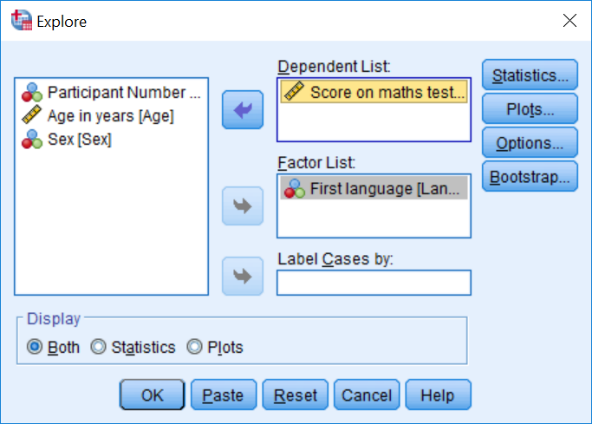
**Exercise 3**

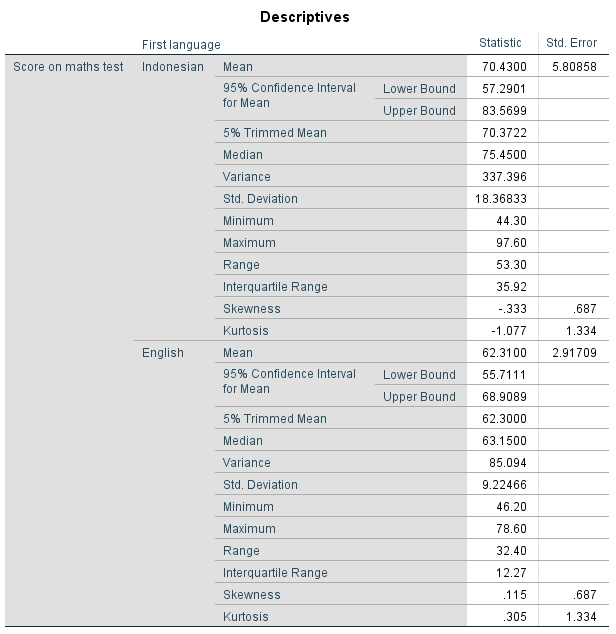
*****Parts 1a and 1b*

**Analyse -> Descriptive Statistics -> Explore**

Move the IV (Language) to the Factor List; and the DV to the Dependent List

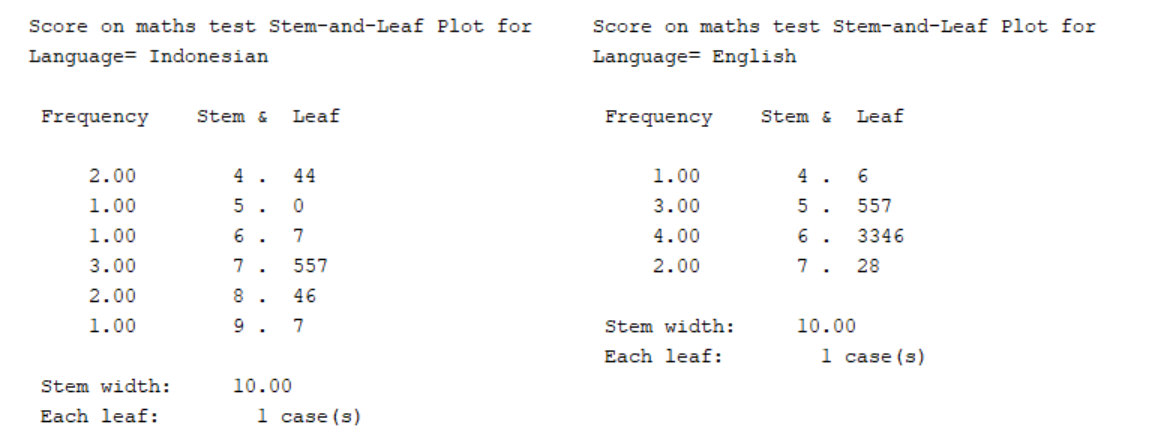
In the Statistics option, ensure ‘Descriptives’ is selected

In the Plots option, make sure Stem and Leaf Plots are ticked



From the output, you can see on the Mean rows and in the Statistics column that the Indonesian speakers had a higher mean score of 70.43 than the English speakers 62.31.

