

Publication quality plots

Heath Blackmon

last updated: 2016-01-29

Contents

Plotting our data is one of the best ways to quickly explore it and the various relationships between variables, and is integral to telling a compelling story in your manuscripts.

1. Graphing environments in R

R has three frequently used graphing environments. Each of these (`base`, `lattice`, `ggplot2`) has strengths and weaknesses. Most of the graphs that you see in publications can be produced using the base package and that is what we will focus on learning today. Regardless of the packages that you use to make your plots you build graphs in R in an iterative or layered approach. First you plot the basic data to be displayed and then you begin to add additional items to this.

2. Types of graphs

3. Generating graphs

First lets load in some example data to work with. This data is from a study of the impact of using computers to take notes in a classroom. It includes the score students made on a test plus the method they used for taking notes and their sex. [scores.csv](#)

```
data <- read.csv("scores.csv")
```

When you first begin to work with a dataset it is often a good idea to do some general exploration of your data. So lets start out by examining the data we have read in and see what its structure looks like.

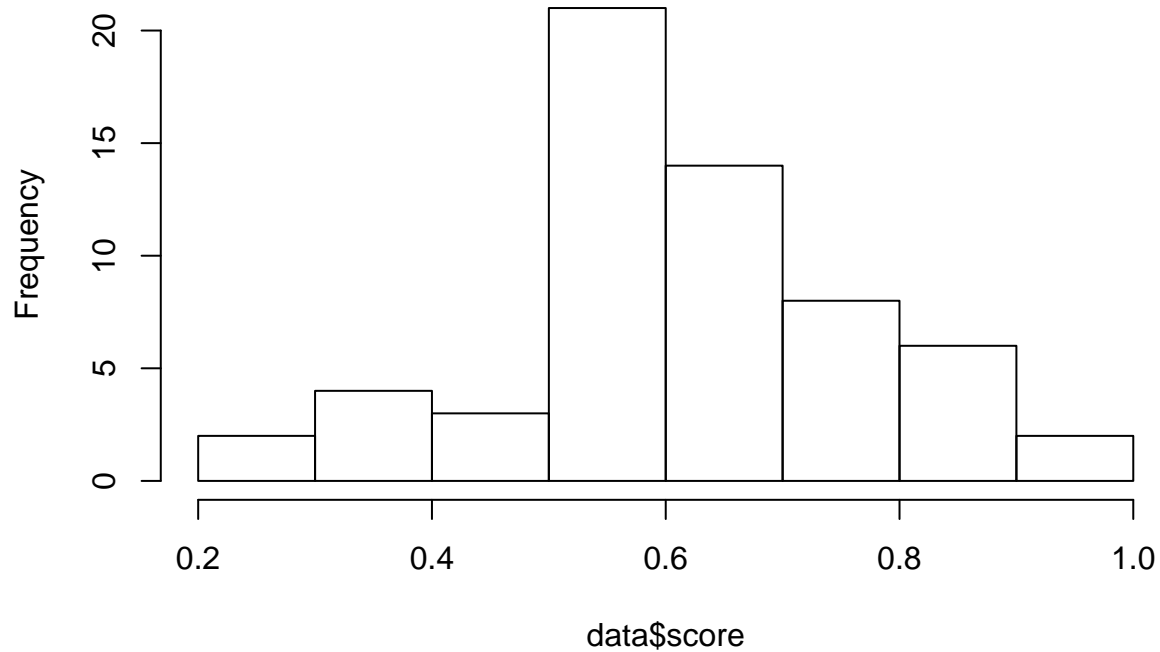
```
head(data)
```

```
##   score method   sex
## 1  0.33  comp.  male
## 2  0.52  comp.  male
## 3  0.20  comp.  male
## 4  0.73  paper  male
## 5  0.32  comp.  male
## 6  0.96  paper female
```

Now lets just look at a simple histogram of the test scores:

```
hist(data$score)
```

