ANT TRUCKS DATA ANALYSIS

GROUP 3

Gande, Avinash

Ganduri, Catherine Christina

Goel, Ashish

Gupta, Ruchika

Jasti, Gnaneshwar

Jawahar Vasagam, Premi

Karur, Vishal

BUSINESS QUESTIONS

Identifying the Optimal Truck Model

Historical MPG values analysis

Risk factor analysis for each truck model

Choosing the best truck model for expansion

Benefits of partnering with the truck company for future purchases and maintenance operations

Identifying High-Risk Drivers

Data analysis approach for driver risk identification

Top 5 high-risk drivers and causal factors

Implications for compliance with traffic rules and regulations

Strategies for hiring and training better employees for safer driving

BUSINESS PLAN

Driver Risk Prediction

- Factors affecting driver risk
- Predictive model development

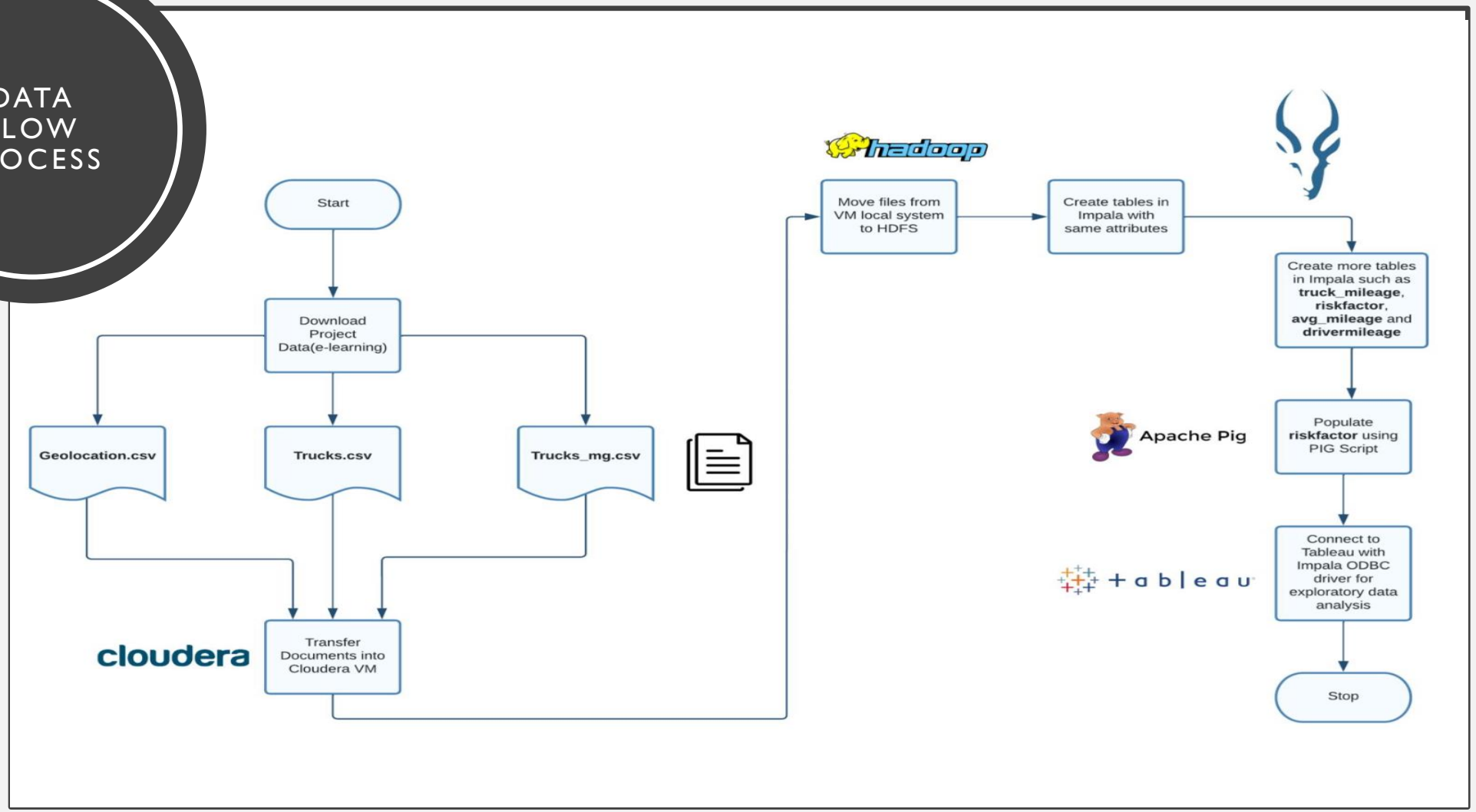
Risky Driver Identification

- Data analysis approach
- Risk factors and characteristics

Accident Reduction for Commercial Trucks

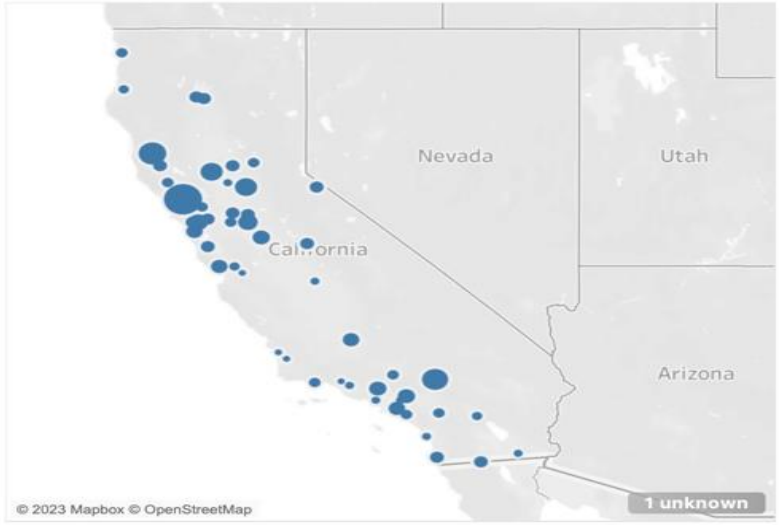
- Driver data analysis for risk identification
- Strategies for reducing accidents
- Impact on road safety

