

# Reproducible Research Project1

## Loading and Preprocessing the Data

```
# Load the data
activity <- read.csv("activity.csv")

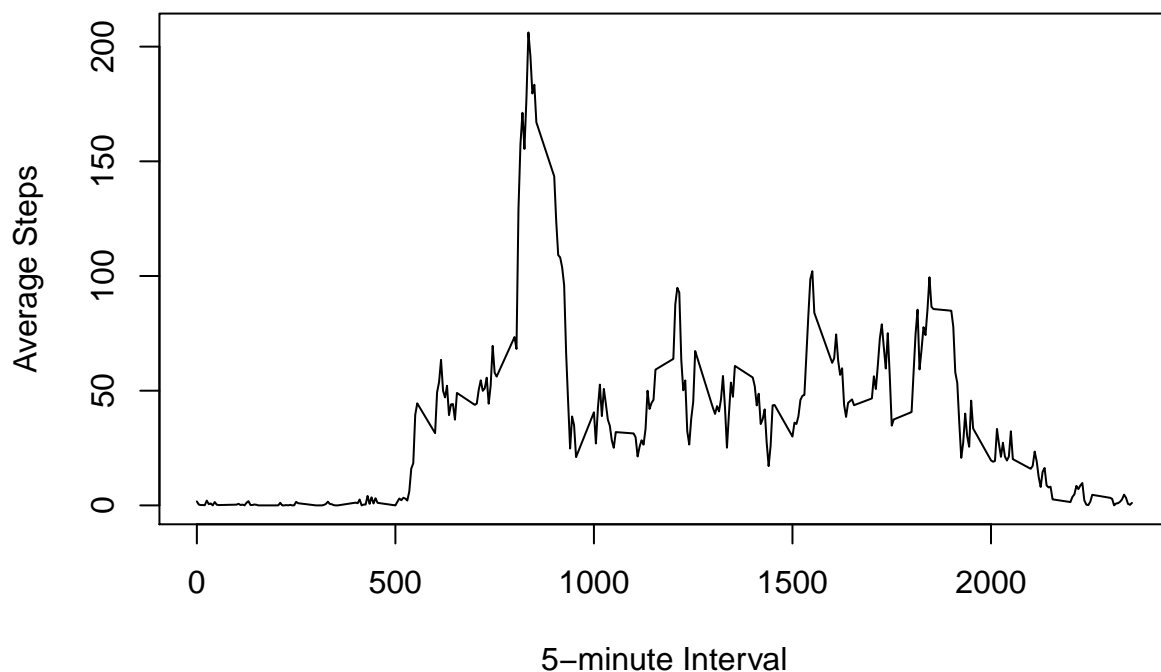
# Process/transform the data
activity$date <- as.Date(activity$date)
```

What is mean total number of steps taken per day?

```
# Calculate average steps for each interval
avg_steps <- aggregate(steps ~ interval, activity, mean, na.rm = TRUE)

# Create time series plot
plot(avg_steps$interval, avg_steps$steps, type = "l",
     main = "Average Daily Activity Pattern",
     xlab = "5-minute Interval", ylab = "Average Steps")
```

### Average Daily Activity Pattern



```
# Find interval with maximum steps
max_interval <- avg_steps$interval[which.max(avg_steps$steps)]
print(paste("Interval with maximum steps:", max_interval))
```

```
## [1] "Interval with maximum steps: 835"
```

## Inputting Missing Values?

```
# Calculate total missing values
total_missing <- sum(is.na(activity$steps))
print(paste("Total missing values:", total_missing))
```

