

Marie Marshall

9/16/2014

Guide To Using Base R

What is R?

R is a free software environment for statistical computing and graphics. R provides a wide variety of statistical (linear and nonlinear modelling, classical statistical tests, time-series analysis, classification, clustering, ...) and graphical techniques.

How to Use R

1) Choose Output Options

R Markdown documents can contain a metadata section that includes both title, author, and date information as well as options for customizing output. Also, R Markdown has built in support for several output formats (HTML, PDF, and MS Word documents as well as Beamer presentations)

2) Initialize The libraries

Add the external community-developed libraries you will be analyzing

3) Determine Data Type/ Simulate Data

There are different types of variables: Factor and Numerical

Factor data

Consists of names representing categories. Example: The gender categories (male & female) of the people where you work or go to school; or the make of cars in the parking lot (categories of Ford, GM, Toyota, Mazda, KIA, etc) is factor data.

Numerical data

Consists of numbers that represent counts or measurements. Example: The number of males & females where you work or go to school; or the number of the make of cars Ford, GM, Toyota, Mazda, KIA, etc is numerical data.

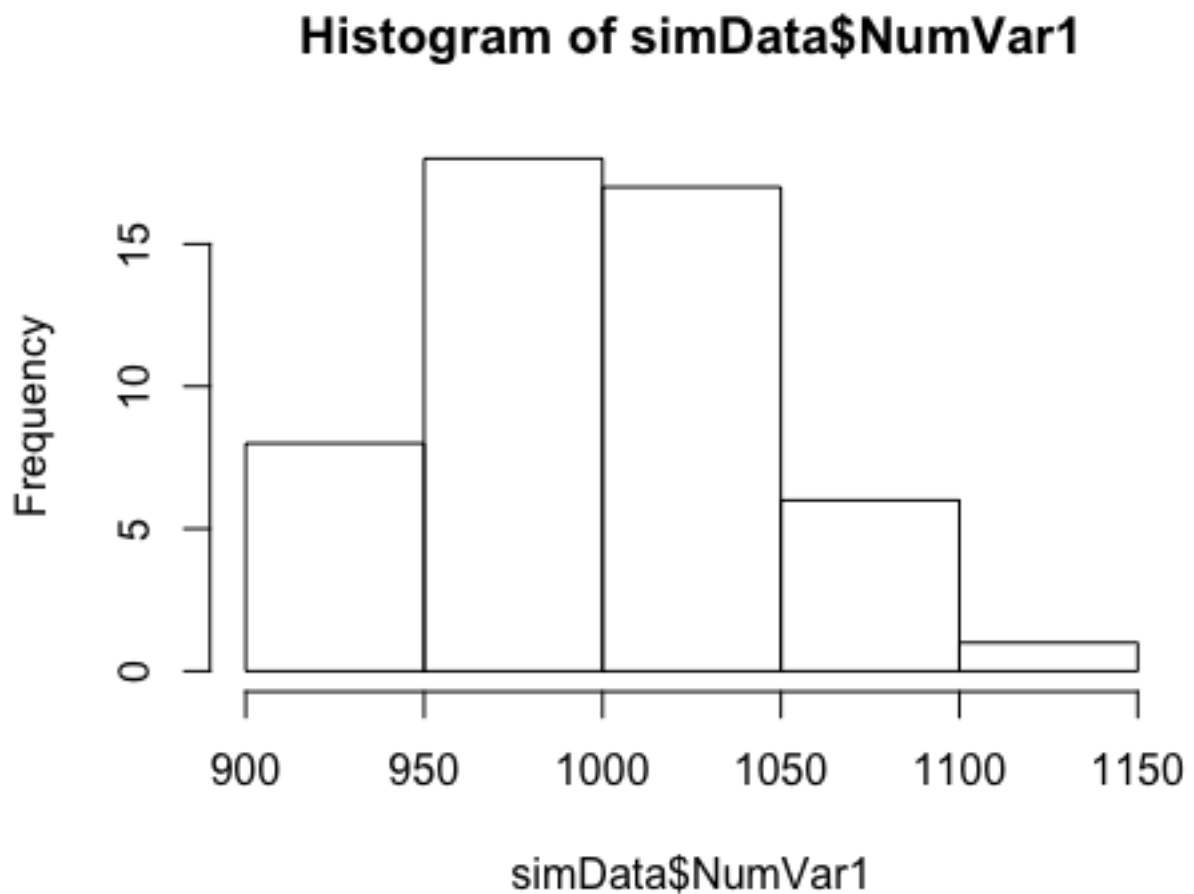
4) Create Graph

One Variable

Numeric Variable

Histogram

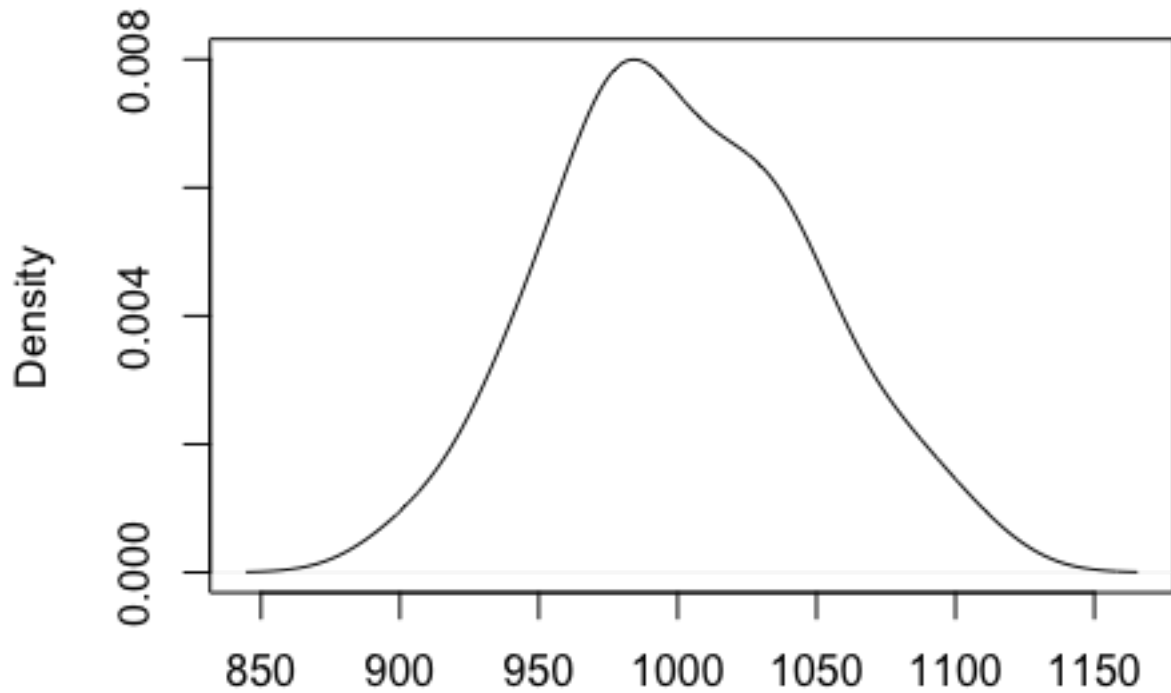
A histogram is a graphical representation of the distribution of data. It is an estimate of the probability distribution of a numeric variable.



Kernel Density Plot

This graph is a way to estimate the probability density function of a numeric variable.

