Road Accident Data Analysis



**Mathematics For Engineers- II**

PROJECT REPORT

SUBMITTED IN PARTIAL FULFILLMENT REQUIREMENT FOR THE AWARD OF

DEGREE OF

**BACHELOR OF TECHNOLOGY**

SUBMITTED BY

RAGHAV ARORA 220370

VIRESH CHAUHAN 220422

NISHANT 220313

ARYYAN AHLAWAT 220338

UNDER THE SUPERVISION OF

Dr. Ranjib Banarjee

**SCHOOL OF ENGINEERING AND TECHNOLOGY**

Logo, company name

Description automatically generated

**BML MUNJAL UNIVERSITY, Gurugram, Haryana-112413**

# **ACKNOWLEDGEMENT**

We would like to extend our sincere gratitude to everyone who helped us complete this project. The completion of this project would not have been possible without their assistance, direction, and encouragement.

First, we want to express our gratitude to Dr. Ranjib Banerjee for his valuable direction, criticism, and ideas throughout the project. His knowledge and perspective were crucial in helping us refine our ideas and make sure our work was of the highest calibre.

Additionally, we want to thank BML Munjal University's faculty and staff for creating a stimulating and encouraging learning environment for us. Their suggestions and criticism were extremely helpful in determining the project's course.

We would also like to express our gratitude to the study's subjects and participants for their participation and contributions. Their participation and cooperation were essential to the project's success.

We also want to express our gratitude to the administrative and technical staff for giving us the tools and assistance we needed to complete this project. They were accommodated with administrative, equipment, and logistical tasks.

Finally, we thank our families and friends for their unwavering encouragement and support throughout the project. Their moral assistance, tolerance, and comprehension have been crucial in helping us concentrate on this project and finish it successfully.

Once again, we extend our sincere thanks to all those who have contributed to this project. Your support and encouragement have been valuable, and we are grateful for all that you have done.

**TABLE OF CONTENTS**

1. **INTRODUCTION …………………………………. 4**
2. **METHODOLOGY ………………………………… 5**

**2.1. Collecting the relevant data**

**2.2. Categorizing and visualization of the data**

**2.3. Creating a prediction model**

**2.4. Creating and testing hypothesis**

**2.5. Relation between the various seasons of a year**

1. **RESULT AND ANALYSIS………………………… 16**

**3.1. Season-based differentiation**

**3.2. Line plot of the total number of accidents over the year**

**3.3. Time-based differentiation**

**3.4. Prediction made by the model**

**3.5 Hypothesis testing**

1. **CONCLUSION……………………………………. 31**
2. **REFERENCES ……………………………………. 32**

**INTRODUCTION**

With our country's population increasing day by day, it is only natural that the traffic we observe on a daily basis should also only increase. Thus, this rising road population, coupled with some other factors such as mediocre road facilities, poor traffic management, and minimal enforcement of the traffic laws by the police, will naturally give rise to more and more road accident cases in India. Analysing road accident data helps identify common risk factors and patterns associated with accidents. This information can be used to develop targeted interventions and preventive measures to mitigate those risks. Through our report, we wish to analyse road accident data from the year 2001 to 2014 and realize/visualize the other factors that might lead to road accidents in India, which can then be addressed further by the concerned authorities, by implementing some countermeasures. For example, the time of accidents can give us some suggestions as to the time of the day the police should be more vigilant in an area, etc. Hence, it can help the authorities in allocating resources effectively to address the areas of greatest need, at the required times. Moreover, this analysis can also enable us to single out the effective traffic-related regulations in an area by looking at the simple trend of road accidents in that region. If a particular policy is effective, the number of accidents should decrease in that area. Finally, we can also use the data and graphs/plots generated to raise public awareness about road accidents.

**METHODOLOGY**

1. **Collecting the relevant data.**

To analyse and make firm predictions related to the road accidents topic (in India) we needed to collect trustworthy, and unbiased data, which we were able to acquire through a site [1], mentioned in the references part of this report. To make the required predictions, we needed relatively recent data regarding the number of accidents in particular seasons for all the states in India, and another dataset highlighting the time of the accidents that have taken place throughout the years. The particular datasets we were able to get, checked all these boxes, being fairly recent (2001-2014), and having some other relevant information as well.

1. **Categorization and Visualization of the data.**

We selected 5 states on the basis of the frequency of road accidents that takes place over that particular state (five states with the most accidents) and proceeded to sort the data into different vectors, depending on the season for the first dataset, while the time for the second, with the help of R. Further, we visualized the data in the form of boxplots, bar graphs, and even pie charts for better understanding. The following code represents these actions:

1. **For dividing the data on the basis of seasons:**

#import the librarires

library(magrittr)

library(dplyr)

library(ggplot2)

#import the data

path <-'C:\\Users\\Aryyan\\Downloads\\only\_road\_accidents\_data\_month2.csv'

#creating a dataframe for the data

road\_accidents\_month<- read.csv(path)

#creating a dataframe for grouping the states

Total\_acidents\_for\_states\_14Y<-road\_accidents\_month%>%group\_by(STATE.UT)%>%summarise(across(c(JANUARY,FEBRUARY,MARCH,APRIL,MAY,JUNE,JULY,AUGUST,SEPTEMBER,OCTOBER,NOVEMBER, DECEMBER), sum))

#sorting the data into Summer, Spring,Autumn and Winter

Data\_season\_wise<-Total\_acidents\_for\_states\_14Y|>group\_by(SEPTEMBER,OCTOBER,NOVEMBER)|>mutate(Autumn=sum(SEPTEMBER,OCTOBER,NOVEMBER))

Data\_season\_wise<-Data\_season\_wise|>group\_by(AUGUST,JULY,JUNE)|>mutate(Summer=sum(AUGUST,JULY,JUNE))

Data\_season\_wise<-Data\_season\_wise|>group\_by(MARCH,APRIL,MAY)|>mutate(Spring=sum(MARCH,APRIL,MAY))

Data\_season\_wise<-Data\_season\_wise|>group\_by(DECEMBER,JANUARY,FEBRUARY)|>mutate(Winter=sum(DECEMBER,JANUARY,FEBRUARY))

Data\_season\_wise<-Data\_season\_wise|>group\_by(Summer,Winter,Autumn,Spring)|>mutate(TOTAL=sum(Summer,Spring,Autumn,Winter))

#creating vectors for Summer, Spring,Autumn and Winter as Percentage

Data\_season\_wise$`SUMMER\_P`<-Data\_season\_wise$Summer/Data\_season\_wise$TOTAL \*100

Data\_season\_wise$`SPRING\_P`<-Data\_season\_wise$Spring/Data\_season\_wise$TOTAL \*100

Data\_season\_wise$`WINTER\_P`<-Data\_season\_wise$Winter/Data\_season\_wise$TOTAL \*100

Data\_season\_wise$`Autumn\_P`<-Data\_season\_wise$Autumn/Data\_season\_wise$TOTAL \*100

#plotting a boxplot to check for median and compare all the various seasons

boxplot(Data\_season\_wise$SUMMER\_P, Data\_season\_wise$SPRING\_P,Data\_season\_wise$WINTER\_P,Data\_season\_wise$Autumn\_P, ylab="Percentage of accidents", labels=c("SUMMER\_P", "SPRING\_P","WINTER\_P", "AUTUMN\_P" ))

#creating a pie graph for composition according to seasons

pie(colSums(Data\_season\_wise[, c("Summer", "Winter", "Autumn", "Spring")]), labels=c("SUMMER", "WINTER", "AUTUMN", "SPRING"), main="Seasonal distribution of all accidents in India(2001-14)")

#sorting the data according to number of accidents

summer\_sorted <- Data\_season\_wise[order(Data\_season\_wise$`Summer`, decreasing = TRUE),]

Autumn\_sorted <- Data\_season\_wise[order(Data\_season\_wise$`Autumn`, decreasing = TRUE),]

winter\_sorted <- Data\_season\_wise[order(Data\_season\_wise$`Spring`, decreasing = TRUE),]

spring\_sorted <- Data\_season\_wise[order(Data\_season\_wise$`Winter`, decreasing = TRUE),]

#plotting bar graphs for the sorted data for the top 5 states

barplot(sort(Data\_season\_wise$`Summer`, decreasing = TRUE)[1:5], names.arg =(summer\_sorted$STATE.UT[1:5]),main = "Highest Summer Accidents")

barplot(sort(Data\_season\_wise$`Winter`, decreasing = TRUE)[1:5],names.arg =(winter\_sorted$STATE.UT[1:5]), main = "Highest Winter Accidents")

barplot(sort(Data\_season\_wise$Autumn, decreasing = TRUE)[1:5], names.arg =(Autumn\_sorted$STATE.UT[1:5]),main = "Highest Autumn Accidents")

barplot(sort(Data\_season\_wise$`Spring`, decreasing = TRUE)[1:5], names.arg =(spring\_sorted$STATE.UT[1:5]),main = "Highest Spring Accidents")

