EVALUATING THE CHANGES IN ENERGY EFFICIENCY RANKINGS OF COMMERCIAL BUILDINGS AFTER AN ENERGY RETROFIT

A Thesis
Presented to
The Academic Faculty

by

Pragadeesh Muthiah

In Partial Fulfillment
of the Requirements for the Degree
Master of Science in the
School of Civil & Environmental Engineering

Georgia Institute of Technology
August 2019

EVALUATING THE CHANGES IN ENERGY EFFICIENCY RANKINGS OF COMMERCIAL BUILDINGS AFTER AN ENERGY RETROFIT

Approved by:

Dr. John E. Taylor, Advisor
School of Civil & Environmental Engineering
Georgia Institute of Technology

Dr. Dennis R. Shelden
School of Architecture
Georgia Institute of Technology

Dr. Eric Marks
School of Civil & Environmental Engineering
Georgia Institute of Technology

Date Approved: July 25th 2019

ACKNOWLEDGEMENTS

I wish to thank my advisor Dr. John Taylor for helping me develop this thesis. I also want to thank Jessica Rose and Facilities Management for funding me by developing a joint GRA position with the school of Civil and Environmental Engineering. Working in this position with Facilities Management helped me get access to better data and facilities for research. I would like to thank my mentor Abby Francisco and NDL lab for helping me through the process of research. I would also like to thank my committee members Dr. Marks and Dr. Shelden for reviewing my thesis work. Finally, I would like to thank my family and friends for their continued support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

EUI	Energy Use Intensity
HVAC	Heating, Ventilation, and Air Conditioning
Georgia Tech	Georgia Institute of Technology, Atlanta Campus
Occ.sch	Occupied School Temporal segment
Occ.sum	Occupied Summer Temporal Segment
Unocc.sch	Unoccupied School Temporal segment
Unocc.sum	Unoccupied Summer Temporal Segment
Peak.sum	Peak Summer Temporal Segment
DF	Degrees of freedom

SUMMARY

Buildings account for more than 60% energy consumption in developed nations. Reducing natural resources and increasing cost of energy utilities makes it critical to reduce the energy consumption of buildings by making them more efficient. Buildings are complex, and it is very difficult to identify energy efficiency opportunities in them. This process becomes further difficult when managing a large portfolio of buildings. Energy retrofit is an effective method to improve the energy efficiency of a building. However, choosing an energy retrofit project, and quantifying the effectiveness of an energy retrofit project is still a challenging task.

Georgia Tech campus was used as a test bed for this thesis. This thesis evaluates the effectiveness of energy retrofit projects installed in buildings on Georgia Tech campus. The test sample was comprised of 36 buildings, with 5 buildings receiving an energy retrofit project between 2016 and 2018. The energy efficiency of 36 buildings was determined using a temporal segmented building energy benchmarking method developed by Francisco et al. [1], for two year-long periods, 2016 and 2018. The change in energy efficiency of this group of buildings between the two periods was determined as a part of this analysis and it was used to understand the effectiveness of an energy retrofit project.

The thesis tested the claim that buildings which received an energy retrofit project shows increased relative energy efficiency in 2018 compared to 2016. The thesis also tests the claim that buildings which did not receive an energy retrofit between the two periods do not show an increase in energy efficiency. The current research did not find enough evidence to support either of these claims. However, it is possible that these retrofitted buildings did not show an increase in energy efficiency due to rebound of other effects. In order to better understand the effectiveness of a retrofit project, the thesis evaluated the change in efficiency in different temporal segments. Evaluating efficiency in different temporal segments further helps in understanding the building efficiency. Comparing trends in EUI and efficiency change was another methodology used in this thesis to better understand the energy performance of a building.

The methodology used in this research follows a top-down approach to evaluate the effectiveness of an energy retrofit project, and provides techniques in identifying future prospects for retrofit projects. Such

a top down approach makes it easier to evaluate buildings in a large portfolio of buildings with less information. This methodology helps facility managers test the effectiveness of retrofit projects implemented and compare the returns against initial estimates. It also helps facility managers narrow the scope in identifying inefficient buildings with future prospective for energy retrofits.

CHAPTER 1. INTRODUCTION

Commercial building is an important sector which accounts for about 18% of the end-use energy in the US [3]. It is estimated that there is potential to improve efficiency for commercial buildings by 30% by 2030 [4]. One way to achieve this energy efficiency is by implementing energy efficiency measures such as retro-commissioning, operational changes, occupant behavioral changes, and energy retrofits, among the many available measures. Energy retrofits are a popular energy conservation measure implemented in commercial buildings to improve the energy efficiency of a building due to their high return on investment.

Building energy benchmarking is a method of comparing the energy efficiency of a building relative to a group of buildings. This thesis explores the changes in the energy efficiency of non-residential buildings on the Georgia Tech campus. Typical benchmarking methods produce benchmarks every year and use it to determine the energy efficiency of a building. However, yearly measurement of these benchmarks results in loss of important information regarding the fluctuation in the efficiency of buildings during different temporal segments in a year. The energy efficiency of commercial buildings this research uses is a temporal segmented building energy benchmarking methodology developed by Francisco et al [1]. This benchmarking methodology follows an approach which is useful in identifying the inefficiencies in a building by evaluating the building energy efficiency in specific temporal segments. Identification of these inefficiencies is important in decision making for top-level decision-makers. These inefficiencies in the building can be reduced using energy retrofit projects. Energy retrofits are investments with the goal of achieving higher energy savings potential. When managing multiple buildings, it is critical to manage the limited funds and identify appropriate energy retrofit projects to be implemented. In order to choose an energy retrofit, it is important to understand the effect of energy retrofits on the energy efficiency of buildings. The thesis evaluates the change in the energy efficiency of buildings after an energy retrofit project. The research focuses on 36 commercial buildings in Georgia Tech campus out of which 5 buildings had an energy retrofit project completed between 2016 and 2018. The thesis also highlights the importance of observing the changes in scores such as EUI (Energy Use Intensity), which measures the energy use per unit area, in combination with energy efficiency scores by exploring the relationship between change in EUI and change in the energy efficiency of buildings.

The thesis starts with a section on literature review. It covers the different energy performance assessments for buildings, and the relation between energy efficiency and energy retrofits. The next section of the thesis explains the different methods used for the analysis in this thesis and introduces the hypothesis tested. The thesis next covers the results of the analyses conducted. The discussion section next focuses on the interpretations of the results and elaborates on the outcomes of the analysis. The limitations and future steps section covers certain shortcomings of the research and the suggested improvements for future research in this area. Finally, the conclusion section summarizes the problem addressed in this thesis and explains the contribution of the research.

CHAPTER 2. BACKGROUND

2.1 Energy performance assessments

In order to identify energy efficiency opportunities in a building, a first step is understanding the energy performance of the building. Energy performance assessments are a common method to understand the energy performance of the building. Building energy classification is one such energy performance assessment which enables assessing multiple buildings at once [5]. Building energy classification helps to compare multiple buildings with different characteristics [6]. Building energy benchmarking is a method of energy classification which involves comparing the building against other relevant benchmarks [5]. Building energy benchmarking is a process that is carried out periodically to compare the energy efficiency of a building relative to other similar buildings in a group. Building energy benchmarking is a top-down approach and it typically tracks quantities such as energy consumption and EUI, using monthly billing data and some other physical information [7]. Some advantages of using energy efficiency benchmarking are that it typically requires simple monthly utility data and the end results are easier to understand for building owners [5].

Determining the energy efficiency of a building can be broadly carried out using three methods: white box method, gray box method, and black box method [8]. A white box method is a bottom-up approach to energy classification which relies on physical information such as design documentation to develop the energy model. The method relies on the use of the first principle of thermodynamics, energy balance. This methodology focuses on using maximum available physical information about the building envelope. It requires a lot of information and is highly effective in energy diagnosis or simulation at an individual building level. However, similar to most bottom-up approaches, such a method becomes infeasible at a large scale, when assessing multiple buildings. Black-box methods are a top-down approach method which relies on data fitting techniques and uses less physical building information in developing the energy model. When dealing with multiple buildings with less data, it is appropriate to use black-box methods as this requires much less physical information about the building. Such a method relies on the use of more granular data to develop a better model. A gray box method is a top-down approach which uses both physical information and data fitting to develop the energy model [8]. It combines both physical information and data fitting to create the energy model. This model takes advantage of available building physical information in developing an energy model. However, this method might be difficult to implement on a portfolio of buildings when physical information may not be known.

Gray box methods help when analyzing a large set of buildings when there is an availability of physical information such as space usage, building age, renovation, occupancy, etc., along with the building energy consumption data. Temporal segmented building energy benchmarking method, which was developed by Francisco et al. [1], is a gray box method and helps to analyze a large number of buildings at once to identify significant inefficiencies in certain buildings. Widespread adoption of smart meters in the US [9] makes easier access to more granular data. Georgia Tech campus has smart meter energy data available for most of the buildings and there is further availability of a database of physical information regarding these buildings. Availability of this combination of data makes the gray box methods a good fit for this thesis work. The benchmarking method uses a regression-based methodology, which was adopted from Chung et al. [10], to develop energy efficiency benchmarks.

The energy efficiency benchmarks that were developed as a part of the temporally segmented buildings energy benchmarking research showed fluctuations in the energy efficiency throughout the year. The energy efficiency for all the buildings in the model was significantly different than the total period in at least one of the temporal periods. It could also be seen that the energy efficiency of some buildings fluctuated a lot through the different temporal periods of the year. These observed fluctuations highlight certain inefficiencies that might exist in the buildings. Energy efficiency methods such as energy retrofits are a good solution to these inefficiencies and are a popular method to improve the energy efficiency of the buildings. Understanding the trends in energy efficiency during different temporal periods can help in evaluating the effectiveness of an energy retrofit project.

2.2 Energy efficiency and retrofits

In a large system of buildings such as universities, multiple buildings are being managed by an organization. Buildings have a variety of issues causing problems resulting in inefficiencies. These inefficiencies cause the building to consume more energy, which results in higher costs to the organization.

Every building may have different reasons for inefficiency, and it is critical to identify them. With limited availability of funds, it is critical to choose energy projects with higher returns on investment. The situation in these cases needs a multi-objective approach to making decisions. Karmellos et al. [11] used pareto analysis, a multi-objective approach using cost and energy savings as two objectives. Some current research has been done to identify energy retrofits at the individual building level [6] [12]. Some methods also used smart meter data at the individual building level to identify energy retrofits in high-performance buildings [13], but there has been limited research in identifying energy retrofits for a group of commercial buildings.

Energy retrofits include different kinds of measures used to improve the energy efficiency of the building by making changes to equipment, systems, or at the assembly level. Energy retrofits are a popular method for improving energy efficiency and they can help realize more savings than energy conservation measures such as operational changes, infiltration reduction, energy awareness programs, etc. which do not involve a retrofit [14]. Energy retrofits can broadly be classified as standard or deep retrofits [15]. Standard retrofits are investment measures at a smaller scale and include Lighting, Envelope, HVAC, and other system level changes [15]. HVAC and Lighting system retrofits are mostly preferred by owners as they tend to return higher savings [16].

Measuring, and quantifying the savings of an energy retrofit project is, however, a challenging task. Savings from energy retrofit projects are usually overestimated and the predicted savings do not match the observed savings [17] [18]. Effects such as rebound effects and, free rider effects are some other factors which make quantification of energy savings complex [19]. There are a variety of different methods currently being used to evaluate energy retrofits. Analyzing bills before and after an energy retrofit is a simple method to evaluate such projects [20]. Energy modeling and simulation is a common method used. However, working with an energy model and simulation is time intensive and is difficult to evaluate for a large set of buildings. Krati et al [21] used neural network methods to determine energy savings. The model was able to predict daytime peaks but was unable to predict certain evening peaks, and again shows the challenges in determining energy savings. The model however is a black-box method and relies less on the available physical information. Lee et al [22] also evaluated 18 different energy retrofit toolkits for commercial buildings, and most of them used smart meter data. The research compares different toolkits available for evaluating an energy retrofit. Some of the privately developed analytical tools help in quick evaluation of energy retrofits. However, the methodologies described lack a holistic analysis to identify if the energy retrofit tackles an appropriate inefficiency in the building. The benchmarking method used in this thesis allows a broader approach and helps to analyze the effects of energy retrofits on many buildings at once by developing energy efficiency scores. The aim of the current research is to test the effectiveness of energy retrofits in buildings across Georgia Tech, by calculating the energy efficiency scores before and after an energy retrofit.

CHAPTER 3. METHODS

This thesis was conducted at the Georgia Institute of Technology (Georgia Tech), Atlanta campus. There were 36 buildings included in the analysis. These buildings were chosen because their heating and cooling are powered by a district water loop, hence removing any bias created by using individual heating or cooling systems. The data used in the analysis measured the average power delivered-received for each building. The data was collected at a 15-min interval using smart meters. A sample of the smart meter data can be found in Appendix A.1.

3.1 Analysis Periods

The first part of the analysis involved the development of energy benchmarks for all the 36 buildings. The benchmarks for these 36 buildings were developed for two different year-long periods. The first period was from 2015-2016, and the second period was in 2018. The time period for the analysis can be classified into two broad periods, 2016, and 2018. The first period containing data ranging from 9/26/15 to 9/25/16, will be referred to as the year 2016 henceforth in this thesis. The second period containing data ranging from 1/1/2018 to 12/13/2018 will be referred to as the year 2018 henceforth in this thesis. In total, 331 days were chosen from the year-long periods. 34 days had to be removed due to missing data fields in certain buildings during different scattered periods. Each of these yearlong periods was further segmented into different temporal periods. The different temporal periods were occupied and unoccupied periods during the school period, occupied and unoccupied periods during summer, and peak summer period. Due to lack of accurate occupancy data, the occupied period was assumed from 8 AM to 8 PM, and the unoccupied period was assumed from 8 PM to 8 AM. School year and summer were defined based on the Georgia Tech campus academic calendar. The summer peak period is based on peak pricing by Georgia Power and ranges from 2 PM – 7 PM on weekdays between June 1st and August 30th. Further details about the temporal periods can be found in Table 1.

Table 1: Temporal Period Details

Time Period 2016			
Temporal Period	Source	State(s)	Days/times
Occupancy shifts	Building hours and consumption trends	Occupied	8AM – 8PM (M-F)
		Unoccupied	8PM – 8AM (M-F)
Seasonal shifts	GT school calendar	School year	9/26/15 – 5/7/16, 8/21/16 – 9/25/16
		Summer	5/8/16 – 8/20/16
Summer peak demand	Georgia Power	Peak billing demand	09/26/15 – 09/30/15, 2PM – 7PM 06/01/16 – 09/25/15, 2PM – 7PM
Time Period 2018			
Temporal Period	Source	State(s)	Days/times
Occupancy shifts	Building hours and consumption trends	Occupied	8AM – 8PM (M-F)
		Unoccupied	8PM – 8AM (M-F)
Seasonal shifts	GT school calendar	School year	1/8/18 – 4/24/18, 8/20/18 – 12/04/18
		Summer	5/14/18 – 7/25/18
Summer peak demand	Georgia Power	Peak billing demand	06/01/18 – 09/30/18, 2PM – 7PM

The data was then used to develop energy efficiency scores to perform daily energy benchmarking. The efficiency scores are developed for each temporal period. The current analysis focuses on the total period for both 2016 and 2018. However, it is important to understand that efficiency scores in other temporal segments might provide useful insights regarding the performance of a building under different conditions and it can be used to further understand the changes in a building's energy consumption. These benchmarks can change drastically when an energy retrofit is installed in a building. In order to understand the effect of energy retrofit on buildings, this thesis compared the energy efficiency scores in 2016, prior to energy retrofit projects, to 2018, after the energy retrofit projects were complete.

3.2 Efficiency Score Model

Energy efficiency scores help in identifying the energy performance of buildings. This score ranks the buildings based on their daily energy efficiency. These energy efficiency scores for buildings were developed using similar methodology as Francisco et al [1]. The first step towards the development of the efficiency score was the development of a regression model. In order to develop the regression model, the dependent variable used was daily average EUI of buildings. The independent variables help normalize the skewed energy use in buildings as a result of different features. The list of independent variables is given in Table 2.

Table 2: Dependent and Independent variables in regression model

Variable	Description
Independent Variables	Floor area
	Building age
	Years since renovation
	Number of floors
	Percent renovated
	Space type: Laboratory Wet
	Space type: Laboratory Dry
	Space type: Office
	Space type: Mechanical
	Space type: General
	Space type: Circulation
	Space type: Service
	Space type: Supply
	Spacetype: Classroom
	Space type: Study
	Space type: Special
Dependent Variable	Daily Average EUI

A multivariate linear regression method was used in this process. The regression model was then simplified into the form:

$$EUI_{norm} = EUI_o - EUI + a \tag{1}$$

here a is the intercept, EUIo is the measured EUI of the building, EUI is the predicted EUI based on the regression model, a is the model intercept, and EUI_{norm} is the building's normalized EUI. EUI_{norm} scores were created daily for all the temporal periods, and for all the buildings presented in the model. The higher the value of EUI_{norm}, the lower the energy efficiency of the building is and vice versa. The EUI_{norm} scores were then scaled between 0 and 1 for each day using the below equation:

$$efficiency = \frac{\left(1 - (EUI_{norm} - \min(EUI_{norm}))\right)}{(\max(EUI_{norm}) - \min(EUI_{norm}))} \tag{2}$$

This resulted in the development of normalized energy efficiency scores. The efficiency score of 1 denotes the most efficient building and an efficiency score of 0 indicates the least efficient building. These efficiency scores are then used to compare the performance of buildings in 2016 and 2018. The model was developed using the statistical tool R. The code used to evaluate the energy efficiency of buildings can be found in Appendix B.1.

3.3 Energy projects and temporally segmented buildings energy benchmarking

There was a total of 36 buildings analyzed using this process. Benchmarks were developed for all 36 buildings for two year-long periods. In order to understand the effect of energy projects on a building's performance, it was important to identify the buildings which received at least one energy retrofit, and the ones that did not receive any energy retrofit between Sep 25, 2016, and Jan 1, 2018. Of the 36 buildings that were included in this analysis, there were a total of 5 buildings which had an energy project that was completed between the two time periods. Table 3 contains a list of buildings that underwent an energy retrofit project and a broad classification of the type of project.

Table 3: Buildings with Energy retrofit between 2016-2018

Building Number	Building Name	Total Number of energy projects	HVAC, Controls, Chiller plant upgrades	Lighting
b031	Success Center	1	1	
b081	Howey School of Physics	3	3	
b101	Knight	1	1	
b144	Love Manufacturing Building (MRDC II)	3	2	1
b146	Petit Biotechnology Building	2	2	

Energy projects are intended to make buildings more efficient. Buildings which received an energy retrofit are expected to perform relatively more efficiently than other buildings in this sample which did not receive an energy project. This can indicate that a building which received an energy project between 2016 and 2018 might show increased efficiency score in 2018 versus 2016. However, buildings which did not receive an energy retrofit between 2016 and 2018 might show no change or decrease in energy efficiency in 2018 versus 2016. There is a possibility of decrease in the energy efficiency due to the relative nature of efficiency benchmarks. This research will test the following hypothesis:

Hypothesis 1: Buildings which received an energy retrofit project between 2016 and 2018 show a significant increase in energy efficiency during 2018 compared to 2016

Hypothesis 2: Buildings which did not receive an energy retrofit project between 2016 and 2018 do not show a significant increase in energy efficiency during 2018 compared to 2016

To test this, efficiency scores for the buildings were created for two year-long periods and were compared using a paired t-test. The method compares the mean of average daily benchmarks in 2016 against average daily benchmarks in 2018. The significance for t-test values is dependent on the p-value and the t-score. Further details regarding this are provided in the next sub-section.

3.4 Mean comparison using T-test

Both the hypotheses in this thesis compare the energy efficiency score in 2016 versus 2018. In order to compare the change in efficiency between the two years a two-tailed paired t-test was conducted over the distribution of daily efficiency scores for every building. The paired t-test is a method that is effective in identifying if there is a significant difference between the daily efficiency scores distribution in 2016 and 2018. The test analysis was performed using the statistical tool R, and the function t.test (x, y) was used for the comparison. The code used to conduct the complete analysis can be found in Appendix B.2. The first step is to identify if the results are significant. The t-score which comes as a result of the test gives an idea about how different both the distributions are. The magnitude of the t-score evaluates the difference in the two samples. A higher score indicates that the first distribution is more significantly different than the other. A negative t-score indicates that the building was more efficient in 2018 compared to 2016. Whereas a positive t-score indicates that the building was less efficient in 2018 compared to 2016.

In order to conduct the paired t-test between energy efficiencies in 2016 and 2018, the first step is calculating the differences between all the pairs. d is used to represent the differences. The mean of difference (d) is represented by m, and the standard deviation of the difference (d) is represented by s. The significance of differences is measured by testing how far the distribution of d is from 0. The t value for the t-test can be calculated using the below formula:

$$t = \frac{m}{s/\sqrt{n}} \tag{3}$$

Here m is the mean of difference (d), s is the standard deviation of the difference (d), and n is the size of d.

Significance of the t score depends on two factors, critical t-score, and p-value. For this analysis, a significance level of p less than 0.05 was used. The critical t-score is determined using p-value and degrees of freedom (df) = n-1. For example, if there are 331 days in the observation. Then degrees of freedom are 330. For a significance level of 0.05, the critical t-value for a two-tailed test is +/- 1.967.

3.5 Validation

3.5.1 Energy projects and temporally segmented buildings energy benchmarking

The temporally segmented building energy benchmarking method is an advanced building energy benchmarking method used in this research to understand the building energy performance. This benchmarking method was used in this research since simple units such as EUI tends to be biased and does not account for various features present in a building. However, simple quantities such as EUI can be combined with efficiency scores to better understand the energy performance of the building. This acts as a first step towards validation of the trends observed in the energy efficiency scores. This first step in validation is carried out by understanding the relation between EUI and energy efficiency of buildings between 2016 and 2018. This involved comparing the change in EUI versus the change in efficiency.

EUI has traditionally been used to measure the energy performance of buildings. However, EUI is biased and it masks the many other factors that also affect the energy use other than the area of the building. Reduction in EUI is considered an improvement in the building performance and usually indicates improved efficiency of buildings. However, it should also be noted that a reduction in EUI does not always mean that the building is more efficient. The benchmarking model used in this analysis shows the efficiency of buildings compared to the group. This means that is possible for a building to become less efficient even though it had a reduction in its EUI. Such a scenario is possible when the group of buildings in the model in an average had a much greater reduction in EUI than the building that is being compared. This trend in EUI can also be used to test the effectiveness of energy retrofits. The EUI change versus efficiency change was looked into closely to understand if there was a significant observation in the energy efficiency change corresponding to a significant observation in EUI change.

3.5.2 Energy efficiency change and temporal segments

In order to determine the change in EUI, a two-tailed paired t-test was performed between daily EUI in 2016 versus the daily EUI in 2018. The methodology followed for this t-test is similar to the one explained in section 2.4. The same set of 331 days used for the total period was used in this analysis. The next step involved examining the trend in significance of EUI change corresponding to the trend in significance of efficiency change. Based on the t-test, EUI change can have three possibilities: insignificant, significant increase, or significant decrease. Similarly, efficiency change can have three possibilities: insignificant, significant increase, or significant decrease. Hence, there are 9 possibilities for this comparison. This comparison was explored for the 5 buildings which received an energy retrofit.

