

Road Traffic Injuries

B. Shilpa - CSE
K. Ashwini - CSE
Rubina Sultana - CSE

BVRIT-H

17-June-2016

What is Dataset

A Dataset is a collection of related data, usually presented in tabular form.

Dataset Chosen :

Road Traffic Injuries

Github Repository :

<https://github.com/ShilpaBatthineni/Road-Traffic-Injuries>

Dataset

File Edit Code View Plots Session Build Debug Tools Help

Go to file/function Addins Project: [None]

map.txt x rtia x year x Untitled1* x Untitled2* x

Filter

	Accident_Index	Longitude	Latitude	Police_Force	Accident_Severity	Number_of_Vehicles	Number_of_Casualties	Date	Month	Year	Day_of_Week	Time	X1st_Road_Class	X1st_Road
1	200901BS70001	-0.201349	51.51227	1	2	2	1	01-01-09	Jan	2009	5	15:11	6	
2	200901BS70002	-0.199248	51.51440	1	2	2	11	05-01-09	Jan	2009	2	10:59	5	
3	200901BS70003	-0.179599	51.48667	1	3	2	1	04-01-09	Jan	2009	1	14:19	3	
4	200901BS70004	-0.203110	51.50780	1	2	2	1	05-01-09	Jan	2009	2	8:10	3	
5	200901BS70005	-0.173445	51.48208	1	2	2	1	06-01-09	Jan	2009	3	17:25	3	
6	200901BS70006	-0.185525	51.49341	1	3	2	3	01-01-09	Jan	2009	5	11:48	6	
7	200901BS70007	-0.178561	51.48018	1	2	2	1	08-01-09	Jan	2009	5	13:58	3	
8	200901BS70008	-0.178524	51.49196	1	3	1	1	02-01-09	Jan	2009	6	13:18	5	
9	200901BS70009	-0.167395	51.49646	1	3	1	2	07-01-09	Jan	2009	4	12:15	6	
10	200901BS70010	-0.183275	51.48115	1	3	1	1	10-01-09	Jan	2009	7	9:52	3	
11	200901BS70011	-0.173445	51.48208	1	3	2	1	07-01-09	Jan	2009	4	0:09	3	
12	200901BS70012	-0.183013	51.49500	1	3	1	1	16-01-09	Jan	2009	6	17:49	4	
13	200901BS70015	-0.206779	51.49878	1	3	2	1	12-01-09	Jan	2009	2	14:00	3	
14	200901BS70016	-0.209082	51.50619	1	3	2	1	09-01-09	Jan	2009	6	8:15	3	
15	200901BS70017	-0.169548	51.49308	1	3	2	1	17-01-09	Jan	2009	7	12:15	3	
16	200901BS70019	-0.173445	51.48208	1	2	2	1	25-01-09	Jan	2009	1	22:05	3	
17	200901BS70020	-0.169724	51.48867	1	3	2	1	26-01-09	Jan	2009	2	17:30	4	
18	200901BS70021	-0.186108	51.48236	1	3	1	1	26-01-09	Jan	2009	2	17:05	6	
19	200901BS70023	-0.176861	51.49391	1	3	2	1	19-01-09	Jan	2009	2	14:27	5	
20	200901BS70024	-0.194837	51.50930	1	3	1	1	27-01-09	Jan	2009	3	0:28	3	
21	200901BS70025	-0.188919	51.50228	1	3	1	1	21-01-09	Jan	2009	4	23:15	3	
22	200901BS70026	-0.184005	51.50750	1	2	1	1	22-01-09	Jan	2009	5	22:15	3	

Showing 1 to 23 of 615,764 entries

Console

Types of Data

Data is broadly classified into 2 types

- Qualitative data.
- Quantitative data

Each is broken down into sub-types:

- Qualitative data can be ordinal and nominal.
- Quantitative data can be discrete (often, integer) , continuous, Interval scale and Ratio scale.

Phases of data

- Requirements
- Collection :

<https://data.gov.uk/dataset/road-accidents-safety-data/resource/1ae84544-6b06-425d-ad62-c85716a80022>

<https://data.gov.uk/dataset/road-accidents-safety-data/resource/0dc990b8-10e9-40d1-a6fd-f34efccd2799>

Size Of Data : **252 MB**

Phases of data...

