Data Visualisation – Approach and Techniques Tasks

Bronwyn Bowles-King

Practical task 1

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.

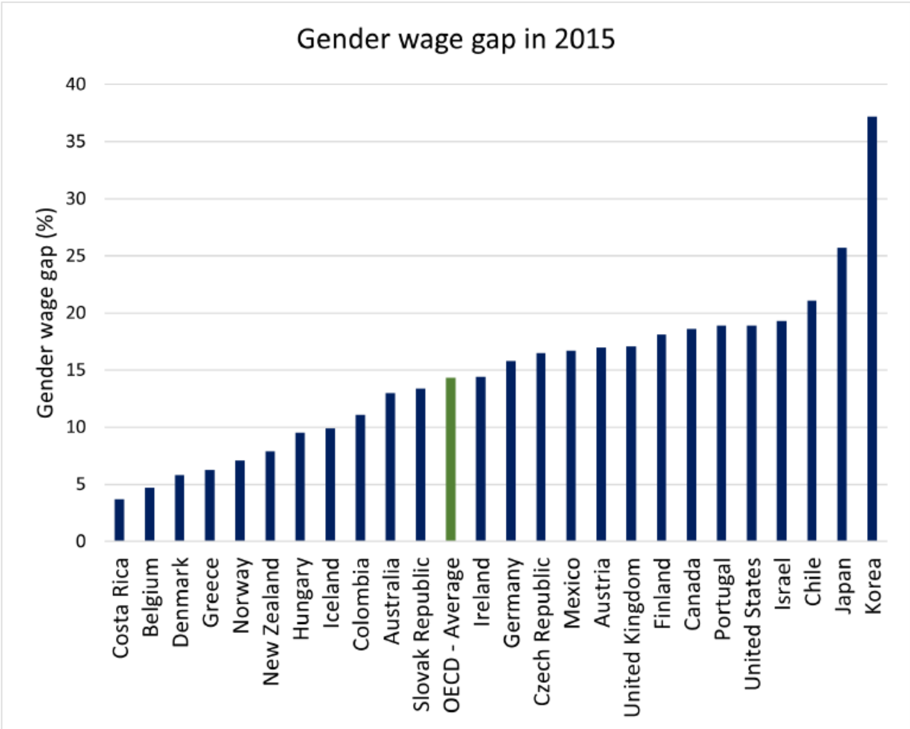


*Figure 1: Gender wage gap in 2015 (**Source**)*

Which three countries have the lowest gender wage gap?

Costa Rica, Belgium and Denmark have the lowest gender wage gap.

Which three countries have the highest gender wage gap?

Korea, Japan and Chile have the lowest gender wage gap.

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).

The IMF (2015) cites Costa Rica's well-developed national policies as promoting gender equality. The gender wage gap is thus low compared to most other countries, although the gap is still present. According to the OECD (2024), this is partly due to women's higher education levels in Costa Rica compared to men. There are also many women working in the Costa Rican public sector or government departments, where salaries are more equitable, and unions ensure fair wages for everyone (OECD, 2024). This pushes down the wage gap for the whole country.

2. The following line graph shows the sale of isopropanol 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 of the lines 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 regards to isopropanol sales (max. 100 words).

Figure 2 is not exactly about isopropanol sales levels in the US. It shows a spike in March 2020 in the *price* for isopropanol, which is not the same as sales. It also seems to be about wholesale prices as opposed to consumer prices. We cannot use direct data from the graph alone to assume sales increased at the same time. In such a case, one would often find sales fall as sudden high prices make people less likely to buy. However, we can apply our own knowledge and look at other sources to find out about sales levels.

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 sanitiser.

Wholesale prices spiked in March 2020 due to the COVID-19 pandemic. We can infer from the source that sales also spiked (ICIS, 2020). With the need for hand sanitising to prevent the spread of the virus, people started using isopropanol products frequently. Producers thus increased prices – initially – without losing sales. However, product prices for consumers in shops did not always spike or remain high. In South Africa, regulators stepped in and industries cooperated rather than solely seeking profit (Competition Commission, 2021; Sasol, 2020). Thus, even as isopropanol sales increased, scarcity did not continue pushing prices up everywhere for everyday consumers.

3. Below, the bubble plot (a scatter plot with variable dot sizes) shows carbon dioxide (CO^2) emissions per person in tonnes versus 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.