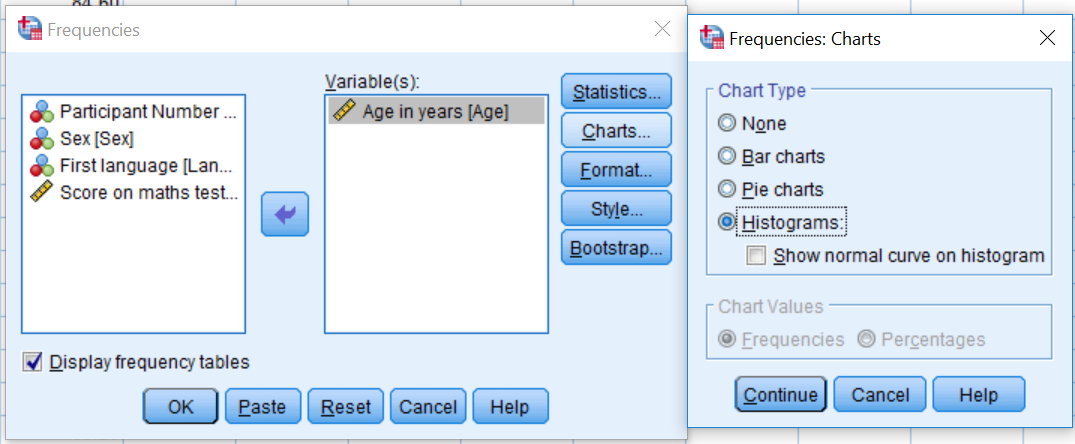
The Stem and Leaf plots are below.



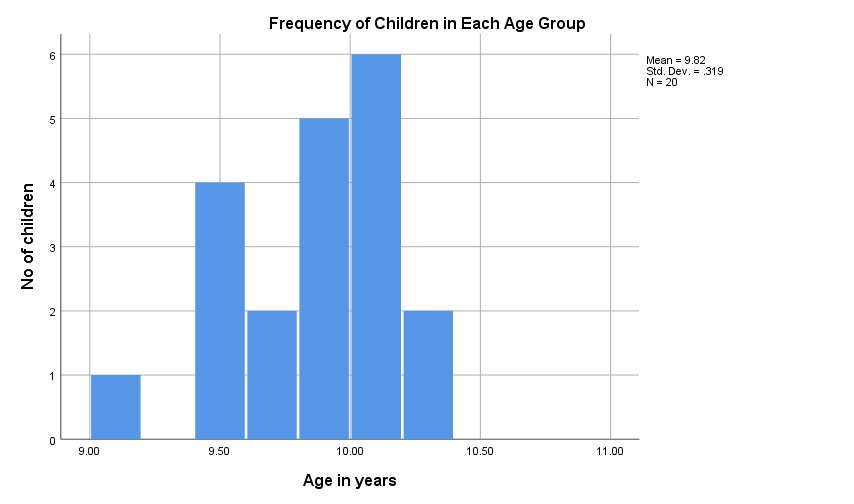
1c. Use the Frequencies command to produce a histogram of the age of the children (both Indonesian and English speaking plotted together). Edit the X‐axis label and give the chart an appropriate title

**Analyse -> Descriptive Statistics -> Frequencies**

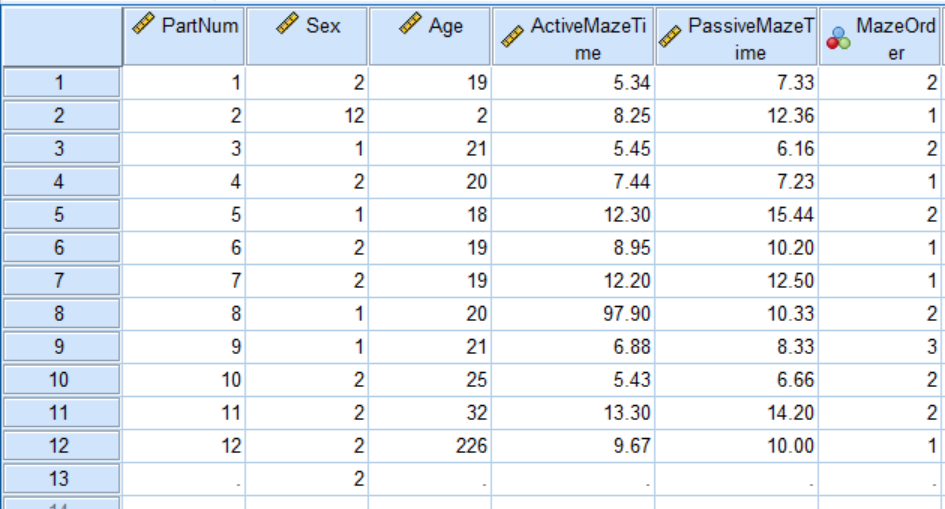
Move the Age in Years variable into the Variables box. Click on the Charts option, and select Histograms.



