**The Codes and Results**

unzip("activity.zip")

initialData <- read.csv("activity.csv", colClasses=c("numeric", "Date", "numeric"))

head(initialData)

|  | **steps**  <dbl> | **date**  <date> | **interval**  <dbl> |  |
| --- | --- | --- | --- | --- |
| 1 | *NA* | 2012-10-01 | 0 |  |
| 2 | *NA* | 2012-10-01 | 5 |  |
| 3 | *NA* | 2012-10-01 | 10 |  |
| 4 | *NA* | 2012-10-01 | 15 |  |
| 5 | *NA* | 2012-10-01 | 20 |  |
| 6 | *NA* | 2012-10-01 | 25 |  |

6 rows

str(initialData)

'data.frame': 17568 obs. of 3 variables:

$ steps : num NA NA NA NA NA NA NA NA NA NA ...

$ date : Date, format: "2012-10-01" "2012-10-01" "2012-10-01" "2012-10-01" ...

$ interval: num 0 5 10 15 20 25 30 35 40 45 ...

data <- initialData[!(is.na(initialData$steps)), ]

totalStepsDay <- aggregate(steps ~ date, data, sum)

head(totalStepsDay)

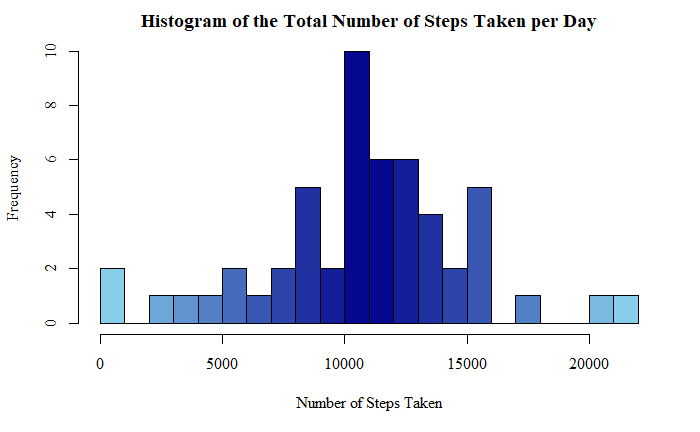
|  | **date**  <date> | **steps**  <dbl> |  |  |
| --- | --- | --- | --- | --- |
| 1 | 2012-10-02 | 126 |  |  |
| 2 | 2012-10-03 | 11352 |  |  |
| 3 | 2012-10-04 | 12116 |  |  |
| 4 | 2012-10-05 | 13294 |  |  |
| 5 | 2012-10-06 | 15420 |  |  |
| 6 | 2012-10-07 | 11015 |  |  |

paletteBlue <- colorRampPalette(c("skyblue", "darkblue", "skyblue"))

hist(totalStepsDay$steps, breaks=20, xlab="Number of Steps Taken",

main="Histogram of the Total Number of Steps Taken per Day",

col=paletteBlue(22), family="serif")



library(dplyr)

totalStepsSummary <- summarise(totalStepsDay, meanOfTotalSteps=mean(totalStepsDay$steps),

medianOfTotalSteps=median(totalStepsDay$steps))

print(totalStepsSummary)

| **meanOfTotalSteps**  <dbl> | **medianOfTotalSteps**  <dbl> |  |  |  |
| --- | --- | --- | --- | --- |
| 10766.19 | 10765 |  |  |  |

meanStepsInterval <- aggregate(steps ~ interval, data, mean)

head(meanStepsInterval)

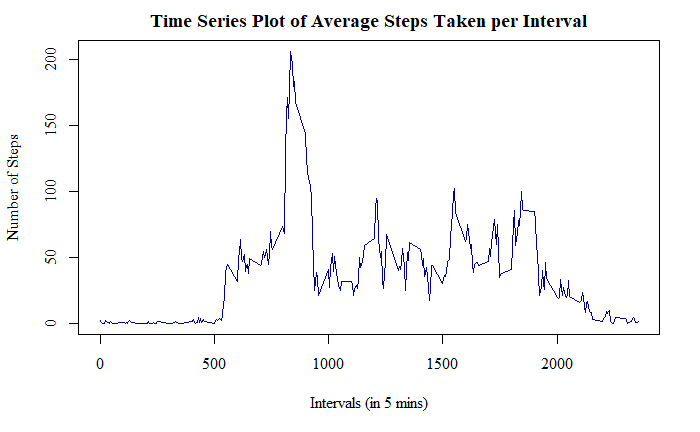
|  | **interval**  <dbl> | **steps**  <dbl> |  |  |
| --- | --- | --- | --- | --- |
| 1 | 0 | 1.7169811 |  |  |
| 2 | 5 | 0.3396226 |  |  |
| 3 | 10 | 0.1320755 |  |  |
| 4 | 15 | 0.1509434 |  |  |
| 5 | 20 | 0.0754717 |  |  |
| 6 | 25 | 2.0943396 |  |  |

