

# Data Curation Guidebook - The Social Life of Neighborhoods

Author: Francine Stephens



# Contents

<b>Preface</b>	<b>5</b>
<b>1 Neighborhood-Census Geographies Crosswalk</b>	<b>7</b>
<b>2 Neighborhood Boundaries</b>	<b>13</b>
<b>3 Crime Data</b>	<b>15</b>
<b>4 Segregation Data</b>	<b>17</b>
4.1 Example one . . . . .	17
4.2 Example two . . . . .	17
<b>5 Gentrification Data</b>	<b>19</b>
<b>6 Immigration Data</b>	<b>21</b>



# Preface

Welcome to The Social Life of Neighborhoods! As the TA, you will be guiding students through the process of using and synthesizing spatial and quantitative data to create maps for their neighborhood assignments. The TA of this course also plays the important role of the *data curator*. When the TA prepares and provides the necessary data for the neighborhood assignments, the students are able to focus more on analyzing the data and drawing connections to the course readings to make sense of the patterns emerging in and around their neighborhoods. This makes for a smoother class experience for the students, TA, and professor!

The responsibility of curating data for the class over multiple weeks of the quarter may seem a little daunting, but the good news is that you do not have to come up with a workflow from scratch. This guidebook along with the R scripts stored in the linked GitHub folder will walk you through the steps to automate a substantial amount of the work.<sup>1</sup> The guidebook is organized in a series of chapters based on the course assignment and the data the students need from you to complete their assignment in ArcGIS Online. A brief description of each chapter is provided in the list below:

1. **Chapter 1** covers Neighborhood Assignment #1 and the crucial steps to complete after students have submitted this assignment and before they begin their first mapping assignment.
2. **Chapter 2** covers how to create the ArcGIS Online class page and students' neighborhood boundaries from the census tracts they submitted in Assignment #1 for the Neighborhood Organizations assignment.
3. **Chapter 3** covers the Theories of Crime/Social Disorder Assignment, which involves collecting and preparing the crime data.
4. **Chapter 4** covers the Segregation assignment and how to prepare data for the cities and metro areas/regions that contain the neighborhoods chosen by students.

---

<sup>1</sup>Some weeks, e.g., Crime Data, will require a little more manual data work on your part because the data are not coming from a single provider like the U.S. Census, and have different formats and attributes that you will have to standardize.

5. **Chapter 5** covers the Gentrification assignment and how to prepare data for the city that contains the neighborhoods chosen by students.
6. **Chapter 6** covers the Globalization assignment and how to prepare the immigration data for the students' neighborhoods and the city that contains their chosen neighborhood.<sup>2</sup>

Each chapter will explain the key materials and steps that you need to curate the data, including the **data sources** needed for the assignment; a description of the **key components of the R script**; the specific **parameters that need to be updated in the script** to run for your class; and the **data/shapefiles that the script will output** for you to upload to the ArcGIS Online class page for the students to use.

### Before You Begin

Before diving into Chapter 1, download the code and boilerplate files from Francine's public GitHub repository to your computer. The data and scripts within the GitHub repository folder will be referenced in the following chapters. *You do not need to set up a GitHub account/page* to run any of the scripts or save the exported data.

---

<sup>2</sup>This assignment was scratched for the 2021 Spring Quarter, so the data collection and preparation steps did not get a "real" test run, but the script and steps are still included in this guidebook to serve as boilerplate code and documentation for future TAs to build upon as needed.

# Chapter 1

## Neighborhood-Census Geographies Crosswalk

This chapter is fundamental to all of the data that you will curate during the quarter. It explains how to create the neighborhood-census geographies crosswalk, which is a master spreadsheet that links each student's neighborhood to various census geographies - census tracts, cities, and metropolitan statistical areas - that will be used throughout the course. This crosswalk file is for you and your future data curation work, not the students. You do not need to share the crosswalk with students in the class.

