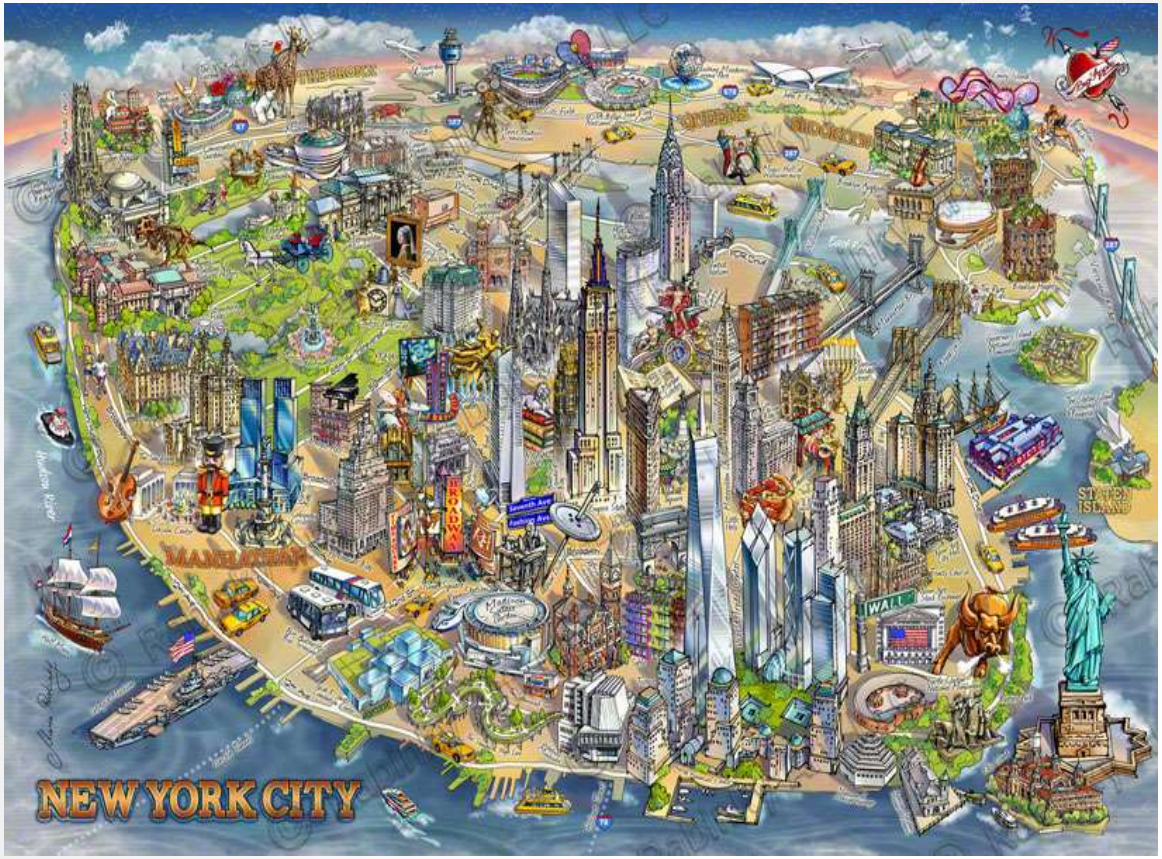
CAR ACCIDENTS IN NEW YORK CITY

Springboard Foundations of
Data Science Capstone Project
By: Katie Wilson

TABLE OF CONTENTS

INTRODUCTION

New York City – the largest city in the United States. With the constantly increasing population in the city that never sleeps, car accidents happen all over the city at all times of the day and night.
Data analysis of any information would make it much easier for the police and the residents to help preemptively address the cause of these accidents.
The NYPD keeps an up-to-date, publicly available dataset online with much of this information.



