Customizing ggplot

Kim Dill-McFarland, kadm@uw.edu

version January 18, 2022

Contents

Overview

In this workshop, we explore some options to customize ggplot. This document contains the basic plot to start and then addtional plots with modifications in response to attendee questions.

Prior to the workshop

Please install R, RStudio, and the following packages.

```
#Data manipulation
#install.packages("tidyverse")
library(tidyverse)
#Example data
#install.packages("devtools")
#devtools::install_github("BIGslu/kimma")
library(kimma)
```

Load data

Briefly, these data are from RNA-sequencing of human dendritic cells cultured with and without virus. Samples are from 3 donors and a random subset of 1000 genes were selected. Expression data are in an limma EList object (named `example.voom`) containing expression (E), sample/patient metadata (`targets`), and gene metadata (`genes`). Expression is expressed as TMM-normalized log2 counts per million (CPM).

We combine the expression, sample, and gene metadata in a single table for use in plotting.

```
dat <- as.data.frame(example.voom$E) %>%
  rownames_to_column("geneName") %>%
  pivot_longer(-geneName, names_to = "libID") %>%
  inner_join(example.voom$targets, by = "libID") %>%
  inner_join(example.voom$genes, by = "geneName")

dat
```

```
## # A tibble: 12,000 x 15
##     geneName      libID value group lib.size norm.factors donorID median_cv_cover~
##     <chr>         <chr> <dbl> <fct>    <dbl>        <dbl> <chr>              <dbl>
##  1 ENSG0000000~ lib1   6.11 1       79646.        1.00  donor1             0.514
##  2 ENSG0000000~ lib2   8.00 1       88008.        0.951 donor1             0.435
##  3 ENSG0000000~ lib3   7.32 1      178020.        1.09  donor2             0.374
##  4 ENSG0000000~ lib4   7.33 1      133836.        0.943 donor2             0.388
##  5 ENSG0000000~ lib5   7.92 1      192547.        1.00  donor3             0.353
##  6 ENSG0000000~ lib6   7.99 1      175144.        0.974 donor3             0.349
##  7 ENSG0000000~ lib7   8.29 1      205377.        1.02  donor4             0.339
##  8 ENSG0000000~ lib8   8.05 1      149311.        0.995 donor4             0.350
##  9 ENSG0000000~ lib9   7.13 1      182080.        1.04  donor5             0.342
## 10 ENSG0000000~ lib10  8.58 1      181755.        0.980 donor5             0.330
## # ... with 11,990 more rows, and 7 more variables: virus <fct>, asthma <chr>,
## #   batch <dbl>, hgnc_symbol <chr>, Previous symbols <chr>,
## #   Alias symbols <chr>, gene_biotype <chr>
```

Additionally, we load table with GSEA results for these data. See the `example_gsea.R` script for the code to run GSEA.

```
gsea <- read_csv("data/example_gsea.csv", show_col_types = FALSE)
class(gsea)
```

```
## [1] "spec_tbl_df" "tbl_df"      "tbl"         "data.frame"
gsea
```

```
## # A tibble: 48 x 8
##     pathway            pval  padj    ES    NES nMoreExtreme  size leadingEdge
##     <chr>             <dbl> <dbl> <dbl>  <dbl>        <dbl> <dbl> <lgl>
##  1 HALLMARK_ADIPOGENE~ 0.0933 0.320 -0.555 -1.38           49    15 NA
##  2 HALLMARK_ALLOGRAFT~ 0.413  0.763  0.412  1.04          190    14 NA
##  3 HALLMARK_ANDROGEN_~ 0.0807 0.298  0.789  1.41           37     4 NA
##  4 HALLMARK_ANGIOGENE~ 0.137  0.411  0.885  1.31           64     2 NA
##  5 HALLMARK_APICAL_JU~ 0.124  0.396 -0.632 -1.38           67     9 NA
##  6 HALLMARK_APICAL_SU~ 0.260  0.568 -0.867 -1.16          132     1 NA
##  7 HALLMARK_APOPTOSIS  0.0521 0.243  0.750  1.48           23     6 NA
##  8 HALLMARK_BILE_ACID~ 0.255  0.568 -0.544 -1.19          139     9 NA
##  9 HALLMARK_CHOLESTER~ 0.470  0.806 -0.634 -1.04          243     3 NA
## 10 HALLMARK_COAGULATI~ 0.715  0.895 -0.397 -0.810         387     7 NA
## # ... with 38 more rows
```
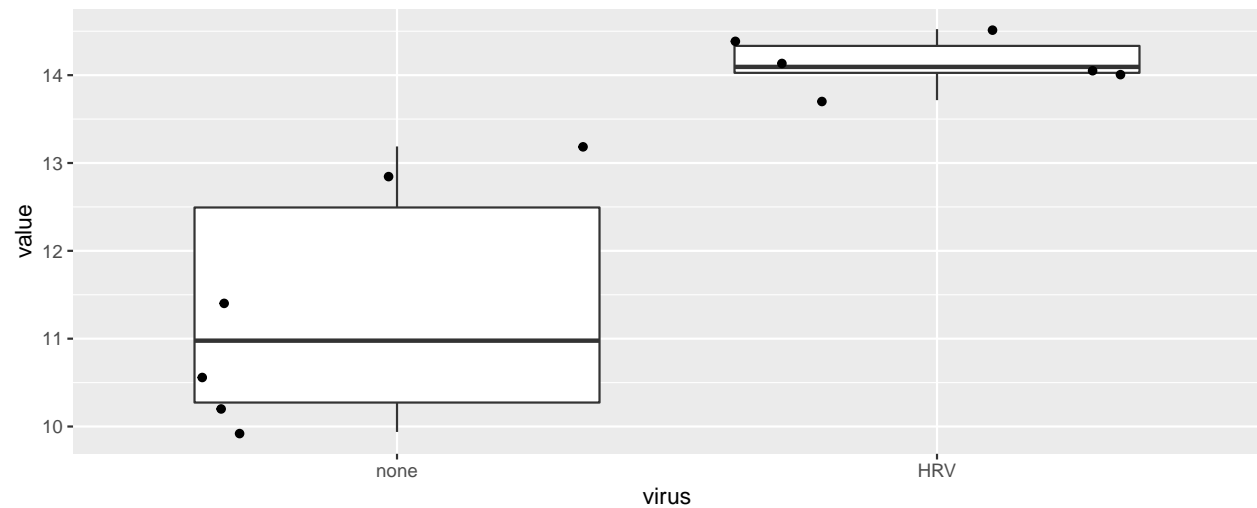
Plots

