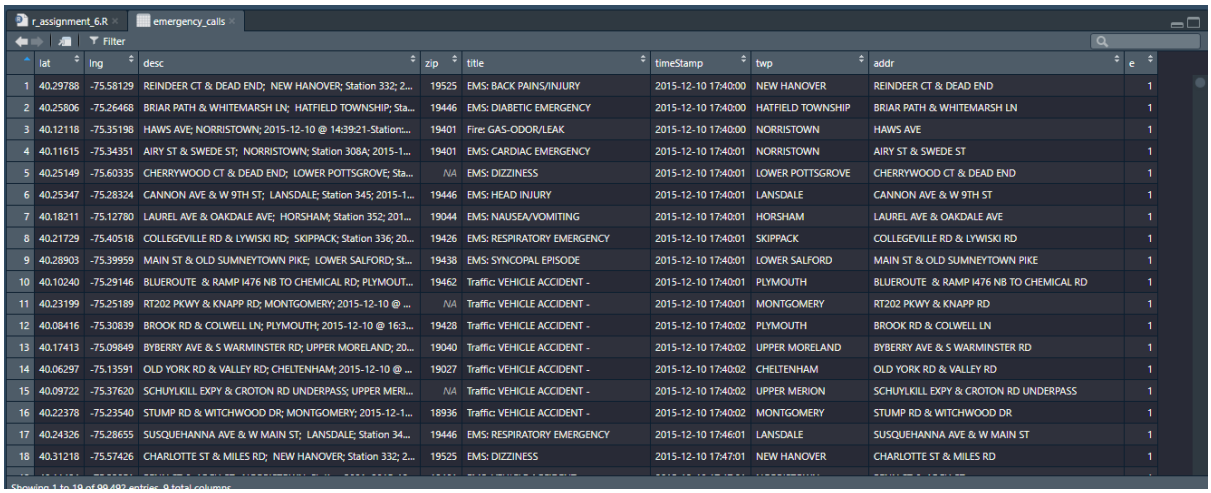
Dhir Thacker | 17070122019 | CSE-1

SD lab assignment 6 (R-assignment 2)

Exploratory data analysis in R

A snippet of the 911.csv dataset



	lat	lng	desc	zip	title	timeStamp	twp	addr	e
1	40.29788	-75.58129	REINDEER CT & DEAD END; NEW HANOVER; Station 332; 2...	19525	EMS: BACK PAINS/INJURY	2015-12-10 17:40:00	NEW HANOVER	REINDEER CT & DEAD END	1
2	40.25806	-75.26468	BRIAR PATH & WHITEMARSH LN; HATFIELD TOWNSHIP; Sta...	19446	EMS: DIABETIC EMERGENCY	2015-12-10 17:40:00	HATFIELD TOWNSHIP	BRIAR PATH & WHITEMARSH LN	1
3	40.12118	-75.35198	HAWS AVE; NORRISTOWN; 2015-12-10 @ 14:39:21-Station:...	19401	Fire: GAS-ODOR/LEAK	2015-12-10 17:40:00	NORRISTOWN	HAWS AVE	1
4	40.11615	-75.34351	AIRY ST & SWEDE ST; NORRISTOWN; Station 308A; 2015-1...	19401	EMS: CARDIAC EMERGENCY	2015-12-10 17:40:01	NORRISTOWN	AIRY ST & SWEDE ST	1
5	40.25149	-75.60335	CHERRYWOOD CT & DEAD END; LOWER POTTSGROVE; Sta...	NA	EMS: DIZZINESS	2015-12-10 17:40:01	LOWER POTTSGROVE	CHERRYWOOD CT & DEAD END	1
6	40.25347	-75.28324	CANNON AVE & W 9TH ST; LANSDALE; Station 345; 2015-1...	19446	EMS: HEAD INJURY	2015-12-10 17:40:01	LANSDALE	CANNON AVE & W 9TH ST	1
7	40.18211	-75.12780	LAUREL AVE & OAKDALE AVE; HORSHAM; Station 352; 201...	19044	EMS: NAUSEA/VOMITING	2015-12-10 17:40:01	HORSHAM	LAUREL AVE & OAKDALE AVE	1
8	40.21729	-75.40518	COLLEGEVILLE RD & LYWISKI RD; SKIPPACK; Station 336; 20...	19426	EMS: RESPIRATORY EMERGENCY	2015-12-10 17:40:01	SKIPPACK	COLLEGEVILLE RD & LYWISKI RD	1
9	40.28903	-75.39959	MAIN ST & OLD SUMNEYTOWN PIKE; LOWER SALFORD; St...	19438	EMS: SYNCOPAL EPISODE	2015-12-10 17:40:01	LOWER SALFORD	MAIN ST & OLD SUMNEYTOWN PIKE	1
10	40.10240	-75.29146	BLUEROUTE & RAMP I476 NB TO CHEMICAL RD; PLYMOUT...	19462	Traffic: VEHICLE ACCIDENT -	2015-12-10 17:40:01	PLYMOUTH	BLUEROUTE & RAMP I476 NB TO CHEMICAL RD	1
11	40.23199	-75.25189	RT202 PKWY & KNAPP RD; MONTGOMERY; 2015-12-10 @ ...	NA	Traffic: VEHICLE ACCIDENT -	2015-12-10 17:40:01	MONTGOMERY	RT202 PKWY & KNAPP RD	1
12	40.08416	-75.30839	BROOK RD & COLWELL LN; PLYMOUTH; 2015-12-10 @ 16:3...	19428	Traffic: VEHICLE ACCIDENT -	2015-12-10 17:40:02	PLYMOUTH	BROOK RD & COLWELL LN	1
13	40.17413	-75.09849	BYBERRY AVE & S WARMINSTER RD; UPPER MORELAND; 20...	19040	Traffic: VEHICLE ACCIDENT -	2015-12-10 17:40:02	UPPER MORELAND	BYBERRY AVE & S WARMINSTER RD	1
14	40.06297	-75.13591	OLD YORK RD & VALLEY RD; CHELTENHAM; 2015-12-10 @ ...	19027	Traffic: VEHICLE ACCIDENT -	2015-12-10 17:40:02	CHELTENHAM	OLD YORK RD & VALLEY RD	1
15	40.09722	-75.37620	SCHUYLKILL EXPY & CROTON RD UNDERPASS; UPPER MERI...	NA	Traffic: VEHICLE ACCIDENT -	2015-12-10 17:40:02	UPPER MERION	SCHUYLKILL EXPY & CROTON RD UNDERPASS	1
16	40.22378	-75.23540	STUMP RD & WITCHWOOD DR; MONTGOMERY; 2015-12-1...	18936	Traffic: VEHICLE ACCIDENT -	2015-12-10 17:40:02	MONTGOMERY	STUMP RD & WITCHWOOD DR	1
17	40.24326	-75.28655	SUSQUEHANNA AVE & W MAIN ST; LANSDALE; Station 34...	19446	EMS: RESPIRATORY EMERGENCY	2015-12-10 17:46:01	LANSDALE	SUSQUEHANNA AVE & W MAIN ST	1
18	40.31218	-75.57426	CHARLOTTE ST & MILES RD; NEW HANOVER; Station 332; 2...	19525	EMS: DIZZINESS	2015-12-10 17:47:01	NEW HANOVER	CHARLOTTE ST & MILES RD	1

