Reproducible Research: Course Project 1

Steve Orosz

August 19, 2018

```
knitr::opts_chunk$set(echo = TRUE)

#Libraries

library(lubridate)
```

```
##
## Attaching package: 'lubridate'
```

```
## The following object is masked from 'package:base':
##
##     date
```

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:lubridate':
##
##     intersect, setdiff, union
```

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

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

```
#Set working directory

setwd("C:/Users/smoro/OneDrive/Desktop/Data Science/John Hopkins University/Reproducible Research/Proje

#getwd()

#Get file

Mydata <- read.csv("C:/Users/smoro/OneDrive/Desktop/Data Science/John Hopkins University/Reproducible R

head(Mydata)
```

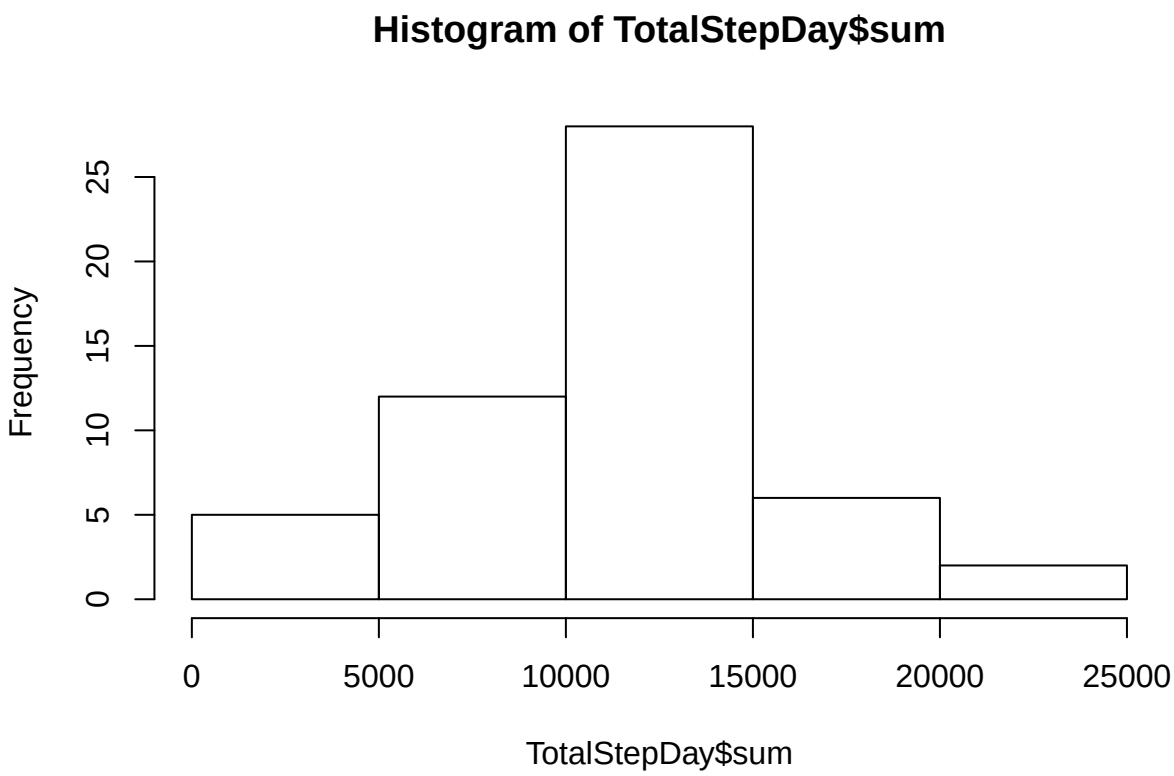
```
##   steps       date interval
## 1    NA 2012-10-01        0
```

```
## 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
```

```
#View(Mydata)

#Create weekday fields
MydataComplete <- Mydata %>% mutate (weekday = weekdays(as.Date(Mydata$date))) %>% mutate(weekdayNum = w

#View(MydataComplete)

#Create data set without misssing values and with missing values

WithOut <-  na.exclude(MydataComplete)
With <- MydataComplete

#View(WithOut)
#View(With)
```

```
#Count total steps per day
ts <- tapply(WithOut$steps, WithOut$date , sum)

TotalStepDay <- data.frame(date = names(ts), sum = ts)

#View(TotalStepDay)
```

```
#Histogram of total steps per day
hist(TotalStepDay$sum)
```



What is mean and median total number of steps taken per day?

