Assignment no 2

Advance topics in information systems

Jayesh Parab 27148572

Executive Summary

This report is aimed at reducing the number of accidents taking place in Victoria by studying the data of the last 5 years. After analyzing the dataset I found that a number of accidents occurred due to poor or no street lights. Such conditions which increase the chances of accidents can be studied and such issues can be solved. The process I followed to minimize the number of accidents was to find the top 5 cities in which accidents occur and then provide demographics of the cities and using this analyzed information suggest government to start street light or implement new street lights in places more prone to accidents.

Contents

Introduction

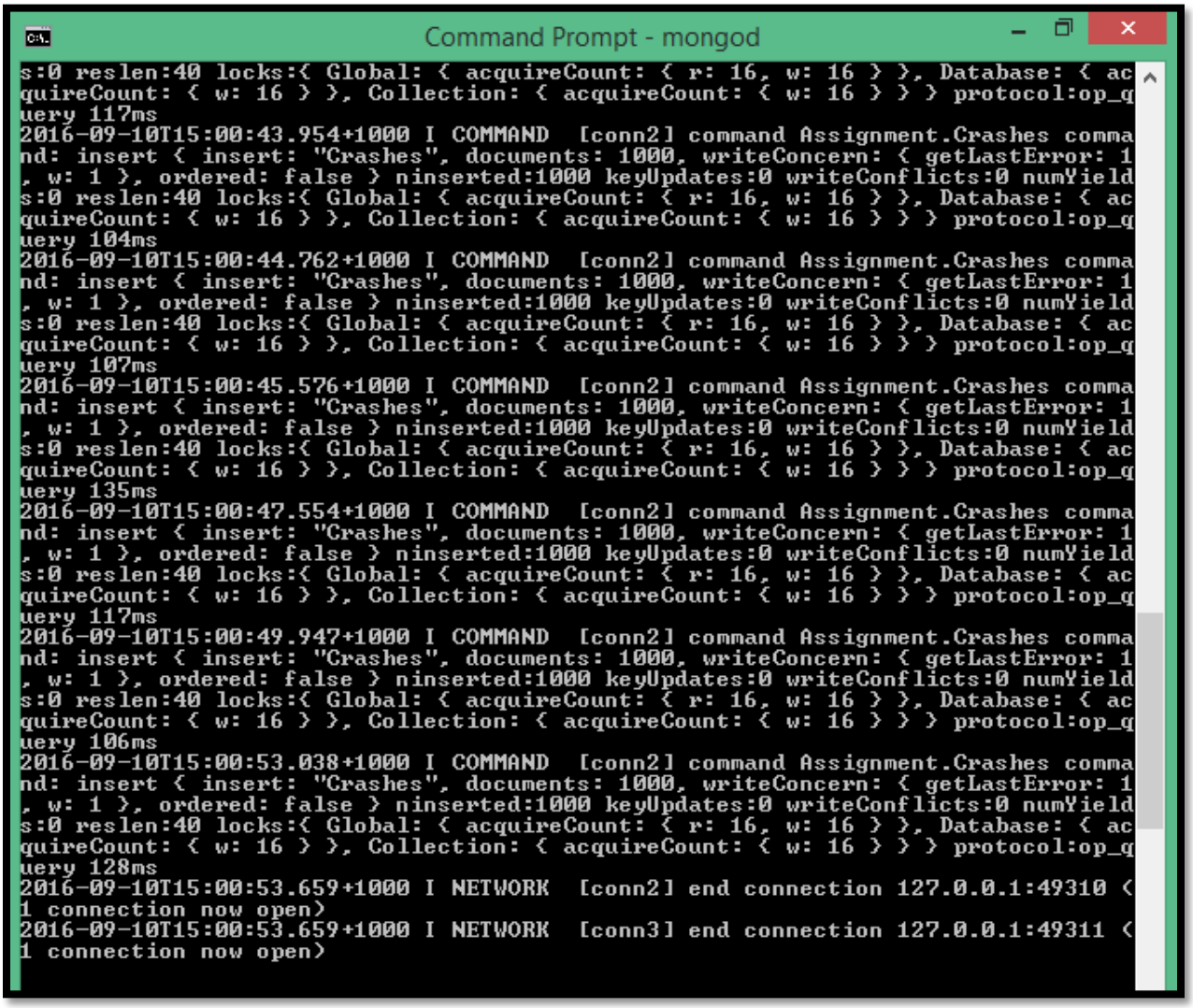
The dataset that we are using is about accidents occurring in last 5 years in Victoria. The aim is to study Mongod, Tableau, R Studio and R language to interpret and visualize the selected dataset and provide solution in order to reduce the number of accidents from occurring.

Business Problem

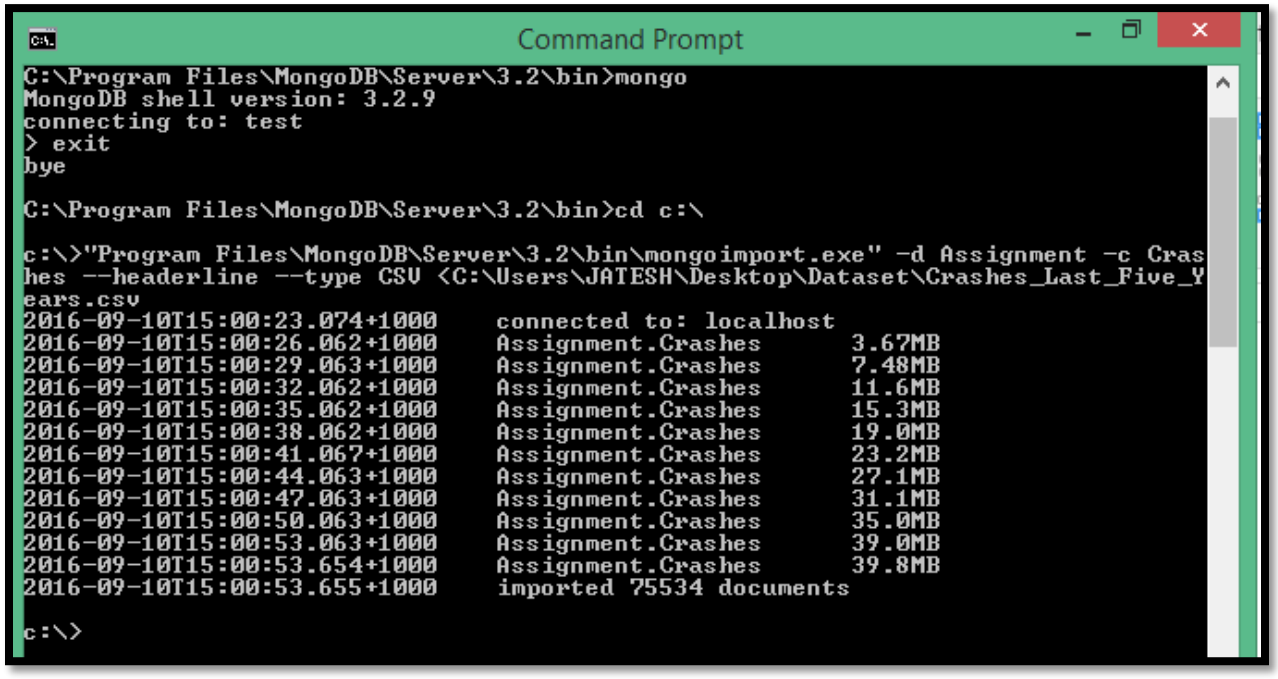
The selected dataset was analyzed and after analyzing, I found that a number of accidents occur due to poor light conditions which affect visibility of drivers. Therefore the business problem for this report is to reduce the number of accidents occurring due to no proper lights offered

Loading Data into Mongod

Starting Mongodb by mongod command:

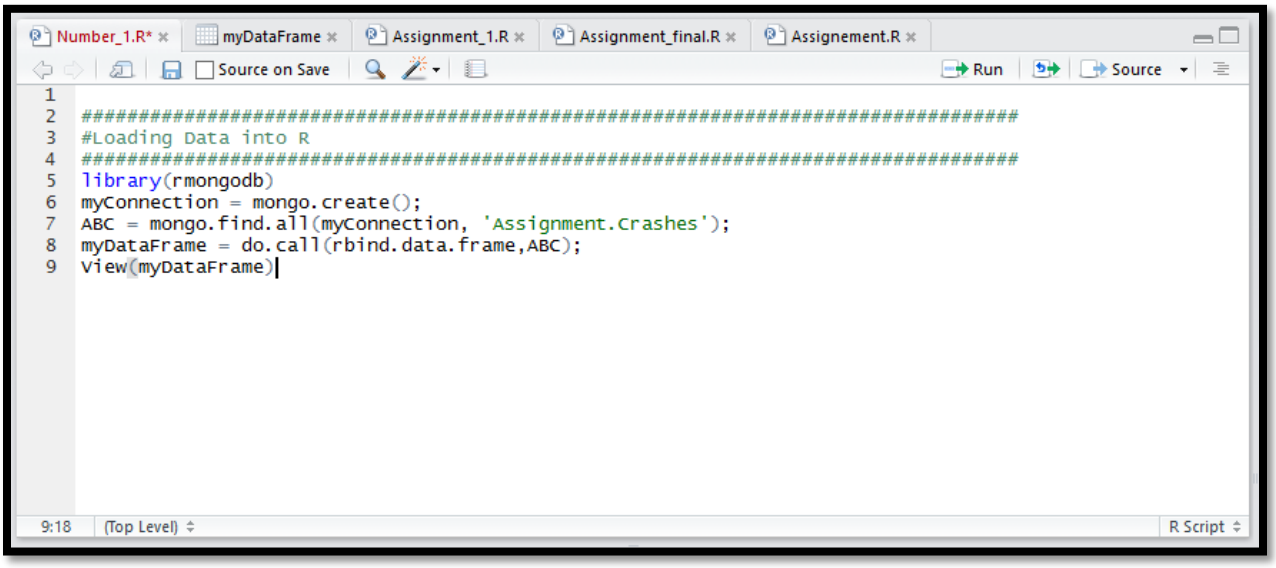


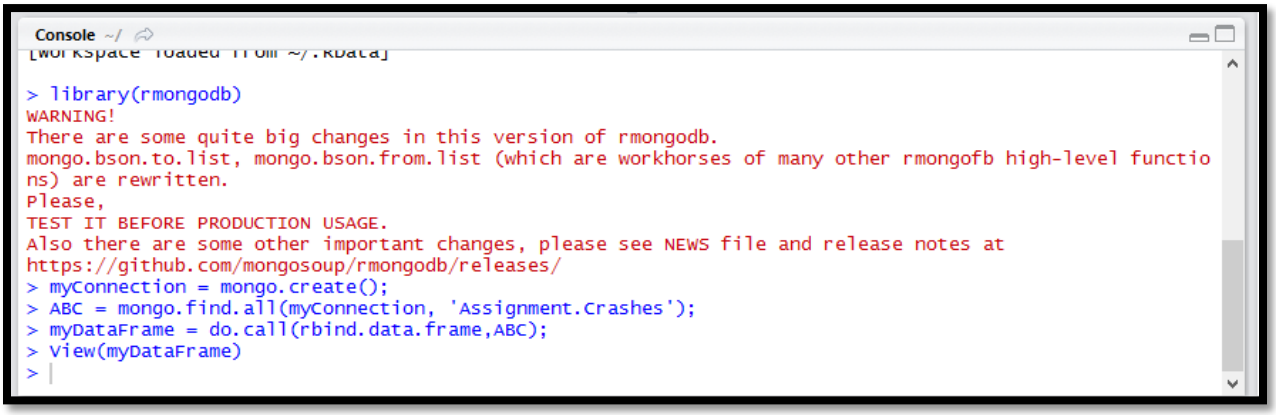
Typing Mongo command in another command prompt and importing CSV file in mongod:

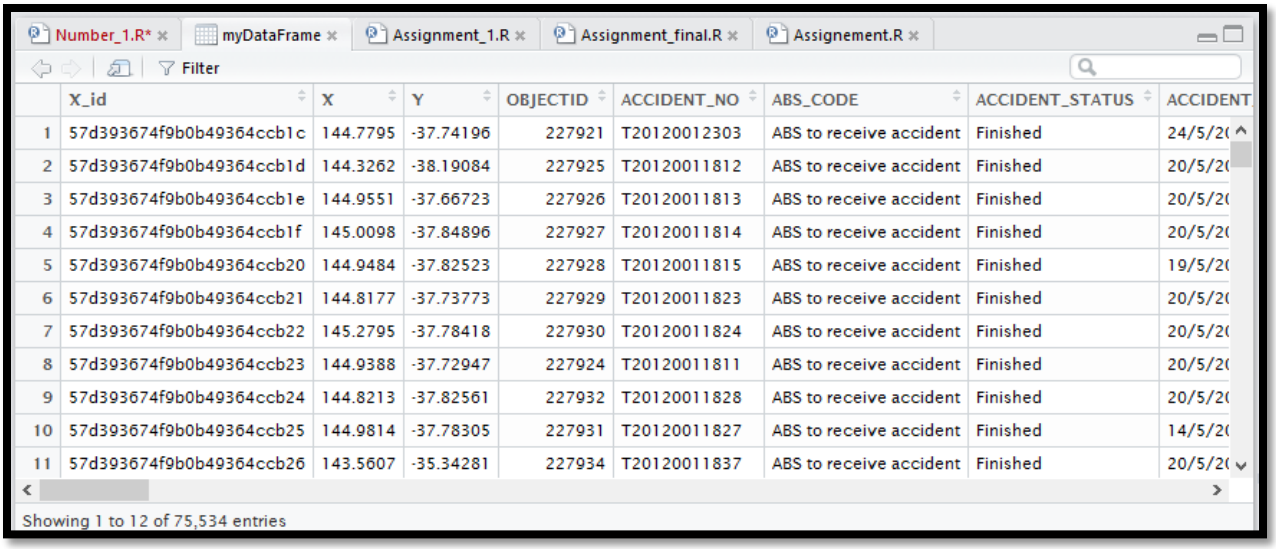


Using the dataset where required file is uploaded

Loading data from mongod to R







Number_1.R* | myDataFrame | Assignment_1.R | Assignment_final.R | Assignement.R

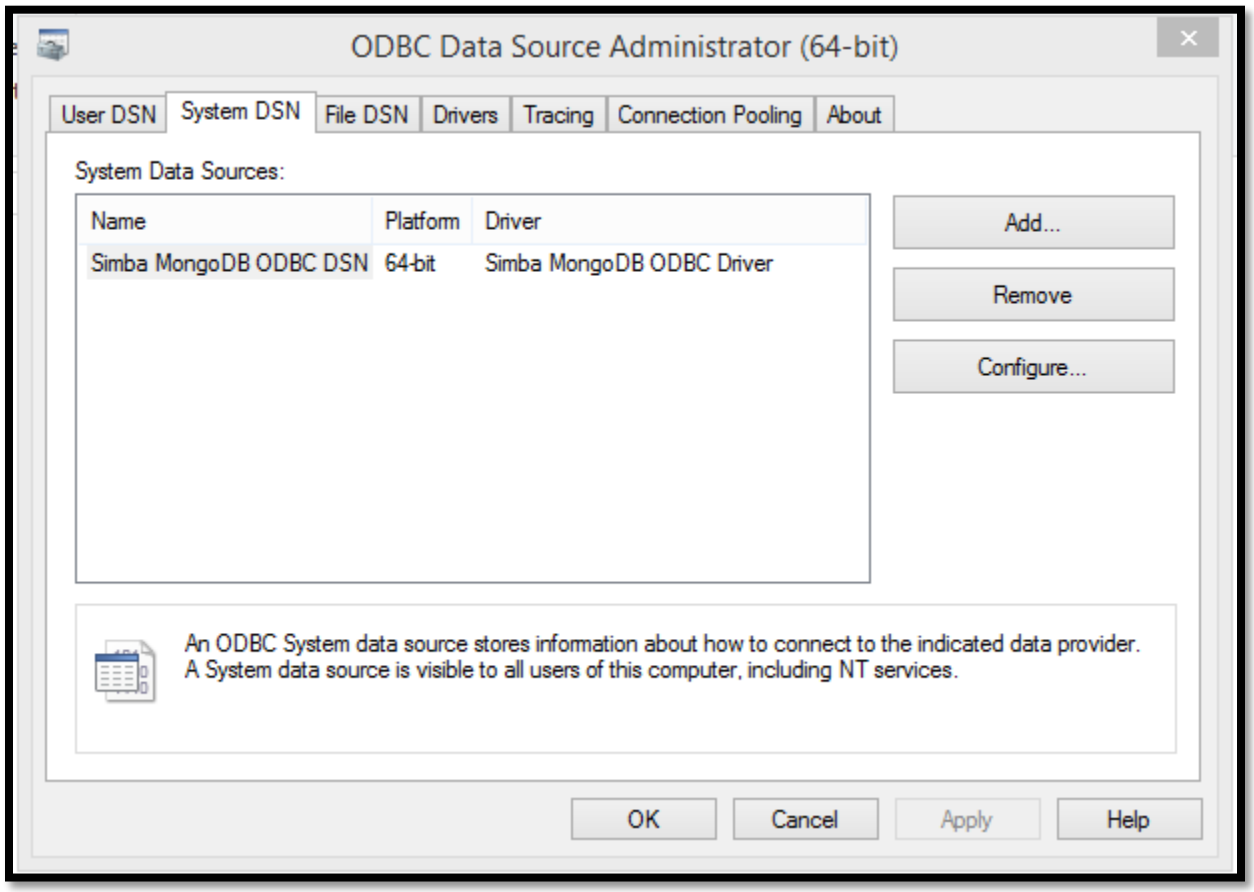
Filter

	X_id	X	Y	OBJECTID	ACCIDENT_NO	ABS_CODE	ACCIDENT_STATUS	ACCIDENT
1	57d393674f9b0b49364ccb1c	144.7795	-37.74196	227921	T20120012303	ABS to receive accident	Finished	24/5/2(
2	57d393674f9b0b49364ccb1d	144.3262	-38.19084	227925	T20120011812	ABS to receive accident	Finished	20/5/2(
3	57d393674f9b0b49364ccb1e	144.9551	-37.66723	227926	T20120011813	ABS to receive accident	Finished	20/5/2(
4	57d393674f9b0b49364ccb1f	145.0098	-37.84896	227927	T20120011814	ABS to receive accident	Finished	20/5/2(
5	57d393674f9b0b49364ccb20	144.9484	-37.82523	227928	T20120011815	ABS to receive accident	Finished	19/5/2(
6	57d393674f9b0b49364ccb21	144.8177	-37.73773	227929	T20120011823	ABS to receive accident	Finished	20/5/2(
7	57d393674f9b0b49364ccb22	145.2795	-37.78418	227930	T20120011824	ABS to receive accident	Finished	20/5/2(
8	57d393674f9b0b49364ccb23	144.9388	-37.72947	227924	T20120011811	ABS to receive accident	Finished	20/5/2(
9	57d393674f9b0b49364ccb24	144.8213	-37.82561	227932	T20120011828	ABS to receive accident	Finished	20/5/2(
10	57d393674f9b0b49364ccb25	144.9814	-37.78305	227931	T20120011827	ABS to receive accident	Finished	14/5/2(
11	57d393674f9b0b49364ccb26	143.5607	-35.34281	227934	T20120011837	ABS to receive accident	Finished	20/5/2(

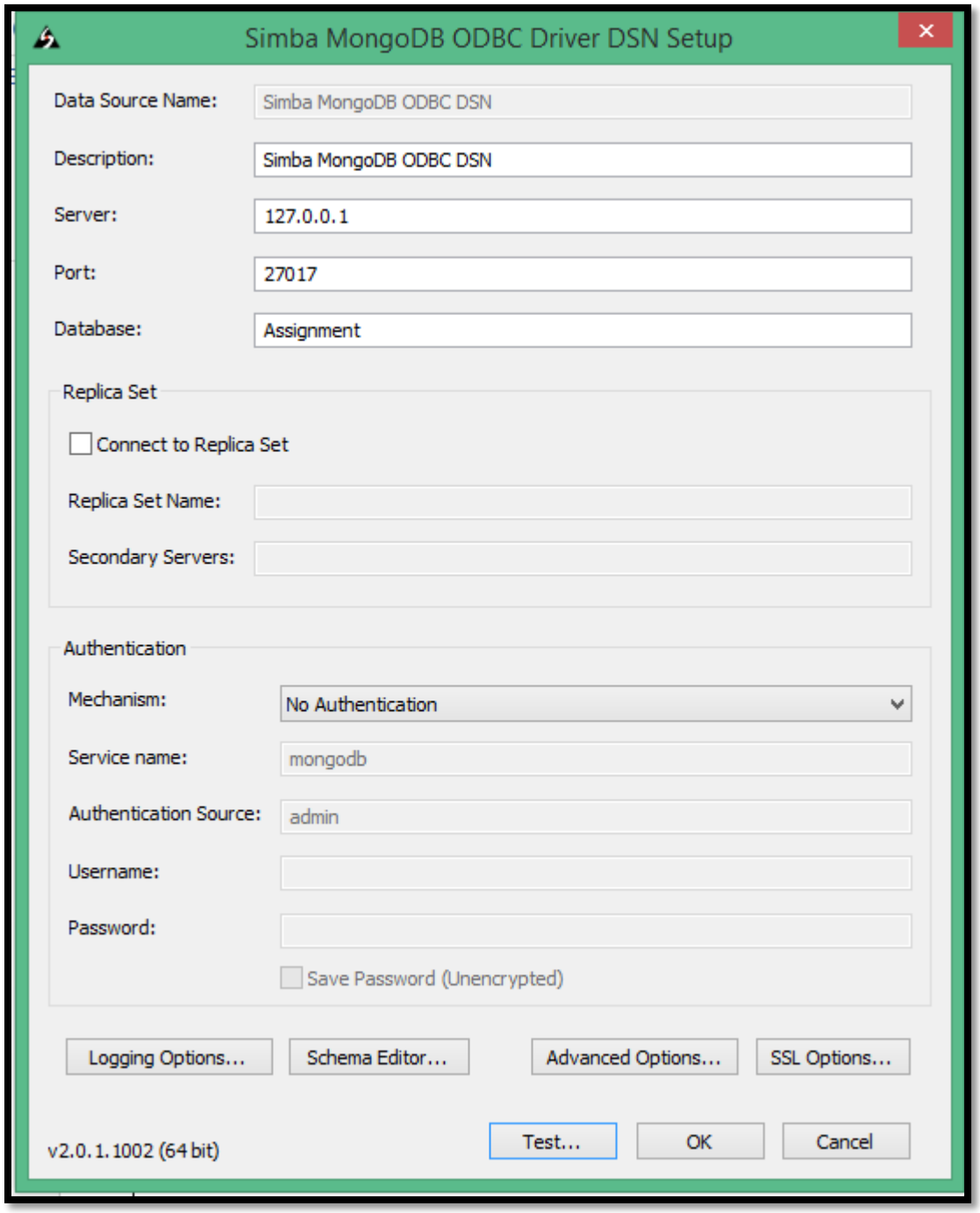
Showing 1 to 12 of 75,534 entries

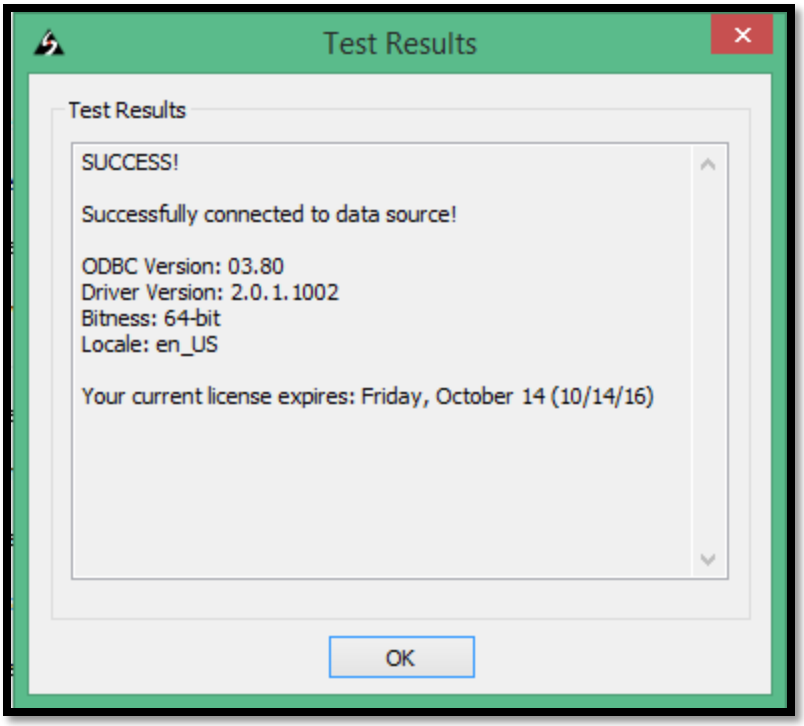
Connecting to Tableau

Starting ODBC



Testing Connection




Test Results
Test Results
SUCCESS!
Successfully connected to data source!
ODBC Version: 03.80
Driver Version: 2.0.1.1002
Bitness: 64-bit
Locale: en_US
Your current license expires: Friday, October 14 (10/14/16)
OK