- Cleaning : Sorted the data and removed all missing values
- Exploring :

Simple Findings

- ① At what time zone more accidents occur.
- ② Severity Of Accidents In a Week and which day of week has more no of accidents.
- ③ Severity Of Accidents In Month and which month has highest rate of Accidents.

Phases of data...

- ① Severity of accidents , no of accidents in each mode and determine which mode has more frequent occurrence of accidents.

Complex Findings

- ① Using Non spatial analysis(chi- Square) we can validate accidents that are distributed uniformly according to the different types variables for the accidents.
- ② Spatial Analysis to specify the locations where the crashes occurred, and to assess specific patterns of distribution through map visualization.

Phases of data...

- Modelling :

Statistical Technique : Correlation is a statistical technique that can show how strongly pairs of variables are related.

- Transforming

- Visualizing :

Bar plots, Ggplots, Ggmaps, Line charts

Findings With Code

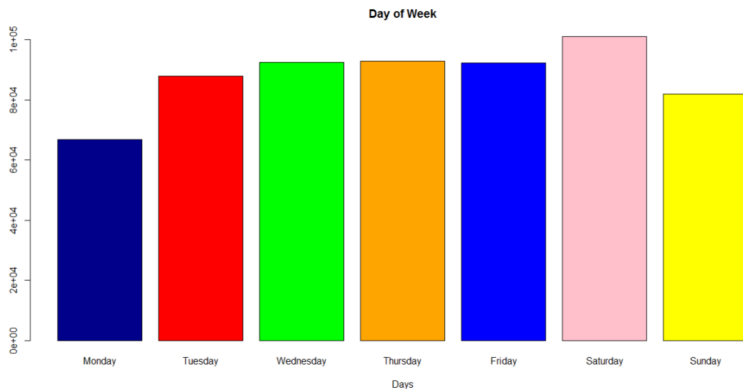


Figure: Days Of Week

Findings With Code...



```
1
2
3 days<-table(DfTRoadSafety_Accidents$Day_of_week)
4 > barplot(days, main="Day of week", xlab="Days",
5   col=c("darkblue","red","green","orange","blue","pink","yellow"),
6   names.arg = c("Monday","Tuesday","wednesday","Thursday","Friday","Saturday","Sunday"))
7
```

The image shows a screenshot of an R script editor window. The window has a title bar with standard icons and a menu bar with 'Source on Save', 'Run', and 'Source'. The script content is as follows:

```
1
2
3 days<-table(DfTRoadSafety_Accidents$Day_of_week)
4 > barplot(days, main="Day of week", xlab="Days",
5   col=c("darkblue","red","green","orange","blue","pink","yellow"),
6   names.arg = c("Monday","Tuesday","wednesday","Thursday","Friday","Saturday","Sunday"))
7
```

The status bar at the bottom indicates '3:1 (Top Level)' and 'R Script'.

Figure: Days Of Week Code

Findings With Code...

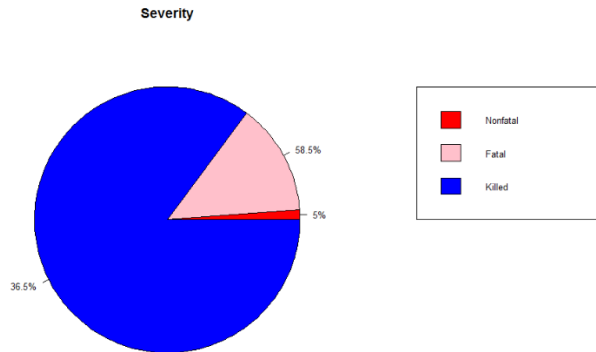
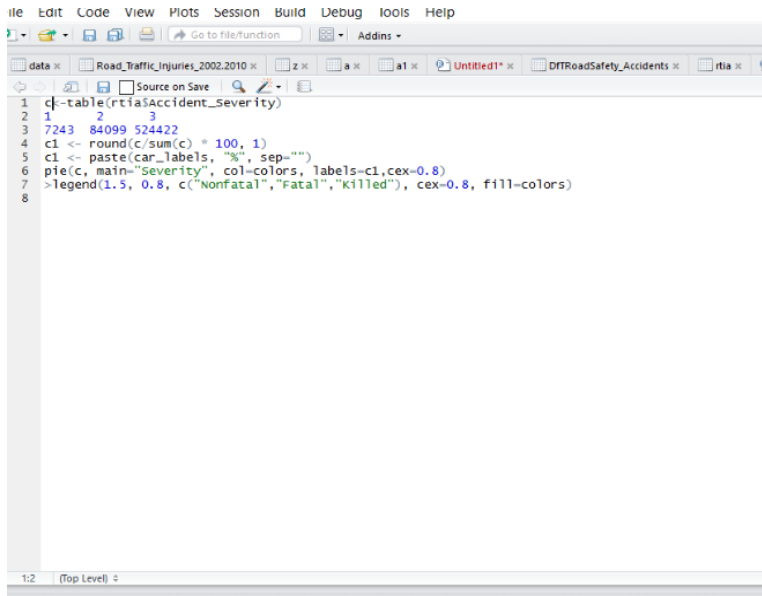


