3.9 Exercise: Relationships between numeric variables

This exercise will enable you to construct graphs of two numeric variables as discussed in the previous video. The skills addressed are:

1. Creating a scatterplot of two numeric variables.
2. Adding a trend line (if suitable) and label any interesting points.
3. Train your eyes to observe and draw an envelope around the scatter.

[*iNZight Lite version linked here*]

INSTRUCTIONS

Follow the instructions below to generate the graphs. Or you may prefer to print these instructions. If you have a problem doing the exercise, scroll down to **Common questions**.

To begin this exercise load the **gapminder_2008** dataset into iNZight using **File > Example data ….** You will find the data set in **Module (package)** *FutureLearn*.

Creating a scatterplot of two numeric variables

Any time we graph two numeric variables, we will get a scatterplot. The variable we are mostly interested in (the **outcome variable**) goes into the **Variable 1** slot and will be on the vertical axis. The **predictor variable** will go in the **Variable 2** slot and will be on the horizontal axis.

We are interested in the relationship between the average level of CO_2 emissions and the amount of energy used per person for countries around the world.

Drag **CO2emissions** into the **Variable 1** slot and **EnergyUsePerPerson** into the **Variable 2** slot.
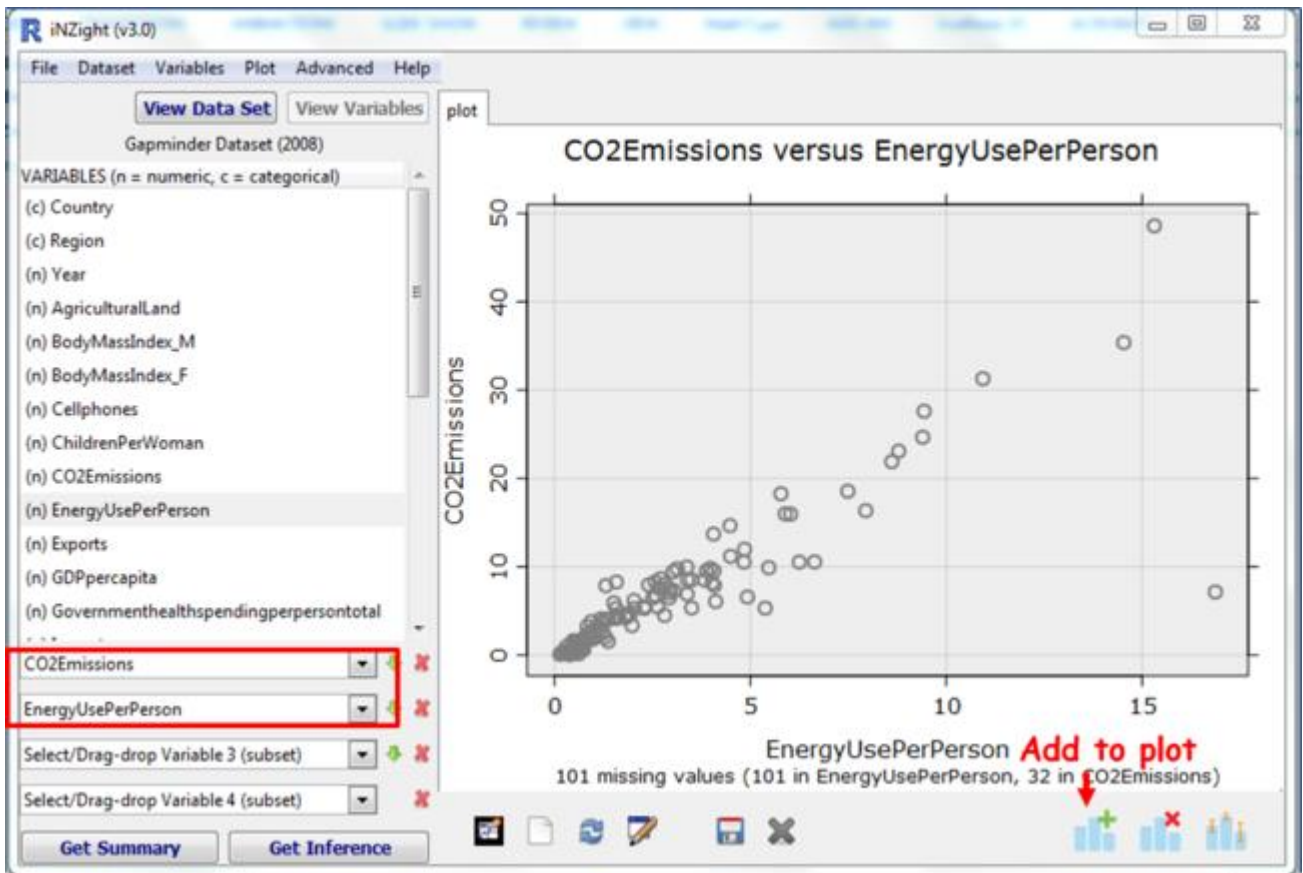
You can change your plot labels by going to **Add to Plot** and then selecting **Axes and Labels**.