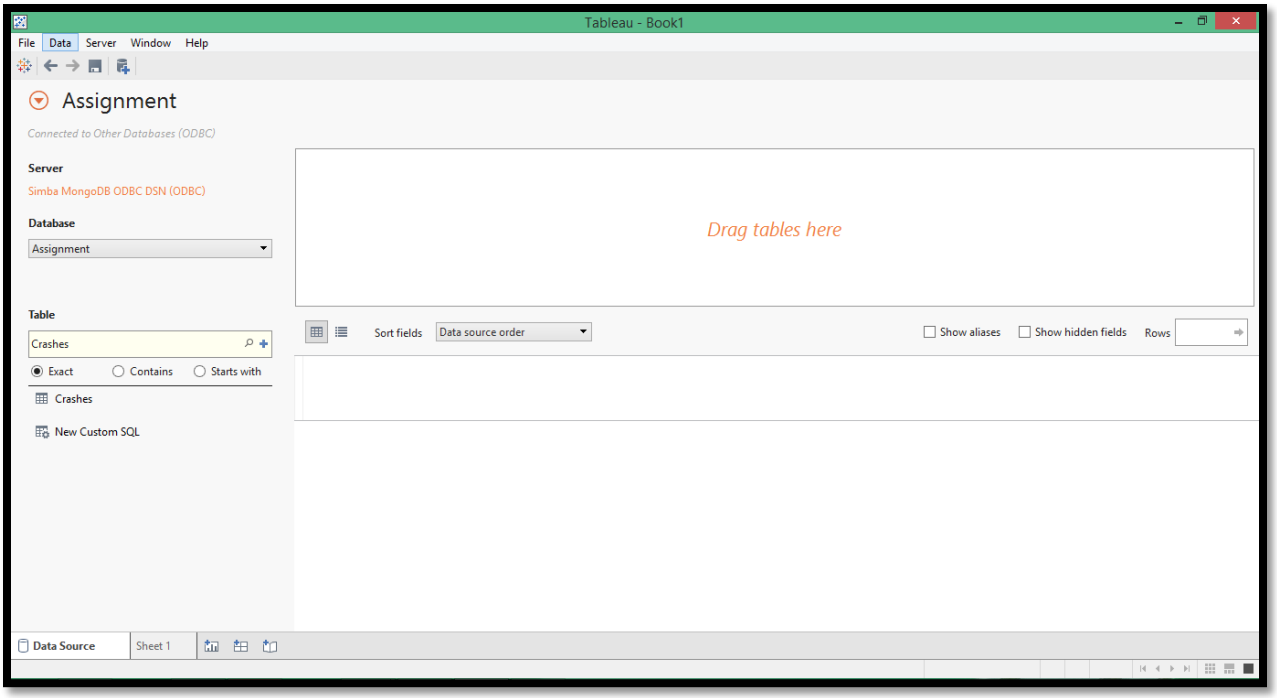


Tableau - Book1
File
Data
Server
Window
Help
Assignment
Connected to Other Databases (ODBC)
Server
Simba MongoDB ODBC DSN (ODBC)
Database
Assignment
Table
Crashes
Exact
Contains
Starts with
Crashes
New Custom SQL
Drag tables here
Sort fields
Data source order
Show aliases
Show hidden fields
Rows
Data Source
Sheet 1

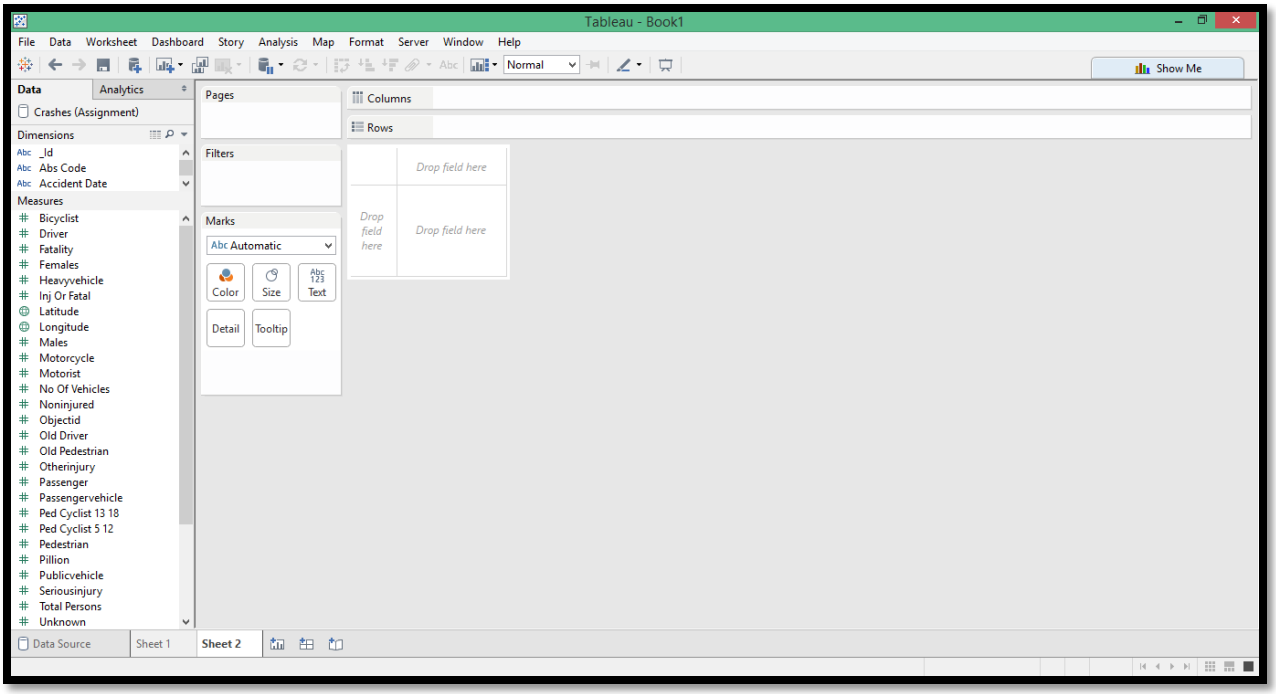
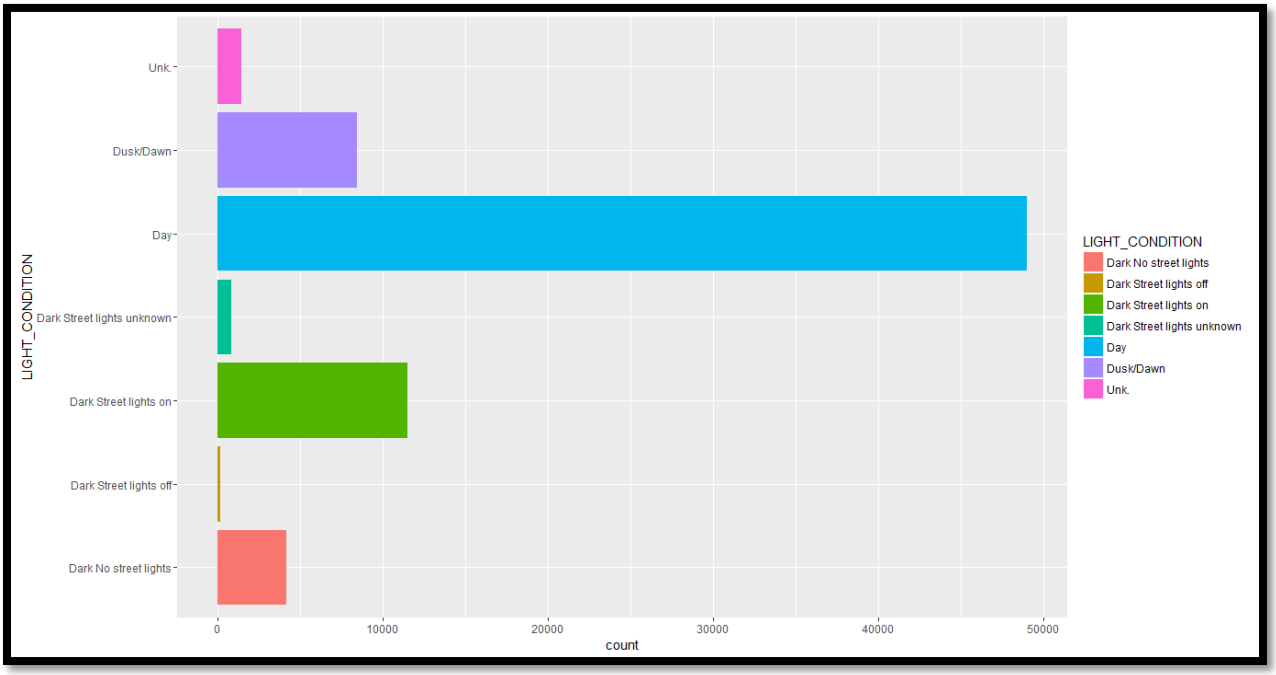
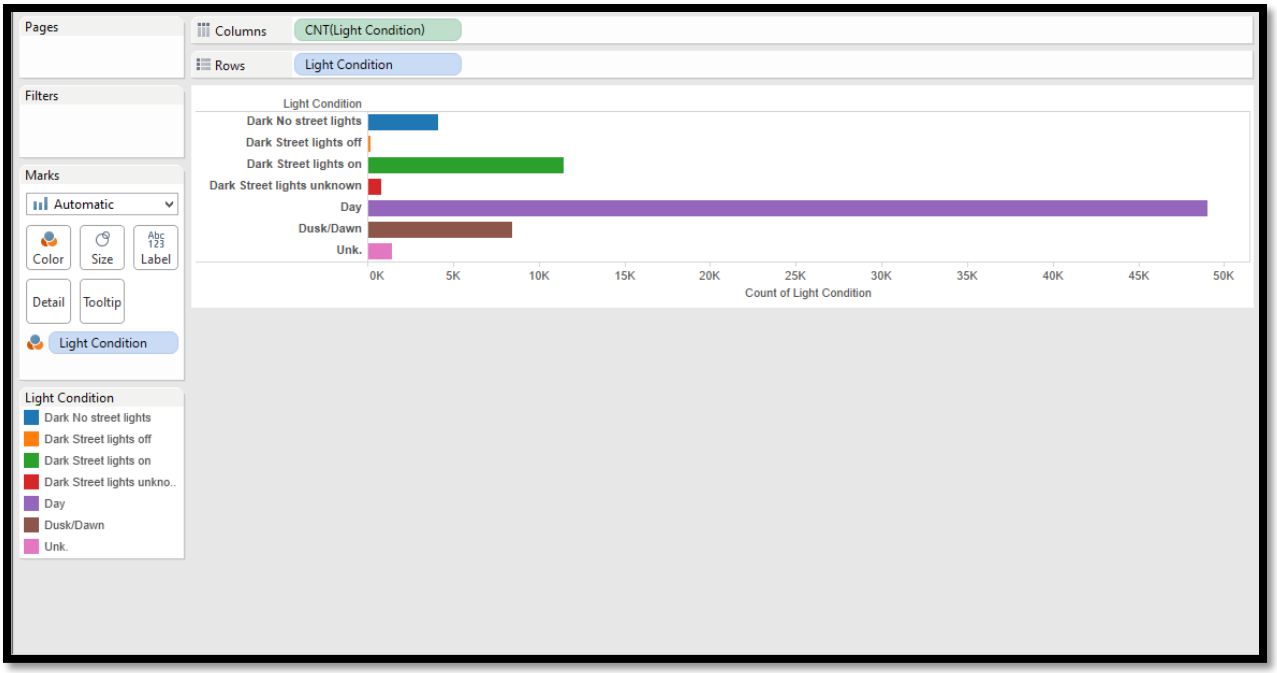


Tableau - Book1
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Normal
Show Me
Data
Analytics
Crashes (Assignment)
Dimensions
_Id
Abs Code
Accident Date
Measures
Bicyclist
Driver
Fatality
Females
Heavyvehicle
Inj Or Fatal
Latitude
Longitude
Males
Motorcycle
Motorist
No Of Vehicles
Noninjured
Objectid
Old Driver
Old Pedestrian
Otherinjury
Passenger
Passengervehicle
Ped Cyclist 13 18
Ped Cyclist 5 12
Pedestrian
Pillion
Publicvehicle
Seriousinjury
Total Persons
Unknown
Pages
Filters
Marks
Automatic
Color
Size
Text
Detail
Tooltip
Columns
Rows
Drop field here
Drop field here
Drop field here
Data Source
Sheet 1
Sheet 2

Analyzing Crashes dataset

```
12  ##############################################################################
13  ##Viewing number of accidents by Accident Type
14  ##############################################################################
15
16  library(ggplot2)
17  library(data.table)
18  ggplot(data = myDataFrame, aes(x = LIGHT_CONDITION,fill =LIGHT_CONDITION)) +
19  geom_bar()+ coord_flip()
```





The above graph shows number of accidents vs light conditions in which accidents took place. Although accidents occurred during day are more. We can take steps in order to prevent accidents occurring in Dark Streets where lights are off and Dark Streets where no Street lights are there.

Case 1: Accidents occurring due to Dark Street light off

Sub setting the data according to Light Type = Dark Street light off

```
24  ################################################################################
25  ##Sub setting data according to light type equal to dark street light off
26  ################################################################################
27  x.sub <- subset(myDataFrame, LIGHT_CONDITION == "Dark Street lights off")
28  View(x.sub)
29
```

Records reduced from 75534 to 164

Number_1.R* | x.sub | myDataFrame | Assignment_1.R | Assignment_final.R | Assignement.R

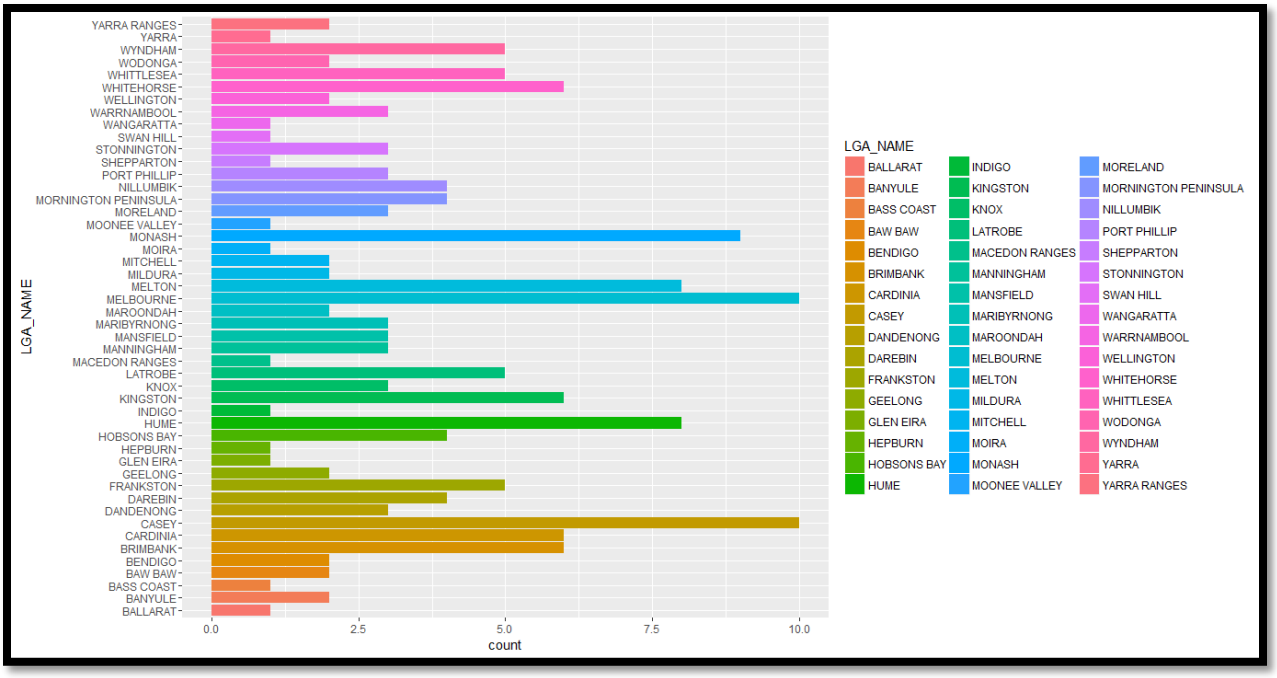
Filter

	HIT_RUN_FLAG	LIGHT_CONDITION	POLICE_ATTEND	ROAD_GEOMETRY	SEVERITY
ICLES IN SAME LANE)	No	Dark Street lights off	Yes	T intersection	Serious injury accident
BROKEN DOWN	No	Dark Street lights off	Yes	Not at intersection	Other injury accident
IDES WITH VEHICLE PARKED ON LEFT OF...	No	Dark Street lights off	Yes	Not at intersection	Other injury accident
TROL ON CARRIAGEWAY (ON STRAIGHT)	No	Dark Street lights off	No	T intersection	Other injury accident
E. PED HIT BY VEHICLE FROM THE RIGHT.	No	Dark Street lights off	Yes	Not at intersection	Serious injury accident
RIAGEWAY INTO OBJECT/PARKED VEHICLE	No	Dark Street lights off	Yes	Not at intersection	Serious injury accident
ICLES IN SAME LANE)	No	Dark Street lights off	No	T intersection	Other injury accident
OVERTAKING)	No	Dark Street lights off	Yes	Not at intersection	Other injury accident
ICLES IN SAME LANE)	No	Dark Street lights off	No	Not at intersection	Other injury accident
VEHICLE	No	Dark Street lights off	Yes	Not at intersection	Other injury accident
ND INTO OBJECT/PARKED VEHICLE	No	Dark Street lights off	Yes	Not at intersection	Serious injury accident

Showing 1 to 12 of 164 entries

Classifying the accidents according to suburb in which they occurred:

```
31  ################################################################################
32  ##Classifying accident count according to suburb
33  ################################################################################
34
35  ggplot(data = x.sub, aes(x = LGA_NAME,fill = LGA_NAME)) + geom_bar() + coord_flip()
36
```



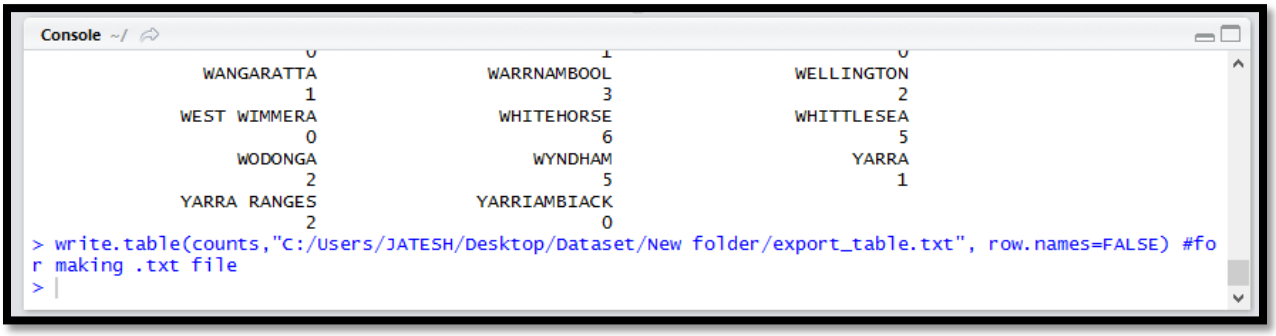
Considering the financial constraint it is incorrect to offer street lights directly where accidents took place so for now we will follow a step by step approach and first offer street lights to top 5 accident prone suburb.

```
37
38  ###############################################################################
39  ##Finding top 5 suburbs
40  ###############################################################################
41  counts <- table(x.sub$LGA_NAME)
42  summary(x.sub$LGA_NAME)
43  write.table(counts,"C:/Users/JATESH/Desktop/Dataset/New folder/export_table.txt", row.names=FALSE)
44  barplot(sort(counts,decreasing=TRUE)[1:5],
45          main="Accidents due to Street Lights OFF", col="blue", xlab="City",
46          ylab="Number of Accidents",
47          border="red", density=c(90, 70, 50, 40, 30))
48  box()
49
46:9  (Top Level)                                                    R Script
```

Output of summary command:

```
> summary(x.sub$LGA_NAME)
                                (FALLS CREEK)          (LAKE MOUNTAIN)
                     0                      0                        0
         (MOUNT BAW BAW)          (MOUNT BULLER)          (MOUNT HOTHAM)
                     0                      0                        0
                ALPINE                 ARARAT                 BALLARAT
                     0                      0                        1
               BANYULE             BASS COAST                  BAW BAW
                     2                      1                        2
               BAYSIDE                BENALLA                 BENDIGO
                     0                      0                        2
           BOROONDARA               BRIMBANK                  BULOKE
                     0                      6                        0
             CAMPASPE              CARDINIA                   CASEY
                     0                      6                       10
  CENTRAL GOLDFIELDS            COLAC OTWAY             CORANGAMITE
                     0                      0                        0
            DANDENONG               DAREBIN          EAST GIPPSLAND
                     3                      4                        0
            FRANKSTON             GANNAWARRA                 GEELONG
                     5                      0                        2
            GLEN EIRA               GLENELG           GOLDEN PLAINS
                     1                      0                        0
              HEPBURN              HINDMARSH             HOBSONS BAY
                     1                      0                        4
              HORSHAM                  HUME                   INDIGO
                     0                      8                        1
             KINGSTON                  KNOX                  LATROBE
                     6                      3                        5
               LODDON        MACEDON RANGES                MANNINGHAM
                     0                      1                        3
```

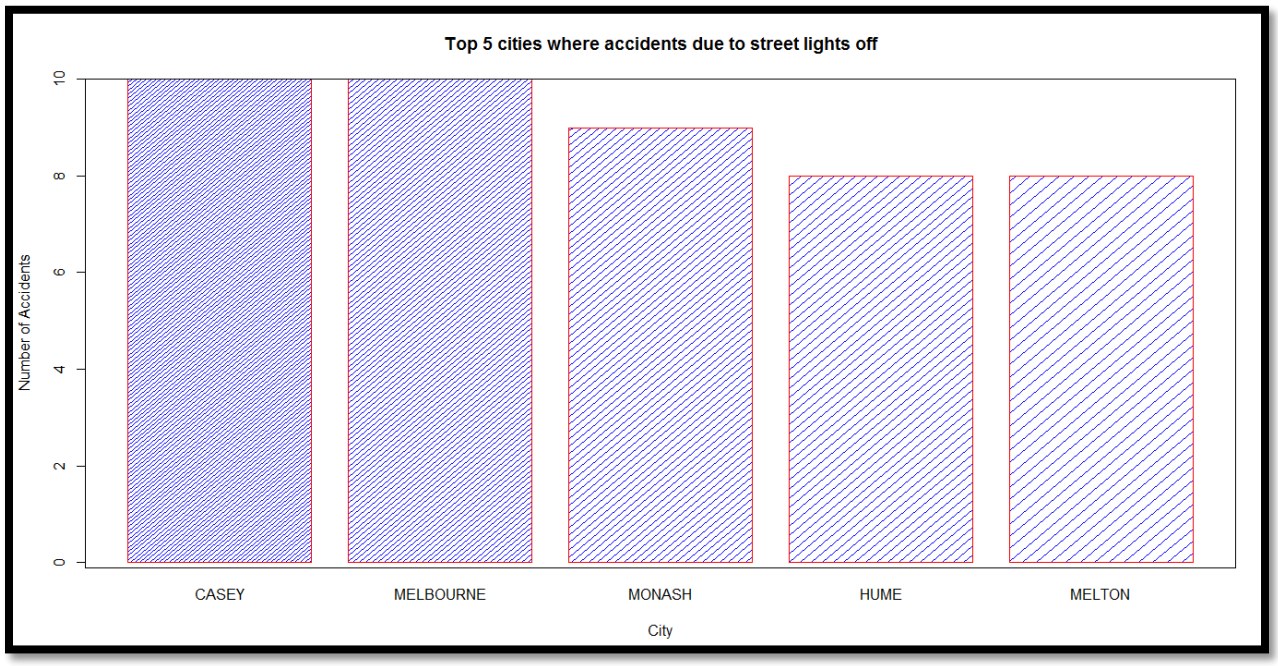
Output of transferring summary output to .txt file

