

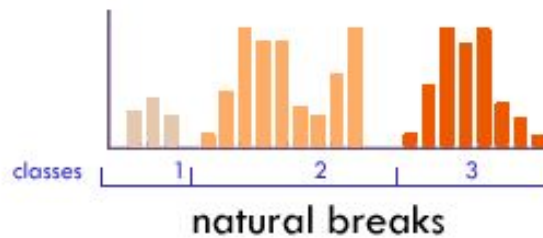
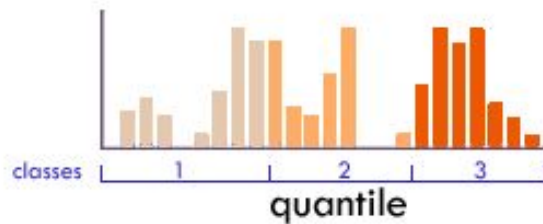
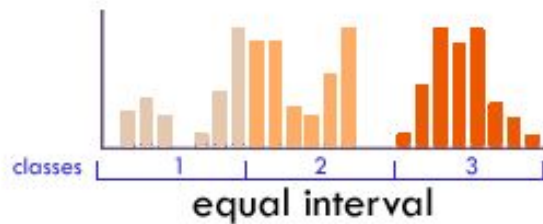
making data work for you

foundations of geospatial analysis

2026.09.14

**different areal
boundaries can
suggest different
conclusions about
the same data**

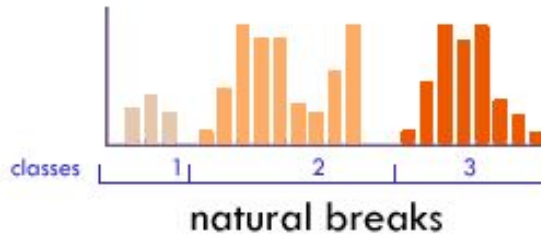
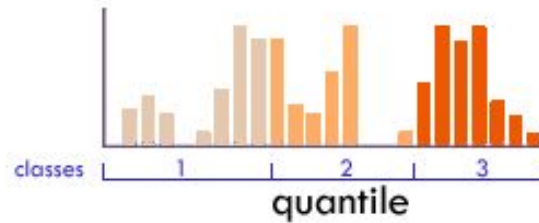
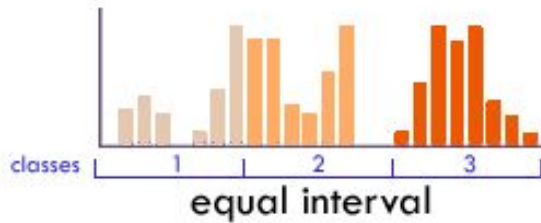
**so can different
data classification
schemes**





EQUAL INTERVAL

divides the data into equal size classes (e.g., 0-10, 10-20, 20-30, etc.) and works best on data that is generally spread across the entire range.

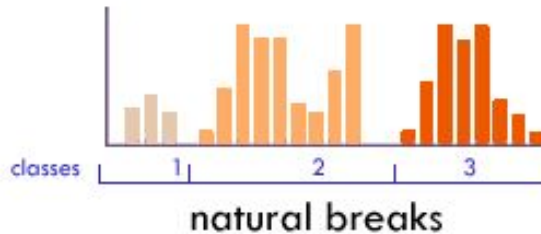
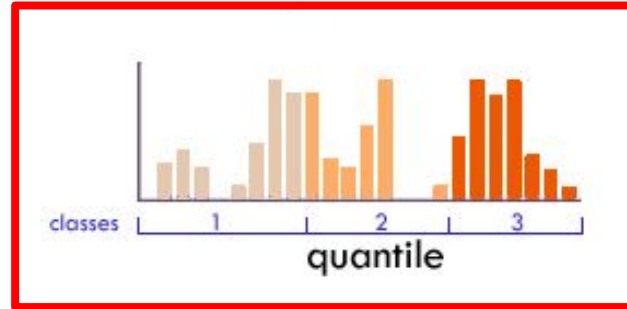
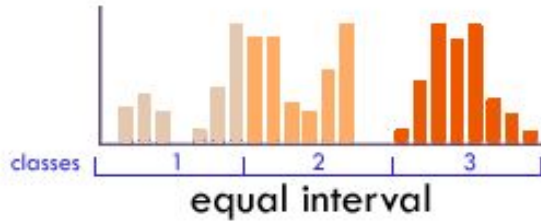