```
#Mean total steps per day
m1 <- tapply(WithOut$steps, WithOut$date , mean)

MeanStepDay <- data.frame(date = names(m1), DateMean = m1)

#View(MeanStepDay)

MeanStepDay
```

```
##                  date   DateMean
## 2012-10-01 2012-10-01         NA
## 2012-10-02 2012-10-02  0.4375000
## 2012-10-03 2012-10-03 39.4166667
## 2012-10-04 2012-10-04 42.0694444
## 2012-10-05 2012-10-05 46.1597222
## 2012-10-06 2012-10-06 53.5416667
## 2012-10-07 2012-10-07 38.2465278
## 2012-10-08 2012-10-08         NA
## 2012-10-09 2012-10-09 44.4826389
## 2012-10-10 2012-10-10 34.3750000
## 2012-10-11 2012-10-11 35.7777778
## 2012-10-12 2012-10-12 60.3541667
```

```
## 2012-10-13 2012-10-13 43.1458333
## 2012-10-14 2012-10-14 52.4236111
## 2012-10-15 2012-10-15 35.2048611
## 2012-10-16 2012-10-16 52.3750000
## 2012-10-17 2012-10-17 46.7083333
## 2012-10-18 2012-10-18 34.9166667
## 2012-10-19 2012-10-19 41.0729167
## 2012-10-20 2012-10-20 36.0937500
## 2012-10-21 2012-10-21 30.6284722
## 2012-10-22 2012-10-22 46.7361111
## 2012-10-23 2012-10-23 30.9652778
## 2012-10-24 2012-10-24 29.0104167
## 2012-10-25 2012-10-25  8.6527778
## 2012-10-26 2012-10-26 23.5347222
## 2012-10-27 2012-10-27 35.1354167
## 2012-10-28 2012-10-28 39.7847222
## 2012-10-29 2012-10-29 17.4236111
## 2012-10-30 2012-10-30 34.0937500
## 2012-10-31 2012-10-31 53.5208333
## 2012-11-01 2012-11-01         NA
## 2012-11-02 2012-11-02 36.8055556
## 2012-11-03 2012-11-03 36.7048611
## 2012-11-04 2012-11-04         NA
## 2012-11-05 2012-11-05 36.2465278
## 2012-11-06 2012-11-06 28.9375000
## 2012-11-07 2012-11-07 44.7326389
## 2012-11-08 2012-11-08 11.1770833
## 2012-11-09 2012-11-09         NA
## 2012-11-10 2012-11-10         NA
## 2012-11-11 2012-11-11 43.7777778
## 2012-11-12 2012-11-12 37.3784722
## 2012-11-13 2012-11-13 25.4722222
## 2012-11-14 2012-11-14         NA
## 2012-11-15 2012-11-15  0.1423611
## 2012-11-16 2012-11-16 18.8923611
## 2012-11-17 2012-11-17 49.7881944
## 2012-11-18 2012-11-18 52.4652778
## 2012-11-19 2012-11-19 30.6979167
## 2012-11-20 2012-11-20 15.5277778
## 2012-11-21 2012-11-21 44.3993056
## 2012-11-22 2012-11-22 70.9270833
## 2012-11-23 2012-11-23 73.5902778
## 2012-11-24 2012-11-24 50.2708333
## 2012-11-25 2012-11-25 41.0902778
## 2012-11-26 2012-11-26 38.7569444
## 2012-11-27 2012-11-27 47.3819444
## 2012-11-28 2012-11-28 35.3576389
## 2012-11-29 2012-11-29 24.4687500
## 2012-11-30 2012-11-30         NA
```

```
#View(MeanStepDay) Removed 0 steps to calculate median
preMedianStepDay <- WithOut %>% filter(WithOut$steps > 0)
```

```r
#Median total steps per day
md <- tapply(preMedianStepDay$steps, preMedianStepDay$date , median)

MedianStepDay <- data.frame(date = names(md), DateMedian = md)

