Traffic Safety of Los Angeles

Analysis, Conclusions and Recommendations

September 17, 2019

Saul E Villarreal sevillar@syr.edu

Corinna Fabre cdfabre@syr.edu

Veasna Oum voum01@syr.edu

Jenn Dandrea jdandrea@syr.edu

Table of Contents

[Project Abstract](#_heading=h.1fob9te) **3**

[Contribution Statement](#_heading=h.3znysh7) **3**

[Introduction](#_heading=h.2et92p0) **6**

[The Dataset](#_heading=h.tyjcwt) **6**

[Data Analysis Methods (Clarify)](#_heading=h.3dy6vkm) **9**

[Analysis Methods and Summary](#_heading=h.1t3h5sf) **9**

[Conclusions](#_heading=h.4d34og8) **26**

[References](#_heading=h.17dp8vu) **26**

[Appendix](#_heading=h.3rdcrjn) **27**

# Project Abstract

This project uses published, historic data from Los Angeles categorizing all reported traffic incidents from 2010-2019. Data exploration included creating visualizing on demographics, location, yearly trends. A SQL database was also constructed to allow interaction between SQL and R. Multiple-Linear Regression models, time-series analysis and visualizations assisted in conclusions and recommendations.

Several basic graphs for analysis include demographic, age, and area location of collisions. Additionally, weather data from a Los Angeles downtown weather station was integrated into the data to assess how weather affects traffic accidents.

The main objective of the data analysis is to provide recommendations to the city council of Los Angeles to help reduce traffic collisions, by finding patterns and predictive models to aid in their planning of allocation of budget funds. Desired outputs were to identify age demographic and common location of repetitive collisions to help increase awareness and additional measures to break the patterns. Suggestions would include traffic cameras, additional police presence, additional traffic lights, or driving restrictions. A key finding identified is the correlation between the reduction of LA public transportation and an increase of collisions.

Conclusions from this project include:

* Public Transportation
  + Partner with local businesses to offer reduced rates to employees
  + Partner with local businesses to establish alternative work schedules or work from home opportunities
  + Focus on drivers ages 20-34
  + Alternate side of street parking to allow for more buses
* Additional Monitoring
  + Reinstate traffic officers who have been transferred
  + Revisit state law prohibiting the use of speed radars
  + Reinvest profits from recent tax increase to overhaul streets

# Contribution Statement

Jenn Dandrea: Applied the ‘cleaned data script’ to start. Created individualized data frames to subset the data. This included data frame looking at the premise descriptions and count of incidents at each type of area, ie, bar, airport, street, school. However, ‘street’ was a dominant location and therefore separating out the other locations was not required. After that exploration, a data frame was created with specific streets provided from another team members analysis to inspect. The data from these specific street were then mapped, over time, to see if there were obvious patterns. It is clear that in 2015, there were fewer incidences. This is seeing on Sepulveda Boulevard on the north end, with a drastic reduction in accidents from all other years. Deeper visualization adds the points to reflect the time of day in which the incident occurred. By analysing the time of day, one can identify targeted solutions to prevent incidents, as discussed in the analysis summary.

Saul: Scripted the clean data which we all worked off of. Cleaning the data entailed tasks such as transforming a character date string to date, converting intra-hour data to the nearest hour, and creating an lapply function to handle the latitude and longitude JSON formatted data. Requested and downloaded data from an LA weather station through the NOAA website. The weather data was in daily format, this required merging the original clean data set data to daily to include weather data such as daily precipitation, average wind speed, and average temperature. Created 3 functions to perform descriptive statistics and percentile analysis. desStats performed descriptive stats to all data given a vector, desStats2 performed descriptive stats given a numeric vector and a categorical subgroup, such as month and incidents or zip codes and incidents. quantTable took a numeric vector and a group category, and returned a table of percentiles from 5-95% in 5% increments. Using these equations and plots I was able to identify that there was a significant difference between the incident data between 2010-2014, and 2015 to 2018 which became really apparent as a similar finding from my other peers. Created a multiple-linear regression equation trying to predict traffic incidents, the equation included wind, temperature, precipitation, and month and year categorical variables. The equation yielded an adjusted r-squared of .82%, and the test data set had an error of - 400 to +200 incidents monthly, with an rmse of 217.8.

Corinna: Performed data exploration on age, location, gender, descent, and neighborhood demographics. Located, integrated, and cleaned public transportation ridership data and overlaid that data on collision linear models to explore correlation. Created various visual representations of the demographics information and timelines of historical collision data. Created a predictive linear model to plot potential collision rates. Created plots that broke down demographics, collision rates, and time. Located, integrated, and cleaned Los Angeles population data and created a visualization of the percent of collisions by descent, as compared to their percentage of the general population of the city.

Veasna: Executed core analysis which was divided into two parts: leveraging the odbc connection package in R to connect to SQL Server for data cleanup, metadata formatting and aggregation. Used SQL connection to share dataset among the group. Used R for further analysis such as looking at the data central tendency, histogram of collisions and time series projection. Specific steps taken for Time Series analysis:

1. Determined the Traffic data Central Tendency
   1. Median collisons count = 8
2. Create a histogram based on the assessment of the median, plotted to narrow down the analysis
3. Further narrowed down findings to the 4th quartile range based on the distribution of collisions
4. Selected the top five streets from the boxplot which was then used in a Time Series forecast. For the current dataset, the top five streets included:
   1. Sepulveda
   2. Sunset
   3. Ventura
   4. Vermont
   5. Western
5. Developed and executed a Regression Integrated Moving Average (RIMA Model)
   1. This model was used to project the next two years of collision rate from step 4

# Introduction

Traffic safety is becoming a never ending battle in large cities. Although cars are getting ‘smarter’ with enhanced safety features, data shows that traffic accidents are increasing. The key issue to solve is how to reduce these traffic collisions to increase public safety in Los Angeles. Regulation is not always the right answer. Using data published over the last decade, the team identified facts from the information to include age, race, weather impacts, and location of higher collisions. The key question to answer is, how can we use the historical data, to predict accident indicators and prevent the problem from occurring.

Additional data pulled into the model include weather data and transportation changes from LA. Key findings included

* + - 1. Public transportation and collisions are correlated
      2. Early drivers are more liking to be in an accident
      3. 58% of collisions involve men
      4. Females between 20 and 30 are the highest at risk for a collision
      5. Most collisions occur during afternoon commuting times
      6. There is an increase in collisions during all months but especially trending up are the January, February, July and August months
      7. 5 key main street w/ collision rate in the top 4th quartile range