Illustrated maps by Rabinky Art: http://www.illustratedmaps.com/IllustratedMaps/Z-NewYorkCity-IllustratedMap.html

DATA:
TABLE BREAKDOWNS

- The NYPD dataset has 29 variables providing a wide potential for analysis
- Showing the number of incidents by borough, top 10 types of vehicles involved, and top 10 contributing factors in each incident included in the dataset from the NYPD database

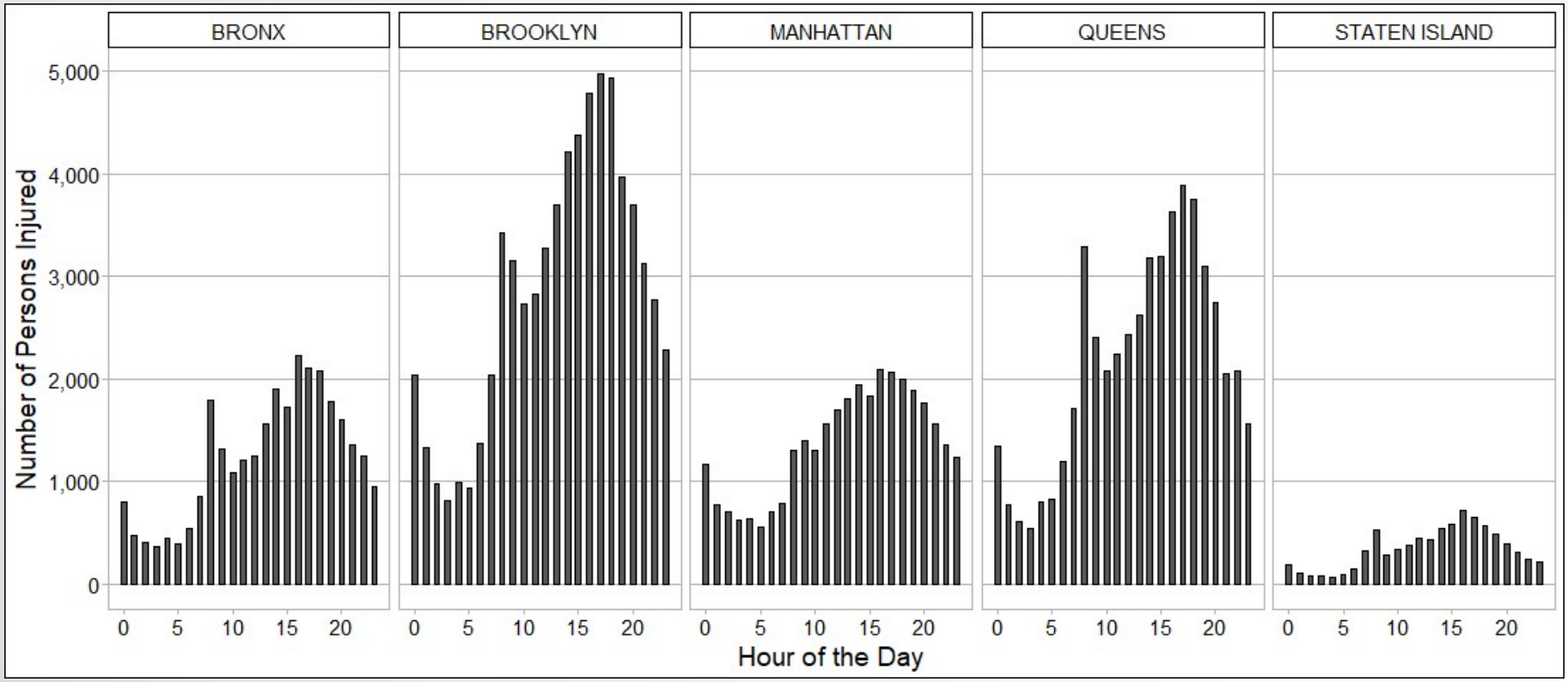
Borough	ValueCount
BROOKLYN	237,929
QUEENS	202,733
MANHATTAN	193,589
BRONX	103,583
STATEN ISLAND	34,185

ContributingFactor	ValueCount
driver inattention/distraction	163,339
failure to yield right-of-way	52,512
fatigued/drowsy	43,472
backing unsafely	35,044
following too closely	30,438
turning improperly	26,862
lost consciousness	18,086
traffic control disregarded	14,922
passing or lane usage improper	14,313
driver inexperience	14,241

VehicleType	ValueCount
passenger vehicle	619,916
sport utility / station wagon	265,100
taxi	42,074
van	23,637
pick-up truck	19,954
small com veh(4 tires)	13,222
large com veh(6 or more tires)	12,464
bus	12,369
livery vehicle	8,419
motorcycle	6,029

