

4.11 Exercise: Advanced scatterplots for deeper analysis

This exercise will enable you to use iNZight to explore more complicated relationships between variables and the explore the effects of a third and fourth variable, enabling you to view changes over time.

The skills addressed are:

1. Create a scatterplot of two numeric variables, subset by a 3rd variable.
2. Explore the effect of a third and fourth variable using colour and size.
3. Extract and explore changes in individual groups within a categorical variable over time.

[iNZight Lite version [linked here](#)]

INSTRUCTIONS

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

[If you are using a Windows machine you need to be in **dual windows mode** for this. See how in **Common questions** below.]

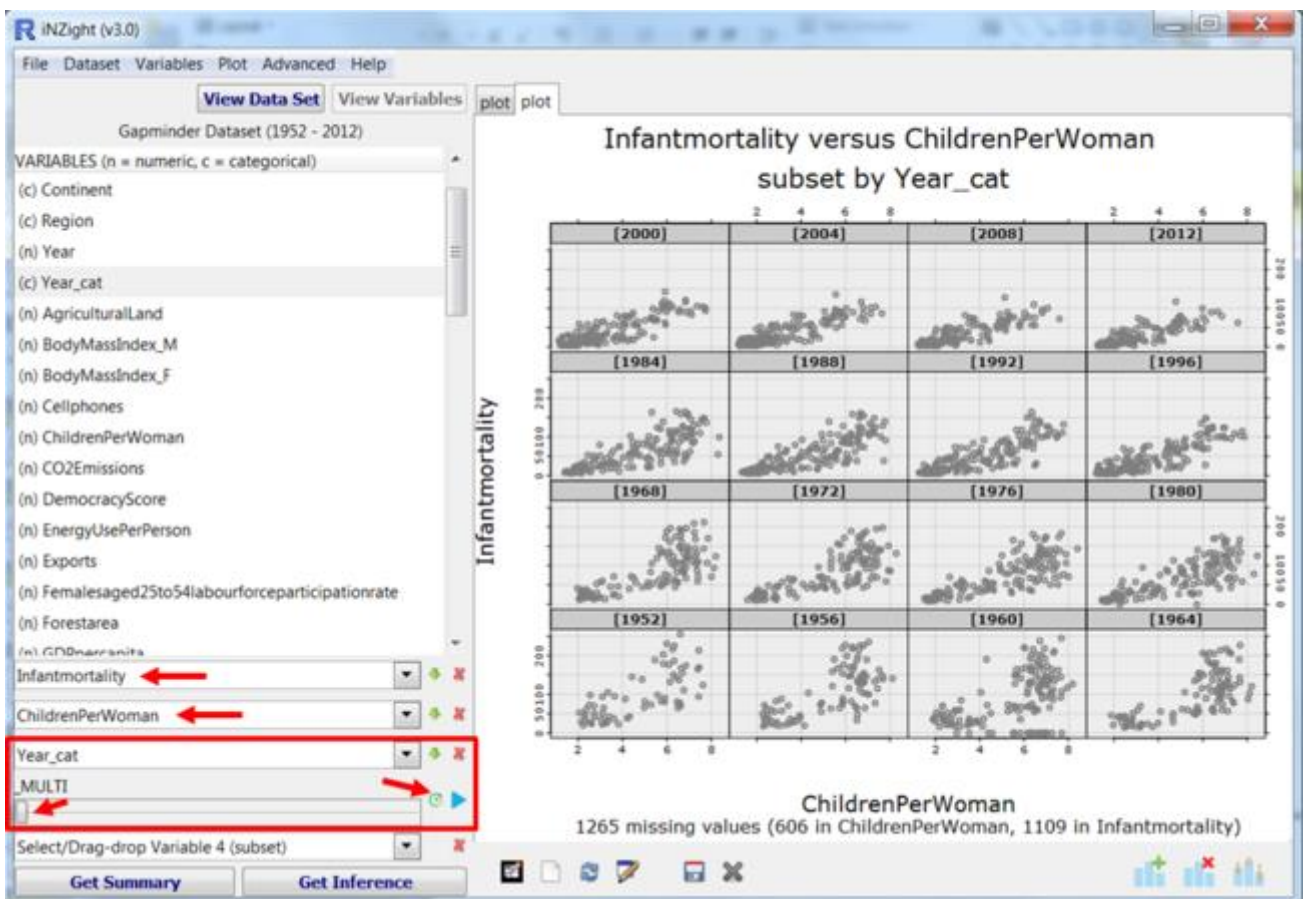
Load the **gapminder** dataset (but **not** *gapminder_2008*) into iNZight using **File > Example data ...**. You will find the data set in **Module (package) FutureLearn**.

