HW2

Dan Fanelli

February 20, 2017

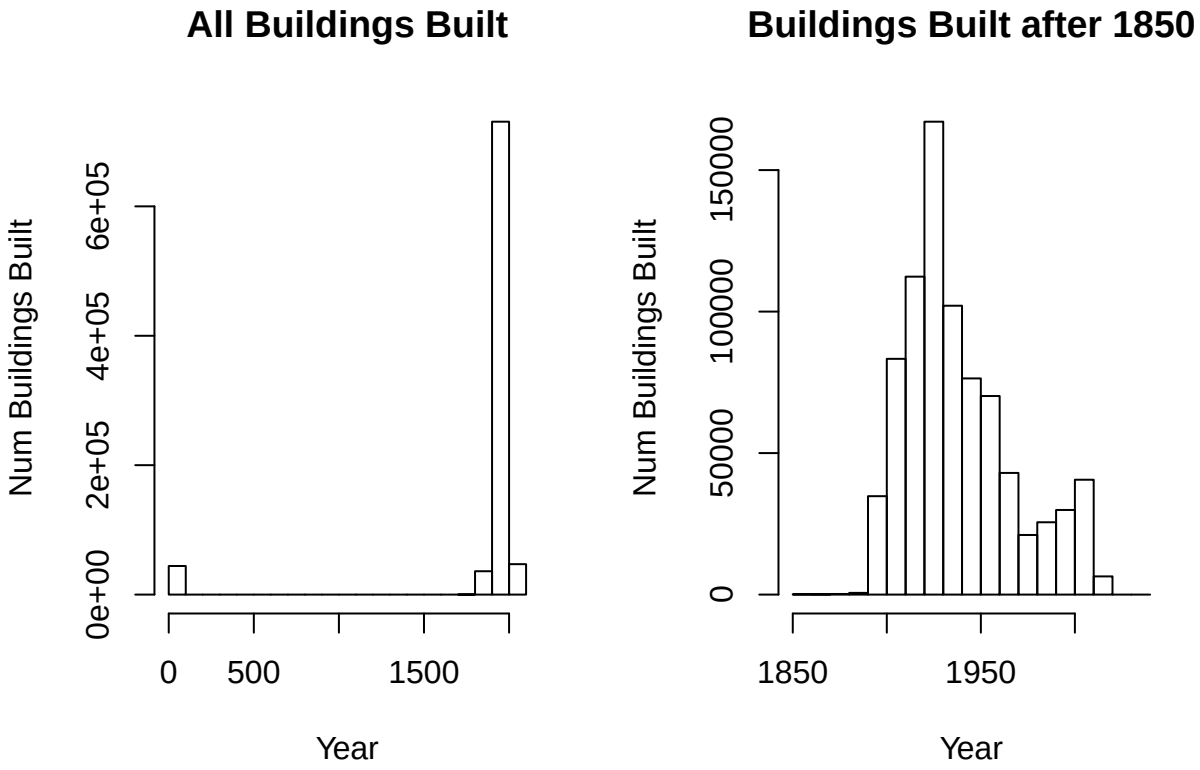
Module 2: NYC Buildings Data Analysis

Table 1: A Look at the NYC Buildings Data

Borough	LotArea	BldgArea	NumBldgs	NumFloors	YearBuilt	YearAlter1	YearAlter2	AssessTot
BK	151930	0	0	0	0	0	0	1164645
BK	19682	154400	1	9	1920	1994	2007	12019950
BK	0	102534	1	12	0	0	0	5864853
BK	387060	0	15	0	0	0	0	0
BK	6384	0	0	0	0	0	0	21150
BK	0	0	0	0	0	0	0	1800

