

Main Project Code

Travis Helm

11/3/2021

Peer Graded Assignment 1

Loading and Processing Data (NOTE: I saved in my folder unzipped).

I'm adding a weekend/weekdy column

```
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

activity <- read.csv('activity.csv')

activity <- activity %>%
  mutate(date = as.Date(activity$date, "%Y-%m-%d"),
         Date_Type = if_else(weekdays(date) %in% c('Saturday','Sunday'), 'weekend','weekday'),
         Weekday = weekdays(date))
head(activity)

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

What is mean total number of steps taken per day?

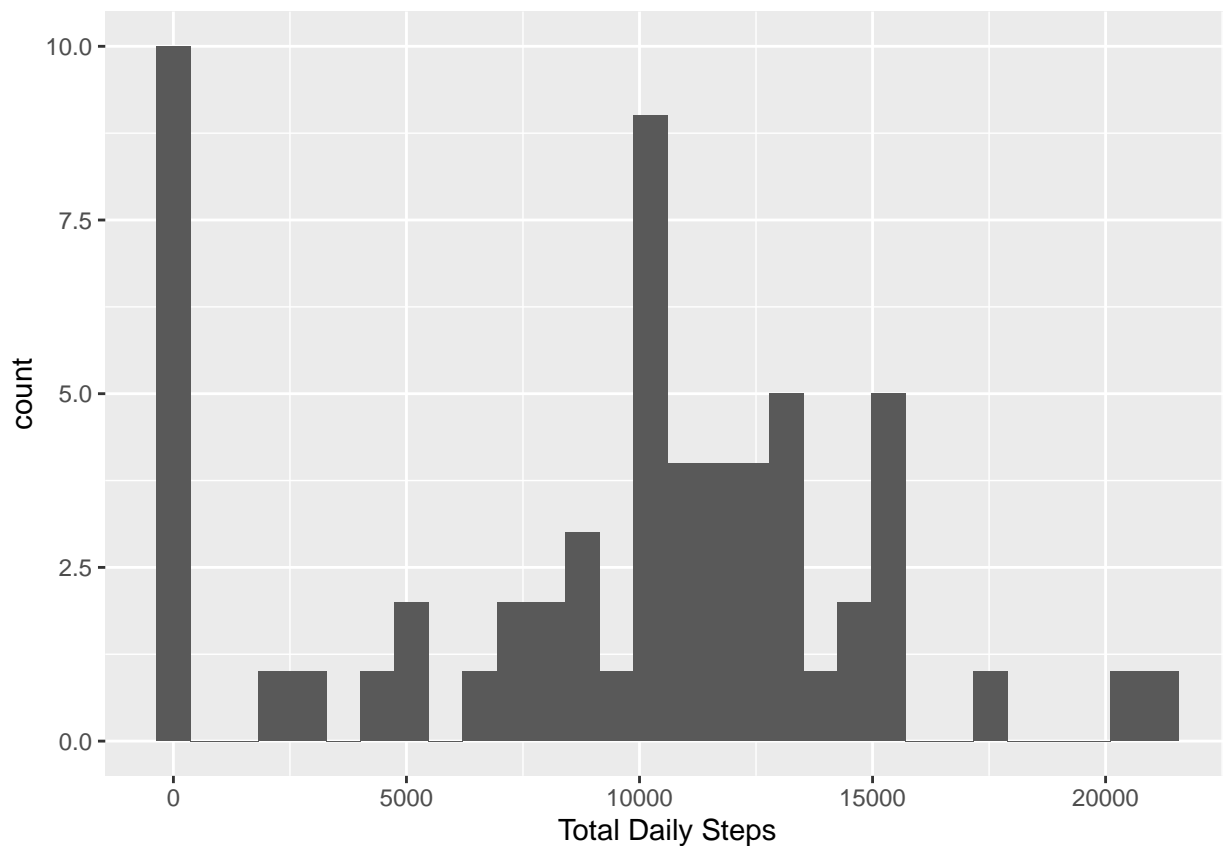
Create Histogram

I group by date and then get the total number of steps

