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()
                     masks stats::lag()
     dplyr::lag()
```

1.1 Functions

```
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_maleVfemale_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'='Male Bias') %>%__
     \rightarrowhead(5)
     down_genes = genes %>% filter(t < 0) %>% mutate('Direction'='Female Bias') %>%_
      \rightarrowhead(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$Symbol[xx] == '',</pre>
                                         as.character(sig_genes$ensemblID[xx]),
                                         as.character(sig_genes$Symbol[xx]))
     }
     sig_genes
```

	gencodeID	ensemblID	Symbol	logFC	\mathbf{t}	adj.P.V
	<chr></chr>	<chr></chr>	<chr $>$	<dbl $>$	<dbl $>$	<dbl></dbl>
	ENSG00000114374.12	ENSG00000114374	USP9Y	8.6836787	100.18087	1.95363
	ENSG00000183878.15	ENSG00000183878	UTY	8.5971522	88.46621	8.13312
	ENSG00000012817.15	ENSG00000012817	KDM5D	8.6930099	88.03682	3.59349
A data frama, 10 v 10	ENSG00000067048.16	ENSG00000067048	DDX3Y	8.5878028	86.79686	5.03518
A data.frame: 10×10	ENSG00000067646.11	ENSG00000067646	ZFY	8.5329011	81.80364	9.10601
	ENSG00000229807.10	ENSG00000229807	XIST	-9.2961371	-100.35608	1.95362
	ENSG00000005889.15	ENSG00000005889	ZFX	-0.6315196	-46.06126	3.88770
	ENSG00000147050.14	ENSG00000147050	KDM6A	-0.5283117	-31.99315	3.01783
	ENSG00000126012.11	ENSG00000126012	KDM5C	-0.4877813	-31.52353	1.7667
	ENSG00000130021.13	ENSG00000130021	PUDP	-0.5965470	-27.93309	1.6502!

1.2.2 Load residualized expression

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."

	ļ	ENSG00000130021.13	ENSG00000005889.15	ENSG00000147050.14	ENSG00
A data frama, 2 v E	ļ	<dbl></dbl>	<dbl></dbl>	<dbl></dbl>	<dbl $>$
A data.frame: 2×5	R12864	1.7296843	2.2011553	0.7197812	0.777743
	R12865	-0.6949843	-0.7467199	-0.5590480	-0.342652

1.2.3 Load pheno data

		BrNum	RNum	Region	RIN	Age	Sex	Race	Dx
A data.frame: 2×8		<fct></fct>	<fct $>$	<fct $>$	<dbl $>$	<dbl $>$	<fct $>$	<fct $>$	<fct $>$
	R12864	Br1303	R12864	Caudate	9.6	42.98	Female	AA	Schizo
	R12865	Br1320	R12865	Caudate	9.5	53.12	Male	AA	Schizo

1.2.4 Merge dataframe

1. 393 2. 19

	rowname	BrNum	RNum	Region	RIN	Age	Sex	Race	Dx
A data frame: 2 × 10	<chr></chr>	<fct $>$	<fct $>$	<fct $>$	<dbl $>$	<dbl $>$	<fct $>$	<fct $>$	<fct $>$
A data.frame: $2 \times 10 \frac{1}{1}$	R12864	Br1303	R12864	Caudate	9.6	42.98	Female	AA	Schizo
2	R12865	Br1320	R12865	Caudate	9.5	53.12	Male	AA	Schizo

1.2.5 Melt data frame