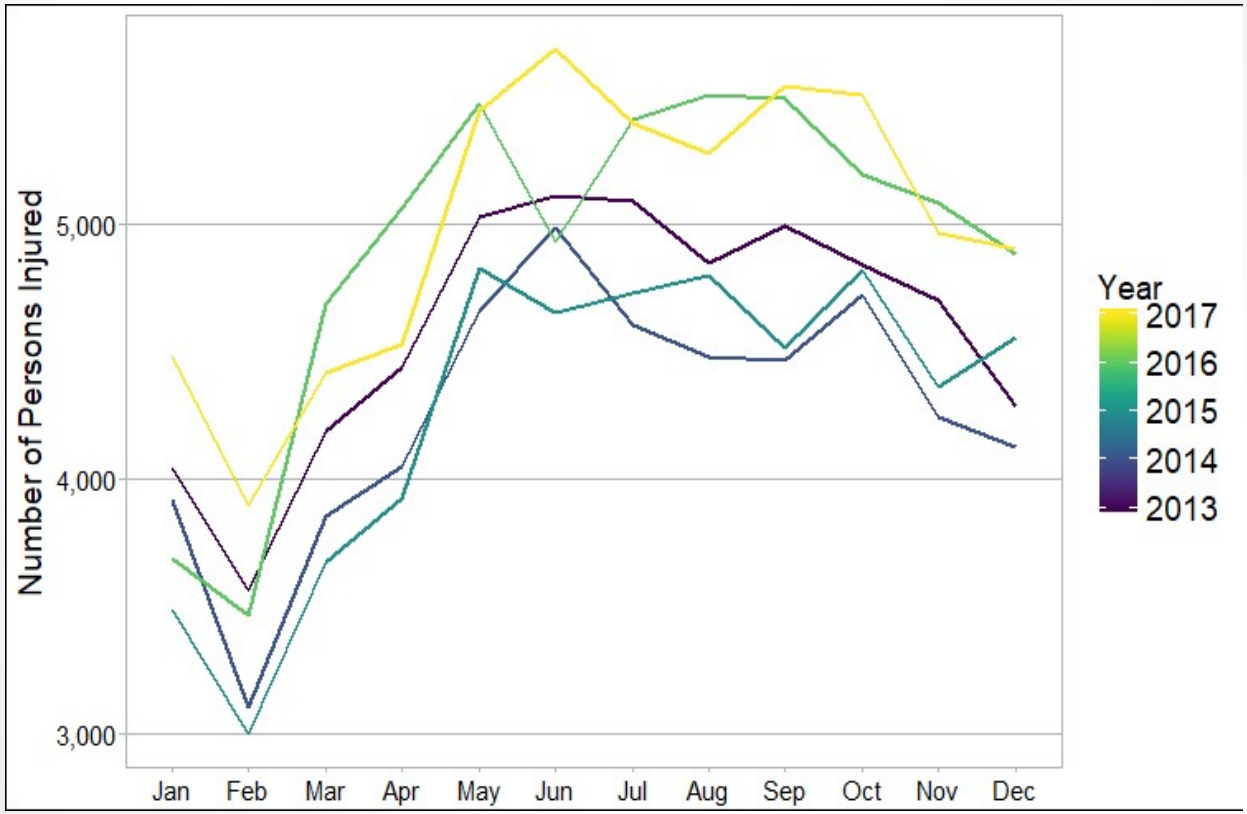
DATA:
INJURIES BY BOROUGH

The NYPD dataset includes the time of each incident. The graph below shows the number of injuries recorded by the hour of the day in which they occurred in each borough.



