

GSE126296

Kazantseva

4/4/2023

#GSE126296 #Does male and female donors respond the same way to sprint? Show unique genes and pathways for female

```
DATA<- getGEO("GSE126296", AnnotGPL = TRUE)[[1]]

## Found 1 file(s)

## GSE126296_series_matrix.txt.gz

GSE126296 <- DATA

#head(pData(DATA))
#head(exprs(DATA))
#experimentData(DATA)

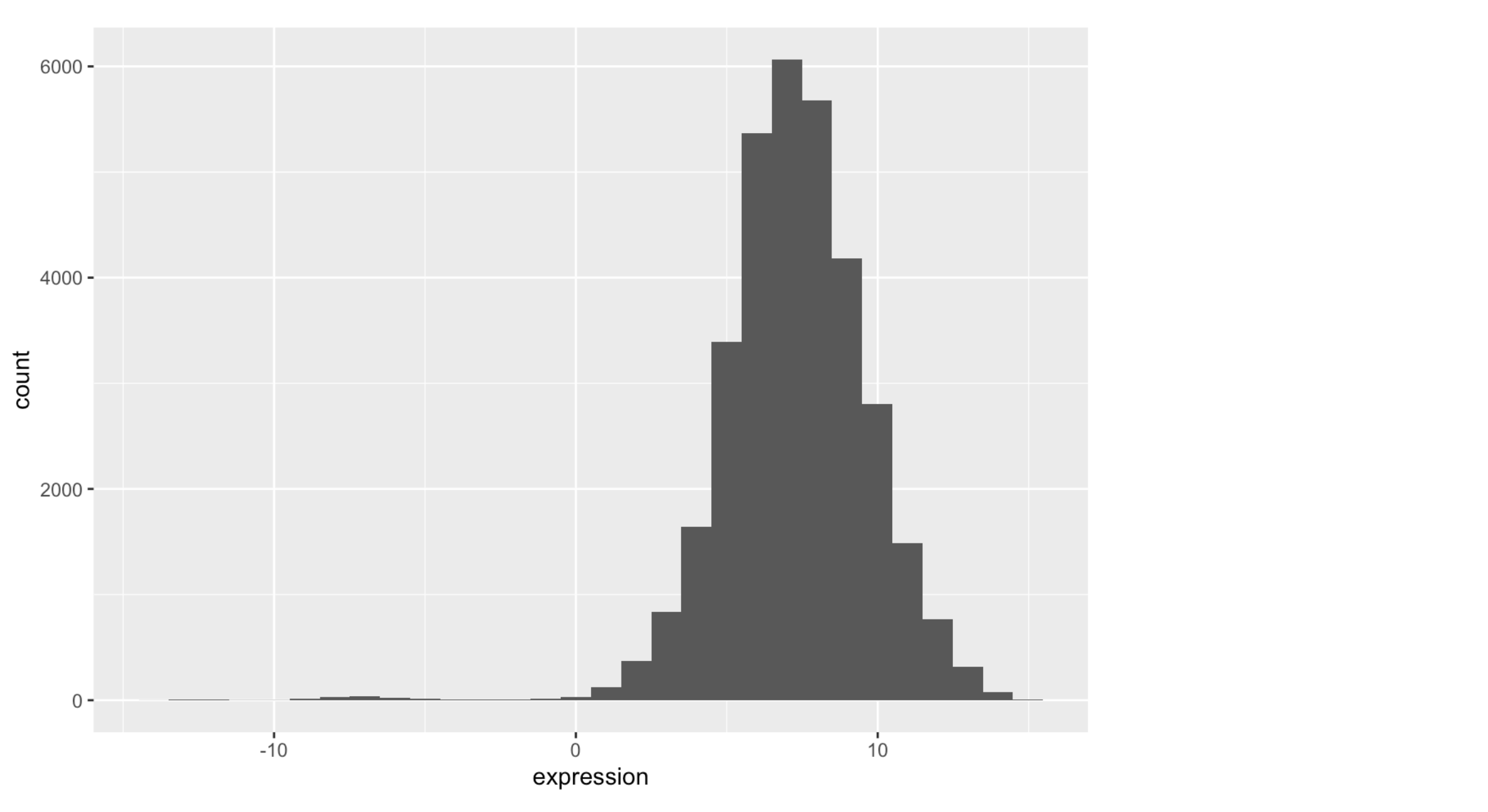
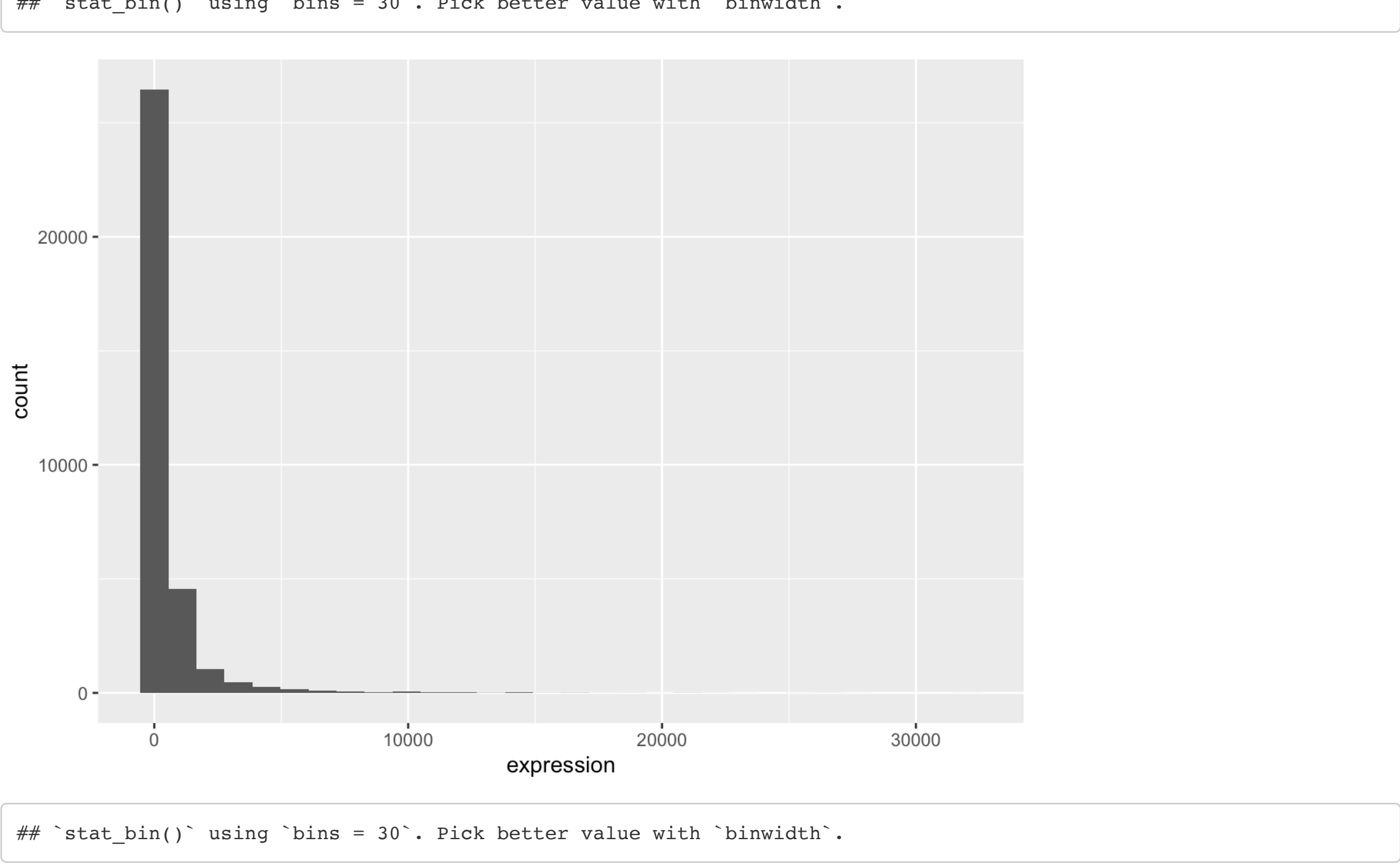
###Filtering

#head(pData(GSE126296))
#head(fData(GSE126296))

pData(GSE126296) <- pData(GSE126296)[, c("title", "gender:chl", "time:chl","individual:chl")]
colnames(pData(GSE126296)) <- c("Title", "Gender", "Time", "Sprint")
#head(pData(GSE126296))
#colnames(fData(GSE126296))

fdata(GSE126296) <- fData(GSE126296)[, c( "Gene symbol", "ID", "Gene ID")]
```

Normalization



```
exprs(GSE126296) <- normalizeBetweenArrays(exprs(GSE126296),
                                             method="quantile")
#head(exprs(GSE122121))

