

Making Data Accessible (and Joyful)



Jenny Ye
Data Director, WNYC Data News Team

Goals:

- present data journalism principles/projects
- listen to data stories
- get started on visualizing data
 - Pivot Tables, Chartbuilder, Carto Maps
- introduce tools for further tinkering

WNYC Data News Team

Tell macro (city/region) and micro (me) stories simultaneously.

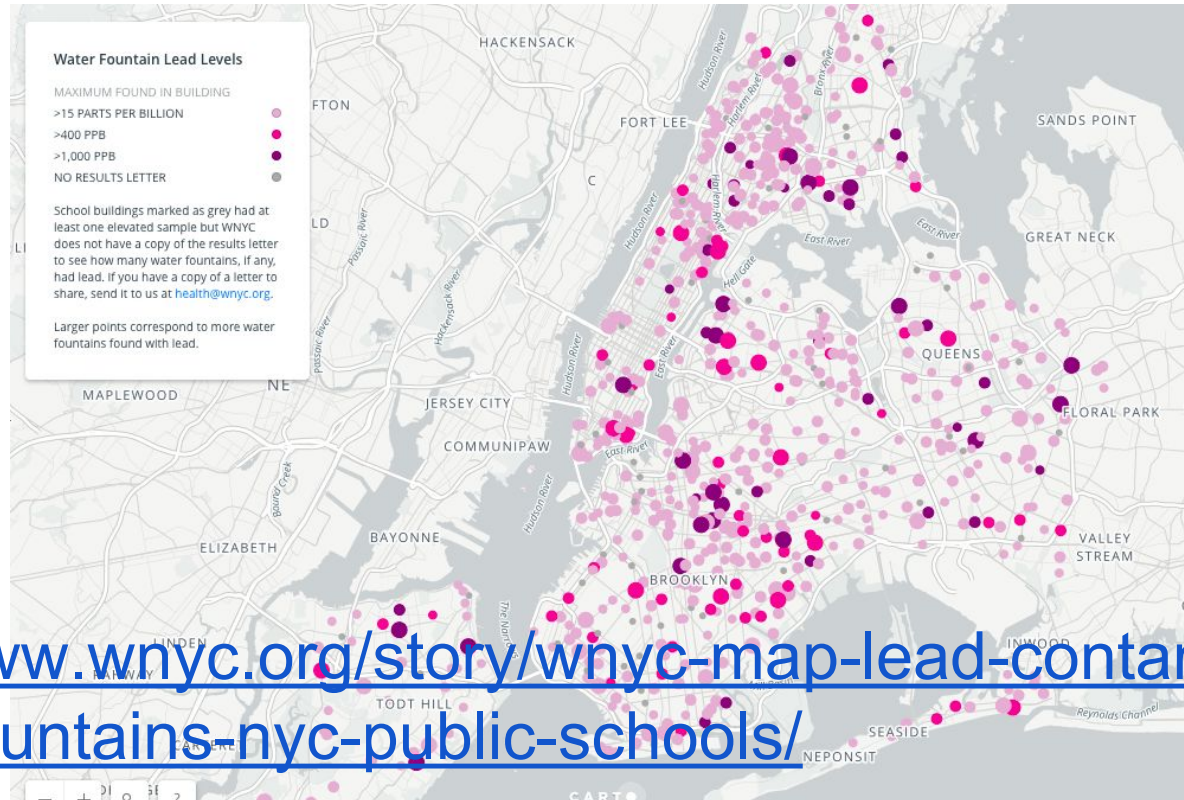
Great projects are a service to our community.

We open our underlying data for reuse and scrutiny.

Help people understand the stories behind the numbers.



Recent Project: Water Tests at NYC public schools



<http://www.wnyc.org/story/wnyc-map-lead-contamination-water-fountains-nyc-public-schools/>

Recent Project: Water Tests at NYC public schools

2015 story:

<http://www.wnyc.org/story/most-new-york-city-schools-have-not-tested-their-water-decade/>

Data:

<http://schools.nyc.gov/AboutUs/schools/watersafety>

(right click, Inspect, Network tab)

→ macro and micro: NYPD Bruised series

Misdemeanor drug possession

About **2.5%** of people charged with misdemeanor drug possession in New York City are also charged with resisting arrest.