Boxplot

```
dat %>%
  filter(hgnc_symbol == "IFIT3") %>%

  ggplot(aes(x = virus, y = value)) +
  geom_boxplot(outlier.shape = NA) +
  geom_jitter()
```
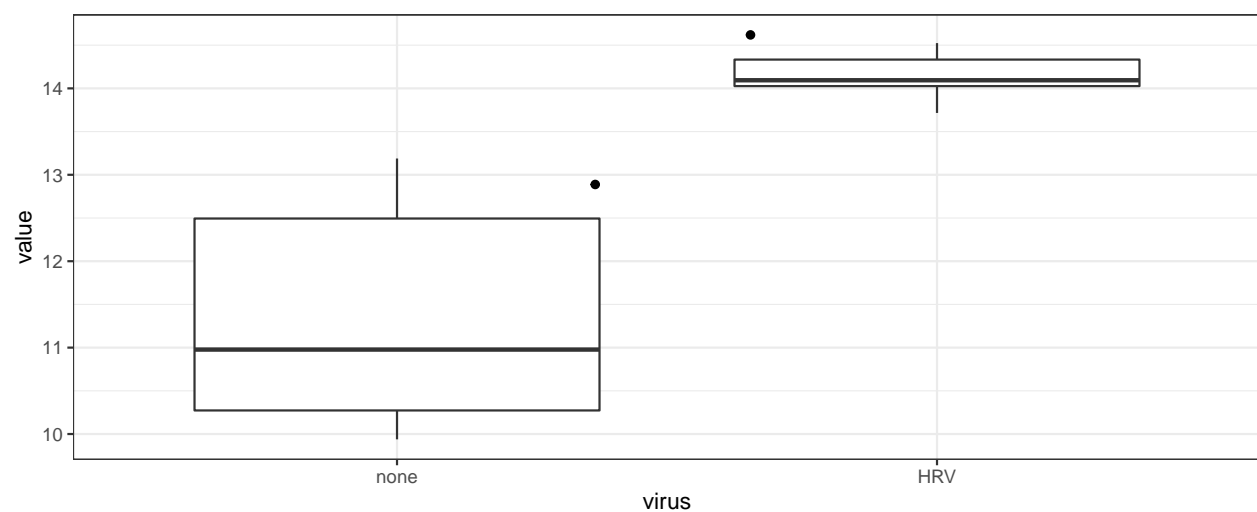

Only show some points

```
dat %>%
  filter(hgnc_symbol == "IFIT3") %>%

  ggplot(aes(x = virus, y = value)) +
  geom_boxplot(outlier.shape = NA) +
  geom_jitter(data=filter(dat,
                     hgnc_symbol == "IFIT3" & donorID=="donor1")) +
#theme_classic() +
  theme_bw()
```


