

Loading and preprocessing the data

```
url <- 'https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2Factivity.zip'
download.file(url, method = "curl", destfile = "activity.zip")
unzip("./activity.zip")
activity <- read.csv("activity.csv", stringsAsFactors = FALSE)
activity$date <- as.POSIXct(activity$date, format="%Y-%m-%d")
summary(activity)
```

```
##           steps           date           interval
##  Min.      :  0.00   Min.      :2012-10-01 00:00:00   Min.      :    0.0
##  1st Qu.:  0.00   1st Qu.:2012-10-16 00:00:00   1st Qu.: 588.8
##  Median :  0.00   Median :2012-10-31 00:00:00   Median :1177.5
##  Mean    : 37.38   Mean    :2012-10-30 23:32:27   Mean     :1177.5
##  3rd Qu.: 12.00   3rd Qu.:2012-11-15 00:00:00   3rd Qu.:1766.2
##  Max.     :806.00   Max.     :2012-11-30 00:00:00   Max.      :2355.0
##  NA's     :2304
```

```
dim(activity)
```

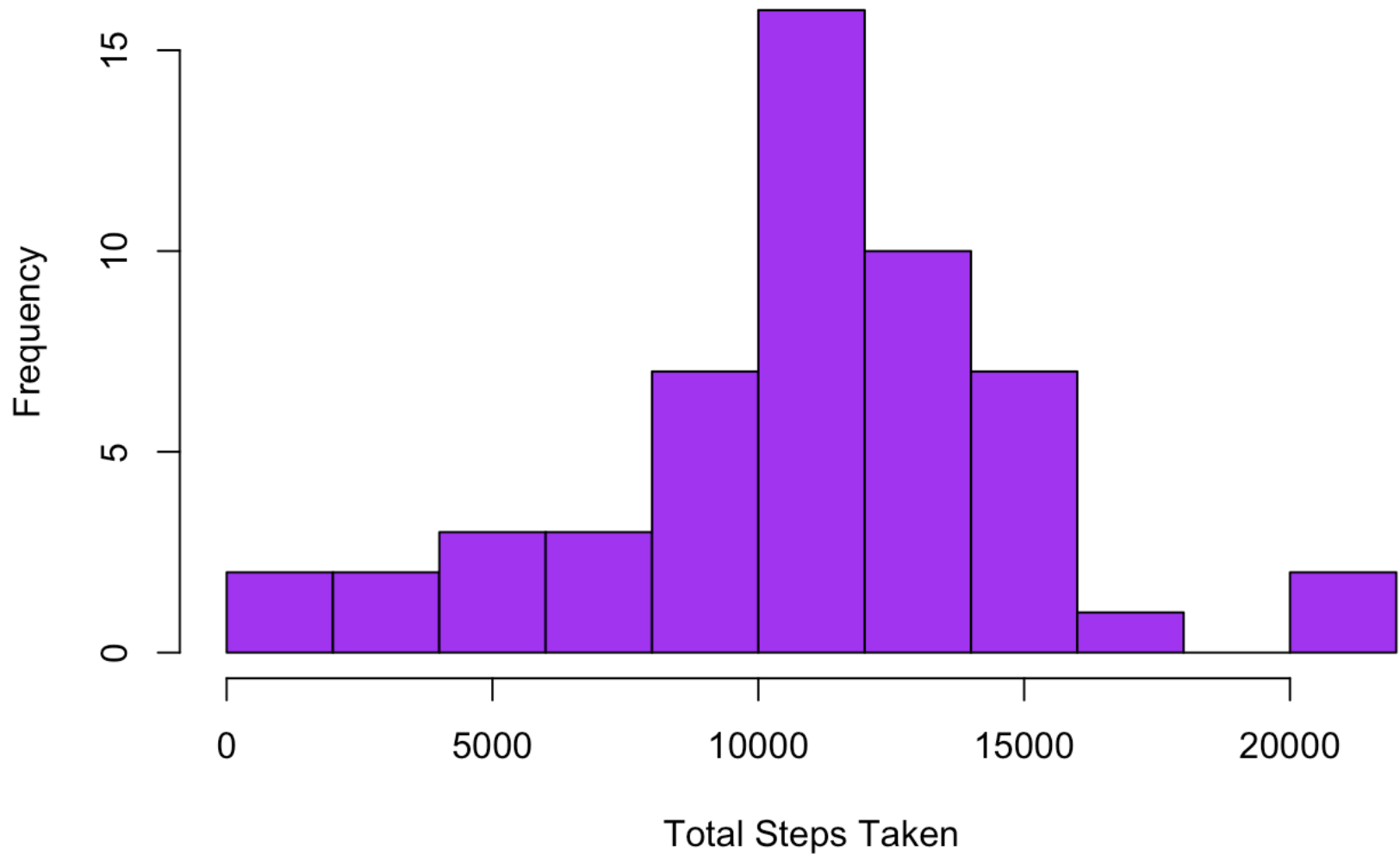
```
## [1] 17568      3
```

1. What is mean total number of steps taken per day?

Calculate the total number of steps taken per day

```
stepsTotal <- aggregate(steps ~ date, data = activity, FUN = sum, na.rm = TRUE)
hist(stepsTotal$steps, col = "purple", xlab = "Total Steps Taken", main = "Steps taken each day", breaks = "FD")
```

Steps taken each day



Calculate and report the mean and median of the total number of steps taken per day

```
mean(stepsTotal$steps)
```

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

```
median(stepsTotal$steps)
```