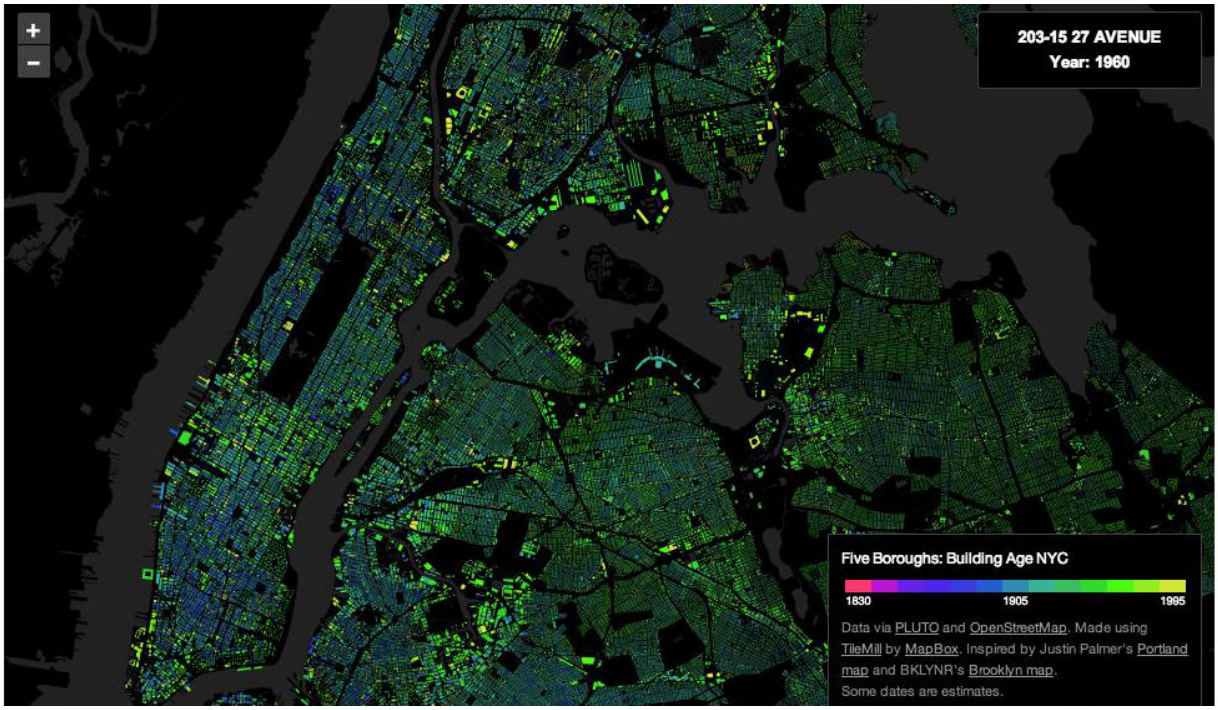
Q1: Accuracy of the Building History Data

Build a graph to help the city determine when most buildings were constructed. Is there anything in the results that causes you to question the accuracy of the data? ? (note: only look at buildings built since 1850)



Seeing such a concentration of buildings built around 1940 at first seems a bit drastic. The image below, taken from here seems to agree with this though, with the majority of NYC having a blue/dark green tone, confirming that the majority of today's buildings were built in that time period. It also makes you see that

there is also a field called "Year Altered", so perhaps the storyline is that the NYC building acreage was all claimed around the 1940s, but then was rebuilt vertically to keep up with demand.



The NYC population by year would probably correlate to the total number of floors by year, and wikipedia's Demographic history of New York City tells us that the NYC population growth did in fact level off around mid-century. (See image below)

Year	Population	White (includes White Hispanics)	% W	Non-Hispanic Whites	% ANG	Black	% B	Asian	% A	Other or Mixed	% O/M	Hispanic/ Latino	% H/L	Foreign born	% FB
1900	3,437,202	3,369,898	98.04	N/A	N/A	60,666	1.76	6,607	0.19	31	0	N/A	N/A	1,270,080	36.95
1910	4,766,883	4,669,162	97.95	N/A	N/A	91,709	1.92	5,669	0.12	343	0.01	N/A	N/A	1,944,357	40.79
1920	5,620,048	5,459,463	97.14	N/A	N/A	152,467	2.71	7,969	0.14	149	0	N/A	N/A	2,028,160	36.09
1930	6,930,446	6,589,377	95.08	N/A	N/A	327,706	4.73	12,972	0.19	391	0.01	N/A	N/A	2,358,686	34.03
1940	7,454,995	6,977,501	93.59	6,856,586	91.97	458,444	6.15	17,986	0.24	1,064	0.01	120,915	1.62	2,138,657	28.69
1950	7,891,957	7,116,441	90.17	N/A	N/A	747,608	9.47	21,441	0.27	6,467	0.08	N/A	N/A	1,784,206	22.61
1960	7,781,984	6,640,662	85.33	N/A	N/A	1,087,931	13.98	43,103	0.55	10,288	0.13	N/A	N/A	1,558,690	20.03
1970	7,894,862	6,048,841	76.62	4,969,749	62.95	1,668,115	21.13	94,499	1.20	83,407	1.06	1,278,630	16.20	1,437,058	18.20
1980	7,071,639	4,294,075	60.72	3,668,945	51.88	1,784,337	25.23	231,501	3.27	761,762	10.77	1,406,024	19.88	1,670,199	23.62
1990	7,322,564	3,827,088	52.26	3,163,125	43.20	2,102,512	28.71	512,719	7.00	880,245	12.02	1,783,511	24.36	2,082,931	28.45
2000	8,008,278	3,576,385	44.66	2,801,267	34.98	2,129,762	26.59	792,477	9.90	1,509,654	18.85	2,160,554	26.98	2,871,032	35.85
2010	8,175,133	3,597,341	44.00	2,722,904	33.31	2,088,510	25.55	1,043,535	12.77	1,445,747	17.68	2,336,076	28.58		

