

CHRISTOPHER G. PRENER, PH.D.

FINAL PROJECT

VERSION: FEBRUARY 13, 2018

SOC 4650 & SOC 5650: INTRODUCTION TO GIS

SPRING, 2018
SAINT LOUIS UNIVERSITY

Contents

<i>Project Overview</i>	5
<i>How the Citizens' Service Bureau Works</i>	9
<i>The Citizens' Service Bureau Data</i>	17
<i>Project Memo</i>	19
<i>Analysis Development</i>	31
<i>1: Organizing Group Work</i>	33
<i>2: Cleaning the Citizen's Service Bureau Data</i>	37
<i>3: Cleaning the Demographic Data</i>	41
<i>4: Cleaning Spatial Data</i>	45
<i>5: Creating Geodatabases</i>	57
<i>6: Wire Framing the Poster</i>	59

7: *Creating Maps 2 to 4* 63

Project Overview

This guide contains instructions for completing the final project for SOC 4650/5650: INTRODUCTION TO GIS, which requires students to clean and map non-emergency call data from the City of St. Louis. These data describe the presence of various types of problems in City neighborhoods. The project brings together many of the scientific computing and GI science skills that are developed over the course of the semester. It also models the collaborative nature of GI science work by asking students to complete some aspects of the final project as a group using the collaborative development (GitHub) and communication (Slack) tools we have learned this semester. This process and project mirrors the steps taken to author a geospatial conference poster presentation (for all students) and a geospatial journal article (for students enrolled in SOC 5650).

Econometrics and Big Data

Among sociologists and urban researchers, there has been a recent desire to produce quantitative metrics that assist in assessing the social characteristics of neighborhoods. This effort, termed “econometrics”, initially relied on the development of systematic quantitative measures derived from neighborhood surveys and observation (Sampson and Raudenbush 1999). The goal of econometrics, like econometrics¹ and psychometrics², has been to produce generalizable techniques that are not specific to a single neighborhood or city, but rather can be used to measure and compare social processes in various cities.

The logic of these instruments has been extended, however, to also encompass novel administrative data sources produced by cities themselves. Unlike earlier econometric instruments, measures derived from administrative data can be obtained for little to no cost and carried out over long time periods (O’Brien, Sampson, and Winship 2015). While “big data” may be faddish and even subject to notable failures (Lazer et al. 2014), large datasets of data produced by

¹ The statistical field of economics that seeks to model economic theory.

² The statistical field of psychology that seeks to measure psychological health and mental illness.

cities may offer us the opportunity measure the health or disorder of neighborhoods in new ways.

The Citizens' Service Bureau

Following O'Brien and colleagues (2015), we will use non-emergency call data collected through a "3-1-1" system (the three digit phone number sometimes used for these services) for the final project. The *Citizens' Service Bureau* (CSB) is St. Louis's version of the 3-1-1 system. They describe their purpose in city government as being "to effectively and efficiently register and route city service requests, answer citizen requests for information, and provide City departments with statistics as needed." These types of requests cover a variety of non-emergent issues - everything from potholes to graffiti to traffic lights that need new bulbs. Our goal with these data is to map the extent of particular types of urban issues that may signal the "health" of neighborhoods within the City.

Final Project Progression

Producing the final project deliverable(s) requires a number of distinct steps that mirror the workflow we have introduced this semester for conducting spatial research. You will need to clean and project tabular data related to the CSB call category of your choice, clean and project demographic data about St. Louis, and produce a number of reference and thematic maps that describe your topic. To produce those maps, you will also need to obtain and clean spatial data that covers the St. Louis area. Instructions for each of these aspects of the project are included in subsequent sections of this document.

Work Groups

All students will be assigned a work group after the initial memos are submitted. We will try to place students in work groups who are working on similar topic areas. Each work group will be responsible for collectively creating the geodatabase and initial set of maps that group members will need for their final projects. The instructions in this document will detail which aspects of the project are to be completed by the group and which aspects each individual student is expected to complete.

Each work group will be assigned a GitHub repository and a Slack channel. Groups are expected to track their progress using the

GitHub Project associated with their work group repo. Work will be uploaded to individual branches within your group repository and then merged through pull requests, with group members providing “peer review” of each others’ work before accepting pull requests. They should also utilize Slack for communication, and can also reach out to both Chris and Kyle from their work group Slack channel.

Students will be asked to submit “peer reviews” of each others’ contributions to the group at the end of the semester. However, if there are issues with group members failing to complete work in a timely fashion, they should be brought to Chris immediately. Failure to participate fully in the work group will be reflected in students’ participation and final project grades.

Waypoints

The final project instructions include “waypoints” - designated check-in times for students to ensure that everyone continues to make consistent forward progress throughout the semester. Waypoints for the Spring, 2018 edition of the course are as follows:

1. **Lecture-03 (February 5th)** - Topic memo submitted as a GitHub issue in each student’s assignment repository (see page 19 for details).
2. **Lecture-05 (February 19th)** - Meeting report posted in each group’s Slack channel. A designated reported should summarize the meeting and include details on who has been assigned what tasks
3. **Lecture-07 (March 5th)** - Progress report from each student due in group’s Slack channel. Each student should provide details of what work they have accomplished so far and what tasks still remain.
 - (a) **SOC 5650 Only** - An annotated bibliography with twenty to twenty-five scholarly, peer-reviewed sources is due at this waypoint.
4. **Lecture-11 (April 2th)** - Draft materials due in each student’s assignment repository. You are expected to have made *substantial* progress at this point - the overall layout and design of the poster should be established and a portion of the maps should be made and included on the poster. This will only be graded for completion - students who do not submit poster drafts or whose poster drafts show minimal progress will receive a deduction applied to their final project grade.

- (a) **SOC 5650 Only** - An draft paper showing *substantial* progress on the Introduction, Literature Review, and Data and Methods sections is due at this waypoint.
- 5. **Lecture-13 (April 16th)** - Response to reviewer due as a GitHub issue in each student's assignment repository. Each student should open a new issue and respond to Chris's feedback with a plan for how to address each substantive point.
- 6. **Lecture-15 (April 30th)** - Progress report from each student due as a GitHub issue in each student's assignment repository. Each student should provide a progress report on each of the substantive points detailed in the response to reviewer.

Updates

This guide will be updated after memos are submitted with additional details about how to complete the final project. Please check the [Final Project](#) page on the course website for updates and additional information - it will continue to be updated with information, tips, and guidance as the project progresses.

How the Citizens' Service Bureau Works

The CSB is a clearing house for the City. Requests for service come primarily from residents or business owners in the City through the CSB's website or by phone. Other City agencies also put requests for service into the CSB, and they take requests via fax, email, and Twitter.

The following pages walk through what a CSB submission via the City [CSB website](#) entails. It is an example of semi-structured data collection. Where possible, fields with a limited number of options are used to limit the types of data or information can be entered. However, there are a number of places where unstructured information can be entered as well. The form fields that appear are flexible. For different problem types, different questions are asked.

The CSB also uses this form to help users identify whether the CSB will actually be able to assist them. For example, they warn submitting residents that graffiti above the second story or on particular types of construction materials cannot be cleaned by the City.

Note how the CSB makes use of spatial data as part of the data collection process. When an address is entered and successfully located, the next page returns parcel and address level information from City datasets overlaid on the Google Maps basemap. This allows submitters to verify that they are correctly locating the problem that they wish to report.

Once the report is addressed, it is marked as "Closed" in the CSB database and its resolution is updated on the CSB's website. Residents who provided their contact information

When to call Citizens' Service Bureau (622-4800)