A positive trend is evident in our plot. As the energy use per person increases, so do the levels of CO2 emissions.

To add a trend line:

- In **Add to Plot** choose **Trend Lines and Curves**.
- Select the type of line or curve you wish to fit to your scatterplot. In this case select a linear trend line.
- Select the colour you would like for your trend line.

A trend line will appear on your graph.
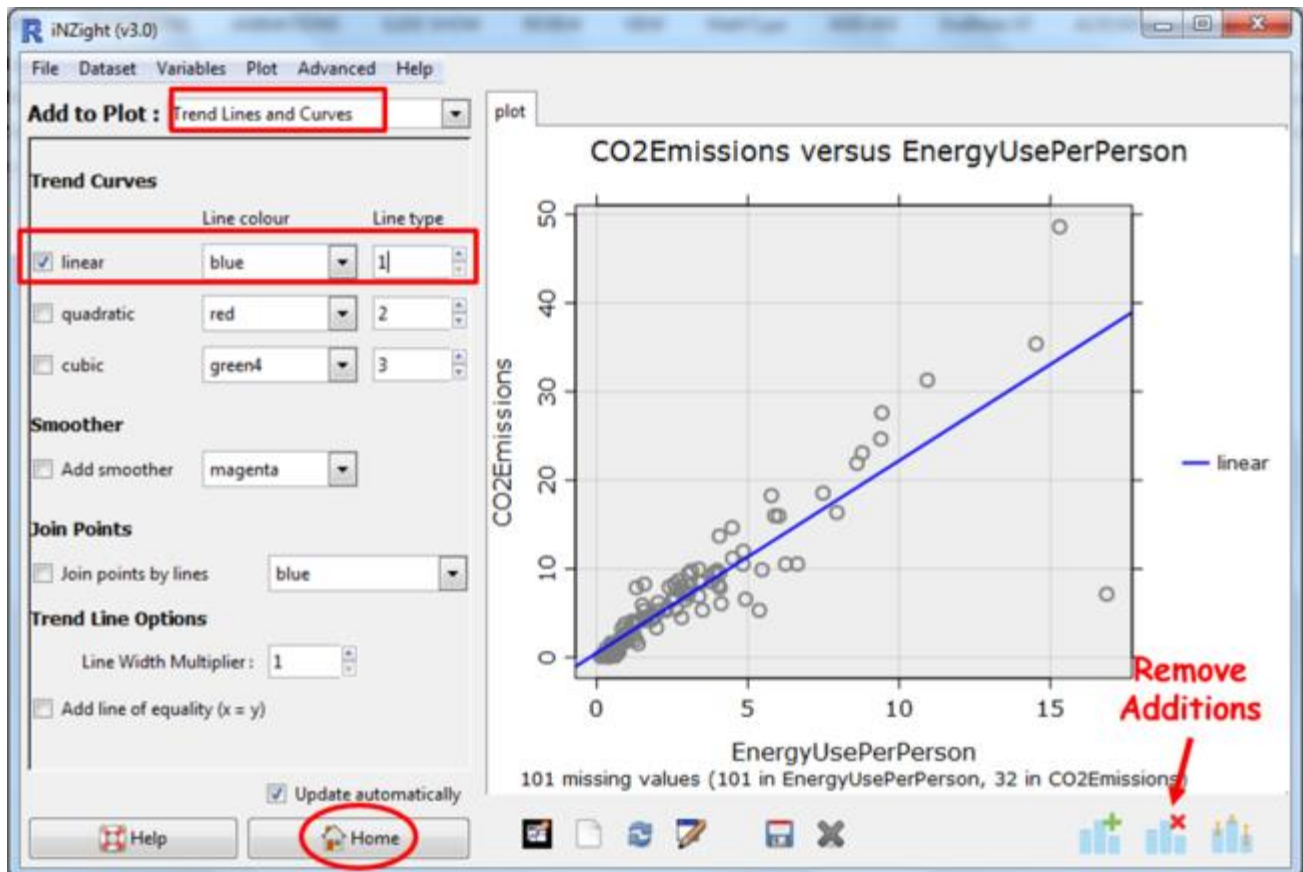
To identify interesting points:

- In **Add to Plot** choose **Identify points**.
- For **Text Labels** select the name of the variable we want to use to label the points – in this case, **Country**.
- For every point you wish to identify, first click **Locate**, and then click on the point you wish to identify on the plot.
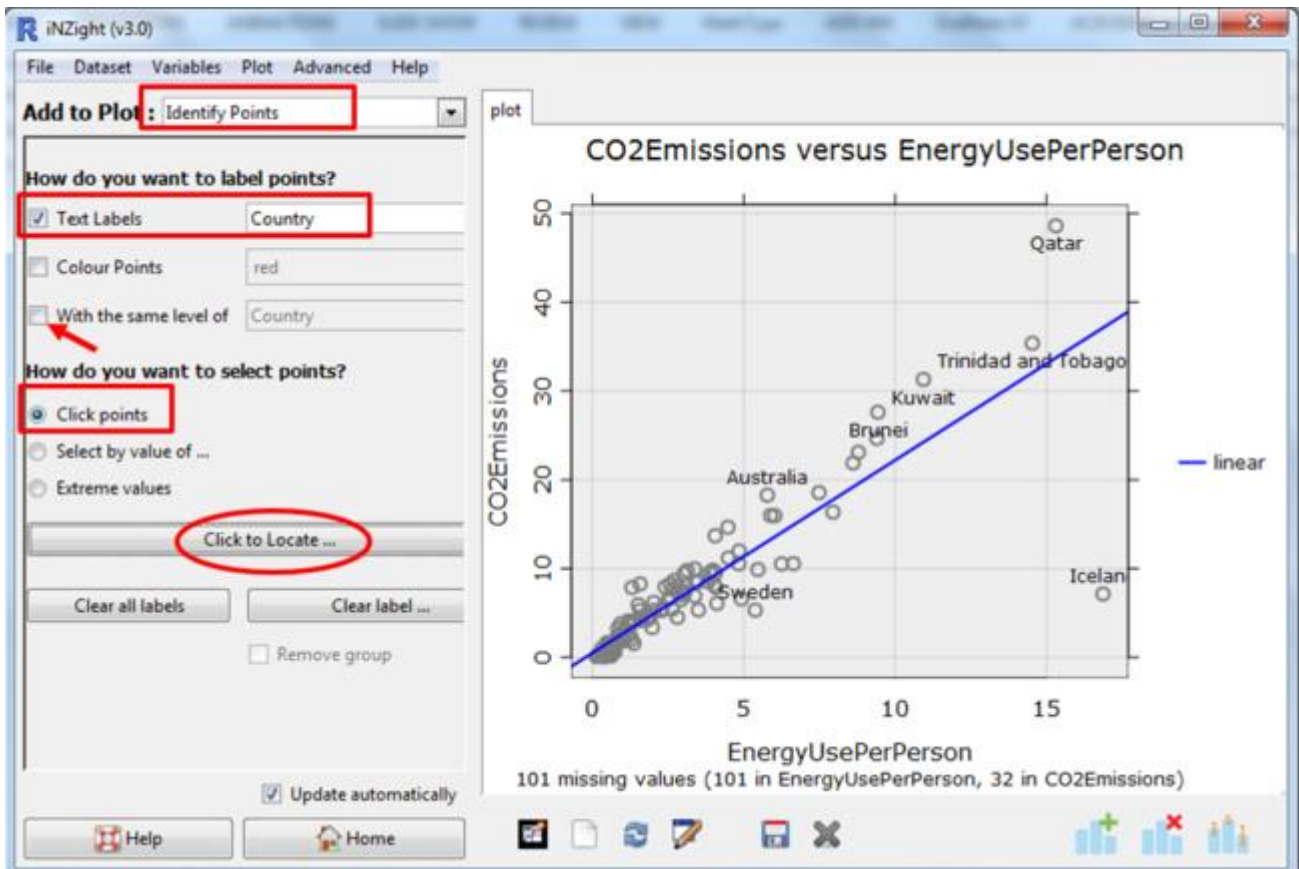
If you click **With the same level of** these same countries will also be labelled on other plots you do.

