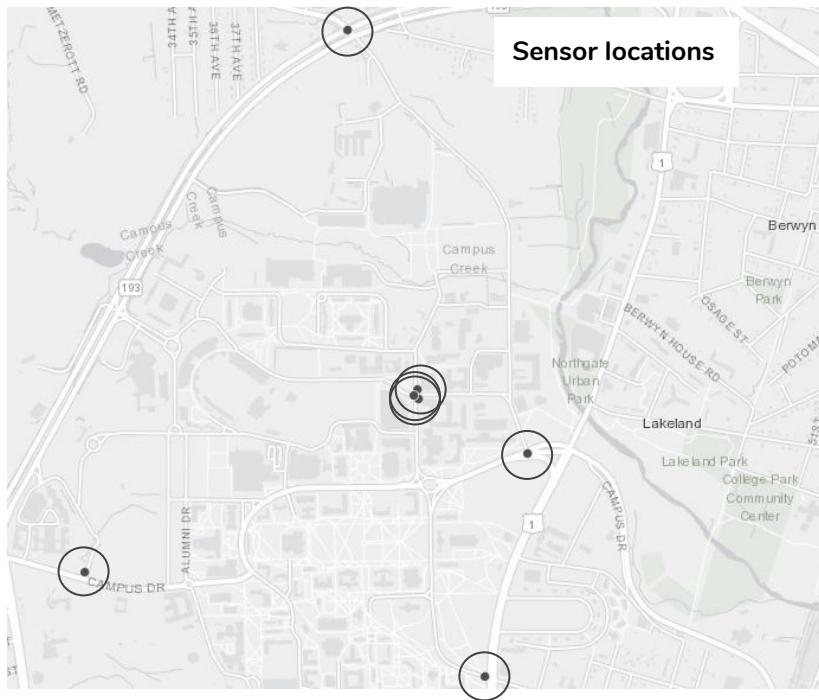


# Dataset: UMD Traffic Count Sensor Data

Team: DC20072



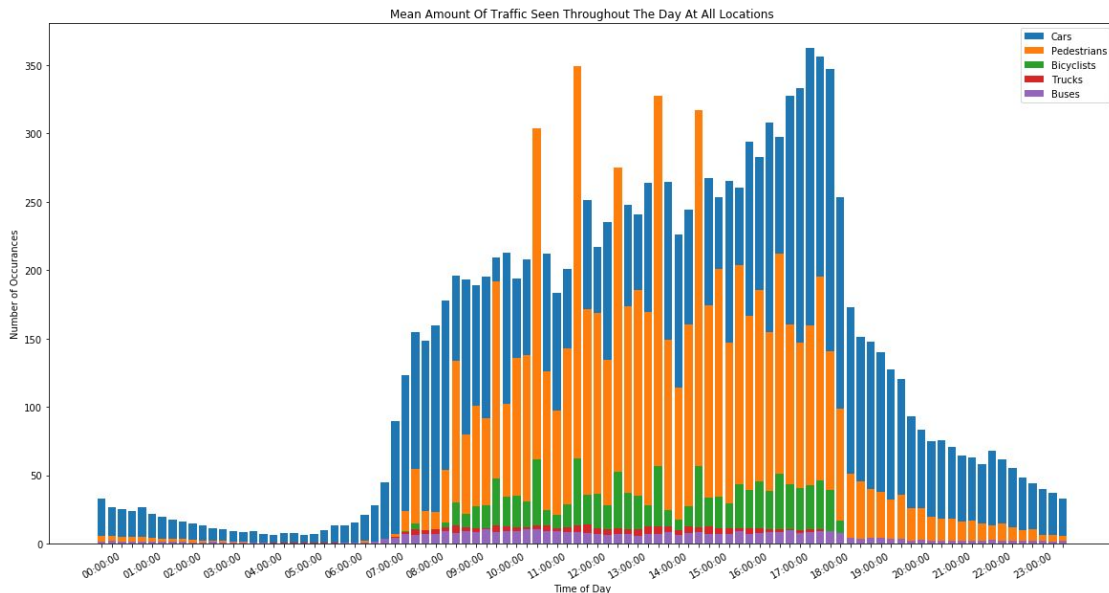
# What Does The Data Look Like?



- Features:
  - Car count (Integer)
  - Truck count (Integer)
  - Bus count (Integer)
  - Bicyclist count (Integer)
  - Pedestrian count (Integer)
  - Timestamp (Date/Time)
  - Sensor location (String)
- How much?
  - 2 weeks worth of data captured at 15 minute intervals
  - 13000 rows



# What preparation steps did we take?

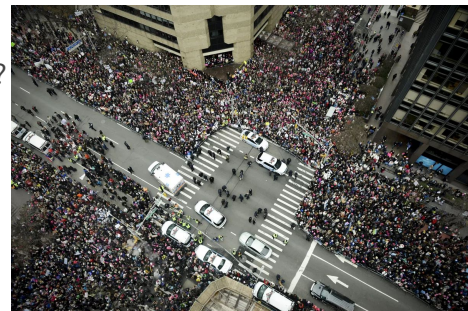


- Gathered weather data for the time that the data was gathered
- Gathered crime data for the areas and time period that matches with the provided dataset
- Created a new learning dataset that separated the variance of time into block of 6 hours each.
- Applied latitude, longitude for sensor locations to allow for arcgis plotting

