main

September 14, 2021

1 Functional enrichment analysis with g:Profiler

```
[1]: library(tidyverse)
     library(gprofiler2)
```

```
── Attaching packages ──────────────────────────── tidyverse
1.3.1 ──

  ggplot2 3.3.5      purrr   0.3.4
  tibble  3.1.4      dplyr   1.0.7
  tidyr   1.1.3      stringr 1.4.0
  readr   2.0.1      forcats 0.5.1

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

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

1.1 Load eQTL results

```
[3]: dt = data.table::fread("../../_m/BrainSeq_sexGenotypes_4features_3regions.txt.
     ↪gz") %>%
         filter(Type == "Gene") %>% mutate(ensemblID=gsub("\\..*", "", gene_id))

     caudate = dt %>% filter(Tissue == "Caudate") %>% distinct(gene_id, .
     ↪keep_all=TRUE)
     dlpfc = dt %>% filter(Tissue == "DLPFC") %>% distinct(gene_id, .keep_all=TRUE)
     hippo = dt %>% filter(Tissue == "Hippocampus") %>% distinct(gene_id, .
     ↪keep_all=TRUE)
```

```
shared = intersect(intersect(caudate$ensemblID, dlpfc$ensemblID),⌷
 ↪hippo$ensemblID)
```

1.2 Calculated enrichment and visual plot

1.2.1 Caudate

```
[4]: gostres <- gost(query=caudate$ensemblID, organism="hsapiens")
     gostres$result %>%
         data.table::fwrite(file = "caudate_functional_enrichment.txt", sep="\t")

     p <- gostplot(gostres, capped = FALSE, interactive = FALSE)
     print(p)
     save_ggplots("caudate_manhattan", p, 9, 5)
```
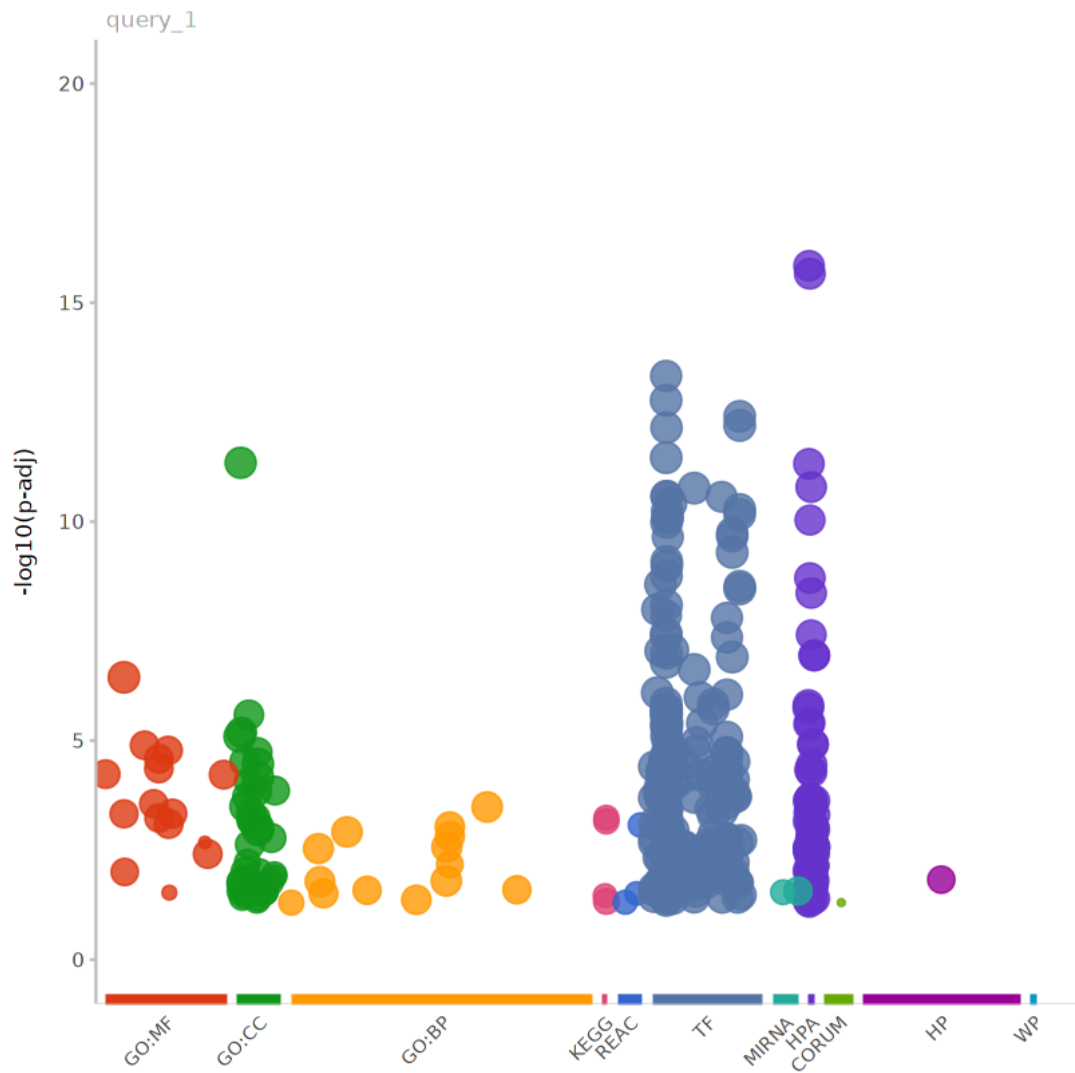
```
[5]: gostres$result %>% arrange(p_value) %>% filter(source == "HPA") %>% head
```

A data.frame: 6 × 14

	query <chr>	significant <lgl>	p_value <dbl>	term_size <int>	query_size <int>	intersection_size <int>	pre <d
1	query_1	TRUE	1.438913e-16	7114	555	455	0.8
2	query_1	TRUE	2.189684e-16	8005	555	491	0.88
3	query_1	TRUE	4.776245e-12	6874	555	432	0.77
4	query_1	TRUE	1.623518e-11	6734	555	424	0.70
5	query_1	TRUE	9.191831e-11	4948	555	335	0.60
6	query_1	TRUE	1.935737e-09	6624	555	412	0.74

