

Historical Urban Ecological (HUE) Data Set
Sanitation Infrastructure Inputting Guide

Center for Population Economics, 2013

<http://hue.uadata.org/>

I. Sources

1. In-street Water and Sewer infrastructure data come from annual reports that include tables and/or written descriptions detailing the construction of sewer and/or water distribution pipes and mains that ran under, in, or just alongside city streets. These reports were created and published by city councils as well as various city departments, commissions, and boards. Sanitation Infrastructure coverage metadata documents issued alongside the Sanitation shapefiles detail the year for which partial or full construction data was found and which sources were consulted.

These annual reports were collected from city, county, and state archives, university libraries, historical societies, and from online repositories such as GoogleBooks, Archive.org, and HathiTrust. CPE staff and researchers looked for records covering the period of inauguration of public sanitation systems, which were typically up and running by the mid-1800s, up through 1930. Due to fires, loss, decay, and lack of access, it was not possible to collect complete data for every year for every city, resulting in incomplete coverage.

The majority of the reports used in the creation of this data contain distribution tables which list each main street along which the sewer or water pipe was laid, the two cross streets where the pipe installation started and stopped, and other pipe information (see Table 1 for example). For the few years for which only textual descriptions are available, the infrastructure information in the shapefile tends to be more limited. This is because fewer installations were reported in the absence of distribution tables (because fewer pipes were installed, only certain projects were emphasized, or because the report was incomplete) and any installations that were described too vaguely were not codeable.

Statement of Location, Size, and Number of Feet of Pipe laid in 1881.				
<small>NOTE.—B indicates Boston; S.B., South Boston; E.B., East Boston; B.H., Boston Highlands; D., Dorchester; W.R., West Roxbury; Bri., Brighton.</small>				
In what Street.	Between what Streets.	District.	Size.	Length.
Commonwealth ave.....	Beacon and W. Chester Park	B.	48	116
	Total 48-inch			116
Francis	Tremont and Brookline ave.....	B.H.	30	632
	Total 30-inch			632

Table 1: Example of Distribution Pipe Table

Table quality, where reports are available, is generally good across cities for this time period. Most reports include construction tables with information on the location of new pipes between two

geographic features such as intersections. This style of reporting is the norm for all cities for which the CPE provides in-street sanitation data. Some reports contain more information than others on contracts, pipe materials, and costs. The variables input by the CPE are described in more detail in the following section.

Unfortunately some reports do not contain complete or even partial annual construction tables, further reducing overall coverage. For Chicago Sewers detail dropped off after 1915 when distribution tables were no longer included in the documents and inputters relied on textual descriptions where available. For Chicago Water after 1922, a section of each table titled "Installed by Special Assessment" did not include pipe diameters. Boston Water data are limited or nonexistent for the period 1908–14 as distribution tables are absent in the available materials. For the most part, the Boston Water textual descriptions of pipe additions were descriptive enough to code some new construction though many fewer installations were documented for these years than for years with distribution tables. Annual reports of the Philadelphia Bureau of Water do not include construction tables after 1899. Some reports contain tables that recapitulate several years of construction. These intricacies and omissions are found in reports across all cities and noted where appropriate in the accompanying metadata.

2. During the infrastructure coding process, additional materials were sometimes used to cross check or confirm street names and locations given in the distribution tables. Table 2 lists major resources used for each city:

All cities	http://sanborn1.proquest.com/index.php
	http://www.davidrumsey.com/
Boston	A Record of the Streets, Alleys, Places, Etc. in the City of Boston (Google Books)
	http://www.historicmapworks.com/ - Various district atlases
	http://dca.lib.tufts.edu/features/bostonstreets/cowpaths/
Chicago	http://www.chsmedia.org/househistory/nameChanges/start.pdf
	http://www.rootsweb.ancestry.com/~itappcnc/pipcnstreetfind.htm (database search)
Baltimore	http://mdhistory.net/msaref07/html/
Philadelphia	http://www.phillyhistory.org/historicstreets/
Manhattan	http://www.nypl.org/node/196237
Brooklyn	http://www.nypl.org/node/196237

Table 2: Additional materials

II. Infrastructure Inputting Methodology

In-street sanitation infrastructure data described in collected reports were added to historically accurate street centerline shapefiles created as part of the Historical Urban Ecological (HUE) Data Set. These street centerlines reflect streets in place c.1930, and exhibit the street names used at the time. These streets were planarized and thus form a new segment at each intersection. Data was input segment by segment by a team of trained research assistants. Additional variables and explanatory information were entered mid-way through the project. The most complete and constant variables are those specifying the earliest known/recorded construction of Sewer or Water pipes in a particular street segment. Information on missing variable codes, descriptive statistics, and frequency distributions can be found in the Sanitation Infrastructure Codebook specific to each city.

1. Data input:

- Sewer and Water infrastructure data were coded on separate street centerline shapefiles to eliminate errors resulting from entry into the wrong set of fields (sewer vs water), then later combined into the final HUE Streets product. SewerDate (the date each street segment in the shapefile first received sewer service) and WaterDate were collected for all construction table entries that could be located in or along a street segment or segments and were specified at over 100 feet in length (where noted in the construction tables).
- Additional fields preserved from the Sewer records are as follows: S_PipeLeng (the length of the sewer pipe installed, in feet), S_Pipe_Width (the diameter of the pipe, in feet), S_Inputter (the initials of the person who coded that segment, internal data only), and S_Comments (any additional information about the pipe or information about replacement pipes in subsequent years, internal data only). The fields preserved for Water records are similar: W_PipeLeng, W_Pipe_in (the pipe diameter in inches), W_Inputter, and W_Comments.
- Note: PipeLeng values do not necessarily correspond to the geometric length of a given street segment since a pipe installation often covers more than one segment, or may not cover a full segment from cross street to cross street. These values were recorded to enable the joining of separate segments that were part of a single installation as specified in the construction table.