# Questions to answer









- What is the relationship between weather and traffic?
  - Does weather decrease one mode of transportation and increase another?
- What is the relationship between sensor location and favored mode of transportation?
  - What are the traffic patterns at each location?
- What time of day has the most traffic? Is this different when taking mode of transportation into account? How about for day of the week?
  - How do major events affect these numbers? Ex: Football games
- Are there any correlations between the modes of transportation?
  - Ex: As amount of cars goes down does that also decrease pedestrians?
- How does crime affect the traffic flow?
  - Which crimes typically occur when more people are around? Less people?
  - What time of day do specific crimes typically occur?
- Can we predict future traffic?
  - Can we create some kind of model to predict traffic in the future?












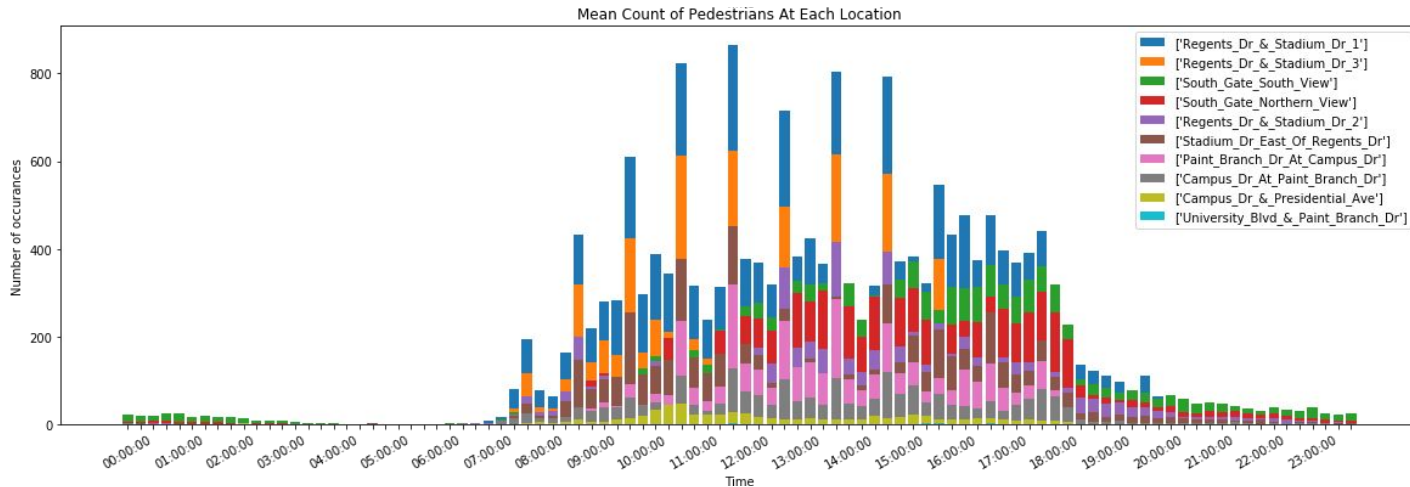
# What is the relationship between weather and traffic?

- Not enough variation in weather to see how rain or very cold temperatures affect changes in traffic

