PA1_template

1.1 Total number of steps taken every day

```
unzip(zipfile = "activity.zip")
activitydata <- read.csv("activity.csv", header=T, na.strings=c("NA"))
activitydata2 <- is.na(activitydata$steps)
stepsperday <- aggregate(activitydata$steps, by=list(activitydata$date), sum)
head(stepsperday)
```

1.2 Histogram of the total number of steps taken each day

```
library(ggplot2)
ggplot(stepsperday, aes(x=x)) +
  geom_histogram() + labs(title="Total steps taken each day",
                x="Number of steps", y="Number of days", na.rm=T)
```

1.3 Mean and median of the total number of steps taken per day

```
meanstepsperday <- mean(stepsperday$x, na.rm=T)
meanstepsperday
medianstepsperday <- median(stepsperday$x, na.rm=T)
medianstepsperday
```

2.1 Time series plot of 5-min interval and average number of steps taken, averaged across all days

```
activitydata <- read.csv(choose.files())
activitydata3 <- activitydata[!activitydata2,]
stepsperinterval <- aggregate(activitydata3$steps, by=list(activitydata3$interval), mean)
ggplot(stepsperinterval, aes(x=Group.1, y=x)) +
  labs(title = "Total steps taken each interval", x = "Interval", y = "Number of steps")+
  geom_line()
```

2.2 5-minute interval that on average across all days in the dataset, contains max number of steps

```
stepsperinterval[which.max(stepsperinterval$average),]
```

3.1 Total number of missing values in the dataset

```
totalNA <- sum(activitydata2)
totalNA
```

3.2/3 Filling in all of the missing values in the dataset

```
completeactivitydata <- aggregate(activitydata3$steps, by=list(activitydata3$interval), sum)
meanvalues <- tapply(activitydata$steps, activitydata$interval, mean, na.rm=T, simplify=T)
completeactivitydata$steps[activitydata2] <- meanvalues[as.character(completeactivitydata$interval[ac
tivitydata2])]
```

3.4 Histogram of the total number of steps taken each day

```
completeactivitydata <- aggregate(activitydata$steps, by=list(activitydata$date), sum)
ggplot(completeactivitydata, aes(x=x)) +
  geom_histogram() + labs(title="Total steps taken each day", x="Numberof steps", y="Number of days
")
mean(completeactivitydata$steps)


mean(as.numeric(completeactivitydata$steps))
median(as.numeric(completeactivitydata$steps))
```

4.1 New factor variable in the dataset with two levels

```
activitydata$weekday <- weekdays(activitydata$date)
activitydata$weekend <- ifelse (activitydata$weekday == "Saturday" | activitydata$weekend == "Sunday
", "Weekend", "Weekday")
weekendactivitydata <- aggregate(activitydata$steps, by=list(activitydata$day, activitydata$interval), m
ean)
head(weekendactivitydata)
ggplot(weekendactivitydata, aes(x=Group.2, y=x, color=Group.1)) +
  geom_line() + facet_grid(Group.1 ~ .) + labs(title = "Mean of Steps by Interval", x = "interval", y = "st
eps")
```