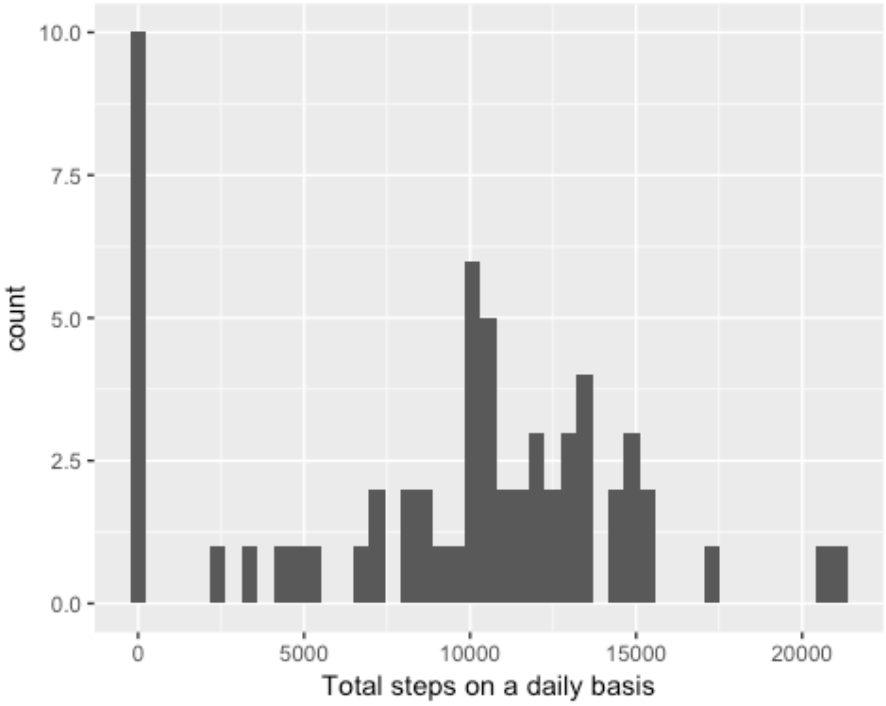
Reproducible Research: Peer Assessment 1

Loading and preprocessing the data

```
setwd("~/Reproducible_research")
unzip("repdata%2Fdata%2Factivity.zip")
activity_data <- read.csv("activity.csv", header= TRUE, sep= ",")
activity_data$date <- as.Date(activity_data$date, "%Y-%m-%d")
library(ggplot2)
```

What is mean total number of steps taken per day?

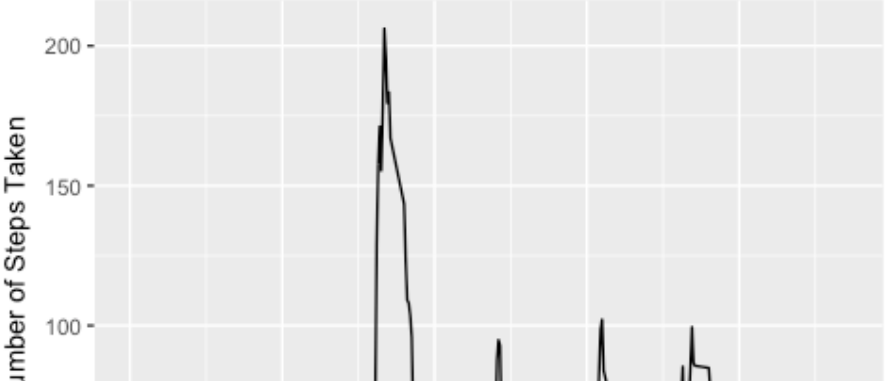
```
daily_steps <- tapply(activity_data$steps, activity_data$date, sum, na.rm= TRUE)
qplot(daily_steps, binwidth= 480, xlab= "Total steps on a daily basis")
```

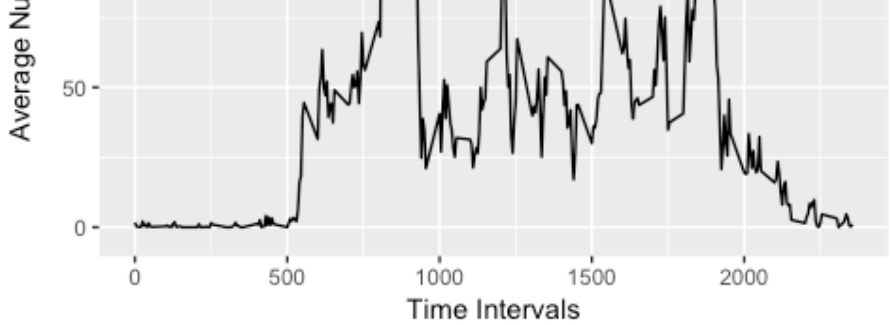



```
#Calculate and report the mean and median of the total number of steps taken per day
steps_mean <- mean(daily_steps)
steps_median <- median(daily_steps)
cbind(steps_mean, steps_median)
##      steps_mean steps_median
## [1,]    9354.23        10395
```