##Filtering

#fData(GSE126296)

GSE126296_F=GSE126296[!grepl("///", fData(GSE126296)$"Gene symbol"), ]

GSE126296_F <- GSE126296_F[fData(GSE126296_F)$"Gene symbol"!= "", ]
fdata(GSE126296_F)$mean_expression <- apply(exprs(GSE126296_F), 1, mean)
GSE126296_F <- GSE126296_F[order(fData(GSE126296_F)$mean_expression, decreasing = TRUE), ]
GSE126296_F <- GSE126296_F[!duplicated(fData(GSE126296_F)$"Gene symbol"), ]

#pData(GSE126296_F)

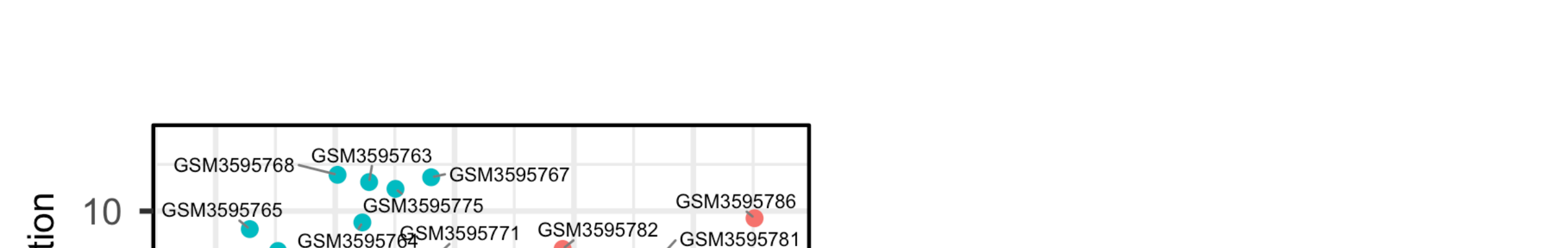
#pData(GSE122121)$Infection
GSE126296_F <- GSE126296_F[seq_len(12000), ]
dim(GSE126296_F)
```

```
## Features Samples
## 12000 28
```

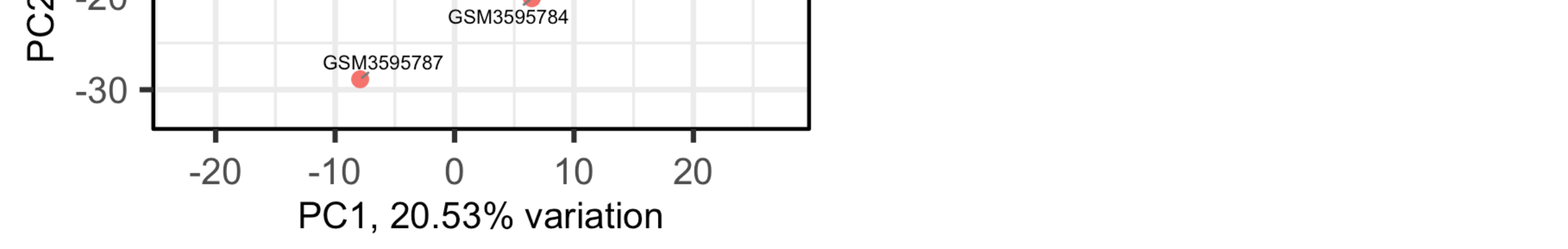
PCA

```
for (i in c(1:28)) {
  if (pData(GSE126296_F)[i,"Gender"]=="male" && pData(GSE126296_F)[i,"Time"]=="pre-exercise") {
    pData(GSE126296_F)[i,"Gender/Time"]="male/pre"
  }
  if (pData(GSE126296_F)[i,"Gender"]=="male" && pData(GSE126296_F)[i,"Time"]=="pre-exercise") {