# The Dataset

This dataset reflects traffic collision reports for all accidents that took place in the city of Los Angeles, beginning in 2010. It should be noted that, according to the city, “This data is transcribed from original traffic reports that are typed on paper and therefore there may be some inaccuracies within the data”

|  |  |
| --- | --- |
| **DR Number** | “Division of Records Number: Official file number made up of a 2-digit year, area ID, and 5 digits” |
| **Date Reported** | Originally formatted as: MM-DD-YYT00:00:00.000  New format: MM-DD-YY |
| **Date Occurred** | Originally formatted as: MM-DD-YYT00:00:00.000  New format: MM-DD-YY |
| **Time Occurred** | Originally formatted in 24-hour military time |
| **Area ID** | “The LAPD has 21 Community Police Stations referred to as Geographic Areas within the department. These Geographic Areas are sequentially numbered from 1-21” |
| **Area Name** | “The 21 Geographic Areas or Patrol Divisions are also given a name designation that references a landmark or the surrounding community that it is responsible for. For example 77th Street Division is located at the intersection of South Broadway and 77th Street, serving neighborhoods in South Los Angeles” |
| **Reporting District** | “A code used in producing reports to group data into geographic sub-areas within an area” |
| **Crime Code** | “Indicates the crime committed. For this dataset - all Crime Code 997” |
| **Crime Code Description** | “Defines the Crime Code provided. For this dataset - all Traffic Collision” |
| **MO Codes** | “Modus Operandi: Activities associated with the suspect in commission of the crime” |
| **Victim Age** | Formatted as a two digit number |
| **Victim Descent** | Originally formatted as a single letter, the descent codes include:  A - Other Asian  B - Black  C - Chinese  D - Cambodian  F - Filipino  G -Guamanian  H - Hispanic Latin Mexican  I - American Indian Alaskan Native  J - Japanese  K - Korean  L - Laotian  O - Other  P - Pacific Islander  S - Samoan  U - Hawaiian  V - Vietnamese  W - White  X - Unknown  Z - Asian Indian  The new format includes full text titles only |
| **Victim Sex** | Originally formatted as a single letter, including:  F - Female  M - Male  X - Unknown  The new format includes full text titles only |
| **Premise Code** | “The type of structure or location where the incident took place” |
| **Premise Description** | “Defines the Premise Code provided” |
| **Address** | “Street address of crime incident rounded to the nearest hundred block to maintain anonymity” |
| **Cross Street** | “Cross Street of rounded Address” |
| **Location** | “The location where the crime incident occurred. Actual address is omitted for confidentiality. XY coordinates reflect the nearest 100 block” |
| **Zip Codes** | Five digital strings |
| **Census Tracts** | Three digit identifying number |
| **Precinct Boundaries** | Three digit identifying number |
| **LA Specific Plans** | One and two digit identifying numbers |
| **Council Districts** | One and two digit identifying numbers |
| **Neighborhood Councils (Certified)** | One and two digit identifying numbers |

Table 1. Features in the dataset.

*\*Please note: descriptions of both original and new formatting are included here to better facilitate cross-referencing and review*

Other data used in this analysis include:

Los Angeles city demographics -- This data structured in two columns, descent name (reformatted to match those included in the collision data) and the number of people in that community.

Public transportation ridership -- This data was structured in two columns, the date (in year and month format) and the number of riders. This data represents all ridership and does not differentiate between bus and rail.

NOAA Los Angeles downtown weather station -- The LA weather station data had 49 columns and included data from 2010-2018. The data used for the analysis was daily precipitation, average wind speed, and average temperature.

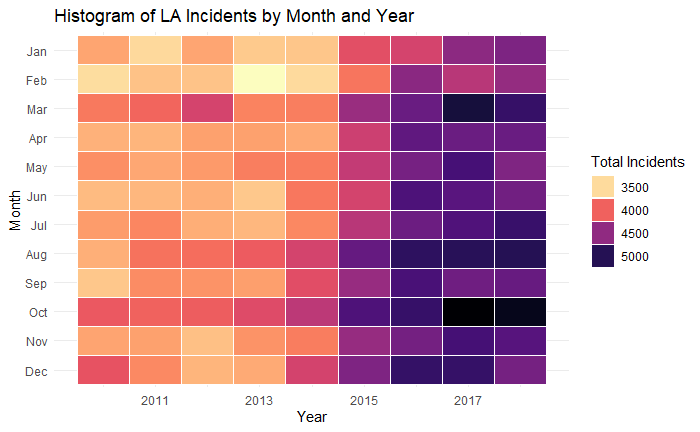
# Data Analysis Methods and Summary

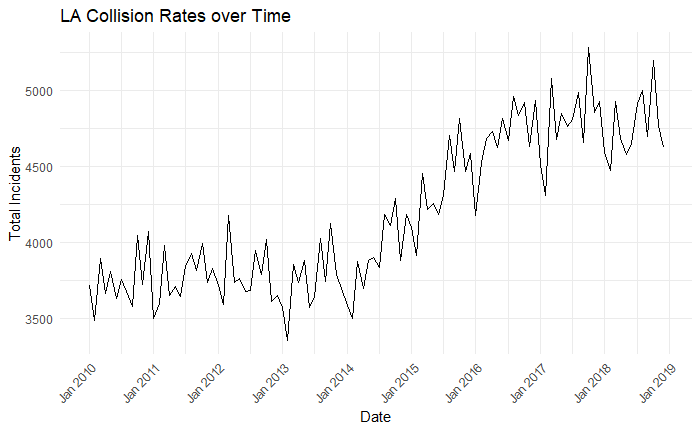
The raw data was obtained as raw data (via Kaggle) from the City of Los Angeles and reformatted to meet the teams needs. Key changes include: reformatting the date occurred, the time occurred, and the date reported to comply with the requirements of the packages used and to make for simpler human analysis. Other changes include removing unnecessary or redundant instances, cleaning text for either unneeded characters or reformatting non-standard characters. Finally, the most labor intensive change was to reformat the location column, which included latitude and longitude and was in a JSON format within the dataframe. This information was unlisted, tagged with appropriate designations, and converted into numbers.

Exploratory data analysis included basic graphing of age, location, sex, time and ethnicity of collision victims. This initial exploration allows for a recalibration of where the most impactful information might be found so further analysis can be made and recommendations can be formed.

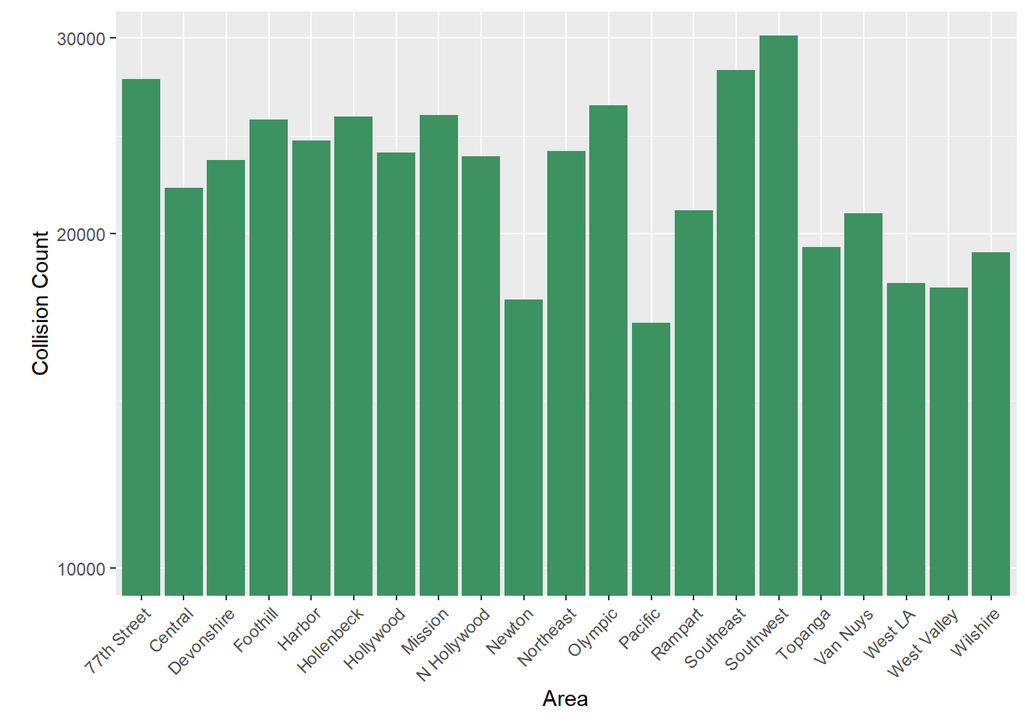
From there, further analysis such as linear modeling and predictive functions were used to breakdown the data to see which elements were the most statistically significant and where reform efforts should be focused. Two main multiple-linear regression models were created, weather vs accidents, and ridership vs accidents.

