main

August 13, 2021

1 Boxplot of top five up- and downregulated genes

```
[1]: suppressMessages({library(ggpubr)
                       library(tidyverse)
                       library(data.table)})
```

```
[2]: save_ggplots <- function(p, fn, w=7, h=7){
         for(ext in c('.svg', '.png', '.pdf')){
             ggsave(p, filename=paste0(fn, ext), width=w, height=h)
         }
     }


     add_symnum <- function(res){
         symnum.args <- list(cutpoints = c(0, 0.0001, 0.001, 0.01, 0.05,  1),
                             symbols = c("****", "***", "**", "*",  "ns"))
         symnum.args$x <- res$adj.P.Val
         pvalue.signif <- do.call(stats::symnum, symnum.args) %>%
             as.character()
         pvalue.format <- format.pval(res$adj.P.Val, digits = 2)
         res <- res %>%
             dplyr::ungroup() %>%
             mutate(FDR = pvalue.format, p.signif = pvalue.signif)
         return(res)
     }
```

1.1 Select most significant DEGs, up- and downregulated

```
[3]: genes = fread('../../metrics_summary/_m/male_specific_DE_genes.txt') %>%
         filter(Type == 'gene') %>% add_symnum()
     up_genes = genes %>% filter(t > 0) %>% mutate('Direction'='Upregulated') %>%
     →head(5)
     down_genes = genes %>% filter(t < 0) %>% mutate('Direction'='Downregulated')
     →%>% head(5)
     sig_genes = bind_rows(up_genes, down_genes) %>% mutate_at(vars(Direction), as.
     →factor)
```

```
for(xx in seq_along(sig_genes$Symbol)){
    sig_genes$New_ID[xx] <- ifelse(sig_genes$Symbol[xx] == '',
                                    as.character(sig_genes$ensemblID[xx]),
                                    as.character(sig_genes$Symbol[xx]))
}
sig_genes
```

	Feature <chr>	gencodeID <chr>	Symbol <chr>	ensemblID <chr>	Chrom <chr>
A data.table: 10×15	ENSG00000119411.10	ENSG00000119411.10	BSPRY	ENSG00000119411	chr9
	ENSG00000159871.14	ENSG00000159871.14	LYPD5	ENSG00000159871	chr19
	ENSG00000163833.7	ENSG00000163833.7	FBXO40	ENSG00000163833	chr3
	ENSG00000158457.5	ENSG00000158457.5	TSPAN33	ENSG00000158457	chr7
	ENSG00000146360.8	ENSG00000146360.8	GPR6	ENSG00000146360	chr6
	ENSG00000231752.5	ENSG00000231752.5	EMBP1	ENSG00000231752	chr1
	ENSG00000164649.19	ENSG00000164649.19	CDCA7L	ENSG00000164649	chr7
	ENSG00000163347.5	ENSG00000163347.5	CLDN1	ENSG00000163347	chr3
	ENSG00000128606.12	ENSG00000128606.12	LRRC17	ENSG00000128606	chr7
	ENSG00000187416.11	ENSG00000187416.11	LHFPL3	ENSG00000187416	chr7

1.2 Load phenotypes

```
[4]: pheno_file = paste0('/ceph/users/jbenja13/projects/sex_sz_ria/',
                         'input/commonMind/phenotypes/combine_files/',
                         '_m/CMC_phenotypes_all.csv')
pheno = fread(pheno_file) %>%
    filter(Sex %in% c("XX", "XY"),
           Dx %in% c("Control", "SCZ")) %>%
    mutate_if(is.character, as.factor) %>%
    select(c(Individual_ID, "RNAseq:Sample_RNA_ID", Dx, Sex))
colnames(pheno)[2] <- "SampleID"
levels(pheno$Sex) <- c("Female", "Male")
levels(pheno$Dx) <- c("CTL", "SZ")
pheno %>% head(2)
```

	Individual_ID <fct>	SampleID <fct>	Dx <fct>	Sex <fct>
A data.table: 2×4	CMC_HBCC_003	CMC_HBCC_RNA_PFC_3158	CTL	Male
	CMC_HBCC_005	CMC_HBCC_RNA_PFC_3152	CTL	Male

