equality as an investment tool, beginning with a gender strategy focus on three key areas:

(1) Supporting companies in facilitating more inclusive hiring practices.
(2) Prioritizing strengthening the local talent pool and stimulating reskilling and capacity-building initiatives in areas where women are disproportionately affected by unemployment.
(3) Bolstering data and reporting mechanisms on gender equality contributes to a better understanding of the causes of the gender gap and enables evidence-based policy measures to rectify it.

These efforts led Costa Rica to rank 13th globally in the World Economic Forum's 2020 annual Gender Gap Report, beating both the UK and the US, which came in at 21 and 53, respectively (UNCTAD, 2020).

2. The following line graph shows the sale of isopropanol (IPA) from May 2019 to March 2020 in the United States of America. The sales are measured using US cents per weight (lb) of the product (US CTS/lb). Focus on the general trend of the three lines on the graph rather than what each line refers to specifically when answering the questions.
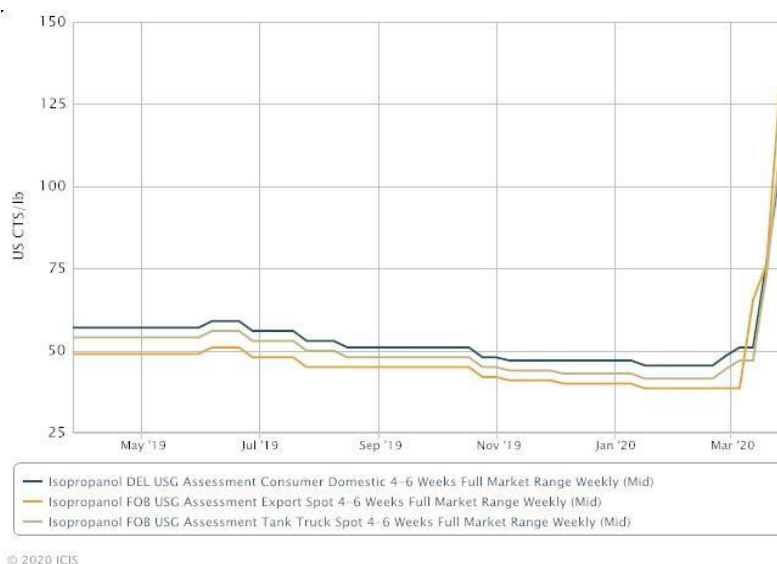


Figure 2: Isopropanol sales from May 2019 to March 2020 (**Source**)

- **Explain what is happening in the graph during March 2020 with regard to isopropanol sales (max. 100 words).**

**Answer:**
As we can see from the line graph, the isopropanol sales had an amazing increase in March 2020. Looking at the US cents per weight (lb) of the product (US CTS/lb), we can presuppose that the increases were something around 100% when we analyze the export and tank truck spots.

- **Describe a possible reason for the observation you made about isopropanol sales in March 2020 (max. 100 words). Hint: Isopropanol is the main ingredient in hand sanitizer.**

**Answer:**
IPA is a solvent used in many products, like cosmetics and those used for personal care. Is also used in hand sanitizers, with the ability to kill germs (Koray, 2020).

On March 11th, 2020, COVID-19 was characterized as a pandemic, a designation recognizing that there were outbreaks of COVID-19 worldwide (PAHO, 2024).

Due to cases of COVID-19 skyrocketing in the US, the domestic prices of IPA increased at unprecedented levels, because of the strong demand for hand sanitizer. It became the first product to sell out in supermarkets and pharmacies, especially with hospitals experiencing a significant increase in COVID-19 patients (Koray, 2020).