CAUTION: Avoid equal interval if your data are skewed to one end or if you have one or two really large outlier values



QUANTILES

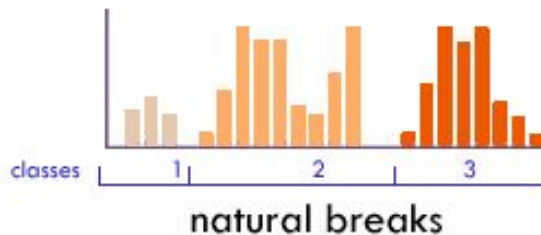
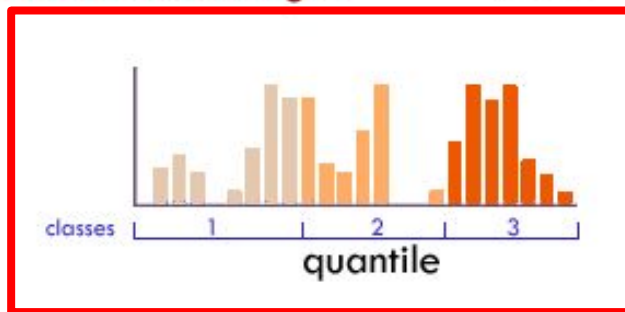
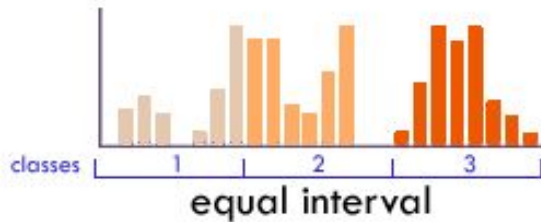
create attractive maps that place an equal number of observations in each class: If you have 30 counties and 6 data classes, you'll have 5 counties in each class.