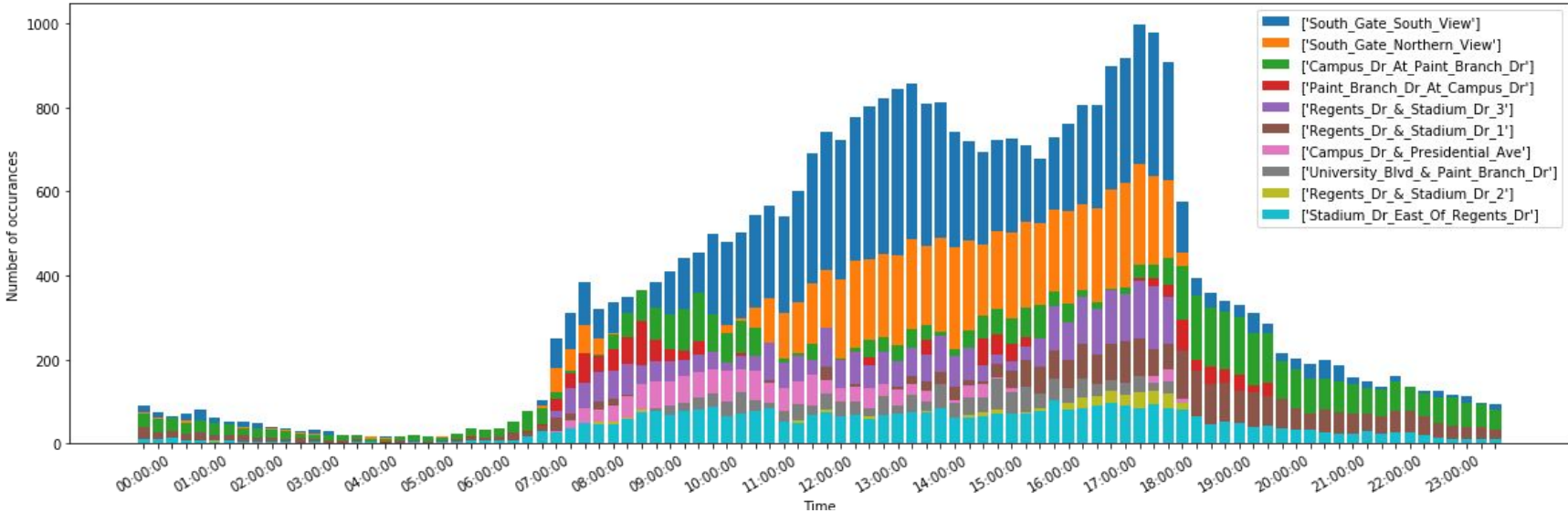
28	29	30	31	1	2
					
Mostly Sunny	Cloudy	Cloudy	Cloudy	Mostly Sunny	Mostly Sunny
Actual: 69°   56°	Actual: 68°   59°	Actual: 67°   60°	Actual: 77°   55°	Actual: 54°   44°	Actual: 55°   38°
1.06 in	0 in	0 in	0.51 in	1 in	0 in

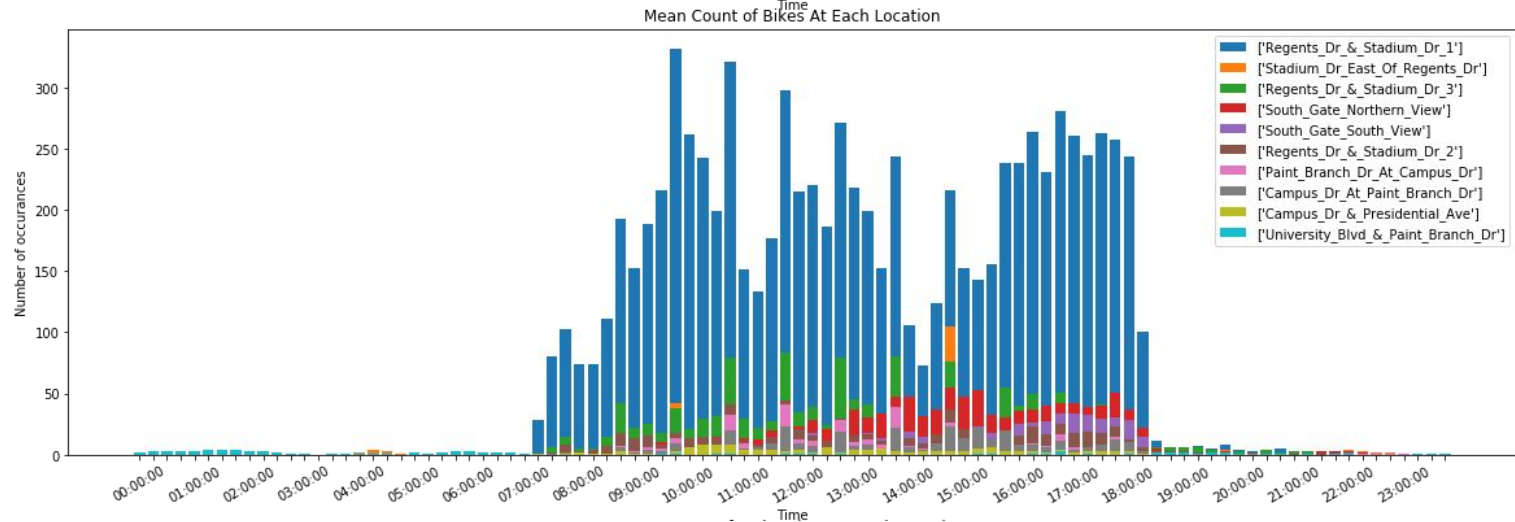
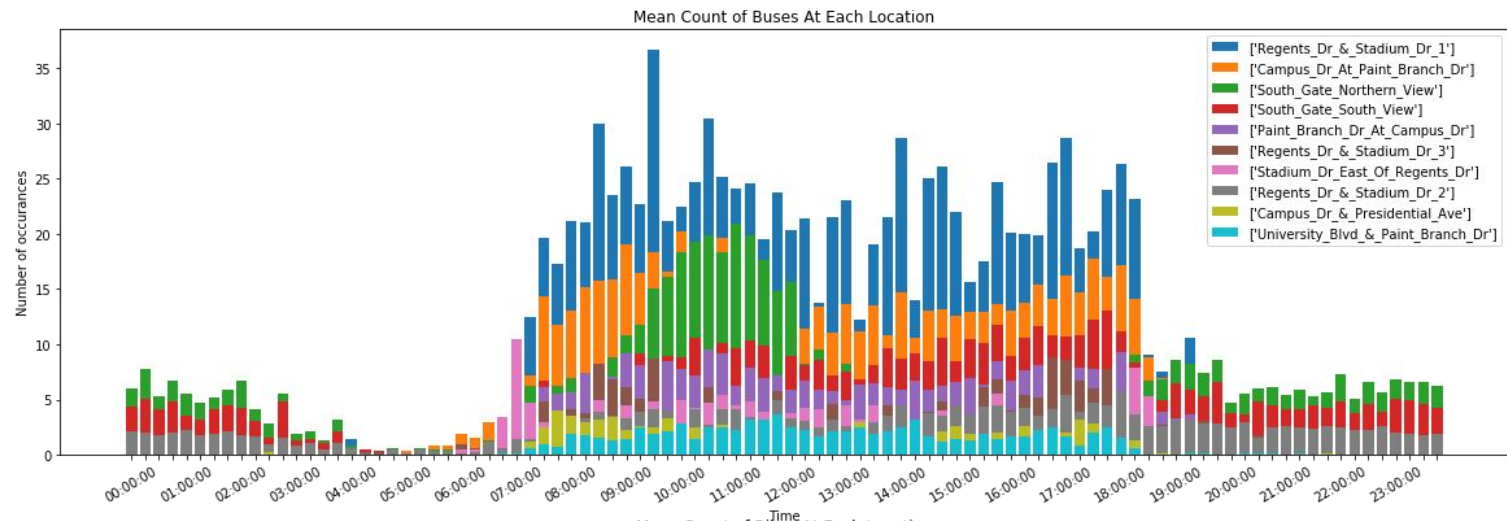
3	4	5	6	7	8	9
						
