

4.7 Exercise: Techniques for scatterplots

– R version

Note: Copying and pasting text (e.g. R code) from a pdf is not reliable. For that reason we have also provided the code in [a text file](#)

This exercise will enable you to become more proficient in creating scatterplots with iNZightPlot. You will learn how to apply the most suitable trend line and use techniques to overcome perceptual problems.

The skills addressed are:

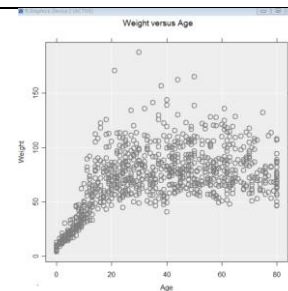
1. Create a scatterplot of two numeric variables and apply a suitable trend line.
2. Use techniques such as jittering, transparency and running quartiles to deal with overprinting.

We are going to explore the relationship between variables **Age** and **Weight** of people in the **nhanes_1000** dataset in the **FutureLearnData** package.

# R code	Output and/or Commentary
<pre># Setup library(iNZightPlots) library(FutureLearnData) data(nhanes_1000)</pre>	

Plot Weight vs Age

```
iNZightPlot(Age, Weight, data=nhanes_1000)
```



Add a trend line

```
iNZightPlot(Age, Weight, data=nhanes_1000, trend="linear")
```



Get some summary information (result depends on trends fitted)

```
getPlotSummary(Age, Weight, data=nhanes_1000, trend="linear")
```

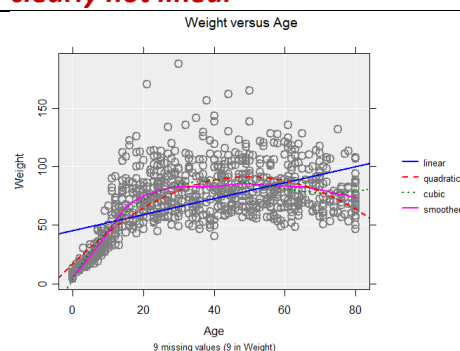
```
getPlotSummary(Age, Weight, data=nhanes_1000, trend="linear",  
summary.type="inference")
```

```
> getPlotSummary(Age, Weight, data=nhanes_1000, trend="linear")
iNZight Summary
-----
Response/outcome variable: Weight (numeric)
Predictor/explanatory variable: Age (numeric)
Total number of observations: 1000
Number omitted due to missingness: 9 (9 in Weight)
Total number of observations used: 991
-----
Summary of Weight versus Age:
-----
Linear trend:
  Weight = 45.17 + 0.6527 * Age
  Linear correlation: 0.53
Rank correlation: 0.52 (using Spearman's Rank Correlation)
-----
```

The inference results make no sense here as the real trend is clearly not linear

Add more trend curves and a smoother

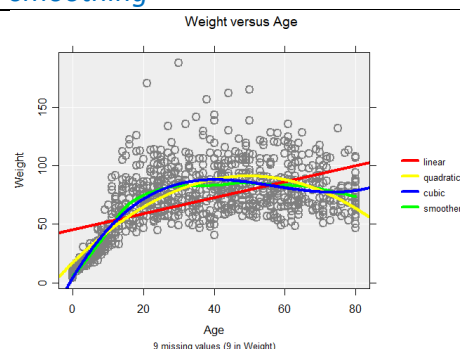
```
iNZightPlot(Age, Weight, data=nhanes_1000, trend=c("linear",  
"quadratic", "cubic"), smooth=.25)
```



The value given for the smooth (between 0 and 1) controls the level of smoothing. Bigger numbers correspond to more smoothing

Make all the lines thicker, all solid lines, and change line colours

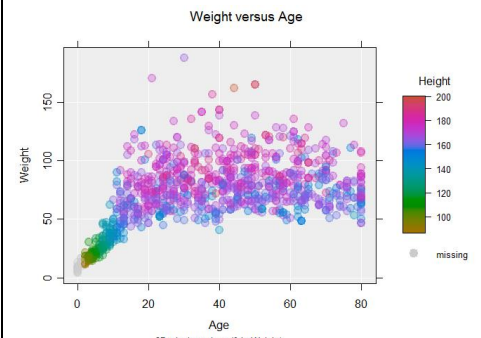
```
iNZightPlot(Age, Weight, data=nhanes_1000, trend=c("linear",  
"quadratic", "cubic"), smooth=.25, lwd=2,  
lty.trend=list(linear=1, quadratic=1, cubic=1),  
col.trend=list(linear="red", quadratic="yellow", cubic="blue"),  
col.smooth="green")
```



