

main

August 16, 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/female_specific_DE_4features.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

```

A data.table: 10 × 15

Feature <chr>	gencodeID <chr>	Symbol <chr>	ensemblID <chr>	Chrom <chr>
ENSG00000070915.9	ENSG00000070915.9	SLC12A3	ENSG00000070915	chr16
ENSG00000263006.6	ENSG00000263006.6	ROCK1P1	ENSG00000263006	chr18
ENSG00000224273.2	ENSG00000224273.2		ENSG00000224273	chr7
ENSG00000003137.8	ENSG00000003137.8	CYP26B1	ENSG00000003137	chr2
ENSG00000167703.14	ENSG00000167703.14	SLC43A2	ENSG00000167703	chr17
ENSG00000111181.12	ENSG00000111181.12	SLC6A12	ENSG00000111181	chr12
ENSG00000249669.8	ENSG00000249669.8	CARMN	ENSG00000249669	chr5
ENSG00000183298.5	ENSG00000183298.5	RPSAP19	ENSG00000183298	chr1
ENSG00000223749.8	ENSG00000223749.8	MIR503HG	ENSG00000223749	chrX
ENSG00000251513.2	ENSG00000251513.2		ENSG00000251513	chr5

1.2 Load phenotypes

```

[4]: pheno_file = '/ceph/projects/v3_phase3_paper/inputs/phenotypes/_m/
      ↪caudate_phenotypes.csv'
pheno = fread(pheno_file) %>% column_to_rownames(var='V1') %>%
      mutate_if(is.character, as.factor)
levels(pheno$Sex) <- c("Female", "Male")
levels(pheno$Dx) <- c("BD", "CTL", "SZ")
pheno %>% head(2)

```

A data.frame: 2 × 8

	BrNum <fct>	RNum <fct>	Region <fct>	RIN <dbl>	Age <dbl>	Sex <fct>	Race <fct>	Dx <fct>
R12864	Br1303	R12864	Caudate	9.6	42.98	Female	AA	SZ
R12865	Br1320	R12865	Caudate	9.5	53.12	Male	AA	SZ

1.3 Load residualized expression

```

[5]: res_file = '../..../interaction_model/caudate/_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("RNum")
resdf0[1:2, 1:5]

```

Warning message in fread(res_file):
 "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		RNum	ENSG00000183298.5	ENSG00000003137.8	ENSG00000251513.2	ENSG00000000000.0
		<chr>	<dbl>	<dbl>	<dbl>	<dbl>
	1	R12864	0.70060990	0.3263843	0.7323149	0.272581
	2	R12865	0.08171483	-1.1774422	-1.0513554	0.692609

1.3.1 Merge and melt dataframe

```
[6]: df <- inner_join(pheno, resdf0, by="RNum") %>%
      select(c('RNum', 'Sex', 'Dx', starts_with('ENSG')) %>%
      pivot_longer(-c(RNum, 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. 3930 2. 19

A tibble: 2 × 10	RNum	Sex	Dx	gencodeID	Res	Feature	Symbol
	<chr>	<fct>	<fct>	<fct>	<dbl>	<chr>	<fct>
	R12864	Female	SZ	ENSG00000183298.5	0.7006099	ENSG00000183298.5	RPSAP1
	R12864	Female	SZ	ENSG00000003137.8	0.3263843	ENSG00000003137.8	CYP26B

