City of New York LL84
Data Analysis & Quality Assessment

FINAL

David Hsu
Assistant Professor
University of Pennsylvania
hsuyd@design.upenn.edu

March 14, 2012

Contents

Chapter 1

Introduction

This report is the first analysis of the City of New York's benchmarking data collected under Local Law 84 (which will be referred to, in this report, as the NYC data and LL84). As part of the City of New York's Greater and Greener Buildings Plan and PlaNYC 2030, NYC LL 84 was passed in 2009 in order to transform the real estate and energy efficiency market. By requiring building owners to gather energy data and obtain a standard building rating score – and then publicly disclose this score – this score allows current owners to better manage their properties by benchmarking their performance against other buildings, and also allows prospective buyers of real estate to evaluate the future benefit of energy efficient buildings, as well as the opportunity for avoiding future costs by making timely investments in energy efficiency.

Starting in August 2011, LL 84 required owners of all commercial and multifamily buildings over 50,000 square feet to submit their energy information to a online benchmarking tool. Due to the extremely high concentration of buildings and people in New York City, over 15,000 properties and 22,000 buildings are expected to be above this threshold. These buildings constitute more than 1.5 billion square feet and roughly 55% of all energy used in New York City.

In aggregate, the NYC data is the largest and most current database of observed building energy performance at any scale. It also uniquely describes energy usage across buildings for the nation's largest city, which is particularly important since the building sector constitutes the largest single sector of energy use in New York City. In comparison, the most recent U.S. CBECS survey constitutes only half as many buildings over all 50 states. Furthermore, since the data was collected through the U.S. Environmental Protection Agency's Portfolio Manager tool, it also includes critical information about building occupancy, space and energy use characteristics which can be used to infer relationships between building characteristics and observed energy performance.

Although the NYC data is an unprecedented opportunity for building energy research, there are a number of key questions that need to be answered about the NYC data. First, because the data is self-reported by building owners, it is not clear how consistently the data has been gathered to common standards. Second, the City of New York has a strong interest in verifying and improving the quality of the data for public use. Third, the NYC data requires checking, cleaning, and further standardization in order to be used in any future research and analysis; giving feedback to owners and consultants will improve the annual collection of the benchmarking data. Only after these key steps, the NYC data can be used to better inform the design of existing and future environmental policies.

This report is the result of extensive collaboration with the City of New York's Office of Long-Term Planning and Sustainability (OLTPS), in order to answer some of these questions about the NYC data. Any information about that could be used to identify the location of an individual

building was removed from the NYC data before handing over to me for analysis. In addition, a Memorandum of Agreement was signed by the University of Pennsylvania and OLTPS to protect the privacy of the NYC data and to allow the publication of subsequent academic research.

The following chapters describe:

- Chapter 2 describes the source, cleaning, and an assessment of the quality of the NYC data
- Chapter 3, immediately following, provides a stand-alone description of the overall building population from the remaining data
- Chapter 4 provides summaries of each various building types within the NYC data, and trends in building typology and observed energy performance

Acknowledgements Many thanks to Laurie Kerr, Hilary Beber, and Donna Hope for organizing this collaboration and enabling access to the NYC data and other data sources. In addition, thanks to Constantine Kontakosta, Adam Hinge, and Alexandra Sullivan for regular discussions of the data and analysis. Finally, many thanks for the assistance of the Greater Philadelphia Innovation Cluster (GPIC) in supporting this research.

Chapter 2

Data

This chapter describes the original NYC data, comprised of 10,016 buildings, and how it has been cleaned and processed for further analysis. These cleaning steps result in a partially cleaned dataset of 8,456 buildings, and a more fully cleaned NYC dataset of 7,401 remaining buildings. Section 2.1 describes the data collection process, as well as the additional information used for the purposes of comparison and validation. Section 2.2 describes the steps used to remove identified errors or omissions in the data. Section 2.2.2 is a quality assessment based on the plausibility of observed distributions within the NYC data itself. Section 2.4 is a quality assessment based on comparison to other datasets or validated subsamples.

2.1 Sources

LL 84 requires owners of buildings to submit information to the U.S. Environmental Protection Agency's EnergyStar Portfolio Manager (PM) tool in order to obtain a benchmarking rating. As such, this includes information about the overall building size, the distribution of space and energy uses within the building, as well as calculations of on-site and source (primary) use and greenhouse gas emissions in the building (the entire dataset will be referred to as the NYC data, but when examining the data in individual fields, I will refer to them as from PM).

The quality of the NYC data was assessed by comparing it to several other sources of information. In particular, the NYC data was joined to the City of New York's Primary Land Use Tax Lot Output (PLUTO) database which includes information about building and lot size, zoning, building age and most recent renovations, as well as assessed tax value[1]. This was joined to the individual buildings in the NYC data by geocodes.

2.2 Cleaning

2.2.1 Initial Cleaning

Initial cleaning reduced the original NYC data from 10,016 to 8,456 buildings. Table 2.1 specifies the steps were taken to remove buildings with data errors that could be identified *a priori*, such as missing or implausible data. A single specific building records that may have multiple errors, but the resulting 8,456 cleaned building records do not have any of the above errors listed in the table.

[1]For more information, see `http://www.nyc.gov/html/dcp/pdf/bytes/plutouserguide.pdf` or `http://www.nyc.gov/html/dcp/pdf/bytes/plutolayout.pdf`

Cleaning step	Bldgs Left
Original dataset	10,016
(-) not in New York State	10,010
(-) duplicate entries (older records for same building IDs)	9,655
(-) minor building types (CBECS:Other and less than 10)	9,599
(-) not in New York zip codes	9,436
(-) not in New York City five counties	9,390
(-) buildings with no energy use reported	8,468
(-) buildings with no floor space reported	8,456
(-) buildings with EUI below 5 or above 1,000 kBtu/psf	8,242
(-) remove top and bottom 5% of EUIs	7,401

Table 2.1: Tally of Remaining Buildings After Each Cleaning Step

2.2.2 Further Cleaning

Another way to determine the data quality is to internally compare from within the NYC data. Similar fields from different databases can be examined to check whether data is consistently reported, as in the reported building sizes. The distributions of energy data, such as in Energy Use Intensity (EUI), can be examined.

Reported Building Sizes Figure 2.1 show the absolute and percentage difference, respectively, of the different areas as reported by PM and PLUTO. Despite the reported differences and any possible inaccuracies this may introduce, these buildings were kept in the cleaned data because it could not be determined which field is usually "correct".

Energy Use Intensity One of the key metrics for further analysis and comparison is the Energy Use Intensity, or EUI, which is calculated as the total energy use of the building divided by floor area in units of kBTU per square foot. Source EUIs were used throughout. Since source EUI allows both a building size and energy use comparison, is was particularly important to remove any buildings with implausible source EUI figures.

First, any buildings with source EUIs less than 5 or more than 1,000 kBTU per square foot were removed from the dataset. Either extreme was considered unrealistic on the basis of engineering knowledge. Second, the distribution of source EUI for each of the major facility types was examined. Extreme values relative to the overall distribution were then removed. Table 2.2 shows the quantiles for source EUI by facility type, with the median at 50% and the top and bottom 5% showing extreme and implausible source EUI figures.

Removing the extreme values below 5 and above 1000 kBTU per square foot removed an additional 214 buildings, leaving 8,242 buildings. Removing 10% of the data based on the source EUI quantiles reduced further the total number of buildings down to 7,401. The overall dataset is summarized by building type in Table 2.2 and further described in Chapter 3.

2.3 Consultant Analysis

Consultants were used extensively by building owners to perform benchmarking and report data. The original dataset features 204 unique consulting firms, but by eliminating near-duplicate names,

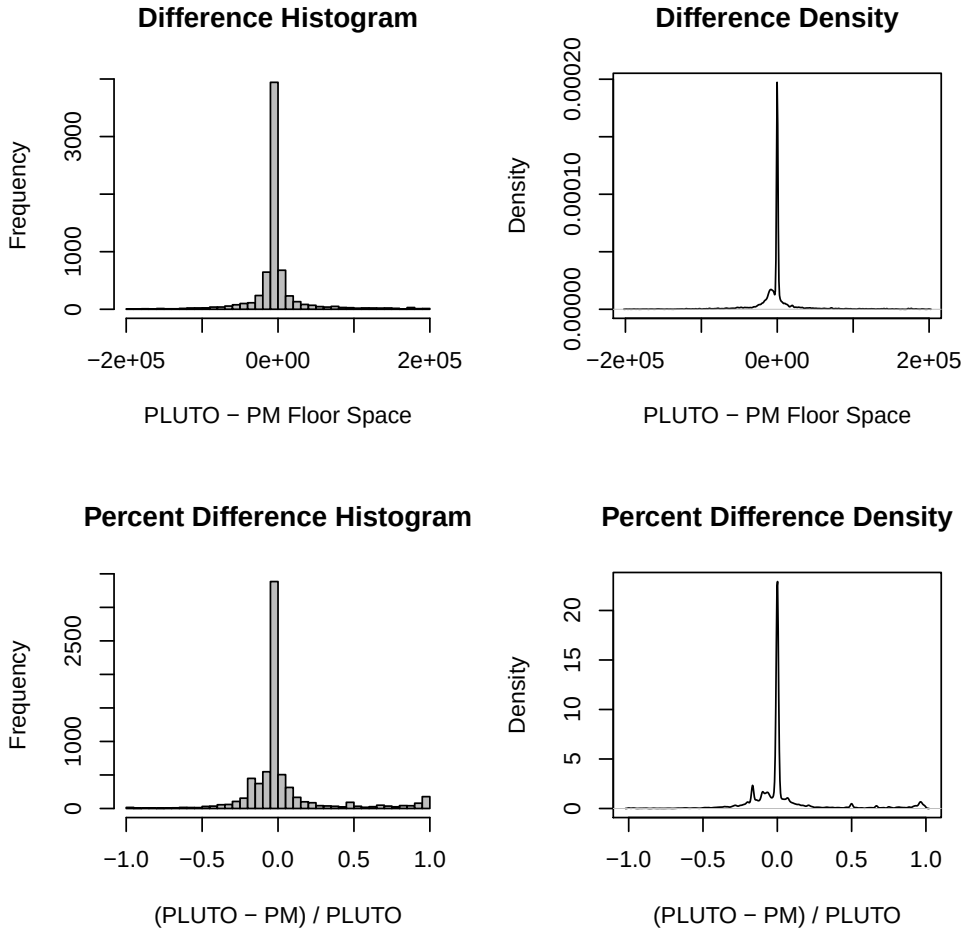


Figure 2.1: Histogram and density plot of the percentage divergence in building areas reported in the NYC PLUTO database and in EPA-PM. 0% report the same PLUTO and PM floor space. Plot idea from Constantine Kontokosta.

this is reduced to 157 unique consulting firms[2].

Inspecting the NYC data revealed that a relatively small number of consultants were responsible for a large proportion of the benchmarked buildings. This has important implications for a process- or supplier-based view of the production of the NYC data, and how it can be cleaned and checked for quality assurance. By identifying data errors through the various checking mechanisms above, it is possible to:

- clarify instruction from the City or PM on how to report data properly,
- identify whether consultants are making systematic errors, and
- give feedback in order to improve subsequent benchmarking.

[2]For example, "Green Consulting" versus "Green Consulting, LLP" versus "GREEN Consulting". Late in the process of analysis, it was discovered that some consultants entered their name under the Portfolio Manager fields of "Full Name" and "Organization", and these were not analyzed because of the difficulty of distinguishing which field was intended by the building owner to indicate the consultant. Future analyses will make additional efforts to clean up the entries of consultant names in these fields as well.

	Facility Type	Number	0%	5%	25%	50%	75%	95%	100%
1	Multifamily Housing	6583	5.0	51.6	106.2	132.1	160.7	225.9	999.7
2	Office	912	7.4	95.0	166.0	212.9	277.7	426.7	905.7
3	Other	156	5.3	31.3	73.6	150.9	313.0	697.8	979.7
4	Hotel	122	11.7	125.0	197.8	246.7	298.0	398.6	561.7
5	Warehouse (Unrefrigerated)	91	11.9	22.9	39.7	71.8	120.8	246.8	683.9
6	Retail	75	8.5	64.7	140.4	196.9	281.0	509.8	841.2
7	K-12 School	59	55.8	71.2	140.4	195.2	222.4	307.5	577.7
8	Residence Hall/Dormitory	49	20.8	98.5	149.5	234.1	355.3	389.6	581.5
9	Senior Care Facility	49	81.5	114.9	212.9	267.5	315.2	451.1	460.2
10	Hospital	45	171.6	311.5	455.6	483.5	602.4	712.5	874.5
11	Supermarket/Grocery	23	111.0	198.7	654.6	716.3	776.8	917.5	926.9
12	College / University	20	103.9	140.0	168.4	270.3	331.1	415.8	466.2
13	Medical Office	19	74.4	121.2	176.4	263.3	334.5	705.8	744.4
14	Bank/Financial Institution	14	126.2	166.9	233.1	286.1	344.7	402.5	430.3
15	Entertainment/Culture	13	16.1	118.5	265.1	334.1	386.9	580.1	847.2
16	Education	11	72.7	80.0	98.3	156.2	226.8	277.1	299.0

Table 2.2: Table Showing Source EUI Quartiles by Facility Type. Buildings with source EUI below 5 and over 1,000 kBTU/sf already removed. Removing the bottom and top 5 percent removes extreme values of source EUI. Median is by definition at 50 percent of source EUI values.

