Energy 396 – Bass Connections in Energy: Innovation and Design

Smart HVAC:

Applying Wi-Fi Occupancy Data to

HVAC Scheduling

Team Members: Tommy Hessel, Lawton Ives, Ivy Jiang, Jack Kochansky,

Elizabeth Lamb, Franco Picone, Ben Williams

26 April 2021

Table of Contents

Acknowledgements

We would like to thank Casey Collins and Wendy Peters from Duke Facilities Management and Mary Clair Thompson from the Duke Office of Information Technology for their help in collecting the abundant data for our analysis. We also want to thank Professor Tim Johnson, Professor Emily Klein, and Professor Josiah Knight for their willingness to answer questions about the complex energy systems on campus. Finally, we would like to acknowledge the Bass Connections and Energy Initiative programs for making our research possible.

1. Executive Summary

Heating, Ventilation, and Air Conditioning (HVAC) systems regulate indoor air flow and temperature control. These systems consume enormous amounts of energy, but efficiency upgrades are often cost-prohibitive. We conduct an analysis using a combination of energy use and Wi-Fi data from Grainger Hall at Duke University to recommend two potential solutions to decrease HVAC usage. One is to allow the temperature to fluctuate more past the set point in periods of low occupancy, and the other is to determine times of near-zero occupancy in order to shut the systems off completely. Through our analysis, we estimate that Grainger Hall can save 23.2% of its energy usage and 19.6% of its carbon footprint. In the wider U.S. commercial building market, we find that applicable buildings may be able to save a range of 5-20% of their energy consumption.

2. Introduction

According to the U.S. Energy Information Administration's (EIA) 2017 Annual Energy Outlook (AEO), the U.S. commercial building sector consumes a total 17.83 Quadrillion Btus (Quads) of energy, and HVAC systems comprise a significant portion of that total consumption at 5.35 Quads—accounting for 30% of commercial building energy use (EIA, 2017).

This energy draw is not only large, but also widely recognized as inefficient. The U.S. Department of Energy acknowledges that amongst technology options to improve energy efficiency, advanced HVAC sensors have the highest technical energy savings potential (US Department of Energy, 2017). Additional studies have found that there is a significant discrepancy between how building energy use is designed and how it is used during realistic

daily operation, indicating a need to evaluate the relationship between building occupants and energy requirements (Naylor et al., 2018, 1).

On a more local scale, Duke University's 2019 Climate Action Plan names "maximizing opportunities for building energy efficiency" and "continuing to invest in the energy efficiency of existing campus buildings with strategies such as HVAC optimization... with a goal of 20% reduction by 2024" as high-priority recommendations (Sustainable Duke, 2019). Ongoing efforts to meet this 20% target include LED lighting retrofits and switching from steam to hot water for heating in certain buildings. Our project can contribute to reduction goals as well: with low costs and significant carbon and financial savings, HVAC optimization can be accomplished with occupancy conditionals for thermostat setting rather than expensive infrastructure upgrades.

2.1 Existing Occupancy-Tracking Efforts

2.1.1 Overview

Efforts to generate HVAC energy savings through tracking occupancy are versatile but imperfect. Existing occupancy-tracking initiatives range from simple door counters to more complex methods like CO2 detection (Naylor et al., 2018, 4). The most popular sensors today are CO2 detection and simple Passive Infrared detectors; often, these methods are used in combination with one another for best results (Naylor et al., 2018, 7). Less sophisticated technology like video feed, temperature change, humidity, and others also hold a place in the smart HVAC sector (Naylor et al., 2018, 3).

A set of common challenges exist among each of these methods: false triggering, return on investment, and lag time. Addressing inaccurate changes in HVAC controls caused by incorrect or delayed detection are primary in making meaningful HVAC changes, and thus these errors reduce the desirability of occupancy-tracking mechanisms as a worthwhile expense. For instance, CO2 detectors are a current market leader, and they can achieve 91% accuracy in binary prediction, or simply detecting whether or not someone is in the room (Arief-Ang et al., 2017, 2). However, accuracy falls to 15% when they are challenged with detecting the exact number of occupants in the room (Arief-Ang et al., 2017, 2). If sellers cannot guarantee that the detector will live up to its promised energy savings, customers may decide to spend their dollars elsewhere, given that CO2 sensors can cost over $200,000 for a 100,000 sq. ft building (Lafond,

2019). The failure of an occupancy detector to accurately estimate the number of people in a room stems from a range of factors like the type of activities performed in the space or even age of the occupants.

2.1.2 CO2

CO2 sensors estimate occupancy by measuring the amount of carbon dioxide in a room. Because people emit carbon dioxide when they exhale, the sensor can detect subtle changes in the CO2 content of a small space and estimate the number of people accordingly (Arief-Ang et al., 2017, 3). The system employs a flow rate formula that takes into account metabolic rate, room size, and exhale rate (Arief-Ang et al., 2017, 3).

CO2-based systems are fairly reliable when used in smaller spaces and in conjunction with known activity levels or other occupancy-counting methods (Naylor et al., 2018, 4). However, these machines tend to be slow to calculate resulting in lag time, are less accurate in larger or open spaces, and can easily miscalculate if ventilation moves carbon dioxide through the room before the machine can finish its calculation (Naylor et al., 2018, 4). Furthermore, the formula employed uses a standardized exhale CO2 quantity and rate in its calculations, when in reality this variable changes from individual to individual depending on height and weight (Arief-Ang et al., 2017, 4). While spaces like office buildings could technically acquire this information, it raises privacy concerns (Arief-Ang et al., 2017, 4). As a result, CO2 calculations usually fall short of their full potential.

2.1.3 Passive Infrared

Passive Infrared (PIR) technology takes advantage of the fact that all objects emit heat. By passively detecting the heat that people emit, PIR sensors can detect the presence of people in a room (Creston, 2017). PIR sensors have near perfect binary detection, and some studies report that they can accurately count 10 people in a room up to 91% of the time (Yun & Lee, 2014, 8061). They are also quite inexpensive, making them the current market leader in single-room applications (Naylor et al., 2018, 4).

Despite this, PIR sensors come with some important limitations. Most of all, the studies that report such high counting accuracies are often executed under near-perfect conditions.

Conditions as simple as someone standing between someone else and the sensor can affect the sensor's ability to detect that person (Dilouie, 2017). Other conditions that can interfere with proper counting include changes in room temperature or improperly placed HVAC vents that can cause errors by blowing cold or hot air over the sensor (Dilouie, 2017). In addition, they only work within single rooms. A large office building or commercial space would need a unique sensor for each room, and they would not execute properly in open spaces, similar to CO2 detection (Naylor et al., 2018, 4).

2.1.4 Other

Other tracking methods exist, such as sifting through video-feeds or using pressure pads in chairs to detect sitting people (Naylor et al., 2018, 4). These might generate higher accuracy than the above methods; however, they present their own unique set of challenges. Both raise privacy concerns, and analyzing video feeds prevents real-time feedback (Naylor et al., 2018, 4).

Temperature sensors, simple door open/close counters, and humidity sensors can provide external information to supplement any of the above sensors. None of these can be used independently, however. Door counters, for example, cannot account for someone who holds the door open for other people (Naylor et al., 2018, 4).

2.1.5 Wi-Fi

In addition to binary detection and precise counting, many methods of occupancy sensing, including those above, also fall short in terms of data collection and location sensing. That is, having a record of how occupants move through the building and where they spend their time can provide a more comprehensive understanding of the way in which occupants affect building energy use and thus lead to greater savings (Naylor et al., 2018, 3). The primary methods that help capture this aspect of occupancy are tracking using radio-frequency, which require all members of a space to wear a trackable tag, and Wi-Fi (Naylor et al., 2018, 4).

Of the two, Wi-Fi tracking has emerged as more feasible as it leverages existing equipment. One can estimate the number of people in a building by looking at the number of devices, such as cell phones or laptops, that connect with Wi-Fi. One can also tell where these people are within a building depending on which Wi-Fi router they connect with. The ability to collect data on these

associations allow facility managers to better understand how people use a building and thus make HVAC adjustments more precise and localized.

Wi-Fi's ability to navigate the challenges outlined above – lag time, false triggering, and return on investment – also promotes it as a viable future option for occupancy sensing. With proper parameters, Wi-Fi can reliably predict a range of people in a location (see "2.1.6 Existing Research"). Additionally, data could conceivably be transmitted in real-time to HVAC control centers, which can be programmed to respond according to the current occupancy. Most importantly, Wi-Fi has virtually zero fixed costs within institutions that already have robust Wi-Fi networks installed. Our project also seeks to provide open-source resources that will make this set-up process a small effort.

2.1.6 Existing Research

Scientific studies have identified Wi-Fi-based occupancy estimation as more accurate and effective than alternative methods. For example, a study published in 2019 conducted by researchers at the University of New Mexico, Albuquerque found that Wi-Fi based occupancy had a 96% accuracy rate (Simma et al., 2019, 498). The same year, a study published by a group of Italian researchers found that a machine learning model leveraging Wi-Fi and Bluetooth connectivity data could reliably estimate building occupancy with an accuracy of over 90% (Longo et al., 2019, 3). An earlier study conducted by researchers at the University of Houston also found that Wi-Fi based occupancy estimation achieved an accuracy of over 90%, and performed better than ambient light sensing (Mohammadmoradi et al., 2017, 7).

3. Technical Design

3.1 Methodology

After establishing an objective and doing research for this project, our focus was to break down larger, complex data sets and problems into simple checkpoints and final goals. The two overarching groups of data sets that were necessary for completing this project were, as visualized in Figure 1, Wi-Fi data and Facilities data – both building energy usage and air handler unit (AHU) operational data. The original concept of utilizing Wi-Fi data is credited to the 2020 Data+ team attempting to map the path of students in the Bryan Center based on their

access point connection patterns, and will be explored further in the Wi-Fi section (Swartzendruber et al., 2021). The difficulties with the acquisition and permissions perspective will also be broken down more in the Wi-Fi and Obtaining Data sections.

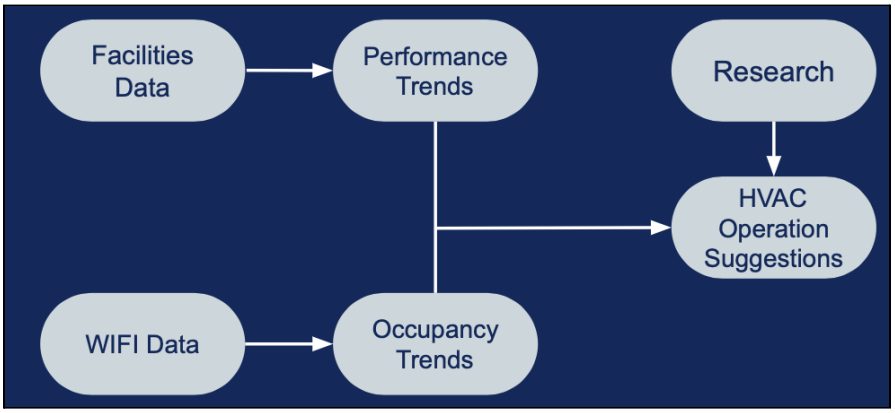


Figure 1. Flowchart of HVAC recommendation design process.

Research was another crucial step within the recommendation development process. This research centered around exploration into how best to make recommendations for improving the performance of the HVAC system and making assumptions about potential savings. These three data streams all feed into our final set of recommendations focused on the Grainger Hall BAS, specifically the hours of operation and set point variation of the AHU. By retracing the path we took to reach this solution, another group could follow using the work we have already completed to improve the operations of other HVAC systems.

3.2 Wi-Fi

3.2.1 Introduction to Large-Scale Wi-Fi Networks

Wi-Fi networks can regularly and reliably record interactions between devices and the network. These interactions happen constantly and instantaneously, and data on these interactions has the potential to help accurately estimate the number of people in a given building, or more granularly, a given room. These independent events can be recorded and stored by IT (information technology) hardware and can be used as a form of “implicit sensing”, to determine the number of devices in a building. The number of devices can then be used as a proxy for total occupancy—a crucial variable to consider when attempting to optimize HVAC operations.

Before presenting the Wi-Fi data used to estimate occupancy, it is important to understand a network's relevant functions and processes. "Wi-Fi enabled devices periodically broadcast Probe Request frames to retrieve information on the surrounding available wireless networks and their characteristics. Such frames are generally transmitted in bursts … with a temporal periodicity which depends on the device manufacturer and operating system, as well as on its activity status (e.g., active vs stand-by)" (Longo et al., 2019, 3). An opportunity exists to leverage these recorded "frames" to estimate a building's occupancy or specific rooms' occupancy. There are two types of regularly recorded requests that occur between Wi-Fi routers or "access points" (APs) and Wi-Fi enabled devices: associations and authentications.

Authentications are interactions in which the device enters some form of credentials to gain access to the AP, or automatically connects to it in the case of an open network (Thompson, 2020). For example, this might be an event that requires a user to access their phone's Wi-Fi settings, select a network, enter a password (if required) and only then will the blue checkmark appear next to the network indicating they have "authenticated" (Thompson, 2020).

Associations, on the other hand, are interactions in which the device simply recognizes an AP or network because it is within range of said network (Thompson, 2020). This is the case, for example, when a user is simply walking down the street coming into contact with Wi-Fi networks from different storefronts and office buildings. As they walk, they will see various Wi-Fi networks appear and disappear from their phone's list of available networks as they enter and leave their range. They have not connected to the networks—rather, they have simply recognized them. These are called "associations". The key distinction here is that a user does not need to connect to the network in order to associate with it, nor do they need to have their Wi-Fi turned on or "airplane mode" off (Thompson, 2020). This presents a potential opportunity for measuring occupancy due to the fact that people can be passively recorded as present at a given access point simply by carrying a Wi-Fi enabled device with them, such as a tablet, smartphone, smartwatch, or laptop, which in today's day and age nearly everyone does.

Wi-Fi-based occupancy is advantageous compared to alternatives due to the affordability of measuring occupancy on more granular, room-to-room levels without having to install sensors in each of these zones. Knowing exactly where within a given building each access point is located, information collected by each access point can be isolated to understand movement patterns and

how people are distributed throughout the building. This presents even greater opportunities for energy and monetary savings by optimizing HVAC operations on a granular basis rather than standardizing settings across entire buildings, as different rooms or parts of a building often have different heating, cooling and ventilation needs, especially in commercial buildings.

3.2.2 Wi-Fi Data

In order to estimate occupancy for Grainger Hall, home to Duke's Nicholas School of the Environment, our team used Wi-Fi connectivity data collected by Duke's Office of Information Technology (OIT) on device connections between February 21, 2021, and April 2, 2021. The dataset contains de-identified observations (authentications and associations) made by all access points in the building. Each observation contains information on the time at which the observation was made, the name of the access point at which the observation was made, the building ID, the building name, and the hashed MAC address, in addition to other Duke specific variables (i.e. netID). Most of these variables are not unique to Duke's Wi-Fi network, and can be collected by Wi-Fi networks in other commercial buildings. This standardization makes Wi-Fi based occupancy estimation promising in commercial applications given the robust Wi-Fi network infrastructure that is already present in many modern buildings. Some of the more relevant variables for our purposes are the "Time", "AP_name", and "MAC address" of a given observation. The "Time" variable is particularly useful because it can be used as a timestamp of when an association was made to extrapolate occupancy estimates at 15 minute intervals. The "AP_name" identifies the access point where the observation occurred which can provide information about the specific location of building occupants. The "MAC address" variable also presents substantial utility as it can be used to ensure accuracy in estimates of occupancy by avoiding double counting devices, given that a device's "MAC address" is a unique identifier.

3.2.3 Data to Occupancy

Our team developed an algorithm to obtain an accurate estimate of building occupancy from the records of individual associations at 15 minute intervals. Due to the low occupancy of Grainger Hall during the Covid pandemic, the algorithm was developed and tested using similar data from the neighboring Levine Science Research Center (LSRC) during a month in the fall semester. Using similar but older data to develop and test the algorithm alleviated concerns from Duke's IT

security team on the privacy and identifiability of recent data from a small building. The algorithm developed by the project team was then transmitted to Duke OIT and run on Duke servers. This procedure allowed the team to get occupancy estimates for Grainger Hall without storing sensitive data outside Duke's systems.

In developing the occupancy estimation algorithm, several important assumptions were made. Having established that associations are the relevant observations for estimating building occupancy, it is important to understand how and when associations are recorded by the Wi-Fi network in order to understand these assumptions. As a user moves throughout a building, associations are recorded as soon as the user walks within range of a given access point (Thompson, 2020). This means that users who move continuously throughout a building will periodically trigger associations as they enter the range of different access points. This also means that users who enter a building, reach their destination, and stay there for extended periods of time will no longer trigger associations, so long as they remain at their "destination". This presents a challenge in that it can be difficult to know when a user is still inside of the building but no longer moving. Another challenge to using Wi-Fi data to estimate occupancy is that it can be difficult to distinguish between users who are actually inside the building and users who are outside of the building and associate to an access point only in passing.

In order to circumvent these issues, the project team established an arbitrary yet rational length of time in which two associations would be counted as being part of the same "trip". This time window was assumed to be two hours. If two *consecutive* associations were made by the same device any more than two hours apart, these associations would be counted as separate "trips". Furthermore, the 2-hour time window was chosen because we can reasonably expect individuals inside a building to move around inside the building at least once every 2 hours to change locations or use the restroom, for example. Additionally, the team established an arbitrary but justified number of associations that must occur during a 2-hour trip, in order for the device to be counted as inside the building. This minimum number of associations was set to three. Thus, in order for a trip to be considered in the occupancy estimate, the associated device must make at least three connections during the trip. This was done to filter out trips that were likely the product of a user simply passing by a building rather than entering it, as in these situations it is expected that the user's device would only associate to one or two access points located near the

exterior of the building. All trips lasting less than 20 minutes were removed from the occupancy estimate, as we can reasonably assume that these trips were users who were again just passing by the outside of the building. Finally, the remaining trips were aggregated to obtain occupancy estimates at 15-minute intervals.

An important secondary goal of the project was to optimize HVAC operations within specific rooms or zones, by adjusting temperature set points at times of low occupancy. In order to achieve these more granular HVAC adjustments, our team developed a similar algorithm specifically for estimating occupancy within a given room or zone of a building. In order to filter trips to obtain localized occupancy estimates, the same steps as previously described were followed. Additionally, all trips that did not spend at least 50% of their time in the zone or room of interest were removed. Images of the Python functions created to achieve this granular occupancy estimation, as well as the more general, building-wide estimation can be found in Appendix A.

In developing this algorithm, special care was taken to maximize the generalizability of the code so that it can be easily modified. The hope is that this will allow our team's code to be easily repurposed for use with Wi-Fi data from other commercial buildings that are prime candidates for Wi-Fi based occupancy estimation. Additionally, parameters such as the 2-hour time window and 3 minimum associations can be easily modified to produce more or less conservative estimates of occupancy. As our knowledge and understanding of Wi-Fi networks expands, we can seamlessly update these assumptions to produce more accurate occupancy estimates.

3.3 HVAC

3.3.1 Introduction to HVAC

Many commercial Heating, Ventilation, and Air Conditioning (HVAC) systems operate differently from the residential systems that are more familiar to many of us. Larger commercial systems typically provide constant ventilation and centralize production of thermal energy to efficiently ensure that buildings are safe and comfortable. This analysis focuses on Variable Air Volume (VAV) systems, which is the dominant strategy for ensuring thermal comfort in an energy-efficient manner. These systems are typically controlled using a Building Automation System (BAS), which uses a type of proprietary code from the HVAC system manufacturer to

program different parts of the system to respond to a variety of inputs from temperature, weather, and system-performance related sensors. Most of the buildings on Duke's West Campus, including Grainger Hall, use VAV systems with chilled water provided from the university's three chilled water plants and two steam plants (Duke University, n.d.; Schramm, 2020).

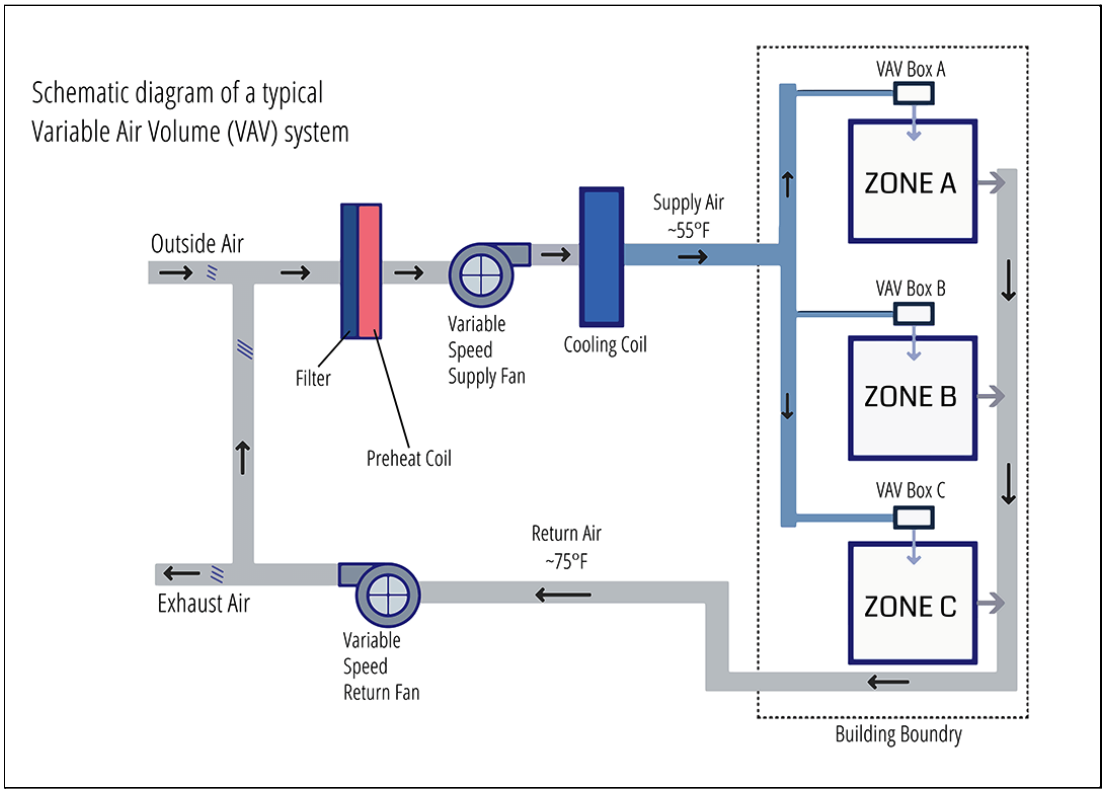


Figure 2. Simplified diagram of a Variable Air Volume (VAV) HVAC system (Severn Group).

Figure 2 shows a simplified diagram explaining the operation of a Variable Air Volume system like the one that cools the classrooms and communal spaces in Grainger Hall. The air handler sits in the building's basement and contains two fans that drive air movement throughout the system (Collins, personal communication, 2021). This air handler runs whenever the system is on (usually when building managers expect the building to be occupied) to provide constant ventilation and centralized cooling for the whole building. This constant cooling and ventilation is provided by two fans working together with a cooling coil and a damper that allows outside air to enter the system (sometimes described as the "economizer damper"). The return fan pulls air from the building into the air handler, where it is filtered and then cooled with chilled water

produced in a centralized plant (Collins, personal communication, 2021). In buildings that exist outside of a campus setting, chilled water can be provided by a chiller dedicated to that building's HVAC system. In both cases, the chilled water is usually created using electricity (Trane Technologies, n.d.). When humidity and temperature are mild, the economizer damper will open in order to take advantage of "free" cooling that the outside air can provide (Collins, personal communication, 2021). Even in winter when temperatures outside are frigid, most buildings have constant demand for cooling. This is because when spaces in the interior of a building are occupied, people's body heat causes the temperature to rise, which leads to cooling demand. The centralized cooling is also used to manage humidity by cooling the air below the dew point and removing condensation. In regions with humid climates, like much of North Carolina, managing humidity is important to reduce mold growth and reduce the likelihood of transmitting airborne pathogens (Consulting Specifying Engineer Staff, 1970).

Once air has been properly cooled and dehumidified, a second fan in the air handler distributes the air into the building's ducts. Just before arriving in a room, the air goes through a Variable Air Volume (VAV) unit; these small units individually control the amount of heating and cooling in each zone in the building (Evans, 2019). A zone is a small area of the building monitored by a single thermostat. Typically, a zone is a single room or hallway, but some larger rooms and common spaces have multiple zones to account for the complex ways that air moves in larger spaces. In Grainger Hall, Field Auditorium is served by two zones and the open common space in the center of the building is served by multiple zones on each floor.

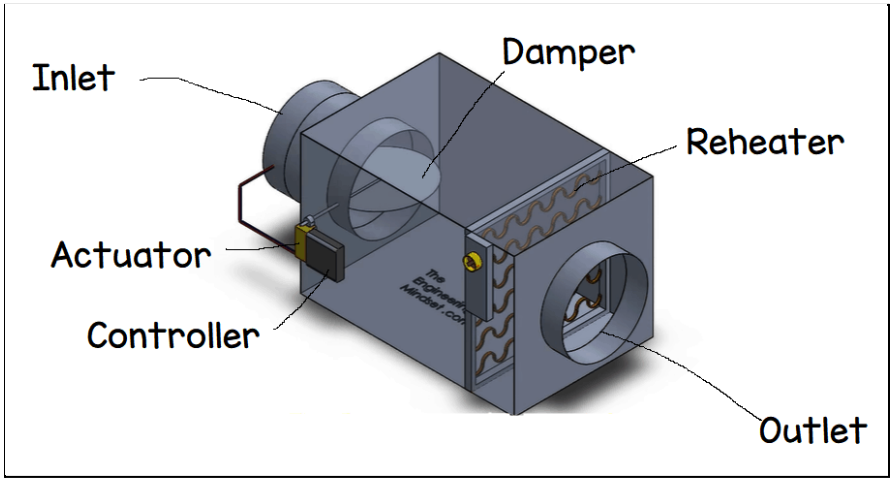


Figure 3. Variable Air Volume (VAV) Unit diagram (TheEngineeringMindset).

The VAV units are relatively simple, with a damper and reheater coil being the most important components (Figure 3). The damper varies the volume of newly-conditioned air being released into the room, which effectively controls cooling and ventilation. The reheater coil uses steam or hot water to provide for any heating needs the room may have. On Duke's campus, steam is produced from natural gas at two centralized steam plants (Duke University, n.d.). In other commercial buildings, steam or hot water is often provided by a dedicated boiler inside the building (Farsetta, n.d.). The VAV units allow building operators to respond differently to temperature changes in each zone, which ensures greater efficiency than making changes at the whole building level.

Controlling commercial HVAC systems—which for a given building might include multiple air handler units and hundreds of zones—is no small task. Building managers typically rely on a Building Automation System (BAS), which includes control panels that are installed around the building near VAV boxes and other mechanical equipment that are programmed to respond to inputs from various sensors and perform other actions on a schedule. The BAS is set up to function with or without internet access and input from human operators. When needed, the BAS can be set to retain data from specific sensors and controls. As explained below, the Facilities Management team at Duke archived a vast amount of BAS information for this project. In the future, real time reports of occupancy from Wi-Fi records could be used to adjust HVAC settings in the BAS. The BAS HVAC system for Grainger Hall is shown in Figures 4, 5, and 6.

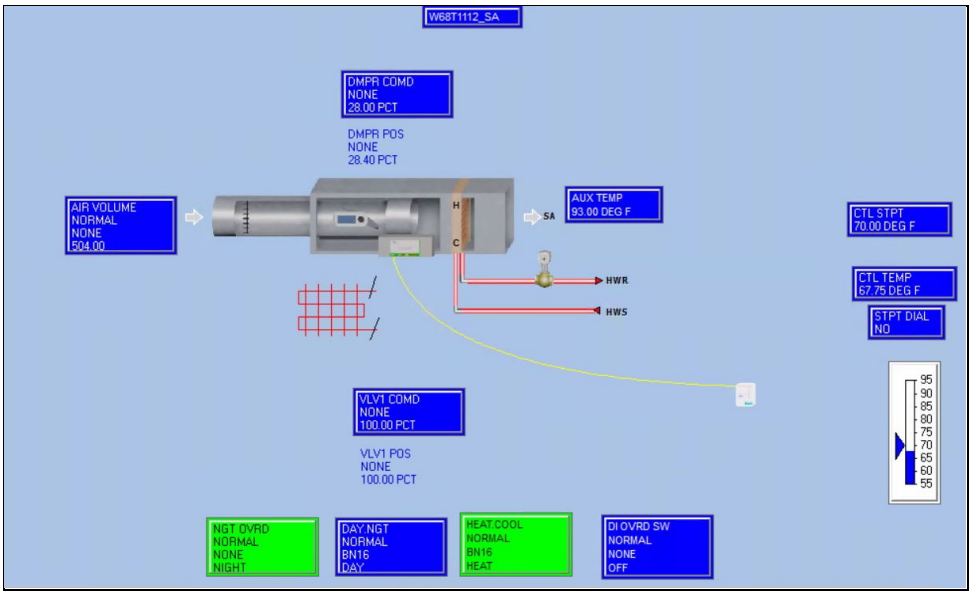


Figure 4. BAS representation of a Variable Air Volume (VAV) Unit in Grainger Hall.

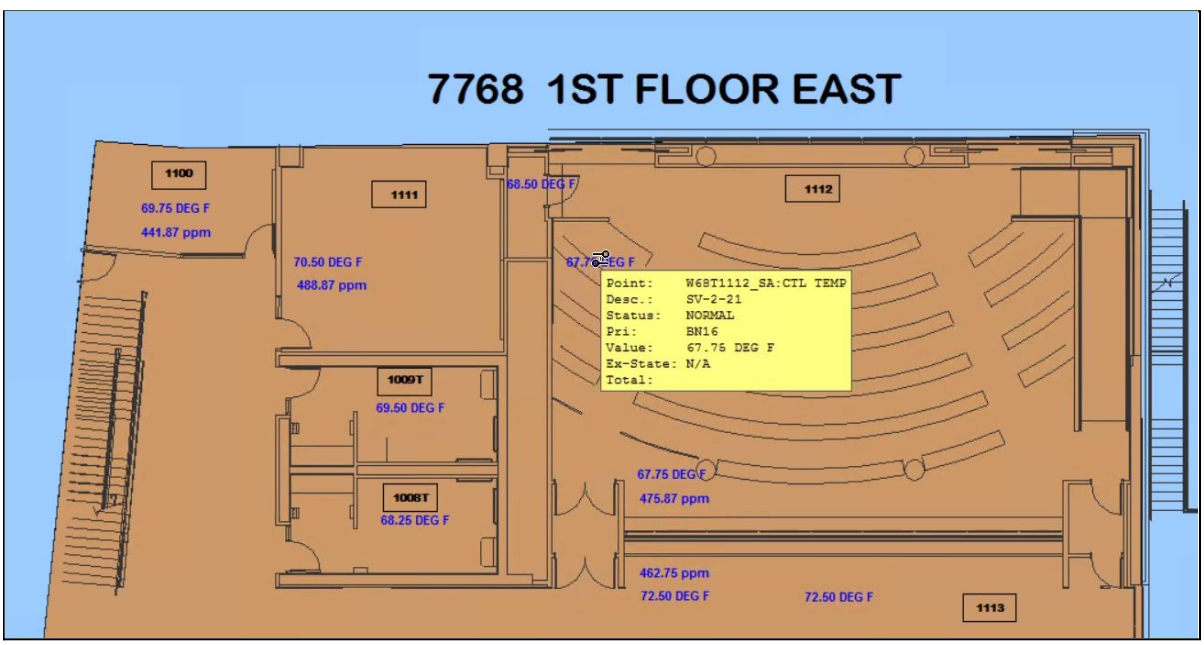


Figure 5. BAS representation of temperatures in various zones in the first floor of Grainger Hall.

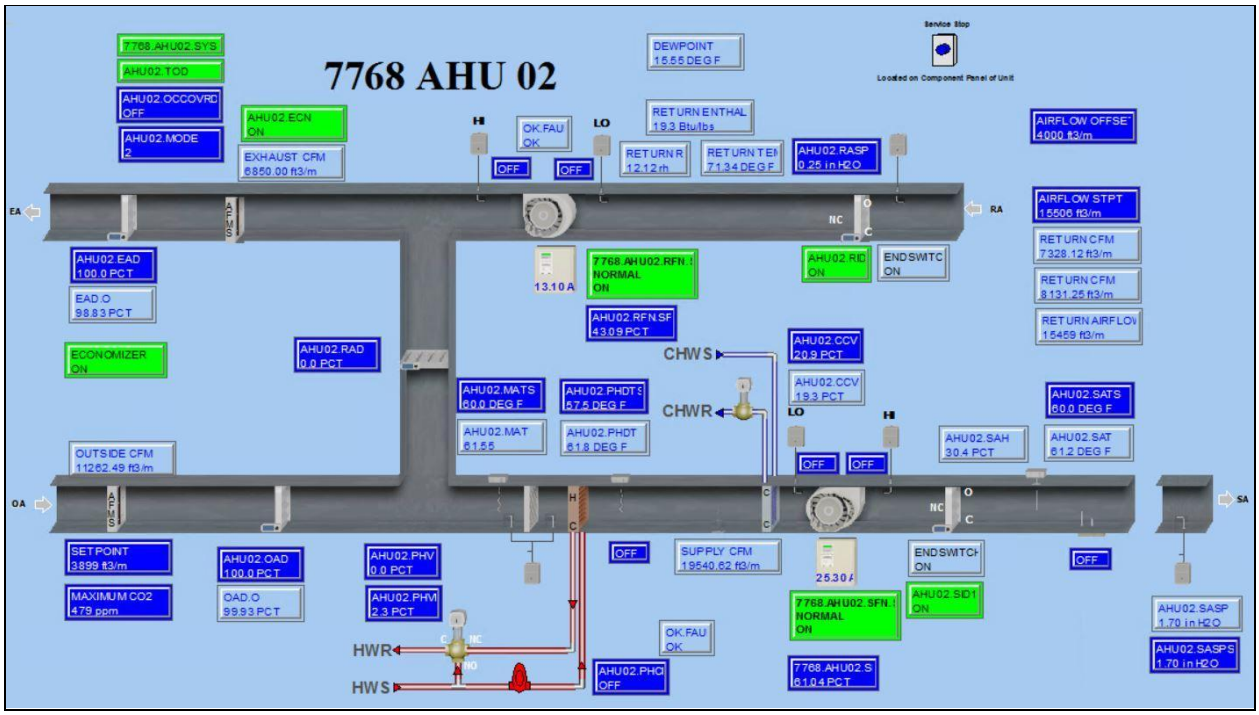


Figure 6. BAS representation of Grainger Hall's largest air handler unit: AHU 02.

3.3.2 Levers for Efficiency Improvement

Our research and conversations with Duke facilities managers guided our team toward two main opportunities for generating savings in the HVAC system. We aimed to identify times when the air handler (and thus the whole HVAC system) could be shut off and other windows of time when it would be appropriate to let the temperature of specific zones move outside their typical range.

Although many ventilation fans in air handler units have variable speeds, the reduction in energy consumption from reduced speeds is negligible. To achieve meaningful savings in ventilation, our team needed to find times when the building is likely to be completely empty. When the building is completely empty, the air handler can be turned off without the risk of stagnant air making the building uncomfortable or spreading pathogens to occupants. Buildings can also withstand periods of time without humidity control as long as they are not excessively long. Beyond the published schedules for academic spaces and some information on work shifts, Duke FMD does not have a clear picture into when buildings are occupied–particularly in the late

evening and early morning. We focused our analysis on these time windows to ensure recommendations were most practical.

Smaller gains in efficiency can also be achieved by determining when it is appropriate to allow some zones to accommodate a wider range of temperatures. The BAS code can be changed to choose temperature setpoints for specific zones that are closer to the outside air temperature than typically used. These settings can be used in times when a space is unlikely to be occupied, even if other parts of the building are still in use. These changes need to be in place for at least 30 minutes for them to be meaningful, and the system performs better when an exceptionally cool zone is not adjacent to an exceptionally warm zone. Identifying times and zones where setpoints could be adjusted in this manner was a secondary goal of the team's analysis.

3.4 Obtaining Data

Our team worked with a variety of campus stakeholders to obtain and leverage the three main datasets for this project: energy consumption, air handler usage, and Wi-Fi occupancy. Each dataset came with its own set of challenges with regards to obtaining, cleaning, and understanding the contained values.

First, we obtained the energy usage dataset to understand the campus energy profile (composed of values such as Ton-Hours of cold water usage for Grainger). We worked with members of the Duke Facility and Management Department (FMD), Wendy Peters and Casey Collins, to collect, format, clean, and ultimately analyze these data. Peters and Collins graciously agreed to work with us to build upon the work of previous Data+ and Bass Connections teams (Swartzandruder et al., 2021). Therefore, we had unfettered access to the campus energy usage metrics. Cleaning these data proved to be difficult. The scale of the dataset—13 attributes (columns) and over 72,000 records (rows)—made cleaning the data by hand impossible, and required that all cleaning be automated. Since the initial data obtained were only reported as cumulative values (e.g., total energy use per day), we had to break out the usage into 15-minute intervals with a python script. Our script also had to handle periods of data loss, unexpected sensor errors (e.g., a negative value or a value lower than one 15 minutes before), and periodic resets both related and not related to the completion of a 24-hour cycle (e.g., daylight savings).

Second, the team continued to work with Collins and Peters to narrow our scope and obtain more granular HVAC data. The inputs we needed to manipulate to directly reduce HVAC energy usage came from the air handler units within the buildings. In determining which building to pull air handler unit data from, we evaluated three main considerations: data availability, size, and building room types. The first criteria was the most restrictive because few buildings on campus have AHUs with accurate, 15-min reporting on multiple air flow and temperature metrics. This first requirement ruled out all the "red bricks" on East Campus and most of the buildings around Abele Quad. For the size, our team wanted to capture a building that had open spaces, but was not so large that it would not respond quickly to temperature changes. With a suggested cap of 100,000 sq. ft (close to the average commercial building size), we ruled out spaces like French Science Center, the Bryan Center, and West Union all with areas greater than 200,000 sq. ft and turned to smaller spaces like sections of the Levine Science Research Center and Grainger Hall with 73,515 sq. ft (EIA, 2015). Finally, our last consideration on building types led us to steer away from buildings with labs, libraries, and food service since these facilities need near constant ventilation. Altogether, Grainger was the building that best fit our criteria and additionally helped ensure our estimates would be conservative since it is already LEED-Platinum efficient (Duke University, 2015).

After initial analysis of air handler building reports, we identified just under 70 sensors or points we wanted to collect data from based on the Grainger Hall diagrams we obtained through Duke Facilities Management (FMD). These ~70 points provided insights into how the building HVAC systems worked and how the building's climate responded to changes in these HVAC inputs like shutting them off or being bound by temperature setpoints. Even with weekly collections of these 15-minute intervals of Grainger's air handler, the dataset only has ~3,000 records, making for easy manipulation and visualization.

3.5 Exploratory Analysis

We began our analysis by using the current supplied to the air handler fans to determine the times at which the system is running and how they align with the building's class schedule. Outlined below (Figure 7) is a chart showing the air handler's schedule for a typical week along with the registrar's class schedule for Grainger Hall.

Another observation within the context of the HVAC data and its trends are how the temperature set points of the system exist in binary. Whenever the system is running, FMD has set the system to keep the temperature in most of the building at 70 degrees. This leaves open the opportunity to refine the system to function on a gradient allowing us to reliably estimate the interior temperature and control heating and cooling on a smaller scale within the building.

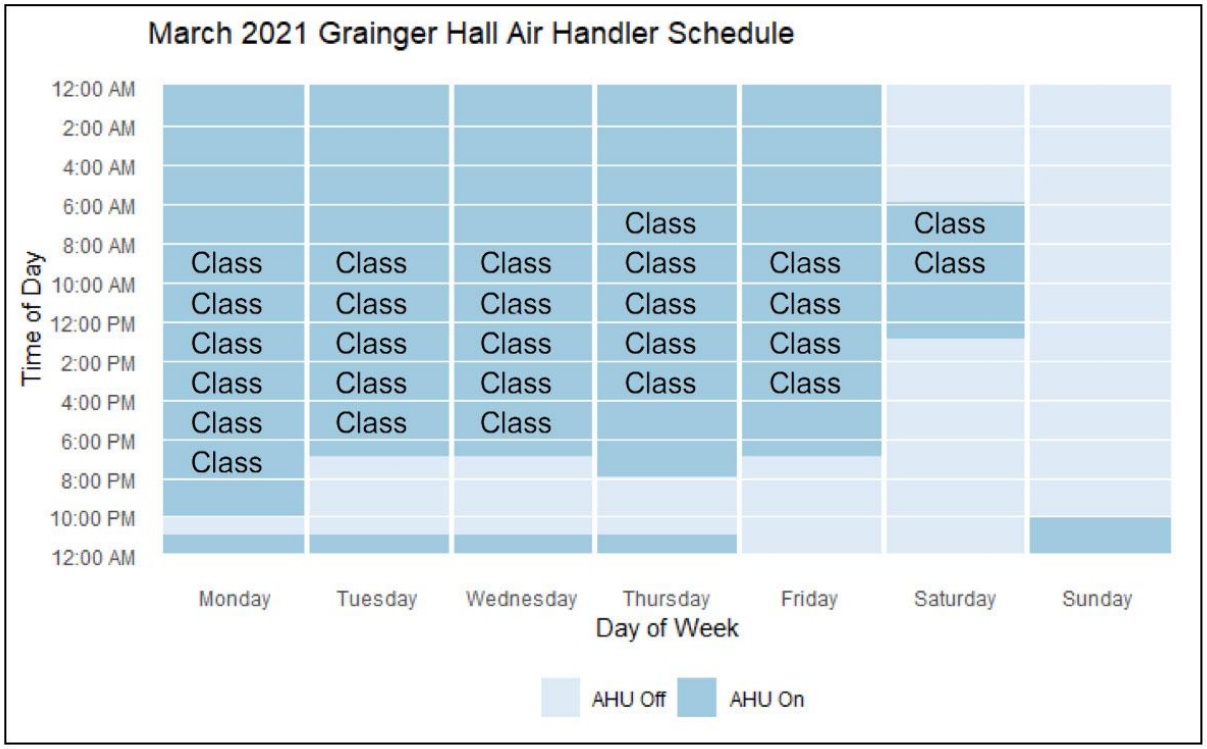


Figure 7. AHU operation schedule in Grainger Hall for March 2021. The system was entirely off at the times shaded in the lighter color and on at the times shaded in the darker color.

3.6 Occupancy Analysis

After acquiring occupancy data from Duke OIT, the team performed a set of exercises to determine the probabilities of occupancy levels being above specific thresholds throughout the week. The team then used these probabilities to make recommendations on when to completely shut off the air handling system, when to allow the system to use wider setpoints, and when to operate in a traditional manner. Lastly, the recommendations were adjusted to leave settings in place long enough for the system to respond and to ensure that they flowed logically.

Due to security and privacy concerns, Duke's OIT team returned results of the occupancy algorithm in five device intervals for a four week period in March 2021. The absence of devices during a given 15-minute window was recorded as zero and all other counts were rounded up to the next five. For example, the data would record a time period in which 13 devices were in the building as 15. A time period with 6 devices would be recorded as 10. In the future, other users applying this method should attempt to use raw data rather than rounded numbers.

All of the data referenced in this section of the report refers to this rounded count of devices as "occupancy." It is important to remember that these counts are a proxy for the true occupancy of the building. Duke OIT suggests that a typical building user might have 1.5 devices that associate with the Wi-Fi network, but further research is needed to tie the device count to a reasonable count of true occupancy (M. Thompson, personal communication, 2021).

The data generally showed expected trends in building occupancy. Figure 8 shows the mean recorded occupancy during the observation period for every 15-minute period during a typical week. Periods of especially high occupancy are clustered between 9:00 am and 4:00 pm Monday through Thursday. These high occupancy times also include more intense hour and a half blocks that appear to correspond with when classes meet in the building. On weekends and in the overnight hours, occupancy is generally low.

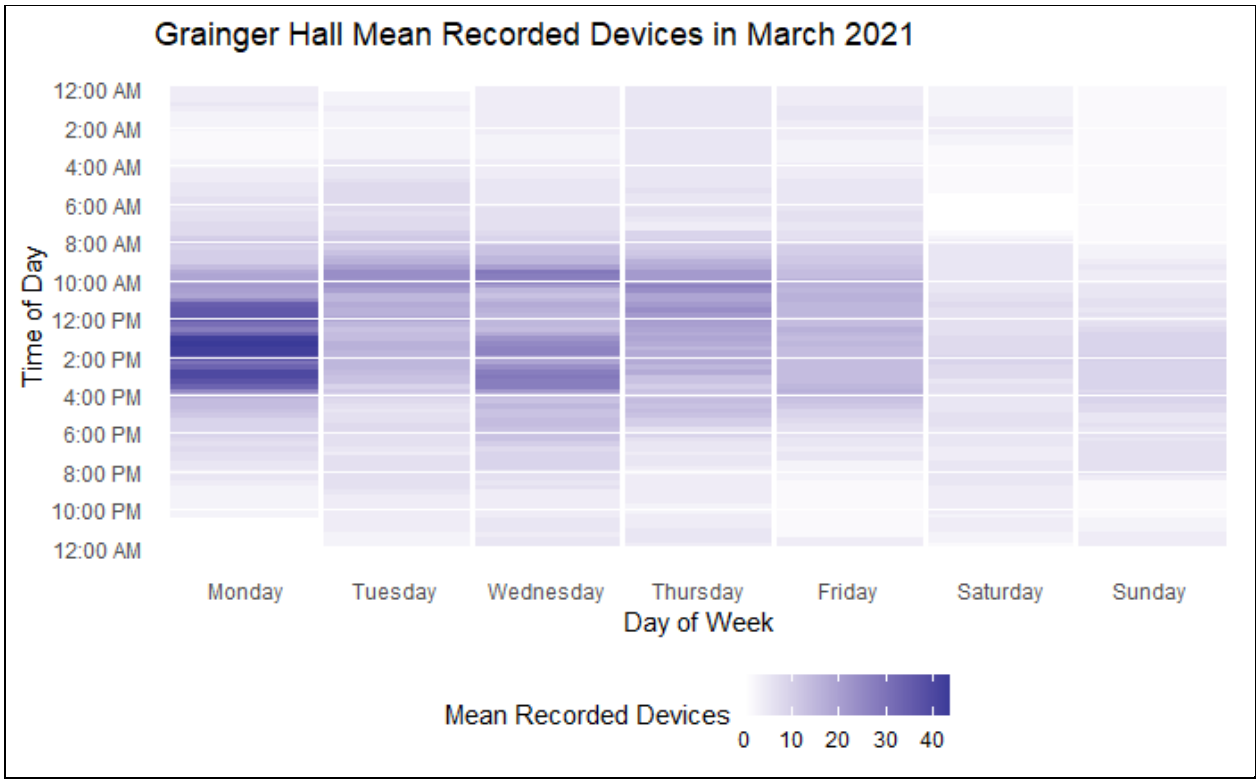


Figure 8. Mean rounded count of individual devices in Grainger Hall during each 15-minute period during a typical week over the trial period.

The team next used the data to determine the probability of occupancy exceeding five or ten devices during a given 15-minute period. Given the limitations of the data, described above, these benchmarks reasonably represent the states in which the building is unoccupied or has low occupancy. Since data recorded as five includes any time that four or fewer devices were present in the building, Grainger hall was likely effectively empty for most of the time in which this value was recorded. Such low counts of devices could represent a single person standing outside the building, or someone making a quick trip into an office. Despite some ethical concerns and unique considerations for the coronavirus pandemic, the team determined that for the purposes of this exercise, it is reasonable to recommend shutting off the air handling system when the probability of the recorded occupancy being higher than five is low. We refer to 15-minute periods that meet this criteria as times of "no occupancy" for ease of discussion.

When the recorded occupancy is between five and ten, it is likely that someone is inside the building. However, the light use makes it unlikely that occupants are congregating in the larger

communal spaces. These times likely represent people working in the building's offices and/or people sitting individually in larger communal spaces. In the interest of saving energy, it is likely appropriate to use wider setpoints during these times of relatively low occupancy. We refer to 15-minute periods that meet this criteria as times of "low occupancy" for ease of discussion.

To determine which times fit the criteria for "no occupancy" and "low occupancy," the project team assumed a normal distribution of true occupancy during a typical week and calculated the cumulative distribution function (cdf) for five and ten occupants during each 15-minute period during a typical week. These values were calculated using Student's t-distribution with an appropriate number of degrees of freedom for the number of observations (typically 3 degrees of freedom). Given the small sample size, the team decided that a cdf value above 0.75 would be necessary for determining that a time period is a period of either "no occupancy" or "low occupancy." Figures 9 and 10 show a calendar view of time periods that meet the criteria for "no occupancy" and "low occupancy," respectively.

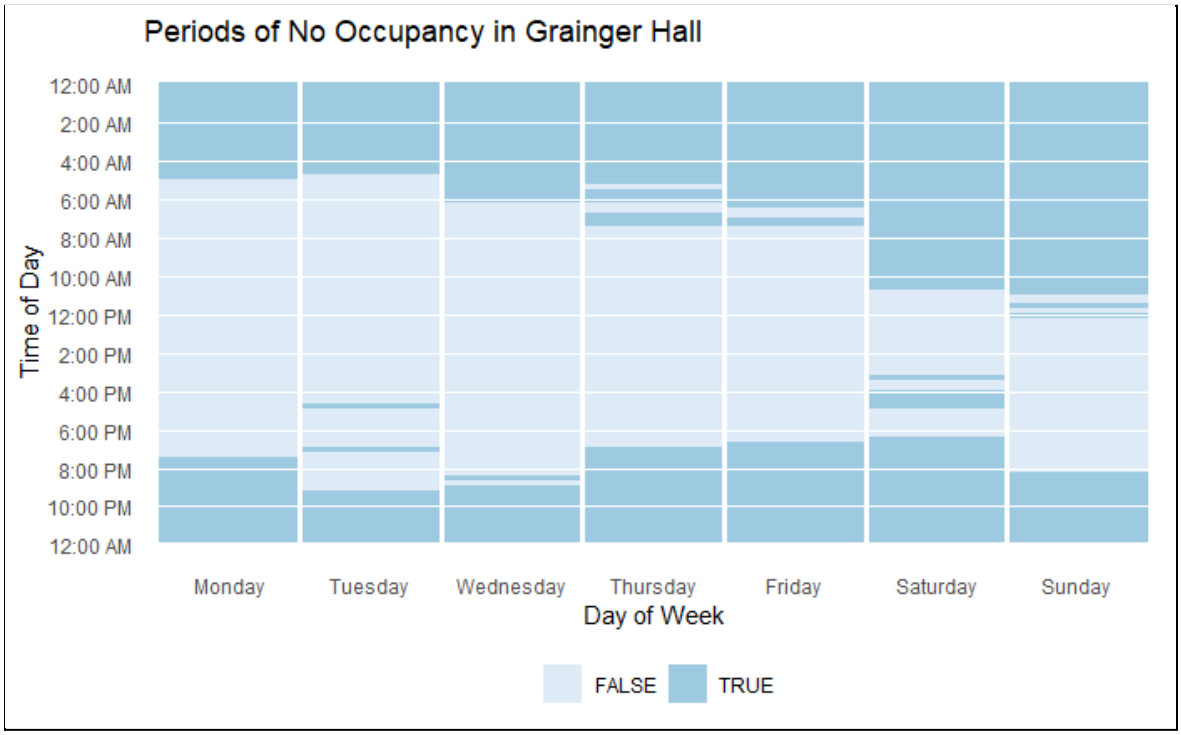


Figure 9. Calendar view of whether 15-minute periods during a typical week meet the criteria for "no occupancy." Darker shaded periods represent time when the probability of *fewer than 5 devices* being recorded as present in the building was higher than 0.75.

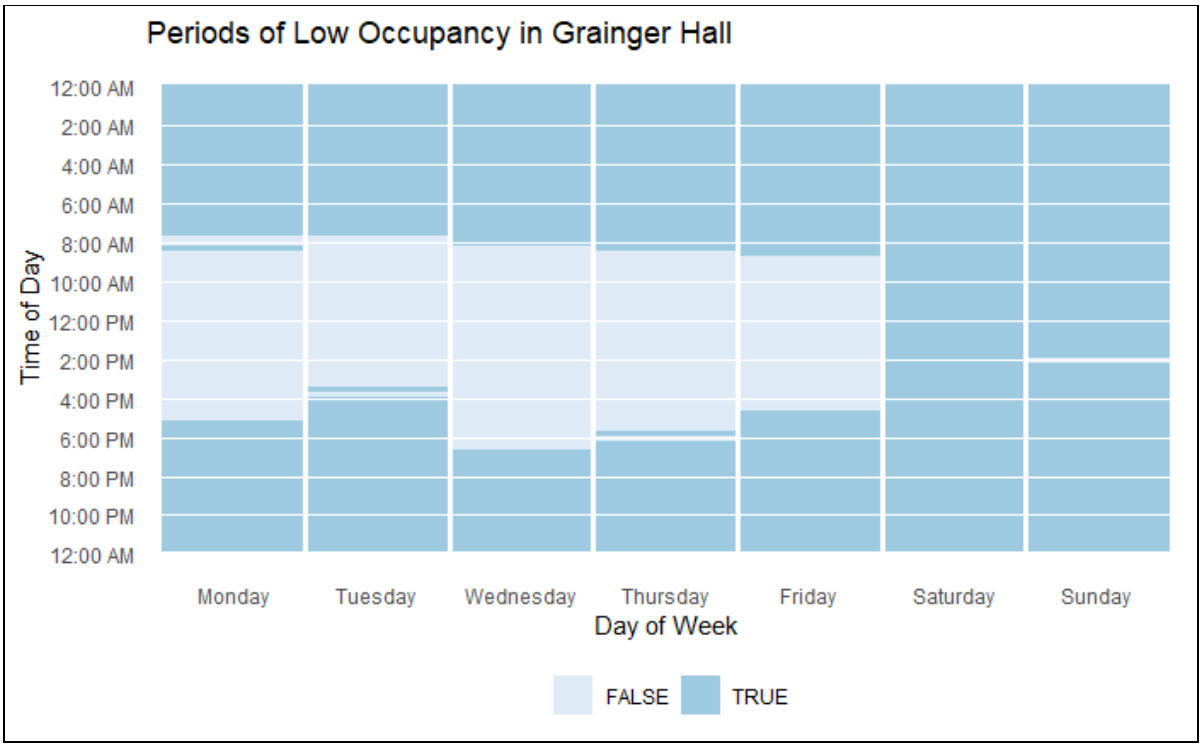


Figure 10. Calendar view of whether 15-minute periods during a typical week meet the criteria for "low occupancy." Darker shaded periods represent time when the probability of *fewer than 10 devices* being recorded as present in the building was higher than 0.75.

3.7 Recommendations

The team combined the results of the probability analysis described in the section above with a set of criteria for ensuring that the HVAC system can respond to changes in settings appropriately to develop a set of recommendations. Initially, the classification results described in Figures 9 and 10 were combined to create a set of "raw" recommendations. A calendar view of these recommendations is available in Figure 11. Time periods that met the criteria for "no occupancy" are labelled as times where the air handler should be turned off and times that only meet the criteria for "low occupancy" are labelled as times that "Wide Setpoints" should be used. Time periods that did not meet the criteria for either category are shown as times that "Normal Operation" should continue.

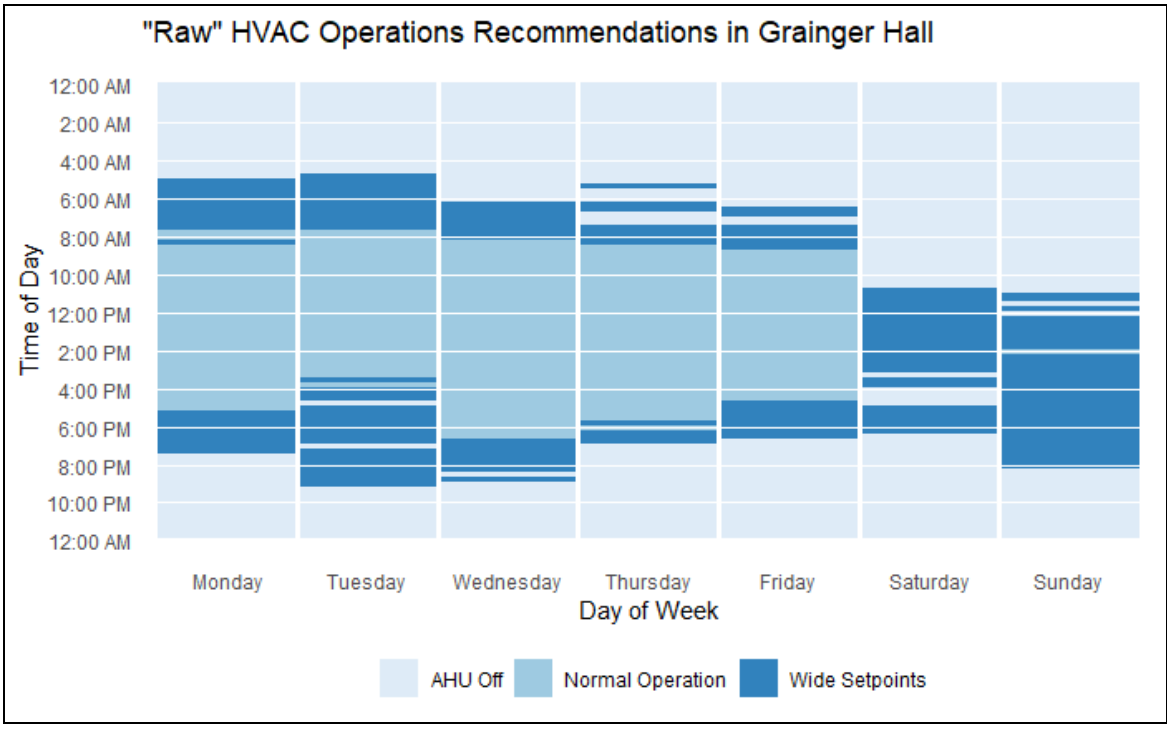


Figure 11. "Raw" operations recommendations for the Grainger's HVAC. These recommendations were created by combining the classifications for "low occupancy" and "no occupancy".

These recommendations generally follow intuition about commuting patterns as periods of no occupancy occur overnight with times of low occupancy as the building transitions in and out of traditional working hours. However, the "raw" recommendations include a variety of times where operators would have to transition between various settings repeatedly in periods of only a few hours.

To avoid inefficiencies when the system shuts off or slows only to have to recondition air a few minutes later, Duke FMD advised the team that periods of wide setpoints needed to last at least 30 minutes and shut offs of the air handler needed to last at least an hour. The recommendations were manually adjusted to ensure that each time period used settings that were at least as comfortable as the "raw" recommendations and compliant with the minimum time rules stated above. Keeping every time period "at least as comfortable" means that AHU Off times could be adjusted to either of the other statuses and Wide Setpoint times could be adjusted to Normal Operations. Time periods designated as Normal Operations in the "raw" recommendations could not be adjusted. The resulting adjusted recommendations are explained in Figure 12.

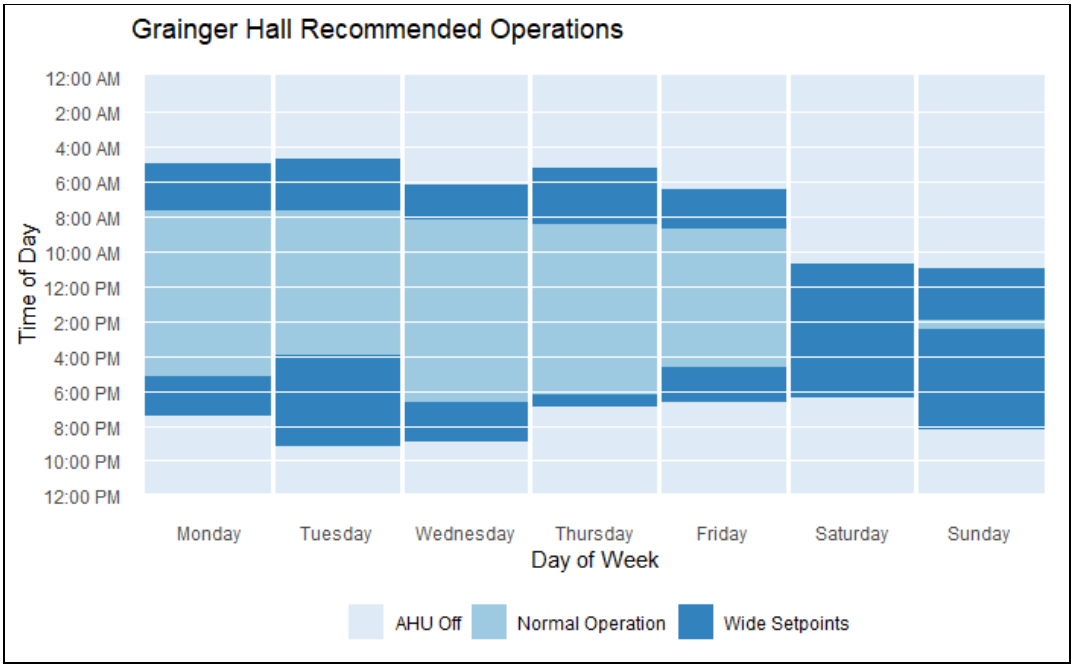


Figure 12. Calendar view of recommended operations for Grainger adjusted to meet the minimum time requirements to prevent losses in efficiency from repeated transitions between statuses.

These recommendations represent a significant departure from the schedule that Duke FMD used during March 2020. Due to the pandemic, Duke has focused on providing ventilation whenever there is a chance that buildings are occupied. In academic spaces like Grainger Hall, this includes providing ventilation overnight for security and housekeeping staff. Figure 13 compares the current schedule to the team's recommendations during the work week.

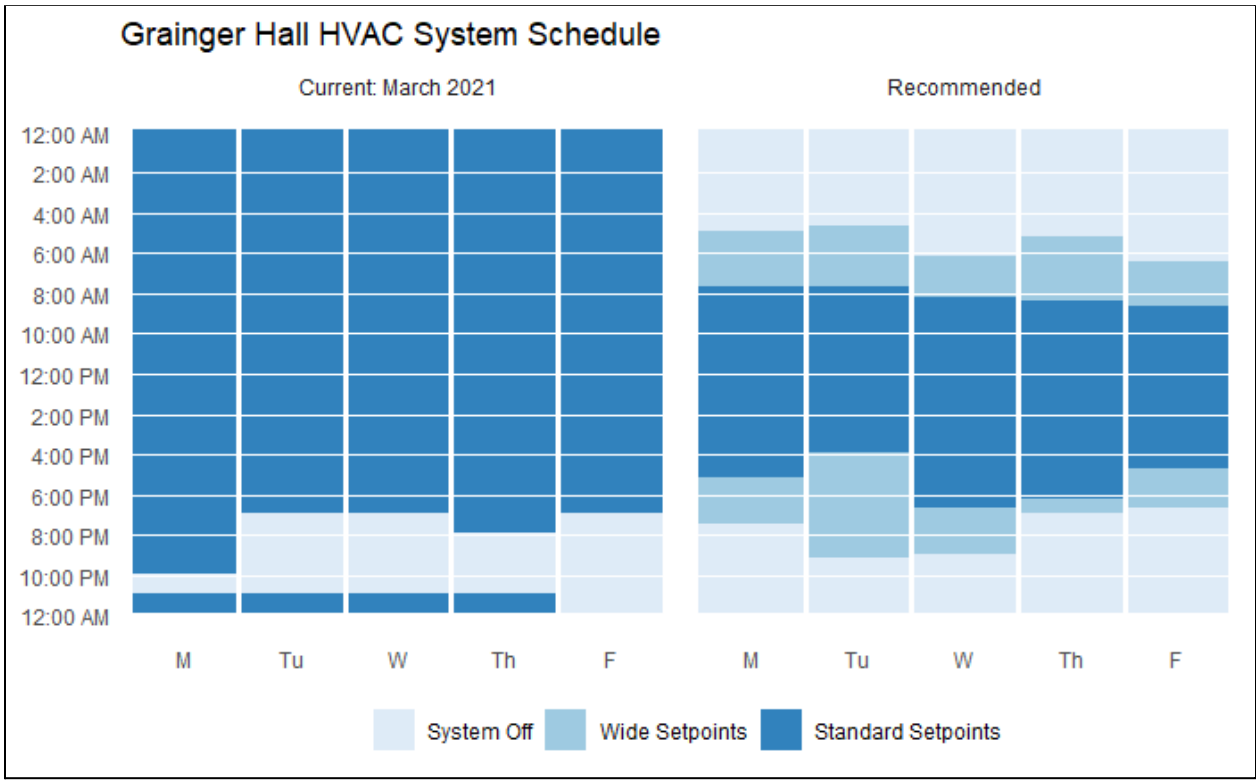


Figure 13. Comparison of the current schedule and the team's recommendations for HVAC operations in Grainger Hall.

During the work week, implementing these recommendations would avoid 31.25 hours of air handler operation and allow wider setpoints to be used for 25.75 hours. A full explanation of the financial and environmental implications of these changes is available in Section 4. These figures ignore weekend days, as the team's algorithms recommend drastic increases in operations on these days that would likely not be implemented by the university.

4. Grainger Environmental Benefit Analysis

Despite Grainger Hall's LEED Platinum Certification, we project that the building can still reduce HVAC energy usage by 23.2%. This reduction can be achieved in two key ways: shutting the air handlers off when occupancy is near zero, and letting the temperature set point wander more when occupancy is low. The first method—turning the air handlers off during times of 5 people or less—can be applied to 31.25 hours during each 5-day weekday period (we focused

our efforts on Monday-Friday, as air handler usage is limited on weekends). Even after accounting for the excess energy demands of re-heating or re-cooling the building after a shutdown period, air handler shutdown is expected to generate 21.8% savings compared with current usage. See the link to the Grainger Environmental Analysis spreadsheet in the Appendix B to see the calculations for this figure.

Permitting a wider set point during low-occupancy hours can produce another bump in energy savings. Currently, Grainger's HVAC system does not wander beyond a tight range of just 1^{O}F around the set point. By permitting this range to expand to two or more degrees, Duke could save energy. To quantify these reductions, we estimated the amount of energy that would be saved by letting the building stay one degree warmer on hot days or one degree cooler on cold days. Our analysis uses 2019 data, which is more complete and typical than 2020 or 2021 figures (as it is unaffected by pandemic occupancy patterns). A full calendar year was used in order to account for the effects of seasonality.

The first step of these calculations involves daily usage of hot water and chilled water in Grainger Hall, standardized to units of kBtu. Next, it involves weather data, finding the average daily temperatures for every day of 2019. Using this temperature data, it is possible to calculate the number of *heating degree days* and *cooling degree days* on each of these days in Durham. One heating degree day represents the number of degrees that average daily outdoor temperature is below 65ºF (EIA, 2020). As such, in the winter, there are many heating degree days in North Carolina, whereas one would expect far fewer in the summer. Instead, there are more cooling degree days in a North Carolina summer. Cooling degree days are the number of degrees by which daily average temperature exceeds 65ºF—meaning that cooling would be needed for indoor spaces (EIA, 2020).

A regression of heating/cooling degree days with heating and cooling energy requirements for Grainger Hall reveals the typical energy intensity of HVAC at different temperatures. In these two regression analyses (one for heating, one for cooling), the slope represents the estimated energy requirement associated with the outdoor temperature being either one degree higher on a hot day or one degree lower on a cold day. See Appendix C for these regressions. Mathematically speaking, this is equivalent to the energy saved by letting the indoor temperature stay one degree higher in hot weather or one degree lower in cold weather. These energy

intensity estimates yield savings of roughly 1.4% from letting the set point wander by one additional degree for 25.75 hours each week. Based on our analysis of Wi-Fi data, there are 25.75 hours per week during which Grainger Hall's occupancy is low (<10) but not so low as to merit a shutdown of the air handlers (<5 people). The reason for this set point adjustment is to save energy while minimizing the impact on occupants' comfort level. Given that there are more than 5 people in the building, we do not want to completely shut off the air handlers but also recognize the opportunity to reduce energy usage.

By combining estimated savings from air handler shutdown and wider set point allowances, we find that our recommendations could reduce HVAC energy usage in Grainger Hall by 23.2% each year. Outlined in Table 1 below are estimates of the financial, energy, and emissions reductions resulting from the aforementioned recommendations in HVAC usage. For the calculations behind these savings estimates, see the spreadsheet linked in the Appendix.

In a given year of our recommended adjustments:		
Cooling savings (kBtu)	1084447.445	per year
Cooling savings ($)	$15,795.29	per year
Heating savings (kBtu)	497096.1091	per year
Heating savings ($)	$9,941.92	per year
Ventilation savings (kBtu)	102249	per year
Ventilation savings ($)	$38,376.16	per year
Energy savings (kBtu)	1683793	per year
Energy savings (%)	23.22%	per year
Financial savings ($)	$64,113.37	per year
Emissions savings (mtCO2e)	66.41	per year
Emissions savings (%)	19.63%	per year

Table 1. Summary of the key findings resulting from making the changes in Grainger Hall HVAC usage outlined in Section 3.7.

5. National Environmental Benefits

5.1 Background

Wi-Fi-based occupancy tracking has the potential to drive massive HVAC energy savings beyond a single building at Duke University. To understand the extent of these savings, our team turned to the Energy Information Administration (EIA) to estimate how many buildings are suited for

Wi-Fi occupancy tracking. The EIA estimates that there are over 5.9 million commercial buildings in the United States, accounting for over 15% of all energy consumed in the country each year (EIA, 2020). Of these several million buildings, nearly all have some form of preexisting Wi-Fi infrastructure that can be leveraged to estimate building occupancy. The EIA has also found that commercial buildings use over 33% of their energy on heating, cooling and ventilation. Ventilation represents the largest energy consumption of all HVAC functions. As seen in Figure 14, the subset of buildings that could potentially use a Wi-Fi based system for optimizing HVAC performance faced a combined $8.6 billion in electricity costs for ventilation alone in 2012 (EIA, 2020). These conditions create a tremendous opportunity to capture significant and widespread cost and energy savings if Wi-Fi based occupancy estimation can be reliably used to optimize commercial HVAC systems.

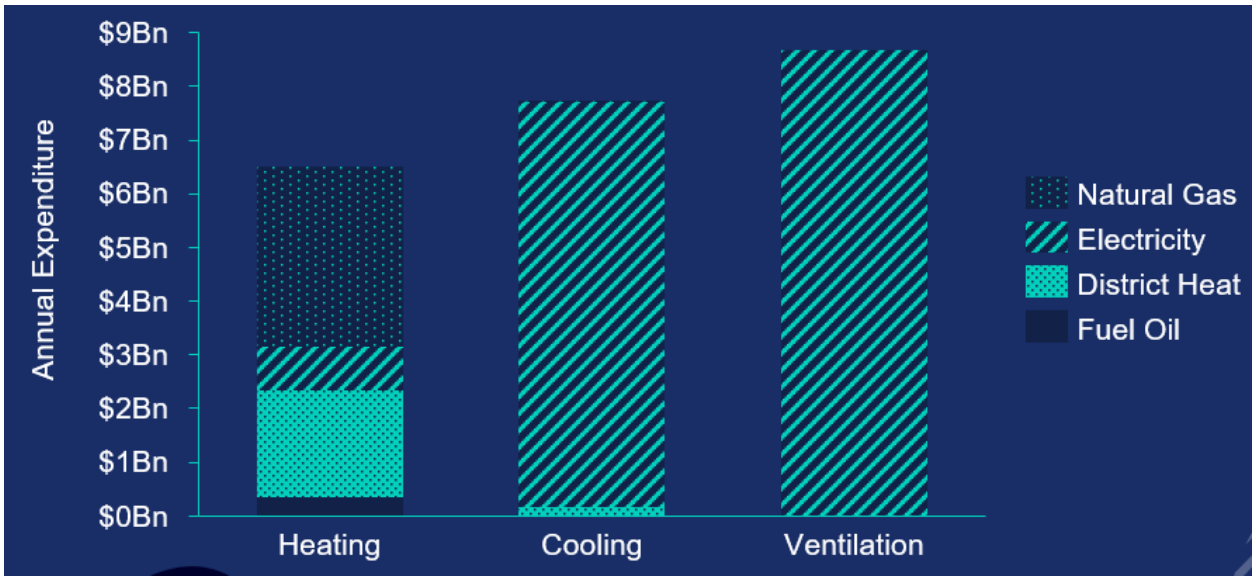


Figure 14. HVAC energy expense for buildings included in market sizing (EIA, 2012).

5.2 Data

In order to size the market and estimate the potential energy and cost savings that our Wi-Fi-based occupancy estimation could produce if used in commercial buildings throughout the United States, our team used data from the Energy Information Administration's 2012 Commercial Buildings Energy Consumption Survey (EIA, 2012). The survey sampled over 6,700 commercial buildings throughout the country, and includes data on the type or use of the

building, its age, size, energy use and several other characteristics. The survey then weighted all data recorded from survey respondents so that it is at the national scale (EIA, 2012).

5.3 Methodology

Leveraging the extensive EIA dataset, our team first reduced the data down to a subset of buildings that reasonably aligned with the building upon which we tested the Wi-Fi-based estimation technology, and that had use cases we deemed most suitable for the technology. For our purposes, the team chose to filter the buildings to include only those that have a BAS, like buildings on Duke University's campus do, for the sake of consistency. Having eliminated all buildings that do not have a BAS, we can reasonably expect to see similar savings in these commercial buildings to those observed in buildings on campus. Next, our team filtered the dataset by the building-use criteria to include only commercial buildings that we felt were best suited to use this technology, specifically those that met the criteria described in Section 5.4. Having reduced the EIA dataset to only include survey respondents meeting the specified criteria, our team used a simple equation to calculate potential energy consumption and expenditure savings at a national scale. Each survey respondent's energy consumption and expenditure, broken down by energy source (e.g, Natural Gas, Electricity, District Heat, and Fuel Oil), were multiplied by the savings percentage for the given scenario and summed to obtain national estimates of total energy consumption and expenditure savings.

5.4 Criteria

5.4.1 Non-Unique Needs

Certain types of buildings especially outside of Duke (e.g., hospitals, research labs, and industrial kitchens) require special ventilation needs in order to meet certain industry standards. In hospitals and research labs, a certain level of consistent ventilation is required to prevent the spread of infectious air molecules or toxic substances. In kitchens, ventilation similarly runs high to carry out smoke and fumes from cooking. Because of this, there are fewer opportunities to reduce HVAC supply when these institutions require constant ventilation, regardless of occupancy. Thus, these types of buildings are not prime candidates for HVAC reduction through occupancy tracking.

5.4.2 Unpredictability

Like with most occupancy-tracking methods, a varying building schedule is the crux of occupancy-tracking usefulness. Buildings with dynamic, unpredictable occupancy have more applications for Wi-Fi occupancy tracking.

5.4.3 Localized Zones

While occupancy (or the lack thereof) can help determine when to turn off the entire building's HVAC system, the opportunities to do so may be rare for busy buildings. Tracking occupancy becomes more useful if the building can be broken down into specific zones. Within these zones, temperature can be adjusted independently, leading to more localized HVAC changes and greater energy savings in the long run. Buildings with larger gathering spaces are ideal as they often sit unused and can lead to greater savings with fewer zone adjustments–as opposed to keeping track of many small rooms at once. Buildings with auditoriums and large conference rooms, for example, are ideal candidates.

Based on the three criteria – non-unique needs, unpredictability, and localized zones – we generated a list of building types best suited for Wi-Fi occupancy tracking: K-12 schools, churches, hotels, fitness studios, theaters, conference centers, office buildings, and retail spaces.

5.5 Calculations

After filtering the data for these building types, we calculated the total energy expenditure ($) and consumption (kBtu) by energy source and end use for the remaining buildings. An external report prepared for the U.S. Department of Energy, which analyzed potential energy savings for occupancy-based control of variable-air-volume systems, calculated that one can expect to see from ~6% to ~18% energy savings in office buildings like Grainger Hall (Pacific Northwest National Laboratory et al., 2013). Based on these numbers, we prepared a base, bull, and bear case of 5%, 10%, and 20% savings to estimate the total national energy and cost savings that could be generated through Wi-Fi-based occupancy tracking, presented below in Figure 10.

Energy Consumption and Expenditure Savings by Source												
	5% savings				10% savings				20% savings			
Source	Consumption (kBtu)		Expenditure ($)		Consumption (kBtu)		Expenditure ($)		Consumption (kBtu)		Expenditure ($)	
Natural Gas	22,039,215,940	$	168,278,339.75		44,078,431,881	$	336,556,679.50		88,156,863,762	$	673,113,359.01	
Electricity	30,288,490,769	$	851,873,282.01		60,576,981,538	$	1,703,746,564.03		121,153,963,076	$	3,407,493,128.05	
District Heat	6,182,151,158	$	107,792,539.40		12,364,302,316	$	215,585,078.80		24,728,604,631	$	431,170,157.59	
Fuel Oil	741,302,993	$	17,665,796.49		1,482,605,987	$	35,331,592.98		2,965,211,973	$	70,663,185.96	
Total	59,251,160,860	$	1,145,609,957.65		118,502,321,721	$	2,291,219,915.31		237,004,643,441	$	4,582,439,830.62	

Table 2. HVAC-related energy & expenditure savings for base, bull & bear scenarios (EIA, 2012).

Lastly, it is important to note that a limitation of using this data set to size the market is that because it is survey data, it is not perfectly random and therefore is not perfectly representative of all commercial buildings in the United States.

5.6 Calculations in Context

Translating this to an equivalency in environmental benefits would provide more clarity on the potential environmental impact.

Recall that the energy consumption by end use estimation was broken down into categories of natural gas, electricity, district heat, and fuel oil. These sources have varying carbon contents, so estimating total carbon emissions savings requires a slightly different approach per source.

First, natural gas has a carbon intensity of 0.0053 mtCO2 per therm (EPA, 2019), so one can derive the metric tons of carbon dioxide savings using the following conversion factor:

$$\left(\frac{1\ mmBtu}{1000\ kBtu}\right) * \left(\frac{1\ therm}{0.1\ mmBtu}\right) * \left(\frac{0.0053\ mtCO2}{1\ therm}\right)$$

5% Savings	20% Savings
2,039,215,940 kBtu → 1.17 million mtCO2	88,156,863,762 kBtu → 4.67 million mtCO2

Table 3. Natural Gas savings under 5% and 20% scenarios.

Next, we can look at electricity carbon intensity. This varies depending on the local utility's energy mix, so we used the 2019 CO2 emissions intensity in pounds per kWh from Duke Energy as a proxy (Duke Energy, 2019).

$$\left(\frac{1\ kWh}{3.412\ kBtu}\right) * \left(\frac{0.86\ lb\ CO2}{1\ kWh}\right) * \left(\frac{1\ mtCO2}{2204.6\ lbs}\right)$$

5% Savings	20% Savings
30, 288, 490, 769kBtu → 3.46 million mtCO2	121, 153, 963, 076kBtu → 13.85 million mtCO2

Table 4. Electricity savings under 5% and 20% scenarios.

District heat is a subcategory of steam plant operations that specifically addresses heating of buildings. Duke's steam plants run on natural gas. This district heat energy usage, however, is calculated as a separate category from natural gas. Thus, we will use the same carbon intensity from the previous natural gas calculation (EPA, 2019).

$$\left(\frac{1\ mmBtu}{1000\ kBtu}\right) * \left(\frac{1\ therm}{0.1\ mmBtu}\right) * \left(\frac{0.0053\ mtCO2}{1\ therm}\right)$$

5% Savings	20% Savings
6, 182, 151, 158kBtu → 328,000 mtCO2	24, 728, 604, 632kBtu → 1.31 million mtCO2

Table 5. District heat savings under 5% and 20% scenarios.

Finally, we can address the carbon savings from fuel oil (EPA, 2019).

$$\left(\frac{1\ barrel}{6287\ kBtu}\right) * \left(\frac{429.61\ kg\ CO2}{1\ barrel}\right) * \left(\frac{1\ mtCO2}{1000\ kg}\right)$$

5% Savings	20% Savings
$741,302,993$kBtu → 50,000 mtCO2	$11,860,847,888$kBtu → 810,000 mtCO2

Table 6. Fuel oil savings under 5% and 20% scenarios.

Summing these carbon savings mitigates approximately 5,000,000 mtCO2 in the 5% energy savings scenario and 20,600,000 mtCO2 in the 20% scenario across applicable commercial buildings in the United States. The 5% scenario is equivalent to 1 million passenger vehicles being taken off the road for 1 year, or for better visualization, if the state of Nevada were to stop driving for a year (Statistica, 2021). The 20% scenario is approximately equivalent to 4.5 million passenger vehicles, or automobile registrations in the state of New York.

6. Business Case

6.1 Comparative Cost of Implementation

Compared to other occupancy sensors, Wi-Fi tracking is fairly inexpensive. The chart below outlines the different costs of installing each different sensor in a 100,000 sq. foot space with ~110 independent rooms like Grainger Hall, as well as the potential energy savings associated with each method. This number of rooms in Grainger Hall was obtained by analyzing the floor plans provided by Casey Collins (Appendix D). These calculations also assume that, like Grainger Hall, the building has existing Wi-Fi infrastructure.

The cost of labor for sifting through and sorting Wi-Fi data was calculated using an estimation of the number of hours one would expect to spend setting up a similar analysis. We reached a number of 40 hours using our own estimation, conversations with IT professionals, and the assumption that future users would have a code and set of instructions. We also used the average salary of a data engineer, $93,000, and broke down the hourly compensation based on a 48 week/year, 50 hour/week work year for a rate of $38.75 per hour (Payscale, 2021). Finally, we assumed that people adopting this method of occupancy tracking already have existing Wi-Fi infrastructure, making fixed costs obsolete.

	PIR	CO2	Wi-Fi
Fixed Costs per Unit	$100 (Fixr, 2021)	NA	NA
Labor per Unit	$163 (Perfect Partnership, 2020)	NA	NA
Total Unit Cost	$263	NA	$1,550
Total Cost per 110 units or 100,000 sqft	$28,930	$233,000 (Lafond, 2019)	$1,550
Potential Savings	5.9% (Snyder, 2019)	9% (Florida Power & Light)	5-20%

Table 7. Occupancy tracker cost and energy savings comparison.

With many tracking methods, the single price above may not be the only cost. Sensors like CO2 and PIR are often paired with each other or other sensors for higher accuracy (Naylor et al., 2018, 1). As a result, these prices are likely just a fraction of what one would actually pay. Wi-Fi tracking, however, achieves an accuracy which would not require back-up sensors (Mohammadmoradi et al., 2017, 7).

It is also important to keep in mind that the potential savings of methods like PIR and CO2 are those advertised under perfect use-cases with limited false triggering and no delayed functioning. This report's Wi-Fi tracking calculations, however, consider a range and even estimate a case study of savings as high as 23.2%, and external researchers cast a net of 5-20% (4. Grainger Hall Environmental Benefits; Pacific Northwest National Laboratory et al., 2013, 29).

This wide net accounts for two unique factors that influenced our calculations. Firstly, the data used in this report was recorded during COVID-19, a time when the majority of Duke University is operating through online platforms. This means that building occupancy is likely lower than it would be during normal functioning, and thus there were more opportunities to turn the air

handler completely off. Secondly, we also note that Grainger Hall is a LEED Platinum Certified building, meaning fewer opportunities to make efficiency gains since the building already takes many precautions to preserve energy. While we are unable to quantify the extent to which this affected our calculations, they are important to keep in mind when applying our method to future applications.

6.2 Government Incentives

To offer more context on the costs and benefits on a national scale, it is valuable to know government and local incentives exist to mitigate some of the cost of energy efficiency. The federal Energy Policy Act of 2005 established a tax deduction for energy efficiency upgrades in commercial buildings. This tax deduction of $1.80 per sq. ft is available to owners of buildings that address lighting, building envelope, or heating, cooling, ventilation, and hot water systems that reduce the building's energy and power cost by 50% or more (DSIRE, 2021).

As an example, a commercial building the size of Grainger Hall would be able to deduct $126,000 from their taxes if they were able to meet 50% energy use reductions (Duke University, 2015).

A more local incentive is the NC Financing Program for Renewable Energy and Energy Efficiency. This legislation authorizes cities and counties to establish revolving loan programs to finance renewable energy and energy efficiency projects. Effectively, it offers loans at low interest (no more than 8%) with loan terms limited to 20 years. Cities and counties are also able to secure funding through federal Energy Efficiency and Conservation Block Grants and other unrestricted revenue sources (DSIRE, 2021).

These incentives exist mainly to offset the cost of expensive infrastructure or technology upgrades. However, our Wi-Fi analysis may not even require this. Given the cost of $1,550 in labor to implement this Wi-Fi analysis, it would only require incredibly low savings of 30,202 kBtu (or 0.35%) in Grainger Hall to break even with costs.

7. Social Benefits

Beyond money saved and emissions avoided, team analysis showed that using Wi-Fi to optimize HVAC usage can bring a number of social benefits. This category includes everything from convenience to comfort to learning, illustrating clear benefits to society that do not fully fit in another bucket.

7.1 Improved Comfort

Imagine Duke University's Cameron Indoor Stadium before the pandemic. At every basketball game, students and fans would pack the courtside, jumping around and cheering, generating significant body heat. In fact, according to Professor Tim Johnson, one person at rest gives off an average of 350 Btu per hour (Johnson, 2021). When combined with 9,000 other people in the tiny basketball stadium, the temperature may rise to uncomfortable levels. Accordingly, Duke FMD has to manually pre-cool Cameron before games. This means that a sudden influx of people raises the temperature back up to a normal level without leaving everyone sweaty and hot.

Stadiums are not the only spaces that can become uncomfortable due to accumulating body heat. Classrooms, stores, and eateries are all examples of venues that can see sudden spikes in occupancy, so it would be more efficient to prepare for such swings and preemptively cool rooms. A program that utilizes Wi-Fi data can automatically do just that. By looking at historical trends of occupancy, as estimated by the number of devices that associate with a local Wi-Fi access point, the system can predict the times when it should preemptively begin cooling. This keeps visitors happy and comfortable, even at rush hour.

7.2 Smoother Demand Management

When spaces are suddenly flooded with people, the HVAC system has no choice but to work at full strength in order to shift the area back to a normal temperature. In many cases, this means that the system draws a large amount of electricity from the local grid at once, increasing the strain and need for instantaneous generation to meet the demand. While the buyer of this electricity may not pay any more for this sudden draw of electricity than they would for the same

amount of power over a longer time period, this spike in demand has costs to the broader electric grid (Tanaka, 2006).

The fundamental difference between the electric sector and other markets is that in electricity, supply must meet demand precisely, at all times. Because electricity storage is not yet economical, the need to suddenly cool a room means that a power plant somewhere is burning more coal or natural gas. Ramping power plants up quickly is expensive, and this cost is passed on to customers. This means that major users of electricity, such as universities, malls, and offices, may inadvertently be raising electricity prices across the board by making heavy, sudden draws on the grid (Tanaka, 2006). As depicted in the graph below of a typical generation stack in the Mid-Atlantic region, generating electricity at peak demand is multiple times more expensive than base load levels, and typically relies on high-emissions power plants such as natural gas, coal, and oil (Figure 15).

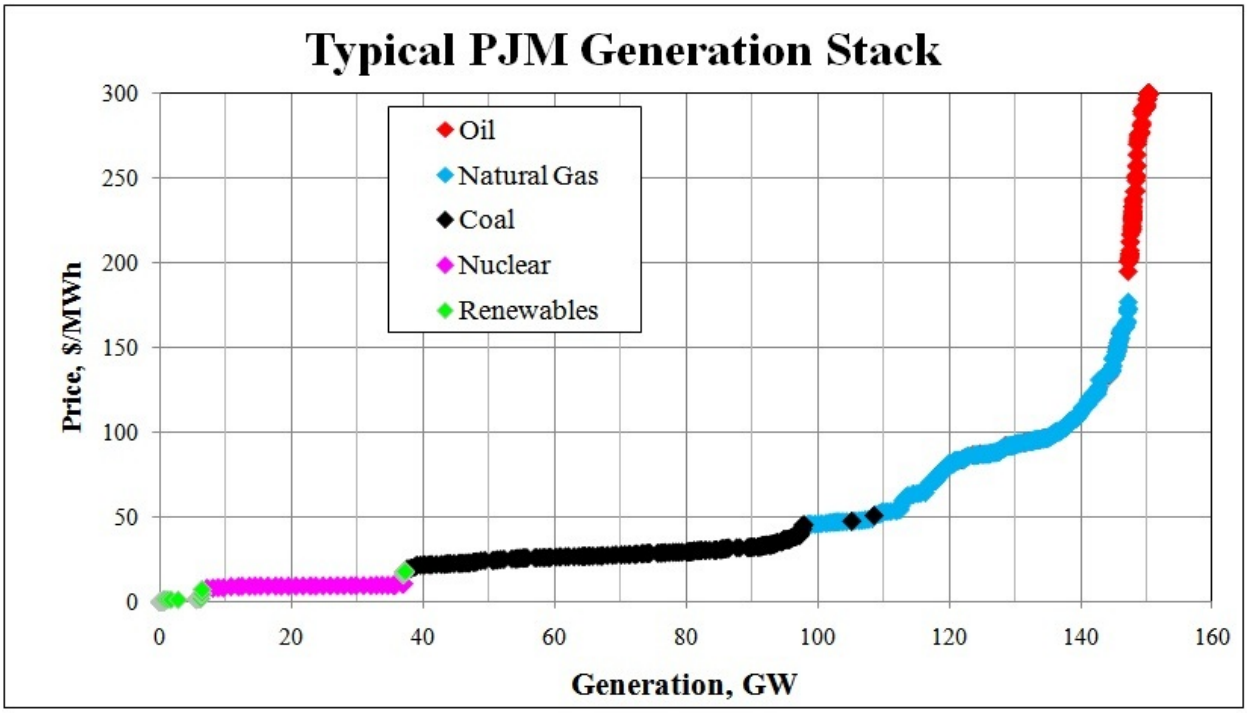


Figure 15. Typical utility electricity generation stack for Pennsylvania, New Jersey, and Maryland, 2008 (Posner, 2017).

Using a Wi-Fi-based adaptive system can help to combat this problem. By predicting high-occupancy times, the system can draw electricity from the grid over a longer duration of time, thereby reducing the need for a sudden spike in generation. This benefit is expected to grow even greater in the future as the generation mix across the country shifts towards non-dispatchable renewable sources of energy, notably wind and solar. Electric systems with fewer dispatchable energy sources are less resilient in the face of shocks in demand, so adaptable management systems will be a necessity for a more sustainable energy future. Pictured below is a graph of average daily electricity demand in California over an eight-year period during which installed residential rooftop solar capacity grew dramatically (Figure 16). It clearly shows that as sector-wide electricity demand hits its peak in the evening, solar generation capacity is diminishing for the day, leading to a sudden need to generate electricity from other sources. Often, these quickly dispatchable sources are expensive, fossil fuel-based power plants, as discussed above (Blumsack, 2018).

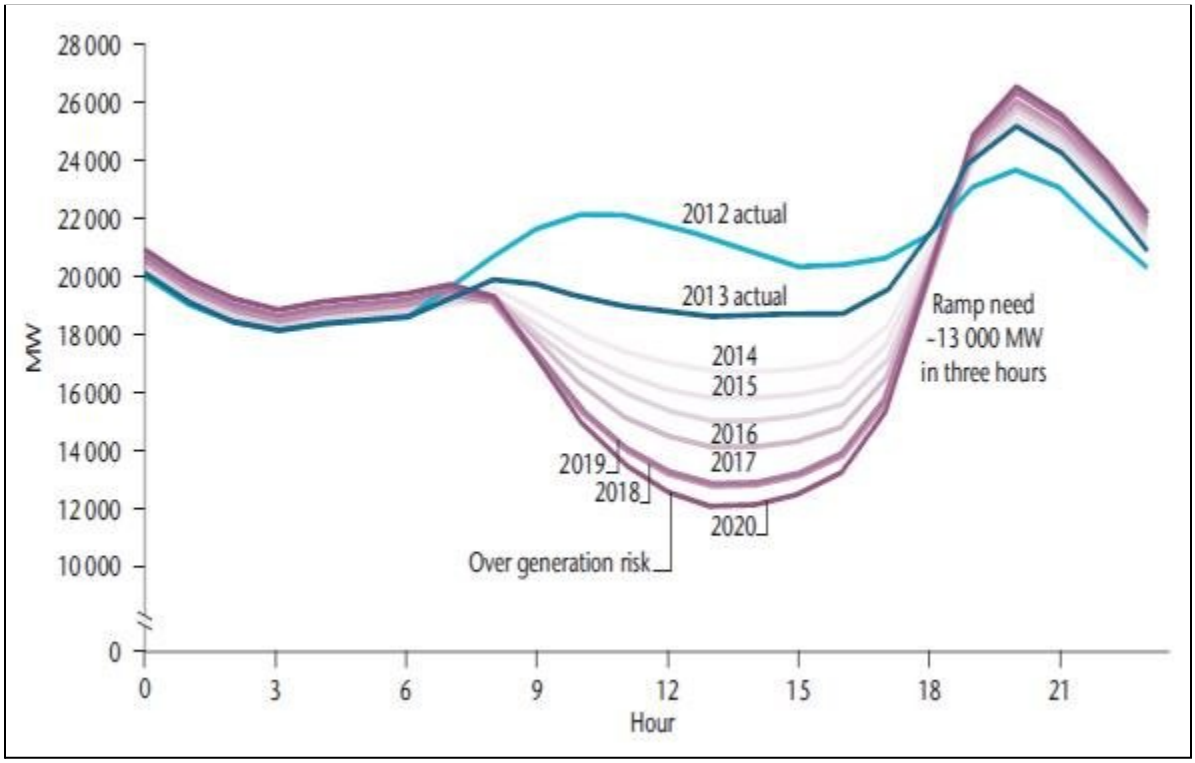


Figure 16. California daily utility electricity demand curves, 2012-2020 (Fraas & Partain, 2015).

7.3 Educational Tool

Finally, future students interested in energy can learn from and build upon our research from this year. The work from a Duke University 2020 Data+ team – "Network Visualization of Foot Traffic Patterns" – which pioneered Wi-Fi data as a proxy for building occupancy in the Bryan Center, was tremendously helpful this year (Duke University, 2020). We hope to pay it forward. There is a lot of potential for efficiency gains in HVAC on Duke's campus; the 2019 Duke Climate Action Plan's first recommendation on energy was HVAC optimization. Student researchers can help drive some of those key improvements, but years of accumulated knowledge are required to make a difference.

For that reason, everything we learned and developed will be available to future research groups. All code will be posted on GitHub, from which interested students can download and apply the tools we developed. We have also documented the unique challenges that we faced throughout this process, as well as the best responses to those obstacles. In particular, accessing confidential data was a hurdle that future students should be aware of. This can help them to learn from our experiences, and hopefully, those groups can continue to make important progress.

8. Conclusion

8.1 Limitations

Like other occupancy-tracking methods, Wi-Fi has some limitations, despite its better performance overall. For example, it assumes that everyone inside the building is carrying a working smartphone or other device. This may overlook a large demographic that may not own a smartphone, leaves their phone at home or keeps it fully shut down most of the time.

Collecting occupancy data via Wi-Fi also presents some privacy concerns, as noted above. The associations can be traced to a unique user, meaning that their behaviors and activities are recorded and could be potentially traced. While the data can be filtered so that these associations are anonymous, the raw data would be preserved for longer than its normal duration (Thompson, 2020). Despite these limitations, Wi-Fi demonstrates promise in its ability to track occupancy, as well as its applications for future use.

8.2 Areas for Improvement

Due to time constraints, we were unable to achieve precision to the degree we would have liked. While our algorithms were informed by data analysis and experts at Duke OIT, we were forced to make some assumptions that reduce the certainty of our conclusions. Future work on this project may include refining the parameters surrounding what qualifies as a single person when they may have multiple devices or are moving through the building and connecting with multiple access points. This work may include physically walking through a building and mapping access points to physical locations to be better able to understand how patterns of Wi-Fi associations correlate with certain movements or behaviors.

8.3 Future Applications

Other applications for Wi-Fi-based occupancy tracking are numerous. For example, researchers could develop a program that optimizes HVAC usage based on real-time Wi-Fi data, rather than historical trends. There are several technical challenges in implementing this idea–it would require a constant, two-way flow of information and constant data analysis–but it certainly holds potential for the future. With more precise measurements and greater fluency between Wi-Fi and HVAC data, real-time occupancy adjustments to heating, cooling, and ventilation could optimize efficiency better even than our proposed system.

Future researchers should also look into machine learning as a means of system optimization. As a field, machine learning holds immense potential to reshape anything it is applied to, and we expect it to play a major role in energy efficiency over the coming decades. By taking cases of days that best model typical days of the week and continuously feeding in new data as they are passed over by OIT (in a hypothetical future), a “living document” set of recommendations could be refined further and further in order to maximize efficiency of the system due to the mostly habitual nature of life on campus. This could be translated into a commercial setting through a very similar approach and would only require basic modifications related to how the building is controlled and how large each of the ventilated spaces are. Seeing how well this information is compiled in our case, it does not require much imagination to make the logical leap to doing this with nearly any building that has an integrated Wi-Fi system.

Teams with more experience in machine learning and sufficient processing power should take this next step. In the past few years, researchers have increasingly applied machine learning to building optimization, with growing promise (Eini et al., 2021). We believe that as future research groups look to involve themselves in energy efficiency on campus, they should certainly consider pursuing this route. In fact, a Duke Data+ team explored using machine learning as a means of forecasting energy demand on campus last year (Swartzendruber et al., 2021). They have published this tool on GitHub, available to the public, which could serve as a foundation for HVAC Wi-Fi machine learning projects in the future, along with our findings (Swartzendruber et al., 2021). Their projections are included below (Swartzendruber et al., 2021).

Another desirable tool that this project did not pursue due to time and resource constraints is a real time dashboard that would allow students and campus officials to check in on resource consumption of buildings such as Grainger Hall. This would incorporate much of the existing progress outlined in this report and consolidating it onto a Duke-hosted webpage. This sort of transparency about consumption could assist students in knowing how their individual choices may affect building energy use.

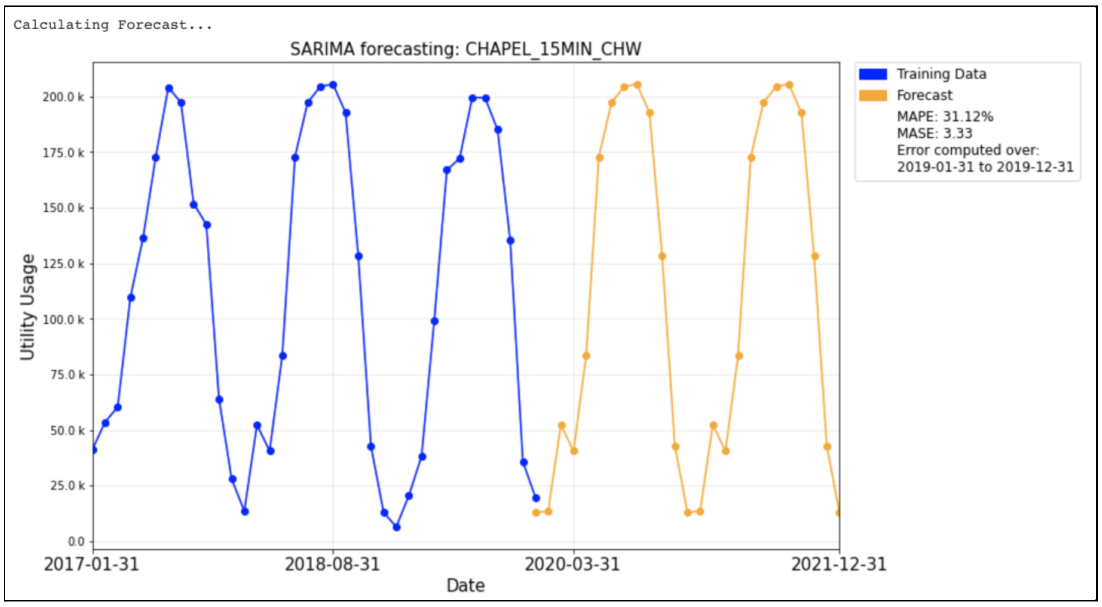


Figure 17. Duke Chapel utility electricity usage per year, 2017-2021 (Swartzendruber et al., 2021).

8.4 Discussion

According to this report's calculations, Grainger Hall has the potential to generate a 23.2% decrease in energy use by using Wi-Fi occupancy data to make HVAC operation adjustments. Though Wi-FI occupancy tracking comes with some challenges, it has the potential to generate ~237 billion kBTU in savings at an extremely low cost in applicable buildings across the United States with existing Wi-Fi infrastructure. Through future research, like developing localized HVAC adjustments and employing machine learning, total savings could increase even more.

Those interested in replicating the research in this report should attempt to make real HVAC adjustments based on Wi-Fi occupancy data rather than just models. Generating a real-life case study in W-Fi occupancy tracking will be integral to achieving widespread adoption and its many benefits.

Bibliography

Arief-Ang, I. B., Salim, F. D., & Hamilton, M. (2017, June). CD-HOC Indoor Human Occupancy Counting Using Carbon Dioxide Sensor Data. *ArXiv*, *NA*(NA), 1-24. Research Gate. NA

Blumsack, S. (2018). *Basic economics of power generation, transmission and distribution | EME 801: Energy Markets, Policy, and Regulation*. Psu.edu; Penn State Department of Energy and Mineral Engineering. https://www.e-education.psu.edu/eme801/node/530

Consulting Specifying Engineer Staff. (1970). Dehumidification Alternatives for Commercial Buildings. *Consulting Specifying Engineer*. Retrieved from https://www.csemag.com/articles/dehumidification-alternatives-for-commercial-buildings

Creston. (2017, October NA). *Occupancy Sensor Placement and Technology*. Creston. Retrieved April 4, 2021, from https://www.crestron.com/getmedia/4114baf7-8de2-41bd-92bb-c45214a36571/mg_bp_occupancy_sensor_placement_technology

Degree Days - Handle with Care! (n.d.). Www.energylens.com; BizEE. https://www.energylens.com/articles/degree-days

Degree-days - U.S. Energy Information Administration (EIA). (2020, August 17). Www.eia.gov; EIA. https://www.eia.gov/energyexplained/units-and-calculators/degree-days.php

Dilouie, C. (2017, August 21). *All About Occupancy and Vacancy Sensors*. Lighting Control Association. Retrieved April 4, 2021, from https://lightingcontrolsassociation.org/2017/08/21/all-about-occupancy-and-vacancy-sensors/

DSIRE. (2021, January 27). *Energy-Efficient Commercial Buildings Tax Deduction*. NC Clean Energy Technology Center. Retrieved April 4, 2021, from https://programs.dsireusa.org/system/program/detail/1271/energy-efficient-commercial-buildings-tax-deduction

Duke Energy. (2019). *Environmental Performance Metrics*. https://sustainabilityreport.duke-energy.com/operations/environmental-performance-metrics/

Duke University. (n.d.). Duke Facilities: Utility Services. Retrieved April 5, 2021, from https://facilities.duke.edu/services/utilities/systems

Duke University. (2015, October 25). *Duke University's Grainger Hall Receives LEED Platinum Certification*. Nicholas school of the Environment. Retrieved April 4, 2021, from https://nicholas.duke.edu/news/duke-universitys-grainger-hall-receives-leed-platinum-certification#:~:text=DURHAM%2C%20N.C.%20%E2%80%93%20Grainger%20Hall%2C,the%20U.S.%20Green%20Building%20Council

Duke University. (2020, NA NA). *Network Visualization of Foot Traffic Patterns*. Data+ (2020). Retrieved April 4, 2021, from https://bassconnections.duke.edu/project-teams/data-2020

EIA. (2012, NA NA). *Commercial Building Energy Consumption Survey*. EIA Independent Statistics & Analysis. Retrieved April 4, 2021, from https://www.eia.gov/consumption/commercial/data/2012/

EIA. (2015, May 08). Average size of new commercial buildings in theUnited States continues to grow. U.S. Energy Information Administration. Retrieved 4 26, 2021, from https://www.eia.gov/todayinenergy/detail.php?id=21152

EIA. (2017, July NA). EIA. Retrieved April 4, 2021, from https://www.eia.gov/outlooks/aeo/assumptions/pdf/0554(2017).pdf

Eini, R., Linkous, L., Zohrabi, N., & Abdelwahed, S. (2021). Smart building management system: Performance specifications and design requirements. *Journal of Building Engineering*, *39*, 102222. https://doi.org/10.1016/j.jobe.2021.102222

EPA. (2019). *Greenhouse Gases Equivalencies Calculator — Calculations and References*. https://www.epa.gov/energy/greenhouse-gases-equivalencies-calculator-calculations-and-references

EPA. (2020, March). *Greenhouse Gas Equivalencies Calculator*. https://www.epa.gov/energy/greenhouse-gas-equivalencies-calculator

Evans, P. (2019). Variable Air Volume – VAV. Retrieved April 4, 2021, from The Engineering Mindset.com website: https://theengineeringmindset.com/variable-air-volume/

Farsetta, J. (n.d.). Commercial Steam Boilers: A Primer. Retrieved April 4, 2021, from Certified Commercial Property Inspectors Association website: https://ccpia.org/commercial-steam-boilers-a-primer/

FIXR. (2021). *How Much Does It Cost to Install Motion Sensor Lights?* FiXR. Retrieved April 25, 2021, from https://www.fixr.com/costs/outdoor-motion-sensor-lights-installation#:~:text=Area%20Reflective%20Motion%20Sensor&text=These%20sensors%20scan%20large%20areas,price%20of%20%2440%20to%20%24100.

Florida Power & Light. (NA, NA NA). *Demand Control Ventilation: An FPL Technical Primer*. FPL. Retrieved April 19, 2021, from https://www.fpl.com/content/dam/fpl/us/en/business/save/programs/pdf/dcv-primer

Fraas, L., & Partain, L. (2015, June). *Displacing California's Coal and Nuclear Generation with Solar PV and Wind By 2022 Using Vehicle-To-Grid Energy Storage*. ResearchGate. https://www.researchgate.net/figure/California-daily-utility-electricity-demand-curves_fig2_278848880

Johnson, T. (2021, April 2). *Tim Johnson - EE HVAC Call* (L. Ives, T. Hessel, E. Lamb, F. Picone, I. Jiang, B. Williams, & J. Kochansky, Interviewers) [Personal communication].

Lafond, A. (2019, August 14). *What Is the Cost of Demand Control Ventilation? Using Occupancy Sensors and More to Optimize Ventilation Airflow*. Foobot. Retrieved April 4, 2020, from https://foobot.io/resources/demand-control-ventilation-cost/

Longo, E. A., Redondi, A., & Cesana, M. (2019, November). Accurate occupancy estimation with Wi-Fi and bluetooth/BLE packet capture. *Computer Networks*, *163*(2019), 1-9. ScienceDirect. https://doi.org/10.1016/j.comnet.2019.106876

Mid-Atlantic Controls. (2017). What Is a Building Automation System? Retrieved April 4, 2021, from https://info.midatlanticcontrols.com/blog/what-is-a-building-automation-system

Mohammadmoradi, H., Yin, S., & Gnawali, O. (2017, July). Room occupancy estimation through Wi-Fi, UWB, and light sensors mounted on doorways. *ICSDE*, *NA*(July 2017), 27- 34. ACM DL. https://doi.org/10.1145/3128128.3128133

Naylor, S., Gillott, M., & Lau, T. (2018, November). A review of occupant-centric building control strategies to reduce building energy use. *Renewable and Sustainable Energy Reviews*, *96*(November 2018), 1-10. Elsevier. https://doi.org/10.1016/j.rser.2018.07.019

Pacific Northwest National Laboratory, Zhang, J., Liu, G., Lutes, R., & Brambley, M. (2013). *Energy Savings for Occupancy- Based Control (OBC) of Variable- Air-Volume (VAV) Systems*. PNNL. https://www.pnnl.gov/main/publications/external/technical_reports/pnnl-22072.pdf

Payscale. (2021, NA NA). *Average Data Engineer Salary*. Payscale. Retrieved April 4, 2021, from https://www.payscale.com/research/US/Job=Data_Engineer/Salary

Perfect Partnership. (2020, NA NA). *Cost of Install a Motion Sensing Switch*. DIY or Not. Retrieved April 4, 2021, from https://diyornot.com/Project.aspx?ndx2=4&Rcd=240

Posner, B. (n.d.). *The Fundamentals of Electricity Markets | EBF 200: Introduction to Energy and Earth Sciences Economics*. Www.e-Education.psu.edu. Retrieved April 4, 2021, from https://www.e-education.psu.edu/ebf200/node/151

Schramm, S. (2020). Providing a cooler future for Duke. *Duke Today*. Retrieved from https://today.duke.edu/2020/09/providing-cooler-future-duke

Simma, K. C., Mammoli, A., & Bogus, S. M. (2019). Real-Time Occupancy Estimation Using Wi-Fi Network to Optimize HVAC Operation. *Procedia Computer Science*, *155*(2019), 495-502. Elsevier. https://doi.org/10.1016/j.procs.2019.08.069

Snyder, J., & Lighting Energy Alliance. (2019, June 25). *Potential opportunities to reduce HVAC energy using lighting sensors in commercial buildings*. Lighting HVAC: Lighting

Research Center. Retrieved April 19, 2021, from https://www.lrc.rpi.edu/programs/energy/pdf/Lighting_HVAC_report_2019-9-30.pdf

Statistica. (2021, March). *U.S. automobile registrations in 2019, by state*. https://www.statista.com/statistics/196010/total-number-of-registered-automobiles-in-the-us-by-state/

Sustainable Duke. (2019, April 1). *2019 Duke University Climate Action Plan Update*. Sustainability Duke. Retrieved April 4, 2021, from https://sustainability.duke.edu/sites/default/files/2019capupdate.pdf.

Swartzendruber, E., Llewellyn, G., & Takeshima, S. (2021, April 3). *Forecasting Duke University Utility Usage*. Rhodes Information Initiative at Duke. https://bigdata.duke.edu/projects/forecasting-campus-energy-usage-improved-energy-management

Tanaka, M. (2006). Real-time pricing with ramping costs: A new approach to managing a steep change in electricity demand. *Energy Policy*, *34*(18), 3634–3643. https://doi.org/10.1016/j.enpol.2005.07.012

Thompson, M. (2020, December 3). *Wi-Fi Walkthrough Interview.* (L. Ives, T. Hessel, E. Lamb, F. Picone, I. Jiang, B. Williams, & J. Kochansky, Interviewers) [Personal communication].

Trane Technologies. (n.d.). Trane Commercial HVAC: Chillers. Retrieved April 5, 2021, from https://www.trane.com/commercial/north-america/us/en/products-systems/chillers.html

US Department of Energy. (2017, December NA). *Energy Savings Potential and RD&D Opportunities for Commercial Building HVAC Systems*. Office of Energy Efficiency & Renewable Energy. Retrieved April 4, 2021, from https://www.energy.gov/sites/prod/files/2017/12/f46/bto-DOE-Comm-HVAC-Report-12-21-17.pdf

Yun, J., & Lee, S.-S. (2014, May 5). Human Movement Detection and Identification Using Pyroelectric Infrared Sensors. *Sensors*, *201*(May 2014), 8057-8081. NCBI. 10.3390/s140508057

Appendix

Appendix A: Occupancy Estimation Functions

```python
def CreateTrips(dfData,lapseHours = 2):
    """ Function that adds columns with distrinct "trip numbers" and connection times to each association.
    Each trip number denotes a set of association from the same MAC address where each association happens within lapseHours hours of another
    association from the same device. Adds connection times column, which gives the time until the next association from the same MAC address.
    Adds isLast column to the  dataframe to indicate when no next association exists (connection times need to be created from estimates).
    """

    #defines a length of time after which a new association is considered a 'new visit'
    lapseTime = datetime.timedelta(hours = lapseHours)

    lastSeen = {}
    deviceTripNum = {}

    isLast = []
    connectionTime = []
    tripNum = []

    tripCounter = 0

    for i in reversed(range(len(dfData))):
        if dfData['macaddr_hashed'].iloc[i] in lastSeen:
            timeSince = -(dfData['dateTime'].iloc[i] - lastSeen[dfData['macaddr_hashed'].iloc[i]])

            if timeSince > lapseTime:
                isLast.append(True)
                connectionTime.append(np.nan)
                tripNum.append(tripCounter)

                deviceTripNum[dfData['macaddr_hashed'].iloc[i]] = tripCounter
                tripCounter += 1

            else:
                isLast.append(False)
                connectionTime.append(timeSince)
                tripNum.append(deviceTripNum[dfData['macaddr_hashed'].iloc[i]])

            lastSeen[dfData['macaddr_hashed'].iloc[i]] = dfData['dateTime'].iloc[i]

        else:
            isLast.append(True)
            lastSeen[dfData['macaddr_hashed'].iloc[i]] = dfData['dateTime'].iloc[i]
            tripNum.append(tripCounter)

            deviceTripNum[dfData['macaddr_hashed'].iloc[i]] = tripCounter
            tripCounter += 1

            #Appends NaN to connection time since time has to be added later
            connectionTime.append(np.nan)

    dfData['isLast'] = list(reversed(isLast))
    dfData['connectionTime']= list(reversed(connectionTime))
    dfData['tripNum'] = list(reversed(tripNum))

    return(dfData)
```

```python
def estimateLastConnectionTimes(dfData,pval = 0.5):
    """ Function that adds estimates for the connection times of associations labelled as last connections.
    Creates estimates for last connection times, called linger times, by taking the upper limit of a (p = 0.5)
    confidence interval for how long the device was likely connected. Function then replaces NaN connection
    times in fullData dataframe with newly estimated linger time.
    """

    notLast = dfData[dfData['isLast'] == False]
    apConnections = {}

    for j in range(len(notLast)):
        if notLast['ap_name'].iloc[j] in apConnections:
            apConnections[notLast['ap_name'].iloc[j]].append(notLast['connectionTime'].iloc[j])
        else:
            apConnections[notLast['ap_name'].iloc[j]] = [notLast['connectionTime'].iloc[j]]

    apLinger = {}

    for i in apConnections.keys():
        if len(apConnections[i]) == 1:
            apLinger[i] = 2 * apConnections[i][0]
        else:
            a = [o.total_seconds() for o in apConnections[i]]
            interval = st.t.interval(pval, len(a)-1, loc=np.mean(a), scale=st.sem(a))
            apLinger[i] = datetime.timedelta(seconds = max(interval))

    fullData = dfData
    fullData = fullData.assign(connectionTime = fullData.connectionTime.fillna(fullData.ap_name.map(apLinger)))
    return(fullData)
```

```python
def filterTrips(fullData):
    """ Function modifies previous fullData dataFrame to filter out all trips that do not have a minimum of
    3 associations, or a length longer than 20 minutes,
     during the previously defined time window (2 hrs). Function creates a new dataFrame,
    dfTrips, containing the final filtered table of all trips that are counted as valid.
    """

    tripNum = []
    startTime = []
    endTime = []
    minAssoc = 3
    for i in range(fullData.tripNum.max()):
        assoc = fullData[fullData['tripNum'] == i]
        beginTrip = assoc['dateTime'].iloc[0]
        endTrip = assoc['dateTime'].iloc[-1] + assoc['connectionTime'].iloc[-1]
        if ((len(assoc) >= minAssoc)):
            startTime.append(beginTrip)
            endTime.append(endTrip)
            tripNum.append(i)
        else:
            continue

    d = {'tripNum': tripNum, 'startTime' : startTime, 'endTime' : endTime}
    dfTrips = pd.DataFrame(data = d)
    dfTrips['tripLen'] = dfTrips['endTime'] - dfTrips['startTime']
    dfTrips =dfTrips[dfTrips['tripLen'] > datetime.timedelta(minutes = 20)]
    dfTrips['tripLen'] = dfTrips['endTime'] - dfTrips['startTime']
    len(dfTrips[dfTrips['tripLen'] > datetime.timedelta(hours = 10)])
    dfTrips.sort_values(by=["startTime"], inplace=True, ascending=True)

    return(dfTrips)
```

```python
def getIntervals(dfTrips, tDelta = datetime.timedelta(minutes=15)):
    """ Function creates 15 minute buckets to group trips by in order to estimate occupancy during each interval.
    First, starting point of the first interval (startdt) and starting point of the last interval (enddt) are
    created MANUALLY. The year, month, day, hours, minutes, and seconds must be CHANGED MANUALLY on the lines
    below to accurately reflect the start of the 15 minute interval in which the first observation/association
    falls, and the start of the 15 minute interval in which the last observation/association falls (because we
    are using forward looking intervals). The function aggregates occupancy during each interval by checking
    how many trips have startTimes before the start of the 15 minute interval and endTimes after the end of the
    15 minute interval. Lastly, function exports dataFrame of occupancy to a csv file.
    """

    eastern = pytz.timezone("US/Eastern")
    startdt = datetime.datetime(2020, 9, 8, 0, 0, 0, tzinfo = eastern)
    enddt = datetime.datetime(2020, 9, 11, 23, 45, 0, tzinfo = eastern)

    startList = list(datetime_range(startdt, enddt, tDelta))
    endList = list(datetime_range(startdt + tDelta, enddt + tDelta, tDelta))

    occupancy = []
    time = []

    for i in range(len(startList)):
        tempOccupancy = len(dfTrips[(dfTrips['startTime'] < startList[i]) & (dfTrips['endTime'] > endList[i])])
        occupancy.append(tempOccupancy)
        time.append(startList[i].strftime("%m/%d/%Y, %H:%M:%S"))

    dfOccupancy = pd.DataFrame(list(zip(time, occupancy)), columns =['Time', 'Occupancy'])

    return(dfOccupancy)
```

```python
def filterTrips(fullData, minAssoc = 2):
    """ Function modifies previous fullData dataFrame to filter out all trips that do not have a minimum of
    minAssoc associations, or a length longer than 20 minutes,
     during the previously defined time window (2 hrs). Function creates a new dataFrame,
    dfTrips, containing the final filtered table of all trips that are counted as valid.
    """

    tripNum = []
    startTime = []
    endTime = []
    zoneStay = []
    #create list of ap names for relevant zone/room - MUST REPLACE AP names w/ desired zone
    zoneAPs = ['lsrc-a150-ap3502i-hc-1','lsrc-a156-ap3702i-hc-1','lsrc-a158-ap3502i-hc-1']

    for i in range(fullData.tripNum.max()):
        assoc = fullData[fullData['tripNum'] == i]
        # add condition that removes trips that do not have an association to one of the APs in relevant zone
        APsInTrip = assoc['ap_name']
        intersectionAPs = list(set(zoneAPs) & set(APsInTrip))
        if ((len(assoc) >= minAssoc) and intersectionAPs):
            #obtain dateTime of first association of trip that connects to AP in relevant zone
            assoc.sort_values(by=["time"], inplace=True, ascending=True)
            firstAssocIndex = assoc['ap_name' in zoneAPs].index[0]
            beginTrip = assoc['dateTime'].iloc[firstAssocIndex]
            #obtain dateTime of first association after the last association of trip that connected to
            #AP in relevant zone
            assoc.sort_values(by=["time"], inplace=True, ascending=False)
            lastAssocIndexQuasi = assoc['ap_name' in zoneAPs].index[0]
            lastAssocIndex = len(assoc) - lastAssocIndexQuasi - 1
            assoc.sort_values(by=["time"], inplace=True, ascending=True)
            for row in assoc.loc[firstAssocIndex:lastAssocIndex].itertuples():
                timeInZone = 0
                timeInZone += assoc['connectionTime'].iloc[row]

            endTrip = beginTrip + timeInZone
            startTime.append(beginTrip)
            endTime.append(endTrip)
            zoneStay.append(timeInZone)
            tripNum.append(i)
        else:
            continue

    d = {'tripNum': tripNum, 'startTime' : startTime, 'endTime' : endTime, 'zoneStay' : zoneStay}
    dfTrips = pd.DataFrame(data = d)
    dfTrips['tripLen'] = dfTrips['endTime'] - dfTrips['startTime']

    #take ratio of total time in zone to total duration of trip
    dfTrips = dfTrips[(dfTrips['zoneStay'] / float(dfTrips['tripLen'])) > 0.5]

    #Removes minimum time to relax definition of trip
    #dfTrips =dfTrips[dfTrips['tripLen'] > datetime.timedelta(minutes = 20)]
    dfTrips.sort_values(by=["startTime"], inplace=True, ascending=True)
    return(dfTrips)
```

Zone-Level Occupancy Estimation Function (modified filterTrips function)

Appendix B: Environmental Benefit Calculations

https://docs.google.com/spreadsheets/d/16DtLDzYm1_xc8l6cknC9G8VrDBrqcPDD6sb_LQIOZFg/edit?usp=sharing

Appendix C: Heating and Cooling Energy Consumption Regression

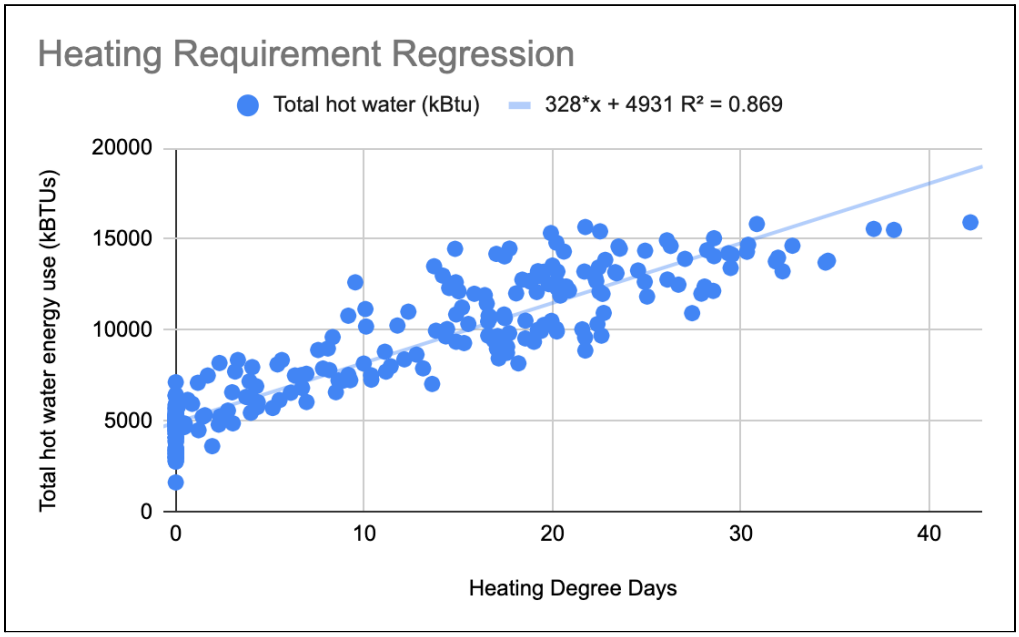


Chart of the regression of heating degree days on hot water energy usage per day, 2019. We used the slope of the regression line, 328, as our estimate for the kBtu of energy that could be saved by letting the temperature set point wander by one degree more on cold days.

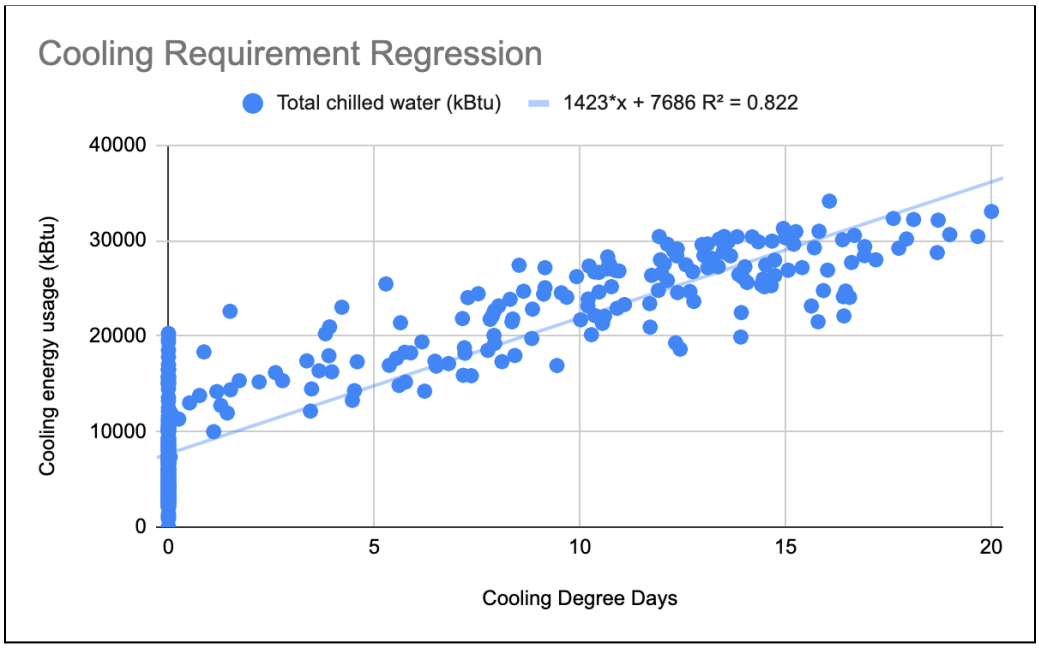
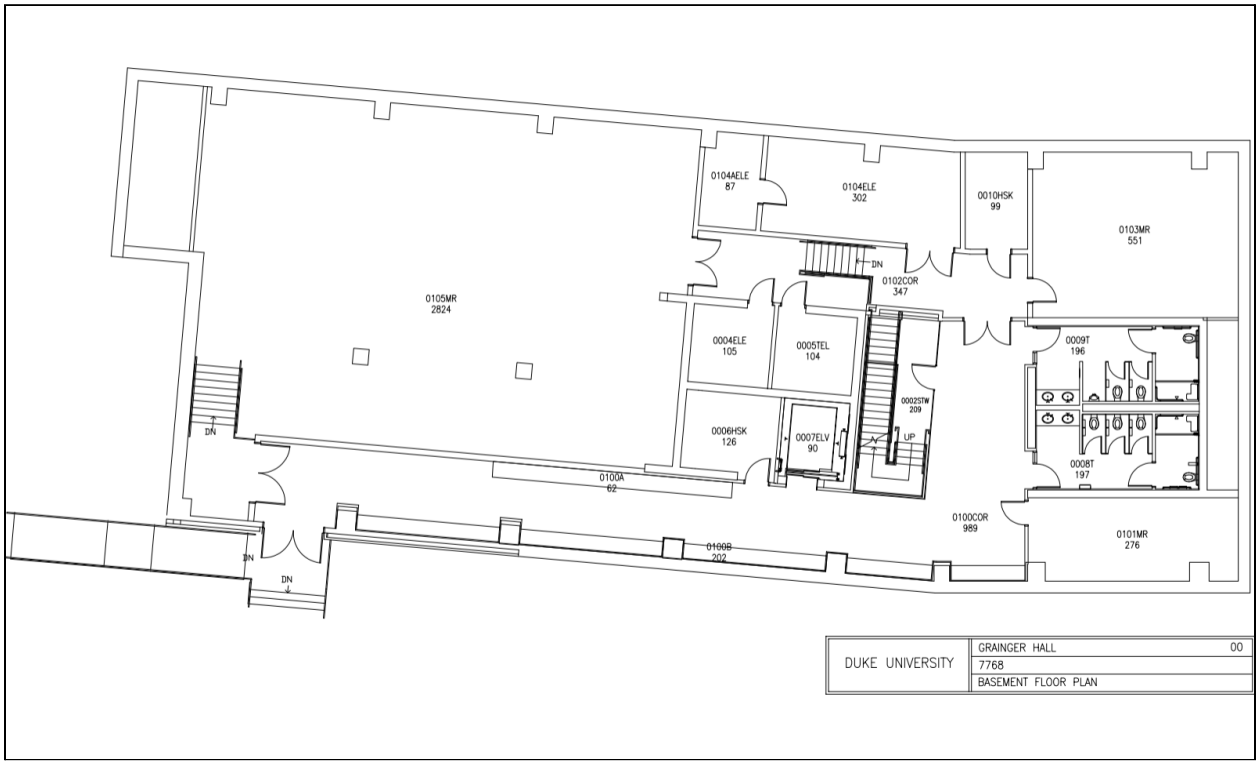
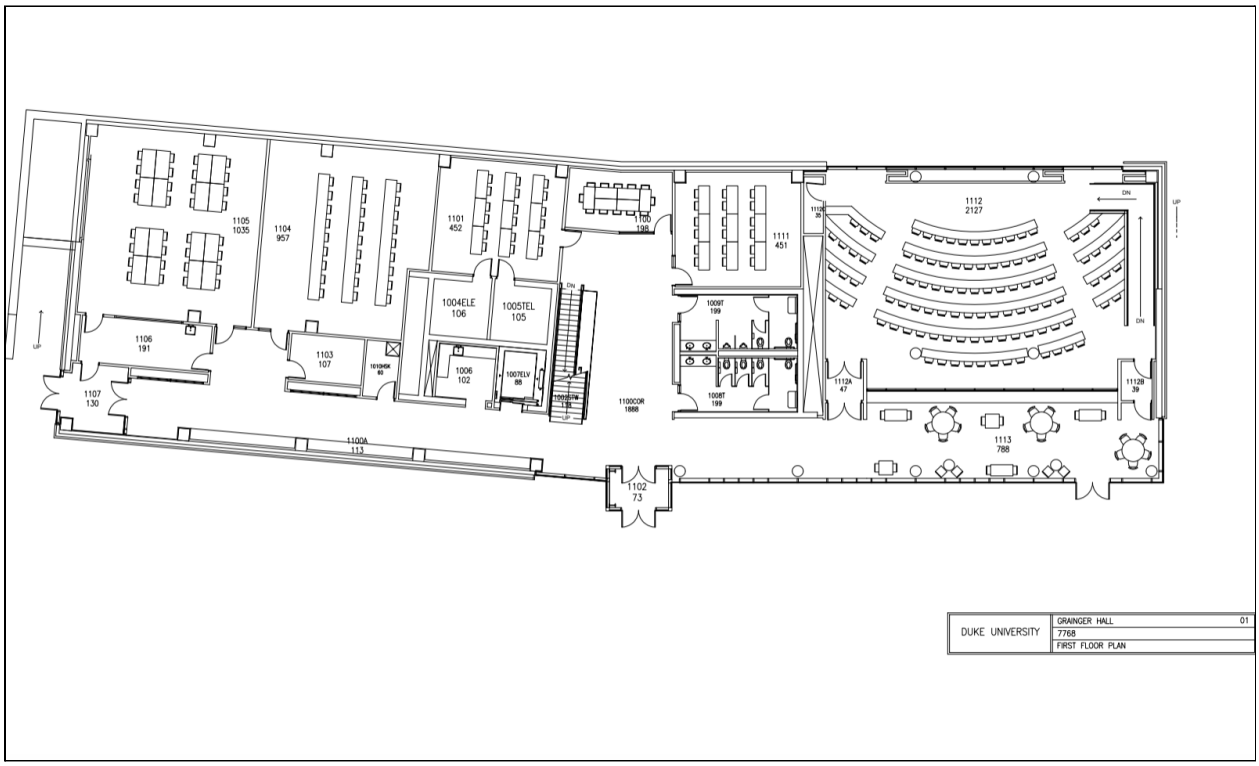
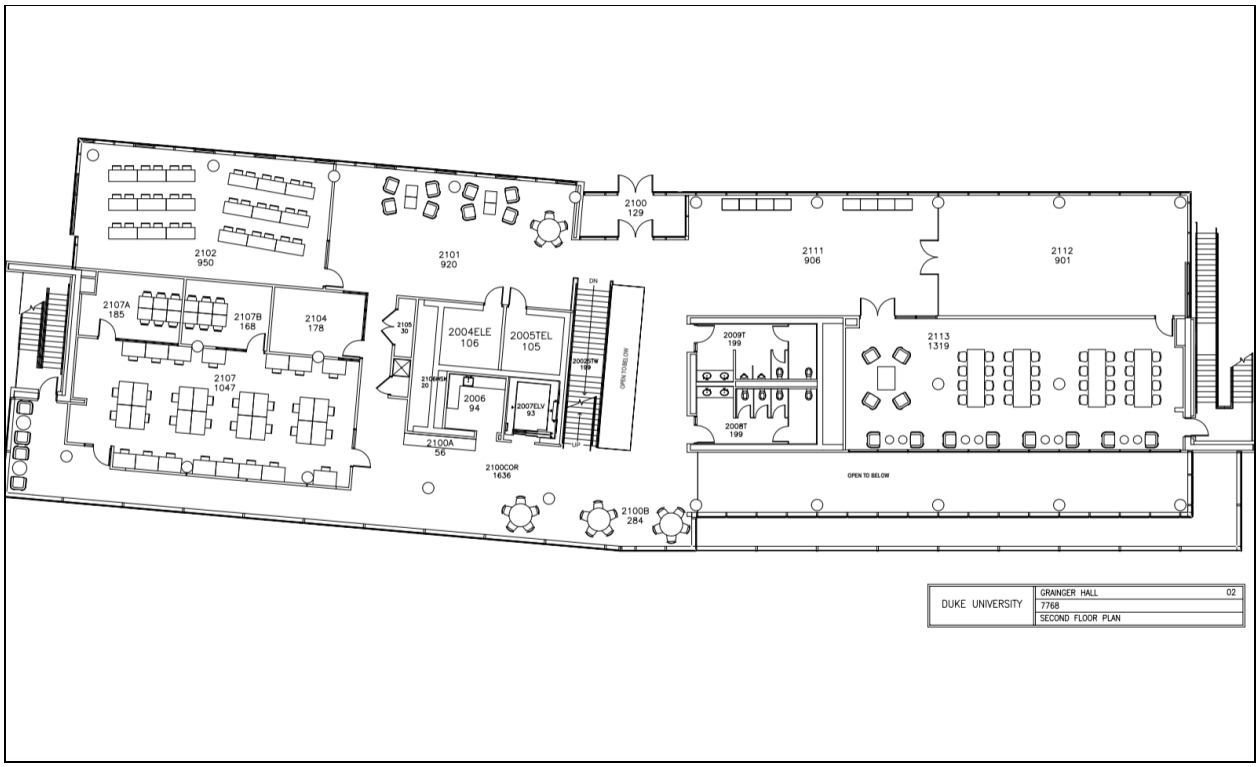


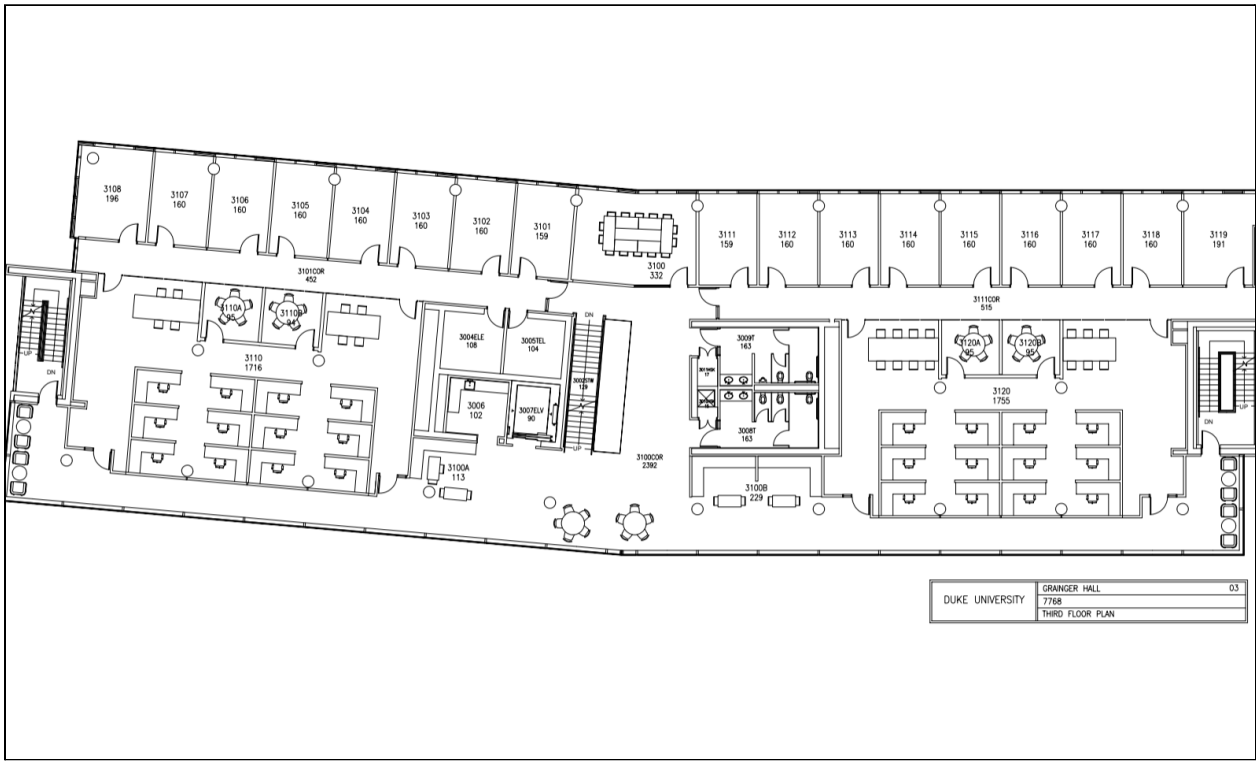
Chart of the regression of cooling degree days on cold water energy usage per day, 2019. We used the slope of the regression line, 1423, as our estimate for the kBtu of energy that could be saved by letting the temperature set point wander by one degree more on hot days.

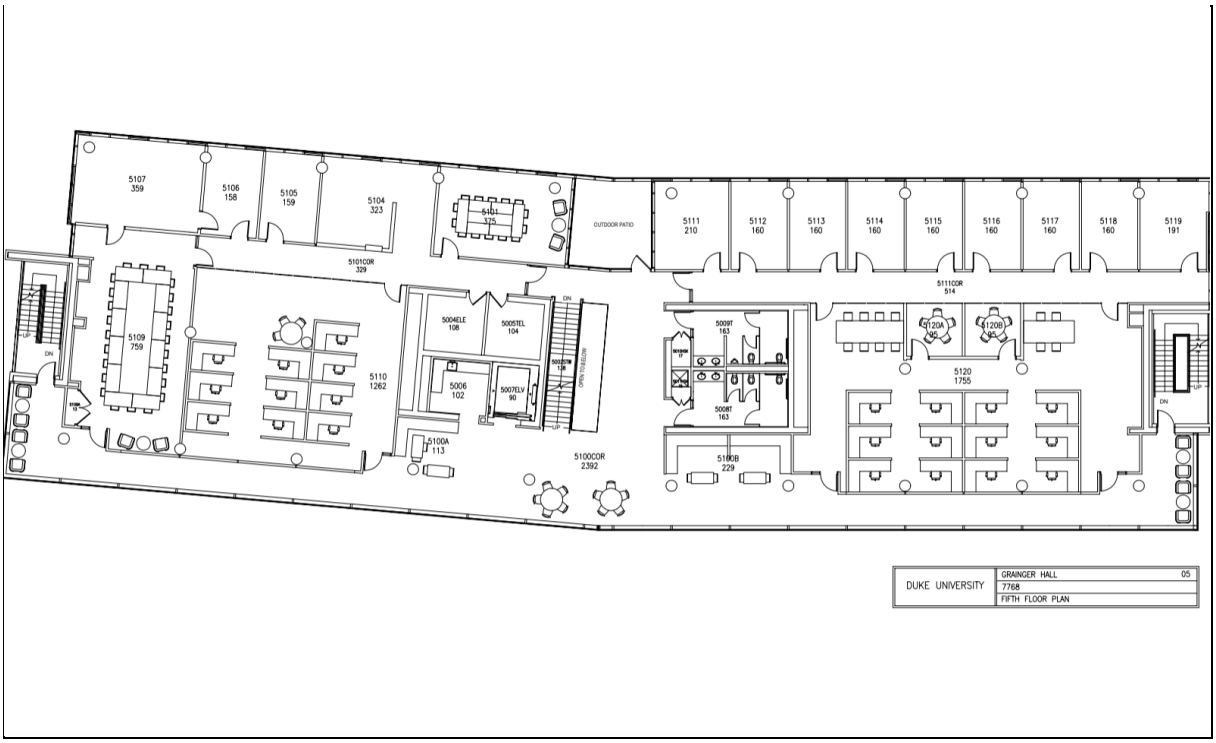
Appendix D: Grainger Hall Floor Plans

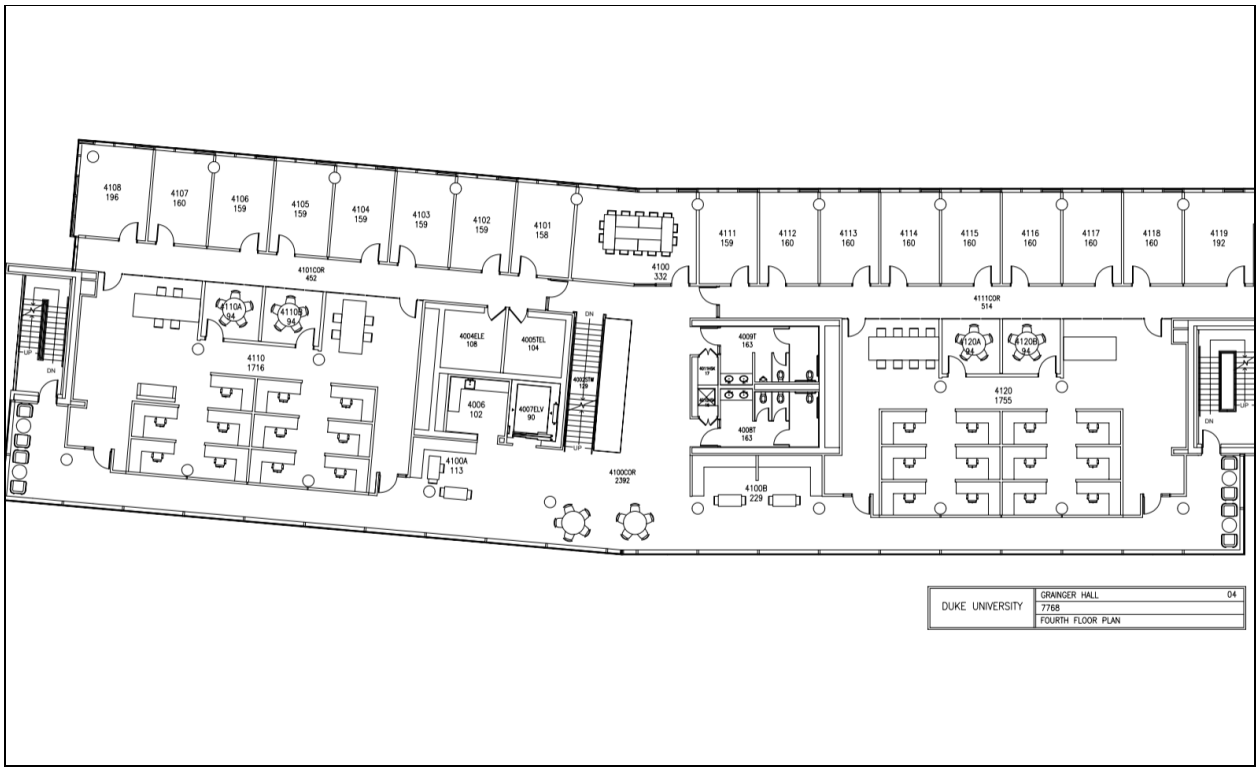


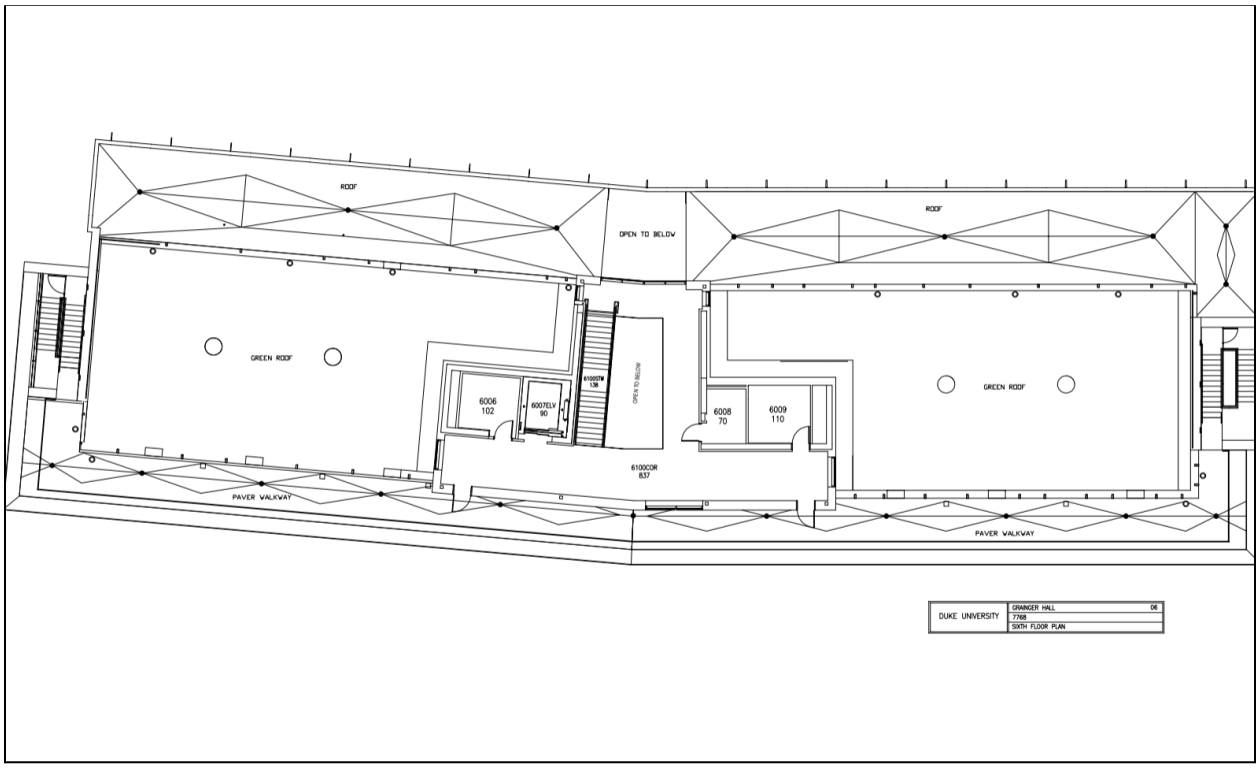



2102
950
2101
920
2100
129
2111
906
2112
901
2107A
185
2107B
168
2104
178
2107
1047
2105
30
2004ELE
106
2005TEL
105
2006
94
2007ELV
93
2100A
56
2100COR
1636
2009T
199
2008T
199
2113
1319
OPEN TO BELOW
2100B
284
DUKE UNIVERSITY
GRAINGER HALL
02
7768
SECOND FLOOR PLAN


3108
196
3107
160
3106
160
3105
160
3104
160
3103
160
3102
160
3101
159
3100
332
3111
159
3112
160
3113
160
3114
160
3115
160
3116
160
3117
160
3118
160
3119
191
3101COR
452
3111COR
515
3110A
95
3110B
94
3110
1716
3004ELE
108
3005TEL
104
3006
102
3007ELV
90
3100A
113
3100COR
2392
3009T
163
3008T
163
3120A
95
3120B
95
3120
1755
3100B
229
DUKE UNIVERSITY
GRAINGER HALL
03
7768
THIRD FLOOR PLAN


5107
359
5106
158
5105
159
5104
323
5101
375
CUTDOOR PATIO
5111
210
5112
160
5113
160
5114
160
5115
160
5116
160
5117
160
5118
160
5119
191
5101COR
329
5111COR
514
5109
759
5110
1262
5004ELE
108
5005TEL
104
5006
102
5007ELV
90
5008STR
108
OPEN TRESSLE
DN
UP
5100A
113
5100COR
2392
5009T
163
5009T
163
5100B
229
5120A
95
5120B
95
5120
1755
DUKE UNIVERSITY
GRAINGER HALL
05
7768
FIFTH FLOOR PLAN


4108
196
4107
160
4106
159
4105
159
4104
159
4103
159
4102
159
4101
158
4101COR
452
4100
332
4111
159
4112
160
4113
160
4114
160
4115
160
4116
160
4117
160
4118
160
4119
192
4111COR
514
4110A
94
4110B
94
4110
1716
4004ELE
108
4005TEL
104
4006
102
4007ELV
90
4100A
113
4100COR
2392
4009T
163
4008T
163
4100B
229
4120A
94
4120B
94
4120
1755
DN
UP
DUKE UNIVERSITY
GRAINGER HALL
04
7768
FOURTH FLOOR PLAN


ROOF
OPEN TO BELOW
ROOF
GREEN ROOF
6006
102
6007ELV
90
6100STW
138
OPEN TO BELOW
6008
70
6009
110
6100COR
837
GREEN ROOF
PAVER WALKWAY
PAVER WALKWAY
DUKE UNIVERSITY
GRAINGER HALL
06
7768
SIXTH FLOOR PLAN

Appendix E: Occupancy Processing R-Markdown File

(see next page)

Appendix E: Occupancy Processing R-Markdown File

Imports the data

```
rawData = read.csv("occupancy-with-mask-new.csv")

#formats dates and sorts sequentially
allData <- rawData %>%
  mutate(date_time = parse_date_time(Time, orders = "mdY HMS")) %>%
  mutate(weekday = weekdays(date_time)) %>%
  arrange(date_time) %>%
  select(-Time) %>%
  relocate(date_time, weekday)
```

Summary stats

```
#creates columns for hours and mins
allData <- allData %>%
  mutate(hours = hour(date_time), minutes = minute(date_time))

#sets treshold for "high occupancy"
highBarLow = 10
highBarHigh = 15

avgOcc <- allData %>%
  group_by(weekday, hours, minutes) %>%
  summarise(n = n(), Mean = mean(occupancy.num),
            sd = sd(occupancy.num),
            Median = median(occupancy.num),
            ZeroValues = sum(occupancy.num == 0),
            HighValues1 = sum(occupancy.num >= highBarLow),
            HighValues2 = sum(occupancy.num >= highBarHigh))
```

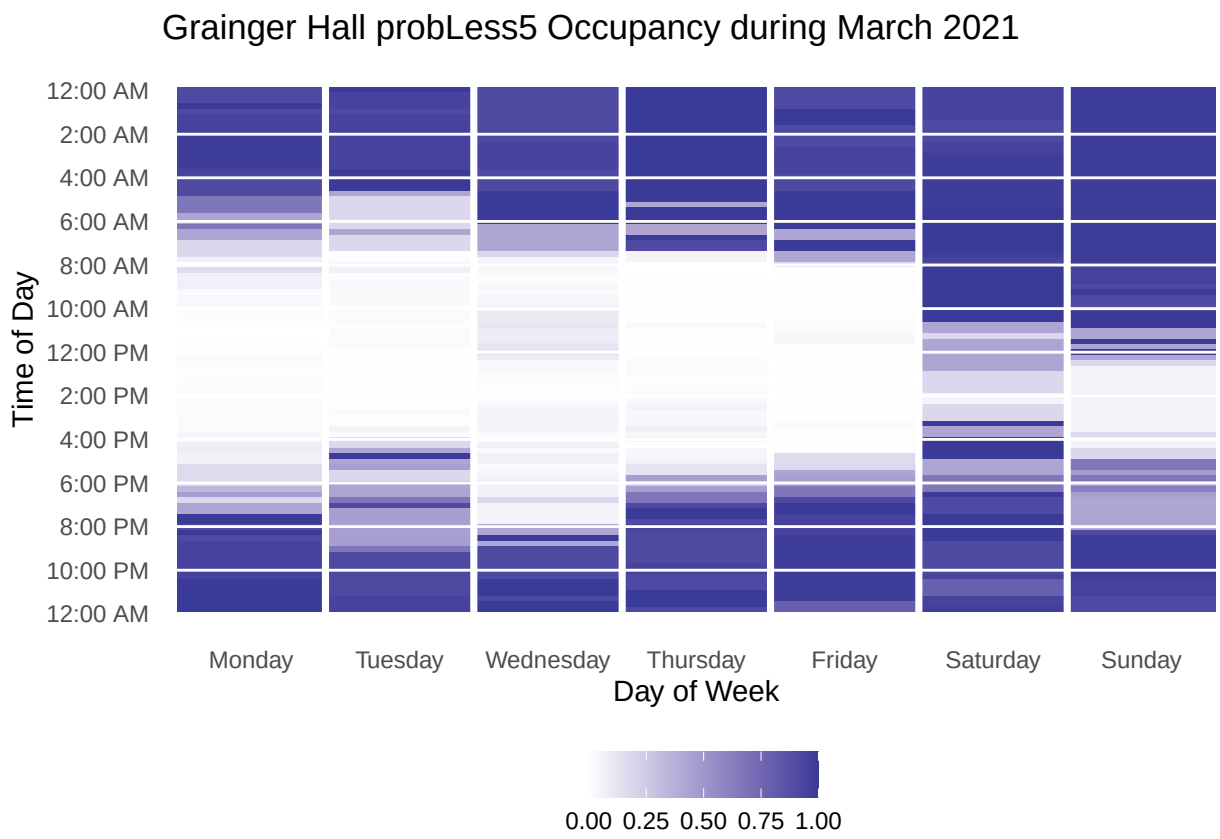
```
## `summarise()` regrouping output by 'weekday', 'hours' (override with `.groups` argument)
```

```
avgOcc <- avgOcc %>%
  mutate(HighValues1pct = HighValues1 / n,
         HighValues2pct = HighValues2 / n,
         ZeroValuespct = ZeroValues / n,
         stdErr = sd / sqrt(n),
         probLess5 = pt((6-Mean)/stdErr, n-1),
         prob5_90 = probLess5 > 0.9,
```

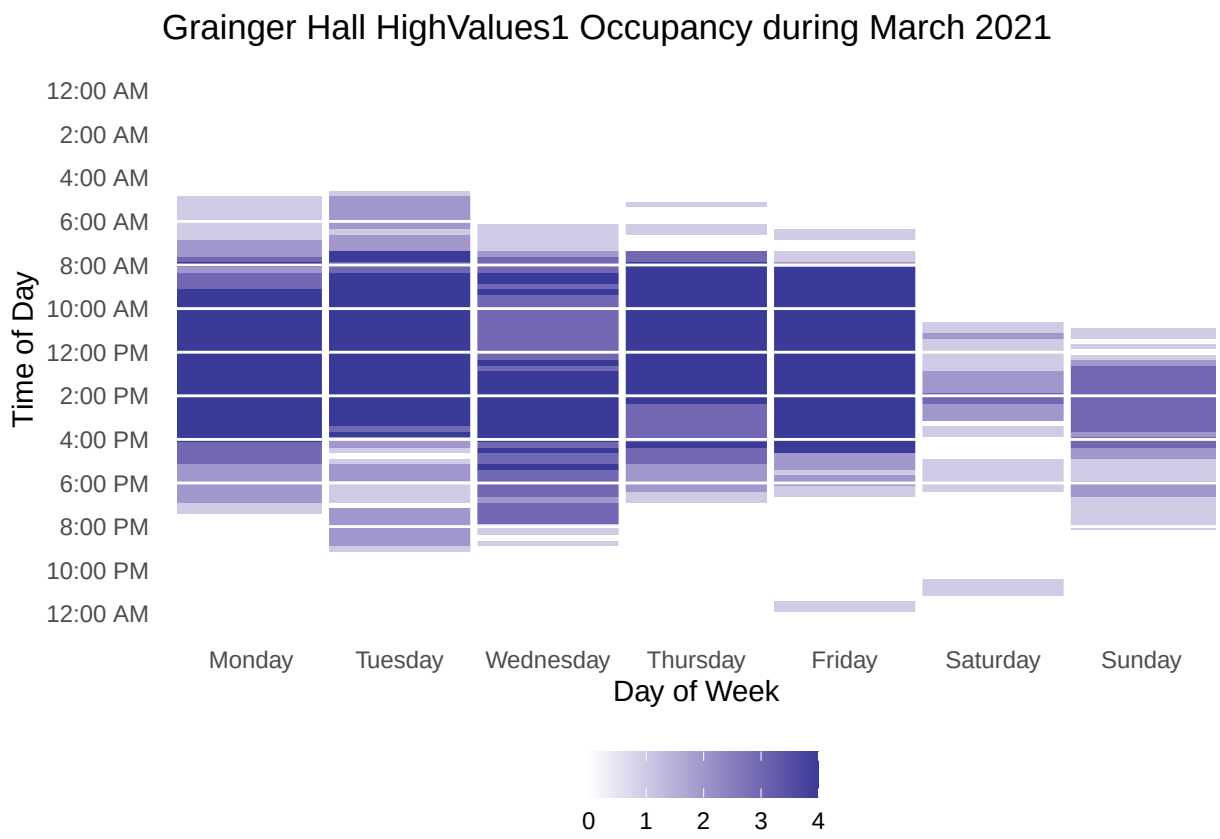
```
         prob5_80 = probLess5 > 0.8,
         prob5_75 = probLess5 > 0.75,
         prob5_50 = probLess5 > 0.5,
         probLess10 = pt((11-Mean)/stdErr, n-1),
         prob10_90 = probLess10 > 0.9,
         prob10_80 = probLess10 > 0.8,
         prob10_75 = probLess10 > 0.75,
         prob10_50 = probLess10 > 0.5)
```

Create average occupancy graph (to mess with plotting styles)

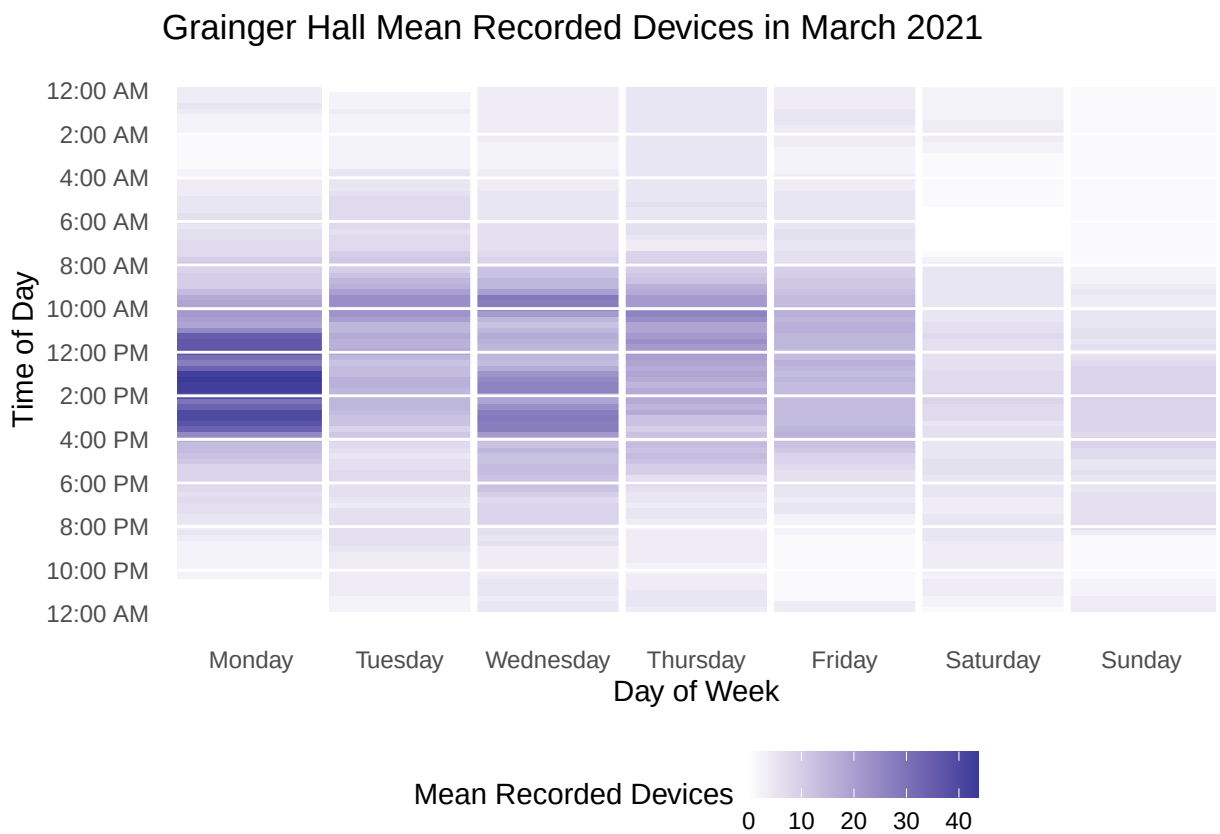
```
plotData <- avgOcc %>%
  mutate(hrSinceMidnight = hours + minutes / 60) %>%
  mutate(weekday = factor(weekday, levels = c("Monday", "Tuesday", "Wednesday", "Thursday", "Friday", "S
  #mutate(timeObj = hms::as_hms(sinceMidnight * 60))


plotWeek <- function(dfData, fieldName, discrete = FALSE, altTitle = NA, legText = NULL){
  ggObj = ggplot(dfData, mapping = aes(x = weekday, y = hrSinceMidnight, fill = eval(parse(text = fieldN
    geom_raster()+
    xlab("Day of Week") + ylab("Time of Day") +
    scale_y_reverse(breaks = seq(0,24, by = 2), labels = c("12:00 AM", paste0(seq(2, 10, by = 2), ":00 A
                              "12:00 PM", paste0(seq(2, 10, by = 2), ":00 PM"), "12:00 AM"),
                    minor_breaks = NULL) +
    geom_vline(xintercept= seq(1.5, 6.5, by = 1), color="white", size=1.25) +
    geom_hline(yintercept= seq(2, 22, by = 2), color="white", size=0.5) +
    ggtitle(replace_na(altTitle, paste("Grainger Hall", fieldName, "Occupancy during March 2021"))) +
    theme_minimal() +
    theme(legend.position = "bottom", panel.ontop = TRUE,
          panel.grid.major.x = element_blank(), panel.grid.major.y = element_blank())

  if(discrete == FALSE){
      ggObj = ggObj + scale_fill_gradient2(name = legText)
    }else{
      ggObj = ggObj + scale_fill_brewer(name = legText)
    }

  ggObj
}


#Work on formatting time and making plot more visually pleasing

plotWeek(plotData, "probLess5")
```




```r
plotWeek(plotData, "HighValues1")
```

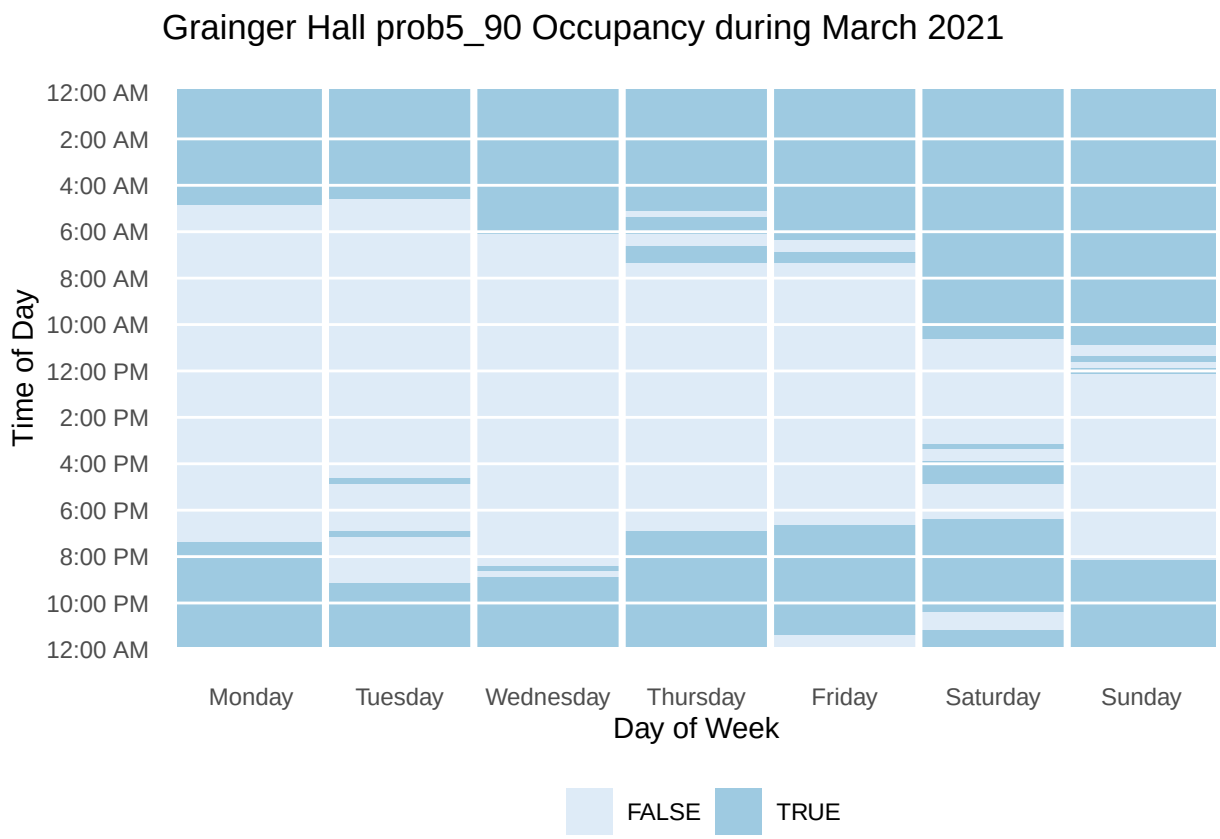



```
plotWeek(plotData, "Mean", legText = "Mean Recorded Devices", altTitle = "Grainger Hall Mean Recorded De
```

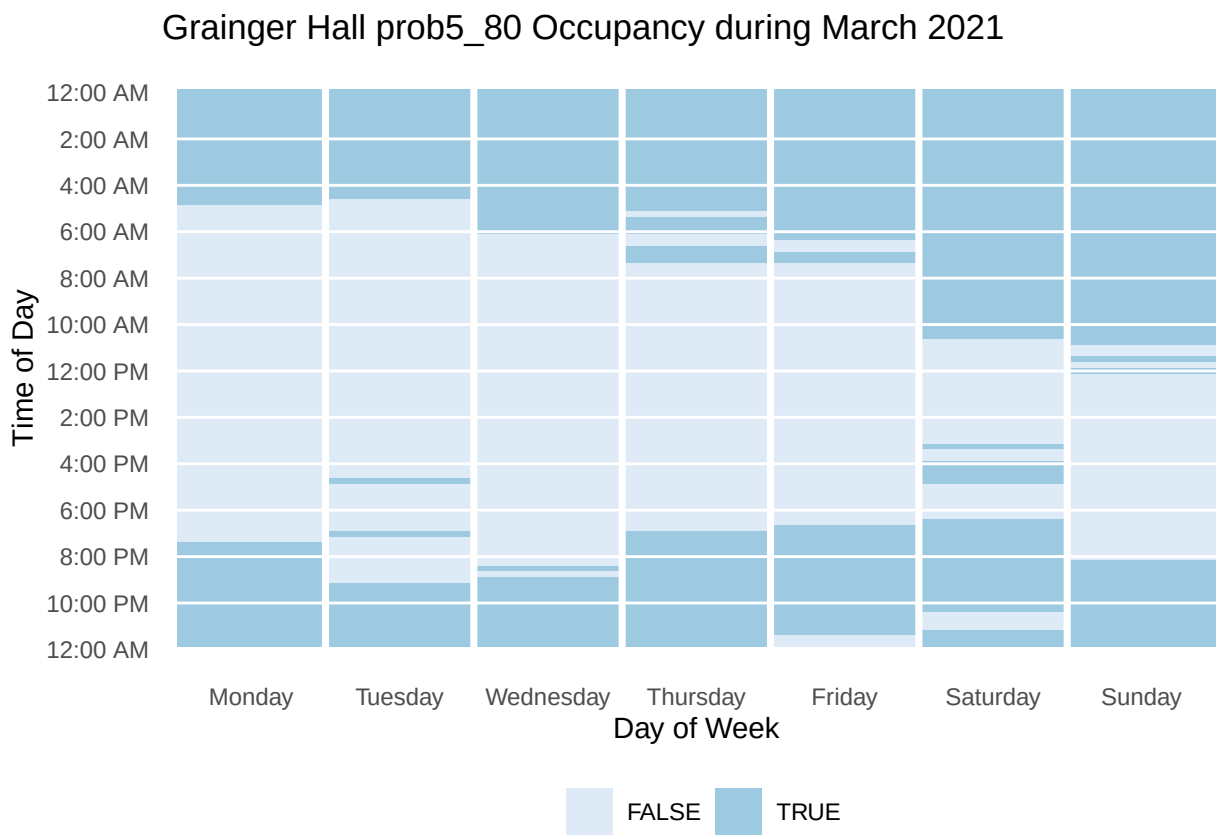


#Probability true mean less than 5

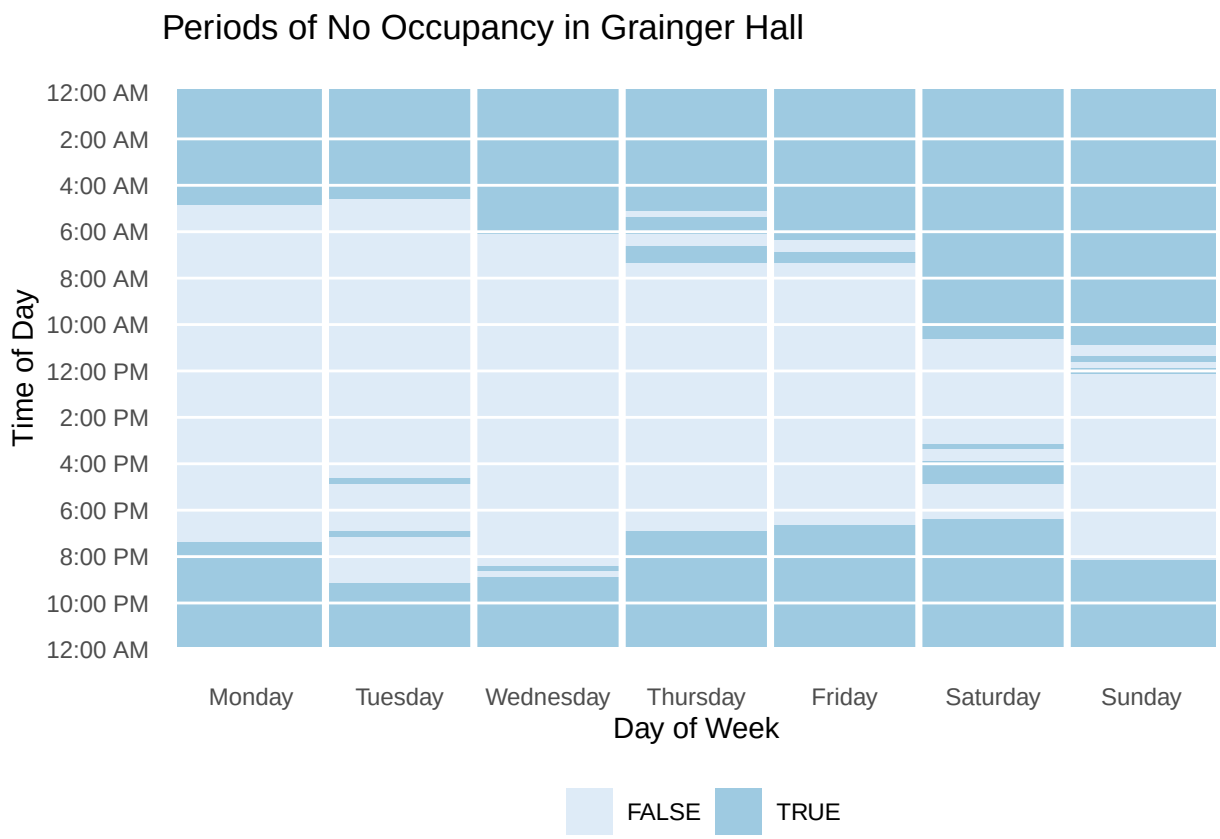
```
plotWeek(plotData, "prob5_90", discrete = TRUE)
```



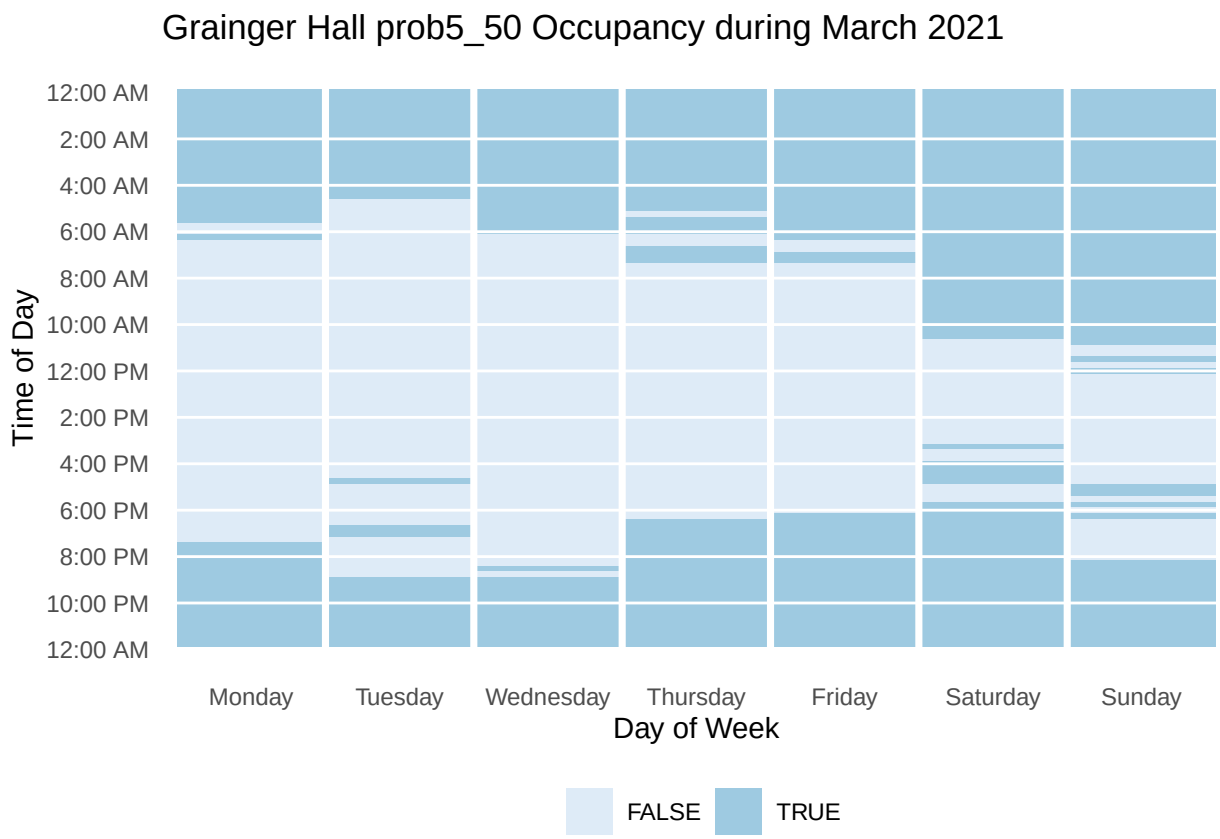

```r
plotWeek(plotData, "prob5_80", discrete = TRUE)
```




```
plotWeek(plotData, "prob5_75", discrete = TRUE, altTitle = "Periods of No Occupancy in Grainger Hall")
```

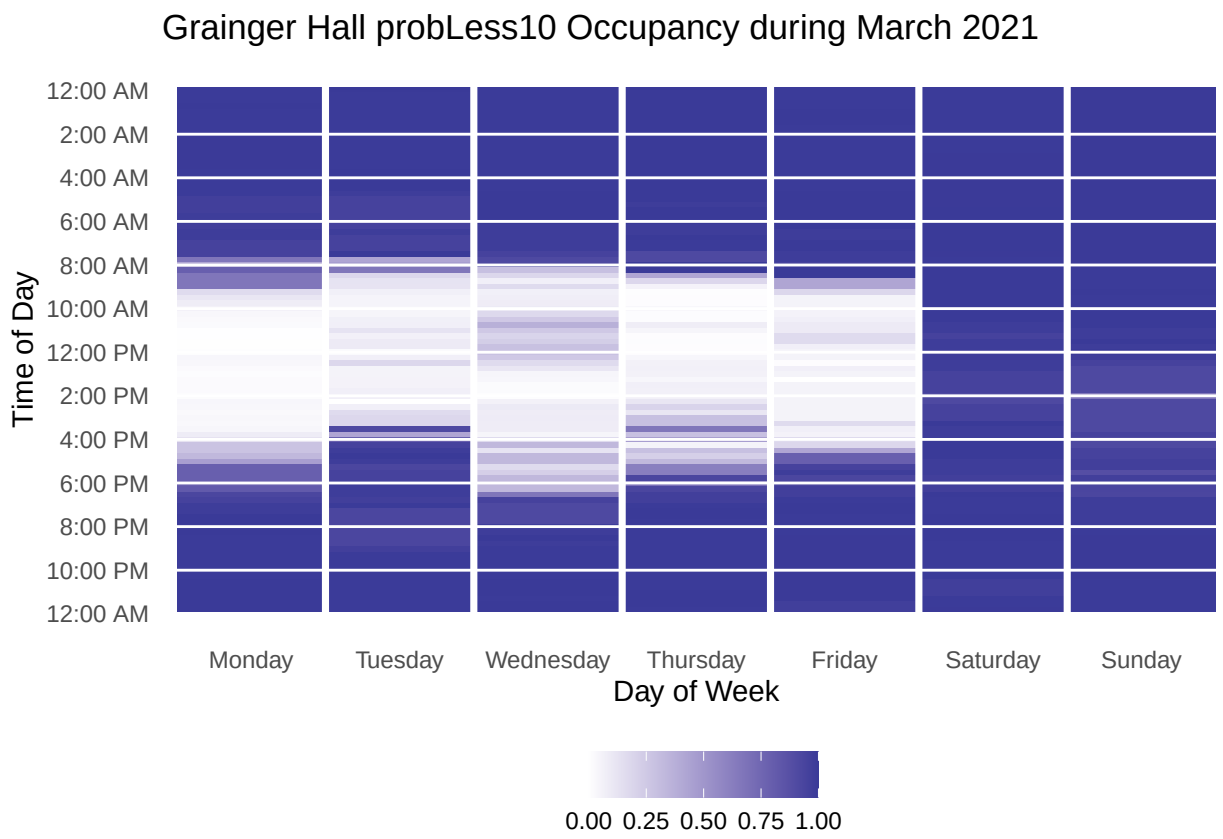



```r
plotWeek(plotData, "prob5_50", discrete = TRUE)
```

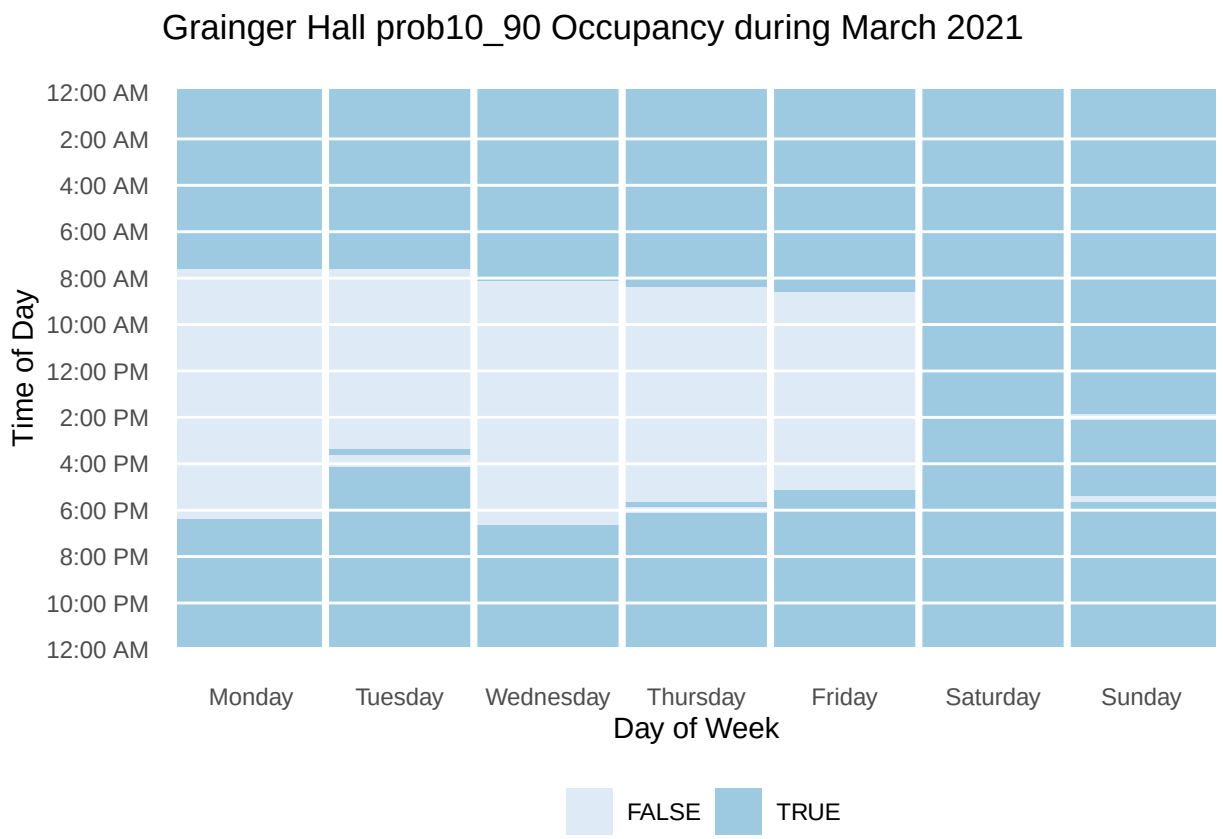


#Probability true mean less than 10

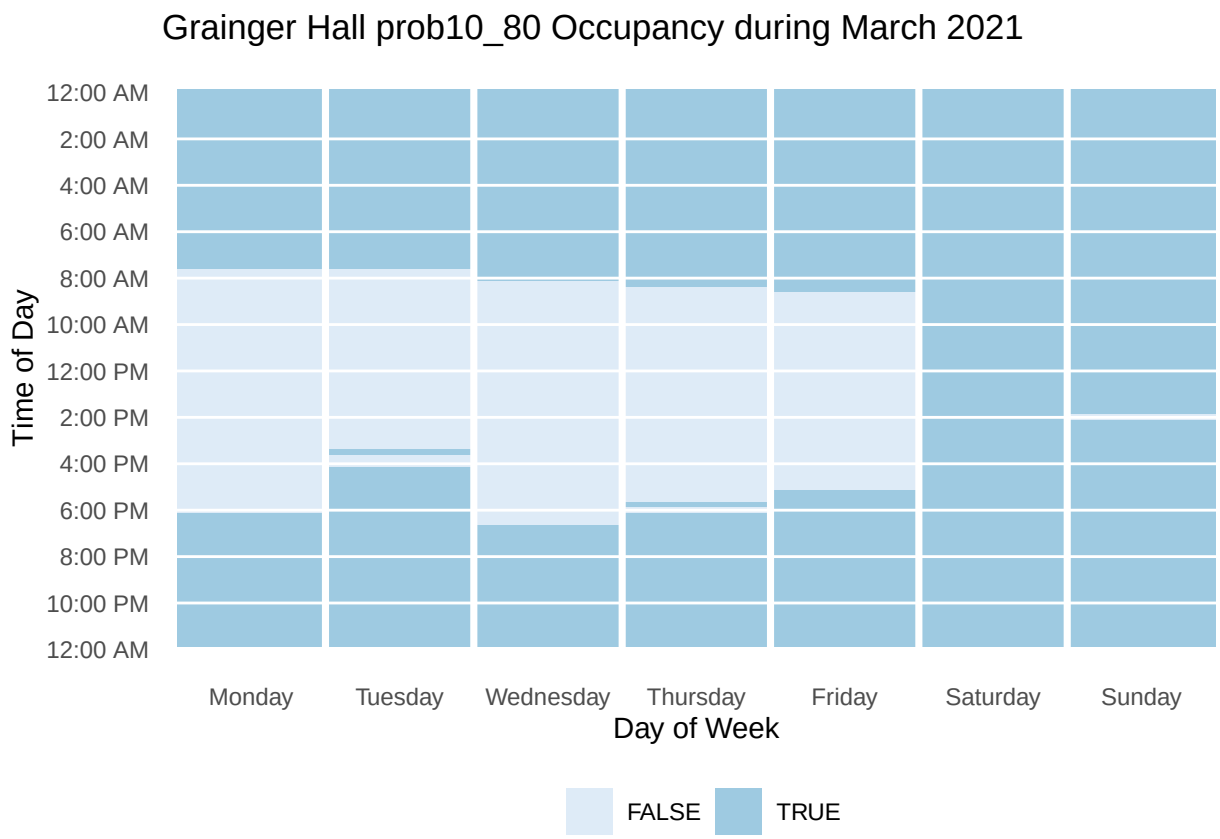
```
plotWeek(plotData, "probLess10")
```



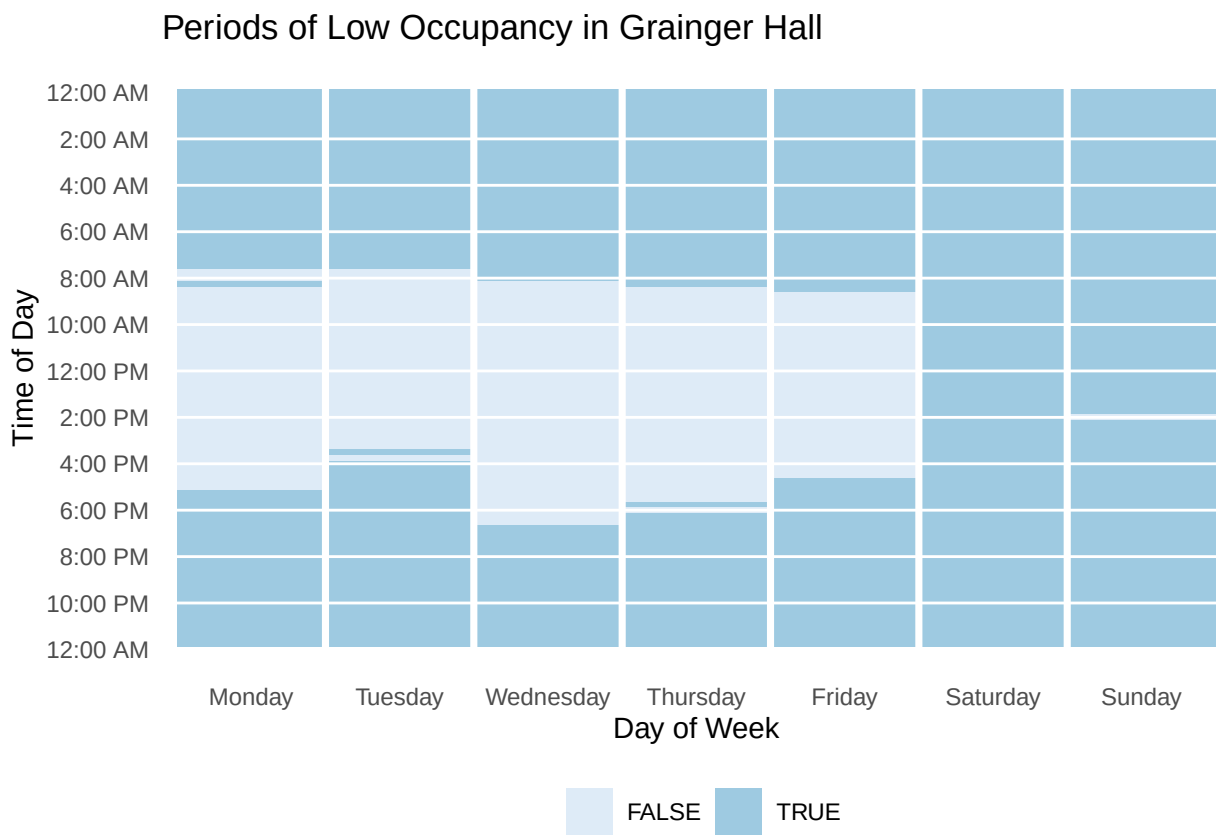

```r
plotWeek(plotData, "prob10_90", discrete = TRUE)
```



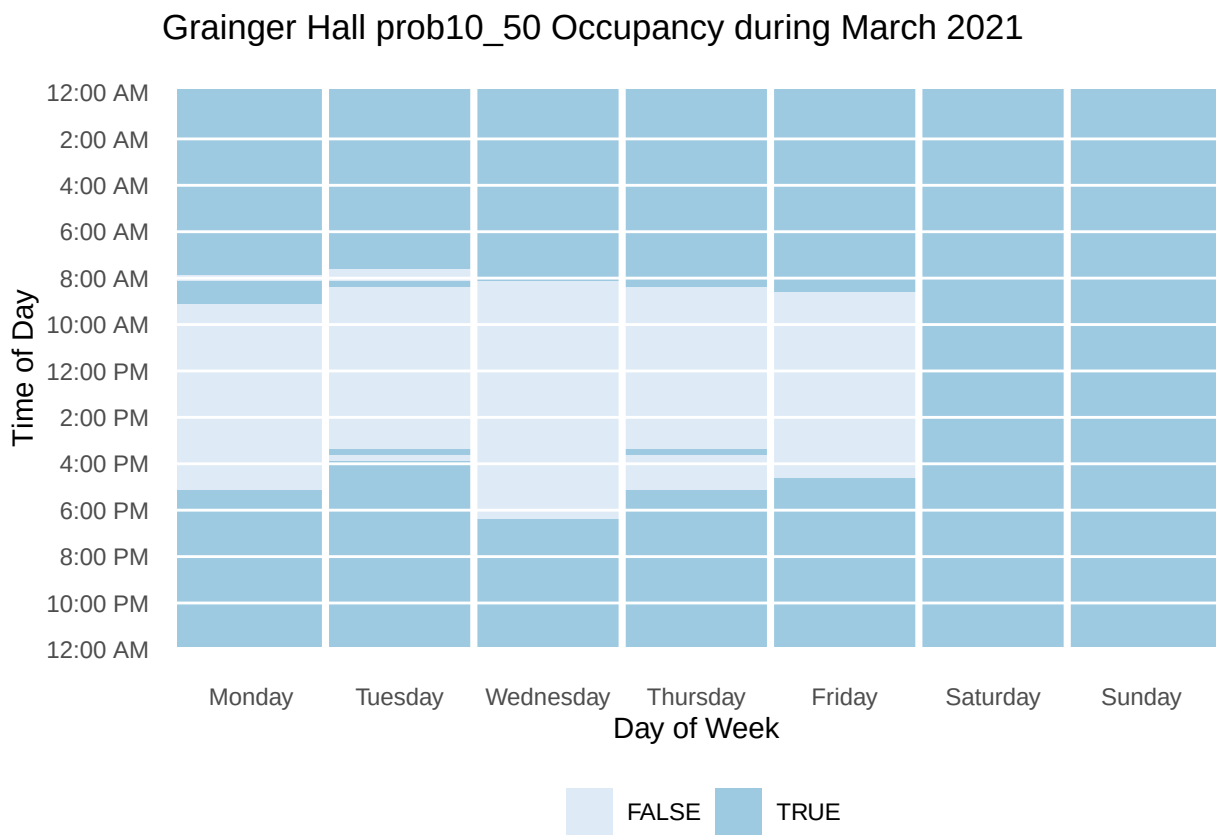

```r
plotWeek(plotData, "prob10_80", discrete = TRUE)
```




```
plotWeek(plotData, "prob10_75", discrete = TRUE, altTitle = "Periods of Low Occupancy in Grainger Hall")
```




```r
plotWeek(plotData, "prob10_50", discrete = TRUE)
```



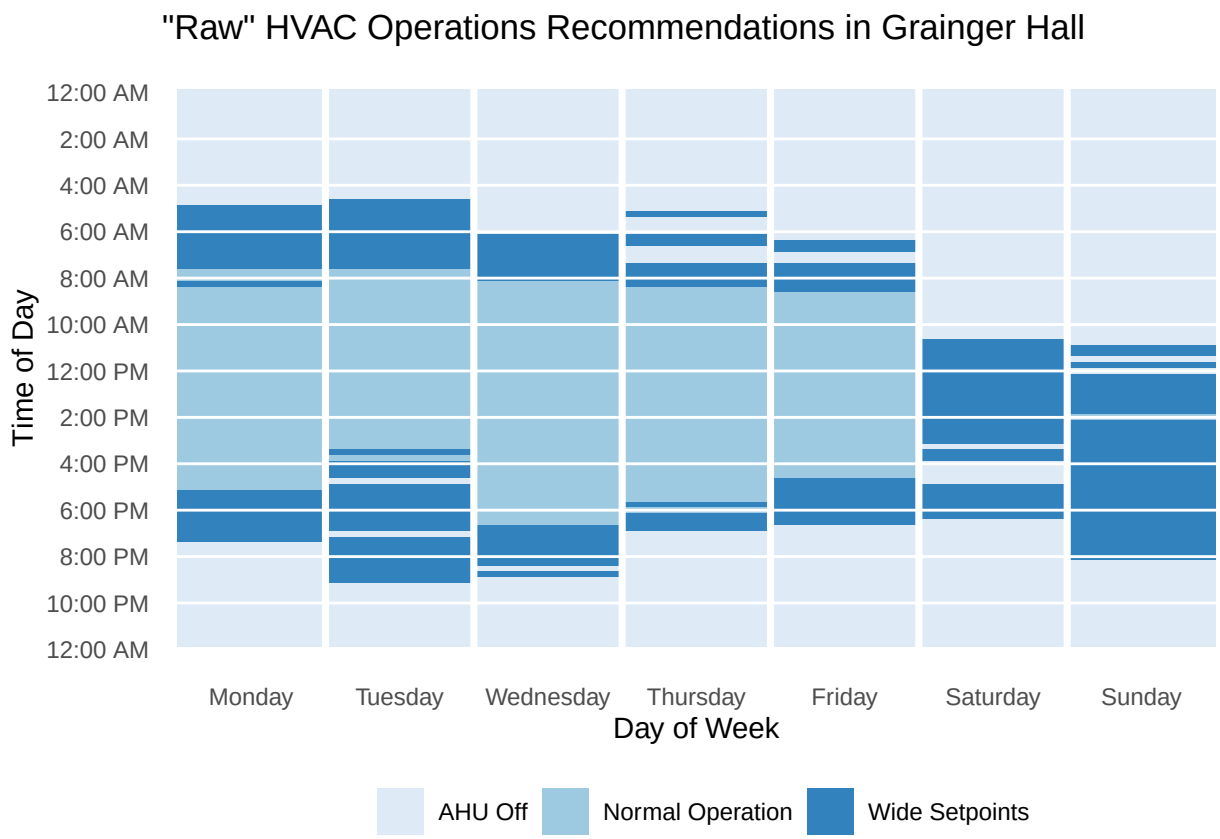
Convert occupancy to recommendations

Using the 75% probability level for <5 occupants (no occuoancy) and <10 occupants (low occupancy)

All changes must be for longer than 30 minutes - err on the conservative side if changes are recommended for <30 mins

All shut downs must be for longer than 1 hour

```
occTable = plotData %>%
  select(weekday, hours, minutes, hrSinceMidnight, prob5_75, prob10_75)

statuses = c("AHU Off", "Wide Setpoints", "Normal Operation")

occTable = occTable %>%
  mutate(recomendation = 3 - (prob5_75 + prob10_75)) %>%
  mutate(occText = statuses[recomendation])%>%
  arrange(weekday, hrSinceMidnight)

plotWeek(occTable, "occText", discrete = TRUE, altTitle = "\"Raw\" HVAC Operations Recommendations in G
```

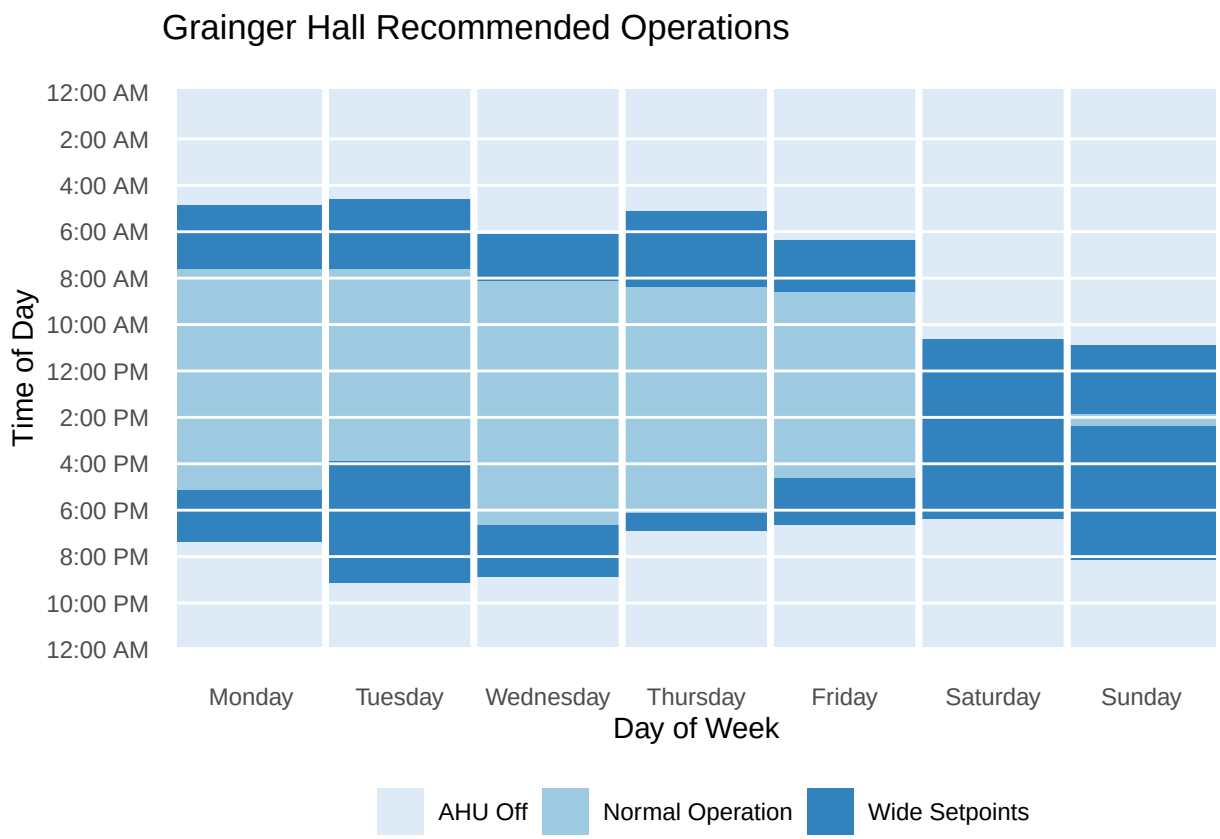


Export data to handle minimum time criteria manually

```
write.csv(occTable, "initialRecs.csv")
```

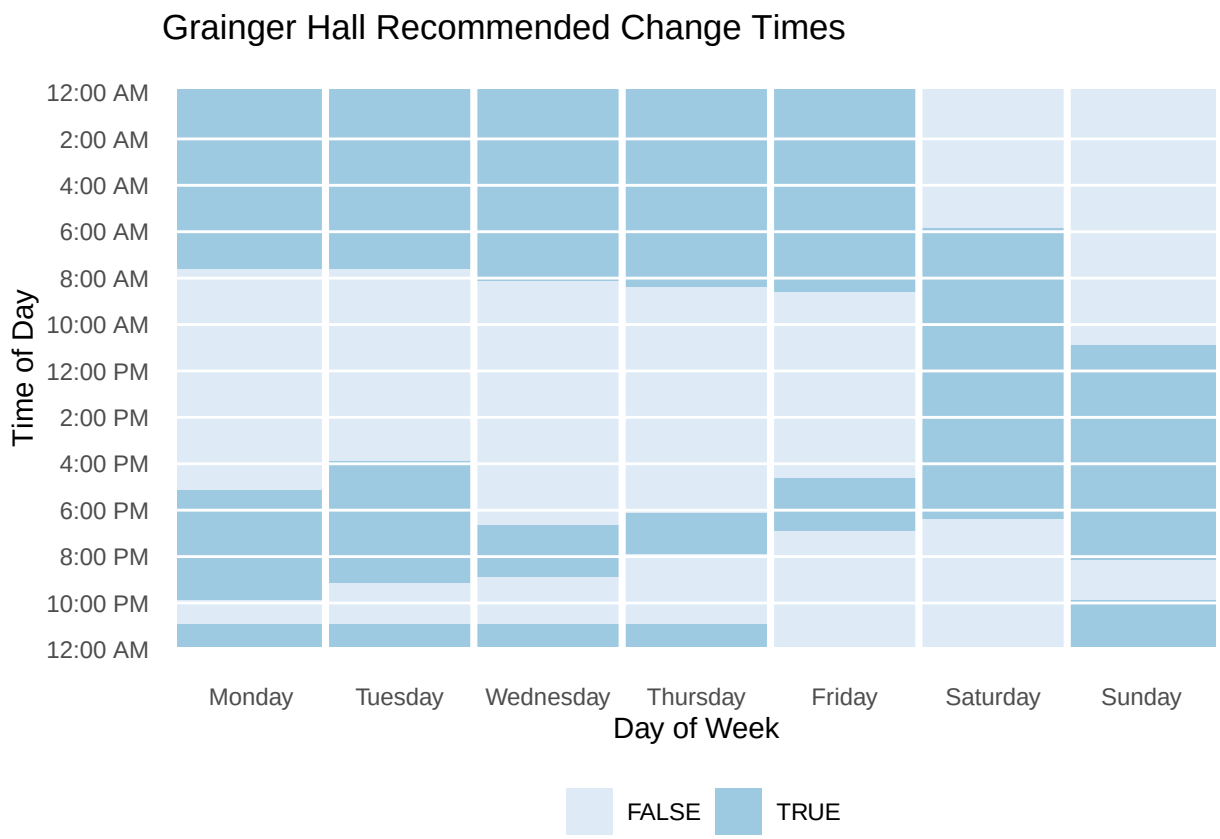
Imports edited recs back

```
newRecs = read.csv("editedRecs.csv")

newRecs$weekday = factor(newRecs$weekday, levels = c("Monday", "Tuesday", "Wednesday", "Thursday", "Fri

plotWeek(newRecs, "smoothedOccText", discrete = TRUE, altTitle = "Grainger Hall Recommended Operations")
```

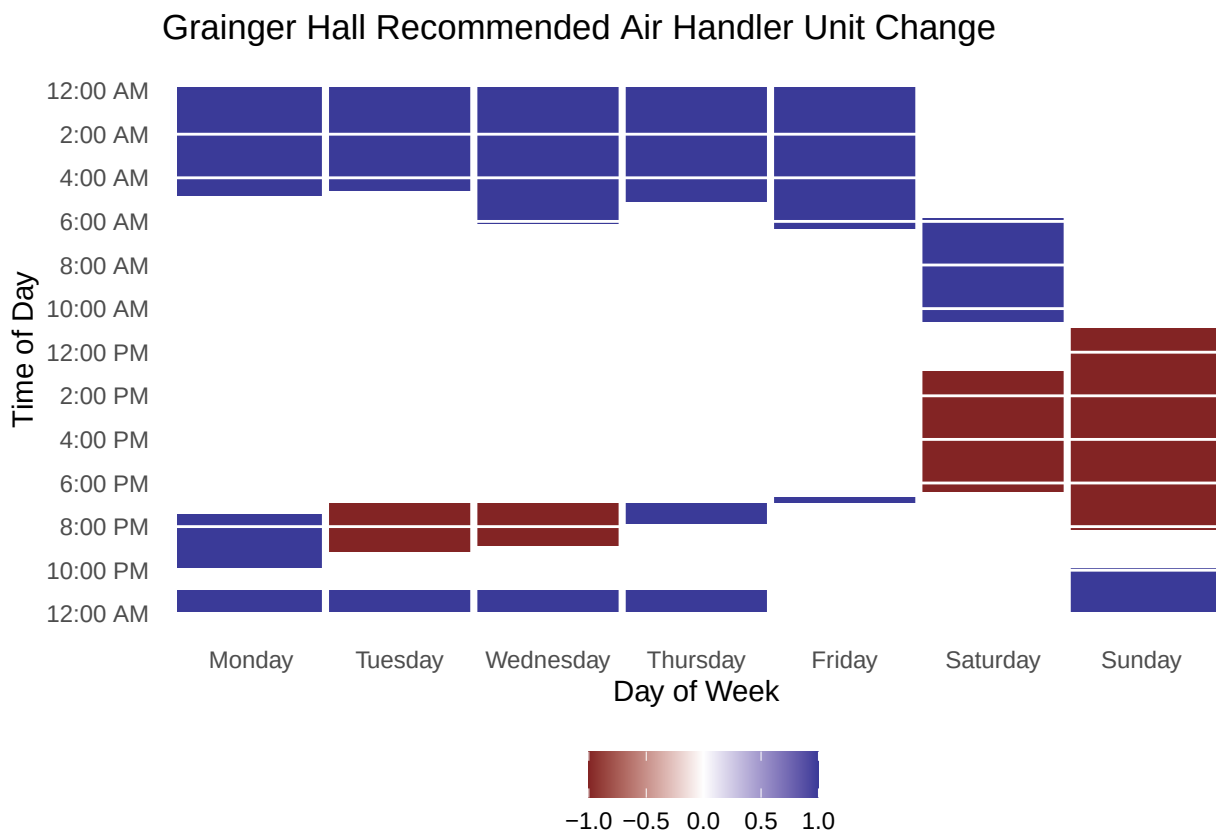


Determine extent of changes

```
newRecs = newRecs %>%
  mutate(changed = (smoothedOccText != Current.Schedule),
         AHUAv1 = ifelse(smoothedOccText == Current.Schedule, 0, ifelse(
           smoothedOccText == "AHU Off" & Current.Schedule != "AHU Off", 1, ifelse(
             smoothedOccText == "Wide Setpoints" & Current.Schedule != "AHU Off", 0, -1))))

plotWeek(newRecs, "changed", discrete = TRUE, altTitle = "Grainger Hall Recommended Change Times")
```

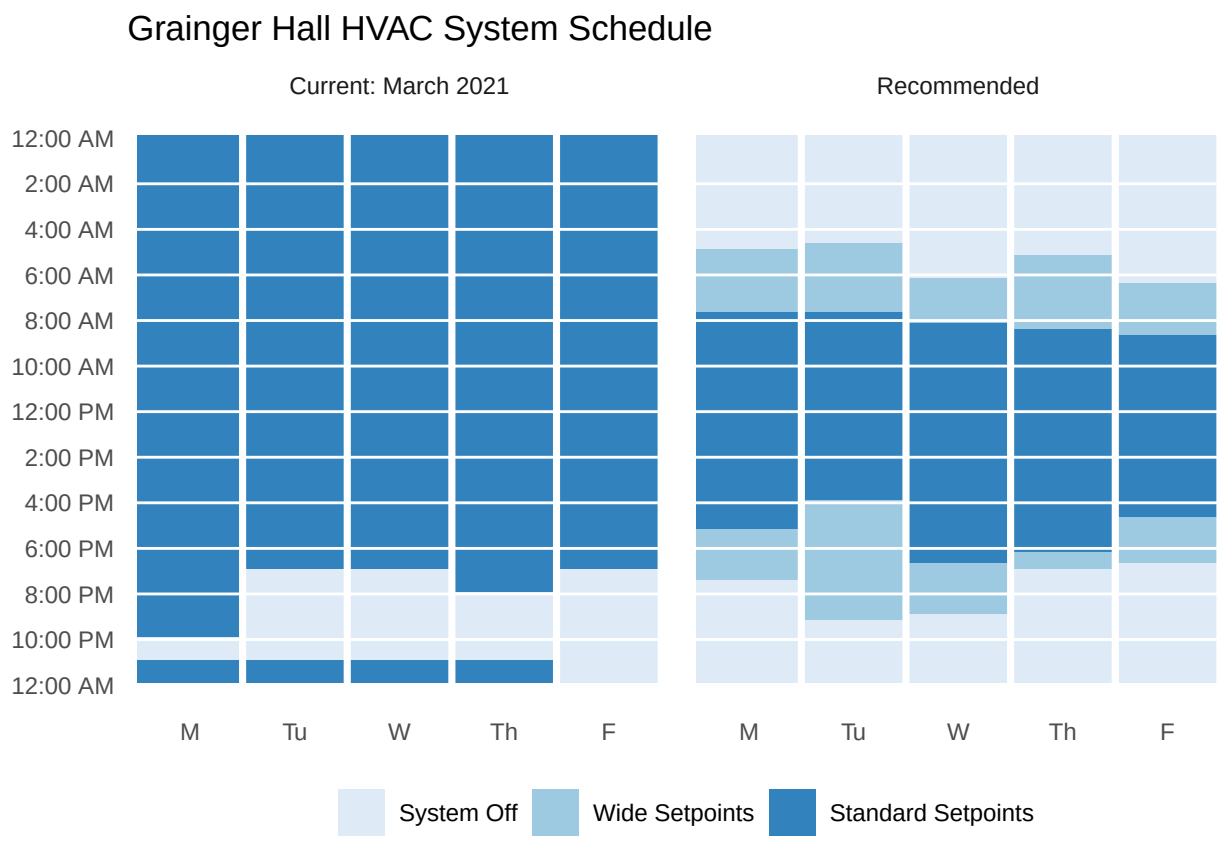



```
plotWeek(newRecs, "AHUAv1", altTitle = "Grainger Hall Recommended Air Handler Unit Change")
```

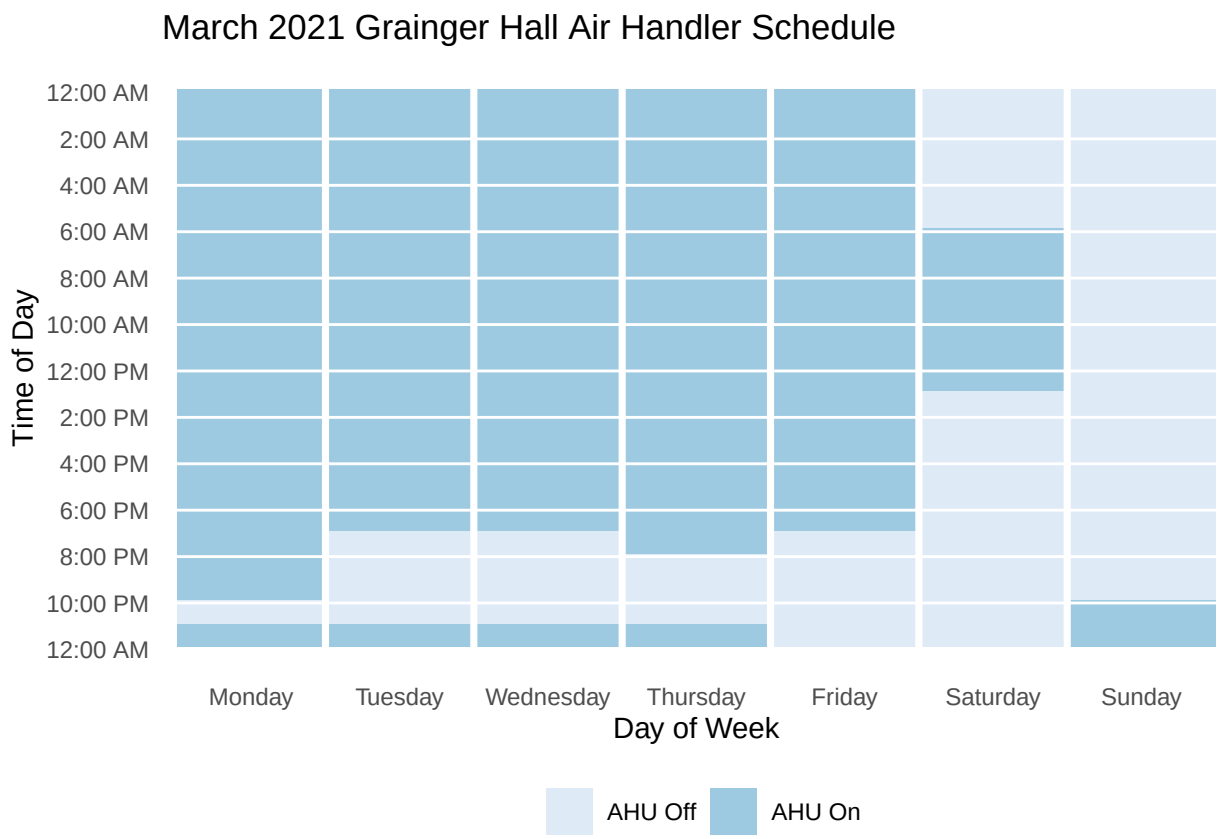



```
fiveDay = newRecs %>%
  filter(weekday != "Saturday" & weekday != "Sunday") %>%
  mutate(smoothedOccText = factor(smoothedOccText, levels = c("AHU Off", "Wide Setpoints", "Normal Opera
  mutate(Current.Schedule = factor(Current.Schedule, levels = c("AHU Off", "Wide Setpoints", "Normal Op

#plotWeekRec(fiveDay, "smoothedOccText", discrete = TRUE, altTitle = "Grainger Hall Recommended Operati

#plotWeekRec(fiveDay, "Current.Schedule", discrete = TRUE, altTitle = "Grainger Hall Current Operations

fiveDayPlot = fiveDay %>%
  select(weekday, hrSinceMidnight, smoothedOccText, Current.Schedule) %>%
  pivot_longer(c(smoothedOccText, Current.Schedule),names_to = "state", values_to = "mode")

state.lab <-  c("Current: March 2021", "Recommended")
names(state.lab) <- c("Current.Schedule", "smoothedOccText")

ggplot(fiveDayPlot, mapping = aes(x = weekday, y = hrSinceMidnight, fill = mode)) +
  facet_grid(cols = vars(state), labeller = labeller(state = state.lab)) +
  geom_raster()+
  scale_x_discrete(labels = c("M", "Tu", "W", "Th", "F"), name = NULL) +
  scale_y_reverse(breaks = seq(0,24, by = 2), labels = c("12:00 AM", paste0(seq(2, 10, by = 2), ":00 AM
                            "12:00 PM", paste0(seq(2, 10, by = 2), ":00 PM"), "12:00 AM"),
                  minor_breaks = NULL, name = NULL) +
  geom_vline(xintercept= seq(1.5, 6.5, by = 1), color="white", size=1.25) +
  geom_hline(yintercept= seq(2, 22, by = 2), color="white", size=0.5) +
  ggtitle("Grainger Hall HVAC System Schedule") +
```

```
  theme_minimal() +
  theme(legend.position = "bottom", panel.ontop = TRUE,
        panel.grid.major.x = element_blank(), panel.grid.major.y = element_blank()) +
  scale_fill_brewer(labels = c("System Off", "Wide Setpoints", "Standard Setpoints"), name = NULL)
```




```
plotCurrent <- newRecs %>%
  mutate(oldSched = factor(Current.Schedule)) %>%
  mutate(oldSched = revalue(oldSched, c("Normal Operation" = "AHU On")))



plotWeek(plotCurrent, "oldSched", discrete = TRUE, altTitle = "March 2021 Grainger Hall Air Handler Sche
```



Total avoided hours

```r
print(paste("Total avoided AHU hours:", sum(newRecs$AHUAv1)/4))
```

```
## [1] "Total avoided AHU hours: 23.25"
```

```r
print(paste("Weekday avoided AHU hours:", sum(fiveDay$AHUAv1)/4))
```

```
## [1] "Weekday avoided AHU hours: 31.25"
```

```r
print(paste("Total wider setpoint hours:", sum(newRecs$smoothedOccText == "Wide Setpoints")/4))
```

```
## [1] "Total wider setpoint hours: 42.25"
```

```r
print(paste("Weekday wider setpoint hours:", sum(fiveDay$smoothedOccText == "Wide Setpoints")/4))
```

```
## [1] "Weekday wider setpoint hours: 25.75"
```