class007.R

Noah

2022-03-01

```r
url <- "https://tinyurl.com/UK-foods"
x <- read.csv(url)
dim(x)
```

```
## [1] 17  5
```
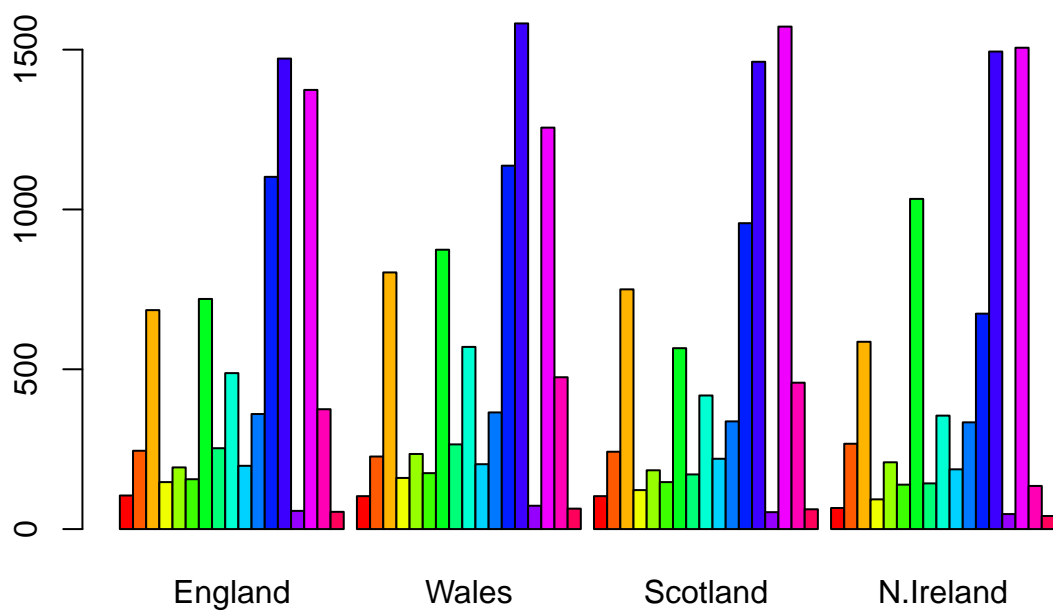
```r
rownames(x) <- x[,1]
x <- x[,-1]
head(x)
```

```
##               England Wales Scotland N.Ireland
## Cheese            105   103      103        66
## Carcass_meat      245   227      242       267
## Other_meat        685   803      750       586
## Fish              147   160      122        93
## Fats_and_oils     193   235      184       209
## Sugars            156   175      147       139
```
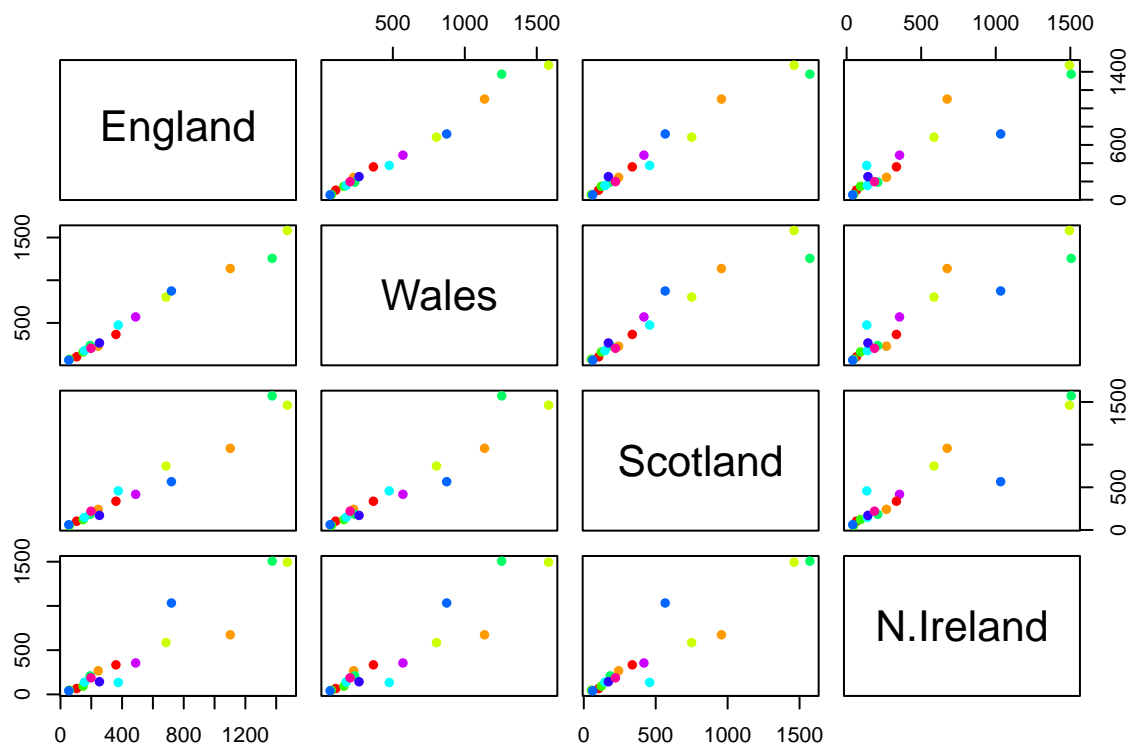
```r