Figure: Severity

Findings With Code...



The screenshot shows the RStudio IDE interface. The menu bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Tools, and Help. The toolbar contains icons for file operations and a search bar. The tab bar shows several open files: data, Road_Traffic_Injuries_2002.2010, z, a, a1, Untitled1*, DfTRoadSafety_Accidents, and rtia. The source editor displays the following R code:

```
1 c<-table(rtia$Accident_Severity)
2 1      2      3
3 7243  84099 524422
4 c1 <- round(c/sum(c) * 100, 1)
5 c1 <- paste(car_labels, "%", sep="")
6 pie(c, main="Severity", col=colors, labels=c1,cex=0.8)
7 >legend(1.5, 0.8, c("Nonfatal","Fatal","Killed"), cex=0.8, fill=colors)
8
```

The status bar at the bottom indicates the current position is 1:2 (Top Level).

Findings With Code...

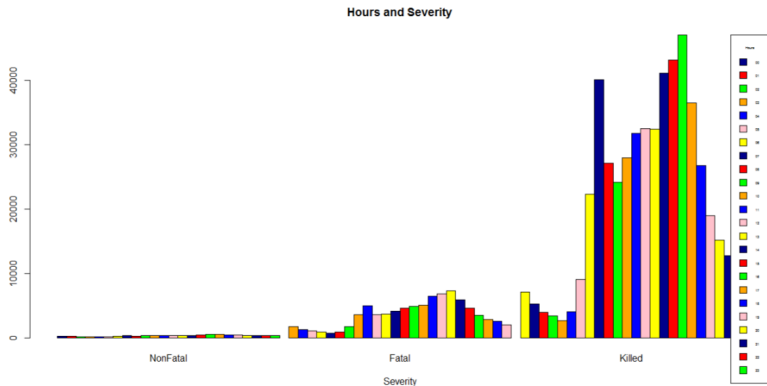


Figure: Hours and Severity

Findings With Code...

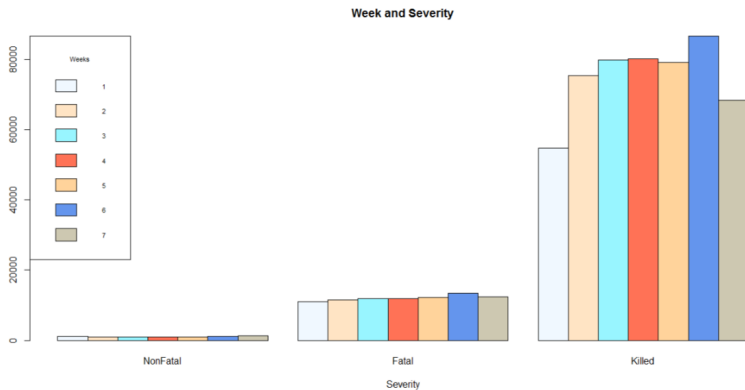


Figure: Weeks and Severity

Findings With Code...