```
[6]: gostres$result %>% arrange(p_value) %>% filter(source %in% c("KEGG", "REAC",␣
     ↪"HP", "CORUM"))
```
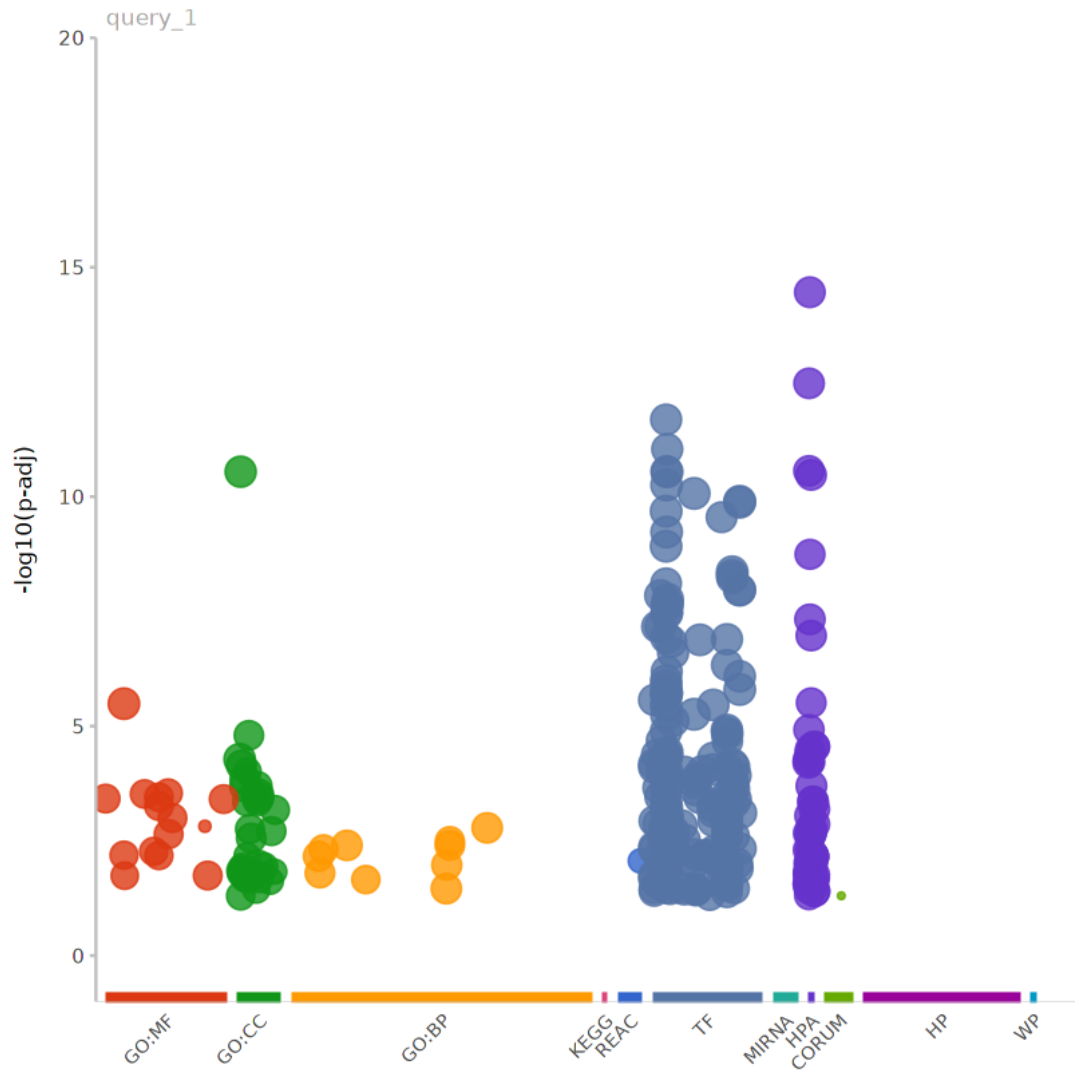
A data.frame: 9 × 14

query <chr>	significant <lgl>	p_value <dbl>	term_size <int>	query_size <int>	intersection_size <int>	precisi <dbl>
query_1	TRUE	0.0005769362	249	377	29	0.0769
query_1	TRUE	0.0006925586	306	377	33	0.0875
query_1	TRUE	0.0008430197	175	492	24	0.0487
query_1	TRUE	0.0148780785	992	236	80	0.3389
query_1	TRUE	0.0307500459	125	492	17	0.0345
query_1	TRUE	0.0344182481	128	377	16	0.0424
query_1	TRUE	0.0468653334	363	377	32	0.0848
query_1	TRUE	0.0488887731	155	492	19	0.0386
query_1	TRUE	0.0498687761	3	222	3	0.0135

1.2.2 DLPFC

