Coursera : Reproducible Research

Peer-graded Assignment: Course Project 1

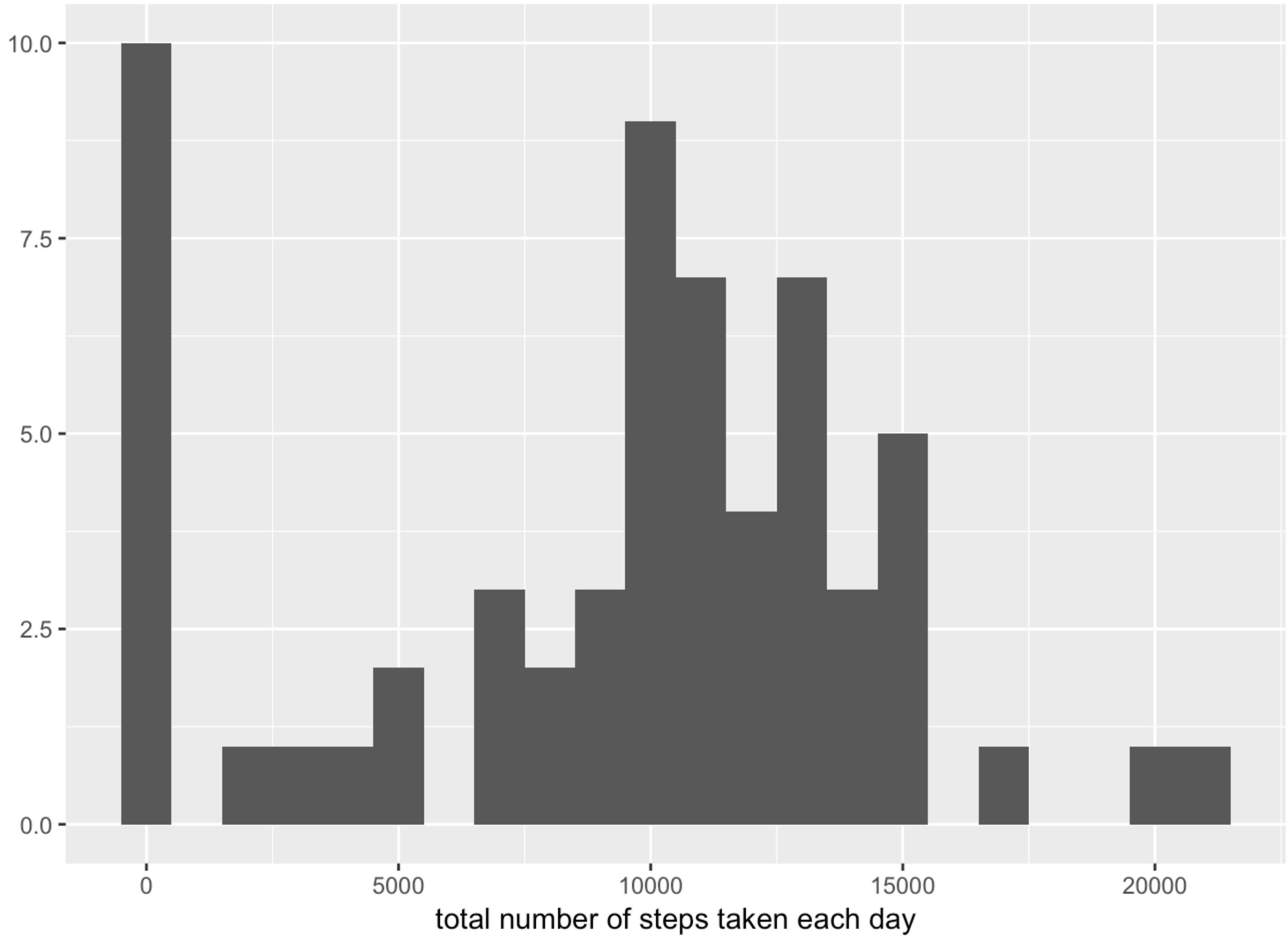
Activity Monitoring Data Analysis

Loading and Preprocessing

```
data <- read.csv("activity.csv")
```

Total steps taken per day

```
library(ggplot2)
total.steps <- tapply(data$steps, data$date, FUN=sum, na.rm=TRUE)
qplot(total.steps, binwidth=1000, xlab="total number of steps taken each day")
```



