ROAD TRAFFIC ACCIDENT DATA ANALYSIS AND VISUALIZATION IN R PROGRAMMING LANGUAGE

VALENTIN SILYANOV[1], JAMSHID SODIKOV[2], QUDRATILLA AZIZOV[2]
[1]*Moscow Automobile And Road Construction State Technical University (MADI)*
[2]*Tashkent Highway Engineering Institute*
Corresponding author's email: silyanov@bk.ru

Abstract: *The paper reviews road traffic accident data analysis and visualization in R programming environment. The aim is to show how to extract meaningful data from raw database and visualize it. The results revealed that hour wise, day wise, month wise and year wise plots allowed observe how road traffic accidents change in timescale. Two types of road traffic accident mainly occurred such as type 1 (collision) and type 5 (collision with pedestrian). Both types of road traffic accidents happened in similar magnitudes across all timescales. Visualization and data analysis of road traffic accidents led to make conclusions which would assist reduce the number of accidents.*

Key words: *Road traffic accidents, data analysis, visualization, R programming language*

I. INTRODUCTION

Road traffic accidents occurs randomly in time and location. This is due to an incident depends on number of factors such as driver, vehicle, and road. These factors may play role in an incident occurrence separately or jointly. Driving behavior changes in response to the vehicle and road besides it heavily depends on driver's physiological condition, age, sex, education and other factors. Therefore, it's complicated to predict where and when traffic accident would happen. Based on historical traffic accident data, it's possible to analyze traffic incident data to find relationship between factors. Traffic accident data visualization on other hand provide detailed insights how it changes over time. The paper focuses on practical issues to prevent traffic accidents. Data analysis and visualization help to observe traffic accidents occurrence and take appropriate measures to increase traffic safety.

In Uzbekistan, about 2,000 people die each year as a result of traffic accidents. According to the Pulitzer Center on Crisis Reporting, it has the lowest figures in road deaths among Central Asian countries, for every 100 thousand people 11.32 people die. In Kazakhstan, it is equal to 20.5, in Kyrgyzstan - 19.2, in Russia - 18.6, in Belarus - 14.4, and in Ukraine - 13.5. Economic losses from traffic accidents are equivalent to 2.8% of Uzbekistan's GDP, which is also the lowest indicator (Tihonov, 2016).

Road accident is an event that arose during the movement of a vehicle on the road, in which people are killed or injured; vehicles or structures are damaged. Road accidents are divided into the following types (Highway Standard of Uzbekistan, 2007):

1. Collision: An incident in which moving vehicles collided with each other or with train. It includes collisions with parked vehicle (in front of traffic lights, traffic congestion or due to technical malfunction) and collision with stopped trains.

2. Rollover is a type of vehicle crash in which a vehicle tips over onto its side or roof.

3. Collision with unattended vehicle: An incident in which a moving vehicle has driven onto a standing vehicle, as well as to a trailer or a semi-trailer.

4. Collision with an obstacle: An incident in which a vehicle hit a stationary object (bridge support, pole, tree, fence, etc.).

5. Collision with pedestrian: An incident in which a vehicle has hit a person or a person hit moving vehicle. it also includes incidents in which pedestrians have injured from a cargo or object carried by a vehicle (boards, containers, rope, etc.).

6. Collision with a cyclist: An incident in which a vehicle has hit a bicyclist or cyclist has hit a moving vehicle.

7. Collision with a cartage: An incident in which a vehicle has hit a cartage, or wagons transported by animals, or carts transported by these animals, hit a moving vehicle.

8. Passenger's fall: An incident in which a passenger has fallen from a moving vehicle or from the cabin of a moving vehicle as a result of a sudden change in speed or trajectory, etc., The fall of a passenger from a non-moving vehicle when boarding (landing) at a stop is not an accident.

9. Other kind of accident: Incidents which are not related to the above mentioned road accident types. It includes the drop off a carried cargo or an object thrown by the wheel to a person, an animal or other vehicle, a hit on persons who are not participants in traffic, a collision with a suddenly appeared obstacle (a fallen cargo, a loosened wheel, etc.), etc.

II. LITERATURE REVIEW

In Japan, traffic accident analysis system using GPS was developed to decrease traffic accidents. The system has following information: road structure, road accessory facilities, and weather information. The system is used to analyze accident frequency, accident rate, and fatality rate (Hirasawa, 2003). In Pakistan, road traffic accident analysis carried out to identify causes of road traffic accident occurrence in terms of hourly wise, daily wise, monthly wise, yearly wise and road traffic accident severity measured (Syyed, 2013). In US, some researchers suggested to using machine learning paradigms in traffic accident analysis (Miao, etc. 2005). They found that among several machine learning paradigms hybrid decision tree-neural network approach outperformed the individual approaches. In Saudi Arabia, researchers (Ansari, etc. 2000) reviewed road traffic accidents from 1971 to 1994 to determine causes and effects. They found that causes are rapid infrastructure development, inflow immigrants with various background and driving habits, most of the traffic accidents are result of over speeding and driver error. In India, relationship between types of lanes, total no. of injuries, accidents, persons killed, types of vehicles, awareness of drivers and accidents, types of highways and number of accidents, persons killed or injured were studied (Pooja, etc. 2017). Recently, in Uzbekistan there were published papers which highlighted new electronic road traffic accident collection system and visualization in google fusion table (Sodikov, 2017).