```
[7]: gostres <- gost(query=dlpfc$ensemblID, organism="hsapiens")
     gostres$result %>%
         data.table::fwrite(file = "dlpfc_functional_enrichment.txt", sep="\t")

     p <- gostplot(gostres, capped = FALSE, interactive = FALSE)
     print(p)
     save_ggplots("dlpfc_manhattan", p, 9, 5)
```

query_1

-log10(p-adj)

GO:MF GO:CC GO:BP KEGG REAC TF MIRNA HPA CORUM HP WP

```
[8]: gostres$result %>% arrange(p_value) %>% filter(source == "HPA") %>% head
```

A data.frame: 6 × 14		query <chr>	significant <lgl>	p_value <dbl>	term_size <int>	query_size <int>	intersection_size <int>	pre <d
	1	query_1	TRUE	3.497935e-15	8005	503	446	0.8
	2	query_1	TRUE	3.379636e-13	7114	503	408	0.8
	3	query_1	TRUE	2.733495e-11	6874	503	393	0.7
	4	query_1	TRUE	3.365502e-11	6734	503	387	0.7
	5	query_1	TRUE	1.799271e-09	4948	503	303	0.6
	6	query_1	TRUE	4.691130e-08	6624	503	372	0.7

```
[9]: gostres$result %>% arrange(p_value) %>% filter(source %in% c("KEGG", "REAC",␣
      ↪"HP", "CORUM"))
```

A data.frame: 2 × 14	query <chr>	significant <lgl>	p_value <dbl>	term_size <int>	query_size <int>	intersection_size <int>	precisio <dbl>
	query_1	TRUE	0.008646936	175	453	21	0.04635
	query_1	TRUE	0.049711572	3	192	3	0.01562

1.2.3 Hippocampus

```
[10]: gostres <- gost(query=hippo$ensemblID, organism="hsapiens")
      gostres$result %>%
          data.table::fwrite(file = "hippocampus_functional_enrichment.txt", sep="\t")

