Lab6_1

2024-05-07

Section 1-3:

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.3.2
```

```
data.df <- read.csv("phagocyte_PCA.csv", row.names=1)
# 5
gene.PCA <- prcomp( t(data.df) )
# 6
PCA.summary <- summary(gene.PCA)
PCA.summary
```

```
## Importance of components:
##                             PC1      PC2       PC3       PC4       PC5
## Standard deviation     880.6774 520.3121 227.79013 201.67594 193.66738
## Proportion of Variance   0.6303   0.2200   0.04217   0.03305   0.03048
## Cumulative Proportion    0.6303   0.8503   0.89243   0.92548   0.95596
##                             PC6       PC7      PC8      PC9     PC10     PC11
## Standard deviation     150.26078 102.12845 84.35153 70.53612 50.99879 34.89438
## Proportion of Variance   0.01835   0.00848  0.00578  0.00404  0.00211  0.00099
## Cumulative Proportion    0.97431   0.98279  0.98857  0.99261  0.99472  0.99571
##                            PC12     PC13    PC14     PC15     PC16     PC17
## Standard deviation     32.69040 29.91605 27.2092 23.79487 21.48151 18.74291
## Proportion of Variance  0.00087  0.00073  0.0006  0.00046  0.00037  0.00029
## Cumulative Proportion   0.99658  0.99731  0.9979  0.99837  0.99875  0.99903
##                            PC18     PC19     PC20     PC21    PC22    PC23
## Standard deviation     17.61497 16.26246 14.86637 13.85835 11.2509 8.76728
## Proportion of Variance  0.00025  0.00021  0.00018  0.00016  0.0001 0.00006
## Cumulative Proportion   0.99928  0.99950  0.99968  0.99983  0.9999 1.00000
##                            PC24
## Standard deviation     1.146e-13
## Proportion of Variance 0.000e+00
## Cumulative Proportion  1.000e+00
```

```
# 7
PCA.importance <- PCA.summary$importance
PCA.importance
```

```
##                             PC1       PC2       PC3       PC4       PC5
## Standard deviation     880.67743 520.31210 227.79013 201.67594 193.66738
## Proportion of Variance   0.63027   0.22000   0.04217   0.03305   0.03048
## Cumulative Proportion    0.63027   0.85027   0.89243   0.92548   0.95596
##                             PC6       PC7       PC8       PC9      PC10      PC11
## Standard deviation     150.26078 102.12845 84.35153 70.53612 50.99879 34.89438
## Proportion of Variance   0.01835   0.00848  0.00578  0.00404  0.00211  0.00099
## Cumulative Proportion    0.97431   0.98279  0.98857  0.99261  0.99472  0.99571
##                            PC12      PC13      PC14      PC15      PC16      PC17
## Standard deviation     32.69040 29.91605 27.20925 23.79487 21.48151 18.74291
## Proportion of Variance  0.00087  0.00073  0.00060  0.00046  0.00037  0.00029
## Cumulative Proportion   0.99658  0.99731  0.99791  0.99837  0.99875  0.99903
##                            PC18      PC19      PC20      PC21      PC22      PC23
## Standard deviation     17.61497 16.26246 14.86637 13.85835 11.25089 8.767283
## Proportion of Variance  0.00025  0.00021  0.00018  0.00016  0.00010 0.000060
## Cumulative Proportion   0.99928  0.99950  0.99968  0.99983  0.99994 1.000000
##                               PC24
## Standard deviation     1.146105e-13
## Proportion of Variance 0.000000e+00
## Cumulative Proportion  1.000000e+00
```

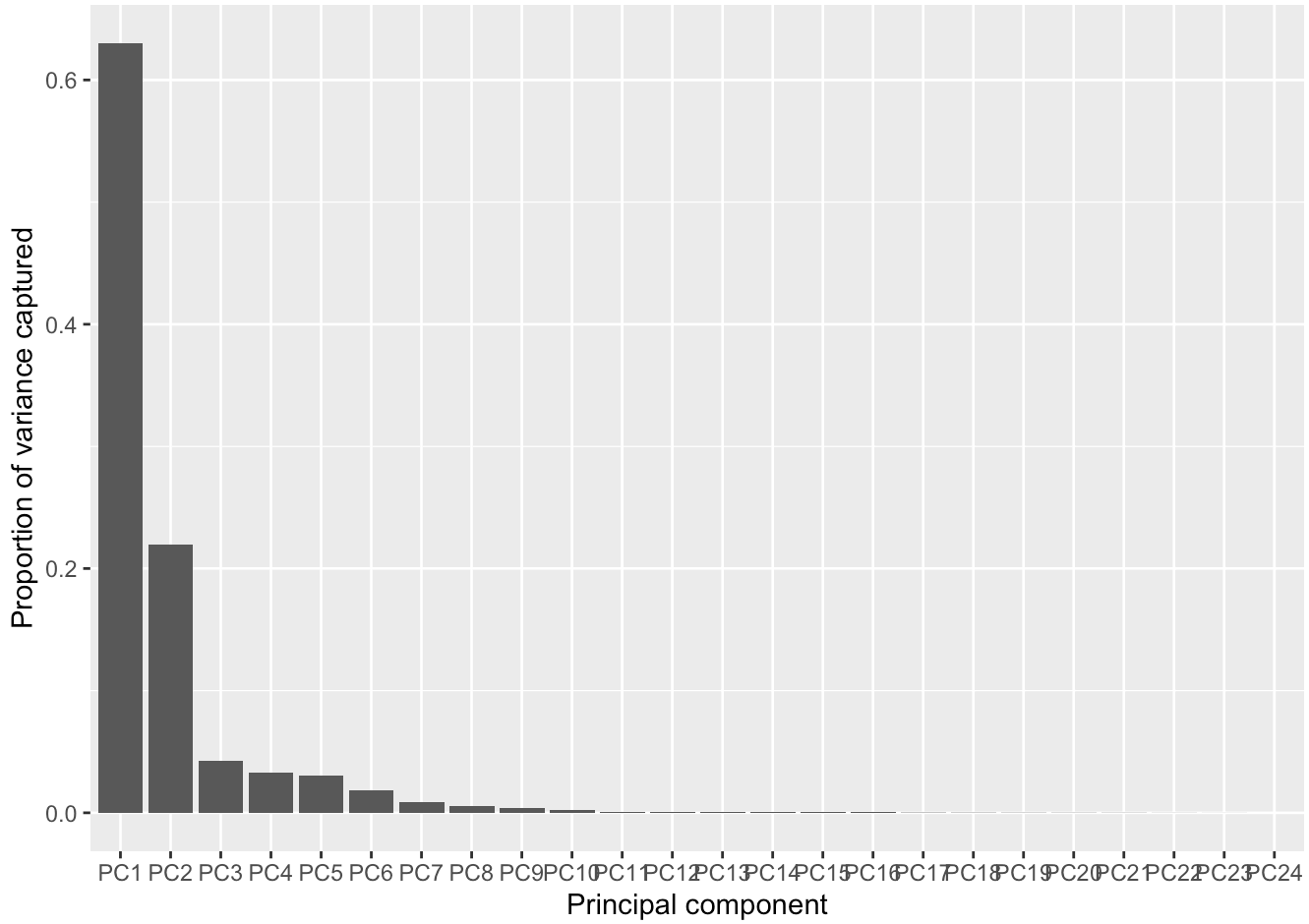
```
# 8
prop.var <- PCA.importance[2,]
prop.var
```

