proj1

Drew

1/6/2021

Project 1: Movement

Submitted by Andrew Chang on Jan 08 2021 for the Johns Hopkins Reproducible Research course for -0.3263944 credits

```
readAct <- read.csv("C:/Users/arkai/Documents/R/projects/DSS05_RepRes/repdata_data_activity/activity.csv
                    colClasses = c("numeric", "POSIXct", "integer"))

?read.csv
```

```
## starting httpd help server ... done
```

```
summary(readAct)
```

```
##      steps             date                          interval
##  Min.   :  0.00   Min.   :2012-10-01 00:00:00   Min.   :   0.0
##  1st Qu.:  0.00   1st Qu.:2012-10-16 00:00:00   1st Qu.: 588.8
##  Median :  0.00   Median :2012-10-31 00:00:00   Median :1177.5
##  Mean   : 37.38   Mean   :2012-10-31 00:25:34   Mean   :1177.5
##  3rd Qu.: 12.00   3rd Qu.:2012-11-15 00:00:00   3rd Qu.:1766.2
##  Max.   :806.00   Max.   :2012-11-30 00:00:00   Max.   :2355.0
##  NA's   :2304
```

```
dim(readAct)
```

```
## [1] 17568     3
```

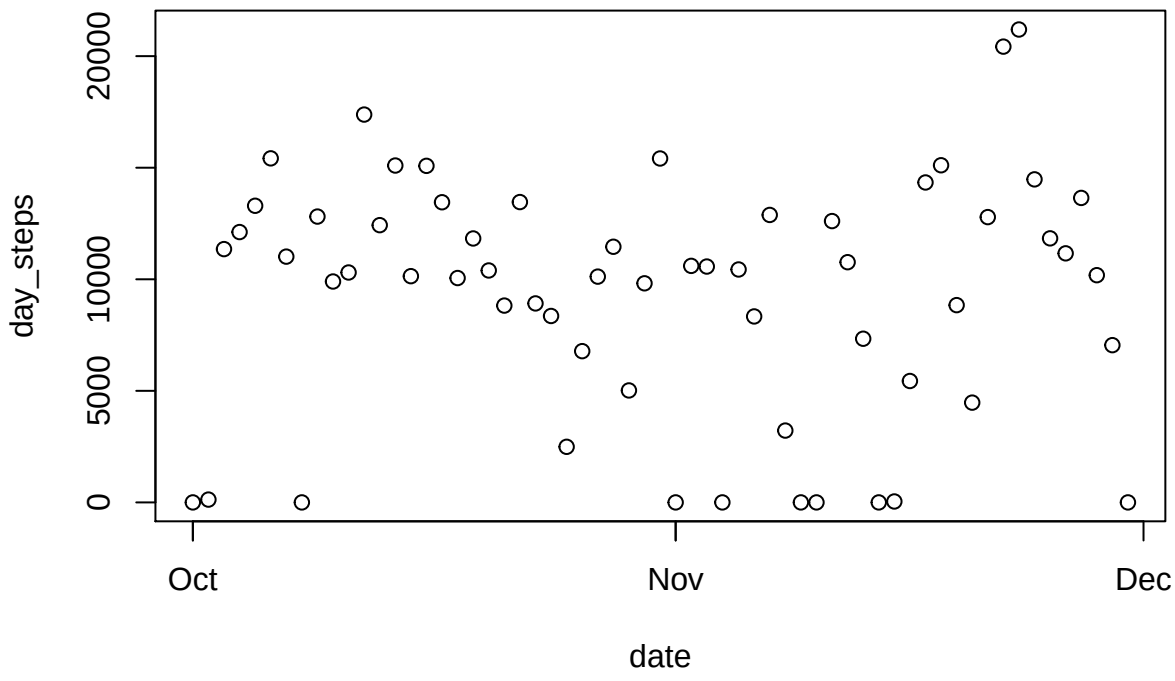
```
head(readAct)
```

```
##   steps       date interval
## 1    NA 2012-10-01        0
## 2    NA 2012-10-01        5
## 3    NA 2012-10-01       10
## 4    NA 2012-10-01       15
## 5    NA 2012-10-01       20
## 6    NA 2012-10-01       25
```

```
actDaily <- readAct %>%
        select(
                date, steps
        ) %>%
        group_by(date) %>%
        summarize(
                day_steps = sum(steps, na.rm = TRUE)
        )
```