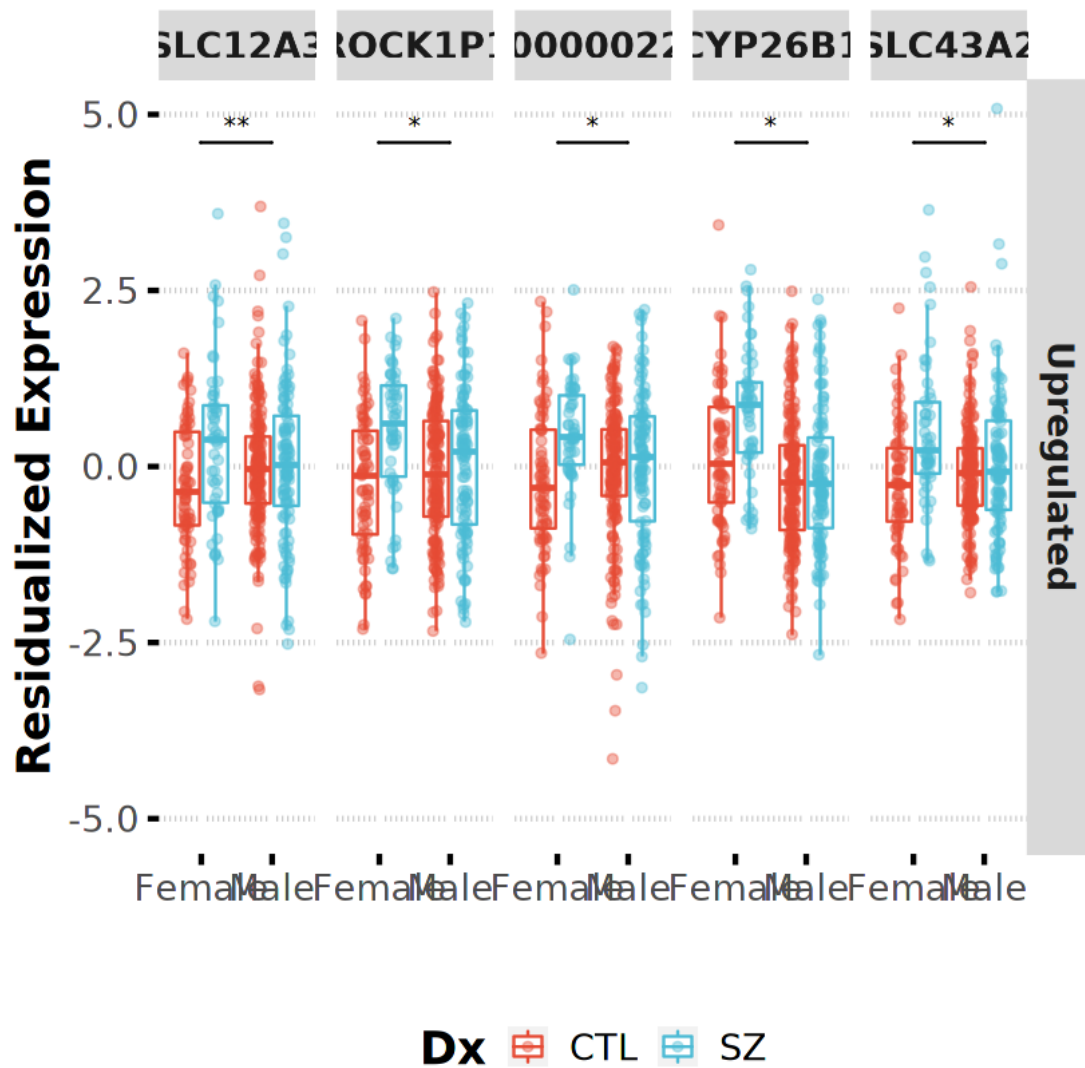
1.3.2 Initial ggplot with ggpubr

```
[7]: tmp = sig_genes %>%
      mutate(group1='Female', group2='Male', y_pos=4.6) %>%
      filter(Direction == 'Upregulated') %>%
      mutate_if(is.character, as.factor)

bxp_up <- df %>% filter(Direction == "Upregulated") %>%
      ggboxplot(x="Sex", y="Res", color="Dx", facet.by=c("Direction", "New_ID"),
      xlab="", add="jitter", palette="npg", ylim=c(-5,5),
      panel.labs.font=list(face='bold'), legend="bottom",
      ylab='Residualized Expression', add.params=list(alpha=0.4),
      ggtheme=theme_pubclean(base_size=20))+
      geom_signif(data=tmp, tip_length = 0, manual=TRUE,
      aes(xmin=group1, xmax=group2, annotations=p.signif, y_position_
      ↪ = y_pos)) +
      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=4.6) %>%
  filter(Direction == 'Downregulated') %>%
  mutate_if(is.character, as.factor)

bxp_down <- df %>% filter(Direction=="Downregulated") %>%
  ggboxplot(x="Sex", y="Res", color="Dx", facet.by=c("Direction", "New_ID"),
    xlab='', palette="npg", add='jitter', ylim=c(-5, 5),
    panel.labs.font=list(face='bold'), legend="bottom",
    ylab='Residualized Expression', add.params=list(alpha=0.4),
    ggtheme=theme_pubclean(base_size=20)) +
  geom_signif(data=tmp, tip_length = 0, manual=TRUE,
```