DATA FLOW PROCESS

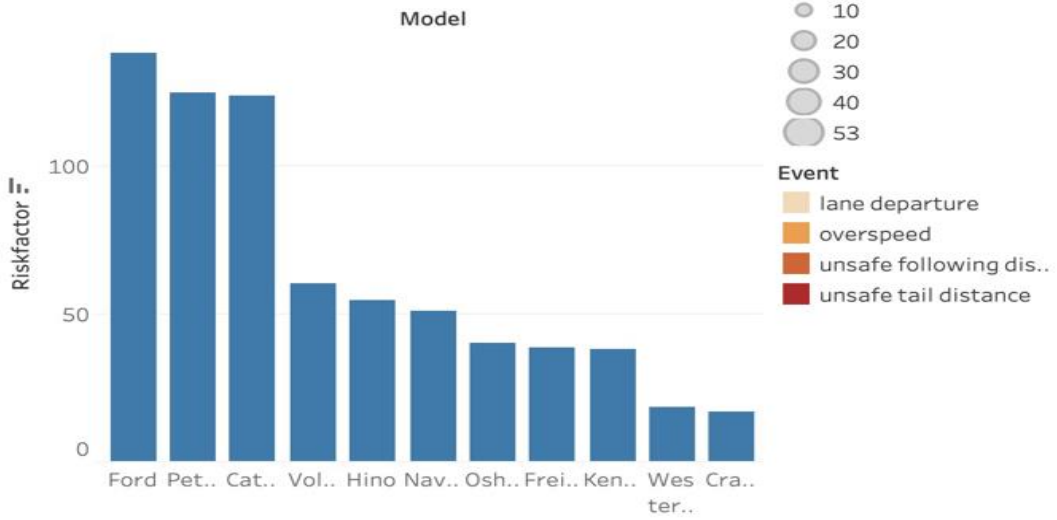


DASHBOARD

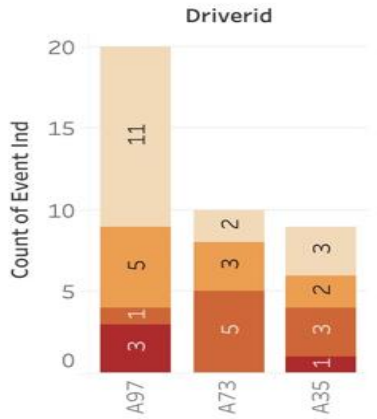
Location of Violations



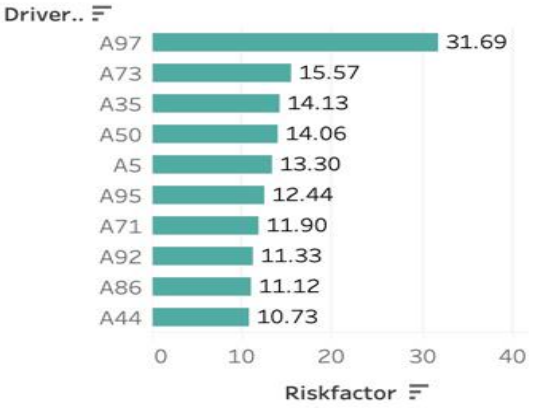
Model vs Risk Factor



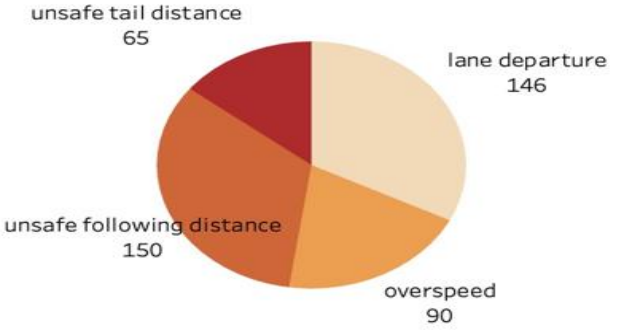
Risky Drivers by Proportion of Violations



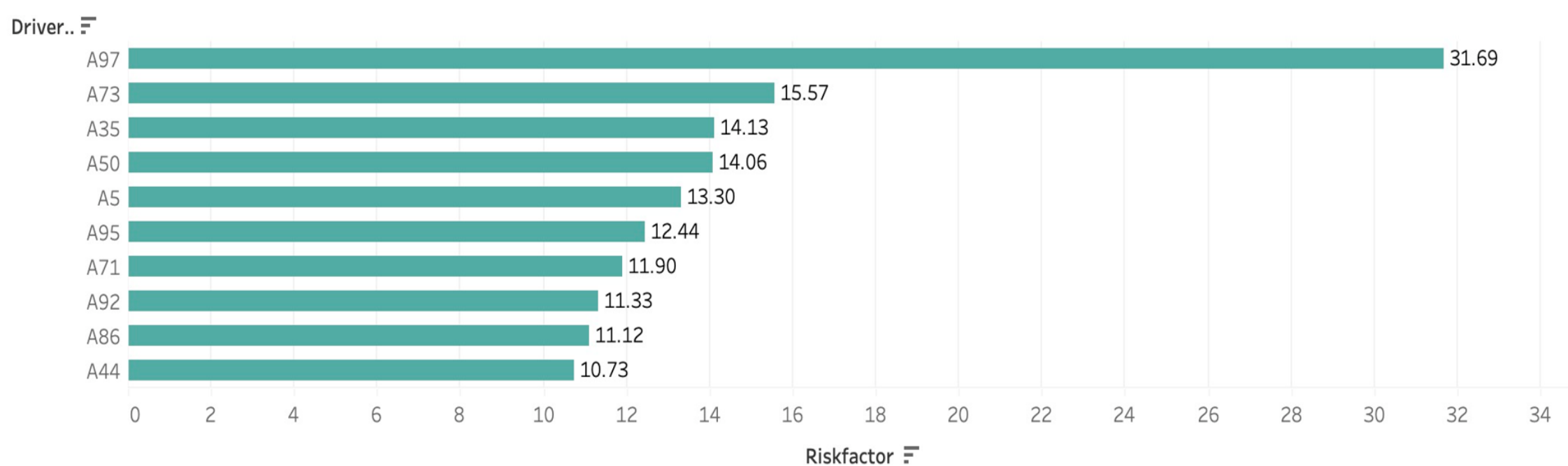
Top 10 Drivers by Risk Factor



Accidents by Event Categories



TOP 10 DRIVERS BY RISK FACTOR



- The above graph observes drivers with highest risk ratings. By analyzing the risk factor computed by the company, it is apparent that Driver "A97" is the most hazardous driver with a risk factor of 31.69.

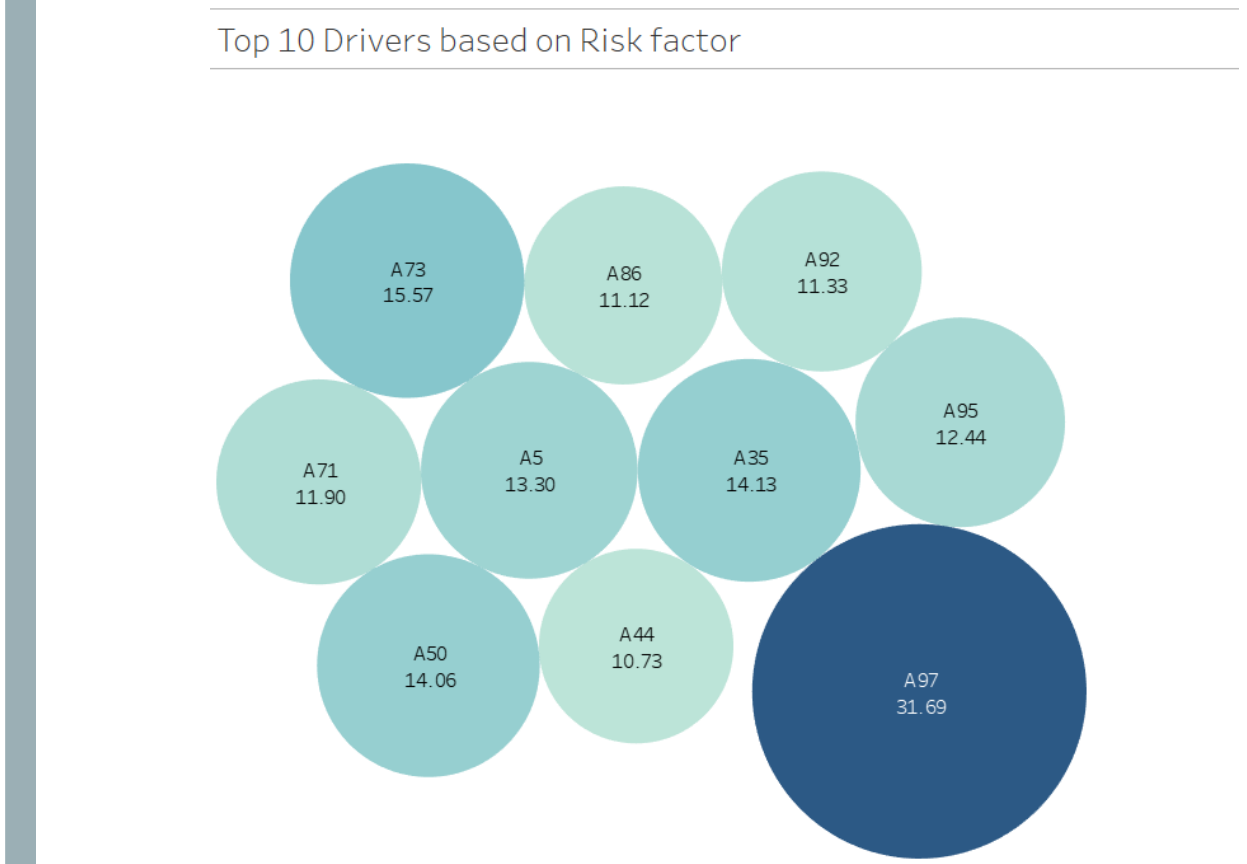
IDENTIFYING DRIVERS WITH HIGHEST RISK FACTOR

Driver A97 has the highest risk factor of 31.69. However, there is a large gap between A97 and other drivers in terms of risk factor

Drivers with a high risk factor fall in the range of 10 and 15

INSIGHTS

Driver A97 could be a one-off instance which can be dealt with separately.