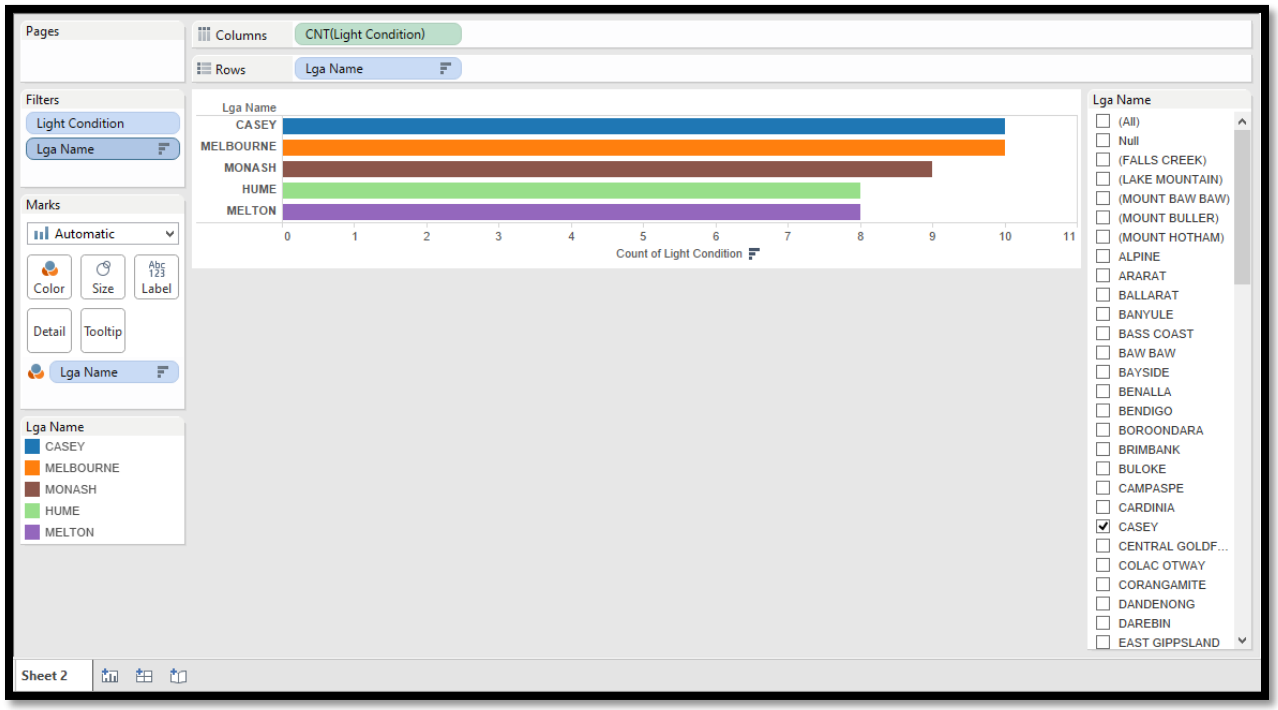


Structure of .txt file:

```
"Var1" "Freq"
"" 0
"(FALLS CREEK)" 0
"(LAKE MOUNTAIN)" 0
"(MOUNT BAW BAW)" 0
"(MOUNT BULLER)" 0
"(MOUNT HOTHAM)" 0
"ALPINE" 0
"ARARAT" 0
"BALLARAT" 1
"BANYULE" 2
"BASS COAST" 1
"BAW BAW" 2
"BAYSIDE" 0
"BENALLA" 0
"BENDIGO" 2
"BOROONDARA" 0
"BRIMBANK" 6
"BULOKE" 0
"CAMPASPE" 0
"CARDINIA" 6
"CASEY" 10
"CENTRAL GOLDFIELDS" 0
"COLAC OTWAY" 0
"CORANGAMITE" 0
"DANDENONG" 3
"DAREBIN" 4
"EAST GIPPSLAND" 0
"FRANKSTON" 5
"GANNAWARRA" 0
"GEELONG" 2
"GLEN EIRA" 1
"GLENELG" 0
"GOLDEN PLAINS" 0
"HEPBURN" 1
"HINDMARSH" 0
"HOBSONS BAY" 4
```

Top 5 cities where accident took place due to street light being off



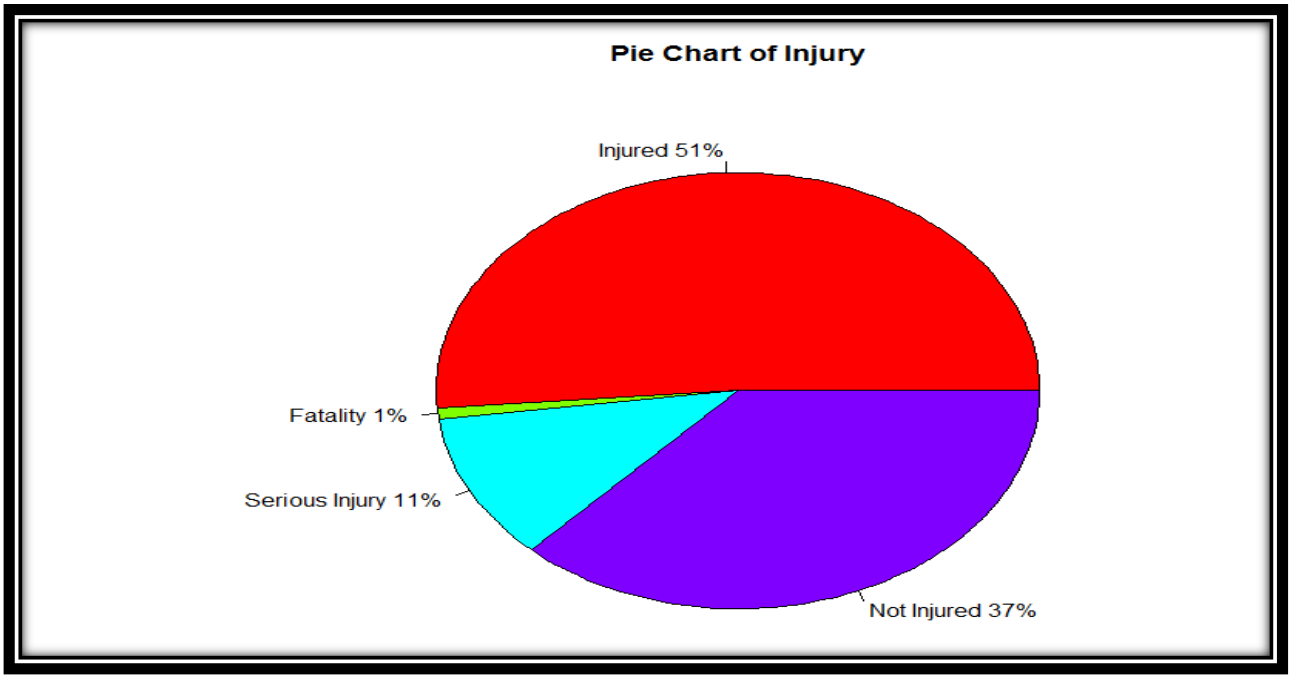


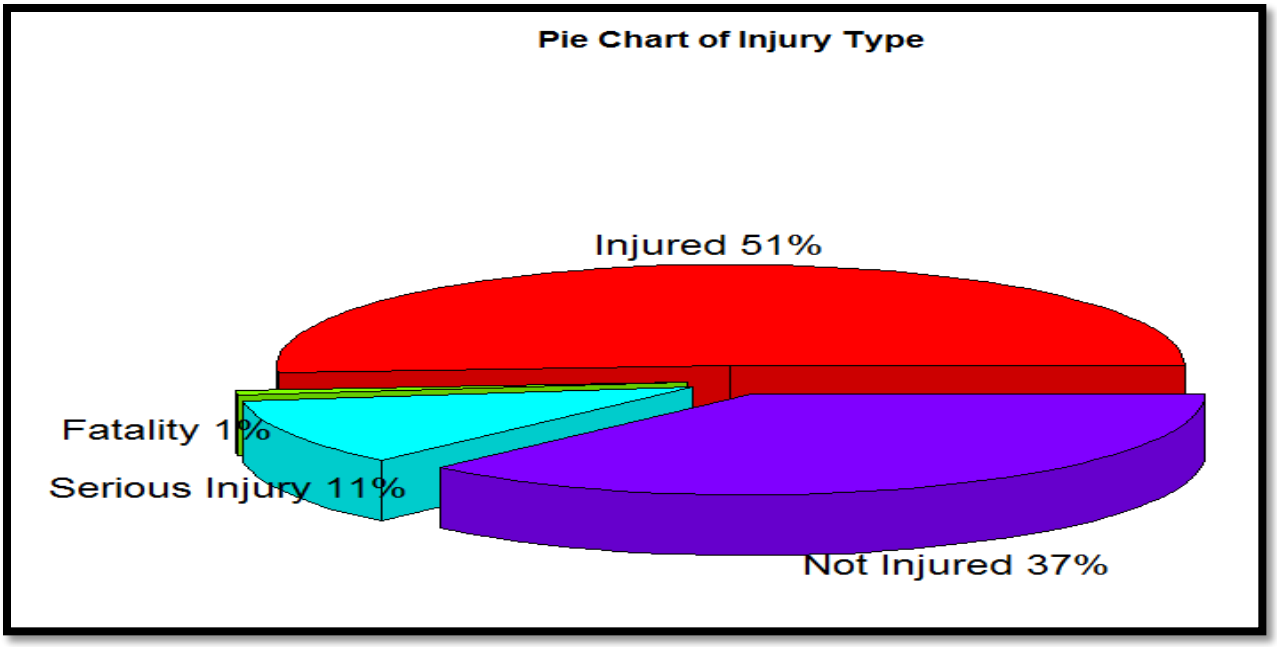
Demographics of top 5 cities where lights were off

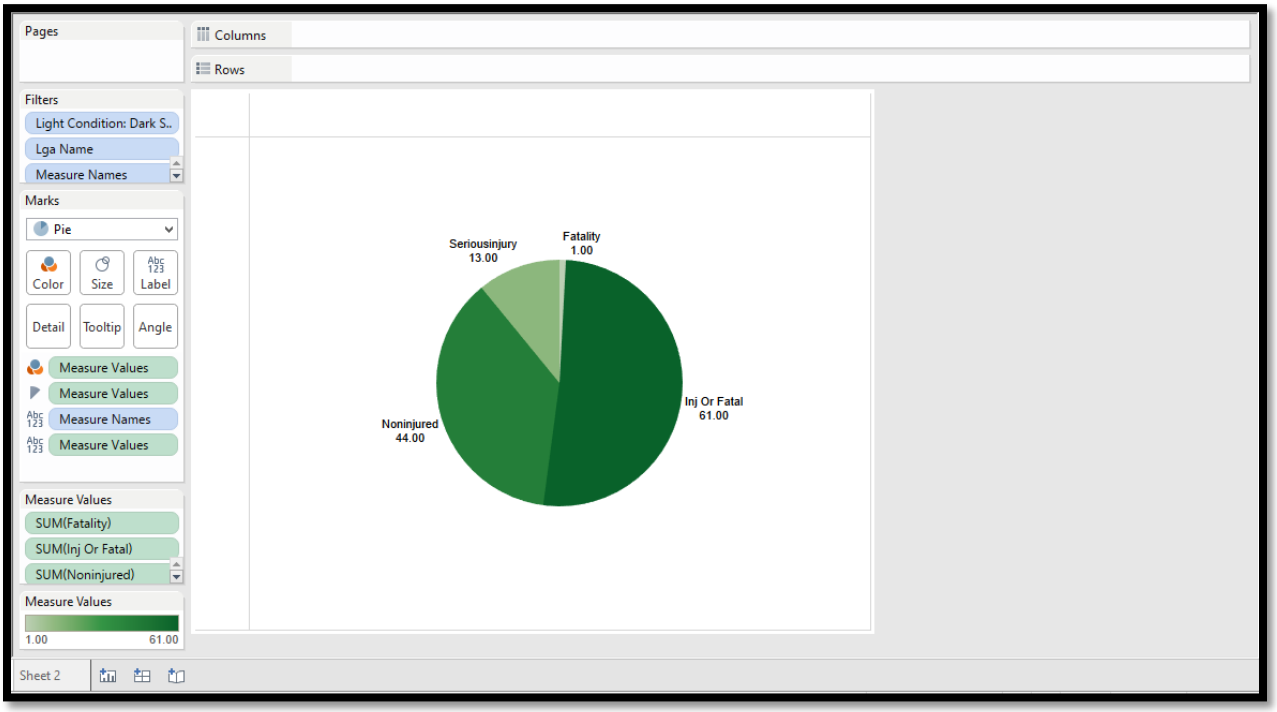
Demographics by Injury type

```
50  ##############################################################################
51  ##Demographics of Injury type in top 5 cities
52  ##############################################################################
53  topcity <-subset(x.sub, LGA_NAME == "CASEY" | LGA_NAME == "MELBOURNE" | LGA_NAME == "MONASH" |
54                     LGA_NAME == "MELTON" | LGA_NAME == "HUME")
55  view(topcity)
56  Slices <- c(sum(topcity$INJ_OR_FATAL),sum(topcity$FATALITY),sum(topcity$SERIOUSINJURY),
57              sum(topcity$NONINJURED))
58  Slices
59  lbls <- c("Injured", "Fatality","Serious Injury", "Not Injured")
60  pct <- round(Slices/sum(Slices)*100)
61  lbls <- paste(lbls, pct) # add percents to labels
62  lbls <- paste(lbls,"%",sep="") # ad % to labels
63  pie(Slices,labels = lbls, col=rainbow(length(lbls)),
64      main="Pie Chart of Injury Type")
65  library(plotrix)
66  pie3D(Slices,labels=lbls,explode=0.1,
67        main="Pie Chart of Injury Type")
```

67:37 (Top Level) R Script





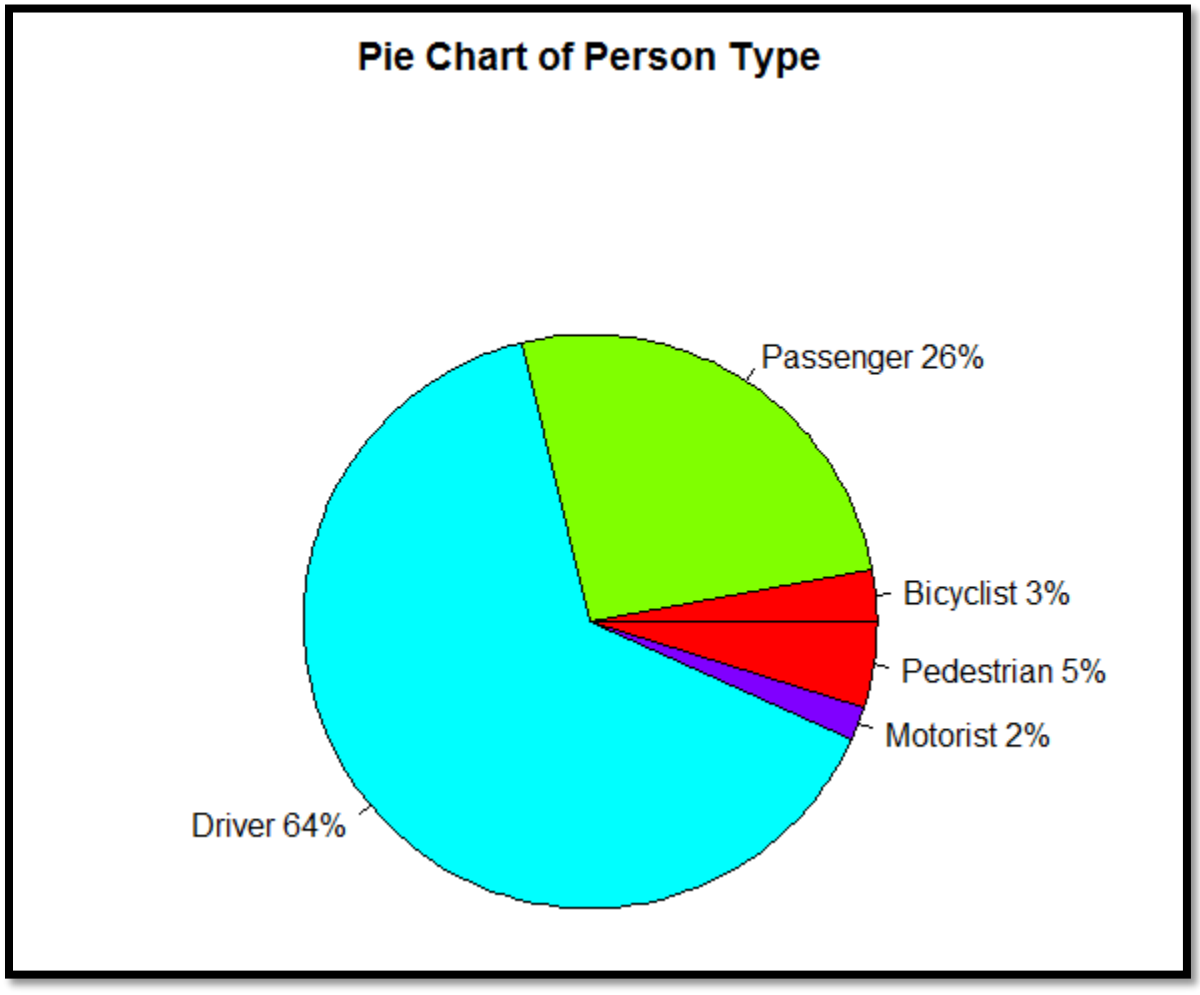


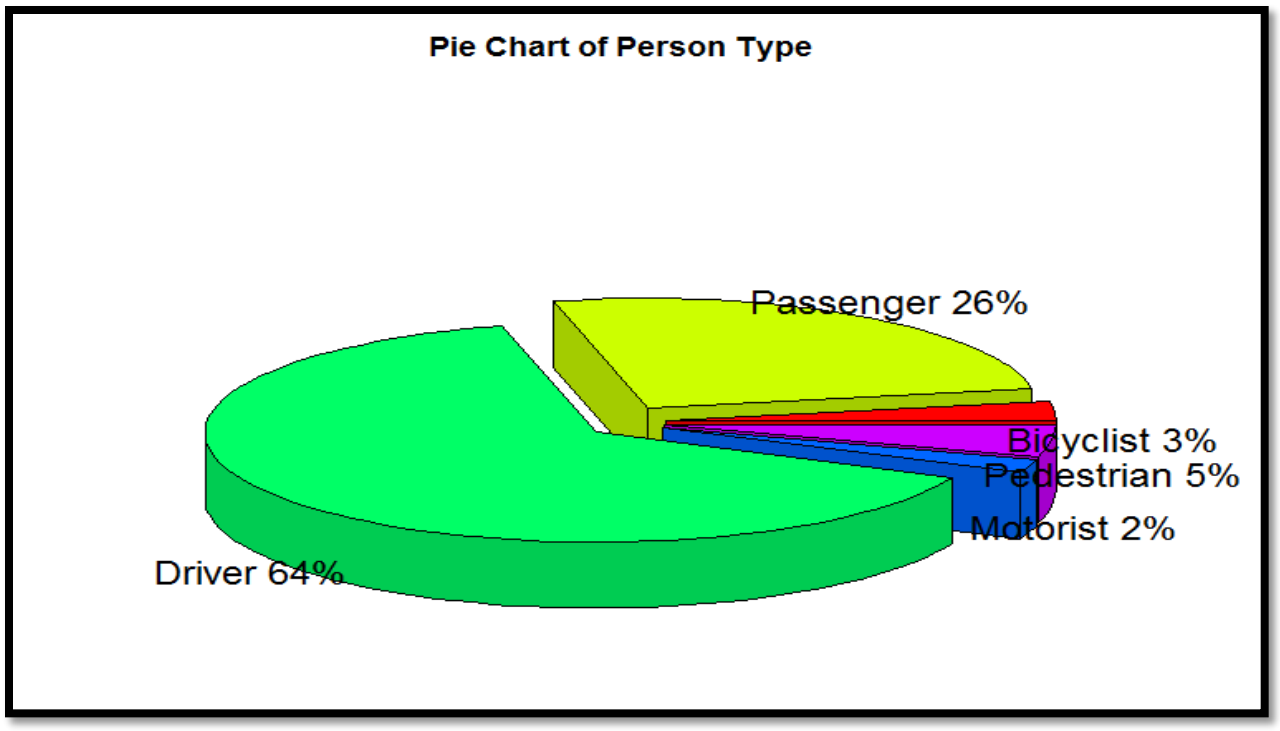
From the above two pie charts it is clear that in accidents of top 5 cities where street lights are off people getting injured is 63% which is very high. This supports the idea of starting the street lights in the area as the probability of injury occurring after accident is high in this cities.

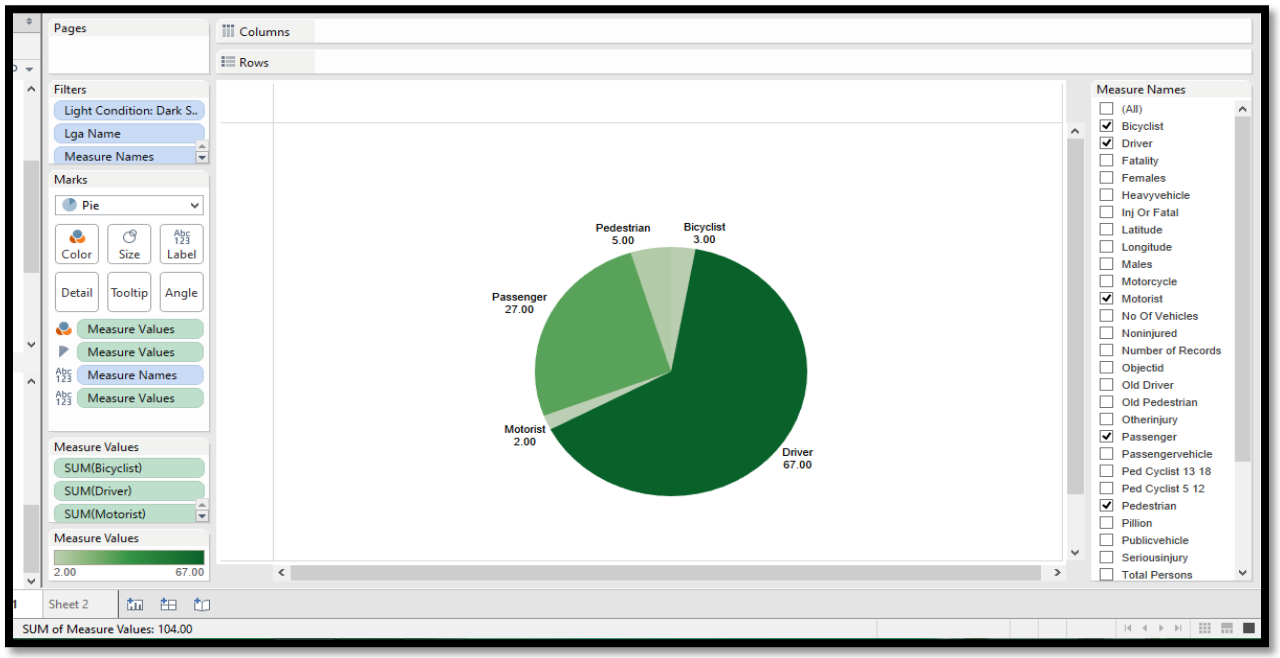
Demographics by People Involved

```
69  ###############################################################################
70  ##Demographics of people type involved in accidents of top 5 cities
71  ###############################################################################
72  slices1 <- c(sum(topcity$BICYCLIST),sum(topcity$PASSENGER),sum(topcity$DRIVER),
73               sum(topcity$PEDESTRIAN),sum(topcity$MOTORIST))
74  slices1
75  lbls1 <- c("Bicyclist", "Passenger","Driver", "Pedestrian", "Motorist")
76  pct <- round(slices1/sum(slices1)*100)
77  pct
78  lbls1 <- paste(lbls1, pct) # add percents to labels
79  lbls1 <- paste(lbls1,"%",sep="") # ad % to labels
80  pie(slices1,labels = lbls1, col=rainbow(length(lbls)),
81      main="Pie Chart of Injury")
82  library(plotrix)
83  pie3D(slices1,labels=lbls1,explode=0.3,
84        main="Pie Chart of Person Type ")
```