Mean and Median of steps taken per day

```
mean(total.steps, na.rm=TRUE)
```

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

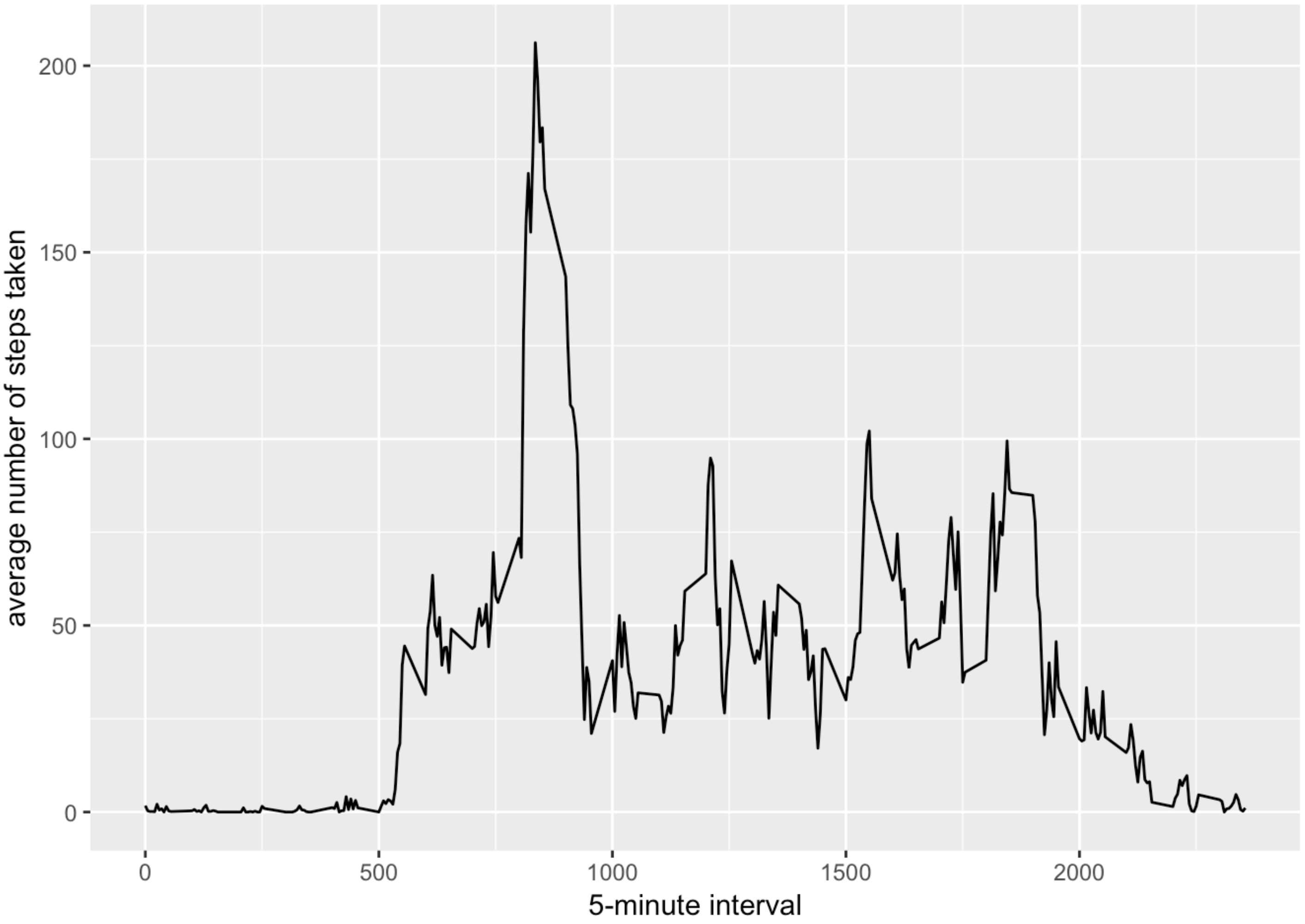
```
median(total.steps, na.rm=TRUE)
```

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

Average number of steps taken per day

```
library(ggplot2)
averages <- aggregate(x=list(steps=data$steps),by=list(interval=data$interval),FUN=me
an, na.rm=TRUE)

ggplot(data=averages, aes(x=interval, y=steps)) +
  geom_line() +
  xlab("5-minute interval") +
  ylab("average number of steps taken")
```