Table 2.3 shows the breakdown of the identified errors by consultant, as identified as "Service and Product Providers" field in PM. Code for the columns in the table indicate flagged consultant data entry errors:

- Rem = Removed from analysis dataset
- E=0 indicated source EUI equal 0
- NoE for no source EUI entered
- EHi indicates source EUI over 500 kBtu per square foot
- ELo indicates source EUI below 30 kBtu per square foot
- A=0 indicates area of zero entered
- Fac indicates no facility type entered.

Finally, Er is the maximum of each of the error columns. There are possibly individual building records that have multiple errors, so the maximum of the columns captures the minimum number of records that contain errors. PercentEr therefore captures the percentage of the original buildings that are have at least one error.

The following errors were not removed from the dataset, but need to be corrected in the future data collection:

- PLUTO indicates that a PLUTO floor area was used for the PM calculation
- R indicates an implausible EnergyStar Rating of 0 or 100

Consult	Total	Rem	E=0	NoE	EHi	ELo	A=0	Fac	PLUTO	R	Er	%Er
None	4829	1173	90	486	149	155	102	91	1652	336	1652	34
C3	369	20	0	0	2	0	0	0	249	2	249	67
C6	255	43	0	3	3	0	0	0	239	3	239	94
C7	201	105	4	59	1	25	3	1	144	5	144	72
C9	174	87	3	76	3	4	3	0	135	5	135	78
C5	278	15	0	0	0	1	0	0	123	0	123	44
C2	484	65	0	7	8	0	0	0	117	0	117	24
C4	347	183	0	7	1	94	0	0	102	3	102	29
C1	513	22	0	0	0	2	0	0	80	0	80	16
C10	135	16	0	0	2	3	0	0	61	2	61	45
C24	63	61	0	2	0	59	0	0	27	0	59	94
C20	73	70	14	56	0	14	14	2	1	38	56	77
C13	114	14	5	6	1	5	5	1	52	2	52	46
C16	95	7	0	0	0	0	0	0	51	0	51	54
C17	94	29	0	17	0	0	0	0	47	0	47	50
C12	117	10	2	4	0	2	2	2	41	2	41	35
C19	76	37	31	3	2	31	31	31	0	36	36	47
C22	66	16	0	3	4	0	0	0	33	0	33	50
C26	57	9	0	0	0	0	0	0	27	0	27	47
C23	66	9	0	1	0	0	0	0	24	0	24	36
C18	93	8	0	0	0	0	0	0	23	0	23	25
C32	28	5	0	5	0	0	0	0	22	0	22	79
C33	22	3	0	0	1	0	0	0	21	0	21	95
C8	176	14	0	5	1	1	0	0	19	5	19	11
C37	18	14	0	2	0	11	0	0	16	1	16	89
C25	60	21	0	1	6	0	0	0	1	14	14	23
C40	16	1	0	0	0	0	0	0	14	0	14	88
C30	33	5	0	0	0	1	0	0	12	3	12	36
C11	126	16	0	6	2	0	0	0	12	0	12	10
C27	55	7	0	0	0	0	0	0	12	0	12	22
C34	22	17	0	5	12	0	0	0	0	0	12	55
C39	16	7	0	3	0	2	0	0	11	0	11	69
C28	53	15	1	1	0	1	1	0	10	2	10	19
C21	71	12	0	1	3	1	0	0	0	7	7	10
C38	18	8	4	1	0	4	4	5	0	7	7	39
C51	9	1	0	0	0	0	0	0	7	0	7	78
C15	108	3	0	0	0	0	0	0	7	0	7	6
C43	14	5	0	2	0	0	0	0	5	1	5	36
C44	13	8	1	5	0	1	1	0	5	0	5	38
C69	5	0	0	0	0	0	0	0	5	0	5	100

Table 2.3: Breakdown of Consultant Data Errors by Type, Sorted by Total Errors, Minimum 5 Total Errors. Firm names are anonymized with numbers. Codes for error types are above in the text.

2.4 Validation by Control Groups

The NYC data was also extensively checked for consistency by adding knowledge of city staff and various consultants. OLTPS identified particular consultants within each building type which it considered to have done a credible job of data reporting, such as consultants who actually visited the building site and took new measurements as part of the benchmarking reporting process (as opposed to those who may have used previously existing data or default settings in the reporting mechanism). Two such control groups were identified for the two major property sectors, multifamily and office buildings.

2.4.1 Multifamily

In the multifamily housing sector, one owner and one consultant were used as a control group for the data, since they were confirmed to have performed much more extensive auditing, which led OLTPS to believe that their benchmarking data is also fairly robust.

Table 2.4 shows the summary statistics for all multifamily buildings, and the control groups from owner 1 and consultant 1.

Figure 2.2 shows the comparison of the log source EUI between the control groups and the general population. Histograms and quantile-quantile ("Q-Q") plots show that the source EUIs from owner 1 most closely resembles the overall population, though at the low end (low quartiles), owner 1 has more buildings that have comparatively lower source EUIs. Consultant 2's buildings have relatively lower source EUIs as compared throughout the entire population distribution.

Figure 2.3 shows the comparison of the log area between the control groups and the general population. Histograms and quantile-quantile ("Q-Q") plots show that the building sizes from consultant 1 most closely resembles the overall population. Owner 1 has a higher proportion of larger and smaller buildings.

2.4.2 Office

Similarly, two control groups – with one owner and one consultant – was used to check the office group.

Table 2.5 shows the summary statistics for all office buildings, and the control groups from owner 2 and consultant 2.

Figure 2.4 shows the comparison of the log source EUI between the control groups and the general population. Histograms and quantile-quantile ("Q-Q") plots show that the source EUI distribution of consultant 2's buildings are a very close match to the overall office population. Owner 2 has buildings with a generally higher distribution of source EUIs.

Figure 2.5 shows the comparison of the log area between the control groups and the general population. Histograms and quantile-quantile ("Q-Q") plots show that consultant 2 again has a distribution of building sizes that closely match the general population. Owner 2 again has a group of buildings that are comparatively larger than the overall population.

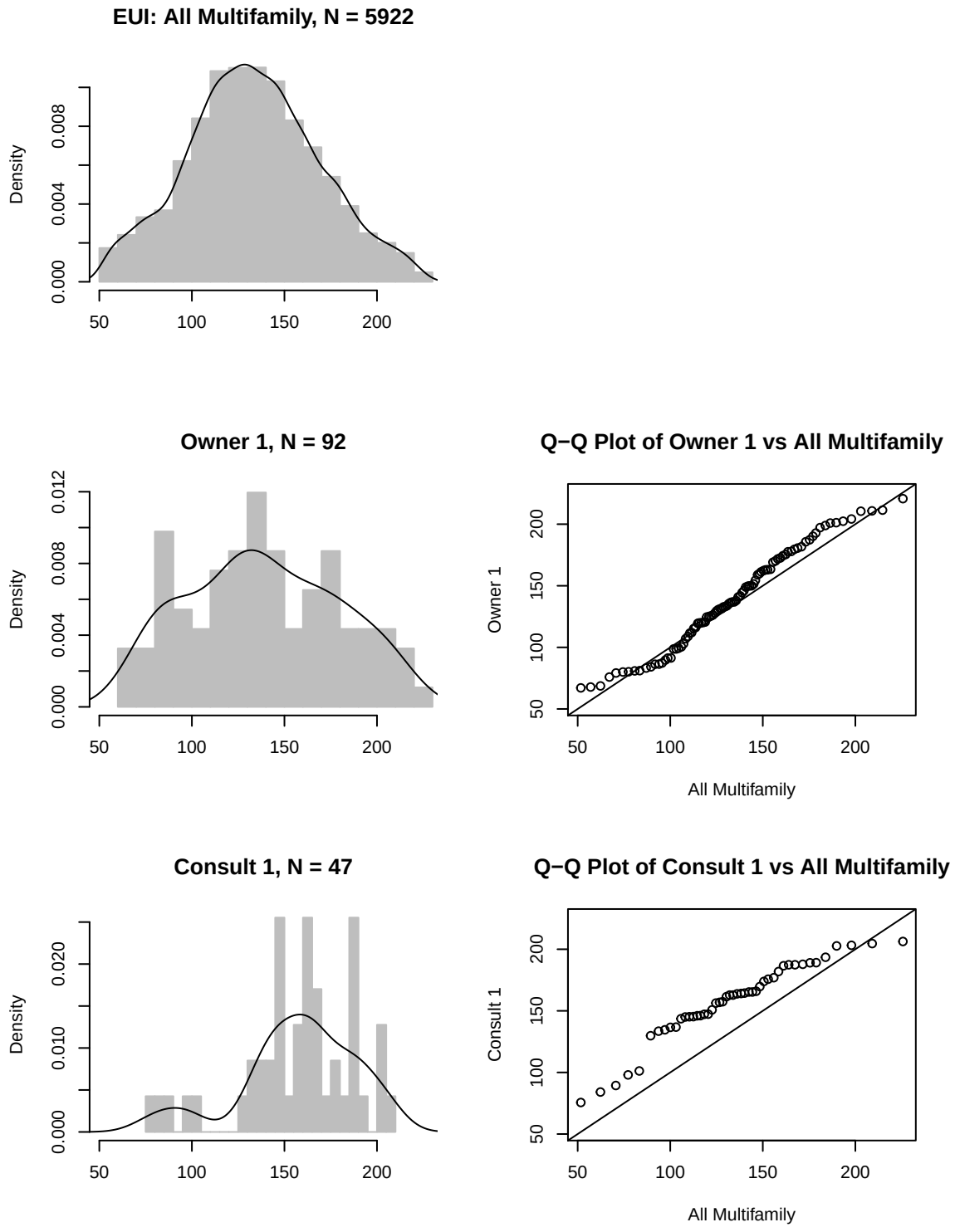


Figure 2.2: EUI comparisons between Control Groups and all MF. Histograms and quantile-quantile ("Q-Q") plots show that the source EUIs from owner 1 most closely resembles the overall population, though at the low end (low quartiles), owner 1 has more buildings that have comparatively lower source EUIs. Consultant 1's buildings have relatively higher source EUIs compared to the entire population distribution.

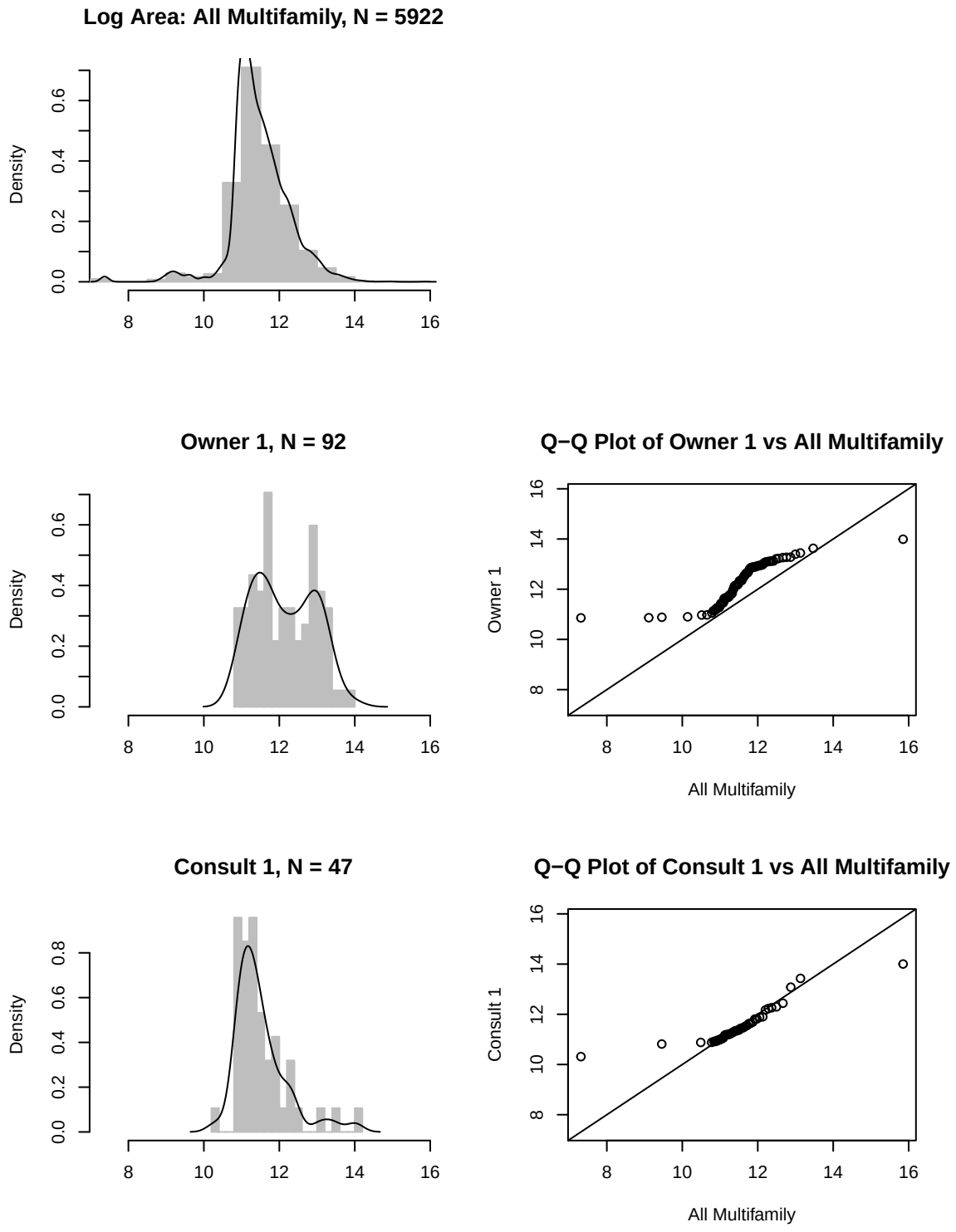


Figure 2.3: Log Area comparisons between Control Groups and all MF. See Table 2.4 for log conversion. Histograms and quantile-quantile ("Q-Q") plots show that the building sizes from consultant 1 most closely resembles the overall population. Owner 1 has a higher proportion of larger and smaller buildings.