Create a scatterplot of two numeric variables, subset by a 3rd variable

We are going to explore the relationship between the variables **Infantmortality** and **ChildrenPerWoman** of countries in the **Gapminder** dataset.

Construct a scatterplot of **Infantmortality** (outcome variable) and **ChildrenPerWoman** (predictor variable).

Drag **Year_cat** into **Variable 3 (subset)**. Year_cat is a categorical version of Year. [This way we get separate graphs for every subset formed by a value of the variable whereas iNZight cuts a numeric variable like Year up into just 4 class intervals.]



Use the slider or push **play** to watch the changes over years. [Speed Tip: Click the slider handle and use the left and right-arrow keys to move it.]

Do you see anything interesting in your plots that might be worth investigating further?

You have done the basics already at this stage. You can do a lot of useful exploration just using the variable drop slots and playing over a 3rd or 4th variable. For those who want to proceed, we are now going to dig in more and add a lot more information but it will be a bit trickier.

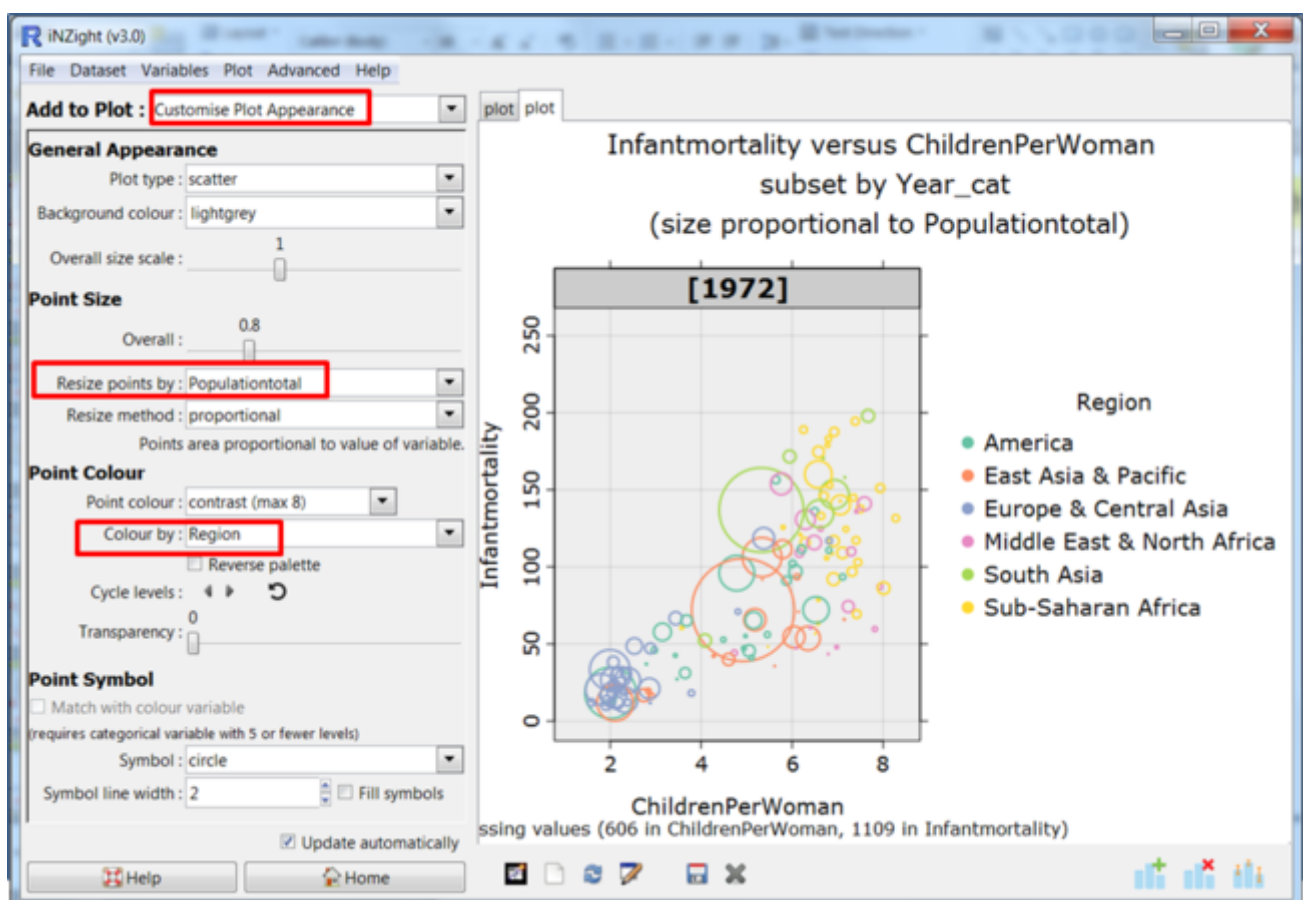
Add more information using colour and size