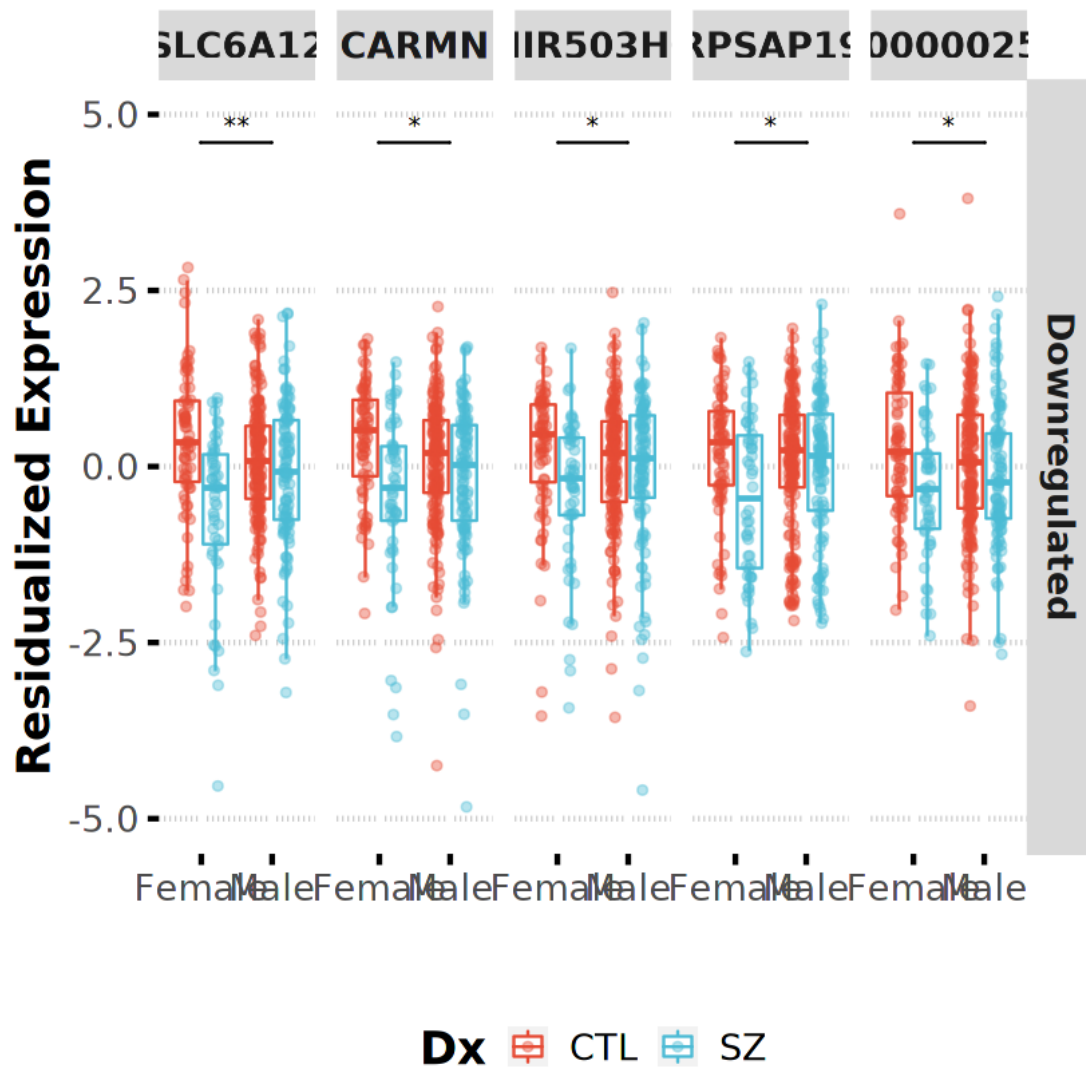
```

aes(xmin=group1, xmax=group2, annotations=p.signif, y_position_
↪= y_pos)) +
font("xy.title", face="bold") + font("legend.title", face="bold")
bxp_down