III. DATA ANALYSIS

For analysis and visualization of traffic accidents in Tashkent city, data used from 2005 to 2012, with a total number of 2053 observations. Table 1 shows variables such as ID, Date, Time, Number of accidents, Number of dead, Number of injured, Type of road traffic accident, and location (r code: dtp[102:112,]). The objective of data analysis is to extract useful data from database and further use for visualization and analysis.

Table 1. *Extracted road traffic accidents data*

ID	Date	Time	No. of accident	No. of Dead	No. of Injured	Type of RTA	Location
102	02.27.2008	8:20:00	1	0	1	1	Tashkent u.yusupov street
103	03.07.2008	21:30:00	1	0	1	1	Tashkent a.kodiriy street
104	03.11.2008	22:00:00	1	0	2	1	Tashkent konservatoriya binosi
105	04.13.2008	22:30:00	1	0	1	1	Tashkent navoiy street
106	04.18.2008	0:25:00	1	0	1	1	Tashkent navoiy street
107	05.23.2008	17:00:00	1	0	1	5	Tashkent navoiy street 1
108	08.31.2008	18:50:00	1	0	1	5	Tashkent a.kodiriy street
109	09.25.2008	10:30:00	1	0	1	1	Tashkent selhozteh petrol
110	10.29.2008	12:30:00	1	0	1	5	Tashkent konservatoriya
111	11.13.2008	17:40:00	1	1	0	5	Tashkent a.kodiriy street
112	12.03.2008	23:20:00	1	0	1	5	Tashkent ezidyor restoran

While analyzing dataset, missing data were removed. Traffic accidents data analyzed from 2007 to 2011 years, from date variable we extracted day of the week, month and year. Data analysis revealed that there are two types of traffic accidents frequently occurred, namely collision (type 1) and collision with pedestrian (type 5), in quantity: 874 and 960 cases correspondingly. The primary goal of the analysis is to observe distribution of the traffic accident data across time during the day, day of the week, month and year.

IV. ROAD TRAFFIC DATA VISUALIZATION

In order to visualize data, we used R programming language. All necessary libraries loaded and script is provided in appendix A. The first part of script produces multiple plots and the second part of the script produces plots for hour wise, day wise, month wise and year wise plots. Let's review each plot and observe how road traffic accidents would change over different time scales (fig 1). All plots separated by type of traffic accident such as type 1 (collision) and type 5 (collision with pedestrian).

From time wise plot we may observe that type 5 (collision with pedestrian) accident mainly occurs from 07:00 to 23:00 with peak values at 18:00 and 16:00. The plot clearly shows that type 5 happens during the morning commute 08:00 (freq = 40) and from 16:00 to 20:00 evening commute. On other hand type 1 (collision) happened relatively frequently when type 5 had low frequency from 00:00 to 07:00. Type 1 accident occurred more than 20 times from 07:00 to 23:00.

The second plot is day wise shows that the majority type 5 accident occurred on Friday (freq = 160), Tuesday(freq = 153), and Wednesday (freq = 148). Lowest value observed on Sunday (freq = 98). Type 1 accident happened on Saturday (freq = 131), followed by Friday (freq = 123) and Sunday (freq =122).

The third plot highlights month wise road traffic data distribution, March (freq = 94) and September (freq = 94) have highest frequency for the type 5 accident and lowest frequency in July (freq = 63). Type 1 accident lowest frequency occurred on February (freq = 52), May (freq = 53), August (freq = 56) and November (freq = 50). Type 1accident approximately evenly distributed across months.

The fourth plot shows year wise road traffic accident distribution; the highest frequency has been in year 2009 for type 5 accident (freq = 225), but for type 1 accident year 2008 has highest frequency (freq = 173).