3.      Below, the bubble plot (a scatter plot with variable dot size) shows carbon dioxide ($CO_2$) emissions per person in tonnes vs the gross domestic product (GDP) per capita (average per person). No unit is given for the GDP per capita; however, the US dollar is typically used when comparing different countries (Callen, n.d.). Each dot represents a country. The colours of the dots refer to the continent to which the country belongs. The size of the dot refers to the size of the population in the country. The larger the dot, the larger the population.
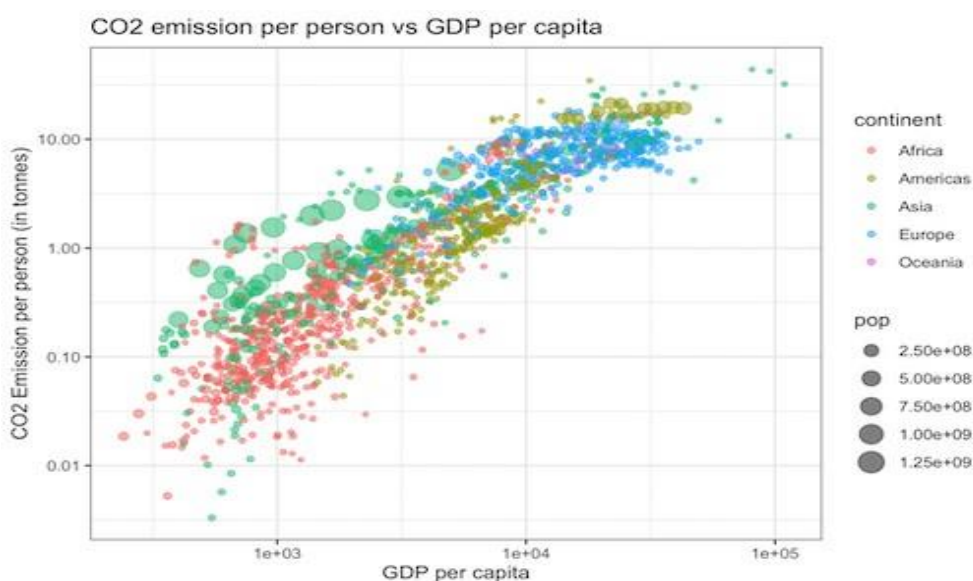


Figure 3: $CO_2$ emissions per person vs GDP per capita

- **Discuss the relationship between $CO_2$ emissions per person and GDP per capita for each continent listed in the figure legend (max. 350 words).**

**Answer:**
The bubble chart above shows the per capita $CO_2$ emissions on the vertical axis against the GDP per capita for each continent listed.

Before analyzing it there are some pieces of information to know about the relationship between $CO_2$ emissions and GDP.

First, is the concept of GDP, which measures the size and health of a country's economy over a period (Bank of England, 2019).

The second is that historically, $CO_2$ emissions have been strongly correlated with how much money a person has: the richer you are, the more $CO_2$ you emit (Ritchie, 2021). According to Roser (2020), people in very poor countries have very low emissions, because of the lack of access to modern energy and technology.

Finally, based on IMF (2023) data, the rank of continents by GDP per capita is:

| Rank | Continent | GDP (US billion) |
|------|-----------|------------------|
| 1. | Asia | 38,435 |
| 2. | North America | 31,603 |
| 3. | Europe | 25,440 |
| 4. | South America | 4,100 |
| 5. | Africa | 2,858 |
| 6. | Oceania | 1,979 |

As we can see in the plot (which shows five continents and is in according with the IMF data), the richest continents, which also are three of the most populous: Asia, Europe, and America, are the ones that emit more $CO_2$ per person, in tonnes, into the atmosphere.

In the continents where people have a GDP per capita higher than $10.000,00 (1e+04), $CO_2$ emissions are between 5,00 and 10,00. On the other side, in countries where GDP is below $1.000,00 (1e+03), the emissions per capita are less than 2,00.

That explains why the African continent, although the second most populous, is on the side of those that emit less. Its population has a GDP per capita of less than $1.000,00, and consequently a deficit in access to modern energy and technology.

Further to Roser (2020), the focus on the rich gives a wrong impression. In the public debate is important the reflect that the sustainable greenhouse gas emissions of most of the world's population are currently too high. The problem is larger for the extremely rich, but it isn't limited to them.

