Miniproject08_Attempt2.R

Noah

2022-03-05

```
wisc.df <- read.csv("WisconsinCancer.csv", row.names=1)
wisc.data <- wisc.df[,-1]
diagnosis <- factor(wisc.data[,1])
colMeans(wisc.data[,-1])
```

```
##            texture_mean          perimeter_mean               area_mean
##            1.928965e+01            9.196903e+01            6.548891e+02
##         smoothness_mean        compactness_mean          concavity_mean
##            9.636028e-02            1.043410e-01            8.879932e-02
##     concave.points_mean          symmetry_mean  fractal_dimension_mean
##            4.891915e-02            1.811619e-01            6.279761e-02
##               radius_se              texture_se            perimeter_se
##            4.051721e-01            1.216853e+00            2.866059e+00
##                 area_se         smoothness_se          compactness_se
##            4.033708e+01            7.040979e-03            2.547814e-02
##            concavity_se       concave.points_se             symmetry_se
##            3.189372e-02            1.179614e-02            2.054230e-02
##    fractal_dimension_se            radius_worst           texture_worst
##            3.794904e-03            1.626919e+01            2.567722e+01
##         perimeter_worst              area_worst        smoothness_worst
##            1.072612e+02            8.805831e+02            1.323686e-01
##       compactness_worst         concavity_worst    concave.points_worst
##            2.542650e-01            2.721885e-01            1.146062e-01
##          symmetry_worst fractal_dimension_worst
##            2.900756e-01            8.394582e-02
```

```
apply(wisc.data,2,sd)
```

```
##             radius_mean            texture_mean          perimeter_mean
##            3.524049e+00            4.301036e+00            2.429898e+01
##               area_mean         smoothness_mean        compactness_mean
##            3.519141e+02            1.406413e-02            5.281276e-02
##          concavity_mean     concave.points_mean           symmetry_mean
##            7.971981e-02            3.880284e-02            2.741428e-02
##  fractal_dimension_mean               radius_se              texture_se
##            7.060363e-03            2.773127e-01            5.516484e-01
##            perimeter_se                 area_se         smoothness_se
##            2.021855e+00            4.549101e+01            3.002518e-03
##          compactness_se            concavity_se       concave.points_se
##            1.790818e-02            3.018606e-02            6.170285e-03
##             symmetry_se    fractal_dimension_se            radius_worst
```

```
##            8.266372e-03            2.646071e-03            4.833242e+00
##           texture_worst         perimeter_worst              area_worst
##            6.146258e+00            3.360254e+01            5.693570e+02
##        smoothness_worst       compactness_worst         concavity_worst
##            2.283243e-02            1.573365e-01            2.086243e-01
##    concave.points_worst          symmetry_worst fractal_dimension_worst
##            6.573234e-02            6.186747e-02            1.806127e-02
```

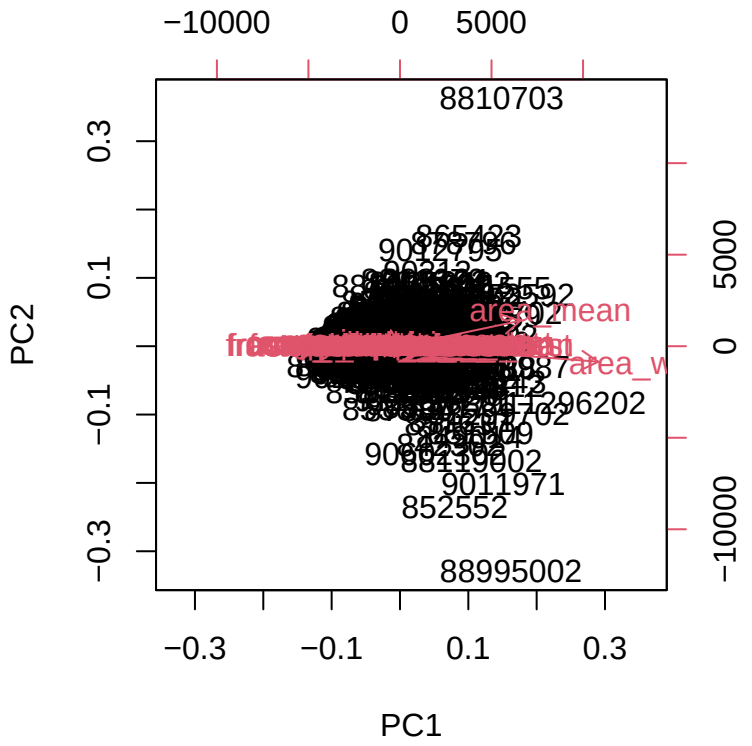
```
wisc.pr <- prcomp(wisc.data[,-1])
summary(wisc.pr)
```

```
## Importance of components:
##                             PC1      PC2      PC3     PC4     PC5     PC6   PC7
## Standard deviation     666.161 85.49544 26.52784 7.38816 6.31187 1.71854 1.347
## Proportion of Variance   0.982  0.01618  0.00156 0.00012 0.00009 0.00001 0.000
## Cumulative Proportion    0.982  0.99822  0.99978 0.99990 0.99999 0.99999 1.000
##                           PC8    PC9   PC10   PC11   PC12   PC13  PC14    PC15
## Standard deviation     0.6079 0.3713 0.2867 0.1618 0.0597 0.0471 0.039 0.02855
## Proportion of Variance 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.000 0.00000
## Cumulative Proportion  1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.000 1.00000
##                           PC16    PC17    PC18    PC19     PC20     PC21
## Standard deviation     0.01981 0.01722 0.01381 0.01344 0.009507 0.007602
## Proportion of Variance 0.00000 0.00000 0.00000 0.00000 0.000000 0.000000
## Cumulative Proportion  1.00000 1.00000 1.00000 1.00000 1.000000 1.000000
##                            PC22     PC23     PC24     PC25     PC26     PC27
## Standard deviation     0.005918 0.005557 0.004023 0.003535 0.001925 0.001698
## Proportion of Variance 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000
## Cumulative Proportion  1.000000 1.000000 1.000000 1.000000 1.000000 1.000000
##                            PC28      PC29
## Standard deviation     0.001416 0.0008387
## Proportion of Variance 0.000000 0.0000000
## Cumulative Proportion  1.000000 1.0000000
```

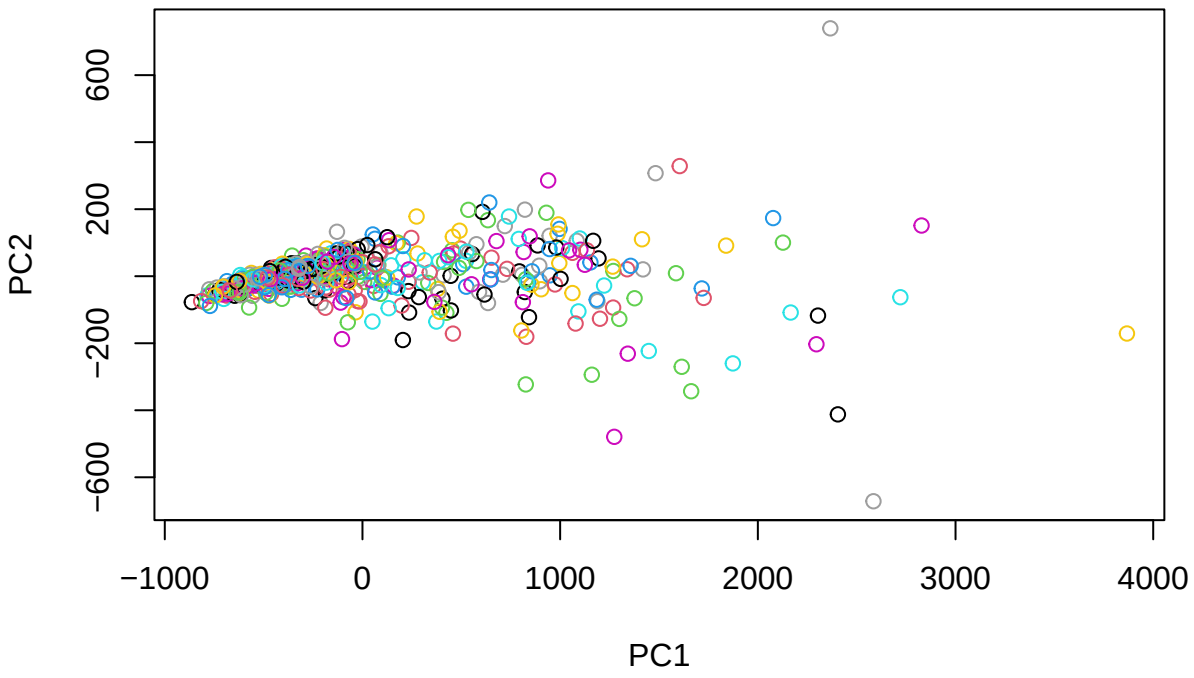
```
biplot(wisc.pr)
```

```
## Warning in arrows(0, 0, y[, 1L] * 0.8, y[, 2L] * 0.8, col = col[2L], length =
## arrow.len): zero-length arrow is of indeterminate angle and so skipped

## Warning in arrows(0, 0, y[, 1L] * 0.8, y[, 2L] * 0.8, col = col[2L], length =
## arrow.len): zero-length arrow is of indeterminate angle and so skipped

## Warning in arrows(0, 0, y[, 1L] * 0.8, y[, 2L] * 0.8, col = col[2L], length =
## arrow.len): zero-length arrow is of indeterminate angle and so skipped

## Warning in arrows(0, 0, y[, 1L] * 0.8, y[, 2L] * 0.8, col = col[2L], length =
## arrow.len): zero-length arrow is of indeterminate angle and so skipped

## Warning in arrows(0, 0, y[, 1L] * 0.8, y[, 2L] * 0.8, col = col[2L], length =
## arrow.len): zero-length arrow is of indeterminate angle and so skipped

## Warning in arrows(0, 0, y[, 1L] * 0.8, y[, 2L] * 0.8, col = col[2L], length =
## arrow.len): zero-length arrow is of indeterminate angle and so skipped
```