### 1.0.1 When should you curate this data?

Start creating the crosswalk *after students have submitted Assignment #1 and before the first mapping assignment* in ArcGIS online (this was Assignment #3 in Spring 2021). It would be the most efficient use of your time to curate this data *while* you are grading or reviewing Assignment #1, because some of the information that students include in this assignment is necessary for building the crosswalk file.

### 1.0.2 What do you need to get started?

The folder `nhood_shapes` contains the file and script that are relevant for creating the crosswalk spreadsheet. Specifically, you should open and view:

- the crosswalk file: `student_neighborhoods_list.csv`
- the R script: `create_neighborhoods.rmd`
- The documents students submitted for Assignment #1 on Canvas.

## 8CHAPTER 1. NEIGHBORHOOD-CENSUS GEOGRAPHIES CROSSWALK

### 1.0.3 Prepare the Crosswalk

#### Step 1. Copy the `student_neighborhoods_list.csv`.

This is the crosswalk file for the Spring 2021 quarter.

	A	B	C	D	E	F	G	H
1	student	nhood_name	tract_id	city	state_ab	place_id	metro_name	metro_id
2	Jessica	Downtown	5.31E+10	Yakima	WA	5380010	Yakima, WA	5349420
3	Pamela	Corktown	2.62E+10	Detroit	MI	2622000	Detroit-Warren-Livonia, MI	2619820
4	Nashira	Montbello	8.03E+09	Denver	CO	820000	Denver-Aurora-Broomfield, CO	819740
5	Nashira	Montbello	8.03E+09	Denver	CO	820000	Denver-Aurora-Broomfield, CO	819740
6	Jordan	Mission District	6.08E+09	San Francis	CA	667000	Bay Area Region	
7	Jordan	Mission District	6.08E+09	San Francis	CA	667000	Bay Area Region	
8	Jordan	Mission District	6.08E+09	San Francis	CA	667000	Bay Area Region	
9	Anthony	College Park	6.07E+09	Chino	CA	613210	Riverside-San Bernardino-Ontario, CA	640140
10	Nicole	Duveneck	6.09E+09	Palo Alto	CA	655282	Bay Area Region	
11	Olivia	City Center	6.08E+09	East Palo A	CA	620956	Bay Area Region	
12	Grace	Haight-Ashbury	6.08E+09	San Francis	CA	667000	Bay Area Region	
13	Grace	Haight-Ashbury	6.08E+09	San Francis	CA	667000	Bay Area Region	
14	Becky	Downtown	6.03E+09	Bakersfield	CA	603526	Bakersfield-Delano, CA	612540
15	Niki	Little Sai Gon	6.06E+09	Westminst	CA	684550	Los Angeles-Long Beach-Santa Ana, CA	631100
16	Claudia	Blue Hills	2.91E+10	Kansas City	MO	2938000	Kansas City, MO-KS	2928140
17	Claudia	Blue Hills	2.91E+10	Kansas City	MO	2938000	Kansas City, MO-KS	2928140
18	Josh	Chinatown	6E+09	Oakland	CA	653000	Bay Area Region	
19	Josh	Chinatown	6E+09	Oakland	CA	653000	Bay Area Region	
20	Betsayada	Boyle Heights	6.04E+09	Los Angele	CA	644000	Los Angeles-Long Beach-Santa Ana, CA	631100
21	Betsayada	Boyle Heights	6.04E+09	Los Angele	CA	644000	Los Angeles-Long Beach-Santa Ana, CA	631100
22	Betsayada	Boyle Heights	6.04E+09	Los Angele	CA	644000	Los Angeles-Long Beach-Santa Ana, CA	631100
23	Sahir	Niles	6E+09	Fremont	CA	626000	Bay Area Region	
24	Ricky	Mission District	6.08E+09	San Francis	CA	667000	Bay Area Region	
25	Ricky	Mission District	6.08E+09	San Francis	CA	667000	Bay Area Region	
26	Ricky	Mission District	6.08E+09	San Francis	CA	667000	Bay Area Region	
27	Andrew	Downtown	6.08E+09	San Mateo	CA	668252	Bay Area Region	
28	Andrew	Downtown	6.08E+09	San Mateo	CA	668252	Bay Area Region	
29	Anushree	Overfelt	6.09E+09	San Jose	CA	668000	Bay Area Region	
30	Anushree	Overfelt	6.09E+09	San Jose	CA	668000	Bay Area Region	
31	Whitney	Oak Knoll	6E+09	Oakland	CA	653000	Bay Area Region	
32	Kyle	Hazelwood	4.2E+10	Pittsburgh	PA	4261000	Pittsburgh, PA	4238300
33	Kyle	Hazelwood	4.2E+10	Pittsburgh	PA	4261000	Pittsburgh, PA	4238300
34	Farah	Overfelt	6.09E+09	San Jose	CA	668000	Bay Area Region	

