Technical University of Denmark

02441 Applied Statistics and Statistical Software

Case 2: Energy performance of buildings

Group 5:
Begoña Bolós Sierra, s193036
Laura Sans Comerma, s192437
Jorge Montalvo Arvizu, s192184

January 2020

Summary

In order to improve the management of the energy consumption in public buildings in Høje Taastrup municipality, it is requested to test and discover which parameters influence energy consumption the most and which buildings should be prioritized for the retrofitting improvements. For that, the statistical analysis of the collected parameters, e.g. temperature, humidity or building type; are analysed in order to find a linear regression model that sheds light on the matter. The final model shows a dependency of heat consumption on several variables and five buildings were identified as retrofitting candidates given its low insulation, which was calculated given the physical equation of heat loss and the coefficients of a general linear model. Finally, some additional analysis could have been performed, taking into account that the comfort temperatures vary depending of the building.

Contents

List of Figures

List of Tables

1 Introduction

The reduction of the energy consumption is one of the biggest challenges for citizens and governments nowadays. There are several reasons to reduce or moderate the consumption of energy, for instance the limitation of some limited energy resources such as oil, natural gas or coal. Furthermore, the production of energy from fossil fuels is having a negative impact in the planet, because of the release of greenhouse gas emissions that are causing an environmental crisis. However, the traditional energy production is not decreasing even with the implementation of new ways to produce energy with renewable sources [1].

There are sectors that consume more energy than others, therefore they are more likely to be the focused of improvements to reduce energy consumption. For instance, in Denmark, public or private buildings represent almost half of consumed energy consumption (Fig. 1) [2].

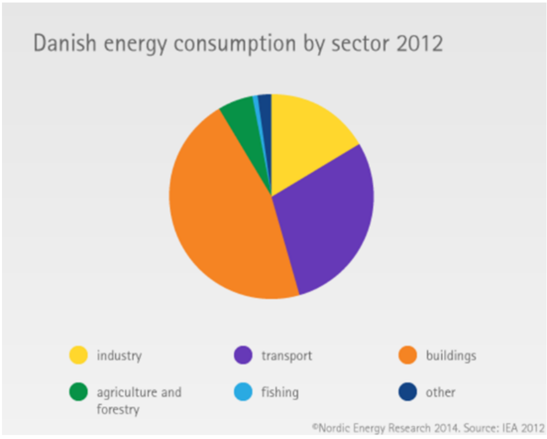

Figure 1: Danish energy consumption by sector (2012).

In the case of Høje Taastrup municipality (HTK), the energy consumption of the public buildings is a great candidate to be selected for the energy consumption improvement. Retrofitting is a solution for high energy-consuming buildings, in order to make them more sustainable and improve their performance and efficiency. There are several energy-efficiency strategies that can be applied [3]. However, these processes can be expensive, for that reason it is interesting to begin the improvements with the buildings that have the poorest performance. Those buildings with low performance are usually large and old, which use a large proportion of the energy for heating.

To determine which buildings are going to be prioritized for the retrofitting improvements, statistical analysis tools are going to be used in order to create a model that sheds light on the problem.

To achieve this, the analysis is going to be performed in two parts: the first part is based on the cleansing of the two data sets (WUnderground data and *meter data*), whilst the second part is the statistical analysis of those cleaned data sets. In order to carry out the second part of the project, the modelling is going to be based on the following physics formula that shows the heat loss through a wall to estimate the insulation parameter of each building:

$$Q_{heat} = U_a(T_{indoor} + T_{outdoor}) \tag{1}$$

The general lineal regression model will be carried out considering the temperature indoor ($T_{indoors}$) as constant at 21ºC[1]. Moreover, the amount of insulation (U_a) will be estimated from this general linear regression model and other weather variables (e.g. humidity and pressure) will be analysed.

After the statistical analysis, we expect to obtain a linear regression model that shows the variables that are significant in the heating loss, together with the determination of the buildings that should be prioritized for the retrofitting from the estimated insulation parameters.

[1] 21 refers to the *Thermal Comfort Standard* that has been set as a number between 20 and 22 ºC degrees by the American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE).

2 Data

2.A Data cleansing

The data cleansing consists on the preparation and later merging of two data sets; UWunderground (WG) data and the *meter data* (HTK consumption reads per day). First of all, the provided files and their structures were inspected.

In the case of WG data, the fixed values columns or pure `NAs` columns were removed. After this for all the continuous variables the means were calculated by date and for the factor variables the mode was also calculated. After this, the results were merged into one single data frame of length 120 observations.

Regarding the *meter data*, the information of the different files were stored in a data frame with 11422 observations and 26 variables. After that, the data frame was reduced to the 3 variables of interest, which are *ID*, *Time* (date and hour) and *Readings*. Afterwards, the buildings (*ID*s) with less than 121 readings were eliminated of the analysis. To obtain the final data frame, the consumption was necessary to be calculated for the different buildings. To achieve this, an interpolation was performed at 11.59pm per each day, and the consumption was calculated as the difference between readings.

Finally, WU data and *meter data* cleaned were merged together to obtain a data frame with 9794 observations and 14 variables. However, 1 shows 12 variables, due to *ID*s and *date* is not included in this summary.

temp	dew_pt	hum	wind_spd
Min. :-1.800	Min. :-3.600	Min. :49.00	Min. : 3.713
1st Qu.: 4.579	1st Qu.: 2.190	1st Qu.:73.10	1st Qu.:11.305
Median : 8.905	Median : 6.833	Median :82.29	Median :15.195
Mean : 8.733	Mean : 6.317	Mean :81.00	Mean :16.367
3rd Qu.:12.833	3rd Qu.: 9.947	3rd Qu.:89.30	3rd Qu.:20.786
Max. :18.500	Max. :15.583	Max. :98.39	Max. :41.929

dir	vis	pressure	cond
South :1081	Min. : 1.965	Min. : 985.8	Scattered Clouds:3395
SW :1079	1st Qu.:11.710	1st Qu.:1011.1	Mist :2575
SE :1076	Median :19.491	Median :1017.7	Clear :1328
West : 908	Mean :21.267	Mean :1016.6	Mostly Cloudy : 748
ESE : 745	3rd Qu.:29.900	3rd Qu.:1022.4	Fog : 663
NE : 666	Max. :50.000	Max. :1040.2	Light Rain : 252

fog	rain	snow	consumption
Min. :0.00000	Min. :0.00000	Min. :0.000000	Min. :0.00000
1st Qu.:0.00000	1st Qu.:0.00000	1st Qu.:0.000000	1st Qu.:0.07572
Median :0.00000	Median :0.05263	Median :0.000000	Median :0.15271
Mean :0.05233	Mean :0.12569	Mean :0.009586	Mean :0.43909
3rd Qu.:0.00000	3rd Qu.:0.18182	3rd Qu.:0.000000	3rd Qu.:0.33878
Max. :0.65000	Max. :0.63636	Max. :0.375000	Max. :8.70171

Table 1: Summary statistics of the merged data WU and *meter data*.

Hence, the merged data (Fig. 1) has the same number of observations than the raw data (Fig. 3), which corresponds to 9794 rows. The number of variables would have been the same if the data frame with the raw data had maintained the variable *snow*. Regarding the variable's values of both summaries the differences are negligible. The variation in the consumption may be due to the interpolation, which was necessary to be done in order to obtain the consumption for each building.

2.B Data inspection

As mentioned above, the data used in the analysis consists of combination of three previously cleaned data sets: HTK building data share, WG and *meter data* , respectively. The period of the metering ranges from September 1, 2018 to December 31, 2018. The process of data cleaning and mining is explained in the **Appendix 6**. The data used in the analysis is structured, in column order, as *date*, *ID*, *consumption*, *temperature*, *dew point*, *humidity*, *wind speed*, *wind direction*, *visibility*, *pressure*, *weather condition*, *fog* and *rain*. The summary statistics of each of the data set attributes can be found in Table

2. Furthermore, a pairs plot was also produced to get an overview of the whole data set, see Appendix 6 - Supplementary Figure 12.

Variable	Status	Type
Date	Independent	Categorical nominal
ID	Independent	Nominal
Consumption	Dependent	Continuous ratio
Temperature	Independent	Continuous ratio
Dew point	Independent	Continuous ratio
Humidity	Independent	Continuous ratio
Wind speed	Independent	Continuous ratio
Wind direction	Independent	Nominal
Visibility	Independent	Continuous ratio
Pressure	Independent	Continuous ratio
Weather condition	Independent	Nominal
Fog	Independent	Discrete binary (0 or 1)
Rain	Independent	Discrete binary (0 or 1)

Table 2: Variable names, status and types.

consumption	temp	dew_pt	hum	wind_spd	dir
Min. :0.00000	Min. :-1.900	Min. :-3.650	Min. :49.00	Min. : 3.84	Length:9794
1st Qu.:0.07465	1st Qu.: 4.556	1st Qu.: 2.333	1st Qu.:72.68	1st Qu.:11.24	Class :character
Median :0.15160	Median : 8.833	Median : 6.600	Median :82.07	Median :15.44	Mode :character
Mean :0.43617	Mean : 8.724	Mean : 6.309	Mean :81.02	Mean :16.37	
3rd Qu.:0.33618	3rd Qu.:12.857	3rd Qu.:10.000	3rd Qu.:89.81	3rd Qu.:20.72	
Max. :8.70266	Max. :18.615	Max. :15.125	Max. :98.61	Max. :42.27	

vis	pressure	cond	fog	rain
Min. : 1.965	Min. : 986.5	Length:9086	Min. :0.00000	Min. :0.00000
1st Qu.:11.706	1st Qu.:1011.2	Class :character	1st Qu.:0.00000	1st Qu.:0.00000
Median :18.878	Median :1017.4	Mode :character	Median :0.00000	Median :0.05263
Mean :20.541	Mean :1016.5		Mean :0.05202	Mean :0.12689
3rd Qu.:29.815	3rd Qu.:1022.5		3rd Qu.:0.00000	3rd Qu.:0.20000
Max. :50.000	Max. :1040.4		Max. :0.65000	Max. :0.61905

Table 3: Summary statistics of the numerical variables (raw data).

After a thorough review of the data set obtained after the cleansing, it is found no missing or atypical values. However, the data set was changed during the analysis. First of all, the date attribute was split to week number and and type of day (i.e. workday (Monday to Friday) or weekend (Saturday and Sunday), added to the data frame as new columns: *week* and *day-type*. After this, the following attributes were factorized: *ID*, *wind direction*, *condition*, *date* and *seasonality*.

After this, the data set was inspected in order to find non-informative attributes and then to remove them from the data frame for the analysis. This also helped to obtain a reduced data set that would make easier to create models with. The variables *visibility* and *wind direction* were removed because the information that they give it is not informative to our heat loss analysis, but it could be used for future work or other type of analysis. The variable *condition* was removed because it describes the weather condition when these conditions were already included in other numerical variables which are better to perform the analysis. The variables *rain* and *fog* were removed due to errors in the cleaning, because they were entered to the data set as a binary attribute (0 = absence of rain/fog and 1 = presence of rain/fog), but then they were treated like a continuous ratio, therefore they were excluded from the analysis.

To fully understand the data set, the HTK building data share file was inspected. This file contains information about the building, such as ID, type of building and address. To better understand and interpret the results, the attribute *type* was used. There are 25 defined types of buildings (Appendix 6, Table 10) plus type "000" which was added since some of the buildings ID were lacking this information.

Regarding outliers, while visualizing the data, there were observations that were clearly away from the rest of the data. For example, the two highest outliers were found in a building (ID: 78185925) corresponding to type 032, Sports and Swimming Centres (see Fig. 2 and 3). These two observations were removed from the data set. In the figures mentioned above, a representation of consumption against difference of temperature between indoors and outdoors can be seen. This difference of temperature is calculated with the formula ($Difference = 21 - T_{OUT}$). This makes sense since the closer to December (winter) the outdoor temperature decreases and the difference between the outdoor and indoor temperature increases.

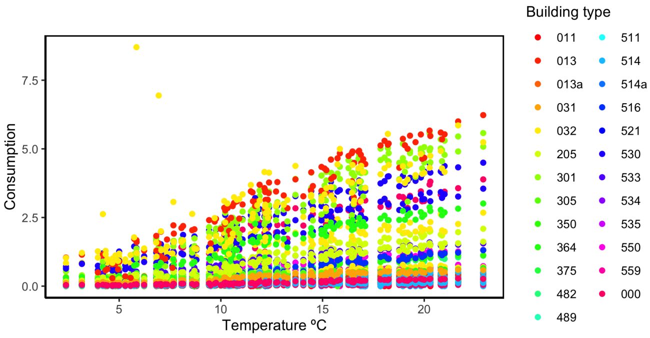

Figure 2: Plot of consumption against difference of temperature indoor and outdoor.

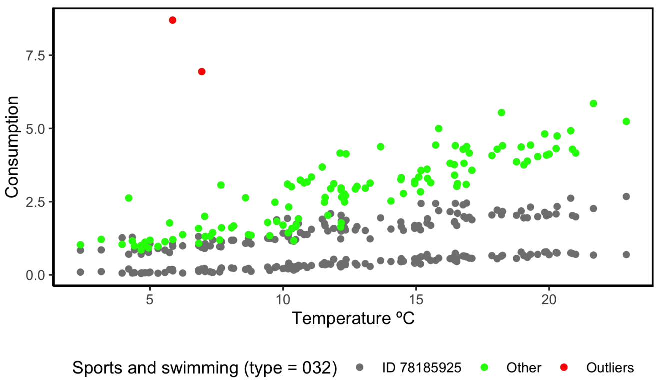

Figure 3: Plot of consumption against difference of temperature indoor and outdoor only Sports and Swimming Centres building type.

To follow with the data visualization, the consumption in the analyzed period regarding the type of building was plotted (Figure 4). There's a clear difference of consumption between building type. The buildings with the highest consumption are swimming pools, schools, kindergartens, and nursing centers.

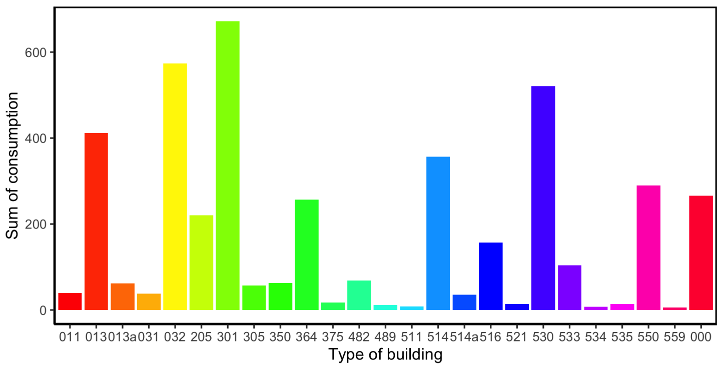

Figure 4: Plot of sum of consumption against the type of building.

The consumption along the analyzed period was also visualized, see Figure 5. It can be seen in the plot that the consumption increases along the days, this is because the date of measuring start is 1^{st} September and it continues until the end of December, where the consumption increases depending on the heat consumption because of the difference of temperature between indoor and outdoor. This plot is interesting because it shows peaks and lows.

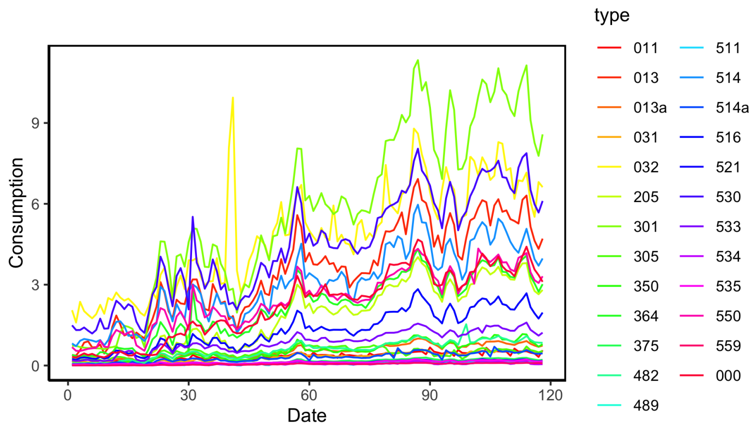

Figure 5: Plot of consumption changes during the 120 day measurements for each type of building.

3 Statistical Analysis

To begin selecting the statistical model, and after the data cleaning and variables removal from the previous section, the data frame used had the following attributes, in column order: *ID*, *consumption*, *temperature*, *wind speed*, *pressure*, *week*, and *daytype*.

The first step was to test with a simple model of consumption against the ID and $I(21 - temp)$, to calculate the insulation or U-value, see Equations 1 and 2.

$$lm_u = lm(consumption \sim ID * I(21 - temp), df) \tag{2}$$

This simple linear model carried out an ANCOVA (since the independent variables are continuous and categorical) and the model diagnostics were tested under four main assumptions:

1. Normality of residuals
2. Variance homogeneity
3. Variance should be independent of the location
4. Linear relationship between independent variable and dependent variable

After that, it was decided to create a model ($\mathbf{lm}_1$) with another variable to represent the date, the model is represented by the following equation:

$$lm_1 = lm(consumption \sim (ID + week) * I(21 - temp), df) \tag{3}$$

The diagnostics of the model were tested to give us the information to decide making a transformation on any variable or investigate interactions between variables. This will be explained in detail in Section 4, but the normalization carried out was the following:

$$adjconsumption_{ID} = consumption_{ID}/mean(consumption_{ID}) \tag{4}$$

Once the normalization was done, the next model $\mathbf{lm}_2$ included the normalized variable *adjconsumption* and all the interactions of all the remaining variables, this was done by using the scope function (see R documentation) and a k factor of $log(n)$,as can be seen in the following equation:

$$\begin{aligned} lm_2 = &lm(adjconsumption \sim ID + hum + windspd + pressure+ \\ &week + I(21 - temp) + ID : I(21 - temp) + week : I(21 - temp)+ \\ &windspd : week + hum : windspd, df) \end{aligned} \tag{5}$$

After analyzing the diagnostics of model $\mathbf{lm}_2$, it was decided to remove some observations and clean outliers and odd buildings, as will be further explained in section 4. Then the model $\mathbf{lm}_3$ was created with the following equation, this model included the variable *daytype*, which was explained in the previous section:

$$\begin{aligned} lm_3 =& lm(adjconsumption \sim ID + hum + windspd + pressure + week + daytype + \\ & I(21 - temp) + ID : I(21 - temp) + ID : daytype + week : daytype + \\ & week : I(21 - temp) + hum : pressure + windspd : I(21 - temp), df) \end{aligned} \tag{6}$$

Subsequent to the diagnostics of model $\mathbf{lm}_3$, these showed still some odd behaviour of buildings that was affecting the model, as will be explained in detail on Section 4. Therefore another iteration of model cleansing was effectuated and the final model $\mathbf{lm}_4$ was created using the following equation:

$$\begin{aligned} lm_4 =& lm(adjconsumption \sim ID + hum + windspd + pressure + daytype + I(21 - temp) + \\ & ID : I(21 - temp) + ID : daytype + daytype : I(21 - temp) + hum : windspd + \\ & windspd : pressure + hum : I(21 - temp) + windspd : I(21 - temp) + \\ & hum : daytype + hum : windspd : I(21 - temp) + hum : daytype : I(21 - temp, df) \end{aligned} \tag{7}$$

With the results of the final model, the insulation parameters were finally calculated by obtaining the coefficients of the final model $\mathbf{lm}_4$ and multiplying with the matrix A, to obtain the estimated parameters (or slopes) of each building dependant on equation 1 and 8:

$$V[A\hat{\theta}] = AV[\hat{\theta}]A^T = A\Sigma_\theta A^T \tag{8}$$

Here $\hat{\theta}$ represents the column-vector of the estimated parameters of the final model, and the variance of those estimated parameters $V[\hat{\theta}]$ are noted as the matrix Σ_θ. This was done because the correlation of each parameter affects each building, so it wasn't possible to just square the standard errors of each building and add them. Furthermore, the confidence intervals were calculated using the following equation:

$$C.I. = \theta \pm t_{0.975,df} * s.e. \tag{9}$$

where $C.I$ indicates the confidence interval, $t_0.975$ indicates the critical t-value at quantile 0.975, and $s.e.$ indicates the standard error.

The estimated parameters $\hat{\theta}$ multiplied by A would give us the final insulation of each building. Then we could identify which buildings were identified as potential suitors of the retrofitting to decrease its energy consumption given heat loss.

4 Results

After running the first model $\mathbf{lm}_u$ it was decided to remove the variable *dew point* from the data frame because it showed a high correlation (corr = 0.95) with the attribute temperature and this could affect our models (see Fig. 6). Pearson's correlation test result can be found in Section 6). Also, the results of the ANCOVA shows that both variables *ID* and *I(21-temp)* were significant (and also the interaction). The diagnostics of the model did not look right, as can be seen in the diagnostics graph on **Section 6**.

	Sum Sq	df	F value	Pr(>F)
ID	4712.23	82	3233.00	0.0000
I(21 - temp)	405.67	1	22822.45	0.0000
ID:I(21 - temp)	1118.30	82	767.25	0.0000
Residuals	171.10	9626		

Table 4: ANCOVA table results for lm_u.