Mostly Sunny	Mostly Sunny	Mostly Cloudy	Mostly Sunny	Cloudy	Mostly Cloudy	Mostly Cloudy
Actual: 57°   43°	Actual: 60°   39°	Actual: 63°   45°	Actual: 59°   45°	Actual: 61°   46°	Actual: 45°   35°	Actual: 45°   31°
0 in	0 in	0 in	0 in	0 in	0.06 in	0 in

# What is the relationship between location and mode of transportation?

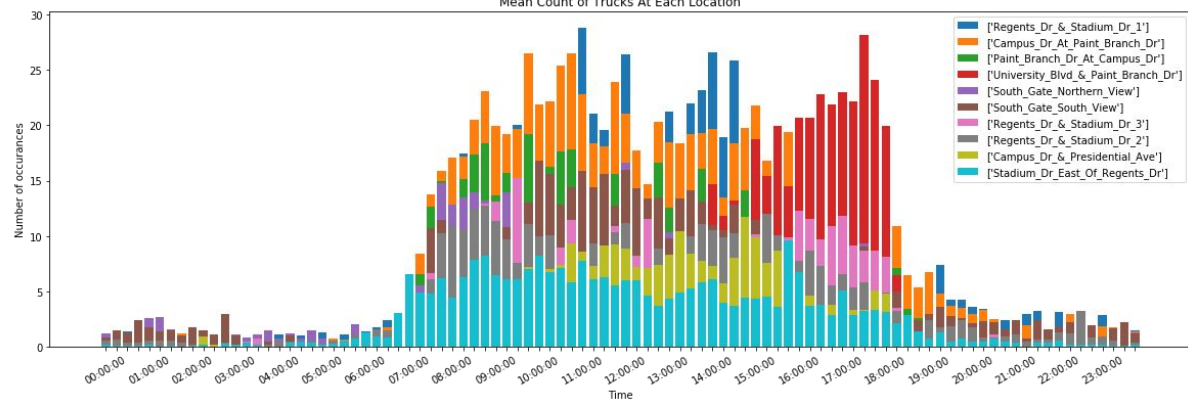
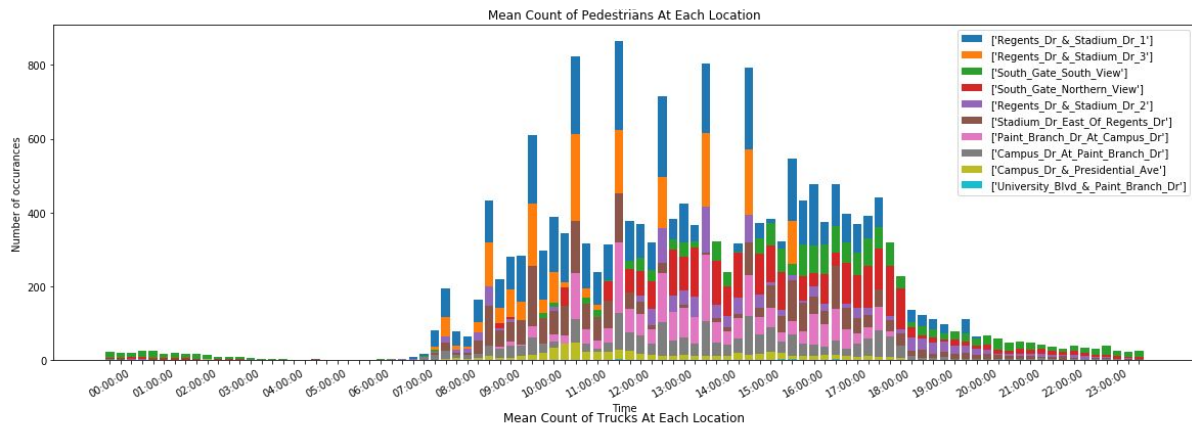


