fitbit0analysis

Deepak Grewal

4/1/2022

```
library(dplyr)
library(tidyverse)
```

Reading in data

```
activity <- read_csv("activity.csv")
```

analyse Activity

What is mean total number of steps taken per day?

```
# derive mean and median steps per day
summSteps <- activity %>%  group_by(date) %>%
  summarise(TotalSteps= sum(steps, na.rm = TRUE)) %>% ungroup() %>%
  mutate(allSteps = sum(TotalSteps),meanPerDay= mean(TotalSteps), medianPerDay= median(TotalSteps))

summSteps$meanPerDay[1]
```

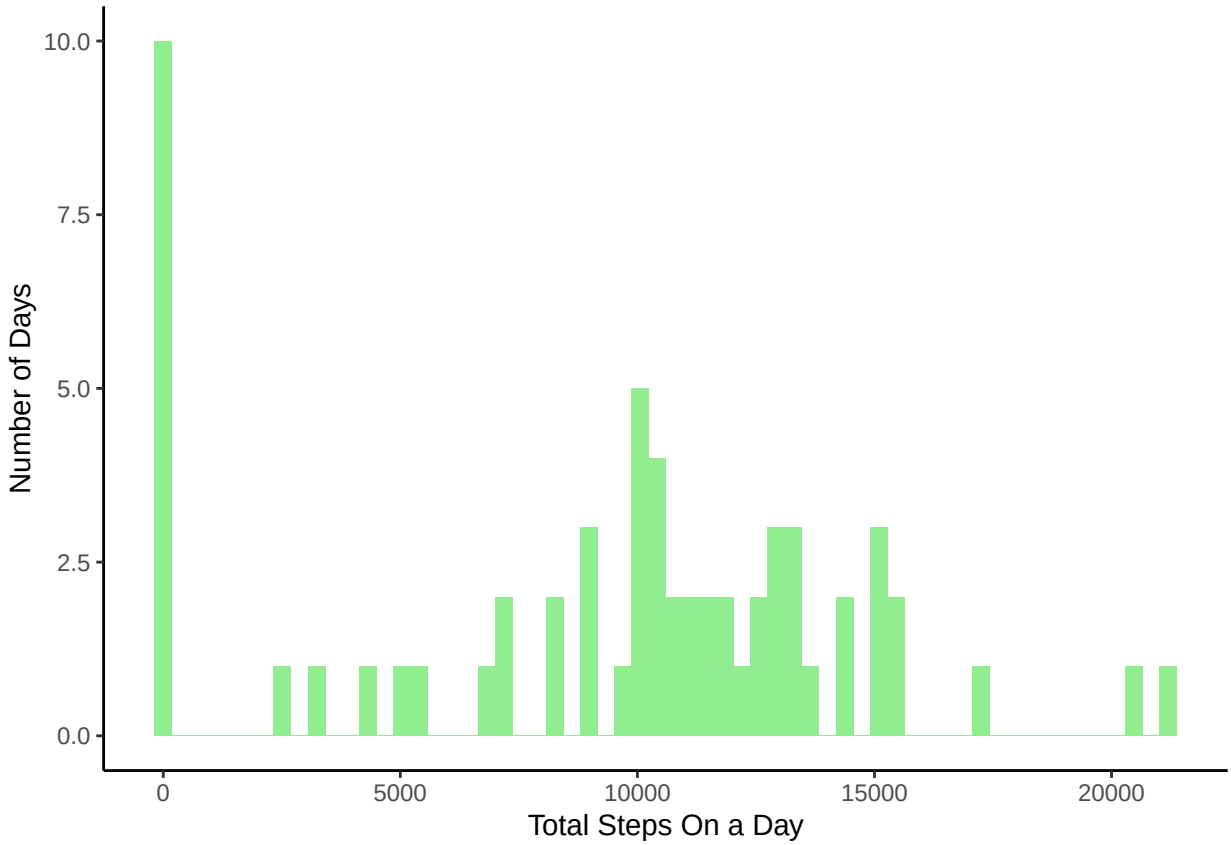
```
## [1] 9354.23
```

Calculate the total number of steps taken per day

```
# calculate total steps on each day
TotalSteps <- activity %>%  group_by(date) %>%
  summarise(TotalSteps= sum(steps, na.rm = TRUE))
```