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

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

```
##     interval    steps
## 104      835 206.1698
```

imputing missing data

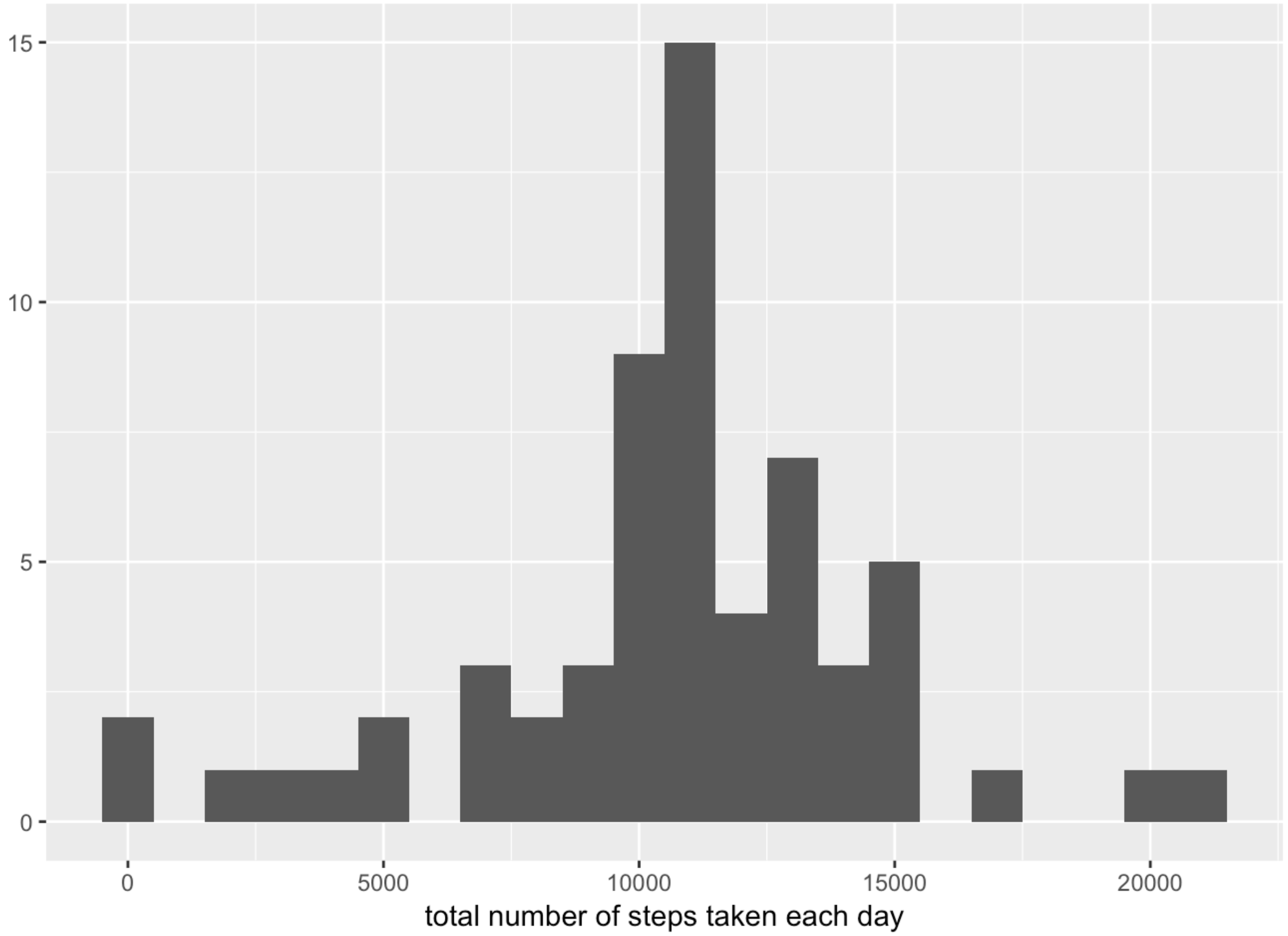
```
missing <- is.na(data$steps)
table(missing)
```

```
## missing
## FALSE  TRUE
## 15264  2304
```

```
# Replace each missing value with the mean value of its 5-minute interval
fill.value <- function(steps, interval) {
  filled <- NA
  if (!is.na(steps))
    filled <- c(steps)
  else
    filled <- (averages[averages$interval==interval, "steps"])
  return(filled)
}
filled.data <- data
filled.data$steps <- mapply(fill.value, filled.data$steps, filled.data$interval)
```

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

```
total.steps <- tapply(filled.data$steps, filled.data$date, FUN=sum)
qplot(total.steps, binwidth=1000, xlab="total number of steps taken each day")
```



```
mean(total.steps)
```