Histogram of data\$score



We might want to know if males and females scored differently on the test:

```
t.test(data$score ~ data$sex)
```

```
##
## Welch Two Sample t-test
##
## data: data$score by data$sex
## t = 1.4555, df = 56.48, p-value = 0.1511
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.02282997 0.14423269
## sample estimates:
## mean in group female mean in group male
## 0.6492308 0.5885294
```

So no difference in males and females. What about differences based on the way that the students took notes?

```
t.test(data$score ~ data$method)
```

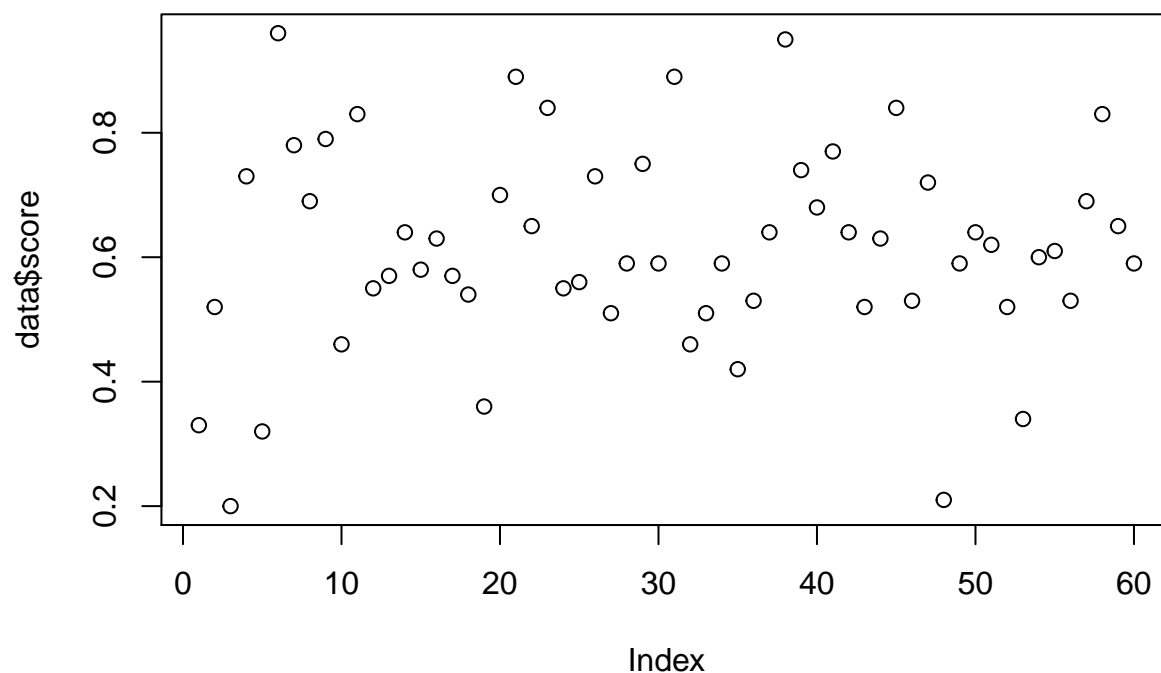
```
##
## Welch Two Sample t-test
##
## data: data$score by data$method
## t = -5.2705, df = 57.775, p-value = 2.11e-06
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.2552685 -0.1147315
```

```
## sample estimates:
## mean in group comp. mean in group paper
##      0.5223333      0.7073333
```

Now we do have a difference in our two groups. This is interesting and is something that we will want to illustrate in a plot for our readers. How should we do this? There is no wrong or right answer here. You want to do this effectively and there are many ways. If you haven't ever read any of Edward Tufte's books I would recompensed reading [Envisioning Infromation](#)

Lets slowly build up a nice graph that illustrates the variance and mean of our data.

```
#
plot(data$score)
```

