YES Spatial Analysts Team: Teaching guide and discussion

Eric Darsow, Computer department, Community College of Allegheny County (CCAC)

Youth Enrichment Services Pb (lead) Exposure Student Research Project | Summer 2017

Table of Contents

Spatial Analyst team project overview

The spatial analysts were a four-student sub-team of the Lead Exposure summer research project coordinated by Youth Enrichment Services (YES) to investigate lead exposure levels and prevention measures in the Lincoln-Lemington neighborhood of Pittsburgh, PA. Our group was tasked with using spatial data and geographic information system software to gain insight into how the neighborhood's built and social environments impact residents' exposure to lead toxins.

The spatial analysts worked together for about 40 hours during July and August, 2017 to learn mapping concepts, build data plans, collect field data, and visualize the results for presentation to a public audience. None of the students had worked with professional mapping software before and most had only rudimentary exposure to social research methods. And yet, over the course of the project, the four spatial analysts developed core competencies in each important knowledge domain and produced professional-quality maps (see Figures 1 and 2 below) and academic posters. Additionally, our group gelled well and we enjoyed our work together.

Curriculum overview

The spatial analysis team of the YES summer research project engaged with each phase of the academic research process from the standpoint of geo-spatial relationships among built and social environmental features. The following table summarizes the core knowledge components explored during the learning sequence. See the "Student Outcomes" section for a detailed discussion of student outcomes in each area.

	Knowledge/Skill area
	Academic research process
	Internalization of core research question
	Research Method design
	Literature review
	Analyzing field data and drawing conclusions/ making recommendations
	Documenting and presenting findings to an audience
	Data collection
	Data schema (table) creation
	Collecting data in field
	Data entry and cleaning
	Technical tool use
	Human and physical mapping fundamentals
	Navigating geographic information system software
	Digitizing features of a paper map and attaching attribute fields to features
	Creating map layouts and choosing appropriate symbolization
	Auxiliary tool use
	Troubleshooting and debugging

Learning sequence overview

The spatial analysis team was tasked with progressing through each of the major phases of an academic inquiry over the period only 1.5 months during the summer of 2017. This learning schedule was very compact, indeed, and required agile teaching and mentoring throughout the process to address the differences in learning rates among the team members as well as the natural preferences each student had for various components of the project.

Learning phase overview

Our learning process progressed—roughly—through the following phases:

Phase #	Description	Length of focus
1	**Introductory fieldwork exploration**: mapping built and social environmental features on paper	4 hours
2	Exploration of **fundamental mapping concepts** using carto.com's online mapping tool	8 hours
3	**"Crash course" in navigating desktop mapping software** package QGIS: navigating the map, adding layers, adjusting layer properties	4 hours
4	Collecting and visualizing geospatial data about the **built environment** in Lincoln-Lemington: age of housing stock	4 hours
5	Collecting and visualizing geospatial data about the **social environment** in Lincoln-Lemington: children and adult gathering place use frequency	4 hours
6	Creating and tweaking **map layouts** from visualized data layers for image export and posting online	6 hours
7	Preparing **presentation aids** and practicing individual and group presentations	2 hours

Lesson structure and sequence: exploration not mastery

The project's life cycle was compacted into about 1.5 months from start to finish—meaning time available for exploring tangential or contextual concepts was extremely limited. In other words, we had to *cut out all the fat* and focus on only the core skills required to get data into the mapping system, conduct simple visualization, and ship the maps out for presentation.

Clearly, such a compacted timeline is not ideal from the standpoint of robust student learning across the major learning components. However, by working within Pittsburgh's Summer At Work program deadlines we were able to provide an opportunity for students with very little previous experience with the academic research process *exposure* to each of the core tasks that would be honed in a larger research endeavor.

In other words, our learning objective was not *mastery* but rather *exploration*. Instead of tests and formal learning requirements, each of our mini-lessons were followed immediately with application of the concepts to the Pb exposure project itself. Students experienced, in effect, a hybrid of a summer-school-like setup (structured lessons and exams) with a summer camp flare. The students embraced this model with gusto: we had near full attendance over the course of the project and they regularly made comments about wanting to devote more time to work in the computer lab/workshop than we had scheduled.

Geographic information system (GIS) tool overview

Geospatial analysis has occupied a central role in many utility and physical plant services long before computer technology progressed to our current stage in which powerful tools and great documentation are available for use by the public—without cost—and with moderate training. The geographic information system (GIS) component of the larger summer research project concerning Pb (lead) exposure involves learning about and using these tools, their features, and mapping fundamentals followed by visualizing and presenting findings.

Each participant in this project's sub-group learned to discuss mapping concepts fluently as a result of repeated practice with several different—but related—mapping tools. The core learning in this group emerged from hand-on experiences with the data gathering and preparation process and the subsequent analysis, visualization, and presentation of findings.

Carto vs. QGIS

Students primarily worked on the QGIS desktop software running on Ubuntu Linux boxes. We also used Carto's (formerly CartoDB's) online mapping tool to get our "feet wet" with the concept of layered data, background layers, and basic map navigation. The following table compares various attributes of each software package so other instructors can make informed choices about the best tools for a teaching endeavor:

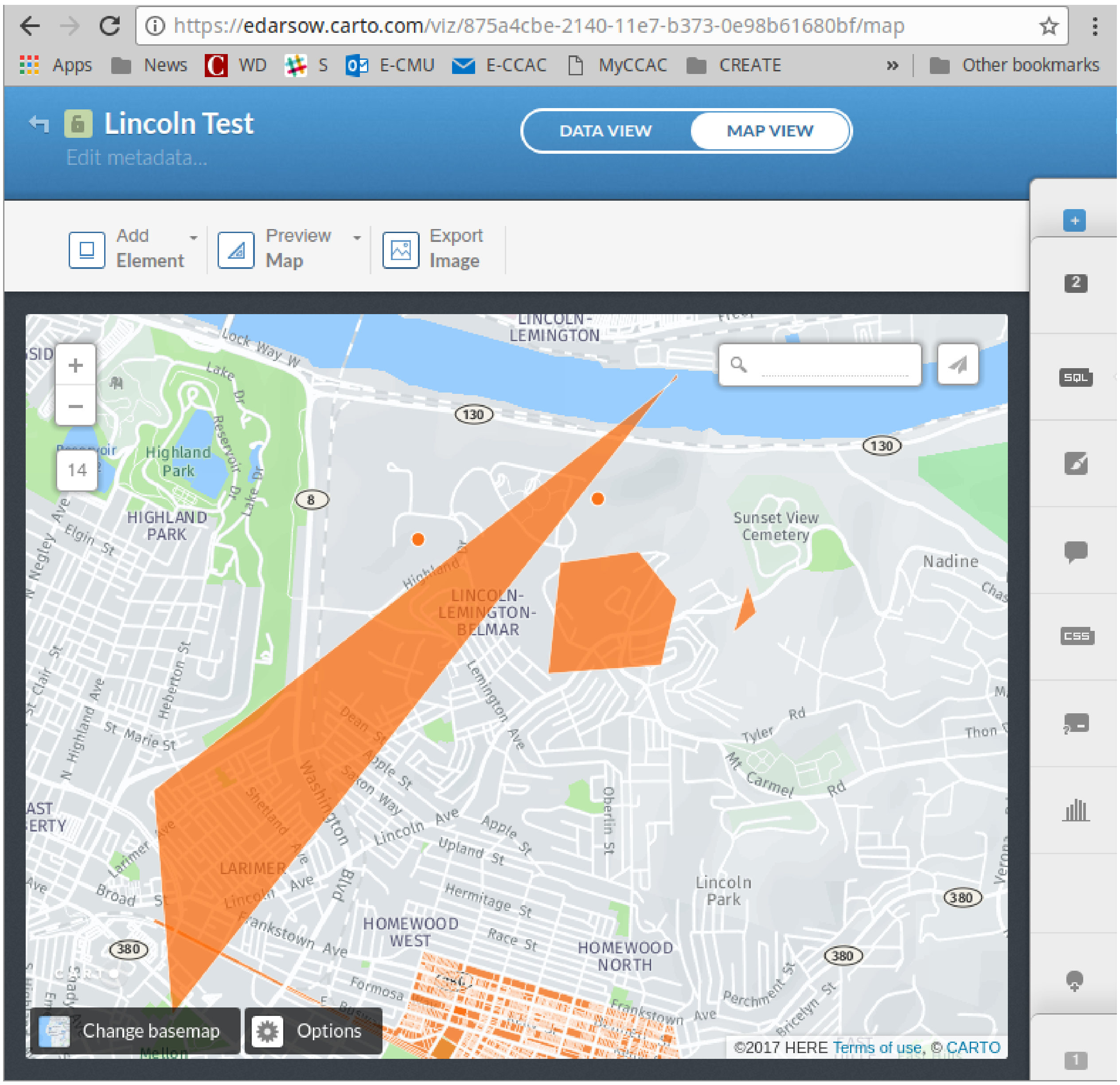
	Carto Online Mapper http://carto.com	**QGIS Desktop** http://qgis.org
Cost	**Free accounts** and many map and data sources can be uploaded for no cost—but all the data is publicly available. Users must pay a fee to restrict access to maps	**No charge for install or upgrade**. Code is "open source" and documentation is licensed under Creative Commons – Attribution share alike.
Platform and requirements	Runs in any **major web browser** (Chrome, Chromium (Linux), Firefox, Safari)	Compatible with **MS Windows, Linux, BSD, and Android.** Software package is "heavy" from a resource standpoint and should be run on systems with at least 8 gigabytes of available RAM to avoid sluggish response times
User interface characteristics	**Simple user interface** with javascript (responsive) controls and expanding tool bars. The new user interface Carto rolled out in 2016 takes some time to get used to	**Advanced user interface** with tool bars, right-click menus, menu bars, and many dialog boxes. Tool tips over menu buttons are very handy
"Horsepower"	**Relatively low**: Carto will not handle data layers with more than a thousand or so records easily. No support for joining tables	**Very high**: QGIS uses a Python interpreter for processing spatial data, which is an industry leading data manipulating programming language that can handle just about the largest data sets you can find.
Analytical tools	**Basic symbology controls**, layer formatting, digitizing, and data layer manipulations for primarily vector type data. Support exists for expanding	**Includes a full spectrum of mapping functions and tools for both vector and raster graphics.** Geoprocessing and geoanalytical tools are all available without cost. Many plugins exist for expanding tools with other open source modules.

	functions with widgets that can be installed on the platform and integrated into the user interface	
Student learning curve	**Relatively low** learning curve with plenty of online documentation for the simple suite of tools.	**Relatively high** since there are a number of settings required to be tweaked to visualize even basic layers. There are lots of online tutorials written by universities and companies who use this fantastic piece of open source software

Carto user interface

Note that Carto runs in a web browser and is account specific. This screen shot is of a map with a custom-made vector data layer (an arbitrary practice polygon). You can see the right-hand side bar is currently collapsed but, when expanded, allows users to manipulate layer-specific properties.