By race



White defendants are charged with resisting arrest in **1.7%** of cases



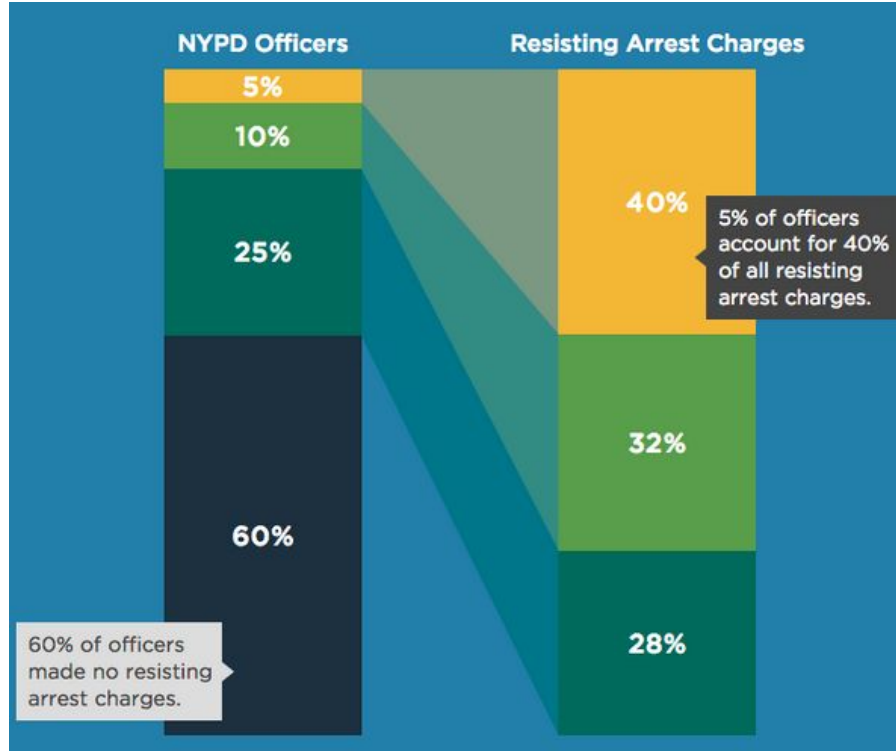
Black defendants are charged with resisting arrest in **3.1%** of cases

That means a black defendant in a misdemeanor drug possession case is **85.4% more likely** to get charged with resisting arrest than a white defendant.

- Lead your analysis with questions, not topics.

<http://www.wnyc.org/story/resisting-arrest-black-white/>

→ macro and micro: resisting arrest



<http://www.wnyc.org/series/nypd-bruised/>

“Reporter Robert Lewis looked for and found telling patterns in the conduct of a minority of NYPD officers, including the one whose chokehold led to Garner’s death. Along the way, one story flows into another, expanding on questions raised earlier in the series and clearly illustrating the problems with both individual stories and big picture data analysis.”

- 2016 duPont award judges

<http://www.wnyc.org/story/can-the-nypd-spot-the-abusive-cop/>

→ service oriented: SchoolBook

SchoolBook Find a school Enter your address or a school name Go

District 2 > Middle schools

English Español 中文

J.H.S. 104 Simon Baruch

Grades 6-8 District 2 Phone 212-674-4545

Principal Rocco Macri School website Address 330 East 21st Street, Manhattan, NY 10010

Admission by Screened details

Stats

Enrollment 1073 students similar schools 368

Class size 28.2 students per class similar schools 26.2

Racial breakdown

Asian White Hispanic Black Other

Academics

State Tests - English	citywide
55% pass	30%
State Tests - Math	citywide
61% pass	35%