#sorting and plotting a line graph for the top 5 states to see increase in total accidents highest\_accident\_states <- Data\_season\_wise[order(Data\_season\_wise$TOTAL, decreasing = TRUE),]

high\_states <- highest\_accident\_states$STATE.UT[1:5]

df4 <- road\_accidents\_month %>% filter(STATE.UT %in% high\_states) %>% select(STATE.UT, YEAR, TOTAL)

ggplot(df4, aes(x = YEAR, y = TOTAL, group = STATE.UT, color = STATE.UT)) + geom\_line() + theme\_bw()

1. **For dividing the data on the basis of time of the accidents:**

#import the libraries

library(magrittr)

library(dplyr)

#import the data

path1 <-'C:\\Users\\Aryyan\\Downloads\\only\_road\_accidents\_data3.csv'

#creating a dataframe for the data

road\_accidents\_time<- read.csv(path1)

#creating a dataframe for grouping the states

Total\_acidents\_for\_states\_14Y\_time<-road\_accidents\_time %>% group\_by(STATE.UT)%>%summarise(across(c(X0.3.hrs...Night.,X3.6.hrs...Night.,X6.9.hrs..Day.,X9.12.hrs..Day.,X12.15.hrs..Day.,X15.18.hrs..Day.,X18.21.hrs..Night.,X21.24.hrs..Night.), sum))

#sorting the data into Morning, Afternoon,Evening and Night time

Data\_time\_wise<-Total\_acidents\_for\_states\_14Y\_time |> group\_by(X0.3.hrs...Night.,X3.6.hrs...Night.,X21.24.hrs..Night.)|>mutate(Night=sum(X21.24.hrs..Night.,X0.3.hrs...Night.,X3.6.hrs...Night.))

Data\_time\_wise<-Data\_time\_wise|> group\_by(X18.21.hrs..Night.)|>mutate(Evening=sum(X18.21.hrs..Night.))

Data\_time\_wise<-Data\_time\_wise|> group\_by(X6.9.hrs..Day.,X9.12.hrs..Day.)|>mutate(Morning=sum(X6.9.hrs..Day.,X9.12.hrs..Day.))

Data\_time\_wise<-Data\_time\_wise|> group\_by(X12.15.hrs..Day.,X15.18.hrs..Day.)|>mutate(Afternoon=sum(X12.15.hrs..Day.,X15.18.hrs..Day.))

#creating a vector for total accidents

Data\_time\_wise<-Data\_time\_wise |> group\_by(Night,Evening,Morning)|>mutate(Total=sum(Morning,Night,Evening,Afternoon))

#creating vectors for Morning, Afternoon, Evening and Night as Percentage

Data\_time\_wise$`Morning\_P` <- Data\_time\_wise$Morning / Data\_time\_wise$Total \*100

Data\_time\_wise$`Afternoon\_P` <- Data\_time\_wise$Afternoon / Data\_time\_wise$Total \*100

Data\_time\_wise$`Evening\_P` <- Data\_time\_wise$Evening / Data\_time\_wise$Total \*100

Data\_time\_wise$`Night\_P` <- Data\_time\_wise$Night / Data\_time\_wise$Total \*100

#plotting a boxplot to check for median and compare all the various times of day

boxplot(Data\_time\_wise$Morning\_P, Data\_time\_wise$Night\_P,Data\_time\_wise$Afternoon\_P,Data\_time\_wise$Evening\_P)

#creating a pie graph for composition according to time of day

pie(colSums(Data\_time\_wise[, c("Evening\_P","Afternoon\_P","Morning\_P","Night\_P")]), labels=c("Evening", "Afternoon", "Morning", "Night"), main="Distribution of all accidents in India according to time(2001-14)")

#sorting the data according to number of accidents

morning\_sorted <- Data\_time\_wise[order(Data\_time\_wise$Morning, decreasing = TRUE),]

afternoon\_sorted <- Data\_time\_wise[order(Data\_time\_wise$Afternoon, decreasing = TRUE),]

evening\_sorted <- Data\_time\_wise[order(Data\_time\_wise$Evening, decreasing = TRUE),]

night\_sorted <- Data\_time\_wise[order(Data\_time\_wise$Night, decreasing = TRUE),]

#plotting the sorted data as bar plot to see the top states

barplot(sort(Data\_time\_wise$Morning, decreasing = TRUE)[1:5], names.arg =(morning\_sorted$STATE.UT[1:5]),main = "Highest Morning-Time Accidents")

barplot(sort(Data\_time\_wise$Afternoon, decreasing = TRUE)[1:5],names.arg =(afternoon\_sorted$STATE.UT[1:5]), main = "Highest Afternoon-Time Accidents")

barplot(sort(Data\_time\_wise$Evening, decreasing = TRUE)[1:5], names.arg =(evening\_sorted$STATE.UT[1:5]),main = "Highest Evening-time Accidents")

barplot(sort(Data\_time\_wise$Night, decreasing = TRUE)[1:5], names.arg =(night\_sorted$STATE.UT[1:5]),main = "Highest Night-Time Accidents")

#creating a dataframe for sorting the highest number of total accidents

highest\_accident\_states\_time <- Data\_time\_wise[order(Data\_time\_wise$Total, decreasing = TRUE),]

#getting the names of the top 5 states with highest accidents

high\_states\_time <- highest\_accident\_states\_time$STATE.UT[1:5]

df4\_time<- road\_accidents\_time %>% filter(STATE.UT %in% high\_states\_time) %>% select(STATE.UT, YEAR,Total)

#plotting a line graph to see progression over the years for each of the top states

library(ggplot2)

ggplot(df4\_time, aes(x = YEAR, y = Total, group = STATE.UT, color = STATE.UT)) + geom\_line() + theme\_bw()

#plotting the line graph for total accidents per year

Total\_over\_years<-road\_accidents\_month %>% group\_by(YEAR)%>%summarise(across(c(TOTAL), sum))

plot(Total\_over\_years, type = "o")

1. **Creating a Prediction Model:**

Finally, we proceeded with the creation of a linear regression model that would be able to predict the trend of road accidents in these five particular states for the data divided on the basis of seasonal accidents. We made a model capable of predicting the number of accidents in the coming years (from 2015-2022,) for all the states, and also 5 different models for each of the selected states. The following code performs the specified tasks:

#making a linear regression model to predict total number of accidents for the next few years

Total\_over\_years<-road\_accidents\_month %>% group\_by(YEAR)%>%summarise(across(c(TOTAL), sum))

model\_for\_total<-lm(TOTAL ~ YEAR ,data = Total\_over\_years)

new\_data <- tibble(YEAR= 2015:2022)

prediction<-predict(model\_for\_total, newdata = new\_data)

plot(prediction, type = "o")

print(prediction)

#making a prediction model for top 5 states

#Tamil Nadu

TN\_over\_years<-road\_accidents\_month[road\_accidents\_month$STATE.UT %in% c('Tamil Nadu'),]

model\_for\_TN<-lm(TOTAL ~ YEAR ,data = TN\_over\_years)

predict(model\_for\_TN, newdata=new\_data)

plot(predict(model\_for\_TN, newdata=new\_data))

#Maharashtra

MH\_over\_years<-road\_accidents\_month[road\_accidents\_month$STATE.UT %in% 'Maharashtra',]

model\_for\_MH<-lm(TOTAL ~ YEAR ,data = MH\_over\_years)

predict(model\_for\_MH, newdata=new\_data)

plot(predict(model\_for\_MH, newdata=new\_data))

#Karnataka

KN\_over\_years<-road\_accidents\_month[road\_accidents\_month$STATE.UT %in% 'Karnataka',]

model\_for\_KN<-lm(TOTAL ~ YEAR ,data = KN\_over\_years)

predict(model\_for\_KN, newdata=new\_data)

plot(predict(model\_for\_KN, newdata=new\_data))

#Andhra Pradesh

AP\_over\_years<-road\_accidents\_month[road\_accidents\_month$STATE.UT %in% 'Andhra Pradesh',]

model\_for\_AP<-lm(TOTAL ~ YEAR ,data = AP\_over\_years)

predict(model\_for\_AP, newdata=new\_data)

plot(predict(model\_for\_AP, newdata=new\_data))

#Kerala

KL\_over\_years<-road\_accidents\_month[road\_accidents\_month$STATE.UT %in% 'Kerala',]

model\_for\_KL<-lm(TOTAL ~ YEAR ,data = KL\_over\_years)

predict(model\_for\_KL, newdata=new\_data)

plot(predict(model\_for\_KL, newdata=new\_data))

1. **Creating and testing the Hypothesis:**

H0: There is no difference in the number of accidents between Morning and Night.

H1: There is a difference in the number of accidents between Morning and Afternoon.

# State the hypothesis

# Null hypothesis: There is no difference in the number of accidents between Morning and Night.

# Alternative hypothesis: There is a difference in the number of accidents between Morning and Afternoon.

#Significance level

alpha <- 0.05

#Calculate the test statistic

t.test(Morning\_accidents, Night\_accidents)

# Determine the p-value

p.value <- t.test(Morning\_accidents, Night\_accidents)$p.value

# Make a decision

if (p.value < alpha) {

# Reject the null hypothesis

print("There is a difference in the number of accidents between morning and afternoon.")

