Reproducible Research: Peer Assessment 1

Environment

```
library(magrittr)
library(dplyr)
library(ggplot2)
library(scales)
library(ggplot2)
library(knitr)
```

Loading and preprocessing the data

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

# Preprocess the data
data %<>%
  mutate(date=as.Date(as.character(date, format="%Y%m%d")))
```

What is mean total number of steps taken per day?

```
# Group steps by day
data_grouped_day <- data %>%
  group_by(date) %>%
  summarize(
    steps=sum(steps, na.rm=TRUE)
  )

# Plot histogram
png("figures/plot1.png")
hist(data_grouped_day$steps, breaks=30, xlab="date", ylab="steps", main="Total number of steps taken ea
dev.off()
```

```
## png
##   2
```

```
hist(data_grouped_day$steps, breaks=30, xlab="date", ylab="steps", main="Total number of steps taken ea
```

```
# Print mean and median steps per day
mean_before_imput <- round(mean(data_grouped_day$steps), 0)
median_before_imput <- round(median(data_grouped_day$steps), 0)
cat(sprintf('The mean steps taken per day is %d\n', mean_before_imput))
```
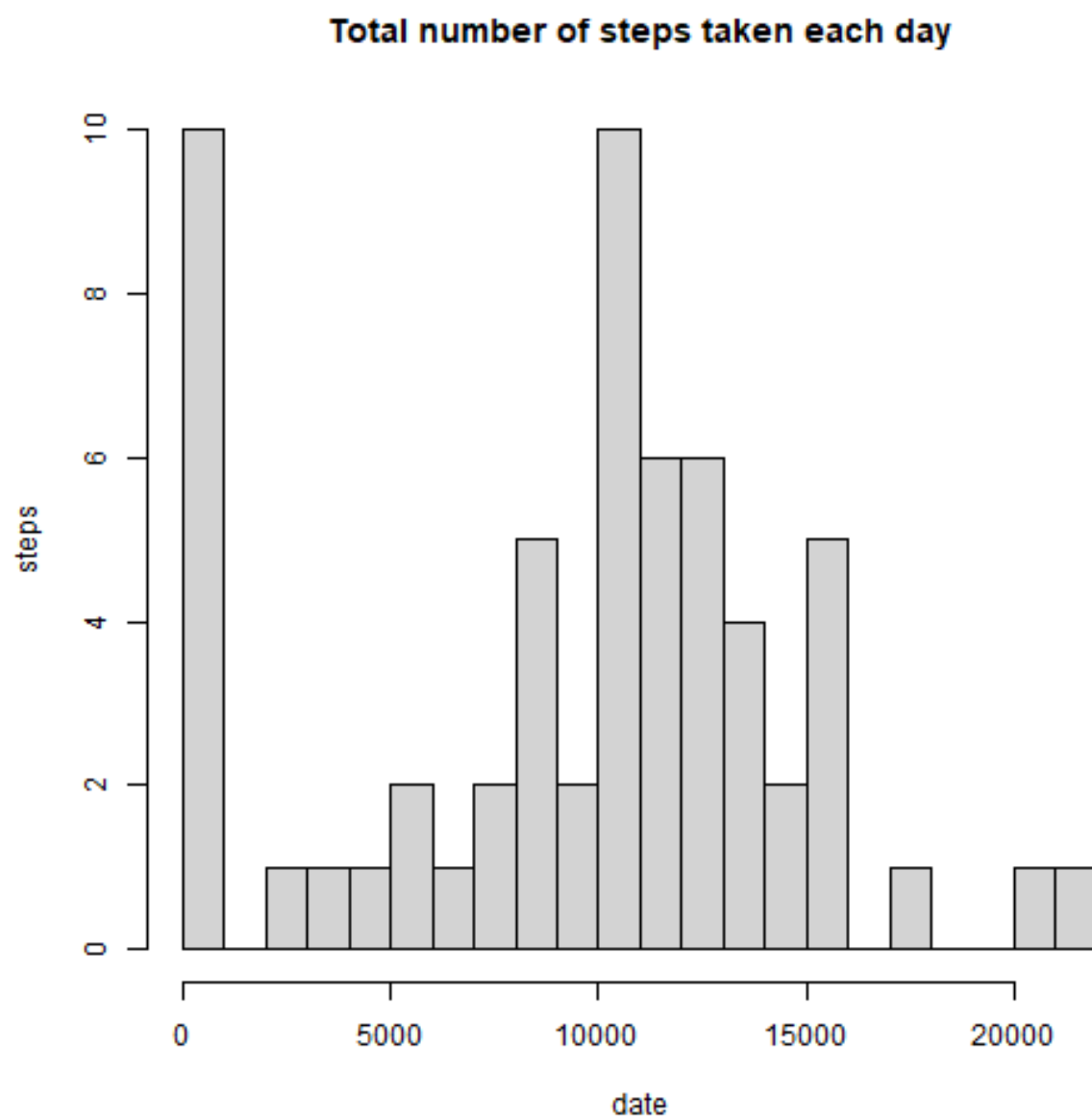
```
## The mean steps taken per day is 9354
```

Figure 1: plot of chunk unnamed-chunk-9