We suspect that there is a lot more information that can be graphed to give us more insight into the rate of infant mortality and the number of children per woman.

- How does the graph change over time?
- What is happening in larger populations?
- Do different regions have similarities?
- Can we track particular countries over time?

You may want to resize your plot window to get a larger graph. Also drag to a particular Year, say 1972 so that the changes you will see below will be more obvious.

We go into **Add to Plot** and colour by **Region** and size the points by **Populationtotal**.

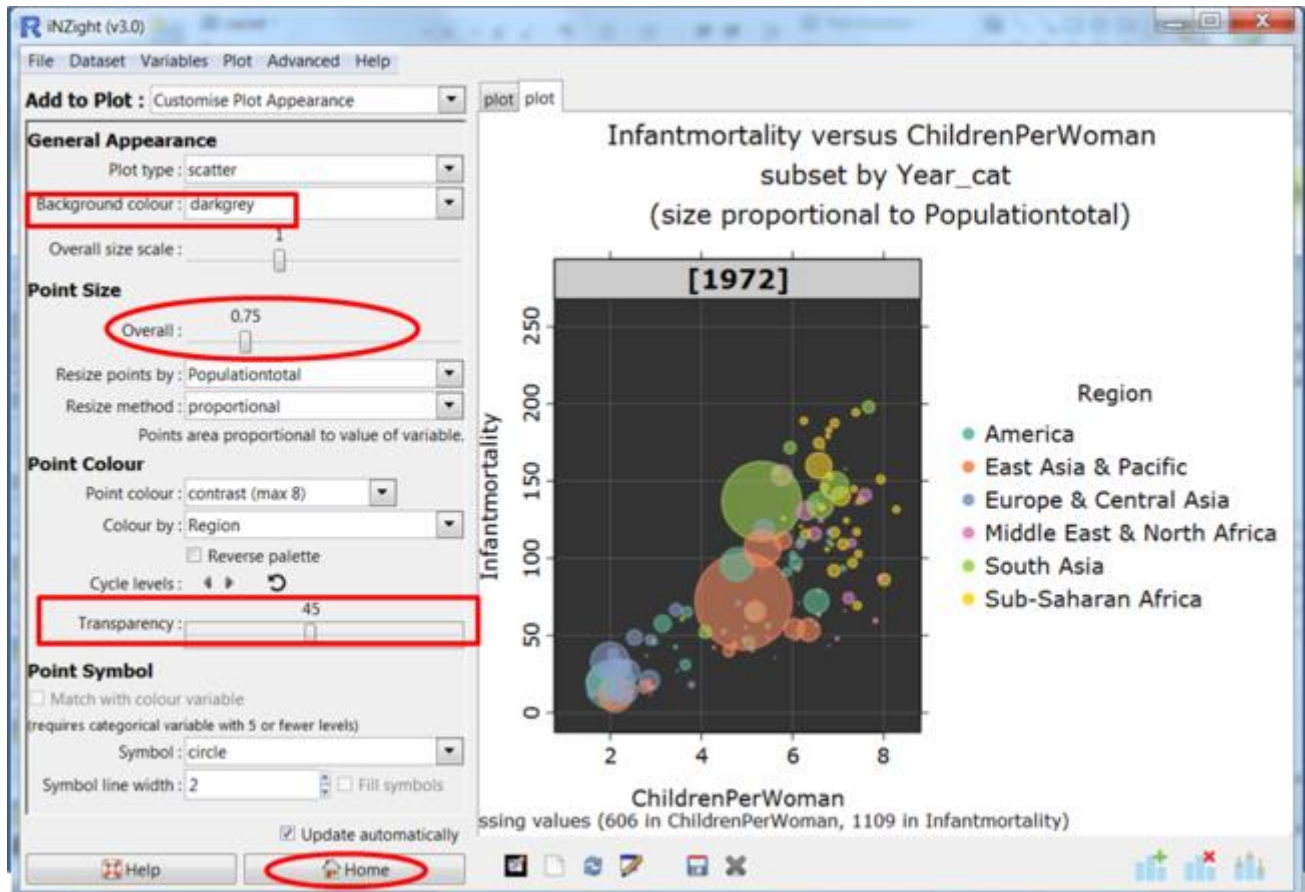


There are problems with this, e.g., the lighter colours are hard to see.

We will now make some changes. Change the **Background colour** to *darkgrey*. How does this help?

Move the **Transparency** slider to about 45%. What does this do?

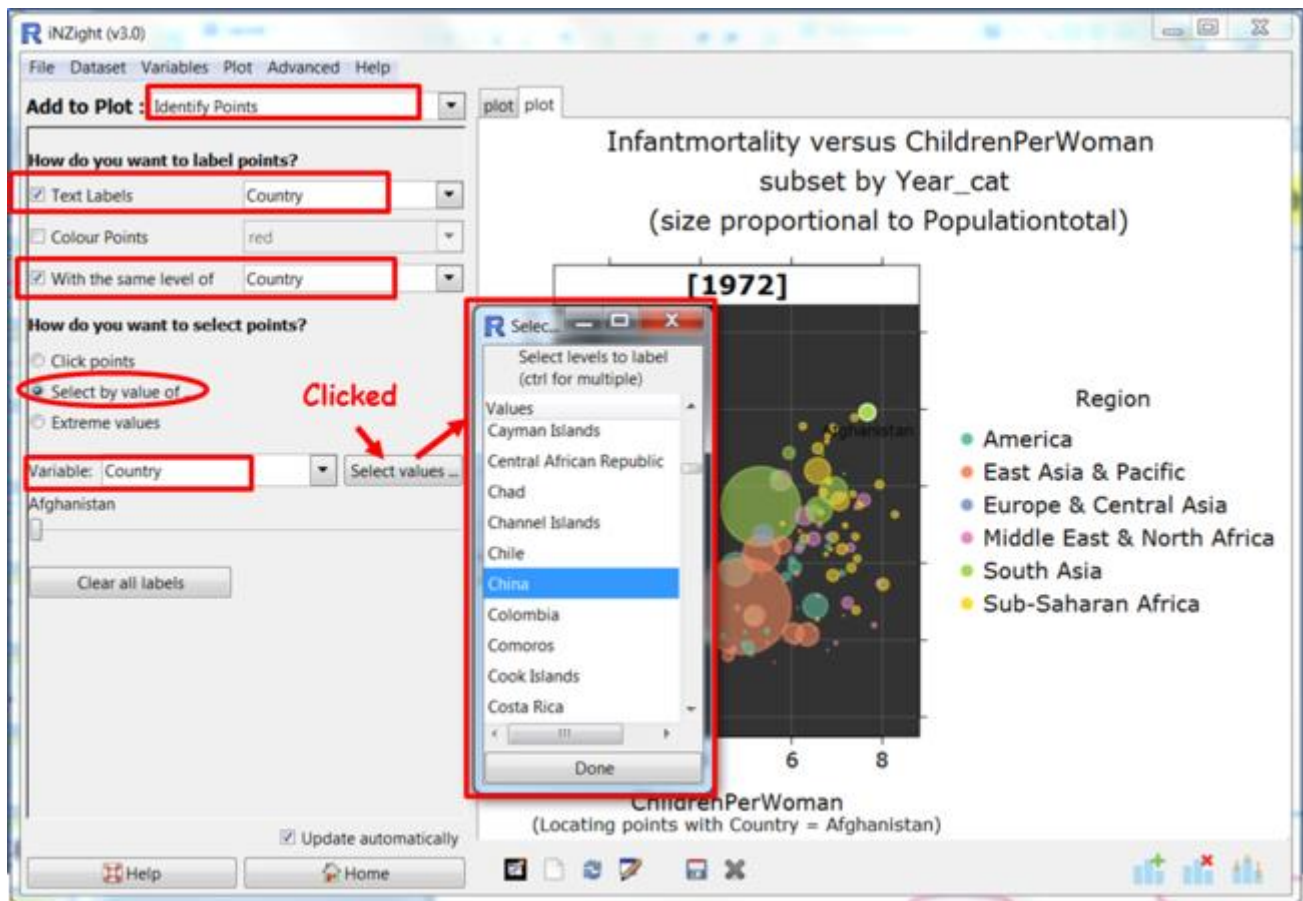
Reduce the overall **Point size**.