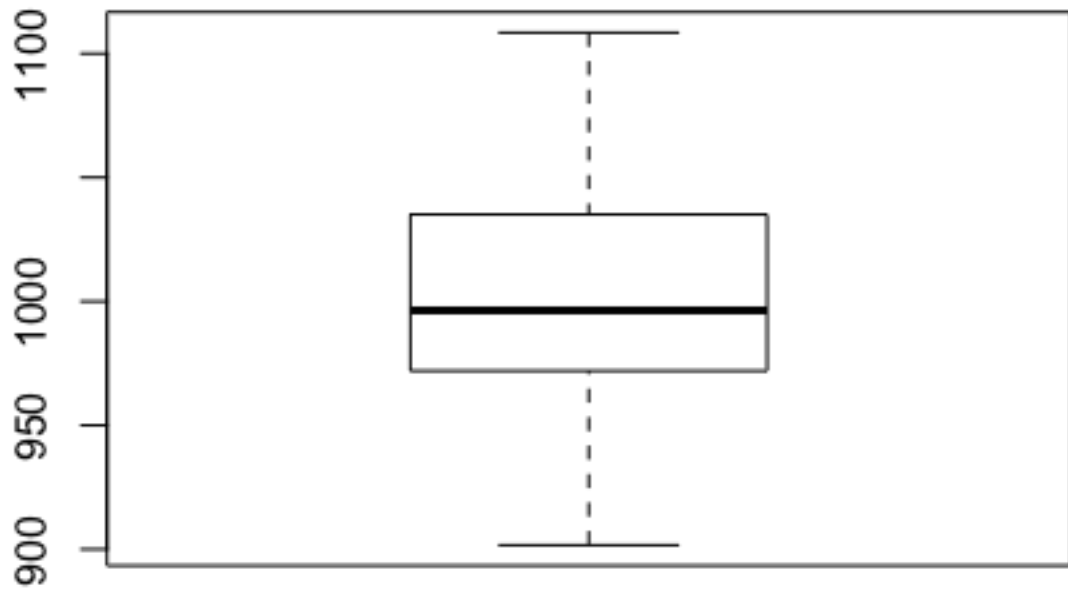
density.default(x = simData\$NumVar1)



N = 50 Bandwidth = 19.05

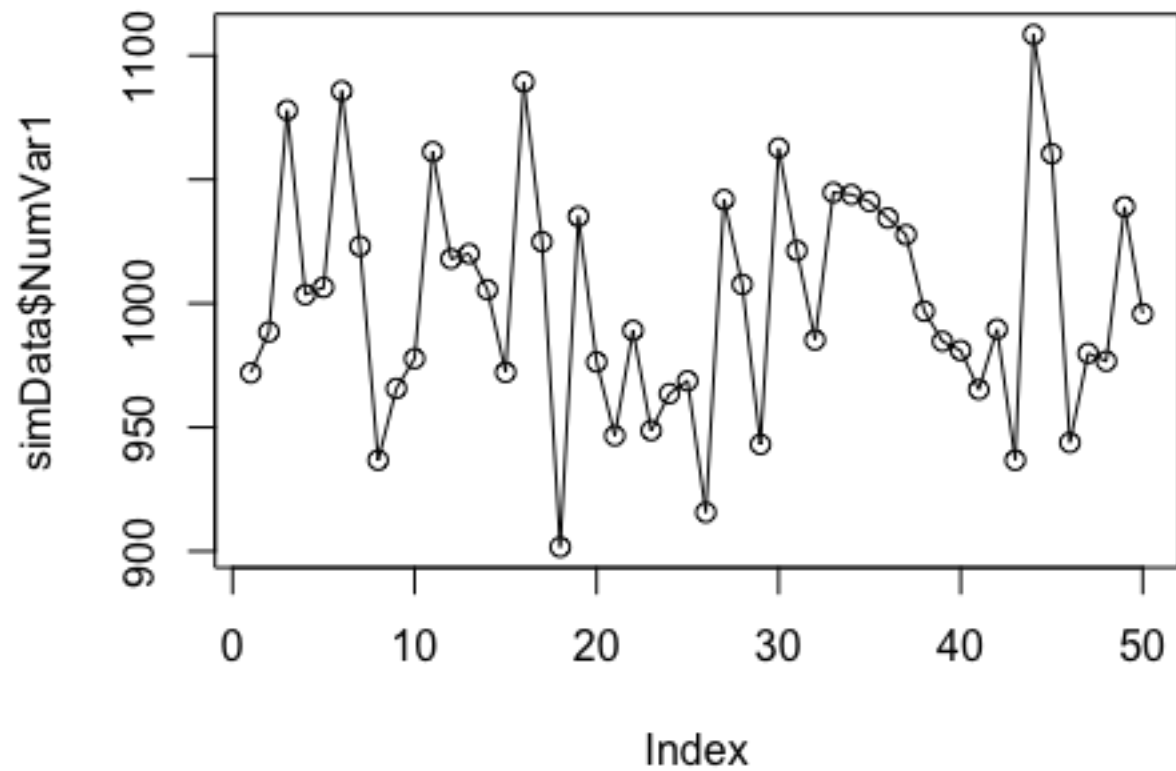
Box Plot

A boxplot is a convenient way of graphically depicting groups of numerical data through their quartiles. The lines extending vertically from the boxes indicate variability outside the upper and lower quartiles. Outliers may be plotted as individual points.



Index Plot

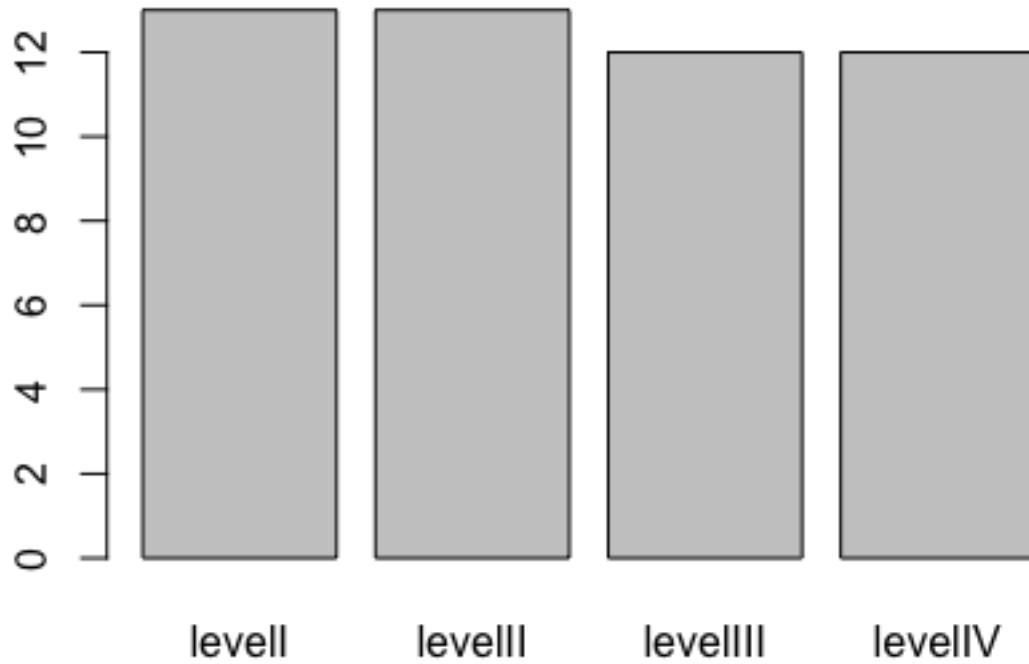
An index plot is a scatterplot of data plotted serially against the observation/case number within the sample.



Factor Variable

Barplot

A bar plot displays the frequency for all observations of a factor variable. (You can create a bar plot with vertical or horizontal bars)



Pie Chart

A pie chart is a circular chart divided into sectors, each sector shows the relative size of each factor. A pie chart should rarely be used because individuals have a hard time determining the size differences of the sections.

For Getting Counts Enter:

