Civic Innovation Fellowship Y3 Boot Camp Curriculum
September 19 - December 9, 2016

A joint project of the Manhattan Borough President's Office,
BetaNYC, CUNY Service Corps, and
Fund for the City of New York.

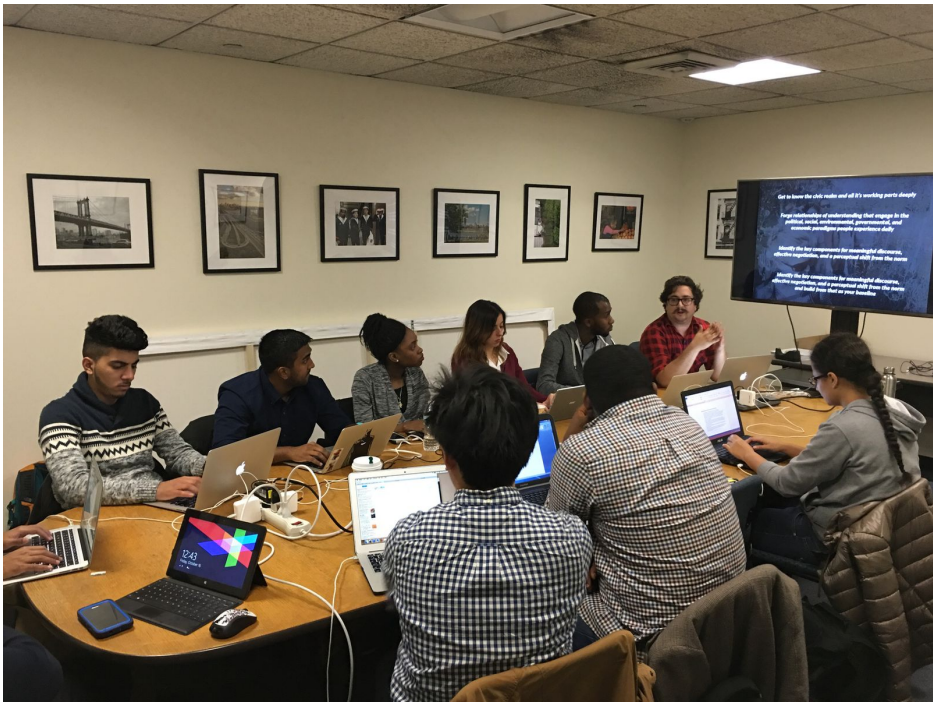
Instructors: Noel Hidalgo, Emily Goldman
Mentors: Bruna Jermann, Eli Rivera, Muhammed Aslam, Shalom Joseph

CIF Y3 Students at the Manhattan Borough President's Office, 2016.

CIF Y2 Students at Data and Society Research Institute, 2015.

CIF Y1 Students at the Manhattan Borough President's Office, 2014.

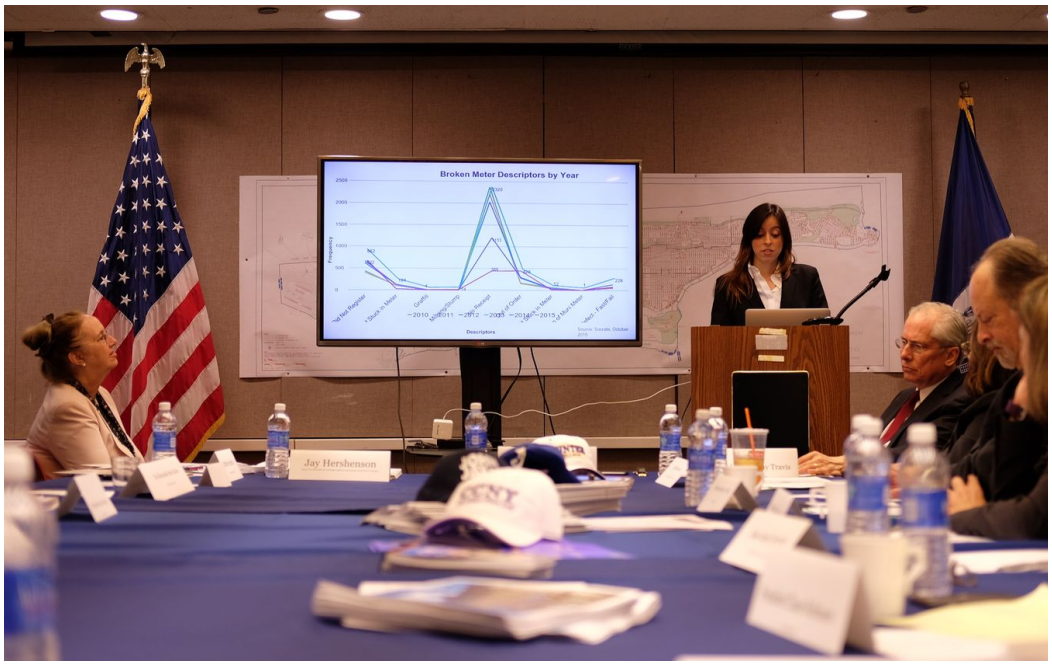
Daniel Latorre, WiseCity founder, teaching 2015 CUNY Service Corps members about media theory and service design.



Students, at the Manhattan Borough President's Office, during the boot camp phase.

Students presenting at Data and Society, 2015.

Students presenting their work at Civic Hall, 2015.



Students presenting their work at the MBPO, 2015.

Students' final presentations, 2015.

Table of Contents

Introduction to the Curriculum

The following pages of educational material, in the form of seven lab assignments, comprise the curriculum designed for Year Three of the Civic Innovation Fellowship program. Video recordings of lectures, and their slide decks, which walk students through the material and supplement the written assignments, are available through BetaNYC's YouTube channel.

We view this curriculum as a complete course on comprehending and using New York City's landscape of Open Data, with free and open source technology. The course begins with a lab that sets students up with several communication tools they need for the course, including Slack, Discourse, and Google Drive: ***Communication is Key: Introduction to the Tools.*** The following week focuses on New York City Civics, including the history of New York City, structure of its government, role of Community Boards, and the evolution of Open data: ***Historical Context of Open Data and Civic Tech in New York City***. During the next week, students explore the physical and virtual landscape of Community Boards: ***Manhattan's Community Boards: On the Earth and on the Web***. Both of these weeks include selected readings to supplement the lectures. In subsequent labs, the course becomes more technical:

In Lab 3, ***Getting to Know the Landscape of NYC Open Data,*** we teach students how to gain information and insights from six Open data portals (other than Socrata).
In Lab 4, ***Getting to Know Socrata & Using its Built-in Features,*** we teach usage of Socrata, as a platform for data analysis, visualization, and making maps.
In Lab 5, ***Analyzing Large Datasets in Spreadsheets,*** we teach how to analyze large datasets in spreadsheets, using Google Sheets as the main spreadsheet platform.
In Lab 6, ***Mapping and Analysis on the Cloud with Carto,*** we teach how to analyze and map data in Carto, a popular and versatile web-mapping software.

Building on the curriculum from last year, this year's course is meant to provide a *strong conceptual and technical foundation for learning and innovating with NYC Open data and Civic tech*. While it was designed for the CUNY fellows, it may also be useful for Community Board members and other city dwellers who have an interest in better understanding their environment and environs through the newly-available data and technology that leaders of the Open data movement have worked so hard to provide to the public.

Civic Innovation Fellowship Y3 Boot Camp
Lab #00:
Due: September 23, 2016

Communication is Key: Introduction to the Tools

This lab is designed to guide you through the tools we have established for team communication. All of these tools are free to download and use, and enable us to communicate with each other in an agile, open, and friendly way. We encourage you to use these tools often, but appropriately. This lab will help you understand their individual proper uses. Since communication is key to this program, you may need to refer back to this lab throughout the year.

These are the tools you will need to communicate with us:

1) Email
2) Slack
3) Discourse (talk.beta.nyc)
4) Google Drive

Lab Objectives:

- Your computer and devices will be set up to run all of the communication tools
- Become familiar with when, how, and why to use them
- Learn how to execute all of the functions required of you, for each tool

Let's Get Started:

Part I: Email!
Email should be reserved for communication with external individuals and private communication with the CIF leadership team. For the entirety of this program, we need you to use a gmail email account.

All external emails should cc CIF@beta.nyc this will go to Emily, Noel, and Lucian. It is important that we are copied on ALL external emails pertaining to this program. This means if you get an email response from us, you still need to cc CIF@beta.nyc. If you need to email your school about your work with us, cc us.

If you would like to contact us privately our emails are:
Noel Hidalgo - noel@beta.nyc
Emily Goldman - emily@beta.nyc
Lucian Reynolds - LReynolds@manhattanbp.nyc.gov

Lab Steps:

1. If you do not already have a gmail account, set one up now. Go to www.gmail.com , and click Create Account. Follow the remaining steps.

2. From your gmail account, send CIF an email with the subject, “Lab 0 Task 1: Hi from your name here” Leave the body of this email blank.

3. Emily is the point person for this lab. Send her a private email with the subject, “Lab 0, Task 2,” and write her a two sentence note introducing yourself.

Part II: Slack!

Slack should be used for all group communications. This is where you will engage with staff, mentors, and fellows **in real time**. While there are many BetaNYC slack channels for communication, you are expected to pay attention to #CIF-Fellows-2016

If you do not see this channel, email cif@beta.nyc with your slack user name.

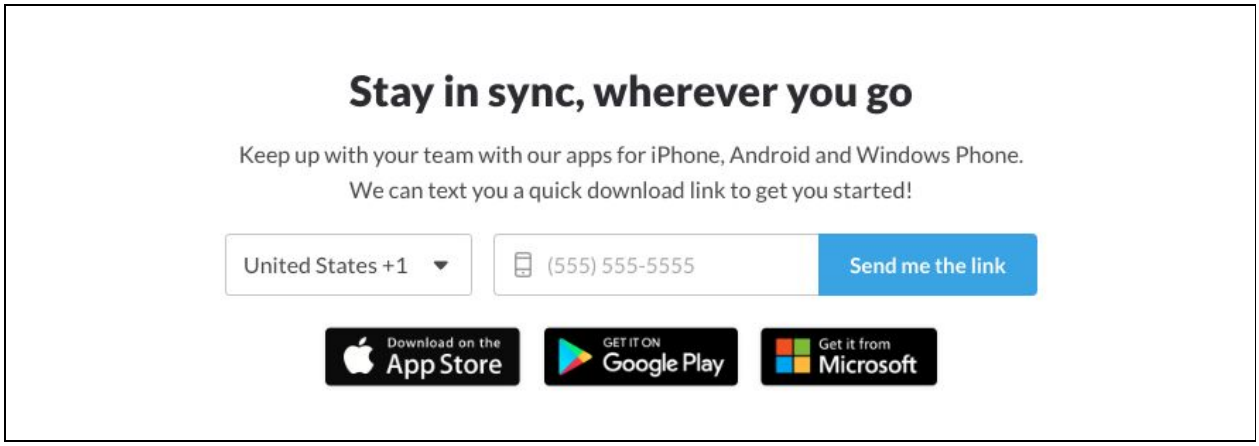
Before you are to arrive at work and while you are working, you are expected to monitor the #CIF-Fellows-2016 slack channel and be aware of the real time messages we are sharing. Group messages will be sent out with the flag @group.

If you need to communicate with the staff, you can find us as:

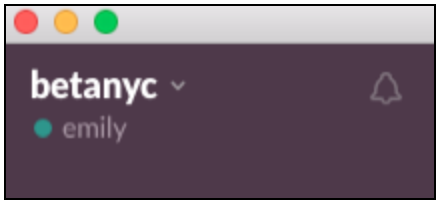
Noel Hidalgo - @noneck
Emily Goldman - @emily
Lucian Reynolds - @lucianreynolds

Lab Steps:

1. Download Slack on your computer at https://slack.com/downloads/, and request that you be sent a link to download the app on your phone. Install it on your phone too.



Log into the BetaNYC Slack team with your existing credentials: username and password. In the upper left corner of the application, you will see:



2. The side bar on the left side is kind of like a table of contents for Slack. Notice under "CHANNELS" you will see "cif-fellows-2016." All fellows, mentors, and instructors are included in this channel. If you click that channel, all existing communication will appear in the central panel of the Slack environment. The right panel provides details about the channel you are currently in. Take a few minutes to become familiar with this environment.

3. Upload a profile picture by clicking your name, and then Profile & account.

4. You can communicate with an individual in Slack in two ways. When you start typing a message, use @username to call attention to a specific member of the channel. Scroll up and check out the use of @username in the existing communication in Slack. No need to include a colon after @username. Just start typing your message.

5. The other way to communicate with an individual is to create a Direct Messages connection. Click the + sign next to Direct Messages, and add someone. Add emily, and send her a message that says, "Lab 0 Task 3--direct message in Slack." Since it's a direct message, you don't have to use @emily.

6. You can use "@channel" if you want to alert all members of the channel of your message immediately. Otherwise, if sending a message to the whole group, you don't need to use @channel.

7. You can react to a message with **emoticons**, and use them in or as your message. Hover over an existing message to see the add emoticon icon. Hover over the emoticon icon on the right side of the message box to add one to your message. React to the direct message you just sent to Emily, and send her another message that contains only an emoticon.

8. Conduct steps 8 - 10 in Direct Messages with yourself. Copy and paste a URL in the message box and notice what happens when you hit enter.

9. Copy and paste a link in the message box to share a link with the team. We will do this often, to share links to Google Docs, so be familiar with this process.

10. You can also upload a file. Click the plus sign on the left side of the message bar, and upload an existing document from your computer. Hit enter.

11. You may not be able to delete a message you post on Slack, so use discretion, and follow the general guidelines and principles for communication, outlined at the beginning of this lab.

12. In the sidebar, you will also see a category, “STARRED.” To star a channel, which will move it up to the top, make sure you are in the channel, and hover over the channel’s name at the top. At outlined star will appear next to it. Click the star to activate it, and notice the channel moves up in the sidebar. That just helps you prioritize the channel.

13. In the panel on the right side, take a look at the Notification Preferences. The default settings are that you will receive a notification on your desktop and your phone when you have been specifically mentioned in a Slack communication (by @your username). If you click Edit Notification Preferences at the bottom, and then click Account Preferences at the bottom of the dialogue box that pops up, you can refine your settings. Explore this as you wish. We just suggest that the default settings for desktop and mobile notifications are retained as ON.

Part III: Discourse!

Discourse is open source software released in 2013 to serve as *an improved Internet discussion forum*, which includes many innovative features, like infinite scrolling and live updates (no need to refresh the page), etc. Discourse powers **talk.beta.nyc (**a.k.a. **“Talk”),** which is our online discussion platform. You will learn a lot by using Talk throughout the year, and can get to know its many features over time.

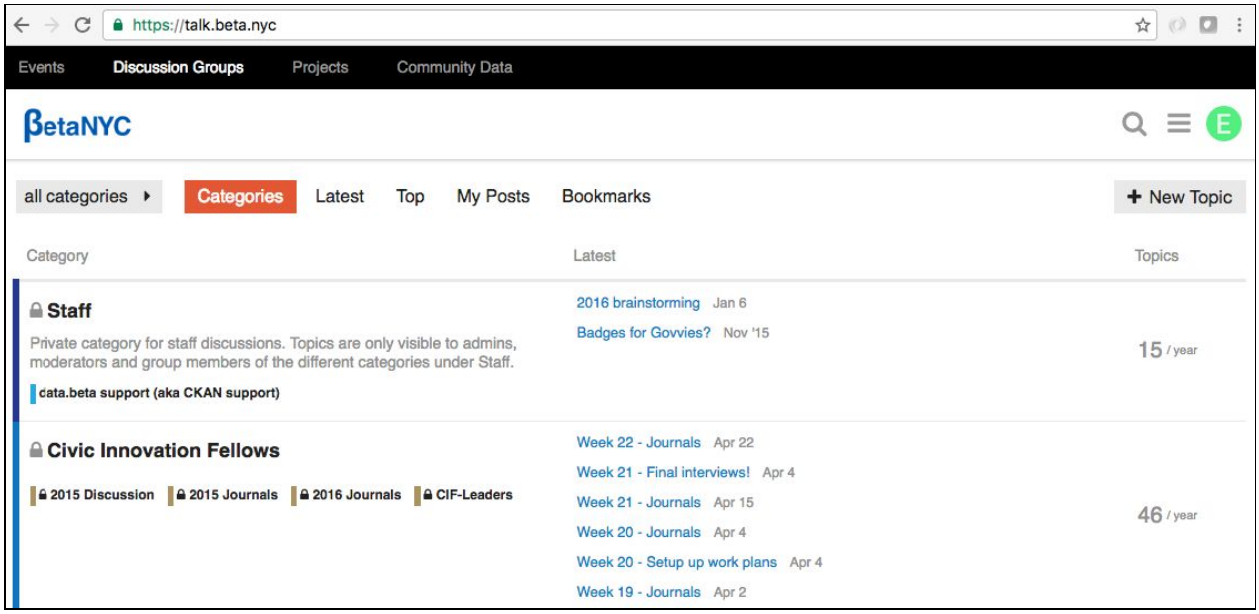
It is also where you will find a copy of the week's objectives and where you are expected to journal your weekly activities.

To get paid, we need you to journal your weekly activities in Talk, in a thread we have set up specifically for this.

Lab Steps:

1. Log into talk.beta.nyc with your existing credentials: username and password. Are you in? If you have any issues logging in, call us over and we may need you to email cif@beta.nyc with your username.

2. In the upper right corner, you will see the first initial of your first name in a green circle. Click it and select the little wheel below it to open preferences. That's where you can upload a good profile picture and fill out other biographical details. Do this!



3. In Discussion Groups, notice the Civic Innovation Fellows group, and locate our private thread for journaling ("2016 Journals"). If you click it, you will see it is currently empty, but that will change soon….

4. Talk has many features. Noel has detailed instructions that YOU WILL RECEIVE AS YOUR FIRST WELCOME EMAIL from Talk. Please read these guidelines, focusing for now on how you will **write a journal entry**, **comment on a thread**, and **like a comment.**

5. About Markdown: as you progress through the BootCamp and Fellowship program, you will learn how to write in Markdown, a simple computer language. See this blog about Markdown,

written by one of its creators back in 2004. For now, watch the ten-minute tutorial on Markdown referenced in the guidelines page (http://commonmark.org/help/tutorial/). Don't worry too much about it though, especially if you do not have prior experience. We will walk you through writing your first journal entry as a group.

Part IV: Drive!

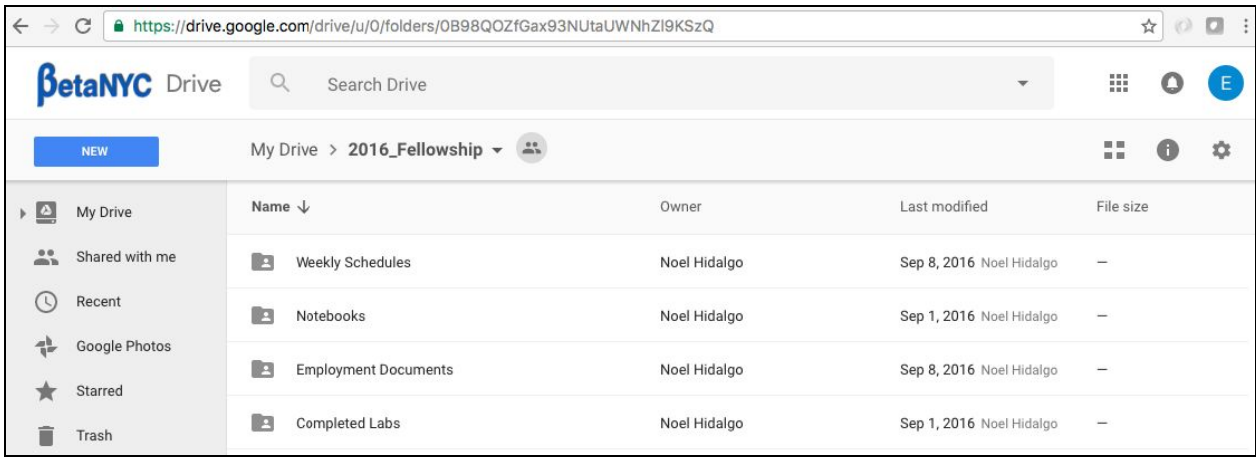
In lieu of learning management software, such as Blackboard, we are repurposing Google Drive.

For security purposes, you must use a google account and be authorized There is one centralized folder for all of Y3 CIF folders. This is the 2016_Fellowship folder. If you do not see this folder, email cif@beta.nyc with your gmail account.

1. Weekly Schedules and Lab Outlines - 2016_Fellowship > Weekly Schedules
    a. Here you will find your weekly work schedule and lab that should be completed by the end of the day.
2. CIF Notebooks - 2016_Fellowship > Notebooks > "your name"
    a. Here you will find your name. We are expecting that when you share working documents, you shared them from this folder.
3. Completed Labs - 2016_Fellowship > Week "##"
    a. When you have completed your weekly lab, you will find the corresponding week and submit your lab with the following file name "CIF16_NAME_Week##", with the ## being the corresponding week.

Lab Steps:

1. Click the blue NEW button on the left side and notice that you can create a new folder, and you can upload an existing file or folder.



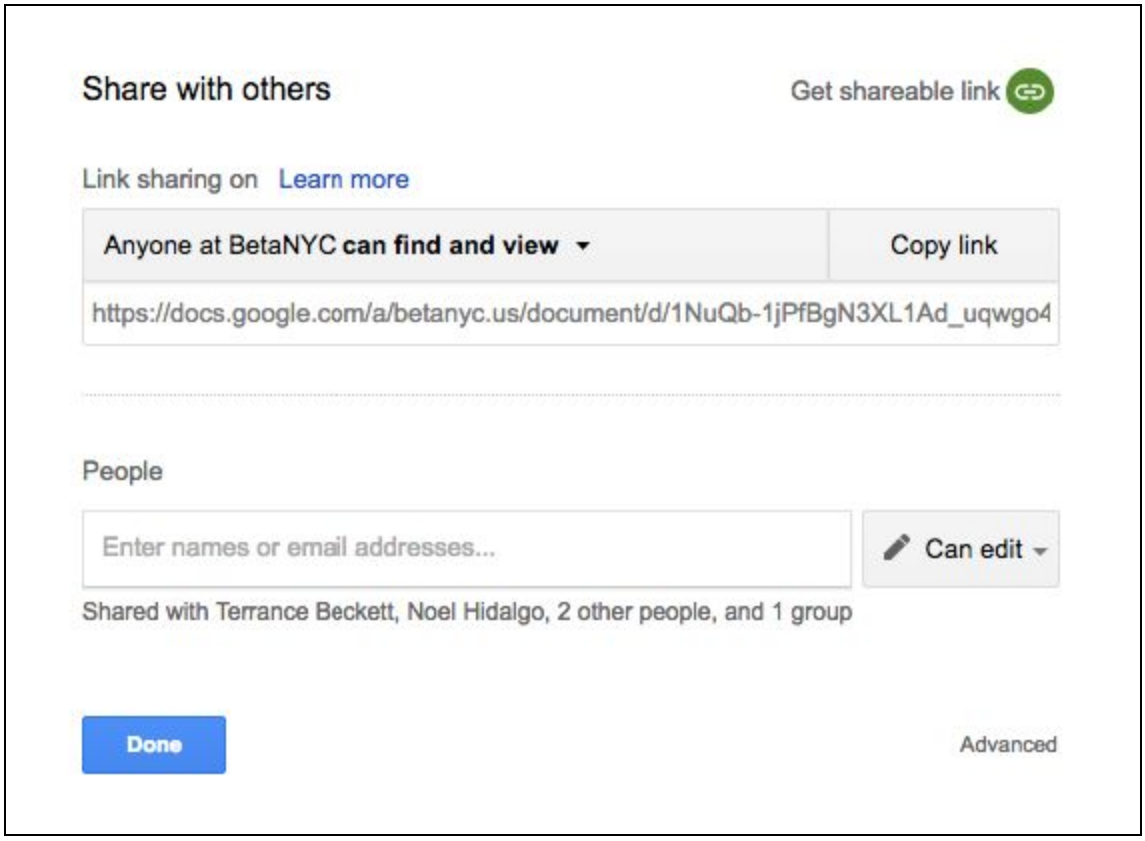
2. Open the Notebooks folder, and notice that each of you already has a folder for your work. You may want to organize your folder with separate folders within it. Play with creating folders within your own folder. You can always delete them later.

3. Back in Drive, when you click the Search bar, it provides a list of file types. If you select a file type, you can then search for a specific file within that type by clicking the down arrow on the right side of the bar, and typing in the file name, or part of the name. Pretty handy search utility, if you ask me!

4. Use the File Upload feature to upload an existing Word doc, Excel spreadsheet, and PowerPoint file from your computer into your BetaNYC Drive. After the uploads are complete, you will notice them in your list of files and folders. Click each file and then click “Open With” to convert the file to Google Docs, Google Sheets, or Google Slides respectively. What happens in your list of files and folders? Notice how Drive retains the original file type in the name of the converted version. You can change hat by clicking Rename.

5. There are a few basic tasks you need to be comfortable with when working in **Google Docs.** Steps 6 - 9 will take you through those tasks: **sharing a file and managing file permissions** (a.k.a “Access Control List” or ACL); **collaborative document editing**; **commenting within a document**; and **inserting hyperlinks**. With Lab 0 open in Docs, click File, Make a Copy, put your name in the file name, and navigate to and save it in your own folder. Use your Copy of this document to execute the following steps.

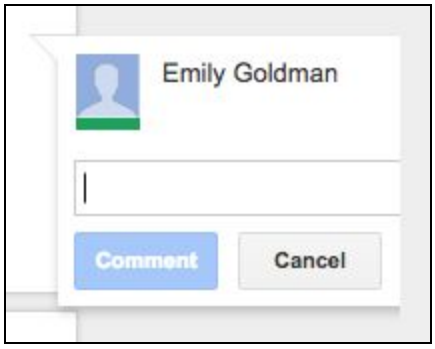
6. To share a Google Doc with others, click the blue Share button in the upper right. Your sharing options will appear as in the box below.



Please take a few minutes to toggle among the different options you have for sharing.

7. If you want to share the document with specific individuals for collaborative editing, you can enter their names in the box under People, and click Done. They will be sent an email with the document that they can then edit. If you wish to Slack them the link too, paste it into a Slack message. Then when they open it, they will be able to edit it. When multiple people are working on a document at the same time, you will see their personal icon on the upper right of the document, next to Comments. See the image below for a document in Emily's Drive that Noel is collaborating on. See Noel?!

8. Another way to collaborate on a document is to insert comments. Click Comments next to Share, and you can type comments in the margins, labeled with your name.



When you hit comment, it will be saved for your collaborators to see. If you wish to notify collaborators of a comment you have completed, click the three dots next to Resolve, and send them the link. When they open the link, the material commented on will be highlighted.

9. Inserting hyperlinks in a Google Doc is pretty easy. With the cursor where you want the link, click Command K, or find and click the link icon on the Toolbar, or as an option in the Insert drop down menu. Paste the link in the box that appears, and pick the text to display in its place. When you hit apply, the text replacing the link will appear underlined in blue.

Deliverables:

Unlike the other labs, apart from the emails and messages we have asked you to send along the way (which will be checked), there are no deliverables for Lab 0. However, we will be checking up on you and your computers shortly to make sure that you have set things up properly. And you shouldn't have to ask us the basic questions covered in the lab.

Concluding Remarks: Congratulations! You have just become acquainted with some of the most up-to-date communication mechanisms out there. COMMUNICATION IS KEY! Communication and Community share 7 letters. Why do you think this is? Include your thoughts about the relationship between these two words in your journal entry for Week 1.

Civic Innovation Fellowship Y3 Boot Camp
Lab #01:
Due: September 23, 2016

YouTube video of lecture: http://bit.ly/2jlIFAN
Link to the slides: http://bit.ly/2iV3a7j

Manhattan's Community Boards: on the Earth and on the Web

Continuing the work from last year, in Lab 1 you will build on and update an existing survey in spreadsheet form (link). This important survey documents thirty-six elements of the web presence and activity of Manhattan's twelve Community Boards. You have each have been assigned a CB. (See Column N here). The introductory lecture given by Noel and Emily has provided the necessary background information, and we hope the inspiration, for the project. Those slides are available on the Drive.