See ggthemes for more

Save to file

```
plot1 <- dat %>%
  filter(hgnc_symbol == "IFIT3") %>%

  ggplot(aes(x = virus, y = value)) +
  geom_boxplot(outlier.shape = NA) +
  geom_jitter()

ggsave(filename = "plot1.png", plot1,
       width=3, height=3)
```
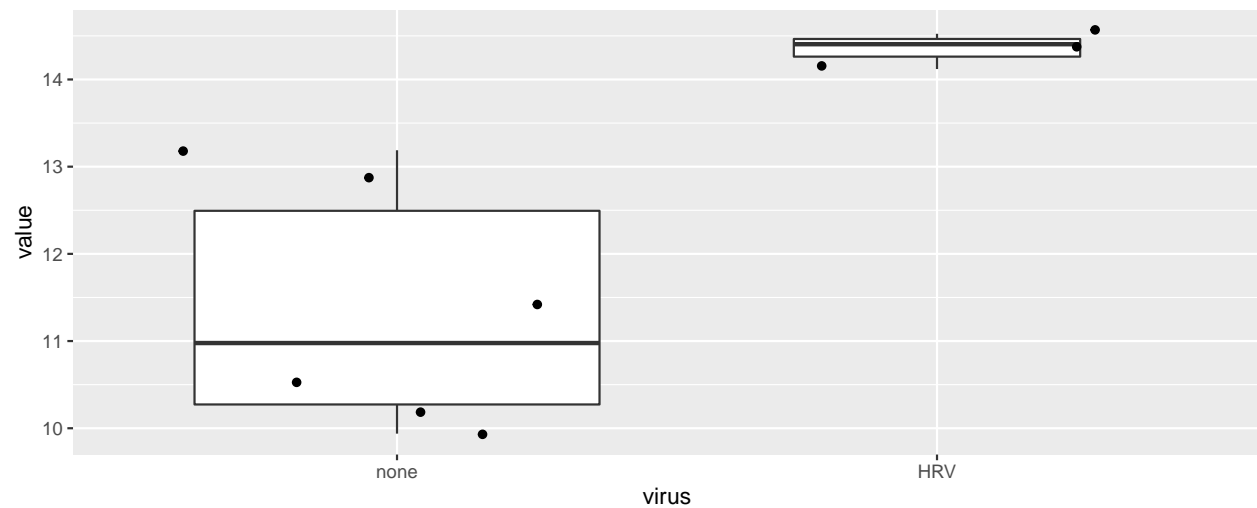
Width of boxplot relative to sample size.

```
dat %>%
  filter(hgnc_symbol == "IFIT3") %>%
  arrange(virus) %>%
  slice_head(n=9) %>%

  ggplot(aes(x = virus, y = value)) +
  geom_boxplot(varwidth = TRUE,
               outlier.shape = NA) +
  geom_jitter()
```
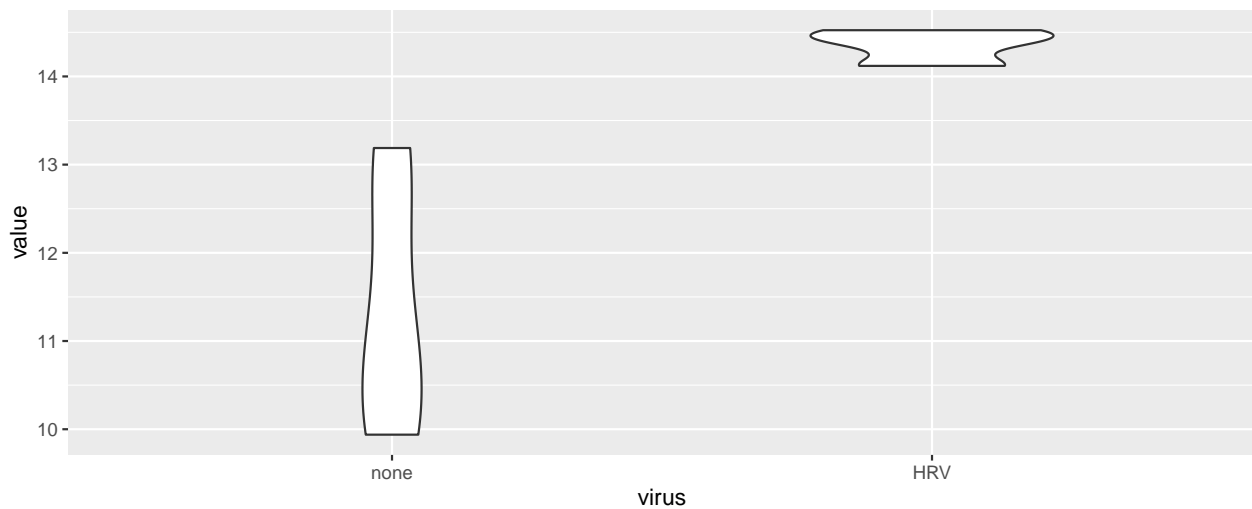