```r
cat(sprintf('The mean steps taken per day is %d', median_before_imput))
```

```
## The mean steps taken per day is 10395
```

What is the average daily activity pattern?

```r
# Group steps by day
data_grouped_interval <- data %>%
  group_by(interval) %>%
  summarize(
    steps=mean(steps, na.rm=TRUE)
  )

# Plot histogram
png("figures/plot2.png")
with(data_grouped_interval, plot(x=interval, y=steps, type="l", xlab="interval", ylab="steps", main="Tim
dev.off()
```

```
## png
##   2
```

```r
with(data_grouped_interval, plot(x=interval, y=steps, type="l", xlab="interval", ylab="steps", main="Tin
```

```r
# Find interval with max steps
max_steps <- subset(data_grouped_interval, steps==max(steps))

# Print mean and median steps per day
cat(sprintf('The 5-minute interval, on average across all the days in the dataset, that contains the ma
```

```
## The 5-minute interval, on average across all the days in the dataset, that contains the maximum numb
```

Imputing missing values

```r
# Print nº of NA
cat(sprintf('The total number of missing values in the dataset is %d\n', sum(is.na(data$steps))))
```

```
## The total number of missing values in the dataset is 2304
```

```r
# Calculate mean per 5-minute interval and impute steps by interval mean
data_mean_interval <- data %>%
  group_by(interval) %>%
  mutate(mean_interval = round(mean(steps, na.rm = TRUE),0)) %>%
  ungroup() %>%
  mutate(steps=ifelse(is.na(steps), mean_interval, steps))

# Group by day
data_grouped_day_imputed <- data_mean_interval %>%
  group_by(date) %>%
  summarize(
    steps=sum(steps, na.rm=TRUE)
  )