MedianStepDay
```

```
##                  date DateMedian
## 2012-10-01 2012-10-01         NA
## 2012-10-02 2012-10-02       63.0
## 2012-10-03 2012-10-03       61.0
## 2012-10-04 2012-10-04       56.5
## 2012-10-05 2012-10-05       66.0
## 2012-10-06 2012-10-06       67.0
## 2012-10-07 2012-10-07       52.5
## 2012-10-08 2012-10-08         NA
## 2012-10-09 2012-10-09       48.0
## 2012-10-10 2012-10-10       56.5
## 2012-10-11 2012-10-11       35.0
## 2012-10-12 2012-10-12       46.0
## 2012-10-13 2012-10-13       45.5
## 2012-10-14 2012-10-14       60.5
## 2012-10-15 2012-10-15       54.0
## 2012-10-16 2012-10-16       64.0
## 2012-10-17 2012-10-17       61.5
## 2012-10-18 2012-10-18       52.5
## 2012-10-19 2012-10-19       74.0
## 2012-10-20 2012-10-20       49.0
## 2012-10-21 2012-10-21       48.0
## 2012-10-22 2012-10-22       52.0
## 2012-10-23 2012-10-23       56.0
## 2012-10-24 2012-10-24       51.5
## 2012-10-25 2012-10-25       35.0
## 2012-10-26 2012-10-26       36.5
## 2012-10-27 2012-10-27       72.0
## 2012-10-28 2012-10-28       61.0
## 2012-10-29 2012-10-29       54.5
## 2012-10-30 2012-10-30       40.0
## 2012-10-31 2012-10-31       83.5
## 2012-11-01 2012-11-01         NA
## 2012-11-02 2012-11-02       55.5
## 2012-11-03 2012-11-03       59.0
## 2012-11-04 2012-11-04         NA
## 2012-11-05 2012-11-05       66.0
## 2012-11-06 2012-11-06       52.0
## 2012-11-07 2012-11-07       58.0
## 2012-11-08 2012-11-08       42.5
## 2012-11-09 2012-11-09         NA
## 2012-11-10 2012-11-10         NA
## 2012-11-11 2012-11-11       55.0
## 2012-11-12 2012-11-12       42.0
## 2012-11-13 2012-11-13       57.0
## 2012-11-14 2012-11-14         NA
```

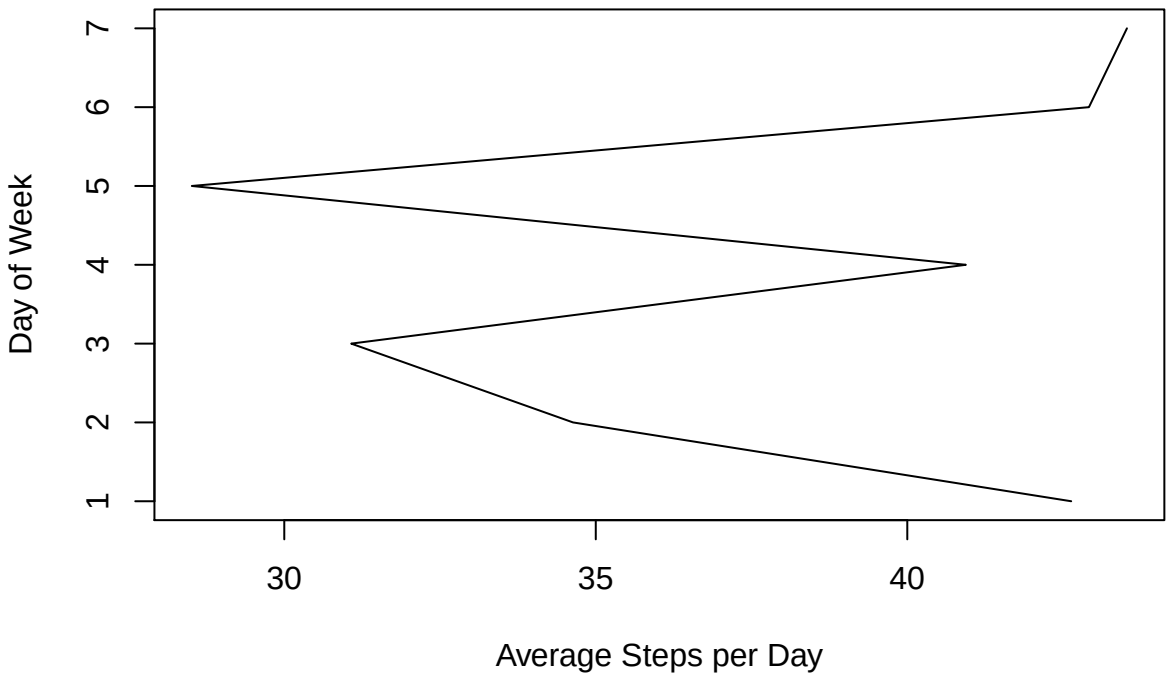
```
## 2012-11-15 2012-11-15       20.5
## 2012-11-16 2012-11-16       43.0
## 2012-11-17 2012-11-17       65.5
## 2012-11-18 2012-11-18       80.0
## 2012-11-19 2012-11-19       34.0
## 2012-11-20 2012-11-20       58.0
## 2012-11-21 2012-11-21       55.0
## 2012-11-22 2012-11-22       65.0
## 2012-11-23 2012-11-23      113.0
## 2012-11-24 2012-11-24       65.5
## 2012-11-25 2012-11-25       84.0
## 2012-11-26 2012-11-26       53.0
## 2012-11-27 2012-11-27       57.0
## 2012-11-28 2012-11-28       70.0
## 2012-11-29 2012-11-29       44.5
## 2012-11-30 2012-11-30         NA
```

