

Reproducible Research - Assignment One

Load the required R Libraries and Surppress Messages

```
suppressMessages(library(dplyr))
suppressMessages(library(ggplot2))
suppressMessages(library(knitr))
suppressMessages(library(lubridate))
suppressMessages(library(scales))
```

Load Dataset and Transform Date

```
activity <- read.csv("activity.csv",header = TRUE, sep = ',')
activity$date <- ymd(activity$date)
steps <- activity %>%
  filter(!is.na(steps)) %>%
  group_by(date) %>%
  summarize(steps = sum(steps))
```

What is mean total number of steps taken per day?

Plot histogram

```
plot.steps.day <- ggplot(activity2, aes(x=daily.step)) +
  geom_histogram(binwidth = 1000, aes(y=..count.., fill=..count..)) +
  geom_vline(xintercept=mean.activity, colour="red", linetype="dashed", size=1) +
  geom_vline(xintercept=median.activity, colour="green", linetype="dotted", size=1) +
  labs(title="Histogram of Number of Steps taken each day", y="Frequency", x="Daily Steps")
plot.steps.day
```

Mean and median total number of steps taken per day

```
mean.steps <- round(mean(steps$steps, na.rm=TRUE))
median.steps <- round(median(steps$steps, na.rm=TRUE))
paste("The Mean number of steps is",mean.steps,"and the Median is",median.steps)
```

What is the average daily activity pattern?

```
activity$interval <- as.POSIXct(strptime(sprintf("%04d", activity$interval), "%H%M"))
interval <- activity %>%
  filter(!is.na(steps)) %>%
  group_by(interval) %>%
  summarize(steps = mean(steps))
```

Time-plot of the 5-minute interval and average number of steps taken, averaged across all days

```
ggplot(interval, aes(x=interval, y=steps)) +  
  geom_line(color = "mistyrose4") +  
  scale_x_datetime(breaks = date_breaks("2 hour"),  
    labels = date_format("%H:%M"),  
    limits = c(interval$Interval[1], interval$Interval[288])) +  
  labs(title = "Average Number of Steps taken (Averaged Across All Days)",  
    x = "Time of Day",  
    y = "Average Steps")
```

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

```
max.interval <- interval[which.max(interval$steps),]  
paste("The interval with the maximum number of steps is",  
  round(max.interval$steps[1]), "at",  
  strftime(max.interval$interval[1], format="%H:%M:%S"))
```

Calculate and report the total number of missing values in the dataset

```
missing.steps <- sum(is.na(activity$steps))  
paste("The number of missing values in the dataset is",missing.steps)
```

Devise a strategy for filling in all of the missing values in the dataset.

```
tofill.activity <- activity  
missing.activity <- is.na(tofill.activity$steps)  
mean.interval <- tapply(tofill.activity$steps,  
  tofill.activity$interval,  
  mean, na.rm=TRUE, simplify=TRUE)
```

Create a new dataset (equal to the original) but with the missing data filled in.

```
tofill.activity$steps[missing.activity] <- mean.interval[as.character(tofill.activity$interval[missing.activity])]
```

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

```
filled_steps <- tofill.activity %>%  
  filter(!is.na(steps)) %>%  
  group_by(date) %>%  
  summarize(steps = sum(steps))  
ggplot(filled_steps, aes(x = steps)) +  
  geom_histogram(fill = "mistyrose3", colour="Black", binwidth = 1000) +  
  labs(title = "Total Number of Steps Each Day (missing values replaced)",  
    x = "Steps per day",  
    y = "Occurances")
```

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

```
mean.steps <- round(mean(filled_steps$steps, na.rm=TRUE))
median.steps <- round(median(filled_steps$steps, na.rm=TRUE))
paste("The Mean number of steps is",mean.steps,"and the Median is",median.steps)
```

Do these values differ from the estimates from the first part of the assignment?

```
imputed <- mutate(tofill.activity,
  weektype = ifelse(weekdays(tofill.activity$date) == "Saturday" |
    weekdays(tofill.activity$date) == "Sunday", "Weekend", "Weekday"))
imputed$weektype <- as.factor(imputed$weektype)
imputed.full <- imputed %>%
  group_by(interval, weektype) %>%
  summarise(steps = mean(steps))
ggplot(imputed.full, aes(x=interval, y=steps, color = weektype)) +
  geom_line(color = "mistyrose4") +
  facet_wrap(~weektype, ncol = 1, nrow=2)+
  scale_x_datetime(breaks = date_breaks("2 hour"),
    labels = date_format("%H:%M"),
    limits = c(imputed.full$Interval[1],
      imputed.full$Interval[288])) +
  labs(title = "Average Number of Steps taken (Weekdays vs. Weekends)",
    x = "Time of Day",
    y = "Average Steps")
```