Violin version with scale to sample size

```
dat %>%
  filter(hgnc_symbol == "IFIT3") %>%
  arrange(virus) %>%
  slice_head(n=9) %>%

  ggplot(aes(x = virus, y = value)) +
  geom_violin(scale = "count")
```
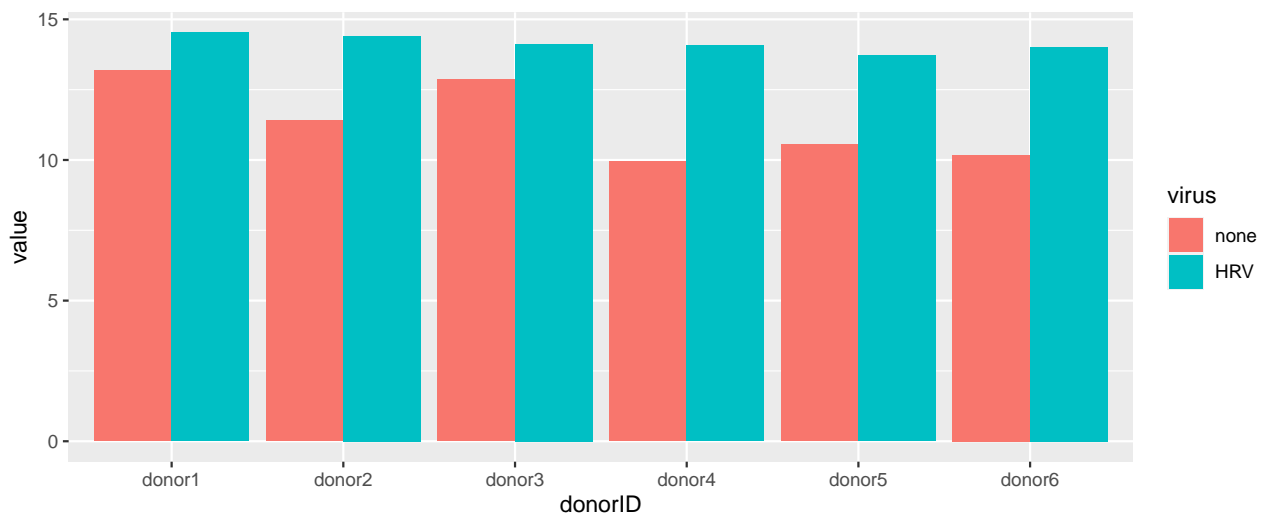
Barplot

```r
dat %>%
  filter(hgnc_symbol == "IFIT3") %>%

  ggplot(aes(x = donorID, y = value)) +
  geom_bar(aes(fill = virus), stat = "identity",
           position = 'dodge')
```
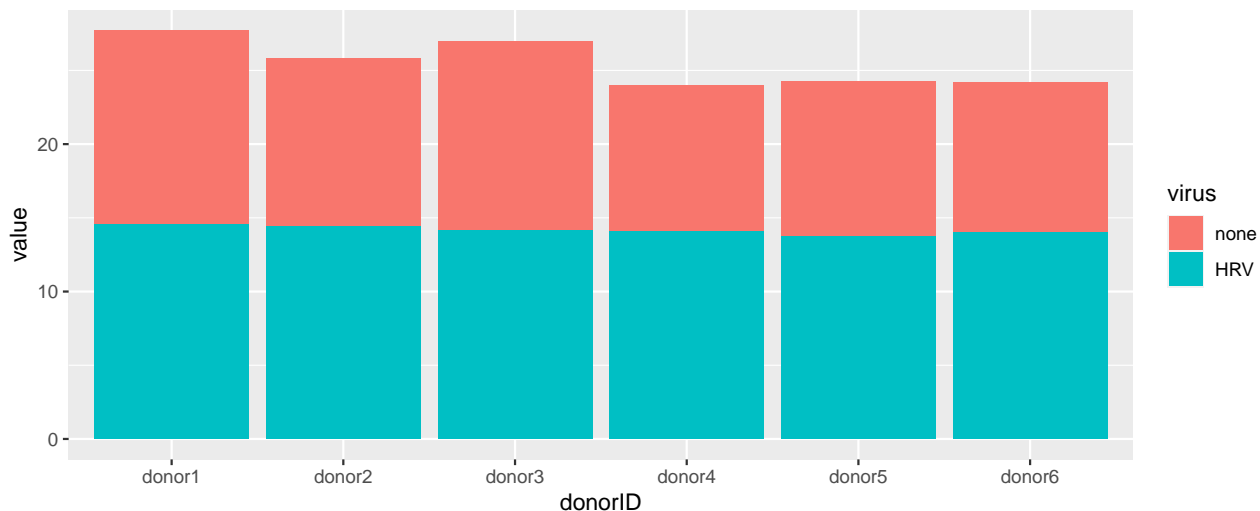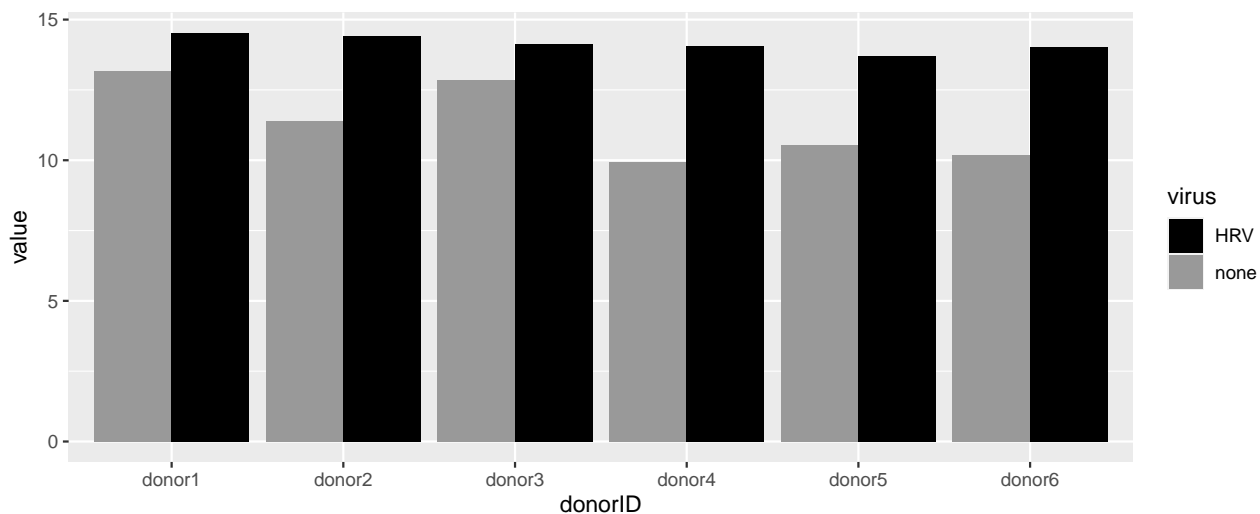

