main

July 9, 2021

1 Boxplot of top five up- and downregulated genes

```
[1]: library(ggpubr)
     library(ggsignif)
     library(tidyverse)
```

Loading required package: ggplot2

Attaching packages   tidyverse
1.3.1

tibble	3.1.2	dplyr	1.0.7
tidyr	1.1.3	stringr	1.4.0
readr	1.4.0	forcats	0.5.1
purrr	0.3.4		

Conflicts
tidyverse_conflicts()
  dplyr::filter() masks stats::filter()
  dplyr::lag()   masks stats::lag()

1.1 Functions

```
[2]: 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)
     }

     save_ggplots <- function(fn, p, w, h){
```

```
    for(ext in c('.pdf', '.png', '.svg')){
        ggsave(paste0(fn, ext), plot=p, width=w, height=h)
    }
}
```

1.2 All genes

1.2.1 Differential expressed genes

```
[3]: genes = read.delim('../../_m/genes/diffExpr_szVctl_full.txt')
     genes = genes %>% filter(adj.P.Val < 0.05) %>%
                 select(gencodeID, ensemblID, Symbol, logFC, t, adj.P.Val)
     genes = add_symnum(genes)

     up_genes = genes %>% filter(t > 0) %>% mutate('Direction'='Up') %>% head(5)
     down_genes = genes %>% filter(t < 0) %>% mutate('Direction'='Down') %>% 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
```

	gencodeID <chr>	ensemblID <chr>	Symbol <chr>	logFC <dbl>	t <dbl>	adj.P <db
	ENSG00000248587.7	ENSG00000248587	GDNF-AS1	0.8015019	12.696887	1.38
	ENSG00000138944.7	ENSG00000138944	KIAA1644	0.5637328	12.073351	1.70
	ENSG00000185052.11	ENSG00000185052	SLC24A3	0.2917627	11.122852	3.97
A data.frame: 10×10	ENSG00000140015.19	ENSG00000140015	KCNH5	0.5156546	10.185331	6.71
	ENSG00000171004.17	ENSG00000171004	HS6ST2	0.3021052	9.670025	3.19
	ENSG00000180176.14	ENSG00000180176	TH	-0.7962020	-8.362986	1.89
	ENSG00000182674.5	ENSG00000182674	KCNB2	-0.2712236	-7.472987	4.97
	ENSG00000214216.10	ENSG00000214216	IQCJ	-0.2860577	-7.315037	1.17
	ENSG00000159640.15	ENSG00000159640	ACE	-0.2366734	-7.176286	2.63
	ENSG00000253734.1	ENSG00000253734	LINC01289	-0.5606620	-6.722036	2.79

1.2.2 Load residualized expression

```
[4]: res_file = '../../_m/genes/residualized_expression.tsv'
     tmp_df = data.table::fread(res_file, data.table=F)

     resdf0 = tmp_df %>% filter(V1 %in% sig_genes$gencodeID) %>%
      →column_to_rownames("V1")
     resdf0 <- data.frame(t(resdf0))
```

```
resdf0[1:2, 1:5]
```

Warning message in data.table::fread(res_file, data.table = F):
"Detected 393 column names but the data has 394 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

	ENSG00000214216.10 <dbl>	ENSG00000248587.7 <dbl>	ENSG00000253734.1 <dbl>	ENSG00000 <dbl>
R12864	-0.1600779	0.9255823	-0.1815321	-0.9924114
R12865	1.2913529	0.2076055	-0.7973130	0.4257132

1.2.3 Load pheno data

```
[5]: pheno_file = '/ceph/projects/v4_phase3_paper/inputs/phenotypes/_m/
     ↪caudate_phenotypes.csv'
pheno = read.csv(pheno_file, row.names=1) %>% mutate_if(is.character, as.factor)
pheno[1:2, 1:10]
```

A data.frame: 2 × 10

	Sex <fct>	Race <fct>	Dx <fct>	Age <dbl>	mitoRate <dbl>	rRNA_rate <dbl>	totalAssignedGe <dbl>
R12864	Female	AA	SZ	42.98	0.03265387	8.675159e-05	0.5148545
R12865	Male	AA	SZ	53.12	0.01978740	6.976684e-05	0.5207789