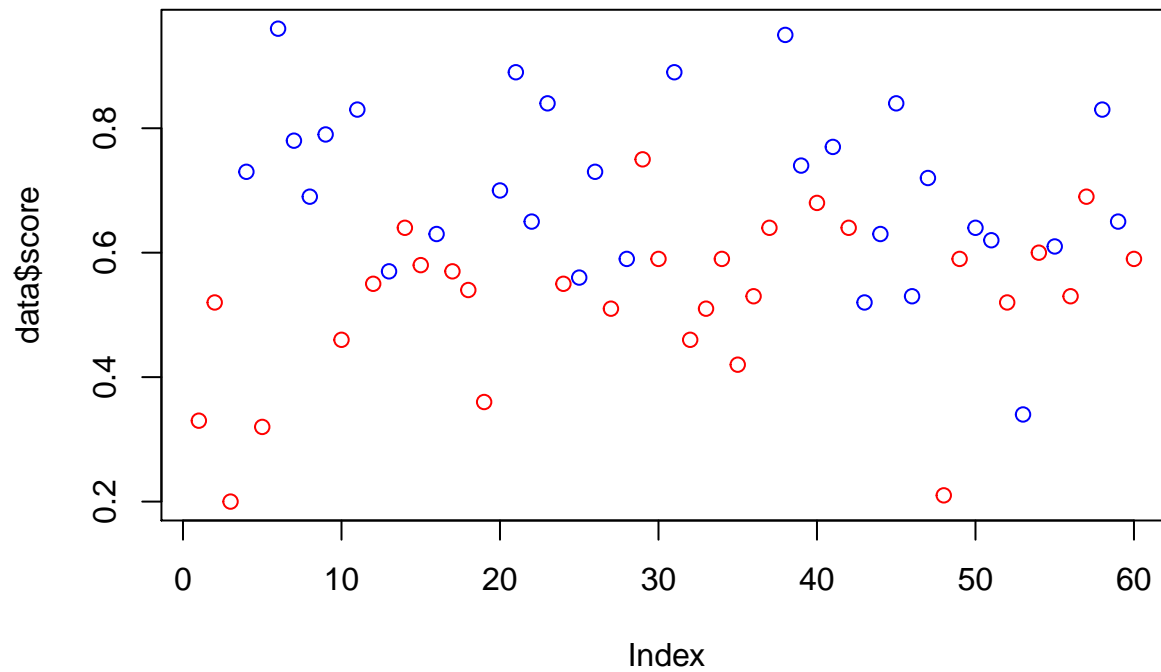


Lets try coloring those circles based on note taking method:

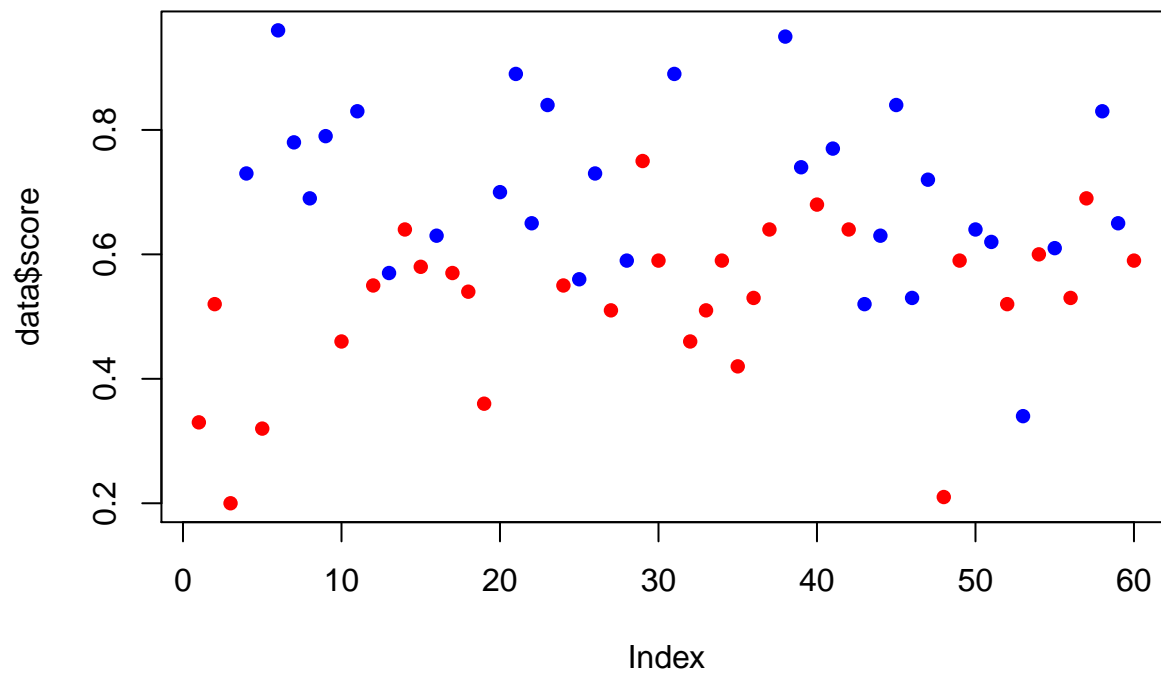
```
data$method
```

```
## [1] comp. comp. comp. paper comp. paper paper paper paper comp. paper
## [12] comp. paper comp. comp. paper comp. comp. comp. paper paper paper
## [23] paper comp. paper paper comp. paper comp. comp. paper comp. comp.
## [34] comp. comp. comp. comp. paper paper comp. paper comp. paper paper
## [45] paper paper paper comp. comp. paper paper comp. paper comp. paper
## [56] comp. comp. paper paper comp.
## Levels: comp. paper
```

```
colors <- c("red", "blue")[data$method]
plot(data$score, col=colors)
```



```
plot(data$score, col=colors, pch=16)
```

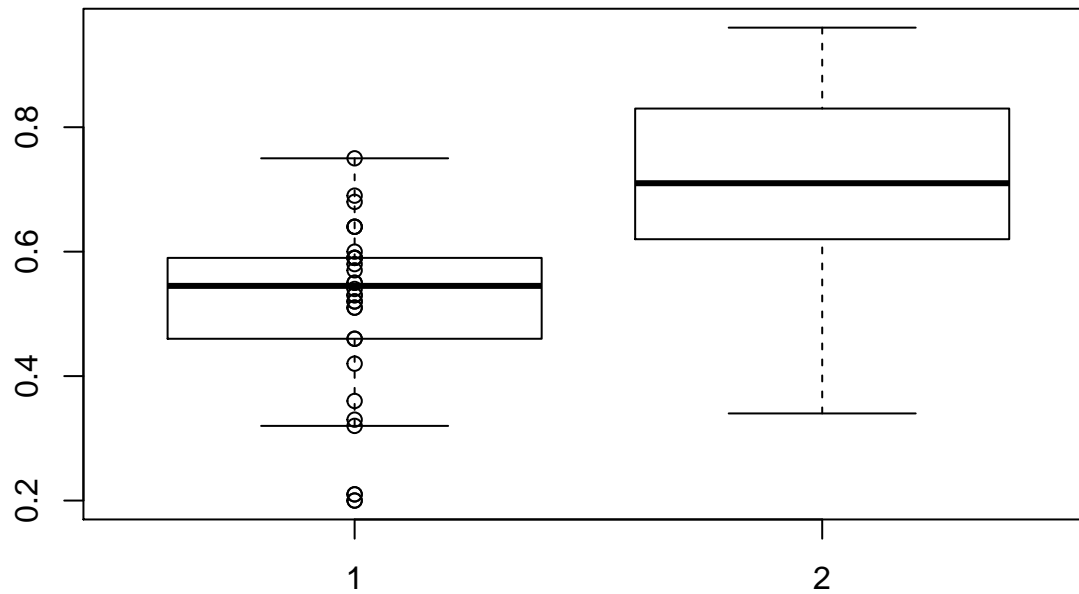


Well that's kind of interesting but we probably want to split these two sets of numbers apart from each other. A good option here might be a strip chart where we have boxplots and individual points combined.