1.3 Load residualized expression

```
[5]: res_file = paste0('/ceph/users/jbenja13/projects/sex_sz_ria/analysis/',
                       'interaction_sex_sz/cmc_dlpfc/_m/genes/residualized_expression.
    ↪tsv')
resdf0 = fread(res_file) %>%
    filter(V1 %in% sig_genes$gencodeID) %>%
    column_to_rownames(var="V1") %>% t %>%
```

```
    as.data.frame %>% rownames_to_column() %>%
    rename(SampleID=rowname) %>%
    mutate_at("SampleID", as.factor)
resdf0[1:2, 1:5]
```

Warning message in fread(res_file):
"Detected 858 column names but the data has 859 columns (i.e. invalid file).
Added 1 extra default column name for the first column which is guessed to be
row names or an index. Use setnames() afterwards if this guess is not correct,
or fix the file write command that created the file to create a valid file."

A data.frame: 2 × 5

	SampleID <fct>	ENSG00000119411.10 <dbl>	ENSG00000128606.12 <dbl>	ENSG0000014... <dbl>
1	MSSM_RNA_PFC_1	-1.2509392	-0.1458972	-0.5281311
2	MSSM_RNA_PFC_2	0.6018737	-1.0077370	0.9476006

1.3.1 Merge and melt dataframe

```
[6]: df <- inner_join(pheno, resdf0, by="SampleID") %>%
    select(c('SampleID', 'Sex', 'Dx', starts_with('ENSG'))) %>%
    pivot_longer(-c(SampleID, Sex, Dx),
                 names_to='gencodeID', values_to="Res") %>%
    inner_join(sig_genes, by='gencodeID') %>%
    mutate_at(vars("New_ID", "Symbol", "gencodeID"), as.factor)
df$New_ID = with(df, reorder(New_ID, adj.P.Val, median))
df$Dx <- fct_drop(df$Dx)
dim(df)
df[1:2, 1:10]
```

1. 8580 2. 19

A tibble: 2 × 10

SampleID <fct>	Sex <fct>	Dx <fct>	gencodeID <fct>	Res <dbl>	Featu... <chr>
CMC_HBCC_RNA_PFC_3158	Male	CTL	ENSG00000119411.10	-0.4446609	ENS...
CMC_HBCC_RNA_PFC_3158	Male	CTL	ENSG00000128606.12	-0.8120611	ENS...

