

Reproducible_research

durvesh

5/6/2021

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.3      v purrr   0.3.4
## v tibble  3.1.1      v dplyr  1.0.5
## v tidyr   1.1.3      v stringr 1.4.0
## v readr   1.4.0      v forcats 0.5.1
```

```
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
```

```
# 1.Code for reading in the dataset
```

```
#zipF<- "C:\\Users\\durvesh\\Downloads\\repdata_data_activity.zip"
#outDir<-"C:\\Users\\durvesh\\Documents\\unzipfolder1"
#unzip(zipF,exdir=outDir)

t1<-read.csv("..\unzipfolder1\\activity.csv")
```

```
#1 Code for Processing the Data Set
class(t1$date)
```

```
## [1] "character"
```

```
typeof(t1$date)
```

```
## [1] "character"
```

```
t1$date<-as.Date(t1$date)

t1$steps[is.na(t1$steps)]<-0

t1$steps <- as.numeric(t1$steps)
d1<-weekdays(t1$date)

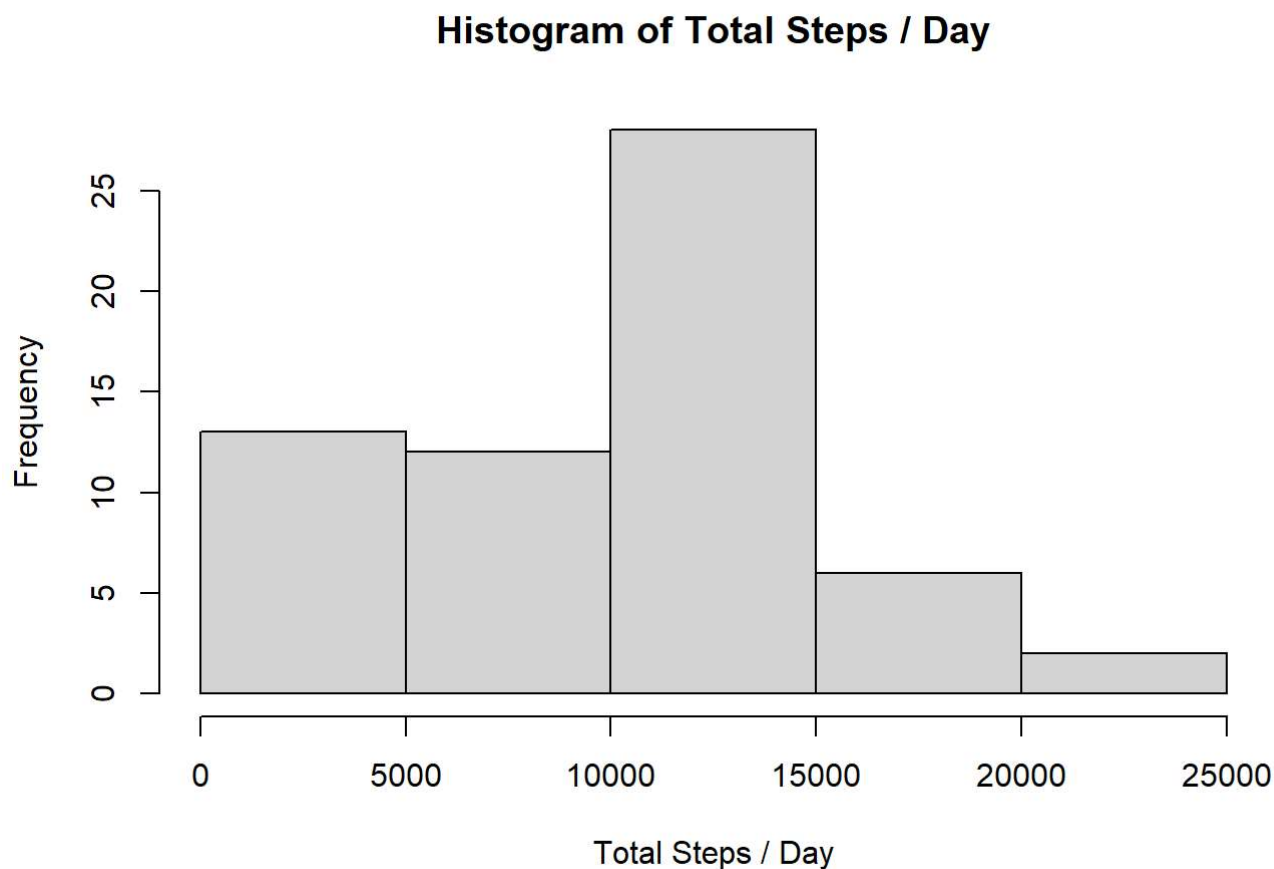
t2 <- cbind(t1,d1)
head(t2)
```

```
##  steps      date interval    d1
## 1    0 2012-10-01         0 Monday
## 2    0 2012-10-01         5 Monday
## 3    0 2012-10-01        10 Monday
## 4    0 2012-10-01        15 Monday
## 5    0 2012-10-01        20 Monday
## 6    0 2012-10-01        25 Monday
```