Another simple test to compare the change in EUI and change in efficiency is using a correlation test. A Pearson's-correlation test was also conducted between the change in EUI versus the change in efficiency in 2016 and 2018. The first variable for this test X is a change in EUI and the second variable Y is a change in the efficiency. The equation is shown below:

$$X = EUI_{2018} - EUI_{2016}\ ; Y = Efficiency_{2018} - Efficiency_{2016} \tag{4}$$

A Pearson correlation test was run between X and Y. The result is a correlation coefficient, r, which can range from -1 to +1. A negative value indicates that an increase in X corresponds to a decrease in Y and vice versa. A positive value indicates that an increase in X corresponds to an increase in Y and vice versa. A value of 0 indicates that there is no observed relation between X and Y.

CHAPTER 4. RESULTS

The analysis results have been outlined in this section. The first part of the section details the regression model that was developed. The regression model was then used to evaluate the efficiency scores. The second part of this section shows the efficiency score results for both the periods 2016 and 2018. The third part of the section shows the results of t-test that compared the efficiency scores in 2016 versus 2018. The final section shows the results of validation showing the relation between EUI and efficiency.

4.1 Regression results

The regression model was developed for both the periods 2016 and 2018. Both the periods had 5 other temporal segments. It is important to understand the R-squared value of the regression model developed. A higher R-squared value indicates that the developed model is a better fit to the data. In this analysis, the regression model was developed for each day for each temporal period. The R-squared values of daily regression model have been visualized in Figure 1 and 2. The graph shows the density distribution of the R-squared value for different temporal segments.

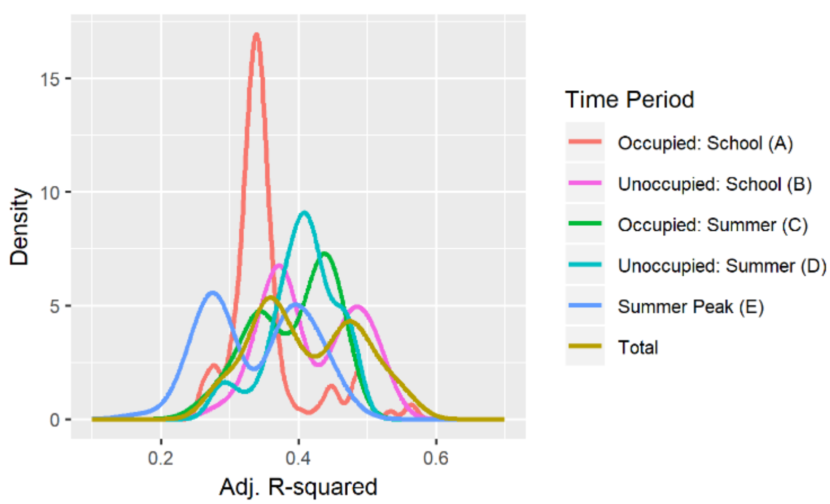


Figure 1: Regression model R-value for 2016

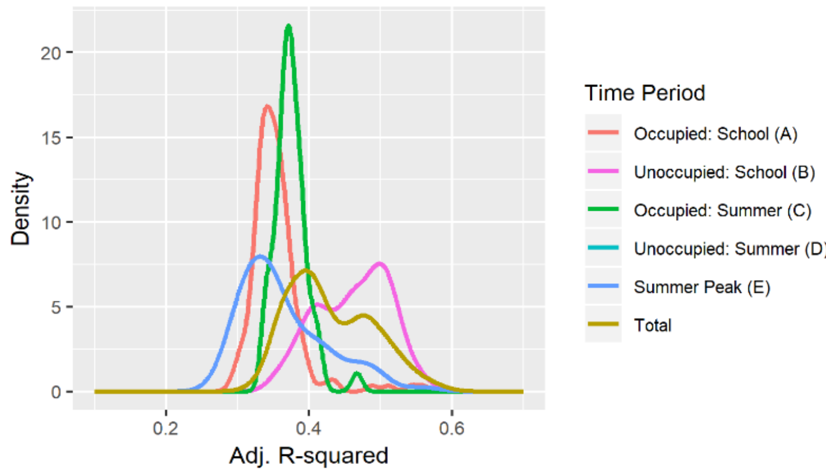


Figure 2: Regression model R-squared value for 2018

The average of R-squared value for temporal segments in 2016 ranged between 0.36 to 0.46. The average R-squared value for the total segment was 0.43. The average of R-squared value for temporal segments in 2018 ranged between 0.34 to 0.42. The average R-squared value for the total segment was 0.41. Additional results regarding the model can be found in Appendix A.2.

4.2 Efficiency score results

The Efficiency score was developed for each building based on the coefficients from the regression model developed for each day. This efficiency score was then used to rank the buildings every day. The rankings of the building were then normalized on a scale from 0 to 1 giving the normalized efficiency rankings, with 1 being the most efficient and 0 being the least efficient. The normalized ranking is also dependent on the magnitude of efficiency. This method of ranking highlights the difference in efficiency scores of the buildings. The current benchmarking model calculates these efficiency scores for 5 different temporal periods. This information can be used to identify inefficiencies in a building. Figure 3 shows the normalized efficiency score for building b002 in 2016 versus the normalized efficiency score for 2018. The current analysis only focuses on the changes in normalized efficiency rankings in the total period. It can be observed in Figure 3 that there is substantial variation in the change of normalized efficiency rankings observed by different temporal periods. Figure 3 also illustrates the variation in efficiency change for different temporal segments between the time periods. Further graphs for all the buildings can be found in Appendix D.

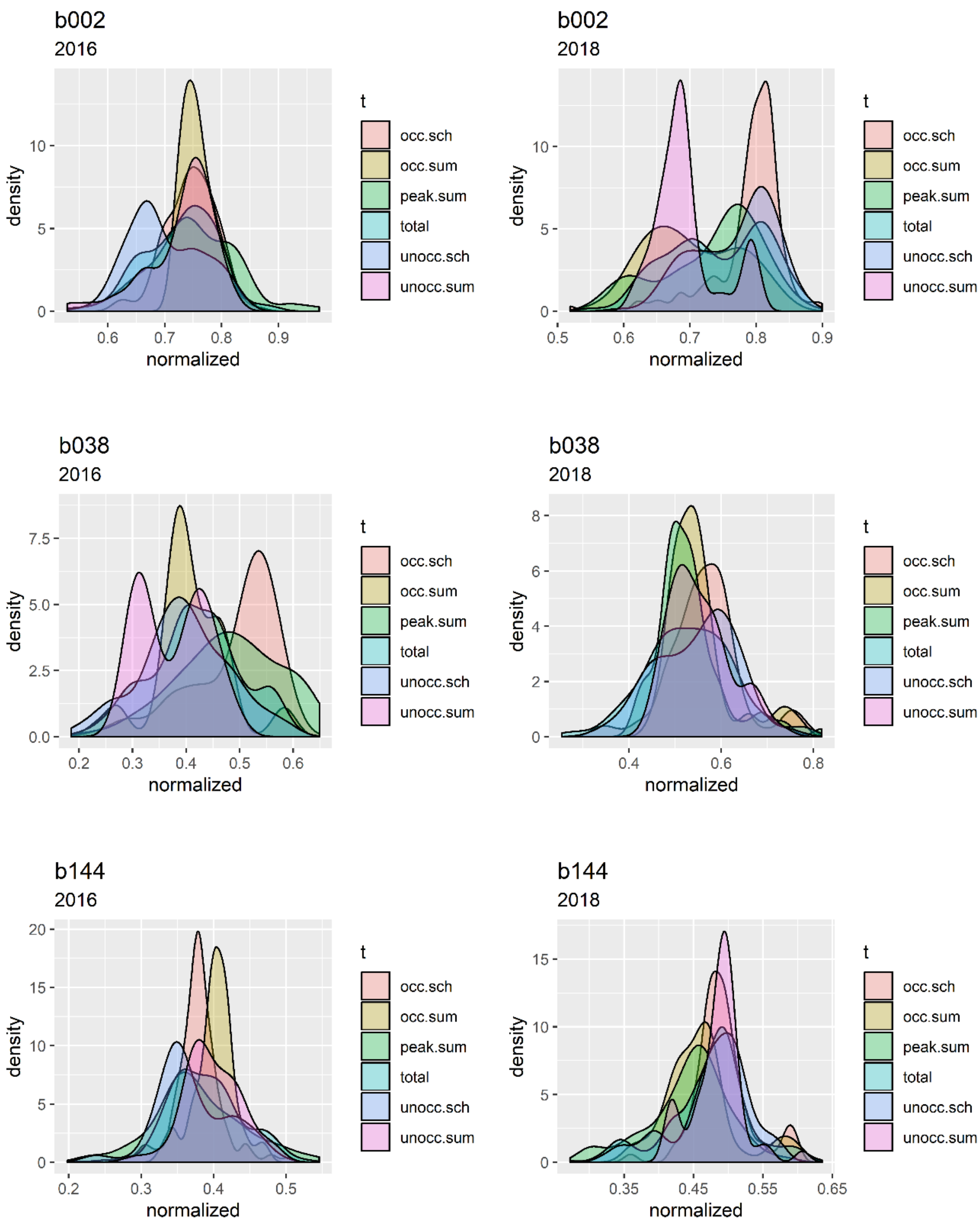


Figure 3: 2016 vs 2018: Energy Efficiency rankings for buildings b002, b038, and b144 for different temporal segments

4.3 T-test results by periods

The paired t-test compares the energy efficiency in 2016 versus normalized efficiency in 2018. The comparison included a total of 36 buildings and their normalized efficiency ranking was compared between both the periods. Energy efficiency scores were calculated for each day for all the 36 buildings included in the analysis. There were a total of 331 days included in the analysis. This results in a distribution of 331 normalized efficiency scores per building each year. The paired t-test compares the change in this distribution from period A to B. In order to conduct a paired t-test it is necessary to compare the observations from same day and month in both periods A and B. For example, normalized efficiency score for building b002 on 10/28/2015 can only be compared with normalized efficiency score on 10/28/2018. The results of the t-test are shown in Table 4. The degree of freedom (DF) for the data in table 5 is 330 which is given by the formula n – 1, where n is the number of observations. The p-value for significance is 0.05. In order to determine the significant cases, it is also necessary to consider the critical t value. The critical t value for significance depends on the p-value and the degree of freedom.

The critical T-Value (two-tailed) = +/- 1.967 (with p = 0.05 and DF = 330)

Table 4: Paired T-test 2016 vs 2018

Building Name	T-score	DF	P-Value	Efficiency change
b144	-20.8219	330	<0.001	-0.09192
b103	-19.8061	330	<0.001	-0.10757
b038	-15.9889	330	<0.001	-0.12714
b040	-9.2187	330	<0.001	-0.08952
b050	-6.48989	330	<0.001	-0.04816
b051	-5.91291	330	<0.001	-0.02503
b101	-5.40146	330	<0.001	-0.04454
b147	-4.96019	330	<0.001	-0.02937
b039	-4.88013	330	<0.001	-0.04328
b123	-4.44057	330	<0.001	-0.04178
b025	-4.08401	330	<0.001	-0.03576
b030	-3.7073	330	<0.001	-0.03633
b104	-2.57279	330	<0.05	-0.03468

b002	-2.35857	330	<0.05	-0.02435
b084	-2.28098	330	<0.05	-0.00666
b099	-1.89104	330	0.059475	-0.02717
b055	-0.47902	330	0.632231	-0.01879
b153	-0.40413	330	0.686374	-0.01876
b124	0.631635	330	0.528052	-0.0102
b111	1.944222	330	0.052698	-0.00016
b066	1.944223	330	0.052698	-0.00021
b045	2.020161	330	<0.05	0.000408
b031	3.039209	330	<0.01	0.004637
b036	3.551422	330	<0.001	0.012268
b058	4.867487	330	<0.001	0.013535
b135	4.967464	330	<0.001	0.012524
b075	5.16713	330	<0.001	0.016837
b061a	5.827163	330	<0.001	0.016343
b029	6.405894	330	<0.001	0.040925
b165	6.987121	330	<0.001	0.021962
b114	7.476982	330	<0.001	0.028848
b081	7.726059	330	<0.001	0.017085
b076	7.803699	330	<0.001	0.047474
b146	9.796013	330	<0.001	0.053095
b061	10.1286	330	<0.001	0.040434
b022	13.9529	330	<0.001	0.04744
	Total Significant Cases			30

Out of the 36 buildings present in the analysis 15 buildings showed a significant increase in energy efficiency, 6 buildings did not show any significant change, and 15 buildings showed a significant decrease in efficiency in period B compared to A. Both the hypotheses, 1 and 2, tested the effect of energy retrofit in building energy efficiency. The null hypothesis for hypotheses 1 and 2 are shown below with the results:

H_{01}: *Buildings that received an energy retrofit between period A and B does not show a significant increase in energy efficiency*

There were 5 buildings which received an energy retrofit. Out of these, 3 (60%) buildings showed a decrease in energy efficiency and 2 (40%) buildings showed a significant increase in energy efficiency. The null hypothesis was only rejected 2 out of 5 times.

H_{02}: Buildings that did not receive an energy retrofit between period A and B show a significant increase in energy efficiency

There were a total of 31 buildings which did not receive an energy retrofit between periods A and B. Out of these 13 (42%) buildings showed a significant increase in energy efficiency, and 18 (58%) buildings did not show a significant increase in energy efficiency. Out of these 18 buildings, 12 buildings showed a significant decrease in energy efficiency, and 6 buildings did not show any significant change in energy efficiency. The null hypothesis was only rejected 18 out of 31 times.

4.4 EUI and energy efficiency

The current analysis focuses on the change in efficiency for 36 buildings. It is also important to understand the relationship between the change in EUI and the change in efficiency. The t-test performed compared the distribution of EUI in 2016 versus EUI in 2018. The complete results of the t-test analysis for EUI can be found in Appendix C.2. Out of the 36 buildings present in the analysis 8 buildings showed a significant increase in the EUI, 4 buildings did not show any significant change, and 24 buildings showed a decrease in EUI in 2018 compared to 2016. Table 5 shows the trend in EUI versus efficiency for buildings which received an energy retrofit.

Table 5: EUI vs Efficiency in buildings with an energy retrofit

Building name	EUI t-score	EUI Trend	Efficiency t-score	Efficiency trend
b146	0.40113062	Not significant	9.796013	Significant decrease
b031	3.10986673	Significant decrease	3.039209	Significant decrease
b081	4.32127917	Significant decrease	7.726059	Significant decrease
b101	8.87817056	Significant decrease	-5.40146	Significant increase

b144	50.5306418	Significant decrease	-20.8219	Significant increase

In order to better understand the relationship between change in EUI and change in efficiency, a Pearson correlation test was conducted between change in EUI and change in efficiency for all the buildings. The results of the test are shown in Table 6. A detailed table can be found in Appendix C.3.

Table 6: EUI change vs Efficiency change

Building Name	X1	DF	Significance	Coefficient
b165	-13.9832	330	<0.001	-0.605945997
b050	-12.4564	330	<0.001	-0.561484791
b111	-12.4154	330	<0.001	-0.560218534
b055	-11.6873	330	<0.001	-0.537045879
b036	-8.54481	330	<0.001	-0.421991156
b061	-7.54623	330	<0.001	-0.38019967
b114	-7.21205	330	<0.001	-0.365659153
b103	-6.52787	330	<0.001	-0.33504348
b099	-5.41277	330	<0.001	-0.282814859
b040	-4.76395	330	<0.001	-0.251188718
b123	-4.61236	330	<0.001	-0.243677618
b124	-4.38898	330	<0.001	-0.232529668
b025	-3.83283	330	<0.001	-0.204380303
b135	-3.62545	330	<0.001	-0.193748881
b031	-3.56895	330	<0.001	-0.190840169
b081	-2.07738	330	0.038523582	-0.112444356
b147	-1.94221	330	0.05294501	-0.105211524
b022	-1.85725	330	0.064148479	-0.100656779
b058	-1.43323	330	0.152720171	-0.077835965
b051	-0.81306	330	0.416755391	-0.044247042
b144	-0.55911	330	0.576460551	-0.030442343
b045	0.776334	330	0.4380961	0.042251863
b146	0.994288	330	0.320796207	0.054083031
b084	1.735628	330	0.083543676	0.094125935
b038	1.870247	330	0.062316213	0.101354217
b101	2.778322	330	<0.05	0.149640766

b039	3.134273	330	<0.05	0.168299372
b029	3.134418	330	<0.05	0.168306925
b002	3.516453	330	<0.001	0.188132967
b061a	3.566559	330	<0.001	0.190716781
b066	3.908568	330	<0.001	0.208245505
b030	4.15088	330	<0.001	0.220545216
b076	5.039699	330	<0.001	0.264735014
b104	5.219391	330	<0.001	0.273479495
b075	5.870469	330	<0.001	0.304589827
b153	6.866727	330	<0.001	0.350346886

4.5 Energy efficiency change and temporally segments

Energy efficiency change was also calculated for different temporal segments during the period. There were five temporal segments – occupied school, unoccupied school, occupied summer, unoccupied summer, and peak summer. Building energy efficiency might be different in these temporal periods than the total period [1] and this gives a better idea about the effect of energy retrofits on building energy efficiency. Table 5 shows the number of significant cases using the t-test for energy efficiency change in each of the temporal periods.

Table 7: Energy efficiency change in temporal segments

Building	Total	Unoccupied Summer	Occupied Summer	Unoccupied School	Occupied School	Peak Summer
b144	Increase	Increase	Increase	Increase	Increase	Increase
b101	Increase			Increase	Increase	
b031	Decrease				Increase	
b081	Decrease		Decrease		Decrease	Decrease
b146	Decrease	Decrease	Decrease		Decrease	Decrease

The energy efficiency changes were closely examined for the 5 buildings which received an energy retrofit in order to understand their effect. Table 7 shows the trend of energy efficiency for these five buildings.

CHAPTER 5. DISCUSSION

Quantifying the savings from an energy retrofit project is a challenge. Savings from a retrofit project are usually exaggerated and often more than the realized savings [17,18]. The main objective of this study was to understand the effect of energy retrofits on the energy efficiency of buildings. This analysis focused on using a top-down approach in order to compare the energy efficiency of buildings in 2016 versus 2018. The top-down approach was used to gather as much information as possible for a large group of buildings with minimal information available. The change in the efficiency score of a building is either caused if there is a change in the efficiency of the building or if there is a change in the efficiency of another building. The normalized efficiency rankings are dependent on the performance of all the buildings. For example, a building's efficiency score can increase either if the building became more efficient or if other buildings became less efficient. Out of 36 buildings that were analyzed, 32 buildings had a significant change in energy efficiency between 2016 and 2018. The study explored if an energy retrofit in 5 of these buildings correlate with an improvement in the energy efficiency of a building. The hypothesis 1 tested the claim that energy retrofits cause an increase in the energy efficiency of a building. The null hypothesis 1 was rejected 2 times out of 5. The hypothesis 2 tested the claim that in the absence of an energy retrofit there will not be an increase in the energy efficiency of a building. The null hypothesis 2 was rejected 20 times out of 31.

5.1 Energy efficiency scores

In the t-test conducted, a positive t-score indicates that the building became less efficient relative to the group in 2018 versus 2016. There are a total of 17 buildings which had a positive t-score. These buildings became less efficient than the rest of the group. This set of buildings are very good candidates for the identification of future energy conservation projects. 3 buildings which had an energy retrofit between 2016 and 2018 fall into this category. This is possible if the energy retrofit did not target the inefficiency in the building with the potential for maximum energy savings. The decreased energy efficiency is also possible due to change in other factors. Buildings with energy retrofits may have an increase in energy consumption due to effects such as rebound effect [17]. Operational changes also heavily affect energy consumption. Buildings b146, b061, and b022 have a relative higher magnitude of t-score. This indicates that these buildings had a substantial decrease in their energy efficiency. Building b146 received an energy retrofit between 2016 and 2018 but it still has a decrease in its efficiency post the installation of the energy retrofit. This can indicate that the facility managers were able to identify that the building had an inefficiency but were not able to identify the exact inefficiency. However, building b146 received an HVAC retrofit and it is also possible that the effects of the energy retrofit are not observed in the electrical energy consumption.

In the t-test conducted, a negative t-score indicates that the building became more efficient relative to the group in 2018 versus 2016. 15 buildings showed a negative t-score. These buildings became more efficient than the rest of the group. 2 buildings among the 15 received an energy retrofit. Building b144 has a very high t-score, indicating that there is a significant difference in the normalized efficiency score. This can indicate that the energy retrofit project was targeted towards the appropriate inefficiency present in the building. Building b144 also received a lighting energy retrofit, which shows a direct reduction in electrical energy consumption. Buildings such as b103 and b038 show a substantial increase in the normalized efficiency score. These buildings did not receive an energy retrofit project. It is very important to understand this increase in efficiency score. However, current data does not show a possible cause for this increase in efficiency. It is possible to have this increase due to inexpensive alternative energy conservation methods such as operational changes, or other energy awareness programs. This also highlights the need to account for other factors influencing energy efficiency.

5.2 Role of energy retrofits

Energy retrofits are measures aimed at reducing the energy consumption of a building. It is expected that the installation of an energy retrofit reduces the energy consumption of a building by replacing an inefficient system with an efficient system. Hypothesis 1 tested the claim that energy retrofits improve the energy efficiency of a building. The null hypothesis H_{01} was rejected only for 2 out of the 5 buildings. This indicates that there was not enough evidence to conclude that energy retrofits cause an increase in energy efficiency. One possible reason that there was no improvement in the efficiency can be the relative nature of efficiency rankings used in the benchmarking model. The benchmarking model measures the efficiency of a building relative to other buildings in the group. It is possible that a building saw a decrease in efficiency rankings even if it had a slight improvement in its energy efficiency. Such a scenario can happen when the rest of the buildings in the group show a much greater improvement in energy efficiency. The current analysis only considered energy retrofits as the factor to affect the energy efficiency of buildings. It is possible that buildings without an energy retrofit showed an increase in energy efficiency due to other factors such as occupancy changes, or other energy awareness programs. Hypothesis 2 tested the claim that buildings which did not receive an energy retrofit did not show an increase in energy efficiency. There were a total of 31 buildings which did not receive an energy retrofit between periods A and B. Out of these 11 (35%) buildings showed a significant increase in energy efficiency, and 20 (65%) buildings did not show a significant increase in energy efficiency. Out of these 20 buildings, 11 buildings showed a significant decrease in energy efficiency, and 9 buildings did not show any significant change in energy efficiency.

Some possible reasons for the uncertainty can be attributed to the complex nature of building energy and the various factors that affect it. However, many other factors might affect the change in energy use that is observed. In order to understand the energy performance of a building, looking at factors other than energy efficiency can help in further understanding the energy performance of a building. EUI is one such quantity which is not difficult to quantify and was evaluated in combination with energy efficiency to understand the energy performance of buildings in the analysis.

5.3 Energy retrofits and temporal segments

Energy efficiency changes were also evaluated for different temporal segments for all the buildings in this analysis. A similar t-test was used to evaluate the change in efficiency in different temporal segments. This information can be used to further understand the effectiveness of an energy project. Out of the 36 buildings in this analysis, only building b144 had a significant increase in energy efficiency during all the temporal segments. The building b144 received 2 HVAC and 1 lighting energy retrofit project. Building b144 had an effective energy retrofit project as it shows a significant increase in the efficiencies in all the temporal periods. Identifying such building is critical for effective facility management while managing a large portfolio of buildings. Understanding the methods and projects used in this building would help facility managers evaluate and formulate best practices for future energy retrofit projects.