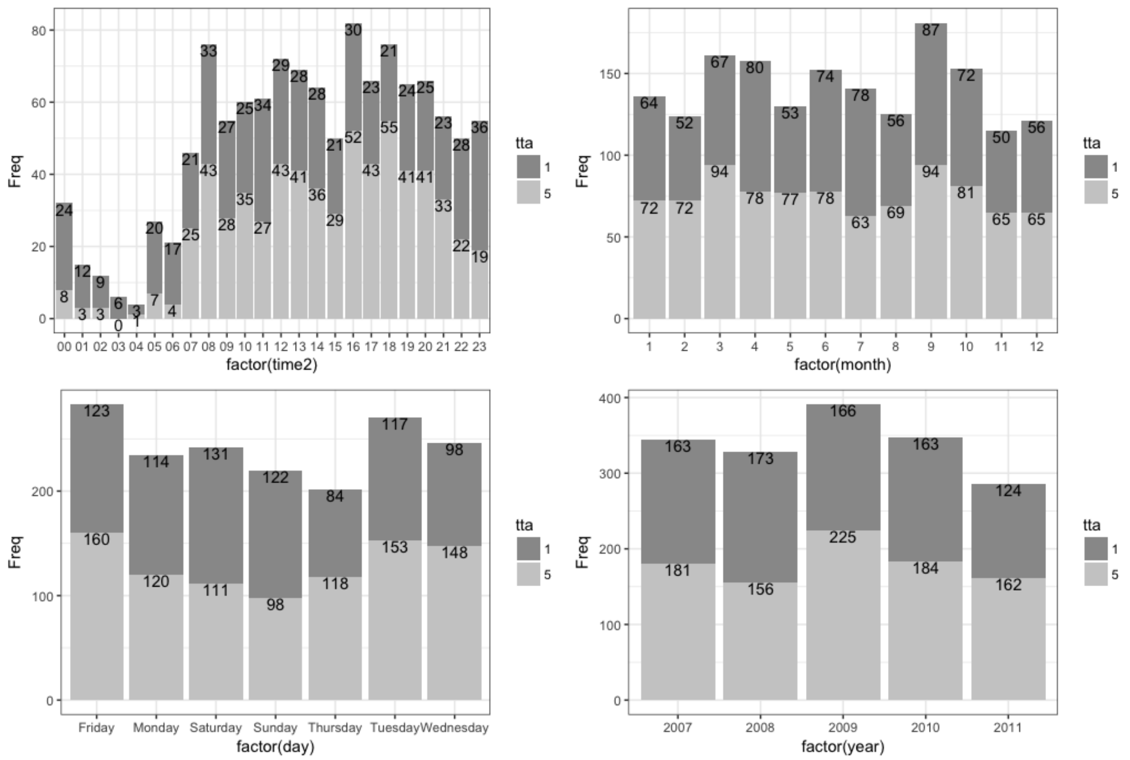

Figure 1. *Traffic accident distribution during the day, weekdays, month, and year*

V. CONCLUSIONS AND RECOMMENDATIONS

Based on observations it can be concluded that two types of road traffic accidents occurred in Tashkent city. They are road traffic accident type 1 (collision) and type 5 (collision with pedestrian). The other type of road traffic accidents was drop out of analysis due to low number of occurrence. Visualization and data analysis of road traffic accidents led to make conclusions which would assist reduce the number of accidents. Following recommendations and measures offered to prevent accidents:

1. Develop a functional classification of roads and streets that will take into account proper speed limits, currently speed limit is 70 km/h in urban area, which is significantly higher compared to developed countries, where speed limit varies from 40 to 50 km/h depending on functional purpose.

2. Prohibit pedestrian crossings on one level on main streets and roads within the city.

3. Increase visibility of pedestrian crossings in day and night, utilize modern traffic equipment.

4. Take measures to calm traffic on roads and streets with low traffic.

5. Development and provision of legal documents on road rules and guidelines. Public awareness campaigns that raise awareness of the risks and penalties for breaking the law. Supporting the enforcement of legislative measures. Adoption and enforcement of internationally harmonized laws requiring the use of seat belts, helmets and safety equipment for children.

6. Urban and transport planning. Population mobility studies, which assists properly organize public transportation and complement the competing modes of transport. Promotion of public transport (BRT - Bus Rapid Transit).

7. Designing safer roads and establishing an independent road safety audit requirement at the project stage (road safety audit);

8. Training specialists in urban and transport planning plays an important role. At present, highway engineers are engaged in transportation planning, but for the effective and efficient organization of transport planning, appropriate specialists should be trained.

In summary, road traffic accident data analysis and visualization assist transportation engineers and police officers to make proper conclusions by observing and drawing conclusions from hour wise, day wise, month wise and year wise plots. The research may further expanded by observing how weather conditions and location would effect to the road traffic accident frequency.

