

Reproducible Research: Course Project 1

Read in the .csv file.

```
data <- read.csv("activity.csv", colClasses = c("numeric", "character", "integer"))
```

Format date and group data by day using group_by function in dplyr

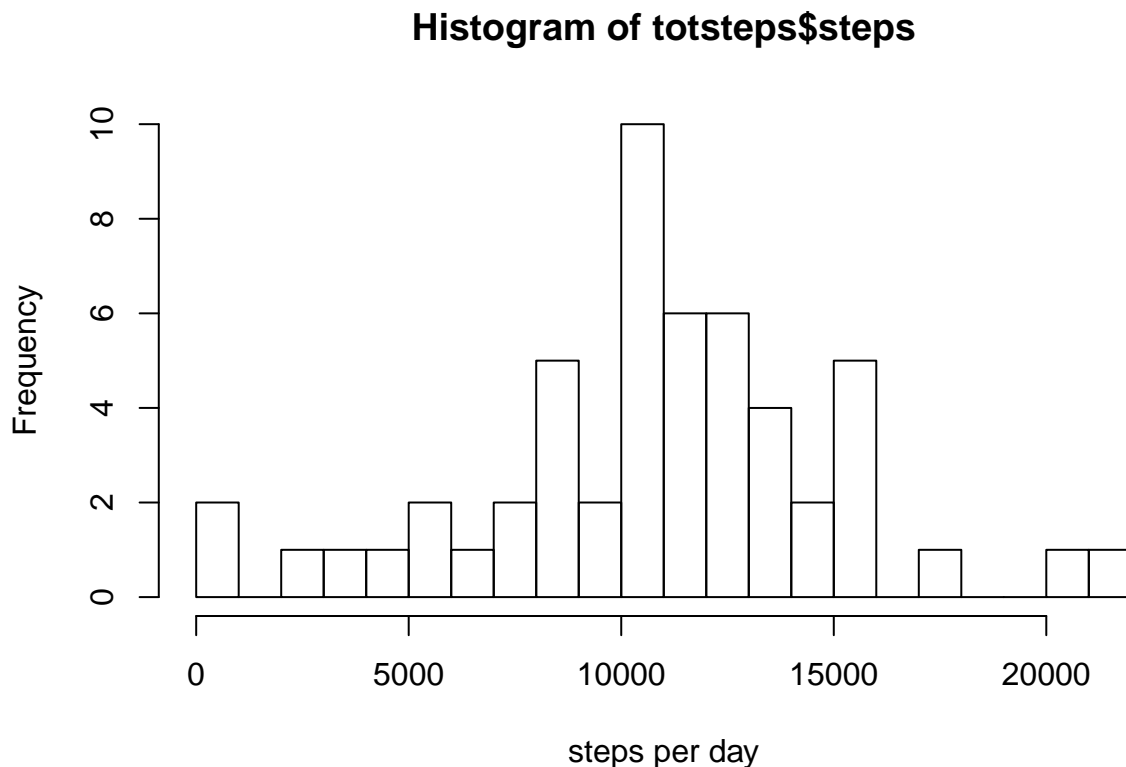
```
library(dplyr)
```

```
##  
## 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
```

```
totsteps <- data %>%  
  filter(!is.na(steps)) %>%  
  group_by(date) %>%  
  summarize(steps = sum(steps))
```

Plot histogram of total steps per day

```
hist(totsteps$steps, breaks=20, xlab='steps per day' )
```



```
dev.copy(png, file="plot_step2.png", width=480, height=480)
```

```
## quartz_off_screen
##      3
```

```
dev.off()
```

```
## pdf
##      2
```

Calculate mean and median steps

```
median(totsteps$steps)
```

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

```
mean(totsteps$steps)
```