dim(x)
```

```
## [1] 17  4
```

```r
barplot(as.matrix(x), beside=T, col=rainbow(nrow(x)))
```

```
pairs(x, col=rainbow(10), pch=16)
```
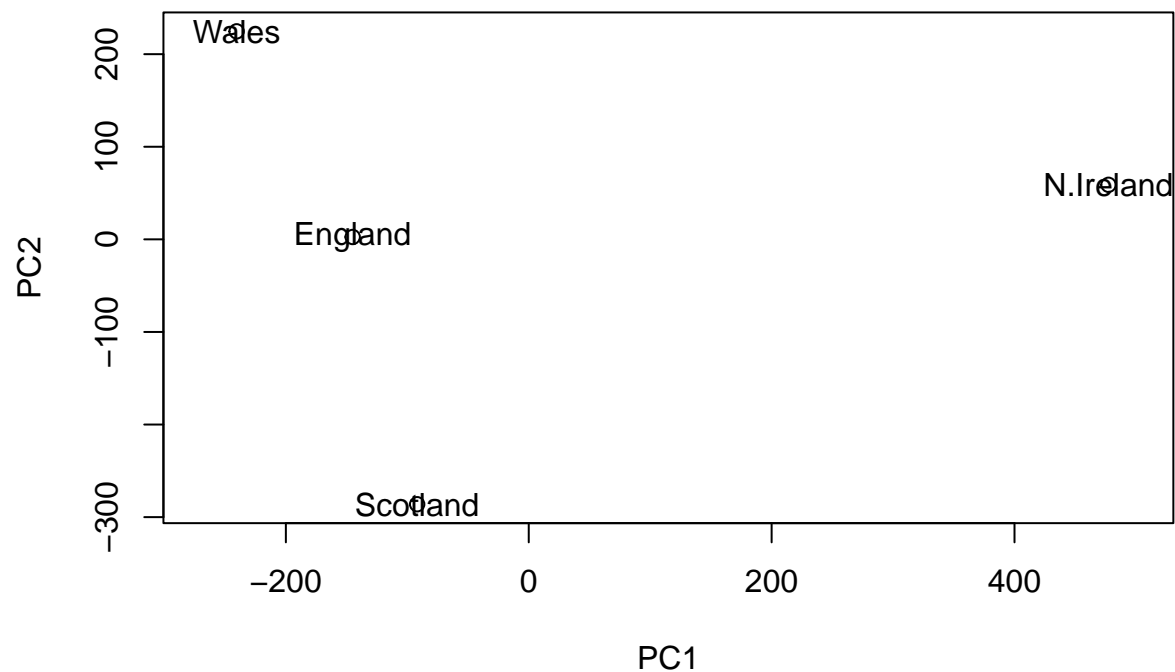
```
pca <- prcomp( t(x) )
summary(pca)
```

```
## Importance of components:
##                           PC1      PC2      PC3       PC4
## Standard deviation    324.1502 212.7478 73.87622 4.189e-14
## Proportion of Variance  0.6744   0.2905  0.03503 0.000e+00
## Cumulative Proportion   0.6744   0.9650  1.00000 1.000e+00
```

```
plot(pca$x[,1], pca$x[,2], xlab="PC1", ylab="PC2", xlim=c(-270,500))
text(pca$x[,1], pca$x[,2], colnames(x))