```
rowname Sex
                                gencodeID
                                                      Res
                                                               ensemblID
                                                                                  Symbol logFC
               <chr>
                        <fct>
                                 <fct>
                                                      <dbl>
                                                               <chr>
                                                                                  <fct>
                                                                                          <dbl>
A tibble: 2 \times 13
              R12864
                                ENSG00000130021.13
                                                     1.729684
                                                               ENSG00000130021
                                                                                  PUDP
                        Female
                                                                                          -0.59654
                        Female ENSG00000005889.15 2.201155
                                                               ENSG00000005889
                                                                                          -0.63151
               R12864
                                                                                 ZFX
```

1.2.6 Initial ggplot with ggpubr

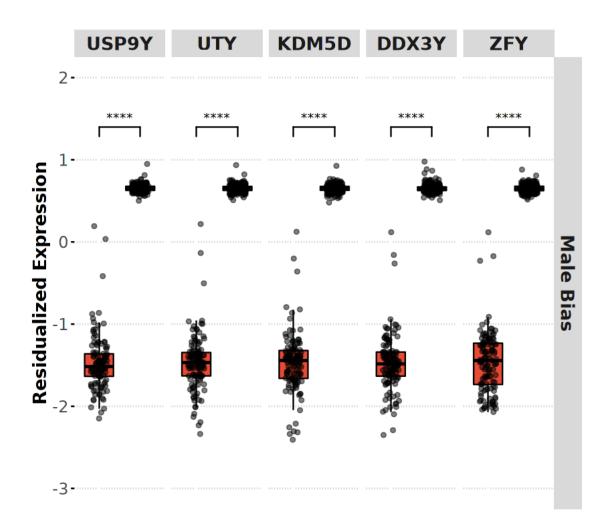
```
[8]: | tmp = sig_genes %>% mutate(group1='Female', group2='Male', y_pos=1.4) %>%__
      →filter(Direction == 'Male Bias') %>%
         mutate if(is.character, as.factor)
     bxp up <- df %>% filter(Direction=="Male Bias") %>%
             ggboxplot(x="Sex", y="Res", facet.by=c("Direction", "New_ID"),__

→fill="Sex", xlab='',
                       panel.labs.font=list(face='bold', size = 16), palette="npg", __
      →outlier.shape=NA,
                       ylab='Residualized Expression', add='jitter', ylim=c(-3, 2),
                       add.params=list(alpha=0.5), legend="bottom", __

→ggtheme=theme_pubclean()) +
              geom_signif(data=tmp,
                           aes(xmin=group1, xmax=group2, annotations=p.signif,__
      \hookrightarrowy_position = y_pos),
                           manual=TRUE) +
              font("xy.title", size=16, face="bold") + font("xy.text", size=14) +
              font("legend.title", size=18, face="bold") +
              font("legend.text", size=18) + rremove("x.text") + rremove("x.ticks")
     bxp_up
```

Warning message:

[&]quot;Ignoring unknown aesthetics: xmin, xmax, annotations, y position"



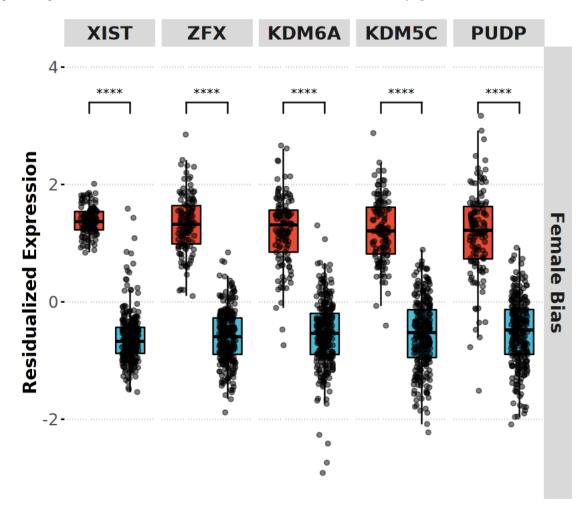
Sex

Female

Male

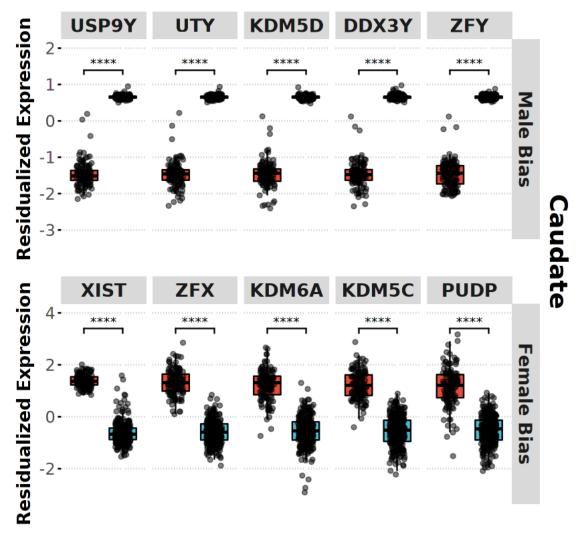
Warning message:

"Ignoring unknown aesthetics: xmin, xmax, annotations, y_position"



Sex **=** Female **=** Male

1.2.7 Annotate figure



Sex

Female

Male