1.3.2 Initial ggplot with ggpubr

```
[7]: tmp = sig_genes %>%
    mutate(group1='Female', group2='Male', y_pos=4.8) %>%
    filter(Direction == 'Upregulated') %>%
    mutate_if(is.character, as.factor)

bxp_up <- df %>% filter(Direction=="Upregulated") %>%
        ggboxplot(x="Sex", y="Res", facet.by=c("Direction", "New_ID"),
                  add='jitter', color="Dx", xlab='', palette="npg",
                  panel.labs.font=list(face='bold'),  legend="",
                  ylab='Residualized Expression', ylim=c(-5, 5),
                  add.params=list(alpha=0.4), outlier.shape=NA,
```
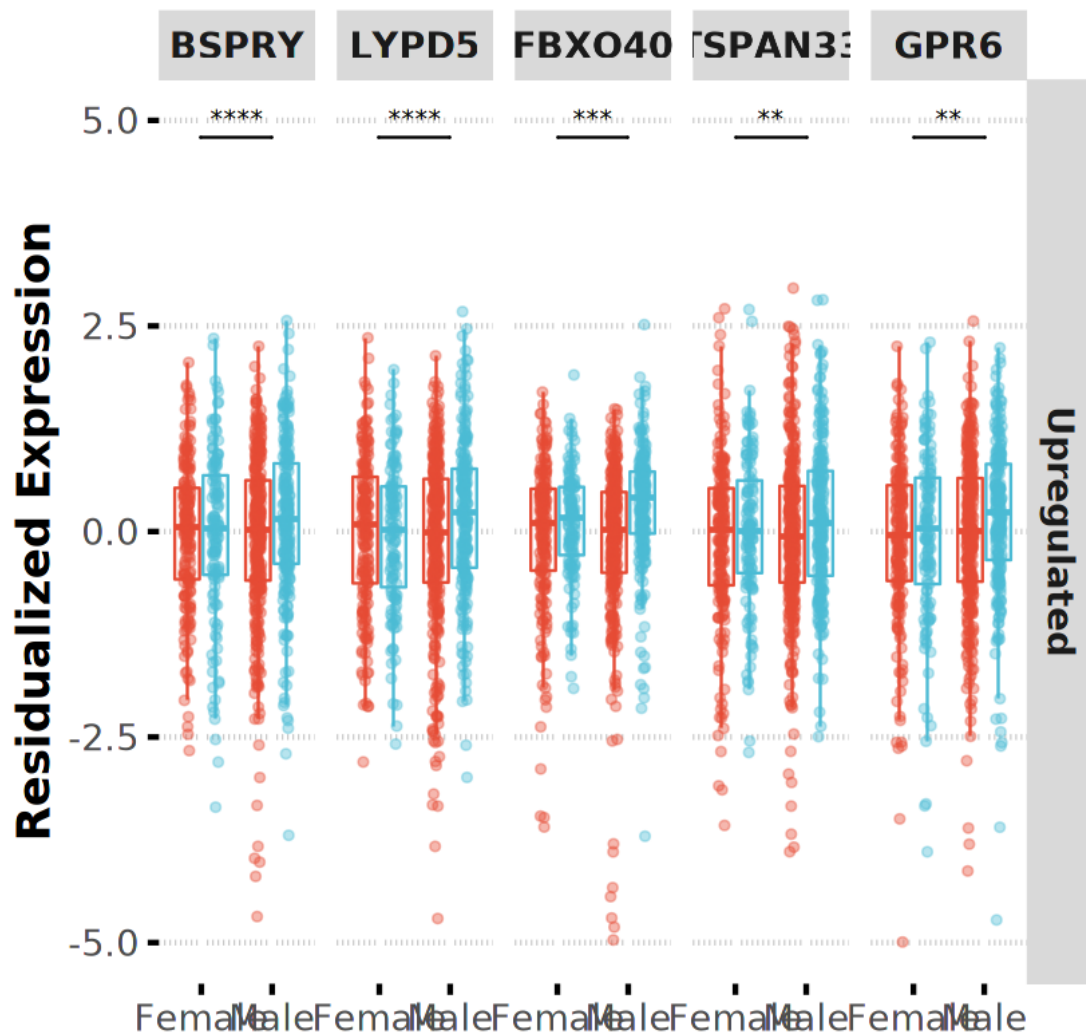
```
                ggtheme=theme_pubclean(base_size=20)) +
        geom_signif(data=tmp, tip_length = 0,
                    aes(xmin=group1, xmax=group2, annotations=p.signif,␣
↪y_position = y_pos),
                    manual=TRUE) +
        font("xy.title", face="bold") + font("legend.title", face="bold")

bxp_up
```