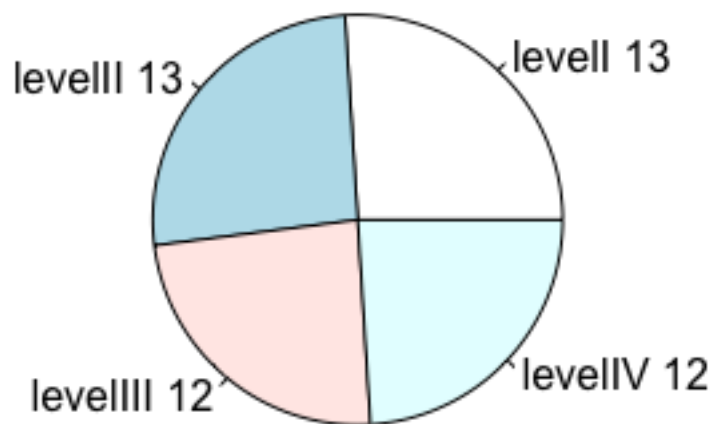
```
counts=table(simData$FacVar3)
```

To Create Tables Enter:

```
labs=paste(simData$FacVar3,counts)
```

To Plot Enter:

```
pie(counts,labels=labs)
```



Plotting Two Variables

Two Numeric Variables

Index Plot

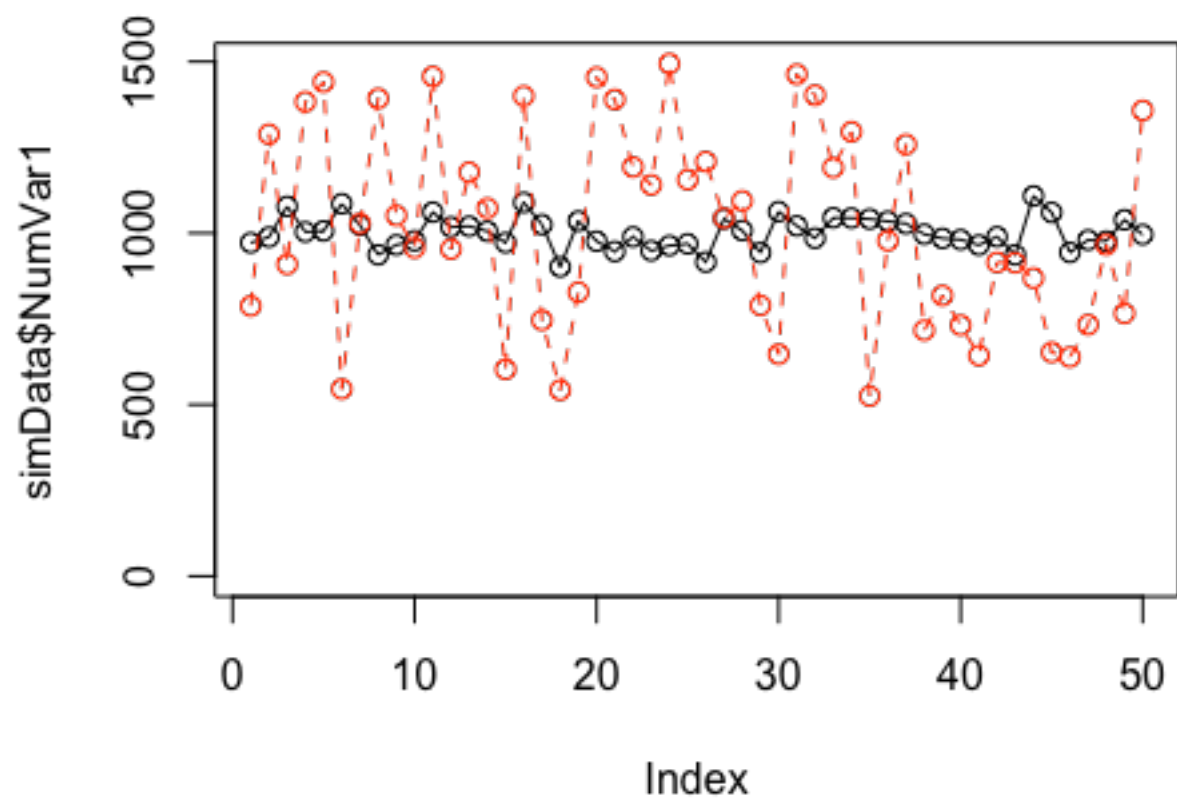
An index plot is a scatterplot of data plotted serially against the observation/case number within the sample.

To plot and index plot with one variable enter:

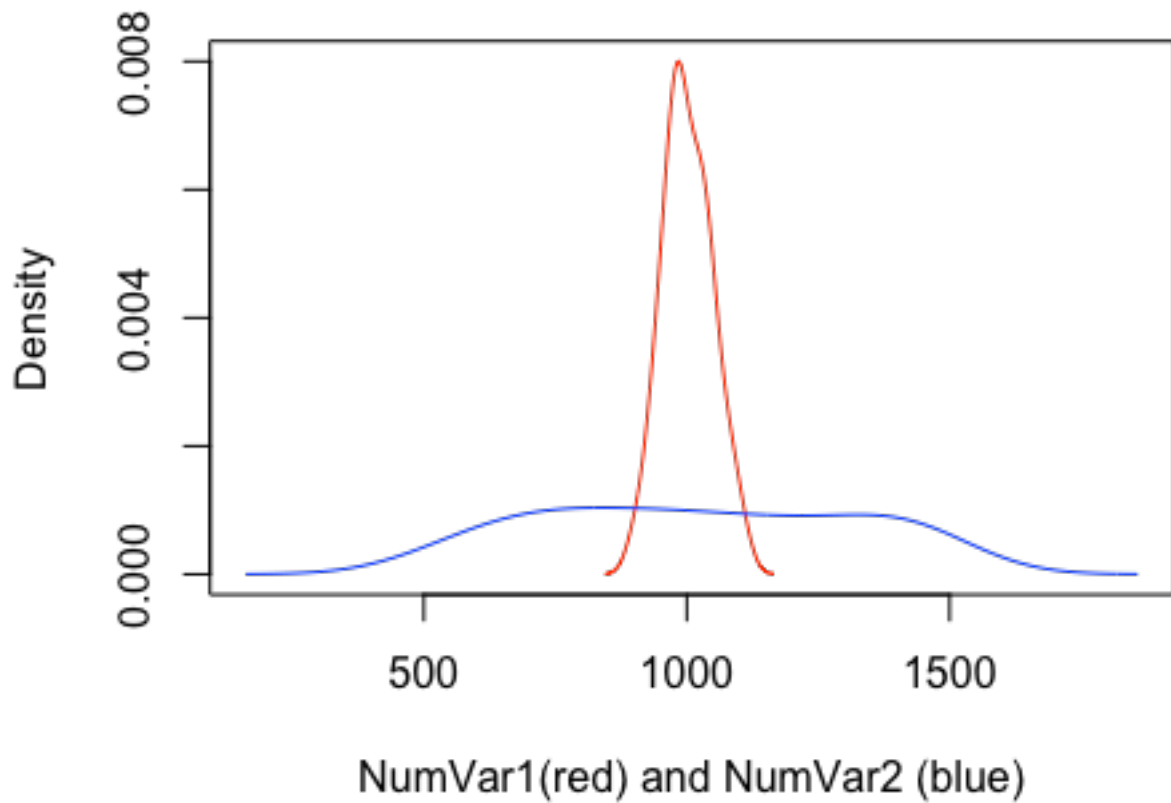
```
plot(simData$NumVar1,type="o",ylim=c(0,max(simData$NumVar1,simData$NumVar2)))
```

To add another variable enter:

```
lines(simData$NumVar2,type="o",lty=2,col="red")
```

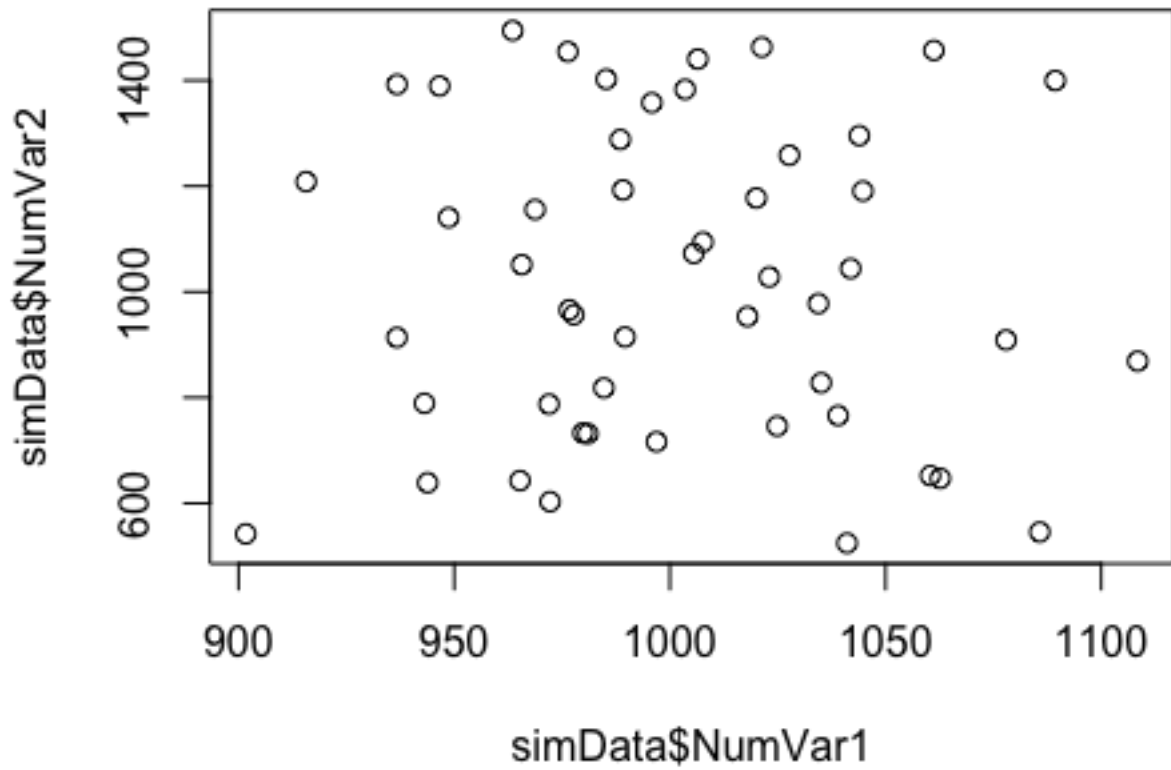


Density Plot



Scatter Plot

A scatter plot is a graph of plotted points that show the relationship between two sets of numeric data. The pattern of the resulting points revealing any correlation present