```
##     PC1     PC2     PC3     PC4     PC5     PC6     PC7     PC8     PC9    PC10
## 0.63027 0.22000 0.04217 0.03305 0.03048 0.01835 0.00848 0.00578 0.00404 0.00211
##    PC11    PC12    PC13    PC14    PC15    PC16    PC17    PC18    PC19    PC20
## 0.00099 0.00087 0.00073 0.00060 0.00046 0.00037 0.00029 0.00025 0.00021 0.00018
##    PC21    PC22    PC23    PC24
## 0.00016 0.00010 0.00006 0.00000
```

```
# 9
scree.df <- data.frame( prop.var )
scree.df
```

```
##      prop.var
## PC1   0.63027
## PC2   0.22000
## PC3   0.04217
## PC4   0.03305
## PC5   0.03048
## PC6   0.01835
## PC7   0.00848
## PC8   0.00578
## PC9   0.00404
## PC10  0.00211
## PC11  0.00099
## PC12  0.00087
## PC13  0.00073
## PC14  0.00060
## PC15  0.00046
## PC16  0.00037
## PC17  0.00029
## PC18  0.00025
## PC19  0.00021
## PC20  0.00018
## PC21  0.00016
## PC22  0.00010
## PC23  0.00006
## PC24  0.00000
```

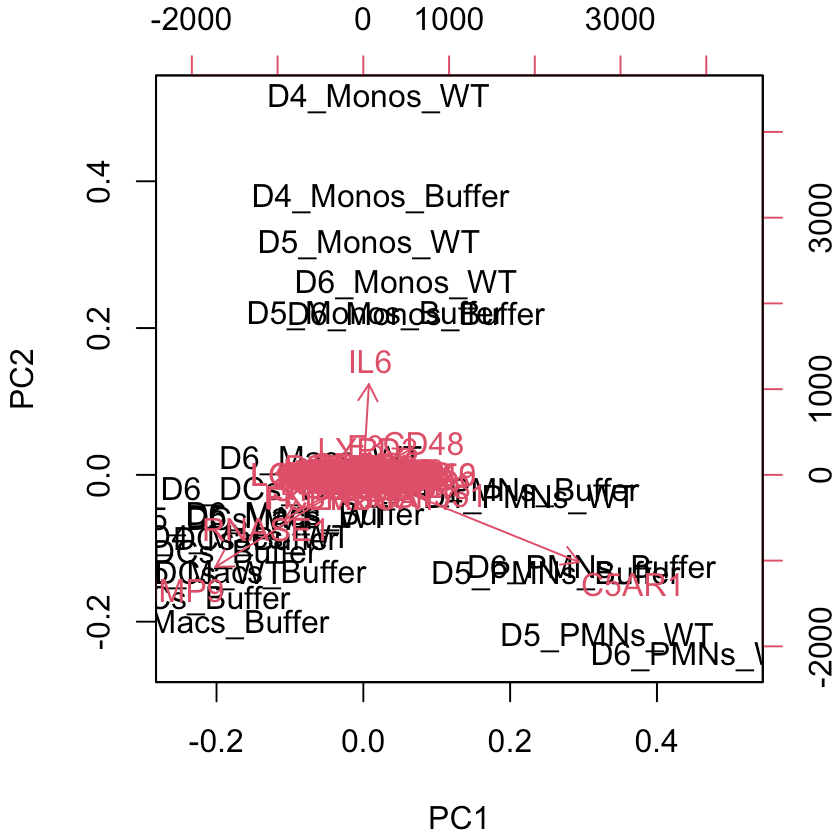
```
# 10
PCs <- factor(rownames(scree.df), levels=unique(rownames(scree.df)))
ggplot(data=scree.df, aes(x = PCs, y = prop.var)) +
  geom_col()+
  ylab("Proportion of variance captured") +
  xlab("Principal component")
```




