Traffic Accident Analysis in Addis Ababa - Case Study

Khalifa Alghafri

2024-11-07

1. Background and problem definition (5pts)

Traffic accidents are among the leading causes of injury and death globally. Most importantly, the study of causes and antecedents of these mishaps will go a long way in enhancing road safety, and hence decreasing the fatality rate. In this research, the Road Traffic Accident Dataset of Addis Ababa City is employed to analyze the correlation between antecedent conditions such as weather conditions, driver's experience, and the Road Alignment and Traffic congestion, with the extent of accident severity.

Specific inputs from the data set include details of over 12000 accidents that occurred between 2017 and 2020, and the circumstances when they occurred including time, location, type of vehicle and conditions at the time of the accident. Particularly, this work will be aimed at identifying specific characteristics of severe accident types because they have the highest impact on public health.

The issue under focus in this study is the analysis of severe road traffic accidents to determine their main antecedents. Thus, working with this dataset, the aim is to identify potential patterns as to how different variables such as weather condition, drivers' experience, road gradients and much more are linked to the severity of the accidents. The findings generated from the analysis may serve to enhance the creation of safer prevention strategies that may assist policymakers, traffic authorities, and the public to minimize the number of fatal accidents on the road.

You can download the dataset from https://www.kaggle.com/datasets/saurabhshahane/road-traffic-accidents?select=RTA+Dataset.csv

Objectives

This project seeks to answer the following research questions:

1. The main causes of RTAs in Addis Ababa The frequency of RTAs in Addis Ababa
2. This study will seek to explain how the Accidents are distributed to the male and female Drivers.
3. Towards the probability of the Accidents, how are they divided on different age band of drivers?
4. What is the relationship between the driver's experience and number of accidents or how does experience affect the number of accidents?
5. How are road accidents distributed by the type of vehicles?
6. Do changes in road surface affect the level of the accident?
7. This raises the next questions: Are there any trends in the accident severity across the types of road alignment such as curves or flat surfaces etc?
8. The Relation Between Cause of Accidents and the Level of Injury.

9. Sex of Driver linked to collision type.
10. On which day of the week are there most number of accidents reported in the various regions.The number of accidents also affects by light conditions.
11. Light conditions contribution to the number of accidents.

2. Data wrangling, munging and cleaning (20pts)

```r
# Loading libraries
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----------------------- tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.2     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
library(lubridate)
library(ggplot2)
library(dplyr)
```

```r
# Loading dataset
accidents_data <- read.csv("RTA Dataset.csv")

# Inspect the dataset
str(accidents_data)
```

```
## 'data.frame':    12316 obs. of  32 variables:
##  $ Time                        : chr  "17:02:00" "17:02:00" "17:02:00" "1:06:00" ...
##  $ Day_of_week                 : chr  "Monday" "Monday" "Monday" "Sunday" ...
##  $ Age_band_of_driver          : chr  "18-30" "31-50" "18-30" "18-30" ...
##  $ Sex_of_driver               : chr  "Male" "Male" "Male" "Male" ...
##  $ Educational_level           : chr  "Above high school" "Junior high school" "Junior high school" "J
##  $ Vehicle_driver_relation     : chr  "Employee" "Employee" "Employee" "Employee" ...
##  $ Driving_experience          : chr  "1-2yr" "Above 10yr" "1-2yr" "5-10yr" ...
##  $ Type_of_vehicle             : chr  "Automobile" "Public (> 45 seats)" "Lorry (41?100Q)" "Public (>
##  $ Owner_of_vehicle            : chr  "Owner" "Owner" "Owner" "Governmental" ...
##  $ Service_year_of_vehicle     : chr  "Above 10yr" "5-10yrs" "" "" ...
##  $ Defect_of_vehicle           : chr  "No defect" "No defect" "No defect" "No defect" ...
##  $ Area_accident_occured       : chr  "Residential areas" "Office areas" "  Recreational areas" "Offic
##  $ Lanes_or_Medians            : chr  "" "Undivided Two way" "other" "other" ...
##  $ Road_allignment             : chr  "Tangent road with flat terrain" "Tangent road with flat terrain
##  $ Types_of_Junction           : chr  "No junction" "No junction" "No junction" "Y Shape" ...
##  $ Road_surface_type           : chr  "Asphalt roads" "Asphalt roads" "Asphalt roads" "Earth roads" .
##  $ Road_surface_conditions     : chr  "Dry" "Dry" "Dry" "Dry" ...
```

```
##  $ Light_conditions           : chr  "Daylight" "Daylight" "Daylight" "Darkness - lights lit" ...
##  $ Weather_conditions         : chr  "Normal" "Normal" "Normal" "Normal" ...
##  $ Type_of_collision          : chr  "Collision with roadside-parked vehicles" "Vehicle with vehicle
##  $ Number_of_vehicles_involved: int  2 2 2 2 2 1 1 2 2 2 ...
##  $ Number_of_casualties       : int  2 2 2 2 2 1 1 1 1 1 ...
##  $ Vehicle_movement           : chr  "Going straight" "Going straight" "Going straight" "Going strai
##  $ Casualty_class             : chr  "na" "na" "Driver or rider" "Pedestrian" ...
##  $ Sex_of_casualty            : chr  "na" "na" "Male" "Female" ...
##  $ Age_band_of_casualty       : chr  "na" "na" "31-50" "18-30" ...
##  $ Casualty_severity          : chr  "na" "na" "3" "3" ...
##  $ Work_of_casuality          : chr  "" "" "Driver" "Driver" ...
##  $ Fitness_of_casuality       : chr  "" "" "" "Normal" ...
##  $ Pedestrian_movement        : chr  "Not a Pedestrian" "Not a Pedestrian" "Not a Pedestrian" "Not a
##  $ Cause_of_accident          : chr  "Moving Backward" "Overtaking" "Changing lane to the left" "Chan
##  $ Accident_severity          : chr  "Slight Injury" "Slight Injury" "Serious Injury" "Slight Injury
```

Featuer selection

The columns are being dropped because they either provide irrelevant information for analyzing the main factors of road traffic accidents or contain data that is not central to understanding accident causes, severity, or patterns. This simplification helps focus on the most impactful variables for the analysis.

```r
# Droping  the  columns below
columns_to_drop <- c("Time", "Educational_level", "Owner_of_vehicle", "Lanes_or_Medians",
                  "Types_of_Junction", "Casualty_class", "Sex_of_casualty",
                  "Age_band_of_casualty", "Casualty_severity", "Work_of_casuality",
                  "Fitness_of_casuality", "Pedestrian_movement")

# Remove the columns from the dataset
new_accidents_data <- accidents_data[, !(names(accidents_data) %in% columns_to_drop)]

# Check the remaining columns
remaining_columns <- colnames(new_accidents_data)
print("Remaining Columns After Dropping:")
```

```
## [1] "Remaining Columns After Dropping:"
```

```r
print(remaining_columns)
```

```
##  [1] "Day_of_week"                 "Age_band_of_driver"
##  [3] "Sex_of_driver"               "Vehicle_driver_relation"
##  [5] "Driving_experience"          "Type_of_vehicle"
##  [7] "Service_year_of_vehicle"     "Defect_of_vehicle"
##  [9] "Area_accident_occured"       "Road_allignment"
## [11] "Road_surface_type"           "Road_surface_conditions"
## [13] "Light_conditions"            "Weather_conditions"
## [15] "Type_of_collision"           "Number_of_vehicles_involved"
## [17] "Number_of_casualties"        "Vehicle_movement"
## [19] "Cause_of_accident"           "Accident_severity"
```

