Reproducible Research - Week Two

merkasaur

2/19/2020

R Markdown

Reading in activity.csv file and renaming to AMD. Histogram is built. Dependencies loaded. Steps by Day Histogram Mean: 10766.19 Median: 10765

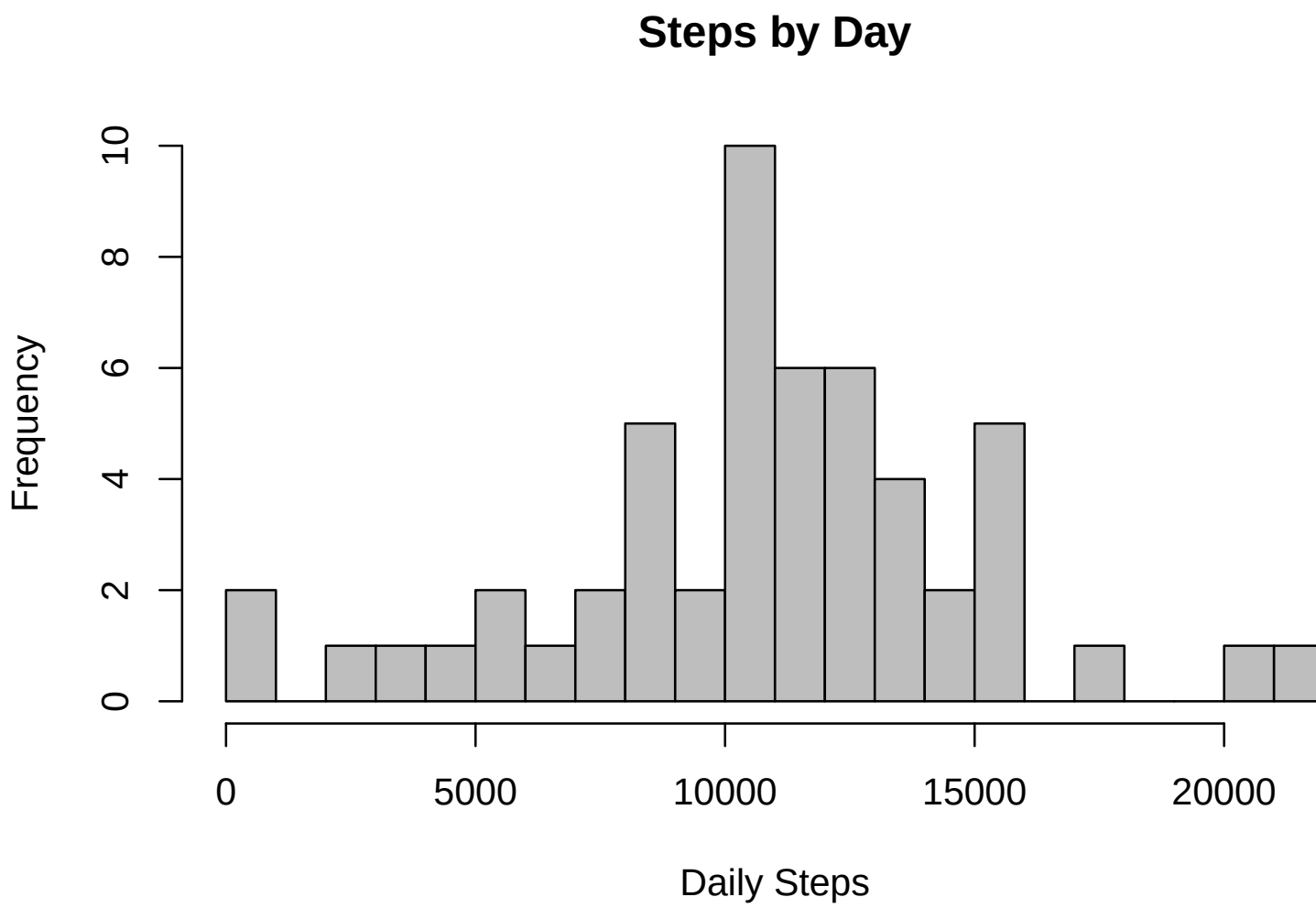
```
#Dependencies
library(magrittr)
library(dplyr)
```

```
##
## 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
```

```
library(ggplot2)
library(readr)
#Read activity.csv.
AMD <- read.csv("activity.csv")
#Histogram
stepsdate <- AMD %>%
      select(date, steps) %>%
      group_by(date) %>%
      summarise(dsteps = sum(steps)) %>%
      na.omit()
hist(stepsdate$dsteps,xlab="Daily Steps",main="Steps by Day",col = c("gray"),breaks = 20)
```




```
#mean
##This is the mean.
mean(stepsdate$dsteps)
```

```
## [1] 10766.19
```

