Shenhao Wang

March – May, 2021

Draft 1

Integrated Smart Chicago Data Set

# Motivation and Background

Data is the fuel to research. However, researchers are increasingly limited by data availability. When the big data era calls for innovative use of data sets, the private industries hold much more data than researchers. But on the other side, a large number of open data sources and data collection tools are created across various channels, presenting unprecedented opportunities for researchers to take advantage of.

This work is designed to improve the author’s knowledge about big data sources and build the capacity of collecting various data, using Chicago as a specific example. It targets to collect a *complete* data set for Chicago, enabling the author to push to the boundary of the publicly available data sets in a specific area. Due to the author’s background, the author pays particular attention to the data sets related to human dynamics & urban mobility, along with the pertinent topics such as energy consumption, air pollution, economic growth, and other impacts of human behavior. The final goal is a data dictionary of the Integrated Smart Chicago Data Set, which can be the foundation for a multitude of research projects. The present draft has incorporated the contents, resources, and remaining questions for 14 different data sources, including (1) meta data set, (2) socio-demographics from Census, ACS, and CTPP, (3) GIS shapefiles, (4) mobility flow, (5) spatiotemporal ridership, (6) general transit feed specification (GTFS), (7) general road network flows, (8) travel survey, (9) images, (10) OpenStreetMap and Point-of-Interests, (11) mobility-related data (COVID, air pollution, energy), (12) web scraping data, (13) private data providers, and (14) others.

# 1 - Meta data sets

Universities, research centers, governments, and individual researchers have collected useful meta information about data sources. Resources include: data libraries from **Northwestern University**[[1]](#footnote-1), data sets by complex and sustainable urban networks (CSUN) lab from **University of Illinois at Chicago**[[2]](#footnote-2), GIS data by **University of Illinois**[[3]](#footnote-3), **City of Chicago Data Portal**[[4]](#footnote-4), **Illinois DOT Open Data Portal**[[5]](#footnote-5), **US Bureau of Transportation Statistics**[[6]](#footnote-6), **US DOT data inventory**[[7]](#footnote-7), **NREL Transport Data Center**[[8]](#footnote-8), and personal websites by **researchers**[[9]](#footnote-9). **Papers** and agents’ **modeling reports** are also entry points to data sets[[10]](#footnote-10).

# 2 - Socio-demographics and commuting data from census, ACS, and CTPP

Contents. **Census** and **American Community Survey (ACS)** can provide the socio-demographics for counties, census tracts, census block groups, and census blocks. Census is conducted every ten years, and ACS has its 1-, 3-, and 5-year indicators. There exists a tradeoff between timeliness and spatiotemporal resolution. Higher the spatiotemporal resolution is inevitably associated with untimeliness: e.g. the infrequent decennial census data have data at census block level, but ACS typically has information at census tracts or block groups level. Census and ACS provide information such as age, gender, income, population, race, housing units, among many others. Based on Census and ACS, State DOTs also fund a cooperative program – **Census Transportation Planning Products (CTPP)** – for transport community. CTPP provides information for home, work locations, and travel flows at relatively high granularity (e.g. down to Census tracts and TAZs) for 2006-2010 and 2012-2016[[11]](#footnote-11),[[12]](#footnote-12). **National Household Travel Survey 2017[[13]](#footnote-13)** is the source for general travel behavior in the US, although its sampling did not cover the whole US. The NHTS 2017 data have about **six state add-on’s** at the census tract level, including **Arizona, California[[14]](#footnote-14), Wisconsin, Georgia,** and **Iowa**[[15]](#footnote-15). For the state add on’s, researchers can apply for geospatial data more granular than the census tract level (e.g. **California[[16]](#footnote-16)**). Since NHTS2017 does not cover the whole US, the **Bureau of Transportation Statistics (BTS)** estimated trips and vehicle miles (**LATCH Survey**[[17]](#footnote-17)) for all the census tracts of the US based on the NHTS2017 data.