5.4 EUI change versus efficiency change

EUI is a very common unit to measure the energy performance of a building. It gives a measure of energy consumed per unit area. However, it is important to understand that EUI is not a good measure to understand the energy performance of a building. EUI does not reflect the effect of variables other than area. It is possible for buildings with a high EUI to be more efficient and a building with a low EUI to be less efficient. Table 4 shows the trend of EUI and efficiency for buildings which received an energy retrofit. It is observed that except building b146, 4 other buildings had a significant decrease in the EUI. However, this does not directly correlate with efficiency. It can be seen that 3 buildings had a significant decrease in energy and efficiency and 2 buildings had a significant increase in energy efficiency. Buildings b146, b031, and b081 had a smaller magnitude of change in EUI and showed a significant decrease in energy efficiency. Buildings b144 and b101 had a greater magnitude of decrease in EUI and showed a significant increase in energy efficiency. This, however, highlights the importance of looking at other indicators when evaluating the success of an energy retrofit project. EUI does not highlight the true performance improvement of a building and such a factor can only be identified by using advanced benchmarking methods. Buildings which had an increase in EUI are expected to show a reduction in energy efficiency, but this was not the case observed. There were 8 buildings which had an increase in EUI. 3 buildings showed a significant increase in energy efficiency, 1 building did not show a significant change in efficiency, and 4 buildings showed a significant decrease in energy efficiency.

Evaluating the changes in EUI is not enough to understand the energy performance of a building. It is generally expected that buildings which have a reduction in EUI have an increase in efficiency. However, energy efficiency using this benchmarking method shows the relative performance of a building compared to a group of buildings. The correlation test between change in EUI and change in efficiency shows a significant positive correlation in 11 buildings, significant negative correlation in 15 buildings, and no significant correlation in 10 buildings. Understanding the relationship between change in EUI and change in the efficiency of buildings can help in showing that it is not sufficient to evaluate just the EUI of a building. Examining other quantities such as energy efficiency in addition to EUI would help us better understand the changes in a building.

5.5 Contenders for future energy retrofit projects

This thesis focuses on the change in the efficiency of buildings during the total period in both the time periods. However, using the benchmarking method used in this thesis, it is possible to determine the energy efficiency scores in different temporal segments and evaluate its changes. A brief summary of the number of buildings which has a significant increase, significant decrease, or no significant change in energy efficiency between the two periods, in different temporal segments, can be found in Table 8. Detailed t-test results for efficiency change in each temporal segments can be found in Appendix C.1.

Table 8: Energy efficiency change by temporal segments

Temporal Segments	Significant increase	No significant change	Significant decrease
Total	15	6	15
Unoccupied Summer	7	17	12
Occupied Summer	9	14	13
Unoccupied School	11	20	5
Occupied School	13	14	9
Peak Summer	7	23	6

An application of this research is in aiding facility managers to identify buildings which are contenders for future energy retrofit projects. In order to identify future prospects for energy retrofit projects, buildings were identified using the efficiency change information and EUI change information. This identified buildings which had a significant decrease in efficiency during 2018 and also had a significant increase in the EUI. Shown in Table 9 are a set of buildings which had a significant decrease in energy efficiency and had a significant increase in the EUI.

Table 9: Buildings with future energy retrofit prospects

Building	Efficiency	EUI
b036	Significant decrease	Significant Increase
b075	Significant decrease	Significant Increase

B029	Significant decrease	Significant Increase
b165	Significant decrease	Significant Increase

CHAPTER 6. LIMITATIONS AND FUTURE STEPS

An important component of the current research was to understand the feasibility of conducting a deeper study related to energy retrofits and their effects on energy efficiency of a building. The current analysis had certain limitations associated with the regression model, energy profile, and the energy retrofits. This section explains the limitations of the current research and explains the future steps to address these.

6.1 Regression model

The average R squared value for the regression model was close to 0.4 for periods A and B. However, this score is lower than other regression analyses present in literature, with R-squared values ranging from 0.7 to 0.8 [1,10,23]. Higher R-squared indicates a better fit of the developed model to the actual data. This would aid in the development of better-normalized efficiency scores, which would help in better understanding of the building's performance. One of the main steps for future research would be improving the benchmarking model by focusing on improving the regression model. It is ideal to have an R-squared value greater than 0.7 consistent with the other research. The regression model may also be improved by using a different gray or black box method [24].

The regression model can also be made more robust by increasing the number of buildings present in the model. Including buildings which are not powered by district energy in the model can be explored in future research. The total number of buildings in the current analysis was 36. The current scope only involves 36 buildings powered by district energy to reduce the bias in energy consumed by energy systems. There are over 100 non-residential buildings on Georgia Tech campus. It can be very useful and significant to include all the buildings in the future analyses.

6.2 Energy profile

The current data analyzed includes only the electrical energy consumption. Electrical energy is a major part of the energy profile but other utilities such as chiller and steam systems have a major contribution to the energy consumption of a building. In order to understand the true change in the energy efficiency of buildings and the effect of energy retrofits, it is important to consider and analyze data from all the utilities. This is however limited by the availability of granular time series data of the different utilities. The energy consumption by the Georgia Tech campus consists of many different utilities. In the year of 2018 electric energy contributed to 58%, chilled water contributed to 19%, steam contributed to 11%, and other utilities contributed to 12% of the total energy consumption. At present only chilled water, steam, and electricity data are available in the 15-minute time interval for all the buildings on Georgia Tech campus. However, the current analysis only consists of electricity data, which contributes to less than 60% of the total energy. Including chilled water and steam data in the model will increase the total energy profile to 90%, which can provide meaningful insights in understanding the overall energy trends in the building. It is also important to consider operational changes in the analysis and the effects of operational changes.

The current study only analyzes the electricity consumption data. It is important to analyze other important utilities such as heating and cooling which are an important component of the energy consumption of a building. HVAC systems consume a major fraction of energy in building [25]. Smart meter data is available for steam, and chiller in most of the buildings on Georgia Tech campus. It is, however, challenging to work with the chiller and steam data due to frequent errors and the presence of a large number of outliers in the dataset. The next step forward would be evaluating the usability of the chiller and steam data collected from the smart meters on the Georgia Tech campus. Another possible alternative would be evaluating electrical consumption for all the commercial buildings in Georgia Tech campus and understanding the effects of district energy in the building energy profile.

6.3 Energy retrofits

Another limitation of the study was the number of buildings which received an energy retrofit project. There were only 5 buildings with an energy retrofit in the current set of 36 buildings which received an energy retrofit between 2016 and 2018. It is also important to understand the type of energy retrofit and its effect on the energy profile. Lighting retrofits consume electrical energy, and installing such retrofits will show more changes to electrical energy consumed. Energy retrofits such as HVAC upgrades will show more effects on chilled water energy consumption. It is highly possible that the change in the energy for certain buildings was not reflected in the current analysis as this research only considered electrical energy. The 5 buildings in this analysis had a mix of different types of energy retrofits, which made it further difficult to understand the effects of a specific type of energy retrofit in the energy efficiency of buildings. It is important to include more buildings which received an energy retrofit project. There are more than 30 buildings which have received an energy project in the past few years. The next step would be to consider including most of these 30 buildings in the model. This would help us better understand the effects of specific type of retrofit on energy efficiency of a building.

The current analysis only examined the trend in total period of 2016 versus 2018. There is less inclusion of temporal periods in our current analysis. Trends in temporal periods can further help in understanding the energy performance of a building. An important next step would be to understand the trend of efficiency scores in these temporal periods. In the current analysis energy retrofits was the only factor considered which could influence the energy efficiency of the buildings. Many other factors such as rebound effects, occupancy changes, operational changes, energy conservation awareness programs can affect the energy efficiency of the building. It is also important to consider such important factors to correctly evaluate the driver of energy efficiency change in the buildings and understand the contribution of energy retrofits in energy savings. Including these factors and understanding their effects on the energy efficiency is a difficult task and an important next step would be to explore the literature to identify methods to adjust for these factors.

CHAPTER 7. CONCLUSION

When it comes to facility management, identifying an energy inefficiency in a commercial building can be a challenging task. It is a further challenging task to find these inefficiencies when managing a large group of buildings. Current methods to find these inefficiencies are either too time-consuming such as energy diagnosis or are insufficient in analyzing multiple buildings at once. Temporally segmented building energy benchmarking methodology developed by Francisco et al [1] has been used in this thesis to identify these inefficiencies. This benchmarking method follows a top-down approach and is effective in finding inefficiencies in buildings by evaluating the trends of efficiency scores in different temporal segments. This method may help facility managers in identifying suitable energy retrofit projects to be installed in a building. In order to further validate the applications of the benchmarking methodology, it is important to understand the effects of energy retrofits on the energy efficiency of buildings. This thesis focused on evaluating the effects of energy retrofits on the energy efficiency of a group of non-residential buildings.

There were 36 buildings which were part of the analysis and 5 buildings received an energy retrofit. The research focused on evaluating the changes in the energy efficiency of a building after an energy retrofit. The current analysis followed a top-down approach to energy benchmarking. The benchmarking model developed a daily normalized efficiency score for a building. This efficiency score measures the daily relative efficiency of the building with respect to the other 36 buildings in the group. Efficiency scores were developed for two periods, 2016 and 2018. The analysis did not find sufficient evidence to demonstrate that there was a significant increase in energy efficiency of buildings which received an energy retrofit. It was also observed that there were multiple buildings which had a significant increase in energy efficiency without an energy retrofit project. This highlights the important fact that it is also necessary to understand other factors which could have driven the efficiency change of a building. However, this can also indicate that buildings which received an energy retrofit were not evaluated holistically before the installation of the energy retrofit project. It is possible that the energy retrofits installed are not targeted at specific important inefficiencies present in a building. It was also observed that all the 5 buildings which received an energy retrofit between 2016 and 2018 did not have a significant increase in the EUI, but it was found that 3 out the

5 buildings had a significant decrease in the energy efficiency. Energy retrofits were found to prevent a significant increase in the EUI of a building but the current study did not find enough evidence to show that the installation of energy retrofits cause a significant increase in the energy efficiency of a building relative to the group. This shows that when looking at the effectiveness of an energy retrofit project in a group of buildings, many energy retrofit projects are ineffective in increasing the energy efficiency of a building, and they need to be further targeted towards specific inefficiencies present in these buildings.

Commercial buildings are very complex and are made of multiple systems. Building systems tend to become inefficient with time and consume more energy. Identifying these inefficiencies can be a challenging task and can be further challenging for facility managers and decision makers when evaluating multiple buildings at once. It is critical to identify and fix these inefficiencies as they greatly help in reducing the energy consumption of the building and reducing the costs associated with it. This is particularly important with reduction in the availability of resources and increasing energy costs. Energy retrofit can be an effective method in targeting these energy inefficiencies. Comparing the energy efficiency of buildings over a period of time using a top-down approach such as temporal segmented building energy benchmarking helps us in understanding the effects of these energy retrofits on the energy efficiency of a building. Evaluating the energy efficiency scores in different temporal segments provides facility managers with additional information regarding the effectiveness of an energy project. This helps in developing best practices for future retrofit projects. Using the combination of EUI and efficiency methodology mentioned in this thesis aids in identifying future prospects for retrofit projects. This helps in evaluating the effectiveness of an energy retrofit project and further helps facility managers in decision making while choosing an appropriate energy conservation project for a building.

APPENDIX A. REGRESSION MODEL DEVELOPMENT

A.1 Smart Meter Data

Smart Meter Data for Meter 022_MH1 for 2 hours

TimestampUTC2	TimestampUTC	Active Energy Delivered-Received
2013-01-02 00:00:00-05:00 EST	1/2/2013 0:00	2986752.5
2013-01-02 00:15:00-05:00 EST	1/2/2013 0:15	2986767
2013-01-02 00:30:00-05:00 EST	1/2/2013 0:30	2986781.75
2013-01-02 00:45:00-05:00 EST	1/2/2013 0:45	2986796.25
2013-01-02 01:00:00-05:00 EST	1/2/2013 1:00	2986810.75
2013-01-02 01:15:00-05:00 EST	1/2/2013 1:15	2986825.25
2013-01-02 01:30:00-05:00 EST	1/2/2013 1:30	2986839.75
2013-01-02 01:45:00-05:00 EST	1/2/2013 1:45	2986854.25
2013-01-02 02:00:00-05:00 EST	1/2/2013 2:00	2986869

A.2 R value for Regression Model

R Value for 2016

Occupied: School (A)	Unoccupied: School (B)	Occupied: Summer (C)	Unoccupied: Summer (D)	Summer Peak (E)	Total
0.3517493	0.4173438	0.3920509	0.40514	0.3389913	0.4123122

Occupied: School (A)	Unoccupied: School (B)	Occupied: Summer (C)	Unoccupied: Summer (D)	Summer Peak (E)	Total
0.0576537	0.0663034	0.0580615	0.0493095	0.0726831	0.0772041

R Value for 2018

Occupied: School (A)	Unoccupied: School (B)	Occupied: Summer (C)	Unoccupied: Summer (D)	Summer Peak (E)	Total

Occupied: School (A)	Unoccupied: School (B)	Occupied: Summer (C)	Unoccupied: Summer (D)	Summer Peak (E)	Total
0.0315977	0.0374303	0.0504409	0.0227101	0.0603181	0.058634

APPENDIX B. CODE WRITTEN IN R

This section contains all the codes written in R for the analysis. Some parts of the code for cleaning data, segmenting data, plots, and regression model development were adapted from code developed by Francisco et al. [1].

B.1 Efficiency Score Development

```
#import data

data  <- read.csv("./spacetype/spacetype.csv",

        stringsAsFactors = FALSE)


#code used to convert some cases where the use case also has codea associated with it


library(readr)

updata <- parse_number(data$Use.Code)


data$Use.Code <- updata


#classification based on FICM codes

data$word_CODE  <- ifelse(data$Use.Code>=100 & data$Use.Code< 200, "Classrooms",
                          ifelse(data$Use.Code>= 200 & data$Use.Code< 225,"Laboratory Dry",
                          ifelse(data$Use.Code>= 225 & data$Use.Code< 300,"Laboratory Wet",
                          ifelse(data$Use.Code>= 300 & data$Use.Code< 400,"Office",
                          ifelse(data$Use.Code>= 400 & data$Use.Code<
                          500,"Study",ifelse(data$Use.Code>= 500 & data$Use.Code< 520,"Special
                          Use",ifelse(data$Use.Code>= 520 & data$Use.Code< 530,"Special Use
                          Athletics",ifelse(data$Use.Code>= 530 & data$Use.Code< 600,"Special Use",
                          ifelse(data$Use.Code>= 600 & data$Use.Code< 700,"General
                          Use",ifelse(data$Use.Code>= 700 & data$Use.Code<
                          740,"Support",ifelse(data$Use.Code>= 740 & data$Use.Code< 750,"Support
                          Parking Deck",ifelse(data$Use.Code>= 750 & data$Use.Code<
                          800,"Support",ifelse(data$Use.Code>= 800 & data$Use.Code< 900,"Health
                          Care",ifelse(data$Use.Code>= 900 & data$Use.Code<
                          1000,"Residential",ifelse(data$Use.Code>= 50 & data$Use.Code< 60,"Inactive
                          Area",ifelse(data$Use.Code>= 60 & data$Use.Code< 70,"Alteration
                          Area",ifelse(data$Use.Code>= 70 & data$Use.Code<
                          80,"Unfinished",ifelse(data$Use.Code>= 10 & data$Use.Code<
                          20,"Custodial",ifelse(data$Use.Code>= 20 & data$Use.Code< 30,"Circulation
                          Area",ifelse(data$Use.Code>= 30 & data$Use.Code<
                          40,"Mechanical",ifelse(data$Use.Code>= 40 & data$Use.Code< 50,"Office",
                          "NAB"))))))))))))))))))))))
```

```
b <- aggregate(data=data, Inside.Wall.Area~Facility+word_CODE, sum)

c <- reshape(b, idvar="Facility", timevar="word_CODE",direction="wide")

newdata <- c[ which(c$Facility=='002' ,) ]

write.csv(c,file='./spacetype/FICM_sorted_details_check.csv')
```

Clean-Data 2016

```
# import libraries
library(tidyverse)
library(lubridate)
library(data.table)
library(padr)
library(plotly)
library(zoo)

# import functions
source("./R/01_clean-data_functions.R")

# define daylight savings times
EST <- c("2015-11-01 05:00:00","2015-11-01 05:15:00","2015-11-01 05:30:00","2015-11-01 05:45:00")

# create cleanData column
start <- as.POSIXct("2015-07-01 01:00", tz= "UTC")
end <- start + as.difftime((365*1)+1, units = "days")
timestamp <- seq(from=start, by = 15*60, to = end)
timestamp <- timestamp[1:length(timestamp)-1]
clean <- data.table(DateTime = as.POSIXct(format(timestamp,tz =
"America/New_York",usetz=TRUE),tz="America/New_York"))

# clean electricity data
# define folders based on where utility data is stored
```

```
folders <- list.dirs(path = "./data", full.names = TRUE)

files <- list.files(path=folders, pattern = "*.csv", full.names=T, recursive=FALSE)


# set correct column names

colNames <- c("002a", "002b", "022", "025", "029", "030", "031", "036", "038", "039", "040", "045",
"050a", "050b", "051", "055", "058", "059", "061", "061a","066", "075", "076", "081", "084", "099", "101",
"103a", "103b", "103c", "103d", "104a", "104b", "111", "114", "123", "124","135", "144a", "144b", "146",
"147a", "147b", "147c", "153a", "153b", "165a", "165b")


# clean electricity data

g <- 1

for (i in files) {

  clean[, paste0("b",colNames[g]) := cleanElec(i,"20150630 2100","20160630 2045")]

  g <- g + 1

}


# combine double meters

clean[, b002 := b002a + b002b][,c("b002a","b002b") := NULL]

clean[, b050 := b050a + b050b][,c("b050a","b050b") := NULL]


clean[, b103 := b103a + b103b + b103c + b103d][,c("b103a","b103b","b103c","b103d") := NULL]

clean[, b104 := b104a + b104b][,c("b104a","b104b") := NULL]

clean[, b144 := b144a + b144b][,c("b144a","b144b") := NULL]

clean[, b147 := b147a + b147b + b147c][,c("b147a","b147b","b147c") := NULL]

clean[, b153 := b153a + b153b][,c("b153a","b153b") := NULL]

clean[, b165 := b165a + b165b][,c("b165a","b165b") := NULL]



# find NA, zeros, negatives, and outliers

m <- data.table(t(clean[, lapply(.SD, function(x) sum(x < 0, na.rm = T)), .SDcols = 2:38]))

clean[clean < 0] <- NA

m[, V2 := t(clean[, lapply(.SD, function(x) sum(x == 0, na.rm = T)), .SDcols = 2:38])]

m[, v3 := t(clean[, lapply(.SD, function(x) sum(x > (mean(x,na.rm = T)+(IQR(x,na.rm = T)*20)), na.rm =
T)), .SDcols = 2:38])]

attach(clean)
```

```
for (i in names(clean)[2:length(clean)]) {
  y <- get(i)
  thresh <- (mean(y,na.rm = T)+(IQR(y,na.rm = T)*20))
  y[y > thresh] <- NA
  clean[,c(i) := y]
}
detach(clean)
m[, V4 := t(clean[, lapply(.SD, function(x) sum(is.na(x))), .SDcols = 2:38])]
setnames(m, c("negative", "zero", "outlier","na"))
n <- names(clean)[2:length(clean)]
m[, name := n]

#outlierKD(clean,b055)
#plot(clean$DateTime,clean$b153)

# subtract/add submeters
clean[, b059 := NULL]

clean[, b099a := NULL]#cleanData this is repeating with b099

# interpolation, max gap interperlated is 6 hours (linear)
clean2 <- clean[, lapply(.SD, function(x) na.approx(x, maxgap = 24)), .SDcols = 2:37]
m <- data.table(t(clean2[, lapply(.SD, function(x) sum(is.na(x)))]))
m[, V2 := t(clean2[, lapply(.SD, function(x) sum(x == 0, na.rm = T))])]
n <- names(clean2)
m[, name := n]
clean2[,DateTime := clean$DateTime]

write.csv(clean2, "./test/cleanElec_2015.csv",row.names = F)
write.csv(m, "./test/missingDataDetail_2015.csv",row.names = F)
```

Clean Data Functions

```
# clean electricity data file
```

```
  # start/end format = "20150101 0000"
  # electricity units must be in kWh
cleanElec <- function(file,start,end) {
  column_names <- c("DateTimeUTC","energy")  # creates variable with column names
  e <- fread(file, select = c(2:3))   # read files and specify columns
  setnames(e,column_names)   # labels columns in the file
  e[, kWh := energy - shift(energy, fill=first(energy))]   # subtract cumulative energy
  e[, DateTimeUTC := as.POSIXct(e$DateTimeUTC, tz="UTC")]
  e[, time := DateTimeUTC - shift(DateTimeUTC, fill=first(DateTimeUTC))]   # subtract cumulative
energy
  if ((sum(e[,time] < 900)) > 8) {
    e <- e %>%
      thicken('15 min')
    e[, DateTimeUTC := DateTimeUTC_15_min]
    e[, DateTimeUTC_15_min := NULL]
  }
  e <- e %>%
    pad(start_val = ymd_hm(start, tz = "UTC"),end_val = ymd_hm(end, tz = "UTC"))  # fill in DateTime
gaps (gets filled with NA)
  e[, DT := as.character(DateTimeUTC)]
  setkey(e,DT)
  e <- e[!EST]
  return(e[,kWh])
}


outlierReplace = function(dataframe, cols, rows, newValue = NA) {
  if (any(rows)) {
    set(dataframe, rows, cols, newValue)
  }
}


outlierKD <- function(dt, var) {
  var_name <- eval(substitute(var),eval(dt))
  na1 <- sum(is.na(var_name))
```

```
  m1 <- mean(var_name, na.rm = T)
  par(mfrow=c(2, 2), oma=c(0,0,3,0))
  boxplot(var_name, main="With outliers")
  hist(var_name, main="With outliers", xlab=NA, ylab=NA)
  outlier <- boxplot.stats(var_name)$out
  mo <- mean(outlier)
  var_name <- ifelse(var_name %in% outlier, NA, var_name)
  boxplot(var_name, main="Without outliers")
  hist(var_name, main="Without outliers", xlab=NA, ylab=NA)
  title("Outlier Check", outer=TRUE)
  na2 <- sum(is.na(var_name))
  cat("Outliers identified:", na2 - na1, "n")
  cat("Propotion (%) of outliers:", round((na2 - na1) / sum(!is.na(var_name))*100, 1), "n")
  cat("Mean of the outliers:", round(mo, 2), "n")
  m2 <- mean(var_name, na.rm = T)
  cat("Mean without removing outliers:", round(m1, 2), "n")
  cat("Mean if we remove outliers:", round(m2, 2), "n")
  response <- readline(prompt="Do you want to remove outliers and to replace with NA? [yes/no]: ")
  if(response == "y" | response == "yes"){
    dt[as.character(substitute(var))] <- invisible(var_name)
    assign(as.character(as.list(match.call())$dt), dt, envir = .GlobalEnv)
    cat("Outliers successfully removed", "n")
    return(invisible(dt))
  } else{
    cat("Nothing changed", "n")
    return(invisible(var_name))
  }
}
```

