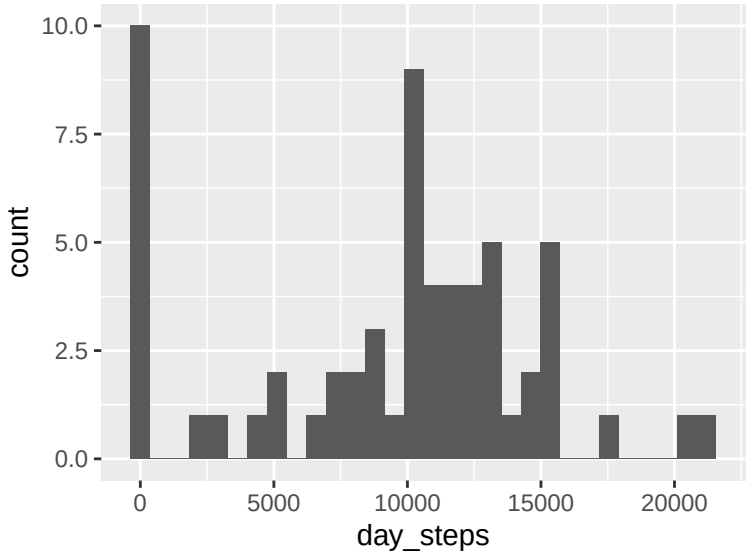
Project 1: Movement

Submitted by Andrew Chang on Jan 10 2021 for the Johns Hopkins Reproducible Research course

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

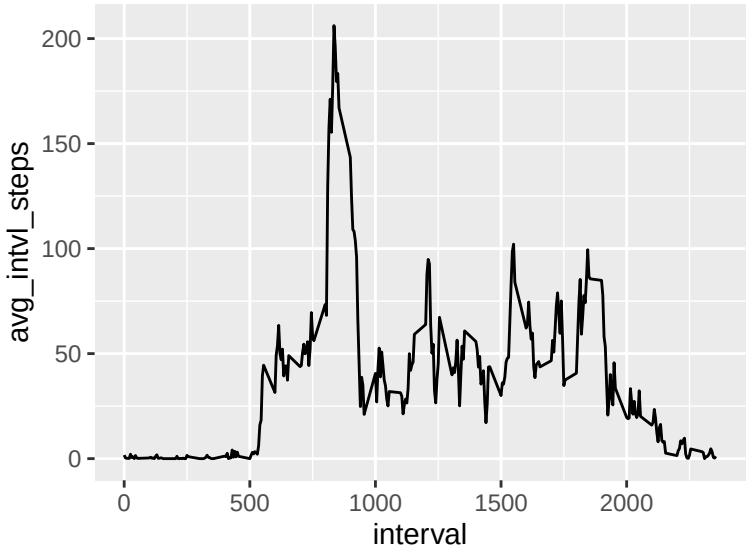
```
actDaily <- readAct %>%
        select(
                date, steps
        ) %>%
        group_by(date) %>%
        summarize(
                day_steps = sum(steps, na.rm = TRUE)
        )

actDailyMean <- round(mean(actDaily$day_steps, na.rm = TRUE), 1)
actDailyMed <- round(median(actDaily$day_steps, na.rm = TRUE), 1)

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



The mean and median number of steps taken each day are 9354.2 and 1.0395×10^4, respectively.

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

maxIntvlSteps <- round(max(actIntvl$avg_intvl_steps, na.rm = TRUE), 1)
maxIntvl <- actIntvl$interval[which.max(actIntvl$avg_intvl_steps)]

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