```

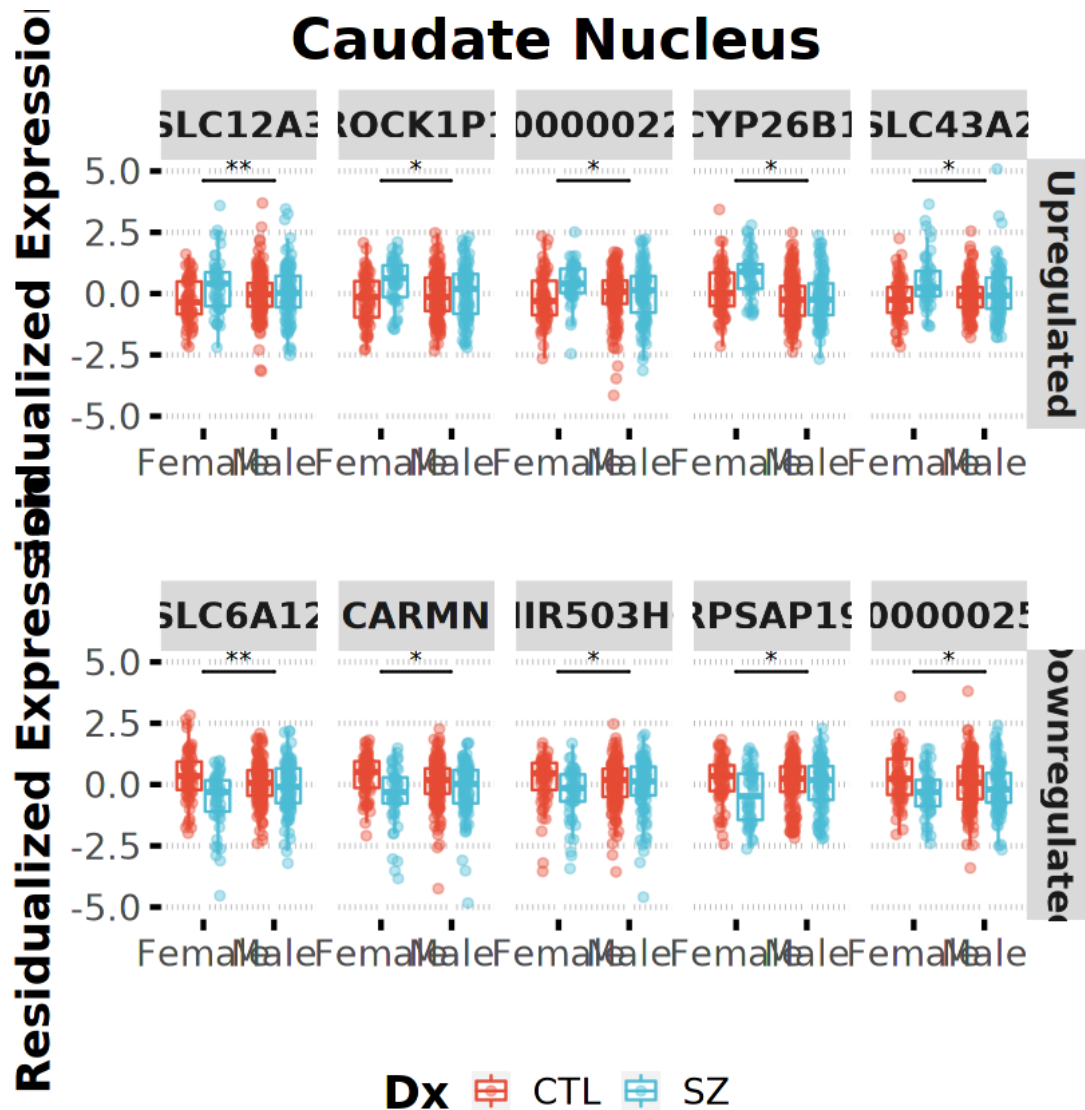
Warning message:

"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("Caudate Nucleus", size=25,
↪face='bold'))
figure
```



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

1.4 Session Info

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

```
[1] "2021-08-16 18:48:36 EDT"
```

```
   user  system elapsed  
19.861   2.865   20.196
```

```
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     2021-08-16
```

```
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