Exploratory Time Analysis

Landon C

August 8, 2015

Mapping

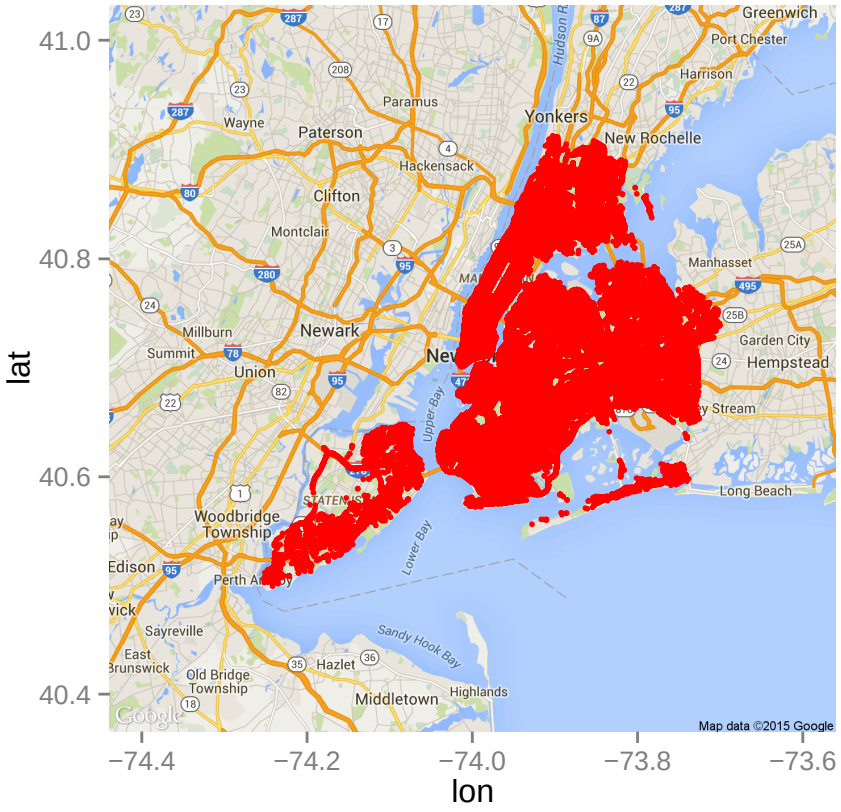
Getting a map with a location of all accidents

```
library(ggmap)
library(lubridate)

accident = read.csv("accident2015.csv", header=TRUE)
al1 = get_map(location = c(lon = -74., lat = 40.7), zoom = 10, maptype = 'roadmap')
```

```
## Map from URL : http://maps.googleapis.com/maps/api/staticmap?center=40.7,-74&zoom=10&size=640x640&sca
```

```
al1MAP = ggmap(al1)

#al1MAP
al1MAP + geom_point(data = accident, aes(x = LONGITUDE, y = LATITUDE), colour = "red", size = 1)
```

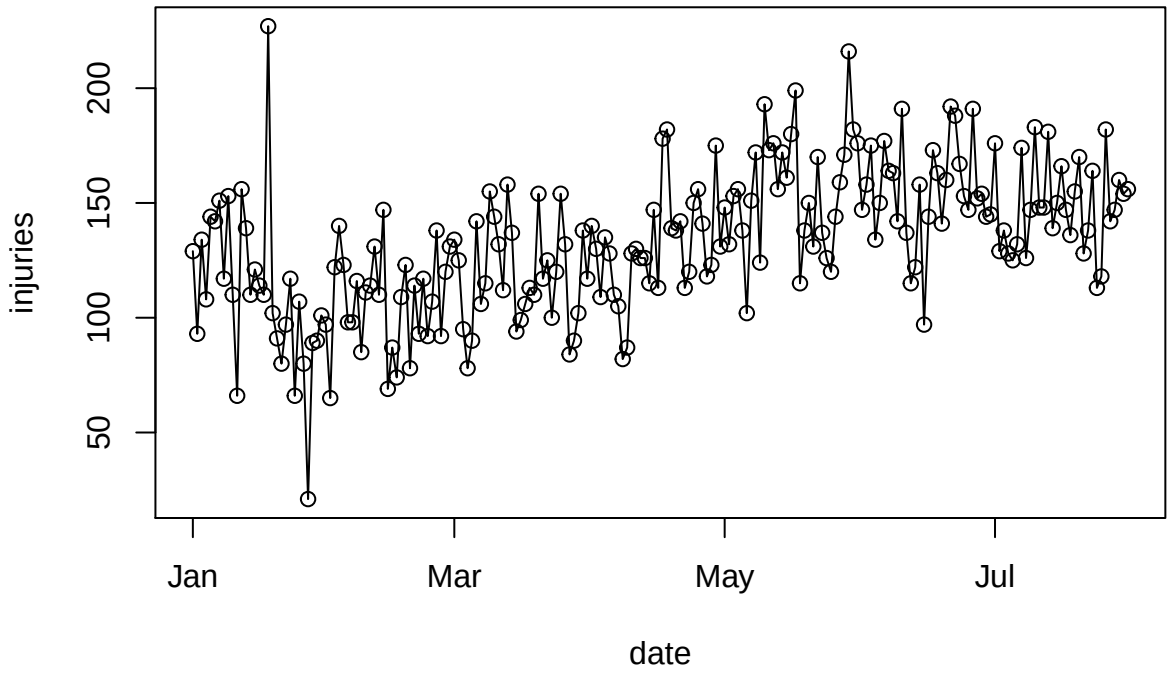
```
## Warning: Removed 20097 rows containing missing values (geom_point).
```



#Investigating Injuries Over time Investigating injuries over time for whole data set. Because the number of fatalities overall is not very high, we are not interested in this statistic. However, the number of injuries per day is high enough to possibly predict.

