Loading and processing the data

data<-read.csv("activity.csv")

What is mean total number of steps taken per day?

*#remove na's*

data\_no\_na<-data[!is.na(data$steps),]

*#Group the steps per day*

steps\_perday<-aggregate(data\_no\_na$steps, list(data\_no\_na$date), sum)

colnames(steps\_perday)<-c("Date","Steps")

*#Plot histogram*

hist(steps\_perday$Steps, main="Histogram of total steps per day", xlab = "Steps per day")

mean\_steps<-mean(steps\_perday$Steps, na.rm=TRUE)

median\_steps<-median(steps\_perday$Steps, na.rm=TRUE)

mean\_steps

## [1] 10766.19

median\_steps

## [1] 10765

What is the average daily activity pattern?

*#Find mean for each interval*

five\_min\_avg<-aggregate(steps~interval, data = data, mean, na.rm=TRUE)

*#plot time series graph*

plot(five\_min\_avg$interval, five\_min\_avg$steps, type="l", main="Time series for 5min intervals", ylab="Avg steps per 5min intervals", xlab="Intervals")

*#Find max steps, then loop to find the interval*

max\_steps <- max(five\_min\_avg$steps)

**for** (i **in** 1:288)

{

**if** (five\_min\_avg$steps[i] == max\_steps)

five\_minute\_interval\_at\_max\_steps <- five\_min\_avg$interval[i]

}

five\_minute\_interval\_at\_max\_steps

## [1] 835

What is the average daily activity pattern?

**library**(ggplot2)

## Warning: package 'ggplot2' was built under R version 3.3.2

subdata = data[!is.na(data$steps), ]

steps\_per\_interval <- aggregate(subdata$steps,

by = list(interval = as.factor(subdata$interval)),

FUN=mean, na.rm=TRUE)

steps\_per\_interval$interval <-

as.integer(levels(steps\_per\_interval$interval)[steps\_per\_interval$interval])

colnames(steps\_per\_interval) <- c("interval", "steps")

ggplot(data=steps\_per\_interval, aes(x=interval, y=steps)) +

geom\_line()

max\_interval <- steps\_per\_interval[which.max(steps\_per\_interval$steps),]

max\_interval

## interval steps

## 104 835 206.1698

Imputing missing values

total\_na <- 0

**for** (i **in** 1:17568)

{

**if**(is.na(data$steps[i]))

total\_na <- total\_na+1

}

total\_na

## [1] 2304

activity\_filled\_in <- data

**for** (i **in** 1:17568)

{

**if**(is.na(activity\_filled\_in$steps[i]))

{

five\_minute\_pointer <- activity\_filled\_in$interval[i]

**for** (j **in** 1:288)

{

**if** (five\_min\_avg$interval[j] == five\_minute\_pointer)

activity\_filled\_in$steps[i] <- five\_min\_avg$steps[j]

}

}

}

total\_steps\_each\_day\_filled\_in <- aggregate(steps~date, data=activity\_filled\_in, FUN=sum, na.rm=TRUE)

hist(total\_steps\_each\_day\_filled\_in$steps)

total\_steps\_each\_day\_mean\_filled\_in <- mean(total\_steps\_each\_day\_filled\_in$steps)

total\_steps\_each\_day\_median\_filled\_in <- median(total\_steps\_each\_day\_filled\_in$steps)

total\_steps\_each\_day\_mean\_filled\_in

## [1] 10766.19

total\_steps\_each\_day\_median\_filled\_in

## [1] 10766.19

Are there differences in activity patterns between weekdays and weekends?

**library**(lubridate)

## Warning: package 'lubridate' was built under R version 3.3.3

##

## Attaching package: 'lubridate'

## The following object is masked from 'package:base':

##

## date

week <- wday(activity\_filled\_in$date)

week\_day <- week

**for** (i **in** 1:17568)

{

**if**(week[i] == 1)

week\_day[i] <- 'weekend'

**if**(week[i] == 2)

week\_day[i] <- 'weekday'

**if**(week[i] == 3)

week\_day[i] <- 'weekday'

**if**(week[i] == 4)

week\_day[i] <- 'weekday'

**if**(week[i] == 5)

week\_day[i] <- 'weekday'

**if**(week[i] == 6)

week\_day[i] <- 'weekday'

**if**(week[i] == 7)

week\_day[i] <- 'weekend'

}

activity\_filled\_in$weekday <-week\_day

weekday <- grep("weekday",activity\_filled\_in$weekday)

weekday\_frame <- activity\_filled\_in[weekday,]

weekend\_frame <- activity\_filled\_in[-weekday,]

five\_minutes\_average\_weekday <- aggregate(steps~interval, data=weekday\_frame, FUN=mean, na.rm=TRUE)

five\_minutes\_average\_weekend <- aggregate(steps~interval, data=weekend\_frame, FUN=mean, na.rm=TRUE)

plot(x = five\_minutes\_average\_weekday$interval, y = five\_minutes\_average\_weekday$steps, type = "l")

plot(x = five\_minutes\_average\_weekend$interval, y = five\_minutes\_average\_weekend$steps, type = "l")