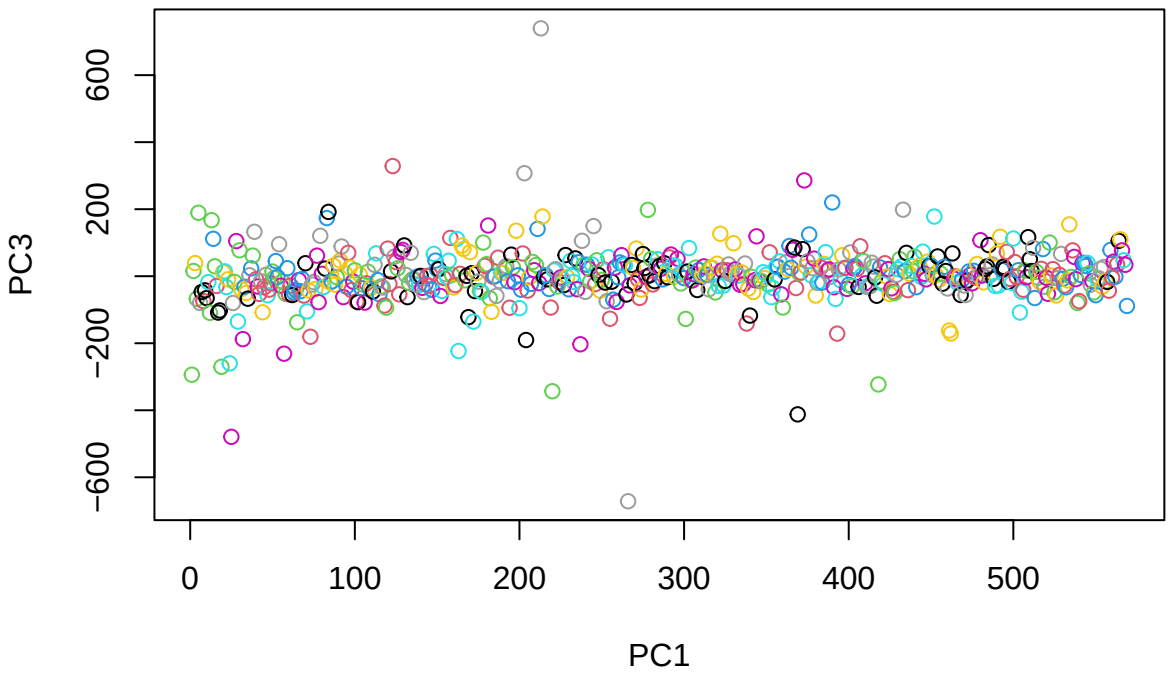
```
## Warning in arrows(0, 0, y[, 1L] * 0.8, y[, 2L] * 0.8, col = col[2L], length =
## arrow.len): zero-length arrow is of indeterminate angle and so skipped

## Warning in arrows(0, 0, y[, 1L] * 0.8, y[, 2L] * 0.8, col = col[2L], length =
## arrow.len): zero-length arrow is of indeterminate angle and so skipped

## Warning in arrows(0, 0, y[, 1L] * 0.8, y[, 2L] * 0.8, col = col[2L], length =
## arrow.len): zero-length arrow is of indeterminate angle and so skipped

## Warning in arrows(0, 0, y[, 1L] * 0.8, y[, 2L] * 0.8, col = col[2L], length =
## arrow.len): zero-length arrow is of indeterminate angle and so skipped

## Warning in arrows(0, 0, y[, 1L] * 0.8, y[, 2L] * 0.8, col = col[2L], length =
## arrow.len): zero-length arrow is of indeterminate angle and so skipped

## Warning in arrows(0, 0, y[, 1L] * 0.8, y[, 2L] * 0.8, col = col[2L], length =
## arrow.len): zero-length arrow is of indeterminate angle and so skipped

## Warning in arrows(0, 0, y[, 1L] * 0.8, y[, 2L] * 0.8, col = col[2L], length =
## arrow.len): zero-length arrow is of indeterminate angle and so skipped

## Warning in arrows(0, 0, y[, 1L] * 0.8, y[, 2L] * 0.8, col = col[2L], length =
## arrow.len): zero-length arrow is of indeterminate angle and so skipped

## Warning in arrows(0, 0, y[, 1L] * 0.8, y[, 2L] * 0.8, col = col[2L], length =
## arrow.len): zero-length arrow is of indeterminate angle and so skipped

## Warning in arrows(0, 0, y[, 1L] * 0.8, y[, 2L] * 0.8, col = col[2L], length =
## arrow.len): zero-length arrow is of indeterminate angle and so skipped

## Warning in arrows(0, 0, y[, 1L] * 0.8, y[, 2L] * 0.8, col = col[2L], length =
## arrow.len): zero-length arrow is of indeterminate angle and so skipped

## Warning in arrows(0, 0, y[, 1L] * 0.8, y[, 2L] * 0.8, col = col[2L], length =
## arrow.len): zero-length arrow is of indeterminate angle and so skipped

## Warning in arrows(0, 0, y[, 1L] * 0.8, y[, 2L] * 0.8, col = col[2L], length =
## arrow.len): zero-length arrow is of indeterminate angle and so skipped

## Warning in arrows(0, 0, y[, 1L] * 0.8, y[, 2L] * 0.8, col = col[2L], length =
## arrow.len): zero-length arrow is of indeterminate angle and so skipped
```




```
plot(wisc.pr$x, col = diagnosis,
     xlab = "PC1", ylab = "PC2")
```




```
plot(wisc.pr$x[, 2 ], col = diagnosis,
     xlab = "PC1", ylab = "PC3")

df <- as.data.frame(wisc.pr$x)
df$diagnosis <- diagnosis

library(ggplot2)
```