On the web

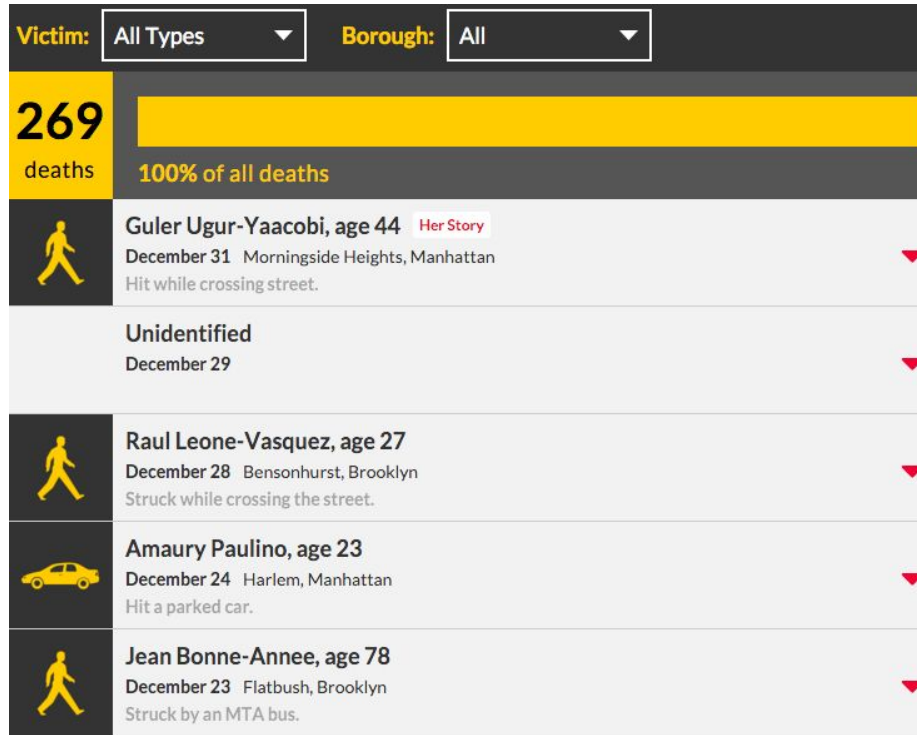
School website

View details for district 2

- school search tool that incorporates multiple public datasets
- reporting tool
- updated annually
- translated into Spanish and Chinese

<https://data.schoolbook.org/>

→ data transparency: Mean Streets



- data collected by Data News on traffic deaths in New York City
- available on our GitHub account as a csv
<https://github.com/datanews/mean-streets>
- more complete than city and state counts
- The New York World at Columbia Journalism School tracked 2015 traffic deaths:
<http://project.wnyc.org/traffic-deaths-2015/>

@NAILBITERBOT: ALERTS FOR CLOSE MARCH MADNESS GAMES



Nailbiter Bot

@NailbiterBot

Following

M: North Carolina (1) v. Villanova (2) is 64-69 with 2:44 left. Watch on [@TBSNetwork](#).
[#MarchMadness](#)



NCAA March Madness Live

Visit March Madness Live to check your bracket, watch every game live and stay up to date with the 2017 Men's NCAA Basketball Tournament.

3.0/5.0 stars – 13,038 ratings

CLOSE GAMES

- * THE SCORE IS WITHIN 8 POINTS WITH BETWEEN 1:30 AND 3:00 LEFT
- * THE SCORE IS WITHIN 6 POINTS WITH BETWEEN :30 AND 1:30 LEFT
- * THE SCORE IS WITHIN 4 POINTS IN THE LAST 30 SECONDS

HOW TO MAKE @NAILBITERBOT

1. FIND A LIVE-ISH GAME SCORES.
2. WRITE A SCRIPT THAT CHECKS EACH GAME TO SEE IT'S ENDING AND CLOSE.
3. EXTEND THE SCRIPT SO THAT IT TWEETS ABOUT EACH NAILBITER IDENTIFIED.
4. SET THAT SCRIPT TO RUN REPEATEDLY.

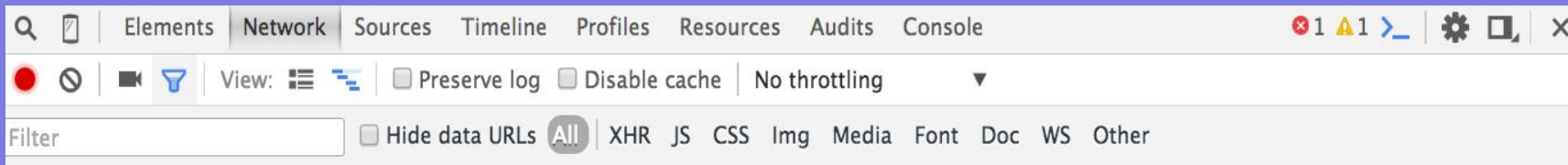
LIVE-ISH SCORES

```
<section id="scoreboard">
  <h3>NCAA Scoreboard</h3>
  <section class="day">

    <div class="day-wrapper">
      <h2>Tuesday, April 7, 2015</h2>
    </div>

  <section class="game final allconf all-conf american-athletic atlantic-coast alpha" id="gamecenter-/game/basketball-women/dl/2015/04/07/notre-dame-uconn">
    <h3>Notre Dame vs Connecticut</h3>
    <div class="game-content">
      <div class="game-status final">Final</div>
      <div class="game-championship">Championship</div>
      <div class="linescore">
        <table class="linescore">
          <tr>
            <th colspan="1"></th>
            <th>1</th>
            <th>2</th>
            <th class="final"></th>
            <th class="status-tray"></th>
          </tr>
          <tr>
            <td class="school" style="background: url(http://i.turner.ncaa.com/dr/ncaa/ncaa7/release/sites/default/files/images/logos/schools/n/notre-dame.24.png) no-repeat 4px 4px"><div class="rank">1</div><div class="team"><a href="/schools/notre-dame">Notre Dame</a></div></td>
            <td>23</td>
            <td>30</td>
            <td class="final score">53</td>
            <td class="status-tray"></td>
          </tr>
          <tr>
            <td class="school" style="background: url(http://i.turner.ncaa.com/dr/ncaa/ncaa7/release/sites/default/files/images/logos/schools/u/uconn.24.png) no-repeat 4px 4px"><div class="rank">1</div><div class="team"><a href="/schools/uconn">Connecticut</a></div></td>
            <td>31</td>
            <td>32</td>
            <td class="final score">63</td>
            <td class="status-tray winner"><span></span></td>
          </tr>
        </table>
      </div>
      <div class="gamecenter-links final">
        <ul class="linklist">
          <li><a href="/game/basketball-women/dl/2015/04/07/notre-dame-uconn" class="gamecenter">GameCenter &raquo;</a></li>
        </ul>
      </div>
    </div>
  </section>
</section>
</section>
```