```
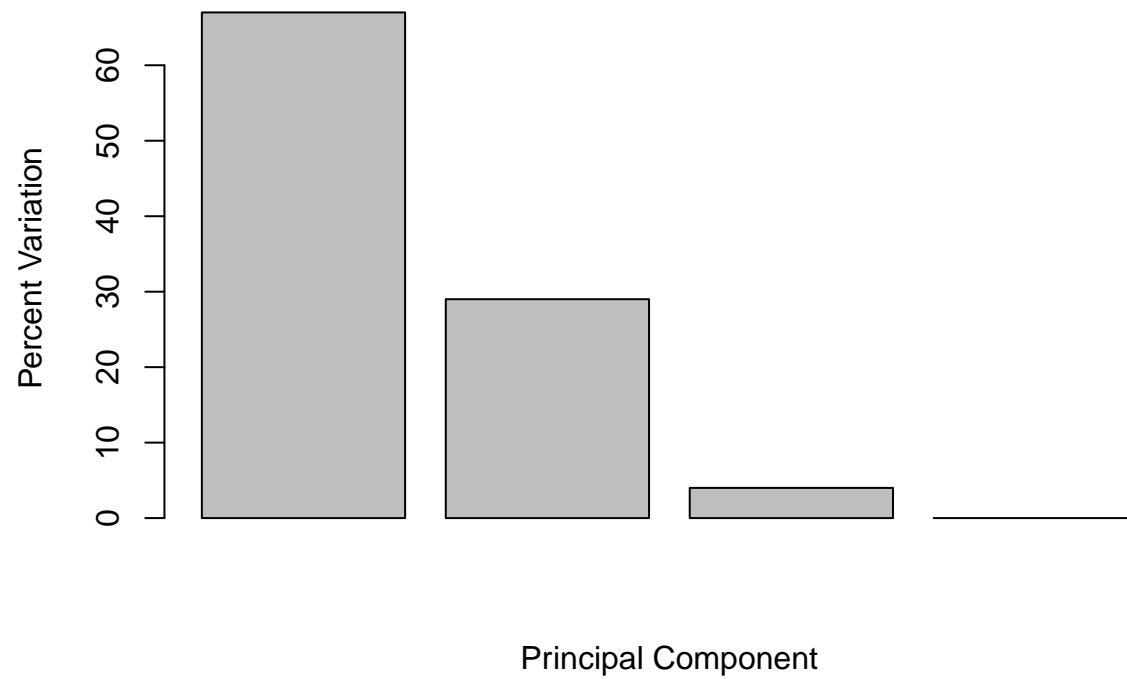
```r
v <- round( pca$sdev^2/sum(pca$sdev^2) * 100 )
v
```

```
## [1] 67 29  4  0
```
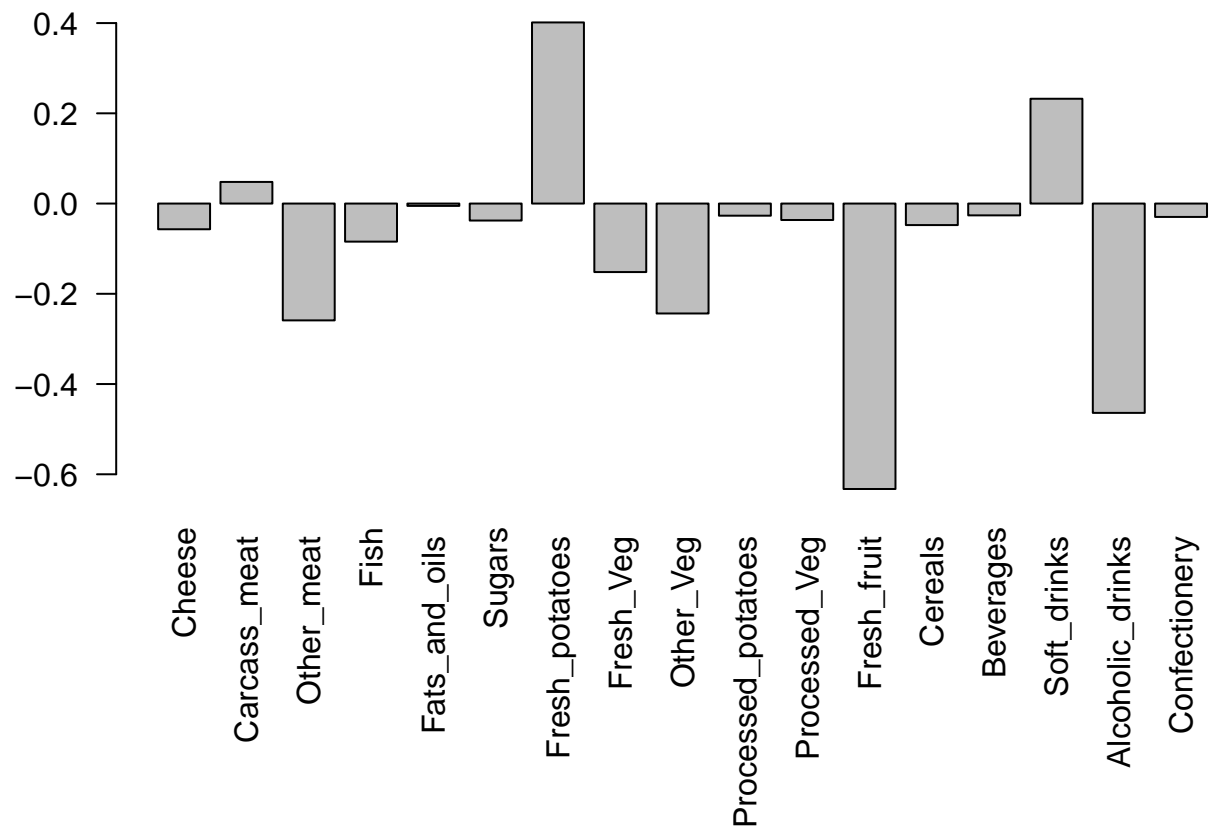
```r
z <- summary(pca)
z$importance
```

```
##                              PC1       PC2      PC3          PC4
## Standard deviation     324.15019 212.74780 73.87622 4.188568e-14
## Proportion of Variance   0.67444   0.29052  0.03503 0.000000e+00
## Cumulative Proportion    0.67444   0.96497  1.00000 1.000000e+00
```
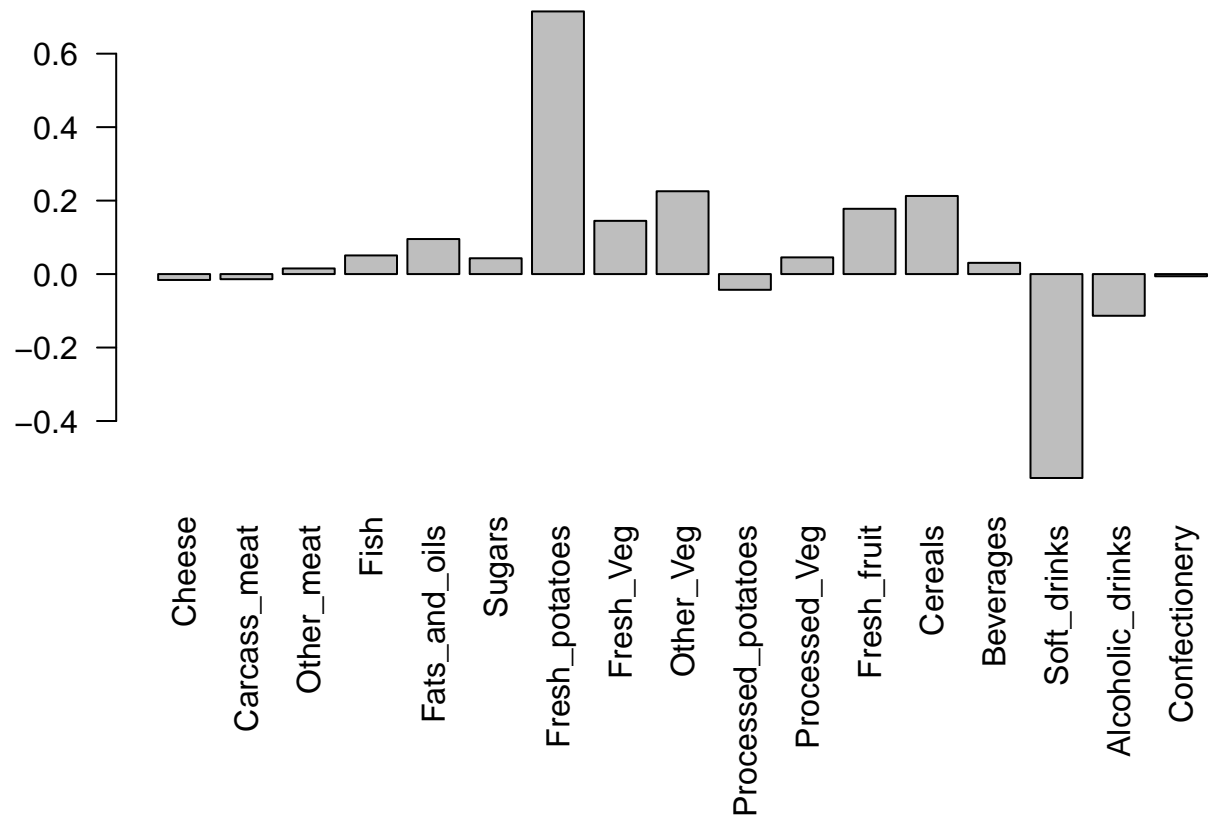
```r
barplot(v, xlab="Principal Component", ylab="Percent Variation")
```