- Loading the dataset and import required libraries and packages

```
# Importing data

emergency_calls <- read.csv("911.csv")
View(emergency_calls)

# Loading packages

install.packages("Rshiny")
require(dplyr)
require(plyr)
require(psych)
require(ggplot2)
require(plotly)
require(tidyr)
library(lubridate)
library(viridis)
install.packages(rshi)
library(chron)
# install.packages("ggmap")
library(ggmap)
library(ggmap)
library(date)
# install.packages("date")
require(date)
```

- Summarizing the data

```
# Summarizing the data

dim(emergency_calls)
summary(emergency_calls)
describe(emergency_calls)
str(emergency_calls)
```

```
> dim(emergency_calls)
[1] 99492     9
> summary(emergency_calls)
      lat              lng              desc               zip             title              timeStamp            twp                addr                e          
 Min.   :30.33   Min.   :-95.60   Length:99492       Min.   :17752   Length:99492       Length:99492       Length:99492       Length:99492       Min.   :1  
 1st Qu.:40.10   1st Qu.:-75.39   Class :character   1st Qu.:19038   Class :character   Class :character   Class :character   Class :character   1st Qu.:1  
 Median :40.15   Median :-75.30   Mode  :character   Median :19401   Mode  :character   Mode  :character   Mode  :character   Mode  :character   Median :1  
 Mean   :40.16   Mean   :-75.32                      Mean   :19238                                                                                    Mean   :1  
 3rd Qu.:40.23   3rd Qu.:-75.21                      3rd Qu.:19446                                                                                    3rd Qu.:1  
 Max.   :41.17   Max.   :-75.00                      Max.   :77316                                                                                    Max.   :1  
                                                     NA's   :12855                                                                                               

> describe(emergency_calls)
          vars     n     mean       sd   median  trimmed      mad      min      max    range  skew kurtosis    se
lat          1 99492    40.16     0.09    40.15    40.16     0.08    30.33    41.17    10.83 -16.64  1637.57  0.00
lng          2 99492   -75.32     0.17   -75.30   -75.31     0.13   -95.60   -75.00    20.60 -21.55  2122.56  0.00
desc*        3 99492 49729.76 28710.13 49729.50 49730.12 36864.11     1.00 99455.00 99454.00   0.00    -1.20 91.02
zip          4 86637 19237.66   345.34 19401.00 19269.04    94.89 17752.00 77316.00 59564.00  55.23  9299.34  1.17
title*       5 99492    69.14    35.39    70.00    71.70    51.89     1.00   110.00   109.00  -0.31    -1.31  0.11
timeStamp*   6 99492 36106.86 20779.90 35920.50 36067.31 26520.01     1.00 72577.00 72576.00   0.02    -1.19 65.88
twp*         7 99492    35.58    19.55    34.00    35.76    23.72     1.00    69.00    68.00  -0.01    -1.07  0.06
addr*        8 99492 10867.71  6324.03 10966.00 10870.41  8148.37     1.00 21915.00 21914.00  -0.01    -1.20 20.05
e            9 99492     1.00     0.00     1.00     1.00     0.00     1.00     1.00     0.00    NaN      NaN  0.00
> str(emergency_calls)
'data.frame':   99492 obs. of  9 variables:
 $ lat      : num  40.3 40.3 40.1 40.1 40.3 ...
 $ lng      : num  -75.6 -75.3 -75.4 -75.3 -75.6 ...
 $ desc     : chr  "REINDEER CT & DEAD END;  NEW HANOVER; Station 332; 2015-12-10 @ 17:10:52;" "BRIAR PATH & WHITEMARSH LN;  HATFIELD TOWNSHIP; Station 345; 2015-
12-10 @ 17:29:21;" "HAWS AVE; NORRISTOWN; 2015-12-10 @ 14:39:21-Station:STA27;" "AIRY ST & SWEDE ST;  NORRISTOWN; Station 308A; 2015-12-10 @ 16:47:36;" ...
 $ zip      : int  19525 19446 19401 19401 NA 19446 19044 19426 19438 19462 ...
 $ title    : chr  "EMS: BACK PAINS/INJURY" "EMS: DIABETIC EMERGENCY" "Fire: GAS-ODOR/LEAK" "EMS: CARDIAC EMERGENCY" ...
 $ timeStamp: chr  "2015-12-10 17:40:00" "2015-12-10 17:40:00" "2015-12-10 17:40:00" "2015-12-10 17:40:01" ...
 $ twp      : chr  "NEW HANOVER" "HATFIELD TOWNSHIP" "NORRISTOWN" "NORRISTOWN" ...
 $ addr     : chr  "REINDEER CT & DEAD END" "BRIAR PATH & WHITEMARSH LN" "HAWS AVE" "AIRY ST & SWEDE ST" ...
 $ e        : int  1 1 1 1 1 1 1 1 1 1 ...
>
```

- Checking variable classes

```
32
33  # Checking the variable class
34
35  class(emergency_calls$zip)
36  class(emergency_calls$timeStamp)
37  class(emergency_calls$title)
38
```

37:29 (Top Level)

Console | Terminal | Jobs

E:/Sem_8/sd_lab/r/r_assignment_2/

```
> # Checking the variable class
>
> class(emergency_calls$zip)
[1] "integer"
> class(emergency_calls$timeStamp)
[1] "character"
> class(emergency_calls$title)
[1] "character"
>
```

- Cleaning data
  - Removing ‘e’ column

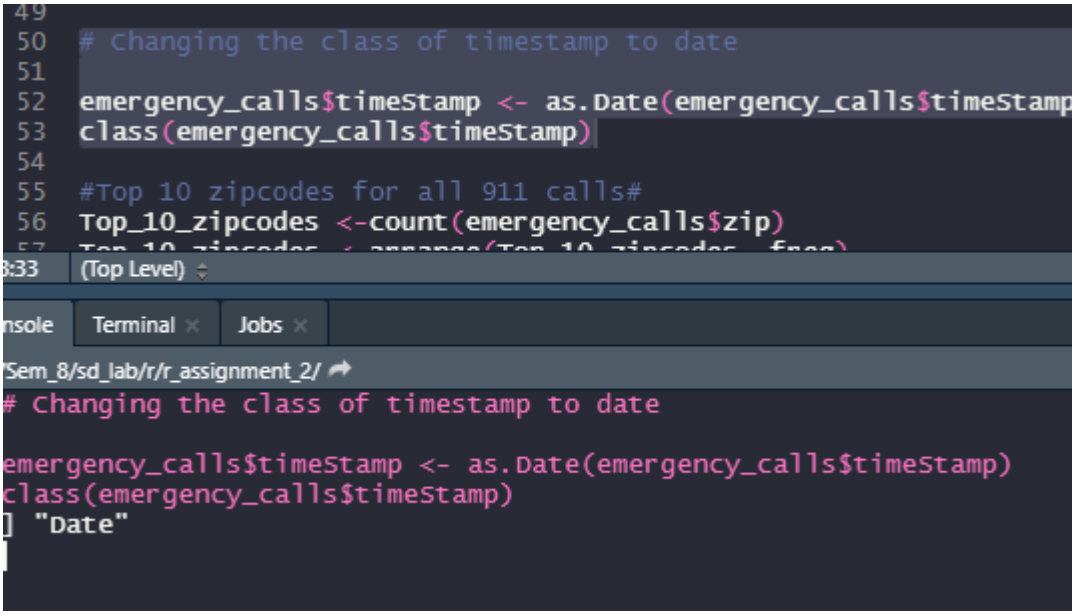
```
42
43  # Cleaning the data
44
45  # Removing the dummy variable ('addr' column and 'e' column)
46
47  emergency_calls <- emergency_calls[,1:8]
48  View(emergency_calls)
49
```