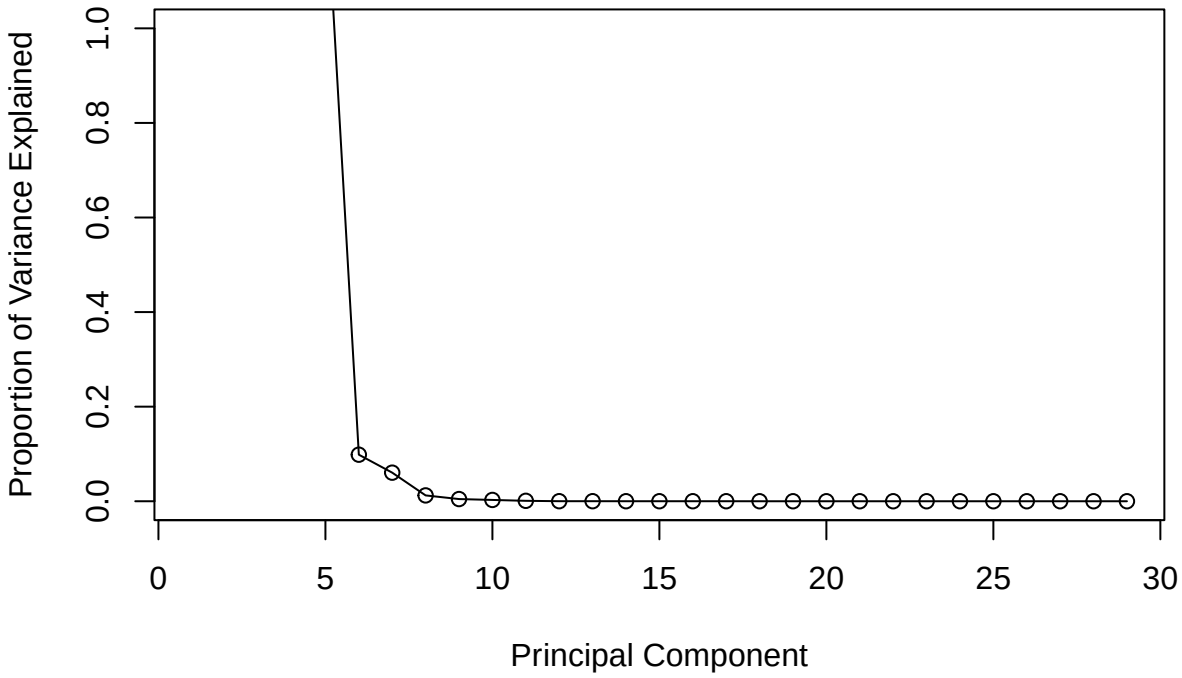

```
ggplot(df) +
  aes(PC1, PC2, col= diagnosis) +
  geom_point()
```

11.75	12.16	12.47	12.86	13.17	13.53	13.82	14.27	14.69	1!
11.76	12.18	12.49	12.87	13.2	13.54	13.85	14.29	14.71	1!
11.8	12.19	12.54	12.88	13.21	13.56	13.86	14.34	14.74	1!
11.81	12.2	12.56	12.89	13.24	13.59	13.87	14.4	14.76	1!
11.84	12.21	12.58	12.9	13.27	13.61	13.88	14.41	14.78	1!
11.85	12.22	12.62	12.91	13.28	13.62	13.9	14.42	14.8	1!
11.87	12.23	12.63	12.94	13.3	13.64	13.94	14.44	14.81	1!
11.89	12.25	12.65	12.95	13.34	13.65	13.96	14.45	14.86	1!
11.9	12.27	12.67	12.96	13.37	13.66	13.98	14.47	14.87	1!
11.93	12.3	12.68	12.98	13.38	13.68	14.02	14.48	14.9	1!
11.94	12.31	12.7	12.99	13.4	13.69	14.03	14.5	14.92	1!
11.95	12.32	12.72	13	13.43	13.7	14.04	14.53	14.95	1!
11.99	12.34	12.75	13.01	13.44	13.71	14.05	14.54	14.96	1!
12	12.36	12.76	13.03	13.45	13.73	14.06	14.58	14.97	1!
12.03	12.39	12.77	13.05	13.46	13.74	14.11	14.59	14.99	1!
12.04	12.4	12.78	13.08	13.47	13.75	14.19	14.6	15	1!
12.05	12.42	12.8	13.11	13.48	13.77	14.2	14.61	15.04	1!
12.06	12.43	12.81	13.14	13.49	13.78	14.22	14.62	15.05	1!
12.07	12.45	12.83	13.15	13.5	13.8	14.25	14.64	15.06	1!

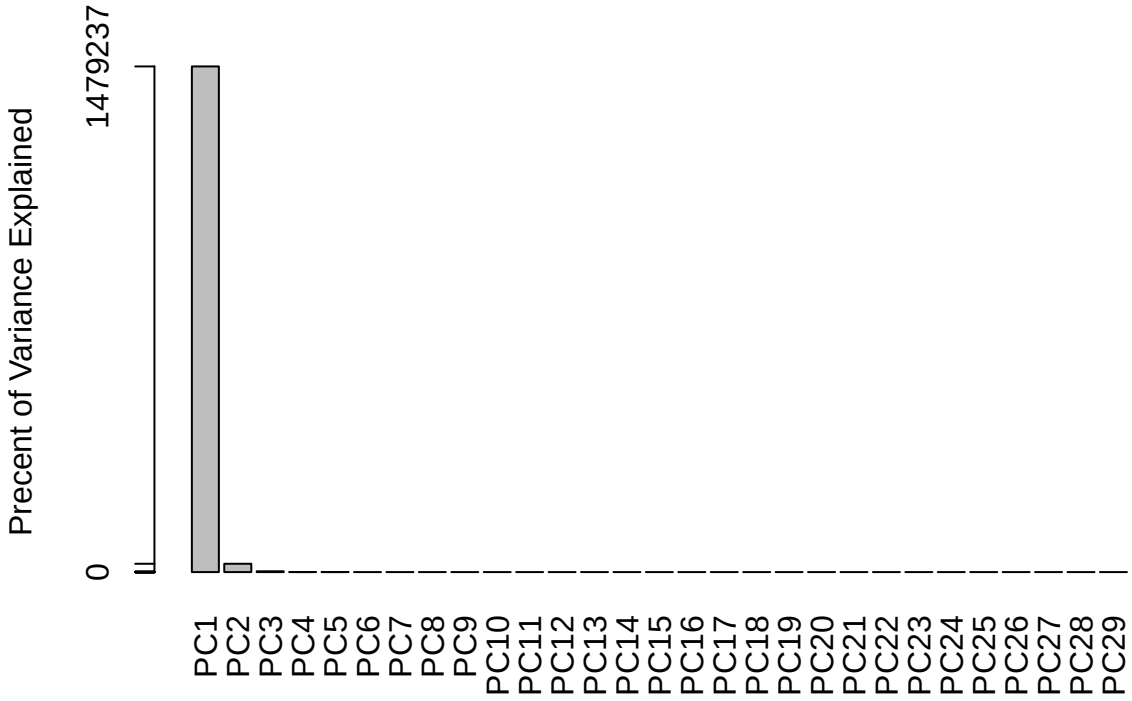
```
pr.var <- wisc.pr$sdev^2
head(pr.var)
```

```
## [1] 4.437711e+05 7.309470e+03 7.037265e+02 5.458498e+01 3.983975e+01
## [6] 2.953395e+00
```