Segment Data 2016

```
# Import ------------------------------------------------------------------

# import libraries

library(data.table)

library(tidyverse)

library(lubridate)

library(dygraphs)

library(xts)  # for plotting with dygraphs


# import data

data <- fread("./test/cleanElec_2015.csv")[, DateTime := as.POSIXct(DateTime)]


#data.remove <- c("b099")

#data <- select(data, -data.remove)


feat <- fread("./output/featPct2.csv")


#feat <- fread("./output/featPct3.csv")


# import functions

source("./R/02_segment-data_functions.R")


# data prep for temporal segmentation -------------------------------------

# clean column names and data types

setcolorder(data, c(data[, sort(names(data))])) #order matters for EUI calcs

colNames <- c("building","area","service","circ","class","general","lab_dry",
"lab_wet","mech","office","res","special","study","supply",

        "age","reno","pctReno","floors","use")

setnames(feat,colNames)

feat[, area := as.numeric(area)][,res := NULL]  # remove residential buildings because they are removed
froms sample


# specify any buildings to remove from analysis

#buildings.remove <- c("b166")
```

```
#data <- select(data, -buildings.remove)
#feat <- filter(feat, building != buildings.remove) %>% as.data.table()

# divide energy by building area
area.vector <- feat[,area]
eui <- sweep(data[,1:(length(data)-1)],MARGIN=2,FUN="/", STATS=area.vector)
data <- data.table(cbind(DateTime = data[,DateTime], eui))

# OPTIONAL: graph energy use
x <- as.xts(data[,c("DateTime","b153")])
dygraph(x) %>% dyRangeSelector()

# temporal segmentation ----------------------------------------------------
#add in Date and Time column
data <- data %>%
  mutate(day_of_week = wday(DateTime, week_start = getOption("lubridate.week.start", 1)),
         weekday = ifelse(day_of_week < 6, "yes", "no"),
         hour = hour(DateTime),
         date = date(DateTime),
         time = as.numeric(hm(strftime(DateTime, format = "%H:%M")))/60,
         working_hours = ifelse(weekday == "yes" & between(time, 480, 1200), "yes", "no"))

# now create energy slice labels for -- occ.sch, unocc.sch, occ.sum, unocc.sum, peak.sum
school <- data %>%
  filter(weekday == "yes" & (DateTime %between% c("2015-08-21 01:00", "2015-09-14 14:01") |
                             DateTime %between% c("2015-09-23 03:29", "2015-10-17 12:46") |
                             DateTime %between% c("2015-10-18 16:29", "2015-12-12 14:01") |# 2015 fall
semester
                             DateTime %between% c("2016-01-10 23:59", "2016-02-29 00:01") |
                             DateTime %between% c("2016-02-29 23:59", "2016-03-10 04:16") |
                             DateTime %between% c("2016-03-10 06:59", "2016-03-22 12:31") |
                             DateTime %between% c("2016-04-02 14:14", "2016-05-05 00:46"))) %>% # 2016
spring semester
  mutate(slice.label = ifelse(working_hours == "yes", "occ.sch", "unocc.sch"))
```

```
summer <- data %>%
  filter(weekday == "yes" & (DateTime %between% c("2015-07-09 00:29", "2015-07-31 23:46") |   #2015
summer
                            DateTime %between% c("2016-05-15 00:00", "2016-06-30 00:46"))) %>% # 2016
summer
  mutate(slice.label = ifelse(working_hours == "yes", "occ.sum", "unocc.sum"))


summerPeak <- data %>%
  filter(weekday == "yes" & (DateTime %between% c("2015-07-09 00:29","2015-09-14 14:01") |
                             DateTime %between% c("2015-09-23 03:29","2015-09-30 23:46") |#2015 summer peak
                           DateTime %between% c("2016-06-01 00:00", "2016-06-30 00:46"))) %>% #2016
summer peak
  mutate_at(vars(b002,b022), funs(mav(.,2))) %>%      # extra step to compute the 30-min rolling average
  filter(between(time, 840, 1140)) %>%
  mutate(slice.label = "peak.sum")



total <- data %>% filter(DateTime %between% c("2015-12-31 23:59", "2016-02-28 23:46") |
                         DateTime %between% c("2016-02-29 23:59", "2016-03-10 04:16")|
                         DateTime %between% c("2016-03-10 06:59", "2016-03-22 12:31")|
                         DateTime %between% c("2016-04-02 14:14", "2016-06-30 00:46")|
                         DateTime %between% c("2015-07-09 00:29", "2015-09-14 14:01")|
                         DateTime %between% c("2015-09-23 03:29", "2015-10-17 12:46")|
                         DateTime %between% c("2015-10-18 16:29", "2015-12-31 23:46")) %>%
  mutate(slice.label = "total")


#combine seperate energy.slice data frames
data.slices <- rbind(school,summer,summerPeak,total)


#create daily values
s.daily <- data.slices %>%
  filter(slice.label != "peak.sum") %>%    # need to do separately because finding the max instead of mean
for peak.sum
  group_by(slice.label, date) %>%
```

```
  summarise_at(vars(b002:b165), funs(mean(., na.rm = T)))

s.daily2 <- data.slices %>%
  filter(slice.label == "peak.sum") %>%
  group_by(date) %>%
  summarise_at(vars(b002:b165), funs(max(., na.rm = T))) %>%
  mutate(slice.label = "peak.sum") %>%
  select(slice.label, everything())

# combine two daily value data.frames
final.segemented.data <- rbind(data.frame(s.daily), data.frame(s.daily2))
#segmented.data.finer <- final.segemented.data

#buildings.remove <- c("b146")
#temp.data <- select(segmented.data.finer, -buildings.remove)
#buildings.remove <- c("b099")
#temp.data <- select(temp.data, -buildings.remove)

segment.final <- temp.data #removing b146 and b099 as they have -inf and na values

# write data --------------------------------------------------------------
write_rds(final.segemented.data, path="./test/segmented-final-data-2015.rds")
write_csv(final.segemented.data, path="./test/segmented-final-data-2015.csv")

write_rds(feat, path="./test/feat-clean-new.rds")  #this has features with correct names so don't have to keep doing when importing
```

Segment Data Function

```
# moving average used for summer peak
mav <- function(x,n){stats::filter(x,rep(1/n,n), sides=2)}
```

Regression Model 2016

```
# Import ------------------------------------------------------------------
# import libraries
library(data.table)
library(tidyverse)

# import data
data <- read_rds("./test/segmented-final-data-2015.rds")
feat <- read_rds("./test/feat-clean-new.rds")

feat$res <- NULL

# import functions
source("./R/03_regression-models_functions.R")

# prep features for regression ---------------------------------------------
# take log10 of area to improve normality
feat[,area := log10(area)]

# scale features
feat.scaled <- feat %>%
  select(area:floors) %>%
  scale() %>%
  as.data.frame()

# check scaling was done correctly; mean = 0, sd = 1
feat.scaled %>%
  summarise_all(funs(mean, sd))

# add in building number
feat.scaled <- cbind(building = feat$building, feat.scaled)
feat.scaled$building <- as.character(feat$building)
```

```
# regression --------------------------------------------------------------

# variable inputs to functions; define here

time.slices <- unique(data$slice.label)


# compute normalized energy use for each building for each day, regression model stats, and regression coef stats

for (i in time.slices) {


  name <- paste("output", i, sep = ".")

  assign(name, normalizedE(i, data))


}


norm <- rbind(output.occ.sch[[1]],output.unocc.sch[[1]],output.occ.sum[[1]],output.unocc.sum[[1]],output.peak.sum[[1]],output.total[[1]])

modelstats <- rbind(output.occ.sch[[2]],output.unocc.sch[[2]],output.occ.sum[[2]],output.unocc.sum[[2]],output.peak.sum[[2]],output.total[[2]])

coefstats <- rbind(output.occ.sch[[3]],output.unocc.sch[[3]],output.occ.sum[[3]],output.unocc.sum[[3]],output.peak.sum[[3]],output.total[[3]])


# compute benchmarks based on normalized energy use

for (i in time.slices) {


  name <- paste("bench", i, sep = ".")

  assign(name, benchmarkE(i, norm))


}


bench.data <- rbind(bench.occ.sch, bench.occ.sum, bench.total, bench.unocc.sch, bench.unocc.sum, bench.peak.sum)

rm(bench.occ.sch, bench.occ.sum, bench.total, bench.unocc.sch, bench.unocc.sum, bench.peak.sum)


# write data files ---------------------------------------------------------
```

```
write_rds(norm, "./test/normdata-2015.rds")
write_rds(bench.data, "./test/benchdata-2015.rds")
write_rds(modelstats, "./test/modelstatsdata-2015.rds")
write_rds(coefstats, "./test/coefstats-2015.rds")

#write csv files
write_csv(bench.data, "./test/benchdata-2015.csv")
write_csv(norm, "./test/normdata-2015.csv")

write_csv(modelstats, "./test/modelstatsdata-2015.csv")
write_csv(coefstats, "./test/coefstats-2015.csv")
```

Regression Model Functions

```
# reference: http://www.stat.columbia.edu/~martin/W2024/R10.pdf
normalizedE <- function(slice, energyData) {

  e <- energyData %>% filter(slice.label == slice)
  day <- e$date
  features <- feat.scaled %>% select(-building)
  allCoef <- c(names(features))

  # define empty dataframes
  norm <- data.table(actualE = 0, predictE = 0, intercept = 0, eNorm = 0, names = 0, day = as.Date(0, origin = "1970-01-01"), slice.label = "")[-1,]
  model.d <- data.table(r = 0, rse = 0, minEUI = 0, meanEUI = 0, maxEUI = 0, day = as.Date(0, origin = "1970-01-01"), slice.label = "")[-1,]
  coefstats <- data.table(coefName = 0, coef = 0, pvalue = 0, day = as.Date(0, origin = "1970-01-01"), slice.label = "")[-1,]
  x <- 1

  for (i in day) { # for each day, benchmark energy consumption
    d <- as.Date(i, "1970-01-01")
    y <- e %>%
      filter(date == d) %>%
```

```
    select(-slice.label:-date)
  z <- t(log10(y*10^5))    #transform EUI to make more normal
  model.data <- data.table(cbind(EUI = z, features))


  mdl <- step(lm(EUI ~ 1, data = model.data), direction = "forward", scope = ~ area + service + circ + class + general + lab_dry + lab_wet + mech + office +
             special + study+ supply + age + reno + pctReno + floors)


  # get regression model overall stats
  r2 <- summary(mdl)$adj.r.squared
  rse <- summary(mdl)$sigma
  minEUI <- min(z)
  meanEUI <- mean(z)
  maxEUI <- max(z)
  day_df <- day[x]
  f <- data.table(r = r2, rse = rse, minEUI = minEUI, meanEUI = meanEUI, maxEUI = maxEUI, day = day_df)
  f[,slice.label := slice]
  model.d <- rbind(model.d,f)


  h <- matrix(nrow = length(feat.scaled$building), ncol=4)


  for (j in 1:length(feat.scaled$building)) {
    predictE <- predict(mdl,features[j,]) # given each building's features, predict energy consumption
    actualE <- model.data[j,EUI]
    intercept <- mdl$coefficients[1] # intercept is the average building performance (b/c features standardized)
    eNorm <- actualE - predictE + intercept #this subtracts the level of performance from the mean, so everthing is in comparison to the mean
    h[j,1] <- actualE
    h[j,2] <- predictE
    h[j,3] <- intercept
    h[j,4] <- eNorm
    h <- data.table(h)
  }
```

```
    setnames(h, old=c("V1","V2","V3","V4"), new=c("actualE", "predictE","intercept","eNorm"))
    h[,"names" := feat.scaled$building]
    h[,day := rep(day[x],length(y))]
    h[,slice.label := rep(slice, length(y))]
    norm <- rbind(norm,h)

    for (j in allCoef) {  #loop through all coefficients
      if (j %in%  names(mdl$coefficients)[-1] == "TRUE") {     #see which coefficiencts for this model are
present
        coefName <- j
        c <- mdl$coefficients[match(coefName,names(mdl$coefficients))]
        p <- summary(mdl)$coefficients[,4][match(coefName,names(mdl$coefficients))]
        #v <- vif(mdl)[match(coefName,names(mdl$coefficients))]
        f <- data.table(coefName = coefName, coef = c, pvalue = p, day = day[x])
        f[,slice.label := slice]
        coefstats <- rbind(coefstats,f)
      }
    }

    x <- x + 1
  }
  dfList <- list(norm, model.d, coefstats)
  return(dfList)

}

benchmarkE <- function(label, normData) {  # eNorm is 3 col-- eNorm, day, names
  x <- normData %>% filter(slice.label == label) %>% select(eNorm:day) %>% data.table()
  #x <- data[,c("eNorm","names","day")]
  bench <- data.table(eNorm = 0, names = 0, day = as.Date(0, "1970-01-01"), efficiency = 0, normalized =
0)[-1,]
  for (i in unique(x$day)) {
    d <- as.Date(i, "1970-01-01")
```

```
    y <- x[ day == d]
    y <- y[order(rank(eNorm))]
    fun.ecdf <- ecdf(y[,eNorm])
    benchmark <- environment(fun.ecdf)$y
    y[,"efficiency" := benchmark]
    y[,"normalized" := (1-(eNorm-min(eNorm))/(max(eNorm)-min(eNorm)))]
    bench <- rbind(bench,y)
  }
  output <- bench[,"t" := rep(label,length(x$day))]
  return(output)
}
```

Clean-Data 2018

```
# import libraries
library(tidyverse)
library(lubridate)
library(data.table)
library(padr)
library(plotly)
library(zoo)

# import functions
source("./R/01_clean-data_functions.R")

# define daylight savings times
EST <- c("2018-11-04 05:00:00","2018-11-04 05:15:00","2018-11-04 05:30:00","2018-11-04 05:45:00")

# create cleanData column
start <- as.POSIXct("2018-01-01 01:00", tz= "UTC")
end <- start + as.difftime((365*1)+1, units = "days")
timestamp <- seq(from=start, by = 15*60, to = end)
timestamp <- timestamp[1:length(timestamp)-1]
```

```
clean <- data.table(DateTime = as.POSIXct(format(timestamp,tz =
"America/New_York",usetz=TRUE),tz="America/New_York"))


# clean electricity data

# define folders based on where utility data is stored

folders <- list.dirs(path = "./data", full.names = TRUE)

files <- list.files(path=folders, pattern = "*.csv", full.names=T, recursive=FALSE)


# set correct column names

colNames <- c("002a", "002b", "022", "025", "029", "030", "031", "036", "038", "039", "040", "045",
"050a", "050b", "051", "055", "058", "059", "061", "061a","066", "075", "076", "081", "084", "099", "101",
"103a", "103b", "103c", "103d", "104a", "104b", "111", "114", "123", "124","135", "144a", "144b", "146",
"147a", "147b", "147c", "153a", "153b", "165a", "165b")


# clean electricity data

g <- 1

for (i in files) {

  clean[, paste0("b",colNames[g]) := cleanElec(i,"20171231 2000","20190101 1945")]

  g <- g + 1

}


# combine double meters

clean[, b002 := b002a + b002b][,c("b002a","b002b") := NULL]

clean[, b050 := b050a + b050b][,c("b050a","b050b") := NULL]


clean[, b103 := b103a + b103b + b103c + b103d][,c("b103a","b103b","b103c","b103d") := NULL]

clean[, b104 := b104a + b104b][,c("b104a","b104b") := NULL]

clean[, b144 := b144a + b144b][,c("b144a","b144b") := NULL]

clean[, b147 := b147a + b147b + b147c][,c("b147a","b147b","b147c") := NULL]

clean[, b153 := b153a + b153b][,c("b153a","b153b") := NULL]

clean[, b165 := b165a + b165b][,c("b165a","b165b") := NULL]


# find NA, zeros, negatives, and outliers

m <- data.table(t(clean[, lapply(.SD, function(x) sum(x < 0, na.rm = T)), .SDcols = 2:38]))
```

```
clean[clean < 0] <- NA

m[, V2 := t(clean[, lapply(.SD, function(x) sum(x == 0, na.rm = T)), .SDcols = 2:38])]

m[, v3 := t(clean[, lapply(.SD, function(x) sum(x > (mean(x,na.rm = T)+(IQR(x,na.rm = T)*20)), na.rm =
T)), .SDcols = 2:38])]

attach(clean)

for (i in names(clean)[2:length(clean)]) {

  y <- get(i)

  thresh <- (mean(y,na.rm = T)+(IQR(y,na.rm = T)*20))

  y[y > thresh] <- NA

  clean[,c(i) := y]

}

detach(clean)

m[, V4 := t(clean[, lapply(.SD, function(x) sum(is.na(x))), .SDcols = 2:38])]

setnames(m, c("negative", "zero", "outlier","na"))

n <- names(clean)[2:length(clean)]

m[, name := n]


#outlierKD(clean,b055)

#plot(clean$DateTime,clean$b153)


# subtract/add submeters

clean[, b059 := NULL]


clean[, b099a := NULL]#cleanData this is repeating with b099


# interpolation, max gap interperlated is 6 hours (linear)

clean2 <- clean[, lapply(.SD, function(x) na.approx(x, maxgap = 24)), .SDcols = 2:37]

m <- data.table(t(clean2[, lapply(.SD, function(x) sum(is.na(x)))]))

m[, V2 := t(clean2[, lapply(.SD, function(x) sum(x == 0, na.rm = T))])]

n <- names(clean2)

m[, name := n]

clean2[,DateTime := clean$DateTime]


write.csv(clean2, "./test/cleanElec.csv",row.names = F)
```

write.csv(m, "./test/missingDataDetail.csv",row.names = F)

Segment Data 2018

```
# Import -------------------------------------------------------------------

# import libraries

library(data.table)

library(tidyverse)

library(lubridate)

library(dygraphs)

library(xts)  # for plotting with dygraphs



# import data

data <- fread("./test/cleanElec.csv")[, DateTime := as.POSIXct(DateTime)]



feat <- fread("./output/featPct2.csv")



# import functions

source("./R/02_segment-data_functions.R")



# data prep for temporal segmentation -------------------------------------

# clean column names and data types

setcolorder(data, c(data[, sort(names(data))])) #order matters for EUI calcs

colNames <-
c("building","area","service","circ","class","general","lab_dry","lab_wet","mech","office","res","special","
study","supply",

            "age","reno","pctReno","floors","use")

setnames(feat,colNames)

feat[, area := as.numeric(area)][,res := NULL]  # remove residential buildings because they are removed
froms sample
```

```
# divide energy by building area

area.vector <- feat[,area]

eui <- sweep(data[,1:(length(data)-1)],MARGIN=2,FUN="/", STATS=area.vector)

data <- data.table(cbind(DateTime = data[,DateTime], eui))


# OPTIONAL: graph energy use

x <- as.xts(data[,c("DateTime","b153")])

dygraph(x) %>% dyRangeSelector()


# temporal segmentation ---------------------------------------------------

#add in Date and Time column

data <- data %>%

  mutate(day_of_week = wday(DateTime, week_start = getOption("lubridate.week.start", 1)),

         weekday = ifelse(day_of_week < 6, "yes", "no"),

         hour = hour(DateTime),

         date = date(DateTime),

         time = as.numeric(hm(strftime(DateTime, format = "%H:%M")))/60,

         working_hours = ifelse(weekday == "yes" & between(time, 480, 1200), "yes", "no"))


# now create energy slice labels for -- occ.sch, unocc.sch, occ.sum, unocc.sum, peak.sum

school <- data %>%

  filter(weekday == "yes" & (DateTime %between% c("2018-01-08 01:00", "2018-03-10 04:16") |

                           DateTime %between% c("2018-03-10 06:59", "2018-03-22 12:31") |

                           DateTime %between% c("2018-04-02 14:14", "2018-04-24 23:46") |# 2018 spring
semester

                           DateTime %between% c("2018-08-20 00:00", "2018-09-14 14:01")|
```

```
                    DateTime %between% c("2018-09-23 03:29", "2018-10-17 12:46")|
                    DateTime %between% c("2018-10-18 16:29", "2018-12-04 00:46"))) %>% # 2016 fall
semester
  mutate(slice.label = ifelse(working_hours == "yes", "occ.sch", "unocc.sch"))

summer <- data %>%
  filter(weekday == "yes" & (DateTime %between% c("2018-05-14 00:00", "2018-06-30 20:46")|
                    DateTime %between% c("2018-07-09 00:29", "2018-07-25 23:46"))) %>%
  mutate(slice.label = ifelse(working_hours == "yes", "occ.sum", "unocc.sum"))

summerPeak <- data %>%
  filter(weekday == "yes" & (DateTime %between% c("2018-06-01 00:00","2018-06-30 20:46")|
                    DateTime %between% c("2018-07-09 00:29","2018-09-14 14:01")|
                    DateTime %between% c("2018-09-23 03:29","2018-09-30 23:46"))) %>%
  mutate_at(vars(b002,b022), funs(mav(.,2))) %>%      # extra step to compute the 30-min rolling average
  filter(between(time, 840, 1140)) %>%
  mutate(slice.label = "peak.sum")


total <- data %>%  filter(DateTime %between% c("2017-12-31 23:59", "2018-03-10 04:16")|
                  DateTime %between% c("2018-03-10 06:59", "2018-03-22 12:31")|
                  DateTime %between% c("2018-04-02 14:14", "2018-06-30 20:46")|
                  DateTime %between% c("2018-07-09 00:29", "2018-09-14 13:59")|
                  DateTime %between% c("2018-09-23 03:29", "2018-10-17 12:46")|
                  DateTime %between% c("2018-10-18 16:29", "2018-12-31 23:46")) %>%
  mutate(slice.label = "total")

#combine seperate energy.slice data frames
```

```
data.slices <- rbind(school,summer,summerPeak,total)


#create daily values

s.daily <- data.slices %>%

  filter(slice.label != "peak.sum") %>%    # need to do separately because finding the max instead of mean for peak.sum

  group_by(slice.label, date) %>%

  summarise_at(vars(b002:b165), funs(mean(., na.rm = T)))


s.daily2 <- data.slices %>%

  filter(slice.label == "peak.sum") %>%

  group_by(date) %>%

  summarise_at(vars(b002:b165), funs(max(., na.rm = T))) %>%

  mutate(slice.label = "peak.sum") %>%

  select(slice.label, everything())


# combine two daily value data.frames

final.segemented.data <- rbind(data.frame(s.daily), data.frame(s.daily2))




# write data ---------------------------------------------------------------

write_rds(final.segemented.data, path="./test/segmented-final-data.rds")


write_csv(final.segemented.data, path="./test/segmented-final-data.csv")


write_rds(feat, path="./test/feat-clean.rds")  #this has features with correct names so don't have to keep doing when importing
```