Figure -1: Carto.com online mapping interface and result of digitization practice

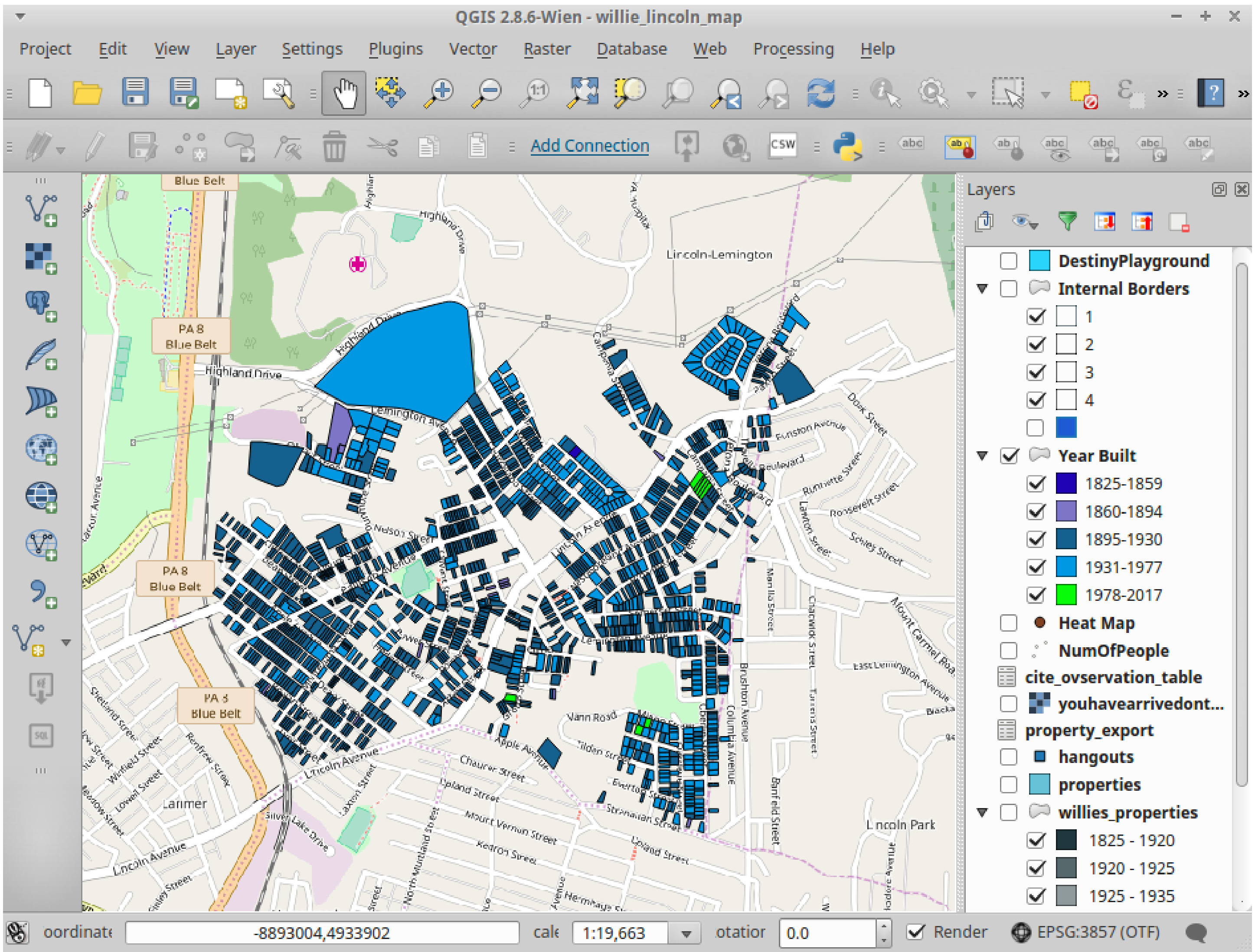


QGIS desktop user interface

QGIS's interface mirrors the basic functionality of the corporate and expensive ArcMap Desktop by ESRI corporation. The layers panel on the right allows users to toggle on and off various layers with ease. The menu and tool bars a the top allow users to manipulate the map view and access layer-specific functions. See qgis.org for links to full documentation.

One of QGIS's many plugins allows QGIS desktop to link to a Carto online account and upload and download layers between the desktop and server seamlessly. This makes pairing Carto and QGIS an ideal set because students can create advanced map layers in QGIS and upload them to Carto directly for accessing on any web-compatible device. Our students did this with their property assessment data and accessed the Carto map via their mobile computing devices (phones).

Figure 0: QGIS User interface and active property assessment layers with color coding by year built



Note on ESRI products

ESRI's ArcMap Desktop GIS software is the "industry leader" in mapping tools. However, ArcMap licensing is exceedingly expensive (thousands of dollars for a single license for a single computer) and notoriously difficult to install and keep running. The system requires an active internet connection during the entire user session in order to maintain links with the ESRI license servers. ESRI has rolled out a new version of the traditional ArcMap Desktop called ArcPro which is terribly buggy and extremely prone to crashing and fickle reverse compatibility issues.

Because QGIS is open source and licensed for open download, alteration, and republication, it is an ideal choice for student training since students can install copies of the software on virtually any system that meets hardware specifications. Additionally, students can market their new skills to businesses without that business needing to pay for a license for any software package: they can download, install, and begin producing "business value" within minutes.

Session Outlines

The following table presents information about each group meeting, including length, guiding questions, learning objectives, and learning exercises. Note that the sessions during July 3-8 are highly structured to provide students exposure to the core concepts that they implemented during the later, once-per-week meetings.

This progression from more to less work structure makes pedagogical sense because students needed a shared foundation of skills that would be used in each of the group's sub-projects. After the foundation was solid, students began working on more individualized sub-projects for which there were not as many shared skills. With more time, each student would have been able to gain proficiency in each skill in each sub-group. Given the actual time constraints, however, students branched off to their own mini-projects and helped one another with discrete skills as needed but less whole-group instruction was appropriate given the different rates in which each of the sub-projects progressed.

Date	Length	General Topic & Guiding Questions	Objectives	Learning Exercises
MON 3 JUL	2.5 hr	**Representing the world symbolically** * *What kinds of information can we convey spatially & symbolically?* * *How might we represent or convey information that we can't convey spatially & symbolically?* * *What principles can we use to choose representations of given data?*	* Students will be able to (SWBAT) make a spatial & symbolic representation of a public area around the classroom space using symbols of their choosing. * SWBAT evaluate the best way to convey information of various kinds. * SWBAT discuss their choices of what data to represent and how they represented it using principles of sound spatial representations. * SWBAT discuss potential shortcomings of spatial symbolic representations and describe steps to avoid them.	* Make a hand-drawn map of a public area. Participants choose the broad category of information they want to try to capture. * Share and discuss each person's map and the data they represented. Question: What is strong about their symbols? What could be improved? * Evaluate a set of data about the dangers of lead poisoning and choose some ways to convey this information.
TUE 5 JUL	4 hr	Map and geospatial data fundamentals * *How can computers help us in visualizing spatial data? How can they get in the way?*	* SWBAT discuss the benefits and drawbacks of computers as a mapping tool. * SWBAT describe the variation in features that exist among common mapping	* Guided practice through using CartoDB to map data from the USCUR Open Data Portal. * Group task: Use CARTO do map another data set on

		* *What software exists to help us? What are the pros and cons of web-server based GIS tools versus those running on a local machine?* * *How can I use a computer to map coordinates? Lines? Polygons?*	tools? * SWBAT use CartoDB to make a map of the housing parcels in the study area. * SWBAT use QGIS to make a map of housing parcels. * SWBAT compare what the tools have in common and what makes them different.	their own relevant to lead levels. * Guided practice through using QGIS to do the same thing we did in Carto. * Group Task: Use QGIS for the same task done as a group in CARTO * Discussion of the differences and similarities in mapping and technical process between the two tools.
WED 6 JUL	4 hr	Mapping conceptual framework * *What is actually happening when we make a flat map of a large, curved surface?* * *When does my knowledge of how maps are made enter into my data visualizing that I might do day-to-day as a GIS professional?*	* SWBAT model (actually do it) how a spherical surface is projected onto a flat one. * SWBAT describe how projections algorithms influence how a map "looks" and map these variances onto actual perceptions of the world. * SWBAT interpret geospatial meta data and tinker with its parameters in both CartoDB and QGIS	* Using blow-up globes and overhead projector film, we'll make two global maps using two different projection algorithms and discuss the tradeoffs in this fundamentally imperfect process.
THU 7 JUL	4 hr	Geospatial mapping "mini project" Part 1: Plan and Gather * *What data might be relevant to gather if we want to understand how a community "functions" or "works"?* * *What format should each type of data be stored in? How should it be captured? How can it*	* SWBAT Design a data gathering plan tied to a specific question about a place. * SWBAT implement a data gathering project by heading out into a public space and gathering both qualitative data from conversations with residents and quantitative or discrete pieces of data.	

		be checked for accuracy? What level of error is acceptable?		
FRI 8 JUL	4 hr	Geospatial mapping “mini project” Part 2: Analyze and visualize * *What do I do now? I have all of this data and a blank map.* * *What is the best way to represent the data I collected?* * *How can I blend both qualitative and quantitative data that is relevant to my inquiry area?*	* SWBAT encode their gathered data in proper formats (text versus spatial versus numeric) and store that data in a computer? * SWBAT analyze their data by focusing on certain variables of interest, learning generally about the data, and then looking for trends within the data. * SWBAT use a mapping tool of their choice to make a spatial and symbolic representation of their data. * SWBAT discuss their findings and presentation choices with the group.	This day is day 2 of the mini-project. Participants will have gathered a chunk of data from field work on Day 1 and will be guided through the process of creating a representation of that data that illuminates any conclusions or questions their data bring. Participants will be given wide latitude in the tools they choose to use, and what features of those tools are most helpful.
TUE 11 JUL	4 hr	**Loosely structured work time:** Loading up the PIE data into QGIS. Visualizing by property. Adding keys to the data table in the spreadsheet. Sharing the spreadsheet on google drive. Printing out walk sheets.		Activities 5-8 as needed
TUE 18 JUL	4 hr	**Loosely structured work time:** Social Environmental mapping: bus stop map, schools and churches, and gathering spots (parks, etc.) We digitized those points and learned how to make map layouts with the print layout tool		Activities 5-8 as needed
TUE 25 JUL	4 hr	**Loosely structured work time:** ***** Creating heat maps based on the social survey. * Continuing work on built environment maps.		Activities 5-8 as needed
TUE 1 AUG	6 hr	**Loosely structured work time:** * Student sub-group work time with intensive facilitator guidance		Activities 5-8 as needed
TUE 8 AUG	6 hr	**Loosely structured work time:** * Student sub-group work time with intensive facilitator guidance		Activities 5-8 as needed

Lesson activity notes

The following sub-sections discuss the structure and outcomes of a handful of discrete lesson activities we undertook over the course of the project. Teachers are encouraged to adapt each of these components to their specific projects and contexts.

Note that some of these activities are more like teaching approaches than discrete chunks of steps that students do with their teacher. This is intended to provide the reader with a flavor of possible “teacher moves” that can be incorporated into any teaching context with virtually any student group set.

Activity 1: Data organization 101—Creating a data table about ourselves

The foundation concept in data science and, by extension, mapping software is the data schema. A schema is a structure for arranging information or data. Humans often represent object and their data attributes in a data table.