r_assignment_6.R | emergency_calls

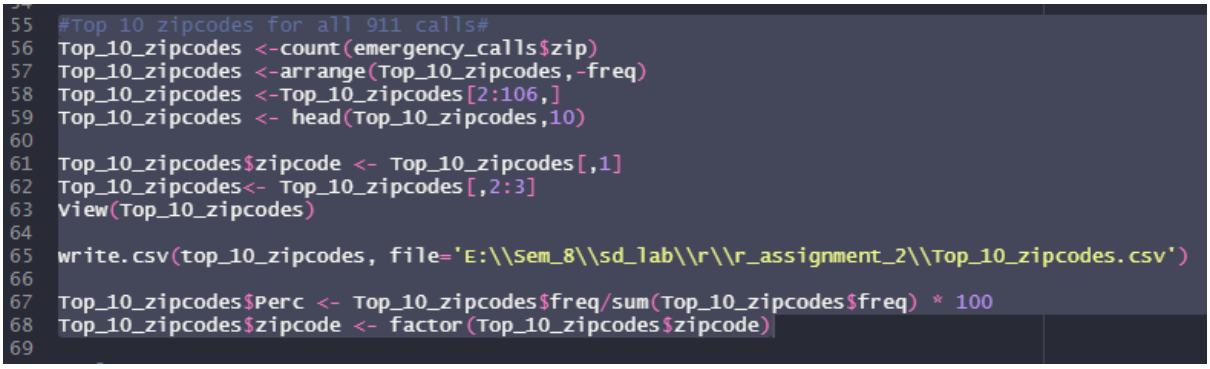
Filter

	lat	lng	desc	zip	title	timeStamp	twp	addr
1	40.29788	-75.58129	REINDEER CT & DEAD END; NEW HANOVER; Station 332; 2...	19525	EMS: BACK PAINS/INJURY	2015-12-10 17:40:00	NEW HANOVER	REINDEER CT & DEAD END
2	40.25806	-75.26468	BRIAR PATH & WHITEMARSH LN; HATFIELD TOWNSHIP; Sta...	19446	EMS: DIABETIC EMERGENCY	2015-12-10 17:40:00	HATFIELD TOWNSHIP	BRIAR PATH & WHITEMARSH LN
3	40.12118	-75.35198	HAWS AVE; NORRISTOWN; 2015-12-10 @ 14:39:21-Station:...	19401	Fire: GAS-ODOR/LEAK	2015-12-10 17:40:00	NORRISTOWN	HAWS AVE
4	40.11615	-75.34351	AIRY ST & SWEDE ST; NORRISTOWN; Station 308A; 2015-1...	19401	EMS: CARDIAC EMERGENCY	2015-12-10 17:40:01	NORRISTOWN	AIRY ST & SWEDE ST
5	40.25149	-75.60335	CHERRYWOOD CT & DEAD END; LOWER POTTSGROVE; Sta...	NA	EMS: DIZZINESS	2015-12-10 17:40:01	LOWER POTTSGROVE	CHERRYWOOD CT & DEAD END
6	40.25347	-75.28324	CANNON AVE & W 9TH ST; LANSDALE; Station 345; 2015-1...	19446	EMS: HEAD INJURY	2015-12-10 17:40:01	LANSDALE	CANNON AVE & W 9TH ST
7	40.18211	-75.12780	LAUREL AVE & OAKDALE AVE; HORSHAM; Station 352; 201...	19044	EMS: NAUSEA/VOMITING	2015-12-10 17:40:01	HORSHAM	LAUREL AVE & OAKDALE AVE
8	40.21729	-75.40518	COLLEGEVILLE RD & LYWISKI RD; SKIPPACK; Station 336; 20...	19426	EMS: RESPIRATORY EMERGENCY	2015-12-10 17:40:01	SKIPPACK	COLLEGEVILLE RD & LYWISKI RD
9	40.28903	-75.39959	MAIN ST & OLD SUMNEYTOWN PIKE; LOWER SALFORD; St...	19438	EMS: SYNCOPAL EPISODE	2015-12-10 17:40:01	LOWER SALFORD	MAIN ST & OLD SUMNEYTOWN PIKE
10	40.10240	-75.29146	BLUEROUTE & RAMP I476 NB TO CHEMICAL RD; PLYMOUT...	19462	Traffic: VEHICLE ACCIDENT -	2015-12-10 17:40:01	PLYMOUTH	BLUEROUTE & RAMP I476 NB TO CHEMICAL RD