```
[11]: save_ggplots('caudate_topSig_boxplot', figure, 12, 8)
```

1.3 Autosomal only

1.3.1 Differential expressed genes

```
[12]: genes = read.csv('../../metrics_summary/_m/autosomal_DEG.csv')
      genes = genes %>% rename("gencodeID" = "Feature") %>%
              filter(adj.P.Val < 0.05) %>%
              select(gencodeID, ensemblID, Symbol, logFC, adj.P.Val)
      genes = add_symnum(genes)
      up_genes = genes %>% filter(logFC > 0) %>% mutate('Direction'='Male Bias') %>%_
      \rightarrowhead(5)
      down_genes = genes %>% filter(logFC < 0) %>% mutate('Direction'='Female Bias')_
       \rightarrow%>% 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(as.character(sig_genes$Symbol[xx]) == '',</pre>
                                          as.character(sig_genes$ensemblID[xx]),
                                          as.character(sig_genes$Symbol[xx]))
      sig_genes
```

	gencodeID	ensemblID	Symbol	logFC	adj.P.Val	FD
	<chr></chr>	<chr $>$	<chr $>$	<dbl $>$	<dbl $>$	<c< td=""></c<>
•	ENSG00000205611.4	ENSG00000205611	LINC01597	1.3079897	6.877867e-20	< :
	ENSG00000283443.1	ENSG00000283443		1.4162533	2.838705e-18	< 5
	ENSG00000149531.15	ENSG00000149531	FRG1BP	0.6715762	1.190836e-17	< 5
A data.frame: 10×9	ENSG00000282826.1	ENSG00000282826	FRG1CP	0.5551378	1.474014e-15	1.5
A data.frame: 10×9	ENSG00000258484.3	ENSG00000258484	SPESP1	0.7304981	4.974455e-12	5.0
	ENSG00000095932.6	ENSG00000095932	SMIM24	-0.9013291	1.022672e-15	1.0
	ENSG00000261600.1	ENSG00000261600		-0.9681869	3.921912e-13	3.9
	ENSG00000236698.1	ENSG00000236698	EIF1AXP1	-0.2509585	2.863151e-12	2.9
	ENSG00000241532.1	ENSG00000241532	AGGF1P3	-0.7277903	8.886125e-11	8.9
	ENSG00000180574.3	ENSG00000180574		-0.2318311	9.268579 e-11	9.3

1.3.2 Load residualized expression

```
[13]: resdf0 = tmp_df %>% filter(V1 %in% sig_genes$gencodeID) %>%⊔

→column_to_rownames("V1")

resdf0 <- data.frame(t(resdf0))

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

		ENSG00000236698.1	ENSG00000261600.1	ENSG00000241532.1	ENSG00000
A data.frame: 2×5		<dbl></dbl>	<dbl></dbl>	<dbl></dbl>	<dbl $>$
	R12864	-0.0355813	0.4617652	0.05548178	2.0144561
	R12865	-1.1127706	-1.1681747	0.50413144	-0.5880534

1.3.3 Merge dataframe

1. 393 2. 19

		rowname	BrNum	RNum	Region	RIN	Age	Sex	Race	Dx
A data.frame: $2 \times 10 -$		<chr></chr>	<fct $>$	<fct $>$	<fct $>$	<dbl $>$	<dbl $>$	<fct $>$	<fct $>$	<fct $>$
	1	R12864	Br1303	R12864	Caudate	9.6	42.98	Female	AA	Schizo
	2	R12865	Br1320	R12865	Caudate	9.5	53.12	Male	AA	Schizo

1.3.4 Melt data frame

		Sex	gencodeID		ensemblID	Symbol	$\log l$
A tibble 2 × 12	<fct $>$	<fct $>$	<fct> ENSG00000236698.1 ENSG00000261600.1</fct>	<dbl $>$	<chr $>$	<fct $>$	<dl
A tibble, 2 × 12	R12864	Female	ENSG00000236698.1	-0.0355813	ENSG00000236698	EIF1AXP1	-0.2
	R12864	Female	ENSG00000261600.1	0.4617652	ENSG00000261600		-0.9