Regression Model 2018

```
# Import ----------------------------------------------------------------

# import libraries
library(data.table)
library(tidyverse)

# import data
data <- read_rds("./test/segmented-final-data.rds")
feat <- read_rds("./test/feat-clean.rds")



# import functions
source("./R/03_regression-models_functions.R")

# prep features for regression ------------------------------------------

# take log10 of area to improve normality
feat[,area := log10(area)]

# scale features
feat.scaled <- feat %>%
  select(area:floors) %>%
  scale() %>%
  as.data.frame()

# check scaling was done correctly; mean = 0, sd = 1
feat.scaled %>%
  summarise_all(funs(mean, sd))

# add in building number
feat.scaled <- cbind(building = feat$building, feat.scaled)
feat.scaled$building <- as.character(feat$building)

# regression ------------------------------------------------------------
```

```
# variable inputs to functions; define here

time.slices <- unique(data$slice.label)


# compute normalized energy use for each building for each day, regression model stats, and regression coef stats

for (i in time.slices) {


  name <- paste("output", i, sep = ".")

  assign(name, normalizedE(i, data))


}


norm <- rbind(output.occ.sch[[1]],output.unocc.sch[[1]],output.occ.sum[[1]],output.unocc.sum[[1]],output.peak.sum[[1]],output.total[[1]])

modelstats <- rbind(output.occ.sch[[2]],output.unocc.sch[[2]],output.occ.sum[[2]],output.unocc.sum[[2]],output.peak.sum[[2]],output.total[[2]])

coefstats <- rbind(output.occ.sch[[3]],output.unocc.sch[[3]],output.occ.sum[[3]],output.unocc.sum[[3]],output.peak.sum[[3]],output.total[[3]])


# compute benchmarks based on normalized energy use

for (i in time.slices) {


  name <- paste("bench", i, sep = ".")

  assign(name, benchmarkE(i, norm))


}


bench.data <- rbind(bench.occ.sch, bench.occ.sum, bench.total, bench.unocc.sch, bench.unocc.sum, bench.peak.sum)

rm(bench.occ.sch, bench.occ.sum, bench.total, bench.unocc.sch, bench.unocc.sum, bench.peak.sum)


#coeff stats analyze

coeff <- read_rds("./output/coefstats.rds")
```

```
total <- subset(coeff, coeff$slice.label == "total")
occsum <- subset(coeff, coeff$slice.label == "occ.sum")
unoccsum<- subset(coeff, coeff$slice.label == "unocc.sum")
peaksum<- subset(coeff, coeff$slice.label == "peak.sum")
occsch<- subset(coeff, coeff$slice.label == "occ.sch")
unoccsch<- subset(coeff, coeff$slice.label == "unocc.sch")
library(plyr)
T <- count(total, "coefName")
T$freq <- T$freq/3.67

OS<- count(occsum, "coefName")
OS$freq <- OS$freq/.53
US<- count(unoccsum, "coefName")
US$freq <- US$freq/.53

PS<- count(peaksum, "coefName")
PS$freq <- PS$freq/.86
OSc<- count(occsch, "coefName")
OSc$freq <- OSc$freq/1.53
USc<- count(unoccsch, "coefName")
USc$freq <- USc$freq/1.54


frequency <- rbind(T,OS,US,PS,OSc,USc)

write_csv(frequency, "./output/frequency.csv")

# write data files -------------------------------------------------------
write_rds(norm, "./test/normdata.rds")
write_rds(bench.data, "./test/benchdata.rds")
write_rds(modelstats, "./test/modelstatsdata.rds")
write_rds(coefstats, "./test/coefstats.rds")
```

```
#write csv files
write_csv(bench.data, "./test/benchdata.csv")
write_csv(norm, "./test/normdata.csv")
write_csv(modelstats, "./test/modelstatsdata.csv")
write_csv(coefstats, "./test/coefstats.csv")
```

B.2 Statistical Tests, Plots, and Validation

Statistical Tests

```
# Import ----------------------------------------------------------------
library(tidyverse)
library(data.table)
source("./R/04_statistical-tests_functions.R")



Difference <- A$b153 - B$b153
hist(Difference,
     col="gray",
     main="Histogram of differences",
     xlab="Difference")



#t-test for all the 36 buildings

A <- read_rds("./test/benchdata-m-2015.rds")
B <- read_rds("./test/benchdata-m.rds")

E <- unique(A$names)
#total is the period
List = list()
for(i in 1:36)
{
```

```
 C <- subset(A, A$names == E[i] & A$t == "total")
  D <- subset(B, B$names == E[i] & B$t == "total")
  List[[length(List)+1]] <- t.test(C$normalized,D$normalized,paired=TRUE,alternative = "two.sided")
  boxplot(C$normalized,D$normalized)
}

hyp <- data.frame(matrix(unlist(List), nrow=length(List), byrow=T))
results <- cbind(E,hyp)

# for occ.sch
ListA =list()
for(i in 1:36)
{
  C <- sample_n(subset(A, A$names == E[1] & A$t == "occ.sch") , 142)
  D <- subset(B, B$names == E[1] & B$t == "occ.sch")
  ListA[[length(ListA)+1]] <- t.test(C$normalized,D$normalized,paired = TRUE, alternative = "two.sided")

}

hypA <- data.frame(matrix(unlist(ListA), nrow=length(ListA), byrow=T))
resultsA <- cbind(E,hypA)

# for unocc.sch
ListB =list()
for(i in 1:36)
{
  C <- sample_n(subset(A, A$names == E[i] & A$t == "unocc.sch") , 143)
  D <- subset(B, B$names == E[i] & B$t == "unocc.sch")
  ListB[[length(ListB)+1]] <- t.test(C$normalized,D$normalized,paired=TRUE,alternative = "two.sided")

}

hypB <- data.frame(matrix(unlist(ListB), nrow=length(ListB), byrow=T))
```

```
resultsB <- cbind(E,hypB)

# for peak.sum
ListC =list()
for(i in 1:36)
{
  C <-subset(A, A$names == E[i] & A$t == "peak.sum")
  D <- sample_n(subset(B, B$names == E[i] & B$t == "peak.sum"), 75)
  ListC[[length(ListC)+1]] <- t.test(C$normalized,D$normalized,paired=TRUE,alternative = "two.sided")

}

hypC <- data.frame(matrix(unlist(ListC), nrow=length(ListC), byrow=T))
resultsC <- cbind(E,hypC)
# for occ.sum
ListD =list()
for(i in 1:36)
{
  C <- sample_n(subset(A, A$names == E[i] & A$t == "occ.sum"), 48)
  D <- subset(B, B$names == E[i] & B$t == "occ.sum")
  ListD[[length(ListD)+1]] <- t.test(C$normalized,D$normalized,paired=TRUE,alternative = "two.sided")

}

hypD <- data.frame(matrix(unlist(ListD), nrow=length(ListD), byrow=T))
resultsD <- cbind(E,hypD)
# for unocc.sum
ListE =list()
for(i in 1:36)
{
  C <- sample_n(subset(A, A$names == E[i] & A$t == "unocc.sum") , 48)
  D <- subset(B, B$names == E[i] & B$t == "unocc.sum")
  ListE[[length(ListE)+1]] <- t.test(C$normalized,D$normalized,paired=TRUE,alternative = "two.sided")
```

```
}

hypE <- data.frame(matrix(unlist(ListE), nrow=length(ListE), byrow=T))
resultsE <- cbind(E,hypE)

write.csv(results,"./test/paired-t-test-m1.csv")
write.csv(resultsA,"./test/paired-t-test-occsch.csv")
write.csv(resultsB,"./test/paired-t-test-unoccsch.csv")
write.csv(resultsC,"./test/paired-t-test-peaksum.csv")
write.csv(resultsD,"./test/paired-t-test-occsum.csv")
write.csv(resultsE,"./test/paired-t-test-unoccsum.csv")

#above code ends here, making box plots now
L = list()

pdf("box-plot-efficiency.pdf")



for(i in 1:36)
{
  C <- subset(A, A$names == E[i] & A$t == "total")
  D <- subset(B, B$names == E[i] & B$t == "total")

    boxplot(C$efficiency,D$efficiency, ylab= 'Normalized efficiency', xlab= E[i], main=' Efficiency in 2015 vs 2018')
}
```

Plots B

```
# Import -----------------------------------------------------------------
library(tidyverse)
library(data.table)
library(stats)
```

```
library(forcats) # for reordering boxplots by median
library(surveillance) # for week # function
library(RColorBrewer)  # for heat map colors
library(plotly)
library(zoo) # for rolling average
library(sp)
library(surveillance)
library(gridExtra)
library(ggplot2)

#benchmarks plotting praga

benchmarksA <- read_rds("./test/benchdata-m-2015.rds")
benchmarksB <- read_rds("./test/benchdata-m.rds")

E <- unique(benchmarksA$names)

plot_A = list()
plot_B = list()

for(i in 1:36)
{

  c <- benchmarksA[benchmarksA$names == E[i],]
  d <- benchmarksB[benchmarksB$names == E[i],]

  p = ggplot(c, aes(normalized, fill = t)) + geom_density(alpha=.3) + ggtitle(E[i], subtitle = "Period A")
  q = ggplot(d, aes(normalized, fill = t)) + geom_density(alpha=.3) + ggtitle(E[i], subtitle = "Period B")
  r = grid.arrange(p, q, ncol = 2)
  file_name = paste("normalized_A_", E[i], ".tiff", sep="")

  ggsave(file_name,plot = r, width = 8, height = 3, dpi = 600)
```

```
}

for (i in 1:36)
  {



  file_name = paste("normalized_A_", E[i], ".tiff", sep="")
  tiff(file_name)
  print(plot_A[[i]])
  dev.off()
}
for (i in 1:36)
{
  file_name = paste("normalized_B_", E[i], ".tiff", sep="")
  tiff(file_name)
  print(plot_B[[i]])
  dev.off()
}

ggplot(d, aes(normalized, fill = t)) +
  geom_histogram(bins = 40, position = "dodge", binwidth = .03)

#benchmarks plotting praga

types <- fread("./output/buildingTypes.csv")

coef.stats <- fread("./output/coefficient_stats-no166.csv")

reg.stats <- fread("./output/modelstatsdata.csv")
```

```
# Edit Data labels -------------------------------------------------------

reg.stats[reg.stats == "OccupiedSchool"] <- "A"

reg.stats[reg.stats == "UnoccupiedSchool"] <- "B"

reg.stats[reg.stats == "OccupiedSummer"] <- "C"

reg.stats[reg.stats == "UnoccupiedSummer"] <- "D"

reg.stats[reg.stats == "Peaks"] <- "E"


# Appendix: distribution of regression model R values ---------------------

reg.stats2 <- reg.stats # cause going to turn one column into a factor

#reg.stats2[, t := factor(t, c("A","B","C","D","E","Total"))]


reg.stats2[reg.stats2 == "A"] <- "Occupied: School (A)"

reg.stats2[reg.stats2 == "B"] <- "Unoccupied: School (B)"

reg.stats2[reg.stats2 == "C"] <- "Occupied: Summer (C)"

reg.stats2[reg.stats2 == "D"] <- "Unoccupied: Summer (D)"

reg.stats2[reg.stats2 == "E"] <- "Summer Peak (E)"

reg.stats2[reg.stats2 == "total"] <- "Total"


reg.stats2[, t := factor(t, c("Occupied: School (A)","Unoccupied: School (B)","Occupied: Summer
(C)","Unoccupied: Summer (D)","Summer Peak (E)","Total"))]


tiff(file="./figs/Figure_rvalue_praga.tiff", units="in", width=5,height=3, res = 300)

ggplot(reg.stats2, aes(r, color = t)) +

  geom_line(stat = "density", size = 1) +

  xlim(0.1,0.7) +

  labs(x = "Adj. R-squared", y = "Density", color = "Time Period") +

  scale_color_manual(values=c(`Occupied: School (A)` = "#F8766D", `Unoccupied: School (B)`=
"#F564E3", `Occupied: Summer (C)` = "#00BA38", `Unoccupied: Summer (D)` = "#00BFC4", `Summer
Peak (E)` = "#619CFF", `Total` = "#B79F00")) +

  theme(text = element_text(size=10))
```

```
dev.off()
```

Validation: EUI

```
# Import ----------------------------------------------------------------
library(tidyverse)
library(data.table)
source("./R/04_statistical-tests_functions.R")
#test
A <- read_csv("./test/eui2015.csv")
B <- read_csv("./test/eui2018.csv")

#difference should follow a normal distribution

Difference <- A$b153 - B$b153
hist(Difference,
     col="gray",
     main="Histogram of differences",
     xlab="Difference")


#t-test for all the 36 buildings

A <- read_rds("./test/benchdata-2015.rds")
B <- read_rds("./test/benchdata.rds")
E <- unique(A$names)
#total is the period
List = list()
for(i in 1:36)
{
 C <- subset(A, A$names == E[i] & A$t == "total")
  D <- subset(B, B$names == E[i] & B$t == "total")
  List[[length(List)+1]] <- t.test(C$normalized,D$normalized,paired=TRUE,alternative = "two.sided")
```

```
  boxplot(C$normalized,D$normalized)
}

hyp <- data.frame(matrix(unlist(List), nrow=length(List), byrow=T))
results <- cbind(E,hyp)

# for occ.sch
ListA =list()
for(i in 1:36)
{
  C <- sample_n( subset(A, A$names == E[i] & A$t == "occ.sch"), 142)
  D <- subset(B, B$names == E[i] & B$t == "occ.sch")
  ListA[[length(ListA)+1]] <- t.test(C$normalized,D$normalized,paired=TRUE,alternative = "two.sided")

}

hypA <- data.frame(matrix(unlist(ListA), nrow=length(ListA), byrow=T))
resultsA <- cbind(E,hypA)

# for unocc.sch
ListB =list()
for(i in 1:36)
{
  C <- sample_n(subset(A, A$names == E[i] & A$t == "unocc.sch") , 143)
  D <- subset(B, B$names == E[i] & B$t == "unocc.sch")
  ListB[[length(ListB)+1]] <- t.test(C$normalized,D$normalized,paired=TRUE,alternative = "two.sided")

}

hypB <- data.frame(matrix(unlist(ListB), nrow=length(ListB), byrow=T))
resultsB <- cbind(E,hypB)

# for peak.sum
```

```
ListC =list()
for(i in 1:36)
{
  C <-subset(A, A$names == E[i] & A$t == "peak.sum")
  D <- sample_n(subset(B, B$names == E[i] & B$t == "peak.sum"), 75)
  ListC[[length(ListC)+1]] <- t.test(C$normalized,D$normalized,paired=TRUE,alternative = "two.sided")

}

hypC <- data.frame(matrix(unlist(ListC), nrow=length(ListC), byrow=T))
resultsC <- cbind(E,hypC)
# for occ.sum
ListD =list()
for(i in 1:36)
{
  C <- sample_n(subset(A, A$names == E[i] & A$t == "occ.sum"), 48)
  D <- subset(B, B$names == E[i] & B$t == "occ.sum")
  ListD[[length(ListD)+1]] <- t.test(C$normalized,D$normalized,paired=TRUE,alternative = "two.sided")

}

hypD <- data.frame(matrix(unlist(ListD), nrow=length(ListD), byrow=T))
resultsD <- cbind(E,hypD)
# for unocc.sum
ListE =list()
for(i in 1:36)
{
  C <- sample_n(subset(A, A$names == E[i] & A$t == "unocc.sum") , 48)
  D <- subset(B, B$names == E[i] & B$t == "unocc.sum")
  ListE[[length(ListE)+1]] <- t.test(C$normalized,D$normalized,paired=TRUE,alternative = "two.sided")

}
```

```
hypE <- data.frame(matrix(unlist(ListE), nrow=length(ListE), byrow=T))
resultsE <- cbind(E,hypE)

write.csv(results,"./test/paired-t-test.csv")
write.csv(resultsA,"./test/paired-t-test-occsch.csv")
write.csv(resultsB,"./test/paired-t-test-unoccsch.csv")
write.csv(resultsC,"./test/paired-t-test-peaksum.csv")
write.csv(resultsD,"./test/paired-t-test-occsum.csv")
write.csv(resultsE,"./test/paired-t-test-unoccsum.csv")

#abobve code ends here, making box plots now
L = list()

pdf("box-plot.pdf")



for(i in 1:36)
{
  C <- subset(A, A$names == E[i] & A$t == "total")
  D <- subset(B, B$names == E[i] & B$t == "total")

    boxplot(C$normalized,D$normalized, ylab= 'EUI', xlab= E[i], main='EUI in 2015 vs 2018')
}
```

Validation: EUI vs Efficiency

```
library(tidyverse)
library(data.table)
library(lubridate)
library(ggpubr)



#filtering only required normalized values
```

```
A <- read_rds("./test/benchdata-2015.rds")
B <- read_rds("./test/benchdata.rds")

C <- cbind(A[,2:3],A[,5:6])
C <- C %>% filter(t == "total")
C <- C[,1:3]

D <- cbind(B[,2:3],B[,5:6])
D <- D %>% filter(t == "total")
D <- D[,1:3]

# converting to wide format
norm2015 <- reshape(C, idvar="day", timevar="names",direction="wide")
norm2018 <- reshape(D, idvar="day", timevar="names",direction="wide")

norm2015 <- norm2015[,order(names(norm2015))]
norm2018 <- norm2018[,order(names(norm2018))]

write.csv(norm2015,"./test/latnorm2015.csv")
write.csv(norm2018,"./test/latnorm2018.csv")

#after removing the row numbers

norm2015 <- read.csv("./test/latnorm2015.csv")
norm2018 <- read.csv("./test/latnorm2018.csv")

norm2015 <- rbind(norm2015[169:339,],norm2015[1:168,])

compare <- norm2015[,2:37]-norm2018[,2:37]
compare <- cbind(norm2018[,1],compare)

eui2015 <- read.csv("./test/eui2015-new.csv")
```

```
eui2018 <- read.csv("./test/eui2018-new.csv")

compareeui <- eui2015[,2:37]-eui2018[,2:37]

compareeui <- compareeui[,order(names(compareeui))]
compareeui <- cbind(eui2018[,1],compareeui)

names(compareeui)[names(compareeui) == 'eui2018[,1]'] <- 'building'

columnname <- colnames(compareeui)
columnname[1] <- "building"    # renaming column
colnames(compareeui) <- columnname
colnames(compare)<- columnname

# pearson test
result <- list()

for(i in 2:37)
{
  X <- compare[,i]
  Y <- compareeui[,i]
  result[[length(result)+1]] <- cor.test(compare[,i],compareeui[,i], method = "pearson")
}

pearson_result <- data.frame(matrix(unlist(result), nrow=length(result), byrow=T))


cn <- data.frame(columnname[2:37])

temp <- cbind(cn,pearson_result)
```

```
# t-tst with updates formattting, true paired test
ListA =list()
for(i in 2:37)
{
  CA <- norm2015[,i]
  DA <- norm2018[,i]
  ListA[[length(ListA)+1]] <- t.test(CA,DA,paired = TRUE, alternative = "two.sided")

}

hypA <- data.frame(matrix(unlist(ListA), nrow=length(ListA), byrow=T))
resultsA <- cbind(cn,hypA)



write.csv(temp,"./test/euivsefficiency.csv")
write.csv(resultsA,"./test/ordered_paired_t-test.csv")
```

APPENDIX C: STATISTICAL TESTS RESULTS

C.1 Paired T-Test For Efficiency Change For All Temporal Segments

Paired T-Test Normalized Efficiency 2016 Vs 2018 – Total Period

Building Name	T-score	DF	P-Value	Efficiency change
b144	-20.8219	330	1.53E-62	-0.09192
b103	-19.8061	330	1.69E-58	-0.10757
b038	-15.9889	330	3.01E-43	-0.12714
b040	-9.2187	330	3.30E-18	-0.08952
b050	-6.48989	330	3.06E-10	-0.04816
b051	-5.91291	330	8.22E-09	-0.02503
b101	-5.40146	330	1.25E-07	-0.04454
b147	-4.96019	330	1.12E-06	-0.02937
b039	-4.88013	330	1.64E-06	-0.04328
b123	-4.44057	330	1.22E-05	-0.04178
b025	-4.08401	330	5.53E-05	-0.03576
b030	-3.7073	330	0.000245	-0.03633
b104	-2.57279	330	0.010514	-0.03468
b002	-2.35857	330	0.018914	-0.02435
b084	-2.28098	330	0.023172	-0.00666
b099	-1.89104	330	0.059475	-0.02717
b055	-0.47902	330	0.632231	-0.01879
b153	-0.40413	330	0.686374	-0.01876
b124	0.631635	330	0.528052	-0.0102
b111	1.944222	330	0.052698	-0.00016
b066	1.944223	330	0.052698	-0.00021
b045	2.020161	330	0.044155	0.000408
b031	3.039209	330	0.002557	0.004637
b036	3.551422	330	0.000438	0.012268
b058	4.867487	330	1.74E-06	0.013535
b135	4.967464	330	1.08E-06	0.012524
b075	5.16713	330	4.07E-07	0.016837
b061a	5.827163	330	1.32E-08	0.016343
b029	6.405894	330	5.01E-10	0.040925

b165	6.987121	330	1.50E-11	0.021962
b114	7.476982	330	6.60E-13	0.028848
b081	7.726059	330	1.28E-13	0.017085
b076	7.803699	330	7.58E-14	0.047474
b146	9.796013	330	4.15E-20	0.053095
b061	10.1286	330	3.12E-21	0.040434
b022	13.9529	330	2.93E-35	0.04744

Significant changes in EUI

Buildings with an energy retrofit

Paired T-Test Normalized Efficiency 2016 Vs 2018 – Unoccupied School

Building Name	T-score	DF	P-Value	Efficiency change
b144	-16.6615	142	3.19E-35	-0.1184
b038	-14.7171	142	2.33E-30	-0.16948
b103	-13.9772	142	1.82E-28	-0.11445
b002	-10.0616	142	2.60E-18	-0.08298
b030	-7.77951	142	1.36E-12	-0.0962
b153	-7.02324	142	8.27E-11	-0.09637
b101	-5.7531	142	5.18E-08	-0.07701
b051	-5.48909	142	1.81E-07	-0.03074
b104	-5.19004	142	7.13E-07	-0.08289
b084	-2.98906	142	0.003299	-0.01439
b147	-2.80367	142	0.00576	-0.03502
b025	-2.47223	142	0.014609	-0.04424
b066	-2.28328	142	0.023898	-0.06197
b045	-2.25891	142	0.025412	-0.05053
b099	-1.99011	142	0.048499	-0.04182
b039	-1.5775	142	0.116905	-0.03404
b061a	-1.4238	142	0.156698	-0.02553
b124	-0.72661	142	0.468662	-0.02935
b123	-0.60621	142	0.545345	-0.02414
b076	-0.06333	142	0.949595	-0.02759
b040	0.194708	142	0.8459	-0.02382
b075	0.333617	142	0.73916	-0.01422
b050	0.711277	142	0.47808	-0.01068
b031	1.095013	142	0.275365	-0.00659

b055	1.573741	142	0.117773	-0.00484
b058	1.768782	142	0.079077	-0.00116
b036	2.027809	142	0.04445	0.000508
b111	2.071762	142	0.040097	0.000905
b135	2.458227	142	0.015166	0.003394
b146	2.50528	142	0.013365	0.004756
b081	2.524043	142	0.012702	0.002244
b029	3.036613	142	0.002847	0.016001
b114	3.632723	142	0.000391	0.012043
b061	5.871608	142	2.92E-08	0.025274
b165	8.64429	142	1.03E-14	0.042768
b022	9.56911	142	4.77E-17	0.038357

Significant changes in EUI

Buildings with an energy retrofit

Paired T-Test Normalized Efficiency 2016 Vs 2018 – Occupied School

Building Name	T-score	DF	P-Value	Efficiency change
b144	-25.5285	141	1.00E-54	-0.12204
b103	-12.3312	141	3.68E-24	-0.08969
b039	-8.29997	141	7.62E-14	-0.07412
b038	-7.71992	141	1.95E-12	-0.09238
b147	-7.34096	141	1.54E-11	-0.04754
b002	-6.82895	141	2.35E-10	-0.05563
b101	-6.54488	141	1.03E-09	-0.06475
b153	-5.87746	141	2.87E-08	-0.05513
b025	-5.56788	141	1.26E-07	-0.0582
b030	-4.40543	141	2.07E-05	-0.05004
b031	-3.20622	141	0.001664	-0.02878
b104	-3.14811	141	0.002006	-0.0358
b123	-2.89822	141	0.004353	-0.04077
b061a	-2.45489	141	0.01531	-0.01933
b099	-2.16966	141	0.031709	-0.03667
b135	-2.04876	141	0.04234	-0.02282
b051	-1.99808	141	0.047632	-0.01711
b084	-1.47794	141	0.141655	-0.00861
b111	-1.26761	141	0.207025	-0.0267
b040	-1.10719	141	0.270095	-0.02957

b075	-0.97607	141	0.330701	-0.01848
b124	-0.57834	141	0.563956	-0.02778
b066	-0.10932	141	0.913107	-0.02108
b055	0.076203	141	0.939366	-0.01405
b050	1.289208	141	0.199437	-0.00452
b045	1.502121	141	0.135302	-0.00535
b114	1.541112	141	0.125531	-0.00316
b076	3.12154	141	0.002183	0.009879
b036	3.255587	141	0.001417	0.012161
b165	3.406389	141	0.000858	0.008215
b081	3.617772	141	0.000413	0.007708
b146	3.666945	141	0.000347	0.013625
b029	3.876374	141	0.000162	0.021663
b058	5.112154	141	1.02E-06	0.018715
b022	6.837958	141	2.24E-10	0.021786
b061	9.125161	141	6.72E-16	0.039697

