Cyclists

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
install.packages('plyr', repos = "http://cran.us.r-project.org")
```

```
## Installing package into 'C:/Users/mucho/Documents/R/win-library/4.0'
## (as 'lib' is unspecified)

##
##   There is a binary version available but the source version is later:
##      binary source needs_compilation
## plyr  1.8.7  1.8.8              TRUE
##
##   Binaries will be installed
## package 'plyr' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\mucho\AppData\Local\Temp\RtmpOguNr5\downloaded_packages
```

```
install.packages("tidyverse",repos = "https://cran.rstudio.com")
```

```
## Installing package into 'C:/Users/mucho/Documents/R/win-library/4.0'
## (as 'lib' is unspecified)

##
##   There is a binary version available but the source version is later:
##           binary source needs_compilation
## tidyverse  1.3.1  1.3.2             FALSE

## installing the source package 'tidyverse'
```

```
install.packages("DataExplorer",repos = "https://cran.rstudio.com")
```

```
## Installing package into 'C:/Users/mucho/Documents/R/win-library/4.0'
## (as 'lib' is unspecified)

## package 'DataExplorer' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\mucho\AppData\Local\Temp\RtmpOguNr5\downloaded_packages
```

```
install.packages("janitor",repos = "https://cran.rstudio.com")
```

```
## Installing package into 'C:/Users/mucho/Documents/R/win-library/4.0'
## (as 'lib' is unspecified)

##
##   There is a binary version available but the source version is later:
##         binary source needs_compilation
## janitor  2.1.0  2.2.0             FALSE

## installing the source package 'janitor'
```

```
install.packages("lubridate",repos = "https://cran.rstudio.com")
```

```
## Installing package into 'C:/Users/mucho/Documents/R/win-library/4.0'
## (as 'lib' is unspecified)

##
##   There is a binary version available but the source version is later:
##           binary source needs_compilation
## lubridate  1.8.0  1.9.1              TRUE
##
##   Binaries will be installed
## package 'lubridate' successfully unpacked and MD5 sums checked

## Warning: cannot remove prior installation of package 'lubridate'

## Warning in file.copy(savedcopy, lib, recursive = TRUE): problem copying C:
## \Users\mucho\Documents\R\win-
library\4.0\00LOCK\lubridate\libs\x64\lubridate.dll
## to C:\Users\mucho\Documents\R\win-
library\4.0\lubridate\libs\x64\lubridate.dll:
## Permission denied

## Warning: restored 'lubridate'

##
## The downloaded binary packages are in
##  C:\Users\mucho\AppData\Local\Temp\RtmpOguNr5\downloaded_packages
```

```
install.packages("geosphere",repos = "https://cran.rstudio.com")
```

```
## Installing package into 'C:/Users/mucho/Documents/R/win-library/4.0'
## (as 'lib' is unspecified)

##
##   There is a binary version available but the source version is later:
##           binary source needs_compilation
## geosphere 1.5-14 1.5-18              TRUE
##
##   Binaries will be installed
## package 'geosphere' successfully unpacked and MD5 sums checked
```

```
## Warning: cannot remove prior installation of package 'geosphere'

## Warning in file.copy(savedcopy, lib, recursive = TRUE): problem copying C:
## \Users\mucho\Documents\R\win-
library\4.0\00LOCK\geosphere\libs\x64\geosphere.dll
## to C:\Users\mucho\Documents\R\win-
library\4.0\geosphere\libs\x64\geosphere.dll:
## Permission denied

## Warning: restored 'geosphere'

##
## The downloaded binary packages are in
##  C:\Users\mucho\AppData\Local\Temp\RtmpOguNr5\downloaded_packages

install.packages('rmarkdown',repos = "https://cran.rstudio.com")

## Installing package into 'C:/Users/mucho/Documents/R/win-library/4.0'
## (as 'lib' is unspecified)

##
##   There is a binary version available but the source version is later:
##          binary source needs_compilation
## rmarkdown   2.13   2.20             FALSE

## installing the source package 'rmarkdown'

library(tidyverse)

## Warning: replacing previous import 'lifecycle::last_warnings' by
## 'rlang::last_warnings' when loading 'pillar'

## -- Attaching packages --------------------------------------- tidyverse
1.3.2 --

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.6     v dplyr   1.0.7
## v tidyr   1.2.0     v stringr 1.4.0
## v readr   2.1.2     v forcats 0.5.1

## Warning: package 'ggplot2' was built under R version 4.0.5

## Warning: package 'tibble' was built under R version 4.0.5

## Warning: package 'tidyr' was built under R version 4.0.5

## Warning: package 'readr' was built under R version 4.0.5

## Warning: package 'purrr' was built under R version 4.0.5

## Warning: package 'dplyr' was built under R version 4.0.5

## Warning: package 'stringr' was built under R version 4.0.5
```

```
## Warning: package 'forcats' was built under R version 4.0.5

## -- Conflicts ---------------------------------------
tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()

library(DataExplorer)

## Warning: package 'DataExplorer' was built under R version 4.0.5

library(janitor)

##
## Attaching package: 'janitor'
##
## The following objects are masked from 'package:stats':
##
##     chisq.test, fisher.test

library(lubridate)

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

library(ggplot2)
library(dplyr)
library(geosphere)

# loading Data

library(tidyverse)
library(DataExplorer)
library(janitor)
library(lubridate)
library(ggplot2)
library(dplyr)
library(geosphere)
path <- "C:/Users/Public/Google Data Analytics/Capstone/Track 1/XSL_file"
cyclistic_2021<-
read.csv(paste(path,"/2021/First_three_month_2021.csv",sep=''))
cyclistic_2022<-
read.csv(paste(path,"/2022/First_three_month_2022.csv",sep=''))