```
# 11
biplot(gene.PCA)
```

```
## 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
```



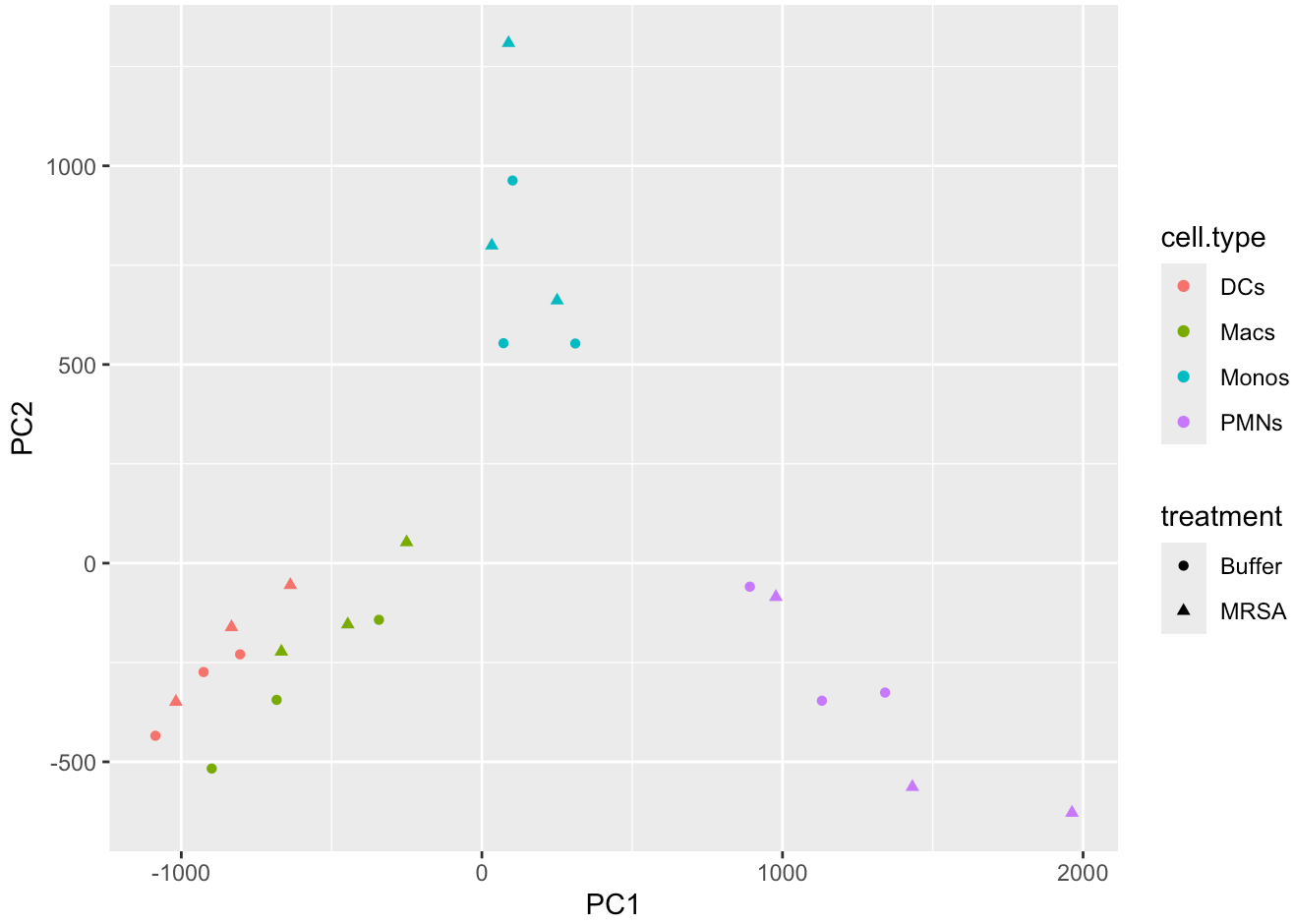