### Mean Count of Cars At Each Location

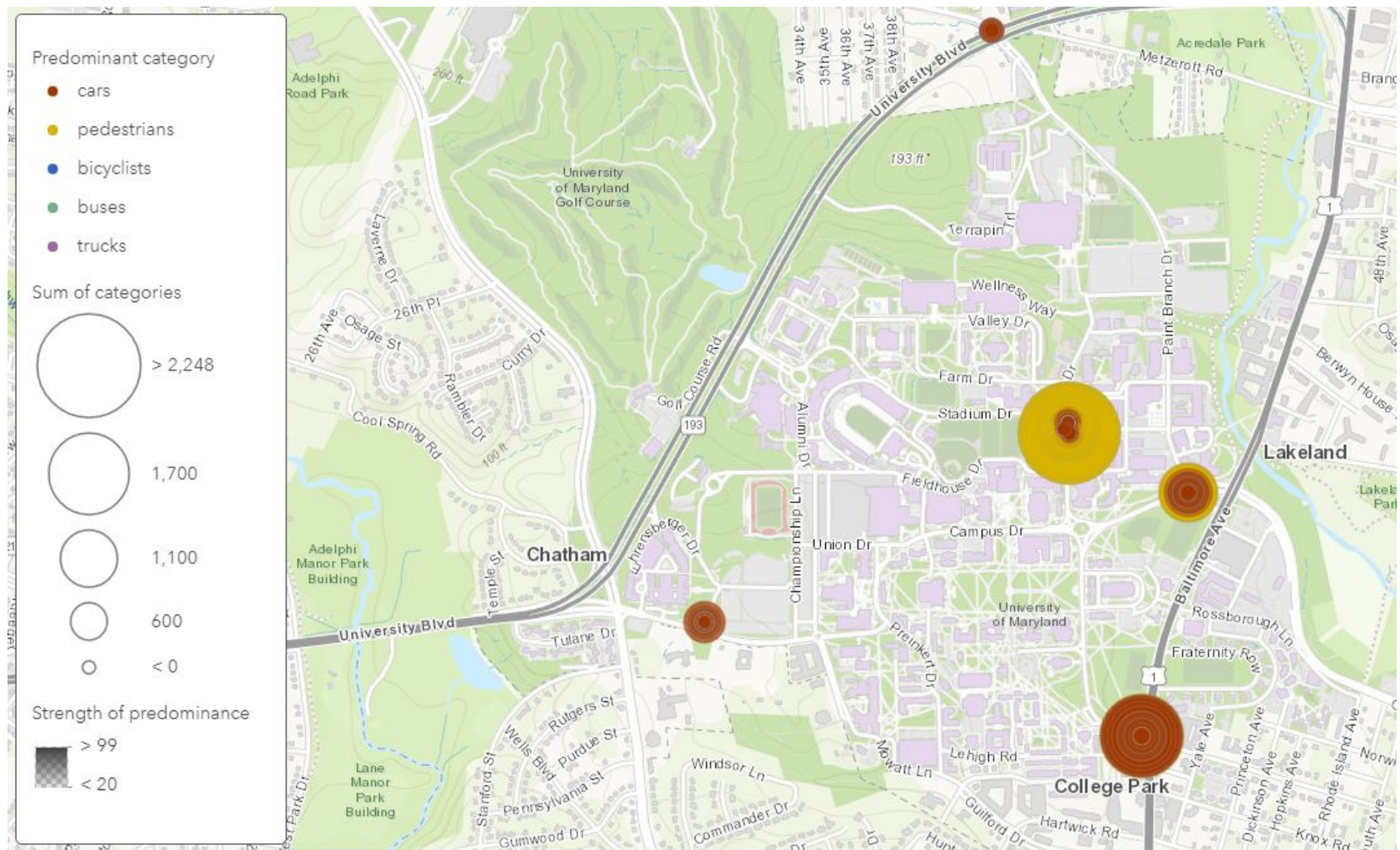




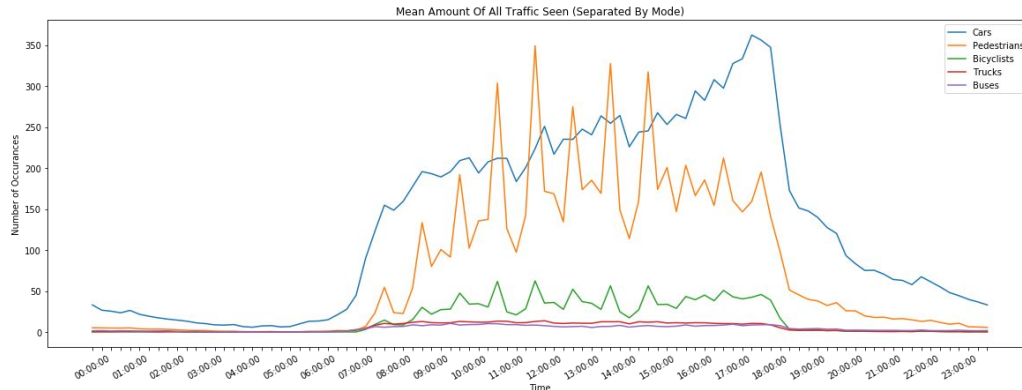
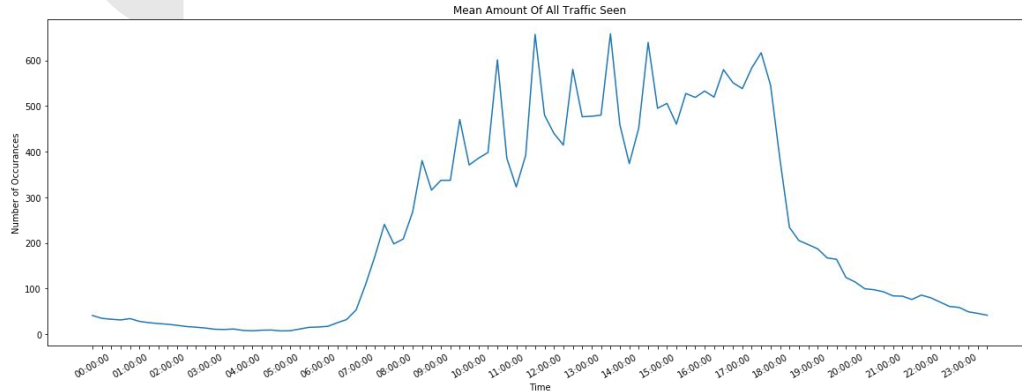




- University Blvd & Point Branch Dr seems to get the least amount of traffic aside from truck traffic
  - Truck traffic increases linearly throughout the week for this street