```
steps_by_day <- aggregate(steps~date,sum,data=t2)
```

#2. Histogram of Total Number of Steps

```
hist(steps_by_day$steps,xlab = "Total Steps / Day",ylab= "Frequency", main="Histogram of Total S  
teps / Day")
```



3. Mean and median number of steps taken each day

```
cat("The mean of number of steps per day is:",mean(steps_by_day$steps))
```

```
## The mean of number of steps per day is: 9354.23
```

```
cat("\n")
```

```
cat("-----")
```

```
## -----
```

```
cat("\n")
```

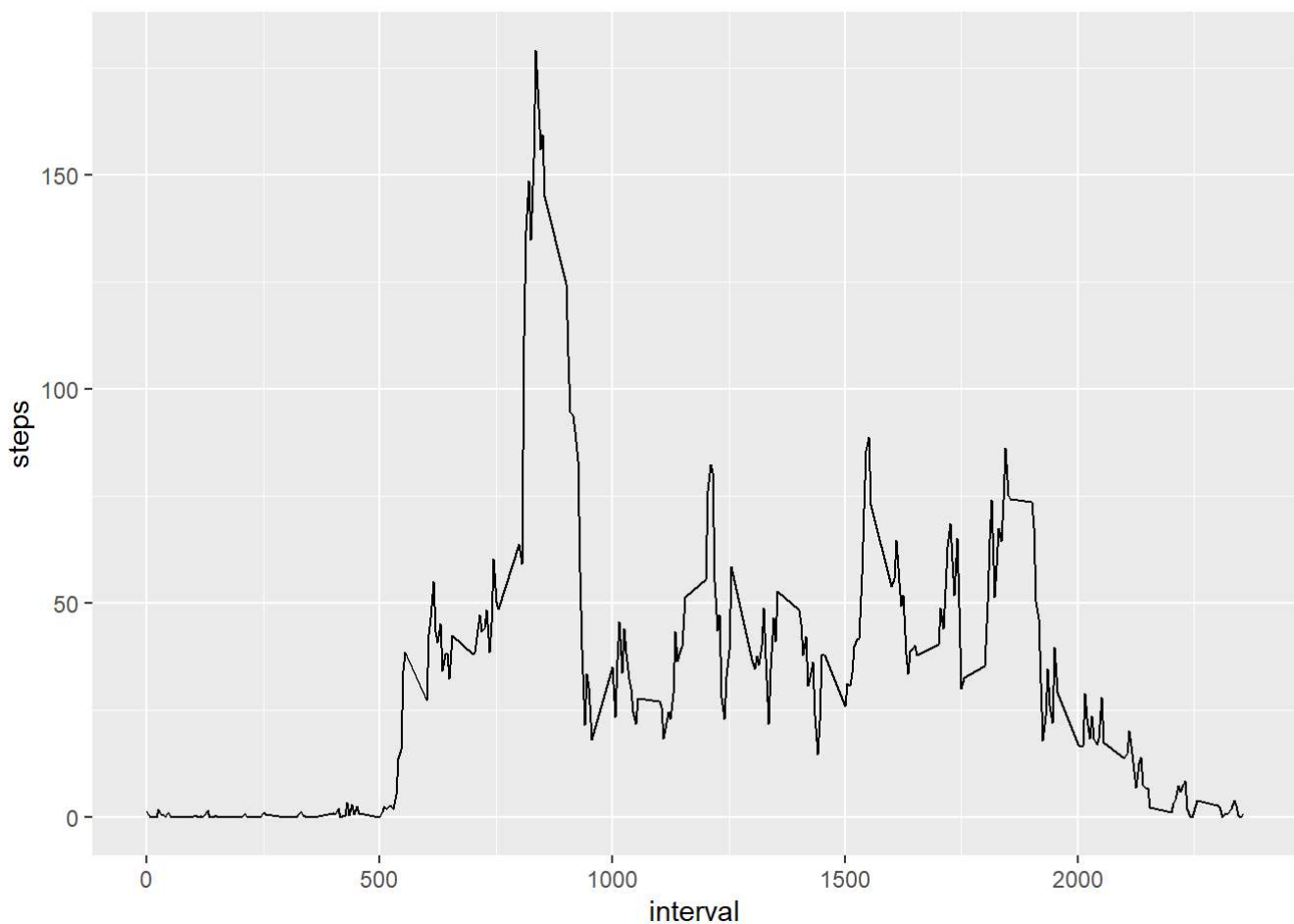
```
cat("The median of number of steps per day is:",median(steps_by_day$steps))
```

```
## The median of number of steps per day is: 10395
```

#4. Time Series Plot of Average No of Steps taken in an Interval

#Aggregate function for mean over all days, for each interval
`agg_interval<-aggregate(steps~interval,data=t2,mean)`

#Plot of Steps by Interval
`ggplot(agg_interval,aes(x=interval,y=steps))+
 geom_line()`



#5 The 5-minute interval that, on average, contains the maximum number of steps

