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 Choosen :
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 | rtia | year | Untitled1* | Untitled2*

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_Roa
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	

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

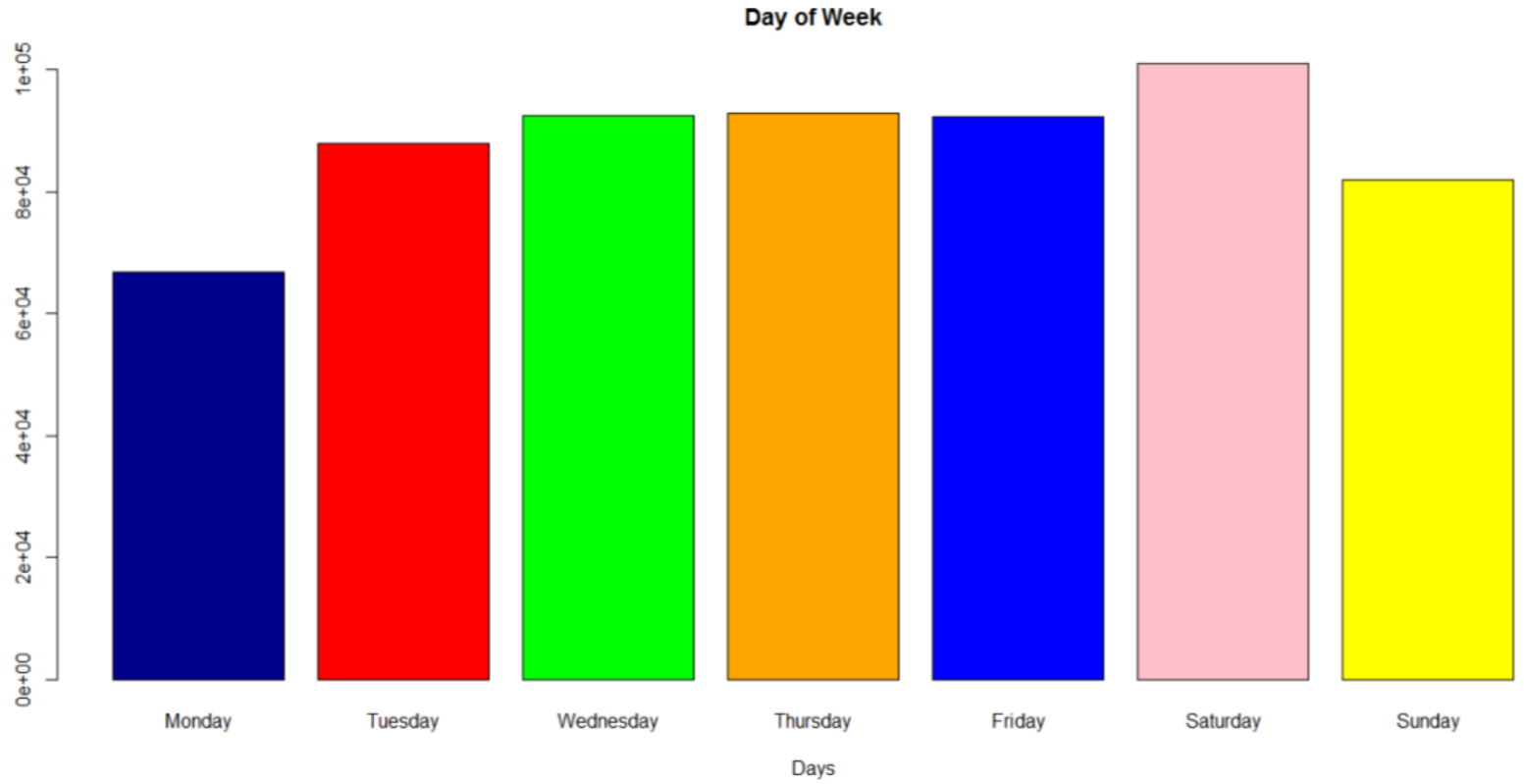


Figure: Days Of Week

Findings With Code...

Figure: Days Of Week Code

Findings With Code...

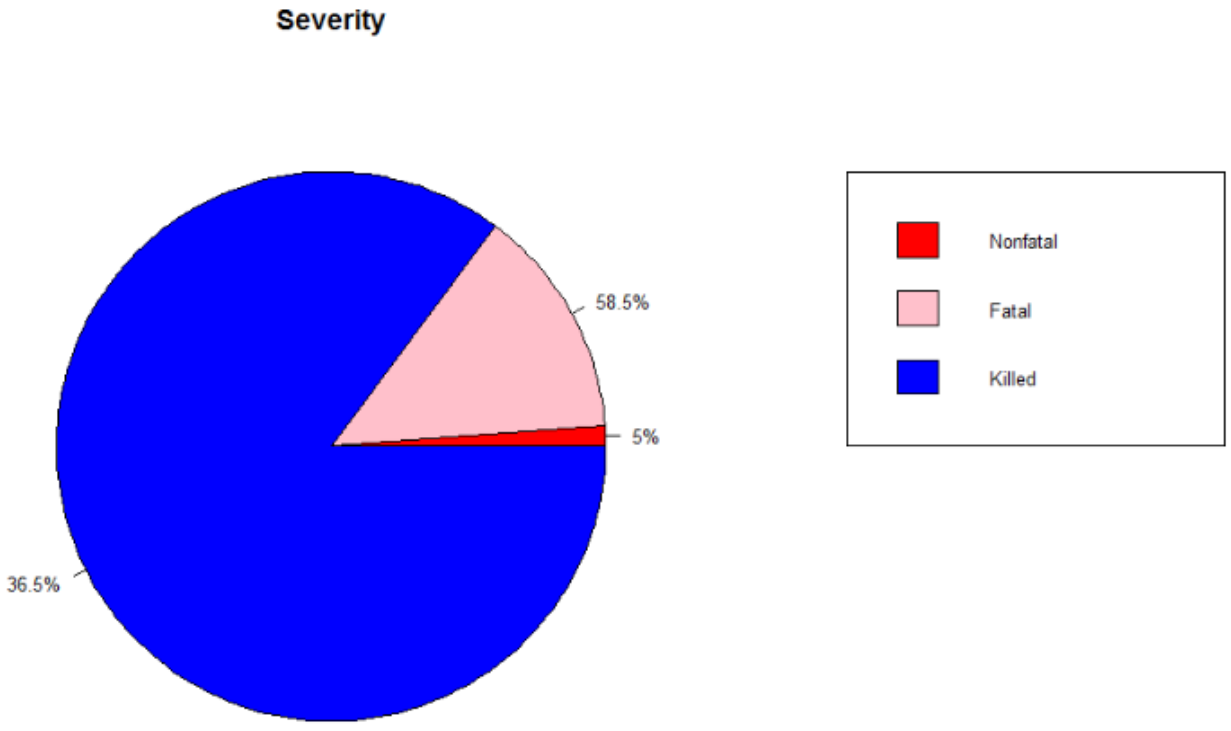
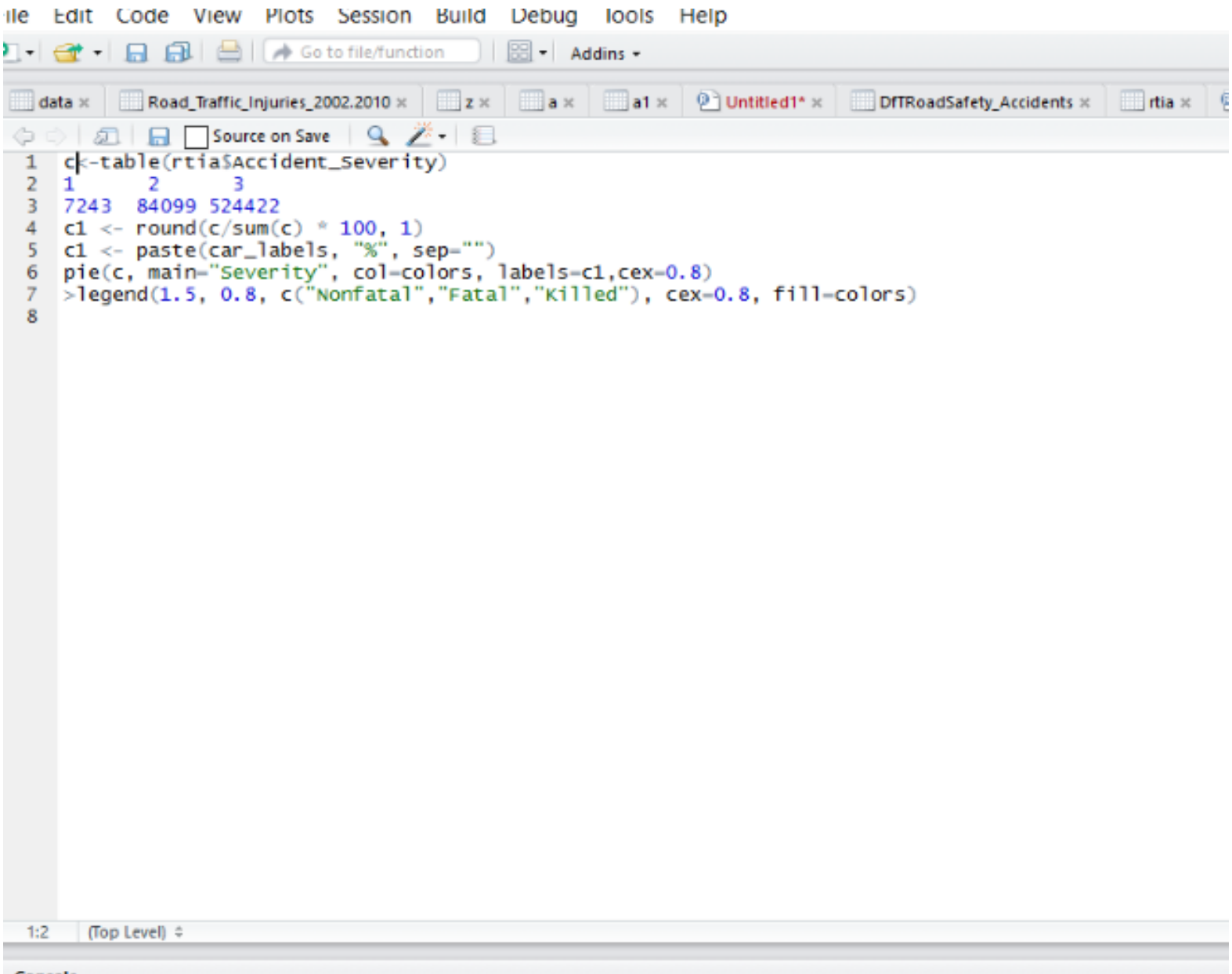


Figure: Severity

Findings With Code...