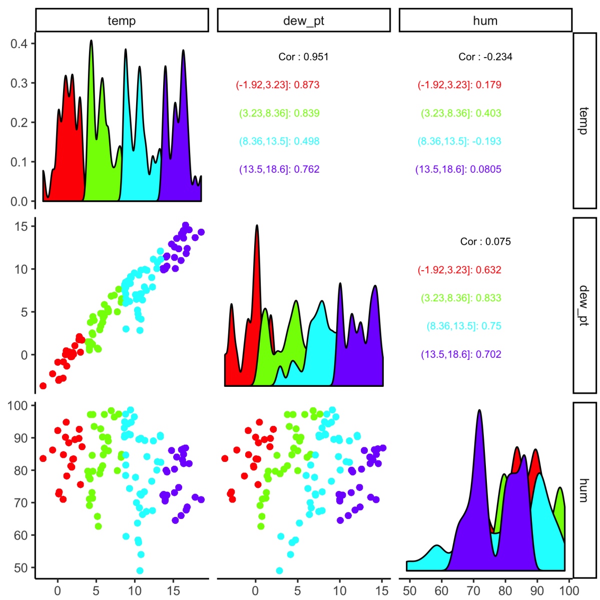

Figure 6: Pairs plot to verify the correlation between the *temperature* and the *dew point*. The plot also included the attribute *humidity*, because the *dew point* is the temperature when the air is completely saturated by the relative humidity.

Then, with model $\mathbf{lm}_1$ there was an addition of the variable *week*, which is a factor variable indicating the number of the week according to the date. The results showed still an statistical significance of all the variables involved and its interactions as can be seen in Table 5. Also, the diagnostics of the model showed an increasing behaviour of the residuals in scale-location plot, as can be seen in Figure 15. This behaviour was further investigated by looking at the variance and mean of each of the buildings.

	Sum Sq	Df	F value	Pr(>F)
ID	4712.52	82	3702.84	0.0000
week	20.22	17	76.63	0.0000
I(21 - temp)	20.15	1	1298.23	0.0000
ID:I(21 - temp)	1117.93	82	878.41	0.0000
week:I(21 - temp)	2.01	17	7.62	0.0000
Residuals	148.87	9592		

Table 5: ANCOVA table results for lm_1

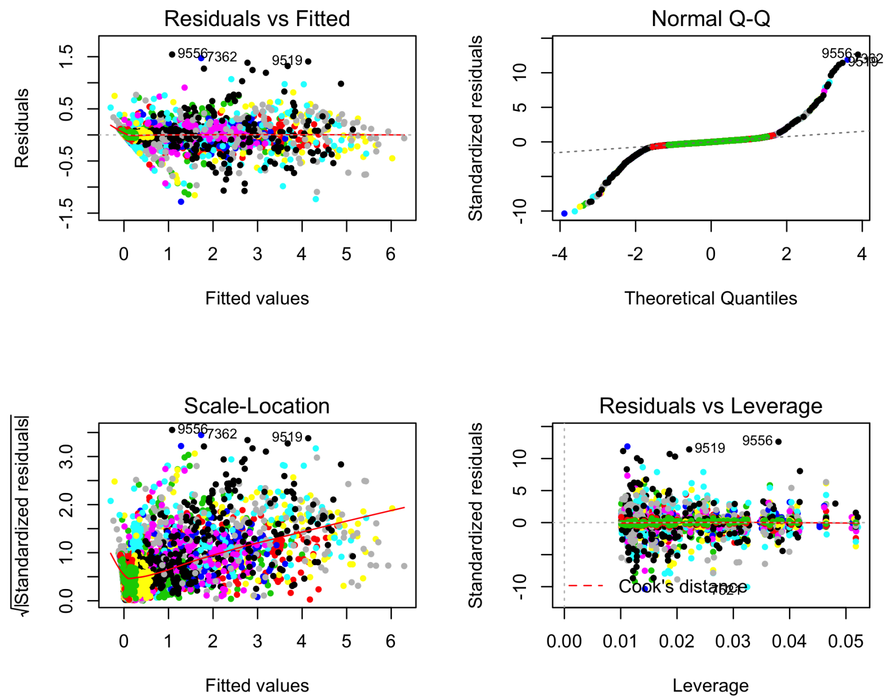

Figure 7: $\mathbf{lm}_1$ model diagnostics

Figure 8 shows the variance against the mean of each building in the left-hand side graph, while the graph on the right shows the consumption against the ID of each building. These two plots shows the relationship between the 'size' of the building and the consumption, where the bigger buildings who consume more (and thus have a bigger mean consumption) had a bigger variance as compared to the small-sized buildings. The data was normalized to remove this relation between variables and the consumption variable was normalized following Equation 4, with this normalization the consumption does not depend anymore on the size of the building.

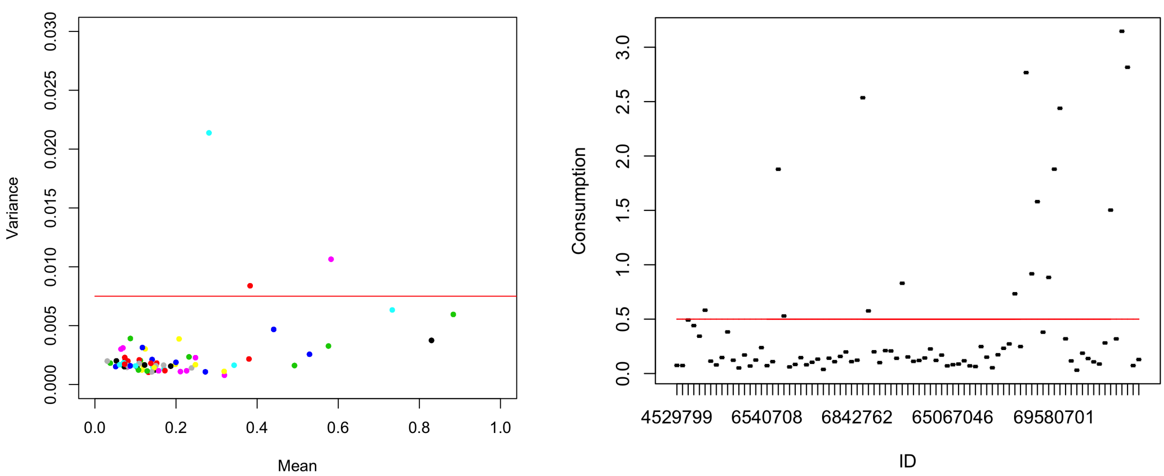

Figure 8: In the left, variance plotted against consumption mean per building. The red line is the threshold for variance set at 0.0075. In the right, the consumption plotted against the IDs, the red line points out the buildings with a mean consumption higher than 0.5.

The model $\mathbf{lm}_2$ shows statistical significance on all the variables and ran in time, thanks to the scope function, as can be seen in Table 6.

	Sum Sq	Df	F value	Pr(>F)
ID	0.01	82	0.00	1.0000
hum	0.71	1	18.78	0.0000
wind_spd	2.07	1	54.98	0.0000
pressure	5.60	1	149.12	0.0000
week	88.16	17	138.01	0.0000
temp	39.74	1	1057.59	0.0000
ID:temp	100.27	82	32.54	0.0000
week:temp	12.58	16	20.93	0.0000
wind_spd:week	11.19	16	18.62	0.0000
hum:wind_spd	0.65	1	17.25	0.0000
Residuals	359.68	9572		

Table 6: ANCOVA results for lm_2

The diagnostics can be seen in Section 6, and they showed still an odd behaviour in the QQ-Plot, where the residuals seemed heavy tailed. With further inspection of the variance and mean of the buildings, there was a hint that the heating pattern of several buildings changed suddenly from September 24, 2018 onwards, as can be seen in Figure 9.

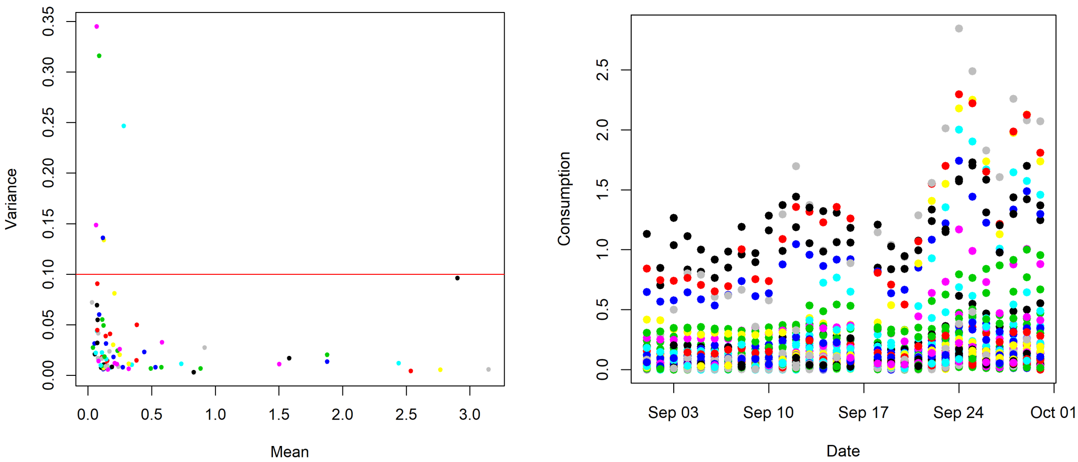

Figure 9: On the left, variance against mean per building, to filter odd-behaving buildings. On the right, the plotted odd-buildings by consumption against date, there is a different heating pattern from September 24, 2018 onwards.

Therefore it was decided to remove all observations before September 24, 2018 for all buildings to exclude the different heating pattern. Also, as seen in Figure 10, there were a couple of odd-behaving buildings that was decided to remove (ID 651118812 and ID 69999051) to prevent these to affect the results and the linearity required by the model to be fitted correctly. The observations of consumption equal to zero and outliers were also cleaned for these buildings with the highest variance.

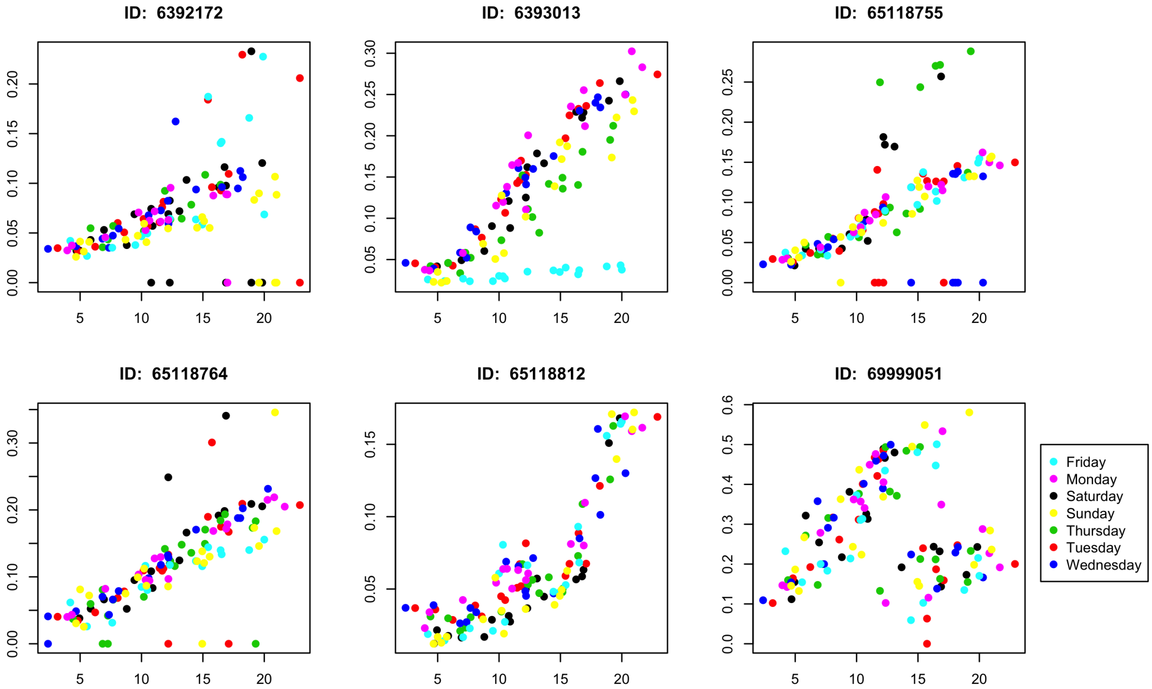

Figure 10: Odd behaving buildings with outliers and odd heating pattern, plot of consumption against difference in temperature. ID 651118812 and ID 69999051 show an odd-behaviour in the heating pattern and ID 6393031 shows a dependancy on the day of the week.

The statistical analysis of the model seemed to be correct (Table 7). However, as seen in model diagnostics on Section 6 the normal distribution still was heavy-tailed, that imply that some of the data is not correctly distributed. As seen in the left-lower panel, the distribution of the points is not right, because it should be spread as a *sky full of stars*, and the values are piled up in the lower part of the plot.

	Sum Sq	Df	F value	Pr(>F)
ID	4.68	80	2.48	0.0000
wind_spd	1.26	1	53.37	0.0000
pressure	0.48	1	20.39	0.0000
week	77.15	17	192.58	0.0000
daytype	9.37	1	397.58	0.0000
temp	45.47	1	1929.61	0.0000
ID:temp	74.45	80	39.49	0.0000
ID:daytype	32.18	80	17.07	0.0000
week:daytype	7.48	15	21.17	0.0000
week:temp	6.68	17	16.68	0.0000
Residuals	216.60	9191		

Table 7: ANCOVA results for lm_3

It was decided to keep removing outliers of buildings with high variance, compared to the other buildings, and remove odd-behaving buildings (7): 4529799, 4529800, 6393013, 69652588, 78185925, and 65118848. Thus, the model ended with $Buildings = 75$ buildings, instead of $Buildings = 83$, once these buildings from this model and the previous model $\mathbf{lm}_2$ were removed.

Lastly, on model $\mathbf{lm}_4$, it was decided to remove the variable *week*, since the observations of odd heating patterns from September were already removed. This model shows statistical significance of all single variables and interactions between them with an upper bound of interactions between three variables, as can be seen in Table 8. The significant interactions in the summary means there's a difference in the slope of the interaction and the dependant variable, e.g. ID and temperature vs. consumption. That means, the slope of each building is different when it is exposed at the same temperature. It can also be seen from the table that there are a few other interactions, e.g. ID and daytype explains the difference of consumption between buildings because of (probably) the type of heating consumption they have - maybe a school is close on weekends and its heat consumption decreases while a residential building increases its heat consumption during the weekends since the tenants are normally at home during weekends.

	Sum Sq	Df	F value	Pr(>F)
ID	6.18	74	4.76	0.0000
hum	5.39	1	307.14	0.0000
wind_spd	10.16	1	579.42	0.0000
pressure	0.31	1	17.80	0.0000
daytype	21.98	1	1253.07	0.0000
I(21 - temp)	930.97	1	53068.74	0.0000
ID:I(21 - temp)	51.96	74	40.02	0.0000
ID:daytype	24.63	74	18.97	0.0000
daytype:I(21 - temp)	3.06	1	174.68	0.0000
hum:wind_spd	2.23	1	126.83	0.0000
wind_spd:pressure	0.50	1	28.31	0.0000
hum:I(21 - temp)	0.56	1	31.99	0.0000
wind_spd:I(21 - temp)	0.31	1	17.76	0.0000
hum:daytype	0.36	1	20.75	0.0000
hum:wind_spd:I(21 - temp)	0.94	1	53.82	0.0000
hum:daytype:I(21 - temp)	2.29	1	130.76	0.0000
Residuals	119.50	6812		

Table 8: ANCOVA results for lm_4

The residuals of the final model **lm4** shows still some lightly-tailed behaviour in the QQ-Plot of Figure 11. However, the residual distribution and variance of each building in the scale-location and residuals plot looks right. It was decided to stop cleaning the outliers and odd buildings since it was starting to get complicated and not very clear which observations were outliers and which were dependant on an 'odd day', i.e. a very cold day.

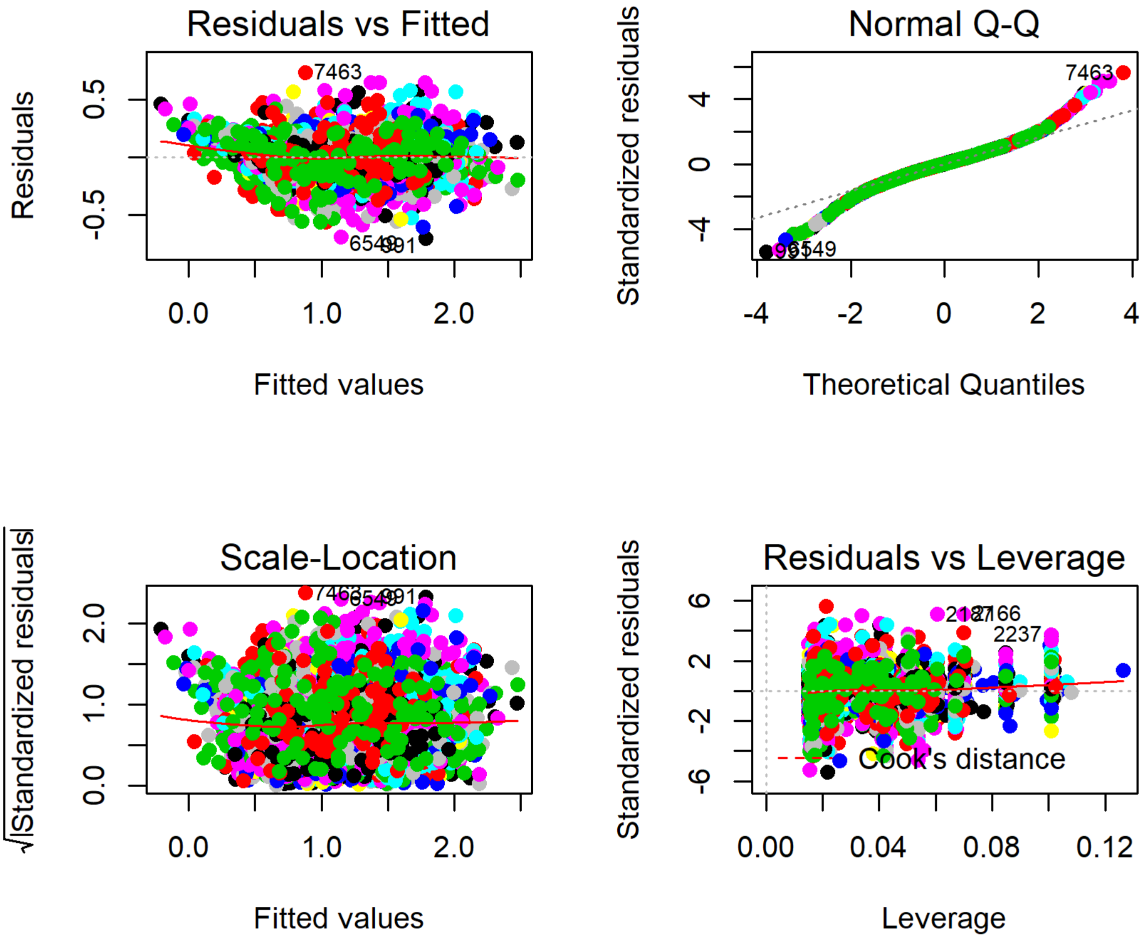

Figure 11: dependency

Lastly, Table 9 and 12 (on Section 6) shows the ID of the identified buildings suited for the retrofitting, i.e. to renovate and upgrade the insulation of the buildings. It can be shown from the table that the insulation is very low at 0.01 compared to the mean of other buildings at 0.06.

	ID	Insulation	Std. Error	CI_Up	CI_Down
9	5325295	0.01	0.01	0.02	-0.01
59	69585544	0.01	0.01	0.03	-0.01
24	65118764	0.01	0.01	0.03	-0.01
17	65014274	0.03	0.01	0.04	0.01
70	7072337	0.03	0.01	0.05	0.01

Table 9: Insulation Parameters of the lowest insulated buildings, these are the potential buildings identified for the retrofitting.

5 Conclusion

Given the statistical analysis performed, it is found that the buildings with the lowest insulation parameter are the ones with the poorest performance and consequently the ones with the highest consumption. In other words, the final general regression model and the ANCOVA analysis have been able to identify the buildings that are less efficient, so they can be prioritized for the retrofitting improvements.

As a future work, the analysis could take into consideration that the mean comfort temperature inside of the building ($T_{indoors}$) varies depending of the building type and the activity that is performed there. For instance, schools and kindergartens have a constant temperature indoors around 21ºC, but swimming pools have a higher comfort temperature indoors (between 26°C and 28°C). The lineal regression model would differ from the type of building depending on this constant. Furthermore, it could be interesting to analyse the building consumption in a whole year, so as to compare between seasons and have a better overview of the variations in the energy consumption.