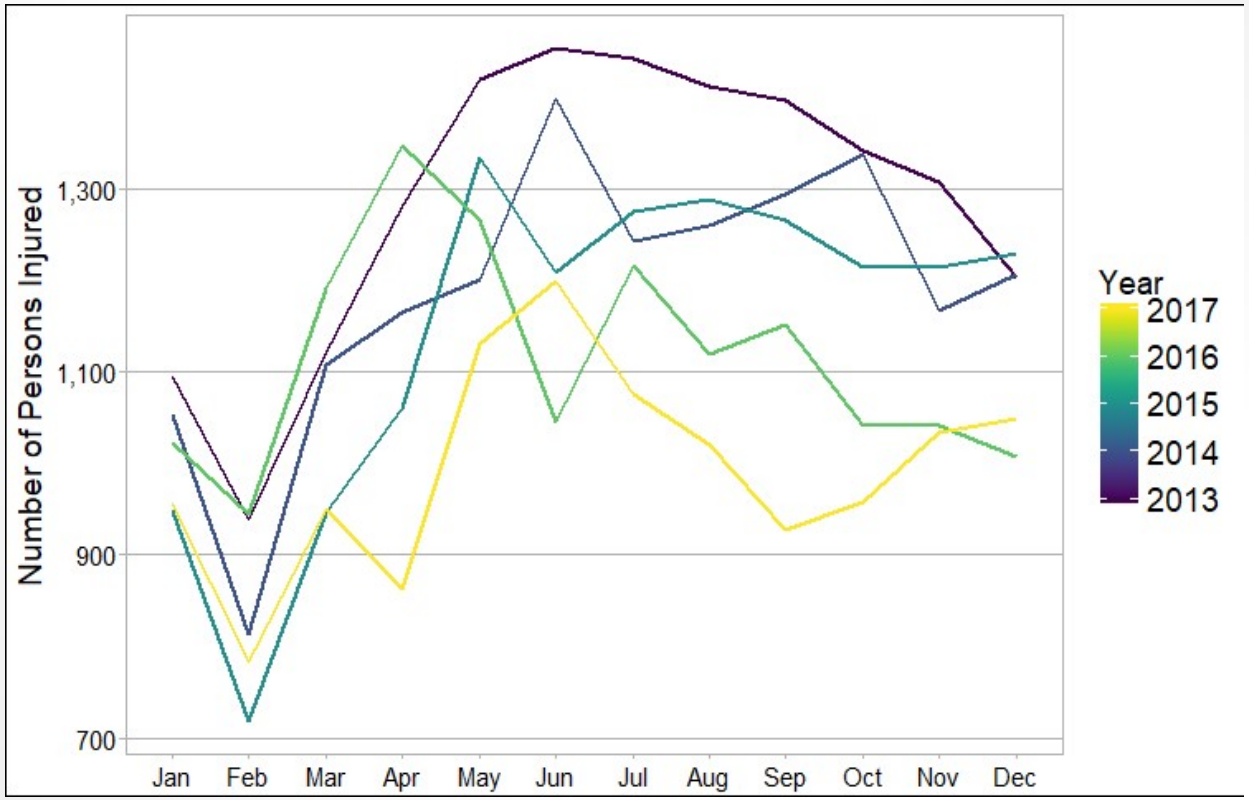
EXPLORATORY DATA ANALYSIS:
INJURIES BY YEAR

- Breaking the injuries down by year, there are clear high points and low points.
- Many factors could contribute to this including tourist season and weather.
- Further exploration of this by borough, contributing factor, vehicle type, and time of day may help understand the distribution.