This Lab will consist of the following:

- Becoming familiar with your assigned Community Board--on the streets, and on the web
- Making sure you understand all the Field Headings in the survey spreadsheet
- Learning how to retrieve the information requested of you
- Updating the entire sheet for your CB to the best of your ability

Lab Objectives:

- Develop familiarity with your assigned Community Board
- Learn how to answer all the questions related to its web presence and activity and fill out the survey to the best of you ability

Tools required:

- All Communication Tools
- GitHub (that's where you found this assignment!)
- Google Docs (where the survey lives)

Before you start:

Make sure you are set with the Communication Tools and know how to retrieve the lab assignment and lecture in GitHub/Google Docs.

Let's get started:

Part I: Navigating your assigned CB--on the streets and on the web

About the Community District

1. What are its boundaries?
2. What are its neighborhoods?
3. What are its icons/landmarks (in your opinion)?
4. Who are the officials in the Community Board?
5. Where is the office located?
6. What are its hours?
7. Get a screenshot of it via "street view"?

About the Community Board Website (these questions relate directly to the survey)

1. What is the Community Board Website URL?
2. Is there a Community Board about page?
3. Do they have a District profile?
4. Do they list Community Board members and staff?
5. Can you easily find the address of district office? If so, where is it located? Yes,
6. Do they post the hours of district office?
7. Is there a contact page?
8. If so, what's on the contact page? How many clicks did it take you to get to the contact page?

News or updates section

1. Listing of News / Updates?
2. Is the news / updates current? (i.e. is there an update within the last 30 days?)
3. Summarize what type of content do they share.

Calendars

1. Calendar of events?
2. Do events have meeting agendas?
3. Meeting Mins / Notes?
4. Is there a listing of Committees?
5. Is there a listing of Task Forces?
6. Is there a listing of Committee Members?
7. What information can you find on the Committee pages?
8. Can you find a collection of archives / resolutions?
9. Can you find a listing of Community Board projects?
10. Do they have a community resources guide on their website?
11. Summarize what type of events they share.

Videos

1. Is there a video archive of meetings?
2. Can you tell if they "live stream" meetings?
3. Do they share their videos via YouTube or Vimeo? If so, what is the URL? Do they allow comments?
4. Summarize what type of videos do they share.

Search

1. Is there a search page or bar?
2. If so, does it work?
3. Summarize what type of content searched for and how did it work.

Maps

1. On their website, do they share maps?
2. If so, what type of maps?
3. How old are the maps?
4. Summarize what types of maps they have on their website.

NYC 311 and Data

1. Does the Community Board website promote NYC 311?

2. Do they promote any other data tools?
3. (If so, which ones?)
4. Summarize what types of data they post on their website and how do they share this.

Newsletters and Email Lists

1. Does the Community Board offer an Email newsletter sign up?
2. Does the Community Board website promote any email discussion lists like google groups or yahoo groups?

Facebook

1. What is the URL of the facebook account?
2. Type of page: Page, Profile, Group?
3. Is the “about” community board filled in?
4. Is there a rating? What is the rating?
5. Address posted?
6. Phone number?
7. Contact information?
8. Office Hours?
9. Accepts comments?
10. Engagement w/ constituents/comments?
11. Posts Meeting Notices?
12. Posts Meeting Video?
13. Posts Community Events?
14. Frequency of posts?

Twitter

1. Twitter URL: What is the URL of the twitter account.
2. Tweets: How many tweets have they tweeted?
3. Following: How many people does this account follow?
4. Followers: How many people follow this account?
5. Likes: How many likes does this account have?
6. Lists: How many lists is this account on?
7. Complete profile: Does the profile have a photo, header image, about statement, address, and URL?
8. Hashtags: Does the twitter account use hashtags? If so, what do you see?
9. Engagement w/ constituents: Does this person tweet at others or responds to other twitter accounts.
10. Tweet Meeting Notices: Is this account used to announce meeting notices.
11. Tweet Video: Is this account used to post / share videos of CB meetings?
12. Retweets: Does this account retweet? If so, what type of retweets?

13. Frequency of posts: How frequent are the posts? Hourly? Daily? Weekly? Bi-monthly? Monthly? Quarterly? If it is sporadic and a mixture of above, please tell us what.
14. What else does the CB post?

Part II: Filling out the survey

Work with the mentors to fill out the survey to the best of your ability!

Deliverables:

1. **All of the above numbered questions answered in clear and concise written form in a Google Doc (your journal entry should provide a link).**
2. **The survey updated to the best of your ability.**
3. **The following miscellaneous questions answered as well, in the same Google Doc.**

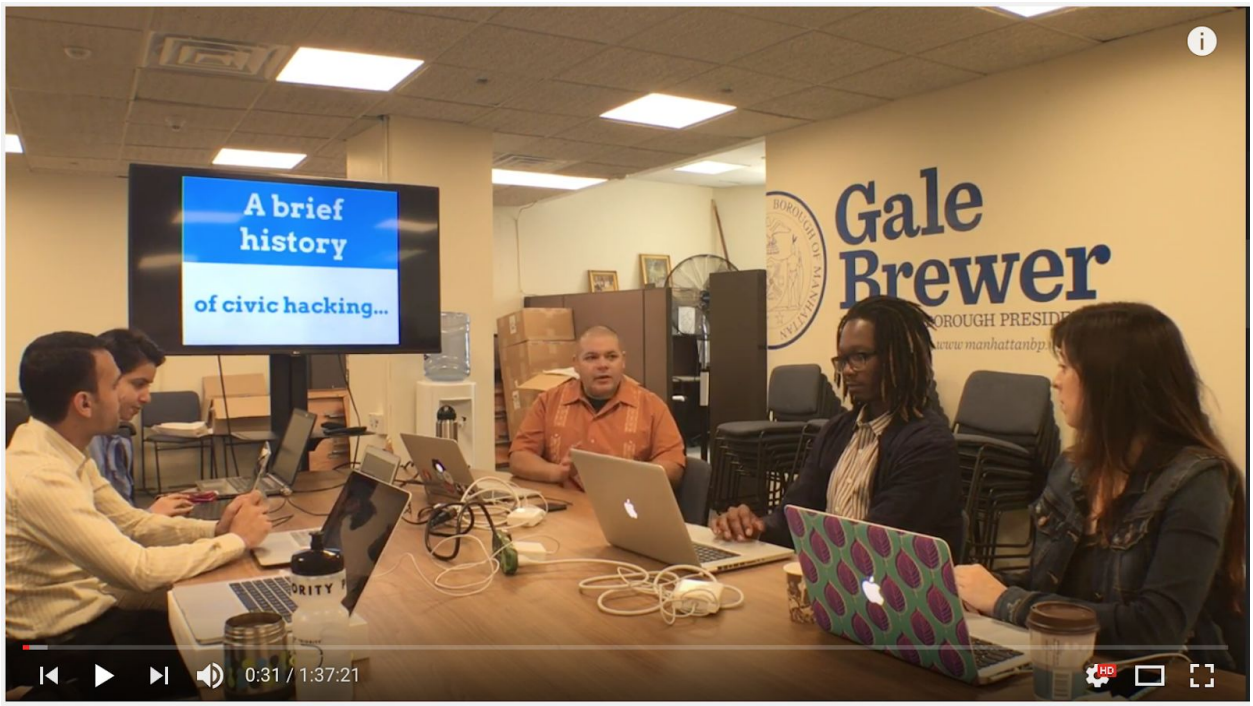
Miscellaneous:

1. Which websites did you use to get the lay of the land of your CB?
2. On a scale of 1-10 (1 being the weakest, 10 being the strongest), how strong do you *think* your CB's digital presence is?
3. Which question(s)/field headings were you *least* familiar with?
4. Which question(s)/columns did you have the *hardest* time answering?
5. Which CB do you live in? Tell us the steps you took to find out.

Concluding Remarks:

Alright! You are now familiar with many aspects of one of Manhattan's twelve Community Board and know how to navigate its digital presence and activity. This lab provided the foundation for the following labs and for your entire Civic Innovation Fellowship Year with us. Congrats!

Civic Innovation Fellowship Y3 Boot Camp
Lab #2:
Due: September 30, 2016



Youtube video of lecture: http://bit.ly/2iaPCWB
Link to the slides: http://bit.ly/2i99ia7

The Historical Context of Open Data and Civic Tech in New York City

This week, we are asking you to develop a strong understanding of the broader context of the Civic Innovation Fellowship program-including key developments in New York City history, the Open data and Civic technology movement, and how BetaNYC, the Manhattan Borough President's Office, and CUNY Service Corps all fit into the whole picture. Noel and Emily both firmly believe that this conceptual foundation is as important as any tool or method you will learn during BootCamp. After you have finished the lab, our hope is that you do too. Slides from lecture are available here, and readings to supplement the lecture are listed below.

This Lab will consist of the following:

- Learning about New York City history, the importance of Progressivism here over time, and the structure of New York City government
- An Introduction to Data; Open Data; NYC Open Data; 311 Data

- An Introduction to Civic Technology; Civic Tech in NYC Government; BetaNYC and other groups alongside government
- Intro to Socrata & Filing (your first?!) 311 Service Request

Lab Objectives:

- Develop familiarity with all the concepts listed above
- Initial exploration of Socrata and experience the process of filing a 311 Service Request
- Learn how to “see data” all around you in the city

Tools required:

- All Communication Tools
- Google Docs (where your submission lives)

Before you start:

Have the slides on hand. The following readings will help you answer the following questions, so have them handy too.

Readings:

Mayor's Office of Community Affairs, About Community Boards
Wikipedia article Government of New York City
New York’s Birth Date: Don’t Go by City’s Seal by Sam Roberts
They Took Manhattan by Kevin Baker

What is a hacker, by Ben Balter
Defining Civic Hacking
The Civic Hacker Hacked, by Mark Headd
Civic Hacking, by Joshua Tauberer

Other books:

Cathedral and the Bizarre - http://www.catb.org/esr/writings/cathedral-bazaar/
The Wealth of Networks - http://www.benkler.org/Benkler_Wealth_Of_Networks.pdf
Starfish and the Spider - http://media.portland.indymedia.org/media/2008/10/380532.pdf
The Boy Who Could Change the World: the Writings of Aaron Swartz - https://monoskop.org/media/text/swartz/
Coding Freedom: The Ethics and Aesthetics of Hacking - http://gabriellacoleman.org/Coleman-Coding-Freedom.pdf

Let’s get started:

Part I: Using the lecture, the readings listed, and your own research, answer the following questions:

1. What does Manhattan mean?
2. Who were Manhattan's first inhabitants?
3. Who were the first European explorers to "document" New York Harbor? -- First -- Second -- Third
4. Who was the first non-Native American inhabitant and where did they come from?
5. When did the first permanent European presence begin and which Nation supported it?
6. In what year did the British rename the city to be New York?
7. Where was the first President of the United States of America sworn in?
8. When was slavery abolished in New York?
9. The Commissioners' Plan of 1811 did what?
10. The modern City of New York was consolidated in what year?
11. What happened at the Seneca Falls Convention? What happened at the Triangle Shirtwaist Factory fire and what important laws grew from this event?
12. What was the main goal of the progressive movement and how has that shaped NYC's government?
13. What is New York City's Charter and when was it last reviewed?
14. What is the City Record and what is its role in the City?
15. What are the five languages of New York City Government?
16. How is NYC government organized?
17. What is fusion balloting?
18. What are New York City's citywide offices and responsibilities?
19. What is the role of New York City's City Council and how many members does it have?
20. What was the Board of Estimates and what did it do?
21. What are Borough Presidents and what are their responsibilities?
22. What is a Community Board and what are their City Charter granted responsibilities?
23. What is a Community District?
24. How do you become a Community Board member?
25. How are Community Boards organized?
26. What is a Borough Board?
27. What is the Borough Service cabinet meeting?
28. What is City agency?
29. What is NYC 311?
30. Who is a constituent?
31. What does ULURP stand for and what is it?
32. What is Open Data and how does it relate to Government?
33. What is Civic Tech?
34. What is Free and Open Source Software (FOSS)?
35. What is Socrata?

Part II: Seeing data and Filing a 311 Service Request (or more than one)

1. Go to the Socrata website here.
2. Type 311 in the Search bar.
3. Open a 311 dataset.
4. When the dataset loads, check out the various **Complaint Types**.
5. Find your own service complaint to file, and do it! Call 311 and report it!

Deliverables:

1. **All of the questions from Part I answered in short written form in a Google Doc (your journal entry should provide a link).**
2. **Answer the following questions about your experience filing a 311 Service Request:**
    a. **Which type of complaint did you request service for?**
    b. **Location of the problem that you called about?**
    c. **Had you ever called 311 before?**
    d. **How did you find the experience of filing a 311 request?**
3. **The following miscellaneous question answered as well, in the same Google Doc.**

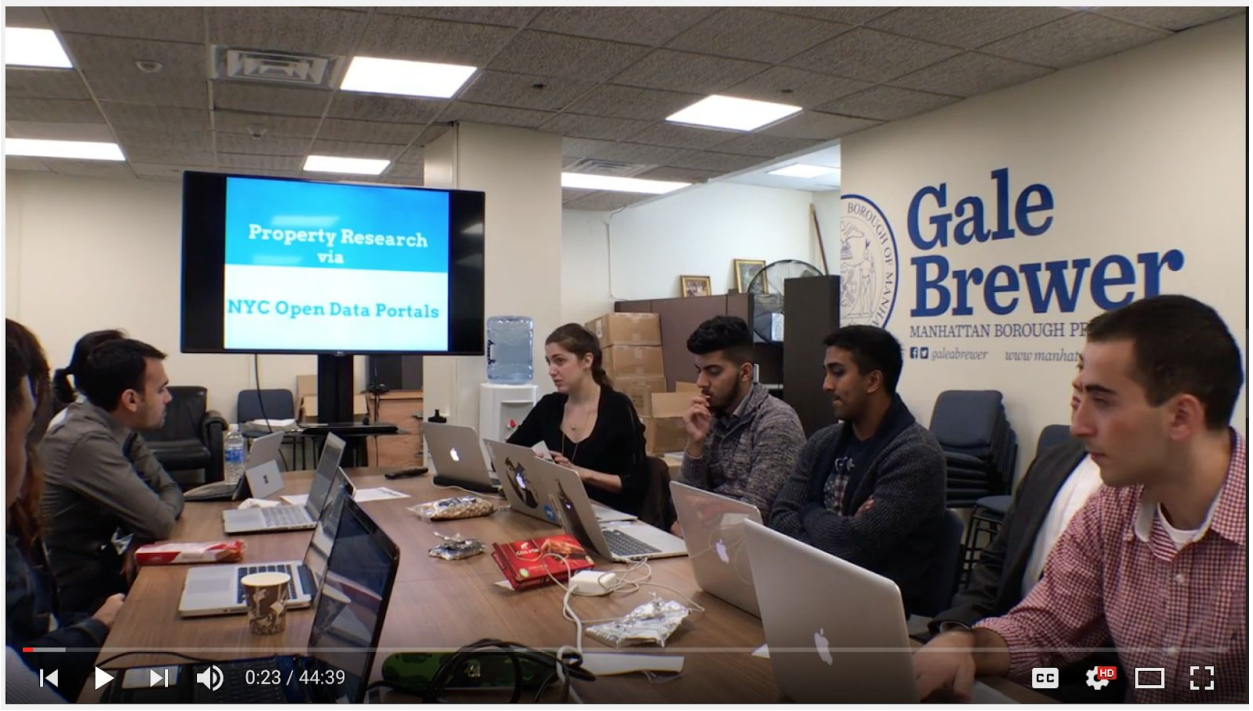
Miscellaneous:

1. Tell us your impression of what data means, and how this lab has affected the way you will "see data" from now on…..

Concluding Remarks:

Woohoo! You now have a true understanding of how Civic Tech and your work this year fit in the MUCH larger picture. Many people are developing tech skills, but few have an understanding of the broader context, which is integral to carrying the work forward in the most useful and innovative ways for society. This lab provided that conceptual foundation for your entire Civic Innovation Fellowship Year. Let it infuse your thinking all year. Congrats!

Civic Innovation Fellowship Y3 Boot Camp
Lab #03:
Due: October 17, 2016



Youtube video of lecture: http://bit.ly/2juRbAC
Link to the slides: http://bit.ly/2i6jPHU

Pick a Property: Getting to Know the Landscape of NYC Open Data

In this lab, you will get to know several portals of information about real property in NYC, learning which resource to use when and for what purposes. You will discover that each resource has some information unique to it.

You will also come to understand that there are different ways to identify property in NYC--e.g. address, BBL (Borough/Block/Lot), BIN, lat/long. These are called **geographic identifiers** and they comprise a key concept in understanding how to conduct effective research about property in NYC.

Have you ever heard of GOAT? Zola? BIS? ACRIS? Well, you will soon :)

As always, make a copy of the lab, place it in your folder in Notebooks. After you have reviewed your completed work with a mentor, you will then place it in the Completed Labs folder in the Drive for Noel and Emily to peruse
. Have fun.

Lab Objectives:

- Learn how to uniquely identify buildings/parcels/apartments in NYC
- Develop a grasp on multiple portals of NYC Open Data
- Successfully conduct extensive research on a building/parcel/apartment of your choice

Let's Get Started:

Part I: Location identifiers and GOAT

If you were to tell a friend where to meet you one day, how would you do this? Are you likely to say the address, to describe the location in terms of intersections and streets, or to reference the latitude and longitude (i.e. "Meet me at 40.730062 , -73.985651"). The truth is that any of these, and many others, may accurately identify your location. Sometimes one makes more sense to use than another, though.

First, pick an address in the five boroughs to explore for this lab. Keep it the same throughout the whole lab, so make a careful choice. It could be your home address, the address of an important friend or family member, an institution you frequent--coffee shop, favorite restaurant or movie theater, etc. or one you are just curious about. You just need to know its street address.

ADDRESS FOR LAB #3 CHOSEN IS:

Go to NYC GOAT and plug in the borough and the address.

Scrutinize the information that appears on the next screen.

Take a screenshot(s) of all the information you get from GOAT. Paste it below.

Report the following:

Property Level

- **Tax Block:**
- **Tax Lot:**
- **BBL:**
- **X Coordinate:**
- **Y Coordinate:**
- **Latitude:**
- **Longitude:**
- **BIN:**

Political Districts:

- **Community Board**
- **City Council:**
- **Assembly District:**
- **Congressional District:**
- **State Senate:**
- **District Attorney:**

Government Services:

- **Police Patrol Borough:**
- **Police Precinct:**
- **Fire Division:**
- **Fire Battalion:**
- **Fire Company:**
- **School District:**
- **DSNY Snow Priority:**

Part II: NYCityMap

What other information are you craving to know about this property? How about a map of it, including the building footprint? This would enhance your understanding of the property in terms of its spatial dimensions.

Go to NYCityMap and type your address in the Search bar.

Question 2: When you typed in the address in the Search bar NYCityMap, how was this different than when you typed it in, in GOAT?

Take a screenshot of NYCMap with your parcel outlined in red. Paste it below.

Question 3: Anything surprising or interesting to note about the shape of your parcel?

Notice the tabs for Building & Property Information, Neighborhood Information, and Elected Official Information on the right-hand side.

Within Neighborhood Information you will see many links listed.

Question 4: Click on several of the links and note the functionality of two of them.

- **Link 1:**

Does it work?

- **Link 2:**

Does it work?

Part III: BIS

Wait, there's still more to learn about my property? Yes. BIS is the Department of Building's Building Information System. Let's find out what this portal has to offer.

On that note, which City agency hosts NYC GOAT and NYCityMap? Ah, those are hosted by the Department of City Planning.

Go to BIS WEB and search for your property.

Notice the multiple ways one can search for a particular property in BIS. Think about how this relates to the concept of geographic identifiers discussed earlier.

Take a screenshot of the information provided by BIS. Paste it here.

Question 5: Some of the information in BIS may not immediately be familiar to you. What are three fields you don't know or would like to learn more about?

- **Field 1:**
- **Field 2:**
- **Field 3:**

Notice toward the bottom there are links for Complaints, Violations-DOB, and Permits In-Process/Issued, and many others.

Go to each of the following links, peruse what you see, and report the number of entries there are for each of the following:

- **Complaints:**
- **Violations-DOB:**
- **Jobs/Filings:**
- **ARA / LAA Jobs:**
- **Actions:**
- **Elevator Records:**

****If there are no complaints or violations against your property, look up 227 West 109th Street, Manhattan instead, and report your findings about that building. Just make a note that you have used this building so we know. CIF discovered last year that it was particularly problematic. Feel free to check out this building too, even if you are satisfied with the results from yours.**

Part IV: Google Maps (Map View)

What additional information will Google Maps provide? Go to Google Maps. Notice where it takes you. Search for your address, and you will fly over to it.

I was able to see the street view of the building. Also, I was able to compare the picture to 2007 and and 2011.

Question 6: What other information does Google Maps provide about your property's neighborhood?

Part V: Google Maps (Street View)

When you click the image of your building on the left, you go to Street View. **Take a screenshot of your building and paste it below. Please note THE DATE it was taken.**

Question 7: Report what ELSE you learn about your building using Street View. What is your perception of the *building's condition* by looking at it in Street View? ALSO--new addition after lecture--how many historical images are available in Street View?

**For fun, check out an art project made with Street View here, called Nine Eyes.

Part VI: ACRIS

ACRIS stands for Automated City Register Information System. Go to ACRIS. Which City agency hosts ACRIS? You'll notice at the top of the website, it is hosted by the Department of Finance.

Go to "Find Addresses and Parcels." Populate the Property Address fields and click "Find BBL."

Then, on the right side, click "Document Search by BBL."

On the next screen, click Search.

Question 8: How many transactions are reported in ACRIS for your property?

Question 9: Can you find the latest DEED under Document Type? If so, what is its date?

Part VII: NYC Census FactFinder

Go to the site. Toggle between Neighborhood Tabulation Area and Census Tract on the left-hand side.

Notice at the top of the map, there are three options: Map, 2010 Census Profile, and 2009-2013 ACS Profile. ACS stands for American Community Survey, a survey distributed to about 3.5 million households yearly. The ACS is a 21st century initiative, launched after the 2000 Decennial Census, in order to capture information about the American people more frequently than every ten years.

Question 9:

- **Name of your property's neighborhood:**
- **Number of its census tract:**

Now, make sure you are in the Neighborhood Tabulation Area view.

Toggle between between the 2010 Census Profile and the ACS Profile.

Question 10: What are the categories of data available through the 2010 Census Profile, and what are the categories available through the ACS? (These are the tabs that appear at the top of the tables). Spend some time noticing some of the characteristics of your property's neighborhood available through each.

Deliverables:

Make sure you have answered the ten (10) questions in this lab, and taken all the screenshots we asked of you.

Concluding Remarks: Congratulations! You have just become acquainted with some of the most important information portals in the landscape of Open Data in NYC! You've learned a lot about a specific property of your choice, and now you can do the same for others.

Civic Innovation Fellowship Y3 Boot Camp
Lab # 04: Learning Socrata / NYC's Open Data portal
Due: October 21, 2016

Youtube video of lecture: http://bit.ly/2iZXJX0
Link to the slides: http://bit.ly/2j0ewtb

Getting to Know Socrata & Using its Built-in Features:

About Socrata:

Socrata, a company founded in 2007, provides cloud-based data analysis and visualization platforms for large datasets; moreover, it is specifically oriented toward *opening government data.* Today, over 1,000 government organizations in the U.S. use Socrata to power and share their Open data, at the Federal, State, County, and City government level--including NYC.

Last week, you explored several other portals of NYC Open data. This week you will get to know how Socrata differs from these, in its analysis and visualization capabilities, and you will gain experience using all of its built-in features. You will focus on a dataset we have already discussed at length: 311. Perhaps this will be your first experience analyzing "big data." Have fun.

Lab Objectives:

- **Create an account**
- **Become familiar with Socrata's environment--quirks and all**
- **Learn how to use the following built-in data analysis and visualization tools:**
    - **Filter**
    - **Sort**
    - **Roll up**
    - **Point map**
    - **Heat map**
- **Execute several complex analyses on 311 data using the techniques you have just learned**

Part I: Using the Portal

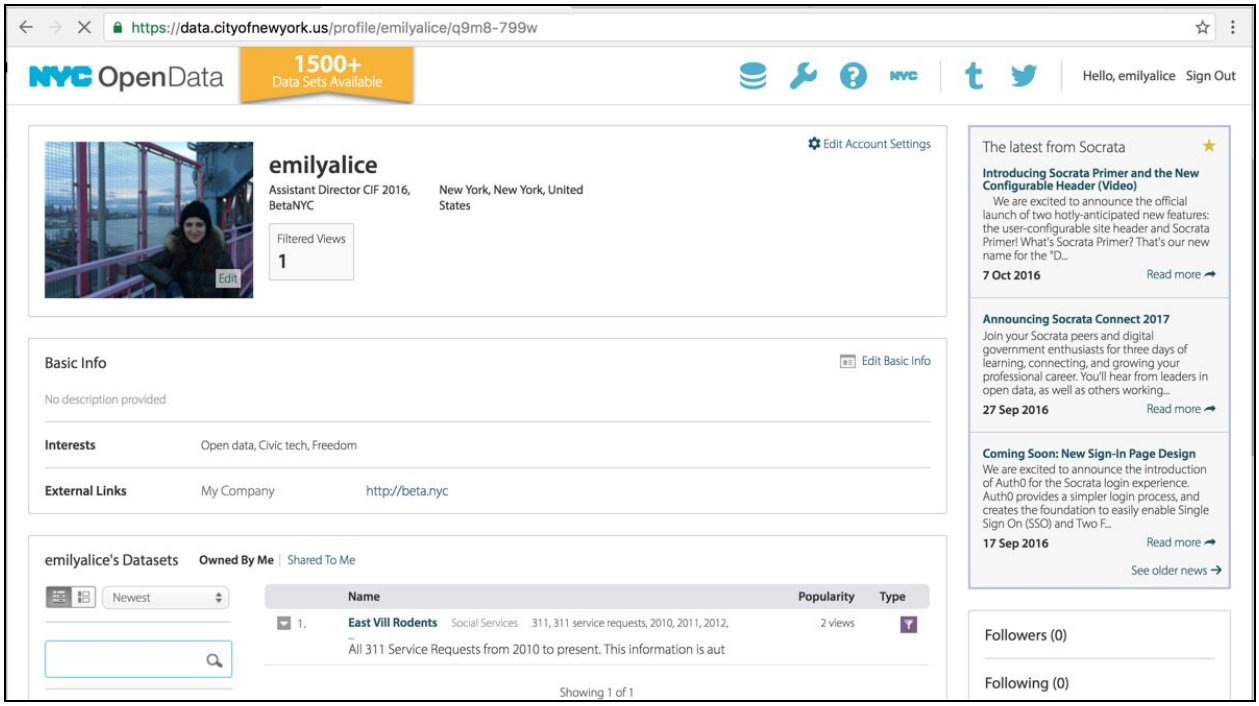
Note - if possible, use Chrome as your browser. Socrata prefers Google Chrome browser!

Lab Steps:

1. First things first: **(Sign up for an account)**

2. Login to socrata: **(Link)**

3. Create a profile page; upload a photo. Paste a screenshot of it here--including the URL as pictured below:



4. Click the datasets icon (resembles a layer cake. Think about why).

5. Tour the Socrata environment:

First observe the domain name in the address bar. Does it say?: https://data.cityofnewyork.us….