    pData(GSE126296_F)[i,"Gender/Time"]="female/pre"
  }
  if (pData(GSE126296_F)[i,"Gender"]=="male" && pData(GSE126296_F)[i,"Time"]=="pre-exercise") {
    pData(GSE126296_F)[i,"Gender/Time"]="male/post"
  }
  if (pData(GSE126296_F)[i,"Gender"]=="male" && pData(GSE126296_F)[i,"Time"]=="pre-exercise") {
    pData(GSE126296_F)[i,"Gender/Time"]="female/post"
  }
}

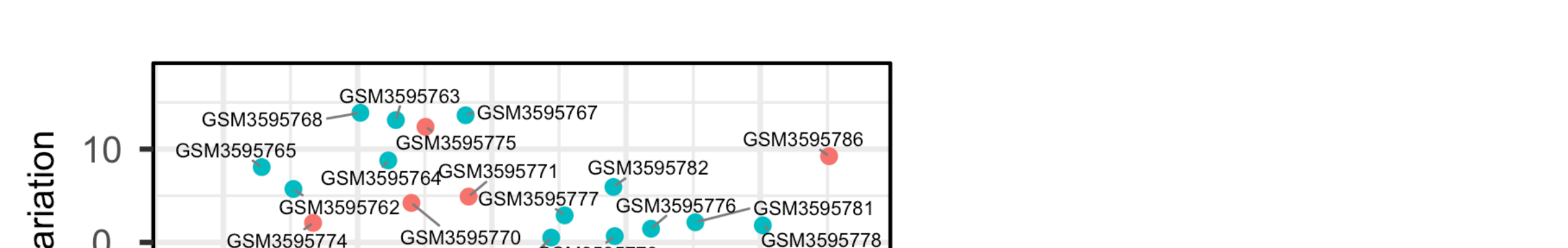
pcaData <- pca(exprs(GSE126296_F), metadata=pData(GSE126296_F))
biplot(pcaData, colby = "Time", legendPosition = "right")
```



```
biplot(pcaData, colby = "Gender", legendPosition = "right")
```



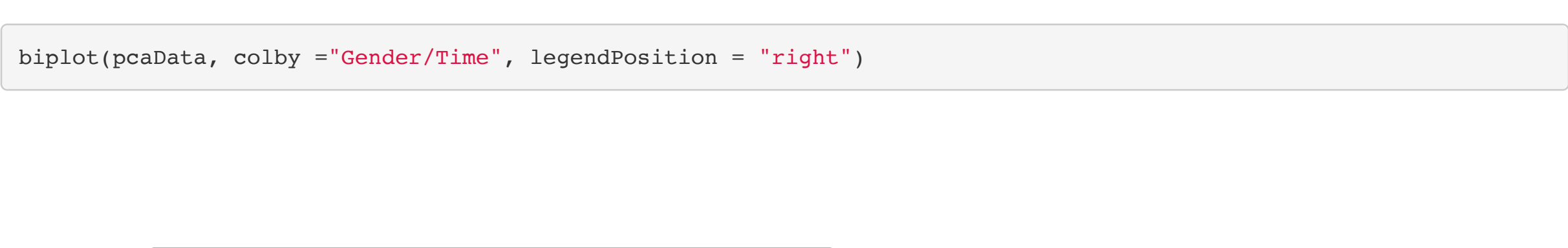
```
biplot(pcaData, colby = "Gender/Time", legendPosition = "right")
```



```
## [1] "RPS4Y1" "ZFY" "EIF1AY" "KDM6A" "ALDH1A1"
## [6] "EIF2S3" "CERS6" "TTY14" "LPL" "IRX3"
## [11] "ZFX" "LOC6446903" "CD99" "ALDH2" "CJ38"
## [16] "SNORD116-20" "VPS72" "NUP210" "SNORD116-15" "IGF2R"
## [21] "PXDC1" "SNORD116-14" "SLC22A5" "OCFRL1" "CCDC69"
## [26] "PUDP" "CXorf36" "ACSM5" "STGALNAC2" "FHL2"
```

```
##Volcano Plot

ggplot(de, aes(x=logFC, y=-log10(adj.P.Val), color=adj.P.Val < 0.05)) +
  geom_point() + theme_bw() + scale_color_manual(values=c("black", "red")) +
  geom_text_repel(data=de %>% dplyr::filter(adj.P.Val < 1e-10), aes(label=Gene.symbol, color=NULL))
```



```
if (!requireNamespace("gage", quietly = TRUE)) BiocManager::install("gage")

##

library("gage")
keggSymbol=kegg.gsets(species = "hsa", id.type = "kegg", check.new=FALSE)
test=keggSymbol$kg.sets
upRegulatedGenes <- de %>% filter(adj.P.Val < 0.05 & logFC > 0) %>% pull("Gene.symbol")
length(upRegulatedGenes)
```

```
## [1] 207

head(upRegulatedGenes)
```

```
## [1] "KDM6A" "EIF2S3" "LPL" "ZFX" "ALDH2" "NUP210"
```

```
#upRegulatedGenes
stats <- de$t

names(stats) <- de$Gene.ID

fgseaResults <- fgseaMultilevel(test, stats, minSize = 15, maxSize = 500)
head(fgseaResults, 3)
```