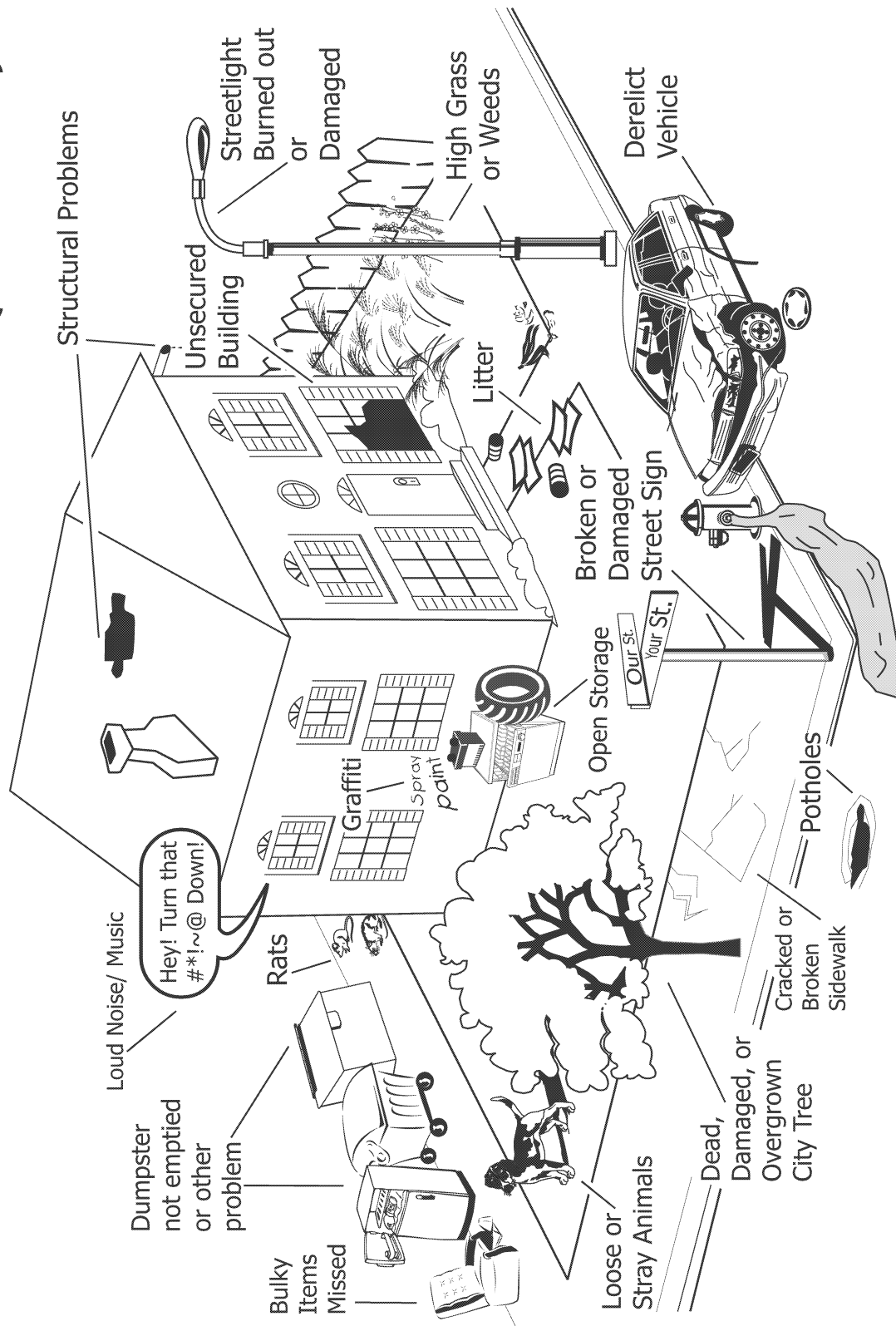
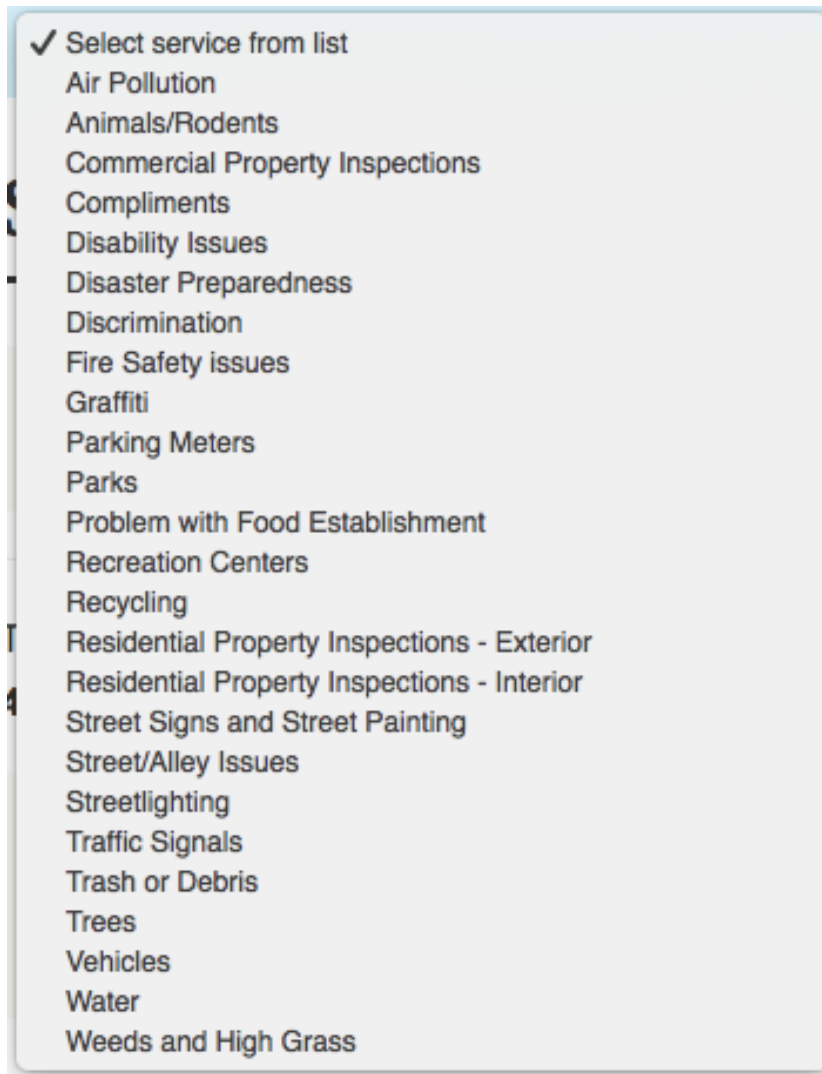


Figure 1: This is the front page of a flier produced by the Citizens' Service Bureau that describes situations the CSB can help residents handle.



✓ Select service from list

- Air Pollution
- Animals/Rodents
- Commercial Property Inspections
- Compliments
- Disability Issues
- Disaster Preparedness
- Discrimination
- Fire Safety issues
- Graffiti
- Parking Meters
- Parks
- Problem with Food Establishment
- Recreation Centers
- Recycling
- Residential Property Inspections - Exterior
- Residential Property Inspections - Interior
- Street Signs and Street Painting
- Street/Alley Issues
- Streetlighting
- Traffic Signals
- Trash or Debris
- Trees
- Vehicles
- Water
- Weeds and High Grass

Figure 2: This is the main list of problems that someone submitting a request to the CSB via their website can select from.

Step 1 of 6: Select The Problem

- Graffiti
 - [Trash container is damaged/has graffiti](#)
 - [Graffiti on park equipment](#)
 - [Graffiti on private property](#)
 - [Graffiti on private/commercial trash dumpster](#)
 - [Graffiti on public property](#)
 - [Inspect city tree \(tree between street and sidewalk\)](#)
 - [Graffiti on a traffic sign/device](#)

Figure 3: For many problem types, users are offered a list of possible subtypes that help the CSB further narrow which City agency will ultimately be responsible for addressing the concern.

Step 2 of 6: Where is the request needed?

By Address

Street Address:

[Search](#)

E.g. 1520 Market

By Intersection

Street 1:

- Select street from list -

Street 2:

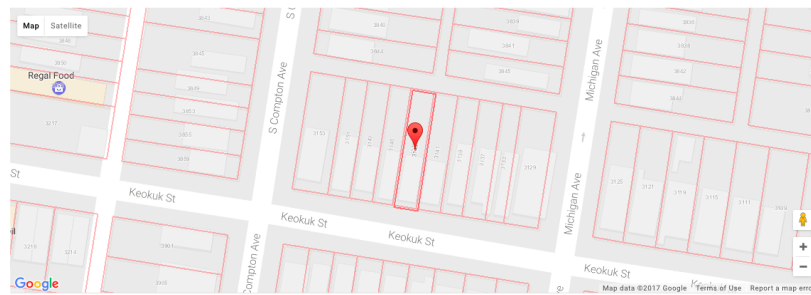
- Select street from list -

[Search](#)

Figure 4: Users can submit requests either by street address or by an intersection.

Step 2 of 6: Confirm the location

You have selected the following location: **3143 KEOKUK ST**



Is This Location Correct?

☒ Yes, continue to step 3

[+ No, try a new address](#)

Figure 5: Note how spatial data are overlaid on the Google Maps basemap. These data include parcel boundaries and street address numbers.

Step 3 of 6: Answer detailed questions about your request

Question 1: How high up is the graffiti?

Note: City cannot remove graffiti higher than the 2nd story. For these cases, the owner is responsible for removing the graffiti. You may instead submit a request for a building inspection

Change to: [inspect for exterior building code violations](#)

☒ Ground level

☐ 2nd story

[Continue](#)

Figure 6: The CSB uses a number of screens to obtain more detailed information about the request and to give users information that may impact how their request is processed. This screen captures the height of the graffiti off the ground.

Step 3 of 6: Answer detailed questions about your request

Question 2: Location of graffiti

☐ On building-rear

☐ On building-front

☒ On building-side

☐ Garage

☐ Other - see details

[Continue](#)

Figure 7: The second screen for the graffiti request. This screen captures the location of the graffiti on the building.

Step 3 of 6: Answer detailed questions about your request

Note: City does not remove graffiti from fiberglass doors or wooden fences. For these cases, the owner is responsible for removing the graffiti. You may instead submit a request for a building inspection

Question 3: Type of surface

- ☒ Unpainted Brick
- ☐ Painted Brick
- ☐ Unpainted Concrete
- ☐ Painted Concrete
- ☐ Unpainted Metal
- ☐ Painted Metal
- ☐ Other - see details

Continue

Figure 8: The third screen for the graffiti request. This screen captures the construction material that the graffiti has been painted on, which can impact how the graffiti is removed.

Step 3 of 6: Answer detailed questions about your request

Question 4: What color is the graffiti?

not to exceed 450 characters

Blue and white

Continue

Figure 9: The fourth screen for the graffiti request. This screen captures details on what the color of the graffiti is.

Step 3 of 6: Answer detailed questions about your request

Question 5: Is the property vacant or occupied?

- ☐ Vacant
- ☒ Occupied
- ☐ Unknown

Continue

Figure 10: The fifth screen for the graffiti request. This screen captures details about the building that has been affected.

Step 3 of 6: Answer detailed questions about your request

Please describe the details regarding your request, such as exact location (front, rear, etc.), description of problem, and any other relevant facts that may help the inspector resolve your issue.

You will have the option to attach a photo to this request in step 5. Even if you plan on providing a photo, it is very important that you provide details below.

Provide Details: (not to exceed 450 characters)

Graffiti is located on the east side of the building in the gangway between this building and 3141 Keokuk Street.

Continue

Figure 11: The sixth screen for the graffiti request. This screen captures details about where the graffiti is on the affected. Note how this is different than FIGURE 6 in terms of the way the data are collected.

Step 4 of 6: Enter your contact information

Please let us know how to contact you in case we need more information to handle your request. (optional)

For Citizens

Last Name

Doe

First Name

Jane

Email Address

jane_doe@gmail.com

For Official Use Only

Account

Please use upper case

Continue

Figure 12: This screen captures details about who is submitting the request. Note that it is *optional*.

Step 5 of 6: Upload Supporting Files (Optional)

Please upload any supporting documents that will clarify your request (optional).

Supported document types: PDF, JPG, JPEG, GIF, PNG

Upload 1

Choose File

graffiti.png

Clear this file

+Add a related document

Submit Request

Figure 13: This screen allows users to upload a photo or document that pertains to the request. This is the last step before submitting the request to the CSB and verifying its accuracy.

Check Status of a Service Request

Find out how your service request is progressing

Your search for service request ID **904156**

[New Search](#)

Service Request (SR)ID:	904156
Problem Location:	3143 KEOKUK ST
Request Initiated:	10/14/2016
Dept To Receive SR:	Operation Brightside
Problem Type:	Graffiti on private property
Dept Response Due By:	11/03/2016
Status of Request:	Closed
Dept Investigation Completed On:	10/25/2016
Dept Resolution:	graffiti removed

Figure 14: This screen shows the resolution of the request.

