July 19, 2022

The results below are generated from an R script.

```
library(tidyverse)

## - Attaching packages ------------------------- tidyverse 1.3.1 -
## v ggplot2 3.3.6     v purrr   0.3.4
## v tibble  3.1.7     v dplyr   1.0.9
## v tidyr   1.2.0     v stringr 1.4.0
## v readr   2.1.2     v forcats 0.5.1
## - Conflicts -------------------------- tidyverse_conflicts() -
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()

library(lubridate)

##
## Attaching package:  'lubridate'
## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union

library(ggplot2)
library(dplyr)
library(skimr)

##Reading the data set
bike_data <- read.csv("total_bike_ride_data.csv")
head(bike_data)

##   rideable_type              started_at                ended_at member_casual start_lat
## 1   docked_bike 2021-10-02 22:26:49 UTC 2021-10-13 10:55:16 UTC        casual  41.85756
## 2   docked_bike 2021-12-26 15:44:59 UTC 2021-12-26 16:21:36 UTC        casual  41.87277
## 3   docked_bike 2021-12-28 18:42:45 UTC 2021-12-28 19:01:27 UTC        casual  41.79664
## 4   docked_bike 2021-12-20 01:19:33 UTC 2021-12-20 01:44:14 UTC        casual  41.89918
## 5   docked_bike 2021-12-31 23:40:47 UTC 2022-01-01 00:26:06 UTC        casual  41.88398
## 6   docked_bike 2021-12-03 07:27:31 UTC 2021-12-03 07:40:22 UTC        casual  41.89435
##   start_lng  end_lat   end_lng
## 1 -87.66154       NA        NA
## 2 -87.62398 41.88398 -87.62468
## 3 -87.62592 41.79949 -87.58645
## 4 -87.67220 41.90327 -87.67843
## 5 -87.62468 41.88103 -87.62408
## 6 -87.62280 41.90799 -87.63150

##Structure check
str(bike_data)
```

```
## 'data.frame': 5860776 obs. of  8 variables:
##  $ rideable_type: chr  "docked_bike" "docked_bike" "docked_bike" "docked_bike" ...
##  $ started_at   : chr  "2021-10-02 22:26:49 UTC" "2021-12-26 15:44:59 UTC" "2021-12-28 18:42:45 UTC"
##  $ ended_at     : chr  "2021-10-13 10:55:16 UTC" "2021-12-26 16:21:36 UTC" "2021-12-28 19:01:27 UTC"
##  $ member_casual: chr  "casual" "casual" "casual" "casual" ...
##  $ start_lat    : num  41.9 41.9 41.8 41.9 41.9 ...
##  $ start_lng    : num  -87.7 -87.6 -87.6 -87.7 -87.6 ...
##  $ end_lat      : num  NA 41.9 41.8 41.9 41.9 ...
##  $ end_lng      : num  NA -87.6 -87.6 -87.7 -87.6 ...

##Data Summary
skim(bike_data)

## Error in kable_latex(x = structure(c("Name", "Number of rows", "Number of columns", :  unused
argument (table.attr = "style='width:  auto;'\n      class='table table-condensed'")

##New column called ride_date and changed its data type
bike_data$ride_date<-as.Date(bike_data$started_at)
##Calculating ride durations
bike_data$started_at<-as_datetime(bike_data$started_at)
bike_data$ended_at<-as_datetime(bike_data$ended_at)

##Creating day , week , month , year output out of the current data
bike_data$month <- strftime(bike_data$ride_date, "%B")
bike_data$day <- strftime(bike_data$ride_date, "%d")
bike_data$year <-strftime(bike_data$ride_date, "%Y")
bike_data$day_of_week <- strftime(bike_data$ride_date, "%A")

##checking column names & data summary
colnames(bike_data)

##  [1] "rideable_type" "started_at"    "ended_at"      "member_casual" "start_lat"
##  [6] "start_lng"     "end_lat"       "end_lng"       "ride_date"     "month"
## [11] "day"           "year"          "day_of_week"

skim(bike_data)

## Error in kable_latex(x = structure(c("Name", "Number of rows", "Number of columns", :  unused
argument (table.attr = "style='width:  auto;'\n      class='table table-condensed'")

##creating new column called lenght of ride
bike_data$length_of_ride=difftime(bike_data$ended_at,bike_data$started_at)

##check
summary(bike_data$length_of_ride)

##   Length    Class     Mode
##  5860776 difftime  numeric

##changing data format into numeric
bike_data$length_of_ride=as.numeric(bike_data$length_of_ride)

##filtering out 0 seconds long rides
bike_data_2<-filter(bike_data,length_of_ride>0)