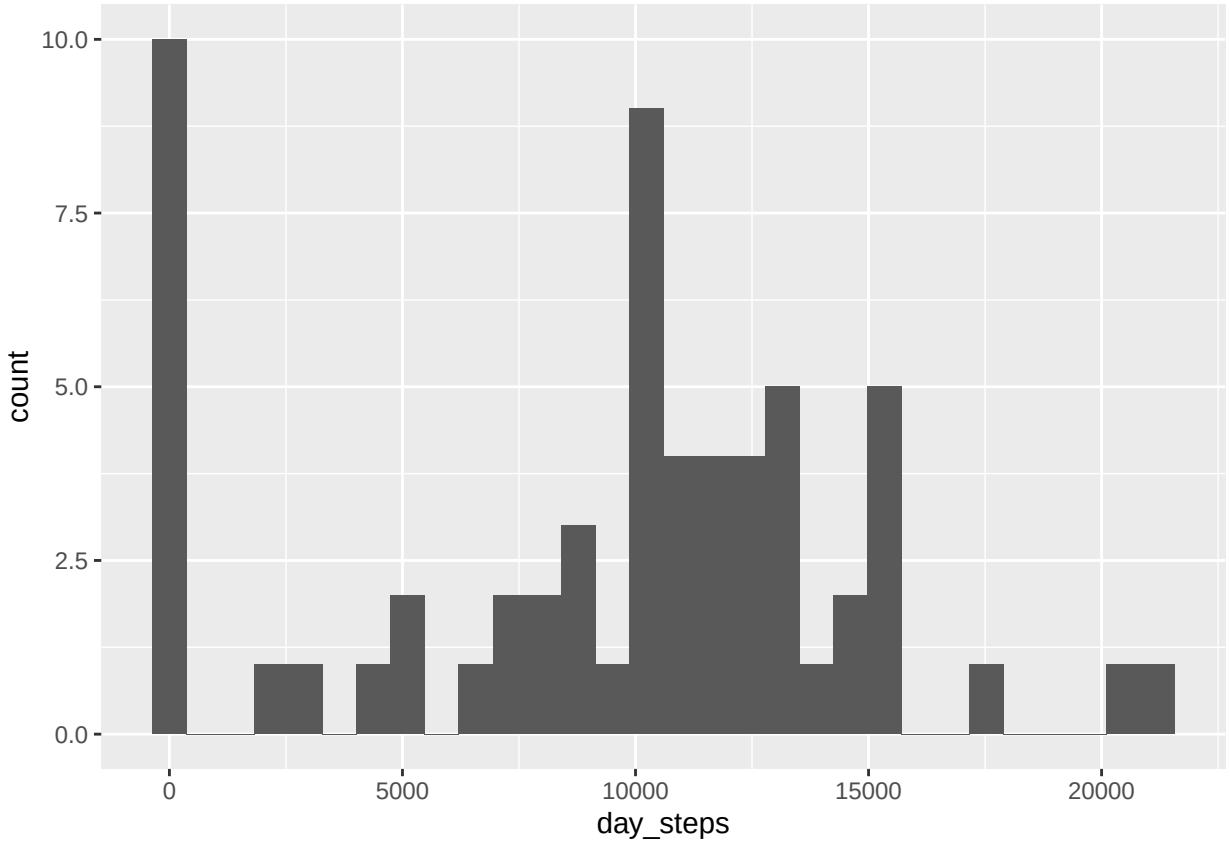
```
actDailyMean <- mean(actDaily$day_steps, na.rm = TRUE)
actDailyMed <- median(actDaily$day_steps, na.rm = TRUE)

plot(actDaily)
```




```
ggplot(data = actDaily, aes(day_steps)) +
       geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

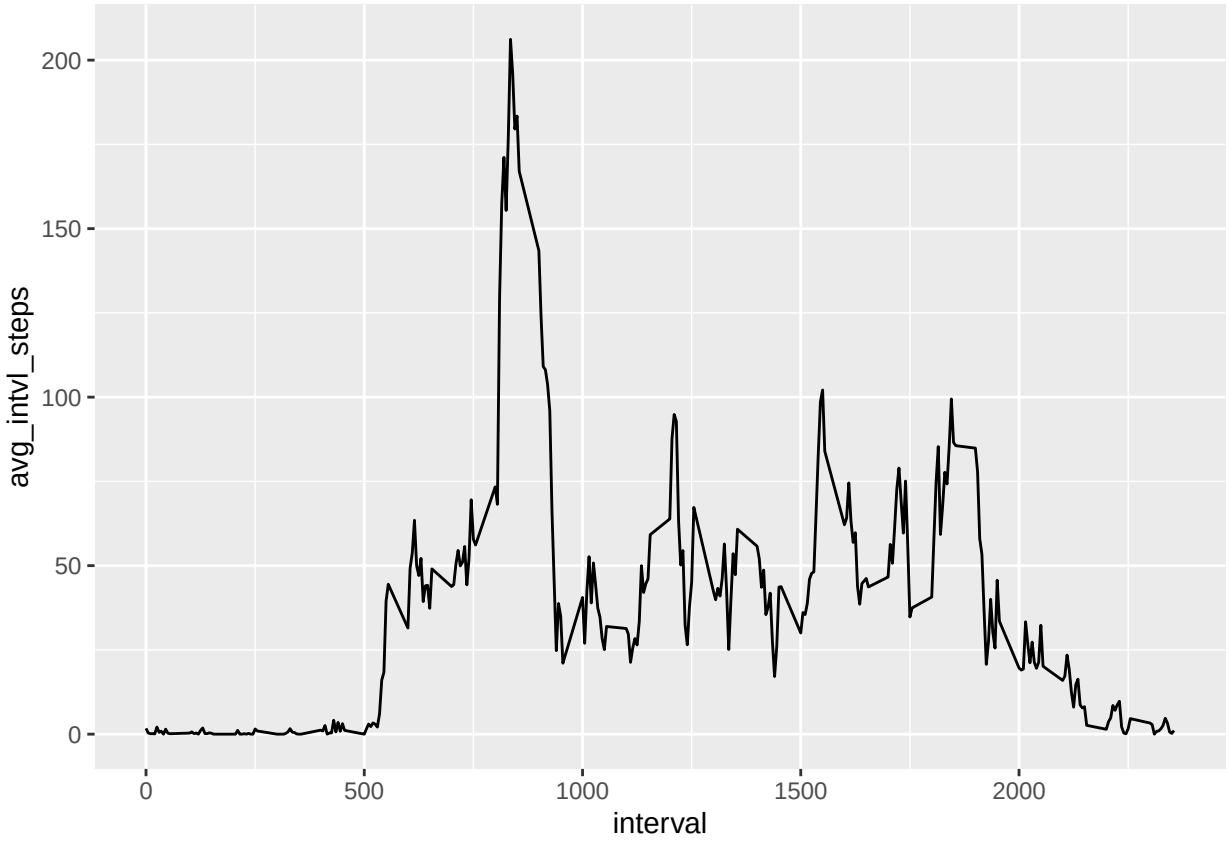



```
actIntvl <- readAct %>%
        select(
                date, steps, interval
        ) %>%
        group_by(interval) %>%
        summarize(
                avg_intvl_steps = mean(steps, na.rm = TRUE)
        )

dim(actIntvl)
```

```
## [1] 288   2
```

```
ggplot(data = actIntvl, aes(interval, avg_intvl_steps)) +
        geom_line()
```




```
nullsteps <- readAct %>%
        select(
                steps, date, interval
        ) %>%
        filter(is.na(steps)) %>%
        count()

nullVals <- nullsteps$n
class(nullVals)
```

```
## [1] "integer"
```

```
# actDaily %>%
#         filter(
#                 day_steps == max(actDaily$day_steps, na.rm = TRUE)
#         )
```

There are 2304 missing values in the dataset.

Including Plots

You can also embed plots, for example:

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