- Go back to the **Home** command panel and explore changes over *Year_cat* with this version.

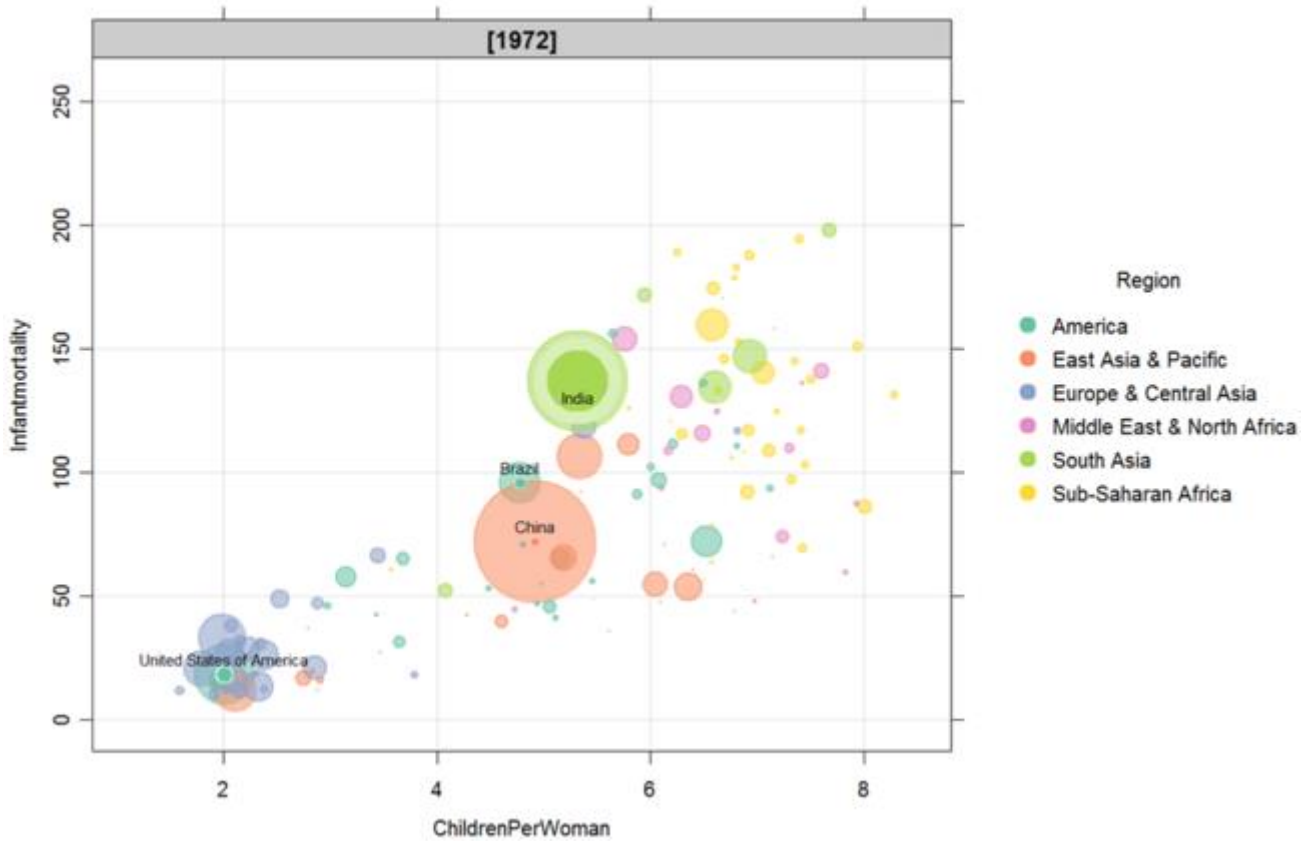
It would be really nice to add a little information about which country is which.