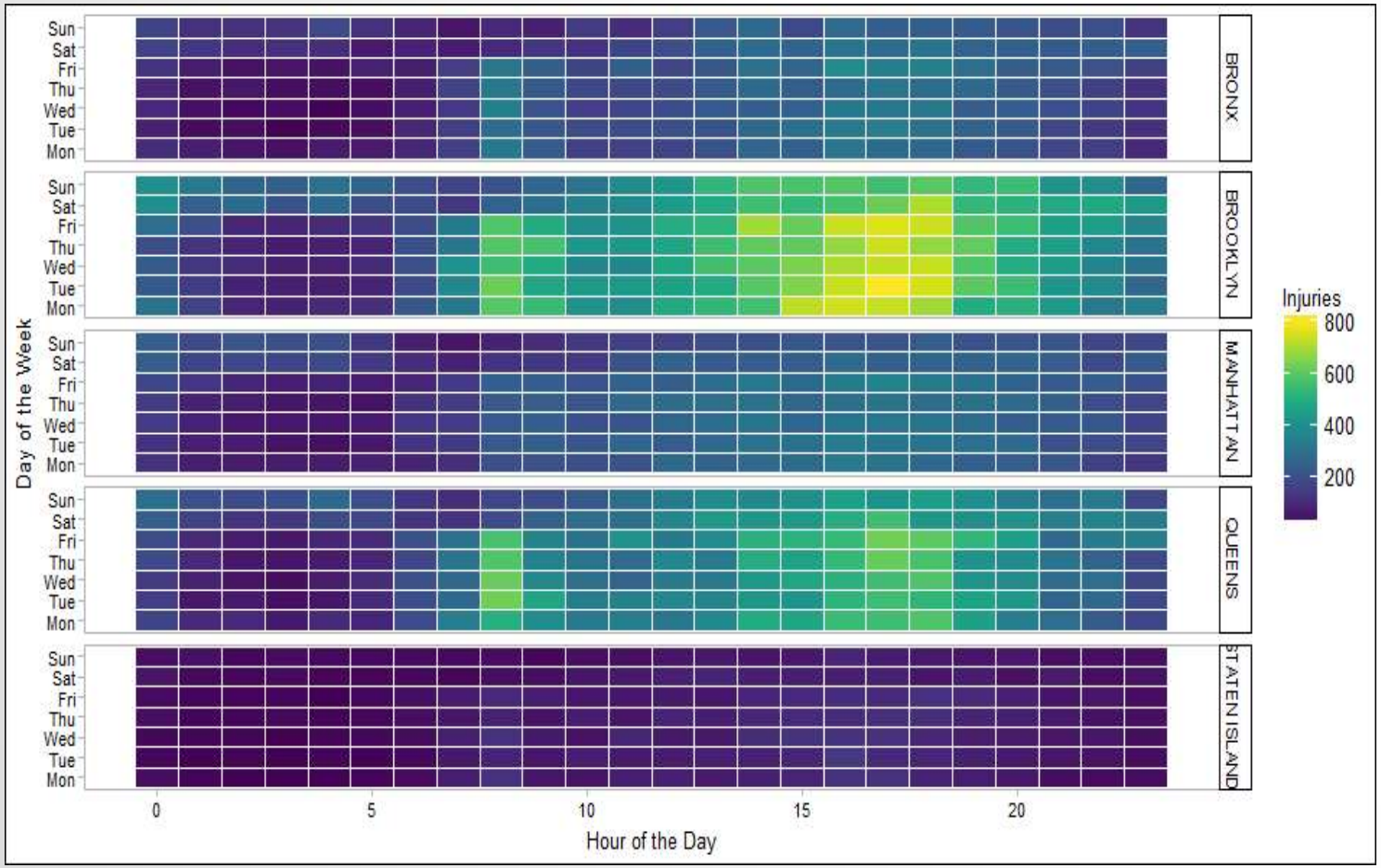


EXPLORATORY DATA ANALYSIS:
BROOKLYN INJURIES BY YEAR

- Similar to the previous page, this breakdown by year is limited to only Brooklyn as the borough with the highest number of recorded incidents.
- While the trend in the full dataset increases from 2013 to 2017, Brooklyn's data decreases within those years.