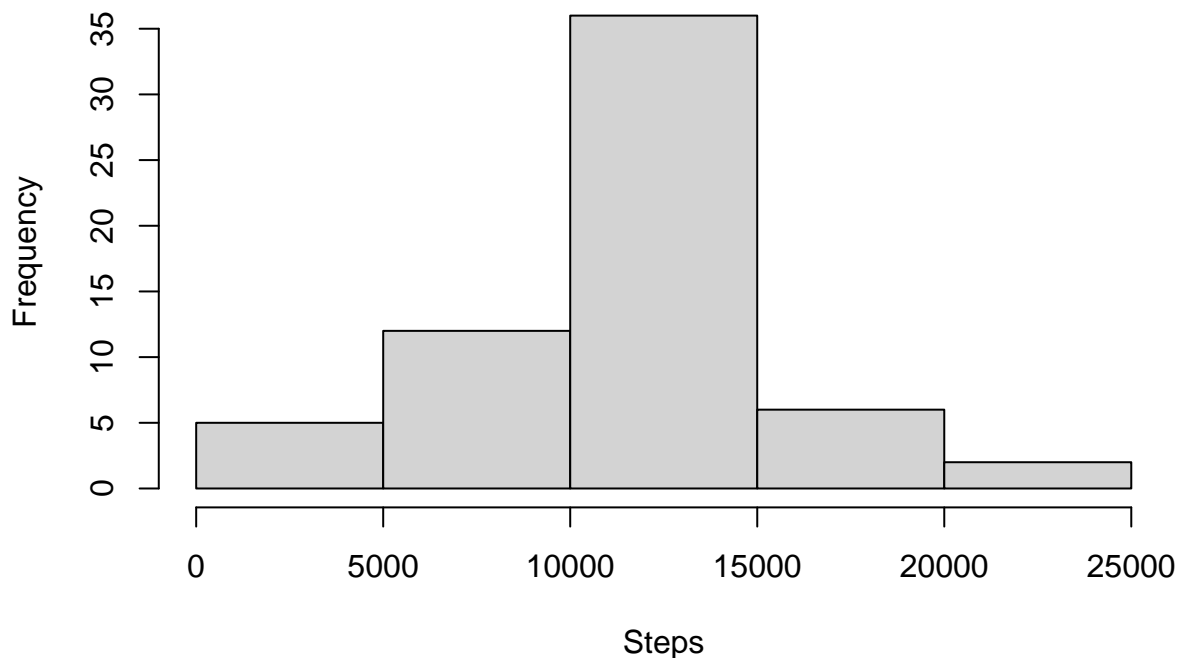
```
## [1] "Total missing values: 2304"
```

```
# Strategy: Use mean for that 5-minute interval
```

```
# Create new dataset with imputed values
activity_imputed <- activity
for (i in 1:nrow(activity_imputed)) {
  if (is.na(activity_imputed$steps[i])) {
    activity_imputed$steps[i] <- avg_steps$steps[avg_steps$interval == activity_imputed$interval[i]]
  }
}
```

```
# Create histogram of imputed data
steps_per_day_imputed <- aggregate(steps ~ date, activity_imputed, sum)
hist(steps_per_day_imputed$steps, main = "Total Steps per Day (Imputed)", xlab = "Steps")
```

### Total Steps per Day (Imputed)



```
# Calculate new mean and median
mean_steps_imputed <- mean(steps_per_day_imputed$steps)
median_steps_imputed <- median(steps_per_day_imputed$steps)

print(paste("New mean steps per day:", round(mean_steps_imputed, 2)))
```

```
## [1] "New mean steps per day: 10766.19"
```

```
print(paste("New median steps per day:", round(median_steps_imputed, 2)))
```

```
## [1] "New median steps per day: 10766.19"
```

Are there differences in activity patterns between weekdays and weekends??

```
# Create weekday/weekend factor
activity_imputed$day_type <- ifelse(weekdays(activity_imputed$date) %in% c("Saturday", "Sunday"), "weekend", "weekday")
activity_imputed$day_type <- as.factor(activity_imputed$day_type)

# Calculate average steps for each interval and day type
avg_steps_by_day <- aggregate(steps ~ interval + day_type, activity_imputed, mean)

# Create panel plot
library(ggplot2)
ggplot(avg_steps_by_day, aes(x = interval, y = steps, color = day_type)) +
  geom_line() +
  facet_wrap(~ day_type, ncol = 1, nrow = 2) +
  labs(title = "Average Steps by Interval and Day Type",
       x = "5-minute Interval", y = "Average Steps")
```