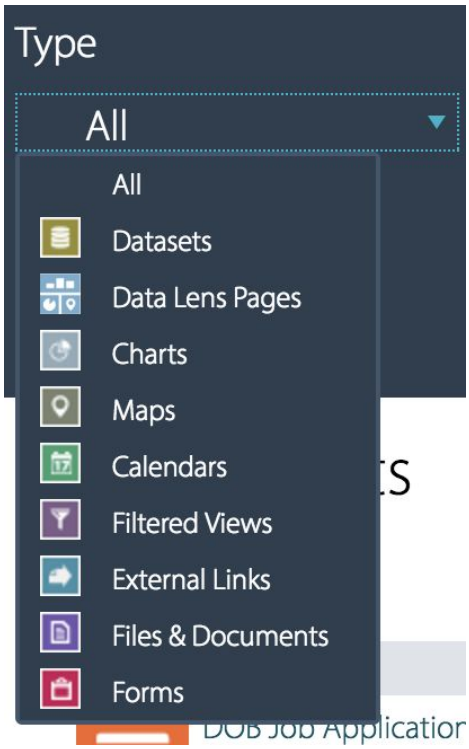
Circle: Yes or No. **If No, what does it say instead?**

Socrata's NYC Open Data portal has at least two potential domain names. Sometimes, a switch between domains may occur, and you may be logged out and lose your work. Therefore, *frequently save your work*, so that if the domain switches, when you log back in, you can resume where you left off.

On the homepage, notice the different ways of refining any given Search. You can search by **Type, Category, Agency, Source**, or through the general Search bar.

Note - Socrata can contain non-government offical filters. If you are looking for the City's official data, make sure that Source is "*Official Data".*

Socrata Data Types:
Notice some of the different types of data available. Datasets, Data Lenses, Filtered Views, etc.



Part 2: Explore Data Lenses.

1. How many different Data Lenses are available in the original list?

Scroll through the complete list of Data Lenses, noticing what topics are covered.

Since the topic of *Links* came up during our Introductory meeting with Aldrin Bonilla, open the LinkNYC Locations Data Lens, and answer the following questions:

2. On which day were fourteen kiosks completed construction? This is the day when 14 were completed in one day.

3. Which two Community Boards have the same number of kiosks?

4. Now, open the 2015 Yellow Taxi Data Lens. Please note the number of Credit Card and Cash transactions in 2015.

5. Next open the 1995 Street Tree Census Data Lens. How many Gingko trees were recorded in NYC in 1995?

6. Finally, open the NYC 311 Service Request Data Lens, noting that it has by far the greatest number of views. Which Community Board displays having the greatest number of 311 calls?

7. According to the Created Date graph, during what month and year was the greatest number of 311 calls recorded?

8. Before moving on, tell us in your own words what the phrase *Data Lens* means to you.

Now, go back to All Data, and explore **Category:** Check out the different options available in the drop down menu under Category.

9. Which Category do you think is most likely to contain 311 data?

10. Now, tell us which Category actually contains 311 data.

With All selected in Type and Category, explore **Agency.** Look at the many agencies that may post data on Socrata.

11. How many datasets are posted by each of the Borough President Offices?:

Bronx:
Brooklyn:
Manhattan:
Queens:
Staten Island:

12. How many datasets are posted by 311?

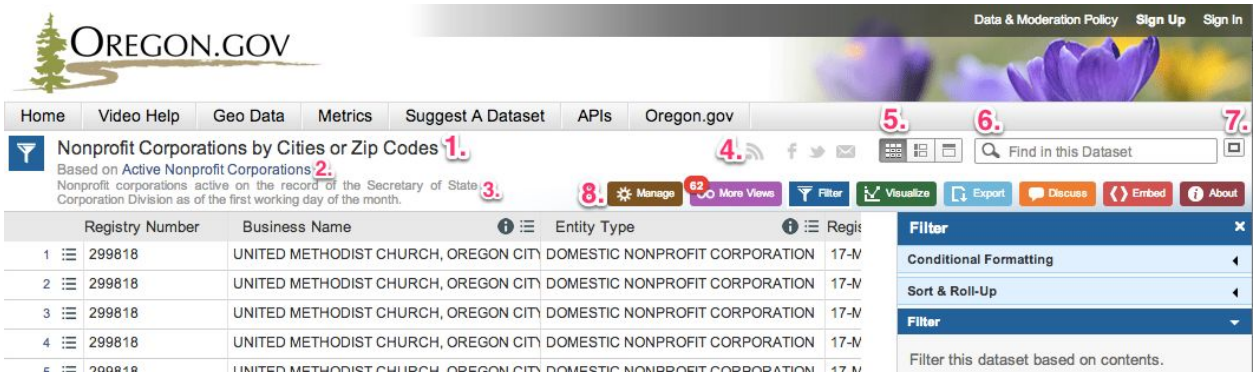
13. Find an agency that has *more than 30* datasets posted.

Under **Source,** notice the two options. For the most part, you will stick with Official Data this year. All Data includes datasets created by regular ol' users like ourselves….potentially muddying the waters.

Okay, we're done exploring Socrata's general environment. Now, open the NYC 311 dataset. You can just click the black & yellow NYC 311 image at the top of the page. This will take you to a dataset containing all Service Requests from 2010 to present. We are going to use 311 data to become familiar with Socrata's built-in analytical features.

Part 3: A dataset in the table view

Let's explore a dataset. This is from Socrata's website.



For reference, "View" means any filter, chart, map, etc that is "Saved" from a dataset.

1. Title: This is the title given to the dataset or view. Views, such as charts, can have a different name from the dataset it is based on.

2. Based On: If you are on a view, this will link to the dataset that the view is based on. For example, if you create a pie chart of the total number of governors for each party, this would link back to the dataset with the numbers for each year.

3. Description: This is the description given to the dataset or view.

4. Social Media: This allows you to interact with this dataset or view through social media:

- RSS Subscription: Subscribe to updates on the dataset
- Facebook: Share the dataset or view on Facebook
- Twitter: Share the dataset or view on Twitter
- Email: Share the dataset or view over email

5. Dataset View: These buttons allow you to switch between the different views of a dataset or view, you can have more than one view splitting the screen vertically at a time

- Detail Row: This view displays each column and row in a tabular format (seen in example above)
- Fat Row: This view displays the details for each row grouped together, for example columns may be in a vertical list.
- Single Row: This view displays all the details for one row at a time, There are arrows to scroll through next and previous rows.
- Chart: If you are viewing a chart a button for the chart view will appear.
- Map: If you are viewing a map, a button fro the map view will appear.

6. Search: Enter in a word or words to search within the dataset for matches.

7. Full Screen: Make the dataset or view full screen on your display.

8. Sidebar Buttons: Each of these buttons opens up a sidebar on the right-hand side of the dataset or view. Not all will appear, depending on if you are logged in and what role you have on the site. For more about user roles, read this article.

- Manage: Click this sidebar to transfer ownership of the dataset or view to another user, delete the dataset or view, share the dataset or view with other users, make the dataset public or private, show and hide columns, and change the column order.

- More Views: Click this sidebar to see the other views created from the same dataset. These might be filtered views, charts, or maps. The number in red reflects the total number of views created from the dataset.
- Filter: Click this sidebar to sort the dataset by columns, group and roll up the dataset, filter the dataset, and set the default filter for that view.
- Visualize: Click this sidebar to set conditional formatting, create a calendar, create a map (if you have location data), and create a chart.
- Export: Click this sidebar to access APi information, print the dataset, and export the dataset.
- Discuss: If commenting is enabled, click this sidebar to comment and read other's comments on the dataset or view. Commenting can be enabled for the dataset and cell level.
- Embed: Click this sidebar to create a form from the dataset and create Social Data Player embeds
- About: Click this sidebar to view the metadata information about the dataset, edit the metadata, view dataset analytics, and contact the dataset owner.

The following is the most important functions you'll need to know:

- **Filter**
    - Sort & Roll-Up
    - Filter
- **Visualize**
    - Calendar
    - Map
    - Chart
- **Export**
    - API information
    - Download
        - CSV
        - CSV for Excel
- **Discuss**
    - Where you can engage with the dataset owner and other dataset users.
- **About:**
    - Dataset description
    - Ratings
    - Specific Metadata
    - Responsible Agency

- Frequency of updating
- Contact Dataset Owner

Part 4: Filter, Roll-Ups & Drill-Downs

According to Wikipedia, a filter in data analysis is:

"A higher-order function that processes a data structure in some order to produce a *new data structure*, containing *exactly those elements of the original data structure for which a given predicate returns the Boolean value true.* " Depending on how geeky you are, you may want to think about this (or not).

In simpler terms, filtering data in a spreadsheet means:

"to set conditions so that *only certain data is displayed*. It is done to make it easier *to focus on specific information in a large database* or table of data."

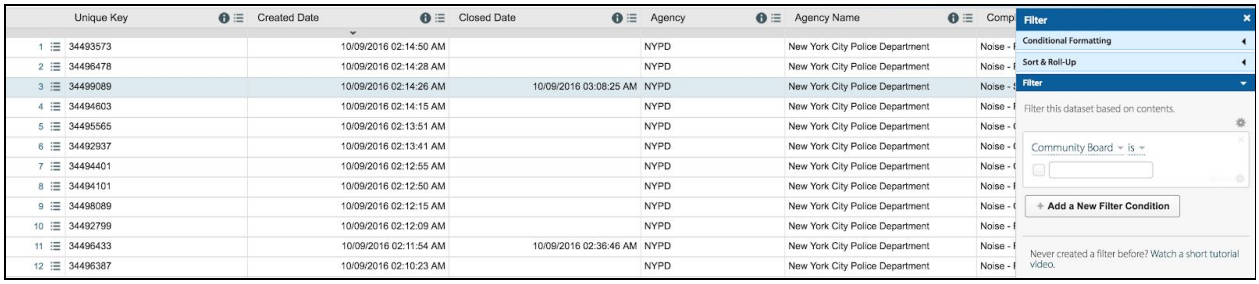
(http://spreadsheets.about.com/od/f/g/2011-05-01-filter-definition.htm, 2011)

Hence, the concept of a filter.

Ok, back to the data. Because the 311 dataset is so large, and analysis on it as a whole can be very slow, let's first *filter* the original dataset to display only the data from your assigned Community Board. This will be a geographic filter.

A Geographic Filter:

On the right side of the Socrata environment, click the blue filter icon, and under where it says, “Filter this dataset based on contents,” change the column from Unique Key to Community Board.



If you leave the operation - is - you will have to type in your Community Board exactly as it appears in the database, case-sensitive. For example, “08 BROOKLYN.”

If you change the operation to - contains - it won’t be case sensitive and typing “08 brooklyn” or *should* yield the same results. However you choose to proceed with your filter, you might as well get used to how Community Boards are labeled in Socrata now.

After typing in your CB, click or hit enter, and the filter process begins. Once the filtered data displays, notice the yellow alert on the left side--”Unsaved View.” Save your filtered data! Save frequently so that if you accidentally get logged out or hit the back button, you don’t have to redo your steps.

Roll-Ups & Drill-Downs:

When we roll-up data, we are aggregating the data *up* by one of its dimensions, often *counting* the total number of rows per this dimension. When we drill-down, we are parsing the data *down* into greater detail.

Let’s start with an example. The table below represents a made-up “311” dataset containing seven rows.

Unique Key	Complaint Type	Descriptor	Borough	Zip Code
1234	Street	Broken Pavement	Man	10024
1235	Noise	Construction	Brooklyn	10261

1236	Housing	Tenant Harassment	Brooklyn	10263
1237	Housing	Insecure Scaffolding	Man	10007
1238	Street	Pothole	Bronx	14760
1239	Street	Pothole	Queens	11365
1240	Noise	Construction	Brooklyn	10265

Let's aggregate the data by **Complaint Type** first, **rolling-up** the individual incidents per Complaint Type by counting the number of incidents for each Type.

Complaint Type	Count Unique Key
Street	3
Noise	2
Housing	2

This simply tells us how many calls were made per Complaint Type.

Next, let's **drill-down** into this dataset by adding another column--say, Descriptor.

Complaint Type	Count Unique Key	Descriptor
Housing	1	Tenant Harrassment
Housing	1	Insecure Scaffolding
Noise	2	Construction
Street	2	Pothole
Street	1	Broken Pavement

Now, let's **drill-down** further by adding a third column--Borough.

Complaint Type	Count Unique Key	Descriptor	Borough

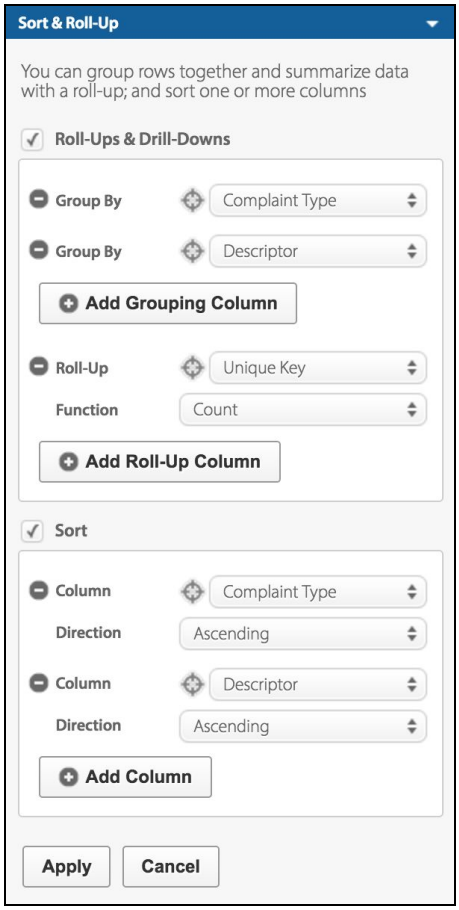
Housing	1	Tenant Harrasment	Brooklyn
Housing	1	Insecure Scaffolding	Man
Noise	2	Construction	Brooklyn
Street	1	Pothole	Bronx
Street	1	Pothole	Queens
Street	1	Broken Pavement	Man

Note: if we drill-down any further into Zip, we would end up back to the original dataset!

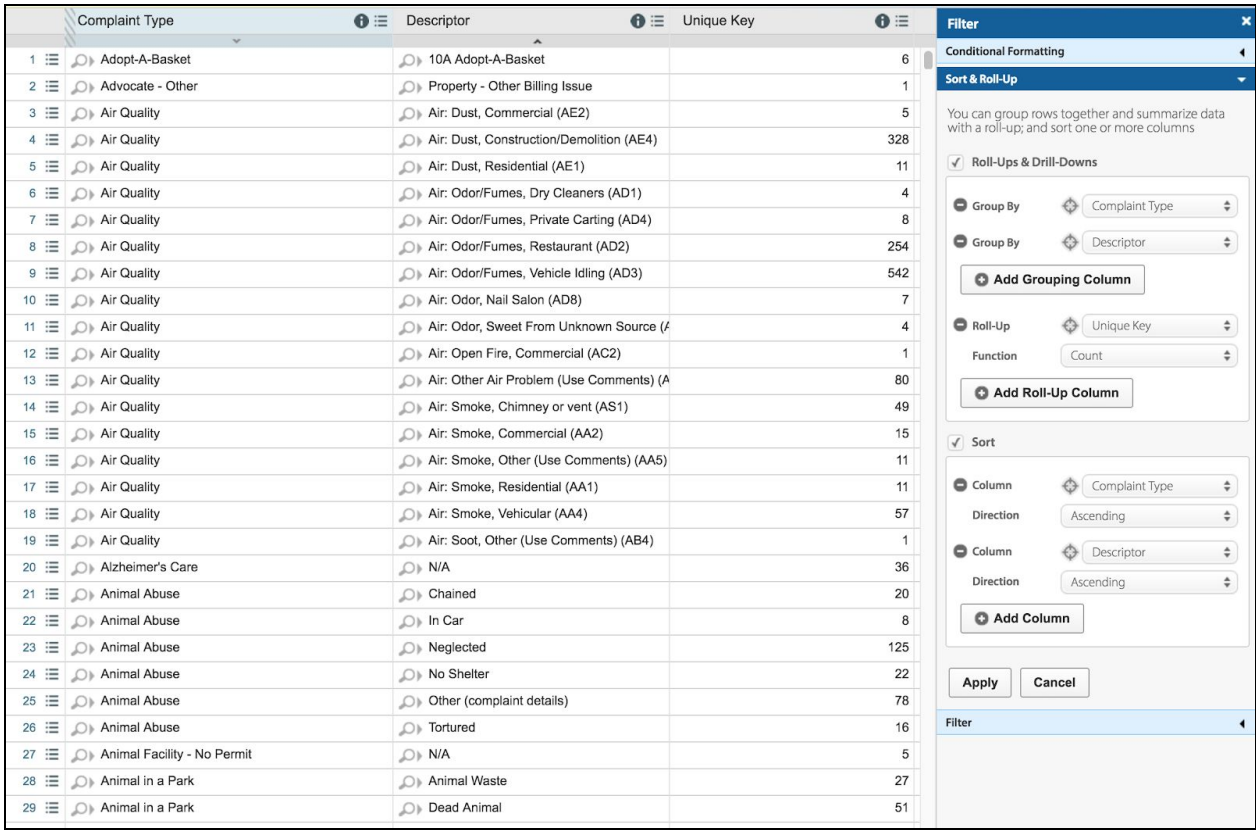
In Socrata, you will typically select a first Grouping column and roll-up by Unique Key using the function Count. You can drill-down into this rolled-up data by adding subsequent Grouping Columns. You'll become expert on using this function by the end of the lab. Okay, back to the real 311 data.

Let's do a simple roll-up on your Community Board data. Using the dataset filtered to your CB, open the Sort and Roll-up Feature on the right side.

1. Group by **Complaint Type** in order to aggregate the data by the primary service requests to which NYC 311 recorded the call.
2. Roll-up the data by the column **Unique Key**, and function **Count** (note this is the only function available) to **count** the number of unique incidents per Grouping Column (in this case, Complaint Type).
3. Drill-down into this data by **adding a second Grouping Column**--Descriptor.
4. Under Sort, we will alphabetize both columns to make it easy to read. First, select **Complaint Type** and select **Ascending**. Then, select **Descriptor** and select **Ascending.**



We get the following results:



	Complaint Type	Descriptor	Unique Key
1	Adopt-A-Basket	10A Adopt-A-Basket	6
2	Advocate - Other	Property - Other Billing Issue	1
3	Air Quality	Air: Dust, Commercial (AE2)	5
4	Air Quality	Air: Dust, Construction/Demolition (AE4)	328
5	Air Quality	Air: Dust, Residential (AE1)	11
6	Air Quality	Air: Odor/Fumes, Dry Cleaners (AD1)	4
7	Air Quality	Air: Odor/Fumes, Private Carting (AD4)	8
8	Air Quality	Air: Odor/Fumes, Restaurant (AD2)	254
9	Air Quality	Air: Odor/Fumes, Vehicle Idling (AD3)	542
10	Air Quality	Air: Odor, Nail Salon (AD8)	7
11	Air Quality	Air: Odor, Sweet From Unknown Source (A	4
12	Air Quality	Air: Open Fire, Commercial (AC2)	1
13	Air Quality	Air: Other Air Problem (Use Comments) (A	80
14	Air Quality	Air: Smoke, Chimney or vent (AS1)	49
15	Air Quality	Air: Smoke, Commercial (AA2)	15
16	Air Quality	Air: Smoke, Other (Use Comments) (AA5)	11
17	Air Quality	Air: Smoke, Residential (AA1)	11
18	Air Quality	Air: Smoke, Vehicular (AA4)	57
19	Air Quality	Air: Soot, Other (Use Comments) (AB4)	1
20	Alzheimer's Care	N/A	36
21	Animal Abuse	Chained	20
22	Animal Abuse	In Car	8
23	Animal Abuse	Neglected	125
24	Animal Abuse	No Shelter	22
25	Animal Abuse	Other (complaint details)	78
26	Animal Abuse	Tortured	16
27	Animal Facility - No Permit	N/A	5
28	Animal in a Park	Animal Waste	27
29	Animal in a Park	Dead Animal	51

Notice how the Complaint Type is listed alphabetically and that the Unique Key column provides a **count** of the total number of unique calls per Complaint Type. Viola, the roll-up function in action!

Now, save this “view” as “NYC311_MN_CB##_Complaint_Detail” and you will have a perpetually updating view of what is going on within your Community Board. Note, ## should be the Community Board number you are working with.

Paste the URL for this data roll-up here:

Part 5: Building a Complex Filter & Sorts

Complex Filter:

Let's build a complex filter upon the geographic filter created earlier. To build a complex filter, click "Add a New Filter Condition," and change Unique Key (the default) to Created Date. You are going to add a Time Range filter to the data now.

Click the operation - is between - and populate the bars to display only the 311 calls in your CB made between the first of October and 17 October.

14. How many calls were made during this ~17 day time range?

Add another New Filter Condition. Change the default Unique Key to Complaint Type. Change the operator from - is - to - contains - and type in "heat." Click or hit enter. Is hasn't been that cold yet in October, so perhaps we would not expect a large number of complaints about heat.

15. How many calls were made between October 1 and October 17 in your CB about a heating issue?

Have you been saving your filters? Get in the habit of doing this, please :) You are going to use this dataset to make a map, so if you have not already, SAVE IT NOW. Call it MAN_CB##_Oct_Heat.

Sorts:

Sorting data allows you to arrange it in an ordered sequence. The opposite of sorting would be shuffling.

Now, click the Sort & Roll-up menu. Check the box next to Sort. Choose a *non-numeric column* to sort by, such as Descriptor, and click Descending or Ascending.

16. Describe how Socrata sorts a non-numeric column in Descending order.

Next choose a numeric column to sort by, such as Created Date.

17. Describe how Socrata sorts a numeric column in Ascending order.

Part 6: Making Maps in Socrata

Point Maps:

In the GIS world, point maps store and represent spatial information consisting of points. Socrata lets us make point maps very easily. Navigate to your saved dataset MAN_CB_##_Oct_Heat. On the right side, click Visualize, then Map. Make sure it indicates the right dataset.

As shown above, give it an Alias, and for Plot Style, choose Point Map. (Leave everything else default for now).

Then hit apply at the bottom to generate a point map of heat complaints so far in October. Click a few individual points to see what further information comes up.

Take a screenshot of your point map and paste it below.

19. Did you find any points that represent multiple rows of data? If so, name an address with multiple heat complaints, and tell us how many heat complaints the point represents.

Heat Maps:

A **heat map** (or **heatmap**) is a graphical representation of data where the individual values contained in a matrix are represented as colors (Wikipedia). They enable us to see “hot spots” or clusters.

Change the Plot Style to Heat Map, and paste a screenshot below of your map.

20. Is there a hot spot of heat complaints? If so, tell us the boundaries of a hot spot.

Part 7: Putting it all together:

Objective - Create a filter of your adopted Community Board

- First
    - Filter by your Community Board by selecting "Community Board", "contains", and typing in "## BOROUGH." If your Community Board is Manhattan 5, you would enter "05 MANHATTAN"
- Second
    - Click "Save As" filter as "CB-Man-##-All"
- Third
    - Paste a URL to this filter in your lab sheet.

Objective - Top 20 Complaint Types for your Community Board, their descriptor, and count of incidents.

- First
    - Open up your saved filter - "CB-Man-##-All"
- Second
    - Select Filter and select "Sort & Roll-Up"
    - Select "Roll-Ups & Drill-Downs"
    - Group by "Complaint Type" & "Descriptor". Select "Add Grouping Column" to get a second selection pull down.
    - Roll-up by "Unique Key", function "Count"
    - Sort by "Unique Key", direction "Descending"
    - Select "Apply"
- Third
    - Be very careful and select "Save As"
    - Name this filter "CB-Man-##-All-Issues-RolledUp"
- Fourth
    - Export as a "CSV for Excel" **(If you have not encountered CSV yet, it means Comma Separated Values, and we will get into this more deeply next week).**
    - Open the exported file in Google Sheets or Microsoft Excel.
    - Cut and paste a table with the top 20 service requests, descriptors, and count of unique keys into the lab's sheet.
    - Cut and paste the filter's URL into your lab sheet.

Objective - What are 2016's top 10 Complaint Types for your Community Board, their descriptor, and count of incidents.

- First
    - Open up your saved filter - "CB-Man-##-All"

- Second
    - Under filter, select "Edit Default Filter"
    - Select "Add a New Filter Condition"
    - Under the pull down, select "Created Date", "is after", and then select "Jan 01 2016"
    - Now, carefully, select "Save As"
    - Name this filter "CB-Man-##-2016-Issues-RolledUp" and click save.
    - Now in the saved view, look at the "Create Date" field and place a check box next to the date "01/01/2016 12:00:00 AM" and let Socrata do its thing.
        - *Note - this filter will continue to aggregate data from 01/01/2016 to the current date. This means its numbers will change. If you are running this lab after 17 Oct, please look at using "is between" and set the close date to 10/17/2016.*
    - With the data filter set to all issues in 2016, select "Save" again. Then, wait for Socrata to do its thing.
- Third
    - Now, select "Sort & Roll-Up"
    - Select "Roll-Ups & Drill-Downs"
    - Group by "Complaint Type" & "Descriptor". Select "Add Grouping Column" to get a second selection pull down.
    - Roll-up by "Unique Key", function "Count"
    - Sort by "Unique Key", direction "Descending"
    - Select "Apply" and let Socrata do its thing.
    - Now, select "Save"
- Fourth
    - Export as a "CSV for Excel"
    - Open the exported file in Google Sheets or Microsoft Excel.
    - Sort the table by "Unique Key" and "Descending"
    - Cut and paste a table with the top 10 service requests, descriptors, and count of unique keys into the lab's sheet.
    - Cut and paste the filter's URL into your lab sheet.
- Fifth
    - Between Compare all Service Requests and Descriptors with the 2016 and note the differences between the two.
    - What is the difference? Write a narrative of what is the difference between the two.

Objective - Let us compare 2015 data with 2016 data.

- First
    - Open up your saved filter - "CB-Man-##-All"
- Second
    - Under filter, select "Edit Default Filter"

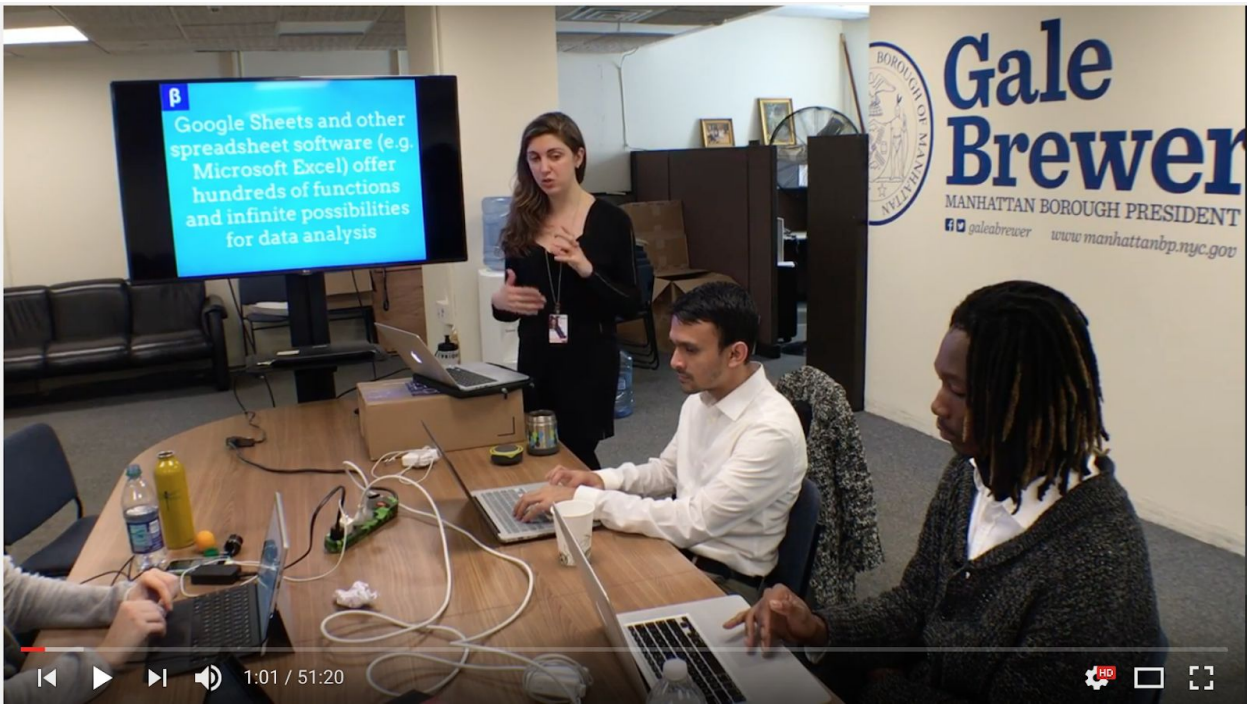
    - Select "Add a New Filter Condition"
    - Under the pull down, select "Created Date", "is between", and then select "Jan 01 2015" and "Oct 17 2015"
    - Now, carefully, select "Save As"
    - Name this filter "CB-Man-##-2015-01to10-Rolledup" and click save.
    - Now in the saved view, look at the "Create Date" field and place a check box next to the date "01/01/2015 12:00:00 AM" and let Socrata do its thing.
    - With the data filter set to all issues in 2015, select "Save" again. Then, wait for Socrata to do its thing.
- Third
    - Now, select "Sort & Roll-Up"
    - Select "Roll-Ups & Drill-Downs"
    - Group by "Complaint Type" & "Descriptor". Select "Add Grouping Column" to get a second selection pull down.
    - Roll-up by "Unique Key", function "Count"
    - Sort by "Unique Key", direction "Descending"
    - Select "Apply" and let Socrata do its thing.
    - Now, select "Save"
- Fourth
    - Export as a "CSV for Excel"
    - Open the exported file in Google Sheets or Microsoft Excel.
    - Sort the table by "Unique Key" and "Descending"
    - Cut and paste a table with the top 10 service requests, descriptors, and count of unique keys into the lab's sheet.
    - Cut and paste the filter's URL into your lab sheet.
- Fifth
    - Compare 2015's Service Requests and Descriptors with 2016's data and note the differences between the two.
    - What is the difference? Write a narrative of what is the difference between the two.
    - Now, look at all three — All vs 2016 vs 2015. What do you see across the top 10 service complaints? Does anything stand out? Note this in your lab.

