Base R Plots: A Guide to Plotting in R

Granger Moch

9/19 /2014

Contents

1 Purpose **3**

2 Data **3**

2.1 Simulate Data . . . 3

2.2 3 Factor Variables . . . 3

2.3 4 Numeric Vars . . . 3

3 Plotting One Numeric Variable **4**

3.1 Index Plot . . . 4

3.2 Histogram . . . 5

3.3 Kernel Density Plot . . . 6

3.4 Box Plot . . . 7

4 Plotting One Factor Variables **8**

4.1 Bar Plot . . . 8

4.2 Pie Chart . . . 9

5 Two Numeric Variables **10**

5.1 Index Plot (with one variable) . . . 10

5.2 Index Plot (with two variables) . . . 11

5.3 Density Plots . . . 12

5.4 Scatter Plots . . . 13

6 Two Factor Variables **14**

6.1 Mosaic Plot . . . 14

6.2 Bar Plots . . . 15

6.2.1 Stacked Bar Plot . . . 16

6.2.2 Stacked Bar Plot (100% scale) . . . 17

7 Two Variables: One Factor and One Numeric **18**

7.1 Box Plots . . . 18

7.2 Data Density Plot . . . 19

7.3 Dotchart . . . 20

7.4 Bar Plot . . . 21

8 Three Variables: Three Factor Variables **22**

8.1 Boxplot 22

9 Three Variables: One Numeric and Two Factor Variables **23**

9.1 Box Plot 23

9.2 Dot Chart of Mean of 1 Numeric over levels of two factor vars 24

9.3 Interaction Plot 25

9.4 some a bar plot 26

10 Three Variables: Two Numeric and One Factor Variable **27**

10.1 Scatter Plot 27

11 Three Variables: Three Numeric Variables **28**

11.1 Plot and Bubble Plot 28

12 Scatterplot Matrix of all Numeric Variables **30**

1 Purpose

I am creating this guide not solely for the assignment, but so that those that are new to R can recognize the extensive functionality it offers. Furthermore, I hope to provide a succinct introduction to several different graph types and their uses. for both my sake as well as those who read this guide.

2 Data

2.1 Simulate Data

2.2 3 Factor Variables

```
FacVar1=as.factor(rep(c("level1","level2"),25))
FacVar2=as.factor(rep(c("levelA","levelB","levelC"),17)[-51])
FacVar3=as.factor(rep(c("levelI","levelII","levelIII","levelIV"),13)[-c(51:52)])
```

2.3 4 Numeric Vars

```
set.seed(123)
NumVar1=round(rnorm(n=50,mean=1000,sd=50),digits=2) ## Normal distribution
set.seed(123)
NumVar2=round(runif(n=50,min=500,max=1500),digits=2) ## Uniform distribution
set.seed(123)
NumVar3=round(rexp(n=50,rate=.001)) ## Exponential distribution
NumVar4=2001:2050

simData=data.frame(FacVar1,FacVar2,FacVar3,NumVar1,NumVar2,NumVar3,NumVar4)
```

We can now use this data frame "simData", as created above, to plot the sample data using a variety of graph types. Each graph type has its advantages and disadvantages, but when used appropriately graphs can be very useful for visualizing data.

For more information on why the set.seed function is used, see this page: http://stackoverflow.com/questions/13605271/reasons-for-using-the-set-seed-function

3 Plotting One Numeric Variable

3.1 Index Plot

An index plot is a scatterplot of values on some measure plotted serially against the associated observation (case) number within the sample. It can help in highlighting extreme values in a sample and in indicating trends when the data are, for instance, in spatial or time order.

```
plot(simData$NumVar1,type="o")
```

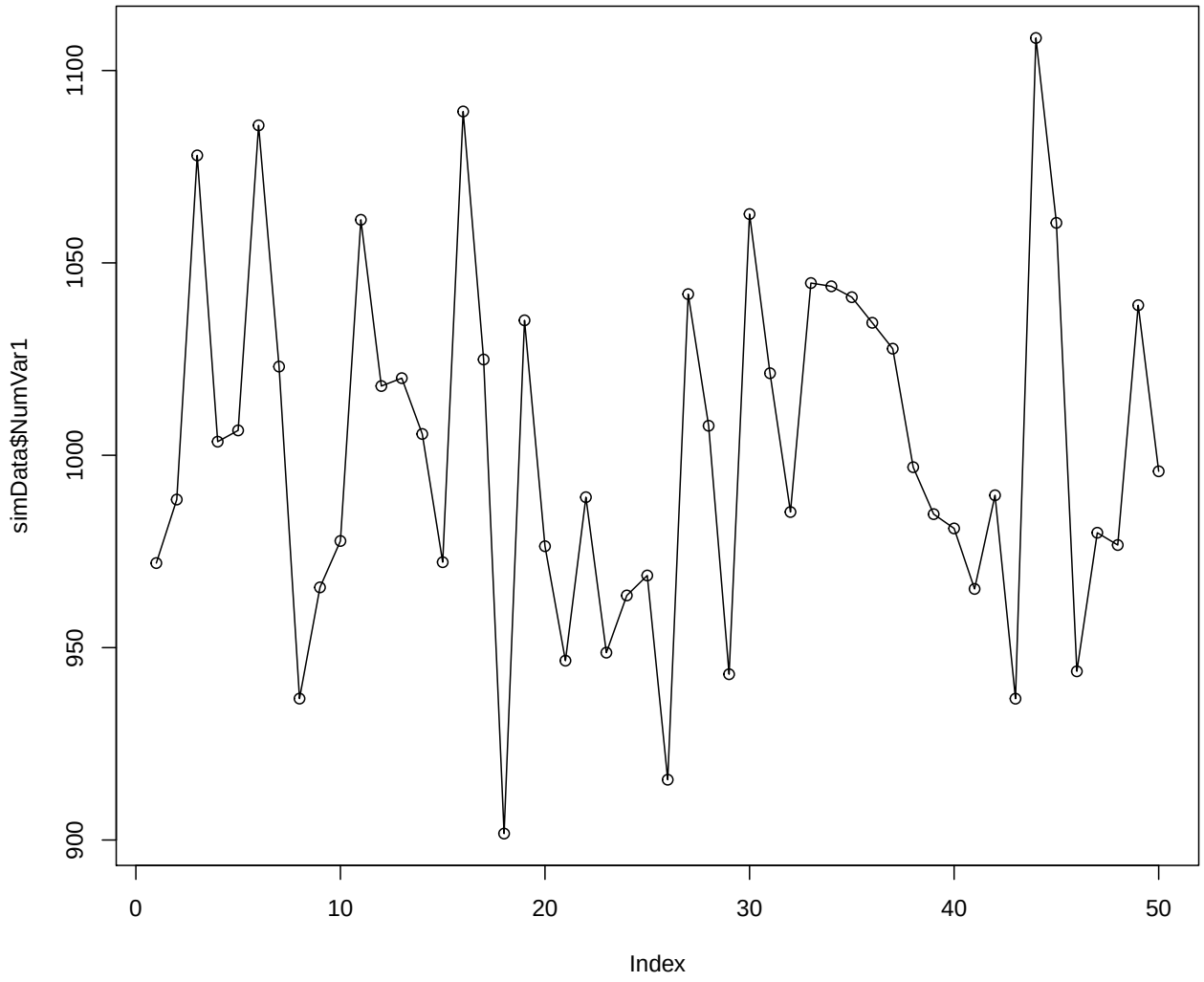


Figure 1: Index Plot of One Numeric Variable

3.2 Histogram

A Histogram is a graphical display of data using bars of various heights. Not to be confused with a bar chart, a histogram groups numbers into ranges of your choice. They are an excellent way to display continuous data, for example weight, minutes, etc.

```
hist(simData$NumVar1)
```

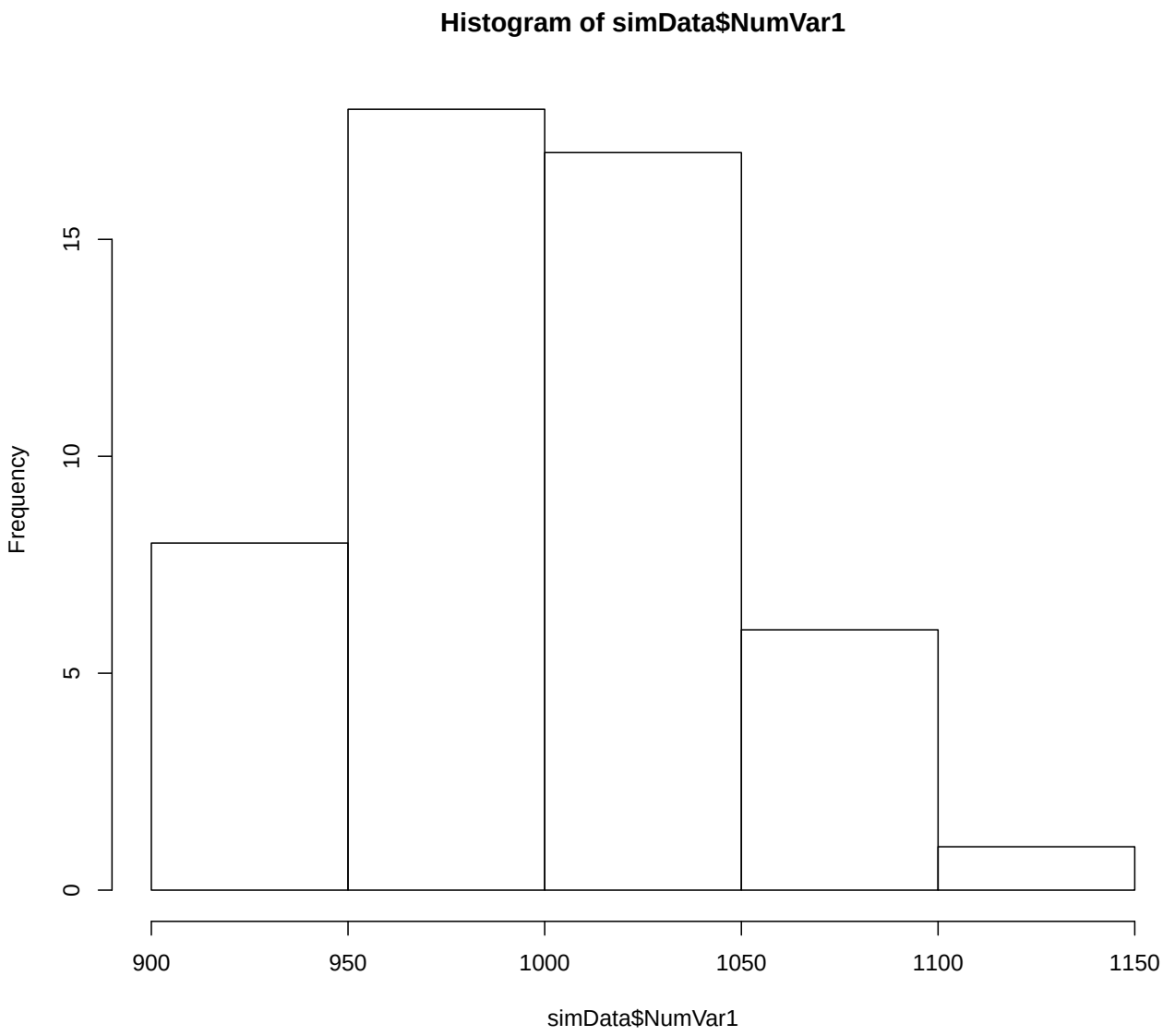


Figure 2: Histogram of One Numeric Variable

3.3 Kernel Density Plot

Kernel density estimation (KDE) is one way to estimate the probability density function of a certain random variable. With this type of plot, inferences can be made about the population.

```
plot(density(simData$NumVar1)) ## Kernel density plot
```

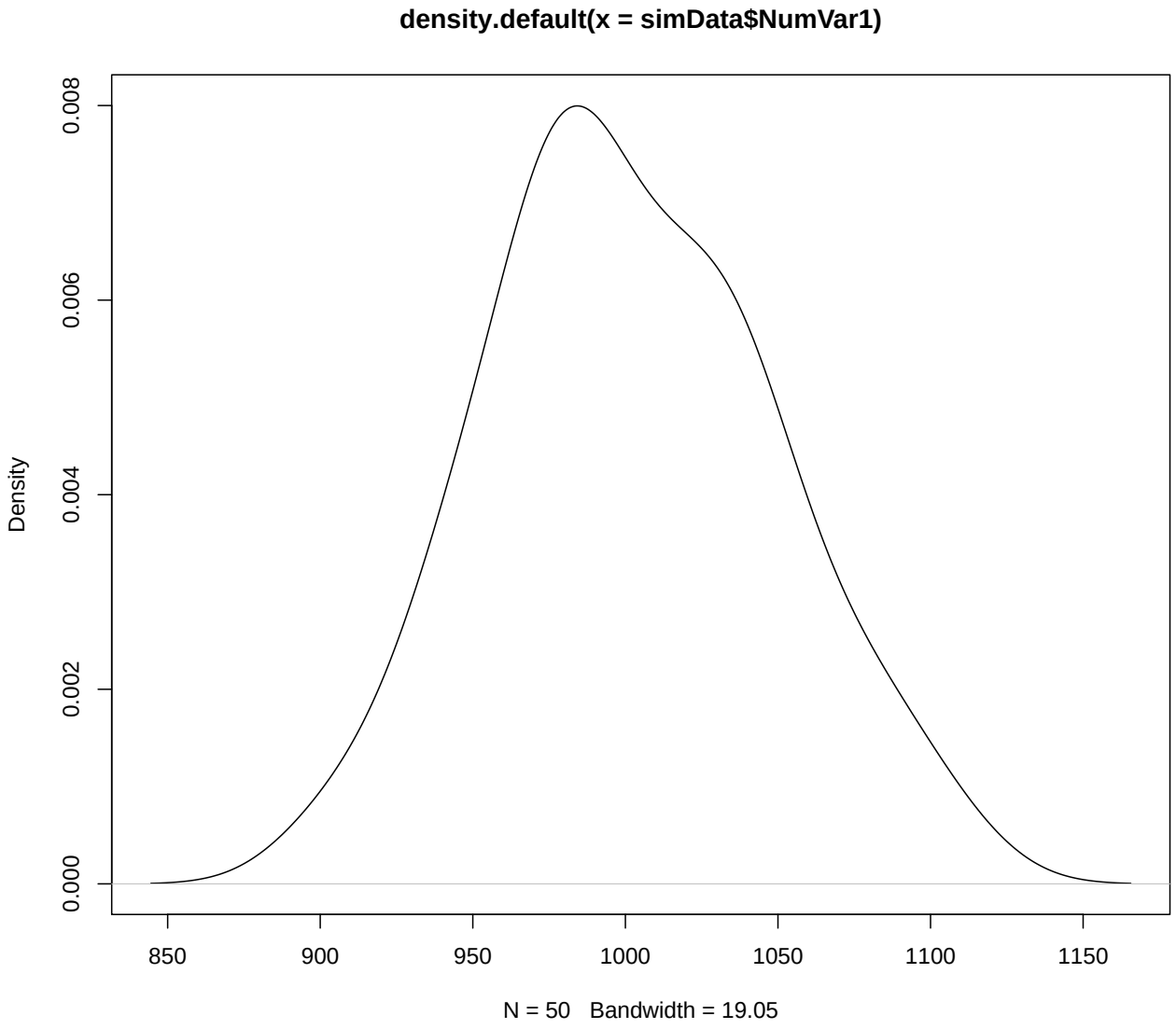


Figure 3: Kernel Density Plot of One Numeric Variable

3.4 Box Plot

The box plot (also known as the box and whisker diagram) is a means of displaying the distribution of data based on a five zone tier from bottom to top : minimum, first quartile, median, third quartile, and maximum.

```
boxplot(simData$NumVar1) ## box plot
```