The Citizens' Service Bureau Data

The core data for this project cover seven years of Citizens Service Bureau (CSB), from 2009 through 2015. The data were originally obtained in February of 2017 from the City of St. Louis's [Public Datasets portal](#).

St. Louis City's Public Datasets - Terms of Use

All of the data that the City of St. Louis makes public, including the CSB data, are available under the following [license](#):

Description *The City of St. Louis strives to enhance public access to and use of data that it collects and publishes. As part of an initiative to improve the accessibility, transparency, and accountability of City government, this catalog offers access to a repository of government-produced, machine-readable data sets. The datasets are organized by originating department. The City provides access to the information free of charge subject to the terms of this agreement. Use of data derived from the datasets, which may appear in formats such as tables and charts, is also subject to these terms.*

Terms *The City of St. Louis reserves the right at any time to update, modify, or discontinue the release of the datasets. The City does not warranty the completeness, accuracy, content, or fitness for any particular purpose or use of any public dataset made available, nor are any such warranties to be implied or inferred with respect to the public datasets furnished therein. The City shall not be responsible or liable for the accuracy, usefulness or availability of any data in the datasets.*

Accountability *When working with the datasets, be aware that these files are raw extracts derived from various data sources. The City of St. Louis is aware that errors exist in these datasets. Contact the originating department if questions/issues arise.*

Data Access

All data can be found in the file `csbCreate.csv`, which is located in the [finalProject](#) repository on GitHub. This repo also contains the code used to create `csbCreate.csv`. **You do not need to execute this script. It is provided for reference only. It is legacy Stata code that**

we will not be using this semester. The output the script generated is included. The total size of the initial dataset is quite large - 87.9MB. It contains $n = 786,355$ observations.

Data Definitions

The following are some basic descriptions of what each variable in 'csbCreate.csv' are supposed to measure:³

- requestid - CSB assigned identification number for each request
- problemcode - general description of the problem associated with the request
- description - (sometimes) more detailed description of the problem associated with the request
- probaddress - address where problem is located
- probcity - city where address is located
- probzip - zipcode where address is located
- datetimeinit - date request made
- srx - UTM x coordinate for address
- sry - UTM y coordinate for address
- status - status of request
- prjcompletedate - date request completed
- reqyear - year request made⁴

³ It is worth noting that a number of these variables, including probcity and probzip, do not actually measure what they are supposed to.

⁴ This is the only variable that was not included in the original data obtained from the City of St. Louis. Chris created this during the initial dataset construction process to facilitate the removal of cases before January 1, 2009 and after December 31, 2016.

Data Quality

You should be aware that these data contain a number of issues. There are numerous duplicate records, the identification variable does not uniquely identify observations, and the variable names are unnecessarily long. The spatial data are also incomplete - some addresses do not have corresponding x,y coordinates, and some x,y coordinates do not have corresponding addresses. All of these issues will need to be resolved in the data cleaning aspect of the project.

Project Memo



Due by February 5th



Submit as a Issue in your GitHub Assignments repo

Memo Assignment

The first task for the final project is selecting a category or categories of interest. The list below reflects the contents of the `problemcode` variable after you have completed the data cleaning steps. This is provided as a preview of the state of the dataset after the data quality problems have been dealt with. Note that the sample size has dropped from $n = 786,355$ to $n = 721,448$. We want you to select a final project topic based not on the initial data release, but on the sample size of observations you will have after the initial data cleaning is complete.

To complete the final project memo, scan the list below and select a problem code or codes that interest you. Your topic should have a minimum of 500 observations and ideally should have several thousand or more. If you find a topic that interests you that has fewer than the ideal number of observations, try to find a number of related categories so that you can pool those observations.

Your memo should be a Markdown formatted GitHub Issue that provides a general statement about your topic and notes the specific problem code or codes you are interested in using along with the number of observations for each code. This should be submitted in your GitHub Assignments repository.

Problem Code Values for Cleaned Data

. tabulate code

code	Freq.	Percent	Cum.
-----+-----			
07-21-2010 Flood	16	0.00	0.00
12-28-2015 Flood	3	0.00	0.00
14th - Tucker Median	1	0.00	0.00
A/C REPAIRS	274	0.04	0.04
ALARM	17	0.00	0.04
APPEARANCE	35	0.00	0.05
ATHLETIC LIGHTS	1	0.00	0.05
Actvt Court Warrant	526	0.07	0.12
Add Block Contact	475	0.07	0.19
Affordable Care Act	3	0.00	0.19
Alter Hist No Permit	289	0.04	0.23
Animal Abuse	91	0.01	0.24
Animal Bite	1,156	0.16	0.40
Animal(s) Surrender	582	0.08	0.48
Asbestos	66	0.01	0.49
Athletic Fields	85	0.01	0.50
BALLAST	3	0.00	0.50
BOILER/FURNACE/RADIA	112	0.02	0.52
BREAKER	9	0.00	0.52
Bad Condition	4,890	0.68	1.20
Bad Drain/Sewer, Ext	222	0.03	1.23
Bad Drain/Sewer, Int	213	0.03	1.26
Bagged Leaf Pickup	474	0.07	1.32
Ballfield Benches	37	0.01	1.33
Bats	652	0.09	1.42
Bed Bugs	474	0.07	1.48
BicycleRackApp	21	0.00	1.49
Bike Lane Issues	4	0.00	1.49
Billboard Inspection	98	0.01	1.50
Birds-Food Cntrl	9	0.00	1.50
Blitz Items	142	0.02	1.52
Building Collapse	1,501	0.21	1.73
Bulky items missed	4,101	0.57	2.30
CEMA	757	0.10	2.40
CFPB Referral	2	0.00	2.40
CHK-JOBSITE	185	0.03	2.43
CIVIL UNREST	2	0.00	2.43
CLEAN AREA	34	0.00	2.43

CLOGGED DRAIN	101	0.01	2.45
CSB Record Request	423	0.06	2.51
Cart dumped incompl	752	0.10	2.61
Cave-Ins	5,183	0.72	3.33
Census Info Call	8	0.00	3.33
Check Welfare	148	0.02	3.35
Child Center	3	0.00	3.35
Child Center, Ext	16	0.00	3.35
Child Center, Int	80	0.01	3.36
City Tree Down	1,116	0.15	3.52
City Tree Limb Down	6,045	0.84	4.36
Clean Picnic Tables	4	0.00	4.36
Cntainr near window	34	0.00	4.36
Cntainr too heavy	834	0.12	4.48
Cntnr blocking	5,995	0.83	5.31
Cntnr knocking wires	1,206	0.17	5.48
Cntnr out of postion	3,140	0.44	5.91
Cntnr overflow/bulky	10,259	1.42	7.33
Cntnr turned	1,942	0.27	7.60
Comml Dumpster, Ext	970	0.13	7.74
Comml coll after hrs	293	0.04	7.78
Const Hist No Permit	169	0.02	7.80
Const No Permit-Ext	1,935	0.27	8.07
Const No Permit-Int	887	0.12	8.19
Construction	4	0.00	8.19
Containr damaged	18,170	2.52	10.71
Containr not level	688	0.10	10.81
Containr smells bad	560	0.08	10.88
Contract Bridge Work	1	0.00	10.88
Contract Park Work	3	0.00	10.89
Contract Pub Bldg Wk	3	0.00	10.89
Contract Street Work	9	0.00	10.89
Curb Repair	1,034	0.14	11.03
Cut FnceLine/TreeLwn	592	0.08	11.11
Cut Median-Pkwy	547	0.08	11.19
DOOR REPAIR	59	0.01	11.20
Dangerous Animal	716	0.10	11.30
Dangerous Dog	2,594	0.36	11.66
Day Care/Childcare	3	0.00	11.66
Dead Animal	5,338	0.74	12.40
Dead Dog-ACC	53	0.01	12.40
Dead Limbs in Park	88	0.01	12.41
Debris-Occupied Bldg	14,819	2.05	14.47
Debris-Vacant Bldg	19,733	2.74	17.20

Debris-Vacant Lot	12,113	1.68	18.88
Debris-alley/street	18,104	2.51	21.39
Defective Plumbing	4	0.00	21.39
Defective Toilet	103	0.01	21.41
Demo Blowing Dust	104	0.01	21.42
Demolition No Permit	115	0.02	21.44
Derelect-Privt Ppty	5,822	0.81	22.24
Derelect-Rt of Way	5,353	0.74	22.99
Did not get cart	2,314	0.32	23.31
Disabled Parking	233	0.03	23.34
Dust/Dirt Emission	8	0.00	23.34
ELECTRICAL INSTALL	52	0.01	23.35
ELECTRICAL TRACE	12	0.00	23.35
ELEVATOR REPAIRS	20	0.00	23.35
EVENT	16	0.00	23.35
EXHAUST	14	0.00	23.36
Electrical Insp	1,826	0.25	23.61
Employment Discrim	7	0.00	23.61
Enforce Ordinace	9,021	1.25	24.86
Evictions	445	0.06	24.92
FH-MAINTENANCE	646	0.09	25.01
FH-RUNNING	380	0.05	25.07
FLAGS	8	0.00	25.07
FLOOR/WALL/CEILING	78	0.01	25.08
FURNITURE REPAIR	44	0.01	25.08
Fill Dirt Request	197	0.03	25.11
Fire Escapes-Exits	63	0.01	25.12
Fire Exit Insp	164	0.02	25.14
Food Estab. Problem	992	0.14	25.28
Food worker ill	9	0.00	25.28
Foodbourne Illness	242	0.03	25.31
Foreman cntct citizn	5,358	0.74	26.06
Fountains	208	0.03	26.09
Fumes/Smells/Odors	72	0.01	26.10
GATE/FENCE	10	0.00	26.10
GENERATOR	2	0.00	26.10
GLASS/WINDOW	10	0.00	26.10
Graffiti-Park Equip	168	0.02	26.12
Graffiti-Private	9,103	1.26	27.38
Graffiti-Prvt Dmpstr	634	0.09	27.47
Graffiti-Public	2,582	0.36	27.83
HAZ MAT	1	0.00	27.83
HEAT Enrollment	12	0.00	27.83
HOLE-DEPRESSION	128	0.02	27.85