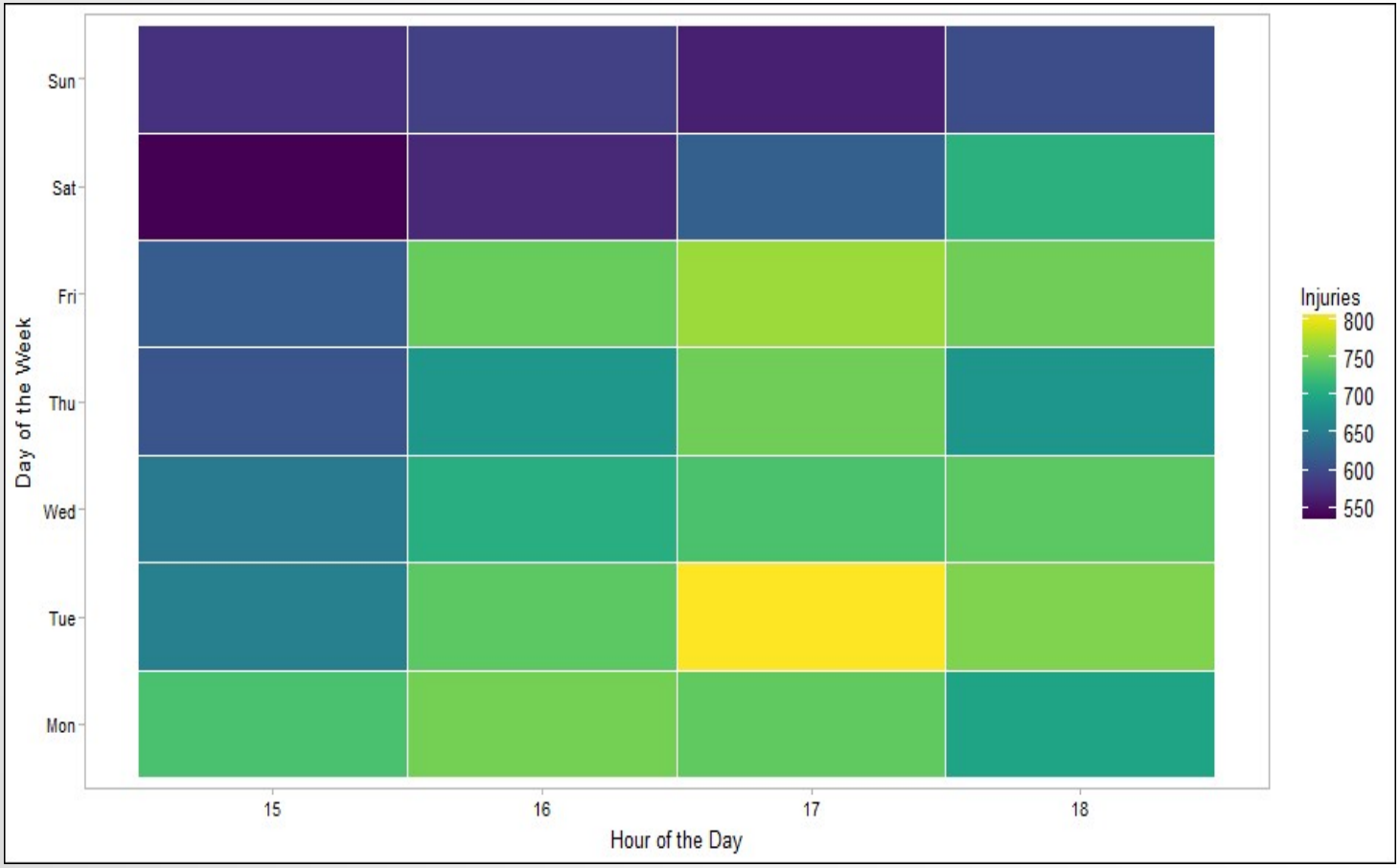


ANALYSIS:
HEAT MAP



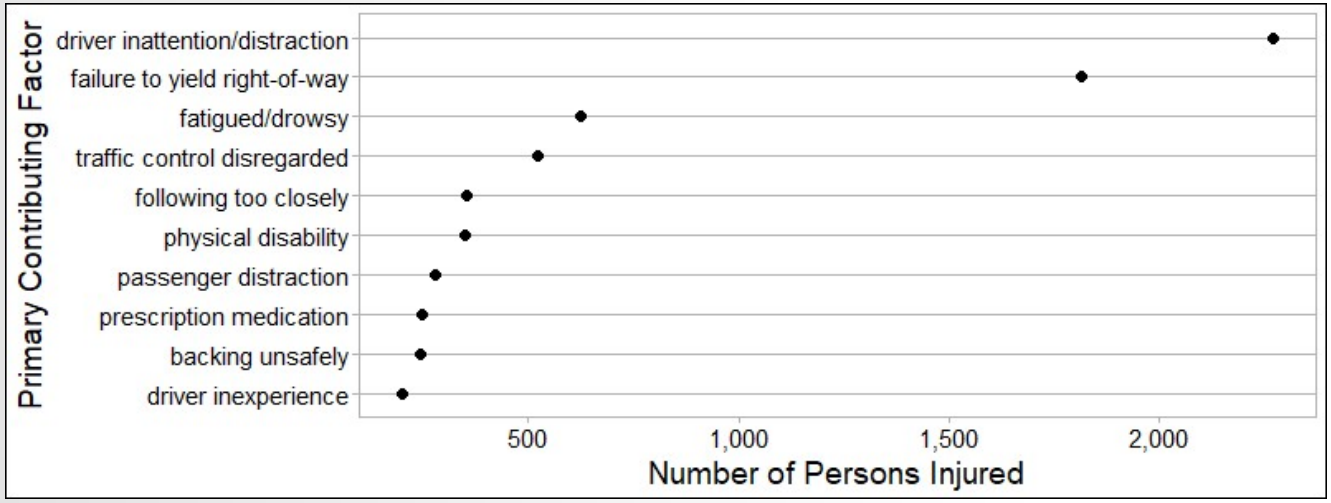
- A breakdown of the injuries over the full dataset again displays the vast amount of accidents that take place in Brooklyn specifically during evening rush hour.
- The rest of the analyses will be done on this specific subset of the data in order to identify the most influential factors.