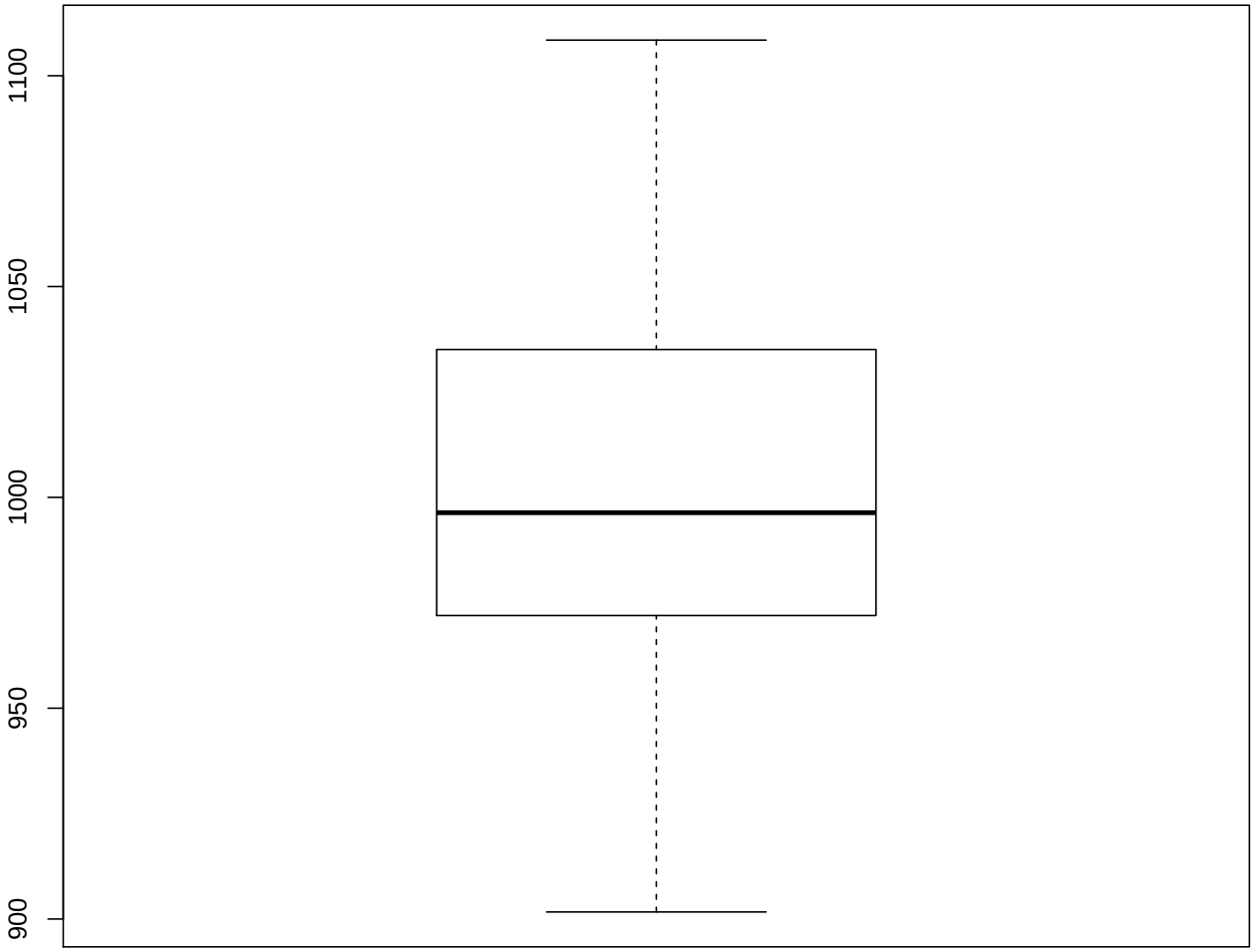


Figure 4: Box Plot of One Numeric Variable

4 Plotting One Factor Variables

4.1 Bar Plot

A bar plot is a chart that uses horizontal (or vertical) bars to show comparisons among different categories. On one axis the specific categories being compared are shown, and on the other axis, a discrete value scale. Bar charts are particularly handy for displaying data that is classified into nominal or ordinal categories. Nominal data are categorised according to descriptive or qualitative information such as state of birth, or ethnicity. Ordinal data is somewhat similar but the different categories can also have the attribute of rank.

```
plot(simData$FacVar3) ## bar plot
```

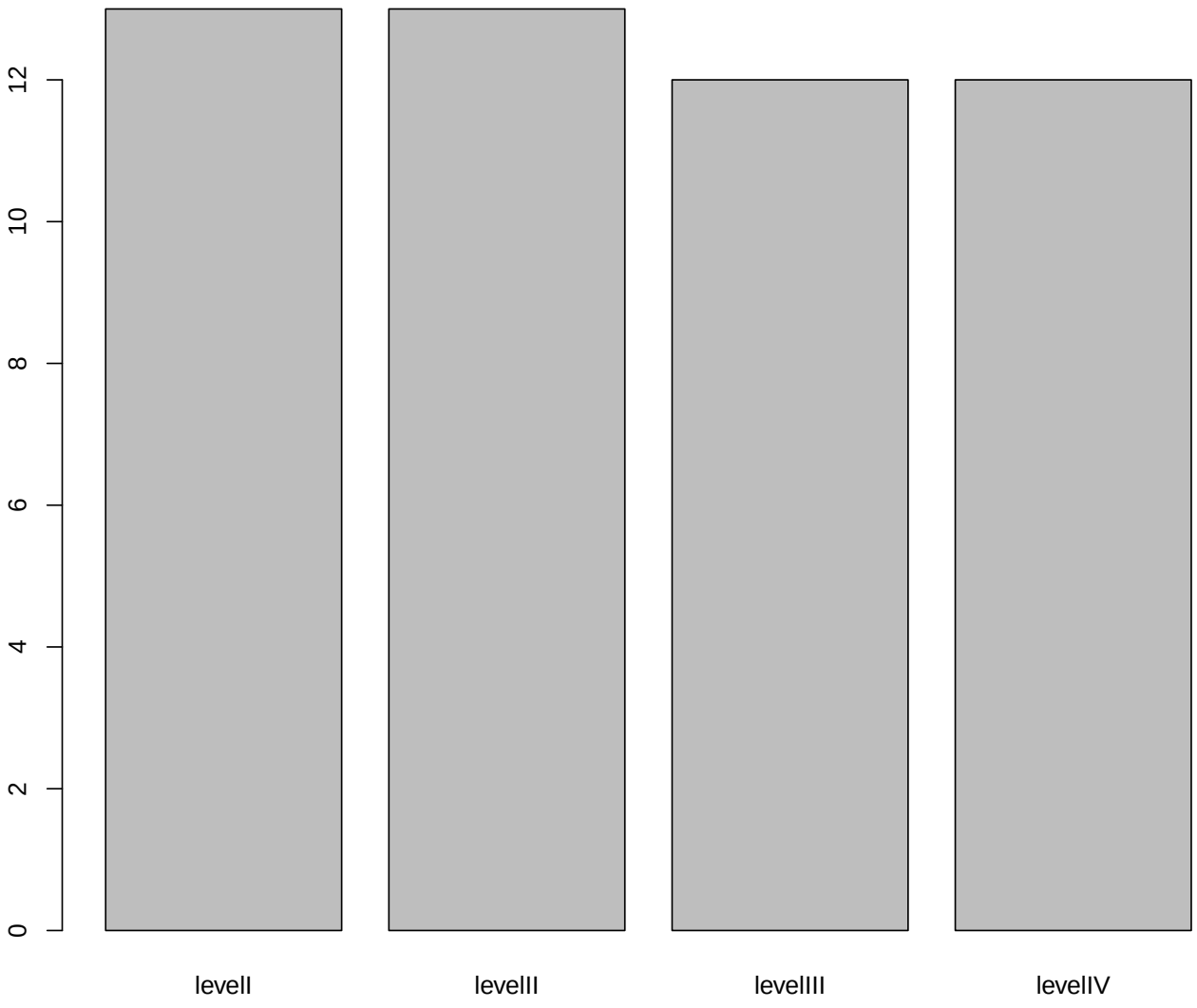


Figure 5: Bar Plot of One Factor Variable

4.2 Pie Chart

"The pie chart is a simple information graphic whose principal purpose is to show the relationship of a part to the whole. It is, by and large, the wrong choice as an exploratory device, and it is certainly not the correct choice when the graph maker or graph reader has a complicated purpose in mind," (Ian Spence, No Humble Pie: The Origins and Usage of a Statistical Chart).

Although the pie chart does have a few upsides, such as providing the ability to quickly identify the largest and smallest portions as they relate to the whole, it should certainly be used cautiously. In many instances, it can be very difficult to tell which portion is larger and estimating HOW much larger can be even tougher. Furthermore, the overall ease of reading decreases as the number of sections increase. Thus, pie charts will not be appropriate for comparing a high number of categories.

```
counts=table(simData$FacVar3) ## get counts
labs=paste(simData$FacVar3,counts)## create labels
pie(counts,labels=labs) ## plot
```

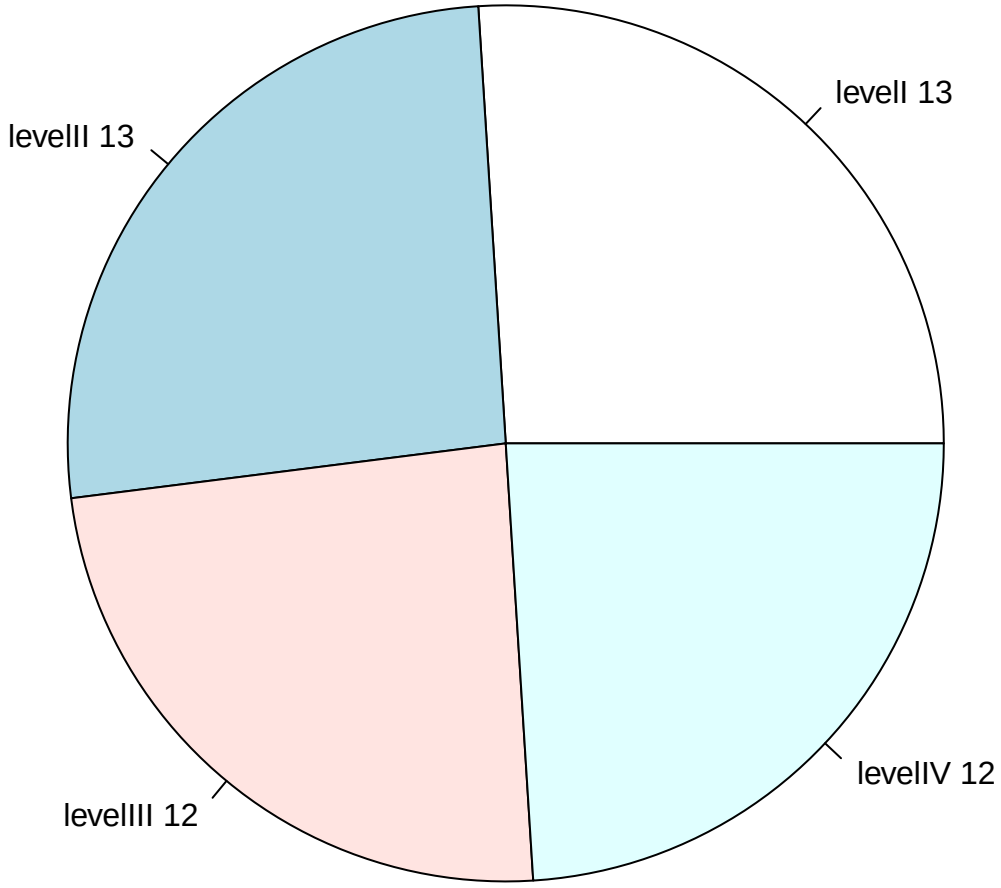


Figure 6: Pie Chart of One Factor Variable

5 Two Numeric Variables

5.1 Index Plot (with one variable)

```
plot(simData$NumVar1,type="o",ylim=c(0,max(simData$NumVar1,simData$NumVar2)))## index plot with one vari
```