LOOK FOR THE DATA FILE



CHROME

TOOLS > DEVELOPER TOOLS AND THEN CLICK ON THE NETWORK TAB

FIREFOX

TOOLS > WEB DEVELOPER > NETWORK.

LOOK FOR NAILBITERS

```
// Needs to be in the 2nd half
if (!game.currentPeriod.trim().match(/^2nd/i)) {
  return false;
}

// Point differential
var diff = Math.abs(homeScore - awayScore);

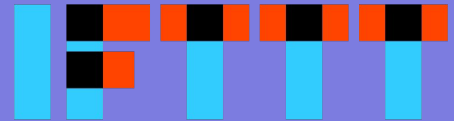
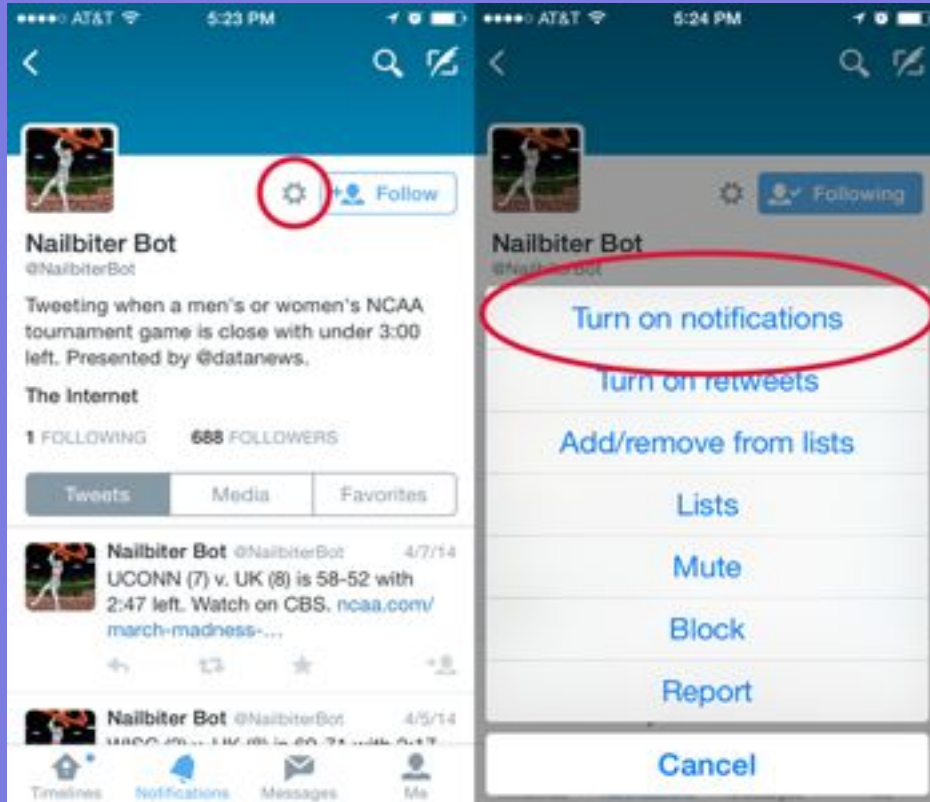
// within 8, between 1:30 and 3 to go
if (min < 3 && min >= 1.5 && diff <= 8) {
  return true;
}

// within 6, between :30 and 1:30 to go
if (min < 1.5 && min >= 0.5 && diff <= 6) {
  return true;
}

// within 4, less than :30 to go
if (min < 0.5 && diff <= 4) {
  return true;
}

return false;
```