Stacked bars

```r
dat %>%
  filter(hgnc_symbol == "IFIT3") %>%

  ggplot(aes(x = donorID, y = value)) +
  geom_bar(aes(fill = virus), stat = "identity")
```

```
dat %>%
  filter(hgnc_symbol == "IFIT3") %>%

  ggplot(aes(x = donorID, y = value)) +
  geom_bar(aes(fill = virus), stat = "identity",
           position = 'dodge') +
  scale_fill_manual(values = c('HRV'='black',
                               'none'='grey60'))
```
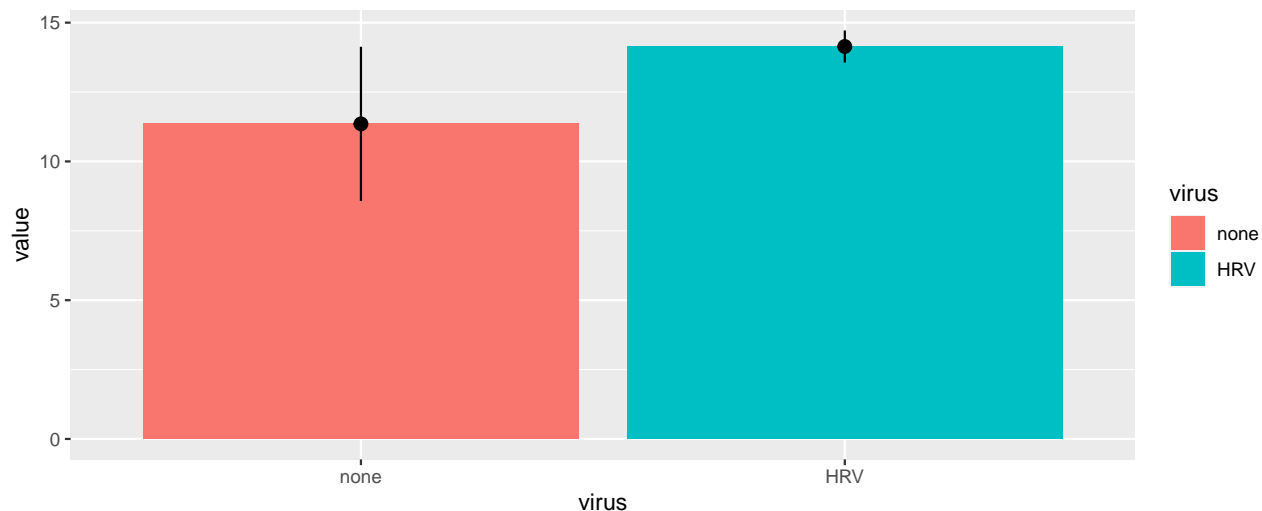