```
#View(MedianStepDay)
```

What is the average daily activity pattern?

Saturday has the highest step average than the rest of the week

```
FiveInterval <- WithOut  %>% mutate(FiveInterval = ((steps / interval) * 1.0) * 5)

m2 <- tapply(FiveInterval$steps, FiveInterval$weekdayNum,  mean)

MeanWeekDays <- data.frame(day = names(m2), mean = m2)

#Display time series of average steps taken over week

plot(  MeanWeekDays$mean, MeanWeekDays$day, type = "l", xlab = "Average Steps per Day", ylab = "Day of W
```



##Missing values Section #What is the total number of rows with missing values?

```
sum(is.na(With$steps))
```

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

Calculate mean and median per interval of steps. This will be used for days with NA for steps

Date could not be used for NA data since no step data for them.

```
#produce mean average
m3 <- tapply(WithOut$steps, WithOut$interval , mean)

MeanStepInterval <- data.frame(interval = names(m3), IntervalMean = m3)

#produce median average

m4 <- tapply(WithOut$steps, WithOut$interval , median)

MedianStepInterval <- data.frame(interval = names(m4), IntervalMedian = m4)

MedianStepInterval
```

```
##      interval IntervalMedian
## 0           0              0
## 5           5              0
## 10         10              0
## 15         15              0
## 20         20              0
## 25         25              0
## 30         30              0
## 35         35              0
## 40         40              0
## 45         45              0
## 50         50              0
## 55         55              0
## 100       100              0
## 105       105              0
## 110       110              0
## 115       115              0
## 120       120              0
## 125       125              0
## 130       130              0
## 135       135              0
## 140       140              0
## 145       145              0
## 150       150              0
## 155       155              0
## 200       200              0
## 205       205              0
## 210       210              0
## 215       215              0
## 220       220              0
## 225       225              0
## 230       230              0
## 235       235              0
## 240       240              0
## 245       245              0
## 250       250              0
## 255       255              0
## 300       300              0
## 305       305              0
## 310       310              0
## 315       315              0
## 320       320              0
## 325       325              0
## 330       330              0
## 335       335              0
## 340       340              0
## 345       345              0
## 350       350              0
## 355       355              0
## 400       400              0
## 405       405              0
## 410       410              0
## 415       415              0
## 420       420              0
```

```
## 425       425              0
## 430       430              0
## 435       435              0
## 440       440              0
## 445       445              0
## 450       450              0
## 455       455              0
## 500       500              0
## 505       505              0
## 510       510              0
## 515       515              0
## 520       520              0
## 525       525              0
## 530       530              0
## 535       535              0
## 540       540              0
## 545       545              0
## 550       550              0
## 555       555              0
## 600       600              0
## 605       605              0
## 610       610              0
## 615       615              0
## 620       620              0
## 625       625              0
## 630       630              0
## 635       635              0
## 640       640              0
## 645       645              0
## 650       650              8
## 655       655             13
## 700       700              7
## 705       705             13
## 710       710             14
## 715       715              0
## 720       720              0
## 725       725             12
## 730       730              0
## 735       735              0
## 740       740             15
## 745       745             19
## 750       750             19
## 755       755             28
## 800       800             41
## 805       805             25
## 810       810             32
## 815       815             13
## 820       820             45
## 825       825             33
## 830       830             37
## 835       835             19
## 840       840             51
## 845       845             60
## 850       850             16
```