1.2.4 Merge dataframe

```
[6]: resdf <- inner_join(rownames_to_column(pheno), rownames_to_column(resdf0),␣
     ↪by="rowname")
dim(resdf)
resdf[1:2, 1:10]
```

1. 393 2. 31

A data.frame: 2 × 10

	rowname <chr>	Sex <fct>	Race <fct>	Dx <fct>	Age <dbl>	mitoRate <dbl>	rRNA_rate <dbl>	totalAssign <dbl>
1	R12864	Female	AA	SZ	42.98	0.03265387	8.675159e-05	0.5148545
2	R12865	Male	AA	SZ	53.12	0.01978740	6.976684e-05	0.5207789

1.2.5 Melt data frame

```
[7]: df = resdf %>% select(c('rowname', 'Dx', starts_with('ENSG'))) %>%
     pivot_longer(-c(rowname, 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))
head(df, 2)
```

A tibble: 2 × 13	rowname <chr>	Dx <fct>	gencodeID <fct>	Res <dbl>	ensemblID <chr>	Symbol <fct>	log <d
	R12864	SZ	ENSG00000214216.10	-0.1600779	ENSG00000214216	IQCJ	-0.
	R12864	SZ	ENSG00000248587.7	0.9255823	ENSG00000248587	GDNF-AS1	0.8

1.2.6 Initial ggplot with ggpubr
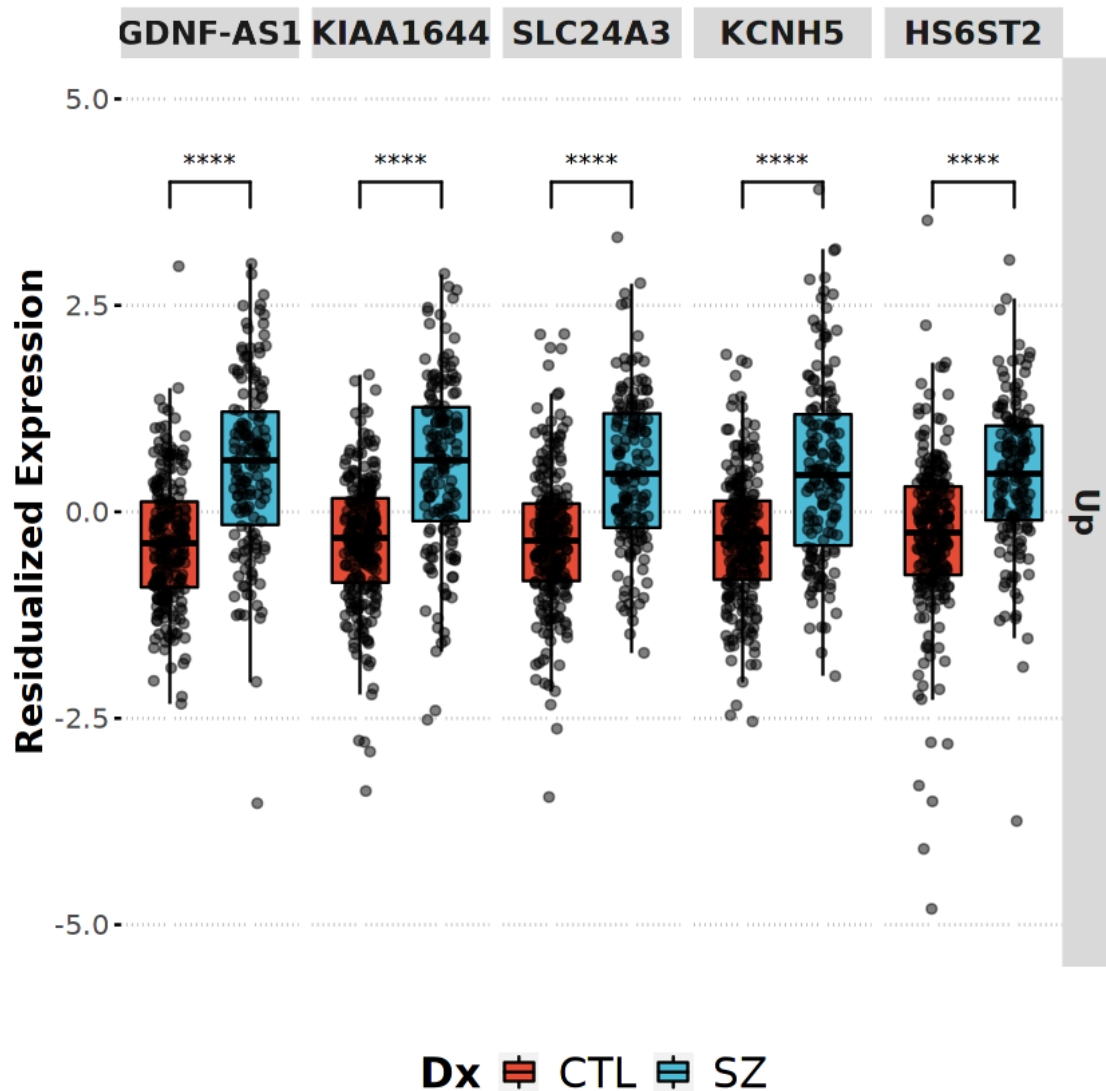
```
[8]: tmp = sig_genes %>% mutate(group1='CTL', group2='SZ', y_pos=4) %>%
         filter(Direction == 'Up') %>%
         mutate_if(is.character, as.factor)

     bxp_up <- df %>% filter(Direction=="Up") %>%
         ggboxplot(x="Dx", y="Res", facet.by=c("Direction", "New_ID"), fill="Dx",
                   xlab='', panel.labs.font=list(face='bold', size = 14),␣
      →palette="npg",
                   outlier.shape=NA, ylab='Residualized Expression', add='jitter',
                   add.params=list(alpha=0.5), legend="bottom", ylim=c(-5, 5),
                   ggtheme=theme_pubclean()) +
         geom_signif(data=tmp,
                     aes(xmin=group1, xmax=group2, annotations=p.signif,
                         y_position=y_pos), manual=TRUE) +
         font("xy.text", size=12) + font("xy.title", size=16, face="bold") +
         font("legend.title", size=18, face="bold") +
         font("legend.text", size=18) + rremove("x.text") + rremove("x.ticks")

     bxp_up
```