Concluding Remarks: Congratulations! Consider yourself an expert user of Socrata. You now have the knowledge and ability to analyze data on its 1,000+ government organization web portals. You've also learned some key concepts in data analysis that will transfer to most other platforms you use in the future.

Deliverables:

Complete the 20 numbered questions which are highlighted in bold, and all the final objectives.

Civic Innovation Fellowship Y3 Boot Camp
Lab # 05: Data Analysis in Spreadsheets
Due: October 28, 2016



Youtube video of lecture: http://bit.ly/2j0ecL1
Link to the slides here: http://bit.ly/2iVSjvO

Analyzing Large Datasets in Spreadsheets:

Excel, Google Sheets, OpenOffice, LibreOffice are all examples of spreadsheet software. For this lab, we will be working with Google Sheets.

Many of the functions and features you will learn transfer from one software to the other, though. And you will have an opportunity later in the year to apply the skills in Excel.

Last week, you learned how to analyze data within the Socrata platform. Super useful for Open Government data particularly. This week, you will learn how to analyze data within software that is used for a tremendous range of purposes, from book-keeping for small businesses, financial modeling for large corporations, and research and policy analysis in academia, non-profits, and government agencies.

This week, we have pre-loaded some datasets into the lab folder that we would like you to work with. Your results should all be comparable, therefore, and you can more easily work together!

Not all the data will be perfectly “clean,” so you will have to tackle working with messy data. That’s great practice because, in reality, that’s how most data is.

Lab Objectives:

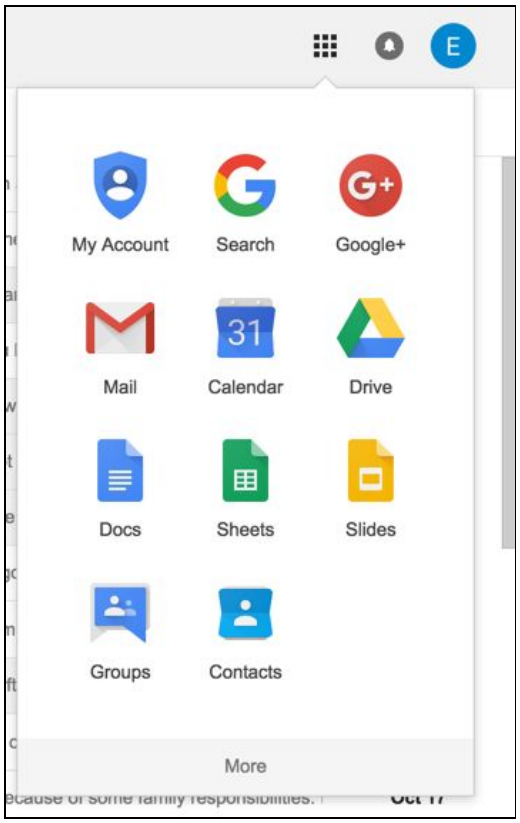
- **Become familiar with Google Sheets spreadsheet environment**
- **Learn how to use the following formulas and functions for data analysis:**
    - **Sort**
    - **Filter**
    - **Split Text to Columns**
    - **SUM**
    - **Percent Change formula**
    - **TEXT**
    - **CONCATENATE**
    - **Find Command**
    - **FIND**
    - **IF**
    - **ISBLANK**
    - ***Pivot Table***
- **Analyze how the total population of your CB has changed since 1970 using DCP data (what does DCP stand for? -- Department of City Planning).**
- **Prepare for analysis some Rent Guidelines Board data using Concat and Text functions.**
- **Clean and analyze the Social Media / Community Board spreadsheet you all created at the beginning of the semester.**
- **Execute several complex analyses on 311 data using the techniques you have just learned.**

Part 1: Become familiar working with data in spreadsheets

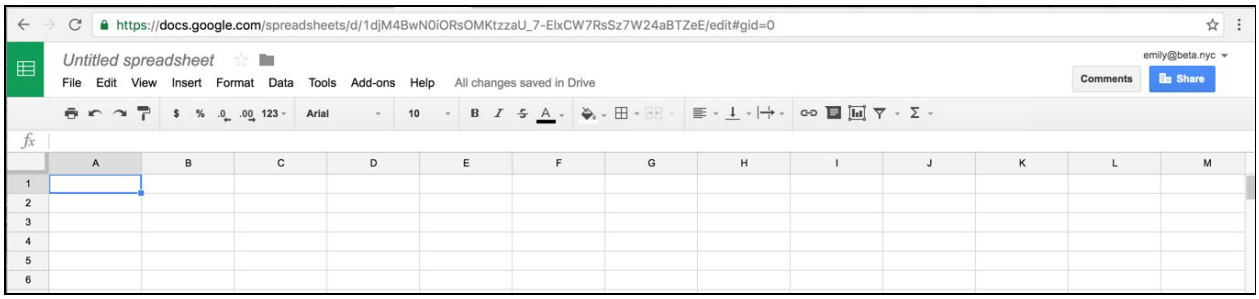
Let’s step back for a minute and define what actually *is* a spreadsheet. A spreadsheet is defined by Wikipedia as “an interactive computer application for organization, analysis, and storage of data in tabular form.” Tabular means that the data is stored in **columns** and **rows**.

Let’s tour the Google Sheets environment:

From your Gmail Account, you can open up a new Google Sheet by clicking on Sheets in the apps menu on the right side.

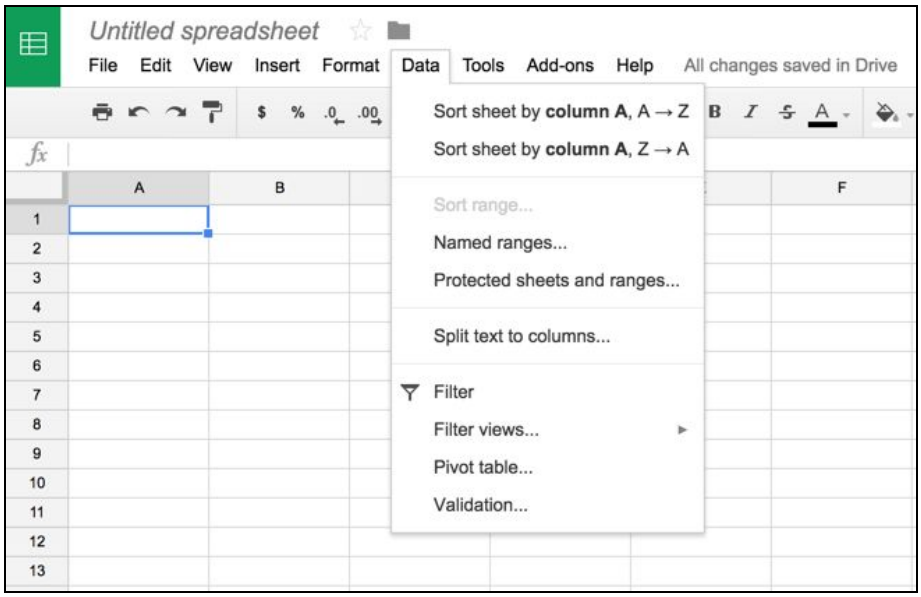


Notice it opens in your web browser.

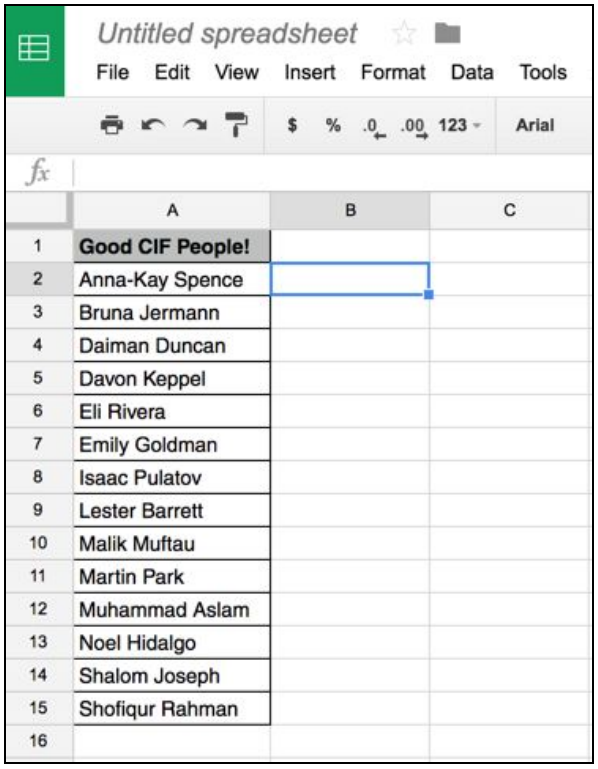


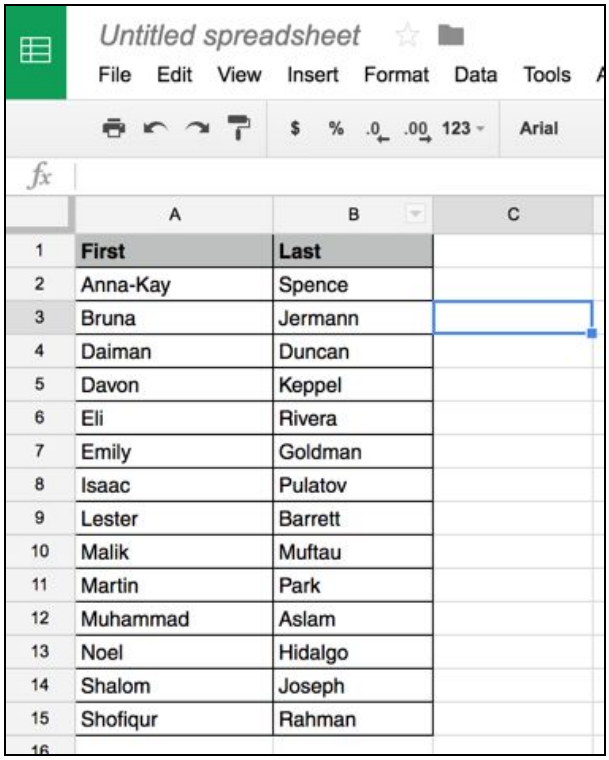
You're probably familiar with most of the tabs and icons pictured above, but let's turn our attention to the Data tab for now. The Data tab is where you will find most of the basic data analysis functions built into Google Sheets. Notice that some of the functions we learned last week from Socrata are available here as well: **Sort**, **Filter,** and **Pivot Table** (Pivot Table is

analogous to the Roll-Up function in Socrata). We will demonstrate how to use Sort and Filter in Google Sheets in the lecture and will teach the Pivot Table function in detail later in the lab.



Split text to columns enables you to split a single column of data into multiple columns depending in an ordered way. This can be very handy. Say you have a single column containing First and Last names of individuals and want to have a column for each. You can use Split text to columns to create two columns, based on where there is a particular separator in the contents of the original (in this case, a space).


Untitled spreadsheet
File Edit View Insert Format Data Tools
$ % .0 .00 123 Arial
fx
A B C
1 Good CIF People!
2 Anna-Kay Spence
3 Bruna Jermann
4 Daiman Duncan
5 Davon Keppel
6 Eli Rivera
7 Emily Goldman
8 Isaac Pulatov
9 Lester Barrett
10 Malik Muftau
11 Martin Park
12 Muhammad Aslam
13 Noel Hidalgo
14 Shalom Joseph
15 Shofiqur Rahman
16


Untitled spreadsheet
File Edit View Insert Format Data Tools
$ % .0 .00 123 Arial
fx
A B C
1 First Last
2 Anna-Kay Spence
3 Bruna Jermann
4 Daiman Duncan
5 Davon Keppel
6 Eli Rivera
7 Emily Goldman
8 Isaac Pulatov
9 Lester Barrett
10 Malik Muftau
11 Martin Park
12 Muhammad Aslam
13 Noel Hidalgo
14 Shalom Joseph
15 Shofiqur Rahman
16

The Fx bar:

It may not be as exciting as a coffee bar or the Genius bar, but the Fx bar at the top of the table is where your functions and formulas appear as you build them.

Okay, that's enough exploring Google Sheets' general environment for now. Let's open up some real data now. The first dataset we are going to use features population data about NYC's Community Districts! We are going to use this population data to become familiar with a few foundational functions and formulas.

Part 2: Analyzing Population Change in Manhattan's 12 CBs!

From Google Sheets, click File-->Open, and navigate to a dataset called CB_Pop_Data in the Lab 5 folder under Weekly Assignments in the Drive. This data comes directly from the BYTES of the Big Apple website, but we prepared it just a bit for the purposes of the lab. Take a look at where we found the original here if you'd like.

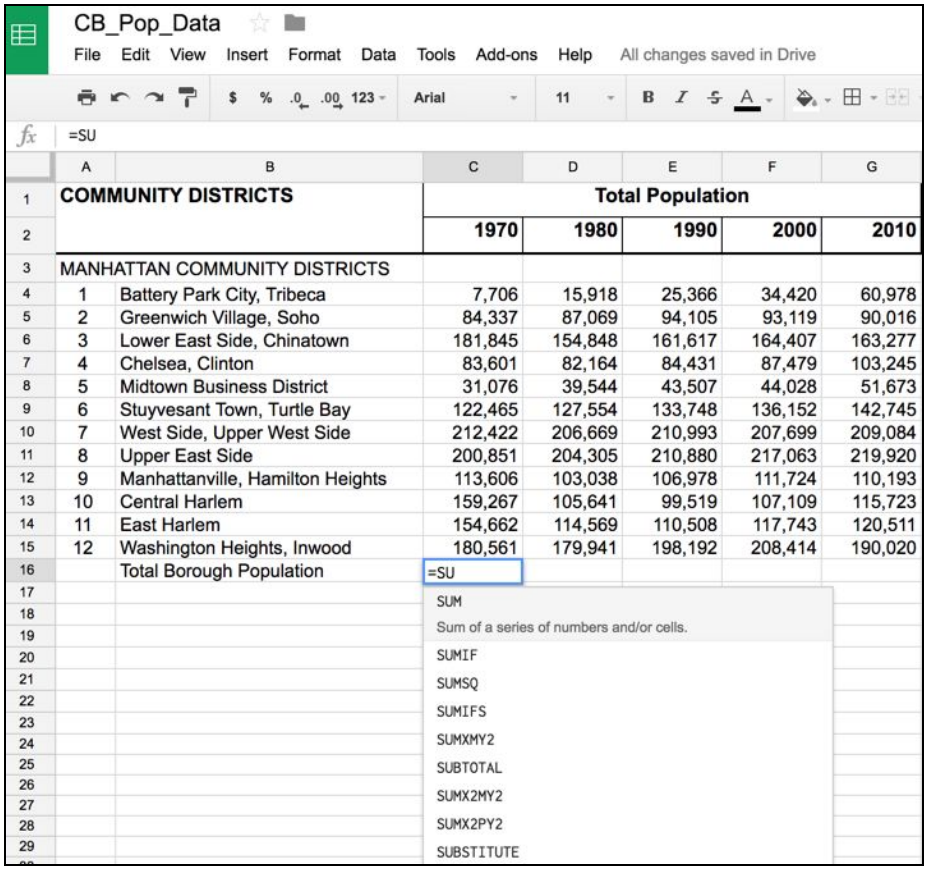
In your analysis of this data, you are going to use FUNCTIONS and FORMULAS. What's the difference you might ask?

Functions and formulas essentially accomplish the same thing, allowing you to perform calculations on your data. But, a function is pre-built into spreadsheet software, while a formula generally is not, and can be much more customized.

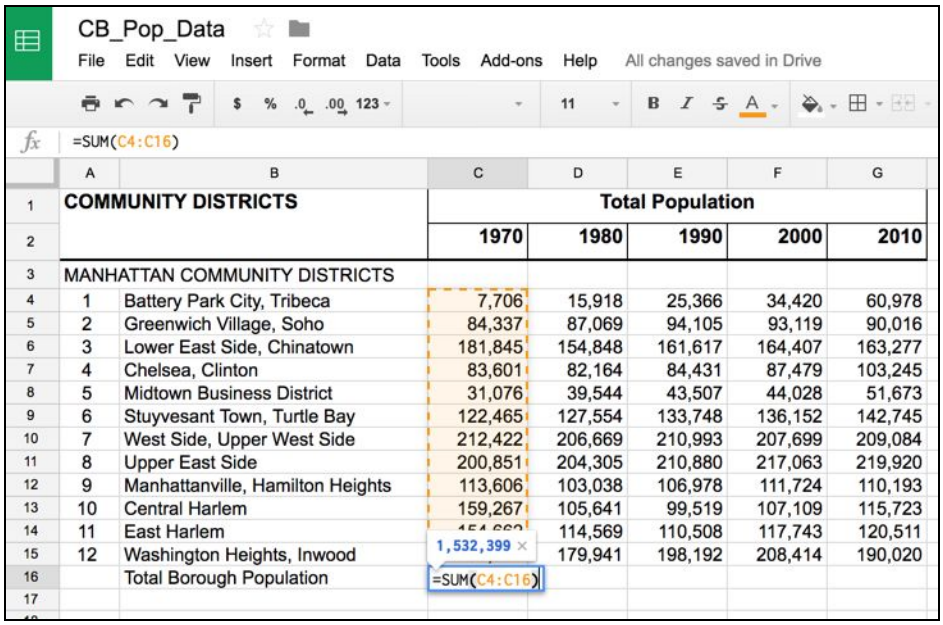
NOTE: Other than Sort, Filter, and Roll-up, Socrata does not lend itself to you performing other functions on your data, while spreadsheet software has **hundreds** of built-in functions, and the capability for you to perform countless customized formulas.

Calculating Total Borough Population for each Decade:

To calculate the total borough population per decade, we'll use the function SUM. To call up a function in Google Sheets, click the cell you wish to populate and type in the = sign. Then, start typing the function's name in CAPITAL LETTERS. A list of functions will appear (in alphabetical order). The function will be built in the Fx bar directly above the table.



After typing or clicking SUM, open parenthesis, and type in the reference number of the cells you wish to sum up, or just click the first and last one, and the whole range will be highlighted.



Click enter and repeat for each decade, or click and drag the lower right corner of the 1970 total population cell to the right through the final column, to perform the same function on the subsequent decades.

Calculating Percentage Population Change:

To calculate the percent population change from 1970 to 2010 for each Community Board, you will use a formula rather than built-in function.

The percentage change formula to compare two years of population data is the following:

((Year 2 population - Year 1 population) / Year 1 population) * 100

Or in more general terms:

Change in Population / Original Population * 100

Click in cell H4 to start your formula for CB 1. As you build the formula, you will see it in the Fx bar directly above the table.

First, click the = sign
Then, open paren.

Then build your formula for percent change in population between 1970 and 2010 by clicking the appropriate cells and typing the correct operators. For guidance, see below:

- Click the cell for Population in 2010 for CB 1
- Subtraction sign
- The cell for Population in 1970 for CB 1
- Close paren
- Division sign (/)
- The cell for Population in 1970 for CB 1
- Multiplication sign (*)
- 100

Click and drag the lower right corner of the result in cell H4 through to the final row in your dataset (Total Borough Population)

Questions:

1 Which CB has the greatest percent change? What is the percent change?

2. Which CB has the smallest (or most negative) percent change? What is that percent change?

3. How about your CB? What is its percent change? Is this surprising? What are your thoughts about this--any ideas for why it gained or lost population?

Part 3: Creating a BBL using CONCAT and TEXT Functions

In the next section of the lab you are going to use two other spreadsheet functions--**Concatenate** and **Text**. You are going to use some Rent Guidelines Board data. The original data was found here. That data, as you can see, is in PDF form. We had to use a Free and Open Source program called Tabula to extract the tabular data into a CSV. Open the dataset called "tabula-2014ManhattanBldgs" in the Lab 5 folder in Google Sheets.

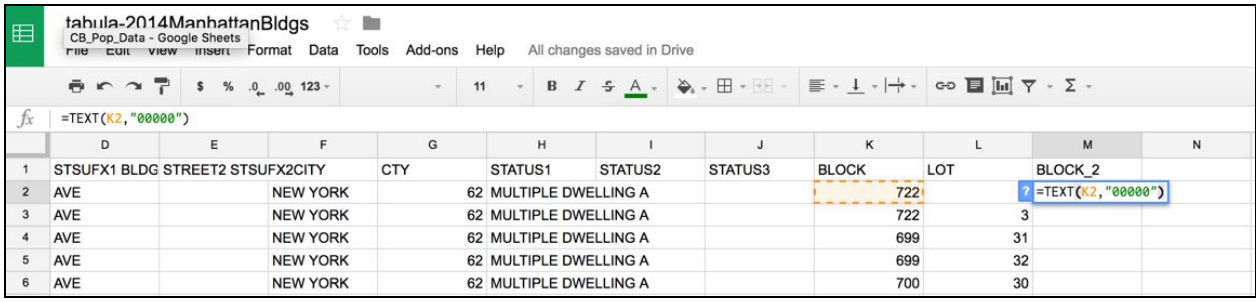
As you can see, there are two columns for Block and Lot. Remember the lesson on Block and Lot? We learned that those numbers are inherently 5 digits, and 4 digits respectively, correct?

FUNCTION: TEXT

Let's use the function TEXT to pad the numbers with the correct number of digits.

Create a new column in Google Sheets called BLOCK_2. Then, click cell M2, and start your TEXT function. As always, start with =.

Then, start typing TEXT, it should appear in a list of built-in functions. After the open paren, click the cell upon which you wish to perform the TEXT function (which is K2), followed by "," and then the number of digits you wish the number to conform to in the form of zeros within quotation marks: "00000." Conceptually does that make sense to you? If not, read more about the TEXT function here and/or here. Click and drag down so the entire M column becomes populated with the correctly formatted data.

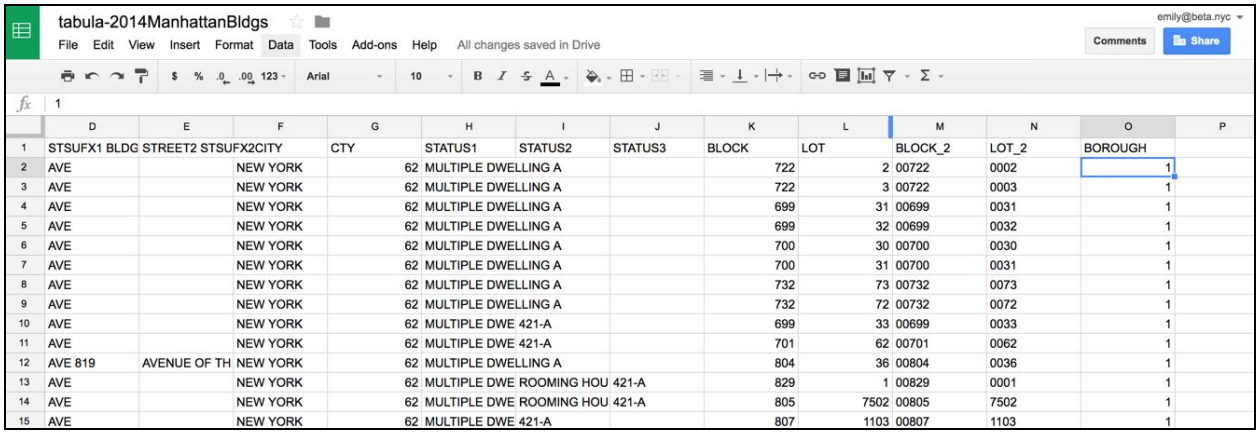


	D	E	F	G	H	I	J	K	L	M	N
1	STSUFX1 BLDG	STREET2 STSUFX2	CITY	CTY	STATUS1	STATUS2	STATUS3	BLOCK	LOT	BLOCK_2	
2	AVE		NEW YORK	62	MULTIPLE DWELLING A			722		=TEXT(K2,"00000")	
3	AVE		NEW YORK	62	MULTIPLE DWELLING A			722	3		
4	AVE		NEW YORK	62	MULTIPLE DWELLING A			699	31		
5	AVE		NEW YORK	62	MULTIPLE DWELLING A			699	32		
6	AVE		NEW YORK	62	MULTIPLE DWELLING A			700	30		

Now, create another column called LOT_2, and do the same, but this time, since LOT is inherently a four digit number, put only 4 zeros in the quotation marks. Drag down to populate the entire column with the correctly formatted data.

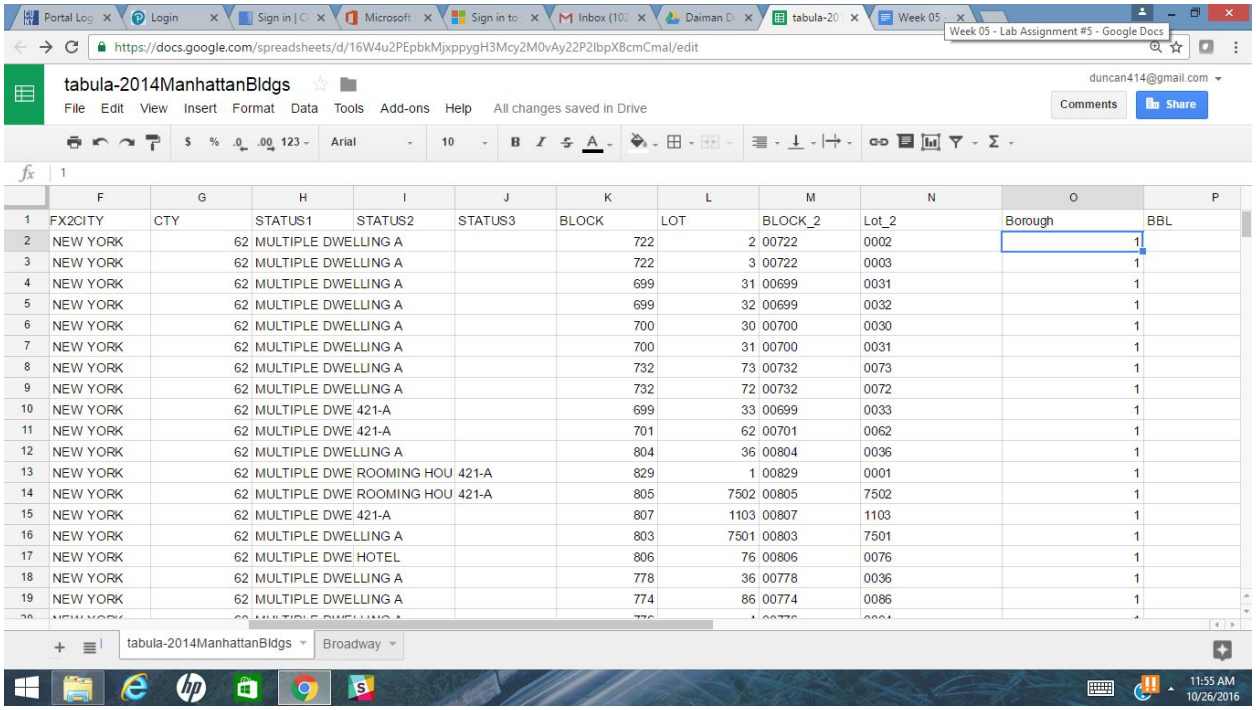
Create another new column called BOROUGH before BBL. Place a 1 in the first cell, which is Manhattan's Borough Code, and drag down to populate the entire column with 1.