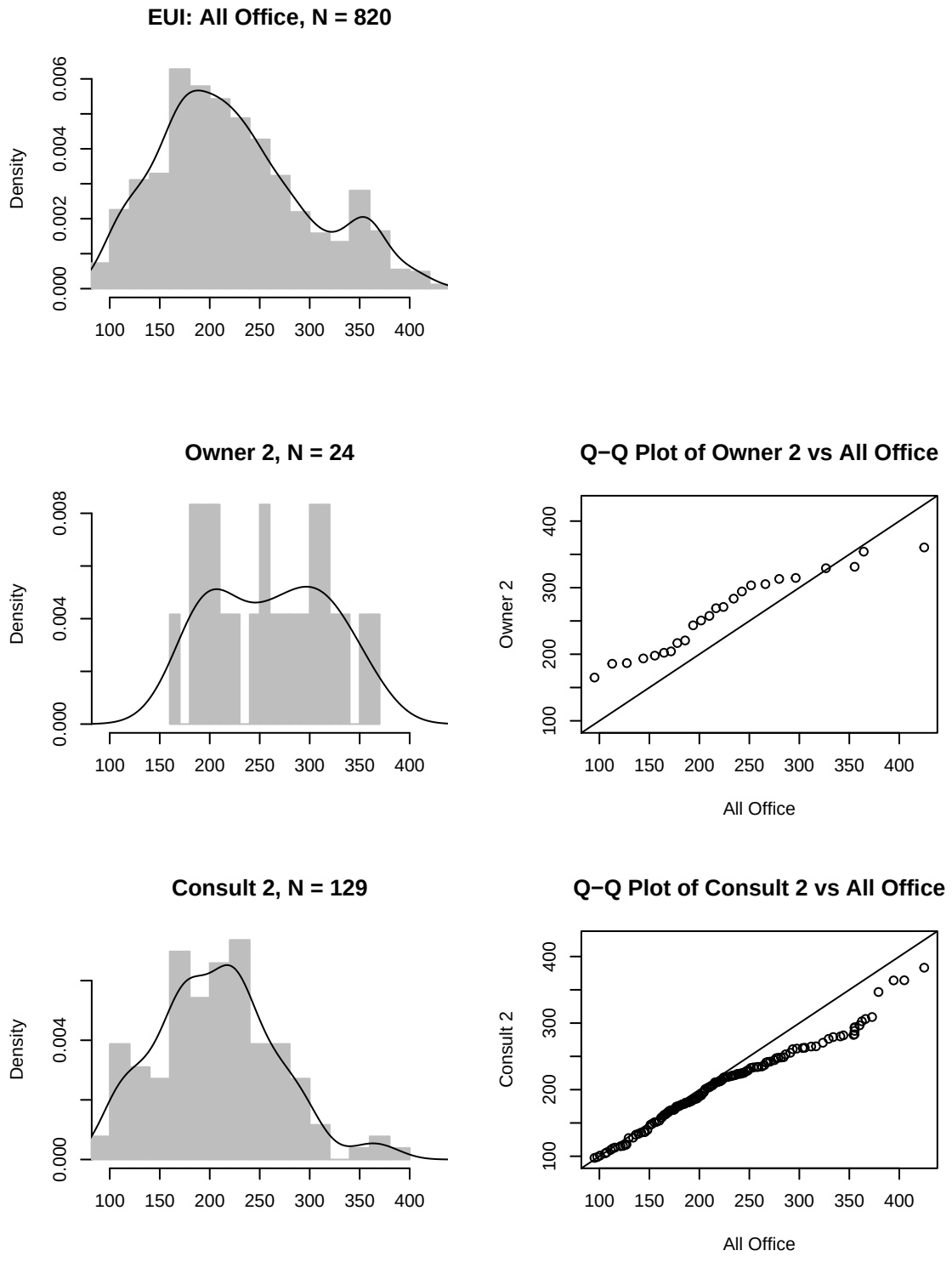


Figure 2.4: Source EUI Comparisons between Control Groups and all Offices. Histograms and quantile-quantile ("Q-Q") plots show that the source EUI distribution of consultant 2's buildings are a very close match to the overall office population. Owner 2 has buildings with a generally higher distribution of source EUIs.

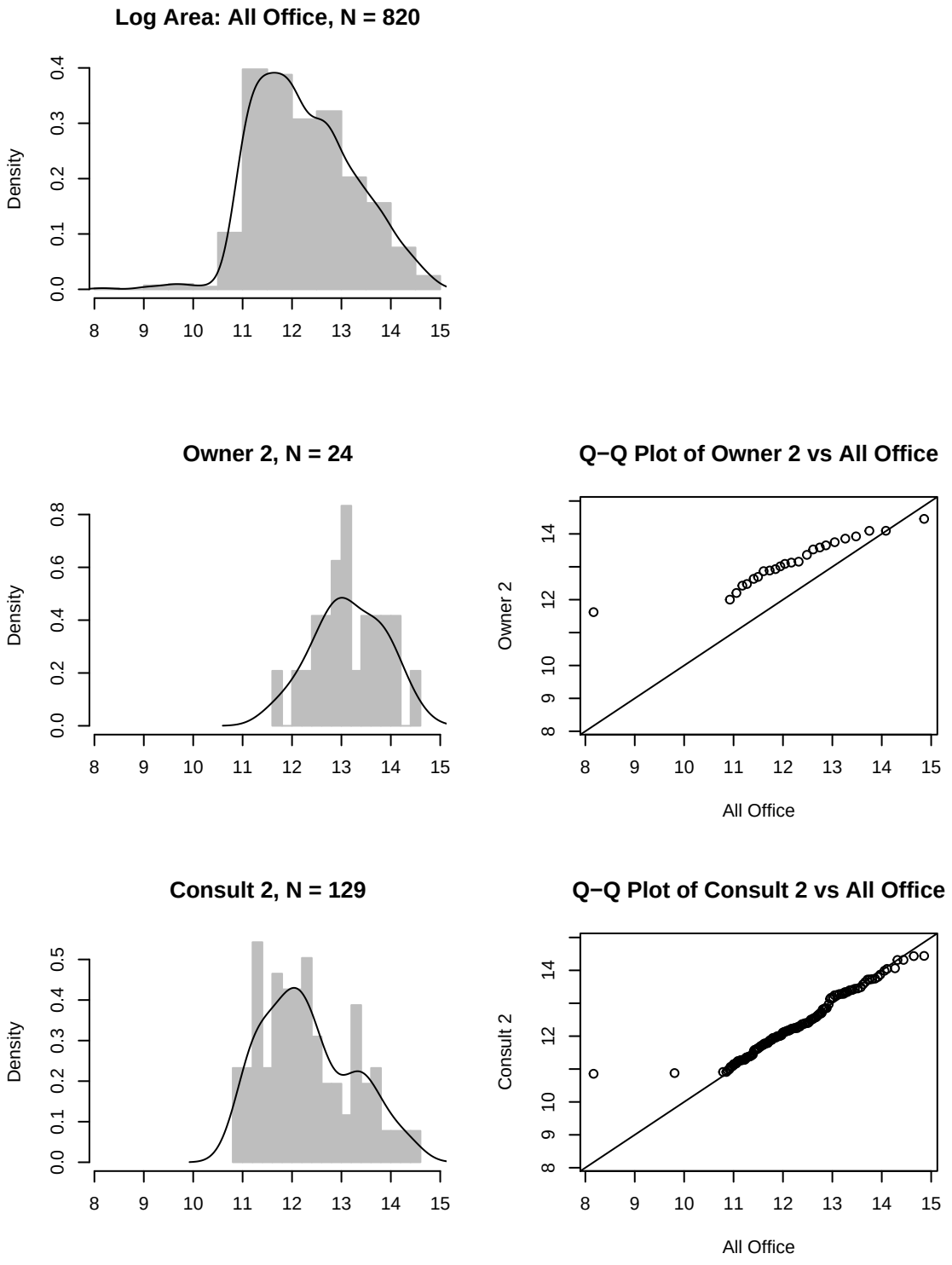


Figure 2.5: Log Area Comparisons between Control Groups and all Offices. See Table 2.5 for log conversion. Histograms and quantile-quantile ("Q-Q") plots show that consultant 2 again has a distribution of building sizes that closely match the general population. Owner 2 again has a group of buildings that are comparatively larger than the overall population.

Metric	Consult 1	Owner 1	MF excl.	All MF
Total number of buildings	47	92	5797	5922
EUI – 0%	76	67	52	52
25%	145	108	109	109
50%	163	136	132	132
75%	179	171	157	157
100%	206	221	363	226
Log EUI – 0%	4.3	4.2	3.9	3.9
25%	5	4.7	4.7	4.7
50%	5.1	4.9	4.9	4.9
75%	5.2	5.1	5.1	5.1
100%	5.3	5.4	5.9	5.4
SF – 0%	30118	51960	1500	1500
25%	61860	93885	62736	62967
50%	83730	186756	88000	88500
75%	126630	394300	145398	146681
100%	1205457	1191180	7661750	7661750
Log SF – 0%	10.3	10.9	7.3	7.3
25%	11	11.4	11	11.1
50%	11.3	12.1	11.4	11.4
75%	11.7	12.9	11.9	11.9
100%	14	14	15.9	15.9
Median EnergyStar Rating	0	0	0	0
Median Year Built	1955	1960	1941	1941

Table 2.4: Summary Statistics for Control Groups from Owner 1 and Consultant 1; Multifamily excluding the control groups; and All Multifamily buildings.

Metric	Owner 2	Consult2	Office excl.	All Office
Total number of buildings	24	129	667	820
EUI – 0%	165	98	95	95
25%	204	168	169	170
50%	263	206	213	213
75%	307	242	274	268
100%	360	383	425	425
Log EUI – 0%	5.1	4.6	4.6	4.6
25%	5.3	5.1	5.1	5.1
50%	5.6	5.3	5.4	5.4
75%	5.7	5.5	5.6	5.6
100%	5.9	5.9	6.1	6.1
SF – 0%	111447	51844	3520	3520
25%	320800	109857	94327	96423
50%	493100	192841	169887	181794
75%	870286	429435	384566	410859
100%	1898730	1866776	2834104	2834104
Log SF – 0%	11.6	10.9	8.2	8.2
25%	12.7	11.6	11.5	11.5
50%	13.1	12.2	12	12.1
75%	13.7	13	12.9	12.9
100%	14.5	14.4	14.9	14.9
Median EnergyStar Rating	68	67	67	67
Median Year Built	1956	1926	1926	1926

Table 2.5: Summary Statistics for Control Groups from Owner 2 and Consultant 2, Office Excluding Control Groups, and All Office.

Chapter 3

Overall Building Population

This chapter describes observed energy performance and building typologies in the fully cleaned dataset.

3.1 Energy Performance

As reported in the NYC dataset, very few buildings have applied for an EnergyStar label. No EnergyStar rating exists for multifamily buildings, which comprise the majority of the NYC dataset. Of the 7,401 buildings in the fully cleaned dataset, only 1,479 buildings (20%) are not multifamily. Of these, only 399 buildings are eligible for an EnergyStar label (e.g. non-multifamily buildings with ratings over 75). Figure 3.1 shows the reported ratings for the non-multifamily buildings.

Figure 3.2 shows the relative breakdown of energy use between different building sectors.

Figure 3.5 shows the distribution of energy and GHG emissions by building sector and quartile.

3.2 Building Typologies

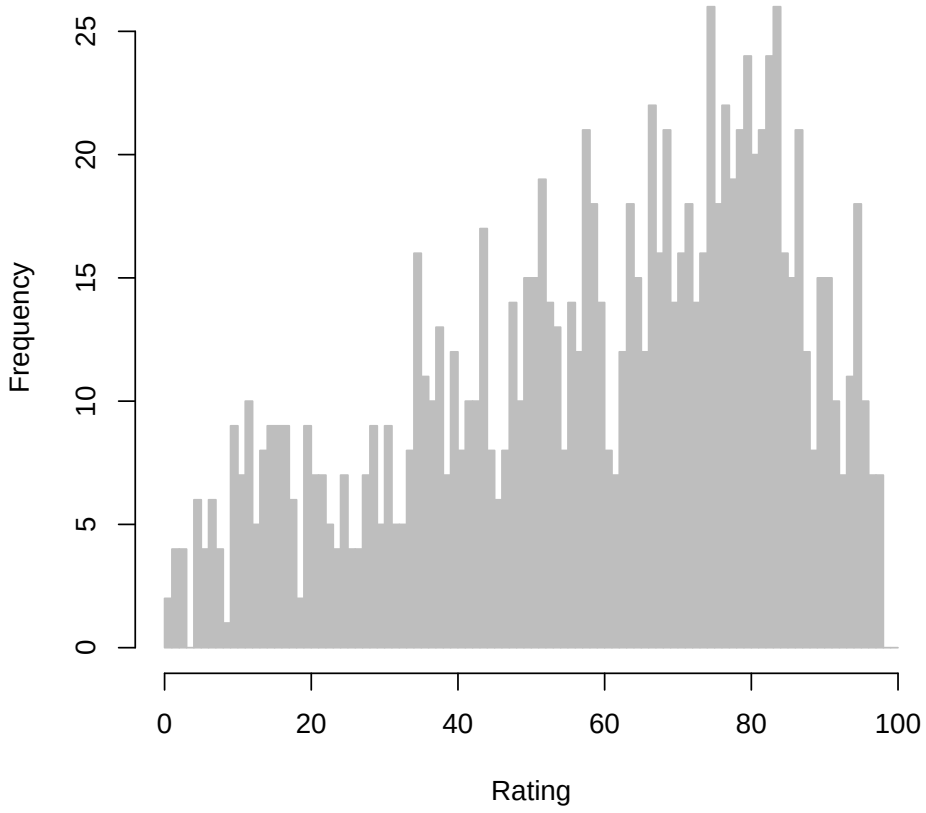


Figure 3.1: EnergyStar Ratings for Eligible Buildings. Only 1,479 (non-multifamily) buildings are ratable. Ratings of 0 and 100 omitted due to likely data errors. Median value is 64.