Resources. It is more efficient to download the data sets using R and Python modules than directly from Census websites. The coding modules include **Python CensusData**[[18]](#footnote-18),[[19]](#footnote-19) and **R tidycensus**[[20]](#footnote-20). Please avoid directly downloading data sets from Census Bureau[[21]](#footnote-21),[[22]](#footnote-22),[[23]](#footnote-23),[[24]](#footnote-24), which does not provide a friendly user interface. The high-level census summary for every state (e.g. Illinois) can be found in the **FTP**[[25]](#footnote-25),[[26]](#footnote-26) by Census Bureau. CTPP data can be downloaded from the **official software**[[27]](#footnote-27) or the **R modules**[[28]](#footnote-28) on Github. The most recent 2020 Census data will be released on April 30, 2021. Stay tuned.

Remaining questions and TBD. (1) Explore the public use microdata from Census. (2) Explore the micro data from CTPP. (3) Why was CTPP stopped after 2016?

# 3 - GIS Shapefiles

Contents. GIS shapefiles are provided by census: **Topologically Integrated Geographic Encoding and Referencing (TIGER)**, which includes the typical hierarchy of census tracts – block groups – blocks. Besides this typical hierarchy, other spatial units can be based on congressional districts, schools, legislative districts, zip codes. Important GIS shapefiles also include the transport and land use systems, such as the road networks, transit stations and lines, building footprints, land use parcels, natural resources.

Resources. TIGER/Line GIS shapefiles can be downloaded from **Census Bureau** with or without the demographic data[[29]](#footnote-29),[[30]](#footnote-30). The important indicators are the GEOIDs, e.g. 11-digit Census Tract FIPS[[31]](#footnote-31). See also the GIS files provided by **universities**, **governments** (**Illinois**), and **City of Chicago data portal, Illinois DOT data portal**[[32]](#footnote-32),[[33]](#footnote-33),[[34]](#footnote-34),[[35]](#footnote-35). The shapefiles of transport networks (e.g. bus, transit, and Metra stations and lines; bike routes and racks[[36]](#footnote-36)) can be found on **City of Chicago Data Portal**[[37]](#footnote-37). For other GIS resources (e.g. buildings, land use, water, natural resources, etc.), please refer to the **OpenStreetMap** section.

Remaining questions. None.

# 4 - Mobility flow

Mobility data (From 4 to 8) always involves the tradeoff between spatial scope, spatial resolution, and temporal resolution. Large spatial scope (e.g. whole US) is often associated with low spatiotemporal resolution (e.g. cross-sectional data + spatial units above census tracts). High spatial resolution (e.g. individual or census blocks) is often associated with small spatial scope (e.g. limited to one city and one travel mode) and/or low temporal resolution (e.g. cross-sectional data). High temporal resolution (e.g. real-time) is often limited to one specific travel mode (e.g. taxi) in a specific area. This tradeoff holds for all the human dynamics data resources.

The mobility flow data refer to the network data sets that include the origin, destination, and the flow counts between OD pairs.

Resources. Two major public resources exist, and both are largely cross-sectional data. The first is the **ACS commuting** data set provided by CTPP[[38]](#footnote-38)[[39]](#footnote-39), which covers the whole US at the census tract level. The data can be efficiently downloaded from the websites on the published papers[[40]](#footnote-40). The second is the **LEHD Origin-Destination Employment Statistics (LODES)**[[41]](#footnote-41),[[42]](#footnote-42), which provides the commuting OD flows for each state at the census block group level, along with work and residential details. LODES is a panel data, collected for every year – the most recent data was the 2017 commuting data, released in Dec 2020. Lastly, a recent Scientific Data paper provides a **multiscale dynamic human mobility flow data set** (daily average) for mobility at the census tract level after the eruption of COVID[[43]](#footnote-43). Besides, the OD flows can be constructed by using travel survey, spatiotemporal ridership, and data sets from private companies.

Remaining questions and TBD. (1) Explore other public large-scale OD flow data sources. (2) What does the LODES data look like? (3) How to obtain a OD matrix for only public transit?

5 – Spatiotemporal ridership

Contents. The spatiotemporal ridership data refer to the **real-time individualized mobility** data sets. The data sets often include the **vehicle ID, departure and arrival time, OD geospatial information** (e.g. census tracts), and others for the travel modes of **bus, subway lines, TNCs, micro-mobility, Divvy, and taxi**.

Resources. They are provided by the **City of Chicago data portal**, for a variety of travel modes, including buses, subway lines, Metra, TNCs[[44]](#footnote-44), micro-mobility[[45]](#footnote-45),[[46]](#footnote-46),[[47]](#footnote-47),[[48]](#footnote-48), Divvy[[49]](#footnote-49), and taxi[[50]](#footnote-50) in the City of Chicago. The real-time bus and subway lines’ ridership is provided by the **collaboration with CTA**.

Remaining questions and TBD. (1) How can we get the individualized spatiotemporal ridership of automobiles?

6 – General Transit Feed Specification (GTFS)

Contents. [TBD].

Resources. The raw GTFS data for CTA and Metra can be found on **OpenMobilityData website**[[51]](#footnote-51),[[52]](#footnote-52). General intro is from Google[[53]](#footnote-53). The most efficient way to download the data set is through the **modules on Github**[[54]](#footnote-54),[[55]](#footnote-55).

Remaining questions and TBD. (1) Explore the exact information in GTFS data set by using the Github modules.

7 – General Road Network Flows

Contents. The typical OD matrix and spatiotemporal ridership data sets are typically not attached to road segments, however, the broad travel management often needs road-specific information. The real-time road-specific information includes **speed**, **congestion**, and **vehicle counts at high spatiotemporal resolution**.

Resources. **City of Chicago Data Portal** provide the traffic congestion estimate for road segments by using the GTFS data[[56]](#footnote-56),[[57]](#footnote-57). **Array of things** provide the data set from 2016 to 2020 by using computer vision techniques from their camera[[58]](#footnote-58). **Uber Movement** also provides high-quality segment-specific information, but unfortunately not for Chicago area[[59]](#footnote-59). **Illinois DOT** provides the traffic counts for the state highways[[60]](#footnote-60), but not the high-resolution road counts in Chicago.

Remaining questions. (1) How to obtain the high-quality network flow data specific to road segment in the urban Chicago area?

8 – Travel Survey

Contents. Travel survey often provides rich cross-sectional data set, which is loosely defined since one-day or one-week travel survey is also common. Travel survey typically includes a complete set of household, person, place, and vehicle attributes. For the travel part, people take the online survey to report their travel modes, OD locations, activity patterns, trip durations, and other mobility variables. In the Chicago case, GPS trackers are used to obtain these data, and transit price information is also offered.

Resources. **Chicago Metropolitan Agency for Planning (CMAP)** implemented a travel survey during 2018-2019[[61]](#footnote-61),[[62]](#footnote-62). To combine alternative-specific variables with travel survey, you need to use **Google Distance Matrix API**.

Remaining questions and TBD. (1) Need to know more specifics about the travel survey. (2) Use the CMAP survey to create the images + numeric data for human dynamics.

9 – Images

Contents. Urban images can be aerial images, street view images, land use images, and nightlight images. Typically they have the coordinate information. Other images are the images from the Websites, such as house images from Zillow, room images from Airbnb, etc. Check the Web-scraping section for more details.

Resources. **Google Maps Static API** can be used to collect aerial images[[63]](#footnote-63), and **Google Static Street View API** can be used to collect street view images[[64]](#footnote-64). Both have some Github tools to facilitate the use[[65]](#footnote-65)[[66]](#footnote-66). Imagery information has been widely used in Remote Sensing, Geography, and many other communities. See also **U.S. Geological Survey (USGS) Earth Explorer**[[67]](#footnote-67), **Google Earth**[[68]](#footnote-68), **Landsat-8 from NASA Earth Observing Systems**[[69]](#footnote-69),[[70]](#footnote-70), **VIIRS from National Centers for Environment Information**[[71]](#footnote-71), University of Oregon libraries[[72]](#footnote-72), and **Xiaohu Zhang’s github repository**[[73]](#footnote-73). See also the papers from the remote sensing and geography communities.

Remaining questions: (1) How to collect high-resolution nightlight images for Chicago area?

10 – OpenStreetMap (OSM) and Point-of-Interests (POIs)

Contents. OSM and POIs are related, but OSM is a data source, while POIs are the data. OSM provide information about road networks, rail networks, nature, water bodies, land use, building footprints, places, and POIs.

Resources. **OSM** itself is a resource[[74]](#footnote-74)[[75]](#footnote-75). While officially we should use **Overpass API** to download information from OSM, researchers have provided tools to facilitate downloading a large amount of data from OSM, including **BBBike**[[76]](#footnote-76), **GeoFabric**[[77]](#footnote-77), and **Overpass-Tubo**[[78]](#footnote-78). However, both Anson Stewart and Justin Anderson told me that the POIs from OSM are not representative. Justin recommended **Google Place API** and **Foursquare**, while Anson recommended checking **Chicago Cityscape**[[79]](#footnote-79)[[80]](#footnote-80) for more details.

Remaining questions. None.

11 – Mobility-Related Topics

Contents. The topics related to human dynamics include **COVID, energy, public health (e.g. air pollution),** and **natural disaster**. These are very broad topics, so I only provide a nutshell about the resources.

Resources. **City of Chicago** provides the vaccine and infection information by Zip codes[[81]](#footnote-81). See also the **multiscale dynamic human mobility flow** paper and **SafeGraph** for mobility under COVID[[82]](#footnote-82). The gas and electricity of buildings are provided by **City of Chicago**[[83]](#footnote-83). See also a recent **Nature Energy paper**, which uses **NHTS data set**[[84]](#footnote-84). **Retail electricity sales** data regarding companies can be found online[[85]](#footnote-85). **Array of Things (AoT)** provide the air quality and the vehicle-pedestrian counts for 4 years with relatively real-time information in Chicago[[86]](#footnote-86),[[87]](#footnote-87), however, Madelaine suggests that the AoT sensors did not work as expected, and **Microsoft** will implement large-scale sensors starting in June, 2021. Besides, **Centers for Disease Control and Prevention** (CDC) conducted a **500 Cities Project (2016-2019)[[88]](#footnote-88)** to measure health outcomes at the census tract level, including tens of health outcomes, unhealthy behavior, and preventions. This 500 Cities Project evolves into the project of **PLACES** in 2020[[89]](#footnote-89). Some flooding information can also be found on **City of Chicago data portal**[[90]](#footnote-90).

Remaining questions and TBD: (1) How to obtain travel-energy data? (2) What is in the retail electricity sales data?

12 – Web scraping data

Contents. Web scraping skills can help to collect the information from websites, such as **Yelp, Bookings, Zillow, Twitter, Amazon reviews, Cars, Airbnb, Wikipedia**, among many others. The possibility is infinite – as long as we can see the information on the web pages, theoretically we can collect the data. For example, we are able to collect the **tweets and likes from Twitter**, **price and images from Zillow**, **comments and stars from Yelp**, etc.

Resources. Most of the websites provide APIs to facilitate data collection, or sometimes directly example data sets. Examples include **Zillow GetSearchResults API**[[91]](#footnote-91), **Twitter developer API**[[92]](#footnote-92), and **Yelp open data sets**[[93]](#footnote-93). Researchers have also attempted to provide efficient **modules on** **Github** or **Medium** to facilitate web scraping, such as **Zillow data**[[94]](#footnote-94),[[95]](#footnote-95), **Twint package**[[96]](#footnote-96), etc. Since websites change their structures regularly, each individual web scraping module can only work for a relatively short period of time. Some web scraping techniques can be found on research papers using them (e.g. Yelp[[97]](#footnote-97)). In addition, there are professional web scraping groups (e.g. **Scrapehero**[[98]](#footnote-98)) to provide professional suggestions and services for web scraping, and web scraping softwares (e.g. **WebHarvy**[[99]](#footnote-99)) with visual interface to facilitate individual data collection efforts. The professional websites teach how to scrape in a large scale without being detected[[100]](#footnote-100),[[101]](#footnote-101).

Remaining questions and TBD. (1) What is the general web scraping rules to follow? (2) How to scrape both current and historical data from Twitter, Yelp, and Zillow? (3) How to scrape the images from the websites (e.g. Yelp, Zillow, and Airbnb)? (4) How to build up a virtual network from Twitter?

13 – Private data providers

Contents. With collaboration, private data providers can give us the data sources that other people to not have access to. Typically the data are the **individual-level trajectory data**, which does not exist in the data sources above.

Resources. Important data providers include **SafeGraph** and **Cuebiq** for individual mobility data, **Womply** for the individual credit card and transaction data, and **Mastercard** for individual transaction data. A Chicago-specific data source is called **Chicago Cityscape**, which is partially open to the public.

Remaining questions and TBD. (1) What is the exact comparative advantage of the data sets in Boston and Chicago from the lab? (2) What do the data sets look like from Cuebiq, Womply, and Mastercard? Do they have the individual socio-demographics? (3) What is in the Chicago Cityscape data set?