Significant changes in EUI

Buildings with an energy retrofit

Paired T-Test Normalized Efficiency 2016 Vs 2018 – Unoccupied Summer

Building Name	T-score	DF	P-Value	Efficiency change
b103	-17.2394	47	5.93E-22	-0.19733
b038	-13.3273	47	1.40E-17	-0.20584
b144	-10.6386	47	4.18E-14	-0.10437
b050	-9.52844	47	1.49E-12	-0.12002
b147	-5.39179	47	2.21E-06	-0.06984
b040	-4.36135	47	7.01E-05	-0.12968
b051	-3.14139	47	0.002908	-0.03613
b039	-2.58044	47	0.013051	-0.04939
b101	-2.5052	47	0.015762	-0.05309
b030	-2.36174	47	0.022385	-0.05827
b124	-2.20611	47	0.032301	-0.07629
b025	-2.14458	47	0.037187	-0.06837
b165	-1.83705	47	0.072529	-0.03016
b055	-1.80074	47	0.078162	-0.06842
b123	-1.31909	47	0.193529	-0.04906
b058	-0.26472	47	0.792383	-0.02646

b066	-0.0666	47	0.94718	-0.03326
b084	0.442811	47	0.659934	-0.00416
b104	0.449598	47	0.655067	-0.03092
b029	0.496983	47	0.621518	-0.02787
b045	1.062784	47	0.29331	-0.01395
b099	1.577098	47	0.121481	-0.00718
b031	1.789746	47	0.079939	-0.00163
b081	2.578871	47	0.013103	0.003192
b036	2.688951	47	0.009887	0.012511
b075	2.823244	47	0.00695	0.008384
b061	3.110338	47	0.003173	0.012046
b146	3.373173	47	0.001496	0.024038
b002	3.643768	47	0.000669	0.016703
b061a	3.79605	47	0.00042	0.013561
b111	4.82457	47	1.52E-05	0.046669
b153	5.21582	47	4.04E-06	0.047689
b135	5.704619	47	7.51E-07	0.035032
b114	6.337072	47	8.30E-08	0.067499
b076	6.355871	47	7.77E-08	0.073879
b022	10.69886	47	3.46E-14	0.084075

Significant changes in EUI

Buildings with an energy retrofit

Paired T-Test Normalized Efficiency 2016 Vs 2018 – Occupied Summer

Building Name	T-score	DF	P-Value	Efficiency change
b103	-11.7649	47	1.31E-15	-0.15074
b040	-10.2962	47	1.24E-13	-0.22383
b038	-8.91697	47	1.13E-11	-0.1744
b050	-8.90355	47	1.18E-11	-0.13142
b144	-7.05792	47	6.68E-09	-0.07749
b123	-6.42225	47	6.16E-08	-0.15913
b039	-5.42325	47	1.98E-06	-0.13253
b051	-4.14489	47	0.000141	-0.05329
b084	-3.58107	47	0.000808	-0.03194
b147	-2.62206	47	0.011742	-0.04377
b031	-2.17853	47	0.034416	-0.03561
b135	-1.42127	47	0.161838	-0.03478
b111	-1.33781	47	0.187396	-0.03724

b061	-0.97255	47	0.335754	-0.03661
b025	-0.80504	47	0.424851	-0.053
b030	-0.63511	47	0.528436	-0.03831
b058	0.087308	47	0.930797	-0.02235
b165	0.546172	47	0.587531	-0.01318
b153	0.568355	47	0.5725	-0.01719
b101	0.702985	47	0.485533	-0.01982
b104	1.277256	47	0.207786	-0.00972
b029	1.382365	47	0.173394	-0.01257
b099	2.310428	47	0.025302	0.005353
b045	2.744659	47	0.008552	0.01403
b036	3.319264	47	0.00175	0.03313
b022	3.656649	47	0.000644	0.017392
b124	3.660154	47	0.000637	0.039429
b066	4.205883	47	0.000116	0.045056
b061a	4.691525	47	2.37E-05	0.02347
b055	4.744533	47	1.98E-05	0.040682
b075	4.970525	47	9.29E-06	0.034649
b002	5.376191	47	2.33E-06	0.033388
b146	5.682316	47	8.12E-07	0.066319
b081	5.718506	47	7.16E-07	0.046694
b114	5.758768	47	6.23E-07	0.042693
b076	8.875556	47	1.30E-11	0.131219

Significant changes in EUI

Buildings with an energy retrofit

Paired T-Test Normalized Efficiency 2016 Vs 2018 – Peak Sum

Building Name	T-score	DF	P-Value	Efficiency change
b103	-12.6587	74	3.39E-20	-0.16783
b040	-6.80908	74	2.23E-09	-0.14218
b144	-6.42947	74	1.12E-08	-0.08561
b038	-3.77116	74	0.000325	-0.07692
b147	-3.67147	74	0.000453	-0.05525
b039	-3.06375	74	0.003048	-0.0953
b025	-2.99454	74	0.003736	-0.06604
b050	-2.51836	74	0.013951	-0.05895
b153	-2.46642	74	0.015963	-0.05245
b123	-1.65492	74	0.102175	-0.0717

b111	-1.65423	74	0.102318	-0.04932
b031	-1.1192	74	0.266676	-0.03194
b030	-1.11384	74	0.268955	-0.04734
b114	-0.0627	74	0.950178	-0.02272
b135	0.117284	74	0.906953	-0.02029
b058	0.201361	74	0.840969	-0.0198
b124	0.414508	74	0.679701	-0.0309
b029	0.732822	74	0.465982	-0.02492
b045	0.734147	74	0.465179	-0.0234
b099	1.108151	74	0.271385	-0.01228
b066	1.124646	74	0.264374	-0.01485
b104	1.152248	74	0.252928	-0.01571
b002	1.281528	74	0.20401	-0.0094
b075	1.296936	74	0.198683	-0.00772
b051	1.312286	74	0.19348	-0.00531
b101	1.596665	74	0.114602	-0.00494
b084	1.726583	74	0.088415	-0.00106
b076	2.112215	74	0.038043	0.001943
b165	2.309089	74	0.023732	0.002853
b036	2.594603	74	0.011412	0.011957
b146	2.727534	74	0.007964	0.011229
b055	2.944795	74	0.004317	0.010379
b061a	3.297135	74	0.001502	0.010266
b061	3.474386	74	0.00086	0.015218
b081	3.732522	74	0.00037	0.021028
b022	5.658205	74	2.73E-07	0.031993

Significant changes in EUI

Buildings with an energy retrofit

C.2 Validation: Paired T-Test EUI 2016 vs 2018

Building name	T-score	DF	P-value	EUI change
b099	-10.0160173	330	7.54E-21	-7.20E-05
b165	-9.08305583	330	9.03E-18	-4.62E-05

b051	-7.77338763	330	9.29E-14	-7.83E-05
b036	-6.84876769	330	3.53E-11	-4.09E-05
b050	-6.49569026	330	2.96E-10	-2.32E-05
b029	-5.00394695	330	9.05E-07	-2.82E-05
b040	-4.38441686	330	1.56E-05	-1.01E-05
b075	-2.09037306	330	0.037331	-9.04E-06
b104	-1.19390081	330	0.233354	-7.33E-06
b146	0.40113062	330	0.688577	3.10E-06
b061	0.79869382	330	0.425029	3.08E-06
b058	1.56955051	330	0.117455	6.81E-06
b025	2.60527233	330	0.009586	1.32E-05
b123	2.72693762	330	0.006726	1.17E-05
b111	2.80270807	330	0.00536	9.20E-06
b055	2.9520527	330	0.003377	2.80E-05
b031	3.10986673	330	0.002031	1.78E-05
b022	4.13260426	330	4.53E-05	1.45E-05
b081	4.32127917	330	2.04E-05	4.04E-05
b045	6.85777305	330	3.34E-11	2.86E-05
b135	7.28330584	330	2.31E-12	3.45E-05
b076	8.45946527	330	8.22E-16	1.09E-05
b101	8.87817056	330	4.06E-17	4.08E-05
b084	9.11082562	330	7.36E-18	4.70E-05
b038	9.21515675	330	3.39E-18	4.84E-05
b147	9.89779849	330	1.89E-20	4.56E-05
b066	10.7571478	330	2.09E-23	5.08E-05
b039	11.6057309	330	1.95E-26	4.06E-05
b114	11.9714837	330	8.98E-28	4.86E-05
b124	13.5673327	330	8.95E-34	3.65E-05
b002	16.8602745	330	1.02E-46	4.41E-05
b153	17.3658398	330	9.78E-49	5.05E-05
b061a	18.9467676	330	4.60E-55	4.70E-05
b103	27.0245638	330	1.80E-86	0.000114
b144	50.5306418	330	1.33E-159	0.000183
b030	63.6460298	330	3.05E-190	0.000526

Significant changes in EUI

	Buildings with an energy retrofit

C.3 Validation Pearson Test: EUI change vs Efficiency Change

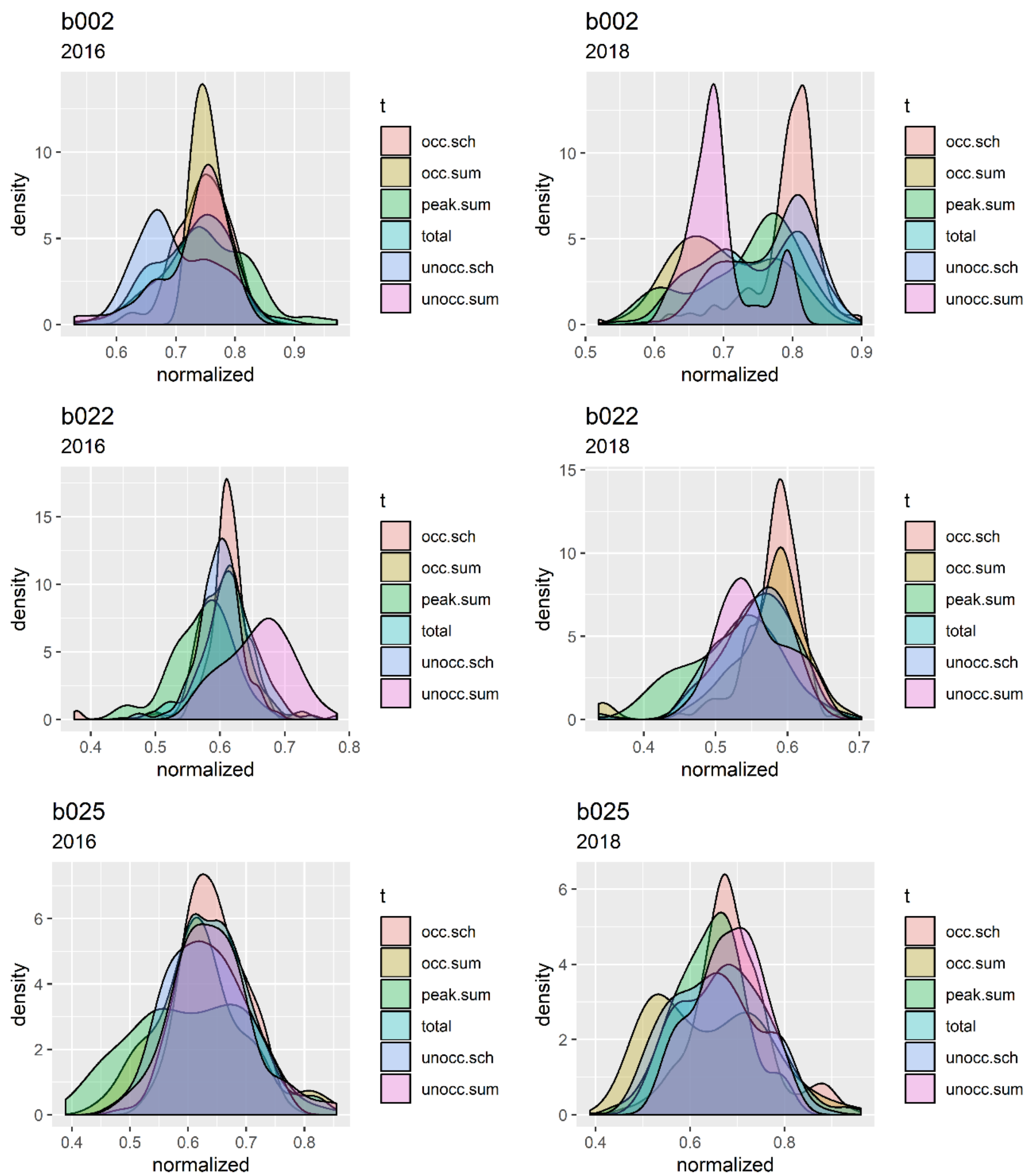
Building Name	X1	X2	Significance	Coefficient
b165	-13.9832	330	2.34E-35	-0.605945997
b050	-12.4564	330	1.48E-29	-0.561484791
b111	-12.4154	330	2.11E-29	-0.560218534
b055	-11.6873	330	1.01E-26	-0.537045879
b036	-8.54481	330	4.53E-16	-0.421991156
b061	-7.54623	330	4.22E-13	-0.38019967
b114	-7.21205	330	3.66E-12	-0.365659153
b103	-6.52787	330	2.45E-10	-0.33504348
b099	-5.41277	330	1.18E-07	-0.282814859
b040	-4.76395	330	2.83E-06	-0.251188718
b123	-4.61236	330	5.66E-06	-0.243677618
b124	-4.38898	330	1.53E-05	-0.232529668
b025	-3.83283	330	0.000151124	-0.204380303
b135	-3.62545	330	0.000332997	-0.193748881
b031	-3.56895	330	0.000410424	-0.190840169
b081	-2.07738	330	0.038523582	-0.112444356
b147	-1.94221	330	0.05294501	-0.105211524
b022	-1.85725	330	0.064148479	-0.100656779
b058	-1.43323	330	0.152720171	-0.077835965
b051	-0.81306	330	0.416755391	-0.044247042
b144	-0.55911	330	0.576460551	-0.030442343
b045	0.776334	330	0.4380961	0.042251863
b146	0.994288	330	0.320796207	0.054083031
b084	1.735628	330	0.083543676	0.094125935
b038	1.870247	330	0.062316213	0.101354217
b101	2.778322	330	0.005770199	0.149640766
b039	3.134273	330	0.001874147	0.168299372
b029	3.134418	330	0.001873249	0.168306925
b002	3.516453	330	0.000497227	0.188132967

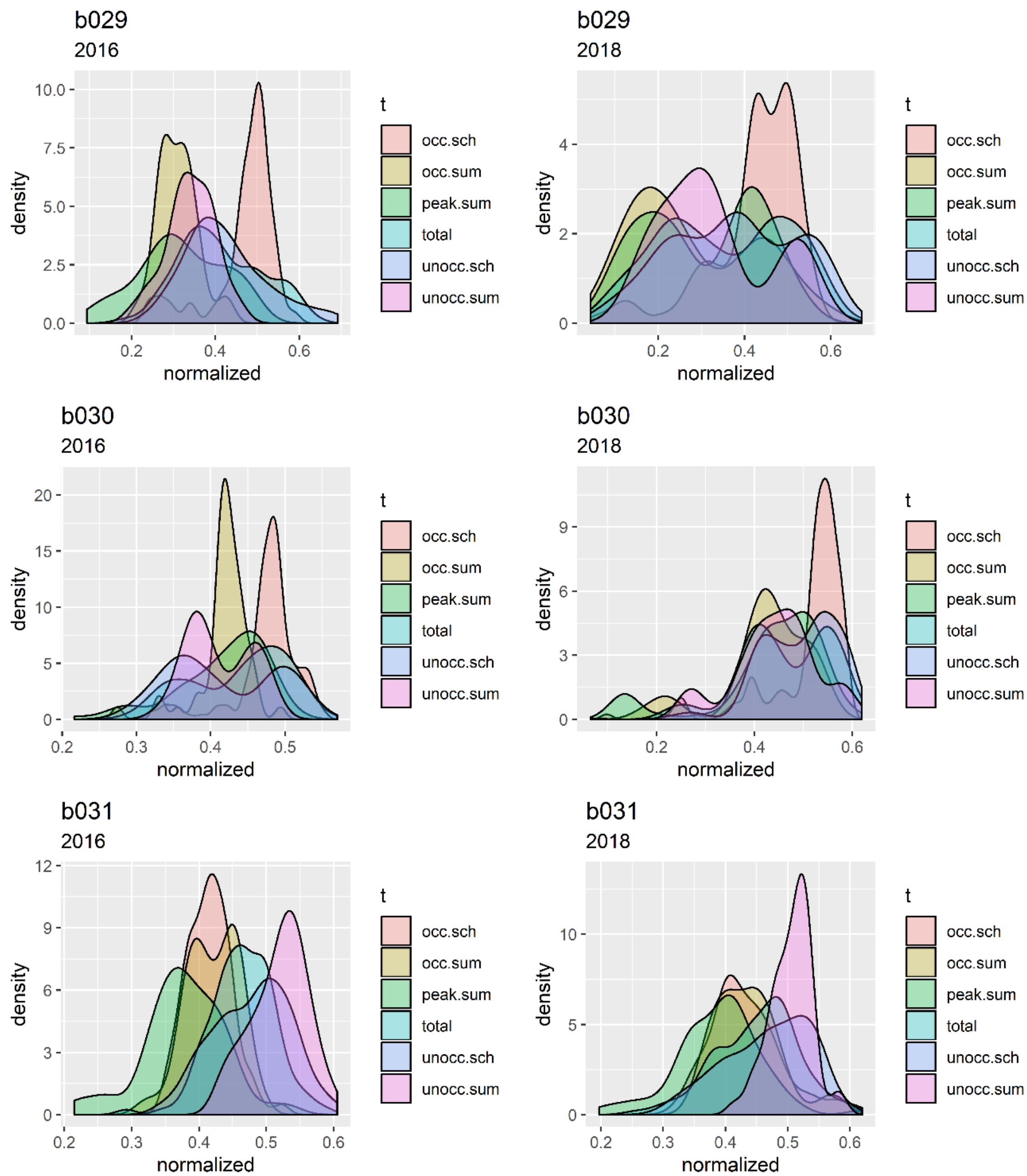
b061a	3.566559	330	0.000414052	0.190716781
b066	3.908568	330	0.000112245	0.208245505
b030	4.15088	330	4.20E-05	0.220545216
b076	5.039699	330	7.62E-07	0.264735014
b104	5.219391	330	3.15E-07	0.273479495
b075	5.870469	330	1.04E-08	0.304589827
b153	6.866727	330	3.17E-11	0.350346886

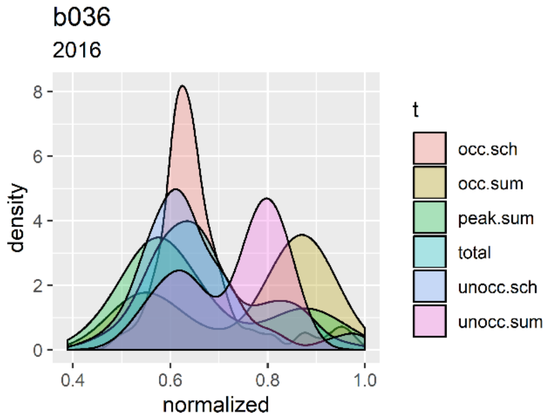
	Significant changes in EUI
	Buildings with an energy retrofit

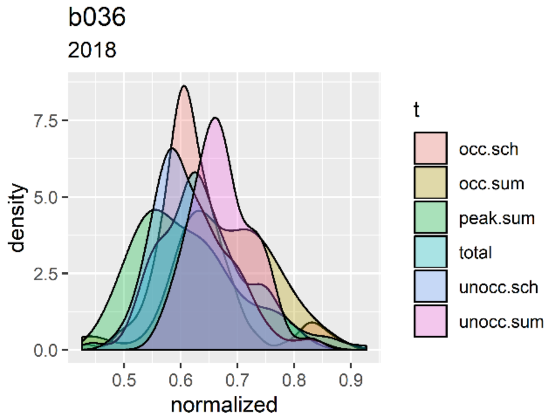
APPENDIX D: PLOTS

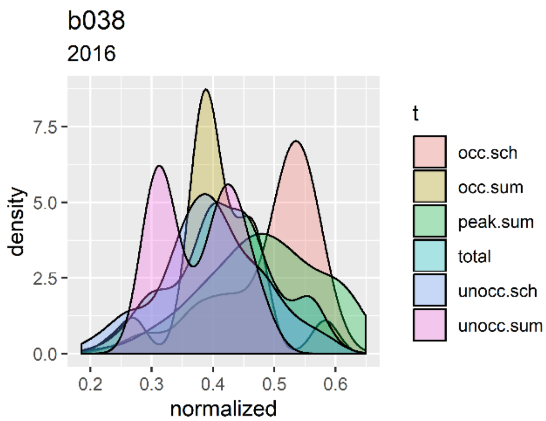
Comparison of normalized energy efficiency rankings between period A and period B

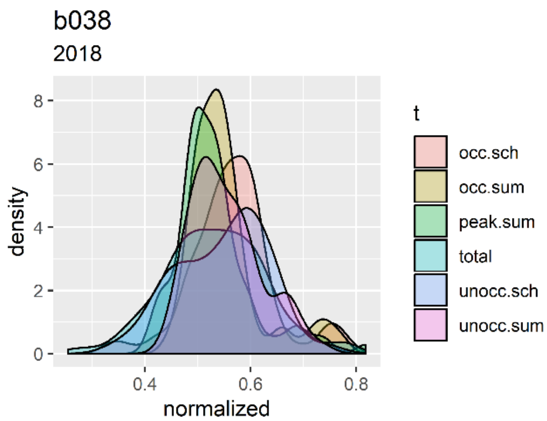


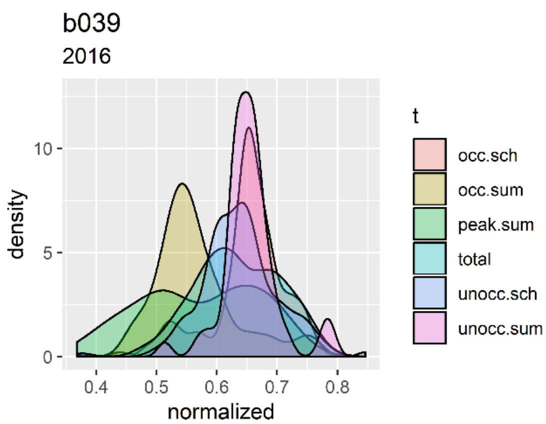

b029
2016
b029
2018
b030
2016
b030
2018
b031
2016
b031
2018
density
normalized
t
occ.sch
occ.sum
peak.sum
total
unocc.sch
unocc.sum