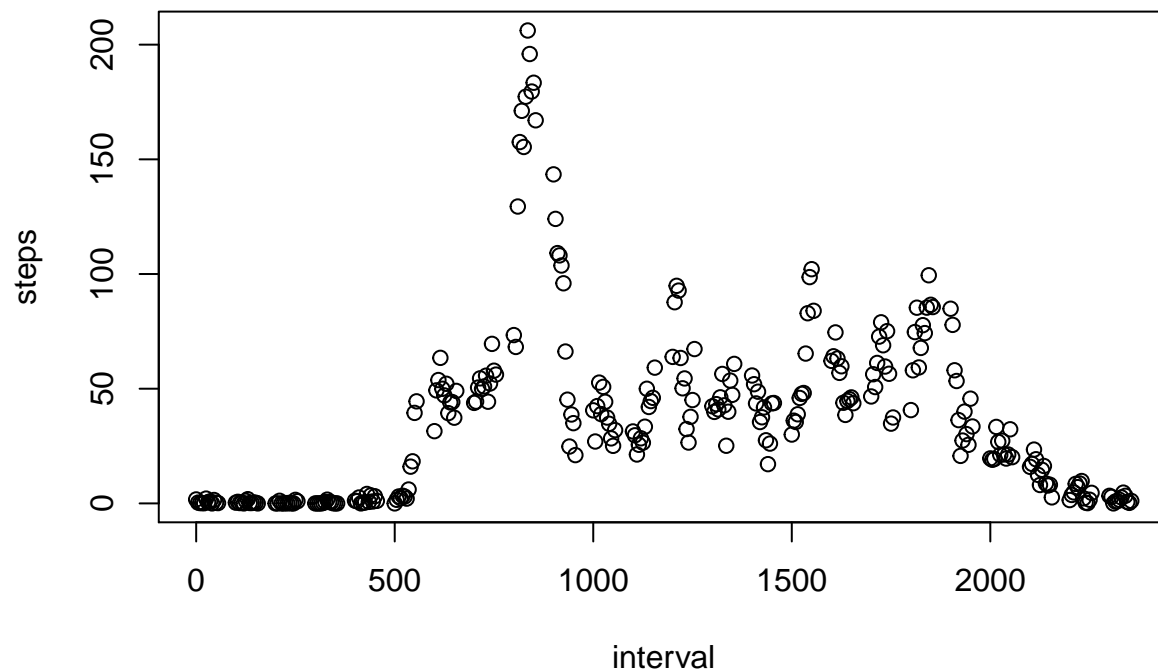
```
## [1] 10766.19
```

Group data by interval and calculate average number of steps

```
int <- data %>%
  filter(!is.na(steps)) %>%
  group_by(interval) %>%
  summarize(steps = mean(steps))
```

Plot steps taken as function of interval

```
plot(int)
```



```
dev.copy(png, file="plot_step4.png", width=480, height=480)
```

```
## quartz_off_screen
##      3
```

```
dev.off()
```

```
## pdf
##      2
```

Find the interval where the maximum steps are taken, on average

```
int[which.max(int$steps),]
```

```
## # A tibble: 1 x 2
##   interval  steps
##   <int>    <dbl>
## 1      835 206.1698
```

Replace missing values with average number of steps per interval, and print total number of missing values

```
alldata <- data
miss <- is.na(alldata$steps)
print(sum(miss))
```

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

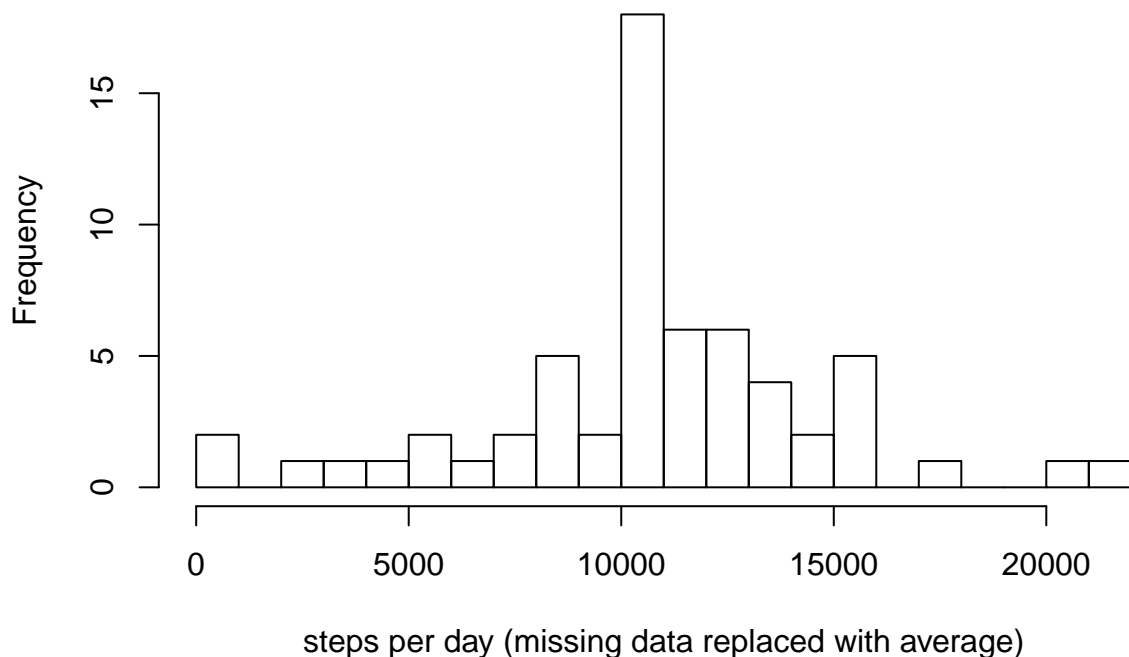
```
avgval <- tapply(alldata$steps, alldata$interval, mean, na.rm=TRUE, simplify=TRUE)
alldata$steps[miss] <- avgval[as.character(alldata$interval[miss])]
```

