Peer-graded Assignment: Course Project 1

Francisco Mimica Porras

2/1/2021

## Setting the data

setwd("C:/Users/Famil/OneDrive/COURSERA COURSES/Reproducible Research/Peer-graded Assignment Course Project 1")  
library(datasets)  
activity <- read.csv("C:/Users/Famil/OneDrive/COURSERA COURSES/Reproducible Research/Peer-graded Assignment Course Project 1/activity.csv")

## Formatting date variable

activity$date <- as.Date(activity$date)

## Including Plots: Histogram without NA values

Histogram of total number of steps taken on each day

sum\_steps<-aggregate(activity$steps,by=list(activity$date),FUN=sum,na.rm=TRUE)   
png(file = "Plot1.png", width = 480, height = 480, units = "px")  
hist(sum\_steps$x,   
 breaks=seq(from=0, to=25000, by=2500),  
 col="yellow",   
 xlab="Total number of steps",   
 ylim=c(0, 20),   
 main="total number of steps taken each day")

## Mean and Median of Steps

Mean and median number of steps taken each day

mean(sum\_steps$x)

## [1] 9354.23

median(sum\_steps$x)

## [1] 10395

## Time Series Plot

Time series plot of the average number of steps taken

avg\_steps<-aggregate(activity$steps,by=list(activity$interval),FUN=mean,na.rm=TRUE)   
colnames(avg\_steps)<-c("interval","steps")  
library(ggplot2)  
png(file = "Plot2.png", width = 480, height = 480, units = "px")  
ggplot(aes(x=interval,y=steps),data=avg\_steps)+geom\_line()

## Maximum Average 5 minute interval

avg\_steps[avg\_steps$steps==max(avg\_steps$steps),1]

## [1] 835

## Imputing NA

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

sum(is.na(activity$steps))

## [1] 2304

Replace NA values with the mean of the steps

activity$steps[is.na(activity$steps)]<-mean(activity$steps,na.rm=TRUE)

Here are some rows of new data set

head(activity)

## steps date interval  
## 1 37.3826 2012-10-01 0  
## 2 37.3826 2012-10-01 5  
## 3 37.3826 2012-10-01 10  
## 4 37.3826 2012-10-01 15  
## 5 37.3826 2012-10-01 20  
## 6 37.3826 2012-10-01 25

## Histogram with Replaced NA values

Histogram of total number of steps taken on each day (replaced NA values)

sum\_steps<-aggregate(activity$steps,by=list(activity$date),FUN=sum,na.rm=TRUE)   
png(file = "Plot3.png", width = 480, height = 480, units = "px")  
hist(sum\_steps$x,   
 breaks=seq(from=0, to=25000, by=2500),  
 col="yellow",   
 xlab="Total number of steps",   
 ylim=c(0, 30),   
 main="Total number of steps taken each day\n(NA replaced by mean)")

## Mean and median number of steps taken each day after replacing NA values with mean

mean(sum\_steps$x)

## [1] 10766.19

median(sum\_steps$x)

## [1] 10766.19

## Differences in activity patterns between weekdays and weekends

Panel plot comparing the average number of steps taken per 5-minute interval across weekdays and weekends

activity$days=tolower(weekdays(activity$date))

# Now categorised days into weekend and weekdays

activity$day\_type<-ifelse(activity$days=="saturday"|activity$days=="sunday","weekend","weekday")

# Take mean steps taken on weekend or weekday in the intervals

avg\_steps<-aggregate(activity$steps,by=list(activity$interval,activity$day\_type),FUN=mean,na.rm=TRUE)  
  
colnames(avg\_steps)<-c("interval","day\_type","steps")

# Create panel plot between average steps and interval seperated by day type

png(file = "Plot4.png", width = 480, height = 480, units = "px")  
ggplot(aes(x=interval,y=steps),data=avg\_steps)+geom\_line()+facet\_wrap(~avg\_steps$day\_type)