bike_data_2<-filter(bike_data,rideable_type!="docked_bike")
```

```
##min and max rides (in seconds)
min(bike_data_2$length_of_ride)

## [1] -3482

max(bike_data_2$length_of_ride)

## [1] 93596

## Average length of ride (in seconds)
bike_data_2%>%summarise(average_length_ride=mean(length_of_ride))

##   average_length_ride
## 1            1057.317

##length of ride by user type (in seconds)
aggregate(bike_data_2$length_of_ride~bike_data_2$member_casual,FUN=mean)

##   bike_data_2$member_casual bike_data_2$length_of_ride
## 1                    casual                  1454.0646
## 2                    member                   782.6263

aggregate(bike_data_2$length_of_ride~bike_data_2$member_casual,FUN=median)

##   bike_data_2$member_casual bike_data_2$length_of_ride
## 1                    casual                        855
## 2                    member                        547

aggregate(bike_data_2$length_of_ride~bike_data_2$member_casual,FUN=max)

##   bike_data_2$member_casual bike_data_2$length_of_ride
## 1                    casual                      93596
## 2                    member                      93594

##Mean length of ride by member type and day of week
aggregate(bike_data_2$length_of_ride~bike_data_2$member_casual+bike_data_2$day_of_week,FUN=mean)

##    bike_data_2$member_casual bike_data_2$day_of_week bike_data_2$length_of_ride
## 1                     casual                  Friday                  1372.0041
## 2                     member                  Friday                   766.9339
## 3                     casual                  Monday                  1463.0431
## 4                     member                  Monday                   758.7097
## 5                     casual                Saturday                  1593.8579
## 6                     member                Saturday                   877.4110
## 7                     casual                  Sunday                  1660.0822
## 8                     member                  Sunday                   887.4403
## 9                     casual                Thursday                  1310.1854
## 10                    member                Thursday                   746.5643
## 11                    casual                 Tuesday                  1295.3247
## 12                    member                 Tuesday                   736.7129
## 13                    casual               Wednesday                  1278.9822
## 14                    member               Wednesday                   738.5013

##Max length of ride by member type and weekday
aggregate(bike_data_2$length_of_ride~bike_data_2$member_casual+bike_data_2$day_of_week,FUN=max)
```

```
##    bike_data_2$member_casual bike_data_2$day_of_week bike_data_2$length_of_ride
## 1                     casual                  Friday                      90025
## 2                     member                  Friday                      89998
## 3                     casual                  Monday                      89997
## 4                     member                  Monday                      89997
## 5                     casual                Saturday                      93596
## 6                     member                Saturday                      93594
## 7                     casual                  Sunday                      90032
## 8                     member                  Sunday                      89996
## 9                     casual                Thursday                      90027
## 10                    member                Thursday                      89997
## 11                    casual                 Tuesday                      90027
## 12                    member                 Tuesday                      89997
## 13                    casual               Wednesday                      89997
## 14                    member               Wednesday                      89998
```