	Sum Sq	Df	F value	Pr(>F)
ID	6.66	80	4.41	0.0000
hum	0.01	1	0.66	0.4178
wind_spd	1.16	1	61.75	0.0000
pressure	0.52	1	27.59	0.0000
week	39.99	13	163.10	0.0000
daytype	8.96	1	475.25	0.0000
I(21 - temp)	35.92	1	1904.51	0.0000
ID:I(21 - temp)	55.08	80	36.51	0.0000
ID:daytype	31.66	80	20.99	0.0000
week:daytype	5.73	12	25.30	0.0000
week:I(21 - temp)	5.27	13	21.50	0.0000
hum:pressure	0.45	1	23.68	0.0000
wind_spd:I(21 - temp)	0.18	1	9.55	0.0020
Residuals	138.54	7346		

	ID	Beta	Std. Error	CI_Up	CI_Down
70	7072337	0.18	0.01	0.20	0.16
17	65014274	0.18	0.01	0.20	0.16
24	65118764	0.20	0.01	0.22	0.18
59	69585544	0.20	0.01	0.22	0.18
9	5325295	0.20	0.01	0.22	0.18

6 Appendix

6.A Supplementary Figures

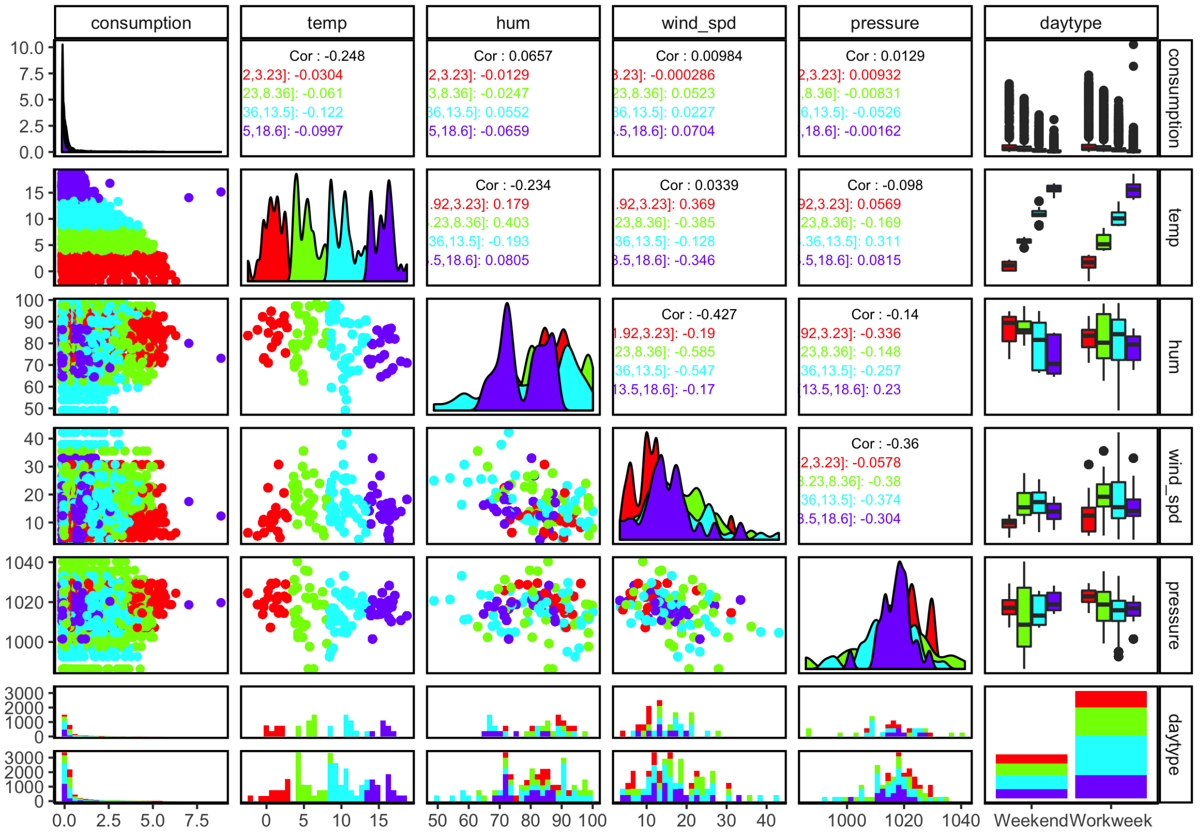

Figure 12: General pairs plot for the numerical attributes plus the day type of the week (workday = Monday to Friday, weekend = Saturday and Sunday).

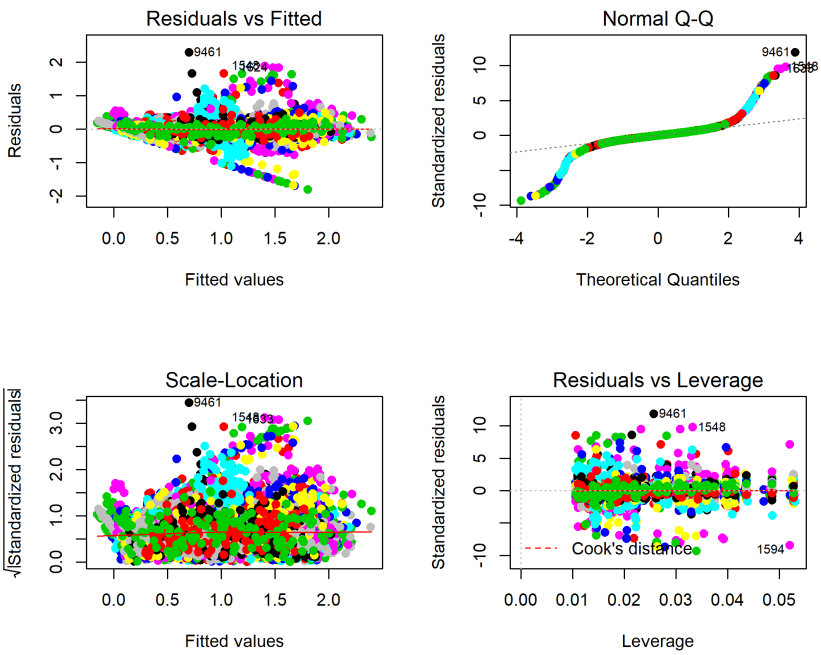

Figure 13: **lm**$_2$ model diagnostics

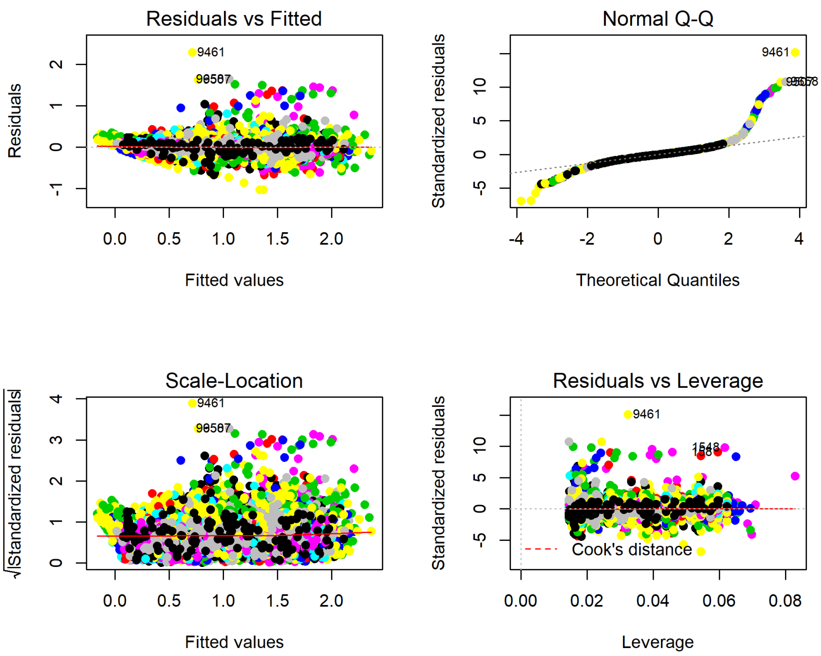

Figure 14: **lm**$_3$ model diagnostics

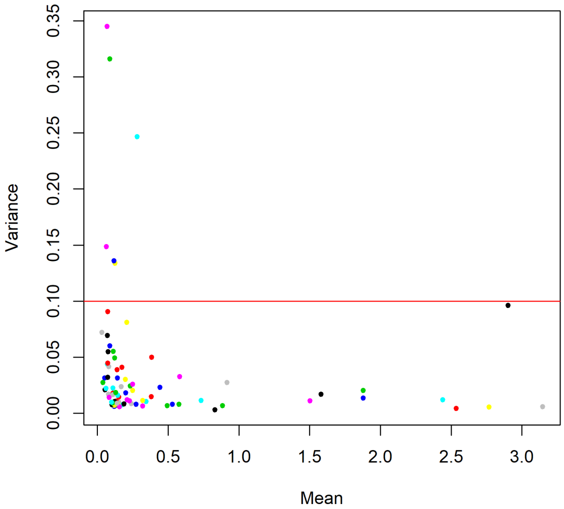

Figure 15: Plot variance vs mean for $\mathbf{lm}_2$ model

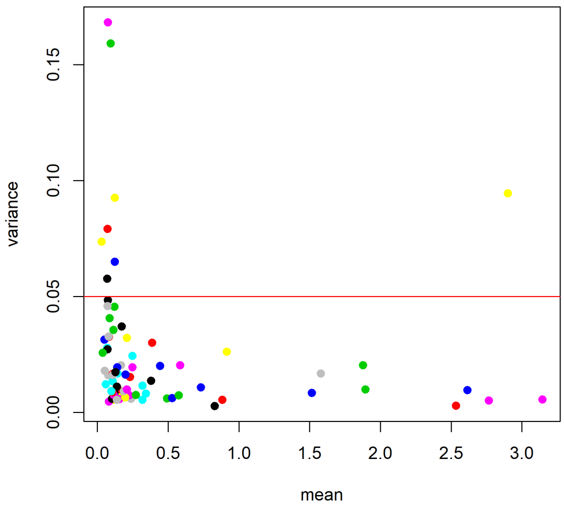

Figure 16: Plot variance vs mean for $\mathbf{lm}_3$ model

6.B Supplementary Tables

```
Pearson's product-moment correlation

data:  df$temp and df$dew_pt
t = 304.53, df = 9790, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
0.9491335 0.9529173
sample estimates:
      cor
0.9510611
```

Correlation test 1: Pearson's correlation test between the temperature and the dew point.

Type code	Type name
000	Non-defined type
011	Beboelsesejendomme
013	Andre ejendomme
013a	Tomme ejendomme
031	Stadion og idrætsanlæg
032	Idræts- og svømmehaller
205	Driftsbygninger
301	Skoler
305	Skolefritidsordninger
350	Biblioteker
364	Andre kulturelle opgaver
375	Fritidsaktiviteter
482	Træningscentre
489	Sundhedstjeneste
511	Dagpleje
514	Integrerede daginstitutioner
514a	Integrerede daginstitution PRIVATE
516	Fritids- og ungdomsklubber
521	Foreb. foranst for børn og unge
525	Særlige dagtilbud og særlige klubber
530	Plejecentre
533	Pleje og omsorg m.v. af ældre og handicappede
534	Institutioner for ældre
535	Hjælpemiddeldepoter
550	Længerevarende botilbud
559	Aktivitets- og samværstilbud

Table 10: Building codes and building names.

6.C Model Supplementary information

6.C.1 Linear model for the insulation (lmu)

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.0079	0.0326	0.24	0.8077
ID4529800	-0.0103	0.0461	-0.22	0.8237
ID4839509	0.2189	0.0461	4.74	0.0000
ID4866195	0.0392	0.0460	0.85	0.3947
ID4887707	-0.0652	0.0461	-1.41	0.1578
ID4962433	-0.1541	0.0461	-3.34	0.0008
ID5037175	-0.0099	0.0461	-0.21	0.8308
ID5093913	-0.0430	0.0461	-0.93	0.3516
ID5093998	-0.0341	0.0461	-0.74	0.4599
ID5140250	0.0403	0.0461	0.88	0.3814
ID5325295	-0.0759	0.0461	-1.65	0.0999
ID6392057	-0.0281	0.0461	-0.61	0.5425
ID6392146	-0.0466	0.0461	-1.01	0.3121
ID6392172	0.0097	0.0460	0.21	0.8327
ID6393013	-0.0343	0.0461	-0.74	0.4569
ID6393014	-0.0398	0.0460	-0.86	0.3877
ID6540708	-0.0249	0.0461	-0.54	0.5890
ID6567326	0.0345	0.0461	0.75	0.4551
ID6618578	-0.5986	0.0461	-12.97	0.0000
ID6618580	-0.2750	0.0461	-5.96	0.0000
ID6627217	-0.0146	0.0461	-0.32	0.7513
ID6627258	0.0033	0.0461	0.07	0.9422
ID6627261	-0.0125	0.0461	-0.27	0.7866
ID6627320	-0.0263	0.0461	-0.57	0.5689
ID6681763	-0.0345	0.0462	-0.75	0.4554
ID6681892	-0.0431	0.0461	-0.93	0.3504
ID6681894	-0.0331	0.0461	-0.72	0.4725
ID6790785	0.0679	0.0461	1.47	0.1409
ID6790798	-0.0115	0.0461	-0.25	0.8034
ID6842413	-0.0205	0.0461	-0.44	0.6574
ID6842421	-0.0517	0.0461	-1.12	0.2625
ID6842603	0.0085	0.0461	0.18	0.8534
ID6842762	-0.0256	0.0461	-0.56	0.5784
ID6921678	-0.0961	0.0461	-2.08	0.0374
ID6940321	-0.2418	0.0461	-5.24	0.0000
ID6940322	-0.0032	0.0461	-0.07	0.9446
ID7072161	-0.0234	0.0461	-0.51	0.6123

ID7072231	-0.0954	0.0461	-2.07	0.0388
ID7072241	-0.0812	0.0461	-1.76	0.0784
ID7072337	-0.0536	0.0461	-1.16	0.2458
ID7183151	-0.1967	0.0461	-4.26	0.0000
ID65005112	-0.0133	0.0461	-0.29	0.7728
ID65012411	-0.0384	0.0460	-0.84	0.4033
ID65014229	-0.0170	0.0461	-0.37	0.7126
ID65014274	-0.0137	0.0461	-0.30	0.7665
ID65052581	-0.0795	0.0461	-1.72	0.0848
ID65063195	-0.0374	0.0461	-0.81	0.4180
ID65063211	-0.0349	0.0461	-0.76	0.4490
ID65063303	-0.0140	0.0461	-0.30	0.7620
ID65067046	-0.0079	0.0461	-0.17	0.8645
ID65118755	-0.0119	0.0460	-0.26	0.7955
ID65118764	-0.0174	0.0459	-0.38	0.7045
ID65118805	-0.0256	0.0461	-0.56	0.5784
ID65118812	-0.0372	0.0461	-0.81	0.4199
ID65118826	-0.0168	0.0461	-0.36	0.7155
ID65118829	-0.0448	0.0461	-0.97	0.3312
ID65118840	-0.0064	0.0461	-0.14	0.8901
ID65118848	-0.0449	0.0461	-0.97	0.3304
ID69001263	-0.0310	0.0461	-0.67	0.5018
ID69001269	-0.0105	0.0461	-0.23	0.8202
ID69089222	-0.0509	0.0461	-1.10	0.2699
ID69250492	-0.0529	0.0461	-1.15	0.2514
ID69429582	-1.0233	0.0461	-22.18	0.0000
ID69469107	-0.6089	0.0461	-13.20	0.0000
ID69478883	0.5062	0.0461	10.97	0.0000
ID69518080	-0.1398	0.0461	-3.03	0.0025
ID69518092	-0.0656	0.0461	-1.42	0.1553
ID69580701	-0.1498	0.0460	-3.25	0.0011
ID69585544	-1.0541	0.0461	-22.85	0.0000
ID69585545	-0.0901	0.0461	-1.95	0.0508
ID69611360	-0.0155	0.0461	-0.34	0.7362
ID69652588	-0.0078	0.0461	-0.17	0.8664
ID69652603	-0.0157	0.0461	-0.34	0.7336
ID69688095	-0.0075	0.0461	-0.16	0.8716
ID69749518	-0.0452	0.0461	-0.98	0.3268
ID69861509	-0.0391	0.0461	-0.85	0.3962
ID69999051	0.2514	0.0462	5.44	0.0000
ID69999094	-0.4423	0.0460	-9.61	0.0000
ID78082613	-0.0526	0.0461	-1.14	0.2543
ID78138095	-0.5149	0.0461	-11.16	0.0000
ID78185925	0.0550	0.0466	1.18	0.2381
ID78443775	-0.0075	0.0461	-0.16	0.8711
ID78673711	-0.0532	0.0461	-1.15	0.2485
I(21 - temp)	0.0054	0.0025	2.21	0.0270
ID4529800:I(21 - temp)	0.0007	0.0035	0.21	0.8322
ID4839509:I(21 - temp)	0.0162	0.0035	4.64	0.0000
ID4866195:I(21 - temp)	0.0267	0.0035	7.67	0.0000
ID4887707:I(21 - temp)	0.0272	0.0035	7.80	0.0000
ID4962433:I(21 - temp)	0.0540	0.0035	15.50	0.0000
ID5037175:I(21 - temp)	0.0040	0.0035	1.15	0.2514
ID5093913:I(21 - temp)	0.0039	0.0035	1.11	0.2685
ID5093998:I(21 - temp)	0.0087	0.0035	2.49	0.0129
ID5140250:I(21 - temp)	0.0219	0.0035	6.28	0.0000
ID5325295:I(21 - temp)	0.0100	0.0035	2.88	0.0039
ID6392057:I(21 - temp)	0.0004	0.0035	0.10	0.9170
ID6392146:I(21 - temp)	0.0116	0.0035	3.33	0.0009
ID6392172:I(21 - temp)	-0.0013	0.0035	-0.37	0.7150
ID6393013:I(21 - temp)	0.0068	0.0035	1.95	0.0511
ID6393014:I(21 - temp)	0.0166	0.0035	4.77	0.0000
ID6540708:I(21 - temp)	0.0018	0.0035	0.53	0.5977
ID6567326:I(21 - temp)	0.0000	0.0035	0.00	0.9981
ID6618578:I(21 - temp)	0.1956	0.0035	56.18	0.0000
ID6618580:I(21 - temp)	0.0594	0.0035	17.06	0.0000
ID6627217:I(21 - temp)	0.0001	0.0035	0.02	0.9850
ID6627258:I(21 - temp)	0.0004	0.0035	0.13	0.8991
ID6627261:I(21 - temp)	0.0068	0.0035	1.96	0.0503
ID6627320:I(21 - temp)	0.0026	0.0035	0.75	0.4548
ID6681763:I(21 - temp)	0.0051	0.0035	1.47	0.1406
ID6681892:I(21 - temp)	0.0082	0.0035	2.35	0.0189
ID6681894:I(21 - temp)	-0.0003	0.0035	-0.09	0.9283

ID6790785:I(21 - temp)	-0.0001	0.0035	-0.04	0.9678
ID6790798:I(21 - temp)	0.0037	0.0035	1.07	0.2860
ID6842413:I(21 - temp)	0.0084	0.0035	2.40	0.0163
ID6842421:I(21 - temp)	0.0143	0.0035	4.10	0.0000
ID6842603:I(21 - temp)	0.0021	0.0035	0.61	0.5411
ID6842762:I(21 - temp)	0.0060	0.0035	1.72	0.0858
ID6921678:I(21 - temp)	0.2082	0.0035	59.81	0.0000
ID6940321:I(21 - temp)	0.0605	0.0035	17.38	0.0000
ID6940322:I(21 - temp)	0.0104	0.0035	3.00	0.0027
ID7072161:I(21 - temp)	0.0040	0.0035	1.14	0.2534
ID7072231:I(21 - temp)	0.0188	0.0035	5.41	0.0000
ID7072241:I(21 - temp)	0.0174	0.0035	5.00	0.0000
ID7072337:I(21 - temp)	0.0097	0.0035	2.78	0.0054
ID7183151:I(21 - temp)	0.0775	0.0035	22.26	0.0000
ID65005112:I(21 - temp)	0.0074	0.0035	2.13	0.0335
ID65012411:I(21 - temp)	0.0062	0.0035	1.78	0.0751
ID65014229:I(21 - temp)	0.0051	0.0035	1.46	0.1441
ID65014274:I(21 - temp)	0.0067	0.0035	1.91	0.0557
ID65052581:I(21 - temp)	0.0188	0.0035	5.40	0.0000
ID65063195:I(21 - temp)	0.0068	0.0035	1.95	0.0515
ID65063211:I(21 - temp)	0.0106	0.0035	3.03	0.0024
ID65063303:I(21 - temp)	0.0008	0.0035	0.23	0.8191
ID65067046:I(21 - temp)	0.0012	0.0035	0.34	0.7347
ID65118755:I(21 - temp)	0.0020	0.0035	0.58	0.5637
ID65118764:I(21 - temp)	0.0049	0.0035	1.41	0.1582
ID65118805:I(21 - temp)	0.0018	0.0035	0.50	0.6136
ID65118812:I(21 - temp)	0.0022	0.0035	0.62	0.5324
ID65118826:I(21 - temp)	0.0155	0.0035	4.44	0.0000
ID65118829:I(21 - temp)	0.0098	0.0035	2.82	0.0049
ID65118840:I(21 - temp)	-0.0013	0.0035	-0.36	0.7194
ID65118848:I(21 - temp)	0.0116	0.0035	3.34	0.0008
ID69001263:I(21 - temp)	0.0153	0.0035	4.40	0.0000
ID69001269:I(21 - temp)	0.0169	0.0035	4.86	0.0000
ID69089222:I(21 - temp)	0.0578	0.0035	16.59	0.0000
ID69250492:I(21 - temp)	0.0184	0.0035	5.29	0.0000
ID69429582:I(21 - temp)	0.3026	0.0035	86.91	0.0000
ID69469107:I(21 - temp)	0.1182	0.0035	33.94	0.0000
ID69478883:I(21 - temp)	0.0814	0.0035	23.38	0.0000
ID69518080:I(21 - temp)	0.0362	0.0035	10.41	0.0000
ID69518092:I(21 - temp)	0.0712	0.0035	20.45	0.0000
ID69580701:I(21 - temp)	0.1595	0.0035	45.87	0.0000
ID69585544:I(21 - temp)	0.2783	0.0035	79.95	0.0000
ID69585545:I(21 - temp)	0.0272	0.0035	7.83	0.0000
ID69611360:I(21 - temp)	0.0046	0.0035	1.33	0.1837
ID69652588:I(21 - temp)	-0.0030	0.0035	-0.85	0.3926
ID69652603:I(21 - temp)	0.0104	0.0035	2.99	0.0028
ID69688095:I(21 - temp)	0.0058	0.0035	1.65	0.0984
ID69749518:I(21 - temp)	0.0063	0.0035	1.82	0.0686
ID69861509:I(21 - temp)	0.0042	0.0035	1.21	0.2272
ID69999051:I(21 - temp)	-0.0037	0.0035	-1.05	0.2942
ID69999094:I(21 - temp)	0.1527	0.0035	43.90	0.0000
ID78082613:I(21 - temp)	0.0241	0.0035	6.94	0.0000
ID78138095:I(21 - temp)	0.2921	0.0035	83.90	0.0000
ID78185925:I(21 - temp)	0.2169	0.0035	61.91	0.0000
ID78443775:I(21 - temp)	0.0005	0.0035	0.14	0.8908
ID78673711:I(21 - temp)	0.0087	0.0035	2.51	0.0122