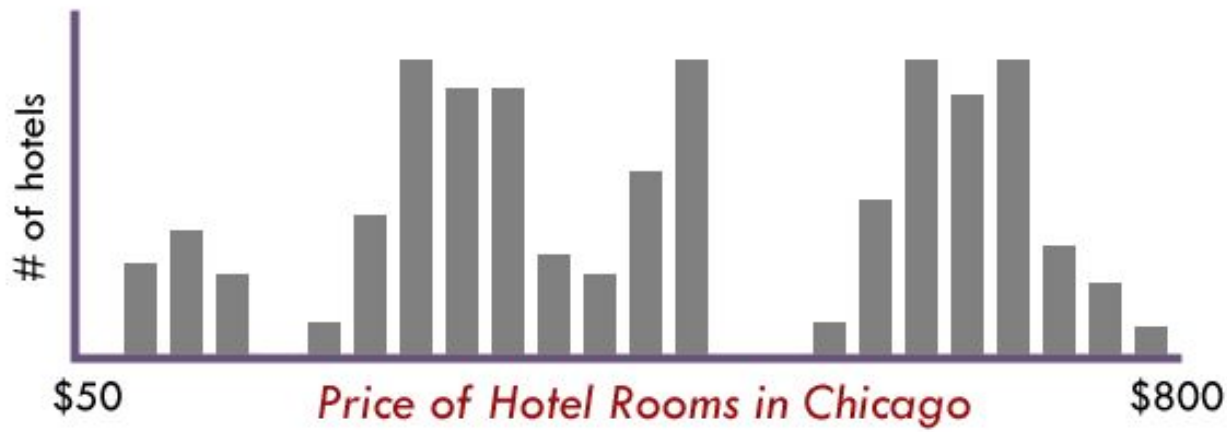




QUANTILES

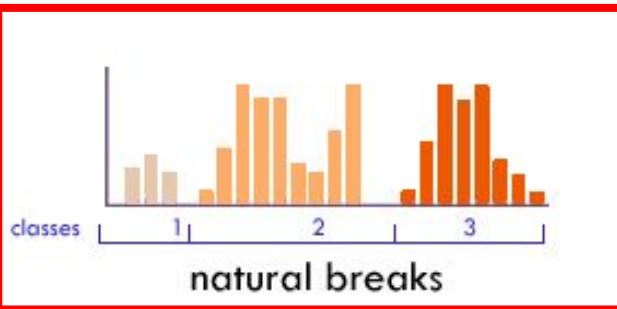
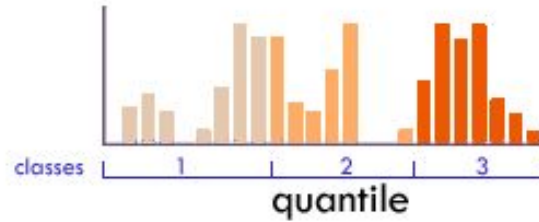
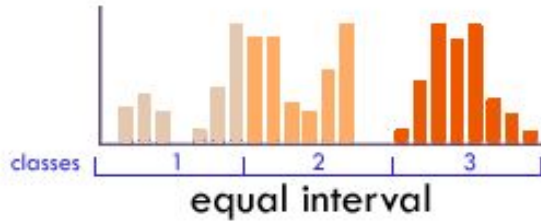
CAUTION: The problem with quantiles is that you can end up with classes that have very different numerical ranges (e.g., 1-4, 4-9, 9-250...the last class is huge)





NATURAL BREAKS

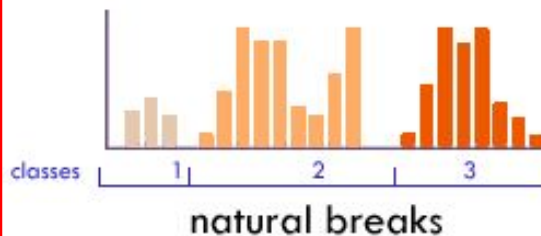
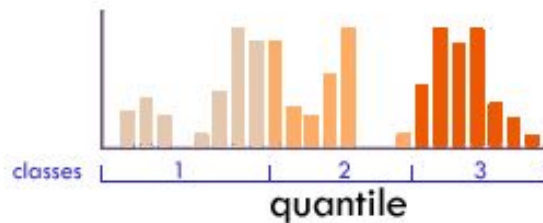
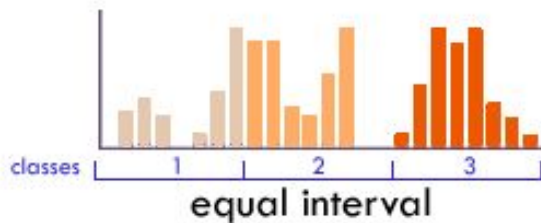
a kind of “optimal” classification scheme that finds class breaks that (for a given number of classes) will minimize within-class variance and maximize between-class differences.





NATURAL BREAKS

CAUTION: each dataset generates a unique classification solution, preventing comparisons across maps

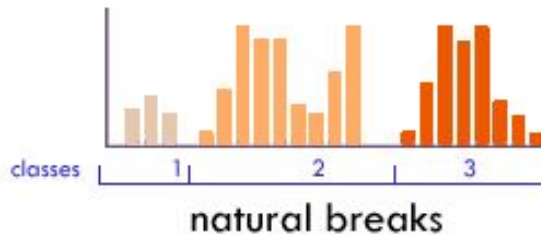
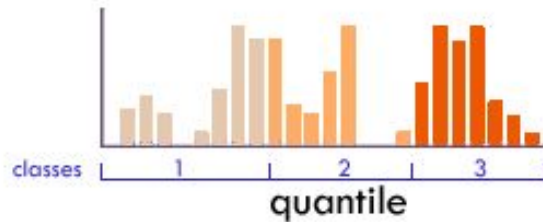
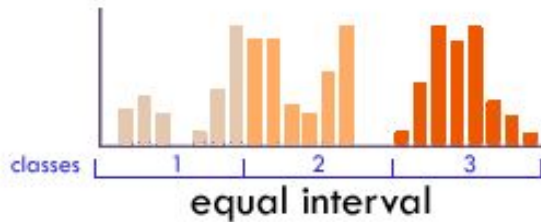




MANUAL

there are many times we need to manually set class breaks. consider:

are there important break points that need to be “hardwired” into your class breaks?



does one of the class breaks need to be the mean? Is this map part of a series?