IDENTIFYING DRIVERS WITH LOWEST RISK FACTOR

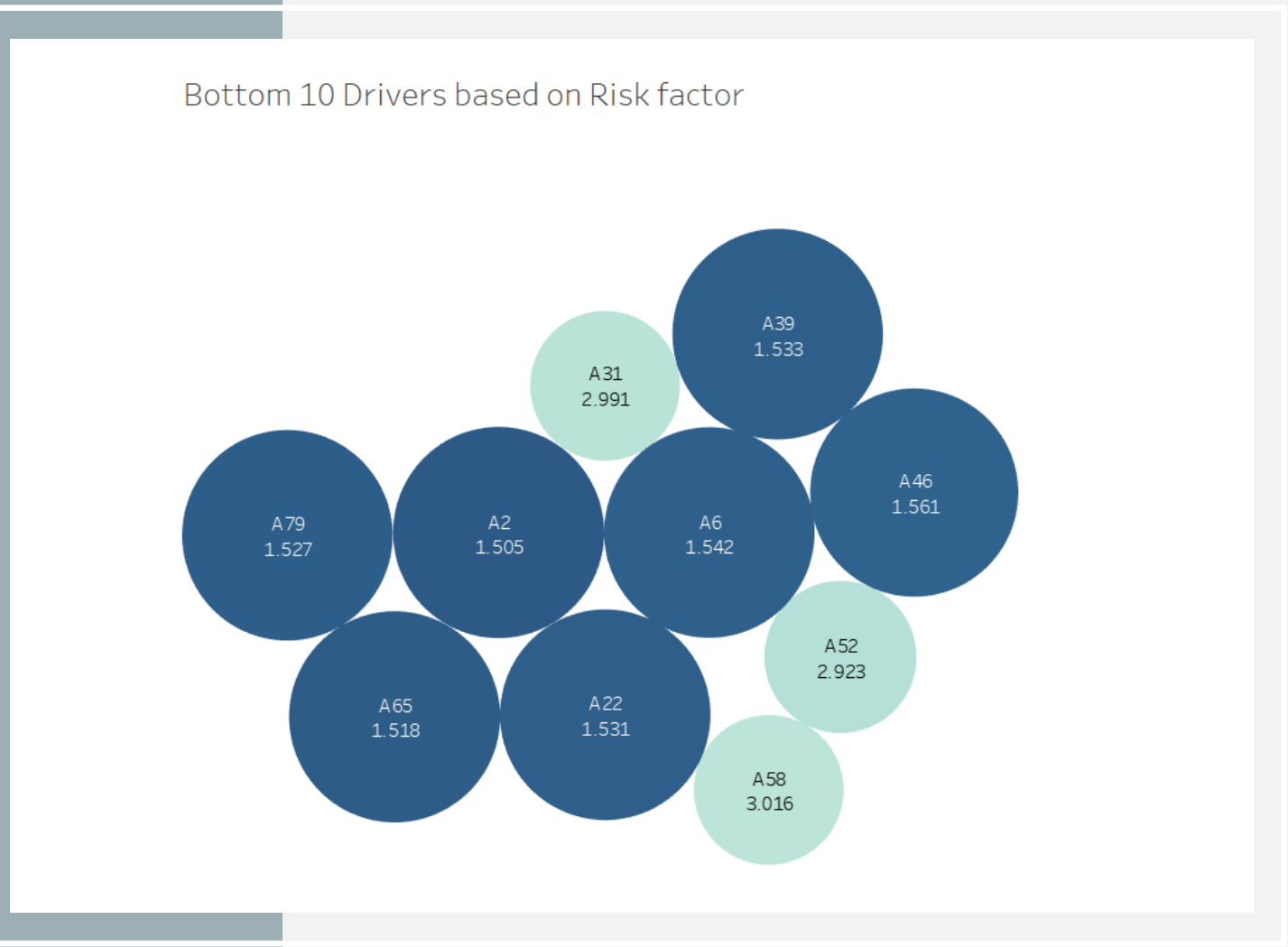
Driver A2 has the least risk factor of 1.505.

Drivers A65, A79 and A22 follow

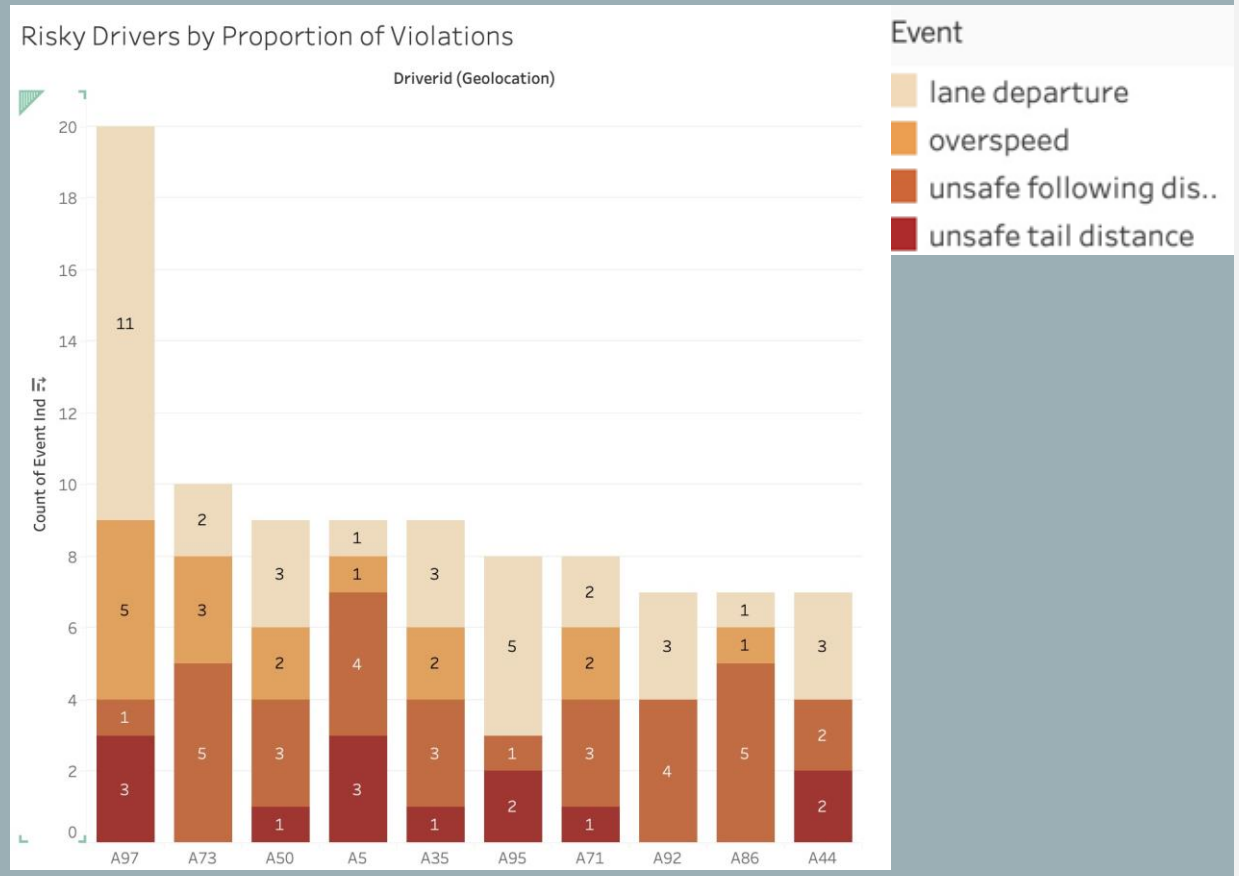
Drivers with a lower risk factor fall in the range of 1.5 and 3