Table 11: Summary for lm_u

	ID	Beta	Std. Error	CI_Up	CI_Down
1	4839509	0.10	0.01	0.12	0.08
2	4866195	0.09	0.01	0.11	0.07
3	4887707	0.05	0.01	0.07	0.03
4	4962433	0.06	0.01	0.07	0.04
5	5037175	0.07	0.01	0.09	0.05
6	5093913	0.03	0.01	0.05	0.01
7	5093998	0.06	0.01	0.08	0.04
8	5140250	0.09	0.01	0.10	0.07
9	5325295	0.01	0.01	0.02	-0.01
10	6392057	0.04	0.01	0.06	0.02
11	6392146	0.05	0.01	0.07	0.03
12	6392172	0.09	0.01	0.10	0.07

13	6393014	0.06	0.01	0.08	0.04
14	65005112	0.05	0.01	0.06	0.03
15	65012411	0.10	0.01	0.12	0.08
16	65014229	0.05	0.01	0.07	0.03
17	65014274	0.03	0.01	0.04	0.01
18	65052581	0.06	0.01	0.08	0.04
19	65063195	0.08	0.01	0.10	0.06
20	65063211	0.07	0.01	0.09	0.05
21	65063303	0.04	0.01	0.06	0.02
22	65067046	0.05	0.01	0.07	0.03
23	65118755	0.05	0.01	0.07	0.03
24	65118764	0.01	0.01	0.03	-0.01
25	65118805	0.11	0.01	0.13	0.09
26	65118826	0.07	0.01	0.09	0.05
27	65118829	0.07	0.01	0.09	0.05
28	65118840	0.05	0.01	0.07	0.03
29	6540708	0.07	0.01	0.09	0.05
30	6567326	0.06	0.01	0.08	0.04
31	6618578	0.07	0.01	0.09	0.05
32	6618580	0.03	0.01	0.05	0.01
33	6627217	0.08	0.01	0.10	0.06
34	6627258	0.05	0.01	0.07	0.04
35	6627261	0.03	0.01	0.05	0.01
36	6627320	0.04	0.01	0.06	0.02
37	6681763	0.04	0.01	0.06	0.02
38	6681892	0.05	0.01	0.07	0.03
39	6681894	0.07	0.01	0.09	0.05
40	6790785	0.05	0.01	0.07	0.04
41	6790798	0.06	0.01	0.08	0.04
42	6842413	0.07	0.01	0.09	0.05
43	6842421	0.04	0.01	0.06	0.02
44	6842603	0.05	0.01	0.07	0.03
45	6842762	0.06	0.01	0.08	0.04
46	69001263	0.05	0.01	0.07	0.03
47	69001269	0.07	0.01	0.09	0.05
48	69089222	0.06	0.01	0.08	0.04
49	6921678	0.07	0.01	0.08	0.05
50	69250492	0.05	0.01	0.06	0.03
51	6940321	0.08	0.01	0.10	0.06
52	6940322	0.05	0.01	0.07	0.03
53	69429582	0.08	0.01	0.10	0.06
54	69469107	0.06	0.01	0.08	0.04
55	69478883	0.07	0.01	0.09	0.05
56	69518080	0.07	0.01	0.09	0.05
57	69518092	0.07	0.01	0.08	0.05
58	69580701	0.04	0.01	0.06	0.03
59	69585544	0.01	0.01	0.03	-0.01
60	69585545	0.09	0.01	0.11	0.07
61	69611360	0.04	0.01	0.06	0.02
62	69652603	0.07	0.01	0.09	0.05
63	69688095	0.06	0.01	0.08	0.04
64	69749518	0.04	0.01	0.06	0.02
65	69861509	0.05	0.01	0.07	0.03
66	69999094	0.06	0.01	0.08	0.04
67	7072161	0.07	0.01	0.09	0.05
68	7072231	0.07	0.01	0.09	0.05
69	7072241	0.05	0.01	0.07	0.03
70	7072337	0.03	0.01	0.05	0.01
71	7183151	0.05	0.01	0.07	0.03
72	78082613	0.05	0.01	0.07	0.04
73	78138095	0.06	0.01	0.08	0.04
74	78443775	0.08	0.01	0.09	0.06
75	78673711	0.03	0.01	0.05	0.02

Table 12: Insulation parameters from lm_4 with confidence intervals

6.C.2 Linear model 1 (lm_1)

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.0862	0.1149	0.75	0.4530
ID4529800	-0.0103	0.0431	-0.24	0.8116
ID4839509	0.2189	0.0431	5.08	0.0000

ID4866195	0.0389	0.0430	0.90	0.3658
ID4887707	-0.0652	0.0431	-1.51	0.1307
ID4962433	-0.1549	0.0431	-3.59	0.0003
ID5037175	-0.0099	0.0431	-0.23	0.8191
ID5093913	-0.0430	0.0431	-1.00	0.3189
ID5093998	-0.0341	0.0431	-0.79	0.4290
ID5140250	0.0400	0.0430	0.93	0.3532
ID5325295	-0.0759	0.0431	-1.76	0.0783
ID6392057	-0.0281	0.0431	-0.65	0.5145
ID6392146	-0.0466	0.0431	-1.08	0.2794
ID6392172	0.0061	0.0430	0.14	0.8869
ID6393013	-0.0343	0.0431	-0.80	0.4260
ID6393014	-0.0401	0.0430	-0.93	0.3520
ID6540708	-0.0249	0.0431	-0.58	0.5632
ID6567326	0.0345	0.0431	0.80	0.4241
ID6618578	-0.5986	0.0431	-13.88	0.0000
ID6618580	-0.2750	0.0431	-6.38	0.0000
ID6627217	-0.0146	0.0431	-0.34	0.7345
ID6627258	0.0033	0.0431	0.08	0.9382
ID6627261	-0.0125	0.0431	-0.29	0.7720
ID6627320	-0.0263	0.0431	-0.61	0.5421
ID6681763	-0.0337	0.0431	-0.78	0.4343
ID6681892	-0.0431	0.0431	-1.00	0.3176
ID6681894	-0.0331	0.0431	-0.77	0.4420
ID6790785	0.0678	0.0431	1.57	0.1158
ID6790798	-0.0114	0.0431	-0.26	0.7915
ID6842413	-0.0205	0.0431	-0.47	0.6351
ID6842421	-0.0517	0.0431	-1.20	0.2305
ID6842603	0.0085	0.0431	0.20	0.8433
ID6842762	-0.0256	0.0431	-0.59	0.5521
ID6921678	-0.0961	0.0431	-2.23	0.0259
ID6940321	-0.2418	0.0431	-5.61	0.0000
ID6940322	-0.0032	0.0431	-0.07	0.9407
ID7072161	-0.0234	0.0431	-0.54	0.5875
ID7072231	-0.0954	0.0431	-2.21	0.0270
ID7072241	-0.0812	0.0431	-1.88	0.0596
ID7072337	-0.0536	0.0431	-1.24	0.2142
ID7183151	-0.1967	0.0431	-4.56	0.0000
ID65005112	-0.0133	0.0431	-0.31	0.7573
ID65012411	-0.0407	0.0430	-0.95	0.3443
ID65014229	-0.0170	0.0431	-0.39	0.6934
ID65014274	-0.0137	0.0431	-0.32	0.7507
ID65052581	-0.0795	0.0431	-1.84	0.0651
ID65063195	-0.0374	0.0431	-0.87	0.3861
ID65063211	-0.0349	0.0431	-0.81	0.4178
ID65063303	-0.0140	0.0431	-0.32	0.7459
ID65067046	-0.0094	0.0431	-0.22	0.8276
ID65118755	-0.0119	0.0430	-0.28	0.7817
ID65118764	-0.0204	0.0429	-0.48	0.6342
ID65118805	-0.0256	0.0431	-0.59	0.5520
ID65118812	-0.0372	0.0431	-0.86	0.3880
ID65118826	-0.0168	0.0431	-0.39	0.6965
ID65118829	-0.0448	0.0431	-1.04	0.2984
ID65118840	-0.0064	0.0431	-0.15	0.8824
ID65118848	-0.0449	0.0431	-1.04	0.2976
ID69001263	-0.0310	0.0431	-0.72	0.4722
ID69001269	-0.0105	0.0431	-0.24	0.8079
ID69089222	-0.0509	0.0431	-1.18	0.2377
ID69250492	-0.0529	0.0431	-1.23	0.2197
ID69429582	-1.0233	0.0431	-23.73	0.0000
ID69469107	-0.6089	0.0431	-14.12	0.0000
ID69478883	0.5062	0.0431	11.74	0.0000
ID69518080	-0.1398	0.0431	-3.24	0.0012
ID69518092	-0.0656	0.0431	-1.52	0.1283
ID69580701	-0.1501	0.0430	-3.49	0.0005
ID69585544	-1.0541	0.0431	-24.45	0.0000
ID69585545	-0.0901	0.0431	-2.09	0.0366
ID69611360	-0.0155	0.0431	-0.36	0.7184
ID69652588	-0.0078	0.0431	-0.18	0.8571
ID69652603	-0.0157	0.0431	-0.36	0.7157
ID69688095	-0.0075	0.0431	-0.17	0.8627
ID69749518	-0.0452	0.0431	-1.05	0.2940
ID69861509	-0.0391	0.0431	-0.91	0.3639

ID69999051	0.2514	0.0431	5.83	0.0000
ID69999094	-0.4426	0.0430	-10.29	0.0000
ID78082613	-0.0526	0.0431	-1.22	0.2225
ID78138095	-0.5149	0.0431	-11.94	0.0000
ID78185925	0.0584	0.0435	1.34	0.1799
ID78443775	-0.0075	0.0431	-0.17	0.8621
ID78673711	-0.0532	0.0431	-1.24	0.2169
week36	0.0709	0.1135	0.62	0.5325
week37	0.0619	0.1135	0.55	0.5855
week38	0.0537	0.1116	0.48	0.6304
week39	0.0673	0.1141	0.59	0.5554
week40	0.1185	0.1168	1.02	0.3100
week41	0.0660	0.1129	0.58	0.5591
week42	0.1027	0.1138	0.90	0.3668
week43	0.1046	0.1130	0.93	0.3546
week44	0.1224	0.1184	1.03	0.3014
week45	0.2780	0.1238	2.25	0.0247
week46	0.2320	0.1159	2.00	0.0454
week47	0.3180	0.1367	2.33	0.0200
week48	0.0626	0.1160	0.54	0.5891
week49	0.1195	0.1176	1.02	0.3097
week50	0.4192	0.1239	3.38	0.0007
week51	0.3060	0.1245	2.46	0.0140
week52	0.1238	0.1186	1.04	0.2966
I(21 - temp)	-0.0191	0.0217	-0.88	0.3785
ID4529800:I(21 - temp)	0.0007	0.0033	0.23	0.8207
ID4839509:I(21 - temp)	0.0162	0.0033	4.97	0.0000
ID4866195:I(21 - temp)	0.0267	0.0032	8.21	0.0000
ID4887707:I(21 - temp)	0.0272	0.0033	8.35	0.0000
ID4962433:I(21 - temp)	0.0540	0.0033	16.60	0.0000
ID5037175:I(21 - temp)	0.0040	0.0033	1.23	0.2197
ID5093913:I(21 - temp)	0.0039	0.0033	1.18	0.2363
ID5093998:I(21 - temp)	0.0087	0.0033	2.66	0.0078
ID5140250:I(21 - temp)	0.0218	0.0033	6.72	0.0000
ID5325295:I(21 - temp)	0.0100	0.0033	3.09	0.0020
ID6392057:I(21 - temp)	0.0004	0.0033	0.11	0.9112
ID6392146:I(21 - temp)	0.0116	0.0033	3.56	0.0004
ID6392172:I(21 - temp)	-0.0009	0.0032	-0.28	0.7794
ID6393013:I(21 - temp)	0.0068	0.0033	2.09	0.0368
ID6393014:I(21 - temp)	0.0166	0.0032	5.11	0.0000
ID6540708:I(21 - temp)	0.0018	0.0033	0.56	0.5723
ID6567326:I(21 - temp)	0.0000	0.0033	0.00	0.9979
ID6618578:I(21 - temp)	0.1956	0.0033	60.13	0.0000
ID6618580:I(21 - temp)	0.0594	0.0033	18.26	0.0000
ID6627217:I(21 - temp)	0.0001	0.0033	0.02	0.9840
ID6627258:I(21 - temp)	0.0004	0.0033	0.14	0.8921
ID6627261:I(21 - temp)	0.0068	0.0033	2.10	0.0362
ID6627320:I(21 - temp)	0.0026	0.0033	0.80	0.4238
ID6681763:I(21 - temp)	0.0051	0.0033	1.56	0.1181
ID6681892:I(21 - temp)	0.0082	0.0033	2.51	0.0120
ID6681894:I(21 - temp)	-0.0003	0.0033	-0.10	0.9233
ID6790785:I(21 - temp)	-0.0001	0.0033	-0.04	0.9702
ID6790798:I(21 - temp)	0.0037	0.0033	1.14	0.2540
ID6842413:I(21 - temp)	0.0084	0.0033	2.57	0.0102
ID6842421:I(21 - temp)	0.0143	0.0033	4.39	0.0000
ID6842603:I(21 - temp)	0.0021	0.0033	0.65	0.5131
ID6842762:I(21 - temp)	0.0060	0.0033	1.84	0.0660
ID6921678:I(21 - temp)	0.2082	0.0033	64.01	0.0000
ID6940321:I(21 - temp)	0.0605	0.0033	18.59	0.0000
ID6940322:I(21 - temp)	0.0104	0.0033	3.21	0.0013
ID7072161:I(21 - temp)	0.0040	0.0033	1.22	0.2216
ID7072231:I(21 - temp)	0.0188	0.0033	5.79	0.0000
ID7072241:I(21 - temp)	0.0174	0.0032	5.36	0.0000
ID7072337:I(21 - temp)	0.0097	0.0033	2.98	0.0029
ID7183151:I(21 - temp)	0.0775	0.0033	23.83	0.0000
ID65005112:I(21 - temp)	0.0074	0.0033	2.28	0.0229
ID65012411:I(21 - temp)	0.0063	0.0032	1.94	0.0521
ID65014229:I(21 - temp)	0.0051	0.0033	1.56	0.1180
ID65014274:I(21 - temp)	0.0067	0.0033	2.05	0.0406
ID65052581:I(21 - temp)	0.0188	0.0033	5.78	0.0000
ID65063195:I(21 - temp)	0.0068	0.0033	2.08	0.0372
ID65063211:I(21 - temp)	0.0106	0.0033	3.24	0.0012
ID65063303:I(21 - temp)	0.0008	0.0033	0.24	0.8067

ID65067046:I(21 - temp)	0.0013	0.0032	0.39	0.6963
ID65118755:I(21 - temp)	0.0020	0.0032	0.61	0.5436
ID65118764:I(21 - temp)	0.0051	0.0032	1.56	0.1178
ID65118805:I(21 - temp)	0.0018	0.0033	0.54	0.5890
ID65118812:I(21 - temp)	0.0022	0.0033	0.67	0.5040
ID65118826:I(21 - temp)	0.0155	0.0033	4.75	0.0000
ID65118829:I(21 - temp)	0.0098	0.0033	3.01	0.0026
ID65118840:I(21 - temp)	-0.0013	0.0033	-0.38	0.7007
ID65118848:I(21 - temp)	0.0116	0.0033	3.57	0.0004
ID69001263:I(21 - temp)	0.0153	0.0033	4.71	0.0000
ID69001269:I(21 - temp)	0.0169	0.0033	5.21	0.0000
ID69089222:I(21 - temp)	0.0578	0.0033	17.75	0.0000
ID69250492:I(21 - temp)	0.0184	0.0033	5.66	0.0000
ID69429582:I(21 - temp)	0.3026	0.0033	93.01	0.0000
ID69469107:I(21 - temp)	0.1182	0.0033	36.32	0.0000
ID69478883:I(21 - temp)	0.0814	0.0033	25.02	0.0000
ID69518080:I(21 - temp)	0.0362	0.0033	11.14	0.0000
ID69518092:I(21 - temp)	0.0712	0.0033	21.89	0.0000
ID69580701:I(21 - temp)	0.1595	0.0032	49.08	0.0000
ID69585544:I(21 - temp)	0.2783	0.0033	85.56	0.0000
ID69585545:I(21 - temp)	0.0272	0.0033	8.38	0.0000
ID69611360:I(21 - temp)	0.0046	0.0033	1.42	0.1549
ID69652588:I(21 - temp)	-0.0030	0.0033	-0.91	0.3603
ID69652603:I(21 - temp)	0.0104	0.0033	3.20	0.0014
ID69688095:I(21 - temp)	0.0058	0.0033	1.77	0.0769
ID69749518:I(21 - temp)	0.0063	0.0033	1.95	0.0513
ID69861509:I(21 - temp)	0.0042	0.0033	1.29	0.1962
ID69999051:I(21 - temp)	-0.0037	0.0033	-1.13	0.2600
ID69999094:I(21 - temp)	0.1527	0.0032	46.98	0.0000
ID78082613:I(21 - temp)	0.0241	0.0033	7.42	0.0000
ID78138095:I(21 - temp)	0.2921	0.0033	89.79	0.0000
ID78185925:I(21 - temp)	0.2167	0.0033	66.19	0.0000
ID78443775:I(21 - temp)	0.0005	0.0033	0.15	0.8832
ID78673711:I(21 - temp)	0.0087	0.0033	2.68	0.0073
week36:I(21 - temp)	-0.0119	0.0221	-0.54	0.5901
week37:I(21 - temp)	-0.0035	0.0219	-0.16	0.8738
week38:I(21 - temp)	-0.0015	0.0217	-0.07	0.9460
week39:I(21 - temp)	0.0065	0.0218	0.30	0.7652
week40:I(21 - temp)	0.0033	0.0219	0.15	0.8799
week41:I(21 - temp)	0.0097	0.0218	0.44	0.6576
week42:I(21 - temp)	0.0044	0.0217	0.20	0.8399
week43:I(21 - temp)	0.0103	0.0216	0.48	0.6344
week44:I(21 - temp)	0.0103	0.0218	0.47	0.6362
week45:I(21 - temp)	-0.0032	0.0221	-0.14	0.8863
week46:I(21 - temp)	0.0010	0.0217	0.05	0.9628
week47:I(21 - temp)	0.0012	0.0220	0.05	0.9565
week48:I(21 - temp)	0.0178	0.0217	0.82	0.4122
week49:I(21 - temp)	0.0145	0.0217	0.67	0.5057
week50:I(21 - temp)	-0.0012	0.0218	-0.05	0.9569
week51:I(21 - temp)	0.0048	0.0218	0.22	0.8249
week52:I(21 - temp)	0.0160	0.0217	0.73	0.4633

Table 13: Summary for lm_1

6.C.3 Linear model 2 (lm_2)

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-5.2998	0.5372	-9.87	0.0000
ID4529800	-0.1379	0.0671	-2.06	0.0398
ID4839509	0.3548	0.0671	5.29	0.0000
ID4866195	0.0015	0.0670	0.02	0.9825
ID4887707	-0.2729	0.0671	-4.07	0.0000
ID4962433	-0.3595	0.0671	-5.36	0.0000
ID5037175	-0.1230	0.0671	-1.83	0.0668
ID5093913	-0.5488	0.0671	-8.18	0.0000
ID5093998	-0.2841	0.0671	-4.23	0.0000
ID5140250	0.0201	0.0670	0.30	0.7640
ID5325295	-0.6625	0.0671	-9.87	0.0000
ID6392057	-0.5003	0.0671	-7.46	0.0000
ID6392146	-0.3333	0.0671	-4.97	0.0000
ID6392172	0.1374	0.0669	2.05	0.0400
ID6393013	-0.3191	0.0671	-4.76	0.0000

ID6393014	-0.2391	0.0670	-3.57	0.0004
ID6540708	-0.3406	0.0671	-5.08	0.0000
ID6567326	0.2815	0.0671	4.20	0.0000
ID6618578	-0.4206	0.0671	-6.27	0.0000
ID6618580	-0.6109	0.0671	-9.11	0.0000
ID6627217	-0.2157	0.0671	-3.22	0.0013
ID6627258	0.0289	0.0671	0.43	0.6669
ID6627261	-0.1373	0.0671	-2.05	0.0407
ID6627320	-0.3341	0.0671	-4.98	0.0000
ID6681763	-0.3620	0.0671	-5.39	0.0000
ID6681892	-0.3722	0.0671	-5.55	0.0000
ID6681894	-0.7726	0.0671	-11.52	0.0000
ID6790785	0.4314	0.0671	6.43	0.0000
ID6790798	-0.1379	0.0671	-2.06	0.0399
ID6842413	-0.1859	0.0671	-2.77	0.0056
ID6842421	-0.3268	0.0671	-4.87	0.0000
ID6842603	0.0443	0.0671	0.66	0.5094
ID6842762	-0.2505	0.0671	-3.73	0.0002
ID6921678	-0.1409	0.0671	-2.10	0.0357
ID6940321	-0.5122	0.0671	-7.63	0.0000
ID6940322	-0.0825	0.0671	-1.23	0.2191
ID7072161	-0.2602	0.0671	-3.88	0.0001
ID7072231	-0.5208	0.0671	-7.76	0.0000
ID7072241	-0.4586	0.0670	-6.84	0.0000
ID7072337	-0.4316	0.0671	-6.43	0.0000
ID7183151	-0.3336	0.0671	-4.97	0.0000
ID65005112	-0.1414	0.0671	-2.11	0.0350
ID65012411	-0.3837	0.0669	-5.74	0.0000
ID65014229	-0.1814	0.0671	-2.70	0.0068
ID65014274	-0.1464	0.0671	-2.18	0.0291
ID65052581	-0.4228	0.0671	-6.30	0.0000
ID65063195	-0.3500	0.0671	-5.22	0.0000
ID65063211	-0.2654	0.0671	-3.96	0.0001
ID65063303	-0.1915	0.0671	-2.86	0.0043
ID65067046	-0.1098	0.0670	-1.64	0.1011
ID65118755	-0.1497	0.0668	-2.24	0.0252
ID65118764	-0.1938	0.0667	-2.90	0.0037
ID65118805	-0.3563	0.0671	-5.31	0.0000
ID65118812	-0.5616	0.0671	-8.37	0.0000
ID65118826	-0.1420	0.0671	-2.12	0.0343
ID65118829	-0.3515	0.0671	-5.24	0.0000
ID65118840	-0.0767	0.0671	-1.14	0.2529
ID65118848	-0.3204	0.0671	-4.78	0.0000
ID69001263	-0.2055	0.0671	-3.06	0.0022
ID69001269	-0.1155	0.0671	-1.72	0.0852
ID69089222	-0.1647	0.0671	-2.46	0.0141
ID69250492	-0.2875	0.0671	-4.29	0.0000
ID69429582	-0.4731	0.0671	-7.05	0.0000
ID69469107	-0.7616	0.0671	-11.35	0.0000
ID69478883	0.2192	0.0671	3.27	0.0011
ID69518080	-0.4531	0.0671	-6.75	0.0000
ID69518092	-0.1714	0.0671	-2.55	0.0107
ID69580701	-0.1810	0.0670	-2.70	0.0069
ID69585544	-0.5351	0.0671	-7.98	0.0000
ID69585545	-0.3636	0.0671	-5.42	0.0000
ID69611360	-0.1717	0.0671	-2.56	0.0105
ID69652588	-0.1004	0.0671	-1.50	0.1346
ID69652603	-0.1477	0.0671	-2.20	0.0277
ID69688095	-0.1026	0.0671	-1.53	0.1261
ID69749518	-0.4534	0.0671	-6.76	0.0000
ID69861509	-0.4636	0.0671	-6.91	0.0000
ID69999051	0.8154	0.0671	12.15	0.0000
ID69999094	-0.3946	0.0670	-5.89	0.0000
ID78082613	-0.2463	0.0671	-3.67	0.0002
ID78138095	-0.2673	0.0671	-3.98	0.0001
ID78185925	-0.0765	0.0677	-1.13	0.2589
ID78443775	-0.0999	0.0671	-1.49	0.1363
ID78673711	-0.4581	0.0671	-6.83	0.0000
hum	-0.0053	0.0009	-5.74	0.0000
wind_spd	-0.0075	0.0046	-1.64	0.1015
pressure	0.0059	0.0005	12.21	0.0000
week36	0.1870	0.1797	1.04	0.2981
week37	-0.0249	0.1908	-0.13	0.8963

week38	0.1547	0.1790	0.86	0.3875
week39	-0.1632	0.2012	-0.81	0.4173
week40	0.3032	0.1855	1.63	0.1022
week41	0.0558	0.1823	0.31	0.7594
week42	0.3978	0.1817	2.19	0.0286
week43	0.3153	0.1922	1.64	0.1009
week44	0.4424	0.1893	2.34	0.0194
week45	0.9890	0.2151	4.60	0.0000
week46	0.9700	0.1929	5.03	0.0000
week47	1.0177	0.3643	2.79	0.0052
week48	0.3221	0.1935	1.66	0.0960
week49	0.8561	0.1977	4.33	0.0000
week50	0.7777	0.2819	2.76	0.0058
week51	1.3961	0.2599	5.37	0.0000
week52	-0.1362	0.2374	-0.57	0.5661
temp	0.0041	0.0354	0.12	0.9068
ID4529800:temp	0.0112	0.0051	2.22	0.0265
ID4839509:temp	-0.0289	0.0051	-5.71	0.0000
ID4866195:temp	-0.0000	0.0051	-0.00	0.9974
ID4887707:temp	0.0222	0.0051	4.39	0.0000
ID4962433:temp	0.0293	0.0051	5.79	0.0000
ID5037175:temp	0.0100	0.0051	1.98	0.0479
ID5093913:temp	0.0447	0.0051	8.83	0.0000
ID5093998:temp	0.0231	0.0051	4.57	0.0000
ID5140250:temp	-0.0015	0.0051	-0.30	0.7673
ID5325295:temp	0.0539	0.0051	10.66	0.0000
ID6392057:temp	0.0407	0.0051	8.05	0.0000
ID6392146:temp	0.0271	0.0051	5.36	0.0000
ID6392172:temp	-0.0111	0.0050	-2.19	0.0283
ID6393013:temp	0.0260	0.0051	5.13	0.0000
ID6393014:temp	0.0196	0.0051	3.88	0.0001
ID6540708:temp	0.0277	0.0051	5.48	0.0000
ID6567326:temp	-0.0229	0.0051	-4.53	0.0000
ID6618578:temp	0.0343	0.0051	6.77	0.0000
ID6618580:temp	0.0498	0.0051	9.83	0.0000
ID6627217:temp	0.0176	0.0051	3.47	0.0005
ID6627258:temp	-0.0024	0.0051	-0.46	0.6423
ID6627261:temp	0.0112	0.0051	2.21	0.0272
ID6627320:temp	0.0272	0.0051	5.38	0.0000
ID6681763:temp	0.0295	0.0051	5.82	0.0000
ID6681892:temp	0.0303	0.0051	5.99	0.0000
ID6681894:temp	0.0629	0.0051	12.43	0.0000
ID6790785:temp	-0.0351	0.0051	-6.94	0.0000
ID6790798:temp	0.0113	0.0051	2.22	0.0261
ID6842413:temp	0.0151	0.0051	2.99	0.0028
ID6842421:temp	0.0266	0.0051	5.26	0.0000
ID6842603:temp	-0.0036	0.0051	-0.71	0.4764
ID6842762:temp	0.0204	0.0051	4.03	0.0001
ID6921678:temp	0.0115	0.0051	2.27	0.0234
ID6940321:temp	0.0417	0.0051	8.24	0.0000
ID6940322:temp	0.0067	0.0051	1.33	0.1847
ID7072161:temp	0.0212	0.0051	4.19	0.0000
ID7072231:temp	0.0424	0.0051	8.38	0.0000
ID7072241:temp	0.0372	0.0051	7.36	0.0000
ID7072337:temp	0.0352	0.0051	6.94	0.0000
ID7183151:temp	0.0272	0.0051	5.37	0.0000
ID65005112:temp	0.0115	0.0051	2.28	0.0229
ID65012411:temp	0.0313	0.0051	6.20	0.0000
ID65014229:temp	0.0148	0.0051	2.92	0.0035
ID65014274:temp	0.0119	0.0051	2.36	0.0185
ID65052581:temp	0.0344	0.0051	6.80	0.0000
ID65063195:temp	0.0285	0.0051	5.63	0.0000
ID65063211:temp	0.0216	0.0051	4.27	0.0000
ID65063303:temp	0.0156	0.0051	3.08	0.0021
ID65067046:temp	0.0089	0.0051	1.77	0.0769
ID65118755:temp	0.0122	0.0050	2.41	0.0158
ID65118764:temp	0.0158	0.0050	3.13	0.0018
ID65118805:temp	0.0290	0.0051	5.73	0.0000
ID65118812:temp	0.0457	0.0051	9.03	0.0000
ID65118826:temp	0.0116	0.0051	2.28	0.0224
ID65118829:temp	0.0286	0.0051	5.65	0.0000
ID65118840:temp	0.0062	0.0051	1.23	0.2172
ID65118848:temp	0.0261	0.0051	5.15	0.0000

ID69001263:temp	0.0167	0.0051	3.31	0.0010
ID69001269:temp	0.0094	0.0051	1.86	0.0632
ID69089222:temp	0.0134	0.0051	2.65	0.0081
ID69250492:temp	0.0234	0.0051	4.63	0.0000
ID69429582:temp	0.0385	0.0051	7.61	0.0000
ID69469107:temp	0.0620	0.0051	12.25	0.0000
ID69478883:temp	-0.0179	0.0051	-3.53	0.0004
ID69518080:temp	0.0369	0.0051	7.29	0.0000
ID69518092:temp	0.0140	0.0051	2.76	0.0058
ID69580701:temp	0.0149	0.0051	2.94	0.0033
ID69585544:temp	0.0436	0.0051	8.61	0.0000
ID69585545:temp	0.0296	0.0051	5.85	0.0000
ID69611360:temp	0.0140	0.0051	2.76	0.0058
ID69652588:temp	0.0082	0.0051	1.62	0.1063
ID69652603:temp	0.0120	0.0051	2.38	0.0175
ID69688095:temp	0.0084	0.0051	1.65	0.0987
ID69749518:temp	0.0369	0.0051	7.29	0.0000
ID69861509:temp	0.0378	0.0051	7.46	0.0000
ID69999051:temp	-0.0664	0.0051	-13.10	0.0000
ID69999094:temp	0.0323	0.0051	6.39	0.0000
ID78082613:temp	0.0201	0.0051	3.96	0.0001
ID78138095:temp	0.0218	0.0051	4.30	0.0000
ID78185925:temp	0.0057	0.0051	1.12	0.2610
ID78443775:temp	0.0081	0.0051	1.61	0.1079
ID78673711:temp	0.0373	0.0051	7.37	0.0000
week36:temp	-0.0099	0.0361	-0.28	0.7832
week37:temp	0.0200	0.0357	0.56	0.5754
week38:temp	0.0068	0.0354	0.19	0.8488
week39:temp	0.0375	0.0360	1.04	0.2977
week40:temp	0.0071	0.0356	0.20	0.8409
week41:temp	0.0525	0.0355	1.48	0.1396
week42:temp	0.0098	0.0355	0.28	0.7816
week43:temp	0.0331	0.0354	0.94	0.3490
week44:temp	0.0004	0.0357	0.01	0.9921
week45:temp	-0.0062	0.0363	-0.17	0.8645
week46:temp	-0.0147	0.0356	-0.41	0.6784
week47:temp	0.0045	0.0386	0.12	0.9063
week48:temp	0.0445	0.0354	1.25	0.2096
week49:temp	0.0165	0.0356	0.46	0.6430
week50:temp	0.0161	0.0365	0.44	0.6588
week51:temp	-0.0057	0.0362	-0.16	0.8739
week52:temp	0.0679	0.0370	1.83	0.0668
wind_spd:week36	-0.0075	0.0030	-2.52	0.0119
wind_spd:week37	-0.0009	0.0039	-0.23	0.8182
wind_spd:week38	-0.0051	0.0033	-1.53	0.1249
wind_spd:week39	0.0023	0.0031	0.76	0.4495
wind_spd:week40	0.0033	0.0026	1.28	0.2005
wind_spd:week41	-0.0084	0.0037	-2.26	0.0238
wind_spd:week42	-0.0115	0.0033	-3.46	0.0005
wind_spd:week43	-0.0070	0.0027	-2.63	0.0085
wind_spd:week44	0.0142	0.0029	4.93	0.0000
wind_spd:week45	-0.0167	0.0031	-5.40	0.0000
wind_spd:week46	-0.0114	0.0042	-2.73	0.0063
wind_spd:week47	-0.0126	0.0034	-3.73	0.0002
wind_spd:week48	-0.0039	0.0026	-1.53	0.1256
wind_spd:week49	-0.0087	0.0027	-3.27	0.0011
wind_spd:week50	-0.0023	0.0035	-0.64	0.5211
wind_spd:week51	-0.0129	0.0038	-3.37	0.0008
hum:wind_spd	0.0002	0.0000	4.15	0.0000

Table 14: Summary lm_2

6.C.4 Linear model 3 (lm_3)

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-1.7713	0.4517	-3.92	0.0001
ID4529800	-0.1581	0.0626	-2.53	0.0115
ID4839509	0.5337	0.0626	8.53	0.0000
ID4866195	0.1924	0.0626	3.08	0.0021
ID4887707	-0.0556	0.0626	-0.89	0.3742
ID4962433	-0.1504	0.0626	-2.40	0.0164
ID5037175	0.0944	0.0626	1.51	0.1315

ID5093913	-0.3653	0.0626	-5.84	0.0000
ID5093998	-0.0756	0.0626	-1.21	0.2271
ID5140250	0.1788	0.0626	2.85	0.0043
ID5325295	-0.5845	0.0626	-9.34	0.0000
ID6392057	-0.3839	0.0626	-6.14	0.0000
ID6392146	-0.1741	0.0626	-2.78	0.0054
ID6392172	-0.1011	0.0661	-1.53	0.1260
ID6393013	-0.5000	0.0626	-7.99	0.0000
ID6393014	-0.0553	0.0626	-0.88	0.3770
ID6540708	-0.2972	0.0626	-4.75	0.0000
ID6567326	0.4869	0.0626	7.78	0.0000
ID6618578	-0.2660	0.0626	-4.25	0.0000
ID6618580	-0.4058	0.0626	-6.49	0.0000
ID6627217	-0.2369	0.0626	-3.79	0.0002
ID6627258	0.0175	0.0626	0.28	0.7792
ID6627261	0.0729	0.0626	1.17	0.2438
ID6627320	-0.3218	0.0626	-5.14	0.0000
ID6681763	-0.1394	0.0628	-2.22	0.0264
ID6681892	-0.1741	0.0626	-2.78	0.0054
ID6681894	-0.5718	0.0626	-9.14	0.0000
ID6790785	0.5411	0.0626	8.64	0.0000
ID6790798	0.0666	0.0626	1.07	0.2869
ID6842413	-0.0347	0.0626	-0.56	0.5789
ID6842421	-0.4257	0.0626	-6.80	0.0000
ID6842603	0.2014	0.0626	3.22	0.0013
ID6842762	-0.0378	0.0626	-0.60	0.5457
ID6921678	0.0762	0.0626	1.22	0.2232
ID6940321	-0.3302	0.0626	-5.28	0.0000
ID6940322	-0.0022	0.0626	-0.04	0.9714
ID7072161	-0.1316	0.0626	-2.10	0.0355
ID7072231	-0.3059	0.0626	-4.89	0.0000
ID7072241	-0.5922	0.0626	-9.46	0.0000
ID7072337	-0.3875	0.0626	-6.19	0.0000
ID7183151	-0.1564	0.0626	-2.50	0.0124
ID65005112	0.1059	0.0626	1.69	0.0904
ID65012411	-0.2737	0.0627	-4.37	0.0000
ID65014229	0.0087	0.0626	0.14	0.8891
ID65014274	-0.0310	0.0626	-0.50	0.6205
ID65052581	-0.3734	0.0626	-5.97	0.0000
ID65063195	-0.1455	0.0626	-2.33	0.0200
ID65063211	-0.2055	0.0626	-3.29	0.0010
ID65063303	-0.2055	0.0626	-3.29	0.0010
ID65067046	0.0645	0.0628	1.03	0.3039
ID65118755	-0.1676	0.0642	-2.61	0.0090
ID65118764	-0.0941	0.0646	-1.46	0.1453
ID65118805	-0.4227	0.0626	-6.76	0.0000
ID65118826	0.0624	0.0626	1.00	0.3184
ID65118829	-0.1088	0.0626	-1.74	0.0819
ID65118840	0.1691	0.0626	2.70	0.0069
ID65118848	-0.2768	0.0626	-4.42	0.0000
ID69001263	-0.2236	0.0626	-3.57	0.0004
ID69001269	0.0849	0.0626	1.36	0.1748
ID69089222	0.0431	0.0626	0.69	0.4912
ID69250492	-0.0503	0.0626	-0.80	0.4213
ID69429582	-0.2971	0.0626	-4.75	0.0000
ID69469107	-0.6304	0.0626	-10.08	0.0000
ID69478883	0.3610	0.0626	5.77	0.0000
ID69518080	-0.2600	0.0626	-4.16	0.0000
ID69518092	0.0470	0.0626	0.75	0.4524
ID69580701	0.0307	0.0626	0.49	0.6235
ID69585544	-0.3760	0.0644	-5.84	0.0000
ID69585545	-0.1764	0.0626	-2.82	0.0048
ID69611360	0.0214	0.0626	0.34	0.7321
ID69652588	0.0788	0.0626	1.26	0.2077
ID69652603	-0.0161	0.0626	-0.26	0.7967
ID69688095	-0.2227	0.0626	-3.56	0.0004
ID69749518	-0.2634	0.0626	-4.21	0.0000
ID69861509	-0.5279	0.0626	-8.44	0.0000
ID69999094	-0.1959	0.0626	-3.13	0.0017
ID78082613	-0.1230	0.0626	-1.97	0.0492
ID78138095	-0.1401	0.0626	-2.24	0.0251
ID78185925	0.1651	0.0626	2.64	0.0083
ID78443775	-0.0058	0.0626	-0.09	0.9259

ID78673711	-0.3361	0.0626	-5.37	0.0000
wind_spd	0.0030	0.0004	7.31	0.0000
pressure	0.0019	0.0004	4.52	0.0000
week36	0.0626	0.1520	0.41	0.6806
week37	0.1026	0.1426	0.72	0.4720
week38	0.3457	0.1478	2.34	0.0193
week39	0.0382	0.1449	0.26	0.7922
week40	0.0920	0.1480	0.62	0.5341
week41	0.1172	0.1420	0.83	0.4089
week42	0.2676	0.1436	1.86	0.0624
week43	-0.3695	0.1828	-2.02	0.0433
week44	-0.0371	0.1524	-0.24	0.8078
week45	0.6454	0.1590	4.06	0.0000
week46	0.7650	0.1652	4.63	0.0000
week47	0.3130	0.2178	1.44	0.1507
week48	0.4540	0.1546	2.94	0.0033
week49	0.3196	0.1550	2.06	0.0393
week50	0.2672	0.1799	1.49	0.1374
week51	0.2008	0.1763	1.14	0.2547
week52	0.0320	0.1516	0.21	0.8326
daytypeWorkweek	0.3327	0.0364	9.13	0.0000
temp	0.0183	0.0272	0.67	0.5000
ID4529800:temp	0.0113	0.0040	2.82	0.0048
ID4839509:temp	-0.0296	0.0040	-7.38	0.0000
ID4866195:temp	-0.0015	0.0040	-0.37	0.7120
ID4887707:temp	0.0214	0.0040	5.33	0.0000
ID4962433:temp	0.0284	0.0040	7.08	0.0000
ID5037175:temp	0.0092	0.0040	2.28	0.0224
ID5093913:temp	0.0440	0.0040	10.96	0.0000
ID5093998:temp	0.0223	0.0040	5.56	0.0000
ID5140250:temp	-0.0011	0.0040	-0.29	0.7753
ID5325295:temp	0.0536	0.0040	13.37	0.0000
ID6392057:temp	0.0403	0.0040	10.04	0.0000
ID6392146:temp	0.0265	0.0040	6.61	0.0000
ID6392172:temp	0.0299	0.0043	6.93	0.0000
ID6393013:temp	0.0267	0.0040	6.66	0.0000
ID6393014:temp	0.0183	0.0040	4.55	0.0000
ID6540708:temp	0.0276	0.0040	6.87	0.0000
ID6567326:temp	-0.0237	0.0040	-5.92	0.0000
ID6618578:temp	0.0336	0.0040	8.39	0.0000
ID6618580:temp	0.0489	0.0040	12.20	0.0000
ID6627217:temp	0.0177	0.0040	4.40	0.0000
ID6627258:temp	-0.0023	0.0040	-0.58	0.5652
ID6627261:temp	0.0104	0.0040	2.58	0.0099
ID6627320:temp	0.0272	0.0040	6.77	0.0000
ID6681763:temp	0.0281	0.0040	6.97	0.0000
ID6681892:temp	0.0295	0.0040	7.36	0.0000
ID6681894:temp	0.0621	0.0040	15.49	0.0000
ID6790785:temp	-0.0336	0.0040	-8.36	0.0000
ID6790798:temp	0.0103	0.0040	2.58	0.0100
ID6842413:temp	0.0145	0.0040	3.63	0.0003
ID6842421:temp	0.0270	0.0040	6.73	0.0000
ID6842603:temp	-0.0042	0.0040	-1.05	0.2923
ID6842762:temp	0.0196	0.0040	4.88	0.0000
ID6921678:temp	0.0106	0.0040	2.65	0.0081
ID6940321:temp	0.0410	0.0040	10.22	0.0000
ID6940322:temp	0.0064	0.0040	1.60	0.1107
ID7072161:temp	0.0207	0.0040	5.16	0.0000
ID7072231:temp	0.0416	0.0040	10.36	0.0000
ID7072241:temp	0.0410	0.0040	10.22	0.0000
ID7072337:temp	0.0350	0.0040	8.72	0.0000
ID7183151:temp	0.0265	0.0040	6.60	0.0000
ID65005112:temp	0.0105	0.0040	2.63	0.0086
ID65012411:temp	0.0302	0.0040	7.51	0.0000
ID65014229:temp	0.0140	0.0040	3.50	0.0005
ID65014274:temp	0.0115	0.0040	2.86	0.0042
ID65052581:temp	0.0342	0.0040	8.54	0.0000
ID65063195:temp	0.0277	0.0040	6.90	0.0000
ID65063211:temp	0.0214	0.0040	5.33	0.0000
ID65063303:temp	0.0157	0.0040	3.90	0.0001
ID65067046:temp	0.0070	0.0040	1.75	0.0805
ID65118755:temp	0.0316	0.0042	7.60	0.0000
ID65118764:temp	0.0209	0.0041	5.08	0.0000