65:17 (Top Level) R Script



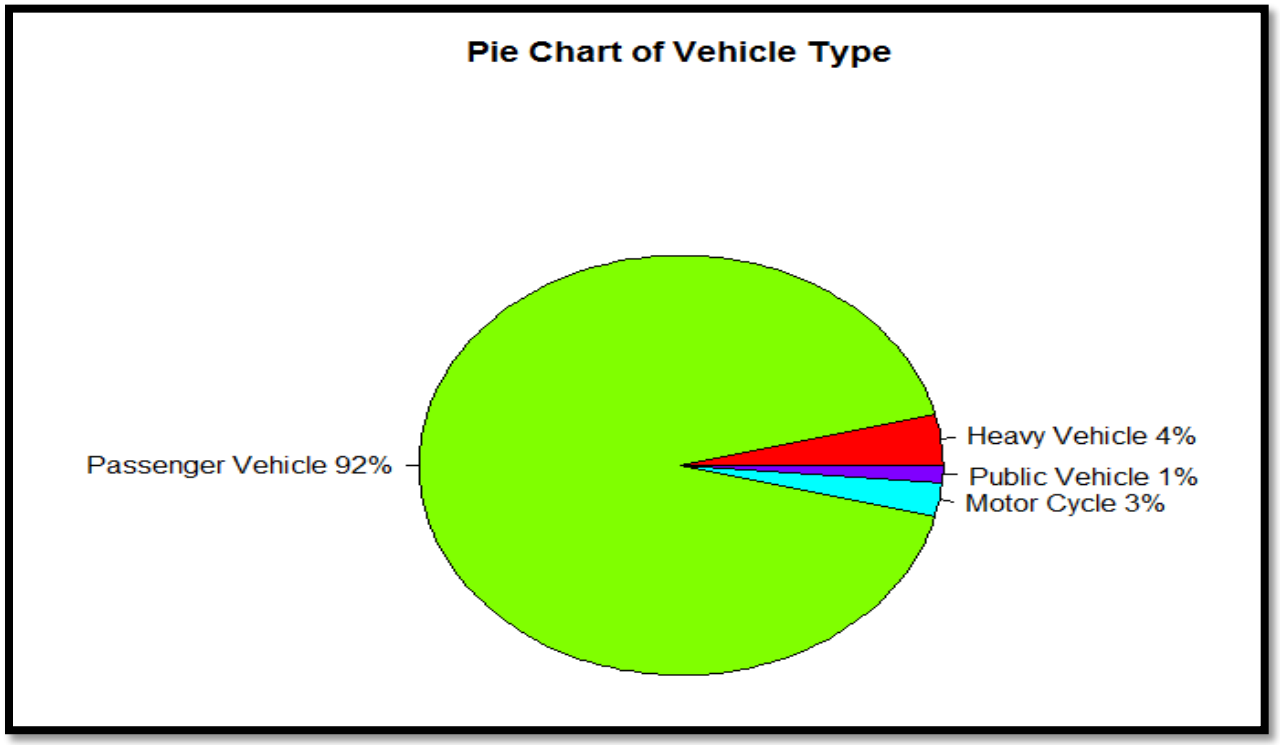


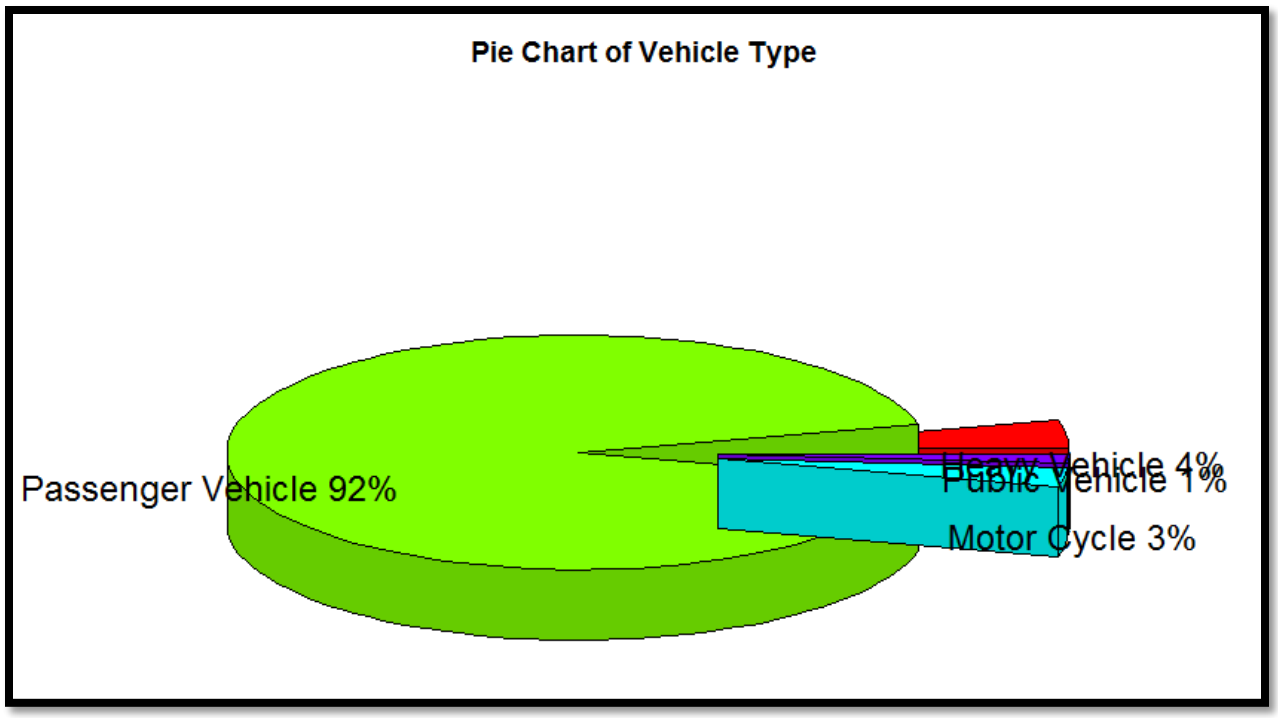


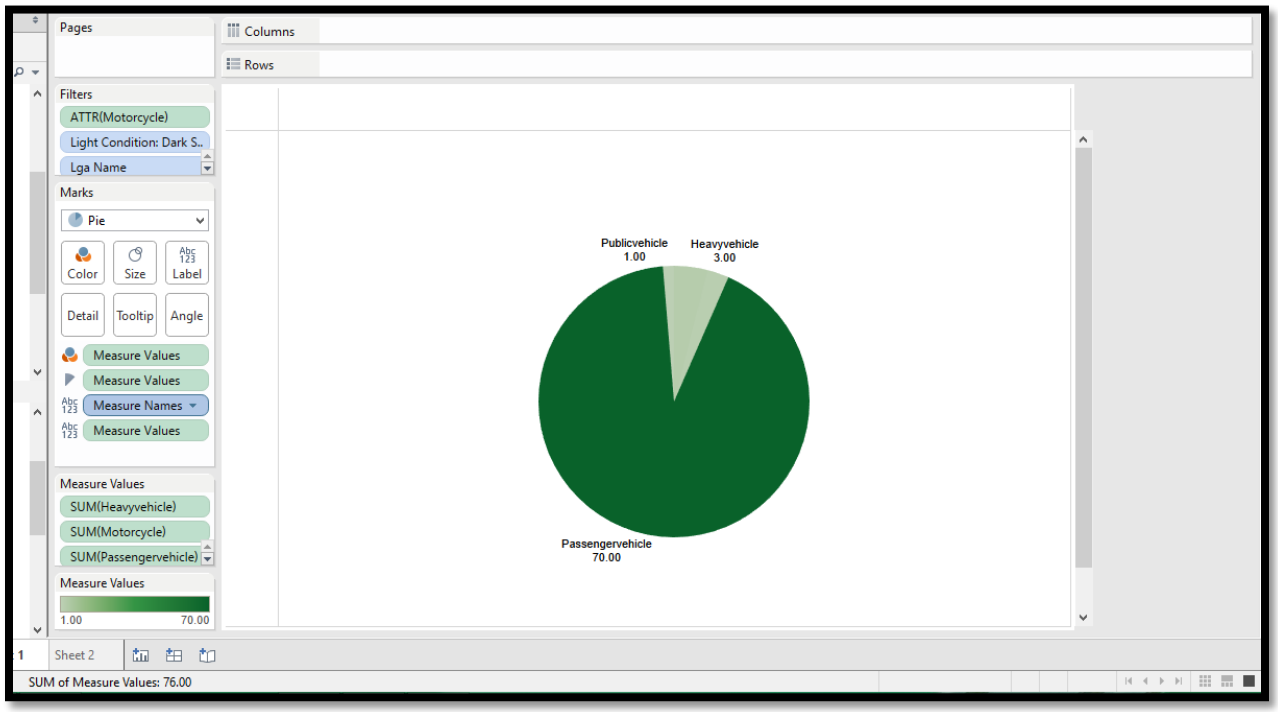
The above pie chart suggest 64% of the drivers getting injured. Normally car accidents cause damage to property and loss of life. Since the percentage of drivers is high a lot of cars accidents must be happening in these areas. So this demographics too support the need to start street lights in the areas.

Demographics by Vehicle Type

```
87  ################################################################################
88  ##Demographics of vehicle type involved in accidents of top 5 cities
89  ################################################################################
90  slices2 <- c(sum(topcity$HEAVYVEHICLE),sum(topcity$PASSENGERVEHICLE),sum(topcity$MOTORCYCLE),
91               sum(topcity$PUBLICVEHICLE))
92  slices2
93  lbls2 <- c("Heavy Vehicle", "Passenger Vehicle","Motor Cycle", "Public Vehicle")
94  pct2 <- round(slices2/sum(slices2)*100)
95  pct2
96  lbls2 <- paste(lbls2, pct2) # add percents to labels
97  lbls2 <- paste(lbls2,"%",sep="") # ad % to labels
98  pie(slices2,labels = lbls2, col=rainbow(length(lbls)),
99      main="Pie Chart of Vehicle Type")
100 library(plotrix)
101 pie3D(slices2,labels=lbls2,explode=0.2,
102       main="Pie Chart of Vehicle Type ")
103 ################################################################################
104
101:40  (Top Level)                                                    R Script
```



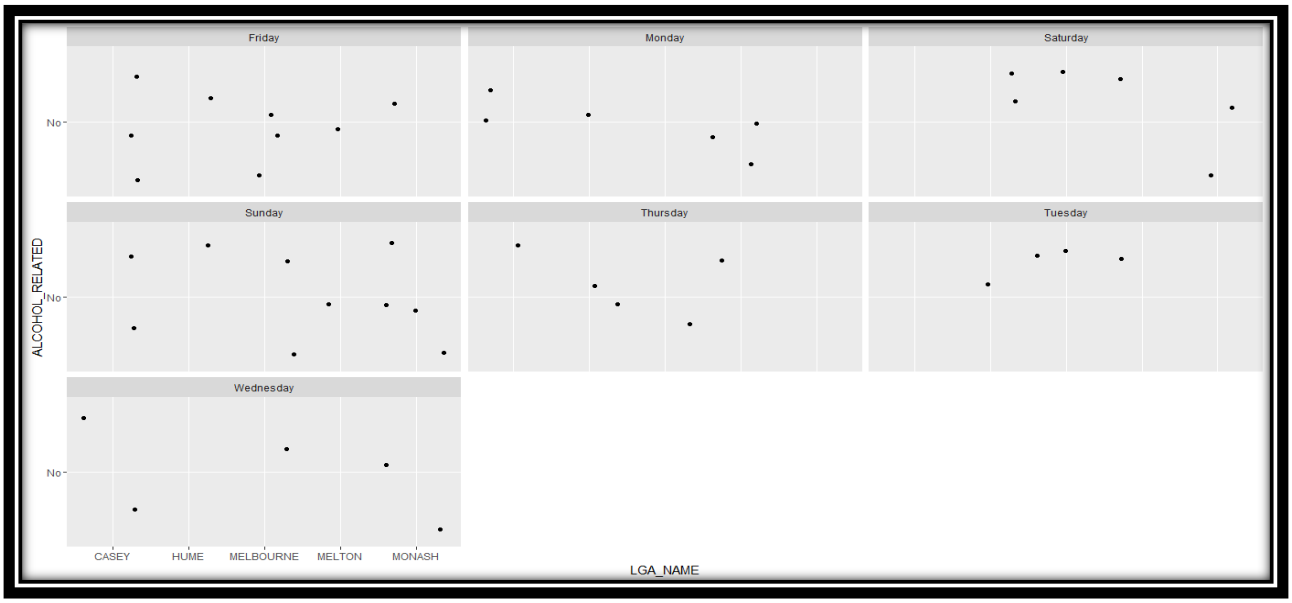


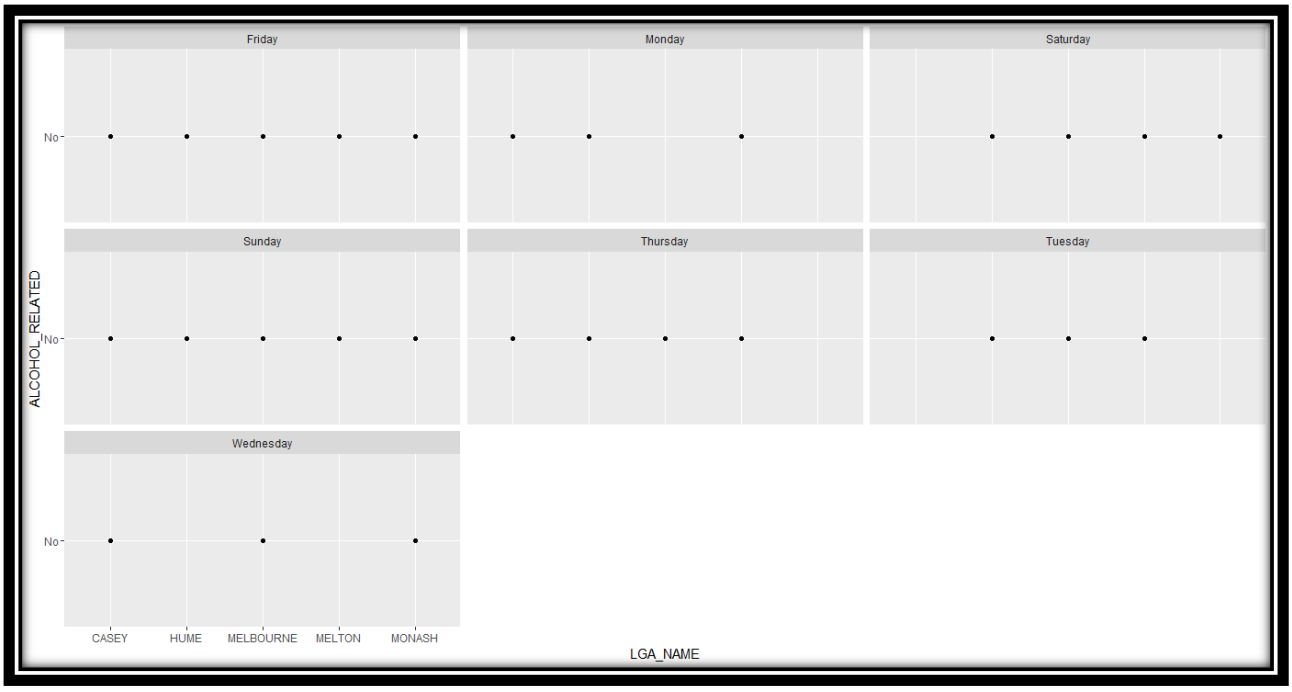


The above demographics also convey most of the accidents involve passenger vehicle.

Impact of Alcohol

```
86  ###############################################################################
87  ##Alcohol Involvement
88  ###############################################################################
89
90  ggplot(topcity, aes(LGA_NAME, ALCOHOL_RELATED)) + geom_jitter()+ facet_wrap(~ DAY_OF_WEEK)
91  ggplot(topcity, aes(LGA_NAME, ALCOHOL_RELATED)) + geom_point()+ facet_wrap(~ DAY_OF_WEEK)
92
93
```

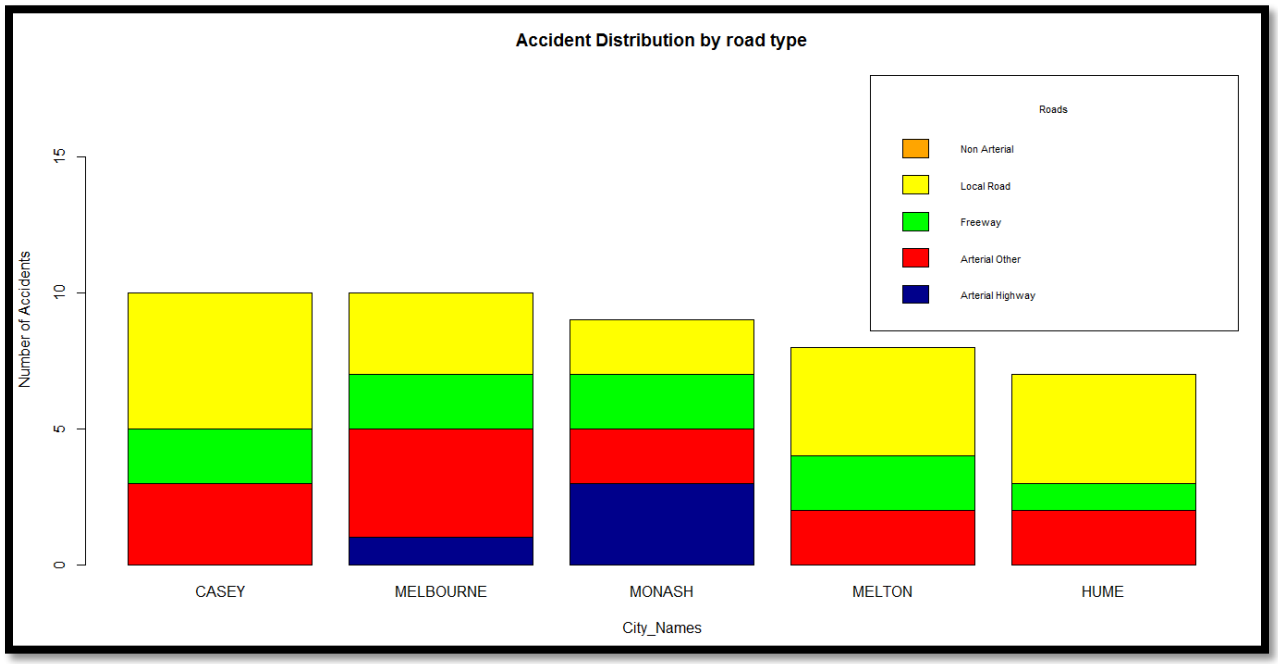


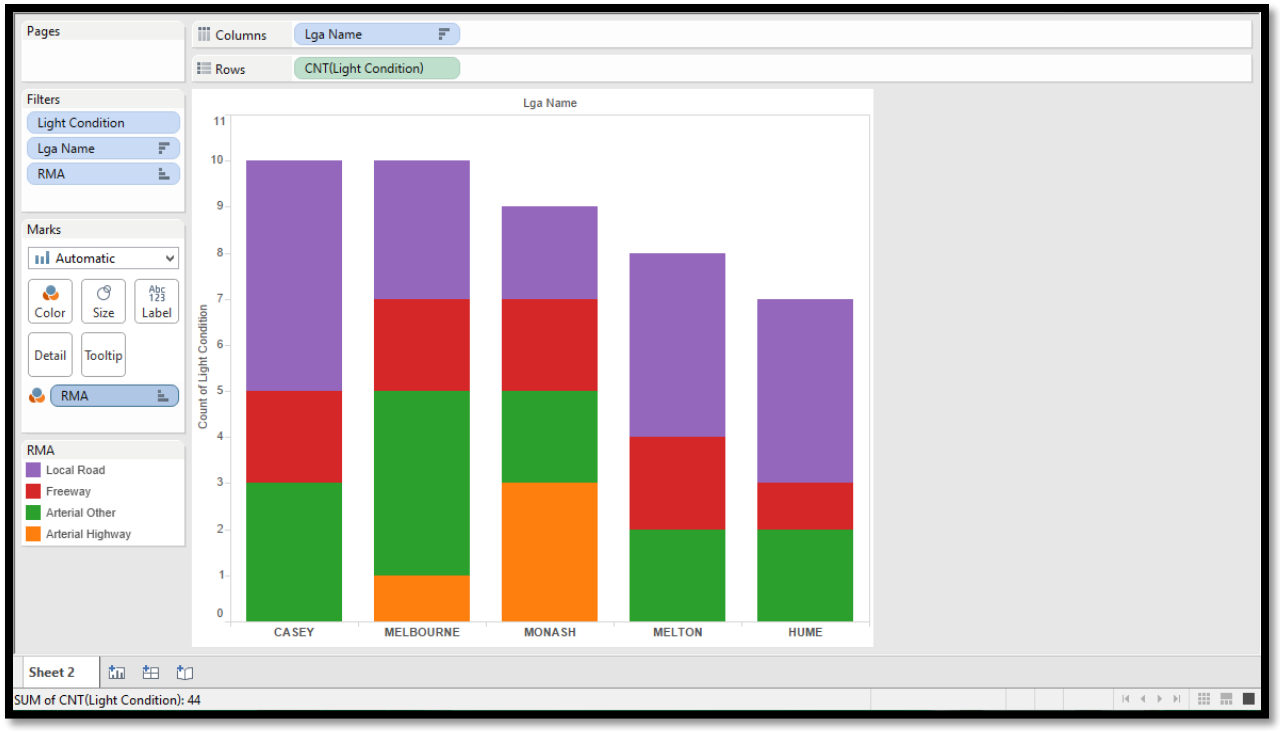


The above two graphs further show that the drivers involved in accidents were not under the influence of alcohol. Therefore improving light conditions could have avoided this accidents.

Accidents by Road Types

```
111 ###############################################################################
112 ##Road Types
113 ###############################################################################
114
115 counts_1 <- table(topcity$RMA, topcity$LGA_NAME)
116 counts_1
117 csel_counts_1 <- counts_1[,c("CASEY", "MELBOURNE","MONASH","MELTON","HUME")];
118 csel_counts_1
119 rsel_counts_1 <- csel_counts_1[c("Arterial Highway","Arterial Other","Freeway","Local Road","Non A
120 rsel_counts_1
121
122 barplot(rsel_counts_1, main="Accident Distribution by road type",
123         xlab="City_Names",ylab = "Number of Accidents", col = c("darkblue","red","green","yellow",
124         legend = c("Arterial Highway", "Arterial Other", "Freeway","Local Road","Non Arterial"),
125         args.legend = list(title = "Roads", x = "topright", cex = .7),ylim = c(0, 18))
126
127
```

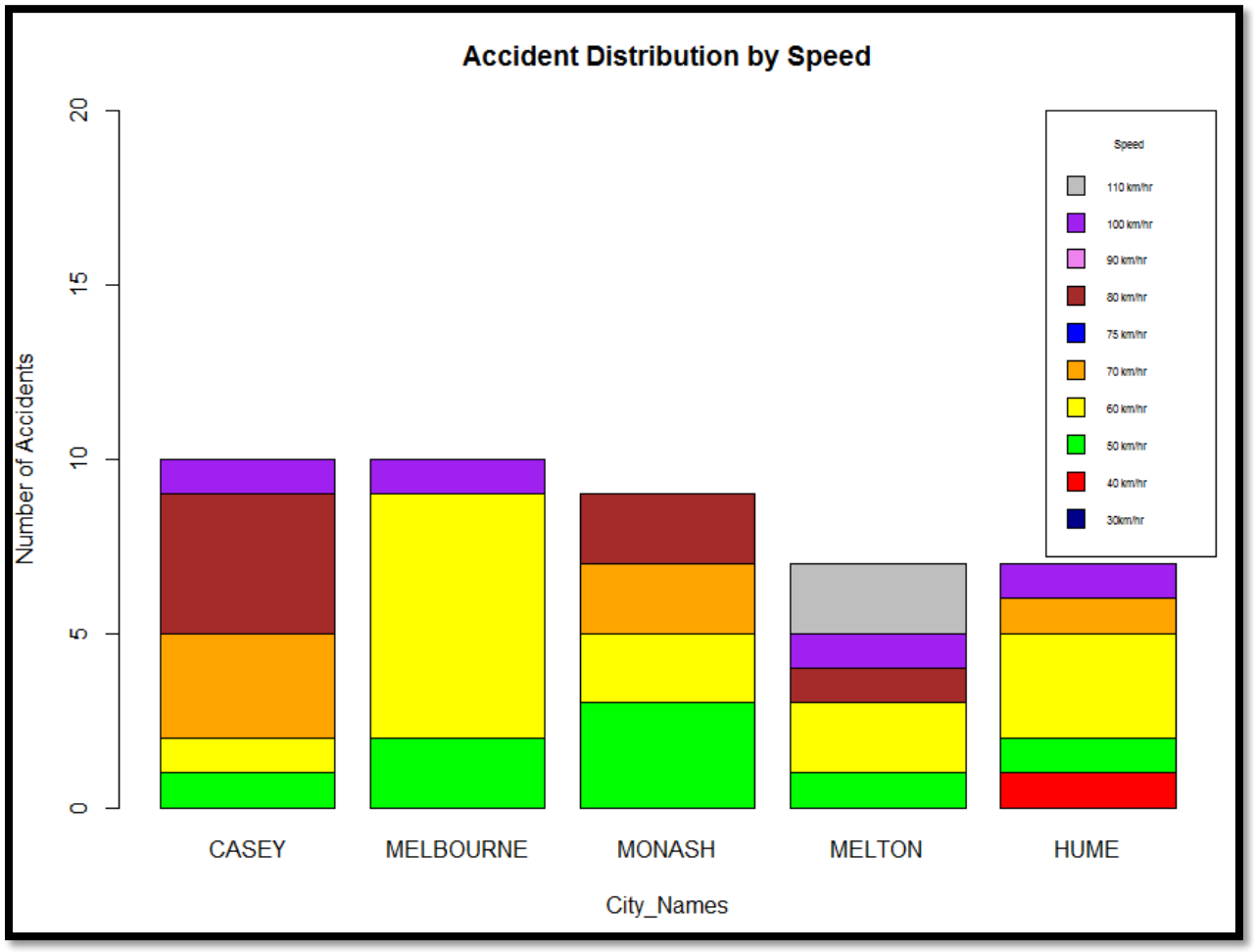


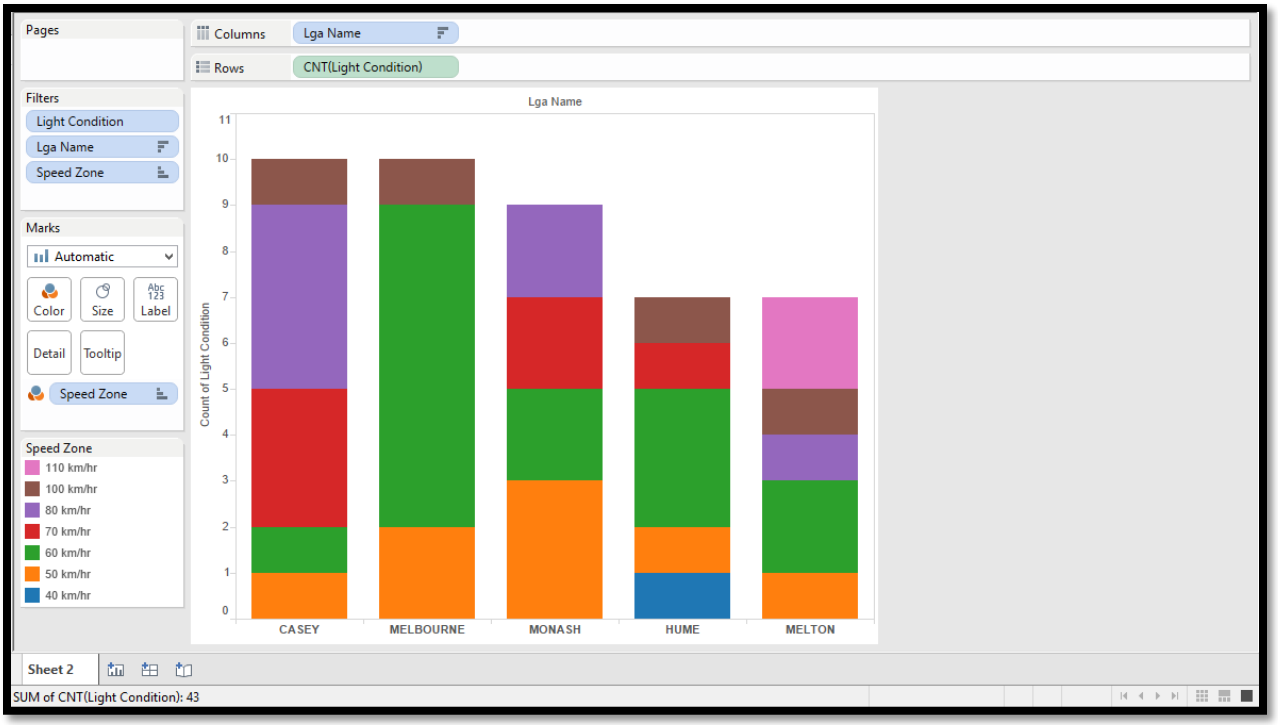


From the above graph it is clears for which city which road types pose danger of accident. For example Casey needs to improve light conditions on Non Arterial roads while Melbourne needs to solve light conditions on Freeway.