INSIGHTS

Many of the drivers are in the lower range of risk factor which means most of them follow the set guidelines while on road



RISKY DRIVERS BY PROPORTION OF VIOLATIONS

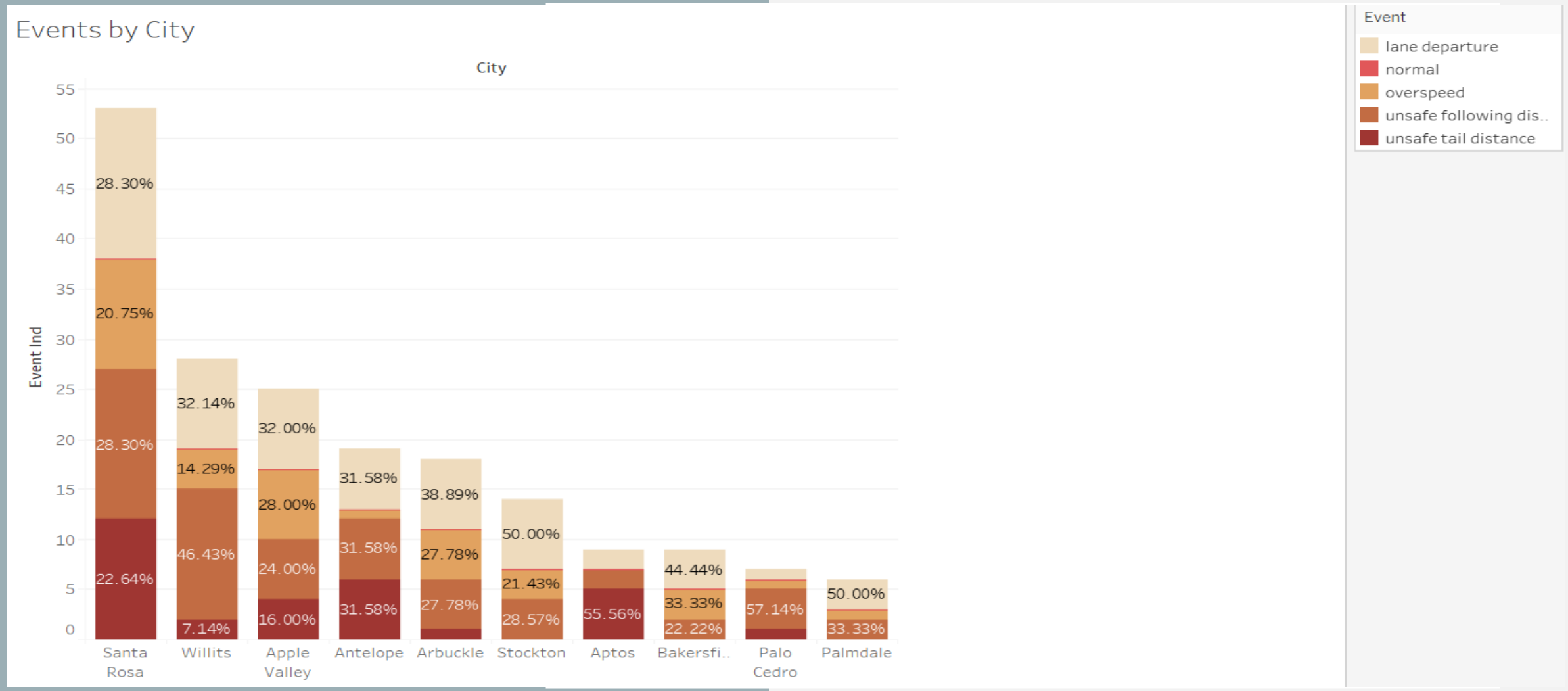


- The graph shows the proportion of violations by the top three risky drivers.
- “Lane Departure” is one of the highest occurring violation with driver “A97” followed by “Overspeed”.
- Similarly, “Overspeed” is the main violation associated with driver “A73” whereas, “lane Departure” and “Overspeed” are the concerns related to driver “A35”.

RECOMMENDATIONS

- Incentivizing Safe Driving: Offering incentives to drivers who maintain safe follow distances and avoid lane departure, as a way to encourage them to adopt safe driving practices.
- Providing training sessions for drivers to improve their driving skills, specifically on maintaining a safe follow distance and avoiding lane departure.

EVENT BREAKDOWN AMONG CITIES WITH HIGHEST EVENTS



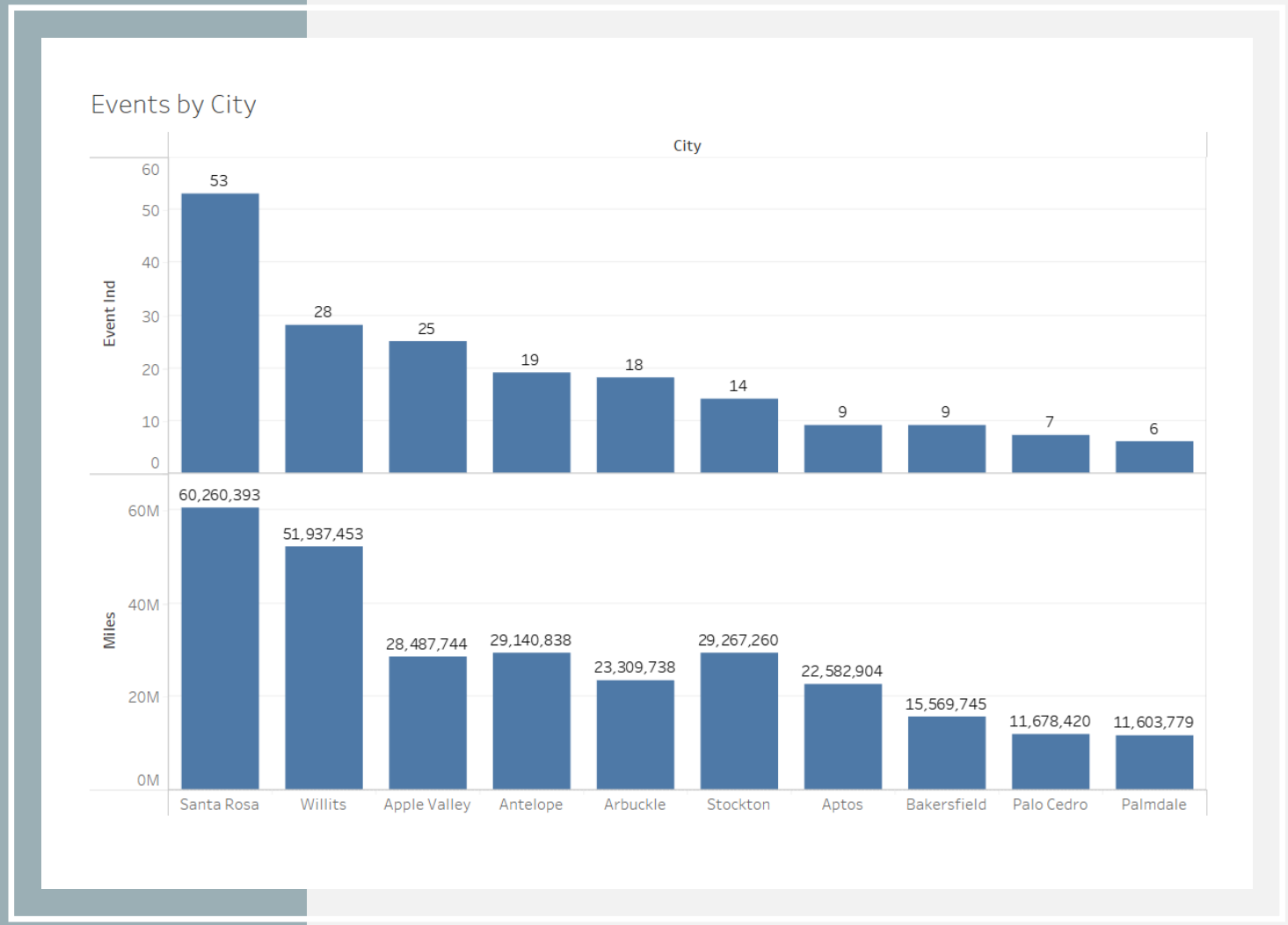
Insight - Out of cities with higher events, drivers are mainly departing their lanes and maintaining unsafe following distance