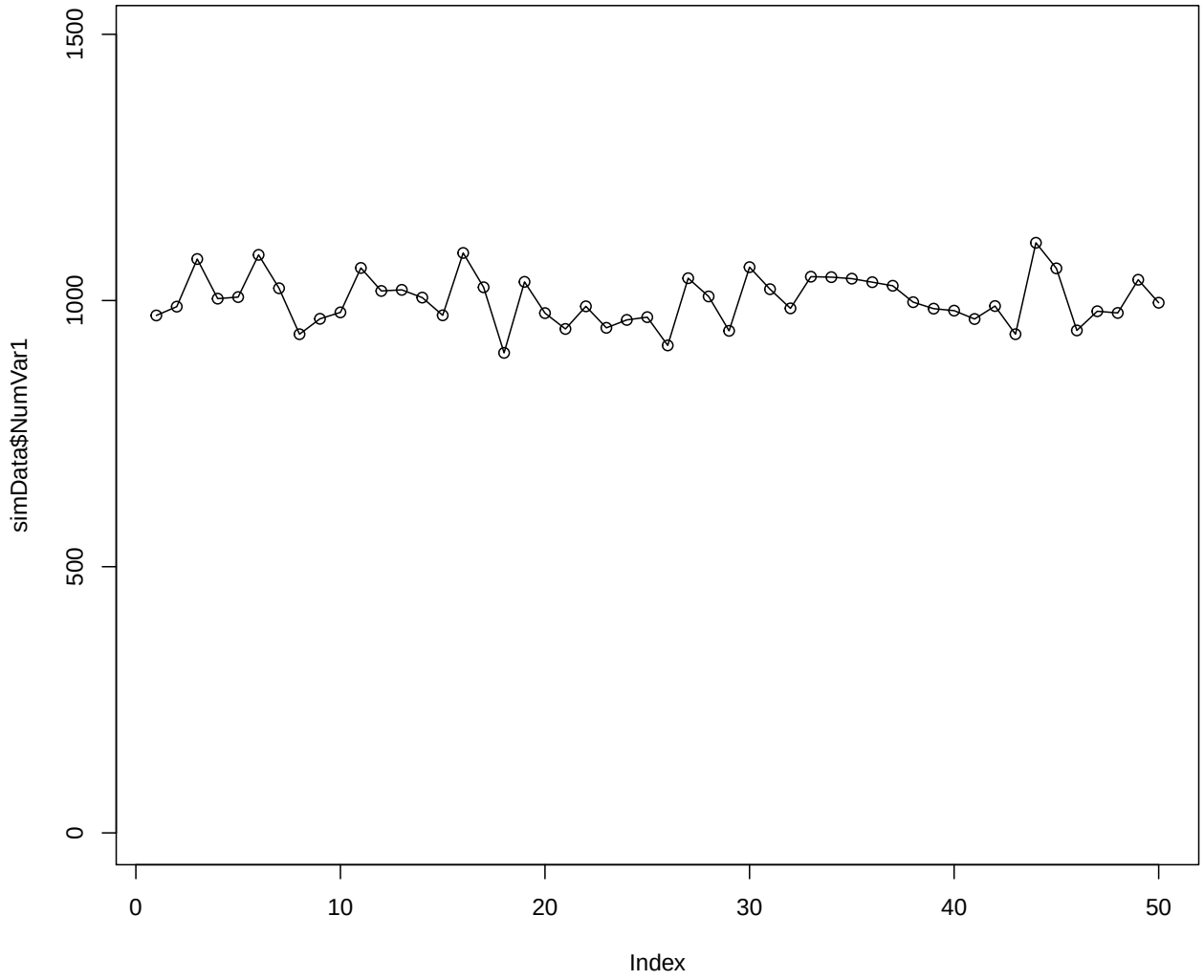


Figure 7: Index Plot of One Factor Variable

5.2 Index Plot (with two variables)

```
plot(simData$NumVar1,type="o",ylim=c(0,max(simData$NumVar1,simData$NumVar2)))
lines(simData$NumVar2,type="o",lty=2,col="red") ## add another variable
```

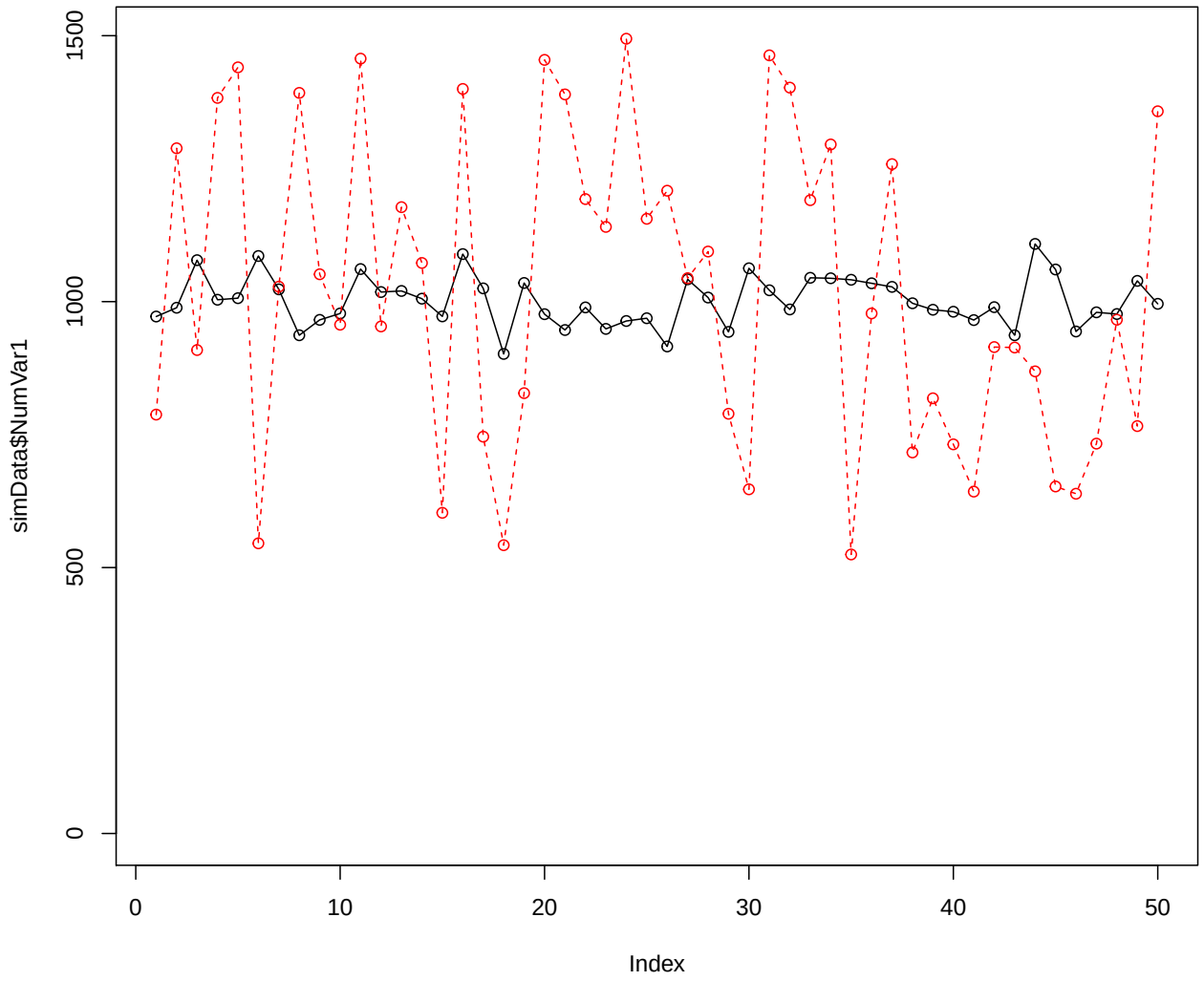


Figure 8: Index Plot of One Factor Variable: Added Variable

5.3 Density Plots

```
dv1=density(simData$NumVar1)
dv2=density(simData$NumVar2)
plot(range(dv1$x, dv2$x),range(dv1$y, dv2$y), type = "n", xlab = "NumVar1(red) and NumVar2 (blue)",
     ylab = "Density")
lines(dv1, col = "red")
lines(dv2, col = "blue")
```

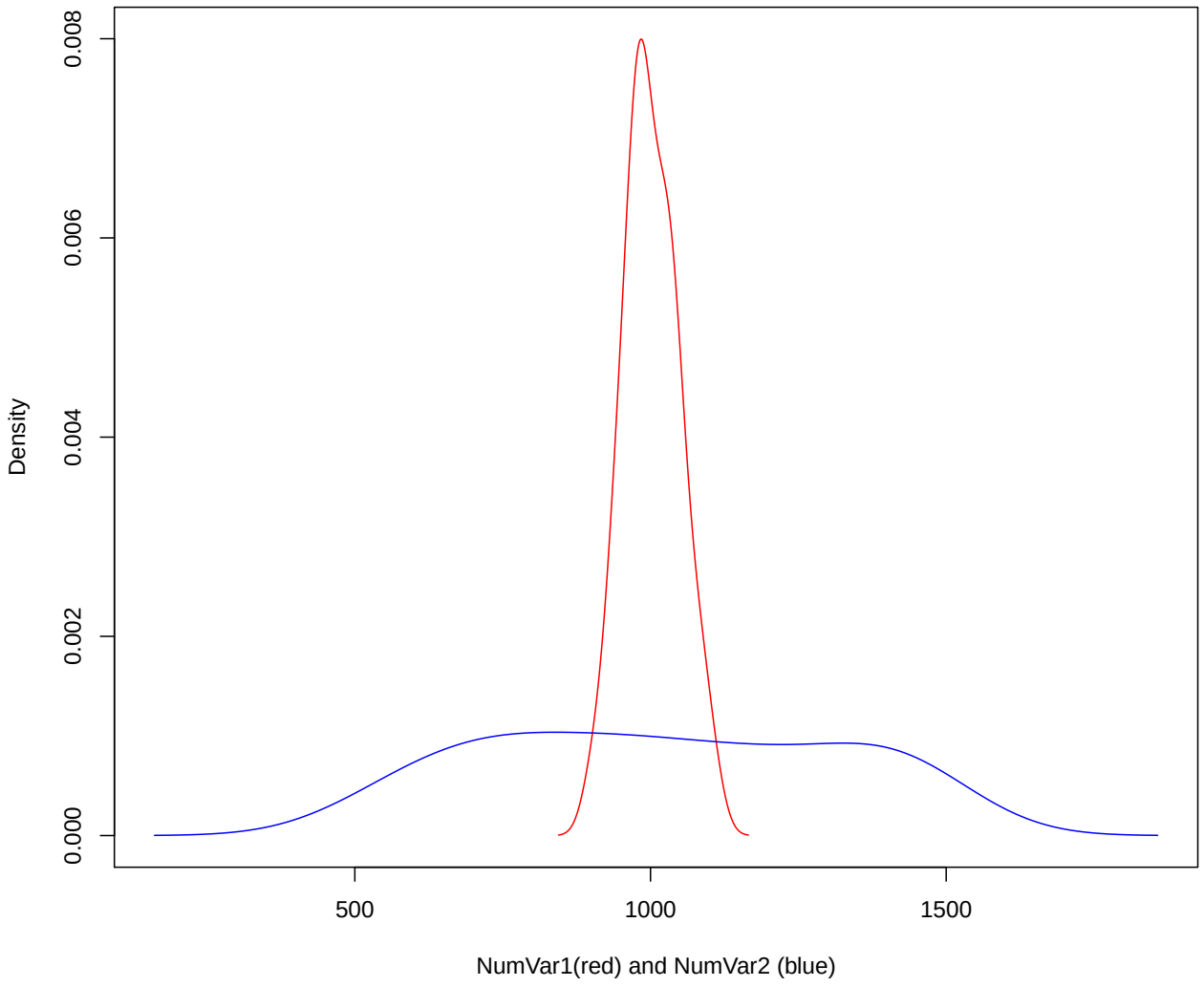


Figure 9: Density Plot of One Factor Variable

5.4 Scatter Plots

Scatter plots are used to plot data points on a horizontal and a vertical axis so as to show the effects of one variable on another. It is useful for identifying relationships, should any exist, between two variables.

```
plot(simData$NumVar1,simData$NumVar2)
```

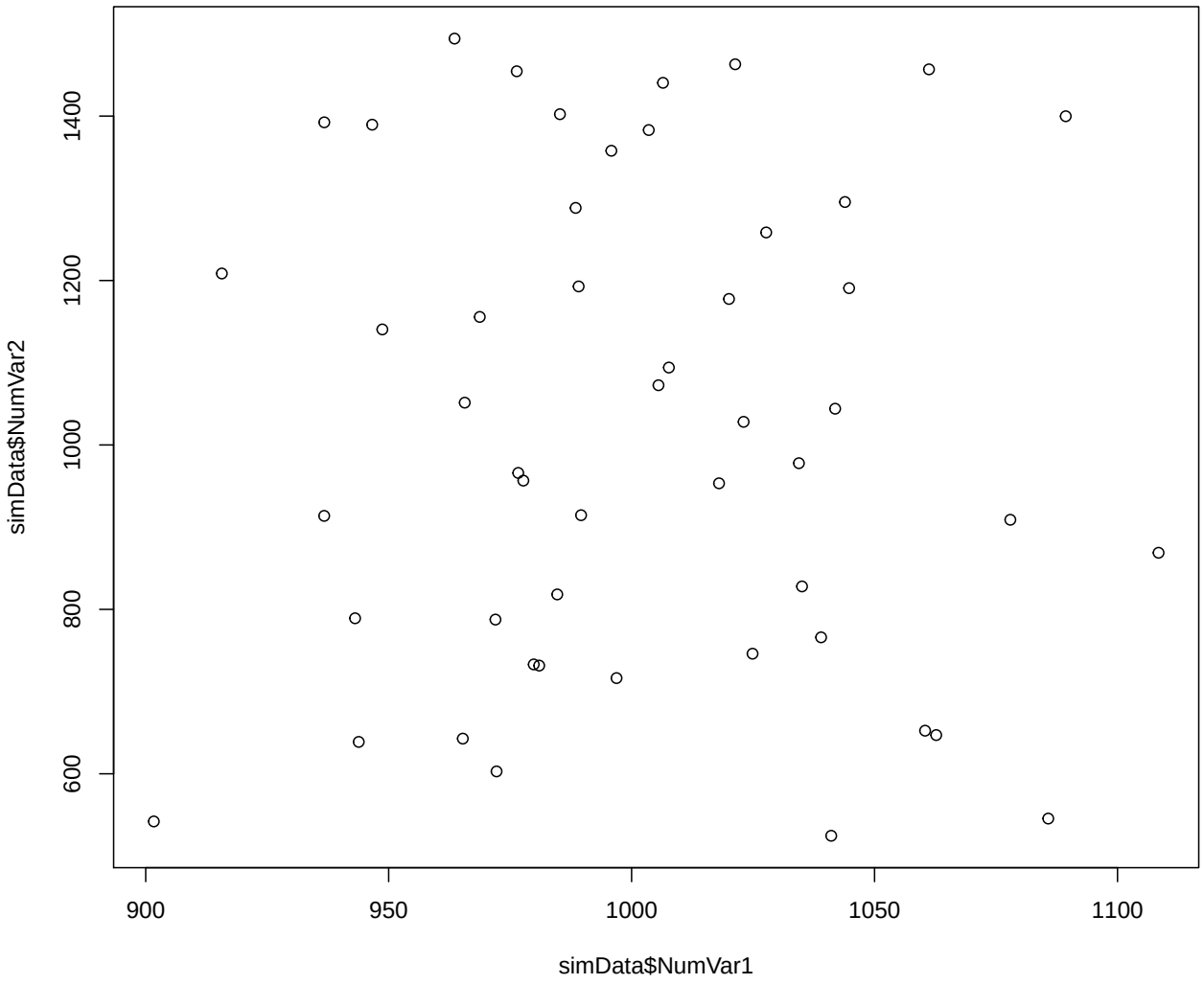


Figure 10: Scatter Plot of One Factor Variable

6 Two Factor Variables

6.1 Mosaic Plot

For enlightenment on Mosaic Plots, reference this page: http://www.vicc.org/biostatistics/LuncheonTalks/DrTsai2.pdf

```
plot(table(simData$FacVar2,simData$FacVar3))
```