Scatter plot coloured by Height and with Transparency

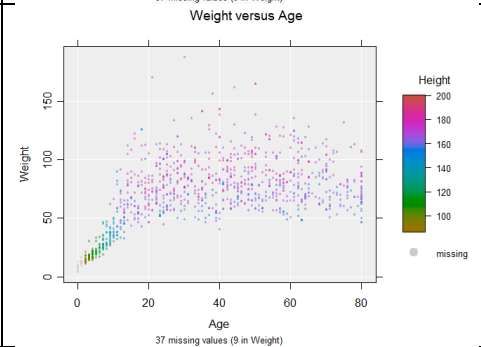
```
iNZightPlot(Age, Weight, data=nhanes_1000, colby=Height, alpha=.3)
```

alpha (0 to 1) controls transparency. Smaller for more transparent



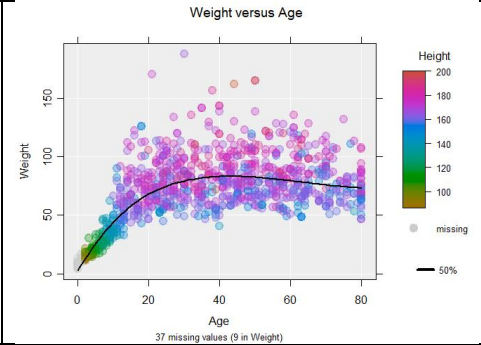
Make the points smaller

```
iNZightPlot(Age, Weight, data=nhanes_1000, colby=Height, alpha=.3,
  cex.pt=.2)
```



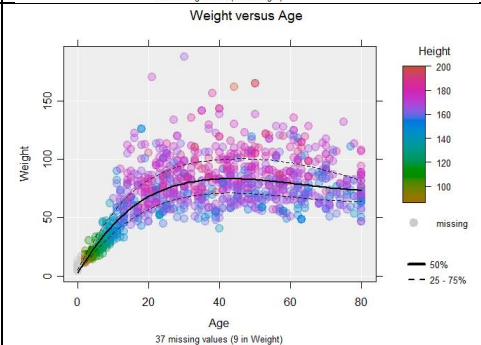
Create a median smoother in black

```
iNZightPlot(Age, Weight, data=nhanes_1000, colby=Height, alpha=.3,
  quant.smooth=.5, col.smooth="black")
```

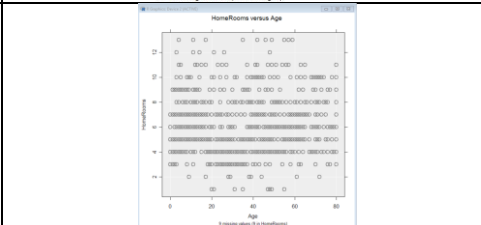


Create a median and quartile smoothers all in black

```
iNZightPlot(Age, Weight, data=nhanes_1000, colby=Height, alpha=.3,
  quant.smooth=c(.25,.5,.75), col.smooth="black")
```

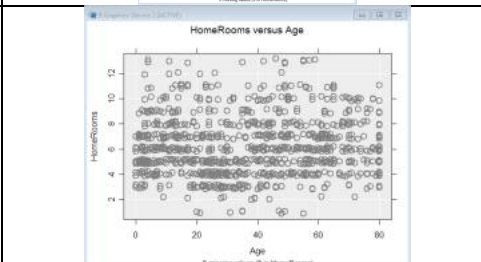


```
iNZightPlot(Age, HomeRooms, data=nhanes_1000)
```



Jitter in the vertical (y) direction

```
iNZightPlot(Age, HomeRooms, data=nhanes_1000, jitter="y")
```



- **Play some more with these settings and try other variables**
- For even more settings, type **?inzpar** into R to get help on the **inzpar**, or type **inzpar** to just get a complete list (last time I looked the help file wasn't entirely complete)

Optional: Try this new feature (interactive web graphics)

We will export an iNZightPlot graph as an *Interactive HTML* file and open this file up in our 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.

Here is sample code:

(works only with single graphs and not with a panel of several graphs)

Make a plot and also store the output in myplot

```
myplot = iNZightPlot(Age, Weight, data=nhanes_1000, trend=c("linear", "quadratic"))
```

Specify a location to store an Interactive HTML file. I will call my file "myintplot.html"

You will have to change the path to the file because this one is to a location on my desktop!

```
filepath = "C:/Users/myusername/Desktop/myintplot.html"
```

```
exportHTML(myplot, filepath)
```

```
browseURL(filepath) #open the file up in my default browser
```

If the plot is truncated make your R plotting window smaller

To add extra variables, e.g. Country ...

```
exportHTML(myplot, filepath, data=nhanes_1000, extra.vars = c("ID", "Gender"))
```

To discuss issues related to this Exercise,

go to <https://gitter.im/iNZightVIT/d2i-R-discussion>

*To be able to post to the list you will have to set up a (free) account on **Github***

<https://github.com/login>

If your question relates to an Exercise, say which one you are talking about!