Warning message:
"Ignoring unknown aesthetics: xmin, xmax, annotations, y_position"
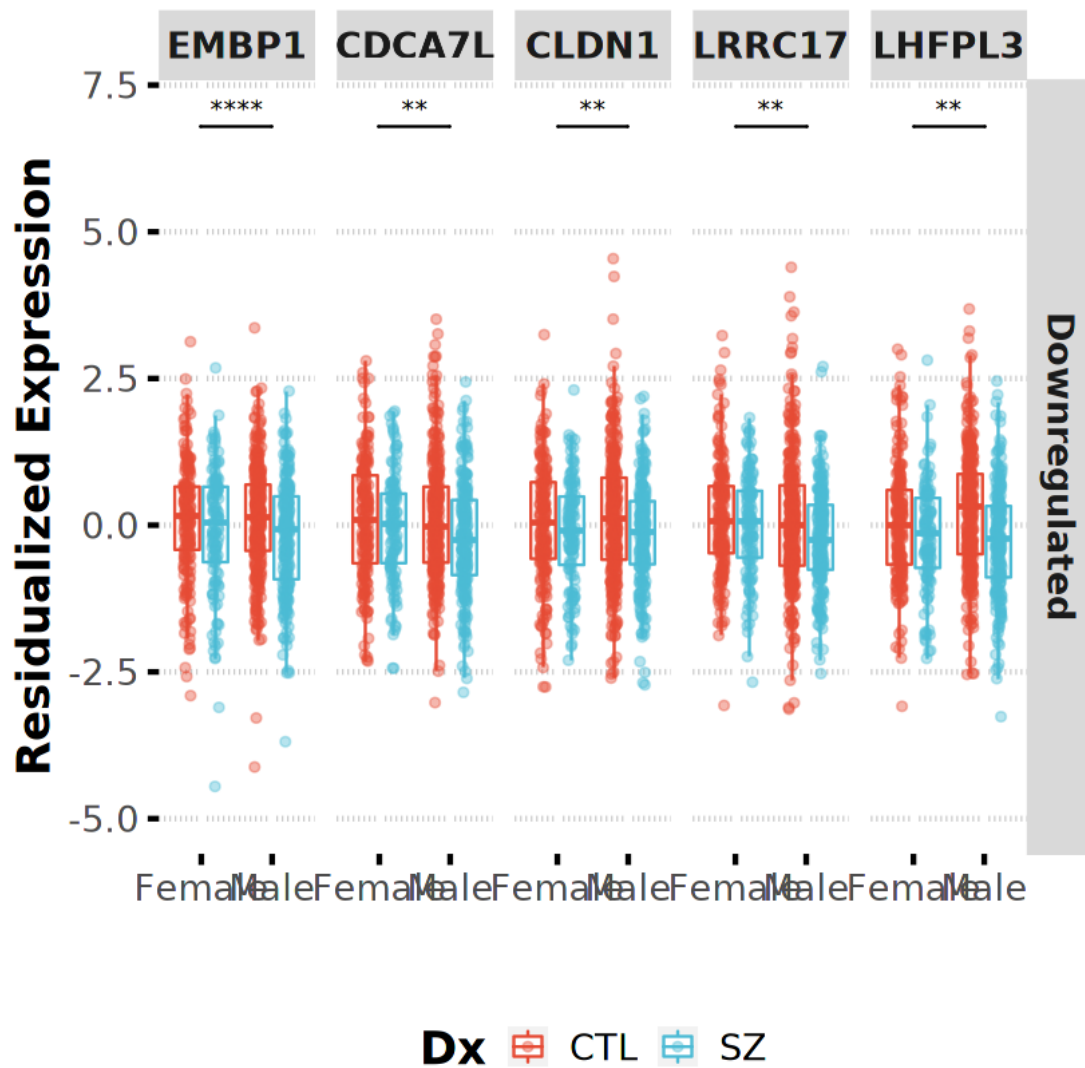
```
[8]: tmp = sig_genes %>%
        mutate(group1='Female', group2='Male', y_pos=6.8) %>%
        filter(Direction == 'Downregulated') %>%
        mutate_if(is.character, as.factor)

     bxp_down <- df %>% filter(Direction=="Downregulated") %>%
             ggboxplot(x="Sex", y="Res", facet.by=c("Direction", "New_ID"),
                     add='jitter', color="Dx", xlab='', palette="npg",
                     panel.labs.font=list(face='bold'), legend="bottom",