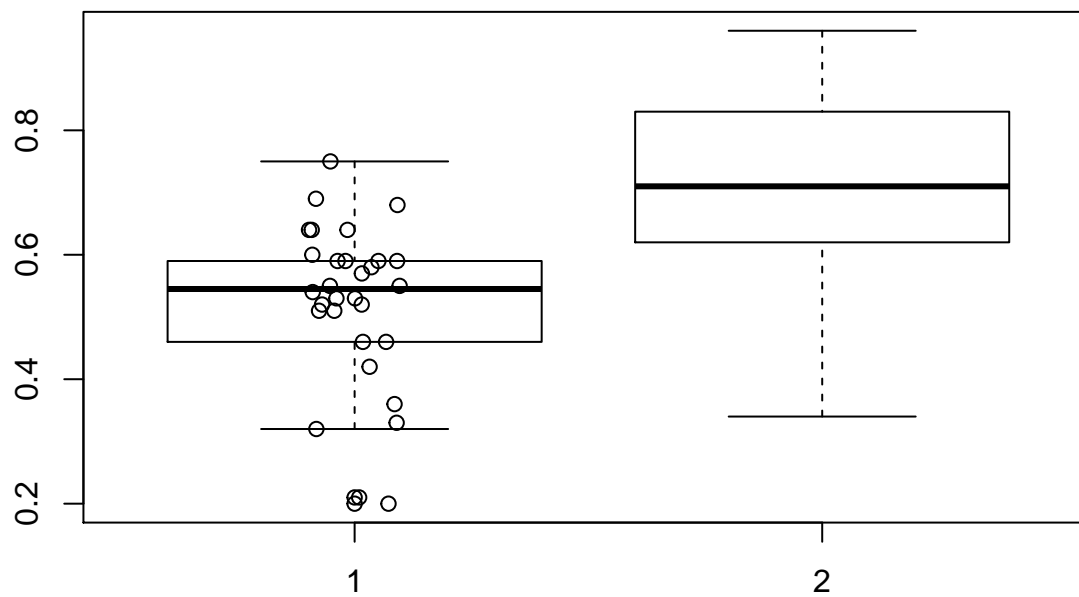
```
comp <- data$score[data$method=="comp."]
paper <- data$score[data$method=="paper"]

#
boxplot(comp, paper)
```

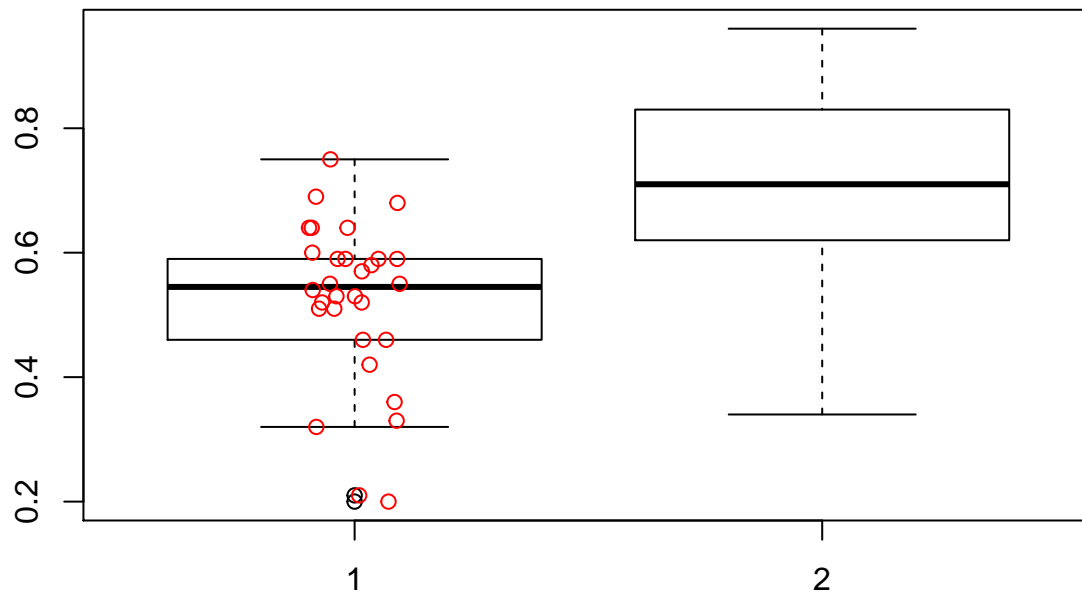
```
#
boxplot(comp, paper)
points(x=rep(1, 30), y=comp)
```