Figure 1.1: The neighborhood-census geography crosswalk

*Column Fields* - The first row is a header containing 8 fields that reflect the key pieces of information for the crosswalk. From left to right, the field names and what they represent are:

1. student - The student's name.
2. nhood\_name - The name of the selected neighborhood that the student provides in Assignment #1.
3. tract\_id - The full 11 digit census tract identifier (FIPS code - state + county + tract identifiers) that the student identifies as being part of their neighborhood in Assignment #1.
4. city - The name of the city where the student's selected neighborhood is located.




5. state\_ab - The abbreviation of the state where the neighborhood and its city are located.
6. place\_id - The 7 digit place identifier for the census place (i.e., city) of the student's chosen neighborhood.<sup>1</sup>
7. metro\_name - The name of the Census Metropolitan Statistical Area (MSA) where the neighborhood and its city are located.
8. metro\_id - The 7 digit identifier (FIPS code) for the Census Metropolitan Statistical Area (MSA).<sup>2</sup>

*Rows* - Each row in this spreadsheet represents student-neighborhood-census tract. In the example crosswalk file, some student and neighborhood names are on multiple rows. This is the case when the student selected more than one census tract to define their neighborhood boundaries in Assignment #1.

**Step 2. Clear the information in the rows below the header and replace with your students' information provided in Assignment #1**

In Assignment #1, the student will name the neighborhood, city, and state of the neighborhood site they choose to study for the quarter. They will also submit a document from the FFIEC Geocode Report that identifies the census tracts in the neighborhood that they have chosen as well as the metropolitan statistical area identifiers. An example image is shown below.


2020 FFIEC Geocode Census Report

Matched Address: 35 LILLIAN ST, SAN FRANCISCO, CA, 94124

MSA: 41884 - SAN FRANCISCO-SAN MATEO-REDWOOD CITY, CA

State: 06 - CALIFORNIA

County: 075 - SAN FRANCISCO COUNTY

Tract Code: 023102

### Summary Census Demographic Information

Tract Income Level	Low
Underserved or Distressed Tract	No
2020 FFIEC Estimated MSA/MD/non-MSA/MD Median Family Income	\$140,900
2020 Estimated Tract Median Family Income	\$37,536
2010 Tract Median Family Income	\$27,647
Tract Median Family Income %	26.64
Tract Population	4667
Tract Minority %	94.73
Tract Minority Population	4421
Owner-Occupied Units	520
1- to 4- Family Units	1081

### Census Income Information

Tract Income Level	Low
2010 MSA/MD/statewide non-MSA/MD Median Family Income	\$103,742
2020 FFIEC Estimated MSA/MD/non-MSA/MD Median Family Income	\$140,900
% below Poverty Line	43.15
Tract Median Family Income %	26.64
2010 Tract Median Family Income	\$27,647
2020 Estimated Tract Median Family Income	\$37,536
2010 Tract Median Household Income	\$31,434