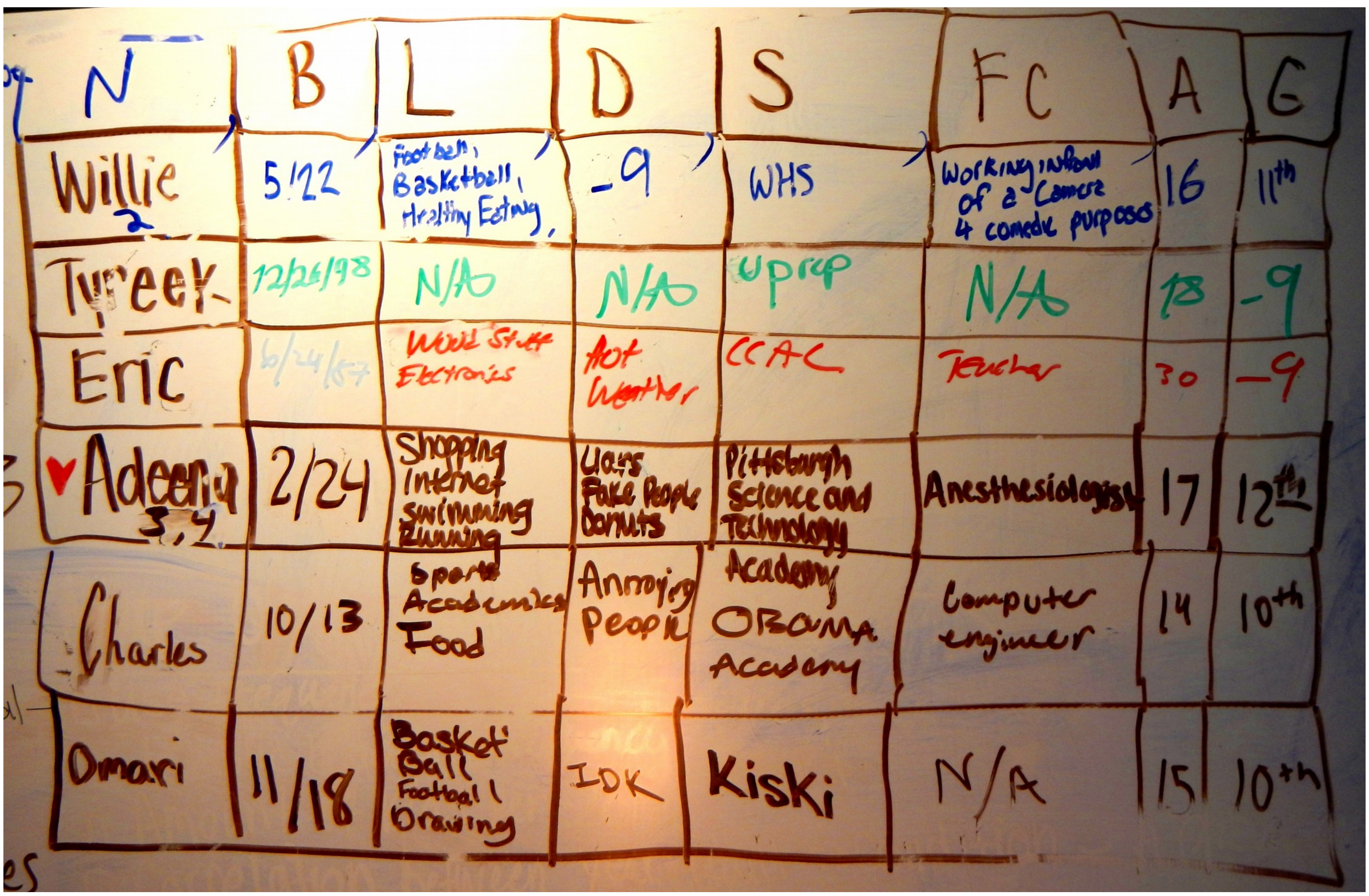
We learned about this concept by creating a data schema for sharing personal information about each group member. First, we brainstormed what we'd like to know about each other: likes, dislikes, age, school, etc. This was done in list form. Then we brainstormed how we could capture this data in an organized way and I facilitated this discussion leading to creating a data table. The most important concept in this process is the notion that the "variables" we captured about each person becomes the header of each column of data. Individual "observations" are group members, and they are stored in individual rows, or records.

The photo below is the completed data table we created with a record (row) for each group member, including myself, the instructor. The variable names (attributes) were abbreviated for ease of transferring to column names. Here is the data dictionary for the attributes:

N: Name | B: Birthday | L: Likes | D: Dislikes | S: School | FC: Future Career | A: Age | G: Grade

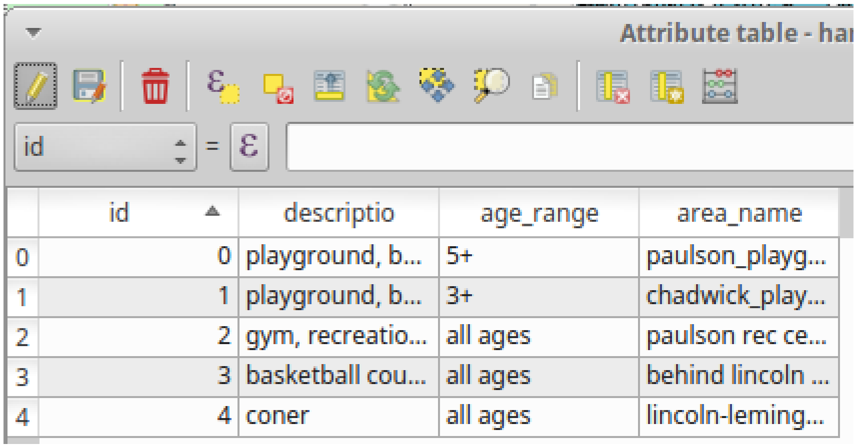
Note that we discussed using a "null value" marker of -9 instead of "n/a" to learn about how we can assist ourselves when sorting data and determining what data values are missing. We can easily sort by numerically ascending values and all the records with a -9 value in the row of interest will be grouped. Using "n/a" not only involves the special "/" character but is text-based, so sorting alphabetically places these null records in the middle of non-null records.

Figure 1: Group personal information data table



Now let’s look at screen shot of the attribute table for the gathering place points that the students digitized in QGIS:

Figure 2: Layer attribute table in QGIS


Attribute table - ha

id = ε

	id	descriptio	age_range	area_name
0	0	playground, b...	5+	paulson_playg...
1	1	playground, b...	3+	chadwick_play...
2	2	gym, recreatio...	all ages	paulson rec ce...
3	3	basketball cou...	all ages	behind lincoln ...
4	4	coner	all ages	lincoln-leming...

Note the similarities between the hand-made data table and the digital version of the data table created for points on a QGIS map representing gathering places. Just as we brainstormed attributes about each person in the group we wanted to share with one another, we created the data schema for gathering places in Lincoln by brainstorming what we wanted to record about each site, and these became the column headers in Figure 2: id number, description, age_range, and area_name.

By first creating a data table with easily identified data (e.g. what school do you attend), students could focus on the organization of the data. When confronted with a technical set of data, such as the gathering places, creating the data schemas on the computer was a straightforward task.

We kept our original table of personal data on the white board in the shop throughout the life of the project such that we could refer back to the concepts we learned easily. As students designed more data tables, I would often have us huddle around the personal data table and remind ourselves how we decided where to put variable names and where to place observations (in that case, an observation = a group member).

Activity 2: Physically projecting a sphere onto a plane

A foundational concept in cartography (mapping) is the notion of a projection from a sphere (the earth) onto a plane (a paper or digital map). This notion is theoretical and somewhat abstract to grasp in one’s brain. To make the process more concrete, we used blow-up globes to represent the spherical earth and overlaid the globes with overhead transparency film folded into cones and cylinders. We then used markers to project points from the sphere onto the overhead film.

The key to projecting a point from a sphere onto a plane is to mark a point on the plane which is located by extending a line from the center of the sphere, through the point on the surface of a sphere, to the planar surface. If some students project a set of points on the globe onto a plane folded into a cone and others project those same points onto a plane folded into a cylinder, we can compare the two projection methods by unfolding the cones and cylinders and overlaying them on one another.

This process reveals that there is no perfectly accurate method for transferring from a sphere on to a plane. This is the nature of the geometric relationships between planes and spheres. Hence, students can physically grasp the notion that how features on a map are related to one another (by distance and angle) is dictated by the method of projection.

Activity 3: From Carto to QGIS

To ease the students into the more complex QGIS features, we started by digitizing and loading data into a Carto map before even touching QGIS. The students are much more familiar with the “slippy map” interface which Carto maps, Google maps, and others use an this allows them to start learning about maps and layers without also having to learn what each of the tool buttons do in QGIS.

After we had created a simple map in Carto by digitizing arbitrary features, we produced a close copy of the Carto map in QGIS. This sequence allows students to navigate QGIS with some conceptual understanding. Developing the foundational knowledge of layers and their types data in Carto lowered the *barrier of entry* into a relatively complex desktop program.

The core principle of this activity: reduce the barriers to entry wherever possible to facilitate maximum ease in learning complicated tasks. This principle can guide lesson sequencing in a variety of disciplines and classes.

Activity 4: Gathering data from the WPRDC & American FactFinder

Pittsburgh is fortunate to be home to the University of Pittsburgh’s Regional Data center which houses and publishes data sets from various public agencies on the city and county level. The Western Pennsylvania Regional Data Center (WPRDC) is the web-based portal for accessing the public data. As an open access site, users can download hundreds of data sets in various formats through a well-designed user interface.

The spatial analysts team was particularly interested in property level data which is published on the WPRDC’s special Property Information Extractor (known as the PIE—see Figure 4). Our group learned how to navigate that system (see Figure 3 below) to retrieve all parcels in Lincoln Lemington and selected attributes for each parcel. The PIE system actually runs Carto in the back end so transferring the government data into one’s own Carto account is a seamless process. Shapefile versions of the data can then be downloaded from the Carto account and loaded into QGIS for processing.