Accidents by Speed

```
111 ##############################################################################
112 ##Speed Analysis
113 ##############################################################################
114 counts_2 <- table(topcity$SPEED_ZONE, topcity$LGA_NAME);
115 counts_2
116 csel_counts_2 <- counts_2[,c("CASEY", "MELBOURNE","MONASH","MELTON","HUME")];
117 csel_counts_2
118 rsel_counts_2 <- csel_counts_2[c("30km/hr","40 km/hr","50 km/hr","60 km/hr","70 km/hr",
119                     "75 km/hr","80 km/hr","90 km/hr","100 km/hr","110 km/hr"),];
120 rsel_counts_2
121 barplot(rsel_counts_2, main="Accident Distribution by Speed",
122         xlab="City_Names",ylab = "Number of Accidents",
123         col = c("darkblue","red","green","yellow","orange","blue",
124                 "brown","violet","purple","grey"),
125         legend = c("30km/hr","40 km/hr","50 km/hr","60 km/hr","70 km/hr",
126                    "75 km/hr","80 km/hr","90 km/hr","100 km/hr","110 km/hr"),
127         args.legend = list(title = "Roads", x = "topright", cex = .5),ylim = c(0, 20))
```

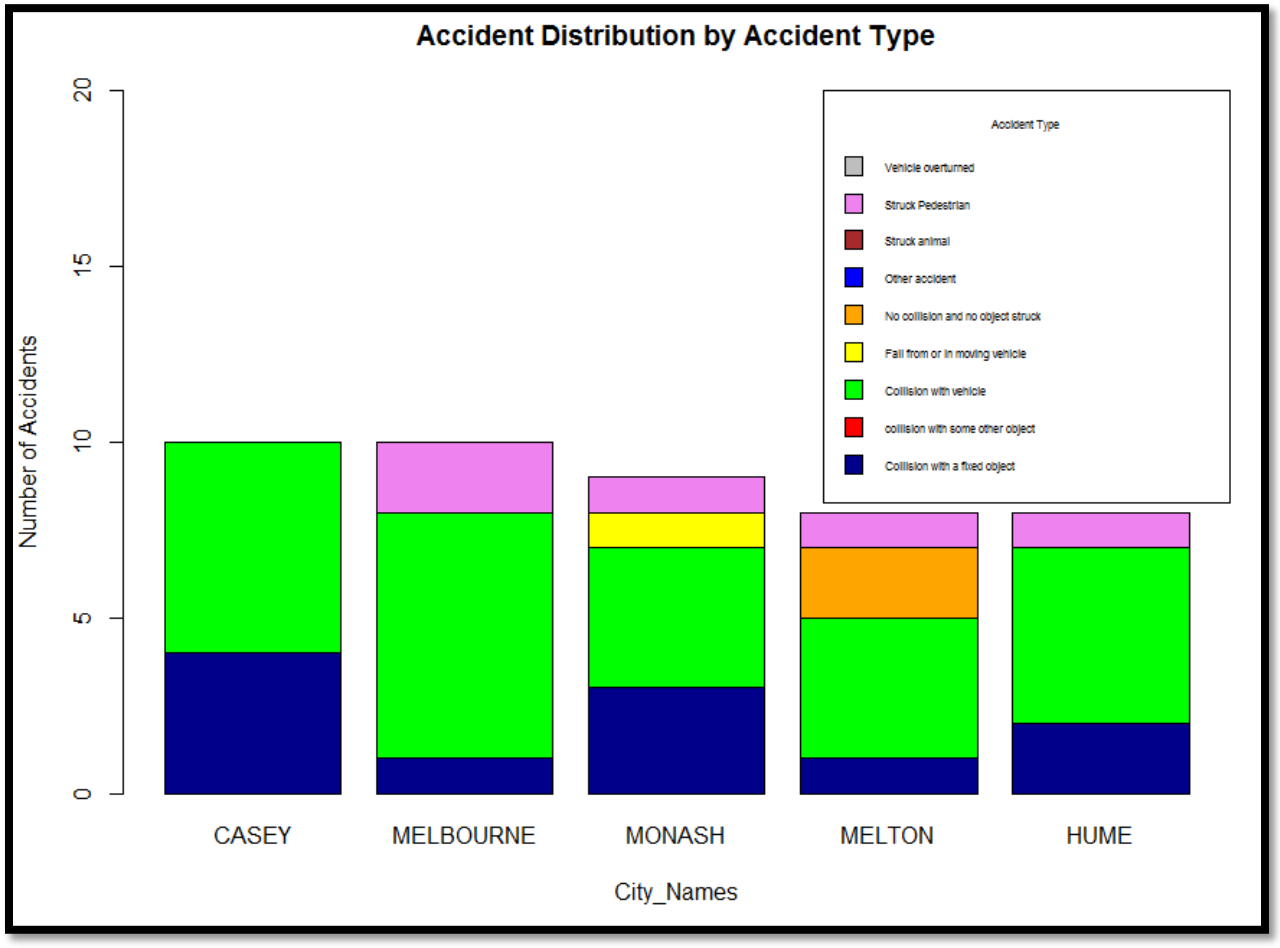


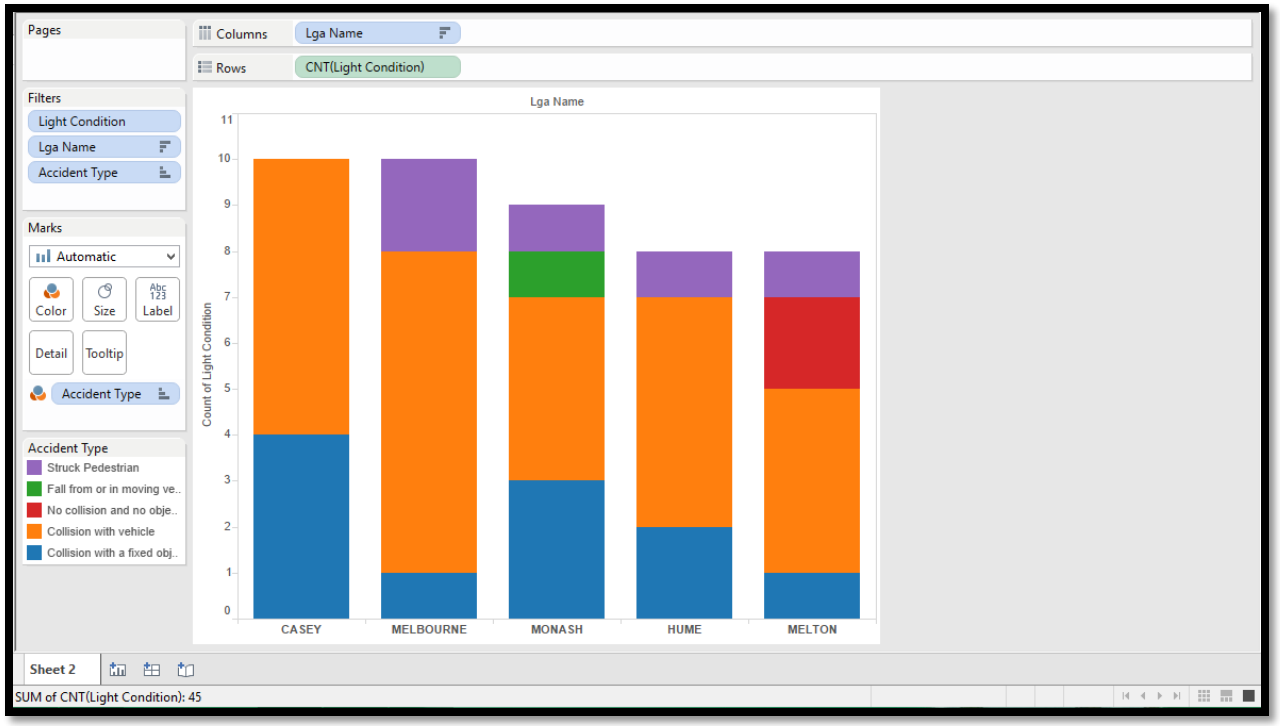


From the above graph it is clear that most number of accidents occur at speeds above 60km/hours so speed limit of 60 can introduced.

Accidents by Accident type

```
146 ###############################################################################
147 ##Accident Type
148 ###############################################################################
149 counts_3 <- table(topcity$ACCIDENT_TYPE, topcity$LGA_NAME);
150 counts_3
151 csel_counts_3 <- counts_3[,c("CASEY", "MELBOURNE","MONASH","MELTON","HUME")];
152 csel_counts_3
153
154 barplot(csel_counts_3, main="Accident Distribution by Accident Type",
155         xlab="City_Names",ylab = "Number of Accidents",
156         col = c("darkblue","red","green","yellow","orange","blue",
157                 "brown","violet","grey"),
158         legend = c("Collision with a fixed object","collision with some other object",
159                    "Collision with vehicle","Fall from or in moving vehicle",
160                    "No collision and no object struck",
161                    "Other accident","Struck animal","Struck Pedestrian","Vehicle overturned"),
162         args.legend = list(title = "Accident Type", x = "topright", cex = .5),ylim = c(0, 20))
146:83  (Top Level)                                                    R Script
```





In all cities most of the accident occur of type collision with vehicle.

Case 2: Accidents occurring due to Dark No street lights

Sub setting according to 'Dark No street lights'

```
177 ################################################################################
178 ##Sub setting data according to light type equal to dark street light off
179 ################################################################################
180 x.sub1 <- subset(myDataFrame, LIGHT_CONDITION == "Dark No street lights")
181 View(x.sub1)
182
183
```

Records reduced to 4143

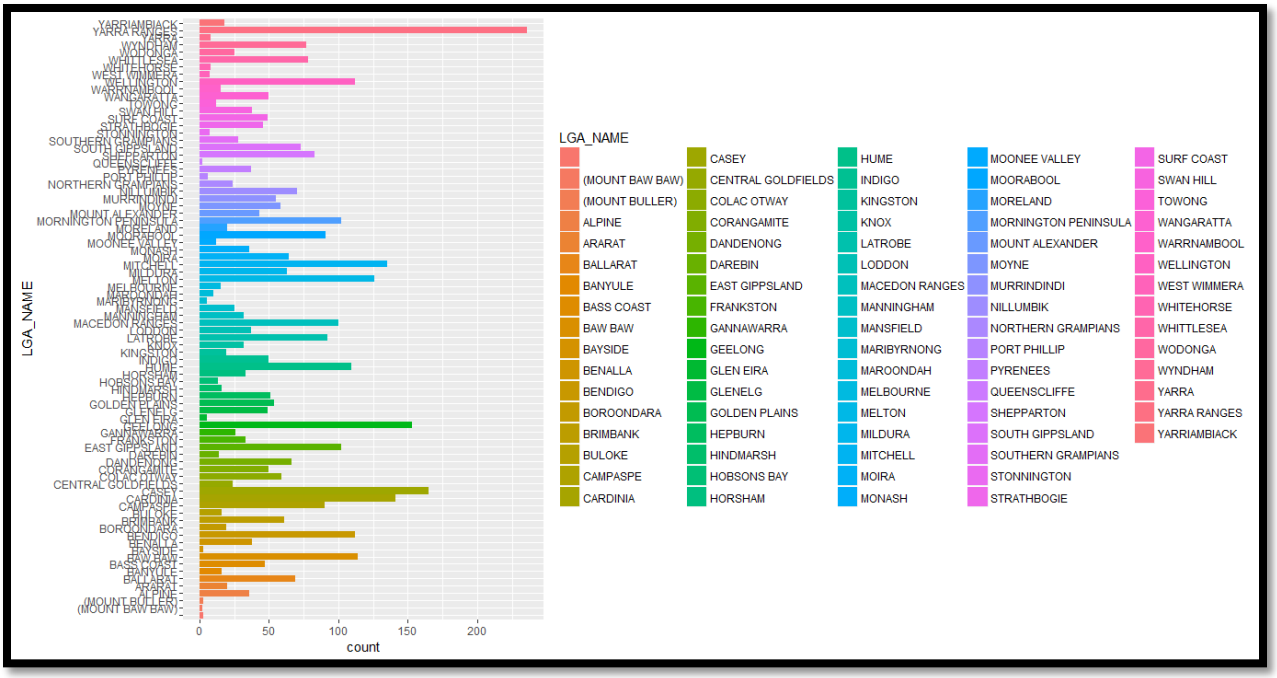
Number_1.R* | x.sub1 | Assignment_1.R | Assignment_final.R | Assignement.R | topcity

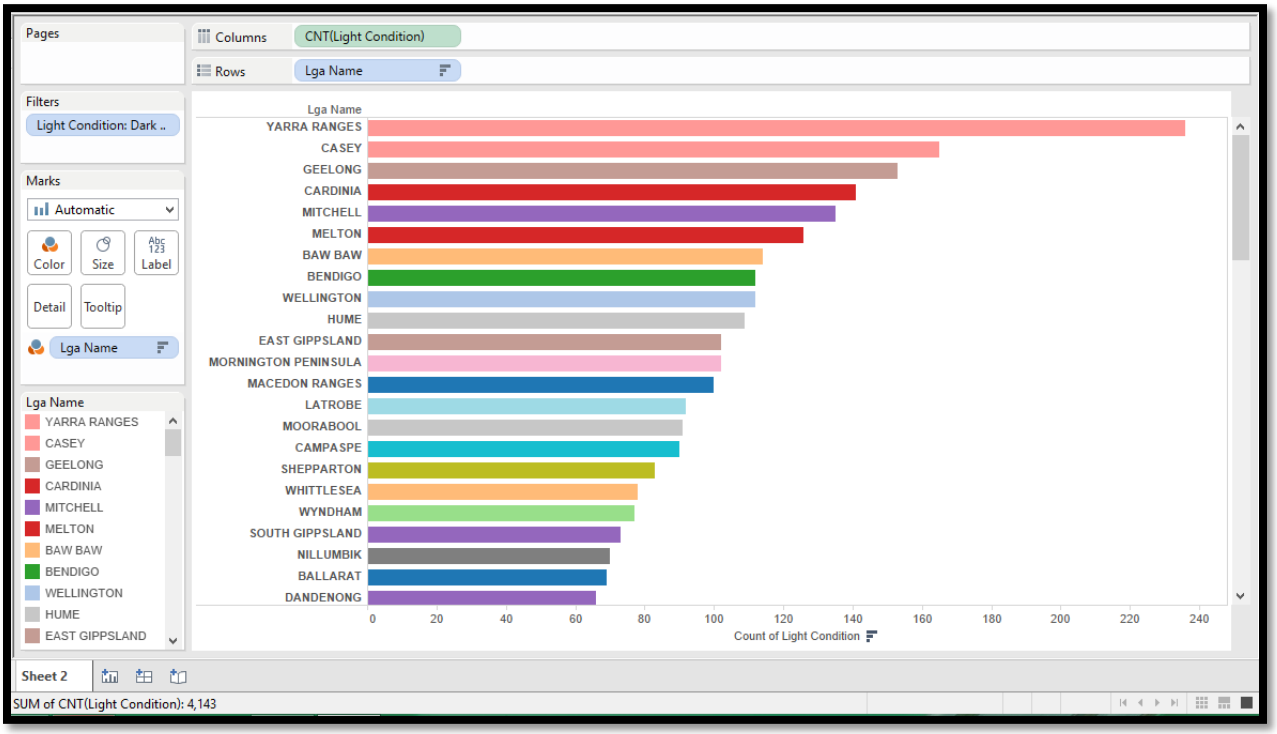
Filter

	X_id	X	Y	OBJECTID	ACCIDENT_NO	ABS_CODE	ACCIDENT_STATUS	ACCIDEN
71	57d393674f9b0b49364ccb62	144.2911	-38.21532	227992	T20120022169	ABS to receive accident	Finished	29/9/
111	57d393674f9b0b49364ccb8a	141.7887	-36.36090	228031	T20120022550	ABS to receive accident	Finished	5/10/
122	57d393674f9b0b49364ccb95	145.2781	-38.06937	228042	T20120022607	ABS to receive accident	Finished	6/10/
142	57d393674f9b0b49364ccba9	144.6534	-38.16824	228061	T20120022669	ABS to receive accident	Finished	6/10/
164	57d393674f9b0b49364ccbbf	141.5206	-37.80797	228084	T20120024738	ABS to receive accident	Finished	2/11/
193	57d393674f9b0b49364ccbdc	147.2191	-36.45622	228113	T20120023440	ABS to receive accident	Finished	16/10,
219	57d393674f9b0b49364ccbf6	144.0268	-37.69653	228139	T20120024785	ABS to receive accident	Finished	2/11/
239	57d393674f9b0b49364ccc0a	146.8439	-36.11486	228159	T20120023480	ABS to receive accident	Finished	25/8/
294	57d393674f9b0b49364ccc41	144.2956	-38.15159	228208	T20120022894	ABS to receive accident	Finished	10/10,
307	57d393674f9b0b49364ccc4e	145.3623	-37.69821	228227	T20120024270	ABS to receive accident	Finished	27/10,
325	57d393674f9b0b49364ccc60	146.3689	-36.35299	228245	T20120025031	ABS to receive accident	Finished	6/11/

Showing 7 to 19 of 4,143 entries

```
184 ################################################################################
185 ##Classifying accident count according to suburb
186 ################################################################################
187 library(ggplot2)
188 ggplot(data = x.sub1, aes(x = LGA_NAME,fill = LGA_NAME)) + geom_bar() + coord_flip()
189
190
```





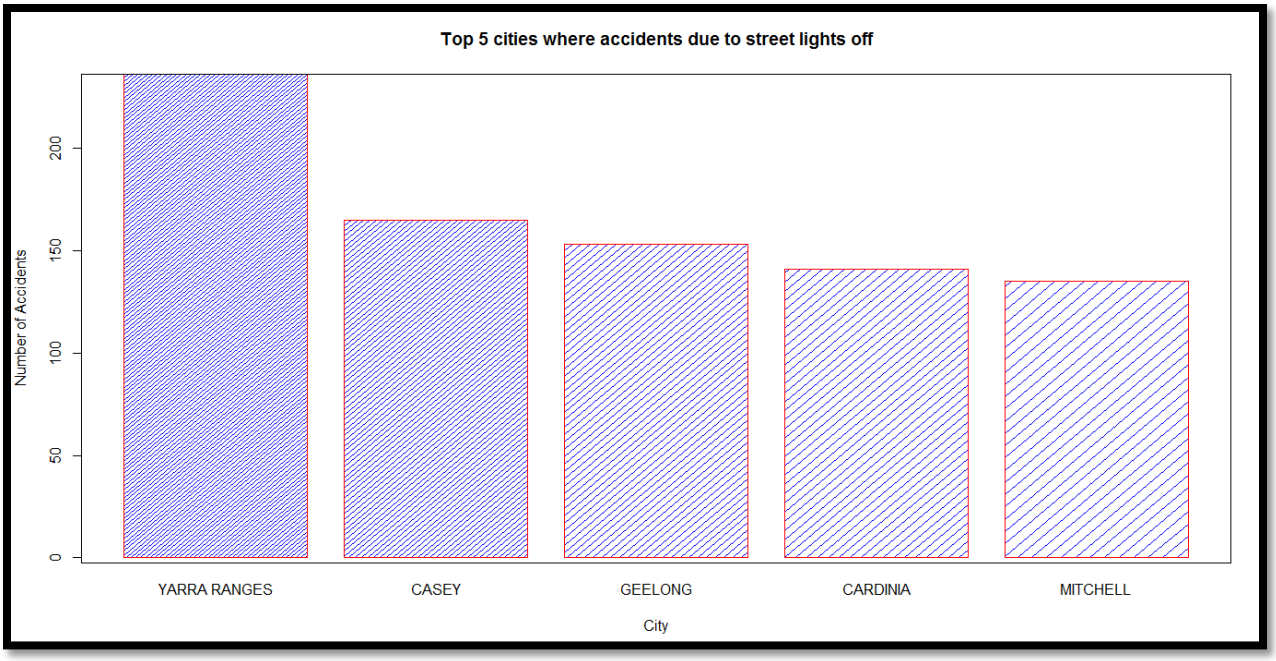
The above graph shows number of accidents occurring in every suburb but we cannot implement street lights everywhere so considering financial constraint choosing top 5 cities.

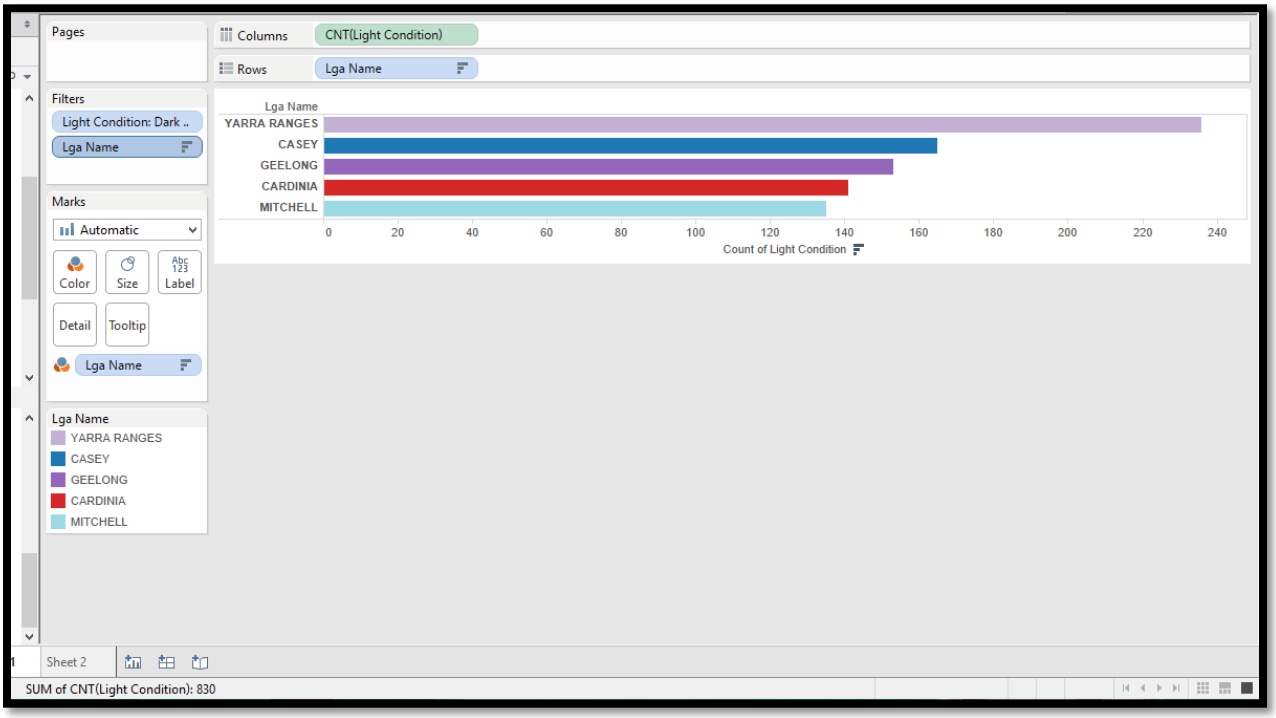
```
190
191 ###############################################################################
192 ##Finding top 5 suburbs
193 ###############################################################################
194 counts_11 <- table(x.sub1$LGA_NAME)
195 summary(x.sub1$LGA_NAME)
196 write.table(counts,"C:/Users/JATESH/Desktop/Dataset/New folder/export_table_1.txt", row.names=FALS
197 barplot(sort(counts,decreasing=TRUE)[1:5],
198         main="Top 5 cities where accidents due to street lights off", col="blue", xlab="City",
199         ylab="Number of Accidents",
200         border="red", density=c(90, 70, 50, 40, 30))
201 box()
202
```

```
> summary(x.sub1$LGA_NAME)
                  (FALLS CREEK)       (LAKE MOUNTAIN)       (MOUNT BAW BAW)
              3               0                     0                     2
(MOUNT BULLER)   (MOUNT HOTHAM)              ALPINE               ARARAT
              3               0                    36                    20
       BALLARAT         BANYULE           BASS COAST              BAW BAW
             69              16                    47                   114
        BAYSIDE         BENALLA             BENDIGO           BOROONDARA
              3              38                   112                    19
       BRIMBANK         BULOKE           CAMPASPE             CARDINIA
             61              16                    90                   141
          CASEY CENTRAL GOLDFIELDS         COLAC OTWAY           CORANGAMITE
            165              24                    59                    50
     DANDENONG         DAREBIN      EAST GIPPSLAND            FRANKSTON
             66              14                   102                    33
     GANNAWARRA        GEELONG          GLEN EIRA              GLENELG
             26             153                     5                    49
  GOLDEN PLAINS         HEPBURN          HINDMARSH          HOBSONS BAY
             54              51                    16                    13
        HORSHAM         HUME             INDIGO             KINGSTON
             33             109                    50                    19
          KNOX         LATROBE            LODDON          MACEDON RANGES
             32              92                    37                   100
     MANNINGHAM        MANSFIELD        MARIBYRNONG             MAROONDAH
             32              25                     5                    10
```