```
#12
PCA.df <- data.frame( gene.PCA$x )
PCA.df
```

```
##                       PC1        PC2          PC3         PC4         PC5
## D4_DCs_Buffer   -1086.26059 -434.34372 -168.0918007  -23.050104   76.363293
## D4_Macs_Buffer   -898.77388 -516.92406 -380.8469217  238.732577   39.881883
## D4_Monos_Buffer   101.95483  962.96582 -170.4603292  -75.480326   -4.691244
## D4_PMNs_Buffer    891.46565  -59.22977   45.4747264  -89.103375 -432.498565
## D5_DCs_Buffer    -926.00540 -274.03063  217.3227077 -371.650618   73.852051
## D5_Macs_Buffer   -682.92173 -344.11325 -167.0428060  231.559060 -114.746006
## D5_Monos_Buffer    71.74558  553.58670  290.9172466  123.187035 -143.008802
## D5_PMNs_Buffer   1131.20191 -346.26804   10.9046835 -130.003762 -225.621096
## D6_DCs_Buffer    -804.41153 -229.47746  181.0590297 -277.395019  165.734514
## D6_Macs_Buffer   -342.63183 -142.57482  -49.1611805  286.553117  -13.503552
## D6_Monos_Buffer   310.70559  552.82570  315.7150464  265.978773  182.412389
## D6_PMNs_Buffer   1341.43290 -325.74027   42.0285344   11.564080  164.992866
## D4_DCs_WT       -1018.14230 -348.94226 -132.1399337   -8.795725   22.982257
## D4_Macs_WT       -667.19622 -222.64829 -175.5614787  127.337528  -14.563627
## D4_Monos_WT        88.62317 1309.02100 -621.3922255 -306.025740  114.694912
## D4_PMNs_WT        978.01649  -85.12709   -9.7001432  -57.798207 -321.233817
## D5_DCs_WT        -833.13604 -161.04936  251.4749683 -292.860025   20.461886
## D5_Macs_WT       -446.25367 -154.14439    0.8810619  170.485080 -137.723448
## D5_Monos_WT        32.97787  799.26038   93.1387746   50.771375  -60.643269
## D5_PMNs_WT       1432.08717 -563.29412 -117.5883976 -153.150056 -119.041228
## D6_DCs_WT        -637.31264  -55.05409  268.5646535 -220.338769   83.726360
## D6_Macs_WT       -250.50647   52.53769   97.7952249  230.934077  -90.356849
## D6_Monos_WT       250.23376  661.03563  281.9142365  230.385482  204.641782
## D6_PMNs_WT       1963.10737 -628.27129 -105.2056778   38.163542  527.887309
##                       PC6        PC7          PC8         PC9        PC10
## D4_DCs_Buffer   -162.068628   10.99440  -30.0232081  110.703008  62.556833
## D4_Macs_Buffer  -149.707372   14.87713 -118.4201739 -108.900655 -13.475798
## D4_Monos_Buffer  -26.054489  -23.47373   48.6883665  -16.918790  35.859717
## D4_PMNs_Buffer  -203.427811  -64.77095   61.5563731  -60.462008  54.696911
## D5_DCs_Buffer    145.294802 -140.59217  -68.3337411 -108.295900   5.041853
## D5_Macs_Buffer   298.317609 -135.78896   57.8686956  115.685844  19.259934
## D5_Monos_Buffer   39.882900   97.61020 -144.8377048   35.256999 -34.191283
## D5_PMNs_Buffer   153.111995  170.32012  -71.6229702   10.548275  86.553026
## D6_DCs_Buffer   -151.231374  153.97080   92.4891284   57.905864 -30.444593
## D6_Macs_Buffer    72.075621   59.18230  121.2558591  -21.234314 -45.100518
## D6_Monos_Buffer  -79.288078  -45.72177    7.2629360   -8.800839  43.432307
## D6_PMNs_Buffer    65.900801   61.59926   36.4950582  -52.815738 113.619415
## D4_DCs_WT       -150.418036  -45.92233  -37.5112271  119.792231  52.039339
## D4_Macs_WT      -121.706001   67.78325  -85.9992905 -133.025636 -17.929902
## D4_Monos_WT       93.988001   14.31380   51.7995660  -11.718114  -5.494418
## D4_PMNs_WT      -272.772931 -202.57806   58.5634169   22.357992 -73.363070
## D5_DCs_WT        184.304429 -140.34040  -32.4586405  -60.099600  -9.678785
## D5_Macs_WT       265.261744  -26.11149   65.0408203    7.616995 -23.937829
## D5_Monos_WT       29.106734   52.86305 -160.3610593   58.010071 -58.721577
## D5_PMNs_WT        90.069960   64.50520  -73.6328523   71.240667 -70.701234
## D6_DCs_WT       -103.275907  128.51150  110.8579666   25.563024 -35.230309
## D6_Macs_WT        58.756920  119.32535  139.1246635  -83.517573 -28.496953
## D6_Monos_WT      -81.324181 -115.05314  -28.5281880   28.475830  35.225117
## D6_PMNs_WT         5.203293  -75.50338    0.7262056    2.632368 -61.518183
##                      PC11       PC12        PC13       PC14        PC15
## D4_DCs_Buffer   -12.349863 -14.9872267   0.6479082  -6.097794  29.1080196
```