Findings With Code...

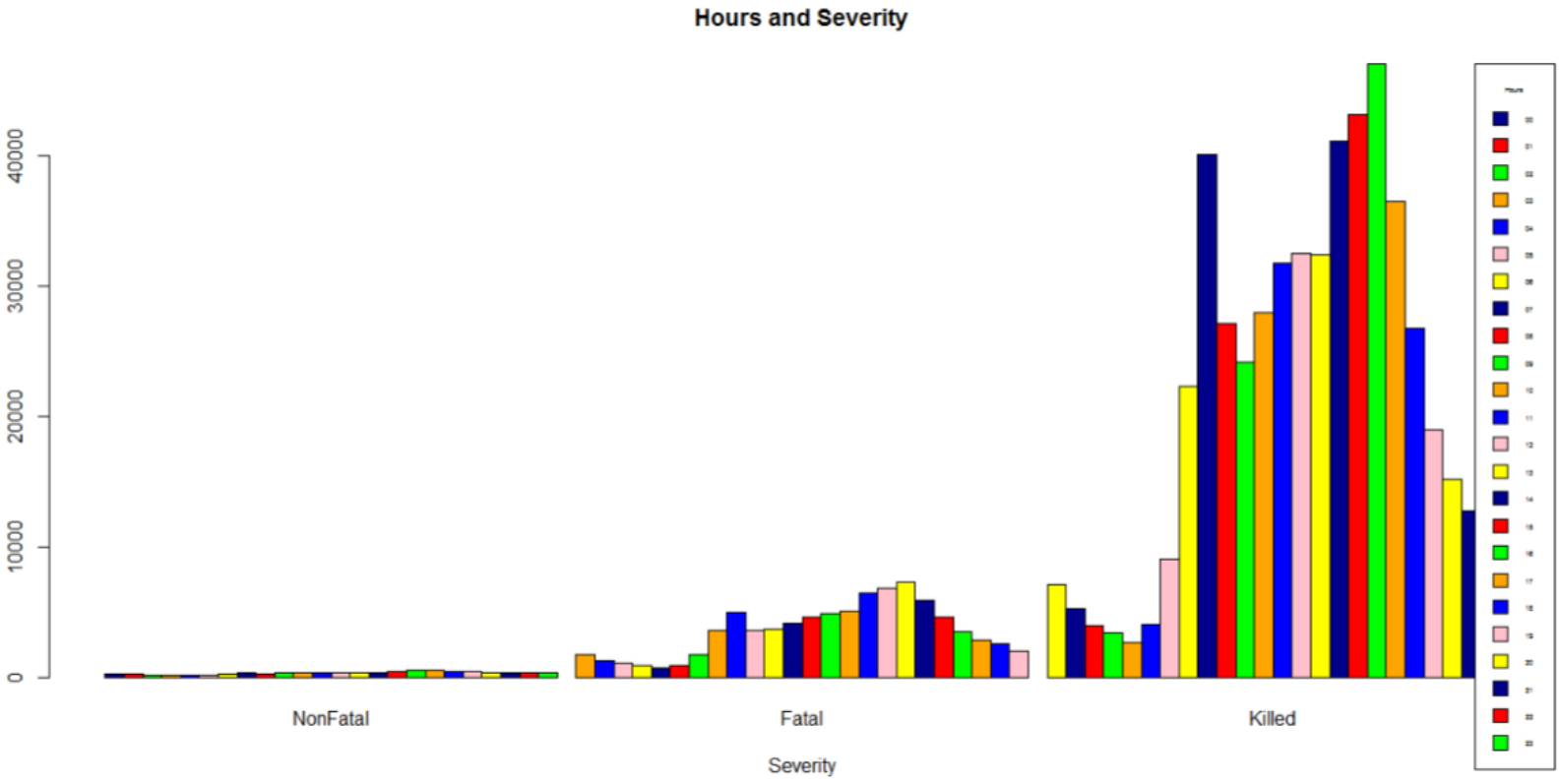


Figure: Hours and Severity

Findings With Code...

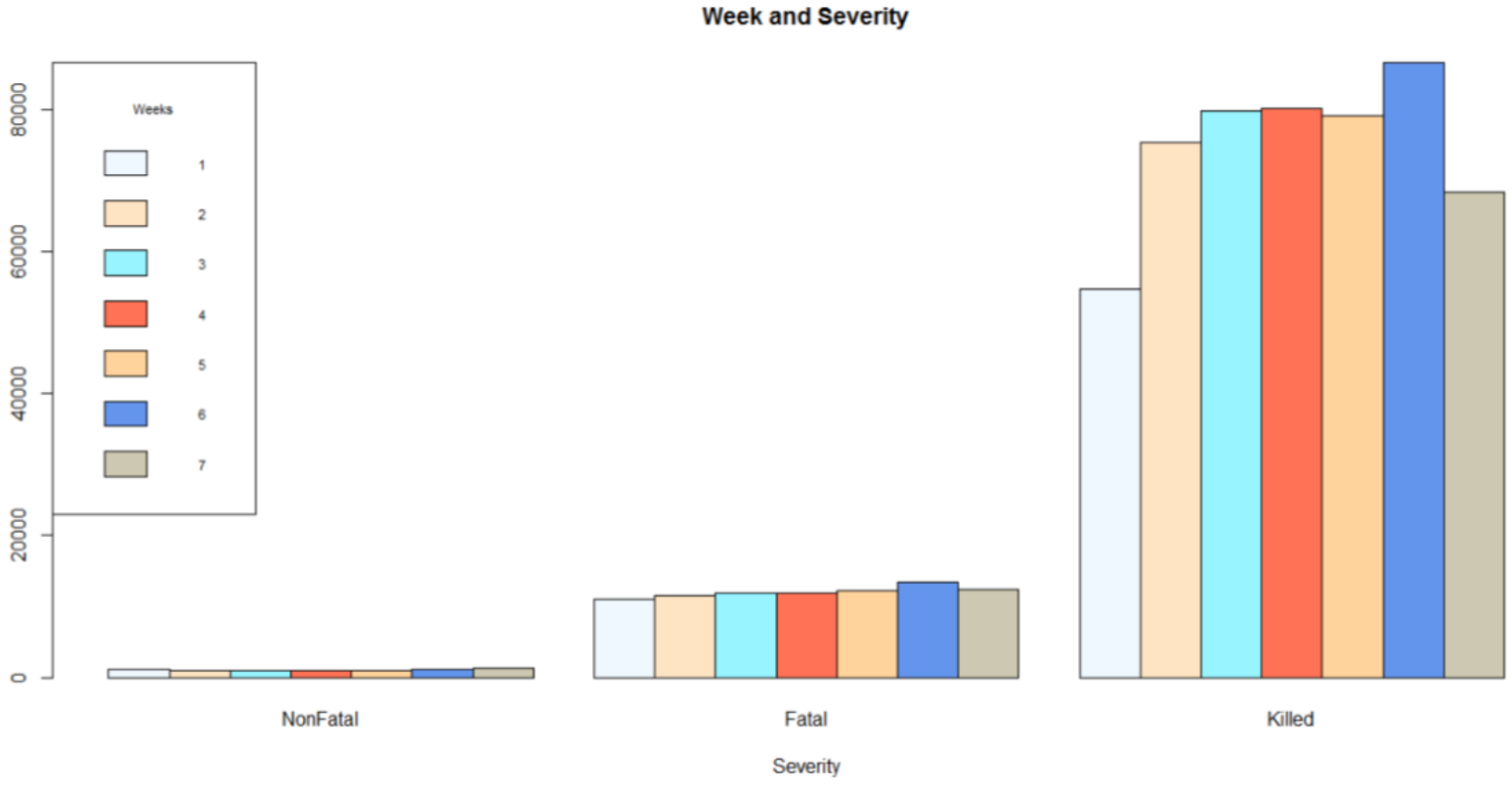


Figure: Weeks and Severity

Findings With Code...