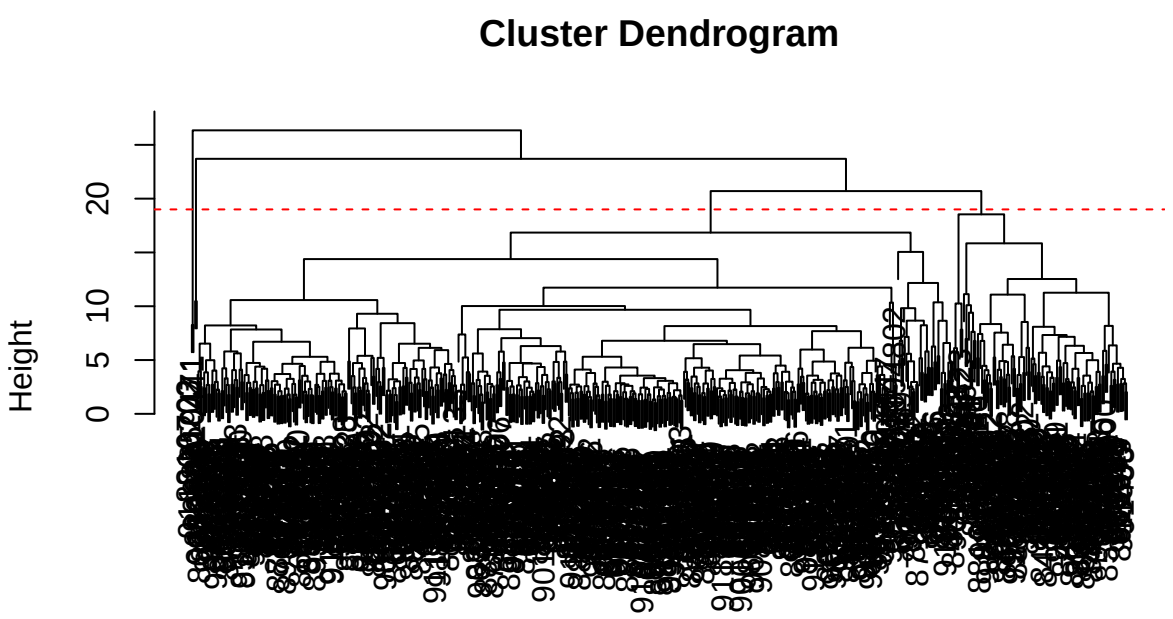
```
pve <- pr.var / 30
plot(pve, xlab = "Principal Component",
     ylab = "Proportion of Variance Explained",
     ylim = c(0, 1), type = "o")
```




```
barplot(pve, ylab = "Precent of Variance Explained",
        names.arg=paste0("PC",1:length(pve)), las=2, axes = FALSE)
axis(2, at=pve, labels=round(pve,2)*100 )
```




```
data.scaled <- scale(wisc.data[,-1])
data.dist <- dist(data.scaled)
wisc.hclust <- hclust(data.dist, method = "complete")
plot(wisc.hclust)
abline(h =19 , col="red", lty=2)
```




```
wisc.hclust.clusters <- cutree(wisc.hclust, k = 4)
table(wisc.hclust.clusters, diagnosis)
```

```
##                     diagnosis
## wisc.hclust.clusters 6.981 7.691 7.729 7.76 8.196 8.219 8.571 8.597 8.598 8.618
##                    1     0     0     0    0     0     0     0     0     0     0
##                    2     1     1     1    1     1     1     1     1     1     1
##                    3     0     0     0    0     0     0     0     0     0     0
##                    4     0     0     0    0     0     0     0     0     0     0
##                     diagnosis
## wisc.hclust.clusters 8.671 8.726 8.734 8.878 8.888 8.95 9 9.029 9.042 9.173
##                    1     0     0     0     0     0    0 0     0     0     0
##                    2     1     1     1     1     1    1 1     0     1     1
##                    3     0     0     0     0     0    0 0     1     0     0
##                    4     0     0     0     0     0    0 0     0     0     0
##                     diagnosis
## wisc.hclust.clusters 9.268 9.295 9.333 9.397 9.405 9.423 9.436 9.465 9.504
##                    1     0     0     0     0     0     0     0     0     0
##                    2     1     1     1     1     1     1     1     1     1
##                    3     0     0     0     0     0     0     0     0     0
##                    4     0     0     0     0     0     0     0     0     0
##                     diagnosis
## wisc.hclust.clusters 9.567 9.606 9.667 9.668 9.676 9.683 9.72 9.731 9.738 9.742
##                    1     0     0     0     0     0     0    0     0     0     0
##                    2     1     1     1     1     1     1    1     0     1     2
```

```
##                     3     0     0     0     0     0     0     0     1     0     0
##                     4     0     0     0     0     0     0     0     0     0     0
##                      diagnosis
## wisc.hclust.clusters 9.755 9.777 9.787 9.847 9.876 9.904 10.03 10.05 10.08
##                     1     0     0     0     0     0     0     0     0     0
##                     2     1     1     1     1     2     1     1     1     1
##                     3     0     0     0     0     0     0     0     0     0
##                     4     0     0     0     0     0     0     0     0     0
##                      diagnosis
## wisc.hclust.clusters 10.16 10.17 10.18 10.2 10.25 10.26 10.29 10.32 10.44 10.48
##                     1     0     0     0    0     0     0     0     0     0     0
##                     2     1     1     1    1     1     3     1     1     1     2
##                     3     0     0     0    0     0     0     0     0     0     0
##                     4     0     0     0    0     0     0     0     0     0     0
##                      diagnosis
## wisc.hclust.clusters 10.49 10.51 10.57 10.6 10.65 10.66 10.71 10.75 10.8 10.82
##                     1     0     0     0    0     0     0     0     0    0     0
##                     2     2     2     2    1     1     1     1     1    2     1
##                     3     0     0     0    0     0     0     0     0    0     0
##                     4     0     0     0    0     0     0     0     0    0     0
##                      diagnosis
## wisc.hclust.clusters 10.86 10.88 10.9 10.91 10.94 10.95 10.96 10.97 11.04 11.06
##                     1     0     0    0     0     0     0     0     0     0     0
##                     2     1     1    1     1     1     1     1     1     2     3
##                     3     0     0    0     0     0     0     0     0     0     0
##                     4     0     0    0     0     0     0     0     0     0     0
##                      diagnosis
## wisc.hclust.clusters 11.08 11.13 11.14 11.15 11.16 11.2 11.22 11.25 11.26 11.27
##                     1     1     0     0     0     0    0     0     0     0     0
##                     2     1     2     1     1     1    1     2     1     2     2
##                     3     0     0     0     0     0    0     0     0     0     0
##                     4     0     0     0     0     0    0     0     0     0     0
##                      diagnosis
## wisc.hclust.clusters 11.28 11.29 11.3 11.31 11.32 11.33 11.34 11.36 11.37 11.41
##                     1     0     0    0     0     0     0     0     0     0     0
##                     2     1     1    1     1     1     1     2     1     1     2
##                     3     0     0    0     0     0     0     0     0     0     0
##                     4     0     0    0     0     0     0     0     0     0     0
##                      diagnosis
## wisc.hclust.clusters 11.42 11.43 11.45 11.46 11.47 11.49 11.5 11.51 11.52 11.54
##                     1     1     0     0     0     0     0    0     0     0     0
##                     2     0     2     1     1     1     1    1     1     2     2
##                     3     0     0     0     0     0     0    0     0     0     0
##                     4     0     0     0     0     0     0    0     0     0     0
##                      diagnosis
## wisc.hclust.clusters 11.57 11.6 11.61 11.62 11.63 11.64 11.66 11.67 11.68 11.69
##                     1     0    0     0     0     0     0     0     0     0     0
##                     2     1    3     1     1     1     1     1     1     1     1
##                     3     0    0     0     0     0     0     0     0     0     0
##                     4     0    0     0     0     0     0     0     0     0     0
##                      diagnosis
## wisc.hclust.clusters 11.7 11.71 11.74 11.75 11.76 11.8 11.81 11.84 11.85 11.87
##                     1    0     0     0     0     0    0     0     1     0     0
##                     2    1     3     2     2     2    2     1     1     1     1
```

```
##                     3    0     0     0     0     0    0     0     0     0     0
##                     4    0     0     0     0     0    0     0     0     0     0
##                      diagnosis
## wisc.hclust.clusters 11.89 11.9 11.93 11.94 11.95 11.99 12 12.03 12.04 12.05
##                     1     0    0     0     0     0     0  0     0     0     0
##                     2     3    1     2     2     1     1  2     1     1     2
##                     3     0    0     0     0     0     0  0     0     0     0
##                     4     0    0     0     0     0     0  0     0     0     0
##                      diagnosis
## wisc.hclust.clusters 12.06 12.07 12.1 12.16 12.18 12.19 12.2 12.21 12.22 12.23
##                     1     0     0    0     0     0     0    0     0     0     0
##                     2     2     1    1     1     3     1    1     2     1     1
##                     3     0     0    0     0     0     0    0     0     0     0
##                     4     0     0    0     0     0     0    0     0     0     0
##                      diagnosis
## wisc.hclust.clusters 12.25 12.27 12.3 12.31 12.32 12.34 12.36 12.39 12.4 12.42
##                     1     0     0    0     0     0     1     0     0    0     0
##                     2     2     2    2     1     1     3     2     1    1     1
##                     3     0     0    0     0     0     0     0     0    0     0
##                     4     0     0    0     0     0     0     0     0    0     0
##                      diagnosis
## wisc.hclust.clusters 12.43 12.45 12.46 12.47 12.49 12.54 12.56 12.58 12.62
##                     1     0     1     1     0     0     0     0     0     0
##                     2     1     1     2     2     1     2     1     1     2
##                     3     0     0     0     0     0     0     0     0     0
##                     4     0     0     0     0     0     0     0     0     0
##                      diagnosis
## wisc.hclust.clusters 12.63 12.65 12.67 12.68 12.7 12.72 12.75 12.76 12.77 12.78
##                     1     0     0     0     1    0     0     0     0     0     0
##                     2     1     1     1     0    1     2     1     2     3     1
##                     3     0     0     0     0    0     0     0     0     0     0
##                     4     0     0     0     0    0     0     0     0     0     0
##                      diagnosis
## wisc.hclust.clusters 12.8 12.81 12.83 12.85 12.86 12.87 12.88 12.89 12.9 12.91
##                     1    0     0     1     0     0     0     0     0    0     0
##                     2    1     1     1     1     2     2     2     3    1     1
##                     3    0     0     0     0     0     0     0     0    0     0
##                     4    0     0     0     0     0     0     0     0    0     0
##                      diagnosis
## wisc.hclust.clusters 12.94 12.95 12.96 12.98 12.99 13 13.01 13.03 13.05 13.08
##                     1     0     0     0     0     0  1     0     0     0     0
##                     2     1     1     1     1     1  2     1     1     3     1
##                     3     0     0     0     0     0  0     0     0     0     0
##                     4     0     0     0     0     0  0     0     0     0     0
##                      diagnosis
## wisc.hclust.clusters 13.11 13.14 13.15 13.16 13.17 13.2 13.21 13.24 13.27 13.28
##                     1     1     0     0     0     1    0     0     0     0     1
##                     2     1     1     1     1     2    2     2     1     2     1
##                     3     0     0     0     0     0    0     0     0     0     0
##                     4     0     0     0     0     0    0     0     0     0     0
##                      diagnosis
## wisc.hclust.clusters 13.3 13.34 13.37 13.38 13.4 13.43 13.44 13.45 13.46 13.47
##                     1    0     1     0     0    1     0     0     0     0     0
##                     2    1     0     1     1    1     1     1     1     2     1
```

```