CITIES WITH HIGHER EVENTS AND MILES DRIVEN

Santa Rosa has the highest number of events followed by Willits and Apple Valley

INSIGHTS

Total number of miles driven by drivers is correlated to the number of incidents

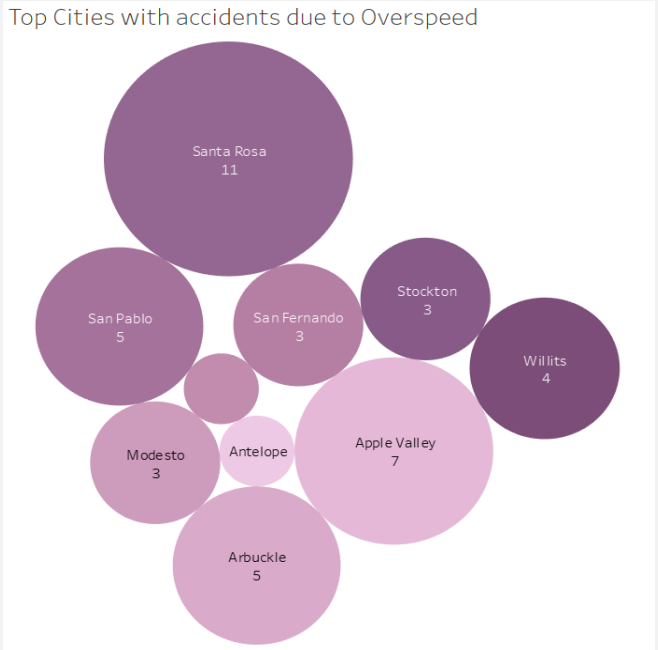


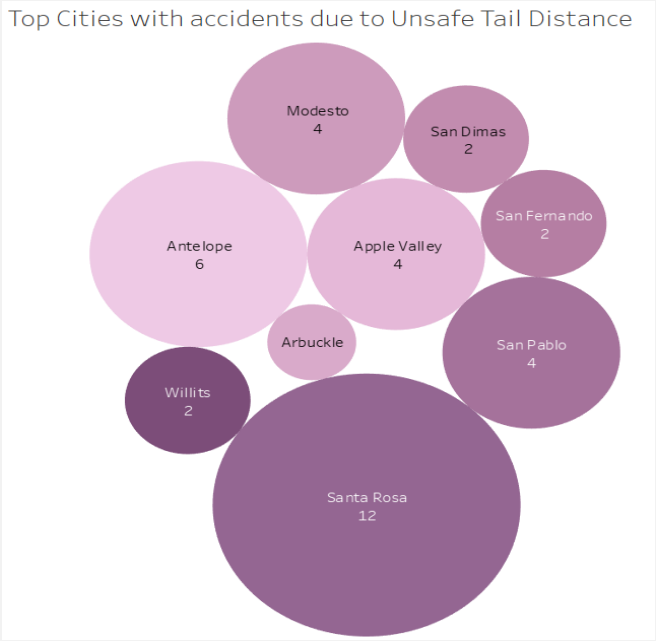
TOP CITIES WITH ACCIDENTS BY EVENT CATEGORIES

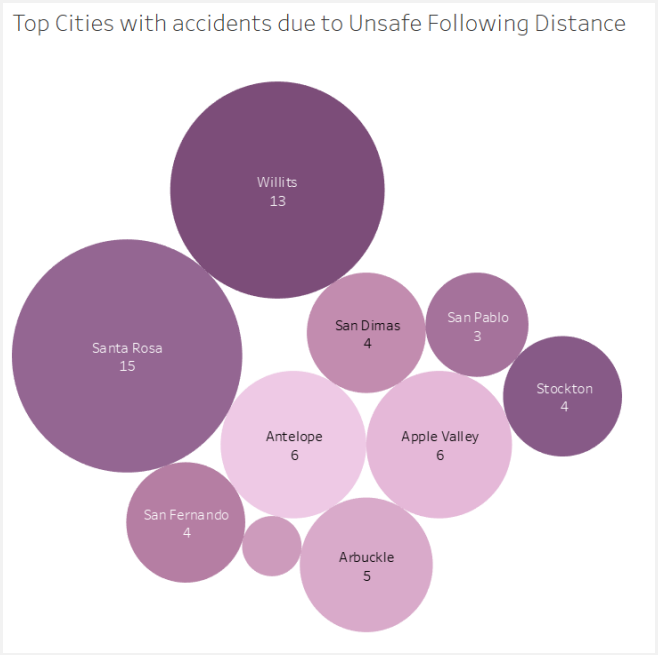
- By observation, Santa Rose in California has the highest number of violations.
- Other cities include San Pablo, San Fernando, Willits.

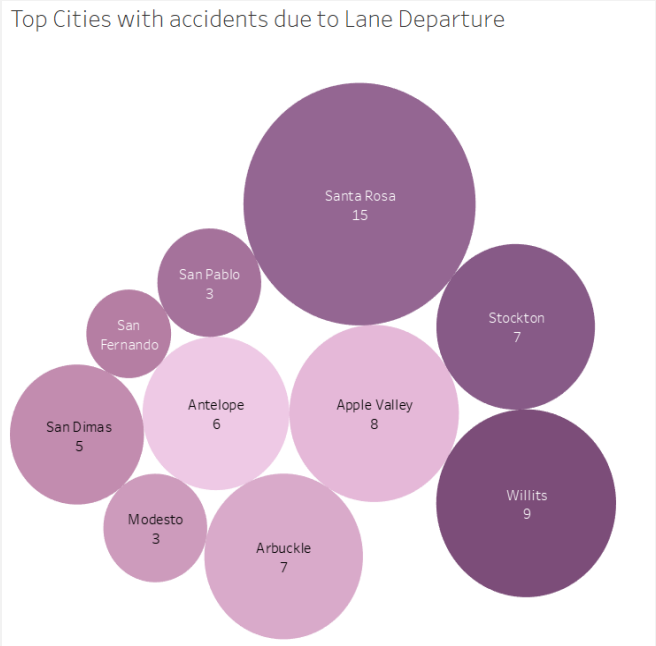
RECOMMENDATIONS

- We can conduct a review of the routes to identify areas that are prone to congestion and accidents and avoid them or reroute the drivers when necessary.