14 – Others

Contents. Other data sets include **weather**[[102]](#footnote-102),[[103]](#footnote-103) and **call detailed records (CDR)** data.

Resources. The weather data is provided by the **National Centers for Environmental Information**. I have not identified any resource for CDR data in Chicago (only Milan[[104]](#footnote-104),[[105]](#footnote-105),[[106]](#footnote-106)).

Remaining questions. (1) How to collect high-resolution spatiotemporal weather information? (2) Is it possible to get CDR data without private collaboration? (3) What happened to city of Milan, where the CDR data is open to the public?

1. <https://libguides.northwestern.edu/c.php?g=114808&p=748275> [↑](#footnote-ref-1)
2. <http://csun.uic.edu/datasets.html> [↑](#footnote-ref-2)
3. <https://www.uis.edu/gis/projects/data/> [↑](#footnote-ref-3)
4. <https://data.cityofchicago.org/> [↑](#footnote-ref-4)
5. <https://gis-idot.opendata.arcgis.com/> [↑](#footnote-ref-5)
6. <https://www.bts.gov/browse-statistical-products-and-data> [↑](#footnote-ref-6)
7. <https://www.transportation.gov/data> [↑](#footnote-ref-7)
8. <https://www.nrel.gov/transportation/secure-transportation-data/> [↑](#footnote-ref-8)
9. <http://urban-computing.com/yuzheng> [↑](#footnote-ref-9)
10. <https://cmap-repos.github.io/cmap_abm_report/#1> [↑](#footnote-ref-10)
11. <https://www.fhwa.dot.gov/planning/census_issues/ctpp/> [↑](#footnote-ref-11)
12. <https://ctpp.transportation.org/> [↑](#footnote-ref-12)
13. <https://nhts.ornl.gov/> [↑](#footnote-ref-13)
14. <https://nhts.dot.ca.gov/> [↑](#footnote-ref-14)
15. <https://www.nrel.gov/transportation/secure-transportation-data/tsdc-cleansed-data.html> [↑](#footnote-ref-15)
16. <https://nhts.dot.ca.gov/> [↑](#footnote-ref-16)
17. <https://www.bts.dot.gov/latch> [↑](#footnote-ref-17)
18. <https://jtleider.github.io/censusdata/> [↑](#footnote-ref-18)
19. <https://github.com/jtleider/censusdata> [↑](#footnote-ref-19)
20. <https://cran.r-project.org/web/packages/tidycensus/tidycensus.pdf> [↑](#footnote-ref-20)
21. <https://data.census.gov/cedsci/all?g=0500000US17031.100000> [↑](#footnote-ref-21)
22. <https://www.census.gov/programs-surveys/geography/technical-documentation/records-layout/tiger-line-demo-record-layouts.html> [↑](#footnote-ref-22)
23. <https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-data.html> [↑](#footnote-ref-23)
24. <https://www2.census.gov/geo/> [↑](#footnote-ref-24)
25. <https://www2.census.gov/geo/pdfs/reference/guidestloc/> [↑](#footnote-ref-25)
26. <https://www2.census.gov/geo/pdfs/reference/guidestloc/17_Illinois.pdf> [↑](#footnote-ref-26)
27. <http://data5.ctpp.transportation.org/ctpp/Browse/browsetables.aspx> [↑](#footnote-ref-27)
28. <https://github.com/gregmacfarlane/ctpp_flows> [↑](#footnote-ref-28)
29. <https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-data.html> [↑](#footnote-ref-29)
30. <https://www.census.gov/geographies/mapping-files/time-series/geo/carto-boundary-file.html> [↑](#footnote-ref-30)
31. <https://www.census.gov/programs-surveys/geography/guidance/geo-identifiers.html> [↑](#footnote-ref-31)
32. <http://www.gis2gps.com/GIS/illcounties/illcounties.html> [↑](#footnote-ref-32)
33. <https://www.uis.edu/gis/projects/data/> [↑](#footnote-ref-33)
34. <https://data.cityofchicago.org/Facilities-Geographic-Boundaries/Boundaries-ZIP-Codes/gdcf-axmw> [↑](#footnote-ref-34)
35. <https://gis-idot.opendata.arcgis.com/> [↑](#footnote-ref-35)
36. <https://data.cityofchicago.org/Transportation/Bike-Racks-Map/4ywc-hr3a> [↑](#footnote-ref-36)
37. <https://data.cityofchicago.org/browse?q=transit&sortBy=relevance&tags=shapefiles> [↑](#footnote-ref-37)
38. <https://www.fhwa.dot.gov/planning/census_issues/ctpp/> [↑](#footnote-ref-38)
39. <https://github.com/gregmacfarlane/ctpp_flows> [↑](#footnote-ref-39)
40. Nelson, G. D. and A. Rae (2016). "An economic geography of the United States: From commutes to megaregions." PloS one 11(11): e0166083. [↑](#footnote-ref-40)
41. <https://lehd.ces.census.gov/data/#top> [↑](#footnote-ref-41)
42. <https://lehd.ces.census.gov/data/lodes/LODES7/LODESTechDoc7.5.pdf> [↑](#footnote-ref-42)
43. Multiscale dynamic human mobility flow dataset in the U.S. during the COVID-19 epidemic. [↑](#footnote-ref-43)
44. <https://data.cityofchicago.org/Transportation/Transportation-Network-Providers-Trips/m6dm-c72p> [↑](#footnote-ref-44)
45. <https://data.cityofchicago.org/Transportation/E-Scooter-Trips-2019-Pilot/2kfw-zvte> [↑](#footnote-ref-45)
46. <https://data.cityofchicago.org/browse?q=scooter&sortBy=relevance> [↑](#footnote-ref-46)
47. <http://dev.cityofchicago.org/open%20data/2020/09/14/scooter-gbfs-public-feeds-2020.html> [↑](#footnote-ref-47)
48. <https://www.chicago.gov/city/en/depts/cdot/supp_info/escooter-share-pilot-project.html> [↑](#footnote-ref-48)
49. <https://data.cityofchicago.org/Transportation/Divvy-Trips/fg6s-gzvg> [↑](#footnote-ref-49)
50. <https://data.cityofchicago.org/browse?q=taxi&sortBy=relevance> [↑](#footnote-ref-50)
51. <https://transitfeeds.com/p/chicago-transit-authority/165?p=1> [↑](#footnote-ref-51)
52. <https://transitfeeds.com/p/metra/169> [↑](#footnote-ref-52)
53. <https://developers.google.com/transit/gtfs> [↑](#footnote-ref-53)
54. <https://github.com/remix/partridge>; <https://github.com/kuanb/peartree> [↑](#footnote-ref-54)
55. <https://github.com/mrcagney/gtfstk>; <https://github.com/mrcagney/gtfs_kit> [↑](#footnote-ref-55)
56. <https://data.cityofchicago.org/Transportation/Chicago-Traffic-Tracker-Congestion-Estimates-by-Se/n4j6-wkkf> [↑](#footnote-ref-56)
57. <https://data.cityofchicago.org/Transportation/Chicago-Traffic-Tracker-Historical-Congestion-Esti/ef4k-dci7> [↑](#footnote-ref-57)
58. <https://api.arrayofthings.org/> [↑](#footnote-ref-58)
59. <https://movement.uber.com/cities?lang=en-US> [↑](#footnote-ref-59)
60. <https://gis-idot.opendata.arcgis.com/search?groupIds=b5143551ce2d41c7a5f89a2400c6af06> [↑](#footnote-ref-60)
61. <https://datahub.cmap.illinois.gov/dataset/mydailytravel-2018-2019-public> [↑](#footnote-ref-61)
62. <https://www.cmap.illinois.gov/data/transportation/travel-survey> [↑](#footnote-ref-62)
63. <https://developers.google.com/maps/documentation/maps-static/overview> [↑](#footnote-ref-63)
64. <https://developers.google.com/maps/documentation/streetview/overview> [↑](#footnote-ref-64)
65. <https://github.com/googlemaps/google-maps-services-python/blob/master/googlemaps/maps.py> [↑](#footnote-ref-65)
66. <https://github.com/robolyst/streetview/tree/master/streetview> [↑](#footnote-ref-66)
67. <https://earthexplorer.usgs.gov/> [↑](#footnote-ref-67)
68. <https://www.google.com/earth/> [↑](#footnote-ref-68)