#Initial Visualization
#For 2021 and 2022
```

```
summary(cyclistic_2021) #statistical summary of data 2021
```

```
##    ride_id          rideable_type       started_at          ended_at
##  Length:594331      Length:594331      Length:594331      Length:594331
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##  start_station_name start_station_id   end_station_name   end_station_id
##  Length:594331      Length:594331      Length:594331      Length:594331
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##    start_lat       start_lng         end_lat         end_lng
##  Min.   :41.64   Min.   :-87.83   Min.   :41.54   Min.   :-88.07
##  1st Qu.:41.88   1st Qu.:-87.66   1st Qu.:41.88   1st Qu.:-87.66
##  Median :41.90   Median :-87.64   Median :41.90   Median :-87.64
##  Mean   :41.90   Mean   :-87.65   Mean   :41.90   Mean   :-87.65
##  3rd Qu.:41.93   3rd Qu.:-87.63   3rd Qu.:41.93   3rd Qu.:-87.63
##  Max.   :45.64   Max.   :-73.80   Max.   :42.08   Max.   :-87.51
##                                   NA's   :647     NA's   :647
##  member_casual
##  Length:594331
##  Class :character
##  Mode  :character
##
##
##
##
```

```
summary(cyclistic_2022) #statistical summary of data 2022
```

```
##    ride_id          rideable_type       started_at          ended_at
##  Length:503421      Length:503421      Length:503421      Length:503421
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##  start_station_name start_station_id   end_station_name   end_station_id
##  Length:503421      Length:503421      Length:503421      Length:503421
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
```

```
##
##
##
##
##    start_lat       start_lng         end_lat         end_lng
##  Min.   :41.65   Min.   :-87.84   Min.   :41.64   Min.   :-87.84
##  1st Qu.:41.88   1st Qu.:-87.66   1st Qu.:41.88   1st Qu.:-87.66
##  Median :41.89   Median :-87.64   Median :41.89   Median :-87.64
##  Mean   :41.90   Mean   :-87.65   Mean   :41.90   Mean   :-87.65
##  3rd Qu.:41.93   3rd Qu.:-87.63   3rd Qu.:41.93   3rd Qu.:-87.63
##  Max.   :45.64   Max.   :-73.80   Max.   :42.07   Max.   :-87.51
##                                   NA's   :429     NA's   :429
##  member_casual
##  Length:503421
##  Class :character
##  Mode  :character
##
##
##
##
```

```
# Cleaning and Preparing Data

sum(is.na(cyclistic_2021))
```

```
## [1] 1294
```

```
sum(is.na(cyclistic_2022))
```

```
## [1] 858
```

```
cyclistic_2021<- na.omit(cyclistic_2021)
cyclistic_2022<- na.omit(cyclistic_2022)


# Create columns as follows: date, month weekday and season

cyclistic_2021$date <- as.Date(cyclistic_2021$started_at)
cyclistic_2021$weekday <- format(as.Date(cyclistic_2021$date), "%A")

cyclistic_2022$date <- as.Date(cyclistic_2022$started_at)
cyclistic_2022$weekday <- format(as.Date(cyclistic_2022$date), "%A")

#riprawdata$season <- ifelse (triprawdata$month %in% c('06','07','08'),
"Summer",
# ifelse (triprawdata$month %in% c('09','10','11'), "Fall",
# ifelse (triprawdata$month %in% c('12','01','02'), "Winter",
#      ifelse (triprawdata$month %in% c('03','04','05'), "Spring", NA))))
```

```
# Checking the new created column and summary
head(cyclistic_2021)

##            ride_id rideable_type          started_at            ended_at
## 1 C2F7DD78E82EC875 electric_bike 2022-01-13 11:59:47 2022-01-13 12:02:44
## 2 A6CF8980A652D272 electric_bike 2022-01-10 08:41:56 2022-01-10 08:46:17
## 3 BD0F91DFF741C66D  classic_bike 2022-01-25 04:53:40 2022-01-25 04:58:01
## 4 CBB80ED419105406  classic_bike 2022-01-04 00:18:04 2022-01-04 00:33:00
## 5 DDC963BFDDA51EEA  classic_bike 2022-01-20 01:31:10 2022-01-20 01:37:12
## 6 A39C6F6CC0586C0B  classic_bike 2022-01-11 18:48:09 2022-01-11 18:51:31
##              start_station_name start_station_id
end_station_name
## 1      Glenwood Ave & Touhy Ave              525          Clark St & Touhy
Ave
## 2      Glenwood Ave & Touhy Ave              525          Clark St & Touhy
Ave
## 3 Sheffield Ave & Fullerton Ave     TA1306000016 Greenview Ave & Fullerton
Ave
## 4      Clark St & Bryn Mawr Ave     KA1504000151     Paulina St & Montrose
Ave
## 5   Michigan Ave & Jackson Blvd     TA1309000002        State St &
Randolph St
## 6         Wood St & Chicago Ave              637       Honore St &
Division St
##   end_station_id start_lat start_lng  end_lat   end_lng member_casual
## 1         RP-007  42.01280 -87.66591 42.01256 -87.67437        casual
## 2         RP-007  42.01276 -87.66597 42.01256 -87.67437        casual
## 3   TA1307000001  41.92560 -87.65371 41.92533 -87.66580        member
## 4   TA1309000021  41.98359 -87.66915 41.96151 -87.67139        casual
## 5   TA1305000029  41.87785 -87.62408 41.88462 -87.62783        member
## 6   TA1305000034  41.89563 -87.67207 41.90312 -87.67394        member
##         date  weekday
## 1 2022-01-13 Thursday
## 2 2022-01-10   Monday
## 3 2022-01-25  Tuesday
## 4 2022-01-04  Tuesday
## 5 2022-01-20 Thursday
## 6 2022-01-11  Tuesday

summary(cyclistic_2021)