Dropping the missing values in data set

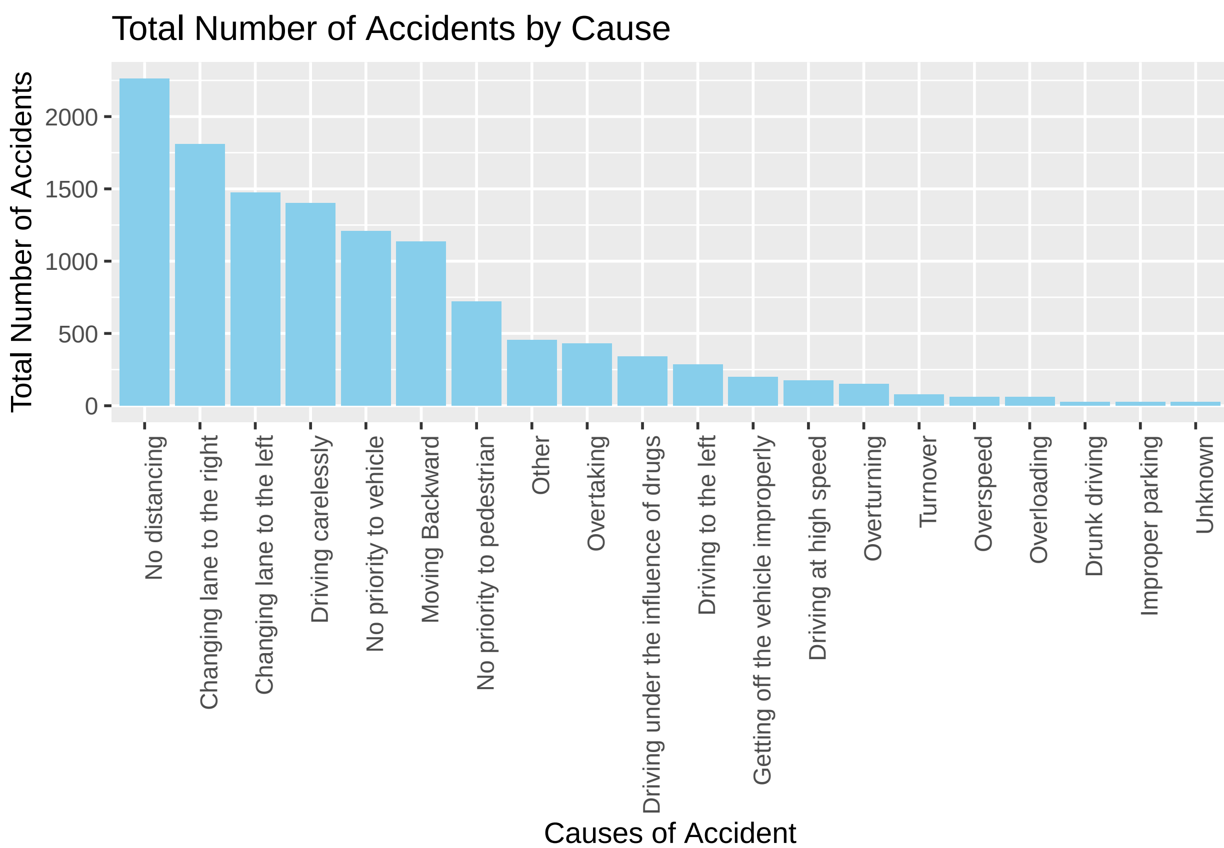
```
# Drop rows with missing values
accidents_clean_data <- na.omit(new_accidents_data)
```

3. Exploratory Data Analysis (30pts) And

4. Data Visualization (15pts)

The main causes of RTAs in Addis Ababa The frequency of RTAs in Addis Ababa

```
# Count  total number of accidents for each cause
cause_accidents_count <- table(accidents_clean_data$Cause_of_accident)

# Convert the table into a data frame for easier manipulation
cause_accidents_df <- as.data.frame(cause_accidents_count)
colnames(cause_accidents_df) <- c("Cause_of_accident", "Total_Accidents")



# Visualize the count of accidents by cause
library(ggplot2)
ggplot(cause_accidents_df, aes(x = reorder(Cause_of_accident, -Total_Accidents), y = Total_Accidents)) +
  geom_bar(stat = "identity", fill = "skyblue") +
  labs(title = "Total Number of Accidents by Cause",
       x = "Causes of Accident",
       y = "Total Number of Accidents") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) # Rotate x labels for better readability
```




```r
# Display the result
print(cause_accidents_df)
```

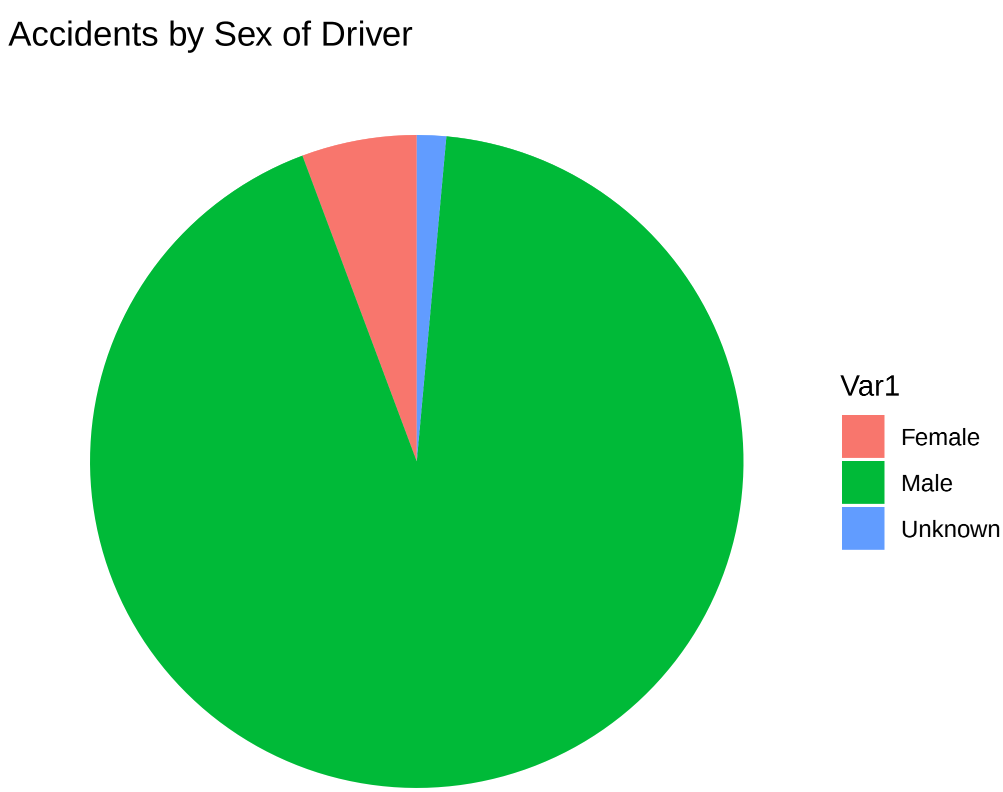
```
##                       Cause_of_accident Total_Accidents
## 1               Changing lane to the left            1473
## 2              Changing lane to the right            1808
## 3                   Driving at high speed             174
## 4                      Driving carelessly            1402
## 5                     Driving to the left             284
## 6  Driving under the influence of drugs             340
## 7                           Drunk driving              27
## 8      Getting off the vehicle improperly             197
## 9                        Improper parking              25
## 10                        Moving Backward            1137
## 11                          No distancing            2263
## 12              No priority to pedestrian             721
## 13                 No priority to vehicle            1207
## 14                                  Other             456
## 15                            Overloading              59
## 16                             Overspeed              61
## 17                             Overtaking             430
## 18                            Overturning             149
## 19                               Turnover              78
## 20                                Unknown              25
```

Findings

The most frequent causes of accidents in Addis Ababa are linked to behaviors like changing lanes improperly, not maintaining adequate distance, and careless driving. Among these, No Distancing is the most prevalent, occurring in over 2,200 incidents. Other common causes include lane changes, careless driving, and lack of priority to vehicles and pedestrians.

How the Accidents are distributed among the male and female Drivers?

```
sex_counts <- table(accidents_clean_data$Sex_of_driver)
ggplot(as.data.frame(sex_counts), aes(x = "", y = Freq, fill = Var1)) +
  geom_bar(stat = "identity", width = 1) +
  coord_polar("y") +
  labs(title = "Accidents by Sex of Driver") +
  theme_void()
```




```
print(sex_counts)
```

