

Reproducible Research: Peer Assessment 1

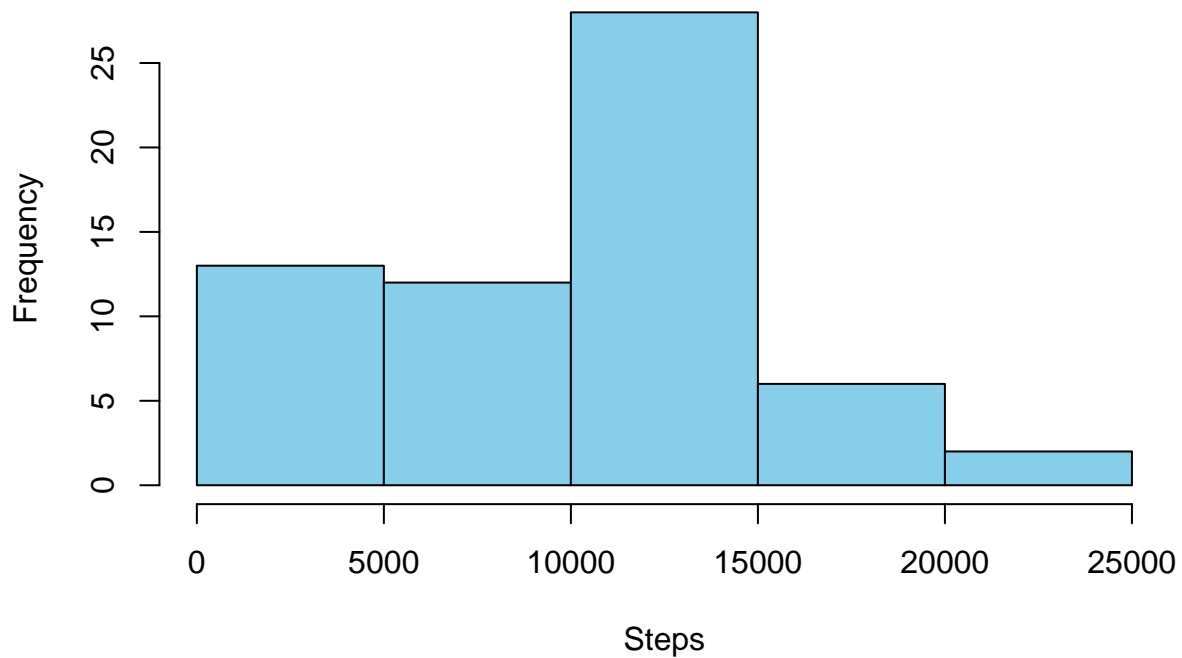
```
activity <- read.csv("activity.csv")
activity$date <- as.Date(activity$date)
head(activity)
```

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

```
total_steps_per_day <- activity %>%
  group_by(date) %>%
  summarize(total_steps = sum(steps, na.rm = TRUE))

hist(total_steps_per_day$total_steps,
     main = "Total Steps Per Day",
     xlab = "Steps",
     col = "skyblue")
```

Total Steps Per Day



```
mean_steps <- mean(total_steps_per_day$total_steps, na.rm = TRUE)
median_steps <- median(total_steps_per_day$total_steps, na.rm = TRUE)
```

```
mean_steps
```

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

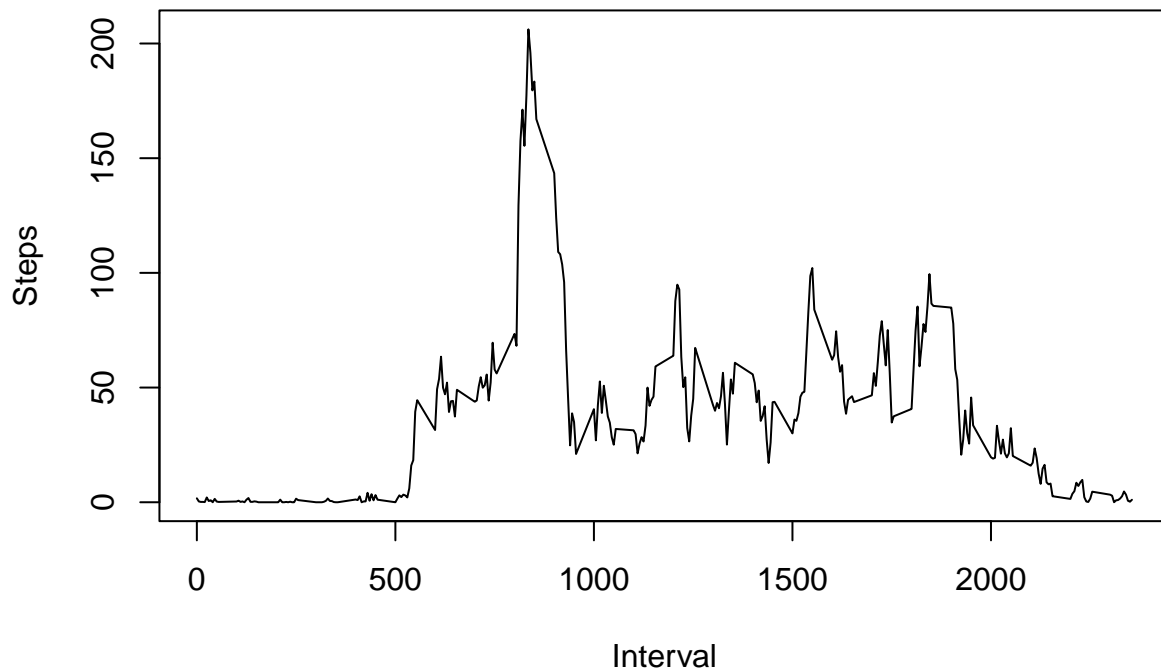
```
median_steps
```

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

```
avg_interval <- activity %>%
  group_by(interval) %>%
  summarize(mean_steps = mean(steps, na.rm = TRUE))

plot(avg_interval$interval, avg_interval$mean_steps,
     type = "l",
     main = "Average Steps per 5-Min Interval",
     xlab = "Interval",
     ylab = "Steps")
```

