ORNL

Impact of Urban Weather on Energy Use

SMC Data Challenge 2018

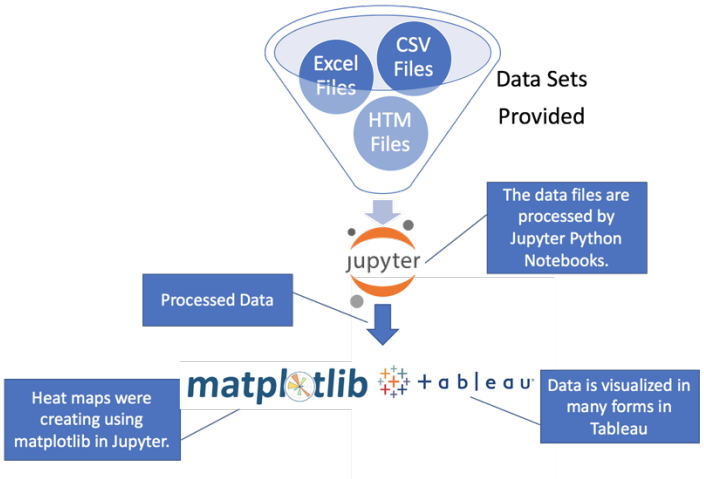
Supriya Chinthavali, Ethan Hicks, Philip Hicks

Introduction: Recent advances in multi-scale integration of models have started to provide unique insights into how interdependent processes influence one another. The effect of these processes is uniquely observable in urban environments. This report provides a brief description of our solutions to the three sub-challenges related to understanding the impact of urban weather on energy use where we apply various machine learning, statistical and visualization techniques including regression, clustering correlation etc.

Solution Description:

Data Wrangling: Firstly, we input all the excel files provided for each building and combined them into a single file using custom python scripts. Also, the data in those files were provided in hour intervals which we aggregated into daily averages and medians to filter and enable efficient and rapid analysis of the dataset. Another python module was developed to extract each building's volume into another file.

Data Analysis and Visualization: After the preliminary processing of the raw datasets, the generated output files were analyzed using several machine learning and statistical techniques such as clustering, correlation(e.g spearman coefficient) etc. Finally, we visualized the results in interesting ways to gain insights to help explain our hypothesis.



1. Are there interesting variations in the weather and building energy use data for the geographic area: To measure the effects of daily wet bulb/dry bulb temperature, humidity on electricity and gas usage, the median temperatures were calculated for each day and correlated with the daily gas and electricity use by computing the Spearman's rank correlation coefficient and Pearson correlation coefficient. While Pearson correlation coefficients measure only linear relationships, Spearman correlation coefficients measure monotonic relationships. The results obtained demonstrate that the buildings can be clustered into three different groups. The first two groups can be referred to as Gas Heaters (Figure 1) and Electric Heaters (Figure 2). Gas Heaters have a strong negative correlation between the gas use and outside temperatures. The electricity use shows a weak positive correlation with the outside temperatures. On the other hand, Electric Heaters have a high negative correlation between the electricity usage and temperatures and a weak negative correlation between gas use and temperatures. The high negative correlation shows that as temperatures drop, power or gas usage increases which implies that they are heating the buildings instead of cooling it. This correlates with Chicago's lower temperatures.

Figure 1 Example Gas Heater

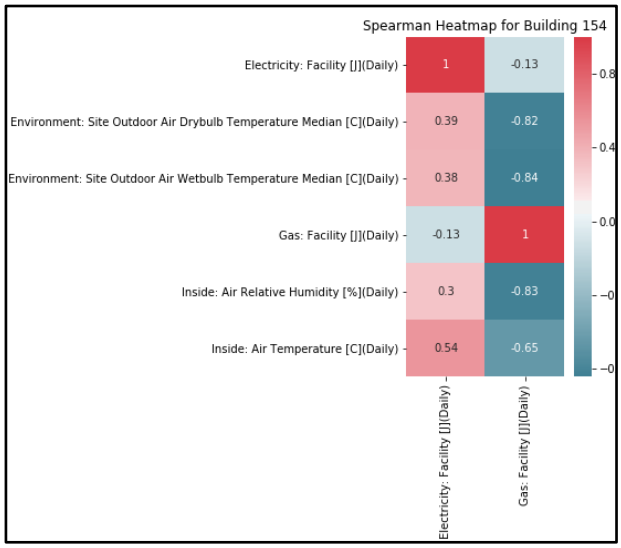
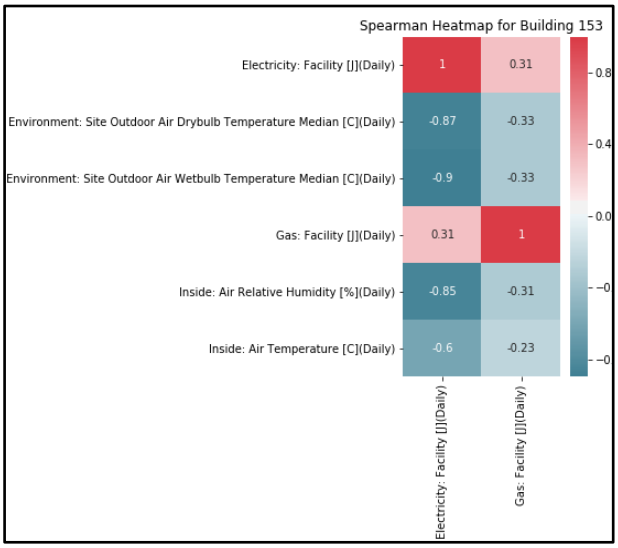


Figure 2 Example Electric Heater



The last classification is Gas Electric Hybrid (Figure 4). These building have medium to strong negative correlation between gas usage and temperature, and the electric to temperature correlation is low to strong negative correlations. What differentiates Hybrids from the Gas Heaters with low electric correlations is the fact that their correlation is negative, instead of positive, and the fact that, like all other Hybrids, they have a medium to high correlation between gas and electricity. This correlation likely means that their heating systems are some sort of combined systems that uses both electricity and gas to heat the buildings.

Figure 3 Example Gas Electric Hybrid

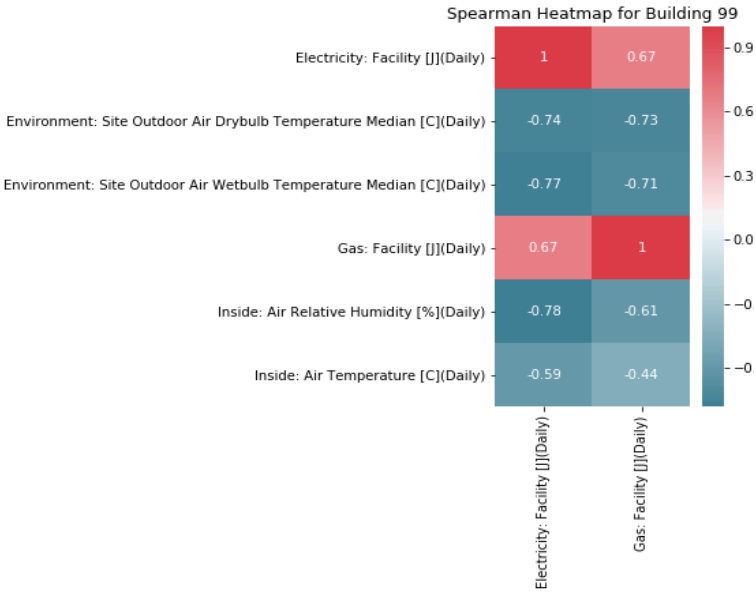


Figure 4 Visual Classification of Buildings

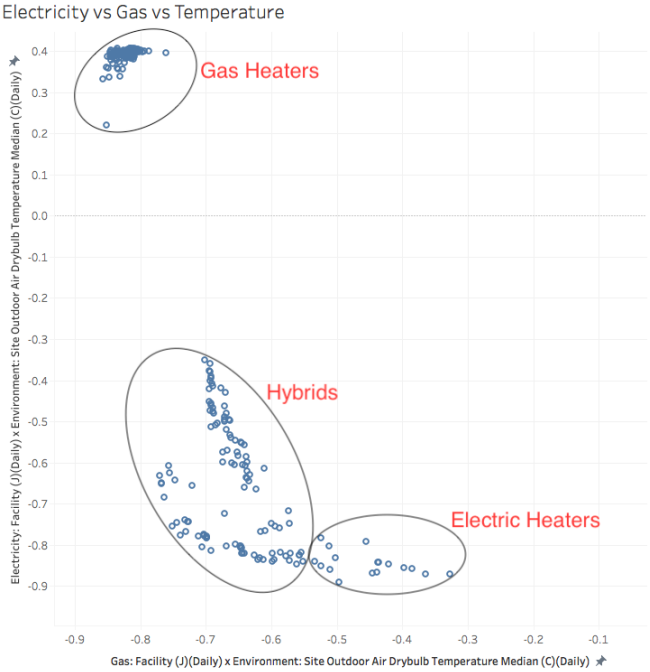


Figure 4 is a visual representation of how many buildings belong to each category and how closely a category relates to each other. Gas Heaters make up the majority with 186 buildings that are very tightly packed together. Hybrids are the second biggest category with 115 buildings and cover the largest range. The smallest category is the Electric Heaters with only 17 buildings.

2. Which buildings in the study have their energy use impacted the most by external factors including the weather? In this task, we first applied k-means clustering to identify different types of buildings based on their energy use data. K-means clustering can be described as an optimization problem which can be posed as: "find the *k* cluster centers and assign the objects to the nearest cluster center, such that the squared distances from the cluster are minimized." Then, for each of these clusters to tried to draw relationships between electricity use and the outside temperatures (Wet and Drybulb) using linear regression. When looking at all the clusters together it is hard to draw relationships between Electricity use and the outside temperatures (Wet and Drybulb). However, when looking at the Daily Gas use we can clearly see for all clusters that as outside temperature increases the use of Gas decreases. For cluster one outside temperature, majority of the buildings electric energy use increases but still a large portion decreases or remains relatively

unaffected by the outside temperature. Cluster two is similar to cluster one but in the opposite way because although a majority show some form of decrease in energy use as temperature increases.

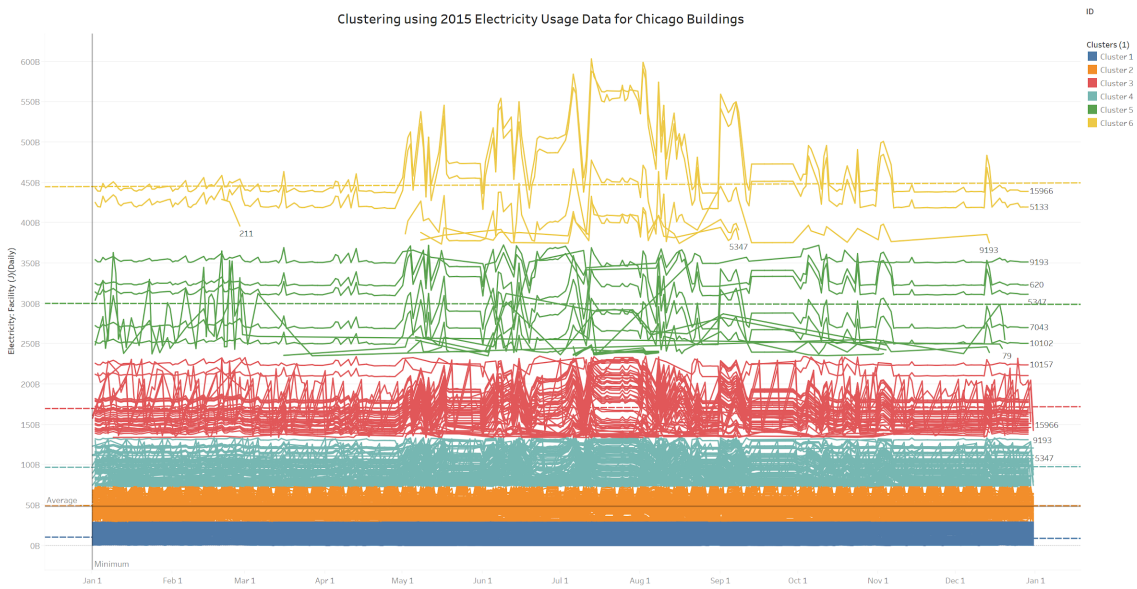


Figure 5 : k-means clustering applied to all the building using 2015 Electricity consumption data.

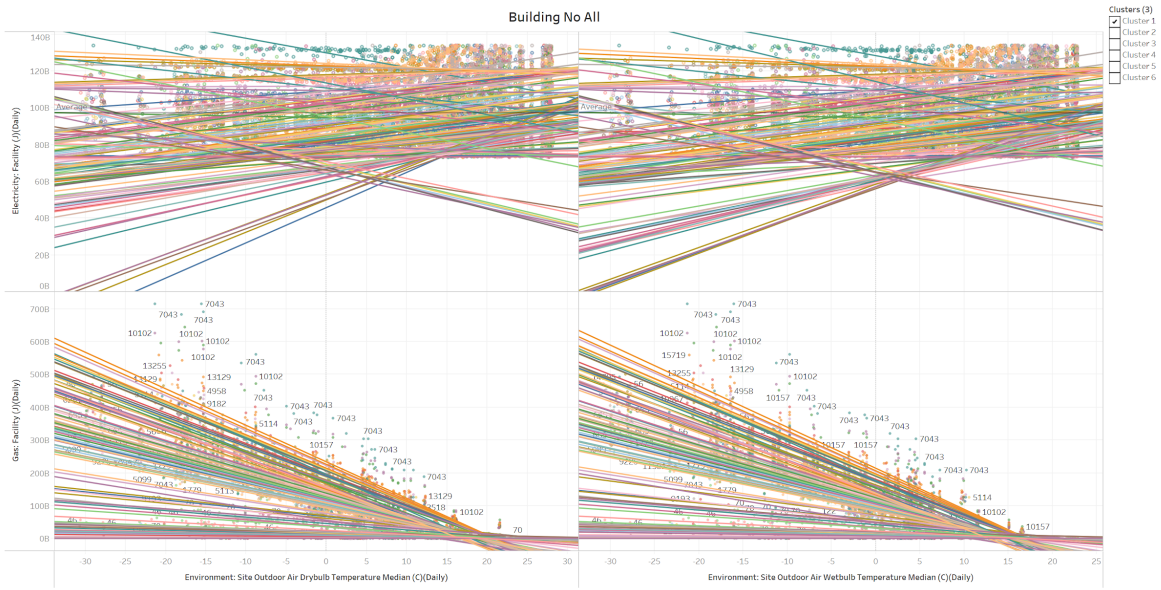


Figure 6 : Linear Regression applied to every building to understand Electricity and Gas use against Dry Bulb and Wet Bulb temperature attributes.

There is still a portion of them that show an increase in energy use as temperature rises. Cluster 3 displays a clear trend where when the outside temperatures increase, buildings consume more energy per day except for a few outliers. No conclusions could be drawn for cluster 4. Cluster 5 only consists of the 6 biggest buildings and we can clearly see the trend that as the outside temperature increases, they use more electricity except for one outlying building. For Cluster 6, we couldn't draw any conclusions from the Dry Bulb temperature data but when we look at the Wet Bulb temperature data, we can notice that these buildings that were showing decreasing trends with dry bulb temperatures is increasing with wet bulb temperatures indicating that humidity also has some impact on energy use. From this we can conclude the humidity had very little effect on determining a buildings energy use since they seemed to be affecting only one of the six clusters of buildings.

3. Are there any interesting visualizations that illustrate the changing dynamics of the simulated urban environment? There is very strong correlation between a building's volume and its daily energy use of electricity. For all large buildings, any building besides a consumer's home, a building's volume almost directly correlates a building's average energy use. The more volume a building has the more electricity it consumers. As with most data sets there are a few outliers as can be seen in **figure 7** and a majority of these are buildings that do not use nearly as much gas as the other buildings as seen in **figure 8.** When we plot the energy use vs the inside air temperature we see some very clear patterns of ups and downs between the two(**F**). During the cooler months, the inside temperatures of the buildings follow a pattern where the inside temperatures stay almost constant (around 20°C) but then over the weekends they let up on their electric and gas use and let the building temperatures fall a few degrees (around 15°C). During the summer the inside temperatures do not follow a clear pattern throughout the week. Also, during the summer very few buildings use gas. We can then draw the conclusion that the buildings in Chicago use electric for both heating and cooling (rarely when it is warm enough) whereas most buildings only use gas for heating their buildings in the cooler months of each year. Looking at the coldest day of the year we can see that it affects multiple buildings by lowering their inside temperatures and causing them to use slightly more energy that day to help offset the drop in temperature (**Figure 11**).

Challenges with the dataset : We were unable to use the high resolution, 90-meter simulated weather data due to several reasons. One reason was we had trouble reading the data with a program where we could perform analytics with it. We used several programs like, MATLAB, Python, etc. but where not able to build any useful visualizations with it. Also, another reason for not using the high resolution data was because it only covered 19 days of weather data whereas the other data sources provided were for the whole year of 2015. It would be hard to make conclusions with data that only spanned 19 days with our other data.

APPENDIX A

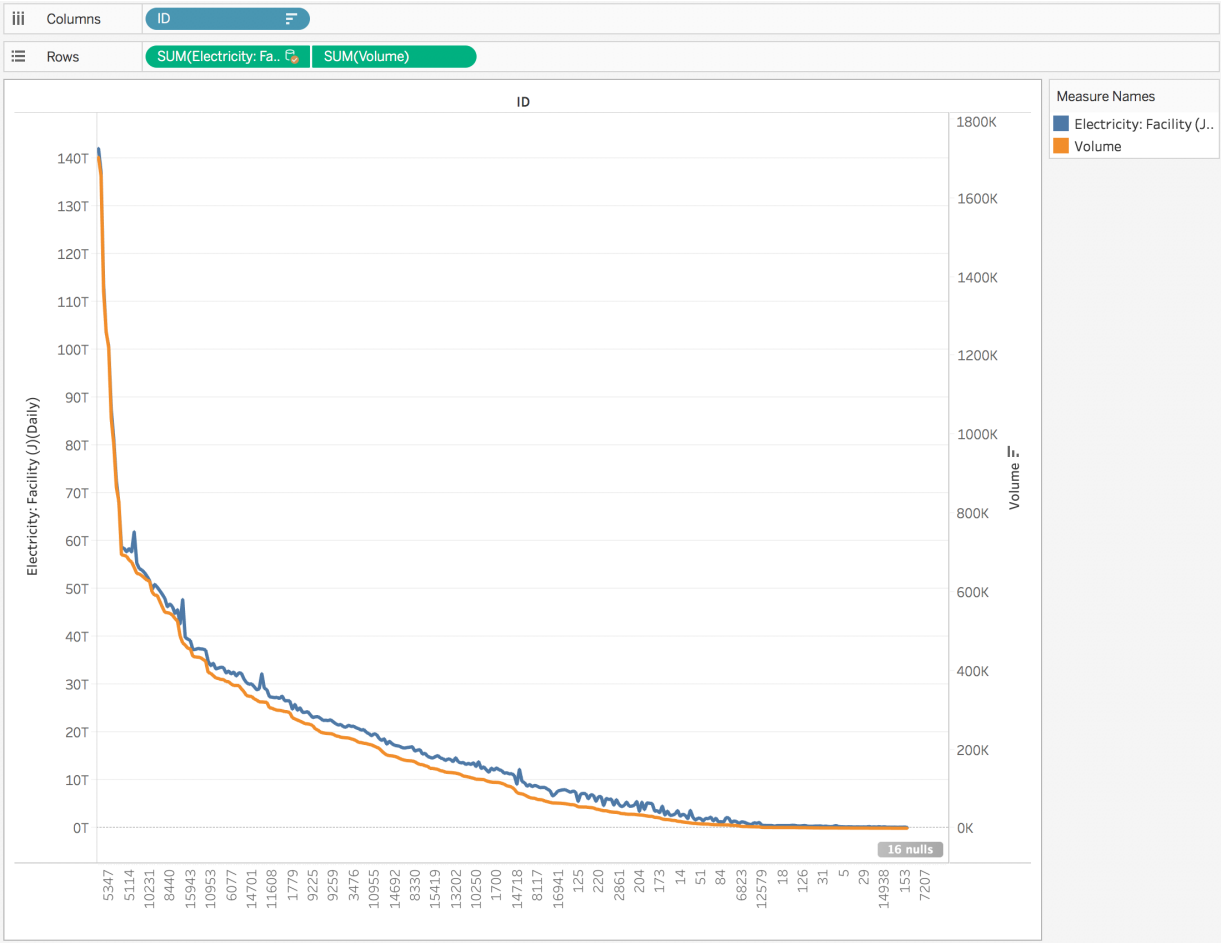


Figure 7: Plots the volumes of buildings vs the sum of their daily electricity use.

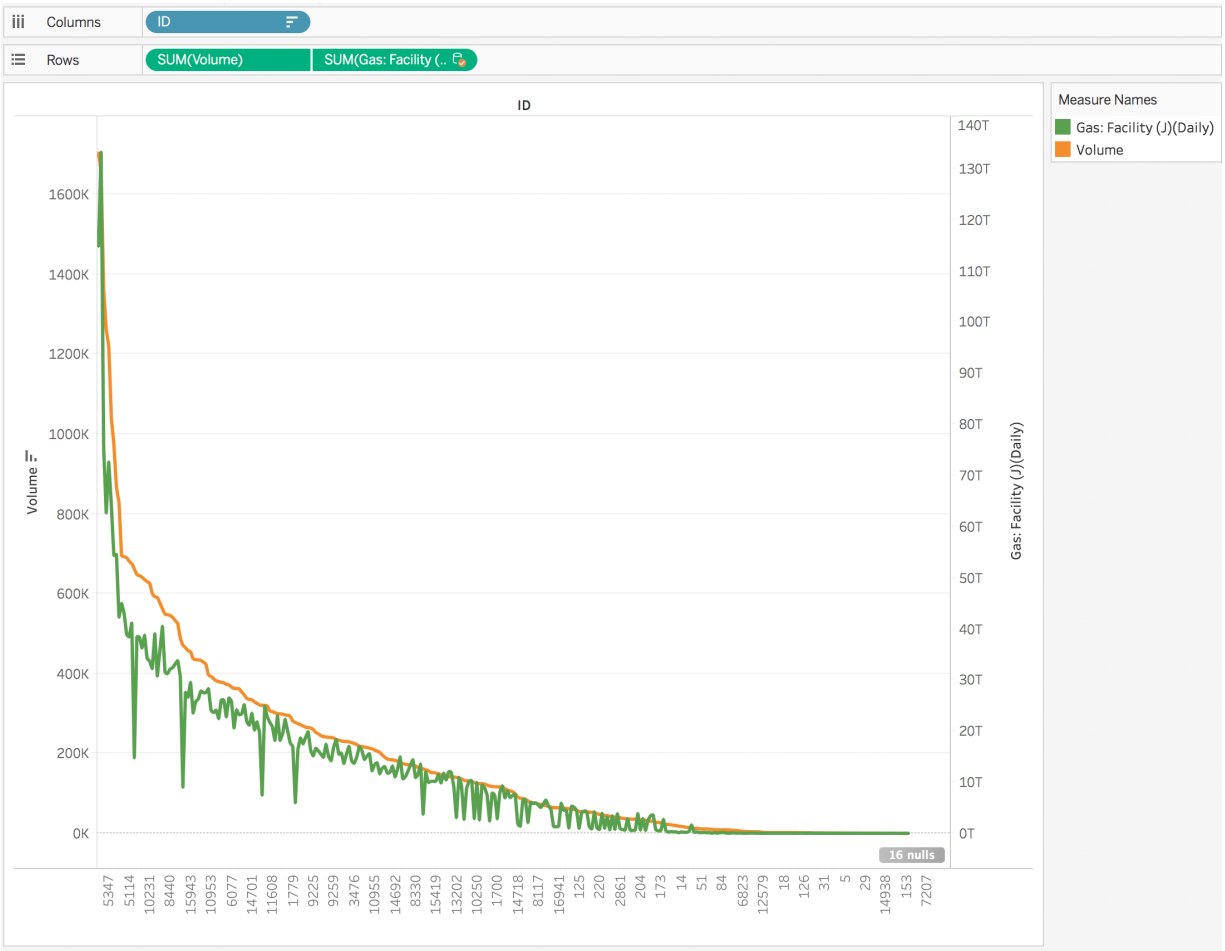


Figure 8: Plots the volumes of buildings vs the sum of their daily gas use.

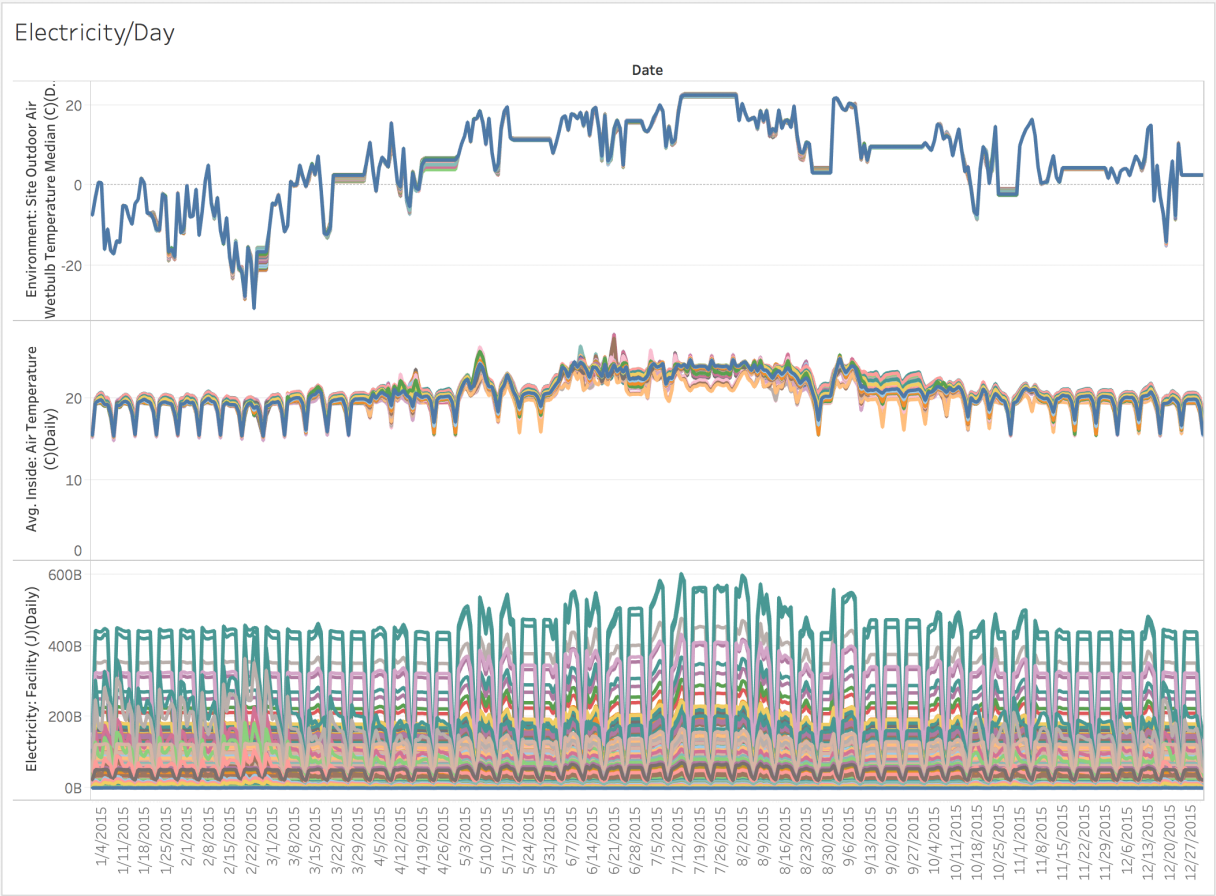


Figure 9: We see how outside temperature affects indoor temperature and electricity use for every building.

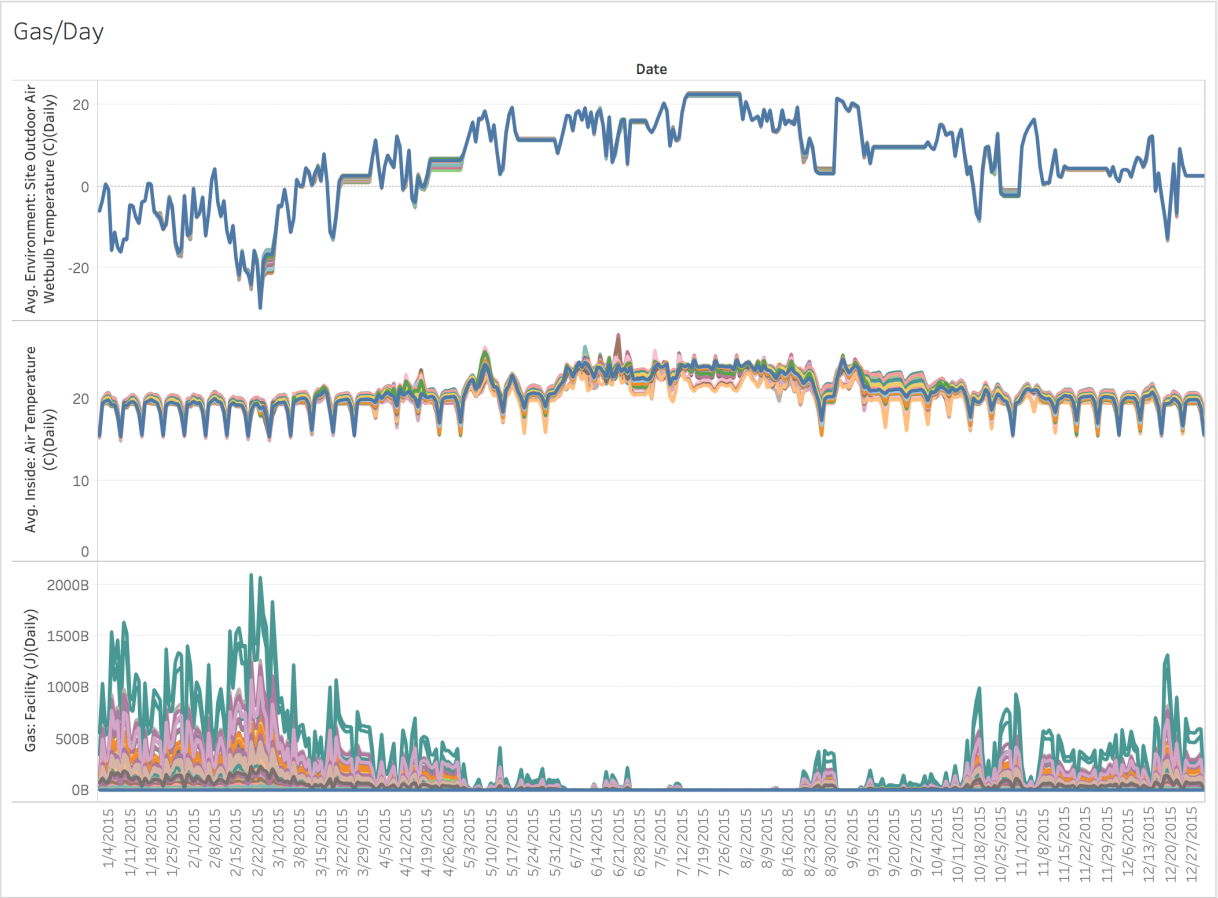


Figure 10: We see how outside temperature affects indoor temperature and gas use for every building.

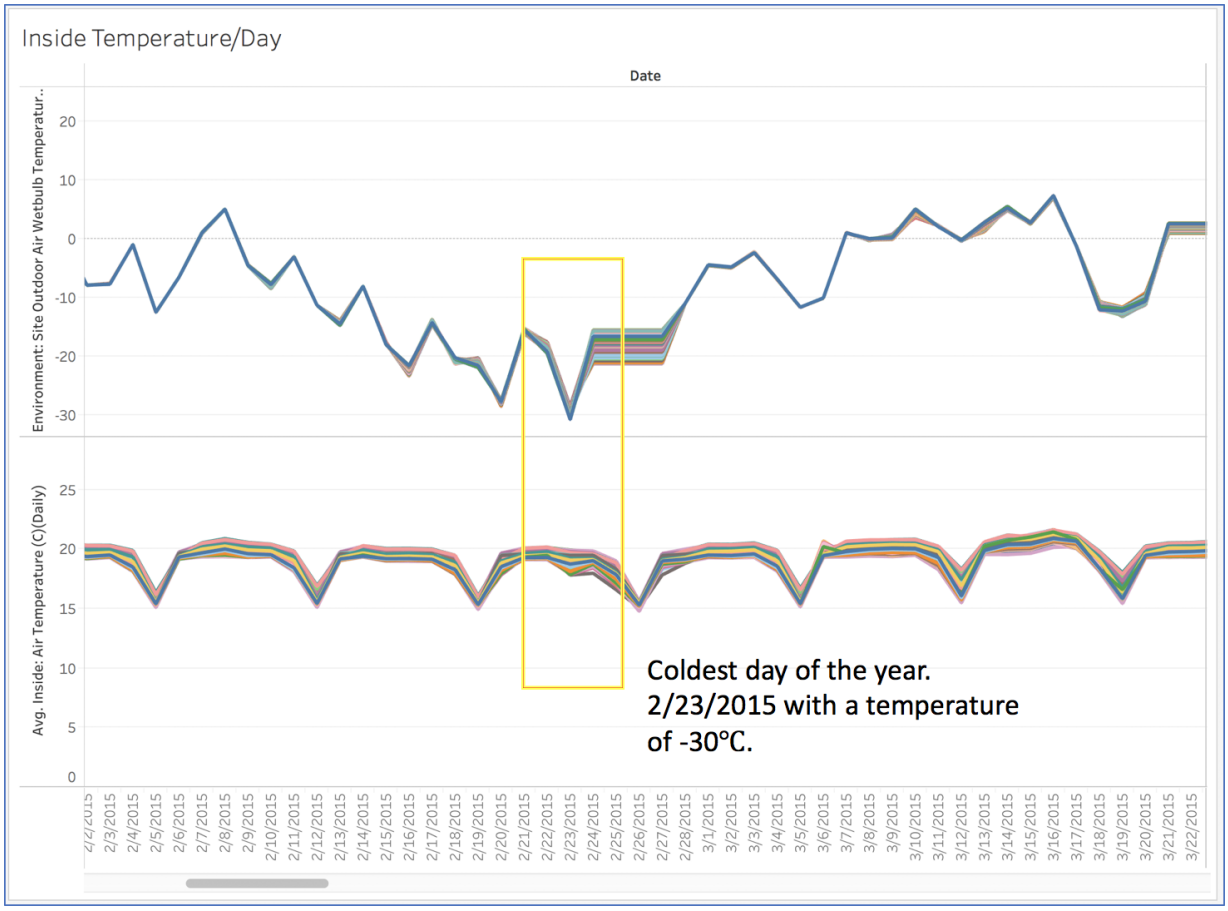


Figure 11: Close up of the coldest day in Chicago and how it lowers the temperatures of some buildings.

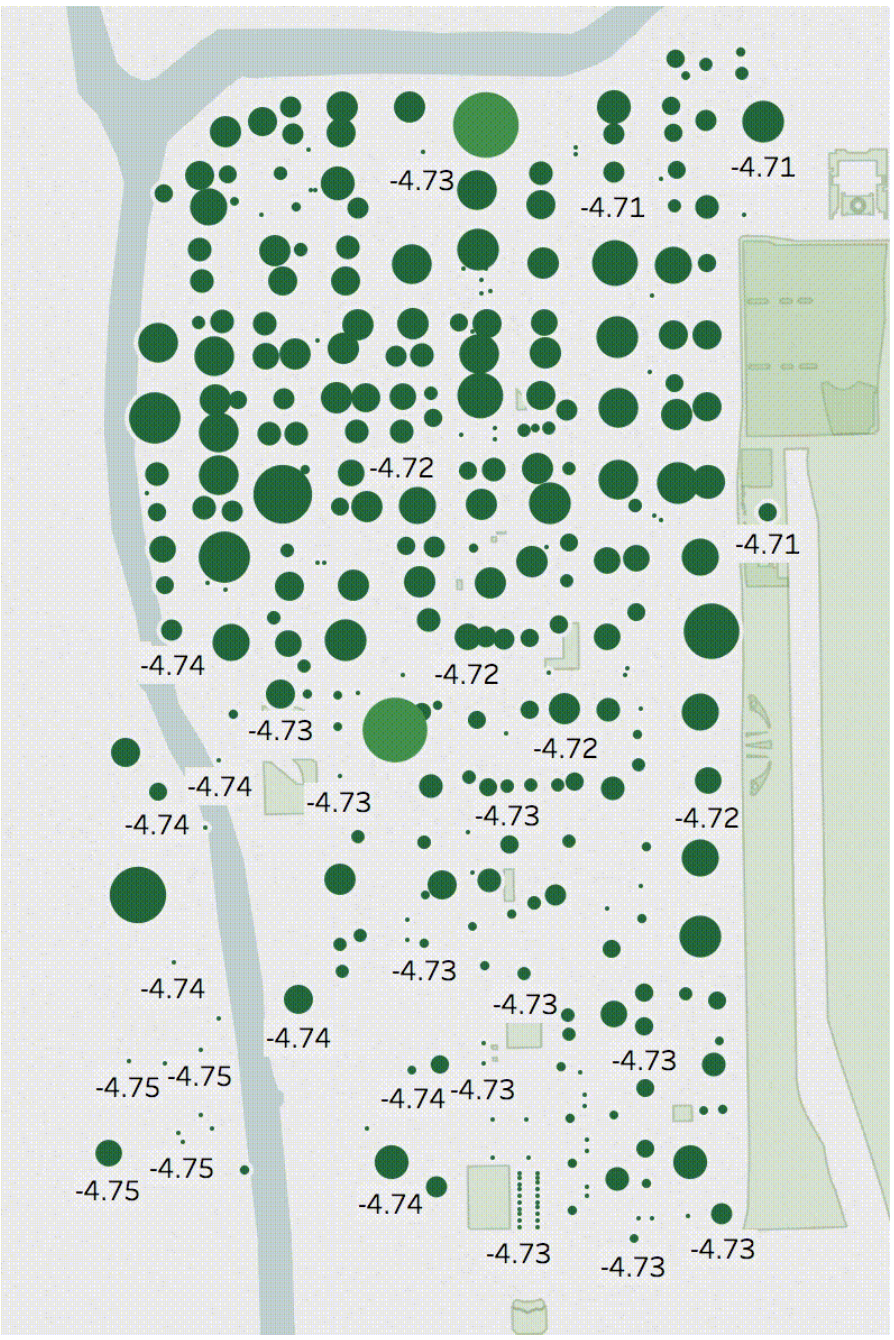


Figure 12: Snapshot from a Time Animation to help understand the geographic correlations with weather and electricity/gas usage. It represents the natural gas usage use state of all the Chicago buildings at a specific instant of time.

References

1. https://nbviewer.jupyter.org/github/cmiller8/EnerNOC-100-Building-Open-Dataset-Analysis/blob/master/EnerNOC%20Sample%20Dataset%20-%20Part%202%20-%20Weather%20Sensitivity%20Screening.ipynb
2. https://www.slideshare.net/EnergyCAP/correlating-energy-use-with-weather-data
3. http://nbviewer.jupyter.org/github/cs109-energy/cs109-energy.github.io/blob/master/iPython/Exploratory%20Analysis.ipynb
4. https://en.wikipedia.org/wiki/Cluster_analysis