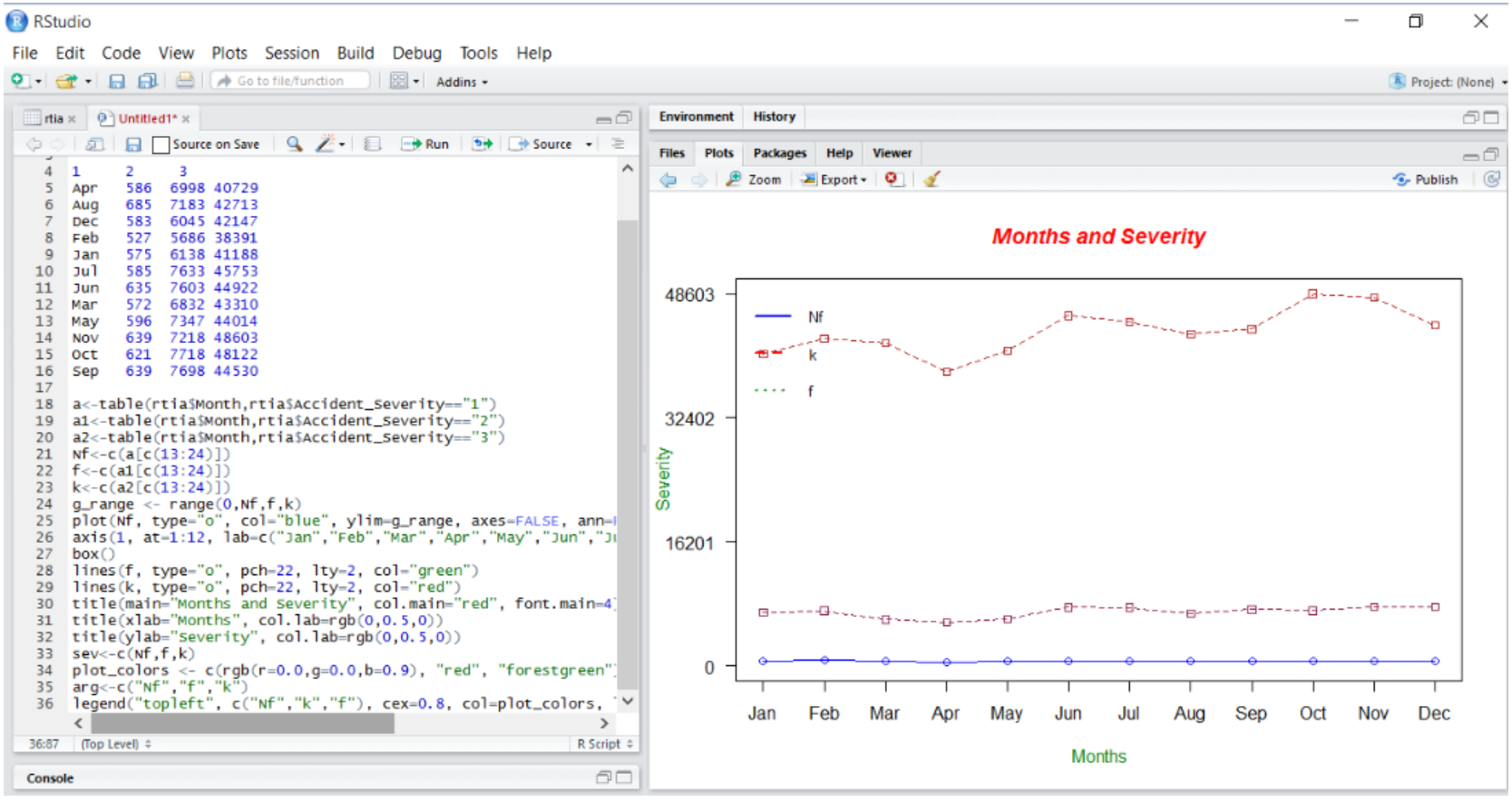


Figure: Month and Severity

Findings With Code...

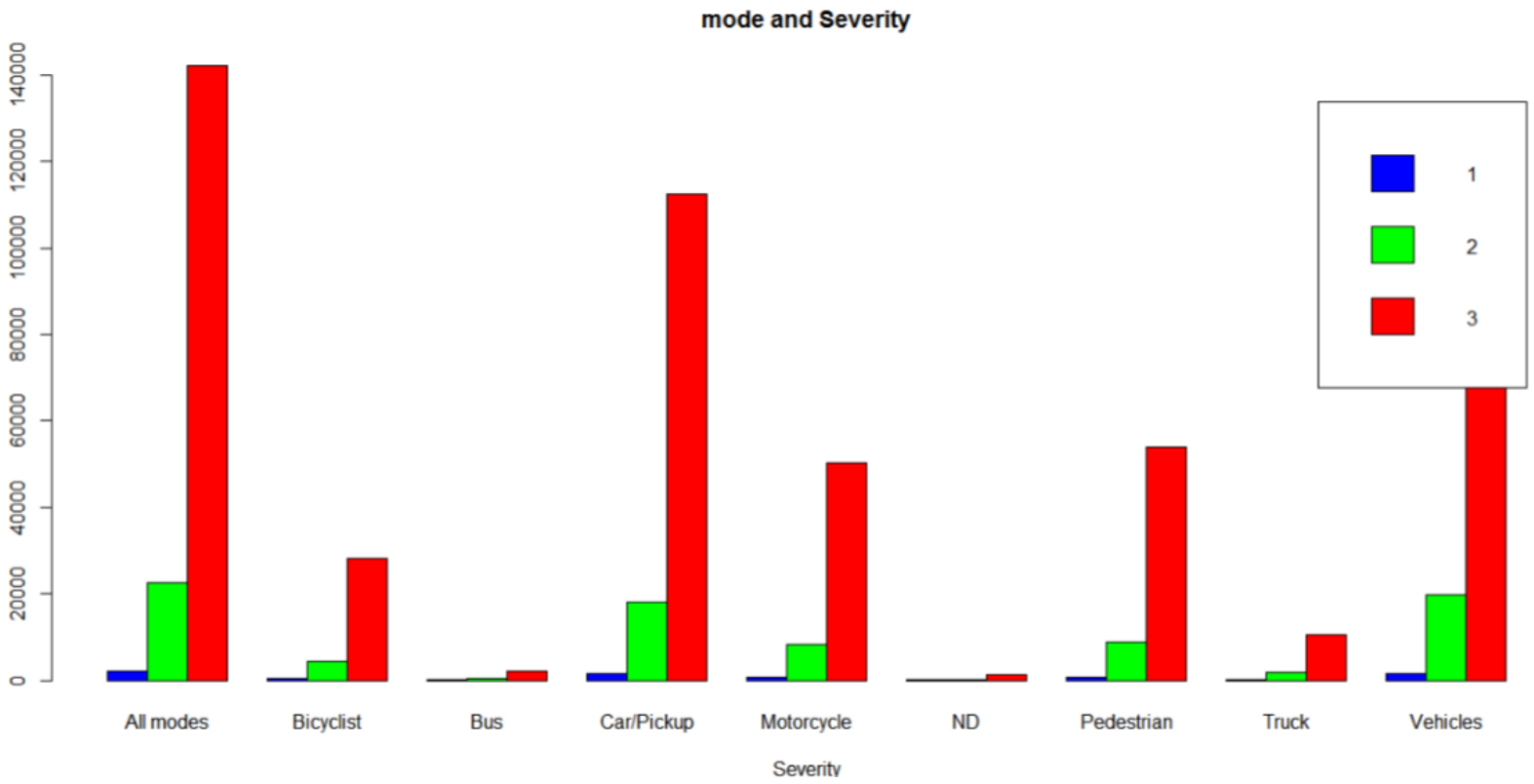


Figure: Modes and Severity

Findings With Code...

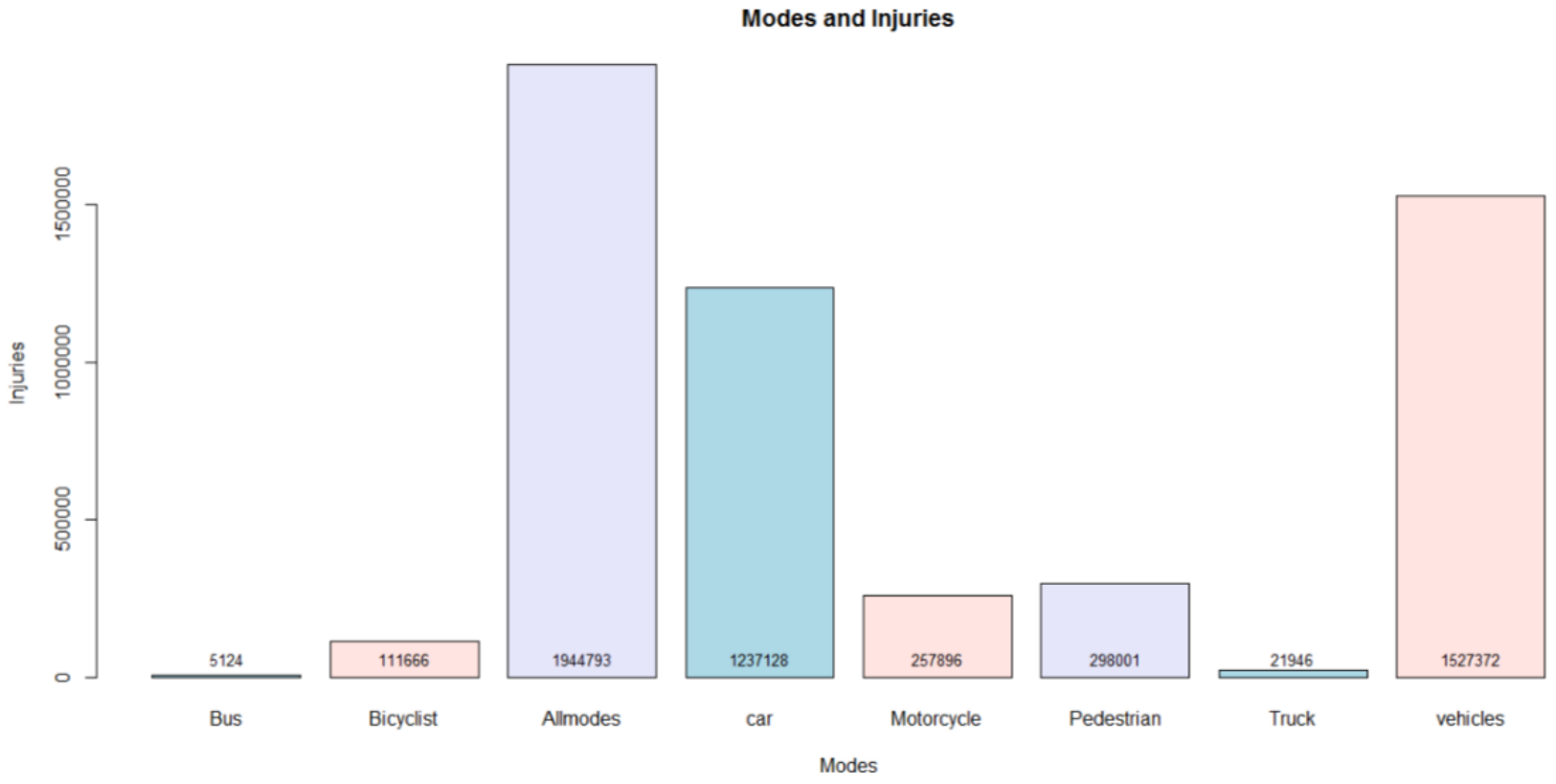


Figure: Modes and Injuries

Findings With Code...