### Census Population Information

Tract Population	4667
Tract Minority %	94.73
Number of Families	970
Number of Households	1547
Non-Hispanic White Population	246
Tract Minority Population	4421
American Indian Population	0
Asian/Hawaiian/Pacific Islander Population	975
Black Population	2559
Hispanic Population	698
Other/Two or More Races Population	189

### Census Housing Information

Total Housing Units	1583
1- to 4- Family Units	1081
Median House Age (Years)	39
Owner-Occupied Units	520
Renter Occupied Units	1027
Owner Occupied 1- to 4- Family Units	520
Inside Principal City?	YES
Vacant Units	36

The numbers highlighted in **green** are the state identifier, county identifier,

<sup>1</sup>The first two digits of the place id are the census state code, which has leading zeros for the first nine states. The script will add the leading zeros that Excel/Google Sheets does not show.

<sup>2</sup>The first two digits of the place id are the census state code, which has leading zeros for the first nine states. The script will add the leading zeros that Excel/Google Sheets does not show. For the segregation assignment, you will use a geographic delineation of the Bay Area presented in the course readings, so do not worry about recording the census name or identifier. The script creates the geography for you.

and tract identifier, respectively. Input them together in the third column. Do not include a decimal point in the tract code (the identifier only contains numbers).

The text highlighted in **yellow** gives the metropolitan statistical area identifier - a 5 digit number - and the name of the metropolitan area, which belong in columns 7 and 8, respectively.<sup>3</sup>

To obtain the place identifier, navigate to Social Explorer. Select the *Tables* option in the left-sidebar menu and then navigate to any survey program (Decennial Census is fine). Select the *Place Geography* option and begin adding the city names provided by students to the drop-down menu as shown in the image below.

Figure 1.2: Social Explorer Geography Selection Page

Once you have added all of the cities to the selection, Proceed to the Tables page and export your results, with all geographic identifiers to a CSV. The CSV will have the place identifier in the first column as is highlighted in **blue** in the image.

**Important Note:** Census geographic identifiers change over time. Therefore, it is important to make sure that the place and metropolitan statistical area geo-identifiers are for the same year as the census tracts. For instance, in the Spring 2021 quarter, the census tracts were still the 2010 boundaries/identifiers, so we used the 2010 boundaries and identifiers for the place and metropolitan statistical area geographies.

<sup>3</sup>Be sure to add the two digit state identifier before the 5 digit MSA identifier when you enter this information in column 7.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	FIPS	Name of Ar Qualifying Nation	State	County	County Sub Census Tract	Place	Combined '2020 Total	Total Popul	Total Populatic				
2	Geo_FIPS	Geo_NAME	Geo_QNAN	Geo_NATIC	Geo_STATE	Geo_COUN	Geo_CS	Geo_CT	Geo_PLACE	Geo_CSA	SE_T003_0	SE_T003_0	SE_T003_003
3	603526	Bakersfield	Bakersfield	0	6				3526		403455	54805.23	15.71928
4	613210	Chino city	Chino City,	0	6				13210		91403	13242.72	16.94304
5	620956	East Palo A	East Palo A	0	6				20956		30034	1878.861	6.673242
6	626000	Fremont cit	Fremont Ci	0	6				26000		230504	16407.55	7.663626
7	644000	Los Angeles	Los Angeles	0	6				44000		3898747	106152.6	2.798943
8	653000	Oakland cit	Oakland Cit	0	6				53000		440646	49843.93	12.75426
9	655282	Palo Alto ci	Palo Alto C	0	6				55282		68572	4142.648	6.429752
10	667000	San Francis	San Francis	0	6				67000		873965	68755.31	8.538809
11	668000	San Jose cit	San Jose Cit	0	6				68000		1013240	61624.78	6.475809
12	668252	San Mateo	San Mateo	0	6				68252		105661	8656.144	8.923413
13	820000	Denver city	Denver City	0	8				20000		715522	115667.4	19.28257
14	2622000	Detroit city	Detroit City	0	26				22000		639111	-74644.6	-10.458
15	2938000	Kansas City	Kansas City	0	29				38000		508090	48352.61	10.51744
16	4261000	Pittsburgh	Pittsburgh	0	42				61000		302971	-2769.17	-0.90573
17	5380010	Yakima city	Yakima City	0	53				80010		96968	5501.193	6.014414