HYDRANT-HIT	665	0.09	27.94
Handicapped Parking	6	0.00	27.94
Hazrds Waste	2	0.00	27.94
Hazrds Waste, Ext	290	0.04	27.98
Hazrds Waste, Int	44	0.01	27.99
Health Center, Int	8	0.00	27.99
High Grass in Park	338	0.05	28.04
Hole-bottom of cont	5,435	0.75	28.79
Holiday Collct Schd	656	0.09	28.88
Homeless Call	13	0.00	28.88
Housing Discrim	15	0.00	28.88
Hsehld Wst Disposal	10	0.00	28.89
Idling	33	0.00	28.89
Illegal Food Estbt	179	0.02	28.92
Illegal Use	2,204	0.31	29.22
Illegal dumpg report	3,404	0.47	29.69
Illegl Use-Opn Storg	16,252	2.25	31.95
Illgal Estbt-Com San	134	0.02	31.96
Imprpr resdnt displ	3,297	0.46	32.42
Inaccessible Facilty	68	0.01	32.43
Inaccessible Program	3	0.00	32.43
Incorrect Bill	3,694	0.51	32.94
Industrial Facility	4	0.00	32.94
Industrial Haz, Ext	35	0.00	32.95
Industrial Haz, Int	5	0.00	32.95
Injured Animal	264	0.04	32.99
Injured/Sick Animal	59	0.01	32.99
Inmate Complaint	26	0.00	33.00
Insects Food Ctrl	174	0.02	33.02
Insp Hotel, Int	49	0.01	33.03
Inspect City Tree	34,410	4.77	37.80
Inspect Private Tree	7,319	1.01	38.81
Instructn hshld/blky	3,553	0.49	39.30
Insufficient Heat	147	0.02	39.33
Intake Form	175	0.02	39.35
KINGSHIGHWAY CLOSURE	4	0.00	39.35
LAWN-CAVEIN	36	0.00	39.36
LEAK	89	0.01	39.37
LEAKING-METER	66	0.01	39.38
LOCK/KEYS	82	0.01	39.39
LOOSE-COVER	310	0.04	39.43
LRA Board up	4,079	0.57	40.00
LRA Tree	2,320	0.32	40.32
Lead Hotline Call	85	0.01	40.33

Lead Insp	3,367	0.47	40.80
Leaf pick up	338	0.05	40.84
Leash Law	114	0.02	40.86
Leaves in Park	34	0.00	40.86
Light Damaged	8,985	1.25	42.11
Light(s) Out	41,580	5.76	47.87
Lights ON Daytime	206	0.03	47.90
Litter Warning	2,040	0.28	48.18
Long Term Initiative	356	0.05	48.23
MISSING-COVER	657	0.09	48.32
MODOT Referral	259	0.04	48.36
Maintain Bike Trail	44	0.01	48.37
Mechanical Insp	583	0.08	48.45
Mice Infstation, Int	324	0.04	48.49
Misc-Air Pollution	30	0.00	48.50
Misc-Animal Care	224	0.03	48.53
Misc-Assessor	4	0.00	48.53
Misc-Bd of Aldermen	6	0.00	48.53
Misc-Bd of Pub Svc	288	0.04	48.57
Misc-Building Div	681	0.09	48.66
Misc-CSB	8	0.00	48.66
Misc-Circuit Clerk	2	0.00	48.66
Misc-City Courts	19	0.00	48.67
Misc-City Emrgcy Mng	22	0.00	48.67
Misc-City Jstce Cntr	1	0.00	48.67
Misc-Civil Rights En	7	0.00	48.67
Misc-Coll of Revenue	4	0.00	48.67
Misc-Comm Dev Admin	1	0.00	48.67
Misc-Comm Sanitation	403	0.06	48.73
Misc-Communications	175	0.02	48.75
Misc-Comptroller	5	0.00	48.75
Misc-Corrections	17	0.00	48.75
Misc-Dept Human Svc	16	0.00	48.76
Misc-Excise	288	0.04	48.80
Misc-Facilities Mgmt	86	0.01	48.81
Misc-Fire/EMS Dept	79	0.01	48.82
Misc-Food Control	103	0.01	48.83
Misc-Forestry Debris	12	0.00	48.84
Misc-Forestry Trees	4	0.00	48.84
Misc-Health Dir Ofc	43	0.01	48.84
Misc-LRA	1,504	0.21	49.05
Misc-License Coll	717	0.10	49.15
Misc-MYR LTR	95	0.01	49.16
Misc-Marshalls Ofc	5	0.00	49.16

Misc-Mayors Ofc	7	0.00	49.17
Misc-Med Security In	2	0.00	49.17
Misc-Nbhd Stblizn Tm	177	0.02	49.19
Misc-Ofc on Disabled	34	0.00	49.19
Misc-Op Brightside	123	0.02	49.21
Misc-PRF Dir Ofc	18	0.00	49.21
Misc-Parks Div	833	0.12	49.33
Misc-Public Safety	4	0.00	49.33
Misc-Recorder Deeds	1	0.00	49.33
Misc-Recreation	26	0.00	49.33
Misc-Recycle Sectn	55	0.01	49.34
Misc-Refuse Div	2,021	0.28	49.62
Misc-StL Dev Corp	2	0.00	49.62
Misc-Stl Agncy Aging	6	0.00	49.62
Misc-Street Dept	1,233	0.17	49.79
Misc-Street Div	2,297	0.32	50.11
Misc-Tow Lot	14	0.00	50.11
Misc-Traffic Div	3,903	0.54	50.65
Misc-Trash Task Frc	631	0.09	50.74
Misc-Treasurer	99	0.01	50.76
Misc-Water Div	503	0.07	50.83
Missed Cut - V Lot	2,270	0.31	51.14
Missed Cut - VBldg	5,402	0.75	51.89
Mold - Food Cntrl	13	0.00	51.89
Mold, Interior	1,975	0.27	52.16
Mosquito Breeding	1,120	0.16	52.32
Mosquito Spraying	3,379	0.47	52.79
Move cont/debrs dump	1,629	0.23	53.01
Move container	3,155	0.44	53.45
NEW CONSTRUCTION	3	0.00	53.45
NOWATER	82	0.01	53.46
NSO Block Request	1,793	0.25	53.71
New Account Insp Req	14	0.00	53.71
New Sign Requested	1,774	0.25	53.96
No Don't Plow Residn	104	0.01	53.97
No Heat, Int	515	0.07	54.05
No Hep A - Food Ctrl	16	0.00	54.05
No Hot Water, Int	313	0.04	54.09
No Interpreter	2	0.00	54.09
No Smoke Detectors	87	0.01	54.10
No Water - Food Ctrl	18	0.00	54.11
No Water Supply, Int	710	0.10	54.20
No comml dumpster	982	0.14	54.34
No response from DNR	4	0.00	54.34

No tub, shower, sink	5	0.00	54.34
Not enough cntnrs	5,521	0.77	55.11
OPERATING COSTS	1	0.00	55.11
OVERHEAD DOOR	59	0.01	55.12
Open Burning/Smoke	57	0.01	55.12
Opt Out Applctn Req	573	0.08	55.20
Other Alley Repair	1,920	0.27	55.47
Other Street Repair	3,722	0.52	55.98
Over 4 Pets	18	0.00	55.99
Over 4 Pets, Ext	853	0.12	56.11
Overcrowding	1,678	0.23	56.34
Ovrflwng littr barel	743	0.10	56.44
PAINT/PLASTER	38	0.01	56.45
PEST CONTROL	38	0.01	56.45
PLUMBING FIXTURE REP	109	0.02	56.47
POWER OUTAGE	24	0.00	56.47
PRESSURE	141	0.02	56.49
PREVENTATIVE MAINTEN	20	0.00	56.49
PROC FRAMING	25	0.00	56.50
Park Lake	49	0.01	56.50
Park Restrooms	91	0.01	56.52
Park Sewer Problem	17	0.00	56.52
Parking Lot Cndition	323	0.04	56.56
Parking Meter	88	0.01	56.57
Parking Ticket	33	0.00	56.58
Pets in restaurant	39	0.01	56.58
Playground Equipment	188	0.03	56.61
Plumbing Insp	1,390	0.19	56.80
Potholes - Alley	2,966	0.41	57.21
Potholes - Major St	4,133	0.57	57.79
Potholes - Residnl	11,517	1.60	59.38
Ppty Maint Code-Ext	23,972	3.32	62.71
Ppty Maint Code-Int	4,992	0.69	63.40
Privt Hauler Cnfrmd	426	0.06	63.46
Problem with food	403	0.06	63.51
Prohibited Animal	161	0.02	63.54
Property Damage-FIRE	8	0.00	63.54
Property Damage-FOR	1,016	0.14	63.68
Property Damage-LRA	278	0.04	63.72
Property Damage-PRK	34	0.00	63.72
Property Damage-REF	2,409	0.33	64.05
Property Damage-SDP	373	0.05	64.11
Property Damage-SDV	682	0.09	64.20
Property Damage-TLD	54	0.01	64.21

Property Damage-TOW	13	0.00	64.21
Pub Accmdtn Discrim	7	0.00	64.21
Public Nuisance Rpt	7,480	1.04	65.25
Public Swim Pool	3	0.00	65.25
Pull Illegal Sign	9,534	1.32	66.57
RECEPTACLE/SWITCH	22	0.00	66.57
RELAMP/REPAIR	113	0.02	66.59
ROOF/TUCKPOINTING	26	0.00	66.59
Rat Bite	5	0.00	66.59
Rat Infestation, Int	541	0.07	66.67
Rats (Exterior)	5,437	0.75	67.42
Rats/Mice Food Ctrl	118	0.02	67.44
Raw Garbage, Ext	4,344	0.60	68.04
Rcy Curbside Cancel	12	0.00	68.04
Rcy Delivery Missed	1,133	0.16	68.20
Rcy Info Youth Grps	3	0.00	68.20
Rcy Pilot 1 inquiry	23	0.00	68.20
Rcy Pilot 2	28	0.00	68.21
Rcy Presentation	30	0.00	68.21
Rcy curb bin damage	11	0.00	68.21
Rcy curb bin missed	104	0.01	68.23
Rcy curb bin missing	17	0.00	68.23
Rcy curb new address	26	0.00	68.23
Rcy curbbin no deliv	7	0.00	68.23
Rcy curbside new add	1	0.00	68.23
Rcy curbside new svc	616	0.09	68.32
Rcy dropoff dumping	5	0.00	68.32
Rcy dropoff litter	4	0.00	68.32
Rcy dropoff no sign	2	0.00	68.32
Rcy dropoff overflow	24	0.00	68.32
Rcy pick up svc info	273	0.04	68.36
Rec Bldg Maintenance	80	0.01	68.37
Rec Equipment	11	0.00	68.37
Rec Housekeeping	23	0.00	68.38
Rec Programming	3	0.00	68.38
Rec-Athletic Lights	48	0.01	68.38
Recycling not collec	2,356	0.33	68.71
Refuse Info Call	3,358	0.47	69.18
Refuse Svc Fee Call	458	0.06	69.24
Refuse not collected	11,955	1.66	70.90
Refuse referl to SDV	73	0.01	70.91
Rehab-roll off cntnr	35	0.00	70.91
Reminder to Trim	85	0.01	70.92
Remove cntnr	1,043	0.14	71.07

Renovation	1	0.00	71.07
Residnl Prvt Dmpstr	194	0.03	71.10
Roach Infestn, Int	859	0.12	71.21
Rollcart stolen/misg	1,722	0.24	71.45
Root Removal	44	0.01	71.46
SIGN INSTALLATION	6	0.00	71.46
SLMPD Referral	43	0.01	71.47
SPRINKLER REPAIR	3	0.00	71.47
STEAM REPAIRS	23	0.00	71.47
STREET-CAVEIN	45	0.01	71.48
SUMP PUMP	13	0.00	71.48
Sewer Lateral Defect	1,614	0.22	71.70
Shelter Comments	2	0.00	71.70
Short Term Initiativ	33,940	4.70	76.41
Sidewalk Ramp	260	0.04	76.44
Sidewalk Repair	5,558	0.77	77.21
Sign needs attention	16,937	2.35	79.56
Signal Bulb Out	1,369	0.19	79.75
Signal Damaged	3,785	0.52	80.27
Signal Dark	1,553	0.22	80.49
Signal Stuck	1,121	0.16	80.64
Signals Flashing	3,760	0.52	81.17
Signals out of sync	1,826	0.25	81.42
Single Light Out	4,933	0.68	82.10
Sirens Not Heard	83	0.01	82.11
Smkng Ban Issue	195	0.03	82.14
Smoke Detectors	38	0.01	82.15
Snow blkg dsabl acs	2	0.00	82.15
Snow/Ice - Major St	267	0.04	82.18
Snow/Ice - Residnl	2,927	0.41	82.59
Soulard Mkt Vendor	3	0.00	82.59
Special Event Recycl	4	0.00	82.59
SpecialEventApp	475	0.07	82.66
Speed Bump Project	3	0.00	82.66
Spray Painting	11	0.00	82.66
Sprinkler Systems	24	0.00	82.66
StRes referl to ACC	148	0.02	82.68
Stagnant Water	7	0.00	82.68
Stray Animal	4,102	0.57	83.25
Stray Animal Cntnd	118	0.02	83.27
Stray Cat	1,623	0.22	83.49
Stray Dog At Large	9,036	1.25	84.75
Stray Dog Cntnd-ACC	1,416	0.20	84.94
Stray Dog Contained	939	0.13	85.07

Street Cleaning	2,430	0.34	85.41
Street Painting Requ	758	0.11	85.51
Stump Removal	2,332	0.32	85.84
Surrender Cat	18	0.00	85.84
Surrender Dog	23	0.00	85.84
Surrender Pet	151	0.02	85.86
Swim Pool Insp, Ext	398	0.06	85.92
Swim Pool Insp, Int	21	0.00	85.92
Swine Flu Call	1	0.00	85.92
Swine Flu Psntn	20	0.00	85.92
TASTE/ODOR	85	0.01	85.94
Tent Encampment	2	0.00	85.94
Tire Storage, Ext	103	0.01	85.95
Too many cntnrs	304	0.04	85.99
Traffic Graffiti	1,817	0.25	86.25
Traffic Ticket	4	0.00	86.25
Trash in the Park	457	0.06	86.31
Trash/Grbage Fd Ctrl	81	0.01	86.32
Tree Inquiry	206	0.03	86.35
Tree Planting request	2,897	0.40	86.75
Truck dropped litter	1,432	0.20	86.95
Unauthrz use contain	2,086	0.29	87.24
Unsanitary Cond, Ext	1,939	0.27	87.51
Unsanitary Cond, Int	1,182	0.16	87.67
Unsanitary Pets, Ext	1,997	0.28	87.95
Unsatisfy Cut - VLot	1,345	0.19	88.13
Unsatisfy Cut -VBldg	1,538	0.21	88.35
VACANT BLDG INITIV	791	0.11	88.46
VBS Inquiry	5,065	0.70	89.16
Vacant Unit Appeal	2,059	0.29	89.44
Vacnt Bldg Unsecured	16,121	2.23	91.68
Vcnt weed check-web	1,902	0.26	91.94
Vehicle Towing	589	0.08	92.02
Vehicle-Rt of Way	4,295	0.60	92.62
WALK-CAVEIN	21	0.00	92.62
WATER COOLER	7	0.00	92.62
WATER-CITZN-PROP-DAM	57	0.01	92.63
WTR-COMING-UP	1,962	0.27	92.90
WTR-VACANT-BLDG	436	0.06	92.96
WTRNBLDG-PIPE-REM	31	0.00	92.97
Wastewater, Basement	549	0.08	93.04
Wastewater, Ext	293	0.04	93.08
Water Leak, Int	219	0.03	93.11
Weeds-Occupied Ppty	26,383	3.66	96.77

Weeds-Vacant Bldg	11,968	1.66	98.43
Weeds-Vacant Lot	3,475	0.48	98.91
Whole Blk Lights Out	3,895	0.54	99.45
Wildlife	7	0.00	99.45
Wildlife Issue	168	0.02	99.48
Yes Plow Residnl	1,057	0.15	99.62
Yrd Wste Not Collect	2,720	0.38	100.00
-----+-----			
Total	721,448	100.00	

Analysis Development

The work for the final project, beyond writing the topic memo, is organized into vignettes. Each vignette has a series of tasks related to one aspect of the final project. They all begin with an overview that includes whether they are group or individual elements of the project, what weeks on the syllabus the skills used correspond to, and end with a note about what data needs to be uploaded to GitHub. Each of the subsequent chapters forms a single vignette.

These vignettes are designed to break the final project down into “bite sized” chunks. They are also designed to be approached in a (generally) chronological manner (see Figure 15).⁵ For each vignette, you should follow the workflow we have discussed in class: plan, organize, document, execute.

⁵ That said, if you are waiting on team members to complete vignettes, you should skip ahead to tasks that do not require their data, such as wire framing your poster.

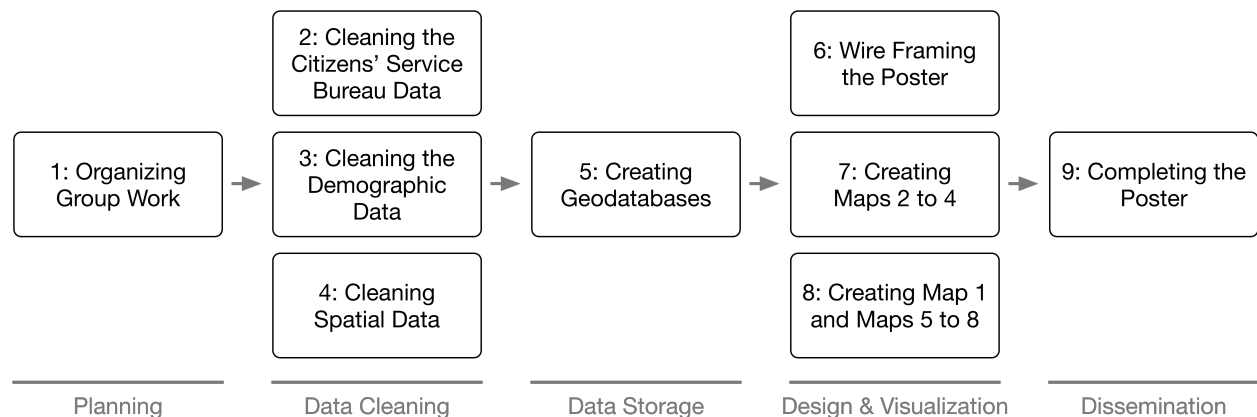


Figure 15: The chronological ordering of the vignettes.

In addition to these steps, which all students will need to complete, there is an additional vignette related to the final paper for students enrolled in soc 5650.

Data Storage

As you proceed through these steps, make sure to save **all** of the raw data, code, and documentation files on your course flash drive using the analysis development approach we have discussed this semester. There is a `FinalProject` directory in each of your repositories for this purpose. You are highly encouraged to use subfolders within `data/`, `docs/`, and `maps/`, and `results/`.

Project Documentation

Throughout the vignettes, there are specific pieces of documentation that we will ask that you produce.

README.md

You will need a read-me file in the top-level of your assignment repository. This file should describe the goal of the repository and, at a high level, how various elements of the project fit together.

RESEARCH_LOG.md

You will need to maintain a research log throughout the final project process. Before proceeding with the vignettes, create a well-formatted and organized research log in RStudio that is organized chronologically. The research log should use markdown syntax and be saved as a markdown file (`.md`). Keep this file updated and have it saved in the top-level of your assignment repository.

For each day you work on the final project (including both individual work and your share of the group work), there should be an entry that describes the work that you have done, what files you have created or modified, and how those modifications took place. Give as much detail as possible.

Think of the research log as step-by-step instructions to yourself that would allow you to recreate all of your work should every final project file you work on be accidentally deleted or lost. For code, you can provide higher level details since, ideally, your literate programming approach means that your code is well annotated and described. For maps, you want to provide as much fine grained detail as possible down to specific font, color, and size choices for data and layout elements.

1: Organizing Group Work

Some of the initial data access and cleaning work, as well as the production of three of maps needed for your posters, is work that we will distribute out across a number of group members. This decreases everyone's workload, and also gives you all experience in working with data in a team setting. Many of the tools we've introduced, like GitHub and Slack, are so powerful in part because they facilitate collaborative data analysis. You'll put these to work on the final project, and use them to help your team manage the data set creation process.

Work groups are also designed to be support structures for you as you complete all aspects of the final project. Your group members should serve as resources as you complete all aspects of the final project. You've been grouped by topic area so that group members will all be working on similar projects.

Overview



Goal: create a plan for the group work portion of the project



Personnel: complete this vignette as a **group**



Due by: lecture-05 (February 19th)



Deliverable: meeting report in group’s Slack channel



Skills: lectures 1-4 plus new Git skills (see below)

Planning

Use GitHub’s “Project” feature on your work group’s GitHub repository to plan out the tasks you need to complete. Create a new project for your team. Assign each vignette that contains group work to a column, and then assign specific tasks and deliverables related to that vignette to individual notes in that column. Place the assigned team member’s name in the note as well to track who has agreed to take on which elements of the project.

Table 1. Work Group Division of Labor

Vignette	Member 1	Member 2	Member 3
3	Race Data	Poverty Data	-
4	Group 1	Group 2	Group 3
5	-	-	Geodatabase Creation
7	Map 3: Race	Map 4: Poverty	Map 2: City Reference

Each team member needs to take responsibility for a column. The third column, for team member 3, does not involve R work for the group aspects of the project but does involve a more complex map design in vignette 7 as well as some geodatabase creation. Keep that in mind as you divide up tasks between the three group members.

Each cell on the table above should be completed on its own branch in GitHub. A branch is a self-contained copy of the repository that allows you to make changes to the data in the repo without impacting the master copy.

Tracking Progress and Peer Review

As a group, you should create a timeline for when you would like group work to be completed. Keep in mind the official check-in dates on page 7. As you complete tasks related to your assigned elements

of vignettes, indicate that they have been completed on the relevant note in your team's GitHub Project.

When the branch is ready to be added to the master branch for your work group, the group member responsible for that branch should open a "pull request" and include a description of what you are proposing to integrate into the master branch.

Both other team members should evaluate each pull request, check the data in the branch and/or deliverables to make sure all of the necessary steps from the vignette have been completed, and comment on the pull request with your approval or your feedback about proposed changes.

If both other team members agree that the necessary steps from the vignette have been completed, one of those team members should accept the pull request and delete the branch.

If the team agrees that changes are warranted, the pull request should be canceled and the agreed-upon changes should be made. The member responsible for that element of the vignette should open a new pull request when they are ready for the work to be re-evaluated.

Once a pull request has been accepted, you may find that additional changes are warranted. If that happens, open up an issue on your work group's repository. Assign the issue to the person who took initial responsibility for that aspect of the project. In the issue you open, describe in detail what the concern is and how you identified it.

If you are assigned an issue, create a new branch or make the necessary changes and, when complete, close the issue and open a pull request to integrate those changes back into the master branch.

What to Store on GitHub

The elements that go into individual branches on GitHub should be the final deliverables listed at the end of each vignette. GitHub will also be used for storing your deliberations about pull requests and issues. Be mindful of organization and file naming - do not choose generic file names because they will cause issues when you merge the branches. `RaceMap.png` is an example of a good filename that will uniquely identify your work as opposed to something like `map.png`, which may not and may be overwritten if someone else also uses that filename in their branch.

Resources

In addition to your work group's GitHub repository, you have also been assigned a Slack channel for discussions.

Evaluation and Grading

When you submit your final poster, you will also need to fill out a short self-evaluation form and a short evaluation for each of your group members. These will be used to inform each students' participation grade for the second half of the semester. If issues arise throughout the group work process, please let Chris know sooner rather than later. It is important to note that what goes on your poster is ultimately each student's individual responsibility (and will be reflected in your grade). Please take the group work element of the final project seriously and provide constructive feedback to the other students in your work group.

2: *Cleaning the Citizen's Service Bureau Data*

As we mentioned previously, there are a number of issues with the data that are included in the `finalProject` repository. Your initial task is to rectify as many issues as possible in the data. Once the data are cleaned, you will need to create a subset of these data that contain only the observations relevant to your topic.

Overview



Goal: create a clean `.csv` file with only the observations that are relevant to your topic.



Personnel: complete this vignette **individually**



Due by: lecture-11 (April 2th)



Deliverables: a knit `.Rmd` notebook for data cleaning as well as cleaned CSB data stored in `.csv` and `.shp` formats.



Skills: lectures 1-3, 5, and 10

Analysis Development: Organization and Documentation

Set-up a well organized .Rmd notebook in your **personal** assignment repository's docs/ folder that uses literate programming techniques to document the process for each of the following steps. Update both your README.md and your RESEARCH_LOG.md files with relevant details as you go through this vignette. The README.md file should include a description of the cleaned, final data since they form the core of this project. The research log should document in greater detail the process for how these data were created.

Initial Cleaning

1. Import the raw data from where they are saved within the data/ folder.
2. Subset these data so that they do not contain three variables that are not needed for the final project: probcity (all observations are in the City of St. Louis), probzip (has lots of missing observations), and prjcompletedate (not relevant).
3. Rename all of the remaining variables so that they have short (ideally less than 9 characters in length), consistent, intuitive names.
4. Drop all observations where the status variable is equal to CANCEL or is empty.
5. Drop all observations with incomplete spatial data. These observations will have fewer than six characters for the x and y coordinates.
6. Identify and drop any duplicate observations that are identical across all of the remaining variables.
7. Confirm that the request identification number variable uniquely identifies observations.
8. There is an additional set of duplicate data - observations that are otherwise identical but have *different* request identification numbers. Identify and drop these observations as well.
9. There is one more set of duplicate data to handle - observations that are otherwise identical but have *different* request identification numbers *and* varying values for the x and y coordinate data. Identify and drop these observations as well.
10. At this point, you should confirm that your sample size is $n = 721,448$.

11. Drop the status variable, as it is no longer needed.
12. Re-order the variables so that the year variable is located directly after (to the right of) the request date variable.⁶
13. Drop all observations that do not have problem codes that match the problem codes that describe your topic.
14. If you are using *multiple* problem codes, create a new categorical variable that simply differentiates between the relevant codes.⁷ For instance, the following values are the focus of a sample project:

⁶ *Hint:* Use the `dplyr::select()` function to do this.

⁷ If you have only one problem code, you may skip this step.

Stray Animal	4,102	0.57	83.25
Stray Animal Cntnd	118	0.02	83.27
Stray Cat	1,623	0.22	83.49
Stray Dog At Large	9,036	1.25	84.75
Stray Dog Cntnd-ACC	1,416	0.20	84.94
Stray Dog Contained	939	0.13	85.07
Surrender Cat	18	0.00	85.84
Surrender Dog	23	0.00	85.84
Surrender Pet	151	0.02	85.86

These nine categories could be summarized as (1) stray animal, loose, (2) stray animal, contained, (3) animal surrender. You would therefore want to make a single categorical variable that properly summarizes each observation's problem code with this simpler set of categories.⁸

⁸ *Hint:* Use the `dplyr::case_when()` function.

15. Conduct a missing data analysis on your final data set and report the findings in your notebook.
16. Export these data as .csv data to your data/ folder.

Shapefile

17. Import your exported .csv file into ArcMap.
18. Project your data using the x and y coordinates already included in the tabular CSB data.
19. Save these data into a shapefile using a projected coordinate system appropriate for mapping citywide data. This shapefile should be stored in your data/ folder.

3: *Cleaning the Demographic Data*

One thing that we emphasized early in the semester was that social phenomena are often located spatially. It is not surprising, then, that we see certain types of social phenomena cluster in areas with particular demographic profiles. Two types of demographic data are needed for the final project: racial data and data on poverty rates.

Overview



Goal: create two shapefiles that contain demographic data on race and poverty respectively



Personnel: team members 1 and 2 from the **group**



Due by: lecture-11 (April 2th)



Deliverables: a knit .Rmd notebook for data access and cleaning, a preview map created using ggplot2, and a shapefile with the demographic data



Skills: lecture-04 and lectures 8-10

Analysis Development: Organization and Documentation

Both team members should create a new branch in the team's GitHub repository and add a top-level folder specific to this part of the project. Within that folder, create a project folder hierarchy with the needed subfolders. Also add a `README.md` file to document the purpose of this specific part of the team's repo. Once your file system is ready, set-up a well organized `.Rmd` notebook in your `docs/` folder that uses literate programming techniques to document the process for each of the following steps.

Separately, you will also want to take more detailed notes in your personal `RESEARCH_LOG.md` file about your process for creating these data.

Demographic Data

1. Depending on your assignment, download 2015 ACS 5-year estimates for table B02001 (Race) or B17021 (Poverty Status of Individuals in the Past 12 Months by Living Arrangement) at the census block group level using the `tidycensus` package.
2. Drop the variables `geoid` and `geodisplaylabel`. Also drop:
 - For table B02001 - drop all estimates and margins of error for non-White and non-African American population groups. Also keep the total population estimate.
 - For table B17021 - drop all estimates and margins of error except the main income below the poverty level and income at or above the poverty level. Also keep the total population estimate.
3. Rename all estimate variables so that they have simple, intuitive names. Leave the variable `geoid` named as such.
4. Ensure that numeric data are all stored in numeric format and that the `geoid` variable is stored in string format without any scientific notation.

Accessing Spatial Data

5. Using the `tigiris` package, download the 2016 census block groups for the City of St. Louis into your global environment.

Joining the Spatial and Tabular Data

6. Join the cleaned demographic data together with the spatial data you have downloaded.
7. Export these newly created and combined data to a shapefile. This shapefile should be stored in your data/ folder.

Modifying Your Shapefiles

8. In ArcGIS, add the 2016 Census Block Groups file for Missouri to a new map document.
9. If necessary, change the projected coordinate system to the coordinate system that the group decides on in Vignette 4.
10. Calculate the area of each Block Group in square miles.
11. In a Markdown formatted “metadata” file, add some notes about where the shapefile data were obtained from, what the final coordinate system is, what data were joined to them, and what the key fields and their meanings are.

Documentation

12. For each of the shapefiles, complete a short Markdown (.md) formatted metadata file that includes the following information. Sample files are included with each of the three shapefiles seeded to you in your work group repository.
 - (a) File name
 - (b) File description - what does it measure?
 - (c) Who created it?
 - (d) When was it created?
 - (e) What is the file’s projected coordinate system?
 - (f) What is the source of the data?
 - (g) How has it been modified?
 - (h) Key fields - what are their names, what type of data do they have, and what do they measure?
13. Save each metadata file along with the shapefile you have created.

Analysis Development: Peer Review

When you have completed committing all of your assigned data and documents to GitHub, open up a pull request so that your team members can peer review your work. Work collaboratively to determine if any shapefiles need further modification. If they do, make those changes and open a new pull request. One of your other team members should take responsibility for accepting your pull request when no further changes are needed and deleting your branch.