  - Changing class of timestamp to date



- Top 10 zipcodes with the most number of 911 calls

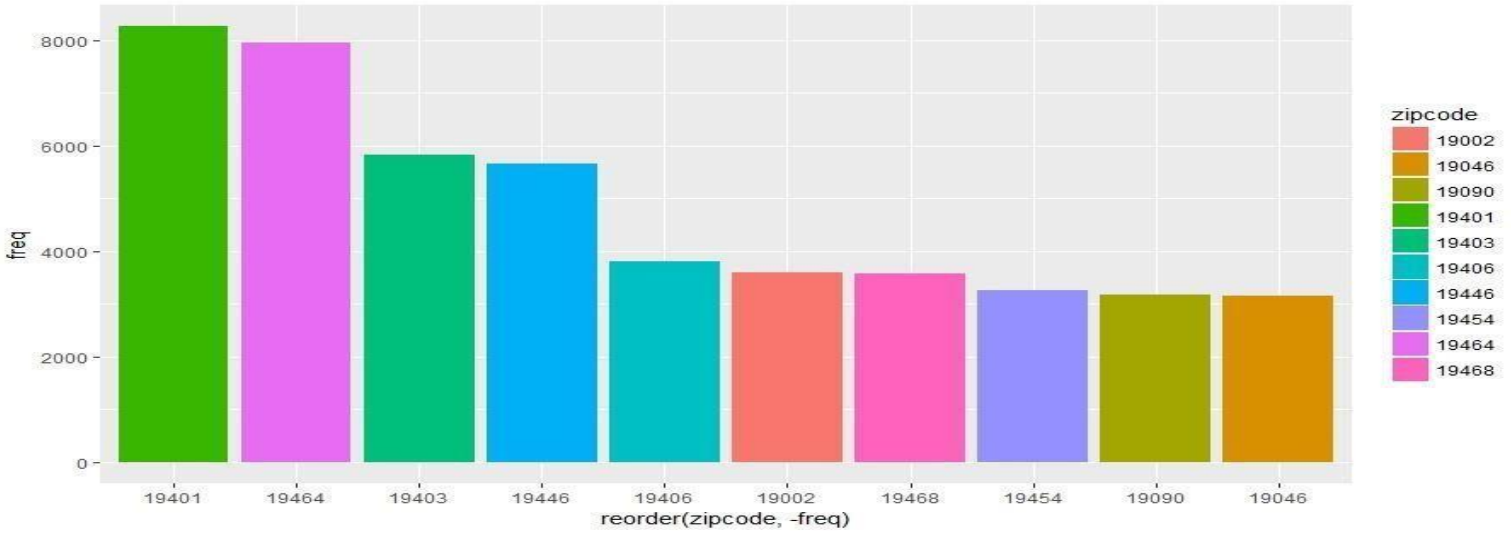


r_assignment_6.R* | Top_10_zipcodes | emergency_calls

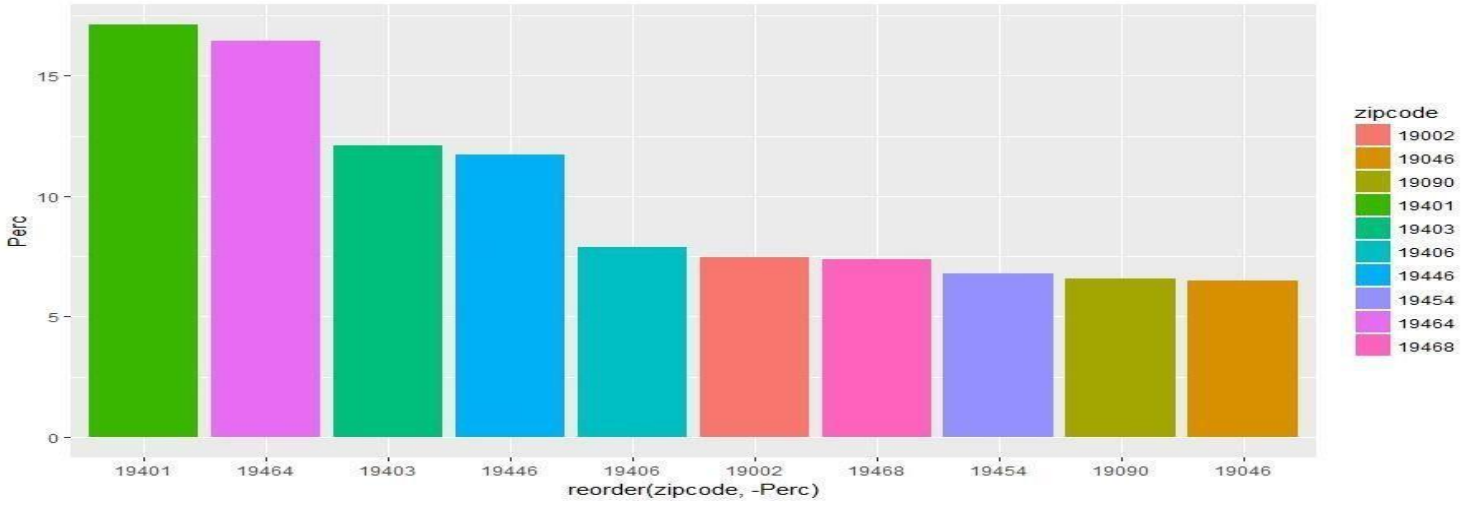
Filter

	freq	zipcode	Perc
2	6979	19401	17.261080
3	6643	19464	16.430055
4	4854	19403	12.005342
5	4748	19446	11.743174
6	3174	19406	7.850218
7	3050	19002	7.543530
8	2990	19468	7.395133
9	2781	19454	6.878215
10	2635	19090	6.517115
11	2578	19038	6.376138

TOP 10 ZIPCODES WITH MOST NO. OF 911 CALLS



TOP 10 ZIPCODES WITH MOST PERCENTAGEOF 911 CALLS



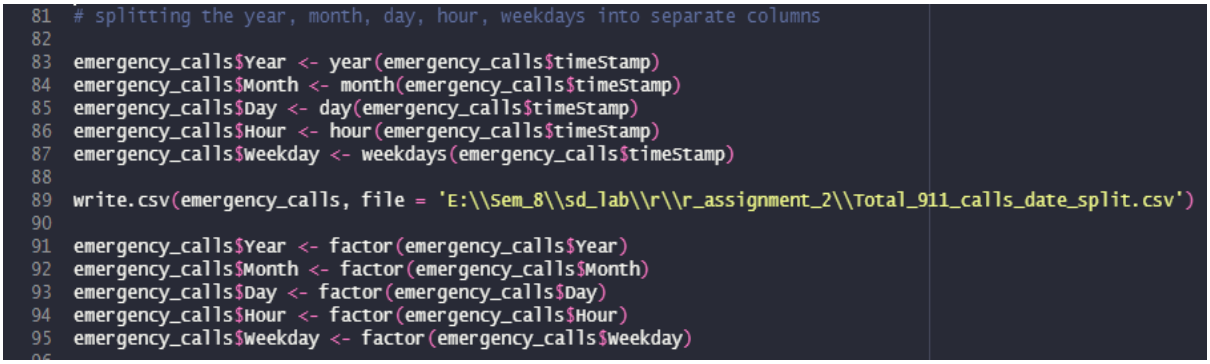
- Splitting title column into type and subtype