# Plot histogram
png("figures/plot3.png")
hist(data_grouped_day_imputed$steps, breaks=30, xlab="date", ylab="steps", main="Total number of steps
```

Time series plot of average number of steps taken by day intervals

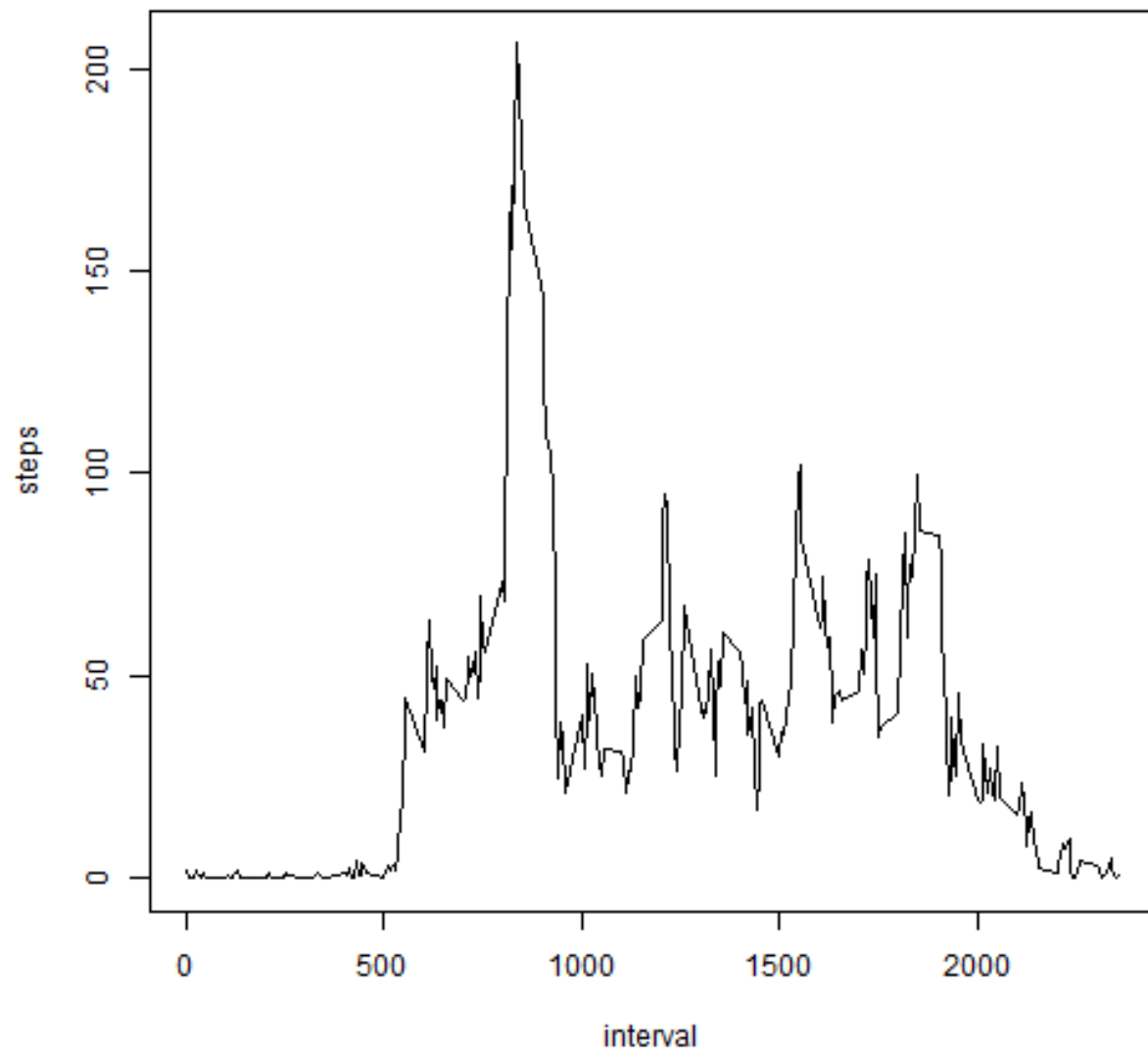Figure 2: plot of chunk unnamed-chunk-10