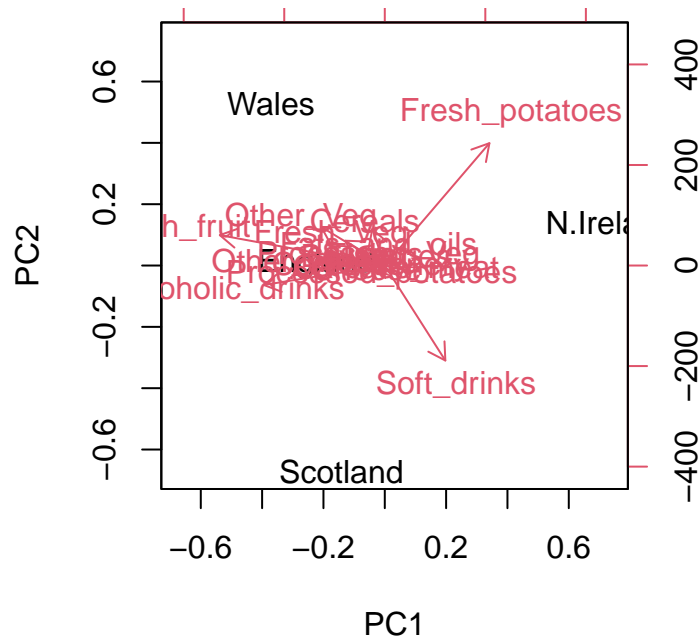
```
par(mar=c(10, 3, 0.35, 0))
barplot( pca$rotation[,1], las=2 )
```

```
par(mar=c(10, 3, 0.35, 0))
barplot( pca$rotation[,2], las=2 )
```

```
biplot(pca)
```

```
url2 <- "https://tinyurl.com/expression-CSV"
rna.data <- read.csv(url2, row.names=1)
head(rna.data)
```

```
##         wt1 wt2  wt3  wt4 wt5 ko1 ko2 ko3 ko4 ko5
## gene1  439 458  408  429 420  90  88  86  90  93
## gene2  219 200  204  210 187 427 423 434 433 426
## gene3 1006 989 1030 1017 973 252 237 238 226 210
## gene4  783 792  829  856 760 849 856 835 885 894
## gene5  181 249  204  244 225 277 305 272 270 279
## gene6  460 502  491  491 493 612 594 577 618 638
```