Group by date again, and plot new histogram with imputed data included

```
totsteps2 <- alldata %>%
  filter(!is.na(steps)) %>%
  group_by(date) %>%
  summarize(steps = sum(steps))
```

```
hist(totsteps2$steps, breaks=20, xlab='steps per day (missing data replaced with average)')
```

Histogram of totsteps2\$steps



```
dev.copy(png, file="plot_step7.png", width=480, height=480)
```

```
## quartz_off_screen
## 3
```

```
dev.off()
```

```
## pdf  
## 2
```

Select data from Saturday, Sunday

```
library(lubridate)
```

```
##
```

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

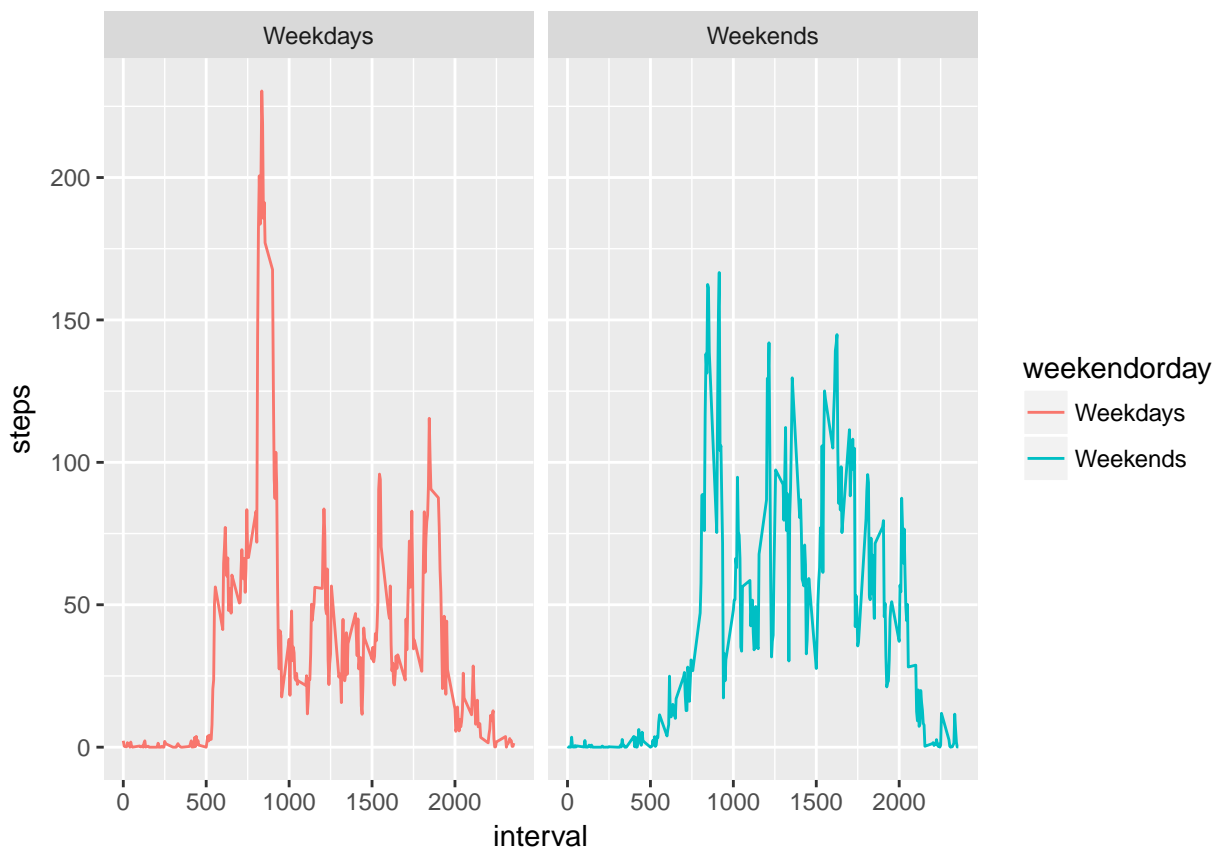
```
## The following object is masked from 'package:base':
```

```
##
```

```
## date
```

```
alldata <- mutate(alldata, weekendorday = ifelse(weekdays(ymd(alldata$date)) == "Saturday" | weekdays(y  
alldata$weekendorday <- as.factor(alldata$weekendorday)
```

```
interval_full <- alldata %>% group_by(interval, weekendorday) %>% summarise(steps = mean(steps))  
library(ggplot2)  
pp <- ggplot(interval_full, aes(x=interval, y=steps, color = weekendorday)) +  
  geom_line() +  
  facet_wrap(~weekendorday, ncol = 2, nrow=1)  
print(pp)
```



```
dev.copy(png, file="plot_step8.png", width=480, height=480)
```

```
## quartz_off_screen  
## 3
```

```
dev.off()
```

```
## pdf
```

```
## 2
```