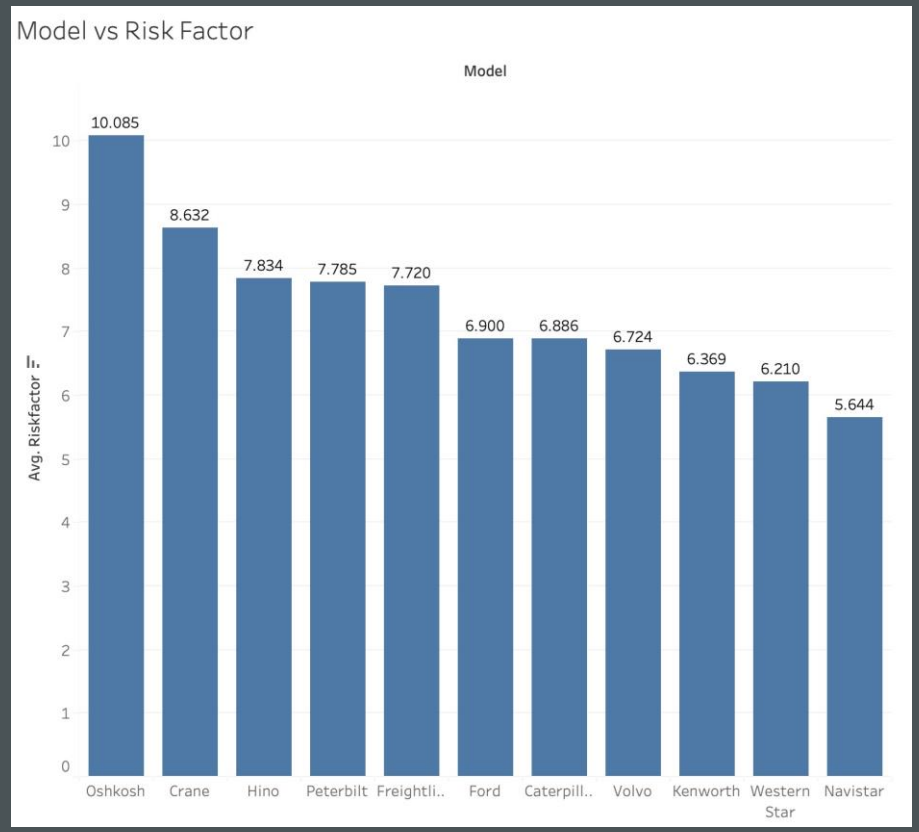


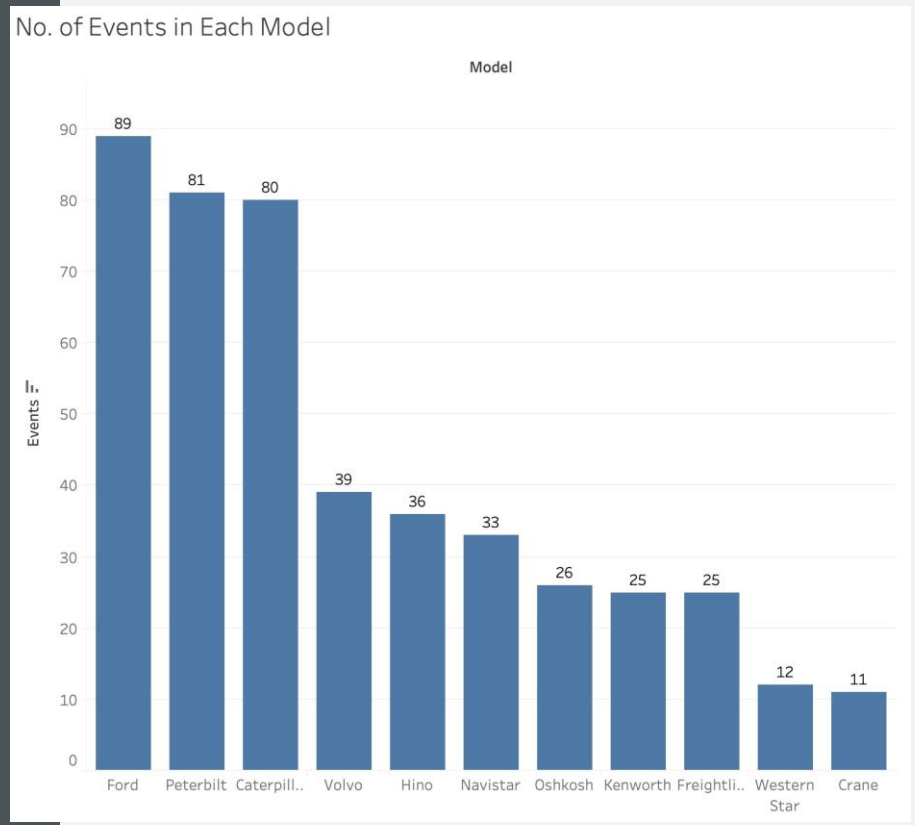


RISK FACTOR BY MODEL

- Although **Ford** has the maximum number of trucks and maximum number of events, **Oshkosh** has the highest risk factor.
- Nevertheless, it should be noted that the risk factor is predominantly influenced by the driver rather than the model.

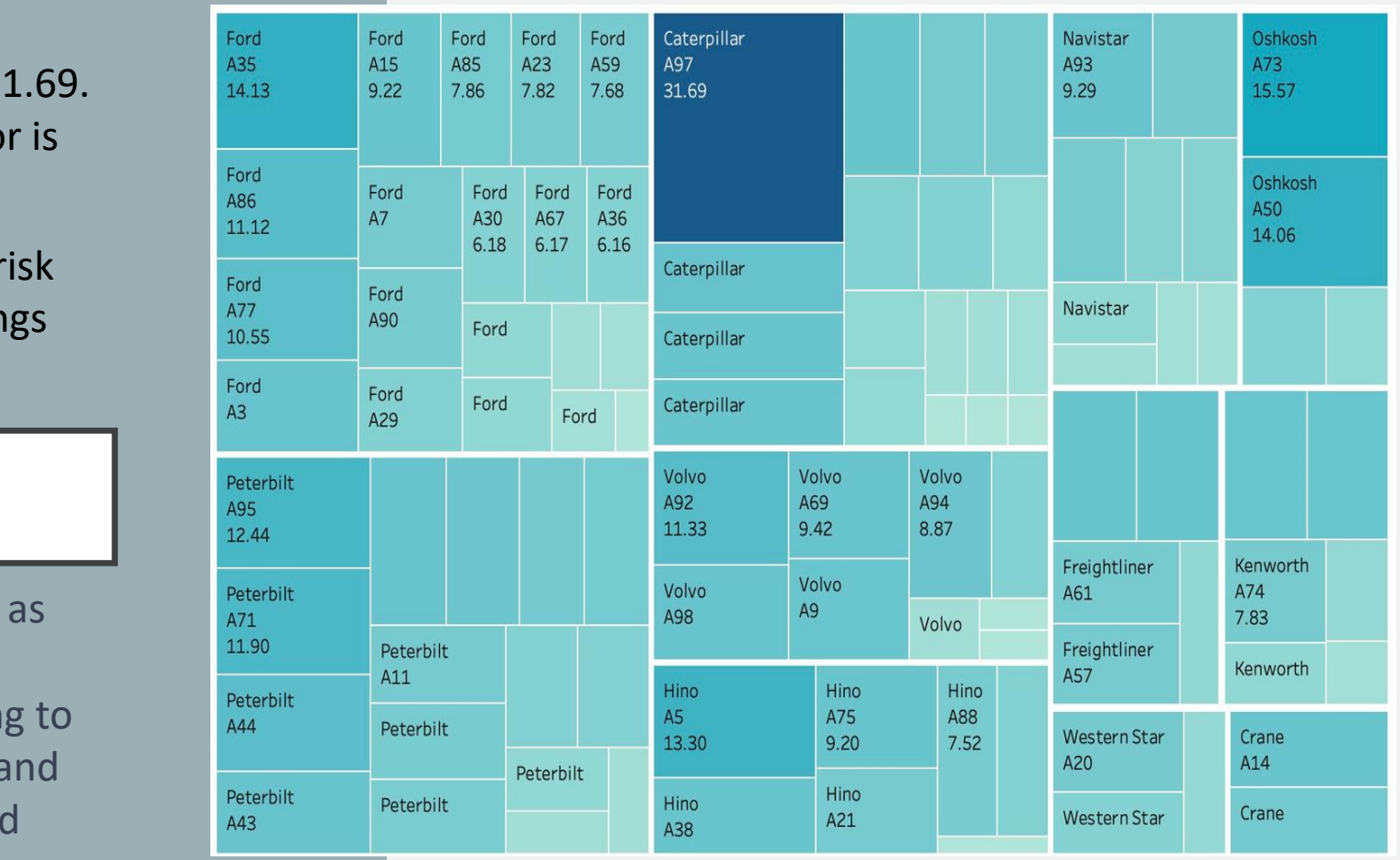
Model	Number of trucks
Ford	20
Caterpillar	19
Peterbilt	16
Navistar	9
Volvo	9
Hino	7
Kenworth	6
Freightliner	5
Oshkosh	4
Western Star	3
Crane	2
Grand Total	**100**