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

GDNF-AS1 | KIAA1644 | SLC24A3 | KCNH5 | HS6ST2

Dx ⊟ CTL ⊟ SZ

```
tmp = sig_genes %>% mutate(group1='CTL', group2='SZ', y_pos=4) %>%
    filter(Direction == 'Down') %>%
    mutate_if(is.character, as.factor)

bxp_down <- df %>% filter(Direction=='Down') %>%
    ggboxplot(x="Dx", y="Res", fill="Dx", palette="npg", xlab='', add="jitter",
            outlier.shape=NA, facet.by=c("Direction", "New_ID"), ylim=c(-5,
→5),
            panel.labs.font=list(face='bold', size=14), legend="bottom",
            add.params=list(alpha=0.5), ylab='Residualized Expression',
            ggtheme=theme_pubclean()) +
    geom_signif(data=tmp,
            aes(xmin=group1, xmax=group2, annotations=p.signif,
```
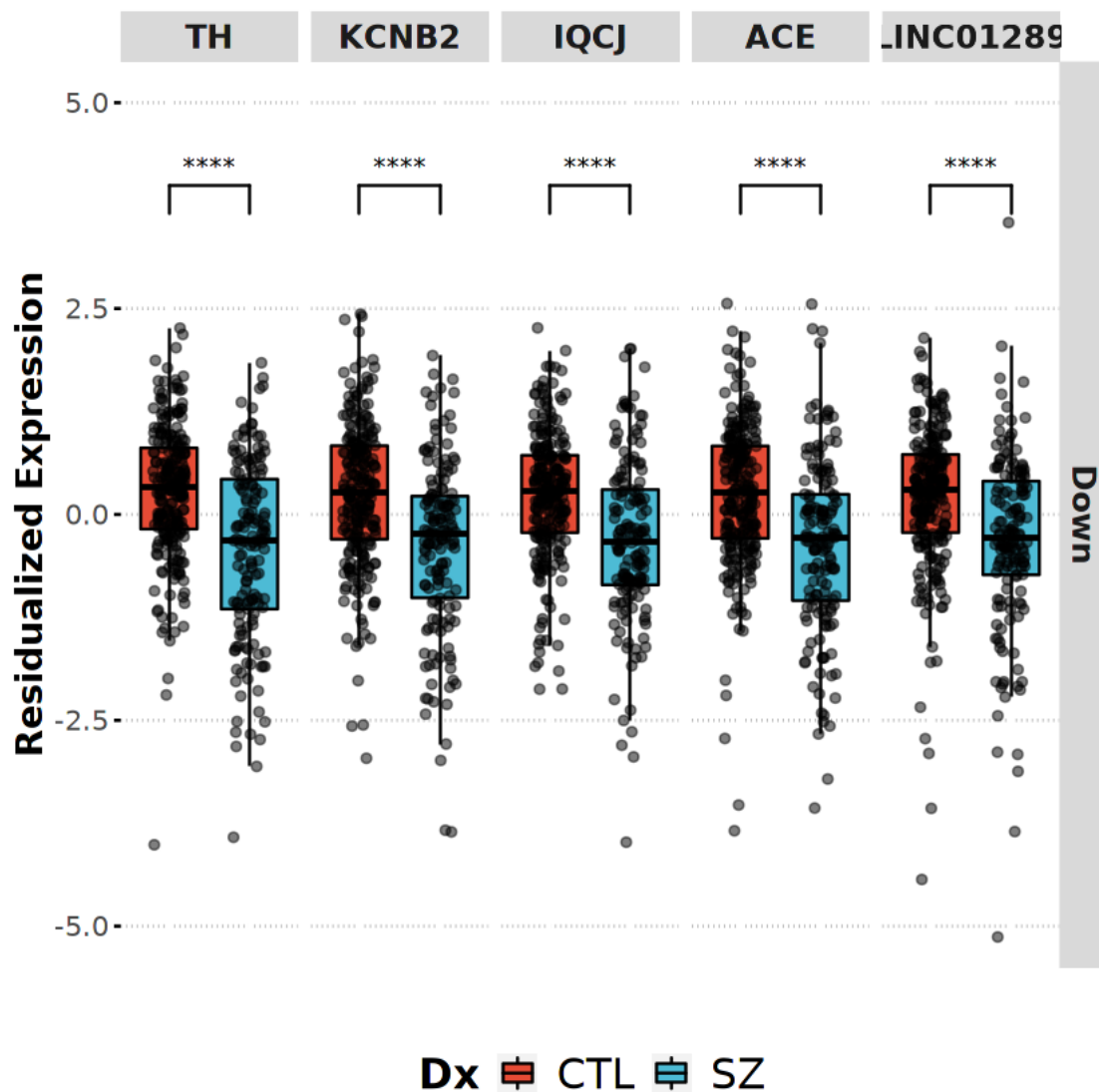
```
                        y_position=y_pos), manual=TRUE) +
    font("xy.text", size=12) + font("xy.title", size=16, face="bold") +