4: *Cleaning Spatial Data*

The spatial data needed for the final project needs to be obtained from a number of different public sources. These data provide important spatial context for both Missouri and St. Louis, describing major bodies of water, the street and highway systems, parks, and other major areas of green space. Each team member will be responsible for obtaining a portion of the needed data, modifying its extent or content, and saving new shapefiles in an appropriate coordinate system. You will also need to create a short metadata file for each shapefile you create or edit. All of the **raw data** should be retained after it is edited.

Overview



Goal: create three sets of shapefiles that have been cleaned and modified so that they are ready to be used in the project's maps



Personnel: each team member should complete their individual tasks



Progression: significant progress should be made by lecture-11 (April 2th) with final steps finished as soon as possible after lecture-13 (April 16th)



Deliverables: all listed shapefiles and corresponding metadata



Skills: lectures 4, 7-8, 10-11, and 13

Analysis Development: Organization and Documentation

All team members should each create a new branch in the team's GitHub repository and add a top-level folder specific to this part of the project. Within that folder, create a data/ folder with two sub-directories - rawData/ and cleanData/. Also add a README.md file to document the purpose of this specific part of the team's repo:

```
workgroup99-18/
  johnShapefiles/
    cleanData/
      STL_BOUNDARY_City/
        STL_BOUNDARY_City.shp

    rawData/
      MissouriCounties/
        original_county_data.shp

  README.md
```

Separately, you will also want to take more detailed notes in your personal RESEARCH_LOG.md file about your process for creating these data.

Projected Coordinate Systems

1. As a team, set an appropriate standard for the projected coordinate system used in mapping statewide data. All statewide shapefiles should be saved in this coordinate system.
2. As a team, set an appropriate standard for the projected coordinate system used in mapping City of St. Louis data. All local shapefiles should be saved in this coordinate system.

Data Included in Work Group Repository

To help get you started, we've seeded each group's repository with some data we've created specifically for this project. Each of these files includes a sample metadata file that you can use as a model for your own documentation.

3. **City of St. Louis Boundary** - this file is *slightly* different from the boundary file that we've included in the course data release. Please make sure to **always** use this boundary file for the final project.
4. **Illinois Hydrology** - The Census TIGER hydrology data is released by county, so bodies of water that lie between jurisdictions end up split. This shapefile contains data on the extent of the Mississippi River in Madison and St. Clair counties for the length of the City of St. Louis.
5. **Illinois River Islands** - The north side of the City sits opposite a group of islands on the Illinois side of the Mississippi, some of which are included in the **Illinois Hydrology** layer. The largest of these islands, Mosenthein Island, is described by this shapefile.

Team Member 1

6. **City of St. Louis Neighborhoods**

- (a) Download the raw data from the City of St. Louis's Public Datasets website (<http://data.stlouis-mo.gov/downloads.cfm>).
- (b) Re-project these data into the coordinate system your group selected in step 2 of this vignette.

7. **Belfontaine/Calvary Cemetery**

- (a) Using your re-projected neighborhood data, select neighborhood "86", which covers two large cemeteries in North St. Louis - the Belfontaine Cemetery and the Calvary Cemetery.
- (b) Create a new shapefile from the selected data and ensure that it is created using the coordinate system your group selected in step 2 of this vignette.

8. **City of St. Louis Street Centerlines**

- (a) On the Census Bureau's TIGER website (<https://www.census.gov/geo/maps-data/data/tiger-line.html>), choose 2016 ▷ Download ▷ Web Interface.

- (b) Using the drop-down menus, download the “All Roads” data for the City of St. Louis.
- (c) Re-project these data into the coordinate system your group selected in step 2 of this vignette.

9. City of St. Louis Blocks

- (a) Download the raw data from the City of St. Louis’s Public Datasets website (<http://data.stlouis-mo.gov/downloads.cfm>).
- (b) Re-project these data into the coordinate system your group selected in step 2 of this vignette.

10. Missouri Botanical Garden

- (a) Using your re-projected city block data, select the polygon with the handle attribute “1410400”, which is the city block that encompasses the Missouri Botanical Garden.
- (b) Create a new shapefile from the selected data and ensure that it is created using the coordinate system your group selected in step 2 of this vignette.

11. Regional Hydrology

- (a) Combine the Illinois hydrology shapefile described above with the STL_HYDRO_MajorBodies shapefile so that you have a single shapefile with all regional hydrology data included in it.
- (b) Combine all of the polygons for the Mississippi River so that it is represented by a single polygon.
- (c) Ensure that this new regional hydrology shapefile is created using the coordinate system your group selected in step 2 of this vignette.

12. Jefferson Expansion National Memorial

- (a) On the National Park Service's Data Store website, go to Park Boundaries and download the file `nps_boundaries.zip`.
- (b) Select the polygons that represent the Jefferson Expansion National Memorial, which is the National Park site that houses the Gateway Arch and the Old Courthouse in Downtown St. Louis.
- (c) Create a new shapefile from the selected data and ensure that it is created using the coordinate system your group selected in step 2 of this vignette.
- (d) Delete the portion of the National Memorial located on the east side of the Mississippi River. See [these instructions](#) for guidance on this task.

Team Member 2

14. City of St. Louis Landmarks

- (a) On the Census Bureau's TIGER website (<https://www.census.gov/geo/maps-data/data/tiger-line.html>), choose 2016 ▸ Download ▸ Web Interface.
- (b) Using the drop-down menus, download the "Area Landmark" data for the City of St. Louis.
- (c) Select the features that lie within the City of St. Louis.
- (d) Create a new shapefile from the selected data and ensure that it is created using the coordinate system your group selected in step 2 of this vignette.

15. City of St. Louis Railroad Centerlines

- (a) Use the drop-down menus on the same Census Bureau TIGER website to download the “U.S. Railroad” data.
- (b) Create a new shapefile from these data that contains only the railroad data within the City of St. Louis and ensure that it is created using the coordinate system your group selected in step 2 of this vignette.

16. City of St. Louis Parks

- (a) Download the raw data from the City of St. Louis’s Public Datasets website (<http://data.stlouis-mo.gov/downloads.cfm>).
- (b) Re-project these data into the coordinate system your group selected in step 2 of this vignette.

17. City of St. Louis Structure Footprints

- (a) On the University of Missouri’s MSDIS website (<http://www.msdis.missouri.edu/index.html>), and choose Data ▸ FTP Download.
- (b) If prompted for log-in credentials (will happen for macOS users), select “Guest”.
- (c) Navigate to Pub ▸ Facilities_Structures and download the file `M0_2014_St_Louis_City_Structure_Footprints_shp.zip`.
- (d) Re-project these data into the coordinate system your group selected in step 2 of this vignette.

18. State of Missouri Boundary

- (a) Copy the Missouri Boundary shapefile in the course data release (Data/CourseData/M0Boundary) into your final project data directory.
- (b) If necessary, re-project these data into the coordinate system your group selected in step 1 of this vignette.

19. State of Missouri Interstate Highway Centerlines

- (a) Use the drop-down menus on the same Census Bureau TIGER website from steps 14 and 15 to download the “Primary Roads” national file.
- (b) Create a new shapefile from these data that contains only the interstate highway data within the State of Missouri and ensure that it is created using the coordinate system your group selected in step 1 of this vignette.
- (c) Create a new shapefile that combines observations that share the same highway name for the FULLNAME attribute.
- (d) Create a new attribute for only the highway number (without the “I-”). Enter values *only* for the following highways: 29, 35, 44, 49, 55, 64, and 70.
- (e) Delete the shapefile that contains Missouri Interstates before observations were combined as it is no longer necessary.

20. State of Missouri Major Cities

- (a) Download the major cities shapefile from the University of Iowa’s GIS website (ftp://ftp.igsb.uiowa.edu/gis_library/USA/us_cities.htm).
- (b) Select all cities within Missouri and create a new layer.
- (c) Select the following “major” cities: Kansas City, St. Louis, Springfield, Columbia, Joplin, Jefferson City, Cape Girardeau, and Hannibal.
- (d) Create a new shapefile from these data that contains only those eight cities and ensure that it is created using the coordinate system your group selected in step 1 of this vignette.

*Team Member 3***21. City of St. Louis Interstate Highway Centerlines**

- (a) On the Census Bureau's TIGER website (<https://www.census.gov/geo/maps-data/data/tiger-line.html>), choose 2016 ▸ Download ▸ Web Interface.
- (b) Using the drop-down menus, download the "Primary Roads" national file.
- (c) Create a new shapefile from these data that contains only the interstate highway data within the City of St. Louis and ensure that it is created using the coordinate system your group selected in step 2 of this vignette.
- (d) Create a new shapefile that combines observations that share the same highway name for the FULLNAME attribute.
- (e) Create a new attribute for only the highway number (without the "I-"). Enter values *only* for the following highways: 44, 55, 64, 70, and 270.
- (f) Delete the shapefile that contains St. Louis interstates before observations were combined as it is no longer necessary.

22. City of St. Louis Secondary Road Centerlines

- (a) On the same Census Bureau TIGER website from step 21, use the drop-down menus to download the "Primary and Secondary Roads" file for Missouri.
- (b) Query the X attribute for state highways (the value "S") to limit your data only to State numbered highways. There are no current non-interstate U.S. highways within the City of St. Louis.
- (c) Create a new shapefile from these data that contains only the state highway data within the City of St. Louis and ensure that it is created using the coordinate system your group selected in step 2 of this vignette.

23. Illinois Interstate Highway Bridges

- (a) Use the same raw “Primary Roads” national data used in step 21 of this vignette.
- (b) Create a new shapefile from these data that contains only the interstate highway data within the Illinois side of the Mississippi River (using the Illinois Hydrology layer released to you in the work group repository). This new shapefile should include three bridges (from north to south) for I-270 (New Chain of Rocks Bridge), I-70 (Stan Musial Veterans Memorial Bridge), and I-64/I-55 (Poplar Street Bridge). Ensure that your new shapefile is created using the coordinate system your group selected in step 2 of this vignette.