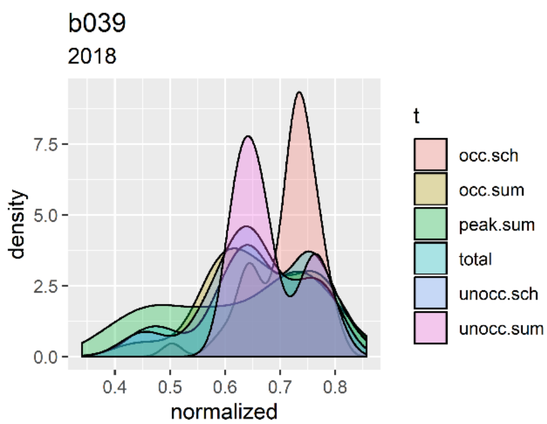

b036
2016
density
normalized
t
occ.sch
occ.sum
peak.sum
total
unocc.sch
unocc.sum

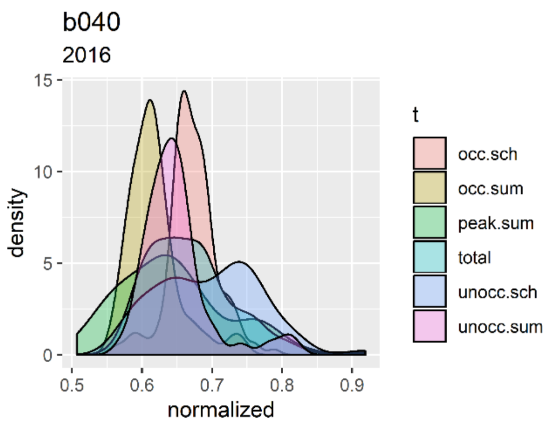

b036
2018
density
normalized
t
occ.sch
occ.sum
peak.sum
total
unocc.sch
unocc.sum

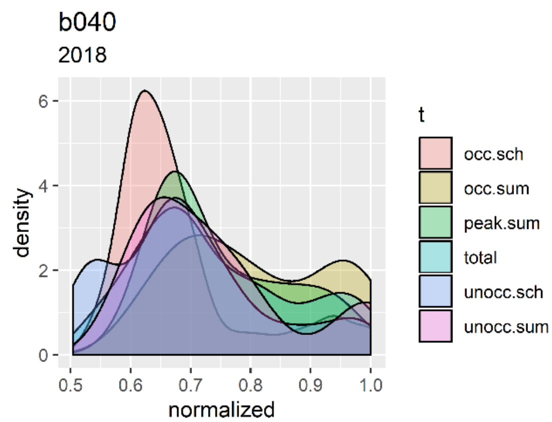

b038
2016
density
normalized
t
occ.sch
occ.sum
peak.sum
total
unocc.sch
unocc.sum

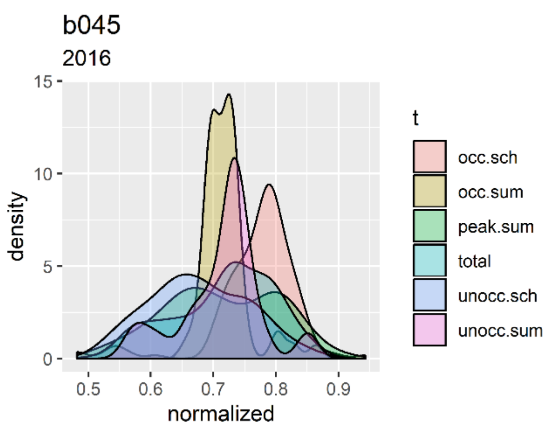

b038
2018
density
normalized
t
occ.sch
occ.sum
peak.sum
total
unocc.sch
unocc.sum

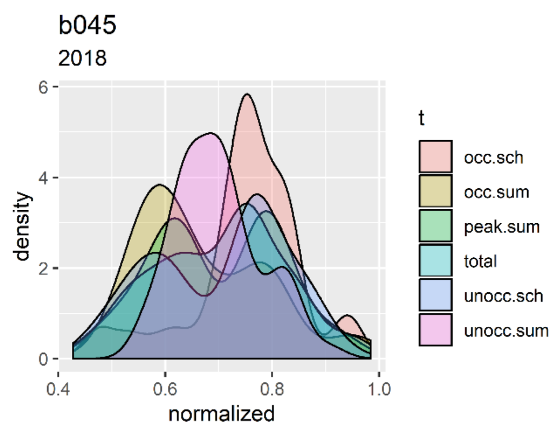

b039
2016
density
normalized
t
occ.sch
occ.sum
peak.sum
total
unocc.sch
unocc.sum

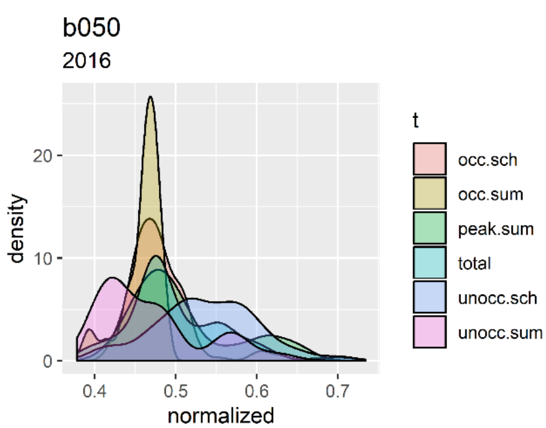

b039
2018
density
normalized
t
occ.sch
occ.sum
peak.sum
total
unocc.sch
unocc.sum

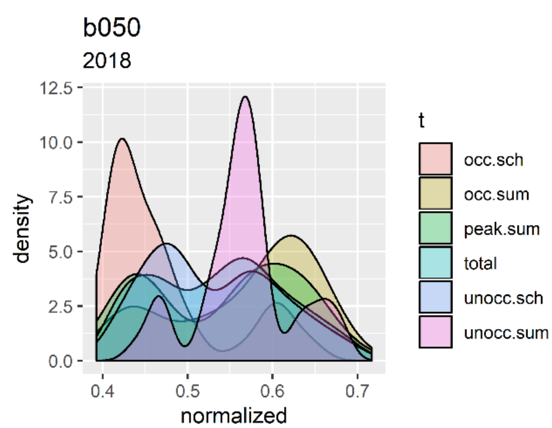

b040
2016
density
normalized
15
10
5
0
0.5
0.6
0.7
0.8
0.9
t
occ.sch
occ.sum
peak.sum
total
unocc.sch
unocc.sum

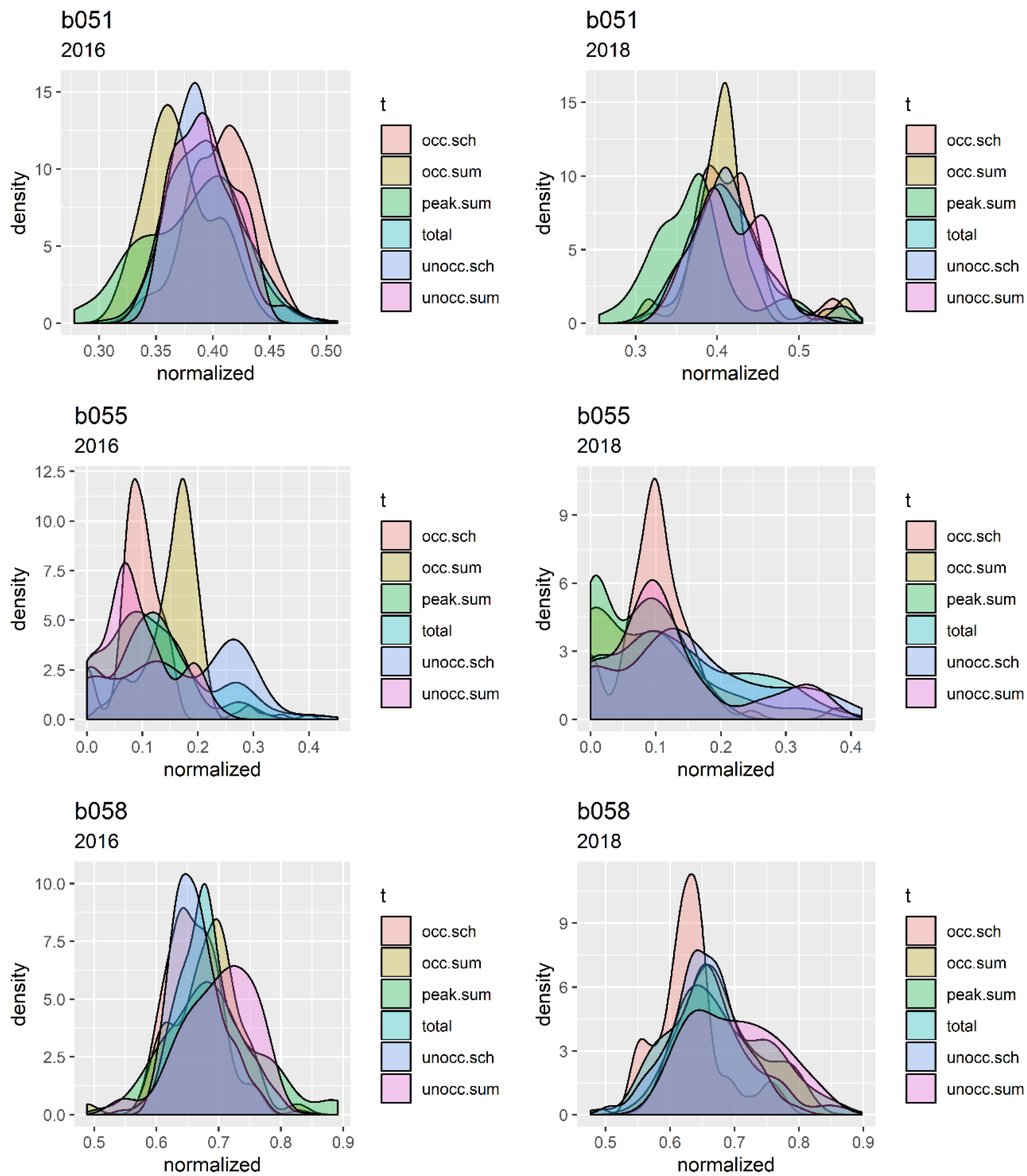

b040
2018
density
normalized
6
4
2
0
0.5
0.6
0.7
0.8
0.9
1.0
t
occ.sch
occ.sum
peak.sum
total
unocc.sch
unocc.sum

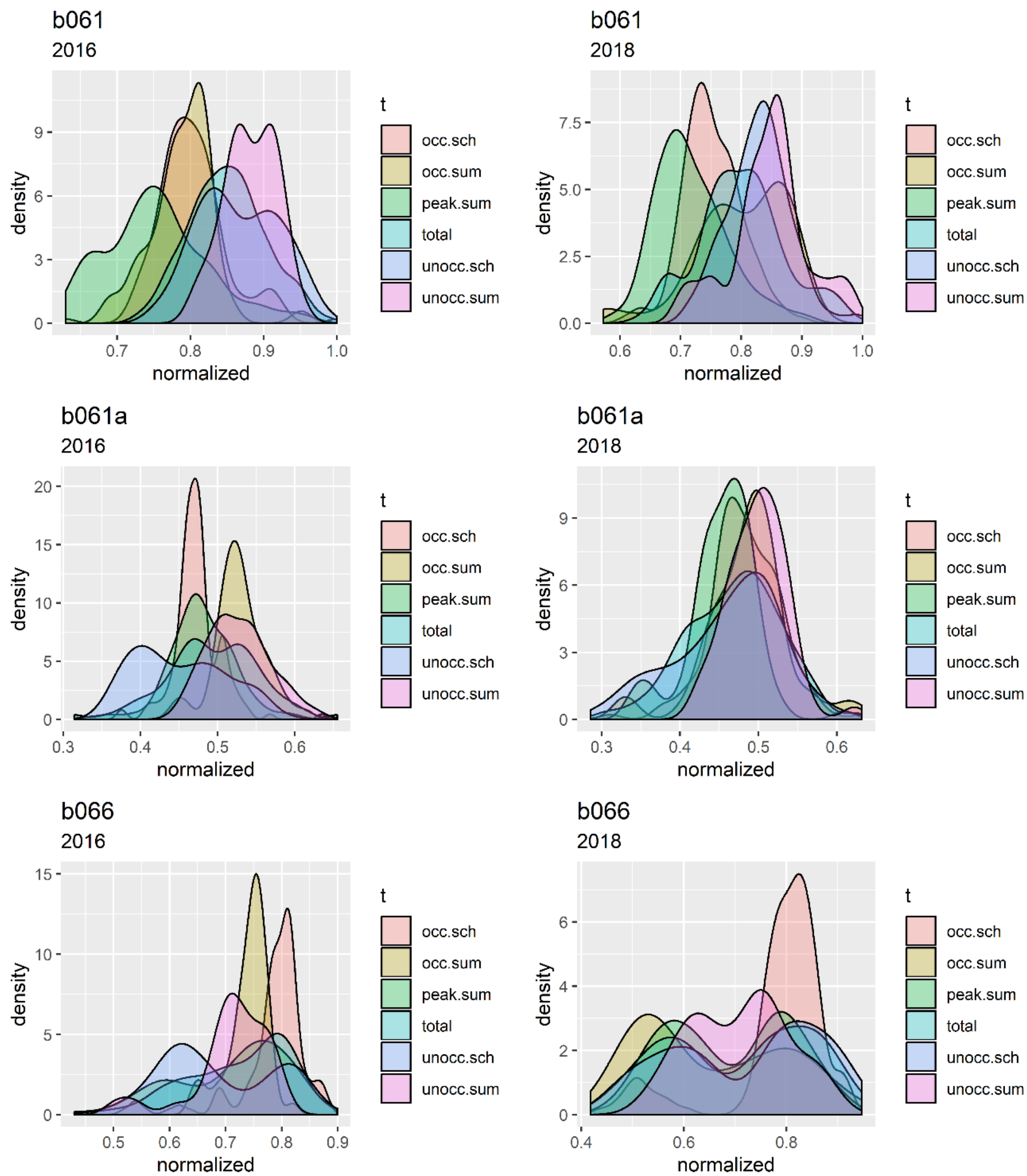

b045
2016
density
normalized
15
10
5
0
0.5
0.6
0.7
0.8
0.9
t
occ.sch
occ.sum
peak.sum
total
unocc.sch
unocc.sum

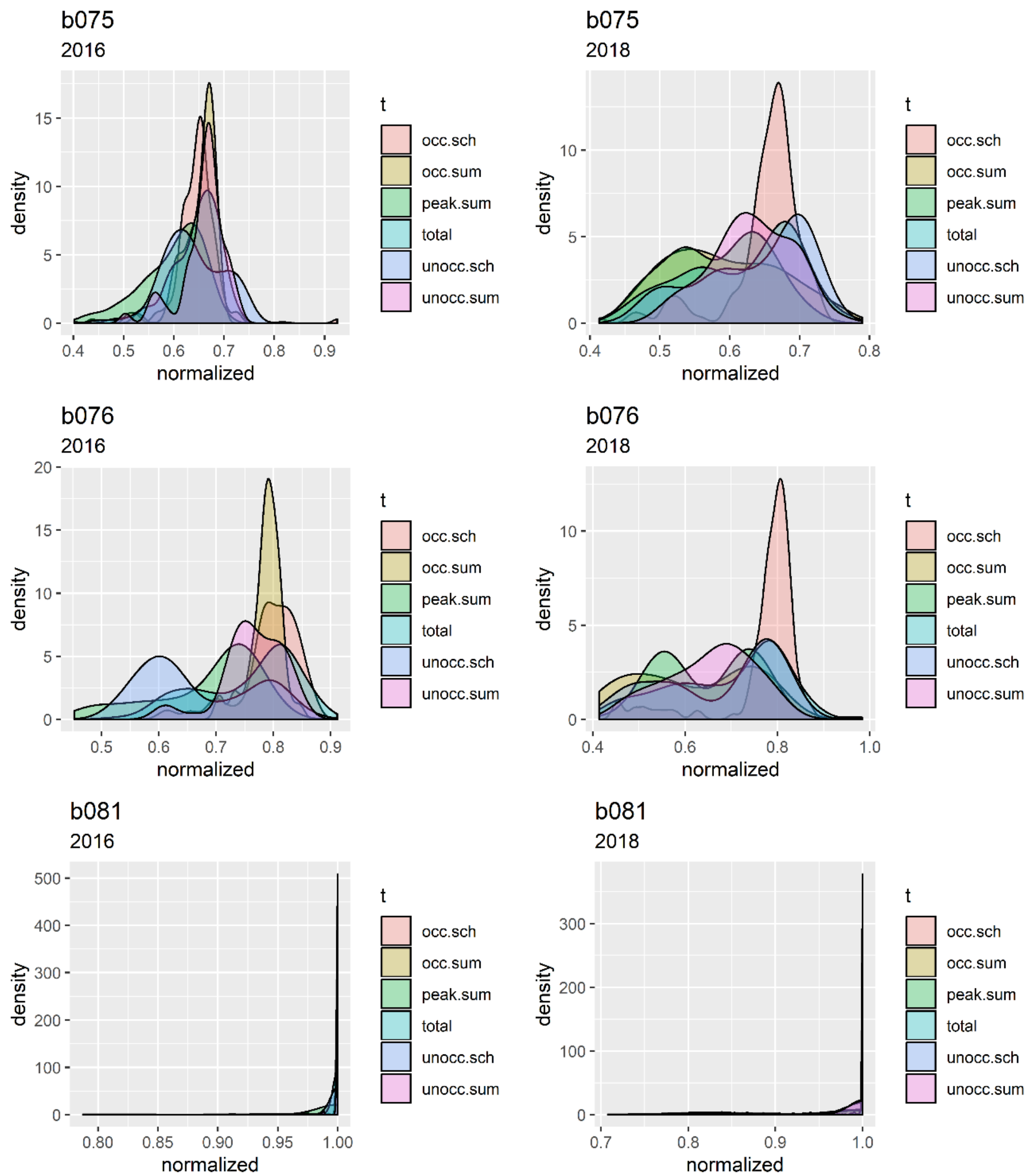

b045
2018
density
normalized
6
4
2
0
0.4
0.6
0.8
1.0
t
occ.sch
occ.sum
peak.sum
total
unocc.sch
unocc.sum

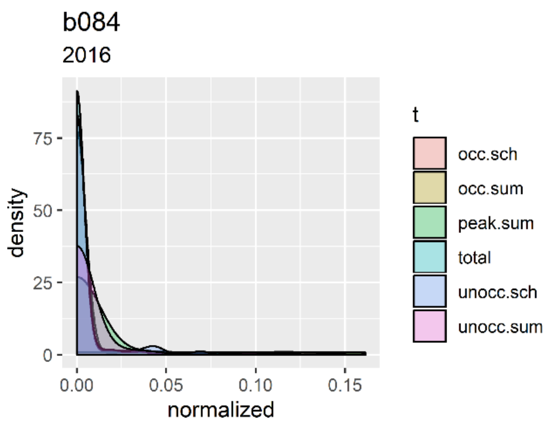

b050
2016
density
normalized
20
10
0
0.4
0.5
0.6
0.7
t
occ.sch
occ.sum
peak.sum
total
unocc.sch
unocc.sum

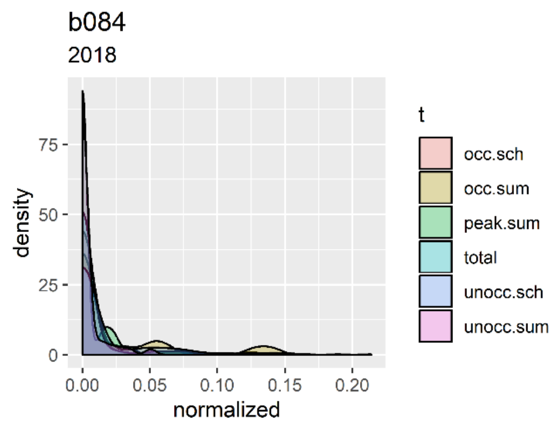

b050
2018
density
normalized
12.5
10.0
7.5
5.0
2.5
0.0
0.4
0.5
0.6
0.7
t
occ.sch
occ.sum
peak.sum
total
unocc.sch
unocc.sum

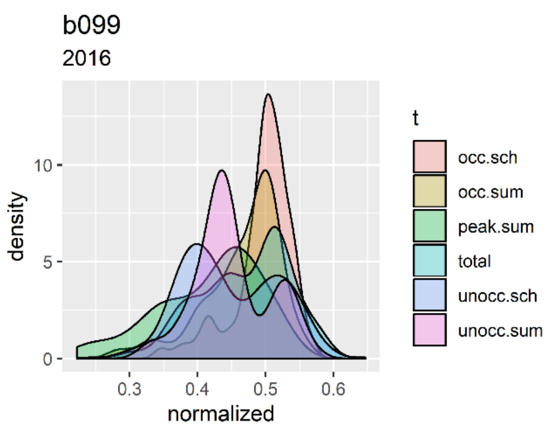

b051
2016
b051
2018
b055
2016
b055
2018
b058
2016
b058
2018
density
normalized
t
occ.sch
occ.sum
peak.sum
total
unocc.sch
unocc.sum

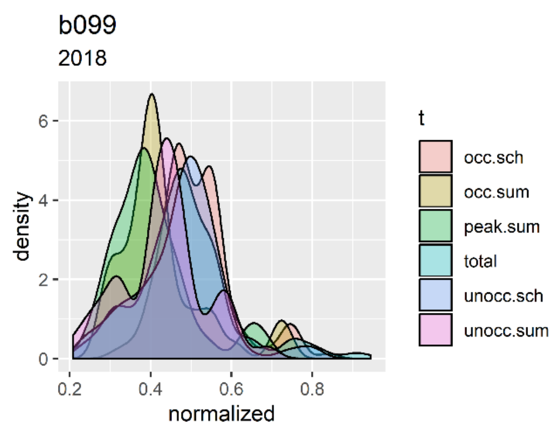

b061
2016
b061
2018
b061a
2016
b061a
2018
b066
2016
b066
2018
density
normalized
t
occ.sch
occ.sum
peak.sum
total
unocc.sch
unocc.sum

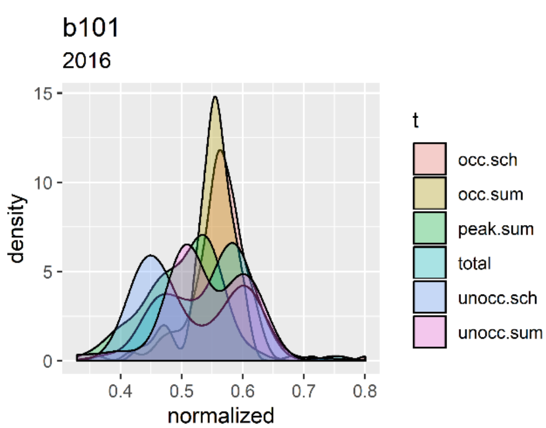

b075
2016
b075
2018
b076
2016
b076
2018
b081
2016
b081
2018
density
normalized
t
occ.sch
occ.sum
peak.sum
total
unocc.sch
unocc.sum