```
##Mean ride length by rider type and months
aggregate(bike_data_2$length_of_ride~bike_data_2$member_casual+bike_data_2$month,FUN=mean)
```

```
##    bike_data_2$member_casual bike_data_2$month bike_data_2$length_of_ride
## 1                     casual             April                  1367.4960
## 2                     member             April                   689.5442
## 3                     casual            August                  1503.1584
## 4                     member            August                   846.0487
## 5                     casual          December                  1096.0712
## 6                     member          December                   660.2963
## 7                     casual          February                  1245.5408
## 8                     member          February                   684.3318
## 9                     casual           January                  1189.9647
## 10                    member           January                   718.8906
## 11                    casual              July                  1543.3992
## 12                    member              July                   854.3591
## 13                    casual              June                  1622.5057
## 14                    member              June                   880.6672
## 15                    casual             March                  1429.2616
## 16                    member             March                   717.4997
## 17                    casual               May                  1494.8911
## 18                    member               May                   802.0006
## 19                    casual          November                  1103.2113
## 20                    member          November                   678.2838
## 21                    casual           October                  1312.2789
## 22                    member           October                   750.1241
## 23                    casual         September                  1440.4613
## 24                    member         September                   824.0674
```

```
##count of total bike rides per month by rider type
bike_data_2%>%count(month,member_casual)
```

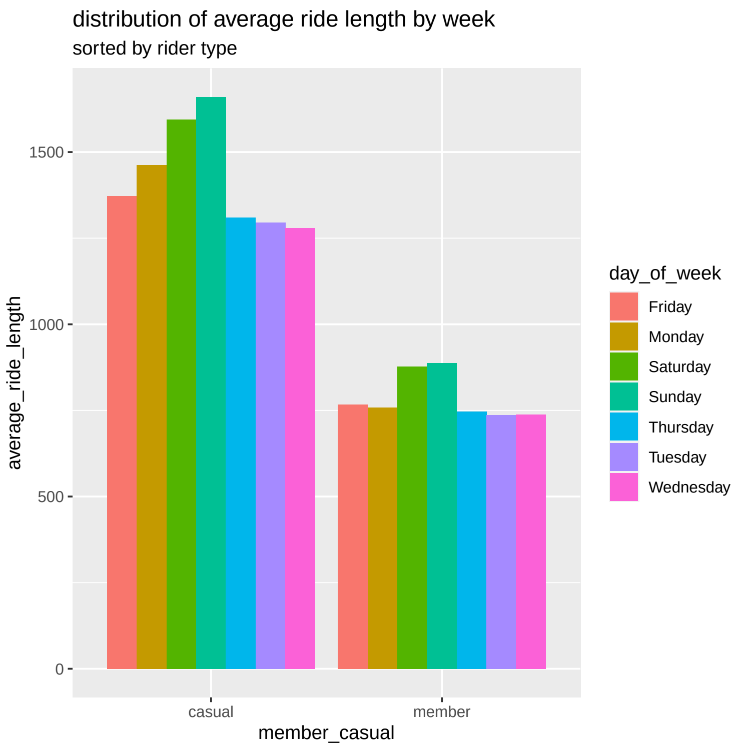
```
##        month member_casual      n
## 1      April        casual 114301
## 2      April        member 244832
## 3     August        casual 367606
## 4     August        member 391681
## 5   December        casual  64810
## 6   December        member 177802
## 7   February        casual  20055
## 8   February        member  94193
## 9    January        casual  17559
## 10   January        member  85250
## 11      July        casual 384358
## 12      July        member 380354
## 13      June        casual 318965
## 14      June        member 358914
## 15     March        casual  81524
## 16     March        member 194160
## 17       May        casual 254006
## 18       May        member 354443
## 19  November        casual  99315
## 20  November        member 253049
## 21   October        casual 234358
## 22   October        member 373984
## 23 September        casual 328553
## 24 September        member 392257
```

```
##usage of each bike type by rider type
bike_data_2%>%count(member_casual,rideable_type)
```

```
##   member_casual rideable_type       n
## 1        casual  classic_bike 1236535
## 2        casual electric_bike 1048875
## 3        member  classic_bike 1981202
## 4        member electric_bike 1319717
```

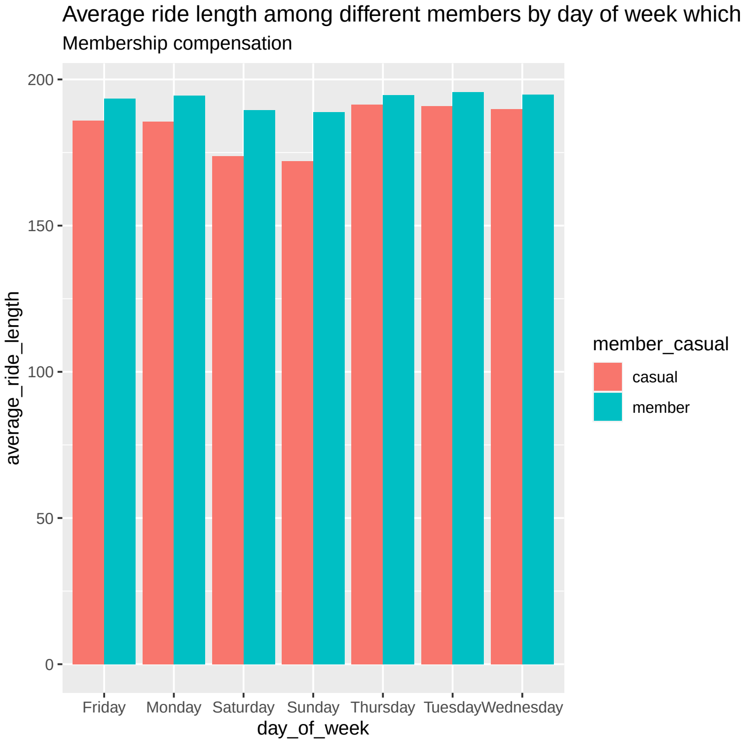
```
##Visualization

##Average ride length per ride type and day
bike_data_2%>%group_by(member_casual,day_of_week)%>%summarise(average_ride_length=mean(length_of_ride))%
  ggplot(aes(x=member_casual,y=average_ride_length,fill=day_of_week))+
  geom_bar(position="Dodge",stat="identity")+
  labs(title="distribution of average ride length by week",subtitle="sorted by rider type")
```