You should have a dataset that looks like this:



	D	E	F	G	H	I	J	K	L	M	N	O	P
1	STSUFX1 BLDG	STREET2 STSUFX2	CITY	CTY	STATUS1	STATUS2	STATUS3	BLOCK	LOT	BLOCK_2	LOT_2	BOROUGH	
2	AVE		NEW YORK	62	MULTIPLE DWELLING A			722	2	00722	0002	1	
3	AVE		NEW YORK	62	MULTIPLE DWELLING A			722	3	00722	0003	1	
4	AVE		NEW YORK	62	MULTIPLE DWELLING A			699	31	00699	0031	1	
5	AVE		NEW YORK	62	MULTIPLE DWELLING A			699	32	00699	0032	1	
6	AVE		NEW YORK	62	MULTIPLE DWELLING A			700	30	00700	0030	1	
7	AVE		NEW YORK	62	MULTIPLE DWELLING A			700	31	00700	0031	1	
8	AVE		NEW YORK	62	MULTIPLE DWELLING A			732	73	00732	0073	1	
9	AVE		NEW YORK	62	MULTIPLE DWELLING A			732	72	00732	0072	1	
10	AVE		NEW YORK	62	MULTIPLE DWE	421-A		699	33	00699	0033	1	
11	AVE		NEW YORK	62	MULTIPLE DWE	421-A		701	62	00701	0062	1	
12	AVE 819	AVENUE OF TH	NEW YORK	62	MULTIPLE DWELLING A			804	36	00804	0036	1	
13	AVE		NEW YORK	62	MULTIPLE DWE	ROOMING HOU	421-A	829	1	00829	0001	1	
14	AVE		NEW YORK	62	MULTIPLE DWE	ROOMING HOU	421-A	805	7502	00805	7502	1	
15	AVE		NEW YORK	62	MULTIPLE DWE	421-A		807	1103	00807	1103	1	

Take a screenshot and paste it below, please!

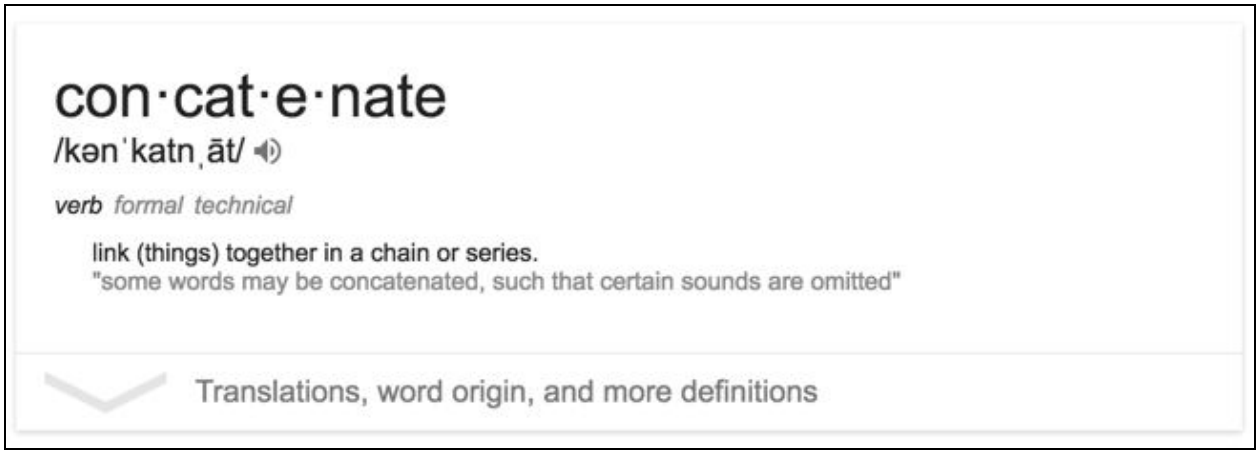


tabula-2014ManhattanBldgs

	F	G	H	I	J	K	L	M	N	O	P
1	FX2CITY	CTY	STATUS1	STATUS2	STATUS3	BLOCK	LOT	BLOCK_2	Lot_2	Borough	BBL
2	NEW YORK	62	MULTIPLE DWELLING A			722	2	00722	0002	1	
3	NEW YORK	62	MULTIPLE DWELLING A			722	3	00722	0003	1	
4	NEW YORK	62	MULTIPLE DWELLING A			699	31	00699	0031	1	
5	NEW YORK	62	MULTIPLE DWELLING A			699	32	00699	0032	1	
6	NEW YORK	62	MULTIPLE DWELLING A			700	30	00700	0030	1	
7	NEW YORK	62	MULTIPLE DWELLING A			700	31	00700	0031	1	
8	NEW YORK	62	MULTIPLE DWELLING A			732	73	00732	0073	1	
9	NEW YORK	62	MULTIPLE DWELLING A			732	72	00732	0072	1	
10	NEW YORK	62	MULTIPLE DWE	421-A		699	33	00699	0033	1	
11	NEW YORK	62	MULTIPLE DWE	421-A		701	62	00701	0062	1	
12	NEW YORK	62	MULTIPLE DWELLING A			804	36	00804	0036	1	
13	NEW YORK	62	MULTIPLE DWE	ROOMING HOU	421-A	829	1	00829	0001	1	
14	NEW YORK	62	MULTIPLE DWE	ROOMING HOU	421-A	805	7502	00805	7502	1	
15	NEW YORK	62	MULTIPLE DWE	421-A		807	1103	00807	1103	1	
16	NEW YORK	62	MULTIPLE DWELLING A			803	7501	00803	7501	1	
17	NEW YORK	62	MULTIPLE DWE	HOTEL		806	76	00806	0076	1	
18	NEW YORK	62	MULTIPLE DWELLING A			778	36	00778	0036	1	
19	NEW YORK	62	MULTIPLE DWELLING A			774	86	00774	0086	1	

FUNCTION: CONCATENATE

What does concatenate mean? See below.



Create a fourth and fifth new column called BL and BBL respectively.

In the column BL, concatenate the data in columns BLOCK_2 and LOT_2 by typing:

=CONCAT(*click your cell1*, *click your cell2*).

Populate the entire column.

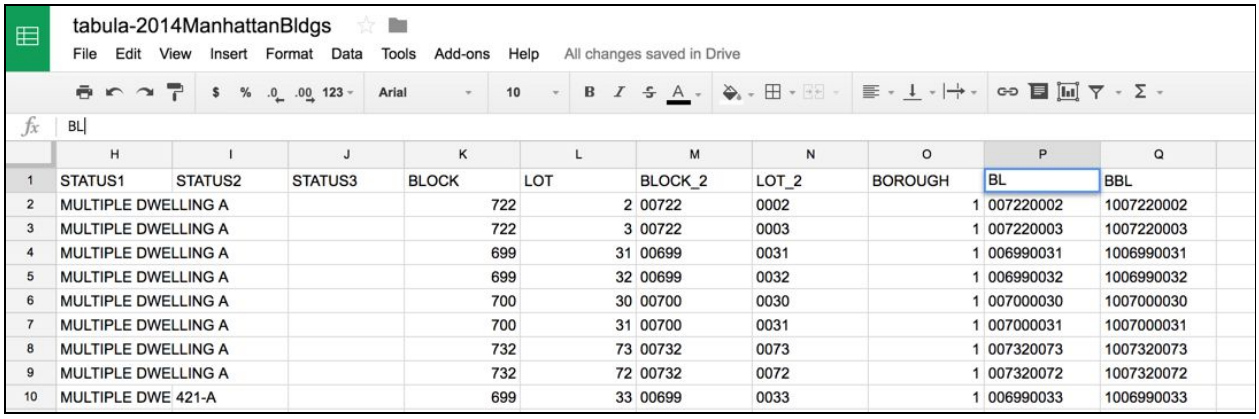
In the BBL column, concatenate the data in columns BOROUGH and BL (IN THAT ORDER PLEASE!) by typing:

=CONCAT(*click your cell1*, *click your cell2*)

Populate the entire column.

Now, you have a column that features the full and accurate BBL of each building in this list of rent-stabilized buildings in Manhattan. It's hard to **overstate** the significance of having **the accurate BBL** for these buildings--for one, it enables you to make a map of those parcels readily, once you know how to do so.

Anyway, the spreadsheet should look like this:



	H	I	J	K	L	M	N	O	P	Q
1	STATUS1	STATUS2	STATUS3	BLOCK	LOT	BLOCK_2	LOT_2	BOROUGH	BL	BBL
2	MULTIPLE DWELLING A			722	2	00722	0002	1	007220002	1007220002
3	MULTIPLE DWELLING A			722	3	00722	0003	1	007220003	1007220003
4	MULTIPLE DWELLING A			699	31	00699	0031	1	006990031	1006990031
5	MULTIPLE DWELLING A			699	32	00699	0032	1	006990032	1006990032
6	MULTIPLE DWELLING A			700	30	00700	0030	1	007000030	1007000030
7	MULTIPLE DWELLING A			700	31	00700	0031	1	007000031	1007000031
8	MULTIPLE DWELLING A			732	73	00732	0073	1	007320073	1007320073
9	MULTIPLE DWELLING A			732	72	00732	0072	1	007320072	1007320072
10	MULTIPLE DWE 421-A			699	33	00699	0033	1	006990033	1006990033

Got it? Oh, also, please note: in other spreadsheet software, you can concatenate many values at once, not just two :)

Part 3: Other useful functions

- FIND
- IF
- ISBLANK

For the following few functions, we are going to use the CB website survey. Make a copy of it now and place it in your Notebooks folder.

FIND:

The Find *command* enables you to search for particular contents across the entire Google Sheet. If you wish to search in a specific column, click Command F, and then in the Find box that appears on the upper right, click the icon with three dots to specify your range.

Using the Website sheet, let's search for *https* in column C containing the URLs, first using the Find command.

Click Find and what do you learn?

Question 4: According to the Find Command, how many CBs have https in their URL?

Now, let's use the FIND function in Google Sheets to double-check our results. We are going to create two new columns, the first of which will tell us if the cell has **http** in the URL, and the second will tell us if the cell has **https** in the URL.

When using the FIND function, *if the string is found, the function will generate a number.*

If the string is not found, the function will generate an error marker in the cell.

You'll see that in action in a few moments.

Insert two new columns next to the URL column (column C). Call column D "http" and column E "https."

In the first cell of column D, type =FIND to summon the FIND function. Then, open paren, and type the following:

=FIND("http", click the C2 cell)

Click and drag down the lower right corner to populate the whole column.

Question 5: What result appears in the cells in Column D? Now, use the Filter function to see if there are other results further down the column. Are there other results?

Now do the same in column E, searching for the string "https."

Question 6: What result appears in the cells? Use the Filter function to see if there are other results further down the column. Are there?

If you're still unsure about these results, at the bottom of the data, create an extra row or two, and give them fake URLs in the C column that do have **https** in them. Now, use the FIND function in column E, and see what happens in those cells. Do you get an error or a number?

IF:

The IF function in spreadsheets is a **very powerful way** to run a logical test on your data, and generate a TRUE or FALSE in a cell in a new column, depending on if the condition is met in the cell in the original column. Make sense? Read more about the IF function here.

Let's execute the IF function on the Twitter Sheet of the Survey data to find out which CBs are "heavy" Twitter users and which are not. The condition on which we will test is whether the number of tweets exceeds 200.

Insert a new column next to Tweets, and call it "Lotta Tweets."

In the first cell of this new column, start your IF function.

Type =IF(click the adjacent cell that contains the number of tweets > 200, TRUE, FALSE)

The syntax of this function means: IF the cell contains a number that is greater than 200, generate a TRUE, **otherwise,** generate a FALSE.

Click and drag down the lower right corner so the function is applied to the entire column. Now, use the Filter function to find out how many CBs qualify as heavy Twitter users (i.e. have a TRUE).

Question 7: How many CBs qualify as heavy Twitter users according to this logical condition?

ISBLANK:

ISBLANK is a function that enables you to find out if the cell is empty, by generating a TRUE or FALSE (TRUE if it is empty). In the Twitter Sheet, insert a new column adjacent to the Complete Profile column and call it, "Profile Info Is Blank."

In the first cell of the new column, start the ISBLANK function. Type:

=ISBLANK(click the adjacent cell in the Complete Profile column). Hit enter. Then click and drag down to populate the entire column.

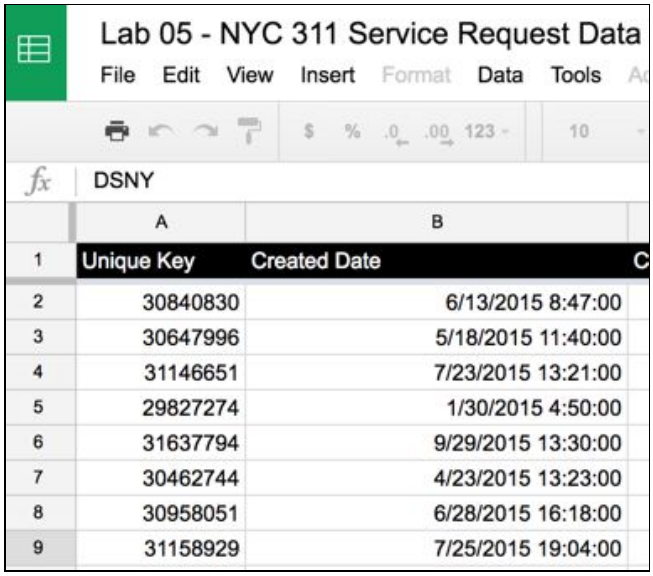
Question 8: How many TRUEs and how many FALSEs are generated in the new column?

Part 4: Pivot Table

Like with Roll-Ups, Pivot tables help you understand large datasets by aggregating and juxtaposing information about particular groupings.

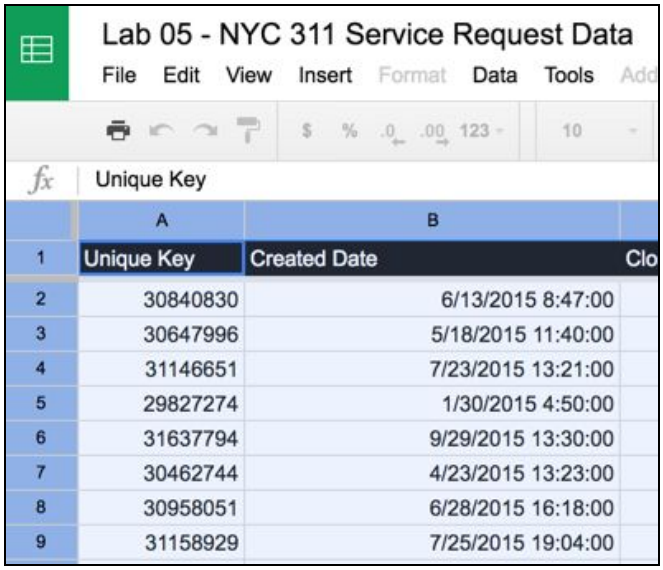
Let's learn the basics of creating a Pivot Table in GoogleSheets.

In general, you want to create your Pivot Table using the entire dataset in your sheet. The most efficient way to make sure you do this is to click the white empty cell above the 1 in the first column, and below the Fx, to the left of the A in the second column.



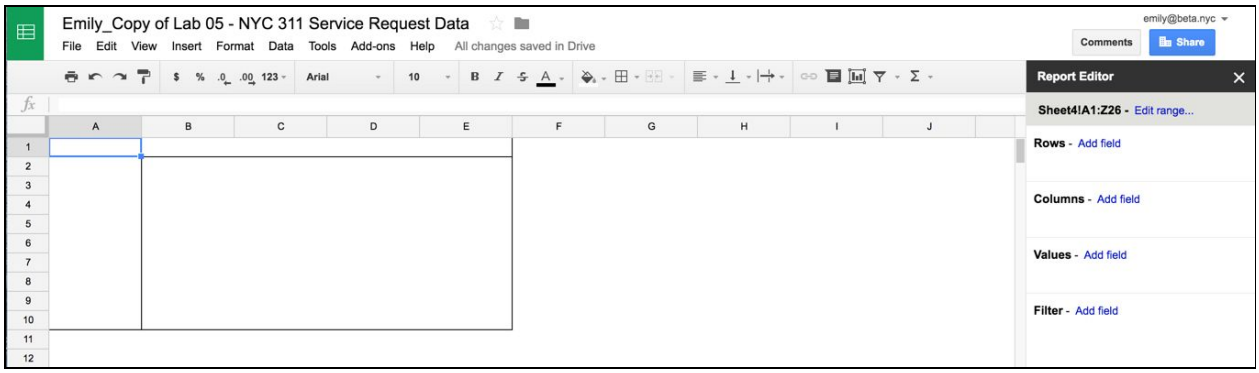
	A	B	
1	Unique Key	Created Date	C
2	30840830	6/13/2015 8:47:00	
3	30647996	5/18/2015 11:40:00	
4	31146651	7/23/2015 13:21:00	
5	29827274	1/30/2015 4:50:00	
6	31637794	9/29/2015 13:30:00	
7	30462744	4/23/2015 13:23:00	
8	30958051	6/28/2015 16:18:00	
9	31158929	7/25/2015 19:04:00	

Then, the entire sheet will become highlighted, and all the data selected.



	A	B	
1	Unique Key	Created Date	Clo
2	30840830	6/13/2015 8:47:00	
3	30647996	5/18/2015 11:40:00	
4	31146651	7/23/2015 13:21:00	
5	29827274	1/30/2015 4:50:00	
6	31637794	9/29/2015 13:30:00	
7	30462744	4/23/2015 13:23:00	
8	30958051	6/28/2015 16:18:00	
9	31158929	7/25/2015 19:04:00	

In the Data menu, click Pivot Table. This will open up a new sheet called Pivot Table X. It will contain an empty table on the left side, and a “Report Editor “on the right. Note, the range of your data appears just beneath Report Editor. You can always edit this if need be.



There are four topics we will cover about Pivot Tables in Google Sheets: Rows, Sort, Values, and Filter.

Rows: Add field

In Pivot Tables, Click “Add field” to “**Rows**” *to select a field to* ***group your data by*****.** This is essentially comparable to choosing a Grouping Column in Socrata (despite the seemingly paradoxical fact that you are within Rows).

All unique entries in that grouping field will then populate the first column, A. You can order them Ascending or Descending if you wish.

Now, turn your attention to:

Values: Add field

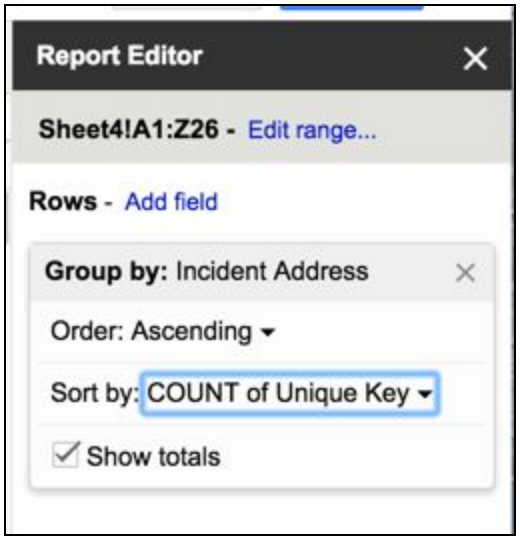
This is how we can **count the number of results per row** displayed in the first column. It is similar to the Roll-Up function in Socrata.

Click “Add field” and choose Unique Key. Then, SUMMARIZE by: COUNT (the default is usually SUM).

Now, in Column B, you will see a Count of all the incidences per row in Column A.

Sort by:

Now that you have a second column containing the Count of incidences per row, you may wish to go back to the box for the field you grouped by, and change the **Sort by** field to COUNT of Unique Key, to see the results in the Pivot Table organized according to COUNT.



Just as we added additional grouping fields in Socrata--drilling-down the data--we can add more fields in Pivot Table in much the same way, by clicking Add Field adjacent to **Rows.**

Filter:

To Filter within the Pivot Table frame, click **Filter**: Add Field, and select the field you wish to filter by. You can check / uncheck to create a filtered Pivot Table.

Part 5: Data Analysis of the CB Survey

Using the skills you learned through the lecture and the lab, go ahead and answer the following questions. Most of them have hints on methods below the question.

- Community Board Websites
    - How many Community Boards have websites?
        - (Filter contents to NO WEBSITE or No, and Count)
    - How many are using http vs https?
        - (Highlight column, Find “https”)
    - Make a pivot table of content management systems and count how many are using them.
        - (Pivot table with COUNTUNIQUE of URLs)
    - Make a pivot table of website metric systems and count how many are using each?
        - (Pivot table with COUNTUNIQUE of URLs)
    - How many have a mobile responsive theme?
    - How many DO NOT have their office address posted to their website
        - (Filter and Count)
    - How many CBs have their address on their website but don’t have office hours posted?
        - (Pivot table, or Filter and Count)
    - How many have an about page
        - (Filter and Count)
    - How many have Board Members listed
        - (Filter and Count)
    - Count how many have Committees & Task Force
    - Count how many have a Calendars
    - Count how many have By laws
    - How many have board meeting presentations?
    - Is Twitter on their homepage?
        - (Pivot or Filter/Count)
    - Is Facebook on their homepage?
        - (Pivot or Filter/Count)
    - Is there a mailing list sign up?
        - (Pivot or Filter/Count)
- Facebook
    - How many Community Boards are on facebook?
        - (Pivot or Filter/Count)
    - How many Manhattan Community Boards are using Facebook?
    - How many are using which type of page?

        - (Pivot)
    - How many have their about section completed?
        - Filter/Count Yes
    - How many engage with their constituents?
    - How many post meeting notices?
- Twitter
    - How many are using twitter?
        - (Filter out the blanks, and Count numbers)
    - How many Manhattan Community Boards are using Twitter?
    - Out of all of the CBs, how many have their profile section completed?
    - How many Community Boards have more than 500 followers?

Part 6: Complex Data Analyses of CB 311 Data

Open the final dataset in the Lab 5 folder to answer the following questions. Primarily you will be using the Pivot Table function extensively in the last part of this lab. Consider yourself an expert Pivot Tabler….in advance :)

Making a copy of NYC 311 Data:

1. Open up this Google Sheet.
2. Make a copy by going to “File” > “Make a copy”
3. Save it to your notebook.

Notes on pivot tables in Google Sheets:

- To make a pivot table in Google Sheets select “Data” > “Pivot Table”
- It might take a minute or two for the Google Sheet to index
- Make sure your range is all cells in the appropriate sheet = “'NYC 311 Service Request Data'!A1:Z38151”
- **NOTE**: IF YOUR PIVOT TABLE DOESN'T SHOW UP
    - TRY SCROLLING DOWN

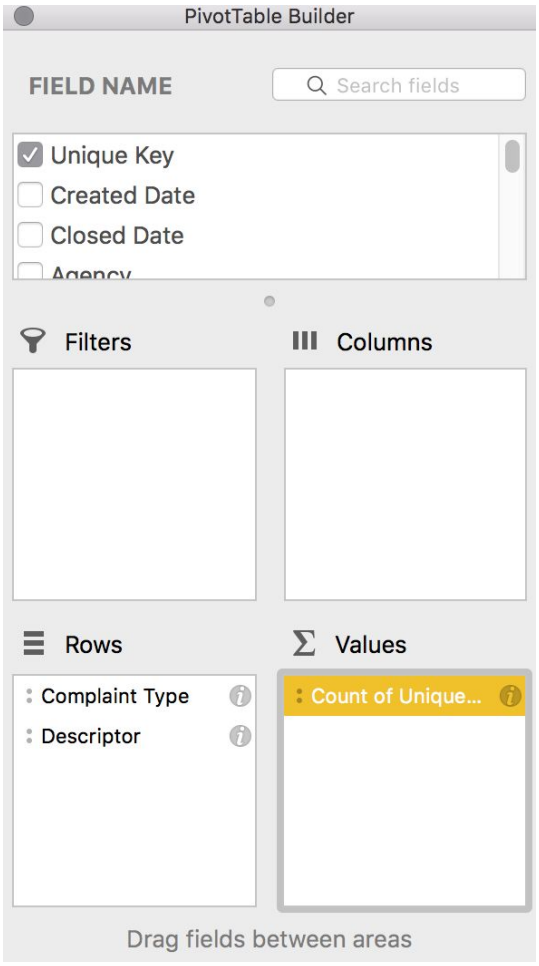
Answer the following questions. As with most data analysis, there are usually multiple ways to do things. Don’t be too concerned about choosing the right way, but more be certain that your way is legitimate, and not a work-around.

Based on the data provided, what is the date range of the data?

Based on the data provided, what Community Board is this data from?

Make a pivot table of Complaint Type, description, and Count Unique key

- What are the top 5 Complaint Types and what are their counts.
- Under the complaint type with the highest count, what are the top 5 descriptors, and what are their numbers?



- Make a pivot table with the following rows:
    - Descriptor
    - Location
    - Incident Address
- Values:
    - Count of Unique Key
- Filter Descriptor by “ENTIRE BUILDING”
- Questions:
    - What is the address with the most number of Heat / Hot Water, Entire Building service requests?
    - What is its lat, long?
    - Go to google street view, and go back to 2007, and take a screen shot and post it in the lab.

    - Using Building’s Department BISWEB, answer the following questions:

        - What is the building’s Tax Block number?
        - What is the building’s Tax Lot number?
        - What is the building’s BIN?
        - How many complaints does this building have? How many complaints are open?
        - What is the oldest complain?
        - What is the most frequent complaint?
        - How many open violations does this building have?

    - Using the City Register’s ACRIS, answer the following:
        - Who is the current property owner?
        - How is the current mortgage lien?

Looking up all NYC 311 complaints for one address:

- Filter the original data to this single Location: (40.776801210358286, -73.9525168463971).
- Copy and paste this filtered data into a new sheet.
- What are the total number of service requests for this location in 2015?
- Make a pivot table with the following
    - rows:
        - Complaint Type
        - Descriptor
    - Values:
        - Unique Key, Count

- What are the other 311 service requests for this location?

Looking up all addresses with one particular service request

- Filter the original data to all requests whose descriptor involves **"JANITOR/SUPER"**
- How many are there?

To answer the following questions you'll need to make a pivot table.

- List all the address in the filtered dataset:
    - FInd the rogue address in the list, and tell us what part of the City it is in. You might use Google Maps.
- Which address has the most service complaints?
- How many of these service requests are still open?
- How many of service requests led to violations?

Pivot Table Key

- Rows:
    - Incident Address
    - Descriptor
    - Resolution Description
    - Created Date
    - Closed Date
- Value
    - Unique Key, Count

- Using BISWEB, answer the following questions:
    - In 2015, how many times did the DOB inspect the complaints of the building?
    - What are the complaints?

Deliverables:

Questions 1 - 8.

All bulleted questions listed in Part 6.

Concluding Remarks:

Congratulations, you have just increased your analytical skills exponentially! You should now feel comfortable diving into other spreadsheet software knowing that functions are relatively straight-forward to learn and make use of. Pivot tables in particular are key to a ton of data analysis.