Color ideas at https://colorbrewer2.org/

Plot mean with error

```
dat %>%
  filter(hgnc_symbol == "IFIT3") %>%

  ggplot(aes(x = virus, y = value)) +
  geom_bar(aes(fill = virus),
           stat = "summary", fun = "mean") +
  stat_summary(fun.data="mean_sdl")
```
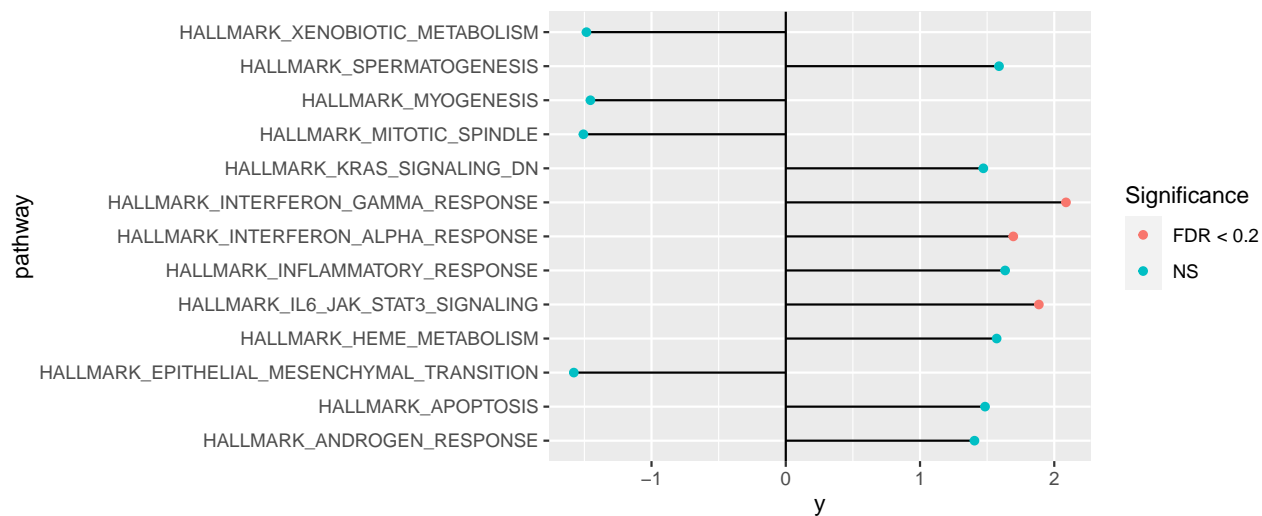
"Lollipop" plot

```
gsea %>%
  filter(padj < 0.3) %>%
  mutate(Significance = ifelse(padj < 0.2, "FDR < 0.2", "NS")) %>%

  ggplot() +
  geom_segment(aes(x=pathway, xend=pathway,
                   y=0, yend=NES)) +
  geom_point(aes(x=pathway, y=NES,
                 color = Significance)) +
  geom_hline(yintercept = 0) +
  coord_flip()
```


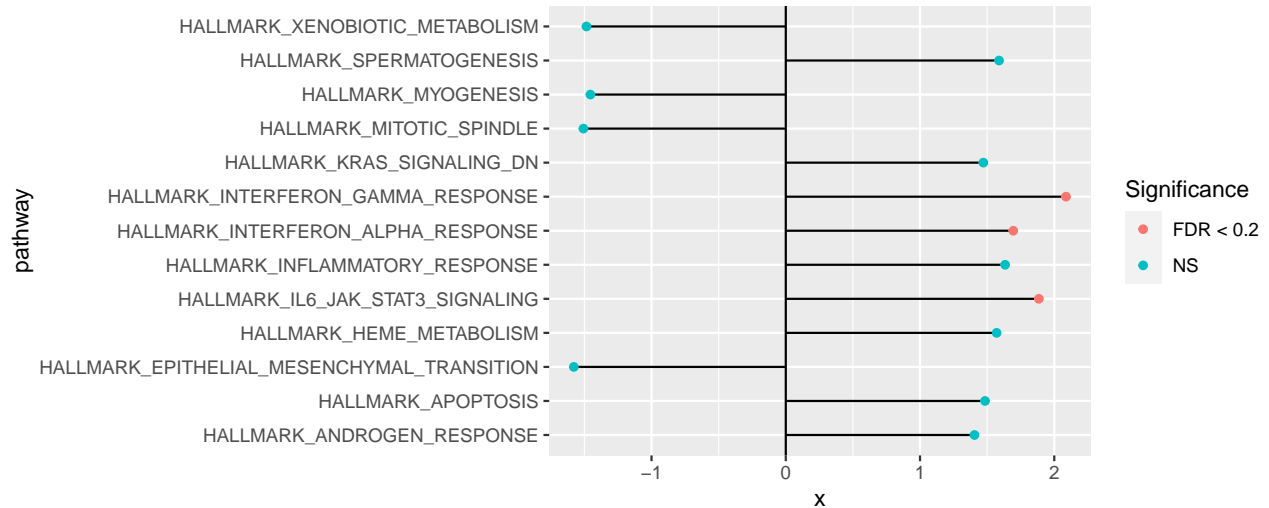Sort

```
gsea %>%
  filter(padj < 0.3) %>%
  mutate(Significance = ifelse(padj < 0.2, "FDR < 0.2", "NS")) %>%

  ggplot() +
```

```
geom_segment(aes(y=pathway, yend=pathway,
                 x=0, xend=NES)) +
geom_point(aes(y=pathway, x=NES,
               color = Significance)) +
geom_vline(xintercept = 0)
```


