Further examples: Histogram and Boxplot

Claire_Miller

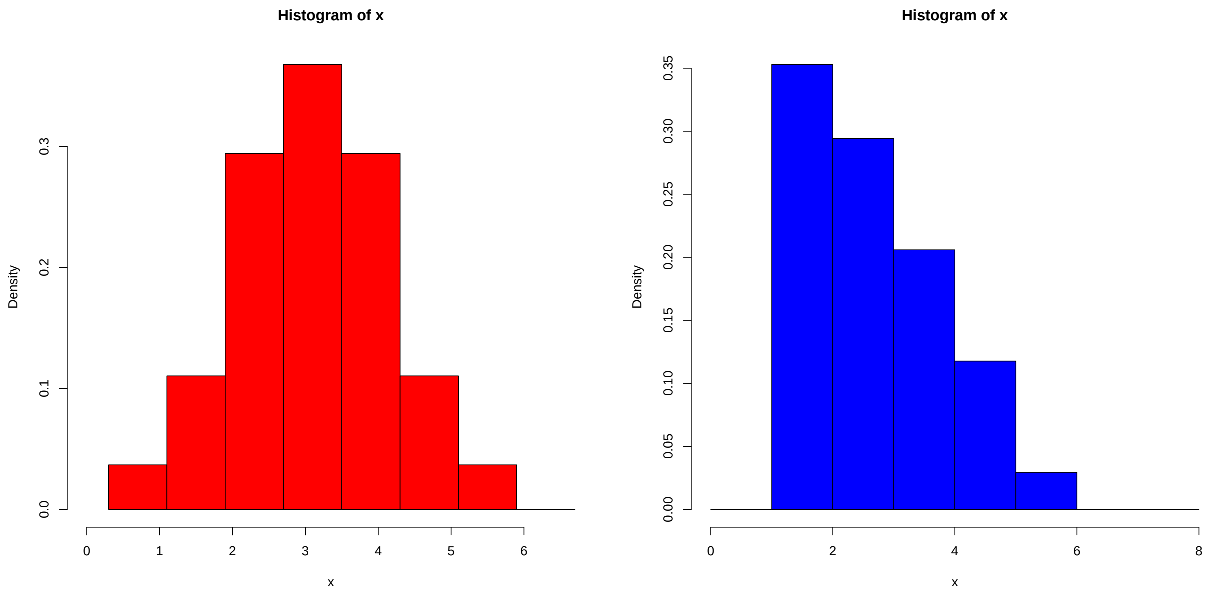
3 January 2018

Histograms and Boxplots

Histograms

We should always be careful when using the `histogram` funtion in R. The results can look totally different depending on the value of the `breaks` argument.

```
par(mfrow=c(1,2))
x <- c(1.03, 1.24, 1.47, 1.52, 1.92, 1.93, 1.94, 1.95, 1.96, 1.97, 1.98,
       1.99, 2.72, 2.75, 2.78, 2.81, 2.84, 2.87, 2.9, 2.93, 2.96, 2.99, 3.6,
       3.64, 3.66, 3.72, 3.77, 3.88, 3.91, 4.14, 4.54, 4.77, 4.81, 5.62)
hist(x,breaks=seq(0.3,6.7,by=0.8),xlim=c(0,6.7),col="red",freq=FALSE)
hist(x,breaks=0:8,col="blue",freq=FALSE)
```

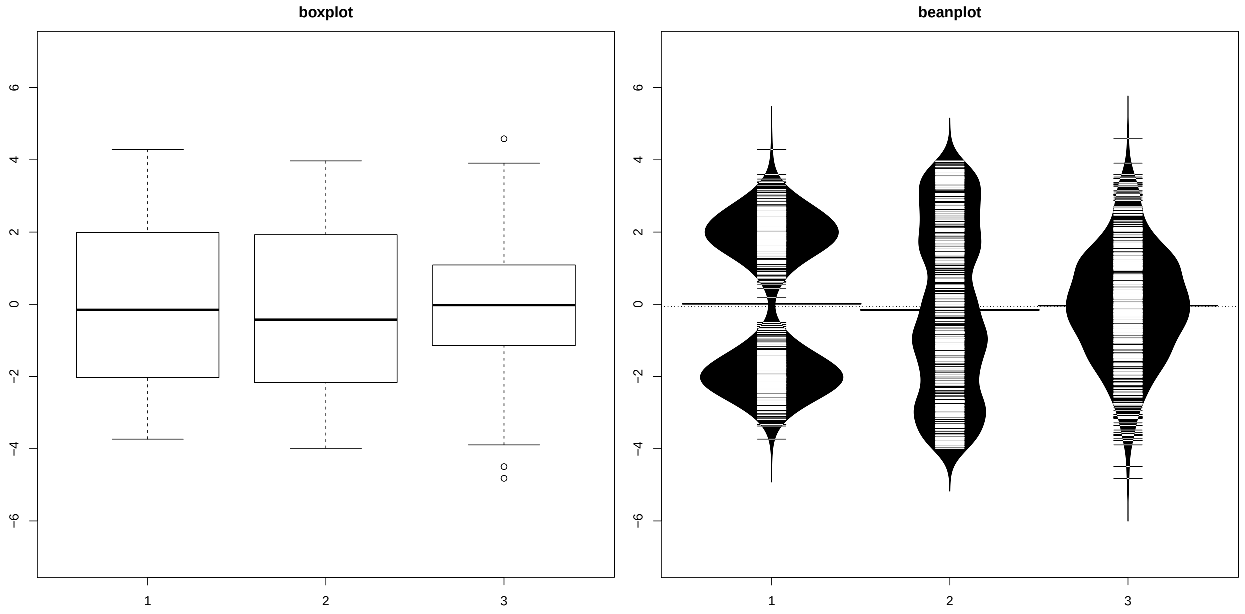


Boxplots

Sometimes the boxplot command can mask features in the data. For example, if our data is made up of subsets of data from different distributions. It is, therefore, important to consider if deeper exploration than just the usual `boxplot` command in R is required. In this case we can also use the `beanplot` package in R. This provides more information about the data sets.

```
library("beanplot")
set.seed(1)
 par(mfrow = c(1, 2), mai = c(0.5, 0.5, 0.5, 0.1))
 mu <- 2
 si <- 0.6
 c <- 500
 bimodal <- c(rnorm(c/2, -mu, si), rnorm(c/2, mu, si))
```

```r
uniform <- runif(c, -4, 4)
normal <- rnorm(c, 0, 1.5)
ylim <- c(-7, 7)
boxplot(bimodal, uniform, normal, ylim = ylim, main = "boxplot", names = 1:3)
beanplot(bimodal, uniform, normal, ylim = ylim, main = "beanplot")
```



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

In this example the full dataset is made up of three subsets of data, and each subset has been drawn from a different distribution.