```
library(ggplot2)
Total_Daily_Steps <- activity %>%
  group_by(date) %>%
  summarise(Total_Steps = sum(steps, na.rm=TRUE)) %>%
  ungroup()

ggplot(data=Total_Daily_Steps, aes(x=Total_Steps)) +
  geom_histogram() +
  labs(x='Total Daily Steps')
```

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



Summary statistics (including the mean and median) are shown below.

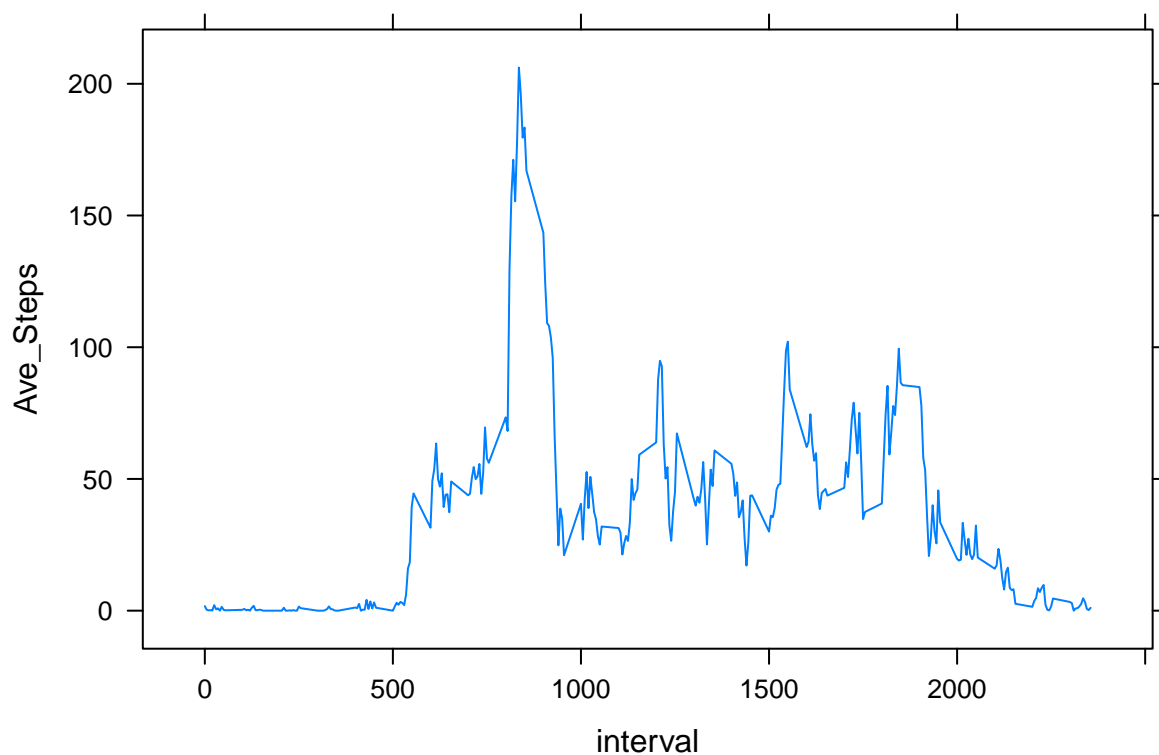
```
summary(Total_Daily_Steps$Total_Steps)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##         0    6778   10395   9354  12811  21194
```

What is the average daily activity pattern?

I group by interval and get the total number of steps

```
Ave_Steps <- activity %>%  
  group_by(interval) %>%  
  summarise(Ave_Steps = mean(steps, na.rm=TRUE)) %>%  
  ungroup()  
  
library(lattice)  
xyplot(Ave_Steps ~ interval, data = Ave_Steps, type = c("l"))
```



What five minute interval contains the most steps?

Interval 835 (see below)

```
max_interval <- Ave_Steps[Ave_Steps$Ave_Steps == max(Ave_Steps$Ave_Steps),]  
max_interval
```

```
## # A tibble: 1 x 2  
##   interval Ave_Steps  
##   <int>     <dbl>  
## 1     835     206.
```

Imputing Missing Values

There are 2304 missing cases

```
## [1] 2304
```

Creating Dataset to replace missing values with average for interval

I create a column called `steps_noNA` that has the average steps for an interval when the steps for that day aren't actually known

```
activity2 <- activity %>%  
  left_join(Ave_Steps, by = 'interval') %>%  
  mutate(steps_noNA = ifelse(is.na(steps), Ave_Steps, steps))
```

Histogram with corrected values

```
Total_Daily_Steps2 <- activity2 %>%  
  group_by(date) %>%  
  summarise(Total_StepsNA = sum(steps, na.rm=TRUE),  
            Total_StepsNoNA = sum(steps_noNA)) %>%  
  ungroup()
```

Mean and Median with and without NAs

The mean and median are lower when missing values are present

```
summary(Total_Daily_Steps2[,c('Total_StepsNA', 'Total_StepsNoNA')])
```

```
## Total_StepsNA Total_StepsNoNA  
## Min.      :    0  Min.      :   41  
## 1st Qu.: 6778  1st Qu.: 9819  
## Median :10395 Median :10766  
## Mean    : 9354  Mean    :10766  
## 3rd Qu.:12811 3rd Qu.:12811  
## Max.    :21194 Max.    :21194
```

```
##Plotting by weekend/weekday
```

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