                     ylab='Residualized Expression', ylim=c(-5, 7),
                     add.params=list(alpha=0.4), outlier.shape=NA,
                     ggtheme=theme_pubclean(base_size=20)) +
             geom_signif(data=tmp, tip_length = 0,
                         aes(xmin=group1, xmax=group2, annotations=p.signif,␣
     ↪y_position = y_pos),
                         manual=TRUE) +
             font("xy.title", face="bold") + font("legend.title", face="bold")

     bxp_down
```
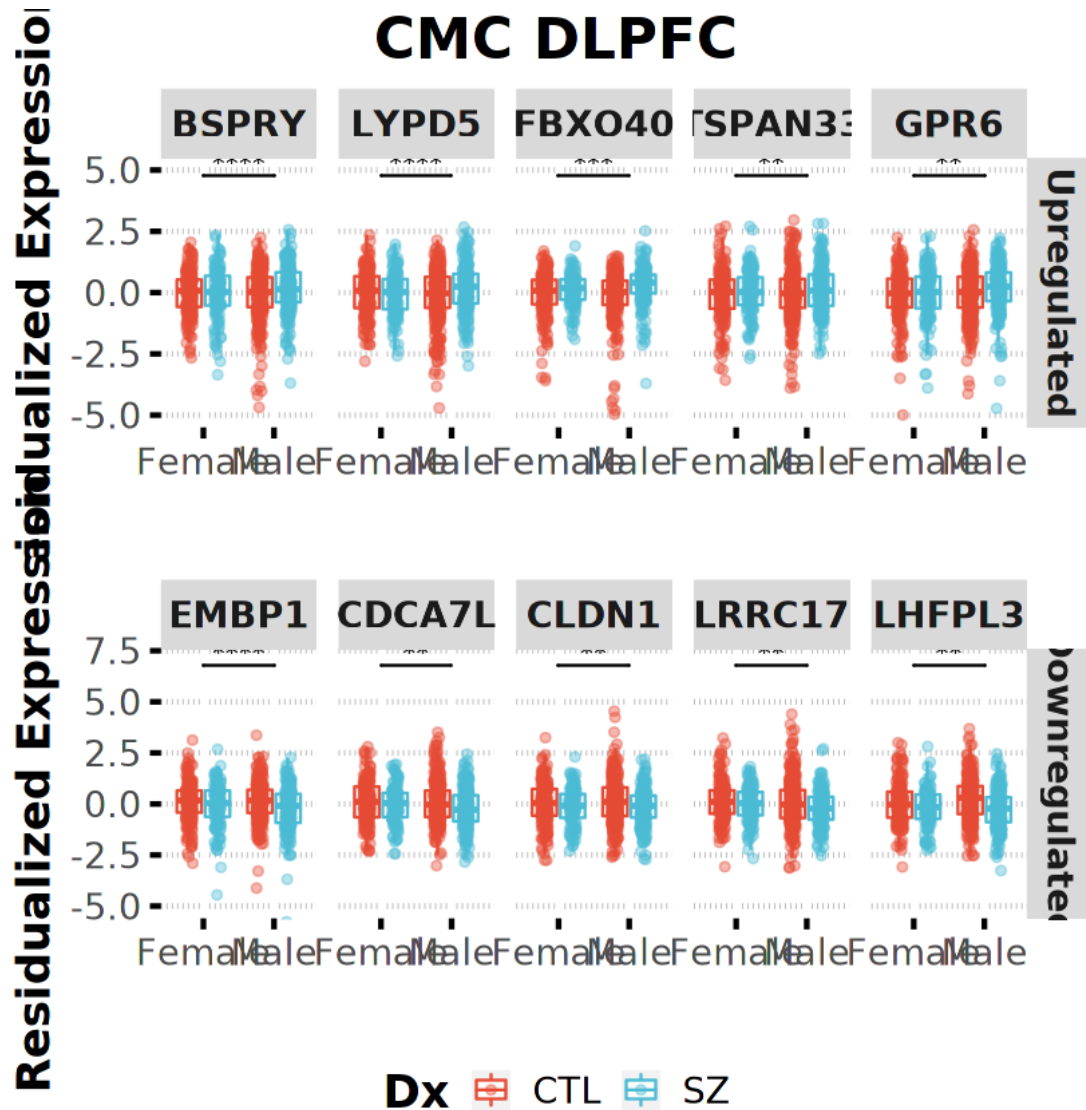
Warning message:
"Ignoring unknown aesthetics: xmin, xmax, annotations, y_position"