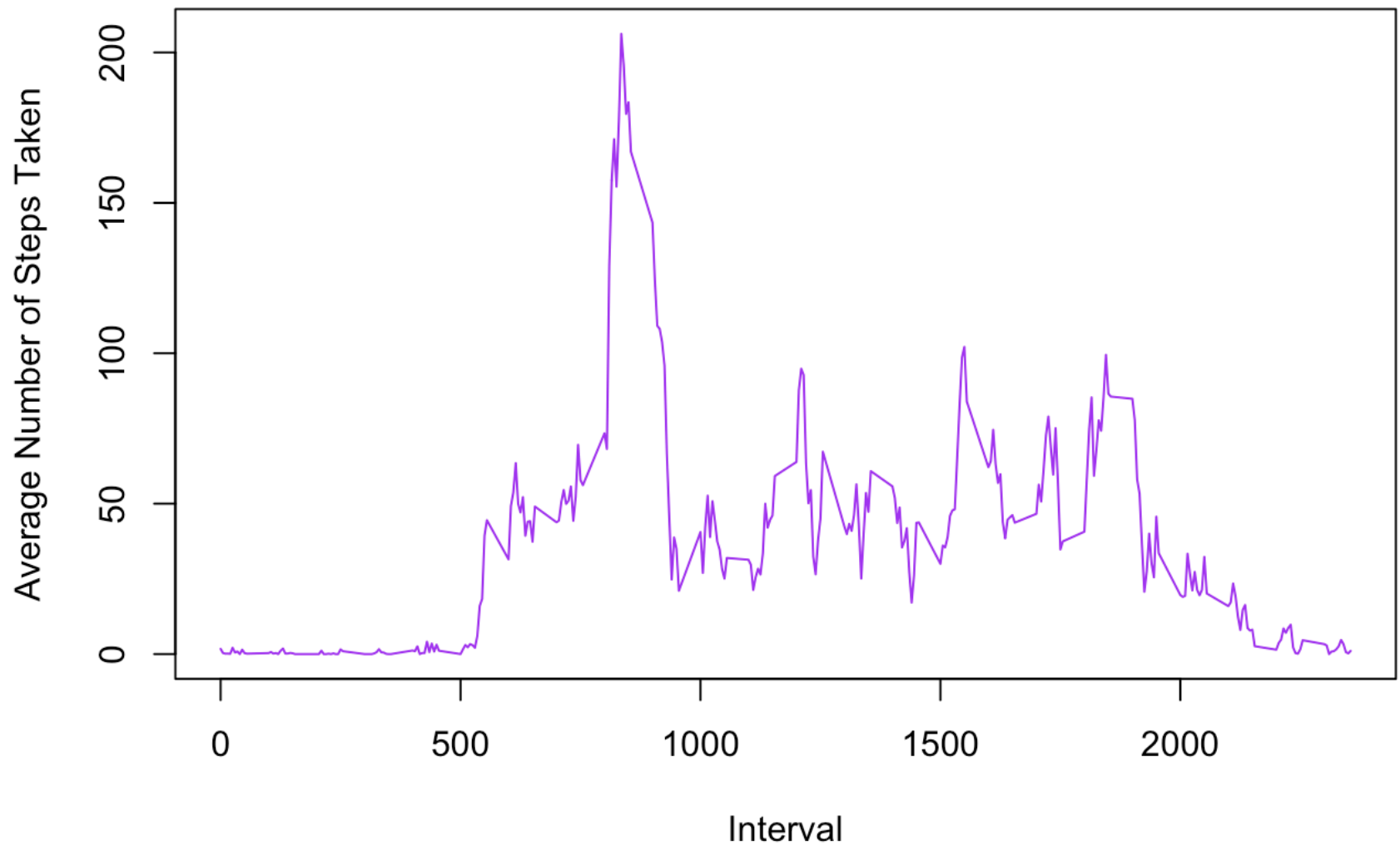
```
## [1] 10765
```

2. What is the average daily activity pattern?

Average steps taken within a day calculated

```
daily <- aggregate(steps ~ interval, data = activity, FUN = mean, na.rm = TRUE)
plot(x = daily$interval, y = daily$steps, type = "l", main = "Average Daily Action",
     ylab = "Average Number of Steps Taken", xlab = "Interval", col = "purple")
```

Average Daily Action



Which 5-minute interval, on average across all the days in the data set, contains the maximum number of steps?

```
daily$interval[which.max(daily$steps)]
```

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

3. Imputing missing values

Calculate and report missing values in the data set

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

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

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

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

As values above are the same, only missing values are in column “steps”

Devise a strategy for filling in all of the missing values in the data set

Step's average value of day & interval will be placed instead of the missing values, as different behavior patterns are not only for weekend or workday but within each day itself. To do so, extra columns depicting name of weekday will be added.

```
activity$day <- weekdays(activity$date)
activity$week <- ""
activity[activity$day == "Saturday" | activity$day == "Sunday", ]$week <- "weekend"
activity[! (activity$day == "Saturday" | activity$day == "Sunday"), ]$week <- "weekday"
```

Step's average value of day calculated bellow.

```
missing <- aggregate(steps ~ interval+day, data = activity, median, na.rm= TRUE)
```