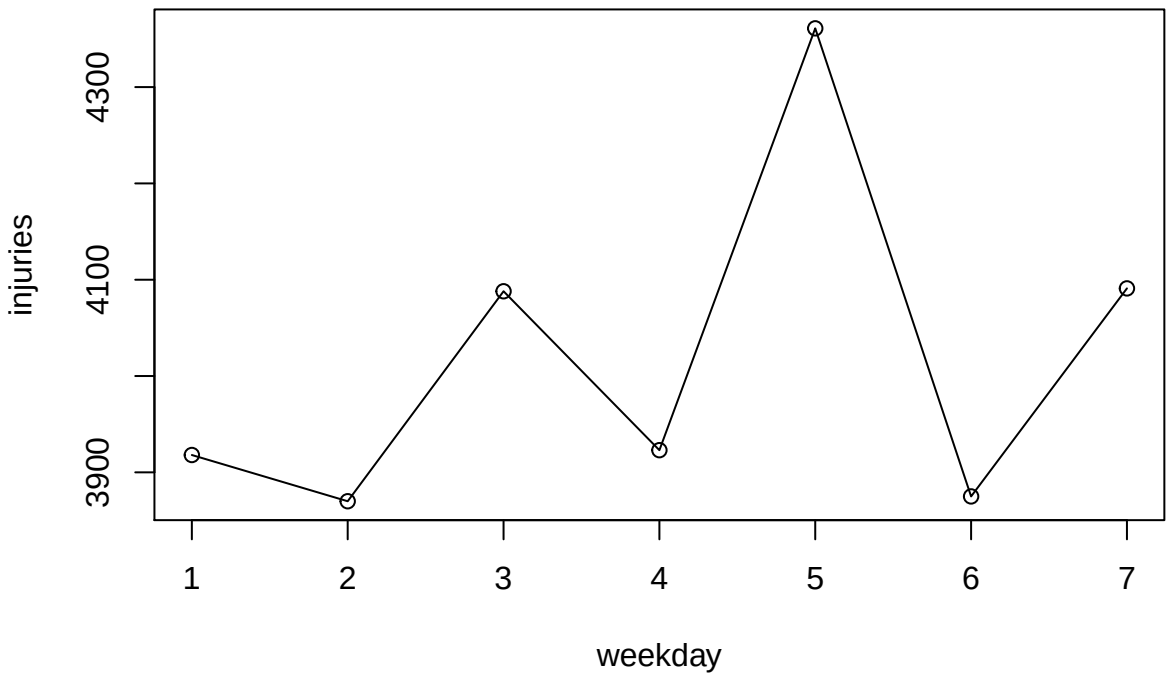
```
####data cleaning
#turn each date into a date object recognizable by computer
accident$DATE = as.Date(accident$DATE, "%m/%d/%Y")

#make month, day, and day of week columns
accident$MONTH = month(accident$DATE)
accident$DAY = day(accident$DATE)
accident$WEEKDAY = weekdays(as.Date(accident$DATE))

#get vector of unique dates as well as months and days, and dayofweek. This will be used when we loop t
dates = unique(accident$DATE)
months = unique(month(dates))
days = unique(day(dates))
weekdays = c("Monday", "Tuesday", "Wednesday","Thursday","Friday","Saturday","Sunday")

#count the total injuries in each day
accident_injuries = 0*c(1:length(dates))
day_index = 1
for (date in dates){
  injuries_day = subset(accident, DATE == date)
  accident_injuries[day_index] =  sum(injuries_day$NUMBER.OF.PERSONS.INJURED)
  day_index = day_index + 1
}
```

```r
#make dataframe of dates and injuries per day plot injuries over dates
date_injuries = data.frame(date = dates, injuries = accident_injuries)
plot(injuries ~ date, date_injuries)
lines(injuries ~ date, date_injuries, type = "l")
```



Investigating Injuries Throughout the Week

```r
#count the total injuries for each day of the week
injuries_ct = 0*c(1:7)
day_index = 1
for (weekday in weekdays){
  injuries_weekday = subset(date_injuries, weekdays(as.Date(date)) == weekday)
  injuries_ct[day_index] =  sum(injuries_weekday$injuries)
  day_index = day_index + 1
}

#plot the injuries over days
day_of_week = c(1:7) #Monday = 1, Tuesday = 2,.., Sunday = 7
daily_injuries = data.frame(weekday = day_of_week, injuries = injuries_ct)
plot(injuries ~ weekday, daily_injuries)
lines(injuries ~ weekday, daily_injuries, type = "l")
```


injuries
3900
4100
4300
1
2
3
4
5
6
7
weekday