```
dev.off()
```

```
## png
##   2
```

```
hist(data_grouped_day_imputed$steps, breaks=30, xlab="date", ylab="steps", main="Total number of steps
```
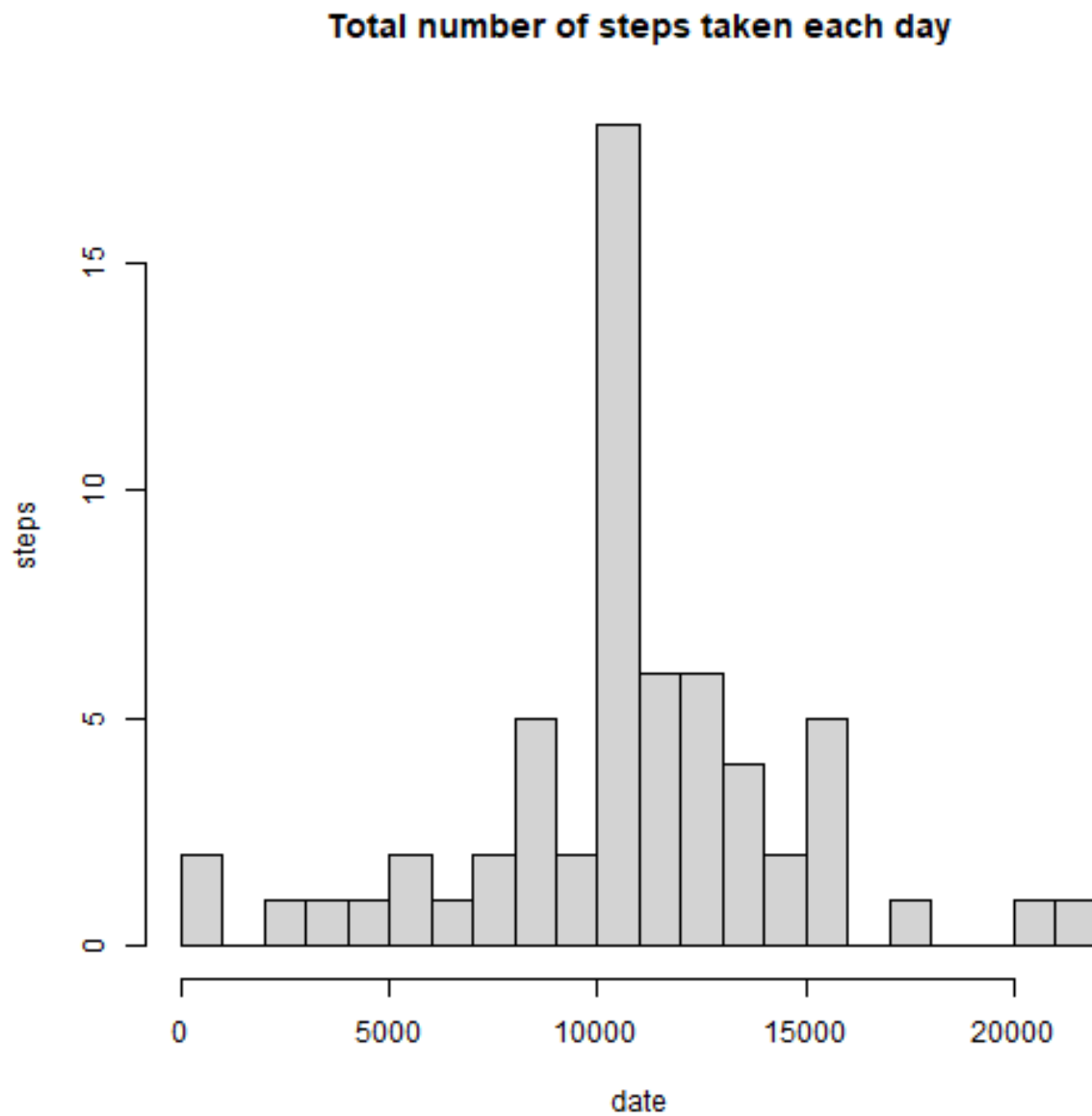
Total number of steps taken each day

Figure 3: plot of chunk unnamed-chunk-11

```
# Print mean and median steps per day after imputing
mean_after_imput <- round(mean(data_grouped_day_imputed$steps), 0)
median_after_imput <- round(median(data_grouped_day_imputed$steps), 0)
cat(sprintf('The mean steps taken per day is %d\n', mean_after_imput))
```

```
## The mean steps taken per day is 10766
```

```r
cat(sprintf('The mean steps taken per day is %d\n', median_after_imput))
```

```
## The mean steps taken per day is 10762
```

```r
# Differences between mean and median before and after imputing
if (mean_after_imput > mean_before_imput) {
  print("Mean after imputation is higher than mean before imputation")
} else if (mean_after_imput < mean_before_imput) {
  print("Mean after imputation is lower than mean before imputation")
} else {
  print("Mean did not change after imputation")
}
```

```
## [1] "Mean after imputation is higher than mean before imputation"
```

```r
if (median_after_imput > median_before_imput) {
  print("Median after imputation is higher than median before imputation")
} else if (median_after_imput < median_before_imput) {
  print("Median after imputation is lower than median before imputation")
} else {
  print("Median did not change after imputation")
}
```

```
## [1] "Median after imputation is higher than median before imputation"
```

Are there differences in activity patterns between weekdays and weekends?

```r
# Create dataframe with type of day
data_grouped_day_imputed_weekday <- data_mean_interval %>%
  mutate(day_type = ifelse(weekdays(date) %in% c("lunes", "martes", "miércoles", "jueves", "viernes"),
  mutate(day_type = as.factor(day_type)) %>%
  select(-mean_interval) %>%
  group_by(interval, day_type) %>%
  summarize(
    steps=mean(steps, na.rm=TRUE)
  )
```

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

```r
# Plot histogram
ggplot(data_grouped_day_imputed_weekday, aes(x = interval, y = steps, col = day_type)) +
  geom_line() +
  facet_grid(. ~ day_type) +
  labs(x = "Interval", y = "Steps", title = "Time series plot of average number of steps taken by day i
  scale_color_manual(values = c("blue", "red")) +
  theme_minimal()
```