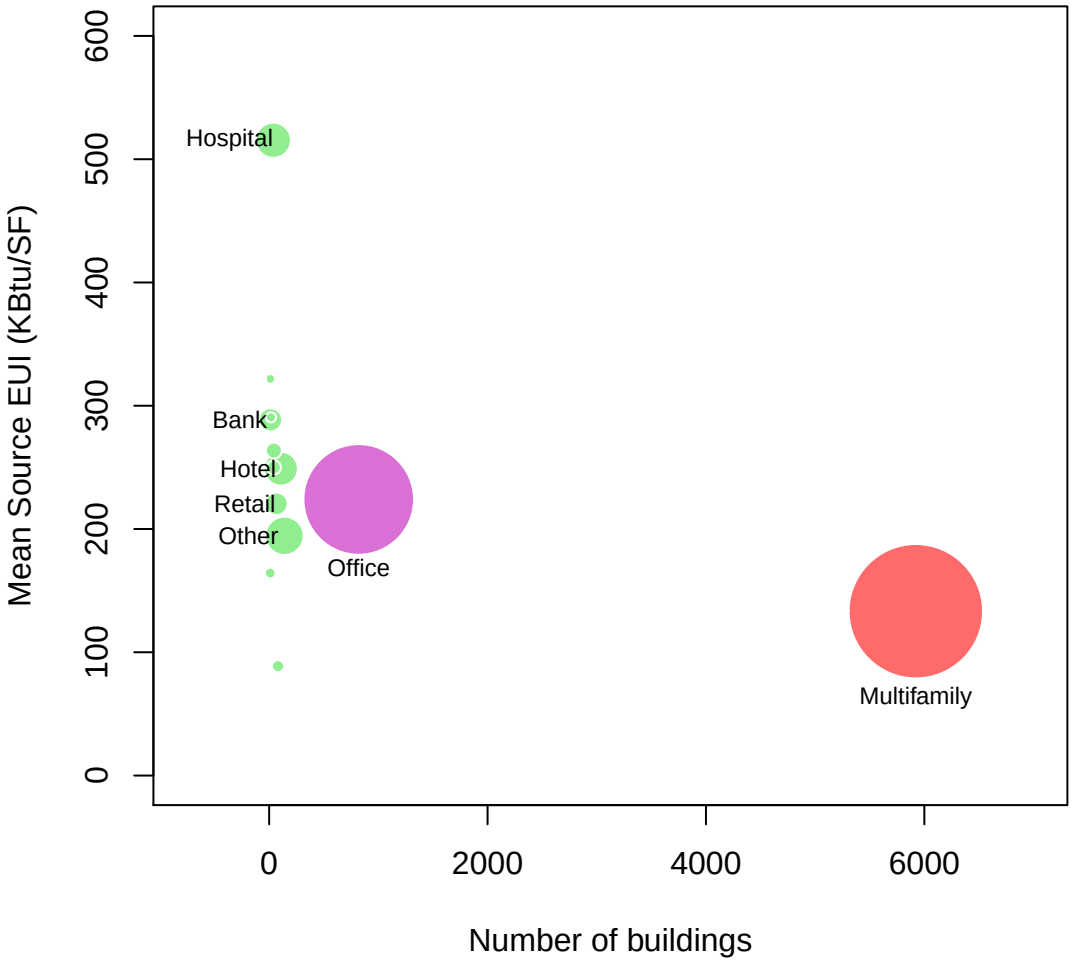


Figure 3.2: Chart of Energy Use by Building Sector. Area of the circles indicates the total amount of energy consumed by sector, plotted against the number of buildings (x-axis) and the mean source EUI in each facility type (y-axis).

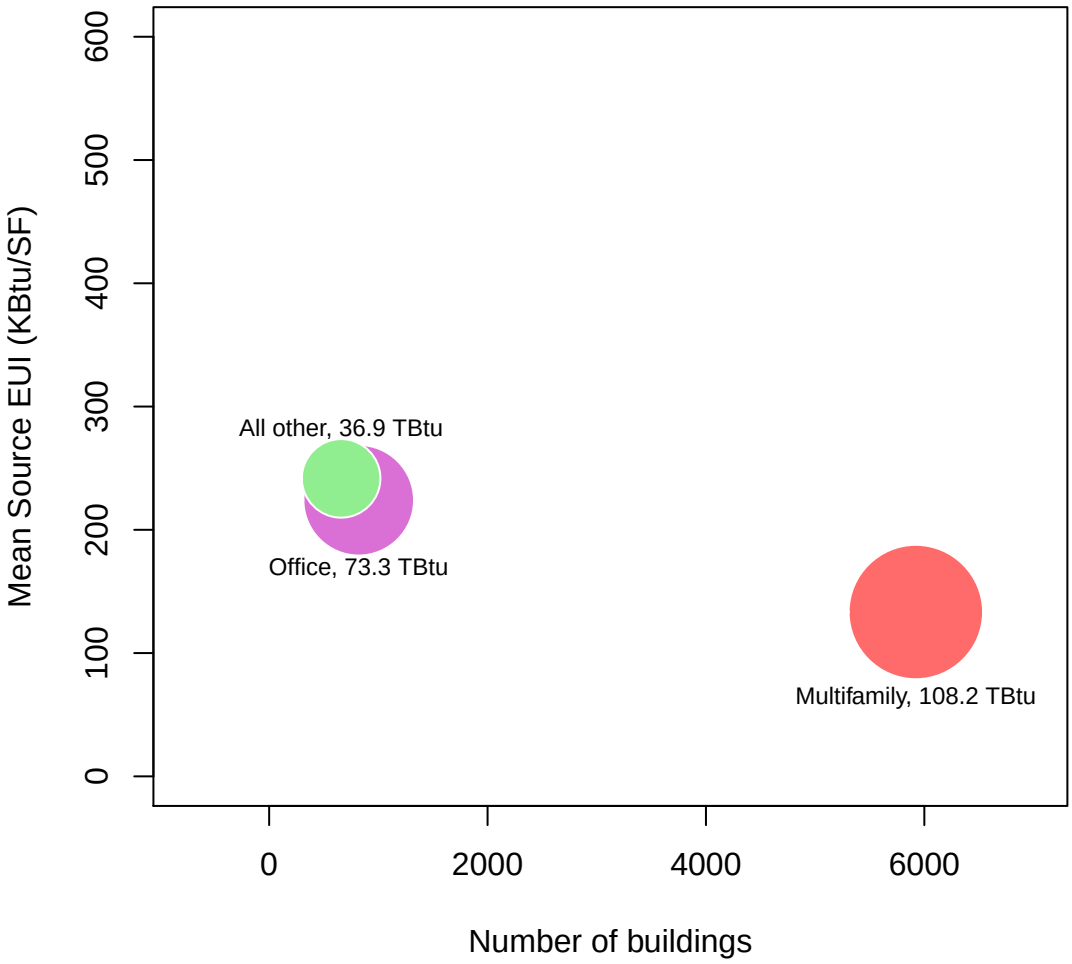


Figure 3.3: Building category versus total energy use. Circle areas indicate total energy use reported in Tera British Thermal Units, or 10^{12} BTUs, plotted versus number of buildings and mean source EUI in each category.

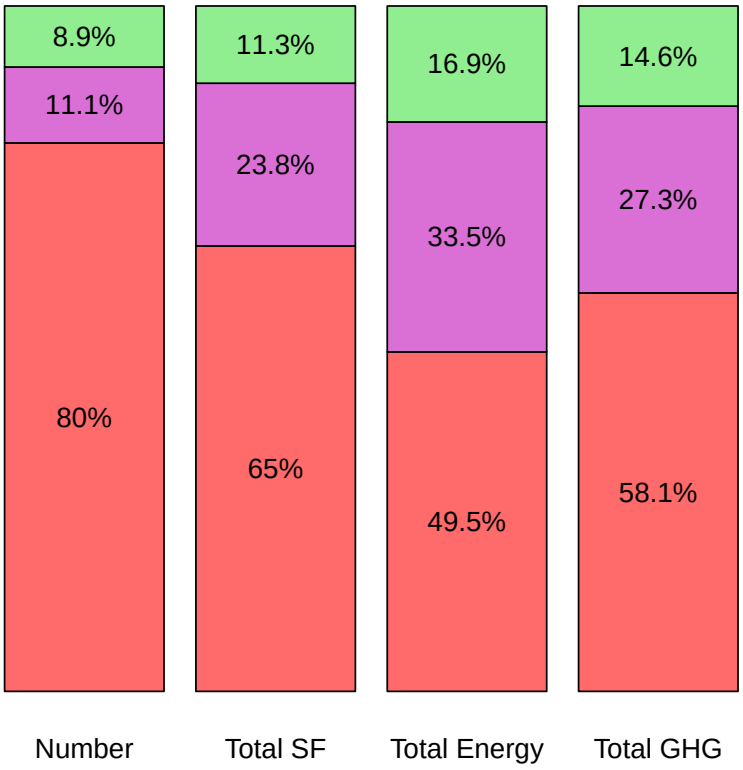
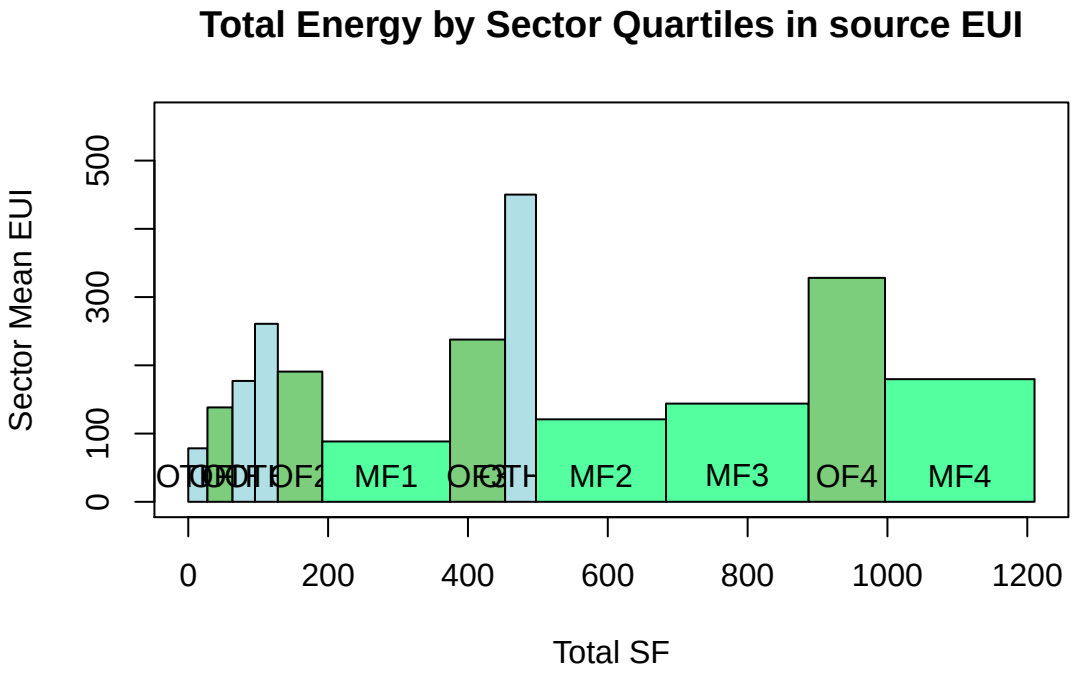


Figure 3.4: Breakdown of Buildings, SF, Energy, and Greenhouse Gas Emissions by Building Sector. Stacked bar charts indicate, from bottom to top, percentage associated with multifamily, office, and all other building types.



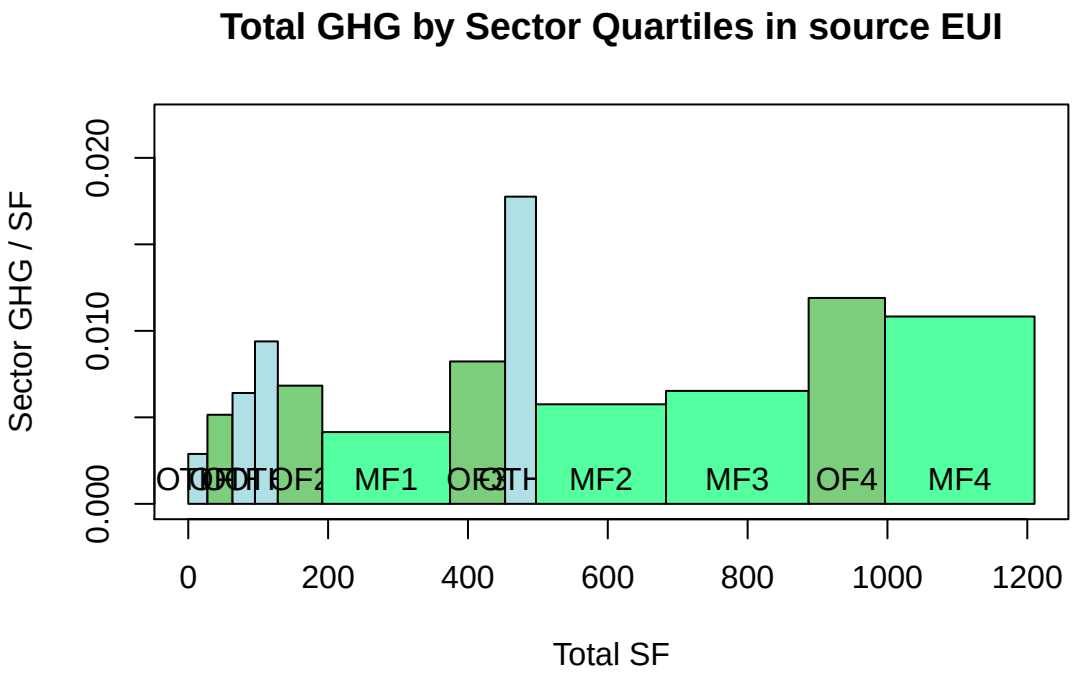


Figure 3.5: Chart of Total Energy Use by Building Sector. Height of each rectangle is the mean source EUI for that building sector and quartile, width is the total square footage, and the total area of each rectangle indicates the total amount of energy consumed by sector. Rectangles are ordered according to the total amount of energy in each rectangle.

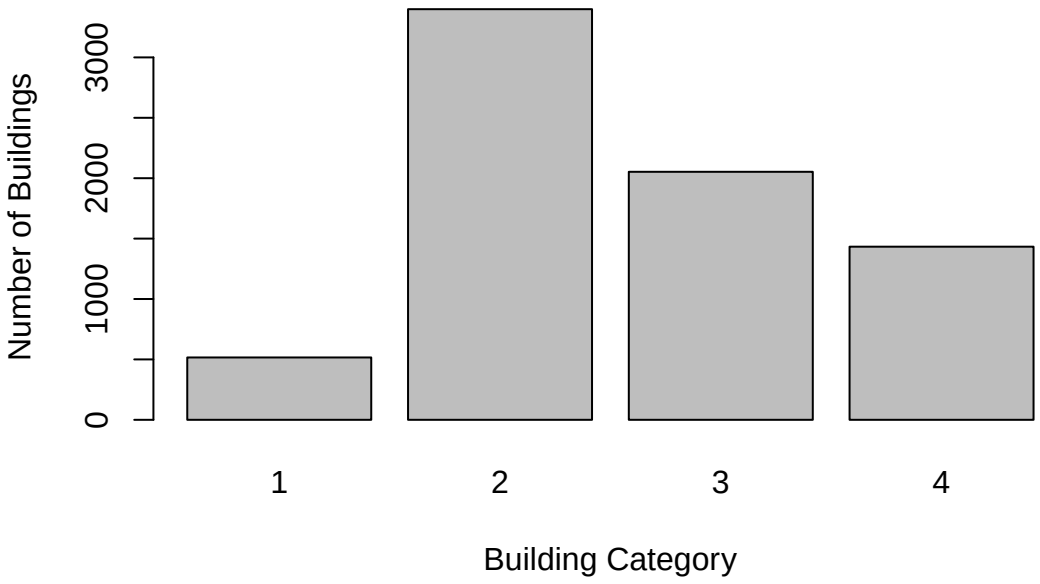


Figure 3.6: Building sizes by category. All categories reported in thousands of square feet. Category 1: buildings less than 50,000 SF shows some reported despite being not required to. Category 2: 50-100,000 SF. Category 3: 100-200,000 SF. Category 4: larger than 200,000 SF.

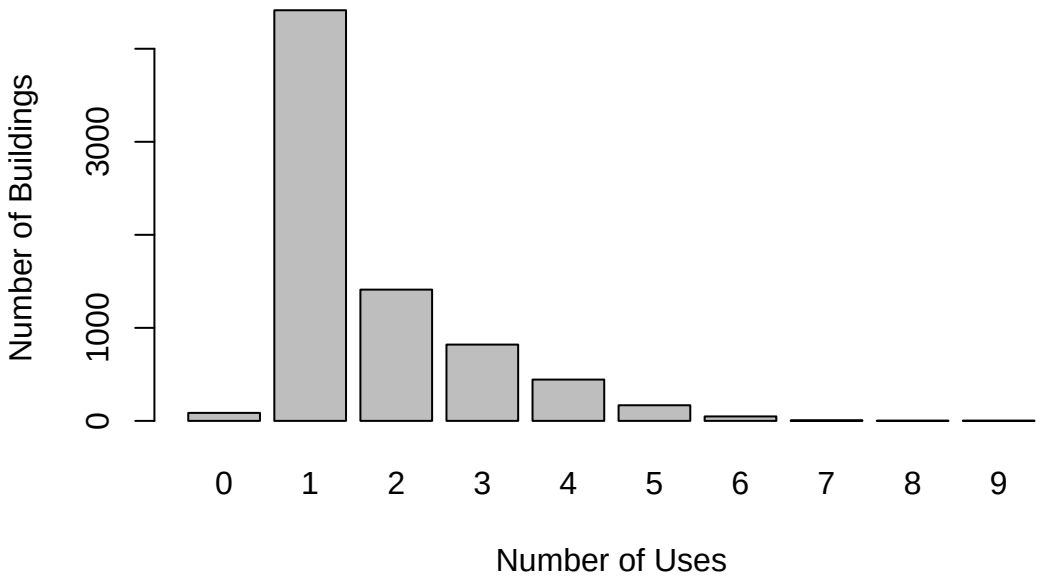


Figure 3.7: Number of different space uses per building. Majority of buildings are single use, with very progressively fewer reporting multiple space uses.

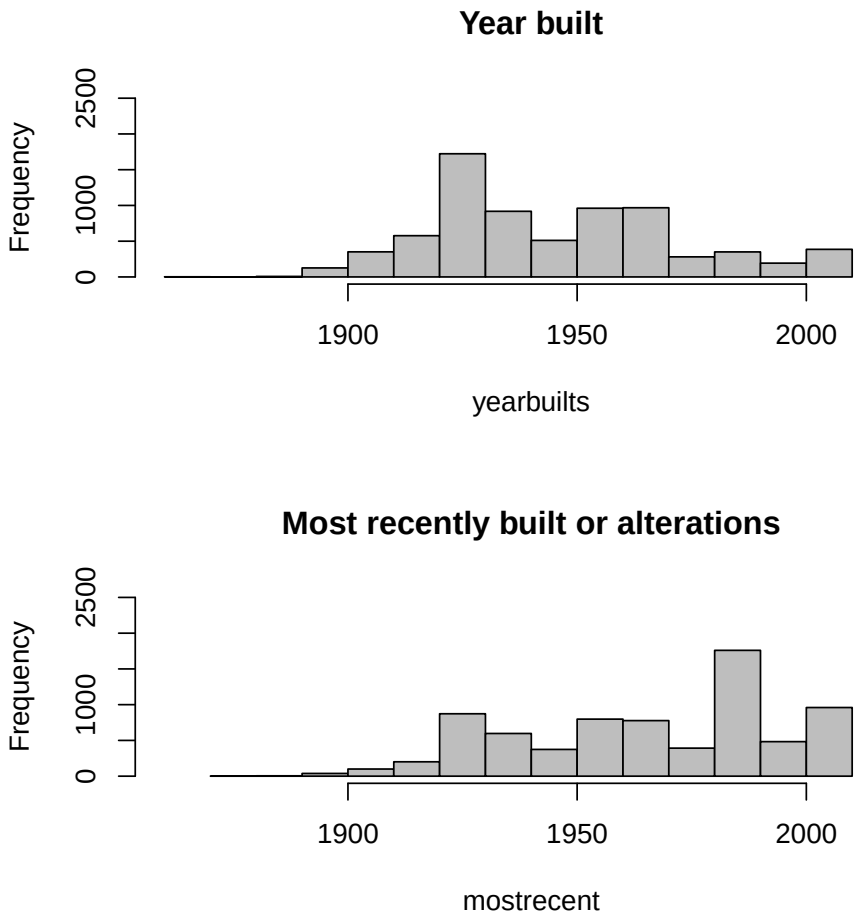


Figure 3.8: Building vintages. Above graph shows building booms in the 1920s, 1950s and 1960s. Below graph shows significant levels of alterations occuring in 1980s and after 2000.

	Building Type	Number	MSF	%Num	%SF
1	Multifamily Housing	5922	786.5	65.0	80.0
2	Office	820	287.5	23.8	11.1
3	Other	140	35.5	2.9	1.9
4	Hotel	108	27.2	2.3	1.5
5	Hospital	38	14.3	1.2	0.5
6	Retail	67	13.1	1.1	0.9
7	Bank/Financial Institution	12	9.9	0.8	0.2
8	Warehouse (Unrefrigerated)	81	9.7	0.8	1.1
9	Senior Care Facility	43	5.7	0.5	0.6
10	Residence Hall/Dormitory	43	5.2	0.4	0.6
11	K-12 School	53	4.3	0.4	0.7
12	Education	9	3.7	0.3	0.1
13	College / University	18	2.6	0.2	0.2
14	Entertainment/Culture	11	2.1	0.2	0.1
15	Medical Office	17	1.9	0.2	0.2
16	Supermarket/Grocery	19	1.0	0.1	0.3

Table 3.1: Breakdown of Square Footage and Numbers of Buildings by PM Facility Type. Square footage in some cases is aggregated or cumulative across multiple buildings. NYC data is dominated by Multifamily and Office buildings in both SF and number.

Chapter 4

Data Summaries by Building Type

4.1 Multifamily Housing

	MSF	Acres	Buildings	Floors
1	7.7	175.9	14	20
2	3.3	75.2	45	9
3	2.9	67.1	39	9
4	2.5	56.4	31	9
5	1.9	43.5	5	20
6	1.5	34.8	5	44
7	1.5	34.6	4	27
8	1.4	33.0	7	25
9	1.4	32.0	6	34
10	1.4	31.8	3	24

Table 4.1: Ten Biggest Multifamily Buildings in NYC Data. Based on cleaned data. Multiple buildings on a lot are clearly affecting the largest reported building records in PM.

	Indicated Use	SF	Percent SF
1	Multifamily	768,210,840	95.66
2	Parking	16,023,646	2
3	Retail	8,398,225	1.05
4	Other	3,727,602	0.46
5	Office	3,380,778	0.42
6	Medical Office	2,095,001	0.26
7	Grocery	446,368	0.06
8	Bank	432,154	0.05
9	School	201,152	0.03
10	Hotel	100,000	0.01

Table 4.2: Top 10 Space Uses Within Multifamily Sector. Based on cleaned data.

