Lab1

Jessica Temporal

August 4, 2016

Download alon.txt: https://drive.google.com/open?id=0B0-8N2fjttG-Xy1sQUNRREk2RUk and read `alon.txt` data

```
alon.data = read.table("~/Documents/methods_bioinformatics/data/lab1/alon.txt",
                       header = T)
```

Often in R, our data frames are read in with the gene names as a data column, instead of a row name. By doing the previous step, we are removing the gene names from a data column and setting them to the row names. (Hint: use dimnames(x)[[1]] on the left side of the assignment and cast the first column to character (as.character()) prior to setting the row names).

Setting the row names to the first column, then removing this first column.

```
rownames(alon.data) <- as.character(alon.data[,1])
alon.data$Gene <- NULL
```

There should be 62 samples. If you have 63 samples, you still have the row names in the first data column. Looking at the dimensions of the data.

```
dim(alon.data)
```

```
## [1] 2000   62
```

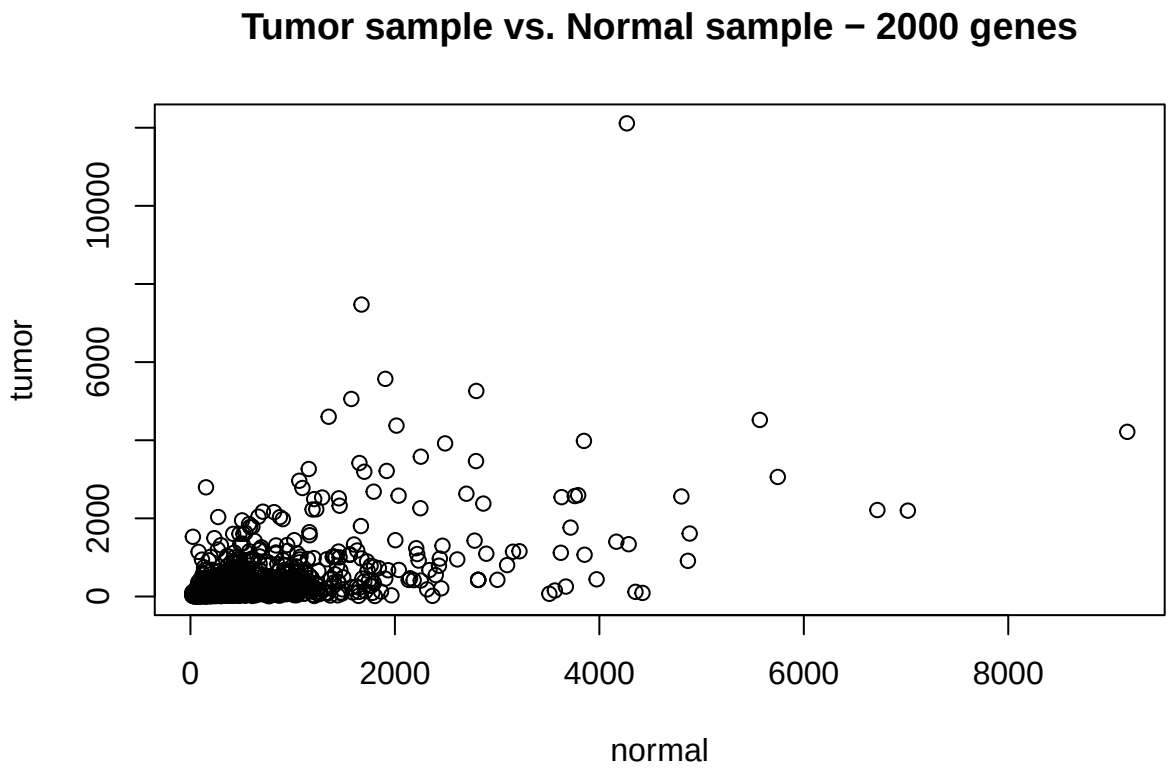
Print the sample names to screen.

```
names(alon.data)
```

```
##  [1] "norm1"   "norm2"   "norm3"   "norm4"   "norm5"   "norm6"   "norm7"
##  [8] "norm8"   "norm9"   "norm10"  "norm11"  "norm12"  "norm13"  "norm14"
## [15] "norm15"  "norm16"  "norm17"  "norm18"  "norm19"  "norm20"  "norm21"
## [22] "norm22"  "tumor1"  "tumor2"  "tumor3"  "tumor4"  "tumor5"  "tumor6"
## [29] "tumor7"  "tumor8"  "tumor9"  "tumor10" "tumor11" "tumor12" "tumor13"
## [36] "tumor14" "tumor15" "tumor16" "tumor17" "tumor18" "tumor19" "tumor20"
## [43] "tumor21" "tumor22" "tumor23" "tumor24" "tumor25" "tumor26" "tumor27"
## [50] "tumor28" "tumor29" "tumor30" "tumor31" "tumor32" "tumor33" "tumor34"
## [57] "tumor35" "tumor36" "tumor37" "tumor38" "tumor39" "tumor40"
```