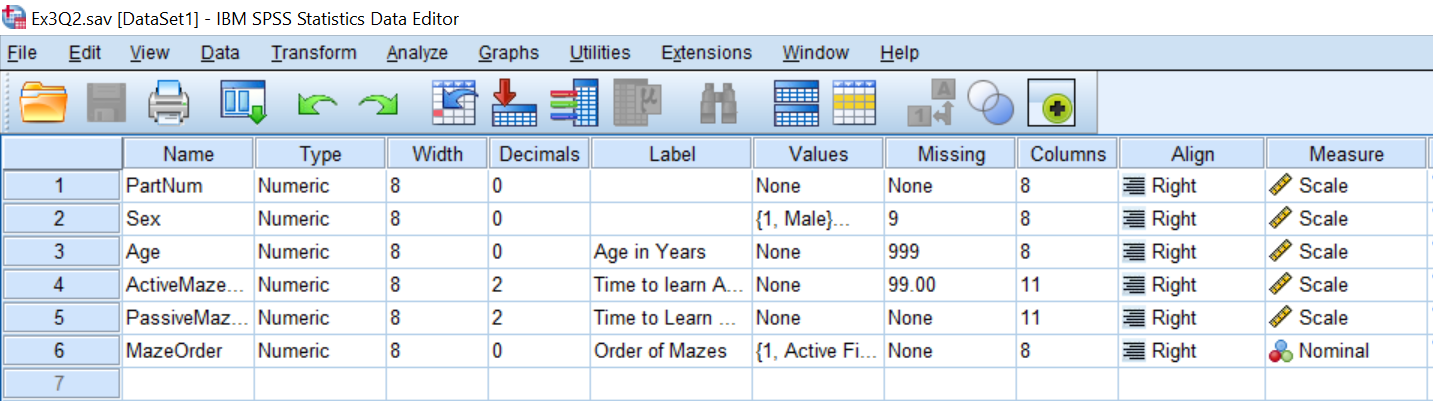
Double click on the chart in the Output window to edit it.



2a)

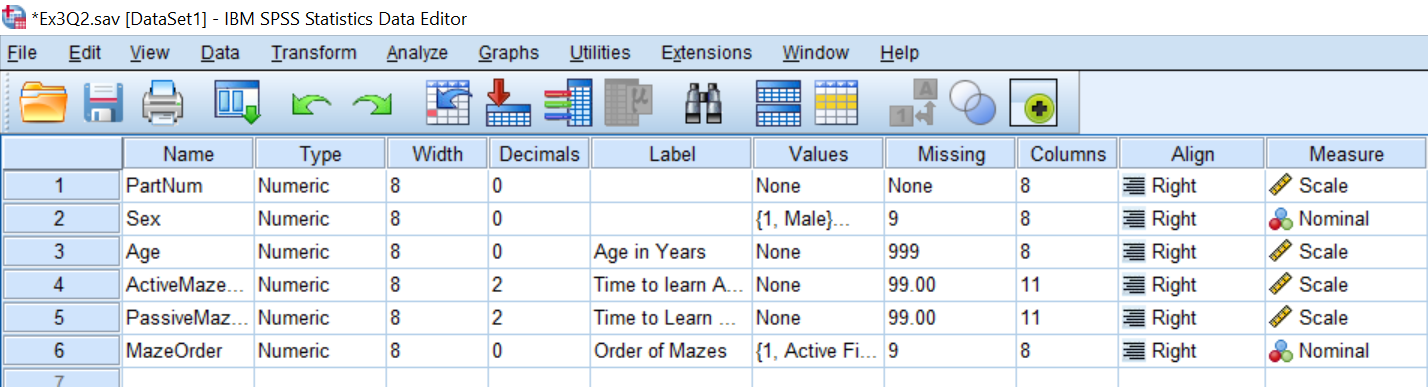


Remove extra 2 from Sex column.