2. Procedure:

- First, an intersection shapefile was created from the c.1930 street centerline file to expedite the coding process and allow the correct street segments to be located easily.
- The "Find" tool in ArcMap was then used to find the location of the desired segment. When the main street for a particular record was searched, the resulting list included all streets intersecting that main street.
- The two cross streets for the record were located in the list and the segment(s) of main street between them were centered in the map viewer.

- All segments of the main street between those cross streets were selected.
- The fields described above were populated using the data in the infrastructure report (not all fields were populated during all stages of data inputting)

Exceptions that came up during infrastructure coding were handled as follows:

- *Main or cross street name did not match any streets in the shapefile:* Using the district/neighborhood listed on the table to narrow the options, the street was located on a historic map and/or the street name was searched in a street name change table. Maps were also used to resolve issues when two streets did not intersect as expected. In areas where significant change had occurred in the streets, historic maps from a year close to the report document's publish date could be used to clarify intersections that varied in the shapefile.
- *A street located on a map was completely missing from the shapefile:* The street lines were drawn based on maps contemporary to 1930. Due to the quality of available historic maps as well as the scope of the project, some small streets and all alleyways were not drawn in the shapefile. Furthermore, streets that appeared for only a brief time in history or were built over prior to 1930 may not appear. Any records in the distribution table that referenced a completely missing street were not coded. Records where one part of the street was missing from the shapefile are denoted with "YYYY/BB" in the comments field. See below for complete descriptions of the quality codes.
- *The pipe length (as specified in the construction table) was less than 100 feet:* The segment was not coded. Most street segments in the shapefile are significantly longer than 100 feet; therefore these small records (which are frequently pipes laid between alleys or small repair segments) were judged to be less significant to the project as a whole and were not searched out.
- *A segment was already coded for a previous year:* In most cases, the segment's comments field was edited to reflect the data for all subsequent installations using the "YYYY/AA" quality code. However, the date of initial pipe installation for any particular street segment was not changed.
- *Instead of two cross streets, the second column of the distribution table lists "From XXX street" or "At XXX street":* If the main street (first column) began/ended at the listed cross street, the segment was coded. If the main street continued on either side of the cross street, not enough information was available to code the pipe. The "At XXX street" records tended to have lengths shorter than 100 feet (in which case they would not be coded) or occasionally, existed between two sides of a divided boulevard and would be coded. (For similar records "East/West/North/South from XXX street", the coding was possible except in cases where the main street did not run in the direction specified.)
- *Instead of a cross street, a neighborhood line ("Newton Line", "Dedham Line", etc) was specified:* If the line could be located on a chronologically close historic map or the line segment in the shapefile near the edge of the city ended approximately where the neighborhood border should be, the installation was coded.

- *Instead of a cross street, a building or landmark ("Asylum", "Mr. XXX's House", "Theatre") was specified:* In the case of some landmarks or buildings, the geographic coordinates of the site could be determined from other online records. Otherwise, the record was not coded.
- *Pipe ends at railroad/water/alley that does not exist in the shapefile/"XX feet from cross street":* The full segment was coded (no segments were split or merged during the infrastructure coding process) and the appropriate quality code was placed in the comments field. If a railroad/railroad bridge was specified but the other cross street was missing or not locatable in the shapefile, the installation could not be coded.
- *One important piece of information (length, width) was missing from the distribution table:* See quality code "8888".
- *Multiple lengths were listed in the distribution table/report for a single pipe:* If the multiple lengths both corresponded to a single pipe diameter, they were added together and coded as a normal width/length pair. If the multiple lengths corresponded to different pipe widths, they were entered in the comments field in a "repeat covering of street segment during the same year" quality code ("YYYY/MM/width1/length1/width2/length2") and the pipe information fields, if not already coded, were populated with either one of the width/length sets if not already coded (otherwise, a "YYYY/AA" quality code was used in addition to the "YYYY/MM" code). Exception- if one of the lengths was less than 100 feet, it was largely disregarded.

3. Quality of infrastructure coding:

- *Missing pipe construction* was judged to be preferable to incorrect pipe coding in cases where the descriptions were vague and incomplete.
- *Coverage.* Lower coverage in certain areas of a city may be due to the city's infrastructure timeline- the areas where infrastructure construction began are likely to have the best overall coverage by the end of the time period-but may also be a result of major differences between the shapefile streets and the streets of various historical maps, especially if a lot of change and expansion occurred during the time period. The Hyde Park area of Boston, for example, underwent significant development and, furthermore, no detailed historical atlases existed for that neighborhood. As a result, when street names in the records did not match street names in the shapefile, most could not be coded.
- *Partial coverage of street segments.* The data do not always cover the entirety of every coded street segment. The lengths specified in the construction tables might only cover just over 100 feet of a particular segment, or might stop 10 feet short of the terminal intersection. As such the data are not appropriate for determining flow dynamics or overall network connectivity. Where the data explicitly indicated that that construction fell short of completing a specific block in question that segment was coded using the "NN" comment code (described in more detail later). Not all segments coded for every city received this code, however. Where early termination was not explicitly stated it was assumed that the pipe in question ran the complete length of the segment.

- *House connections.* Comprehensive information on individual house connections were not available for all cities for which sanitation infrastructure was input, nor were records found adequate to describe the timing of all house connections across the city. The installation of pipes denotes *the earliest possible availability* of access to the public sanitation system, but not the exact date of entry into that system
- *Pipes not located in or adjacent to street segments.* If a sewer or water pipe was not laid in relation to a street and thus its construction could not be attributed to a particular street segment, it was not included in the sanitation infrastructure data.