##                     3    0     0     0     0    0     0     0     0     0     0
##                     4    0     0     0     0    0     0     0     0     0     0
##                      diagnosis
## wisc.hclust.clusters 13.48 13.49 13.5 13.51 13.53 13.54 13.56 13.59 13.61 13.62
##                     1     0     0    0     0     0     0     0     0     0     0
##                     2     1     1    1     1     1     1     1     2     2     1
##                     3     0     0    0     0     0     0     0     0     0     0
##                     4     0     0    0     0     0     0     0     0     0     0
##                      diagnosis
## wisc.hclust.clusters 13.64 13.65 13.66 13.68 13.69 13.7 13.71 13.73 13.74 13.75
##                     1     0     0     0     0     0    0     1     1     0     0
##                     2     2     1     2     1     1    1     1     0     1     1
##                     3     0     0     0     0     0    0     0     0     0     0
##                     4     0     0     0     0     0    0     0     0     0     0
##                      diagnosis
## wisc.hclust.clusters 13.77 13.78 13.8 13.81 13.82 13.85 13.86 13.87 13.88 13.9
##                     1     0     0    0     1     1     0     1     0     0    0
##                     2     2     1    1     0     0     3     0     2     1    2
##                     3     0     0    0     0     0     0     0     0     0    0
##                     4     0     0    0     0     0     0     0     0     0    0
##                      diagnosis
## wisc.hclust.clusters 13.94 13.96 13.98 14.02 14.03 14.04 14.05 14.06 14.11
##                     1     0     0     0     0     0     0     0     0     0
##                     2     1     1     1     1     1     1     1     1     1
##                     3     0     0     0     0     0     0     0     0     0
##                     4     0     0     0     0     0     0     0     0     0
##                      diagnosis
## wisc.hclust.clusters 14.19 14.2 14.22 14.25 14.26 14.27 14.29 14.34 14.4 14.41
##                     1     0    0     1     1     0     0     0     0    0     0
##                     2     1    1     1     1     2     1     1     1    1     1
##                     3     0    0     0     0     0     0     0     0    0     0
##                     4     0    0     0     0     0     0     0     0    0     0
##                      diagnosis
## wisc.hclust.clusters 14.42 14.44 14.45 14.47 14.48 14.5 14.53 14.54 14.58 14.59
##                     1     0     0     0     0     0    0     0     1     1     0
##                     2     2     1     1     1     1    1     2     0     1     1
##                     3     0     0     0     0     0    0     0     0     0     0
##                     4     0     0     0     0     0    0     0     0     0     0
##                      diagnosis
## wisc.hclust.clusters 14.6 14.61 14.62 14.64 14.68 14.69 14.71 14.74 14.76 14.78
##                     1    0     0     0     0     0     0     1     0     0     1
##                     2    1     1     1     2     1     1     0     1     1     0
##                     3    0     0     0     0     0     0     0     0     0     0
##                     4    0     0     0     0     0     0     0     0     0     0
##                      diagnosis
## wisc.hclust.clusters 14.8 14.81 14.86 14.87 14.9 14.92 14.95 14.96 14.97 14.99
##                     1    0     0     0     1    1     0     1     0     0     0
##                     2    1     1     2     1    0     1     1     1     2     2
##                     3    0     0     0     0    0     0     0     0     0     0
##                     4    0     0     0     0    0     0     0     0     0     0
##                      diagnosis
## wisc.hclust.clusters 15 15.04 15.05 15.06 15.08 15.1 15.12 15.13 15.19 15.22
##                     1  0     0     0     1     0    0     0     0     0     1
##                     2  1     1     1     0     1    2     1     1     1     0
```

```
##                     3  0     0     0     0     0    0     0     0     0     0
##                     4  0     0     0     0     0    0     0     0     0     0
##                      diagnosis
## wisc.hclust.clusters 15.27 15.28 15.3 15.32 15.34 15.37 15.46 15.49 15.5 15.53
##                     1     0     0    1     1     0     0     2     1    1     1
##                     2     1     1    0     0     1     1     1     0    0     0
##                     3     0     0    0     0     0     0     0     0    0     0
##                     4     0     0    0     0     0     0     0     0    0     0
##                      diagnosis
## wisc.hclust.clusters 15.61 15.66 15.7 15.71 15.73 15.75 15.78 15.85 16.02 16.03
##                     1     0     1    0     0     0     1     2     0     0     1
##                     2     1     0    1     1     1     1     0     1     1     0
##                     3     0     0    0     0     0     0     0     0     0     0
##                     4     0     0    0     0     0     0     0     0     0     0
##                      diagnosis
## wisc.hclust.clusters 16.07 16.11 16.13 16.14 16.16 16.17 16.24 16.25 16.26
##                     1     0     0     2     0     0     0     1     0     0
##                     2     1     1     0     1     1     1     0     1     1
##                     3     0     0     0     0     0     0     0     0     0
##                     4     0     0     0     0     0     0     0     0     0
##                      diagnosis
## wisc.hclust.clusters 16.27 16.3 16.35 16.46 16.5 16.6 16.65 16.69 16.74 16.78
##                     1     0    0     1     1    0    0     0     0     1     0
##                     2     1    1     0     0    1    1     1     1     0     1
##                     3     0    0     0     0    0    0     0     0     0     0
##                     4     0    0     0     0    0    0     0     0     0     0
##                      diagnosis
## wisc.hclust.clusters 16.84 17.01 17.02 17.05 17.06 17.08 17.14 17.19 17.2 17.27
##                     1     0     0     1     1     0     0     1     0    1     0
##                     2     1     1     0     0     1     1     0     1    0     1
##                     3     0     0     0     0     0     0     0     0    0     0
##                     4     0     0     0     0     0     0     0     0    0     0
##                      diagnosis
## wisc.hclust.clusters 17.29 17.3 17.35 17.42 17.46 17.47 17.54 17.57 17.6 17.68
##                     1     1    0     0     0     0     1     0     0    1     0
##                     2     0    1     1     1     1     0     1     1    0     1
##                     3     0    0     0     0     0     0     0     0    0     0
##                     4     0    0     0     0     0     0     0     0    0     0
##                      diagnosis
## wisc.hclust.clusters 17.75 17.85 17.91 17.93 17.95 17.99 18.01 18.03 18.05
##                     1     0     0     1     0     0     1     0     0     1
##                     2     1     1     0     1     1     1     1     1     0
##                     3     0     0     0     0     0     0     0     0     0
##                     4     0     0     0     0     0     0     0     0     0
##                      diagnosis
## wisc.hclust.clusters 18.08 18.22 18.25 18.31 18.45 18.46 18.49 18.61 18.63
##                     1     0     0     0     0     0     0     0     0     1
##                     2     1     2     1     2     1     1     1     1     0
##                     3     0     0     0     0     0     0     0     0     0
##                     4     0     0     0     0     0     0     0     0     0
##                      diagnosis
## wisc.hclust.clusters 18.65 18.66 18.77 18.81 18.82 18.94 19 19.02 19.07 19.1
##                     1     1     0     0     0     0     0  0     0     1    1
##                     2     0     1     1     1     1     1  1     1     0    0
```

```
##                     3     0     0     0     0     0     0  0     0     0    0
##                     4     0     0     0     0     0     0  0     0     0    0
##                      diagnosis
## wisc.hclust.clusters 19.16 19.17 19.18 19.19 19.21 19.27 19.4 19.44 19.45 19.53
##                     1     1     1     0     1     0     1    0     1     0     1
##                     2     0     0     1     0     1     0    2     0     1     1
##                     3     0     0     0     0     0     0    0     0     0     0
##                     4     0     0     0     0     0     0    0     0     0     0
##                      diagnosis
## wisc.hclust.clusters 19.55 19.59 19.68 19.69 19.73 19.79 19.8 19.81 19.89 20.09
##                     1     2     1     1     1     1     1    1     0     0     1
##                     2     0     1     0     0     0     0    0     1     1     0
##                     3     0     0     0     0     0     0    0     0     0     0
##                     4     0     0     0     0     0     0    0     0     0     0
##                      diagnosis
## wisc.hclust.clusters 20.13 20.16 20.18 20.2 20.26 20.29 20.31 20.34 20.44 20.47
##                     1     0     0     2    1     1     0     0     1     0     1
##                     2     1     1     0    0     0     1     1     0     1     0
##                     3     0     0     0    0     0     0     0     0     0     0
##                     4     0     0     0    0     0     0     0     0     0     0
##                      diagnosis
## wisc.hclust.clusters 20.48 20.51 20.55 20.57 20.58 20.59 20.6 20.64 20.73 20.92
##                     1     0     0     1     0     1     1    1     0     1     1
##                     2     1     1     0     1     0     0    0     1     0     0
##                     3     0     0     0     0     0     0    0     0     0     0
##                     4     0     0     0     0     0     0    0     0     0     0
##                      diagnosis
## wisc.hclust.clusters 20.94 21.09 21.1 21.16 21.37 21.56 21.61 21.71 21.75 22.01
##                     1     1     1    0     0     0     1     1     1     1     1
##                     2     0     0    1     1     1     0     0     0     0     0
##                     3     0     0    0     0     0     0     0     0     0     0
##                     4     0     0    0     0     0     0     0     0     0     0
##                      diagnosis
## wisc.hclust.clusters 22.27 23.09 23.21 23.27 23.29 23.51 24.25 24.63 25.22
##                     1     1     1     1     0     1     1     1     1     1
##                     2     0     0     0     1     0     0     0     0     0
##                     3     0     0     0     0     0     0     0     0     0
##                     4     0     0     0     0     0     0     0     0     0
##                      diagnosis
## wisc.hclust.clusters 25.73 27.22 27.42 28.11
##                     1     1     1     0     0
##                     2     0     0     0     0
##                     3     0     0     0     0
##                     4     0     0     1     1
```

```
grps <- cutree(wisc.hclust, k=2)
table(grps)
```

```
## grps
##   1   2
## 567   2
```