Make a histogram of the total number of steps taken each day Use the data containing total steps

```
# use histogram to display steps
TotalSteps %>%  ggplot(aes(x= TotalSteps)) +
  geom_histogram(bins = 60, fill='lightgreen')+
  labs(x= "Total Steps On a Day" , y= "Number of Days")+
  theme_classic()
```



Calculate and report the mean and median of the total number of steps taken per day

```
# derive mean and median steps per day
summSteps <- activity %>%  group_by(date) %>%
  summarise(TotalSteps= sum(steps, na.rm = TRUE)) %>% ungroup() %>%
  mutate(allSteps = sum(TotalSteps),meanPerDay= mean(TotalSteps), medianPerDay= median(TotalSteps))
```

Mean steps Per Day

```
summSteps$meanPerDay[1]
```

```
## [1] 9354.23
```

Median steps Per Day

```
summSteps$medianPerDay[1]
```

```
## [1] 10395
```

What is the average daily activity pattern? Make a time series plot for the 5-minute interval (x-axis) and the average number of steps taken, averaged across all days (y-axis)

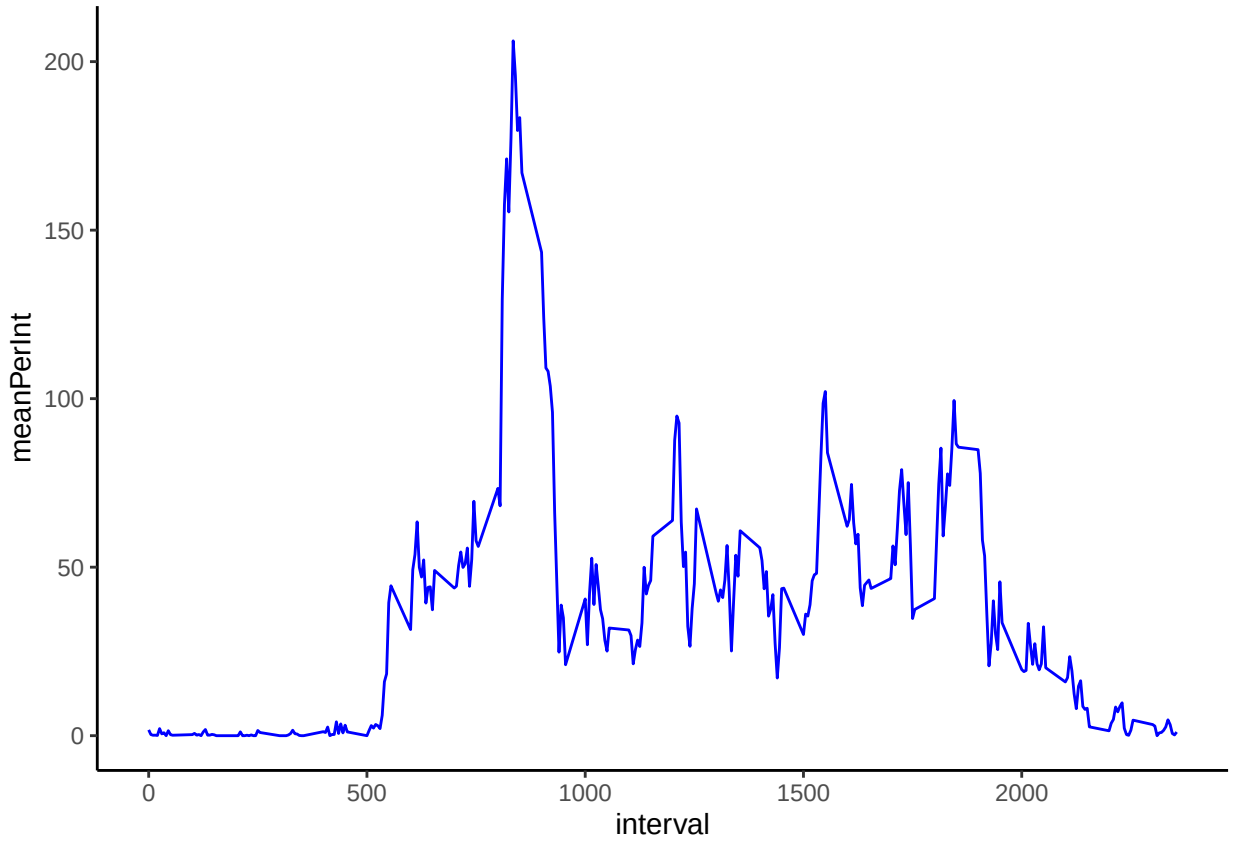
first get average steps for the intervals

```
# derive mean and median steps per 5 minute interval
summStepsInt <- activity %>%  group_by(interval) %>%
  summarise(TotalSteps= sum(steps, na.rm = TRUE),meanPerInt= mean(steps, na.rm = TRUE))
```