##    ride_id          rideable_type       started_at          ended_at
##  Length:593684      Length:593684      Length:593684      Length:593684
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##  start_station_name start_station_id   end_station_name   end_station_id
##  Length:593684      Length:593684      Length:593684      Length:593684
```

```
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##    start_lat       start_lng         end_lat         end_lng
##  Min.   :41.64   Min.   :-87.83   Min.   :41.54   Min.   :-88.07
##  1st Qu.:41.88   1st Qu.:-87.66   1st Qu.:41.88   1st Qu.:-87.66
##  Median :41.90   Median :-87.64   Median :41.90   Median :-87.64
##  Mean   :41.90   Mean   :-87.65   Mean   :41.90   Mean   :-87.65
##  3rd Qu.:41.93   3rd Qu.:-87.63   3rd Qu.:41.93   3rd Qu.:-87.63
##  Max.   :45.64   Max.   :-73.80   Max.   :42.08   Max.   :-87.51
##  member_casual          date              weekday
##  Length:593684      Min.   :2021-01-01   Length:593684
##  Class :character   1st Qu.:2021-03-01   Class :character
##  Mode  :character   Median :2021-03-22   Mode  :character
##                     Mean   :2021-07-01
##                     3rd Qu.:2022-01-20
##                     Max.   :2022-02-28

head(cyclistic_2022)

##            ride_id rideable_type          started_at            ended_at
## 1 C2F7DD78E82EC875 electric_bike 2022-01-13 11:59:47 2022-01-13 12:02:44
## 2 A6CF8980A652D272 electric_bike 2022-01-10 08:41:56 2022-01-10 08:46:17
## 3 BD0F91DFF741C66D  classic_bike 2022-01-25 04:53:40 2022-01-25 04:58:01
## 4 CBB80ED419105406  classic_bike 2022-01-04 00:18:04 2022-01-04 00:33:00
## 5 DDC963BFDDA51EEA  classic_bike 2022-01-20 01:31:10 2022-01-20 01:37:12
## 6 A39C6F6CC0586C0B  classic_bike 2022-01-11 18:48:09 2022-01-11 18:51:31
##              start_station_name start_station_id
end_station_name
## 1      Glenwood Ave & Touhy Ave              525          Clark St & Touhy
Ave
## 2      Glenwood Ave & Touhy Ave              525          Clark St & Touhy
Ave
## 3 Sheffield Ave & Fullerton Ave     TA1306000016 Greenview Ave & Fullerton
Ave
## 4      Clark St & Bryn Mawr Ave     KA1504000151     Paulina St & Montrose
Ave
## 5   Michigan Ave & Jackson Blvd     TA1309000002        State St &
Randolph St
## 6         Wood St & Chicago Ave              637       Honore St &
Division St
##   end_station_id start_lat start_lng  end_lat   end_lng member_casual
## 1         RP-007  42.01280 -87.66591 42.01256 -87.67437        casual
## 2         RP-007  42.01276 -87.66597 42.01256 -87.67437        casual
## 3   TA1307000001  41.92560 -87.65371 41.92533 -87.66580        member
## 4   TA1309000021  41.98359 -87.66915 41.96151 -87.67139        casual
## 5   TA1305000029  41.87785 -87.62408 41.88462 -87.62783        member
## 6   TA1305000034  41.89563 -87.67207 41.90312 -87.67394        member
```

```
##         date  weekday
## 1 2022-01-13 Thursday
## 2 2022-01-10   Monday
## 3 2022-01-25  Tuesday
## 4 2022-01-04  Tuesday
## 5 2022-01-20 Thursday
## 6 2022-01-11  Tuesday
```

```
summary(cyclistic_2022)
```

```
##    ride_id          rideable_type       started_at          ended_at
##  Length:502992      Length:502992      Length:502992      Length:502992
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##  start_station_name start_station_id   end_station_name   end_station_id
##  Length:502992      Length:502992      Length:502992      Length:502992
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##    start_lat       start_lng         end_lat         end_lng
##  Min.   :41.65   Min.   :-87.84   Min.   :41.64   Min.   :-87.84
##  1st Qu.:41.88   1st Qu.:-87.66   1st Qu.:41.88   1st Qu.:-87.66
##  Median :41.89   Median :-87.64   Median :41.89   Median :-87.64
##  Mean   :41.90   Mean   :-87.65   Mean   :41.90   Mean   :-87.65
##  3rd Qu.:41.93   3rd Qu.:-87.63   3rd Qu.:41.93   3rd Qu.:-87.63
##  Max.   :45.64   Max.   :-73.80   Max.   :42.07   Max.   :-87.51
##  member_casual          date              weekday
##  Length:502992      Min.   :2022-01-01   Length:502992
##  Class :character   1st Qu.:2022-02-08   Class :character
##  Mode  :character   Median :2022-03-04   Mode  :character
##                     Mean   :2022-02-25
##                     3rd Qu.:2022-03-17
##                     Max.   :2022-03-31
```