} else {

# Do not reject the null hypothesis

print("There is no difference in the number of accidents between morning and afternoon.")

}

1. **Finding a relation between the various seasons of a year:**

We can use a One-way ANOVA test (Analysis Of Variance) to test if there is a significant difference between the number of accidents happening throughout the various seasons of the year.

# Perform one-way ANOVA test

fit <- aov(TOTAL ~ Summer + Spring + Autumn + Winter, data = Data\_season\_wise)

summary(fit)

**RESULT AND ANALYSIS**

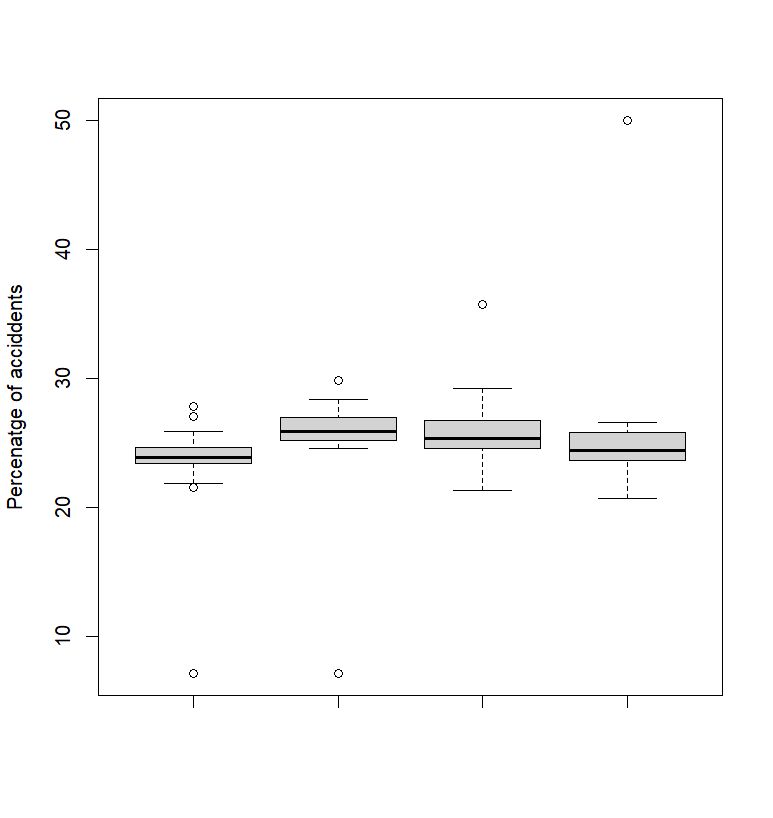
**For season-based differentiation of road accidents:**

The dataset provided contained the distribution of road accidents for every month from 2001 until 2014. By sorting the data on the basis of seasons, and through the visualization of data in the form of bar graphs and pie charts we could easily see and quantify what states have a greater number of road accidents. The pie chart, highlighting the percentage contribution of each season in the total number of accidents for all states, is approximately equal for each state. Thus, this can be loosely interpreted as an indication of the fact that seasons don’t usually have a relation with the number of accidents. Furthermore, the line graph provides us with a loose idea of the trend of road accidents going on in all of the states in India. To get a better idea of the same, we applied a linear regression model to the data:

1. For all the states, the model predicts that the number of road accidents will only increase until the specified year-2022.
2. While, for the 4 states out of 5 that we had earlier selected: Tamil Nadu, Maharashtra, Karnataka, and Andhra Pradesh, the model predicts that the trend will continue, and the number of accidents will only increase as the years go on. But, for Kerala, we can see that according to the model, the number of accidents will actually decrease as the years go by.

Following are the plots and other related graphs that we were able to get through the code:

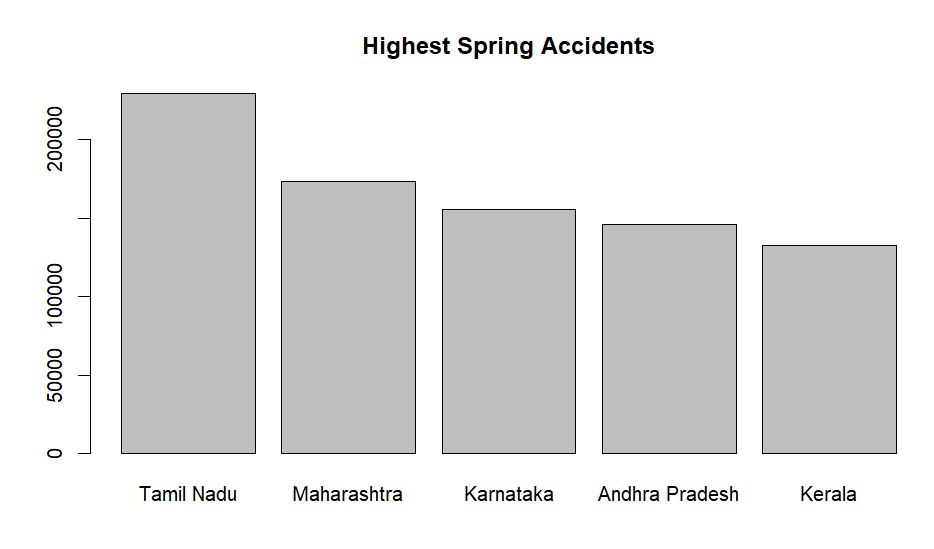
**The box plot:**

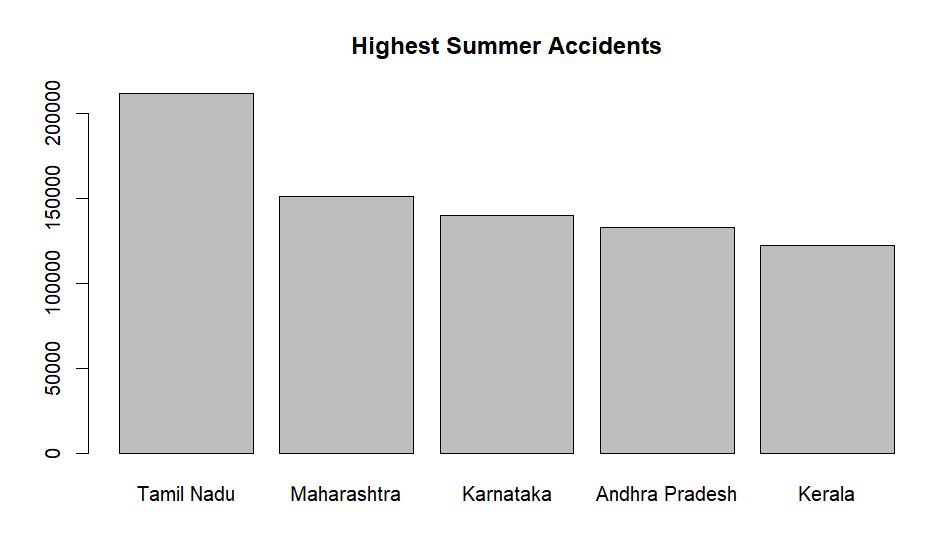


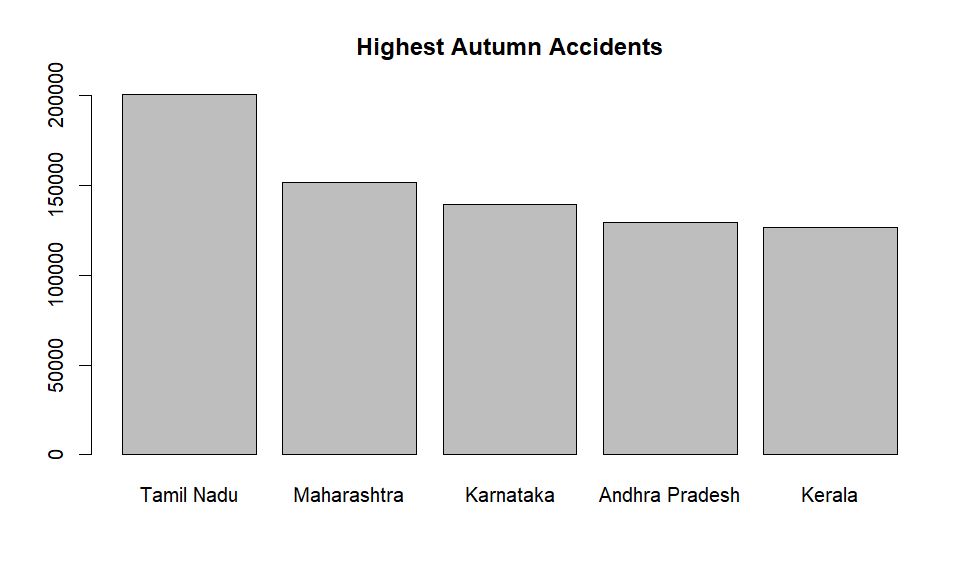
The plot above shows the boxplots of the percentage of accidents in summer, spring, winter, and autumn respectively.

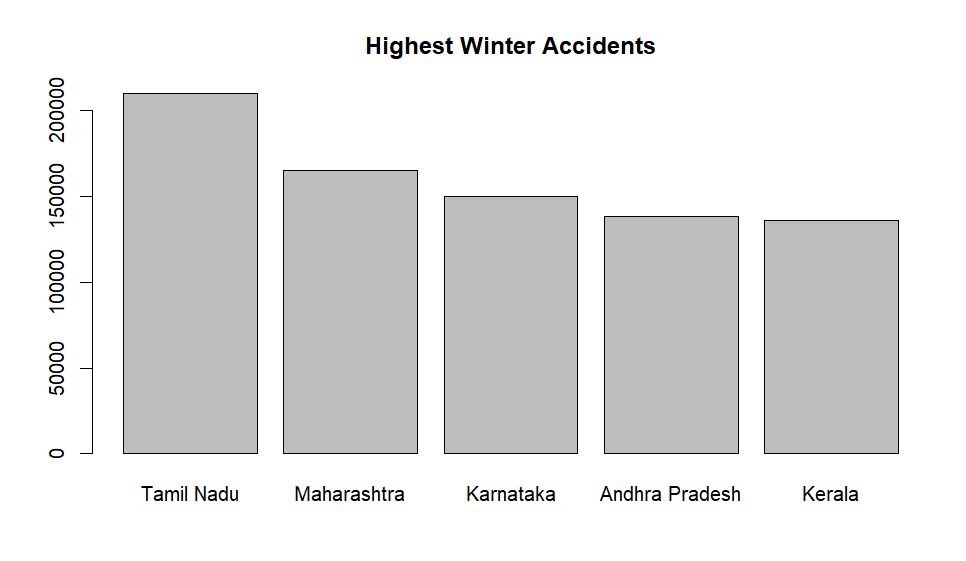
We can see from the plot that the median percentage of spring accidents is slightly higher than in other seasons.

**The Bar graphs:**





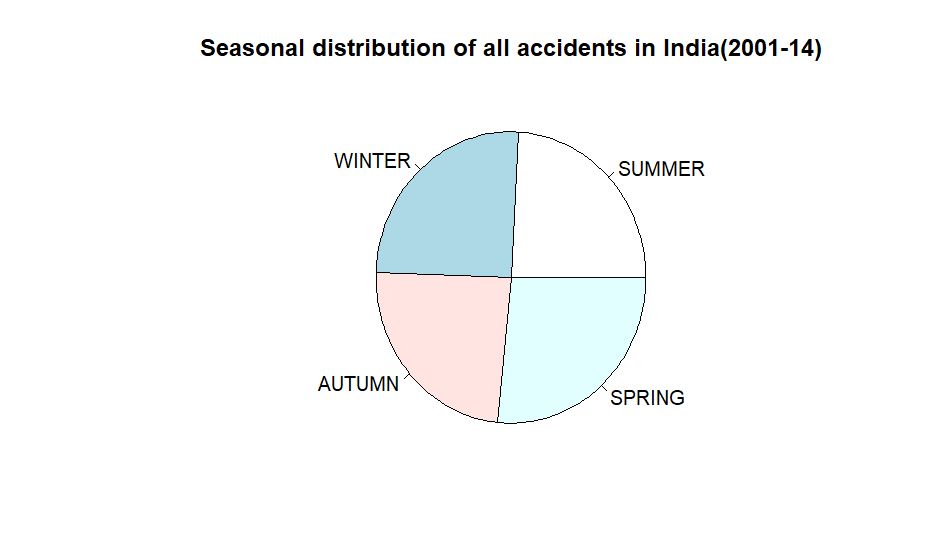




The bar graphs give us a summary of the top 5 states with the highest number of accidents in each season.

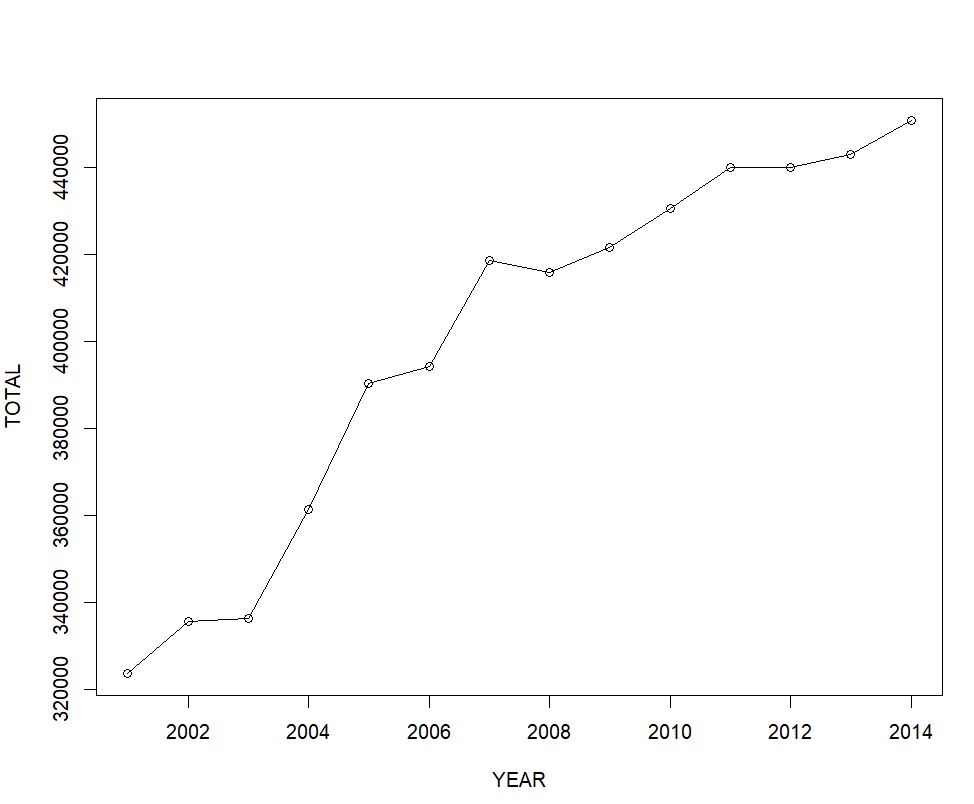
We observe that no matter what the season is, the top 5 states always remain the same with approximately the same range (somewhere around 1,50,00 to 2,00,00 or above) of accidents. This also supports our hypothesis that there is no significant difference between seasons and the number of accidents happening as stated and tested below.

**The Pie Chart:**



The Pie-Chart provides an estimate of the composition of various Seasons.

**Line plot of the total number of accidents over the year:**

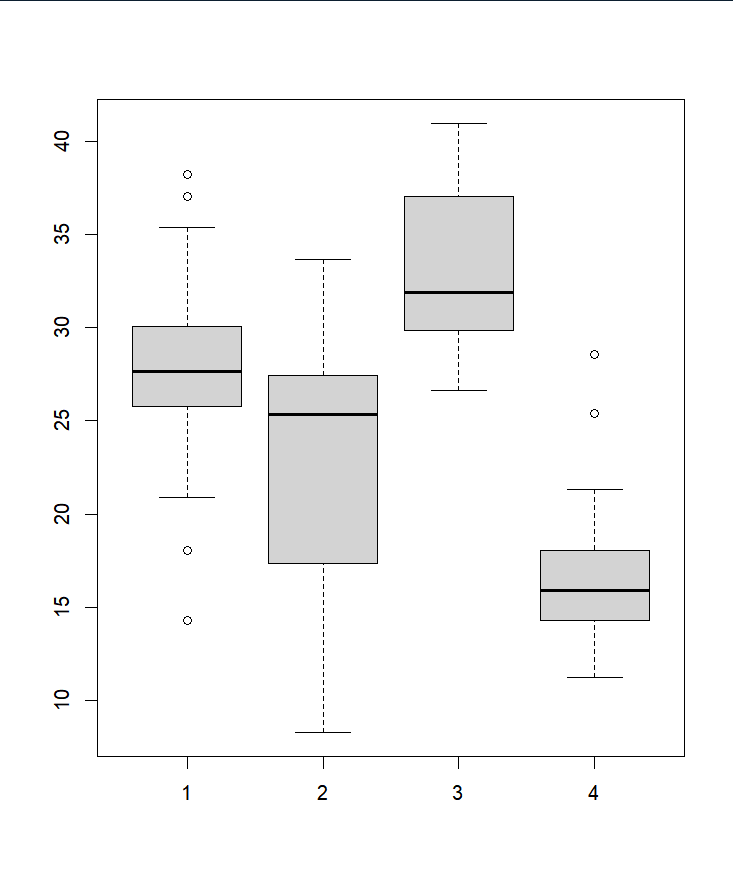


The Line graph shows the progression of the total number of accidents happening over all 35 states and UTs throughout the 14 years. We can use this data to observe the patterns of increase and decrease in the number of accidents and help to predict the outcome for the next several years.

**For time-based differentiation of data:**

Similar to the case with seasons, we were able to analyse the number of road accidents depending on the time of day by first sorting them into 4 different categories: afternoon, evening, night, and morning. Then, we visualized all the data in the form of a boxplot and pie chart, which gave us the median along with a loose idea of the distribution of accidents among the four categories. It is evident from the pie chart that the afternoon and morning proportion of accidents occupy most of the total accidents, thus marking these times as the 'peak' time for accidents to take place during a normal day. The bar graphs simply provide visual aid, helping us clearly quantify the frequency of accidents for each time of the day for the same five states. The line graph again provides us with the same insight as the previous, overall, one, just distributed over the five states. The followings are the graphs/plots:

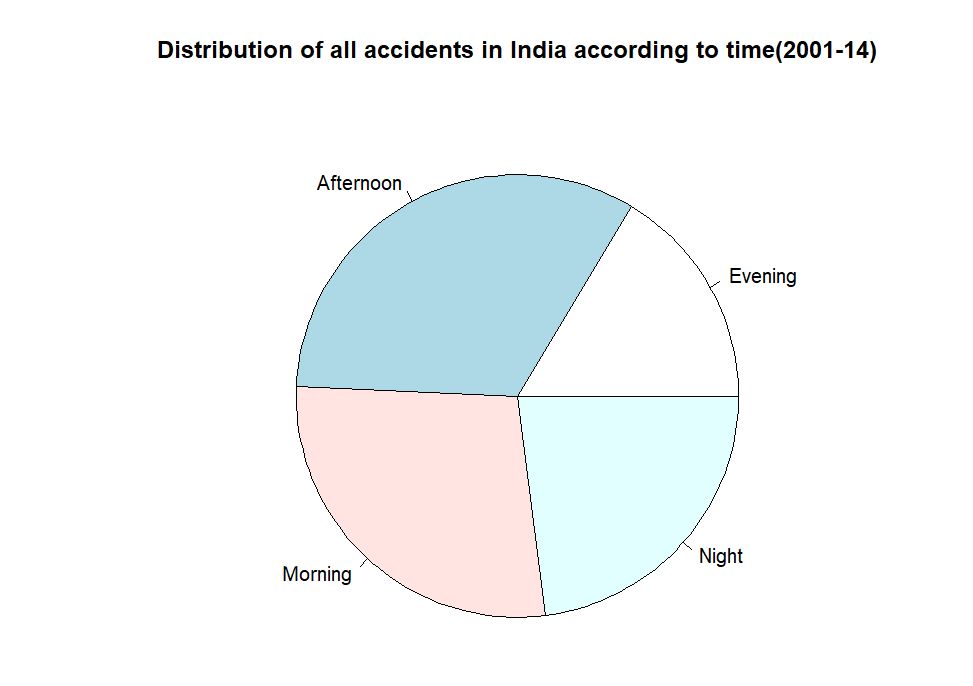
**The Box plot:**



(Here 1,2,3, and 4 represent the times of the day, with 1 = Morning, 2= Night, 3= Afternoon, and 4= Evening.)

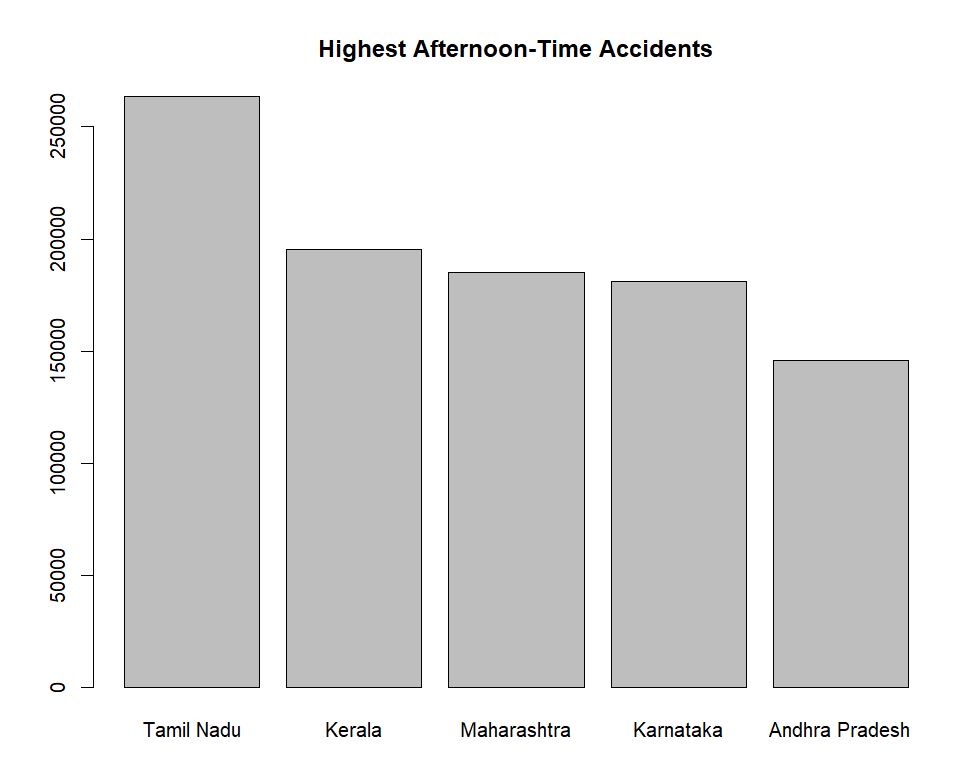
The bar plot shows us the distribution of percentages of accidents happening at various times of the day. We can observe that the median of the afternoon accidents is quite higher than the others while the median of the evening accidents is lower as compared to the others.

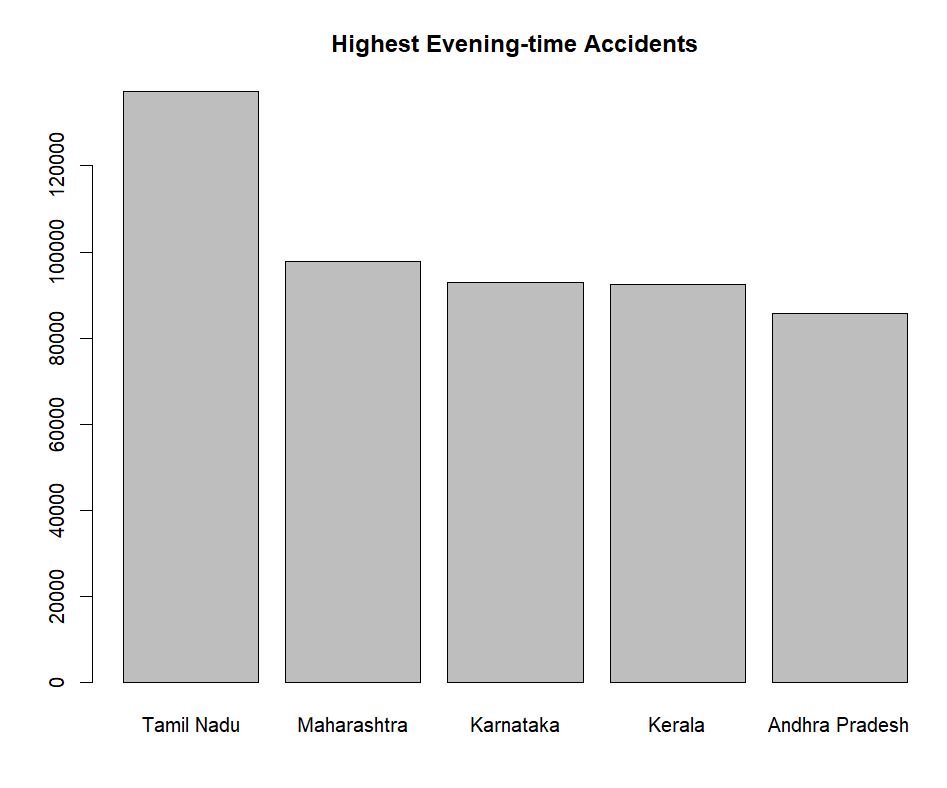
**The Pie Chart:**

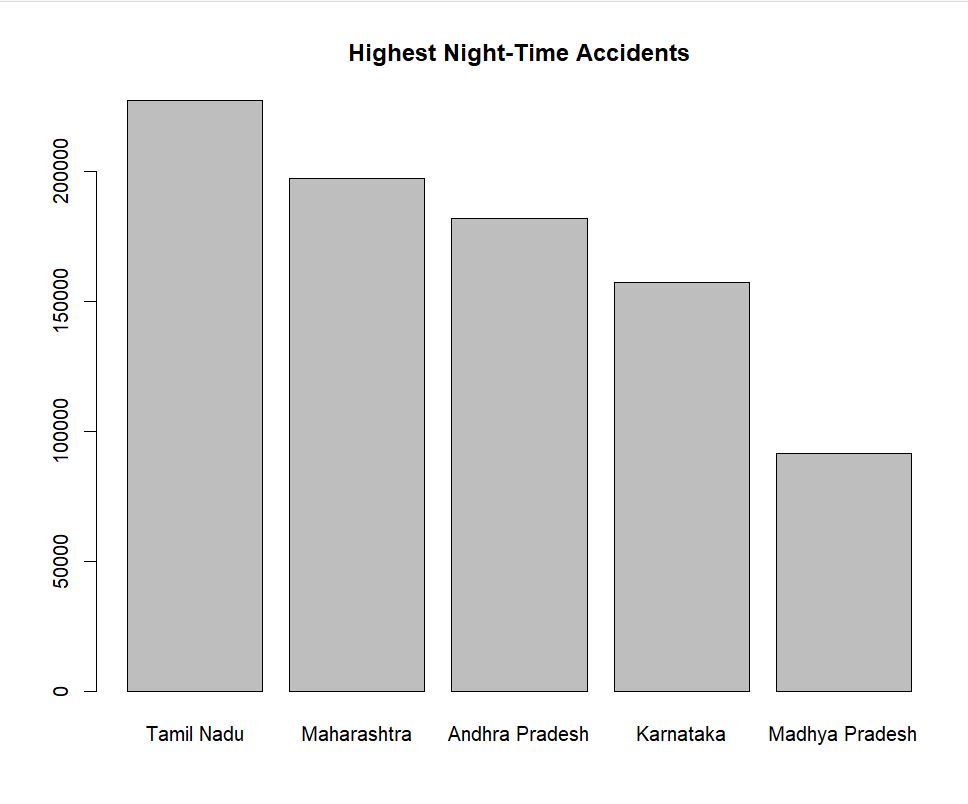


The Pie Chart gives us an estimate of the composition of accidents occurring at various times of the day and we can observe that the accidents happening in the morning and afternoon are significantly more than the ones happening at night and the lowest accidents are happening at the evening time.

**The Bar Graphs:**



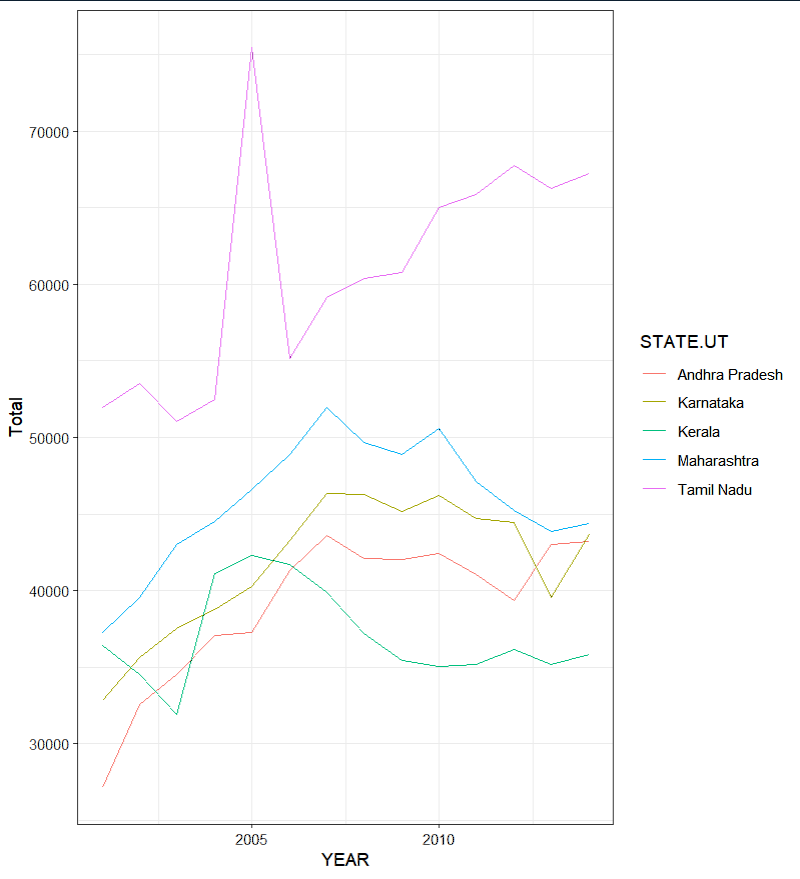




The bar graphs support our previous observations, the range of the number of accidents happening is the same even here and we observe that the states with the highest number of accidents stay the same, this validates our previous observations. However, the order of the highest accidents changes with the change in time. We also observe that Madhya Pradesh is also on the list for the highest night-time accidents.

**The Line Graph:**

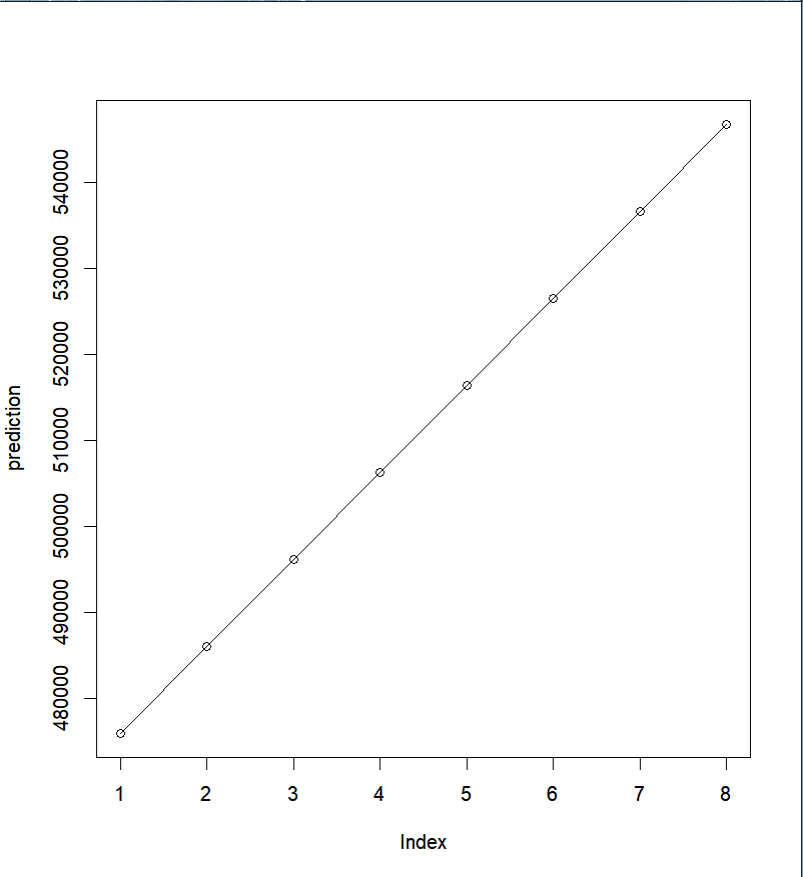
(The line graph(s) is unrelated to the time of the day, and only aims to establish a trend/pattern in the way accidents have been increasing/decreasing in the five selected states with the most accidents in India.)



By this line graph, we can see clearly that there was an initial spike in the number of accidents in each of these states around 2005 however that has been a steady and slow decline in the number of total accidents throughout almost all of these states except Tamil Nadu where there is still a significant increase in the number of accidents. This also supports our previous Graph where we observed that there has been a steady decline in the number of road accidents throughout the country.

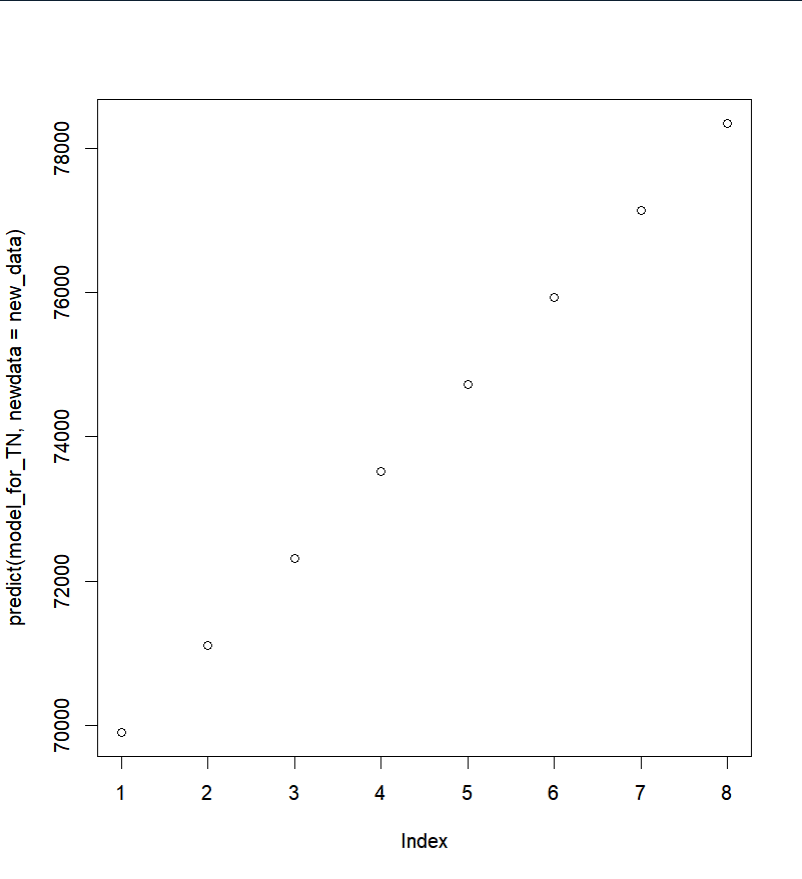
**The Predictions made by the model:**

Overall (for all states):

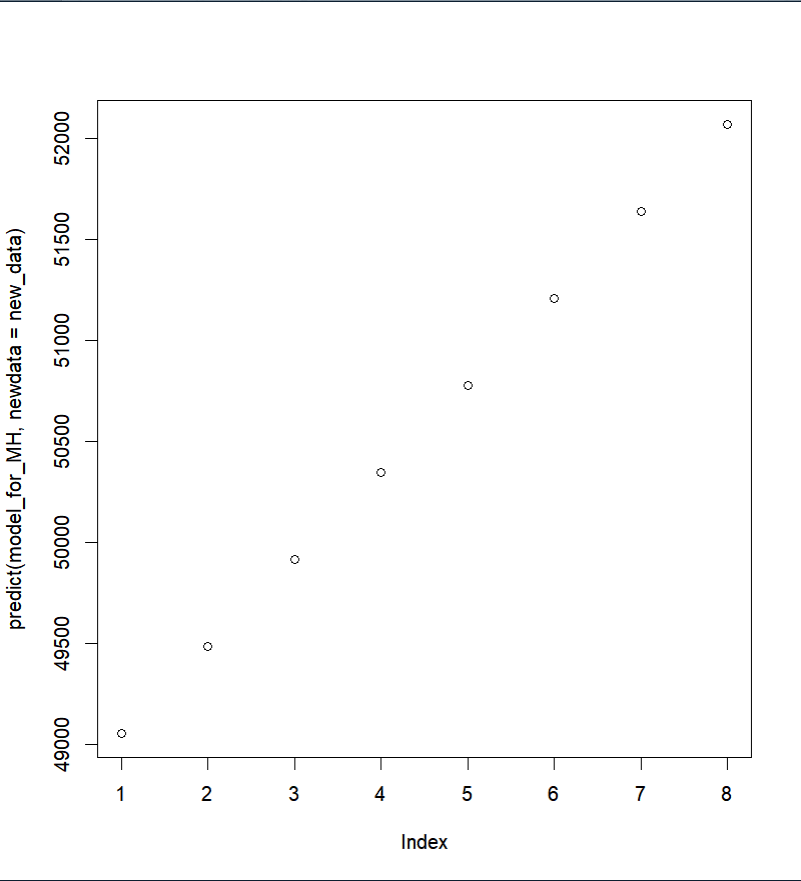


(Here, the Index represents the number of years since 2014, hence 1 =2015, 2=2016, and so on.)