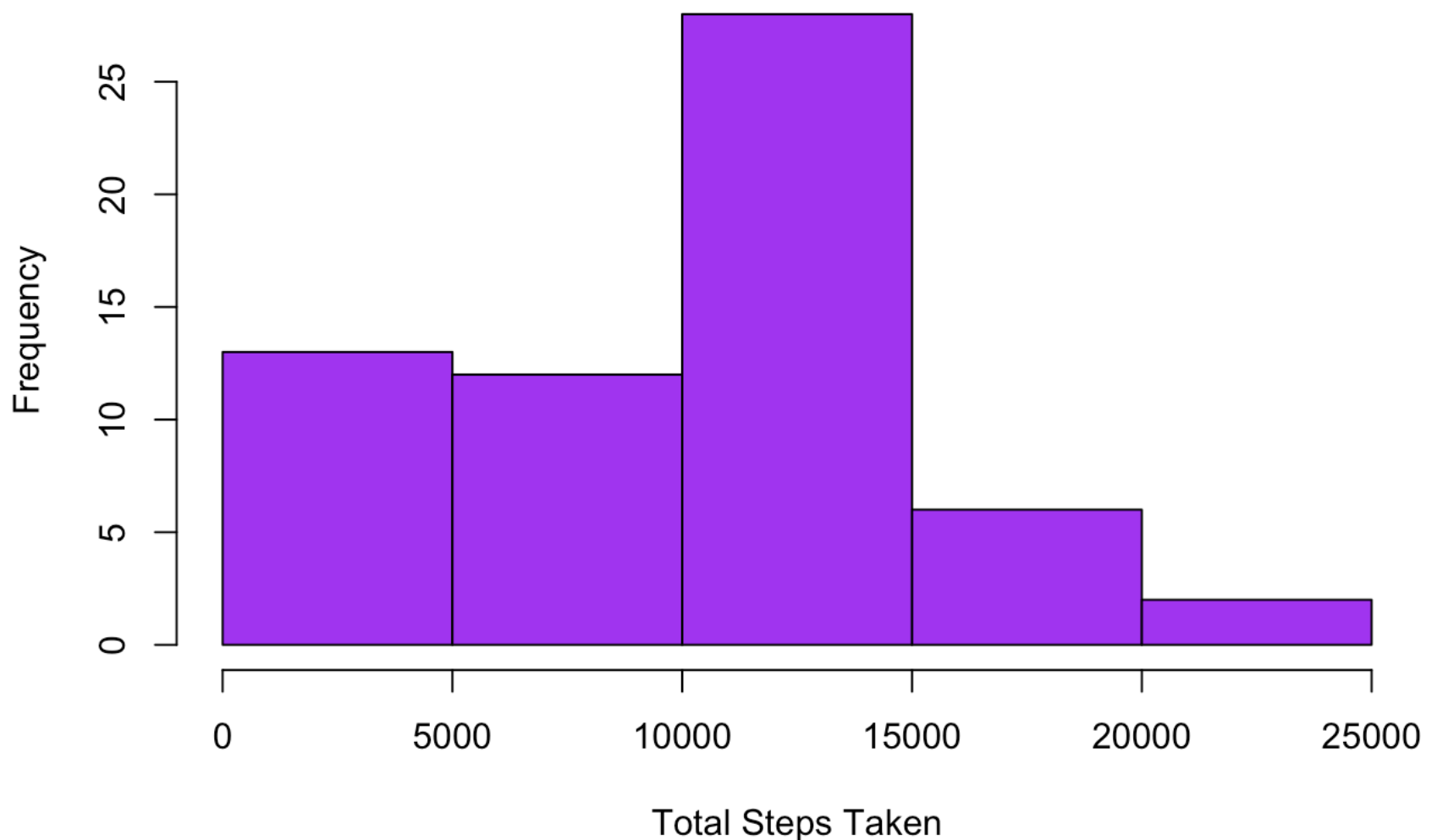
A new data set is created, merging existing set with average steps across day & interval. NA values are replaced with average in column Steps.

```
activity2 <- merge(activity, missing, by = c("day", "interval"))
activity2$stepsNew <- ifelse(is.na(activity2$steps.x), activity2$steps.y, activity2$steps.x)
```

Make a histogram of the total number of steps taken each day

```
stepsTotal2 <- aggregate(stepsNew ~ date, data = activity2, FUN = sum)
hist(stepsTotal2$stepsNew, col = "purple", xlab = "Total Steps Taken", main = "Steps taken each day (With Imputed Values)", breaks = "FD")
```

Steps taken each day (With Imputed Values)



Do these values differ from the estimates from the first part of the assignment? What is the impact of imputing missing data on the estimates of the total daily number of steps?

Total count of steps increased by:

```
sum(activity2$stepsNew) - sum(activity$steps, na.rm = TRUE)
```

```
## [1] 21411.5
```

Also comparing two histograms, with added averages, the frequency increased and optimal bin count reduced. AS averages were applied, histogram itself become more smooth, less edgy.

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

“weekday” & “weekend” variable created above. Average value for factor variables computed

```
seven <- aggregate(stepsNew ~ interval+week, activity2, mean)
```

Chart created

```
library(lattice)
par(mfrow=c(2, 1))
```

```
xyplot(stepsNew ~ interval | week, seven, type = "l", layout = c(1,2), ylab = "Average Number of Steps Taken", col = "purple")
```