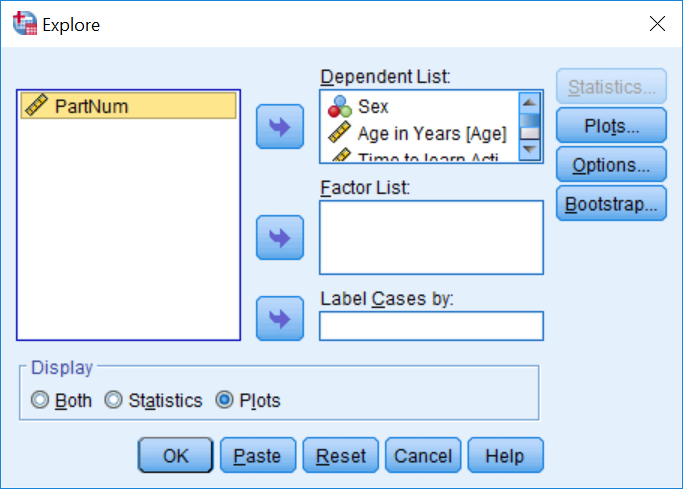


Sex is a Nominal variable, but has been set as Scale.

Missing data values are missing for the Passive Maze condition, and Maze Order. Add these in, like this:



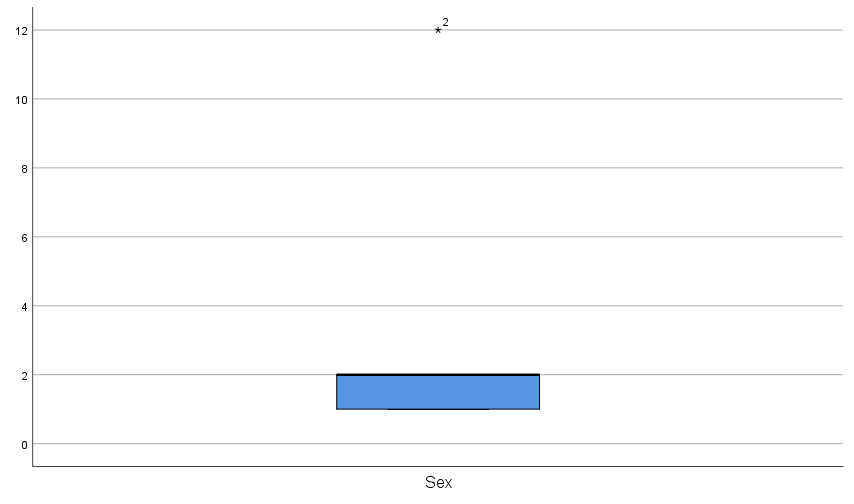
2b) As this is a small dataset, we can easily spot erroneous data…. But a good way to check for outliers more formally (or spurious data points lying outside rangers you would expect), is by producing boxplots in the explore function.

**Analyse -> Descriptive Statistics -> Explore**

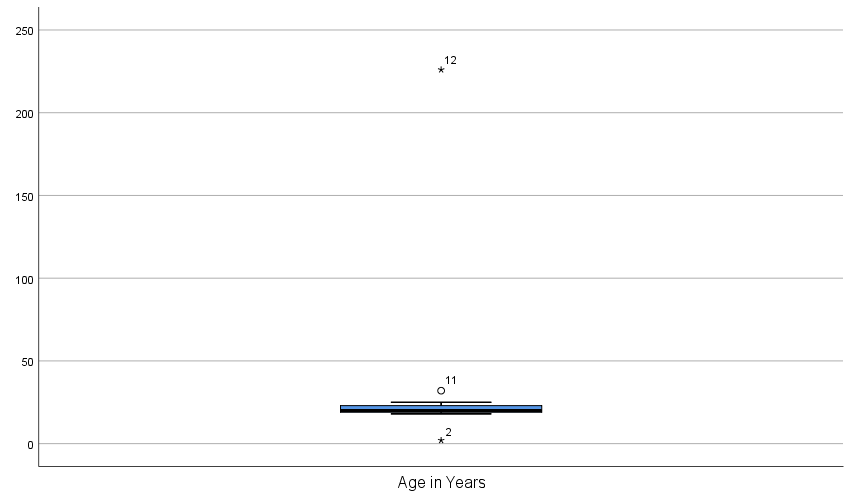
Move all of the variables across to the Dependents box.

Select Plots in the Display option, and click OK.