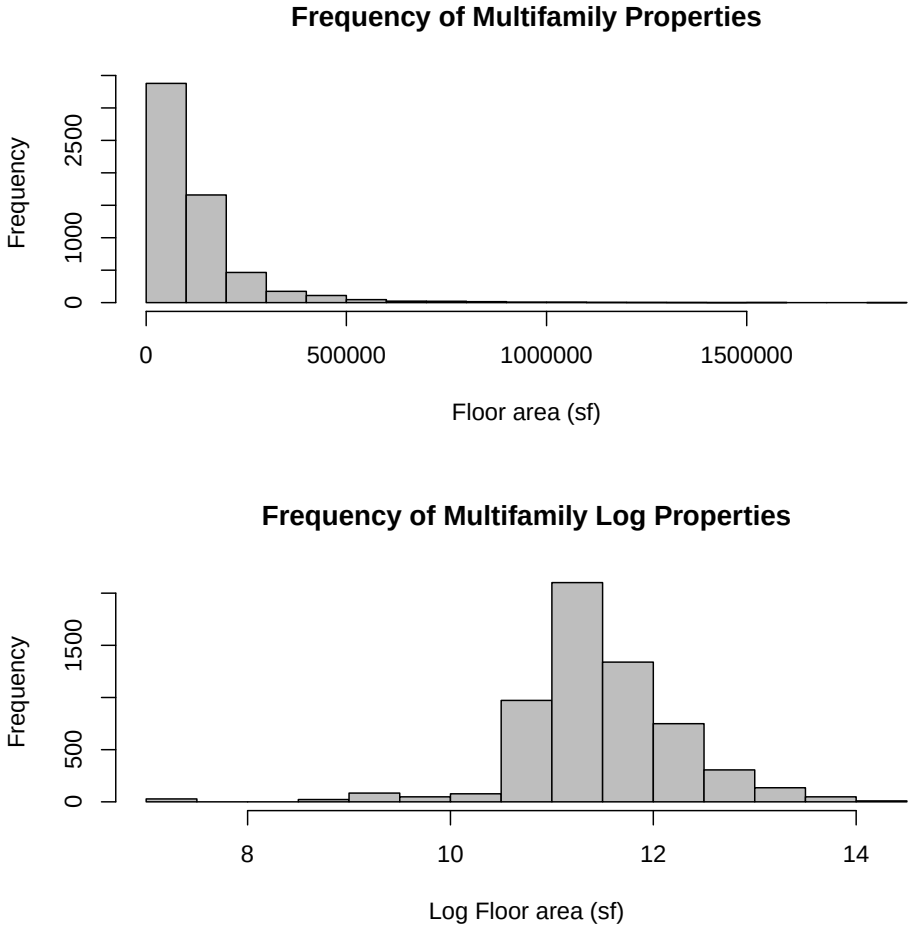


Figure 4.1: Histogram of building sizes, absolute and log

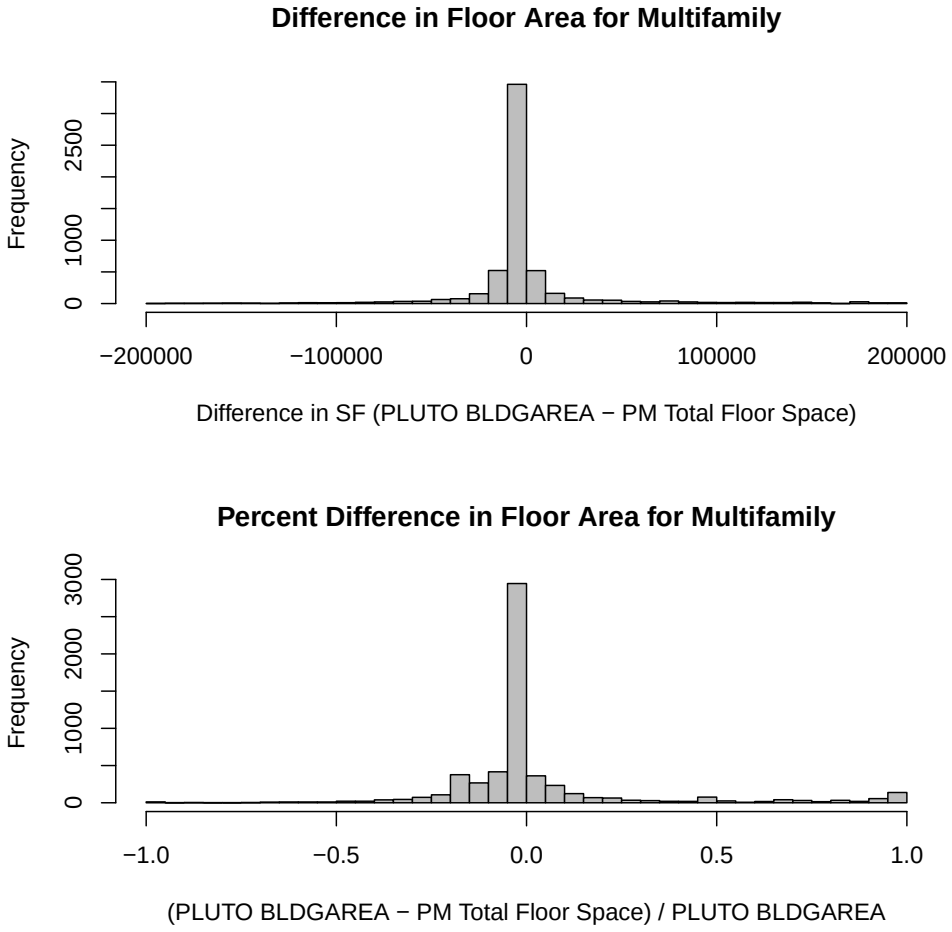


Figure 4.2: Histogram of Differences Between PLUTO and PM Total Floor Space. 43.57% buildings have the exact same PLUTO and PM floor space. Idea for plot from Constantine Kontokosta.

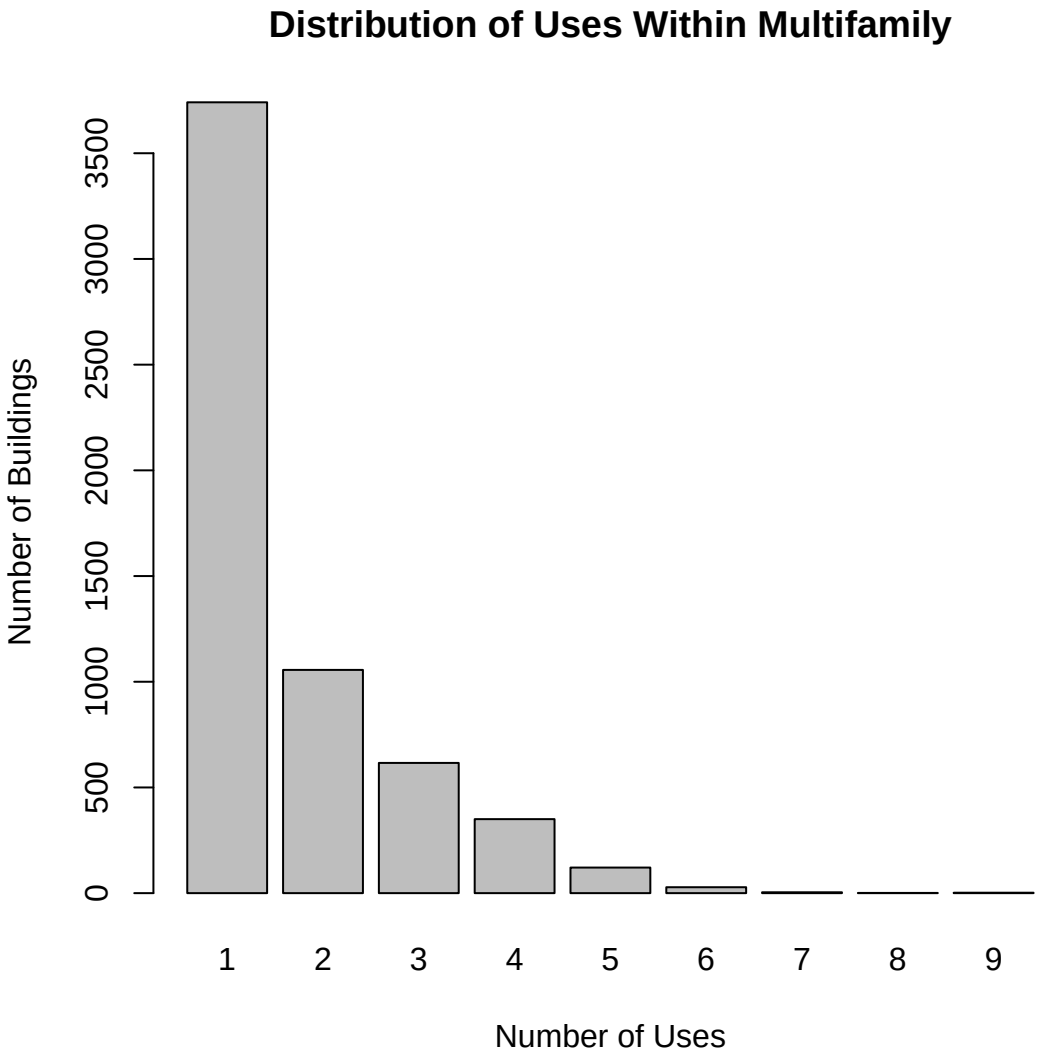


Figure 4.3: Number of Different Uses Indicated in Each Building

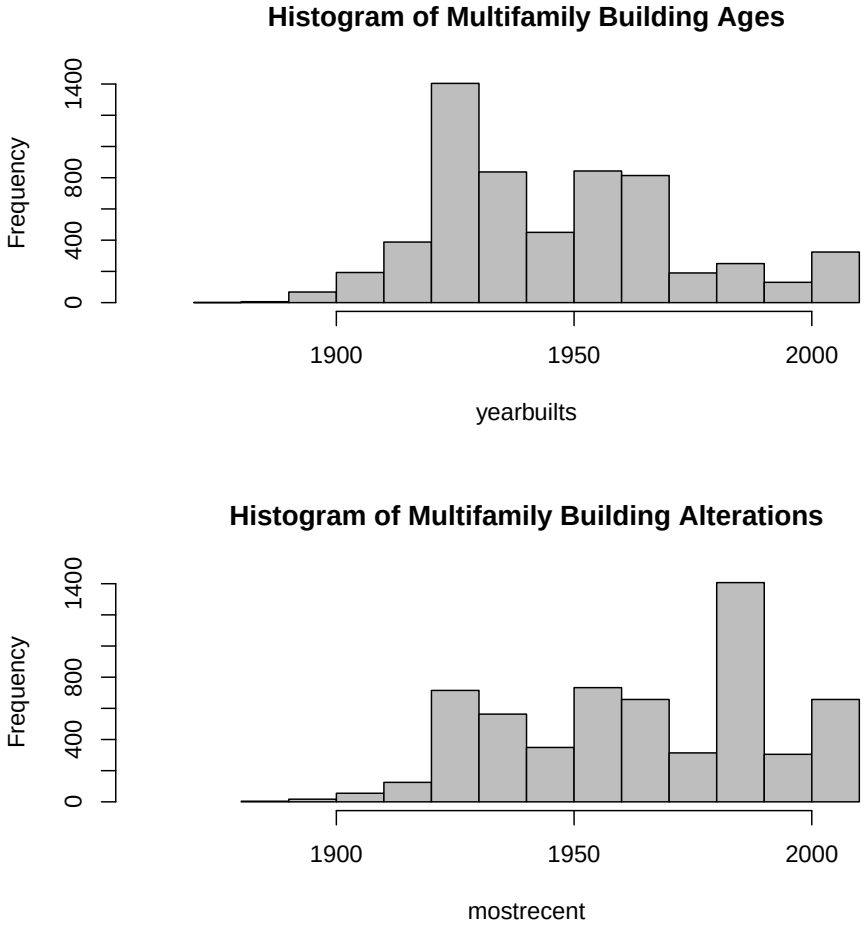


Figure 4.4: Year Built and Alterations from PLUTO

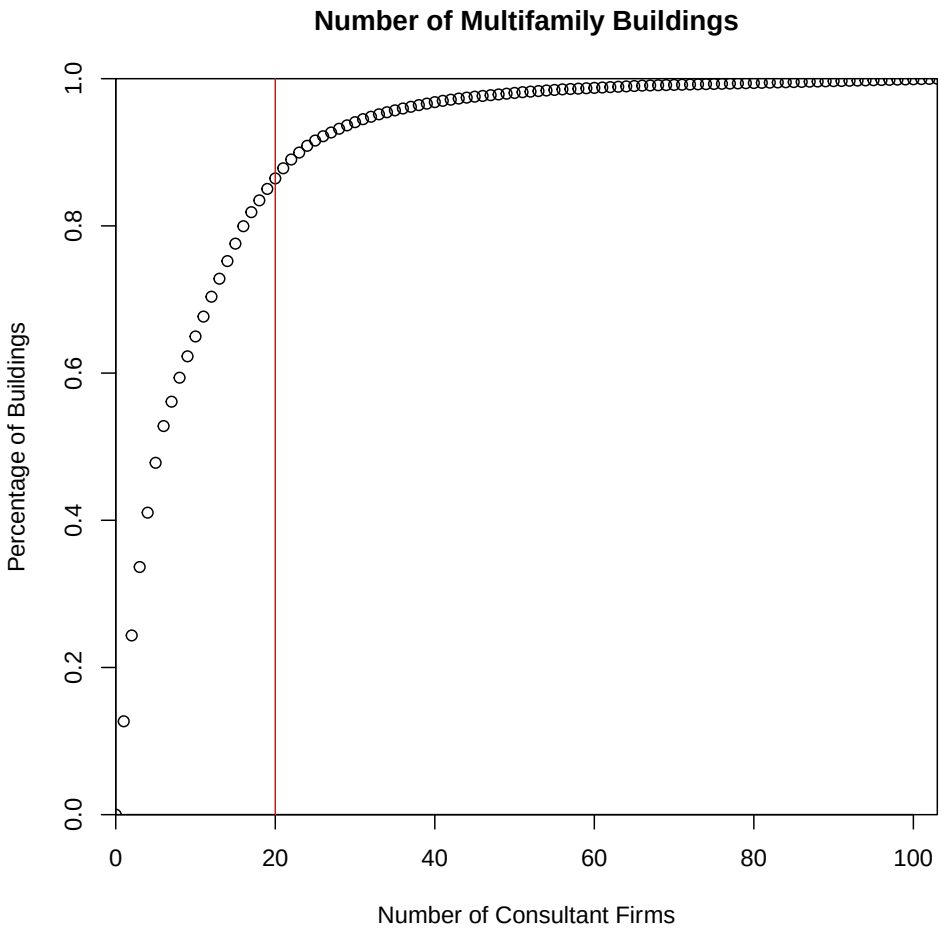


Figure 4.5: Consultants by SF and Number of Buildings. For the approximately 50% of buildings that indicate a consultant, graphs show that 80% of the building square footage and numbers can be reached (and corrected) by contacting the top 20 consultants.

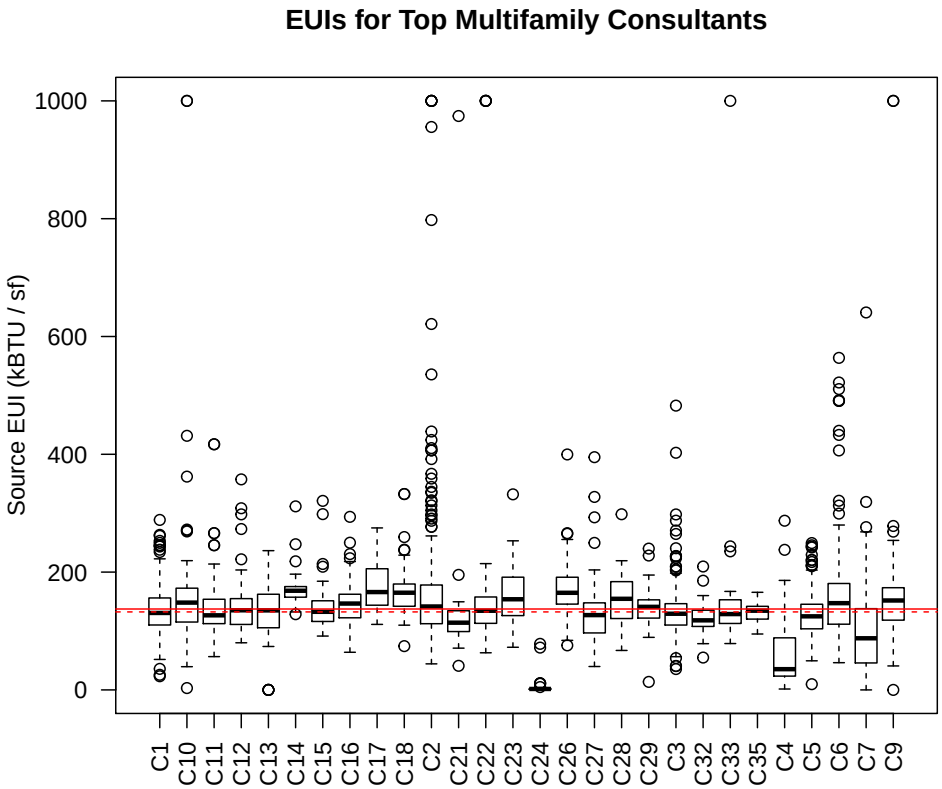


Figure 4.6: EUIs plotted by Consultant. Based on original, uncleaned data, though all EUIs greater than 1,000 kBTU/sf are set to 1000. Red solid line indicates mean across all buildings, red dashed line indicates median.