```
dim(rna.data)
```

```
## [1] 100  10
```

```
pca <- prcomp(t(rna.data), scale=TRUE)
plot(pca$x[,1], pca$x[,2], xlab="PC1", ylab="PC2")
```

```
summary(pca)
```
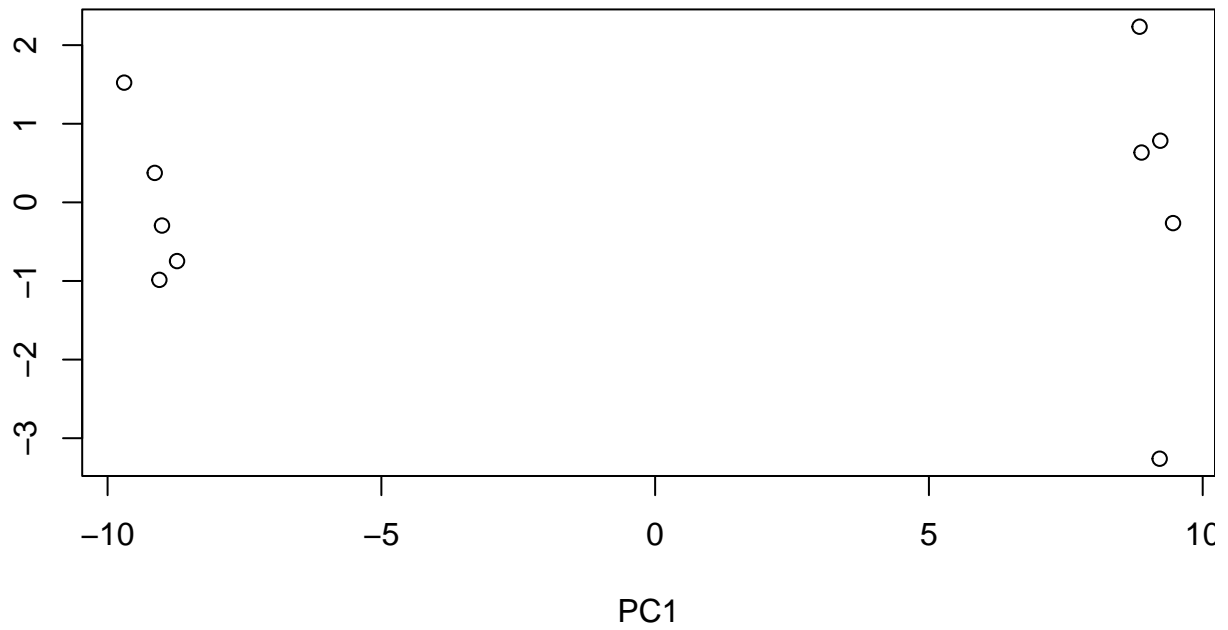
```
## Importance of components:
##                            PC1    PC2     PC3     PC4     PC5     PC6     PC7
## Standard deviation      9.6237 1.5198 1.05787 1.05203 0.88062 0.82545 0.80111
## Proportion of Variance  0.9262 0.0231 0.01119 0.01107 0.00775 0.00681 0.00642
## Cumulative Proportion   0.9262 0.9493 0.96045 0.97152 0.97928 0.98609 0.99251