Two Factor Variables

Mosaic Plot

The mosaic plot starts as a square with length one. The square is divided first into horizontal bars whose widths are proportional to the probabilities associated with the first factor variable. Then each bar is split vertically into bars that are proportional to the conditional probabilities of the second factor variable. (Additional splits can be made if wanted using a third, fourth variable, etc.)

table(simData\$FacVar2, simData\$FacVar3)

	levelA	levelB	levelC
levelI			
levelII			
levelIII			
levelIV			

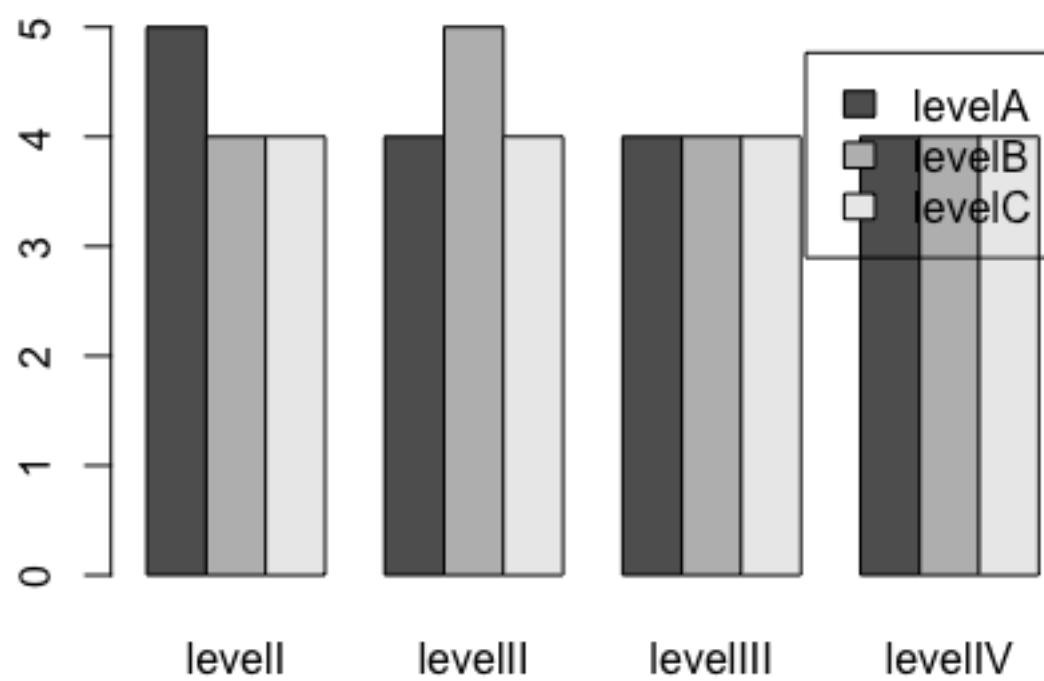
Barplots

To get the cross tab enter:

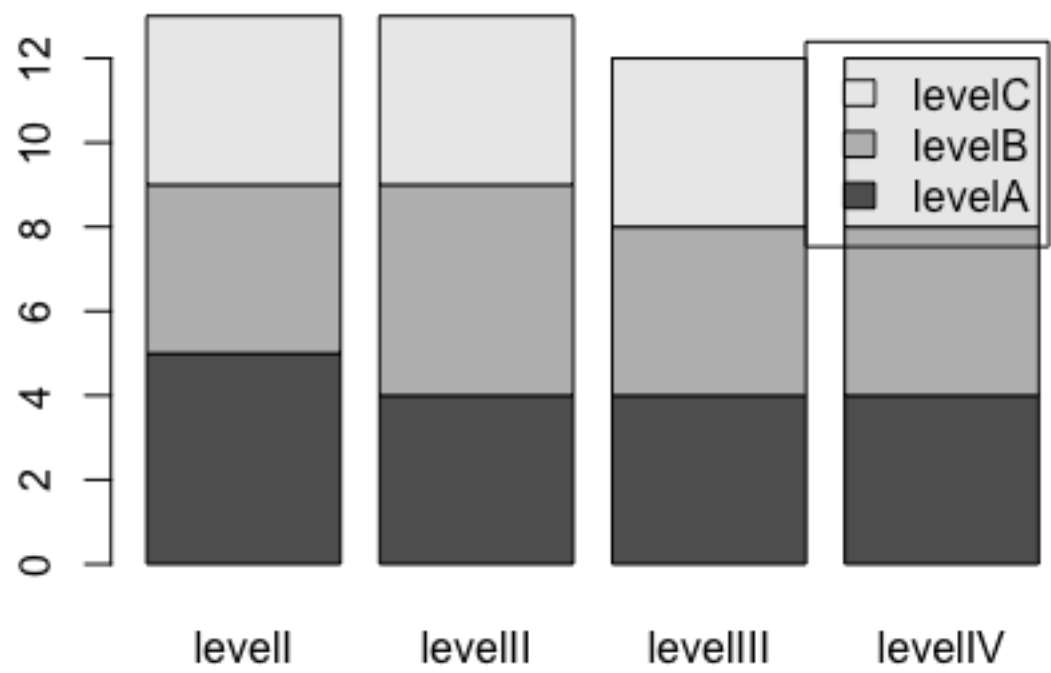
```
bartable=table(simData$FacVar2,simData$FacVar3)
```

To Plot Enter:

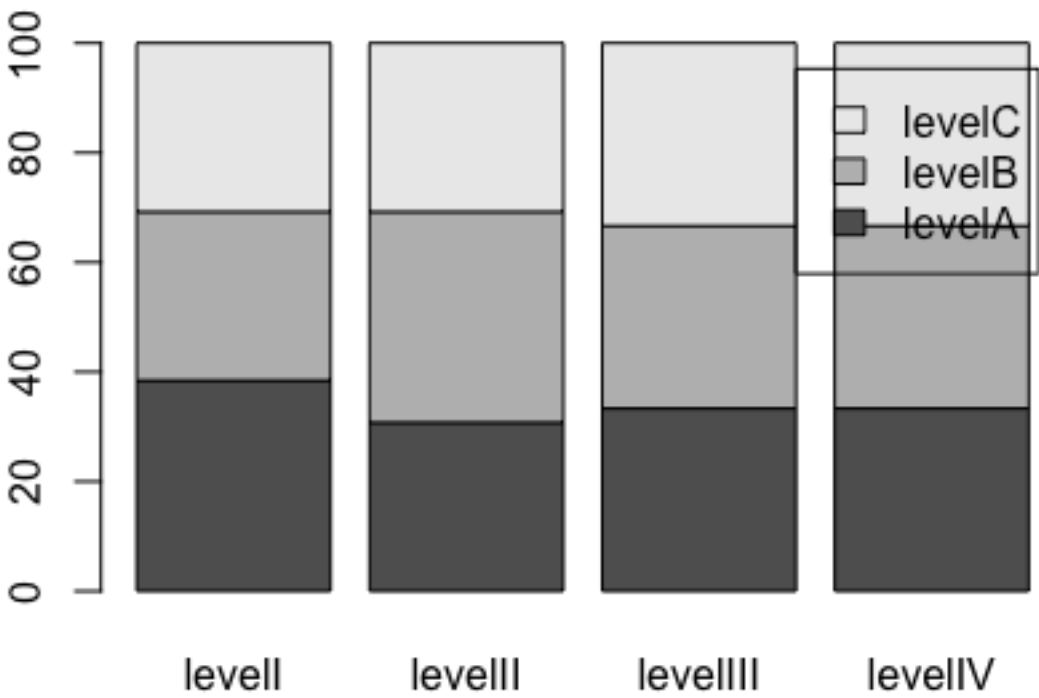
```
barplot(bartable,beside=TRUE, legend=levels(unique(simData$FacVar2)))
```