1.3.5 Initial ggplot with ggpubr

```
aes(xmin=group1, xmax=group2, annotations=p.signif, u

y_position = y_pos),

manual=TRUE) +

font("xy.title", size=16, face="bold") + font("xy.text", size=14) +

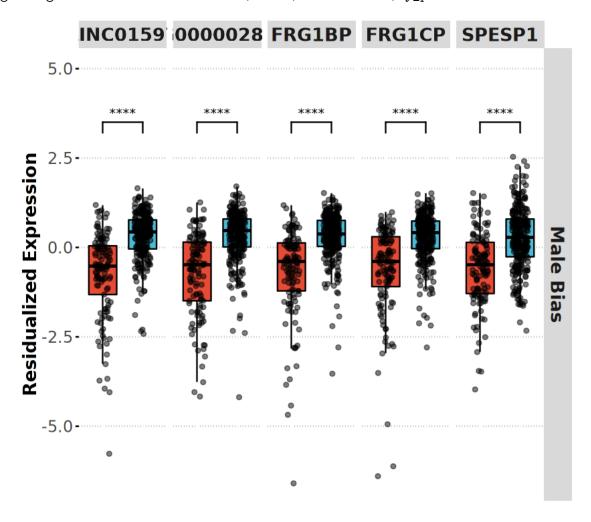
font("legend.title", size=18, face="bold") +

font("legend.text", size=18) + rremove("x.text") + rremove("x.ticks")

bxp_up
```

Warning message:

"Ignoring unknown aesthetics: xmin, xmax, annotations, y_position"

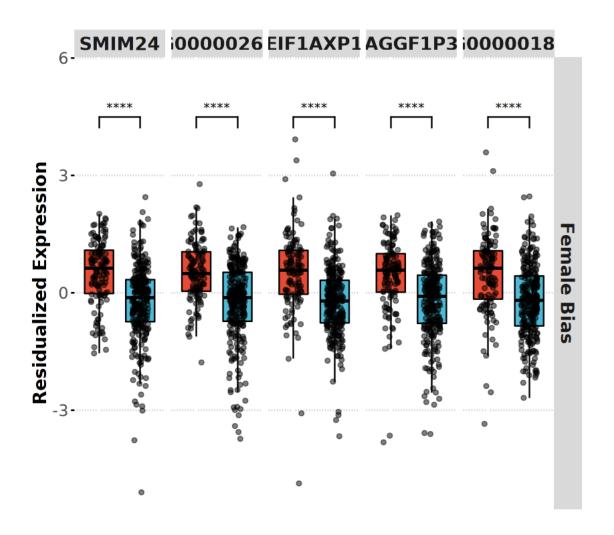


Sex ₱ Female **₱** Male

```
[17]: tmp = sig_genes %>% mutate(group1='Female', group2='Male', y_pos=4.5) %>%
          filter(Direction == 'Female Bias') %>%
          mutate_if(is.character, as.factor)
      bxp_down <- df %>% filter(Direction=="Female Bias") %>%
              ggboxplot(x="Sex", y="Res", fill="Sex", palette="npg", xlab='', outlier.
       \rightarrowshape=NA,
                        facet.by=c("Direction", "New_ID"),
                        panel.labs.font=list(face='bold', size = 16),
                        ylab='Residualized Expression', add='jitter', ylim=c(-5, 5.5),
                        add.params=list(alpha=0.5), legend="bottom", u
       →ggtheme=theme_pubclean()) +
               geom_signif(data=tmp,
                            aes(xmin=group1, xmax=group2, annotations=p.signif,__
       \rightarrowy_position = y_pos),
                           manual=TRUE) +
               font("xy.title", size=16, face="bold") + font("xy.text", size=14) +
               font("legend.title", size=18, face="bold") +
               font("legend.text", size=18) + rremove("x.text") + rremove("x.ticks")
      bxp_down
```