Joey Cherdarchuk - Obumbratta

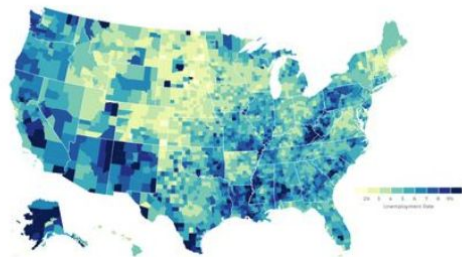
@obumbratta.com

+ Follow

Your map's legend can be turned into a histogram, helping your audience understand the distribution in addition to decoding the colours.



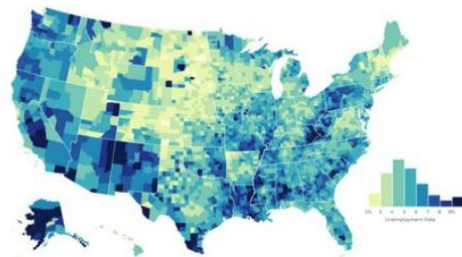
US Unemployment Rates
2016 Annual Averages by County



Source: US Bureau of Labor Statistics



US Unemployment Rates
2016 Annual Averages by County



Source: US Bureau of Labor Statistics

Your map's legend can be turned into a histogram, helping your audience understand the distribution in addition to decoding the colours.