MOBILE NOTIFICATIONS, SMS

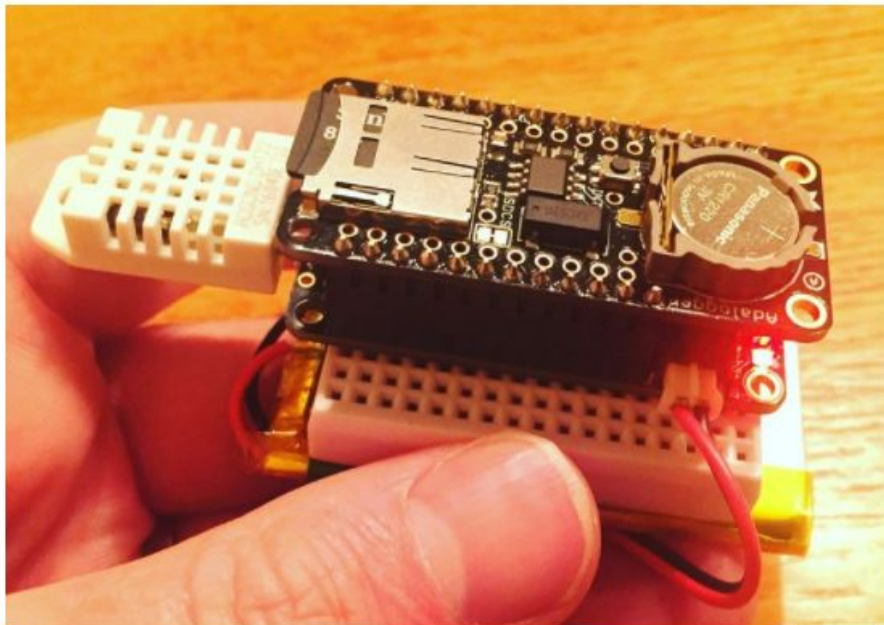


Using sensors to collect data:



Using sensors to collect data:

Harlem Heat Project



The devices that WNYC and its partners will use for the Harlem Heat Project contain a battery, microchip and heat and humidity sensor. (John Keefe/WNYC)

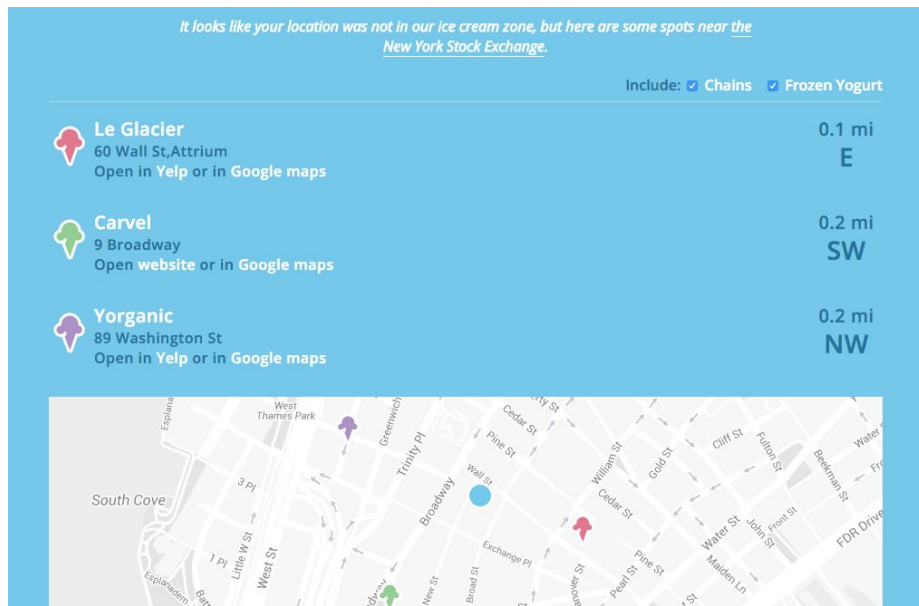
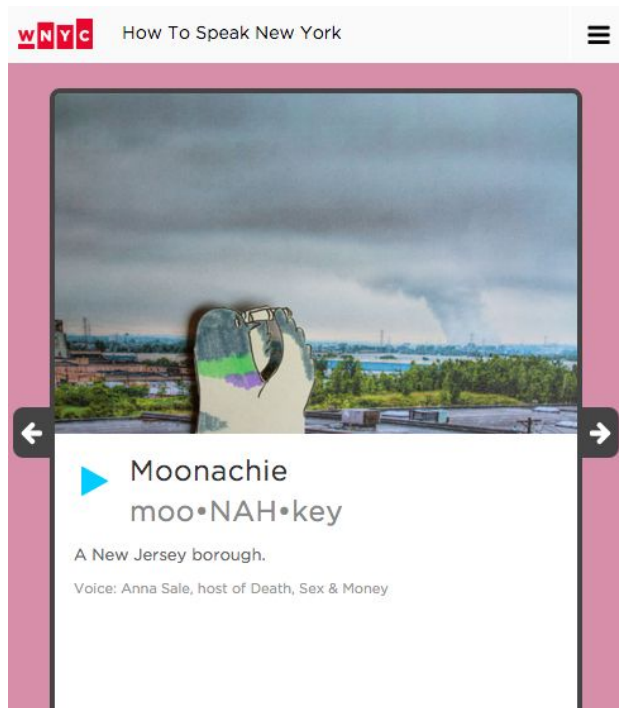
every 15 minutes