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

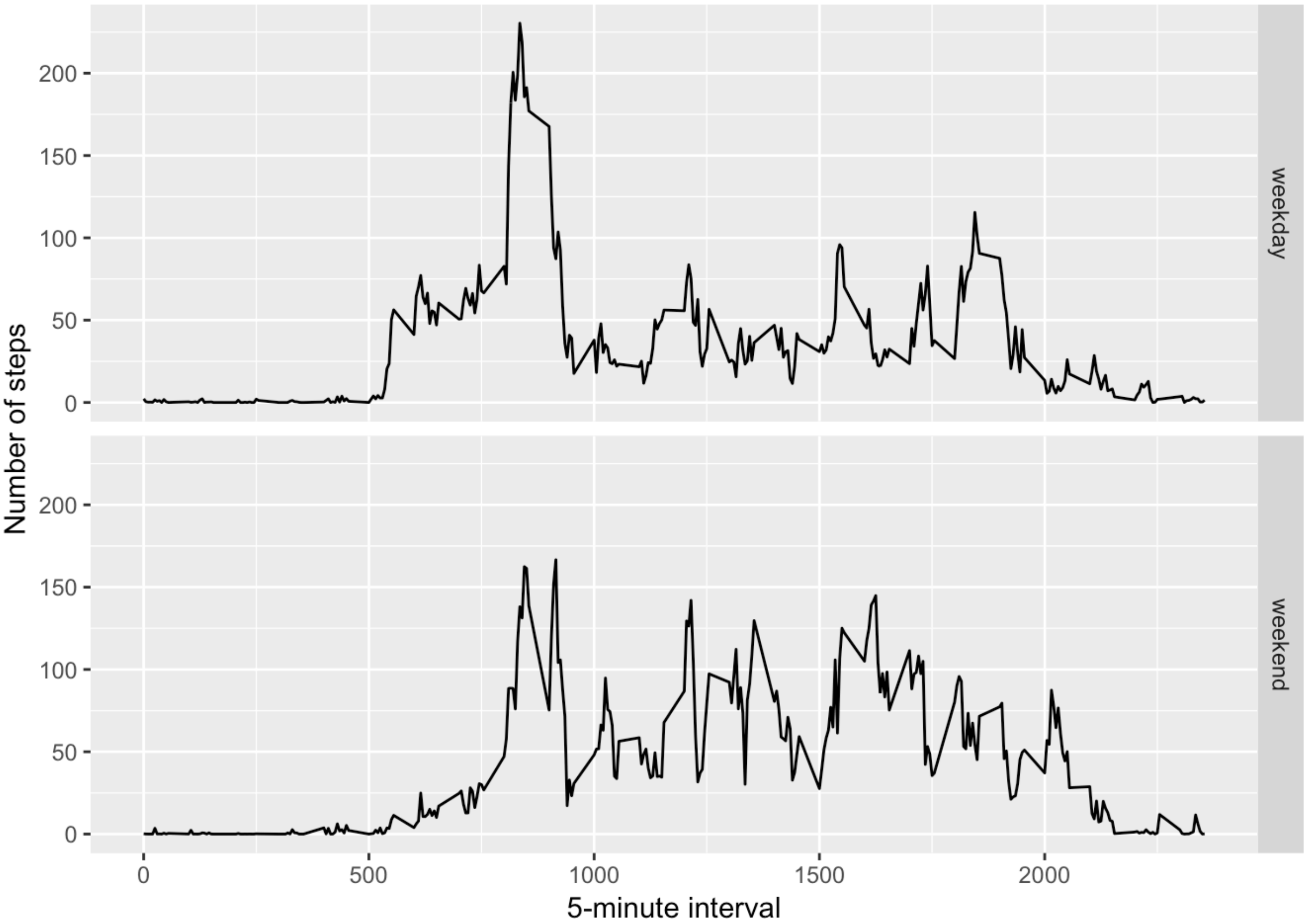
```
median(total.steps)
```

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

Comparing the average number of steps taken per 5-minute interval across weekdays and weekends

```
weekday.or.weekend <- function(date) {
  day <- weekdays(date)
  if (day %in% c("Monday", "Tuesday", "Wednesday", "Thursday", "Friday"))
    return("weekday")
  else if (day %in% c("Saturday", "Sunday"))
    return("weekend")
  else
    stop("invalid date")
}
filled.data$date <- as.Date(filled.data$date)
filled.data$day <- sapply(filled.data$date, FUN=weekday.or.weekend)

averages <- aggregate(steps ~ interval + day, data=filled.data, mean)
ggplot(averages, aes(interval, steps)) + geom_line() + facet_grid(day ~ .) +
  xlab("5-minute interval") + ylab("Number of steps")
```


Number of steps
200
150
100
50
0
weekday
200
150
100
50
0
weekend
0
500
1000
1500
2000
5-minute interval