	Consultant	Number	KSF	MeanEUI	MedianEUI
1	C1	498	68972.9	135.5	130.7
2	C2	458	61822.8	164.3	141.6
3	C3	365	43583.4	132.5	129.3
4	C4	290	13377.6	61.1	35.2
5	C5	266	26784.6	129.0	124.9
6	C6	196	32249.0	160.9	147.2
7	C10	130	12634.9	155.8	148.2
8	C7	128	11265.8	96.2	87.7
9	C14	114	1241.5	168.0	168.3
10	C11	106	10326.9	139.0	126.6
11	C13	106	11537.0	133.5	135.0
12	C15	106	10546.7	137.1	132.6
13	C12	96	27176.5	141.6	135.2
14	C16	94	11359.0	146.6	146.6
15	C18	93	7973.3	167.4	165.0
16	C9	93	8447.7	167.3	151.8
17	C17	75	34453.9	176.5	166.1
18	C22	63	9520.0	190.8	134.4
19	C24	61	7656.0	4.5	1.6
20	C26	56	10111.9	175.0	164.8
21	C27	54	6626.4	137.8	127.0
22	C23	47	8379.7	160.9	153.9
23	C29	37	4606.6	138.5	141.0
24	C28	36	6181.7	156.0	154.7
25	C21	28	5993.6	144.8	114.1
26	C32	23	2903.0	122.8	118.1
27	C33	20	2095.2	178.6	128.8
28	C35	20	1672.2	132.4	134.4

Table 4.3: Consultant Breakdown for Multifamily Buildings. Ordered by number of buildings for each consultant. Based on original, uncleaned data, though all EUIs greater than 1,000 kBTU/sf are set to 1000.

4.2 Office

	Indicated Use	SF	Percent SF
1	Office	263,870,371	91.2
2	Retail	7,814,069	2.7
3	Bank	5,396,587	1.87
4	Other	4,983,548	1.72
5	Parking	2,543,111	0.88
6	Data center	2,109,800	0.73
7	Medical Office	728,113	0.25
8	Multifamily	633,815	0.22
9	Dormitory/School	497,465	0.17
10	School	447,696	0.15

Table 4.4: Top 10 Space Uses Within Office Sector. Based on cleaned data.

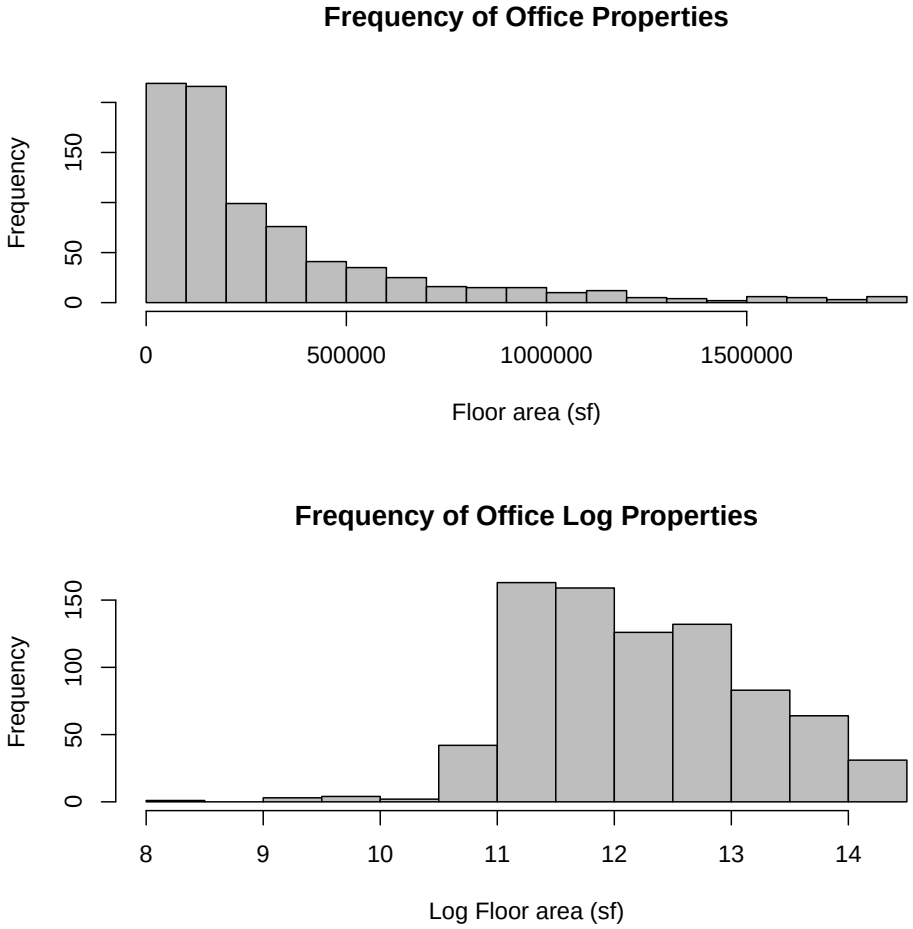


Figure 4.7: Histogram of building sizes, absolute and log

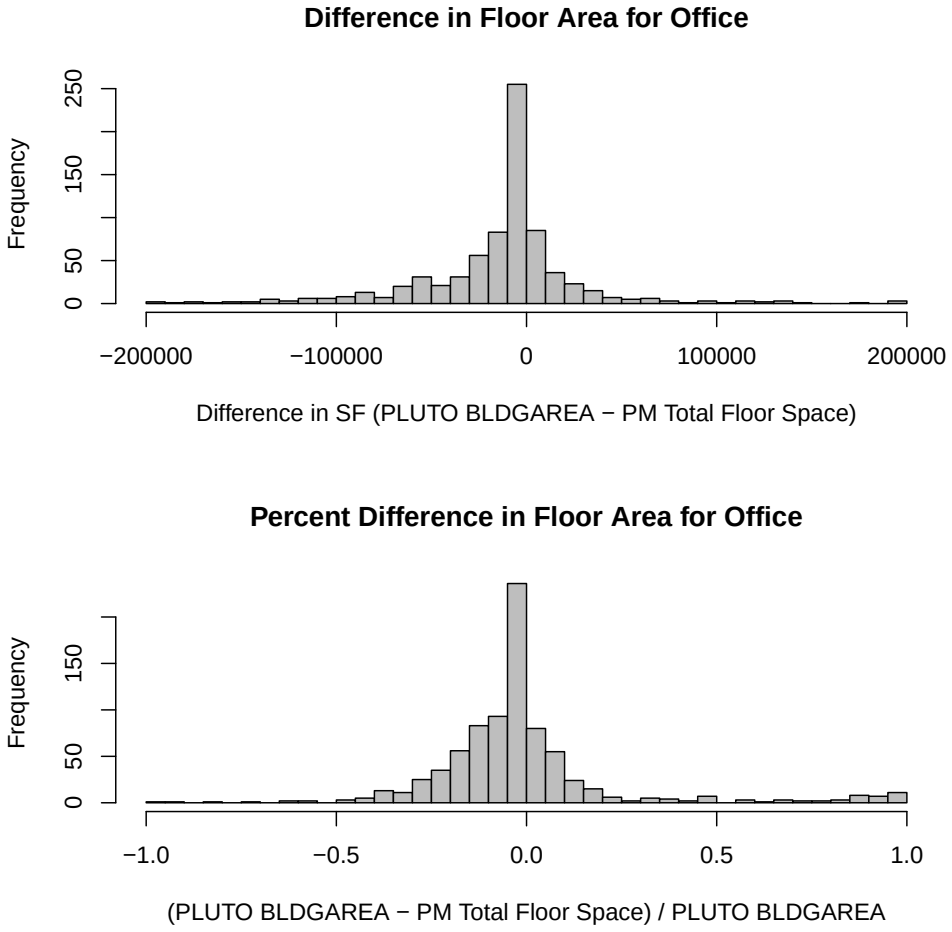


Figure 4.8: Histogram of Differences Between PLUTO and PM Total Floor Space. 12.68% buildings have the exact same PLUTO and PM floor space. Idea for plot from Constantine Kontokosta.

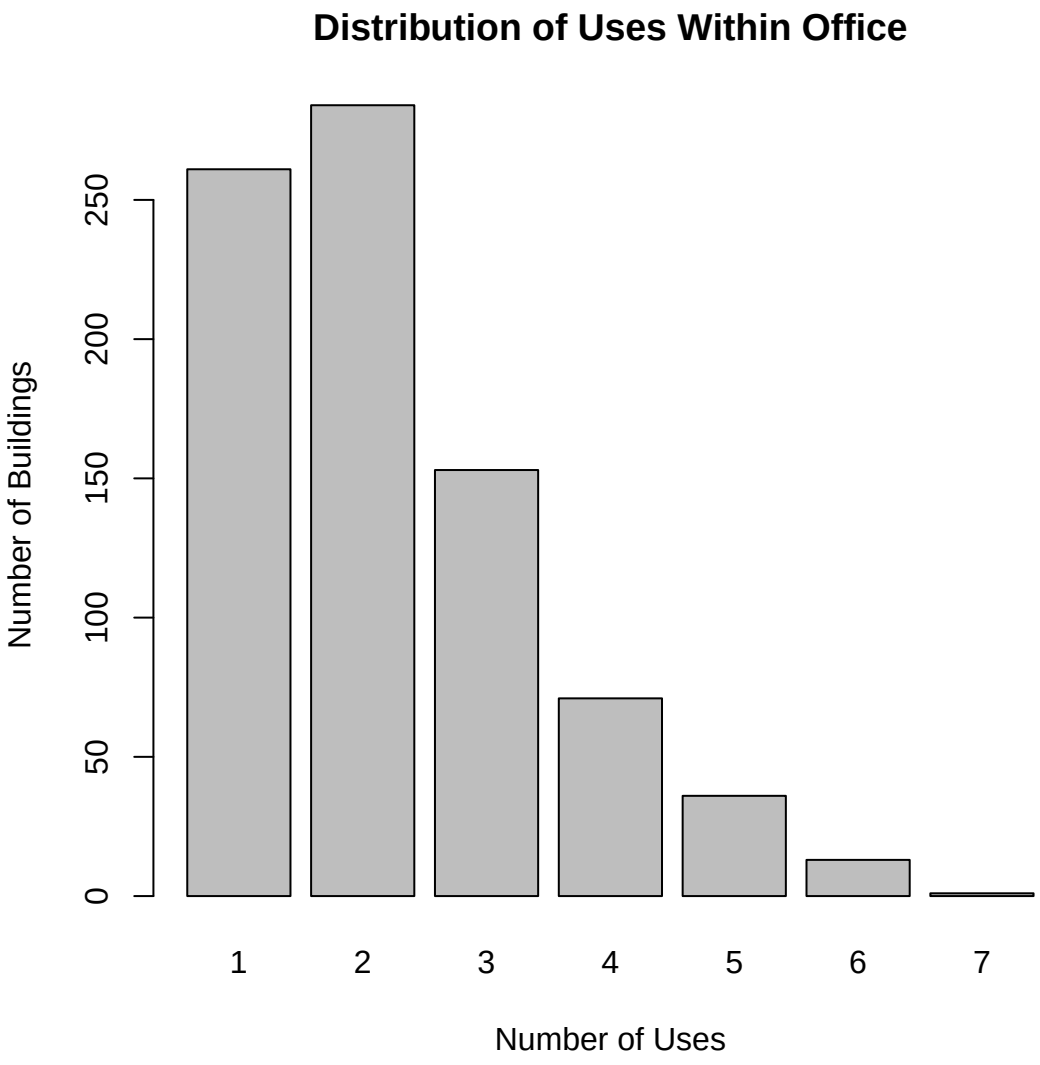


Figure 4.9: Number of Different Uses Indicated in Each Building. In the office sector, multiple uses are quite common.