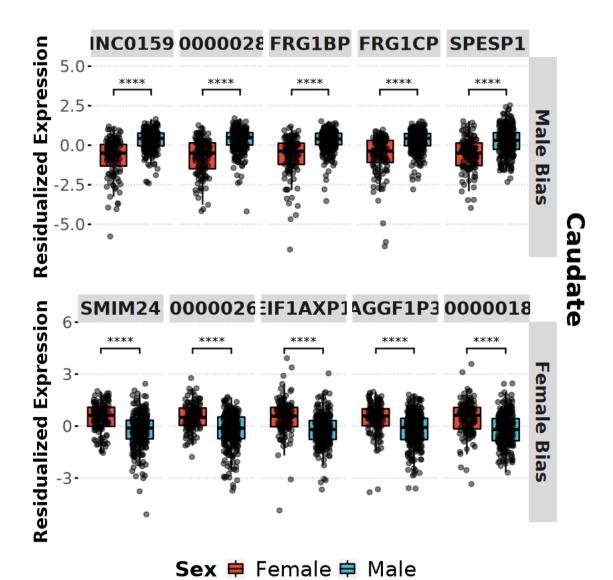
Warning message:

[&]quot;Ignoring unknown aesthetics: xmin, xmax, annotations, y_position"



Sex **=** Female **=** Male

1.3.6 Annotate figure



```
[19]: save_ggplots('caudate_topSig_autosomal_boxplot', figure2, 12, 8)
```

1.4 Session Info

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

[1] "2021-07-09 20:49:18 EDT"
```

user system elapsed 21.941 1.930 22.776

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

1 ackages			_			
package	*	version	date		sour	
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)
•						

```
2.4.1
                        2021-04-23 [1] CRAN (R 4.0.3)
haven
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)
               1.4.2
                        2020-07-20 [1] CRAN (R 4.0.2)
httr
IRdisplay
              1.0
                        2021-01-20 [1] CRAN (R 4.0.2)
               1.2
                        2021-05-11 [1] CRAN (R 4.0.3)
IRkernel
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)
                        2021-02-15 [1] CRAN (R 4.0.3)
lifecycle
              1.0.0
lubridate
              1.7.10
                        2021-02-26 [1] CRAN (R 4.0.3)
              2.0.1
                        2020-11-17 [1] CRAN (R 4.0.2)
magrittr
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)
                        2021-05-16 [1] CRAN (R 4.0.3)
pillar
              1.6.1
              2.0.3
                        2019-09-22 [1] CRAN (R 4.0.2)
pkgconfig
            * 0.3.4
                        2020-04-17 [1] CRAN (R 4.0.2)
purrr
R6
              2.5.0
                        2020-10-28 [1] CRAN (R 4.0.2)
               1.0.7
                        2021-07-07 [1] CRAN (R 4.0.3)
Rcpp
                        2020-10-05 [1] CRAN (R 4.0.2)
readr
            * 1.4.0
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)
                        2021-04-02 [1] CRAN (R 4.0.3)
reprex
              2.0.0
rio
              0.5.27
                        2021-06-21 [1] CRAN (R 4.0.3)
              0.4.11
                        2021-04-30 [1] CRAN (R 4.0.3)
rlang
                        2021-02-13 [1] CRAN (R 4.0.3)
rstatix
              0.7.0
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)
                        2018-11-05 [1] CRAN (R 4.0.2)
sessioninfo
              1.1.1
              1.6.2
                        2021-05-17 [1] CRAN (R 4.0.3)
stringi
stringr
            * 1.4.0
                        2019-02-10 [1] CRAN (R 4.0.2)
                        2021-02-20 [1] CRAN (R 4.0.3)
svglite
              2.0.0
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)
            * 1.3.1
                        2021-04-15 [1] CRAN (R 4.0.3)
tidyverse
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)
              0.3.8
                        2021-04-29 [1] CRAN (R 4.0.3)
vctrs
              2.4.2
                        2021-04-18 [1] CRAN (R 4.0.3)
withr
xm12
              1.3.2
                        2020-04-23 [1] CRAN (R 4.0.2)
              2.2.0
                        2021-05-31 [1] CRAN (R 4.0.3)
zip
```

^{[1] /}home/jbenja13/R/x86_64-pc-linux-gnu-library/4.0

^{[2] /}usr/lib/R/library