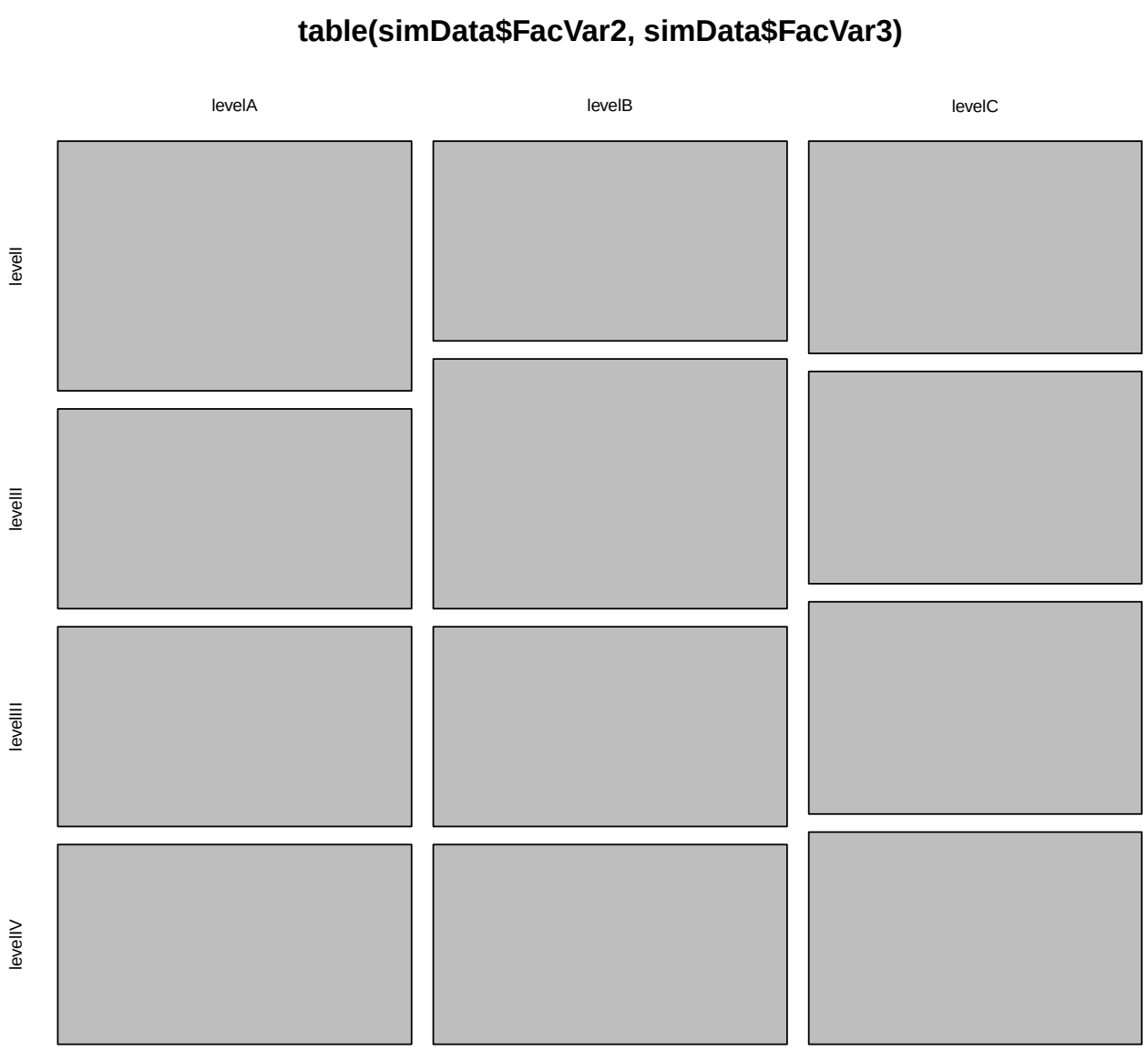


Figure 11: Mosaic Plot of Two Factor Variables

6.2 Bar Plots

```
bartable=table(simData$FacVar2,simData$FacVar3) ## get the cross tab
barplot(bartable,beside=TRUE, legend=levels(unique(simData$FacVar2))) ## plot
```

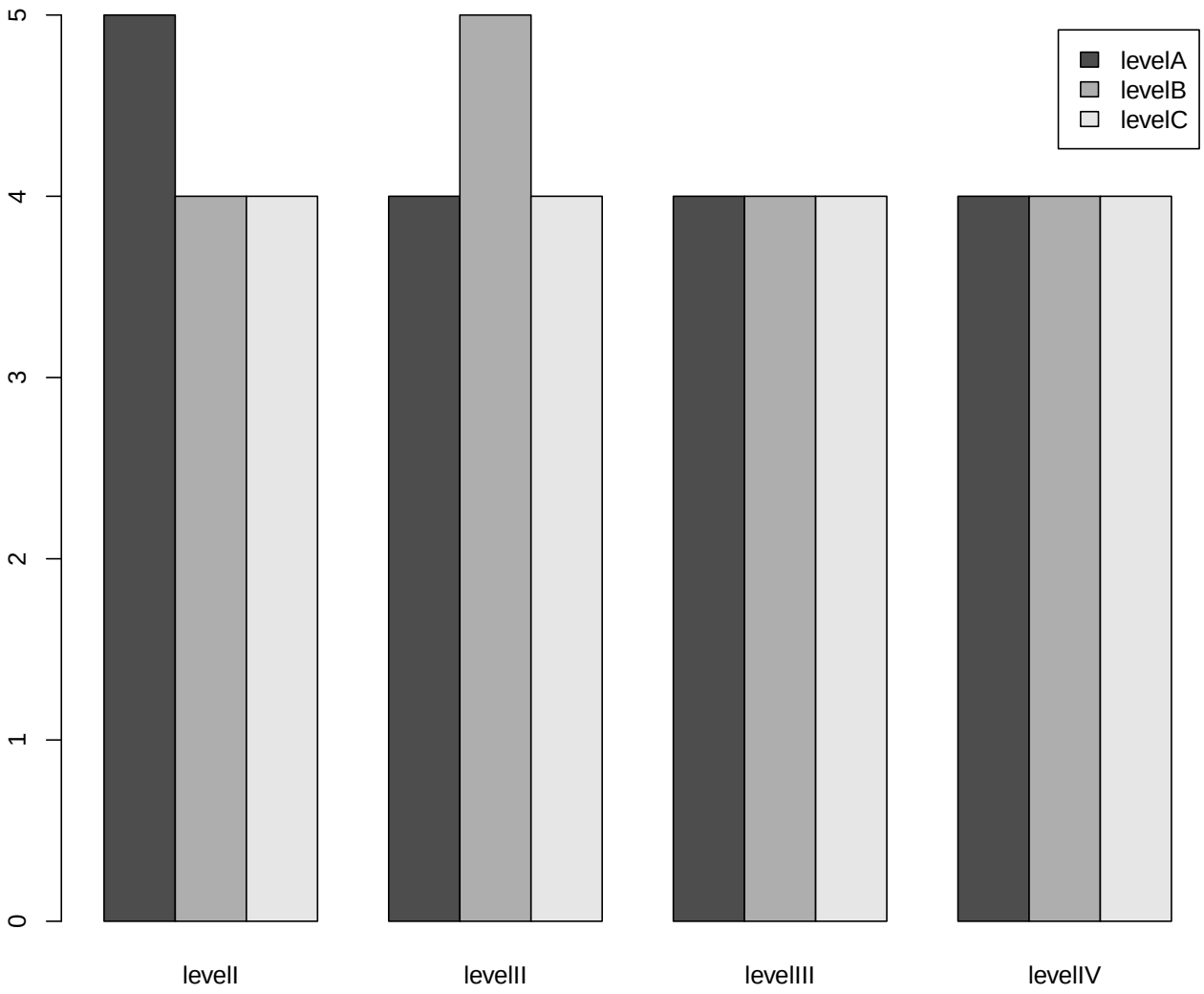


Figure 12: Bar Plots of Two Factor Variables

6.2.1 Stacked Bar Plot

```
barplot(bartable, legend=levels(unique(simData$FacVar2))) ## stacked
```

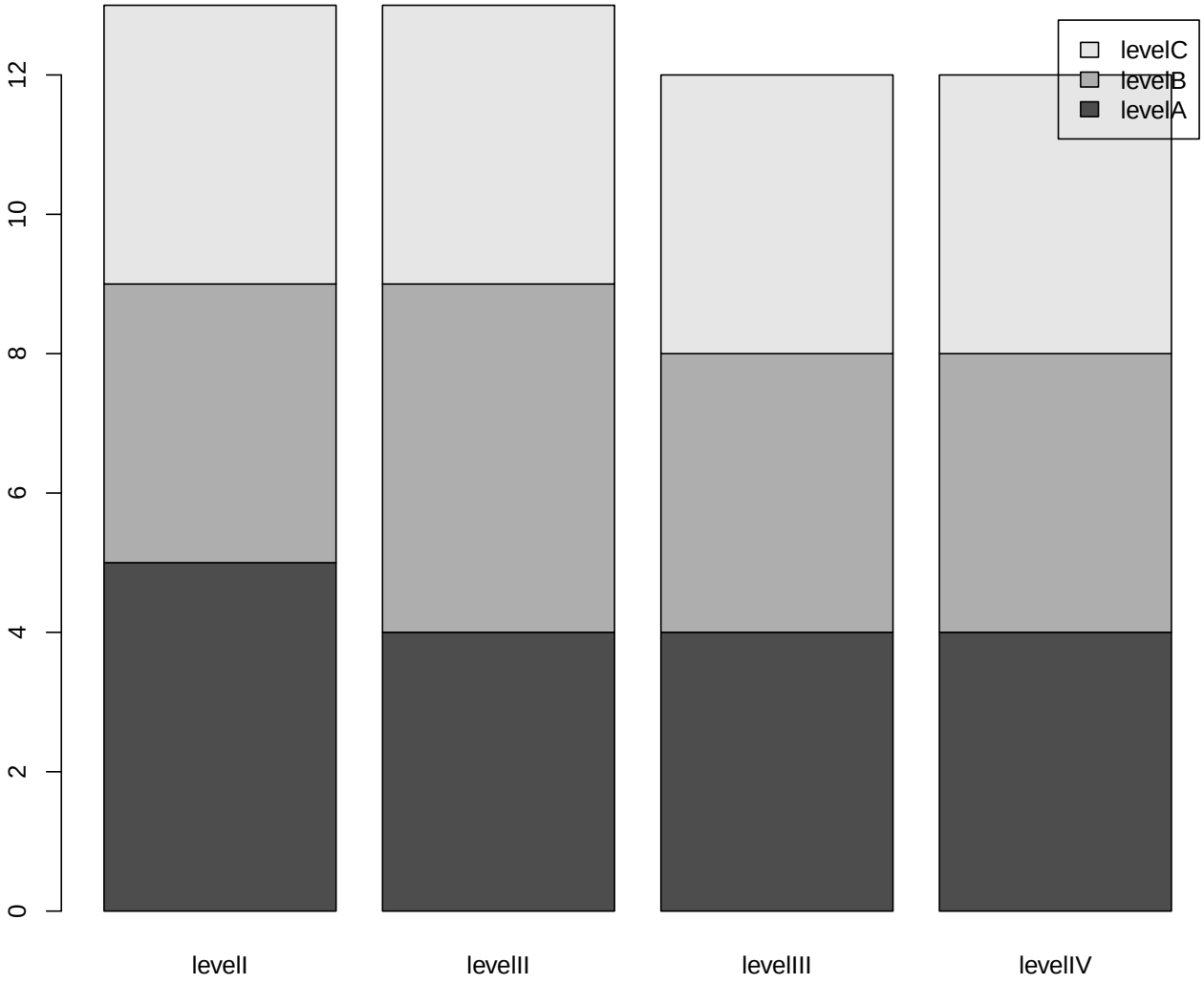


Figure 13: Stacked Bar Plots of Two Factor Variables

6.2.2 Stacked Bar Plot (100% scale)

```
barplot(prop.table(bartable,2)*100, legend=levels(unique(simData$FacVar2))) ## stacked 100%
```

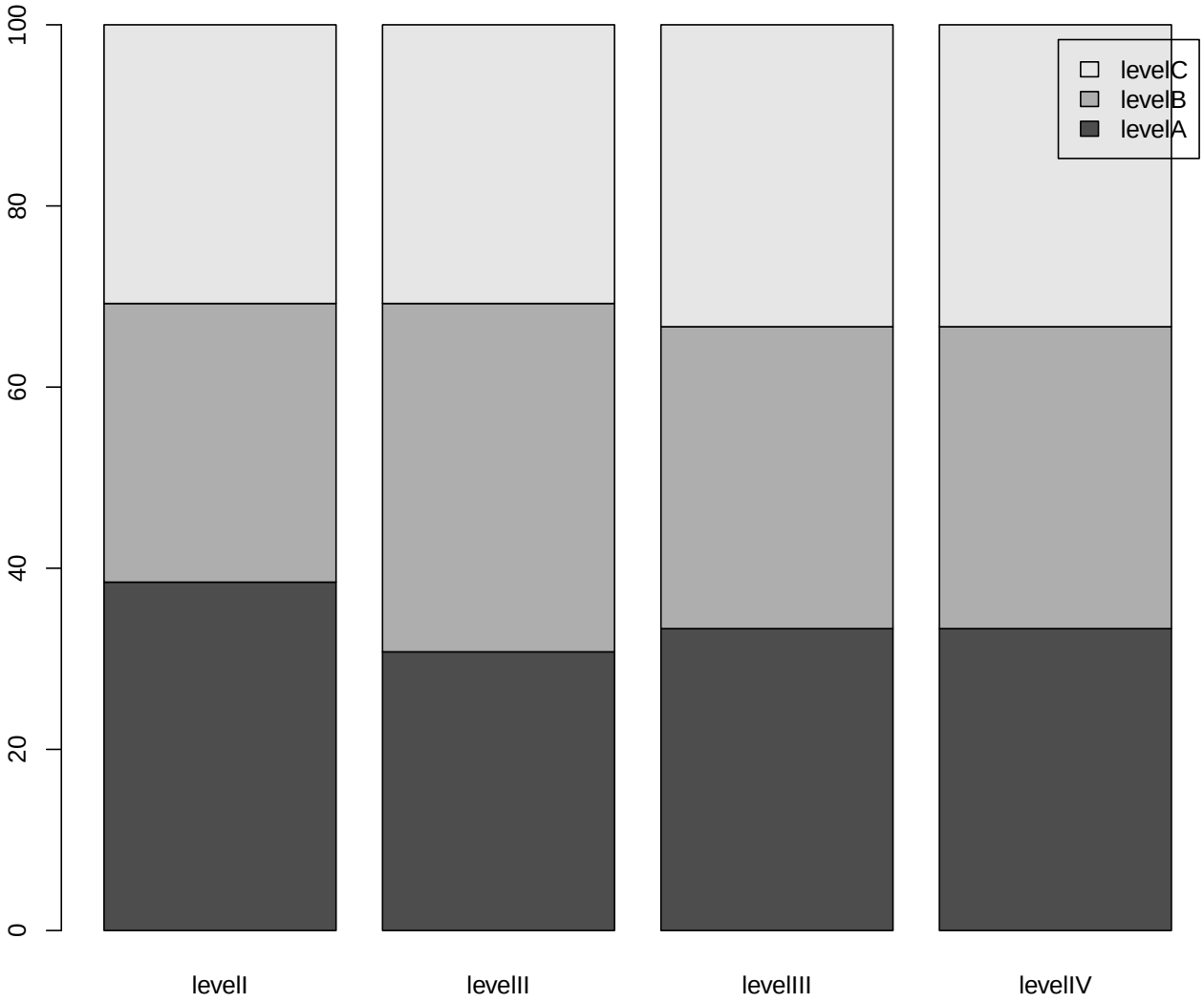


Figure 14: Stacked Bar Plots of Two Factor Variables (100% scale)

7 Two Variables: One Factor and One Numeric

7.1 Box Plots

```
## Box plots for the numeric var over the levels of the factor var
plot(simData$FacVar1,simData$NumVar1)
```

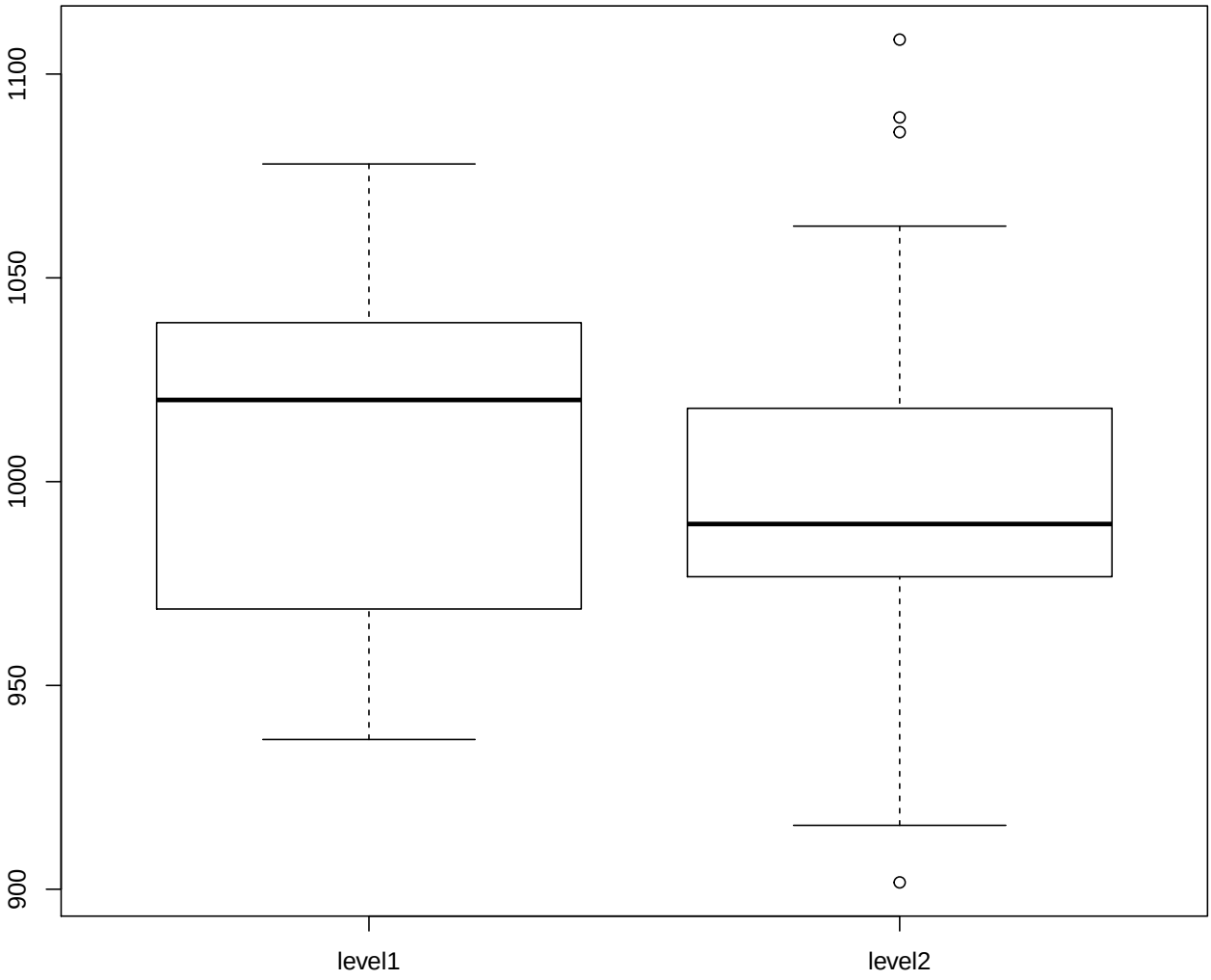


Figure 15: Box Plots of One Factor Variable and One Numeric Variable

7.2 Data Density Plot

```
level1=simData[simData$FacVar1=="level1",]
level2=simData[simData$FacVar1=="level2",]

dv3=density(level1$NumVar1)
dv4=density(level2$NumVar1)

plot(range(dv3$x, dv4$x),range(dv3$y, dv4$y), type = "n", xlab = "NumVar1 at Level1 (red) and NumVar1 a
lines(dv3, col = "red")
lines(dv4, col = "blue")
```

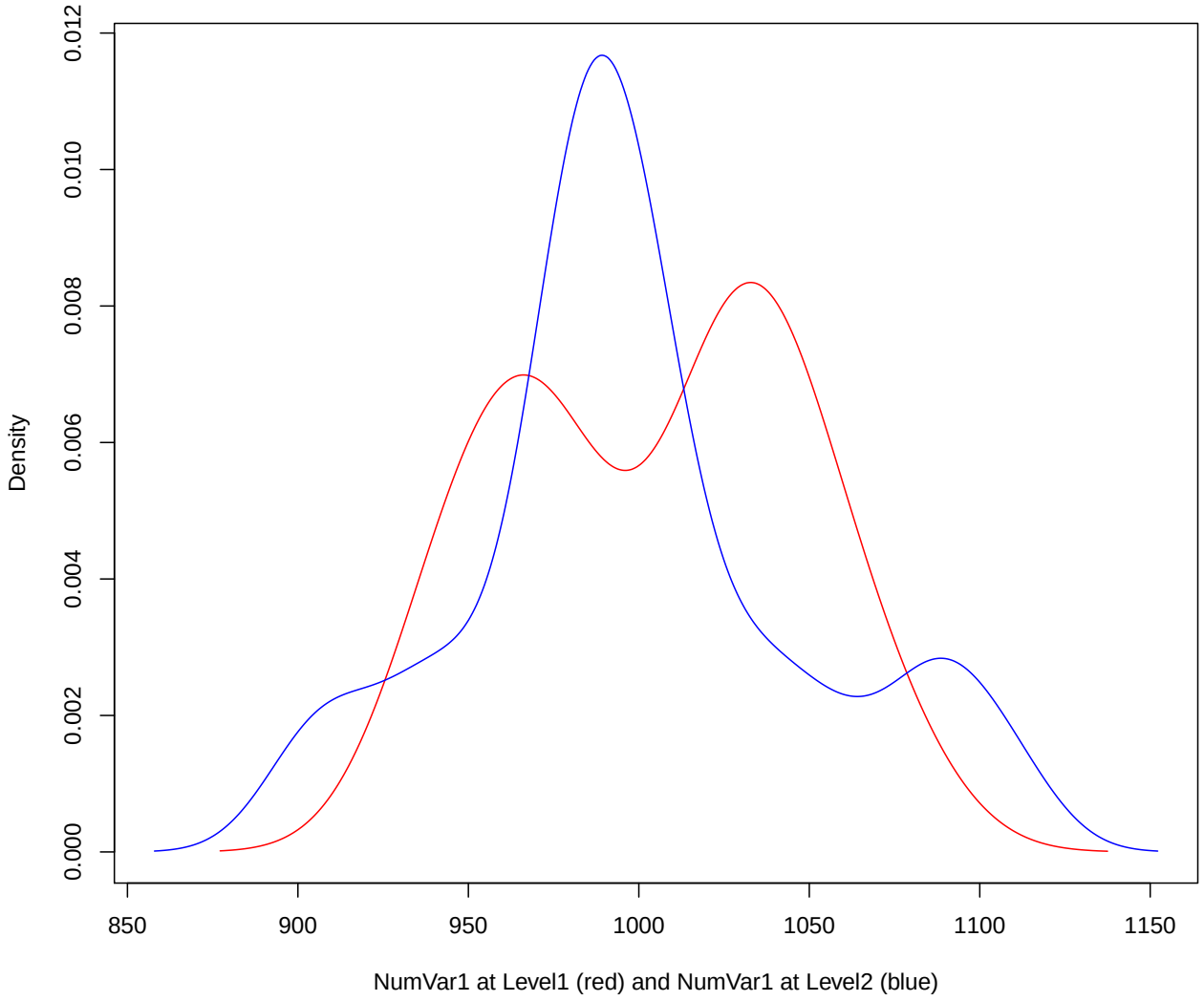


Figure 16: Data Density Plot of One Factor Variable and One Numeric Variable

7.3 Dotchart

```
meanagg=aggregate(simData$NumVar1, list(simData$FacVar3), mean) #Mean of one numeric var over levels of
```

```
dotchart(meanagg$x,labels=meanagg$Group.1)
```

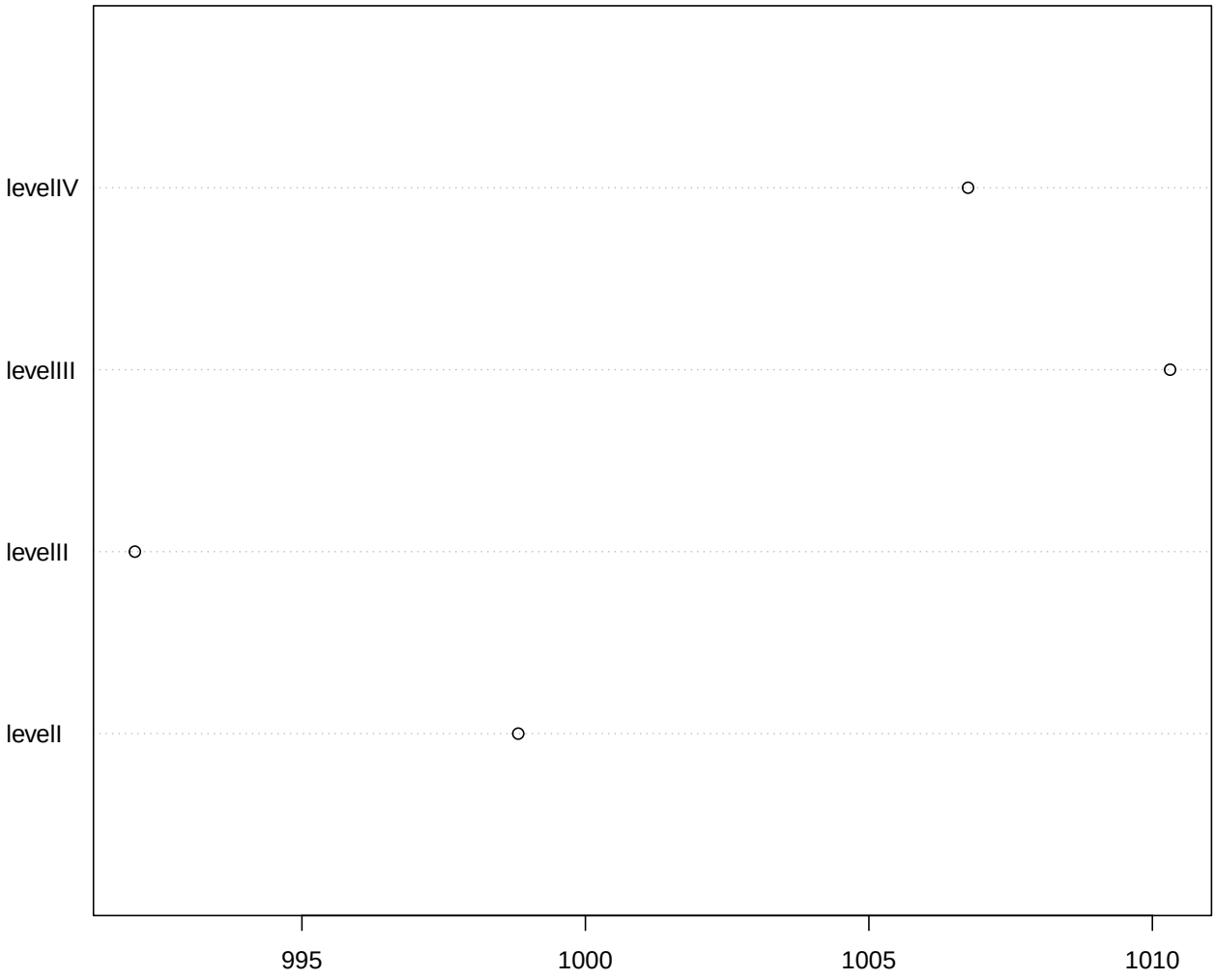
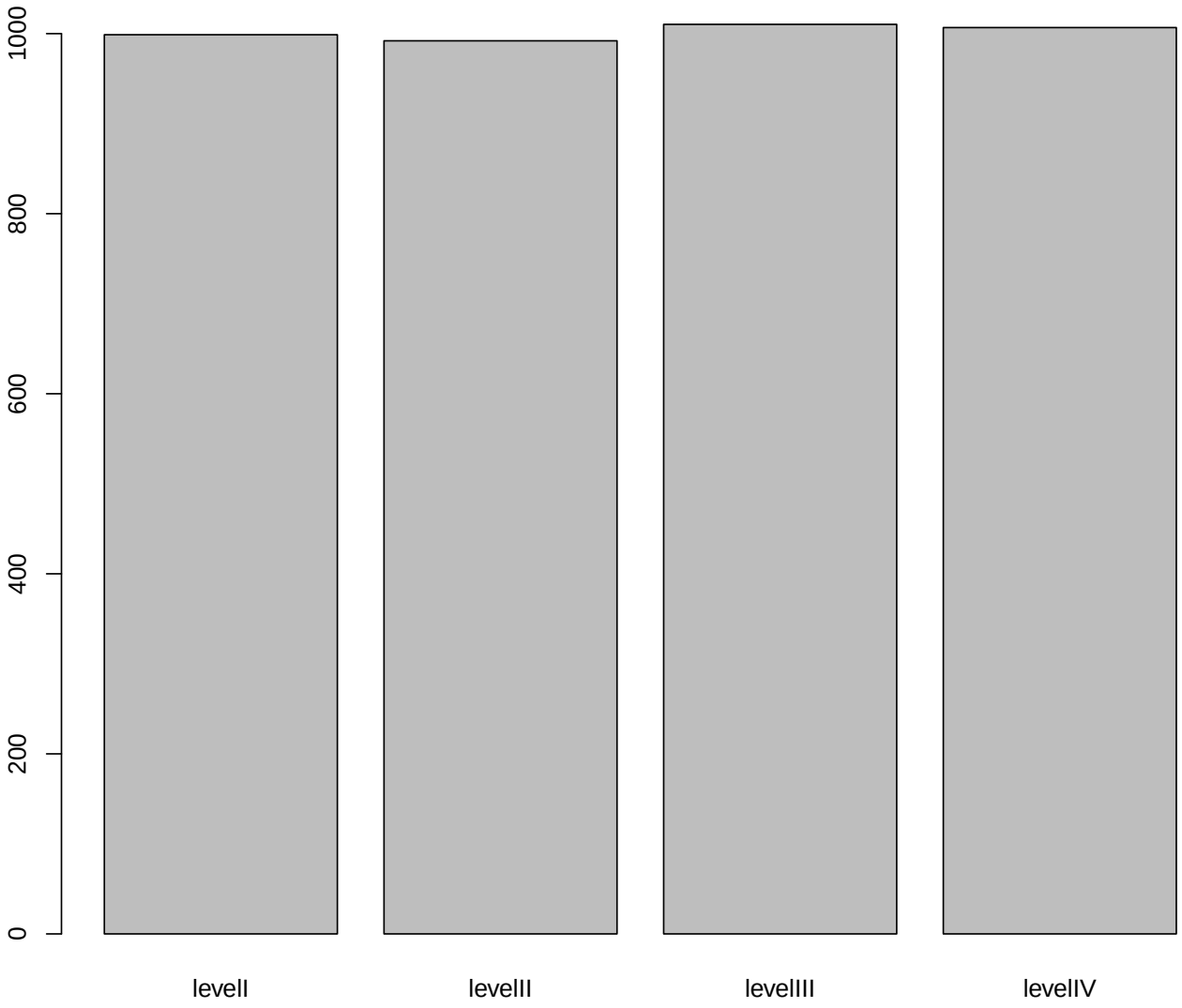


Figure 17: Dot Chart of One Factor Variable and One Numeric Variable

7.4 Bar Plot

```
barplot(meanagg$x,names.arg=meanagg$Group.1)
```



8 Three Variables: Three Factor Variables

8.1 Boxplot

```
par(mfrow=c(1,2))

bar1table=table(level1$FacVar2,level1$FacVar3)
barplot(bar1table,beside=TRUE, main="FacVar1=level1")

bar2table=table(level2$FacVar2,level2$FacVar3)
barplot(bar2table,beside=TRUE, main="FacVar1=level2", legend=levels(unique(level2$FacVar2)))
```

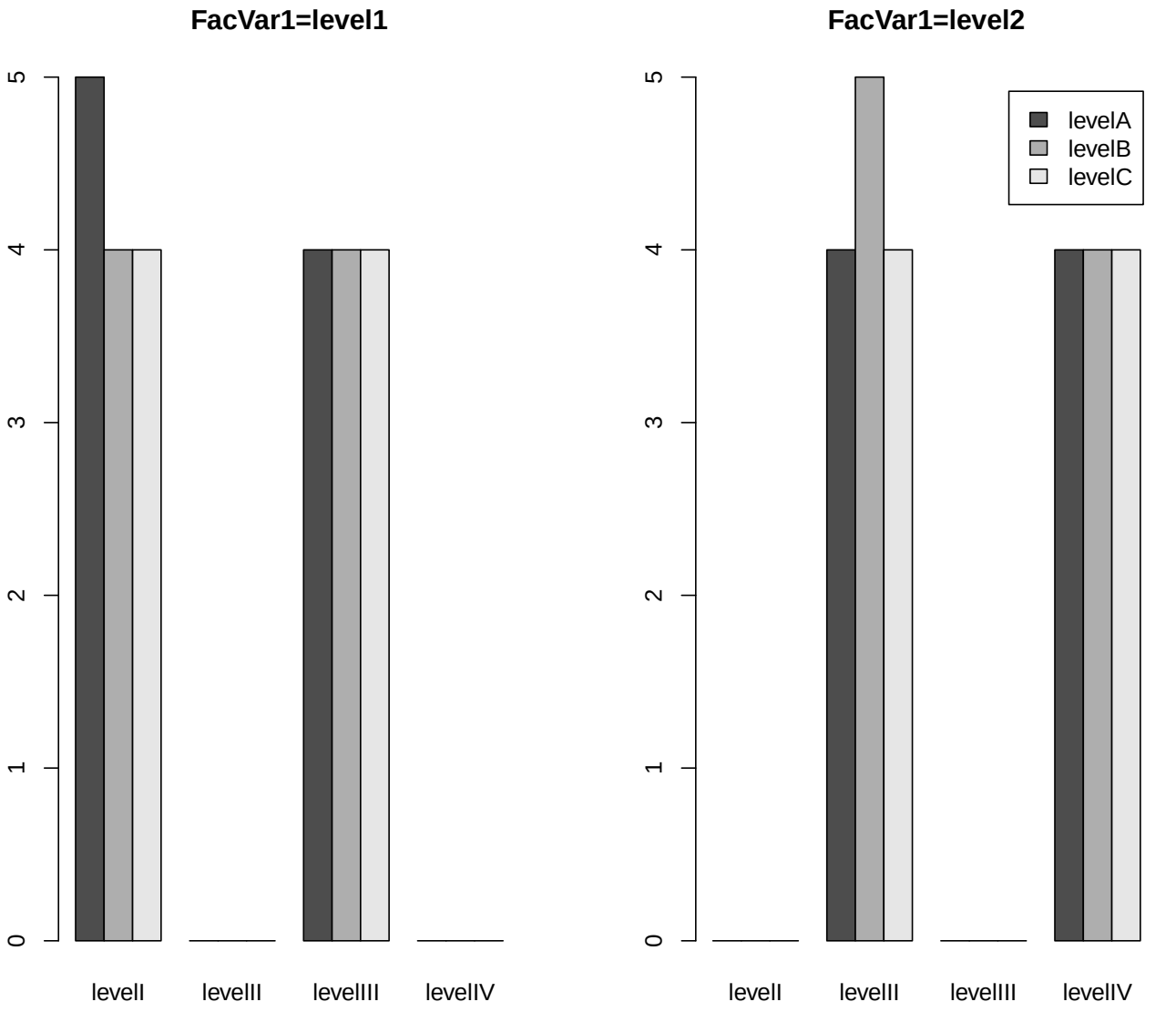


Figure 18: Box Plots of Three Factor Variables

9 Three Variables: One Numeric and Two Factor Variables

9.1 Box Plot

```
par(mfrow=c(1,1))
## boxplot of NumVar1 over an interaction of 6 levels of the combination of FacVar1 and FacVar2
boxplot(NumVar1~interaction(FacVar1,FacVar2),data=simData)
```

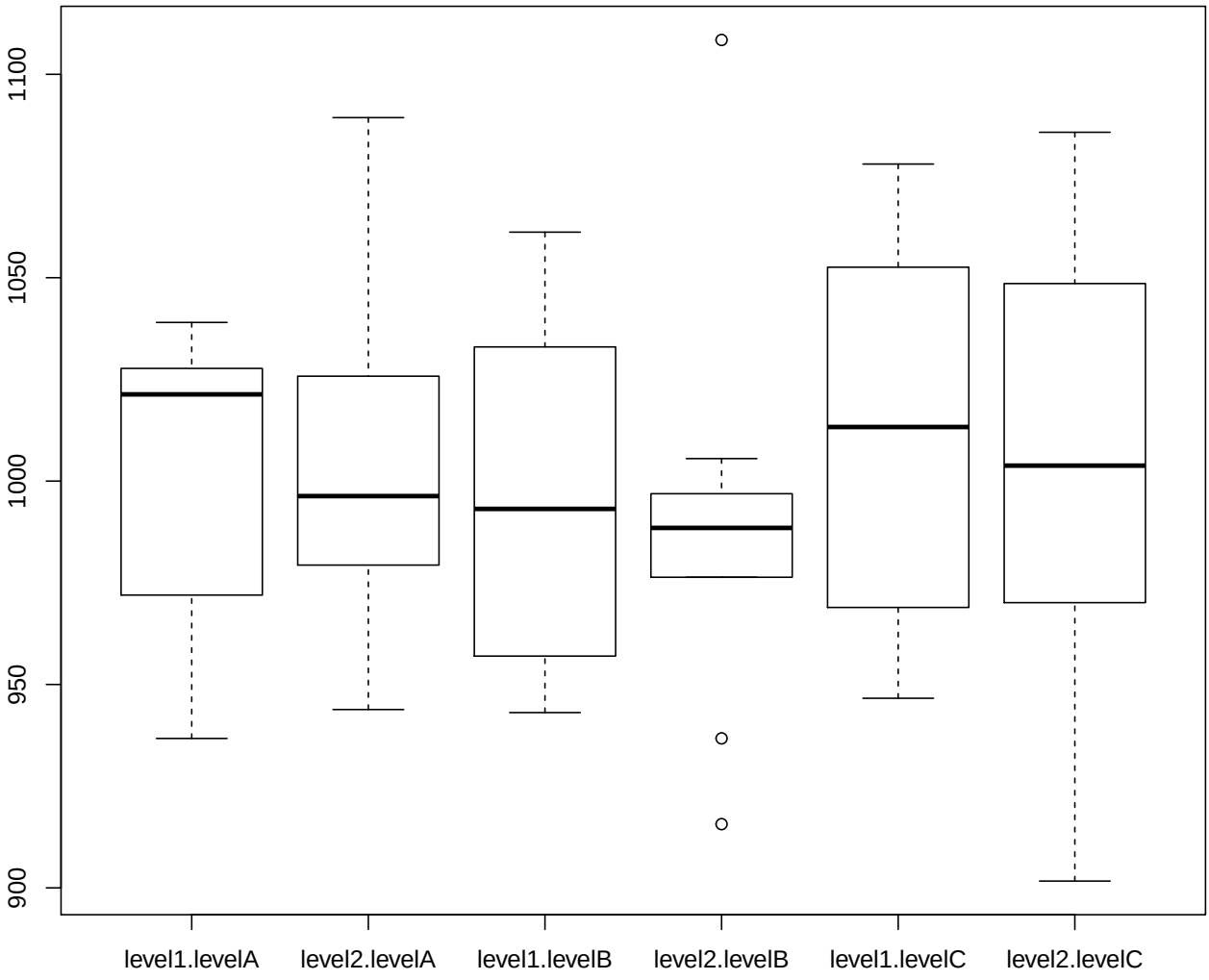


Figure 19: Box Plots of One Numeric Variable and Two Factor Variables Variables

9.2 Dot Chart of Mean of 1 Numeric over levels of two factor vars

```
meanaggg=aggregate(simData$NumVar1, list(simData$FacVar1,simData$FacVar2), mean)
meanaggg=meanaggg[order(meanaggg$Group.1),]
meanaggg$color[meanaggg$Group.2=="levelA"] = "red"
meanaggg$color[meanaggg$Group.2=="levelB"] = "blue"
meanaggg$color[meanaggg$Group.2=="levelC"] = "darkgreen"

dotchart(meanaggg$x,labels=meanaggg$Group.2, groups=meanaggg$Group.1,color=meanaggg$color) ## dotchart
```