```
## 855       855             43
## 900       900             20
## 905       905              8
## 910       910             31
## 915       915             15
## 920       920             16
## 925       925              0
## 930       930              0
## 935       935              0
## 940       940              0
## 945       945              0
## 950       950              0
## 955       955              0
## 1000     1000              0
## 1005     1005              0
## 1010     1010              0
## 1015     1015              0
## 1020     1020              0
## 1025     1025              0
## 1030     1030              0
## 1035     1035              0
## 1040     1040              0
## 1045     1045              0
## 1050     1050              0
## 1055     1055              0
## 1100     1100              0
## 1105     1105              0
## 1110     1110              0
## 1115     1115              0
## 1120     1120              0
## 1125     1125              0
## 1130     1130              0
## 1135     1135              0
## 1140     1140              0
## 1145     1145              0
## 1150     1150              0
## 1155     1155              0
## 1200     1200              0
## 1205     1205              0
## 1210     1210              6
## 1215     1215             10
## 1220     1220              0
## 1225     1225              0
## 1230     1230              0
## 1235     1235              0
## 1240     1240              0
## 1245     1245              0
## 1250     1250              0
## 1255     1255              0
## 1300     1300              0
## 1305     1305              0
## 1310     1310              0
## 1315     1315              0
## 1320     1320              0
```

```
## 1325     1325             0
## 1330     1330             0
## 1335     1335             0
## 1340     1340             0
## 1345     1345             0
## 1350     1350             0
## 1355     1355             0
## 1400     1400             0
## 1405     1405             0
## 1410     1410             0
## 1415     1415             0
## 1420     1420             0
## 1425     1425             0
## 1430     1430             0
## 1435     1435             0
## 1440     1440             0
## 1445     1445             0
## 1450     1450             0
## 1455     1455             0
## 1500     1500             0
## 1505     1505             0
## 1510     1510             0
## 1515     1515             0
## 1520     1520             0
## 1525     1525             0
## 1530     1530             0
## 1535     1535             0
## 1540     1540             0
## 1545     1545             0
## 1550     1550             0
## 1555     1555             0
## 1600     1600             0
## 1605     1605             0
## 1610     1610             0
## 1615     1615             0
## 1620     1620             0
## 1625     1625             0
## 1630     1630             0
## 1635     1635             0
## 1640     1640             0
## 1645     1645             0
## 1650     1650             0
## 1655     1655             0
## 1700     1700             0
## 1705     1705             0
## 1710     1710             0
## 1715     1715             7
## 1720     1720             7
## 1725     1725             0
## 1730     1730             7
## 1735     1735             7
## 1740     1740            26
## 1745     1745             7
## 1750     1750             0
```