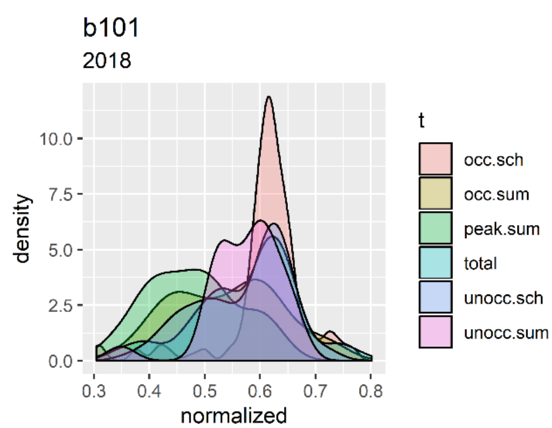

b084
2016
density
75
50
25
0
0.00
0.05
0.10
0.15
normalized
t
occ.sch
occ.sum
peak.sum
total
unocc.sch
unocc.sum

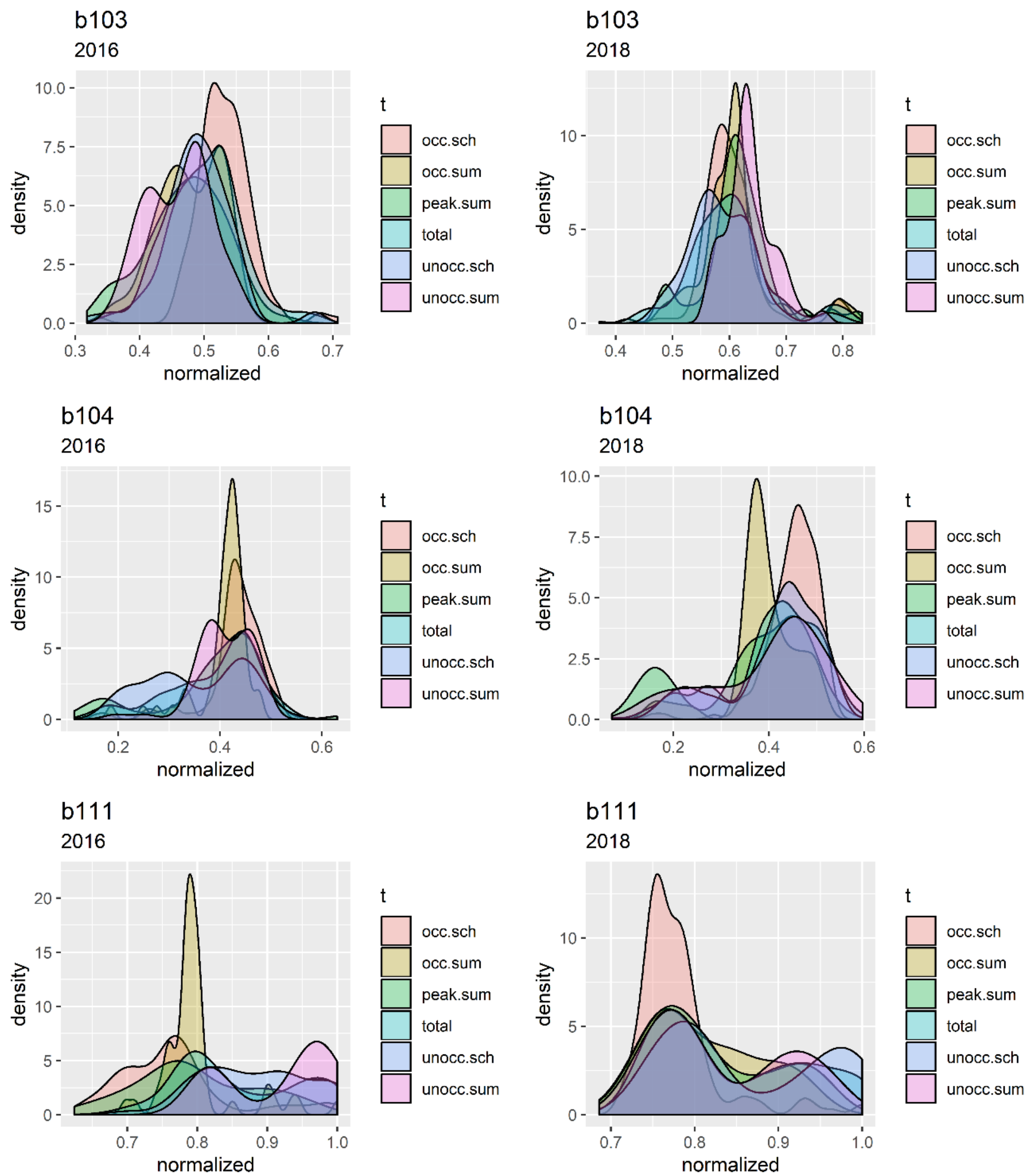

b084
2018
density
75
50
25
0
0.00
0.05
0.10
0.15
0.20
normalized
t
occ.sch
occ.sum
peak.sum
total
unocc.sch
unocc.sum


b099
2016
density
10
5
0
0.3
0.4
0.5
0.6
normalized
t
occ.sch
occ.sum
peak.sum
total
unocc.sch
unocc.sum

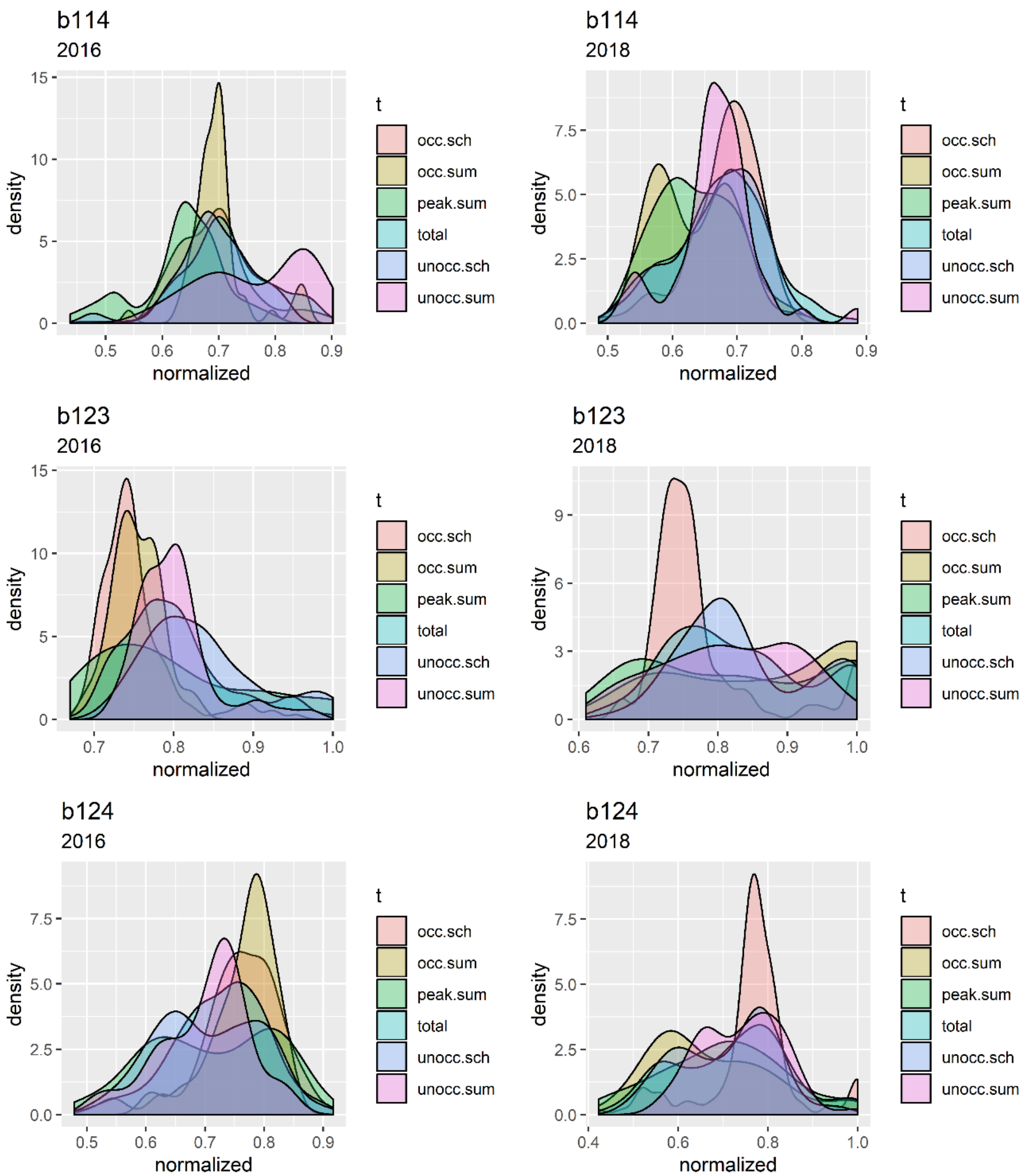

b099
2018
density
6
4
2
0
0.2
0.4
0.6
0.8
normalized
t
occ.sch
occ.sum
peak.sum
total
unocc.sch
unocc.sum

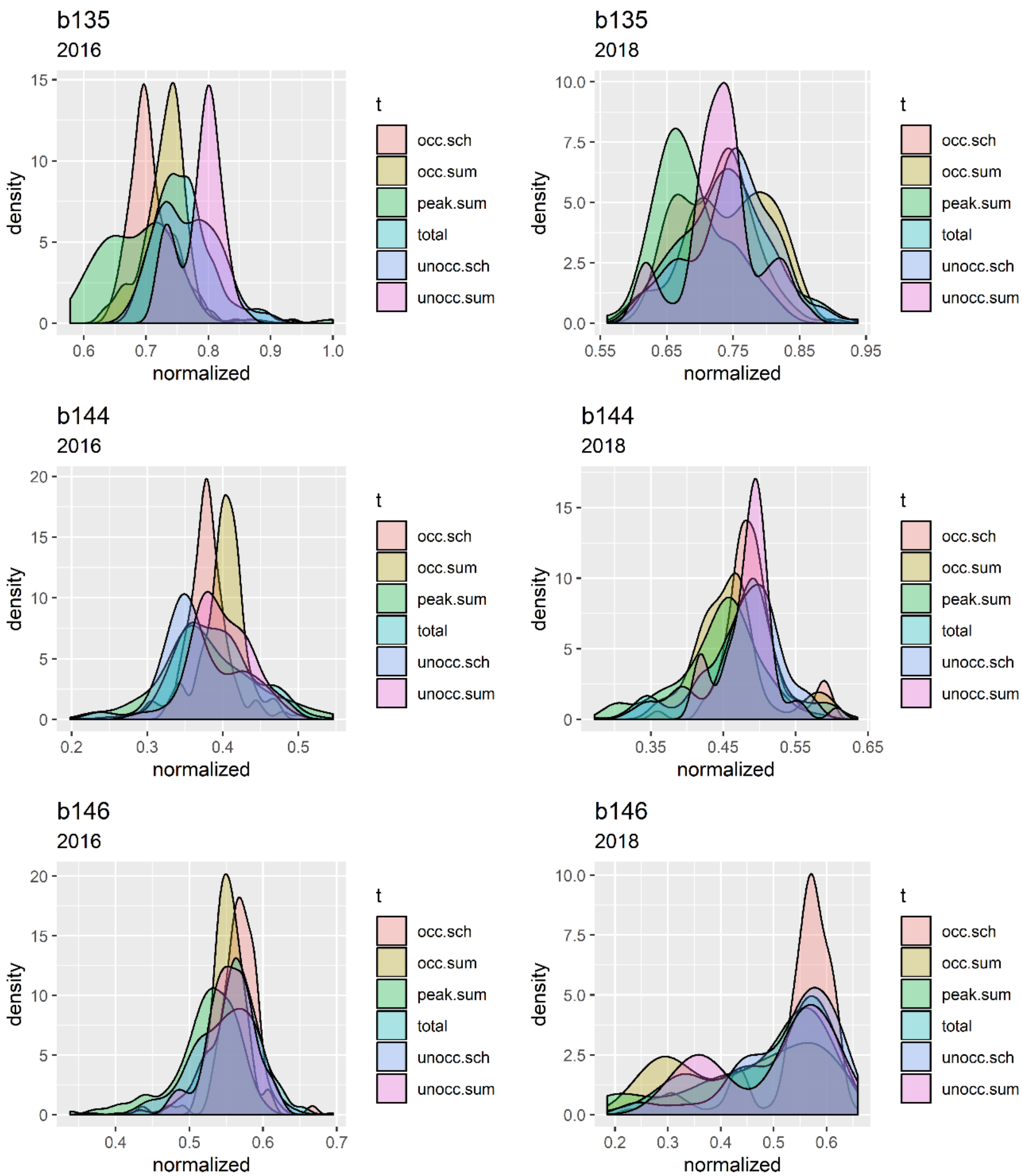

b101
2016
density
15
10
5
0
0.4
0.5
0.6
0.7
0.8
normalized
t
occ.sch
occ.sum
peak.sum
total
unocc.sch
unocc.sum

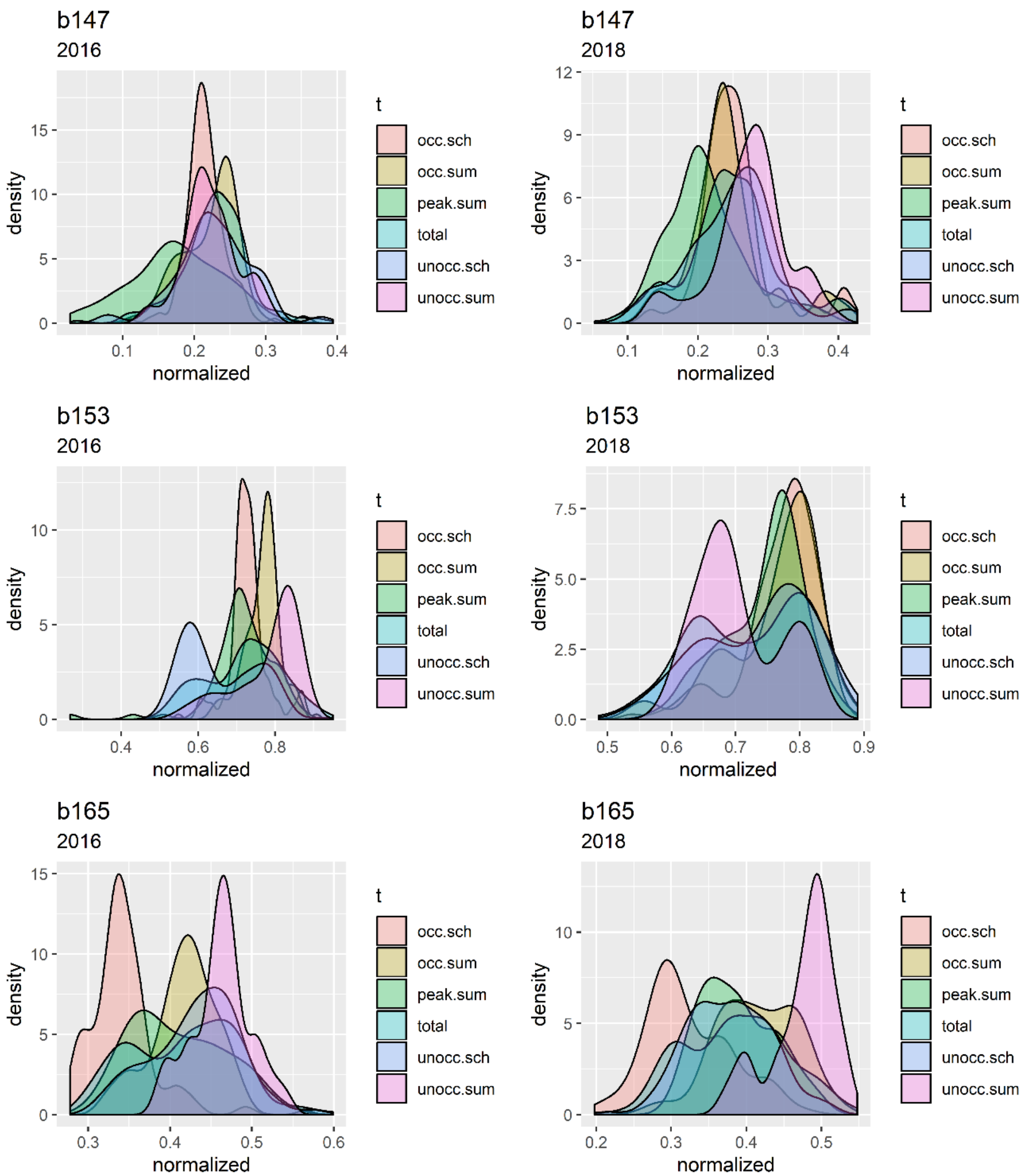

b101
2018
density
10.0
7.5
5.0
2.5
0.0
0.3
0.4
0.5
0.6
0.7
0.8
normalized
t
occ.sch
occ.sum
peak.sum
total
unocc.sch
unocc.sum



b103
2016
b103
2018
b104
2016
b104
2018
b111
2016
b111
2018
density
normalized
t
occ.sch
occ.sum
peak.sum
total
unocc.sch
unocc.sum


b114
2016
b114
2018
b123
2016
b123
2018
b124
2016
b124
2018
density
normalized
t
occ.sch
occ.sum
peak.sum
total
unocc.sch
unocc.sum


b135
2016
density
normalized
0.6
0.7
0.8
0.9
1.0
0
5
10
15
t
occ.sch
occ.sum
peak.sum
total
unocc.sch
unocc.sum
b135
2018
density
normalized
0.55
0.65
0.75
0.85
0.95
0.0
2.5
5.0
7.5
10.0
t
occ.sch
occ.sum
peak.sum
total
unocc.sch
unocc.sum
b144
2016
density
normalized
0.2
0.3
0.4
0.5
0
5
10
15
20
t
occ.sch
occ.sum
peak.sum
total
unocc.sch
unocc.sum
b144
2018
density
normalized
0.35
0.45
0.55
0.65
0
5
10
15
t
occ.sch
occ.sum
peak.sum
total
unocc.sch
unocc.sum
b146
2016
density
normalized
0.4
0.5
0.6
0.7
0
5
10
15
20
t
occ.sch
occ.sum
peak.sum
total
unocc.sch
unocc.sum
b146
2018
density
normalized
0.2
0.3
0.4
0.5
0.6
0.0
2.5
5.0
7.5
10.0
t
occ.sch
occ.sum
peak.sum
total
unocc.sch
unocc.sum


b147
2016
density
normalized
t
occ.sch
occ.sum
peak.sum
total
unocc.sch
unocc.sum
b147
2018
b153
2016
b153
2018
b165
2016
b165
2018

REFERENCES

[1] Francisco A, Mohammadi N, Taylor JE. Evaluating Temporal Shifts in City Scale Building Energy Benchmarks. Constr. Res. Congr. 2018, n.d., p. 450–60. doi:10.1061/9780784481301.045.

[2] Fernández JE. REVIEW Materials for Aesthetic, Energy-Efficient, and Self-Diagnostic Buildings 2007:1807–11.

[3] Administration USEI. Monthly Energy Review March 2017. vol. 0035. 2017.

[4] Energy MG. Unlocking Energy Efficiency in the U.S. Economy. Energy 2009:1–165.

[5] Wang S, Yan C, Xiao F. Quantitative energy performance assessment methods for existing buildings. Energy Build 2012;55:873–88. doi:10.1016/j.enbuild.2012.08.037.

[6] Pérez-Lombard L, Ortiz J, González R, Maestre IR. A review of benchmarking, rating and labelling concepts within the framework of building energy certification schemes. Energy Build 2009;41:272–8. doi:10.1016/j.enbuild.2008.10.004.

[7] Department of Energy (DOE). Designing a benchmarking plan 2013:1–11.

[8] Li Z, Han Y, Xu P. Methods for benchmarking building energy consumption against its past or intended performance: An overview. Appl Energy 2014;124:325–34. doi:10.1016/j.apenergy.2014.03.020.

[9] EIA U. How many smart meters are installed in the United States, and who has

them? 2015.

[10] Chung W, Hui Y V., Lam YM. Benchmarking the energy efficiency of commercial buildings. Appl Energy 2006;83:1–14. doi:10.1016/j.apenergy.2004.11.003.

[11] Karmellos M, Kiprakis A, Mavrotas G. A multi-objective approach for optimal prioritization of energy efficiency measures in buildings: Model, software and case studies. Appl Energy 2015;139:131–50. doi:10.1016/j.apenergy.2014.11.023.

[12] Chidiac SE, Catania EJC, Morofsky E, Foo S. A screening methodology for implementing cost effective energy retrofit measures in Canadian office buildings. Energy Build 2011;43:614–20. doi:10.1016/j.enbuild.2010.11.002.

[13] Hong T, Yang L, Hill D, Feng W. Data and analytics to inform energy retrofit of high performance buildings. Appl Energy 2014;126:90–106. doi:10.1016/j.apenergy.2014.03.052.

[14] Alajmi A. Energy audit of an educational building in a hot summer climate. Energy Build 2012;47:122–30. doi:10.1016/j.enbuild.2011.11.033.

[15] Pacific Northwest National Laboratory, PECI. Advance Energy Retrofit Guide: Practical Ways to Improve EnergyPerformance 2011.

[16] Greely KM, Harris JP, Hatcher AM. Measured Savings and Cost-Effectiveness of Conservation Retrofits in Commercial Buildings. Volume 1: Analysis and Results 1990;1.

[17] Liang J, Qiu Y, James T, Ruddell BL, Dalrymple M, Earl S, et al. Do energy

retrofits work? Evidence from commercial and residential buildings in Phoenix. J Environ Econ Manage 2017;92:726–43. doi:10.1016/j.jeem.2017.09.001.

[18] Fowlie M, Greenstone M, Wolfram C. DO ENERGY EFFICIENCY INVESTMENTS DELIVER ? EVIDENCE FROM THE WEATHERIZATION ASSISTANCE Energy efficiency investments are widely believed to offer a rare win-win opportunity . Detailed engineering projections , such as those summarized by the well-known Mc. Q J Econ 2018:1–48. doi:10.1093/qje/qjy005.Advance.

[19] Rosenow J, Galvin R. Evaluating the evaluations: Evidence from energy efficiency programmes in Germany and the UK. Energy Build 2013;62:450–8. doi:10.1016/j.enbuild.2013.03.021.

[20] Scheer J, Clancy M, Hógáin SN. Quantification of energy savings from Ireland's Home Energy Saving scheme: An ex post billing analysis. Energy Effic 2013;6:35–48. doi:10.1007/s12053-012-9164-8.

[21] Krarti M, Kreider JF, Cohen D, Curtiss P. Estimation of Energy Savings for Building Retrofits Using Neural Networks. J Sol Energy Eng 1998;120:211. doi:10.1115/1.2888071.

[22] Lee SH, Hong T, Piette MA, Taylor-Lange SC. Energy retrofit analysis toolkits for commercial buildings: A review. Energy 2015;89:1087–100. doi:10.1016/j.energy.2015.06.112.

[23] Xuchao W, Priyadarsini R, Siew Eang L. Benchmarking energy use and greenhouse gas emissions in Singapore's hotel industry. Energy Policy 2010. doi:10.1016/j.enpol.2010.04.006.

[24] Li Z, Han Y, Xu P. Methods for benchmarking building energy consumption against its past or intended performance: An overview. Appl Energy 2014;124:325–34. doi:10.1016/j.apenergy.2014.03.020.

[25] Pérez-Lombard L, Ortiz J, Pout C. A review on buildings energy consumption information. Energy Build 2008;40:394–8. doi:10.1016/j.enbuild.2007.03.007.