

DaV report- homework 3

Lab 4-5

Zofia Kochańska

The following report contains 15 stationary plots all of which were created using python. Above each type of plot there is a short information about the placement of their elements and their origin. All of the figures are presenting the same dataset containing the information about the temperatures across the years in 8 different countries. Depicted plots can be divided into four categories:

- Scatter plots
- Box plots
- Time series
- Multiple subplots

Scatter plots

The first type of presented plots is a scatter plot showing the average temperatures across the years. The X-axis shows consecutive years while the Y-axis contains information about the temperature values presented in Celsius. The following plots each have some new elements added when compared to the previous one. The first one- showed in Fig.1, is the most basic one, where the points are presented as circles with black edges. In the following plot named Fig.2 the circles are replaced with the black dots. Next Figure is slightly more readable as the dots are more transparent which allows for the easier analysis of the data. The last of the scatter plots has some color added, but apart from this feature is identical to the Fig.3.

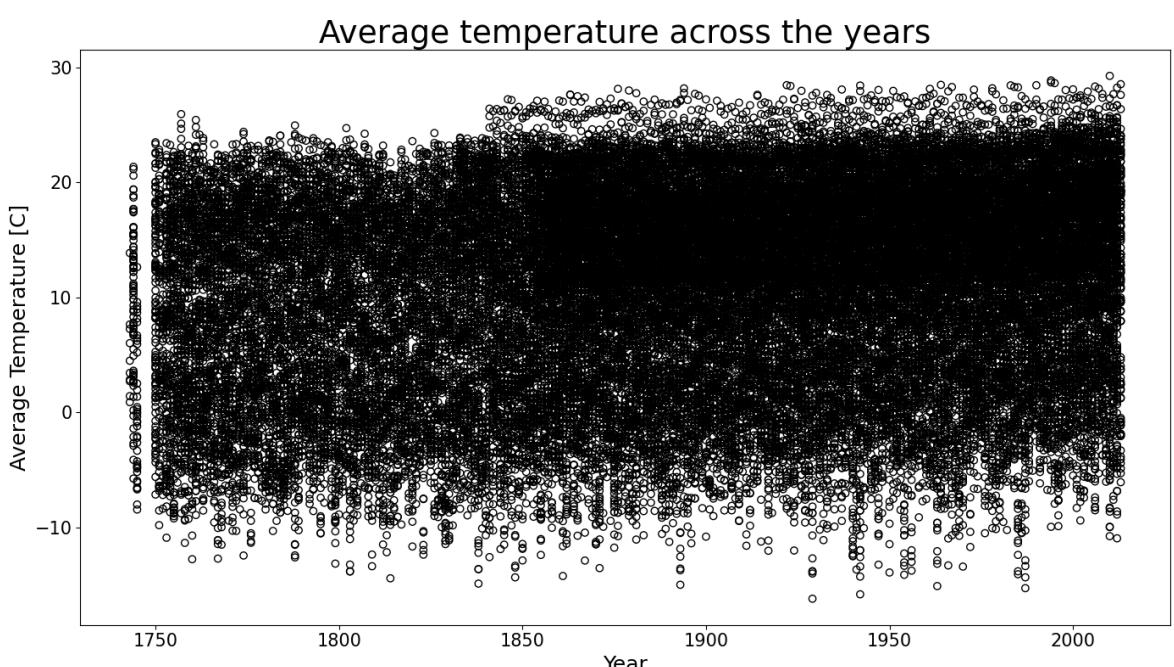


Fig.1 Plain scatter plot with no added features

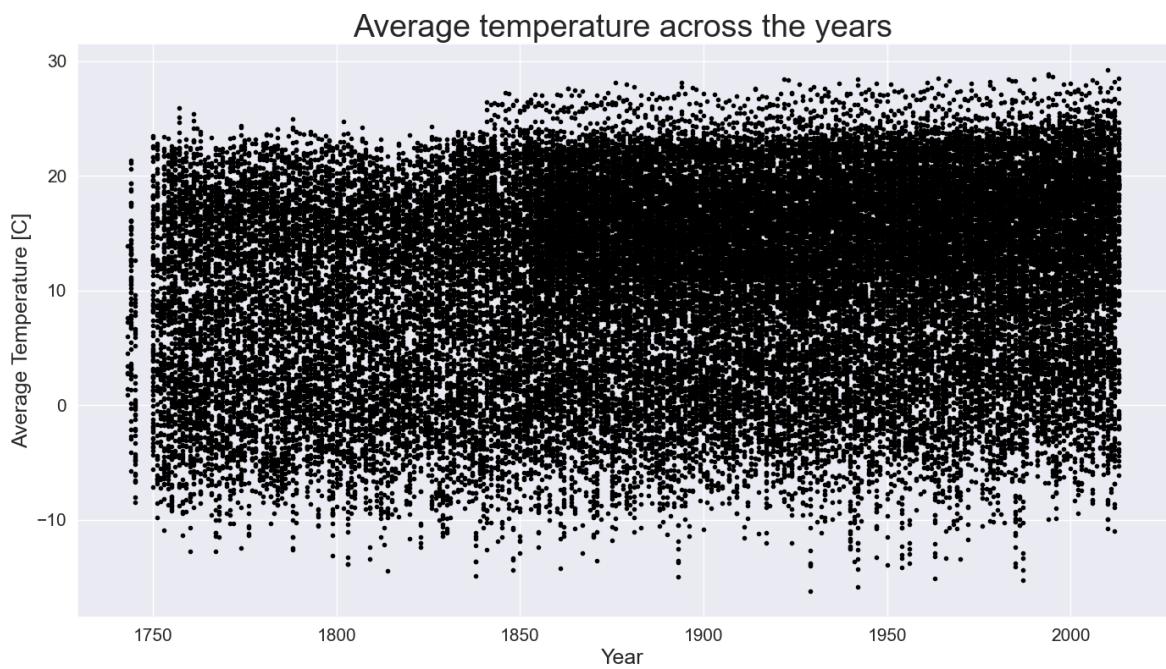


Fig.2 Scatter plot with changed markers representing different temperatures

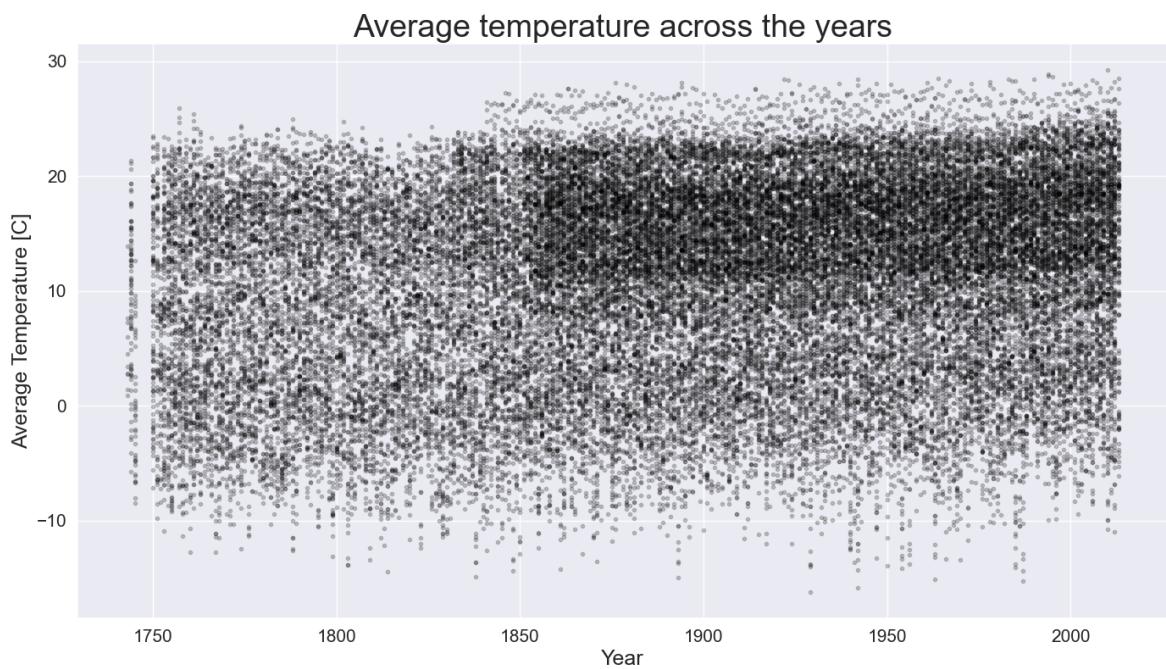


Fig.3 Scatter plot with increased transparency of different data points

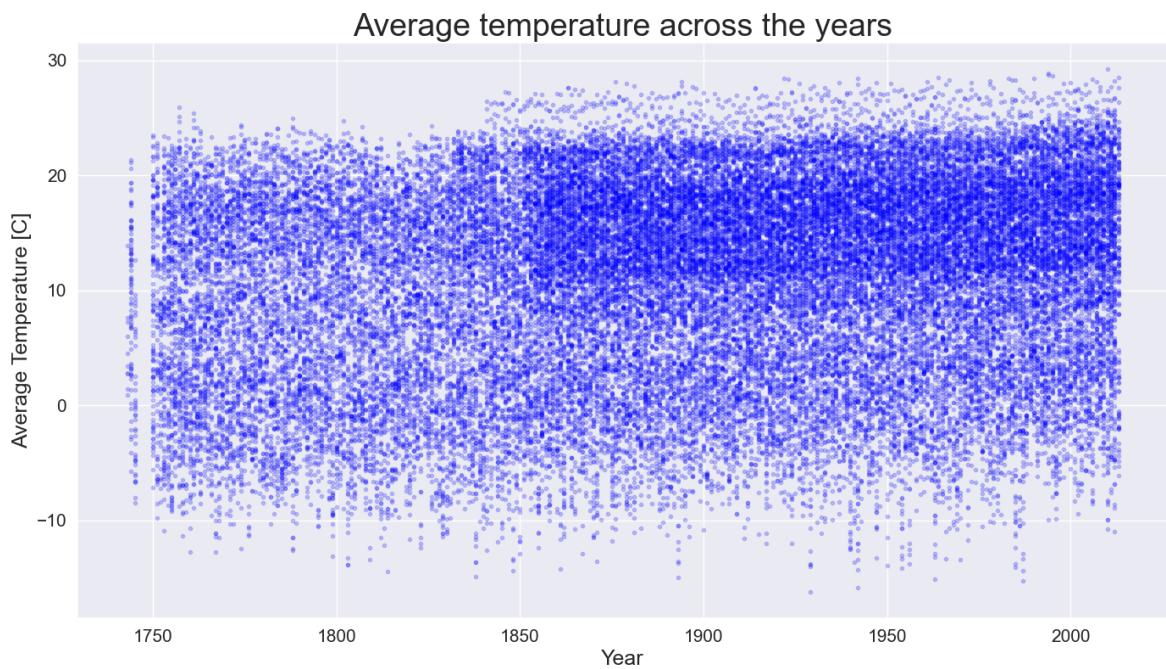


Fig.4 Scatter plot with added color

Box plots

The next type of presented plots is a box plot, where the analyzed countries are placed on the X axis while observed temperatures are visible on the Y axis. The first of the included box plots has no distinguishing features but still manages to visualize the distribution and key features of the data. The median is represented by a line in the middle of the box, dividing it into two halves. It is a measure of central tendency that shows the value that divides