24. Illinois Street Bridges

- (a) On the same Census Bureau TIGER website from step 21, use the drop-down menus to download the “All Roads” layers for Madison County, Illinois and St. Clair County, Illinois.
- (b) For each county, create a new shapefile from the raw data that contains only the street centerline data within the Illinois side of the Mississippi River (using the Illinois Hydrology layer released to you in the work group repository). For Madison County, this should include three bridges (from north to south) for I-270 (New Chain of Rocks Bridge), the Chain of Rocks Pedestrian Bridge, and McKinley Street (McKinley Bridge). For St. Clair County, this should include four bridges (from north to south) for I-70 (Stan Musial Veterans Memorial Bridge), the Martin Luther King Bridge, Washington Avenue (the Eads Bridge), and I-64/I-55 (Poplar Street Bridge). Ensure that your new shapefiles are created using the coordinate system your group selected in step 2 of this vignette.

25. Illinois Railroad Bridges

- (a) On the same Census Bureau TIGER website from step 21, use the drop-down menus to download the “U.S. Railroad” data.
- (b) Create a new shapefile from these data that contains only the railroad data within the Illinois side of the Mississippi River (using the Illinois Hydrology layer released to you in the work group repository). This new shapefile should include three bridges (from north to south): the Merchants Bridge, the Eads Bridge, and the MacArthur Bridge. Ensure that your new shapefiles are created using the coordinate system your group selected in step 2 of this vignette.

26. **City of St. Louis Boundary and Illinois River Islands** - Re-project these two shapefiles seeded to your group in the repository from NAD 1983 to the coordinate system your group has selected.

Documentation

27. For each of the shapefiles you create, complete a short Markdown formatted metadata file that includes the following information. Sample files are included with each of the three shapefiles seeded to you in your work group repository.
 - (a) File name
 - (b) File description - what does it measure?
 - (c) Who created it?
 - (d) When was it created?
 - (e) What is the file's projected coordinate system?
 - (f) What is the source of the data?
 - (g) How has it been modified?
 - (h) Key fields - what are their names, what type of data do they have, and what do they measure?
28. Save each metadata file along with the shapefile you have created.
29. Update your RESEARCH_LOG.md files with relevant details. The research log should document in greater detail than the metadata file the process for how these data were created.

Analysis Development: Peer Review

When you have completed committing all of your assigned shapefiles to GitHub, open up a pull request so that your team members can peer review your work. Work collaboratively to determine if any shapefiles need further modification. If they do, make those changes and open a new pull request. One of your other team members should take responsibility for accepting your pull request when no further changes are needed and deleting your branch.

5: Creating Geodatabases

All of the data committed to your work group repository on GitHub should be converted to geodatabase feature classes.

Overview



Goal: create two geodatabases and a well-organized metadictionary file



Personnel: team member 3 from the **group**



Progression: significant progress should be made by lecture-11 (April 2th) with final steps finished as soon as possible after lecture-13 (April 16th)



Deliverables: both geodatabases and the metadictionary



Skills: lecture-07

Analysis Development: Organization and Documentation

The third team member should each create a new branch in the team's GitHub repository and add a top-level folder specific to this part of the project. Within that folder they should add both geodatabases, the metadictionary, and a README.md file to document the purpose of this specific part of the team's repo:

```
workgroup99-18/
  geodatabases/
    cityGeodatabase.gdb
    metadictionary.xlsx
    stateGeodatabase.gdb
    README.md
```

Separately, you will also want to take more detailed notes in your personal RESEARCH_LOG.md file about your process for creating these data.

The State Geodatabase

1. Using ArcCatalog, create a new file geodatabase for storing your statewide extent data.
2. Import each of the statewide shapefiles into this geodatabase. There should be a total of three feature classes: the boundary (#18), interstate highways (#19), and major cities (#20).

The City Geodatabase

3. Using ArcCatalog, create a new file geodatabase for storing your citywide extent data.
4. Import each of the citywide shapefiles into this geodatabase. There should be a total of eighteen feature classes from steps #6 through #17 and #21 through #27 of the fourth vignette. There should be two additional feature classes added from the third vignette.

Documentation

5. Create a well-organized metadictionary file that lists the feature classes included in each geodatabase. This should be completed in a Microsoft Excel file.

Analysis Development: Peer Review

When you have completed committing all of your assigned shapefiles to GitHub, open up a pull request so that your team members can peer review your work. Work collaboratively to determine if any geodatabases or the accompanying documentation need further modification. If they do, make those changes and open a new pull request. One of your other team members should take responsibility for accepting your pull request when no further changes are needed and deleting your branch.

6: Wire Framing the Poster

Before you begin creating maps with the data you've cleaned, you'll want to wire frame out a design for your poster. If you'd like some inspiration, you can drop by the entrance to the Sociology Department's office suite, where a poster from a student last year is on display.

Overview



Goal: create a wireframe template for your research poster



Personnel: complete this vignette **individually**



Due by: lecture-11 (April 2th)



Deliverables: a .pdf of your PowerPoint or Keynote document



Skills: lecture-06

Analysis Development: Documentation and Organization

Create a folder within the docs/ folder of your **personal** assignments repository that is specific to this element of the assignment. Update both your README.md and your RESEARCH_LOG.md files with relevant details as you progress through this vignette. The README.md file should include a description of the final files. The research log should document in greater detail the process for how these files were created.

Poster Elements

Your poster should contain the following elements:

1. A title across the top of the poster.
2. Information about authorship, the course number, and the date of the final project presentation session.
3. Three text boxes:
 - (a) An introduction box that briefly addresses each of the following three questions - what is the Citizen's Service Bureau, what are you mapping, and why is it important? This introductory box will contain Map 1 (see below).
 - (b) A data and methods box that describes your data - what sources did you obtain data from? What have you done to modify these data?
 - (c) A discussion box that provides some brief discussion of your findings - where do you see the issues you found in the Citizen's Service Bureau data occurring in St. Louis? Does there appear to be overlap with the demographic map you produced?
4. If possible, a relevant photo that illustrates your topic.
5. Eight maps:⁹
 - (a) *Map 1*: A reference map to be included in your introductory box showing St. Louis's location relative to the rest of Missouri.
 - (b) *Map 2*: A reference map of the City of St. Louis.
 - (c) *Maps 3 & 4*: Two thematic demographic maps (race and poverty) showing data at the Census Block Group level.
 - (d) *Map 5*: A thematic map showing the number of CSB incidents for your topic at the Precinct level.

⁹ Keep in mind that St. Louis is tall and narrow when you estimate the size of your maps. It may be useful to export an image of the City boundary shapefile to aid in the wire framing process.

- (e) *Map 6*: A set of small multiples showing the number of CSB incidents for your topic per year at the Precinct level.
- (f) *Map 7*: An inset map showing the location of your specific ward, selected by you because of the high frequency of CSB incidents for your topic there, relative to the rest of the City.
- (g) *Map 8*: A detailed map of your selected ward showing point locations for CSB incidents for your topic along with local contextual details.

Creating Your Wire Frame

6. Using Microsoft PowerPoint or Apple Keynote, create a wire frame that lays out these elements. If you use Microsoft PowerPoint, set your slide size to 44" wide by 40" tall. If you use Apple Keynote, set your slide size to 3168pt wide by 2880pt tall.
7. Add wire frame boxes and reference text describing each box's dimensions for each element listed above.
8. Export your wire frame as a .pdf file.

7: Creating Maps 2 to 4

Three of the maps needed for the final poster will be produced collaboratively with your team members. These maps include a reference map for the City of St. Louis and two demographic thematic maps. See page Work with your group members to determine appropriate sizes for the following three maps. These sizing considerations should take into account all three wire frames that have been created by you and the other members of your group.

Overview



Goal: create three well designed demographic and reference maps that can be used by all group members on their posters



Personnel: each team member should complete their individual tasks



Due by: lecture-11 (April 2th)



Deliverables: map documents (.mxd) files, the needed data, and final exported map images



Skills: lectures 4-6

Analysis Development: Organization and Documentation

Each team member should create a new branch in the team's GitHub repository and add a top-level folder specific to this map. Within that folder, create a project folder hierarchy with the needed subfolders. Also add a README.md file to document the purpose of this specific part of the team's repo. Separately, you will also want to take more detailed notes in your personal RESEARCH_LOG.md file about your process for creating this map.

Map 2: City Reference Map

This map should give a basic overview of the City of St. Louis. Since many of your maps will be thematic maps that have limited spatial context, this map will be critical for allowing your audience to interpret where patterns on the thematic maps are located in relation to well-known landmarks.

The map should include data on the various types of streets (interstates, state highways, and surface streets), hydrology (major bodies of water), major areas of greenspace (parks, cemeteries, and other recreation sites), and the border between Missouri and Illinois. Be sure to include street data that spans the entire width of the Mississippi River to avoid the “bridge to nowhere” phenomenon.

Design Guidelines: There should not be any ground layers outside of the extent of the City boundary and the Mississippi River data. Appropriately label features to help provide additional context for your audience. When you export the map image, make sure that the map view is designed to limit Missouri's boundary from appearing to far above and below the City boundary. This map should be exported to match the background color of each group member's poster (which may entail the production of multiple map images) at 500dpi.

Map 3: Thematic Demographic Map, Race

This map should illustrate one of the defining demographic features of the city - the extent of racial residential segregation in St. Louis.

Design Guidelines: It should use a 5-class color ramp from Color Brewer to map *appropriately normalized* demographic data relating to race at the Census Block Group level. Work with your group members to select a color ramp that is unique from the other demographic map but does not clash with it. It should not have other layers outside of the thematic data. This map should be accompanied by a legend and should be exported to match the background color of each group member's poster (which may entail the production of multiple map images) at 500dpi.

Map 4: Thematic Demographic Map, Poverty

This map should illustrate another defining demographic feature of the city - the widespread poverty (or relative wealth) that is concentrated in particular parts of St. Louis.

Design Guidelines: It should use a 5-class color ramp from Color Brewer to map *appropriately normalized* demographic data relating to poverty at the Census Block Group level. Work with your group members to select a color ramp that is unique from the other demographic map but does not clash with it. It should not have other layers outside of the thematic data. This map should be accompanied by a legend and should be exported to match the background color of each group member's poster (which may entail the production of multiple map images) at 500dpi.

Analysis Development: Peer Review

When you have completed committing your map image to GitHub, open up a pull request so that your team members can peer review your work. Work collaboratively to determine if the map needs further modification. If it does, make those changes and open a new pull request. One of your other team members should take responsibility for accepting your pull request when no further changes are needed and deleting your branch.