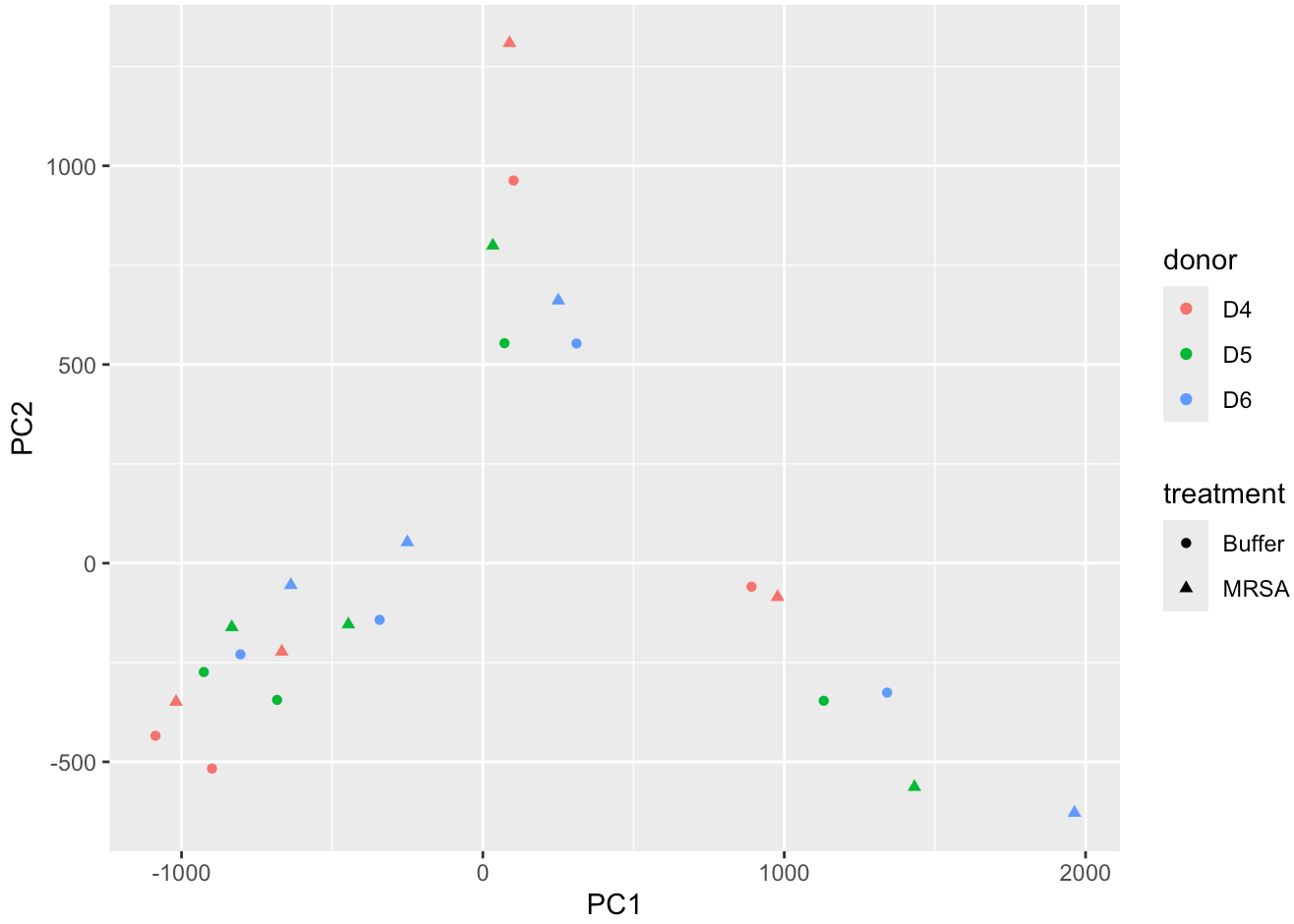
```
## D4_Macs_Buffer  -17.371754  34.2653688  25.9063928  30.018218  -2.4230377
## D4_Monos_Buffer  27.544216 -40.1875166   4.4841942  14.103019  48.2136322
## D4_PMNs_Buffer  -78.143257   0.8307577 -34.7386751  35.712532 -27.0689306
## D5_DCs_Buffer     2.965369  18.7389204  -0.8826358  -1.818678  42.5296043
## D5_Macs_Buffer  -27.019604  10.8345129  44.3206713  18.132246   0.6131858
## D5_Monos_Buffer -19.902901 -77.2955939  20.9673936  24.808682  13.3426718
## D5_PMNs_Buffer    2.310111  28.3802560 -39.1301624 -16.510451  32.2427573
## D6_DCs_Buffer   -30.186519  40.6981438  26.0205384  22.564863   9.2853386
## D6_Macs_Buffer   -2.888303  15.3496646 -34.7065990   5.465610  27.8298005
## D6_Monos_Buffer  44.103309   2.3565281   1.9095178  62.848631  -9.1770641
## D6_PMNs_Buffer    4.821081 -11.4770156  78.8544430 -32.126230 -23.7106498
## D4_DCs_WT        37.639727 -42.4759787 -47.7602161 -25.379978 -29.8674295
## D4_Macs_WT       29.080421 -15.3509399  -1.2286008 -20.031317 -13.8651145
## D4_Monos_WT      -3.068728  10.3900550  -4.7999037   3.710327 -21.7936431
## D4_PMNs_WT       29.946657   0.1403025  41.5247012 -36.775344  21.1932618
## D5_DCs_WT         8.668070 -24.7316388 -11.7544655   1.651255 -34.5232886
## D5_Macs_WT       -8.837224  -3.6276731 -13.0823739  -6.991436 -16.6809013
## D5_Monos_WT     -55.538175  28.4505718  12.4637680 -43.743638 -13.4869175
## D5_PMNs_WT       81.536737  29.6390978  -2.2937708  38.913492 -16.7490867
## D6_DCs_WT         1.392085  -8.1569410   8.5710851  -3.536407 -22.7960030
## D6_Macs_WT       23.786949 -17.2101709 -14.3064662 -35.983619   1.9099007
## D6_Monos_WT      14.627637  76.1776218 -28.3671412 -25.807565  -4.0151368
## D6_PMNs_WT      -53.116042 -40.7511060 -32.6196031  -3.126418   9.8890309
##                       PC16       PC17        PC18         PC19       PC20
## D4_DCs_Buffer    14.8881315 -17.136942   8.1040069   8.95491331 -36.994590
## D4_Macs_Buffer  -18.3672541  -4.408282 -14.0848086  21.62946350   9.234952
## D4_Monos_Buffer  25.9863659 -31.564351 -35.5819664   0.06706518  16.144401
## D4_PMNs_Buffer   24.9785086   6.619780  -4.0977343  -5.74585468  -4.044646
## D5_DCs_Buffer    24.6584239  25.683368  19.8777541  -9.33877030  -7.808993
## D5_Macs_Buffer   -8.0749589  11.543494 -12.9069604 -35.55308501   9.580811
## D5_Monos_Buffer  -1.1441016  30.965649  11.6477572  14.74475976  13.885008
## D5_PMNs_Buffer  -47.5300705 -11.508049   4.4389717  -6.53343478   9.592618
## D6_DCs_Buffer     1.3869117  -6.555884   9.1164739   1.24413017  25.107403
## D6_Macs_Buffer  -10.4178292  31.720944 -15.4966896  22.60106528 -18.446626
## D6_Monos_Buffer -28.1307256 -14.794853  22.5308176 -14.68742868 -20.269236
## D6_PMNs_Buffer   13.7848260  12.935422  -2.6090602  17.04812177  -5.043624
## D4_DCs_WT        -0.9978543  21.509119   0.6884525   1.90799869  14.847016
## D4_Macs_WT       10.2376619  -8.323051  12.5345736 -35.74050154   9.076861
## D4_Monos_WT     -16.5376095  19.489329  21.2503898   2.78200908  -5.106050
## D4_PMNs_WT      -30.4088483  -1.900497  13.6957238   6.24807921   1.098943
## D5_DCs_WT       -32.4656220 -17.866655 -31.6308891  16.77904315  -4.689603
## D5_Macs_WT       23.2850045 -41.361594  41.3084926  26.55977772  11.437460
## D5_Monos_WT       6.7108809 -16.138329 -13.4250782  -9.90866432 -21.566768
## D5_PMNs_WT       35.8828309   3.352030 -11.5652113   1.31787008  -8.291938
## D6_DCs_WT       -10.3996000  -1.029277 -11.0810509 -10.89463843  -4.369804
## D6_Macs_WT        8.4330072   4.541663  -6.2048537 -13.08790599  -7.772732
## D6_Monos_WT      17.5196920  11.038688  -7.4245751   8.93873266  19.743973
## D6_PMNs_WT       -3.2777712  -6.811718   0.9154643  -9.33274586   4.655166
##                       PC21         PC22         PC23          PC24
## D4_DCs_Buffer   -24.2930543 -10.08123641   2.12864100  4.955246e-13
## D4_Macs_Buffer   -6.4211308  19.55460403  -3.88069651 -1.214935e-14
## D4_Monos_Buffer  11.0132977   2.60867696  -3.49389213  1.076509e-13
```