Figure 3: Western Pennsylvania Regional Data Center download portal

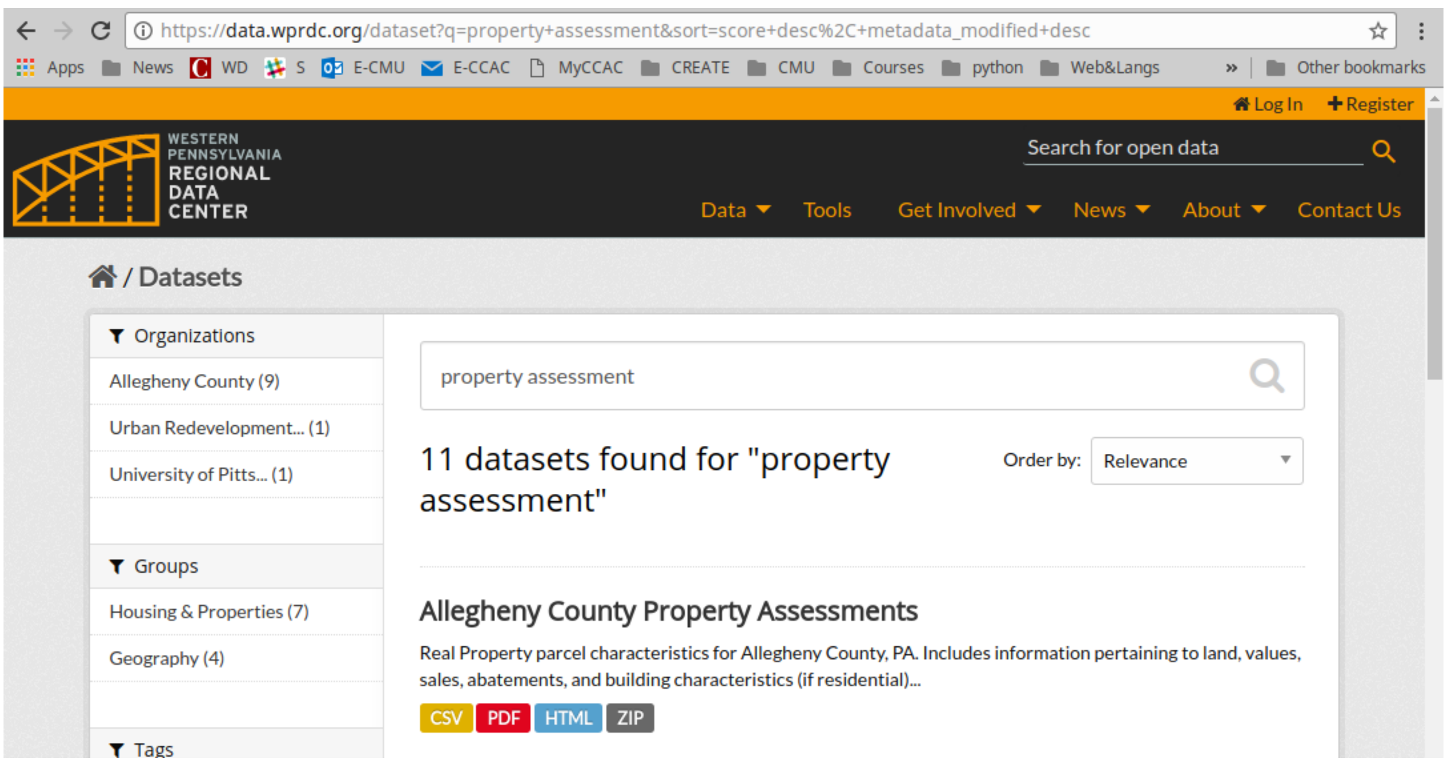
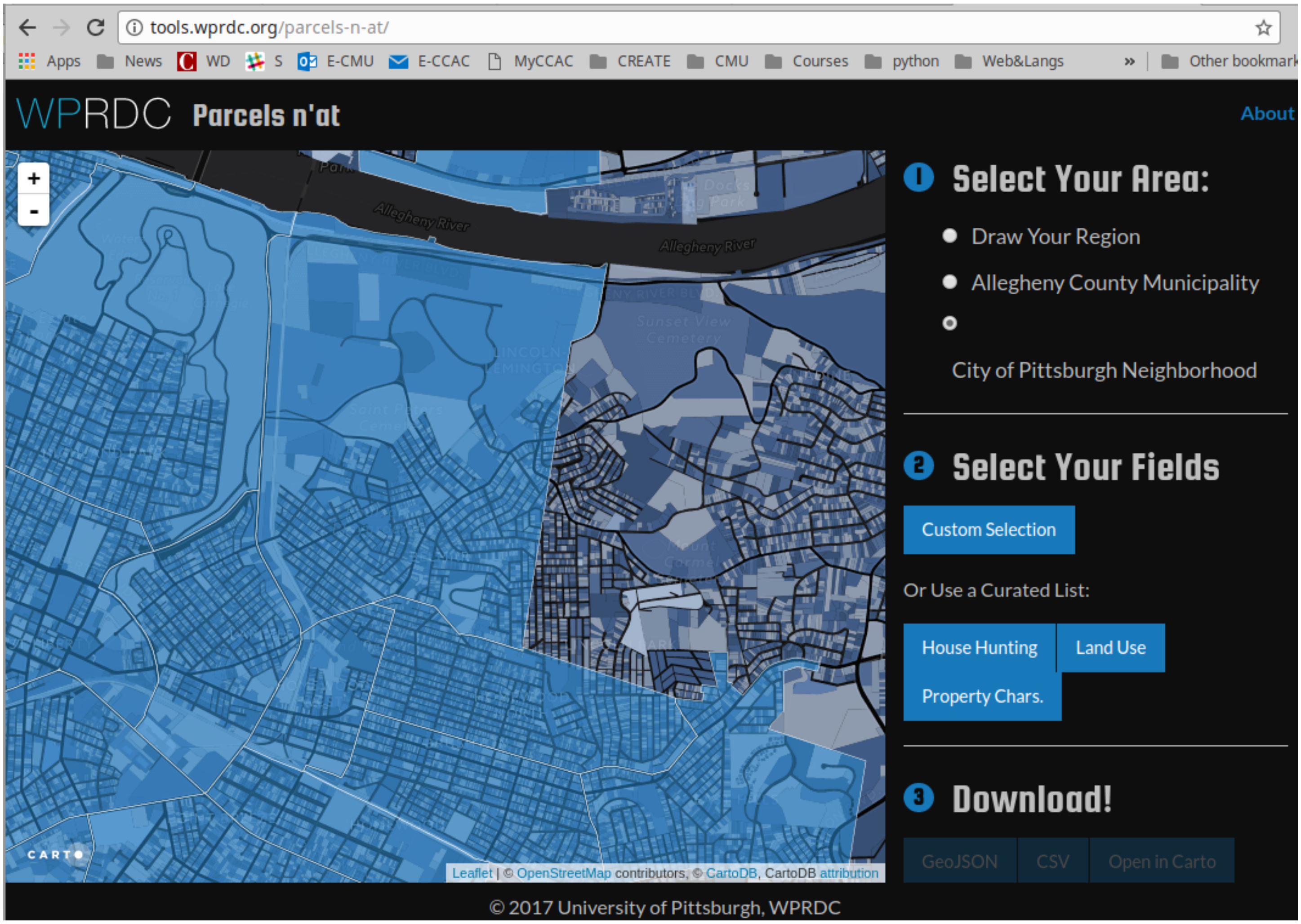
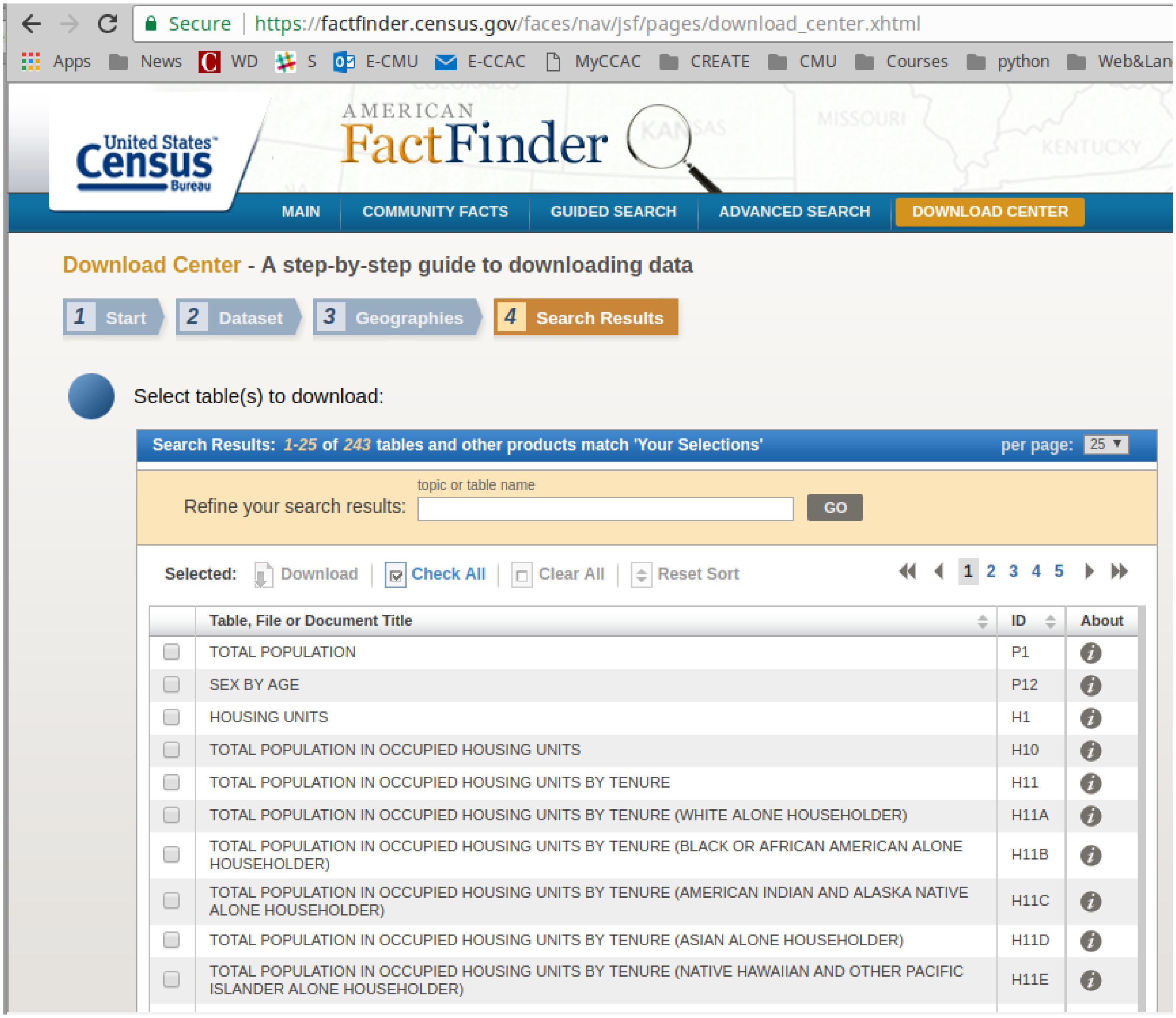


Figure 4: WPRDC's Property Information Extractor for retrieving county property assessment data



One group member who was exploring the distribution of children across the neighborhood of Lincoln-Lemington. With assistance, he accessed US Census Summary File 1 data to retrieve block-level data on all census blocks in the neighborhood. This data is open access and can be downloaded with some somewhat tricky clicking and reading through the American Factfinder data download tool (See Figure 5).

Figure 5: American FactFinder data download portal. This screen shot shows a data search results page containing a list of Summary File 1 data for blocks in Allegheny County



Activity 5: Collaborative—but independent—work time

After our first week of foundational activities, each subsequent Tuesday's work time involved what I term independent but collaborative work time. It is independent in the sense that each student was tackling a sub-component of our larger mapping project. It is collaborative because as the facilitator, I brokered moments of students teaching certain skills to his/her peers during this work time.

For example, Wilie was our resident expert in digitizing features into a data layer. He and I worked together to develop these skills. When another student needed to digitize content for an independent project, I would connect him/her to Willie and Willie would provide peer guidance through the process. This strategic pairing of students on a task-by-task basis is very effective at coping with skill differences in the group.

Ideally, each student would become a "mini-expert" in one technical area and can comfortably teach and guide other students when those particular skills are needed. This frees he facilitator to cultivate new skills with one student while other group members are teach one another what they've already learned.

Activity 6: Iterative design and practice

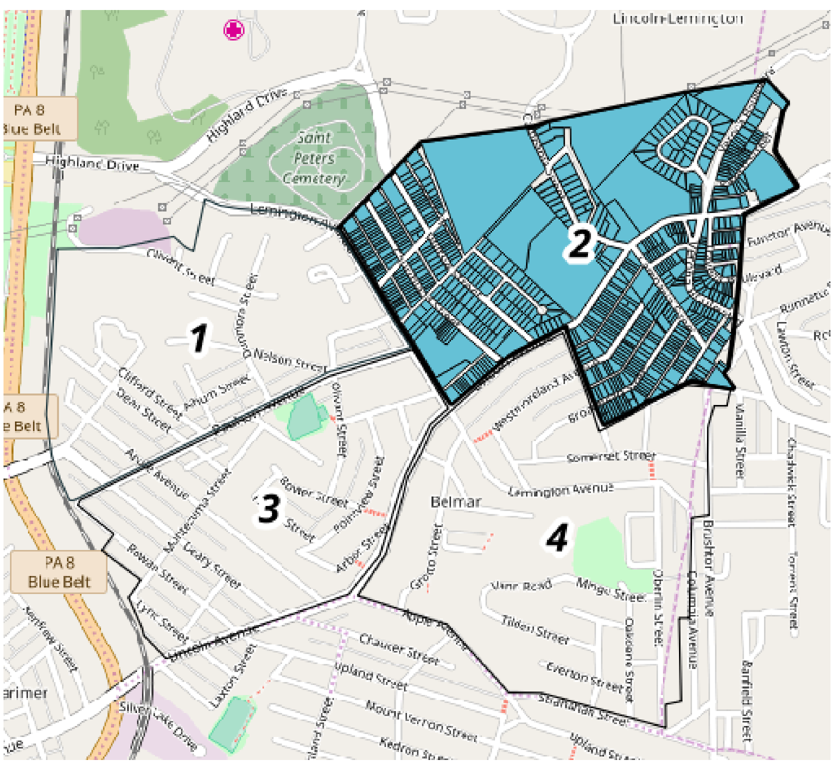
Learning how to load maps with data, visualize that data, and prepare the products for sharing is a complicated, multi-step process that, if pursued from the start with a high quality expectation, will quickly lead to burnout and frustration among students. The key to avoiding this negative experience, is to actively facilitate iterative design. Iteration refers to creating a product of decent quality, assessing that product, and deciding on a next set of improvement to implement on the next iteration.

We used this approach continually in this project. A sample of the iterative design sequence that emerged as we digitized our research zones is included below.