PRACTICE (~5 min)

Choose some countries that look interesting to you and identify them. Now consider other numeric variables in this dataset that may be associated with CO_2 emissions of countries around the world and explore those using scatterplots.

Have you found anything interesting? Post your findings on the discussion.

Optional: *Try this new feature* (*interactive web graphics*)

Save your graph as **File Type:** *Interactive HTML* (you will have to supply a name for the file). The file will open up in your default browser. If that is a modern browser like Chrome, Firefox or Safari (but not Internet Explorer) this will then give you an interactive version of the graph that lets you query it in various ways like hovering over the points, or a trend line, or clicking them, or selecting more than one using the Ctrl or Shift keys, or by dragging.

The save process can be slow if there are a lot of dots to be drawn.

The save window allows other variables to be exported along with the plot. This is particularly useful for hover-over if you have a variable that gives the names of the people or objects.

You can give such files to others. They do not need to be connected to iNZight to work.

Continued over …

Observations of scatterplots

Not all scatterplots show obvious trends with uniform scatter.

For each scatterplot below, train your eye to look for trends, clusters, outliers and draw an envelope around the scatter. You may want to print them out so that you can draw on the above graphs.

For each graph, what do you notice? Post one of your findings on the discussion.

Olympic Gold Medal 100m times

Head dimensions of boys

Fuel consumption rate per month
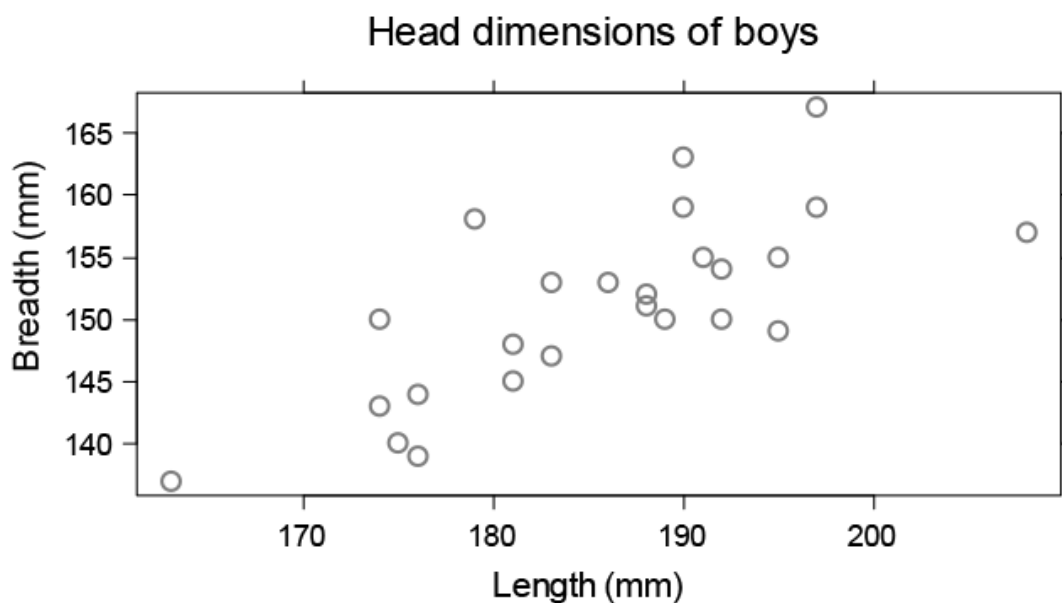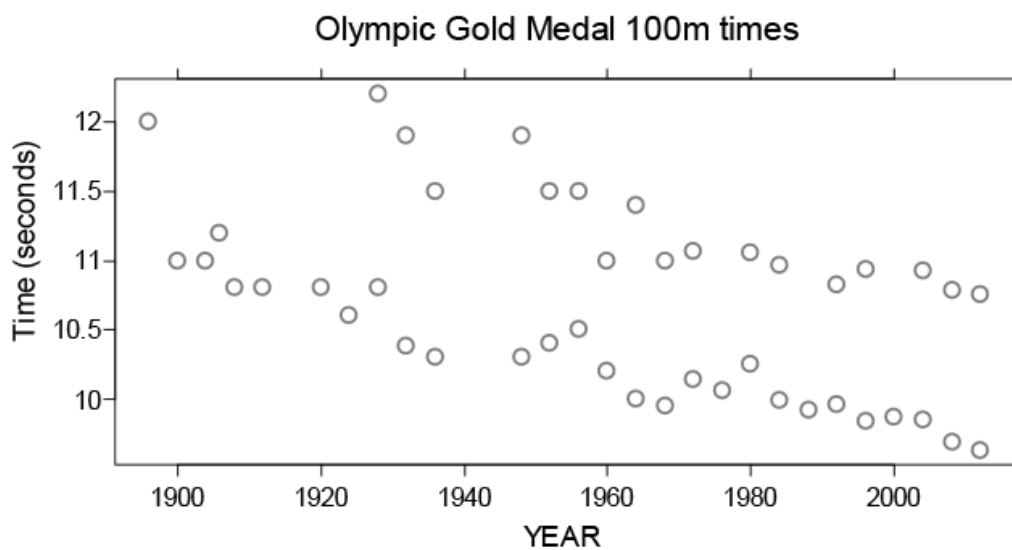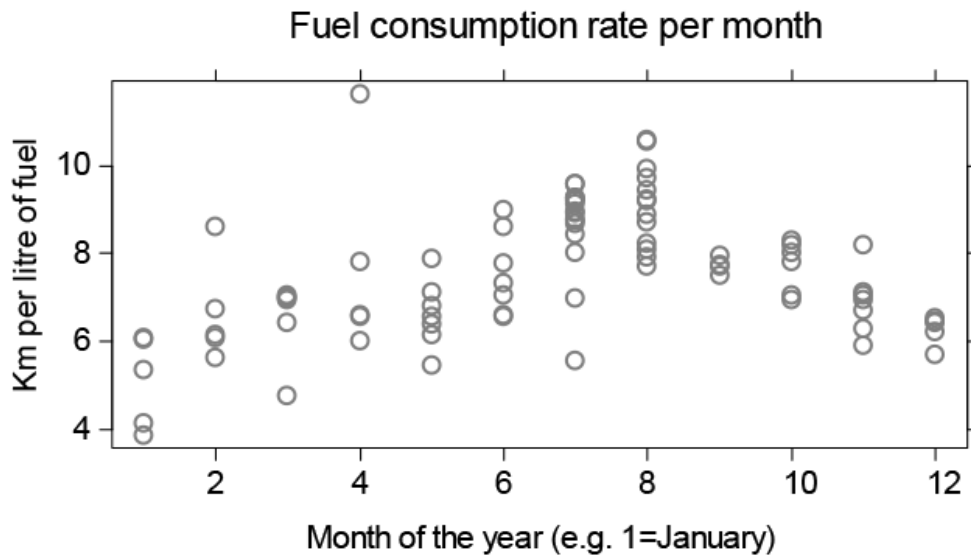
Common questions

How do I identify more than one point on my graph?
Click the locate button and then click on the point you wish to identify. Repeat this process for each point you wish to identify.

I see several ways of identifying points. Where is the help documentation?
Clicking the 'Help' button at the bottom of the **Identify points** command window takes you to an online help page.

Sometimes the data don't seem to follow a straight line, is this a trend?
That's okay. We don't expect all data to behave in a certain way. We are trying to identify any form of a relationship between two variables, whether it is a straight line, clusters of points, or even the shape of a curve. More about this next week.

Can I download the dataset used in the example here? Click here for the Olympic Gold Medal 100m times dataset.