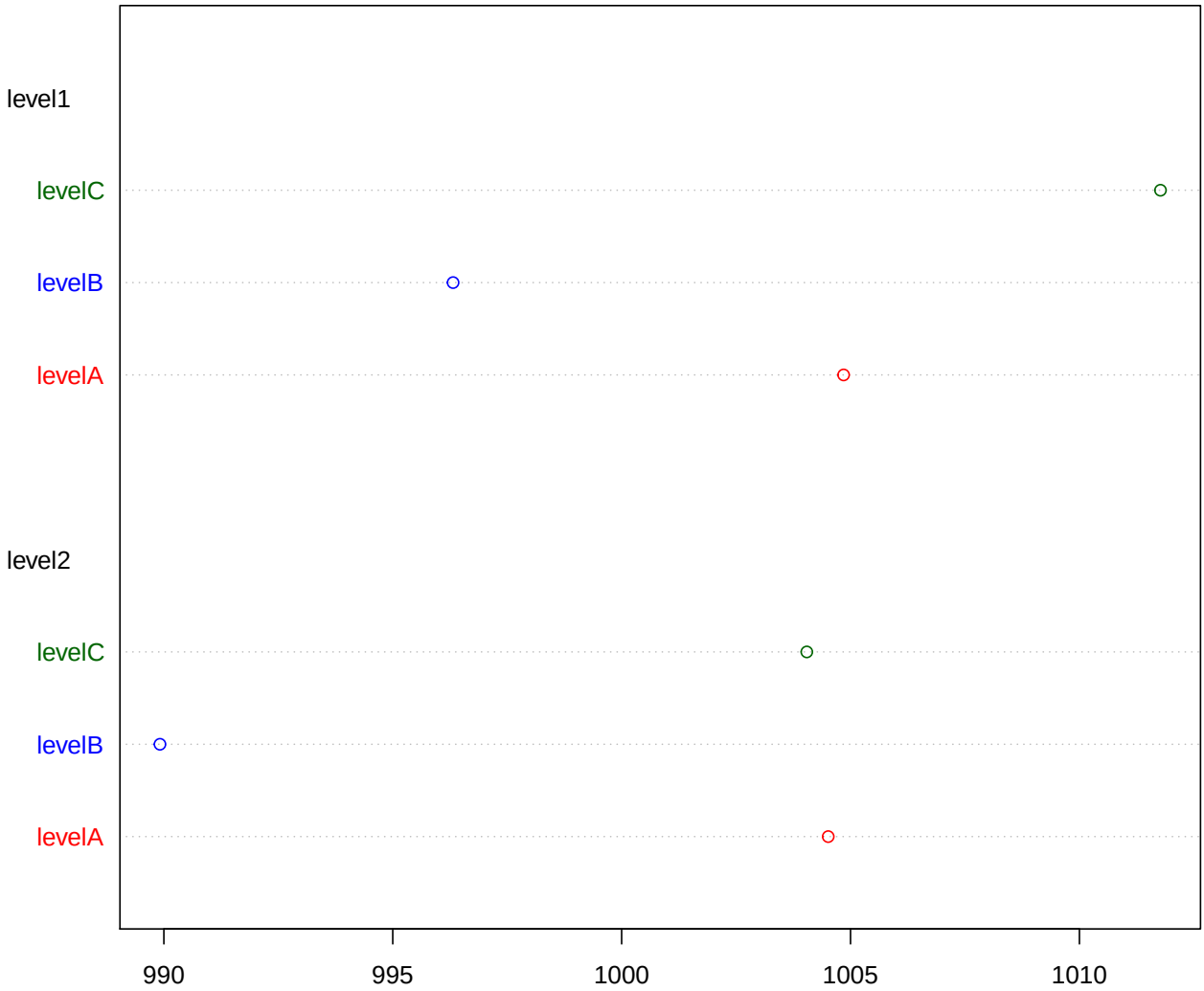


Figure 20: Dot Chart of One Numeric Variable and Two Factor Variables

9.3 Interaction Plot

```
interaction.plot(meanaggg$Group.2,meanaggg$Group.1,meanaggg$x,type="b", col=c(1:2),pch=c(18,24)) ## int
```

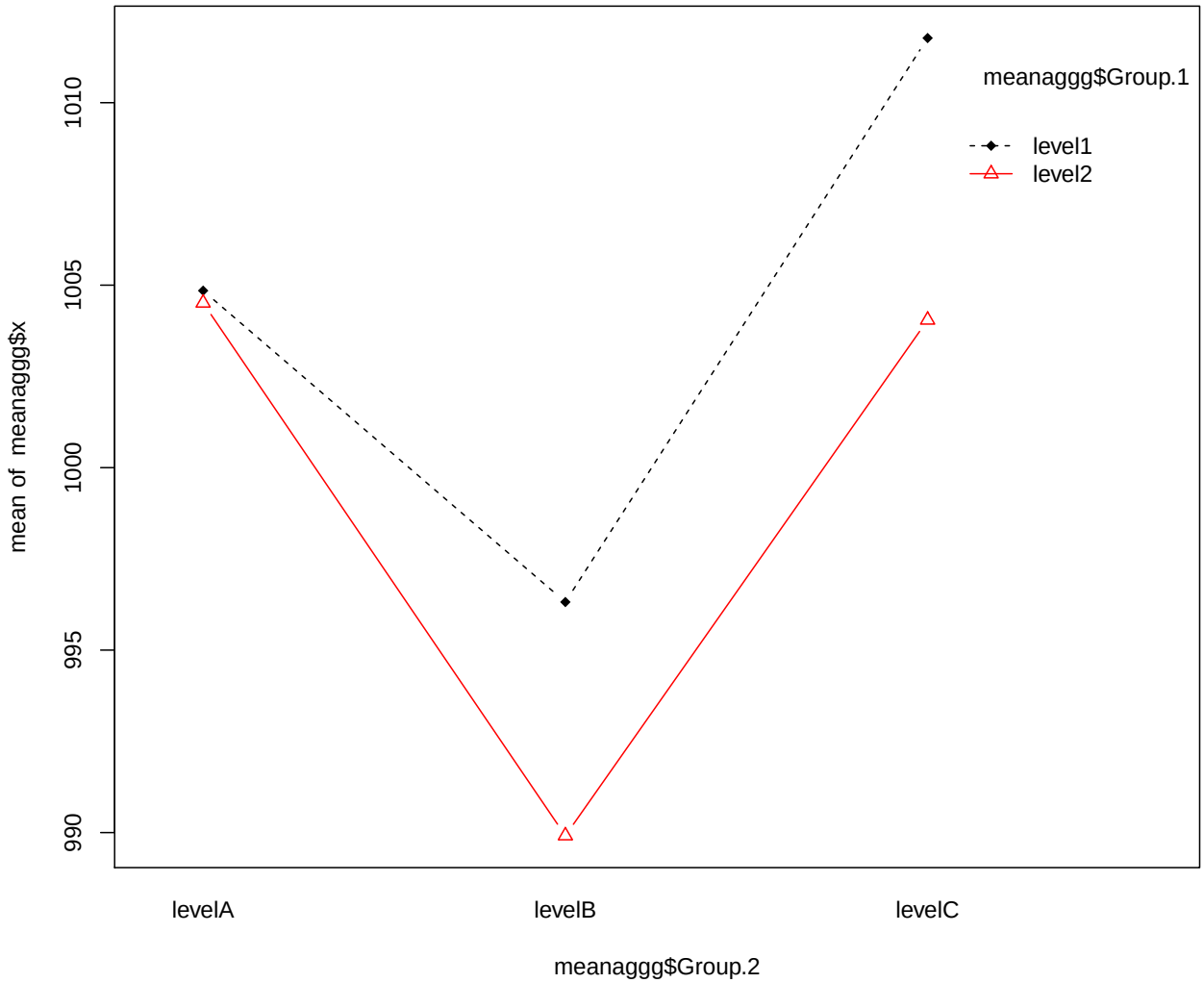


Figure 21: Interaction Plots of One Numeric Variable and Two Factor Variables

9.4 some a bar plot

```
par(mfrow=c(1,2))

level1=meanaggg[meanaggg$Group.1=="level1",]
level2=meanaggg[meanaggg$Group.1=="level2",]

barplot(level1$x,names.arg=level1$Group.2, main="FacVar1=level1")
barplot(level2$x,names.arg=level2$Group.2, main="FacVar1=level2")
```

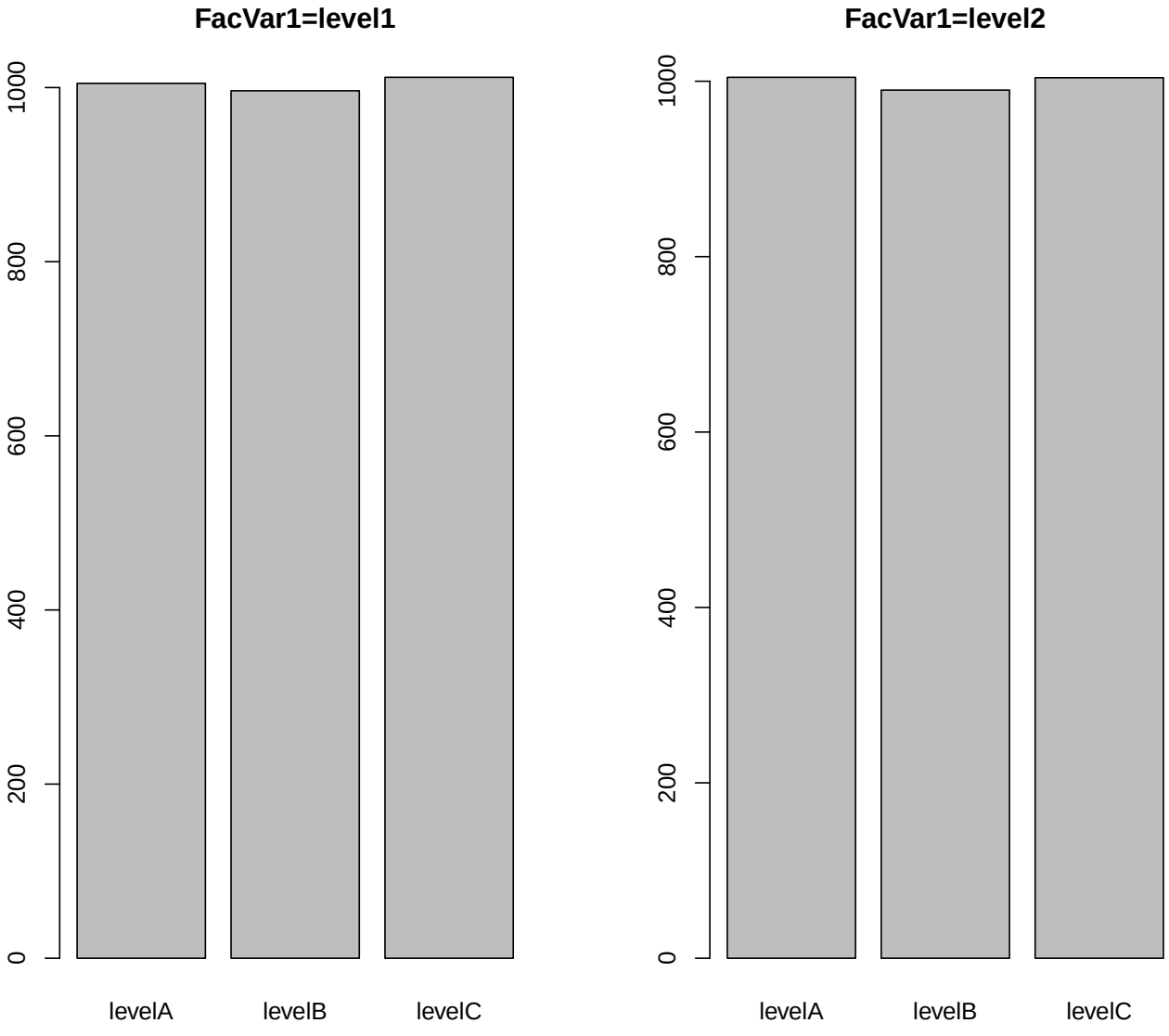


Figure 22: plot of chunk unnamed-chunk-26

10 Three Variables: Two Numeric and One Factor Variable

10.1 Scatter Plot

```
par(mfrow=c(1,1))
plot(simData$NumVar1,simData$NumVar2, col=simData$FacVar1)
legend("topright",levels(simData$FacVar1),fill=simData$FacVar1)
```

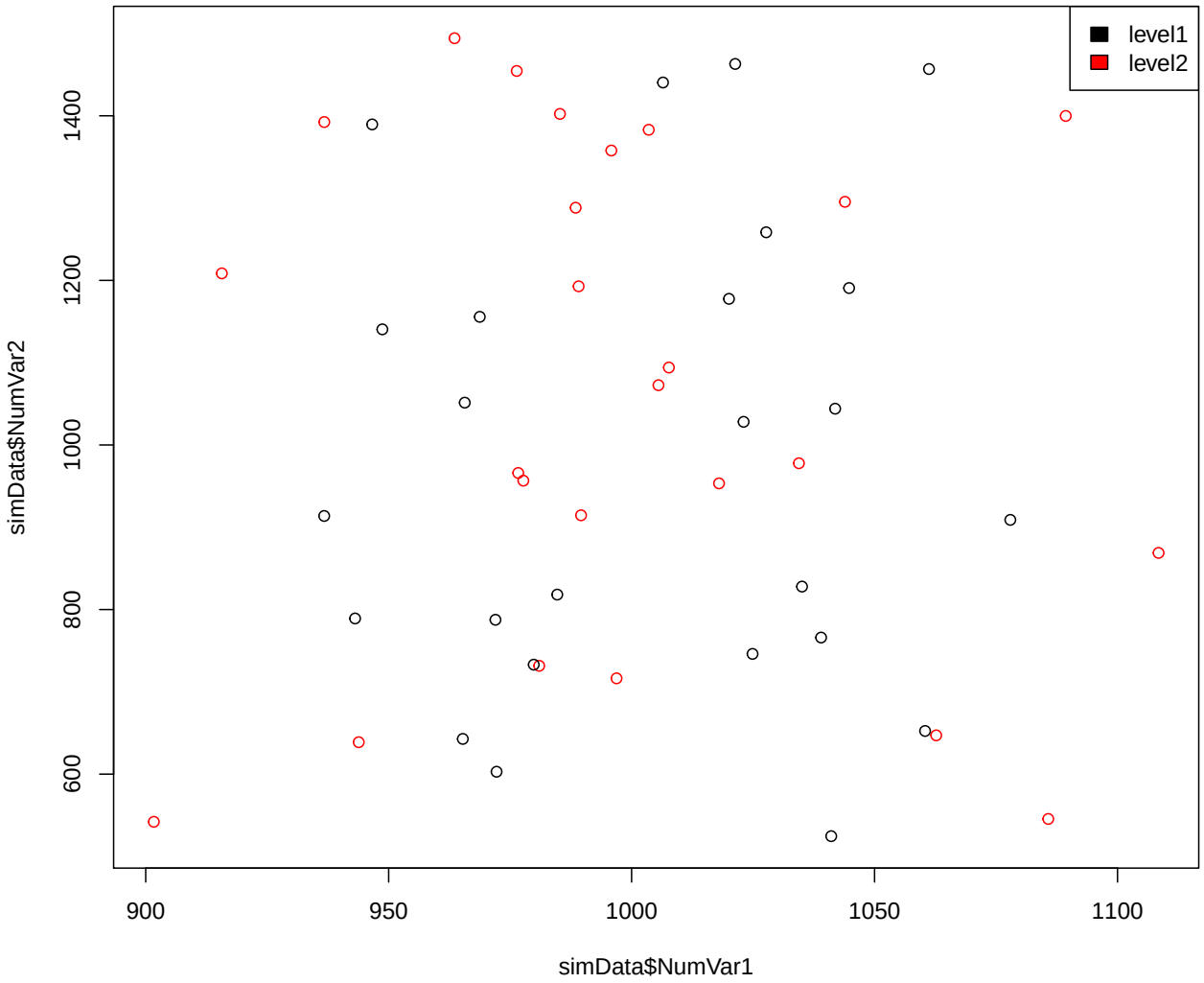


Figure 23: Scatter plot with color identifying the factor variable

11 Three Variables: Three Numeric Variables

11.1 Plot and Bubble Plot

```
## NumVar4 is 2001 through 2050... possibly, a time variable - use that as the x-axis
plot(simData$NumVar4,simData$NumVar1,type="o",ylim=c(0,max(simData$NumVar1,simData$NumVar2)))## join do

lines(simData$NumVar4,simData$NumVar2,type="o",lty=2,col="red")## add another line
```

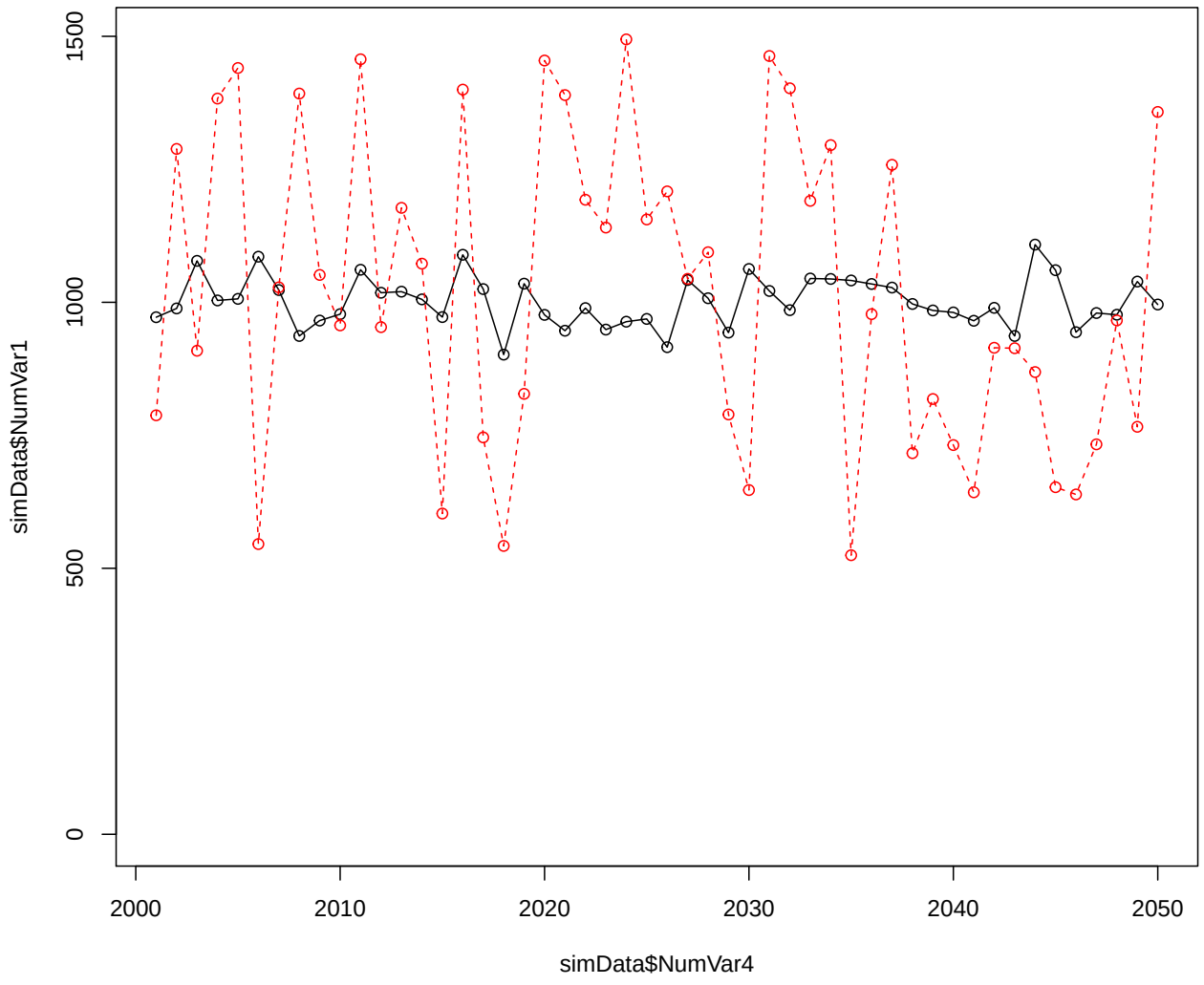


Figure 24: plot of chunk unnamed-chunk-28

```
## Bubble plot - scatter plot of NumVar1 and NumVar2 with individual observations sized by NumVar3
# http://flowingdata.com/2010/11/23/how-to-make-bubble-charts/
radius <- sqrt( simData$NumVar3/ pi )
symbols(simData$NumVar1,simData$NumVar2,circles=radius, inches=.25,fg="white", bg="red", main="Sized by
```

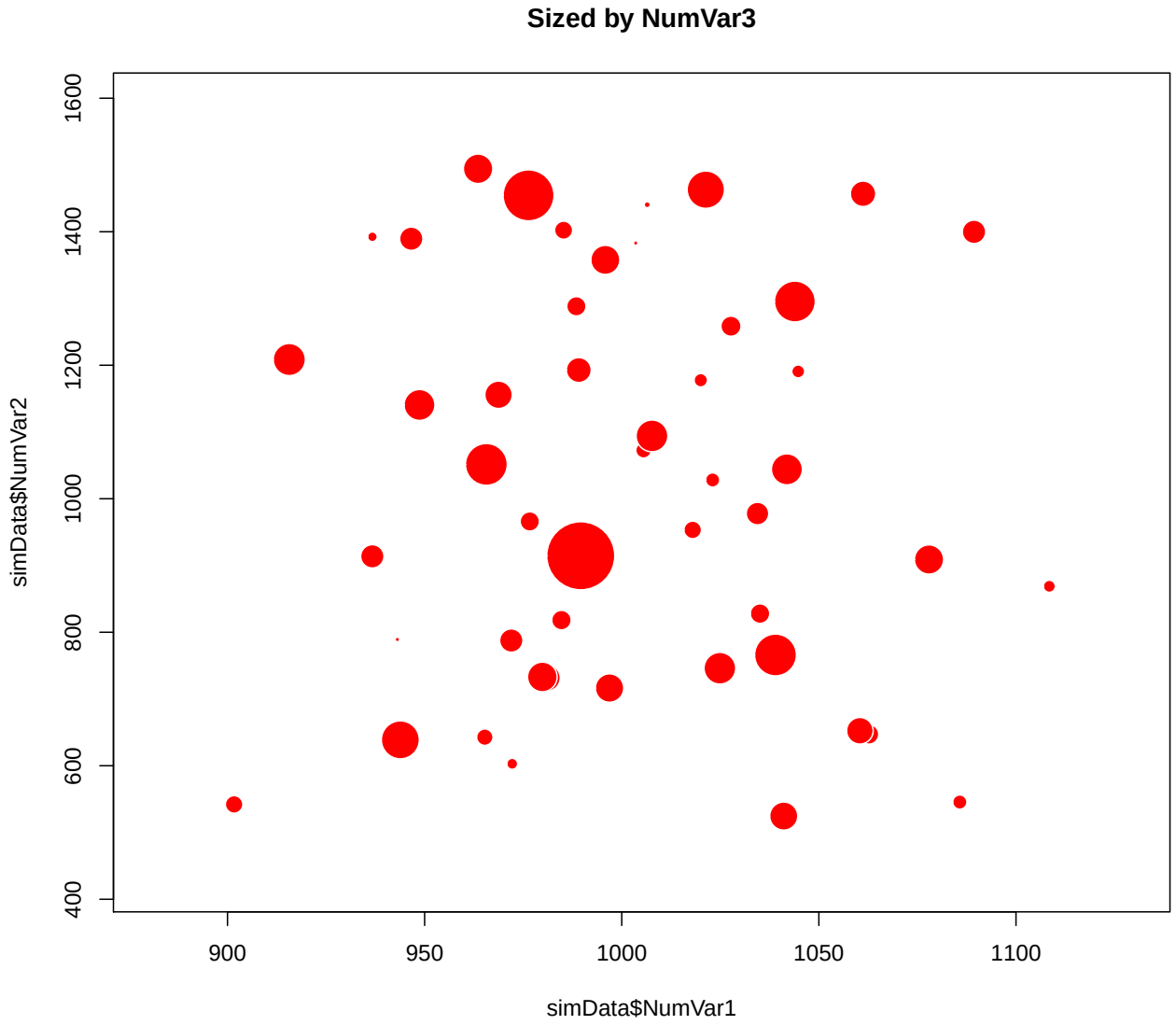


Figure 25: plot of chunk unnamed-chunk-28

12 Scatterplot Matrix of all Numeric Variables

```
pairs(simData[,4:7], col=simData$FacVar1)
```

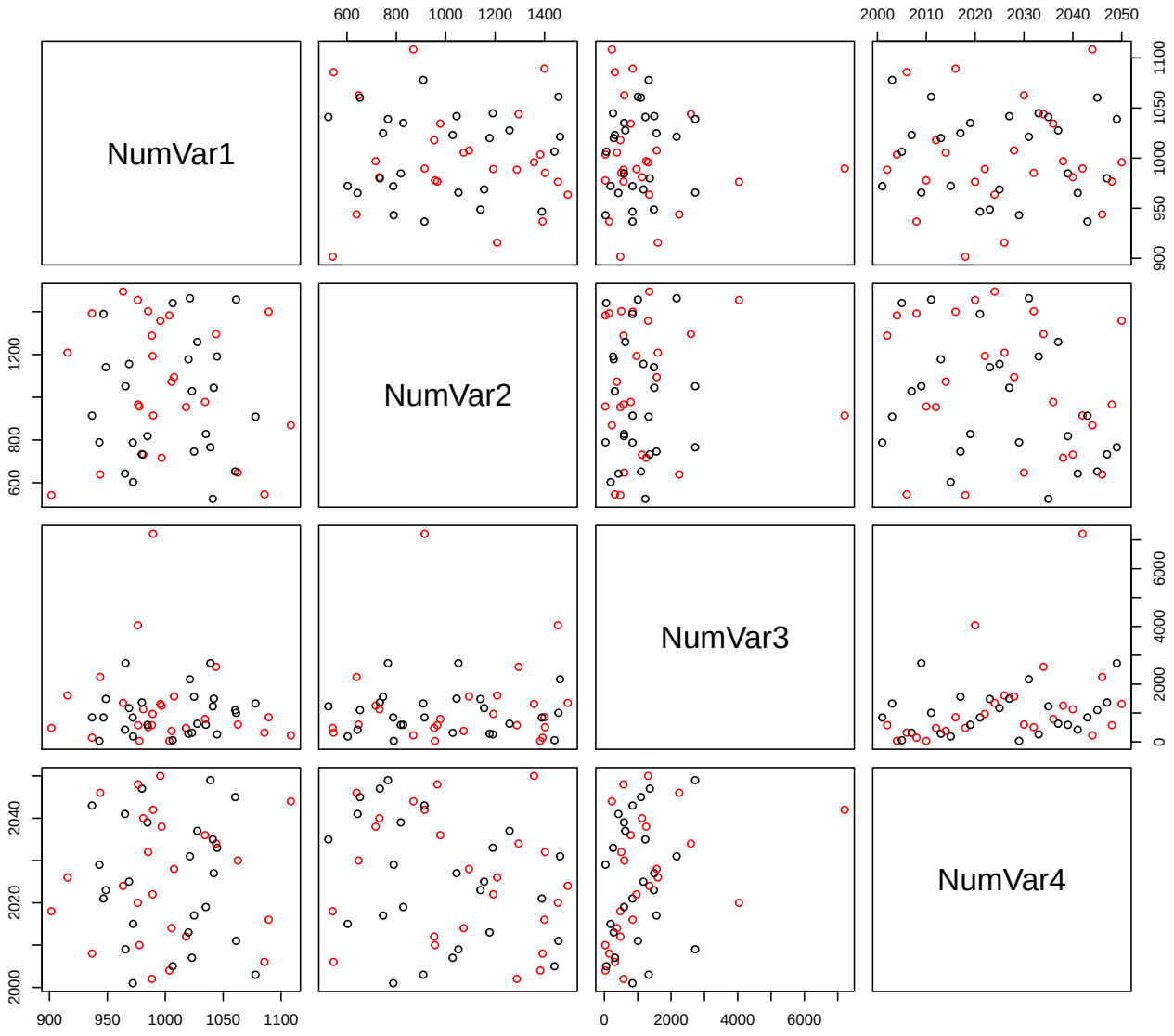


Figure 26: Scatterplot Matrix of all Numeric Variables

```
##colored by a factor Variable
```