- South Gate has by far the most car traffic
  - Makes sense given that it is adjacent to a large road
  - Cars most likely aren't actually entering campus
- Regents Dr & Stadium Dr 1 sensor counted a huge amount of bikes compared to other sensors



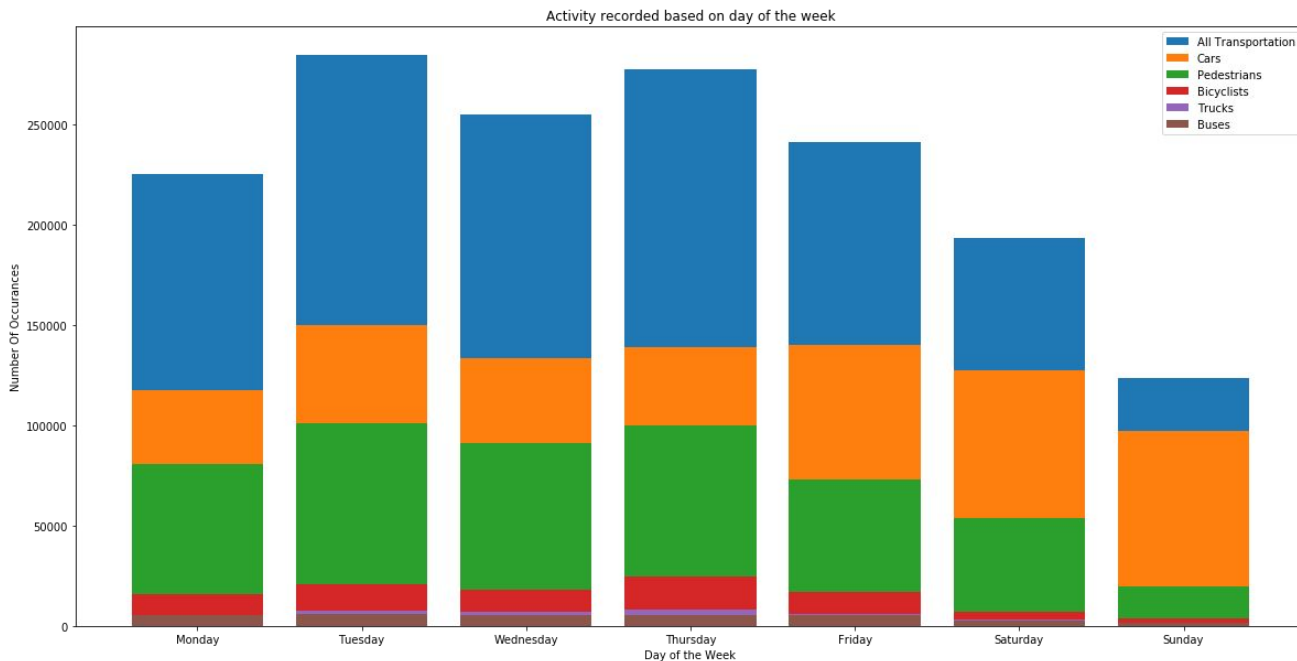
# What time of day has the most traffic?