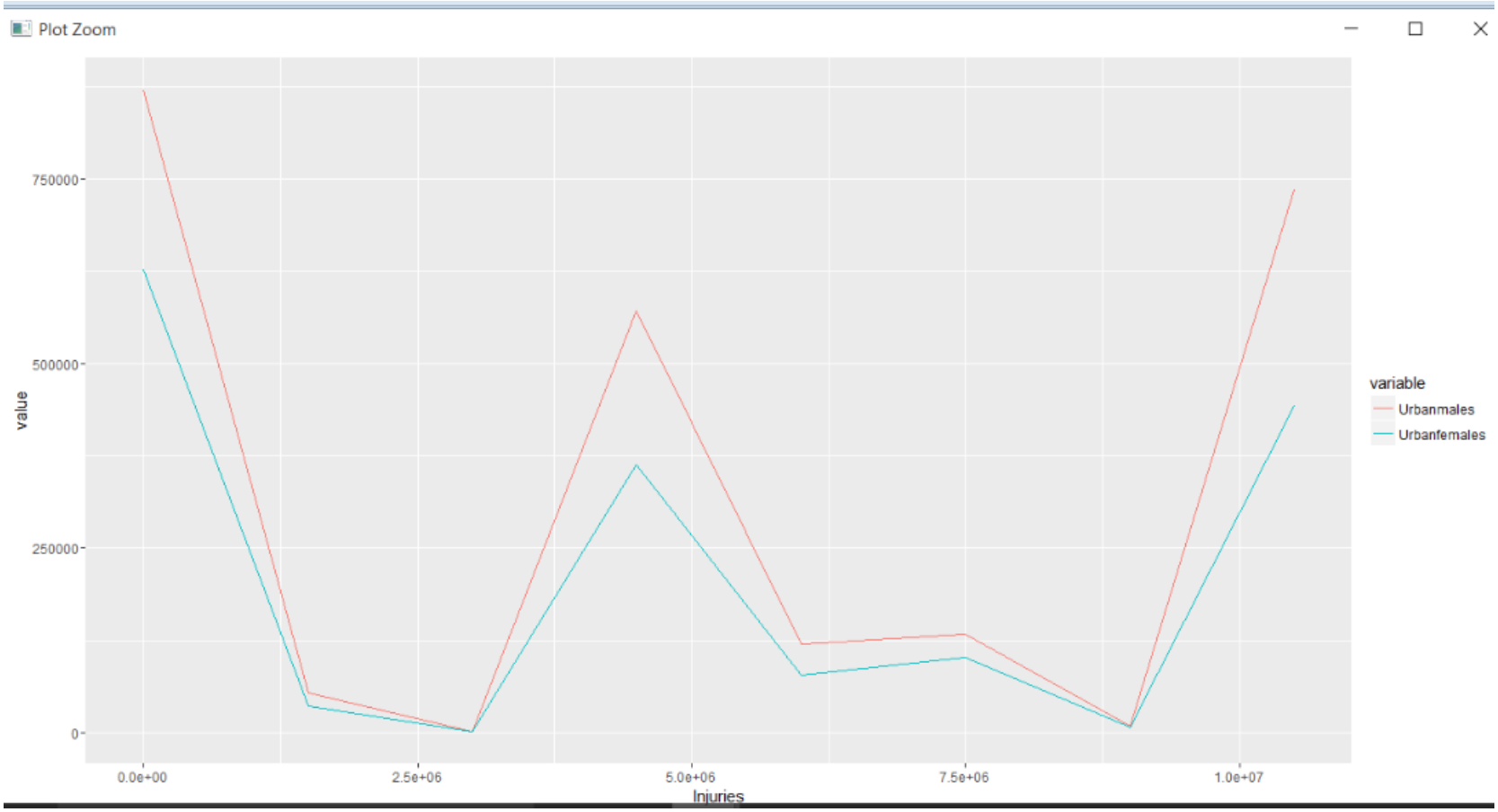


Figure: Urban

Findings With Code...

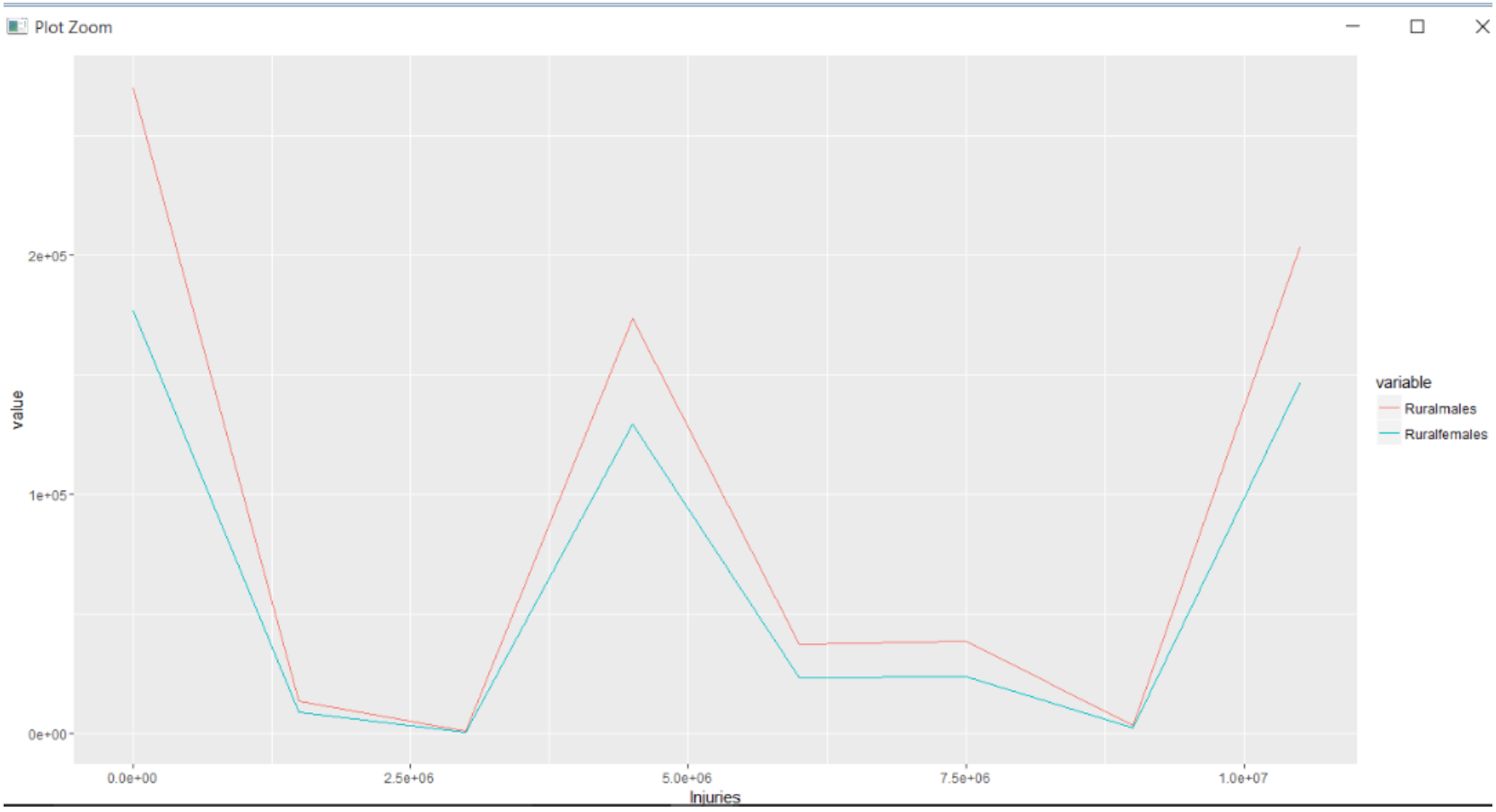
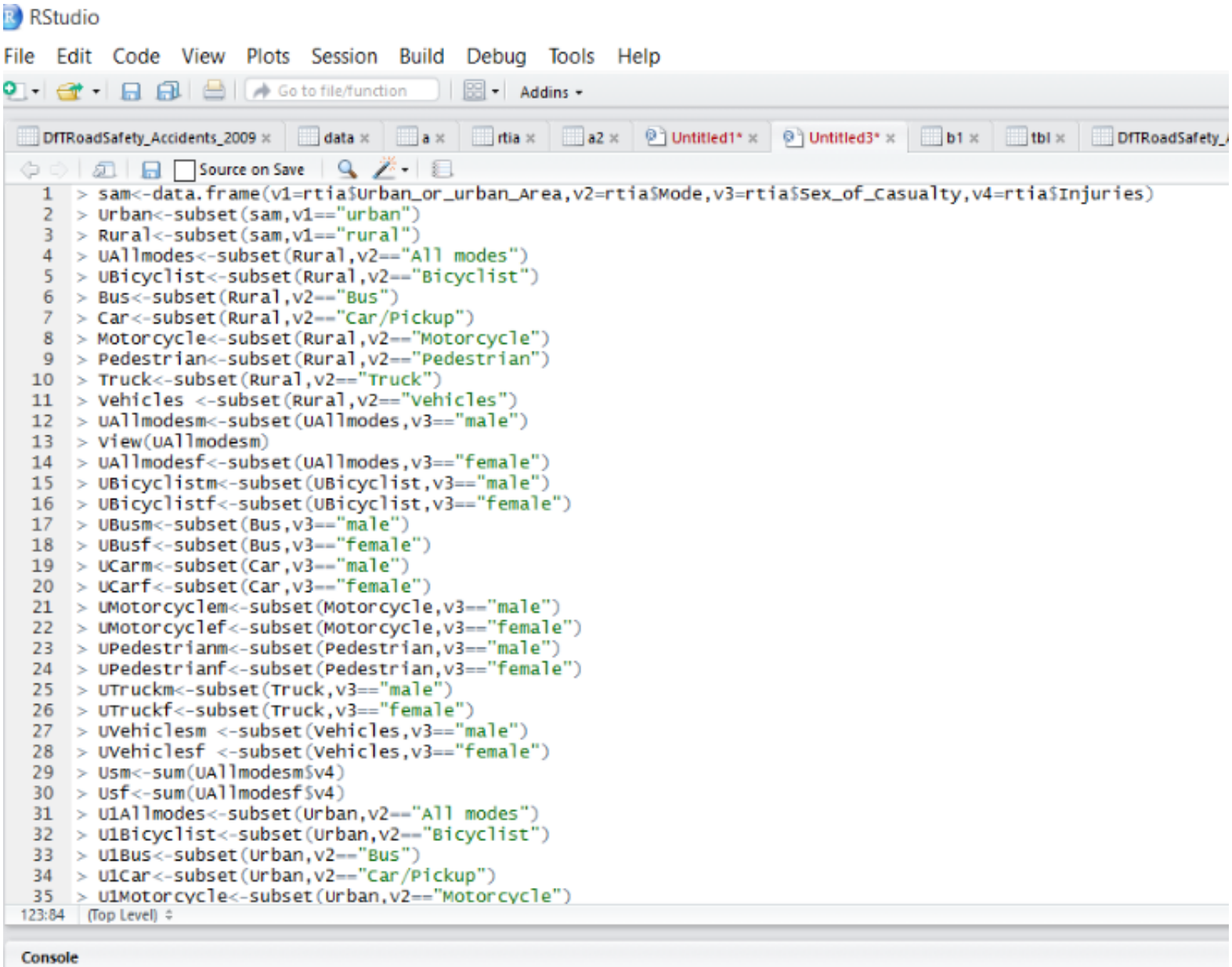