OBUMBRATTA
Interactive Visualization

#VizTips
@obumbratta.com

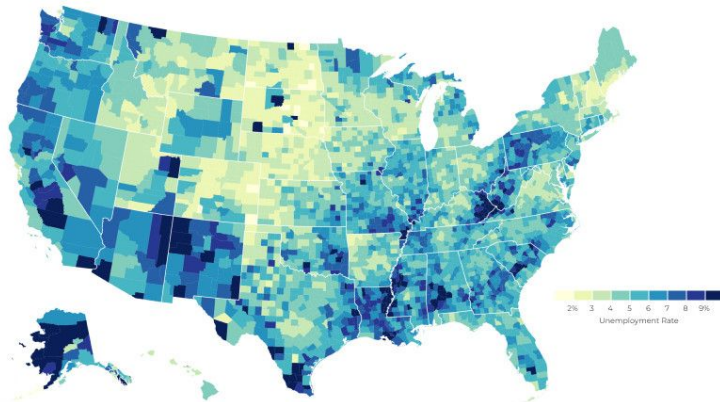
ALT

November 25, 2024 at 1:38 PM Everybody can reply



US Unemployment Rates

2016 Annual Averages by County

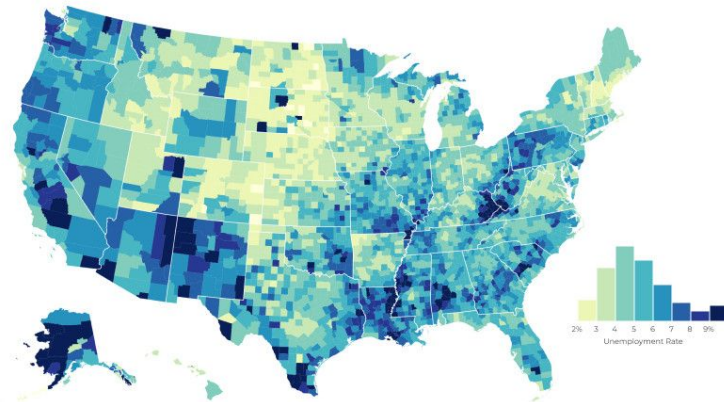


Source: US Bureau of Labor Statistics



US Unemployment Rates

2016 Annual Averages by County



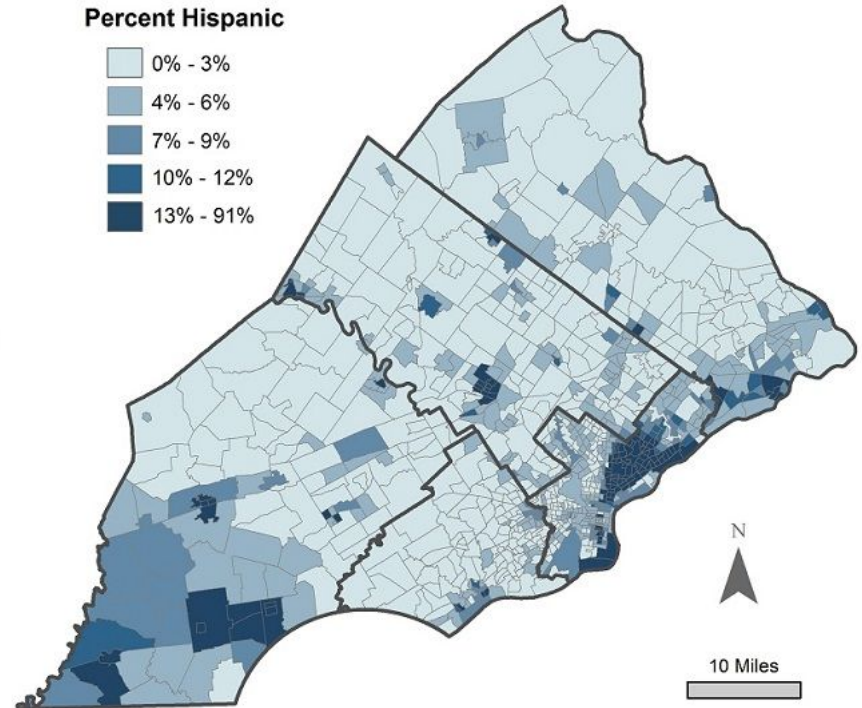
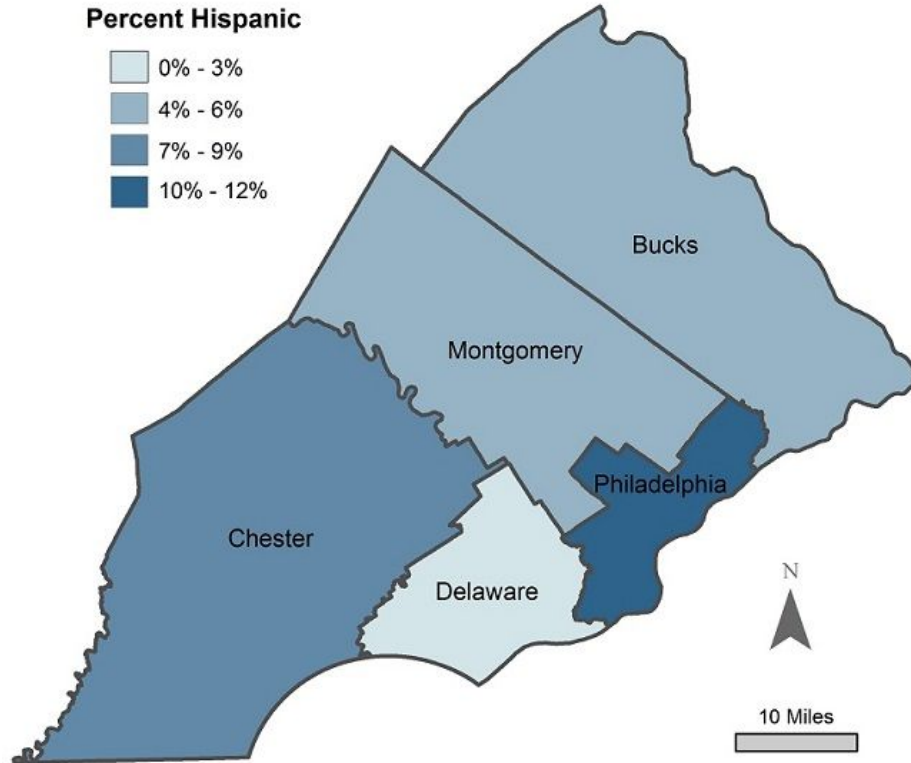
Source: US Bureau of Labor Statistics

Your map's legend can be turned into a histogram, helping your audience understand the distribution in addition to decoding the colours.

