PA1_template

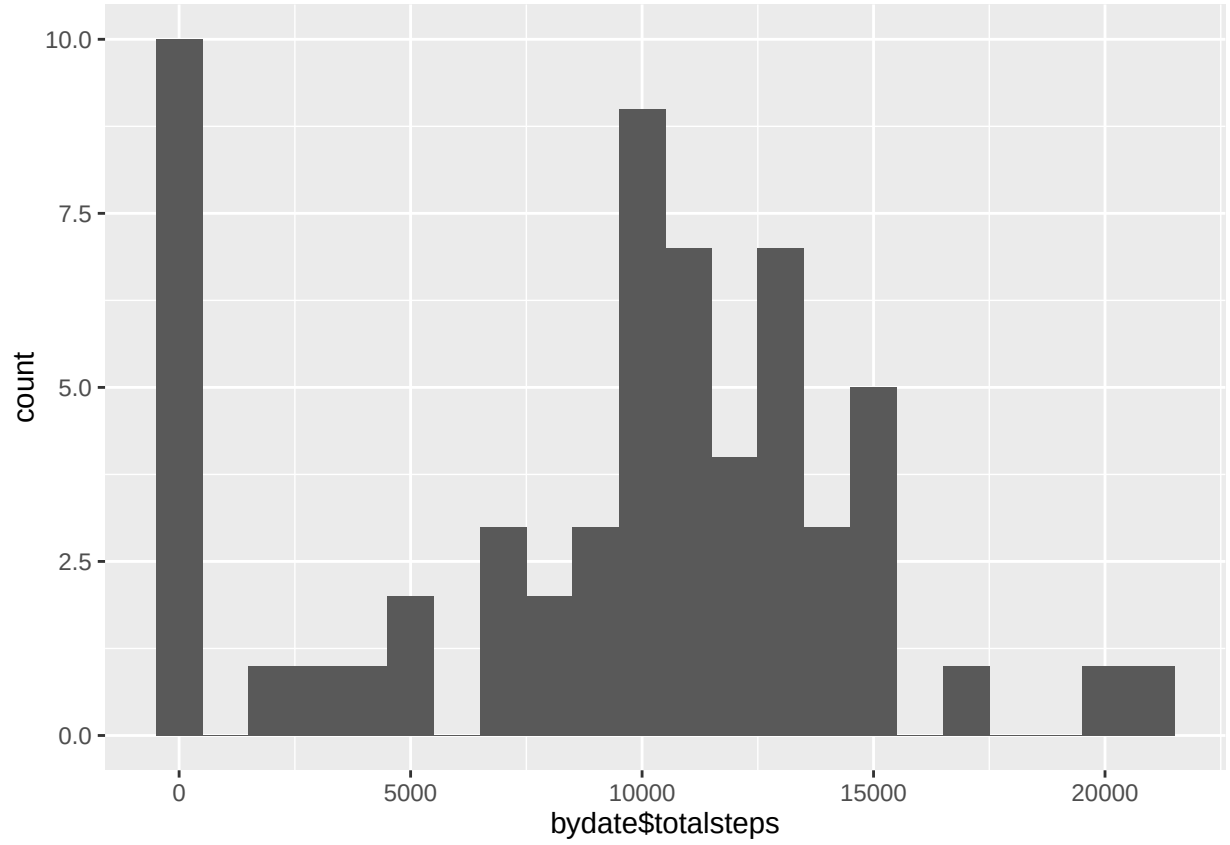
Add the necessary items from the library

Loading and preprocessing the data.

```
sd <- read.csv("activity.csv")
```

What is the mean total number of steps taken per day?

```
bydate <- ddply(sd, ~date, summarise, totalsteps = sum(steps,na.rm = TRUE))
qplot(bydate$totalsteps, binwidth = 1000)
```