# Practical Task 2

The following scatterplot matrix is from the Boston Housing dataset. It contains US census data concerning houses in various areas around the city of Boston. The data includes the Median House Value, Rooms per Dwelling, Crime Rate, and Percentage of Units built before 1940. Examine this graph and answer the questions that follow in your **data_viz** document:
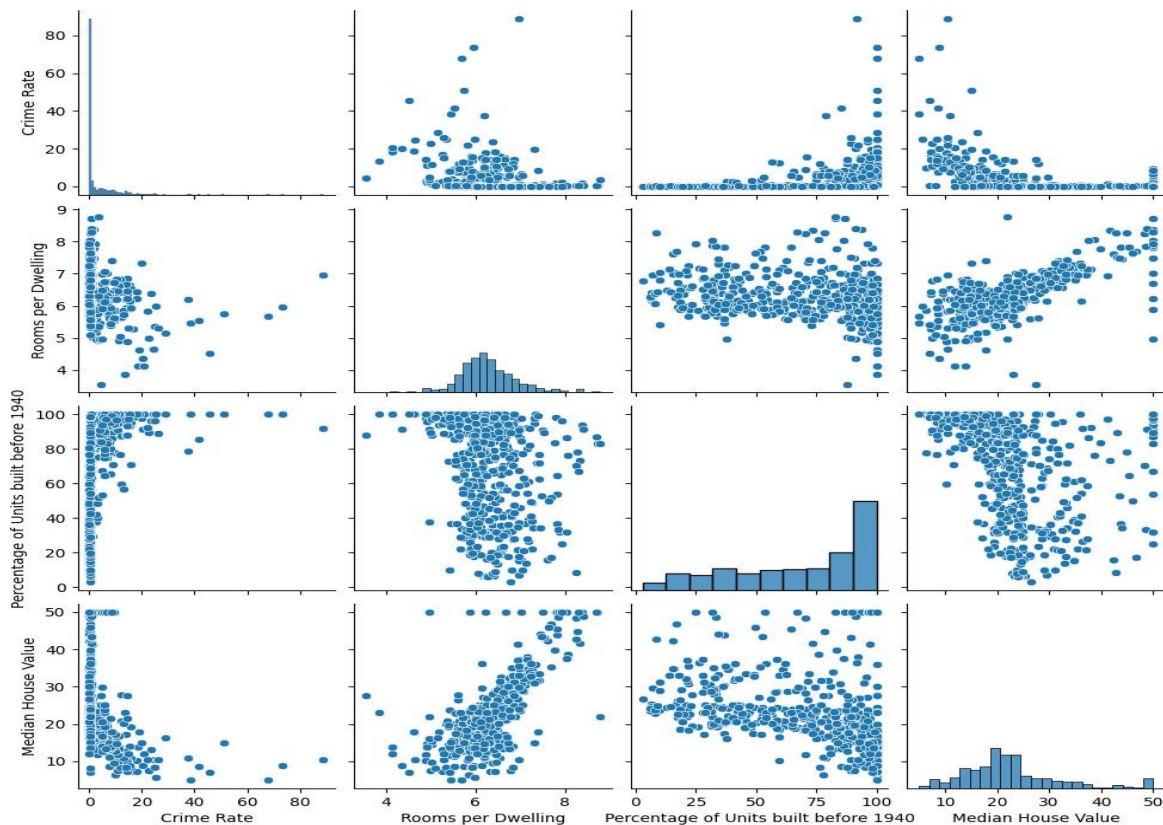


- **Which graph best represents the population of Boston as a whole? Based on this graph, explain whether Boston has a high or low crime rate.**

**Answer:**
The graph that best represents the population of Boston is the histogram (Line 2; column 2) of the Rooms per Dwelling. The average number of rooms per dwelling could be used to estimate housing types and density A histogram is a graphical representation that displays the distribution of a dataset. The example shows a more equitable distribution of the data.

Still based on the graph, we could infer that Boston doesn't have a high crime rate. That inference is possible when we compare the crime rate and the Median House Value. As we can see on the chart it seems there isn't a positive correlation between those two variables, this is, there seems to be a weak and negative correlation between Crime Rate and Median House Value.