Stacked Barplot

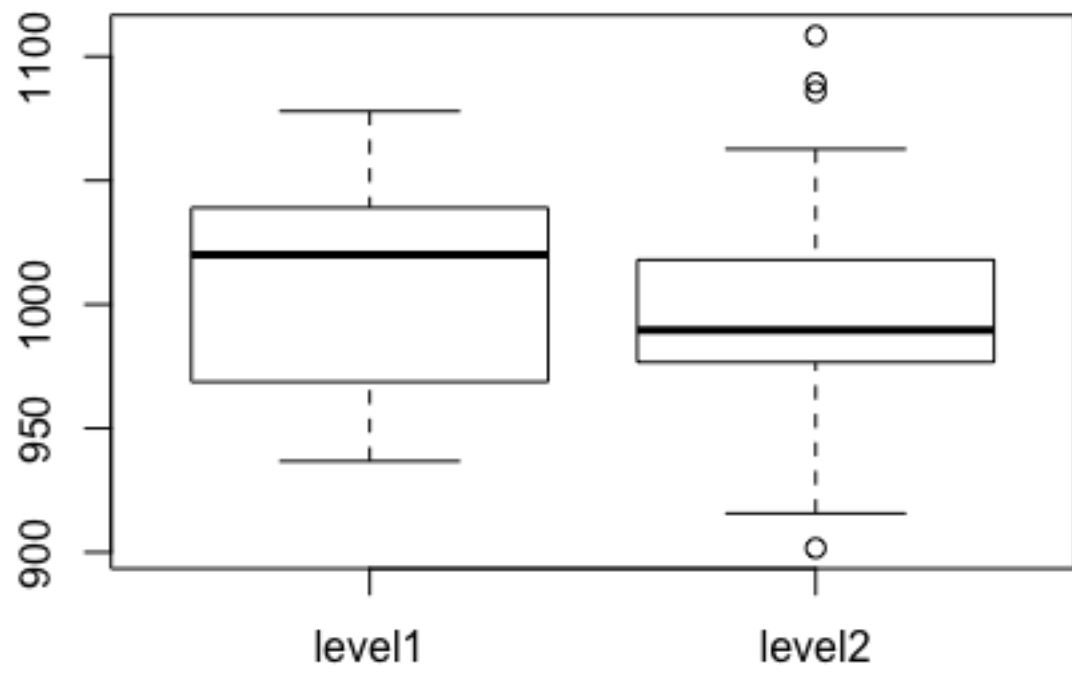


100% Stacked Barplot

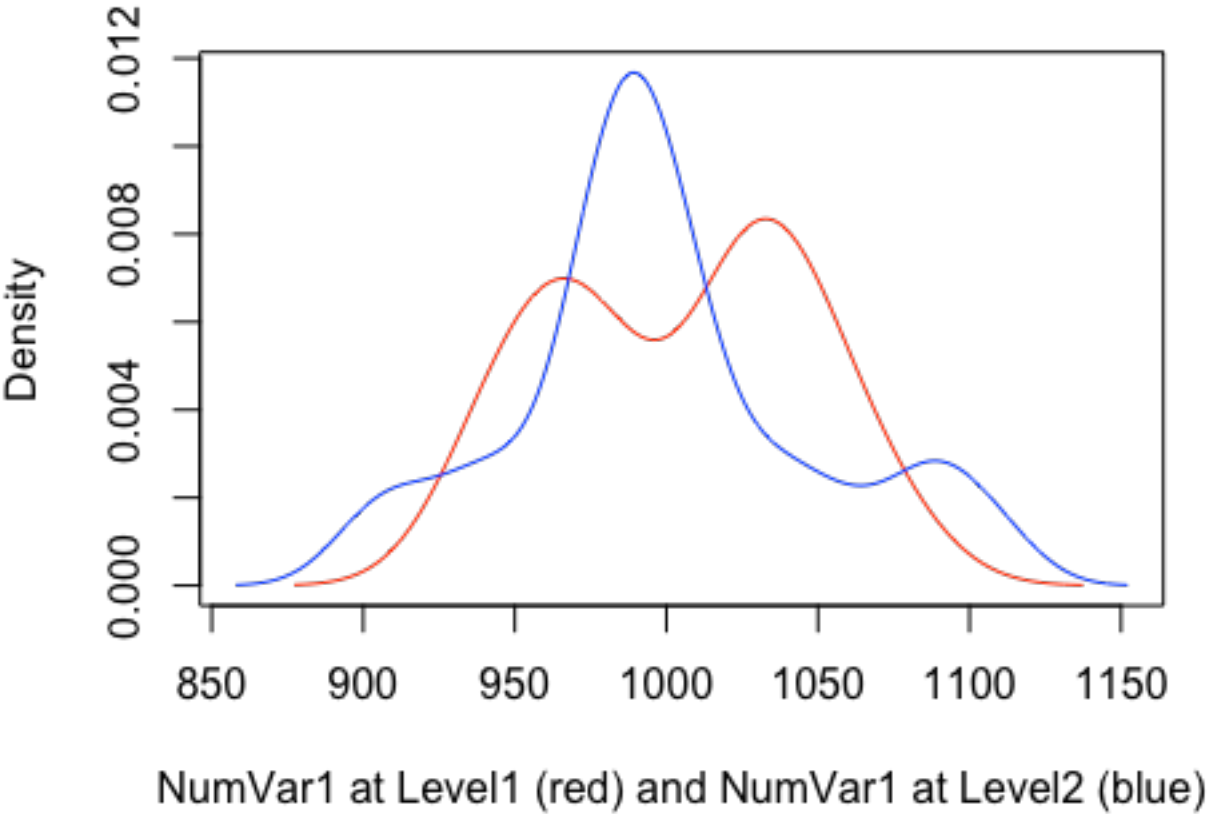


One Factor and One Numeric

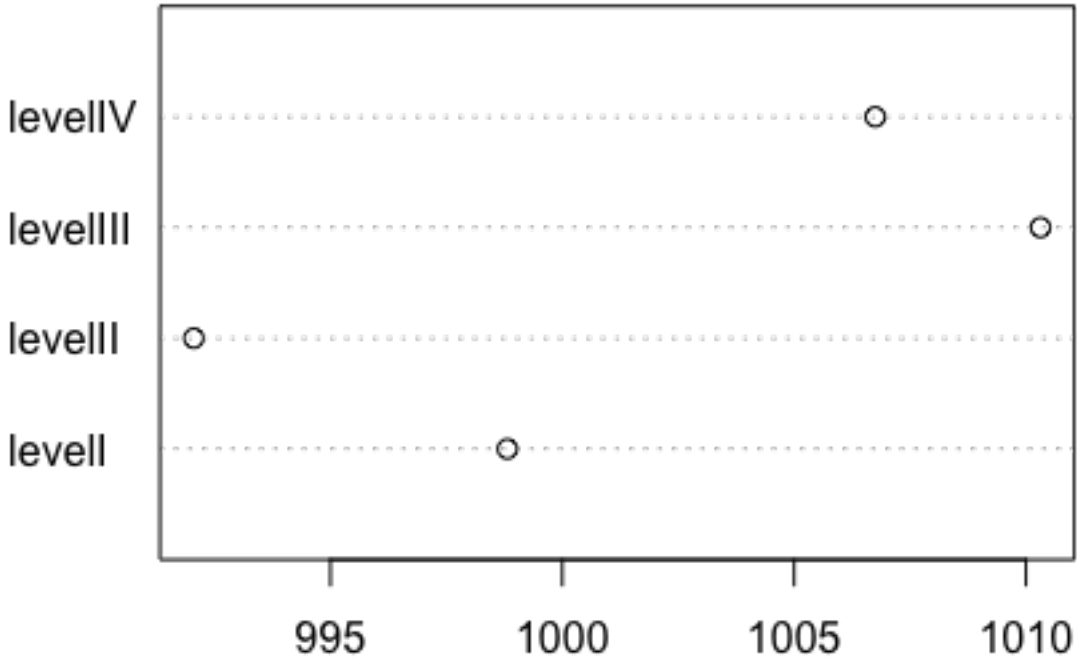
Boxplot



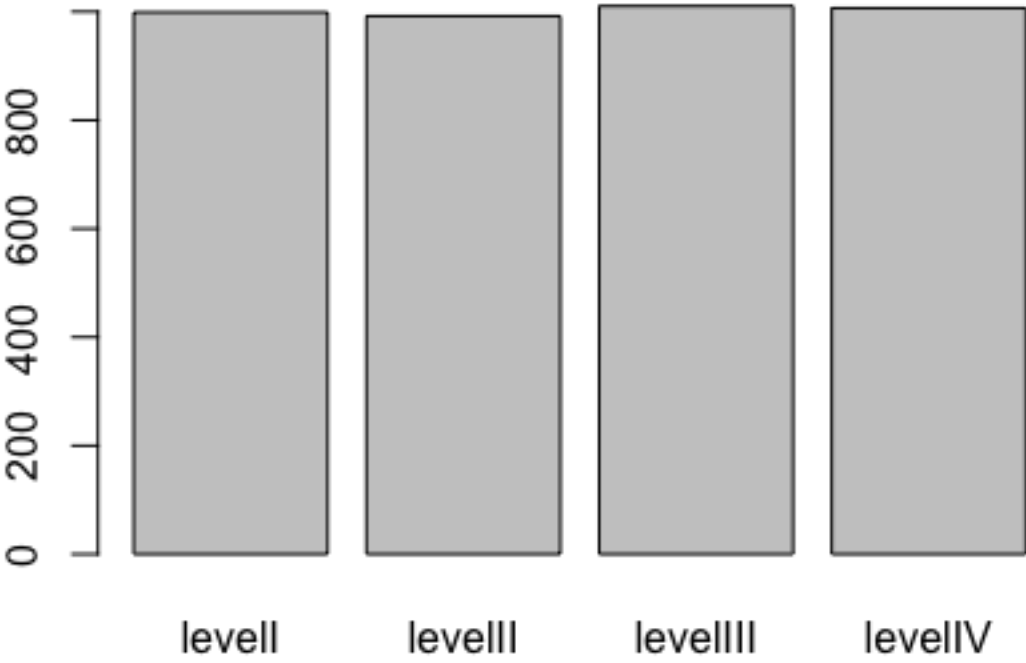
Density Plot



Dot Chart