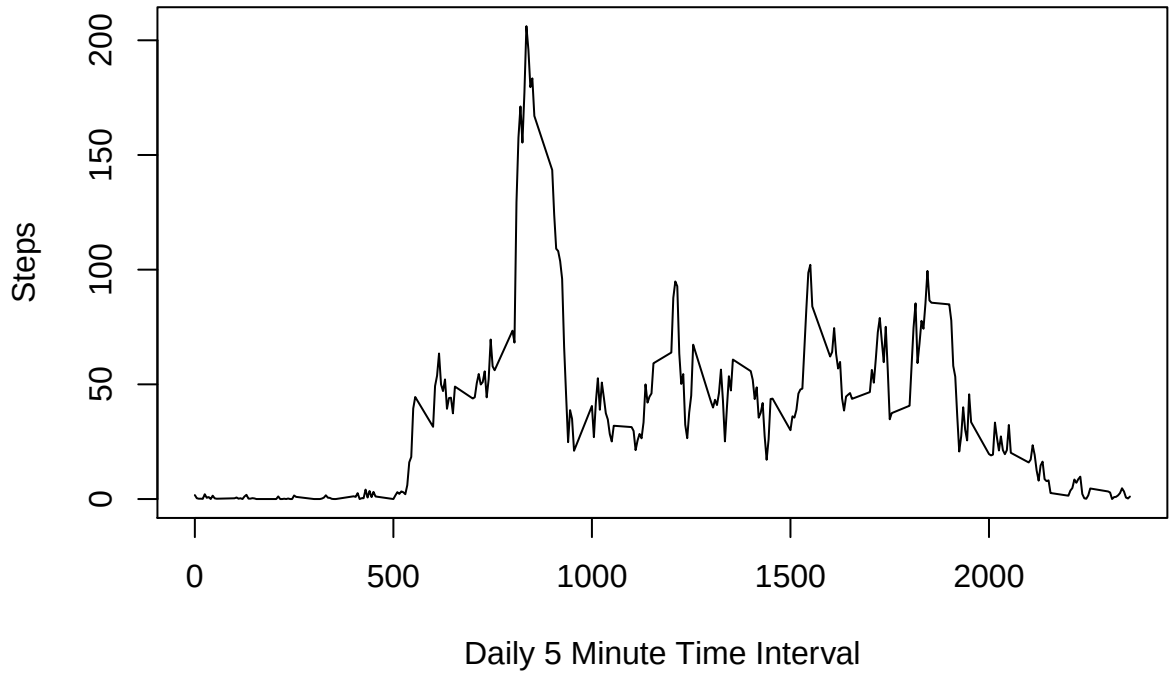

```
meanst <- mean(bydate$totalsteps, na.rm = TRUE)
medianst <- median(bydate$totalsteps, na.rm = TRUE)
paste("Mean = ", round(meanst,2), ", Median = " ,medianst)
```

```
## [1] "Mean =  9354.23 , Median =  10395"
```

As you can see the mean and median are 9354.2295082 and 10395 respectively.

What is the average daily activity pattern?

```
byinterval <- ddply(sd, ~interval, summarise, avgsteps = mean(steps,na.rm = TRUE))
plot(x = byinterval$interval, y = byinterval$avgsteps, type="l", xlab = "Daily 5 Minute Time Interval",
```




```
byinterval[which.max(byinterval$avgsteps),]
```

```
##     interval avgsteps
## 104      835 206.1698
```

As you can see the max above is 'r max(byinterval$avgsteps)'.