1.3.3 Annotate figure

```
[9]: figure <- ggarrange(bxp_up, bxp_down, ncol = 1, nrow = 2,
                         common.legend = TRUE, legend="bottom")
     figure <- annotate_figure(figure, top=text_grob("CMC DLPFC", size=25,␣
     ↪face='bold'))
     figure
```

CMC DLPFC

```
[10]: save_ggplots(figure, 'maleSpecific_SZ', 12, 10)
```

1.4 Session Info

```
[11]: Sys.time()
      proc.time()
      options(width = 120)
      sessioninfo::session_info()
```

```
[1] "2021-08-13 10:37:33 EDT"

    user  system elapsed
  14.964   2.153  29.108
```

```
 Session info
setting   value
version   R version 4.0.3 (2020-10-10)
os        Arch Linux
system    x86_64, linux-gnu
ui        X11
language  (EN)
collate   en_US.UTF-8
ctype     en_US.UTF-8
tz        America/New_York
date      2021-08-13

 Packages
package      * version  date       lib source
abind          1.4-5    2016-07-21 [1] CRAN (R 4.0.2)
assertthat     0.2.1    2019-03-21 [1] CRAN (R 4.0.2)
backports      1.2.1    2020-12-09 [1] CRAN (R 4.0.2)
base64enc      0.1-3    2015-07-28 [1] CRAN (R 4.0.2)
broom          0.7.8    2021-06-24 [1] CRAN (R 4.0.3)
Cairo          1.5-12.2 2020-07-07 [1] CRAN (R 4.0.2)
car            3.0-11   2021-06-27 [1] CRAN (R 4.0.3)
carData        3.0-4    2020-05-22 [1] CRAN (R 4.0.2)
cellranger     1.1.0    2016-07-27 [1] CRAN (R 4.0.2)
cli            3.0.0    2021-06-30 [1] CRAN (R 4.0.3)
colorspace     2.0-2    2021-06-24 [1] CRAN (R 4.0.3)
cowplot        1.1.1    2020-12-30 [1] CRAN (R 4.0.2)
crayon         1.4.1    2021-02-08 [1] CRAN (R 4.0.3)
curl           4.3.2    2021-06-23 [1] CRAN (R 4.0.3)
data.table   * 1.14.0   2021-02-21 [1] CRAN (R 4.0.3)
DBI            1.1.1    2021-01-15 [1] CRAN (R 4.0.2)
dbplyr         2.1.1    2021-04-06 [1] CRAN (R 4.0.3)
digest         0.6.27   2020-10-24 [1] CRAN (R 4.0.2)
dplyr        * 1.0.7    2021-06-18 [1] CRAN (R 4.0.3)
ellipsis       0.3.2    2021-04-29 [1] CRAN (R 4.0.3)
evaluate       0.14     2019-05-28 [1] CRAN (R 4.0.2)
fansi          0.5.0    2021-05-25 [1] CRAN (R 4.0.3)
farver         2.1.0    2021-02-28 [1] CRAN (R 4.0.3)
forcats      * 0.5.1    2021-01-27 [1] CRAN (R 4.0.2)
foreign        0.8-80   2020-05-24 [2] CRAN (R 4.0.3)
fs             1.5.0    2020-07-31 [1] CRAN (R 4.0.2)
generics       0.1.0    2020-10-31 [1] CRAN (R 4.0.2)
ggplot2      * 3.3.5    2021-06-25 [1] CRAN (R 4.0.3)
ggpubr       * 0.4.0    2020-06-27 [1] CRAN (R 4.0.2)
ggsci          2.9      2018-05-14 [1] CRAN (R 4.0.2)
ggsignif       0.6.2    2021-06-14 [1] CRAN (R 4.0.3)
glue           1.4.2    2020-08-27 [1] CRAN (R 4.0.2)