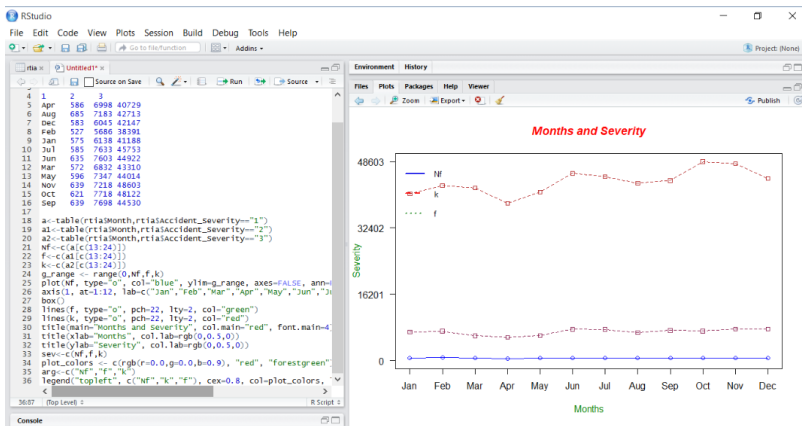


Figure: Month and Severity

Findings With Code...

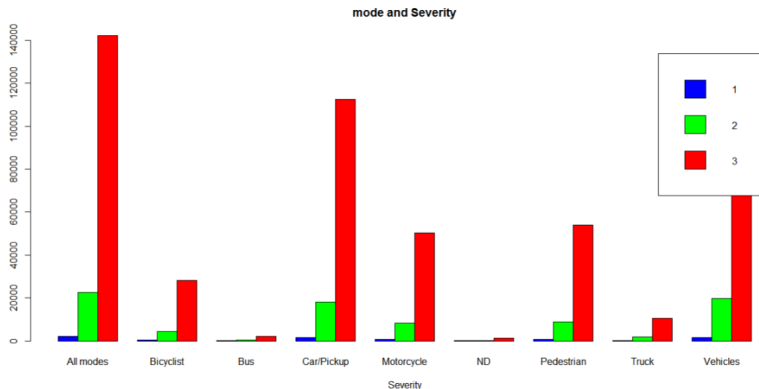


Figure: Modes and Severity

Findings With Code...

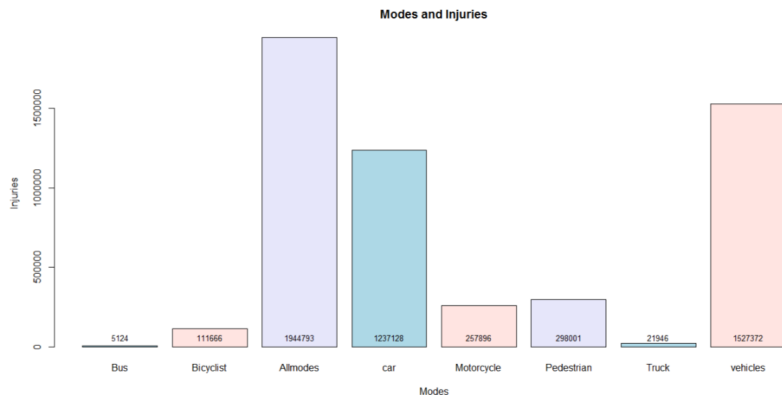


Figure: Modes and Injuries

Findings With Code...

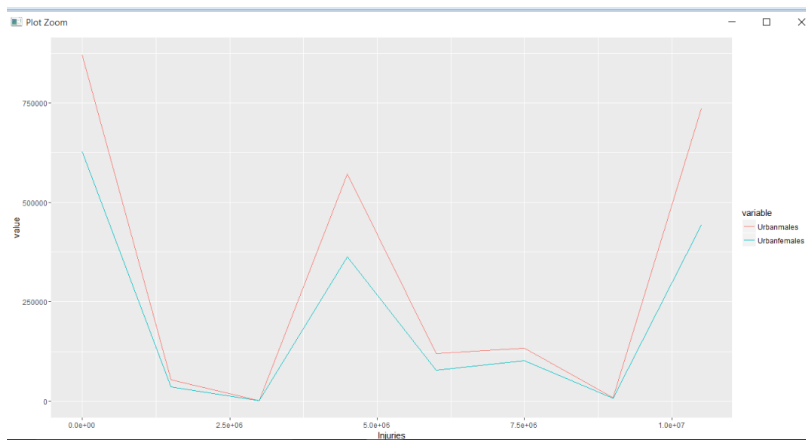


Figure: Urban

Findings With Code...