```
##
## 1: hsa00010 Glycolysis / Gluconeogenesis 0.0402531033 0.13859875 0.3217759
## 2: hsa00020 Citrate cycle (TCA cycle) 0.0008695638 0.01254656 0.4772708
## 3: hsa00030 Pentose phosphate pathway 0.3780760626 0.54811984 0.1013507
## ES NES size leadingEdge
## 1: 0.3797461 1.439273 46 217,3099,55902,3945,5214,125,...
## 2: 0.5667706 1.928217 28 2271,3419,50,1738,3420,3417,...
## 3: 0.3337533 1.053394 21 5214,5226,22934,5211,7086,2539,...
```

```
###Top Pathways

topPathwaysUp <- fgseaResults[ES > 0, ][head(order(pval), n=5), pathway]
topPathwaysDown <- fgseaResults[ES < 0, ][head(order(pval), n=5), pathway]
topPathways <- c(topPathwaysUp, rev(topPathwaysDown))
```

```
#Top PathwaysUP

topPathwaysUp

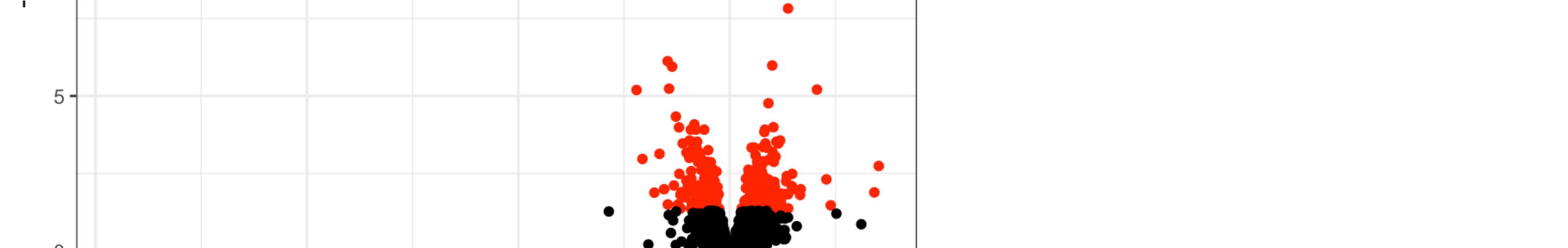
## [1] "hsa05032 Morphine addiction"
## [2] "hsa00280 Valine, leucine and isoleucine degradation"
## [3] "hsa04512 ECM-receptor interaction"
## [4] "hsa04510 Focal adhesion"
## [5] "hsa04015 Rap1 signaling pathway"
```

```
#Top PathwaysDown

topPathwaysDown

## [1] "hsa03050 Proteasome"
## [2] "hsa03010 Ribosome"
## [3] "hsa00515 Mannose type O-glycan biosynthesis"
## [4] "hsa05132 Salmonella infection"
## [5] "hsa05171 Coronavirus disease - COVID-19"
```

```
plotGseaTable(test[topPathways], stats, fgseaResults, gseaParam = 0.5)
```



Pathway	Gene ranks	NES	pval	padj
hsa05032 Morphine addiction		2.25	1.2·10 ⁻⁶	1.4·10 ⁻⁴
Valine, leucine and isoleucine degradation		2.30	2.2·10 ⁻⁶	1.4·10 ⁻⁴
hsa04512 ECM-receptor interaction		2.22	2.4·10 ⁻⁶	1.4·10 ⁻⁴
hsa04510 Focal adhesion		1.92	2.4·10 ⁻⁶	1.4·10 ⁻⁴
hsa04015 Rap1 signaling pathway		1.91	5.9·10 ⁻⁶	3.0·10 ⁻⁴
hsa05171 Coronavirus disease - COVID-19		-1.37	4.2·10 ⁻²	7.3·10 ⁻¹
hsa05132 Salmonella infection		-1.46	1.7·10 ⁻²	7.2·10 ⁻²
15 Mannose type O-glycan biosynthesis		-1.78	1.5·10 ⁻²	6.0·10 ⁻²
hsa03010 Ribosome		-1.51	1.1·10 ⁻²	6.1·10 ⁻²
hsa03050 Proteasome		-2.38	1.1·10 ⁻⁶	1.4·10 ⁻⁴