```
## `summarise()` has grouped output by 'member_casual'.  You can override using the `.groups`
## argument.
```



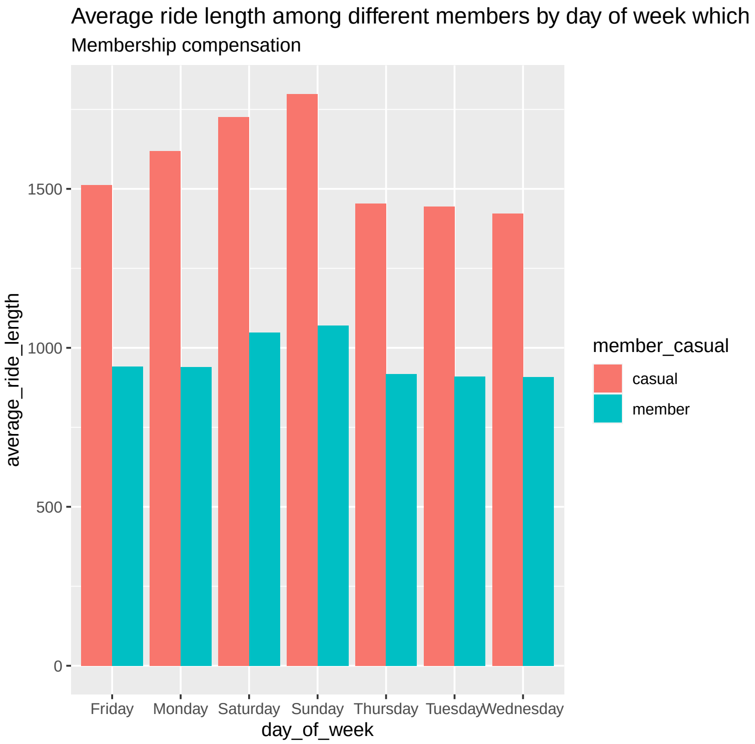
```
##Trips less than 5 minutes
bike_data_2%>%group_by(day_of_week,member_casual)%>%filter(length_of_ride<300)%>%summarise(average_ride_
  ggplot(aes(x=day_of_week,y=average_ride_length,fill=member_casual))+
  geom_bar(position='Dodge',stat='identity')+
  labs(title="Average ride length among different members by day of week which is less than 5 mins",subt
```