Figure 6: Iterative design sequence for digitizing research zones

1. Learn how to digitize an arbitrary polygon on a map (see Figure -1)
2. Secure the paper version of the four research zones from the project facilitator. Digitize the rough boundaries between zones
3. Show the draft digitization to the project facilitator for feedback and adjustments
4. Re-edit the digitized layer based on facilitator feedback
5. Try using the zone polygons for selecting parcel polygons inside each. Students determined that the digitized boundaries they originally created were not carefully placed on the outside of all enclosed parcel polygons. Adjust the boundaries again.
6. Re-attempt to use the polygons to select parcel polygons: success!
7. Format final zone map for printing and distribution to the team. See Figure 7 for final zone map

Figure 7: Research zone polygons digitized from paper version of the zones. Note that Zone 2 contains parcel polygons that were selected using a geoprocessing tool and saved as their own layer



Activity 7: Mini-tutorials

An important principle of learning how to use complicated technical tools is to break down information and skills into small, useful chunks. In context of learning how to use QGIS, I worked with the students together or sub-groups of students to learn how to do a small skill or learn about a small concept. I call these mini-tutorials because they are delivered as stand-alone modules and don't take very much time. They should be delivered when they are needed and not front-loaded before students need to use them because the students will likely lose the knowledge before it's needed.

In terms of preparation, the teacher should be familiar with the core modules that likely will need be taught such that each one can be delivered confidently and smoothly when the time arises. Choosing when to start a mini-tutorial should be made from a clear knowledge of what steps each student currently tackling.

The following list of concepts are sample topics I taught in mini-tutorials by are by no means exhaustive. Dozens of these types of teaching "interventions" were conducted over the course of the project.

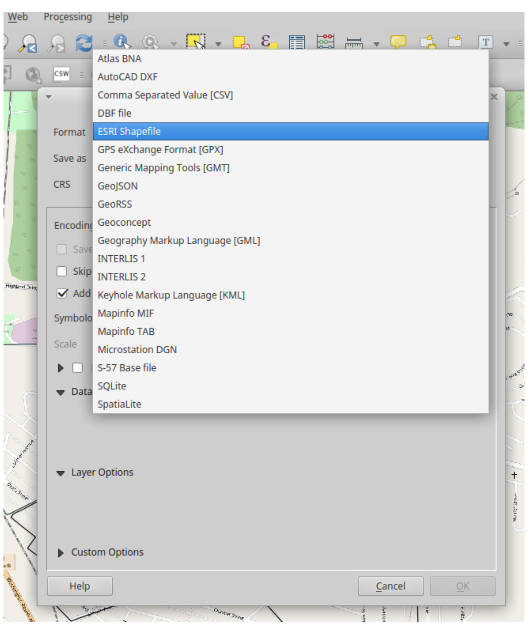
File extensions

Today's mobile computers and many Windows and OSX systems mask the entire notion of file extensions in order to create a "cleaner" user experience. The dreary downside of this trend toward of ease-of-use is that folks who use computers all day long often do not have a concept of file types and their relationship to a file extension (such as .docx for MS Word documents or .png for graphics).

The file extensions mini-tutorial was given to the whole group for a few minutes as we dove into QGIS at first, but had to be re-administered during specific sub-tasks that were file extension dependent, such as saving an exported layer as an ESRI Shapefile.

The reason this is considered a mini-tutorial and not just a "telling of a command" is that students benefit greatly from understanding the concept behind file extensions (i.e. extensions provide clues about file types) as they work on many mapping related tasks. I was sure to show examples of how the programs we were using required files to have certain file extension for various operations and to demonstrate what the error messages look like when there is a file extension error.

Figure 8: Choosing a file extension for a layer export operation



Importing and exporting data across software tools

One critical—but potentially confounding—element of the mapping analysis process is the use of multiple software packages to manipulate data and prepare products for sharing and display. The most critical of these accessory tools is, of course, spreadsheets which allow the user to assemble and easily edit tabular data which is exportable in a variety of formats compatible with QGIS.

Since there are many dimensions to multi-application work flows, this mini-tutorial took place iteratively over the course of the project. When we were entering data from field work into a spreadsheet, we learned how comma separated value files are generated and how QGIS uses delimiter characters to re-assemble the tabular data exported from a spreadsheet system.

Other accessory tools we used were document viewers, image viewers and editors, web browsers, and file explorers.

Students these days also don't have much experience with the notion of a file explorer because phones and tablets provide a very simple, non-tree-structure file browser. A mini-tutorial on file explorers, for example, should include some experimenting with directory trees, the difference between "back" and "up" navigation buttons, creating directories, deleting files, renaming files, and checking file names for extensions in order to route a file to a useful program for viewing and editing the files.

Map symbology best practices

Choosing how to color and shape features on a map is as much an art as it is a science. Mini-tutorials on symbology occurred when students had data entered into the system and loaded into a data layer in QGIS. The core entry into symbology settings is the layer properties dialog box. Students learned how to choose symbology types (graduate colors, graduated size, etc.) by experimenting with the various color options and growing familiar with the somewhat non-intuitive interface of the properties dialog box. See Figure 9 below. Given repeated practice, students quickly became familiar with the ways in which colors can be selected and adjusted in the properties dialog box.

The mini-tutorial also needs to include a discussion of how to use the color options to draw out the meaning of the data we were trying to visualize. For example, Figure 10 shows a screen shot of Willie's symbology properties dialog box for the parcel data layer which includes a field "yearBuilt" variable for each property.

The goal was to create a color scheme to draw attention to the age of the housing stock as it relates to Pb (lead) exposure. Figure 10, however, is a first draft attempt at creating a meaningful symbol scheme. The grey scale band is an appropriate firs choice but does not help the viewer of the map grasp any significance in the map's pattern other than "darker gray is older".

Figure 9, however, shows the final iteration of this design. Not only do the coloring bands include a junction in 1978 (the year paint with Pb was made illegal), but the color scheme for pre-1987 houses is made of blues and purples and the pre-1978 properties were colored in green which allow the viewer to easily notice which parcels were built before and after 1978.

Figure 9: The layer properties dialog box in QGIS for selecting colors for parcel years constructed

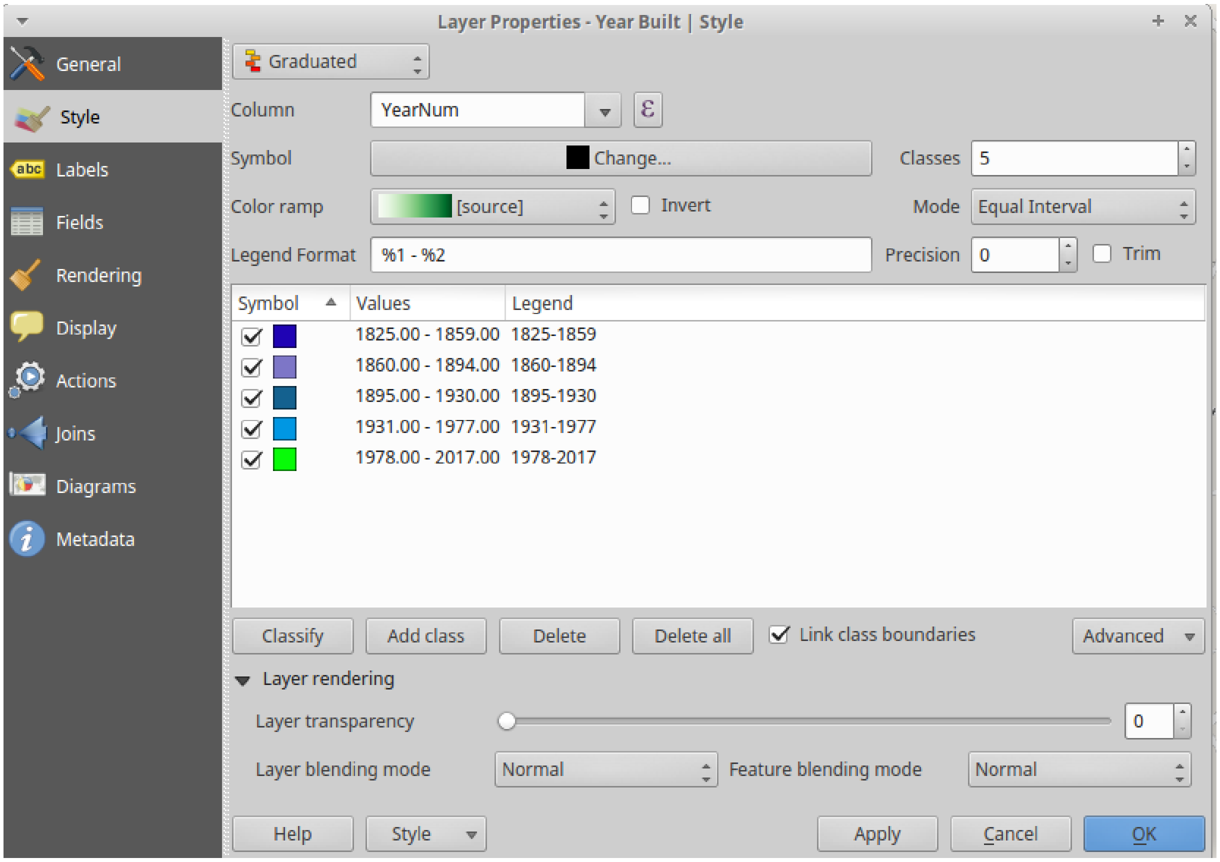
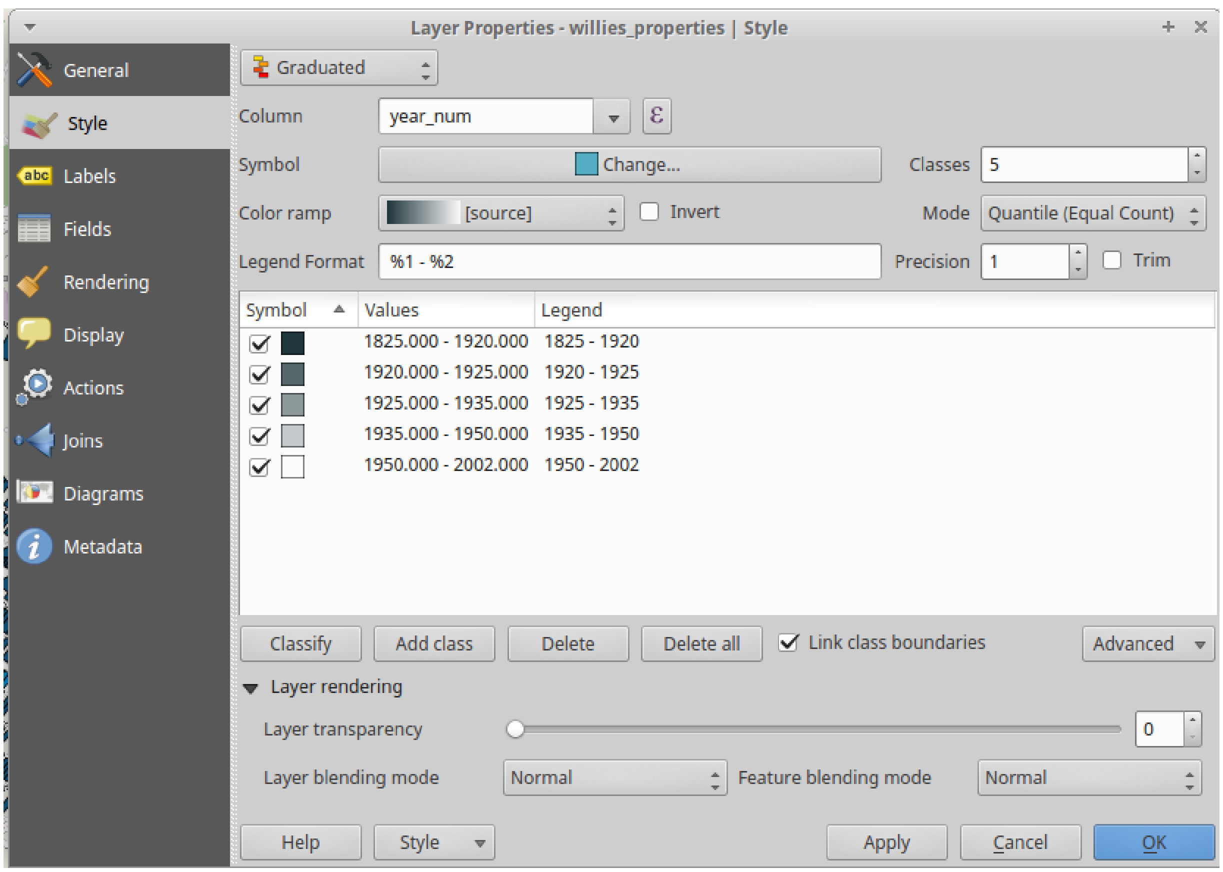