RISK FACTOR BY MODEL & DRIVER ID

- As we can see in the heat map, A97 driver has the highest risk factor of 31.69. By observation, the highest risk factor is associated with Caterpillar model.
- Other drivers with significantly high risk rating are A73 and A35. The risk ratings are 15.57 and 14.13 respectively.

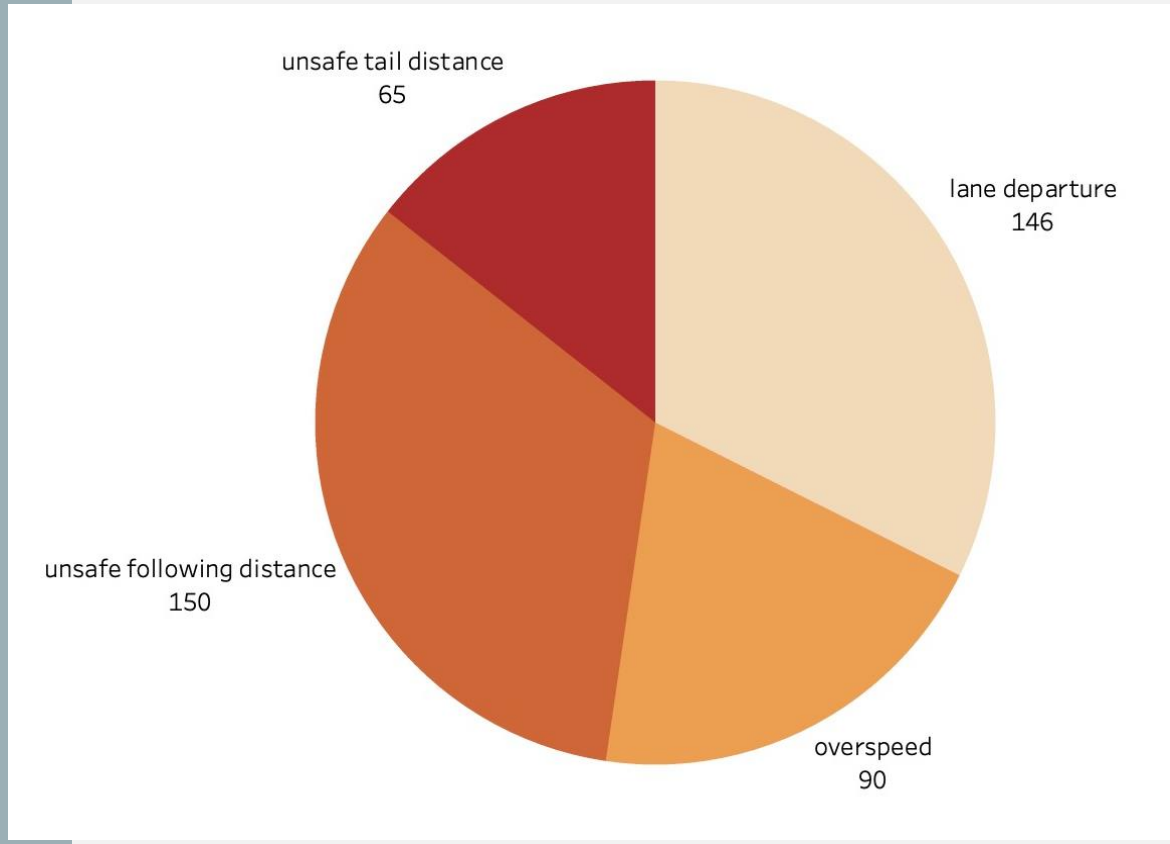


RECOMMENDATION

- For drivers with high risk ratings such as A73 and A35, it may be beneficial to provide additional training or coaching to help improve their driving behaviors and reduce the risk of incidents. This could include defensive driving techniques, avoiding distractions while driving, and adhering to safe driving practices.

ACCIDENTS BY EVENT CATEGORIES

- The graph indicates that the highest number of violations occur as a result of "Unsafe Follow Distance," specifically 150 violations.
- Additionally, violations resulting from "Lane Departure" are also notably high, at 146.
- The remaining violations are attributed to "Unsafe Tail Distance" and "Overspeed."



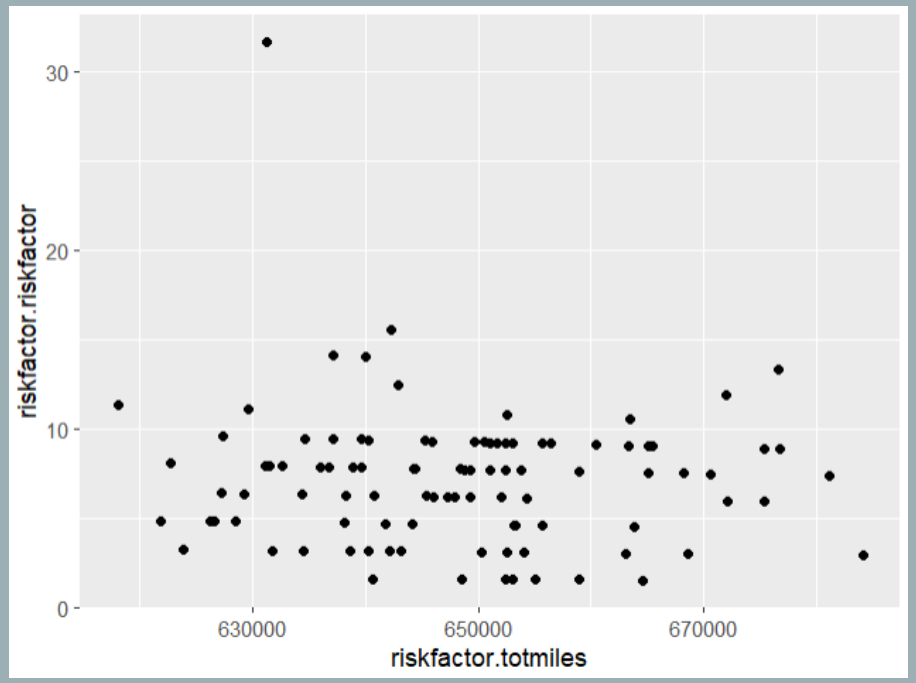
RECOMMENDATIONS

- Vehicle Maintenance: Ensuring that all vehicles are maintained regularly to avoid any mechanical issues that could cause lane departure or sudden stops leading to unsafe follow distances.
- Real-time Monitoring: Installing real-time monitoring devices that alert drivers when they violate safe driving practices such as following distance, lane departure, and overspeeding.