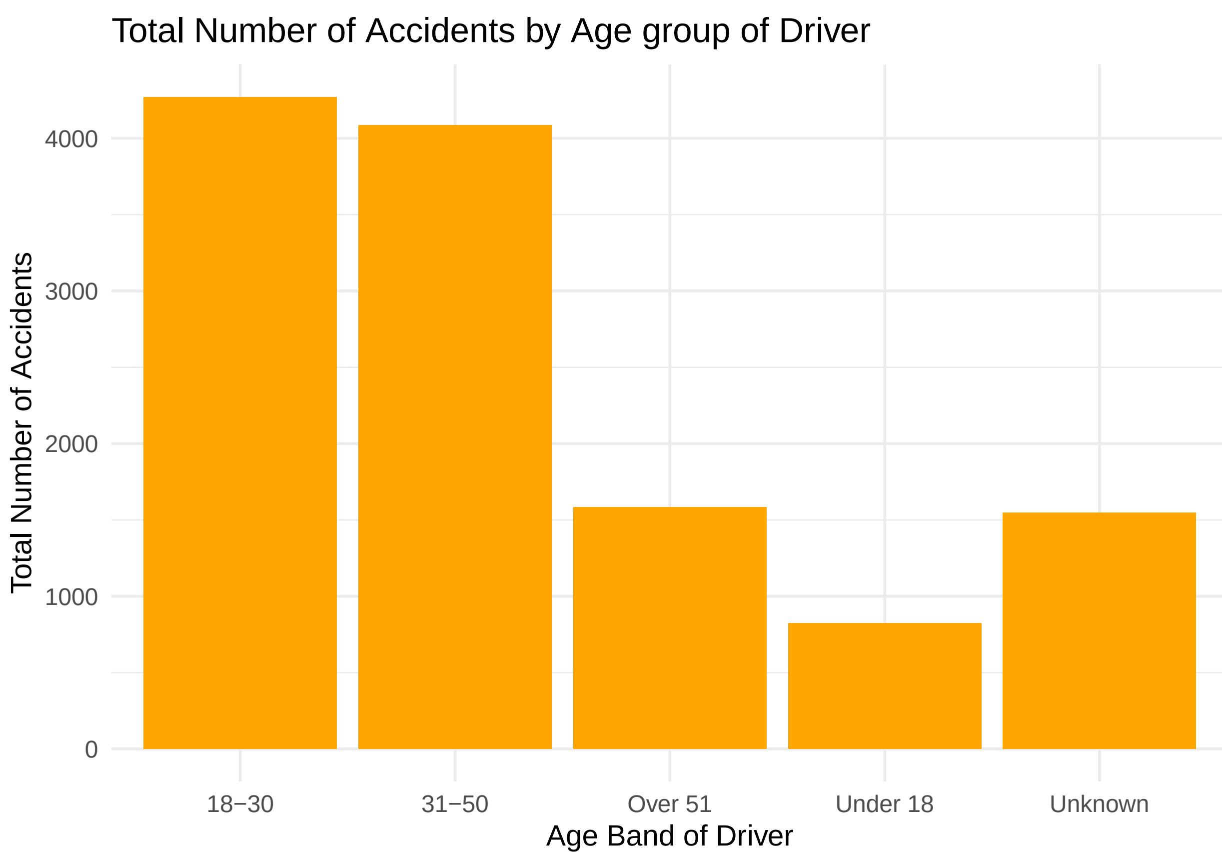
```
##
##  Female    Male Unknown
##     701   11437     178
```

Findings

The majority of road traffic accidents in Addis Ababa involve male drivers, accounting for a significantly higher number of accidents compared to female drivers. The unknown category also represents a small portion of the total accidents

How the Accidents are distributed among the Age Band of Drivers?

```
ggplot(accidents_clean_data, aes(x = Age_band_of_driver)) +
  geom_bar(fill = "orange") +
  labs(title = "Total Number of Accidents by Age group of Driver", x = "Age Band of Driver", y = "Total
  theme_minimal()
```




```
# Print total accidents by age band
table(accidents_clean_data$Age_band_of_driver)
```

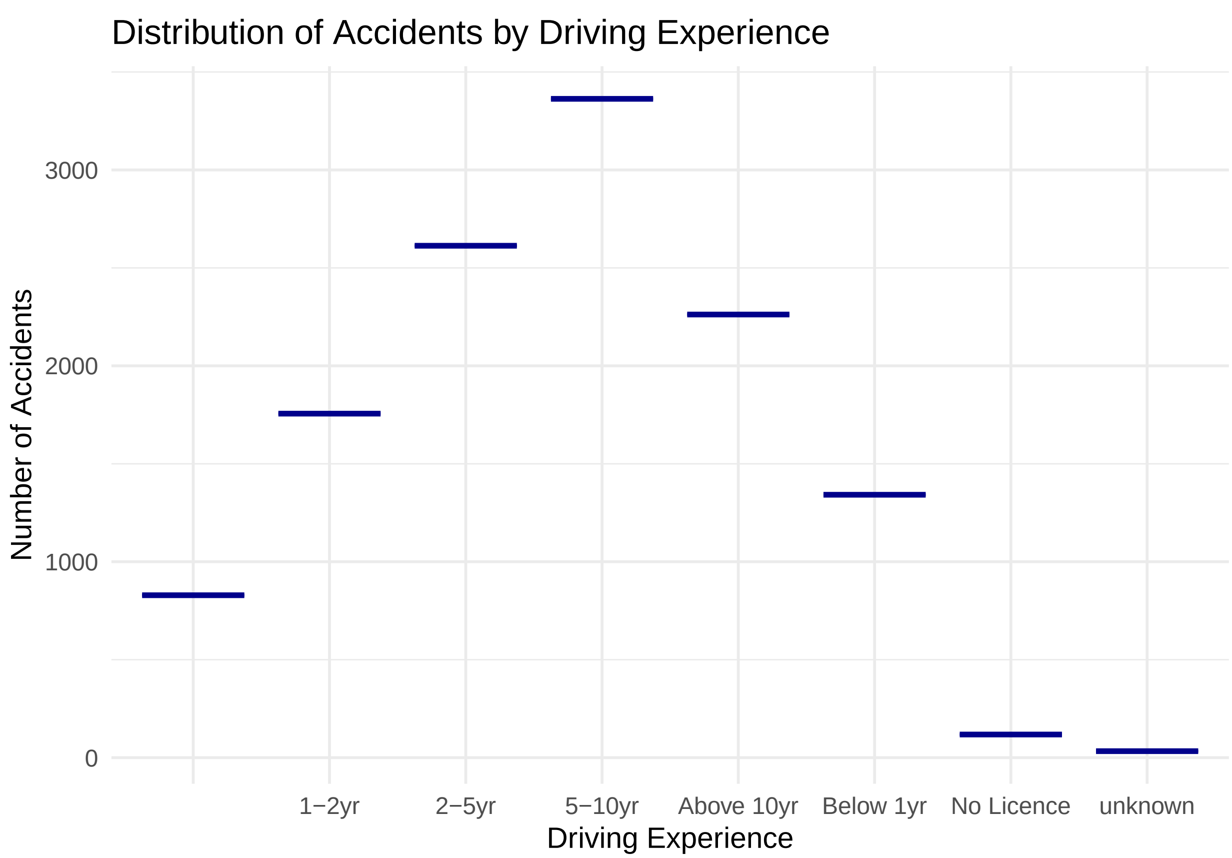
```
##
##    18-30    31-50  Over 51 Under 18  Unknown
##     4271     4087     1585      825     1548
```

Findings

The data shows a higher number of accidents among younger drivers (18-30 years) and those in the 31-50 years group. This suggests that driving experience (or lack of it) is a significant factor in accident occurrence, with younger drivers facing higher risks

What is the realtion between the experince of driver and the number of accidents?

```
# Create a new data frame to count accidents per Driving Experience
accidents_by_driver_experience <- accidents_clean_data %>%
  group_by(Driving_experience) %>%
  summarize(Number_of_accidents = n())

# Now we create a boxplot (if relevant), here we use geom_boxplot to show distribution
ggplot(accidents_by_driver_experience, aes(x = Driving_experience, y = Number_of_accidents)) +
  geom_boxplot(fill = "lightblue", color = "darkblue") +
  labs(title = "Distribution of Accidents by Driving Experience",
       x = "Driving Experience",
       y = "Number of Accidents") +
  theme_minimal()
```




```
# Print total accidents by driving experience
table(accidents_clean_data$Driving_experience)
```

```
##
##                   1-2yr      2-5yr     5-10yr Above 10yr  Below 1yr No Licence
```