Figure: Rural

Findings With Code...



Findings With Code...

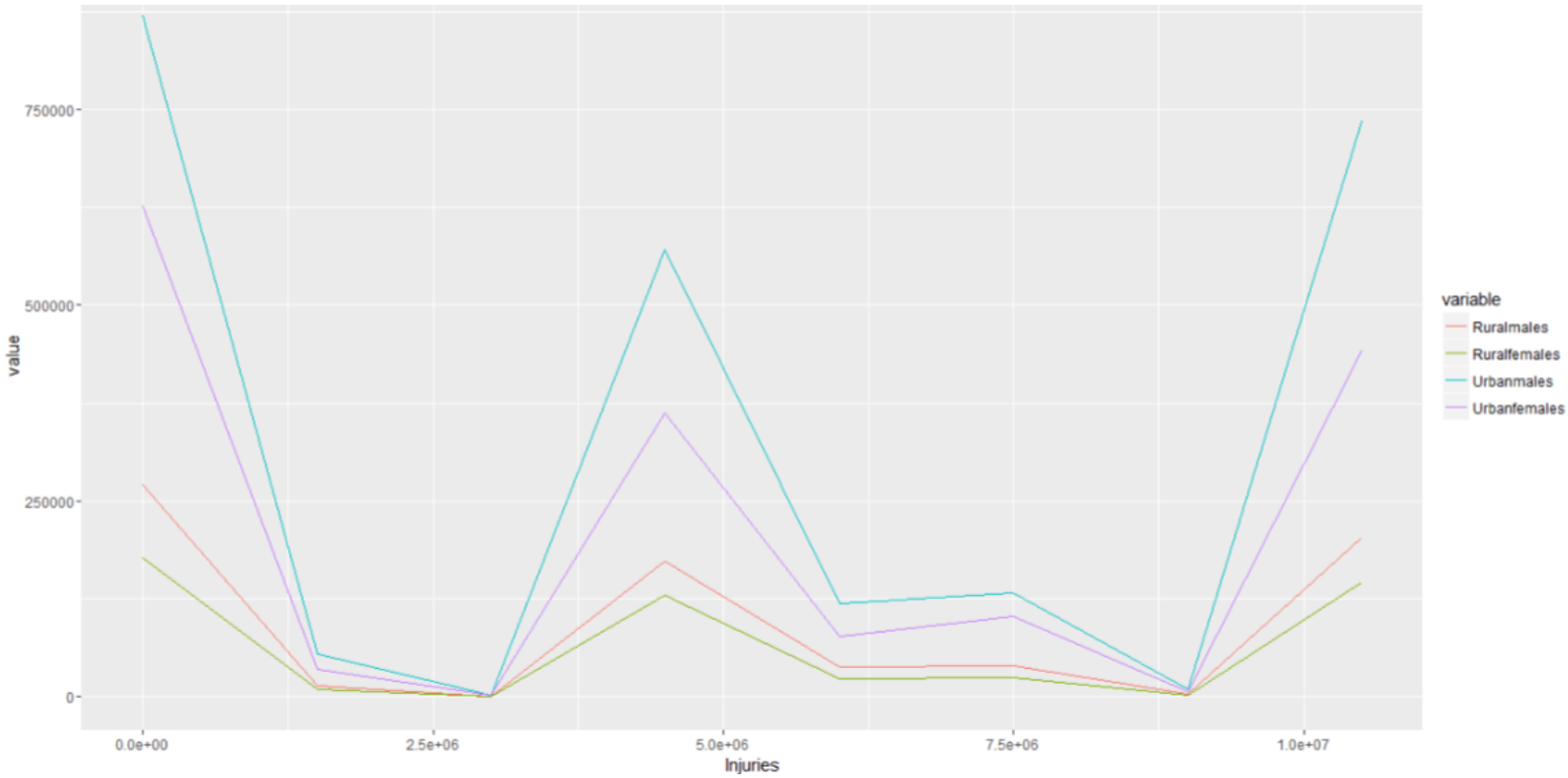


Figure: Ages,Gender And Injuries

Findings With Code...

Figure: correlation

Findings With Code...

```
b<-table(rtia$Weather_Conditions,rtia$Accident_Severity)
> b

1      2      3
-1      3      7     96
1    6016  69813 419752
2     686   8472  60487
3      43    537   5256
4     117   1100   6157
5     112   1072   6801
6       4     84    670
7      61    456   2635
8     123   1534  12487
9      78   1024  10081
> scatter.smooth(b)
> scatter.smooth(b,xlab="weather",ylab = "Severity")
cor.test(rtia$Weather_Conditions,rtia$Accident_Severity)

Pearson's product-moment correlation

data:  rtia$Weather_Conditions and rtia$Accident_Severity
t = 18.45, df = 615760, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
0.02100967 0.02600232
sample estimates:
cor
0.02350614

>
```

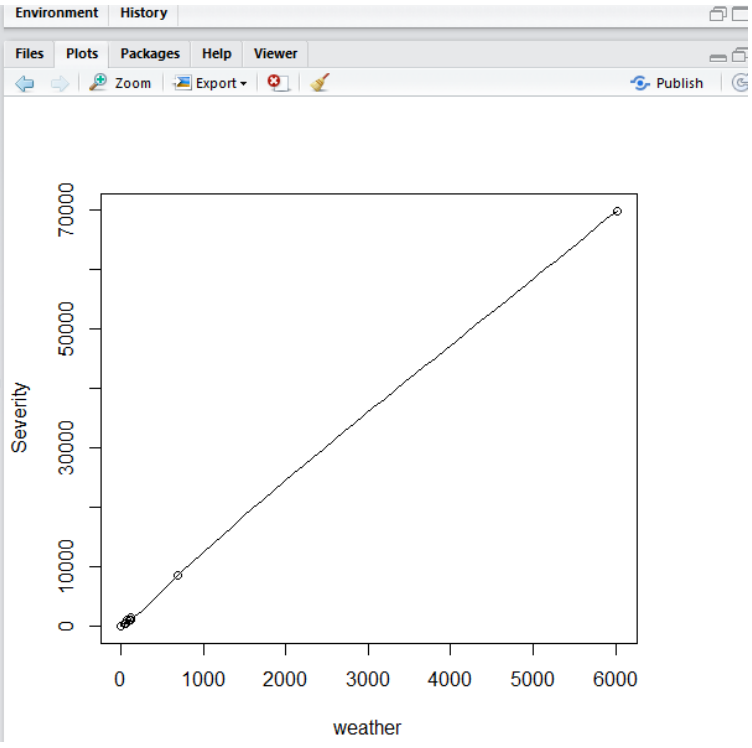


Figure: correlation

Analysis

Analysis is classified into 6 types :

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

Analysis

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...