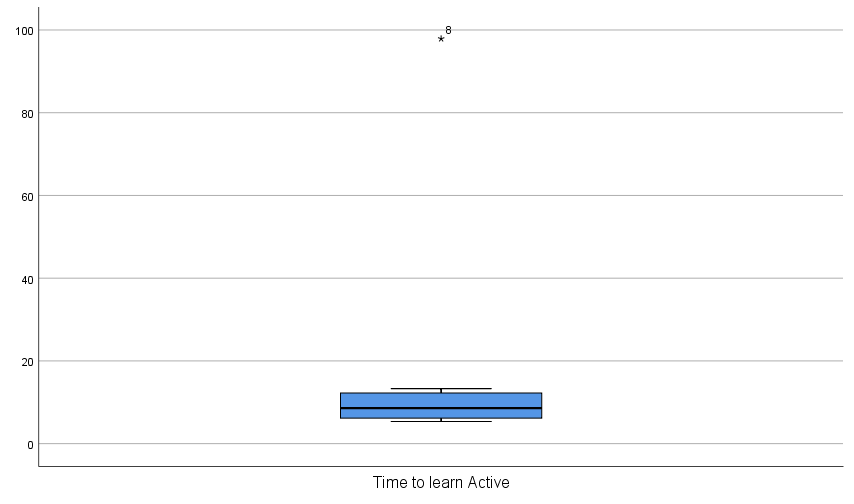
2c)

This shows you participant 2 is an outlier, with a score of 12 for sex. As these scores should only be 1 or 2, it is important to go back to your data to check what this should be, and change it accordingly.

In this case, as we have no way of knowing what this was supposed to be, code it as missing data.

Similarly, participant (or case) 12 is an outlier with an age of over 200. Again, this is unlikely, so you will need to go back to your data to check what this should be, and change it accordingly.

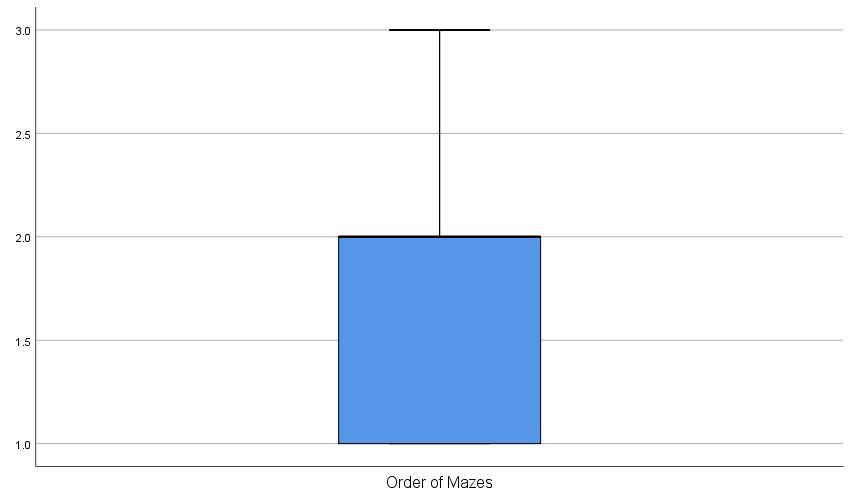
Again, in this case, as we have no way of knowing what this was supposed to be, code it as missing data.

In the Active condition, case 8 is an outlier.

In the real world… if this was a typo, you could go back to your data and replace it with the correct score.

If the time was correct, you would need to decide what to do with this case (remove it, or keep it in). Usually we would remove extreme outliers, as they are unlikely to be representative of most people…

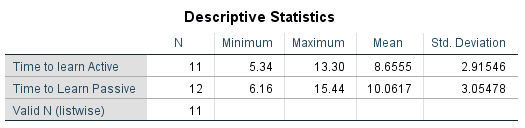
In this case, let’s recode this as missing data.

There doesn’t seem to be anything odd happening with the Time to learn in the Passive condition (boxplot not shown here).

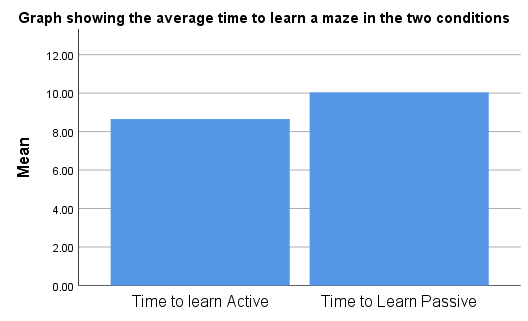
But the range of maze order scores is 1-3, when scores should only be 1 (active first) or 2 (passive first).

Again, let’s recode this to treat this as missing data.

3a) Recoding all of the outliers as missing data, and running **Analyse -> Descriptive Statistics -> Descriptives** we get the following:

Mean time to learn the maze in the active condition = 8.66

Mean time to learn the maze in the passive condition = 10.06



3b) You could draw something like this: