

3.9 Exercise: Relationships between numeric variables (*iNZight Lite version*)

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.

To begin this exercise, load the dataset provided called **Gapminder-2008**.

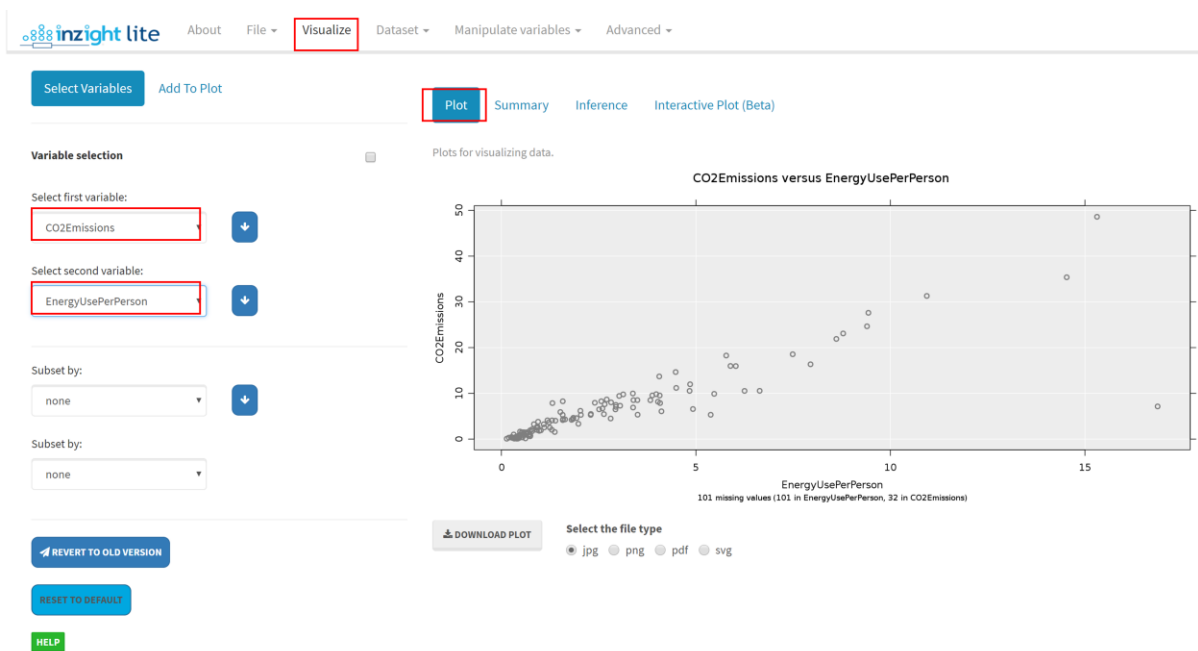
Import the **Gapminder-2008** dataset into iNZight Lite:

- Select **File > Dataset Examples**
- Select Data set category: **Future-Learn**
- Select **Gapminder-2008**
- Click on **Select Set**.

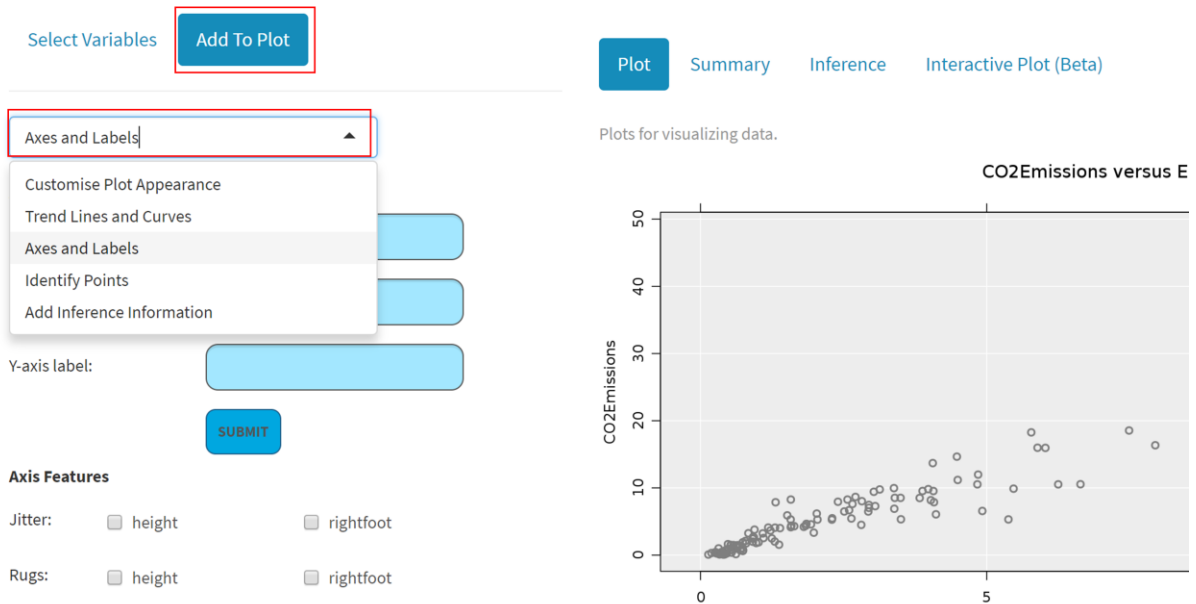
Creating a scatterplot of two numeric variables

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 **first variable** slot and will be on the vertical axis. The **predictor variable** will go in the **second variable** slot and will be on the horizontal axis.

We are interested in the relationship between the average level of CO₂ emissions and the amount of energy used per person for countries around the world. Select **CO2emissions** in the **first variable** slot and **EnergyUsePerPerson** in the **second variable** slot.



You can change your plot labels by clicking **Add to Plot** at the top left of the screen. This will display more options where you can select **Axes and labels** from the drop down box.



Adding a trend line and identifying interesting points

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:

- Click the **Add to Plot** tab
- Select **Trend Lines and Curves**.
- Check the box or the type of line or curve you wish to fit to your scatterplot. In this case we are fitting a linear trend line.
- Select the colour you would like for your trend line.

A trend line will appear on your graph.

Select Variables **Add To Plot**

Trend Lines and Curves

Trend Curves

linear

quadratic

cubic

Line Width Multiplier:

Smoother

Add smoother

Join points

Join points

Trend Line Options

Add y=x line

Line colour

blue

Line type

solid

red

solid

green4

solid

1

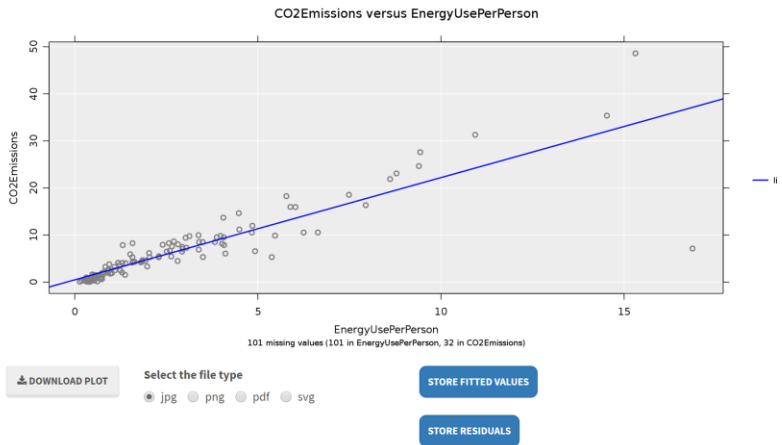
magenta

blue

black

Plot Summary Inference Interactive Plot (Beta)

Plots for visualizing data.



To identify interesting points:

- Select Option: **identify points**.
- Click on **Text label** and select the variable **Country**
- Select method of selection: **Extremes**
- Use the slider to adjust the number of points with labels.

Select Variables **Add To Plot**

Identify Points

How do you want to label points?

Text label

Country

Colour

With the same level of

Select method of selection

Select by value

Extremes

Range of values

Number of points

0 25 225

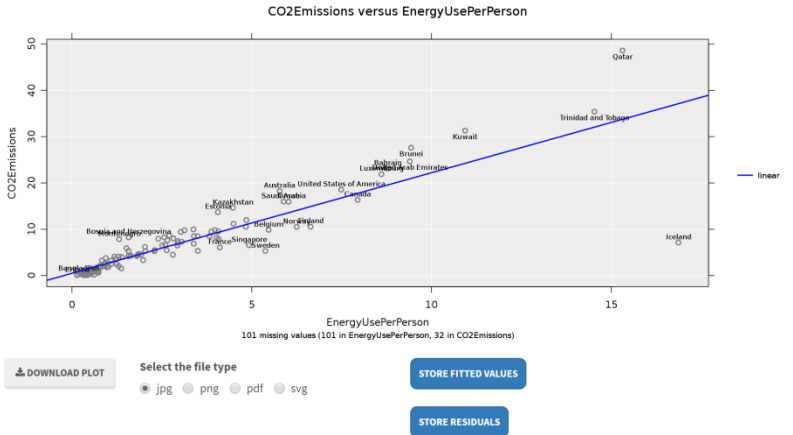
Show stored

STORE SELECTED

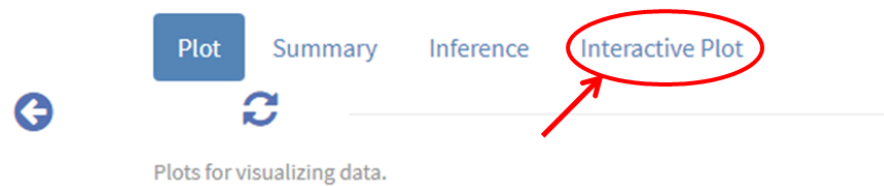
FORGET STORED

Plot Summary Inference Interactive Plot (Beta)

Plots for visualizing data.



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



Click on the **Interactive Plot** tab. This will give you an interactive version of your 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. You have to Click **Produce Plot** for it to appear. This is a protection as the conversion process can be slow if there are many points to be drawn.

A click button (**Select additional variables**) *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 plotted.

You can download these plots as Interactive HTML files which you can give to others. They do not need to be connected to iNZight Lite to work.

PRACTICE (~5 min)

Identify some points that look interesting to you and identify them. Now consider other numeric variables in this dataset that may be associated with CO₂ emissions of countries around the world and explore those using scatterplots.

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

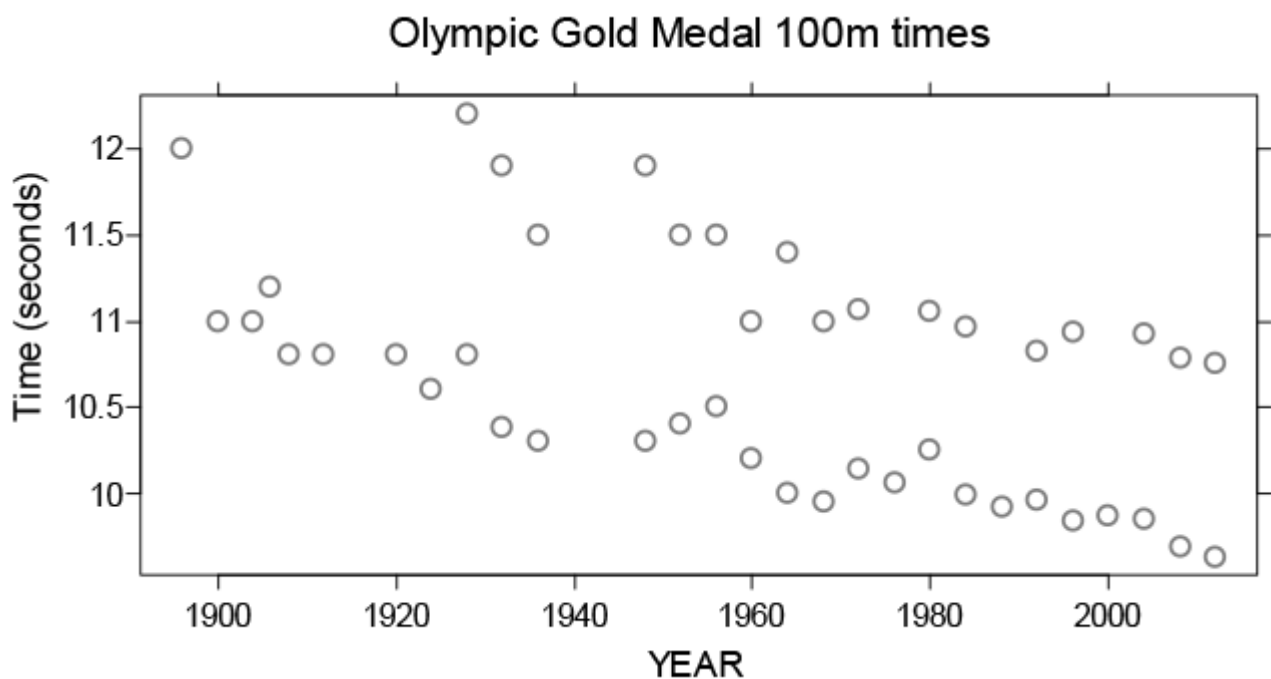
Continued over ...

Observations of scatterplots

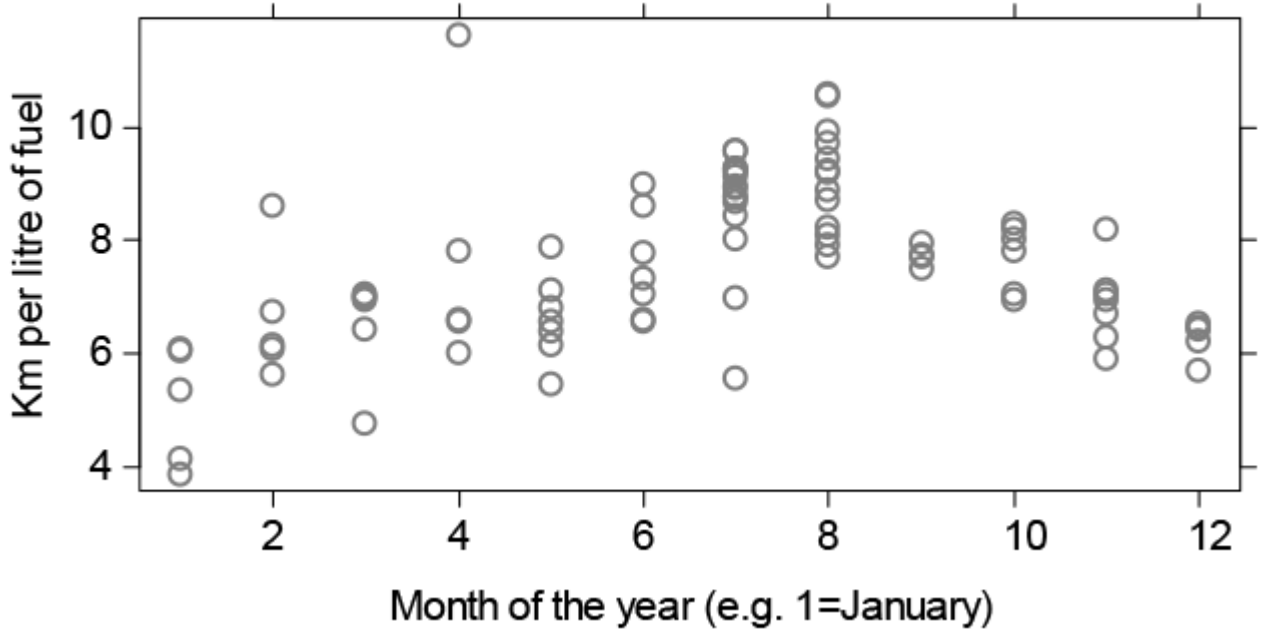
Not all scatterplots show obvious trends with uniform scatter.

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 graphs.

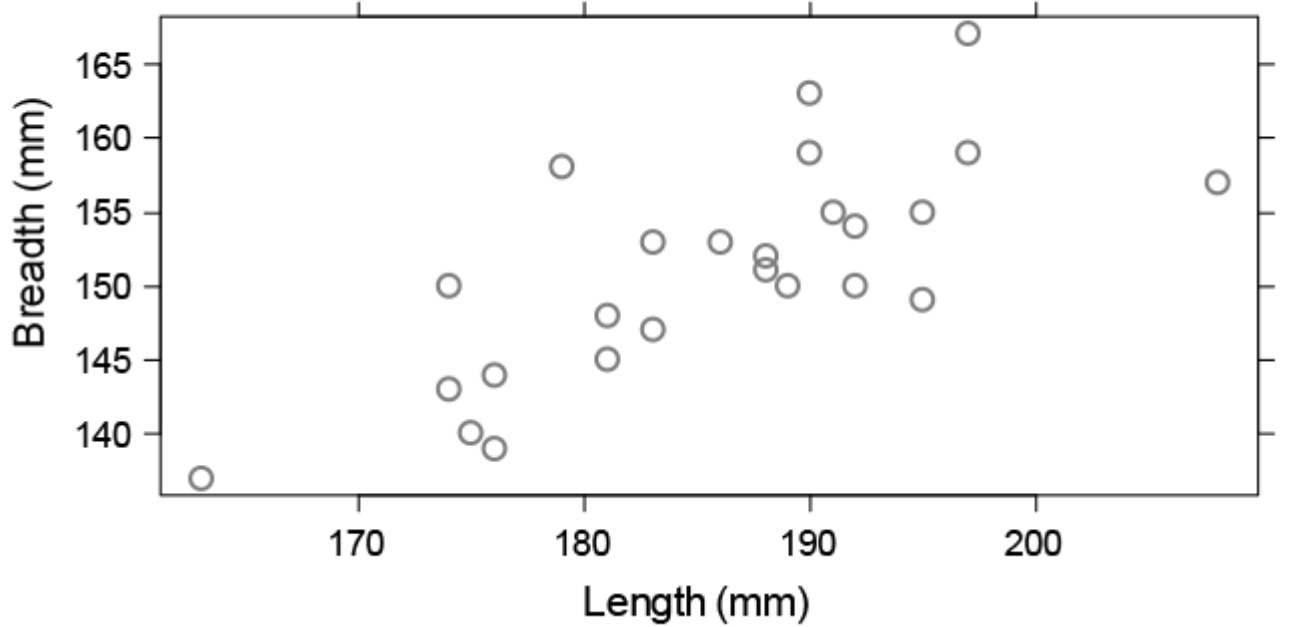
For each graph, what do you notice? Post one of your findings on the discussion.



Fuel consumption rate per month



Head dimensions of boys



Common questions

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.