r_assignment_6.R* | Top_10_zipcodes | emergency_calls

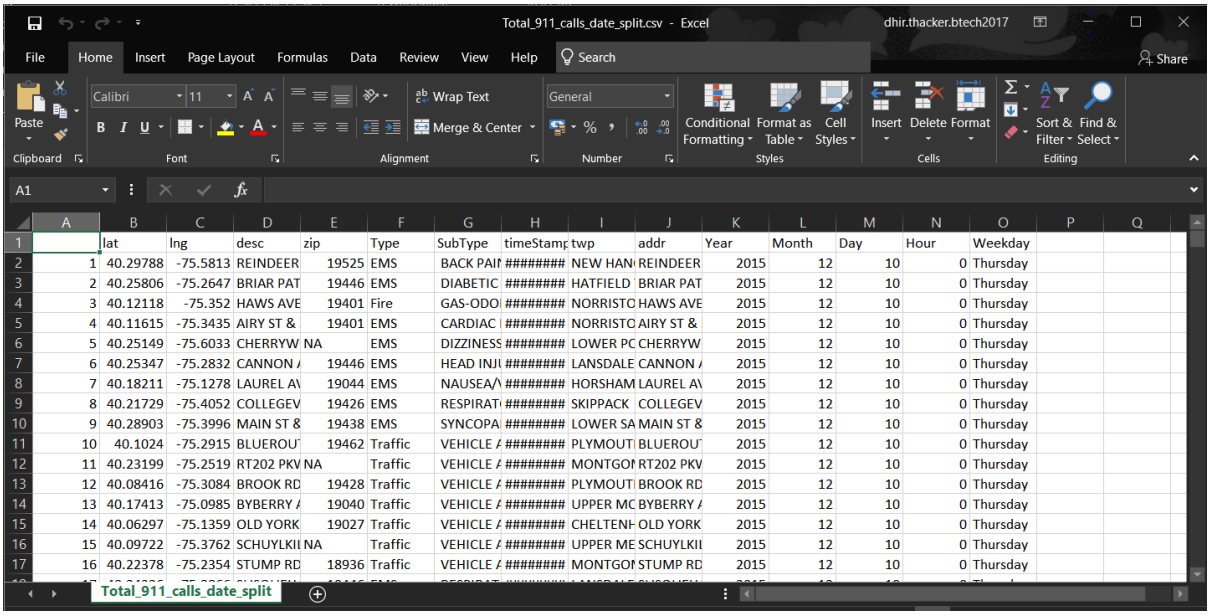
Filter

	lat	lng	desc	zip	Type	SubType	timeStamp	twp	addr
1	40.29788	-75.58129	REINDEER CT & DEAD END; NEW HANOVER; Station 332; 2...	19525	EMS	BACK PAINS/INJURY	2015-12-10	NEW HANOVER	REINDEER CT & DEAD END
2	40.25806	-75.26468	BRIAR PATH & WHITEMARSH LN; HATFIELD TOWNSHIP; Sta...	19446	EMS	DIABETIC EMERGENCY	2015-12-10	HATFIELD TOWNSHIP	BRIAR PATH & WHITEMARSH LN
3	40.12118	-75.35198	HAWS AVE; NORRISTOWN; 2015-12-10 @ 14:39:21-Station:...	19401	Fire	GAS-ODOR/LEAK	2015-12-10	NORRISTOWN	HAWS AVE
4	40.11615	-75.34351	AIRY ST & SWEDE ST; NORRISTOWN; Station 308A; 2015-1...	19401	EMS	CARDIAC EMERGENCY	2015-12-10	NORRISTOWN	AIRY ST & SWEDE ST
5	40.25149	-75.60335	CHERRYWOOD CT & DEAD END; LOWER POTTSGROVE; Sta...	NA	EMS	DIZZINESS	2015-12-10	LOWER POTTSGROVE	CHERRYWOOD CT & DEAD END
6	40.25347	-75.28324	CANNON AVE & W 9TH ST; LANSDALE; Station 345; 2015-1...	19446	EMS	HEAD INJURY	2015-12-10	LANSDALE	CANNON AVE & W 9TH ST
7	40.18211	-75.12780	LAUREL AVE & OAKDALE AVE; HORSHAM; Station 352; 201...	19044	EMS	NAUSEA/VOMITING	2015-12-10	HORSHAM	LAUREL AVE & OAKDALE AVE
8	40.21729	-75.40518	COLLEGEVILLE RD & LYWISKI RD; SKIPPACK; Station 336; 20...	19426	EMS	RESPIRATORY EMERGENCY	2015-12-10	SKIPPACK	COLLEGEVILLE RD & LYWISKI RD
9	40.28903	-75.39959	MAIN ST & OLD SUMNEYTOWN PIKE; LOWER SALFORD; St...	19438	EMS	SYNCOPAL EPISODE	2015-12-10	LOWER SALFORD	MAIN ST & OLD SUMNEYTOWN PIKE
10	40.10240	-75.29146	BLUEROUTE & RAMP I476 NB TO CHEMICAL RD; PLYMOUT...	19462	Traffic	VEHICLE ACCIDENT -	2015-12-10	PLYMOUTH	BLUEROUTE & RAMP I476 NB TO CHEMICAL RD

- Splitting date column into year, month, day, hour, weekdays columns



```
81  # splitting the year, month, day, hour, weekdays into separate columns
82
83  emergency_calls$Year <- year(emergency_calls$timeStamp)
84  emergency_calls$Month <- month(emergency_calls$timeStamp)
85  emergency_calls$Day <- day(emergency_calls$timeStamp)
86  emergency_calls$Hour <- hour(emergency_calls$timeStamp)
87  emergency_calls$Weekday <- weekdays(emergency_calls$timeStamp)
88
89  write.csv(emergency_calls, file = 'E:\\Sem_8\\sd_lab\\r\\r_assignment_2\\Total_911_calls_date_split.csv')
90
91  emergency_calls$Year <- factor(emergency_calls$Year)
92  emergency_calls$Month <- factor(emergency_calls$Month)
93  emergency_calls$Day <- factor(emergency_calls$Day)
94  emergency_calls$Hour <- factor(emergency_calls$Hour)
95  emergency_calls$Weekday <- factor(emergency_calls$Weekday)
```



- Type analysis
  - List of total calls by type

```
# list of total calls by Type

Total_calls_by_type <- summarise(group_by(emergency_calls, Type), Total_calls=n())
write.csv(Total_calls_by_type, file='E:\\Sem_8\\sd_lab\\r\\r_assignment_2\\Total_calls_by_type.csv')
Total_calls_by_type$Perc <- Total_calls_by_type$Total_calls/sum(Total_calls_by_type$Total_calls)*100
```

	Type	Total_calls
1	EMS	48877
2	Fire	14920
3	Traffic	35695

Total_calls_by_type

- List of total calls by zip

```
# list of total calls by zip

Total_calls_by_zip  <- summarise(group_by(emergency_calls, zip), Total_calls=n())
write.csv(Total_calls_by_zip, file='E:\\Sem_8\\sd_lab\\r\\r_assignment_2\\Total_calls_by_zip.csv')
```

	zip	Total_calls
1	17752	1
2	18036	2
3	18041	414
4	18054	326
5	18056	6
6	18070	54
7	18073	736
8	18074	435
9	18076	306
10	18092	14
11	18103	4
12	18914	66
13	18915	118
14	18927	2
15	18932	11
16	18936	254
17	18951	10
18	18960	34
19	18964	1287
20	18966	44
21	18969	672
22	18974	215
23	18976	45
24	19001	1514
25	19002	3050

Total_calls_by_zip

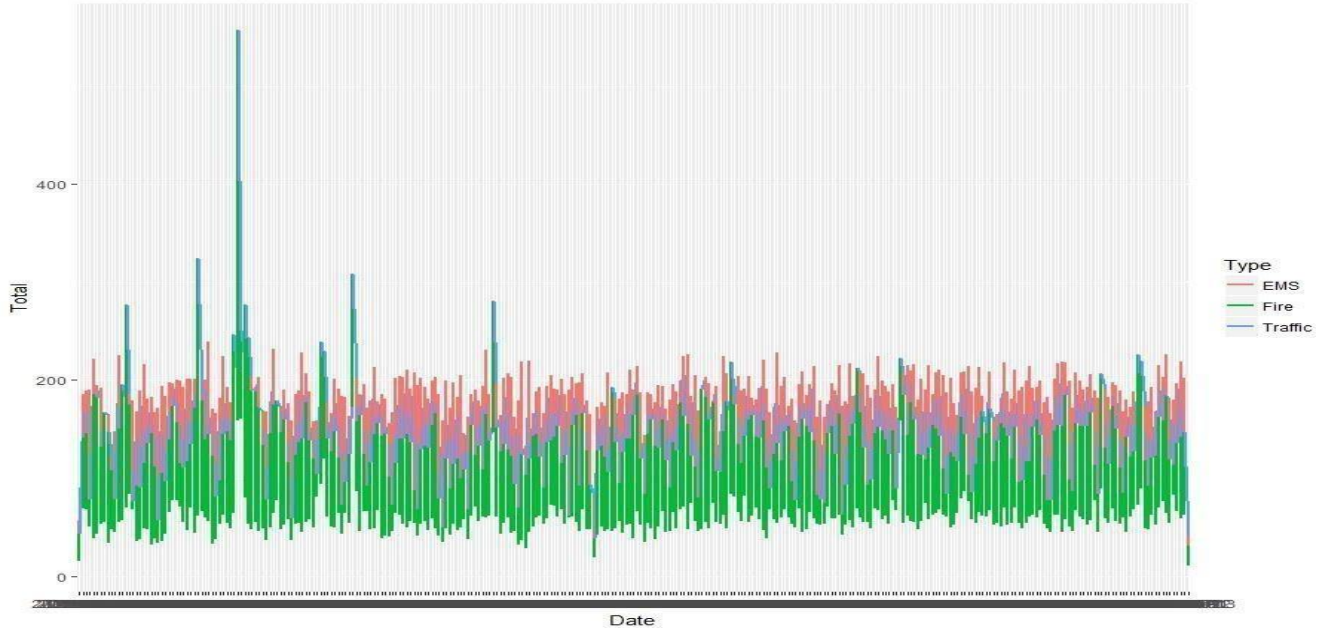
- List of total calls by date

```
# list of total calls by date

Total_calls_by_date <- summarise(group_by(emergency_calls, Year,Month,Day), Total_calls = n())
write.csv(Total_calls_by_date, file='E:\\Sem_8\\sd_lab\\r\\r_assignment_2\\Total_calls_by_date.csv')
```

