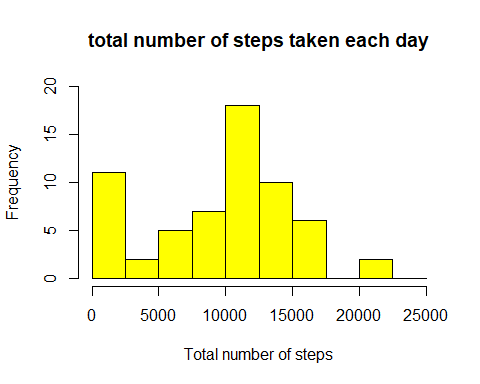
title: “histogram without na” output: word\_document: {r setup, include=FALSE} knitr::opts\_chunk$set(echo = TRUE)

```r  
##Loading Data  
  
setwd("D:/Private/Data\_Sceince/Reproducible Search/Projects/repdata\_data\_activity")  
data<-read.csv("activity.csv")  
##Preprocessing of data  
data$date<-as.Date(data$date)  
##Histogram of total number of steps taken on each day  
no\_of\_steps<-aggregate(data$steps,by=list(data$date),FUN=sum,na.rm=TRUE)

hist(no\_of\_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 number of steps   
mean(no\_of\_steps$x)

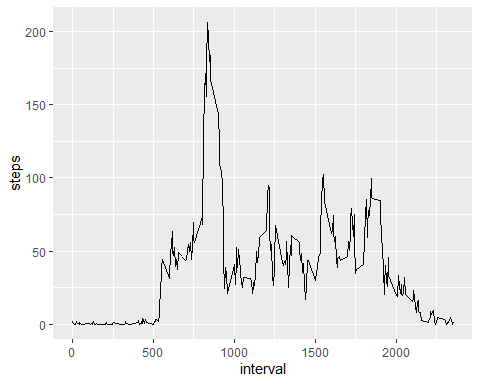
## [1] 9354.23

## [1] 9354.23  
median(no\_of\_steps$x)

## [1] 10395

## [1] 10395

##Time Series Plot  
avg\_steps<-aggregate(data$steps,by=list(data$interval),FUN=mean,na.rm=TRUE)  
  
colnames(avg\_steps)<-c("interval","steps")  
  
  
library(ggplot2)  
ggplot(aes(x=interval,y=steps),data=avg\_steps)+geom\_line()



## Average 5 minute interval  
avg\_steps[avg\_steps$steps==max(avg\_steps$steps),1]

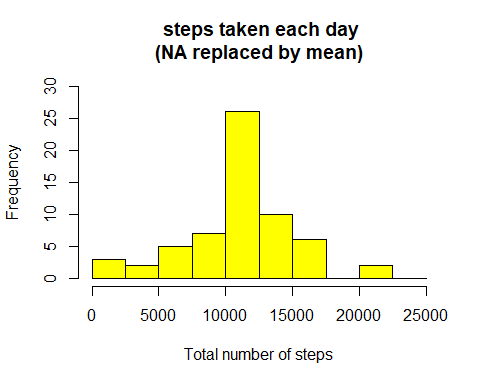
## [1] 835

## [1] 835  
  
##Imputing NA  
  
sum(is.na(data$steps))

## [1] 2304

## [1] 2304  
  
##Replace NA values with the mean of the steps  
  
data$steps[is.na(data$steps)]<-mean(data$steps,na.rm=TRUE)

##Histogram with Repaced NA values  
  
sum\_steps<-aggregate(data$steps,by=list(data$date),FUN=sum,na.rm=TRUE)   
  
hist(sum\_steps$x,   
 breaks=seq(from=0, to=25000, by=2500),  
 col="yellow",   
 xlab="Total number of steps",   
 ylim=c(0, 30),   
 main=" 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

## [1] 10766.19  
median(sum\_steps$x)

## [1] 10766.19

## [1] 10766.19   
  
## date into weekdays  
  
data$days=tolower(weekdays(data$date))  
  
##days into weekend and weekdays  
  
data$day\_type<-ifelse(data$days=="saturday"|data$days=="sunday","weekend","weekday")

## mean steps on weekend or weekday   
  
avg\_steps<-aggregate(data$steps,by=list(data$interval,data$day\_type),FUN=mean,na.rm=TRUE)  
  
colnames(avg\_steps)<-c("interval","day\_type","steps")  
  
## Create panel plot between average steps and interval  
  
ggplot(aes(x=interval,y=steps),data=avg\_steps)+geom\_line()+facet\_wrap(~avg\_steps$day\_type)