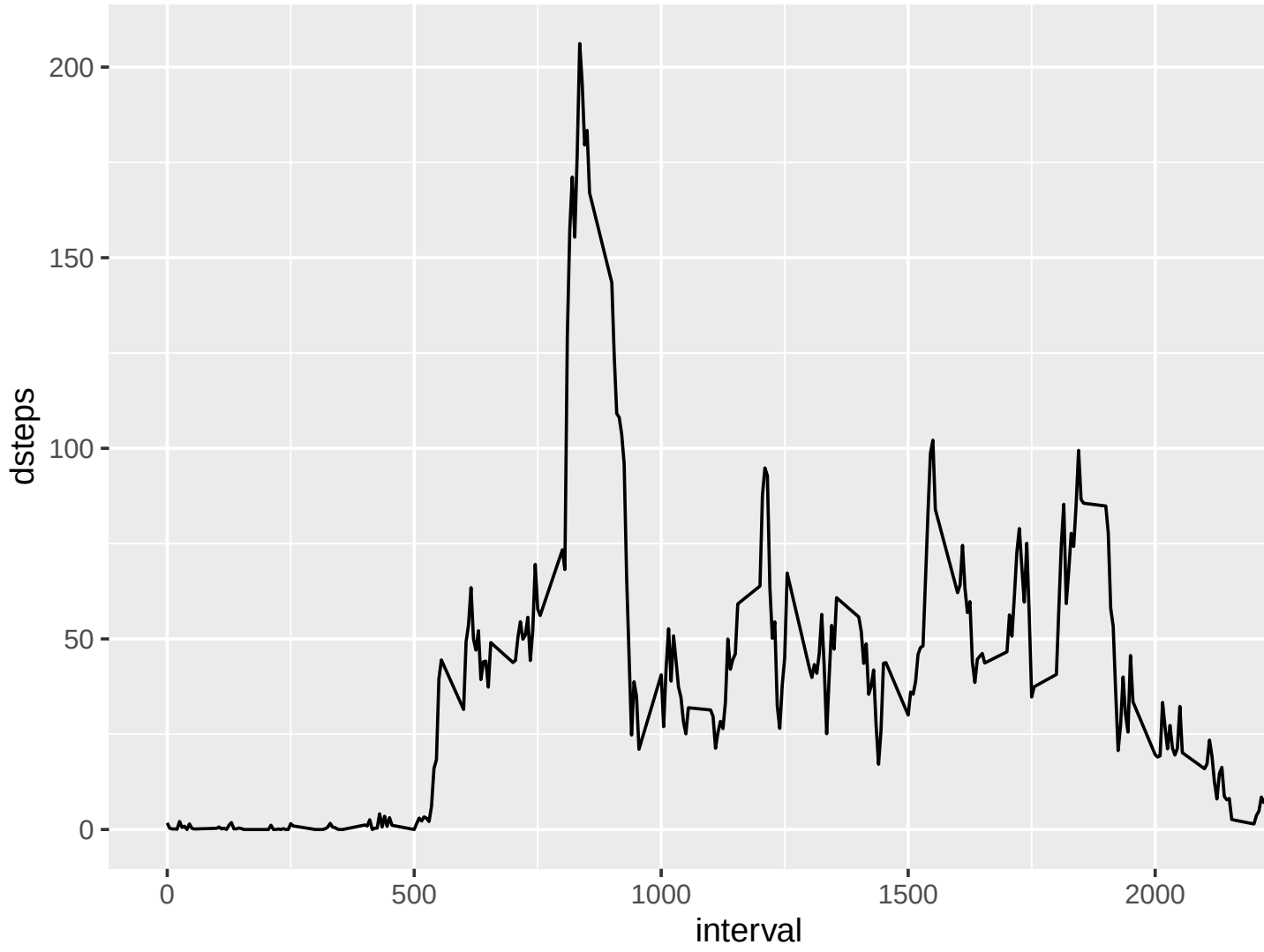
```
##This is the median.
median(stepsdate$dsteps)
```

```
## [1] 10765
```