Average Steps per 5-Min Interval



```
max_interval <- avg_interval$interval[which.max(avg_interval$mean_steps)]
max_interval
```

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

```
sum(is.na(activity$steps))
```

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

```
imputed_activity <- activity
for (i in 1:nrow(imputed_activity)) {
  if (is.na(imputed_activity$steps[i])) {
    interval_value <- imputed_activity$interval[i]
    imputed_activity$steps[i] <- avg_interval$mean_steps[avg_interval$interval == interval_value]
  }
}

sum(is.na(imputed_activity$steps))
```

```
## [1] 0
```

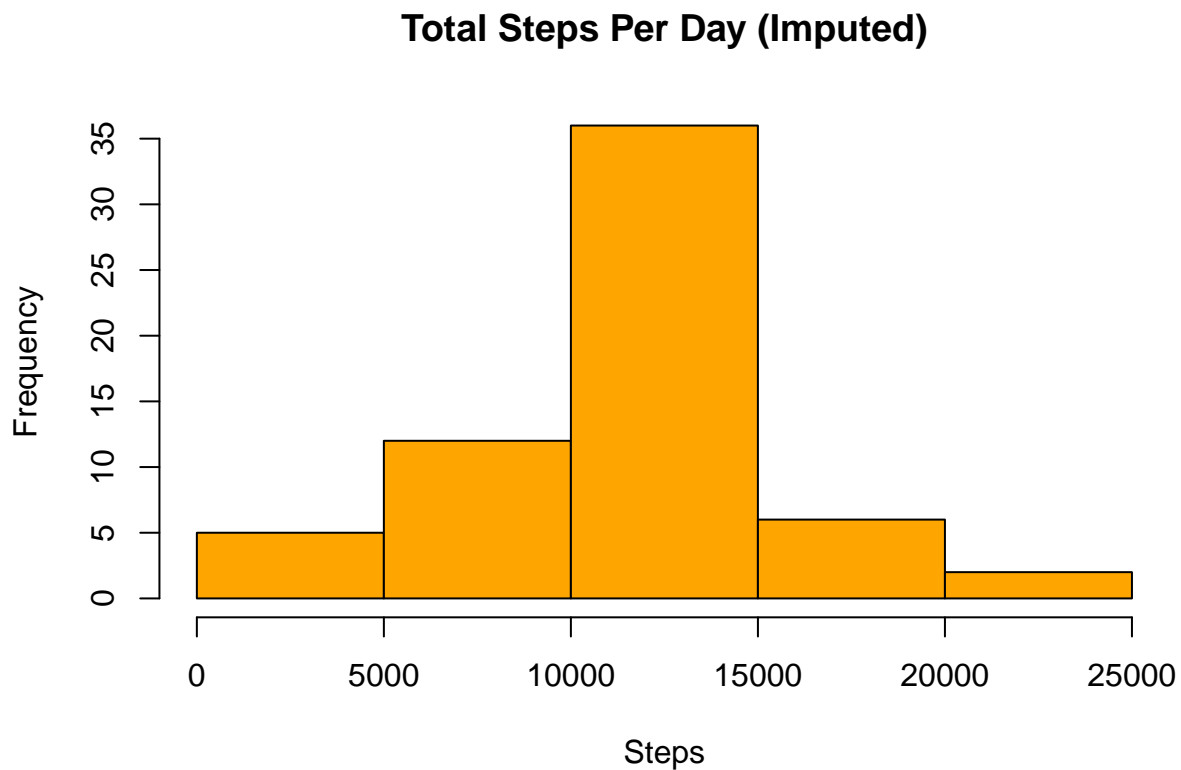
```
total_steps_imputed <- imputed_activity %>%
  group_by(date) %>%
```

```

summarize(total_steps = sum(steps))

hist(total_steps_imputed$total_steps,
      main = "Total Steps Per Day (Imputed)",
      xlab = "Steps",
      col = "orange")

```



```
mean(total_steps_imputed$total_steps)
```

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

```
median(total_steps_imputed$total_steps)
```

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

```

imputed_activity$daytype <- ifelse(weekdays(imputed_activity$date) %in%
                                   c("Saturday", "Sunday"),
                                   "weekend", "weekday")

avg_daytype <- imputed_activity %>%
  group_by(daytype, interval) %>%
  summarize(mean_steps = mean(steps))

```

```
## 'summarise()' has grouped output by 'daytype'. You can override using the  
## '.groups' argument.
```

```
ggplot(avg_daytype, aes(interval, mean_steps)) +  
  geom_line() +  
  facet_wrap(~ daytype, nrow = 2, ncol = 1) +  
  labs(title = "Activity Pattern: Weekday vs Weekend")
```