```
## 1755      1755              10
## 1800      1800              15
## 1805      1805              18
## 1810      1810              26
## 1815      1815              25
## 1820      1820              24
## 1825      1825               9
## 1830      1830              33
## 1835      1835              26
## 1840      1840              34
## 1845      1845              42
## 1850      1850              33
## 1855      1855              30
## 1900      1900              33
## 1905      1905              30
## 1910      1910               8
## 1915      1915               8
## 1920      1920               7
## 1925      1925               0
## 1930      1930               0
## 1935      1935               0
## 1940      1940               0
## 1945      1945               0
## 1950      1950               0
## 1955      1955               0
## 2000      2000               0
## 2005      2005               0
## 2010      2010               0
## 2015      2015               0
## 2020      2020               0
## 2025      2025               0
## 2030      2030               0
## 2035      2035               0
## 2040      2040               0
## 2045      2045               0
## 2050      2050               0
## 2055      2055               0
## 2100      2100               0
## 2105      2105               0
## 2110      2110               0
## 2115      2115               0
## 2120      2120               0
## 2125      2125               0
## 2130      2130               0
## 2135      2135               0
## 2140      2140               0
## 2145      2145               0
## 2150      2150               0
## 2155      2155               0
## 2200      2200               0
## 2205      2205               0
## 2210      2210               0
## 2215      2215               0
## 2220      2220               0
```

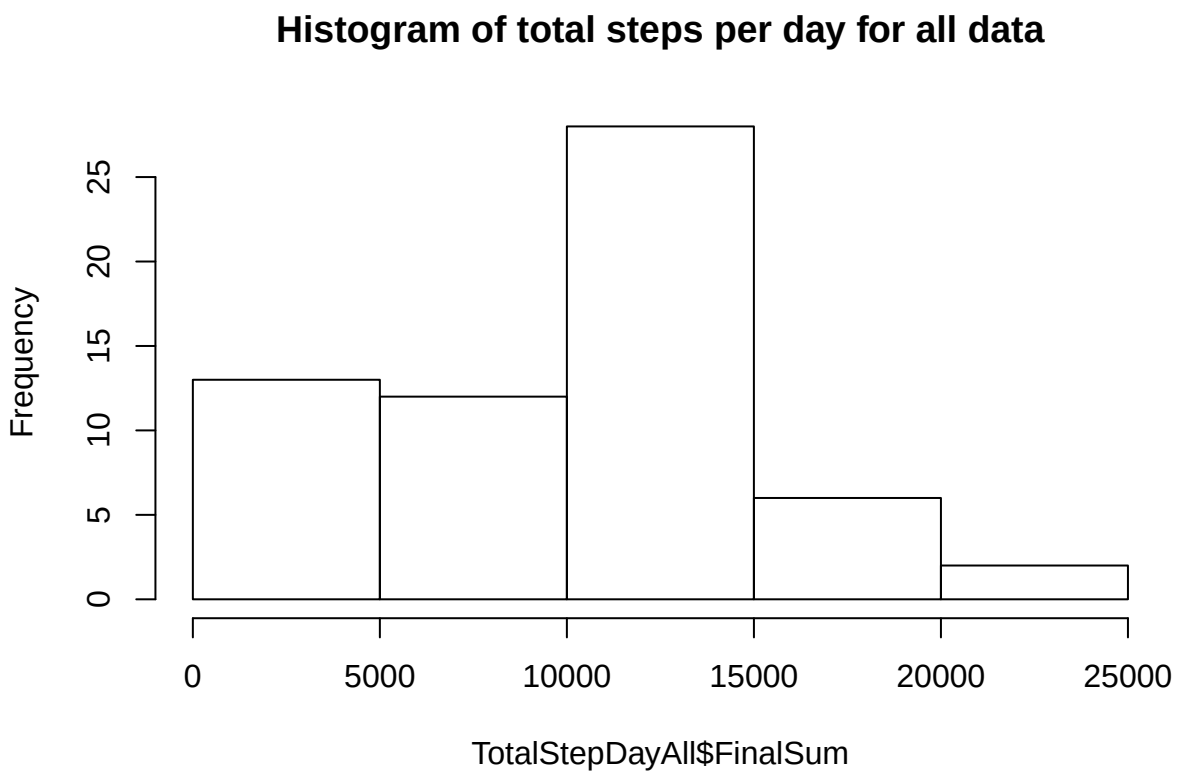
```
## 2225     2225             0
## 2230     2230             0
## 2235     2235             0
## 2240     2240             0
## 2245     2245             0
## 2250     2250             0
## 2255     2255             0
## 2300     2300             0
## 2305     2305             0
## 2310     2310             0
## 2315     2315             0
## 2320     2320             0
## 2325     2325             0
## 2330     2330             0
## 2335     2335             0
## 2340     2340             0
## 2345     2345             0
## 2350     2350             0
## 2355     2355             0
```

```
#Calculate Final Mean average for each day
CombinedFile <- merge(With, MeanStepInterval, by = "interval")
CombinedFile <- merge(CombinedFile, MeanStepDay, by = "date")
CombinedFile <- mutate(CombinedFile, FinalMean = ifelse(is.na(CombinedFile$DateMean),
            CombinedFile$IntervalMean, CombinedFile$DateMean))

#Calculate Final Median average for each day
CombinedFile <- merge(CombinedFile, MedianStepInterval, by = "interval")
CombinedFile <- merge(CombinedFile, MedianStepDay, by = "date")
CombinedFile <- mutate(CombinedFile, FinalMedian = ifelse(is.na(CombinedFile$DateMedian),
            CombinedFile$IntervalMedian, CombinedFile$DateMedian))