```
## `summarise()` has grouped output by 'day_of_week'.  You can override using the `.groups`
## argument.
```

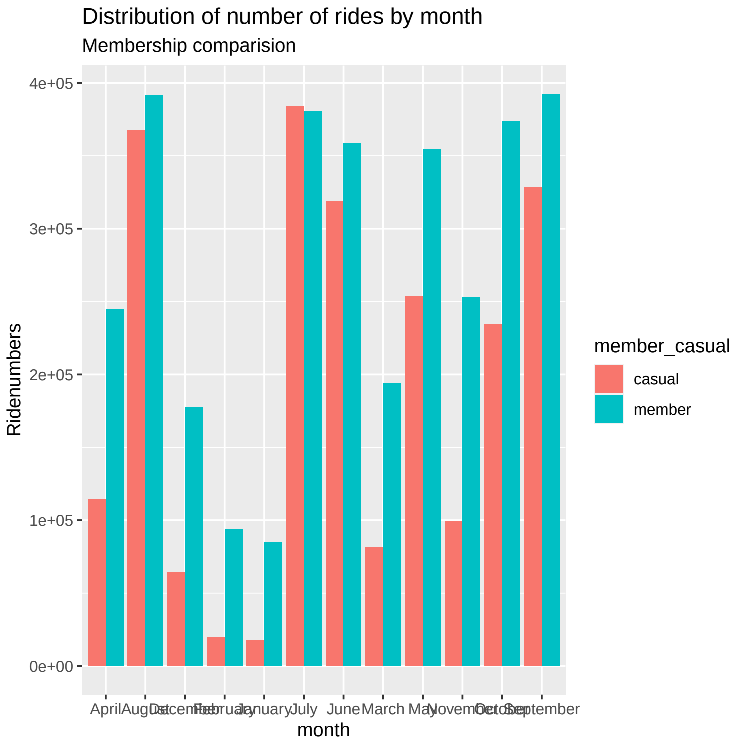


```
##Trips more than 5 mins
bike_data_2%>%group_by(day_of_week,member_casual)%>%filter(length_of_ride>300)%>%summarise(average_ride_
  ggplot(aes(x=day_of_week,y=average_ride_length,fill=member_casual))+
  geom_bar(position='Dodge',stat='identity')+
  labs(title="Average ride length among different members by day of week which is more than 5 mins",subt
```

```
## `summarise()` has grouped output by 'day_of_week'.  You can override using the `.groups`
## argument.
```



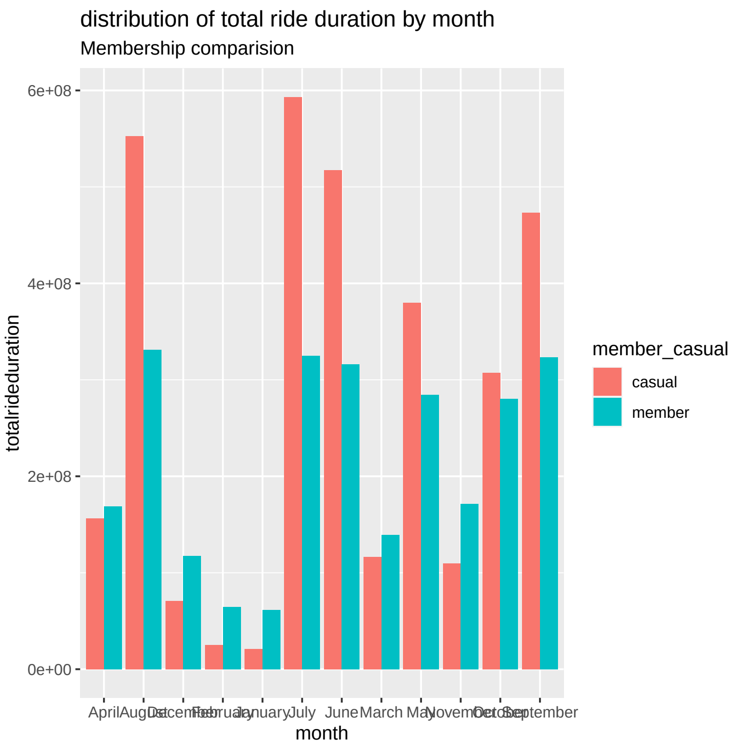
```
##Total Rides per month by rider type
bike_data_2%>%group_by(month,member_casual)%>%summarise(Ridenumbers=n())%>%
  ggplot(aes(x=month,y=Ridenumbers,fill=member_casual))+
  geom_bar(position='Dodge',stat='identity')+
  labs(title="Distribution of number of rides by month",subtitle="Membership comparision")