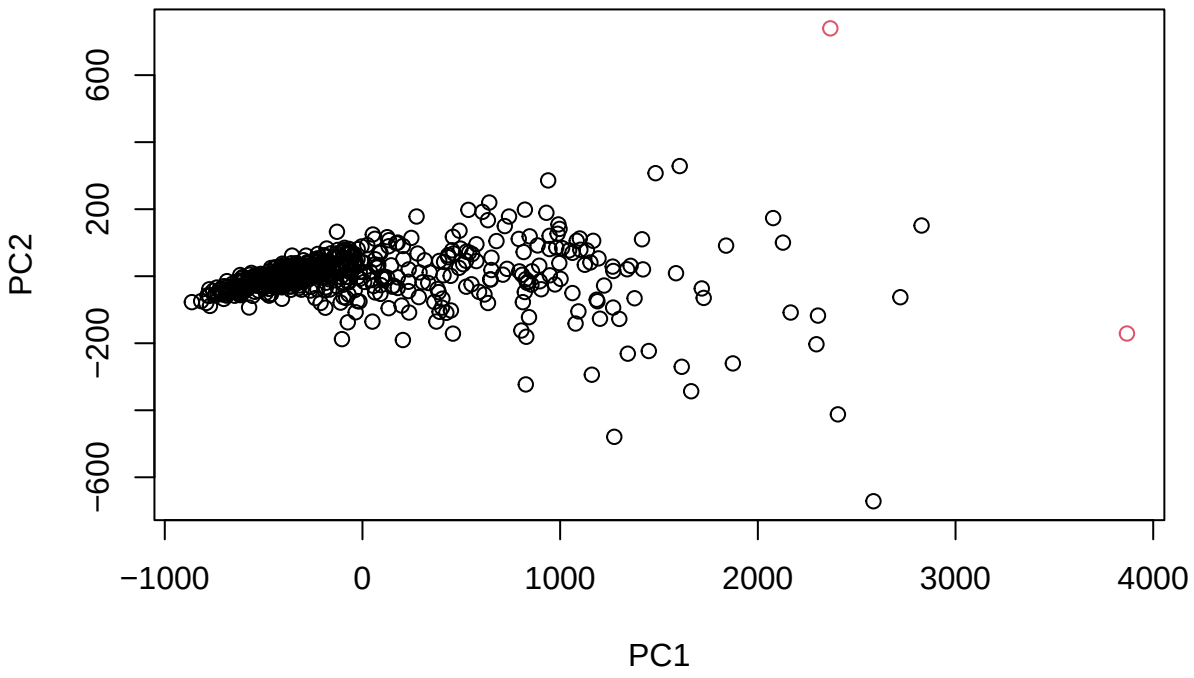
```
table(grps, diagnosis)
```

```
##     diagnosis
## grps 6.981 7.691 7.729 7.76 8.196 8.219 8.571 8.597 8.598 8.618 8.671 8.726
##    1     1     1     1    1     1     1     1     1     1     1     1     1
##    2     0     0     0    0     0     0     0     0     0     0     0     0
##     diagnosis
## grps 8.734 8.878 8.888 8.95 9 9.029 9.042 9.173 9.268 9.295 9.333 9.397 9.405
##    1     1     1     1    1 1     1     1     1     1     1     1     1     1
##    2     0     0     0    0 0     0     0     0     0     0     0     0     0
##     diagnosis
## grps 9.423 9.436 9.465 9.504 9.567 9.606 9.667 9.668 9.676 9.683 9.72 9.731
##    1     1     1     1     1     1     1     1     1     1     1    1     1
##    2     0     0     0     0     0     0     0     0     0     0    0     0
##     diagnosis
## grps 9.738 9.742 9.755 9.777 9.787 9.847 9.876 9.904 10.03 10.05 10.08 10.16
##    1     1     2     1     1     1     1     2     1     1     1     1     1
##    2     0     0     0     0     0     0     0     0     0     0     0     0
##     diagnosis
## grps 10.17 10.18 10.2 10.25 10.26 10.29 10.32 10.44 10.48 10.49 10.51 10.57
##    1     1     1    1     1     3     1     1     1     2     2     2     2
##    2     0     0    0     0     0     0     0     0     0     0     0     0
##     diagnosis
## grps 10.6 10.65 10.66 10.71 10.75 10.8 10.82 10.86 10.88 10.9 10.91 10.94 10.95
##    1    1     1     1     1     1    2     1     1     1    1     1     1     1
##    2    0     0     0     0     0    0     0     0     0    0     0     0     0
##     diagnosis
## grps 10.96 10.97 11.04 11.06 11.08 11.13 11.14 11.15 11.16 11.2 11.22 11.25
##    1     1     1     2     3     2     2     1     1     1    1     2     1
##    2     0     0     0     0     0     0     0     0     0    0     0     0
##     diagnosis
## grps 11.26 11.27 11.28 11.29 11.3 11.31 11.32 11.33 11.34 11.36 11.37 11.41
##    1     2     2     1     1    1     1     1     1     2     1     1     2
##    2     0     0     0     0    0     0     0     0     0     0     0     0
##     diagnosis
## grps 11.42 11.43 11.45 11.46 11.47 11.49 11.5 11.51 11.52 11.54 11.57 11.6
##    1     1     2     1     1     1     1    1     1     2     2     1    3
##    2     0     0     0     0     0     0    0     0     0     0     0    0
##     diagnosis
## grps 11.61 11.62 11.63 11.64 11.66 11.67 11.68 11.69 11.7 11.71 11.74 11.75
##    1     1     1     1     1     1     1     1     1    1     3     2     2
##    2     0     0     0     0     0     0     0     0    0     0     0     0
##     diagnosis
## grps 11.76 11.8 11.81 11.84 11.85 11.87 11.89 11.9 11.93 11.94 11.95 11.99 12
##    1     2    2     1     2     1     1     3    1     2     2     1     1  2
##    2     0    0     0     0     0     0     0    0     0     0     0     0  0
##     diagnosis
## grps 12.03 12.04 12.05 12.06 12.07 12.1 12.16 12.18 12.19 12.2 12.21 12.22
##    1     1     1     2     2     1    1     1     3     1    1     2     1
##    2     0     0     0     0     0    0     0     0     0    0     0     0
##     diagnosis
## grps 12.23 12.25 12.27 12.3 12.31 12.32 12.34 12.36 12.39 12.4 12.42 12.43
##    1     1     2     2    2     1     1     4     2     1    1     1     1
```

```
##    2     0     0     0    0     0     0     0     0     0    0     0     0
##     diagnosis
## grps 12.45 12.46 12.47 12.49 12.54 12.56 12.58 12.62 12.63 12.65 12.67 12.68
##    1     2     3     2     1     2     1     1     2     1     1     1     1
##    2     0     0     0     0     0     0     0     0     0     0     0     0
##     diagnosis
## grps 12.7 12.72 12.75 12.76 12.77 12.78 12.8 12.81 12.83 12.85 12.86 12.87
##    1    1     2     1     2     3     1    1     1     2     1     2     2
##    2    0     0     0     0     0     0    0     0     0     0     0     0
##     diagnosis
## grps 12.88 12.89 12.9 12.91 12.94 12.95 12.96 12.98 12.99 13 13.01 13.03 13.05
##    1     2     3    1     1     1     1     1     1     1  3     1     1     3
##    2     0     0    0     0     0     0     0     0     0  0     0     0     0
##     diagnosis
## grps 13.08 13.11 13.14 13.15 13.16 13.17 13.2 13.21 13.24 13.27 13.28 13.3
##    1     1     2     1     1     1     3    2     2     1     2     2    1
##    2     0     0     0     0     0     0    0     0     0     0     0    0
##     diagnosis
## grps 13.34 13.37 13.38 13.4 13.43 13.44 13.45 13.46 13.47 13.48 13.49 13.5
##    1     1     1     1    2     1     1     1     2     1     1     1    1
##    2     0     0     0    0     0     0     0     0     0     0     0    0
##     diagnosis
## grps 13.51 13.53 13.54 13.56 13.59 13.61 13.62 13.64 13.65 13.66 13.68 13.69
##    1     1     1     1     1     2     2     1     2     1     2     1     1
##    2     0     0     0     0     0     0     0     0     0     0     0     0
##     diagnosis
## grps 13.7 13.71 13.73 13.74 13.75 13.77 13.78 13.8 13.81 13.82 13.85 13.86
##    1    1     2     1     1     1     2     1    1     1     1     3     1
##    2    0     0     0     0     0     0     0    0     0     0     0     0
##     diagnosis
## grps 13.87 13.88 13.9 13.94 13.96 13.98 14.02 14.03 14.04 14.05 14.06 14.11
##    1     2     1    2     1     1     1     1     1     1     1     1     1
##    2     0     0    0     0     0     0     0     0     0     0     0     0
##     diagnosis
## grps 14.19 14.2 14.22 14.25 14.26 14.27 14.29 14.34 14.4 14.41 14.42 14.44
##    1     1    1     2     2     2     1     1     1    1     1     2     1
##    2     0    0     0     0     0     0     0     0    0     0     0     0
##     diagnosis
## grps 14.45 14.47 14.48 14.5 14.53 14.54 14.58 14.59 14.6 14.61 14.62 14.64
##    1     1     1     1    1     2     1     2     1    1     1     1     2
##    2     0     0     0    0     0     0     0     0    0     0     0     0
##     diagnosis
## grps 14.68 14.69 14.71 14.74 14.76 14.78 14.8 14.81 14.86 14.87 14.9 14.92
##    1     1     1     1     1     1     1    1     1     2     2    1     1
##    2     0     0     0     0     0     0    0     0     0     0    0     0
##     diagnosis
## grps 14.95 14.96 14.97 14.99 15 15.04 15.05 15.06 15.08 15.1 15.12 15.13 15.19
##    1     2     1     2     2  1     1     1     1     1    2     1     1     1
##    2     0     0     0     0  0     0     0     0     0    0     0     0     0
##     diagnosis
## grps 15.22 15.27 15.28 15.3 15.32 15.34 15.37 15.46 15.49 15.5 15.53 15.61
##    1     1     1     1    1     1     1     1     3     1    1     1     1
##    2     0     0     0    0     0     0     0     0     0    0     0     0
##     diagnosis
```

```
## grps 15.66 15.7 15.71 15.73 15.75 15.78 15.85 16.02 16.03 16.07 16.11 16.13
##    1     1    1     1     1     2     2     1     1     1     1     1     2
##    2     0    0     0     0     0     0     0     0     0     0     0     0
##     diagnosis
## grps 16.14 16.16 16.17 16.24 16.25 16.26 16.27 16.3 16.35 16.46 16.5 16.6 16.65
##    1     1     1     1     1     1     1     1    1     1     1    1    1     1
##    2     0     0     0     0     0     0     0    0     0     0    0    0     0
##     diagnosis
## grps 16.69 16.74 16.78 16.84 17.01 17.02 17.05 17.06 17.08 17.14 17.19 17.2
##    1     1     1     1     1     1     1     1     1     1     1     1    1
##    2     0     0     0     0     0     0     0     0     0     0     0    0
##     diagnosis
## grps 17.27 17.29 17.3 17.35 17.42 17.46 17.47 17.54 17.57 17.6 17.68 17.75
##    1     1     1    1     1     1     1     1     1     1    1     1     1
##    2     0     0    0     0     0     0     0     0     0    0     0     0
##     diagnosis
## grps 17.85 17.91 17.93 17.95 17.99 18.01 18.03 18.05 18.08 18.22 18.25 18.31
##    1     1     1     1     1     2     1     1     1     1     2     1     2
##    2     0     0     0     0     0     0     0     0     0     0     0     0
##     diagnosis
## grps 18.45 18.46 18.49 18.61 18.63 18.65 18.66 18.77 18.81 18.82 18.94 19 19.02
##    1     1     1     1     1     1     1     1     1     1     1     1  1     1
##    2     0     0     0     0     0     0     0     0     0     0     0  0     0
##     diagnosis
## grps 19.07 19.1 19.16 19.17 19.18 19.19 19.21 19.27 19.4 19.44 19.45 19.53
##    1     1    1     1     1     1     1     1     1    2     1     1     2
##    2     0    0     0     0     0     0     0     0    0     0     0     0
##     diagnosis
## grps 19.55 19.59 19.68 19.69 19.73 19.79 19.8 19.81 19.89 20.09 20.13 20.16
##    1     2     2     1     1     1     1    1     1     1     1     1     1
##    2     0     0     0     0     0     0    0     0     0     0     0     0
##     diagnosis
## grps 20.18 20.2 20.26 20.29 20.31 20.34 20.44 20.47 20.48 20.51 20.55 20.57
##    1     2    1     1     1     1     1     1     1     1     1     1     1
##    2     0    0     0     0     0     0     0     0     0     0     0     0
##     diagnosis
## grps 20.58 20.59 20.6 20.64 20.73 20.92 20.94 21.09 21.1 21.16 21.37 21.56
##    1     1     1    1     1     1     1     1     1    1     1     1     1
##    2     0     0    0     0     0     0     0     0    0     0     0     0
##     diagnosis
## grps 21.61 21.71 21.75 22.01 22.27 23.09 23.21 23.27 23.29 23.51 24.25 24.63
##    1     1     1     1     1     1     1     1     1     1     1     1     1
##    2     0     0     0     0     0     0     0     0     0     0     0     0
##     diagnosis
## grps 25.22 25.73 27.22 27.42 28.11
##    1     1     1     1     0     0
##    2     0     0     0     1     1
```