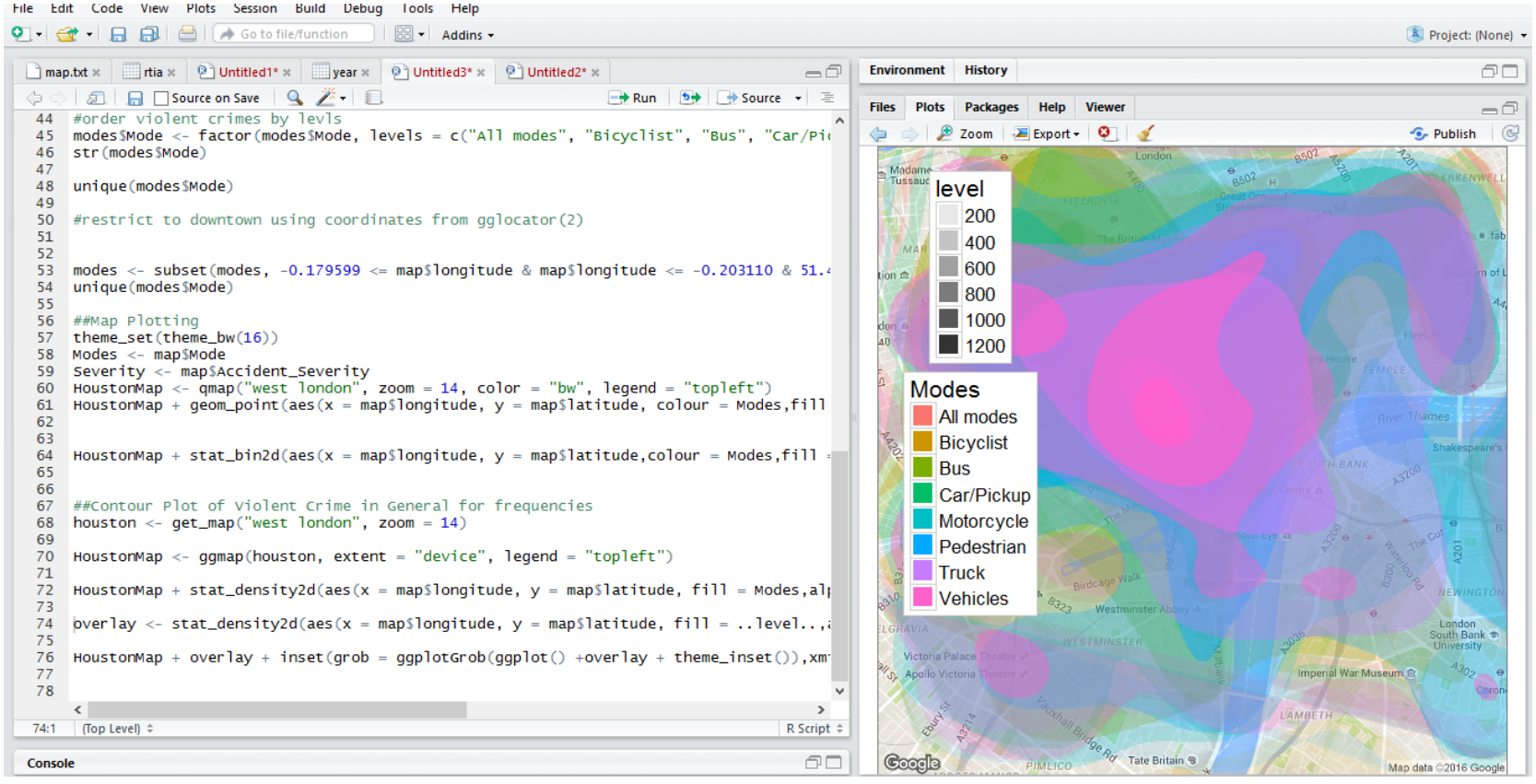


Figure: Spatial analysis

Findings With Code...

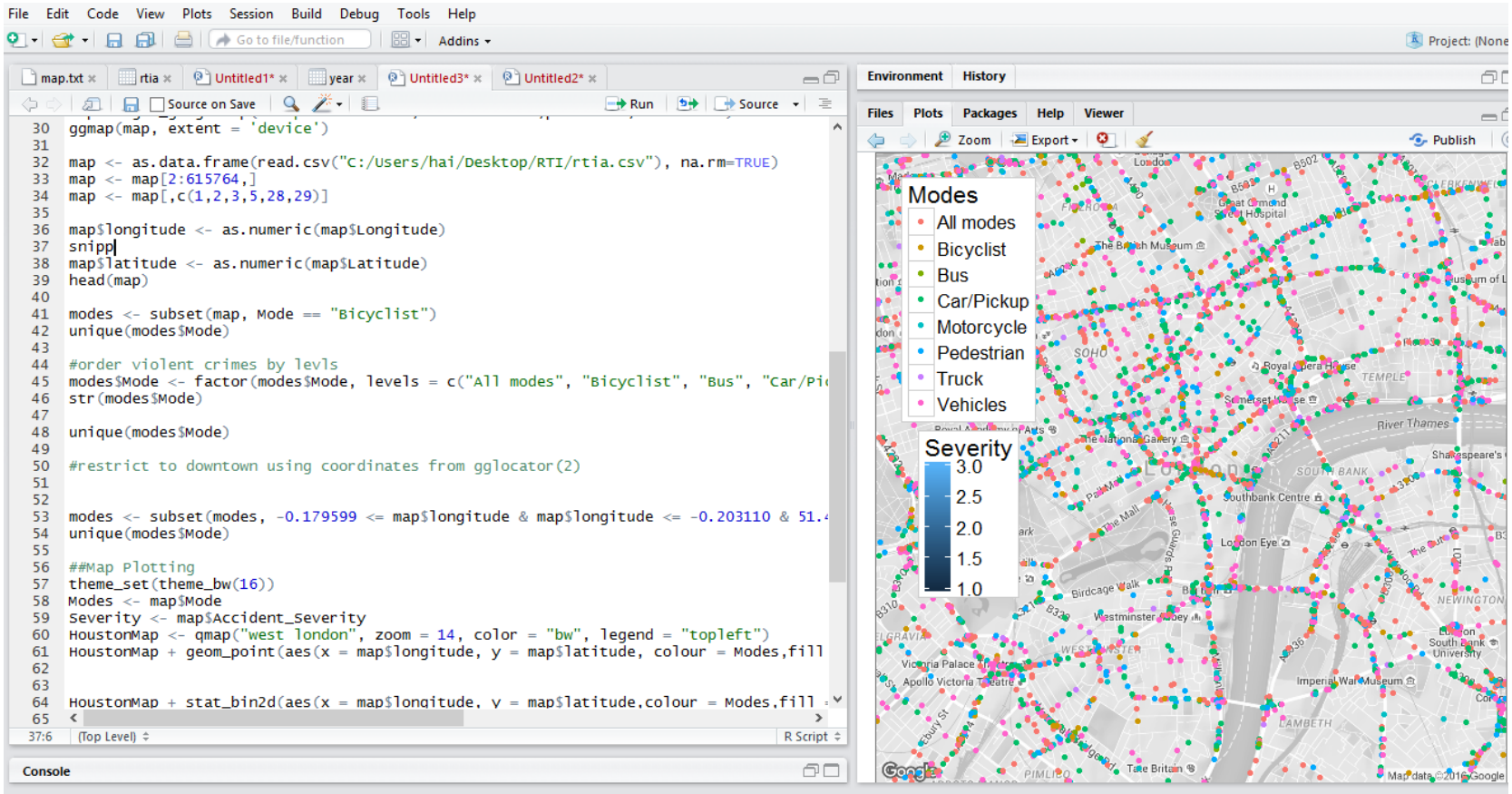


Figure: Spatial analysis

Findings With Code...