```
cat("The 5 min Interval with max steps:",agg_interval$interval[which.max(agg_interval$steps)])
```

```
## The 5 min Interval with max steps: 835
```

```
t <-read.csv("..\unzipfolder1\\activity.csv")
tn<- t %>%
  group_nest(t$interval)

impute_missing<-function(df){
  df$step_impute =
    if_else(is.na(df$steps),mean(df$steps,na.rm=TRUE),as.double(df$steps))
}

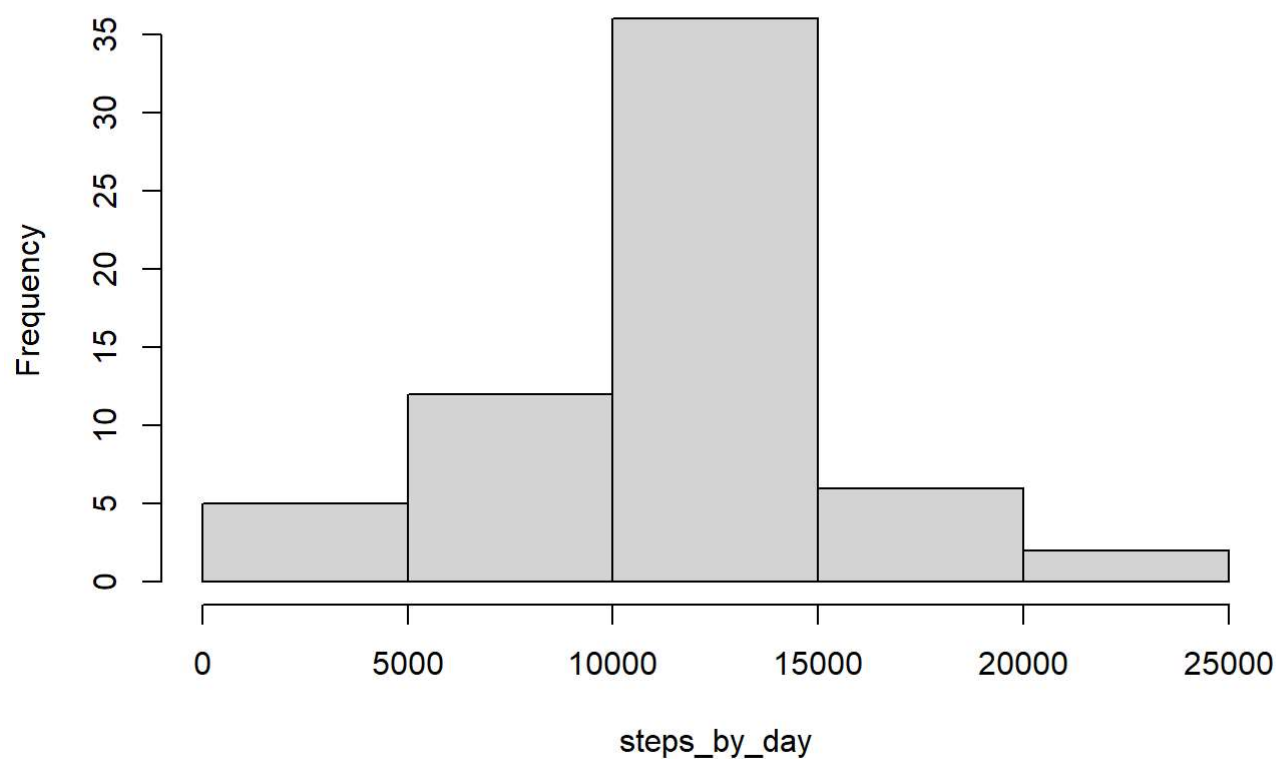
tn <-tn %>%
  mutate(steps_impute = map(data, impute_missing)) %>%
  unnest(cols = c(data, steps_impute))
```

#7. Histogram of the total number of steps taken each day after missing values are imputed

```
steps_by_day_imputed <-aggregate(steps_impute~date,sum,data=tn)
```

```
hist(steps_by_day_imputed$steps_impute,xlab="steps_by_day",main="Histogram with Mean Imputation  
of Steps")
```

Histogram with Mean Imputation of Steps



#8. Panel plot comparing the average number of steps taken per 5-minute interval across weekdays and weekends

```
tn$dateType <- ifelse(as.POSIXlt(tn$date)$wday %in% c(0,6), 'weekend', 'weekday')
```

```
Avgsteps_int_dateType<-aggregate(steps_impute~dateType+interval,data=tn,mean)
```

```
ggplot(Avgsteps_int_dateType, aes(interval, steps_impute)) +  
  geom_line() +  
  facet_grid(dateType ~ .) +  
  xlab("5-minute interval") +  
  ylab("Average number of steps")
```