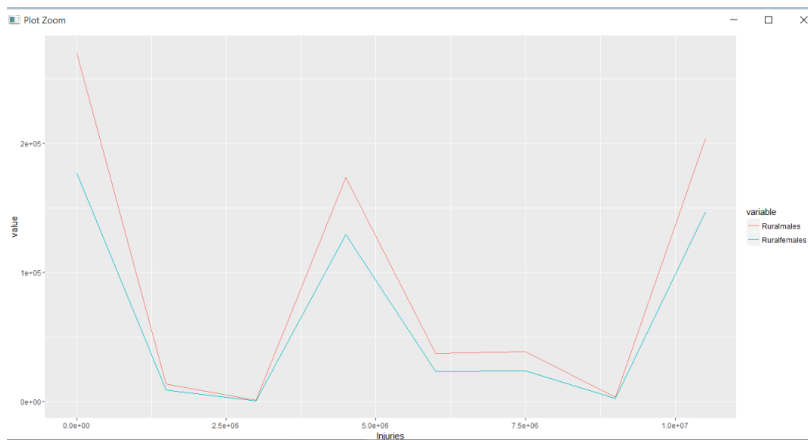


Figure: Rural

Findings With Code...



File Edit Code View Plots Session Build Debug Tools Help

```
1 > sam<-data.frame(v1=rtia$urban_or_urban_Area,v2=rtia$Mode,v3=rtia$Sex_of_casualty,v4=rtia$Injuries)
2 > Urban<-subset(sam,v1=="urban")
3 > Rural<-subset(sam,v1=="rural")
4 > UAllmodes<-subset(Rural,v2=="All modes")
5 > UBicyclist<-subset(Rural,v2=="Bicyclist")
6 > Bus<-subset(Rural,v2=="Bus")
7 > Car<-subset(Rural,v2=="Car/Pickup")
8 > Motorcycle<-subset(Rural,v2=="Motorcycle")
9 > Pedestrian<-subset(Rural,v2=="Pedestrian")
10 > Truck<-subset(Rural,v2=="Truck")
11 > vehicles <-subset(Rural,v2=="vehicles")
12 > UAllmodesm<-subset(UAllmodes,v3=="male")
13 > view(UAllmodesm)
14 > UAllmodesf<-subset(UAllmodes,v3=="female")
15 > UBicyclistm<-subset(UBicyclist,v3=="male")
16 > UBicyclistf<-subset(UBicyclist,v3=="female")
17 > UBusm<-subset(Bus,v3=="male")
18 > UBusf<-subset(Bus,v3=="female")
19 > UCarf<-subset(Car,v3=="female")
20 > UCarf<-subset(Car,v3=="female")
21 > UMotorcyclef<-subset(Motorcycle,v3=="female")
22 > UMotorcyclef<-subset(Motorcycle,v3=="female")
23 > UPedestrianm<-subset(Pedestrian,v3=="male")
24 > UPedestrianf<-subset(Pedestrian,v3=="female")
25 > UTruckm<-subset(Truck,v3=="male")
26 > UTruckf<-subset(Truck,v3=="female")
27 > Uvehiclesm <-subset(vehicles,v3=="male")
28 > Uvehiclesf <-subset(vehicles,v3=="female")
29 > Usm<-sum(UAllmodesm$v4)
30 > Usf<-sum(UAllmodesf$v4)
31 > U1Allmodes<-subset(Urban,v2=="All modes")
32 > U1Bicyclist<-subset(Urban,v2=="Bicyclist")
33 > U1Bus<-subset(Urban,v2=="Bus")
34 > U1Car<-subset(Urban,v2=="Car/Pickup")
35 > U1Motorcycle<-subset(Urban,v2=="Motorcycle")
```

123:84 (Top Level) ↕

Console



Findings With Code...

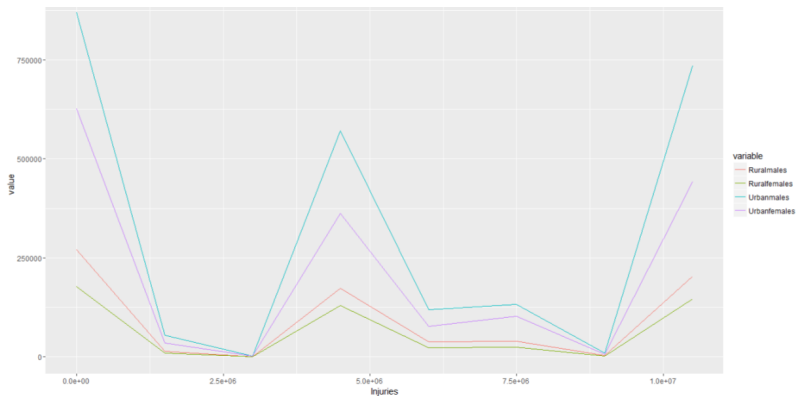


Figure: Ages, Gender And Injuries

Findings With Code...