Txt file

```
"Var1" "Freq"
"" 3
"(FALLS CREEK)" 0
"(LAKE MOUNTAIN)" 0
"(MOUNT BAW BAW)" 2
"(MOUNT BULLER)" 3
"(MOUNT HOTHAM)" 0
"ALPINE" 36
"ARARAT" 20
"BALLARAT" 69
"BANYULE" 16
"BASS COAST" 47
"BAW BAW" 114
"BAYSIDE" 3
"BENALLA" 38
"BENDIGO" 112
"BOROONDARA" 19
"BRIMBANK" 61
"BULOKE" 16
"CAMPASPE" 90
"CARDINIA" 141
"CASEY" 165
"CENTRAL GOLDFIELDS" 24
"COLAC OTWAY" 59
"CORANGAMITE" 50
"DANDENONG" 66
"DAREBIN" 14
"EAST GIPPSLAND" 102
"FRANKSTON" 33
"GANNAWARRA" 26
"GEELONG" 153
"GLEN EIRA" 5
"GLENELG" 49
"GOLDEN PLAINS" 54
"HEPBURN" 51
"HINDMARSH" 16
"HOBSONS BAY" 13
```

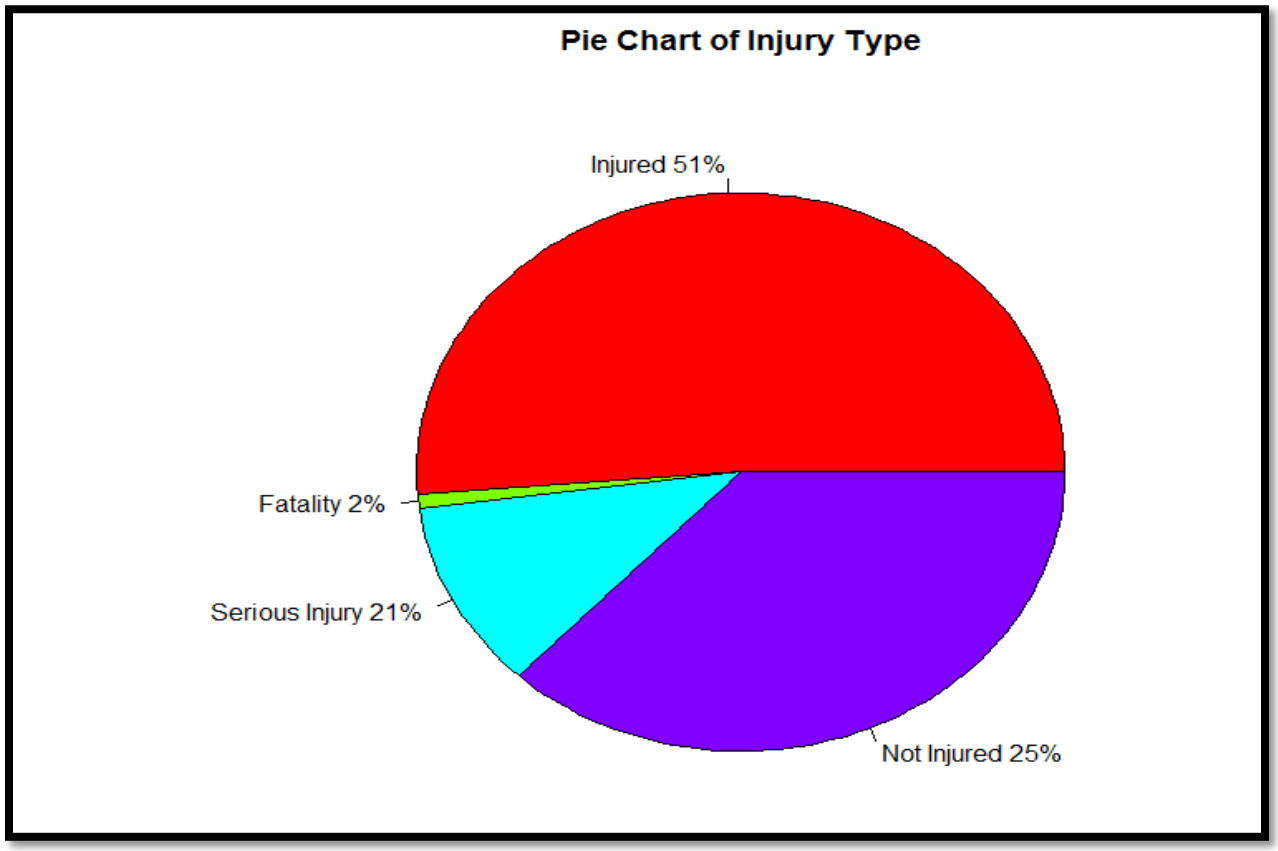

Top 5 cities where accidents due to street lights off
Number of Accidents
0
50
100
150
200
YARRA RANGES
CASEY
GEELONG
CARDINIA
MITCHELL
City

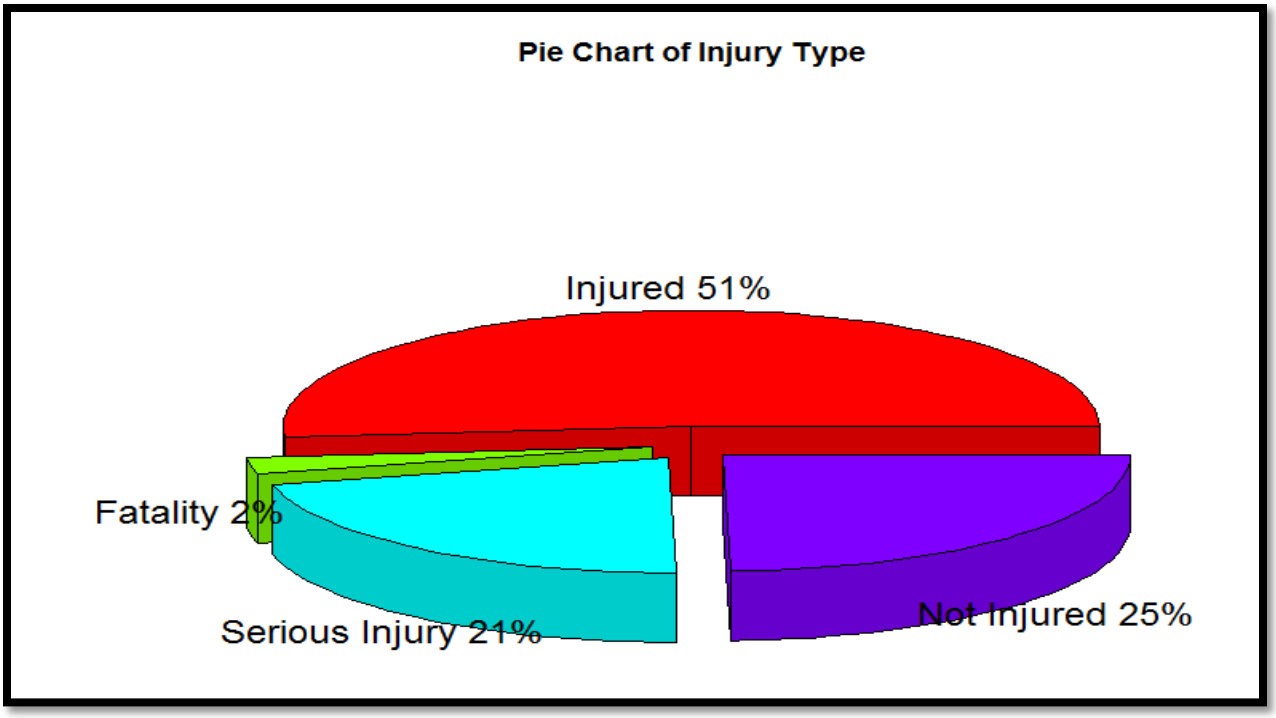

Pages
Columns
CNT(Light Condition)
Rows
Lga Name
Filters
Light Condition: Dark ..
Lga Name
Marks
Automatic
Color
Size
Label
Detail
Tooltip
Lga Name
Lga Name
YARRA RANGES
CASEY
GEELONG
CARDINIA
MITCHELL
0
20
40
60
80
100
120
140
160
180
200
220
240
Count of Light Condition
Sheet 2
SUM of CNT(Light Condition): 830

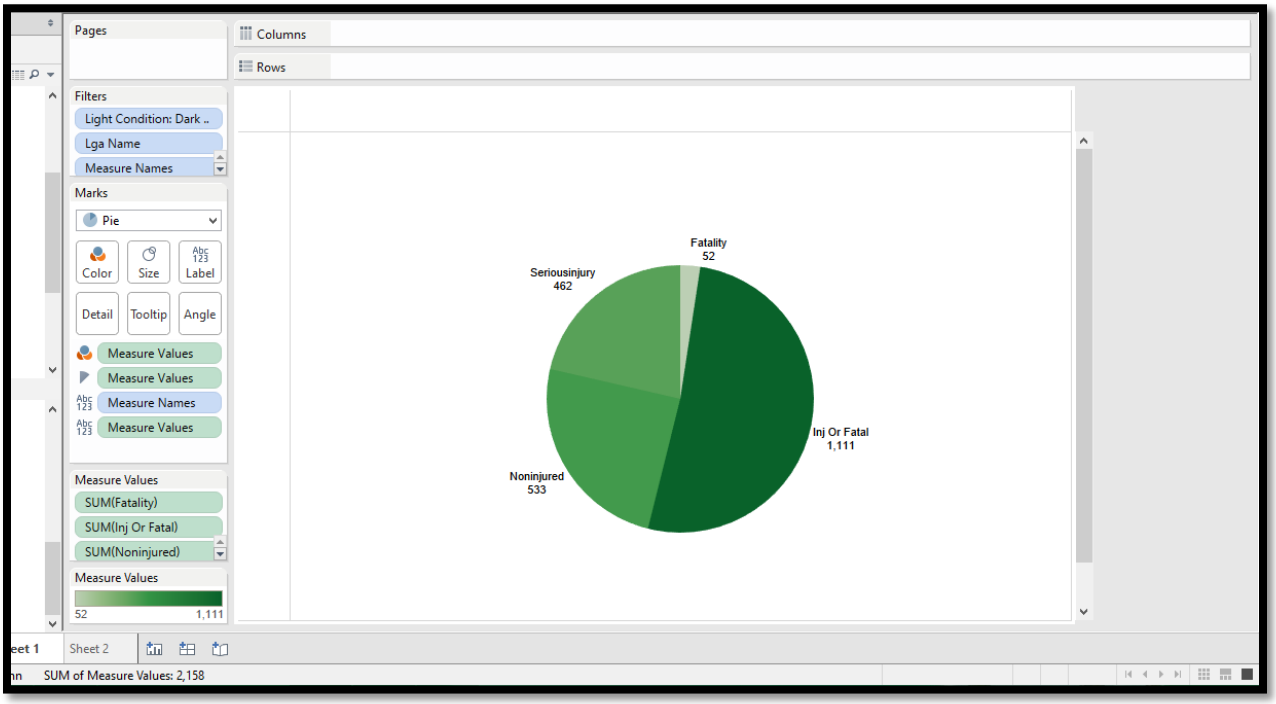
Demographics for top 5 cities

Demographics by Injury Type

```
203 ################################################################################
204 ##Demographics of Injury type in top 5 cities
205 ################################################################################
206 topcity1 <-subset(x.sub1, LGA_NAME == "YARRA RANGES" | LGA_NAME == "CASEY" | LGA_NAME == "GEELONG"
207                   LGA_NAME == "CARDINIA" | LGA_NAME == "MITCHELL")
208 view(topcity1)
209 slices_C2 <- c(sum(topcity1$INJ_OR_FATAL),sum(topcity1$FATALITY),sum(topcity1$SERIOUSINJURY),
210              sum(topcity1$NONINJURED))
211 slices_C2
212 lbls_C2 <- c("Injured", "Fatality","Serious Injury", "Not Injured")
213 pct_C2 <- round(slices_C2/sum(slices_C2)*100)
214 lbls_C2 <- paste(lbls_C2, pct_C2) # add percents to labels
215 lbls_C2 <- paste(lbls_C2,"%",sep="") # ad % to labels
216 pie(slices_C2,labels = lbls_C2, col=rainbow(length(lbls_C2)),
217     main="Pie Chart of Injury Type")
218 library(plotrix)
219 pie3D(slices_C2,labels=lbls_C2,explode=0.1,
220       main="Pie Chart of Injury Type")
221
222 ################################################################################
223
212:1  last                                                              R Script
```



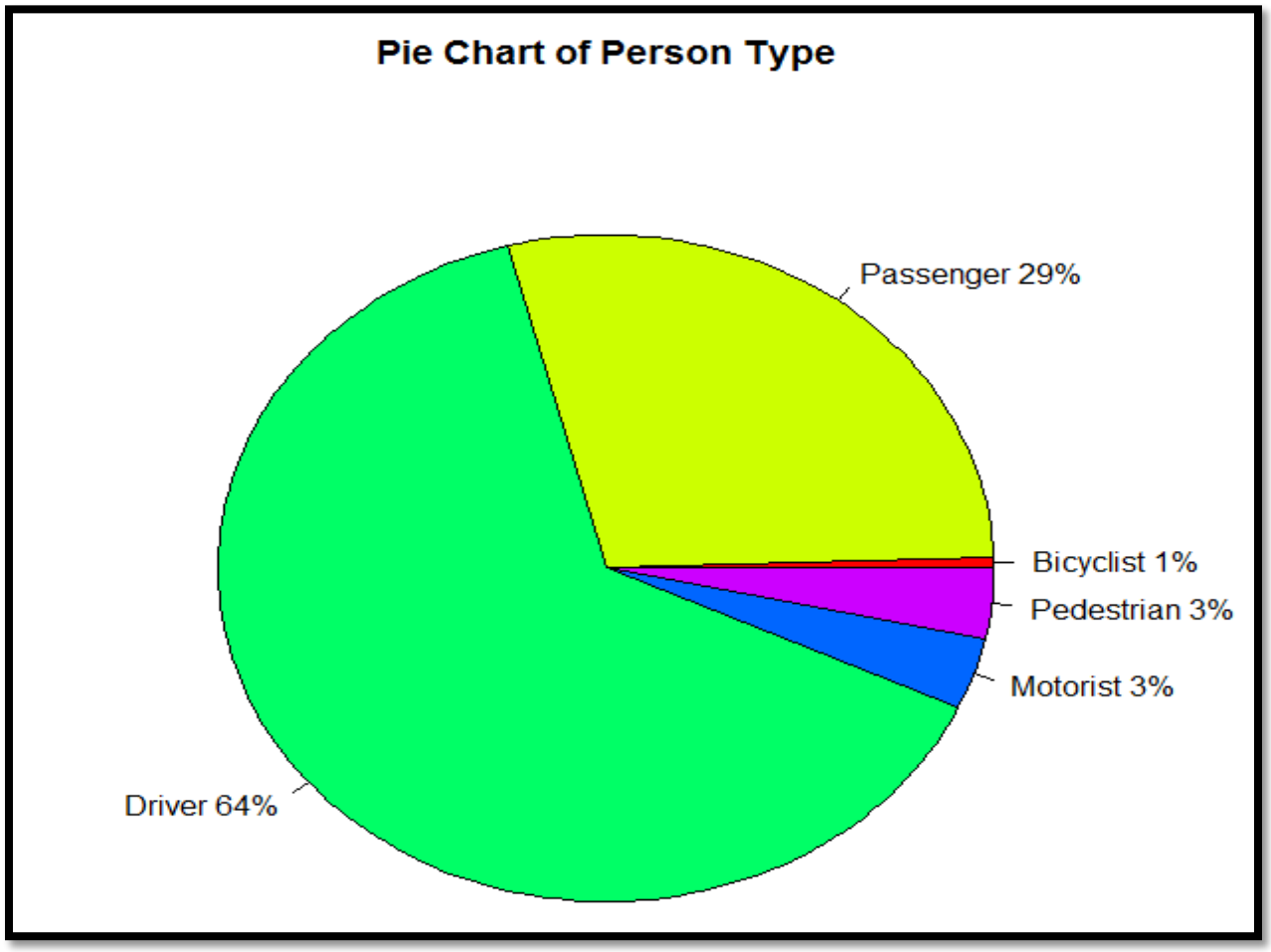


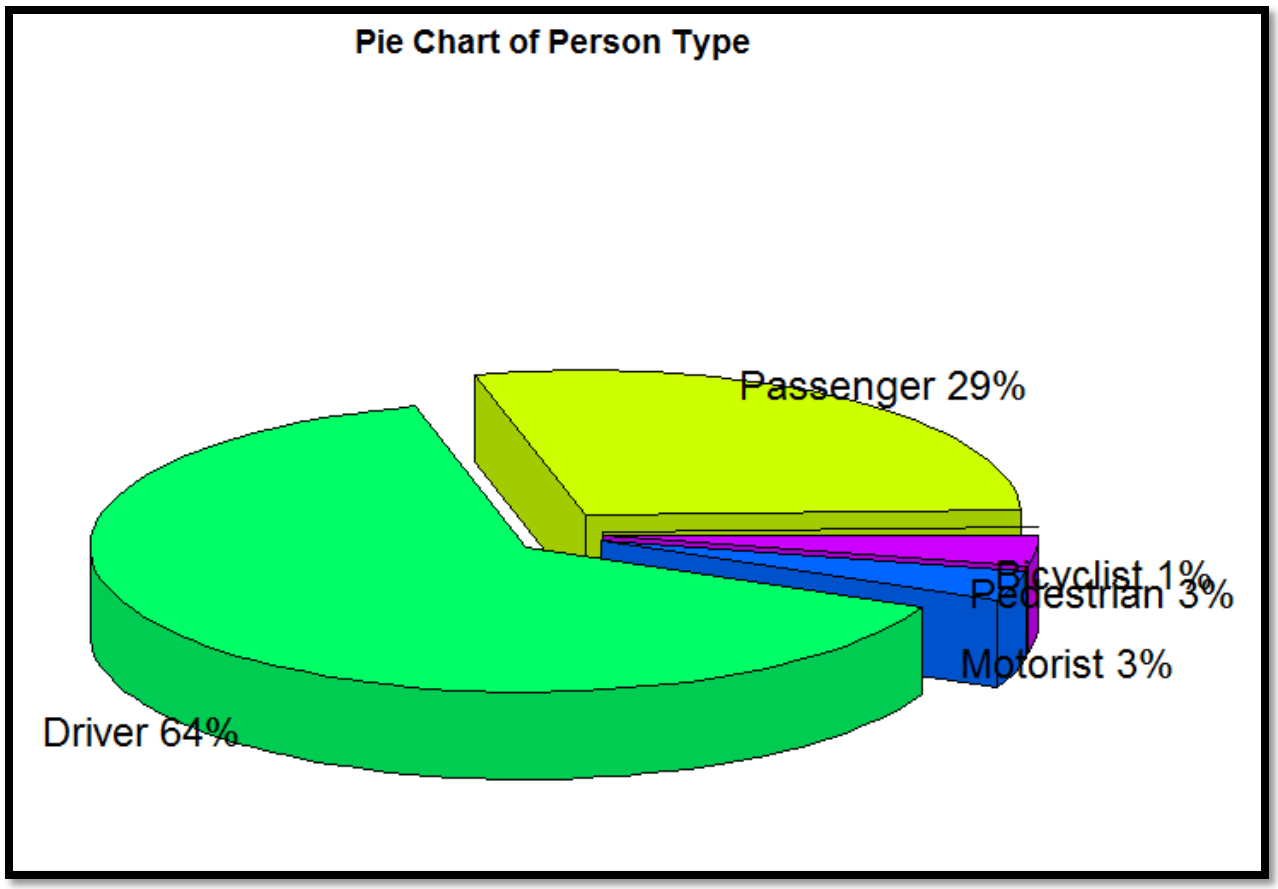


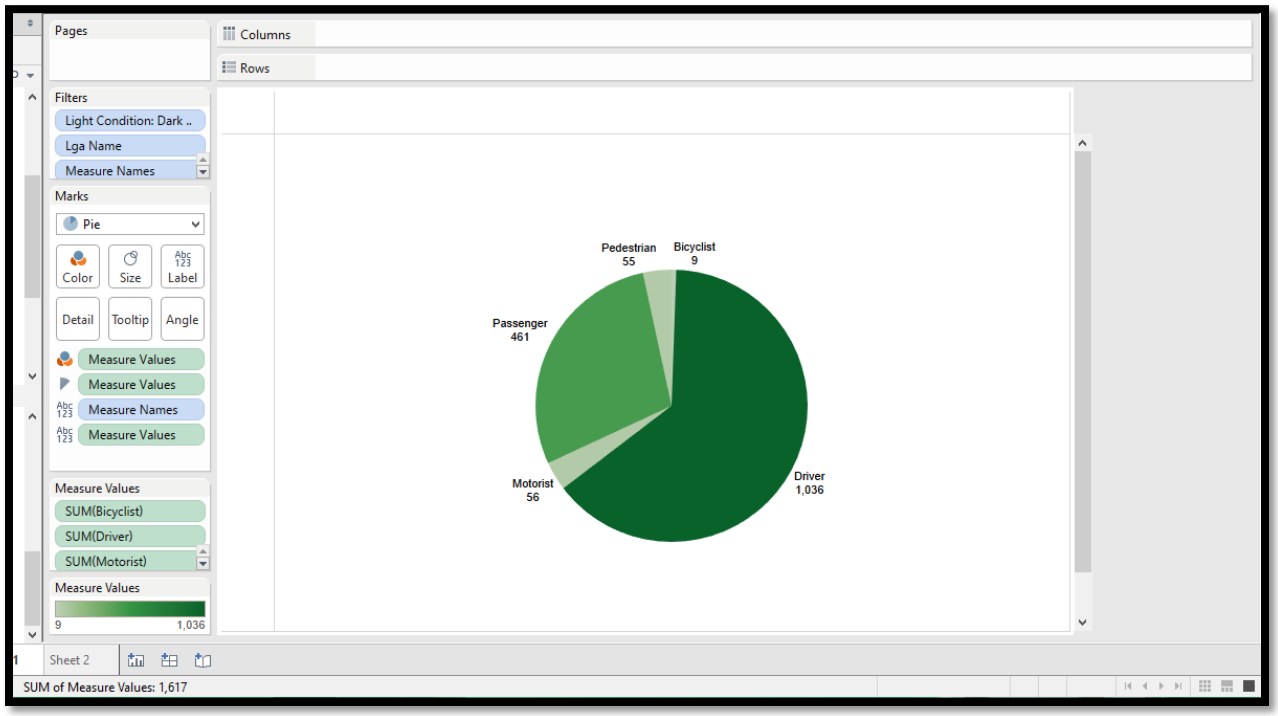
The above demographics suggest people getting injured is almost 75% so the demographics support the idea of implementing street lights.