Therefore, aside from data such as NumFloors being zero, it seems that this data is realistic.

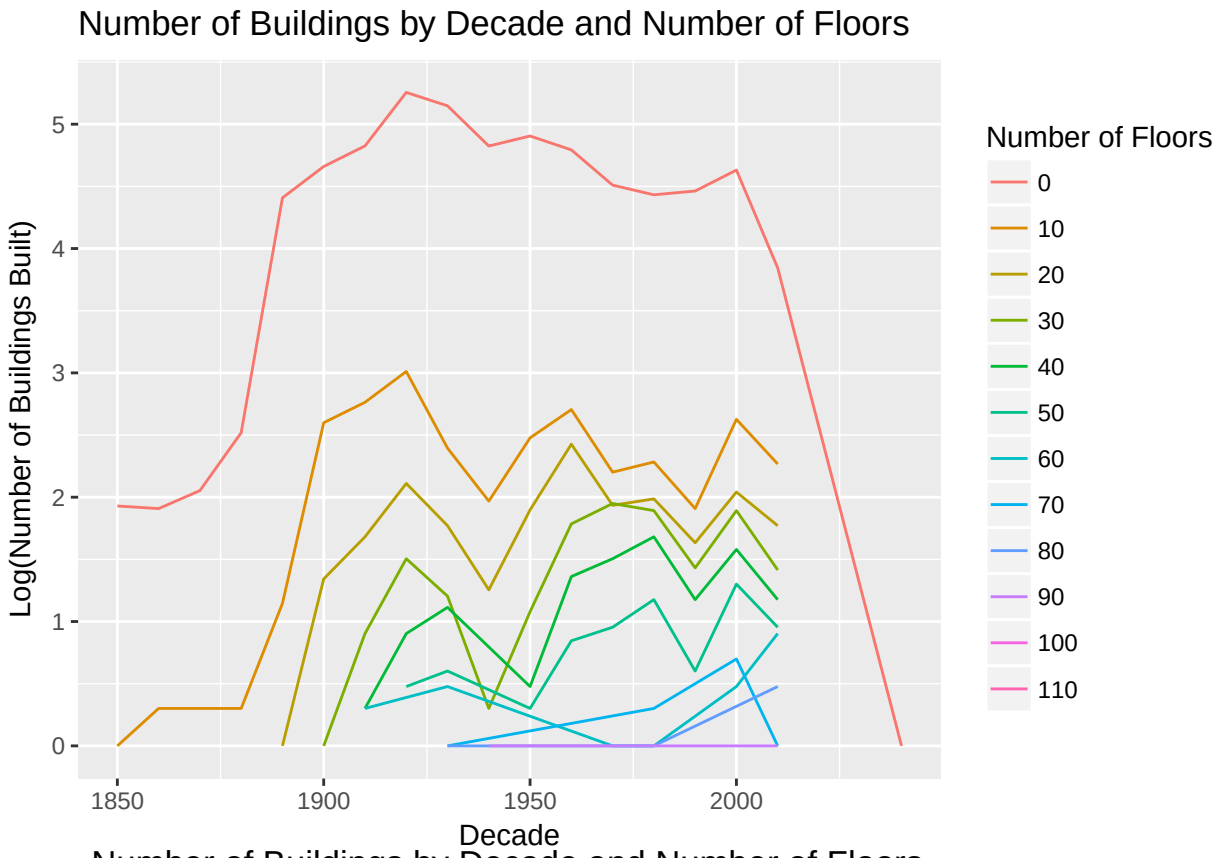
Q2: Number of Floors by Year

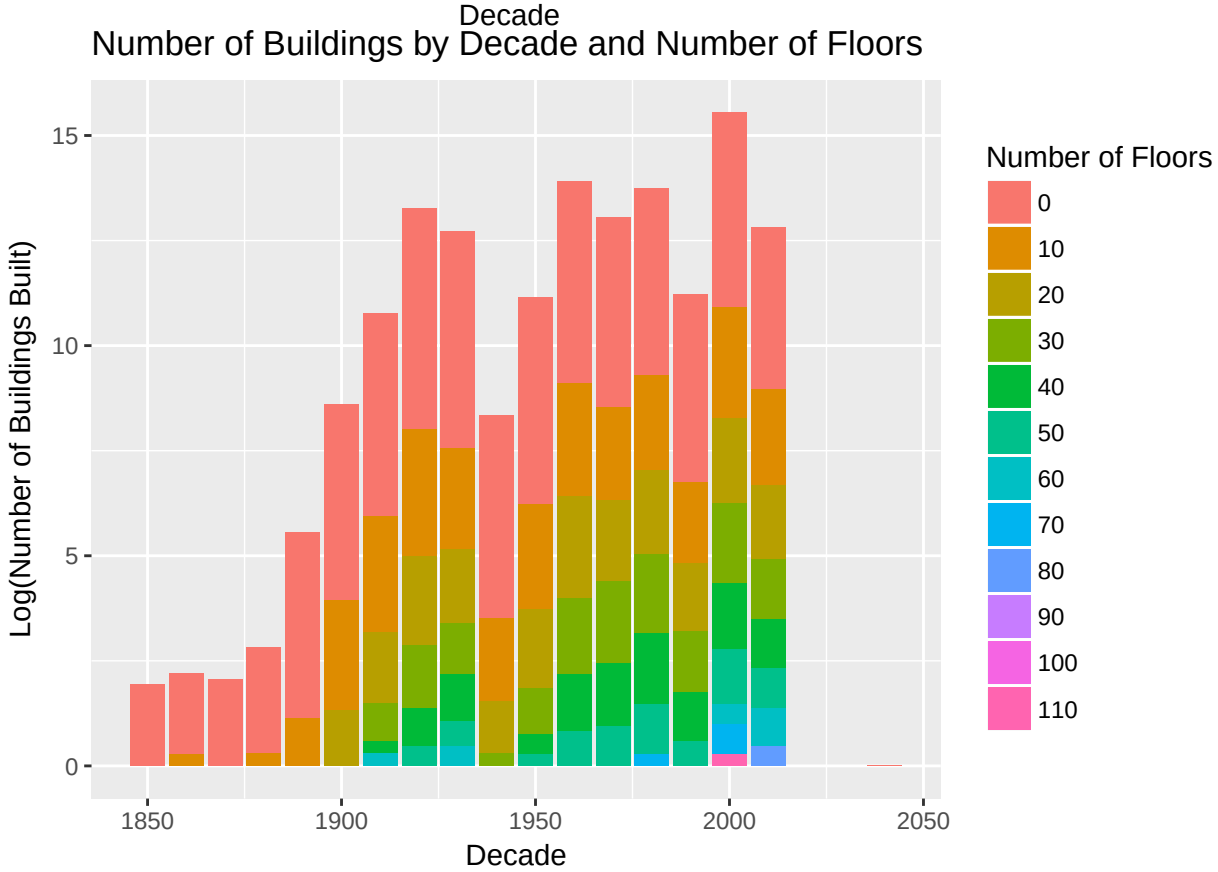
The city is particularly worried about buildings that were unusually tall when they were built, since best-practices for safety hadn't yet been determined. Create a graph that shows how many buildings of a certain

number of floors were built in each year (note: you may want to use a log scale for the number of buildings). It should be clear when 20-story buildings, 30-story buildings, and 40-story buildings were first built in large numbers.