the data into two equal parts. The box of the plot represents the interquartile range (IQR) of the data, which is the range of the middle 50% of the data. The bottom of the box represents the first quartile (Q1), which is the value below which 25% of the data falls. The top of the box represents the third quartile (Q3), which is the value below which 75% of the data falls. Vertical whiskers of the plot extend from the box and represent the range of the data outside the IQR. The length of the whiskers varies as they extend to the smallest and largest data points that fall within 1.5 times the IQR from the nearest quartile. Data points that fall outside the whiskers are considered outliers and are plotted as individual points. The feature added to second plot is jitter- a small amount of individual data point. The last plot showed in this section is a violin plot, where the shape of the plot represents the distribution of the data.

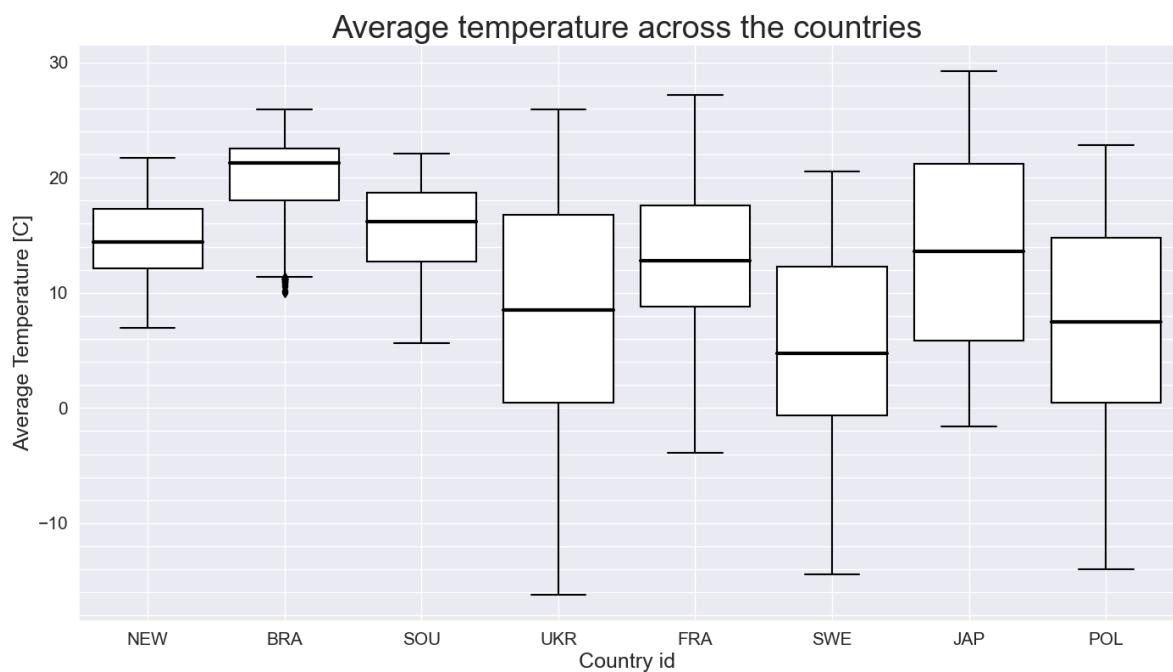


Fig.5 Plain box plot showing temperatures across the years in different countries

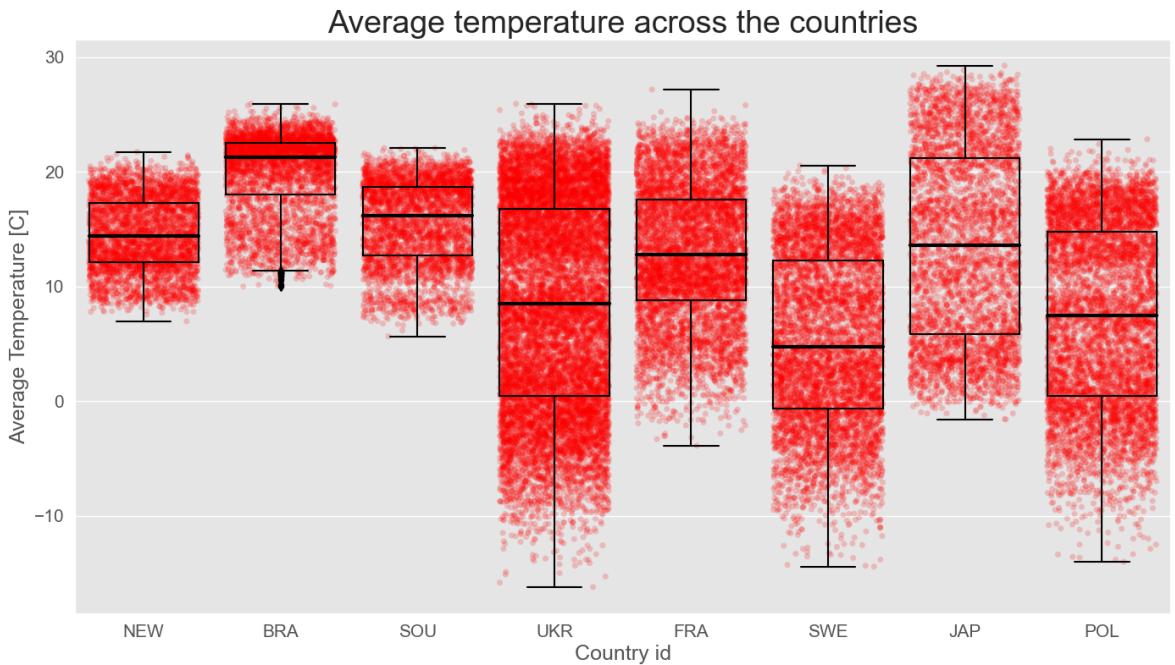


Fig.6 Box plot showing temperatures across the years in different countries with added jitter showing the data points

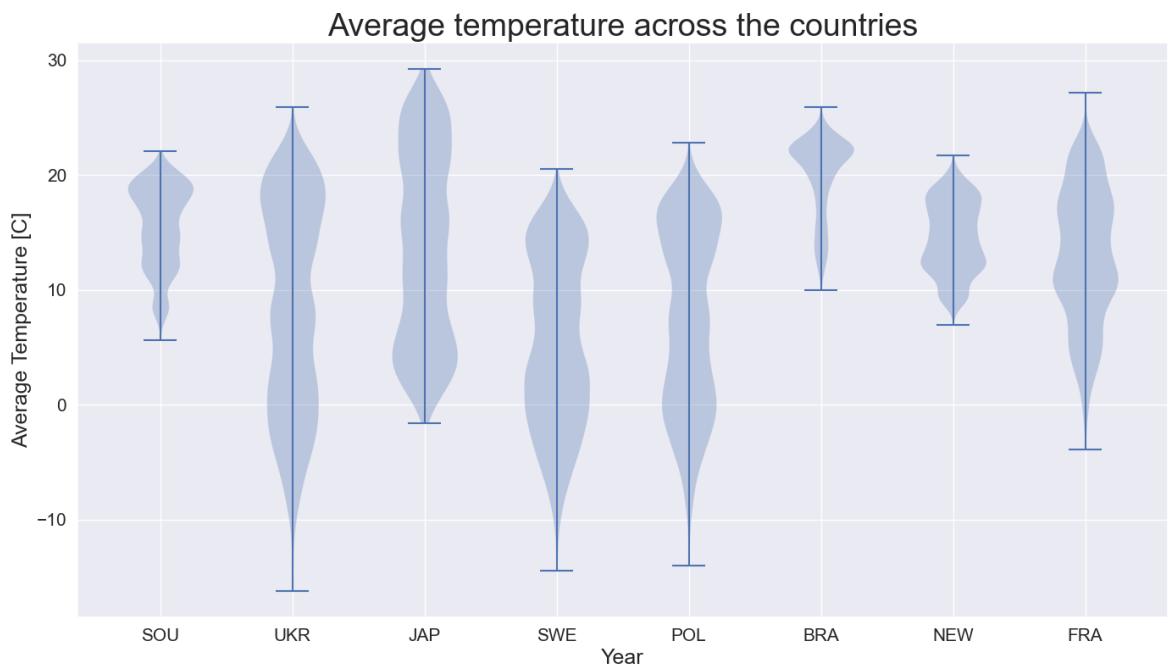


Fig.7 Violin plot showing temperatures across the years in different countries

Time series

This type of plot is showing the mean of the temperatures noted in different countries as lines, where the X axis represents the years. The first plot, showed in Fig.8 is unreadable due to many points on the Y axis being plotted for the same point on the X axis. This issue is fixed in the next plot, where the temperatures noted in different countries are represented by separate, black lines. In the final Time series plot (Figure 10), the lines are additionally colored to represent different countries, and a legend is included to enable users to check their values.

Average temperature across the years

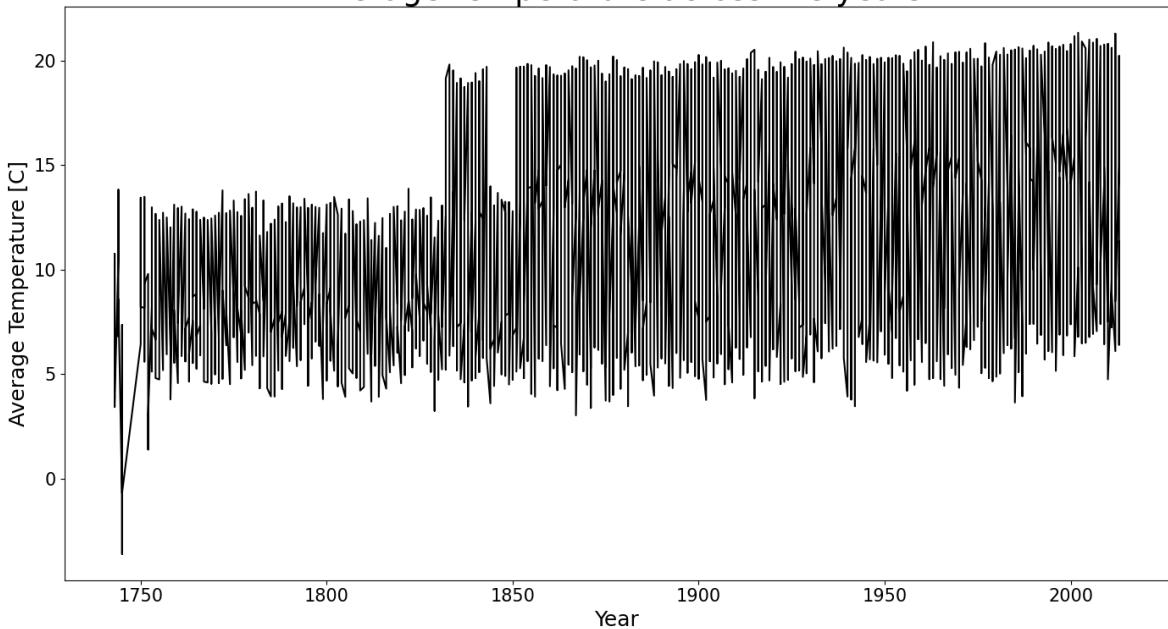


Fig.8 Time series plot showing mean temperatures observed in different countries across the years

Average temperature across the years

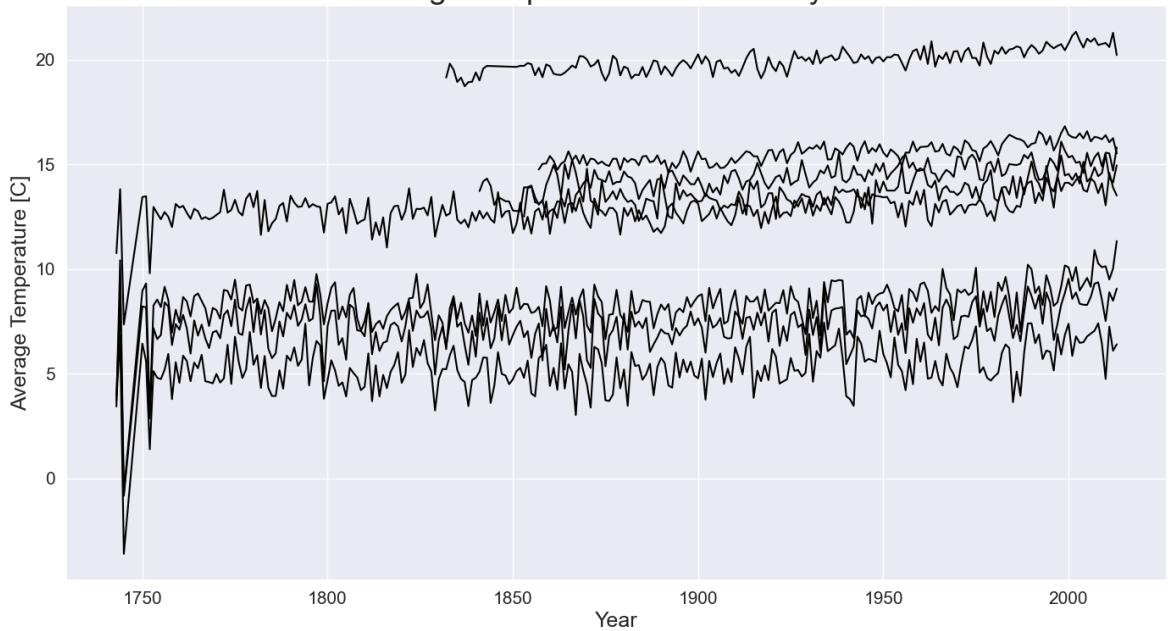


Fig.9 Time series plot showing mean temperatures observed in different countries across the years with separate lines

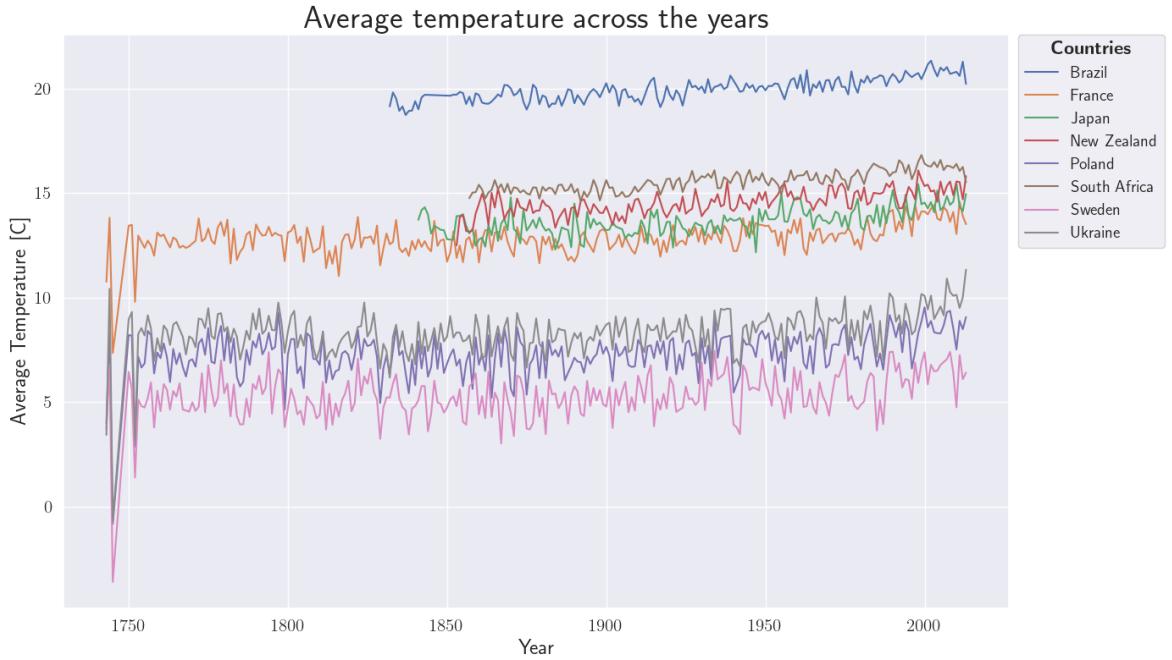


Fig.10 Time series plot showing mean temperatures observed in different countries across the years with separate colored lines

Multiple subplots

The final type of plots is similar to the Time Series plot shown in the previous section, with the only difference being the inclusion of multiple subplots in the following plots. This addition allows for better readability and the possibility of displaying more data. Figure 11 is the simplest of these plots and shows separate subplots for each line representing the temperatures in different countries. The next plot is slightly more complex and displays each individual line representing the mean temperature recorded in different cities. The only difference between Fig. 12 and 13 is the lack of a grey background in the latter plot. In the following Figure, different colors are used to represent each city for which the temperatures were observed. The last plot was further modified by adjusting the sizes and placements of labels, titles, and legends for the best possible readability.

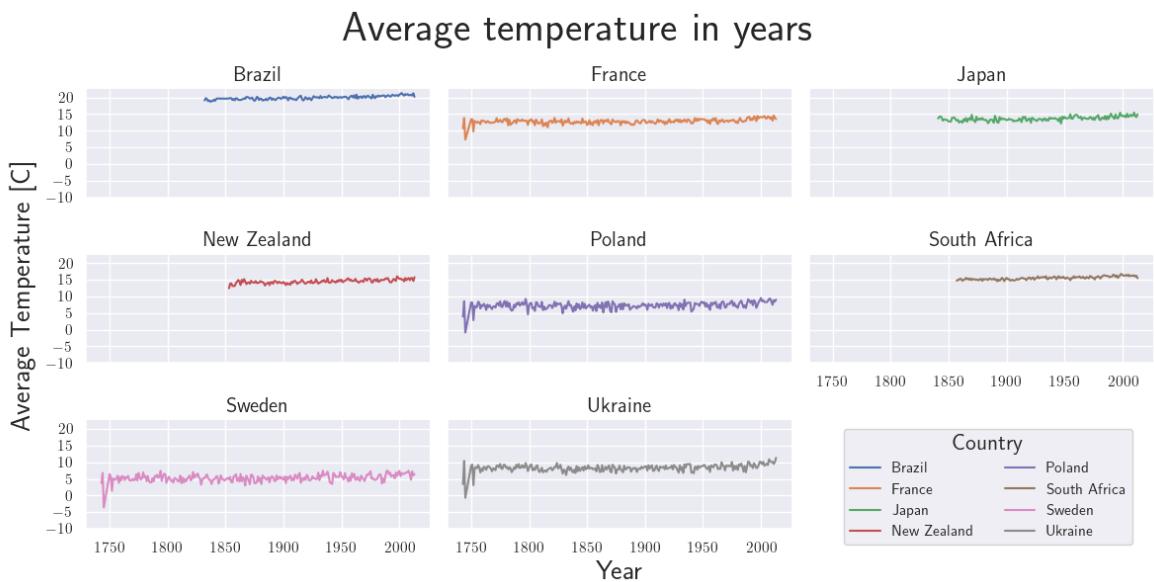


Fig.11 Plot with multiple subplots showing mean temperatures in different countries

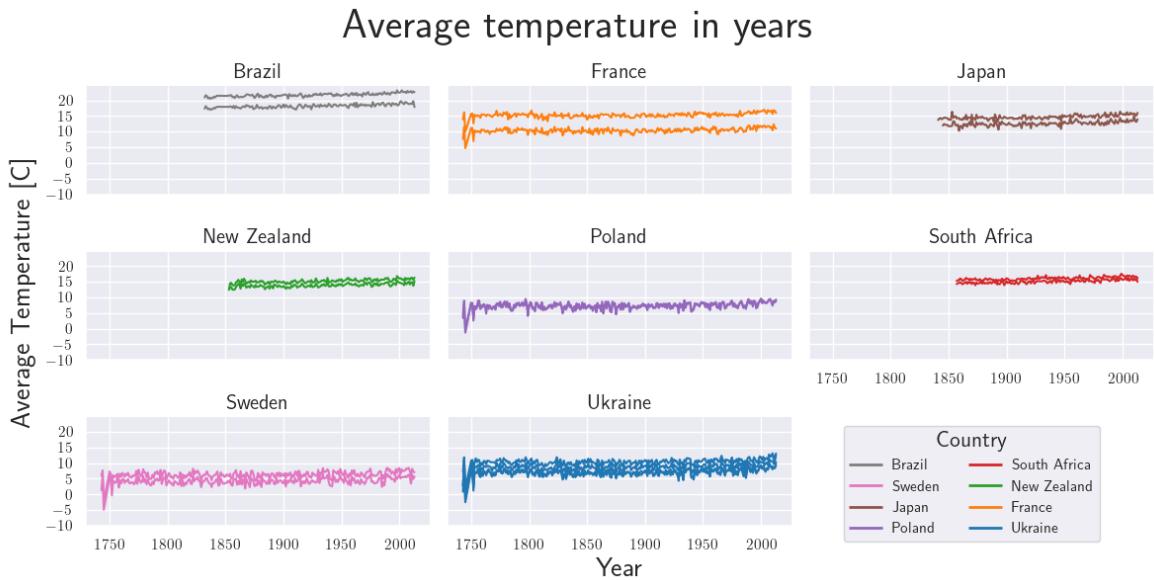


Fig.15 Plot with multiple subplots showing mean temperatures in different countries and cities

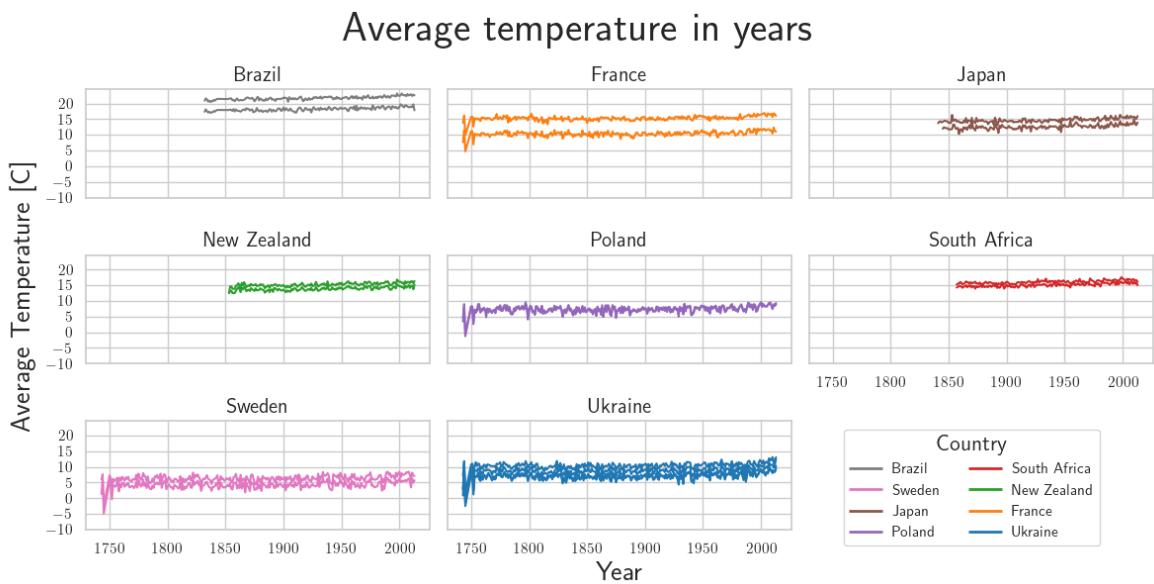


Fig.14 Plot with multiple subplots showing mean temperatures in different countries and cities without the grey background

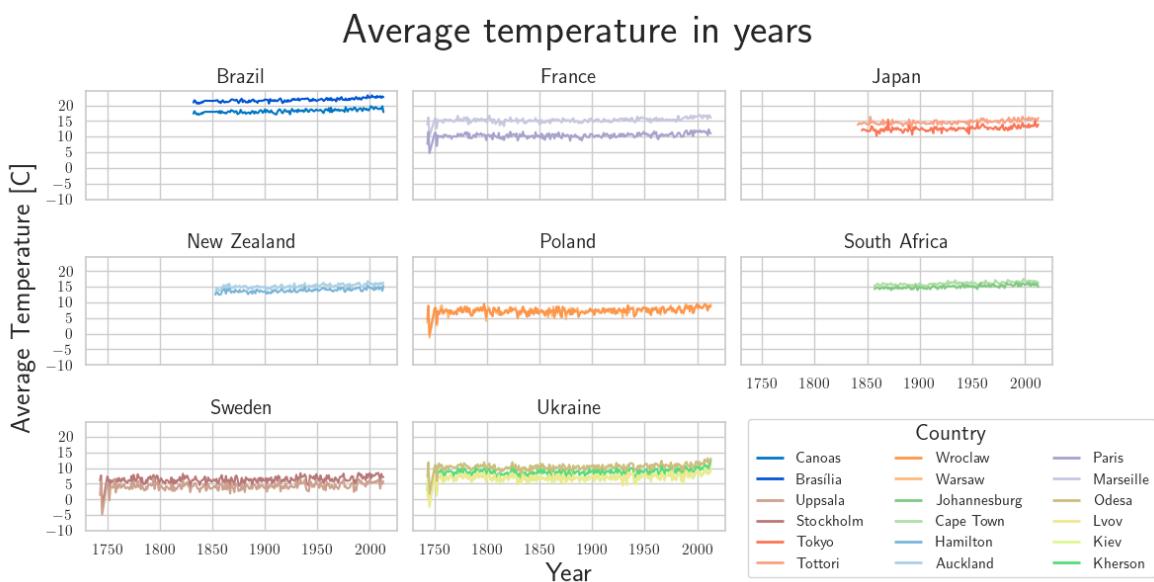


Fig.15 Plot with multiple subplots showing mean temperatures in different countries and cities represented by different colors

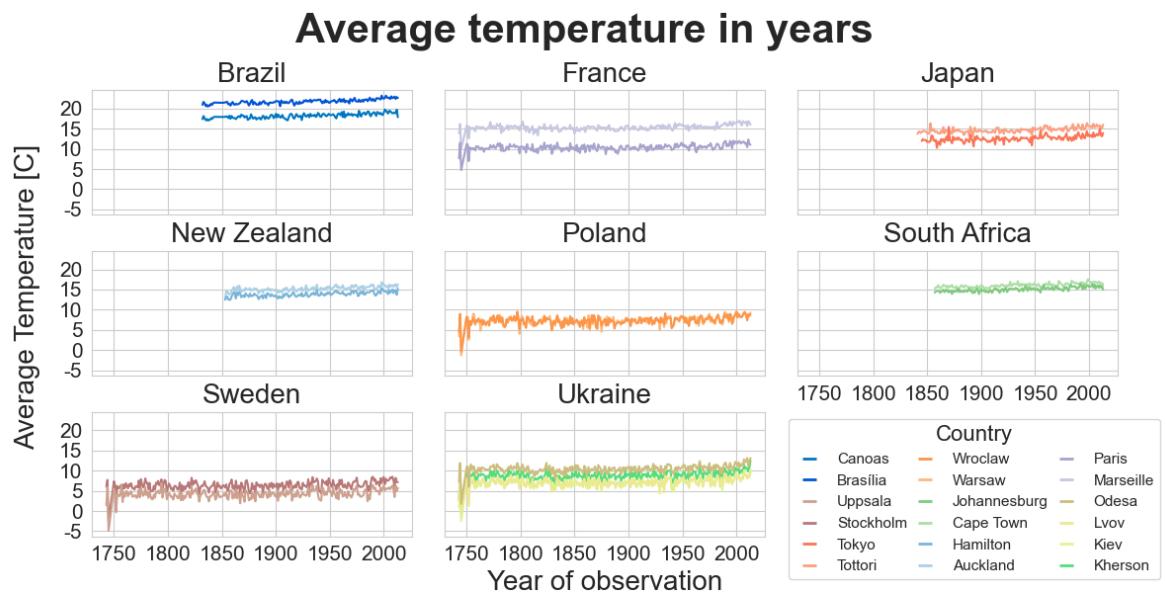


Fig.16 Plot with multiple subplots showing mean temperatures in different countries and cities represented by different colors with increased readability