gridExtra      2.3      2017-09-09 [1] CRAN (R 4.0.2)
gtable         0.3.0    2019-03-25 [1] CRAN (R 4.0.2)
```

```
haven        2.4.1    2021-04-23 [1] CRAN (R 4.0.3)
hms          1.1.0    2021-05-17 [1] CRAN (R 4.0.3)
htmltools    0.5.1.1  2021-01-22 [1] CRAN (R 4.0.2)
httr         1.4.2    2020-07-20 [1] CRAN (R 4.0.2)
IRdisplay    1.0      2021-01-20 [1] CRAN (R 4.0.2)
IRkernel     1.2      2021-05-11 [1] CRAN (R 4.0.3)
jsonlite     1.7.2    2020-12-09 [1] CRAN (R 4.0.2)
labeling     0.4.2    2020-10-20 [1] CRAN (R 4.0.2)
lifecycle    1.0.0    2021-02-15 [1] CRAN (R 4.0.3)
lubridate    1.7.10   2021-02-26 [1] CRAN (R 4.0.3)
magrittr     2.0.1    2020-11-17 [1] CRAN (R 4.0.2)
modelr       0.1.8    2020-05-19 [1] CRAN (R 4.0.2)
munsell      0.5.0    2018-06-12 [1] CRAN (R 4.0.2)
openxlsx     4.2.4    2021-06-16 [1] CRAN (R 4.0.3)
pbdZMQ       0.3-5    2021-02-10 [1] CRAN (R 4.0.3)
pillar       1.6.1    2021-05-16 [1] CRAN (R 4.0.3)
pkgconfig    2.0.3    2019-09-22 [1] CRAN (R 4.0.2)
purrr      * 0.3.4    2020-04-17 [1] CRAN (R 4.0.2)
R6           2.5.0    2020-10-28 [1] CRAN (R 4.0.2)
Rcpp         1.0.7    2021-07-07 [1] CRAN (R 4.0.3)
readr      * 1.4.0    2020-10-05 [1] CRAN (R 4.0.2)
readxl       1.3.1    2019-03-13 [1] CRAN (R 4.0.2)
repr         1.1.3    2021-01-21 [1] CRAN (R 4.0.2)
reprex       2.0.0    2021-04-02 [1] CRAN (R 4.0.3)
rio          0.5.27   2021-06-21 [1] CRAN (R 4.0.3)
rlang        0.4.11   2021-04-30 [1] CRAN (R 4.0.3)
rstatix      0.7.0    2021-02-13 [1] CRAN (R 4.0.3)
rstudioapi   0.13     2020-11-12 [1] CRAN (R 4.0.2)
rvest        1.0.0    2021-03-09 [1] CRAN (R 4.0.3)
scales       1.1.1    2020-05-11 [1] CRAN (R 4.0.2)
sessioninfo  1.1.1    2018-11-05 [1] CRAN (R 4.0.2)
stringi      1.7.3    2021-07-16 [1] CRAN (R 4.0.3)
stringr    * 1.4.0    2019-02-10 [1] CRAN (R 4.0.2)
svglite      2.0.0    2021-02-20 [1] CRAN (R 4.0.3)
systemfonts  1.0.2    2021-05-11 [1] CRAN (R 4.0.3)
tibble     * 3.1.2    2021-05-16 [1] CRAN (R 4.0.3)
tidyr      * 1.1.3    2021-03-03 [1] CRAN (R 4.0.3)
tidyselect   1.1.1    2021-04-30 [1] CRAN (R 4.0.3)
tidyverse  * 1.3.1    2021-04-15 [1] CRAN (R 4.0.3)
utf8         1.2.1    2021-03-12 [1] CRAN (R 4.0.3)
uuid         0.1-4    2020-02-26 [1] CRAN (R 4.0.2)
vctrs        0.3.8    2021-04-29 [1] CRAN (R 4.0.3)
withr        2.4.2    2021-04-18 [1] CRAN (R 4.0.3)
xml2         1.3.2    2020-04-23 [1] CRAN (R 4.0.2)
zip          2.2.0    2021-05-31 [1] CRAN (R 4.0.3)

[1] /home/jbenja13/R/x86_64-pc-linux-gnu-library/4.0
[2] /usr/lib/R/library
```