Figure 10: First attempt at creating a color scheme for property year bu

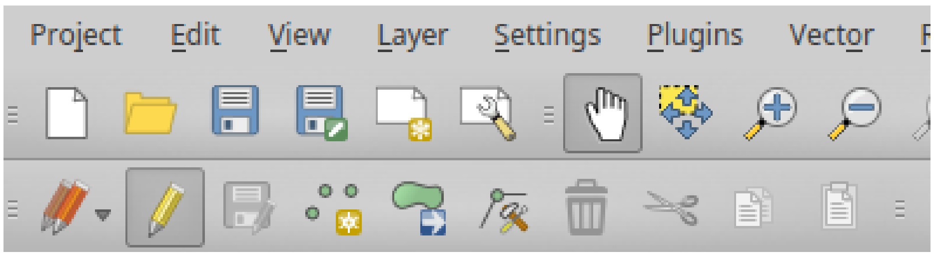


Digitizing features

Learning to digitize features requires patience and practice. Unlike many “drag and drop” application functions, digitizing features involves learning a set of about 5 non-intuitive keyboard and mouse click combinations to enable the user to pan the map and zoom while still editing polygon nodes.

The value of the mini-tutorial is that students learn about each of the tools at the very moment they need to use them—not stacked in the middle of a longer set of instructions about a larger set of tools. Increasing proximity of the tool tutorial to the time of use is critical for effective technical instruction.

Figure 11: The layer editing toolbar (lower row of buttons) which was the subject of a series of mini-tutorials on how to digitize a feature in a map layer



Activity 8:Work process logs

A work process log is an informal document created and maintained by the creator of a project to document the specific work steps taken throughout the endeavor. The raw work process log is generally not published or distributed along with the overall work outputs. The immeasurable value of a detailed work process log stems from its use in tracing back any steps taken to allow for effective editing and adjustment of the work in the future.

During this project, we made the mistake of not starting our work process logs until midway through our work. This was partly in the interest of increasing the pace of work during the heavy learning phase, but, upon reflection, should have been an expectation from the start of the project.

Even if students may not ever return to their work and need to review a work process log, the act of writing, in sentences and/or bullets, the steps taken forces the students’ mind to achieve clarity about what is actually being done. A work process log should be in a clear format that is comfortable to the student making the log. Some students are most comfortable assembling steps in a word processor, while others might choose to use a spreadsheet or even a database.

Activity 9: Troubleshooting troubleshooting

A most important component of technical education involves teaching students how to help themselves when they get stuck. This could be as simple as pointing students to the technical documentation for a given tool and guiding them through the organization scheme of the sections. When a more complicated problem arises, students also need guidance on how to effectively search for an answer to their question which involves the following skills:

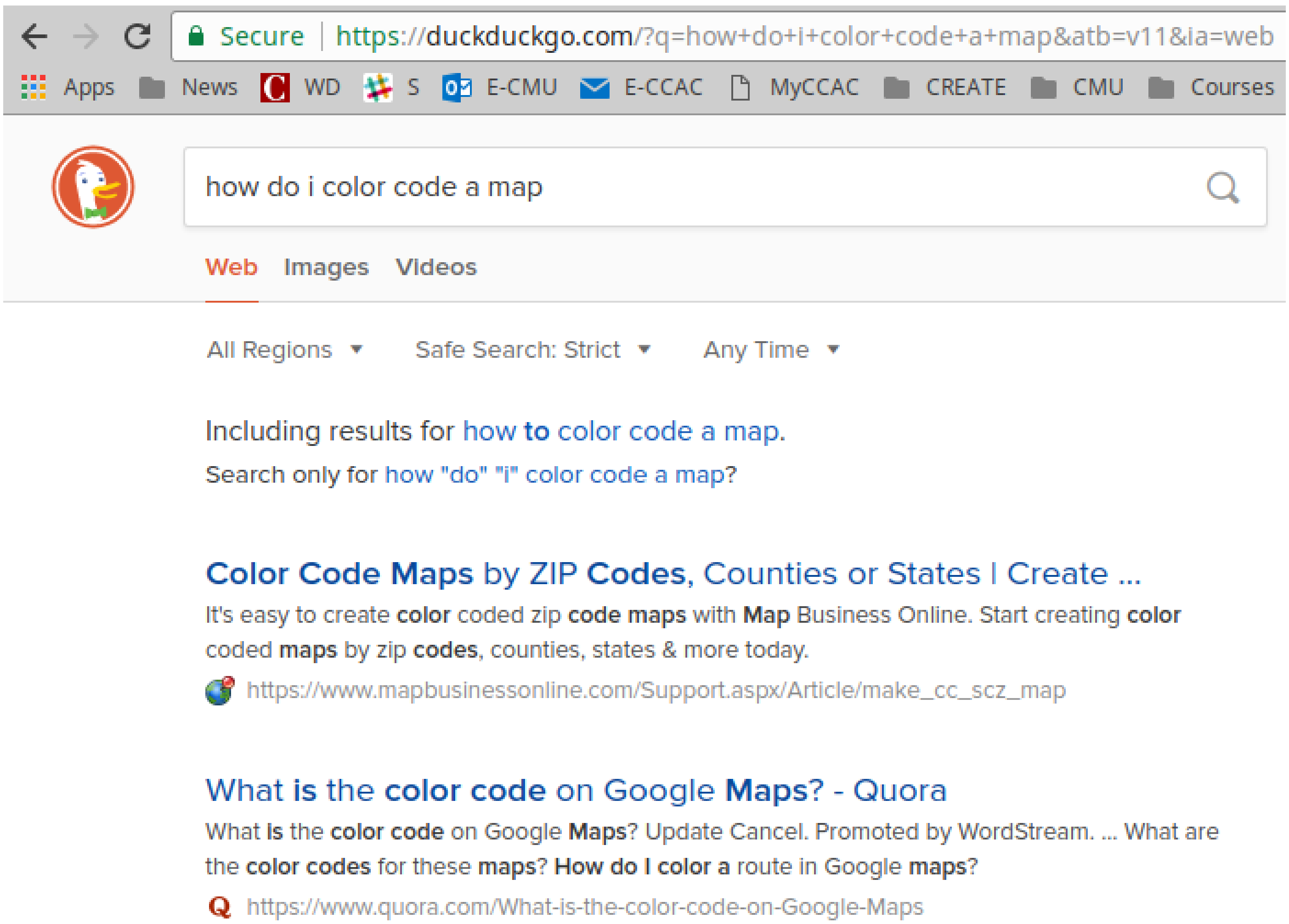
1. Formulating a clear search string (the words one types into the search box in a search engine) that will produce the most relevant results first
2. Assessing the credibility of the source of information

3. Sifting through question and answer forums, such as Stack Overflow, to locate an answer that is reflecting a scenario closest to the issue at hand
4. Adapting the solution found online to the student's actual problem.

This seems like a straightforward task, but students are often used to searching for answers to relatively simple questions, such as the number of feet in a mile. When looking up tutorials for QGIS, for example, the first few search results may be erroneous, and students can sink a tremendous amount of time trying to decode an answer to a problem that is too different from their actual issue to be helpful, causing great frustration.

For example, let's imagine that a student has parcel polygons and year-built data loaded into QGIS and she wants to color older houses a dark red and newer houses a green color. If a student is simply told: do an online search for your problem, she might type "how do I color code a map" into a search engine, yielding these results (see figure 12):

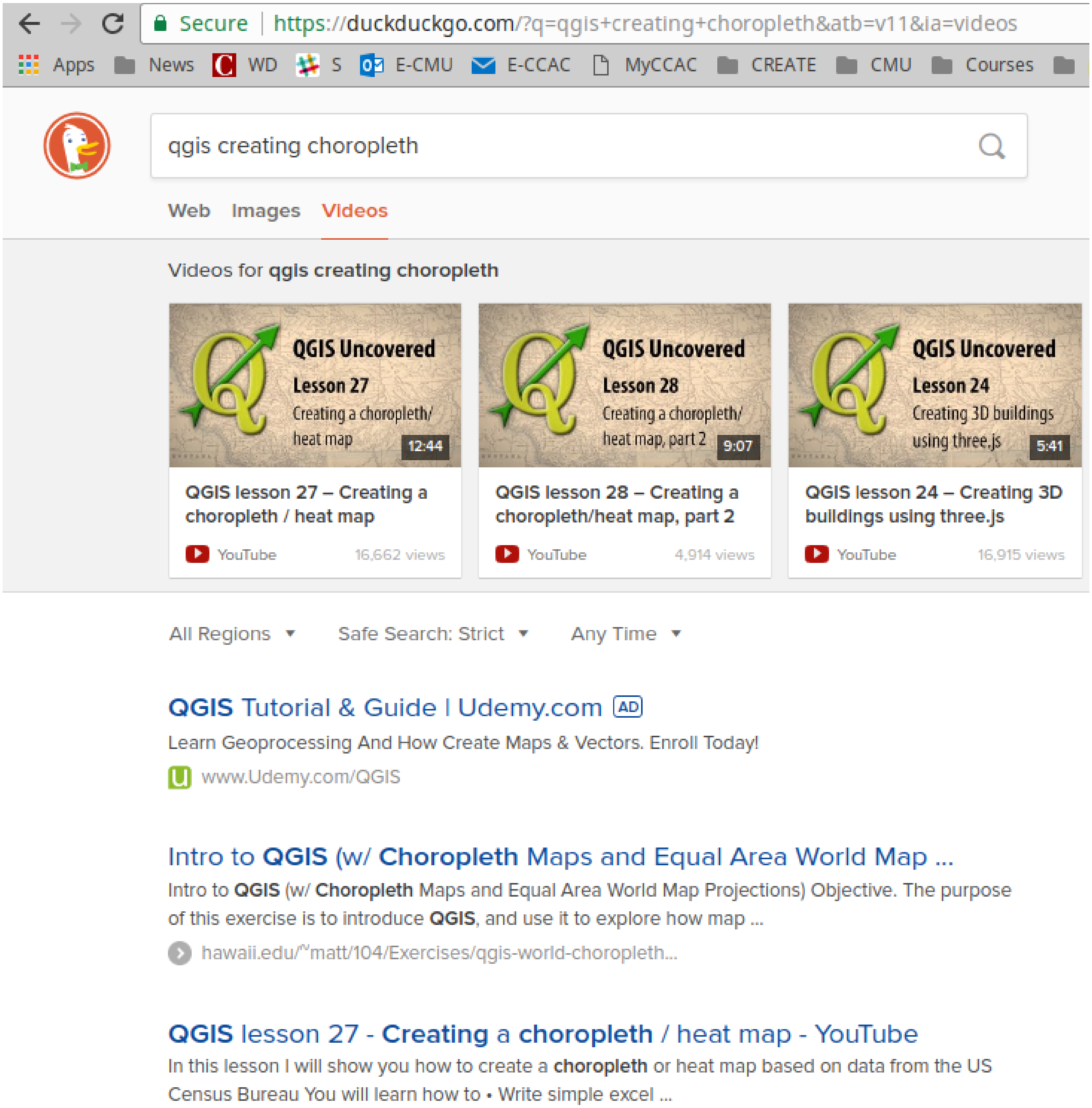
Figure 12: A poorly designed search string to find a tutorial on creating a choropleth map



Note that these results are not specific to QGIS nor are they even specific to the more common ESRI Arc Map. This should be fixed by adding the name of the software package to the search string: "qgis." Next, the student phrased the search in terms of a question, which is a great way to get a search result list filled with yucky sites like quora.com which simply aggregates actual questions and answers to millions of random issues. Quora requires logins and is full of advertisements.