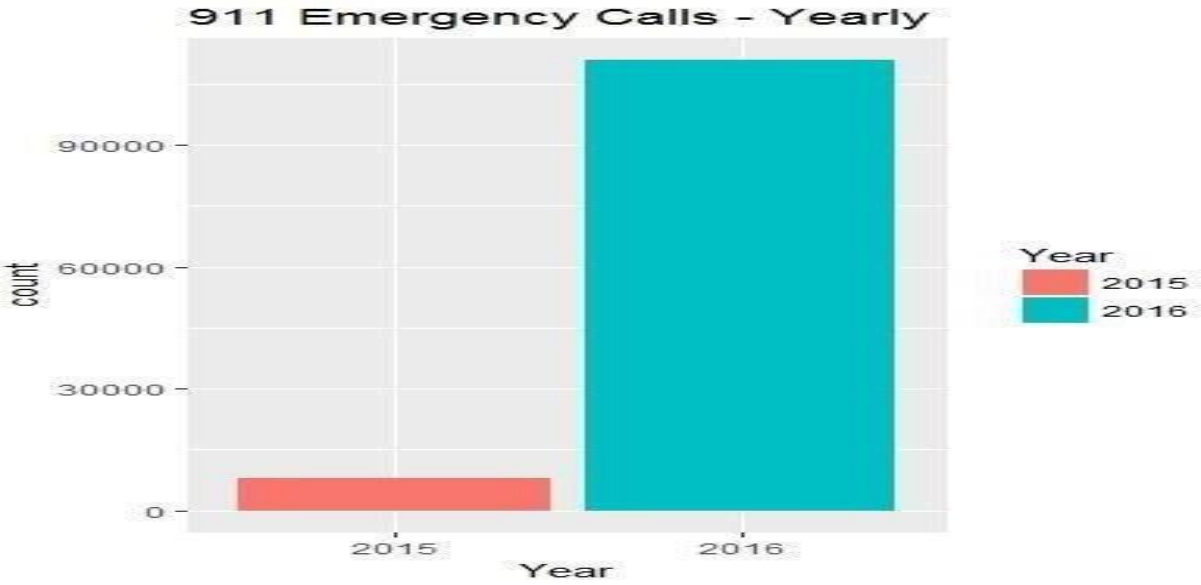
	Year	Month	Day	Total_calls
1	2015	12	10	115
2	2015	12	11	396
3	2015	12	12	403
4	2015	12	13	319
5	2015	12	14	447
6	2015	12	15	421
7	2015	12	16	377
8	2015	12	17	388
9	2015	12	18	346
10	2015	12	19	279
11	2015	12	20	271
12	2015	12	21	430
13	2015	12	22	448
14	2015	12	23	524
15	2015	12	24	445
16	2015	12	25	323