Civic Innovation Fellowship Y3 Boot Camp
Lab # 06: Mapping & Analysis in Carto
Due: November 4, 2016

Youtube video of lecture: http://bit.ly/2jlgWPt
Link to the slides: http://bit.ly/2jv4qRz

Mapping and Analysis on the Cloud with Carto:

This lab is an introduction to web-based mapping and analysis with **Carto**. Biologists originally founded the company, which today has headquarters in both Madrid, Spain, and Brooklyn, New York. Some popular GIS software is for the desktop, like ArcGIS and QGIS, while Carto is on the cloud. As with Socrata and Google Sheets, it opens in your browser. For web-mapping, it is considered one of the best, most versatile products available.

In this lab, you will learn how to **make maps** out of some of the datasets you are already working with in tabular form alone, as well as an opportunity to explore the specific analytical capabilities of Carto. As with Socrata and Google Sheets, you will learn some of the built-in functions of Carto, including **filters**, **merges**, and making **categorical** and **choropleth** maps.

Just like last week, we have pre-loaded some datasets in the lab folder that you will work with for the duration of the lab. We may ask you to go fetch some the data yourself too, though. Have fun!

Lab Objectives:

- **What is a map? What is spatial data?**
- **Become familiar with the Carto environment**
- **Mapping 311 data**
- **Mapping Census data**
- **NOTE: You had to understand the work from previous weeks in order to move on to mapping this data with confidence and conceptual assurance.**
- **Learn how to use the following functions for spatial analysis:**
    - **Filter**
    - **Table Merge**
    - **Categorical map**
    - **Choropleth map**

Part 1: What is a Map? What is Spatial data?

Part 2: The Carto Environment--setting up an account

Part 3: Mapping 311 Data--importing CSVs into Carto

Part 4: Categorical maps

Part 5: Filters in Carto

Part 6: Mapping Census Data--importing Shapefiles and Demographic data

Part 7: Table merges in Carto

Part 8: Choropleth maps

Part 1: What is a Map? What is Spatial data?

Map:

A map is a **representation**, usually on a flat surface, of a whole or part of **an area**. The job of a **map** is to show or describe spatial relationships of specific features that the **map** aims to represent. There are many different types of **maps** that attempt to represent specific things.

(Source: http://geography.about.com/od/understandmaps/a/whatisamap.htm).

I love this definition, it's flexible and open. Note the terms:

1. "representation"
2. "usually"
3. "aims to"
4. "attempt to"

Spatial data:

Spatial data are data that have a spatial component and contain (or **store**) information about that place. It means that data are connected to a place in the Earth.

Go back to the Lab 5 data folder, and answer the following question:

1. **Out of the four pre-loaded datasets, which ones can be considered spatial data? (For the Social Media survey data, just consider the Website sheet). Justify your answer.**

2. **Is the following data spatial data?**

Vendor	# Booths in NYC	Seasons	Apple types	Net sales in 2016
Happy Orchard	**17**	**Fall, Winter**	**Cortland, Empire, Macs, Honeycrisp**	**$2,567.78**
Red Jacket	**24**	**Winter, Spring, Summer, Fall**	**Cortland, Empire, Macs, Honeycrisp, Golden Eye, Gala**	**$8,986.45**

It should be clear that tables can store spatial data, correct? Spatial data can also come in the form of a **shapefile**.

Shapefile:

A **shapefile** is a popular format for **storing** the geometric location and attribute information of geographic features. Geographic features in a **shapefile** can be represented by points, lines, or polygons (areas).

GIS:

A **Geographic Information System** integrates hardware, software, data, and people to capture, manipulate (Isaac!!!!), analyze and display all forms of geographically-referenced information or spatial data.

A GIS allows see, understand, consult and interpret data to reveal relationships, patterns and trends.

Why use GIS?

- Most of the human activities are linked directly or indirectly to location.
- There is an assumption that up to 80% of all activities is linked to location.
- Statistics are related to territory and it means that they could be linked to a specific location --> they are spatial data
- GIS adds value to the traditional “table based” statistics

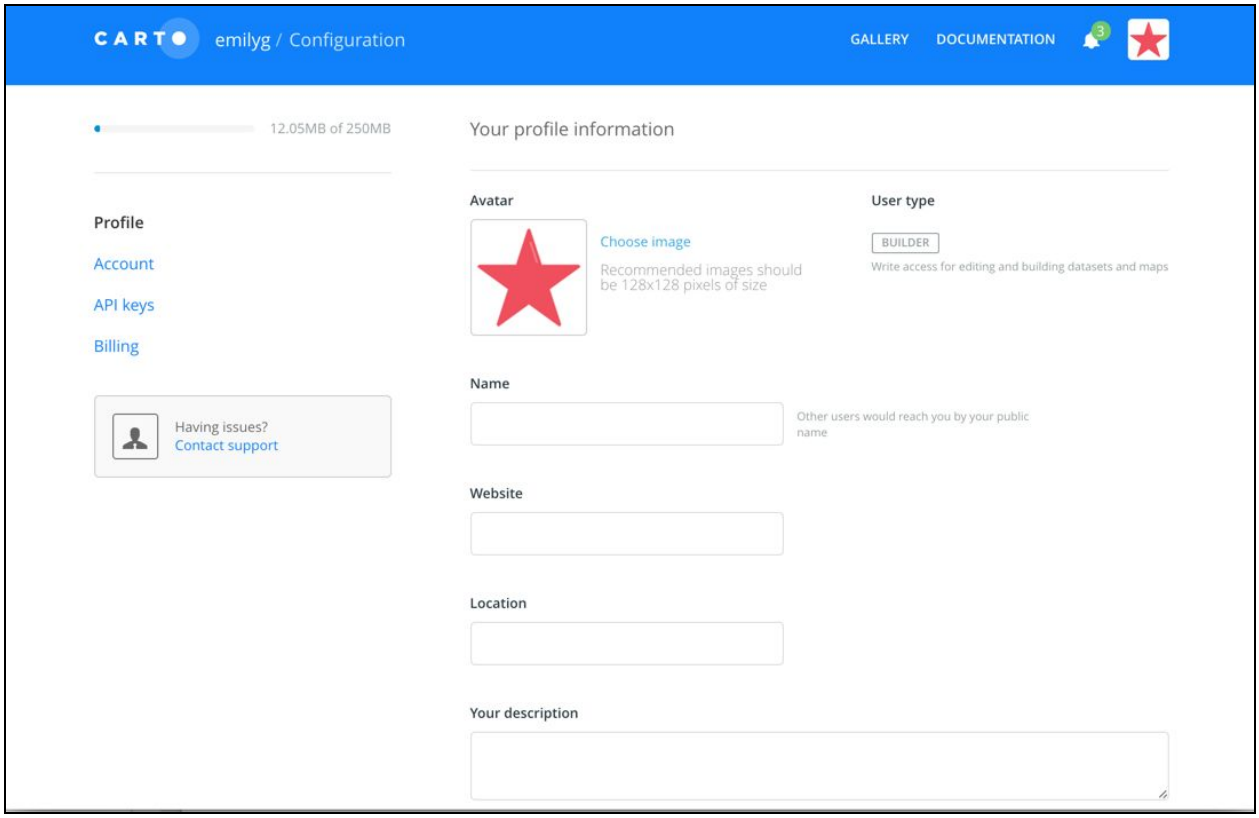
(Source: http://www.cepal.org/celade/noticias/paginas/9/43689/Session10-Brazil.pdf).

Part 2: The Carto Environment--setting up an account

Let’s get started:

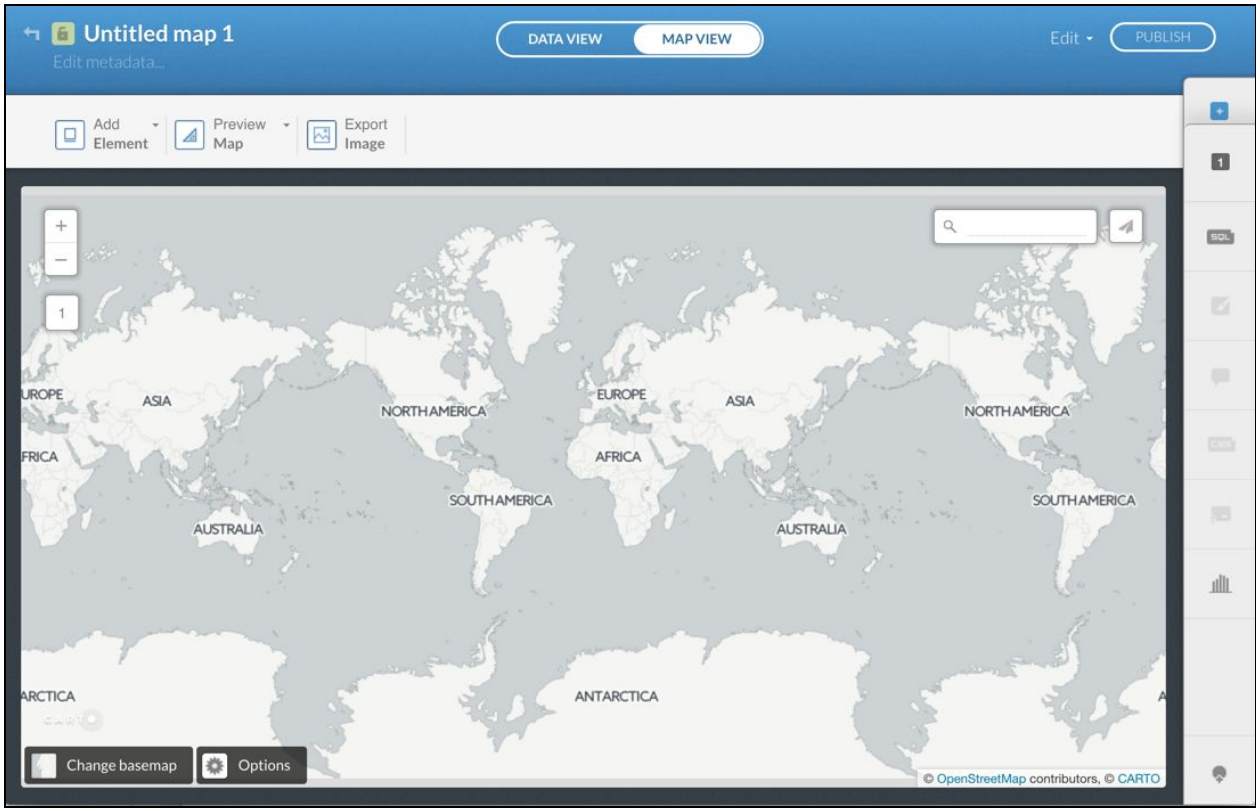
1. Sign up for a Carto account here:

2. Fill out your profile information:



3. Once you have an account and profile, click your username in the upper left corner. Then, click Your Maps next to your username, and now let's create a new, empty map.

4. Skip the step when it says add point, lines, or polygons. Without any data in your map, you should see a map of the world.

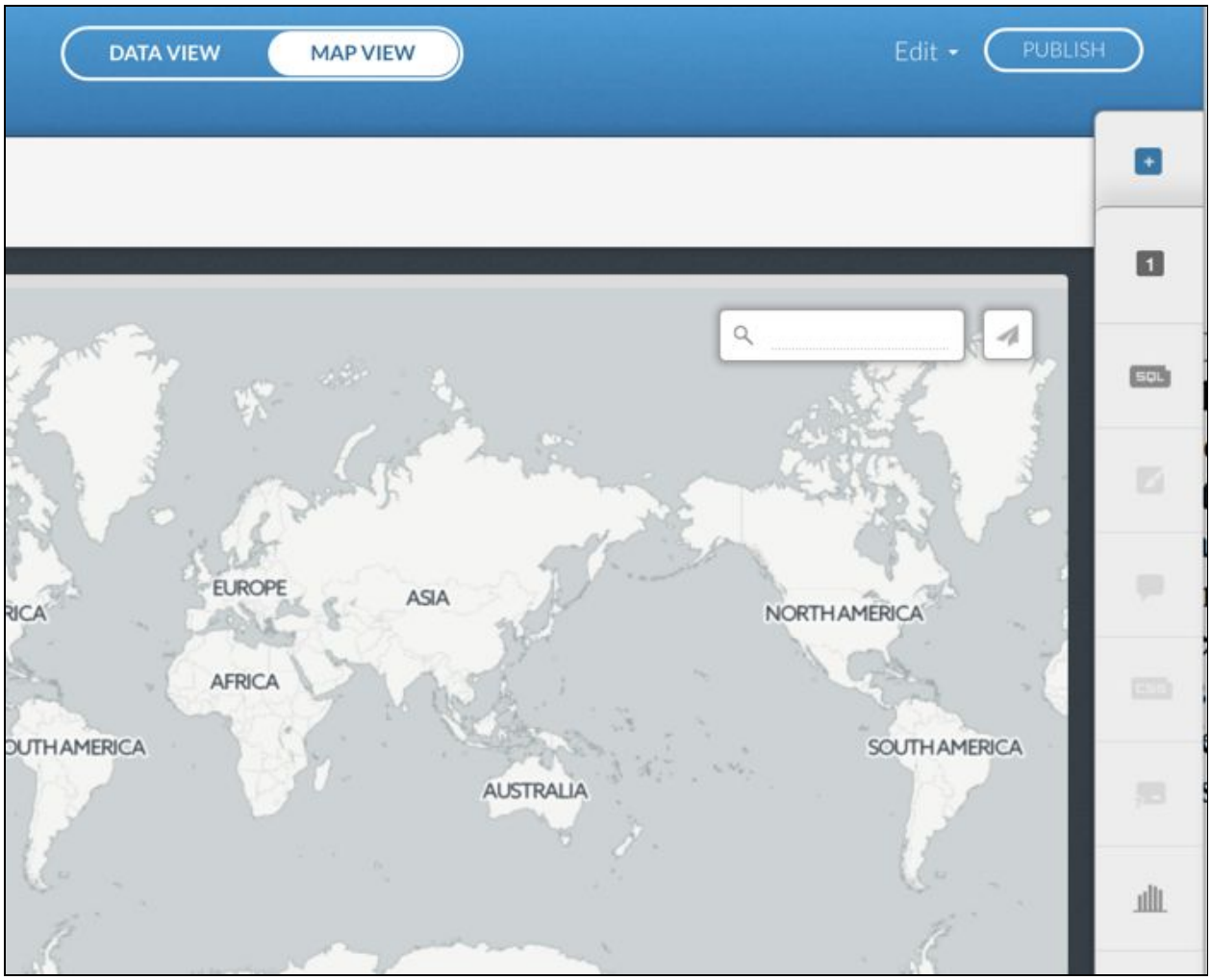


5. Cool. At the top of the screen in the blue band, notice that your map has a title on the left (that you can change simply by clicking it), and in the middle, there are two options: DATA VIEW and MAP VIEW. In the image above, we are in MAP VIEW. Use these to toggle between seeing your data in tabular form and seeing it in map form.

6. Explore Change basemap and Add Element. Okay, enough for now. We'll get to the rest as we go.

Part 3: Mapping 311 Data--importing CSVs into Carto

1. Click the Add Layer icon (+) at the top of the vertical menu on the right.

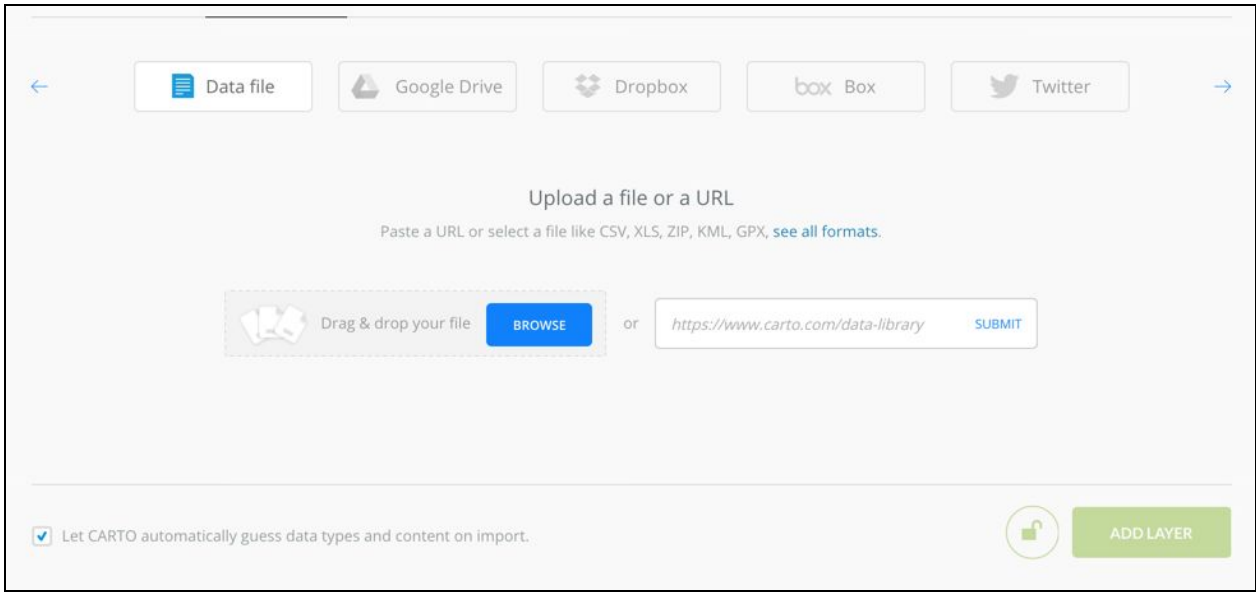


2. Carto can import and map .csvs **if they have latitude and longitude data**, so we are going to add our **CB 8 311 data from 2015** (the same dataset we worked with last week to learn the functions in Google Sheets).

3. Click CONNECT DATASET.

4. **Now you have a few options for importing this dataset.**

5. **Try different ways.**



6. Highlight Google Drive and Search for the dataset within Google Drive. It is called **CB-08-MANHATTAN-2015**. Did you find it? It should be 25 MB. Click CHOOSE and then Add Layer.

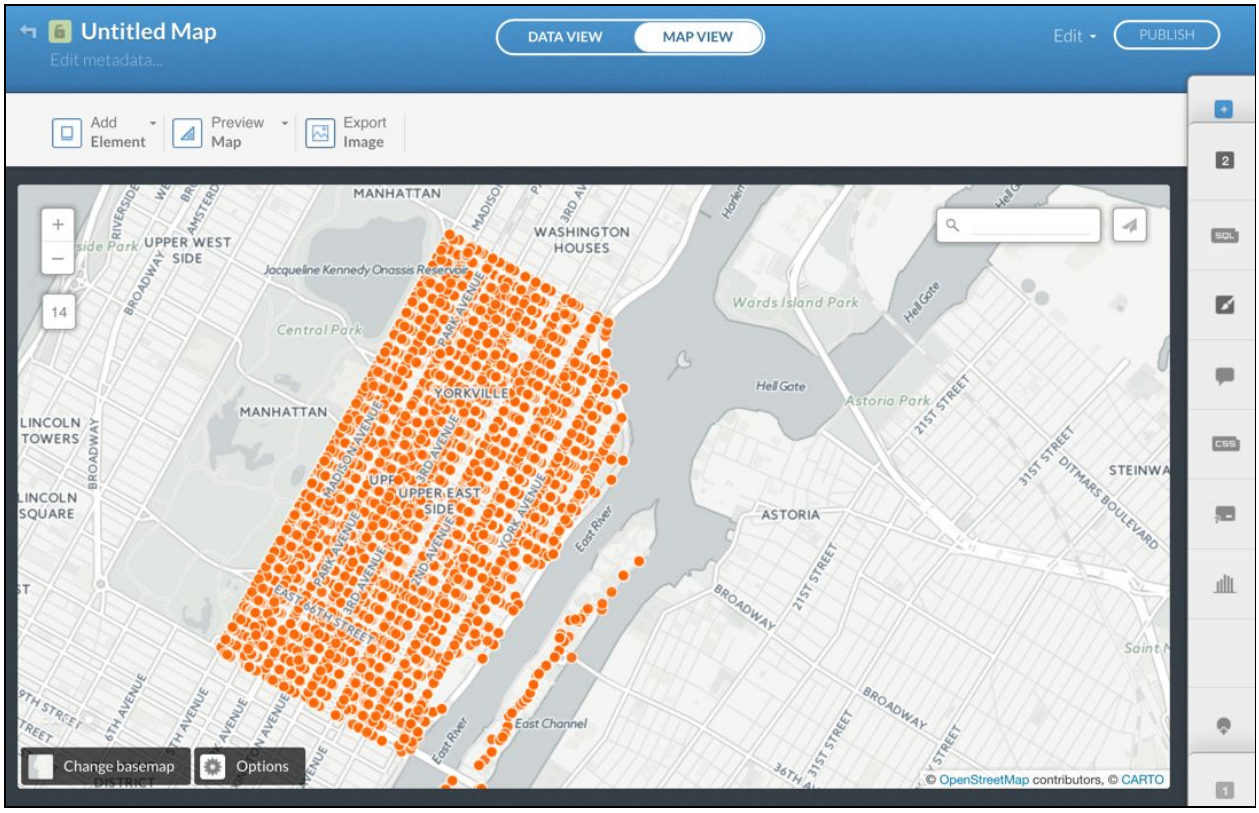
7. Now, import the same data by navigating to the Lab 6 data folder and downloading this dataset to your computer.

8. Then, highlight Data file, and Drag and drop or Browse for the data in your Downloads.

9. Now you may have two copies of the dataset in your Datasets. Carto may have differentiated the name of the file with an underscore 1. Note, Carto also changes the case to lower-case. It should look like this:

10. Click the dataset, and then click Add Layer at the bottom of the screen, and Carto will start the process of importing the data, and then will take you to your map view. Carto RECOGNIZED that the data had lat and long fields, and displayed it in map form accordingly. Does CB 8 fill up with orange dots?

11. It should look like this:



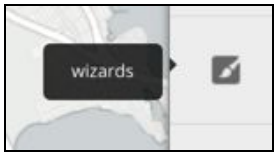
12. Take a screenshot of your map and paste it here.

13. Congrats, you just successfully imported data straight from .csv format into Carto and have a map you can now share. Got it?

Part 4: Categorical maps

1. Click the 1 in the frame on the right side, to start working with your dataset.

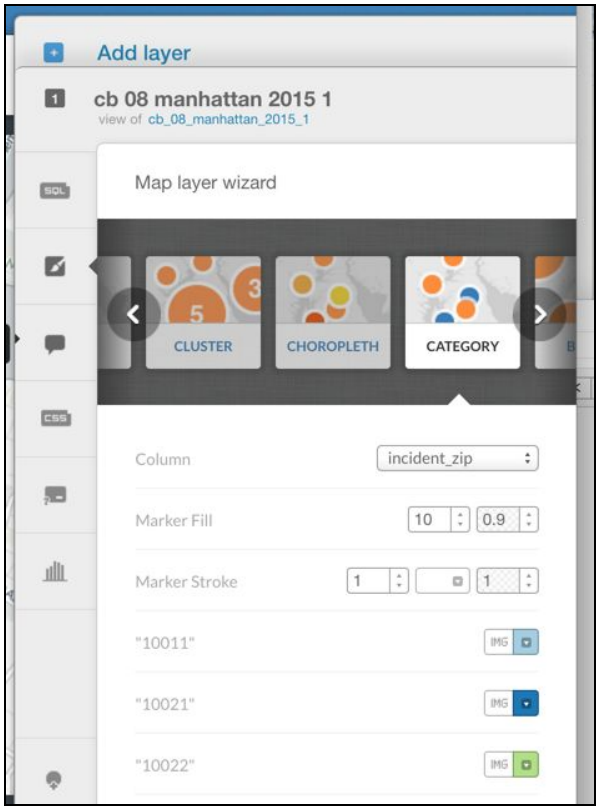
2. Open the Map Wizard--you'll spend a lot of time working with the Wizard.

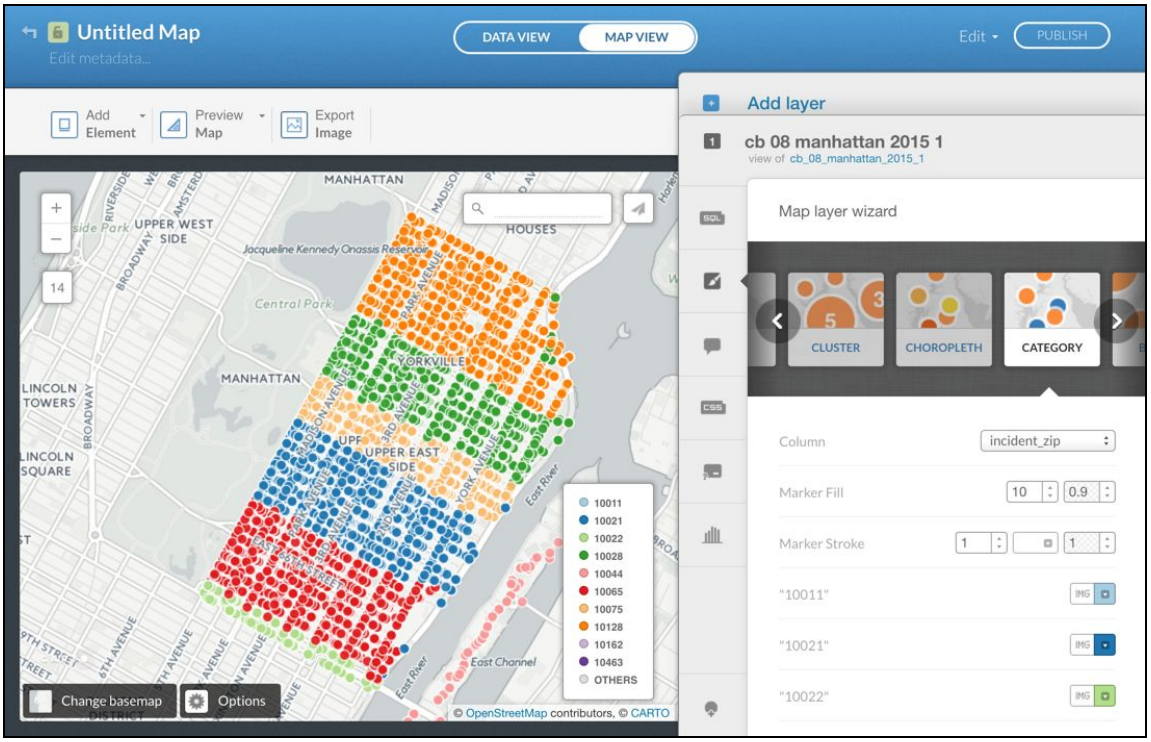


3. In Map Wizard, from "Simple," click the right arrow to scroll over to and click "Category."

4. Now, we are going to categorize this map by Zip Code. This is so we are able to group and view the data in that field colored by the different entries in that field.

5. If you have not already, choose “incident_zip” from the drop down menu next to Column.





Question 3: Making a categorical map in Carto is actually similar to Grouping by in Socrata/ Google Sheets. Think about the similarities and differences between grouping the data by a particular column in Socrata / Google Sheets, versus how we have just categorized the data in Carto.

First, describe the specific steps you take to "Group By" in Socrata (Hint: Roll-Up)

For example, I might say: To group my data by a particular column in Socrata, I open the Filter and Roll-Up frame on the right side of the dataset. Then, I choose my first Grouping Column. Then, I roll-up the data by Count / Unique Key, etc. etc. This enables me to….you go on from there!

Now, do the same for Google Sheets (Hint: Pivot Table)

Now, do the same for Carto (Hint: Category)

What information do you gain with Socrata / Google Sheets about the data in that column/field, and what information do you gain with Carto about the data in that column/field?