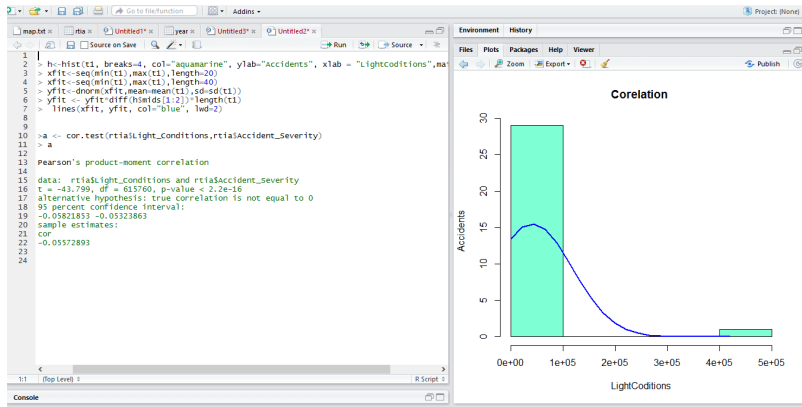


Figure: correlation

Findings With Code...

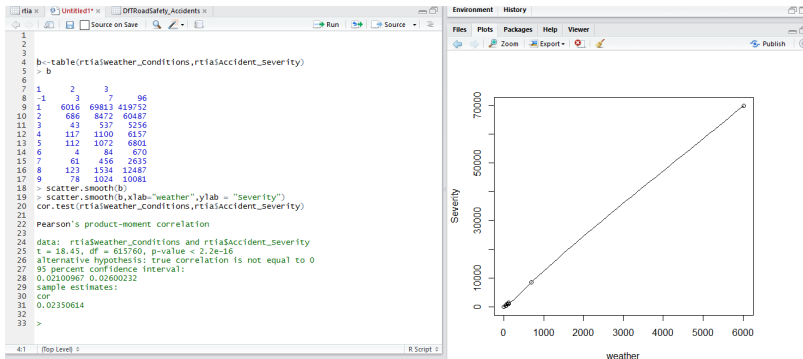


Figure: correlation

Analysis

Analysis is classified into 6 types :

- ① Descriptive
- ② Exploratory
- ③ Inferential
- ④ Predictive
- ⑤ Casual
- ⑥ Mechanist

Spatial Analysis :

Spatial analysis was used to geographically specify the locations where the crashes occurred, and to assess specific patterns of distribution through map visualization.

Non Spatial Analysis :

Using Non spatial analysis(chi- Square) we can validate accidents that are distributed uniformly according to the different types variables and whether or not the variables have significant relationships with them.

Findings With Code...

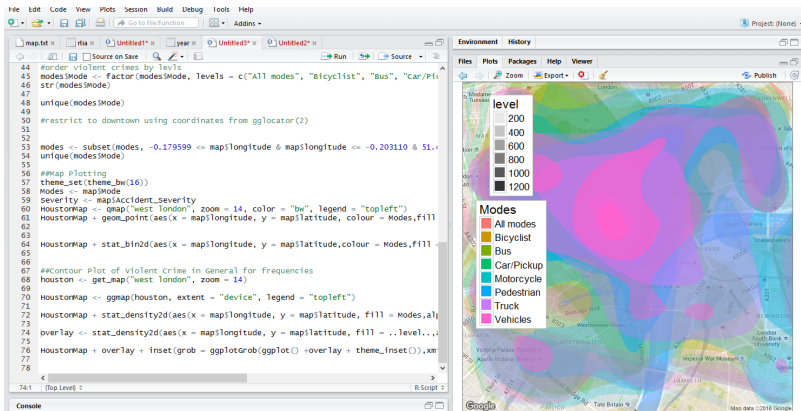


Figure: Spatial analysis

Findings With Code...

