Task 2

b) The train set has 80 entries, ranging in observational values from 1 to 110. The median is 55.5, the mean is 54.5. It has no dimensions and no column names.

The test set has 31 entries, with observational values ranging from 6 to 111. The median is 56.0, the mean is 59.9. It also has no dimensions and no column names.

Ozone has 111 entries in 4 columns named “ozone”, “radiation”, “temperature” and “wind”. Correspondingly, its dimensions are [111 4]. The column “ozone” ranges from 1 to 168, with a median at 31.0 and a mean at 42.1. The column “radiation” ranges from 7 to 334, with a median at 207.0 and a mean at 184.8. The column “temperature” ranges from 57.0 to 97.0, with a median at 79.0 and a mean at 77.79. The column “wind” ranges from 2.3 to 20.7, with the median at 9.7 and the mean at 9.94.

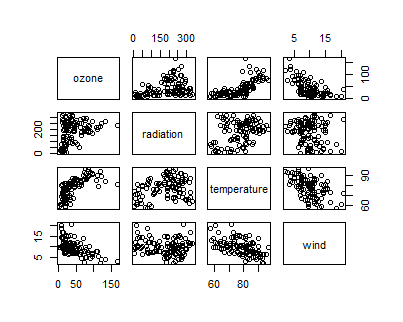
c) You can see the scatterplot matrix for every feature in the dataset below.

Figure 1: Scatterplot for every pair of variables in the dataset

The Pearson correlation coefficient between ozone and radiation is 0.35, between ozone and temperature 0.7 and between ozone and wind -0.61. The correlation between radiation and temperature is 0.29 and between radiation and wind -0.13. The correlation between temperature and wind is -0.5.

In general, the range of the correlation coefficient is -1 to 1. A correlation coefficient of 0 means that there is no correlation between the two variables. Positive values mean positive correlation, negative values mean negative correlation.

In the data we can see that the highest positive correlation is between ozone and temperature, which can be seen in the scatterplot as a distribution with a positive slope. The most negative correlation occurs between ozone and wind, which can be seen in the scatterplot as a downward trend. The next most negative correlation is between temperature and wind, which also show a general downward trend in the scatterplot. The lowest absolute correlation value occurs between radiation and wind, which can be visually confirmed by looking at the scatterplot in which the values seem to display no trend. The correlation coefficients between ozone and radiation as well as radiation and temperature are also quite low.

d) The column “ozone” ranges from 1 to 168, with the mean at 42.1 and a variance of 1107.29.

The column “radiation” ranges from 7 to 334, with the mean at 184.8 and a variance of 8398.74. The column “temperature” ranges from 57.0 to 97.0, with the mean at 77.79 and a variance of 90.82. The column “wind” ranges from 2.3 to 20.7, with the mean at 9.94 and a variance of 12.67.

e) 