Collisions over time: These graphs plot the number of collisions over time in two ways: as a line graph and as a heat map. Both show a marked and continuous increase in total collisions month-over-month. Although this information doesn’t provide us with any actionable insights, it helps create a foundation of information to further analyze, and gives a faster, more impactful understanding of the frequency rates.

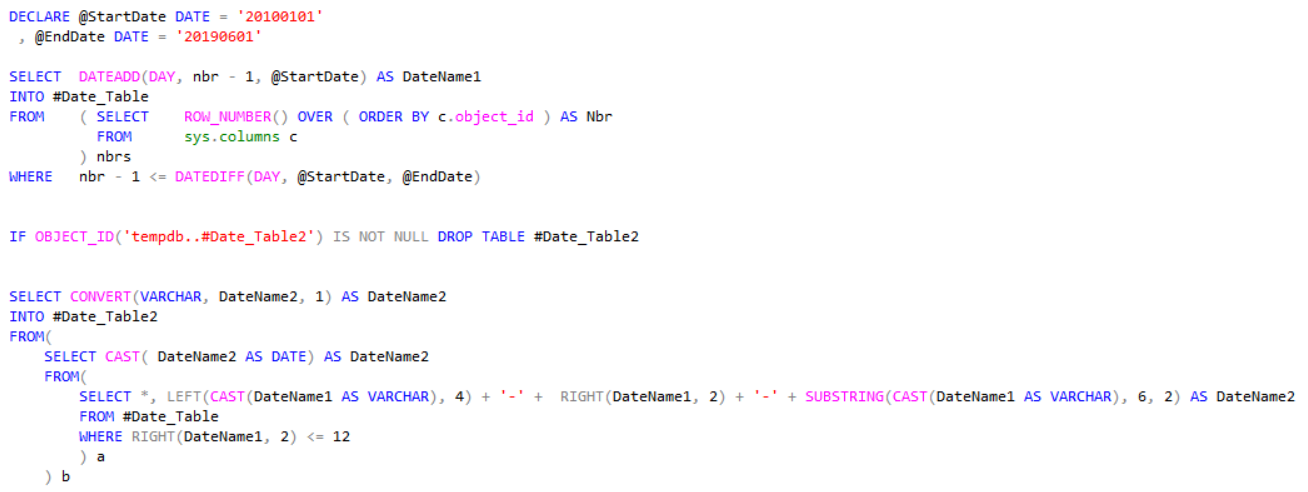




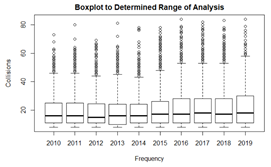
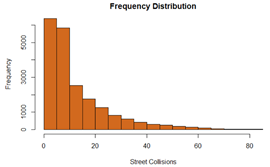
Collision count by area: This data pulls the count of collision, grouped by area in which the collision occurred. This data is pulled from the metadata, over a 10 year timeframe. This visualization can be used alongside the SQL function that was created to help monitor the top 4th quartile of problem areas to further investigate the root causes.



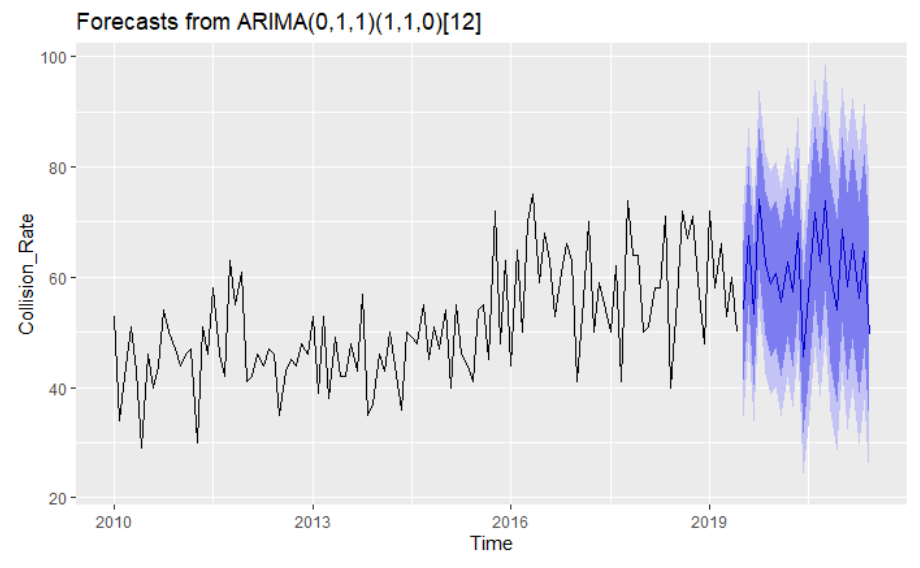
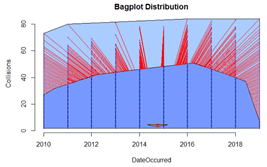
SQL & Time-Series Analysis: SQL database was used to clean, transform and aggregate the csv file to prep for R Studio analysis . Frequency distribution analysis and upper percentile analysis was used to determine and narrow down the significant collisions consideration in ARIMA model. It’s powerful forecast capability can project the accidents prediction in the next few years for the LA city council to review its city traffic and ensure that they use their limited funds/budget optimally.



*SQL Code Snippet to show some of the data cleanup/munging that was done to prepare the data for R*