- On **Add to Plot** go to **Identify points**.
- For **Text Labels** select **Country**.
We could click on points as we did in the past, but here it is easier to:
- Click **Select by value of** and for **Variable** choose **_Country_**, then click on **Select values** and a selection list pops up. I've used that to chose Brazil, China, India and USA.



There was a problem of black labels being hard to see against a dark grey background so I changed the **Background colour** to *white*.

Infantmortality versus ChildrenPerWoman subset by Year_cat (size proportional to Populationtotal)



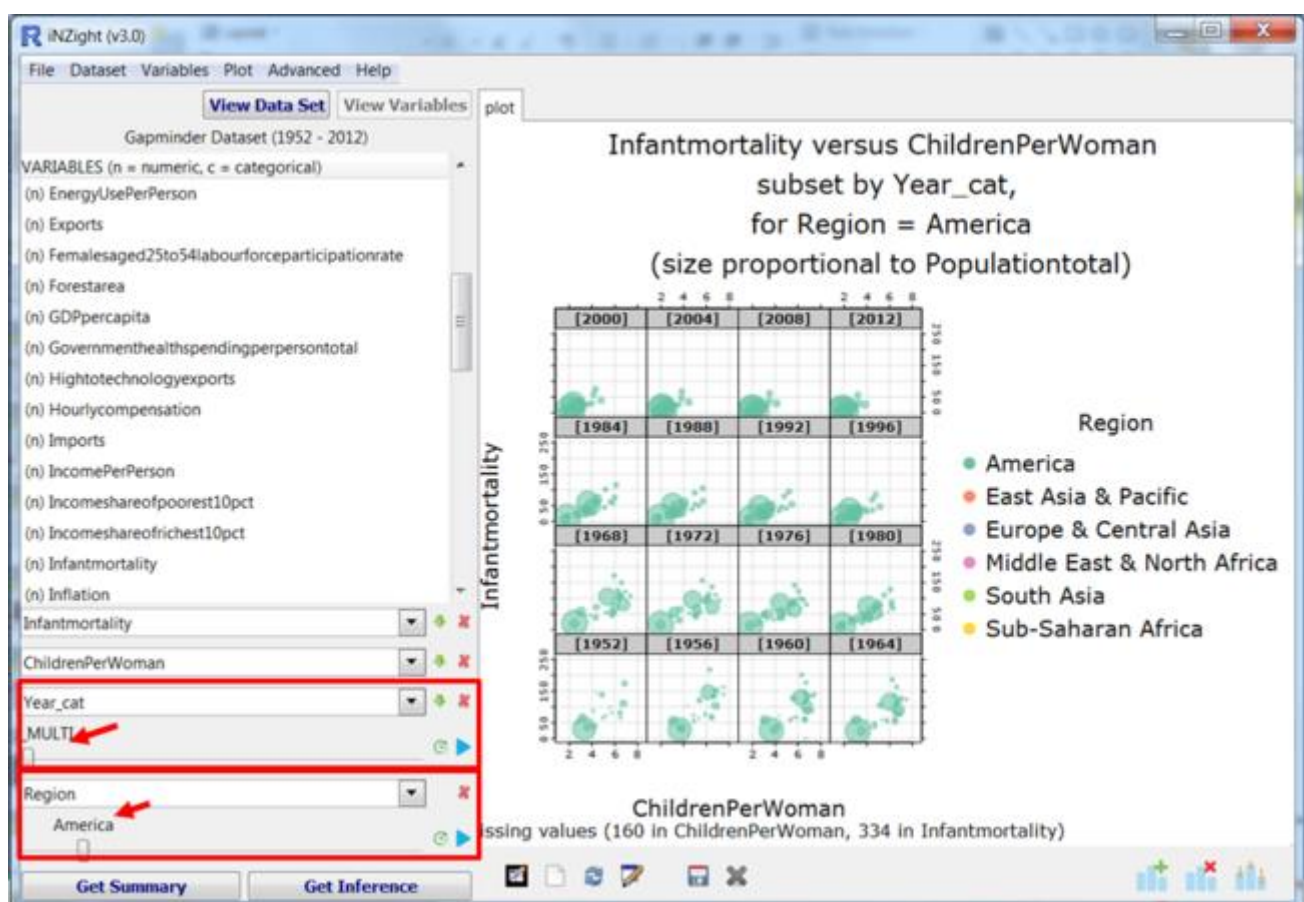
- Go back **Home** and explore changes over Year_cat with this version.
- What are the most interesting changes you see?
- How else would you like to investigate this?