6. You can now take some time to change your **basemap**, and **add elements** to your map to customize it. Take a screenshot of your final map here:

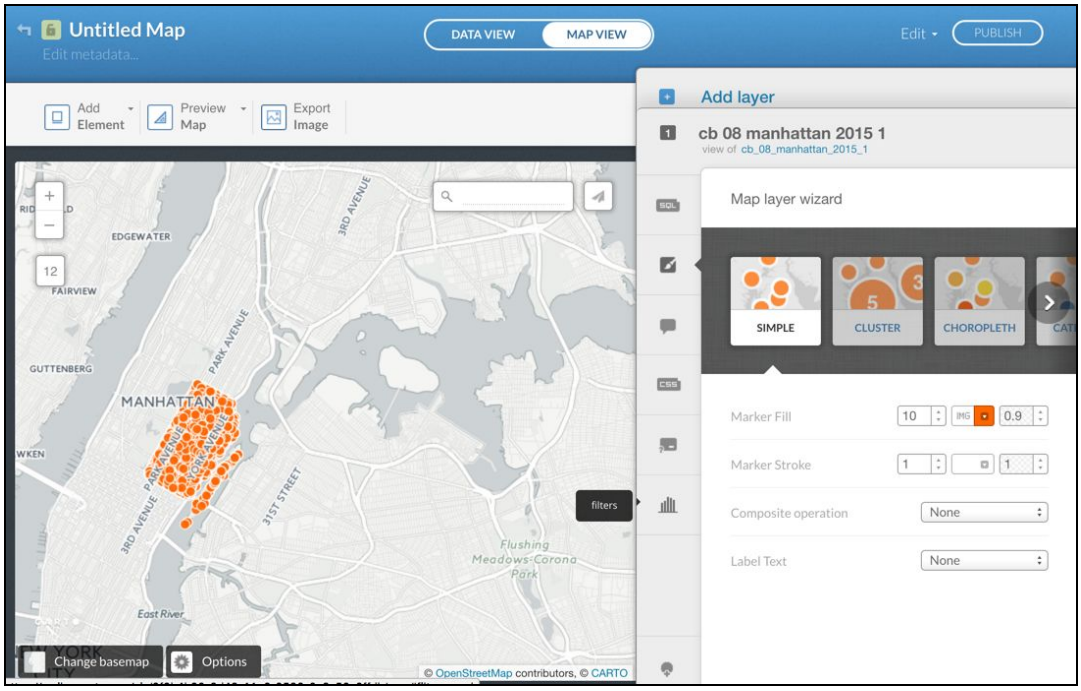
Part 5: Filters in Carto

You **probably** recall doing filters in Socrata and Google Sheets--HA!!!!! Of course you do! Well, let's do filters in Carto, too.

Question 4: Think of a filter you'd like to apply to this dataset. What is it? Tell us why you're interested in seeing just that data.

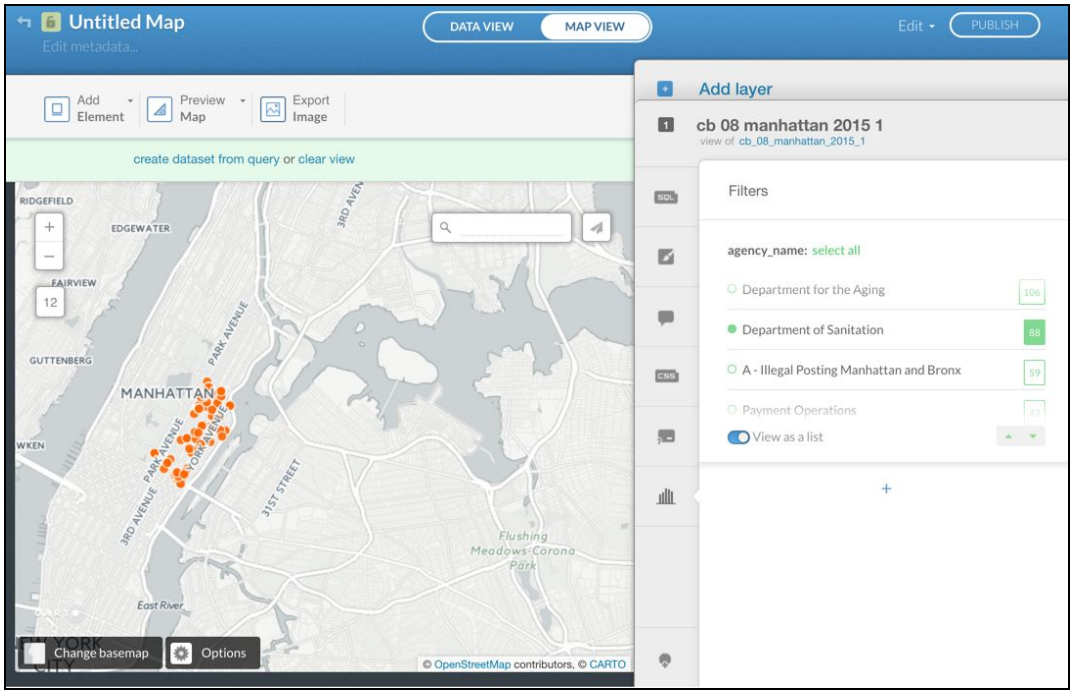
For example: I want to filter the data to only HPD complaints within CB 8 because I'm interested in where the housing issues are located. To apply this filter, I would need to filter by column agency or agency_name, and select HPD or Department of Housing Preservation and Development, respectively. Pick a different filter, and answer in your own words!

1. From your categorized map by Zip, now go back to a simple map where all your points are orange (by default). Open the Map Wizard, and click Simple. It should look like this:



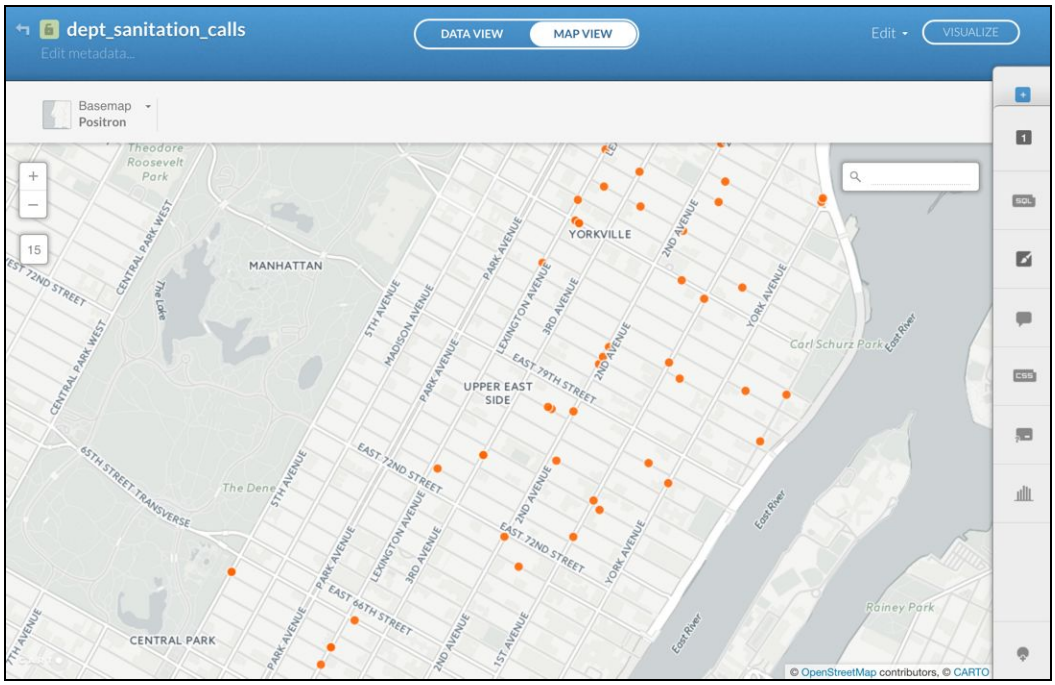
2. Now, we are going to apply filters to the data.

3. Say we wanted to filter our data to only the 311 calls directed to the **Department of Sanitation**. Click the filter tab shown in the image above, and let's filter by the column, agency_name, and then **Select Only**, Department of Sanitation. Carto will load this request, and generate a new map that should look like this:



4. Take a screenshot and paste it here. Remember your map Title should have your name in it!

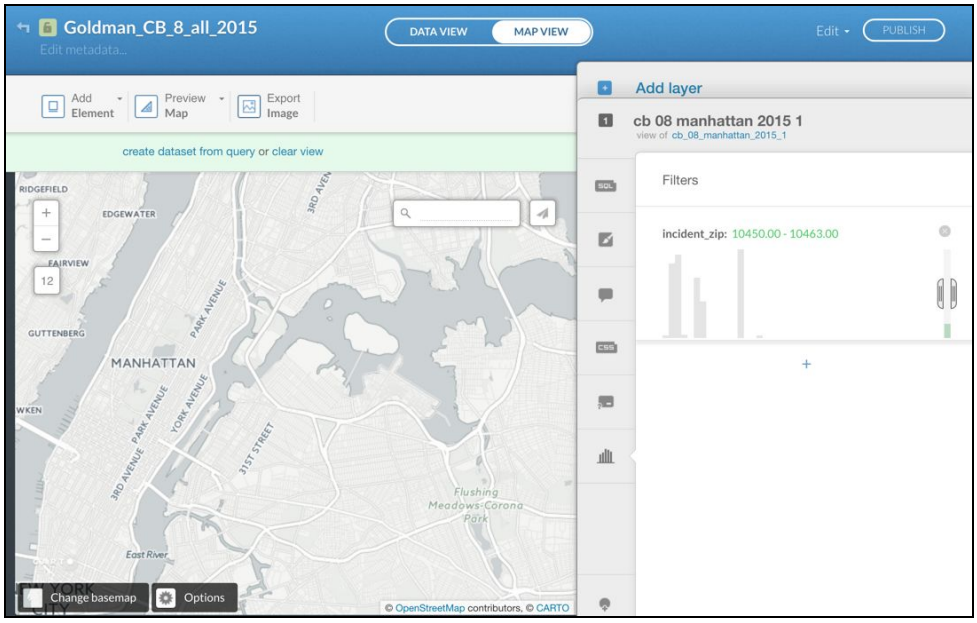
5. At the top of the screen, click “create dataset from query,” to save this filtered dataset. Notice in DATA VIEW, the agency name is exclusively Department of Sanitation. Give the dataset a new name to reflect this. When you toggle back to MAP VIEW, you should see this:



6. Click VISUALIZE in the upper right corner so you can then share and publish your map. In the next screen, click Publish, and paste the link to your map here:

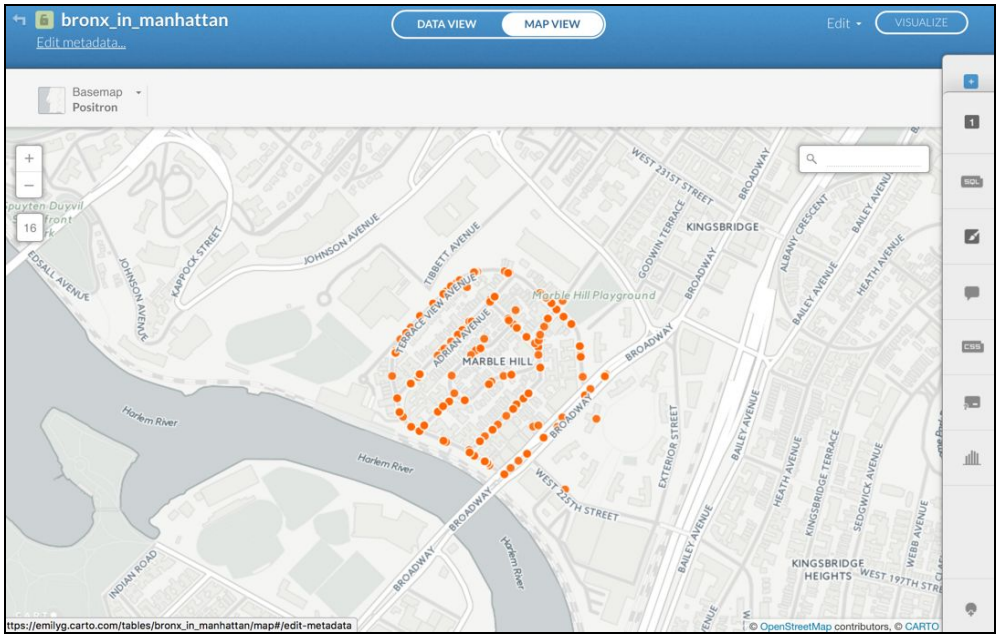
7. Now, let’s go back to our original map to do a different filter, one based on a numeric column. To go to your maps, click the back arrow in the left corner of the blue band at the top of the screen.

8. In your original map with all the 311 data, let’s do another filter by Zip. Open the filter tab, and choose incident_zip as your column. Notice what you see next. Filter to the “rogue” zip in the Bronx. Should look like this:



9. Scroll up to the Bronx to see your calls.

10. Click create dataset from query, rename to reflect these are the calls from the rogue zip code only (and include your first or last name in the title), and take a screenshot of your filtered dataset, and paste it below the following image. Should look like this:



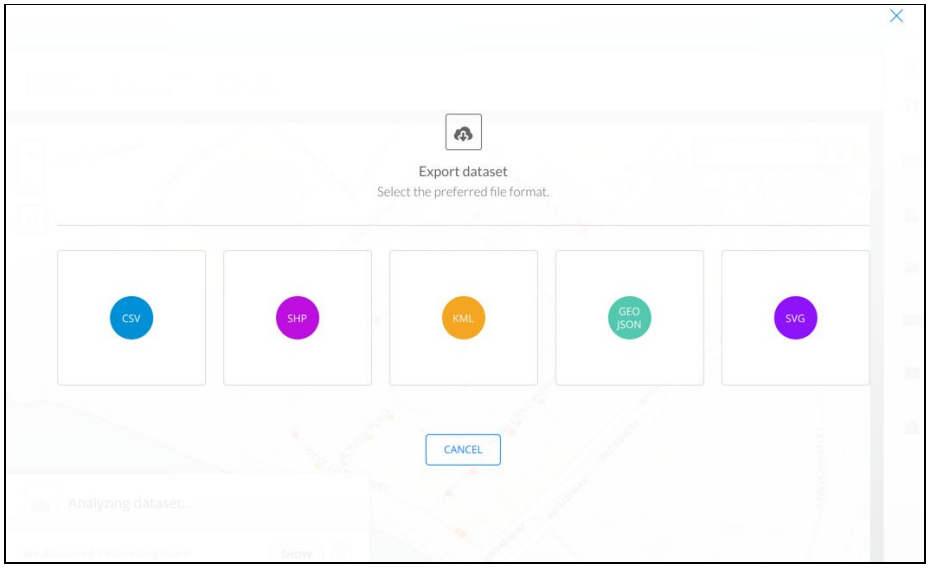
11. Click VISUALIZE in the upper right to turn this data into a map. On the next screen, you'll see PUBLISH in the same corner. Now you know you have a map to share with the world.

12. Okay, you have just filtered the data in two ways and made two new maps. The first was filtered to only Department of Sanitation calls, and the second was filtered to only calls from the zip codes in the Bronx.

Question 5: What do you think about the Filter function in Carto, compared to the Filter function in other resources you've worked with?

Deliverables:

1. Combining Filters and Category Maps: Take your map of calls in the Bronx, and categorize it according to Agency_name. Notice the legend. Take a screenshot of the map here.
2. Adding another filter, Exporting to .csv, Counting rows: Now, filter to only HPD calls. Click create dataset from query, Rename it, Visualize it, and then under the Edit tab, Export layer to .csv.



3. Open the csv., and report how many rows are in this dataset?

Part 6: Mapping Census Data--importing Shapefiles

1. Let's refresh our memory about what a shapefile is:

Shapefile:

A **shapefile** is a popular format for **storing** the geometric location and attribute information of geographic features. Geographic features in a **shapefile** can be represented by points, lines, or polygons (areas).

Please note: a shapefile actually contains 5-6 files within it. In Carto, we need to compress the data into a single .zip file before importing.

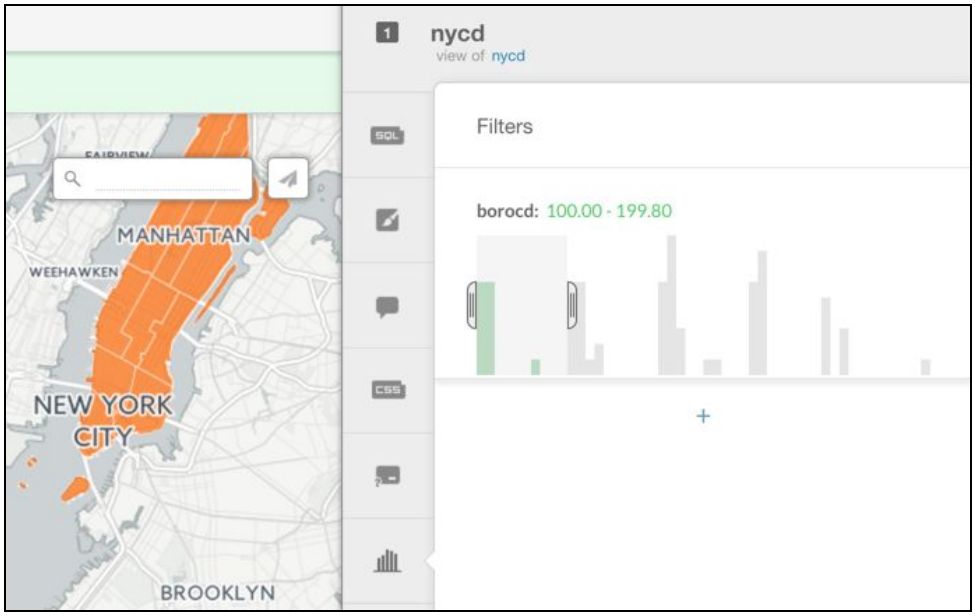
2. In the Lab 6 data folder, find the shapefile for NYC Community Districts compressed to a .zip file. It's called nycd_16c.zip.

3. Back in Carto, start a new map, and import the .zip file by connecting to the dataset.

4. Now, let's filter this dataset to just Manhattan CBs.

5. Open the filter tab, and select the borocd column to filter by. This is a numeric field. Determine how to filter only to Manhattan.

Hint: do you remember that CBs all have a leading digit for Borough? Do you remember what Manhattan's Borough code is? It is 1! Use the filter tool to select only those CBs whose borocd is between 100 and 200.



6. Now, click “create dataset from query.”
7. Give it a new title: put your first or last name in it.
8. Click Save
9. Click OK RENAME!
10. Click VISUALIZE

Question 6: How many CBs *should be* in this dataset? How many actually are in this dataset? How did you find out?

11. Now, let’s turn this boring orange map into a categorical map. In Map Wizard, choose Category, and select the borocd Column. Take a screenshot and paste your categorized map here.

Question 7: Critique the resulting map and legend. What do you like about it? How would you change it to make it better represent the information in it?

12. Now, let’s add some Census data! First we will add the Census tract shapefile, which will display the shapes (i.e. the boundaries) of Manhattan’s Census tracts. Later, we’ll add some more Census data about populations in the form of a .csv.

13. Add layer: Connect to the Manhattan_tracts shapefile located in the Lab 6 data folder.

14. Now with both layers in your map, drag the CB layer on top of the tracts later. Also, open the Wizard for the CB layer and adjust the **Polygon Stroke** (i.e. fiddle around with the options!) to increase the thickness of and/or change the color of the CB boundaries to see them better. You can also increase the layer’s transparency with **Polygon Fill** in the Wizard. You may end up with something like this:

15. Zoom to your CB. Approximately how many tracts are in your CB? Discuss what you notice about the alignment/lack of alignment between CB and tract boundaries.

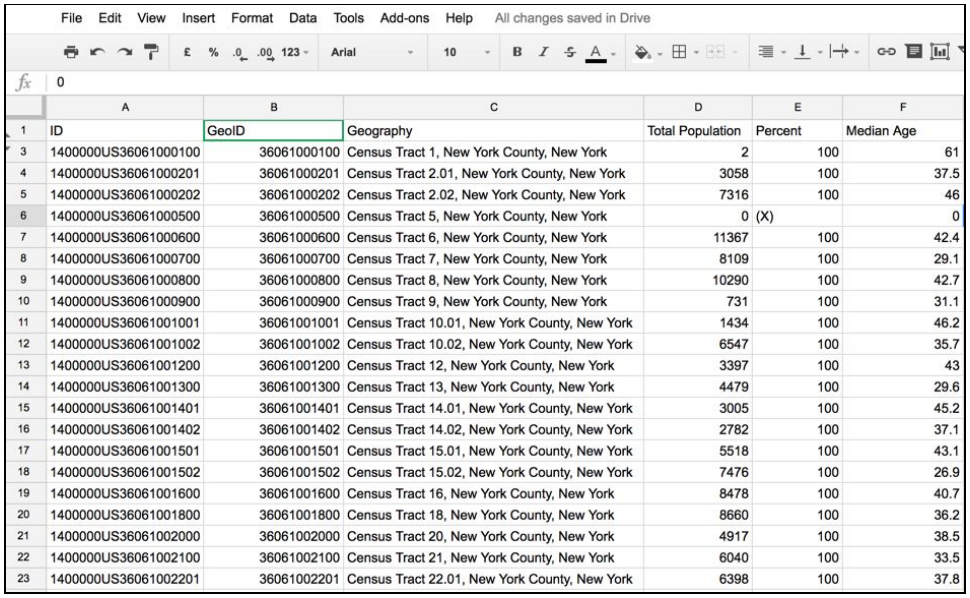
Note: there is a function in Carto (and most GIS software) called **Spatial Join** which enables you to more systematically align these two datasets and count the number of tracts that align with the CB boundaries. We'll come back to this at some point later.

Part 7: Table merges in Carto

Now, let's look at some **demographic data** about Manhattan's Census tracts. These are statistics that describe **population characteristics** aggregated to the Census tract level. We got the data from the U.S. Census Bureau website, using a tool called American FactFinder.

We are going to import the demographic data into Carto, and **merge** the demographic data and the shapefile using a code that uniquely identifies each tract in both datasets—this code is called the GeoID (for **Geo**graphic **Id**entifier--surprise surprise :)).

Note in the following tables the GeoID column from the demographic data, and from the Census tract shapefile data.

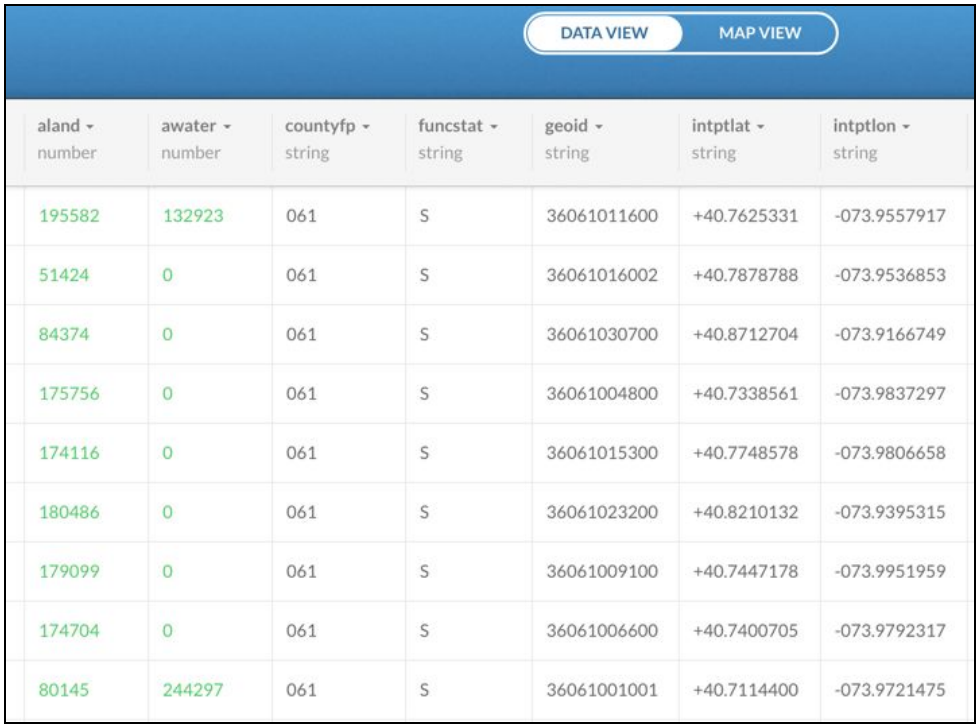


File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

Arial 10

fx 0

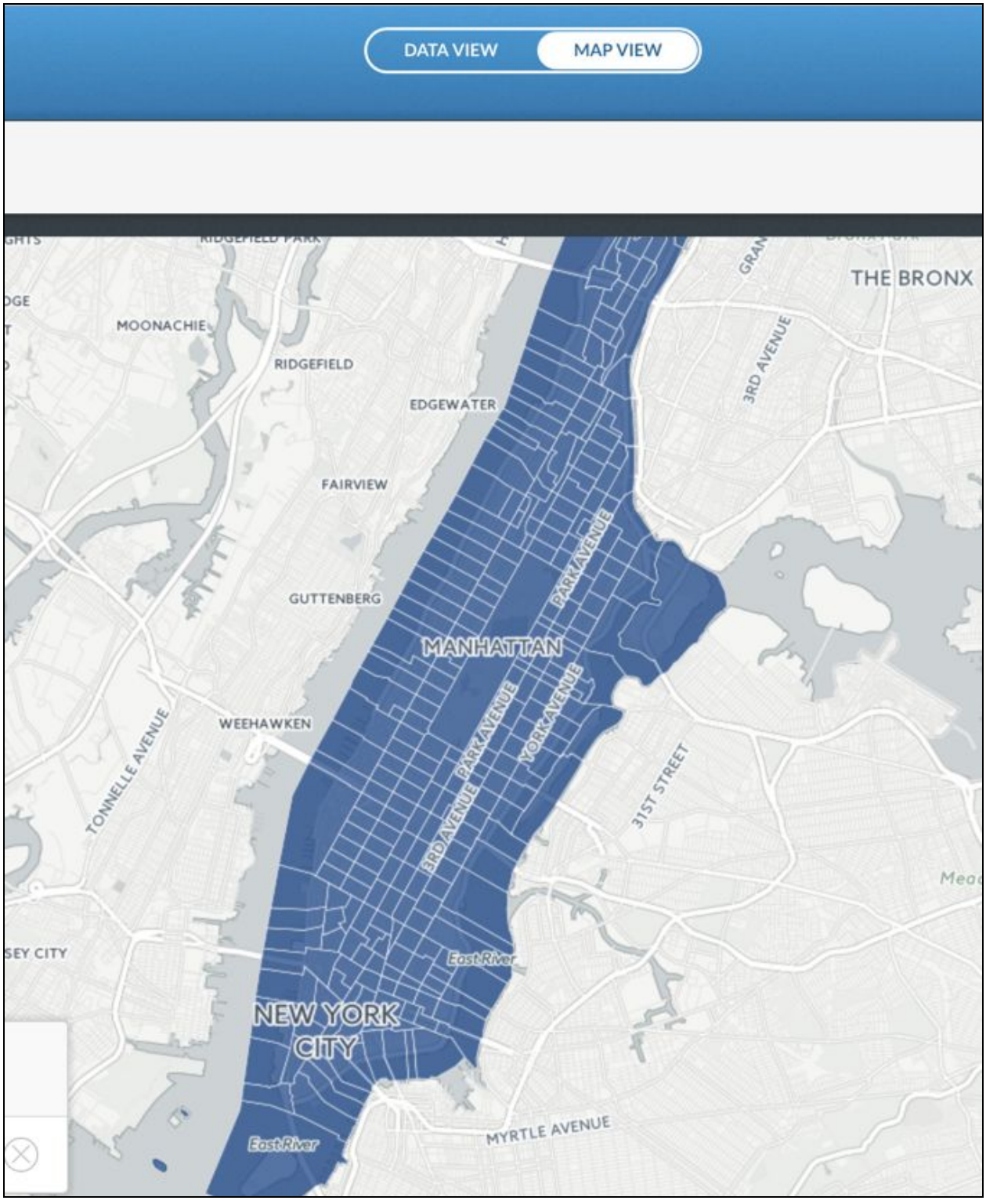
	A	B	C	D	E	F
1	ID	GeoID	Geography	Total Population	Percent	Median Age
3	1400000US36061000100	36061000100	Census Tract 1, New York County, New York	2	100	61
4	1400000US36061000201	36061000201	Census Tract 2.01, New York County, New York	3058	100	37.5
5	1400000US36061000202	36061000202	Census Tract 2.02, New York County, New York	7316	100	46
6	1400000US36061000500	36061000500	Census Tract 5, New York County, New York	0	(X)	0
7	1400000US36061000600	36061000600	Census Tract 6, New York County, New York	11367	100	42.4
8	1400000US36061000700	36061000700	Census Tract 7, New York County, New York	8109	100	29.1
9	1400000US36061000800	36061000800	Census Tract 8, New York County, New York	10290	100	42.7
10	1400000US36061000900	36061000900	Census Tract 9, New York County, New York	731	100	31.1
11	1400000US36061001001	36061001001	Census Tract 10.01, New York County, New York	1434	100	46.2
12	1400000US36061001002	36061001002	Census Tract 10.02, New York County, New York	6547	100	35.7
13	1400000US36061001200	36061001200	Census Tract 12, New York County, New York	3397	100	43
14	1400000US36061001300	36061001300	Census Tract 13, New York County, New York	4479	100	29.6
15	1400000US36061001401	36061001401	Census Tract 14.01, New York County, New York	3005	100	45.2
16	1400000US36061001402	36061001402	Census Tract 14.02, New York County, New York	2782	100	37.1
17	1400000US36061001501	36061001501	Census Tract 15.01, New York County, New York	5518	100	43.1
18	1400000US36061001502	36061001502	Census Tract 15.02, New York County, New York	7476	100	26.9
19	1400000US36061001600	36061001600	Census Tract 16, New York County, New York	8478	100	40.7
20	1400000US36061001800	36061001800	Census Tract 18, New York County, New York	8660	100	36.2
21	1400000US36061002000	36061002000	Census Tract 20, New York County, New York	4917	100	38.5
22	1400000US36061002100	36061002100	Census Tract 21, New York County, New York	6040	100	33.5
23	1400000US36061002201	36061002201	Census Tract 22.01, New York County, New York	6398	100	37.8



DATA VIEW MAP VIEW

aland number	awater number	countyfp string	funcstat string	geoid string	intptlat string	intptlon string
195582	132923	061	S	36061011600	+40.7625331	-073.9557917
51424	0	061	S	36061016002	+40.7878788	-073.9536853
84374	0	061	S	36061030700	+40.8712704	-073.9166749
175756	0	061	S	36061004800	+40.7338561	-073.9837297
174116	0	061	S	36061015300	+40.7748578	-073.9806658
180486	0	061	S	36061023200	+40.8210132	-073.9395315
179099	0	061	S	36061009100	+40.7447178	-073.9951959
174704	0	061	S	36061006600	+40.7400705	-073.9792317
80145	244297	061	S	36061001001	+40.7114400	-073.9721475

Looking at the above data in MAP VIEW, this GeoID uniquely identifies each and every Census tract in Manhattan. Got it?