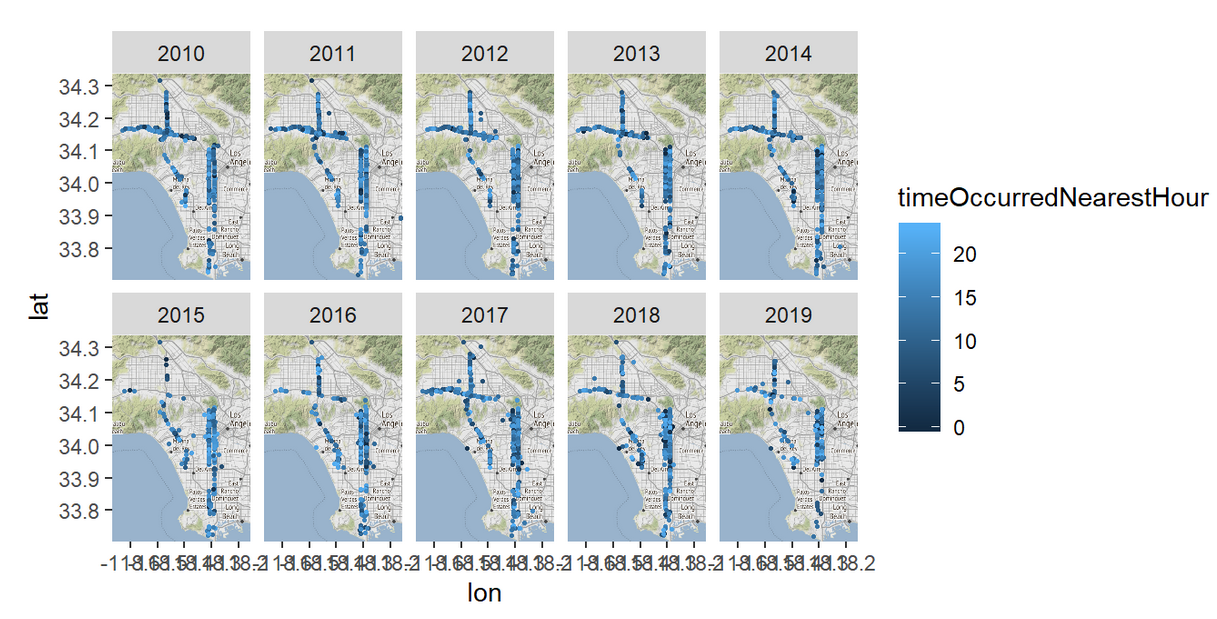


*Frequency distribution and Boxplot to determine the areas the analysis should focus on*

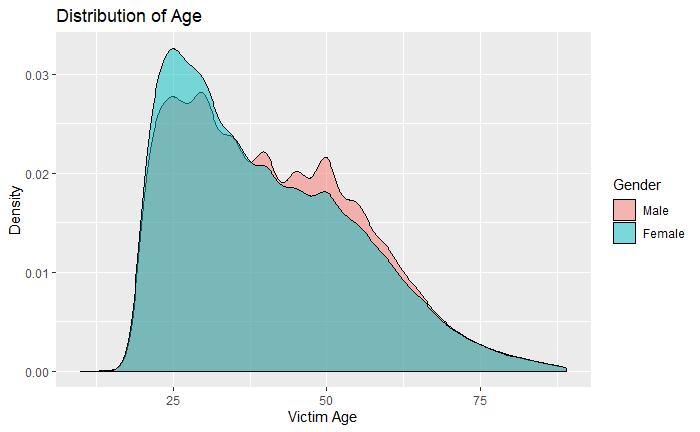


*The visualization on the right is the ARIMA model used to project the collisons outlook for the next two years*

The analyses were based on a subset of the data, focusing only on five key streets: Sepulveda, Ventura, Vermont, and Western. The lower right chart depicts the collisions forecast two years ahead. The trends were showing an increase between 18% - 25% with an 80% interval confidence for 18% spike of collisions and lower and 90% confidence interval for 18% or more spike of collisions. The data is visually plotted onto a Google street map, separated by year. The graphic below compares years from 2010 to 2019 for specific streets. The color of the dots represent the time of day of the accident. Diving deeper on location and time patterns, the city council can leverage the results to install accident reducing solutions. One key solution would be to use red light camera at key intersections. Additionally, seeing patterns in the time of day which causes more accidents, the city can encourage business to use alternate working hours to reduce the amount of people on the roads, at any given time. This would help relieve congestion and probability of more incidences.

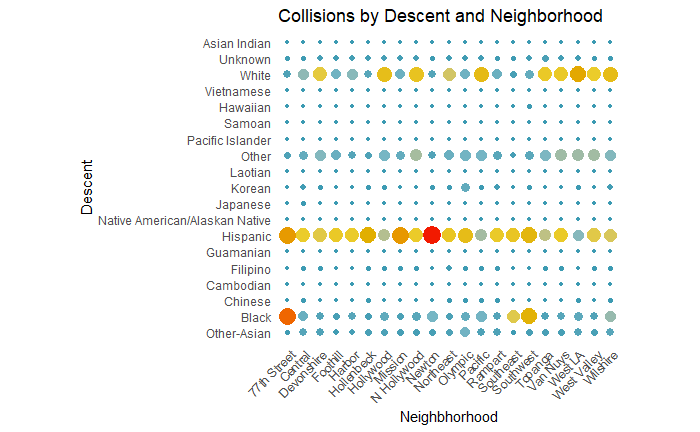


Distribution of Ages: This graph shows the distribution of ages, grouped by sex, and the count of collision from each, over a 10 year span from the metadata. Even without diving into the details, we can see at first blush that women tend to be involved in collisions when they are younger, as compared to their male counterparts. However, as age increases, we see that trend reverse and eventually level out. It should be noted that initial graphs saw a huge spike for men between the ages of 90-100. Those ages were filtered out because, upon review, it became clear that those ages were, either assigned to records that didn’t have the information when being transcribed, or wasn’t disclosed by the victim at the site of the accident.



*Density plot of collisions by age and gender*

Ethnicity: This heat map shows the frequency (plotted) collisions broken down by area and the descent of the victim. Although it clearly demonstrates marked increase for some neighborhoods and races, we found that when analysed alongside the breakdown of Los Angeles’s population by descent, no demographic stood out as being overly represented, which could have signaled a problem area that could be addressed. Instead, the breakdowns held fairly true to the general population.



*Heat map of descent and neighborhood where size and color equate to frequency of collisions*

Weather: The weather data was used to determine if there were any relationships between the number of incidents and wind speed, precipitation, and temperature. The analysis essentially had three steps: exploratory data analysis by looking at descriptive statistics and percentiles, visualizations with different sub-groups, and finally multiple-regression analysis.

Table of descriptive statistics of incidents by year:

mean median std.dev min max skewness kurtosis n

2010 3757.17 3723.00 175.81 3490.00 4073.00 0.48 -0.93 12

2011 3773.00 3780.00 155.28 3506.00 3998.00 -0.06 -1.34 12

2012 3783.25 3728.50 177.99 3598.00 4179.00 0.95 -0.42 12

2013 3752.75 3742.00 208.68 3360.00 4127.00 0.04 -0.72 12

2014 3912.50 3884.50 244.57 3502.00 4291.00 -0.07 -1.26 12

2015 4373.67 4384.50 258.56 3915.00 4813.00 0.02 -1.07 12

2016 4710.50 4709.00 216.76 4178.00 4961.00 -0.97 0.41 12

2017 4810.00 4828.50 256.06 4312.00 5284.00 -0.11 -0.55 12

2018 4751.33 4691.00 216.17 4477.00 5196.00 0.58 -0.93 12

Table of percentiles 5-95% of incidents by year:

## 5% 15% 25% 35% 45% 55% 65% 75% 85% 95%

## 2010 3541 3615 3660 3678 3719 3727 3766 3832 3951 4059

## 2011 3559 3631 3654 3705 3738 3822 3835 3867 3948 3990

## 2012 3606 3641 3674 3685 3717 3739 3768 3833 3976 4093

## 2013 3477 3576 3629 3692 3737 3747 3804 3865 3934 4074

## 2014 3548 3660 3804 3873 3882 3888 3937 4129 4183 4232

## 2015 4017 4154 4208 4252 4309 4458 4468 4500 4627 4754

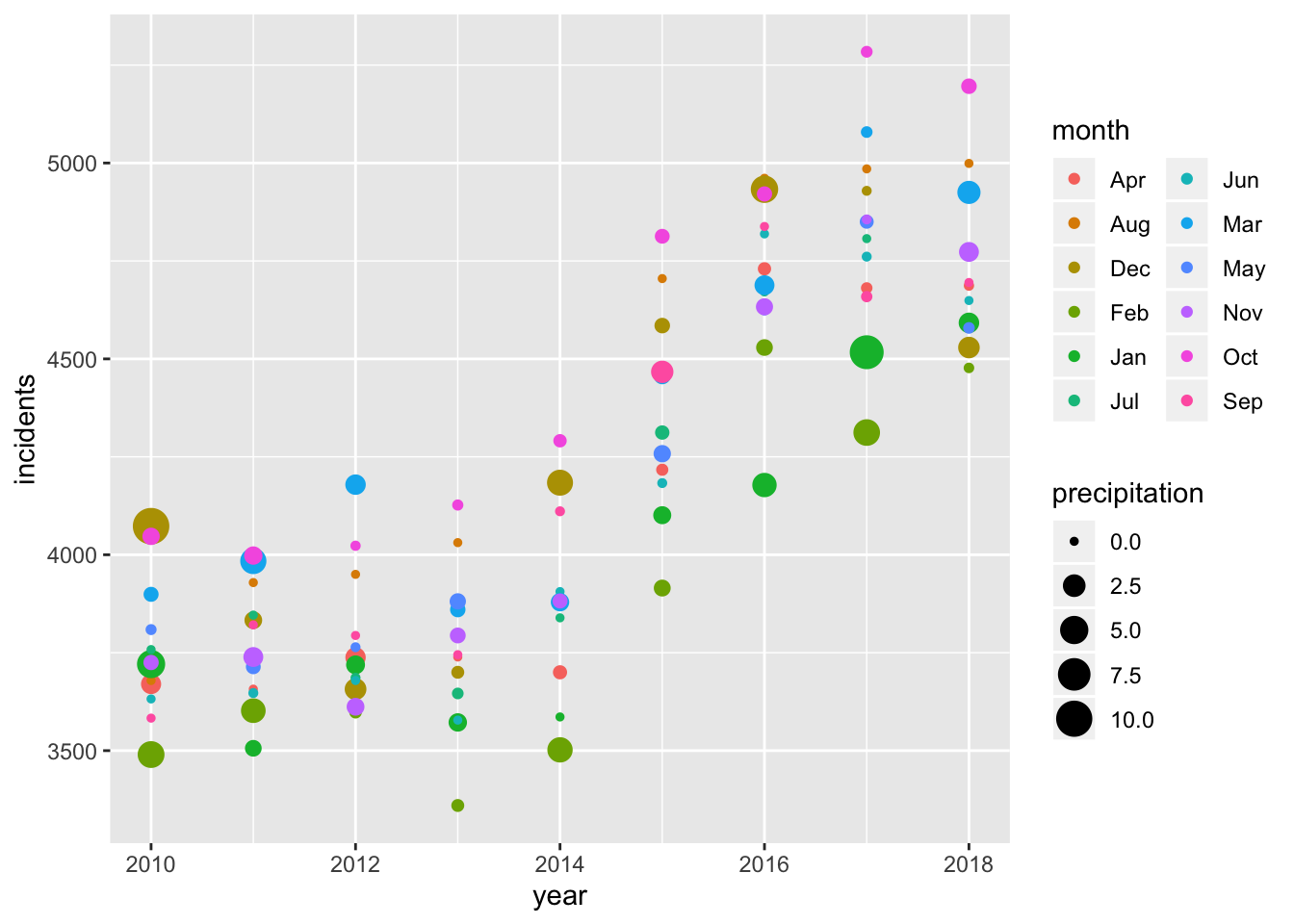
## 2016 4371 4591 4631 4665 4687 4734 4822 4859 4925 4946

## 2017 4425 4609 4676 4749 4805 4850 4867 4943 5018 5171

## 2018 4506 4562 4589 4640 4685 4699 4794 4918 4951 5088

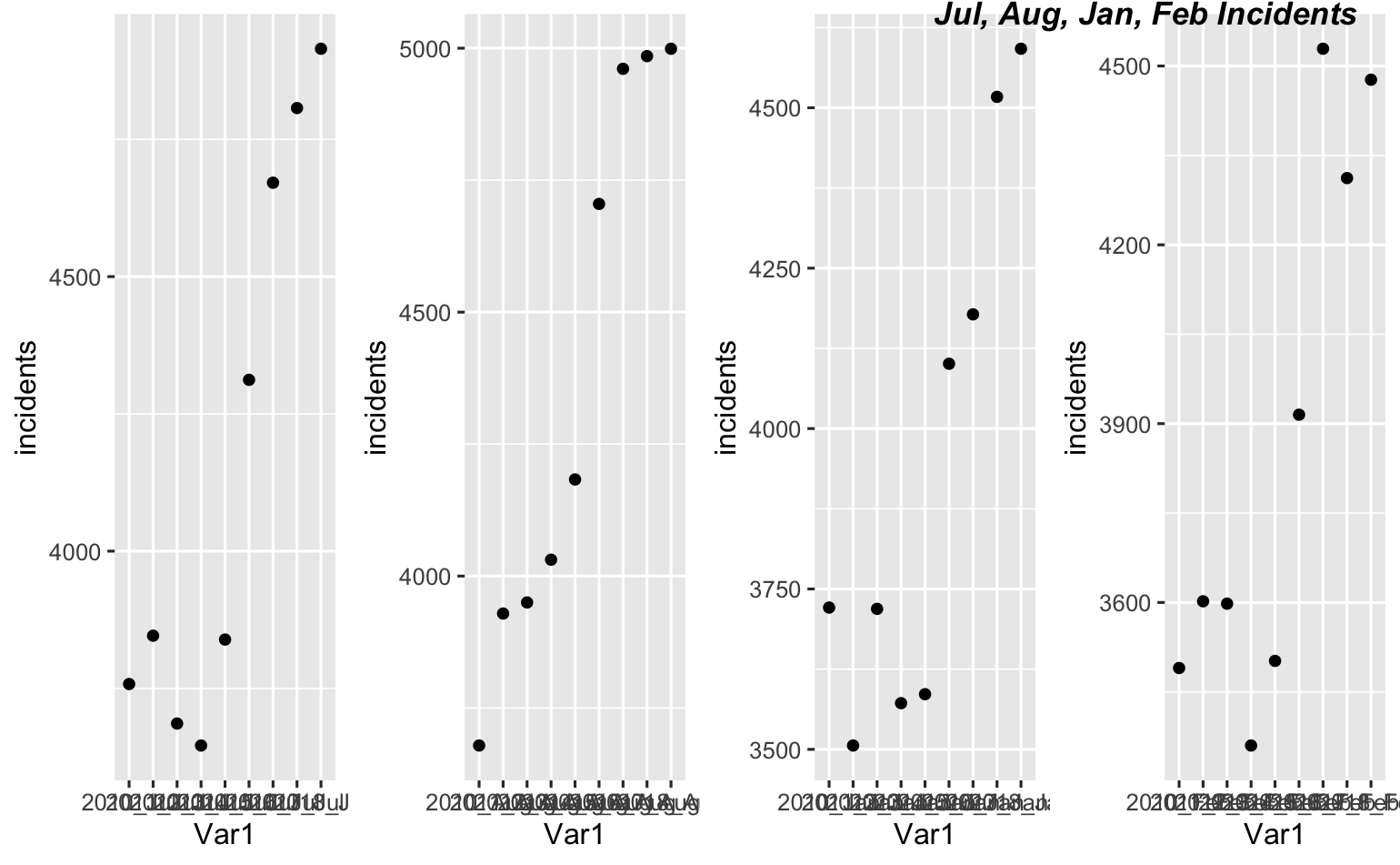
From the two tables above we can see that there is a cut-off in the data set between the years 2010-2014 and 2015-2018. Even more interesting, the 95% in 2011 is higher than the 5% in 2015, showing us a need in the change of the planning for accident prevention as the historical data seems to belong to different distributions.