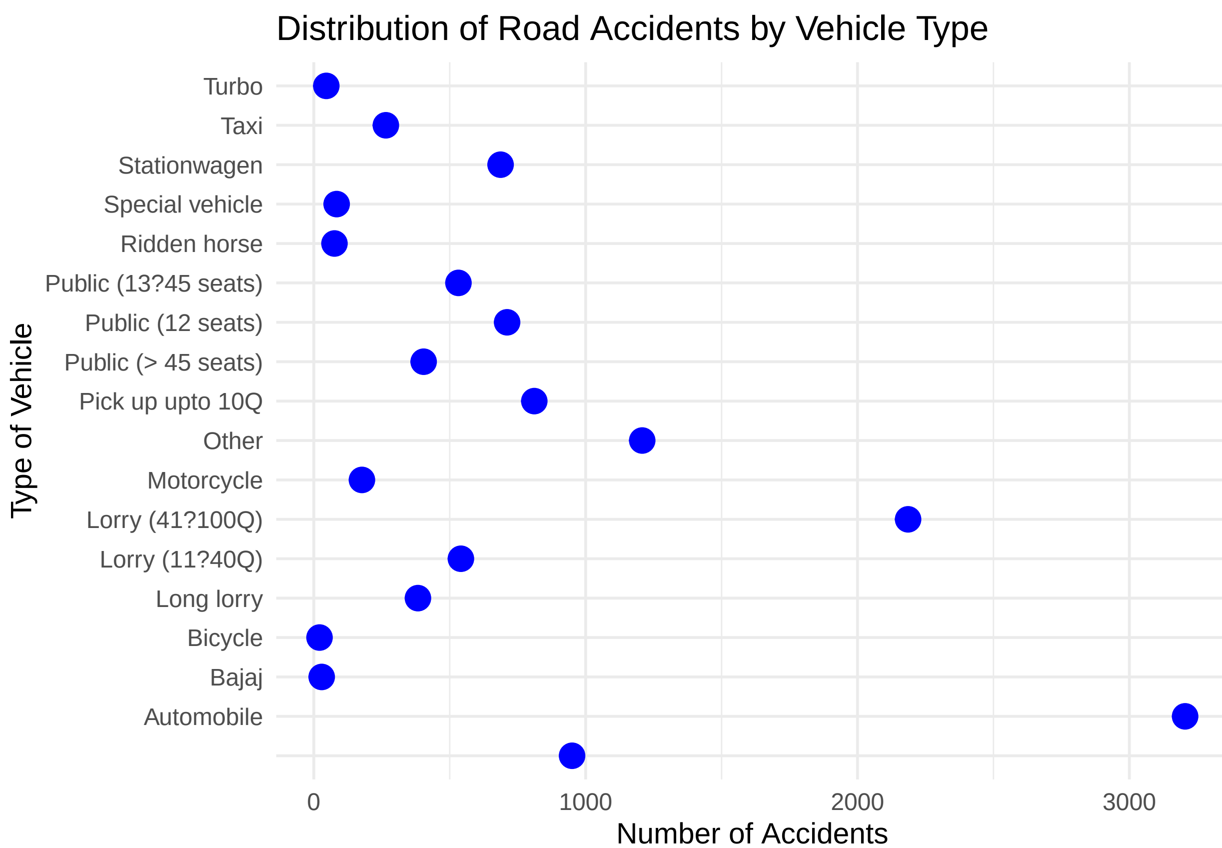
```
##        829        1756        2613        3363        2262        1342         118
##    unknown
##         33
```

Findings

The data shows that both inexperienced drivers (below 1 year) and very experienced drivers (above 10 years) are involved in the highest number of accidents. This suggests that driving experience is a key factor in accident occurrence, with both ends of the spectrum (new and highly experienced drivers) potentially facing higher risks. The middle range (2-5 years and 5-10 years) shows a moderate number of accidents, suggesting these drivers are somewhat more experienced and likely make fewer errors on the road.

What is the distribution of road accidents by the type of vehicle involved?

```r
# Group by 'Type_of_vehicle' and count the number of accidents
vehicle_accidents <- accidents_clean_data %>%
  group_by(Type_of_vehicle) %>%
  summarise(accidents_count = n())



ggplot(vehicle_accidents, aes(x = accidents_count, y = Type_of_vehicle)) +
  geom_point(size = 4, color = "blue") +
  theme_minimal() +
  labs(title = "Distribution of Road Accidents by Vehicle Type",
       x = "Number of Accidents", y = "Type of Vehicle")
```




```r
# View the result
print(vehicle_accidents)
```

```
## # A tibble: 18 x 2
##    Type_of_vehicle        accidents_count
##    <chr>                            <int>
##  1 ""                                 950
##  2 "Automobile"                      3205
##  3 "Bajaj"                             29
##  4 "Bicycle"                           21
##  5 "Long lorry"                       383
##  6 "Lorry (11?40Q)"                   541
##  7 "Lorry (41?100Q)"                 2186
##  8 "Motorcycle"                       177
##  9 "Other"                           1208
## 10 "Pick up upto 10Q"                 811
## 11 "Public (12 seats)"                711
## 12 "Public (13?45 seats)"             532
## 13 "Public (> 45 seats)"              404
## 14 "Ridden horse"                      76
## 15 "Special vehicle"                   84
## 16 "Stationwagen"                     687
## 17 "Taxi"                             265
## 18 "Turbo"                             46
```

Findings

Automobiles and Lorries (especially larger ones) dominate the number of accidents in Addis Ababa. Public transport vehicles also contribute notably to accident statistics. Motorcycles and smaller vehicles like Bajajs have a smaller presence but still contribute to the overall accident count.

Correlation between road surface conditions and accident severity.

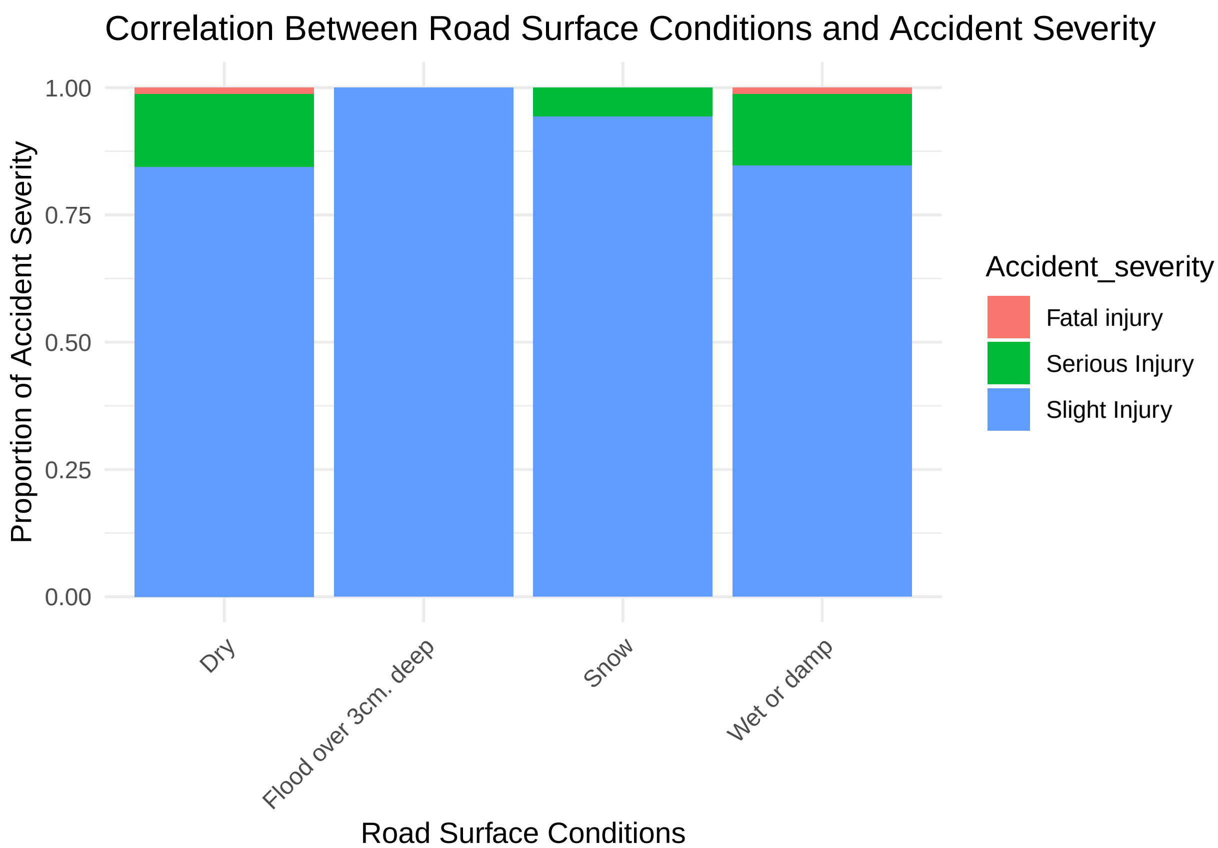
```r
# Create a contingency table for Road Surface Conditions vs Accident Severity
contingency_table <- table(accidents_clean_data$Road_surface_conditions, accidents_clean_data$Accident_s

# show the contingency table
print(contingency_table)
```

```
##
##                          Fatal injury Serious Injury Slight Injury
##   Dry                             121           1332          7887
##   Flood over 3cm. deep              0              0             2
##   Snow                              0              4            66
##   Wet or damp                      37            407          2460
```

```r
# Perform a Chi-squared test
chi_square_test <- chisq.test(contingency_table)
```

```
## Warning in chisq.test(contingency_table): Chi-squared approximation may be
## incorrect
```

```r
# Visualization - Bar plot
ggplot(accidents_clean_data, aes(x = Road_surface_conditions, fill = Accident_severity)) +
  geom_bar(position = "fill") +
  theme_minimal() +
  labs(title = "Correlation Between Road Surface Conditions and Accident Severity",
       x = "Road Surface Conditions", y = "Proportion of Accident Severity") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```