ID65118805:temp	0.0293	0.0040	7.30	0.0000
ID65118826:temp	0.0108	0.0040	2.68	0.0073
ID65118829:temp	0.0277	0.0040	6.90	0.0000
ID65118840:temp	0.0053	0.0040	1.32	0.1883
ID65118848:temp	0.0259	0.0040	6.46	0.0000
ID69001263:temp	0.0168	0.0040	4.19	0.0000
ID69001269:temp	0.0086	0.0040	2.15	0.0318
ID69089222:temp	0.0126	0.0040	3.14	0.0017
ID69250492:temp	0.0225	0.0040	5.60	0.0000
ID69429582:temp	0.0378	0.0040	9.43	0.0000
ID69469107:temp	0.0615	0.0040	15.33	0.0000
ID69478883:temp	-0.0184	0.0040	-4.59	0.0000
ID69518080:temp	0.0361	0.0040	9.01	0.0000
ID69518092:temp	0.0131	0.0040	3.26	0.0011
ID69580701:temp	0.0134	0.0040	3.34	0.0008
ID69585544:temp	0.0400	0.0042	9.55	0.0000
ID69585545:temp	0.0289	0.0040	7.20	0.0000
ID69611360:temp	0.0132	0.0040	3.30	0.0010
ID69652588:temp	0.0075	0.0040	1.86	0.0626
ID69652603:temp	0.0115	0.0040	2.87	0.0041
ID69688095:temp	0.0088	0.0040	2.20	0.0277
ID69749518:temp	0.0362	0.0040	9.02	0.0000
ID69861509:temp	0.0380	0.0040	9.48	0.0000
ID69999094:temp	0.0309	0.0040	7.72	0.0000
ID78082613:temp	0.0196	0.0040	4.88	0.0000
ID78138095:temp	0.0213	0.0040	5.30	0.0000
ID78185925:temp	-0.0054	0.0040	-1.36	0.1751
ID78443775:temp	0.0078	0.0040	1.94	0.0528
ID78673711:temp	0.0368	0.0040	9.18	0.0000
ID4529800:daytypeWorkweek	0.0270	0.0442	0.61	0.5408
ID4839509:daytypeWorkweek	-0.2392	0.0442	-5.42	0.0000
ID4866195:daytypeWorkweek	-0.2344	0.0442	-5.30	0.0000
ID4887707:daytypeWorkweek	-0.2905	0.0442	-6.58	0.0000
ID4962433:daytypeWorkweek	-0.2636	0.0444	-5.94	0.0000
ID5037175:daytypeWorkweek	-0.2905	0.0442	-6.58	0.0000
ID5093913:daytypeWorkweek	-0.2454	0.0442	-5.56	0.0000
ID5093998:daytypeWorkweek	-0.2788	0.0442	-6.31	0.0000
ID5140250:daytypeWorkweek	-0.2031	0.0442	-4.59	0.0000
ID5325295:daytypeWorkweek	-0.1043	0.0442	-2.36	0.0182
ID6392057:daytypeWorkweek	-0.1556	0.0442	-3.52	0.0004
ID6392146:daytypeWorkweek	-0.2129	0.0442	-4.82	0.0000
ID6392172:daytypeWorkweek	-0.1419	0.0456	-3.11	0.0019
ID6393013:daytypeWorkweek	0.2418	0.0442	5.48	0.0000
ID6393014:daytypeWorkweek	-0.2270	0.0442	-5.14	0.0000
ID6540708:daytypeWorkweek	-0.0580	0.0442	-1.31	0.1892
ID6567326:daytypeWorkweek	-0.2745	0.0442	-6.22	0.0000
ID6618578:daytypeWorkweek	-0.2067	0.0442	-4.68	0.0000
ID6618580:daytypeWorkweek	-0.2742	0.0442	-6.21	0.0000
ID6627217:daytypeWorkweek	0.0283	0.0442	0.64	0.5218
ID6627258:daytypeWorkweek	0.0152	0.0442	0.34	0.7315
ID6627261:daytypeWorkweek	-0.2810	0.0442	-6.36	0.0000
ID6627320:daytypeWorkweek	-0.0165	0.0442	-0.37	0.7091
ID6681763:daytypeWorkweek	-0.2873	0.0442	-6.50	0.0000
ID6681892:daytypeWorkweek	-0.2648	0.0442	-6.00	0.0000
ID6681894:daytypeWorkweek	-0.2685	0.0442	-6.08	0.0000
ID6790785:daytypeWorkweek	-0.1659	0.0442	-3.75	0.0002
ID6790798:daytypeWorkweek	-0.2598	0.0442	-5.88	0.0000
ID6842413:daytypeWorkweek	-0.2021	0.0442	-4.58	0.0000
ID6842421:daytypeWorkweek	0.1322	0.0442	2.99	0.0028
ID6842603:daytypeWorkweek	-0.2100	0.0442	-4.76	0.0000
ID6842762:daytypeWorkweek	-0.2843	0.0442	-6.44	0.0000
ID6921678:daytypeWorkweek	-0.2902	0.0442	-6.57	0.0000
ID6940321:daytypeWorkweek	-0.2432	0.0442	-5.51	0.0000
ID6940322:daytypeWorkweek	-0.1072	0.0442	-2.43	0.0152
ID7072161:daytypeWorkweek	-0.1720	0.0442	-3.90	0.0001
ID7072231:daytypeWorkweek	-0.2872	0.0442	-6.51	0.0000
ID7072241:daytypeWorkweek	0.1437	0.0442	3.25	0.0012
ID7072337:daytypeWorkweek	-0.0589	0.0442	-1.33	0.1820
ID7183151:daytypeWorkweek	-0.2368	0.0442	-5.36	0.0000
ID65005112:daytypeWorkweek	-0.3307	0.0442	-7.49	0.0000
ID65012411:daytypeWorkweek	-0.1060	0.0443	-2.39	0.0167
ID65014229:daytypeWorkweek	-0.2542	0.0442	-5.76	0.0000
ID65014274:daytypeWorkweek	-0.1543	0.0442	-3.50	0.0005

ID65052581:daytypeWorkweek	-0.0661	0.0442	-1.50	0.1343
ID65063195:daytypeWorkweek	-0.2733	0.0442	-6.19	0.0000
ID65063211:daytypeWorkweek	-0.0801	0.0442	-1.81	0.0698
ID65063303:daytypeWorkweek	0.0187	0.0442	0.42	0.6717
ID65067046:daytypeWorkweek	-0.1954	0.0442	-4.42	0.0000
ID65118755:daytypeWorkweek	-0.1075	0.0455	-2.36	0.0181
ID65118764:daytypeWorkweek	-0.1470	0.0455	-3.23	0.0012
ID65118805:daytypeWorkweek	0.0888	0.0442	2.01	0.0443
ID65118826:daytypeWorkweek	-0.2732	0.0442	-6.19	0.0000
ID65118829:daytypeWorkweek	-0.3244	0.0442	-7.35	0.0000
ID65118840:daytypeWorkweek	-0.3285	0.0442	-7.44	0.0000
ID65118848:daytypeWorkweek	-0.0583	0.0442	-1.32	0.1868
ID69001263:daytypeWorkweek	0.0242	0.0442	0.55	0.5842
ID69001269:daytypeWorkweek	-0.2679	0.0442	-6.07	0.0000
ID69089222:daytypeWorkweek	-0.2778	0.0442	-6.29	0.0000
ID69250492:daytypeWorkweek	-0.3171	0.0442	-7.18	0.0000
ID69429582:daytypeWorkweek	-0.2353	0.0442	-5.33	0.0000
ID69469107:daytypeWorkweek	-0.1754	0.0442	-3.97	0.0001
ID69478883:daytypeWorkweek	-0.1896	0.0442	-4.29	0.0000
ID69518080:daytypeWorkweek	-0.2582	0.0442	-5.85	0.0000
ID69518092:daytypeWorkweek	-0.2919	0.0442	-6.61	0.0000
ID69580701:daytypeWorkweek	-0.2639	0.0442	-5.97	0.0000
ID69585544:daytypeWorkweek	-0.1462	0.0445	-3.29	0.0010
ID69585545:daytypeWorkweek	-0.2503	0.0442	-5.67	0.0000
ID69611360:daytypeWorkweek	-0.2581	0.0442	-5.85	0.0000
ID69652588:daytypeWorkweek	-0.2396	0.0442	-5.43	0.0000
ID69652603:daytypeWorkweek	-0.1759	0.0442	-3.98	0.0001
ID69688095:daytypeWorkweek	0.1606	0.0442	3.64	0.0003
ID69749518:daytypeWorkweek	-0.2540	0.0442	-5.75	0.0000
ID69861509:daytypeWorkweek	0.0860	0.0442	1.95	0.0516
ID69999094:daytypeWorkweek	-0.2483	0.0442	-5.62	0.0000
ID78082613:daytypeWorkweek	-0.1647	0.0442	-3.73	0.0002
ID78138095:daytypeWorkweek	-0.1699	0.0442	-3.85	0.0001
ID78185925:daytypeWorkweek	-0.1381	0.0442	-3.13	0.0018
ID78443775:daytypeWorkweek	-0.1258	0.0442	-2.85	0.0044
ID78673711:daytypeWorkweek	-0.1630	0.0442	-3.69	0.0002
week36:daytypeWorkweek	-0.1062	0.0302	-3.52	0.0004
week37:daytypeWorkweek	-0.1698	0.0242	-7.01	0.0000
week38:daytypeWorkweek	-0.2880	0.0382	-7.53	0.0000
week39:daytypeWorkweek	-0.1077	0.0237	-4.54	0.0000
week40:daytypeWorkweek	0.0058	0.0248	0.23	0.8161
week41:daytypeWorkweek	-0.1314	0.0281	-4.67	0.0000
week42:daytypeWorkweek	-0.1556	0.0247	-6.30	0.0000
week43:daytypeWorkweek	0.0573	0.0467	1.23	0.2196
week44:daytypeWorkweek	0.0562	0.0260	2.16	0.0304
week45:daytypeWorkweek	-0.0459	0.0252	-1.82	0.0685
week46:daytypeWorkweek	-0.2067	0.0304	-6.81	0.0000
week47:daytypeWorkweek	-0.0710	0.0282	-2.52	0.0117
week48:daytypeWorkweek	0.0317	0.0305	1.04	0.2988
week49:daytypeWorkweek	-0.0141	0.0236	-0.60	0.5496
week50:daytypeWorkweek	-0.0067	0.0264	-0.25	0.7997
week36:temp	-0.0135	0.0286	-0.47	0.6355
week37:temp	0.0008	0.0274	0.03	0.9761
week38:temp	-0.0278	0.0274	-1.02	0.3099
week39:temp	0.0201	0.0273	0.73	0.4631
week40:temp	0.0116	0.0273	0.42	0.6718
week41:temp	0.0234	0.0275	0.85	0.3955
week42:temp	0.0030	0.0272	0.11	0.9120
week43:temp	0.0525	0.0279	1.88	0.0597
week44:temp	0.0326	0.0273	1.19	0.2329
week45:temp	-0.0162	0.0278	-0.58	0.5608
week46:temp	-0.0153	0.0276	-0.55	0.5793
week47:temp	0.0218	0.0285	0.76	0.4444
week48:temp	0.0167	0.0274	0.61	0.5426
week49:temp	0.0221	0.0273	0.81	0.4190
week50:temp	0.0275	0.0277	0.99	0.3215
week51:temp	0.0305	0.0276	1.11	0.2683
week52:temp	0.0403	0.0273	1.48	0.1394

Table 15: Summary lm_3

6.C.5 Linear model 4 (lm_4)

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	5.7025	0.5257	10.85	0.0000
ID4866195	-0.1475	0.0744	-1.98	0.0474
ID4887707	-0.5799	0.0743	-7.80	0.0000
ID4962433	-0.5551	0.0746	-7.44	0.0000
ID5037175	-0.3328	0.0743	-4.48	0.0000
ID5093913	-0.8266	0.0743	-11.12	0.0000
ID5093998	-0.4742	0.0743	-6.38	0.0000
ID5140250	-0.2049	0.0746	-2.75	0.0060
ID5325295	-1.3415	0.0743	-18.05	0.0000
ID6392057	-0.7935	0.0746	-10.64	0.0000
ID6392146	-0.6930	0.0743	-9.33	0.0000
ID6392172	-0.3558	0.0801	-4.44	0.0000
ID6393014	-0.5491	0.0743	-7.39	0.0000
ID6540708	-0.8546	0.0743	-11.50	0.0000
ID6567326	-0.0378	0.0743	-0.51	0.6113
ID6618578	-0.5740	0.0743	-7.72	0.0000
ID6618580	-0.9011	0.0743	-12.13	0.0000
ID6627217	-0.7833	0.0743	-10.54	0.0000
ID6627258	-0.5192	0.0743	-6.99	0.0000
ID6627261	-0.3095	0.0743	-4.17	0.0000
ID6627320	-0.9338	0.0743	-12.57	0.0000
ID6681763	-0.5615	0.0743	-7.56	0.0000
ID6681892	-0.5815	0.0743	-7.83	0.0000
ID6681894	-1.1436	0.0743	-15.39	0.0000
ID6790785	0.0946	0.0745	1.27	0.2041
ID6790798	-0.3416	0.0744	-4.59	0.0000
ID6842413	-0.4279	0.0743	-5.76	0.0000
ID6842421	-0.9724	0.0743	-13.09	0.0000
ID6842603	-0.4229	0.0743	-5.69	0.0000
ID6842762	-0.5165	0.0743	-6.95	0.0000
ID6921678	-0.3629	0.0743	-4.88	0.0000
ID6940321	-0.8988	0.0743	-12.10	0.0000
ID6940322	-0.3775	0.0743	-5.08	0.0000
ID7072161	-0.6462	0.0743	-8.70	0.0000
ID7072231	-0.8128	0.0743	-10.94	0.0000
ID7072241	-1.0334	0.0744	-13.89	0.0000
ID7072337	-0.8771	0.0743	-11.80	0.0000
ID7183151	-0.5881	0.0743	-7.91	0.0000
ID65005112	-0.3944	0.0743	-5.31	0.0000
ID65012411	-0.6369	0.0744	-8.57	0.0000
ID65014229	-0.5015	0.0743	-6.75	0.0000
ID65014274	-0.3757	0.0743	-5.06	0.0000
ID65052581	-0.8669	0.0743	-11.67	0.0000
ID65063195	-0.6426	0.0743	-8.65	0.0000
ID65063211	-0.6296	0.0752	-8.37	0.0000
ID65063303	-0.9500	0.0743	-12.78	0.0000
ID65067046	-0.4285	0.0743	-5.76	0.0000
ID65118755	-0.4608	0.0760	-6.06	0.0000
ID65118764	-0.5230	0.0746	-7.01	0.0000
ID65118805	-0.9228	0.0743	-12.42	0.0000
ID65118826	-0.1556	0.0743	-2.09	0.0363
ID65118829	-0.5812	0.0743	-7.82	0.0000
ID65118840	-0.1415	0.0743	-1.90	0.0570
ID69001263	-0.6745	0.0743	-9.08	0.0000
ID69001269	-0.3406	0.0743	-4.58	0.0000
ID69089222	-0.2874	0.0743	-3.87	0.0001
ID69250492	-0.3502	0.0743	-4.71	0.0000
ID69429582	-0.6959	0.0743	-9.37	0.0000
ID69469107	-1.2437	0.0743	-16.74	0.0000
ID69478883	-0.1751	0.0743	-2.36	0.0185
ID69518080	-0.7214	0.0743	-9.71	0.0000
ID69518092	-0.3258	0.0743	-4.38	0.0000
ID69580701	-0.4879	0.0744	-6.56	0.0000
ID69585544	-0.7858	0.0743	-10.57	0.0000
ID69585545	-0.5856	0.0743	-7.88	0.0000
ID69611360	-0.4974	0.0743	-6.69	0.0000
ID69652603	-0.4432	0.0743	-5.96	0.0000
ID69688095	-0.7084	0.0743	-9.53	0.0000
ID69749518	-0.5947	0.0743	-8.00	0.0000
ID69861509	-1.2163	0.0743	-16.37	0.0000
ID69999094	-0.6039	0.0744	-8.12	0.0000
ID78082613	-0.6496	0.0743	-8.74	0.0000

ID78138095	-0.5607	0.0743	-7.55	0.0000
ID78443775	-0.3910	0.0747	-5.23	0.0000
ID78673711	-0.8921	0.0743	-12.00	0.0000
hum	-0.0241	0.0015	-15.58	0.0000
wind_spd	-0.2067	0.0267	-7.74	0.0000
pressure	-0.0033	0.0005	-6.51	0.0000
daytypeWorkweek	-1.3537	0.1061	-12.75	0.0000
I(21 - temp)	-0.1033	0.0095	-10.89	0.0000
ID4866195:I(21 - temp)	0.0167	0.0045	3.68	0.0002
ID4887707:I(21 - temp)	0.0505	0.0045	11.15	0.0000
ID4962433:I(21 - temp)	0.0475	0.0045	10.49	0.0000
ID5037175:I(21 - temp)	0.0315	0.0045	6.97	0.0000
ID5093913:I(21 - temp)	0.0703	0.0045	15.52	0.0000
ID5093998:I(21 - temp)	0.0436	0.0045	9.62	0.0000
ID5140250:I(21 - temp)	0.0178	0.0045	3.92	0.0001
ID5325295:I(21 - temp)	0.0970	0.0045	21.42	0.0000
ID6392057:I(21 - temp)	0.0632	0.0045	13.89	0.0000
ID6392146:I(21 - temp)	0.0547	0.0045	12.07	0.0000
ID6392172:I(21 - temp)	0.0179	0.0051	3.53	0.0004
ID6393014:I(21 - temp)	0.0450	0.0045	9.93	0.0000
ID6540708:I(21 - temp)	0.0575	0.0045	12.71	0.0000
ID6567326:I(21 - temp)	0.0048	0.0045	1.07	0.2864
ID6618578:I(21 - temp)	0.0512	0.0045	11.31	0.0000
ID6618580:I(21 - temp)	0.0770	0.0045	17.00	0.0000
ID6627217:I(21 - temp)	0.0465	0.0045	10.27	0.0000
ID6627258:I(21 - temp)	0.0259	0.0045	5.72	0.0000
ID6627261:I(21 - temp)	0.0300	0.0045	6.63	0.0000
ID6627320:I(21 - temp)	0.0602	0.0045	13.30	0.0000
ID6681763:I(21 - temp)	0.0512	0.0045	11.32	0.0000
ID6681892:I(21 - temp)	0.0522	0.0045	11.52	0.0000
ID6681894:I(21 - temp)	0.0948	0.0045	20.95	0.0000
ID6790785:I(21 - temp)	-0.0096	0.0045	-2.11	0.0348
ID6790798:I(21 - temp)	0.0310	0.0045	6.83	0.0000
ID6842413:I(21 - temp)	0.0356	0.0045	7.87	0.0000
ID6842421:I(21 - temp)	0.0558	0.0045	12.32	0.0000
ID6842603:I(21 - temp)	0.0317	0.0045	7.00	0.0000
ID6842762:I(21 - temp)	0.0452	0.0045	9.98	0.0000
ID6921678:I(21 - temp)	0.0344	0.0045	7.60	0.0000
ID6940321:I(21 - temp)	0.0730	0.0045	16.14	0.0000
ID6940322:I(21 - temp)	0.0246	0.0045	5.44	0.0000
ID7072161:I(21 - temp)	0.0483	0.0045	10.67	0.0000
ID7072231:I(21 - temp)	0.0702	0.0045	15.50	0.0000
ID7072241:I(21 - temp)	0.0642	0.0045	14.17	0.0000
ID7072337:I(21 - temp)	0.0618	0.0045	13.65	0.0000
ID7183151:I(21 - temp)	0.0508	0.0045	11.22	0.0000
ID65005112:I(21 - temp)	0.0363	0.0045	8.01	0.0000
ID65012411:I(21 - temp)	0.0489	0.0045	10.80	0.0000
ID65014229:I(21 - temp)	0.0419	0.0045	9.25	0.0000
ID65014274:I(21 - temp)	0.0297	0.0045	6.57	0.0000
ID65052581:I(21 - temp)	0.0607	0.0045	13.40	0.0000
ID65063195:I(21 - temp)	0.0542	0.0045	11.97	0.0000
ID65063211:I(21 - temp)	0.0433	0.0045	9.53	0.0000
ID65063303:I(21 - temp)	0.0564	0.0045	12.47	0.0000
ID65067046:I(21 - temp)	0.0349	0.0045	7.70	0.0000
ID65118755:I(21 - temp)	0.0423	0.0047	9.07	0.0000
ID65118764:I(21 - temp)	0.0377	0.0046	8.19	0.0000
ID65118805:I(21 - temp)	0.0574	0.0045	12.69	0.0000
ID65118826:I(21 - temp)	0.0208	0.0045	4.59	0.0000
ID65118829:I(21 - temp)	0.0536	0.0045	11.83	0.0000
ID65118840:I(21 - temp)	0.0212	0.0045	4.69	0.0000
ID69001263:I(21 - temp)	0.0400	0.0045	8.83	0.0000
ID69001269:I(21 - temp)	0.0313	0.0045	6.91	0.0000
ID69089222:I(21 - temp)	0.0298	0.0045	6.58	0.0000
ID69250492:I(21 - temp)	0.0380	0.0045	8.40	0.0000
ID69429582:I(21 - temp)	0.0597	0.0045	13.18	0.0000
ID69469107:I(21 - temp)	0.0968	0.0045	21.38	0.0000
ID69478883:I(21 - temp)	0.0103	0.0045	2.28	0.0228
ID69518080:I(21 - temp)	0.0623	0.0045	13.76	0.0000
ID69518092:I(21 - temp)	0.0327	0.0045	7.22	0.0000
ID69580701:I(21 - temp)	0.0424	0.0045	9.35	0.0000
ID69585544:I(21 - temp)	0.0621	0.0045	13.71	0.0000
ID69585545:I(21 - temp)	0.0518	0.0045	11.45	0.0000
ID69611360:I(21 - temp)	0.0421	0.0045	9.30	0.0000