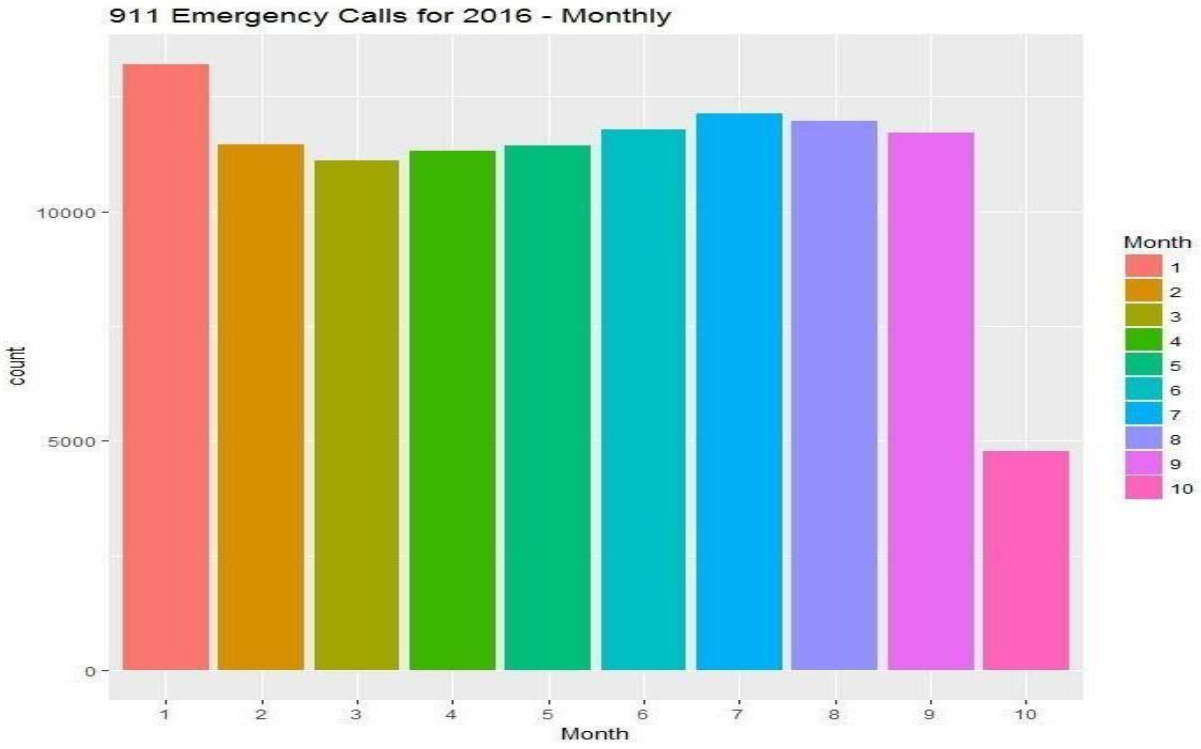
Total_calls_by_date

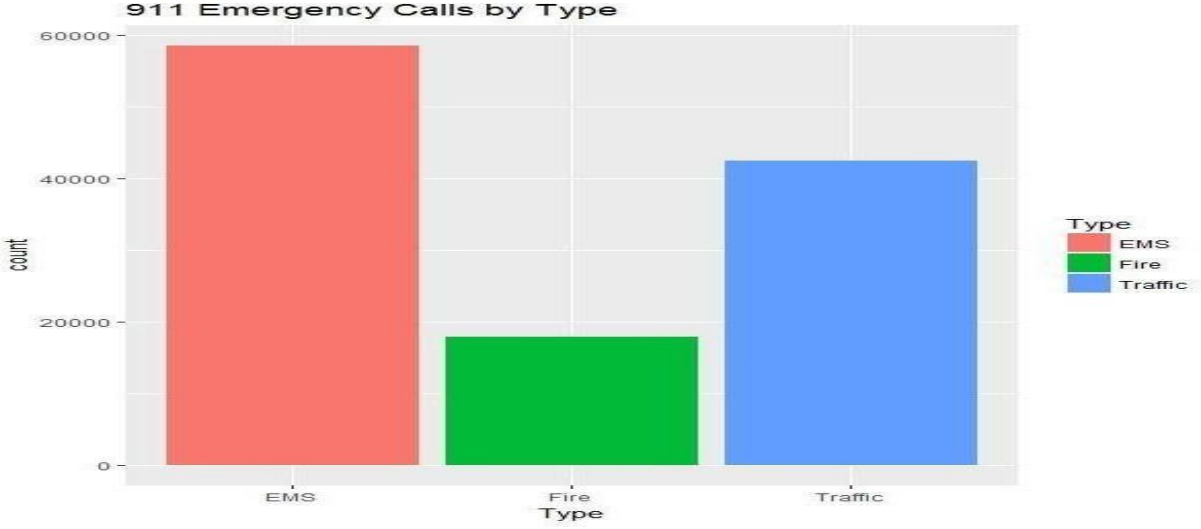
TOTAL NO.OF 911 CALLS MADE EACH DAY BY DIFFERENT TYPES