Plot the data

```
#plot average steps across the time points
summStepsInt %>%  ggplot(aes(x= interval, y= meanPerInt)) + geom_line(color='blue') +
```

```
  theme_classic()
```



Which 5-minute interval, on average across all the days in the dataset, contains the maximum number of steps?

Maximum Steps Interval:

```
# get the maximum average step time interval

maxPerInt <- summStepsInt %>% filter(meanPerInt== max(meanPerInt))

maxPerInt$interval
```

```
## [1] 835
```

Impute Missing Data and Analyse activity

Following Number of records are missing

```
#get total records missing records
missing_rec <- activity %>% filter(is.na(steps) | is.na(date) | is.na(interval) ) %>% count()
missing_rec$n
```

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

Replace missing with median values

```
# replace the missing steps for the interval by median steps for that interval
#1- calculate the median steps per interval
summStepsInt2 <- activity %>%  group_by(interval) %>%
```

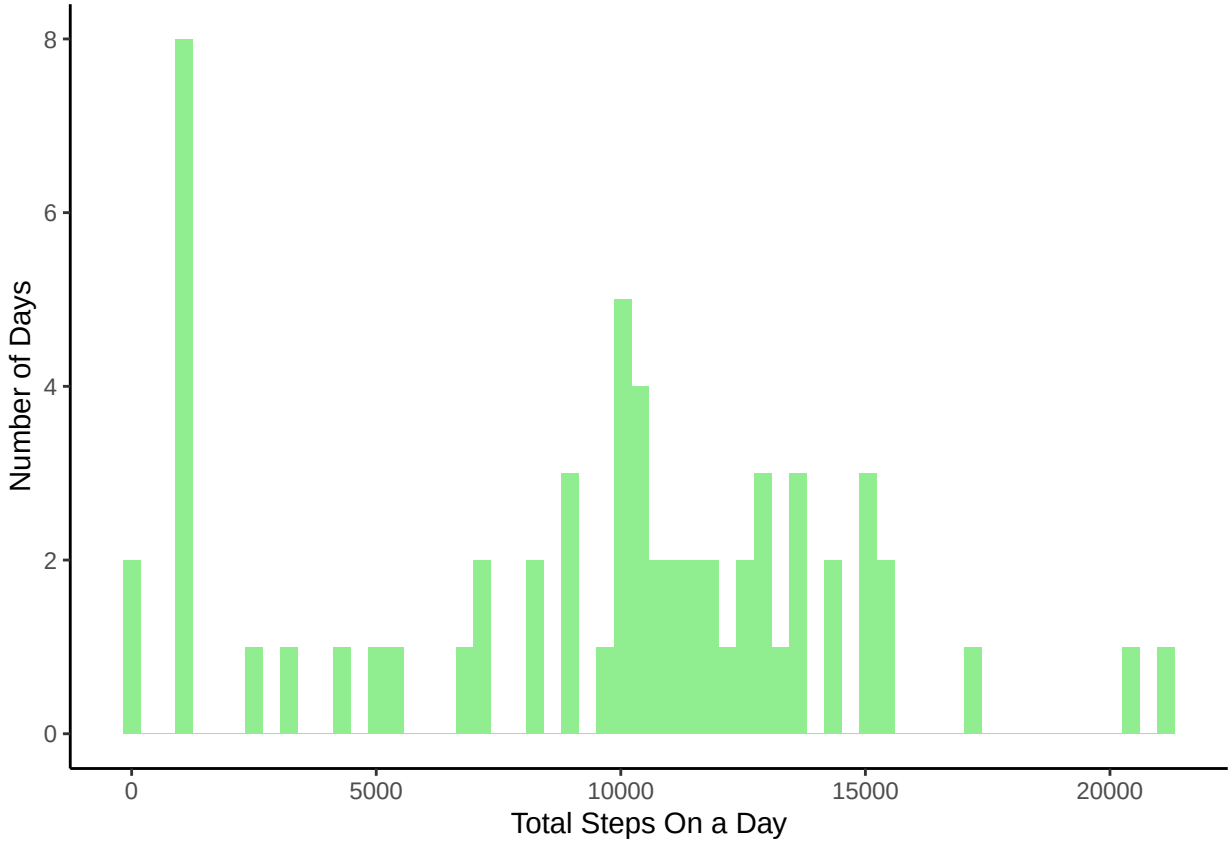
```
  summarise(medianPerInt= median(steps, na.rm = TRUE))
#2- replace missing with median value

activityImputed <- activity %>% left_join(summStepsInt2 , by= c("interval")) %>%
  mutate(steps=if_else(is.na(steps),medianPerInt, steps )) %>% select(-c(medianPerInt))
```

histogram of Imputed steps

```
# calculate total steps on each day
TotalSteps <- activityImputed %>%  group_by(date) %>%