##Time Series Plot

Time series plot is generated. Data is organized for five minute intervals and missing values are fixed.

```r
#Timea series plot
datatimeseries <- AMD %>%
      select(interval, steps) %>%
      na.omit() %>%
      group_by(interval) %>%
      summarize(dsteps=mean(steps))
#plot time series
ggplot(datatimeseries,aes(x=interval,y=dsteps)) +
      geom_line()
```




```r
#Five minute intervals
datatimeseries[which(datatimeseries$dsteps==max(datatimeseries$dsteps)),]
```

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

```r
#Input missing values
missingNAs <- sum(is.na(data))
#Fill in missing values
replaceNA <-function(z) replace(z, is.na(z), mean(z, na.rm=TRUE))
MD <- AMD %>%
  group_by(interval) %>%
  mutate(steps= replaceNA(steps))
```

Histogram generation of Total steps by day with Mean/Median values.

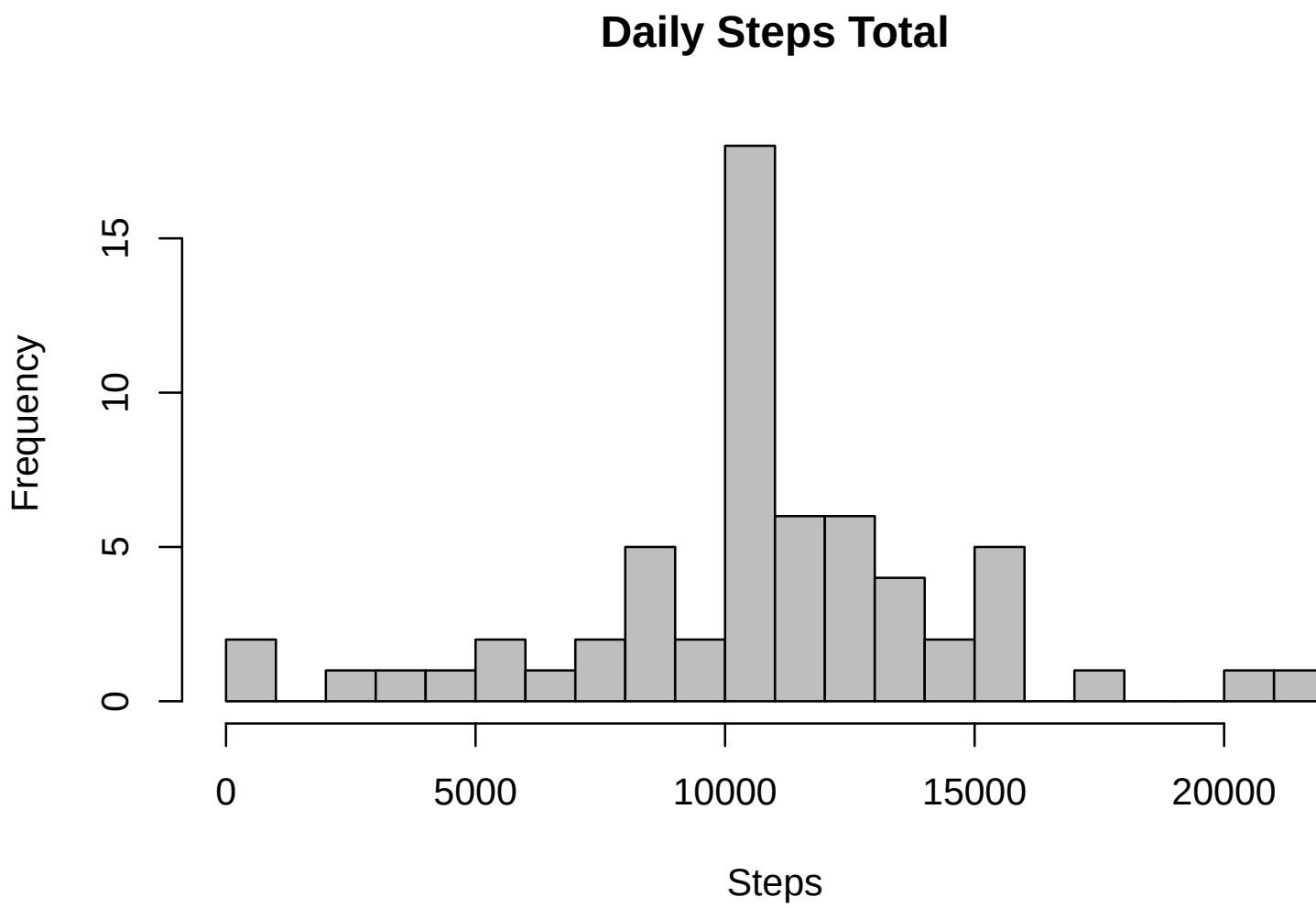
```r
#Histogram Total steps by day Mean/median

totalstepsbyday <- aggregate(MD$steps, by=list(MD$date), sum)
names(totalstepsbyday)[1] ="date"
names(totalstepsbyday)[2] = "completesteps"

#Mean & median of Steps by Day
summary(totalstepsbyday)
```

```
