Assigmnent_latest

Jameaba

October 18, 2015

An Analysis of Data on Activity

```
**Reading data into R Studio**


```r
stappes = read.table("activity.csv", header = T, sep =",")
stappdf = data.frame(stappes)
stappdf1 <- na.omit(stappdf)
```
A summary of variables in the activity dataset

```r
summary(stappes)
```


```
##      steps                  date            interval
## Min.   :  0.00   2012-10-01:  288   Min.   :   0.0
## 1st Qu.:  0.00   2012-10-02:  288   1st Qu.: 588.8
## Median :  0.00   2012-10-03:  288   Median :1177.5
## Mean   : 37.38   2012-10-04:  288   Mean   :1177.5
## 3rd Qu.: 12.00   2012-10-05:  288   3rd Qu.:1766.2
## Max.   :806.00   2012-10-06:  288   Max.   :2355.0
## NA's   :2304     (Other)   :15840
```
 ##Reading file into R Studio


```r
stappes = read.table("activity.csv", header = T, sep =",")
```

Creating a dataframe of the activity dataset


```r
stappdf = data.frame(stappes)
```
Total number of steps taken per day and naming the categorized   variables


```r
stappes2 <- aggregate.data.frame(stappes$steps,
by = list(stappes$date), FUN = sum)
names(stappes) <- c("date", "sumsteps")
```

Histogram of sum of steps across days


```r
library(ggplot2)
```


```
## Warning: package 'ggplot2' was built under R version 3.2.2
```


```r
hist.ggplot <- ggplot(stappes2, aes(x = x,color="blues9"))
g.his <- hist.ggplot +
geom_histogram()
print(g.his)
```


```
```

```
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
```



```r
stappes_msvas=sapply(stappes, function(x) sum(is.na(x)))
stappes_msvas
```

```
##     date sumsteps     <NA>
##     2304        0        0
```

Reported number of mssing values is **2304** for steps variable
#Filling in the missing values using the mean

Reported number of mssing values is **2304** for steps variable

*Filling in the missing values using the mean (steps) and median
for interval(?), respectively*

After filling in MA zero MAs reported

```r
stappes_msvas=sapply(stappes, function(x) sum(is.na(x)))
stappes_msvas
```

```
##     date sumsteps     <NA>
##     2304        0        0
```

##Average daily activity pattern

```r
stappes_msvas=sapply(stappes, function(x) sum(is.na(x)))
```
stappes_msvas
steps   date interval
**2304   0   0 **

```r
stappes_misva2<-colSums(is.na(stappes))
```
stappes_misva2
steps   date interval
**2304   0   0 **

# data frame with MAs steps and interval variables filled
frame(stappes$steps, stappes$date, stappes$interval)
```
*Histogram of the steps taken by day with MA filled*

*The mean and medium of total steps by day*
**Mean of steps is 37.38; median is zero**

*The interval that has the largest number of steps occurs on          2012-10-06*

##Activity patterns on weekends and weekdays

*Converting dates into weekends and weekdays*

 datefilled <-data.frame(stappes_ma_filled[,2])

**mydates <- as.Date(datefilled) mydadates1 <-weekdays(mydates)**

```
str(stappes_ma_filled)
creating 'differentiators', the two levels and labels

*mydates2<- c("Monday", "Tuesday", "Wednesday",
 "Thursday","Friday"*

**mydates3 <- factor((weekdays(mydates) %in% mydates2),
levels=c(FALSE, TRUE), labels = c("Weekend", "weekday"))**
```
*Creation of dataframe with a new factor variable*

**{r, echo = TRUE}
stappeplete1<-data.frame(stappes_ma_filled, mydate4$date_by_day)
names(stappeplete1) <- c("steps",  "date",  "interval",
"day_week")**

*Making things abit less cluttered*

**{r, echo=TRUE}
stapt <- data.frame(stappeplete1)**
```
Creating subsets of weekend days and weekday days

**{r, echo= TRUE}

(stapt2 <- stapt[stapt$day_week == "weekend",])

(stapt3 <- stapt[stapt$day_week == "weekday",])**

Plot intervals by weekends and weekdays

**{r, fig.width=7, fig.height=5, echo=TRUE}
windows()
par(mfrow = c (2,2))
plot(stapt2$interval, stapt2$steps,type="l", xlab   ="Interval",   ylab ="Steps on weekends", col = "blue")
plot(stapt3$interval, stapt3$steps, type="l", xlab = "Interval",   ylab="Steps on weekdays", col ="magenta"
)**
```

*There is apparently a marked difference in the pattern of activity   on weekdays and weekends. On weekdays a ctvity is very intensive
during early mmorning, peaks somewhat quickly and starts to fall
off, albeit with intermitent spurts along the way. For the weekends
activity seems intense during later hours. It is also important to
note that the concentration of pockets of time of intense activity
seems to occur more in batches with respect to weekends than on
weekdays. Another important difference between the two categories
lies in the high concentration of activity (more compact) compared
with somewhat slightly paced activity on weekends.
On weekdays and weekends. On weekdays activity is very intensive
during early morning, peaks somewhat quickly and starts to fall
off, albeit with intermitent spurts along the way. For the weekends
activity seems intense during later hours. It is also important to
note that the concentration of pockets of time of intense activity
seems to occur more in batches with respect to weekends than on
weekdays. Another important difference between the two categories
lies in the high concentration of activity (more compact) compared
with somewhat slightly paced activity on weekends.*