RECOMMENDATIONS & CONCLUSION

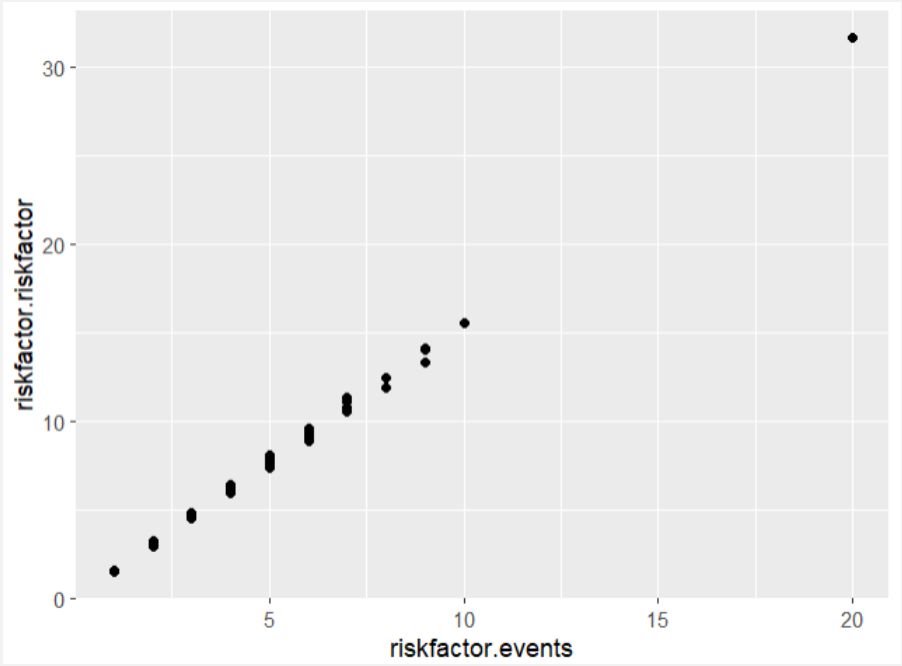
- The risk factor was determined not only by the driver, but also by external factors such as truck model, geographic location, and total miles driven
- It is observed that most of the violations occurred in the northwestern part of California.
- A comprehensive approach that includes a combination of driver training, proactive maintenance, and real-time monitoring of driver behavior and equipment performance can help reduce risk factors and improve safety on the road.
- To further improve safety and reduce risk factors, it may also be helpful to implement a proactive maintenance program for all equipment, including regular inspections and repairs to identify and address potential issues before they become major problems. Additionally, implementing telematics systems that track driver behavior and equipment performance in real-time can help identify high-risk behaviors and address them before they result in incidents.

SCATTER PLOT IN R TO DETERMINE RELATION BETWEEN VARIABLES




```
> cor(rf_model$riskfactor.totmiles,rf_model$riskfactor.riskfactor)
[1] -0.08454524
```

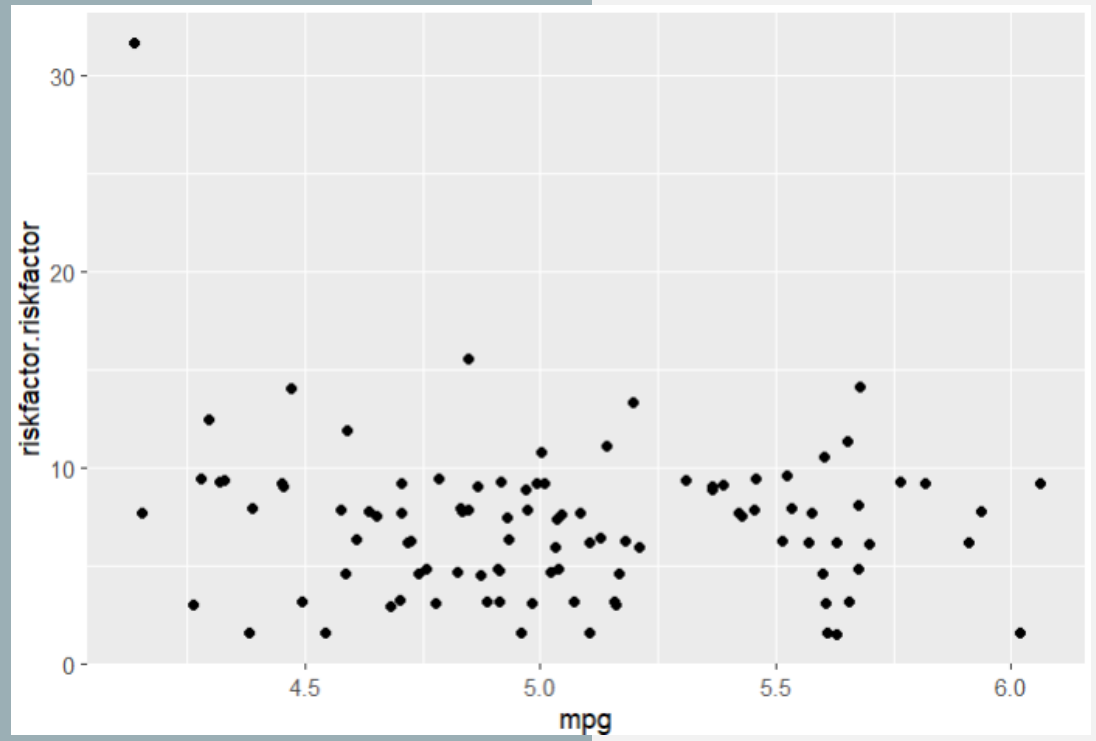
- In this exercise, we wanted to check if there is a correlation between risk factor and total miles
- As we can see, there wasn't any correlation




```
> cor(rf_model$riskfactor.events,rf_model$riskfactor.riskfactor)
[1] 0.9989056
```

- There is a high correlation between events and risk factor which is intuitive since risk factor is derived from events

SCATTER PLOTS IN R TO DETERMINE RELATION BETWEEN VARIABLES




```
> cor(rf_model$mpg,rf_model$riskfactor.riskfactor)
[1] -0.1419668
```

- In this exercise, we wanted to check if there is a correlation between risk factor and total miles per gallon
- As we can see, there is no significant correlation

FOOD FOR THOUGHT – DOES CARRYING OUT REGRESSION ANALYSIS HOLD VALUE?

```
> rf.lm <- lm(riskfactor.riskfactor~ ., data = rf_model)
> summary(rf.lm)

Call:
lm(formula = riskfactor.riskfactor ~ ., data = rf_model)

Residuals:
     Min       1Q   Median       3Q      Max
-0.28876 -0.03917 -0.00580  0.02946  0.35641

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)          7.343e+00  4.006e-01  18.329  < 2e-16 ***
riskfactor.events    1.558e+00  3.511e-03 443.911  < 2e-16 ***
riskfactor.totmiles -1.138e-05  6.153e-07 -18.491  < 2e-16 ***
modelCrane          -4.350e-02  6.438e-02  -0.676  0.50110
modelFord           -1.615e-02  2.784e-02  -0.580  0.56339
modelFreightliner   -6.785e-02  4.336e-02  -1.565  0.12137
modelHino           -7.974e-02  3.837e-02  -2.078  0.04065 *
modelKenworth       -3.015e-02  4.047e-02  -0.745  0.45832
modelNavistar       -1.630e-02  3.507e-02  -0.465  0.64321
modelOshkosh        -2.173e-02  4.788e-02  -0.454  0.65111
modelPeterbilt      -7.906e-02  2.951e-02  -2.679  0.00884 **
modelVolvo          -3.014e-02  3.503e-02  -0.861  0.39188
modelWestern Star   -4.497e-03  5.343e-02  -0.084  0.93312
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.08562 on 86 degrees of freedom
Multiple R-squared:  0.9996,    Adjusted R-squared:  0.9995
F-statistic: 1.744e+04 on 12 and 86 DF,  p-value: < 2.2e-16
```

- The team carried out regression analysis in an attempt to estimate the risk factor by using factors like events, model and number of miles driven
- From the R2 squared values, we can see that the model is nearly perfect fit.