##          date    completesteps
##  2012-10-01: 1   Min.   :   41
##  2012-10-02: 1   1st Qu.: 9819
##  2012-10-03: 1   Median :10766
##  2012-10-04: 1   Mean   :10766
##  2012-10-05: 1   3rd Qu.:12811
##  2012-10-06: 1   Max.   :21194
##  (Other)   :55
```

```r
#Histogram code
hist(totalstepsbyday$completesteps, xlab="Steps", ylab = "Frequency", main = "Daily Steps Total", col= 
```



Mean Values calculated and information generated for weekday vs weekend information. GGplot actions occuring.

```
# Old to New Mean
firstmean <- mean(stepsdate$dsteps, na.rm = TRUE)
secondmean <- mean(totalstepsbyday$dsteps)
#Old to new median
firstmedian <- median(stepsdate$dsteps, na.rm = TRUE)
secondmedian <- median(totalstepsbyday$dsteps)
#Weekdays vs Weekends
MD$date <- as.Date(MD$date)
MD$weekday <- weekdays(MD$date)
MD$weekend <- ifelse(MD$weekday=="Saturday" | MD$weekday=="Sunday", "Weekend", "Weekday")
MDbothwkwkd <- aggregate(MD$steps, by = list(MD$weekend, MD$interval), na.omit(mean))
names(MDbothwkwkd) <- c("weekend", "interval", "steps")
ggplot(MDbothwkwkd, aes(x=interval, y=steps))+
      geom_line()+
      facet_grid(weekend~.)+
      xlab("Interval")+
```

```
      ylab("Steps Mean")+
      ggtitle("Average steps by Interval")
```