```
# Checking for Date Consistency

IsDate <- function(mydate,date.format = "%m/%d/%y") {
  tryCatch(!is.na(as.Date(mydate,date.format)),
           error = function(err) {FALSE})
}

z_2021<-IsDate(cyclistic_2021$date)
table(z_2021)["FALSE"]
```

```
## <NA>
##   NA
```

```r
z_2022<-IsDate(cyclistic_2022$date)
table(z_2022)["FALSE"]
```

```
## <NA>
##   NA
```

```r
# Can count both true and false from  table(z) command

#Measuring Ride length and Ride Duration
library(tidyverse)
library(DataExplorer)
library(janitor)
library(lubridate)
library(ggplot2)
library(dplyr)
library(geosphere)
cyclistic_2021$ride_duration <- as.numeric(difftime(cyclistic_2021$ended_at,
cyclistic_2021$started_at))
cyclistic_2021$ride_length <-
distHaversine(cbind(cyclistic_2021$start_lng,cyclistic_2021$start_lat),cbind(
cyclistic_2021$end_lng,cyclistic_2021$end_lat))

cyclistic_2022$ride_duration <- as.numeric(difftime(cyclistic_2022$ended_at,
cyclistic_2022$started_at))
cyclistic_2022$ride_length <-
distHaversine(cbind(cyclistic_2022$start_lng,cyclistic_2022$start_lat),cbind(
cyclistic_2022$end_lng,cyclistic_2022$end_lat))


#Removing negative values of Ride length, Ride Duration, Missing Start and
End Station Names