  summarise(TotalSteps= sum(steps, na.rm = TRUE))

# use histogram to display steps
TotalSteps %>%  ggplot(aes(x= TotalSteps)) +
  geom_histogram(bins = 60, fill='lightgreen')+
  labs(x= "Total Steps On a Day" , y= "Number of Days")+
  theme_classic()
```



Mean and Median of imputed data

```
# derive mean and median steps per day
summSteps <- activityImputed %>%  group_by(date) %>%
  summarise(TotalSteps= sum(steps, na.rm = TRUE)) %>% ungroup() %>%
  mutate(allSteps = sum(TotalSteps),meanPerDay= mean(TotalSteps), medianPerDay= median(TotalSteps))

summSteps$meanPerDay[1]
```

```
## [1] 9503.869
```

```
summSteps$medianPerDay[1]
```

```
## [1] 10395
```

As you can see above median Did not change(as median was used to replace imputed) , but mean changed

Analyse Weekend Vs Weekdays

1. add factor column for type of Day- Weekday or Weekend and derive data for plot

```
# add factor
activityDay <- activity %>% mutate(day=as.factor(if_else(
  weekdays(date) %in% c("Monday" ,   "Tuesday"   , "Wednesday", "Thursday" , "Friday"), "Weekday", "Weeke


# derive mean and median steps per 5 minute interval
summStepsInt <- activityDay %>%  group_by(interval, day) %>%
  summarise(TotalSteps= sum(steps, na.rm = TRUE),meanPerInt= mean(steps, na.rm = TRUE))
```

2. Display plot

```
#plot average steps across the time points
summStepsInt %>%  ggplot(aes(x= interval, y= meanPerInt, color=day)) + geom_line() +
  theme_classic()
```