Question 8: Remember analogies from the SAT? <u>GeoID</u> is to Census tract as _____ is to individual 311 calls; as _____ is to each

Community Board in the Website page of the big Social Media survey? (Hint: what field uniquely identifies each CB?)

16. Navigate to the Population dataset in the Lab 6 data folder.

17. With your map of CBs and Census tracts open, add layer, and connect to the Population dataset, called carto_pop_data_2. You can do this by first downloading it, and then browsing, or dragging and dropping. Or by finding it in your Drive and connecting to it directly.

18. How does the data look in Carto in DATA VIEW? Columns are ordered a bit differently, but otherwise, should be well-preserved, looking something like this:

nycd_copy 1
Edit metadata...

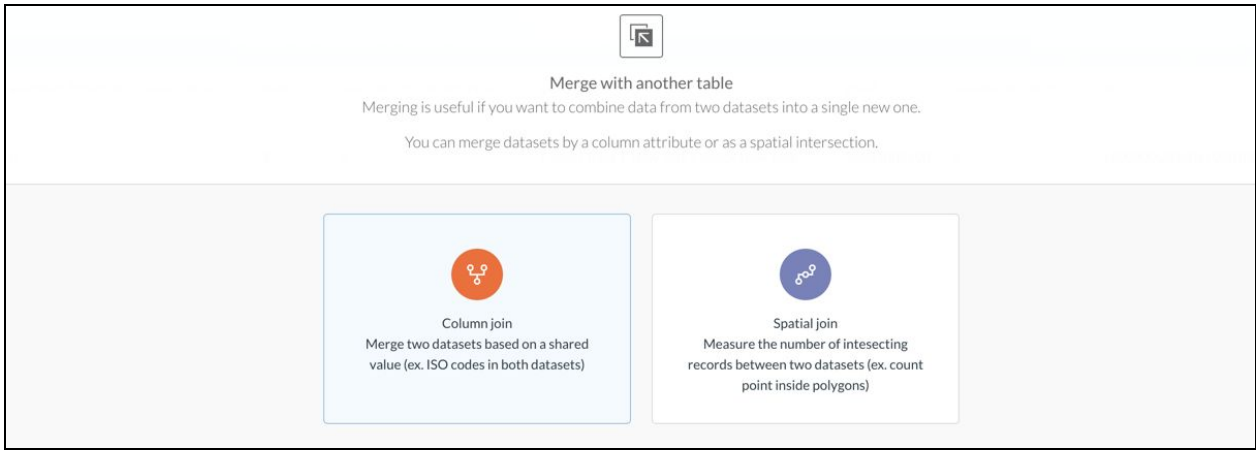
DATA VIEW | MAP VIEW

the_geom GEO geometry	american_indian_and_alaska_native string	asian string	black_or_african_american string	geography string	geoid string	hispanic_or_latino string
null	0	2	0	Census Tract 1, New York County, New York	36061000100	0
null	16	1266	433	Census Tract 2.01, New York County, New York	36061000201	1264
null	41	1648	1118	Census Tract 2.02, New York County, New York	36061000202	2965
null	0	0	0	Census Tract 5, New York County, New York	36061000500	0
null	48	7253	1006	Census Tract 6, New York County, New York	36061000600	2417
null	12	1680	273	Census Tract 7, New York County, New York	36061000700	586
null	1	9069	83	Census Tract 8, New York County, New York	36061000800	215
null	2	88	24	Census Tract 9, New York County, New York	36061000900	50
null	2	93	53	Census Tract 10.01, New York County, New York	36061001001	211
null	74	963	1626	Census Tract 10.02, New York County, New York	36061001002	4200
null	22	476	150	Census Tract 12, New York County, New York	36061001200	1089
null	8	948	188	Census Tract 13, New York County, New York	36061001300	339
null	3	341	107	Census Tract 14.01, New York County, New York	36061001401	410

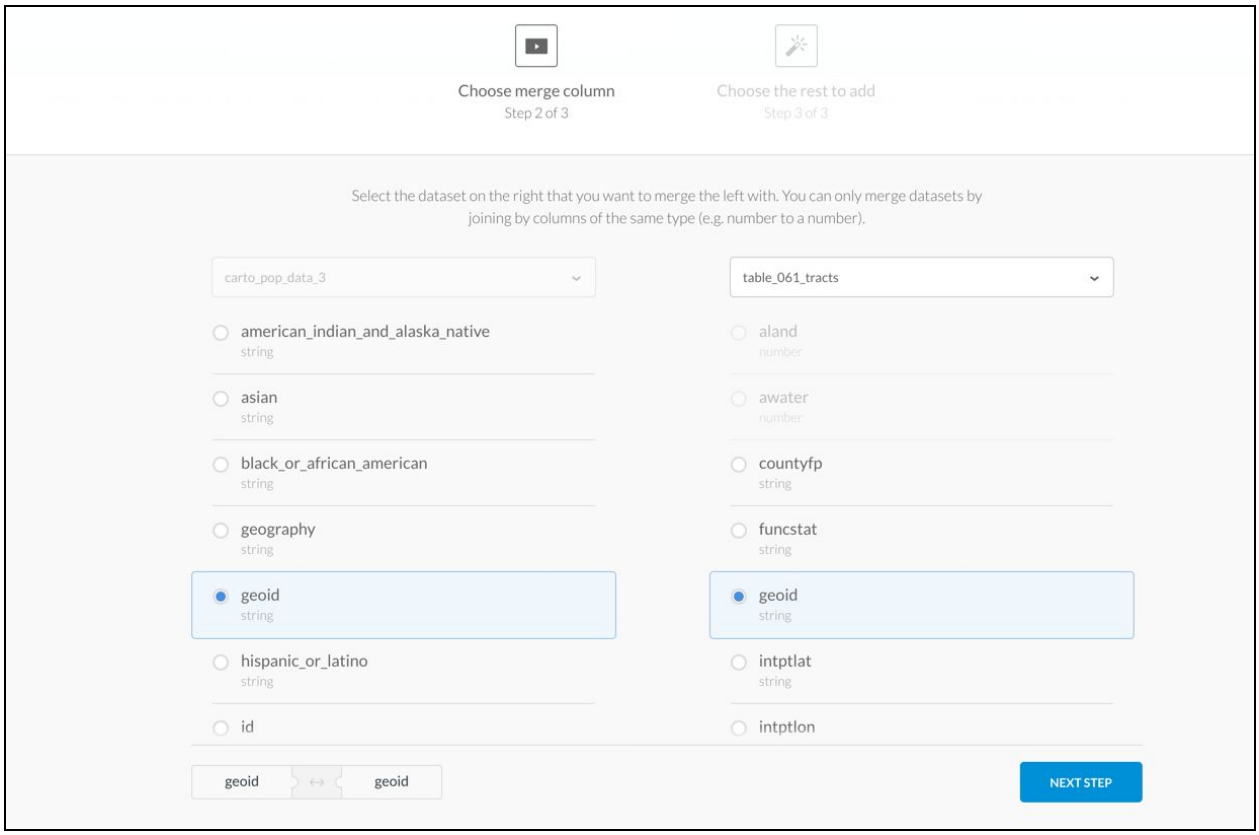
19. Now, let's **merge** the demographic data with the shapes of the tract, eh?

20. Notice the merge icon while in DATA VIEW of the demographic data.

21. Click the icon above for merge, and then in the following screen, click Column join. Read the statement beneath Column join to see what exactly you are doing!



22. In the next screen, you are going to choose two datasets to merge, and pick your column on which to join them. That column is….the unique identifier we discussed earlier….the geoid. It should look like this:



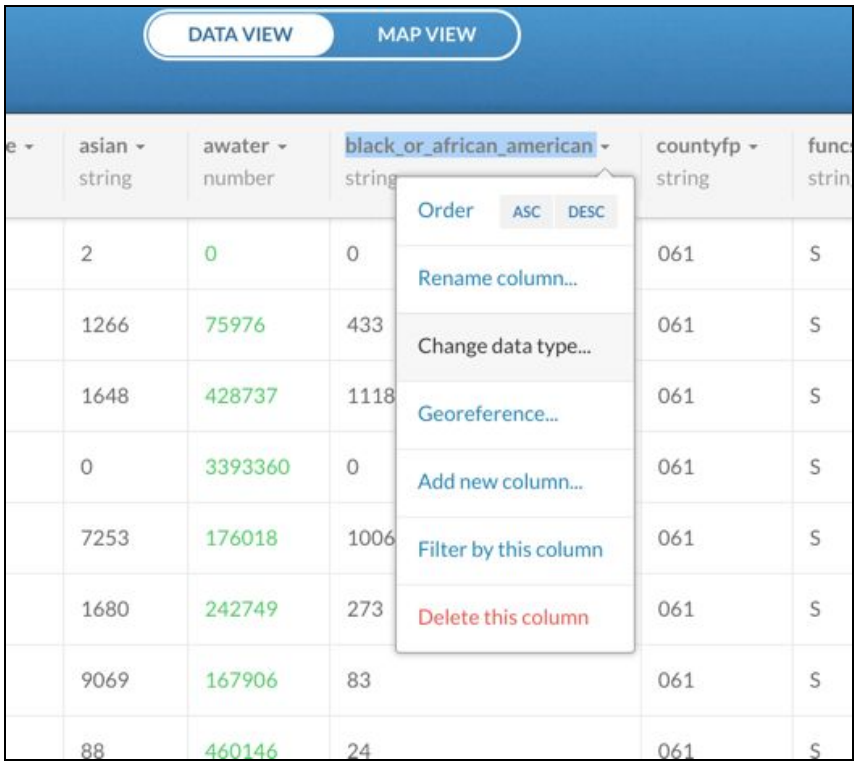
23. Beautiful! After clicking NEXT STEP, accept the defaults, and click Merge.

24. Take a look at your new merged dataset in DATA VIEW. The order of the fields is pretty funky, but oh well, the data is there.

25. Now go to MAP VIEW, and click VISUALIZE. Rename this map, “Demographic map-Manhattan-yourname” (or something like that).

26. Final step before creating our exciting CHOROPLETH maps. In order to make choropleth maps with this data (to be explained in the next section), we need to change the data type of the columns in the demographic data from string (text) to numeric.

27. For every column of **demographic data (meaning total pop, median age, sex, race, hispanic/latino)** all of which are currently of the “string” data type, please change the data type to numeric. Click the field name, and then using the drop down arrow, click…(surprise surprise)...**Change data type**.



28. Then, choose **number** from the drop down list.

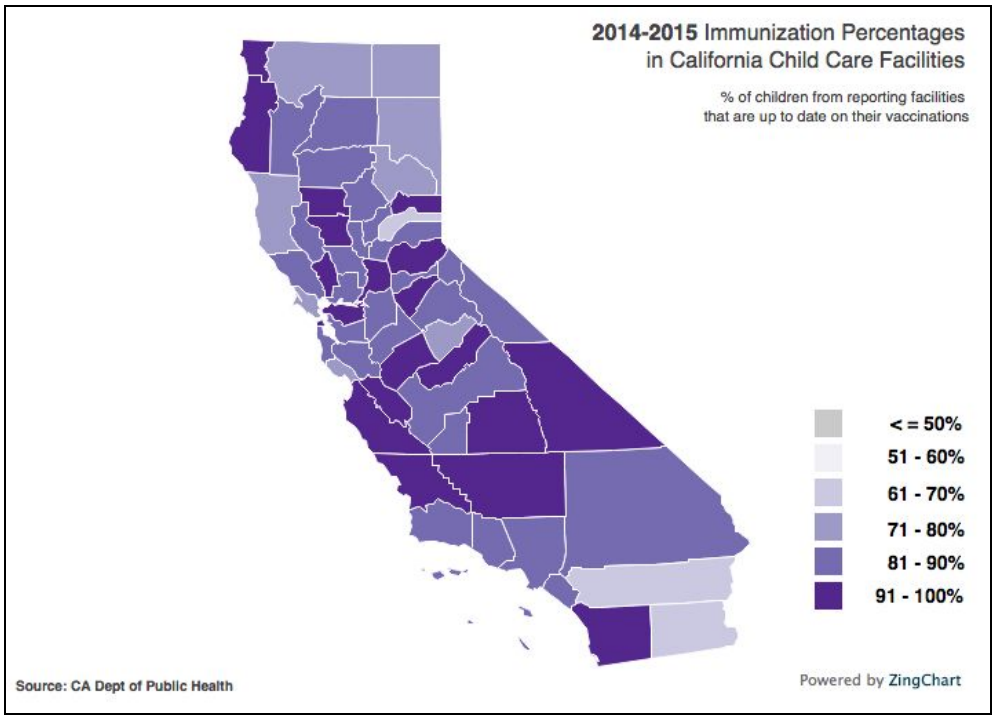
29. Please repeat for all demographic fields in this dataset--tedious perhaps, but it has to be done.

Part 8: Choropleth maps

30. Okay, we are almost ready to make choropleth maps. Now would be a good time to discuss what one is.

A **choropleth map** (from Greek choro ("area/region") + pleth ("multitude")) uses **pre-defined regions** which are **shaded** in proportion to the measurement of a statistical variable. The choropleth map provides an easy way to visualize **how a measurement varies across a geographic area** or **it shows the level of variability within a region.**

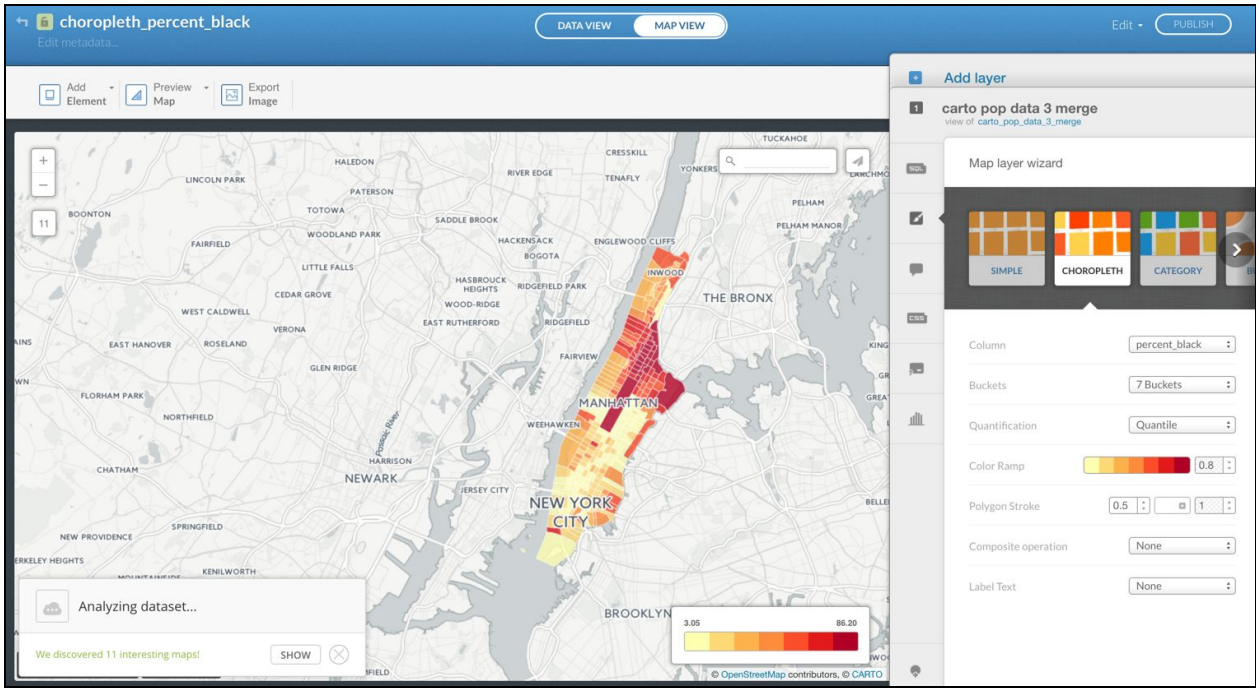
(Source: paraphrased from Wikipedia)

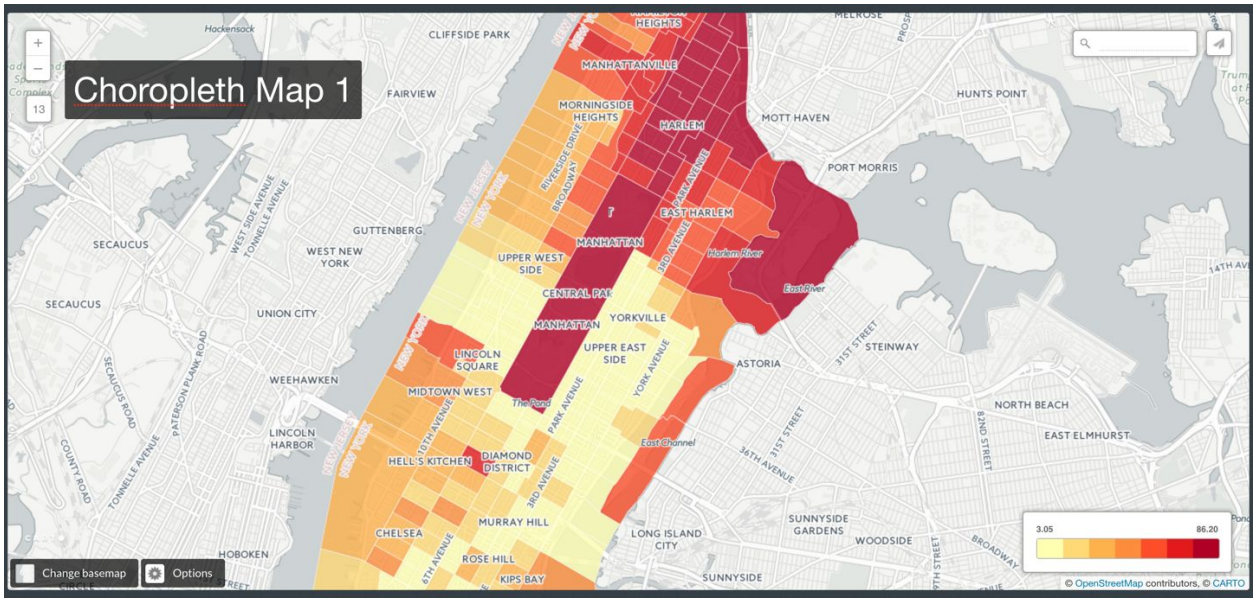


31. Go here to look at more. Notice how they all feature pre-defined geographic regions which are shaded according to a particular measurement!

32. Now, back in Carto, open your Map Wizard, and click choropleth. For column, choose **percent** white or **percent** black or **percent** of any other race/ethnicity, and take a screenshot of your map below.

Note: if you haven't changed data type to number for the column you wish to map, you won't find it available.





33. In the map above, notice the stark changes between adjacent neighborhoods. Feel free to change the basemap, add elements, explore other functions in Carto to improve your map. For example, you might want to try to remove the Central Park census tract...

Deliverables:

Answer the 8 questions in this lab, and take all the screenshots requested.

That's enough for now. Just spend some time absorbing everything you just learned about mapping in Carto. In your own words describe what the following functions do in Carto:

Filter

Table merge (i.e. Column join)

Visualization

Category

Choropleth

Change data type

Bios of Instructors

Noel Hidalgo: Mr. Hidalgo stands at the crossroads of technology, government, community, and impact. He believes in participatory communities and uses technology to improve people's lives. His work has been achieved through patience and organizing problem-solving teams. Mr. Hidalgo is a published author and known as an effective organizer who can walk between worlds.

Currently, he is a Technology and Democracy fellow at Harvard Kennedy School's Ash Center for Democratic Governance and Innovation and is the Executive Director of BetaNYC.

Since 2009, he has organized BetaNYC to be a driving force to improve New York City's use of technology and share its data. BetaNYC, with over 3,800 members, has advocated for the passage of seven government transparency laws, including the city's transformative open data law and city record online law. BetaNYC runs the New York City Civic Innovation Fellows program in partnership with the Manhattan Borough President Gale A. Brewer and curates the NYC School of Data community conference.

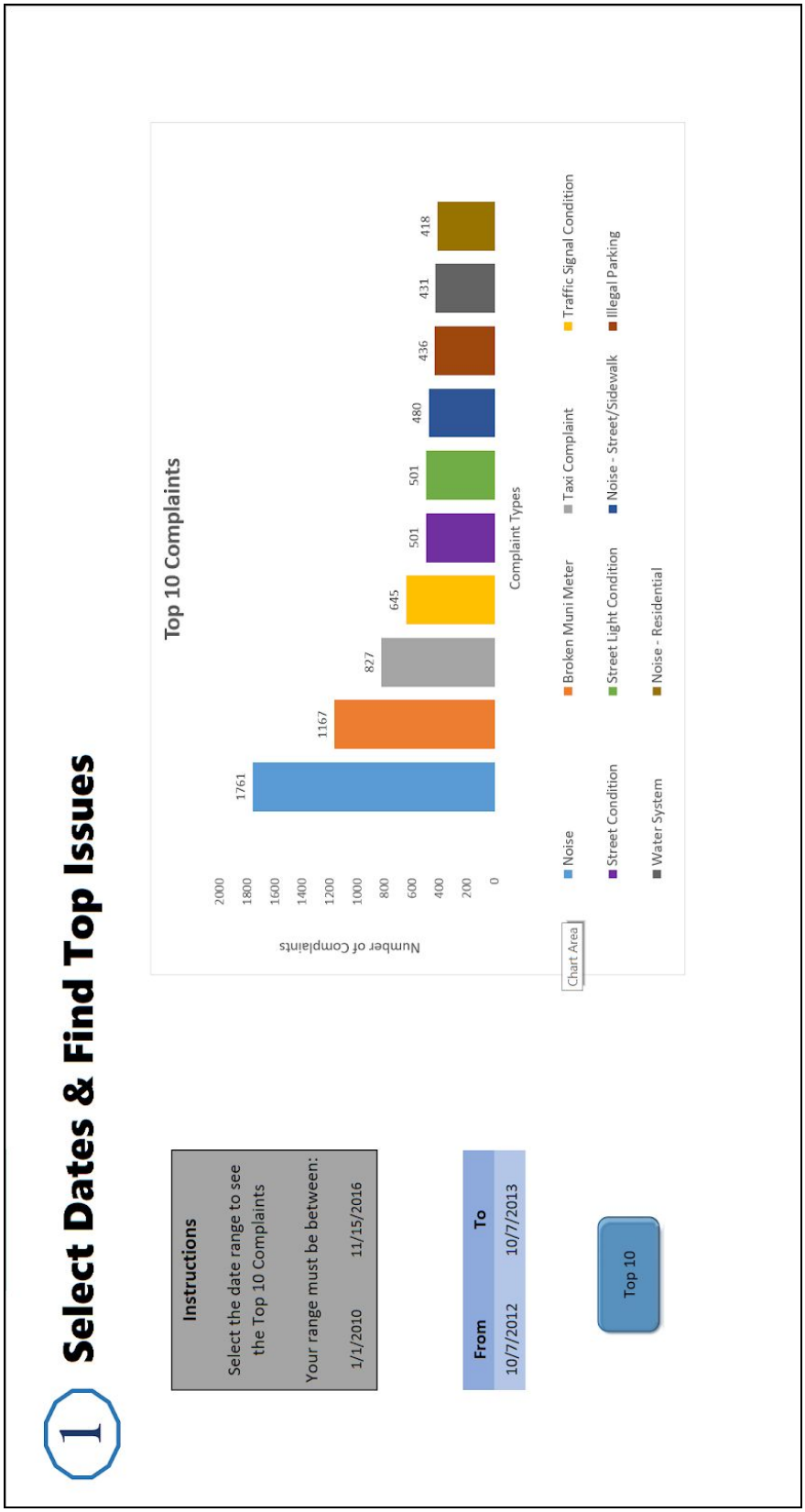
Mr. Hidalgo is an Eagle Scout, an elected member of Code for America's National Advisory Council, and a former fellow at Data & Society Research Institution.

Emily Goldman: Emily finished her Ph.D. in City & Regional Planning from Cornell University in August, 2016. During her time at Cornell, she was a teaching assistant for five semesters, guest-lectured regularly, taught GIS as a Visiting Lecturer, and developed six weeks of curriculum on GIS and data analytics.

Prior to starting her Ph.D. her professional background and experience was in Historic Preservation, working for four years at the NYC Landmarks Preservation Commission. Her dissertation, which combines these interests, is called *Preserving the Social Fabric of Community: A Mixed Methods Analysis of Brooklyn's Newly-Designated Historic Districts (2007 - 2015).* She is happy and honored to be contributing her knowledge and experiences with NYC government, GIS, data analytics, and curriculum development toward the third year of the Civic Innovation Fellowship program.

Fund for the City of New York: The Fund for the City of New York was established by the Ford Foundation in 1968 with the mandate to improve the quality of life for all New Yorkers. For over three decades, in partnership with government agencies, nonprofit institutions and foundations, the Fund has developed and helped to implement innovations in policy, programs, practices and technology in order to advance the functioning of government and nonprofit organizations in New York City and beyond.

Addendum: Preview materials from Alpha version of *Board Stat*



2 Explore The Issue...

Water System

GO

Customize time range here

(hold ctrl button to select more than one month/year at the time)