#View(CombinedFile)
```

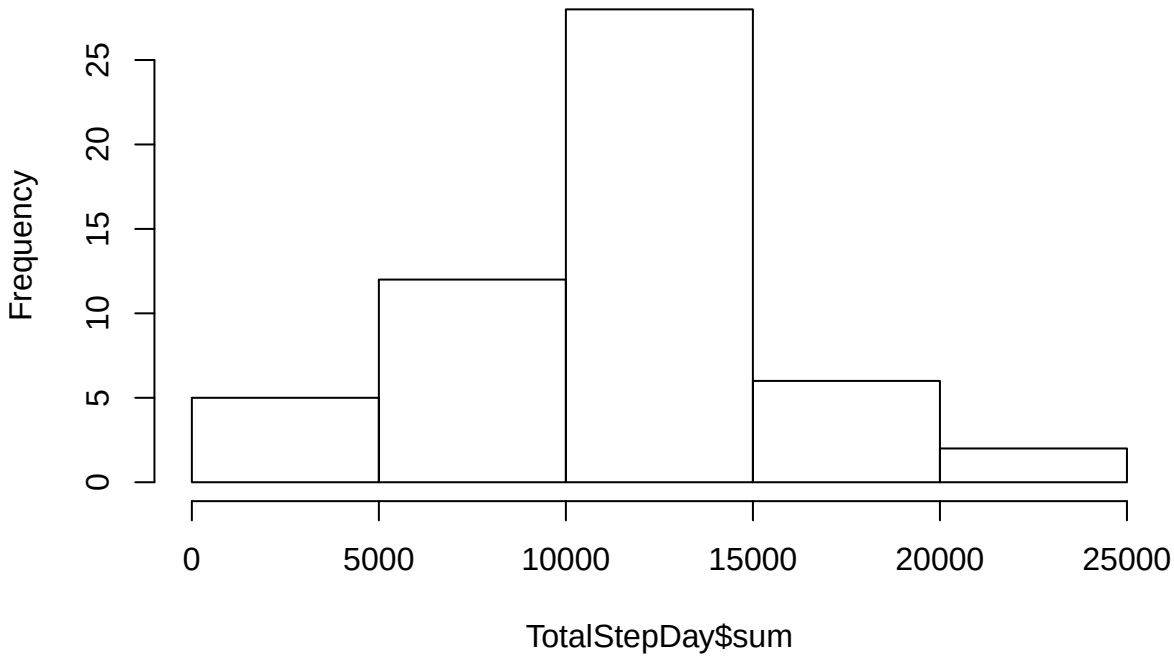
```
#Count total steps per day
ts2 <- tapply(With$steps, With$date , sum)

TotalStepDayAll <- data.frame(date = names(ts2), sum = ts2)

TotalStepDayAll<- mutate(TotalStepDayAll, FinalSum = ifelse(is.na(TotalStepDayAll$sum),
            0, TotalStepDayAll$sum))



#View(TotalStepDayAll)
```

The frequency of total steps per day with NA data that was imputed shows more frequency counts in 0 - 5000 step range than before. The rest of buckets are identical.

```r
#Histogram of total steps per day
hist(TotalStepDayAll$FinalSum, main = "Histogram of total steps per day for all data")
```



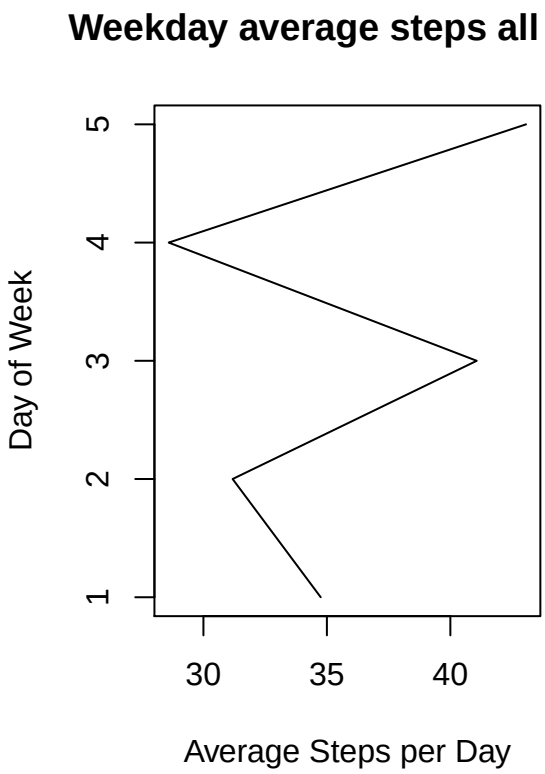