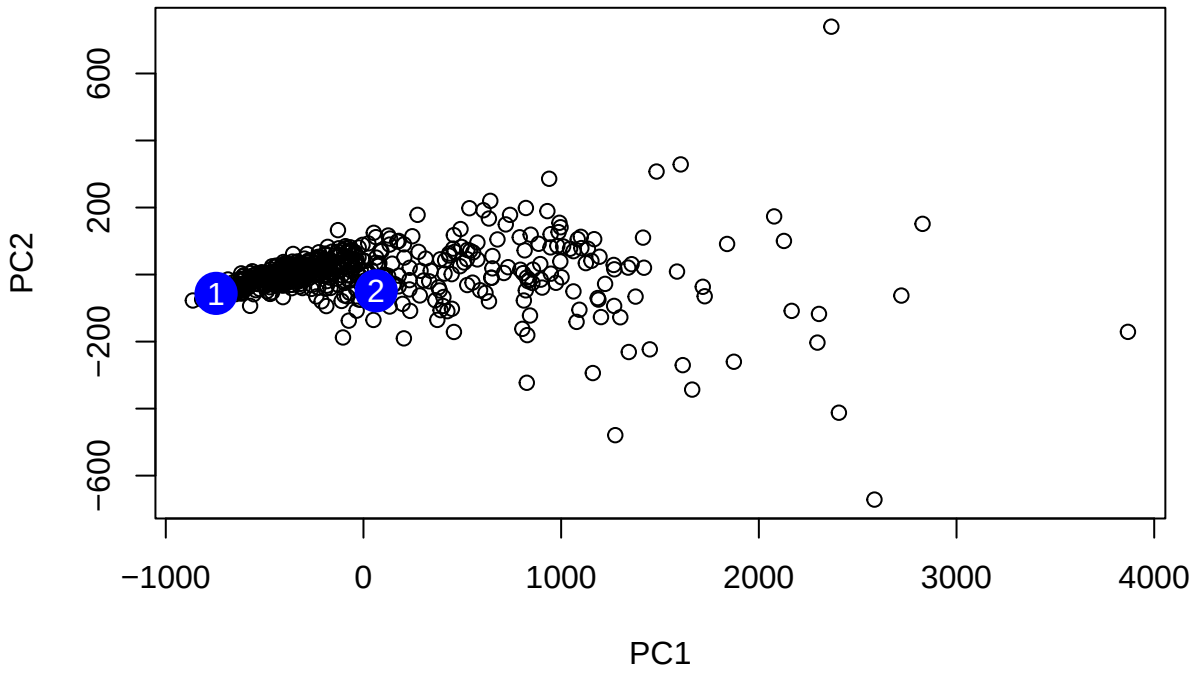
```
plot(wisc.pr$x[,1:2], col=grps)
```




```
url <- "https://tinyurl.com/new-samples-CSV"
new <- read.csv(url)
npc <- predict(wisc.pr, newdata=new)
npc
```

```
##             PC1       PC2       PC3       PC4       PC5       PC6       PC7
## [1,] -745.58228 -56.15114 -21.14308 -3.269575  9.342062  2.265565 -1.156815
## [2,]   64.40863 -48.46994 -15.93951 12.065347 -4.676632 -1.080107 -0.277376
##            PC8        PC9       PC10        PC11        PC12        PC13
## [1,] 0.7563343 -0.1032757  0.0688537 -0.09202633  0.06375044 -0.01669926
## [2,] 0.7260377  0.2276038 -0.7097099 -0.30447319 -0.01883250  0.11744318
##              PC14         PC15         PC16        PC17         PC18
## [1,] -0.023223923  0.009950655 -0.008509815 -0.01948388 -0.007793169
## [2,] -0.007291041 -0.018657551 -0.011000041 -0.02647465  0.000611957
##              PC19         PC20       PC21          PC22         PC23
## [1,] -0.031830403 -0.005101987 0.01845962 -0.0005420926  0.006199072
## [2,]  0.009654825 -0.018181800 0.01048644  0.0088785987 -0.007183325
##              PC24         PC25          PC26         PC27          PC28
## [1,]  0.005901854 -0.005404442 -0.0001697478 -0.003258087 -0.0012304929
## [2,] -0.003020084 -0.003319119 -0.0021016368  0.002101994 -0.0009562871
##              PC29
## [1,] 0.0009908106
## [2,] 0.0008586404
```

```
plot(wisc.pr$x[,1:2])
points(npc[,1], npc[,2], col="blue", pch=16, cex=3)
text(npc[,1], npc[,2], c(1,2), col="white")
```




```
sessionInfo()
```

```
## R version 4.1.2 (2021-11-01)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 19043)
##
## Matrix products: default
##
## locale:
## [1] LC_COLLATE=English_United States.1252
## [2] LC_CTYPE=English_United States.1252
## [3] LC_MONETARY=English_United States.1252
## [4] LC_NUMERIC=C
## [5] LC_TIME=English_United States.1252
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
## [1] ggplot2_3.3.5
##
```

```
## loaded via a namespace (and not attached):
##  [1] pillar_1.7.0     compiler_4.1.2   highr_0.9        tools_4.1.2
##  [5] digest_0.6.29    evaluate_0.15    lifecycle_1.0.1  tibble_3.1.6
##  [9] gtable_0.3.0     pkgconfig_2.0.3  rlang_1.0.2      cli_3.1.1
## [13] rstudioapi_0.13  yaml_2.3.5       xfun_0.30        fastmap_1.1.0
## [17] withr_2.5.0      stringr_1.4.0    dplyr_1.0.8      knitr_1.37
## [21] generics_0.1.2   vctrs_0.3.8      grid_4.1.2       tidyselect_1.1.2
## [25] glue_1.6.2       R6_2.5.1         fansi_1.0.2      rmarkdown_2.12
## [29] farver_2.1.0     purrr_0.3.4      magrittr_2.0.2   scales_1.1.1
## [33] ellipsis_0.3.2   htmltools_0.5.2  colorspace_2.0-3 labeling_0.4.2
## [37] utf8_1.2.2       stringi_1.7.6    munsell_0.5.0    crayon_1.5.0
```