ID69652603:I(21 - temp)	0.0345	0.0045	7.62	0.0000
ID69688095:I(21 - temp)	0.0331	0.0045	7.30	0.0000
ID69749518:I(21 - temp)	0.0545	0.0045	12.04	0.0000
ID69861509:I(21 - temp)	0.0750	0.0045	16.57	0.0000
ID69999094:I(21 - temp)	0.0536	0.0045	11.83	0.0000
ID78082613:I(21 - temp)	0.0491	0.0045	10.85	0.0000
ID78138095:I(21 - temp)	0.0444	0.0045	9.80	0.0000
ID78443775:I(21 - temp)	0.0277	0.0046	6.07	0.0000
ID78673711:I(21 - temp)	0.0694	0.0045	15.34	0.0000
ID4866195:daytypeWorkweek	-0.0162	0.0433	-0.37	0.7087
ID4887707:daytypeWorkweek	-0.0539	0.0432	-1.25	0.2119
ID4962433:daytypeWorkweek	0.0123	0.0439	0.28	0.7788
ID5037175:daytypeWorkweek	-0.0427	0.0432	-0.99	0.3236
ID5093913:daytypeWorkweek	-0.0225	0.0432	-0.52	0.6031
ID5093998:daytypeWorkweek	-0.0406	0.0432	-0.94	0.3470
ID5140250:daytypeWorkweek	0.0138	0.0434	0.32	0.7504
ID5325295:daytypeWorkweek	0.1514	0.0432	3.50	0.0005
ID6392057:daytypeWorkweek	0.0675	0.0432	1.56	0.1188
ID6392146:daytypeWorkweek	0.0364	0.0432	0.84	0.4000
ID6392172:daytypeWorkweek	0.2155	0.0460	4.69	0.0000
ID6393014:daytypeWorkweek	0.0248	0.0432	0.57	0.5669
ID6540708:daytypeWorkweek	0.2100	0.0432	4.86	0.0000
ID6567326:daytypeWorkweek	-0.0285	0.0432	-0.66	0.5092
ID6618578:daytypeWorkweek	0.0030	0.0432	0.07	0.9442
ID6618580:daytypeWorkweek	-0.0468	0.0432	-1.08	0.2786
ID6627217:daytypeWorkweek	0.3052	0.0432	7.06	0.0000
ID6627258:daytypeWorkweek	0.2917	0.0432	6.75	0.0000
ID6627261:daytypeWorkweek	-0.0340	0.0432	-0.79	0.4316
ID6627320:daytypeWorkweek	0.2592	0.0432	6.00	0.0000
ID6681763:daytypeWorkweek	-0.0592	0.0432	-1.37	0.1705
ID6681892:daytypeWorkweek	-0.0408	0.0432	-0.94	0.3447
ID6681894:daytypeWorkweek	-0.0355	0.0432	-0.82	0.4109
ID6790785:daytypeWorkweek	0.0510	0.0433	1.18	0.2394
ID6790798:daytypeWorkweek	-0.0106	0.0433	-0.25	0.8061
ID6842413:daytypeWorkweek	0.0319	0.0432	0.74	0.4609
ID6842421:daytypeWorkweek	0.4088	0.0432	9.46	0.0000
ID6842603:daytypeWorkweek	0.0221	0.0432	0.51	0.6085
ID6842762:daytypeWorkweek	-0.0357	0.0432	-0.83	0.4082
ID6921678:daytypeWorkweek	-0.0545	0.0432	-1.26	0.2075
ID6940321:daytypeWorkweek	-0.0070	0.0432	-0.16	0.8705
ID6940322:daytypeWorkweek	0.1606	0.0432	3.72	0.0002
ID7072161:daytypeWorkweek	0.0833	0.0432	1.93	0.0538
ID7072231:daytypeWorkweek	-0.0561	0.0432	-1.30	0.1940
ID7072241:daytypeWorkweek	0.3897	0.0433	9.00	0.0000
ID7072337:daytypeWorkweek	0.1874	0.0432	4.34	0.0000
ID7183151:daytypeWorkweek	-0.0169	0.0432	-0.39	0.6963
ID65005112:daytypeWorkweek	-0.0580	0.0432	-1.34	0.1795
ID65012411:daytypeWorkweek	0.1104	0.0434	2.55	0.0109
ID65014229:daytypeWorkweek	-0.0086	0.0432	-0.20	0.8426
ID65014274:daytypeWorkweek	0.0748	0.0432	1.73	0.0836
ID65052581:daytypeWorkweek	0.1873	0.0432	4.33	0.0000
ID65063195:daytypeWorkweek	-0.0177	0.0432	-0.41	0.6818
ID65063211:daytypeWorkweek	0.1725	0.0436	3.96	0.0001
ID65063303:daytypeWorkweek	0.2966	0.0432	6.87	0.0000
ID65067046:daytypeWorkweek	0.0153	0.0433	0.35	0.7233
ID65118755:daytypeWorkweek	-0.0763	0.0445	-1.71	0.0866
ID65118764:daytypeWorkweek	0.1237	0.0441	2.80	0.0050
ID65118805:daytypeWorkweek	0.3134	0.0432	7.25	0.0000
ID65118826:daytypeWorkweek	-0.0442	0.0432	-1.02	0.3069
ID65118829:daytypeWorkweek	-0.0868	0.0432	-2.01	0.0446
ID65118840:daytypeWorkweek	-0.1046	0.0432	-2.42	0.0155
ID69001263:daytypeWorkweek	0.2916	0.0432	6.75	0.0000
ID69001269:daytypeWorkweek	-0.0235	0.0432	-0.54	0.5860
ID69089222:daytypeWorkweek	-0.0454	0.0432	-1.05	0.2935
ID69250492:daytypeWorkweek	-0.0914	0.0432	-2.11	0.0345
ID69429582:daytypeWorkweek	-0.0052	0.0432	-0.12	0.9036
ID69469107:daytypeWorkweek	0.0614	0.0432	1.42	0.1553
ID69478883:daytypeWorkweek	0.0699	0.0432	1.62	0.1058
ID69518080:daytypeWorkweek	-0.0337	0.0432	-0.78	0.4353
ID69518092:daytypeWorkweek	-0.0558	0.0432	-1.29	0.1969
ID69580701:daytypeWorkweek	-0.0373	0.0433	-0.86	0.3885
ID69585544:daytypeWorkweek	0.0924	0.0432	2.14	0.0326
ID69585545:daytypeWorkweek	-0.0278	0.0432	-0.64	0.5202

ID69611360:daytypeWorkweek	-0.0183	0.0432	-0.42	0.6720
ID69652603:daytypeWorkweek	0.0601	0.0432	1.39	0.1642
ID69688095:daytypeWorkweek	0.4486	0.0432	10.38	0.0000
ID69749518:daytypeWorkweek	-0.0400	0.0432	-0.93	0.3547
ID69861509:daytypeWorkweek	0.3821	0.0432	8.84	0.0000
ID69999094:daytypeWorkweek	-0.0270	0.0433	-0.62	0.5328
ID78082613:daytypeWorkweek	0.0749	0.0432	1.73	0.0829
ID78138095:daytypeWorkweek	0.0570	0.0432	1.32	0.1870
ID78443775:daytypeWorkweek	0.0875	0.0433	2.02	0.0435
ID78673711:daytypeWorkweek	0.0478	0.0432	1.11	0.2689
daytypeWorkweek:I(21 - temp)	0.1037	0.0081	12.86	0.0000
hum:wind_spd	0.0009	0.0001	9.87	0.0000
wind_spd:pressure	0.0001	0.0000	5.32	0.0000
hum:I(21 - temp)	0.0017	0.0001	15.56	0.0000
wind_spd:I(21 - temp)	0.0043	0.0006	7.71	0.0000
hum:daytypeWorkweek	0.0159	0.0013	12.30	0.0000
hum:wind_spd:I(21 - temp)	-0.0001	0.0000	-7.34	0.0000
hum:daytypeWorkweek:I(21 - temp)	-0.0011	0.0001	-11.44	0.0000