```
# show the Chi-squared test result
print(chi_square_test)
```

```
##
##  Pearson's Chi-squared test
##
## data:  contingency_table
## X-squared = 5.7209, df = 6, p-value = 0.4552
```

Findings

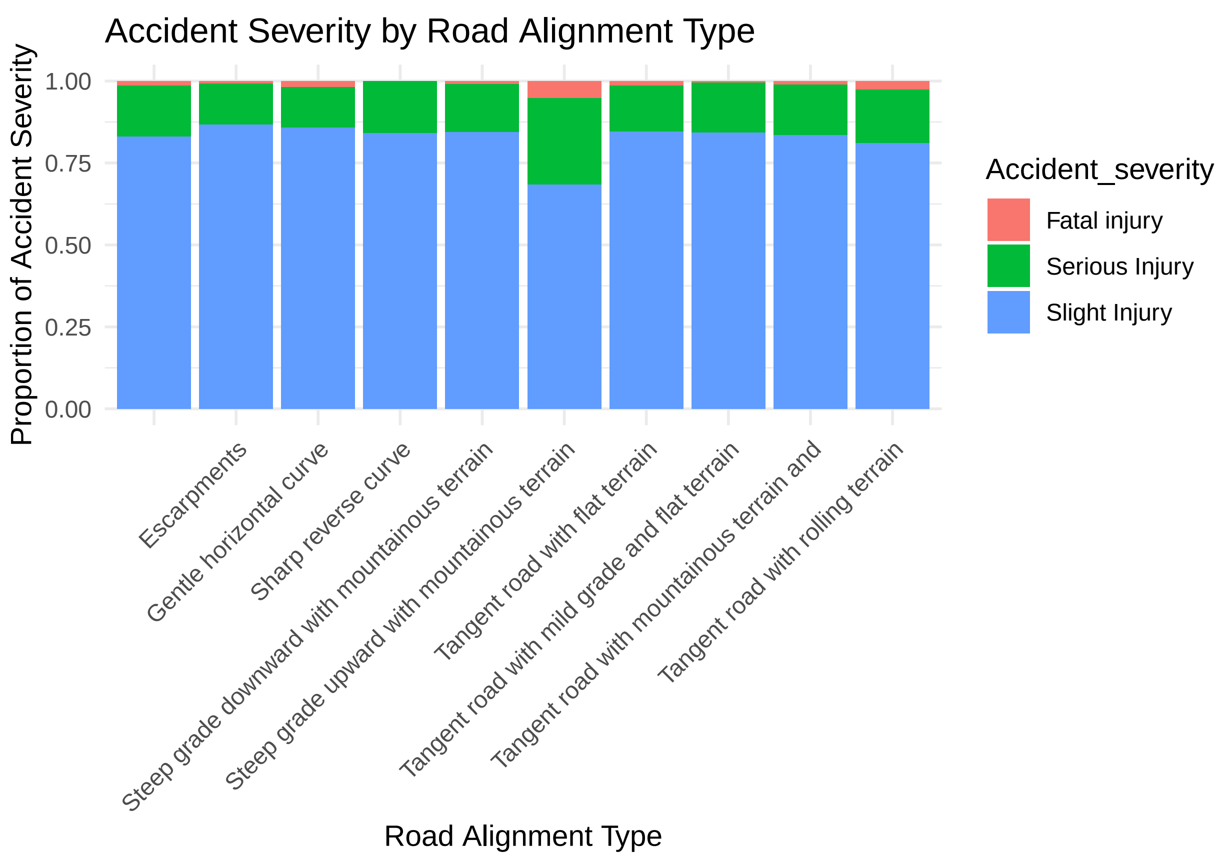
The Graph above shows that accidents on Roads that are Dry or Wer/damp casues serious injury as comapred to those of others.

Trends in accident severity by road alignment type (e.g., curved, flat terrain).

```
# make a contingency table for Road Alignment vs Accident Severity
contingency_table_alignment <- table(accidents_clean_data$Road_allignment, accidents_clean_data$Acciden



# do a Chi-squared test
chi_square_test_alignment <- chisq.test(contingency_table_alignment)
```

```
## Warning in chisq.test(contingency_table_alignment): Chi-squared approximation
## may be incorrect
```

```r
# Visualization - Bar plot
ggplot(accidents_clean_data, aes(x = Road_allignment, fill = Accident_severity)) +
  geom_bar(position = "fill") +
  theme_minimal() +
  labs(title = "Accident Severity by Road Alignment Type",
       x = "Road Alignment Type", y = "Proportion of Accident Severity") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```




```r
# View the contingency table
print(contingency_table_alignment)
```

```
##
##                                                     Fatal injury Serious Injury
##                                                                2             22
##    Escarpments                                                 1             14
##    Gentle horizontal curve                                     3             20
##    Sharp reverse curve                                         0              9
##    Steep grade downward with mountainous terrain               4             63
##    Steep grade upward with mountainous terrain                 1              5
##    Tangent road with flat terrain                            140           1466
##    Tangent road with mild grade and flat terrain               2             77
##    Tangent road with mountainous terrain and                   4             61
```

```
##   Tangent road with rolling terrain                        1              6
##
##                                                      Slight Injury
##                                                                118
##   Escarpments                                                    98
##   Gentle horizontal curve                                       140
##   Sharp reverse curve                                            48
##   Steep grade downward with mountainous terrain                 362
##   Steep grade upward with mountainous terrain                    13
##   Tangent road with flat terrain                               8853
##   Tangent road with mild grade and flat terrain                 422
##   Tangent road with mountainous terrain and                     331
##   Tangent road with rolling terrain                              30
```

```r
# Output the Chi-squared test result
print(chi_square_test_alignment)
```

```
##
##  Pearson's Chi-squared test
##
## data:  contingency_table_alignment
## X-squared = 13.214, df = 18, p-value = 0.7787
```

Findings

As shown in the graph "Steep grade upward with mountainous terrain" have most of serious injuries in accidents. followed by "Tangent road wih rolling terrain" and "sharp reverse curve".