Table 2: Floors by Decade with Log10

	years_tens	floors_tens	the_count	log_count
90	2010	50	9	0.9542425
91	2010	60	8	0.9030900
92	2010	70	1	0.0000000
93	2010	80	3	0.4771213
94	2010	90	1	0.0000000
95	2040	0	1	0.0000000



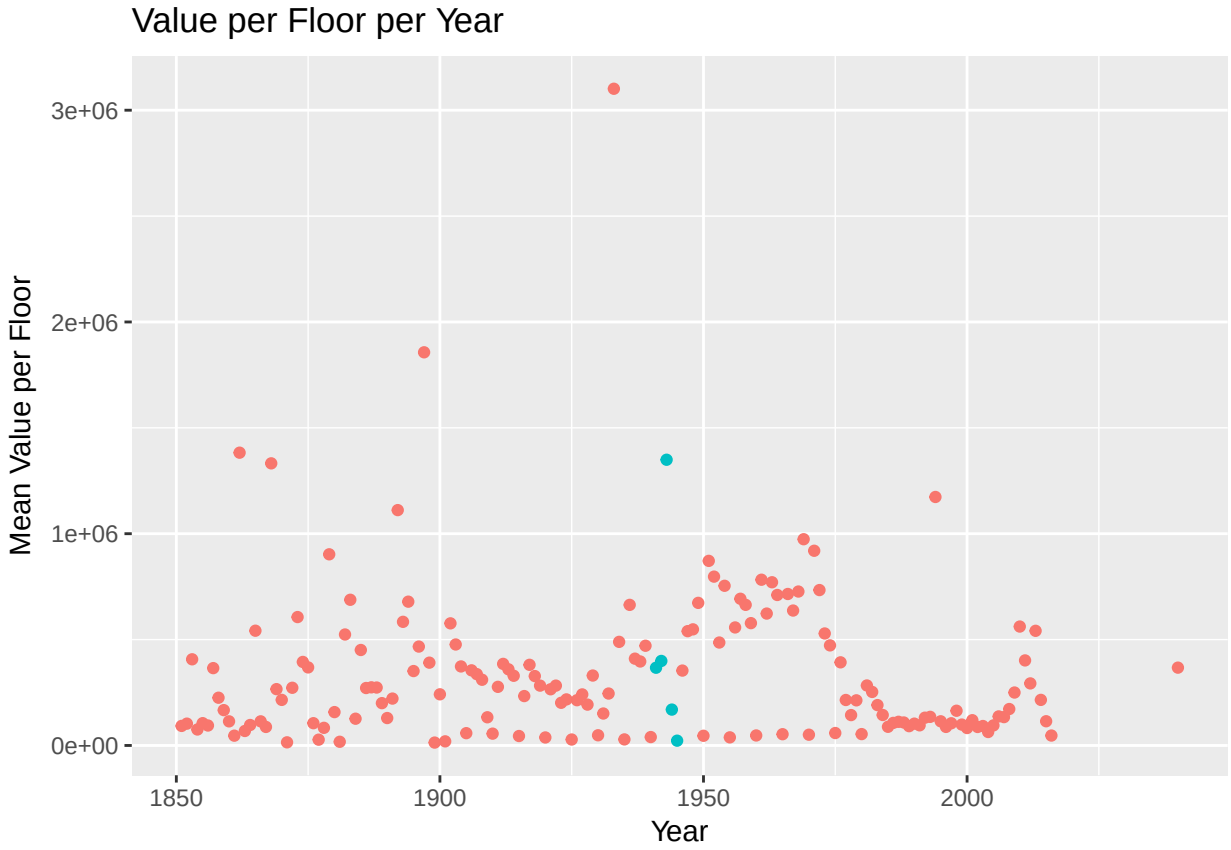


Bear in mind that we have removed the zero floor rows initially from the data, so these zero floors really are

showing us buildings with floors 1-9, and the floor() function translated these to zero, ie less than 10.

Q3: Value per floor per year

Your boss suspects that buildings constructed during the US's involvement in World War II (1941-1945) are more poorly constructed than those before and after the war due to the high cost of materials during those years. She thinks that, if you calculate assessed value per floor, you will see lower values for buildings at that time vs before or after. Construct a chart/graph to see if she's right.



In this case, I would tell my boss that her conclusion is not validated by this data.