Reorder by NES

```
gsea %>%
  filter(padj < 0.3) %>%
  mutate(Significance = ifelse(padj < 0.2, "FDR < 0.2", "NS")) %>%

  ggplot() +
  geom_segment(aes(y=reorder(pathway,NES), yend=reorder(pathway,NES),
                   x=0, xend=NES)) +
  geom_point(aes(y=reorder(pathway,NES), x=NES,
                 color = Significance)) +
  geom_vline(xintercept = 0)
```
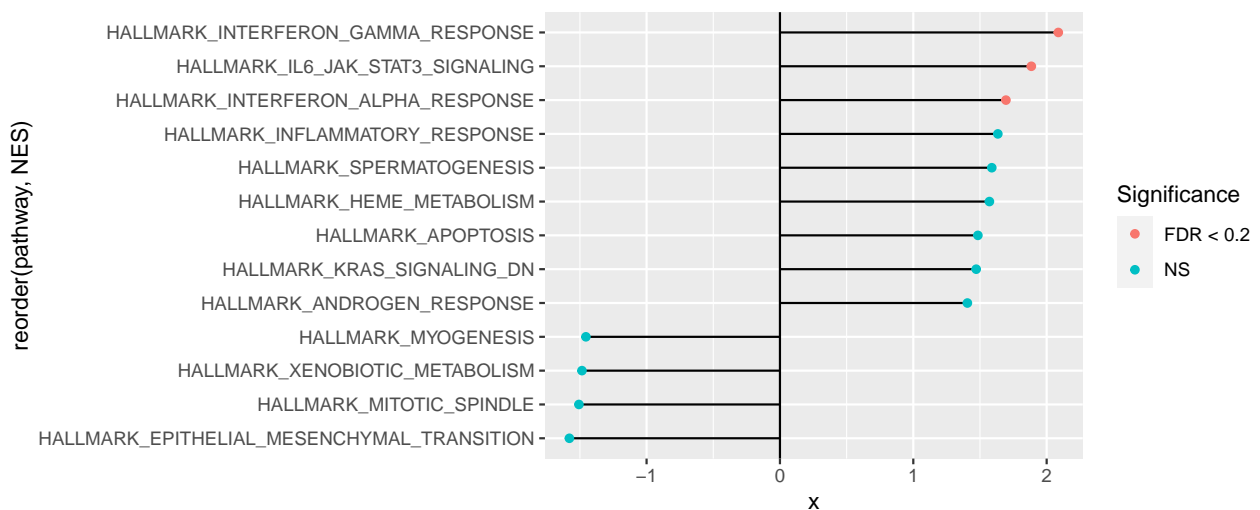


```
gsea %>%
  filter(padj < 0.3) %>%
  mutate(Significance = ifelse(padj < 0.2, "FDR < 0.2", "NS")) %>%
```

```
ggplot() +
geom_segment(aes(y=reorder(pathway,NES),
                 yend=reorder(pathway,NES),
                 x=0, xend=NES), size=1) +
geom_point(aes(y=reorder(pathway,NES), x=NES,
               color = Significance), size=2) +
geom_vline(xintercept = 0, lty="dashed") +
theme_bw() +
theme(legend.position = c(0.1,0.9),
      legend.direction = "vertical",
      plot.title = element_text(hjust = 0.5),
      legend.justification = "left") +
labs(x="Normalized enrichment score",
     y="", title="Hallmark pathways") +
lims(x=c(-2.1,2.1))
```
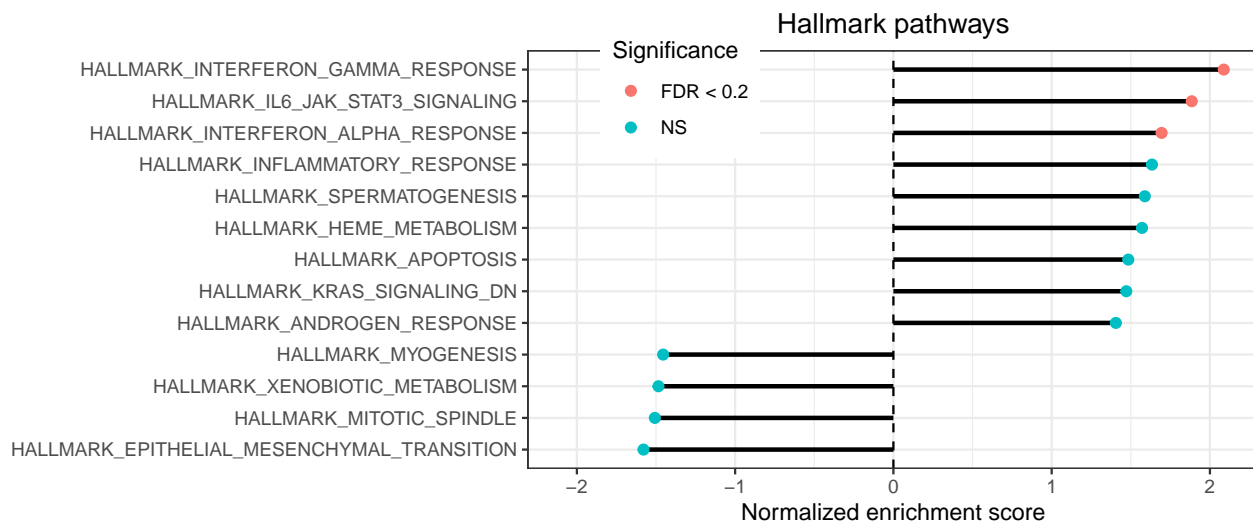