ANALYSIS:
BROOKLYN HEAT MAP



- Similar to the graph on the previous page, this is a breakdown of that four hour period in Brooklyn where the majority of the injuries had occurred.
- Clearly, the time right after the work day ends from Monday through Friday are the most dangerous times in this borough.
- Now that this has been established, what are the highest contributing factors to this trend?

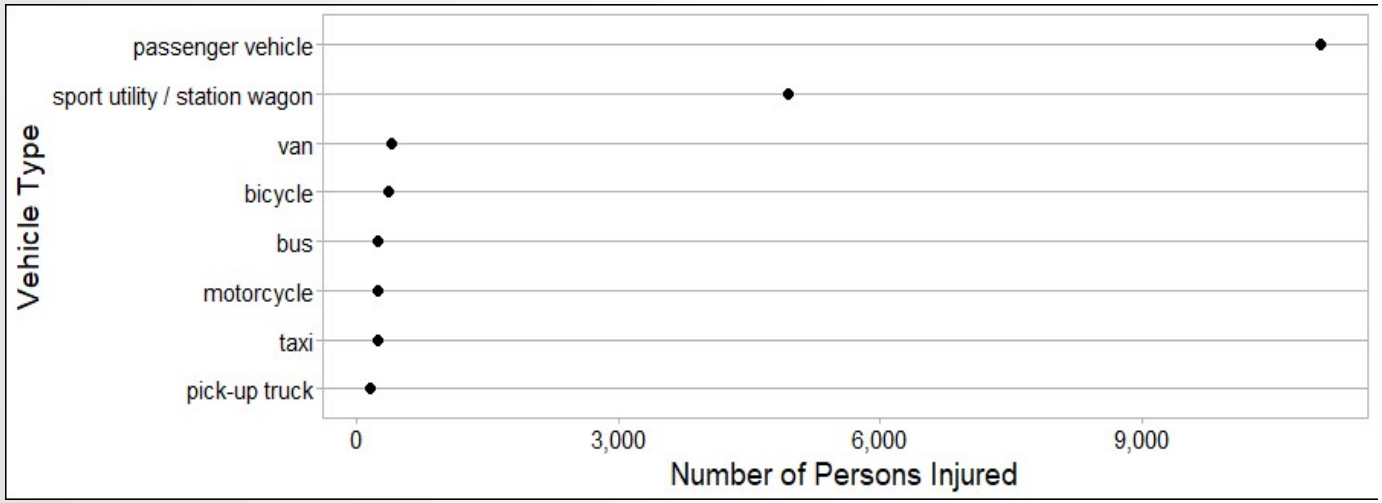
RESULTS:
CONTRIBUTING FACTORS - BROOKLYN



Contributing Factor	Injuries
driver inattention/distraction	2,268
failure to yield right-of-way	1,814
fatigued/drowsy	624
traffic control disregarded	525
following too closely	354
physical disability	351
passenger distraction	279
prescription medication	249
backing unsafely	245
driver inexperience	202

- Using that subset of data from rush hour in Brooklyn, the top contributing factors above are laid out in a table and a graph.
- Clearly the "driver inattention/distraction" category is the largest as is the case with the full dataset. However, unlike the full dataset, the gap between the first and second factors is much smaller in Brooklyn than in the city of New York as a whole.
- What types of vehicles are involved in these injuries most often?