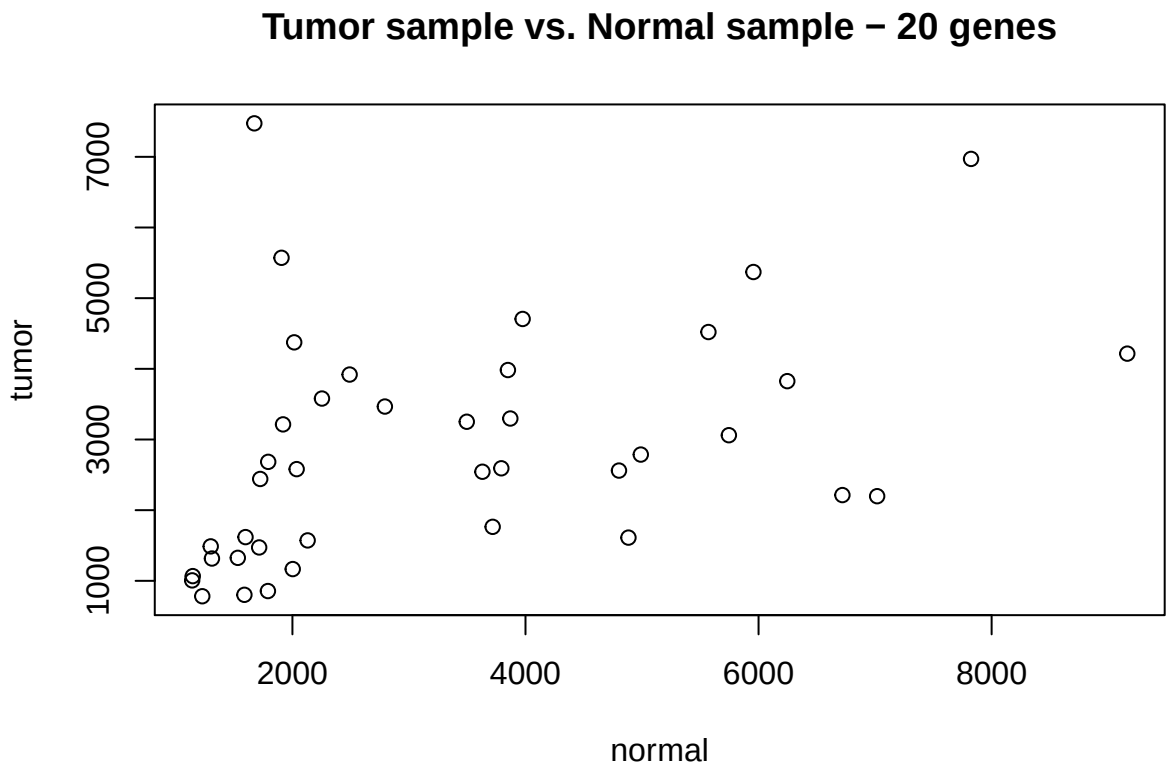
Plotting one of the tumor samples versus one of the normal samples in an xy scatter plot. Remember that the first argument is the x vector. Label the x and y-axes as 'normal' and 'tumor', respectively. Title the plot, 'Tumor sample vs. Normal sample - 2000 genes'.

```
plot(alon.data$norm1, alon.data$tumor1, xlab = "normal", ylab = "tumor",
     main = "Tumor sample vs. Normal sample - 2000 genes")
```