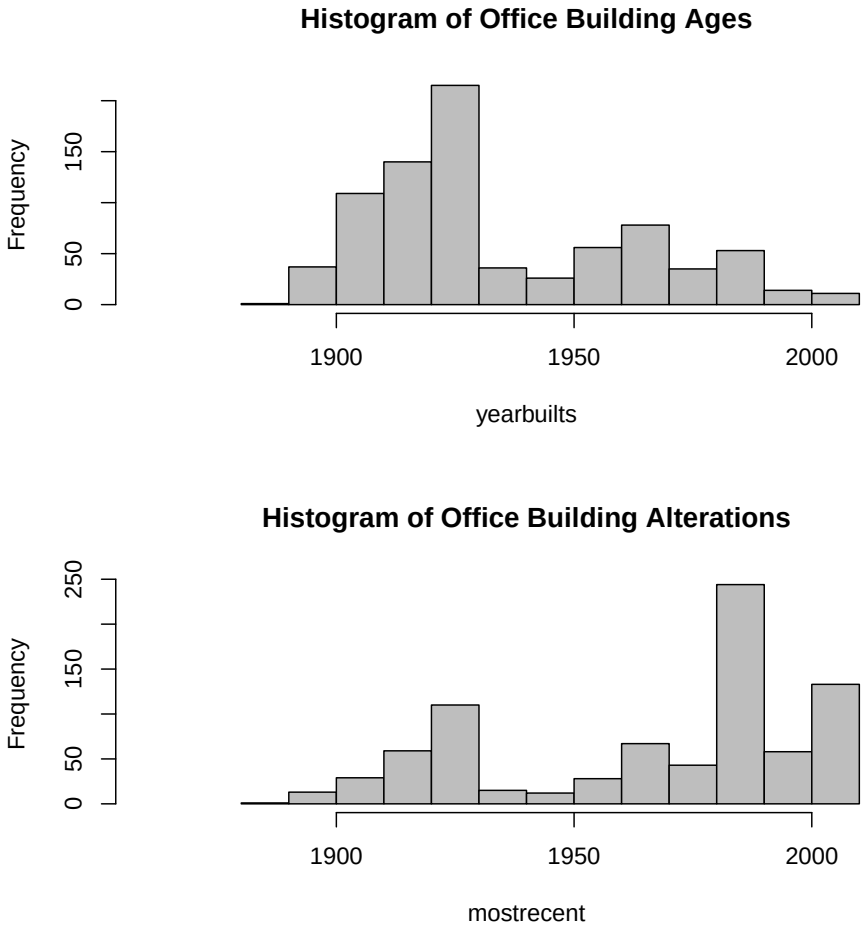


Figure 4.10: Year Built and Alterations from PLUTO

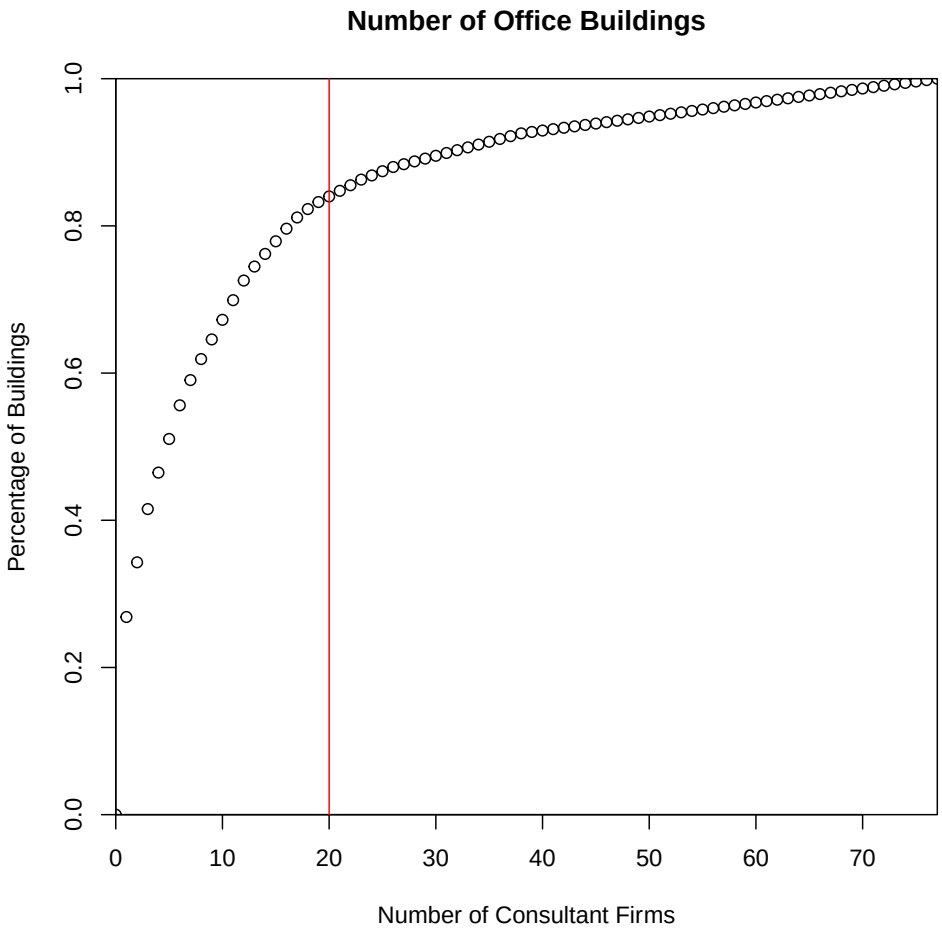


Figure 4.11: Consultants by SF and Number of Buildings. For the approximately 50% of buildings that indicate a consultant, graphs show that 80% of the building square footage and numbers can be reached (and corrected) by contacting the top 20 consultants.

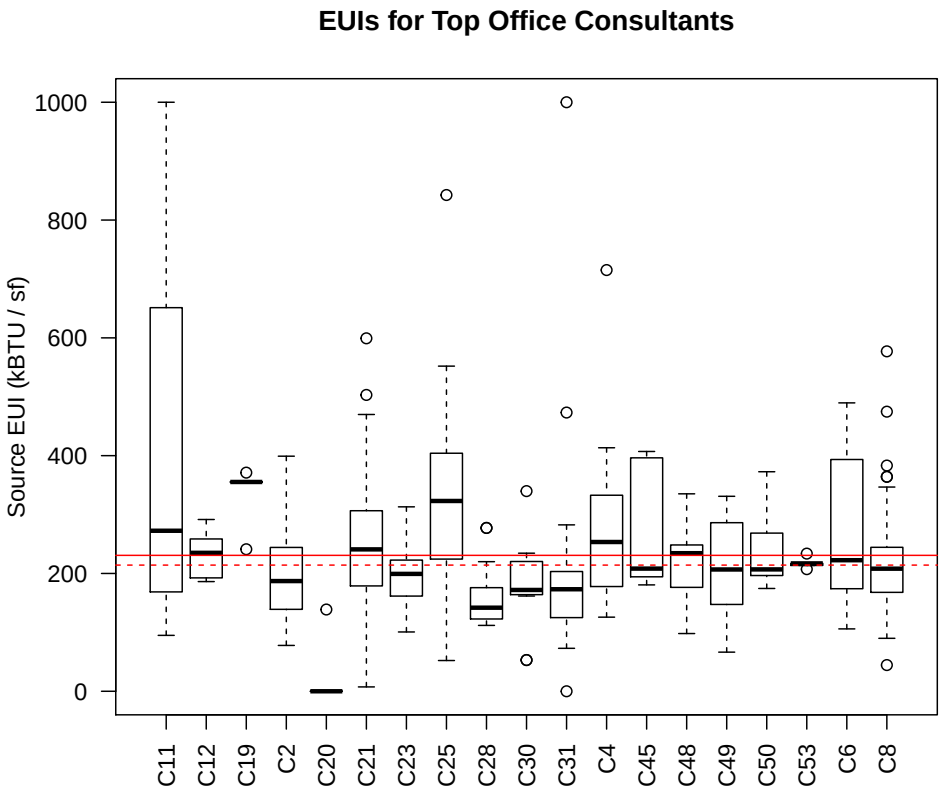


Figure 4.12: EUIs plotted by Consultant. Red solid line indicates mean across all buildings, red dashed line indicates median. Based on original, uncleaned data, though all EUIs greater than 1000 kBTU/sf have been set to 1000.

	Consultant	Number	KSF	MeanEUI	MedianEUI
1	C8	141	47682.4	209.5	208.1
2	C21	39	13542.1	257.9	240.9
3	C6	38	4820.7	258.2	222.5
4	C19	26	4860.9	351.5	355.3
5	C25	24	7868.1	330.7	323.1
6	C31	24	8514.1	207.8	173.2
7	C2	18	2348.3	196.8	187.1
8	C20	15	351.6	9.3	0.0
9	C23	14	3216.5	195.0	199.2
10	C28	14	3889.1	159.3	141.9
11	C30	14	4085.6	179.5	172.0
12	C4	14	1706.4	281.0	253.4
13	C49	10	2279.0	206.1	206.8
14	C12	9	2868.4	231.8	235.0
15	C48	9	5866.3	219.6	234.4
16	C50	9	3430.0	242.2	206.9
17	C11	8	866.8	410.0	272.5
18	C45	6	3826.6	265.8	208.2
19	C53	5	608.8	217.8	217.1

Table 4.5: Consultant Breakdown for Office Buildings. Ordered by number of buildings for each consultant. Based on original, uncleaned data, though all EUIs greater than 1,000 kBTU/sf are set to 1000.

4.3 Other

	Indicated Use	SF	Percent SF
1	Other	21,917,178	63.12
2	Office	5,100,231	14.69
3	Multifamily	1,834,183	5.28
4	Parking	1,804,096	5.2
5	Retail	1,585,752	4.57
6	Hotel	815,612	2.35
7	Dormitory/School	688,654	1.98
8	Medical Office	406,296	1.17
9	Data center	245,914	0.71
10	School	187,667	0.54

Table 4.6: Top 10 Space Uses Within Other Sector. Based on cleaned data.

	Consultant	Number	KSF	MeanEUI	MedianEUI
1	C8	8	1888.1	212.2	195.8
2	C29	7	2280.6	327.9	325.1
3	C4	6	468.5	193.3	237.5
4	C45	6	945.6	277.8	296.0
5	C1	5	575.9	268.7	222.9
6	C30	5	796.4	94.3	33.6
7	C11	4	2202.6	273.8	273.8
8	C25	4	3136.5	822.5	941.9
9	C3	4	1526.0	699.9	699.9
10	C57	3	688.9	29.1	35.4
11	C59	3	762.4	847.9	790.6
12	C68	3	2455.7	769.9	747.5
13	C7	3	644.4	116.2	125.0
14	C9	3	67.5	192.4	0.0

Table 4.7: Consultant Breakdown for Other Buildings. Ordered by number of buildings for each consultant. Based on original, uncleaned data, though all EUIs greater than 1,000 kBTU/sf are set to 1000.

4.4 Hotel

	Indicated Use	SF	Percent SF
1	Hotel	25,721,638	93.65
2	Other	457,801	1.67
3	Dormitory/School	444,301	1.62
4	Office	336,116	1.22
5	Parking	238,111	0.87
6	Multifamily	168,012	0.61
7	Retail	77,046	0.28
8	Grocery	20,361	0.07
9	Bank	1,200	0
10	Hospital	0	0

Table 4.8: Top 10 Space Uses Within Hotel Sector. Based on cleaned data.

	Consultant	Number	KSF	MeanEUI	MedianEUI
1	C4	14	3999.0	187.4	194.9
2	C71	4	1219.5	281.5	258.7

Table 4.9: Consultant Breakdown for Hotel Buildings. Ordered by number of buildings for each consultant. Based on original, uncleaned data, though all EUIs greater than 1,000 kBTU/sf are set to 1000.

4.5 Warehouse

	Indicated Use	SF	Percent SF
1	Other	730,974	48.39
2	Office	475,262	31.46
3	Retail	151,375	10.02
4	Parking	110,745	7.33
5	Warehouse, refrigerated	42,000	2.78
6	Data center	150	0.01
7	Bank	0	0
8	Hospital	0	0
9	Hotel	0	0
10	School	0	0

Table 4.10: Top 10 Space Uses Within Warehouse (Unrefrigerated) Sector. Based on cleaned data.

	Consultant	Number	KSF	MeanEUI	MedianEUI
1	C30	7	695.7	67.9	68.5
2	C8	4	310.4	96.0	83.2
3	C1	3	746.3	75.0	100.1
4	C7	3	241.2	109.9	106.6

Table 4.11: Consultant Breakdown for Warehouse Buildings. Ordered by number of buildings for each consultant. Based on original, uncleaned data, though all EUIs greater than 1,000 kBTU/sf are set to 1000.

4.6 Retail

	Indicated Use	SF	Percent SF
1	Retail	12,466,068	90.79
2	Parking	635,232	4.63
3	Office	350,212	2.55
4	Other	146,973	1.07
5	Multifamily	104,560	0.76
6	Bank	13,140	0.1
7	Grocery	11,628	0.08
8	Medical Office	2,188	0.02
9	Data center	300	0
10	School	292	0

Table 4.12: Top 10 Space Uses Within Retail Sector. Based on cleaned data.

	Consultant	Number	KSF	MeanEUI	MedianEUI
1	C5	3	292.8	281.4	310.4

Table 4.13: Consultant Breakdown for Retail Buildings. Ordered by number of buildings for each consultant. Based on original, uncleaned data, though all EUIs greater than 1,000 kBTU/sf are set to 1000.

Chapter 5

Conclusions

Chapter 2 described the source of the NYC data, as well as the external data that was joined or compared to the NYC data. Section 2.2 described the steps taken to clean the data, using *a priori* knowledge as well as internal comparisons to the NYC data. Section 2.4 described the visual tests used to compare identified control groups from owners and consultants to broader sub-groups within the data, and for both multifamily and office buildings, identified one control group which appears to be very similar to the broader population.

Section 2.3 pointed out a very powerful aspect of the analysis, and of the entire benchmarking process. By performing comparisons across consultants, it was possible to highlight systematic differences between the consultants. This analysis provides valuable feedback to the consultants, and it is hoped that this will greatly improve the quality of data collected next year.

This report is the first description of the overall NYC data; there is much more analysis and description that can be done. After cleaning of the NYC data, Table 2.2 describes the breakdown of buildings by facility type, and the associated source EUI distributions. Other energy performance characteristics are summarized in Section 3.1, and overall building typology data can be found in Section 3.2. Finally, much of this analysis is also repeated for different sub-groups, by facility type, in Chapter 4.