##                            PC8    PC9      PC10
## Standard deviation      0.62065 0.60342 3.348e-15
## Proportion of Variance  0.00385 0.00364 0.000e+00
## Cumulative Proportion   0.99636 1.00000 1.000e+00
```
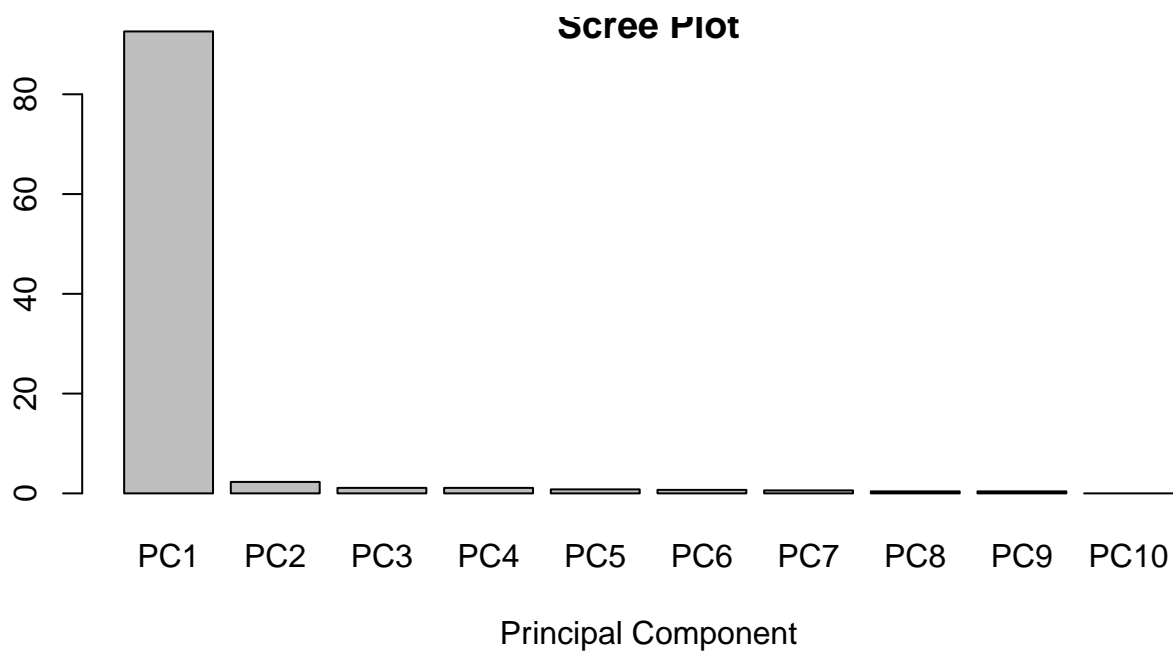
```
plot(pca, main="Quick scree plot")
```

Quick scree plot
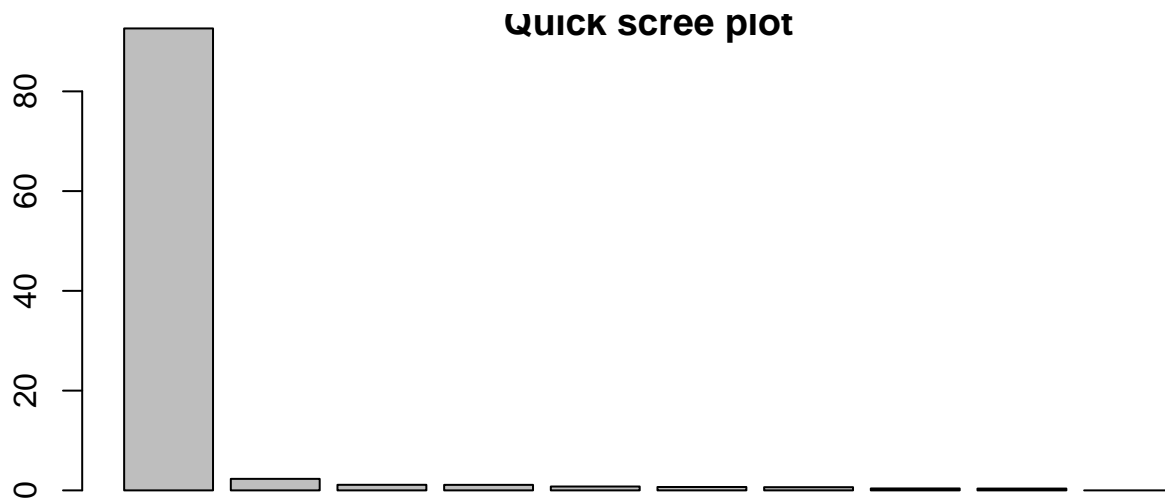
```
pca.var <- pca$sdev^2
pca.var.per <- round(pca.var/sum(pca.var)*100, 1)
pca.var.per
```

```
## [1] 92.6  2.3  1.1  1.1  0.8  0.7  0.6  0.4  0.4  0.0
```

```
barplot(pca.var.per, main="Scree Plot",
        names.arg = paste0("PC", 1:10),