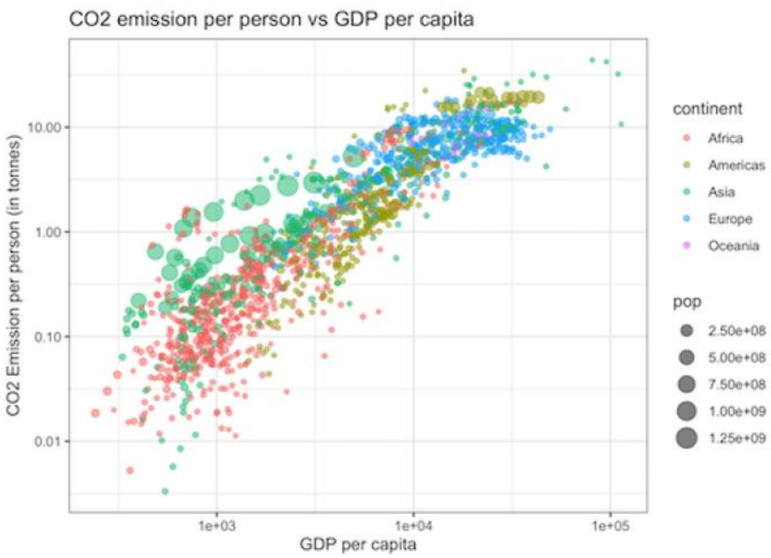


Figure 3: CO_2 emissions per person versus GDP per capita

Discuss the relationship between CO2 emissions per person and GDP per capita for each continent listed in the figure legend (max. 350 words).

The scatter plot in Figure 3 illustrates the relationship between CO_2 emissions per person and Gross Domestic Produce (GDP) per capita for around the world. The plot reveals a positive correlation between GDP rise and pollution for all continents.

African countries (red) are concentrated in the lower left, representing low GDP and emissions per person. Most African nations emit less than one ton of CO_2 per person annually, reflecting lower levels of industrialisation and energy consumption compared to other regions.

Countries in the Americas (green) are spread across a range of GDP and emissions values. Some exhibit high emissions per person at relatively high incomes, indicating economies that are likely using more fossil fuels. Greater variation is thus evident, with both low- and high-emission countries in the region.

Asian countries (yellow-green) are distributed throughout the plot, from low to high incomes and emissions. Several have moderate production but higher emissions, likely due to industrialisation and large populations. The region includes countries with very high incomes and correspondingly high pollution.

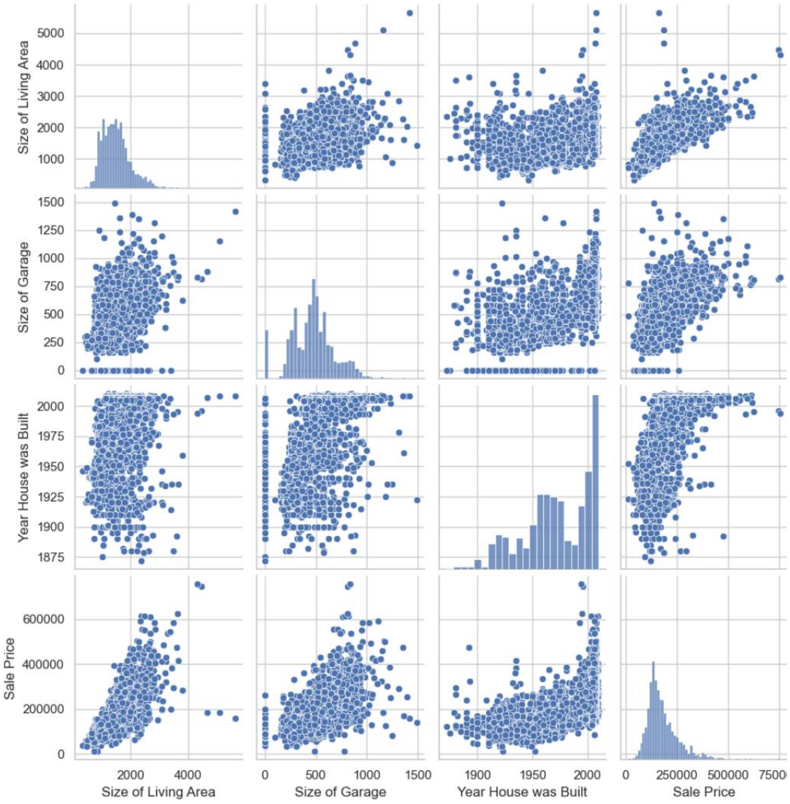
European nations (blue) generally have high GDP and high emissions per person. The latter is unexpected as there are strict EU (2025) environmental policies and a strong move towards cleaner energy. The clustering of blue dots in Figure 3 suggests greater uniformity in incomes and emissions across European nations.