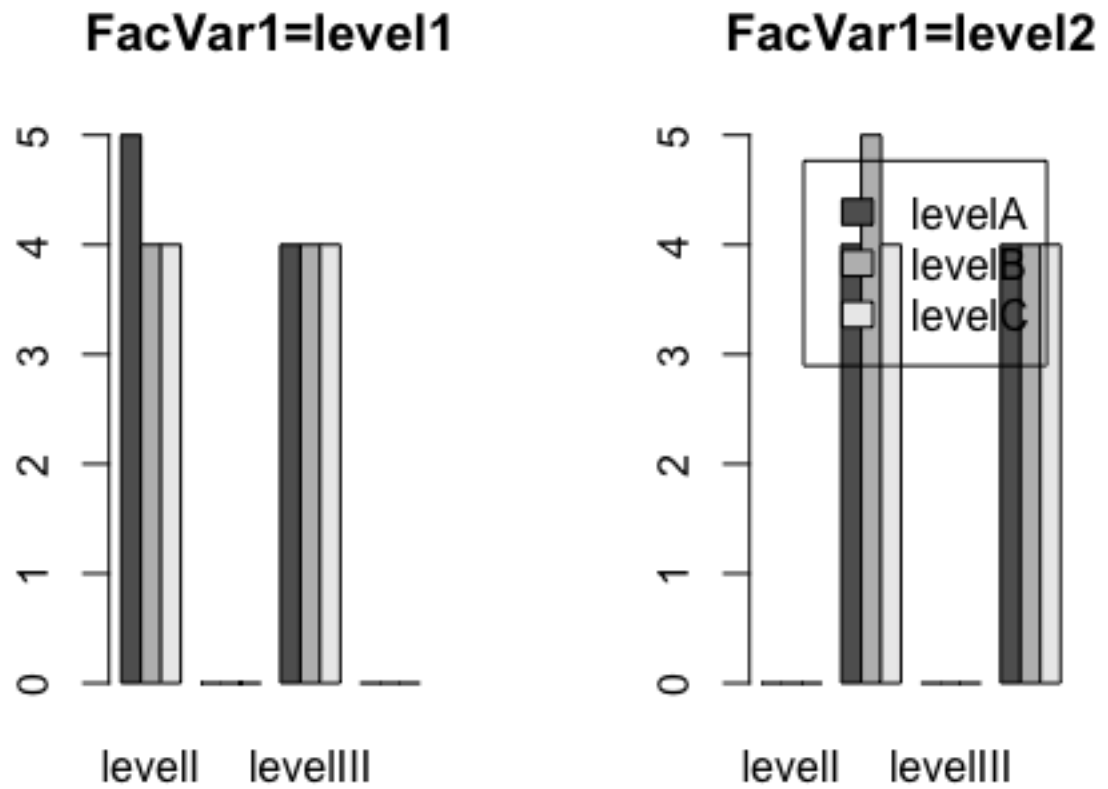


Bar Plot



Three Variables

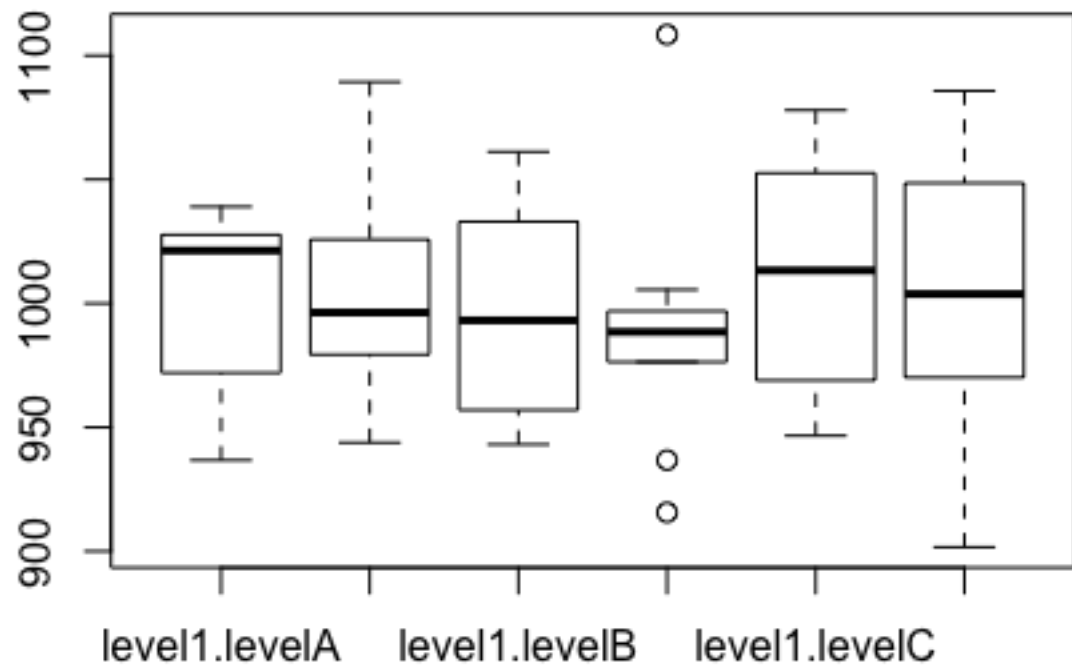
Three Factor Variables



One Numeric and Two Factor Variables

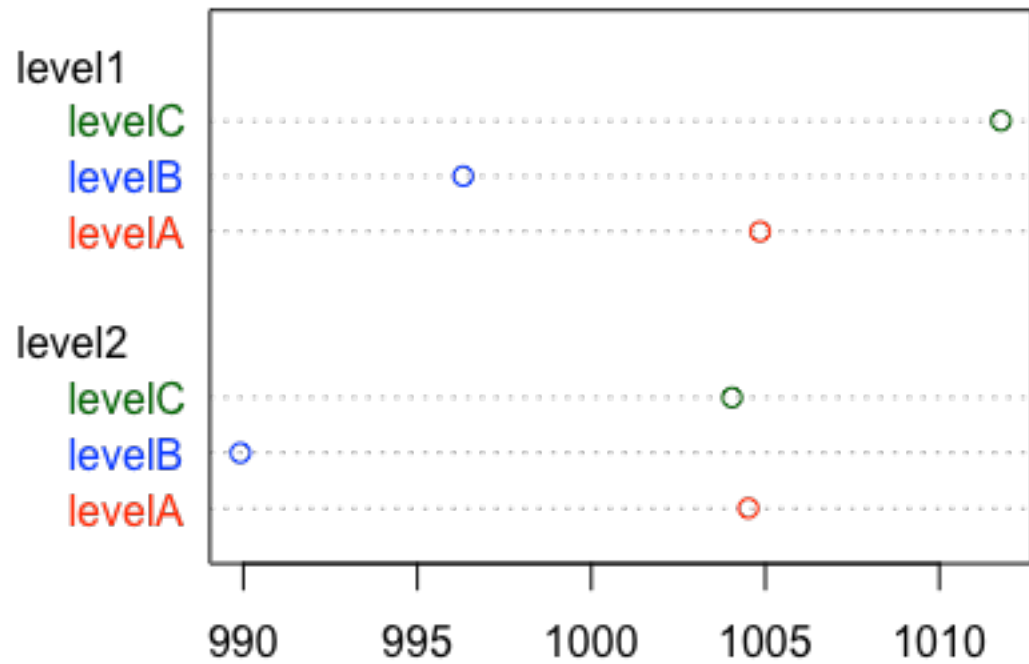
Boxplot

(NumVar1 over an interaction of 6 levels of the combination of FacVar1 and FacVar2)