- **What is the average number of rooms per dwelling in Boston?**

**Answer:**
The average number of rooms per dwelling can be calculated by summing up the number of rooms for all dwellings and dividing by the total number of dwellings. Based again on the histogram (line 2; column 2), as we don't have all the necessary values to calculate the average, we can infer that in this plot the average number of rooms per dwelling in Boston is approximately 6.

- **What is the correlation between Rooms per Dwelling and Median House Value?**

**Answer:**
Based on the chart in line 2 - column 4, we can say that there seems to be a strong and positive correlation between Rooms per Dwelling and Median House Value.

- **By determining what percentage of units were built before 1940, are the majority of the houses relatively old or new?**

**Answer:**
We can determine if most of the houses are relatively old or new by examining the percentage of units built before 1940. Again, as we don't have all the necessary values to calculate the percentage, by seeing the graph which represents the units that were built before 1940 (line 4; column 3), we can infer that the majority of the houses are older, once it was built before 1940.

- **What is the correlation between the Percentage of Units built before 1940 and the Median House Value?**

**Answer:**
By analyzing the scatterplot, and the distribution of the points on the chart, based on "The Data Visualization Catalogue" (Ribecca, 2024) it seems that the correlation strength of those two variables is none since the majority of the houses were built before 1940.

## REFERENCES

Bank of England (2019). What is GDP? Retrieved 30 January 2024, from: https://www.bankofengland.co.uk/explainers/what-is-gdp [Online Resource].

Callen, T. (n.d.). Purchasing Power Parity: Weights Matter. Retrieved 08 March 2023, from https://www.imf.org/en/Publications/fandd/issues/Series/Back-to-Basics/Purchasing-Power-Parity-PPP#.

Holtz, Y and Healy, C. (n.d.) The Gender Wage Gap. Retrieved 08 March 2023, from https://www.data-to-viz.com/story/OneNumSevCatSubgroupOneObsPerGroup.html.

International Monetary Fund – IMF (2023). "GDP (Nominal), current prices". International Monetary Fund. Retrieved 28 January 2024, from: https://www.imf.org/external/datamapper/NGDPD@WEO/OEMDC/ADVEC/WEOWORLD/AFQ [Online Resource].

Investment Monitor (2021). Closing the gap: Costa Rica prioritises gender equality as an investment tool. Published online at Investmentmonitor.ai. Retrieved 26 January 2024, from: https://www.investmentmonitor.ai/sponsored/closing-the-gap-costa-rica-prioritises-gender-equality-as-an-investment-tool/ [Online Resource].

Jones, T. (2018). An introduction to data science. Retrieved 25 August 2020, from https://www.ibm.com/developerworks/library/ba-intro-data-science-1/index.html.

Koray, D. (2020). US IPA prices rise on unprecedented hand sanitizer demand. Retrieved 13 October 2022, from https://www.icis.com/explore/resources/news/2020/03/26/10487220/us-ipa-prices-riseon-unprecedented-hand-sanitizer-demand/.

PAHO/WHO (2024). Coronavirus Disease (COVID-19). Retrieved 28 January 2024, from https://www.paho.org/en/topics/coronavirus-infections/coronavirus-disease-covid-19-pandemic [Online Resource].

Ribecca, S. (2024). Scatterplot. Published online at datavizcatalogue.com. Retrieved 01 February 2024, from: https://datavizcatalogue.com/methods/scatterplot.html [Online Resource].

Ritchie, H. (2021). Many countries have decoupled economic growth from CO2 emissions, even if we take offshored production into account. Published online at OurWorldInData.org. Retrieved 28 January 2024, from: https://ourworldindata.org/co2-gdp-decoupling [Online Resource].

Roser, M. (2020). "The world's energy problem". Published online at OurWorldInData.org. Retrieved 30 January 2024, from: https://ourworldindata.org/worlds-energy-problem [Online Resource].

United Nations Conference on Trade and Development – UNCTAD (2020). Mainstreaming gender equality in investment promotion. IPA Observer 10 – 2020. Published online at Unctad.org. Retrieved 26 January 2024, from: https://unctad.org/publication/mainstreaming-gender-equality-investment-promotion [Online Resource].