- wake up
- record temperature and humidity
- go back to sleep

A few more lighter projects:

<http://project.wnyc.org/pronunciation/>

<http://project.wnyc.org/ice-cream/>



HANDS ON PRACTICE: PIVOT TABLES AND CHARTS

<http://bit.ly/fullspectrum2017data>

Pivot tables help you group data by one or more other variables (count, sum, average, find min/max of groups)

HANDS ON PRACTICE: PIVOT TABLES AND CHARTS

<https://quartz.github.io/Chartbuilder/> **(NO CODE NECESSARY)**

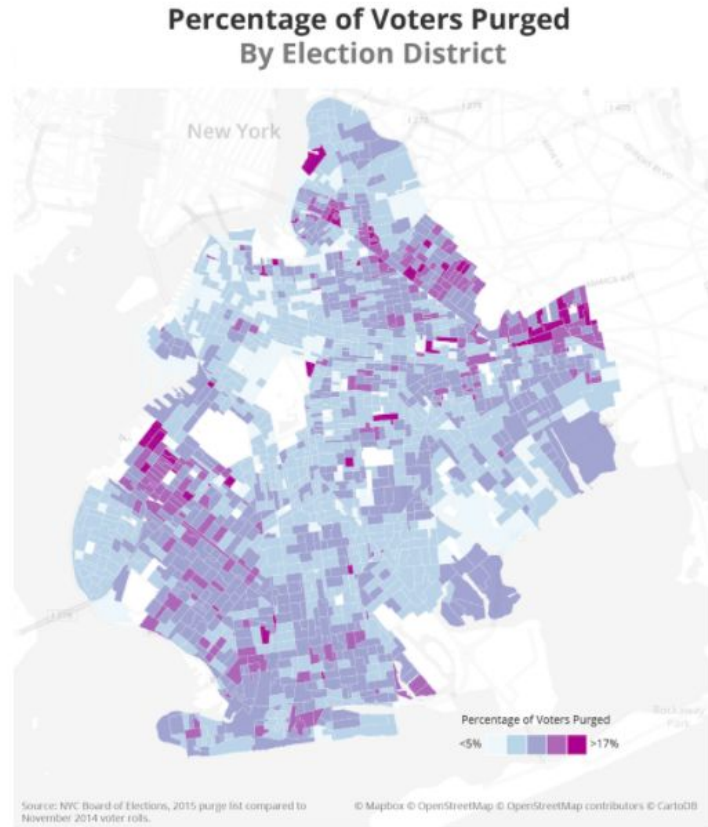
Copy and paste organized data from Excel (include column headings). Add labels. Bar charts should have Y-axes that start at 0, line charts can have Y-axes that start elsewhere. See stats blog post in Slide 17 for more detail.

TRANSPARENCY AND TRUST

- have someone else double check your numbers
- timestamps everywhere
- be clear about what's missing
- bring data back to people affected

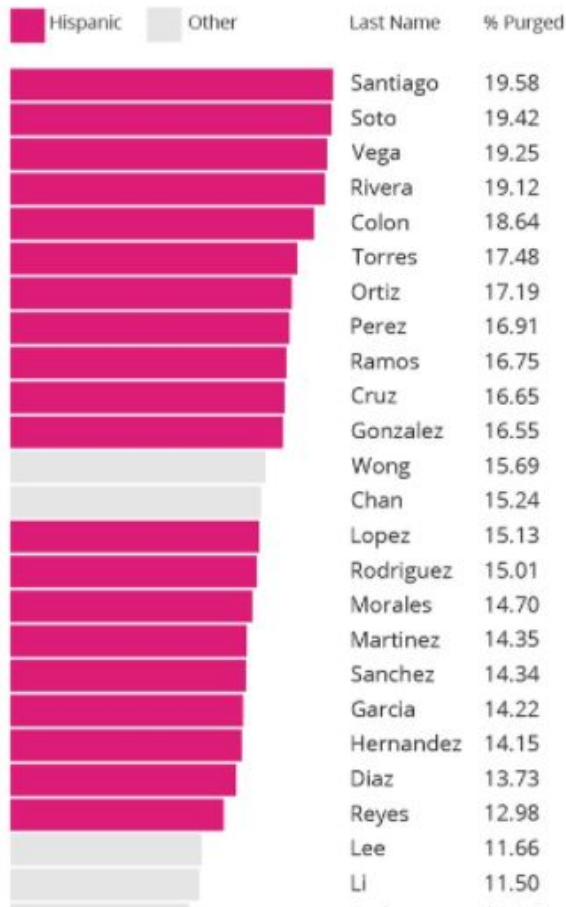
DESCRIBE METHODOLOGY IMMEDIATELY

<http://www.wnyc.org/story/brooklyn-voter-purge-hit-hispanics-hardest/>



The concentrations of purged voters generally align with election districts where the majority of the population is Hispanic, based on the population of individual blocks that make up each election district in the 2010 Census.