Figure 1.3: Social Explorer Download with Place ID highlighted



## Chapter 2

# Neighborhood Boundaries

Assignment 2 or 3 (CHECK). Create neighborhood boundaries and export. Upload to ArcGIS Online.

Figures and tables with captions will be placed in **figure** and **table** environments, respectively.

```
par(mar = c(4, 4, .1, .1))  
plot(pressure, type = 'b', pch = 19)
```

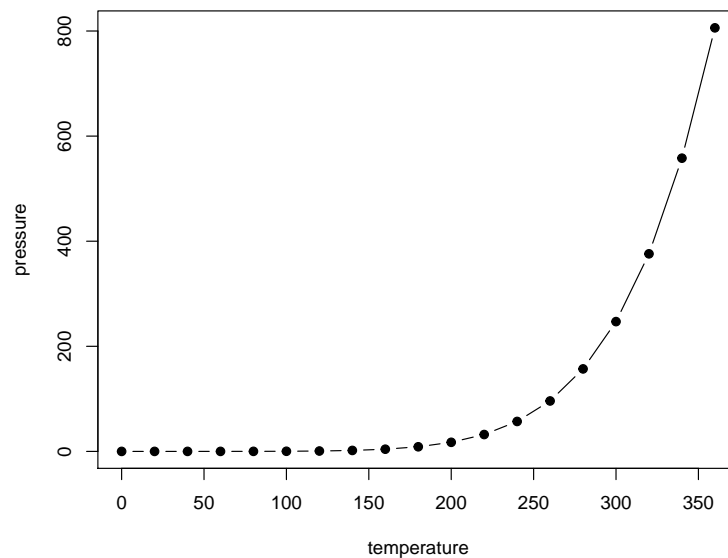


Figure 2.1: Here is a nice figure!

Table 2.1: Here is a nice table!

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
5.1	3.5	1.4	0.2	setosa
4.9	3.0	1.4	0.2	setosa
4.7	3.2	1.3	0.2	setosa
4.6	3.1	1.5	0.2	setosa
5.0	3.6	1.4	0.2	setosa
5.4	3.9	1.7	0.4	setosa
4.6	3.4	1.4	0.3	setosa
5.0	3.4	1.5	0.2	setosa
4.4	2.9	1.4	0.2	setosa
4.9	3.1	1.5	0.1	setosa
5.4	3.7	1.5	0.2	setosa
4.8	3.4	1.6	0.2	setosa
4.8	3.0	1.4	0.1	setosa
4.3	3.0	1.1	0.1	setosa
5.8	4.0	1.2	0.2	setosa
5.7	4.4	1.5	0.4	setosa
5.4	3.9	1.3	0.4	setosa
5.1	3.5	1.4	0.3	setosa
5.7	3.8	1.7	0.3	setosa
5.1	3.8	1.5	0.3	setosa

Reference a figure by its code chunk label with the `fig:` prefix, e.g., see Figure 2.1. Similarly, you can reference tables generated from `knitr::kable()`, e.g., see Table 2.1.

```
knitr::kable(
  head(iris, 20), caption = 'Here is a nice table!',
  booktabs = TRUE
)
```

## Chapter 3

# Crime Data

How to obtain and prepare crime data.





## Chapter 4

# Segregation Data

Some *significant* applications are demonstrated in this chapter.

### 4.1 Example one

### 4.2 Example two



## Chapter 5

# Gentrification Data

We have finished a nice book.



## Chapter 6

# Immigration Data