


Spring June 6th, 2019

Machine Learning with Multi-class Regression and Neural Networks: Analysis and Visualization of Crime Data in Seattle

Erkin David George
Seattle Pacific University

Follow this and additional works at: <https://digitalcommons.spu.edu/honorsprojects>

 Part of the [Business Analytics Commons](#), [Civil Law Commons](#), [Computational Engineering Commons](#), [Criminal Law Commons](#), and the [Other Engineering Commons](#)

Recommended Citation

George, Erkin David, "Machine Learning with Multi-class Regression and Neural Networks: Analysis and Visualization of Crime Data in Seattle" (2019). *Honors Projects*. 106.

<https://digitalcommons.spu.edu/honorsprojects/106>

This Honors Project is brought to you for free and open access by the University Scholars at Digital Commons @ SPU. It has been accepted for inclusion in Honors Projects by an authorized administrator of Digital Commons @ SPU.

Machine Learning with Multi-class Regression and Neural Networks: Analysis and
Visualization of Crime Data in Seattle

by

ERKIN DAVID GEORGE

FIRST READER, CARLOS ARIAS

SECOND READER, JOHN HOSSLER

A project submitted in partial fulfillment
of the requirements of the University Scholars Honors Program

Seattle Pacific University

2019

Approved _____

Date _____

Machine Learning with Multiclass Regression and Neural Networks: Analysis and Visualization of Crime data in Seattle (June 2019)

Primary Erkin David George, First Reader Dr. Carlos Arias, Second Reader Dr. John Hossler

ABSTRACT

This article examines the implications of machine learning algorithms and models, and the significance of their construction when investigating criminal data. It uses machine learning models and tools to store, clean and analyze data that is fed into a machine learning model. This model is then compared to another model to test for accuracy, biases and patterns that are detected in between the experiments. The data was collected from data.seattle.gov and was published by the City of Seattle Data Portal and was accessed on September 17, 2018. This research will be looking into how machine learning models can be used to generate predictions and how the data management will introduce a bias that is unavoidable. This bias will be discussed, as well as the importance of understanding this bias for sensitive data, such as this crime data.

Extended abstract

Looking at criminal data in Seattle, we see how machine learning can be construed to demonstrate the unreliability of setting up machine learning models with the hope of them producing objective results. This will also show how many ethical concerns there are with these machine learning algorithms, as the majority of companies employing these algorithms have not had experience in the past to make sure they do not make mistakes. This could be a loan that should have been given to someone in need, but there was unintentional racial profiling in the data. Or in the case of Seattle Police reports, there could be a disproportionate number of minority arrests compared to the average Seattleite. In this study, the data being investigated contains information such as the location of a crime report, the type of crime reported and the time of the incident. Using this data in pre-defined geographical spaces, we will investigate data patterns in the actual and the predicted data from this experiment. The data we are looking at has been sampled due to the size of the dataset. This is done by using selecting a random 20 percent of the data to train the machine learning models, as investigating with counts and more advanced techniques is outside the scope of this research. Then 2000 generated datapoints for each visualization are selected and graphed as seen below.

1 Machine Learning and its significance

Machine learning is a discipline that combines the fields of mathematics, statistics and computer science, with the aim to make programs that learn to perform tasks efficiently, without being given explicit instructions on how to do so. Rather, the programs learn by using algorithms and large datasets, and combine these methods with previous answers and data sources to guess the next answer for a given problem. Machine learning enables pattern recognition in large datasets that would be impossible to analyze manually [13]. It is classified as a subset of artificial intelligence, and is primarily used in business and production applications, in contrast to the layman imagination of Terminator-like AI. Machine learning manages banking decisions, housing prices, supermarket layouts and medical diagnoses in our day-to-day lives, yet the process is often obfuscated or incomprehensible to the average person [16]. As such, machine learning is important in our day-to-day lives, but we often do not understand how the decisions are made by the machines that we have set up to govern many parts of our lives.

2 Criminal data in Seattle

The data being examined is from the City boundaries of Seattle, including neighborhoods such as Queen Anne, Magnolia, Ballard and other neighborhoods that are patrolled by the Police Department of Seattle [1]. This data originally contained the following columns:

- Report Number. This is the unique identifier for the crime that has been filed in the data portal of Seattle.
- Occurred Date. This is the day that the crime occurred on.
- Occurred Time. This is the time that the crime was approximately committed, reported in the 24-hour time system.
- Reported Date. This is the day that the crime was officially reported to the Seattle Police Department.
- Reported Time. This is the time of day that the crime was officially reported to the Seattle Police Department.
- Crime Subcategory. This is an explanation of the crime that was committed.
- Primary Offense Description. This is an explanation of the crime committed, with more detailed to differentiate between variations of crimes (i.e. burglary with force vs. burglary with no force)
- Precinct. This is the precinct that the crime occurred in, which is defined by the Police Department of Seattle.
- Sector. This is the sector that the police force responsible for handling this crime patrols.
- Beat. This is the specific police route in a given sector.
- Neighborhood. This is the pre-defined geographical boundaries as defined by the City of Seattle.

From these columns, Crime Subcategory, Precinct, Sector, Neighborhood and Occurred Time were used in this experiment to train the machine learning models. Report Number was not included, as this is a data value that is not tied to the nature of the data, as it merely indexes each entry based on the time that this data was gathered. The Reported Time values were dropped, as this value is unpredictable due to people waiting to report crime, delays in processing paperwork and other miscellaneous factors. Beats were dropped because they are closely related to sectors and would need to be investigated with more information than is currently present in this experiment and publicly available. Crime Subcategory and primary Offense Description are treated the same in this experiment, as they describe the crime in slightly differing levels of detail that are beyond the scope of this project for further analysis.

Previous versions of this data have withheld 40% of the crime [1], and this updated record includes all records since 2008. One thing to note is that offense reports are constricted to the beat level, as this was done to preserve community and data privacy. However, this will not be discussed further, as the quantity of data is sufficient for purposes of this experiment, as there is a total of 492,436 reports that are used to train the models in this experiment.

3 Azure Machine Learning and Tools used

3.1 Overview of Azure ML

Azure Machine Learning, or Azure ML, is a drag-and-drop tool that was created by Microsoft as a cloud-based service to be used for making apps and devices smarter [13]. This tool can be used to create features of products available in our day to day lives, such as fraud detection in banking, predictive maintenance or image classification. This studio supports predictive models, data processing and other essential tools in handling large quantities of data, using the data to create programs and tools to support the goals of institutions, businesses and individuals [2].

3.2 Python Scripts in Azure

Graphing and cleaning of the data was done in the programming language Python, as this is a well-documented and powerful tool for data science and machine learning. Within Azure, there is a Python module that allows for the use of Python 4.0, the most modern version of Python [12]. This version of Python allows for a variety of tools to be imported and used to assist the research. In this project, the Pandas library was used, as well as the Matplotlib library, which are tools that are used in machine learning as visualization and analysis tools [3]. Specifically, Pandas is used to manipulate and analyze data, offering data structures and operations that make processing data efficient, easy and reliable. Matplotlib is a library that makes plotting graphs and visualizing data much easier than the default Python methods of data visualization. Additionally, these visualizations can be customized, unlike the auto-generated graphs from Azure. This allows for colored graphs, precise scales, and graph labels that describe what is being plotted. The results from these scripts are seen later in the appendix and section 5.

4 Explanation of Models used

4.1 Machine Learning Models

A machine learning model can be defined as a mathematical construction that aims to learn about data given to it, with some pre-defined learning objective. The learning objectives are generally focused on gaining insight into the data-generating process or the factors that influence this data's creation, so that more informed decisions can be made regarding the data [10].

With our models, we are constructing what is defined as a supervised learning model [11]. This means that the mathematical computations done via this module require that the data being used is labeled and categorized, rather than general data such as just numbers. In contrast, an unsupervised learning model would cluster, aggregate or move data together without giving significance to factors such as the data type, name or classification that is provided. These are created by the use of the Azure modules that have taken care of the mathematical calculations and corrections that would need to be accounted for in defining a machine learning model. In 4.2 and 4.3, we discuss some of the input that the user of these modules can give, so that the experiments can be customized for the users. These are used alongside the data from section 2 to define the models that we are using to create predictive results.

4.2 Multiclass Logistic Regression

Logistic regression is a method in statistics that is used to predict the probability of an outcome, given data that fits it to a logistic function. This is typically a dichotomous relationship between one dependent variable, and one or more independent variables. In Azure, there is a module that automatically is configured to perform Multiclass Logistic Regression on a data set that is provided, given a predefined set of parameters [14]. This module assumes there are three or more outputs that from the data that is provided. The parameters that can be adjusted include the following adjustable settings:

- Trainer mode for the model
 - Configuration of the model to select a single feature, or to sweep for a best parameter. For this experiment, the mode was Single Parameter, as we are training by Crime Subcategory.
- The optimization tolerance
 - This sets a threshold value for stopping the number of iterations the model runs. Once the improvement is less than the threshold, the model returns the current iteration.
- L1 regularization weight
 - This can be used for sparse models, where data is lacking. For this experiment, this matters less, as our model has an abundance of data to use for training.
- L2 regularization weight

- Used for data that is not sparse. One method of finding this is trial and error. For this experiment, an arbitrary value was picked.
- Memory size for L-BFG
 - This determines how much memory is used to store data from a past computation for the next step. More memory is more accurate but results in slower training.
- Random seed number
 - Specifying a number for this field allows you to get repeatable results from the experiment, provided all other fields remain the same. This field accounts for the pseudorandom nature of number generation in computers (9). By selecting a number, the computer generates the same numbers for the algorithm to use.
- Allow unknown categorical levels. Any values passed in the test dataset that are not present in the training data are marked as unknown.

This Logistic Regression model requires numerical values for it to operate. As such, the categorical values in our columns are converted into a numerical representation internally, so that the model can use the values.

Additionally, we have arbitrarily picked the number 42 as the random number seed, so that the results produced from this experiment could be reproduced.

4.3 Multiclass Neural Network

Neural Networks are defined as a set of patterns and algorithms in layers that aim to loosely model the human brain [15]. The first layer is the inputs, and the output layer produces an acyclic graph with weights and nodes corresponding to the input values. In Azure, there is a module that automatically configures itself once it is provided the following configurations:

- Trainer mode for the model
 - Configuration of the model to select a single feature, or to sweep for a best parameter. For this experiment, the mode was Single Parameter, as we are training by Crime Subcategory.
- Hidden layer specification
 - Selects the network architecture to use for training. This is either the fully connection case, or a custom script that is defined in the Net # language [2]. If the custom script is selected, you must also configure the neural network definition.
- Number of hidden nodes
 - Allows you to customize the number of hidden nodes.
- The learning rate
 - Defines the step taken at each learning iteration before corrections occur.
- The number of learning iterations
 - Specify the number of times the algorithm should run to train.
- The initial learning weights diameter
 - Specify the node weights at the beginning of the training process.
- The momentum
 - Specify a weigh to apply to nodes while training from previous iterations.
- The type of normalizer
 - Selects method for feature normalization
 - Binning normalizer. Creates groups of equal sizes, and then normalizes each value in the groups, divided by the total number of groups.
 - Gaussian normalizer. Rescales the values of each feature to have mean 0 and variance 1.
 - Min-max normalizer. Rescales all features to a 0 to 1 interval
 - Do not normalize. No normalization is performed.
- Shuffle examples
 - If selected, this keeps the cases in the exact same order between iterations of the nodes that are selected and run for the experiment.

- Random number seed
 - See 4.1
- Allow unknown categorical levels. See 4.1
- Creates groupings on unknown values. May be less precise but can provide better predictions for new unknown values.

Most of these settings were left as the default. This means we trained by a Single Parameter, which was Crime Subcategory. Our layer specification was set to the fully connected case, which means that we had one hidden layer, with a default of 100 nodes. Our learning weight was set to 0.1, and our momentum was 0, meaning our previous iterations do not affect the weighting process. The random seed was set to 42, as done in section 4.1. Our binning process is the default of the min-max normalizer, where every 0 to 1 interval, the features are linearly rescaled.

5 Other modules

5.1 Training Model

The training model is used to train either a classification or regression model in Azure in this experiment. These are both supervised models, that require precise setup. The regression model is set up to process all the labels, and then attempt to predict the Crime Subcategory. This model executes the training once the model and its parameters have been set, which allows an easy switch between classification algorithms. This can be used to retrain an existing model with new data, keeping the old configuration so that the changes to the setup of the experiment are controlled and predictable. The label column selected for our experiment is Crime Subcategory, so the studio proceeds to attempt to predict this type based off of the rest of the data that has been fed into the model. The neural network model performs the same basic operations, but instead of using regression to train, it uses a neural network to learn.

5.2 Score Model

Score model generates predictions based on the results of the training model. For this experiment that focuses on classifying and predicting data, our results come with a predicted value for the class, as well as the probability of the predicated value. The scoring is taken care of by the Azure platform and its resources, and the details of how this works is beyond the scope of this paper [14]. From the results of this training, our probabilities add up to 1, as the data that is generated appears in the following form:

B	C	D	E	F	G
					
0.050837	0.064371	0.037023	0.030613	0.055986	0.044502
0.050827	0.06436	0.037016	0.030607	0.055976	0.044495
0.050818	0.064348	0.037009	0.030602	0.055966	
0.050809	0.064337	0.037002	0.030596	0.055957	0.044481
0.050799	0.064326	0.036995	0.030591	0.055947	0.044474
0.05079	0.064315	0.036987	0.030585	0.055937	0.044467
0.050781	0.064304	0.03698	0.030579	0.055928	0.04446
0.050771	0.064293	0.036973	0.030574	0.055918	0.044453

The rows represent the times, which are not shown, as that would result in a widely spread graph that would be difficult to view. As such, the graphs further in the section and in the appendix present a sample of the data in a more visually appealing manner.

5.3 Graphs and Data Results

5.3.1 Python Code Explanation and Visualization Explanation

The graphs below were generated using Python and Matplotlib, as mentioned in section 3.2. The opacity of the data has been set to 0.3. This is measured on a scale of 0 to 1 where 0 is renders the data transparent and 1 is totally opaque. This makes the data easier to understand in the visualization, due to the excess of data being generated by these models.

The scale of these graphs has been designed to have probability as the y-axis, and time of day as the x-axis. Probability is measured on a 0 to 1 scale, where 0 is when an event is guaranteed not to occur, and 1 is a guaranteed event. Time of day is measured in military time, where the day starts at 0 and ends at 2400.

The probability on the y-axis has been limited to 0.05 as the maximum, as this range has been found to contain the majority of the resulting probabilities that each set produces. The x-axis is from 0 to 2400, which represents the time in military type, as this was how the data was recorded in the original dataset. This also allows for direct numerical scale, avoiding the need to create a graph that accounts for the AM and PM definition of time. There is one notable exception to this scale with the Car Prowl subcategory, as it was necessary to set the probability scale for this crime from 0 to 0.5, as the probability for this crime was substantially higher than all the other crimes.

The reason for this scale and design of the graph is because the data that was collected was time and location based, and we are attempting to map what the probability of a certain crime is at a given time and place. As such, the crimes have been colored differently per graph for differentiation, per the keys attached to the graphs.

5.3.2 Explanation of the Visualization and Keys

To understand the data in the graphs, we need to note that the data being presented has been plotted unusually. The x-axis is the time of day for the crime, measured from 0 to 2400. The y-axis is the probability of the event occurring, as selected from the entire dataset generated by the model. As such, the probability number may seem low, but this is due to the sampling of the data, as presenting all the data that aggregates to a probability of 1 would result in an overly complicated visualization.

Each graph has its own key, which should be used only for that graph. This is necessary because of the number of graphs, and the limited number of colors that were easy to combine together.

Looking at the graphs, there are points where we appear to have duplicate data that has a different probability for a given time. This is due to the data being combined and reduced, as the original data set had too many labels to comprehend in the scope of this project. The combination of crime types is outlined in the next section.

For a complete visualizing of the data from this project, see the appendix for all graphs generated by the models.

5.3.3 Combination of Crime types

Burglary:	Robbery:	Theft:	Assault:	Sexual crime:	Conduct crime:
Burglary-commercial-secure parking	Robbery-commercial	Motor vehicle theft	Aggravated assault-dv	Sex offense other	Disorderly conduct
Burglary-residential	Robbery-residential	Theft shop-lift	Aggravated assault	Rape	Liquor law violation

Burglary-residential-secure parking.	Robbery-street	Theft-building		Prostitution	Loitering
Burglary-commercial		Theft-bicycle		Pornography	Gamble
		Theft-all other			Trespass

The crime types above were combined based on the similarity of their description. Further analysis of how these crimes being aggregated in this manner would influence the predictive results requires more criminology and sociological knowledge. As such, this decision should be considered an arbitrary decision for this experiment.

Below is a list of the neighborhoods that were used for the graphs with neighbors listed:

- Central Area/Squire Park
- Queen Anne
- Northgate
- Unknown
- Lakewood/Seward Park
- Alki
- High Point
- Roxhill/Westwood/Arbor Heights
- Brighton/Dunlap
- Slu/Cascade
- Alaska Junction
- South Park
- North Admiral
- Capitol Hill
- Lakecity
- Claremont/Rainier Vista
- Roosevelt/Ravenna
- Hillman City
- Sodo
- South Beacon Hill
- Highland Park
- Wallingford
- Fauntleroy Sw
- Sandpoint
- Rainier View
- North Beacon Hill
- Morgan
- Magnolia
- Bitterlake
- South Delridge
- North Delridge
- Greenwood
- Rainier Beach
- Mid Beacon Hill
- Mount Baker
- Phinney Ridge

- Chinatown/International District
- Belltown
- University
- Pioneer Square
- Genesee
- Ballard North
- Miller Park
- Downtown Commercial
- Ballard South
- Madrona/Leschi
- Judkins Park/North Beacon Hill
- Fremont
- Columbia City
- Montlake/Portage Bay
- New Holly
- Madison Park
- Georgetown
- Eastlake - West
- Commercial-Harbor Island
- Commercial Duwamish
- Pigeon Point
- Eastlake - East

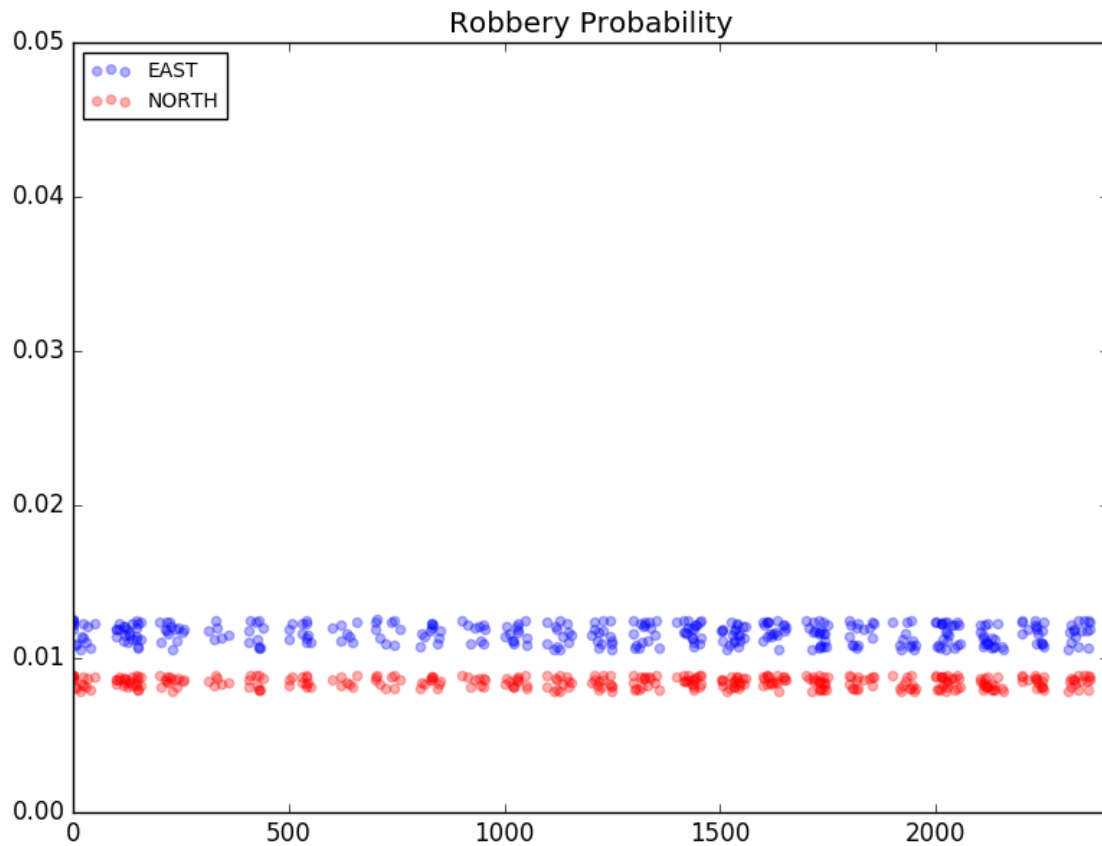
There are this many relevant sectors in our dataset. We consider them more relevant because of their incidence in the dataset. The other sectors were dropped due to lack of data in the original dataset for these sectors or incomplete data. Below is a list of relevant sectors that were used for all the graphs with sectors listed:

- B
- C
- D
- E
- F
- G
- J
- K
- L
- M
- N
- O
- Q
- R
- S
- U
- W

These sectors are divided into three beats, such as B1, B2, B3 or C1, C2, C3. For our analysis of the data, we ignore the individual beats and instead focus on passing in the names of the sectors to the model.

Test Results from Multi-Linear Regression

Figure 1



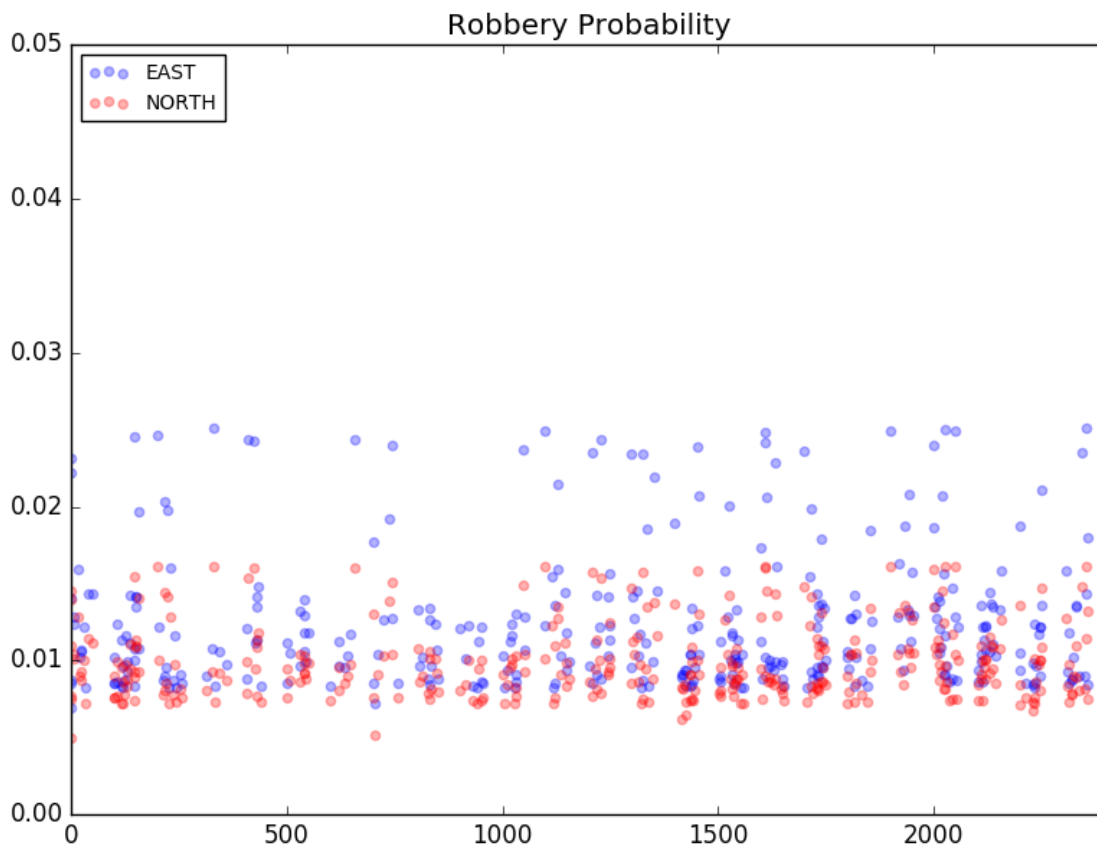
The figure above is a selection of the data generated by the multi-linear regression model in our experiment. It contains the results of 2000 rows from the data. This number was chosen arbitrarily, as there is an excess of data to select, and the data that is selected is randomly selected by using the line of code below:

```
Dataframe1.sample=(n=2000, random_state=1)
```

The random state of 1 is an arbitrarily chosen number, but by having this parameter, we ensure that the exact same process is run on the data between the runs. This way, we do not end up with different graphs when the model produces the same results.

Test Results from Multi-Class Neural Network

Figure 2



6 Results and Analysis

6.1 Understanding Police Patrols

The data that used in this experiment was reported by police patrols that operate in three distinct ways:

- Patrol vehicles
- On foot, known as foot beats
- On bike, known as bike patrols

These patrols operate 24/7 year-round. This is done with by rotating schedule as listed below:

- First Watch: 03:30 - 12:30 (3-12 PM)
- Second Watch: 11:30 - 20:30 (11-8 PM)
- Third Watch: 19:30 - 04:30 (7-4 AM)

Seattle is divided into five geographic areas, which are broken into five precincts that contain police stations. These are North, East, South, West and Southwest. Each of these precincts contains smaller areas called Sectors. In Seattle, there is a total of 17 sectors. Within each Sector is a three-part section, which are beats that individual patrol officers are responsible for [17].

6.2 Comparison and Contrast of Results

6.2.1 Assumptions

One factor to keep in mind is that these models have all learned from criminal data that has already been reported. This means that with the assumption that a crime has happened, we end up with the probabilities that we see above. This assumption is made because the model was trained with data that reports the crimes that have occurred. Lack of crime is not calculated into and non-criminal incidents are not factored into the original data, and this influences what our results look like.

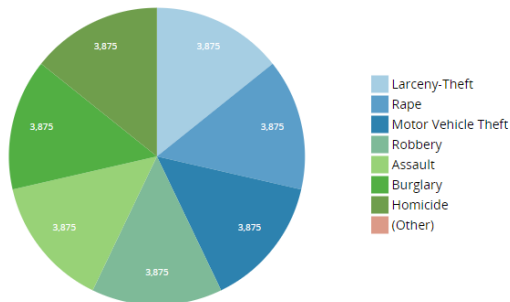
Looking at the graphs above, there are multiple points where the probability for a crime is different, but the time is exactly the same. This is due to the overlap of data that was necessary to condense the incredibly large dataset, where we have the different crimes being trained in our original model, but then we view the results in a less precise manner. Section 6.2.3 explains how the crimes were combined together, due to their similarities.

6.2.2 Comparison between the Regression and Neural results

Comparing the results for Robbery from the Multiclass Logistical Regression model, we see a wide difference between the results for the Neural Network (Figure 2) and the Regression model (Figure 1). The ranges for the results from Neural Model range from 0 to nearly 0.03, while the Regression Model stays under the 0.02 probability mark.

One problem with this data visualization is that we do not get a sum of values that add up to 1, which would normally indicate that something is wrong with the predictive model. However, this is due to the nature of the experiment, where we needed to sample the values from the predictions. There are approximately 492,436 possible rows to select data from, with 10 columns representing the combined crime type from the original 33 columns. With Pandas, the data was cleaned to reduce the data to 2000 rows, with the crime column plotted against the time and probability value of the given neighborhood. This number was chosen after trying out various other values, which either led to the graphs being too filled with data to understand the trends, or overly sparse, making it difficult to understand trends.

With these results, we can see that the predictions generated by the neural network tend to be more spread out and varied, while the regression model tends to linearize the data. One interesting note is that while the probability varies between these models vertically, we notice that the gaps in the horizontal axis tend to stay the same. This trend in the gaps of data at certain times continues in the rest of the graphs in the appendixes, regardless of the crime or the model that was used. The probabilities for the crimes tend to remain in the same range, with the notable exception of Car Prowls. This crime is predicted to occur at nearly half the time, based on graphs in the appendixes. This does not match with data visualization from the Seattle Police department, seen below:



As such, this shows that the machine learning model has likely been influenced by biases introduced as the model was constructed [17].

6.2.3 Patterns in the data and comparisons

One interesting feature of the graphs that we can see is that there are gaps in data, which do not necessarily correspond with the watch changes, as seen in section 7.1. This is likely because of the method of training, where the time label is not considered the primary training method, as we are more concerned with the type of crime that was committed. As such, it appears that the results give an approximation of the gaps in time, as would be accurate with the changing of the watch for a police patrol [17]. This allows for the human error in recording data, the necessity of going over a shift if there is an ongoing crime and other external factors that is not recorded explicitly in this data.

7 Implications for Machine Learning Use in Criminal Studies

To understand Machine Learning, it is necessary to study not only science and programming, but also philosophy, sociology and other disciplines that are looked over, as they are labeled as humanities. Weisburd explains the important understanding of how the places we examine define our results in studying crime in a city. Traditionally, criminology theories have focused on the person [6], attributing the crime to various aspects about the person such as biology, psychology, or sociology factors as some of the main points of interest. However, it would seem that this is not necessarily a metric that should be followed strictly. Criminal records and tracking, while valuable, does not account for certain locations being known for higher crime, as locations do not possess attributes that we would call criminal. Rather, Weisburd argues that the offenders come into contact with their targets more often at specific locations, as a result of repeated interactions [6]. From the results of the experiment, we may note that there is no description of the individuals who committed these various crimes, but this should not be a cause for concern in the scope of this paper. The data instead gives general descriptions of what may occur in a location, based on redacted data that aims to not expose the identities of individuals involved with these incidents. The predictions show that while this may be a viable approach, there is an excess of unknown variables that were not analyzed with this machine learning example, such as socioeconomic status of the offenders, their gender and other details that were kept from the public.

In stark contrast, one modern example of machine learning being used in a concerning manner is the State of China. Yi Shu Ng examines how China is using predictive analytics to notify their police department of potential criminals, based on behavioral patterns. An example of the data being examined is the shopping purchases, where combinations of items such as a hammer and a sack increases the suspicion of an individual [4]. Additionally, traveling outside the country to certain locations and being associated with others who are suspicious can result in an individual being marked as suspect. This is done by using the extensive surveillance technology already present in China and is aimed at predicting a crime in a preventative manner so that the police can apprehend the individuals before they can commit the crime. The companies investing in these technologies allow the Chinese police receive alerts when the marked individuals appear on a surveillance camera [4].

Therefore, it is clear that an objective description of what crime looks like is a nigh impossible objective to reach, as every bit of data that could be possibly interpreted is not available or should be used in machine learning models. If China is able to produce results from their mass surveillance, how might this incentivize other governments to follow their lead, in hopes of creating a more stable state? Pursuing the knowledge that comes from machine learning is certainly important, but the process behind this must be made clearer and should be reviewed so that biases are not introduced to the analysis. Otherwise we end with either a machine learning process that is not effective in predicting crime, or we start developing dystopian technology that can be wielded against the public. This may be more realistic and terrifying than any Terminator movie or technology dystopia that we have imagined, as this is already happening behind closed doors.

8 Bibliography

- [1] data.seattle.gov. 2018. Crime Data. (September 2018). Retrieved May 9, 2019 from <https://catalog.data.gov/dataset/crime-data-76bd0>
 - [2] J-Martens. Azure Machine Learning Documentation – Tutorials, API reference. Retrieved May 9, 2019 from <https://docs.microsoft.com/en-us/azure/machine-learning/>
 - [3] Matplotlib development team. 2019. Matplotlib Website. (April 2019). Retrieved May 9, 2019 from <https://matplotlib.org/>
 - [4] Yi Shu Ng and Yi Shu Ng. 2017. China is using AI to predict who will commit crime next. (July 2017). Retrieved May 9, 2019 from <https://mashable.com/2017/07/24/china-ai-crime-minority-report/>
 - [5] Anon. 2018. NIBRS. (September 2018). Retrieved May 10, 2019 from <https://www.fbi.gov/services/cjis/ucr/nibrs>
 - [6] Anon. Python Data Analysis Library¶. Retrieved May 9, 2019 from <https://pandas.pydata.org/>
David Weisburd. 2016. *Place Matters: Criminology for the Twenty-First Century*, New York: Cambridge University Press.
 - [7] Anon. 2015. An interactive machine-learning-based electronic fraud and abuse detection system in healthcare insurance. (July 2015). Retrieved May 24, 2019 from <https://www.sciencedirect.com/science/article/pii/S1568494615004585>
 - [8] Anon. 2009. Intrusion detection by machine learning: A review. (May 2009). Retrieved May 24, 2019 from <https://www.sciencedirect.com/science/article/pii/S0957417409004801>
 - [9] Anon. A Million Random Digits. Retrieved May 24, 2019 from https://www.rand.org/pubs/monograph_reports/MR1418/index2.html
 - [10] Carsten Riggelsen. 2008. *Approximation methods for efficient learning of Bayesian networks*, Amsterdam, The Netherlands: IOS Press.
 - [11] Devin Soni and Devin Soni. 2018. Supervised vs. Unsupervised Learning. (March 2018). Retrieved May 24, 2019 from <https://towardsdatascience.com/supervised-vs-unsupervised-learning-14f68e32ea8d>
 - [12] Anon. Welcome to Python.org. Retrieved May 24, 2019 from <https://www.python.org/>
Anon. Machine Learning Studio. Retrieved May 23, 2019 from <https://azure.microsoft.com/en-us/services/machine-learning-studio/>
 - [13] Anon. Machine Learning Studio. Retrieved May 23, 2019 from <https://azure.microsoft.com/en-us/services/machine-learning-studio/>
 - [14] Anon. Data.gov. Retrieved February 2, 2019 from <https://www.data.gov/>
 - [15] Christos Stergiou and Dimitrios Siganos. Retrieved May 23, 2019 from https://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/cs11/report.html
- 16 Ted G. Lewis and Peter J. Denning. 2018. Learning machine learning. *Communications of the ACM* 61, 17
17 data.seattle.gov. Seattle Police Beats 2015-2017. Retrieved May 24, 2019 from <http://data-seattlecitygis.opendata.arcgis.com/datasets/seattle-police-beats-2015-2017?geometry=-124.087,47.454,-120.595,47.778>

8.1 Author's Reflection

Our world is full of fascinating connections between all of the various aspects that govern our lives, be it the place we live, the livelihood we occupy or any number of variables that change who we are. While much of who we are is quantifiable and measurable, it may be wise to tread carefully before we make decisions based off of what we consider objective data. Many of the decisions we make in life are made off of what many would call a gut instinct, or a feeling. This extends anywhere from choosing a job to deciding where to go to school, and often is the metric that we use in making important decisions. While

8.2 All graphs produced by the models

The graphs that are produced below have been sorted into two sections: One is for the graphs that follow the regular scale for this experiment, and the others are scaled so that the data is visible.

8.3 Neural Network Graphs

Figure 3

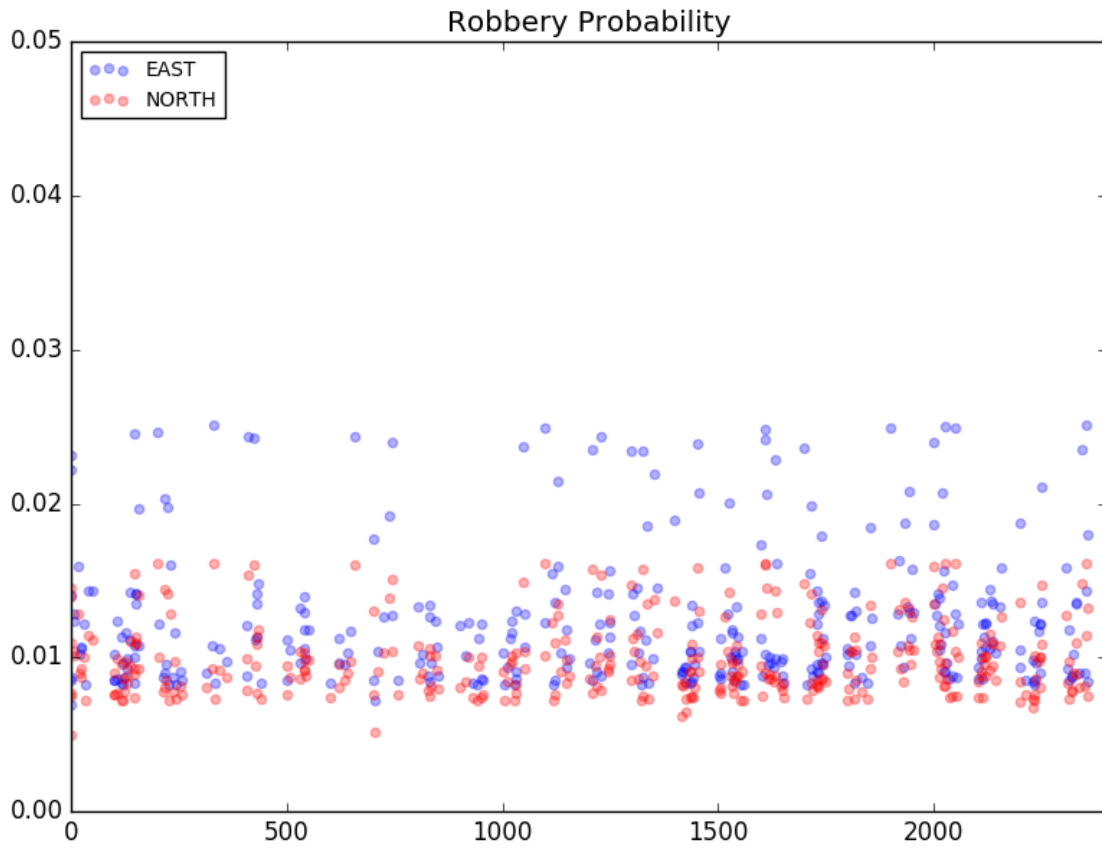


Figure 4
Robbery Probability

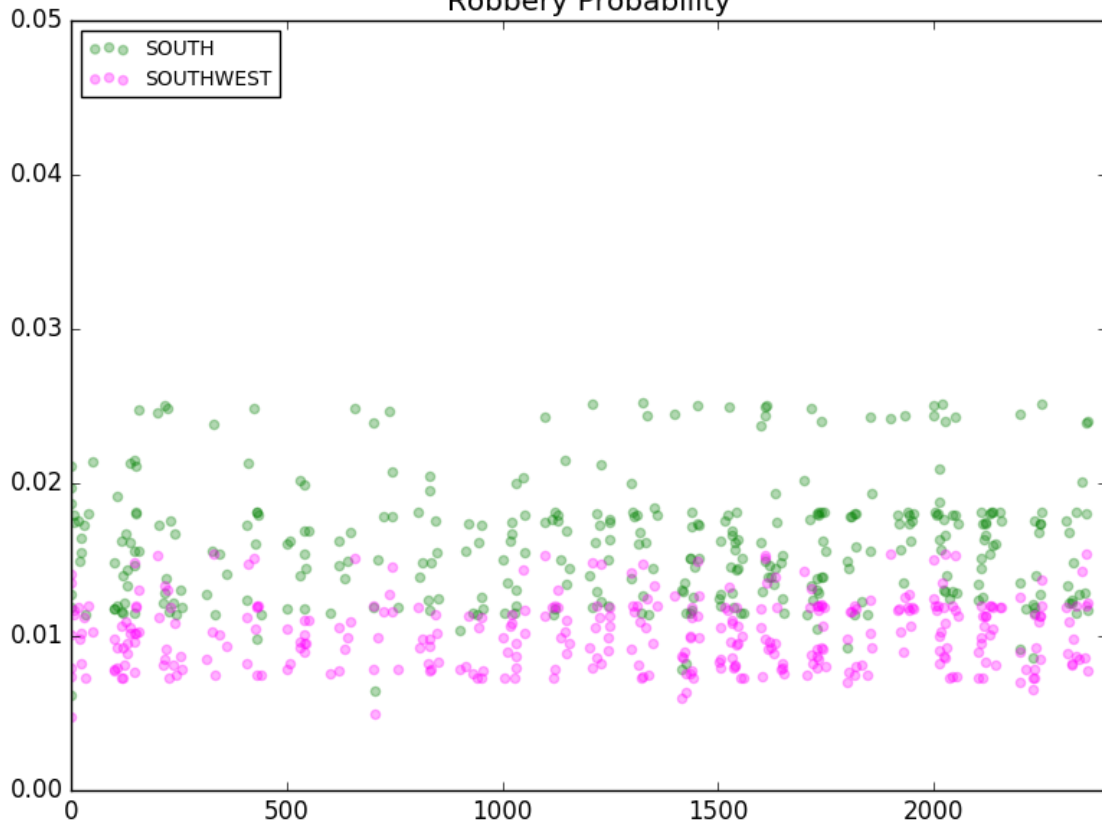


Figure 5
Arson Probability

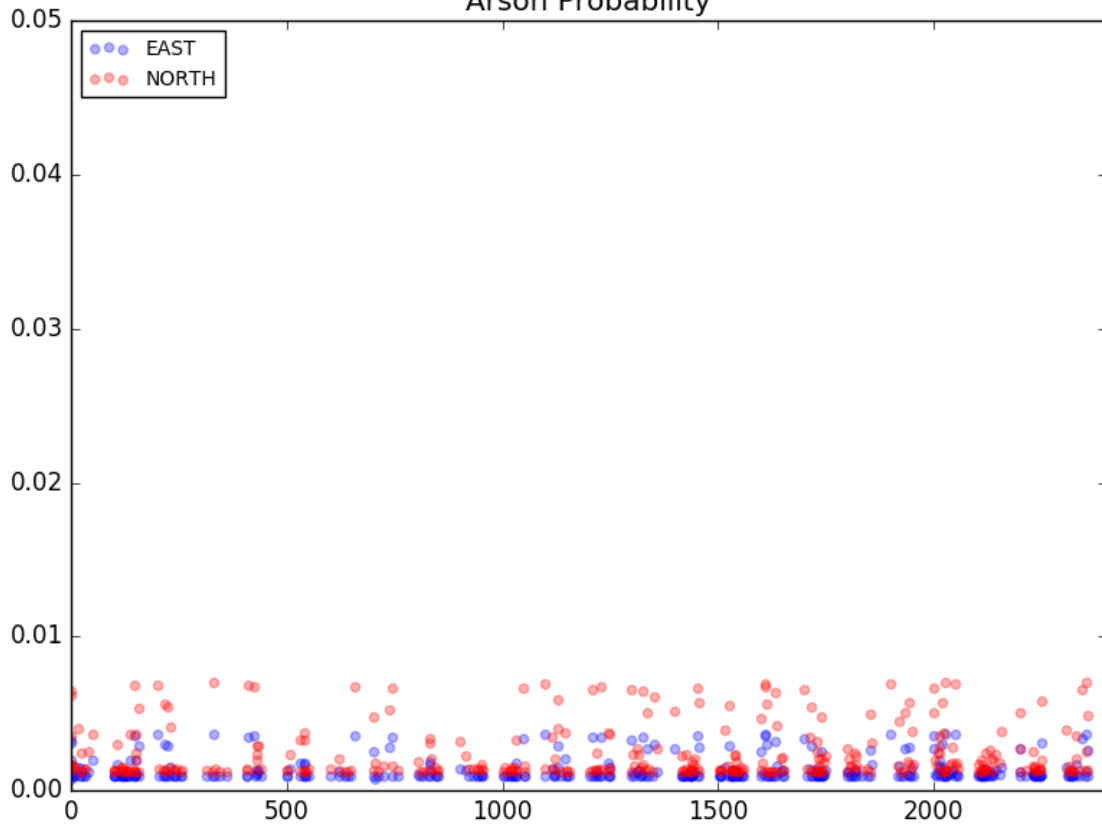


Figure 6
Arson Probability

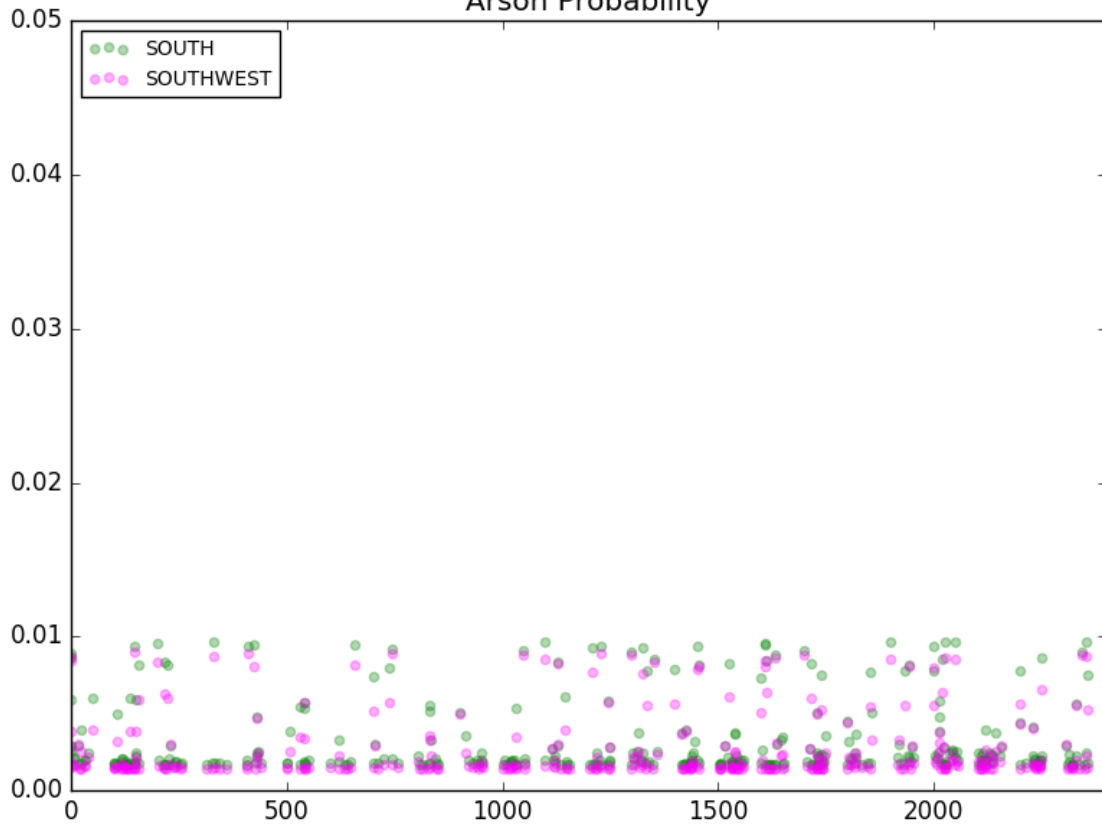


Figure 7
Burglary Probability

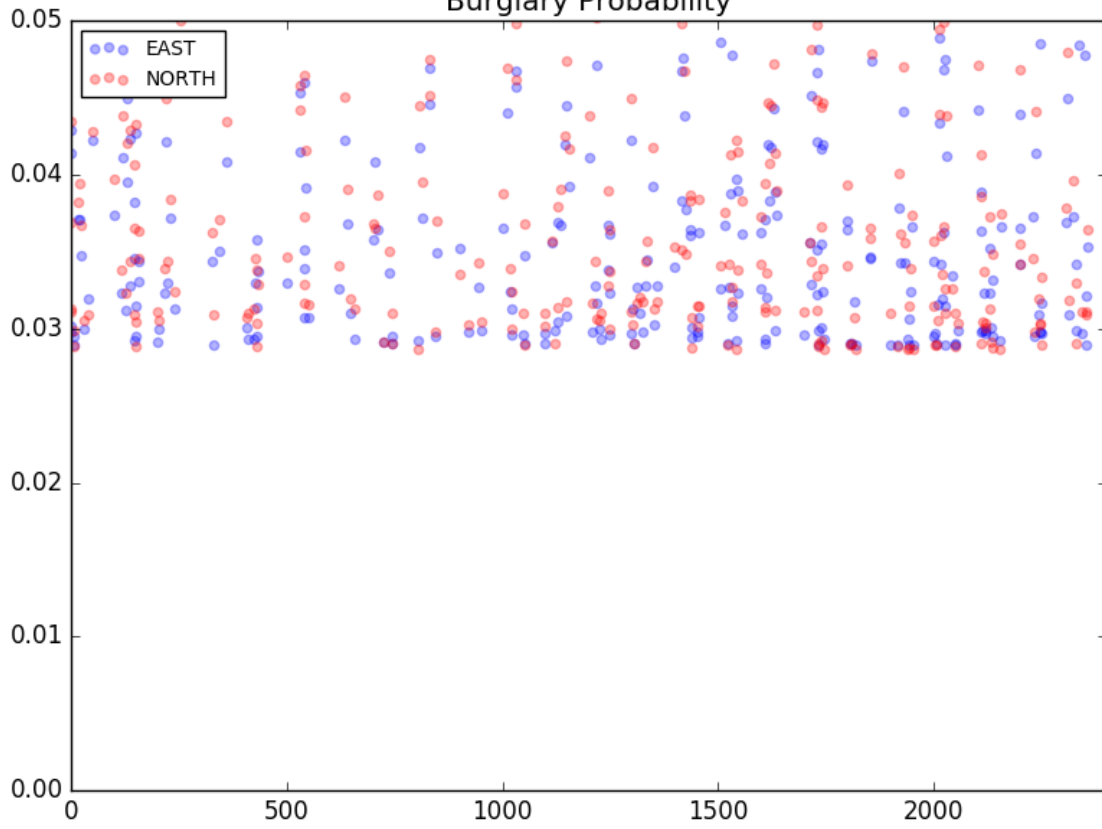


Figure 8
Burglary Probability

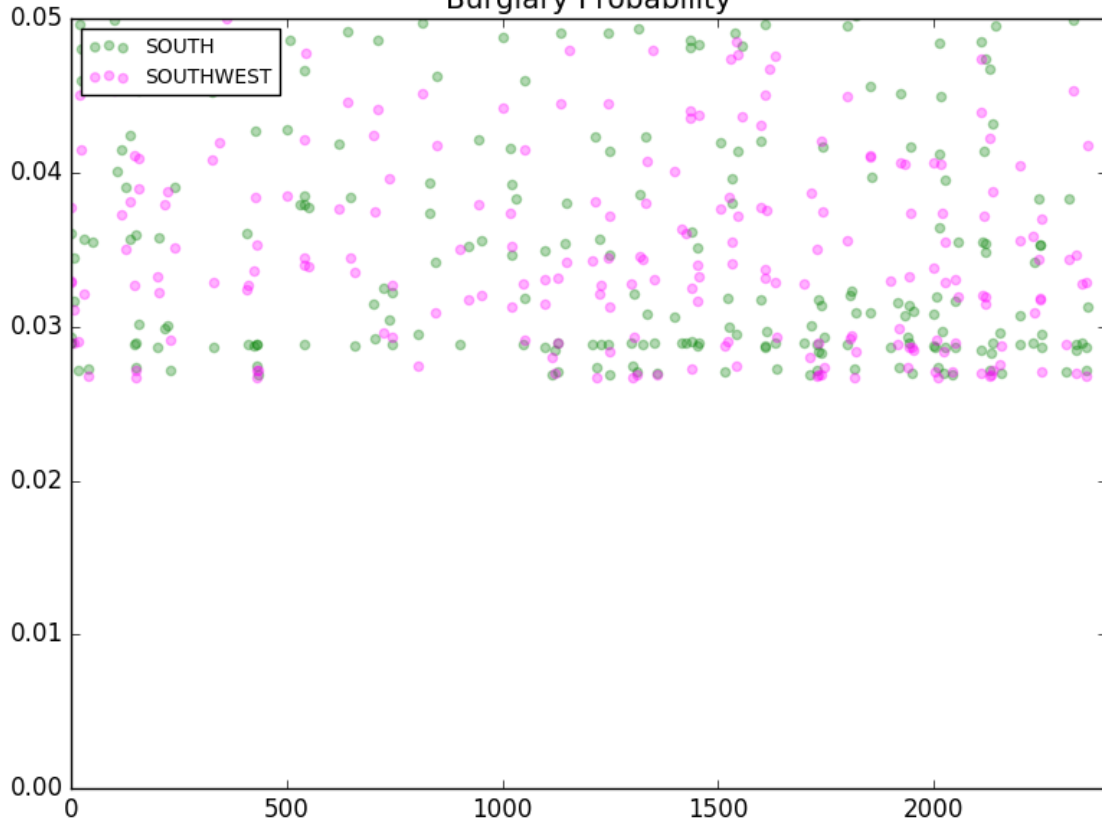


Figure 9
Assault Probability

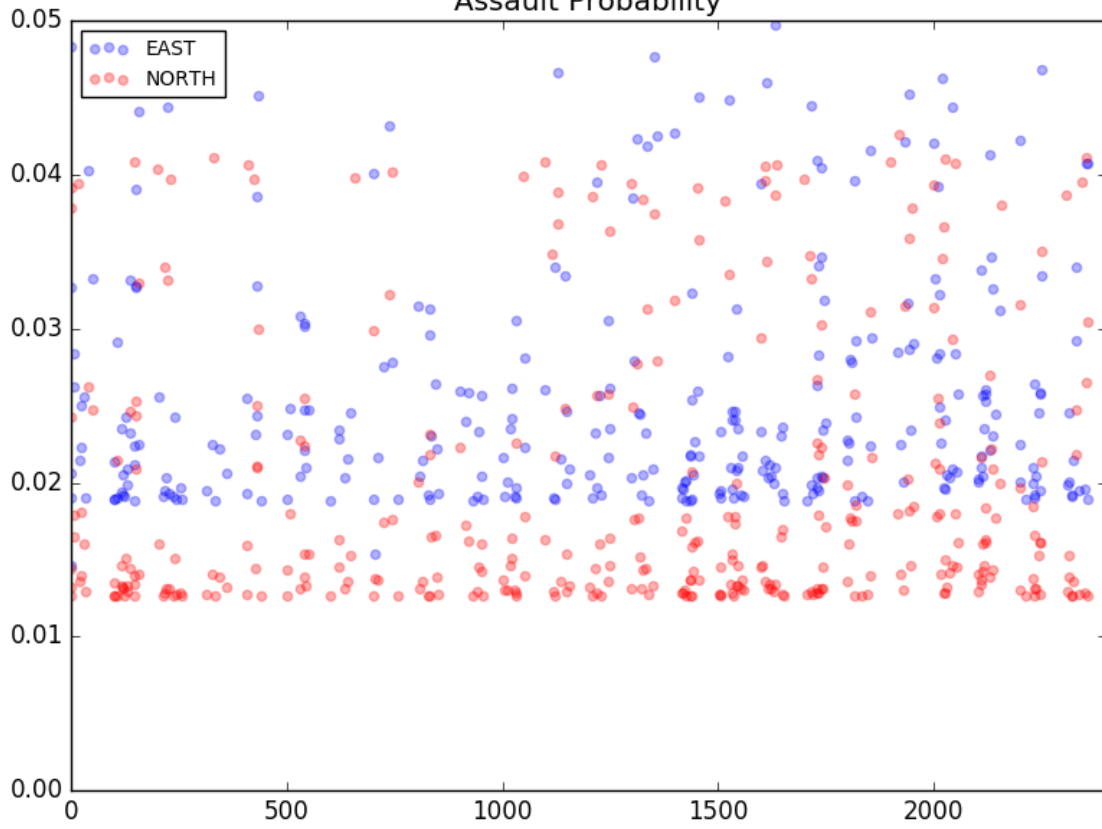


Figure 10
Assault Probability

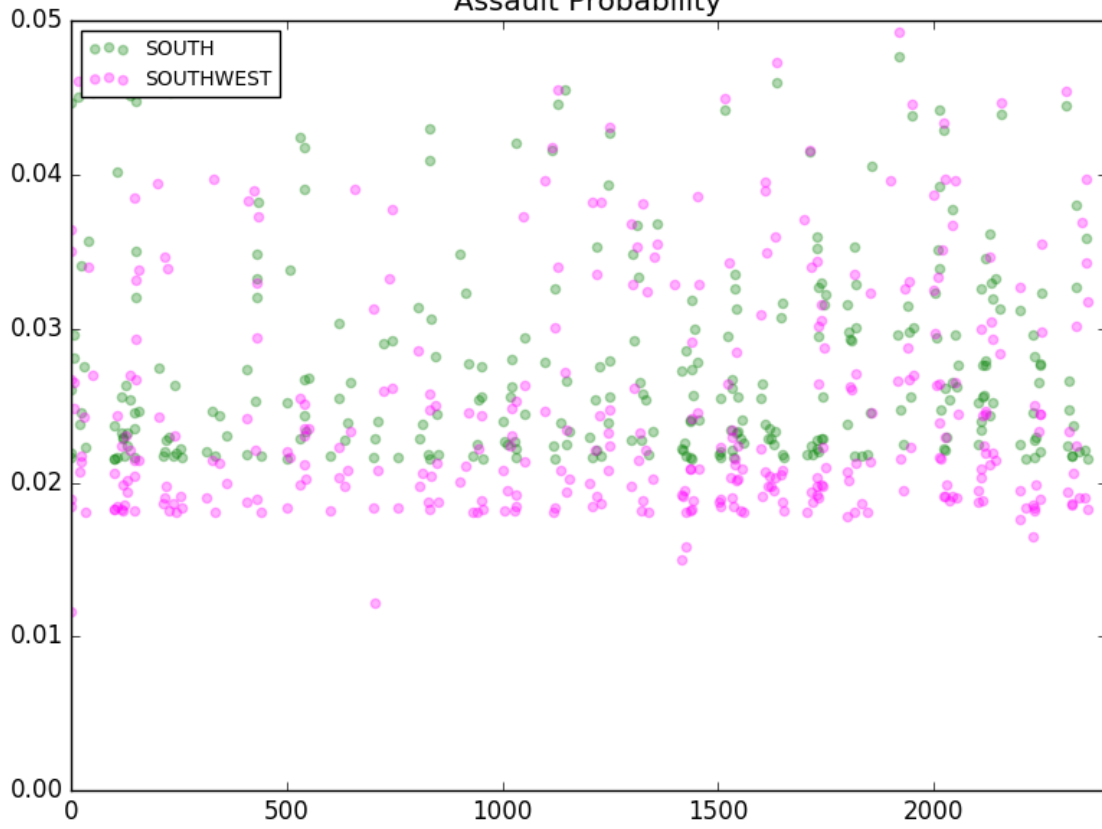


Figure 11
Misdemeanor Probability

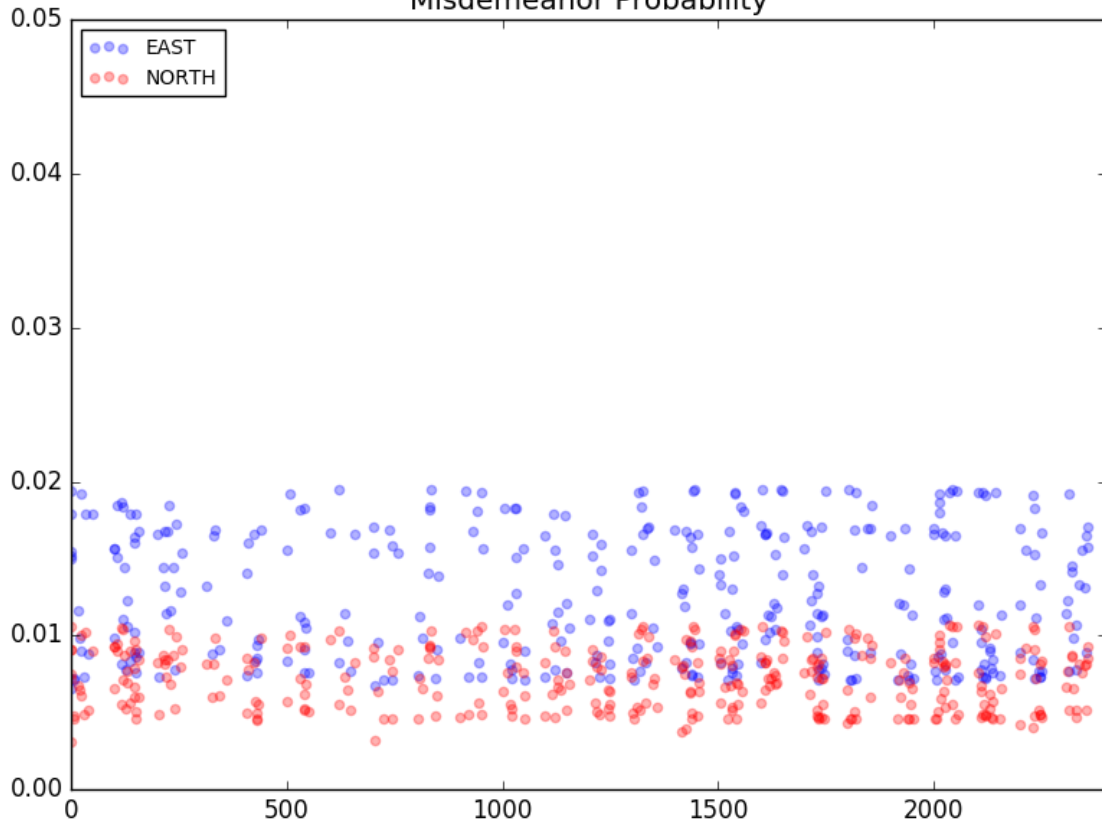


Figure 12
Misdemeanor Probability

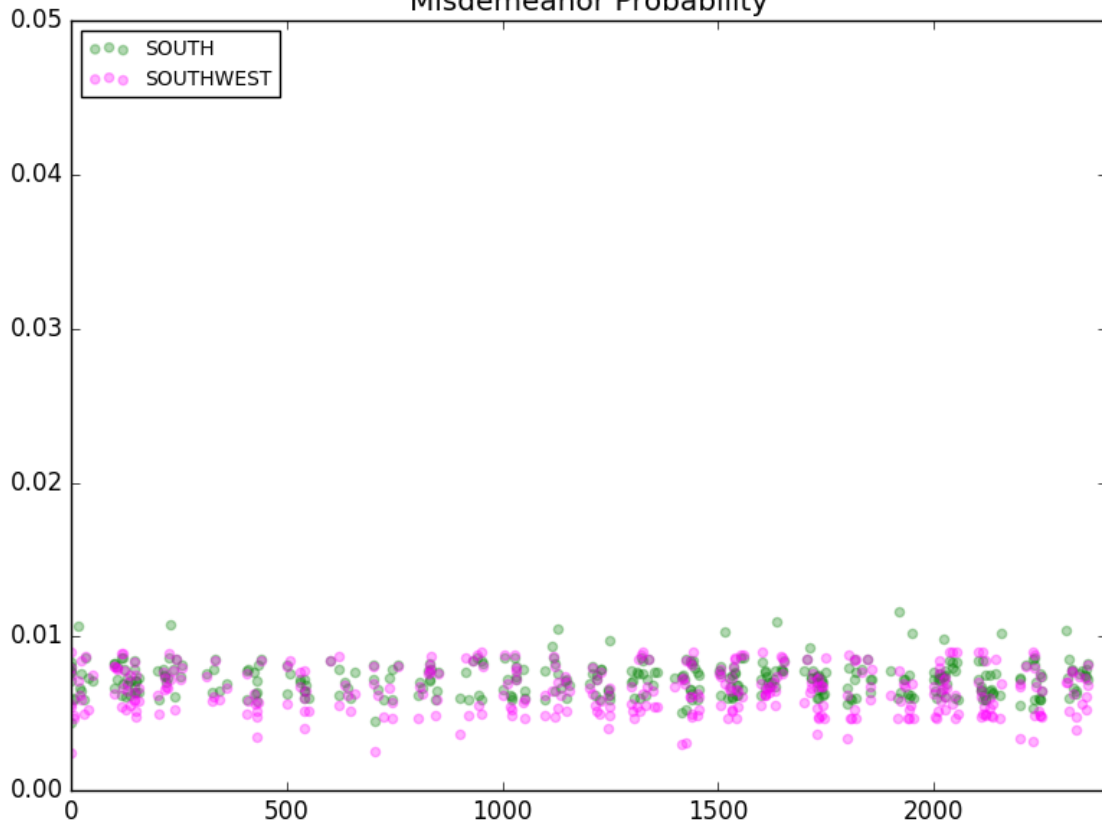


Figure 13
Sex Offense Probability

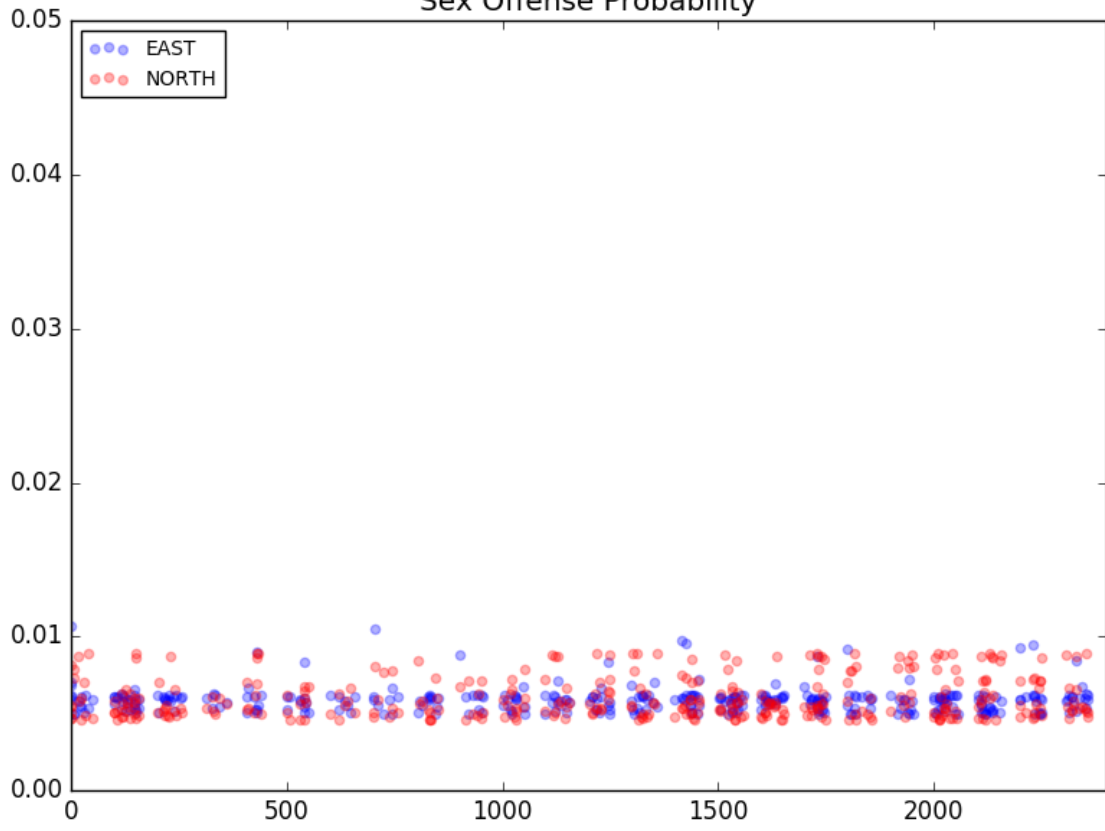


Figure 14
Sex Offense Probability

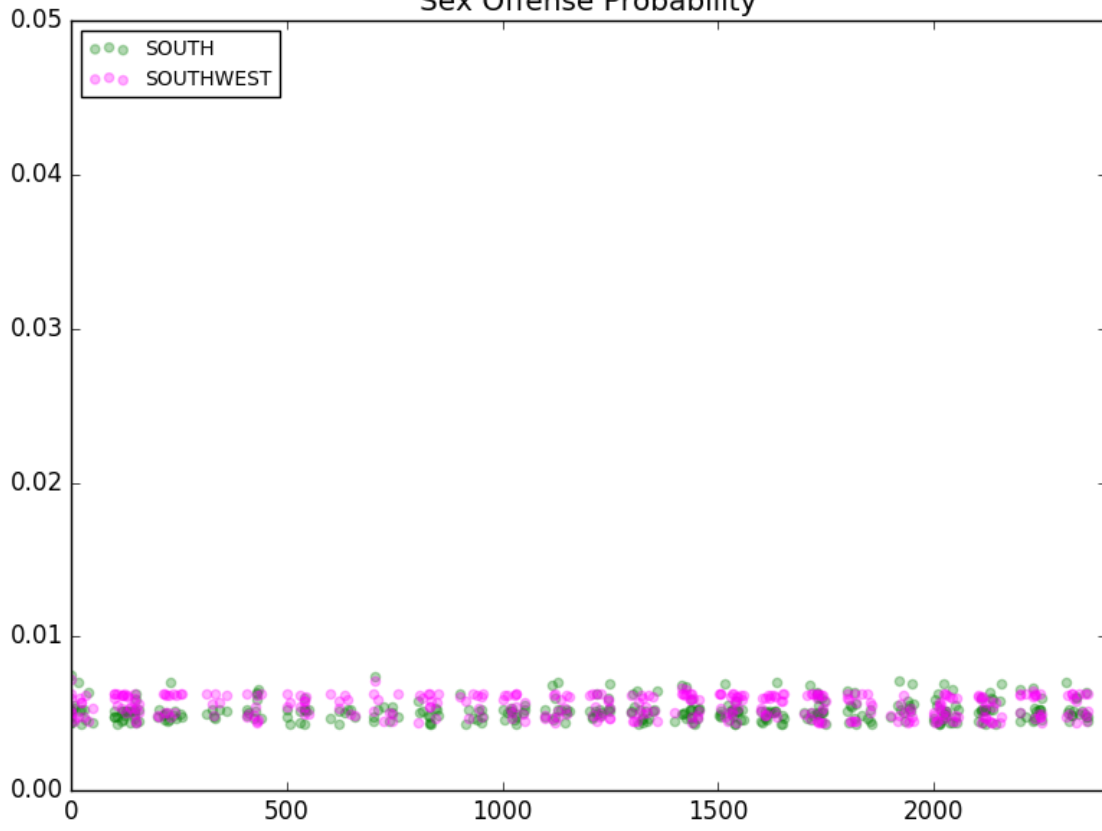


Figure 15
Car Prowl Probability

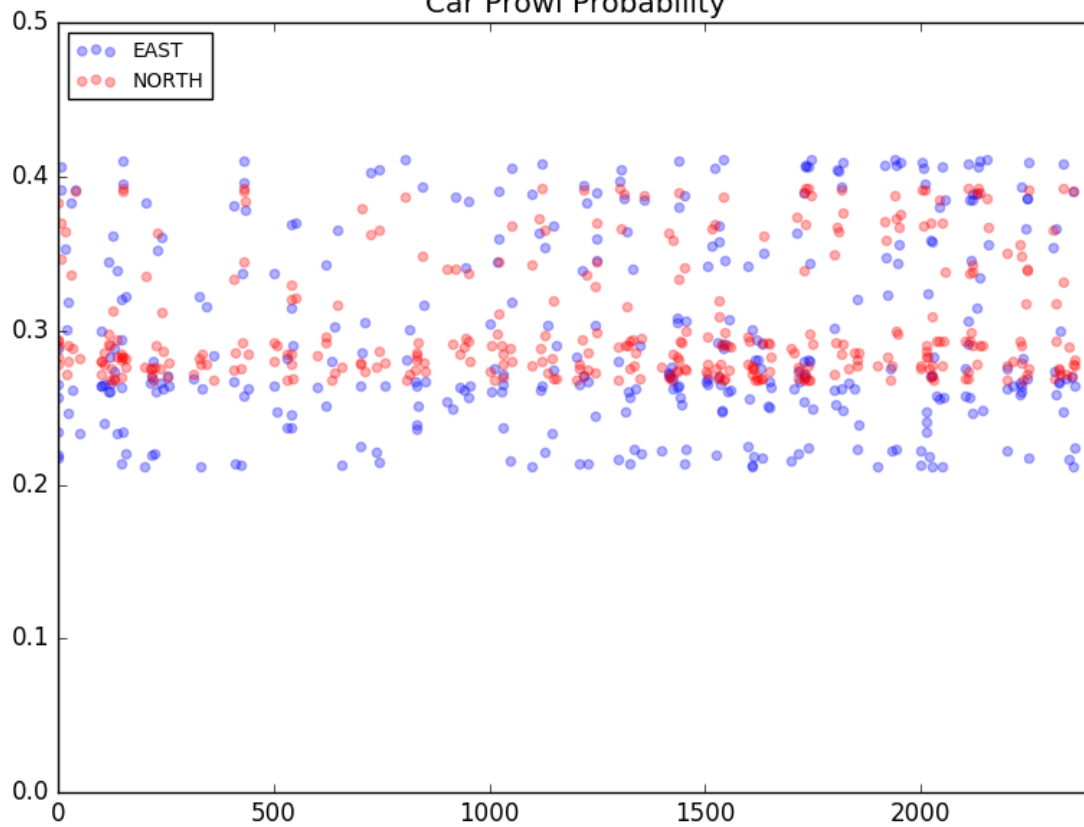


Figure 16
Car Prowl Probability

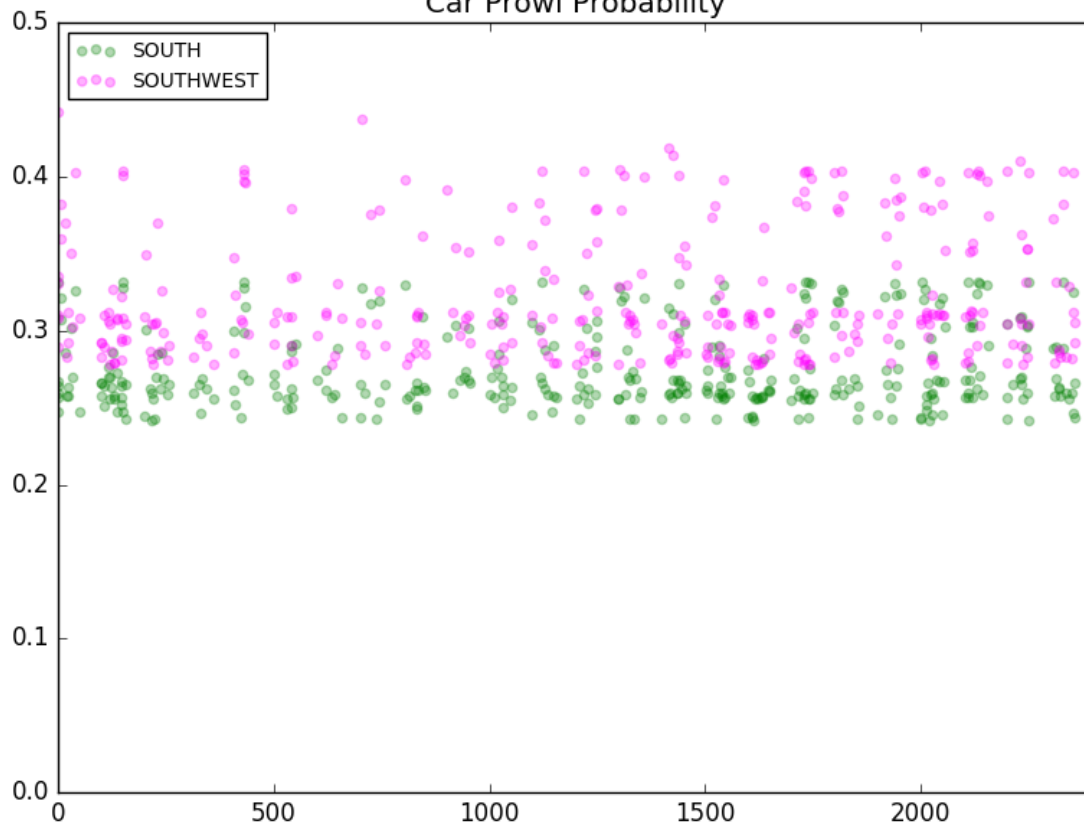


Figure 17
Theft Probability

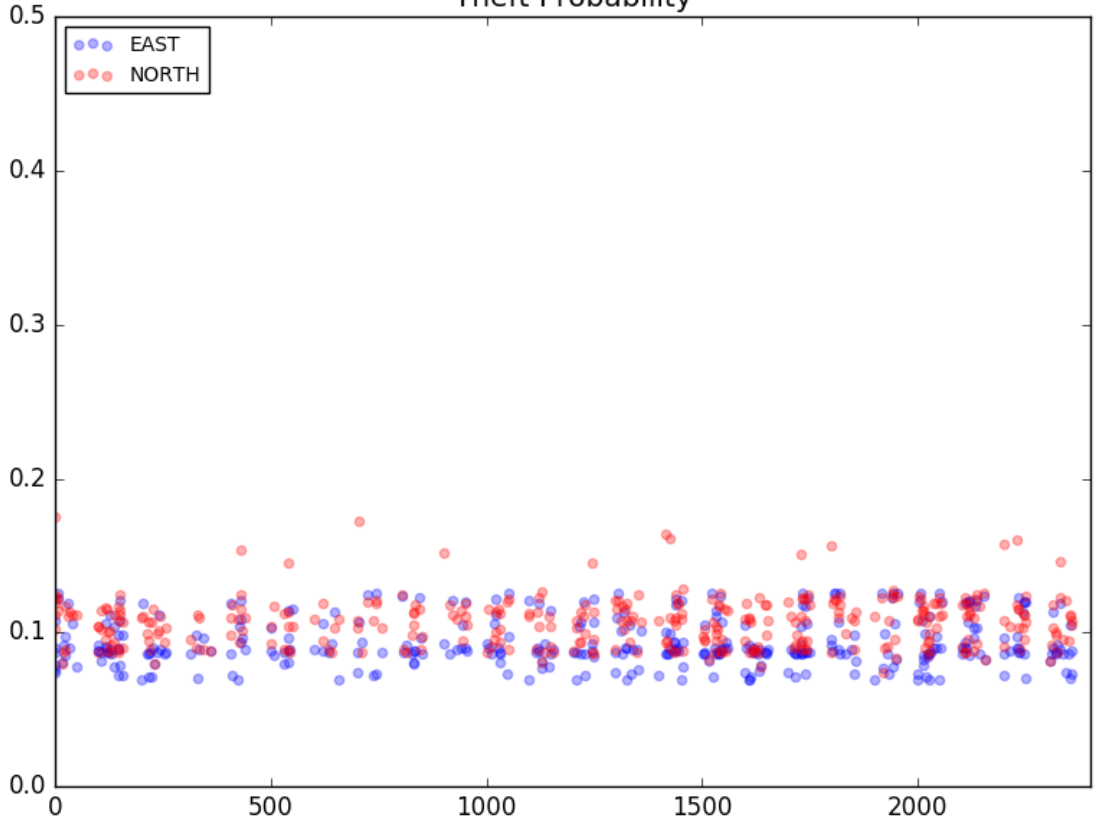


Figure 18
Theft Probability

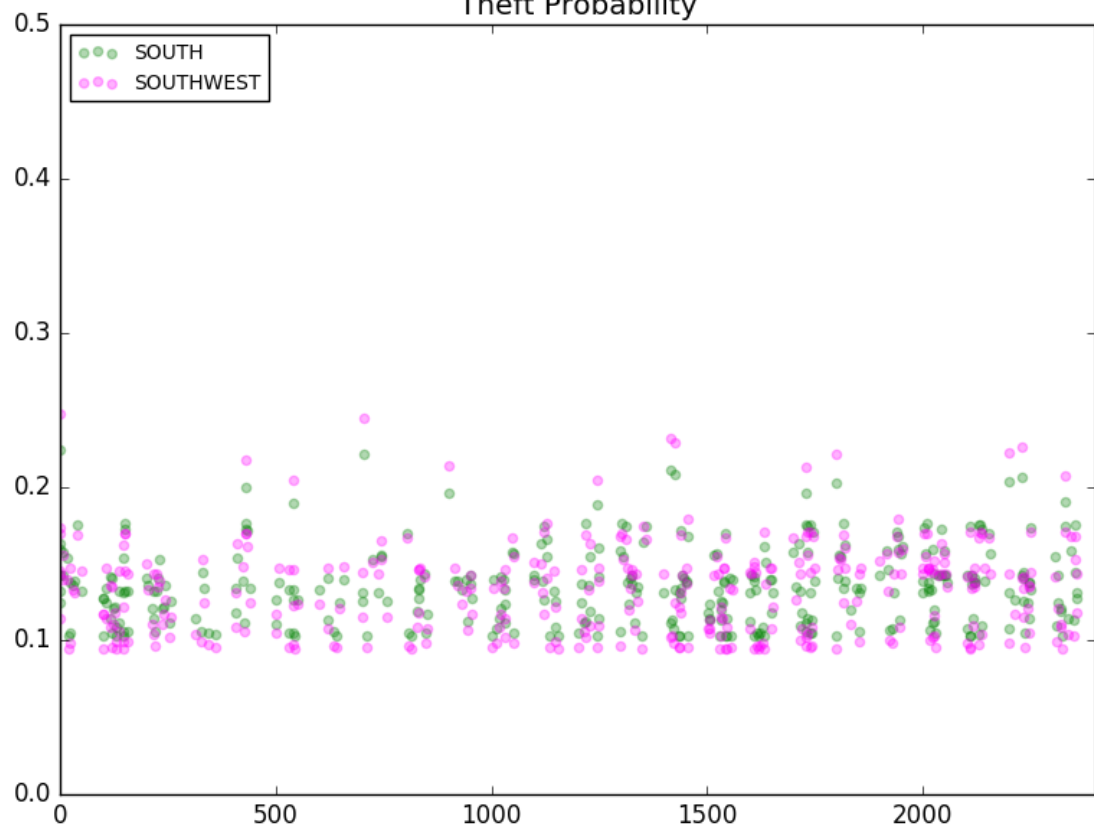


Figure 19

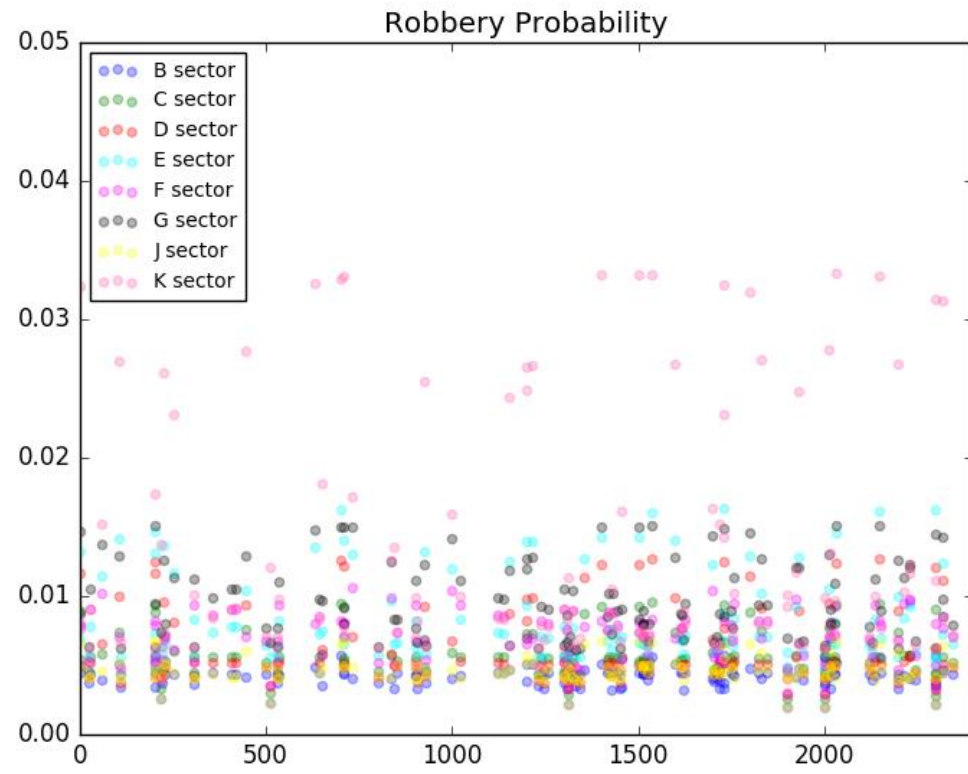


Figure 20

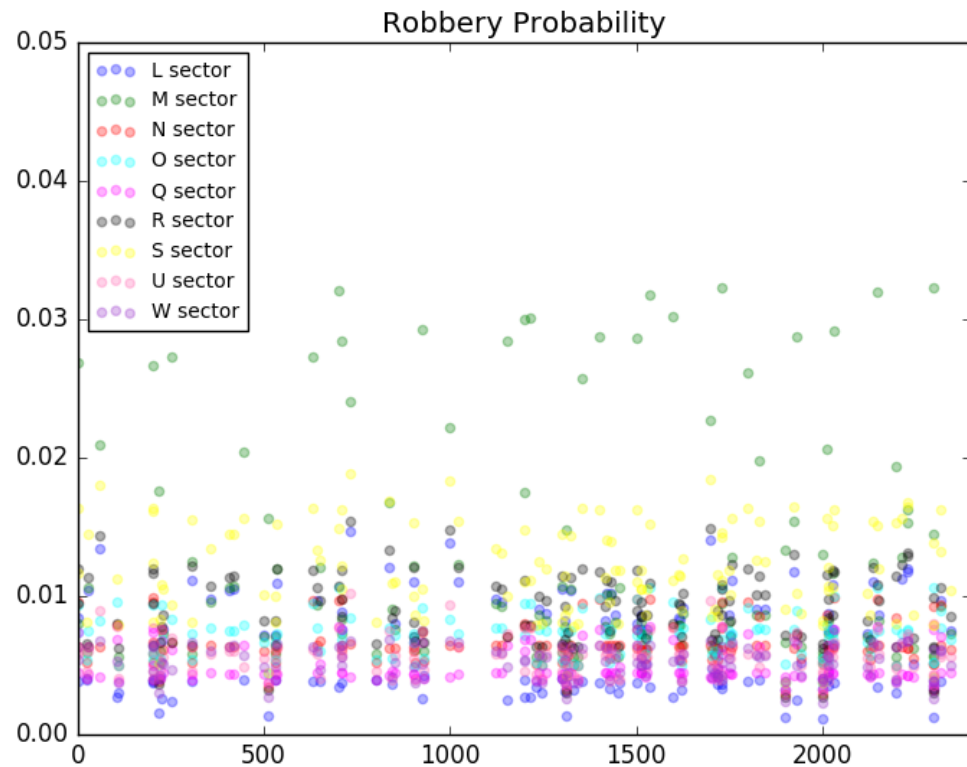


Figure 21

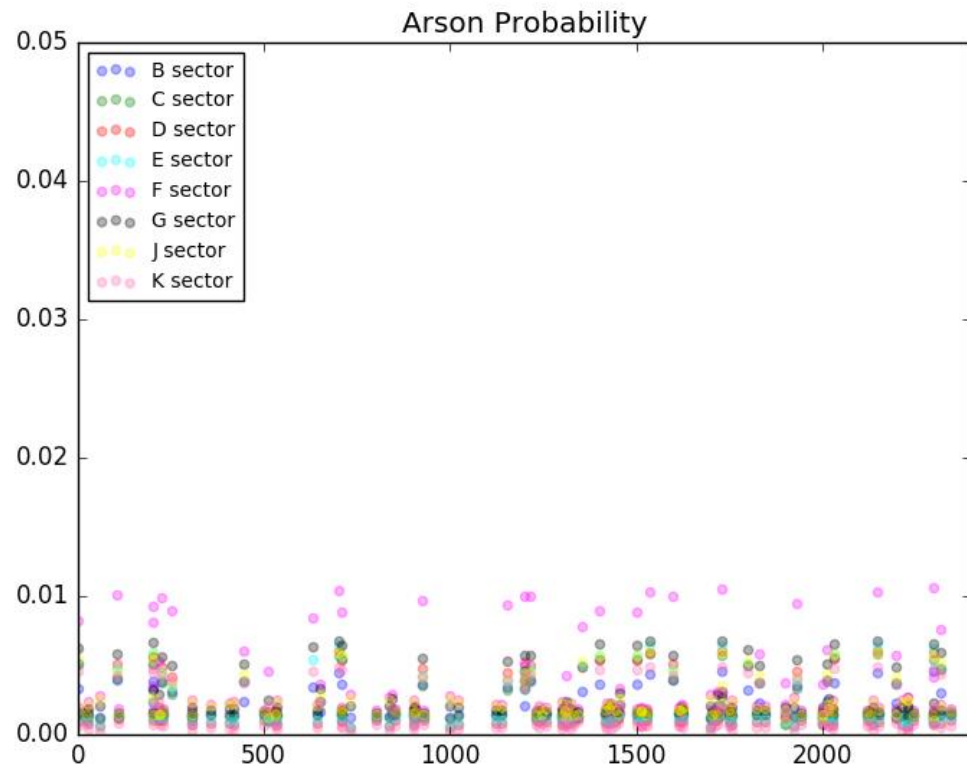


Figure 22

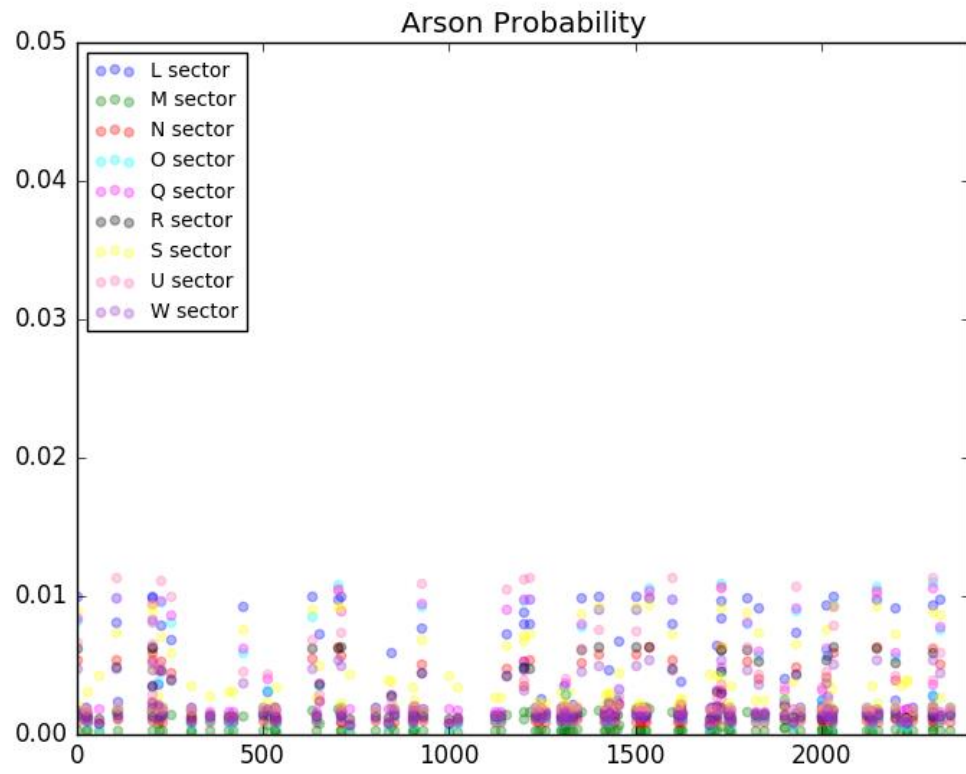


Figure 23

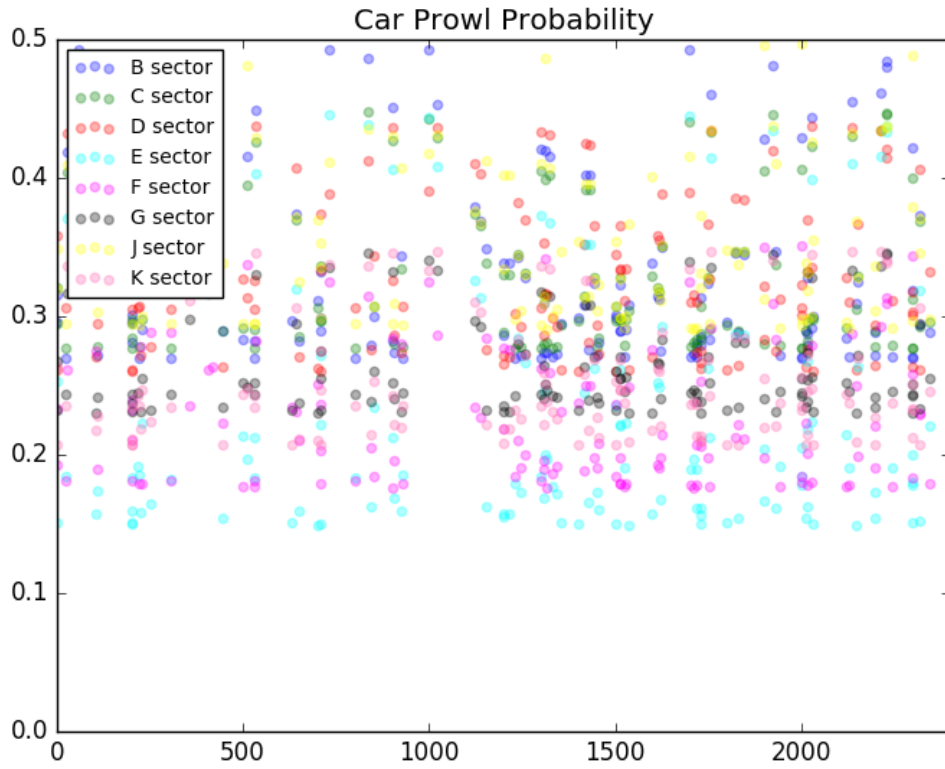


Figure 24

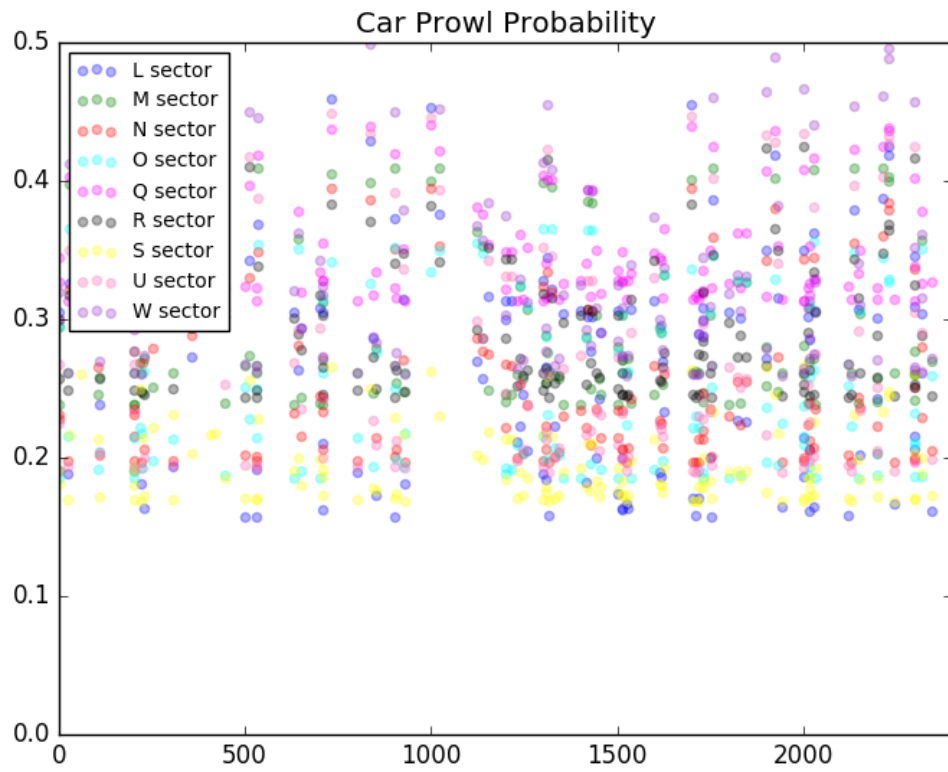


Figure 25

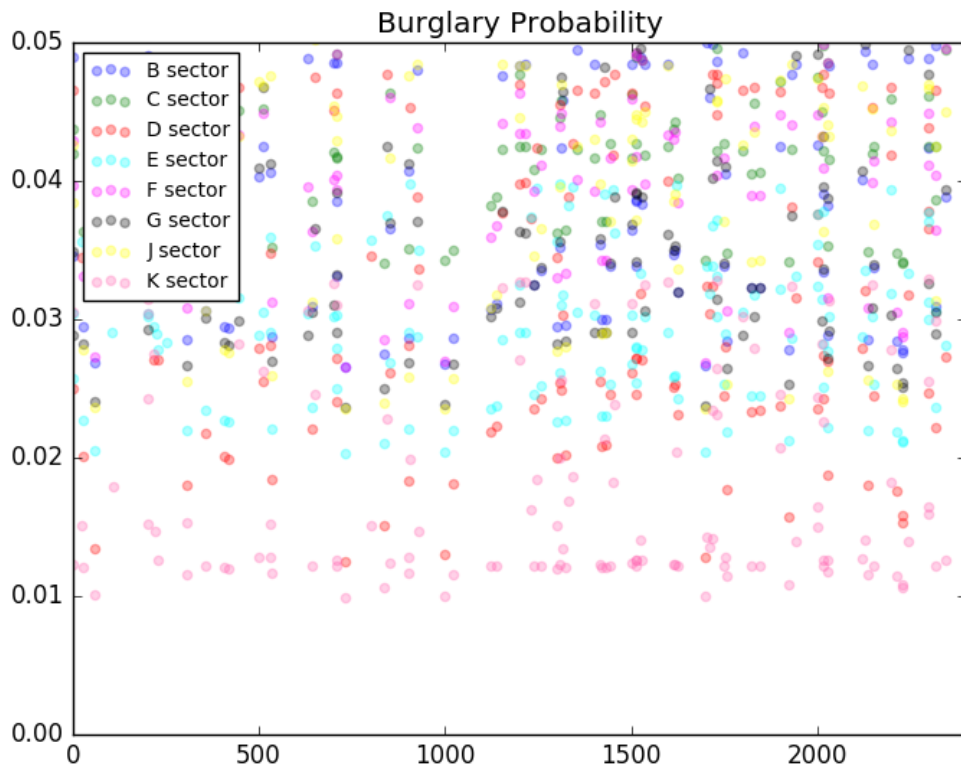


Figure 26

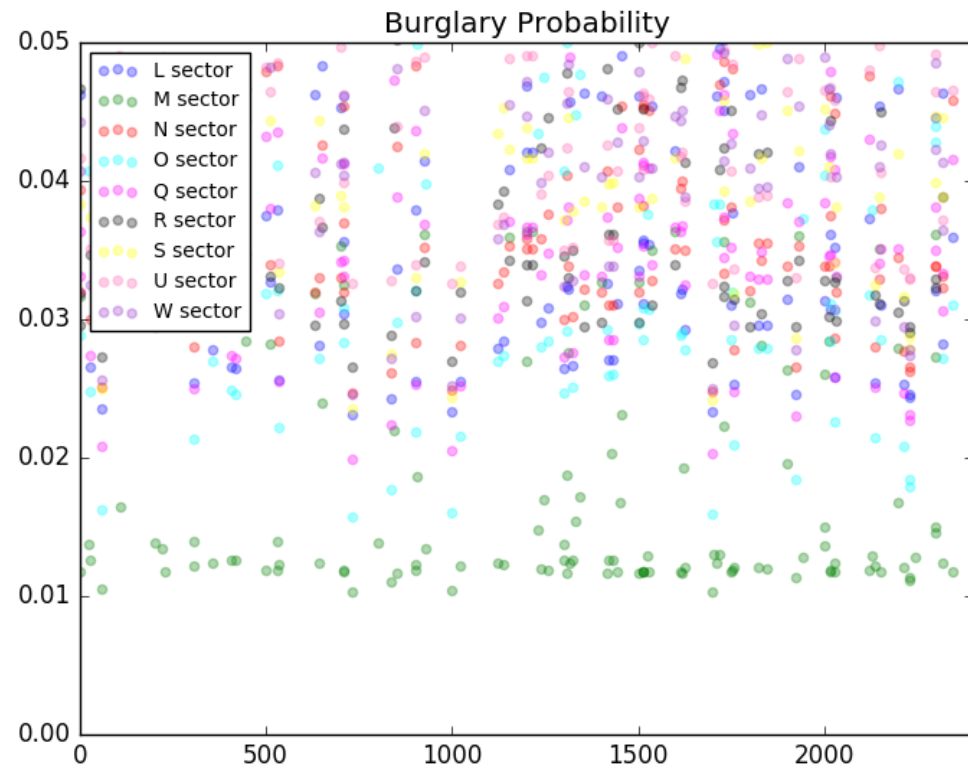


Figure 27

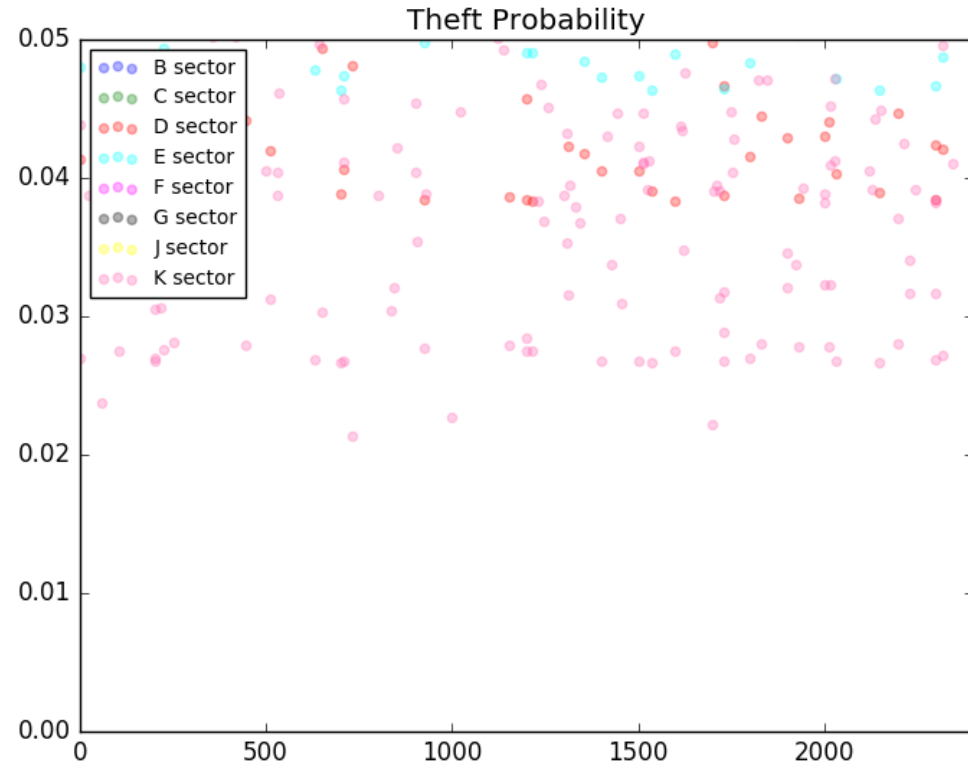


Figure 28

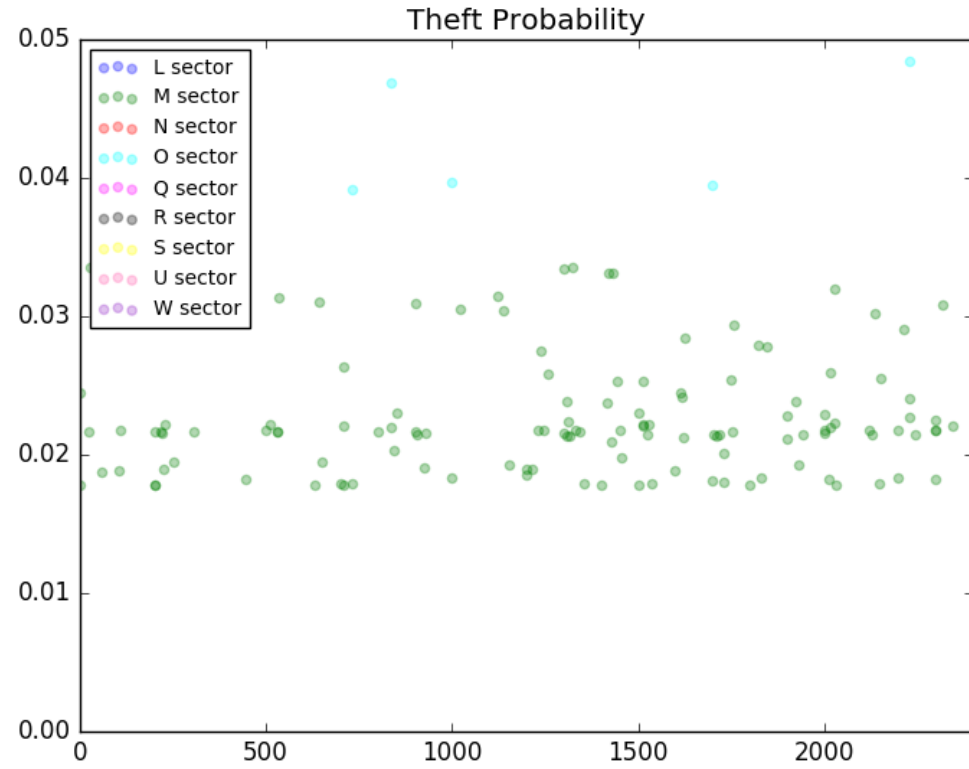


Figure 29

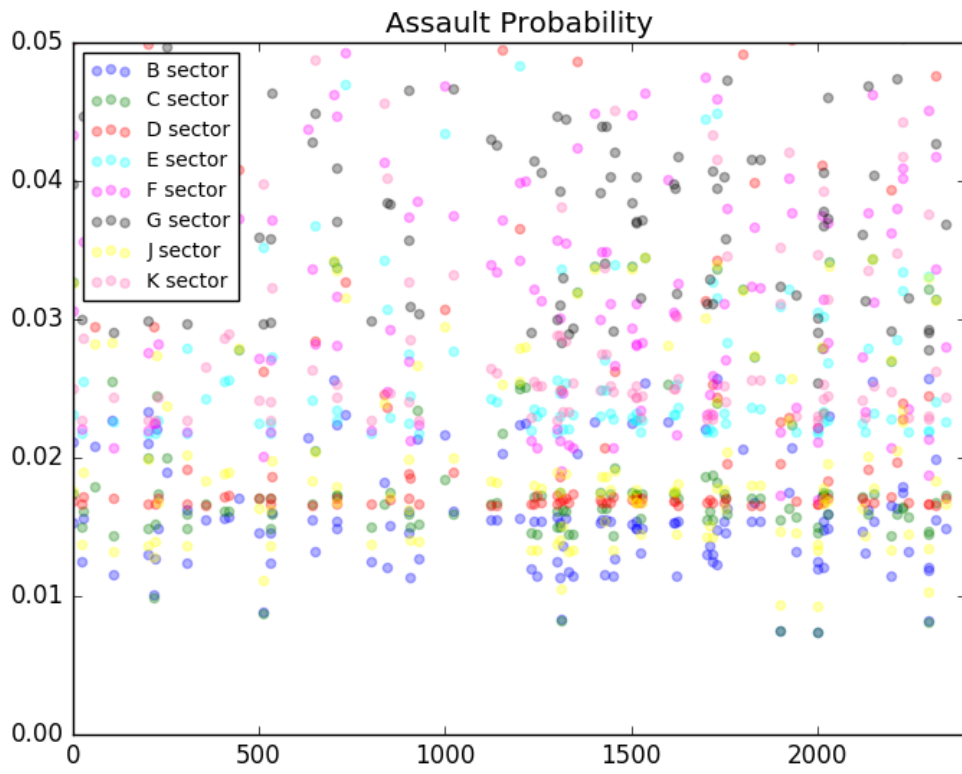


Figure 30

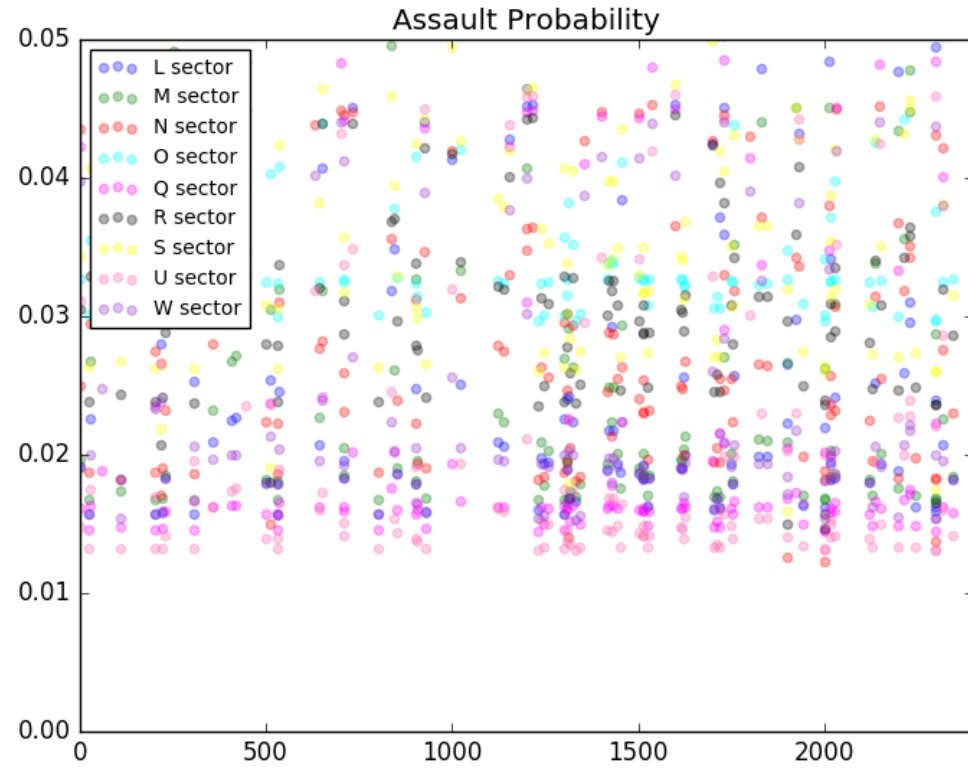


Figure 31

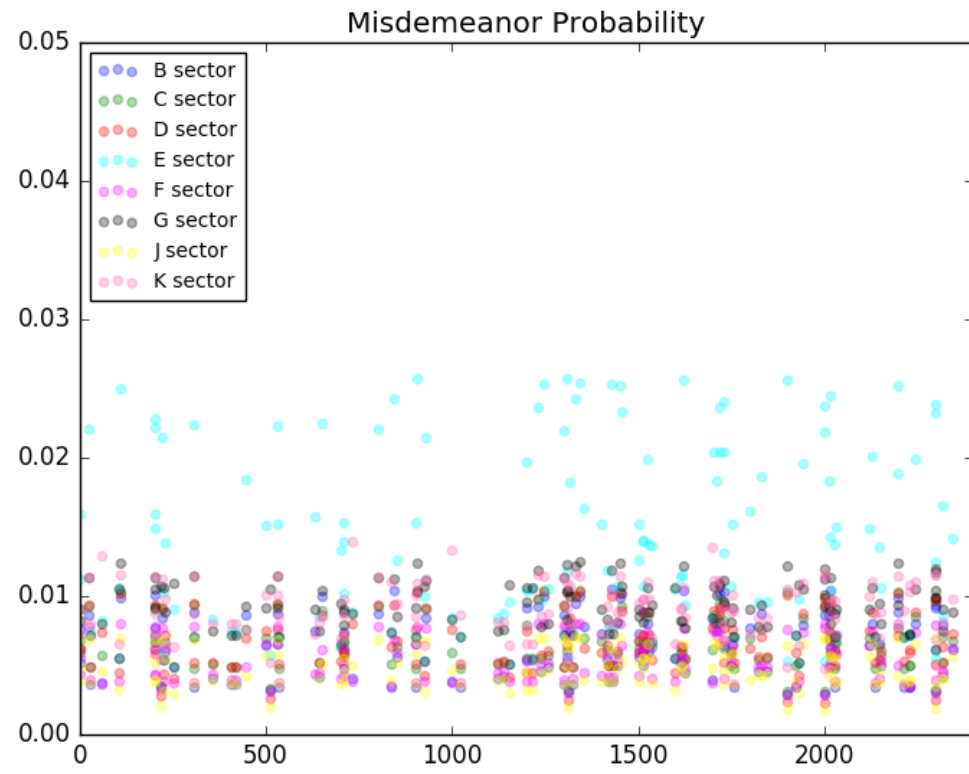


Figure 32

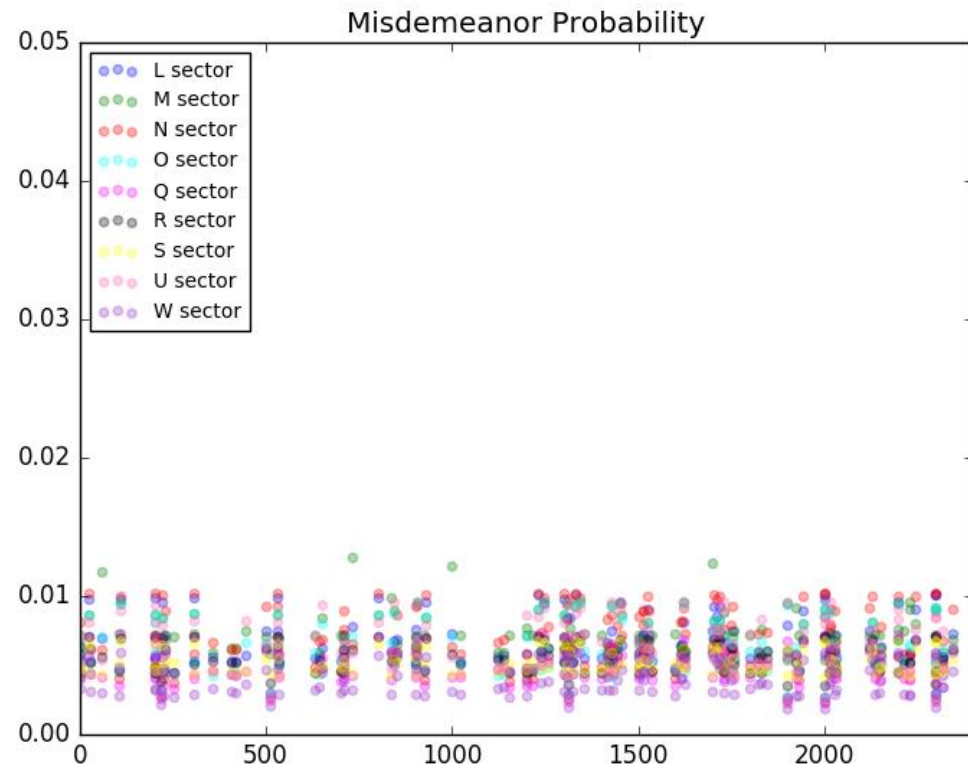


Figure 33

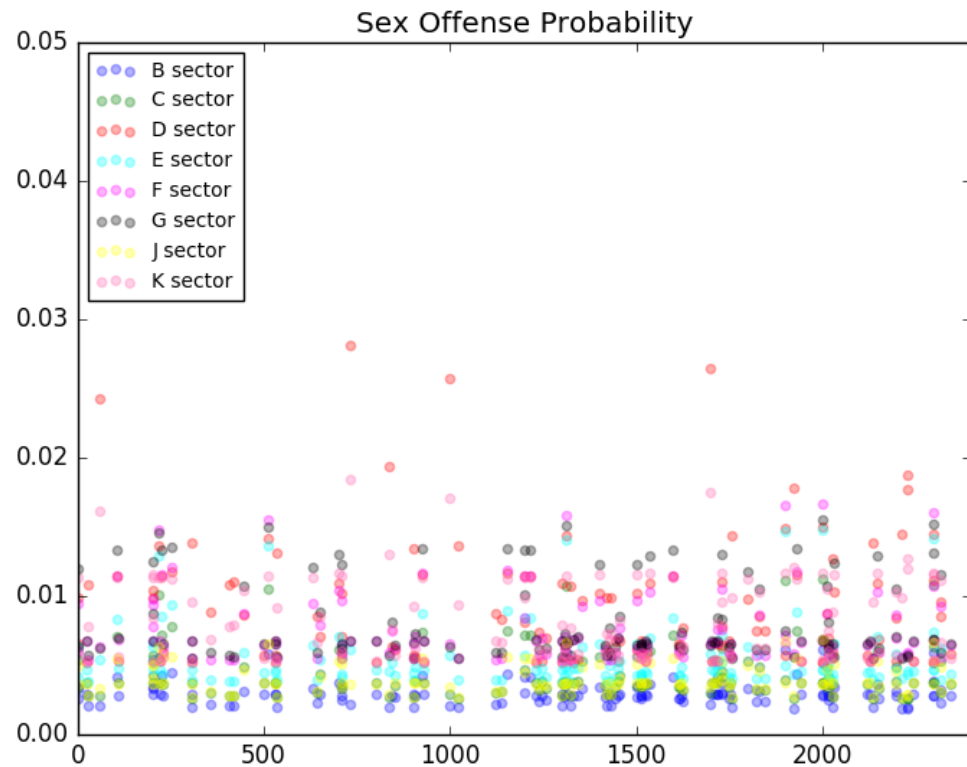


Figure 34

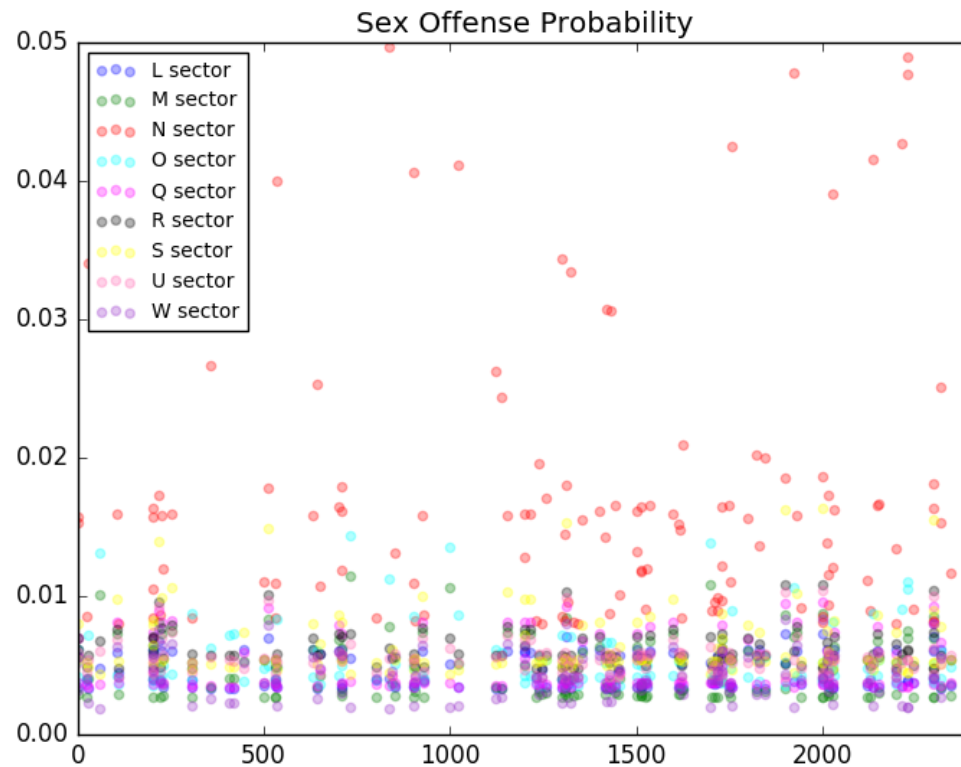


Figure 35

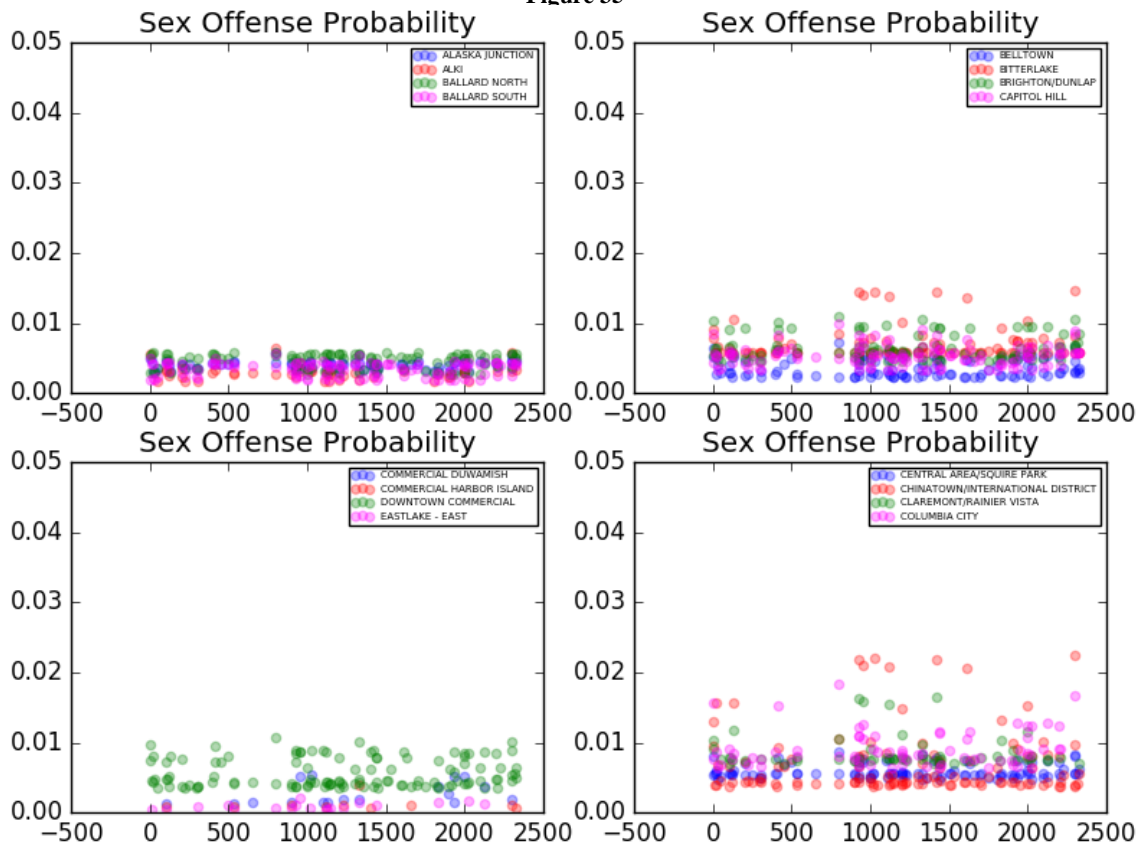


Figure 36

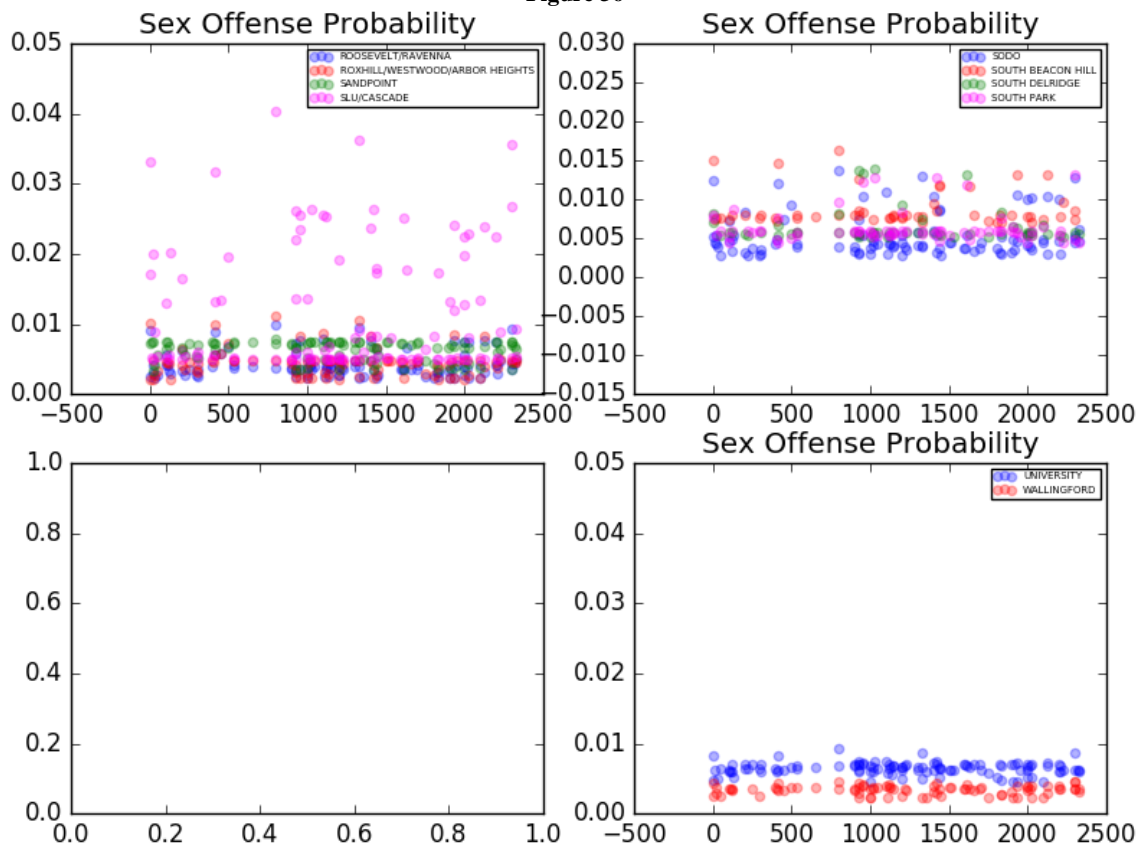


Figure 37

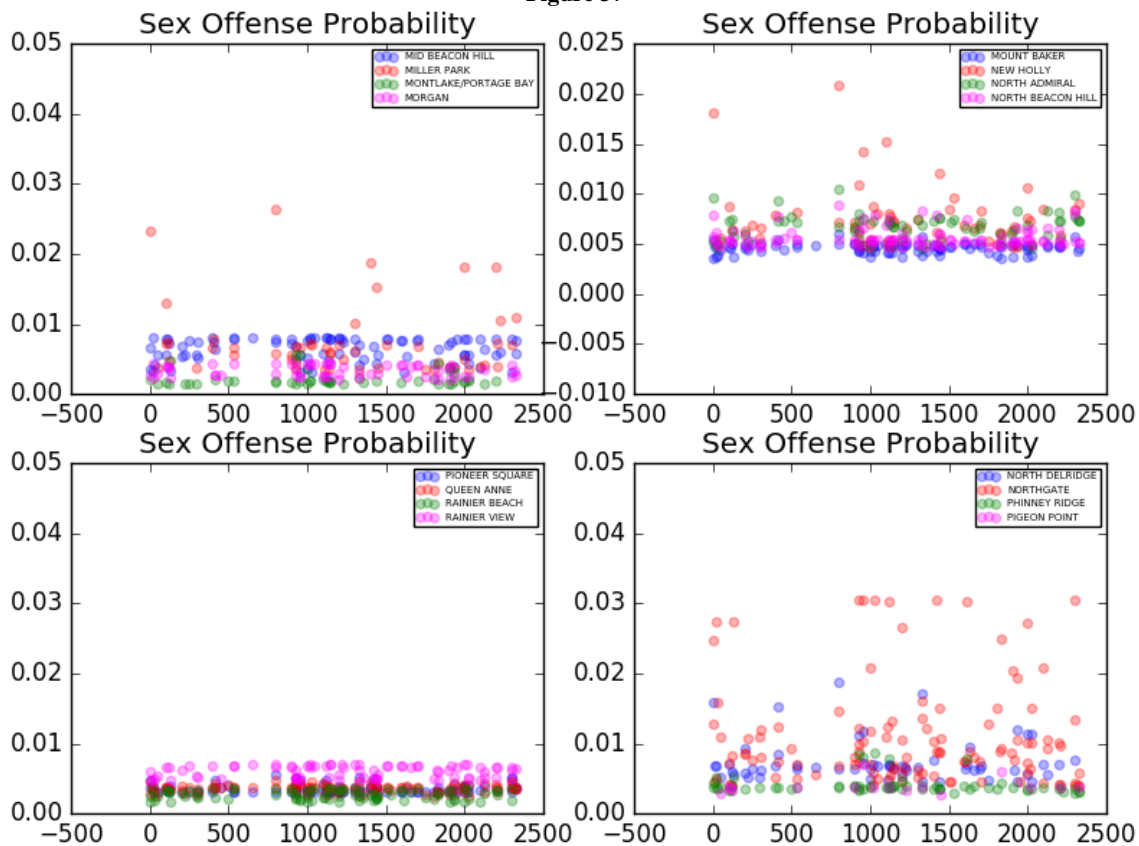


Figure 38

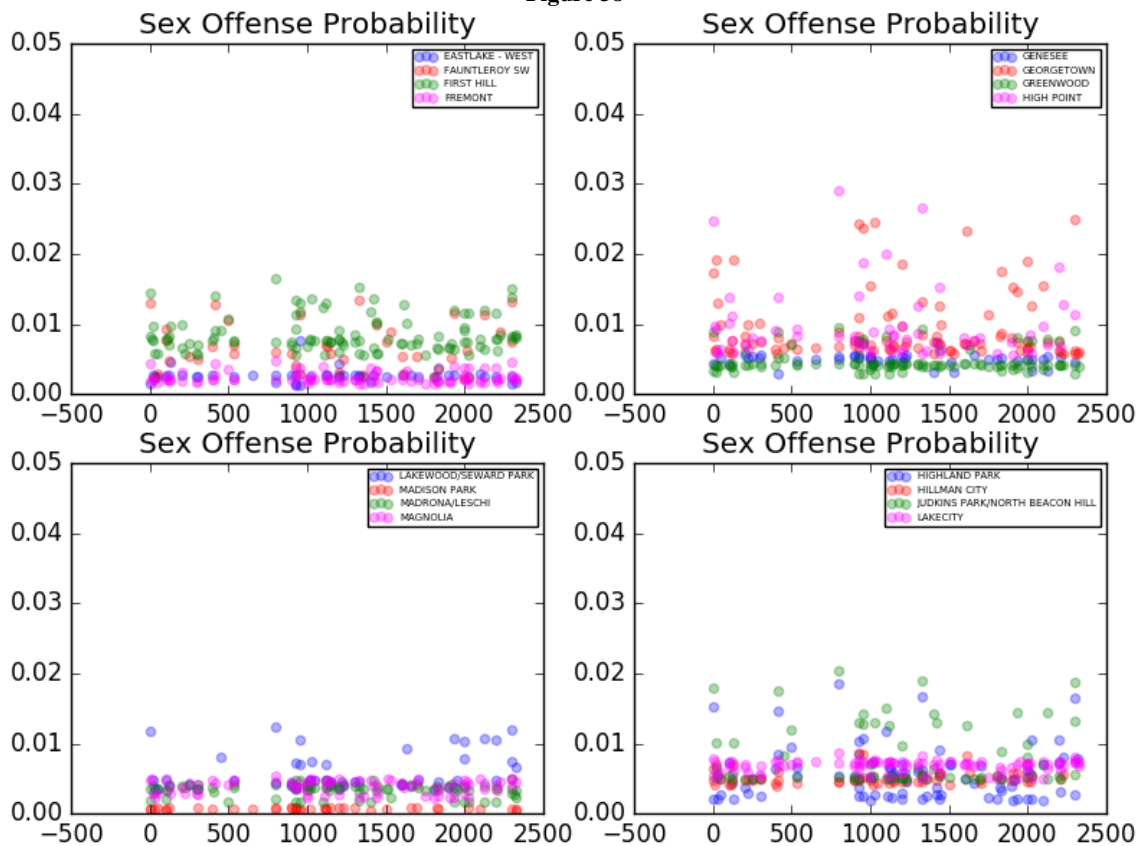


Figure 39

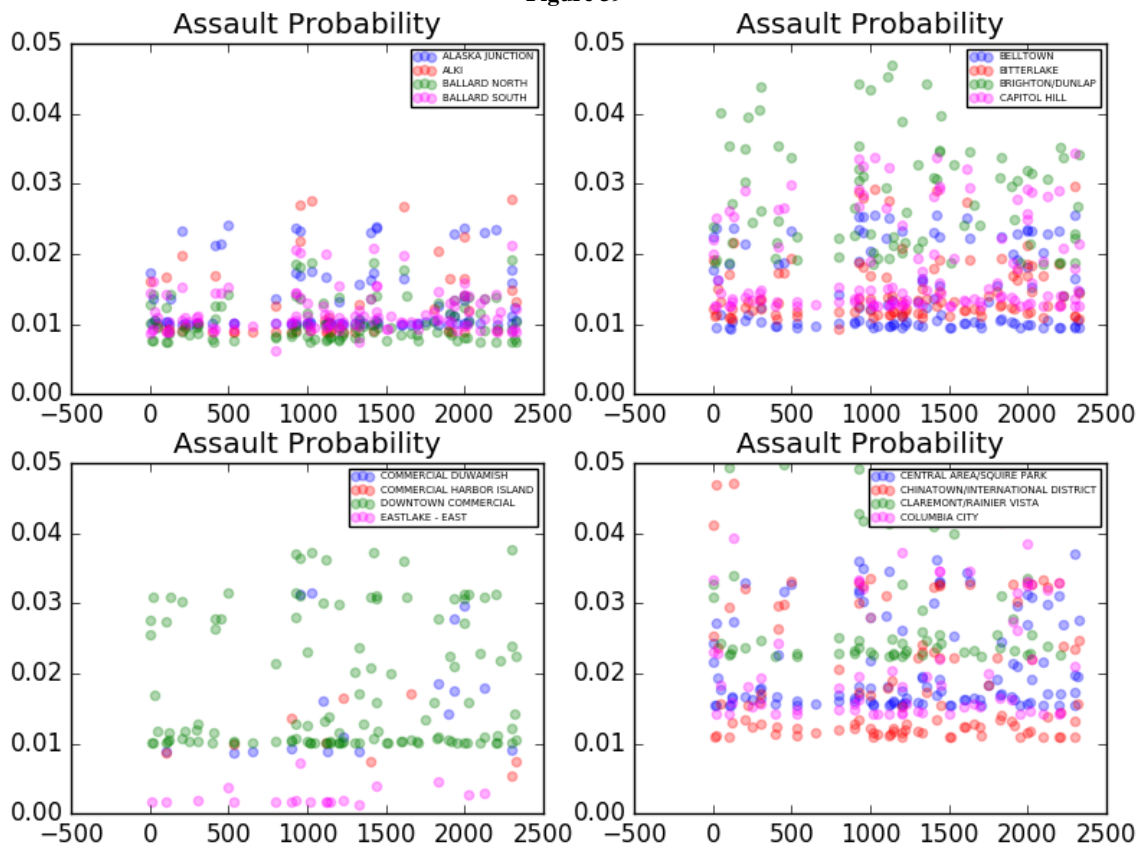


Figure 40

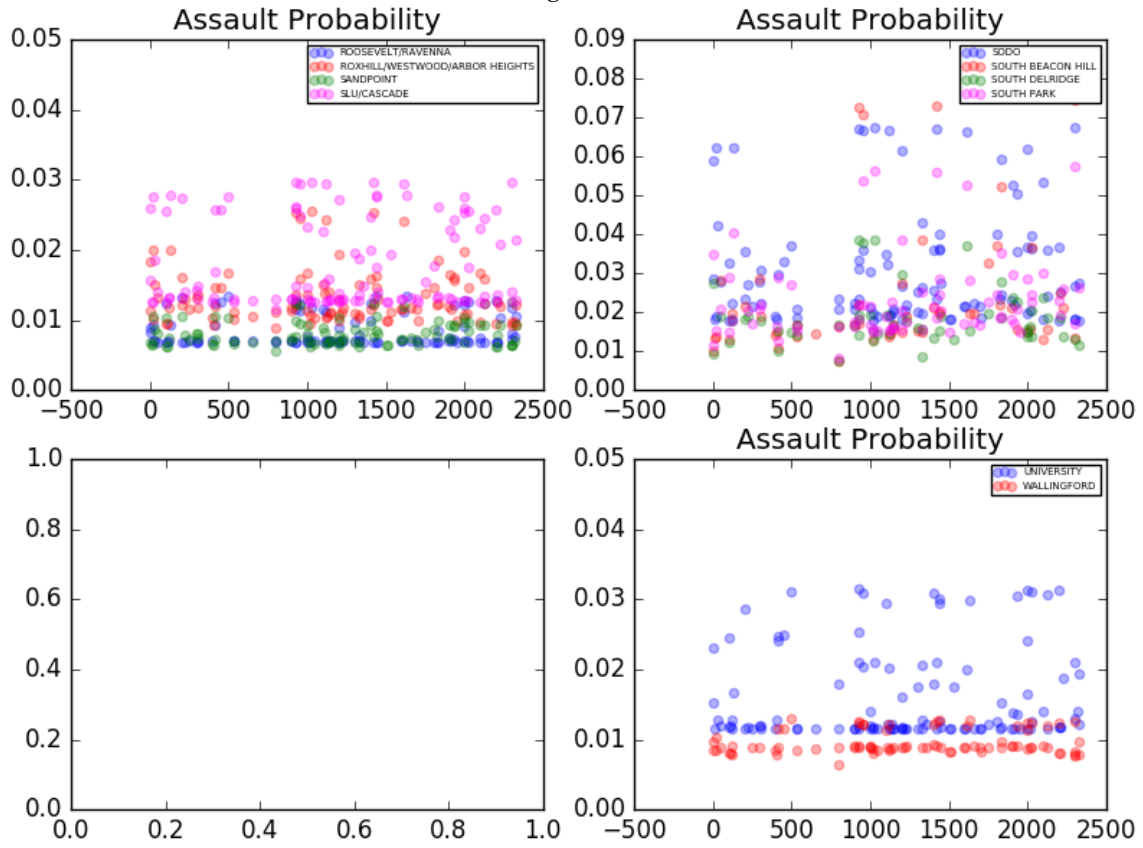


Figure 41

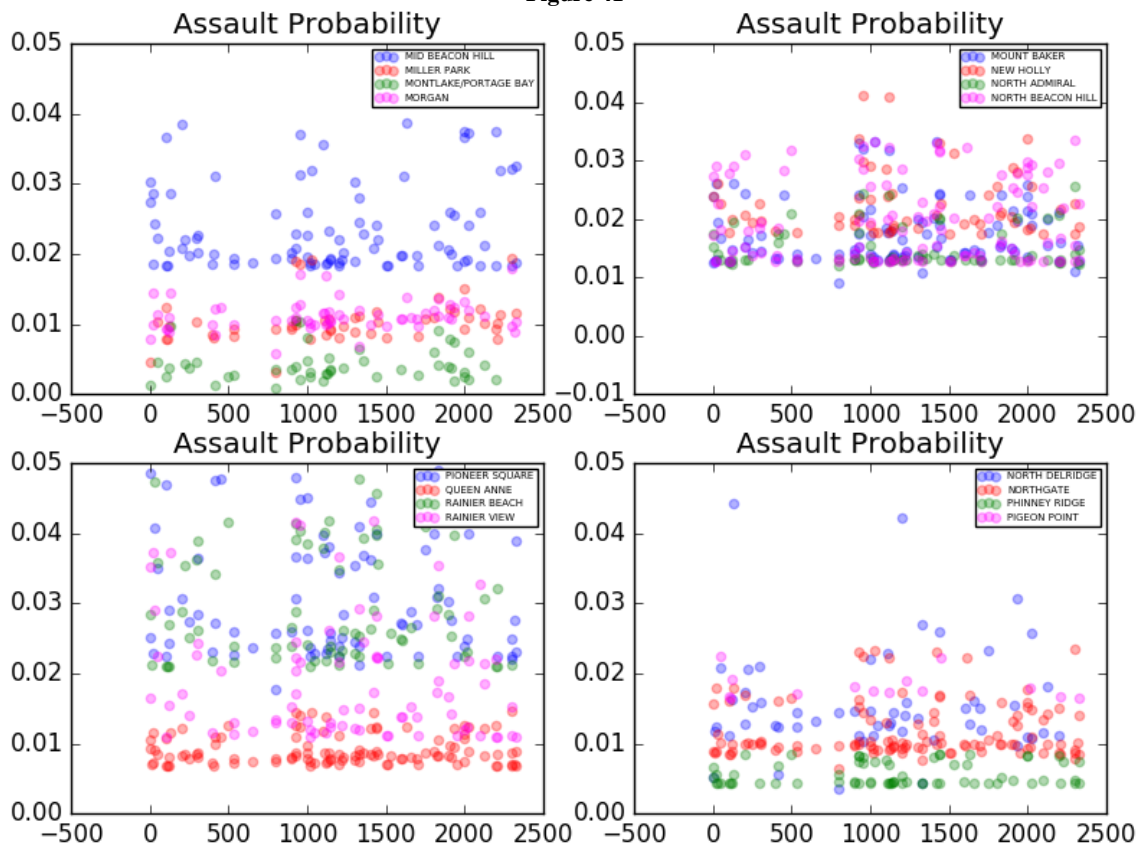


Figure 42

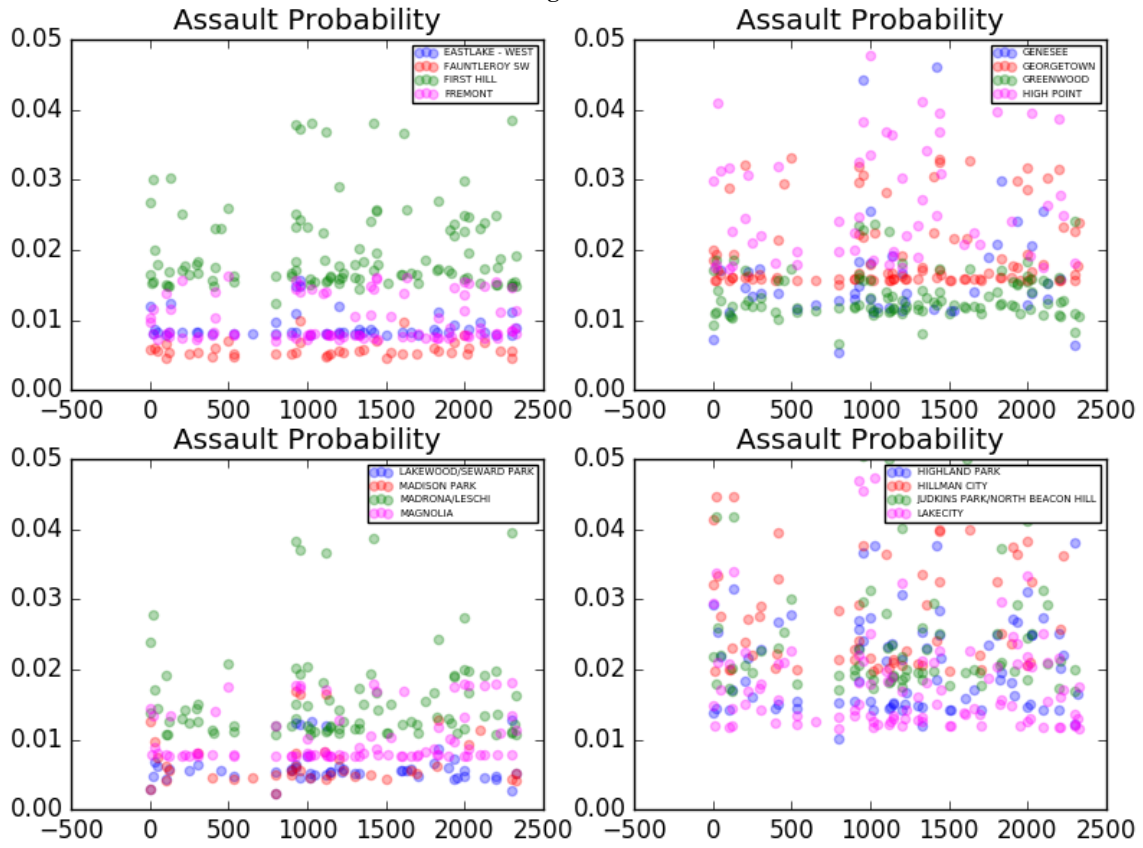


Figure 43

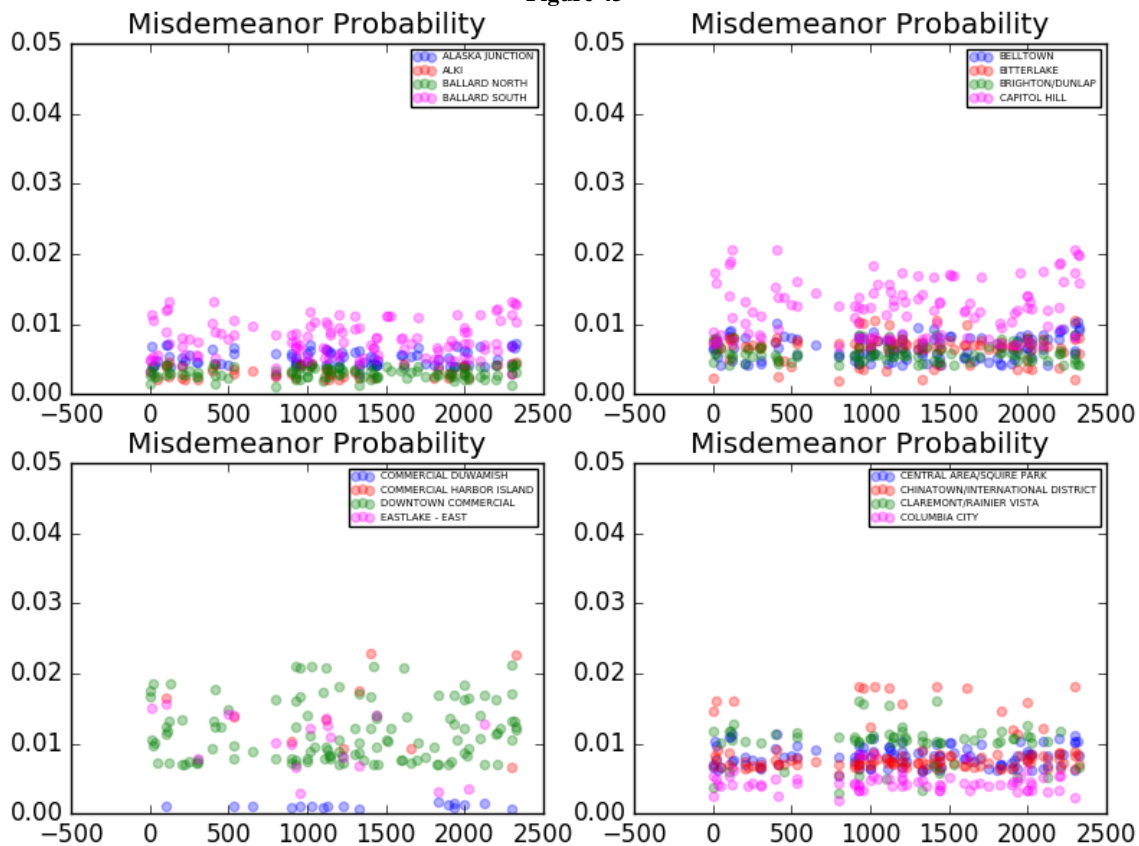


Figure 44

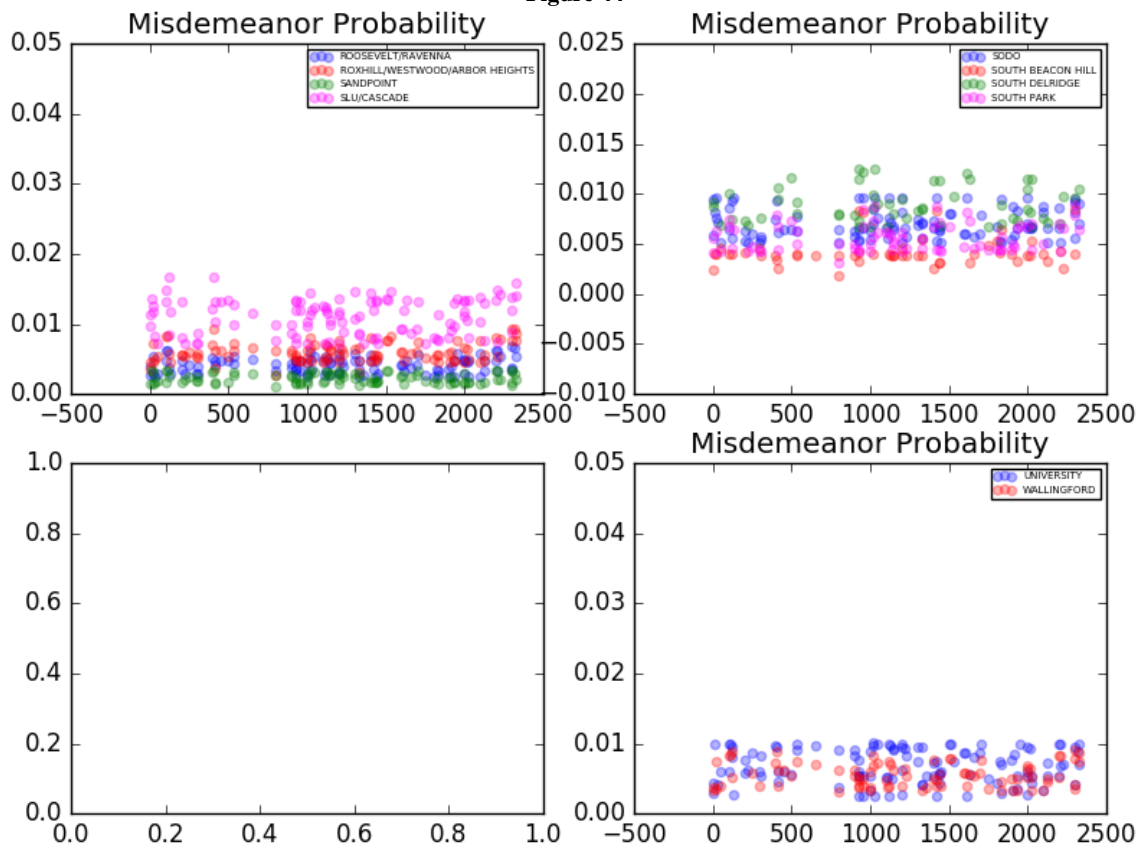


Figure 45

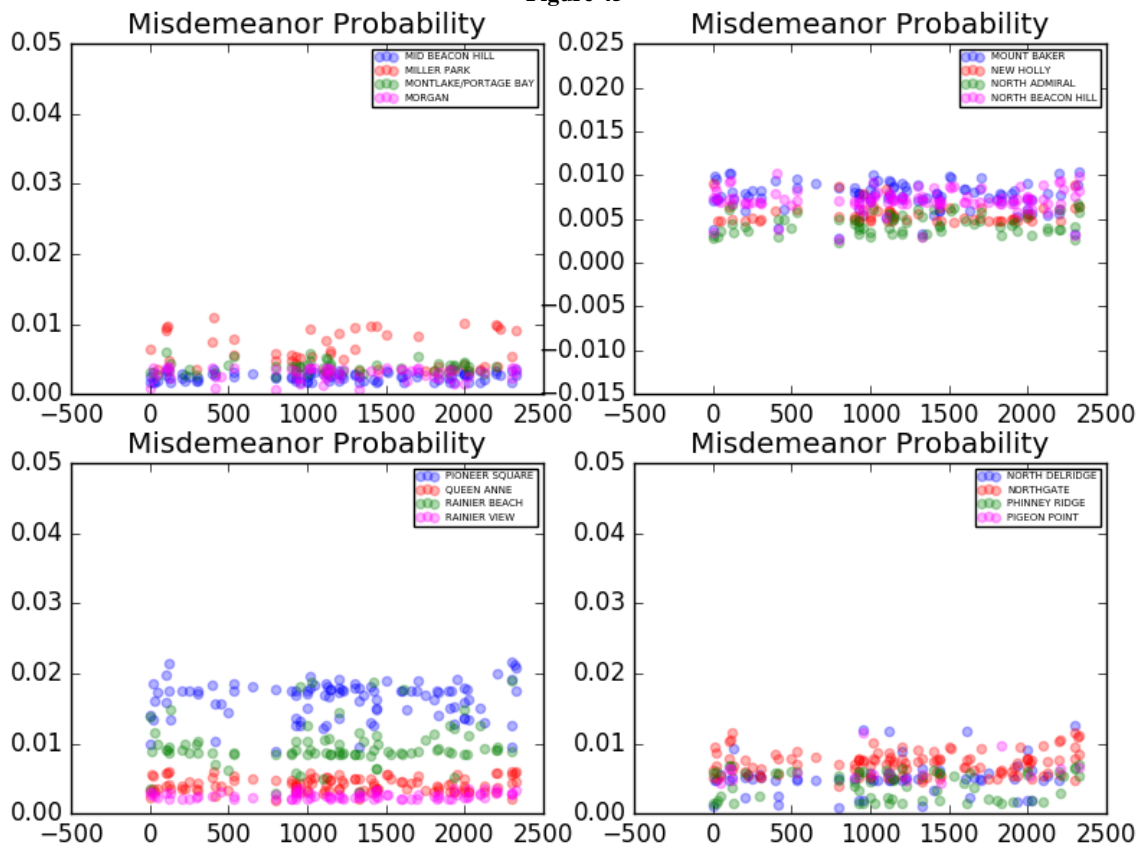


Figure 46

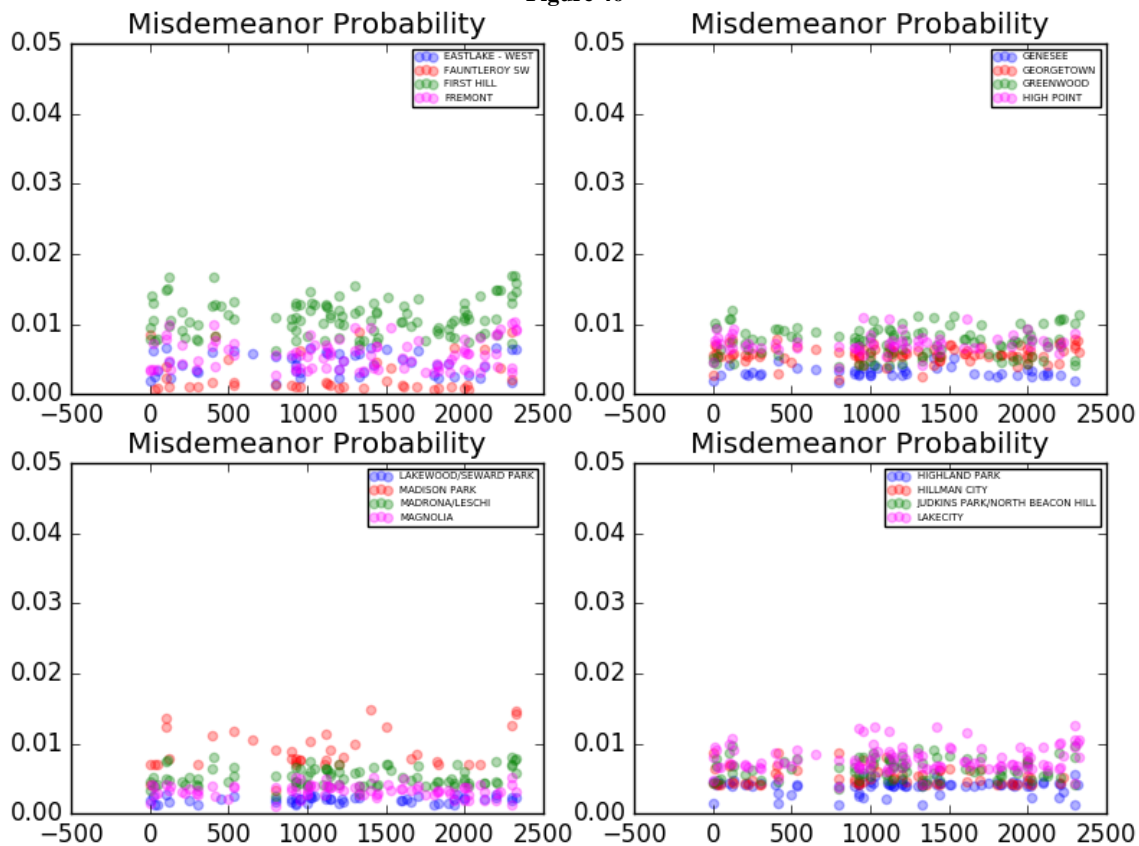


Figure 47

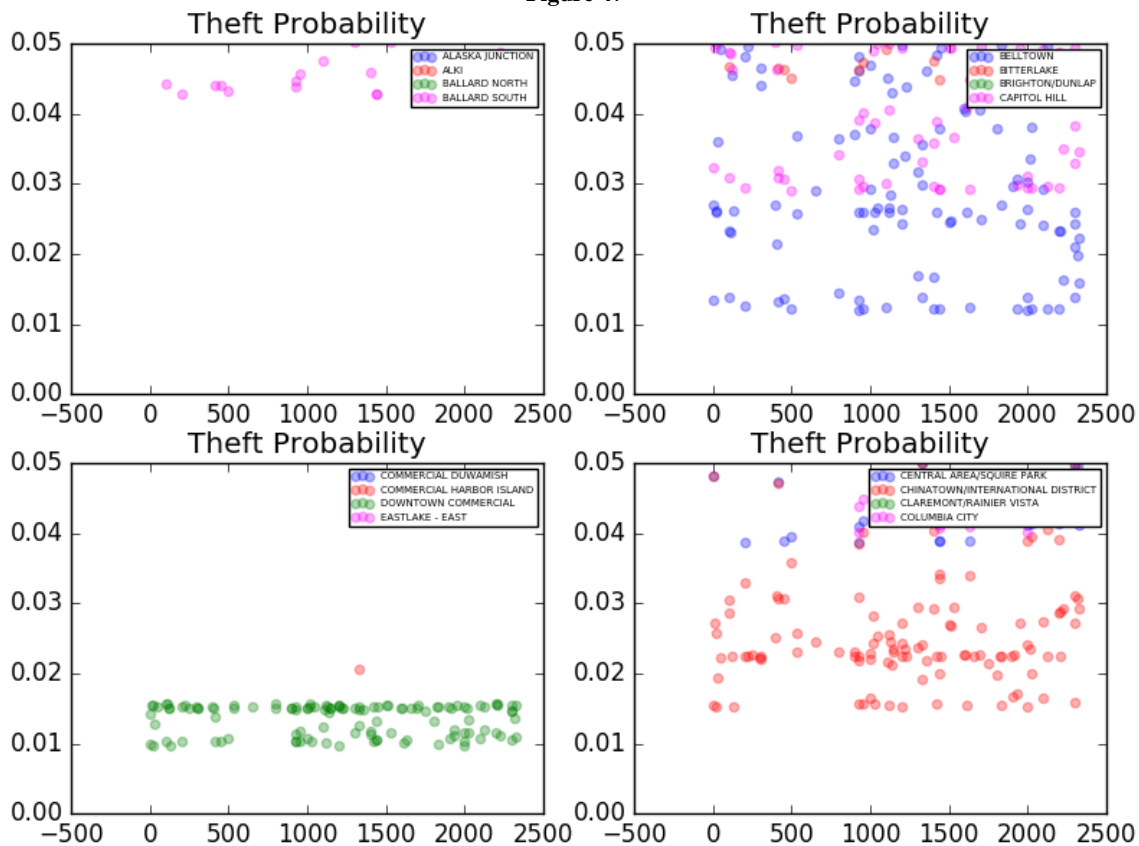


Figure 48

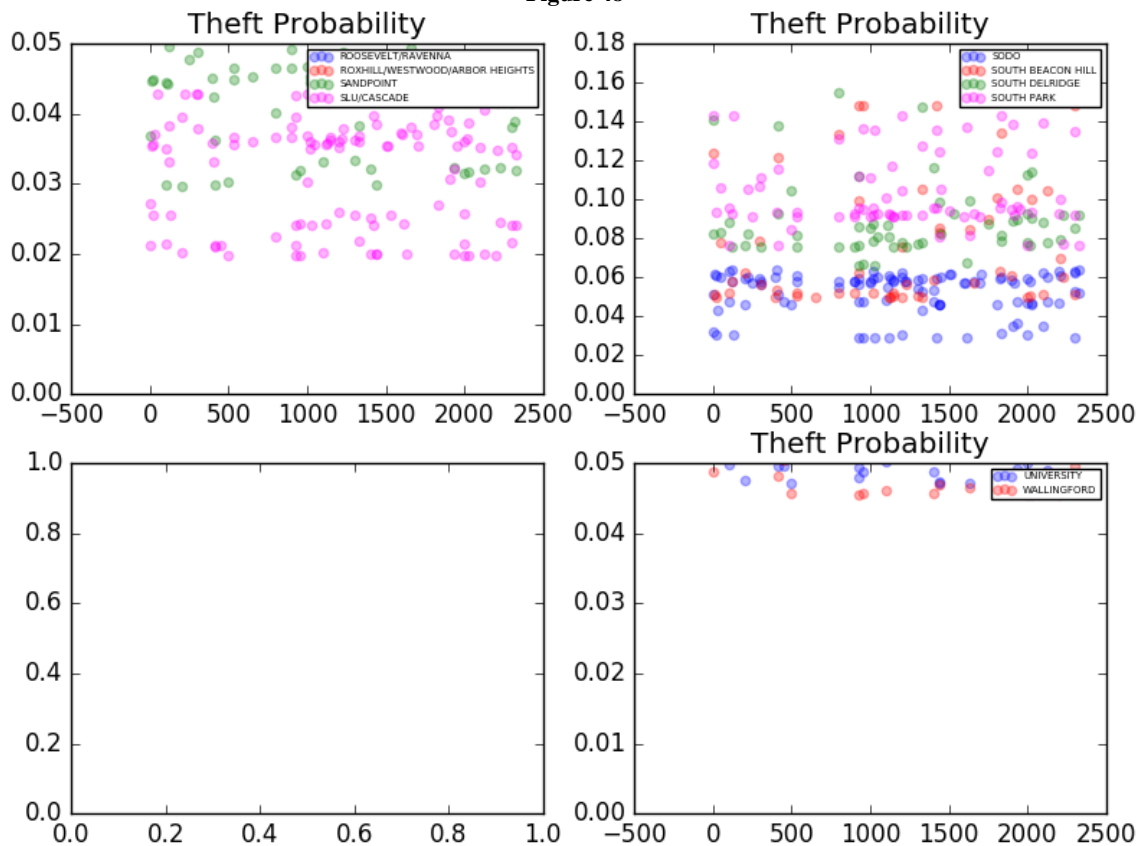


Figure 49

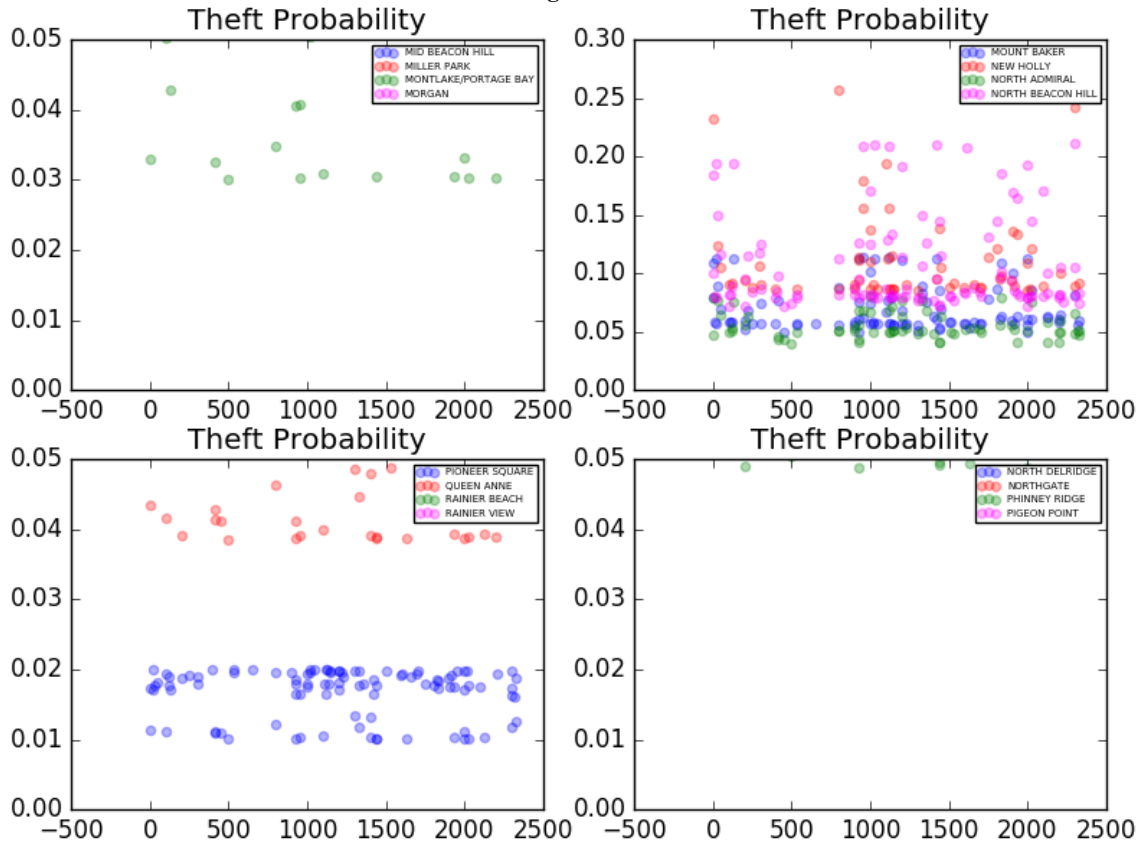


Figure 50

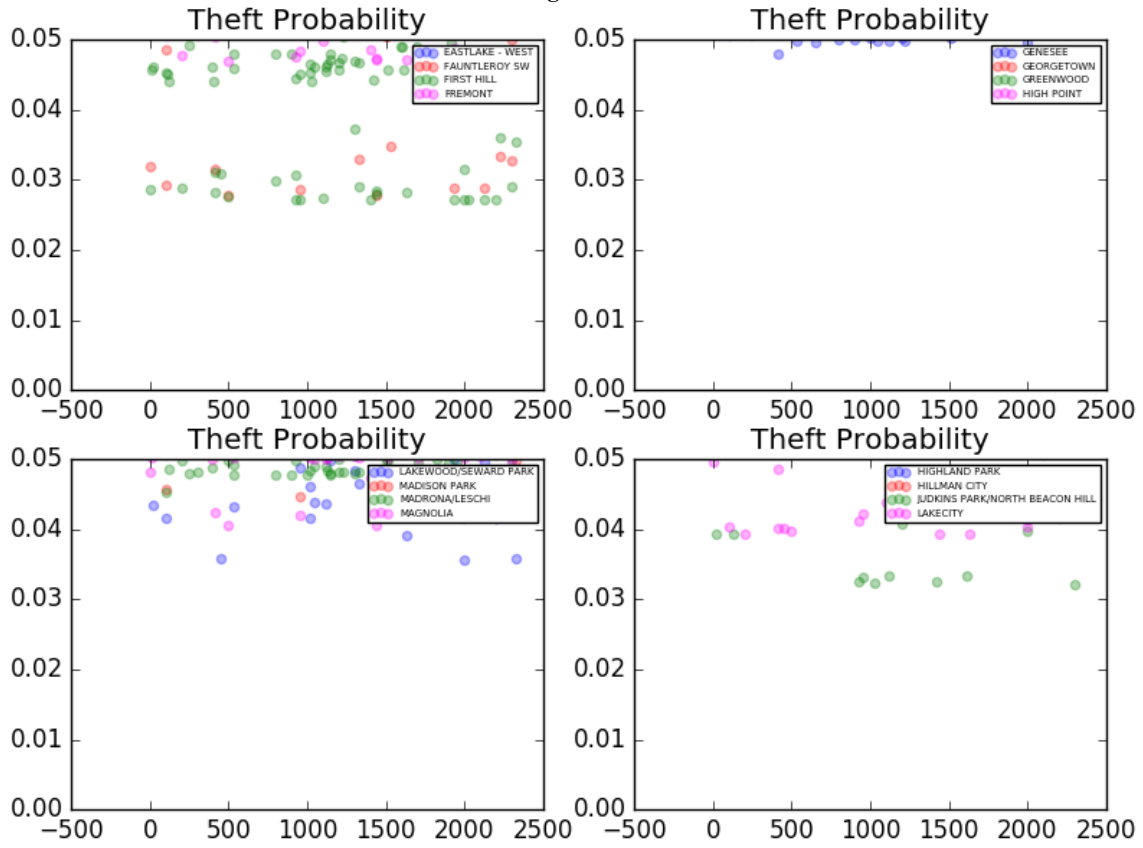


Figure 51

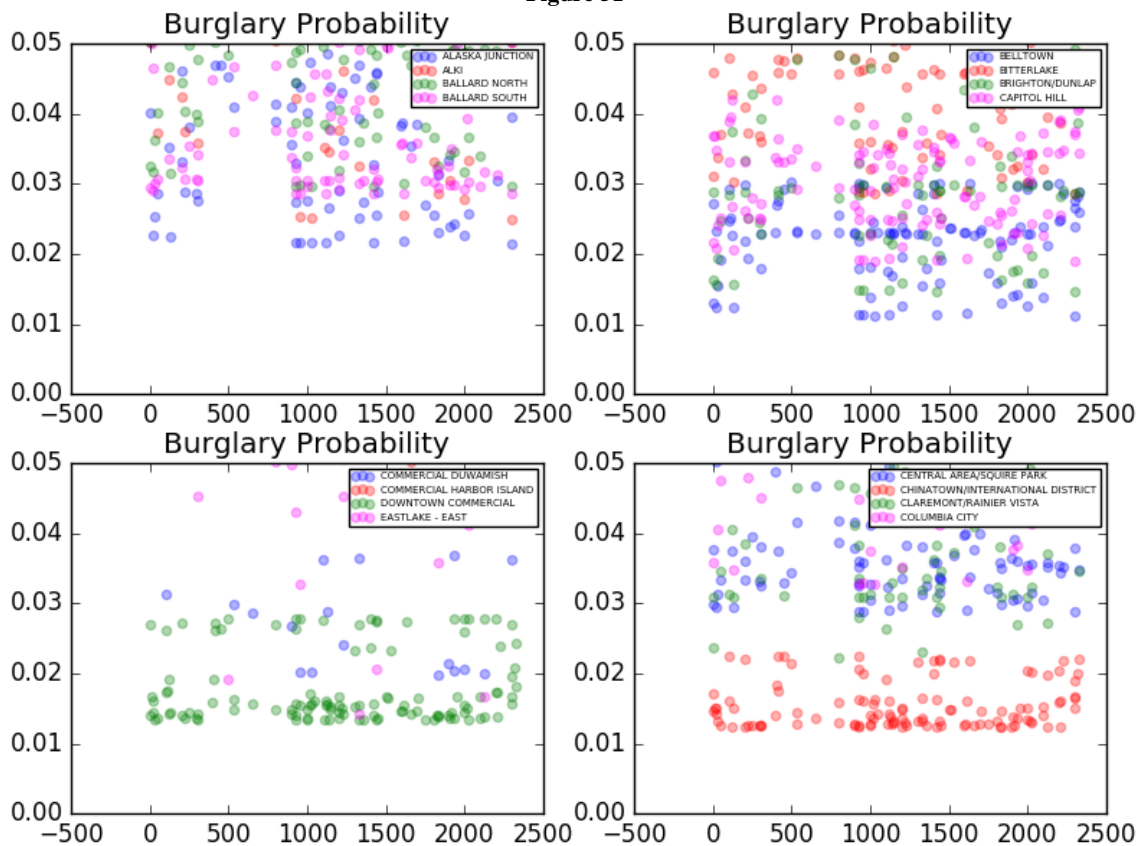


Figure 52

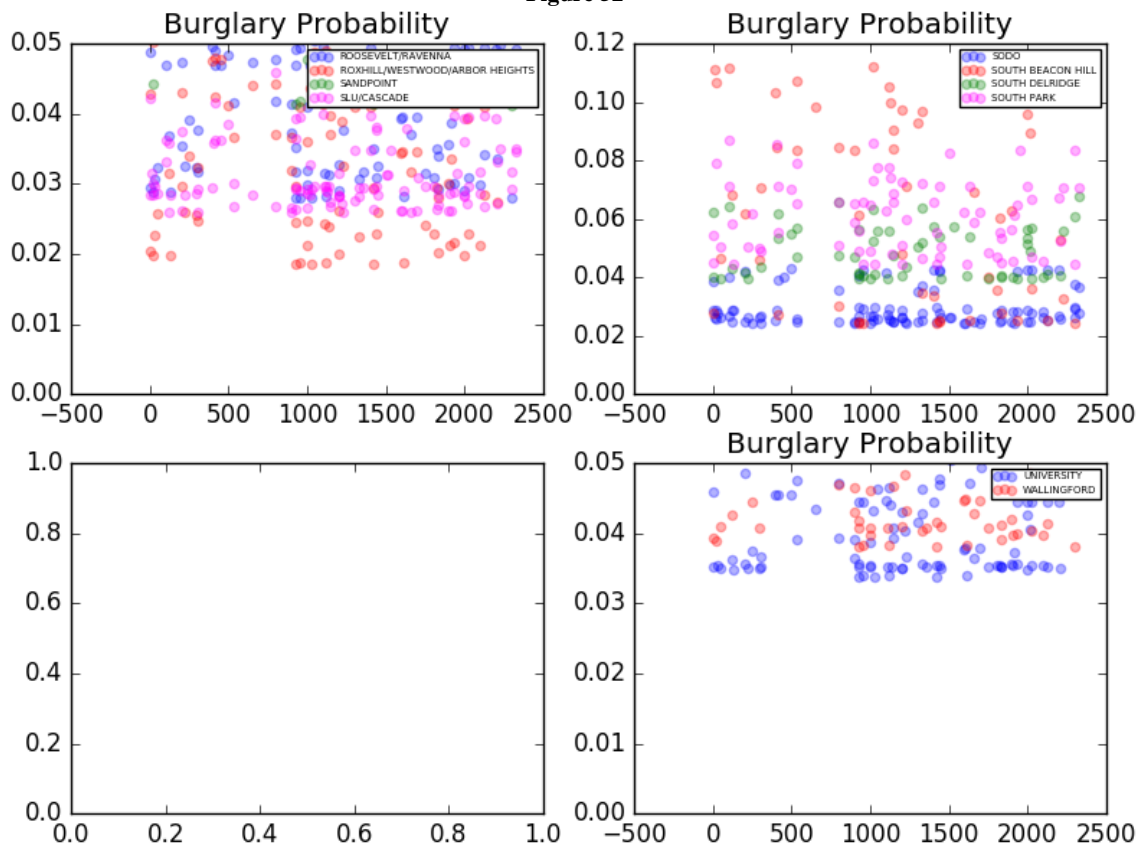


Figure 53

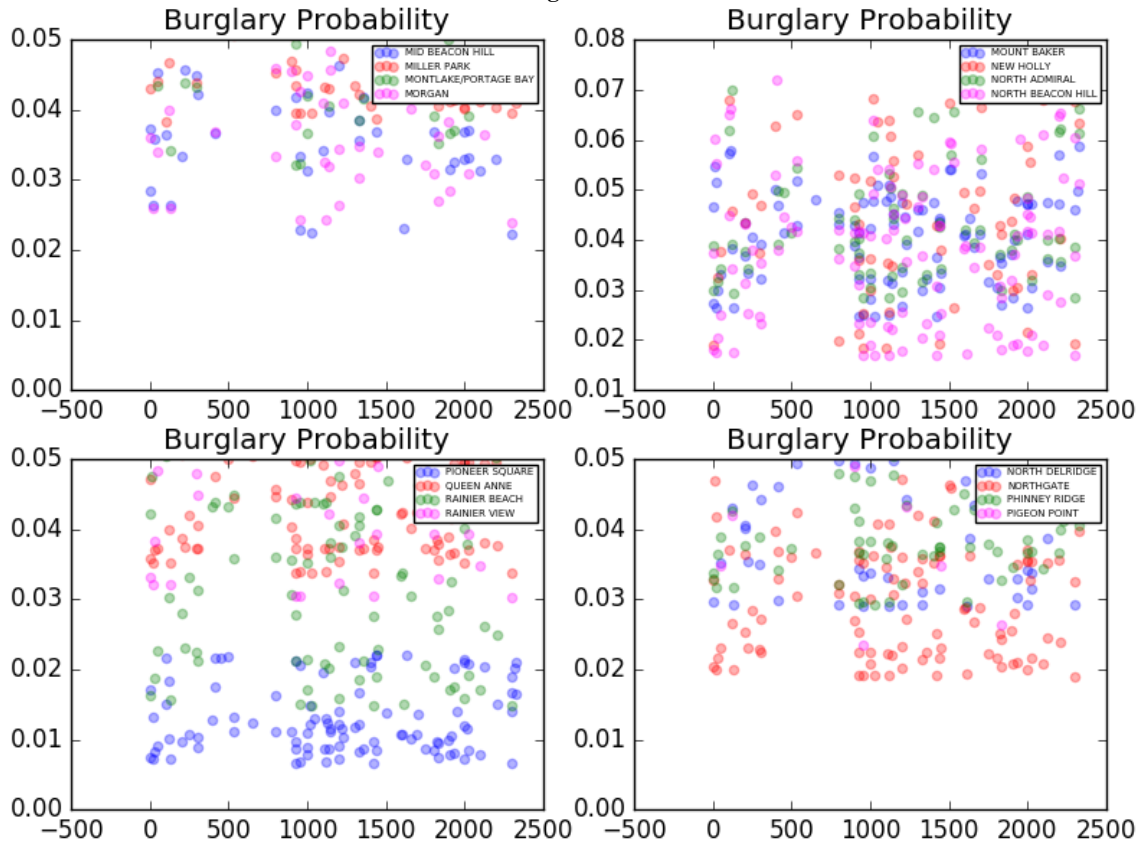


Figure 54

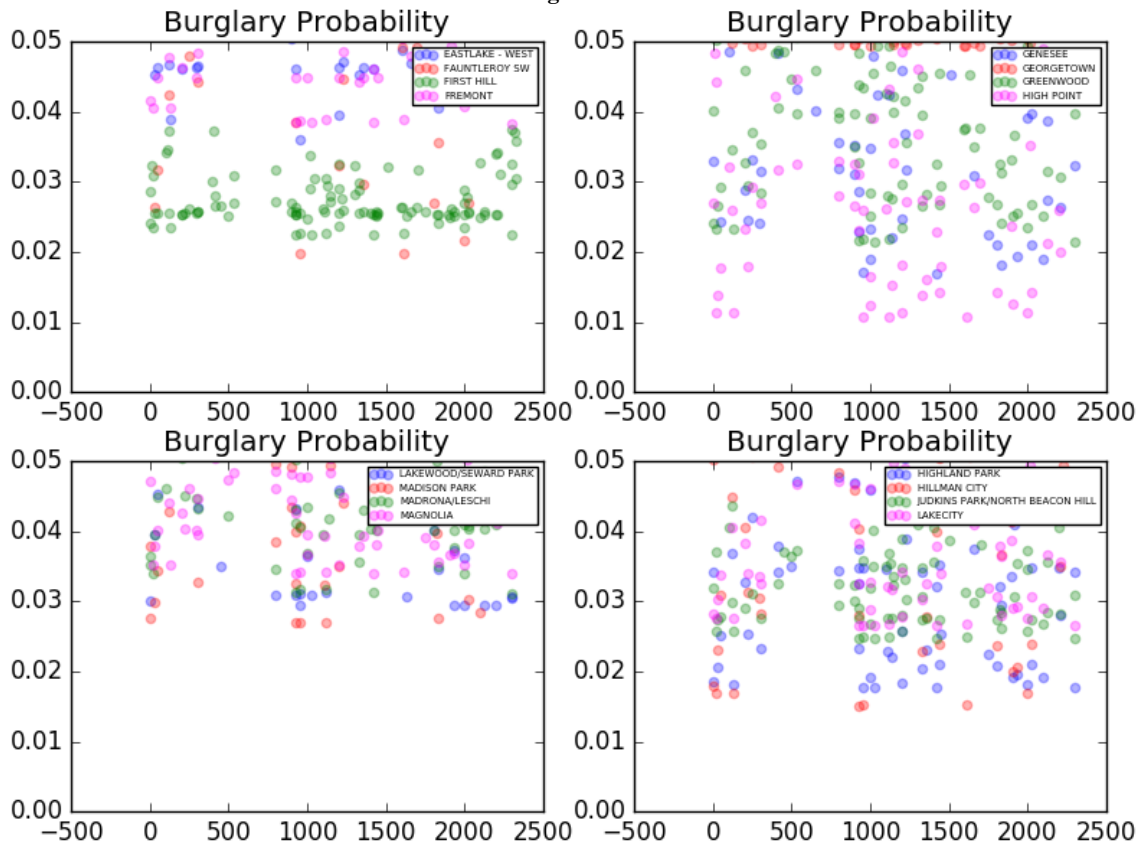


Figure 54

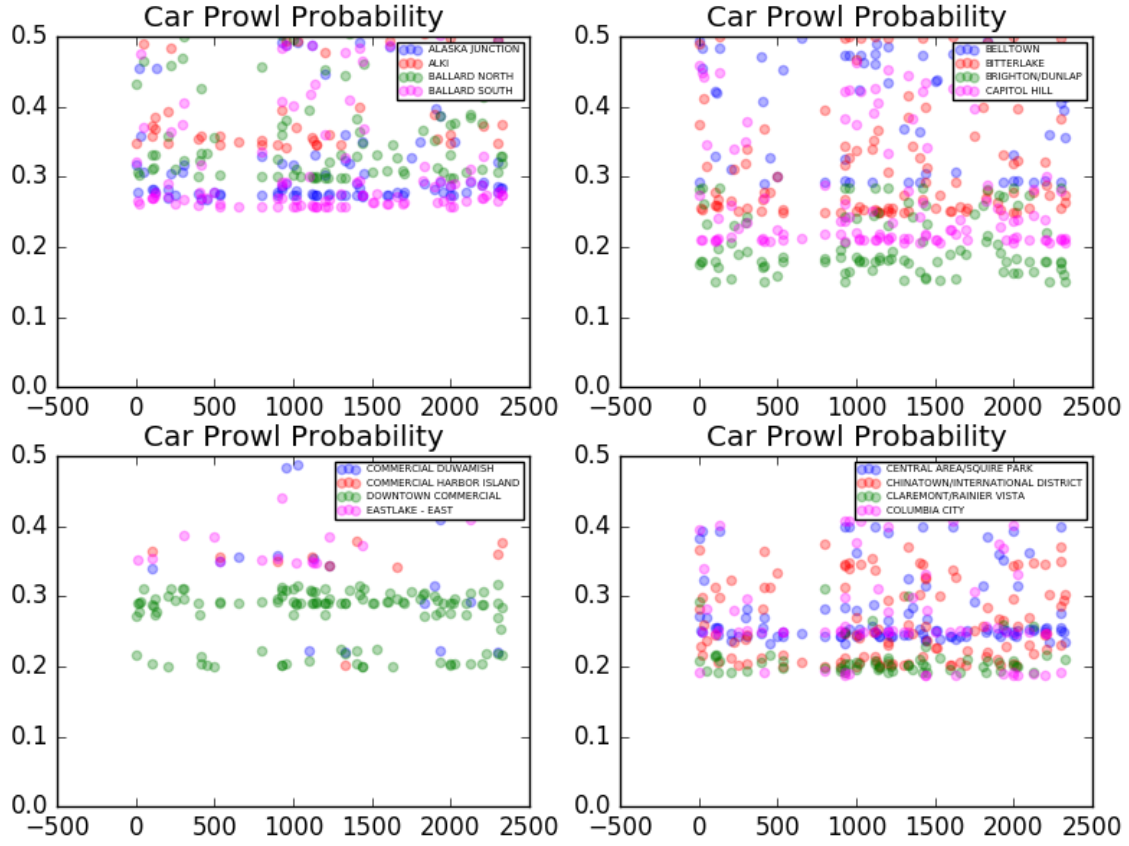


Figure 55

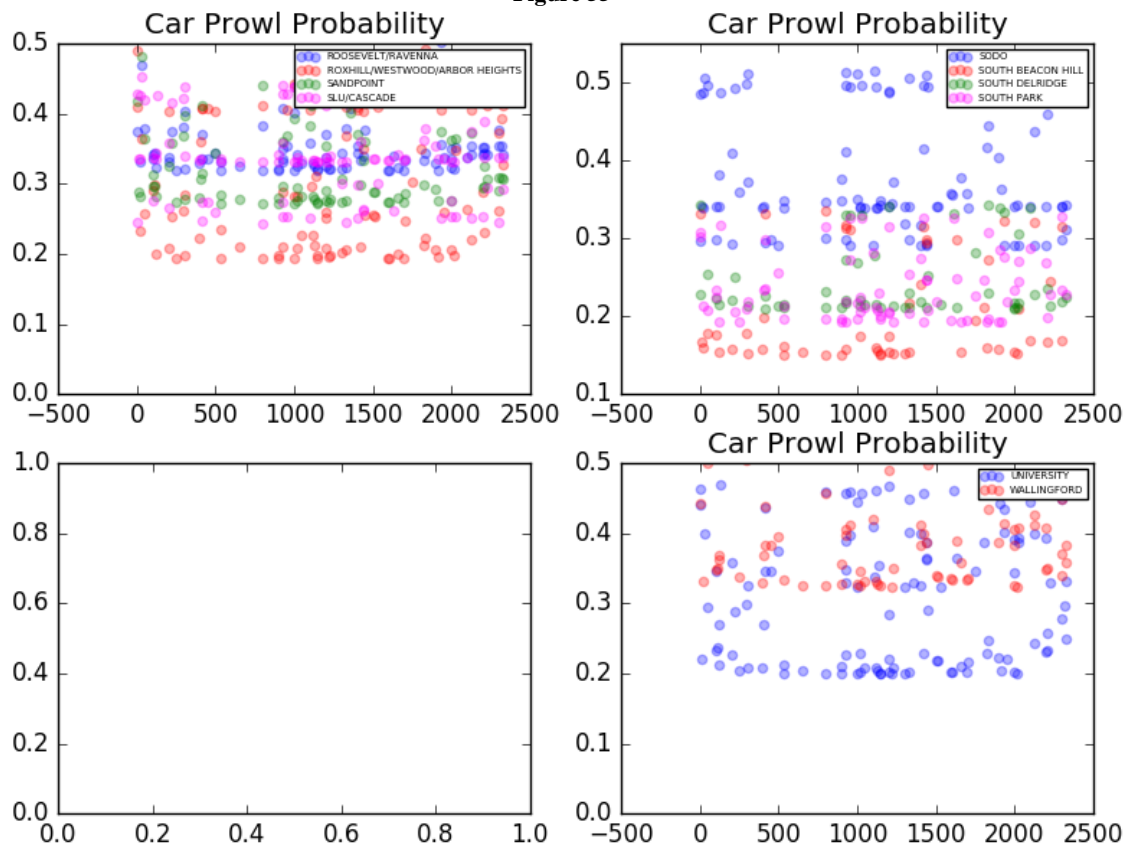


Figure 56

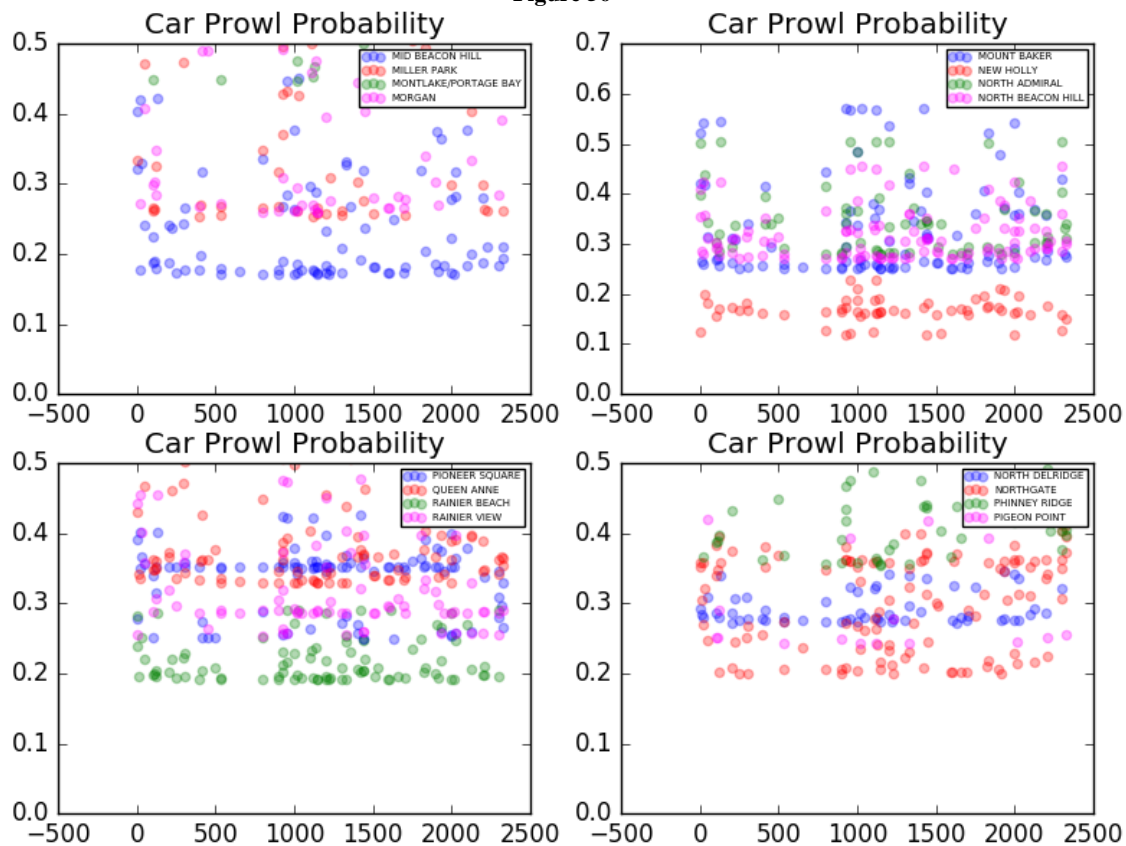


Figure 57

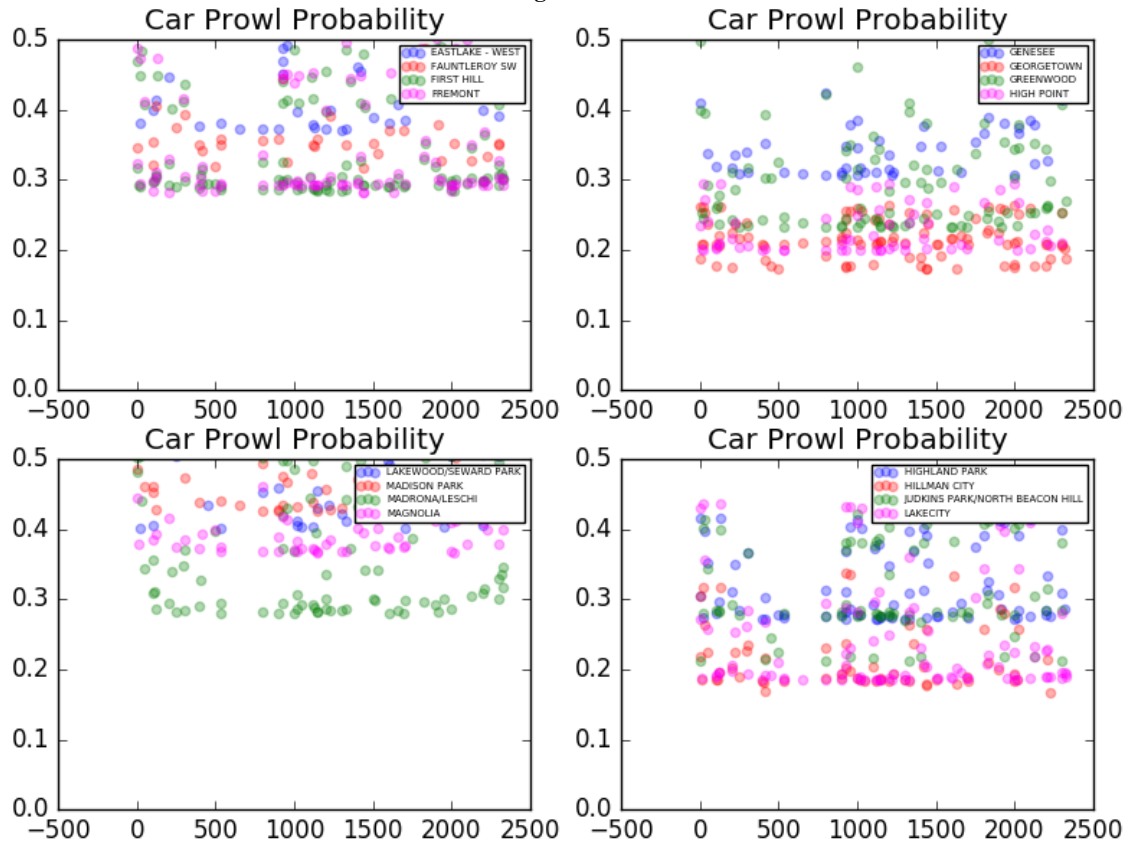


Figure 58

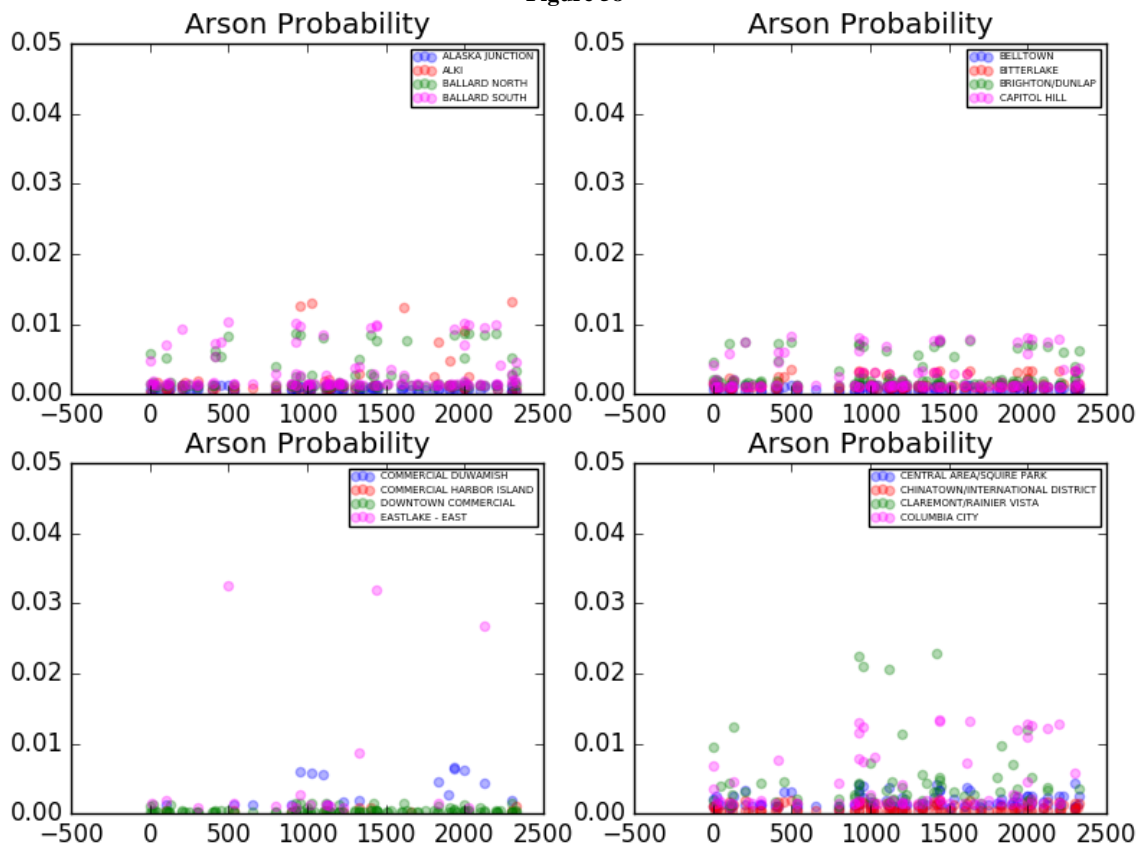


Figure 59

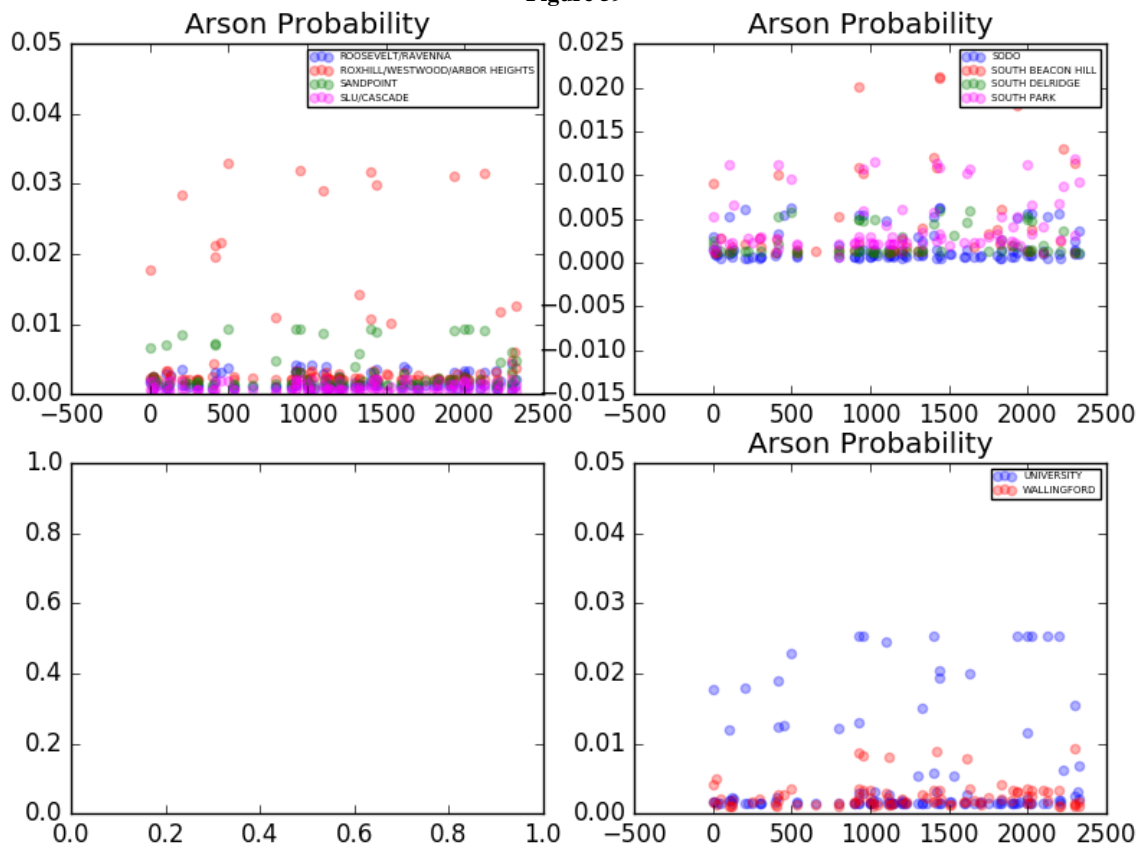


Figure 60

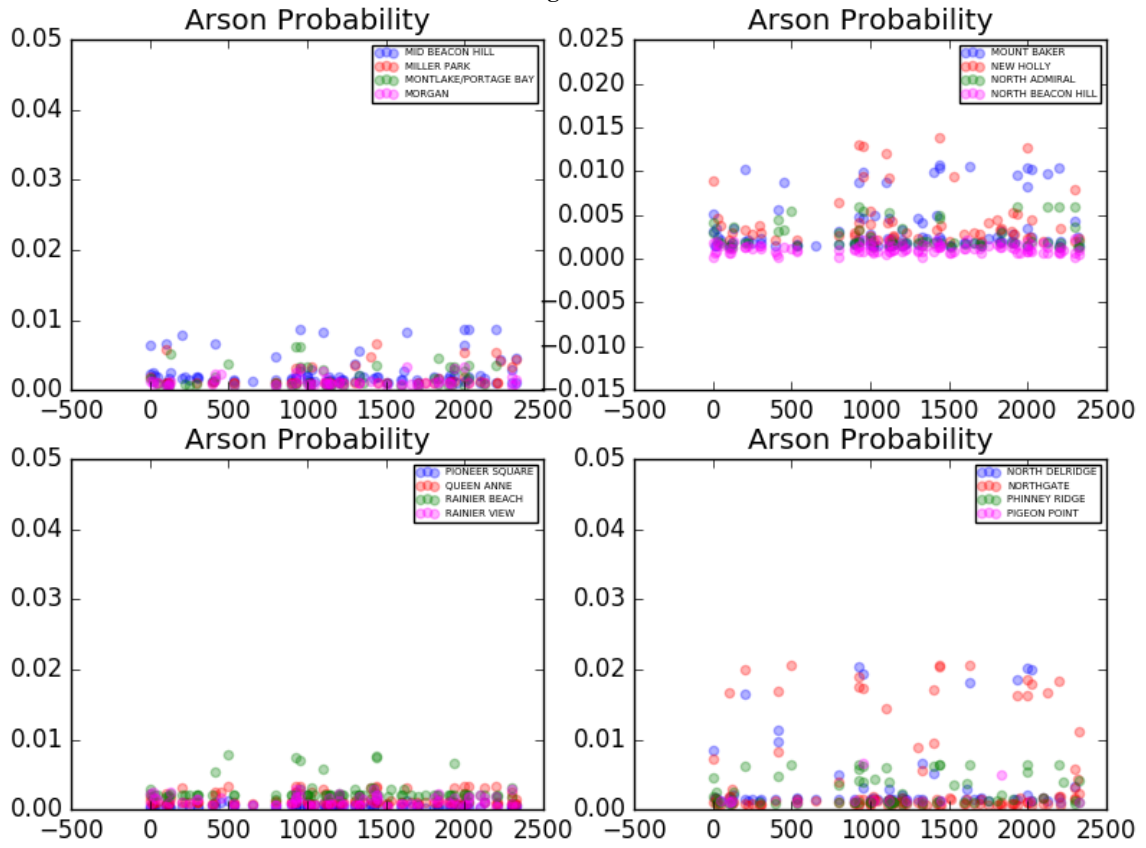


Figure 61

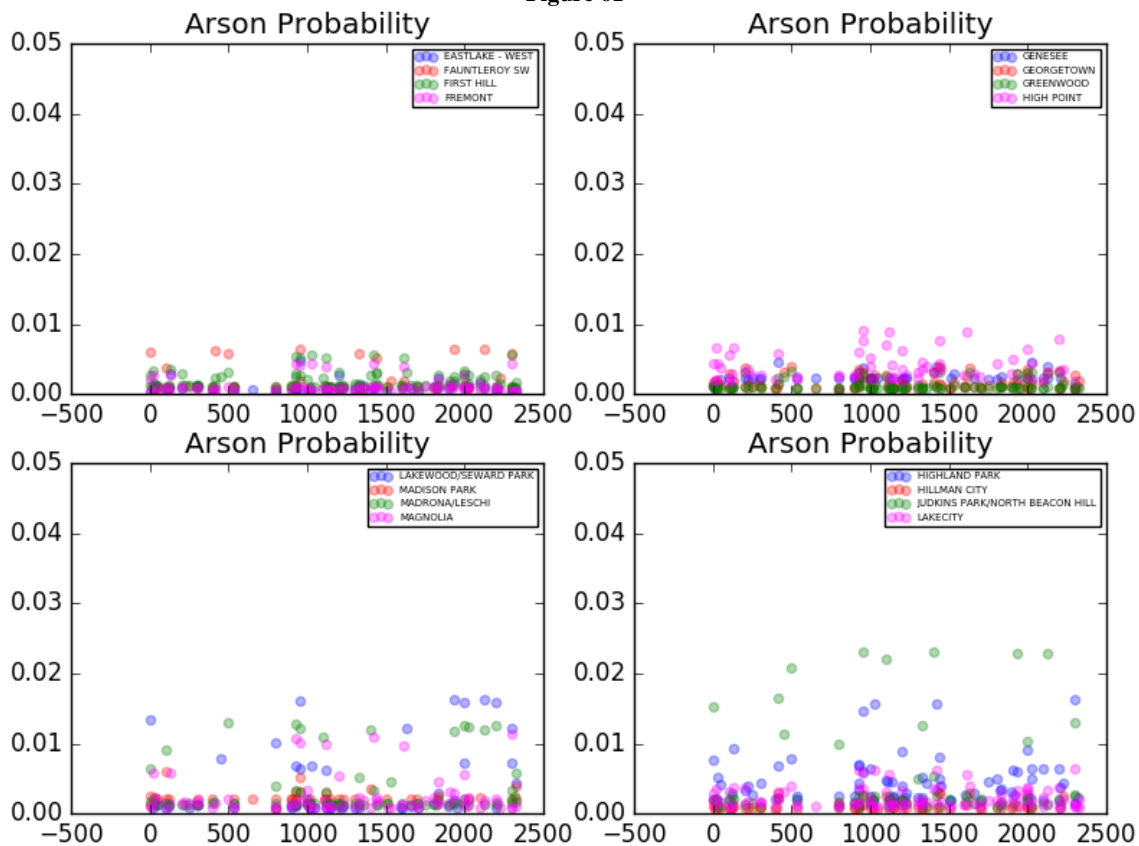


Figure 62

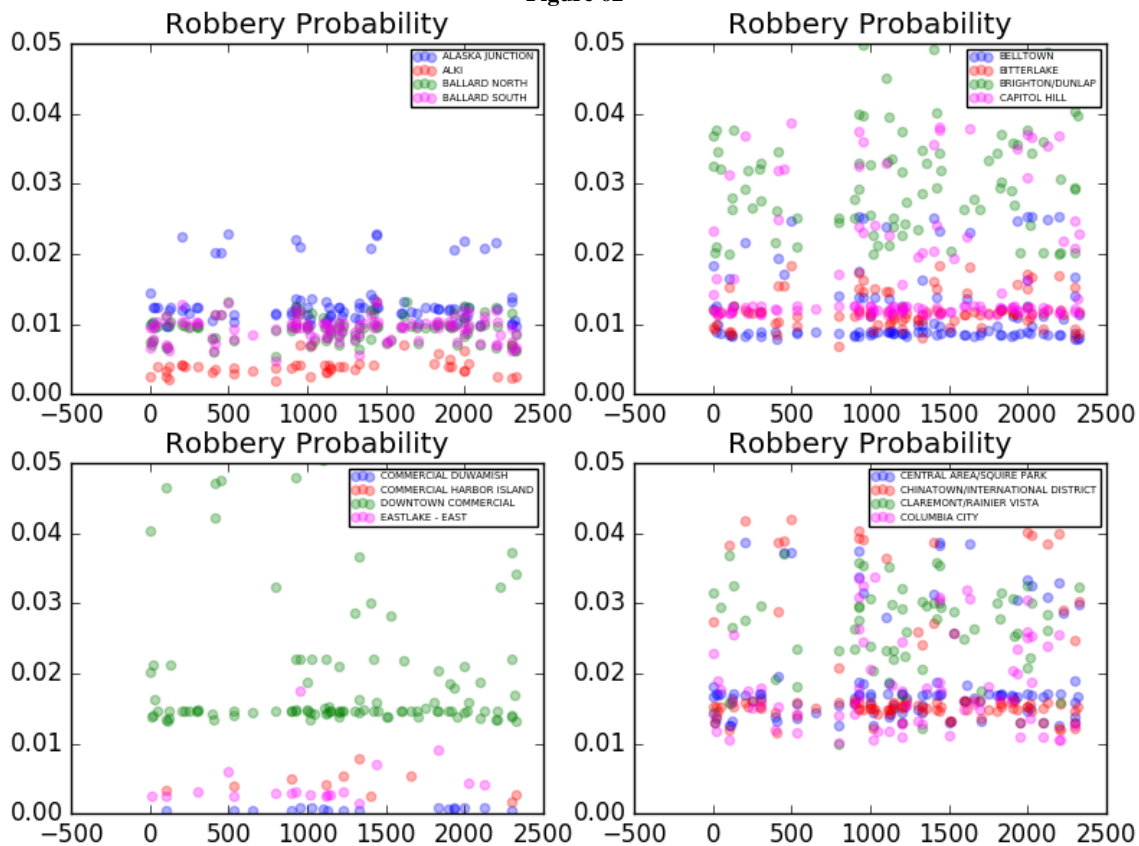


Figure 63

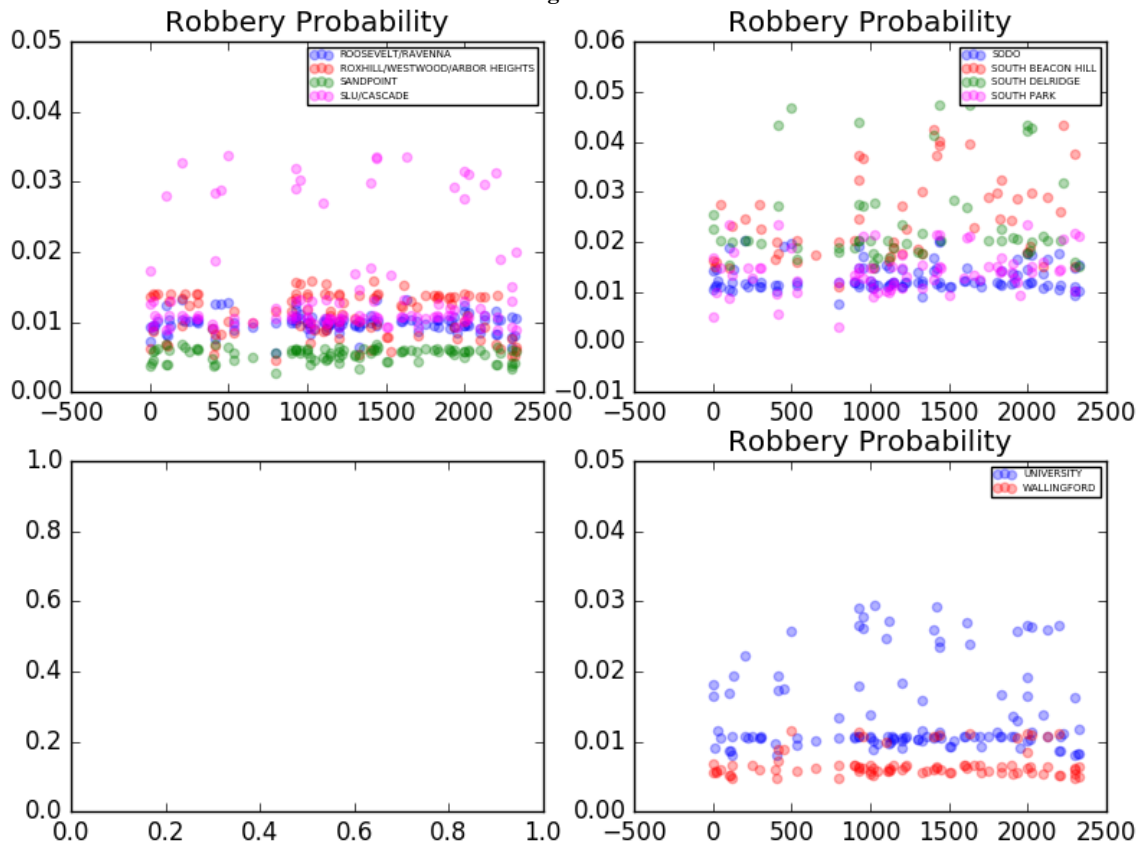


Figure 64

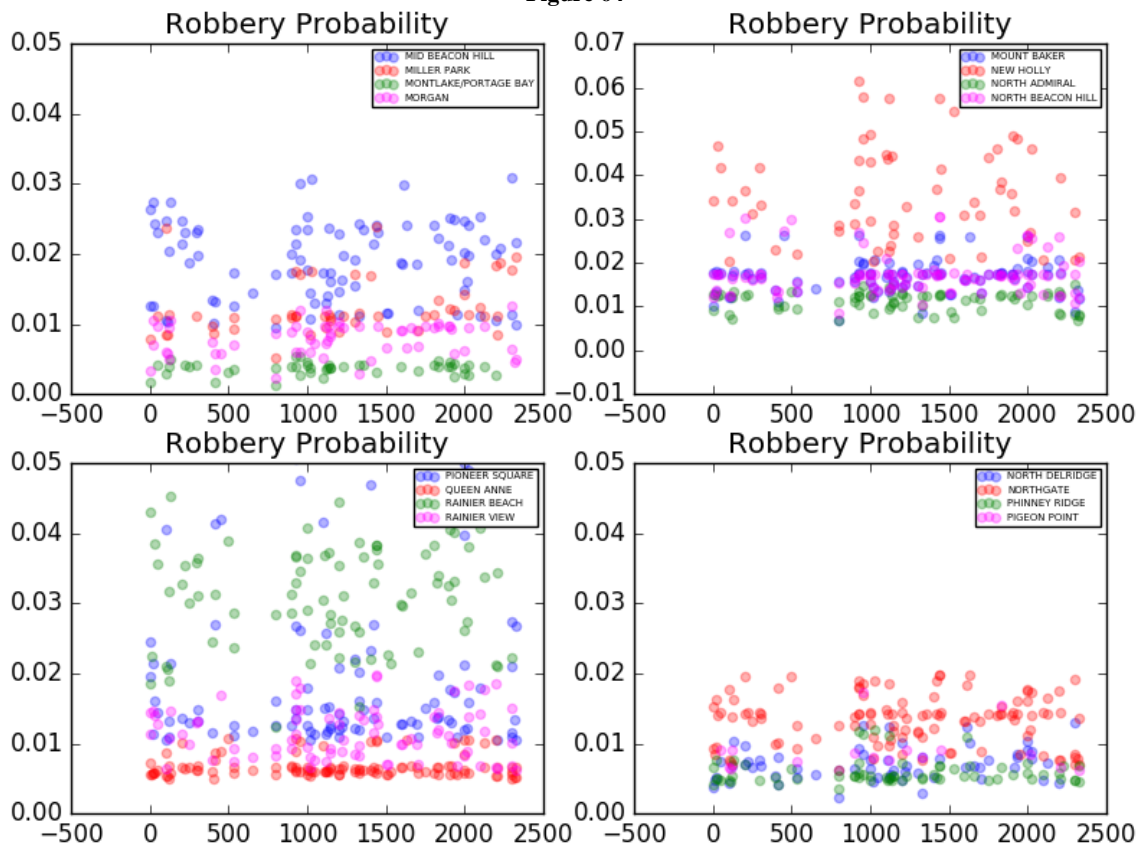
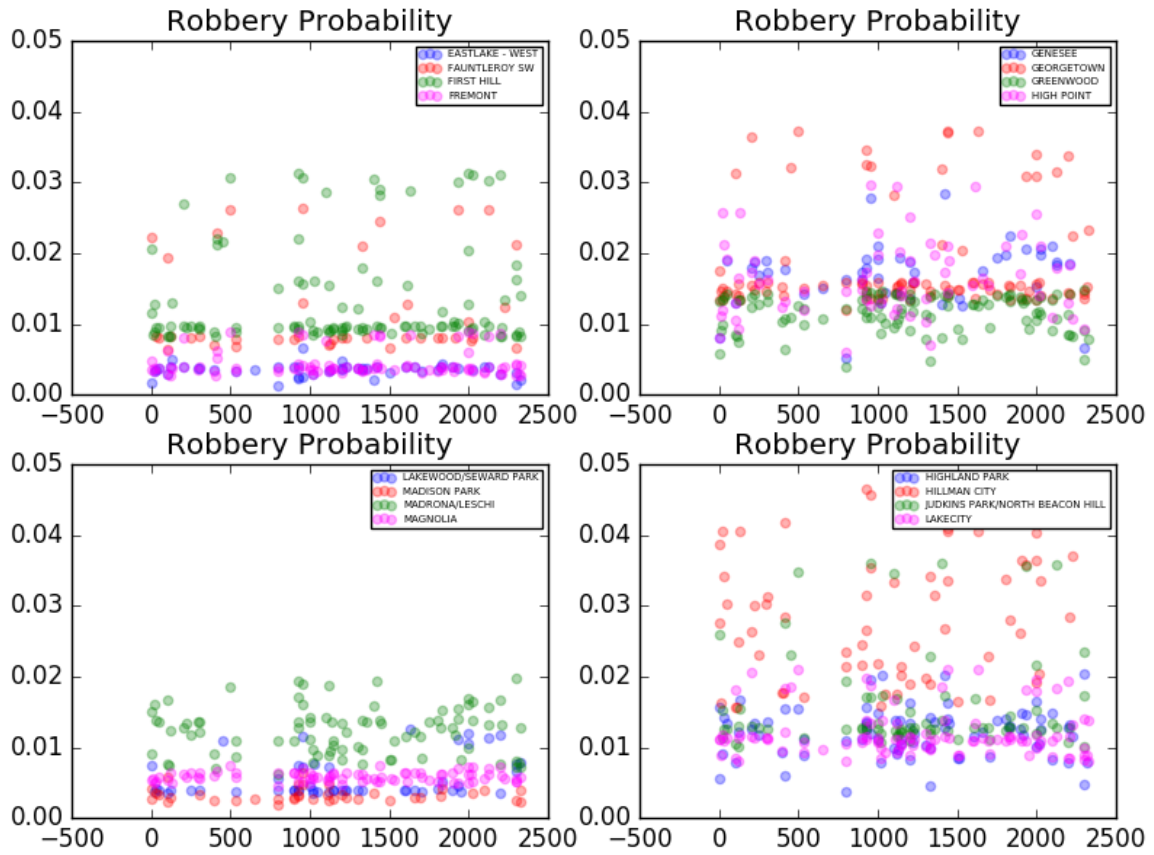


Figure 65



8.4 Multiclass Regression

Figure 66

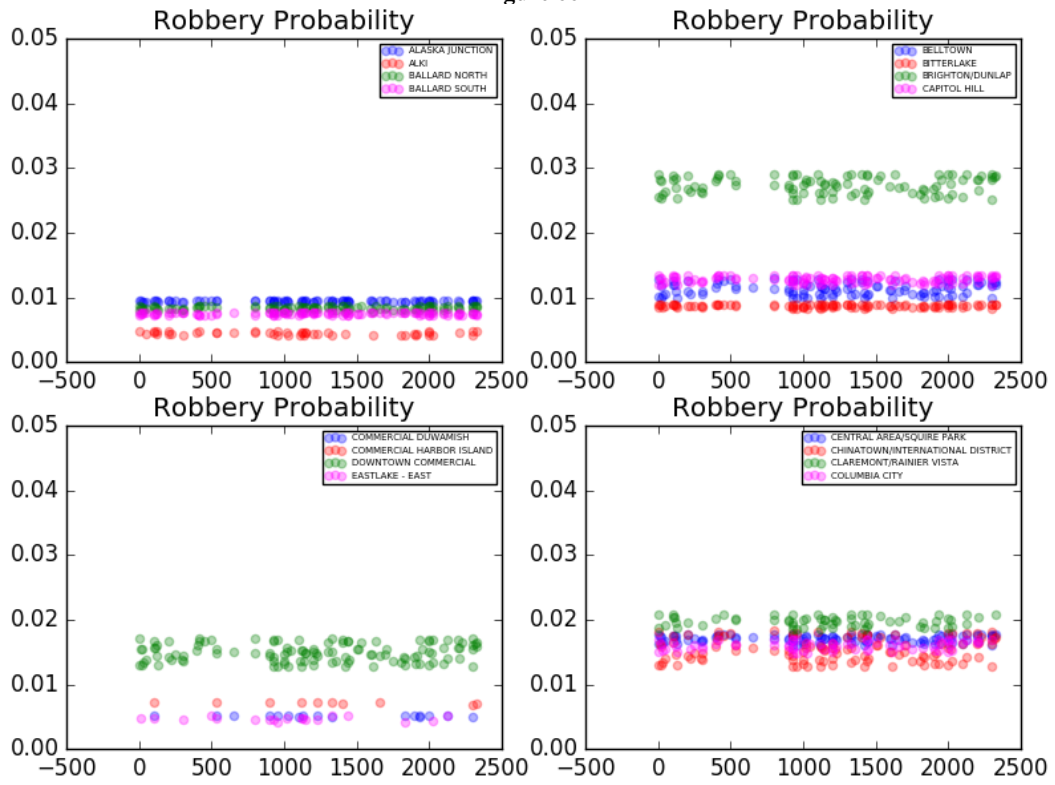


Figure 67

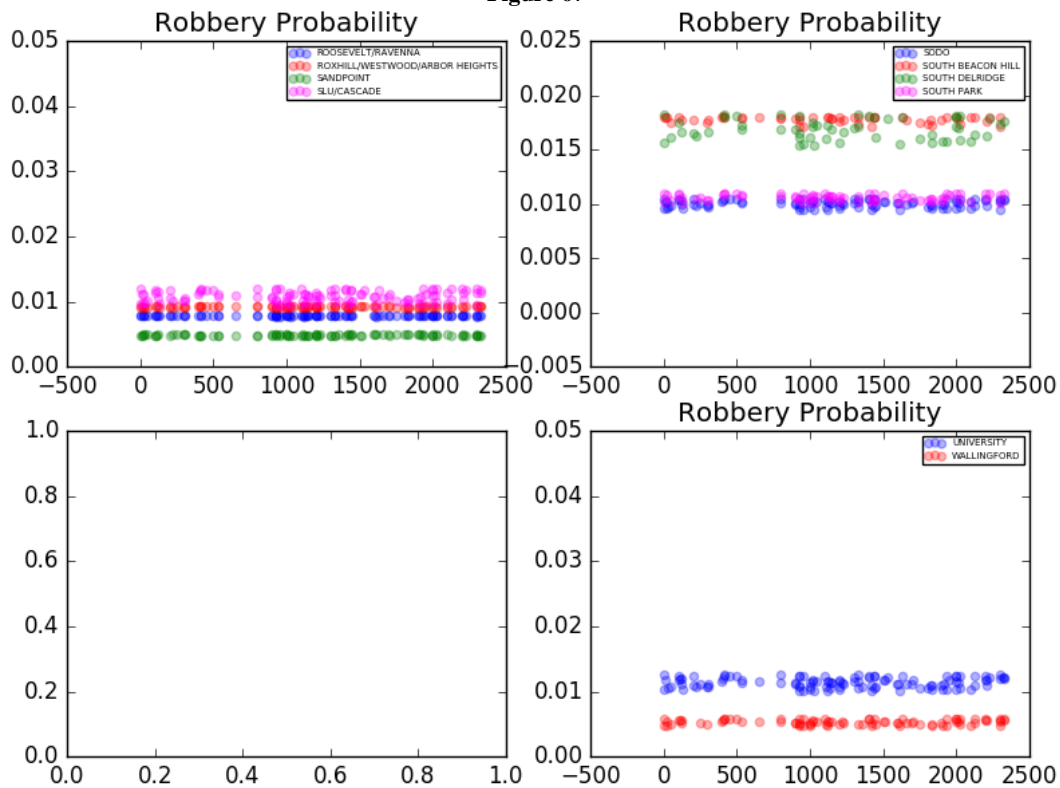


Figure 68

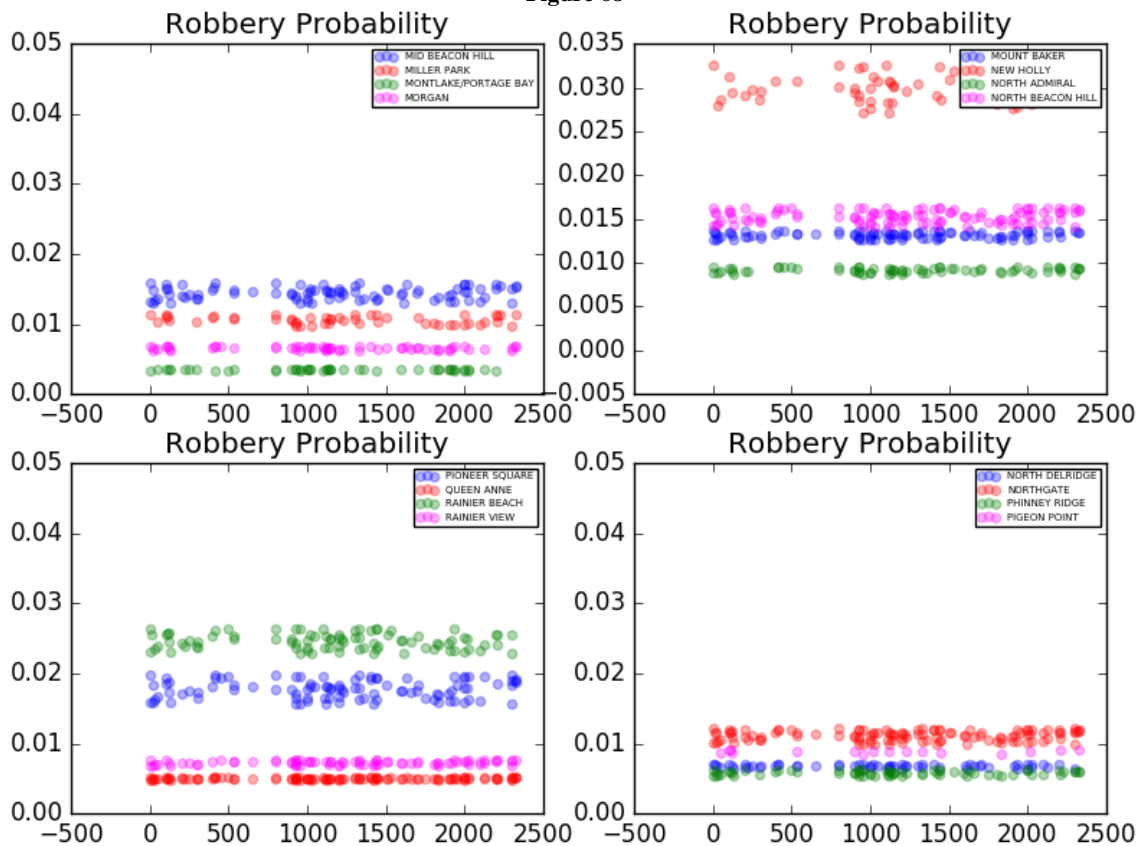


Figure 69

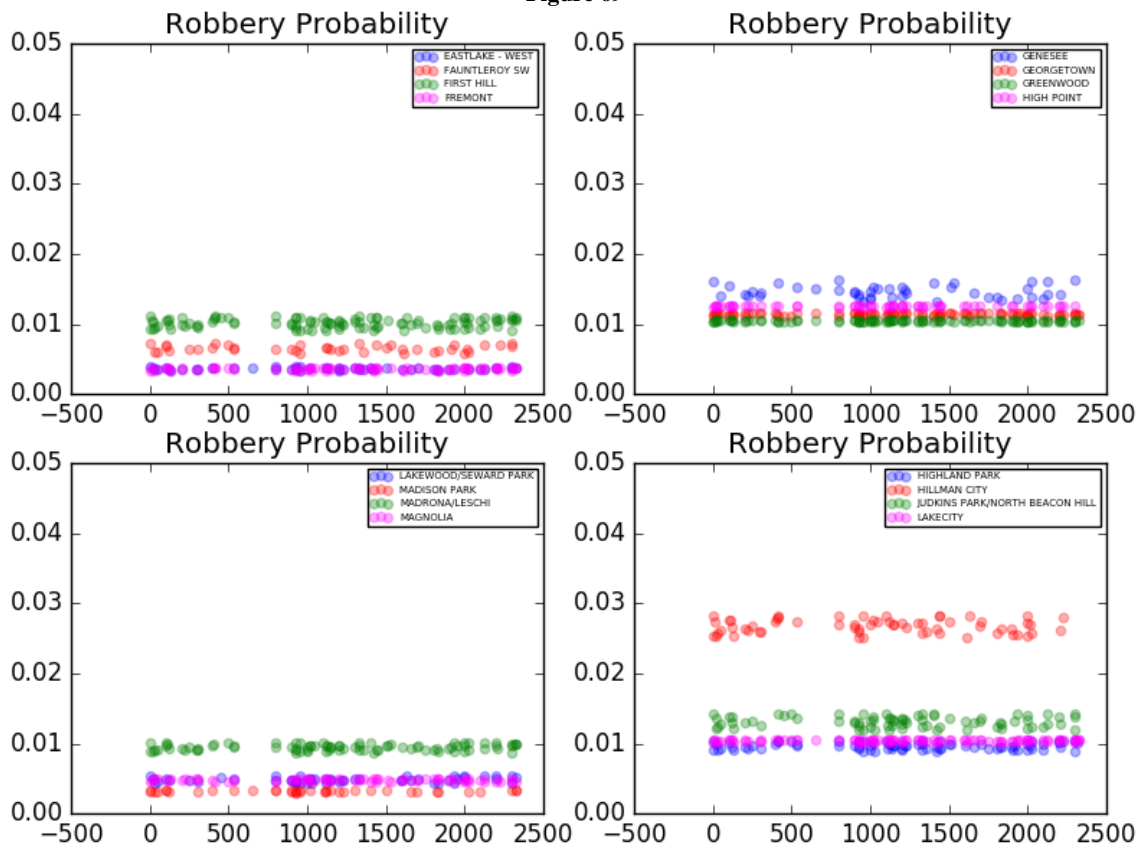


Figure 70

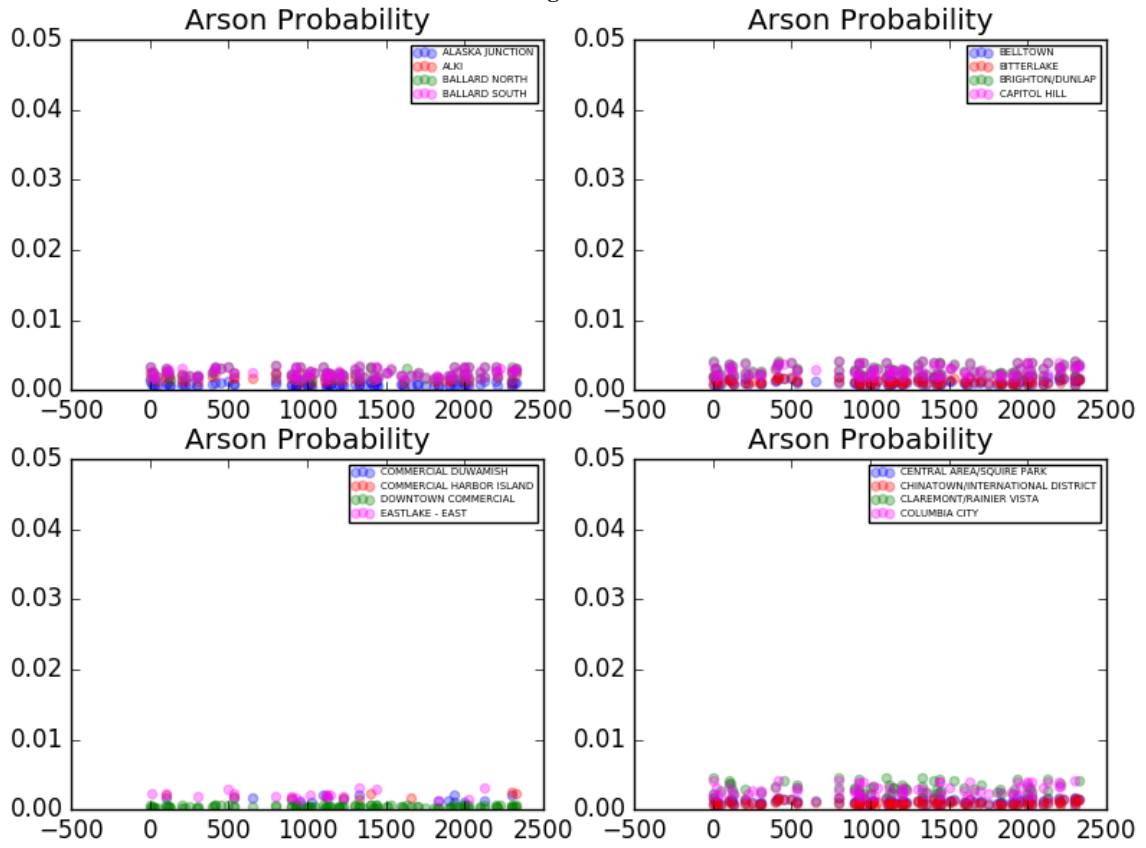


Figure 80

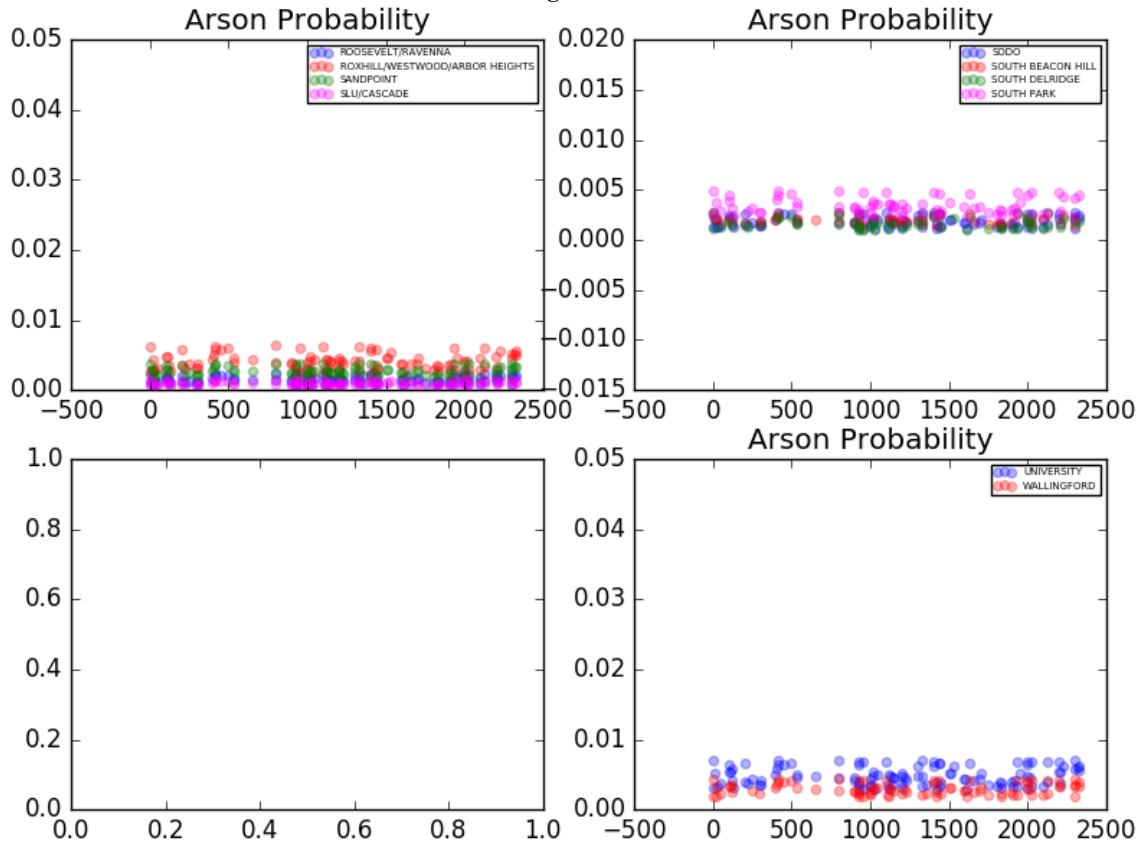


Figure 81

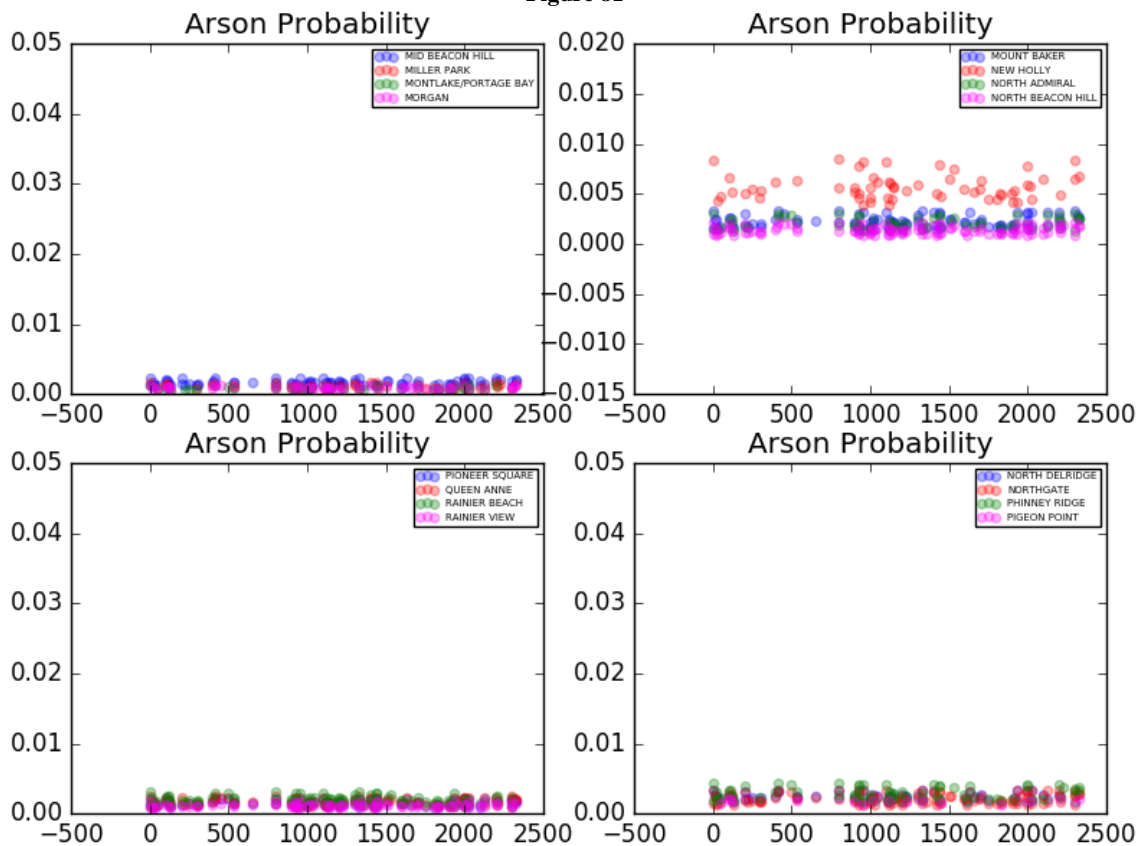


Figure 82

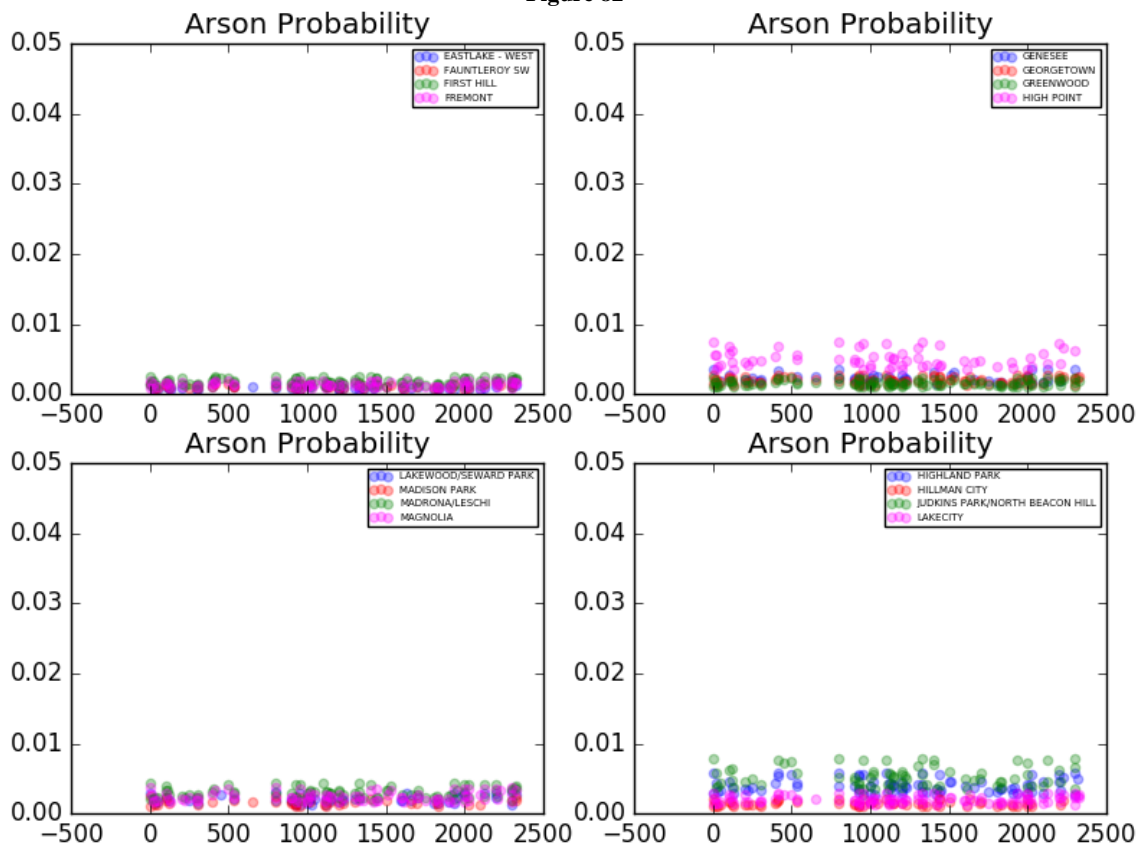


Figure 83

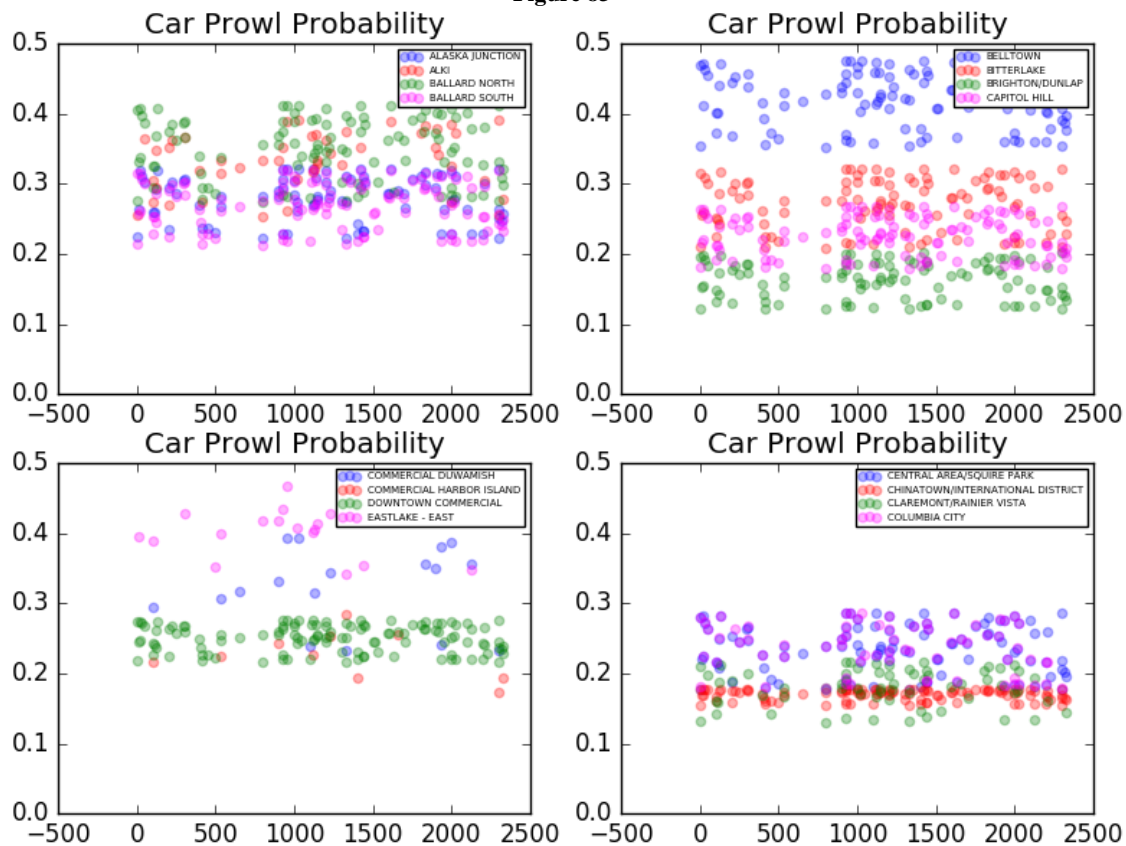


Figure 84

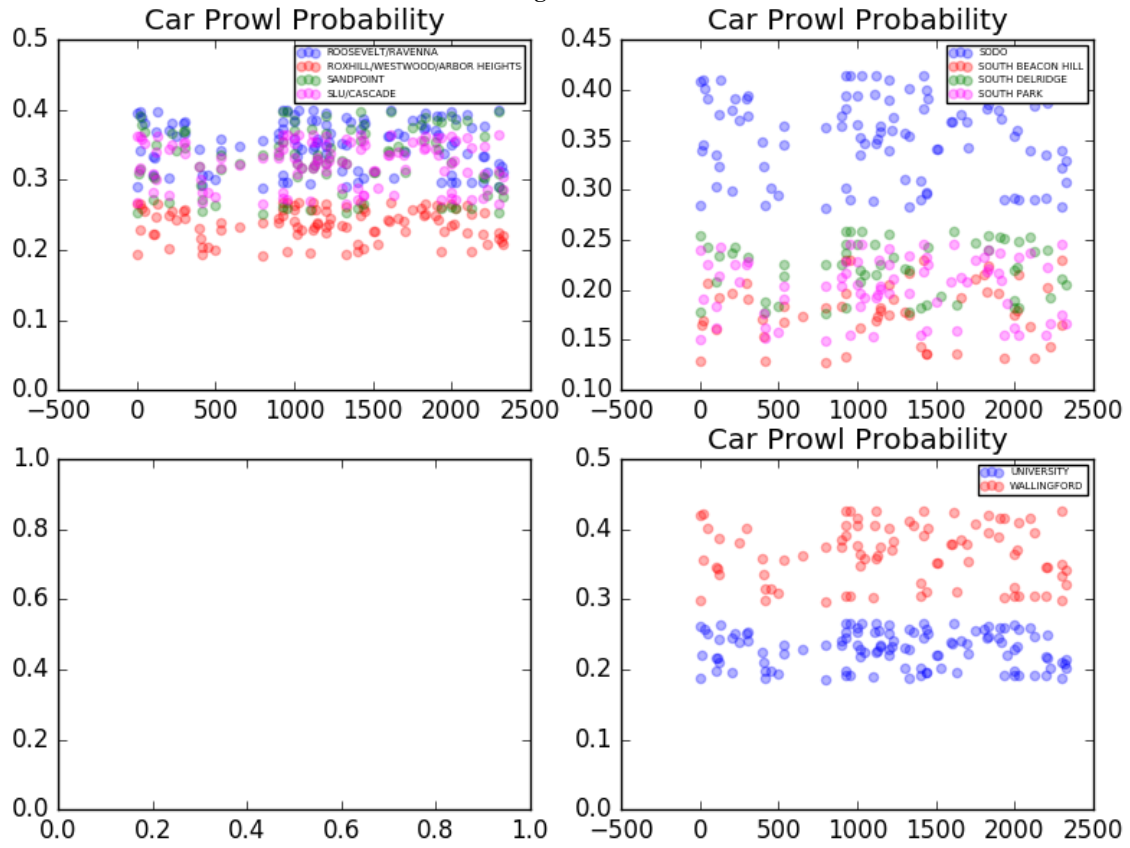


Figure 85

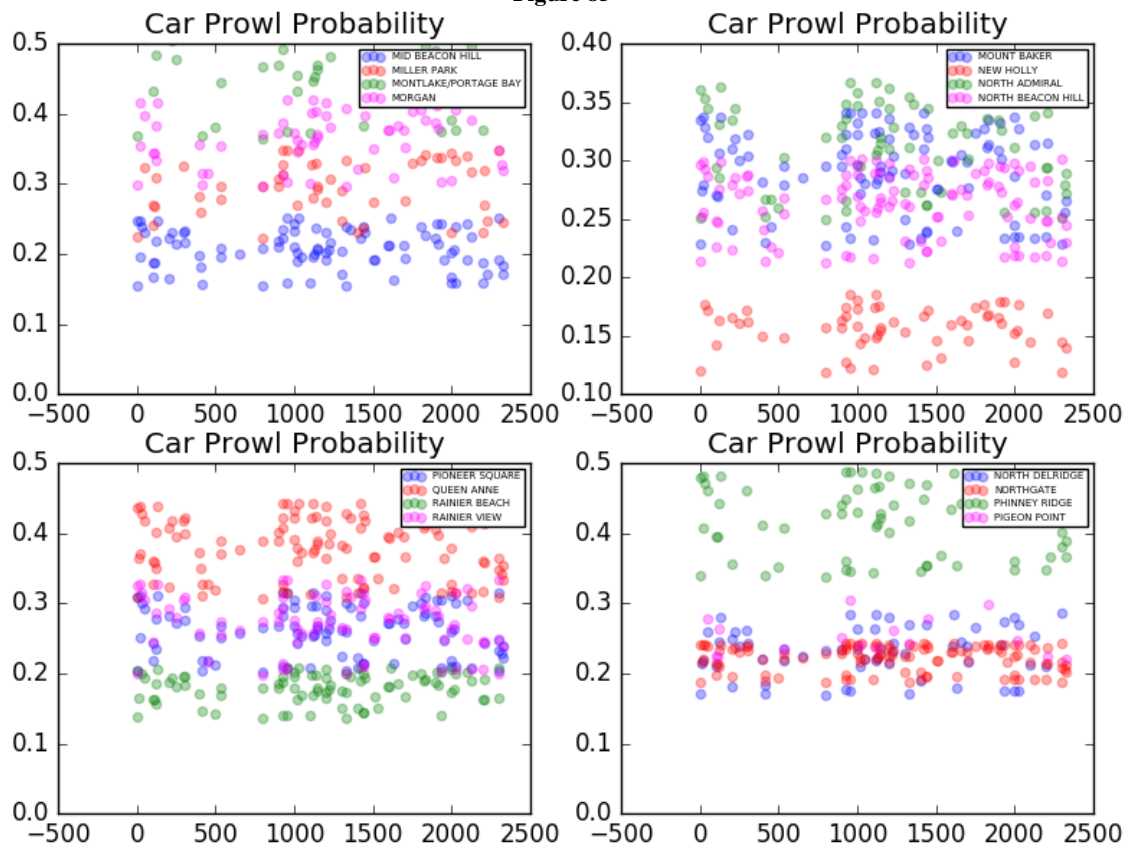


Figure 86

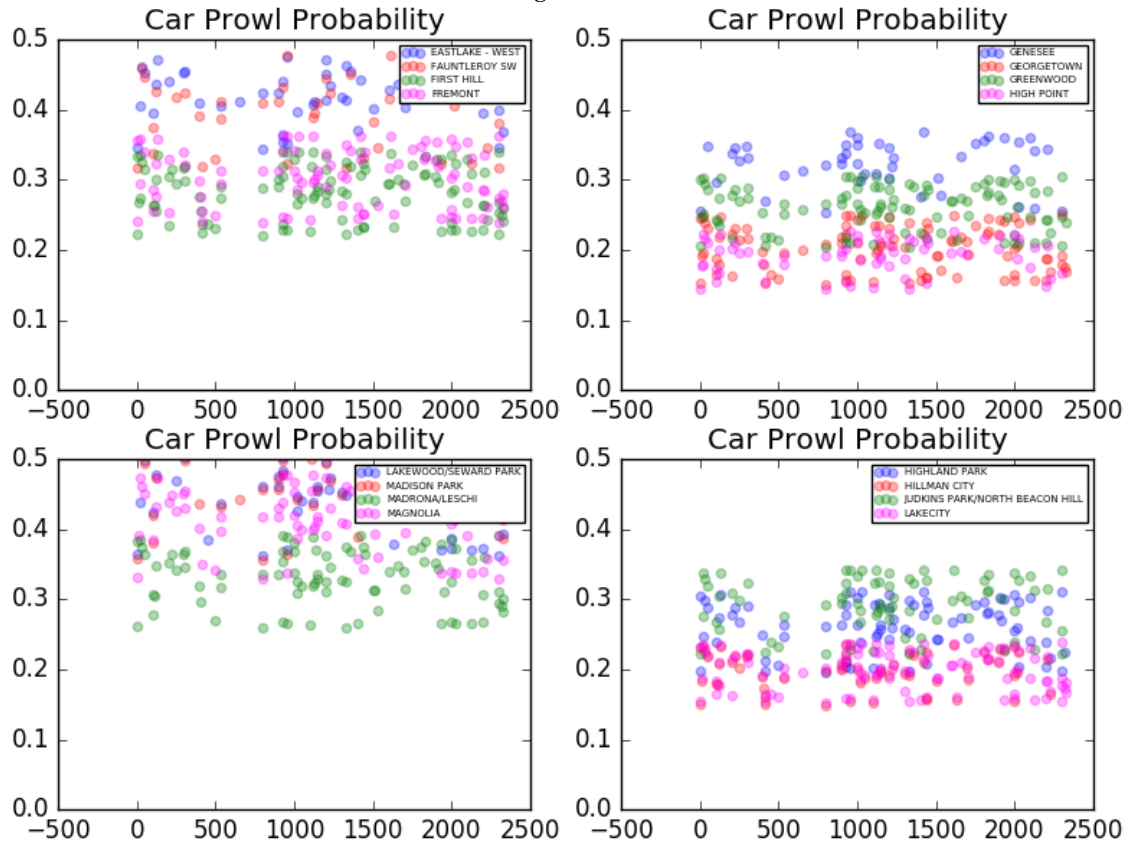


Figure 87

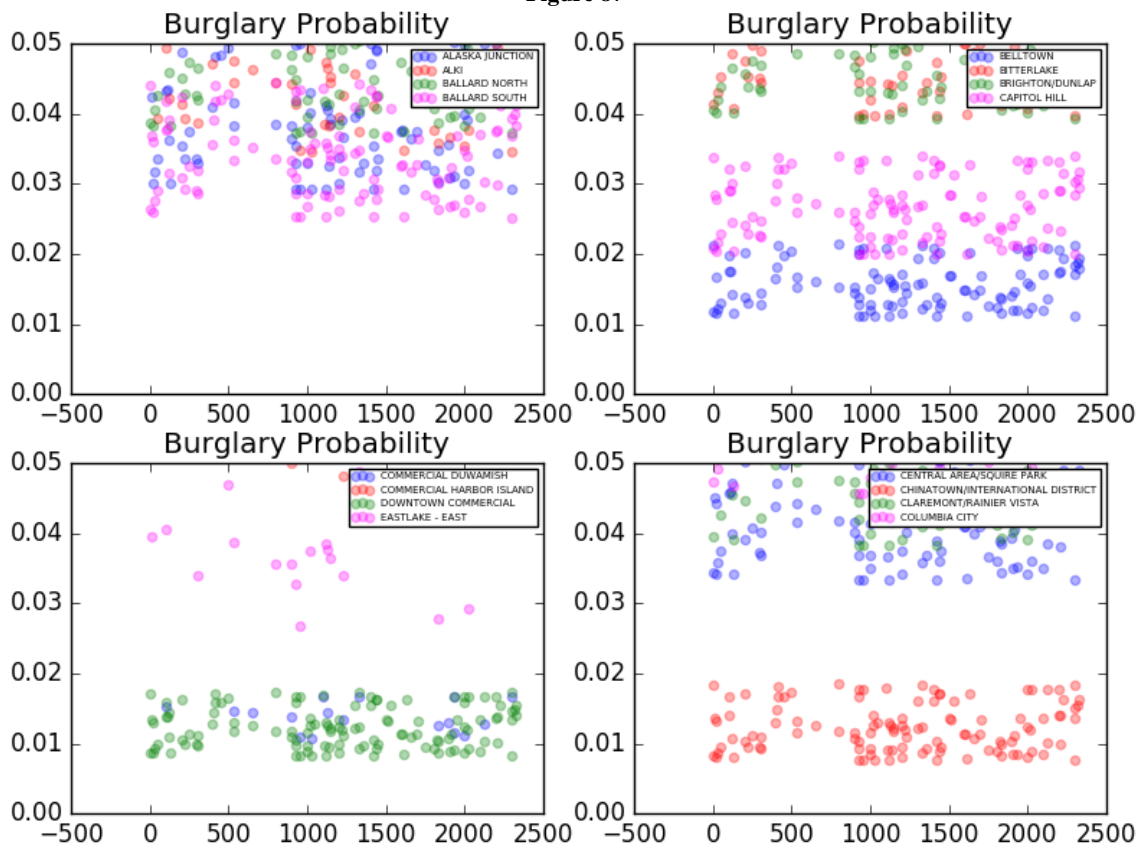


Figure 88

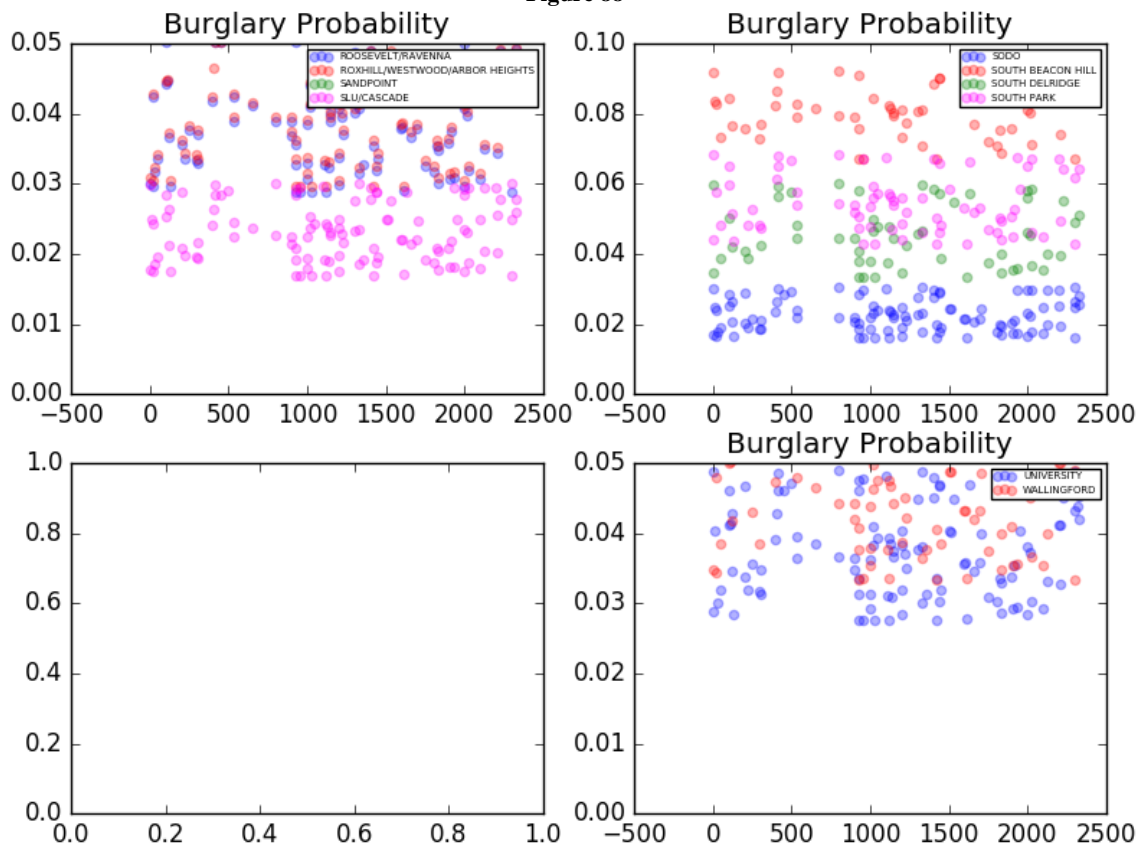


Figure 89

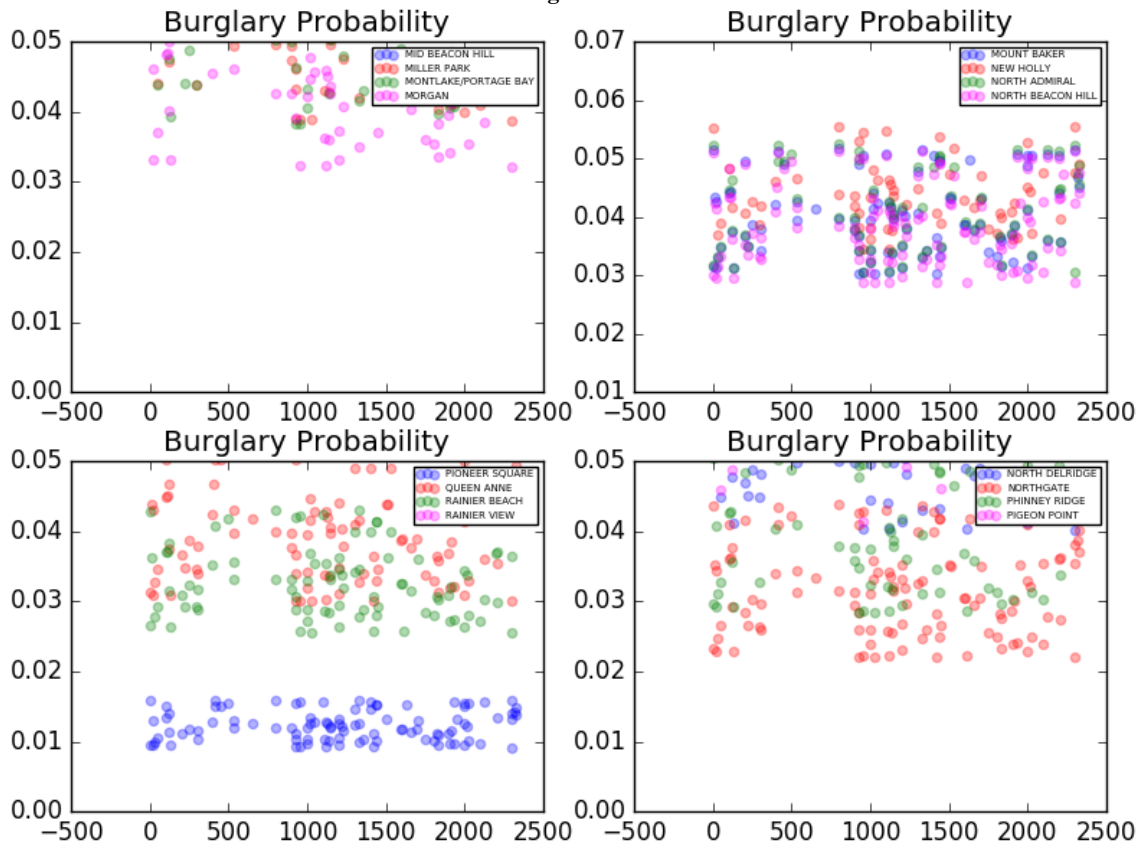


Figure 90

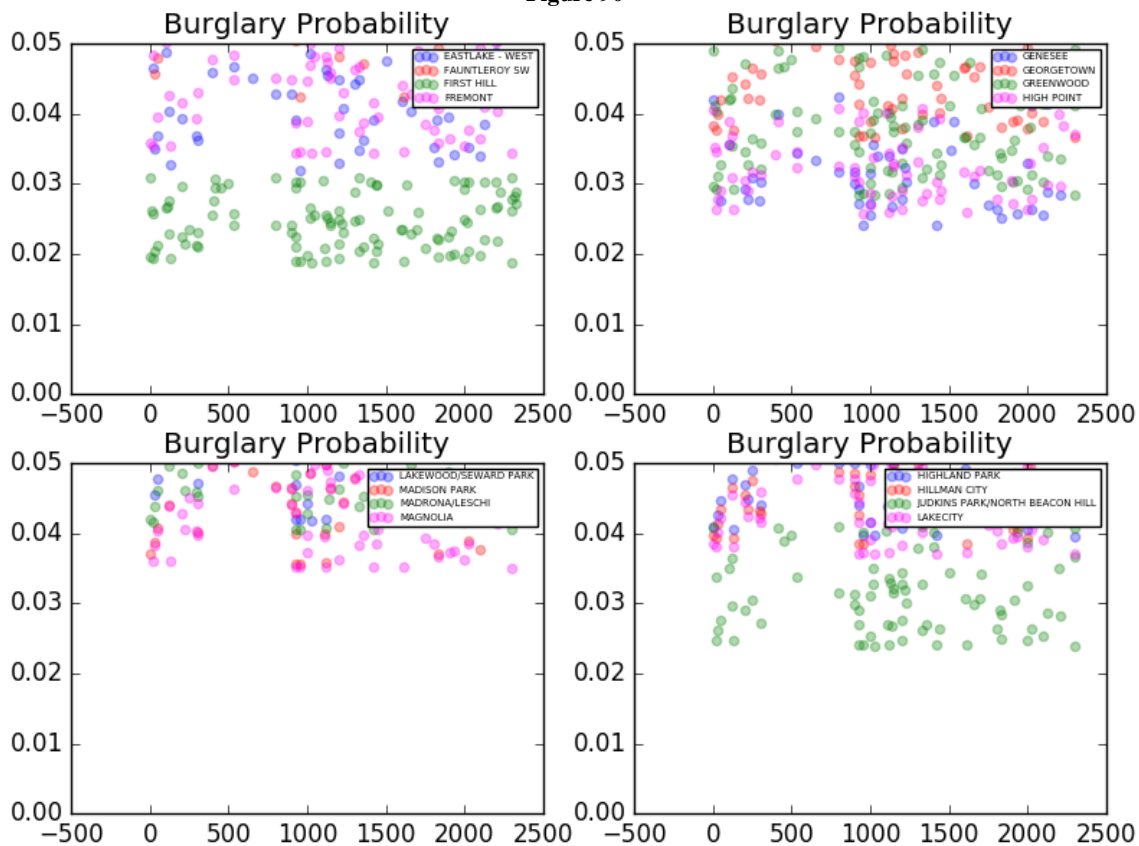


Figure 91

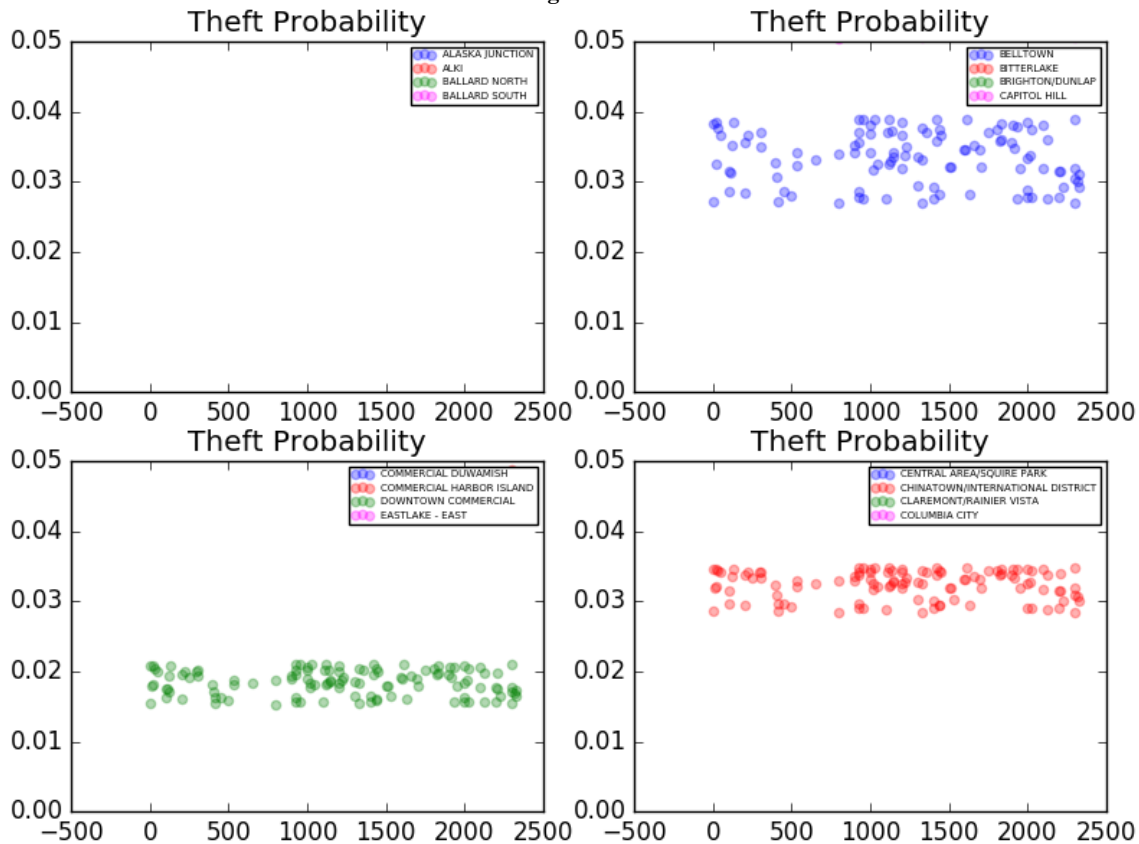


Figure 92

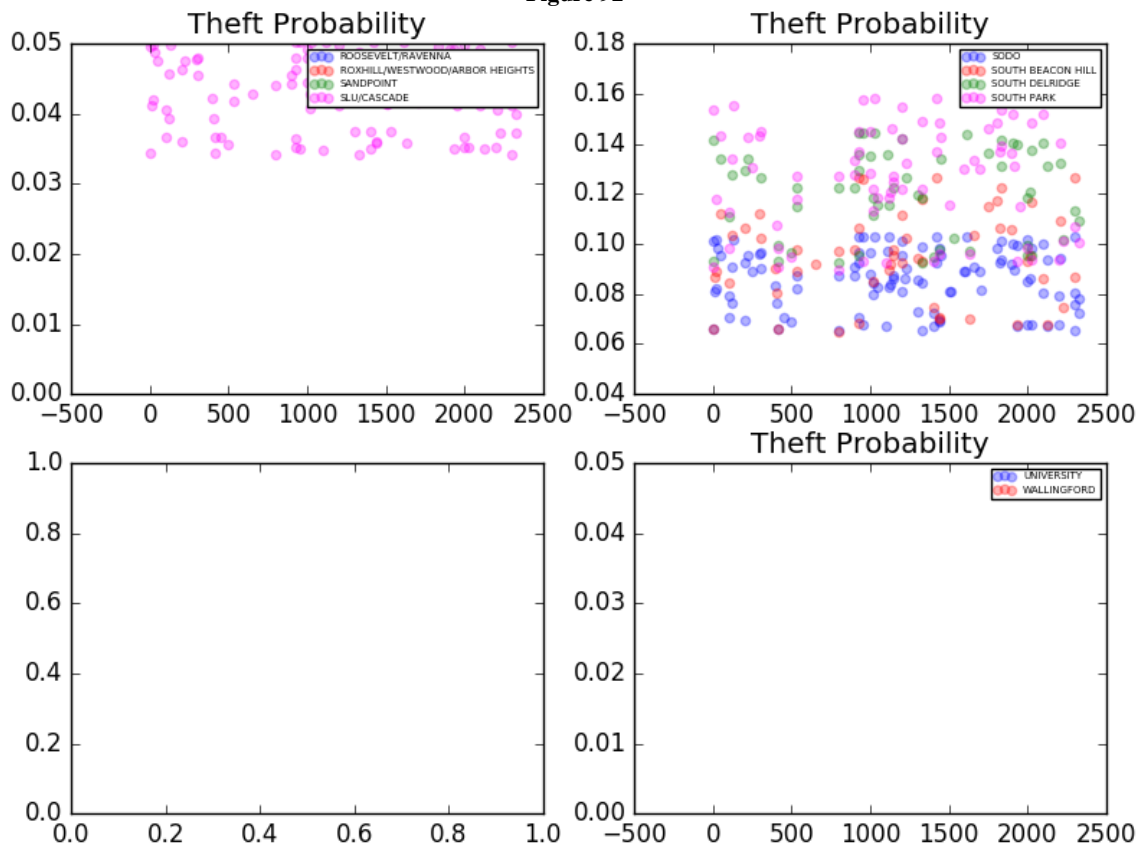


Figure 93

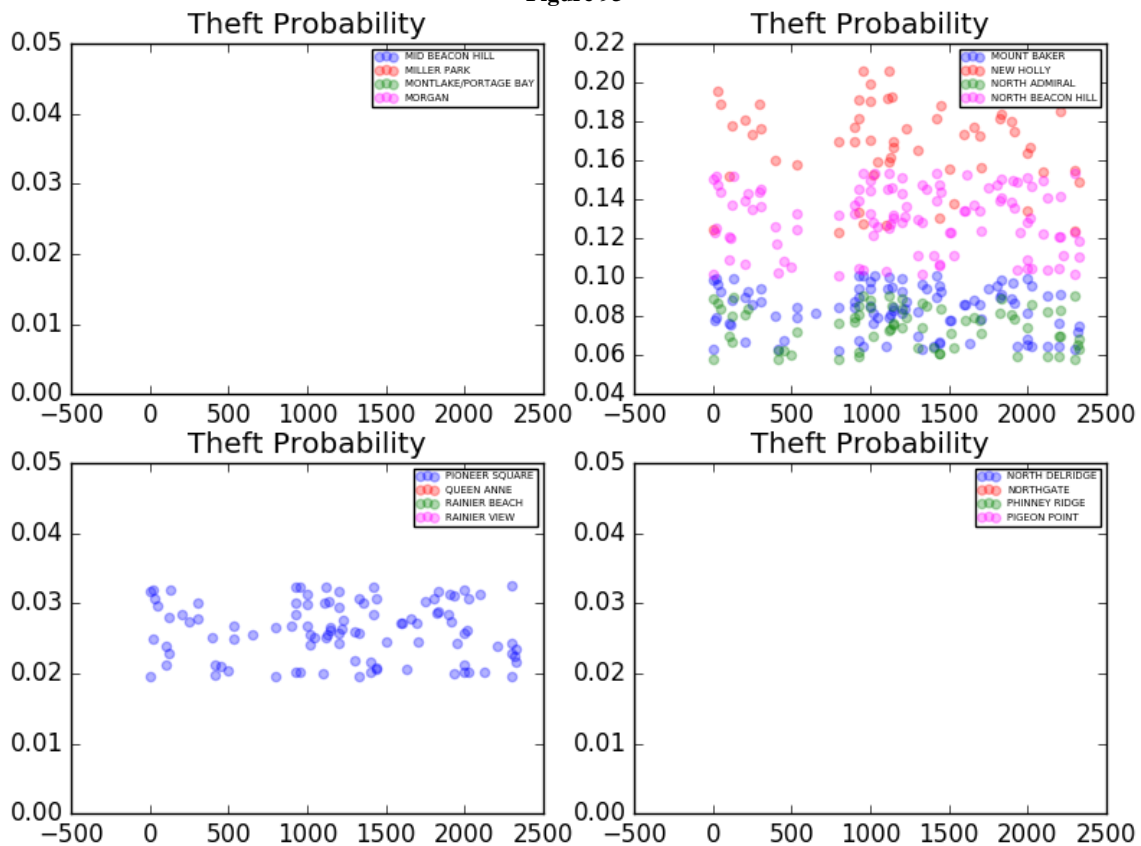


Figure 94

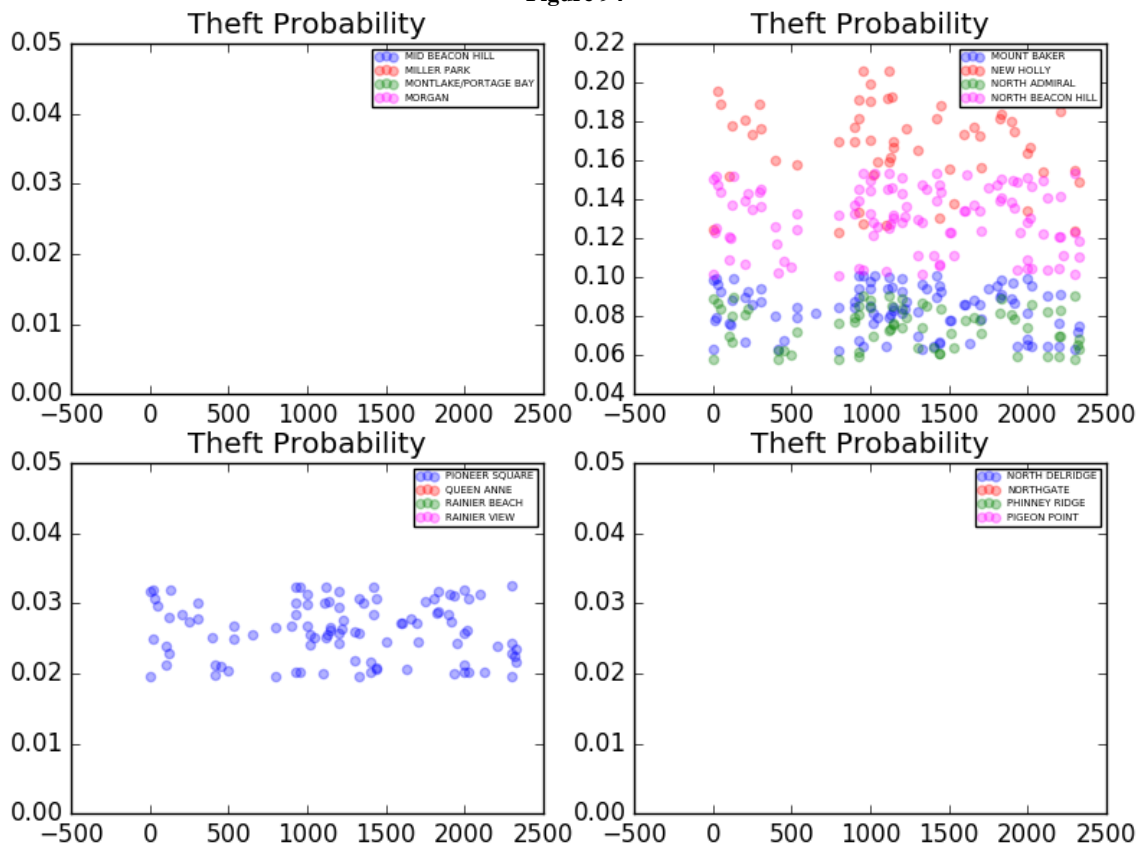


Figure 95

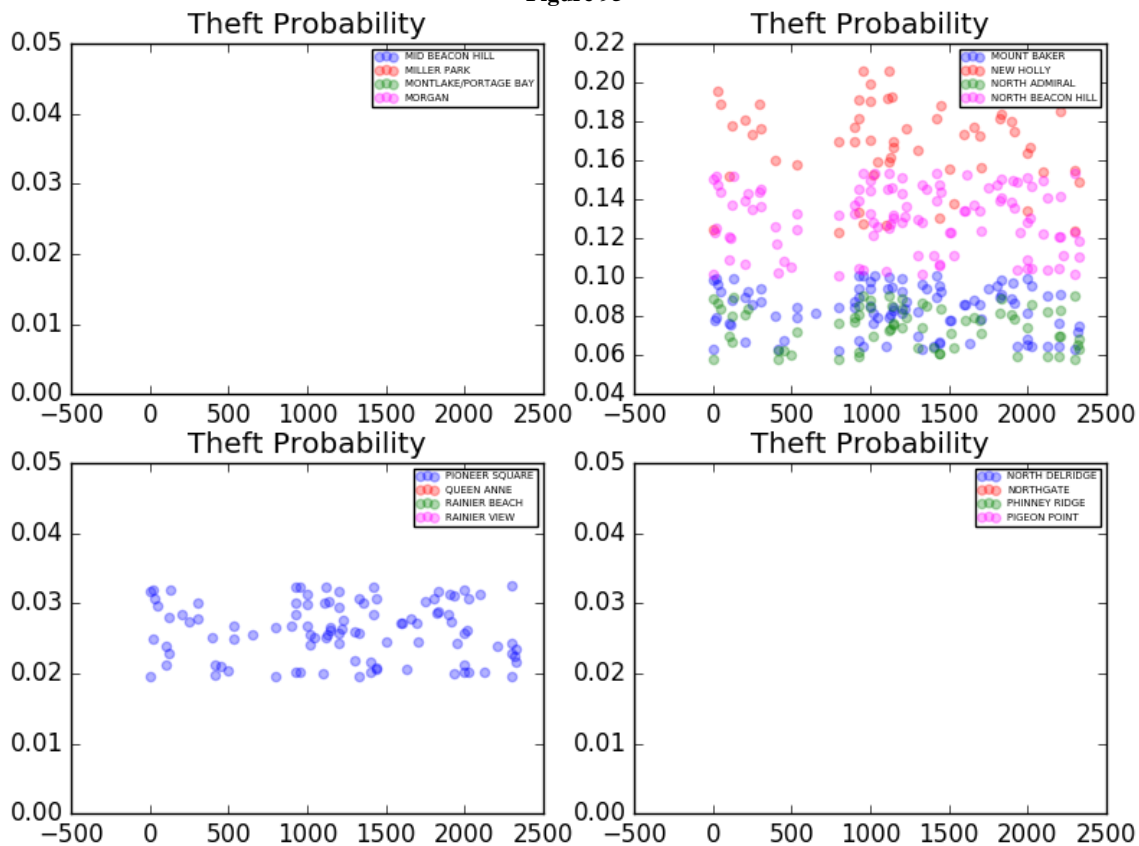


Figure 96

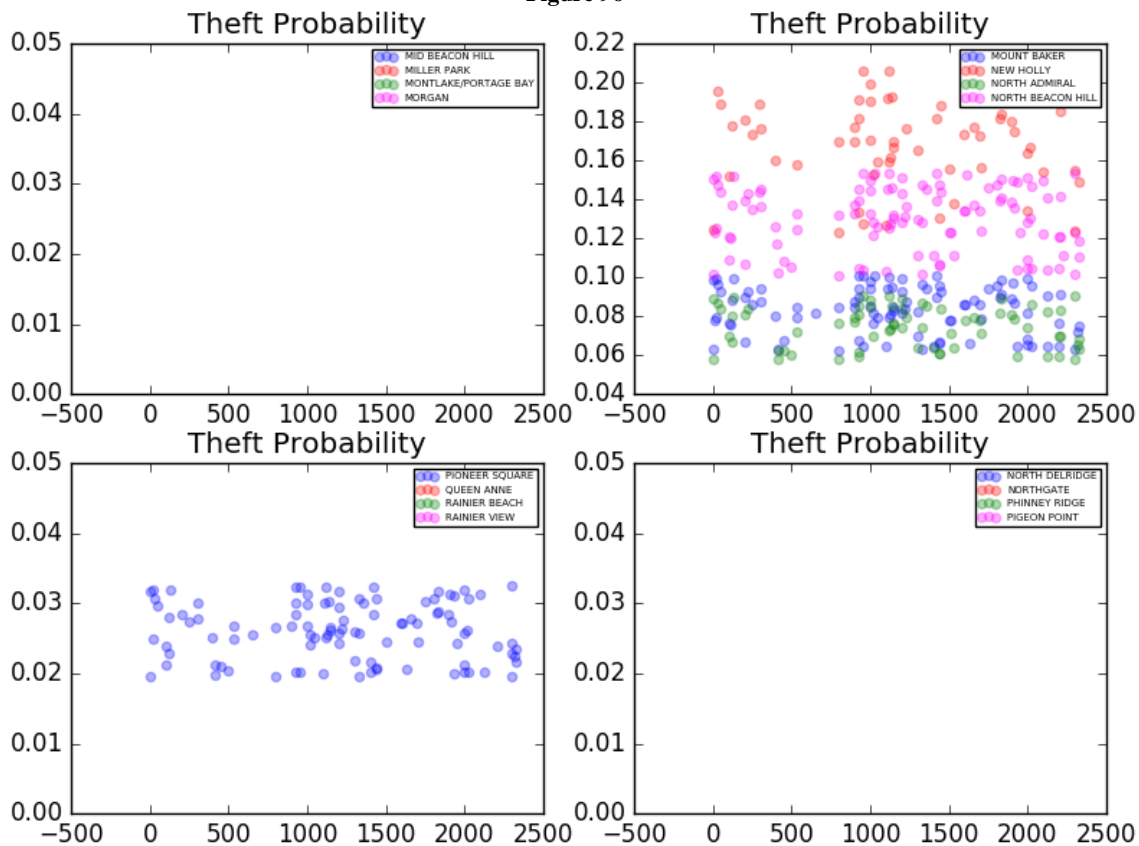


Figure 97

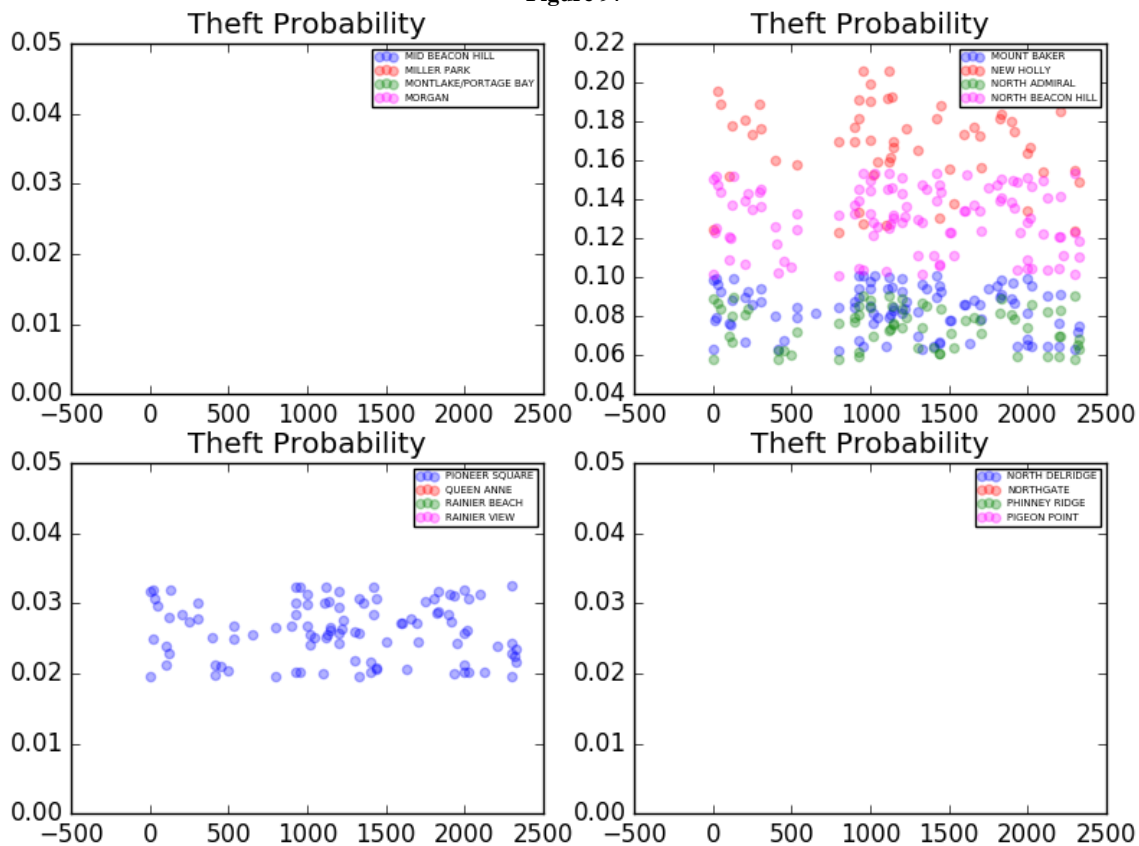


Figure 98

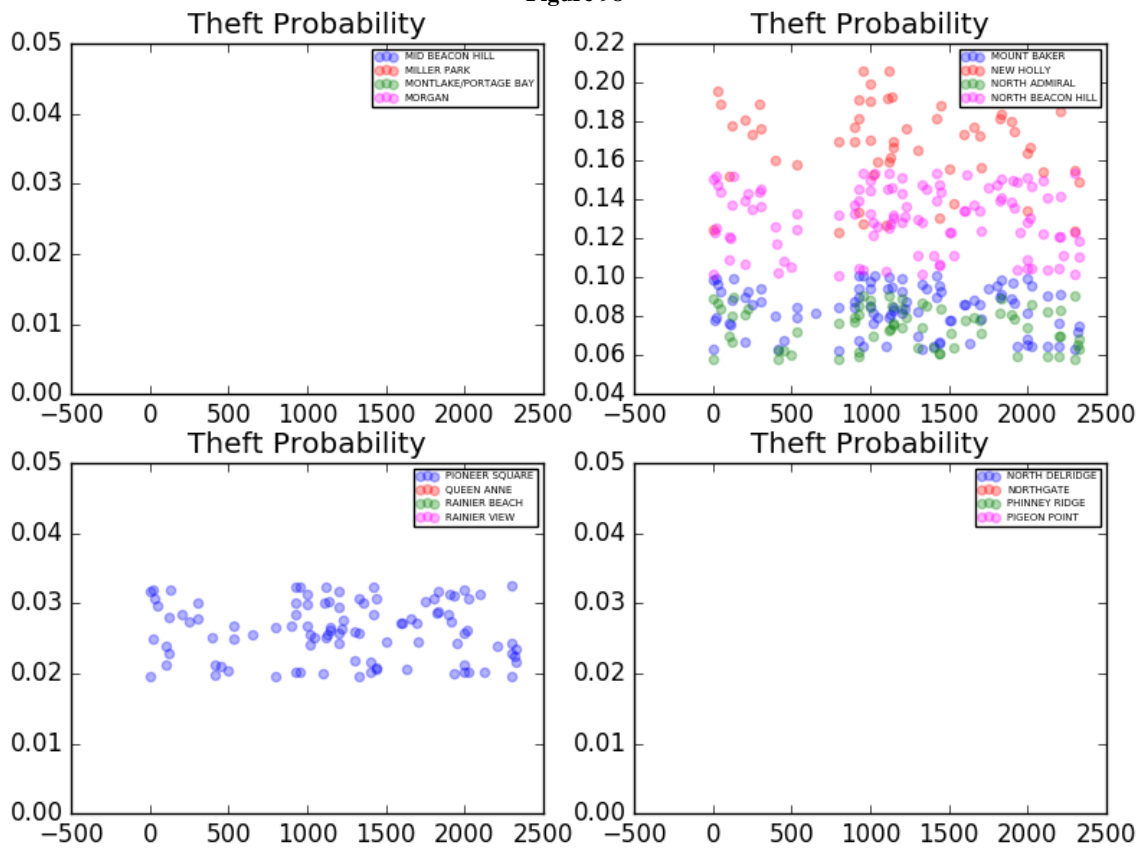


Figure 99

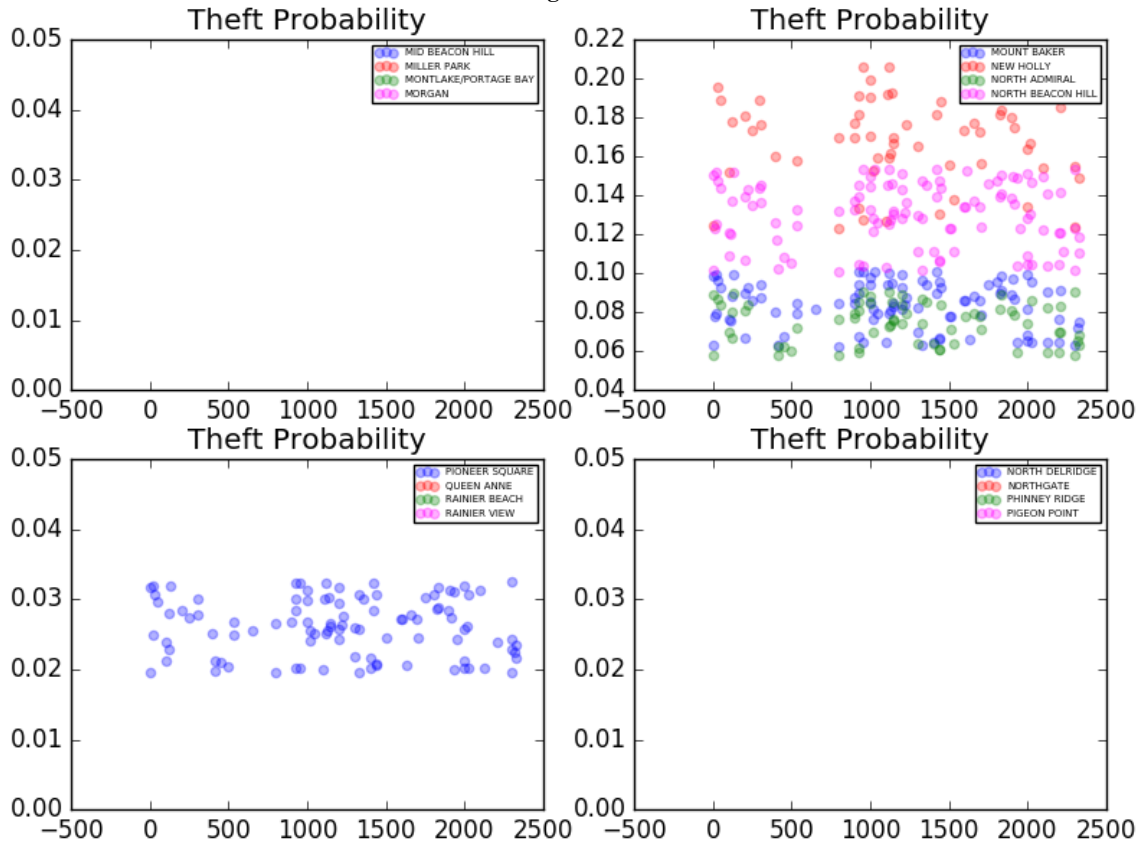


Figure 100

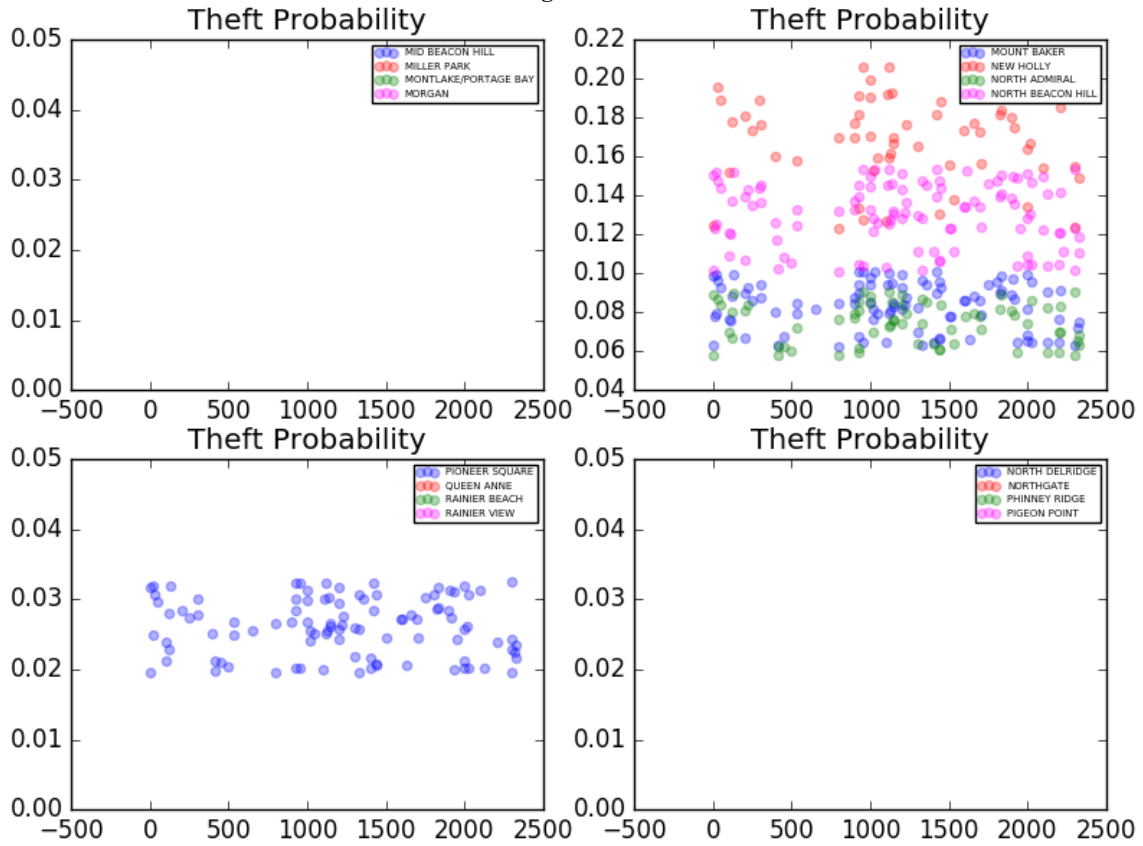


Figure 101

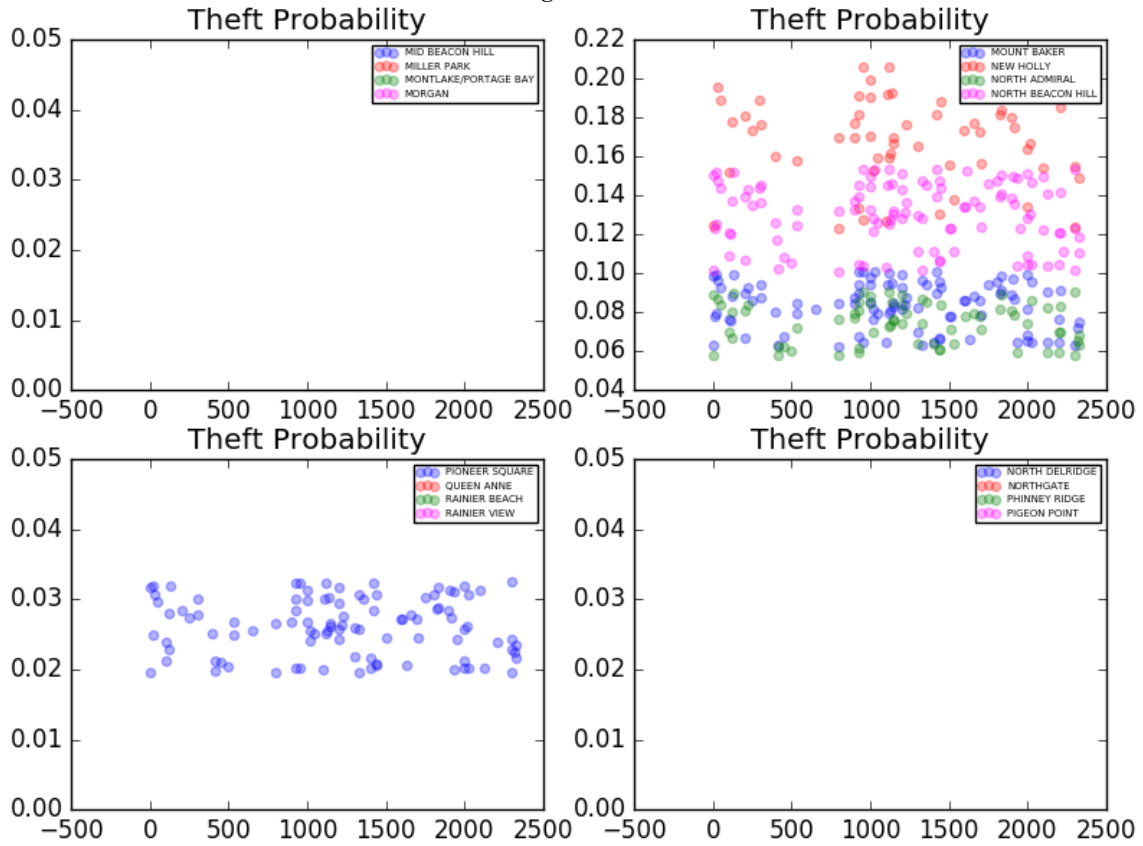


Figure 102

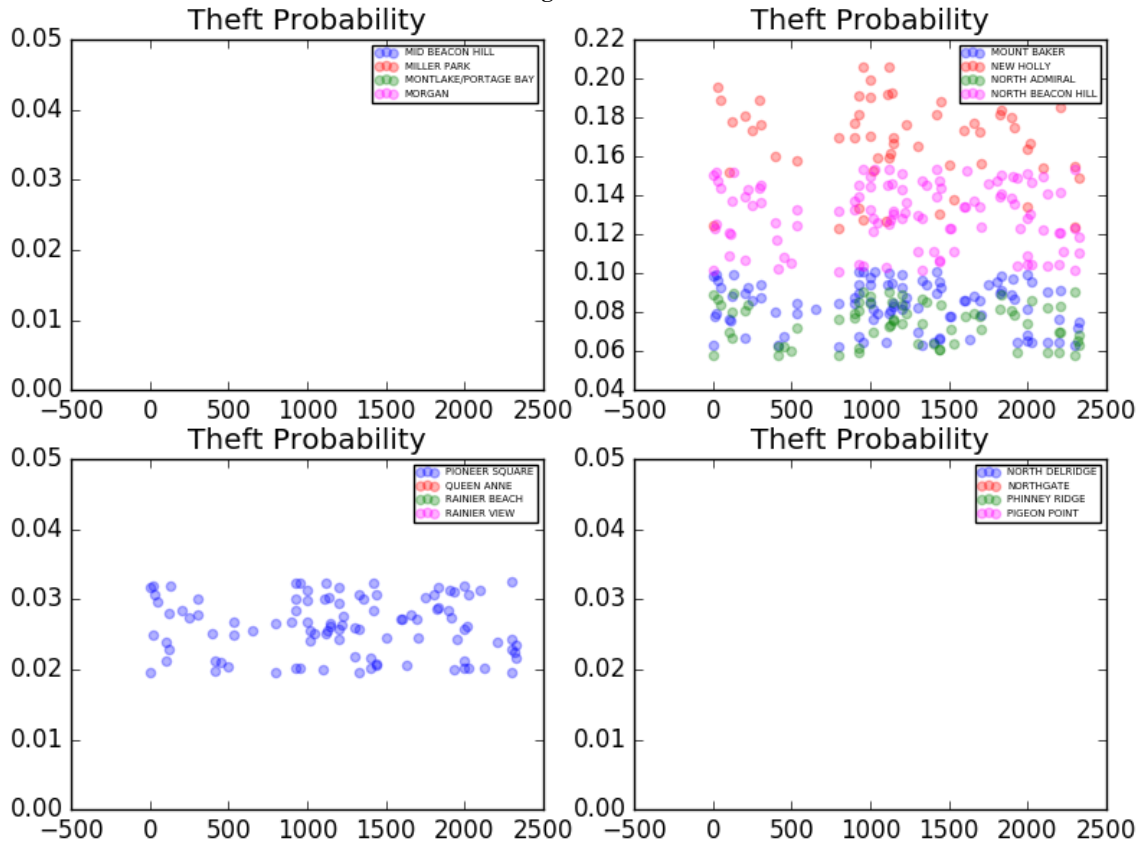


Figure 103

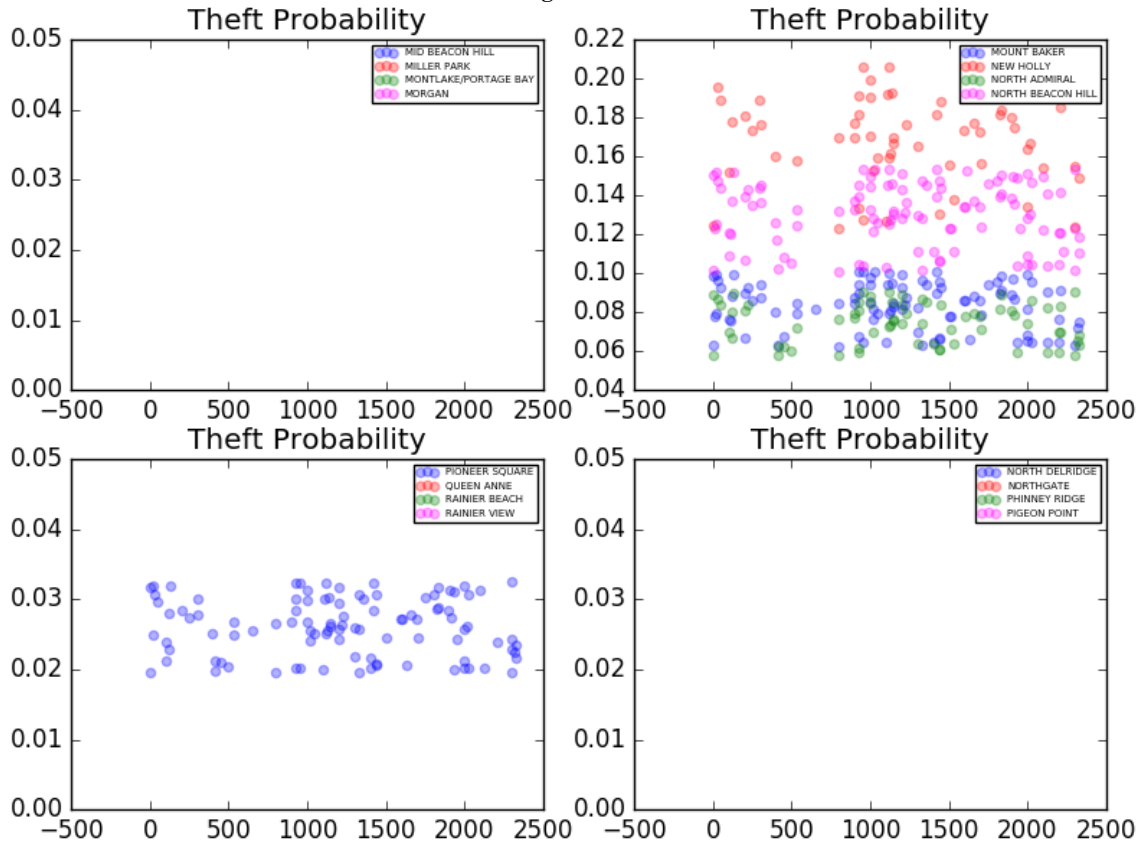


Figure 104

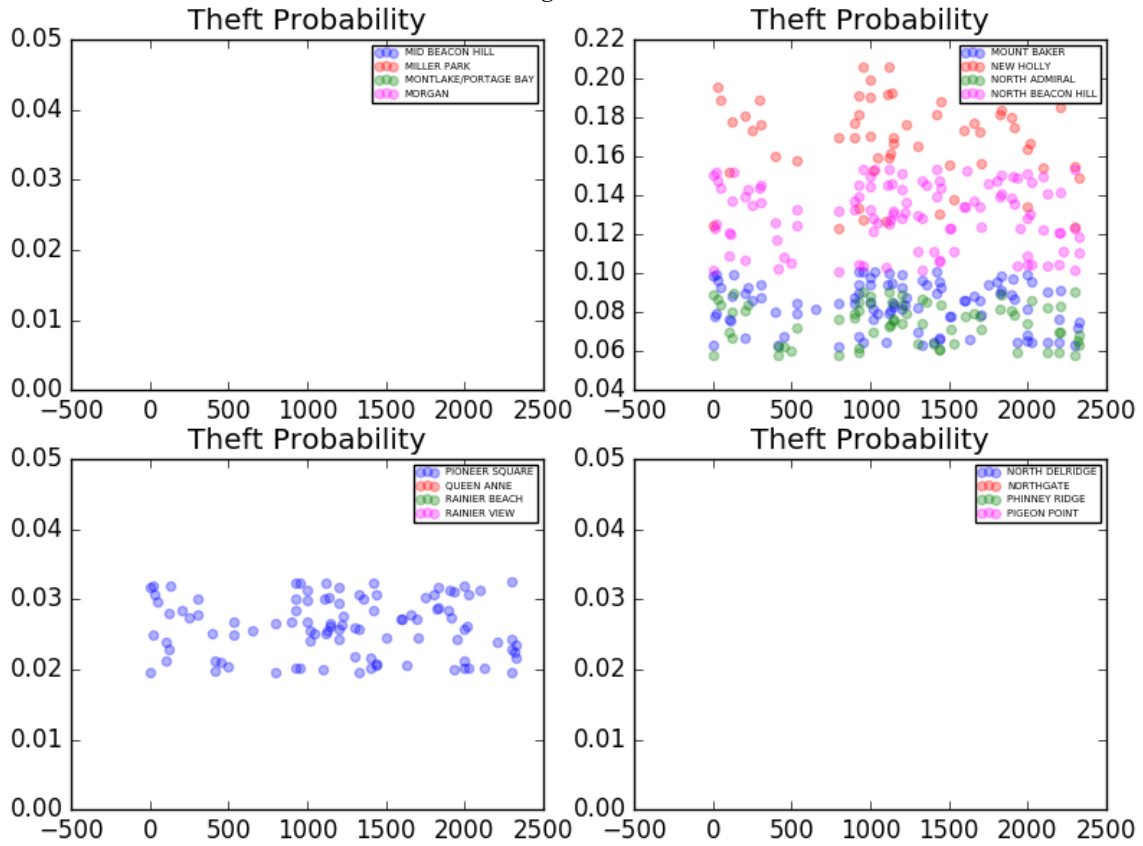


Figure 105

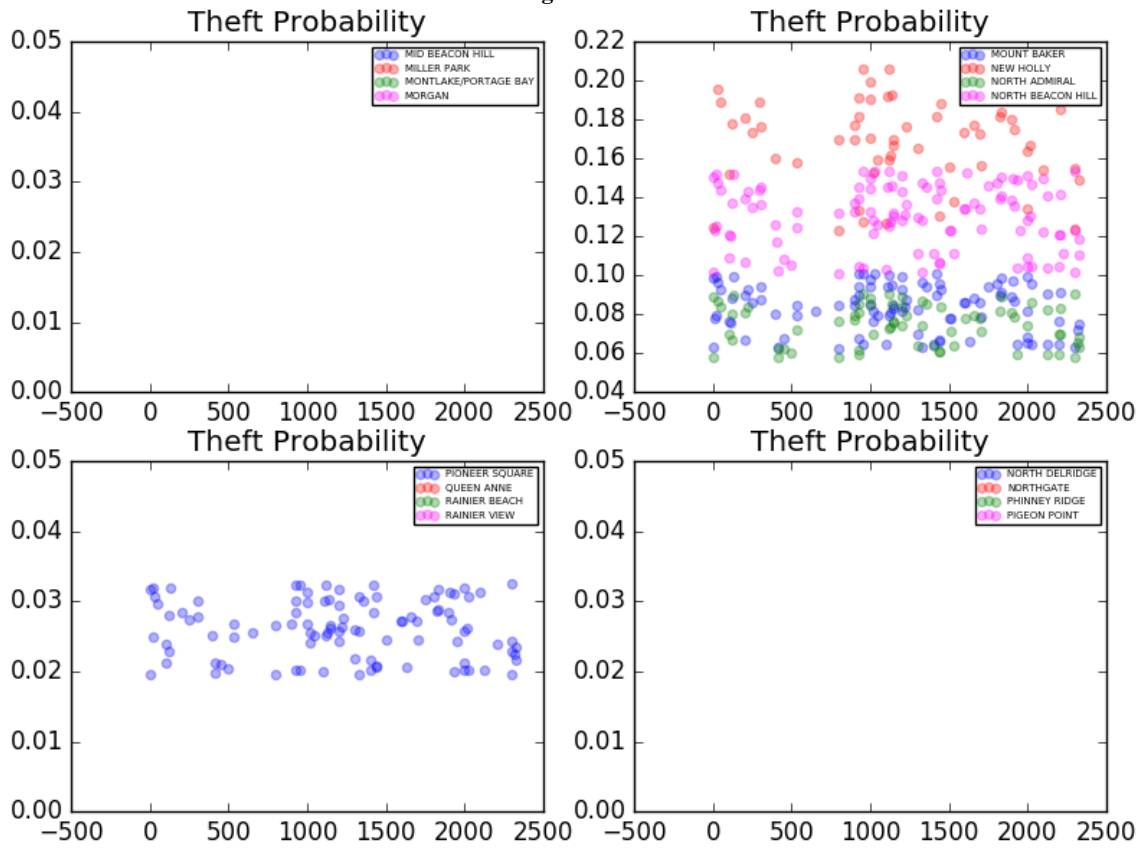


Figure 106

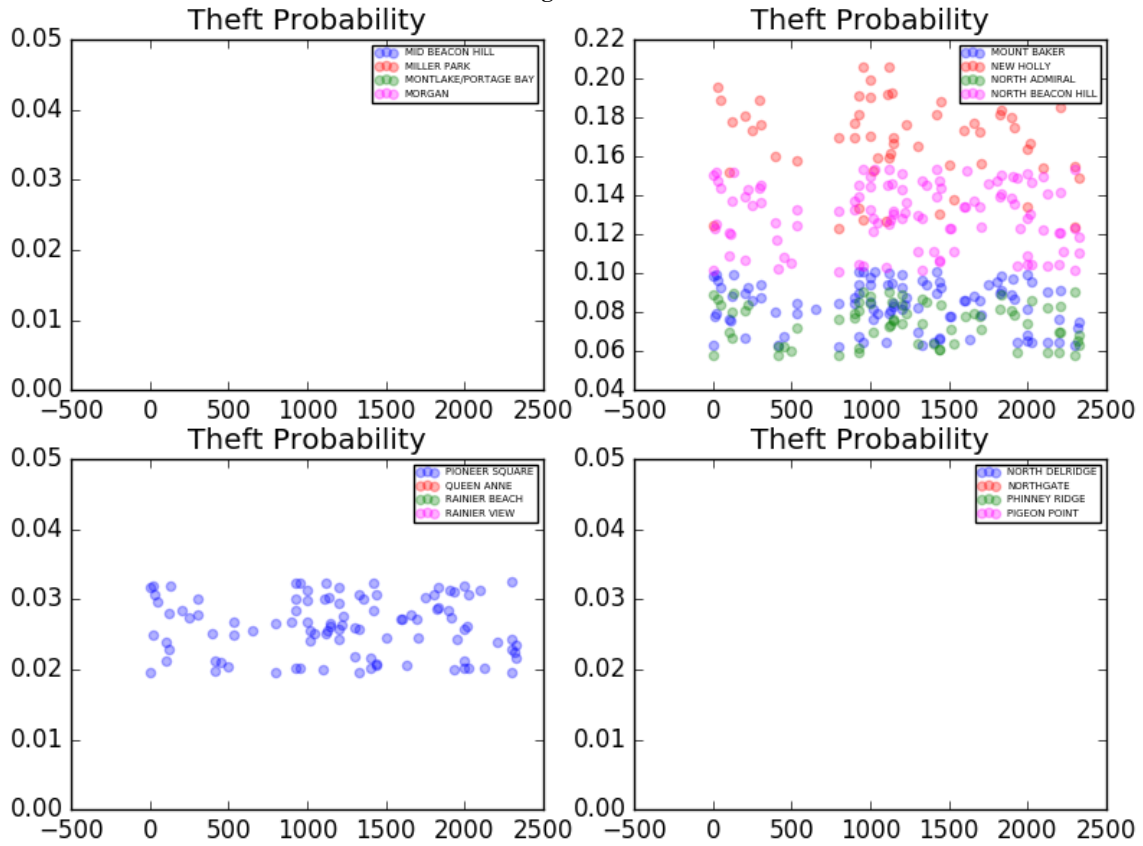


Figure 107

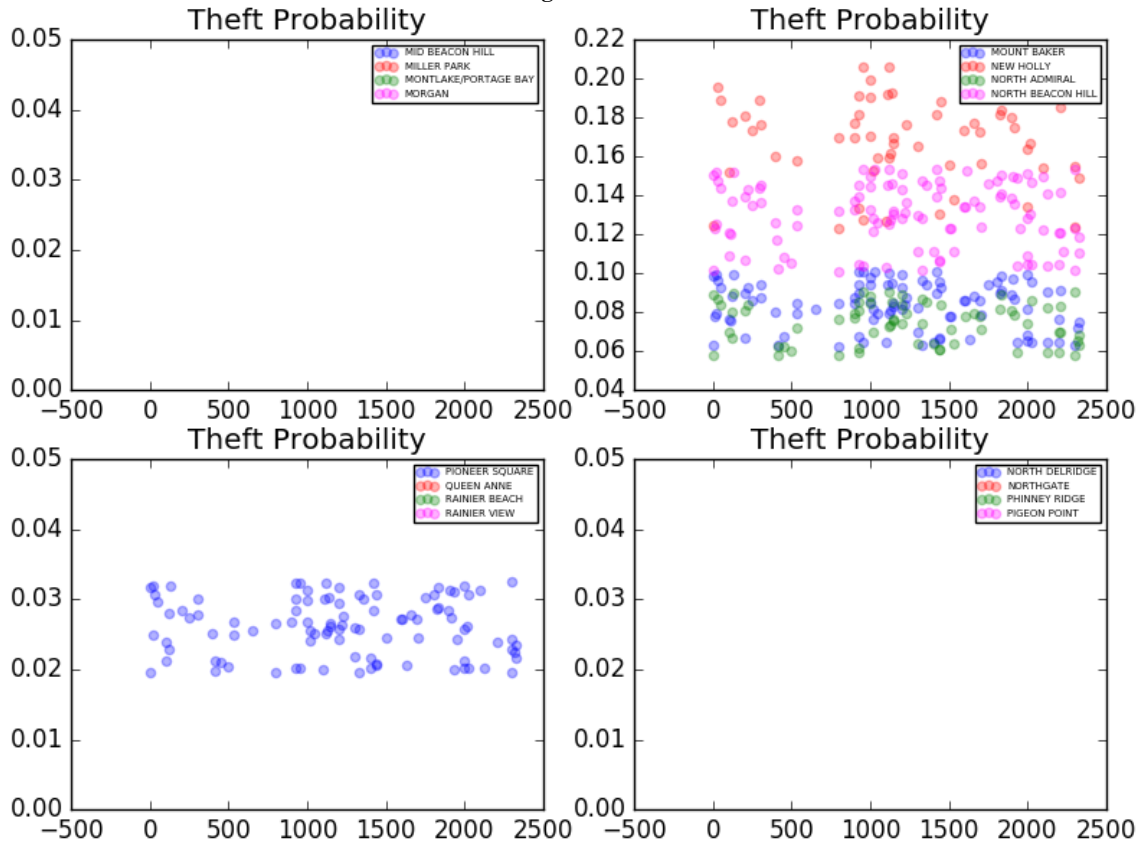


Figure 108

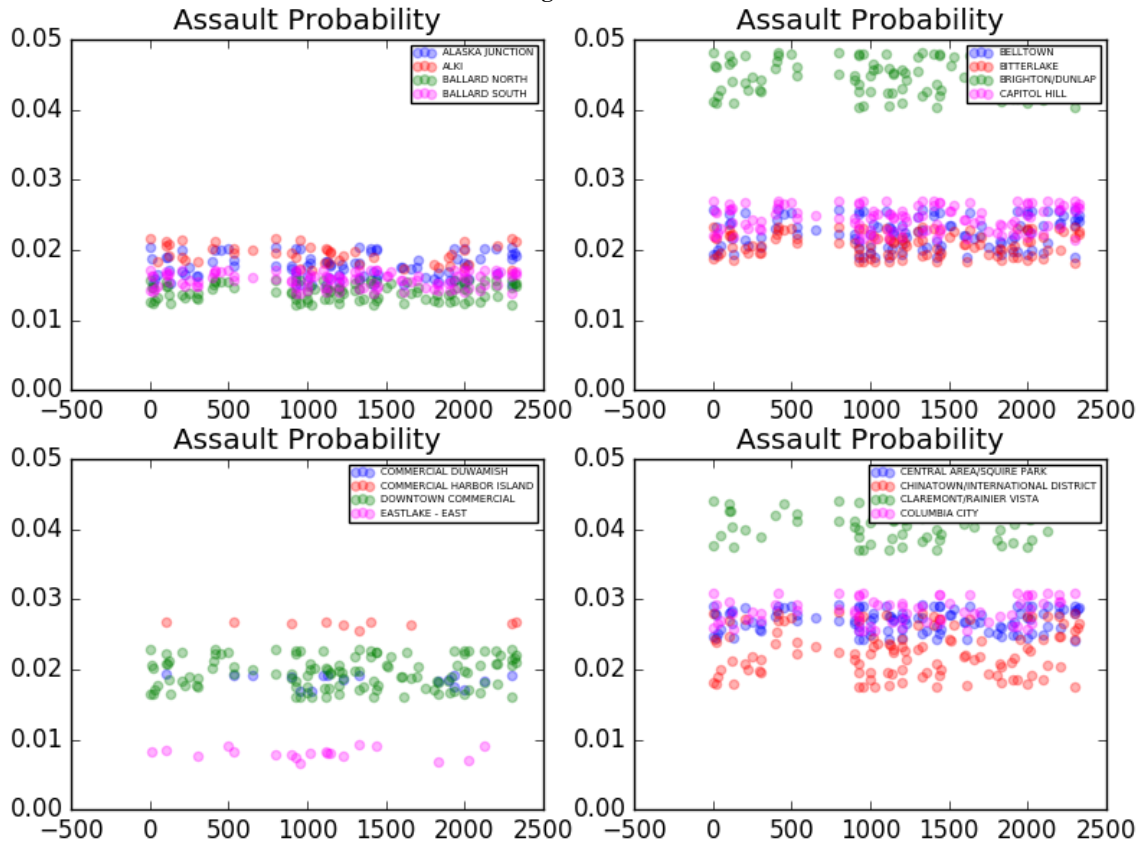


Figure 109

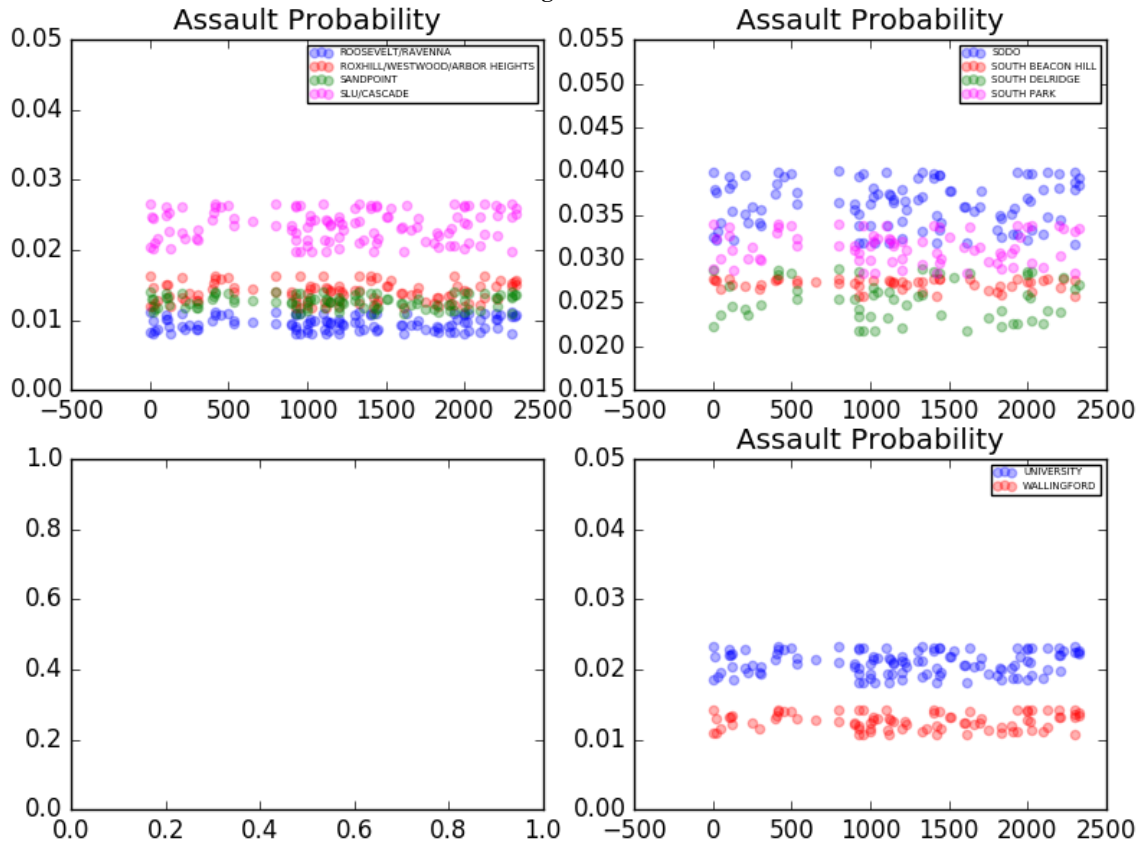


Figure 110

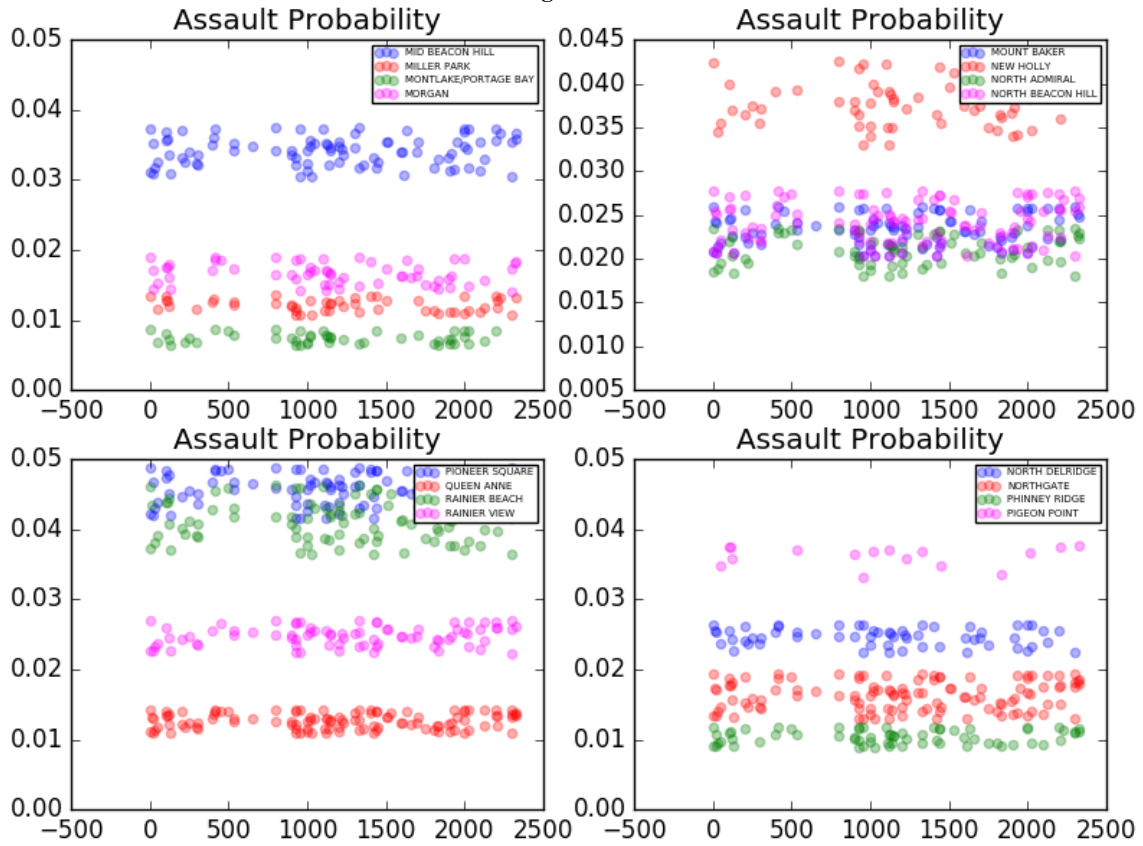


Figure 111

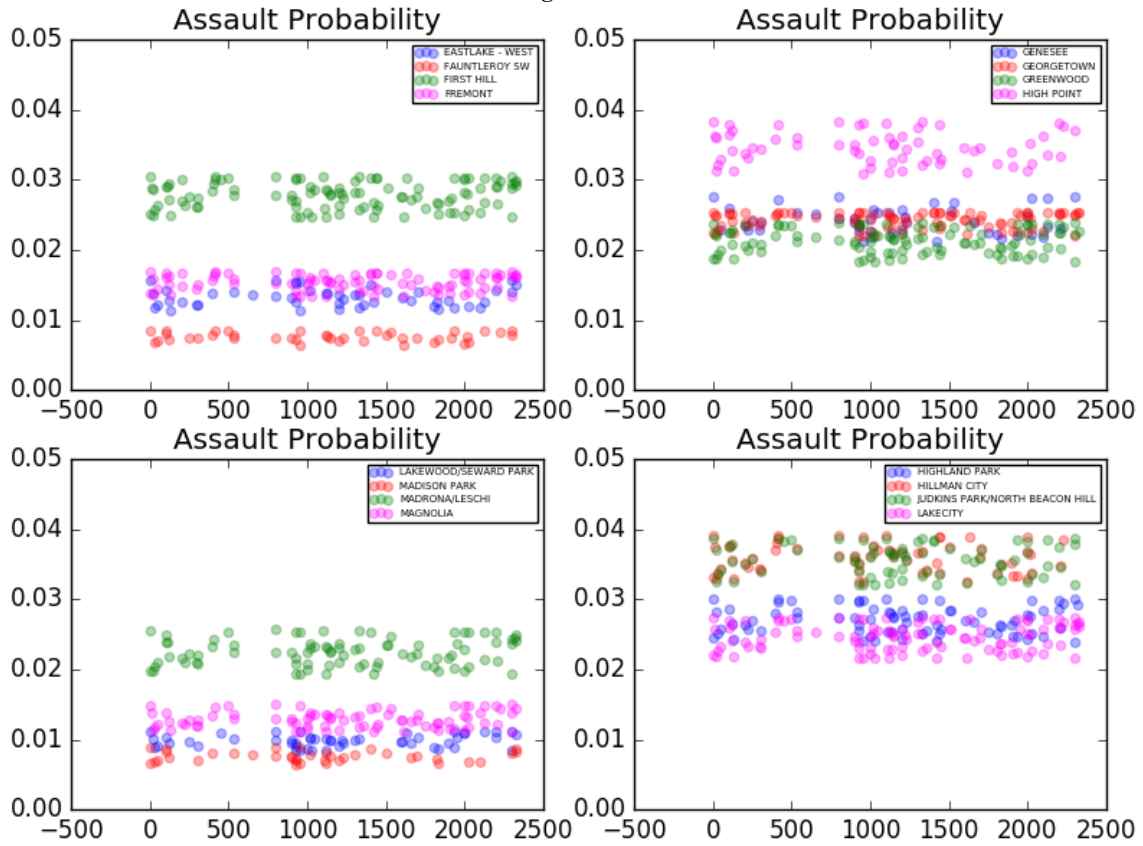


Figure 112

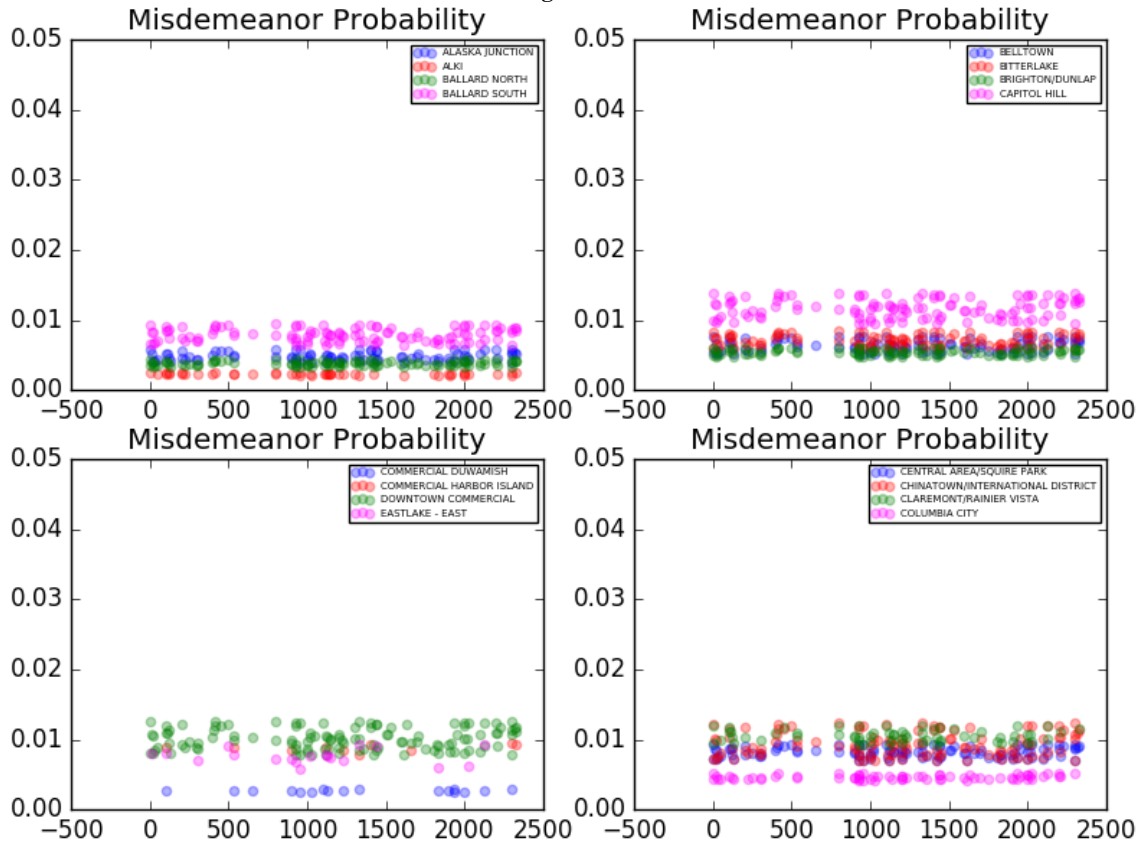


Figure 113

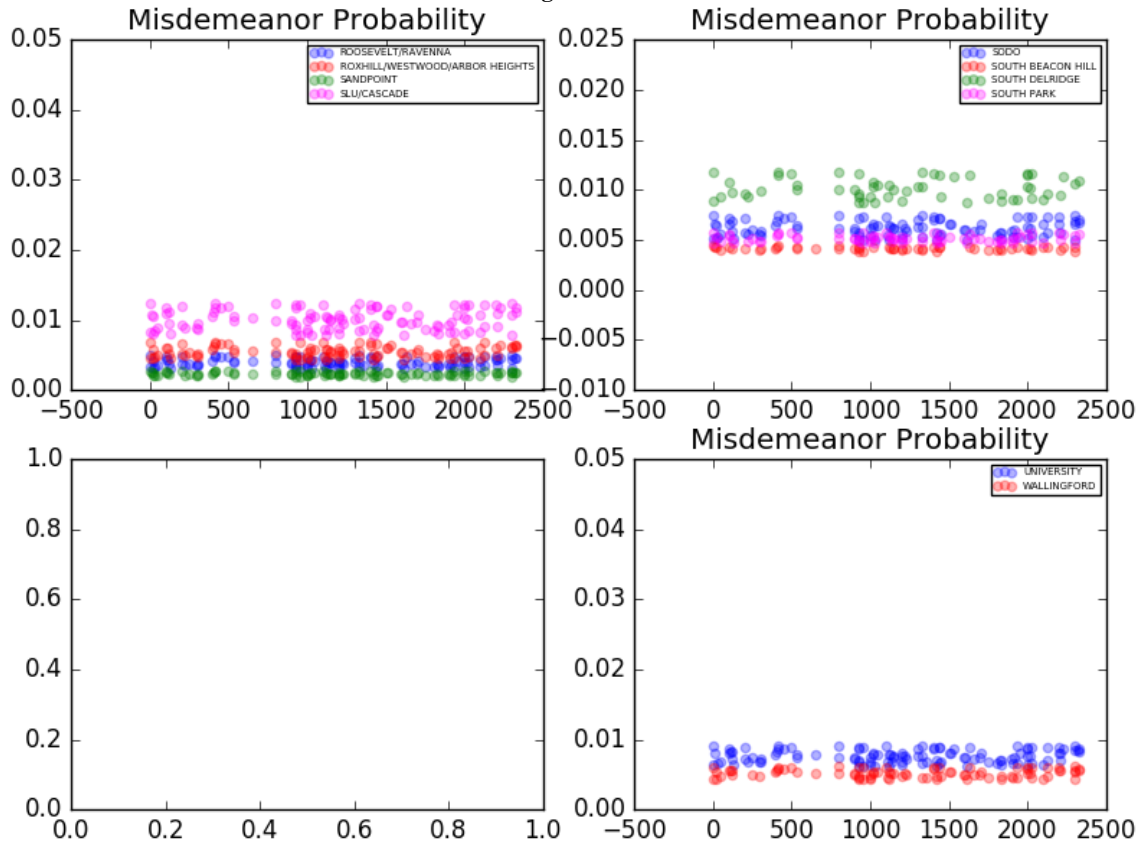


Figure 114

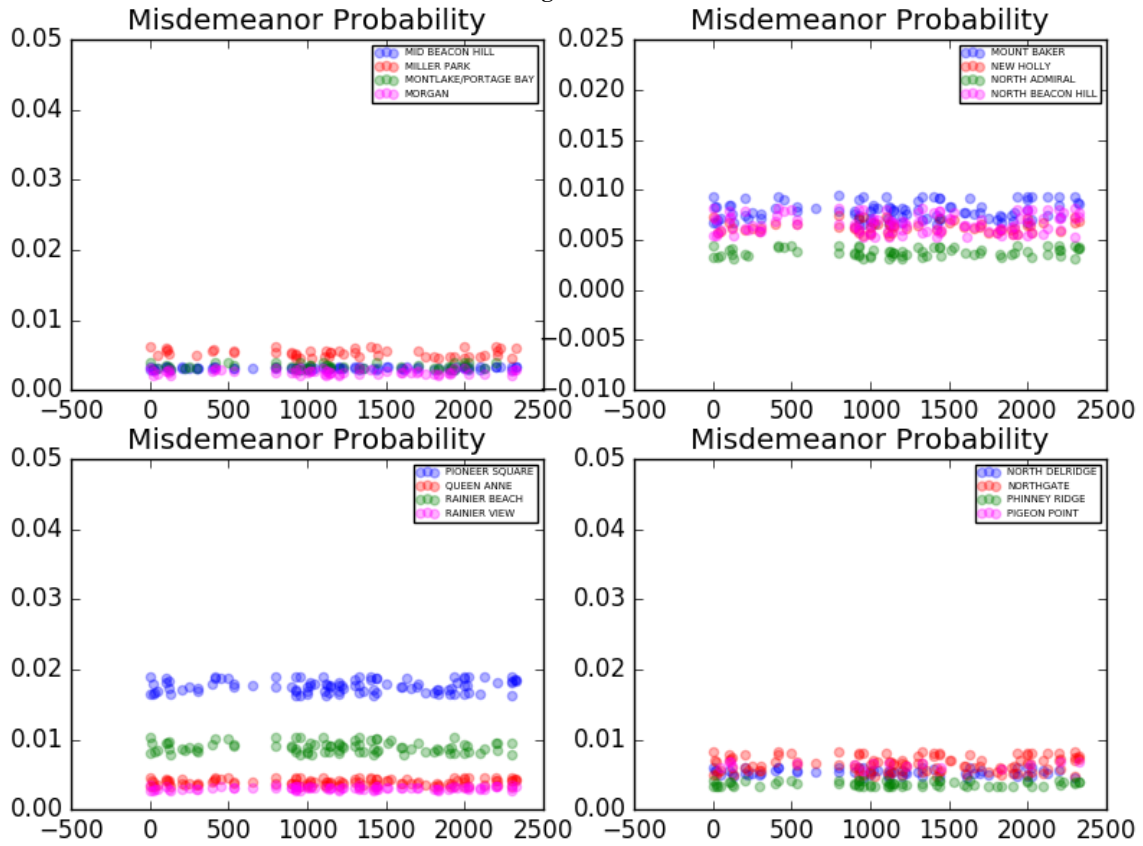


Figure 115

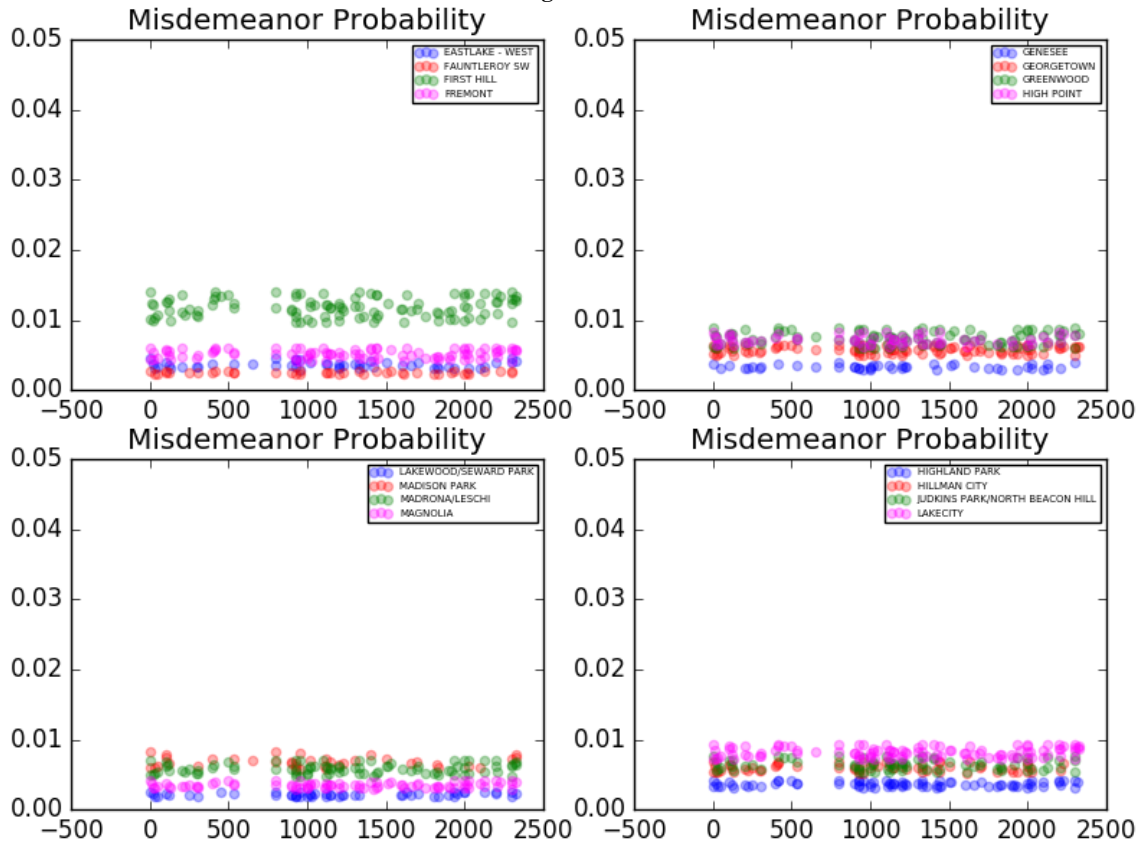


Figure 116

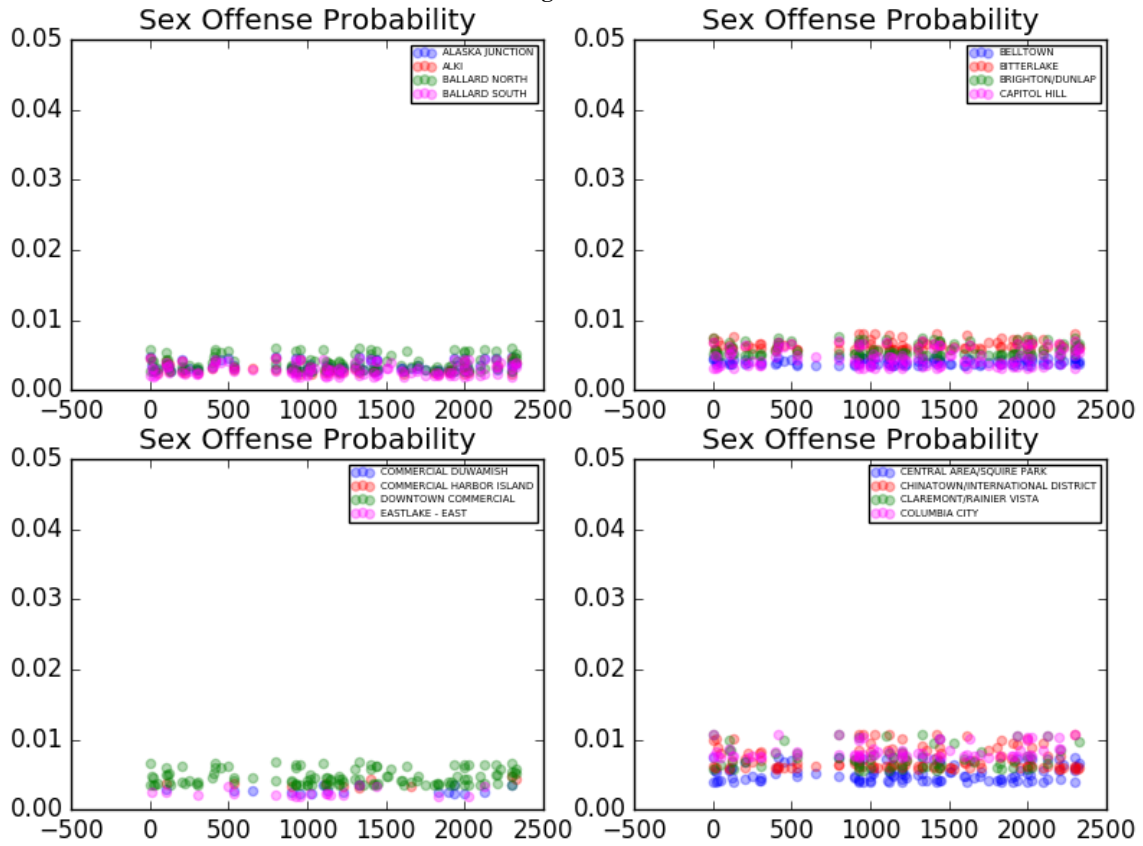


Figure 117

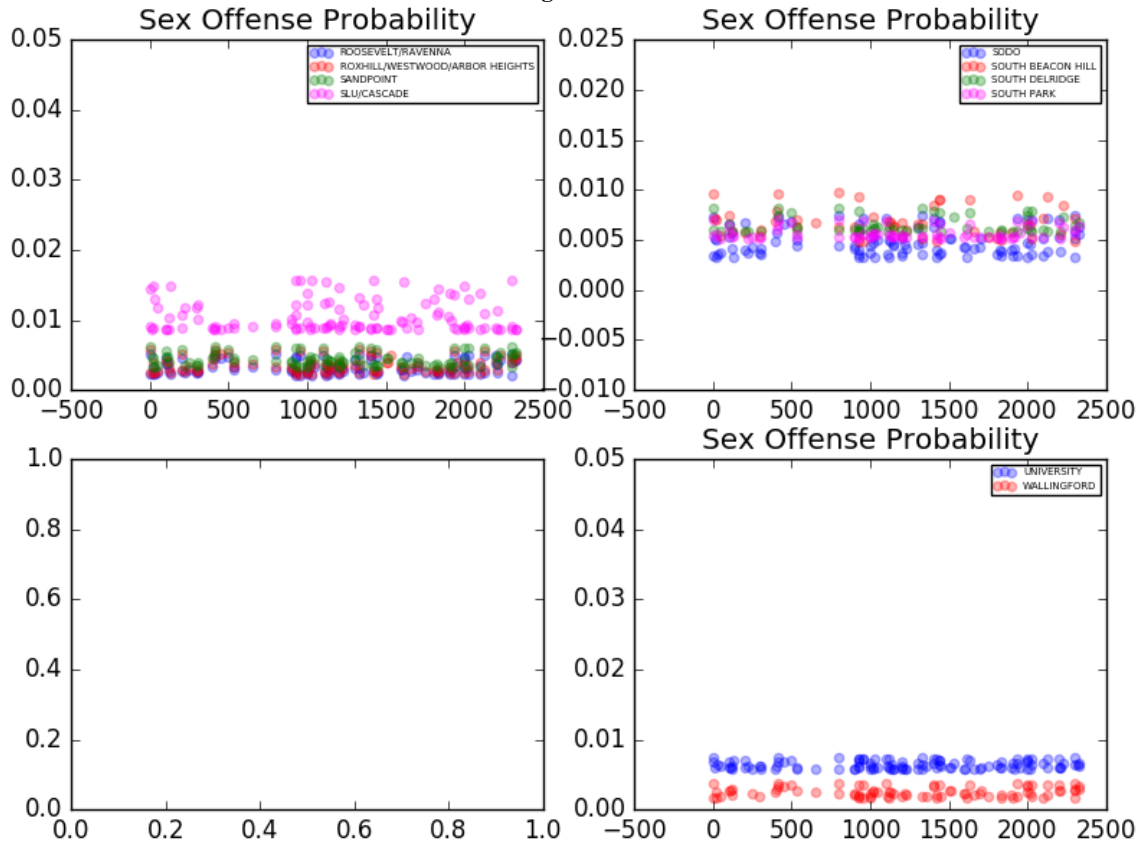


Figure 118

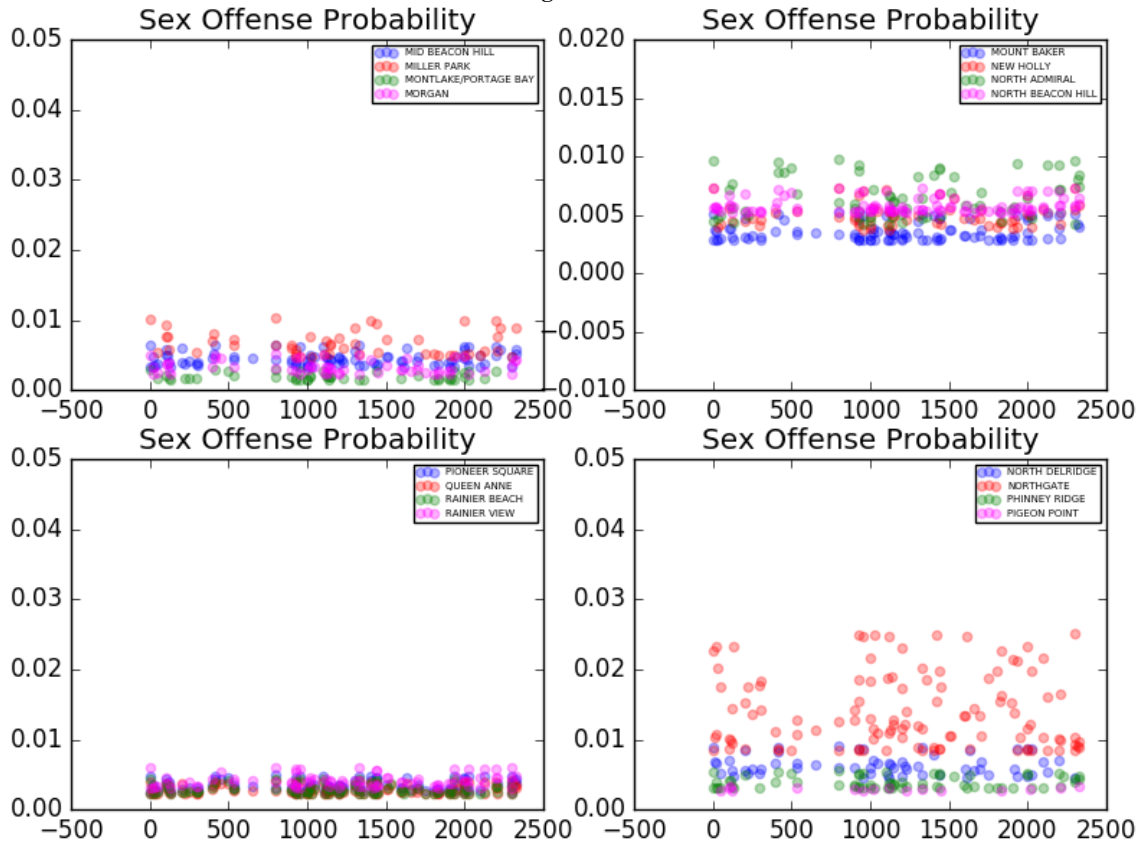


Figure 119

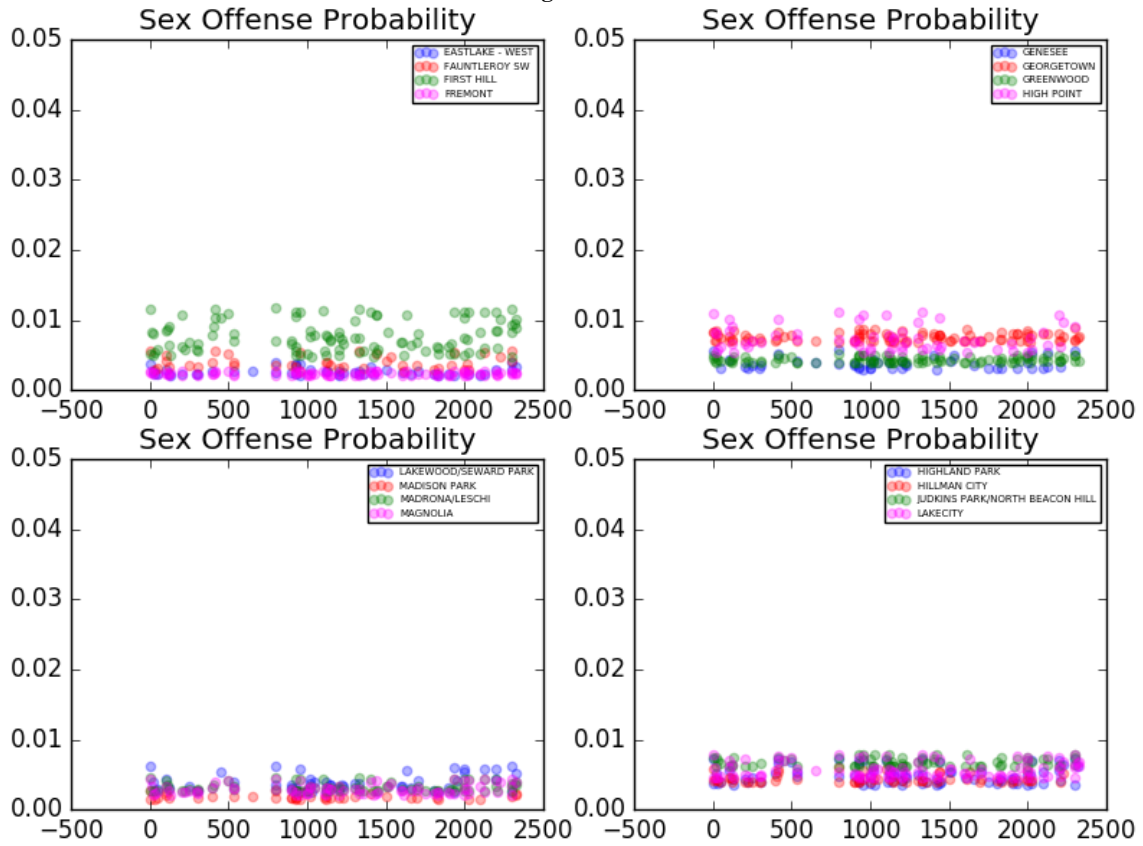


Figure 120
Robbery Probability

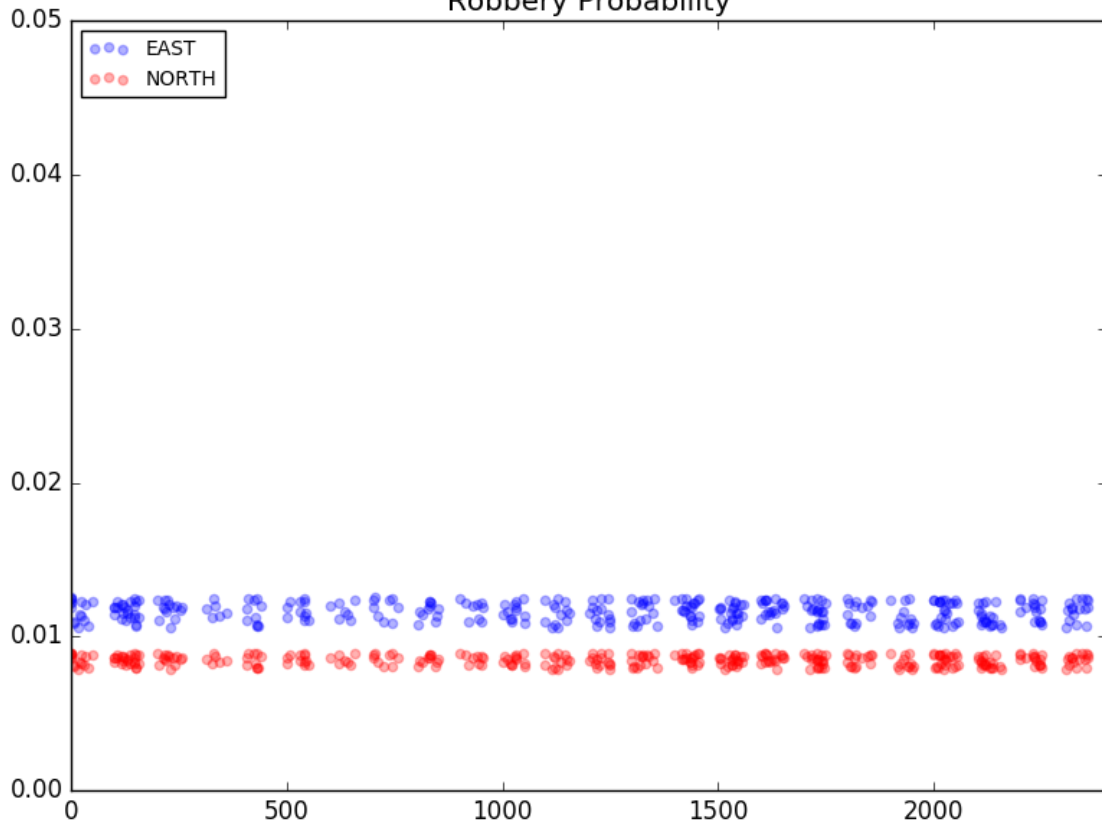


Figure 121
Robbery Probability

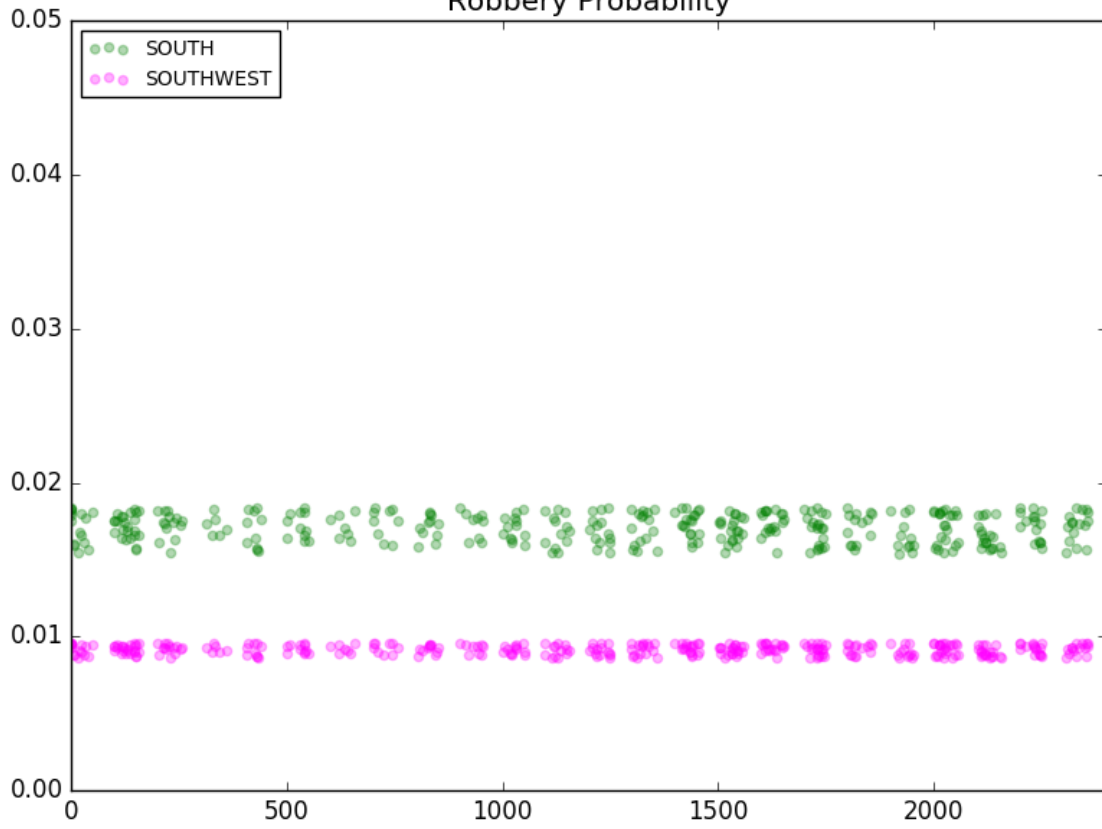


Figure 122
Arson Probability

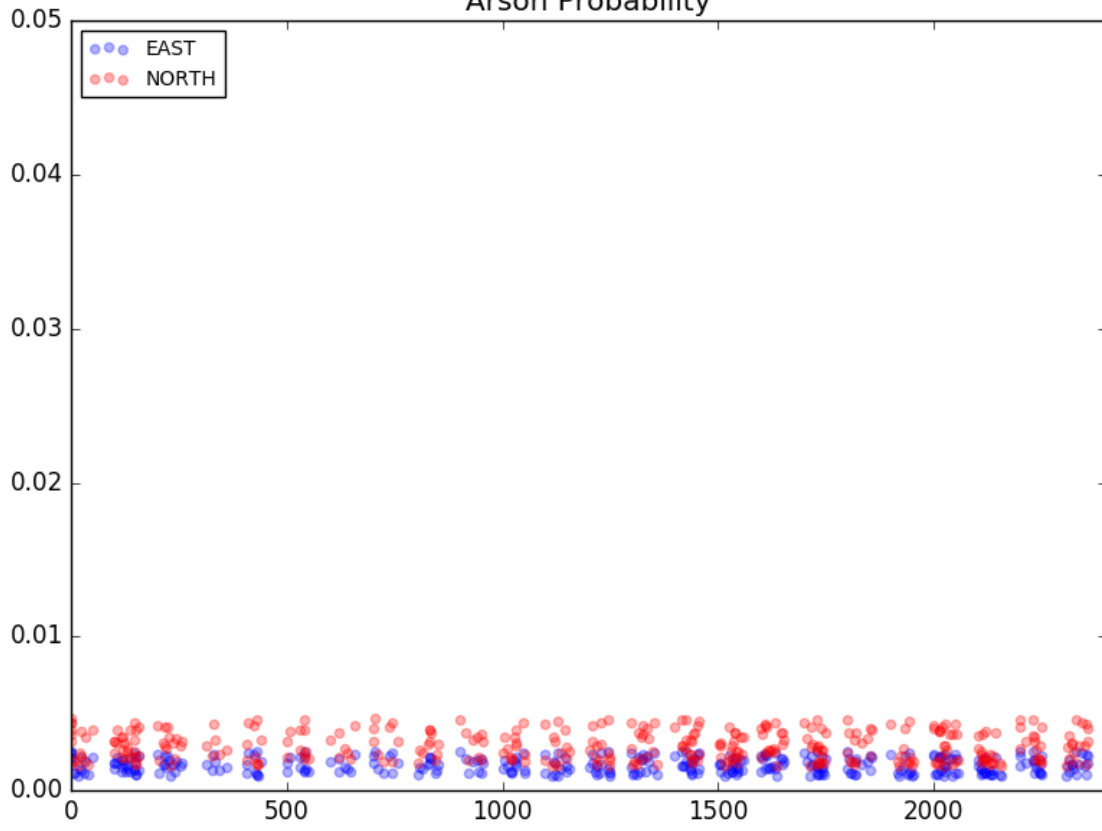


Figure 123
Arson Probability

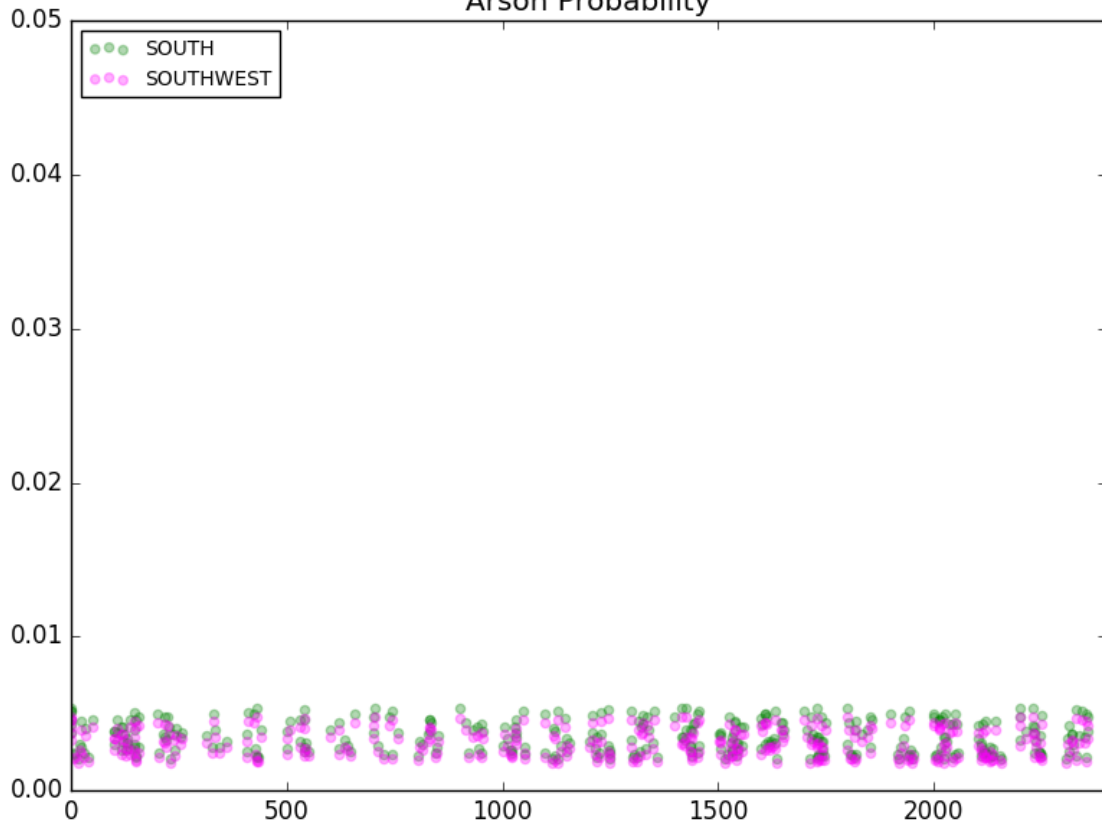


Figure 124
Car Prowl Probability

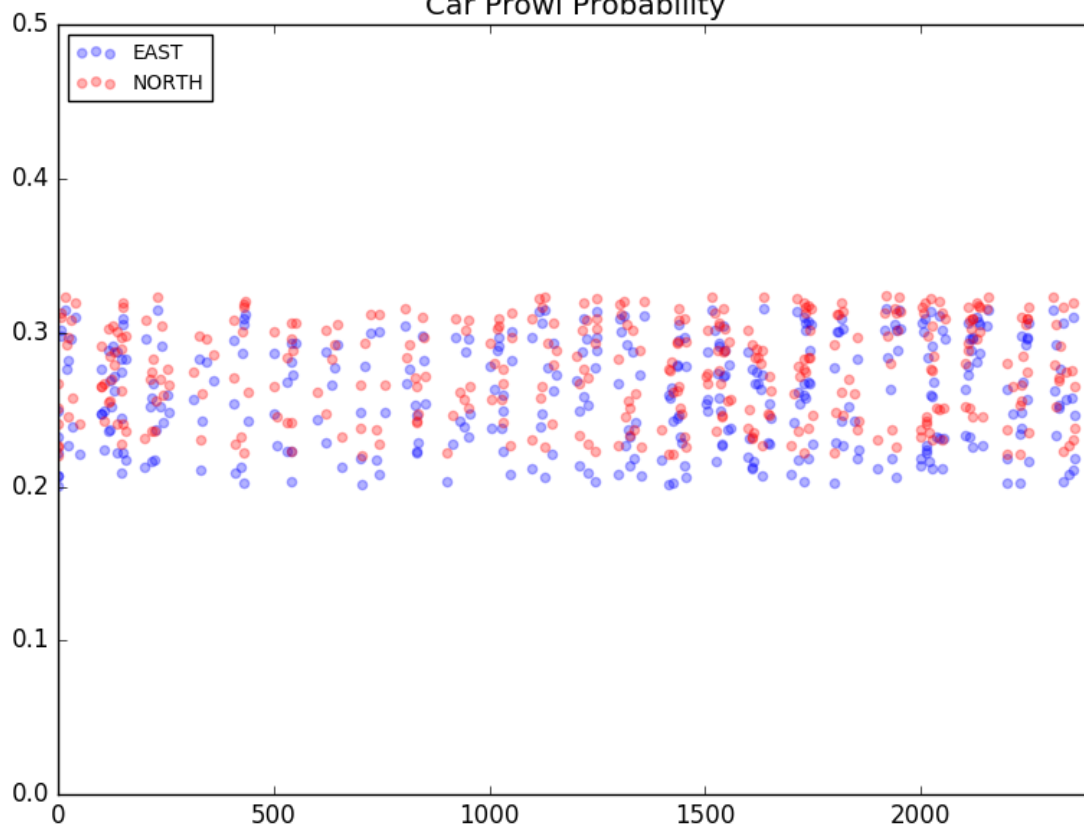


Figure 125
Arson Probability

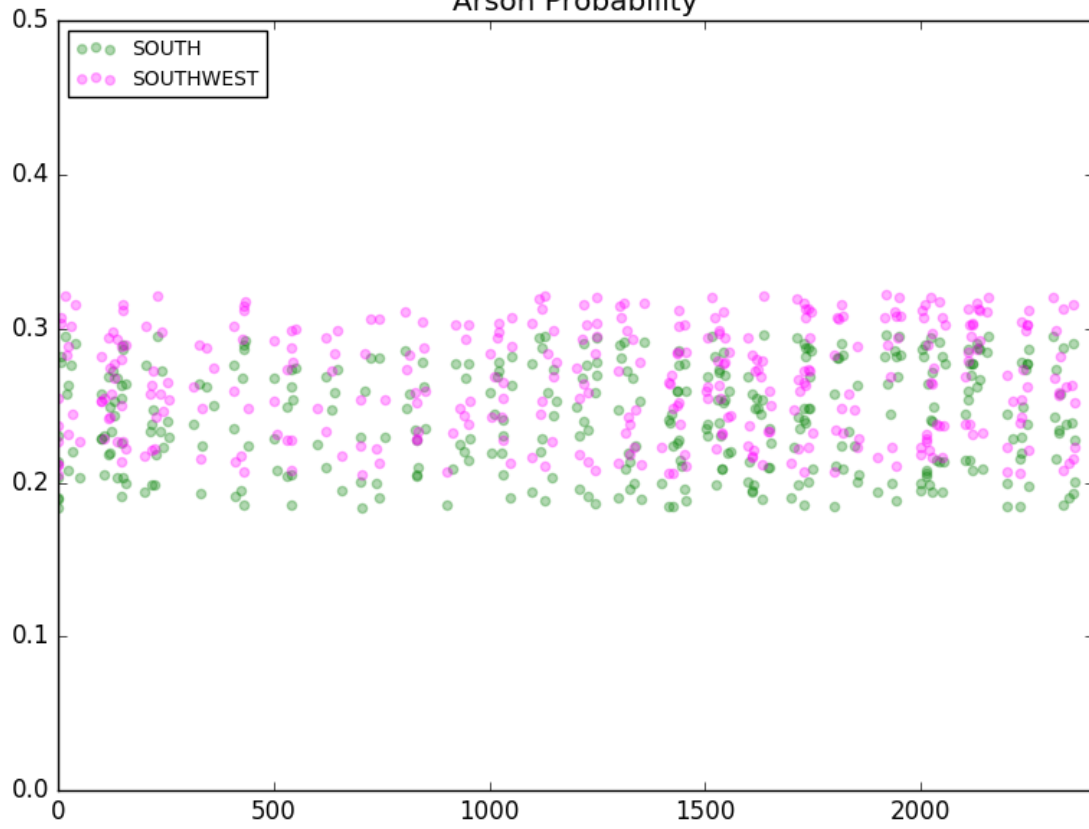


Figure 126
Burglary Probability

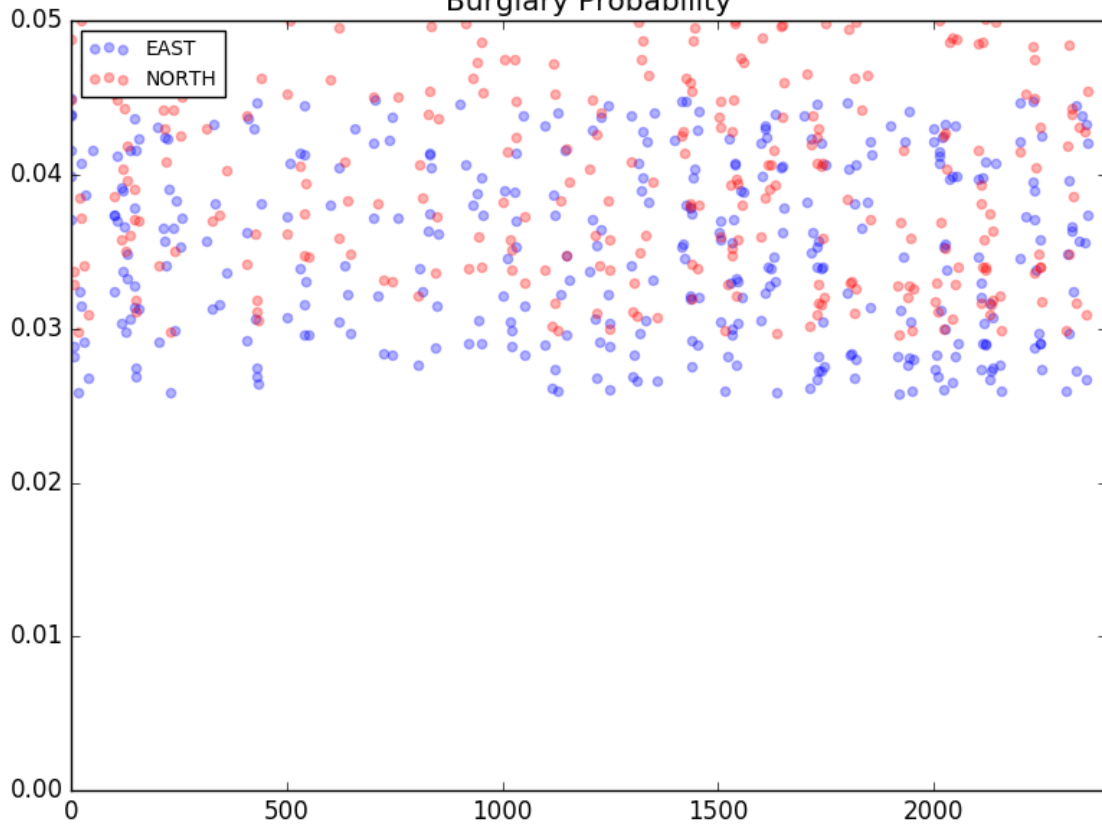


Figure 127
Burglary Probability

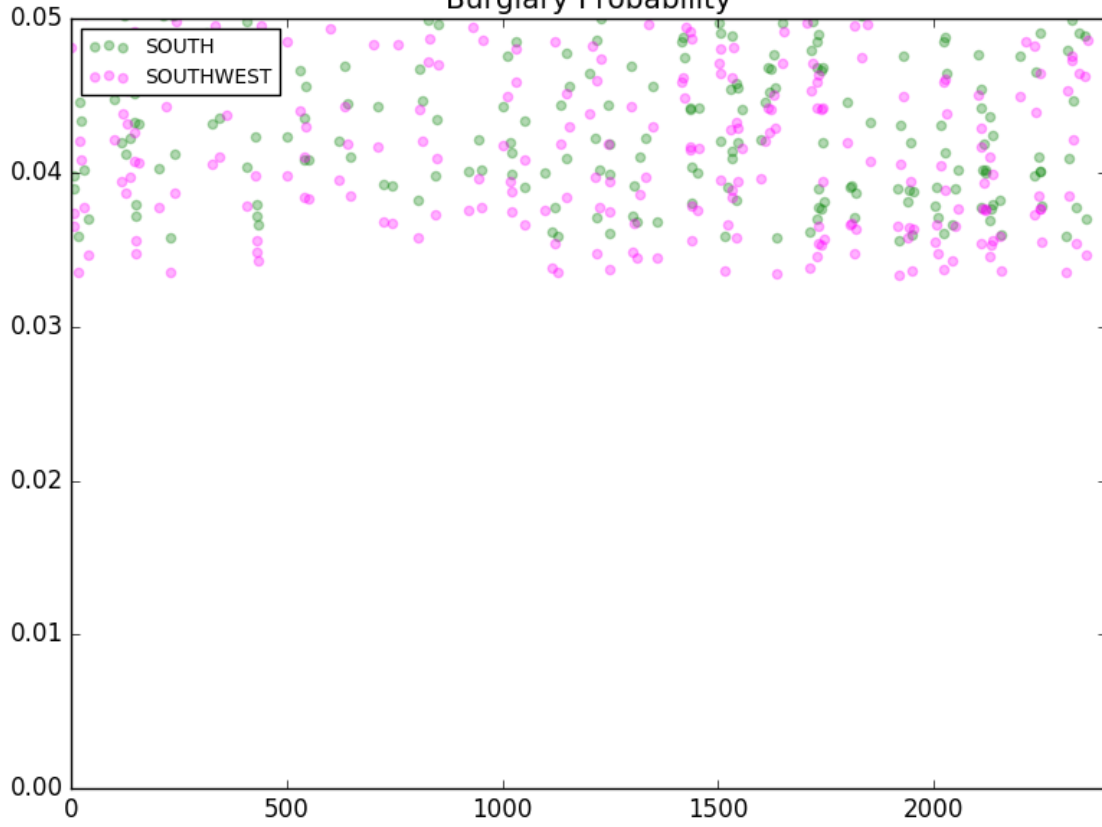


Figure 128
Theft Probability

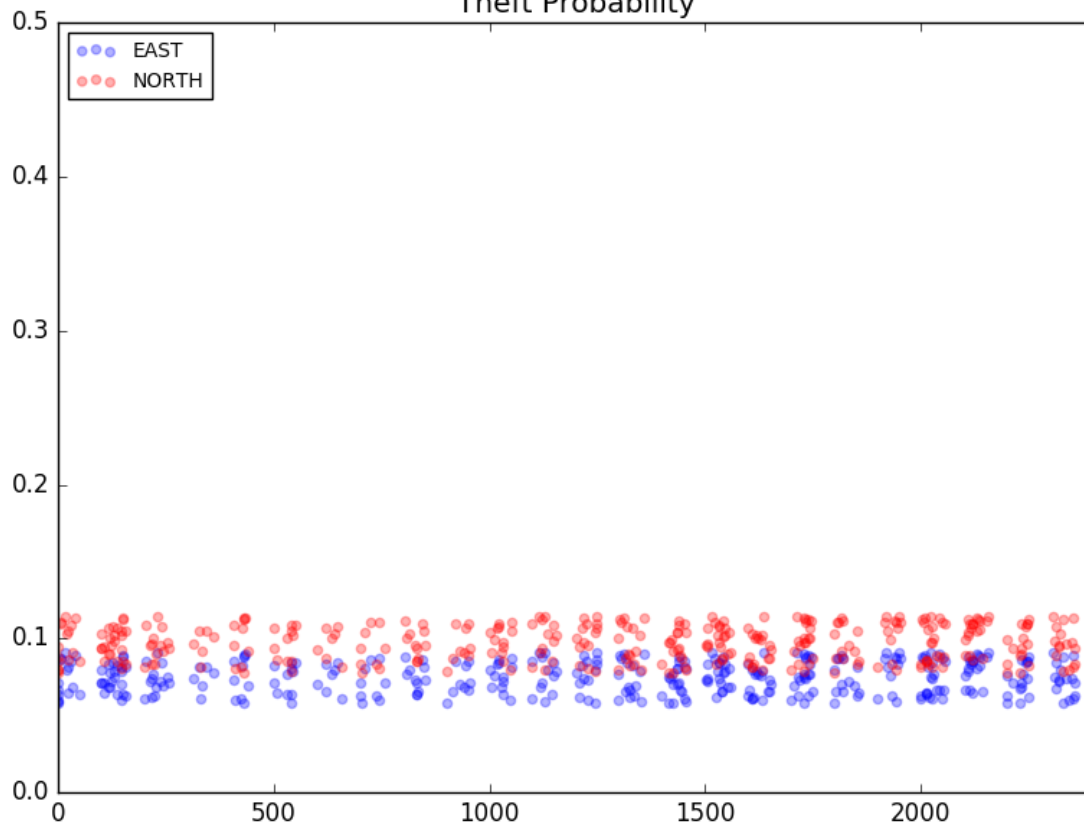


Figure 129
Theft Probability

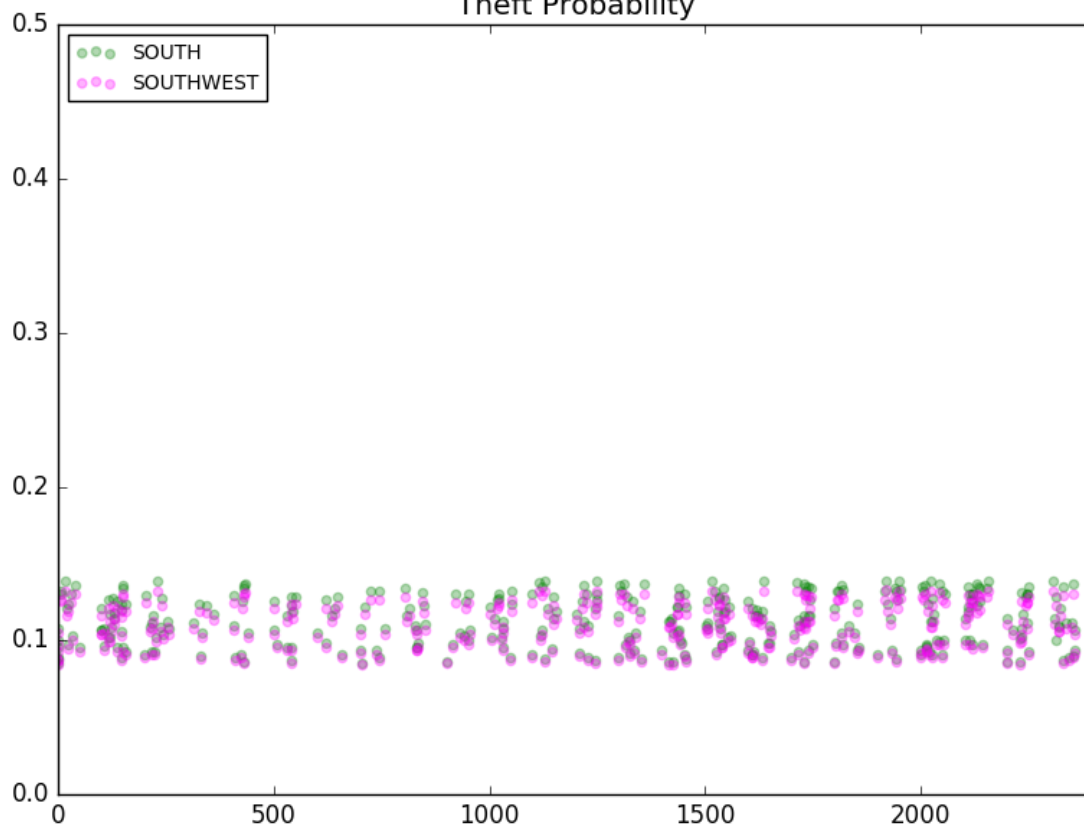


Figure 130
Assault Probability

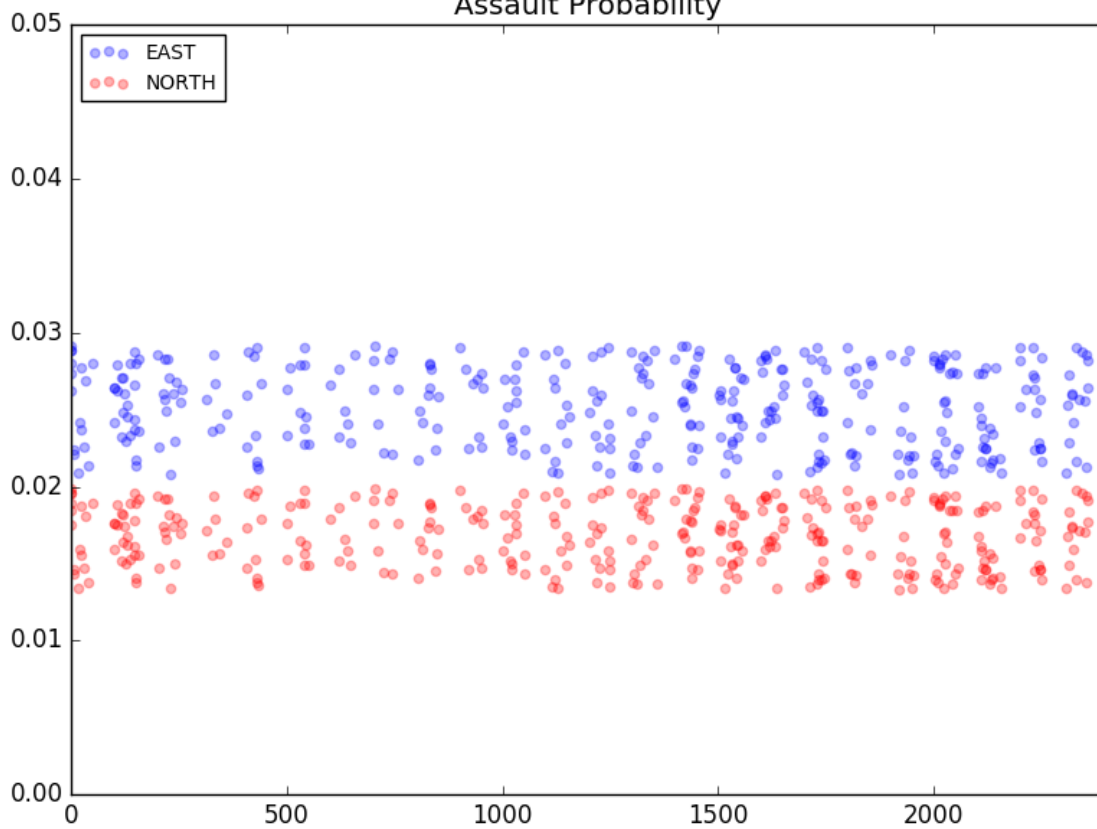


Figure 131
Assault Probability

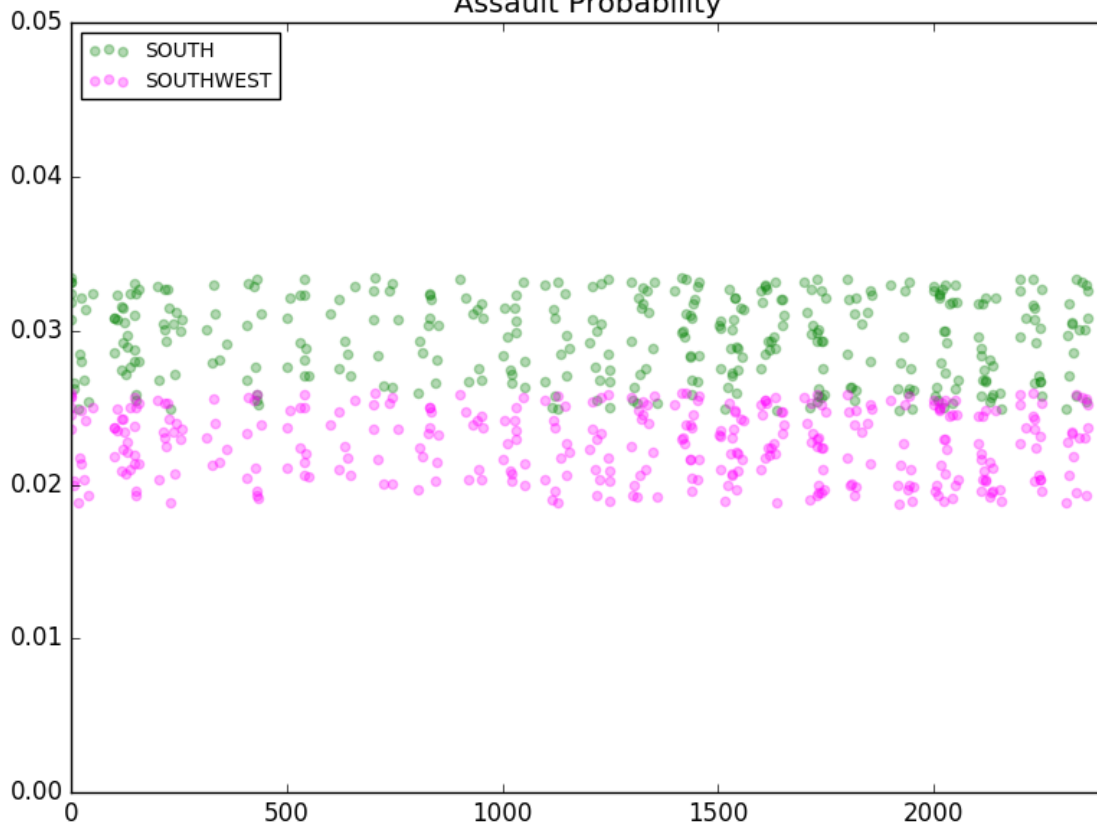


Figure 132
Misdemeanor Probability

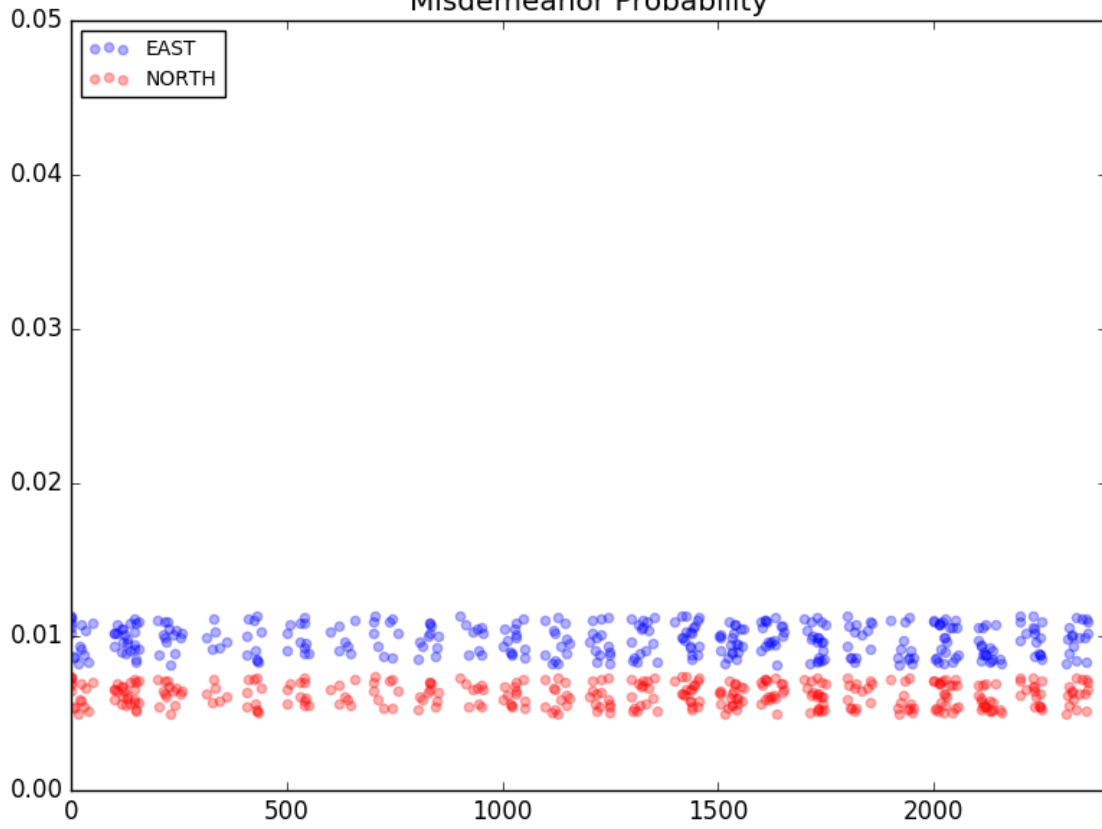


Figure 133
Misdemeanor Probability

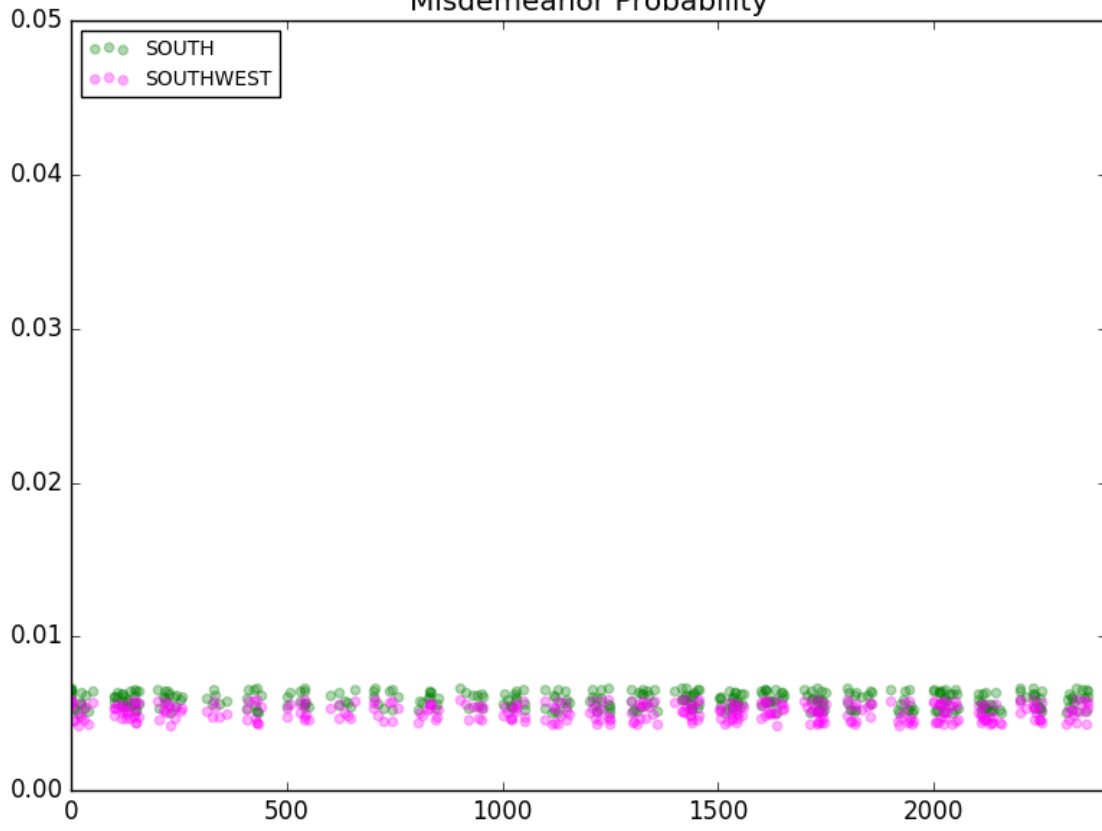


Figure 134
Sex Offense Probability

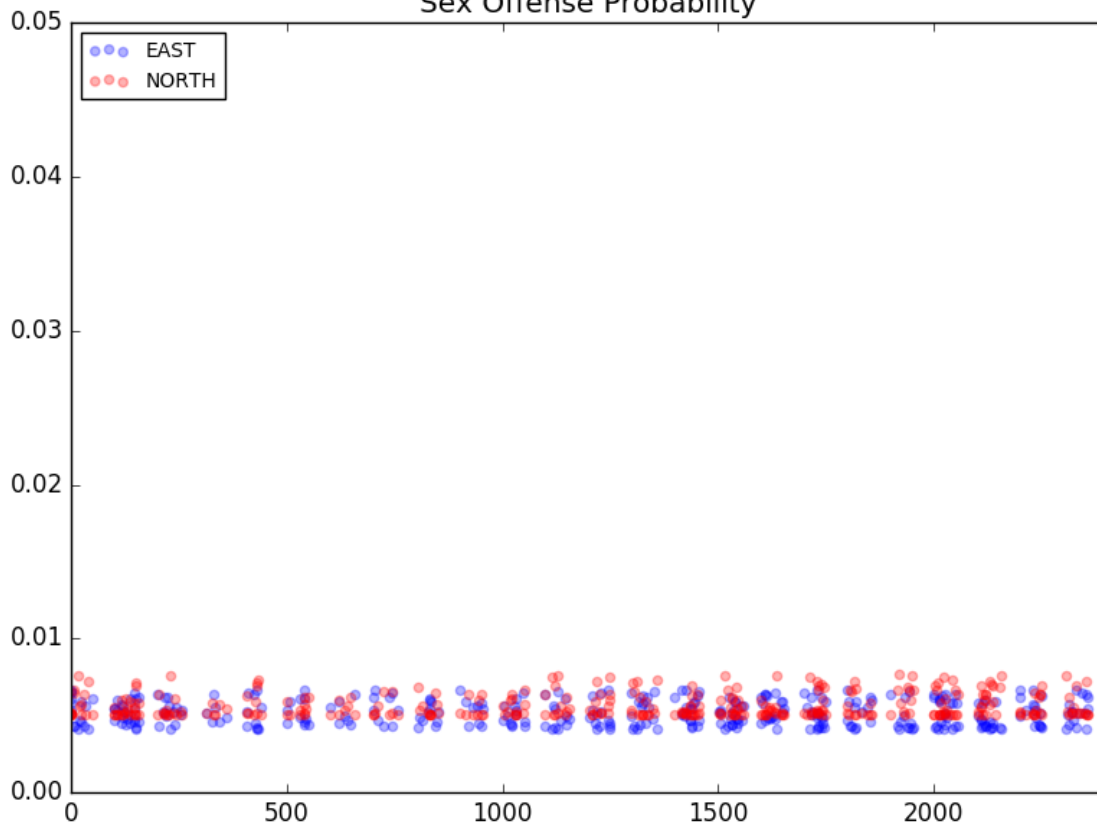


Figure 135
Sex Offense Probability

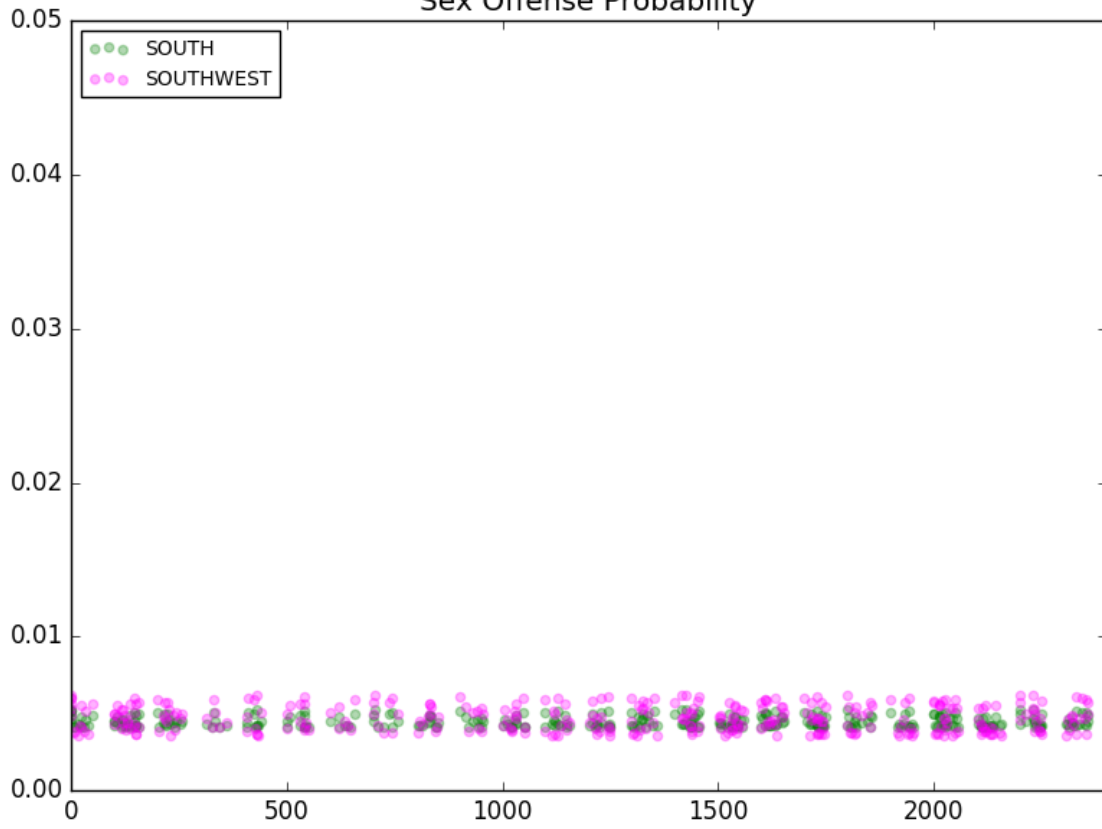


Figure 136

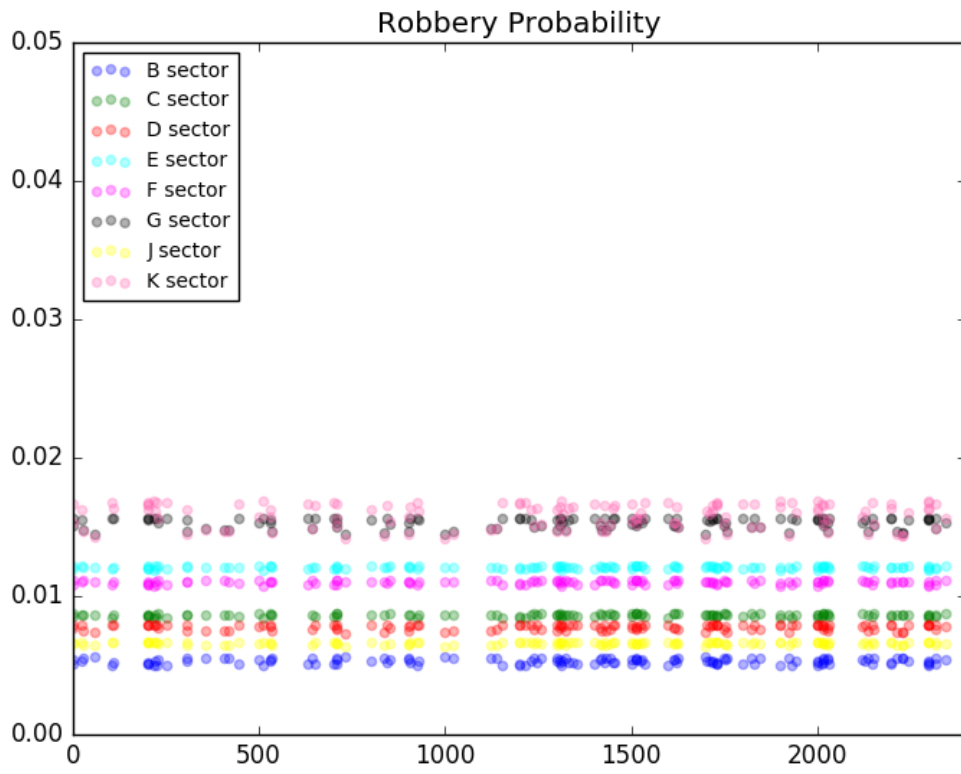


Figure 137

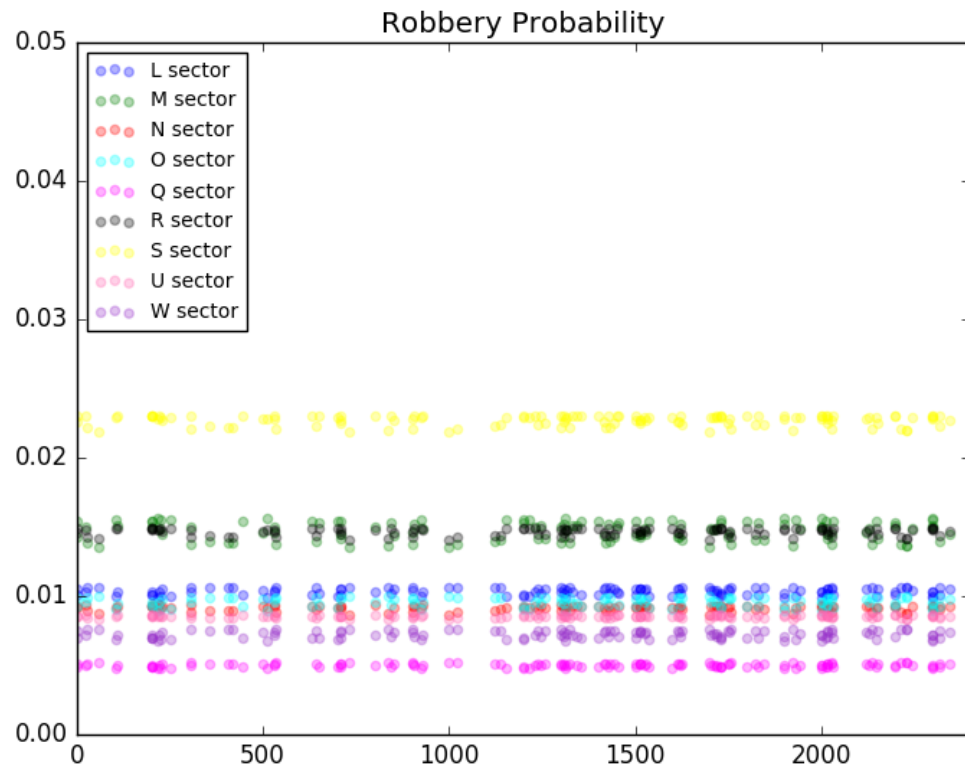


Figure 138

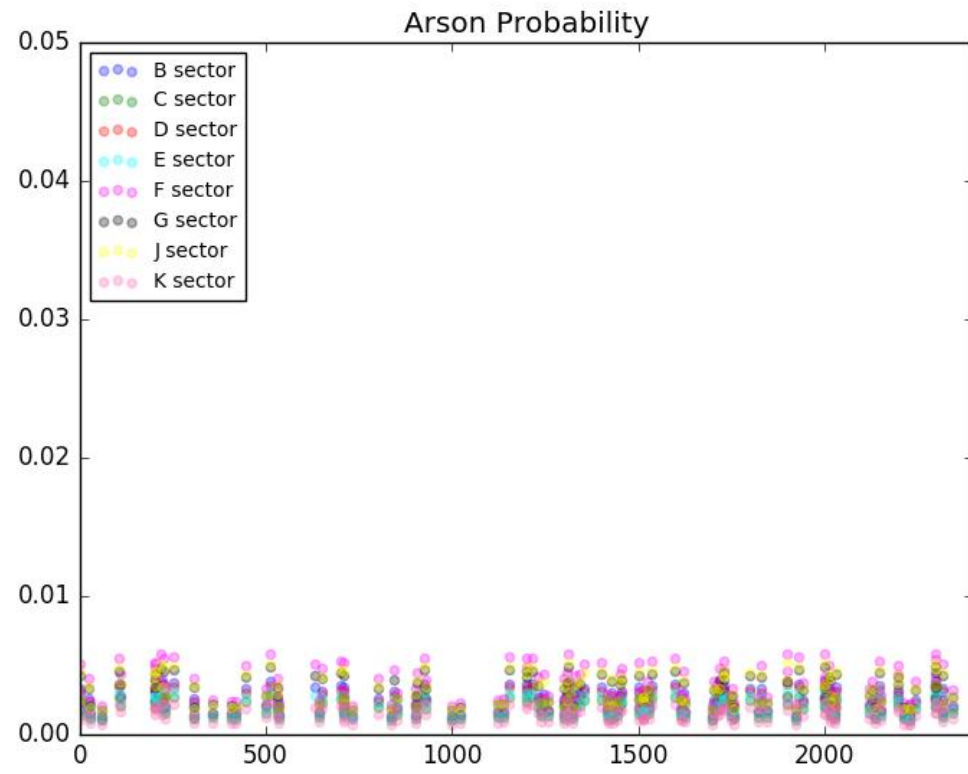


Figure 139

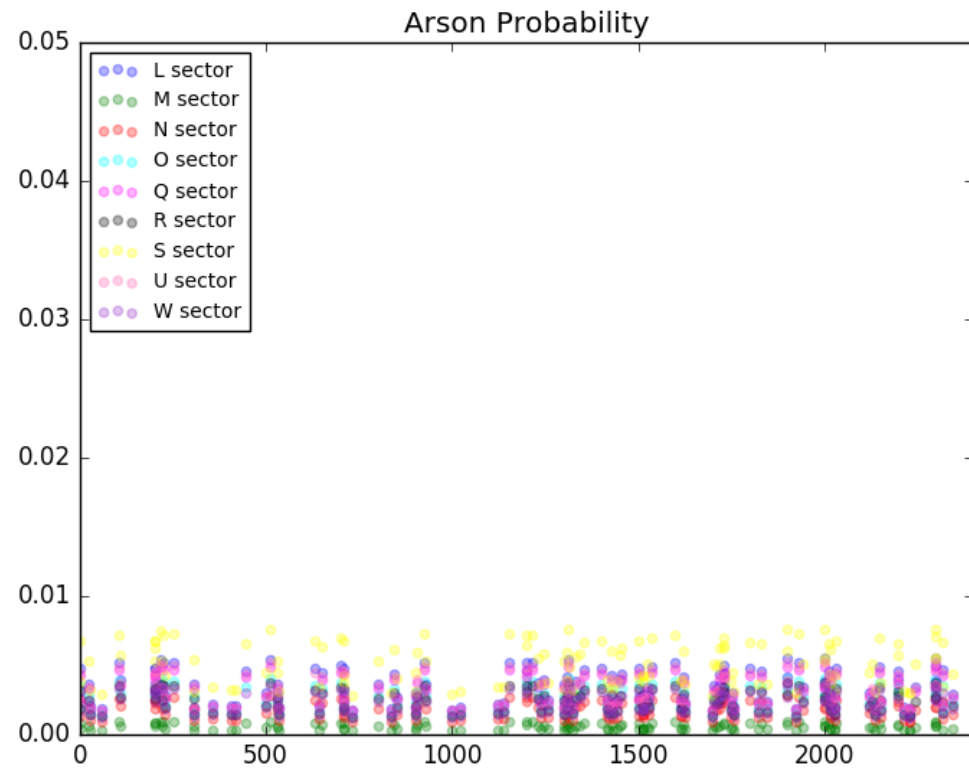


Figure 140

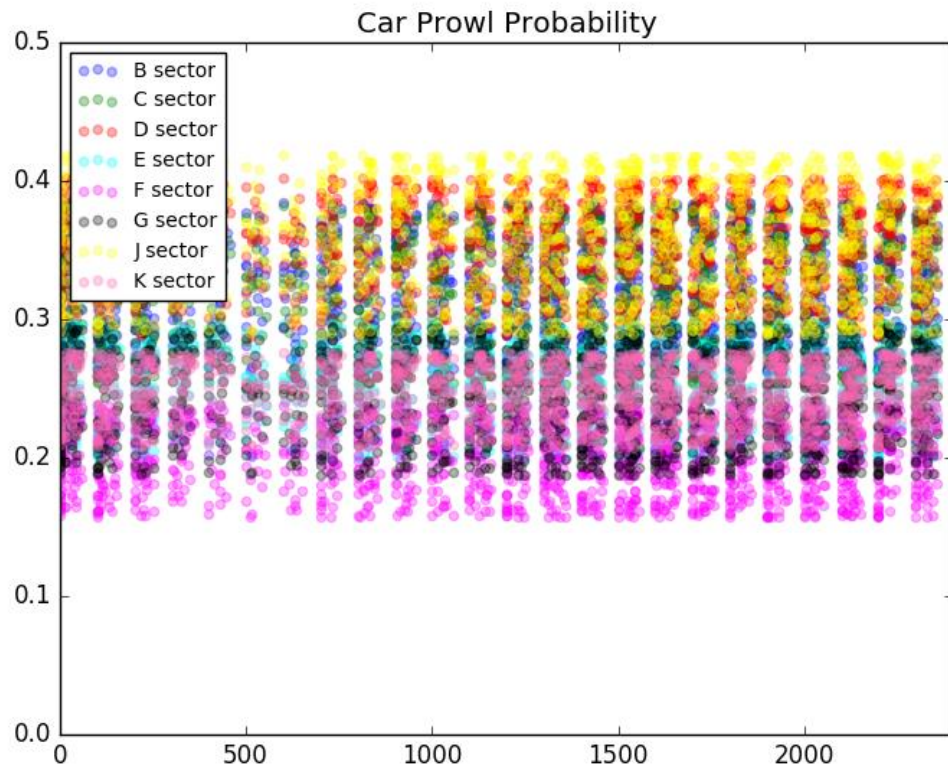


Figure 141

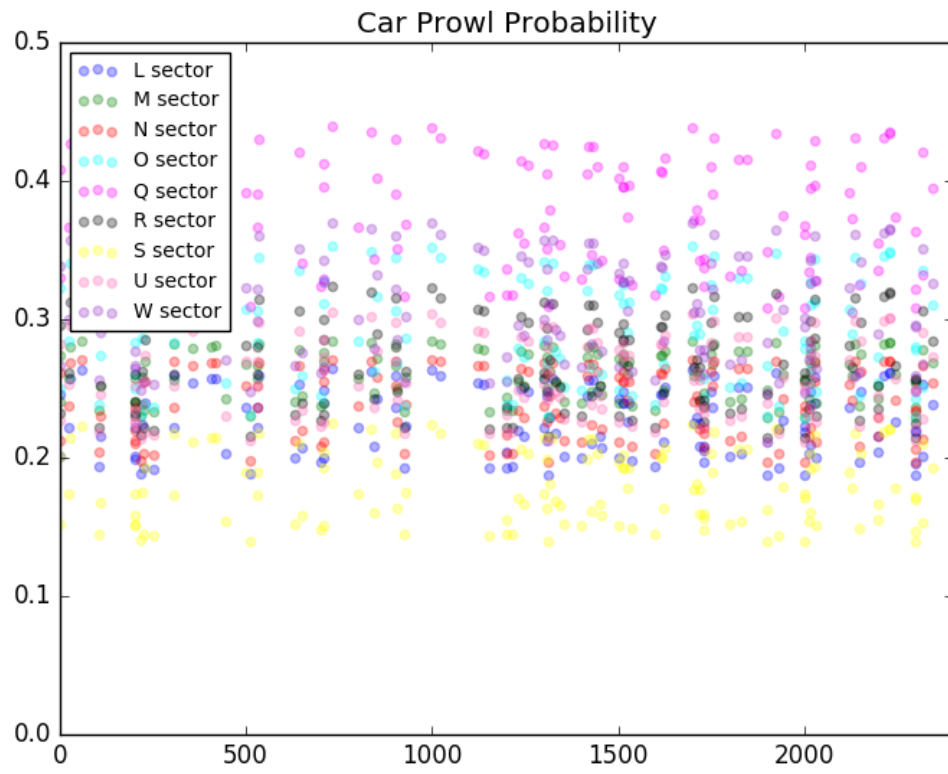


Figure 142

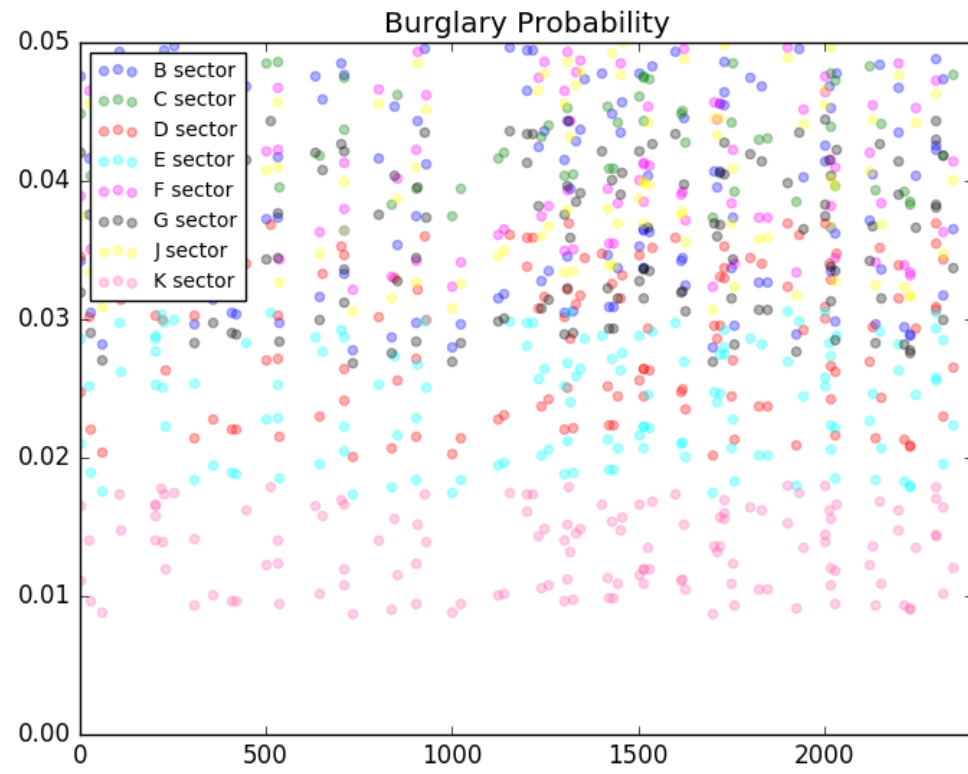


Figure 143

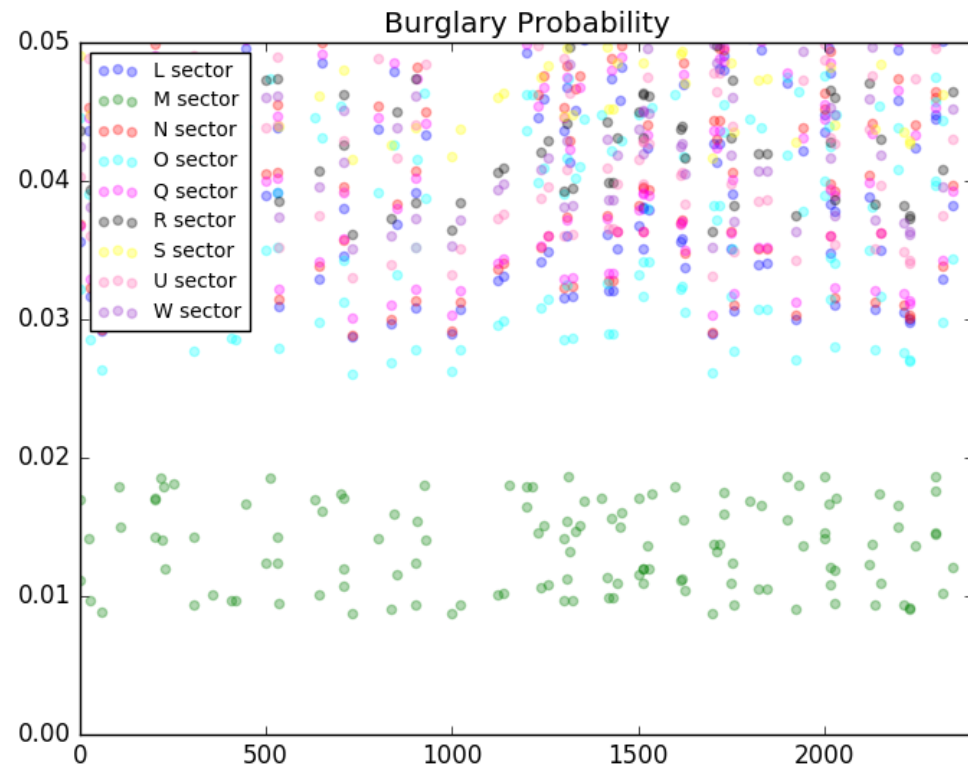


Figure 144

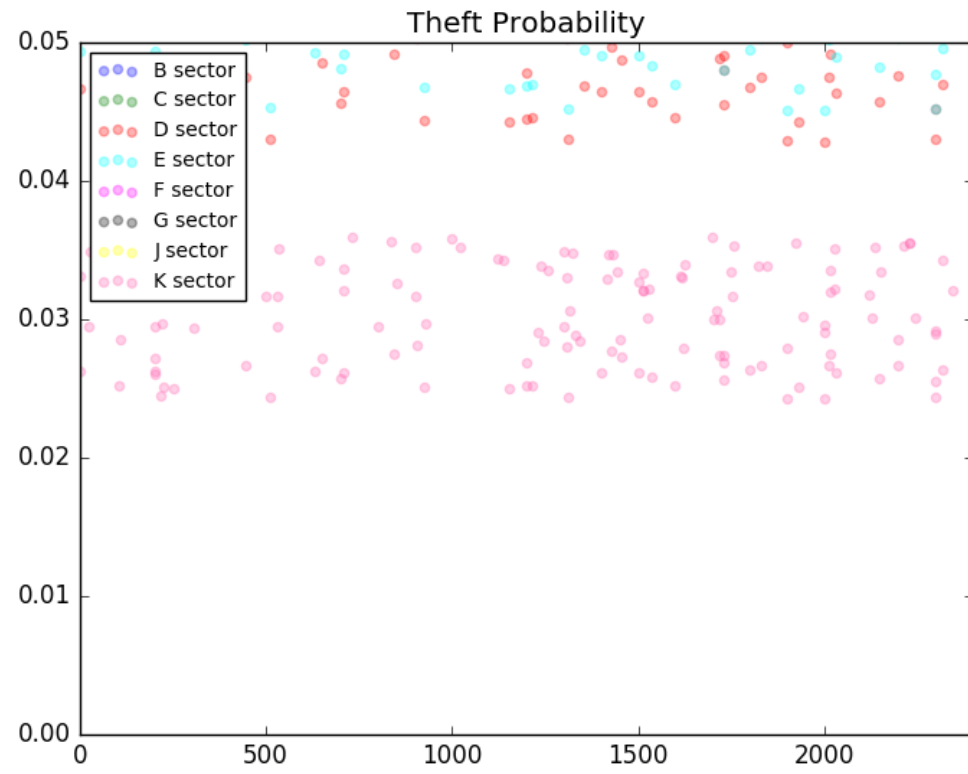


Figure 145

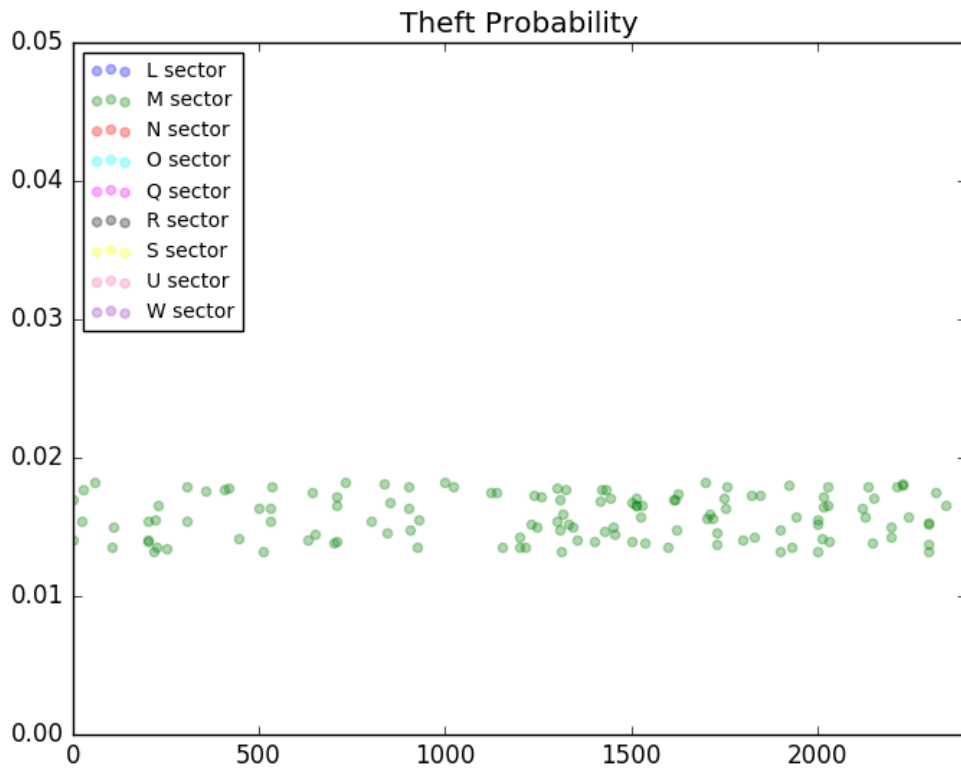


Figure 146

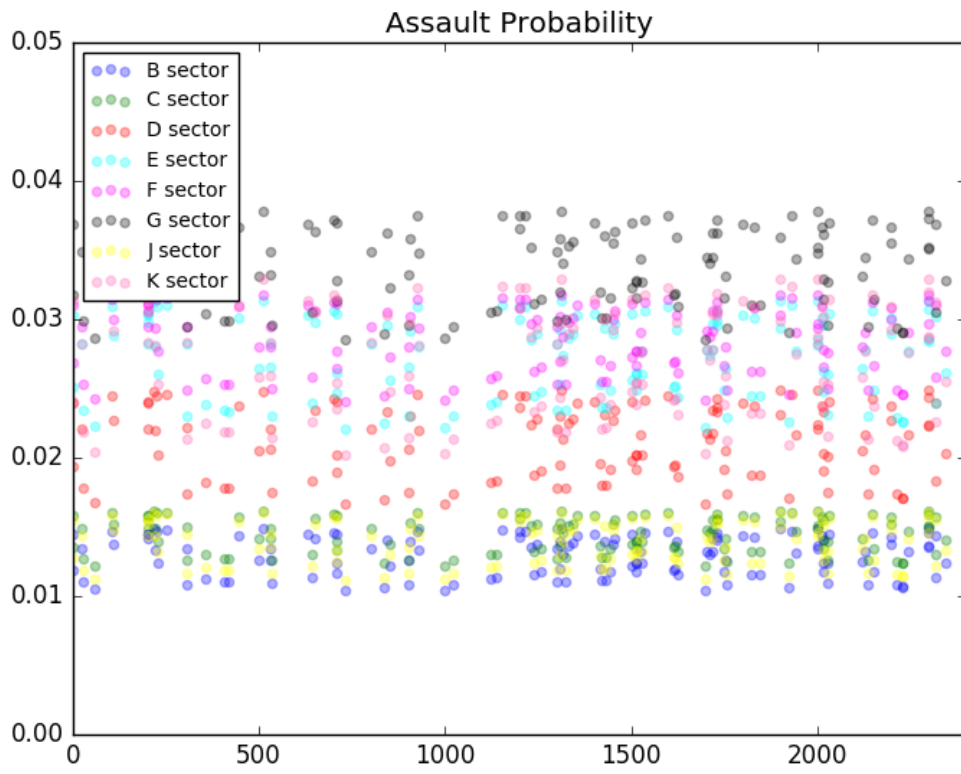


Figure 147

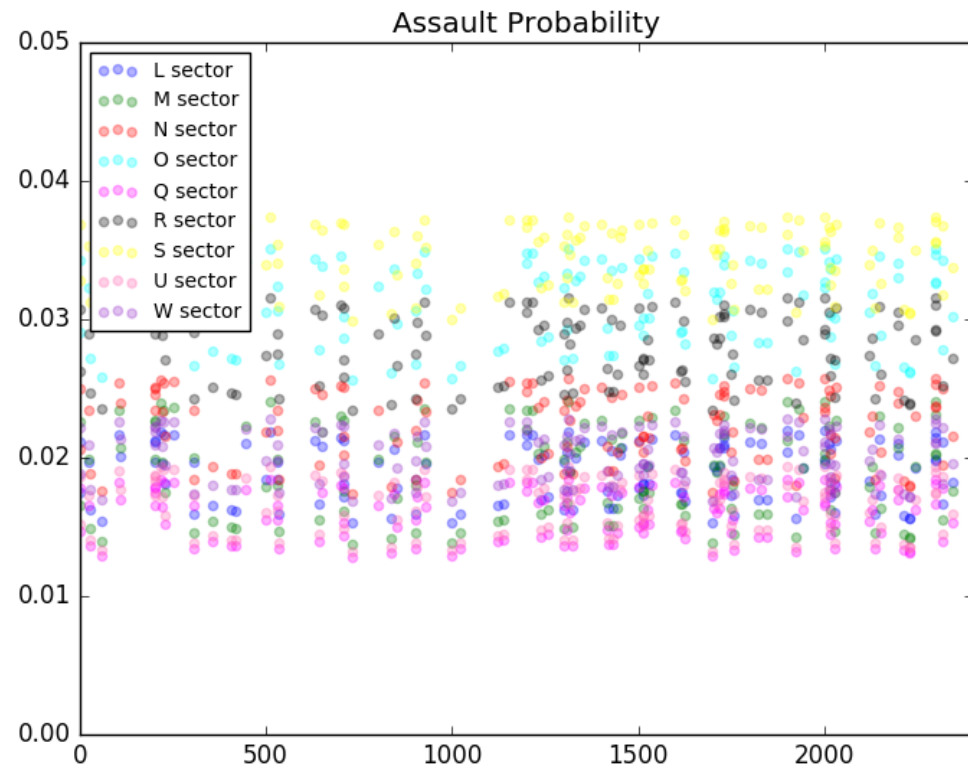


Figure 148

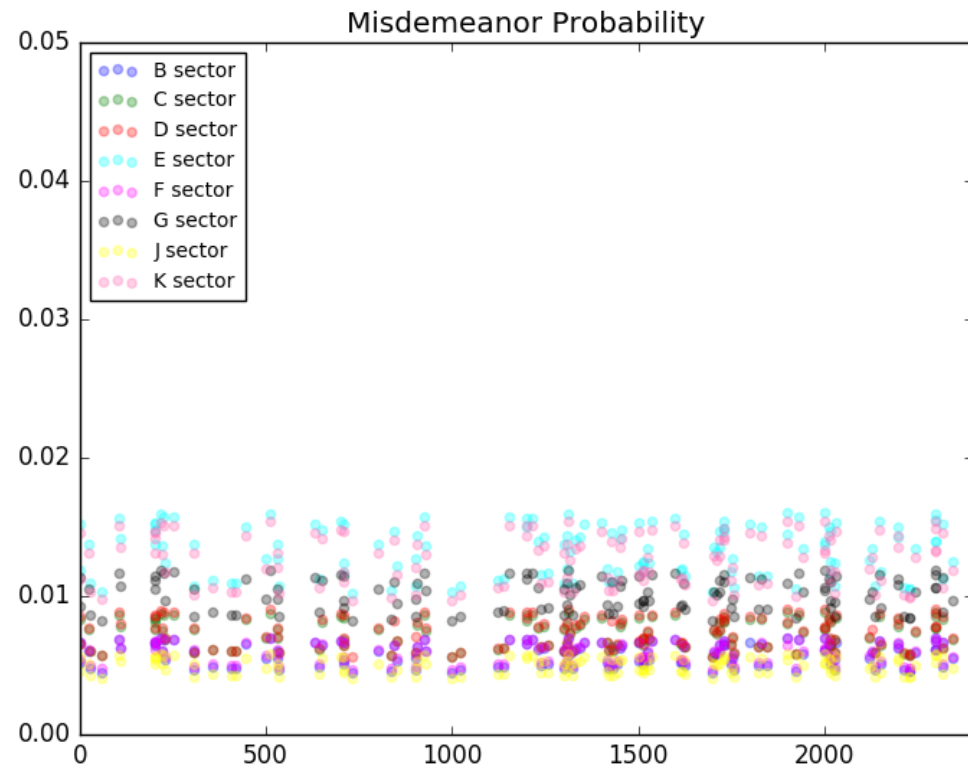


Figure 149

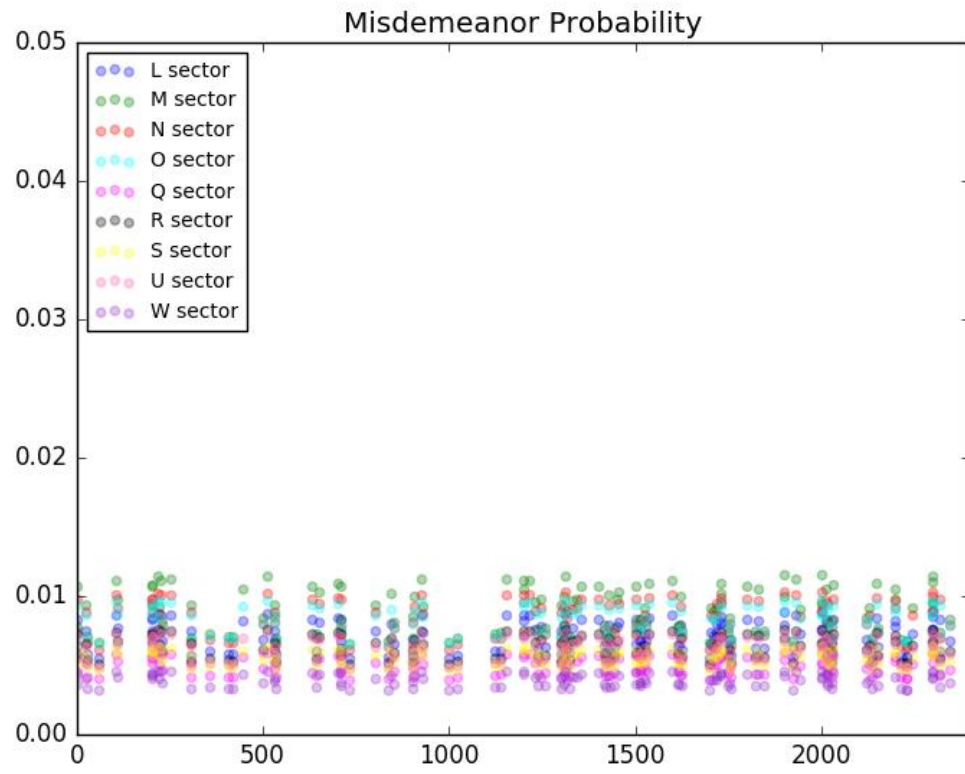


Figure 150

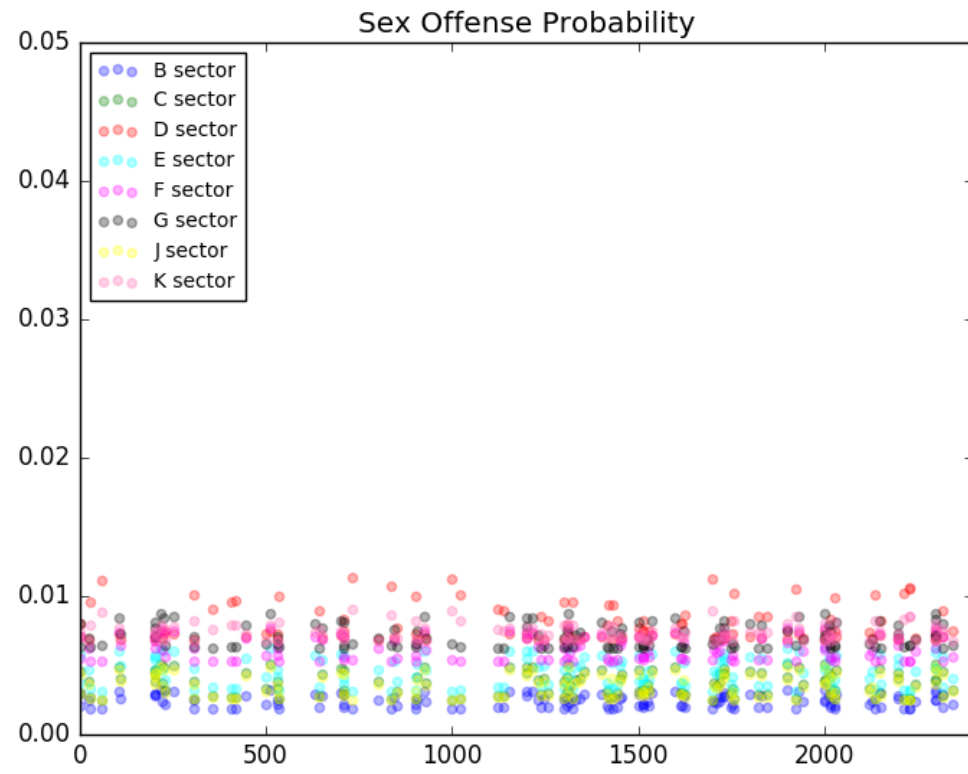


Figure 151