```r
ggsave("figures/plot4.png", plot = last_plot(), device = "png", dpi = 300)
```
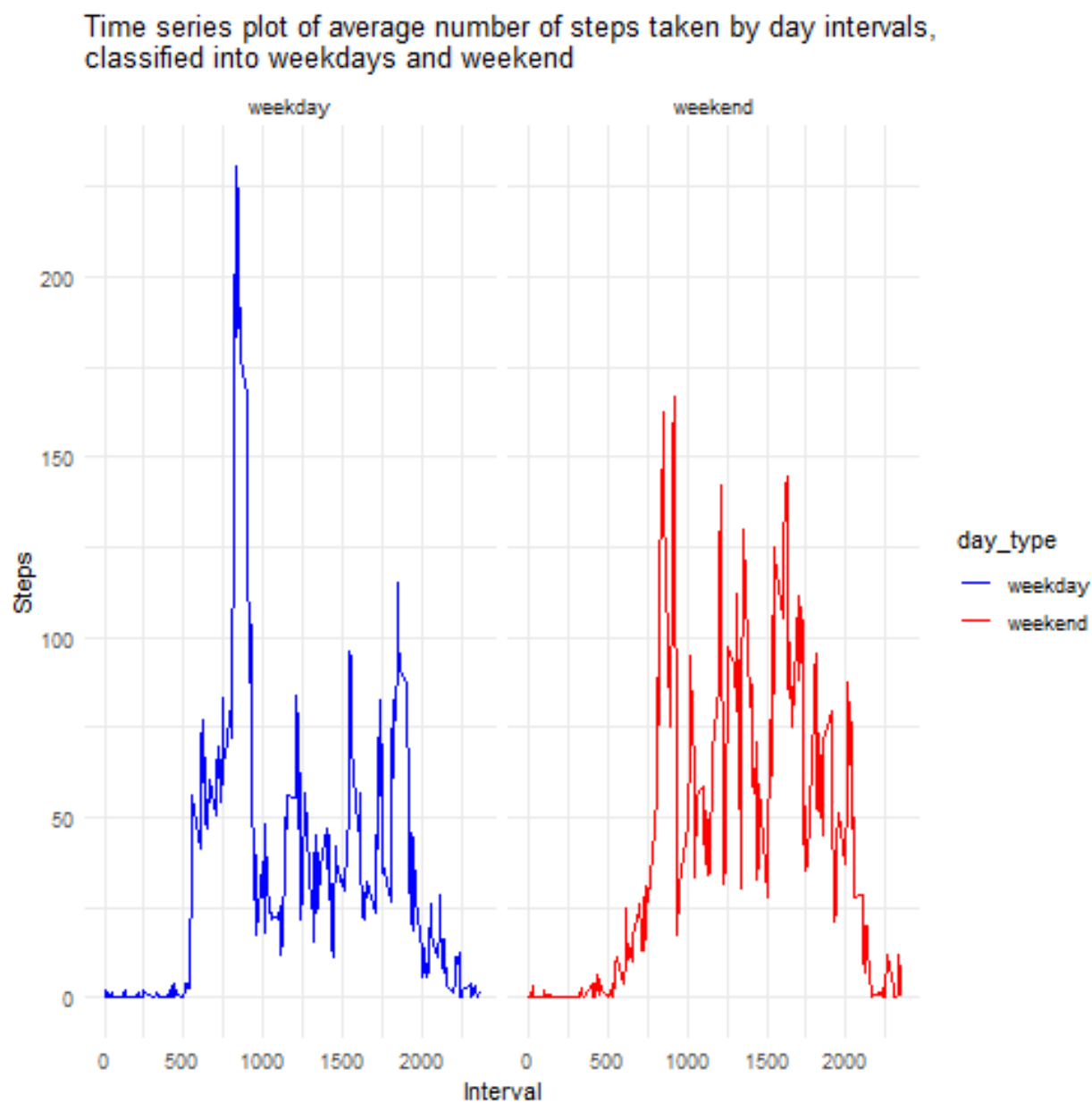
```
## Saving 7 x 7 in image
```

Figure 4: plot of chunk unnamed-chunk-12