Demographics of People Involved

```
222 ###############################################################################
223 ##Demographics of people type involved in accidents of top 5 cities
224 ###############################################################################
225 Slices_C21 <- c(sum(topcity1$BICYCLIST),sum(topcity1$PASSENGER),sum(topcity1$DRIVER),
226                sum(topcity1$MOTORIST),sum(topcity1$PEDESTRIAN))
227 Slices_C21
228 lbls_C21 <- c("Bicyclist", "Passenger","Driver", "Motorist", "Pedestrian")
229 pct_C21 <- round(Slices_C21/sum(Slices_C21)*100)
230 pct_C21
231 lbls_C21 <- paste(lbls_C21, pct_C21) # add percents to labels
232 lbls_C21 <- paste(lbls_C21,"%",sep="") # ad % to labels
233 pie(Slices_C21,labels = lbls_C21, col=rainbow(length(lbls_C21)),
234     main="Pie Chart of Person Type")
235 library(plotrix)
236 pie3D(Slices_C21,labels=lbls_C21,explode=0.1,
237       main="Pie Chart of Person Type ")
238
239
240
236:46  last                                                          R Script
```



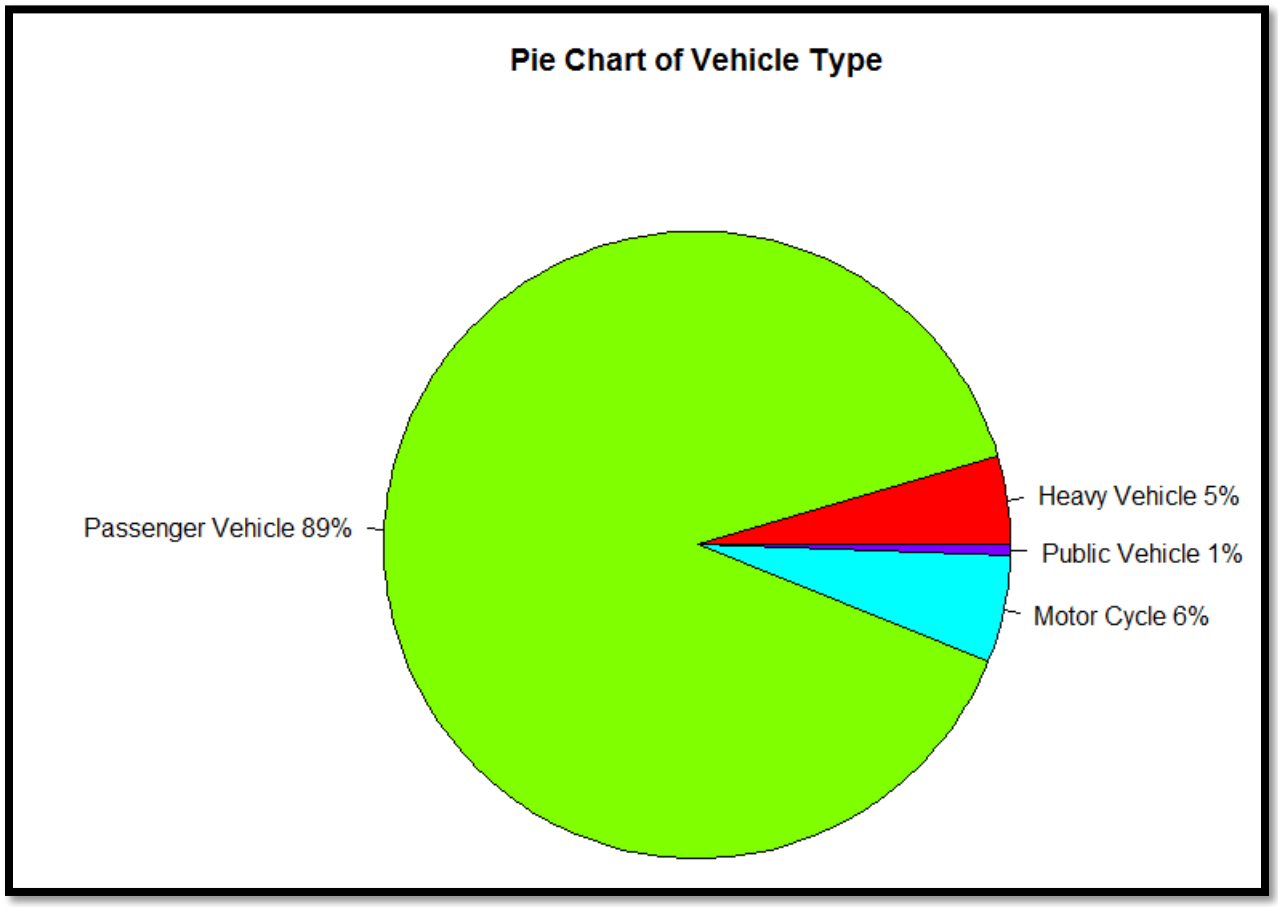


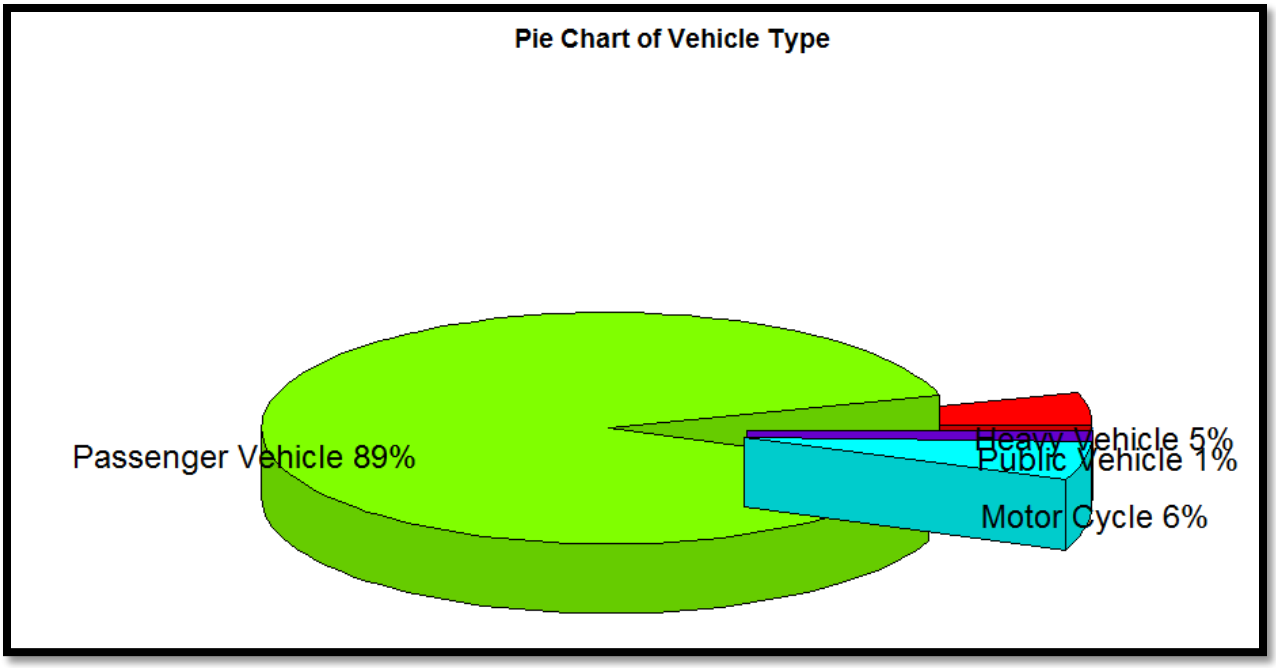


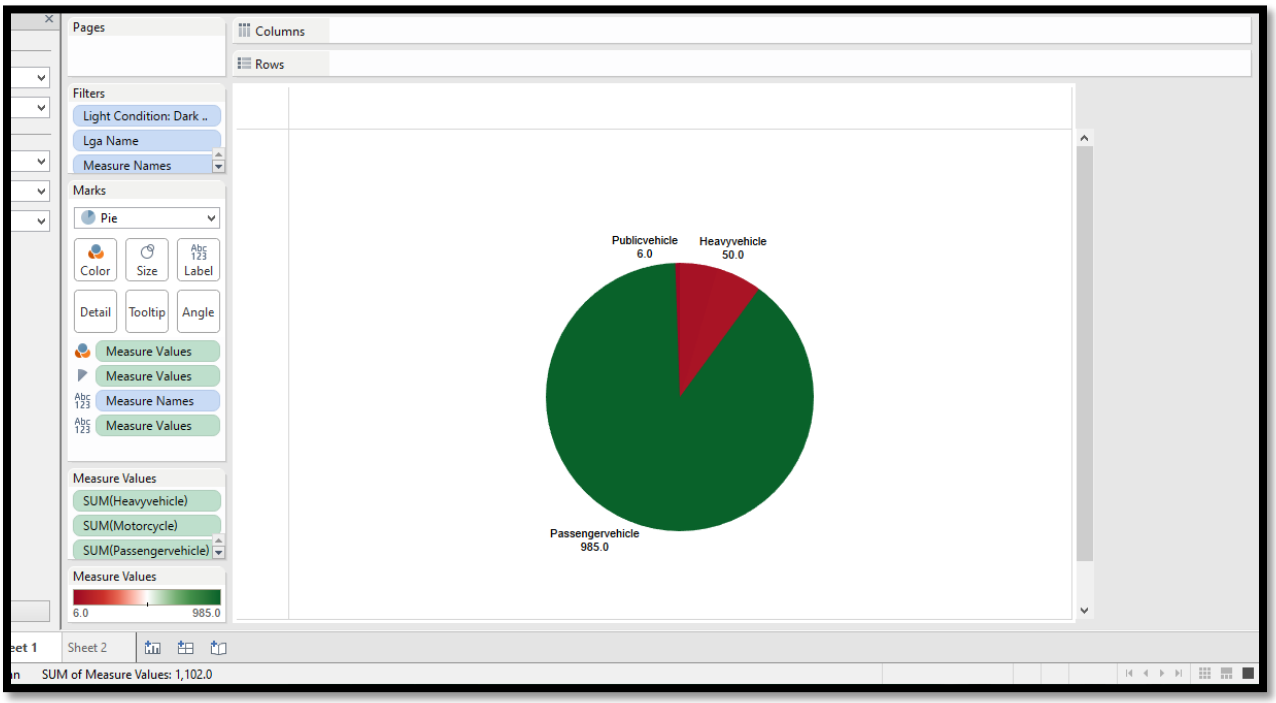
The above demographics suggest most of the car accidents took place

Demographics by Vehicle Type

```
240 ##############################################################################
241 ##Demographics of vehicle type involved in accidents of top 5 cities
242 ##############################################################################
243 slices_C21 <- c(sum(topcity1$HEAVYVEHICLE),sum(topcity1$PASSENGERVEHICLE),sum(topcity1$MOTORCYCLE)
244             sum(topcity1$PUBLICVEHICLE))
245 slices_C21
246 lbls_C22 <- c("Heavy Vehicle", "Passenger Vehicle","Motor Cycle", "Public Vehicle")
247 pct_C22 <- round(slices_C21/sum(slices_C21)*100)
248 pct_C22
249 lbls_C22 <- paste(lbls_C22, pct_C22) # add percents to labels
250 lbls_C22 <- paste(lbls_C22,"%",sep="") # ad % to labels
251 pie(slices_C21,labels = lbls_C22, col=rainbow(length(lbls_C22)),
252     main="Pie Chart of Vehicle Type")
253 library(plotrix)
254 pie3D(slices_C21,labels=lbls_C22,explode=0.2,
255       main="Pie Chart of Vehicle Type ")
```



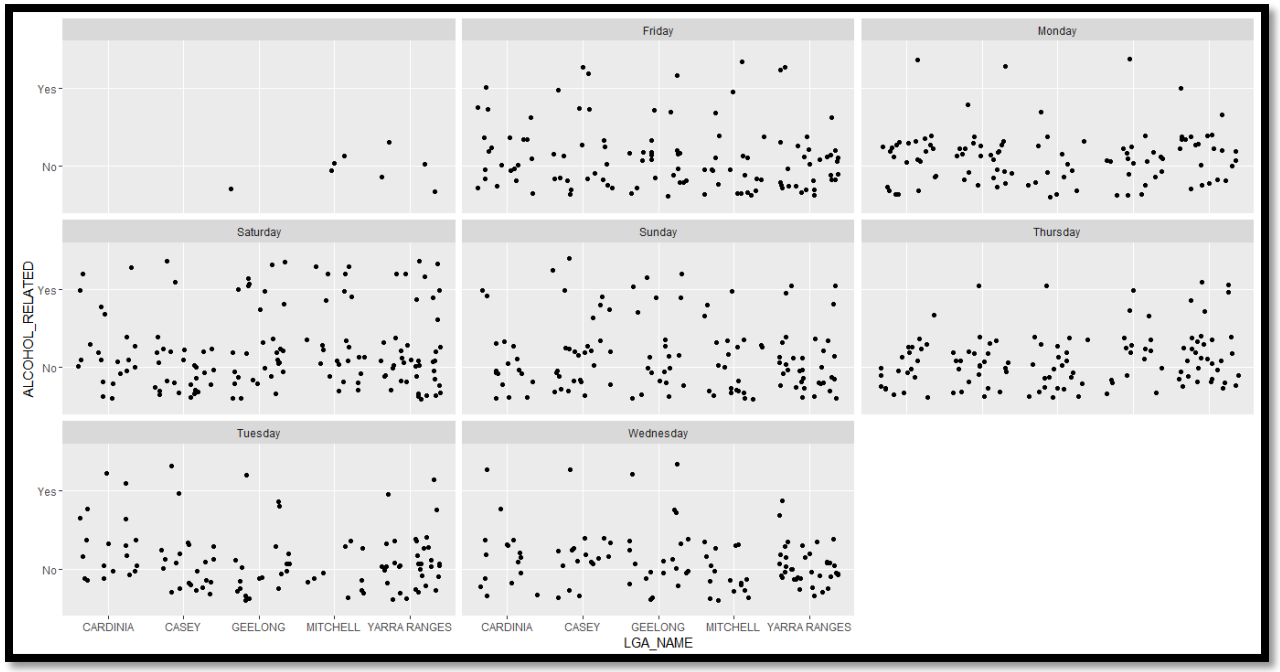




The above demographics confirms most of the car accidents taking place.

Impact of Alcohol

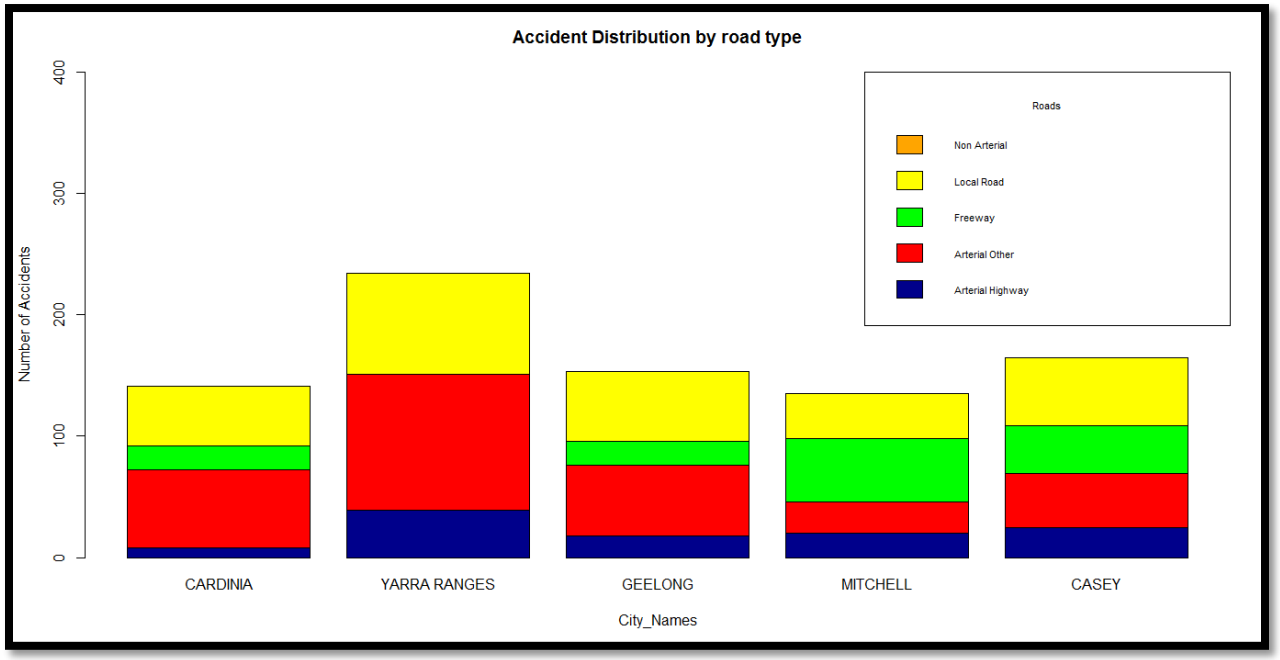
```
256 ###############################################################################
257 ##Alcohol Involvement
258 ###############################################################################
259
260 ggplot(topcity1, aes(LGA_NAME, ALCOHOL_RELATED)) + geom_jitter()+ facet_wrap(~ DAY_OF_WEEK)
261 ggplot(topcity1, aes(LGA_NAME, ALCOHOL_RELATED)) + geom_point()+ facet_wrap(~ DAY_OF_WEEK)
262
263
```

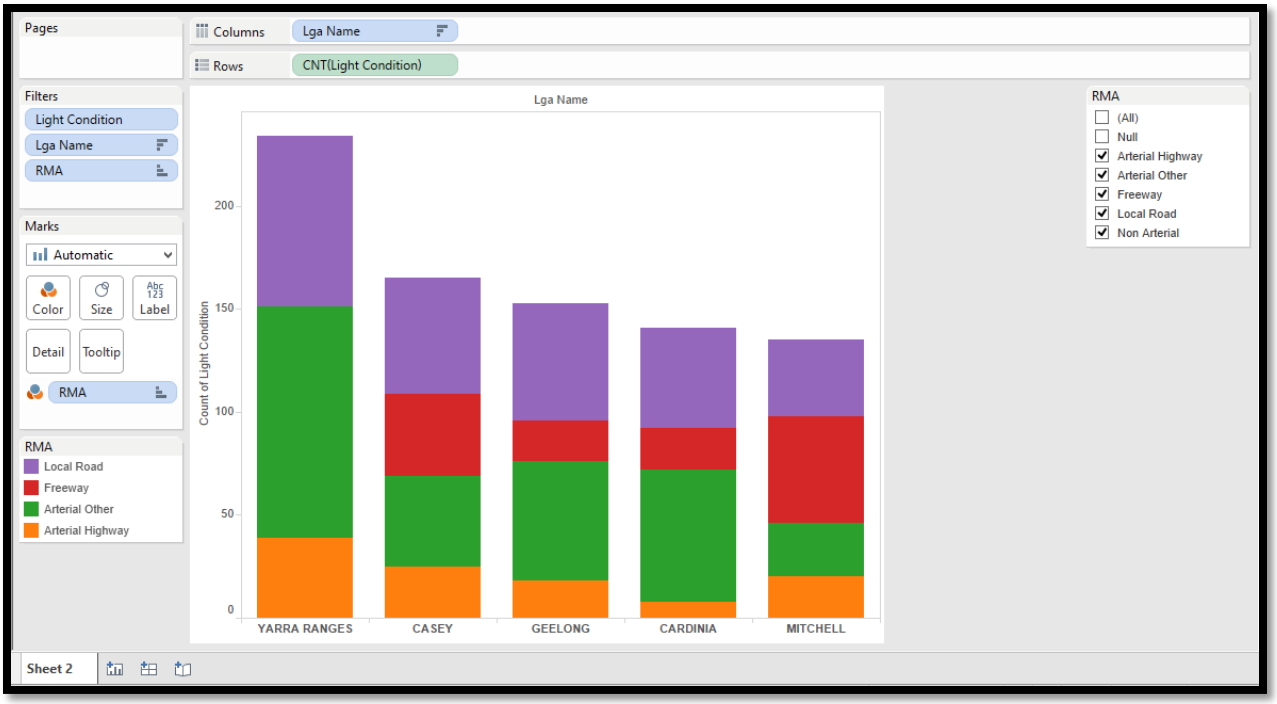


The above graphs suggest accidents occurring sue to alcohol consumption so we can set up alcohol checking points on Friday, Saturday and Sunday in cities to create fear among people to stop consuming alcohol and driving.

Accidents by Road Type

```
264 ###############################################################################
265 ##Road Types
266 ###############################################################################
267
268 counts_11 <- table(topcity1$RMA, topcity1$LGA_NAME)
269 counts_11
270 csel_counts_11 <- counts_11[,c("CARDINIA", "YARRA RANGES","GEELONG","MITCHELL","CASEY")];
271 csel_counts_11
272 rsel_counts_11 <- csel_counts_11[c("Arterial Highway","Arterial Other","Freeway",
273                                    "Local Road","Non Arterial"),];
274 rsel_counts_11
275
276 barplot(rsel_counts_11, main="Accident Distribution by road type",
277         xlab="City_Names",ylab = "Number of Accidents", col = c("darkblue","red","green","yellow",
278         legend = c("Arterial Highway", "Arterial Other", "Freeway","Local Road","Non Arterial"),
279         args.legend = list(title = "Roads", x = "topright", cex = .7),ylim = c(0, 400))
280
```

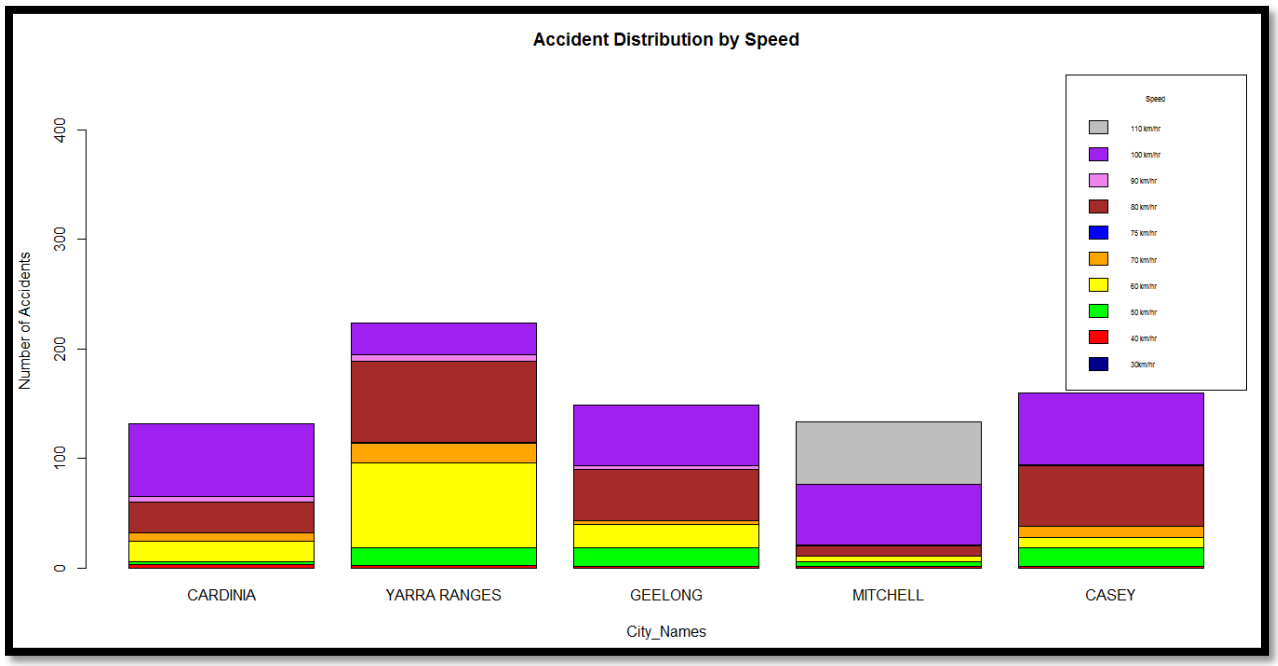


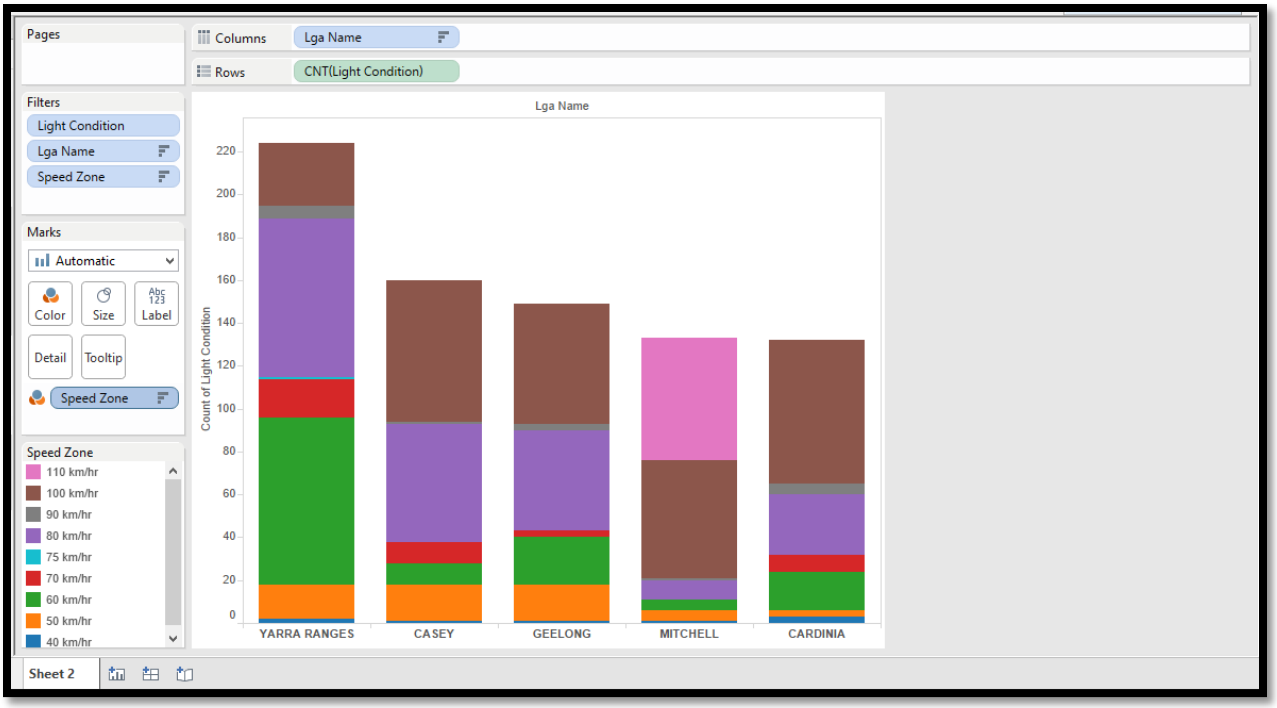


The above graph suggests most of the accidents occurring on local roads and arterial other.