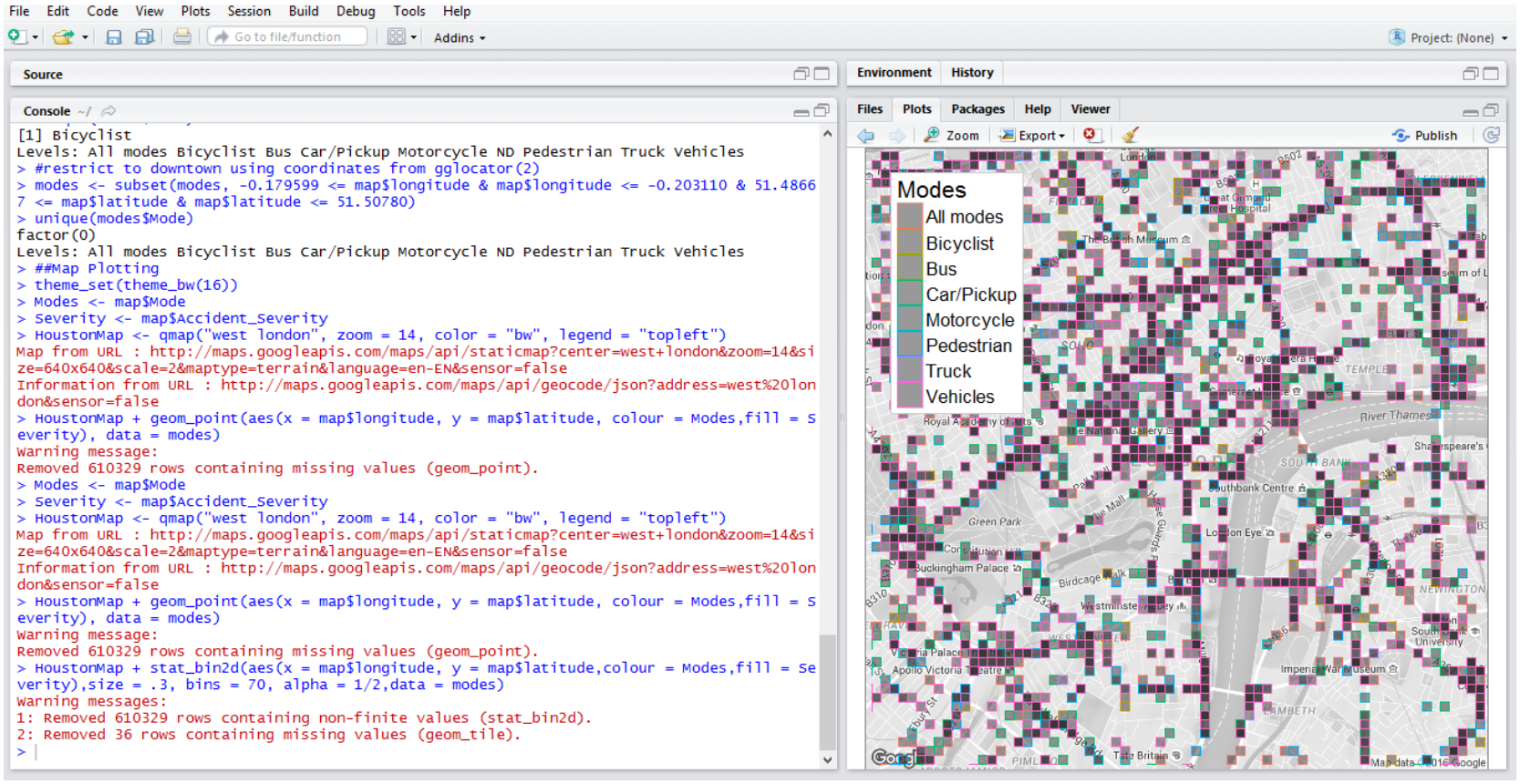


Figure: Spatial analysis

Findings With Code...

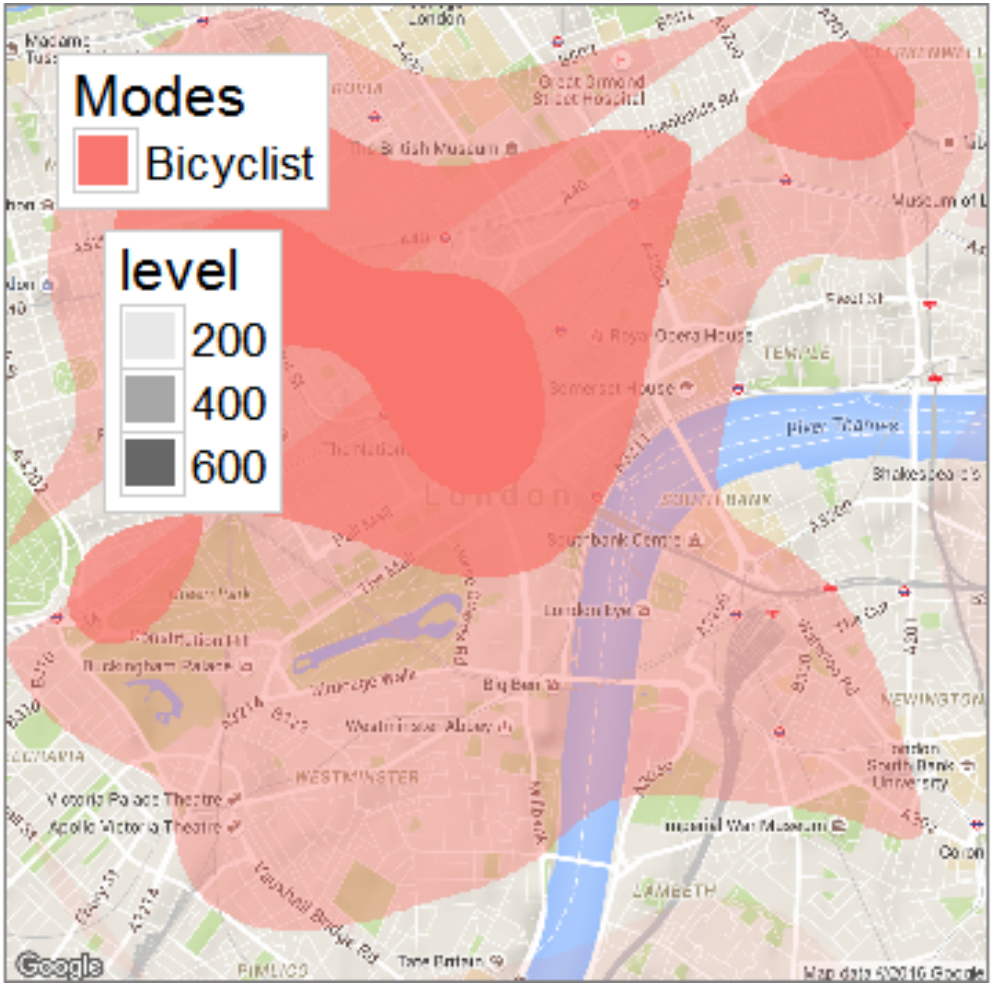


Figure: Spatial analysis

Findings With Code...

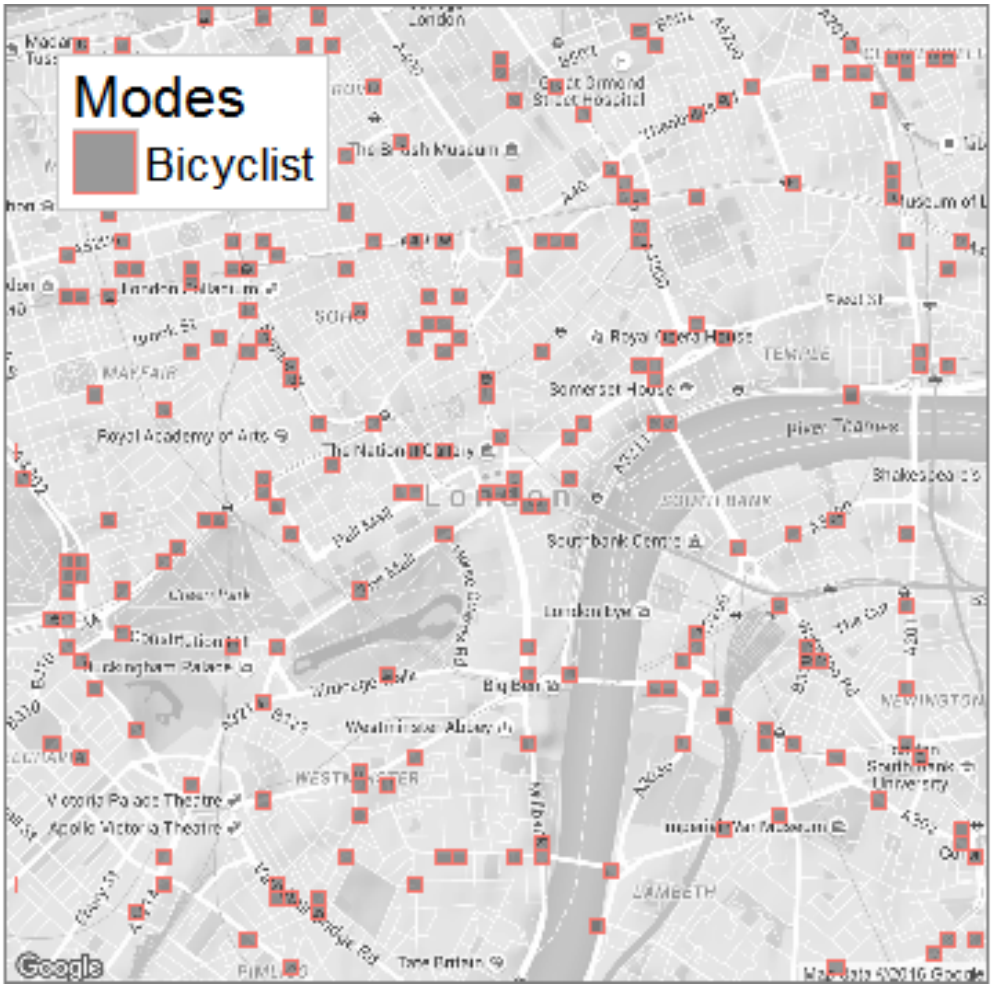


Figure: Spatial analysis

Findings With Code...