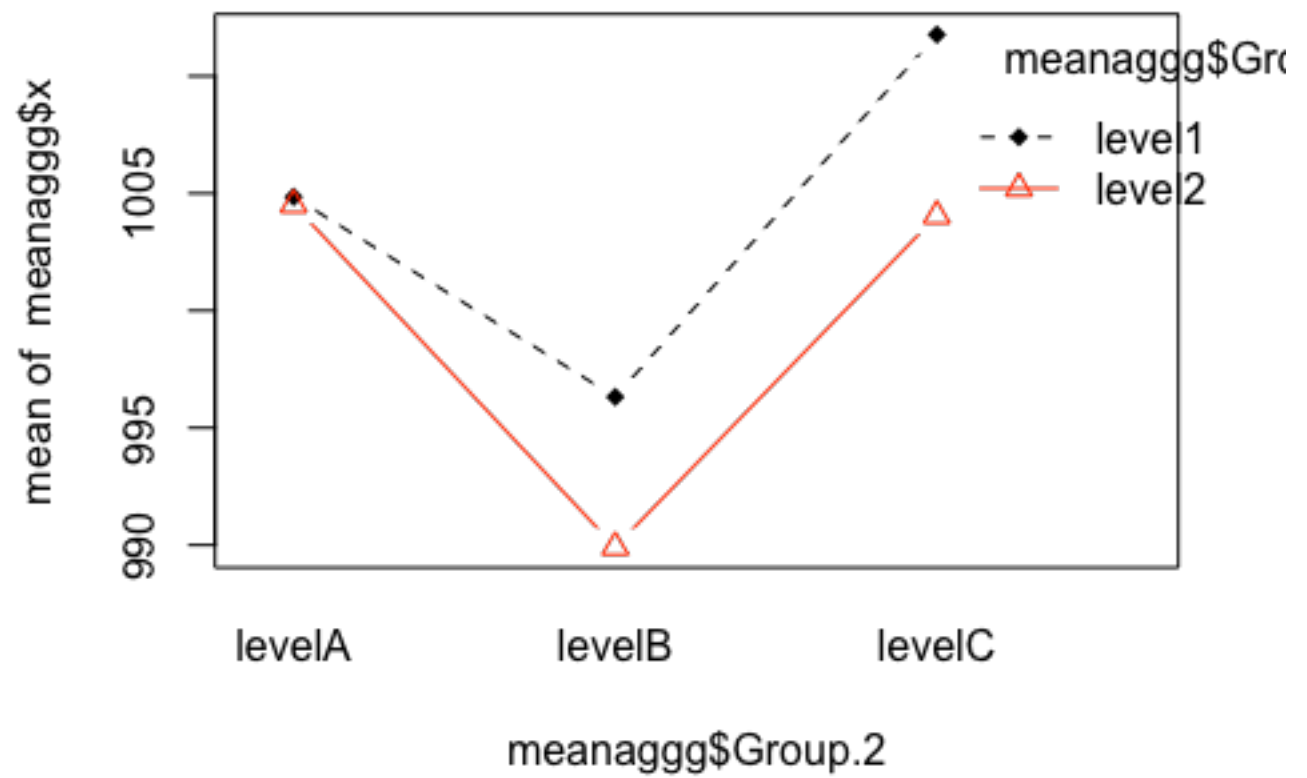
Dot Chart

(Mean of 1 Numeric over levels of two factor vars)

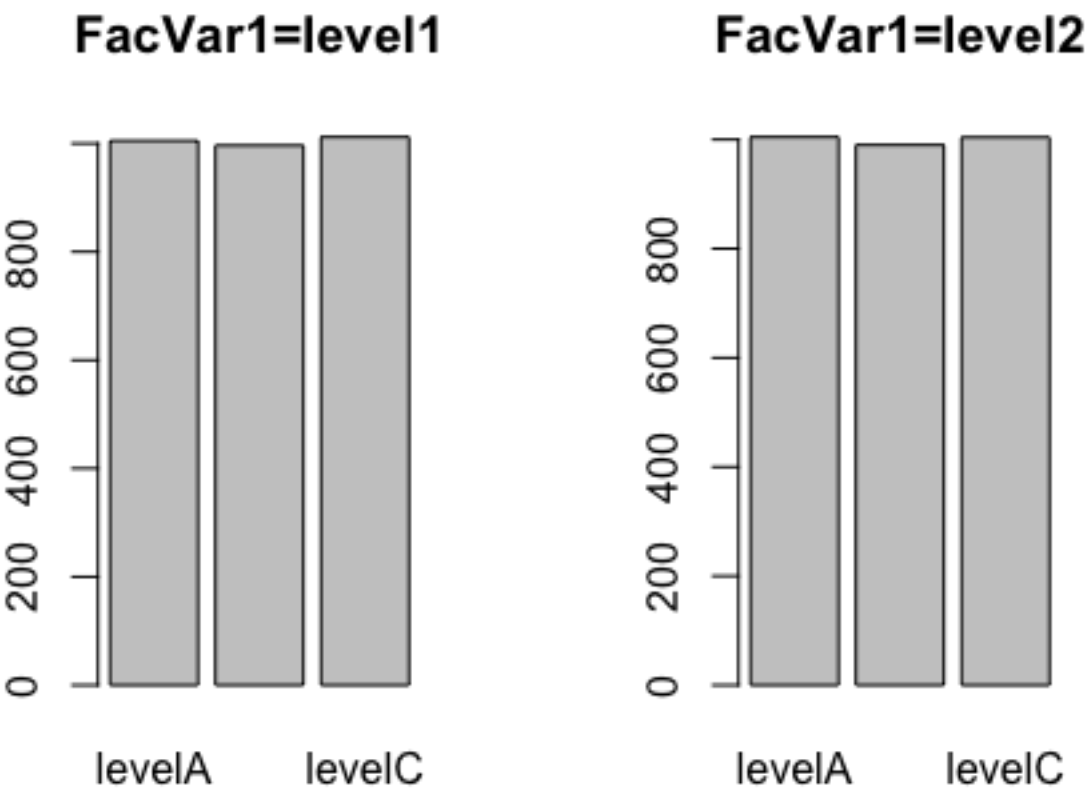


Interaction Chart

An interaction chart plots the mean (or other summary) of the response for a combination of factors, thereby illustrating possible interactions.