Table 16: Summary lm_4

6.D R Code

```r
############################################################################
## Case 2:                                                                ##
## HTK Case: Energy performance of buildings                              ##
############################################################################

# Authors: Begona Bolos Sierra , Laura Sans Comerma, Jorge Montalvo Arvizu


# Load Packages -----------------------------------------------------------

require("car")
require("tidyverse")
library("stringr")
require("readxl")
require("lubridate")
library("dplyr")

# Visualization packages
require("xtable")
require("ggpubr")
require("ggplot2")
require("ggExtra")
require("GGally")
require('ggcorrplot')
require("gridExtra")


# Load data ---------------------------------------------------------------

htk <- read_excel("~/Github/02441_Applied_Statistics/Case2/2_Data/HTK_building_data_share.xlsx")
load("~/Github/02441_Applied_Statistics/Case2/2_Data/WUndergroundHourly.RData")
files <- dir("~/Github/02441_Applied_Statistics/Case2/2_Data/meterdata", pattern="*.txt", full.names=TRUE)

# WUnderground ------------------------------------------------------------

summary(WG)

# Remove NA columns
data_0 <- Filter(function(x)!all(is.na(x)), WG)

# Check removed columns
setdiff(names(WG),names(data_0))

# Remove columns with fixed values
data <- Filter(function(x) length(unique(x))!=1, data_0)

# Check removed columns
setdiff(names(data_0),names(data))

# Change full date to short date
day <- data.frame(str_split_fixed(data$date, " ", 2))
day <- day[,-2]
data<- cbind(day, data)
data <- data[,-2]

# Check summary and structure of data
summary(data)
str(data)

# Factorize cond, and dir
data$cond <- factor(data$cond)
data$dir <- factor(data$dir)

# Sanity-check
str(data)

# Calculate mean value for continuous and mode for factor variables
# Create a mode function
getmode <- function(v) {
```

```
  uniqv <- na.omit(unique(v))
  uniqv[which.max(tabulate(match(v, uniqv)))]
}

# Create a data frame for the means and the modes
mean_mode <- cbind.data.frame(day)
# it has only unique values, remove repeated dates
mean_mode <- unique(mean_mode)

# get the column names
names <- colnames(data)

# Empty cells --> NA
data$cond[data$cond==""]  <- NA
data$dir [data$dir ==""]  <- NA

# Calculate the mean and mode for each colum of the df
for (i in 2:ncol(data)){
  if (is.numeric(data[,i]) == FALSE){
    values <- cbind.data.frame(data$day, data[,i])
    colnames(values) <- c("date","value")
    mode_value<- aggregate(values$value,list(values$date), getmode)
    mean_mode <- cbind.data.frame(mean_mode, mode_value)
  }
  if (is.numeric(data[,i]) == TRUE){
    values <- cbind.data.frame(data$day, data[,i])
    colnames(values) <- c("date","value")
    mean_value <- aggregate(values$value,list(values$date), mean)
    mean_mode <- cbind.data.frame(mean_mode, mean_value)
  }
}

# Erase duplicate dates
mean_mode <- mean_mode[!duplicated(as.list(mean_mode))]
# Erase an extra column
mean_mode <- mean_mode[,-1]
# change column names
colnames(mean_mode) <- names


# Meter ------------------------------------------------------------------------

# Read all data into a single dataframe
df_raw <- do.call(rbind, lapply(files, read.table, sep=";", dec= ","))
df_raw <- df_raw[,c(1,2,4)]

# Keep only columns 1, 2 and 4 ("ID", "Time" and "Reading")
names(df_raw) <- c("ID", "Time", "Reading")
df_raw$Reading <- as.numeric(df_raw$Reading)


# Exclude meters with less than 121 records
count_list <- count(df_raw, ID)
count_list <- count_list[count_list$n < 121,]

for (j in 1:nrow(count_list)){
  df_raw<- df_raw[!(df_raw$ID == count_list$ID[j]), ]
}

# Interpolation
library(stringr)
df_raw$Time <- as.POSIXct(strptime(df_raw$Time, format = "%d-%m-%Y %H.%M"))

# df split in IDs
split_data = split(df_raw, df_raw$ID)

# define the df with Date and ID
day_2 <- data.frame(str_split_fixed(df_raw$Time, " ", 2))
date <- day_2[,1]
df <- cbind(date, df_raw)
df <- df[,c(-3)]

# create new df
```

```r
df_new <- data.frame("ID"= 0,"date" = 0, "consumption" = 0)

# run for every ID and calculate the consumption x day
for (i in split_data){
  x <- i
  x <- x[order(x$Time),]
  readings <- x[,c(3)]
  ex <- x[,c(2)]
  day <- data.frame(str_split_fixed(x$Time, " ", 2))
  date <- day[,1]
  hour <- day[,2]
  plot(readings~ex, xlab="Date", ylab="Readings", col=("Red"))
  x.inter <- list()
  for (element in as.character(date)){
    dat <- as.POSIXct(paste(element, "23:59:00"), format="%Y-%m-%d %H:%M:%S")
    x.inter <- append(x.inter, dat)
  }
  inter.result <- approx(x = ex, y = readings, xout=x.inter)
  points(inter.result$x, inter.result$y, pch = 2)
  legend("topleft", legend = c("data", "interpolated"), pch = c(1,2), col=c("Red", "Black"))
  diff_buil <- diff(inter.result$y) # get the consumption
  # start adding results to the correct df
  date_2 <- date[1:120]
  id_2 <- x$ID[1:120]
  date_diff = data.frame(id_2,date_2, diff_buil)
  colnames(date_diff) <- c('ID','date','consumption')
  df_new <- rbind.data.frame(df_new, date_diff)
  date_diff = data.frame()
}
# Remove first row
df_new <-df_new[-1,]

# Merge with WU
colnames(mean_mode) <- c("date", "temp","dew_pt","hum","wind_spd","dir","vis","pressure","cond","fog","rain","snow")
merged_df <- merge(mean_mode, df_new, by = "date")
# Summary of the merged df
summary(merged_df)

# Load Data ----------------------------------------------------------------
# Load CampusNet Merged Data
df <- read_csv("~/Github/02441_Applied_Statistics/Case2/2_Data/merged_data.csv")
summary(df)
# Set new directory for output files
setwd("~/Github/02441_Applied_Statistics/Case2/4_Images")


# Analysis -----------------------------------------------------------------
# Inspect Data
str(df)
summary(df)
sum_df <- summary(df)
print(xtable(sum_df, type = "latex"), file = "summary_df.tex")

# Date to workweek and weekend, per month
df$date <- as.Date(df$date)
df$day <- weekdays(df$date)
df$week <- week(df$date)
df$daytype <- weekdays(df$date)
workweek <- c("Monday", "Tuesday", "Wednesday", "Thursday", "Friday")
weekend <- c("Saturday", "Sunday")
for (i in workweek) {
  df$daytype[df$day == i] <- "Workweek"
}
for (i in weekend) {
  df$daytype[df$day == i] <- "Weekend"
}

# Factorize variables
df$day <- factor(df$day)
df$daytype <- factor(df$daytype)
df$week <- factor(df$week)
```

```r
df$ID <- factor(df$ID)
df$dir <- factor(df$dir)
df$cond <- factor(df$cond)
df$date <- factor(df$date)
str(df)

# Removing direction, visibility, condition, fog, and rain
plot(fog~cond, df)
plot(rain~cond, df) #it doesn't seem that condition gives interpretable info
df <- df[,-c(8,9,11,12,13)]

# Split building type from HTK file
type <- data.frame(str_split_fixed(htk$Anvendelse, " ", 2))

# Get a building type df only
type_building <-type
type_building <- unique(type_building)
colnames(type_building) <- c("type", "name") # rename columns
type_building <- type_building[order(type_building$type),]
zero_type <- data.frame("000", "not defined") # add the row not defined for some
    buildings
names(zero_type) <- c("type", "name")
type_building <- rbind(type_building,zero_type)

# data frame with only type code
type <- type[,-2]
id_type <- cbind.data.frame(htk$MAlernr, type)  # merge ID and type
colnames(id_type) <- c("ID", "type") # rename columns

df_missing <- data.frame(setdiff(unique(df$ID),unique(htk$MAlernr)),rep("000",6))
colnames(df_missing) <- c("ID", "type") # rename columns
id_type <- rbind(id_type,df_missing)

# Table of building types for appendix
print(xtable(type_building, type = "latex"), file = "type_building.tex")

# Outlier investigation
outliers <- df[c(3282,3357),]
df <- df[-c(3282,3357),]
#plot(df$temp, df$consumption, type="p", col=df$ID, pch=19)
#plot(consumption~temp, subset(df, ID==78185925), pch=19, col=2)

# Now add new type column to the df
df$ID <- factor(df$ID)
df <- merge(df, id_type ,by="ID")

# Calculating insulation
lm_u <- lm(consumption~ID*I(21-temp),df)
Anova(lm_u)
summary(lm_u)
# Store in LaTex table
lm_u_a <- Anova(lm_u)
print(xtable(lm_u_a, type = "latex"), file = "lm_u_anova.tex")
lm_u_sum <- summary(lm_u)
print(xtable(lm_u_sum, type = "latex"), file = "lm_u_summary.tex")

df_u <- data.frame(data.frame(lm_u$coefficients)[c(84:166),],row.names=levels(df$ID))
colnames(df_u) <- "u"
df_u$u <- df_u$u+df_u$u[1]
df_u$u[1] <- df_u$u[1]/2
df_u$ID <- levels(df$ID)
u_sum <- summary(df_u)
print(xtable(u_sum, type = "latex"), file = "u_summary.tex")

# Check dew collinearity
#pairs(subset(df, select=c(4:6)))
cor.test(df$temp, df$dew_pt)
temp_interval <- cut(df$temp, 4) # divide temperature in intervals to colour

png(filename="corr_dewpt.png", width=1750, height=1750, res=300)
p00 <- ggpairs(df[, c(4,5,6)], aes(colour = temp_interval),upper = list(continuous = wrap
    ("cor", size = 2.5)))
for(i in 1:p00$nrow) {
```

```
  for(j in 1:p00$ncol){
    p00[i,j] <- p00[i,j] +
      scale_fill_manual(values = rainbow(4)) +
      scale_color_manual(values = rainbow(4))
  }
}
p00 + theme_classic()
dev.off()

df <- df[,-5] # Correlation very high at 0.95, thus remove dew_pt

# Data visualization ---------------------------------------------------------
# Pairs plot
temp_interval <- cut(df$temp, 4) # divide temperature in intervals to colour
df$month <- as.numeric(as.character(as.Date(df$date, format = "%Y-%m-%d"), format="%m"))
df$month <- factor(df$month)
png(filename="pair_plot_whole.png", width=2500, height=1750, res=300)
p0 <- ggpairs(df[, c(3,4,5,6,7,10)], aes(colour = temp_interval),upper = list(continuous = wrap("cor", size = 2.5)))
for(i in 1:p0$nrow) {
  for(j in 1:p0$ncol){
    p0[i,j] <- p0[i,j] +
      scale_fill_manual(values = rainbow(4))+
      scale_color_manual(values =  rainbow(4))
  }
}
p0 + theme_classic() + theme(panel.border = element_rect(colour = "black", fill=NA, size=1))
dev.off()
#pairs(subset(df, select=c(3:8), col=df$ID))


# Consumption - Temp by type
png(filename="cons-temp.png", width=2250, height=1050, res=300)
p1 <- ggplot() + geom_point(data=df, aes(x=(21-temp), y=consumption, col= type))
p1 + scale_color_manual(guide = guide_legend(),values=rainbow(26), name="Building type") + xlab("Temperature C") + ylab("Consumption") + theme_classic() + theme(plot.margin = margin(1, 1,0.1, 1, "cm"),panel.border = element_rect(colour = "black", fill=NA, size=1))
dev.off()

#Building plot type 032 / ID 78185925
# select type and make also building subset
type032 <- subset(df, col=df$type, type =="032", select = ID:type)
id78185925 <- subset(type032, col=type032$ID, ID =="78185925", select = ID:type)

# Outlier investigation
#outliers <- df[c(3282,3357),]
#plot(df$temp, df$consumption, type="p", col=df$ID, pch=19)
#plot(consumption~temp, subset(df, ID==78185925), pch=19, col=2)

png(filename="78185925.png", width=1750, height=1050, res=300)
p2 <- ggplot() + geom_point(data=type032, aes(x=(21-temp), y=consumption, col="ID 78185925")) + geom_point(data=id78185925, aes(x=(21-temp), y=consumption, col="Other")) + geom_point(data=outliers, aes(x=(21-temp), y=consumption, col="Outliers"))
p2 + scale_color_manual(guide = guide_legend(),values=c("#808080","#00FF00FF", "#FF0000FF" ), name="Sports and swimming (type = 032)") + xlab("Temperature C") + ylab("Consumption") + theme_classic() + theme(legend.position ="bottom",legend.box = 'horizontal',panel.border = element_rect(colour = "black", fill=NA, size=1) )
dev.off()

# After visualizing, remove outliers
#df <- df[-c(3282,3357),] # Removing outliers 3282 and 3357

# Plot consumption sum vs type of building
# aggregate consumption SUM
consumption_sum <- aggregate(df$consumption,list(df$type), sum)
colnames(consumption_sum) <- c("type", "cons") # rename columns

png(filename="consum_type.png", width=2050, height=1050, res=300)
p3 <- ggplot(data=consumption_sum, aes(x=type, y=cons, fill=type)) + geom_bar(stat="identity",show.legend = FALSE)
p3 + scale_fill_manual(values=rainbow(25)) + xlab("Type of building") + ylab("Sum of
```

```
consumption") + theme_classic() + theme(panel.border = element_rect(colour = "black",
 fill=NA, size=1))
dev.off()

# plot Consumption - date
cons_date_sum <- aggregate(df$consumption, list(id11= df$type, id12= df$date), sum)
colnames(cons_date_sum) <- c("type", "date", "cons") # rename columns
cons_date_sum$date <- as.numeric(cons_date_sum$date)
cons_date_sum$rank <- rank(cons_date_sum$date)

png(filename="consum_type_date.png", width=2050, height=1050, res=300)
p4 <- ggplot(data=cons_date_sum, aes(x=date, y=cons, col=type)) + geom_line()
p4 + scale_color_manual(values=rainbow(25)) + xlab("Date") + ylab("Consumption") + theme_
    classic() +theme(plot.margin = margin(1, 1,0.1, 1, "cm"),panel.border = element_rect(
    colour = "black", fill=NA, size=1))
dev.off()


# Test simple model
lm1 <- lm(consumption~(ID+week)*I(21-temp), df)
Anova(lm1)
summary(lm1) # Comment in the report

# Store in latex table
lm1_a <- Anova(lm1)
print(xtable(lm1_a, type = "latex"), file = "lm1_anova.tex")
lm1_sum <- summary(lm1)
print(xtable(lm1_sum, type = "latex"), file = "lm1_summary.tex")

# Residual plots lm1
#plot1 <- qplot(1)
#plot2 <- qplot(1)
#grid.arrange(plot1, plot2, ncol=2)
png(filename="lm1_14residuals.png", width=2050, height=1750, res=300)
par(mfrow=c(2,2))
plot(lm1, col=df$ID, pch=19,cex = 0.6)
dev.off()

png(filename="lm1_residuals.png", width=2050, height=1750, res=300)
par(mfrow=c(1,1))
plot(lm1$residuals~I(21-temp),df, col=df$ID, pch=19,cex = 0.6, ylab="lm1 Residuals")
dev.off()
# residuals have different variance per building


# Calculate residuals variance per building
df_variance <- aggregate(unname(lm1$residuals), list(df$ID), var)
colnames(df_variance) <- c("ID","variance")
df <- merge(df, df_variance, by = "ID")

# Calculate mean consumption by building
df_mean <- aggregate(df$consumption, list(df$ID), mean)
colnames(df_mean) <- c("ID","mean")
df <- merge(df, df_mean, by = "ID")

# Plot variance per building against mean consumption per building
png(filename="variance_vs_mean_building.png", width=1750, height=1750, res=300)
plot(variance ~ mean, df, col=df$ID, pch=19, cex = 0.6,ylab="Variance", xlab="Mean")
lines(seq(-1,4,length.out=100),rep(0.03,100), col=2) # we propose a threshold of 0.03
    variance to identify odd buildings
dev.off()

# Data frame of odd buildings (10) at variance greater than 0.03
df_oddvariance1 <- data.frame(subset(df, variance>0.03))
png(filename="oddbuildings10_var003.png", width=1750, height=1750, res=300)
plot(consumption~I(21-temp), type="p", df_oddvariance1, col=df$ID, pch=19,cex=0.6, ylab="
    Consumption")
dev.off()
df_oddvariance1 <- droplevels(df_oddvariance1)
oddvariance1 <- unique(df_oddvariance1$ID)

# Plot variance per building against mean consumption per building at lower variance
png(filename="var_mean_zoom.png", width=1750, height=1750, res=300)
```

```r
plot(variance ~ mean, df,col=df$ID, cex=0.6, pch=19, xlim=c(0,1), ylim=c(0,0.03),ylab="Variance", xlab="Mean")
lines(seq(0,2,length.out=100),rep(0.0075,100), col=2) # 10+3 more outliers with a proposed threshold of 0.0075 variance to identify odd buildings
dev.off()

# Data frame of odd buildings (13) at variance greater than 0.0075
df_oddvariance2 <- data.frame(subset(df, variance>0.0075))
png(filename="oddbuildings13_var00075.png", width=1750, height=1750, res=300)
plot(consumption~I(21-temp), type="p", df_oddvariance2, col=df$ID, cex=0.6, pch=19, ylab="Consumption")
dev.off()
df_oddvariance2 <- droplevels(df_oddvariance2)
oddvariance2 <- unique(df_oddvariance2$ID)

# Plot mean consumption vs. ID
png(filename="mean_ID.png", width=1750, height=1550, res=300)
plot(mean~ID, df, col=df$ID, ylab="Consumption")
lines(df$ID,rep(0.5,length(df$ID)), col=2) # We propose a threshold at mean of 0.5 to "label" odd-buildings
dev.off()

# Data frame of odd buildings (16) at mean greater than 0.5
df_oddmean <- data.frame(subset(df, mean>0.5))
png(filename="oddbuildings16_mean05.png", width=1750, height=1750, res=300)
plot(consumption~I(21-temp), df_oddmean, col=ID, pch=19, ylab="Consumption", cex=0.6)
dev.off()
df_oddmean <- droplevels(df_oddmean)
oddmeanID5 <- unique(df_oddmean$ID)

# Create data frames to check
df_minus13 <- data.frame(subset(df, variance<0.0075))
df_minus13 <- droplevels(df_minus13)

# Plot consumption against date of odd variance buildings
png(filename="odd_var_vs_date.png", width=1750, height=1550, res=300)
par(mfrow=c(1,1), oma =c(1,2,0,4.5), mar = c(3,2,2,2))
plot(consumption~as.numeric(date), df_oddvariance2, col=ID, pch=19, cex=0.6, xlab="Date", ylab="Consumption")
par(xpd=NA)
legend(x=130, y=4.80, legend=levels(df_oddvariance2$ID), pch=19, col=unique(df_oddvariance2$ID), cex=0.8)
dev.off()
# There are some weird buildings and outliers that we could check

# Plot consumption against date of normal variance buildings
png(filename="normal_var_vs_date.png", width=1750, height=1750, res=300)
par(mfrow=c(1,1))
plot(consumption~as.numeric(date), df_minus13, col=ID, pch=19, cex=0.6, xlab="Temperature C", ylab="Consumption")
dev.off()
# There are some weird buildings and outliers that we could check

### Same as before but with temperature

# Plot consumption against temp of odd variance buildings
png(filename="odd_var_vs_temp.png", width=1750, height=1550, res=300)
par(mfrow=c(1,1), oma =c(0,0,0,4.5), mar = c(5,5,2,2))
plot(y=df_oddvariance2$consumption, x=(21-df_oddvariance2$temp), col=df_oddvariance2$ID, pch=19, cex=0.6, xlab="Temperature ", ylab="Consumption")
par(xpd=NA)
legend(x=25, y=5.2, legend=levels(df_oddvariance2$ID), pch=19, col=unique(df_oddvariance2$ID), cex=0.8)
dev.off()
# There are some weird buildings and outliers that we could check

# Plot consumption against date of normal variance buildings
png(filename="normal_var_vs_temp.png", width=1750, height=1750, res=300)
par(mfrow=c(1,1))
plot(y=df_minus13$consumption, x=(21-df_minus13$temp), col=df_minus13$ID, pch=19, cex=0.6, xlab="Temperature C", ylab="Consumption")
dev.off()
# There are some weird buildings and outliers that we could check
```

```r

# Plot each building with odd variance
par(mfrow=c(3,5))
for (i in unique(df_oddvariance2$ID)) {
  plot(consumption~as.numeric(date),subset(df,ID==i), pch=19, col=ID, main=paste("ID: ",i), xlab="date")
}
# Here we can identify some weird behaving buildings like 69478883 and 69999051, also remove some outliers and adjust factor to day or week or month
# the consumption has some peaks during the metered period, how can we 'tell' our statistical model to adjust for this

# Plot each building with odd variance per day
par(mfrow=c(3,5))
for (i in unique(df_oddvariance2$ID)) {
  set1 <- subset(df,ID==i)
  plot(consumption~as.numeric(date),set1, pch=19, col=ID, main=paste("ID: ",i), type="n", xlab="date")
  z <- 1
  for (j in unique(df_oddvariance2$day)) {
    set2 <- subset(set1,day==j)
    points(consumption~as.numeric(date),set2, pch=19, col=z)
    z <- z+1
  }
}
# per day doesn't really give us a real difference

# Plot each building with odd variance per day type
par(mfrow=c(3,5))
for (i in unique(df_oddvariance2$ID)) {
  set1 <- subset(df,ID==i)
  plot(consumption~as.numeric(date), set1, pch=19, col=ID, main=paste("ID: ",i), type="n", xlab="date")
  z <- 1
  for (j in unique(df_oddvariance2$daytype)) {
    set2 <- subset(set1,daytype==j)
    points(consumption~as.numeric(date), set2, pch=19, col=z)
    z <- z+1
  }
}
# Doesn't really fixes the peaks but helps to identify that maybe per week is the best factor

# Plot each building with odd variance per week
par(mfrow=c(1,1))
for (i in unique(df_oddvariance2$ID)) {
  set1 <- subset(df,ID==4939509)
  plot(consumption~as.numeric(date), set1, pch=19, col=ID, main=paste("ID: ",i), type="n", xlab="date")
  z <- 1
  for (j in unique(df_oddvariance2$week)) {
    set2 <- subset(set1,week==j)
    points(consumption~as.numeric(date), set2, pch=19, col=z)
    z <- z+1
  }
}
# Great fit for weekly peaks
# two outliers and two odd buildings in this model




# Adjust consumption to eliminate 'size' of buildings in the model
df$adjconsumption <- df$consumption/df$mean

# Set final data frame
df_model <- df[,-c(2,3,8,10,11,12)]



# Full model with interactions (with adjusted consumption and date as week)
lm2 <- step(lm(adjconsumption~.-temp+I(21-temp), df_model), scope=~.^3, k=log(nrow(df_model)), trace=FALSE)
```

```r
anova2 <- Anova(lm2)
par(mfrow=c(2,2))
plot(lm2, col=df$ID, pch=19)
sum2 <- summary(lm2, correlation=TRUE)
corr2 <- data.frame(sum2$correlation)



# Clean df variance and mean again
df <- df[,-c(11,12)]

# Calculate residuals variance per building
df_variance <- aggregate(unname(lm2$residuals), list(df$ID), var)
colnames(df_variance) <- c("ID","variance")
df <- merge(df, df_variance, by = "ID")

# Calculate mean consumption by building
df_mean <- aggregate(df$consumption, list(df$ID), mean)
colnames(df_mean) <- c("ID","mean")
df <- merge(df, df_mean, by = "ID")

# Plot variance per building against mean consumption per building
par(mfrow=c(1,1))
plot(variance ~ mean, df, col=df$ID, pch=19)
lines(seq(-1,4,length.out=100),rep(0.1,100), col=2) # we propose a threshold of 0.1
    variance to identify odd buildings

# Data frame of odd buildings (10) at variance greater than 0.1
df_oddvariance <- data.frame(subset(df, variance>0.1))
plot(consumption~I(21-temp), type="p", df_oddvariance, col=ID, pch=19)
df_oddvariance <- droplevels(df_oddvariance)
oddvariance <- unique(df_oddvariance$ID)





# Create data frames to check
df_minus6 <- data.frame(subset(df, variance<0.1))
df_minus6 <- droplevels(df_minus6)





# Plot consumption against date of odd variance buildings
par(mfrow=c(1,1))
plot(consumption~as.numeric(date), df_oddvariance, col=ID, pch=19)
legend("topleft", legend=levels(df_oddvariance$ID), pch=19, col=unique(df_oddvariance$ID)
    , cex=0.8)
# There are some weird buildings and outliers that we could check

# Plot consumption against date of normal variance buildings
plot(consumption~as.numeric(date), df_minus6, col=ID, pch=19)

# Plot each building with odd variance per day
par(mfrow=c(2,3))
for (i in unique(df_oddvariance$ID)) {
  set1 <- subset(df,ID==i)
  plot(consumption~as.numeric(date),set1, pch=19, col=ID, main=paste("ID: ",i), type="n",
      xlab="date")
  legend("topleft", legend=levels(df_oddvariance$day), pch=19, col=unique(df_oddvariance$
      day))
  z <- 1
  for (j in unique(df_oddvariance$day)) {
    set2 <- subset(set1,day==j)
    points(consumption~as.numeric(date),set2, pch=19, col=z)
    z <- z+1
  }
}
```

```r
# Remove outliers

par(mfrow=c(1,1))

plot(consumption~as.numeric(date),subset(df,ID==6392172), col=day, pch=19)
legend("topleft", legend=levels(df$day), pch=19, col=unique(df$day))

plot(consumption~as.numeric(date),data=subset(df,ID==65118755), col=day, pch=19)
legend("topleft", legend=levels(df$day), pch=19, col=unique(df$day))

plot(consumption~as.numeric(date),subset(df,ID==65118764), col=day, pch=19)
legend("topleft", legend=levels(df$day), pch=19, col=unique(df$day))

outliers <- c
    (1558,1552,1544,1611,1592,1568,1600,1629,1612,1603,3061,3045,3065,2971,2961,2955,3044,2995,3034,3055

df <- df[-outliers,]

# Remove zeros (because of shutting down or starting consumption later)
df <- subset(df, consumption!=0)

# Remove odd buildings
df <- subset(df, ID!=65118812)
df <- subset(df, ID!=69999051)
df <- droplevels(df)

# Remove september observations
df$date <- as.Date(df$date)

png(filename="september.png", width=1700, height=1700, res=300)
par(mfrow=c(1,1))
plot(consumption~date,subset(df,date<"2018-10-01"), pch=19, col=ID, ylab="Consumption",
    xlab="Date")
dev.off()


par(mfrow=c(1,1))
plot(consumption~date,subset(df,date<"2018-10-01"), pch=19, col=ID, ylab="Consumption",
    xlab="Date")
df <- subset(df,date>"2018-09-24")
df$date <- factor(df$date)

# Set final data frame
df_model <- df[,-c(2,3,8,12,13)]




# Full model with interactions
lm3 <- step(lm(adjconsumption~.-temp+I(21-temp), df_model), scope=~.^3, k=log(nrow(df_
    model)), trace=FALSE)
anova3 <- Anova(lm3)
par(mfrow=c(2,2))
plot(lm3, col=df$ID, pch=19)
sum3 <- summary(lm3, correlation=TRUE)
corr3 <- data.frame(sum3$correlation)






# Clean df variance and mean again
df <- df[,-c(12,13)]

# Calculate residuals variance per building
df_variance <- aggregate(unname(lm3$residuals), list(df$ID), var)
colnames(df_variance) <- c("ID","variance")
df <- merge(df, df_variance, by = "ID")

# Calculate mean consumption by building
df_mean <- aggregate(df$consumption, list(df$ID), mean)
```

```r
colnames(df_mean) <- c("ID","mean")
df <- merge(df, df_mean, by = "ID")

# Plot variance per building against mean consumption per building
par(mfrow=c(1,1))
plot(variance ~ mean, df, col=df$ID, pch=19)
lines(seq(-1,4,length.out=100),rep(0.03,100), col=2) # we propose a threshold of 0.4 variance to identify odd buildings


# Data frame of odd buildings (10) at variance greater than 0.1
df_oddvariance <- data.frame(subset(df, variance >0.03))
plot(consumption~I(21-temp), type="p", df_oddvariance, col=ID, pch=19)
df_oddvariance <- droplevels(df_oddvariance)
oddvariance <- unique(df_oddvariance$ID)





# Create data frames to check
df_minus17 <- data.frame(subset(df, variance <0.03))
df_minus17 <- droplevels(df_minus17)





# Plot consumption against date of odd variance buildings
par(mfrow=c(1,1))
plot(consumption~as.numeric(date), df_oddvariance, col=ID, pch=19)
legend("topleft", legend=levels(df_oddvariance$ID), pch=19, col=unique(df_oddvariance$ID), cex=0.8)
# There are some weird buildings and outliers that we could check

# Plot consumption against date of normal variance buildings
plot(consumption~as.numeric(date), df_minus19, col=ID, pch=19)

# Plot each building with odd variance per day type
par(mfrow=c(4,5))
for (i in unique(df_oddvariance$ID)) {
  set1 <- subset(df,ID==i)
  plot(consumption~as.numeric(date), set1, pch=19, col=ID, main=paste("ID: ",i), type="n", xlab="date")
  z <- 1
  for (j in unique(df_oddvariance$daytype)) {
    set2 <- subset(set1,daytype==j)
    points(consumption~as.numeric(date), set2, pch=19, col=z)
    z <- z+1
  }
}




# Remove outliers

par(mfrow=c(1,1))

plot(consumption~as.numeric(date),subset(df,ID==4962433), col=day, pch=19)
legend("topleft", legend=levels(df$day), pch=19, col=unique(df$day))

plot(consumption~as.numeric(date),subset(df,ID==6790785), col=day, pch=19)
legend("topleft", legend=levels(df$day), pch=19, col=unique(df$day))

plot(consumption~as.numeric(date),subset(df,ID==7072241), col=day, pch=19)
legend("topleft", legend=levels(df$day), pch=19, col=unique(df$day))

plot(consumption~as.numeric(date),subset(df,ID==6392057), col=day, pch=19)
legend("topleft", legend=levels(df$day), pch=19, col=unique(df$day))

plot(consumption~as.numeric(date),subset(df,ID==4866195), col=day, pch=19)
legend("topleft", legend=levels(df$day), pch=19, col=unique(df$day))
```

```
plot(consumption~as.numeric(date),subset(df,ID==5140250), col=day, pch=19)
legend("topleft", legend=levels(df$day), pch=19, col=unique(df$day))

plot(consumption~as.numeric(date),data=subset(df,ID==78443775), col=day, pch=19)
legend("topleft", legend=levels(df$day), pch=19, col=unique(df$day))

plot(consumption~as.numeric(date),subset(df,ID==65012411), col=day, pch=19)
legend("topleft", legend=levels(df$day), pch=19, col=unique(df$day))

plot(consumption~as.numeric(date),subset(df,ID==65067046), col=day, pch=19)
legend("topleft", legend=levels(df$day), pch=19, col=unique(df$day))

outliers <- c(531,4121,6905,930,915,303,1105,7473,7462,1641,1683,2907,2968,5775,4149,6663,2124,2260)

df <- df[-outliers,]

# Remove odd buildings
df <- subset(df, ID!=4529799)
df <- subset(df, ID!=4529800)
df <- subset(df, ID!=6393013)
df <- subset(df, ID!=69652588)
df <- subset(df, ID!=78185925)
df <- subset(df, ID!=65118848)
df <- droplevels(df)



# Set final data frame
df_model <- df[,-c(2,3,8,9,12,13)]



# Final model
lm4 <- step(lm(adjconsumption~.-temp+I(21-temp), df_model), scope=~.^3, k=log(nrow(df_model)), trace=FALSE)
anova4 <- Anova(lm4)
par(mfrow=c(2,2))
plot(lm4, col=df$ID, pch=19)

png(filename="lm4_residuals.png", width=1700, height=1500, res=300)
par(mfrow=c(2,2))
plot(lm4, col=df$ID, pch=19)
dev.off()

print(xtable(anova4,type="latex"),file="anova4.tex")

sum4 <- summary(lm4, correlation=TRUE)
print(xtable(sum4,type="latex"),file="sum4.tex")

# Get betas
coef4 <- data.frame(sum4$coefficients)
alfa <- coef4[1:75,1]
coef4 <- coef4[80:154,1]
A <- diag(75)
A[,1] <- 1
beta <- A  coef4
alfa <- A  alfa

# Get standard error
corr4 <- sum4$correlation
corr4 <- corr4[80:154,80:154]

cov4 <- sum4$cov.unscaled
cov4 <- cov4[80:154,80:154]

sig4 <- sum4$sigma^2
var_beta <- A  cov4  t(A) * sig4

se <- sqrt(diag(var_beta))
```

```
CI_up <- beta+qt(0.975,lm4$df.residual)*se
CI_down <- beta+qt(0.025,lm4$df.residual)*se

df_final <- data.frame(unique(df$ID),beta,se,CI_up,CI_down)
colnames(df_final) <- c("ID","Beta", "Std. Error","CI_Up", "CI_Down")

df_final2 <- df_final[order(df_final$Beta),]
df_final2 <- df_final2[1:5,]

df_final3 <- df_final[order(df_final$`Std. Error`),]
df_final3 <- df_final3[1:5,]

print(xtable(df_final,type="latex"),file="df_final.tex")
print(xtable(df_final2,type="latex"),file="df_final2.tex")
print(xtable(df_final3,type="latex"),file="df_final3.tex")
```

References

[1] Wikipedia., "World energy consumption."

[2] N. E. R. 2014, "Nordic energy research. source:."

[3] R. Paradis, "Retrofitting existing buildings to improve sustainability and energy performance.."