Methods for Public Land Records Memo

David Hackett

**Data Gathering, Verification, and Cleaning**

The public land records were obtained from the Illinois State Archives from on the ilsos.gov website here: <https://www.ilsos.gov/departments/archives/databases/data_lan.html>. Because the website did not offer the data as a downloadable csv, the records from Cook County were scraped using a headless browser and put into a csv file. To verify the accuracy of the records they were compared to photos of the original documents containing forty land purchases.

In the cleaning process duplicate records for the same lot purchased by the same person were dropped. Additionally, purchases were marked as voided (but not dropped) if they contained a V, VO, or VOID at the end of the Aliquot Parts or Lot column. Finally, purchaser names were matched with other similar names using string matching and then looked over by hand to find names of purchasers that might have been spelled different on purpose or by accident. Rows that seemed to be matches were given a single cleaned name so that they could be grouped for later analysis of individuals.

**Subdividing Sections**

The lots sold were most often sold a quarters or halves of quarter sections. To join these purchases more accurately to geometry from a shapefile, the sections were subdivided into sixteenths, quarters of quarters. The original shapefile came from Cook County Open data: <https://datacatalog.cookcountyil.gov/GIS-Maps/Historical-ccgisdata-Section-2014/jw39-s5cv>, where the sections were already subdivided into quarter sections. These quarter sections were further divided into quarters. Sections that were not a complete mile by mile were extended to be a mile-by-mile section, quartered, then the additional parts of the section were removed for example sections may only have part of a NWNW and SWSW quarter.

**Joining Records to Quarter of Quarters**

While the records gave fairly accurate data in a standardized form down to the section level, the Aliquot Parts or Lot column, which told the quarter or lot within the section that was sold, was not in a standardized format. However, the logic for indicating the part or lot was similar enough in a large portion of the records that with only a small amount of cleaning and few assumptions most of the records were joined with the parts created from subdividing the sections programmatically. For example, a lot with NWSE meant that the land sold was the Northwest quarter of the Southeast quarter and ENE meant the land sold was the East half of the Northeast Quarter.

**Matching Records by Hand**

For the second round of matching, records could not be matched programmatically but could be easily matched by hand were joined to the correct geometry. A digital copy of Mitchell's real estate map of Chicago and suburbs from 1862 was obtained from the University of Chicago Library (<https://www.lib.uchicago.edu/e/collections/maps/chi1900/>). This map was used as a reference for some of the more oddly shaped land purchases. Unfortunately, much of the land had changed hands when the map was made so it was of limited usefulness.

**Matching Records to Blocks and Lots**

The final stage of matching was to match the single lot or block purchases. Lots and Blocks represented the smallest units of land purchased but were the common unit of purchase in the original town of Chicago as well the surrounding parts. For the purchases in the original town of Chicago a plat map of the town from 1836 was used as reference for the lots: <https://en.wikipedia.org/wiki/History_of_Chicago#/media/File:1836_Chicago_Map_by_Mesier.jpg>. For other lots a real estate map of Chicago from 1899 was used: <https://commons.wikimedia.org/wiki/File:1899_Mitchell%27s_new_real-estate_map_of_Chicago.jpg>. In both cases the maps were georeferenced, and blocks were traced based on the maps. For convenience the lots were then created programmatically by splitting the blocks based on the pattern the lots followed in the blocks. This may have resulted in small inaccuracies in both the size and placement of some lots. After this records were matched programmatically by extracting the block number and lot number from the Aliquot Parts or Lot column. In some cases multiple matches were found and the join defaulted to the first match.