- Traffic doesn't usually spike up until around 0700
- At its heaviest around 1600-1800
- Interesting to see huge spikes in traffic every hour starting at 1000
  - Due to students getting to class
  - Modes of traffic that increase are pedestrian traffic and bicycles (almost certainly students)
- Spike at 1730
  - Likely due to faculty & staff returning home
  - Why is there no spike like this in the morning?

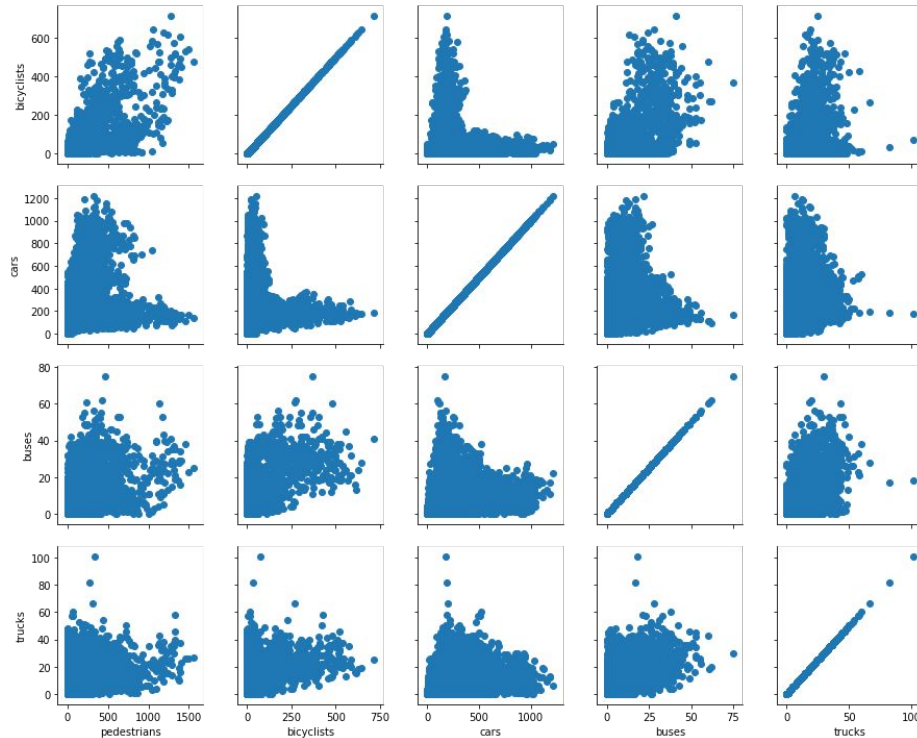


# How about day of the week?



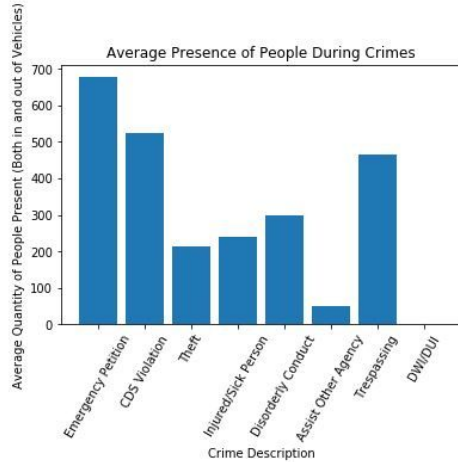
- Tuesday and Thursday seem to be days with most traffic overall
  - Likely to do with how class schedules line up
- Weekend has least amount of traffic

# Correlations between modes of transportation?

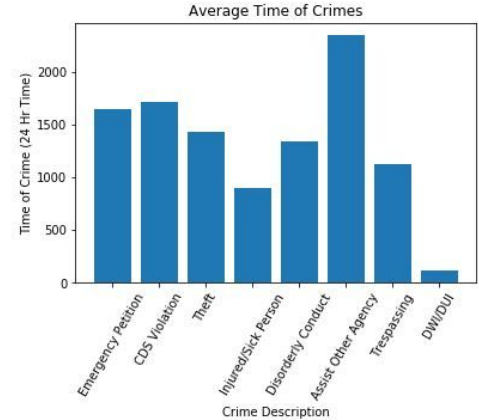


- Strong correlation in amount of cars by bicyclists
  - As the amount of cars on the road decrease, amount of bicyclists increase and other way around is true
- Some correlation in amount of cars by pedestrians
  - As more pedestrians travel, less cars do
- Some correlation in amount of bicyclists by pedestrians
  - The more bicyclists on the road, roughly translates to more pedestrians
  - Makes sense when you think about classes getting out (more pedestrians and bicyclists)

# How does crime affect the traffic flow?



	event	avgTime	avgTraffic
0	Emergency Petition	1645	677
1	CDS Violation	1715	525
2	Theft	1430	215
3	Injured/Sick Person	0900	239
4	Disorderly Conduct	1345	298
5	Assist Other Agency	2345	49
6	Trespassing	1130	465
7	DWI/DUI	0115	1

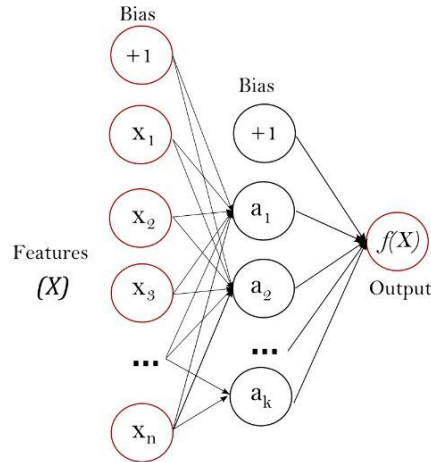


- Interesting to see correlation between how many potential witnesses are present for specific crime types and how that relates to the time of day
- It makes sense that DUI's typically occur late at night around 0115 when bars close and there are few people around for late night crimes like DUI's and Assist Other Agency's
- In the two week period we have data for, there aren't that many crimes so some correlations are fairly loose. There was a higher volume of theft crimes so its results are more accurate:
  - According to ASECURELIFE, most burglaries occur [between the hours of 1000-1500](#), and this data shows that most thefts occurred at 1430 on average



# Building Predictive Models

- Linear Regression
  - Fits a linear model to minimize the sum of squares between observed targets in the dataset and the target values of the dataset.
- Multilayer Perceptron Regressor
  - An iteratively trained regressor than can be used to fit models using non-linear, hidden layers.





# Building Predictive Models

- Both sets of models were trained on the same set of inputs and outputs.
- Input vector example:
  - [pedestrians, bicyclists, buses, trucks, earlyMorning, midMorning, midAfternoon, lateEvening, CampusAndPresidential, CampusAndPaintBranch, RegentsAndStadium, SouthGate, UniversityAndPaintBranch]
- Output example
  - [cars]
- We trained several models on both methods with varying input vectors and output values
- Loose correlations between some input values drove the need to include several modes of transportation as input.





# Building Predictive Models

## Method

- Dealing with time variance in the model
  - Break 24 hour day into 6-hr blocks (early morning, mid morning, late afternoon, and late evening) and store each 6-hr block as a feature for every datapoint
- Splitting training and testing data
  - Shuffling the dataset gave a good spread of locations and time periods for the training and testing sets respectively
- Dealing with sensors sharing an intersection
  - Group the sensors by intersection, integrate new features to the dataset representing the intersection that the datapoint was recorded at



# Building Predictive Models

## Root Mean Squared Error

$$RMSE(y, \hat{y}) = \sqrt{\frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}-1} (y_i - \hat{y}_i)^2}$$

A risk metric corresponding to the expected value of the error or loss.<sup>1</sup>

## Coefficient of Determination ( $R^2$ )

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

where  $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$

Represents the proportion of variance (of  $y$ ) that has been explained by the independent variables. It provides an indication of goodness of fit.<sup>1</sup>

<sup>1</sup>scikit-learn 3.3, Metrics and scoring: quantifying the quality of predictions



# Building Predictive Models

## Results

<b>Linear Regression</b>	<b>Pedestrians</b>	<b>Bicyclists</b>	<b>Cars</b>	<b>Trucks</b>	<b>Buses</b>
<b>RMSE</b>	87.28	31.54	110.55	4.35	5.32
<b>R<sup>2</sup></b>	0.67	0.66	0.63	0.65	0.62

<b>MLP Reg</b>	<b>Pedestrians</b>	<b>Bicyclists</b>	<b>Cars</b>	<b>Trucks</b>	<b>Buses</b>
<b>RMSE</b>	75.08	26.65	93.48	4.17	5.11
<b>R<sup>2</sup></b>	0.76	0.76	0.73	0.67	0.65



# Building Predictive Models

## *Analysis*

- Both methods were good at predicting numbers of trucks and buses, probably because trucks and buses operate on a much more regular schedule with less variability than cars or pedestrians.
- Cars, pedestrians, and bicyclists were hard to predict because of this same fact. Their variability made training a model more difficult
- The MLP Regressor outperformed the Linear Regression model on all metrics, most likely due to this model's ability to fit non-linear data.



# Building Predictive Models

## *Application*

- The MLP Regressor models are reasonably good at predicting snapshots.
- Given pedestrians, bicyclists, trucks, buses, time, and location, these models can predict with reasonable accuracy, the number of cars present.
- Models were created for each transportation method, so one can swap in pedestrians for the target value.