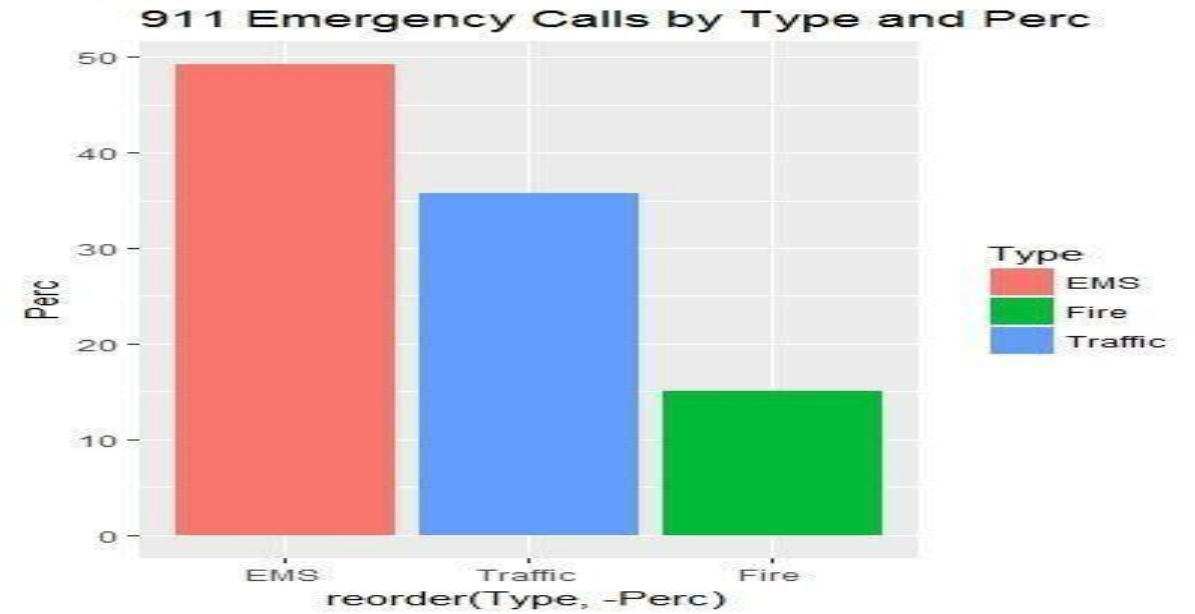


Important analysis plots



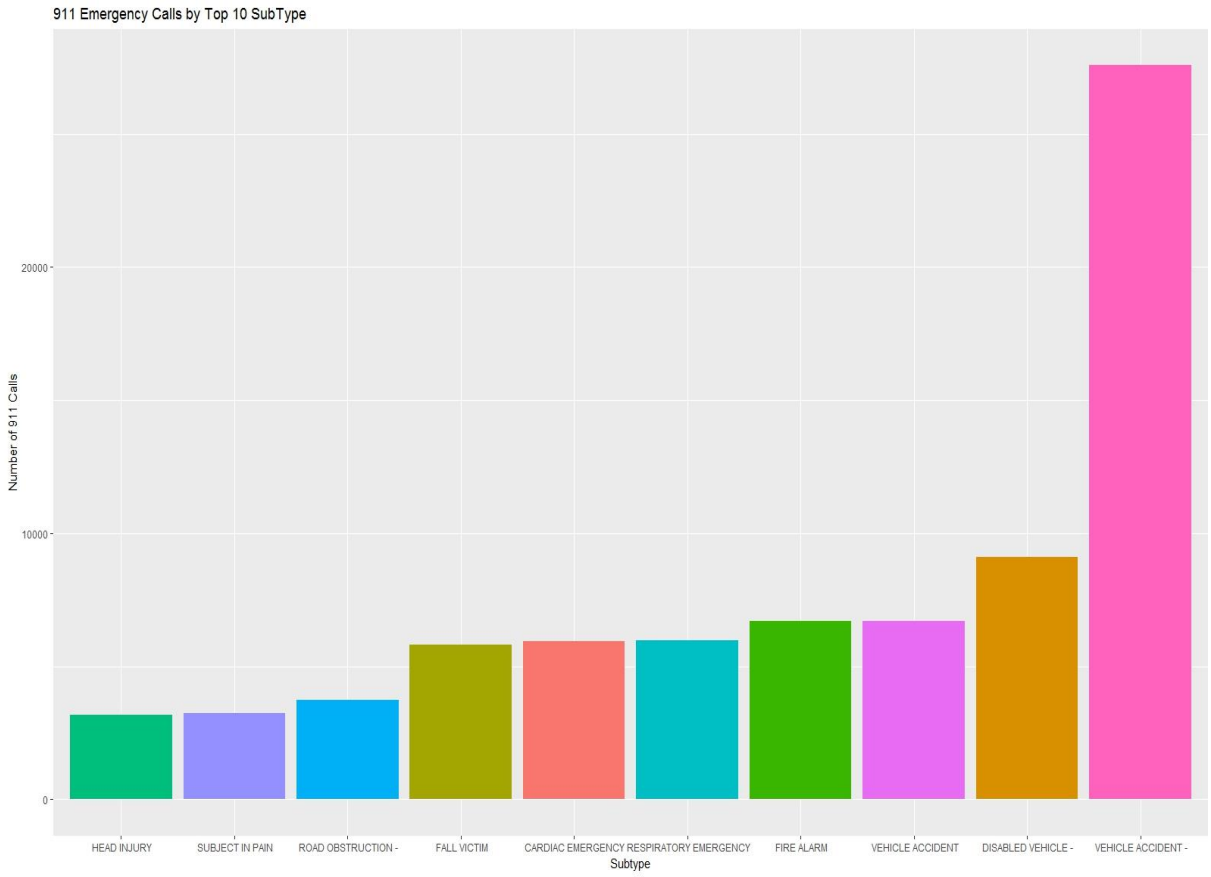






- Subtype analysis
  - Top 10 subtypes overall

```r
# Top 10 Subtypes overall

Top_10_subtypes <- arrange(Total_calls_by_SubType,-Total_calls)
Top_10_subtypes <- head(Top_10_subtypes,10)
View(Top_10_subtypes)


Top_10_subtypes$Perc <- Top_10_subtypes$Total_calls/sum(Top_10_subtypes$Total_calls) * 100
View(Top_10_subtypes)

ggplot(Top_10_subtypes, aes(reorder(SubType, Total_calls), Total_calls, fill = SubType)) +
  geom_bar(stat = "identity") + theme(legend.position = "none") +
  ggtitle("911 Emergency Calls by Top 10 SubType") + xlab("Subtype") + ylab("Number of 911 Calls")
```



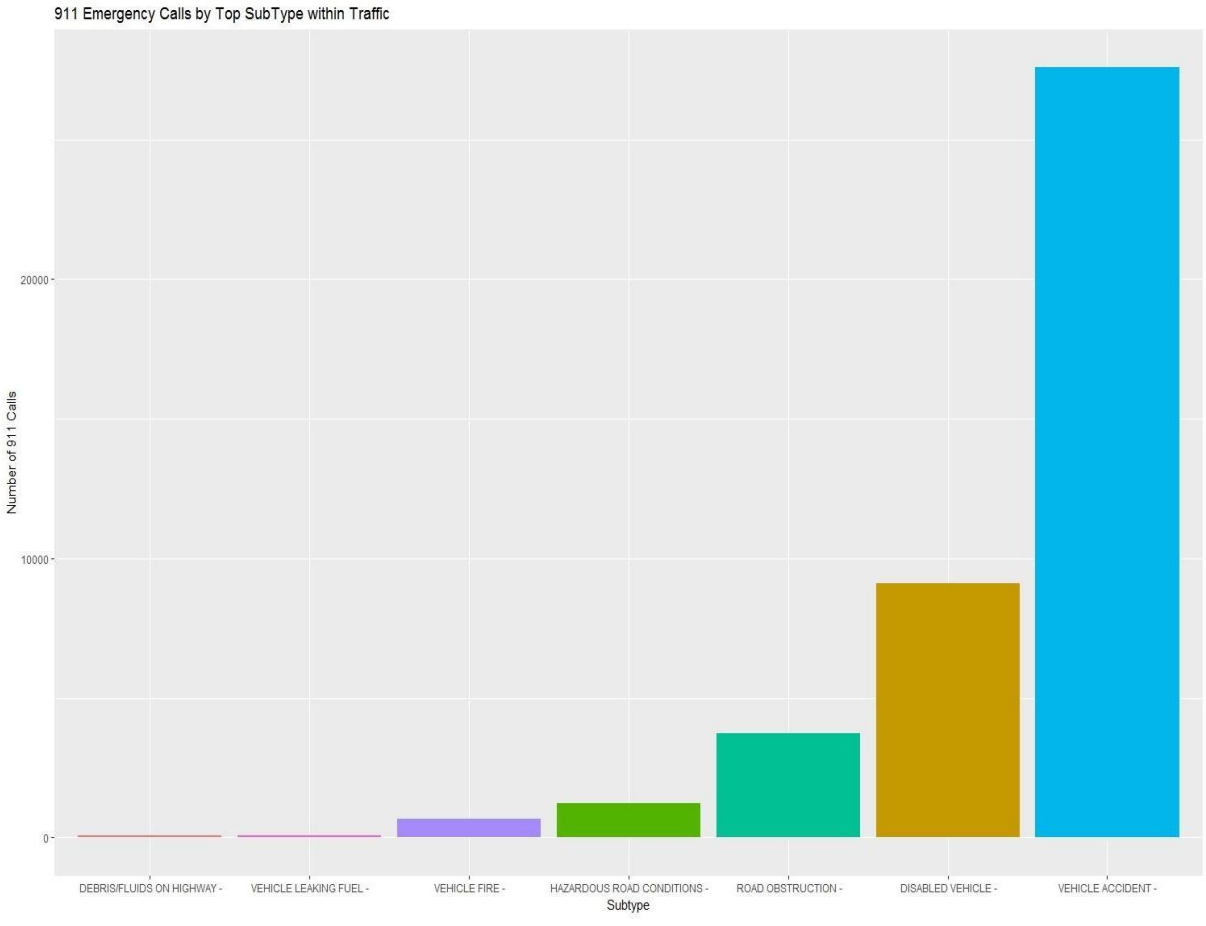
- Top subtypes by traffic

```
# Top subtypes by Traffic#

Total_calls_by_Traffic <- summarise(group_by(Emergency_Calls,Type,SubType),Total = n())
Total_calls_by_Traffic <- subset(Total_calls_by_Traffic,Total_calls_by_Traffic$Type=="Traffic")
View(Total_calls_by_Traffic)

ggplot(Total_calls_by_Traffic, aes(reorder(SubType, Total), Total, fill = SubType)) +
  geom_bar(stat = "identity") + theme(legend.position = "none") +
  ggtitle("911 Emergency Calls by Top SubType within Traffic") + xlab("Subtype") + ylab("Number of 911 Calls")



Subtype_by_weekdays <- summarise(group_by(Emergency_Calls,Weekday, SubType), Total = n())

Subtype_by_weekdays <- subset(Subtype_by_weekdays,Subtype_by_weekdays$Weekday != "Sunday"
                              & Subtype_by_weekdays$Weekday !="Saturday")
```

r_assignment_6.R* | Total_calls_by_Traffic | Total_calls

Filter

	Type	SubType	Total
1	Traffic	DEBRIS/FLUIDS ON HIGHWAY -	67
2	Traffic	DISABLED VEHICLE -	7702
3	Traffic	HAZARDOUS ROAD CONDITIONS -	1086
4	Traffic	ROAD OBSTRUCTION -	3144
5	Traffic	VEHICLE ACCIDENT -	23066
6	Traffic	VEHICLE FIRE -	553
7	Traffic	VEHICLE LEAKING FUEL -	77



- Subtype by weekdays

r_assignment_6.R* | Subtype_by_weekdays | Total_calls_by_T

Filter

	Weekday	SubType	Total
1	Friday	ABDOMINAL PAINS	200
2	Friday	ACTIVE SHOOTER	1
3	Friday	ALLERGIC REACTION	81
4	Friday	ALTERED MENTAL STATUS	203
5	Friday	AMPUTATION	3
6	Friday	ANIMAL BITE	10
7	Friday	APPLIANCE FIRE	31
8	Friday	ASSAULT VICTIM	131
9	Friday	BACK PAINS/INJURY	93
10	Friday	BUILDING FIRE	151
11	Friday	BURN VICTIM	13
12	Friday	CARBON MONOXIDE DETECTOR	80
13	Friday	CARDIAC ARREST	90
14	Friday	CARDIAC EMERGENCY	727

- Hourly spike by Traffic subtype