Plot of Incidents with month and precipitation variables



From looking at this plot we can see that December and October are the months that generally have had higher number of accidents yearly.

July, August, January and February show increased positive trends in the number of accidents in the previous three years as shown:



A multiple-linear regression model for trying to predict the number of accidents given weather vs accidents data was created. The model included the following variables: month, year, precipitation, wind, and temperature. The adjusted r-squared of the model was .82 as follows:

Call:

lm(formula = incidents ~ wind + temperature + precipitation +

month + year, data = yearMonthDF)

Residuals:

Min 1Q Median 3Q Max

-489.99 -143.88 11.34 154.00 378.57

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -3.009e+05 1.870e+04 -16.088 < 2e-16 \*\*\*

wind 2.035e+02 8.947e+01 2.275 0.025237 \*

temperature 9.099e+00 1.032e+01 0.881 0.380454

precipitation 2.882e+01 1.664e+01 1.732 0.086687 .

monthAug 3.152e+02 1.521e+02 2.073 0.040995 \*

monthDec 3.271e+02 1.482e+02 2.207 0.029779 \*

monthFeb -9.989e+01 1.287e+02 -0.776 0.439649

monthJan 3.124e+01 1.506e+02 0.207 0.836082

monthJul 8.426e+01 1.417e+02 0.594 0.553664

monthJun 4.637e+01 1.153e+02 0.402 0.688518

monthMar 2.972e+02 1.086e+02 2.738 0.007421 \*\*

monthMay 8.867e+01 1.012e+02 0.876 0.383085

monthNov 2.402e+02 1.246e+02 1.928 0.056922 .

monthOct 5.418e+02 1.384e+02 3.915 0.000173 \*\*\*

monthSep 1.439e+02 1.529e+02 0.941 0.349046

year 1.509e+02 9.458e+00 15.956 < 2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

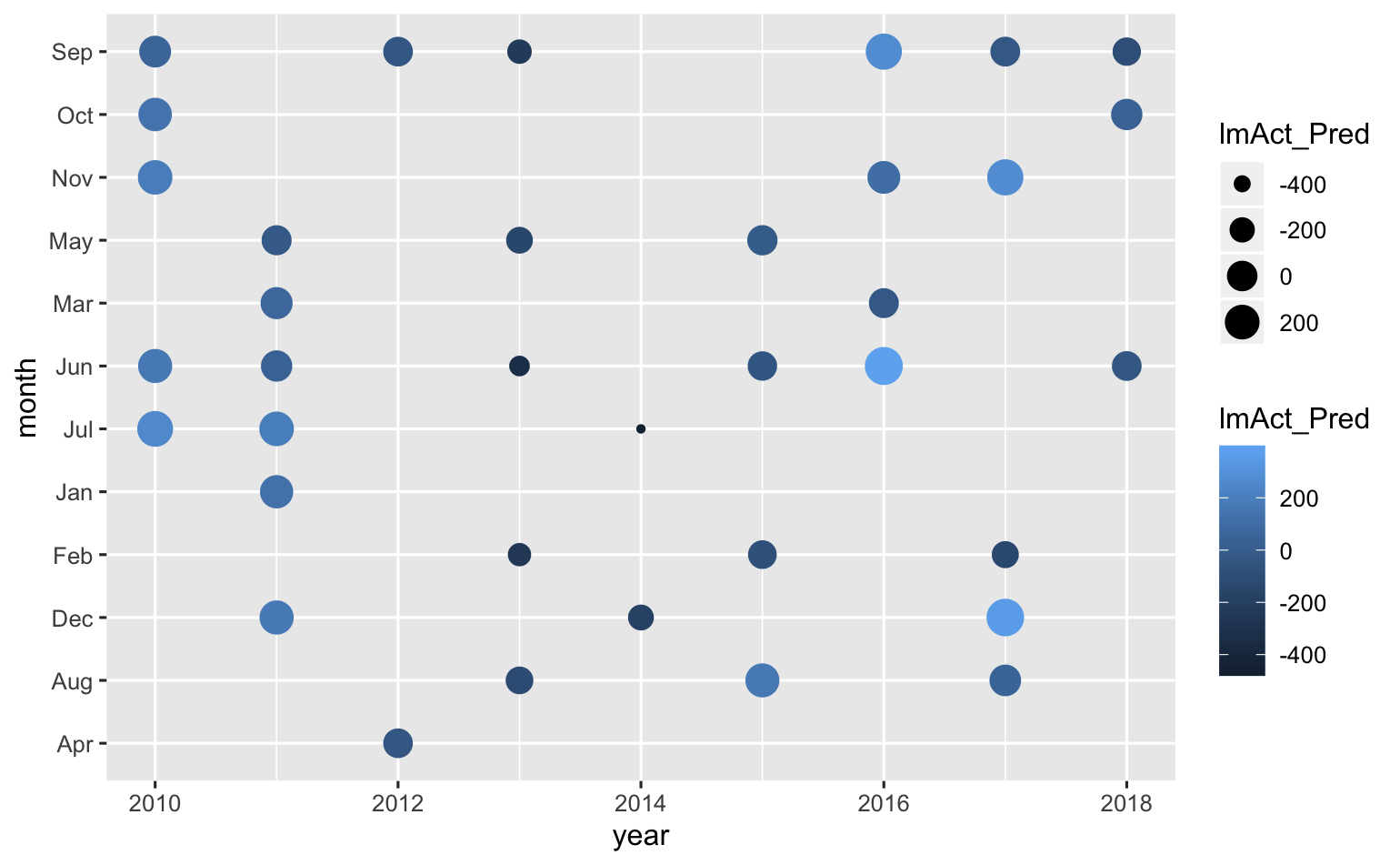
Residual standard error: 209.6 on 92 degrees of freedom

Multiple R-squared: 0.8454, Adjusted R-squared: 0.8202

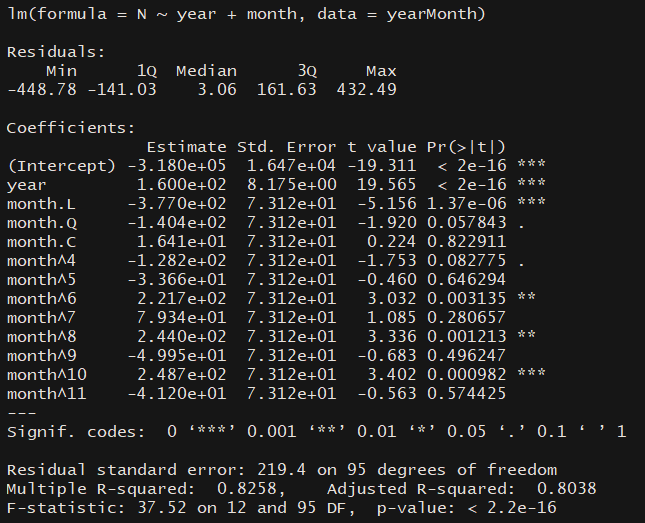
F-statistic: 33.55 on 15 and 92 DF, p-value: < 2.2e-16