R session

```
sessionInfo()
```

```
## R version 4.1.1 (2021-08-10)
## Platform: x86_64-apple-darwin17.0 (64-bit)
## Running under: macOS Big Sur 10.16
##
## Matrix products: default
## BLAS:   /Library/Frameworks/R.framework/Versions/4.1/Resources/lib/libRblas.0.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/4.1/Resources/lib/libRlapack.dylib
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
```

```
## [1] limma_3.48.3    kimma_1.0.0    forcats_0.5.1   stringr_1.4.0
## [5] dplyr_1.0.7     purrr_0.3.4    readr_2.1.0     tidyr_1.1.4
## [9] tibble_3.1.6    ggplot2_3.3.5  tidyverse_1.3.1
##
## loaded via a namespace (and not attached):
##  [1] fs_1.5.0              lubridate_1.8.0    bit64_4.0.5
##  [4] RColorBrewer_1.1-2    httr_1.4.2         tools_4.1.1
##  [7] backports_1.3.0       utf8_1.2.2         R6_2.5.1
## [10] rpart_4.1-15          Hmisc_4.6-0        DBI_1.1.1
## [13] colorspace_2.0-2      nnet_7.3-16        withr_2.4.2
## [16] tidyselect_1.1.1      gridExtra_2.3      bit_4.0.4
## [19] compiler_4.1.1        textshaping_0.3.6  cli_3.1.0
## [22] rvest_1.0.2           htmlTable_2.3.0    xml2_1.3.2
## [25] labeling_0.4.2        checkmate_2.0.0    scales_1.1.1
## [28] systemfonts_1.0.2     digest_0.6.28      foreign_0.8-81
## [31] rmarkdown_2.11        base64enc_0.1-3    jpeg_0.1-9
## [34] pkgconfig_2.0.3       htmltools_0.5.2    dbplyr_2.1.1
## [37] fastmap_1.1.0         highr_0.9          htmlwidgets_1.5.4
## [40] rlang_0.4.12          readxl_1.3.1       rstudioapi_0.13
## [43] farver_2.1.0          generics_0.1.1     jsonlite_1.7.2
## [46] vroom_1.5.6           magrittr_2.0.1     Formula_1.2-4
## [49] Matrix_1.3-4          Rcpp_1.0.7         munsell_0.5.0
## [52] fansi_0.5.0           lifecycle_1.0.1    stringi_1.7.5
## [55] yaml_2.2.1            grid_4.1.1         parallel_4.1.1
## [58] crayon_1.4.2          lattice_0.20-45    haven_2.4.3
## [61] splines_4.1.1         hms_1.1.1          knitr_1.36
## [64] pillar_1.6.4          codetools_0.2-18   reprex_2.0.1
## [67] glue_1.5.0            evaluate_0.14      latticeExtra_0.6-29
## [70] data.table_1.14.2     modelr_0.1.8       vctrs_0.3.8
## [73] png_0.1-7             tzdb_0.2.0         foreach_1.5.1
## [76] cellranger_1.1.0      gtable_0.3.0       assertthat_0.2.1
## [79] xfun_0.28             broom_0.7.10       ragg_1.1.3
## [82] survival_3.2-13       iterators_1.0.13   cluster_2.1.2
## [85] ellipsis_0.3.2
```