neg_length<-sign(cyclistic_2021$ride_length)
sum(neg_length<0)
```

```
## [1] 0
```

```r
neg_duration<-sign(cyclistic_2021$ride_duration)
sum(neg_duration<0)
```

```
## [1] 3
```

```r
sum(!is.na(cyclistic_2021$end_station_name)==TRUE)
```

```
## [1] 593684
```

```r
sum(!is.na(cyclistic_2021$end_station_name)==TRUE)
```

```
## [1] 593684
```

```r
sum(cyclistic_2021$end_station_name=="")
```

```
## [1] 69997
```

```
neg_length<-sign(cyclistic_2022$ride_length)
sum(neg_length<0)
```

```
## [1] 0
```

```
neg_duration<-sign(cyclistic_2022$ride_duration)
sum(neg_duration<0)
```

```
## [1] 2
```

```
sum(!is.na(cyclistic_2022$end_station_name)==TRUE)
```

```
## [1] 502992
```

```
sum(!is.na(cyclistic_2022$end_station_name)==TRUE)
```

```
## [1] 502992
```

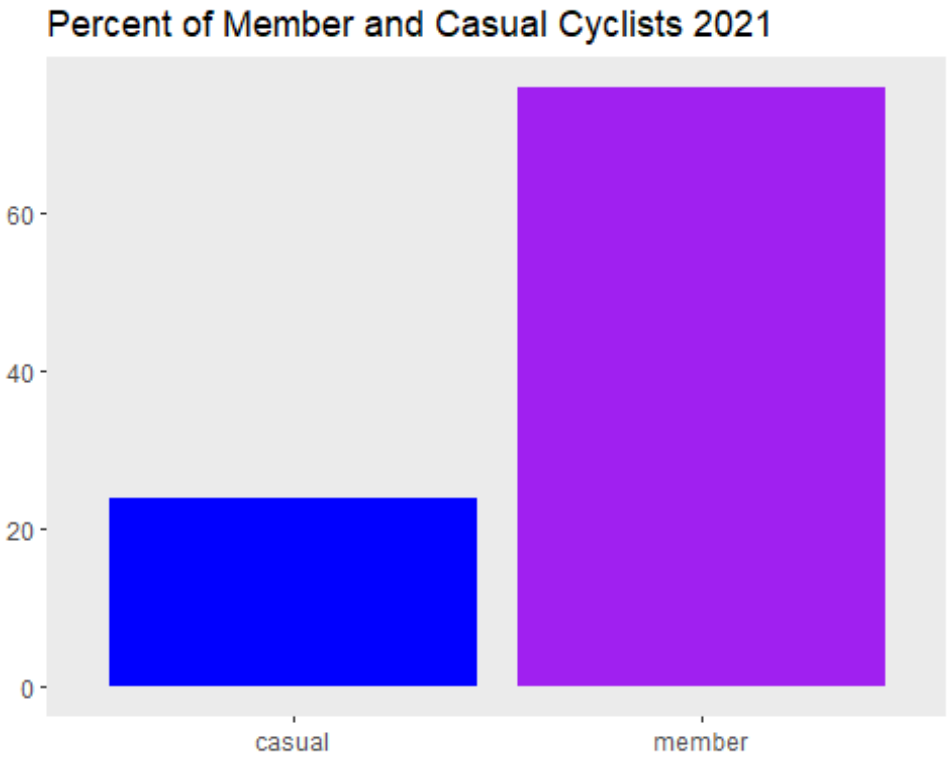
```
sum(cyclistic_2022$end_station_name=="")
```

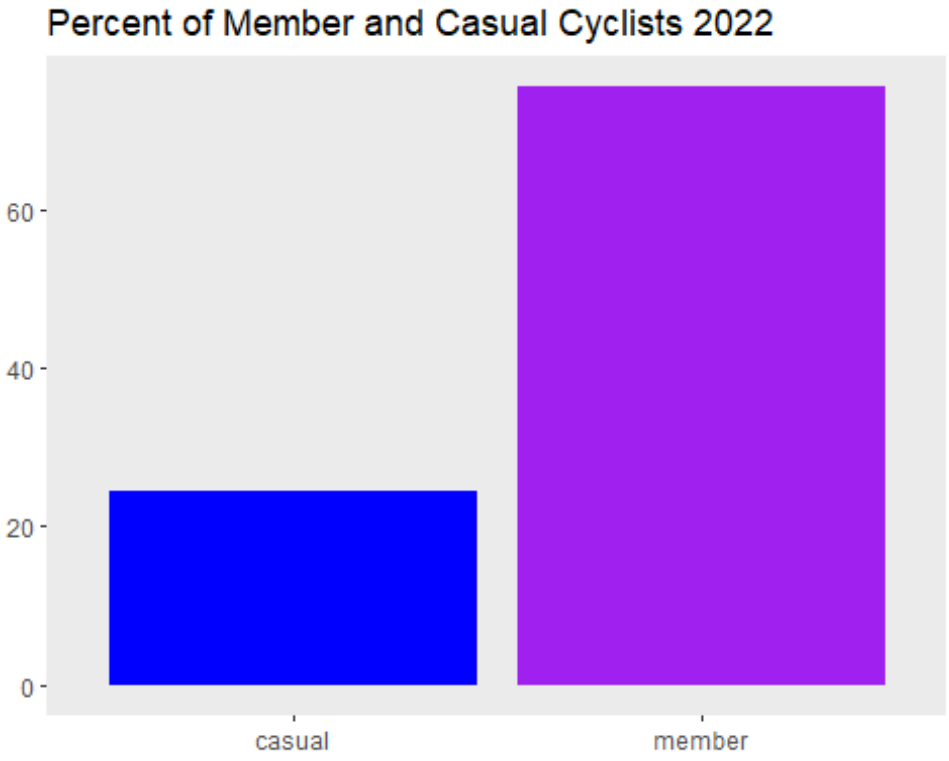
```
## [1] 89010
```

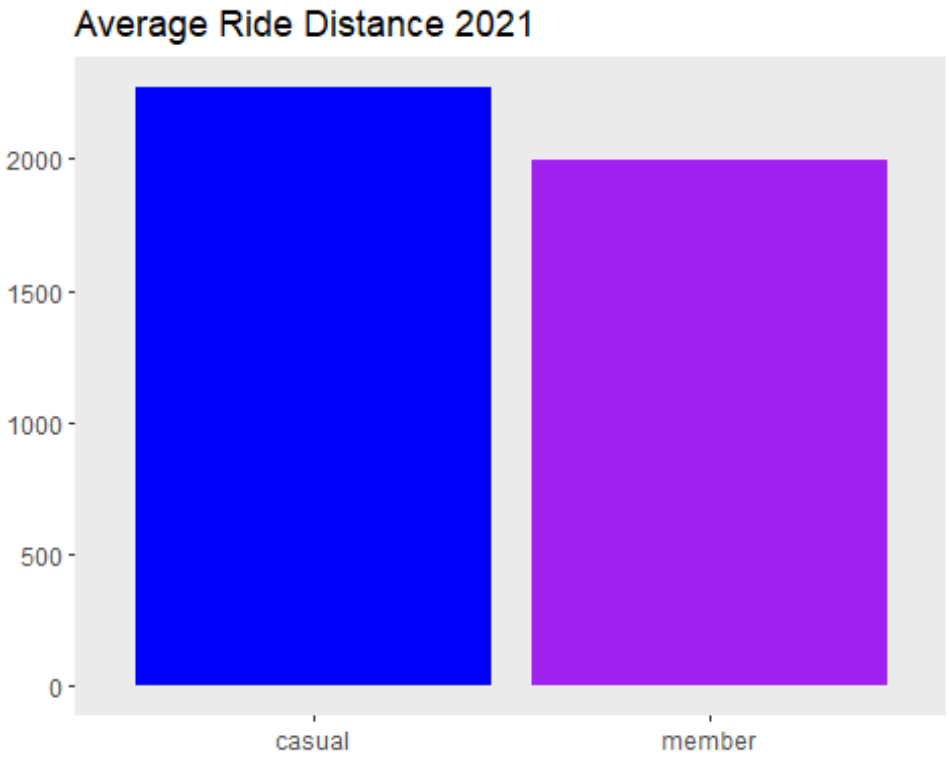
```
preparing_cyclistic_2021<-cyclistic_2021%>%
  filter(ride_duration>0)%>%
  filter(ride_length>0)%>%
  filter(!is.na(start_station_name))%>%
  filter(!is.na(end_station_name))%>%
  filter(!start_station_name == "")


preparing_cyclistic_2022<-cyclistic_2022%>%
  filter(ride_duration>0)%>%
  filter(ride_length>0)%>%
  filter(!is.na(start_station_name))%>%
  filter(!is.na(end_station_name))%>%
  filter(!start_station_name == "")


# Analysis

preparing_cyclistic_2021$weekday <- ordered(preparing_cyclistic_2021$weekday,
levels=c("Sunday", "Monday", "Tuesday", "Wednesday", "Thursday", "Friday",
"Saturday"))