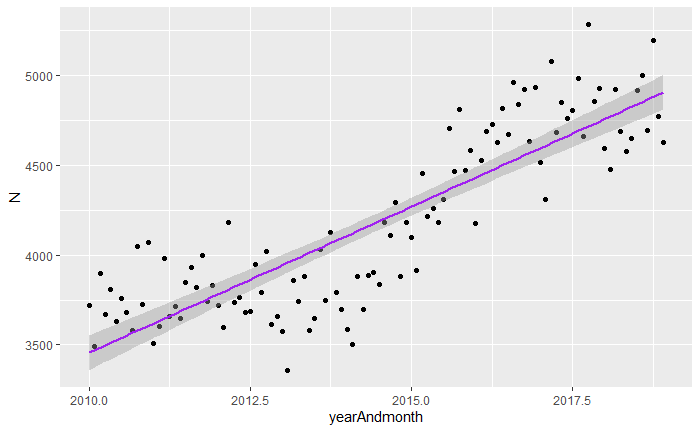
A test and train data-set was created to check the accuracy of the multiple-linear regression model. The rmse of the model was 217.8 and the predicted - actuals ranged between -400 to +200, which one can argue it is an adequate model for planning.

Test data set error plot

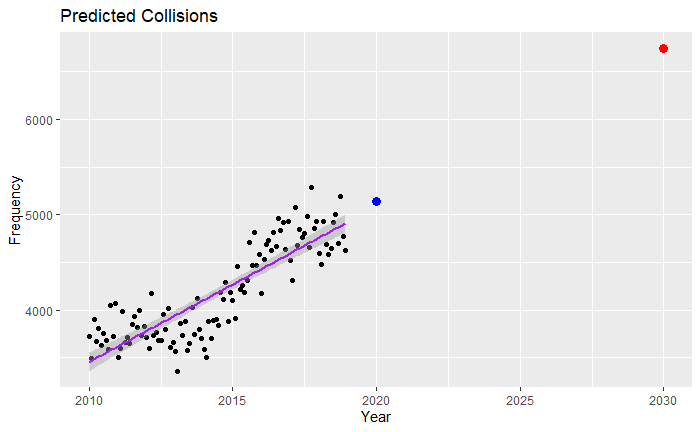


Predicted Collisions: This graph was based on a linear model of year and month and the number of collisions; it also includes a representation of the standard deviation of the model and although many of the instances don’t appear at first blush to be closely related, the model shows a low p-value, adjusted r-squared of just over 80%, and several notations of significance on each variable. This model was used to predict future collision rates, which were plotted over the original graph and show significant, maintained increases in frequency if no interventions are made.



*Linear model output of year, month, and frequency*

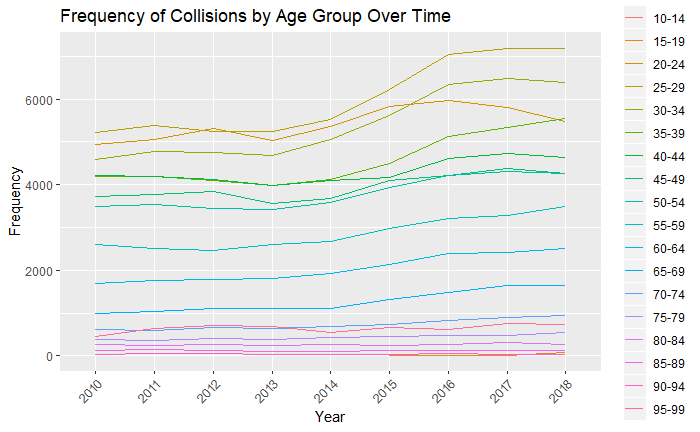
*Linear model graph of year, month, and frequency*



*Linear model graph of year, month, frequency and predicted plots for 2020 and 2030*

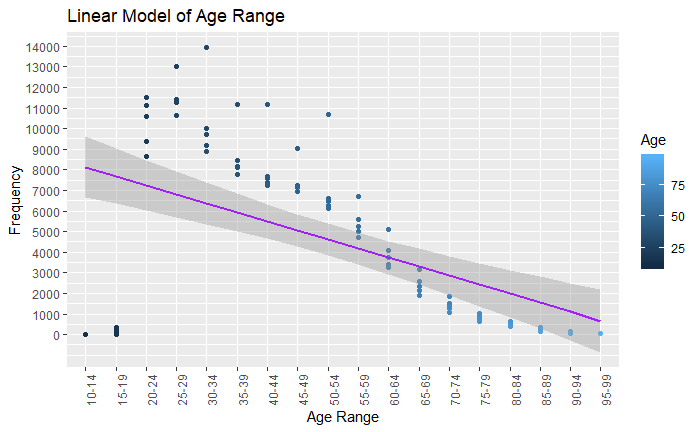
Frequency of Collisions by Age Group: This line graph breaks apart collisions by age group

In order to do this the age information given in the original dataset needed to be converted to ranges; this was done by defining a sequence that defined ranges from 0 to 95 by steps of 5 years at a time. This sequence was applied to each instance to create a new column of only age groups. This allowed for the creation of a line graph that plots collision frequency by year, with each line (further defined by color) representing an individual group. This graph shows that none of the age groups saw a decrease in collisions, some stayed fairly flat, but most saw a marked increase over the 9 years graphed.

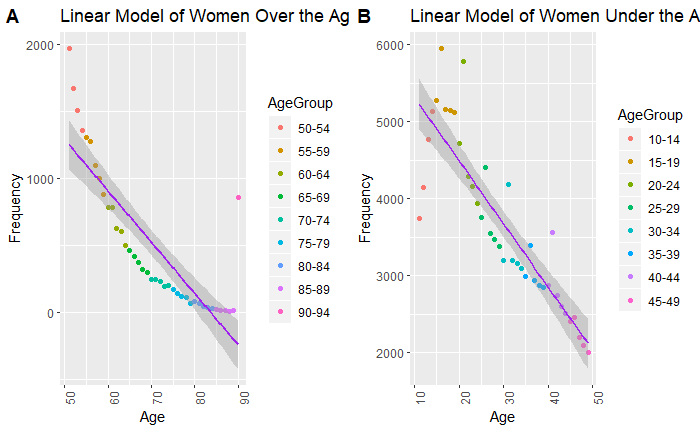


*Line graph of collision frequency over time, separated by age group*

Linear Models of Age: One attempt at correlating and predicting models through linear modeling looked at the impact of a victim’s age and the date the collision occurred on the frequency of collisions. Although several versions were attempted - including limiting the dataset by gender and age range - no clear correlation could be found. Although the linear models reported each variable include as significant, and had low p-values, correlation (viewed as adjusted r-squared) remained at 70% or lower, and displayed large swaths of outliers. Although this can be seen as demonstrating some links, we were looking for stronger relationships than demonstrated here.



