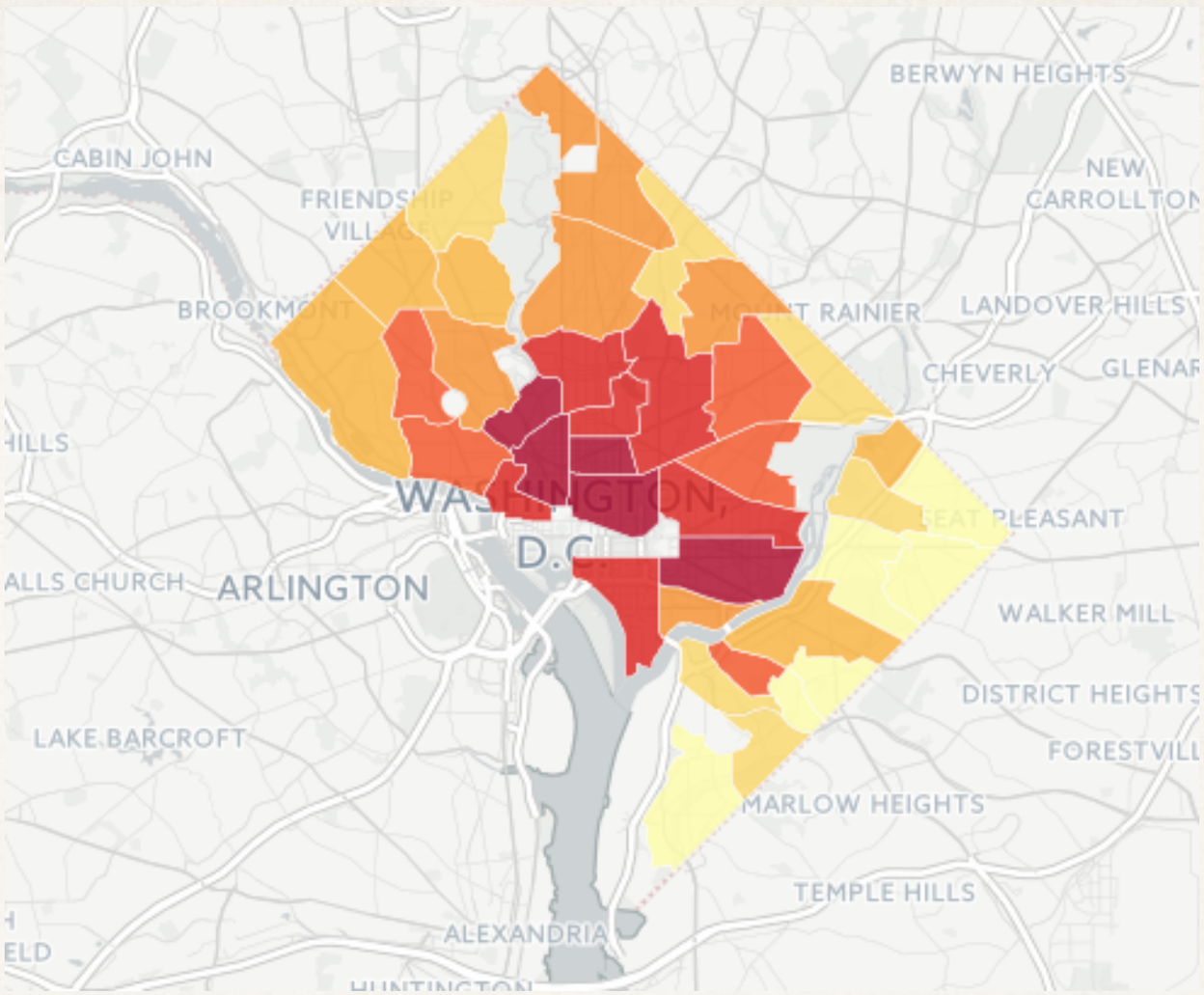


Source: DataLensDC

Visualizing DC's Open Data

Open Data Day DC 2016

@datalensdc

What we're covering

High level start to finish roadmap for data visualization when working with open data.

Select Your Data → Transform Your Data → Visualize Your Data

Who am I?

I started DataLensDC (www.datalensdc.com / @datalensdc) last summer, a website that visualizes the trends and characteristics of the District. My work has been published in The Washingtonian, Washington City Paper, and The Atlantic's CityLab, among other.

Also…

We have some really awesome TAs here to help you out today!

Where is Open Data?

Lots of places!

Mainly: data.gov (for federal data)
opendata.dc.gov (for District data)
But also District agency sites (like OSSE, WMATA), and non government actors.

Google ferociously and often.

Select Your Data ➡ Transform Your Data ➡ Visualize Your Data

Most Common Datatypes

CSV - comma separated values.

Most universally accepted data format, easily accessed in Excel and programming languages.

JSON/GeoJSON - javascript object notation

Data storage through name-value pairing. Common output from APIs, readable in all coding languages.

Just want a CSV? It's easy! http://konklone.io/json/

PDF (not machine readable)

But there's an app for that! http://tabula.technology/

Select Your Data ➡ Transform Your Data ➡ Visualize Your Data

Should I viz this data?

Visualizations require good data.

So consider the limitations of your data.

Select Your Data ➡ Transform Your Data ➡ Visualize Your Data

Should I viz this data?

How accurate are these estimations?

	Census Tract 1, District of Columbia, District of Columbia	
	Estimate	Margin of Error
Median earnings in the past 12 months --		
Total:	101,318	+/-19,523
Car, truck, or van - drove alone	124,265	+/-38,622
Car, truck, or van - carpooled	83,333	+/-132,788
Public transportation (excluding taxicab)	107,404	+/-85,293
Walked	102,384	+/-43,043
Taxicab, motorcycle, bicycle, or other means	128,125	+/-64,195
Worked at home	63,810	+/-18,315

Source: Census American Community Survey

Margin of error shows the source's confidence in it's estimations. In this example:
Census is 90% confident the right number is between (estimate - margin of error) and (estimate + margin of error)

Select Your Data → Transform Your Data → Visualize Your Data

Should I viz this data?

How accurate are these estimations?

	District of Columbia	
	Estimate	Margin of Error
Median earnings in the past 12 months --		
Total:	53,157	+/-831
Car, truck, or van - drove alone	59,939	+/-1,628
Car, truck, or van - carpooled	55,899	+/-3,151
Public transportation (excluding taxicab)	48,187	+/-1,266
Walked	52,114	+/-2,275
Taxicab, motorcycle, bicycle, or other means	56,536	+/-3,121
Worked at home	50,322	+/-3,542

Source: Census American Community Survey

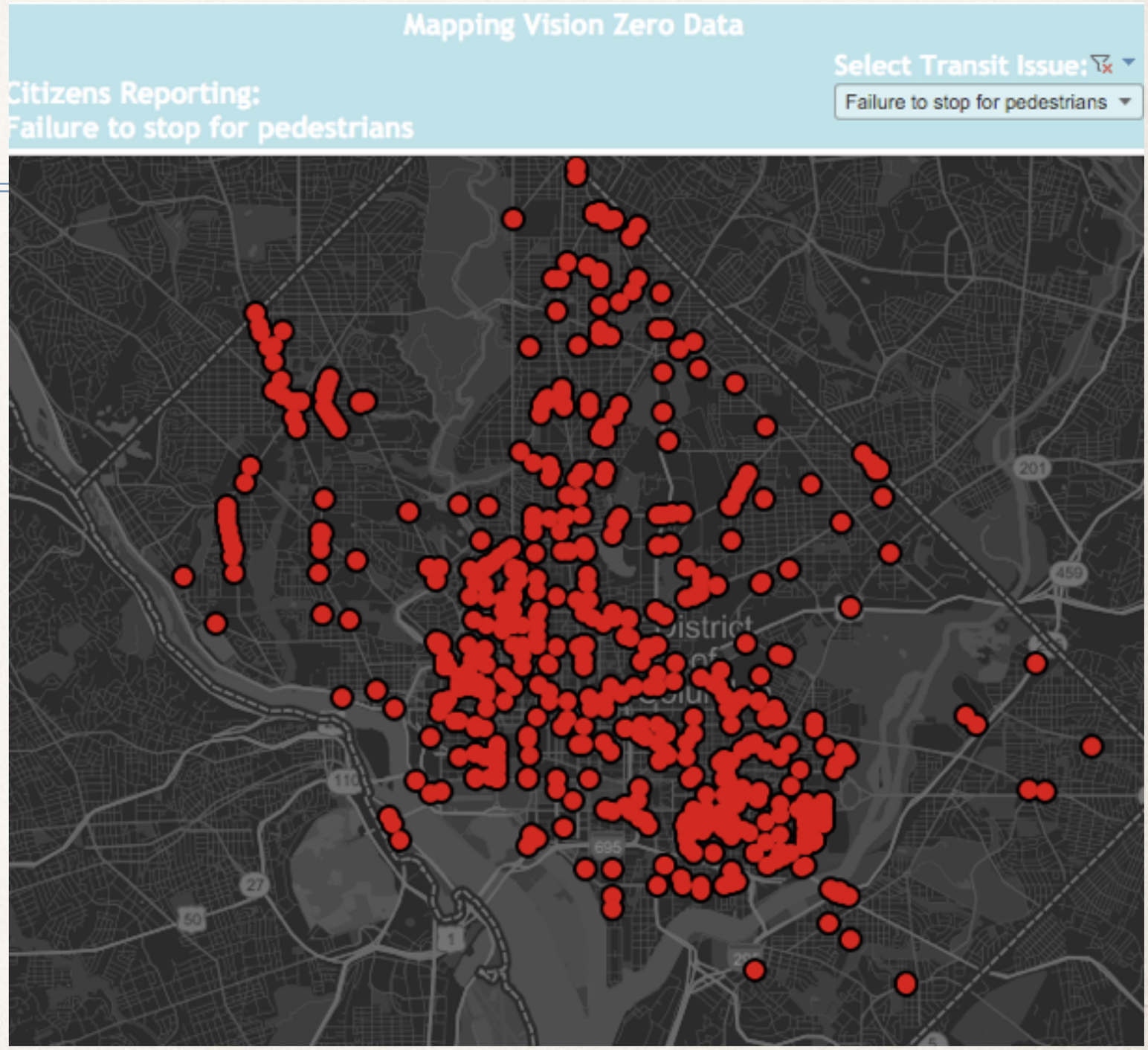
Generally, the margin of error decreases as the population size gets bigger. This may mean looking at a larger geography, a more general group of people or data collected over a larger time span.

Select Your Data ➔ Transform Your Data ➔ Visualize Your Data

Should I viz this data?

Is the data representative of the population?

The Vision Zero Safety Map shows locations people have self-reported online as hazardous.



Source: DataLensDC

Select Your Data → Transform Your Data → Visualize Your Data

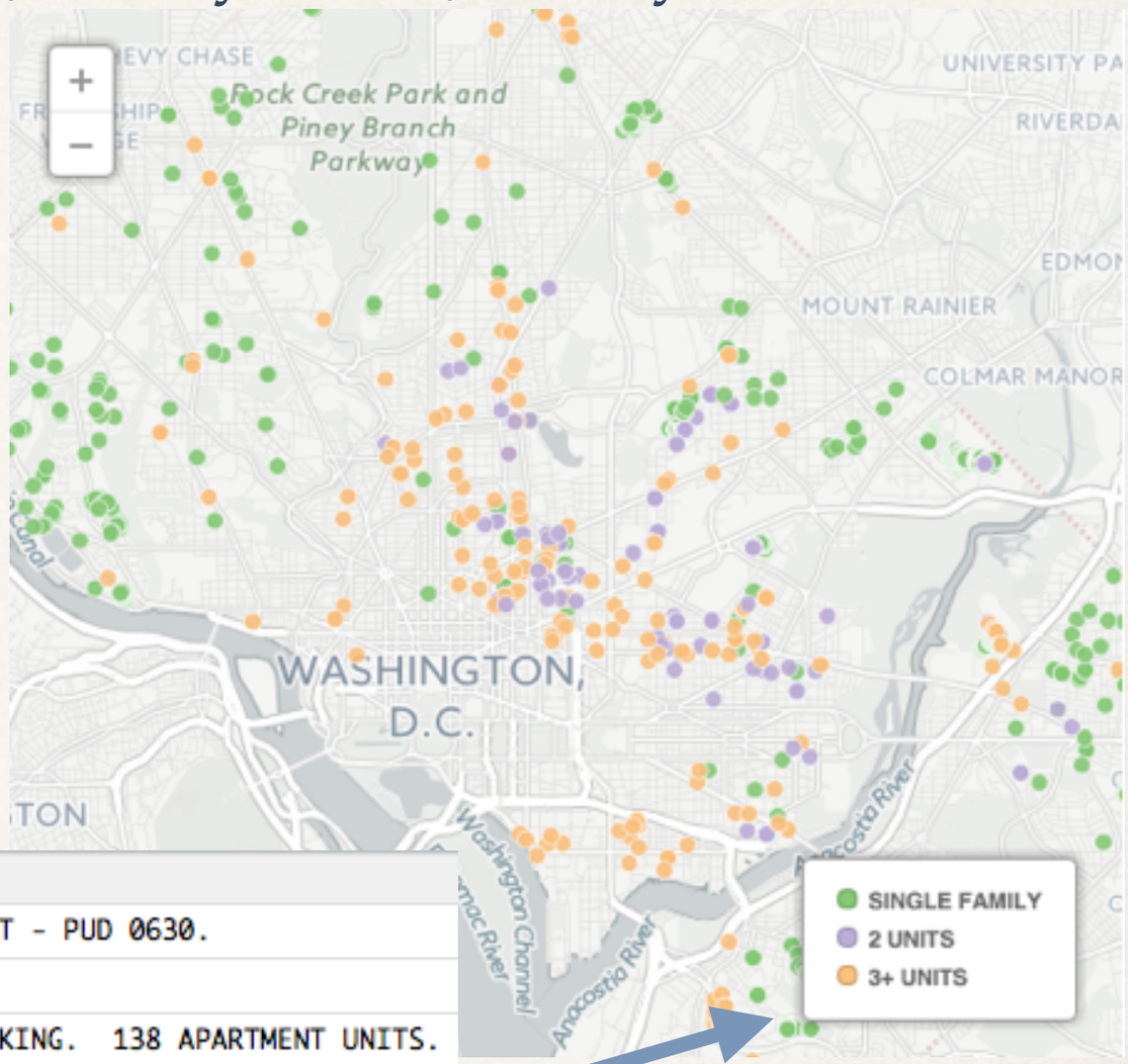
Common Data Transformations

Text analysis.

Messy open text fields sometimes hold valuable information.

DESC_OF_WORK
NEW SINGLE FAMILY DWELLING IN THE POLLIN MEMORIAL COMMUNITY DEVELOPMENT - PUD 0630.
ADDITION OF BOXING ANNEX TO BALD EAGLE REC CENTER.
NEW SIX STORY MULTI-FAMILY DWELLING WITH TWO UNDERGROUND LEVELS OF PARKING. 138 APARTMENT UNITS.
NEW CONSTRUCTION OF A 12 UNIT CONDOMINIUM BUILDING
2 UNITS 3 BEDROOM AND 2 1/2 BATH
NEW CONSTRUCTION OF A MULTI UNIT APARTMENT BUILDING
A NEW CONSTRUCTION 2 UNIT FLAT.
A NEW CONSTRUCTION 2 UNIT FLAT.

Permits for New Residential Buildings, January 2012- January 2016



Source: DataLensDC

You can check out the full code behind this here.

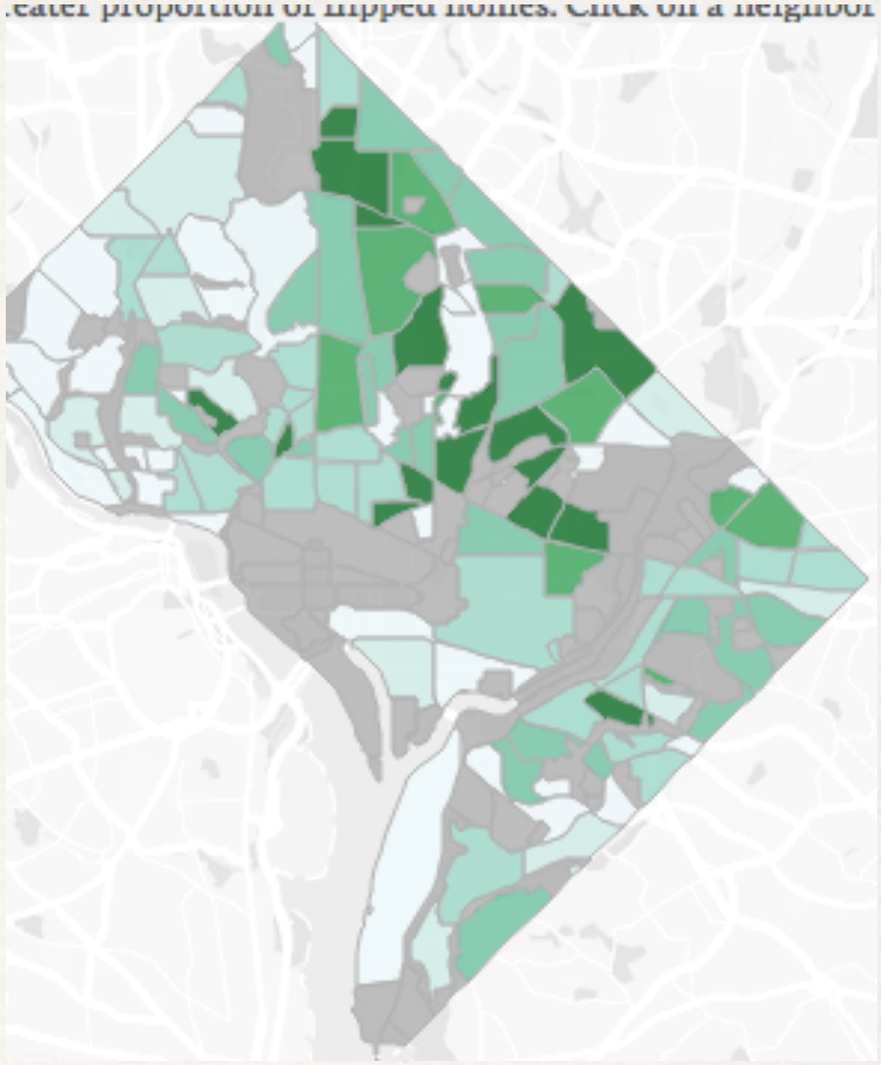
Select Your Data → Transform Your Data → Visualize Your Data

Common Data Transformations

Neighborhood Assignment.

Geolocating an address and then using the resulting latitude and longitude to place that location in a geographic area - ward, zip code, neighborhood, etc.

Flipped Homes as % of Homes on Market, by Neighborhood
2013-2015



Source: DataLensDC

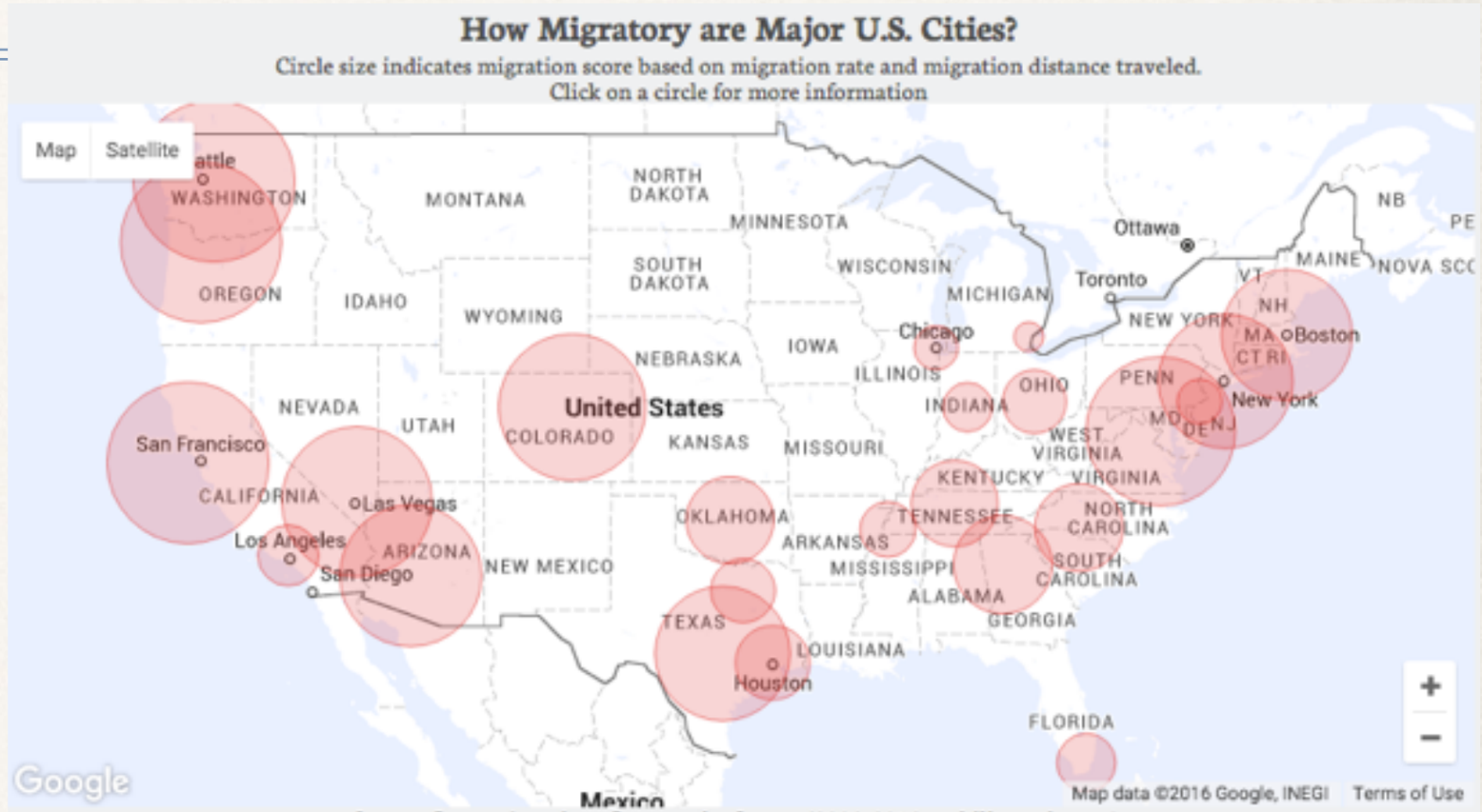
You can check out the full code behind this here.

Select Your Data → Transform Your Data → Visualize Your Data

Common Data Transformations

All the little (un) glamorous bits.

There is often data cleaning to do - numeric fields with commas stored as text, checks for duplicates, and more.



Source: DataLensDC

You can check out the full code behind this here.

Select Your Data ➡ Transform Your Data ➡ Visualize Your Data

Commonly Used Data Analysis Tools + Learning Resources

R	Python	Other
Advanced R by Hadley Wickham	Learn Python the Hard Way	SAS
Swirl. Learn R in R	Learnpython.org	SPSS
		Stata
		MatLab
		Julia
		Excel
		…

Resources for R, Python, and More

datacamp.com codeschool.com

udemy.com kaggle.com

codeacademy.com coursera.org

lynda.com (free through DC public library)

Select Your Data → Transform Your Data → Visualize Your Data

Active DC Data Community

Data Community DC

dcpython™

HACK AND TELL

DataKind

WOMEN WHO CODE

HEAR ME CODE

pyladies

Select Your Data → Transform Your Data → Visualize Your Data

Many Data Viz Tools to Choose From

Point and Click	Statistical Programming	JavaScript
Tableau	**R**	**D3**

Mapping

CartoDB, Leaflet
Mapbox

Select Your Data ➡ Transform Your Data ➡ Visualize Your Data

Many Data Viz Tools to Choose From

Point & Click

Statistical Programming

JavaScript

Tableau

R

D3

- Of all tools, point and click is the easiest to learn if no prior programming experience
- Creates fast and malleable data viz via drag and drop
- Provides a wide range of built-ins but not fully flexible, more bespoke viz styles may not be possible or involve a difficult series of “hacks”

Select Your Data ➡ Transform Your Data ➡ Visualize Your Data

Many Data Viz Tools to Choose From

Point & Click

Tableau

Statistical Programming

R

JavaScript

D3

- If new to R, then steeper learning curve. Best route only if previously familiar with specific programming language
- Greater flexible through programmatic chart design
- Limited ability to create interactive data viz

Many Data Viz Tools to Choose From

Point & Click

Tableau

Statistical Programming

R

JavaScript

D3

- Steep learning curve, challenging even to those with front end experience
- Extremely flexible in design
- Most of the interactive data visualization you see at the NY Times or Washington Post was created D3

Select Your Data ➡ Transform Your Data ➡ Visualize Your Data

Many Data Viz Tools to Choose From

Mapping

CartoDB, Leaflet, MapBox

- Most data visualization tools have mapping, but some tools are mapping specific and very in-depth
- May provide geo-location and well stylized / customized basemaps
- CartoDB and MapBox both have point and click as well as Javascript capabilities. Leaflet is a Javascript language.

Let's Viz Some Data!

A CartoDB Breakout

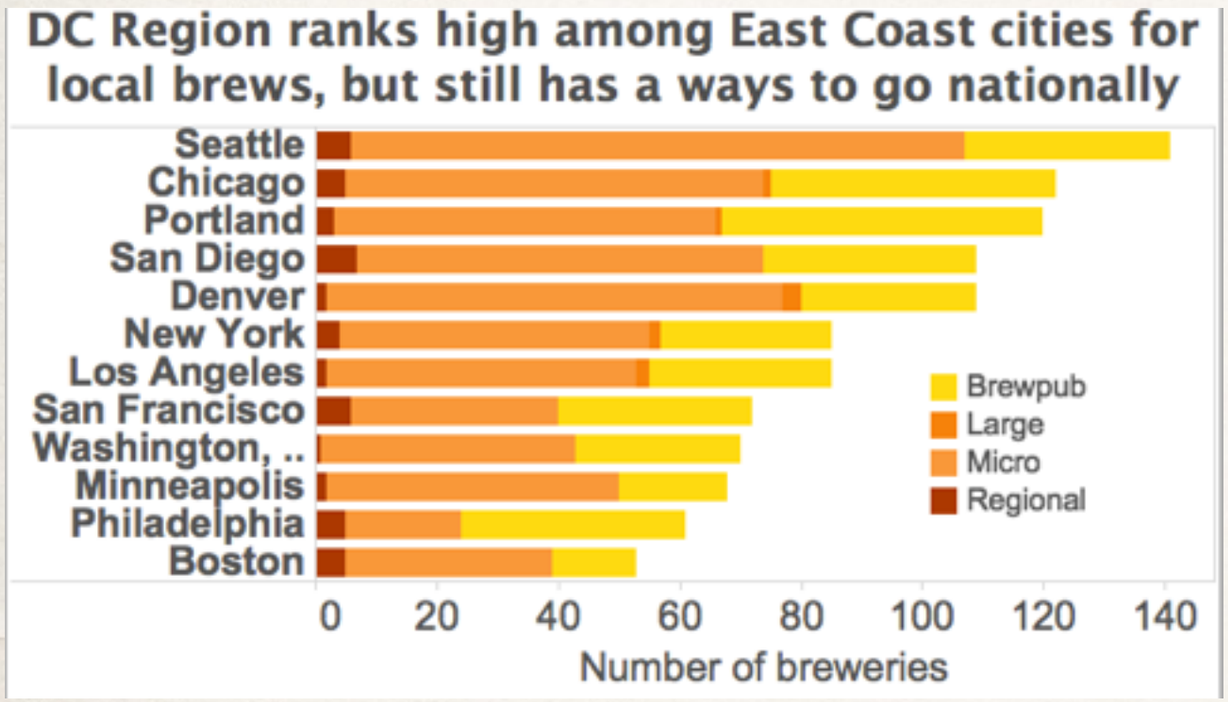
www.cartodb.com

For future reference this Guide to CartoDB from a recent MaptimeDC event is great.

Select Your Data ➡ Transform Your Data ➡ Visualize Your Data

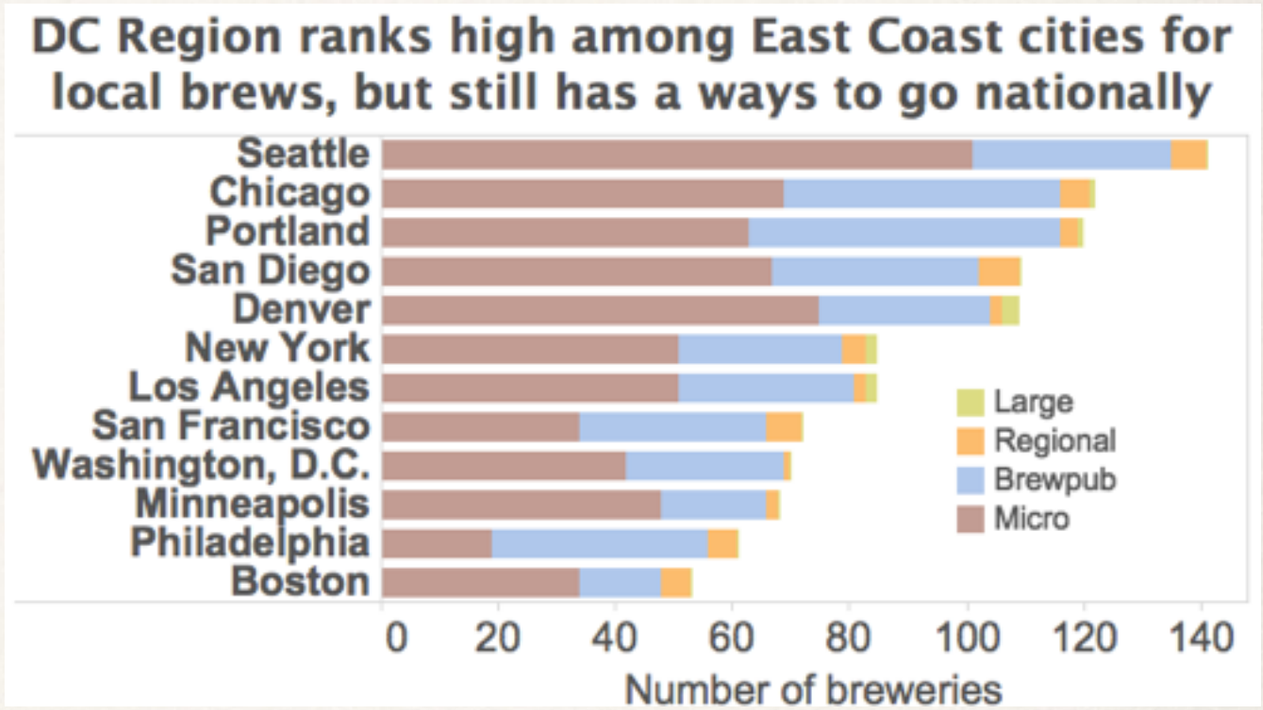
Color Matters

BAD



Source: DataLensDC

BETTER



Source: DataLensDC

At it's worst color can repeal or confuse the reader. At it's best color can draw the reader in and help tell the story.

Select Your Data ➔ Transform Your Data ➔ Visualize Your Data

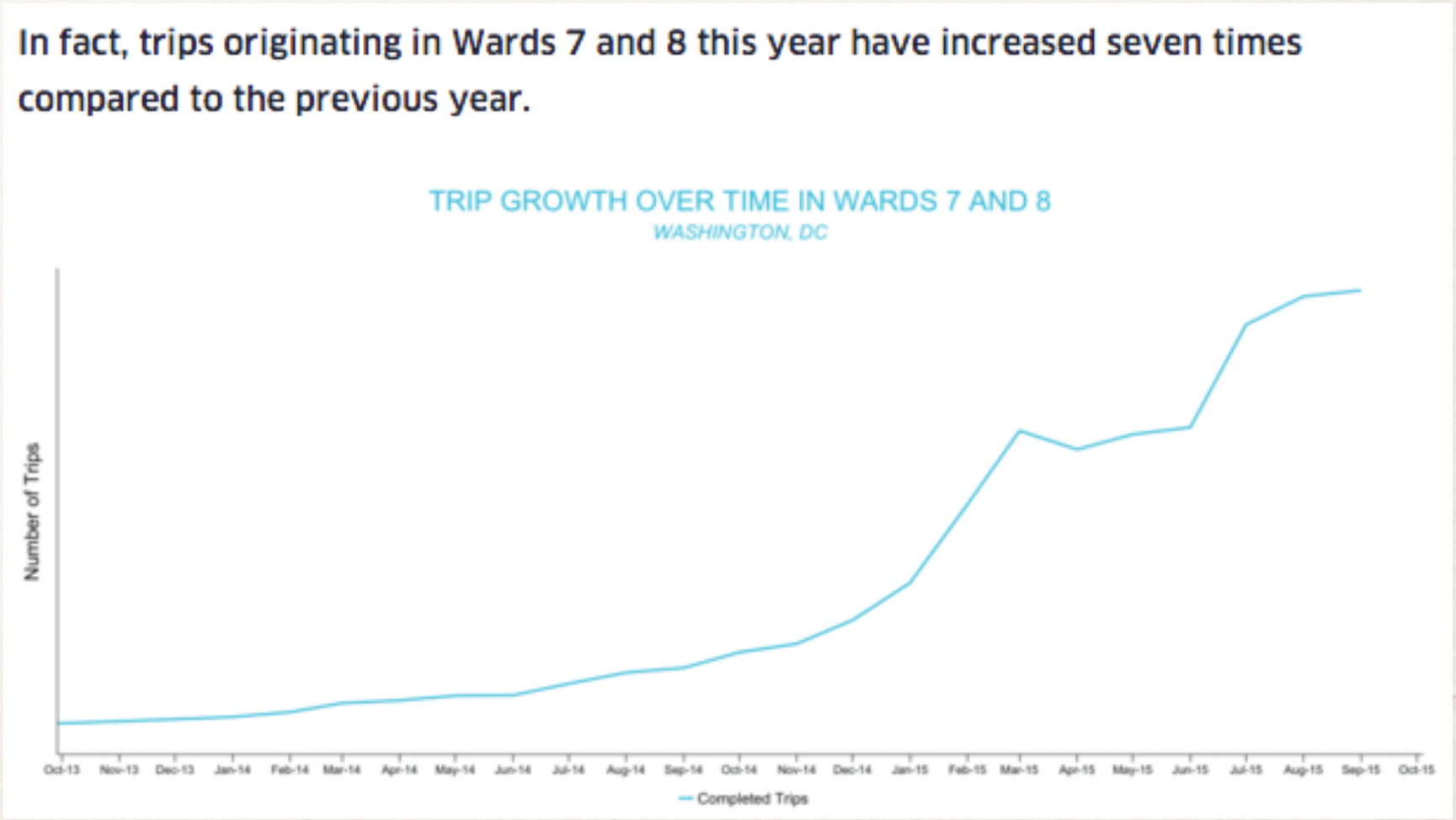
Color Matters

- Use different colors for different categories (qualitative), but not different values across a range (quantitative).
- But never more than six colors.
- Be mindful of color blindness.
- For multiple graphs, a connected color scheme.
- When in doubt, use ColorBrewer2 at colorbrewer2.org

Label Well and Often

Labeling can occasionally be overlooked when focusing on the content of the graph, but are crucial for readers to get the facts

Lack of proper labeling can confuse or mislead, forcing the reader to guess of do “visual math.”

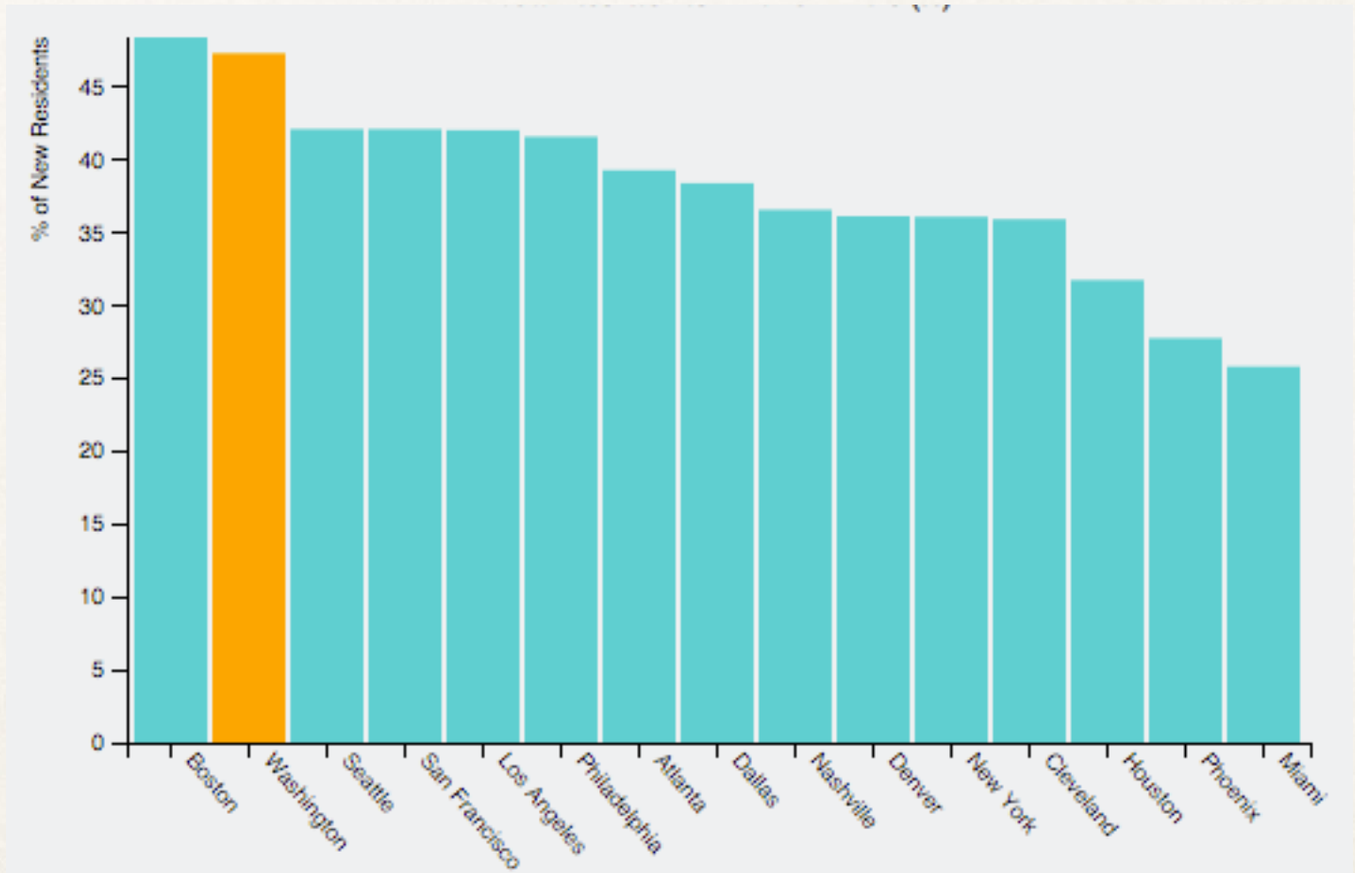


Source: Uber

Titles To Tell Stories

Bad Title: New Residents aged 20-29 (%)

Better Title: Nearly Half of New DC Residents in their 20s



Source: Census American Community Survey (ACS), 2010-2014.

Select Your Data → Transform Your Data → Visualize Your Data

Same Data, Different Story

Mike Hickinbotham @M_Hickinbotham · 20h
Communicating two different messages with the same data @acotgreave #da

Select Your Data ➡ Transform Your Data ➡ Visualize Your Data

Let's Viz Some Data!

A Tableau Breakout

To get started:

https://public.tableau.com/s/download

For future reference Tableau has a ton of free learning resources and a very active community

Select Your Data ➡ Transform Your Data ➡ Visualize Your Data

Awesome Blogs about Data Viz

EVERGREENDATA
INTENTIONAL REPORTING & DATA VISUALIZATION

eagereyes
Visualization and Visual Communication

FLOWINGDATA

DATA STORIES

information is beautiful
ideas, issues, knowledge, data — visualized!

visualising data

JUNK CHARTS
Recycling chartjunk as junk art

information aesthetics. *Where form follows data.*

Select Your Data → Transform Your Data → Visualize Your Data