69. <https://eos.com/landsat-8/> [↑](#footnote-ref-69)
70. <https://neo.sci.gsfc.nasa.gov/view.php?datasetId=VIIRS_543D> [↑](#footnote-ref-70)
71. <https://ngdc.noaa.gov/eog/download.html> [↑](#footnote-ref-71)
72. <https://researchguides.uoregon.edu/c.php?g=777133&p=5574231> [↑](#footnote-ref-72)
73. <https://github.com/sunnyqywang/plosm> [↑](#footnote-ref-73)
74. <https://wiki.openstreetmap.org/wiki/Downloading_data> [↑](#footnote-ref-74)
75. <https://wiki.openstreetmap.org/w/images/a/a5/OSMToolsDecisionTree.png> [↑](#footnote-ref-75)
76. <https://download.bbbike.org/> [↑](#footnote-ref-76)
77. <http://download.geofabrik.de/north-america/us.html> [↑](#footnote-ref-77)
78. <https://overpass-turbo.eu/> [↑](#footnote-ref-78)
79. <https://github.com/chicagocityscape/api> [↑](#footnote-ref-79)
80. <https://www.chicagocityscape.com/> [↑](#footnote-ref-80)
81. <https://data.cityofchicago.org/Health-Human-Services/COVID-19-Cases-Tests-and-Deaths-by-ZIP-Code/yhhz-zm2v> [↑](#footnote-ref-81)
82. Multiscale dynamic human mobility flow dataset in the U.S. during the COVID-19 epidemic [↑](#footnote-ref-82)
83. <https://www.chicago.gov/city/en/depts/mayor/supp_info/chicago-energy-benchmarking/Chicago_Energy_Benchmarking_Reports_Data.html> [↑](#footnote-ref-83)
84. <https://www.nature.com/articles/s41560-020-00752-y#Sec15> [↑](#footnote-ref-84)
85. <https://www.eia.gov/electricity/data/eia861m/> [↑](#footnote-ref-85)
86. <https://arrayofthings.github.io/> [↑](#footnote-ref-86)
87. <https://aot-file-browser.plenar.io/data-sets/chicago-complete> [↑](#footnote-ref-87)
88. <https://www.cdc.gov/places/about/500-cities-2016-2019/index.html> [↑](#footnote-ref-88)
89. <https://www.cdc.gov/places/index.html> [↑](#footnote-ref-89)
90. <https://data.cityofchicago.org/Environment-Sustainable-Development/Smart-Green-Infrastructure-Monitoring-Sensors-Hist/ggws-77ih> [↑](#footnote-ref-90)
91. <https://www.zillow.com/howto/api/GetSearchResults.htm> [↑](#footnote-ref-91)
92. <https://developer.twitter.com/en/docs/twitter-api> [↑](#footnote-ref-92)
93. <https://www.yelp.com/dataset> [↑](#footnote-ref-93)
94. <https://gist.github.com/scrapehero/5f51f344d68cf2c022eb2d23a2f1cf95> [↑](#footnote-ref-94)
95. <https://medium.com/dev-genius/scraping-zillow-with-python-and-beautifulsoup-bbc7e581c218> [↑](#footnote-ref-95)
96. <https://github.com/twintproject/twint> [↑](#footnote-ref-96)
97. The geography of taste: Using Yelp to study urban culture. (Page 5) [↑](#footnote-ref-97)
98. <https://gist.github.com/scrapehero> [↑](#footnote-ref-98)
99. <https://www.webharvy.com/articles/zillow-extraction.html> [↑](#footnote-ref-99)
100. <https://www.scrapehero.com/how-to-build-and-run-scrapers-on-a-large-scale/> [↑](#footnote-ref-100)
101. <https://www.scrapehero.com/how-to-scrape-yelp-com-for-business-listings/> [↑](#footnote-ref-101)
102. <https://www.ncdc.noaa.gov/cdo-web/> [↑](#footnote-ref-102)
103. <https://www.ncdc.noaa.gov/cdo-web/orders?email=qingyiw@mit.edu&id=2406408> [↑](#footnote-ref-103)
104. <https://www.itu.int/en/ITU-D/Statistics/Pages/facts/default.aspx> [↑](#footnote-ref-104)
105. <https://www.solarwinds.com/free-tools/call-detail-record-tracker> [↑](#footnote-ref-105)
106. Check “A multi-source dataset of urban life in the city of Milan” [↑](#footnote-ref-106)