Top Names Purged in Brooklyn



DESCRIBE METHODOLOGY IMMEDIATELY

■ ■ ■

Name categorization is based on a U.S. Census Bureau analysis that calculated the way people who used each surname identified themselves in the 2000 Census. WNYC looked at the list of voters purged and the Brooklyn voter rolls from November 2014 and counted as "Hispanic" any last name where 80% of the people using that name identified as Hispanic in the Census analysis. Only names with 200 or more voters on the purge list are included. More at wnyc.org.








RESPOND TO COMMENTS QUICKLY

WNYC Town Hall Tracker

Home address Go

Congressional town halls in NYC metro area (Feb 18 - Feb 26)

REP. YVETTE CLARKE NY 9th District, Democrat	REP. JOE CROWLEY NY 14th District, Democrat	REP. DAN DONOVAN NY 11th District, Republican
 Hosting a meeting on Trump's first 30 days	 No town hall scheduled, but will be available at a Black History Month event	 No town hall scheduled
Event Details Feb 22 6:30 pm Union Temple 17 Eastern Parkway, Brooklyn, NY 11238	Event Details Feb 22 6:00 pm 88-25 Astoria Blvd, East Elmhurst, NY 11369	Contact (718) 635-5277 ddonovan.house.gov @RepDanDonovan
 Contact (718) 287-1142 clarke.house.gov @RepYvetteClarke	 Contact (718) 921-1400 crowley.house.gov @RepJoeCrowley	

Ann Rubin from Bedford Stuyvesant, Brooklyn, NY

I am not a tweeter, nor does my comment fit into 140 characters, so am hoping the WNYC news team is reading these comments as well:

Why are you continuing to give Rep. Hakeem Jeffries a free pass re crediting this past Saturday's ACA rally event as his providing some kind of town hall accessibility to his constituents? Did you not send any reporter to this event? If you had sent anyone to cover that event in full, you would have observed that Jeffries spoke near the beginning of the rally and left -- before any possible Q&A with constituents -- or at least anyone in attendance. Why is WNYC enabling this subterfuge on Jeffries' part? Speechifying at, and then cutting out of a rally before Q&A, provides optics -- but in no way, shape, or form provides the direct accountability to constituents that a "town hall" is meant to serve.

Jenny Ye from WNYC

Hi all,

I'm Jenny Ye from the WNYC Data News team.

Ann - Thanks for your useful information. We are trying to verify this and if we do, we will make a note of the format of the event.

Lindsey - We're only tracking meetings happening through the end of this weeklong recess (Feb 26) so that one won't be included. Thanks for the info.

Rob, Amy, NYCD-16 - We added that information ahead of tonight's town hall. Thank you.

Feb 22, 2017, 5:33 PM



Published by
WNYC News

Trying to Get Around in the Snow? Here's Our Transit Tracker



Feb 9, 2017 · by [The WNYC Data News Team](#)

New York City has been covered with a blanket of snow today, and [National Weather Service](#) meteorologist Carlie Buccola says it's best to stay inside if you can. According to the Service, areas around the city and Long Island received 6 to 9 inches by noon. If you do decide to leave home, WNYC will continue to monitor the latest information about New York and New Jersey transportation outages:



Transit Tracker



The latest info on transportation outages in New York and New Jersey. Updated by WNYC staff starting Thursday, February 9.



Subways

Service change and delays



Buses

All buses running with delays, with some service changes.

[Details](#) +



NJ Transit

Expect 15 min delays system-wide, going both ways



LIRR

System-wide delays during PM rush



Metro-North

THE TRACKER IS NOT
CURRENTLY BEING
UPDATED.

BE CLEAR ABOUT
DATA FRESHNESS

DATA -> AUDIO

Numbers on air:

Metal Detectors Remain In Schools Despite Drop in Crime

<http://www.wnyc.org/story/metal-detectors-remain-in-schools-despite-drop-in-crime/> (list of weapons here adds too many numbers, they are raw numbers, not rates)

Resisting Arrest in Black and White

<http://www.wnyc.org/story/resisting-arrest-black-white/>

(micro to macro, Eric Garner to citywide. Just two stats - black vs. white defendants citywide, and on Staten Island)

More examples: <http://www.wnyc.org/series/data-news-updates/1>

Numbers on air:

- **Lead with and focus on one statistic.**
- **No more than two numbers in a sentence.**
- **Compare apples to apples, fractions to fractions, percentages to percentages.**
- **Consider raw numbers AND rates. Rates are usually more important.**

Sonifying Data



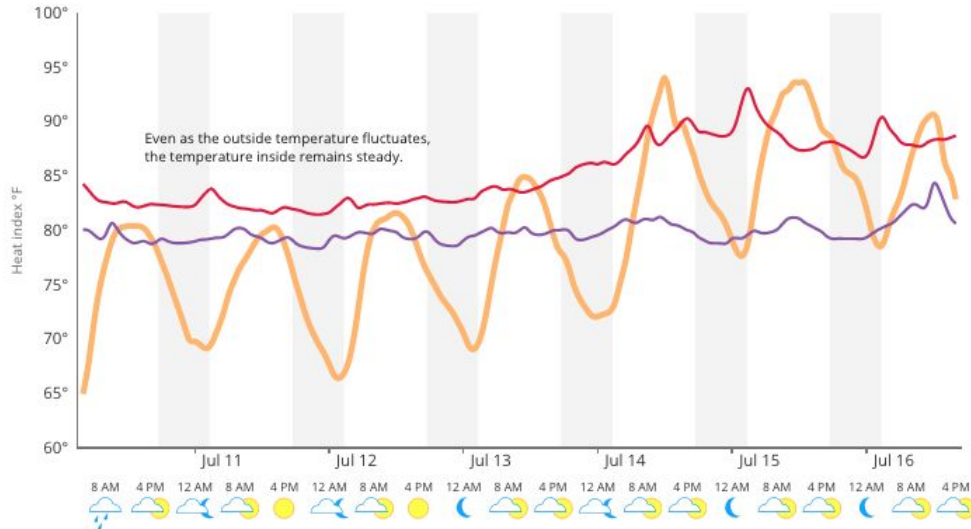
Harlem Heat



■ First floor of a 10-story building, with south-facing windows. Apartment has an A/C, but it's in a different room from the sensor.

■ Top floor of a 7-story building with southeast-facing windows. Apartment doesn't have A/C.

■ Outside heat index



<http://www.wnyc.org/story/harlem-he-at-song>

higher note == higher temperature

Sonifying Data

Reveal's story on Oklahoma earthquakes

<https://www.revealnews.org/article/listen-to-the-music-of-seismic-activity-in-oklahoma/>

Examples by Brian Foo:

<https://mic.com/articles/177877/data-sonification-lets-you-literally-hear-income-inequality#.jcf6sla25>

Guided Tutorials: Try on your own!

CartoDB Mapping: Start with Election Data Tutorial

<https://carto.com/docs/tutorials>

Building Bots to Assist Reporting:

- <https://drive.google.com/folderview?id=0ByS5FG4Zedw8eHBTMk0xVWlXMnM&usp=sharing>
- ONA 2015 session audio:
<http://ona15.journalists.org/sessions/botreporting/#.VtHO3TYrJE5>
- If This Then That Alerts: <https://ifttt.com/>

Styling your visualizations

Check for color accessibility:

<http://webaim.org/resources/contrastchecker/>

Choose colors for your maps:

<http://tristen.ca/hcl-picker/#/hlc/4/1/21313E/EFEE69>

Data Sources:

● Census Reporter

- <https://censusreporter.org/>
- Search by address or place name. Uses the most recent American Community Survey or Census, allows you to save and embed charts, download data. Be careful not to use numbers with this symbol † because of a large margin of error. Small areas will have higher margins of error in ACS data.
- If you need Census data (closest to a full count) use <https://factfinder2.census.gov>

More data tips / tools

PDF -> CSV: Ideally you get your data as Excel, .csv or other machine readable formats. (PDFs are bad) But **Tabula** (tabula.nerdpower.org) can help you try to get data in PDF tables into a CSV. Make sure to do some spot checking.

Try to get data in the most granular form. The smaller geography, the better. Records are better than summaries. **Use forms as your guides for data requests. Fields in a form usually translate to fields in a database.**

Basic Statistics Review: <https://source.opennews.org/en-US/learning/know-your-stats/>

Guide to bad data (& what to do about it) from Quartz: <https://github.com/Quartz/bad-data-guide>

When Maps Shouldn't Be Maps:

<http://www.ericson.net/content/2011/10/when-maps-shouldnt-be-maps/>

More data tips / tools

Secure messaging (Signal, PGP)

<http://www.teenvogue.com/story/how-to-keep-messages-secure>

<https://github.com/hackshackers/hhnyc-crypto/blob/master/tipsheet.md>

Thanks! Questions? Tips?

Jenny Ye

@thepapaya / jye@wnyc.org

Data News Team:

@datanews / datanews.tumblr.com