RESULTS:
VEHICLE TYPE - BROOKLYN



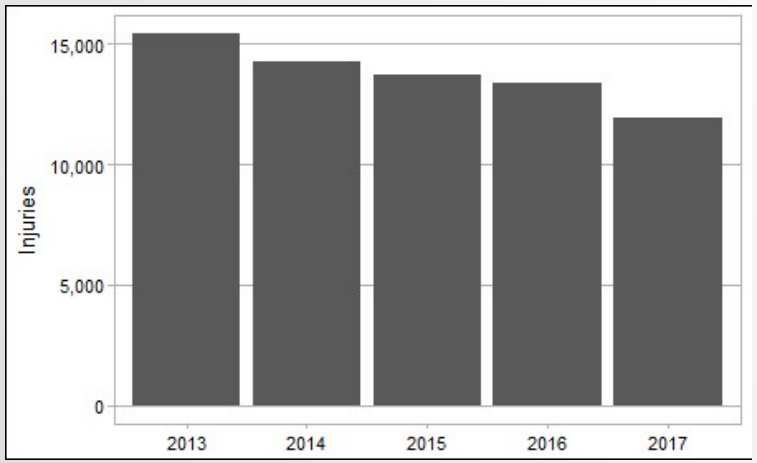
Vehicle Type	Injuries
passenger vehicle	11,035
sport utility / station wagon	4,954
van	411
bicycle	370
bus	249
motorcycle	244
taxi	243
pick-up truck	160

- The breakdown above shows the vehicle types most involved in the injuries during rush hour in Brooklyn.
- Using the data from the last two pages in conjunction, it seems that people driving home from work in their passenger vehicles cause accidents while texting and driving.

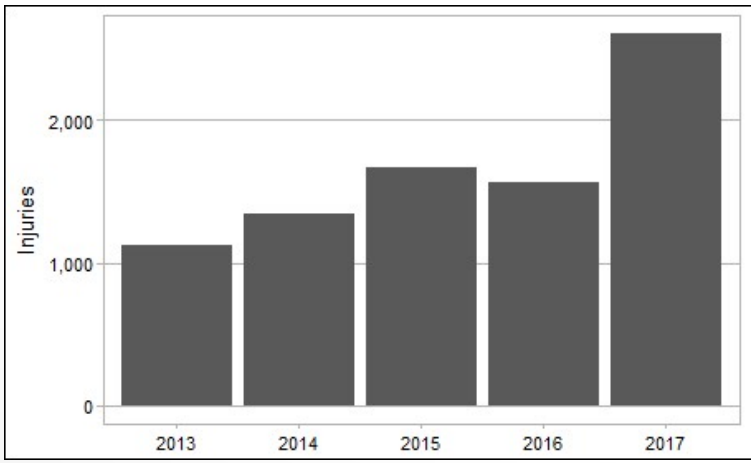
CONCLUSION

- Because this data comes from the NYPD, a recommendation could be made for the police to enforce laws surrounding distracted driving more heavily in Brooklyn.
- According to the Governor's Traffic Safety Committee, a law was implemented in the state of New York to prohibit distracted driving involving cell phones by fining the offender and resulting in a mandatory 120-day driver license or permit suspension.*
- While the total number of accidents has decreased in Brooklyn over the years, the number of accidents involving distracted driving has increased even since that law was implemented in 2014.

Total Accidents in Brooklyn



Driver Inattention/Distraction Accidents in Brooklyn



*Stats and information from the New York state Governor's Traffic Safety Committee website: http://www.safeny.ny.gov/phon-ndx.htm

POTENTIAL FUTURE WORK

Many factors influence the driving conditions in any city. The elements discussed in this report only cover a few of the topics. Other categories to be considered for future work include:

- Precipitation
- Population density
- Location of tourist attractions
- Implementation of Uber/Lyft
- Number of taxis in use in the city
- Speed limits

Any and all of the above list would contribute greatly to the depth of this work. With a goal to create an app available to all New York residents, the police department, and visitors to the city consisting of current and constantly updating data, all of these factors would need to be considered and included.