Solutions to task

Output for task 1

Histogram for 00AJ: Assaults in 09/11 [
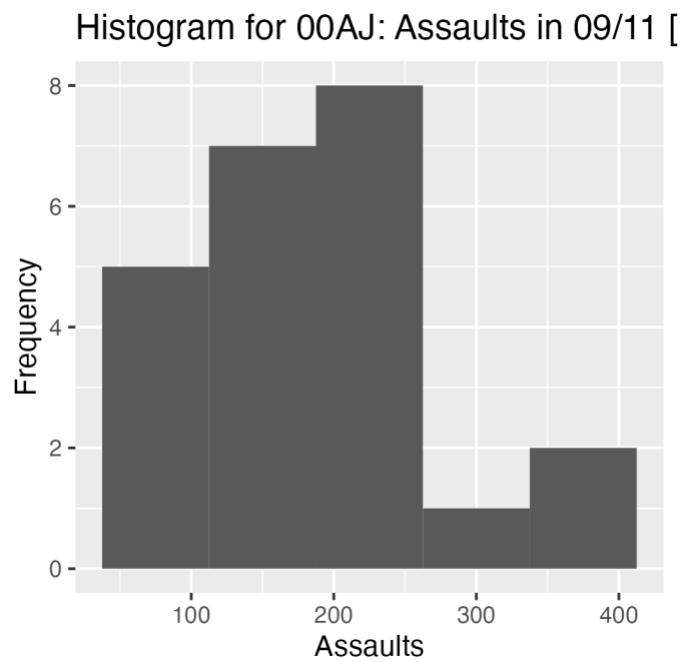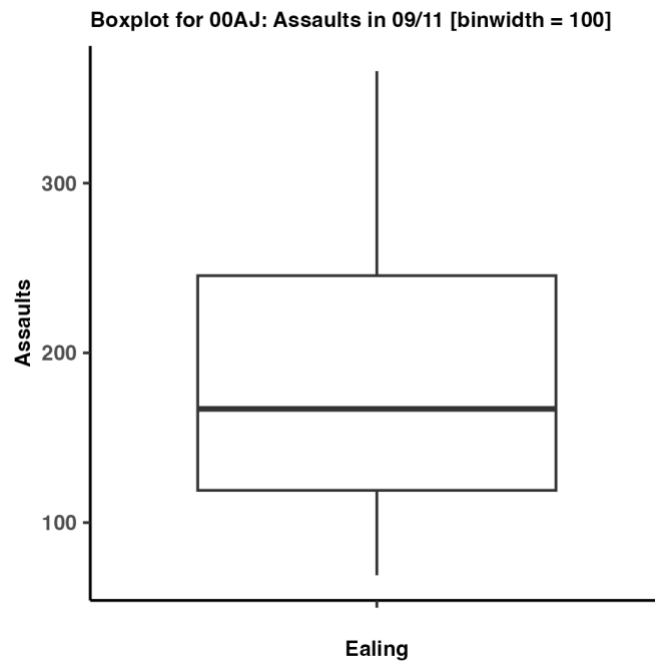
It looks quite broken with titles not sitting within the frame etc.

Output for task 2

Histogram for 00AJ: Assaults in 09/11 [binwidth = 100]
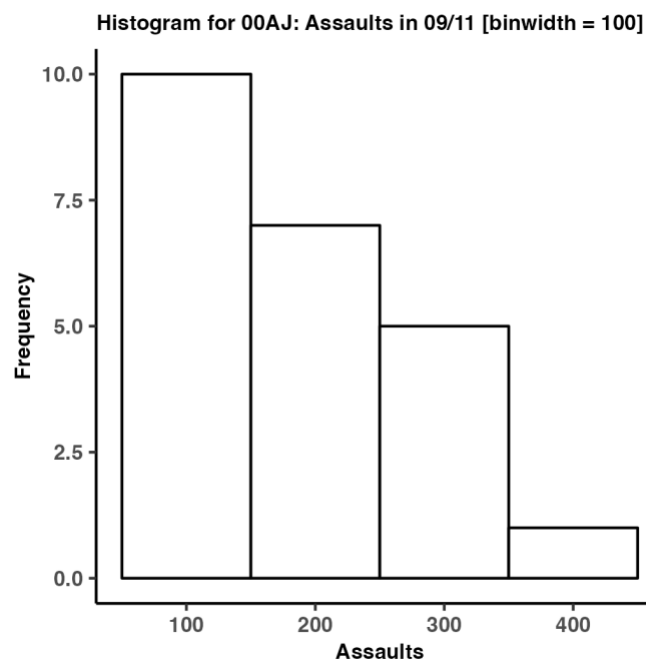
It looks a lot cleaner with title and labels formatted. The appearance of the image has better quality compared to the first plot.

Output for task 3

Boxplot for 00AJ: Assaults in 09/11 [binwidth = 100]

Again, it's a clean box plot with title and labels formatted.

Solutions to questions

Question 1

Graphing technique	Data type	Usage
Scatter plot	Continuous data	A dependent and independent variable, both continuous, are needed for a scatter plot. This helps us how an independent variable plotted on an x-axis impacts or effects the dependent variable which is plotted on the y-axis
Histogram	Continuous data	This is used to assess the shape or distribution of a single continuous variable. It let us see the frequency of a value(s) grouped in bin widths.
Box plot	Continuous data (which can be broken down a categorical variable)	This is used to provide a summary of the data set. It is akin to using the summary() function which reports these descriptive statistics (i.e., median, Q1, Q3, min and max. However, it presents them visually.
Line Chart	Continuous data (that is measured temporally)	It is best for assessing trends or changes in a continuous outcome over time

Question 2

Croydon: Assaults [binwidth = 75]

Croydon: Assaults [binwidth = 100]
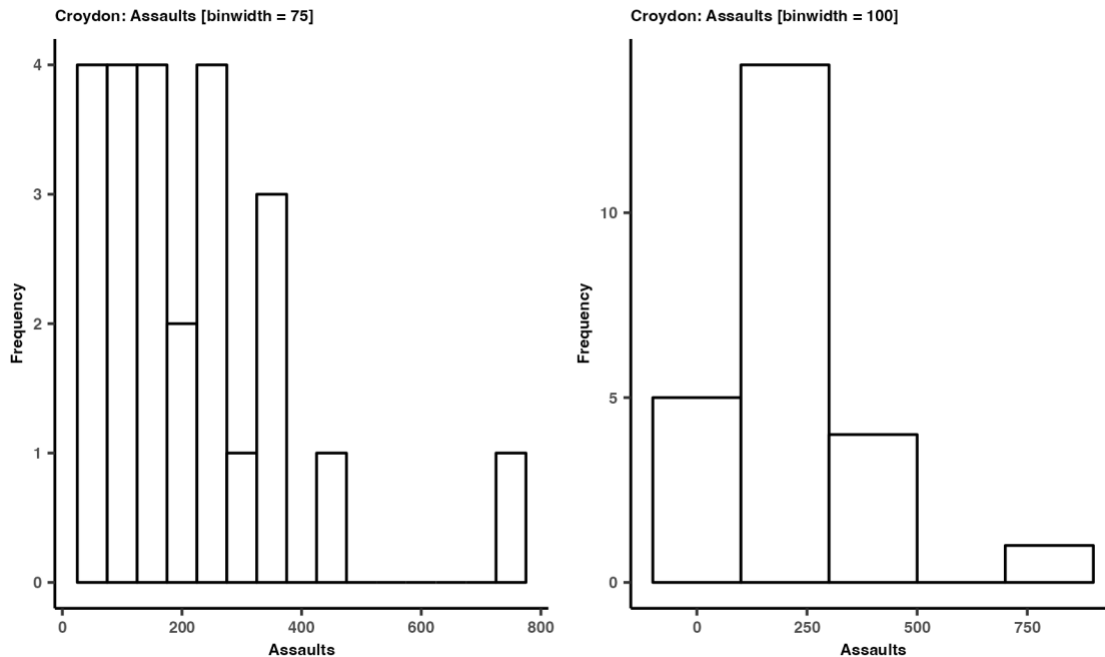
You may want to be cautious on bin width choice – note that a histogram will always show continuous data by grouping data into "bins" of equal width. Note that a bin is plotted as a bar whose height corresponds to how many data points fall within bin. To small bin values will have fewer data points, and very large bin values will indeed have too many data points – and so it's all about have the right balance and using some bin value that is sensible.

Note that the term bins are also sometimes interchangeable with "intervals", "classes", or "buckets".