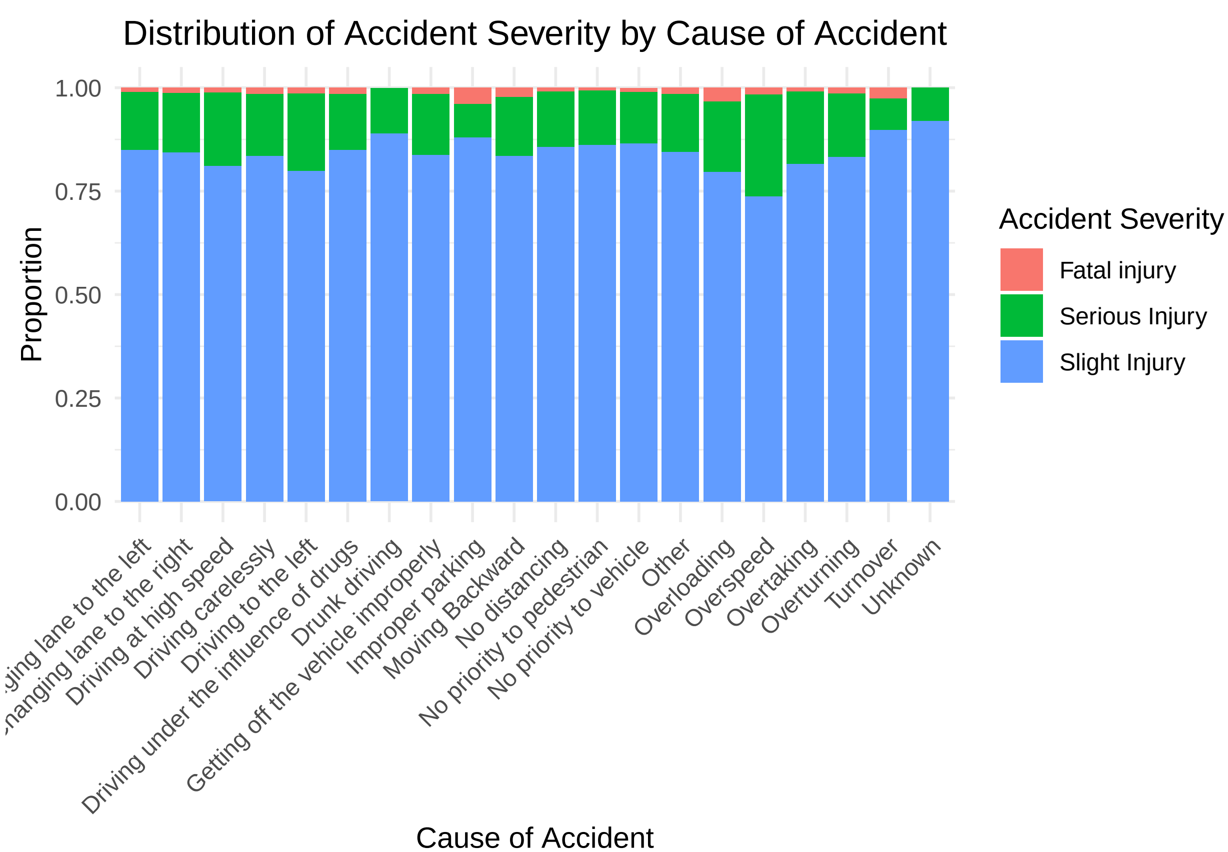
Relation between Cause of Accidents and Injury Severity

```r
# make a contingency table for Cause_of_accident and Accident_severity
contingency_table_cause_severity <- table(accidents_clean_data$Cause_of_accident, accidents_clean_data$



# Perform a Chi-Square test to test for independence between Cause_of_accident and Accident_severity
chisq_test_cause_severity <- chisq.test(contingency_table_cause_severity)
```

```
## Warning in chisq.test(contingency_table_cause_severity): Chi-squared
## approximation may be incorrect
```

```r
# Load necessary library for visualization

# Create a stacked bar chart to show Accident Severity distribution for each Cause of Accident
ggplot(accidents_clean_data, aes(x = Cause_of_accident, fill = Accident_severity)) +
  geom_bar(position = "fill") +  # "fill" makes it a stacked proportion chart
  labs(
    title = "Distribution of Accident Severity by Cause of Accident",
    x = "Cause of Accident",
    y = "Proportion",
```

```r
    fill = "Accident Severity"
  ) +
  theme_minimal() +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1),  # Rotate x-axis labels for readability
    plot.title = element_text(hjust = 0.5)  # Center the title
  )
```




```r
# Display the contingency table to view accident severity counts for each cause
contingency_table_cause_severity
```

```
##
##                                          Fatal injury Serious Injury
##    Changing lane to the left                       16            206
##    Changing lane to the right                      23            260
##    Driving at high speed                            2             31
##    Driving carelessly                              22            209
##    Driving to the left                              4             53
##    Driving under the influence of drugs             5             46
##    Drunk driving                                    0              3
##    Getting off the vehicle improperly               3             29
##    Improper parking                                 1              2
##    Moving Backward                                 26            162
##    No distancing                                   20            303
##    No priority to pedestrian                        5             95
```

```
##   No priority to vehicle                             13            149
##   Other                                               7             64
##   Overloading                                         2             10
##   Overspeed                                           1             15
##   Overtaking                                          4             75
##   Overturning                                         2             23
##   Turnover                                            2              6
##   Unknown                                             0              2
##
##                                           Slight Injury
##   Changing lane to the left                        1251
##   Changing lane to the right                       1525
##   Driving at high speed                             141
##   Driving carelessly                               1171
##   Driving to the left                               227
##   Driving under the influence of drugs              289
##   Drunk driving                                      24
##   Getting off the vehicle improperly                165
##   Improper parking                                   22
##   Moving Backward                                   949
##   No distancing                                    1940
##   No priority to pedestrian                         621
##   No priority to vehicle                           1045
##   Other                                             385
##   Overloading                                        47
##   Overspeed                                          45
##   Overtaking                                        351
##   Overturning                                       124
##   Turnover                                           70
##   Unknown                                            23
```

```
# Display the results of the Chi-Square test
chisq_test_cause_severity
```

```
##
##  Pearson's Chi-squared test
##
## data:  contingency_table_cause_severity
## X-squared = 49.08, df = 38, p-value = 0.1075
```

Findings

Based on the provided data, here's the analysis for the highest and lowest accident types leading to **serious injuries**, **fatal injuries**, and **slight injuries**:

Serious Injuries:

- **Highest**: Number one reason is no distancing which resulted in 303 serious injuries. This also increases serious injuries with a total of 260 to the *"Changing lane to the right"* signal and 206 to the signal *"Changing lane to the left"*.
- **Lowest**: Thanks to 'drunk driving', which claimed no more than 3 severe cases. This was followed by physical violence with, *Improper parking* and *Unknown* accounting for only 2 and 2 on the severe injuries respectively.

Fatal Injuries:

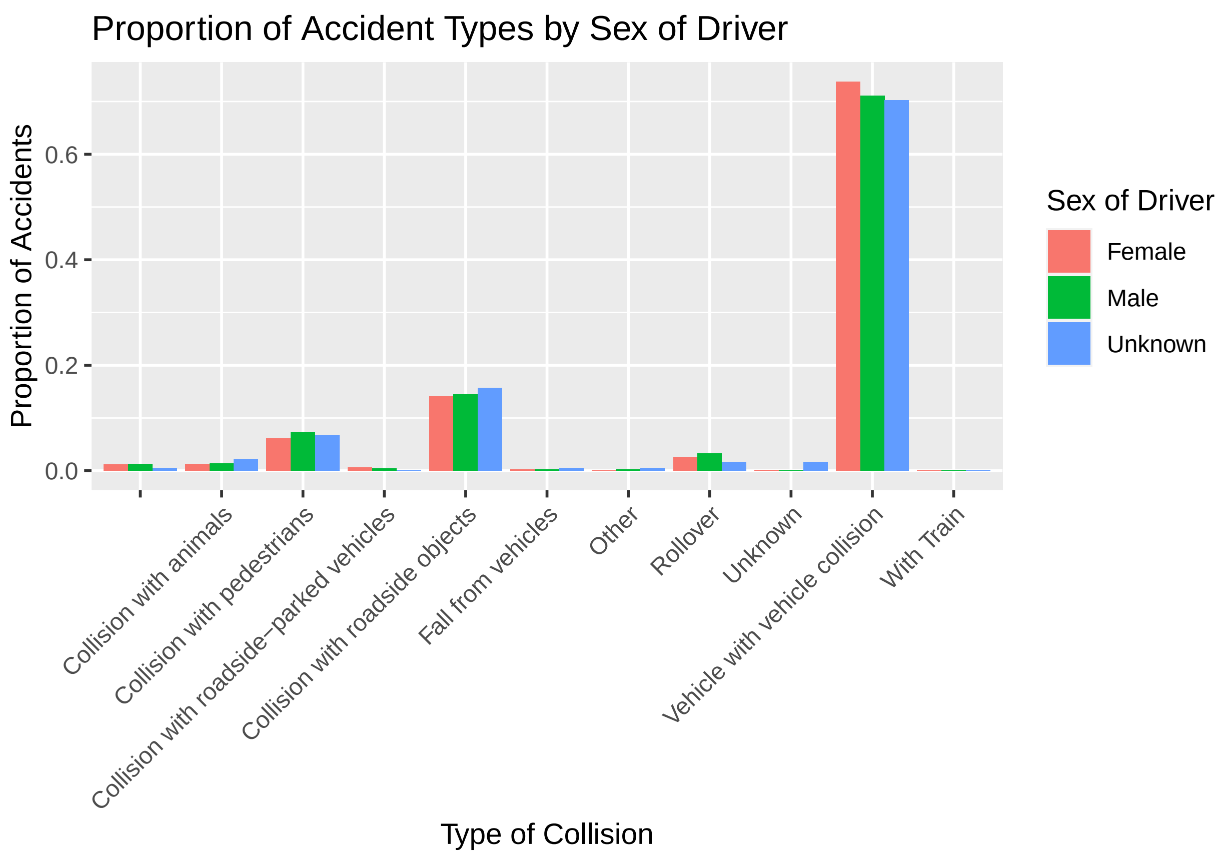
- **Highest**: According to the source, the highest number of fatal injuries, 26, is recorded under the scenario, *"Moving Backward". Other signal words, such as "Changing lane to the right" as well as "Changing lane to the left" are also quite dangerous, where 23 and 16 of the fatal car injuries respectively where reported.
- **Lowest**: The offence of *"drunk driving"* recorded no fatalities (0 fatalities). *Misplacement* and *Unknown* also recorded the lowest average number of fatal injuries; three and none respectively.

Slight Injuries:

- **Highest**: This involves 1,940 slight injuries which recorded for the cases that involved *no distancing. Turning right and turning left* are also responsible for minor crashes and 1,525 and 1,251 slight injuries were reported respectively.
- **Lowest**: Convictions on the offense of ''drunk driving" only recorded 24 minor accidents. also recorded 22 slight injuries while *"Unknown"* recorded 23 slight injuries.