Accidents by Speed

```
282 ###############################################################################
283 ##Speed Analysis
284 ###############################################################################
285 counts_22 <- table(topcity1$SPEED_ZONE, topcity1$LGA_NAME);
286 counts_22
287 csel_counts_22 <- counts_22[,c("CARDINIA", "YARRA RANGES","GEELONG","MITCHELL","CASEY")];
288 csel_counts_22
289 rsel_counts_22 <- csel_counts_22[c("30km/hr","40 km/hr","50 km/hr","60 km/hr","70 km/hr",
290                                    "75 km/hr","80 km/hr","90 km/hr","100 km/hr","110 km/hr"),];
291 rsel_counts_22
292 barplot(rsel_counts_22, main="Accident Distribution by Speed",
293         xlab="City_Names",ylab = "Number of Accidents",
294         col = c("darkblue","red","green","yellow","orange","blue",
295                 "brown","violet","purple","grey"),
296         legend = c("30km/hr","40 km/hr","50 km/hr","60 km/hr","70 km/hr",
297                    "75 km/hr","80 km/hr","90 km/hr","100 km/hr","110 km/hr"),
298         args.legend = list(title = "Speed", x = "topright", cex = .5),ylim = c(0, 450))
299
```

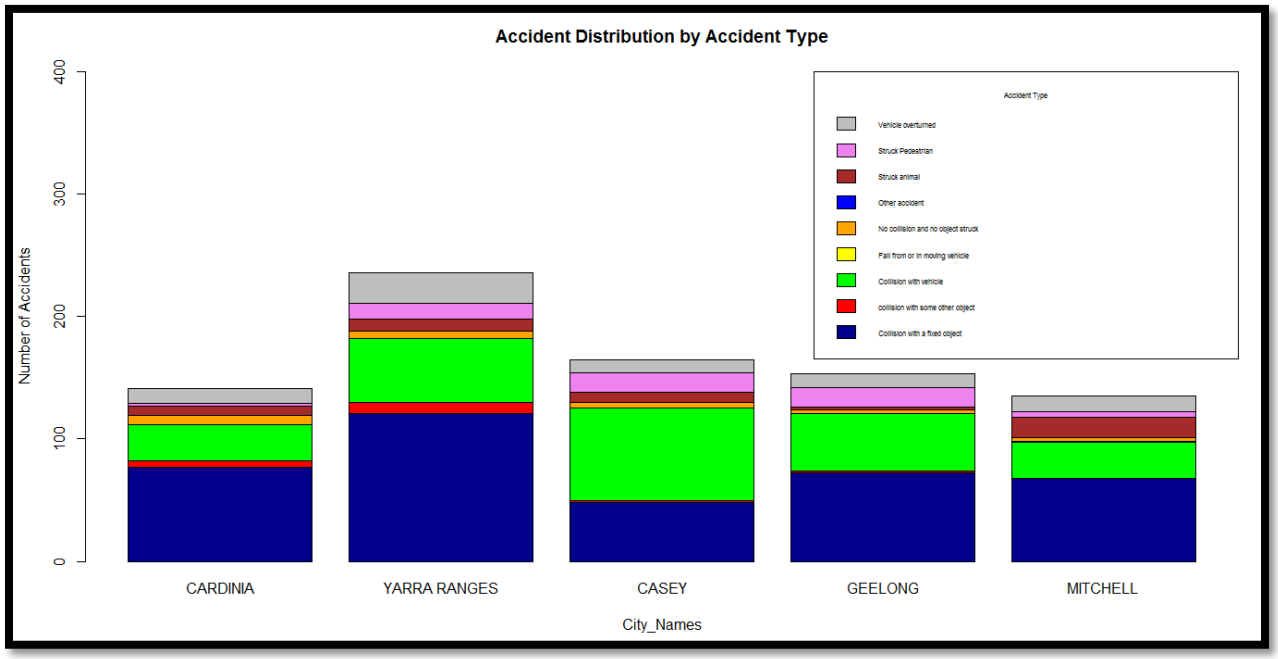


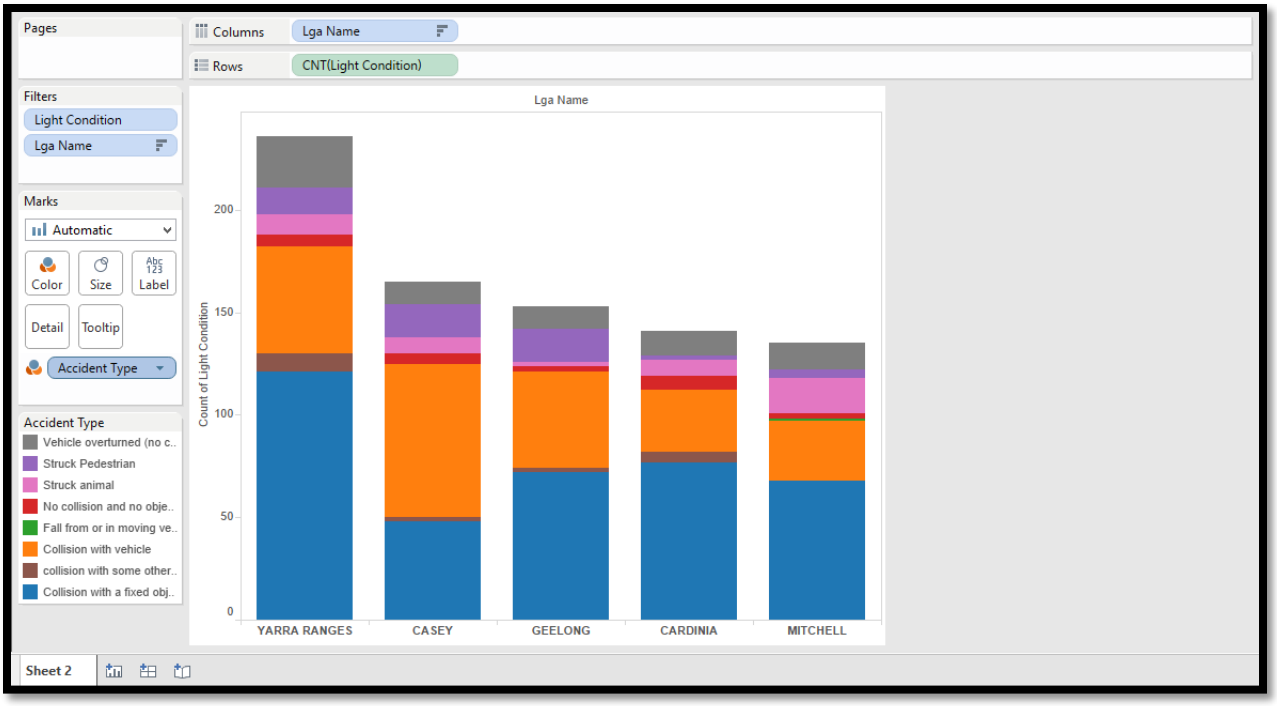


The above graph concludes most of the accidents occurred at speeds above 60km per hour so speed limits of 60 km per hours together.

Accidents by Accident type

```
300  ##############################################################################
301  ##Accident Type
302  ##############################################################################
303  counts_33 <- table(topcity1$ACCIDENT_TYPE, topcity1$LGA_NAME);
304  counts_33
305  csel_counts_33 <- counts_33[,c("CARDINIA", "YARRA RANGES","CASEY","GEELONG","MITCHELL")];
306  csel_counts_33
307
308  barplot(csel_counts_33, main="Accident Distribution by Accident Type",
309          xlab="City_Names",ylab = "Number of Accidents",
310          col = c("darkblue","red","green","yellow","orange","blue",
311                  "brown","violet","grey"),
312          legend = c("Collision with a fixed object","collision with some other object",
313                     "Collision with vehicle","Fall from or in moving vehicle",
314                     "No collision and no object struck",
315                     "Other accident","Struck animal","Struck Pedestrian","Vehicle overturned"),
316          args.legend = list(title = "Accident Type", x = "topright", cex = .5),ylim = c(0, 400))
300:83  last                                                                  R Script
```

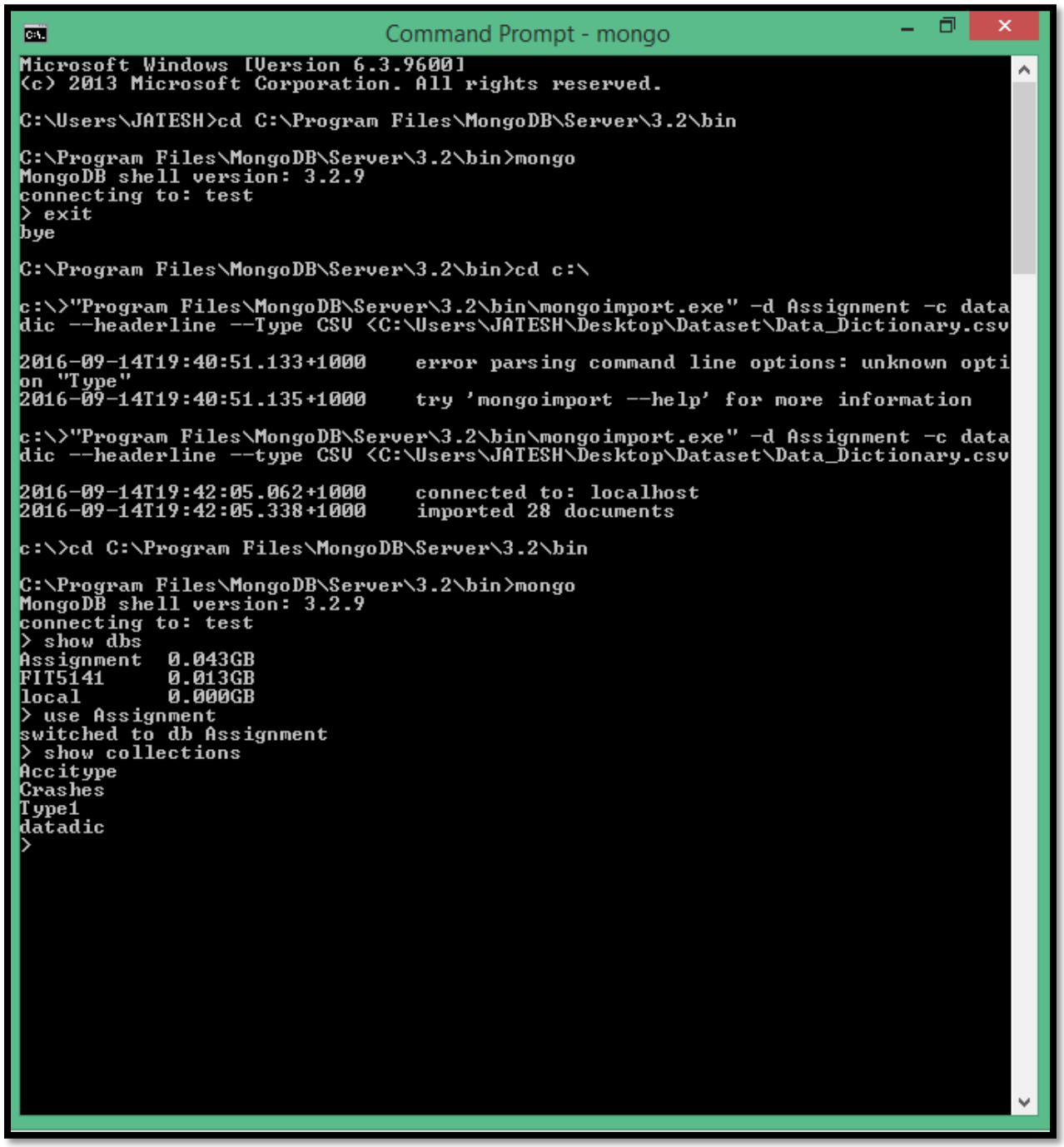




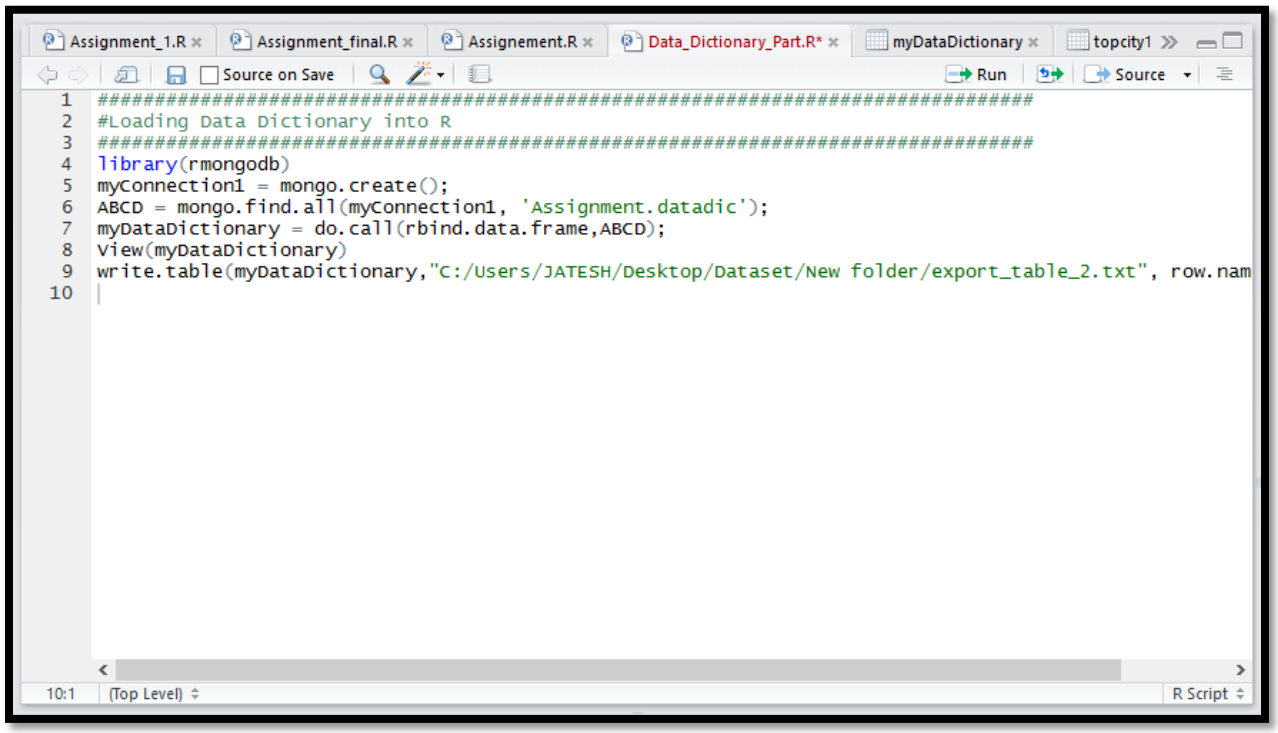
The accident type was collision with a fixed object or collision with vehicle most of the time.

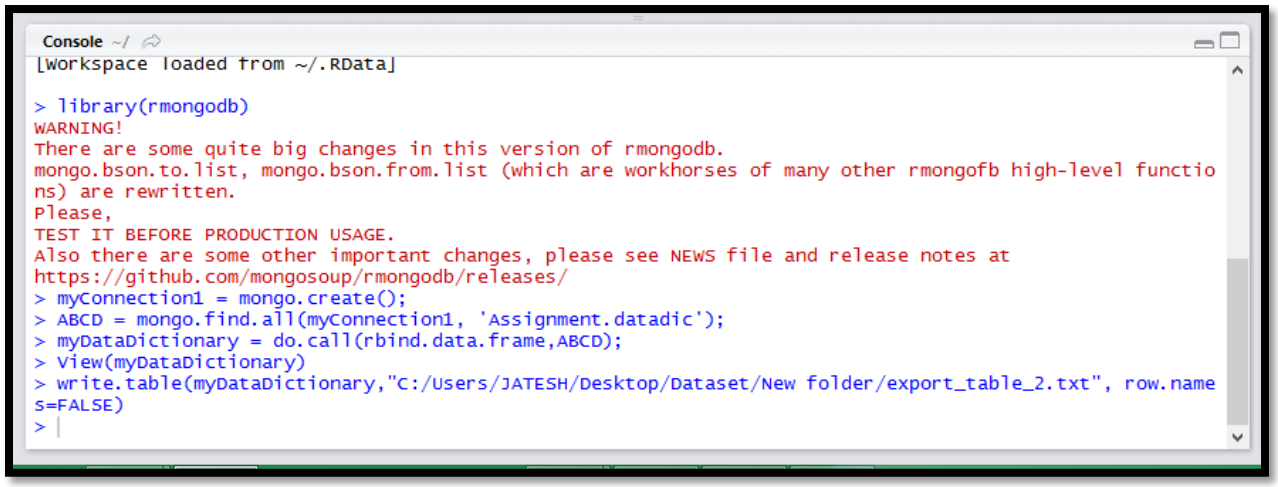
Data Dictionary

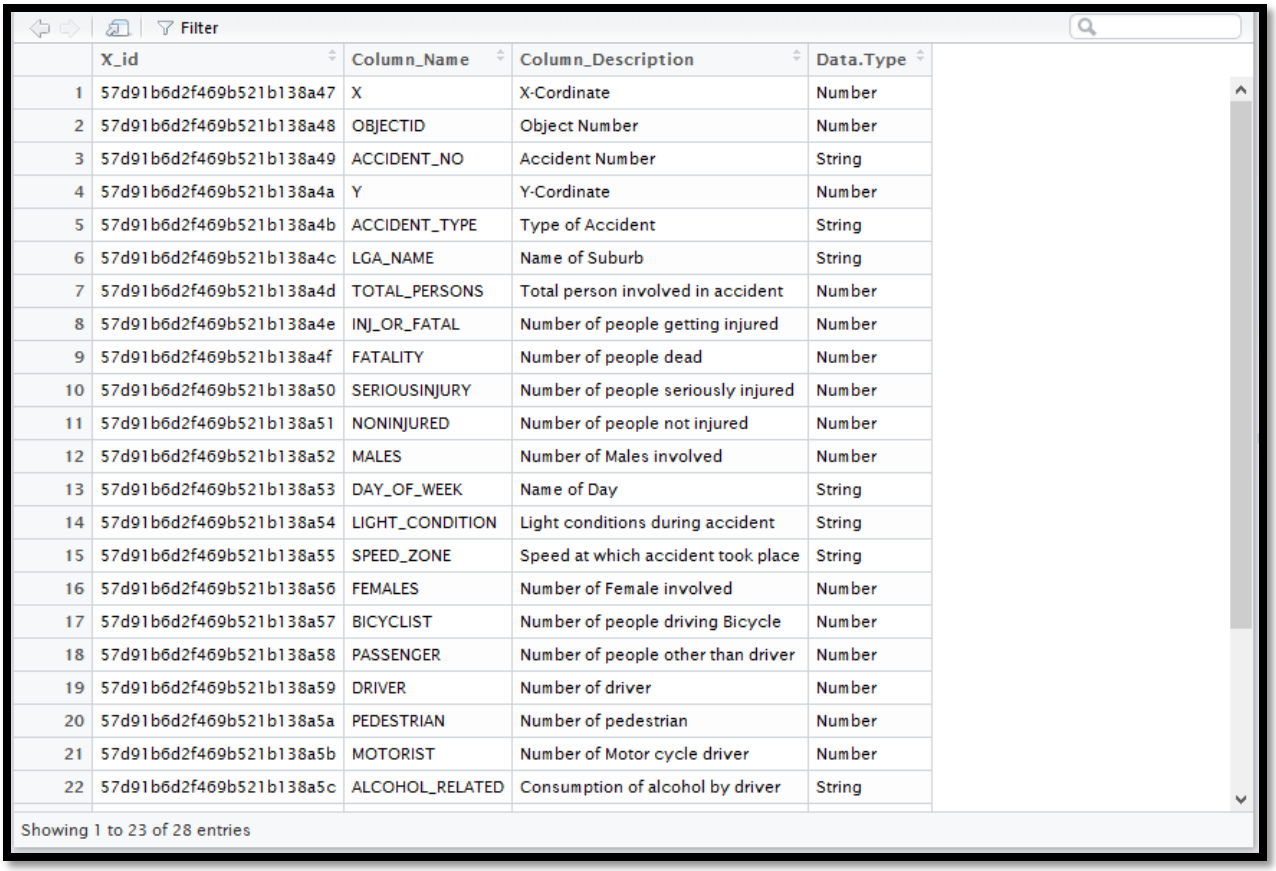
Uploading to Mongod



Uploading to R







Filter

	X_id	Column_Name	Column_Description	Data.Type
1	57d91b6d2f469b521b138a47	X	X-Cordinate	Number
2	57d91b6d2f469b521b138a48	OBJECTID	Object Number	Number
3	57d91b6d2f469b521b138a49	ACCIDENT_NO	Accident Number	String
4	57d91b6d2f469b521b138a4a	Y	Y-Cordinate	Number
5	57d91b6d2f469b521b138a4b	ACCIDENT_TYPE	Type of Accident	String
6	57d91b6d2f469b521b138a4c	LGA_NAME	Name of Suburb	String
7	57d91b6d2f469b521b138a4d	TOTAL_PERSONS	Total person involved in accident	Number
8	57d91b6d2f469b521b138a4e	INJ_OR_FATAL	Number of people getting injured	Number
9	57d91b6d2f469b521b138a4f	FATALITY	Number of people dead	Number
10	57d91b6d2f469b521b138a50	SERIOUSINJURY	Number of people seriously injured	Number
11	57d91b6d2f469b521b138a51	NONINJURED	Number of people not injured	Number
12	57d91b6d2f469b521b138a52	MALES	Number of Males involved	Number
13	57d91b6d2f469b521b138a53	DAY_OF_WEEK	Name of Day	String
14	57d91b6d2f469b521b138a54	LIGHT_CONDITION	Light conditions during accident	String
15	57d91b6d2f469b521b138a55	SPEED_ZONE	Speed at which accident took place	String
16	57d91b6d2f469b521b138a56	FEMALES	Number of Female involved	Number
17	57d91b6d2f469b521b138a57	BICYCLIST	Number of people driving Bicycle	Number
18	57d91b6d2f469b521b138a58	PASSENGER	Number of people other than driver	Number
19	57d91b6d2f469b521b138a59	DRIVER	Number of driver	Number
20	57d91b6d2f469b521b138a5a	PEDESTRIAN	Number of pedestrian	Number
21	57d91b6d2f469b521b138a5b	MOTORIST	Number of Motor cycle driver	Number
22	57d91b6d2f469b521b138a5c	ALCOHOL_RELATED	Consumption of alcohol by driver	String

Showing 1 to 23 of 28 entries

```
"X_id" "Column_Name" "Column_Description" "Data.Type"
"57d91b6d2f469b521b138a47" "X" "X-Cordinate" "Number"
"57d91b6d2f469b521b138a48" "OBJECTID" "Object Number" "Number"
"57d91b6d2f469b521b138a49" "ACCIDENT_NO" "Accident Number" "String"
"57d91b6d2f469b521b138a4a" "Y" "Y-Cordinate" "Number"
"57d91b6d2f469b521b138a4b" "ACCIDENT_TYPE" "Type of Accident" "String"
"57d91b6d2f469b521b138a4c" "LGA_NAME" "Name of Suburb" "String"
"57d91b6d2f469b521b138a4d" "TOTAL_PERSONS" "Total person involved in accident" "Number"
"57d91b6d2f469b521b138a4e" "INJ_OR_FATAL" "Number of people getting injured" "Number"
"57d91b6d2f469b521b138a4f" "FATALITY" "Number of people dead" "Number"
"57d91b6d2f469b521b138a50" "SERIOUSINJURY" "Number of people seriously injured" "Number"
"57d91b6d2f469b521b138a51" "NONINJURED" "Number of people not injured" "Number"
"57d91b6d2f469b521b138a52" "MALES" "Number of Males involved" "Number"
"57d91b6d2f469b521b138a53" "DAY_OF_WEEK" "Name of Day" "String"
"57d91b6d2f469b521b138a54" "LIGHT_CONDITION" "Light conditions during accident" "String"
"57d91b6d2f469b521b138a55" "SPEED_ZONE" "Speed at which accident took place" "String"
"57d91b6d2f469b521b138a56" "FEMALES" "Number of Female involved" "Number"
"57d91b6d2f469b521b138a57" "BICYCLIST" "Number of people driving Bicycle" "Number"
"57d91b6d2f469b521b138a58" "PASSENGER" "Number of people other than driver" "Number"
"57d91b6d2f469b521b138a59" "DRIVER" "Number of driver" "Number"
"57d91b6d2f469b521b138a5a" "PEDESTRIAN" "Number of pedestrian" "Number"
"57d91b6d2f469b521b138a5b" "MOTORIST" "Number of Motor cycle driver" "Number"
"57d91b6d2f469b521b138a5c" "ALCOHOL_RELATED" "Consumption of alcohol by driver" "String"
"57d91b6d2f469b521b138a5d" "NO_OF_VEHICLES" "Number of Vehicle" "Number"
"57d91b6d2f469b521b138a5e" "HEAVYVEHICLE" "Number of Heavy Vehicle" "Number"
"57d91b6d2f469b521b138a5f" "PASSENGERVEHICLE" "Number of Passenger Vehicle" "Number"
"57d91b6d2f469b521b138a60" "MOTORCYCLE" "Number of Motor cycle" "Number"
"57d91b6d2f469b521b138a61" "PUBLICVEHICLE" "Number of Public Vehicle" "Number"
"57d91b6d2f469b521b138a62" "RMA" "Type of Road" "String"
```

Conclusions

Case 1

Considering the above facts the street lights of the five cities should be repaired in order to avoid accidents as not taking measures may result in more accidents and loss of property. Following are the conclusions of all graphs in case 1:

- Top 5 cities where accident occur due to Street lights off are Casey, Melbourne, Monash, Melton and Hume.
- Demographics suggest most of the accidents are caused by passenger car in which most of the drivers get injured.
- Accidents occurring due to consumption of alcohol are almost none.
- Classification table for top 5 cities:

City	Maximum accident occurring speed	Road type with maximum accidents	Accident Type with maximum accidents
Casey	80	Local Road	Collision with Vehicle
Melbourne	60	Arterial Other	Collision with Vehicle
Monash	50	Arterial Highway	Collision with Vehicle
Melton	60	Local Roads	Collision with Vehicle
Hume	60	Local Roads	Collision with Vehicle

Case 2

Considering the above facts the street lights should be implemented in the five cities to reduce number of accidents and loss of property.

Following are the conclusions of all graphs in case 2:

- Top 5 cities where accident occur due to Street lights off are Casey, Cardinia, Yarra Ranges, Geelong and Mitchell.
- Demographics suggest most of the accidents are caused by passenger car in which most of the drivers get injured.
- Accidents occurring due to consumption of alcohol occur mostly on Friday, Saturday and Sunday. Therefore alcohol checking points should be set up on this days.
- Classification table for top 5 cities:

City	Maximum accident occurring speed	Road type with maximum accidents	Accident Type with maximum accidents
Casey	100	Local Road	Collision with Vehicle
Cardinia	100	Arterial Other	Collision with a fixed object
Yarra Ranges	60	Arterial Other	Collision with a fixed object
Geelong	100	Arterial Other	Collision with a fixed object
Mitchell	100	Freeway	Collision with a fixed object