Appendix A. R code for road traffic data visualization

```
library(ggplot2)
# Multiple plot function
# ggplot objects can be passed in ..., or to plotlist (as a list of ggplot objects)
# - cols:   Number of columns in layout
# - layout: A matrix specifying the layout. If present, 'cols' is ignored.
# If the layout is something like matrix(c(1,2,3,3), nrow = 2, byrow = TRUE),
# then plot 1 will go in the upper left, 2 will go in the upper right, and
# 3 will go all the way across the bottom.
multiplot <- function(..., plotlist=NULL, file, cols=1, layout=NULL) {
```

```
  library(grid)
  # Make a list from the ... arguments and plotlist
  plots <- c(list(...), plotlist)
  numPlots = length(plots)
  # If layout is NULL, then use 'cols' to determine layout
  if (is.null(layout)) {
    # Make the panel
    # ncol: Number of columns of plots
    # nrow: Number of rows needed, calculated from # of cols
    layout <- matrix(seq(1, cols * ceiling(numPlots/cols)),
                    ncol = cols, nrow = ceiling(numPlots/cols))
  }
  if (numPlots==1) {
    print(plots[[1]])
  } else {
    # Set up the page
    grid.newpage()
    pushViewport(viewport(layout = grid.layout(nrow(layout), ncol(layout))))
    # Make each plot, in the correct location
    for (i in 1:numPlots) {
      # Get the i,j matrix positions of the regions that contain this subplot
      matchidx <- as.data.frame(which(layout == i, arr.ind = TRUE))
         print(plots[[i]], vp = viewport(layout.pos.row = matchidx$row,
                                    layout.pos.col = matchidx$col))
    }
  }
}
# Here we plot time wise, day wise, month wise and year wise
# tta is type of traffic accident, res1 is data without any missing value
y = xtabs(~ day + tta, res1)
p1 <- ggplot(as.data.frame(y),aes(x=factor(day),y=Freq,fill=tta)) +
  geom_bar(stat="identity",position="stack")+
  geom_text(aes(label=Freq),position="stack",vjust=1)+
  scale_fill_manual(values=c("grey60","grey80"))+
  theme_bw()
y1 = xtabs(~ month + tta, res1)
```

```
p2 <- ggplot(as.data.frame(y1),aes(x=factor(month),y=Freq,fill=tta)) +
  geom_bar(stat="identity",position="stack")+
  geom_text(aes(label=Freq),position="stack",vjust=1)+
  scale_fill_manual(values=c("grey60","grey80"))+
  theme_bw()
y2 = xtabs(~ year + tta, res1)
p3 <- ggplot(as.data.frame(y2),aes(x=factor(year),y=Freq,fill=tta)) +
  geom_bar(stat="identity",position="stack")+
  geom_text(aes(label=Freq),position="stack",vjust=1)+
  scale_fill_manual(values=c("grey60","grey80"))+
  theme_bw()
y3 = xtabs(~ time2 + tta, res_no_na)
p4 <- ggplot(as.data.frame(y3),aes(x=factor(time2),y=Freq,fill=tta)) +
  geom_bar(stat="identity",position="stack")+
  geom_text(aes(label=Freq),position="stack",vjust=1)+
  scale_fill_manual(values=c("grey60","grey80"))+
  theme_bw()
multiplot(p4,p1,p2,p3, cols = 2)
```

References

[1]. *Ansari, S. F Akhdar, M. Mandoorah, K. Moutaery*, (2000) Causes and effects of road traffic accidents in Saudi Arabia, Public Health, Volume 114, Issue 1, pp. 37-39.
[2]. Highway Standard of Uzbekistan (2007), МКН 15-2007 «Правила учета и анализа дорожно-транспортных происшествий на автомобильных дорогах», Ташкент 2007, 30 стр.
[3]. *Hirasawa, M. and M. Asano,* (2003). Development of Traffic Accident Analysis System Using GIS. Proceedings of the Eastern Asia Society for Transportation Studies. Vol.4, pp.1193-1199.
[4]. *Miao C., A. Ajith, P. Marcin,* (2005) Traffic Accident Analysis Using Machine Learning Paradigms. Informatica 29, pp. 89–98
[5]. *Pooja S., R. Ahsan, N. Dhapekar and G. Bhatt*. Hypothesis Testing of Road Traffic Accident Data in India. International Journal of Civil Engineering and Technology, 8(6), 2017, pp. 430–435.
[6]. *Sodikov J.* (2017) Электронный Учет Дорожно-Транспортных Происшествий С Применением Геоинформационных Технологий. Вестник Ташкентского Архитектурно-Строительного Института, Номер 2.
[7]. *Sodikov J.* (2017) Визуализация ДТП с помощью Google Fusion Tables. Вестник ТАДИ, Номер 2.
[8]. *Syyed, A. and K. Afaq*, (2013) Road Traffic Accident Analysis of Motorways in Pakistan. International Journal of Engineering and Technical Research, 2(11).
[9]. *Tihonov M.* (2016). Ежегодно в результате ДТП в Узбекистане погибает около 2 тыс. человек. Онлайн агентство новостей. Http://podrobno.uz/cat/proisshestviya/ezhegodno-v-rezultate-dtp-v-uzbekistane-pogibaet-okolo-2-tys-chelovek-/Last accessed: 11/28/17.