With some prompting from the teacher, students can learn that the kind of map that has “color-coded” features is called a choropleth map (see Figure 0 for an example). If they simply add the name of the software package to the term “choropleth”, many useful results pop to the top of the search results, as shown in figure 13 below.

Figure 13: A useful search string yielding useful tutorials on creating a choropleth map in QGIS



Notice that when the software package name and the technical term is sent to the search engine, all of the firs three results are relevant, and the second result is hosted by a *.edu host, which is likely to involve few adds and is likely to be written by somebody with decent technical know-how.

The teacher should guide this process through a series of questions to help the student first identify what they believe the problem is and how to phrase that issue in technical language that aligns to the terms used in technical documentation and Q&A forums.

A few tips to guide this process:

1. Start search strings with the name and version of the software being used. For QGIS, this is straightforward: type in “qgis” and then search terms corresponding to the issue
2. If an error message is produced by the software, students can often copy and paste that data directly into a search engine and prepending the message with the name of the program. Often, tools will produce an error code and an error message corresponding to that code. In this case, often typing in the name of the program and then the error code is sufficient.
3. When using Q&A forums like Stack Overflow, do not assume that the question the original poster is asking is directly relevant to the real issue at hand. Instead, at least skim the original question to verify that the poster is using the same software package and is having a similar issue to the real issue.
4. When using Q&A forums, do not assume that the first (or highest voted) answer is the most relevant to the issue at hand. Take a moment to read several answers and review the comments on the answers which often clarify murky points or actually disagree with the poster of that particular answer.
5. Learn the technical names for the tools or utilities being used within a given software package. For example, the map area displayed in QGIS is called the “map extent”. Students may not know this precise term and therefore will not be able to help a search engine find relevant results for the issue at hand. In general, therefore, teachers should help students learn the technical names for parts of a program to aid them in finding answers to problems on their own in the future.
6. Teachers should resist the temptation to simply tell students the answer to their questions if the teacher is reasonably confident that the students can, with a few minutes of work, locate the answer themselves. Teaching folks how to help themselves sets them up for the ability to do work on their own in the future. This takes more class time, for sure, but those invested resources pay large dividends in the future.

Student work products

Each student (or student pair) produced a printable map layout based on the data they captured and visualized in map layers. The final layouts of each student are show in Figures 14, 15, and 16 below. A brief discussion of the strengths and areas for improvement of each product follows the images.

Figure 14: Final Product: "Heat Map" depicting areas of high children concentration based on a field-based enumeration of public space users

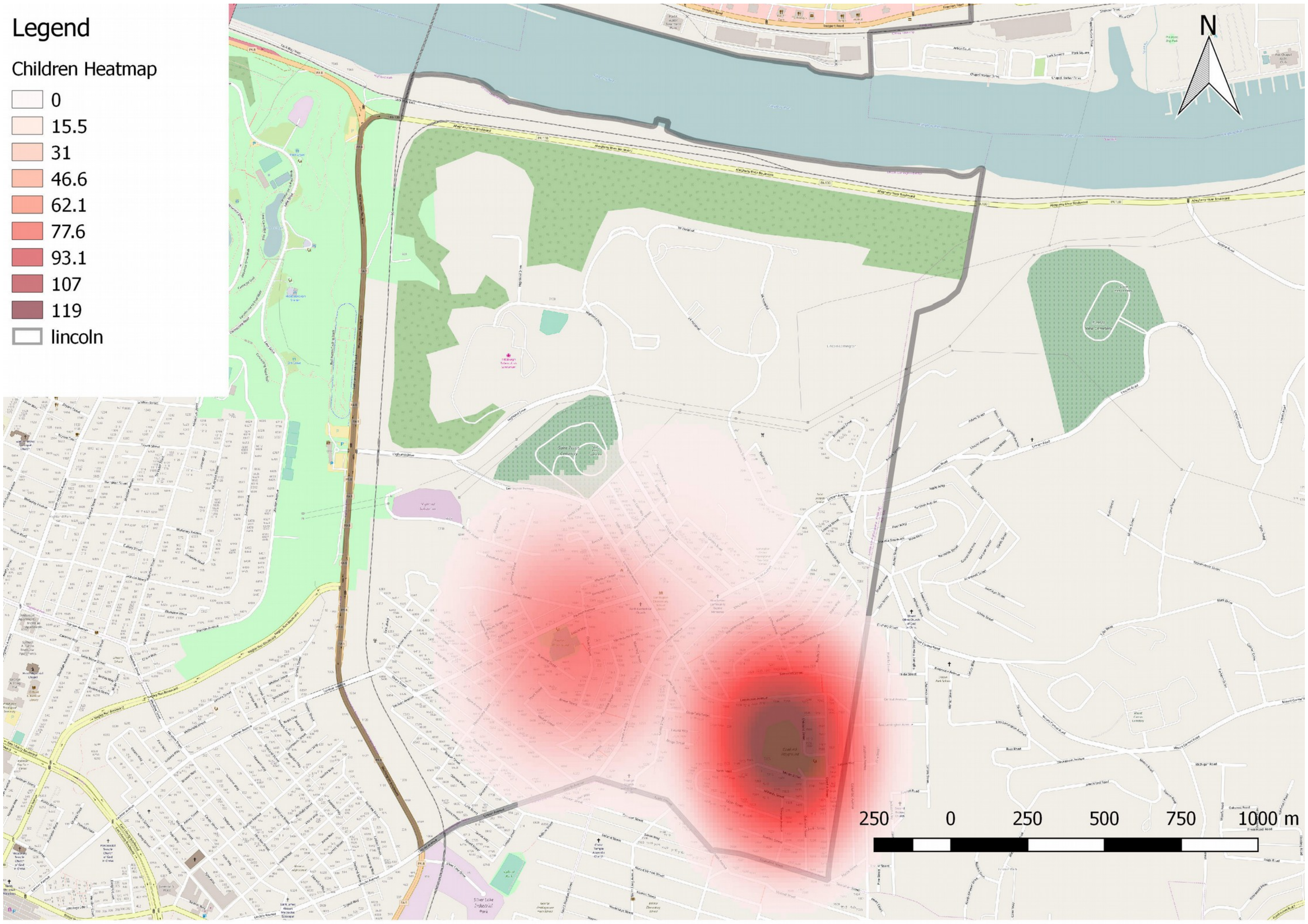


Figure 15: Final Product – Age of housing stock in Lincoln-Lemington

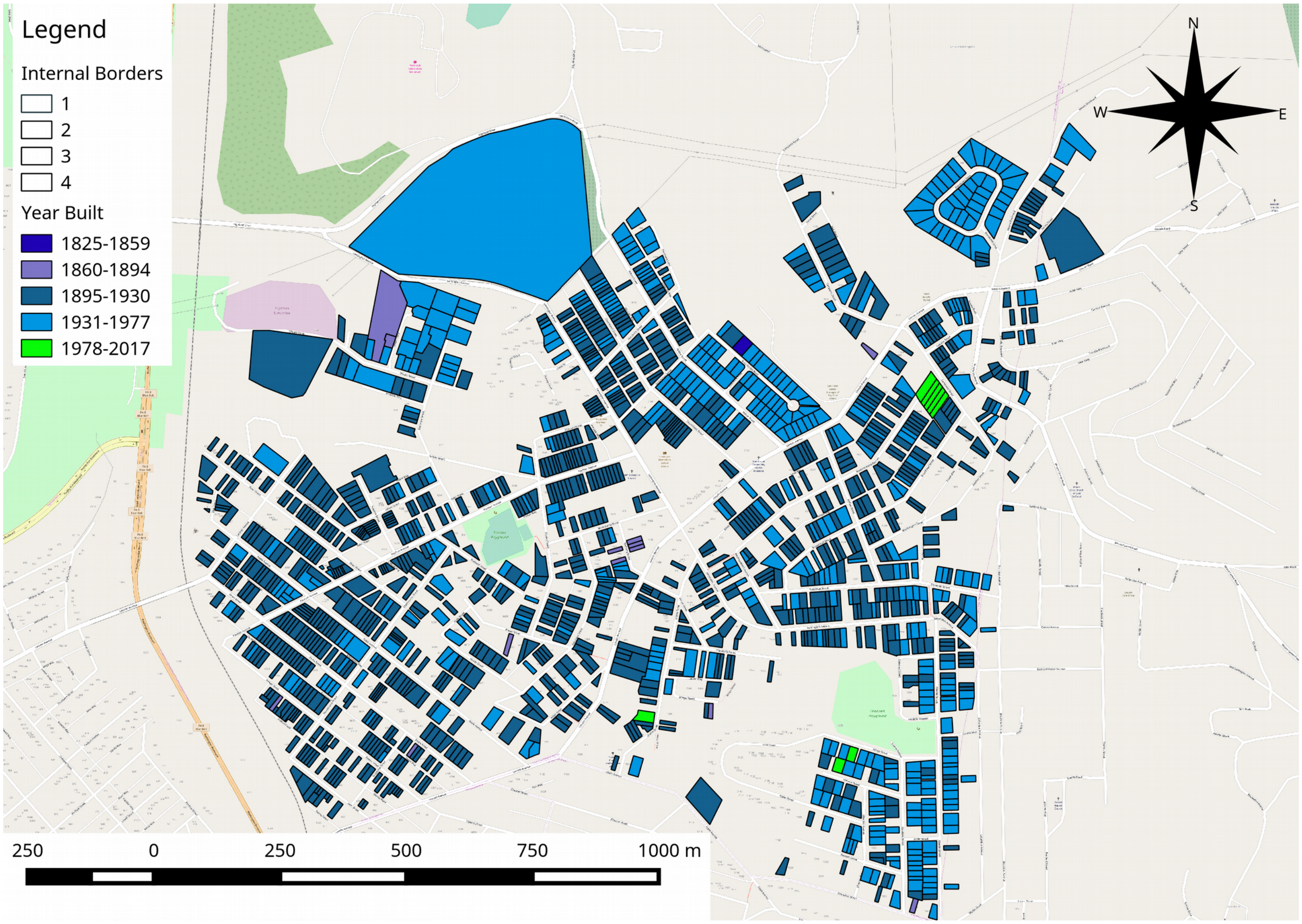
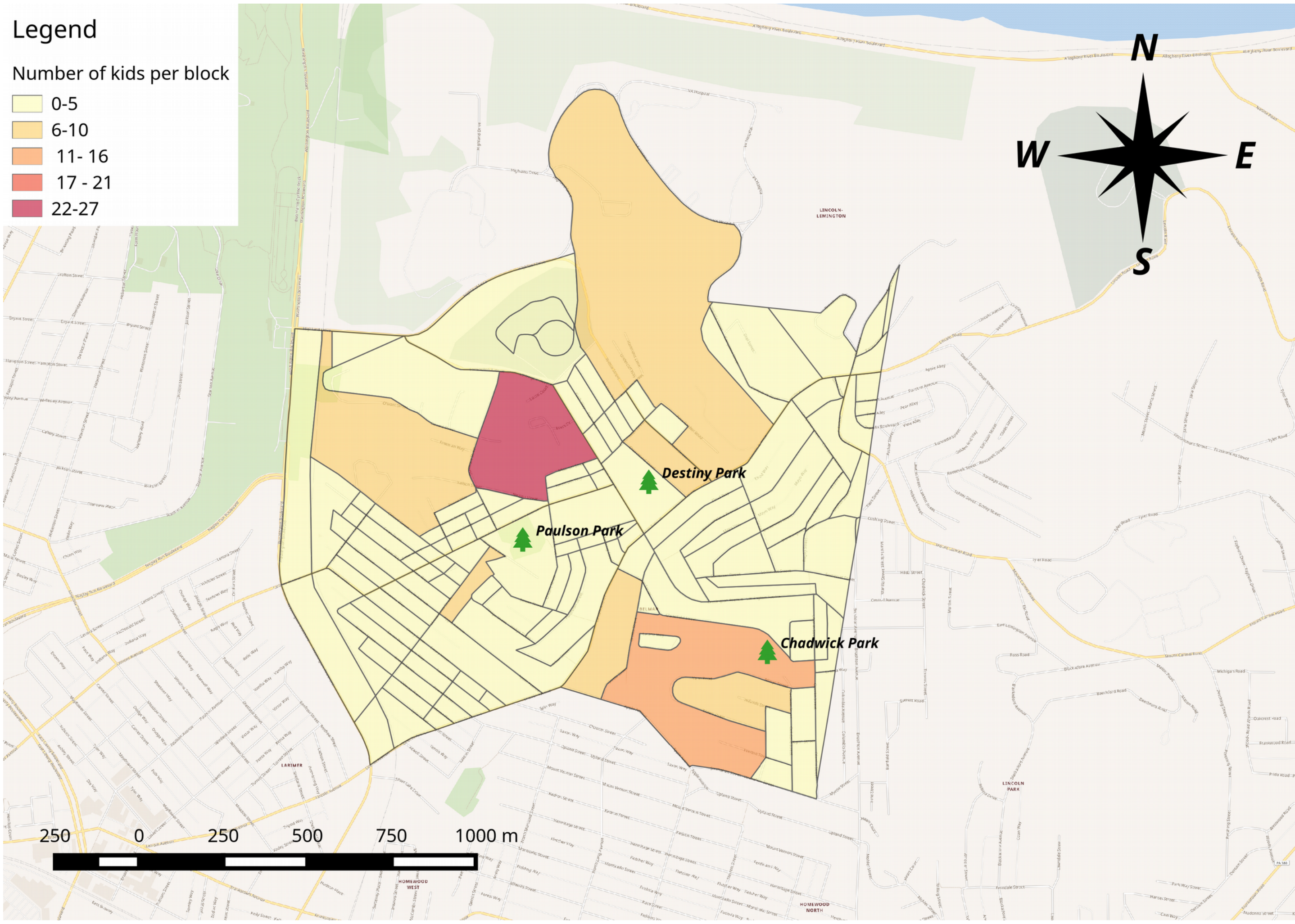