plot(x=meanStepsInterval$interval, y=meanStepsInterval$steps, type="l",

main="Time Series Plot of Average Steps Taken per Interval",

ylab="Number of Steps", xlab="Intervals (in 5 mins)",

col="darkblue", lwd=1.5, family="serif")



meanStepsInterval[grep(max(meanStepsInterval$steps), meanStepsInterval$steps), ]

|  |
| --- |
|  |
|  | **interval**  <dbl> | **steps**  <dbl> |  |  |
| 104 | 835 | 206.1698 |  |  |

anyNA(initialData)

TRUE

data.frame(steps=sum(is.na(initialData$steps)),

interval=sum(is.na(initialData$interval)),

date=sum(is.na(initialData$date)))

|  |
| --- |
|  |
| **steps**  <int> | **interval**  <int> | **date**  <int> |  |  |
| 2304 | 0 | 0 |  |  |

imputedData <- initialData

for(x in 1:17568) {

if(is.na(imputedData[x, 1])==TRUE) {

imputedData[x, 1] <- meanStepsInterval[meanStepsInterval$interval %in% imputedData[x, 3], 2]

}

}

head(imputedData)

|  |
| --- |
|  |
| **steps**  <dbl> | **date**  <date> | **interval**  <dbl> |  |
| 1 | 1.7169811 | 2012-10-01 | 0 |  |
| 2 | 0.3396226 | 2012-10-01 | 5 |  |
| 3 | 0.1320755 | 2012-10-01 | 10 |  |
| 4 | 0.1509434 | 2012-10-01 | 15 |  |
| 5 | 0.0754717 | 2012-10-01 | 20 |  |
| 6 | 2.0943396 | 2012-10-01 | 25 |  |

imputedTotalStepsDay <- aggregate(steps ~ date, imputedData, sum)

head(imputedTotalStepsDay)

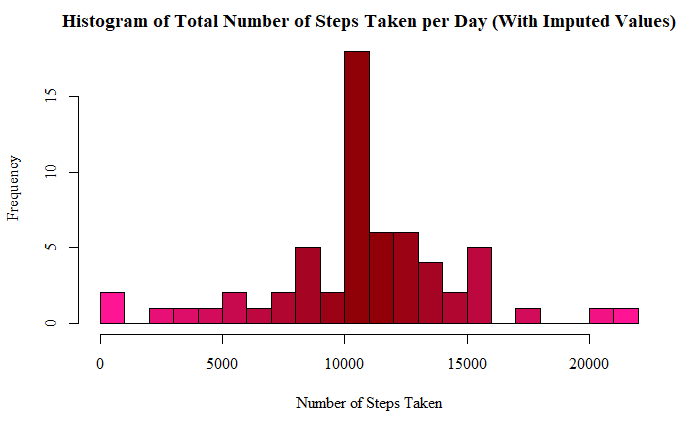
|  |
| --- |
|  |
| **date**  <date> | **steps**  <dbl> |  |  |
| 1 | 2012-10-01 | 10766.19 |  |  |
| 2 | 2012-10-02 | 126.00 |  |  |
| 3 | 2012-10-03 | 11352.00 |  |  |
| 4 | 2012-10-04 | 12116.00 |  |  |
| 5 | 2012-10-05 | 13294.00 |  |  |
| 6 | 2012-10-06 | 15420.00 |  |  |

paletteRed <- colorRampPalette(c("deeppink", "darkred", "deeppink"))

hist(imputedTotalStepsDay$steps, breaks=20, xlab="Number of Steps Taken",

main="Histogram of Total Number of Steps Taken per Day (With Imputed Values)",

col=paletteRed(22), family="serif")



imputedStepsSummary <- summarise(imputedTotalStepsDay,

meanOfTotalSteps=mean(imputedTotalStepsDay$steps),

medianOfTotalSteps=median(imputedTotalStepsDay$steps))