        xlab="Principal Component", ylab="Percent Variation")
```
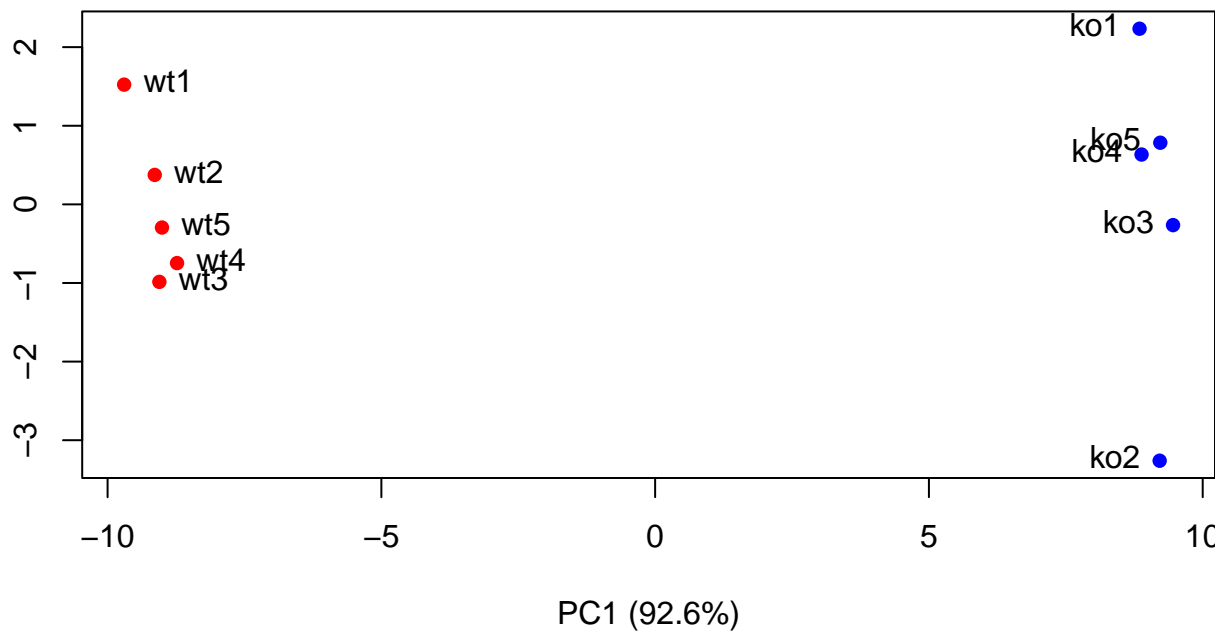
Scree Plot

Principal Component

```
colvec <- colnames(rna.data)
colvec[grep("wt", colvec)] <- "red"
colvec[grep("ko", colvec)] <- "blue"

plot(pca$x[,1], pca$x[,2], col=colvec, pch=16,
     xlab=paste0("PC1 (", pca.var.per[1], "%)"),
     ylab=paste0("PC2 (", pca.var.per[2], "%)"))

text(pca$x[,1], pca$x[,2], labels = colnames(rna.data), pos=c(rep(4,5), rep(2,5)))

library(ggplot2)
```
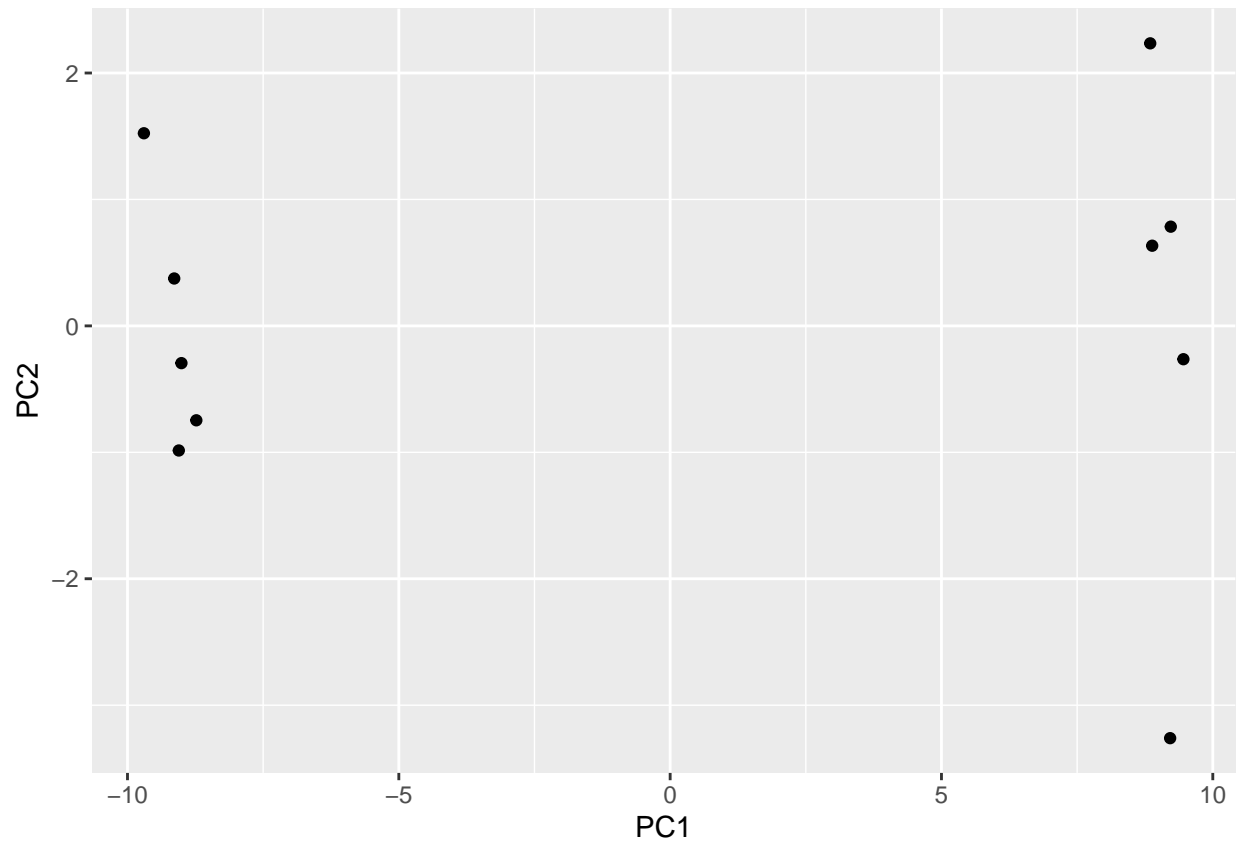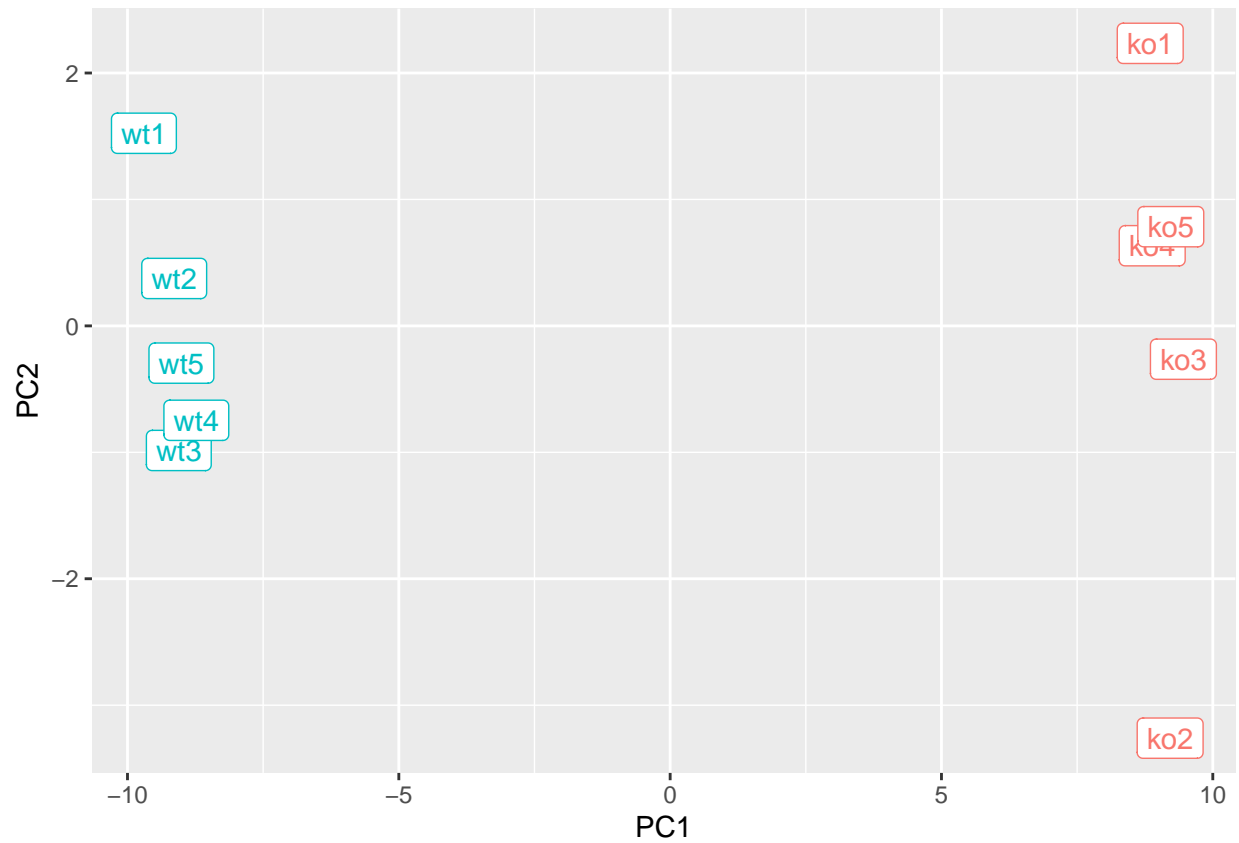
```
df <- as.data.frame(pca$x)

ggplot(df) +
  aes(PC1, PC2) +