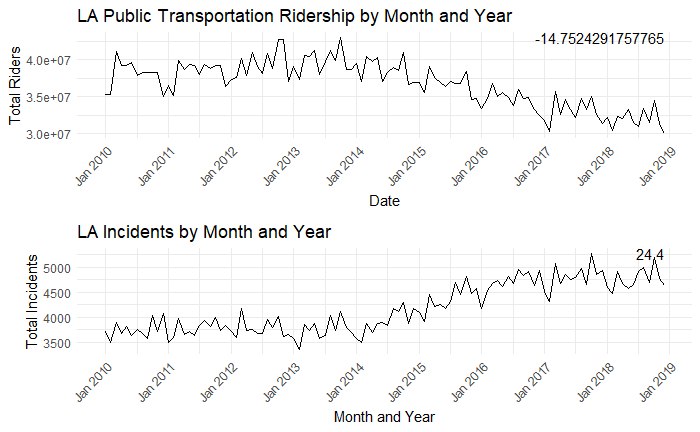
*Linear model of age ranges and collision frequency*

**

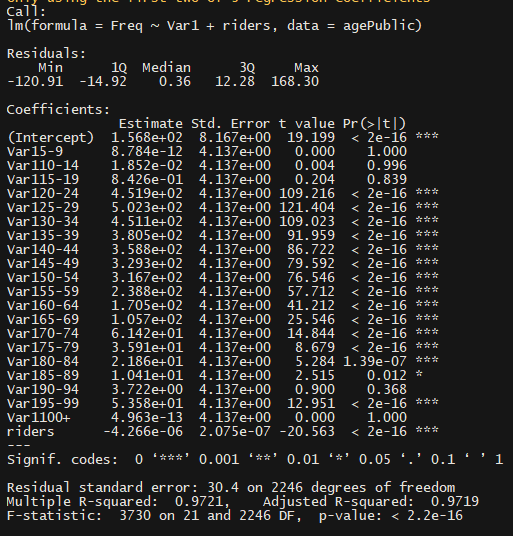
*Comparing linear models of women over and under 50 and collision frequency*

Linear Models of Age and Public Transportation Information: The most successful attempt at correlating variables with collision frequency came when public transportation ridership was pulled in. Over the past decade the city has focused on expanding the rail system and creating faster options that can move even more riders. However, despite the increased options, public transportation ridership continues to decrease; one report proposes that this has to do with the decrease in bus frequency (which was done in an effort to encourage rail ridership), and the higher rail prices -- the combination of the two has forced people back to driving themselves instead.

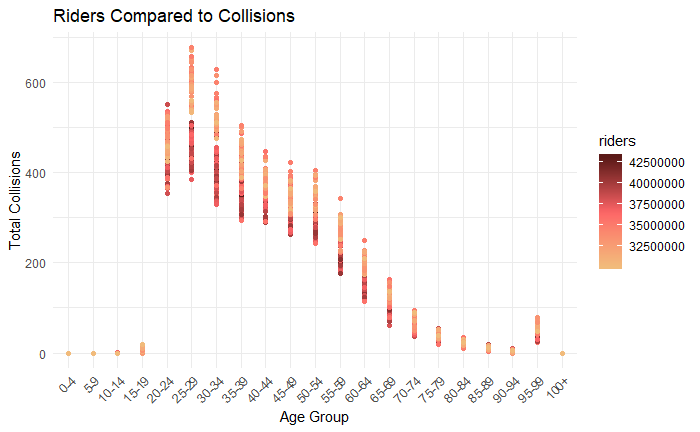
We can see this correlation in two line graphs plotted side by side, as below; public transportation ridership has decreased by nearly 15% over the past decade, whereas collisions have increased by over 24%. Although we can make a logical supposition that these two are related, we can confirm this hypothesis when reviewing the linear model, which has an adjusted r-squared of more than 97%, shows the significance of several ages and ridership variables, and a low p-value. In the final graph we again see how higher collisions are correlated to both age range and and ridership numbers.



*Comparing plots of total public ridership over time and collision frequency over time; the number in the top right corner represents change over time*

**

*Linear model of collision frequency, ridership and age range*

**

*Collision ranges by age group*

# Conclusions

Traffic collisions in Los Angeles have increased significantly between the years of 2010-2018. This report provides a comprehensive look at the data available including the LA DOT, LA weather station, and LA ridership datasets to provide insights to the LA city council on how they should reallocate their budget, and target the appropriate initiatives and programs to reduce or control the traffic accidents in the city. The team believes that the city council should focus on two big areas of impact, Public Transportation and Additional Monitoring.

This analysis provided the demographics the city council should focus their efforts on, it provided a time-series model that the LA city council can use to identify their most severe collision streets yearly to create programs that can address these, and several multiple-linear regression models that can help with the planning and forecasting of programs for accident prevention in the city of Los Angeles.

The objective of this report was to provide the city council insights on the where they should refocus their budget and efforts to make the City of Los Angeles a safer city, the two biggest areas of impact addressing this should be Public Transportation and Additional Monitoring, and the city council now has the tools and results to move their initiatives forward.

# Looking Forward

This analysis can be further improved by including population total data by month or year, including more weather stations since LA covers a huge geographical area, and having this team of data-scientist do this analysis every few years to make sure that the city council can plan appropriately when disruptive changes happen as it did in the year 2015.

Full analysis of the data

*A good analysis is repetitive. You know the intricacies of your work in and out, but your audience does not. You’ve told your readers in your abstract (or introduction, if you prefer) what you had ventured to do and even what you end up finding and the content lays this all out for them. In the conclusions section, you hit them with it again. At this point, they’ve seen the relevant data you’ve carefully chosen to support your theory so it’s time to formally draw your conclusions. Your readers can decide if they agree or not.*

*Speaking of being repetitive, after making your conclusions, you again remind your readers of the objective(s) of this report. Restate them again and help your readers help you―what do you expect now? What feedback would you like? What decision-making can happen now that* your report is presented and the insights have been shared? In my work, I often collaborate with strategists to develop a set of recommendations for our clients. Typically I'll take a stab at it based on the expertise I've gained in working with the data and a strategist will refine using their business insights.

* Restate the questions from your introduction.
* Restate important results.
* Include any recommendations for additional data as needed.

# References

Census Reporter. (2017). Los Angeles, CA [Data file]. Retrieved from: <https://censusreporter.org/profiles/16000US0644000-los-angeles-ca/>

Chiland, E. (2019, February 13). LA is the fifth worst U.S. city for traffic, report says. Retrieved from <https://la.curbed.com/2019/2/13/18222225/los-angeles-traffic-worst-nation-hours>

City of Los Angeles. (2019). Los Angeles Traffic Collision Data [Data file]. Retrieved from: <https://www.kaggle.com/cityofLA/los-angeles-traffic-collision-data>

Metro. (2019). Interactive Estimated Ridership Stats [Data file]. Retrieved from: <http://isotp.metro.net/MetroRidership/YearOverYear.aspx>

Nelson, L., & Smith, D. (2017, April 3). The number of pedestrians, cyclists and drivers killed in L.A. traffic rose sharply in 2016. Retrieved from <https://www.latimes.com/local/lanow/la-me-ln-2016-traffic-deaths-20170403-story.html>

Panagiotopoulos, V. (2017). The Los Angeles metro is great – so why aren't people using it? Retrieved from: <https://www.citymetric.com/transport/los-angeles-metro-great-so-why-aren-t-people-using-it-2742>

NOAA. Los Angeles Downtown Weather Station GHCND:USW00093134

(2019-08-04), OrderID #1827343. <https://www.ncdc.noaa.gov/cdo-web/orders?email=sevillar@syr.edu&id=1827343>

# Appendix

Please see the attached link to the R-scripts for the team: