NOISE MODELING USING INTERNET OF THINGS

by

ANUPAMA DEODHAR

A THESIS

Presented to the Department of Computer & Information Science
and the Robert D. Clark Honors College
in partial fulfillment of the requirements for the degree of
Bachelor of Science

June 2017

An Abstract of the Thesis of

Anupama Deodhar for the degree of Bachelor of Science
in the Department of Computer & Information Science to be taken June 2017

Title: Noise Modeling Using Internet of Things

Approved: ______________________________________

Boyana Norris

Data collected through Internet of Things (IoT) technology has begun to revolutionize the utilization of buildings. Having a wealth of information about noise, lighting, temperature and more collected through accessible, low-cost means allows buildings to be readily customized to increase efficiency and reduce energy costs. The purpose of this project is to prove the feasibility of creating a predictive noise model using low-cost, low-power sensor hardware. Previous research has not adequately addressed how IoT methodologies can be implemented to create noise models, but rather focused on other tools and methods. Furthermore, related research largely takes place outside of the United States suggesting a void in both collected data and research surrounding noise and its applications in America. Noise data was collected in the Knight Library using microphone sensors and the Intel Edison, an IoT device. Results were visualized through a web application, which highlighted relationships between location, time, and noise levels. The resulting models indicated an ability to predict trends over time.

Within a university scope, students can use the resulting models to locate quiet study locations. Outside of a student-oriented scope, having access to noise models in a visual and easily-digestible way provides valuable feedback to inform future building design, improve campus efficiency, and spark discussion about hosting smart buildings on campus.

Acknowledgements

I would like to thank my primary advisor, Dr. Boyana Norris, for her endless encouragement, support, and patience throughout the entire thesis process. Her enthusiasm for both work and life is truly contagious, and I hope to emulate this in the future. I often entered her office overwhelmed with uncertainties about research, this project, and the pile of tasks to complete, but left with a renewed sense of motivation. Despite her busy schedule, her ability to always find time to work with me was, and will always be immensely appreciated.

I would also like to thank Professor Steve Fickas and Professor Samantha Hopkins for their accommodation and valuable feedback throughout the process.

Finally, to my friends and family, I sincerely thank you for all you have given me. I am where I am today because of your unbounded supply of love, encouragement and advice.

Table of Contents

Section 1: Introduction 1
1.1 - Internet of Things (IoT) Technology 1
1.2 - Smart Buildings and Environments 2
1.3 - Noise Policy, Mapping and IoT 2
1.4 - Noise Mapping at University of Oregon 3
Section 2: Research 5
2.1 - Proposed Argument 5
2.2 - Current Research 5
2.2.1 - Urban Noise Mapping 6
2.2.2 - Indoor Building Surveillance 7
Section 3: Methods 9
3.1 - Instrumenting Data Collection 9
3.1.1 - Hardware Implementations 9
3.1.2 - Software Implementation 15
3.2 - Website Creation 18
Section 4: Results/Data Analysis 21
4.1 - Initial Statistical Analysis 21
4.2 - Heat Maps 25
4.2.1 - Days of the Week 29
4.2.2 - Times of Day 29
4.2.3 - Locations 30
4.3 - K-Means Clustering Analysis 32
4.4 - Linear Regression Analysis 35
Section 5: Conclusions 46
5.1 - Summary of Key Findings 46
5.1.4: Local Polynomial Regression: 47
5.2 - Scope & Limitations 47
5.3 - Future Works 48
Appendix A: Hourly Heat Map Tables by Device 51
Appendix B: Local Polynomial Regression Fitting Graphs 58
Per Minute Resolution 58
Finals Week 58

Spring Break 63
Week 1 66
All Weeks 70
Per Hour Resolution 74
Finals Week 74
Spring Break 79
Week 1 83
All Weeks 86
Bibliography 91

List of Accompanying Materials

1. Website: http://ix.cs.uoregon.edu/~adeodhar/index.php

List of Figures

Figure 1: Intel® Edison & Arduino Board 10
Figure 2: Microphone & Light Sensors 10
Figure 3: Custom 3D-printed Shells 12
Figure 4: Map of Edison Placement in Library 14
Figure 5: Example Log File 17
Figure 6: Example of Website Visualization 20
Figure 7: Volume Level Per Hour versus Device ID 31
Figure 8: Duration in Minutes versus Volume Level 33
Figure 9: Duration in Hours versus Volume Level 34
Figure 10: Light versus Volume Level 35
Figure 11: Local Polynomial Regression Fitting Example 44
Figures 12 – 20: Local Polynomial Regressions Per Minute – Finals Week 58
Figures 21 – 27: Local Polynomial Regressions Per Minute – Spring Break 63
Figures 28 –33: Local Polynomial Regressions Per Minute – Week 1 66
Figures 34 – 42: Local Polynomial Regressions Per Minute – All Weeks Combined 70
Figures 43 – 51: Local Polynomial Regressions Hourly – Finals Week 74
Figures 52 – 59: Local Polynomial Regressions Hourly – Spring Break 79
Figures 60 – 66: Local Polynomial Regressions Hourly – Week 1 83
Figures 67 – 75: Local Polynomial Regressions Hourly – All Weeks Combined 86

List of Tables

Table 1: Edison Locations & Nearby Features 22

Table 2: Initial Statistics 23

Table 3: Hourly Heat Map Example for Edison 1 26

Table 4: Hourly Heat Map – All Devices 27

Table 5: Regression Results, One Minute Resolution 37

Table 6: Regression Results, One Hour Resolution 38

Table 7: Local Polynomial Regression Results & Standard Error – Per Minute 41

Table 8: Local Polynomial Regression Results & Standard Error - Hourly 42

Tables 9 – 17: Hourly Heat Maps by Location/Device 51

Section 1: Introduction

1.1 - Internet of Things (IoT) Technology

Have you ever wished your alarm clock could signal your coffee machine to start brewing in the morning? Or that your exercise shirt could monitor your heartrate and exercise intensity and compile that information in the cloud? Or that you could afford to buy either? Although these ideas sound like novelties of the future, they are becoming increasingly plausible through the applications of Internet of Things (IoT) technology.

The concept of IoT refers to a network of sensors or electronic devices that are interconnected via the internet, allowing them to exchange data. The overarching goal of IoT technology is to incorporate existing infrastructure with the use of sensors in order to improve something, whether that is efficiency, quality of life, or any other concepts that could be automated.

Let us consider an example to further clarify the concept. For example, you want to turn your pantry into a “smart pantry”. When you are running low on certain items, your pantry sends that data to your phone via the internet, which compiles a list of items to purchase. Perhaps your phone sends that data to your local grocery store and keeps those items ready for you to pick up. This idea would be simple to implement with the use of IoT technology by lacing the pantry shelves with weight sensors, which would send a data signal when an item has been depleted. Here, the existing infrastructure would be the pantry and shelves, and the only addition would be connecting the sensors and destination (e.g., your phone or grocery store).

1.2 - Smart Buildings and Environments

Data collected through IoT technology has begun to revolutionize the way buildings and environments are used. For example, a building built in Oregon will have different needs for energy efficiency than a building built in New York. Having a wealth of information about noise, lighting, temperature, ventilation, electricity use, and more all collected through low-cost, low-power sensors allows buildings to be easily customized for particular geographic regions, energy standards, and many other metrics of customization (U.S. General Services, "GSA"). The ability to use this data to reduce energy costs and environmental impacts, while increasing efficiency are just a few of the many important applications of IoT technology (U.S. General Services, "GSA").

1.3 - Noise Policy, Mapping and IoT

One increasingly important application of IoT technology is noise mapping. As population growth and urban expansion continues, noise pollution is an extremely relevant issue. In densely populated areas especially, noise exposure is a major player in quality of life for residents. These issues become especially relevant to the elderly, or families with young children who often spend more time sleeping than the average adult. Having a wealth of data about noise patterns or noise pollution can be useful in many applications, including influencing noise policy change.

Environmental noise pollution can have adverse health impacts, including "sleep disturbance, annoyance, noise-induced hearing loss (NIHL), cardiovascular disease, endocrine effects, and increased incidence of diabetes" (Hammer, "EHP"). According to a 1981 study in the United States by the Environmental Protection Agency (EPA), nearly 50 percent of the American population was exposed to harmful levels of traffic noise, but because noise is often deemed less harmful than chemical or radiological pollutants, "Congress has not seriously discussed environmental noise in [more than] 30 years, although noise exposure is a large public concern" (Hammer, "EHP"). It is likely that this number has increased in the last 35 years due to rapid urban growth. It is speculated that this lack of discussion surrounding noise policy is due to a lack of data, or inconsistent data. Creating urban noise maps that provide real, tangible data can encourage dialogue about improving noise policies in the United States.

1.4 - Noise Mapping at University of Oregon

The goal of this project is to combine the ideas of smart buildings and noise mapping to the University of Oregon campus. In this project, noise data will be collected in Knight Library and analyzed and visualized through a browser application.

Although the scale of this project is appropriate within the given time constraints, it is essential to consider the applications of this research on a broader scale. First, within the scope of a university setting, noise mapping would be useful in several ways. Students looking for a quiet study space would be able to access the browser application and identify trends regarding which building areas are quietest at a given time of day or week, and how these trends differ throughout the term. Noise in a university building could also be used to improve building efficiency, such as noise-based temperature control as opposed to motion-controlled actions, for example. Although precise levels of noise are not static and will vary greatly from day-to-day, data that is collected may represent a likely scenario that can be extrapolated to predict future noise levels to an extent. In addition, it is possible that this research may spark discussion about moving towards hosting smart buildings on campus to modernize and improve campus efficiency and help inform future building designs.

Outside of the scope of a university setting, the applications are vast. One critical application is in the field of real estate. Say Sally is interested in buying property in a new city. Everything looks great and she is ready to purchase, but what she does not realize is that a passing train blares its horns at 3am each night. Would she still consider buying the property if she had known beforehand? This problem could be resolved by noise mapping the neighborhood and presenting the data in an application that is rich with visual aids and easily accessible to potential buyers. Another application of noise mapping would be to assist businesses in analyzing customer traffic throughout the day. Business owners could then identify the optimal hours of operation or number of employees to maintain at a given time of day. As mentioned in section 1.3, sparking discussion about noise policy change in the United States would also be a powerful application of this project. The main benefit of analyzing data through smart buildings would be to have a repository of visualized information available to apply to various situations.

Section 2: Research

2.1 - Proposed Argument

The goals of this research are twofold. The first and main goal is to show the feasibility of creating a predictive noise model using low-cost, low-power infrastructure. Using inexpensive hardware and low-voltage equipment introduces many challenges, and determining the viability of noise mapping buildings using these constraints is a large part of the goal. Noise data is collected through use of IoT hardware. The noise sensors being used in this experiment are very basic and do not have the capability of collecting information about noise frequency, amplitude, or recording any contents of speech. They simply sense and output noise levels[1] as a scalar value. The second goal of this project is to exhibit the data in a way that can be easily accessed, understood, and used by a variety of users for diverse purposes. As discussed, the applications of collected data can be vast, and this proof-of-concept study may encourage more cities, business owners and building managers to noise map their domains.

2.2 - Current Research

This thesis project combines two important applications of IoT technology: noise mapping and the development of smart buildings. Because of this, it is necessary to consider existing literature in both areas to frame the purpose and relevance of this project.

[1] Noise data is not collected in decibels, but rather a relative noise scale used by the sensor.

2.2.1 - Urban Noise Mapping

There have been numerous studies that create urban noise maps of cities in recent years. This is partly due to the necessity of ameliorating noise pollution, especially as cities become larger and as adverse health effects from noise are increasing. Studies conducted in China (Min, “GIS-based City central”, P. Tao, “GIS-based city noise”), Egypt (H. Hossan, “Noise mapping”), as well as Europe (P. Tao, “GIS-based city noise”) have used some form of geographic information system (GIS) to collect and visualize noise data. These projects have focused more on creating comprehensive and accurate noise maps using precision tools as opposed to producing similar results through widely accessible means or resource constraints, as this project aims. Additionally, these studies appear to target policy makers or urban planning professionals, as opposed to the general public or unspecialized individuals. One study conducted in Suez, Egypt, has a pointed focus on using the collected data to “develop criteria for the maximum safe noise exposure levels, and to promote noise assessment or noise management policies” (H. Hossan, “Noise mapping”). In fact, this study outlines three main approaches to policy making that may help address noise issues. This discussion of policy is important; however, it does not imply that the collected data will be available to the general public. Exposing the public to collected data will encourage a wide breadth of applications, such as real estate, and is a critical difference in the Suez study and this project.

Another study conducted in Europe (Enda, “Chapter 7”) chooses noise mitigation as its focus, like the Suez study. In this study, GIS is utilized once again as a means for data collection and visualization, but the results appear to be directed towards policy makers and architects. They discuss the positive implications of having noise data available to urban planners and architects in order to improve specific areas inside or outside buildings that are affected by noise, but make no mention of having collected data available to others.

As mentioned, this work differs in some fundamental ways. First, it is important to note that none of these studies took place in the United States, and the difficulty of finding similar noise mapping studies from the United States hints at the extent to which the United States falls behind in researching this topic. Next, the studies discussed above did not utilize IoT technology, but instead used GIS methodologies. One study conducted in Australia (J. Jin, "An Information Framework") did utilize IoT technology, however, focused more on deliberating the effectiveness and accuracy of different types of IoT technology, such as wireless sensor networks and mobile infrastructure instead of creating a predictive model.

Lastly, these studies all focused on outdoor noise monitoring, whereas this work will focus on noise mapping of indoor areas. Through indoor noise mapping, we will first determine the feasibility of creating models using low-cost, low-power infrastructure, and through future work, show the project is scalable to outdoor mapping as well. Because of this, smart buildings and indoor monitoring are equally important to discuss.

2.2.2 - Indoor Building Surveillance

Using sensor technology to create smart buildings is a far newer concept than addressing noise pollution. For this reason, there is less research specifically focused on noise monitoring in buildings. The high-level idea that creating smart environments is an important application of IoT technology is well acknowledged in textbooks and research studies (Bahga, “Internet of Things”), however, specific studies about implementation details are sparse. The most common application of monitoring buildings using IoT technology is for the purposes of energy efficiency. Some studies have proposed IoT networks and control systems (J. Pan, “An Internet of Things Framework”) that show the feasibility of IoT building monitoring, in addition to showing a real improvement in energy efficiency. Another study aimed to build a framework for creating smart environments that emphasizes the customizability of IoT building monitoring (O. Evangelatos, “A Framework).

This work aims to fill the gap in research about indoor noise mapping in particular, in addition to extrapolating existing studies. Specifically, that IoT noise monitoring can be customized based on metrics specific to a purpose or geographic region to effectively utilize the wealth of collected noise data, while keeping in mind that the project can later be scaled to outdoor noise mapping to address noise pollution and policy as a whole.

Section 3: Methods

This research was broken into two main stages. The first stage involved instrumenting the data collection process and collecting data, and the second stage involved interpreting, analyzing, and visualizing the results.

3.1 - Instrumenting Data Collection

3.1.1 - Hardware Implementations

The main hardware that was utilized throughout this project was the Intel® Edison and an Arduino board. The Edison is a small embedded processor that is used primarily for developing wearable technology or other small IoT projects. The Edison was attached to an Arduino board, which provided power supply, digital and analog pins where sensors were connected, and Micro USB connectors to which a computer could be attached. Essentially, the Edison was the brain of the operation, and the Arduino board provided the muscles to complete the task. From this point onwards, any reference to "Edison" will assume an attached Arduino board. The Edison is seen in Figure 1. Other involved hardware includes noise sensors (microphone) and light sensors[2], seen in Figures 2 and 3.

[2] Note that light data was collected and analyzed to determine whether there are correlations between light and noise levels. However, light was not an integral part of the findings of this work, and therefore is only included briefly in the analysis section.

Figure 1: Intel® Edison & Arduino Board

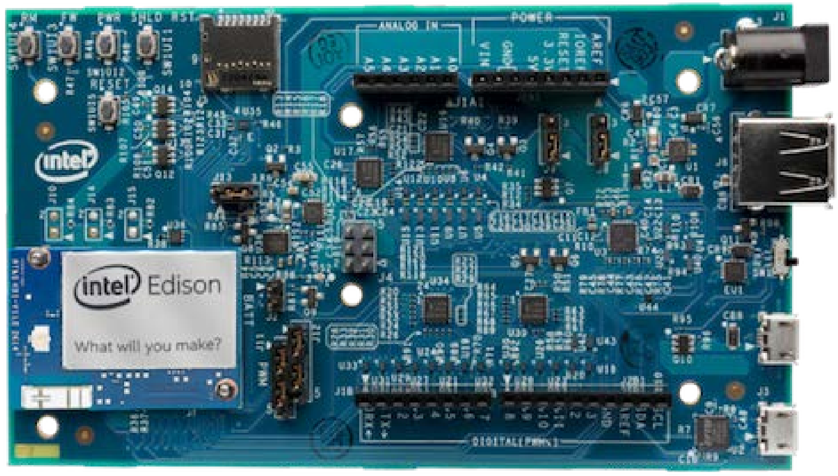

Figure 1: This image shows the Intel® Edison attached to an Arduino board. (Image source: www.developer.android.com).

Figure 2: Microphone & Light Sensors

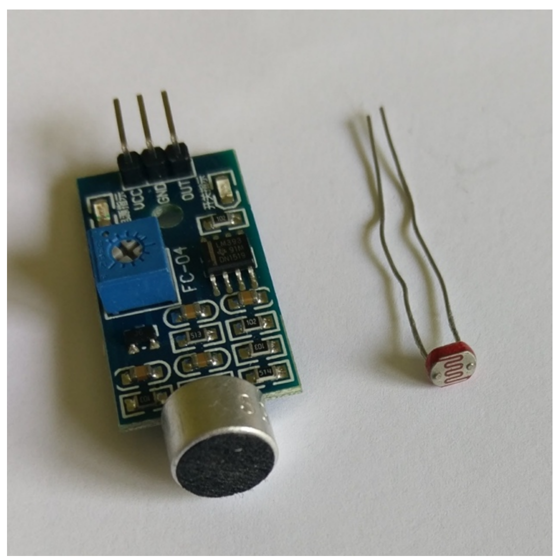

Figure 2: This image shows the microphone sensor (left) and the light sensor (right) used in this work.

A total of ten Edison boards were placed throughout the first floor of the Knight Library. Although seemingly simple, this involved many challenges. The first challenge was disguising the Edisons. Because they were to be placed unmonitored in a public building for three weeks, it was necessary to disguise the boards as to not draw attention. The solution was to utilize a hard-shell case to conceal the Edison and sensors. Due to a lack of inconspicuous ready-to-purchase shells, the best option was to 3D print a custom shell. Once a 3D-printable design was prepared, ten of these custom shells were printed in a black, opaque plastic material. A shell is shown in Figure 3.

Figure 3: Custom 3D-printed Shells

Figure 3.1: A side-view of the shell shows cut-outs for USB connections & power supply.

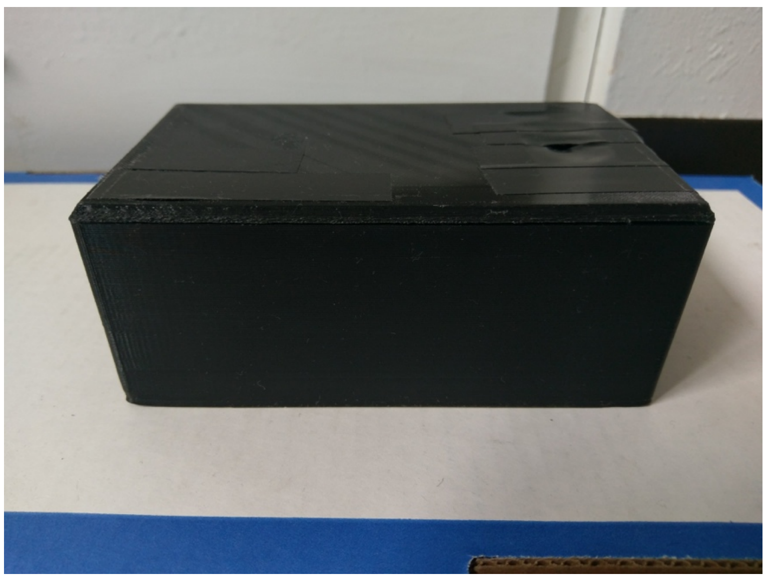

Figure 3.2: Another side-view of the shell. Black electrical tape on the lid covered the imprinted logo, as well as covering large gaps to leave just enough room for the sensors.

As seen in Figure 3, the case included holes for the power cable and microphone sensor, while leaving enough room to conceal all related hardware and wires. The holes are essential for sensors because if they are blocked, they will be significantly less sensitive which will skew the data. The lid and base of the case were held together using black electrical tape. Since the goal of the case was to protect the hardware from both damage and theft, a few other minor precautions were taken; a “Property of University of Oregon” sticker was placed on the front of each case, and strong velcro strips were used to attach the case to the desk surface in the library to discourage removal by curious patrons. Furthermore, a disclaimer was placed near the hardware that explained that noise data was being collected and provided contact information for any questions. Overall, use of the case was successful in the data collection period.

The Edisons were placed throughout the first floor of Knight Library, as seen in Figure 4.

Figure 4: Map of Edison Placement in Library

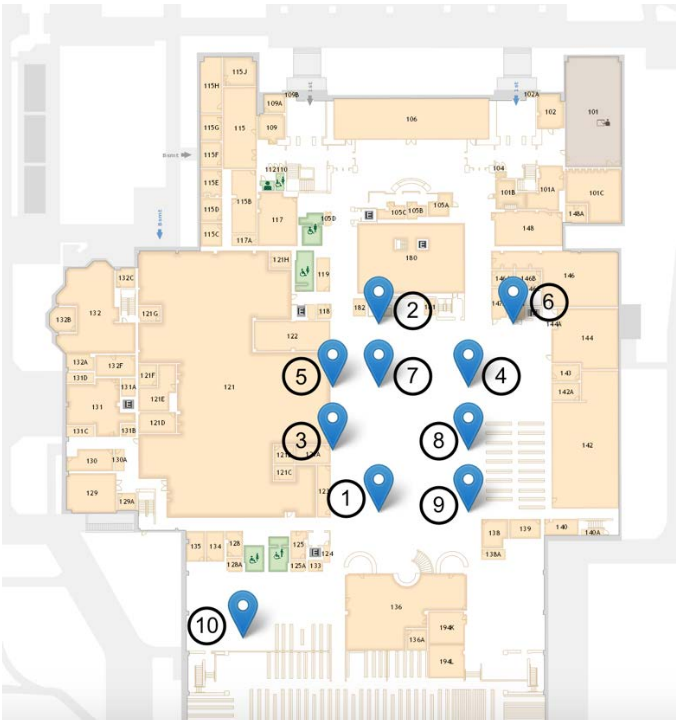

Figure 4: This figure shows a map of the first floor of Knight Library where the Edisons were placed. This image is taken from the browser application developed for this thesis project (see Accompanying Materials in the Table of Contents for the URL).

The placement of Edisons were dependent on a few factors. First, power supply was an important consideration. Because the Edisons are not battery-operated and come with relatively short power cables of approximately 4.5 feet, they could only be placed in locations with accessible outlets. Similarly, it was important to place them where other outlets were available for library patrons to utilize, because more outlets around an area would generally mean a lesser likelihood of somebody unplugging the device. The second factor for placement involved choosing locations that would likely yield a diverse set of data. Edisons were placed near doors, classrooms, stand-up and sit-down computers, printers, and main corridors in order to develop a rich dataset.

3.1.2 - Software Implementation

Software is used to communicate to hardware components how to accomplish a given task. In this case, we created software for collecting, storing, and analyzing data. Before the data collection process began, it was necessary to have the collection and storage software completed.

The Python programming language was the optimal choice for writing the data collection code. Unlike other programming languages such as C++, the software did not need to be re-uploaded to the hardware each time an edit was made.

Due to resource and Knight Library management constraints, the Edisons were not connected to the internet. Because of this, it was not possible to remotely log into the devices to ensure that the data collection process was continuing as planned. Instead, a software daemon was developed to self-manage the process. A daemon is a process that runs in the background of a computer and remains unattended by any specific user throughout its lifetime (IBM, "Glossary of z/OS Terms"). With non-daemon programs, a user explicitly commands the execution of their program, and the program will eventually finish and exit execution on its own. Contrastingly, a daemon will continue to run in the background until an end condition is met, or until the program is explicitly stopped.

The daemon, called MyDaemon[3], was essential due to its properties as discussed, but also because of the location in which it was placed in the Edison. All computing devices have a "boot process", or a process that is executed as soon as the

---

[3] The daemon developed for this project was based upon existing source code (Marechal, "A Simple Unix/Linux Daemon in Python")

device is turned on or restarted. Boot processes can have many functions and components, the most basic of which is loading the operating system into memory, enabling a computer to function on the most basic level. Placing MyDaemon in the Edison's boot process allowed the data collection program to begin execution as soon as the Edison was turned on, without any explicit instruction. This was essential to the success of the data collection process; by automating the process to begin as soon as the device received power, the devices were able to continuously collect data without constant supervision. Furthermore, the devices could self-mange situations in which there was a loss of power, either due to natural causes such as power fluctuations or due to library patrons unplugging the device. In the case of a loss of power, as long as the device was plugged in again, it would resume data collection where it had left off.

MyDaemon had a run() function which facilitated the relationship between the daemon, sensors and the Edison. In essence, this function defined which analog port on the Edison the microphone would be connected to and directed the sensor's output to a log file. Data was collected at a rate of approximately 5 samples per second, and each datum was written to the log file, along with a timestamp, the identification number of the Edison, the light value (lux), and noise level. An example of a log file is seen in Figure 5.

Figure 5: Example Log File

```
log1
2017-03-21 19:18:04.60 | ID: 1 | Light raw: 510, lux: 12 | Vol: 8
2017-03-21 19:18:04.79 | ID: 1 | Light raw: 509, lux: 12 | Vol: 14
2017-03-21 19:18:04.98 | ID: 1 | Light raw: 510, lux: 12 | Vol: 16
2017-03-21 19:18:05.18 | ID: 1 | Light raw: 503, lux: 12 | Vol: 12
2017-03-21 19:18:05.37 | ID: 1 | Light raw: 505, lux: 12 | Vol: 11
2017-03-21 19:18:05.56 | ID: 1 | Light raw: 509, lux: 12 | Vol: 13
2017-03-21 19:18:05.76 | ID: 1 | Light raw: 506, lux: 12 | Vol: 8
2017-03-21 19:18:05.95 | ID: 1 | Light raw: 503, lux: 12 | Vol: 12
2017-03-21 19:18:06.14 | ID: 1 | Light raw: 509, lux: 12 | Vol: 11
2017-03-21 19:18:06.35 | ID: 1 | Light raw: 503, lux: 12 | Vol: 9
2017-03-21 19:18:06.53 | ID: 1 | Light raw: 503, lux: 12 | Vol: 6
2017-03-21 19:18:06.73 | ID: 1 | Light raw: 510, lux: 12 | Vol: 7
2017-03-21 19:18:06.92 | ID: 1 | Light raw: 508, lux: 12 | Vol: 8
2017-03-21 19:18:07.13 | ID: 1 | Light raw: 505, lux: 12 | Vol: 11
2017-03-21 19:18:07.32 | ID: 1 | Light raw: 511, lux: 12 | Vol: 12
2017-03-21 19:18:07.52 | ID: 1 | Light raw: 508, lux: 12 | Vol: 15
2017-03-21 19:18:07.71 | ID: 1 | Light raw: 505, lux: 12 | Vol: 18
2017-03-21 19:18:07.90 | ID: 1 | Light raw: 506, lux: 12 | Vol: 12
2017-03-21 19:18:08.09 | ID: 1 | Light raw: 509, lux: 12 | Vol: 10
2017-03-21 19:18:08.29 | ID: 1 | Light raw: 503, lux: 12 | Vol: 8
2017-03-21 19:18:08.48 | ID: 1 | Light raw: 504, lux: 12 | Vol: 9
2017-03-21 19:18:08.67 | ID: 1 | Light raw: 510, lux: 12 | Vol: 6
2017-03-21 19:18:08.86 | ID: 1 | Light raw: 506, lux: 12 | Vol: 3
2017-03-21 19:18:09.05 | ID: 1 | Light raw: 510, lux: 12 | Vol: 4
2017-03-21 19:18:09.25 | ID: 1 | Light raw: 510, lux: 12 | Vol: 5
2017-03-21 19:18:09.45 | ID: 1 | Light raw: 504, lux: 12 | Vol: 5
2017-03-21 19:18:09.64 | ID: 1 | Light raw: 510, lux: 12 | Vol: 5
2017-03-21 19:18:09.83 | ID: 1 | Light raw: 507, lux: 12 | Vol: 5
2017-03-21 19:18:10.03 | ID: 1 | Light raw: 508, lux: 12 | Vol: 7
```

Figure 5: This figure shows an example of a log file, which stores raw data collected from each device.

Once the hardware and software implementations were complete, data collection began in Knight Library. Data was collected for a total of three weeks, starting on March 17, and ending on April 7, 2017. The duration of three weeks was chosen in part due to resource constraints, such as occupying power outlets in Knight Library for an extended period, but also due to the importance of collecting a substantial amount of data. One day would likely be unconducive to a robust predictive model. One week may have been sufficient, but three weeks provided much more data, improving the likelihood of creating an accurate predictive model. Furthermore, the weeks during which data was collected significantly contributed to the diversity of data; the first week, March 17-24, was the week of final examinations, March 25-April 2 was spring break, and April 3-7 was the first week of spring term[4]. The weeks of final exams, break, and the first week of each term offers perhaps the greatest amount of variability in noise levels, as one may hypothesize that certain weeks associated with the university calendar produce different noise levels.

3.2 - Website Creation

Once the data collection process was complete, there were approximately 94 million[5] data points to be analyzed in a useful way. As described in section 3.1.2, each Edison was maintaining a sampling rate of approximately 5 or 6 samples per second, quickly aggregating to millions of data points. Once the log files had been transferred from each Edison to a computer, the initial process involved reading in the data and storing it in a database. These steps were necessary for the development of a website[6] that visualizes the data.

As mentioned in section 2.1, a main goal of this work is to not only collect data, but also to have it accessible and digestible in such a way that it can be used for diverse purposes. A website is both easily accessible, interactive, and gives leeway for users to select exactly which data they want to view to match their purposes. Likewise, a website can elegantly host interactive visualizations of data which otherwise becomes cumbersome using other software tools.

---

[4] Note that in the analysis section below, weekends are removed from the dates, defining finals week as March 20-24, spring break as March 27-31, and the first week of spring term as April 3-7.

[5] This is an estimate based upon the average sampling rate of each Edison, which was 5 samples per second. This rough estimate assumes that data collection was not interrupted for any period throughout the entire collection period.

[6] Website URL: http://ix.cs.uoregon.edu/~adeodhar/index.php

Although the amount of data that was gathered did not quite qualify as "big data[7]", the size was substantial, and needed to be reduced considerably to be visualized on a webpage. By experimenting with various open-source visualization libraries, it was determined that 200,000 data points or less would return reasonable loading times for a webpage while still maintaining significant detail and integrity of the data. From there, it was necessary to write software that would read in the log file and shrink the data.

To reduce the size of the data, we created software that read in each line of the raw log file and averaged the noise levels down to one reading per minute. Because each line of the log file had an associated timestamp, it was possible to iterate through the raw data, keeping track of which data points were recorded in a certain minute, calculating the average, and writing the average value to a file that maintained the processed data. This reduced the number of data points to approximately 30,000 points over the span of 3 weeks, which was well under the 200,000-point target.

The webpage has two main visualizations. The first page (homepage) visualizes where on the first floor each Edison was placed, and allows users to locate various geographic features close to each device. This is seen earlier in Figure 4. The second visualization was specific to each device, and showed the entirety of the collected data in a graph, as shown in Figure 6.

[7] In computing, big data refers to extremely large amounts of data that may inundate businesses or developers. As a loose generalization, big data is considered that of 1 terabyte or more, but may change depending on the context. (SAS, "What is Big Data")

Figure 6: Example of Website Visualization

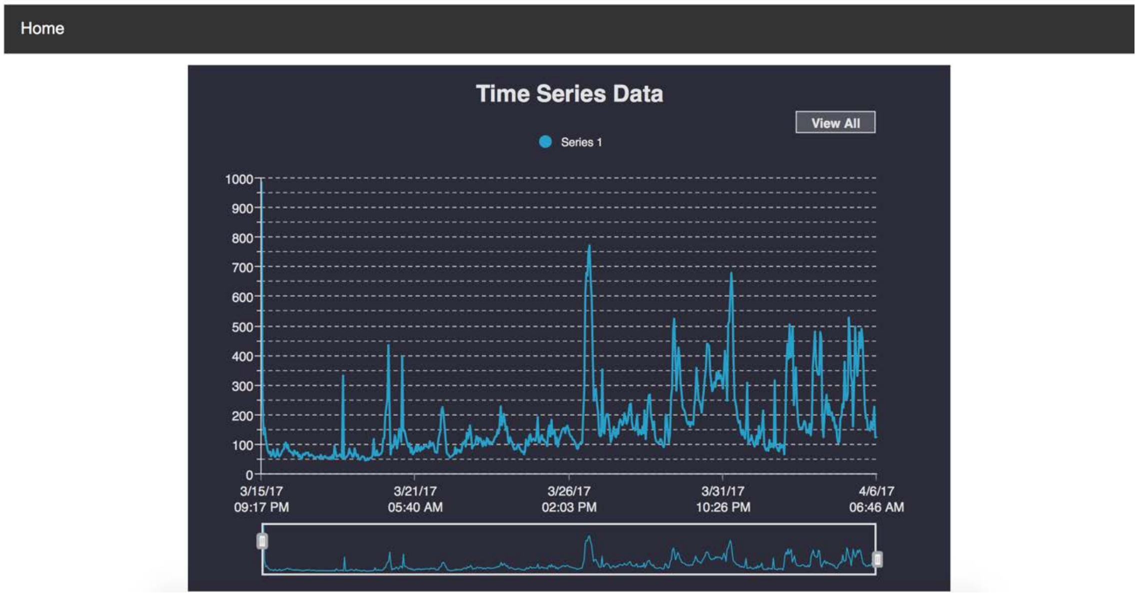

Figure 6: This figure shows an example of the interactive graphical visualization displayed on the webpage. This graph can zoom to an extremely detailed resolution of one point per minute.

Once the raw data files had been reduced in size, the next step was to store them in a database. A database is a collection of organized data that is used for "rapid search and retrieval", and is capable of storing large amounts of data (Merriam-Webster, "Database"). In this case, a MySQL database was utilized to store the filtered data. Data is loaded into MySQL using columns, where each column represents an attribute of that data point. For example, suppose Sally has a database for her refrigerator, which contains apples, oranges, green peppers, and cheddar cheese. Each of these items maintains color, longevity, and date_of_purchase attributes. Sally could query her database with a statement like "select all items from refrigerator where color = orange", which would return oranges and cheddar cheese. For this project, the database consists of the date and time, device ID, raw and adjusted light values, and volume.

Section 4: Results/Data Analysis

With data collected and processed, we next analyzed the data with respect to the goals introduced in Section 2.1. For this analysis, weekend days were filtered out of the dataset. This decision was motivated by the goal of testing the feasibility of making an accurate predictive noise model, and weekend noise patterns may differ significantly from weekday patterns. Comparing weekday and weekend noise patterns may weaken the overall accuracy of the model, and for this reason, only weekday data was utilized.

Analysis of the data was conducted through multiple statistical methods, specifically k-means clustering as well as linear and local polynomial regression fittings. Because readily-interpretable data visualizations are a crux of this work, the results are primarily shown using graphs, tables, and heat map tables. Heat map tables allow one to scan the results and quickly draw conclusions based upon the color of each data cell. In this case, red values will represent loud, high volume levels, whereas green will represent lower volume levels.

4.1 - Initial Statistical Analysis

First, the most basic statistical analysis was applied to each of the 9 datasets[8]: calculating the minimum, maximum, and average values across all weekday values. It was hypothesized that there would be variability amongst finals week, spring break and the first week of spring term, in addition to differences in average sound levels across locations. If the results of this initial test were extremely similar, it may be an indication

[8] The dataset for Edison 9 was removed before analysis began. This device was moved around multiple times by library patrons, and landed in the library lost-and-found. After replacing the device to its original location, it was later found unplugged at the end of the 3-week period. Some data was collected, but not enough to reasonably draw conclusions about that location.

that a predictive noise model based upon this data may be inaccurate, since each Edison was placed in locations where distinct noise patterns would be expected due to varying geographic features. Table 1 shows the identification number of each Edison, in addition to identifying features in the area it was placed (see Figure 4 for a map).

Table 1: Edison Locations & Nearby Features

Edison Identification Number	Identifying Geographic Features
1	• Attached to a table with printers on it
2	• Next to the reference desk
3	• Near the door of Room 121 (meeting space) • Attached to a table with stand-up PC machines
4	• Attached to a table with Mac machines • Near the door of room 142
5	• Near the door of Room 122 (meeting space) • Attached to a table with stand-up PC machines
6	• Next to door of Edmiston Classroom
7	• Next to the reference desk, on opposite side of Edison 2
8	• Next to spiral staircase • Attached to a table with stand-up Mac machines
9	• Attached to a table with Mac machines • Along main walking corridor • [This Edison was unplugged and moved around multiple times by library patrons, so the data collected from this device is not used]
10	• In the back room of the first floor • Attached to a table with PC machines

Table 1: This table provides information about the geographic features of the area in which each Edison device was placed.

The initial analysis, however, did show diversity in noise levels, corroborating the idea that different surrounding environments may yield varying results, and giving primary confirmation that creating a predictive noise model is feasible to some extent using IoT methodology. Table 2 shows the minimum, maximum, and average values for the noise levels.

Table 2: Initial Statistics

Edison ID	Week	Minimum	Maximum	Average
1	Finals	0	955	386.205134
	Break			
	Week 1	0	534	21.655738
	All	0	955	386.19491
2	Finals	0	295	0.19818
	Break	0	411	0.142402
	Week 1	0	286	0.441792
	All	0	411	0.237224
3	Finals	0	954	494.376933
	Break	1	884	115.498716
	Week 1	1	907	186.454778
	All	0	954	276.038614
4	Finals	2	837	171.254237
	Break	166	944	717.491926
	Week 1	216	945	883.926698
	All	2	945	551.504202
5	Finals	0	343	0.298148
	Break	0	427	0.553109
	Week 1	0	390	0.317291
	All	0	427	0.39129
6	Finals	0	419	1.253388
	Break	0	499	1.695891
	Week 1	0	409	5.735485
	All	0	499	2.522008
7	Finals	449	951	878.502694
	Break	172	937	749.936839
	Week 1	168	945	731.373756

	All	168	951	793.757657
8	**Finals**	0	376	1.498517
	Break			
	Week 1			
	All	0	376	1.498517
9	**Finals**	3	945	813.713858
	Break			
	Week 1			
	All	3	954	813.713858
10	**Finals**	0	365	0.351614
	Break	0	429	0.217624
	Week 1	0	349	0.274068
	All	0	429	0.28203

Table 2: Minimum, Maximum, & Average Values organized by Edison ID and Week. Grayed out cells represent weeks where data was not collected by the device for undetermined reasons.

Taking advantage of the heat map coloration of Table 2, it is possible to see that Edisons 2, 5, 6, 8, and 10 were much quieter, with average values being well under 5. By contrast, Edisons 1, 3, 4, and 7 were much louder, with volume levels rising to almost 900. Interestingly enough, there is no strong pattern between weeks; for example, one may assume finals week is much louder than spring break or week 1. However, such patterns were not identified through this data.

It is important to note that the results in Table 2 are calculated using only raw data values; aside from filtering out weekends, no data processing was applied. In the remainder of this section, raw data was first normalized and then processed to an appropriate granularity for the tests to give optimal results. The data was normalized using the minimum and maximum raw data values seen in Table 2. For this reason, the forthcoming analyses will show volume levels on a scale from zero to one. Normalizing the data allows one to compare values from the different datasets on a common scale, eliminating error from comparing values that lie on different scales due to calibration differences among the sensors. The data were normalized prior to averaging and processing.

Visualizing millions of data points would demonstrate neither the most useful nor the most interesting results, and for that reason, the raw data was processed into two different granularities. The first resolution was per minute, meaning every data point collected in a given minute was averaged, and the average value was then written to a file with its associated timestamp. The same process was conducted for an hourly resolution. The k-means and linear regression analyses below will display and discuss results from both granularities to determine if correlations can be drawn between accuracy of predictive models and the resolution of processed data.

4.2 - Heat Maps

As mentioned above, heat maps can aid in quickly discerning patterns among variables. In this case, heat map tables highlight patterns among time of day and volume level across the nine different locations. Table 3 shows an example of hourly volume levels for weekdays collected by Edison 1. The remaining heat map tables for Edisons 2-10 can be found in Appendix A. Each value is an average of finals week, spring break, and the first week of spring term. It is important to note that Table 3 and the tables in Appendix A are comparing values only within their own dataset, and not across all locations. For this reason, Table 4 combines all nine data sets and assigns heat map values based upon the entire data, and not simply data collected at one location. Therefore, Table 4 clearly distinguishes which locations were the loudest and the quietest.

Table 3: Hourly Heat Map Example for Edison 1

	Monday	Tuesday	Wednesday	Thursday	Friday
0:00	0.61	0.84	0.07	0.96	0.34
1:00	0.61	0.87	0.09	0.97	0.28
2:00	0.67	0.79	0.11	0.96	0.22
3:00	0.63	0.71	0.15	0.95	0.16
4:00	0.60	0.58	0.1	0.88	0.15
5:00	0.61	0.41	0.12	0.86	0.18
6:00	0.60	0.42	0.17	0.84	0.15
7:00	0.63	0.36	0.15	0.88	0.16
8:00	0.64	0.37	0.15	0.90	0.17
9:00	0.77	0.28	0.18	0.90	0.37
10:00	0.72	0.30	0.14	0.82	0.48
11:00	0.65	0.19	0.09	0.49	0.72
12:00	0.73	0.17	0.12	0.16	0.84
13:00	0.70	0.06	0.12	0.07	0.73
14:00	0.49	0.04	0.1	0.04	0.73
15:00	0.61	0.06	0.14	0.04	0.65
16:00	0.75	0.03	0.09	0.04	0.49
17:00	0.79	0.03	0.11	0.04	0.51
18:00	0.82	0.02	0.1	0.07	0.26
19:00	0.75	0.01	0.15	0.18	0.17
20:00	0.58	0.04	0.32	0.38	0.17
21:00	0.66	0.06	0.25	0.52	0.14
22:00	0.70	0.06	0.39	0.53	-
23:00	0.78	0.05	0.96	0.39	-

Table 3: This table shows hourly weekday data for Edison #1. Note that this device only collected data for finals week, so the values shown above are unique to finals week. Furthermore, this device did not capture data later in the day on Friday.

Table 4: Hourly Heat Map – All Devices

Edison #	1					2					3					4					5				
Day	**M**	**T**	**W**	**R**	**F**	**M**	**T**	**W**	**R**	**F**	**M**	**T**	**W**	**R**	**F**	**M**	**T**	**W**	**R**	**F**	**M**	**T**	**W**	**R**	**F**
0:00	0.61	0.84	0.07	0.96	0.34	0.00	0.00	0.00	0.00	0.00	0.28	0.35	0.67	0.67	0.27	0.72	0.77	0.71	0.54	0.40	0.00	0.00	0.00	0.00	0.00
1:00	0.61	0.87	0.09	0.97	0.28	0.00	0.00	0.00	0.00	0.00	0.27	0.33	0.57	0.79	0.23	0.76	0.85	0.75	0.61	0.38	0.00	0.00	0.00	0.00	0.00
2:00	0.67	0.79	0.11	0.96	0.22	0.00	0.00	0.00	0.00	0.00	0.19	0.32	0.29	0.68	0.17	0.80	0.76	0.73	0.67	0.42	0.00	0.00	0.00	0.00	0.00
3:00	0.63	0.71	0.15	0.95	0.16	0.00	0.00	0.00	0.00	0.00	0.14	0.45	0.21	0.63	0.31	0.82	0.79	0.68	0.67	0.41	0.00	0.00	0.00	0.00	0.00
4:00	0.60	0.58	0.1	0.88	0.15	0.00	0.00	0.00	0.00	0.00	0.09	0.44	0.17	0.62	0.28	0.78	0.83	0.67	0.69	0.43	0.00	0.00	0.00	0.00	0.00
5:00	0.61	0.41	0.12	0.86	0.18	0.00	0.00	0.00	0.00	0.00	0.12	0.37	0.17	0.48	0.25	0.74	0.89	0.68	0.68	0.40	0.00	0.01	0.00	0.00	0.00
6:00	0.60	0.42	0.17	0.84	0.15	0.00	0.00	0.00	0.00	0.00	0.08	0.47	0.09	0.50	0.32	0.83	0.83	0.69	0.53	0.37	0.00	0.00	0.00	0.00	0.00
7:00	0.63	0.36	0.15	0.88	0.16	0.00	0.00	0.00	0.00	0.00	0.08	0.51	0.07	0.51	0.26	0.88	0.78	0.71	0.50	0.29	0.00	0.00	0.00	0.00	0.00
8:00	0.64	0.37	0.15	0.90	0.17	0.00	0.00	0.00	0.00	0.00	0.07	0.40	0.10	0.45	0.18	0.82	0.78	0.69	0.41	0.23	0.03	0.00	0.00	0.00	0.00
9:00	0.77	0.28	0.18	0.90	0.37	0.00	0.00	0.00	0.00	0.00	0.17	0.37	0.10	0.39	0.19	0.78	0.87	0.69	0.33	0.23	0.00	0.00	0.00	0.00	0.00
10:00	0.72	0.30	0.14	0.82	0.48	0.00	0.00	0.00	0.00	0.00	0.29	0.47	0.16	0.32	0.25	0.71	0.86	0.68	0.27	0.21	0.00	0.00	0.00	0.00	0.00
11:00	0.65	0.19	0.09	0.49	0.72	0.00	0.00	0.00	0.00	0.00	0.29	0.57	0.17	0.22	0.16	0.70	0.87	0.68	0.27	0.19	0.00	0.00	0.00	0.00	0.00
12:00	0.73	0.17	0.12	0.16	0.84	0.00	0.00	0.00	0.00	0.00	0.27	0.49	0.12	0.14	0.14	0.70	0.82	0.69	0.25	0.18	0.00	0.00	0.00	0.00	0.00
13:00	0.70	0.06	0.12	0.07	0.73	0.00	0.00	0.00	0.00	0.00	0.27	0.39	0.12	0.11	0.10	0.67	0.72	0.56	0.27	0.19	0.00	0.00	0.00	0.00	0.00
14:00	0.49	0.04	0.10	0.04	0.73	0.00	0.00	0.00	0.00	0.00	0.30	0.36	0.21	0.14	0.10	0.67	0.64	0.51	0.28	0.22	0.00	0.00	0.00	0.00	0.00
15:00	0.61	0.06	0.14	0.04	0.65	0.00	0.00	0.00	0.00	0.00	0.31	0.39	0.23	0.10	0.12	0.66	0.62	0.51	0.26	0.19	0.00	0.00	0.00	0.00	0.00
16:00	0.75	0.03	0.09	0.04	0.49	0.00	0.00	0.00	0.00	0.00	0.31	0.41	0.25	0.12	0.11	0.64	0.58	0.49	0.26	0.19	0.00	0.00	0.00	0.00	0.00
17:00	0.79	0.03	0.11	0.04	0.51	0.00	0.00	0.00	0.00	0.00	0.32	0.36	0.25	0.12	0.12	0.62	0.54	0.51	0.26	0.18	0.00	0.00	0.00	0.00	0.00
18:00	0.82	0.02	0.10	0.07	0.26	0.00	0.00	0.00	0.00	0.00	0.35	0.36	0.26	0.12	0.12	0.62	0.58	0.60	0.27	0.17	0.00	0.00	0.00	0.00	0.00
19:00	0.75	0.01	0.15	0.18	0.17	0.00	0.00	0.00	0.00	0.00	0.32	0.31	0.26	0.13	0.11	0.62	0.62	0.64	0.30	0.16	0.00	0.00	0.00	0.00	0.00
20:00	0.58	0.04	0.32	0.38	0.17	0.00	0.00	0.00	0.00	0.00	0.32	0.26	0.26	0.13	0.11	0.62	0.63	0.66	0.33	0.17	0.00	0.00	0.00	0.00	0.00
21:00	0.66	0.06	0.25	0.52	0.14	0.00	0.00	0.00	0.00	0.00	0.33	0.35	0.23	0.12	0.12	0.63	0.63	0.66	0.36	0.18	0.00	0.00	0.00	0.00	0.00
22:00	0.70	0.06	0.39	0.53	-	0.00	0.00	0.00	0.00	0.00	0.36	0.45	0.29	0.13	0.14	0.65	0.66	0.64	0.38	0.22	0.00	0.00	0.00	0.00	0.00
23:00	0.78	0.05	0.96	0.39	-	0.00	0.00	0.00	0.00	0.00	0.37	0.58	0.48	0.14	0.16	0.67	0.68	0.55	0.39	0.28	0.00	0.00	0.00	0.00	0.00

Table 4: This table (and the following table which is a continuation) shows all nine datasets in a single heat map. Note that only two places after the decimal are shown, however, the values are not rounded and Microsoft Excel assigns colored values based upon the whole value.

Edison #	6					7					8					10				
	M	T	W	R	F	M	T	W	R	F	M	T	W	R	F	M	T	W	R	F
0:00	0.01	0.01	0.01	0.01	0.00	0.94	0.89	0.91	0.85	0.82	0.02	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1:00	0.02	0.01	0.00	0.01	0.00	0.94	0.90	0.89	0.86	0.75	0.03	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00
2:00	0.02	0.01	0.01	0.03	0.00	0.90	0.82	0.85	0.89	0.79	0.04	0.01	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00
3:00	0.02	0.02	0.01	0.03	0.00	0.80	0.79	0.83	0.86	0.81	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4:00	0.01	0.02	0.01	0.01	0.00	0.61	0.76	0.74	0.80	0.80	0.01	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
5:00	0.01	0.02	0.01	0.01	0.00	0.51	0.80	0.76	0.76	0.80	0.01	0.01	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
6:00	0.01	0.01	0.01	0.01	0.00	0.49	0.82	0.81	0.73	0.78	0.01	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
7:00	0.01	0.01	0.01	0.00	0.00	0.44	0.81	0.82	0.69	0.74	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
8:00	0.00	0.01	0.00	0.00	0.00	0.40	0.76	0.71	0.80	0.74	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
9:00	0.00	0.01	0.00	0.00	0.00	0.50	0.74	0.73	0.78	0.72	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
10:00	0.00	0.01	0.00	0.00	0.00	0.70	0.84	0.79	0.73	0.68	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
11:00	0.00	0.00	0.00	0.00	0.00	0.79	0.88	0.84	0.75	0.69	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
12:00	0.00	0.00	0.00	0.00	0.00	0.76	0.85	0.84	0.80	0.74	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
13:00	0.00	0.00	0.00	0.00	0.00	0.78	0.82	0.78	0.83	0.80	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
14:00	0.00	0.00	0.00	0.00	0.00	0.84	0.85	0.78	0.86	0.81	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
15:00	0.00	0.00	0.00	0.00	0.00	0.87	0.84	0.82	0.86	0.81	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
16:00	0.00	0.00	0.00	0.00	0.00	0.88	0.83	0.80	0.87	0.79	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
17:00	0.00	0.00	0.00	0.00	0.00	0.87	0.83	0.77	0.87	0.79	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
18:00	0.01	0.01	0.01	0.00	0.00	0.84	0.83	0.75	0.88	0.77	0.00	0.02	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00
19:00	0.01	0.01	0.01	0.00	0.00	0.82	0.85	0.76	0.88	0.78	0.03	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00
20:00	0.00	0.00	0.00	0.00	0.00	0.86	0.88	0.79	0.89	0.79	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
21:00	0.00	0.00	0.00	0.00	0.00	0.86	0.88	0.86	0.89	0.79	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
22:00	0.01	0.01	0.01	0.00	0.00	0.87	0.88	0.88	0.88	0.80	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
23:00	0.01	0.01	0.01	0.00	0.00	0.87	0.90	0.84	0.86	0.82	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table 4 (continued): This table (and the previous table) shows all nine datasets in a single heat map. Note that only two places after the decimal are shown, however, the values are not rounded and Microsoft Excel assigns colored values based upon the whole value.

The heat maps above give insight into various patterns across locations, days of the week, and times of the day. These visualizations provide significant information about how noise patterns vary throughout the week, and are useful in creating a noise model that predicts to the granularity of an hour or larger. First, patterns seen in individual locations (Tables 9 – 17 in Appendix A) will be discussed, and later, patterns across all devices will be considered (Table 4).

4.2.1 - Days of the Week

Six devices (Tables 10, 11, 12, 14, 16, and 17, Appendix A) show that Thursdays and Fridays are particularly quiet in comparison to the early and middle parts of the week. Four of these devices (Tables 10, 14, 16, 17, Appendix A) recorded the lowest volume levels for the entirety of Friday, suggesting that Friday is the quietest day of the week. Thursdays show a slightly different pattern, whereby half the day maintains mid-level volume as seen by shades of yellow and light orange, but slowly transition to quieter green shades by the mid-morning to early afternoon. This pattern is observed in all six locations mentioned above.

For other devices, such as Edison # 1 (Table 3), the latter half of Tuesday and the entirety of Wednesday showed quiet volumes. As mentioned in the caption for Table 3 however, Edison # 1 only recorded data for finals week, so the heat map table only visualizes data for that week.

4.2.2 - Times of Day

There are several locations which show high volumes in the early hours of the day, mainly from midnight to approximately 10am. Table 3 (see Tuesday, Thursday), Table 11 (see Wednesday and Thursday), Table 12 (Monday, Tuesday), Table 15 (see Monday, Tuesday, Wednesday), and Table 16 (see Monday) all exhibit similar patterns of high volumes during this timeframe. All of these devices (Edisons #1, 3, 4, 7, & 8) except for Edison #7 (Table 15) lie along main corridors of the first floor. This suggests that there may be some late-night or early morning activities that occur primarily along the main corridors of the first floor, such as vacuuming or other maintenance activities.

Interestingly enough, there are few time ranges that show similar volume patterns across all devices. For example, one may hypothesize that the hours of midnight and 6 am may be quiet in most locations. However, as previously mentioned, this is indeed not the case, and suggests that each location experiences distinct noise patterns throughout the day.

Some devices did experience quieter hours consistently throughout the entire week. For example, Edison #10 (Table 17) shows consistently quiet hours at 11 am, 5 pm, and 8-10 pm across all days of the week. Similarly, Edison #2 shows uniformly quiet hours at 11 am, 2 pm, 5 pm, and 8 pm. Furthermore, Edison #8 (Table 16) shows a few particularly quiet hours across the week, as does Edison #6 (Table 14) and #5 (Table 13).

4.2.3 - Locations

Considering Table 4, it is possible to see how noise at different locations compared to the whole dataset. The heat map clearly identifies quiet and loud locations; in comparison to the entire dataset, Edisons #4 and 7 were the loudest, 1 and 3 were quieter but still quite loud, 6 and 8 were relatively quiet, and 2, 5 and 10 were the quietest. This can also be visualized by Figure 7, which plots the volume of hourly data versus the identification number of each device.

Figure 7: Volume Level Per Hour versus Device ID

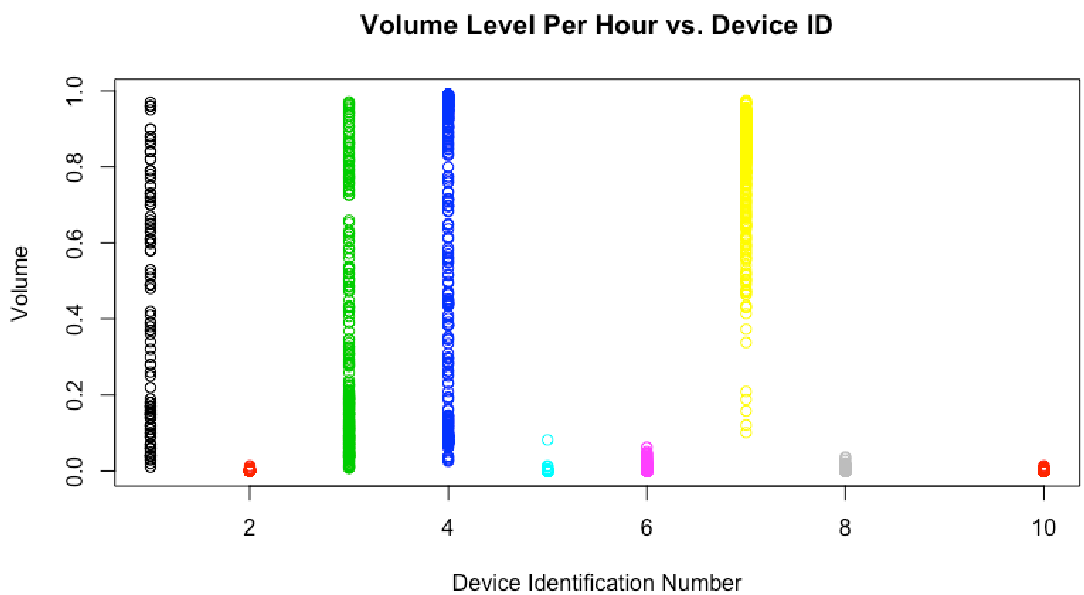

Figure 7: This graph shows differences in volume levels across all devices. The visualization clearly distinguishes between quiet and loud locations.

With Figure 7 and Table 4 in mind, it is possible to cross reference the results with the geographical features surrounding each device (Table 1) in attempt to draw conclusions about what types of features may have correlations with increased volume. For example, Edison #1 was attached to a desk which had a printer on it. Since there are limited printers on the first floor of the library, this area has a unique feature that is bound to attract more traffic. One can hypothesize that more print jobs will be queued during finals week, therefore making that area especially noisy. Edisons #3 and 4 were placed near classrooms or meeting rooms. It is not possible to definitively conclude that areas close to doors are louder, as Edison #6 was also placed near a classroom door but was relatively quiet, however, it may be that some rooms simply have more traffic than others.

Devices on the quieter end of the spectrum tended to be attached to tables with stand-up computers, such as Edisons #5 and 8, or tucked away from the main space of the first floor, such as Edison #10. Locations near the reference desk were inconclusive; both Edison #2 and 7 were placed on either side of the desk, and Edison #2 was very quiet, while Edison #7 was one of the loudest.

4.3 - K-Means Clustering Analysis

The next tool utilized was k-means clustering. K-means clustering divides a series of *n* observations into *k* clusters, based upon each value's proximity to the nearest mean value, or centroid. Clusters identify groups of data that are similar to one another (Piech, "K Means"), and these similarities can then be used to draw conclusions about the dataset.

As mentioned, k-means are utilized to highlight relationships between two variables. However, visualizing time of day versus volume levels would be redundant when considering the heat maps shown above. The heat maps give a sufficient idea of what times of day were the loudest and quietest. Instead, k-means can be used to uncover relationships that are harder to discern upon first glance of the data. One such relationship is volume and duration of noise. For example, when there is a loud noise, does it last a long time, or does it quickly dissipate? Similarly, when a location is quiet, does it remain quiet for extended periods? Figures 8 and 9 below attempt to answer these questions.

Figure 8: Duration in Minutes versus Volume Level

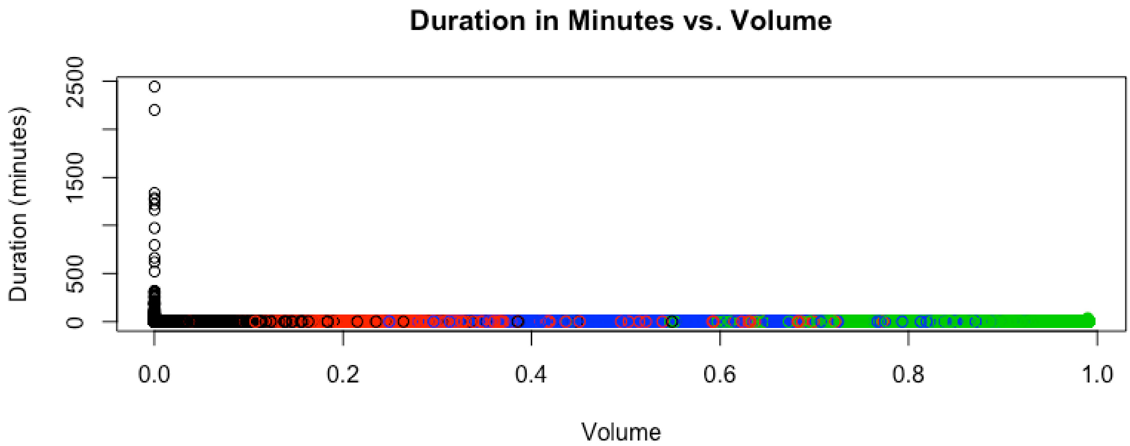

Figure 8: This graph shows the duration of noise in minutes versus the recorded volume. Different colors represent the different clusters, otherwise known as the value of *k* (*k* = 4).

Figure 8 visualizes volume level versus duration in minutes. As seen in this figure, when volume levels are extremely low, or at zero, the duration of this volume level tends to be extensive. As seen by the visualization, quiet periods extended anywhere from zero minutes to almost 2500. This suggests that quiet periods tend to be long-lasting, especially in comparison to loud sounds, which do not show similar correlations. At the loudest volume level of 1, there is the slightest uptick in duration, however it is not significant enough to conclude that loud noises last extended periods of time. Figure 9 shows a similar relationship between low volumes and high time durations in hours. However, Figure 9 does show slightly more correlation between higher volumes and higher duration. This is primarily seen in volumes above 0.9, where they may last approximately 1 or 2 hours.

Figure 9: Duration in Hours versus Volume Level

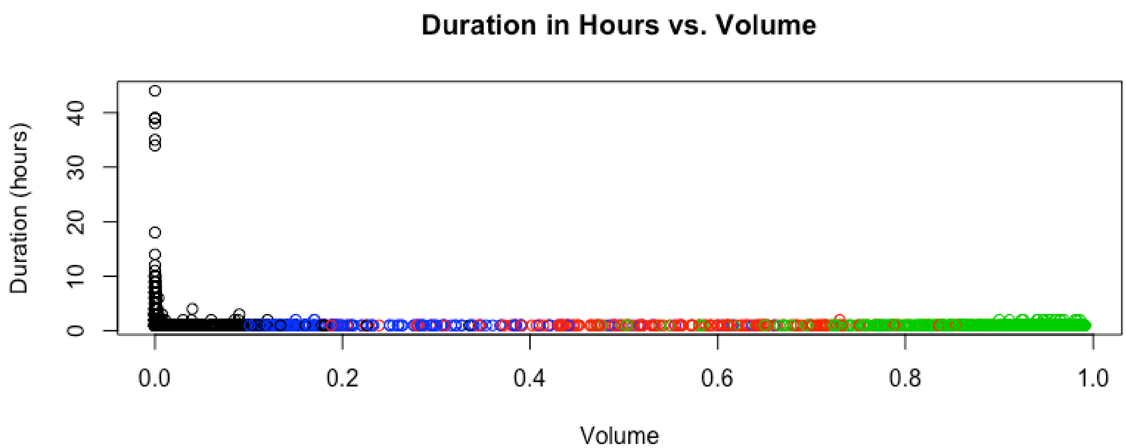

Figure 9: This graph shows the duration of noise in hours versus the recorded volume. Different colors represent the different clusters, otherwise known as the value of *k* (*k* = 4).

As mentioned, k-mean analysis can be used to determine relationships between variables. Because light data was also collected throughout this process, light versus volume data is shown in Figure 10.

Figure 10: Light versus Volume Level

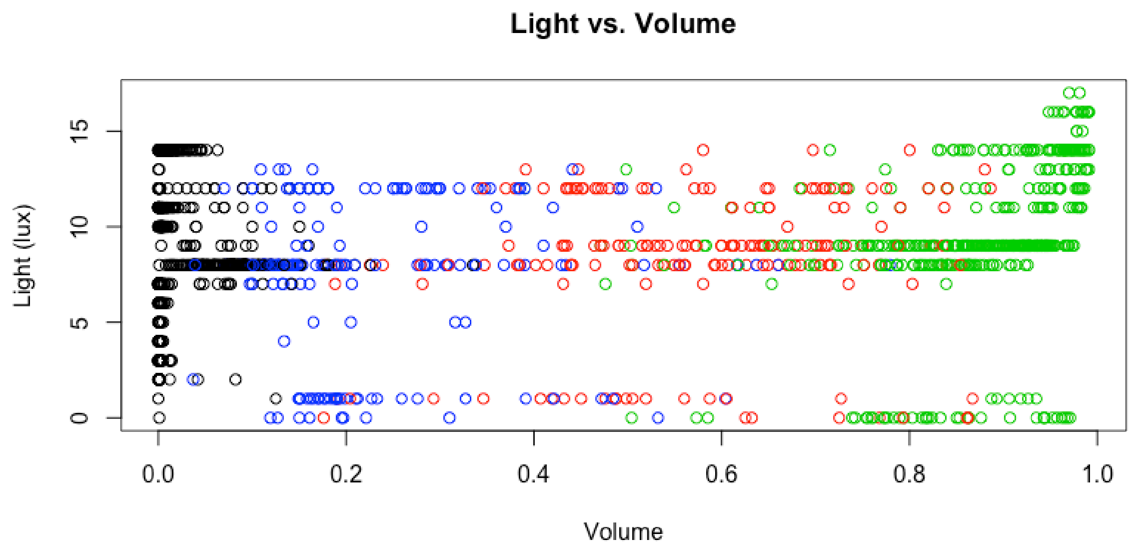

Figure 10: This figure visualizes light values in lux versus volume levels collected across the nine datasets.

Figure 10 does not show any distinct clusters that may be used to conclude a significant relationship between light and volume levels. However, one may be able to hypothesize based upon this visualization that areas with high lux values may also have increased volume levels, or vice versa, as seen by the green cluster in the top right hand corner.

4.4 - Linear Regression Analysis

Next, it is possible to use linear regression analysis to discuss the extent to which these results may be conducive to an accurate predictive noise model. Linear regression models the relationship between an explanatory variable (time) and a dependent variable (volume) by fitting a linear equation to the data values (Yale University, “Linear Regression”).

As noted in the section 3.1.2, each device recorded data at a rate of about 5 samples per second. In order to determine if one resolution yields a more accurate predictive noise model, we smoothed the data into courser resolutions. This regression analysis uses data in hourly and per minute resolutions, and the two resolutions are compared to see if different results were produced. Once again, strictly normalized data values are used.

This regression analysis uses 60 percent of each dataset as training data to create a best-fit linear model. The remaining 40 percent of test data was then analyzed in comparison to the linear model to determine whether the training data could predict the test data. To determine the accuracy of this model, this analysis will focus on the R-squared value. R-squared indicates how close the data is to the regression line, which in turn indicates how accurate the model is (Dass, “Regression”). R-squared values lie between 0 and 100 percent and demonstrate what percent of the variance seen in the dependent variable can be explained by the explanatory variable, which in this case translates to what percent of the variance seen in volume can be explained by time. Ideally, R-squared values should be high, indicating that changes in the dependent variable can be explained to a large extent by the explanatory variable. There is no formal cutoff for what defines a favorable R-squared value, although many suggest considering values above 50 percent.

Tables 5 and 6 below show the p-values and R-squared results for linear regressions conducted in both per minute and hourly resolutions.

Table 5: Regression Results, One Minute Resolution

Edison #	Week	P-Value	Multiple R-Squared (%)
1	Finals	2.00E-16	3.062%
	Break	-	-
	Week 1	-	-
	All	1.07E-06	0.337%
2	Finals	4.58E-06	0.292%
	Break	0.28700	0.016%
	Week 1	0.00016	0.302%
	All	0.01333	0.032%
3	Finals	0.62300	0.003%
	Break	2.00E-16	3.205%
	Week 1	4.86E-08	0.638%
	All	1.31E-06	0.123%
4	Finals	2.00E-16	13.170%
	Break	2.20E-16	10.240%
	Week 1	2.00E-16	4.342%
	All	2.20E-16	3.296%
5	Finals	0.09190	0.039%
	Break	0.00039	0.175%
	Week 1	0.88620	0.000%
	All	0.00027	0.063%
6	Finals	3.60E-07	0.359%
	Break	3.38E-11	0.609%
	Week 1	2.00E-16	5.067%
	All	2.20E-16	1.510%
7	Finals	2.00E-16	7.335%
	Break	2.00E-16	4.088%
	Week 1	2.00E-16	3.196%
	All	2.20E-16	3.134%
8	Finals	9.22E-10	0.520%
	Break	-	-
	Week 1	-	-
	All	9.22E-10	0.520%
9	Finals	-	-
	Break	-	-
	Week 1	-	-
	All	-	-
10	Finals	0.54980	0.005%
	Break	0.16720	0.027%
	Week 1	0.70480	0.003%
	All	2.22E-05	0.007%

Table 5: This table shows linear regression results for a one-minute resolution. The "All" category indicates that linear regression was conducted on all 3 weeks together (finals week, spring break, and week 1 of spring term). Grayed out cells indicate that no data was collected for those weeks for undetermined reasons. Green cells indicate the larger R-squared values.

Table 6: Regression Results, One Hour Resolution

Edison #	Week	P-Value	Multiple R-Squared (%)
1	Finals	0.01220	5.335%
	Break	-	-
	Week 1	-	-
	All	0.01220	5.335%
2	Finals	0.02475	4.235%
	Break	0.62700	0.202%
	Week 1	0.07756	4.042%
	All	0.14040	0.691%
3	Finals	0.33860	0.783%
	Break	0.00051	9.856%
	Week 1	0.02839	6.245%
	All	0.02987	1.498%
4	Finals	2.07E-05	14.410%
	Break	0.00035	10.410%
	Week 1	0.02342	6.663%
	All	0.00084	3.505%
5	Finals	0.3218	0.839%
	Break	0.3788	0.663%
	Week 1	0.4484	0.523%
	All	0.4049	0.199%
6	Finals	0.02423	4.265%
	Break	0.04021	3.548%
	Week 1	1.75E-05	21.420%
	All	2.19E-06	6.881%
7	Finals	0.005744	6.338%
	Break	0.05674	3.068%
	Week 1	0.55770	0.454%
	All	0.01698	1.801%
8	Finals	2.67E-02	4.127%
	Break	-	-
	Week 1	-	-
	All	2.67E-02	4.127%
9	Finals	-	-
	Break	-	-
	Week 1	-	-
	All	-	-
10	Finals	0.11710	2.086%
	Break	0.58420	0.257%
	Week 1	0.54670	0.480%
	All	0.12720	0.739%

Table 6: This table shows linear regression results for a one hour resolution. The “All” category indicates that linear regression was conducted on all 3 weeks together (finals, spring break, and week 1). Grayed out cells indicate no data was collected for those weeks for undetermined reasons. Green cells indicate the larger R-squared values.

Considering a per-minute resolution as seen in Table 5, the regression results show very low R-squared values, the highest of which is 13 percent, with most results falling significantly below five percent. If we adopt a 50 percent benchmark for R-squared significance, the calculated values are much lower. This suggests that the variance seen in volume can be explained by time only to a very small extent using this linear regression model. Similar results are seen in Table 6 which utilizes an hourly resolution. Table 6 shows slightly higher R-squared values than Table 5, which may be explained by the resolution. An hourly resolution smooths data to a greater extent than a per-minute resolution, perhaps leading to a better fitting best-fit line. However, the results from both resolutions suggest that the best-fit lines in this model do not accurately capture the changes in volume throughout a 24-hour period, indicating that a predictive noise model based upon linear regression would be quite inaccurate.

Although these results suggest that coarser resolutions may yield better predictive noise models, one must consider the hazard of overly-smoothed data, as this may lead to a model that is not useful, and that smooths out too many features. This is a tradeoff that must be considered when deciding upon resolutions, and it may be beneficial to compare multiple resolutions as demonstrated herein to decide upon the appropriate granularity.

One notable issue with using this linear regression model is that it accounts for the entire dataset while creating a best-fit line. Since the data fluctuates greatly throughout each 24-hour period, the best fit line cannot accurately account for all the changes in noise levels, therefore producing a line that is the best-fit for only a small portion of the dataset, while the rest of the data has a significant margin of error.

Because of this, it is possible to apply other regression methods to create a far more accurate noise model which can account for fluctuations, and counter the issues with using the entire dataset at once.

4.5: Local Polynomial Regression

As discussed in Section 4.4, a linear regression model is not ideal for this dataset because it cannot sufficiently account for fluctuations in noise levels, and the best-fit line is created using the entire dataset, which produces large error margins.

Instead, utilizing a local polynomial regression fitting provides significantly more accuracy for two main reasons. First, this regression fitting uses local data instead of the global dataset. The model first breaks the whole dataset into smaller subsets, and then calculates the best fit within the smaller regions. In other words, for a given point *p*, the fit is created by using points near *p*, which are weighted by their distance from *p* (The R Manual, “Local Polynomial Regression”). The second reason this model provides more accuracy is because it uses a polynomial fit rather than a linear one. This fit provides far more accommodation for varying data patterns and fluctuations, in turn, increasing the accuracy of the predictive model and decreasing error margins.

Like the linear regression fitting shown in Section 4.4, this local polynomial regression also utilizes 60 percent of the dataset for training data, and the remaining 40 percent for test data. Likewise, this analysis was also conducted on both hourly and per-minute resolutions.

Tables 7 and 8 show the Residual Standard Error for each of the calculations in per-minute and hourly resolutions.

Table 7: Local Polynomial Regression Results & Standard Error – Per Minute

Edison #	Week	Residual Standard Error
1	Finals	0.2944
	Break	
	Week 1	
	All	0.2944
2	Finals	0.0025
	Break	0.0082
	Week 1	0.0051
	All	0.0047
3	Finals	0.3016
	Break	0.1223
	Week 1	0.1838
	All	0.2870
4	Finals	0.1650
	Break	0.2209
	Week 1	0.1176
	All	0.3689
5	Finals	0.0035
	Break	0.0142
	Week 1	0.0059
	All	0.0105
6	Finals	0.0103
	Break	0.0211
	Week 1	0.0189
	All	0.0173
7	Finals	0.0564
	Break	0.1544
	Week 1	0.1381
	All	0.1515
8	Finals	0.0160
	Break	
	Week 1	
	All	0.0160
9	Finals	
	Break	
	Week 1	
	All	
10	Finals	0.0054
	Break	0.0111
	Week 1	0.0046
	All	0.0076

Table 7: This table shows the results for a local polynomial regression fitting using a per-minute resolution. Darker shades of green indicate cells with higher standard error.

Table 8: Local Polynomial Regression Results & Standard Error - Hourly

Edison #	Week	Residual Standard Error
1	Finals	0.2942
	Break	
	Week 1	
	All	0.2942
2	Finals	0.0006
	Break	0.0021
	Week 1	0.0011
	All	0.0012
3	Finals	0.3048
	Break	0.1097
	Week 1	0.2039
	All	0.2739
4	Finals	0.1666
	Break	0.2339
	Week 1	0.0877
	All	0.3678
5	Finals	0.0006
	Break	0.0099
	Week 1	0.0007
	All	0.0059
6	Finals	0.0029
	Break	0.0106
	Week 1	0.0087
	All	0.0084
7	Finals	0.0558
	Break	0.1601
	Week 1	0.1193
	All	0.1414
8	Finals	0.0060
	Break	
	Week 1	
	All	0.0060
9	Finals	
	Break	
	Week 1	
	All	
10	Finals	0.0012
	Break	0.0024
	Week 1	0.0011
	All	0.0015

Table 8: This table shows the results for a local polynomial regression fitting using an hourly resolution. Darker shades of green indicate cells with higher standard error.

A perfect fit for any model would have a residual standard error of exactly zero, but realistically the residual standard error should be small and as close to zero as possible. As seen in Tables 7 and 8, the standard errors are quite small. Although Edisons 1, 3, 4 and 7 show higher error values in comparison to the five remaining devices, the errors are still low. Unlike the linear regression model, there does not appear to be any significant difference between the hourly and per-minute resolutions. In fact, corresponding values in Tables 7 and 8 are quite similar, which may be attributed to the local fitting of data as opposed to the global fitting in the linear regression.

Figure 11 shows an example of a local polynomial regression fitting using a per-minute resolution.

Figure 11: Local Polynomial Regression Fitting Example

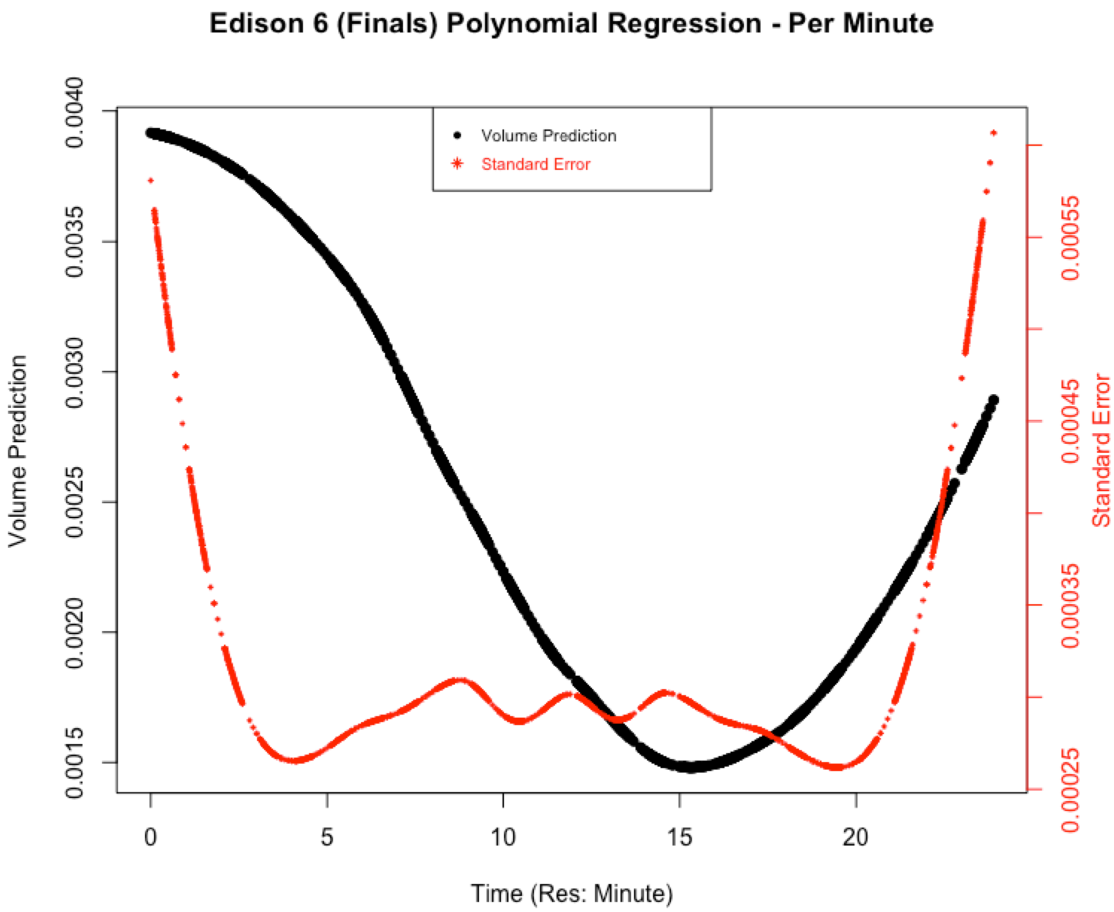

Figure 1l: This figure shows a local polynomial regression fitting for Edison 6 during finals week. Note the standard error axis has a separate scale. As indicated by the legend, the black circles indicate the fitted line, and the red stars signify the standard error.

Figure 11 shows the local polynomial regression fitting for Edison #6 during finals week, and the remainder of the devices and weeks can be found in Appendix B. As seen in Figure 11, the error remains quite low for the majority of the 24-hour period (note that the scale on the right y-axis is different, and much smaller than the left y-axis). A trend across devices and weeks is the standard error peaking during the late evening and early morning hours. This indicates that these times of day may be more unpredictable than others, perhaps due to larger variability of noise levels during those hours as discussed in Section 4.2.2.

Due to low standard errors in this model, we can say with greater confidence that a local polynomial regression fit yields a more accurate model than a linear fit, and furthermore, that it is indeed possible to create a predictive noise model using the methodology presented in this study.

Section 5: Conclusions

5.1 - Summary of Key Findings

A summary of key findings from each of the statistical analyses is as follows.

5.1.1 - Heat Maps:

1. Thursdays and Fridays tended to be the quietest days of the week, with more than half the locations exhibiting this pattern.

2. Five devices showed heightened volume in the late-night to early morning hours a few days throughout the week, approximately ranging from midnight to 10am. Four of these devices were placed along main corridors of the first floor.

3. Certain locations were much louder than others, with approximately 2 locations being very loud, 3 locations being very quiet, and 4 locations lying between loud and quiet.

4. Particular geographic features such as printers, classrooms, and doors tended to be louder, whereas back-corners and stand-up computers were quieter. Devices near the reference desk were inconclusive.

5.1.2 - K-Means:

1. There is a correlation between volume levels and duration, when looking at both hourly and per minute resolutions.

2. Periods of quiet tend to be extensive, whereas there is an absence of comparable periods of high volume levels, suggesting that high volume levels tend to be short-lived.

3. There is no strong relationship between volume and light levels.

5.1.3 - Linear Regression:

1. The hourly resolution yields slightly higher R-squared values, which may be attributed to data that is smoother than a per-minute resolution.

2. Both per-minute and hourly resolutions yield very small R-squared values, indicating that a predictive noise model based upon linear regression may be quite inaccurate.

5.1.4: Local Polynomial Regression:

1. A locally fitted model provides greater accuracy in comparison to a globally fitted model.

2. A polynomial fitting provides more accommodation for noise patterns and fluctuations than a linear one.

3. Error values throughout the day remain quite low. Higher errors are seen in the late night and early morning hours, indicating that these times may be more unpredictable.

4. A predictive noise model with moderate accuracy is possible using the methodology presented in this study when using an appropriately fitted predictive model.

Using conclusions from each of the statistical analyses, it is possible to conclude the methodology presented in this study does create a predictive noise model with reasonable accuracy. Not only does this method provide a repository of information about how the library is used during finals week, spring break, and week 1, but the results also suggest that extrapolating these noise patterns to other weeks is possible.

From the discussion in Sections 4.4 and 4.5 regarding the differences in linear regression and local polynomial regression fittings, the importance of finding the correct fit becomes clear. It is important to keep in mind that with a larger dataset, it may be possible to only use linear regression to determine the feasibility of creating an accurate noise model. However, in this study, that was not the case.

5.2 - Scope & Limitations

Perhaps the most significant limitation of this work is the length of data collection. Ideally, data would be collected for an extended period, such as multiple months, or an entire academic year. A larger dataset may stabilize random fluctuations, giving a clearer idea of how noise levels vary over time. For example, data collected for an academic year may give an idea of how noise varies from fall term to spring term, or from month-to-month. However, for the scope and goals of this project, a three-week collection period provided an appropriate amount of data for this proof-of-concept study.

Furthermore, the ability to visualize the results to the utmost level of detail were constrained by the amount of data. Reducing the data into per-minute or hourly granularities does introduce some error from averaging the data across a given range. However, a way to procedurally decide upon the appropriate resolution for the dataset is outside the scope of this work. Instead, a trial-and-error methodology is recommended based upon the size of the dataset. For example, similar to how per-minute and hourly resolutions were appropriate for a dataset consisting of a few weeks, daily or weekly resolutions may be appropriate for datasets consisting of multiple months or years.

5.3 - Future Work

The future of IoT technology is vast, and this work is the tip of a large, undiscovered iceberg. This project addresses critical topics within IoT, such as instrumenting accessible means for data collection and visualizing them in a useful way, but there are many more future works that can be based upon this starting point.

As the IoT continues to grow, one extremely relevant issue is how WiFi will handle increasing numbers of IoT devices connecting to the wireless network. It is estimated that there will be approximately 50 billion connected devices by 2020, a large majority of which rely on WiFi to communicate (Nordrum, "Popular Internet of Things”). This steep increase in IoT devices and embedded processors is likely to cause significant congestion on WiFi networks, as WiFi was not designed to handle an immense number of devices (Gates, “Build WiFi”). If the capacity of WiFi will plateau or begin to decrease when large number of devices continue to consume increasing bandwidth, other means of communications must be sought to accommodate IoT devices. One solution is using Bluetooth functionality to allow IoT devices to communicate and exchange data. Bluetooth is a method of communication which utilizes radio waves as opposed to physical, wired connections to connect devices together within short distances. Under the overarching umbrella of Bluetooth is Bluetooth Low Energy (BLE), which is an ultra-efficient version of Bluetooth that is fundamentally designed for small sensors. BLE consumes the minimal amount of energy necessary for these small sensors, allowing BLE-equipped sensors to last an extended period when using a small coin-cell battery (Bluetooth Special Interest Group, “How It Works”).

Increasing numbers of devices are coming pre-equipped with Bluetooth capabilities, and utilizing this method of data exchange will lessen the load on WiFi in the coming years. In turn, the sustainability of IoT devices will be maintained. For this reason, a relevant future trajectory for this work is utilizing BLE. One method would be to create a network of devices, with one device designated as the Bluetooth “hub”. The rest of the devices would all connect to the hub, which would then exchange data over WiFi. Instead of having multiple devices all streaming information over WiFi, there will be only one device with this role.

Furthermore, there are devices that are solely equipped with BLE capabilities that are equally, if not more, inexpensive than the Edisons; building future works around these devices will delve further into studying the feasibility of BLE communication. The Texas Instruments SimpleLink SensorTag is a small 1 x 2-inch package which currently only offers BLE capabilities in lieu of WiFi. This package retails for approximately 29 US Dollars (USD) and comes with a multitude of sensors including light, humidity, accelerometer, and more. Unfortunately, this device does not have sufficient software support for the microphone sensor and for that reason was not utilized in this work. However, creating a network of the SensorTag devices which all connect to a hub device with WiFi capabilities, such as the Edison, would be a plausible model for future research which would extend the work presented herein.

Appendix A: Hourly Heat Map Tables by Device

Tables 9 –17 below show heat map tables from each of the nine Edison devices. The heat maps shown highlight patterns amongst time of day and volume levels across the 9 different locations. Please refer to Table 4 in Results & Data Analysis for a table which combines all nine devices into one heat map table, highlighting volume differences between locations in the entire dataset.

Tables 9 – 17: Hourly Heat Maps by Location/Device

	Monday	**Tuesday**	**Wednesday**	**Thursday**	**Friday**
0:00	0.61	0.84	0.07	0.96	0.34
1:00	0.61	0.87	0.09	0.97	0.28
2:00	0.67	0.79	0.11	0.96	0.22
3:00	0.63	0.71	0.15	0.95	0.16
4:00	0.60	0.58	0.1	0.88	0.15
5:00	0.61	0.41	0.12	0.86	0.18
6:00	0.60	0.42	0.17	0.84	0.15
7:00	0.63	0.36	0.15	0.88	0.16
8:00	0.64	0.37	0.15	0.90	0.17
9:00	0.77	0.28	0.18	0.90	0.37
10:00	0.72	0.30	0.14	0.82	0.48
11:00	0.65	0.19	0.09	0.49	0.72
12:00	0.73	0.17	0.12	0.16	0.84
13:00	0.70	0.06	0.12	0.07	0.73
14:00	0.49	0.04	0.1	0.04	0.73
15:00	0.61	0.06	0.14	0.04	0.65
16:00	0.75	0.03	0.09	0.04	0.49
17:00	0.79	0.03	0.11	0.04	0.51
18:00	0.82	0.02	0.1	0.07	0.26
19:00	0.75	0.01	0.15	0.18	0.17
20:00	0.58	0.04	0.32	0.38	0.17
21:00	0.66	0.06	0.25	0.52	0.14

22:00	0.70	0.06	0.39	0.53	-
23:00	0.78	0.05	0.96	0.39	-

Table 9: This table shows hourly weekday data for Edison #1. Note that this device only collected data for finals week, so the values shown above are unique to finals week. Furthermore, this device did not capture data later in the day on Friday.

	Monday	**Tuesday**	**Wednesday**	**Thursday**	**Friday**
0:00	0.00063	0.00100	0.00067	0.00067	0.00000
1:00	0.00067	0.00167	0.00133	0.00100	0.00000
2:00	0.00100	0.00200	0.00133	0.00133	0.00000
3:00	0.00133	0.00200	0.00100	0.00167	0.00000
4:00	0.00100	0.00167	0.00100	0.00100	0.00000
5:00	0.00067	0.00167	0.00100	0.00033	0.00000
6:00	0.00067	0.00133	0.00033	0.00100	0.00000
7:00	0.00033	0.00067	0.00067	0.00000	0.00000
8:00	0.00033	0.00033	0.00033	0.00050	0.00000
9:00	0.00033	0.00000	0.00000	0.00000	0.00000
10:00	0.00067	0.00000	0.00000	0.00000	0.00000
11:00	0.00000	0.00000	0.00000	0.00000	0.00000
12:00	0.00033	0.00000	0.00033	0.00000	0.00000
13:00	0.00033	0.00000	0.00000	0.00000	0.00000
14:00	0.00000	0.00000	0.00000	0.00000	0.00000
15:00	0.00067	0.00067	0.00067	0.00000	0.00000
16:00	0.00033	0.00000	0.00000	0.00000	0.00000
17:00	0.00000	0.00000	0.00000	0.00000	0.00000
18:00	0.00467	0.00067	0.00433	0.00000	0.00000
19:00	0.00333	0.00133	0.00233	0.00000	0.00000
20:00	0.00000	0.00000	0.00000	0.00000	0.00000
21:00	0.00000	0.00000	0.00033	0.00000	0.00000
22:00	0.00067	0.00067	0.00067	0.00000	0.00000
23:00	0.00100	0.00133	0.00100	0.00000	0.00000

Table 10: This table shows hourly weekday data for Edison #2. Averaged over finals week, spring break, & week 1 of spring term.

	Monday	**Tuesday**	**Wednesday**	**Thursday**	**Friday**
0:00	0.27809	0.35367	0.67300	0.67433	0.26700
1:00	0.27133	0.32500	0.57400	0.79200	0.22900
2:00	0.18760	0.32433	0.28933	0.68233	0.16550
3:00	0.13866	0.44933	0.20867	0.63233	0.31450

4:00	0.09400	0.44300	0.16767	0.61600	0.27550
5:00	0.11767	0.37267	0.17367	0.48267	0.25100
6:00	0.08366	0.46533	0.09233	0.49650	0.32200
7:00	0.07666	0.50633	0.06700	0.50600	0.26000
8:00	0.06767	0.39733	0.09700	0.45450	0.18250
9:00	0.17433	0.36666	0.09733	0.38700	0.18800
10:00	0.29400	0.46567	0.16167	0.32200	0.24950
11:00	0.29166	0.56667	0.16967	0.22150	0.15700
12:00	0.26933	0.49266	0.12467	0.13500	0.14050
13:00	0.27067	0.38633	0.12300	0.10500	0.10400
14:00	0.30033	0.36067	0.21467	0.13650	0.10300
15:00	0.31000	0.38567	0.22800	0.10400	0.11750
16:00	0.31266	0.41167	0.24633	0.11750	0.11150
17:00	0.32067	0.36266	0.24633	0.11550	0.11750
18:00	0.35067	0.35933	0.25500	0.12350	0.11800
19:00	0.32167	0.31433	0.26100	0.13200	0.11050
20:00	0.32467	0.26467	0.26167	0.13450	0.11250
21:00	0.32967	0.35133	0.22967	0.12050	0.12200
22:00	0.35567	0.44567	0.28567	0.12900	0.13600
23:00	0.37333	0.58267	0.48200	0.13850	0.15900

Table 11: This table shows hourly weekday data for Edison #3. Averaged over finals week, spring break, & week 1 of spring term.

	Monday	Tuesday	Wednesday	Thursday	Friday
0:00	0.72305	0.76967	0.70700	0.53700	0.39650
1:00	0.75900	0.85000	0.74500	0.60933	0.37800
2:00	0.80033	0.76400	0.72533	0.66500	0.41750
3:00	0.82067	0.78900	0.68033	0.67367	0.40650
4:00	0.77833	0.83333	0.67333	0.68666	0.42600
5:00	0.74467	0.88933	0.68000	0.68133	0.40350
6:00	0.83300	0.83133	0.69033	0.52800	0.36500
7:00	0.87967	0.78133	0.70800	0.50300	0.28500
8:00	0.82300	0.77800	0.68833	0.40500	0.22750
9:00	0.77500	0.86660	0.68633	0.32900	0.22800
10:00	0.71300	0.85600	0.68400	0.26600	0.21350
11:00	0.70067	0.87133	0.67767	0.26600	0.18650
12:00	0.69867	0.82267	0.68767	0.25265	0.18200
13:00	0.67167	0.72333	0.56233	0.26900	0.19200
14:00	0.66933	0.63867	0.50700	0.27500	0.21600

15:00	0.66033	0.61867	0.50867	0.26100	0.19250
16:00	0.64233	0.57700	0.48967	0.26050	0.18800
17:00	0.62333	0.54433	0.51133	0.25650	0.17700
18:00	0.62067	0.58433	0.59667	0.27400	0.17000
19:00	0.61600	0.62433	0.63933	0.29650	0.16050
20:00	0.62000	0.63333	0.65767	0.33350	0.17200
21:00	0.62767	0.63167	0.66300	0.36200	0.17850
22:00	0.65467	0.65533	0.63967	0.38250	0.22400
23:00	0.66767	0.67833	0.55200	0.39050	0.27500

Table 12: This table shows hourly weekday data for Edison #4. Averaged over finals week, spring break, & week 1 of spring term.

	Monday	**Tuesday**	**Wednesday**	**Thursday**	**Friday**
0:00	0.00016	0.00067	0.00000	0.00033	0.00000
1:00	0.00033	0.00033	0.000333	0.00000	0.00033
2:00	0.00067	0.00100	0.000666	0.00067	0.00033
3:00	0.00000	0.00000	0.000333	0.00000	0.00000
4:00	0.00000	0.00033	0.00000	0.00000	0.00000
5:00	0.00166	0.00533	0.00400	0.00067	0.00467
6:00	0.00433	0.00067	0.001666	0.00033	0.00100
7:00	0.00000	0.00067	0.00000	0.00000	0.00000
8:00	0.02733	0.00000	0.00000	0.00000	0.00000
9:00	0.00033	0.00033	0.00033	0.00033	0.00033
10:00	0.00100	0.00067	0.00067	0.00067	0.00067
11:00	0.00100	0.00100	0.00067	0.00067	0.00067
12:00	0.00133	0.00100	0.00133	0.00167	0.00067
13:00	0.00100	0.00100	0.00100	0.00100	0.00067
14:00	0.00100	0.00100	0.00200	0.00100	0.00067
15:00	0.00100	0.00133	0.00133	0.00100	0.00067
16:00	0.00133	0.00067	0.00100	0.00100	0.00200
17:00	0.00167	0.00033	0.00133	0.00000	0.00000
18:00	0.00033	0.00067	0.00100	0.00033	0.00000
19:00	0.00067	0.00000	0.00000	0.00067	0.00050
20:00	0.00033	0.00000	0.00033	0.00033	0.00000
21:00	0.00067	0.00000	0.00000	0.00033	0.00000
22:00	0.00067	0.00000	0.00000	0.00000	0.00000
23:00	0.00033	0.00033	0.00000	0.00000	0.00050

Table 13: This table shows hourly weekday data for Edison #5. Averaged over finals week, spring break, & week 1 of spring term.

	Monday	Tuesday	Wednesday	Thursday	Friday
0:00	0.01236	0.01100	0.00533	0.00700	0.00000
1:00	0.01900	0.01367	0.00467	0.01133	0.00000
2:00	0.01733	0.01367	0.00500	0.02633	0.00000
3:00	0.01566	0.01967	0.00633	0.03033	0.00000
4:00	0.01100	0.01600	0.00666	0.01167	0.00000
5:00	0.00800	0.01567	0.00767	0.00833	0.00000
6:00	0.00633	0.00700	0.00500	0.00733	0.00000
7:00	0.00633	0.00633	0.00567	0.00150	0.00000
8:00	0.00333	0.00933	0.00233	0.00100	0.00000
9:00	0.00400	0.00567	0.00133	0.00050	0.00000
10:00	0.00400	0.00500	0.00233	0.00000	0.00000
11:00	0.00467	0.00267	0.00233	0.00000	0.00000
12:00	0.00400	0.00200	0.00133	0.00100	0.00000
13:00	0.00400	0.00167	0.00200	0.00000	0.00000
14:00	0.00133	0.00033	0.00167	0.00000	0.00000
15:00	0.00133	0.00200	0.00167	0.00000	0.00000
16:00	0.00200	0.00233	0.00167	0.00000	0.00000
17:00	0.00133	0.00133	0.00067	0.00000	0.00000
18:00	0.01433	0.00600	0.01133	0.00000	0.00000
19:00	0.01433	0.00600	0.01000	0.00000	0.00000
20:00	0.00166	0.00100	0.00100	0.00000	0.00000
21:00	0.00400	0.00233	0.00233	0.00000	0.00000
22:00	0.00533	0.00733	0.00767	0.00000	0.00000
23:00	0.00733	0.01067	0.00833	0.00000	0.00000

Table 14: This table shows hourly weekday data for Edison #6. Averaged over finals week, spring break, & week 1 of spring term.

	Monday	Tuesday	Wednesday	Thursday	Friday
0:00	0.94229	0.88533	0.90666	0.85433	0.81800
1:00	0.93600	0.89833	0.89300	0.86267	0.75050
2:00	0.90067	0.81500	0.85100	0.88633	0.78650
3:00	0.79867	0.79133	0.83330	0.85733	0.81150
4:00	0.60700	0.76167	0.73733	0.80000	0.79600
5:00	0.51266	0.80467	0.76100	0.76000	0.80050
6:00	0.48633	0.81933	0.81100	0.73200	0.78050
7:00	0.44233	0.80800	0.81767	0.69100	0.74250
8:00	0.39800	0.76100	0.71300	0.79600	0.73650

9:00	0.50200	0.74000	0.72900	0.78150	0.72100
10:00	0.69733	0.83500	0.79433	0.73250	0.68050
11:00	0.79200	0.88400	0.83533	0.74600	0.68600
12:00	0.76367	0.85200	0.84433	0.79750	0.74200
13:00	0.78333	0.81666	0.77600	0.82950	0.79600
14:00	0.83600	0.84666	0.78167	0.85550	0.81200
15:00	0.87067	0.84233	0.82167	0.85700	0.80750
16:00	0.88367	0.82933	0.80367	0.87100	0.79450
17:00	0.87000	0.83200	0.76933	0.87450	0.79100
18:00	0.84433	0.82767	0.75067	0.88000	0.77300
19:00	0.81633	0.85467	0.76433	0.87750	0.77600
20:00	0.86267	0.88100	0.79467	0.88600	0.79050
21:00	0.85700	0.88200	0.86367	0.88550	0.79100
22:00	0.86967	0.88433	0.88033	0.87700	0.80350
23:00	0.86933	0.90067	0.83900	0.85700	0.81650

Table 15: This table shows hourly weekday data for Edison #7. Averaged over finals week, spring break, & week 1 of spring term.

	Monday	**Tuesday**	**Wednesday**	**Thursday**	**Friday**
0:00	0.01686	0.01000	0.00800	0.00400	0.00000
1:00	0.02800	0.00900	0.00900	0.00500	0.00000
2:00	0.03700	0.01100	0.00200	0.00400	0.00000
3:00	0.01400	0.00800	0.00200	0.00400	0.00000
4:00	0.01000	0.02100	0.00500	0.00300	0.00000
5:00	0.00600	0.00900	0.00200	0.00500	0.00000
6:00	0.01400	0.01600	0.00100	0.00300	0.00000
7:00	0.00700	0.00300	0.00100	0.00200	0.00000
8:00	0.00200	0.00300	0.00100	0.00200	0.00000
9:00	0.00400	0.01000	0.00100	0.00000	0.00000
10:00	0.00500	0.00300	0.00100	0.00000	0.00000
11:00	0.00300	0.00600	0.00200	0.00000	0.00000
12:00	0.00200	0.00100	0.00000	0.00200	0.00000
13:00	0.00800	0.00600	0.00200	0.00100	0.00000
14:00	0.00200	0.00200	0.00100	0.00000	0.00000
15:00	0.00400	0.00500	0.00400	0.00000	0.00000
16:00	0.00100	0.00200	0.00000	0.00000	0.00000
17:00	0.00000	0.00100	0.00000	0.00000	0.00000
18:00	0.00200	0.01600	0.02000	0.00000	0.00000
19:00	0.03400	0.00100	0.00200	0.00000	0.00000

20:00	0.00100	0.00100	0.00000	0.00000	0.00000
21:00	0.01700	0.00300	0.00100	0.00000	0.00000
22:00	0.00300	0.00800	0.00300	0.00000	0.00000
23:00	0.00600	0.01400	0.00400	0.00000	0.00000

Table 16: This table shows hourly weekday data for Edison #8. Note that this device only collected data for finals week, so the values shown above are unique to finals week.

	Monday	**Tuesday**	**Wednesday**	**Thursday**	**Friday**
0:00	0.00084	0.00067	0.00067	0.00267	0.00000
1:00	0.00100	0.00067	0.00100	0.00067	0.00000
2:00	0.00600	0.00133	0.00133	0.00067	0.00000
3:00	0.00167	0.00133	0.00133	0.00067	0.00000
4:00	0.00167	0.00133	0.00067	0.00033	0.00000
5:00	0.00133	0.00133	0.00133	0.00033	0.00000
6:00	0.00100	0.00100	0.00133	0.00033	0.00000
7:00	0.00067	0.00100	0.00100	0.00050	0.00000
8:00	0.00033	0.00033	0.00033	0.00100	0.00000
9:00	0.00000	0.00033	0.00000	0.00050	0.00000
10:00	0.00033	0.00000	0.00000	0.00000	0.00000
11:00	0.00000	0.00000	0.00000	0.00000	0.00000
12:00	0.00033	0.00000	0.00000	0.00000	0.00000
13:00	0.00033	0.00033	0.00000	0.00000	0.00000
14:00	0.00033	0.00033	0.00000	0.00000	0.00000
15:00	0.00100	0.00067	0.00067	0.00000	0.00000
16:00	0.00100	0.00100	0.00100	0.00000	0.00000
17:00	0.00000	0.00000	0.00000	0.00000	0.00000
18:00	0.00400	0.00300	0.00433	0.00000	0.00000
19:00	0.00533	0.00133	0.00200	0.00000	0.00000
20:00	0.00000	0.00000	0.00000	0.00000	0.00000
21:00	0.00000	0.00000	0.00000	0.00000	0.00000
22:00	0.00033	0.00033	0.00000	0.00000	0.00000
23:00	0.00067	0.00100	0.00167	0.00000	0.00000

Table 17: This table shows hourly weekday data for Edison #10. Averaged over finals week, spring break, & week 1 of spring term.

Appendix B: Local Polynomial Regression Fitting Graphs

This section shows local polynomial regression fittings for each of the nine locations in both a per-minute and hourly resolution. Please refer to Section 4.5 for analysis and further information.

Per Minute Resolution

Finals Week

Figures 12 – 20: Local Polynomial Regressions Per Minute – Finals Week

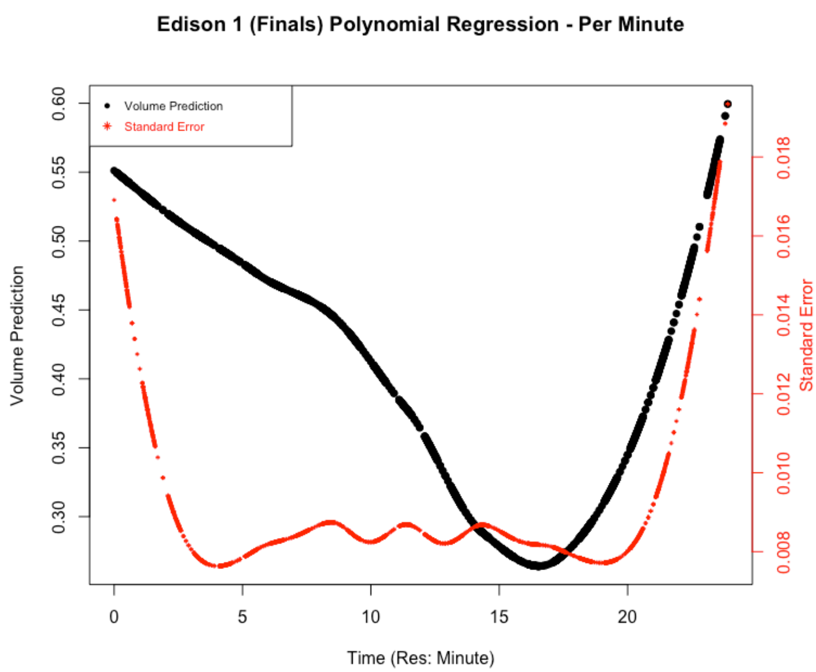

Figure 12: This figure shows the local polynomial regression fitting and standard error for Edison 1 during finals week using a per-minute resolution.

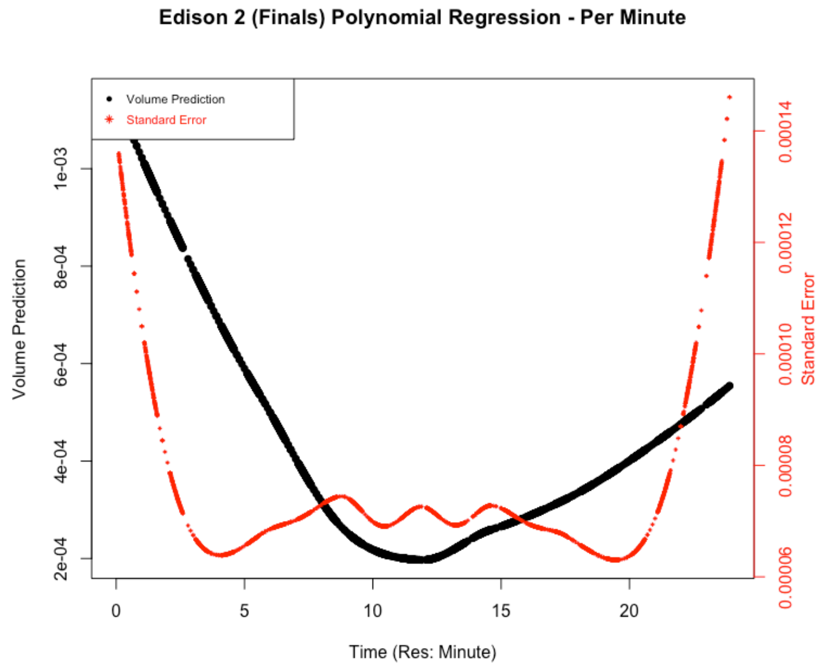

Figure 13: This figure shows the local polynomial regression fitting and standard error for Edison 2 during finals week using a per-minute resolution.

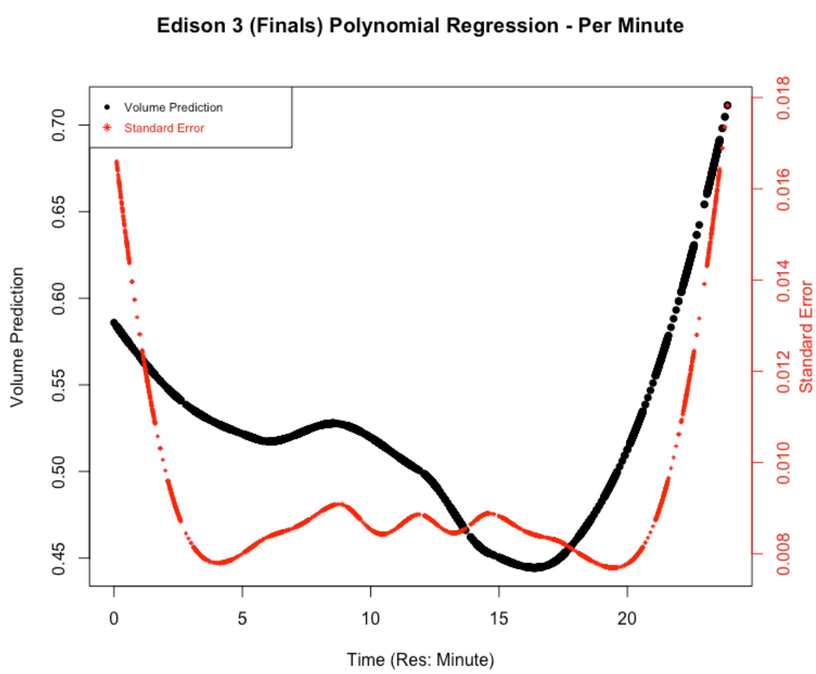

Figure 14: This figure shows the local polynomial regression fitting and standard error for Edison 3 during finals week using a per-minute resolution.

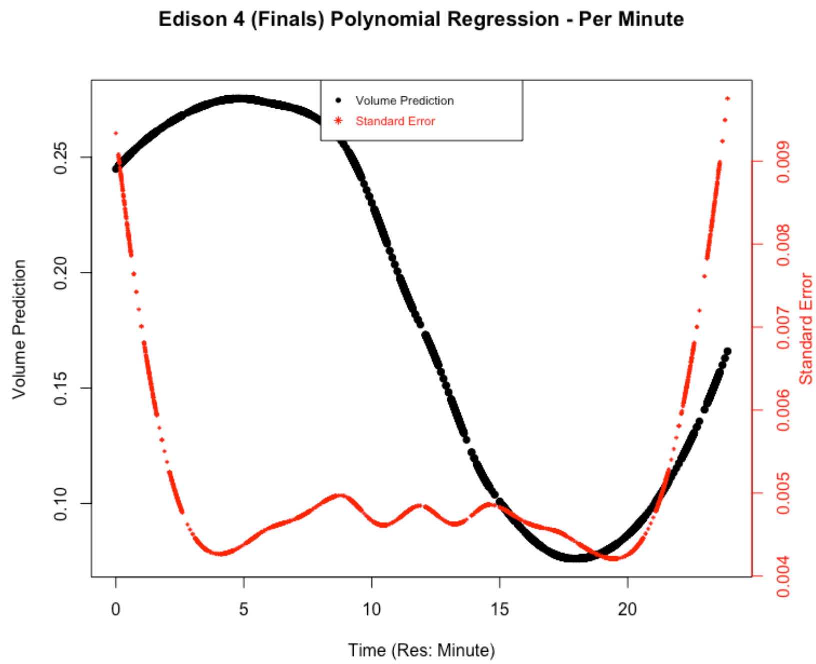

Figure 15: This figure shows the local polynomial regression fitting and standard error for Edison 4 during finals week using a per-minute resolution.

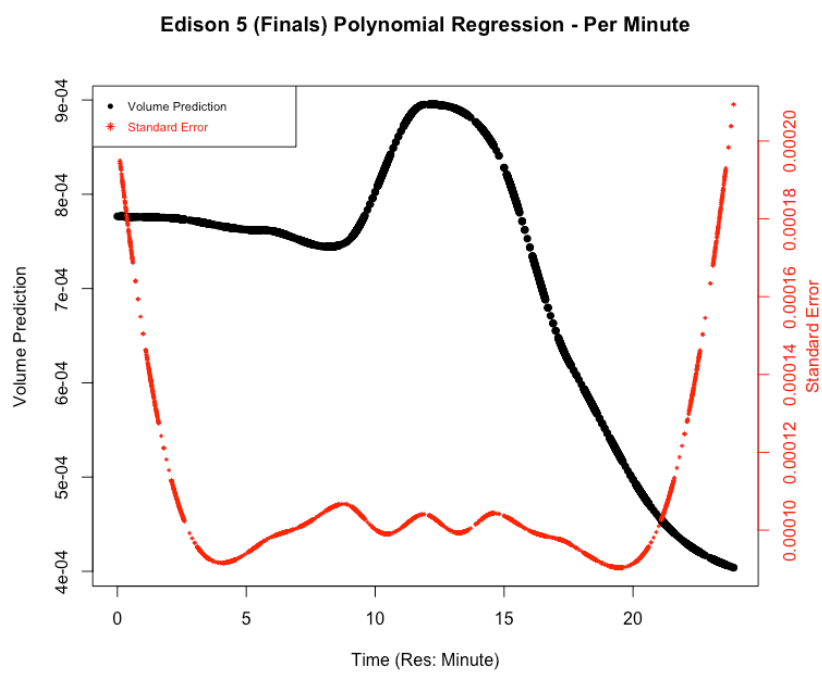

Figure 16: This figure shows the local polynomial regression fitting and standard error for Edison 5 during finals week using a per-minute resolution.

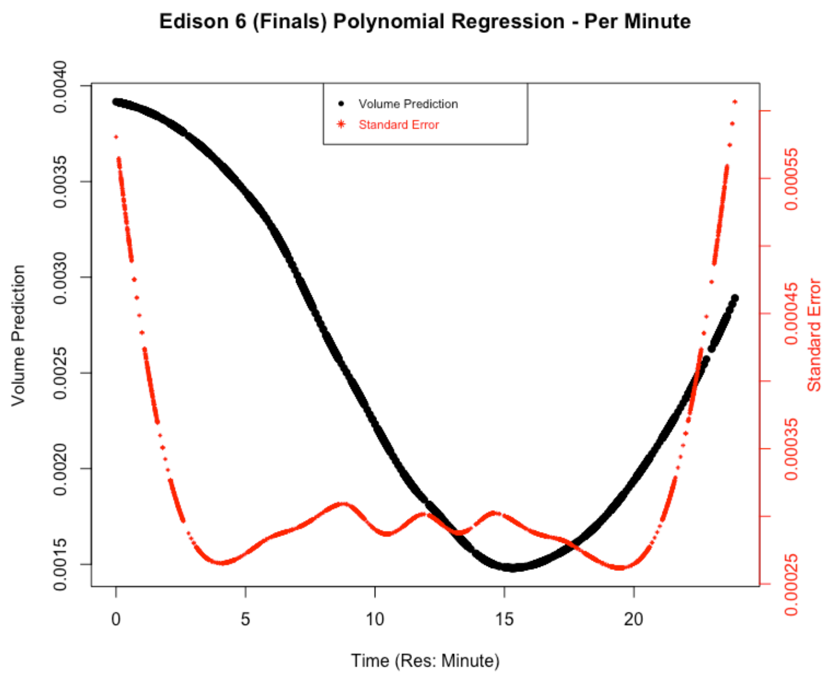

Figure 17: This figure shows the local polynomial regression fitting and standard error for Edison 6 during finals week using a per-minute resolution.

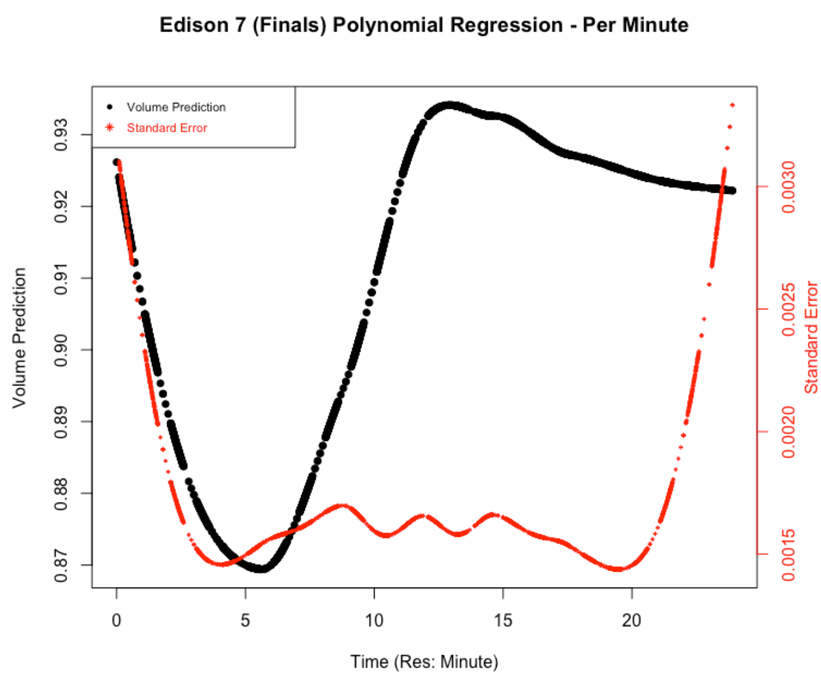

Figure 18: This figure shows the local polynomial regression fitting and standard error for Edison 7 during finals week using a per-minute resolution.

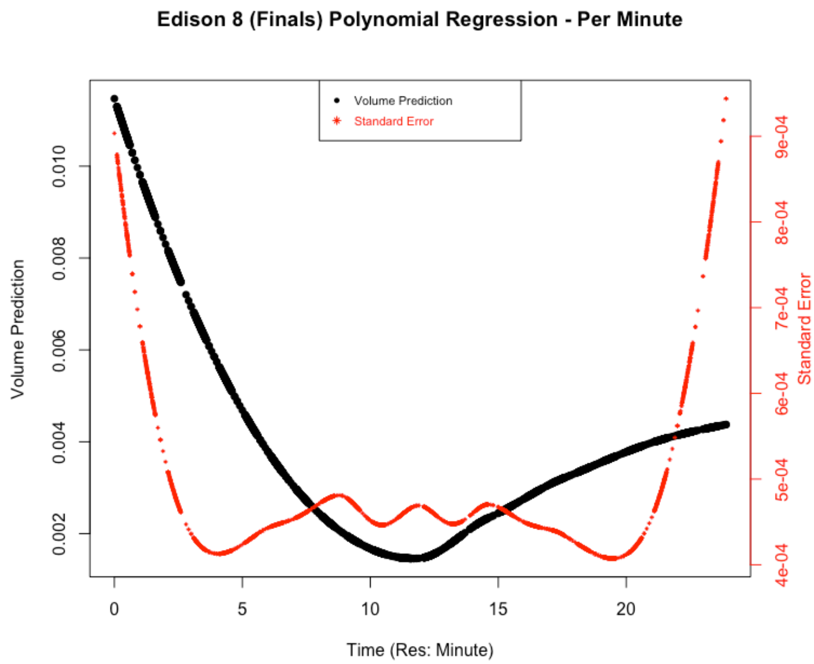

Figure 19: This figure shows the local polynomial regression fitting and standard error for Edison 8 during finals week using a per-minute resolution.

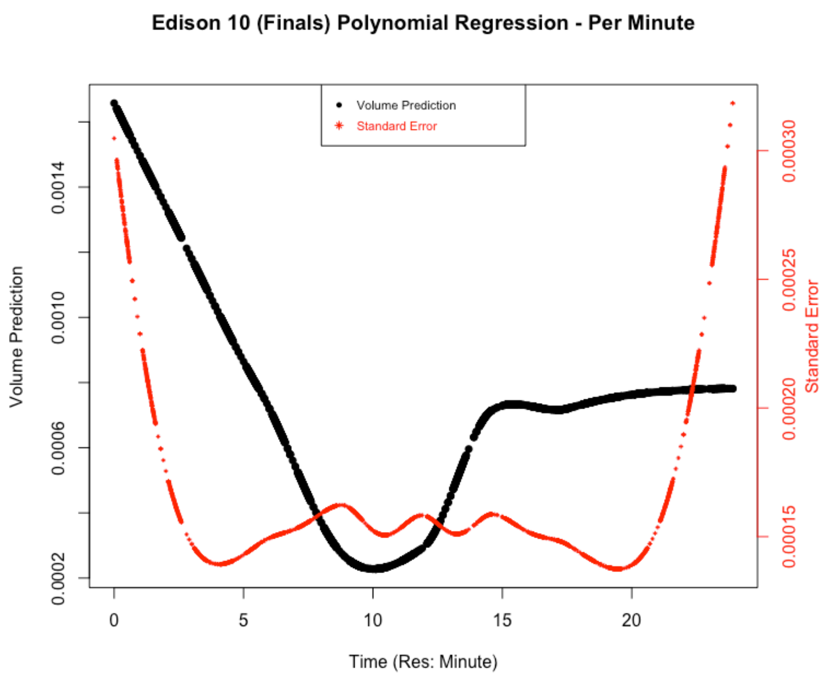

Figure 20: This figure shows the local polynomial regression fitting and standard error for Edison 10 during finals week using a per-minute resolution.

Spring Break

Figures 21 – 27: Local Polynomial Regressions Per Minute – Spring Break

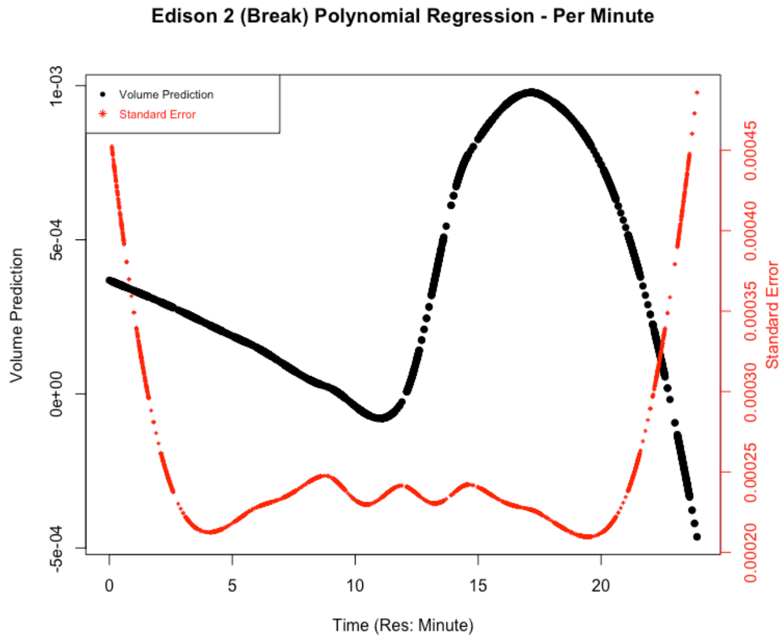

Figure 21: This figure shows the local polynomial regression fitting and standard error for Edison 2 during spring break using a per-minute resolution.

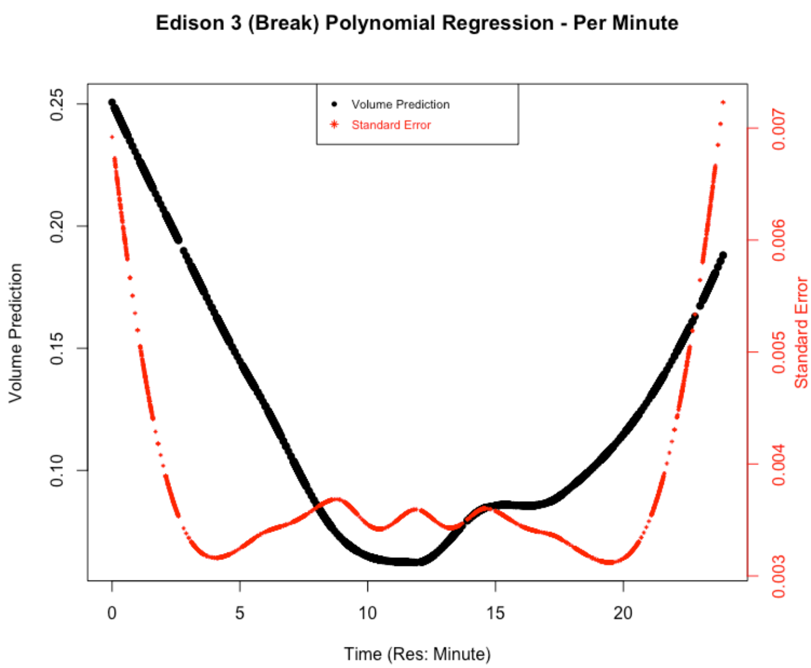

Figure 22: This figure shows the local polynomial regression fitting and standard error for Edison 3 during spring break using a per-minute resolution.

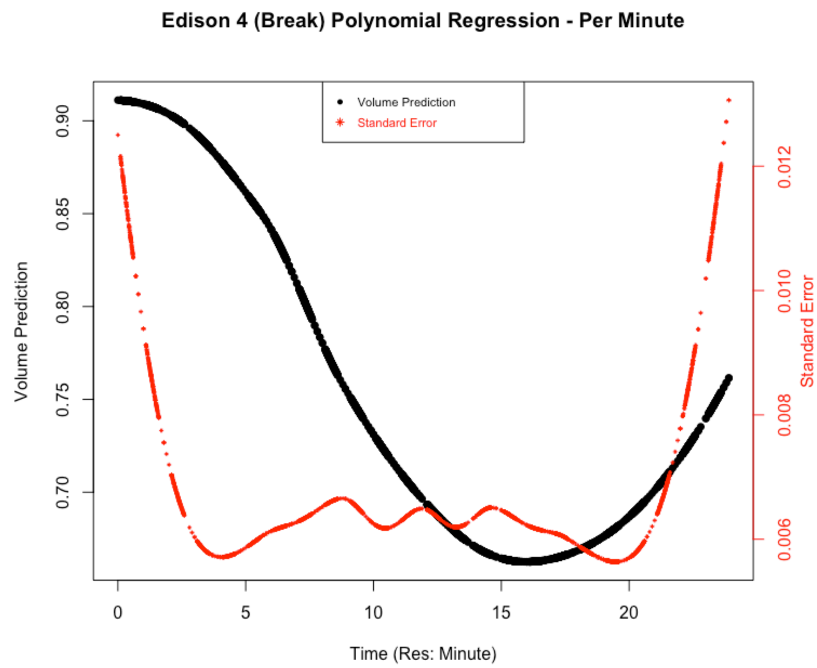

Figure 23: This figure shows the local polynomial regression fitting and standard error for Edison 4 during spring break using a per-minute resolution.

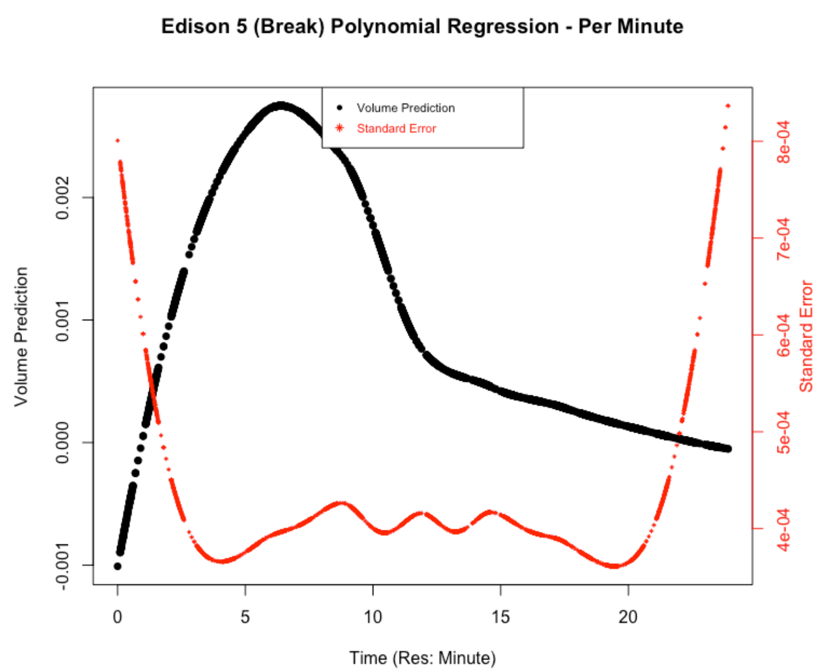

Figure 24: This figure shows the local polynomial regression fitting and standard error for Edison 5 during spring break using a per-minute resolution.

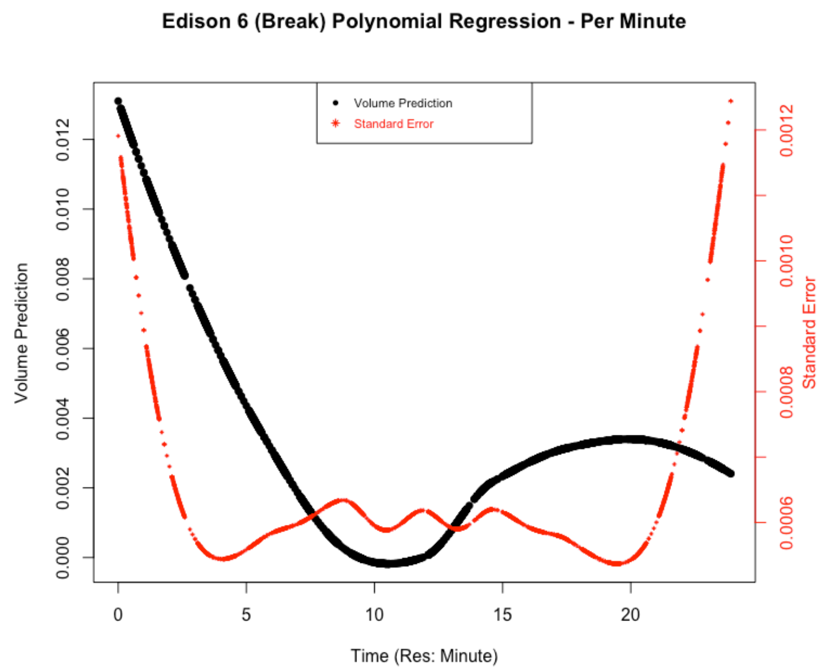

Figure 25: This figure shows the local polynomial regression fitting and standard error for Edison 6 during spring break using a per-minute resolution.

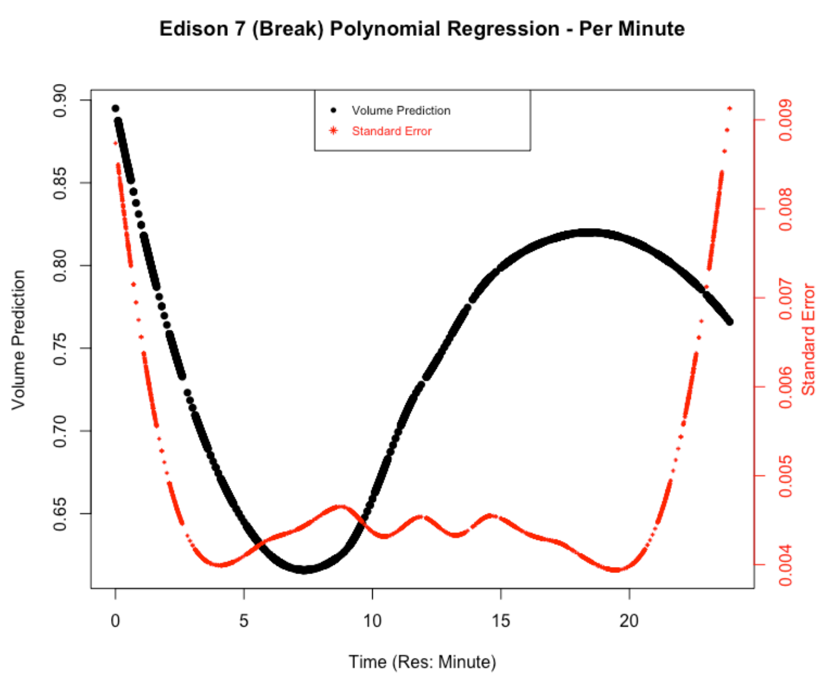

Figure 26: This figure shows the local polynomial regression fitting and standard error for Edison 7 during spring break using a per-minute resolution.

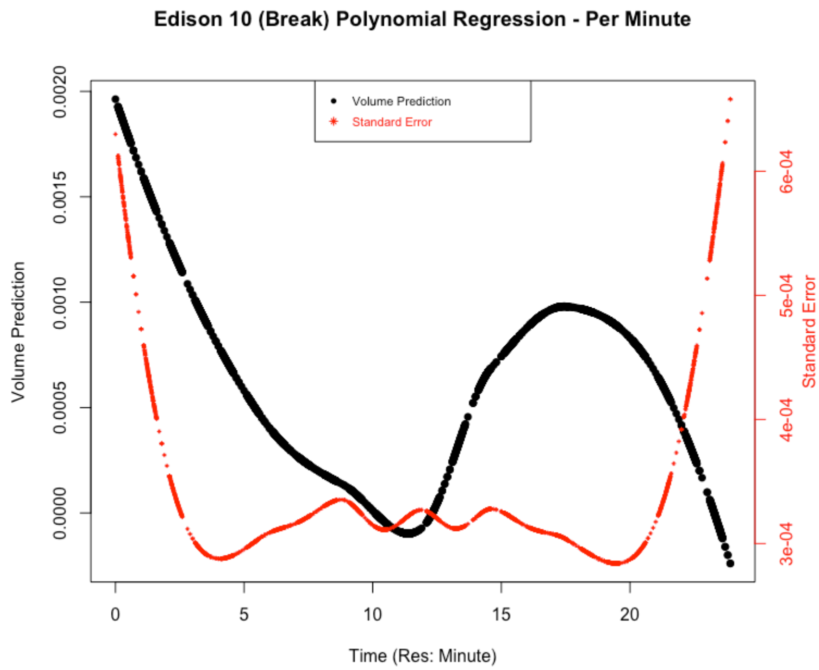

Figure 27: This figure shows the local polynomial regression fitting and standard error for Edison 10 during spring break using a per-minute resolution.

Week 1

Figures 28 –33: Local Polynomial Regressions Per Minute – Week 1

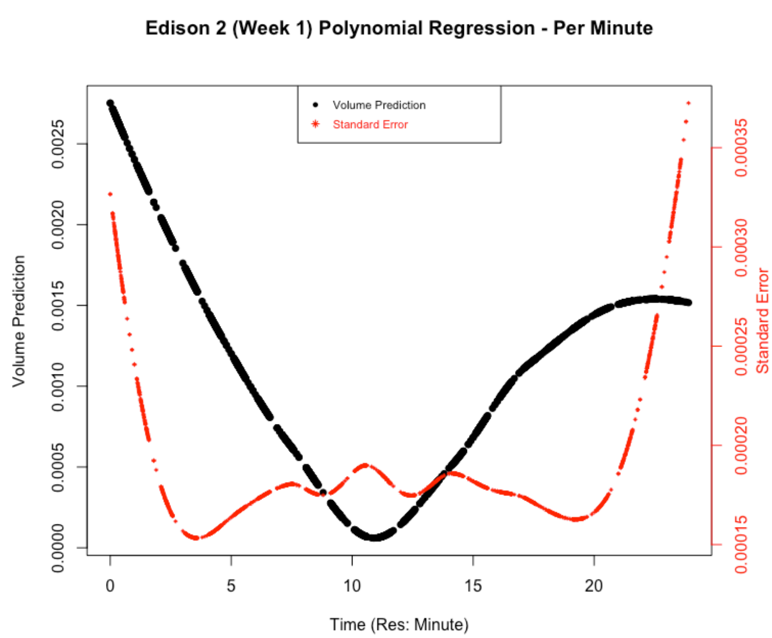

Figure 28: This figure shows the local polynomial regression fitting and standard error for Edison 2 during week 1 using a per-minute resolution.

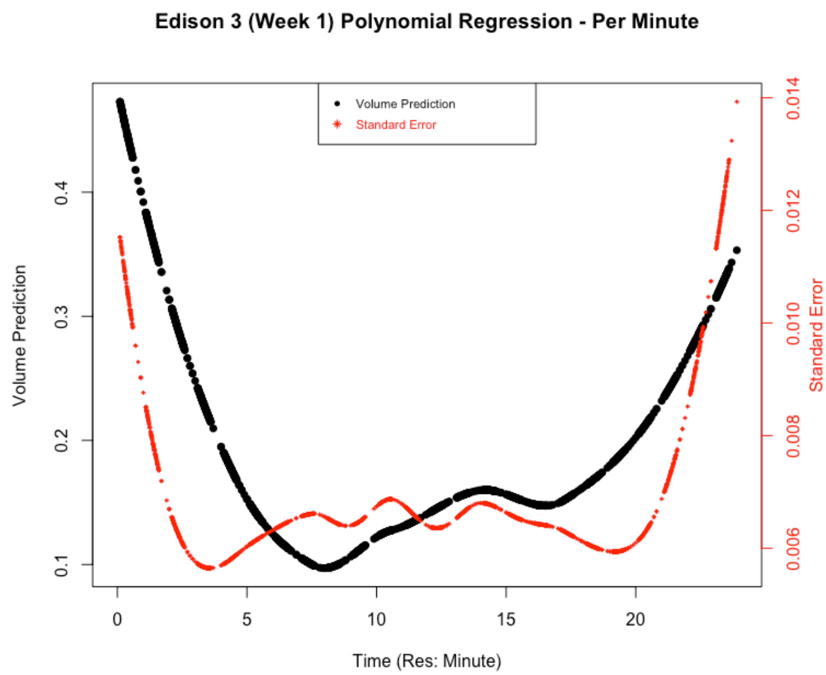

Figure 29: This figure shows the local polynomial regression fitting and standard error for Edison 3 during week 1 using a per-minute resolution.

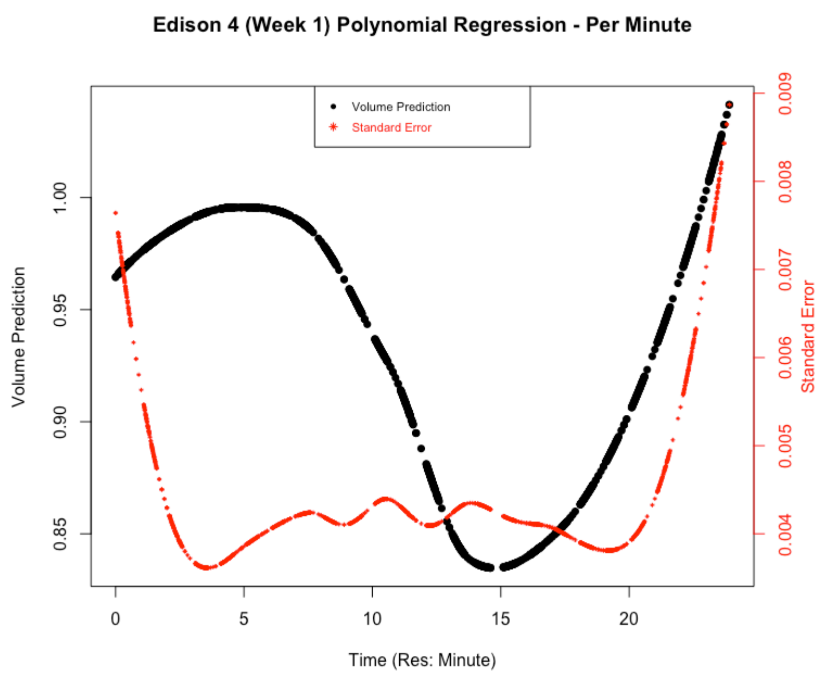

Figure 30: This figure shows the local polynomial regression fitting and standard error for Edison 4 during week 1 using a per-minute resolution.

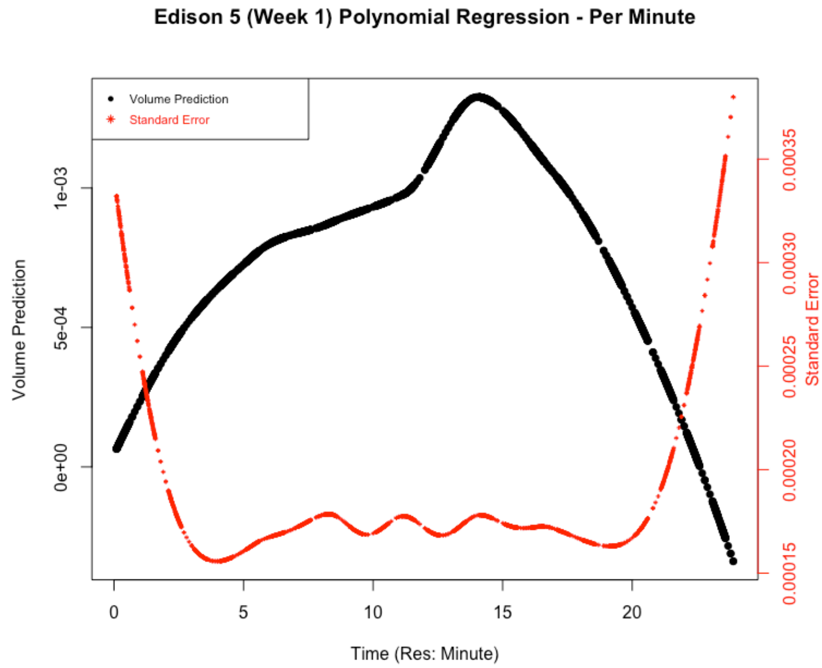

Figure 31: This figure shows the local polynomial regression fitting and standard error for Edison 5 during week 1 using a per-minute resolution.

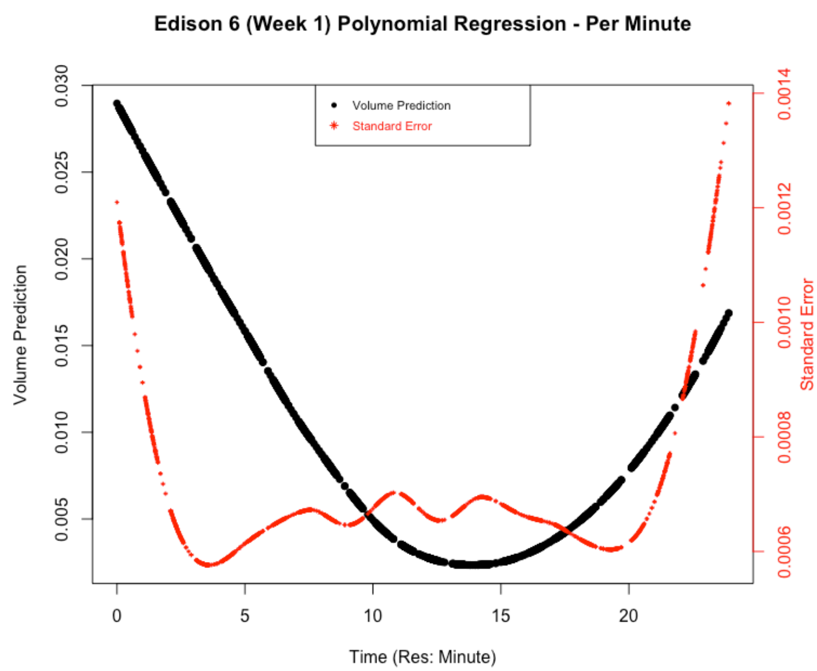

Figure 32: This figure shows the local polynomial regression fitting and standard error for Edison 6 during week 1 using a per-minute resolution.

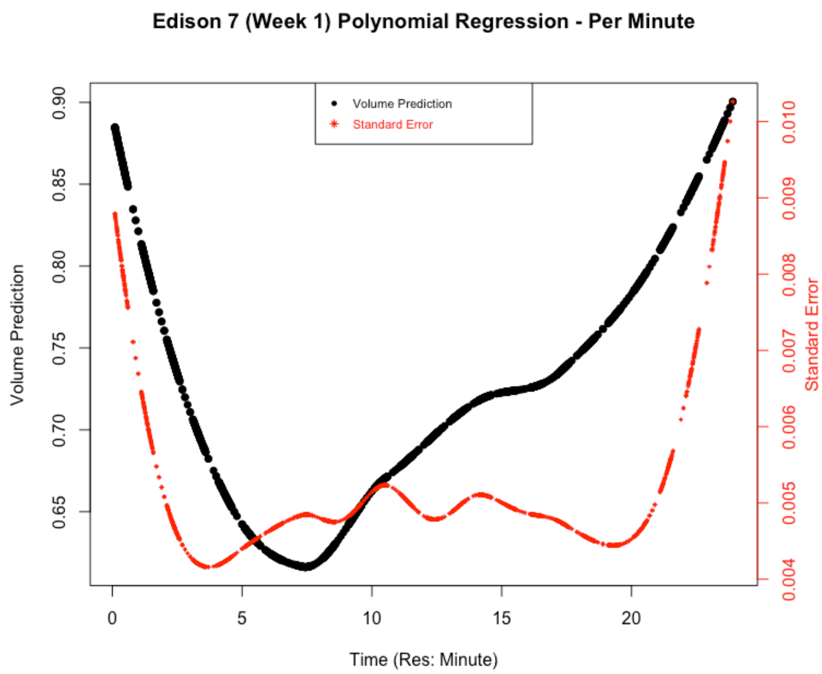

Figure 33: This figure shows the local polynomial regression fitting and standard error for Edison 7 during week 1 using a per-minute resolution.

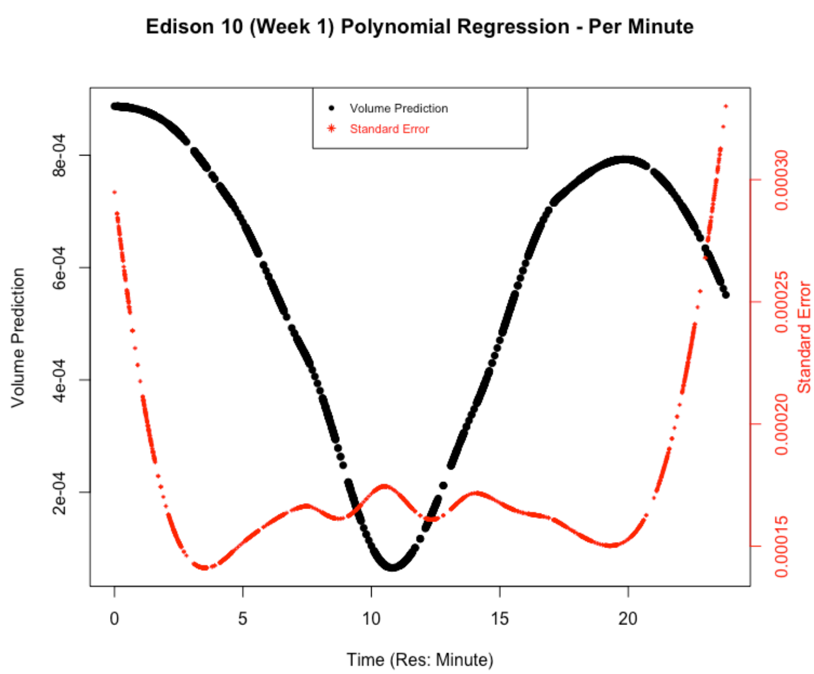

Figure 33: This figure shows the local polynomial regression fitting and standard error for Edison 10 during week 1 using a per-minute resolution.

All Weeks

Figures 34 – 42: Local Polynomial Regressions Per Minute – All Weeks Combined

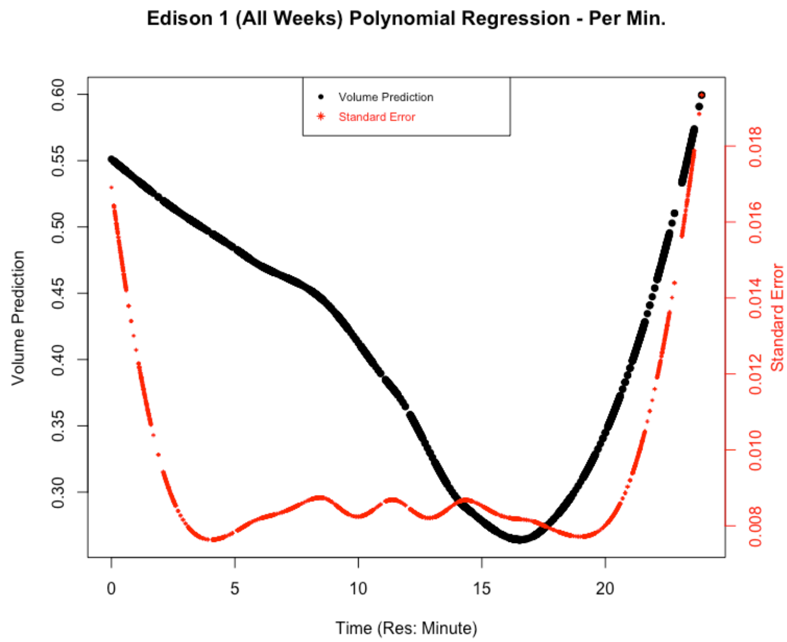

Figure 34: This figure shows the local polynomial regression fitting and standard error for Edison 1 across all weeks using a per-minute resolution.

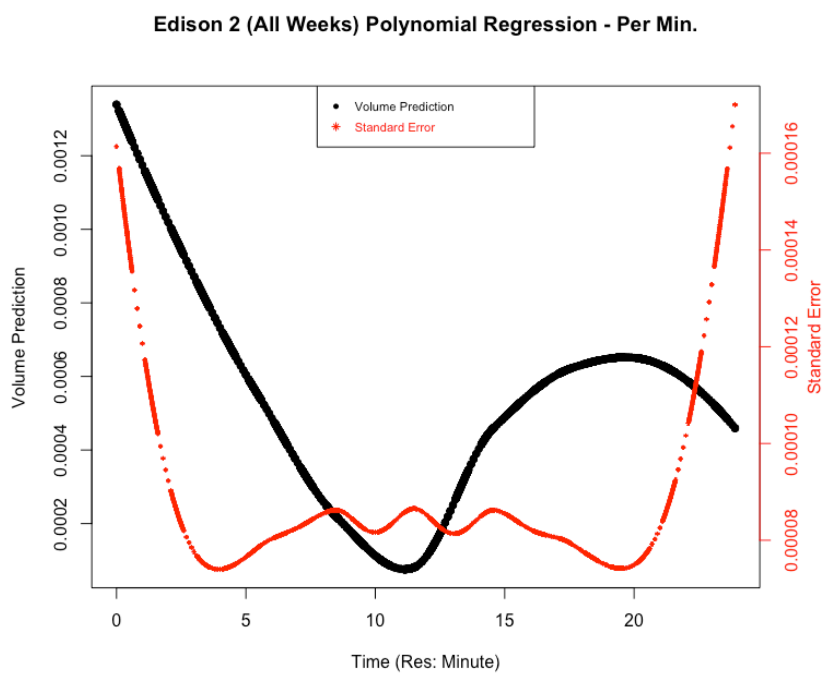

Figure 35: This figure shows the local polynomial regression fitting and standard error for Edison 2 across all weeks using a per-minute resolution.

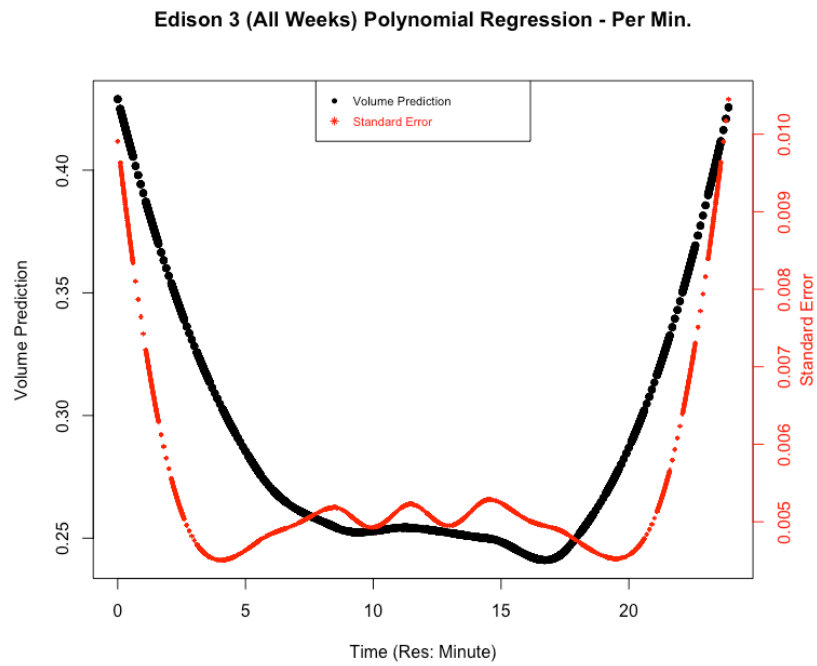

Figure 36: This figure shows the local polynomial regression fitting and standard error for Edison 3 across all weeks using a per-minute resolution.

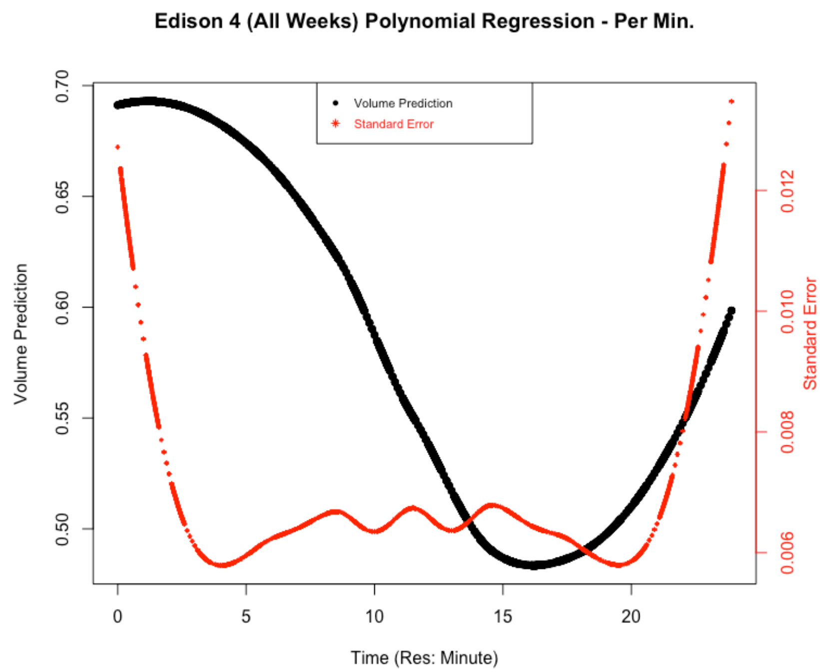

Figure 37: This figure shows the local polynomial regression fitting and standard error for Edison 4 across all weeks using a per-minute resolution.

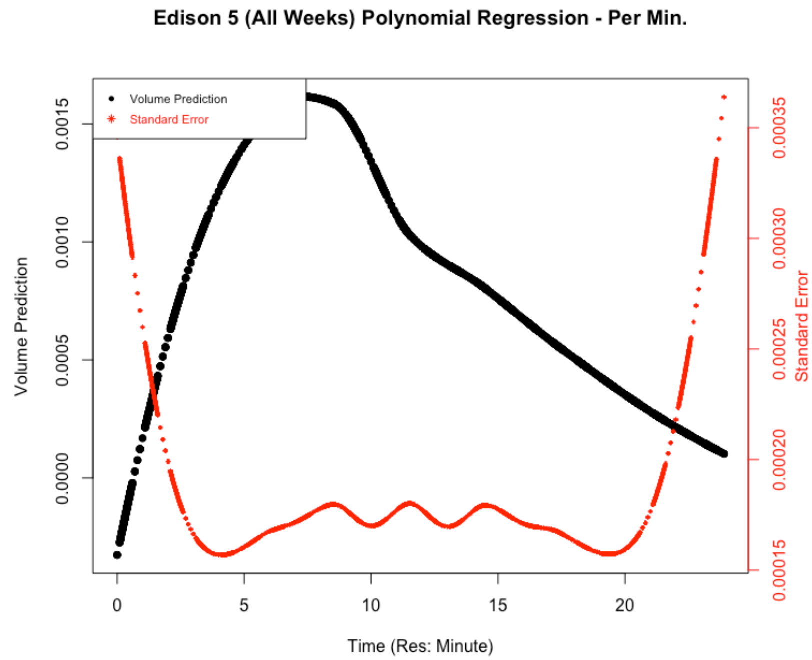

Figure 38: This figure shows the local polynomial regression fitting and standard error for Edison 5 across all weeks using a per-minute resolution.

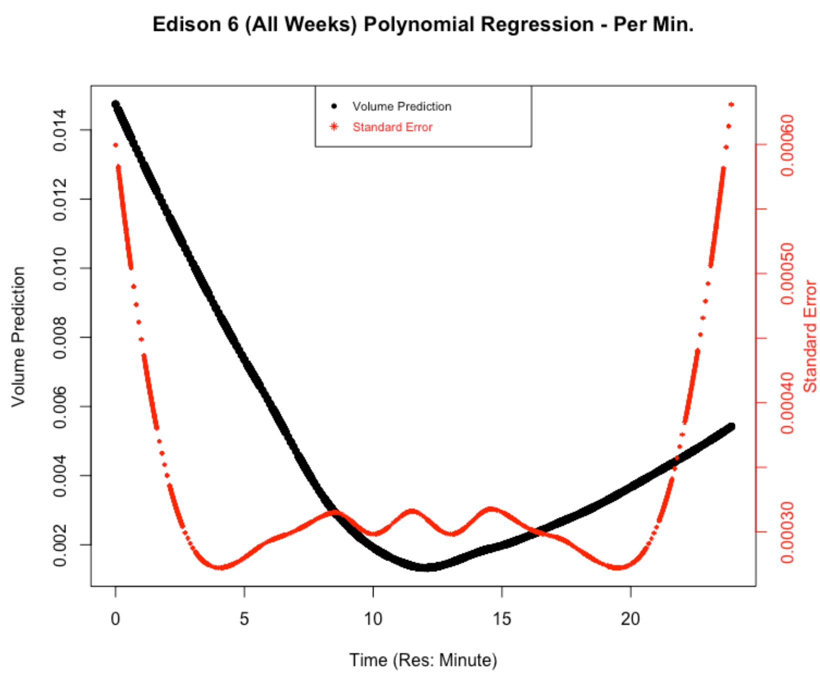

Figure 39: This figure shows the local polynomial regression fitting and standard error for Edison 6 across all weeks using a per-minute resolution.

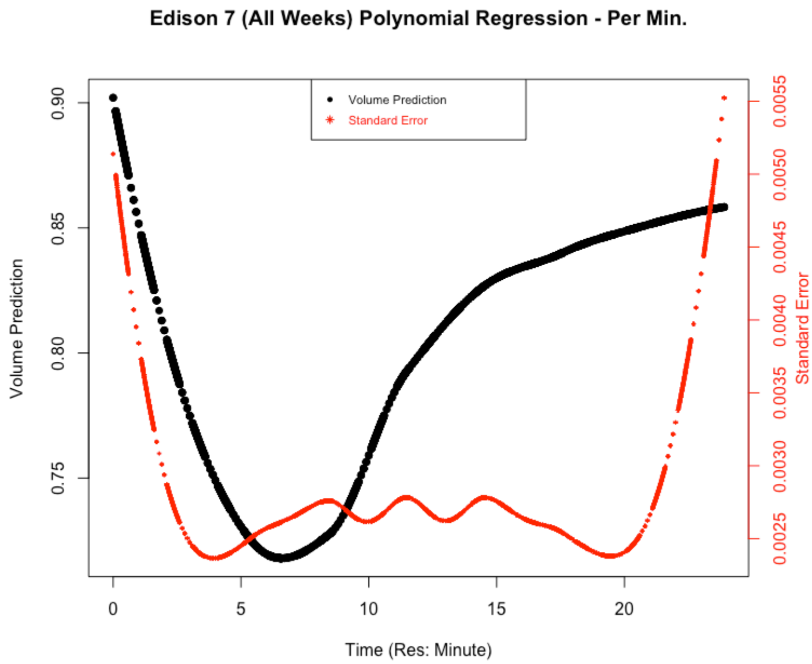

Figure 40: This figure shows the local polynomial regression fitting and standard error for Edison 7 across all weeks using a per-minute resolution.

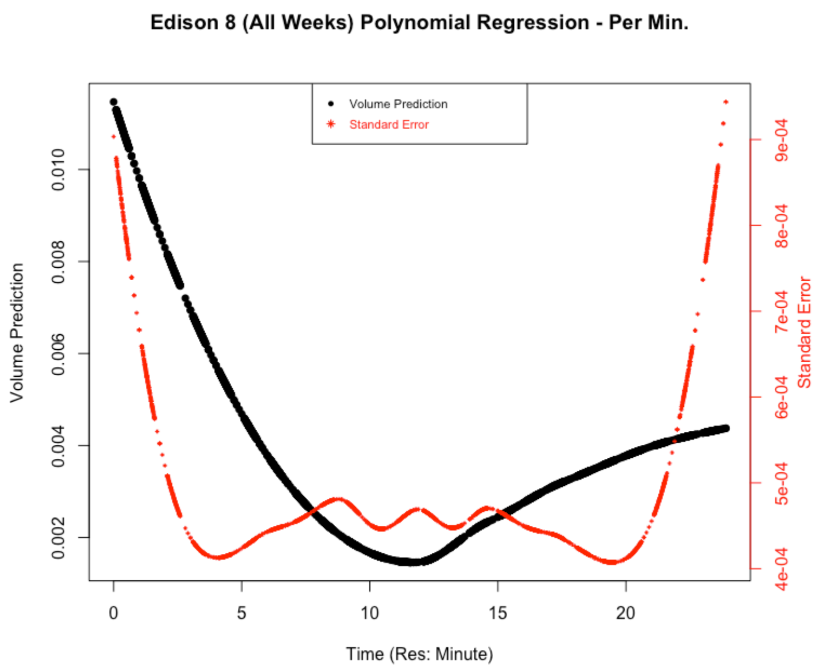

Figure 41: This figure shows the local polynomial regression fitting and standard error for Edison 8 across all weeks using a per-minute resolution.

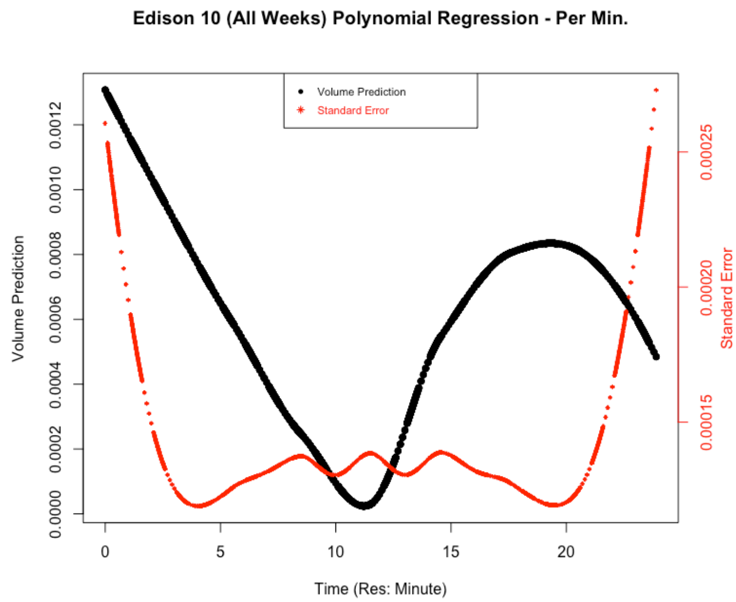

Figure 42: This figure shows the local polynomial regression fitting and standard error for Edison 10 across all weeks using a per-minute resolution.

Per Hour Resolution

Finals Week

Figures 43 – 51: Local Polynomial Regressions Hourly – Finals Week

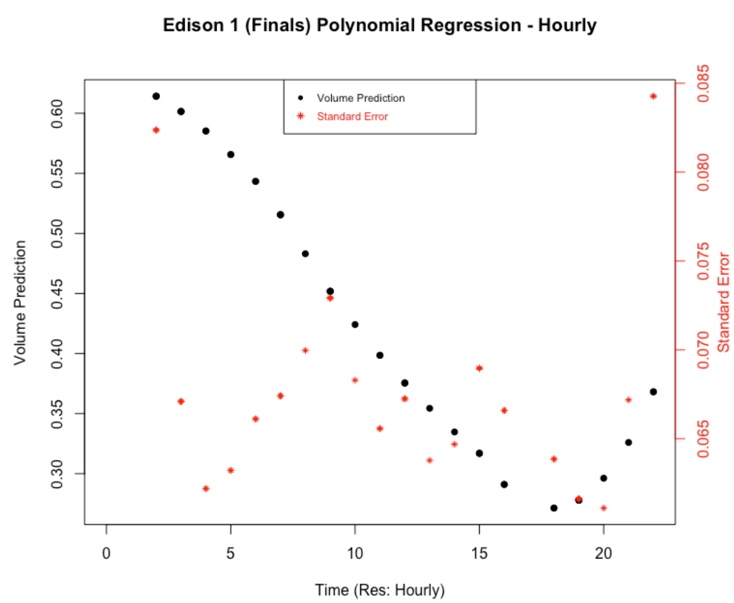

Figure 43: This figure shows the local polynomial regression fitting and standard error for Edison 1 during finals week using an hourly resolution.

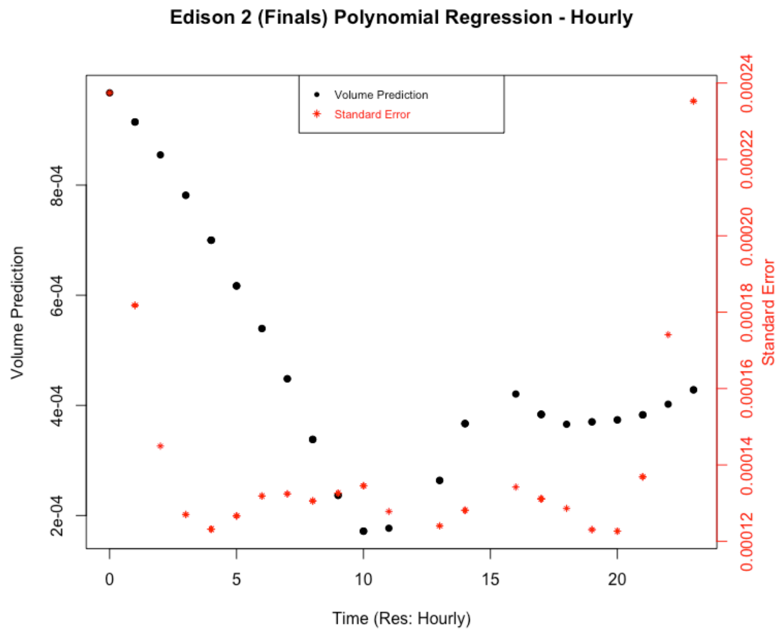

Figure 44: This figure shows the local polynomial regression fitting and standard error for Edison 2 during finals week using an hourly resolution.

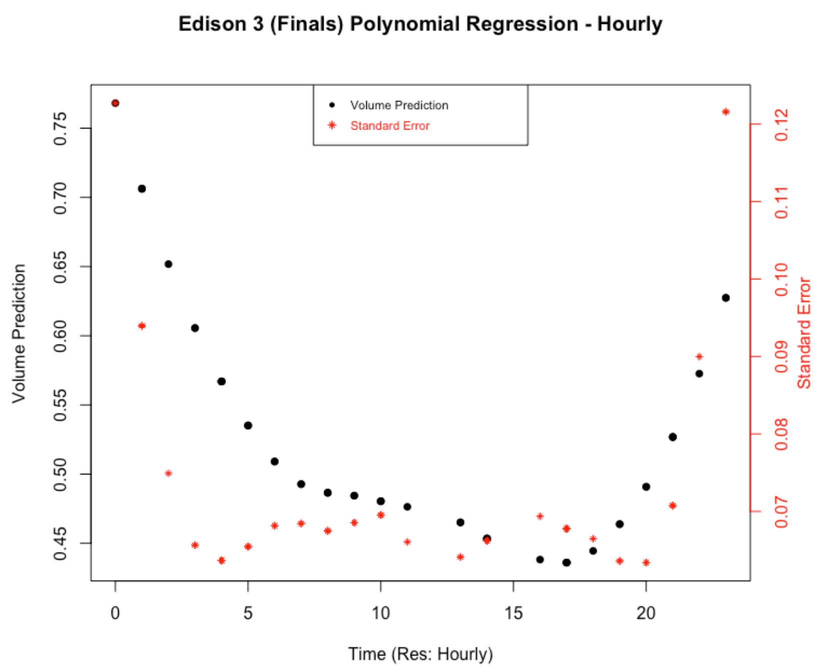

Figure 45: This figure shows the local polynomial regression fitting and standard error for Edison 3 during finals week using an hourly resolution.

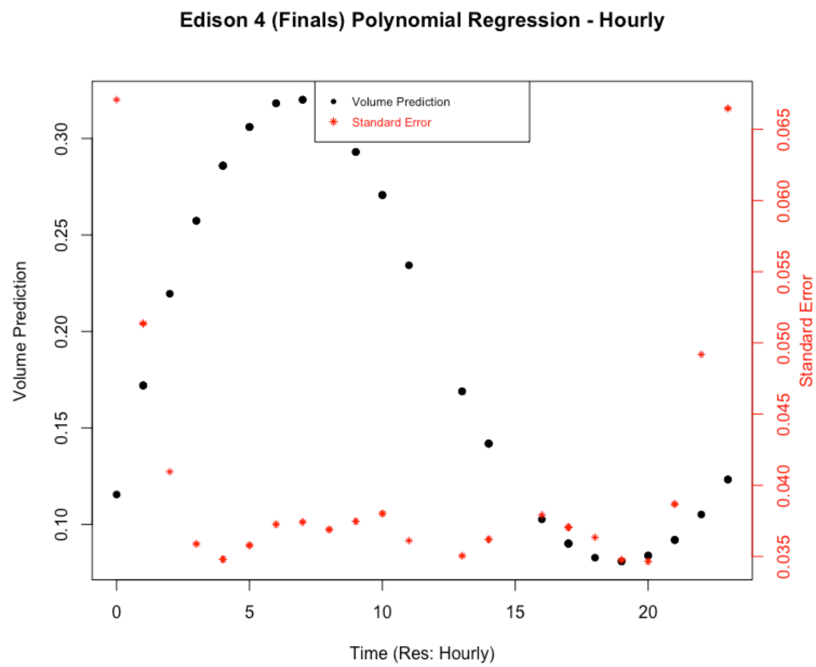

Figure 46: This figure shows the local polynomial regression fitting and standard error for Edison 4 during finals week using an hourly resolution.

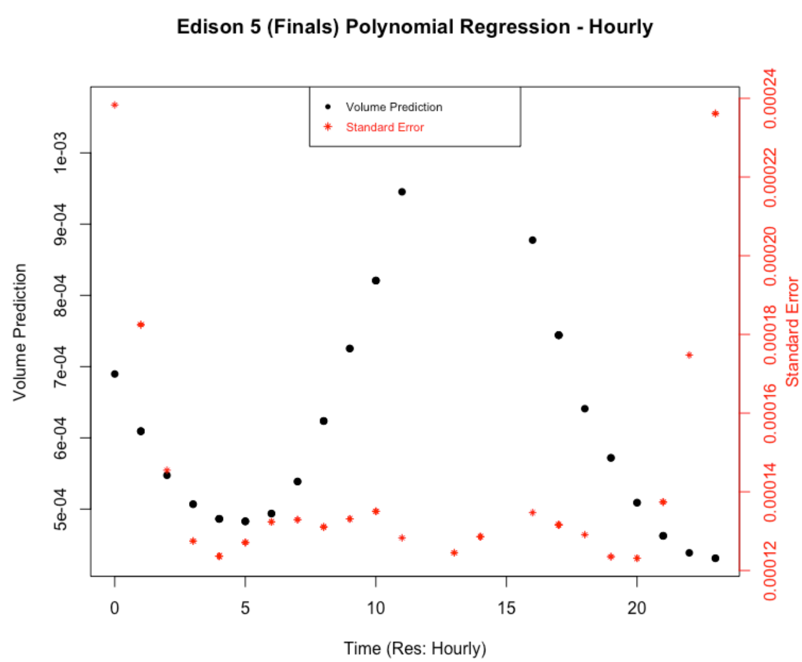

Figure 47: This figure shows the local polynomial regression fitting and standard error for Edison 5 during finals week using an hourly resolution.

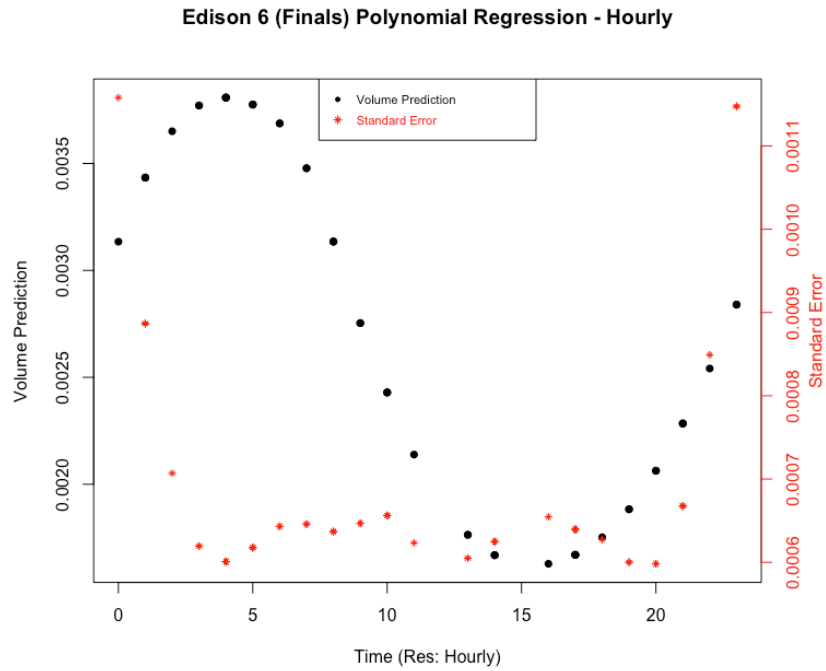

Figure 48: This figure shows the local polynomial regression fitting and standard error for Edison 6 during finals week using an hourly resolution.

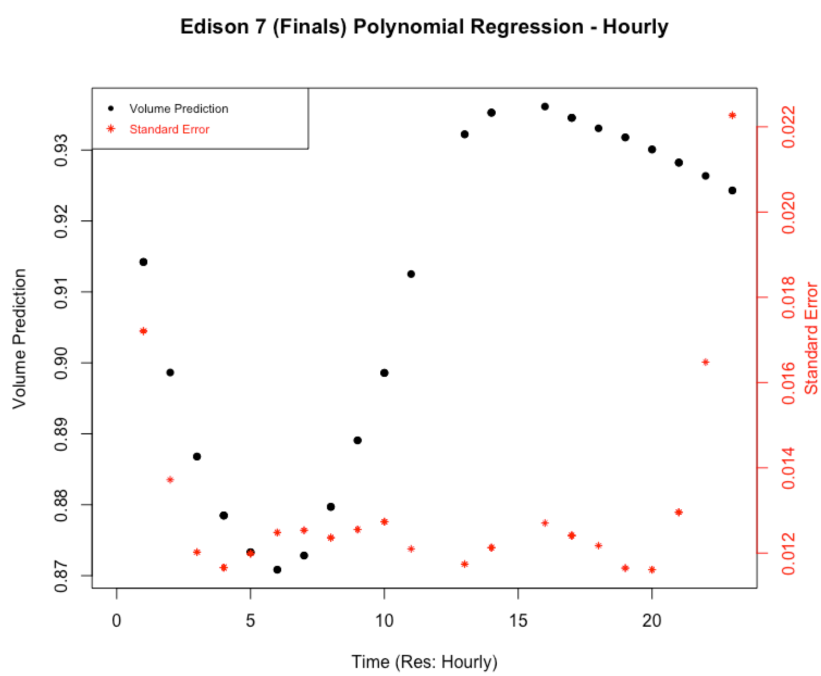

Figure 49: This figure shows the local polynomial regression fitting and standard error for Edison 7 during finals week using an hourly resolution.

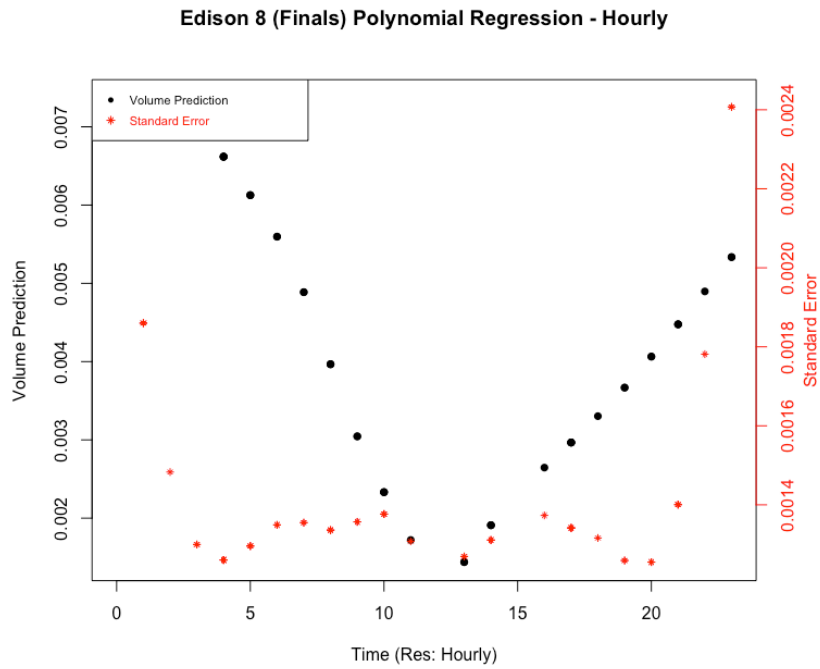

Figure 50: This figure shows the local polynomial regression fitting and standard error for Edison 8 during finals week using an hourly resolution.

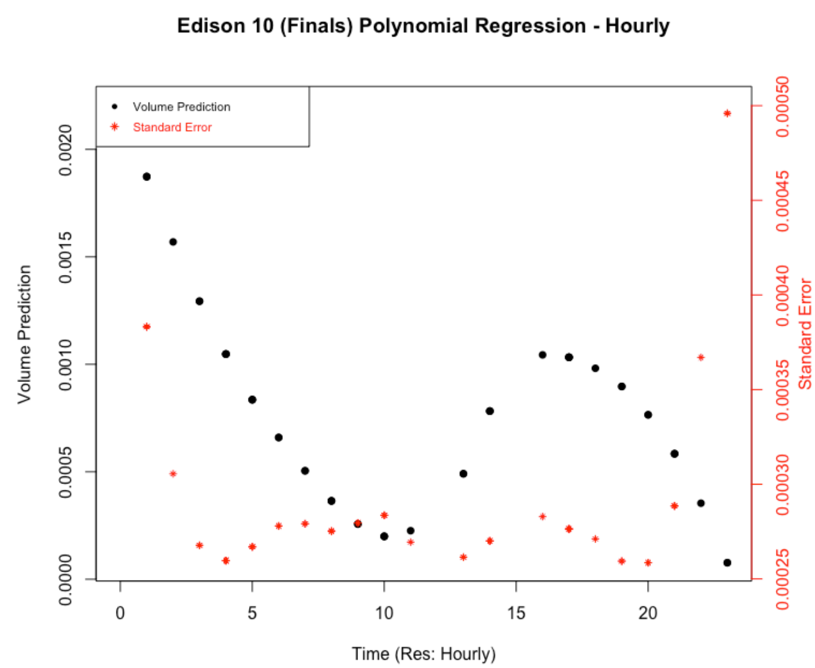

Figure 51: This figure shows the local polynomial regression fitting and standard error for Edison 10 during finals week using an hourly resolution.

Spring Break

Figures 52 – 59: Local Polynomial Regressions Hourly – Spring Break

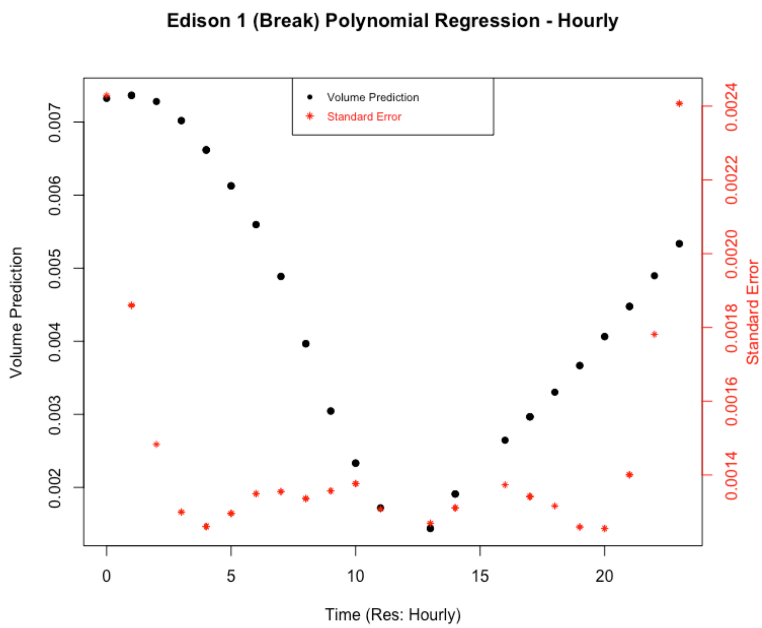

Figure 52: This figure shows the local polynomial regression fitting and standard error for Edison 1 during spring break using an hourly resolution.

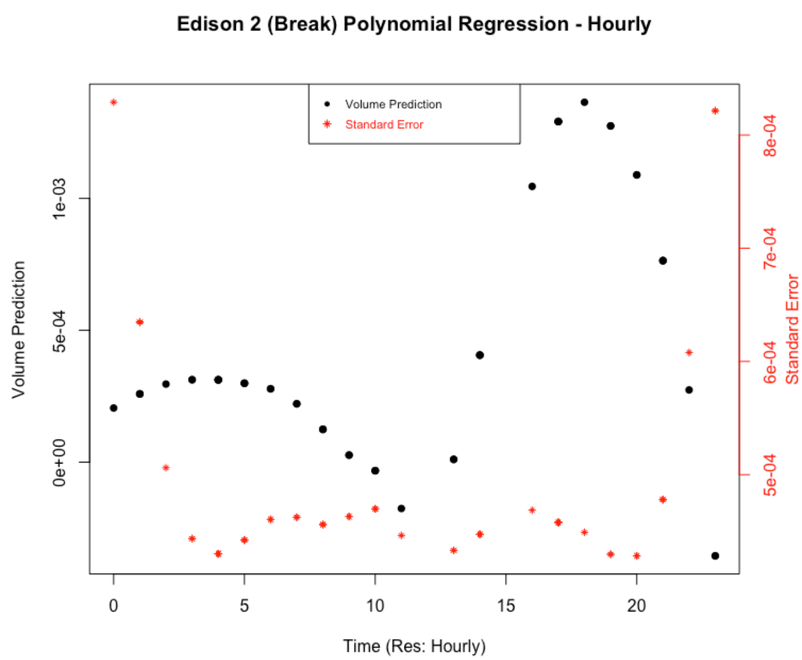

Figure 53: This figure shows the local polynomial regression fitting and standard error for Edison 2 during spring break using an hourly resolution.

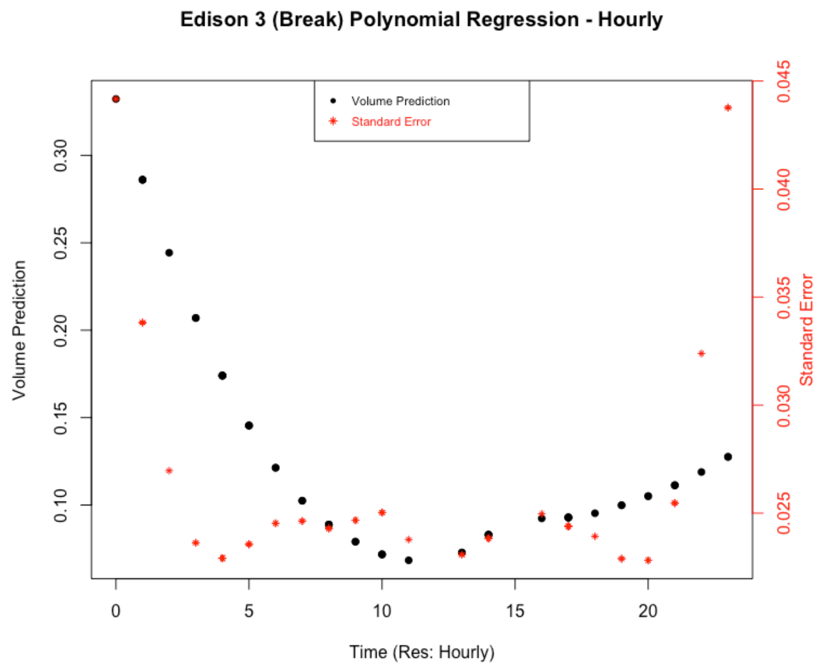

Figure 54: This figure shows the local polynomial regression fitting and standard error for Edison 3 during spring break using an hourly resolution.

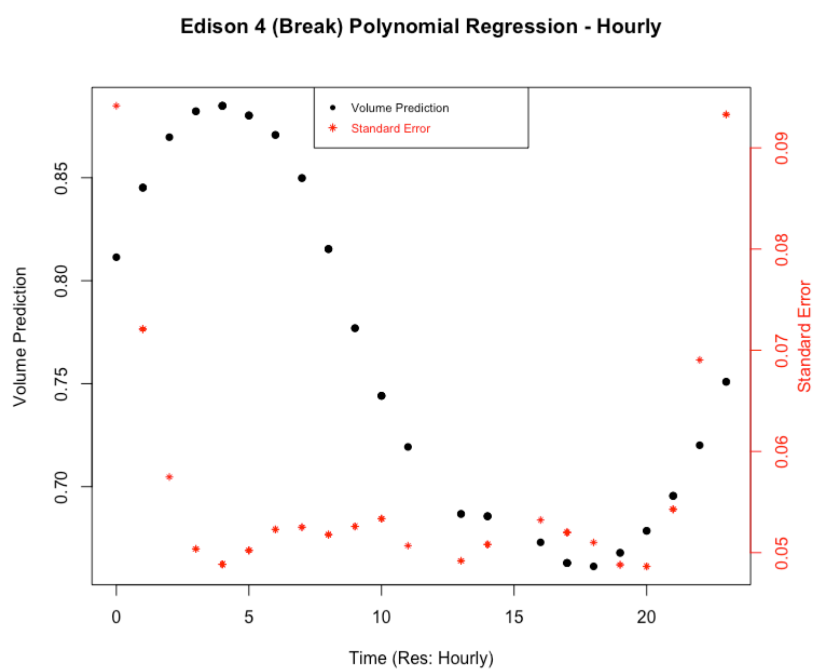

Figure 55: This figure shows the local polynomial regression fitting and standard error for Edison 4 during spring break using an hourly resolution.

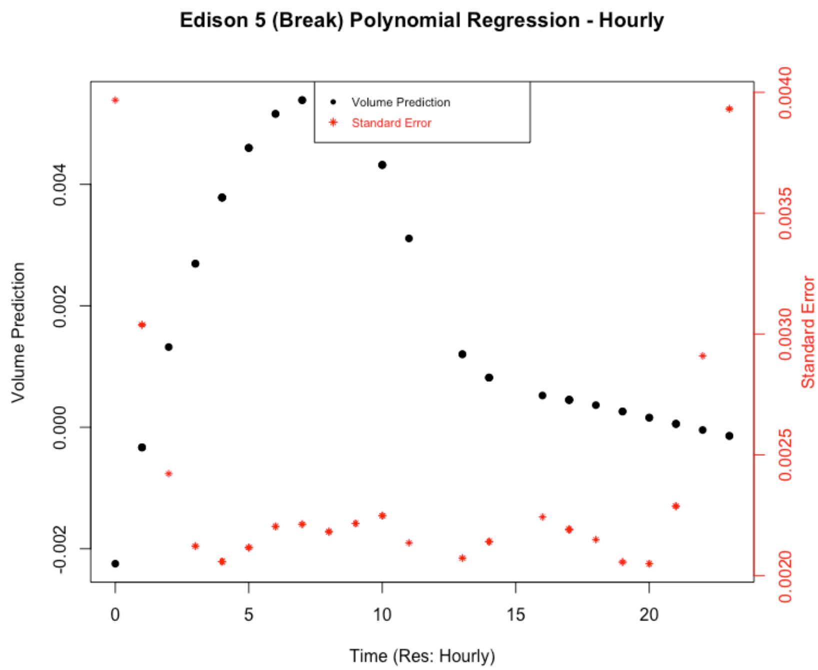

Figure 56: This figure shows the local polynomial regression fitting and standard error for Edison 5 during spring break using an hourly resolution.

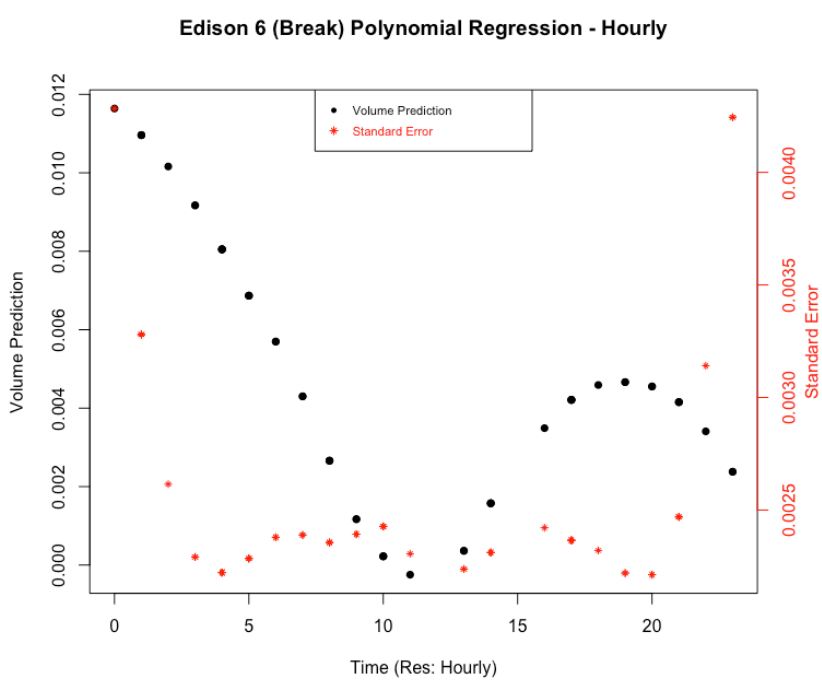

Figure 57: This figure shows the local polynomial regression fitting and standard error for Edison 6 during spring break using an hourly resolution.

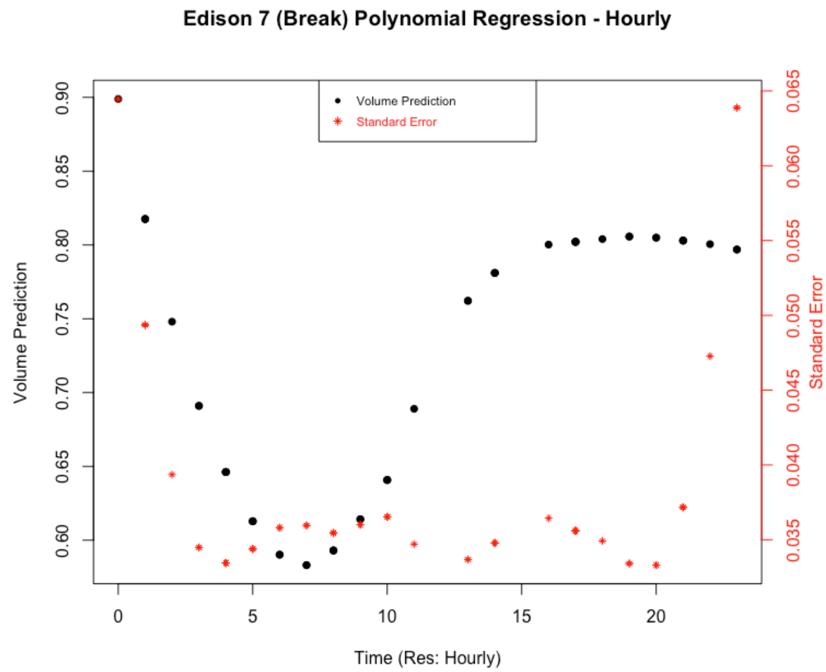

Figure 58: This figure shows the local polynomial regression fitting and standard error for Edison 7 during spring break using an hourly resolution.

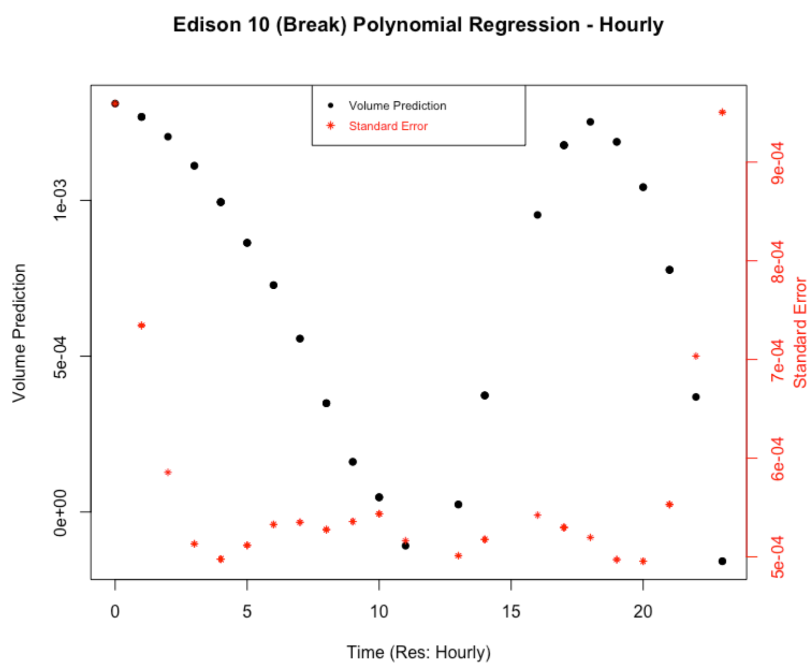

Figure 59: This figure shows the local polynomial regression fitting and standard error for Edison 10 during spring break using an hourly resolution.

Week 1

Figures 60 – 66: Local Polynomial Regressions Hourly – Week 1

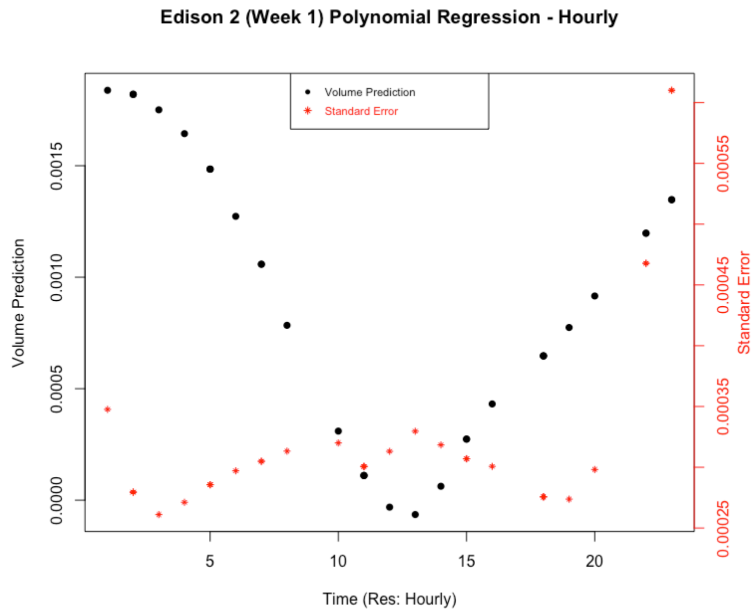

Figure 60: This figure shows the local polynomial regression fitting and standard error for Edison 1 during week 1 using an hourly resolution.

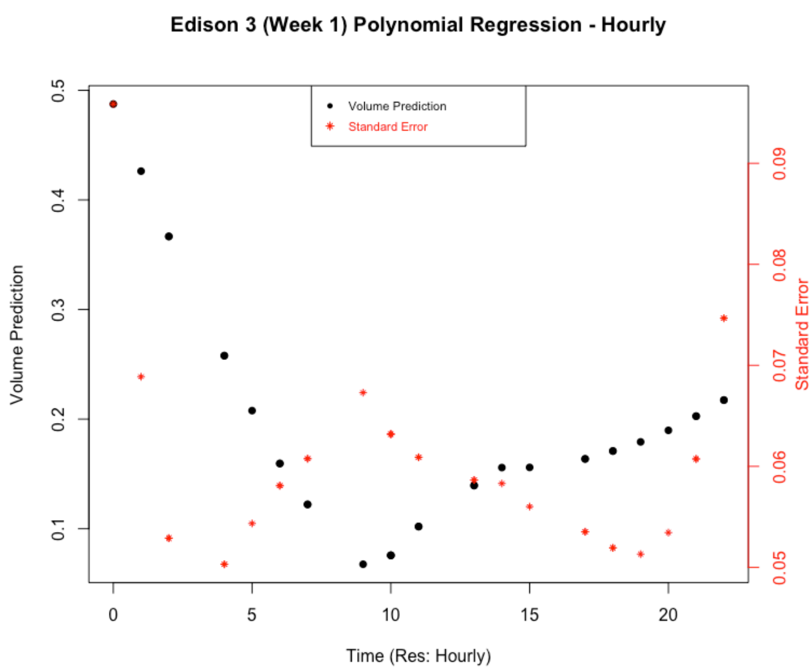

Figure 61: This figure shows the local polynomial regression fitting and standard error for Edison 3 during week 1 using an hourly resolution.

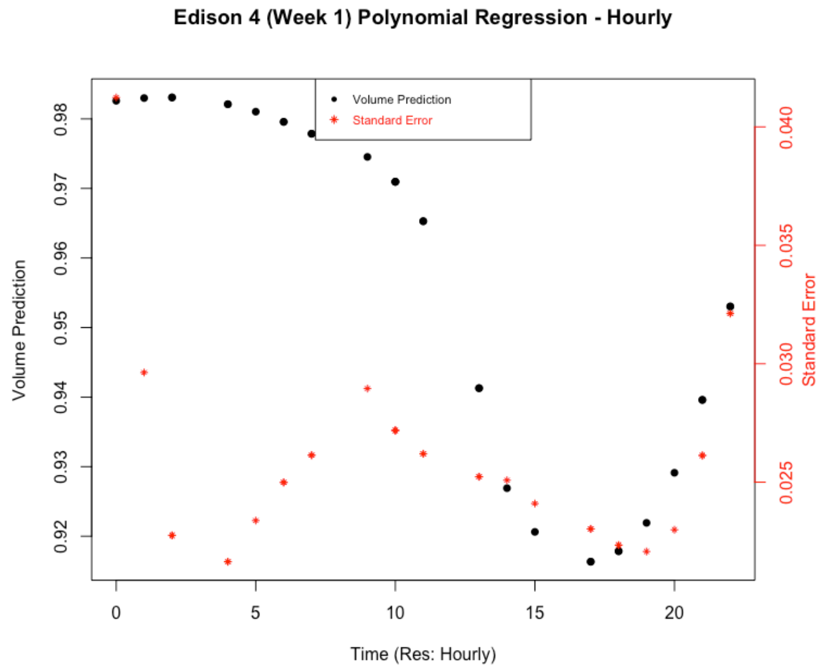

Figure 62: This figure shows the local polynomial regression fitting and standard error for Edison 4 during week 1 using an hourly resolution.

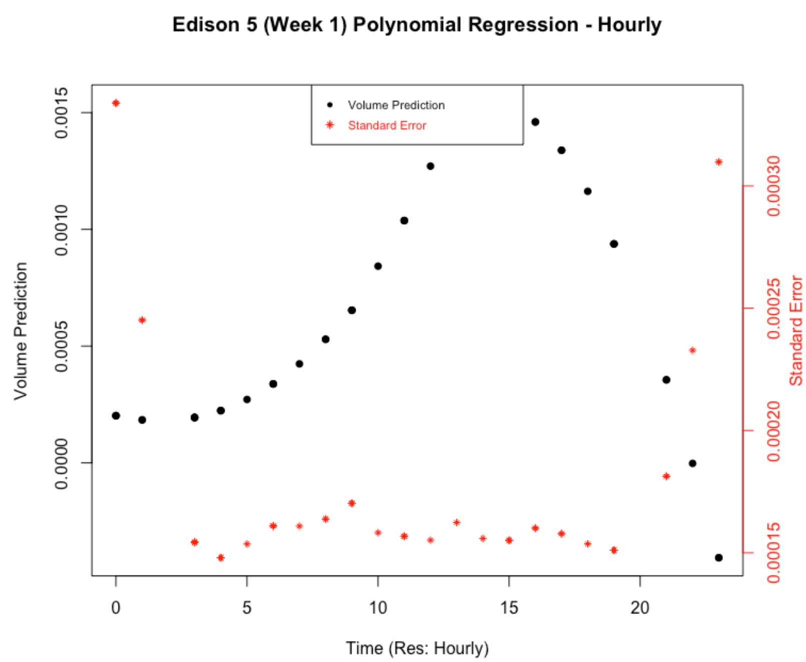

Figure 63: This figure shows the local polynomial regression fitting and standard error for Edison 5 during week 1 using an hourly resolution.

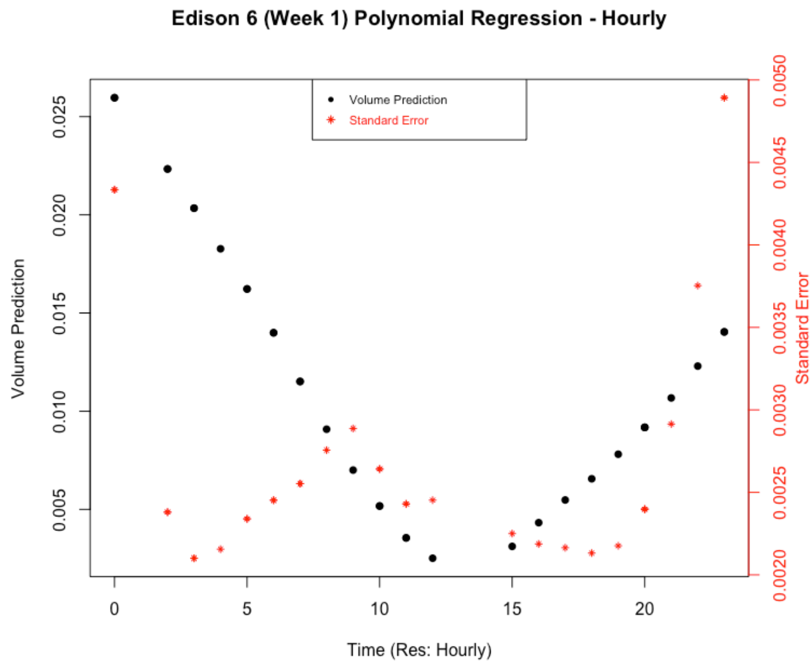

Figure 64: This figure shows the local polynomial regression fitting and standard error for Edison 6 during week 1 using an hourly resolution.

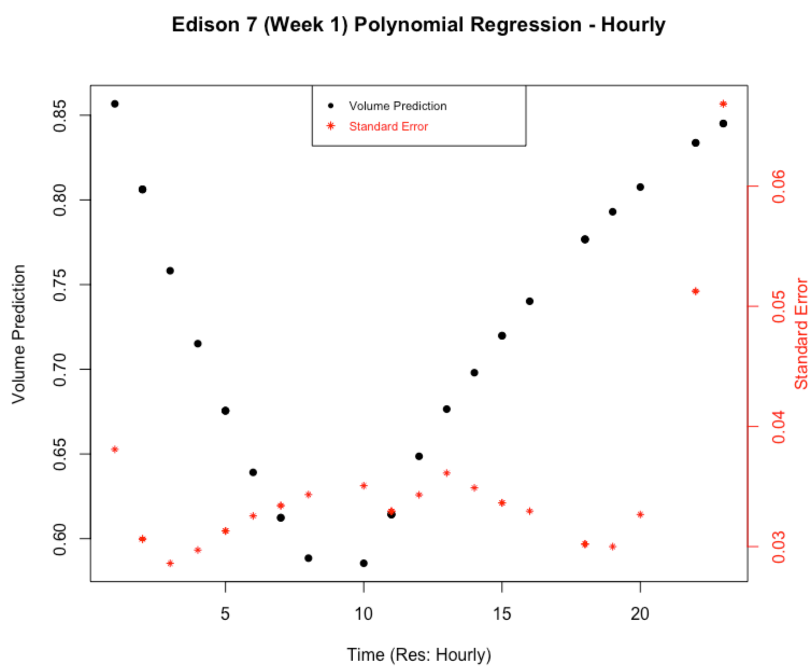

Figure 65: This figure shows the local polynomial regression fitting and standard error for Edison 7 during week 1 using an hourly resolution.

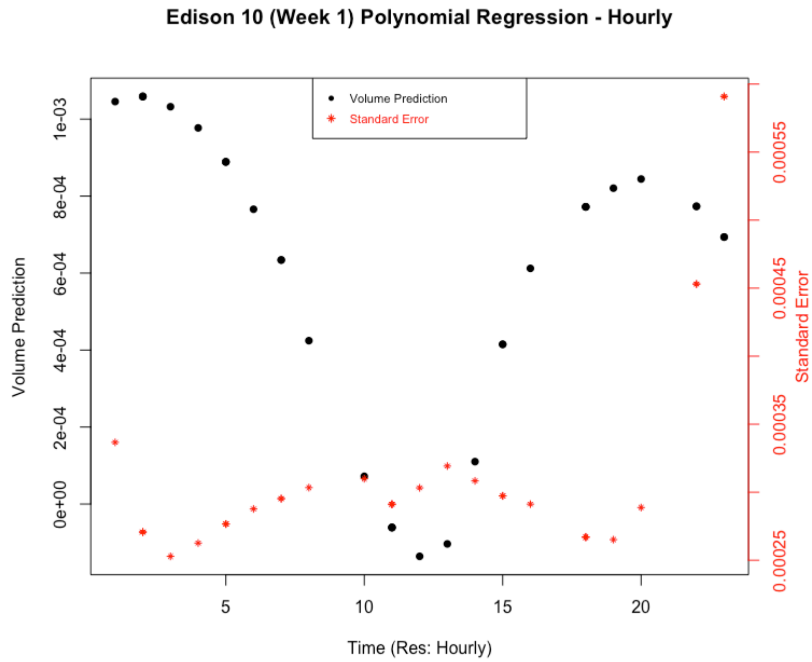

Figure 66: This figure shows the local polynomial regression fitting and standard error for Edison 10 during week 1 using an hourly resolution.

All Weeks

Figures 67 – 75: Local Polynomial Regressions Hourly – All Weeks Combined

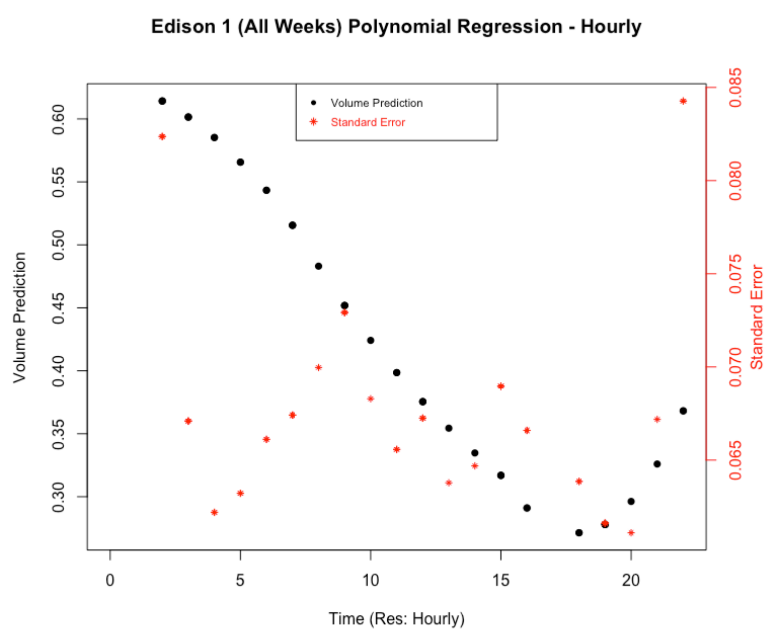

Figure 67: This figure shows the local polynomial regression fitting and standard error for Edison 1 across all weeks using an hourly resolution.

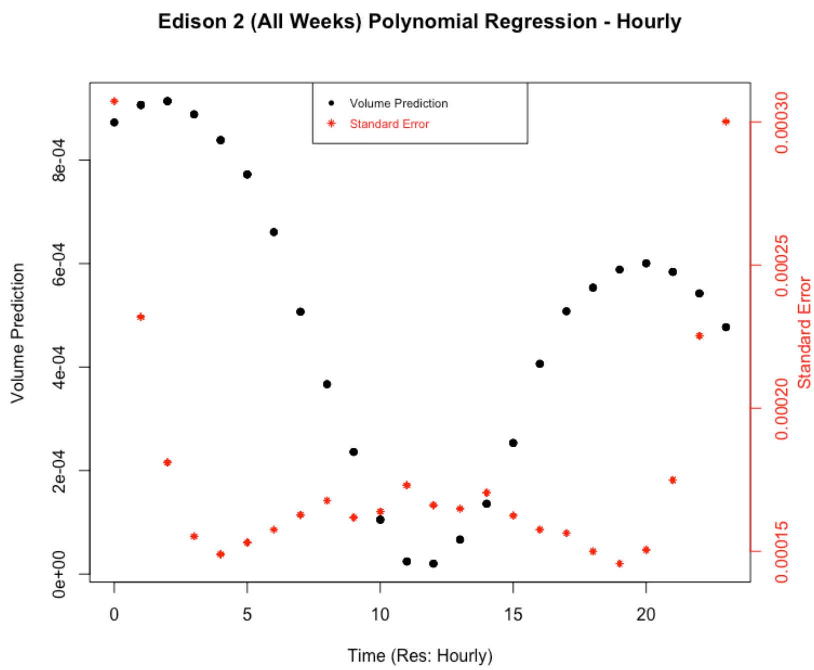

Figure 68: This figure shows the local polynomial regression fitting and standard error for Edison 2 across all weeks using an hourly resolution.

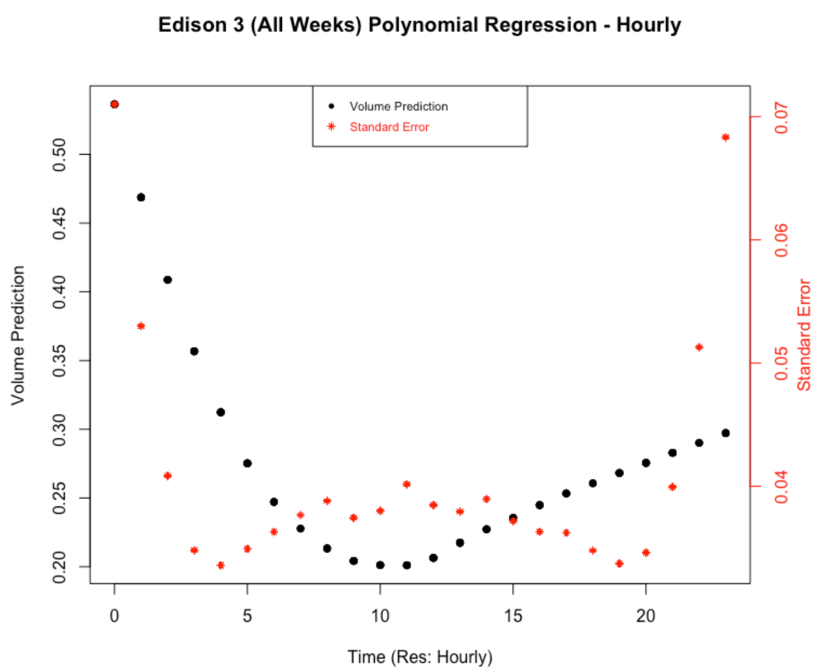

Figure 69: This figure shows the local polynomial regression fitting and standard error for Edison 3 across all weeks using an hourly resolution.

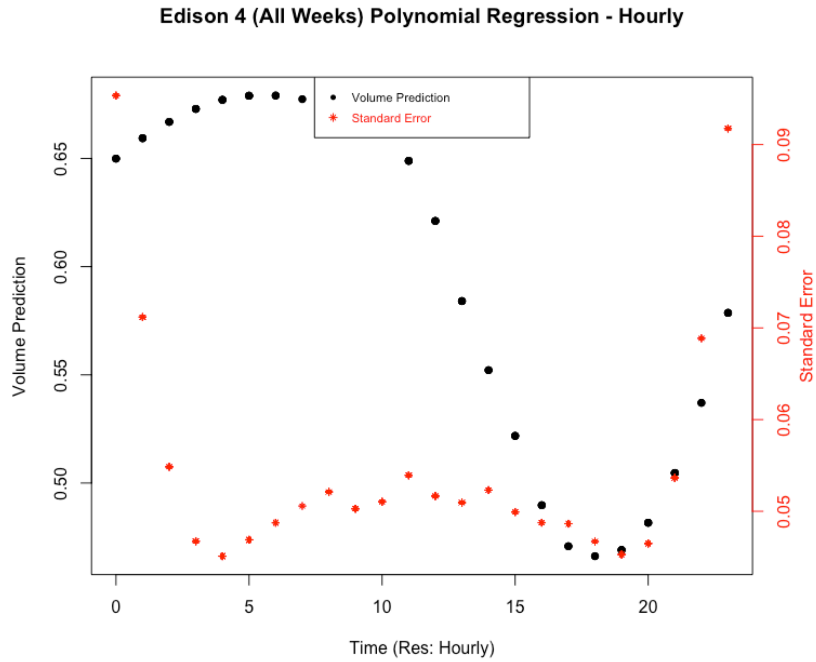

Figure 70: This figure shows the local polynomial regression fitting and standard error for Edison 4 across all weeks using an hourly resolution.

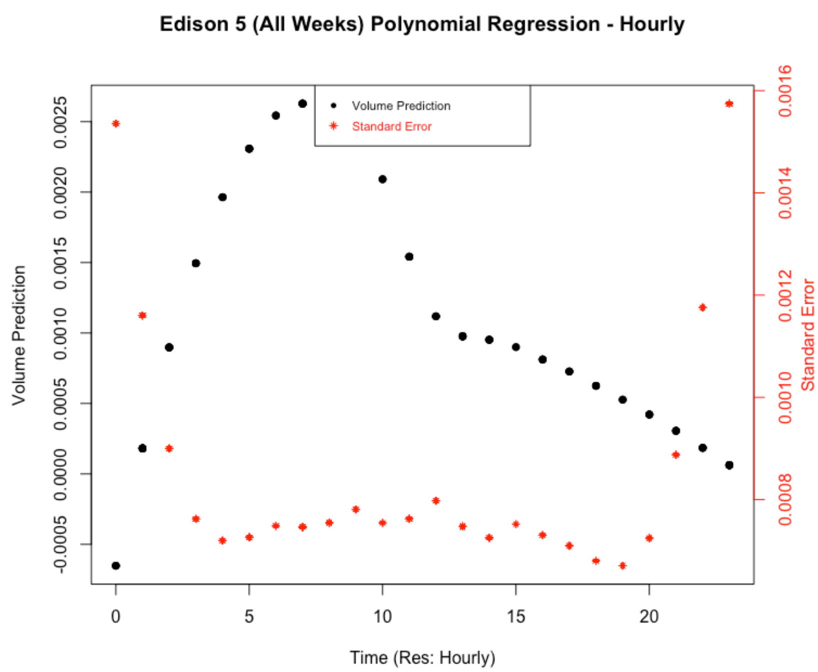

Figure 71: This figure shows the local polynomial regression fitting and standard error for Edison 5 across all weeks using an hourly resolution.

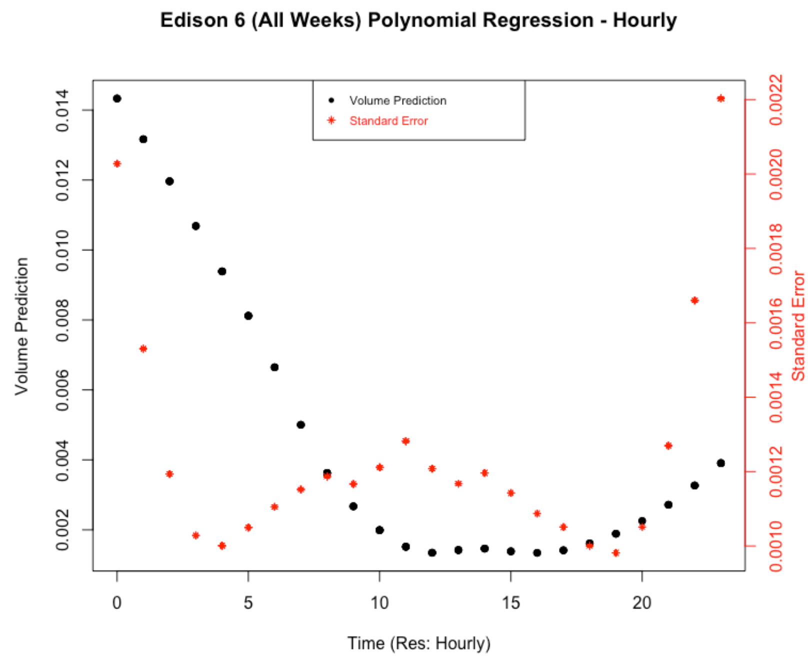

Figure 72: This figure shows the local polynomial regression fitting and standard error for Edison 6 across all weeks using an hourly resolution.

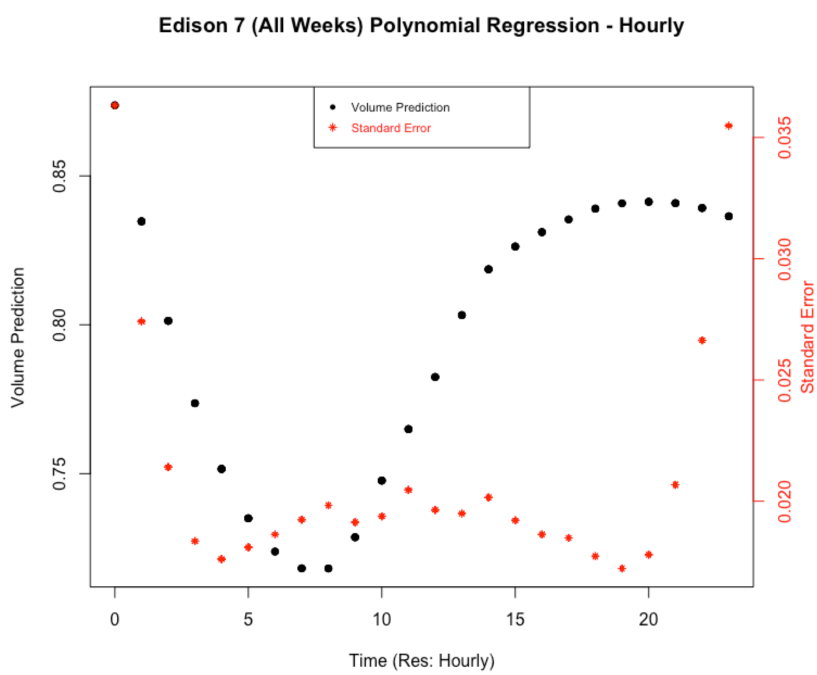

Figure 73: This figure shows the local polynomial regression fitting and standard error for Edison 7 across all weeks using an hourly resolution.

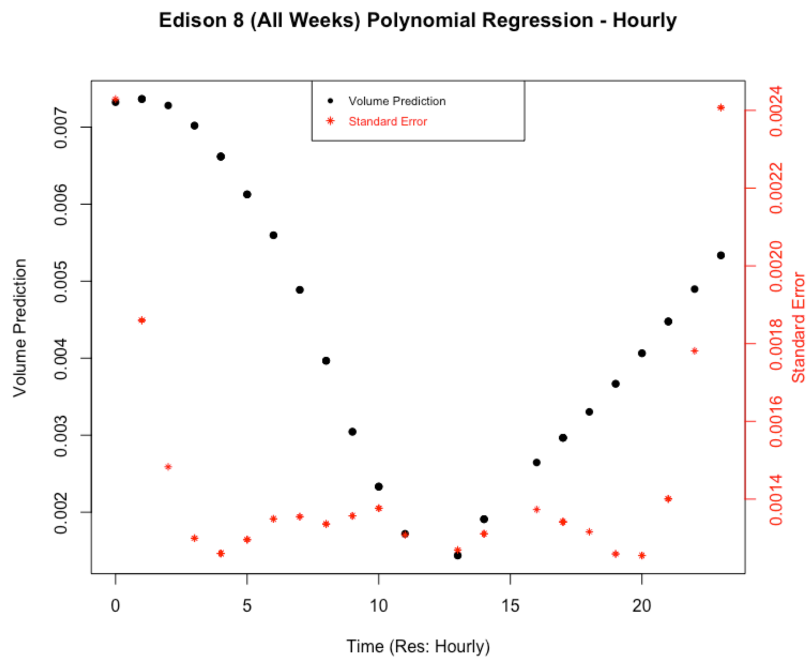

Figure 74: This figure shows the local polynomial regression fitting and standard error for Edison 8 across all weeks using an hourly resolution.

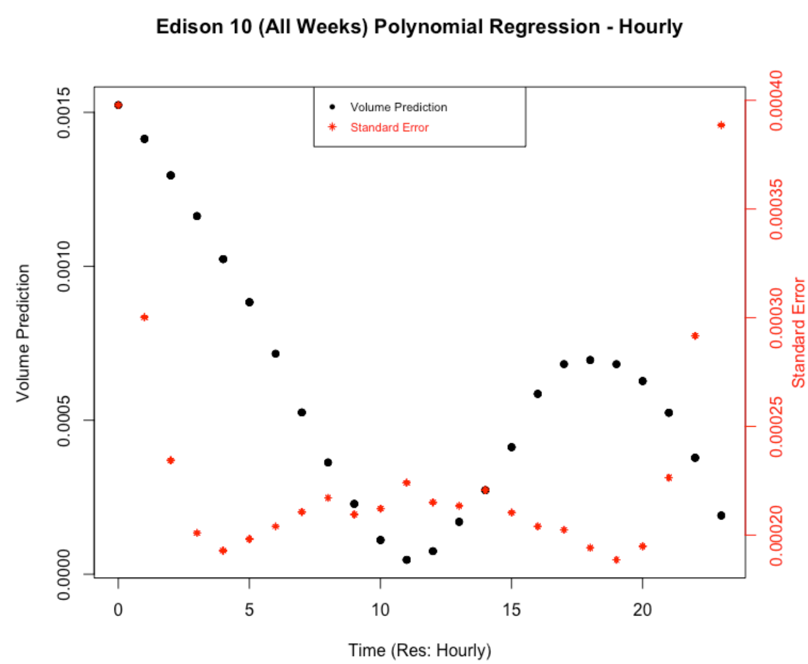

Figure 75: This figure shows the local polynomial regression fitting and standard error for Edison 10 across all weeks using an hourly resolution.

Bibliography

Bahga, Arshdeep, and V. Madisetti. *Internet of Things: A Hands-on Approach*. N.p.: n.p., 2014. Print.

Dass, Gaurhari. "Regression Analysis: How Do I Interpret R-squared and Assess the Goodness-of-Fit?" LinkedIn, 22 Apr. 2015. Web. <https://www.linkedin.com/pulse/regression-analysis-how-do-i-interpret-r-squared-assess-gaurhari-dass>.

"Database." *Merriam-Webster*. Merriam-Webster, n.d. Web. 05 May 2017. <https://www.merriam-webster.com/dictionary/database?utm_campaign=sd&utm_medium=serp&utm_source=jsonld>.

Enda Murphy and Eoin A. King, Chapter 7 - Noise Mitigation Approaches, In Environmental Noise Pollution, Elsevier, Boston, 2014, Pages 203-245, ISBN 9780124115958, http://dx.doi.org/10.1016/B978-0-12-411595-8.00007-0. (http://www.sciencedirect.com/science/article/pii/B9780124115958000070)

Gates, Dirk. "Build WiFi To Weather The IoT Onslaught." *Network Computing*. Network Computing, 28 May 2015. Web. 05 May 2017. <http://www.networkcomputing.com/wireless/build-wifi-weather-iot-onslaught/1680833857>.

"Glossary of Z/OS Terms and Abbreviations." *IBM Knowledge Center*. IBM, n.d. Web. 05 May 2017. <https://www.ibm.com/support/knowledgecenter/zosbasics/com.ibm.zglossary.doc/zglossary.html>.

"GSA & Smart Buildings." U.S. General Services Administration, n.d. Web. <http://www.gsa.gov/portal/content/193249>.

H. Hossam Eldien, "Noise mapping in urban environments: Application at Suez city center," *Computers & Industrial Engineering, 2009. CIE 2009. International Conference on,* Troyes, 2009, pp. 1722-1727. URL: http://ieeexplore.ieee.org.libproxy.uoregon.edu/stamp/stamp.jsp?tp=&arnumber=5223696&isnumber=5223483

Hammer, Monica, Tracy Swinburn, and Richard Neitzel. "EHP – Environmental Noise Pollution in the United States: Developing an Effective Public Health Response." Environmental Health Perspectives, Feb. 2014. Web. <http://ehp.niehs.nih.gov/1307272/>.

"How It Works | Bluetooth Technology Website." *Bluetooth*. Bluetooth Special Interest Group, n.d. Web. 05 May 2017. <https://www.bluetooth.com/what-is-bluetooth-technology/how-it-works>.

J. Jin, J. Gubbi, S. Marusic and M. Palaniswami, "An Information Framework for Creating a Smart City Through Internet of Things," in *IEEE Internet of Things Journal*, vol. 1, no. 2, pp. 112-121, April 2014. URL: http://ieeexplore.ieee.org.libproxy.uoregon.edu/stamp/stamp.jsp?tp=&arnumber=6702523&isnumber=6811169

J. Pan, R. Jain, S. Paul, T. Vu, A. Saifullah and M. Sha, "An Internet of Things Framework for Smart Energy in Buildings: Designs, Prototype, and Experiments," in *IEEE Internet of Things Journal*, vol. 2, no. 6, pp. 527-537, Dec. 2015. URL: http://ieeexplore.ieee.org.libproxy.uoregon.edu/stamp/stamp.jsp?tp=&arnumber=7061425&isnumber=7331244

"Linear Regression." *Linear Regression*. Yale University, n.d. Web. 05 May 2017. <http://www.stat.yale.edu/Courses/1997-98/101/linreg.htm>.

"Local Polynomial Regression Fitting." *R: Local Polynomial Regression Fitting*. The R Manual, n.d. Web. 09 May 2017. <https://stat.ethz.ch/R-manual/R-devel/library/stats/html/loess.html>.

Marechal, Sander. "A Simple Unix/linux Daemon in Python." *Lone Wolves - Web, Game, and Open Source Development*. N.p., 2 Feb. 2007. Web. 05 May 2017. <http://web.archive.org/web/20131017130434/http://www.jejik.com/articles/2007/02/a_simple_unix_linux_daemon_in_python/>.

Nordrum, Amy. "Popular Internet of Things Forecast of 50 Billion Devices by 2020 Is Outdated." *IEEE Spectrum: Technology, Engineering, and Science News*. IEEE Spectrum, 18 Aug. 2016. Web. 05 May 2017. <http://spectrum.ieee.org/tech-talk/telecom/internet/popular-internet-of-things-forecast-of-50-billion-devices-by-2020-is-outdated>.

O. Evangelatos, K. Samarasinghe and J. Rolim, "Syndesi: A Framework for Creating Personalized Smart Environments Using Wireless Sensor Networks," *2013 IEEE International Conference on Distributed Computing in Sensor Systems*, Cambridge, MA, 2013, pp. 325-330. URL: http://ieeexplore.ieee.org.libproxy.uoregon.edu/stamp/stamp.jsp?tp=&arnumber=6569451&isnumber=6569394

P. Tao, Y. Chen and L. Tong, "GIS-based city noise mapping research and development," *2012 IEEE International Geoscience and Remote Sensing Symposium*, Munich, 2012, pp. 6729-6732. URL: http://ieeexplore.ieee.org.libproxy.uoregon.edu/stamp/stamp.jsp?tp=&arnumber=6352340&isnumber=6350328

Piech, Chris, and Andrew Ng. "K Means." *Stanford CS221*. Stanford University, n.d. Web. 05 May 2017. <http://stanford.edu/~cpiech/cs221/handouts/kmeans.html>.

"What Is Big Data and Why It Matters." *SAS US*. SAS, n.d. Web. 05 May 2017. <https://www.sas.com/en_us/insights/big-data/what-is-big-data.html>.