  geom_point()
```

```
df$samples <- colnames(rna.data)
df$condition <- substr(colnames(rna.data),1,2)

p <- ggplot(df) +
  aes(PC1, PC2, label=samples, col=condition) +
  geom_label(show.legend = FALSE)
p
```
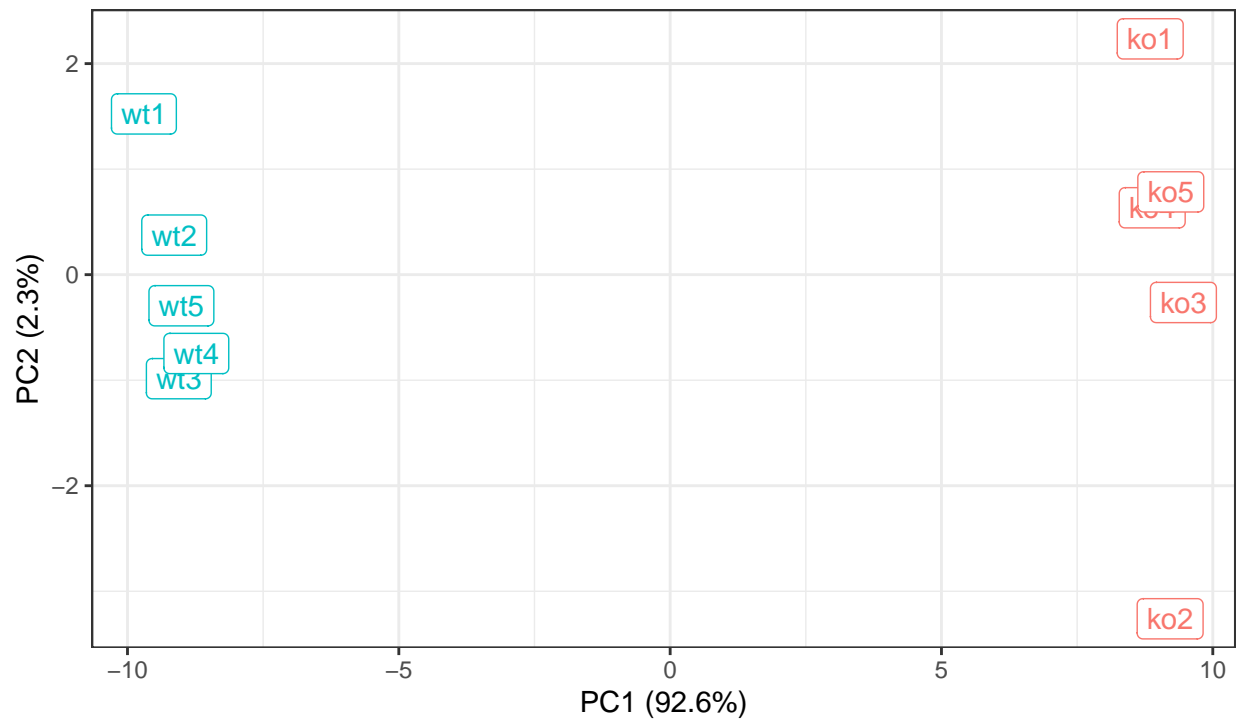
```
p + labs(title="PCA of RNASeq Data",
        subtitle = "PC1 clealy seperates wild-type from knock-out samples",
        x=paste0("PC1 (", pca.var.per[1], "%)"),
        y=paste0("PC2 (", pca.var.per[2], "%)"),
        caption="BIMM143 example data") +
  theme_bw()
```

PCA of RNASeq Data

PC1 clealy seperates wild−type from knock−out samples

BIMM143 example data