

Module 2

We will be working with data about every tax lot in NYC, in a zip file. Please download it from <http://www1.nyc.gov/site/planning/data-maps/open-data/dwn-pluto-mappluto.page>. You will need to combine the data from each boro into one file called 'all_PLUTO_data.R' in order to complete this assignment. I encourage you to do this via code but it is not necessary to submit such code.

- This assignment must be done in a single R script with ggplot2. Use of bigvis is encouraged but not required.
- Please zip your code and images into a file, named 'Last_First_hw2.zip' (mine would be 'Ferrari_Charley_hw2.zip').
- I have created some codes in the sample code file for you- this assumes the data is in a file called 'all_PLUTO_data' which you will need to create (as mentioned above).

Questions

1. After a few building collapses, the City of New York is going to begin investigating older buildings for safety. However, the city has a limited number of inspectors, and wants to find a 'cut-off' date before most city buildings were constructed. 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)

2. 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.

3. 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 way 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.