What is the average daily activity pattern?

```
interval_mean <- as.numeric(tapply(activity_data$steps, activity_data$interval, mean, na.rm = TRUE))
plot2 <- data.frame(unique(activity_data$interval), interval_mean)
colnames(plot2) <- c("interval", "mean")
g2 <- ggplot(plot2, aes(x = plot2$interval, y = plot2$mean))
g2 + geom_line() + labs(x = "Time Intervals", y = "Average Number of Steps Taken")
```





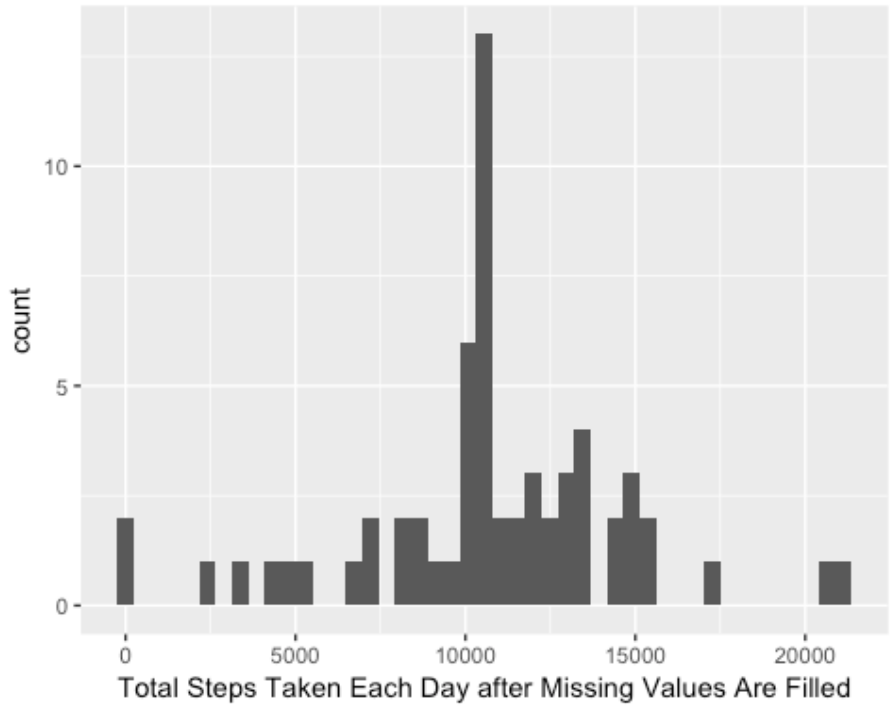
Imputing missing values

```
missing_values <- sum(is.na(activity_data$steps))
print(paste("The total number of missing value is", missing_values, "."))
## [1] "The total number of missing value is 2304 ."
#Devise a strategy for filling in all of the missing values in the dataset.

original_data <- activity_data
for (i in 1:nrow(activity_data)){
if(is.na(activity_data[i, 1]) == TRUE){
activity_data[i, 1] <- plot2[plot2$interval == activity_data[i, 3], 2]
} else {
next
}
}

daily_steps2 <- tapply(activity_data$steps, activity_data$date, sum)

#Make a histogram of the total number of steps taken each day
qplot(daily_steps2, binwidth = 480, xlab = "Total Steps Taken Each Day after Missing Values Are Filled")
```




```
#comparing the mean and median before and after the imputing
mean2 <- mean(daily_steps2)
median2 <- median(daily_steps2)
mat <- matrix(c(mean, mean2, median, median2), nrow = 2, ncol = 2)
table <- as.data.frame(mat, row.names = c("Origianl", "Imputed"))
colnames(table) <- c("Mean", "Median")
table
##                                                  Mean
## Origianl function (x, ...) , UseMethod("mean")
## Imputed                                     10766.19
##                                                Median
## Origianl function (x, na.rm = FALSE, ...) , UseMethod("median")
## Imputed                                       10766.19
```

Are there differences in activity patterns between weekdays and

Are there differences in activity patterns between weekdays and weekends?

```
activity_data$type <- weekdays(as.Date(activity_data$date))
for (i in 1:nrow(activity_data)){

if (activity_data[i, 4] %in% c("Monday", "Tuesday","Wednesday", "Thursday", "Friday")){
activity_data[i, 4] <- "Weekday"
} else {
activity_data[i, 4] <- "Weekend"
}
}
```

#Make a panel plot containing a time series plot (i.e. `type = "l"`) of the 5-minute interval (x- axis) and the average number of steps taken, averaged across all weekday days or weekend days (y-axis).

```
weekdays <- subset(activity_data, activity_data$type == "Weekday")
wdTime <- tapply(weekdays$steps, weekdays$interval, mean)
wddf <- data.frame(unique(weekdays$interval), wdTime, rep("Weekday", times =
length(wdTime)))
colnames(wddf) <- c("interval", "step", "type")
weekends <- subset(activity_data, activity_data$type == "Weekend")
weTime <- tapply(weekends$steps, weekends$interval, mean)
wedf <- data.frame(unique(weekends$interval), weTime, rep("Weekend", times =
length(weTime)))
colnames(wedf) <- c("interval", "step", "type")
plot4 <- rbind(wddf, wedf)
g4 <- ggplot(plot4, aes(x = interval, y = step))
g4 + geom_line() + facet_grid(type ~.) + labs(x = "Interval", y = "Number of Steps")
```