**Practical Task 1**

1. According to the information in graph provided, Costa Rica, Belgium and Denmark had the lowest gender wage gap in 2015 and Korea, Japan and China had the highest gender wage gap. However, by comparing the [source data](https://stats.oecd.org/index.aspx?queryid=54751) with the [given table on Github](https://github.com/holtzy/data_to_viz/blob/master/Example_dataset/9_OneNumSevCatSubgroupOneObs.csv), observe that the author of that table has made a number of errors whilst copying the data. For example, the gender wage gap at median for Chile in 2015 is 14.3 but is wrongly recorded on Github as 21.1. Similarly, Colombia’s median gender wage gap for the years 2010 and 2015 are 12.7 and 14.2 respectively but are wrongly recorded as 6.4 and 11.1 respectively. More significantly, Costa Rica’s 2015 median gender wage gap is 9.1 but is recorded incorrectly as 3.7. The correct graph should look like this:  
     
   Belgium achieved the lowest gender pay gap, most likely because of widespread collective bargaining on wages and the instruction of a new law in 2012 which made it mandatory for the gender pay gap to be specifically considered when unions and employers negotiate their wage agreements.

1. In March 2020, the sale of isopropanol increases very significantly with sales more than doubling, suggesting a sudden surge in demand or a squeeze in supply, leading to a higher market price. A possible reason for the spike in isopropanol sales in March 2020 could be due to increased demand for sanitizing products because of the COVID-19 pandemic. Isopropanol is a key ingredient in hand sanitizers and disinfectants, products that saw a dramatic increase in use as the public and institutions sought to prevent the spread of the virus.
2. For each continent, there is a positive correlation between CO2 emissions per person and GDP per Capita. Africa (red points): Most African countries are clustered in the lower left corner of the graph, which suggests that CO2 emissions per person are relatively low, and the same is true for GDP per capita. This indicates that, on average, African countries have lower economic output per person and are associated with lower levels of CO2 emissions per person.

The points for the Americas are spread across a wide range of the graph, suggesting significant variation among countries in this continent. Some countries in the Americas have low GDP per capita and low CO2 emissions per person, like many African countries. However, there are also countries with high GDP per capita and higher CO2 emissions per person, likely including the United States and Canada.

Asian countries also show a wide distribution, with many points concentrated in the middle of the GDP per capita range and a moderate level of CO2 emissions per person. This suggests a mix of developing and developed economies, with corresponding levels of CO2 emissions.

European countries tend to have higher GDP per capita and a wide range of CO2 emissions per person. The distribution suggests that while economic output is generally high, Europe has a mix of countries with both high and low CO2 emissions per person, which may reflect diverse energy policies and levels of industrialization.

There are fewer points representing Oceania, and they tend to have higher GDP per capita and vary in CO2 emissions per person. Given the smaller sample size, it's difficult to generalize, but it can include countries like Australia and New Zealand, which have high living standards and consequently higher emissions.

The overall trend suggests that there is a positive correlation between GDP per capita and CO2 emissions per person, meaning that as the economic output of an individual increases, so generally does their carbon footprint. However, this relationship is not uniform across or within continents, reflecting different levels of economic development, industrialization, energy sources, and efficiency technologies.

**Practical Task 2**

**NB:** I assume that “Rooms per Dwelling” should read “Average number of rooms per dwelling” given that each point in the scatterplot matrix represents aggregated data for a particular area or neighbourhood in the Boston area, not an individual house. Further, I assume that the first question is really asking which graph best represents the *crime rate* of the population of Boston as a whole, since the question which follows presumes that said graph would be useful in explaining whether Boston has a high or low crime rate.

1. That graph would be the histogram located in the first row, first column because this is a distribution of the crime rate variable alone. From this graph, it appears that Boston has a generally low crime rate for most areas, as indicated by the high frequency of areas with near-zero crime rates. However, there is a long tail to the right, which indicates that there are a few areas with a very high crime rate.
2. I cannot provide an accurate answer to the question of what the average number of rooms per dwelling in Boston is because I only have access to a small histogram, not the original dataset. However, from a cursory observation of the histogram, the average number of rooms per dwelling in Boston is approximately six.
3. The correlation between Rooms per Dwelling and Median House Value is, as one would expect, a positive correlation.
4. The histogram in the third row, third column shows the distribution of the percentage of units built before 1940. A significant number of units have a high percentage (close to 100%), which suggests that the majority of the houses are relatively old.
5. There is a very weak negative correlation between the Percentage of Units built before 1940 and the Median House Value. The scatterplot in the third row, last column illustrates the relationship between the Percentage of Units built before 1940 and the Median House Value. The correlation appears to be slightly negative, indicating that areas with a higher percentage of older units tend to have a lower median house value, although this trend is not very strong.