```
## D4_PMNs_Buffer   -0.3885125  -0.82273607   0.97366790 -3.323626e-13
## D5_DCs_Buffer     8.1038533  12.57879223  -3.41394617  3.201490e-13
## D5_Macs_Buffer   -5.0570293  -2.11651602  -1.01477836  1.404163e-13
## D5_Monos_Buffer -14.7241470  -4.70727320  -0.28844870 -2.378440e-13
## D5_PMNs_Buffer   -2.3675790   1.12393636  -2.83253396 -1.212745e-13
## D6_DCs_Buffer    10.1050797  -6.48169437  18.02629781  2.260518e-14
## D6_Macs_Buffer   24.9335751 -17.70935886  -3.91355816  8.540477e-15
## D6_Monos_Buffer  12.5793884   6.29854458   3.24511874 -1.307461e-13
## D6_PMNs_Buffer   11.0570920  -2.77627285   0.16969309 -1.113510e-13
## D4_DCs_WT        22.6642543  14.12935878   3.34576167  5.205975e-13
## D4_Macs_WT        4.2207021 -28.39446669  -2.41958642  9.687997e-14
## D4_Monos_WT      -9.4637524  -1.65849919   1.34595114 -1.101588e-12
## D4_PMNs_WT        0.5059784  -0.98760457   0.21897722  1.987633e-13
## D5_DCs_WT        -7.8882684 -13.12819625   9.78422490  9.867454e-14
## D5_Macs_WT        5.0329652   2.08791332  -7.19506344 -1.742638e-13
## D5_Monos_WT      22.2808317   7.65963778   2.98713156  8.809040e-14
## D5_PMNs_WT       -4.4408967   0.09292492   1.74616172 -1.805075e-13
## D6_DCs_WT        -9.2075109   7.51273232 -29.35189184  2.055847e-13
## D6_Macs_WT      -22.9352320  21.16168262  17.58760173 -1.623109e-13
## D6_Monos_WT     -22.4107094  -8.62190790  -3.78181595 -5.466270e-13
## D6_PMNs_WT       -2.8991950   2.67695850   0.02698314  7.456405e-13
```

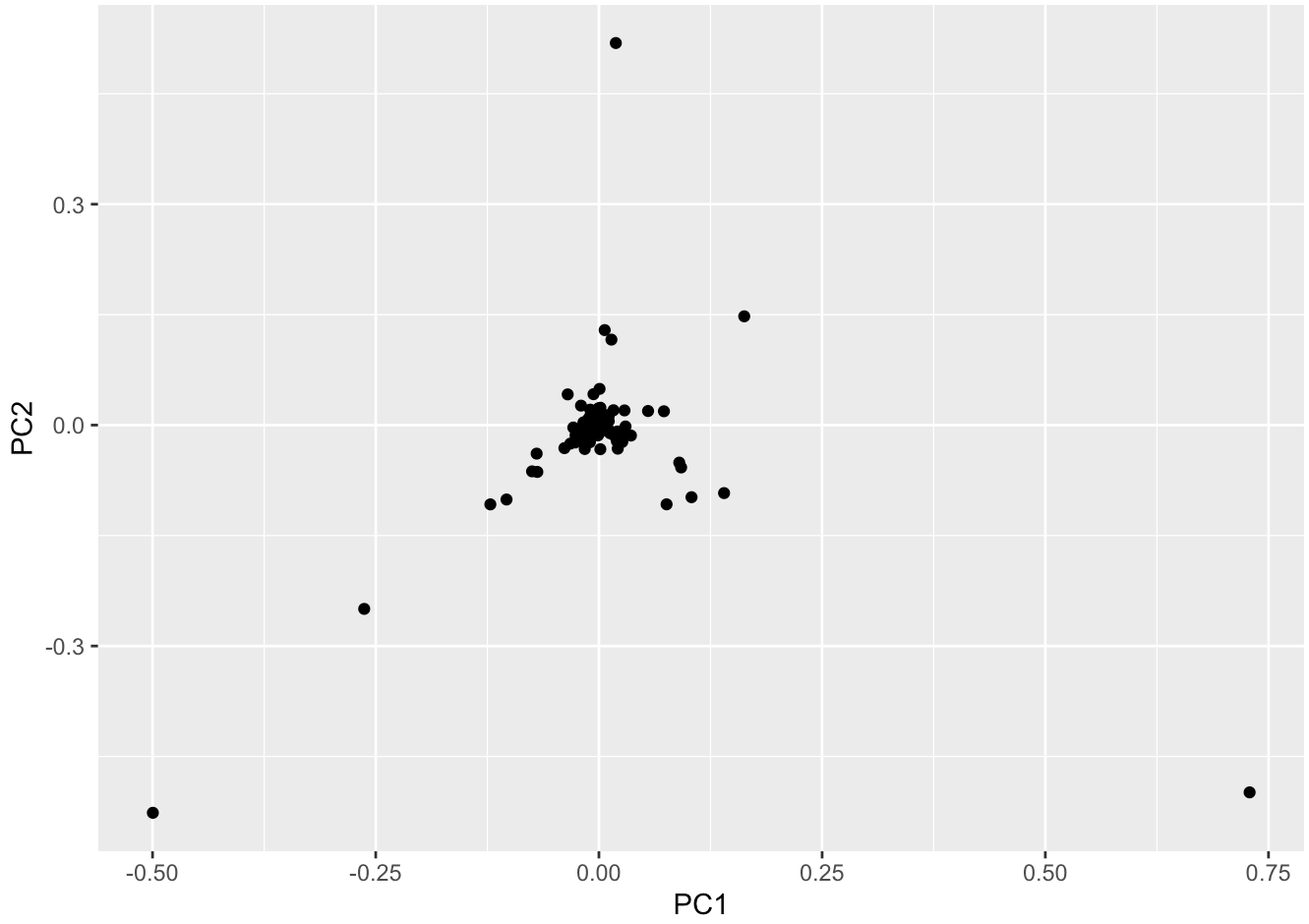
```
# 13
donor <- rep(c(rep("D4", 4), rep("D5", 4), rep("D6",4)), 2)
cell.type <- rep(c("DCs", "Macs", "Monos", "PMNs"), 6)
treatment <- c(rep("Buffer", 12), rep("MRSA", 12))