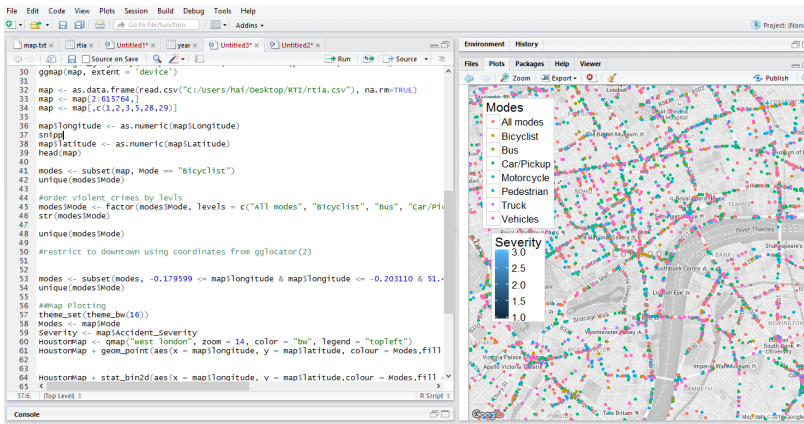


Figure: Spatial analysis

Findings With Code...

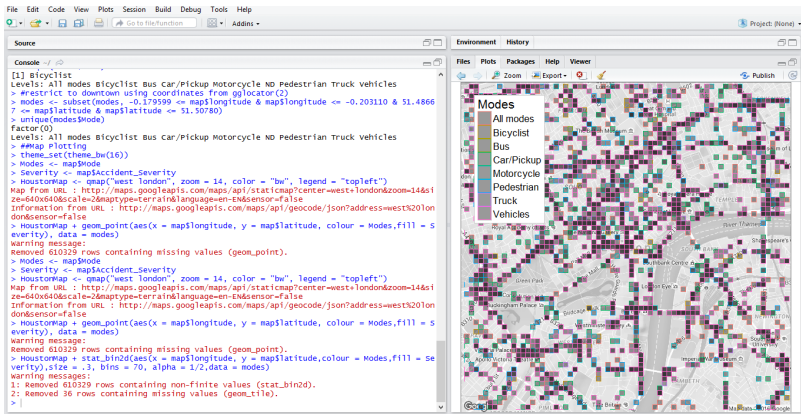


Figure: Spatial analysis

Findings With Code...

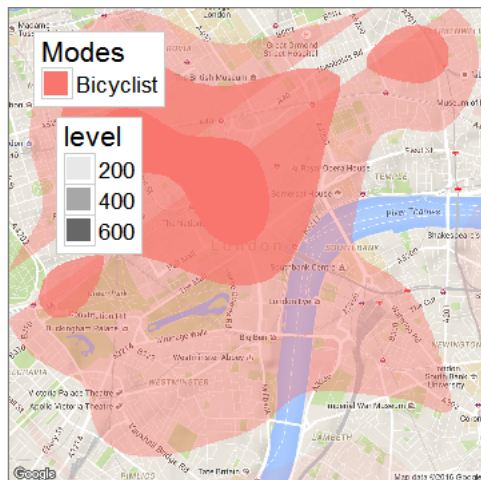


Figure: Spatial analysis

Findings With Code...

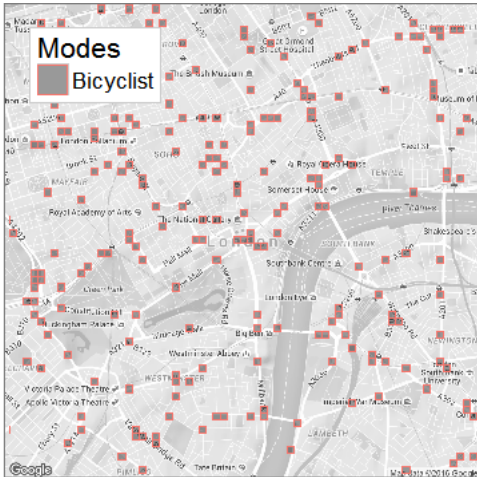


Figure: Spatial analysis

Findings With Code...