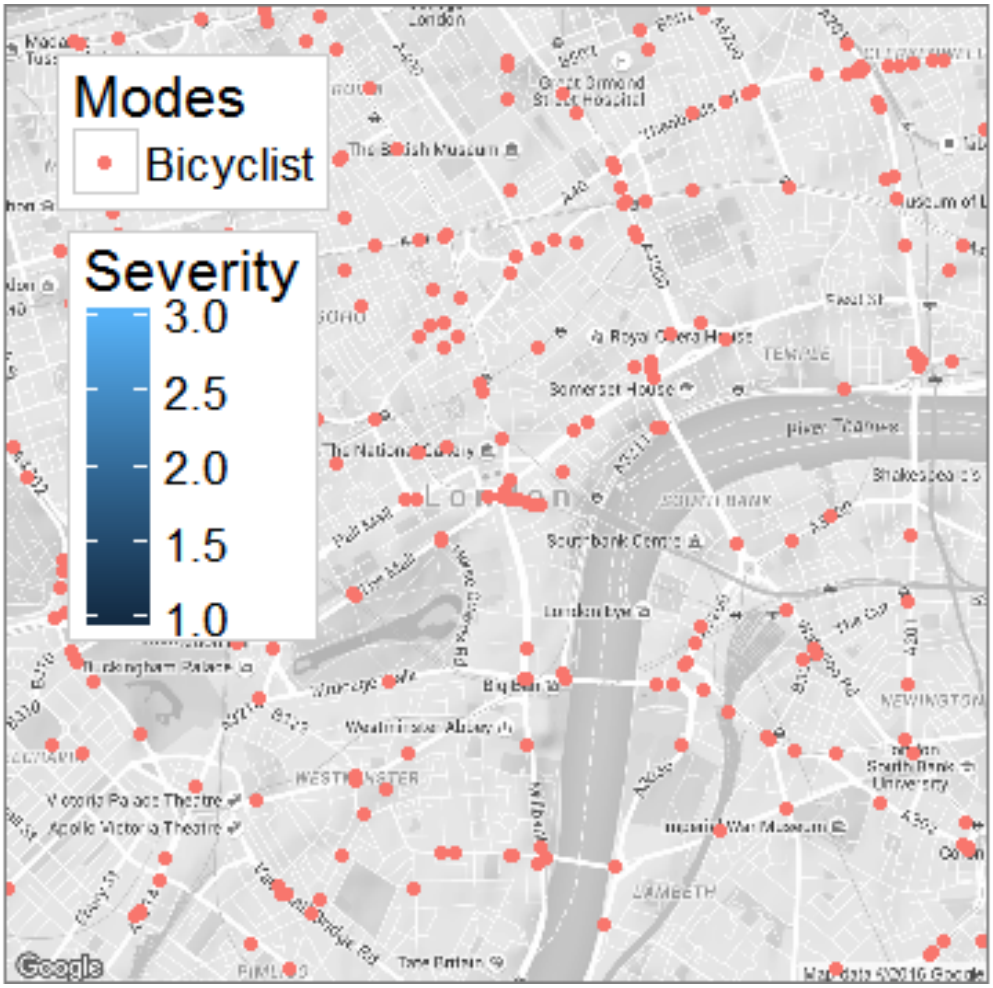


Figure: Spatial analysis

Findings With Code...

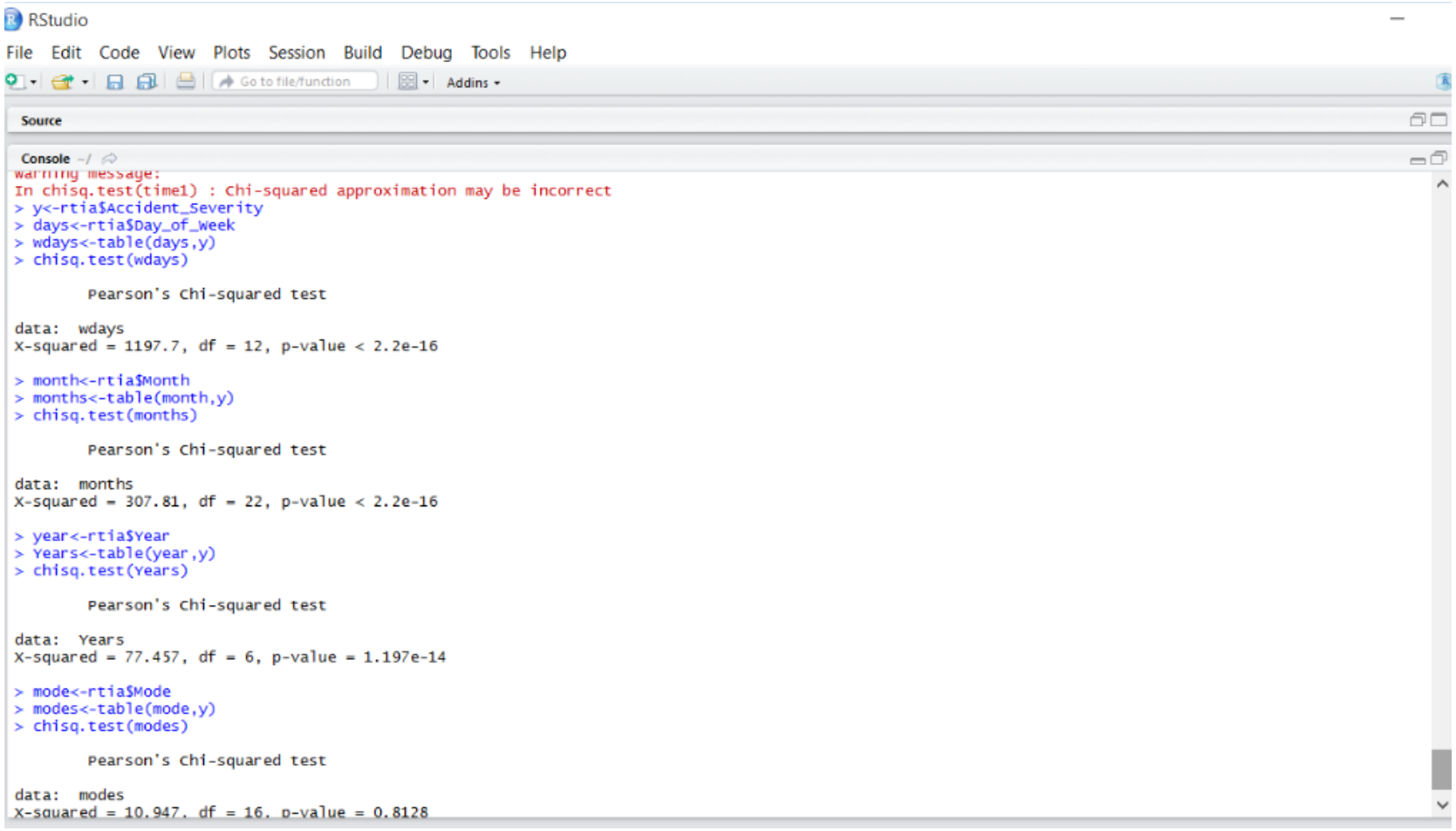


Figure: Non Spatial analysis

Findings With Code...

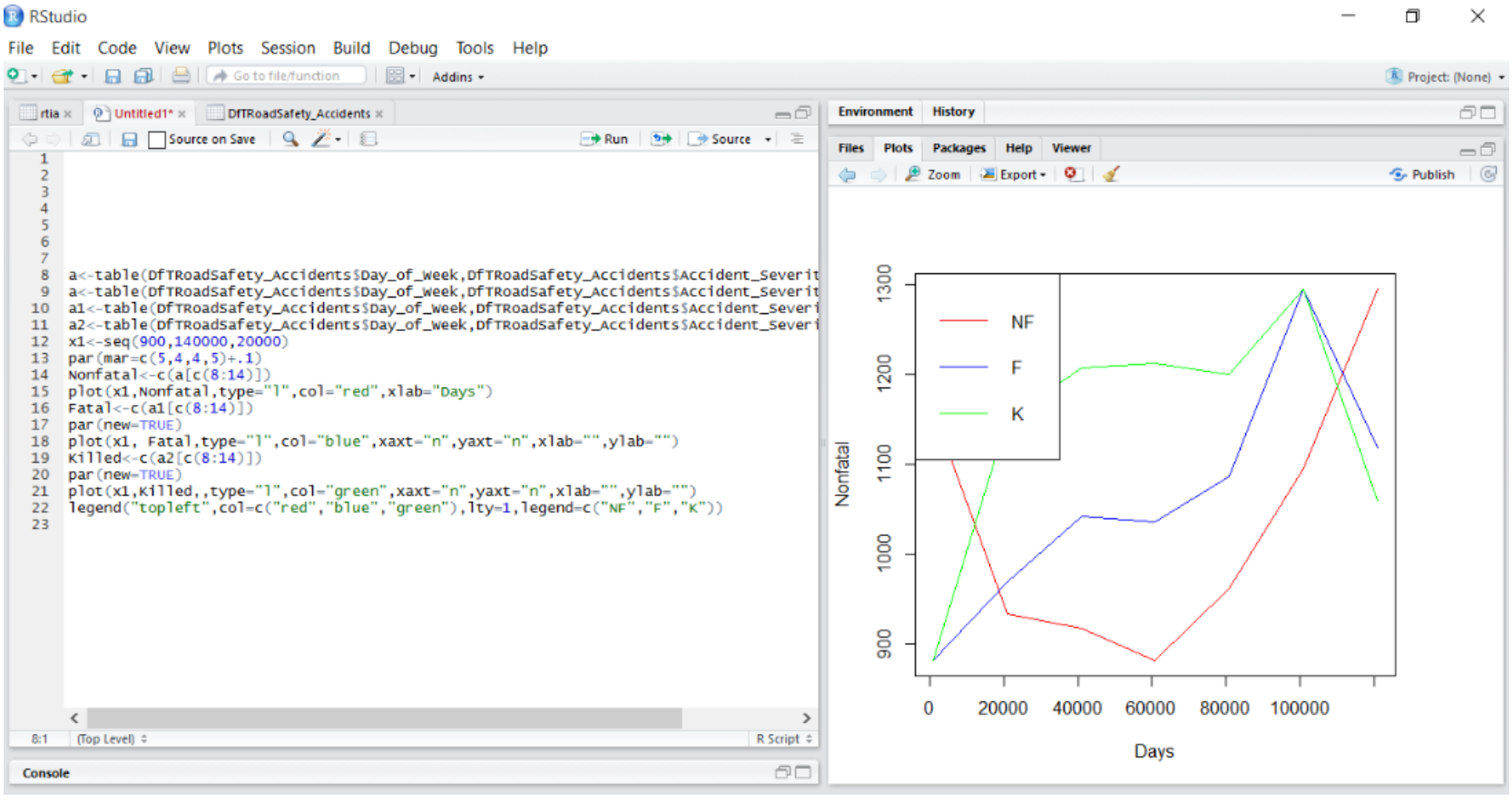


Figure: Non Spatial analysis — Chi Square test

Thank You