Extract and explore changes in individual groups within a categorical variable over time

Sometimes we may just want to look at the changes in each **Region** separately. To do this:

- Pull down the variable name, **Region** into the **Variable 4 (Subset)** slot.
- Move the slider for **Region** onto the region you are interested in.
- Move the slider for **Year_cat** to **MULTI** or push play.

Here is a tiled plot of the **Region America**:



What can you see on your series of plots about the changes in the rate of infant mortality (deaths per 1000 live births) and the number of children per woman?

Look for something interesting in the data, post a comment and read what other people have discovered.

Move both sliders to **MULTI**. (Left hand side for the first subset slot, Right hand side for the second subset slot) . To see the result well, ask for a **New graphics Window** and maximize it (full screen).

What happens?

Optional

If you want to play with a more recent version of the data download [Gapminder_1952-2016_by_4yrs.csv](https://www.stat.auckland.ac.nz/~wild/data/gapminder_datasets/) from https://www.stat.auckland.ac.nz/~wild/data/gapminder_datasets/. [Some variable names may differ somewhat from those in the Exercise above.]

Common questions

How do I put iNZight into dual window mode?

This is the preferred mode if you are working in Windows. Go **File > Preferences** and click **Use dual-window display mode** and **Save**. It doesn't take effect until you shut down iNZight and start it up again.

How can I edit iNZight graphics files saved from iNZight?

If you can use graphics software such as Adobe Illustrator (professional), Affinity Designer (inexpensive) or the free alternative Inkscape, you can use these to import pdf files saved from iNZight and edit the graphics in the graphics program.

Windows users can also edit iNZight graphics in PowerPoint.

If you are not in dual window mode, then click on the **New graphics window** icon (or go **_ Plot > New Plot Window**) This will give you another copy of your graph in a separate graphics window. In this separate window, right-click on the plot and select **Copy as a Metafile**.

Then you can **Paste special** into PowerPoint as an *Enhanced Metafile*. Right-click on the plot and **Ungroup** until it won't ungroup any more. At this stage the plot is composed of ordinary PowerPoint objects that are easy to change. [Please note, however, this is currently *not working with dot plots* (works for other types of graph). It is being caused by double plotting to figure out the sizes and positions of the dots for stacking. We still have not figured out a solution.]