      p <- gostplot(gostres, capped = FALSE, interactive = FALSE)
      print(p)
      save_ggplots("hippocampus_manhattan", p, 9, 5)
```
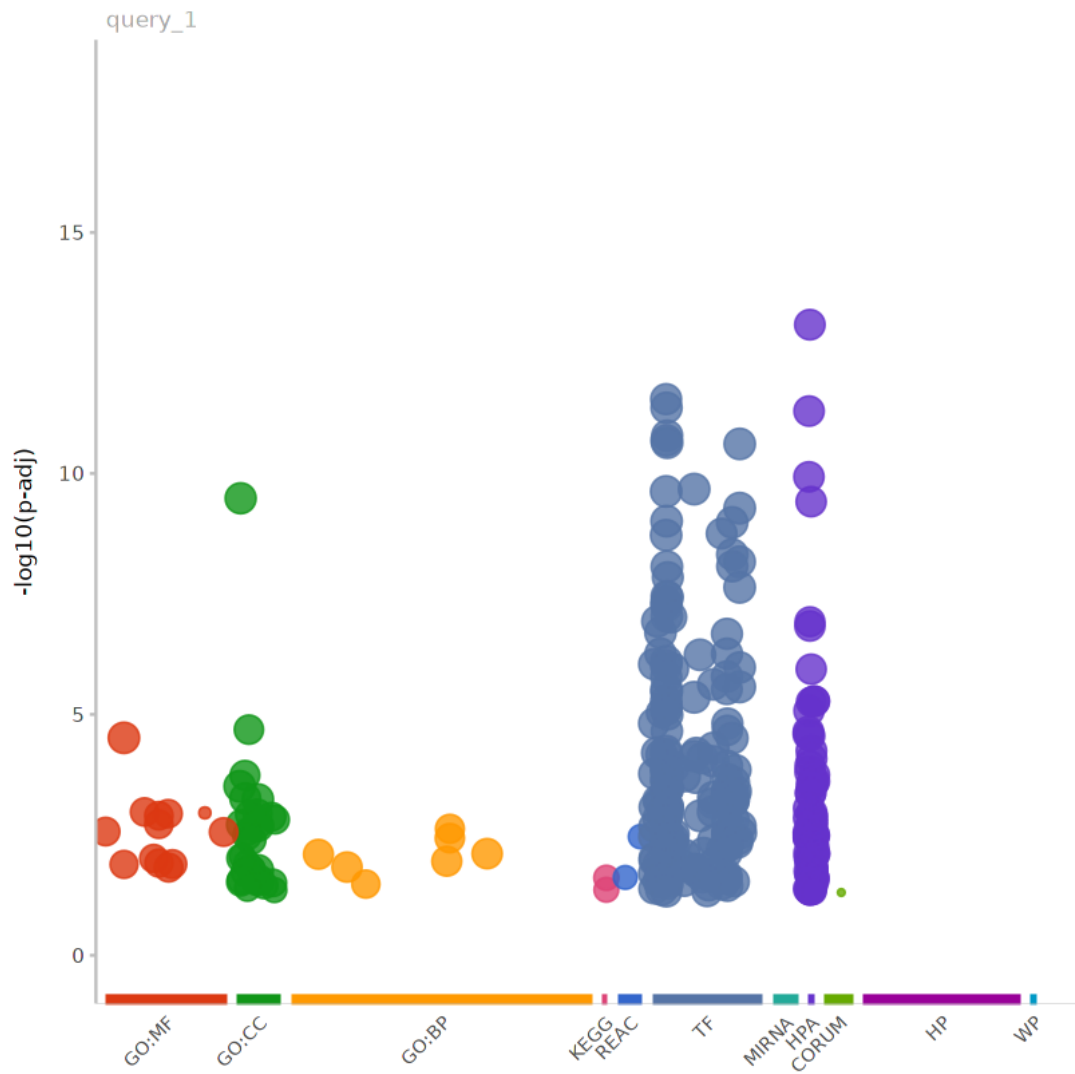
```
[11]: gostres$result %>% arrange(p_value) %>% filter(source == "HPA") %>% head
```

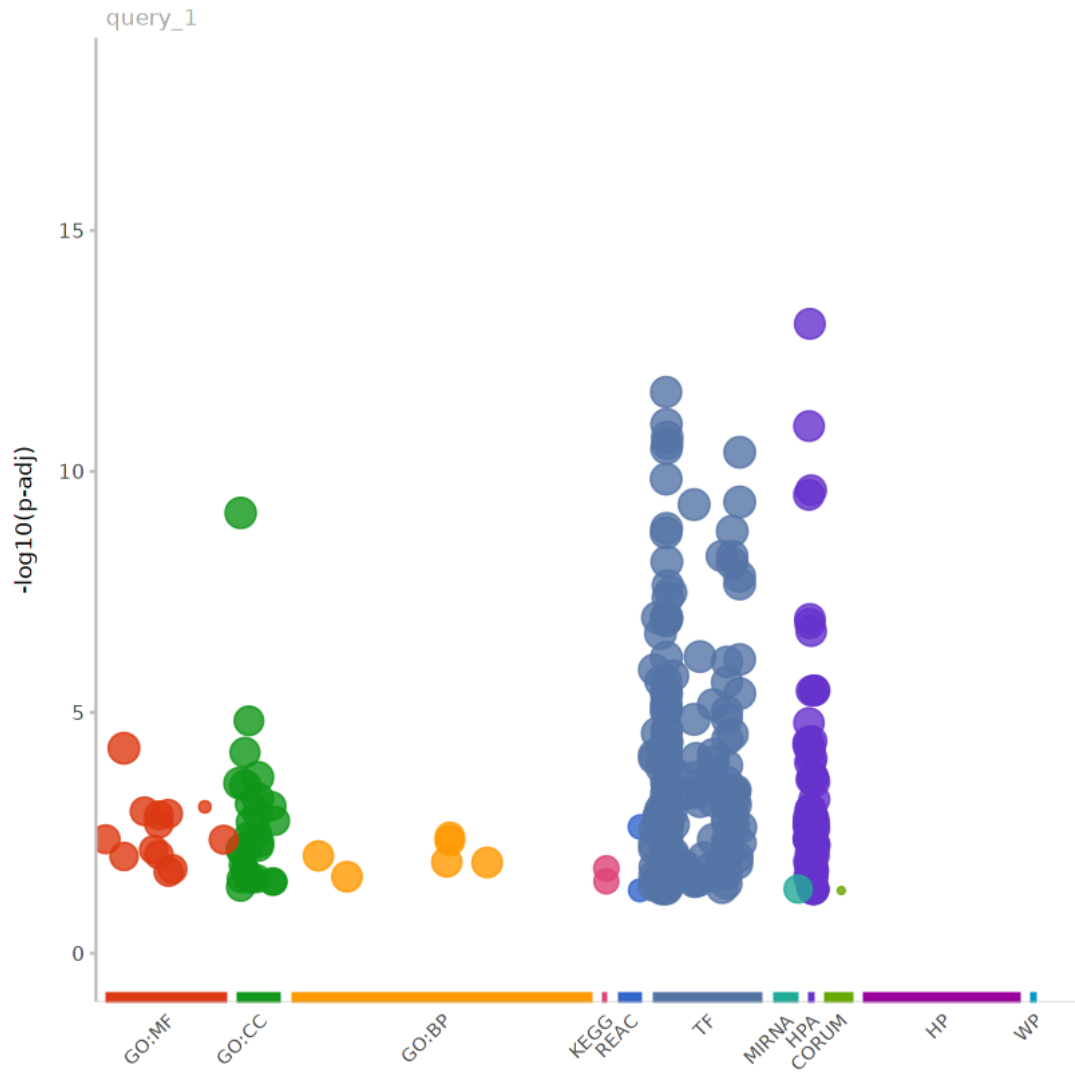|		query	significant	p_value	term_size	query_size	intersection_size	pre
		<chr>	<lgl>	<dbl>	<int>	<int>	<int>	<d
A data.frame: 6 × 14	1	query_1	TRUE	8.160630e-14	8005	476	421	0.8
	2	query_1	TRUE	5.070567e-12	7114	476	385	0.8
	3	query_1	TRUE	1.172854e-10	6874	476	372	0.7
	4	query_1	TRUE	3.846421e-10	6734	476	365	0.7
	5	query_1	TRUE	1.193443e-07	4948	476	282	0.5
	6	query_1	TRUE	1.464580e-07	6624	476	352	0.7

```
[12]: gostres$result %>% arrange(p_value) %>% filter(source %in% c("KEGG", "REAC",␣
      ↪"HP", "CORUM"))
```

	query	significant	p_value	term_size	query_size	intersection_size	precisio
	<chr>	<lgl>	<dbl>	<int>	<int>	<int>	<dbl>
A data.frame: 5 × 14	query_1	TRUE	0.003447673	175	427	21	0.04918
	query_1	TRUE	0.024203252	155	427	18	0.04215
	query_1	TRUE	0.024629586	306	321	26	0.08099
	query_1	TRUE	0.043721843	249	321	22	0.06853
	query_1	TRUE	0.049705759	3	178	3	0.01685

1.2.4 Shared si-eQTL (eGenes)

```
[13]: gostres <- gost(query=shared, organism="hsapiens")
gostres$result %>%
    data.table::fwrite(file = "shared_functional_enrichment.txt", sep="\t")

p <- gostplot(gostres, capped = FALSE, interactive = FALSE)
print(p)
save_ggplots("shared_manhattan", p, 9, 5)
```

query_1

-log10(p-adj)

GO:MF   GO:CC   GO:BP   KEGG REAC   TF   MIRNA HPA   HP   WP
                                              CORUM

[14]: `gostres$result %>% arrange(p_value) %>% filter(source == "HPA") %>% head`

A data.frame: 6 × 14

	query <chr>	significant <lgl>	p_value <dbl>	term_size <int>	query_size <int>	intersection_size <int>	pre <d
1	query_1	TRUE	8.630194e-14	8005	465	412	0.88
2	query_1	TRUE	1.143182e-11	7114	465	376	0.80
3	query_1	TRUE	2.448280e-10	6734	465	358	0.70
4	query_1	TRUE	3.066105e-10	6874	465	363	0.78
5	query_1	TRUE	1.146056e-07	6624	465	345	0.74
6	query_1	TRUE	1.429749e-07	4948	465	276	0.59

[15]: `gostres$result %>% arrange(p_value) %>% filter(source %in% c("KEGG", "REAC",`
`↪"HP", "CORUM"))`

7

		query	significant	p_value	term_size	query_size	intersection_size	precisio
		<chr>	<lgl>	<dbl>	<int>	<int>	<int>	<dbl>
A data.frame: 5 × 14		query_1	TRUE	0.002387899	175	417	21	0.05035
		query_1	TRUE	0.017302403	306	314	26	0.08280
		query_1	TRUE	0.032181657	249	314	22	0.07006
		query_1	TRUE	0.049307318	84	417	12	0.02877
		query_1	TRUE	0.049773371	3	175	3	0.01714

1.3 Session Info

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