Oceania (pink) is less densely represented as there are fewer and smaller nations in the region. There is a pattern of higher GDP per capita with a narrower but higher range of emission levels. Although the individual nations are not shown in Figure 3, Australia likely has higher GDP and higher emissions, while smaller island nations may have lower values for both metrics within the Oceania cluster.

Across all continents, higher GDP per capita is mostly associated with higher CO_2 emissions per person. However, the strength and nature of this relationship vary by region, influenced by factors such as industrialisation, resources available, policy enforcement, and incentives to change to cleaner energy (Ritchie & Roser, 2020).

Practical task 2

The following scatterplot matrix is from the Ames Housing dataset. It contains data collected by the Ames City Assessor's Office describing 2930 property sales which occurred in Ames, Iowa between 2006 and 2010. The data includes the sale price ($), the year the house was built, the size of the garage (ft^2), and the size of the living area (ft^2).



What do the graphs along the diagonal represent?

The diagonal blocks, moving downwards from left to right, are histograms for individual variables. For example, the histogram in the bottom right corner shows only sales price data.

However, most figures in the matrix above are scatterplots showing data from two of the four variables (living area size, garage size, year of construction, and sale price), allowing us to see how different pairs of variables relate to each other. Due to the crossover of variables, some scatterplots are reversed images of each other, such as the top-right corner and bottom-left corner figures.

Are most garages in Ames larger or smaller than 1000 ft^2?

From all the figures showing this data, we can see that most garages are under 1000 ft^2.

Are the most expensive houses in Ames built before or after 1950?

The more expensive houses were built after 1950.

Describe the relationship between “Size of Living Area” and “Sale Price”.

Larger living areas equate with higher prices in the majority of cases, as shown in the scatterplot in the bottom-left corner. The data points are clustered in the lower left of this mini-plot, indicating that more houses have living areas between about 1 000 and 2 500 ft^2 and sell for between $100 000 and $350 000. As one moves to the right, where there are larger homes, sale prices also rise, with some homes exceeding $500 000 and living areas above 3 000 ft^2.

There are a few outliers and a point at which a large living area can become less attractive and equates with a lower price. There are three examples showing the largest living areas of the homes in the dataset (over 4 000 ft^2), yet all their prices are low (under $200 000). This can be related to the costs of maintenance for a large home, and possibly to location. However, two homes with only slightly smaller living areas command the highest prices in the dataset (over $700 000). These two homes were probably well-maintained and could be in a more attractive area.

References

Competition Commission. (2021). Consent Agreement: Competition Commission and Crest Chemicals (Pty) Ltd. Case No: CO176Dec20. http://blogs.sun.ac.za/ccle/files/2021/06/NOM-CC-And-Crest-Chemicals-Non-confidential-CO176Dec20.pdf

EU. (2025). Environment policy: General principles and basic framework. https://www.europarl.europa.eu/factsheets/en/sheet/71/environment-policy-general-principles-and-basic-framework

HyperionDev. (2025). Data Visualisation – Approach and Techniques. Course materials. Private repository, GitHub.

ICIS. (2020). US IPA prices soar on rising global demand and supply shortage. https://www.icis.com/explore/resources/news/2020/03/19/10484600/us-ipa-prices-soar-on-rising-global-demand-and-supply-shortage

IMF. (2015). Fair Play: More Equal Laws Boost Female Labor Force Participation. https://www.imf.org/external/pubs/ft/sdn/2015/sdn1502.pdf

OECD. (2024). Gender Equality in Costa Rica. https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/07/gender-equality-in-costa-rica_06ebece0/a25e8f8a-en.pdf

Ritchie, H. & Roser, M. (2020). CO_2 emissions. Our World in Data. https://ourworldindata.org/co2-emissions

Sasol. (2020). Sasol responds to increased demand for alcohols used in sanitisers and disinfectants in South Africa. https://www.sasol.com/sasol-responds-increased-demand-alcohols-used-sanitisers-and-disinfectants-south-africa-3