preparing_cyclistic_2022$weekday <- ordered(preparing_cyclistic_2022$weekday,
levels=c("Sunday", "Monday", "Tuesday", "Wednesday", "Thursday", "Friday",
"Saturday"))
```

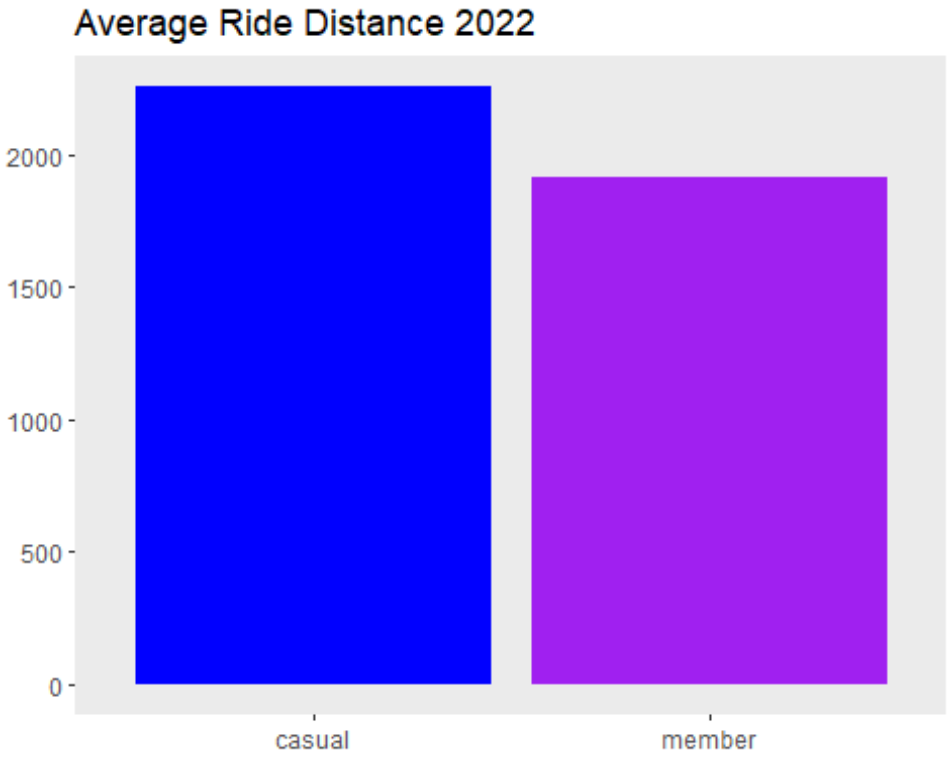
Including Plots

You can also embed plots, for example:

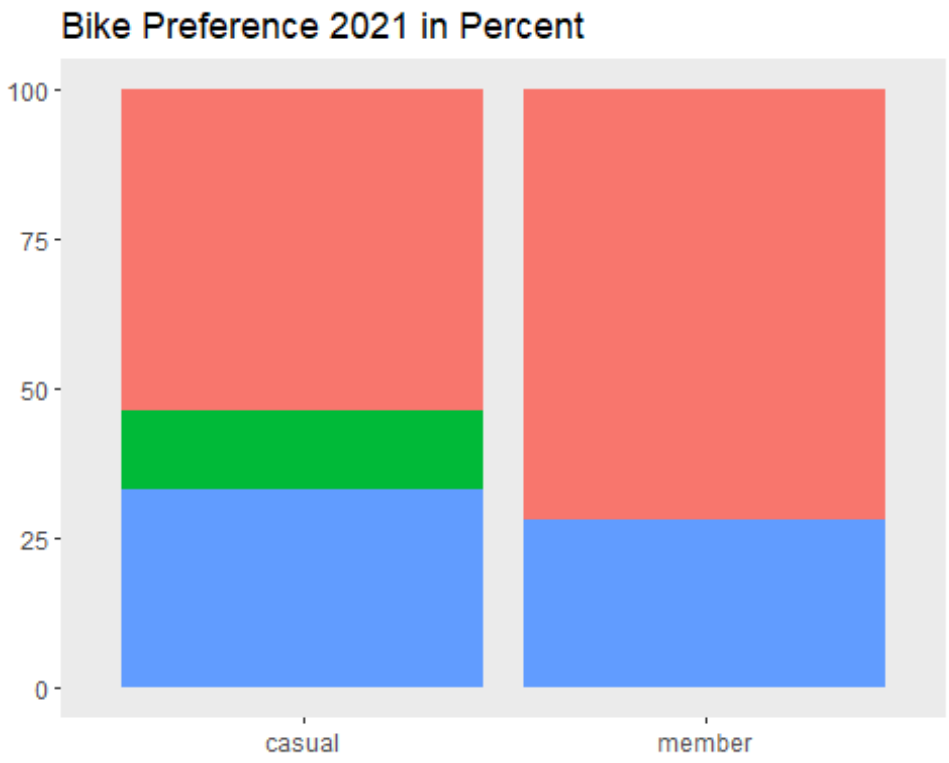




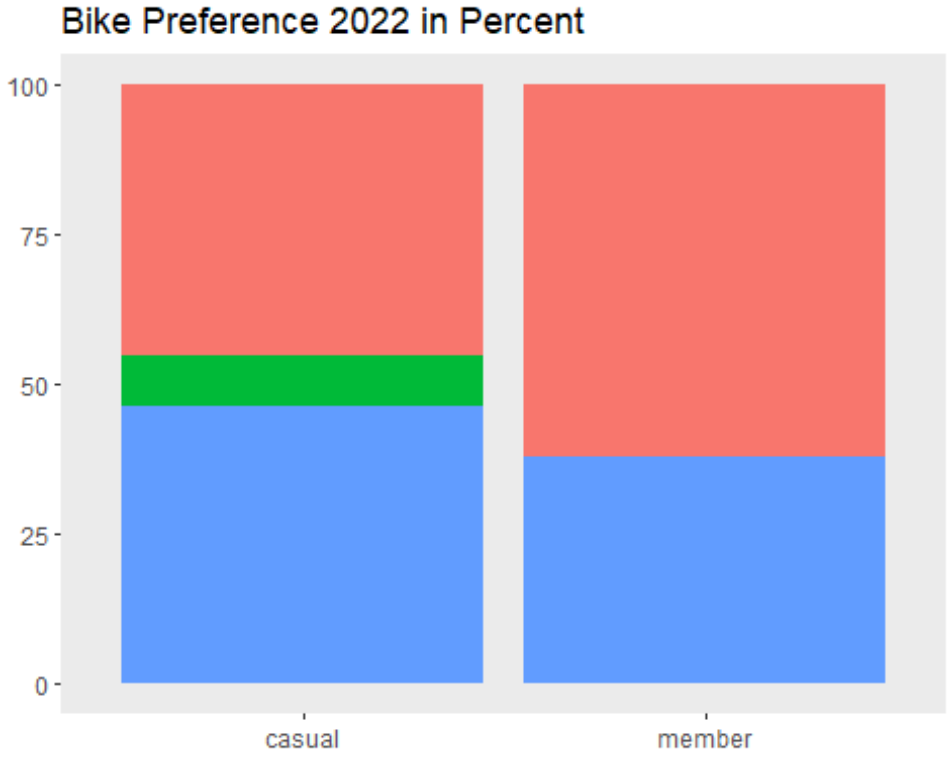




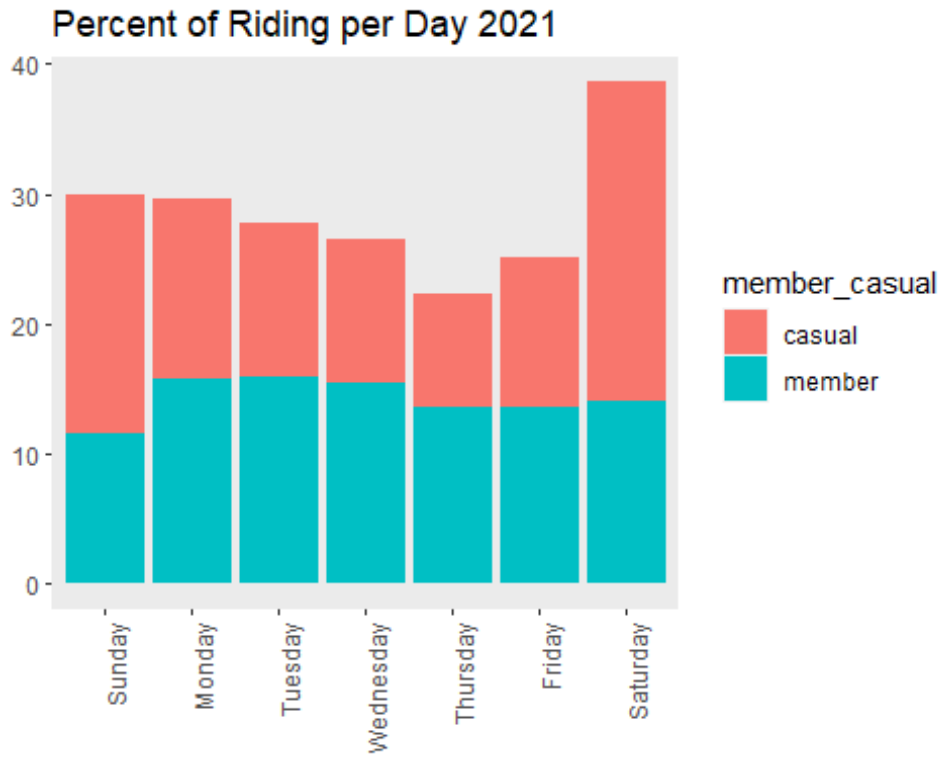