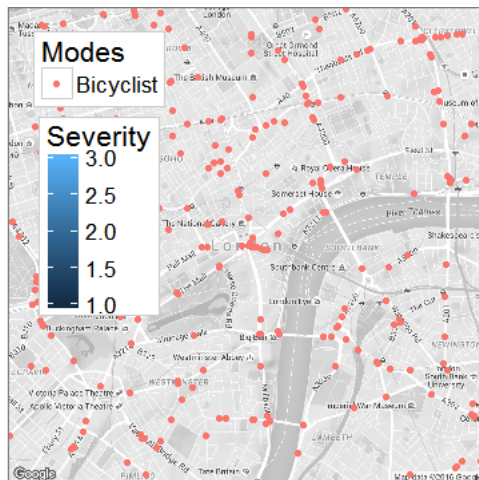
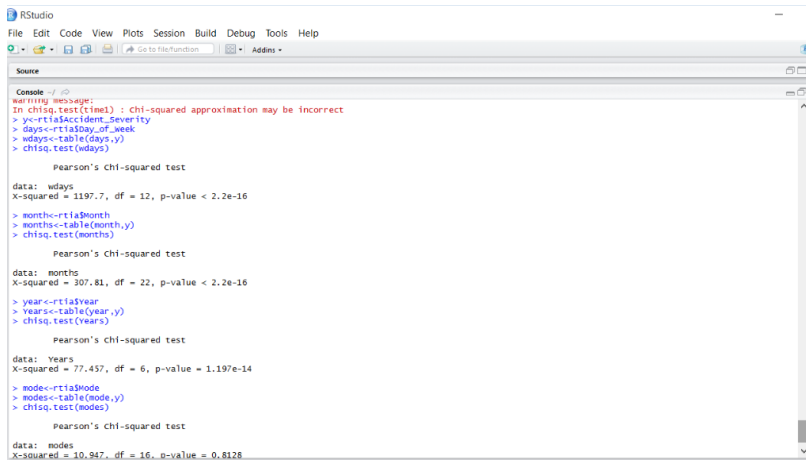


Figure: Spatial analysis

Findings With Code...



```
RStudio
File Edit Code View Plots Session Build Debug Tools Help
Go to file/function Addins

Source
Console
Warning message:
In chisq.test(time1) : Chi-squared approximation may be incorrect
> y<-rtia$Accident_Severity
> days<-rtia$Day_of_week
> wdays<-table(days,y)
> chisq.test(wdays)

Pearson's Chi-squared test

data: wdays
X-squared = 1197.7, df = 12, p-value < 2.2e-16

> month<-rtia$Month
> months<-table(month,y)
> chisq.test(months)

Pearson's Chi-squared test

data: months
X-squared = 307.81, df = 22, p-value < 2.2e-16

> year<-rtia$Year
> years<-table(year,y)
> chisq.test(years)

Pearson's Chi-squared test

data: years
X-squared = 77.457, df = 6, p-value = 1.197e-14

> mode<-rtia$Mode
> modes<-table(mode,y)
> chisq.test(modes)

Pearson's Chi-squared test

data: modes
X-squared = 10.947, df = 16, p-value = 0.8128
```

Figure: Non Spatial analysis

Findings With Code...

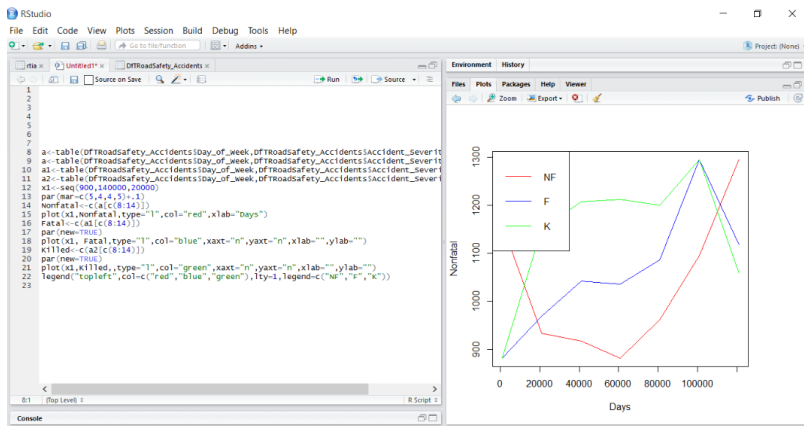


Figure: Non Spatial analysis — Chi Square test

Thank
You