```r
hist(TotalStepDay$sum, main = "Histogram of total steps per day for only complete data")
```

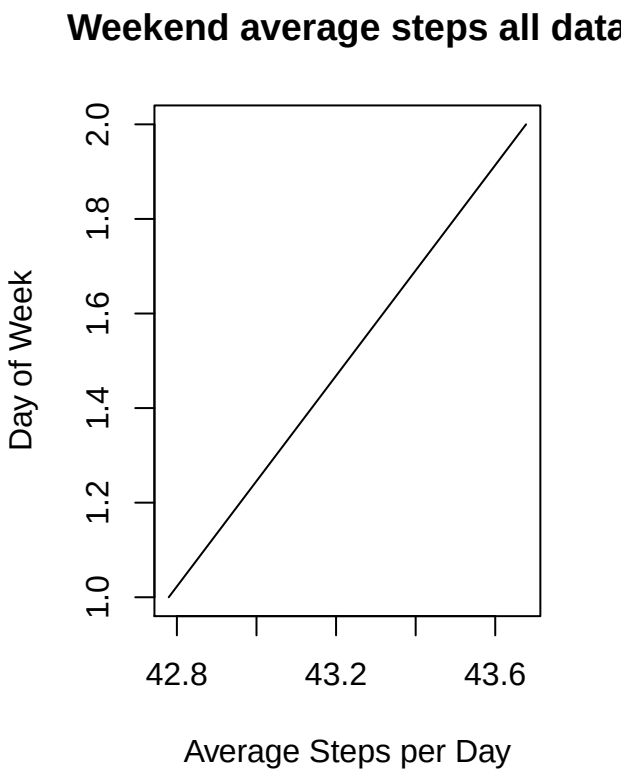



```
#CombinedFile

#Create weekend subset
SatSun <- c("Saturday", "Sunday")
weekend <- subset(CombinedFile, weekday %in% (SatSun))

#Create weekday subset
Wkdays <- c("Monday","Tuesday","Wednesday","Thursday","Friday")
weekdays <- subset(CombinedFile, weekday %in% (Wkdays))
```

```
#Get the 5 minute interval mean for NA rows
naSet <- CombinedFile[is.na(CombinedFile$steps),] %>% select(steps,date,interval, weekday, weekdayNum, I
#naFiveInterval <- subset(naSet, interval == 5)

#Calculate five minute interal for NA file
naFiveInterval <- naSet  %>% mutate(FiveInterval = ((FinalMean / interval) * 1.0) * 5) %>% select(steps

#Combine the two five interval files into one
CombinedFiveInterval <- rbind(FiveInterval, naFiveInterval)

CombinedFiveIntervalFinal  <- na.exclude(CombinedFiveInterval)
CombinedFiveIntervalFinal <- subset(CombinedFiveIntervalFinal, interval > 0 )


#Create weekday subset
Wkdays <- c("Monday","Tuesday","Wednesday","Thursday","Friday")
```

```
weekdays <- subset(CombinedFiveIntervalFinal, weekday %in% (Wkdays))

w1 <- tapply(weekdays$steps, weekdays$weekdayNum,  mean)

weekdayAverage <- data.frame(day = names(w1), mean = w1)

#View(weekdayAverage)

#Create weekend subset
SatSun <- c("Saturday", "Sunday")
weekend <- subset(CombinedFiveIntervalFinal, weekday %in% (SatSun))

w2 <- tapply(weekend$steps, weekend$weekdayNum,  mean)

weekendAverage <- data.frame(day = names(w2), mean = w2)

#View(weekendAverage)




#Display time series of average steps for week days and then weekend days
par(mfrow=c(1,2))

plot(weekdayAverage$mean, weekdayAverage$day, type = "l", xlab = "Average Steps per Day", ylab = "Day o


plot(weekendAverage$mean, weekendAverage$day, type = "l", xlab = "Average Steps per Day", ylab = "Day o
```


Weekday average steps all data
Day of Week
1
2
3
4
5
30
35
40
Average Steps per Day


Weekend average steps all data
Day of Week
1.0
1.2
1.4
1.6
1.8
2.0
42.8
43.2
43.6
Average Steps per Day