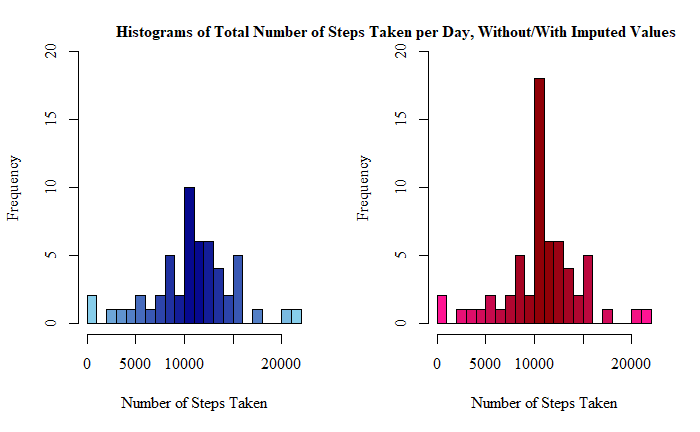
print(imputedStepsSummary)

| **meanOfTotalSteps**  <dbl> | **medianOfTotalSteps**  <dbl> |  |  |  |
| --- | --- | --- | --- | --- |
| 10766.19 | 10766.19 |  |  |  |

rbind(totalStepsSummary, imputedStepsSummary)

|  |
| --- |
|  |
| **meanOfTotalSteps**  <dbl> | **medianOfTotalSteps**  <dbl> |  |  |  |
| 10766.19 | 10765.00 |  |  |  |
| 10766.19 | 10766.19 |  |  |  |

|  |
| --- |
|  |
| **meanOfTotalSteps**  <dbl> | **medianOfTotalSteps**  <dbl> |  |  |  |
| 10766.19 | 10765.00 |  |  |  |
| 10766.19 | 10766.19 |  |  |  |



naByDate <- initialData[is.na(initialData$steps), ]

table(naByDate$date)

2012-10-01 2012-10-08 2012-11-01 2012-11-04 2012-11-09 2012-11-10 2012-11-14 2012-11-30

288 288 288 288 288 288 288 288

length(unique(data$interval))

288

daysData <- imputedData

daysData$days <- weekdays(daysData$date)

daysData$weekday <- as.character(rep(0, times=17568))

for(x in 1:17568) {

if(daysData[x, 4] %in% c("Saturday", "Sunday")) {

daysData[x, 5] <- "weekend"

} else {

daysData[x, 5] <- "weekday"

}

}

daysData$weekday <- factor(daysData$weekday)

head(daysData)

|  | **steps**  <dbl> | **date**  <date> | **interval**  <dbl> | **days**  <chr> | **weekday**  <fctr> |
| --- | --- | --- | --- | --- | --- |
| 1 | 1.7169811 | 2012-10-01 | 0 | Monday | weekday |
| 2 | 0.3396226 | 2012-10-01 | 5 | Monday | weekday |
| 3 | 0.1320755 | 2012-10-01 | 10 | Monday | weekday |
| 4 | 0.1509434 | 2012-10-01 | 15 | Monday | weekday |
| 5 | 0.0754717 | 2012-10-01 | 20 | Monday | weekday |
| 6 | 2.0943396 | 2012-10-01 | 25 | Monday | weekday |

weekdayData <- daysData[daysData$weekday=="weekday", ]

weekendData <- daysData[daysData$weekday=="weekend", ]

weekdayMean <- aggregate(steps ~ interval, weekdayData, mean)

weekendMean <- aggregate(steps ~ interval, weekendData, mean)

par(mfrow=c(2, 1), mar=c(4, 4.1, 3, 2.1))

plot(weekdayMean$interval, weekdayMean$steps, type="l",

main="Time Series Plot of Average Steps Taken per Interval, for Weekdays",

xlab="Intervals (in 5 mins)", ylab="Number of Steps", family="serif",

col="darkred", lwd=1.5, ylim=c(0, 230))

plot(weekendMean$interval, weekendMean$steps, type="l",

main="Time Series Plot of Average Steps Taken per Interval, for Weekends",

xlab="Intervals (in 5 mins)", ylab="Number of Steps", family="serif",

col="darkblue", lwd=1.5, ylim=c(0, 230))