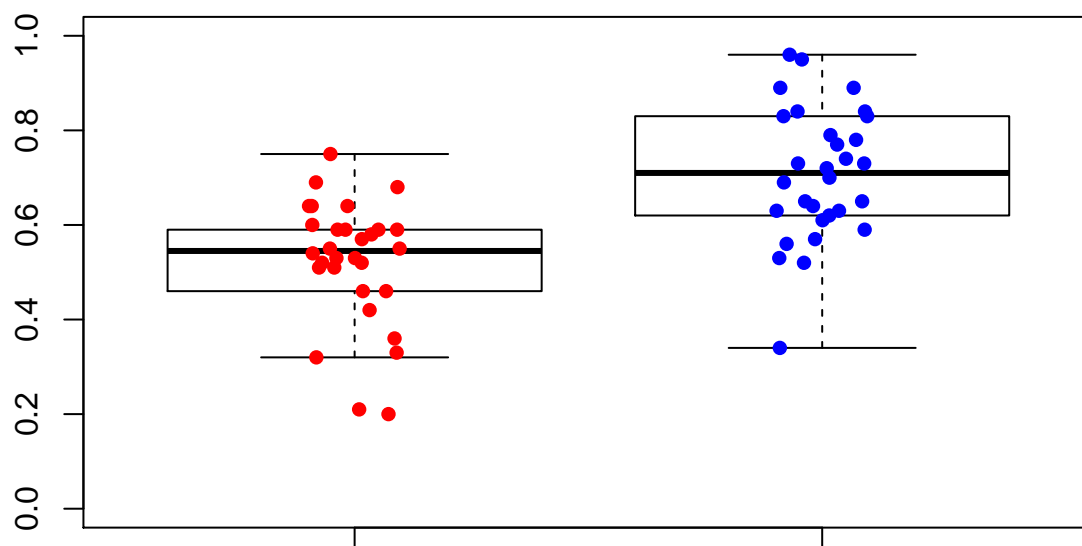
```
boxplot(comp, paper)
noise <- jitter(rep(1, 30), factor=5)
points(x=0+noise, y=comp)
```



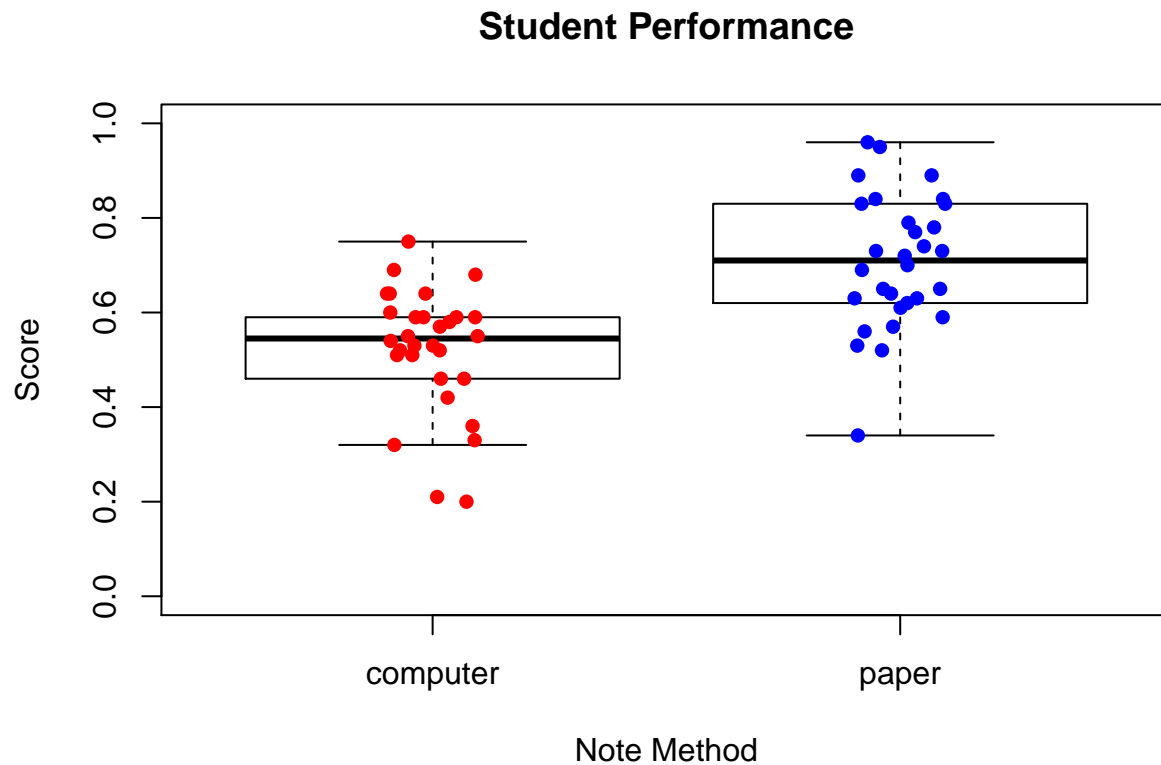
```
boxplot(comp, paper)
points(x=0+noise, y=comp, col="red")
```



```
boxplot(comp, paper, outline=F, ylim=c(0,1))
points(x=0+noise, y=comp, col="red", pch=16)
points(x=1+noise, y=paper, col="blue", pch=16)
```

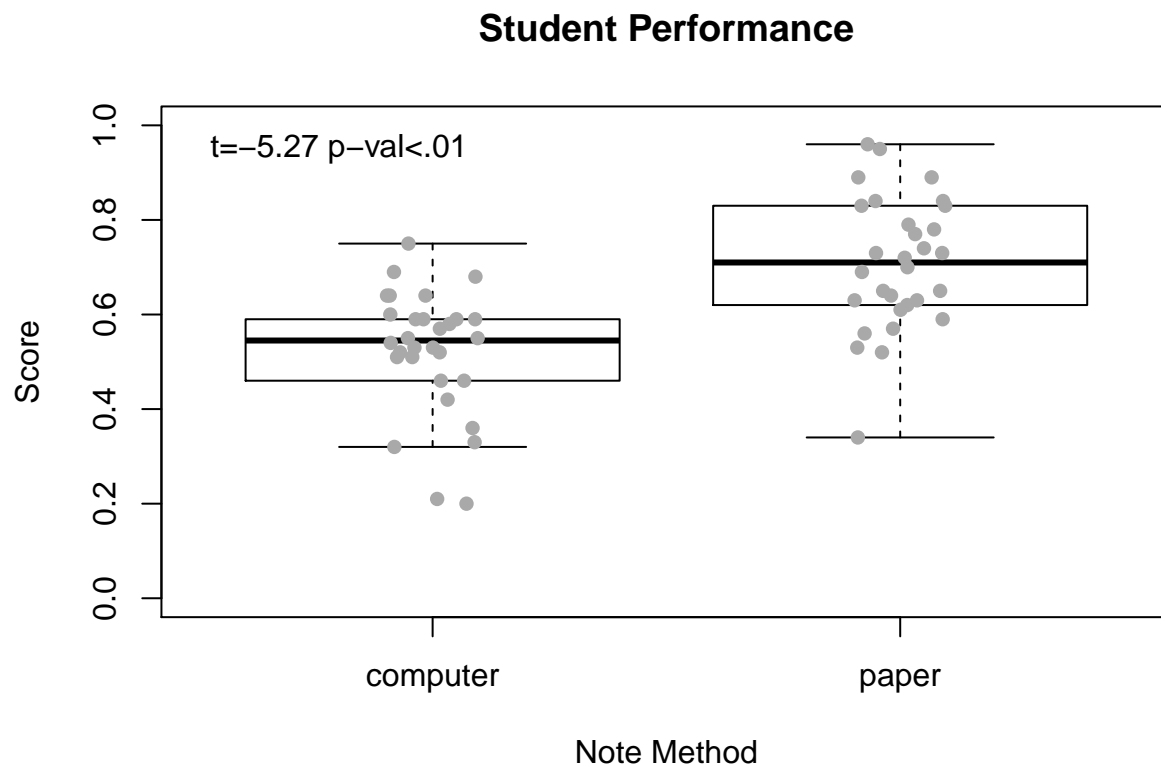


```
boxplot(comp, paper, outline=F, xlab="Note Method", ylab="Score", main="Student Performance", names=c("1", "2"))
points(x=0+noise, y=comp, col="red", pch=16)
points(x=1+noise, y=paper, col="blue", pch=16)
```