# 14
PC1 <- PCA.df$PC1
PC2 <- PCA.df$PC2
ggplot(data=PCA.df, aes(x = PC1, y = PC2, colour=cell.type, shape=treatment)) +
  geom_point()+
  ylab("PC2") +
  xlab("PC1")
```



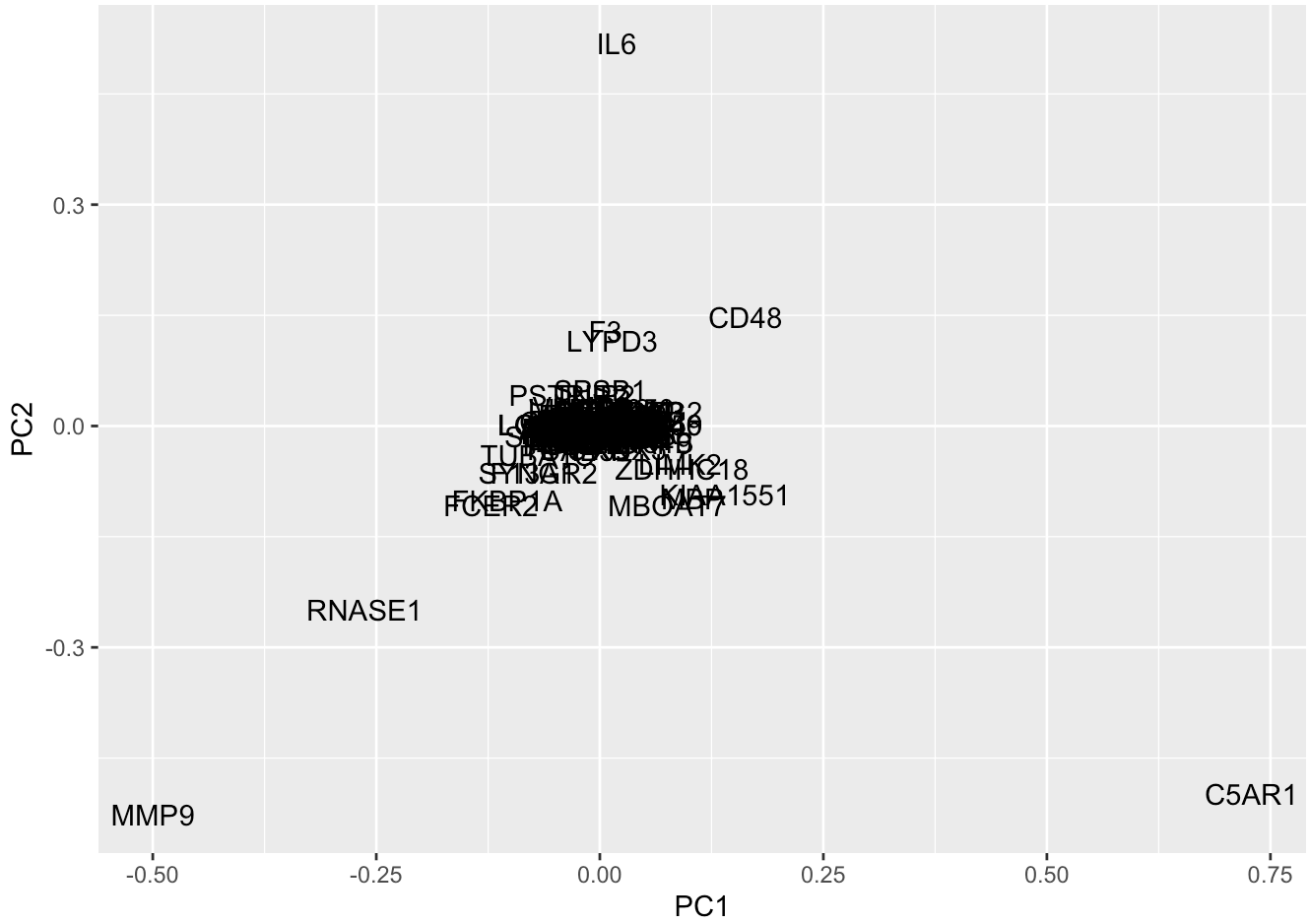

