Stat 123 Homework 2

Jonathan Wilson

January 6, 2019

```
knitr::opts_knit$set(root.dir =
"C:\\Users\\jon\\Documents\\School\\R\\HW\\HW2")
```

Using the file given in the instructions, read the daily interest in lds.org for 2014-06-09 through 2014-09-06 into an data.frame.

```
#setwd("C:\\Users\\jon\\Documents\\School\\R\\HW\\HW2")
lds <- read.table("reportclean.csv", sep=",", header=TRUE,
                  stringsAsFactors=FALSE)
#lds
#str(lds)
#lds <- read.csv("reportclean.csv", stringsAsFactors=FALSE)
```

Name the variables in the data.frame “date” and “interest”.

```
names(lds) <- c("date","interest")
```

Replace the “date” character vector in the data.frame into a vector of Date objects of the same name.

```
lds$date <- as.Date(lds$date, format="%m/%d/%Y")#, format = "%m/%d/%Y"
```

Use the “weekdays” function to define a new variable in the data.frame called “weekday” from the “date” variable.

```
lds$weekday <- weekdays(as.Date(lds$date))
```

Find the mean of the interest variable across all Mondays.

```
mondays <- lds[lds$weekday=="Monday",]
mean.monday <- mean(mondays$interest)
mean.monday

## [1] 26.76923
```

Find the mean of the interest variable for the weekends.

```
weekends <- lds[lds$weekday %in% c("Sunday","Saturday"),]
mean.weekends <- mean(weekends$interest)
mean.weekends

## [1] 54.28
```

Find the mean of the interest variable for all weekdays.

```r
#By weekdays I am assuming that you mean only Mon-Fri. I could easily run a
mean for either.
weekdays <- subset(lds, !(lds$weekday %in% c("Sunday","Saturday")))
mean.weekdays <- mean(weekdays$interest)
mean.weekdays
```

```
## [1] 21.13846
```

Comment on whether you believe there is a statistically significant difference between the mean interest on weekdays and weekends? You don't have to do a formal test. We're asking for just your intuition after seeing the data. Since your response won't be executable code, do not include it in an r chunk, or if you do, prefix your comments with R's comment character, #.

What are the median, minimum, and maximum interest?

```r
#The mean is much higher on the weekends. I would say it is statistically
significant since more people are on
#during the weekends. That is my idea.
median.interest <- median(lds$interest)
median.interest
```

```
## [1] 22
```

```r
min.interest <- min(lds$interest)
min.interest
```

```
## [1] 14
```

```r
max.interest <- max(lds$interest)
max.interest
```

```
## [1] 100
```

What is the correlation between interests one day apart?

```r
lag <- lds[1:89,]
x <- rep(c("2014-05-09",0,"Sunday"), ncol(lag))
lag <- rbind(x, lag)#Create lag in rows by one day
lag$interest <- as.integer(lag$interest)
cor(lds$interest, lag$interest)
```

```
## [1] 0.05095246
```