We might even want to add the results of our statistical test to the plot as well:

```
results <- t.test(x=comp, y=paper)
boxplot(comp, paper, outline=F, xlab="Note Method", ylab="Score", main="Student Performance", names=c("computer", "paper"))
points(x=0+noise, y=comp, col="darkgray", pch=16)
points(x=1+noise, y=paper, col="darkgray", pch=16)
text(x=.8, y=.95, paste("t=", round(results$statistic, digits=2), " p-val<.01", sep=""))
```



So now we have a plot that we like and we need to get it out of R and into our Latex, Word or document. You have lots of options here but make sure that you comment how you do it.

1. If you use latex I recommend exporting pdf files:

```
pdf(file="fig1.pdf", width = 6, height = 4)

results <- t.test(x=comp, y=paper)
boxplot(comp, paper, outline=F, xlab="Note Method", ylab="Score", main="Student Performance", names=c("computer", "paper"))
points(x=0+noise, y=comp, col="darkgray", pch=16)
points(x=1+noise, y=paper, col="darkgray", pch=16)
text(x=.8,y=.95, paste("t=", round(results$statistic,digits=2), " p-val<.01", sep=""))

dev.off()
```

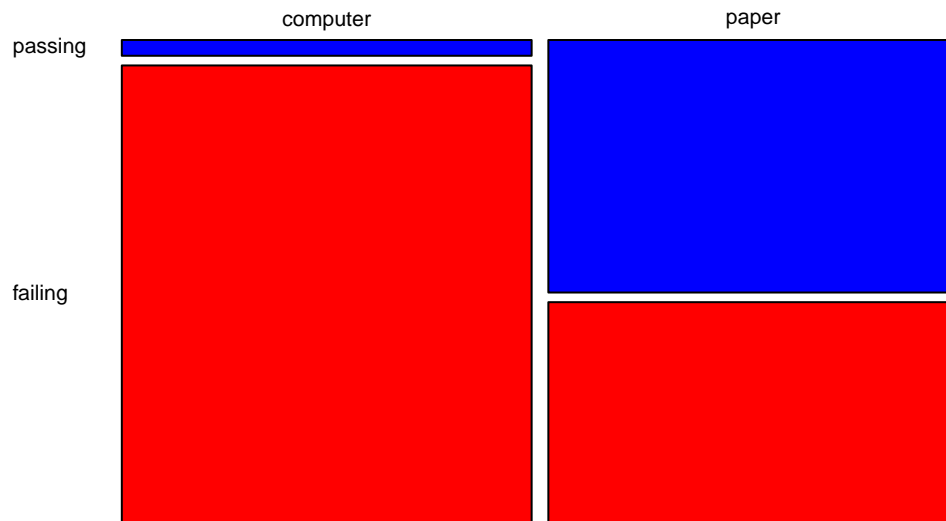
```
## pdf
## 2
```

2. If you use Word I recommend that you export a PDF file, but you won't be able to drop this into your word document. Instead you can keep the pdf file for submitting to the journal (almost all journals accept this) and for your working document you can just take a high resolution screen capture.

3. Challenge 1

Make a mosaic plot that shows the relationship between passing (at least 70%) and using a computer. You will have to munge your data to make a contingency table and then use `mosaicplot`

Effect of note taking methods



4. Challenge 2

Load the old faithful dataset `data(faithful)` and use what you have learned to create a plot. Try using `plot`, `abline`, `lm`. Make sure to make it more awesome than this:

Geyser eruption patterns