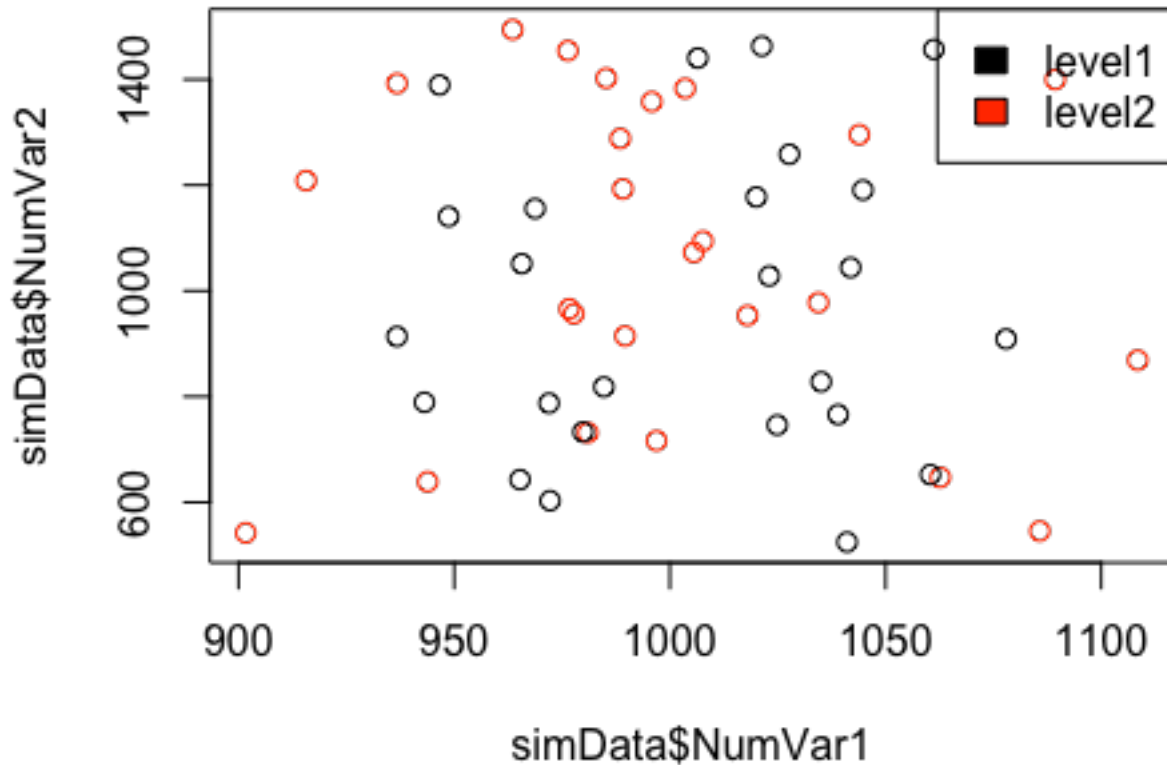


Some A Bar Plot



Two Numeric and One Factor Variables

Scatter Plot (with color identifying the factor variable)



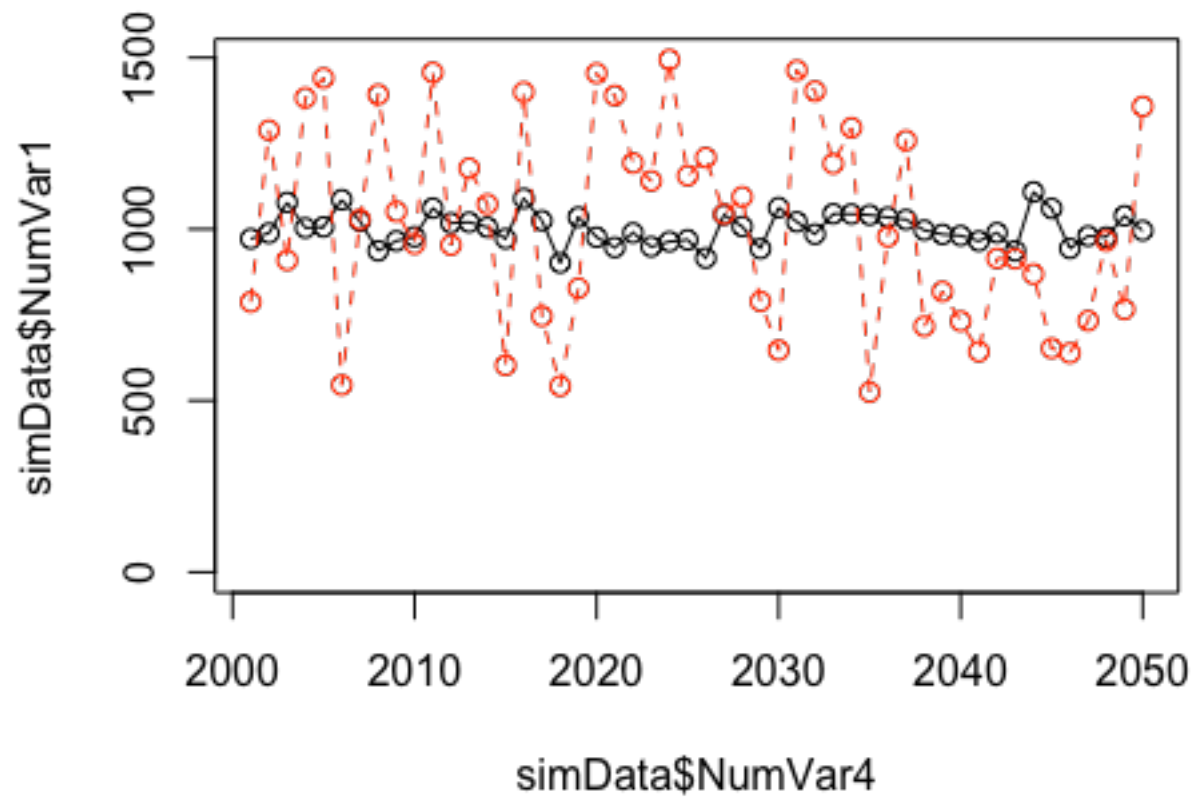
Three Numeric Variables

To Join Dots With the Lines Enter:

```
plot(simDataNumVar4, simDataNumVar1, type="o", ylim=c(0, max(simDataNumVar1, simDataNumVar2)))
```

To Add Another Line Enter:

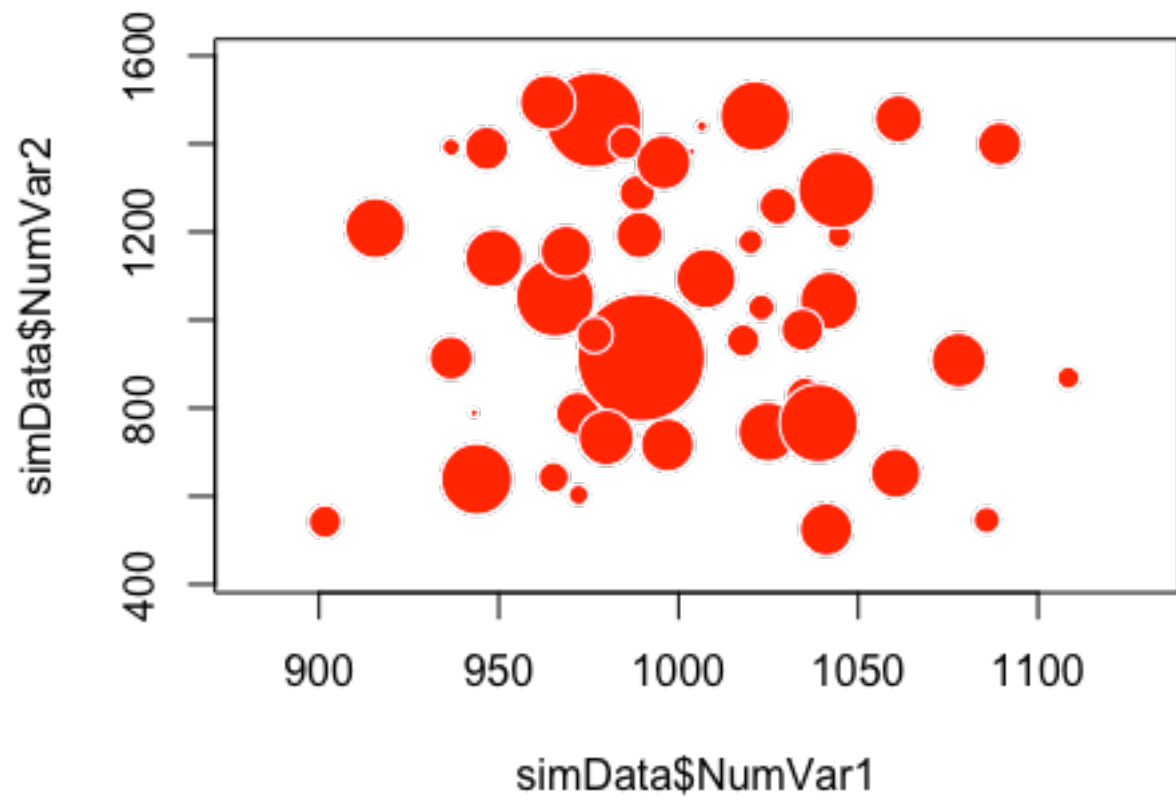
```
lines(simDataNumVar4, simDataNumVar2, type="o", lty=2, col="red")
```

Bubble plot

A bubble chart is a type of chart that displays three dimensions of data. Each entity with its triplet of associated data is plotted as a disk that expresses two of the values. Then, through the disk's xy location and the third through its size.

Sized by NumVar3



Scatterplot Matrix of All Numeric Vars (colored by a Factor variable)