Figure 16: Final Product – Census count of children by census block with park points overlay



Discussion of final products

Each of the final product maps include a carefully tuned legend, a well-positioned map frame to include the entire study area of Lincoln-Lemington, a scale bar, and a compass rose. Note that the complexity of the maps visually does not necessarily correspond to the complexity of creating the map on the back end. For example, Figure 14 is Adeena's heat map based on the children enumeration (count) conducted at two parks in the neighborhood. While her map can be reduced to something like "a pair of bleeding red dots" the tool she used to generate that heat map is complex and requires tinkering with tool settings a number of times in succession to optimize the size of the "heat points". She learned an incredible amount about stamina and patience with computers when they aren't producing the desired visual outcome.

Figure 15 includes the final image of Willie and Charles' final map of housing stock age. Their choropleth map has very logical color breaks before and after 1978. Willie chose the color and hue for each of the color bands (year ranges) to convey the idea visually that most of the houses in Lincoln are at risk for lead paint exposure.

Figure 16 is an image of Omari's final map layout. The orange and yellow shading is based on US Census data which enumerates the number of children under 18 years old per census block. Due to time constraints and logistical complications, Omari was unable to attach the results of his playground equipment condition field data.

Project reflections

Facilitating vs. teaching vs. guiding

The structure of this project required careful alignment of my teaching style with the needs of the students, which were very fluid from moment to moment as they moved from one type of task to anoher. The core principle I employed is that student learning is maximized when each learner is challenged at a level that is not so low as to induce boredom but not too challenging to create discouragement. (The so-called zone of proximal development.) Given the students' varied exposure to academic research and technical competencies, creating an optimized challenge level required a unique approach for each student which I attempted to create with the following strategies.

1. **The students did all of the computer "driving."** No matter how complicated the task, the students maintained control over the systems and I stood next to the screen and offered directions as needed. This allows the students to absorb the work processes even if I didn't stop and explain each and every click and procedure (due to time constraints). If we had more time, I would have supported students on a practice procedure and encouraged them to try the procedure again on the "real" data by themselves. The only time I drove the computer was to experiment quickly with an unusual behavior that may have required adjusting settings or tinkering with menus that the students would not ordinarily encounter over the course of the project.
2. **I injected whole-group discussion and instruction judiciously throughout our work time.** After the first week of introductory lessons, most of the students were working more or less independently on their component of the project. This meant that there were not many instances in which what I would teach would be relevant to the entire group at that time. My approach during this work time was to conduct the mini-tutorials (Activity 7) with individual students or pairs of students and then facilitating their sharing learning with the other group members on an as-needed basis.
3. **I offered sympathetic support when the computer systems revealed bugs or frustrating aspects**—but I didn't take over control of the computers. Even if there was a crash in QGIS (which only happened two or three times), I resisted the temptation to jump in and take over the recovery process. Students need to learn that computers are not perfect machines and when possible, steps should be taken pro-actively to avoid loss in the event of an error. I would certainly be present as they checked for lost data and redid any lost work, but they were responsible for carrying out the recovery.
4. **Adeena was the official group leader, and where possible, I deferred to her judgment** when deciding what the core tasks of the group should be so as to not undermine her role as a peer leader. She exuded a natural tendency toward group leadership and this made my work as an adult facilitator relatively straightforward because I could say "Adeena—you're the leader here!" and she would confidently step into her role as a peer leader. Student leaders with less assertive personalities may need more mentoring and guidance by the adult facilitator to make the peer leadership process smooth and meaningful.

Race and the politics of community data

Background on community data in Pittsburgh

The Pittsburgh area is home to hundreds of discrete communities, some of which are administratively autonomous and others are classified as neighborhoods of a larger urban area, the largest of which is the city of Pittsburgh. Many formerly industrial neighborhoods have failed to thrive following the catastrophic exodus of steel manufacturing jobs and support industry jobs over the past forty years. Alternatively, some neighborhoods have grown new positions of strength by shifting their focus from steel manufacturing to another industry, such as technology—as is the case in the East Liberty neighborhood.

This differential in neighborhood and small town vitality has spawned an thriving civil society sector in town as foundations, nonprofits, and educational institutions implementing a wide variety of projects aimed at community betterment and revitalization. The politics of this process, are, as one would expect, often contentious and fraught with tensions within and between communities, organizations, and institutions.

One axis of contention relates directly to the content of this project: community data gathering, reporting, and ownership. Core questions which must be navigated during any community revitalization effort include:

1. Who makes decisions about what data should be collected from community residents?
2. Who should collect the data and how should it be stored and shared in its raw form?
3. What analysis tools should be used to process the data? Who knows how to use those tools? Who is taught to use the tools?
4. What is the healthiest role of “outside” groups who wish to contribute to the community revitalization efforts but who do not have any substantive ties to the communities in question?
5. How should racial differences be navigated during the data gathering, dissemination, and decision making process?

Communities versus institutions

In Pittsburgh, a concerning trend has emerged: corporate foundations fund consultancies and nonprofits which are contracted to offer technical and data-related assistance in community revitalization efforts. A community assessment process is usually undertaken in which a survey instrument is created and administered with varying degrees of community input and participation. Once the data has been collected, folks with data-related degrees and institutional affiliations (from Carnegie Mellon University or the University of Pittsburgh) are contracted to provide technical analytical services on behalf of the communities. The data crunchers, however, are almost never residents of the communities of interest, and often do not share racial or socio/economic backgrounds with those about whom the data allegedly speaks.

This skill differential is justified by citing the power of compelling data visualizations to offer insight into the workings of a given community. Technical tools are thought of as the exclusive domain of folks with technical training in statistics, computer science, and web technology. In Pittsburgh, most of the communities targeted for revitalization efforts are majority folks of color and the technically skilled folk are mostly white men who do not live in depressed areas of town.

While technical gurus can certainly carry out work to the specification of the hiring organization, they are rarely in active conversation with the community residents themselves and may not even have any interest at all in maintaining ties to the communities in question following the conclusion of the project. This dynamic effectively creates a massive gap in accountability between the community residents whose well-being is the alleged focus of a project and the individuals who have the technical skills to craft visualizations which depict the quality of life in those communities.

Value of the YES approach

This YES summer research project represents a notably different approach than the norm in Pittsburgh: the spatial analysts themselves were high school students, all African American, and some of whom live in Lincoln-Lemington. In other words, the data analysts were also community members and citizens who share critical axes of identity with the residents of the community of focus.

It is true that the program manager for the spatial analysts, your author, identifies as a white male and has occupied the role of an “outside techie” crunching data on behalf of a community home to mostly folks of color. The YES project, however, was designed and managed by YES staff who share racial an geographic proximity to the Lincoln-Lemington community. My role as a mapping facilitator meant that the strategic decisions regarding what data is collected, who collects it, and how it is presented were the domain of the project managers and YES staff—a dynamic that felt much healthier and less fraught than previous community revitalization collaborations. I supported the endeavor and offered targeted technical guidance, but the fundamental decisions were not mine, nor should they have been.

Another important note about the YES project is that the high school students learned a great deal of skills and knowledge during the summer program such that their final products were of a reputable quality and added substantive value to the ongoing efforts to reduce lead exposure on folks in Pittsburgh. The spatial analysis didn’t have a masters degree, or undergraduate degrees—they had a few Linux computers running open source software, smart phones, and a great facilitation team to help them through the research process. While the value of solid academic training should not be diminished, the threshold for robust data-based research programs may actually be much lower than those with privileged institutional affiliations would like the public to believe.

Hence, the YES project is a testament to the undue weight we as a society place on degrees and institutional affiliations. With only a few months of collaboration and training, a team of interested folks from—nearly regardless of their academic backgrounds—could join together to plan, administer, analyze a data-driven approach to community revitalization. The control over which data is gathered, how it is gathered, and how that data is used to influence policy and programs could be retained by members of the communities in question. This would likely lead to much more organic and inclusive growth strategies for Pittsburgh’s neighborhoods because:

1. **Communication between those with intimate knowledge of the community data and the residents of the communities would be smoother and more honest**. If community members have existing relationships with folks who also are working on data-based revitalization, discussions about intentions and uses of the data can exist with more trust, which should lead to better outcomes for everybody.
2. **The “cut and run” effect would be greatly diminished** since those working with the data aren’t going to leave for another job in a far-away city or become consumed with yet another project for an entirely different community. These projects often take years to run their course, and maintaining strong ties between those who have worked with the data throughout this process is invaluable to the success of the projects.

Ideas for future community mapping projects

This project was the first summer research program I’ve worked on and, as a result, I have many thoughts about how to improve the process for similar endeavors in the years to come:

1. **Extend the overall timeline of the project by a week or two.** Students should certainly have some time off in the summers, and projects like this would benefit from a little more breathing room. By the end of this project, we were somewhat crunched for time and our scheduled weekly meetings came to an end before we had time to create the final presentations together and practice them.
2. **I should spend more time in the field with the students.** My work with the spatial analysts was primarily at my shop/lab in Swissvale, PA. Since we had a field data component in the project, I should have blocked off more time to be with the students as we walked up and down streets, counted folks

going into stores, and checking conditions of the playgrounds, etc. My reason for less time in the field was mostly due to ongoing time constraints on other projects, and these should be diminished.

3. **Include a community member or two who is an adult in the group** where possible to promote the dissemination of skills from folks like myself with institutional affiliations to community members for whom these skills are best used.