## `summarise()` has grouped output by 'month'.  You can override using the `.groups`
## argument.
```



Distribution of number of rides by month

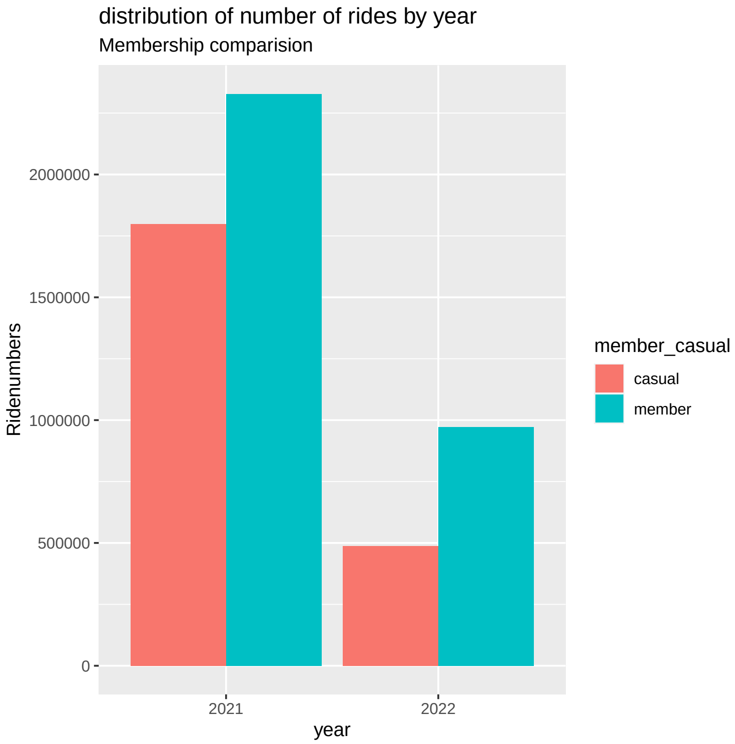
Membership comparision

```
##Total length of Ride per month by rider types
bike_data_2%>%group_by(month,member_casual)%>%summarise(totalrideduration=sum(length_of_ride))%>%
  ggplot(aes(x=month,y=totalrideduration,fill=member_casual))+
  geom_bar(position='Dodge',stat='identity')+
  labs(title="distribution of total ride duration by month",subtitle="Membership comparision")

## `summarise()` has grouped output by 'month'.  You can override using the `.groups`
## argument.
```



```
##Total Rides per year by rider type
bike_data_2%>%group_by(year,member_casual)%>%summarise(Ridenumbers=n())%>%
  ggplot(aes(x=year,y=Ridenumbers, fill=member_casual)) +
  geom_bar(position='Dodge',stat='identity') +
  labs(title="distribution of number of rides by year",subtitle="Membership comparision")
```

```
## `summarise()` has grouped output by 'year'.  You can override using the `.groups`
## argument.
```



The R session information (including the OS info, R version and all packages used):

```
sessionInfo()
```

```
## R version 4.2.1 (2022-06-23 ucrt)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 19044)
##
## Matrix products: default
##
## locale:
## [1] LC_COLLATE=English_Canada.utf8  LC_CTYPE=English_Canada.utf8
## [3] LC_MONETARY=English_Canada.utf8 LC_NUMERIC=C
## [5] LC_TIME=English_Canada.utf8
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
##  [1] skimr_2.1.4     lubridate_1.8.0 forcats_0.5.1   stringr_1.4.0   dplyr_1.0.9
##  [6] purrr_0.3.4     readr_2.1.2     tidyr_1.2.0     tibble_3.1.7    ggplot2_3.3.6
## [11] tidyverse_1.3.1
##
## loaded via a namespace (and not attached):
##  [1] tidyselect_1.1.2 xfun_0.31        repr_1.1.4       haven_2.5.0      colorspace_2.0-3
##  [6] vctrs_0.4.1      generics_0.1.3   htmltools_0.5.2  base64enc_0.1-3  utf8_1.2.2
## [11] rlang_1.0.2      pillar_1.8.0     glue_1.6.2       withr_2.5.0      DBI_1.1.3
## [16] dbplyr_2.2.1     modelr_0.1.8     readxl_1.4.0     lifecycle_1.0.1  munsell_0.5.0
## [21] gtable_0.3.0     cellranger_1.1.0 rvest_1.0.2      evaluate_0.15    labeling_0.4.2
## [26] knitr_1.39       tzdb_0.3.0       fastmap_1.1.0    fansi_1.0.3      highr_0.9
## [31] broom_1.0.0      scales_1.2.0     backports_1.4.1  jsonlite_1.8.0   farver_2.1.0
## [36] fs_1.5.2         hms_1.1.1        digest_0.6.29    stringi_1.7.6    grid_4.2.1
## [41] cli_3.3.0        tools_4.2.1      magrittr_2.0.3   crayon_1.5.1     pkgconfig_2.0.3
## [46] ellipsis_0.3.2   xml2_1.3.3       reprex_2.0.1     assertthat_0.2.1 httr_1.4.3
## [51] rstudioapi_0.13  R6_2.5.1         compiler_4.2.1
```

```
Sys.time()
```

```
## [1] "2022-07-19 11:25:39 PDT"
```