Relationship between sex and type o faccident

```
# Create a contingency table for Sex_of_driver and Type_of_collision
contingency_table_sex_collision <- table(accidents_clean_data$Sex_of_driver, accidents_clean_data$Type_o


# Calculate the proportion of each type of collision by sex
prop_sex_collision <- prop.table(contingency_table_sex_collision, 1)



# Plot the data using a bar plot to visualize the proportions
library(ggplot2)
contingency_data <- as.data.frame(prop_sex_collision)
colnames(contingency_data) <- c("Sex_of_driver", "Type_of_collision", "Proportion")

ggplot(contingency_data, aes(x = Type_of_collision, y = Proportion, fill = Sex_of_driver)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(title = "Proportion of Accident Types by Sex of Driver",
       x = "Type of Collision",
       y = "Proportion of Accidents",
       fill = "Sex of Driver") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```




```r
print(contingency_table_sex_collision)
```

```
##
##              Collision with animals Collision with pedestrians
##   Female    8                     9                         43
##   Male    146                   158                        841
##   Unknown   1                     4                         12
##
##          Collision with roadside-parked vehicles
##   Female                                       4
##   Male                                        50
##   Unknown                                      0
##
##          Collision with roadside objects Fall from vehicles Other Rollover
##   Female                              99                  2     0       18
##   Male                              1659                 31    25      376
##   Unknown                             28                  1     1        3
##
##          Unknown Vehicle with vehicle collision With Train
##   Female       1                            517          0
##   Male        10                           8132          9
##   Unknown      3                            125          0
```

```r
# View the proportions
print(prop_sex_collision)
```

```
##
##                        Collision with animals Collision with pedestrians
##   Female  0.0114122682          0.0128388017               0.0613409415
##   Male    0.0127655854          0.0138148116               0.0735332692
##   Unknown 0.0056179775          0.0224719101               0.0674157303
##
##           Collision with roadside-parked vehicles
##   Female                              0.0057061341
##   Male                                0.0043717758
##   Unknown                             0.0000000000
##
##           Collision with roadside objects Fall from vehicles       Other
##   Female                       0.1412268188       0.0028530670 0.0000000000
##   Male                         0.1450555216       0.0027105010 0.0021858879
##   Unknown                      0.1573033708       0.0056179775 0.0056179775
##
##               Rollover      Unknown Vehicle with vehicle collision   With Train
##   Female  0.0256776034 0.0014265335                   0.7375178317 0.0000000000
##   Male    0.0328757541 0.0008743552                   0.7110256186 0.0007869196
##   Unknown 0.0168539326 0.0168539326                   0.7022471910 0.0000000000
```

Findings

The data reveal the numbers of RTAs by the type of collision, by driver's sex, and by the ratio of each type of collision for each sex.x. Here's a breakdown of the key findings:

Relative Frequency (Rate of Occurrence):

The second part of the data provides the **relative frequency** of each type of collision for each sex, expressed as a proportion of all accidents in that category: - **Collision with Animals**: Male appears slightly more frequently than Female, 0.0128 for Male and 0.0114 for Female. - **Collision with Pedestrians**: Males are again seen to have a higher relative frequency (0.0138) than females (0.0128). - **Collision with Roadside-Parked Vehicles**: Males are less frequently related to this disorder than females (0.0044 > 0.0057). - **Collision with Roadside Objects**: Frequencies of male workers are relatively less frequent (number of males per 1000 is 0.1451) than female workers (number of females per 1000 is 0.1412). - **Fall from Vehicles**: It turns out that the parameters are not very distinct from each other in terms of the probabilities for both male and female participants (males: 0.0027, females: 0.0029). - **Other Types of Collisions**: The relative frequencies of most types of accidents such as roll over are however slightly higher in males than in females with 0.0329 and 0.0257 respectively.

For the **"Vehicle with Vehicle Collision"** compares the relative frequency of the males and female and the former has relatively a higher value 0.711 than the females 0.737.

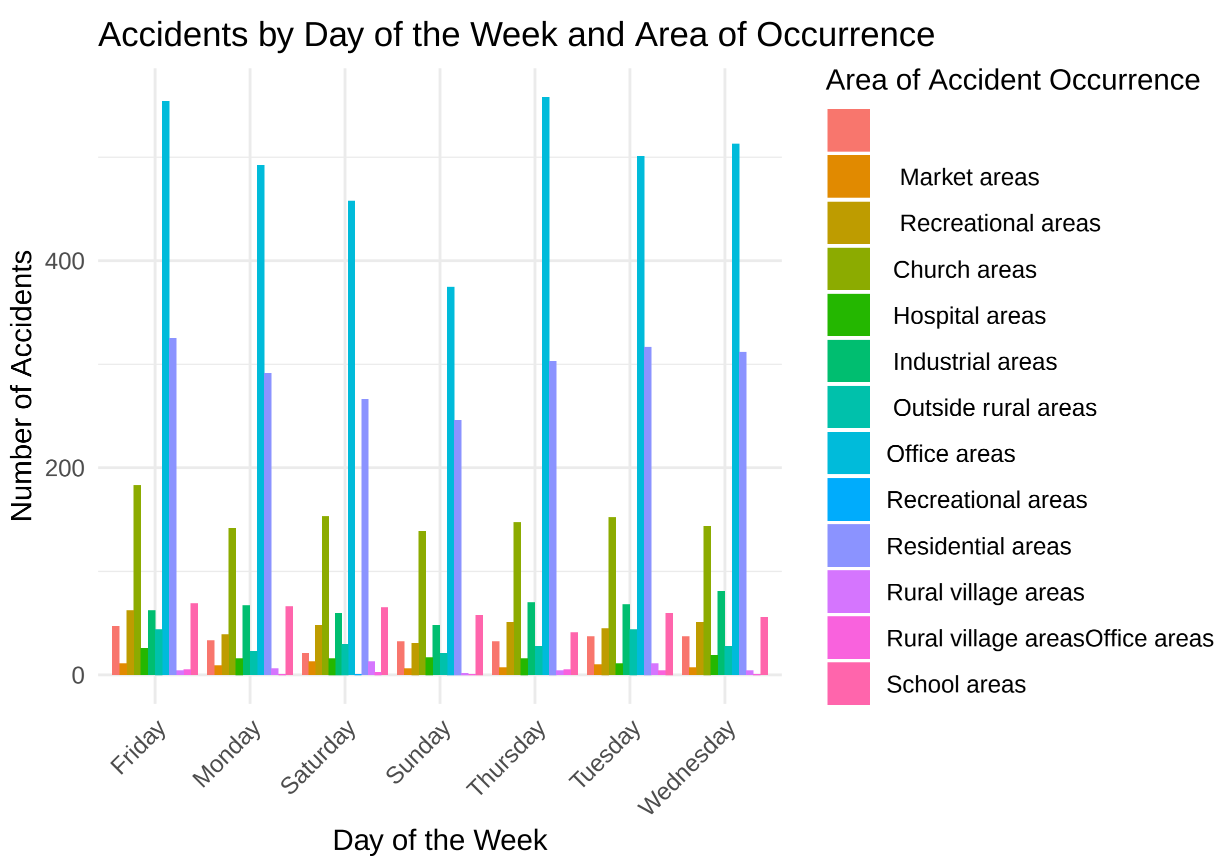
Men are also more engaged in all forms of the crash than women especially the ones that; involve pedestrian, roadside furniture and other cars.
Thus, relative frequency rates are indicating that while males have more overall accidents, **the accident rate** is the same for both sexes for certain kinds of accidents, for example, "Fall from vehicles" or "Collision with animals." **Unknown sex** often goes hand by hand with minimal occurrences of the incidents inclusive of all categories of collisions and articulated that some of the collision types may either lack records or are rarely reported.

Which day of the week has most number of accidents recorded in different areas.