```
# 15
ggplot(data=PCA.df, aes(x = PC1, y = PC2, colour=donor, shape=treatment)) +
  geom_point()+
  ylab("PC2") +
  xlab("PC1")
```




```
# 16
rotation <- data.frame(gene.PCA$rotation)
ggplot(data=rotation, aes(x = PC1, y = PC2)) +
  geom_point()+
  ylab("PC2") +
  xlab("PC1")
```




```
ggplot(rotation, aes(x = PC1, y = PC2)) +
  geom_text( label=row.names(rotation) )+
  ylab("PC2") +
  xlab("PC1")
```



Section 4: Assessment

Q1:

PC1 capture 0.63027, PC2 capture 0.22, so first two components capture 0.63027 + 0.22 = 0.85027.

Q2:

The minimum number of components is 4 as 0.63027+0.22+0.04217+0.03305 > 0.9

Q3:

Too much overlapping, can't see the label very clear.

Q4:

Cell type is most distinctly clustered by PCA as we can see from the plot that each different cell type are most spatially grouped by their category. This indicate the there is a strong correlation between gene expression patterns and the categories represented by the cell types. It means that gene expression of the samples within each category in cell type are very similar to each other and they are distinct from gene expression in the other categories.

Q5:

C5AR1 has the most weight in PC1 and IL6 and MMP9 have the most weight in PC2 as they are very far very the other data point.

Q6:

We can see the mystery data point is cluster into the group of Monos, so it is likely to be Monos.

Q7:

In our PCA, we only use about 250 randomly selected genes, but in the figure 1B, they used 1000 genes. So their genes selected for PCA are likely to capture a broader range of variation across samples, leading to more distinct clusters. So they have more support evident to ensure their cluster and make a clear cluster and the group are more distinct.

Q8:

This paper seek answer for the question that how does Staphylococcus aureus suppress the human immune response during infection.

Q9:

These 2 additional treatment group work as control group, it can be used to compare the result with the cell infected with the MRSA to see if those cells can get a similar response. These groups allows the author to determine if the gene expression changes were specific to MRSA infection or from a general inflammatory response.

Q10:

Figure 3E provided the evident in patients.

Q11:

It is in the GEO database: gene expression data from myeloid cells and whole blood exposed to S. aureus have been deposited in GEO under accession number GSE193219. The cytokine profile studies are publicly available (accession no. GSE131990). Clinical cohort data were obtained from publicly available studies (accession nos. GSE40396 and GSE30119).