```
        geom_line()
```



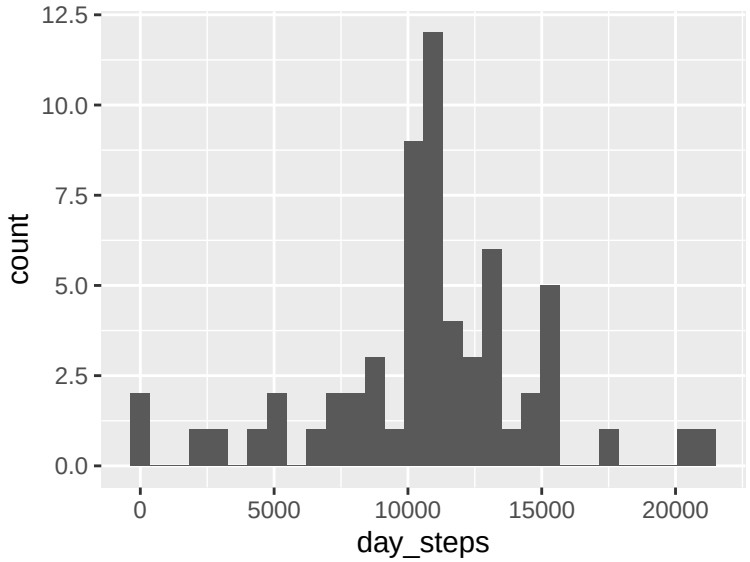

```
nullStep <- readAct %>%
        select(
                steps, date, interval
        ) %>%
        filter(is.na(steps))

nullStepRows <- nullStep %>%
        count()

nullStepCount <- nullStepRows$n
```

The interval with the highest average daily step count is 835 with 206.2 steps.

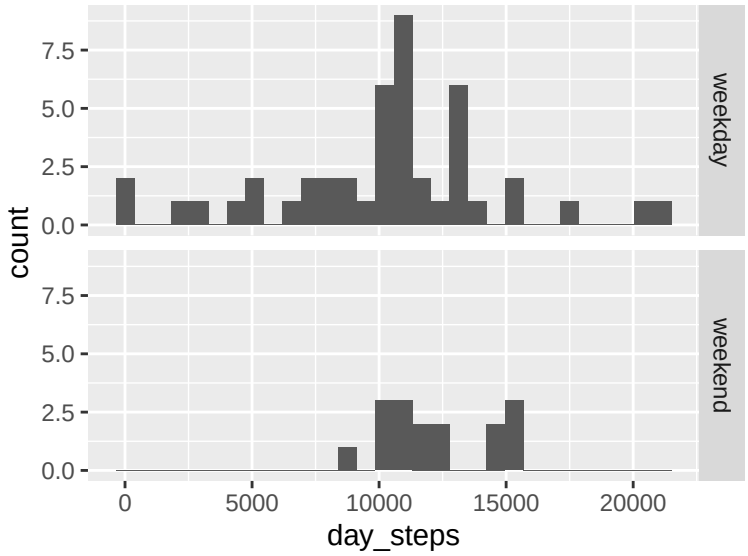
There are 2304 missing values in the dataset.

```
avgSteps <- nullStep %>%
        inner_join(actIntvl, by = "interval") %>%
        select(
                date,
                interval,
                avg_intvl_steps
        )

readAct0 <- readAct %>%
        left_join(avgSteps, by = c("date", "interval")) %>%
        mutate(
                steps = ifelse(is.na(steps), avg_intvl_steps, steps)
        )

nullRA0 <- readAct0 %>%
        select(
                steps, date, interval
        ) %>%
        filter(is.na(steps))
```

```
actDaily_impAvg <- readAct0 %>%
        select(
                date, steps
        ) %>%
        group_by(date) %>%
        summarize(
                day_steps = sum(steps, na.rm = TRUE)
        )

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




```
readAct_wkpt <- readAct0 %>%
        mutate(
                weekpart = factor(ifelse(weekdays(date) %in% c("Saturday", "Sunday"), "weekend", "weekda
        )

actDaily_wkptAvg <- readAct_wkpt %>%
        select(
                date, weekpart, steps
        ) %>%
        group_by(date, weekpart) %>%
        summarize(
                day_steps = sum(steps, na.rm = TRUE)
        )

ggplot(data = actDaily_wkptAvg, aes(day_steps)) +
        geom_histogram() +
        facet_grid(rows = vars(weekpart))
```


count
7.5
5.0
2.5
0.0
weekday
weekend
0
5000
10000
15000
20000
day_steps