```
# Filter out the "Other" and "Unknown" categories from Area_accident_occured
accidents_clean_filtered <- accidents_clean_data %>%
  filter(!Area_accident_occured %in% c("Other", "Unknown"))

# Group by 'Day_of_week' and 'Area_accident_occured', and count the number of accidents
accidents_by_day_area <- accidents_clean_filtered %>%
  group_by(Day_of_week, Area_accident_occured) %>%
  summarise(accident_count = n()) %>%
  arrange(desc(accident_count))
```

```
## 'summarise()' has grouped output by 'Day_of_week'. You can override using the
## '.groups' argument.
```

```
# Create a bar plot to visualize accidents by Day_of_week and Area_accident_occured
ggplot(accidents_by_day_area, aes(x = Day_of_week, y = accident_count, fill = Area_accident_occured)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(title = "Accidents by Day of the Week and Area of Occurrence ",
       x = "Day of the Week",
       y = "Number of Accidents",
       fill = "Area of Accident Occurrence") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))  # Rotate x-axis labels for better readabili
```

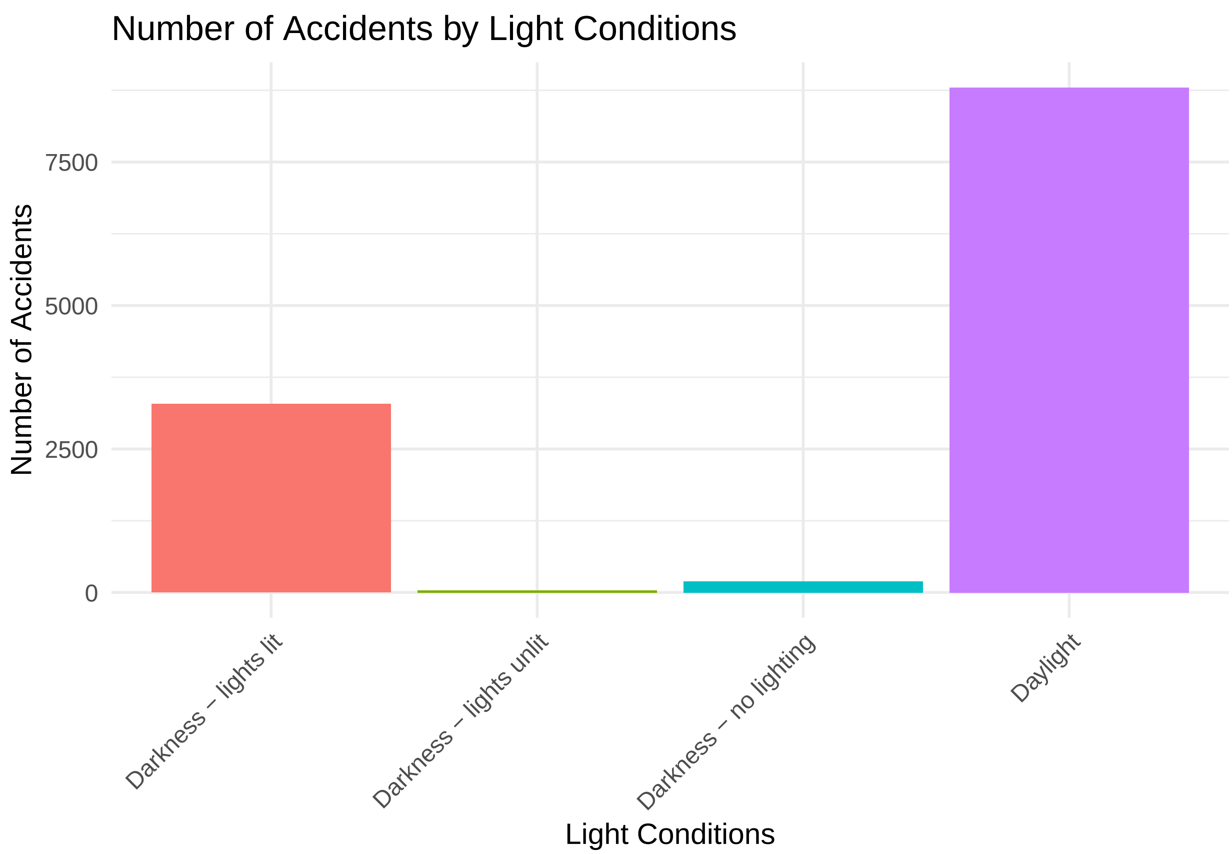


Findings

Observing the above graph we can conclude that higher number of accidents occurs in Office areas followed by residential areas and church areas, While the lowest number of accidents include areas such as Rural village areas,School areas also have low rate of accidents as compared to others,this sugget focusing on busy areas of the city might reduce the rate of accidents occuring.

Light conditions contribution to the number of accidents.

```
# Filter out the "Other" and "Unknown" categories from Light_conditions
accidents_clean_filtered_light <- accidents_clean_data %>%
  filter(!Light_conditions %in% c("Other", "Unknown"))

# Group by 'Light_conditions' and count the number of accidents
accidents_by_light_condition <- accidents_clean_filtered_light %>%
  group_by(Light_conditions) %>%
  summarise(accident_count = n()) %>%
  arrange(desc(accident_count))


# Create a bar plot to visualize accidents by Light_conditions
ggplot(accidents_by_light_condition, aes(x = Light_conditions, y = accident_count, fill = Light_conditi
  geom_bar(stat = "identity", show.legend = FALSE) +
  labs(title = "Number of Accidents by Light Conditions",
       x = "Light Conditions",
       y = "Number of Accidents") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))  # Rotate x-axis labels for better readabili
```




```r
# Display descriptive statistics (summary of data)
summary(accidents_by_light_condition)
```

```
##  Light_conditions   accident_count
##  Length:4           Min.   :  40
##  Class :character   1st Qu.: 154
##  Mode  :character   Median :1739
##                     Mean   :3079
##                     3rd Qu.:4664
##                     Max.   :8798
```

```r
# Display the first few rows of the result for better understanding
head(accidents_by_light_condition)
```

```
## # A tibble: 4 x 2
##   Light_conditions        accident_count
##   <chr>                            <int>
## 1 Daylight                          8798
## 2 Darkness - lights lit             3286
## 3 Darkness - no lighting             192
## 4 Darkness - lights unlit             40
```

Finidngs

The above analysis proved that the **most frequent** accidents happen at **daylight**, this can be explained by the fact that traffic is always more intensified during the daylight hours. Night- time accidents are still frequent especially where streets are well lit . **Darkness no light** as well as **unlit light** has lower occurrence though risky due to poor illumination. The minimum number of accidents is registered in **night with no lighting**; it reveals that, although safely threatening, such accidents are rare, probably because such conditions are met less often or there is little movement on the roads. Comparing the findings of this analysis to the previous research it can be concluded that traffic safety measures especially in the night with good lighting could help minimize situations that lead to accidents.

Final Words

Based on the findings of road traffic accidents in Addis Ababa, it is -possible to identify several important facts that will help to explain the main reasons behind frequent accidents in the city. This paper is able to present that the key causes of the accidents include behaviours that include the improper lane change, failure to maintain distance, and carelessness among others. Moreover, it has been established that male drivers, young drivers, drivers with little experience, and also over experienced drivers are likely to be involved in an accident. Some types of vehicles including automobiles and lorries are also frequently involved in accidents.

Traffic conditions and or The roads conditions and alignment, and road surfaces are among the areas most associated with accident severity where dry or wet surfaces and area that has a mountainous terrain are associated with serious injury. Day time is the peak period in road accidents while saturation of lighting is an important factor in night time accidents. The given data also provides information on the association of the accident types with gender; men are more inclined towards accidents, and especially towards collisions with such objects as pedestrians, roadside objects, and other vehicles.

In sum, these findings give better understanding about the determinants of road safety in Addis Ababa and thus will create the basis for future risk reduction and enhanced road safety intervention in this city.

Amends That Can Be Suggested Upon Analysis

1. **Addressing Careless Driving and Lane Change Accidents**:

   - Concentrating on drivers' awareness, especially novice drivers, regarding proper distance and safe way of lane changing.
     They should develop specific pattern driving tests in an effort to minimize how dangerous the roads are.

2. **Strengthen Enforcement on Speed and Distancing**:
   Intonate more frequent, unpredictable speeding checker and breath analyzer tests in an effort to reduce speeding and drunken driving.
   – Implement even harsher consequences for non adherence to social distancing and drinking and driving cases.

3. **Improving Road Infrastructure**:
   –consolidate road design in frontier areas (example; curves, steep grades) by use of barriers, clear indications and efficient drainage system.
   traffic calming measure should be adopted to prevent rapid speed in areas that experienced high crashes.

4. **Better Street Lighting**:
   – Maintain pedestrian ways well lit particularly during the night in areas of high human trafficked and or accident frequencies.

- Use of motion sensing or solar powered lights in areas where they are almost dark.

5. **Targeted Driver Training**:
   The following courses are as follows; Protective education and training programs for young and inexperienced drivers Protecting education and training programs for older and overconfident drivers.

6. **Pedestrian Safety**:

   - Pedestrian structure should be enhanced for better crossing areas, and pedestrian bridges. Enhance the extent of risk that is associated with and related to safe pedestrian behavior.

7. **Public Transport Safety**:
   – Check up and safety education of the public transport drivers.
   – Supervise observation of safety measures to lessen the public transportation incidence of crashes.

Hence as it expands its interventions in these targeted solutions, Addis Ababa will be able to minimize road accidents and enhance safety in its entire area.