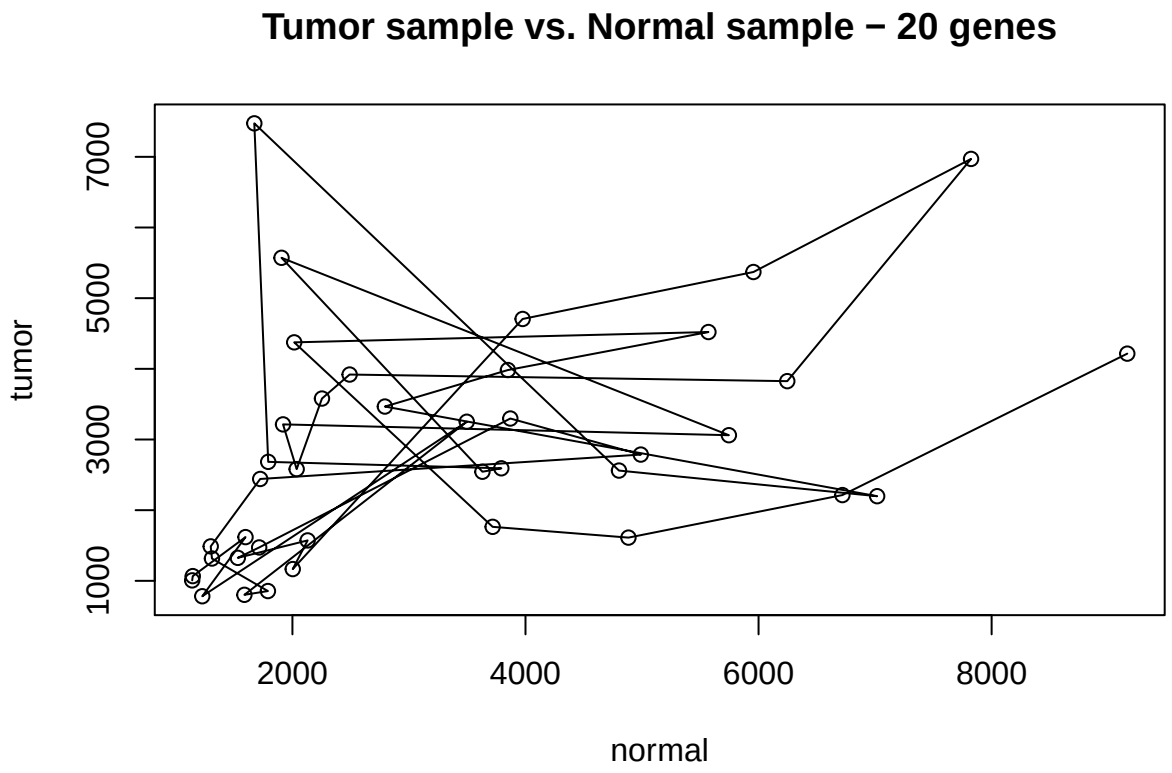
Now do the same with 2 normal samples, adjusting the axes labels and title, but pick only 20 genes.

```
norm1and2 = cbind(alon.data$norm1[1:20], alon.data$norm2[1:20])
tumor1an2 = cbind(alon.data$tumor1[1:20], alon.data$tumor2[1:20])
plot(norm1and2, tumor1an2, xlab = "normal", ylab = "tumor",
     main = "Tumor sample vs. Normal sample - 20 genes")
```



Add a line to connect the points

```
plot(norm1and2, tumor1an2, xlab = "normal", ylab = "tumor",
     main = "Tumor sample vs. Normal sample - 20 genes")
lines(norm1and2, tumor1an2)
```



Take the ratio of gene 5 to gene 15 and plot the profile of the gene across all samples. Label each point with the sample name (see text() help and use cex=1).

```
ratio5and15 <- alon.data[5,]/alon.data[15,]
plot(1:62, ratio5and15, xlab = "samples", main = "Gene profile across samples")
text(1:62, ratio5and15, labels = names(alon.data), cex = 1)
```