```
## `summarise()` has grouped output by 'member_casual'. You can override
using the
## `.groups` argument.
```



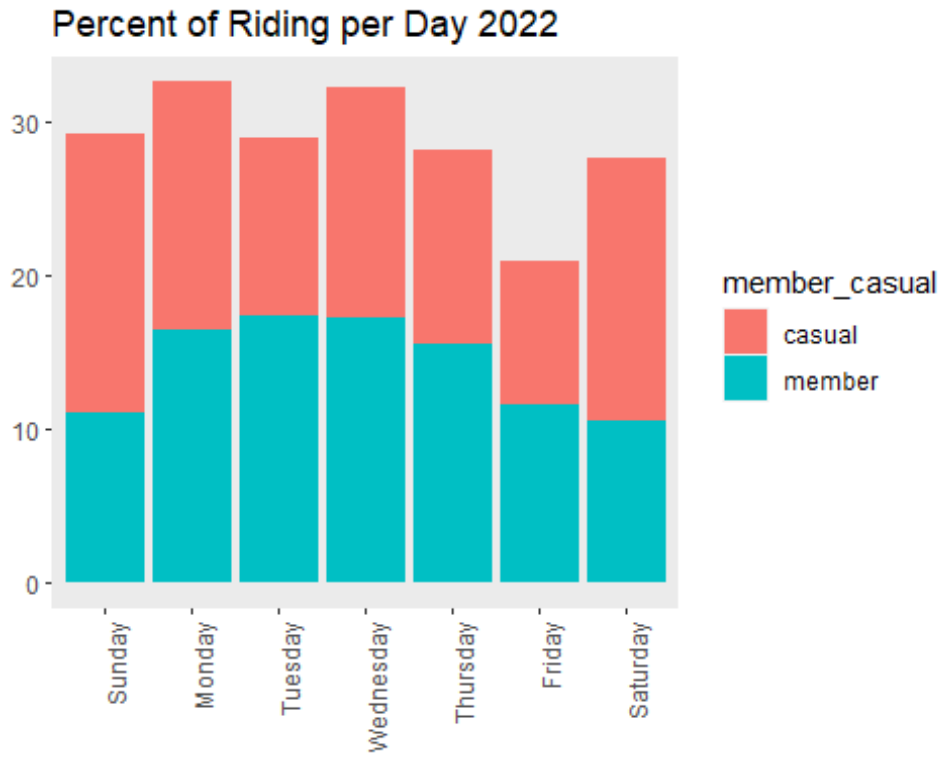