    font("legend.title", size=18, face="bold") +
    font("legend.text", size=18) + rremove("x.text") + rremove("x.ticks")

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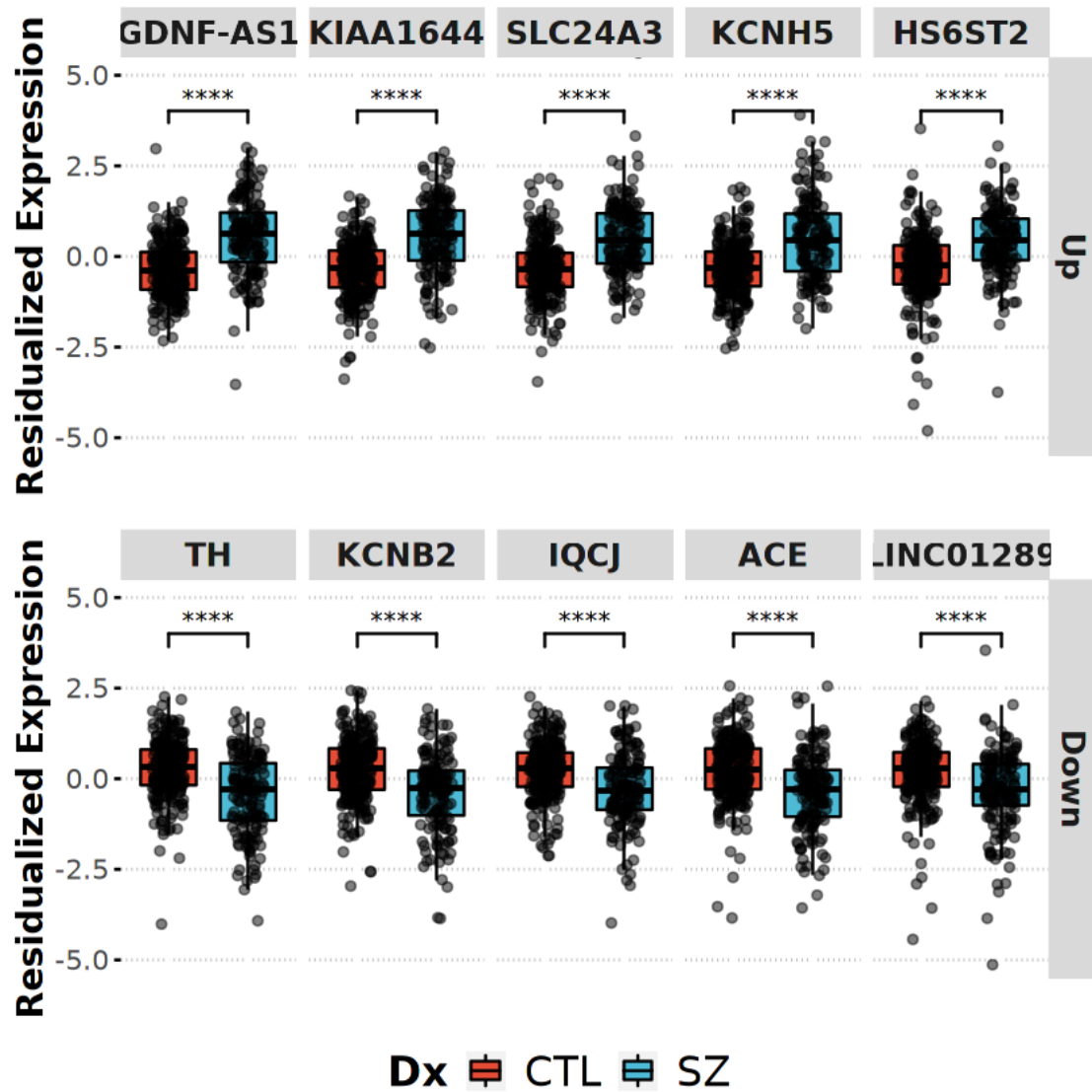
Warning message:
"Ignoring unknown aesthetics: xmin, xmax, annotations, y_position"

1.2.7 Annotate figure

```
[10]: figure <- ggarrange(bxp_up, bxp_down, ncol=1, nrow=2, common.legend=TRUE,
      ↪legend="bottom")
      figure
```



```
[11]: save_ggplots('caudate_topSig_boxplot', figure, 10, 8)
```

1.3 Session Info

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

```
[1] "2021-07-09 14:53:31 EDT"
```

```
   user  system elapsed
 10.894   1.645  12.726
```

```
 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     2021-07-09
```

```
 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        1.6.2    2021-05-17 [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
```