```
> library(tidyverse)
> library(caret)
> library(dslabs)
> data("tissue_gene_expression")
> class(tissue_gene_expression)
[1] "list"
> names(tissue_gene_expression)
[1] "x" "y"
> dim(tissue_gene_expression$x)
[1] 189 500
> dim(as.matrix(tissue_gene_expression$y))
[1] 189    1
>
>
> pc <- prcomp(tissue_gene_expression$x)
> data.frame(pc_1 = pc$x[,1], pc_2 = pc$x[,2],
+ tissue = tissue_gene_expression$y) %>%
+ ggplot(aes(pc_1, pc_2, color = tissue)) +
+ geom_point()
>
>
>
> avgs <- rowMeans(tissue_gene_expression$x)
> data.frame(pc_1 = pc$x[,1], avg = avgs,
+ tissue = tissue_gene_expression$y) %>%
+ ggplot(aes(avgs, pc_1, color = tissue)) +
+ geom_point()
cor(avgs, pc$x[,1])
> cor(avgs, pc$x[,1])
[1] 0.5969088
>
>
>
> x <- with(tissue_gene_expression, sweep(x, 1, rowMeans(x)))
> pc <- prcomp(x)
> data.frame(pc_1 = pc$x[,1], pc_2 = pc$x[,2],
+ tissue = tissue_gene_expression$y) %>%
+ ggplot(aes(pc_1, pc_2, color = tissue)) +
+ geom_point()
>
>
>
> for(i in 1:10){
+ boxplot(pc$x[,i] ~ tissue_gene_expression$y, main = paste("PC", i))
+ }
>
>
>
> plot(summary(pc)$importance[3,])
>
```