Imputting missing values

```
sum(is.na(sd$steps))
```

```
## [1] 2304
```

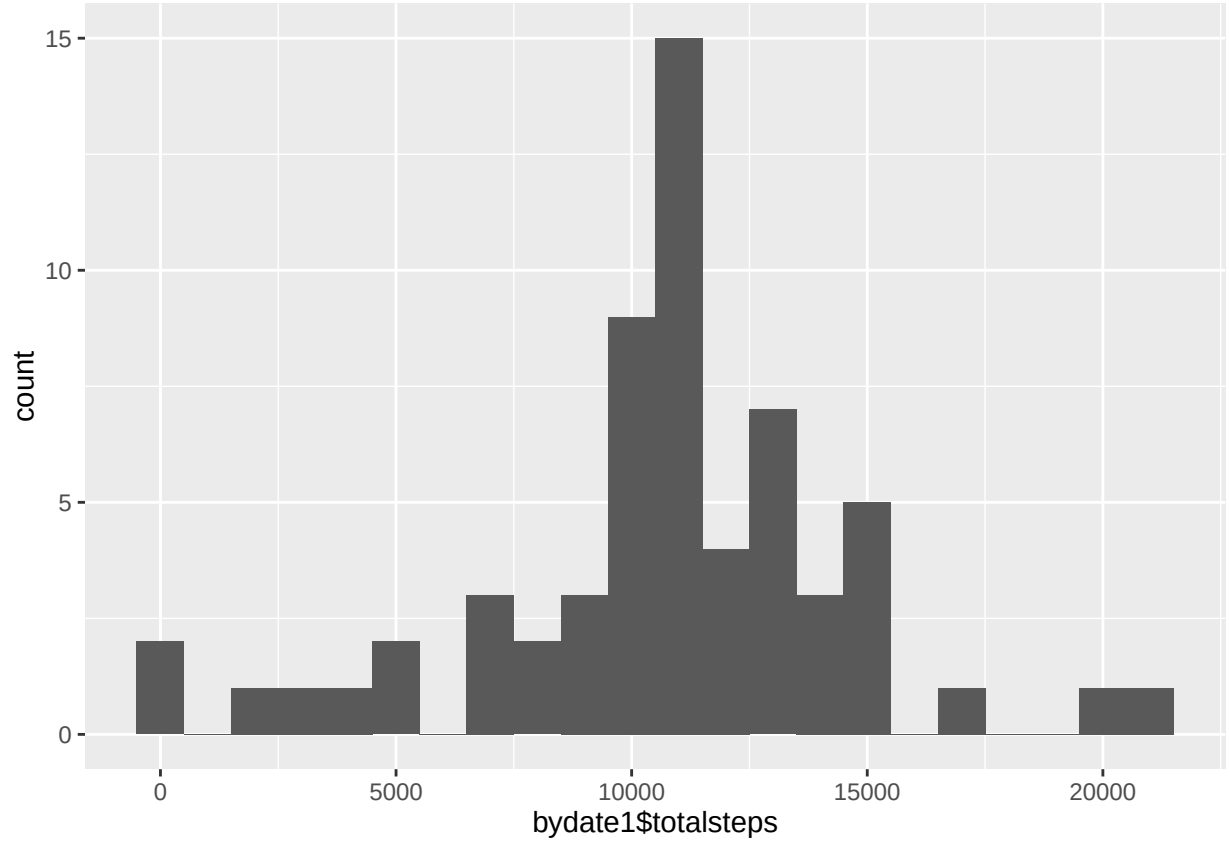
There are 2304 missing values in the database

To fill these values we will use the average for that time of day throughout the rest of the data.

```
## Loop through sd1 and replace any na's with the average for that interval
sd1 <- sd
for(i in 1:length(sd$steps))
  {
    if(is.na(sd$steps[i]))
    {
      sd1$steps[i] <- byinterval$avgsteps[i %% (24*12)+1]
    }
  }
```

Recalculate the mean and median and compare to the non-imputed data

```
bydate1 <- ddply(sd1, ~date, summarise, totalsteps = sum(steps,na.rm = TRUE))
qplot(bydate1$totalsteps, binwidth = 1000)
```




```
meanst1 <- mean(bydate1$totalsteps, na.rm = TRUE)
medianst1 <- median(bydate1$totalsteps, na.rm = TRUE)
data.frame(baseData = c(meanst, medianst), imputedData = c(meanst1, medianst1), row.names = c("Mean","M
```

```
##        baseData imputedData
## Mean    9354.23    10766.19
## Median 10395.00    10766.19
```

As you can see the mean and median both increased with our imputation of data. This is because some of the days had many na's that increased to the average values when we imputed the data.

Please note that the median and mean are the same for the imputed data. This is because with replacing na's with the average data from that interval has had the effect of creating some days with exactly the mean steps. These days ended up being the 50th percentile and becoming the mean.

Are there differences in activity patterns between weekdays and weekends?

```
sd1$date <- as.POSIXct(sd1$date)
sd1 <- transform(sd1, wday = format(sd1$date, "%w"))

sd1$wday <- as.character(sd1$wday)
sd1$wday[sd1$wday == "0" | sd1$wday =="6"] <- "Weekend"
sd1$wday[sd1$wday != "Weekend"] <- "Weekday"
sd1$wday <- as.factor(sd1$wday)
```

```
byint <- sd1 %>%
group_by(interval, wday) %>%
summarise_each(funs(mean))

p <- ggplot(byint, aes(x = interval, y = steps))
p + geom_line() + facet_grid(wday ~ .)
```