- **Q1**. How many observations are in this dataset?

  569 observations

- **Q2**. How many of the observations have a malignant diagnosis?

  212 malignant diagnoses

- **Q3**. How many variables/features in the data are suffixed with _mean?

  10 variables

- **Q4**. From your results, what proportion of the original variance is captured by the first principal components (PC1)?

  98.2%

- **Q5**. How many principal components (PCs) are required to describe at least 70% of the original variance in the data?

  1

- **Q6**. How many principal components (PCs) are required to describe at least 90% of the original variance in the data?

  1

- **Q7.** What stands out to you about this plot? Is it easy or difficult to understand? Why

  This biplot is unreadable and needs to be cleaned up in order to understand. All the numbers and text is on top of each other.

- **Q8.** Generate a similar plot for principal components 1 and 3.
  What do you notice about these plots?

  The diagnosis B tends towards the right of the plot with diagnosis M towards the left.

- **Q9.** For the first principal component, what is the component of the loading vector (i.e. `wisc.pr$rotation[,1]`) for the feature `concave.points_mean`?

- **Q10.** What is the minimum number of principal components required to explain 80% of the variance of the data?

  1 principal component

- **Q11.** Using the `plot()` and `abline()` functions, what is the height at which the clustering model has 4 clusters?

  19

- **Q12.** Can you find a better cluster vs diagnoses match by cutting into a different number of clusters between 2 and 10?

  Yeah, you can probably find a better cluster match with 2 clusters with one for malignant cells and one fore benign

- **Q13.** Which method gives your favorite results for the same data.dist dataset? Explain your reasoning.

  I like the “complete” method because it gives you everything and from there you can interpret the clustering procedure

- Q18. Which of these new patients should we prioritize for follow up based on your results?

  We should prioritize patient 1