```
[1] "2021-09-14 08:11:02 EDT"

   user  system elapsed
 11.510   0.448  26.075

 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     2021-09-14

 Packages
 package     * version  date       lib source
 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)
 bitops        1.0-7    2021-04-24 [1] CRAN (R 4.0.3)
 broom         0.7.9    2021-07-27 [1] CRAN (R 4.0.3)
 Cairo         1.5-12.2 2020-07-07 [1] CRAN (R 4.0.2)
 cellranger    1.1.0    2016-07-27 [1] CRAN (R 4.0.2)
 cli           3.0.1    2021-07-17 [1] CRAN (R 4.0.3)
 colorspace    2.0-2    2021-06-24 [1] CRAN (R 4.0.3)
 crayon        1.4.1    2021-02-08 [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)
fastmap        1.1.0     2021-01-25 [1] CRAN (R 4.0.2)
forcats      * 0.5.1     2021-01-27 [1] CRAN (R 4.0.2)
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)
glue           1.4.2     2020-08-27 [1] CRAN (R 4.0.2)
gprofiler2   * 0.2.1     2021-08-23 [1] CRAN (R 4.0.3)
gtable         0.3.0     2019-03-25 [1] CRAN (R 4.0.2)
haven          2.4.3     2021-08-04 [1] CRAN (R 4.0.3)
hms            1.1.0     2021-05-17 [1] CRAN (R 4.0.3)
htmltools      0.5.2     2021-08-25 [1] CRAN (R 4.0.3)
htmlwidgets    1.5.3     2020-12-10 [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)
lazyeval       0.2.2     2019-03-15 [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)
pbdZMQ         0.3-5     2021-02-10 [1] CRAN (R 4.0.3)
pillar         1.6.2     2021-07-29 [1] CRAN (R 4.0.3)
pkgconfig      2.0.3     2019-09-22 [1] CRAN (R 4.0.2)
plotly         4.9.4.1   2021-06-18 [1] CRAN (R 4.0.3)
purrr        * 0.3.4     2020-04-17 [1] CRAN (R 4.0.2)
R.methodsS3    1.8.1     2020-08-26 [1] CRAN (R 4.0.3)
R.oo           1.24.0    2020-08-26 [1] CRAN (R 4.0.3)
R.utils        2.10.1    2020-08-26 [1] CRAN (R 4.0.3)
R6             2.5.1     2021-08-19 [1] CRAN (R 4.0.3)
Rcpp           1.0.7     2021-07-07 [1] CRAN (R 4.0.3)
RCurl          1.98-1.4  2021-08-17 [1] CRAN (R 4.0.3)
readr        * 2.0.1     2021-08-10 [1] CRAN (R 4.0.3)
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         2.0.1     2021-08-05 [1] CRAN (R 4.0.3)
rlang          0.4.11    2021-04-30 [1] CRAN (R 4.0.3)
rstudioapi     0.13      2020-11-12 [1] CRAN (R 4.0.2)
rvest          1.0.1     2021-07-26 [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        1.7.4     2021-08-25 [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       * 3.1.4    2021-08-25 [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)
tzdb           0.1.2    2021-07-20 [1] CRAN (R 4.0.3)
utf8           1.2.2    2021-07-24 [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)
viridisLite    0.4.0    2021-04-13 [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)

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