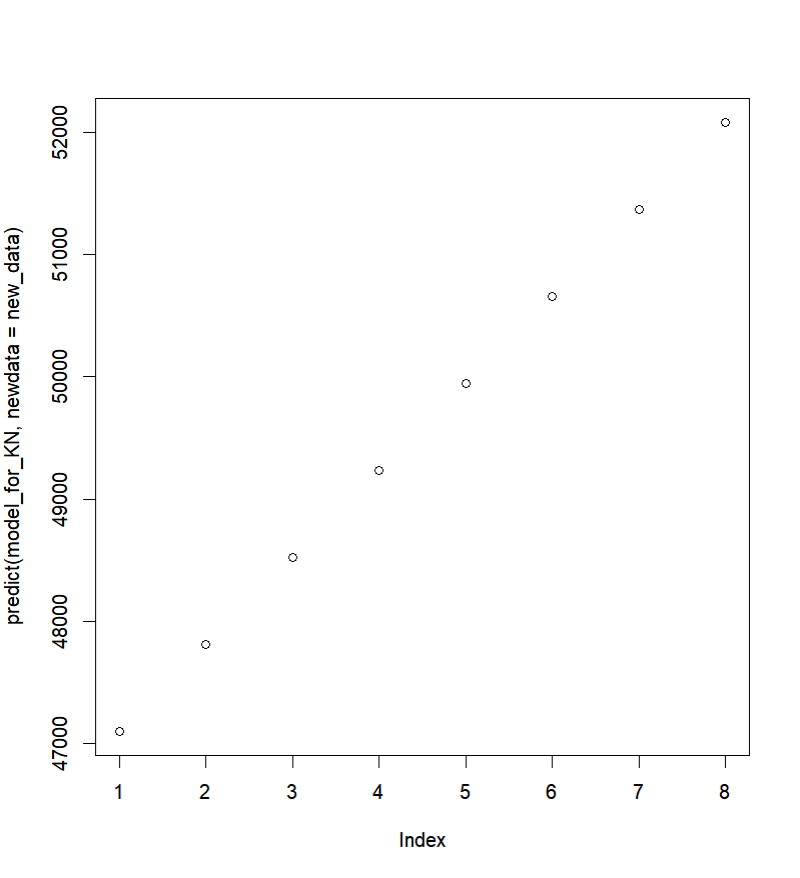
**For Tamil Nadu:**



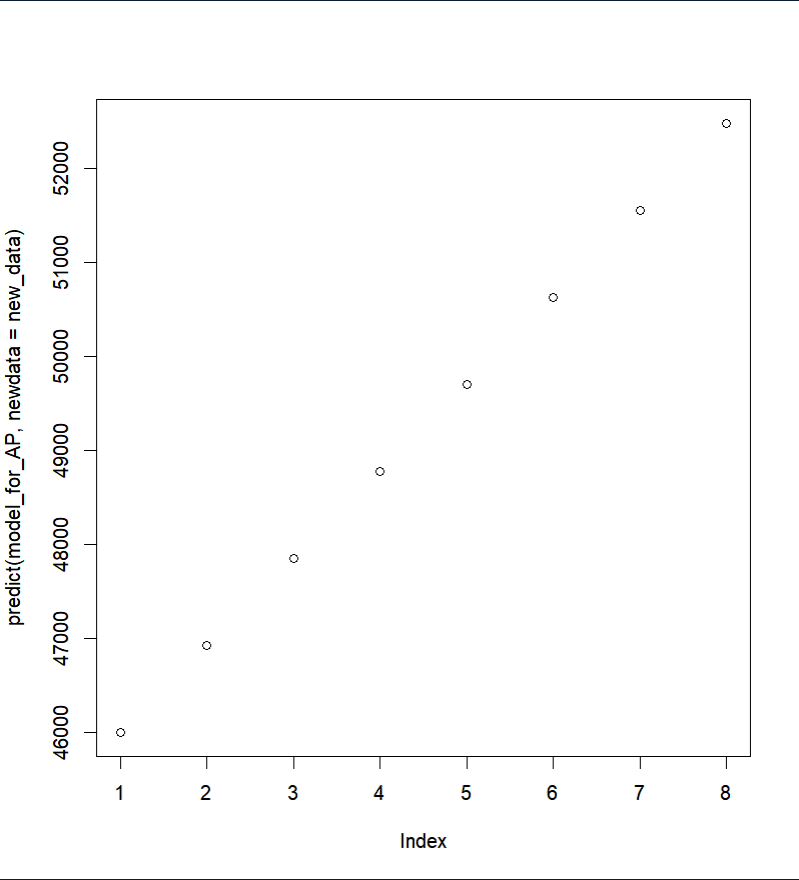
**For Maharashtra:**



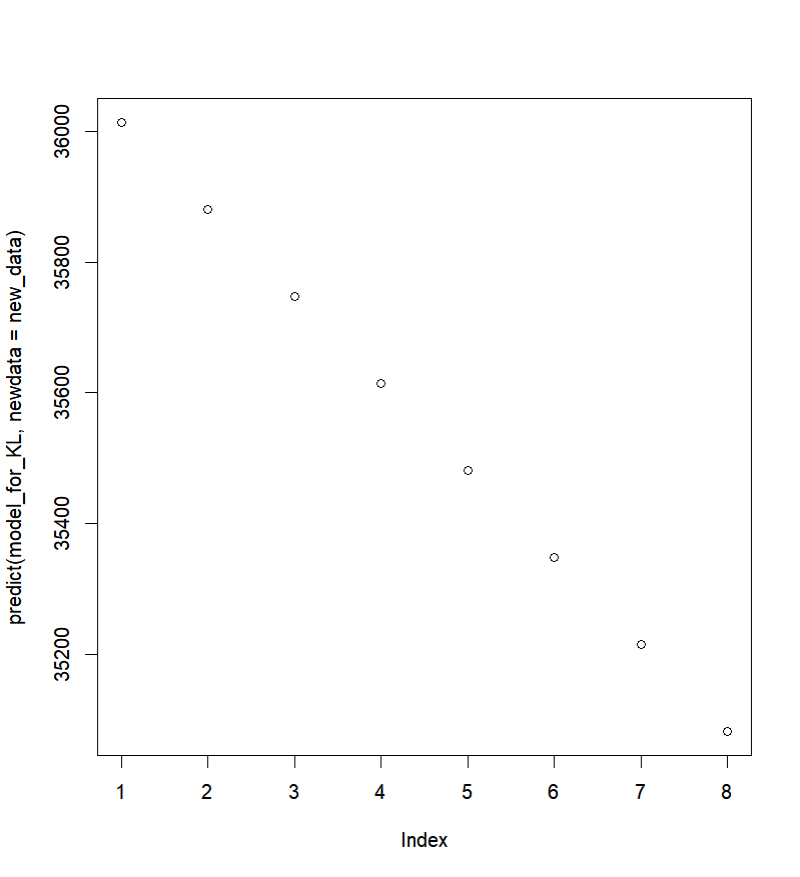
**For Karnataka:**



**For Andhra Pradesh:**



**For Kerala:**



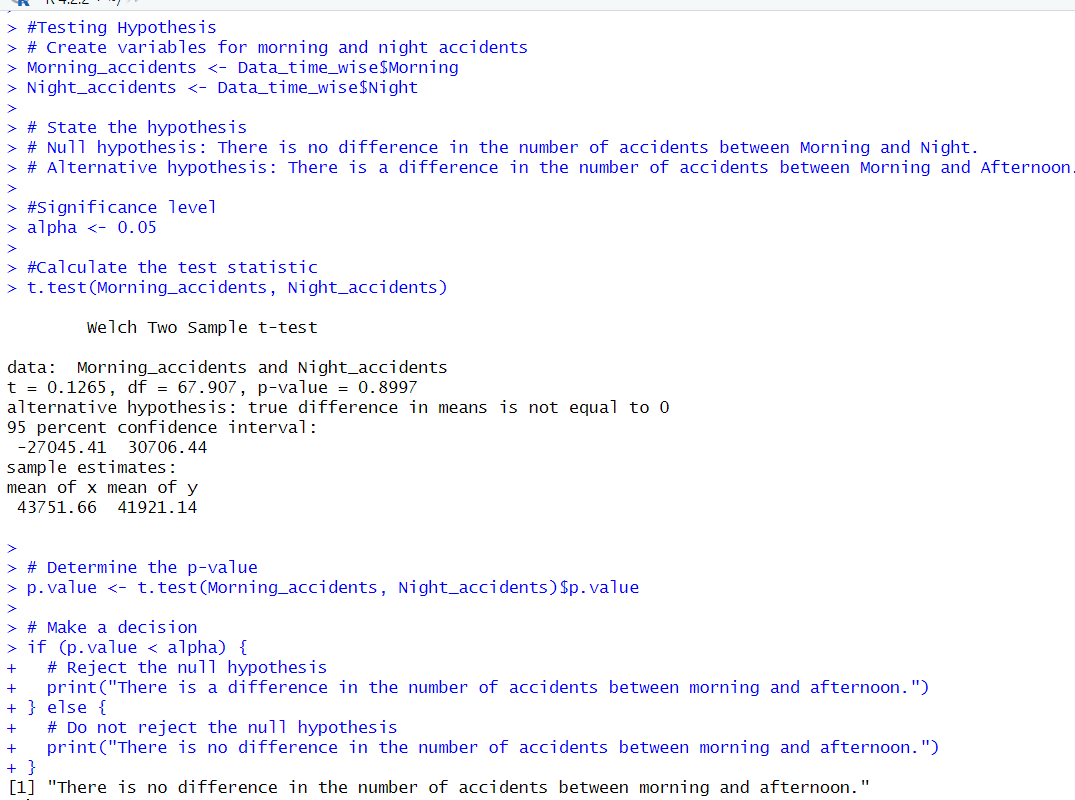
We are using a Linear Regression model that takes the data from the previous years (2001-2014) for each of these states and predicts the value for the next 8 years. We observe that compared to the last 14 years there is a decrease in the rate at which road accidents used to occur in almost all the top 5 states except for Kerala. In Kerala, we see there is an actual decrease in the number of accidents.

**Hypothesis Testing:**

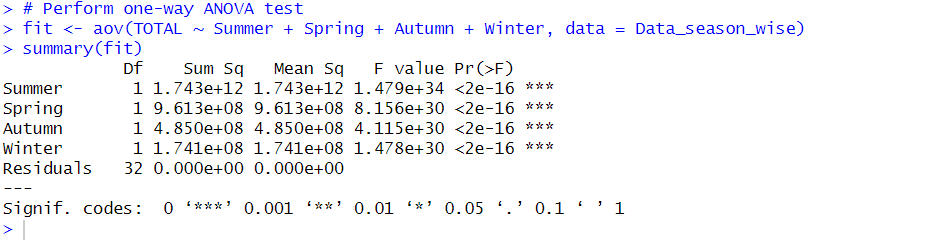
H0: There is no difference in the number of accidents between Morning and Night.

H1: There is a difference in the number of accidents between Morning and Afternoon.

**Results:**



We also performed an ANOVA test to see the relation between the various seasons of the year and observe if there is any significant difference between them.



Since the P value of all of the seasons came out to be less than our significance level 𝝰=0.05, we can say there isn’t any significant difference between the seasons.

**CONCLUSION**

In conclusion, our analysis of road accidents based on season and time of day provides important insights into the trends and patterns of road accidents in India. Our regression model predicts that the number of road accidents will continue to increase in the future, with four out of five states (Tamil Nadu, Maharashtra, Karnataka, and Andhra Pradesh) predicted to have more accidents due to an increase in the accident rate. However, the model predicted a decrease in accidents for Kerala. These findings suggest that there is an urgent need to address the underlying factors that contribute to road accidents in order to reduce their occurrence.

Our analysis also revealed that there is no significant difference in the number of accidents between seasons. This finding suggests that factors other than season, such as driver behaviour, road conditions, and vehicle quality, may have a greater impact on the occurrence of accidents. However, our analysis of the time of day revealed that there is a peak time for accidents during the morning and afternoon. This finding suggests that there is a need for increased focus on enforcing traffic rules and regulations during these periods to reduce the number of accidents.

To reduce the number of accidents during peak times, we recommend that the government impose more traffic rules and regulations during these periods. These rules could include speed limits, seat belt laws, and restrictions on mobile phone use while driving. It is also important to acknowledge potential limitations in our analysis, such as incomplete data or missing variables that may impact the accuracy of our findings. Therefore, further research is needed to better understand the underlying factors that contribute to road accidents.

Additionally, our findings have broader implications for the economic and social costs of road accidents, as well as the physical and mental well-being of individuals and families affected by them. Road accidents can have significant economic costs, such as medical expenses, property damage, and lost productivity. They can also have significant social costs, such as increased mortality rates, reduced quality of life, and reduced social cohesion. Therefore, reducing the number of accidents is not only important for improving road safety but also for promoting economic development and social well-being.

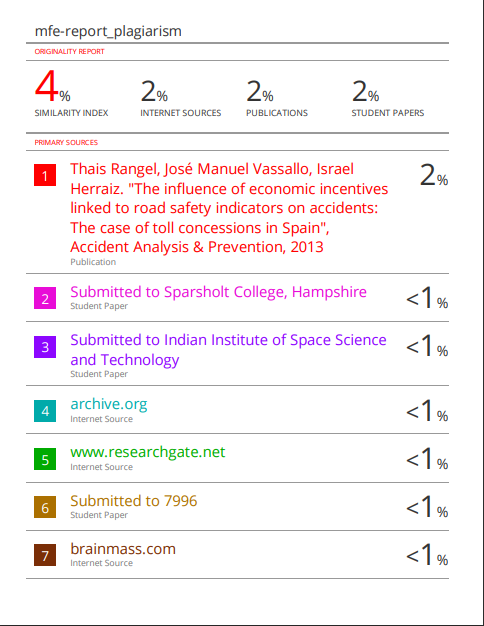
Future research could explore the reasons for the peak times of accidents and the types of accidents that occur during these times. Further investigation could also examine the effectiveness of different types of traffic rules and regulations in reducing accidents. For example, research could investigate the impact of stricter enforcement of traffic laws, public education campaigns, and technological interventions such as automated speed cameras.

Overall, our analysis provides important insights into the trends and patterns of road accidents in India and highlights the need for continued efforts to improve road safety. By implementing targeted interventions and conducting further research, we can reduce the number of accidents, promote economic development, and improve the well-being of individuals and families affected by road accidents.

**REFERENCES**

[**https://www.kaggle.com/code/pratimtalukdar/road-accidents-in-indian-states-2001-2014**](https://www.kaggle.com/code/pratimtalukdar/road-accidents-in-indian-states-2001-2014) **[1]**

[**https://www.statisticshowto.com/probability-and-statistics/hypothesis-testing/anova/**](https://www.statisticshowto.com/probability-and-statistics/hypothesis-testing/anova/)**– used as a reference as to how to conduct and use ANOVA.**



A screenshot of a computer

Description automatically generated with low confidence