```
## `summarise()` has grouped output by 'member_casual'. You can override
using the
## `.groups` argument.
```

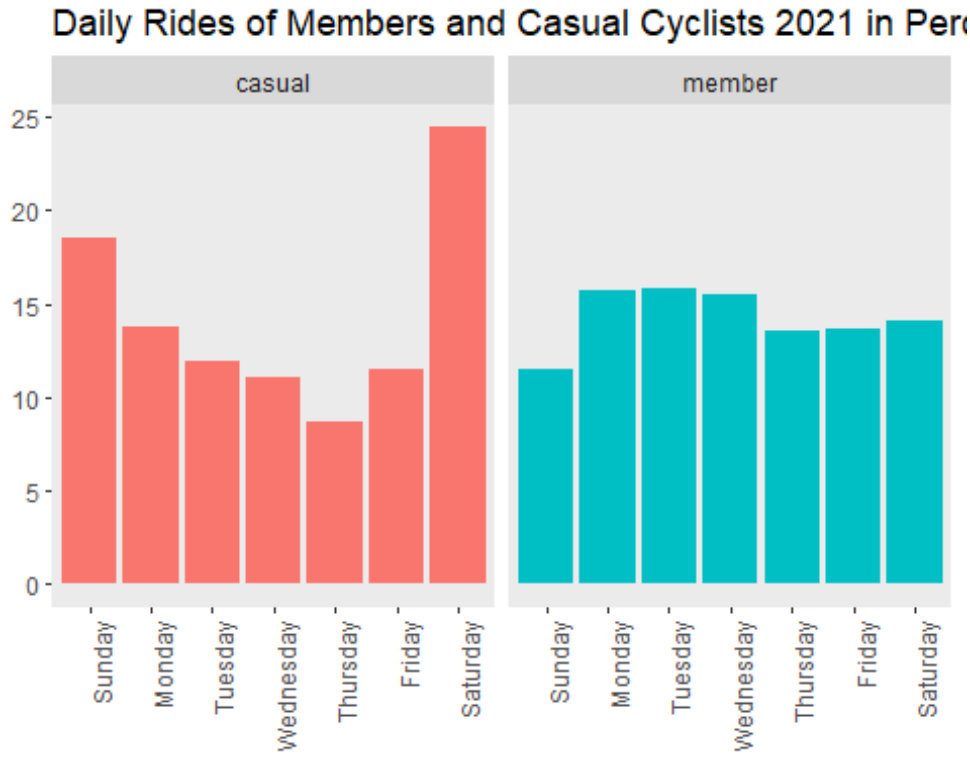


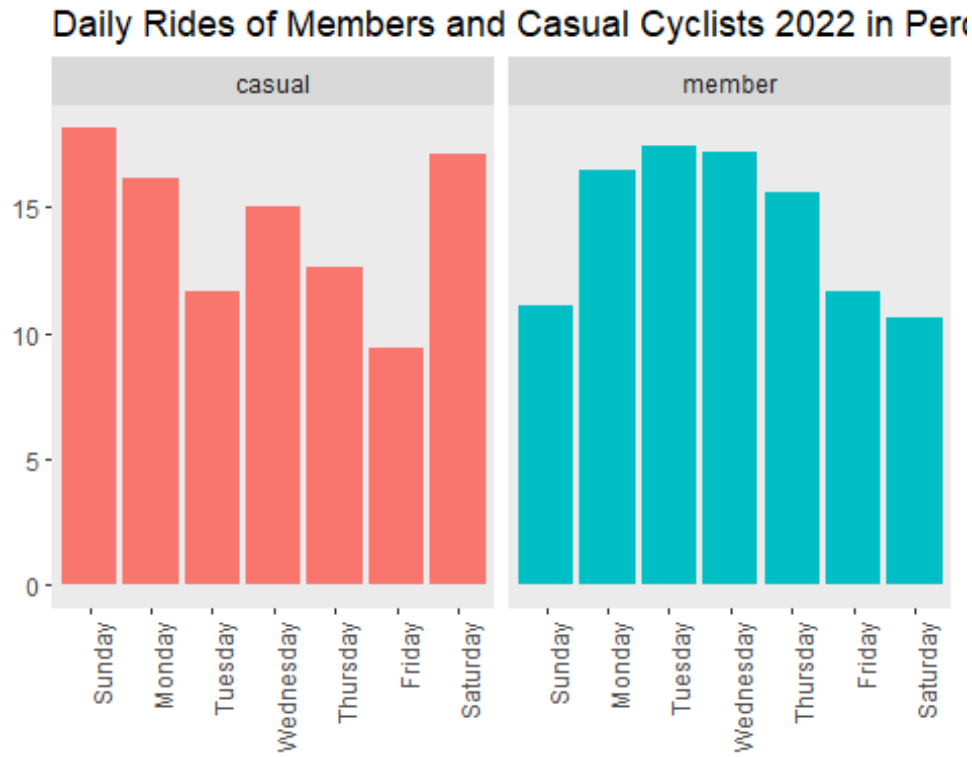
```
## `summarise()` has grouped output by 'member_casual'. You can override
using the
## `.groups` argument.
```




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


Percent of Riding per Day 2022
30
20
10
0
Sunday
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
member_casual
casual
member


Daily Rides of Members and Casual Cyclists 2021 in Per
casual
member
25
20
15
10
5
0
Sunday
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.