**classification can
lead and mislead**

**but so can the
geography itself**

which of these is correct? both? neither??



modifiable areal unit problem

MAUP is a source of statistical bias in aggregated data

it states that patterns in data can be a result of the choice of district boundaries **as much as the real-world patterns in the phenomena being studied**

MAUP is particularly problematic in spatial analysis and choropleth maps

modifiable areal unit problem

MAUP is a source of statistical bias in aggregated data

it states that patterns in data can be a result of the choice of district boundaries **as much as the real-world patterns in the phenomena being studied**

MAUP is particularly problematic in spatial analysis and choropleth maps

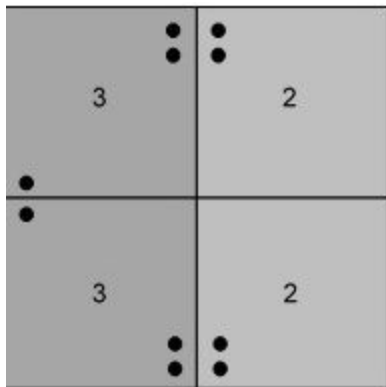
modifiable areal unit problem

MAUP is a source of statistical bias in aggregated data

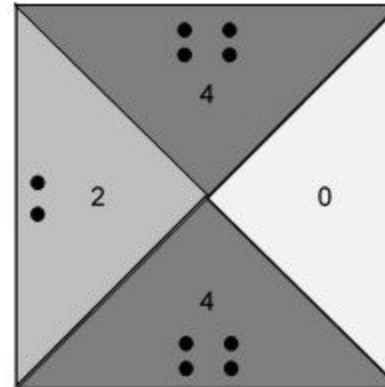
it states that patterns in data can be a result of the choice of district boundaries **as much as the real-world patterns in the phenomena being studied**

MAUP is particularly problematic in spatial analysis and choropleth maps

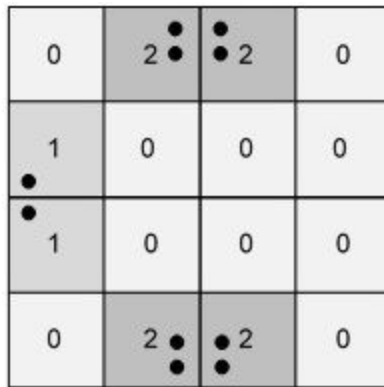
four square districts



four triangular districts

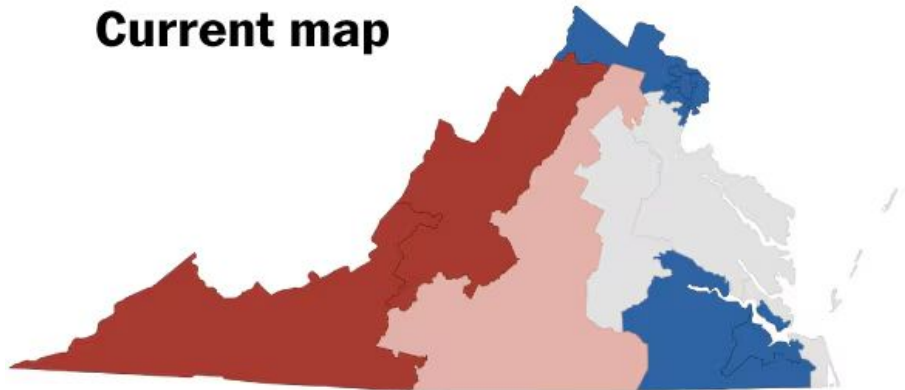


sixteen square districts

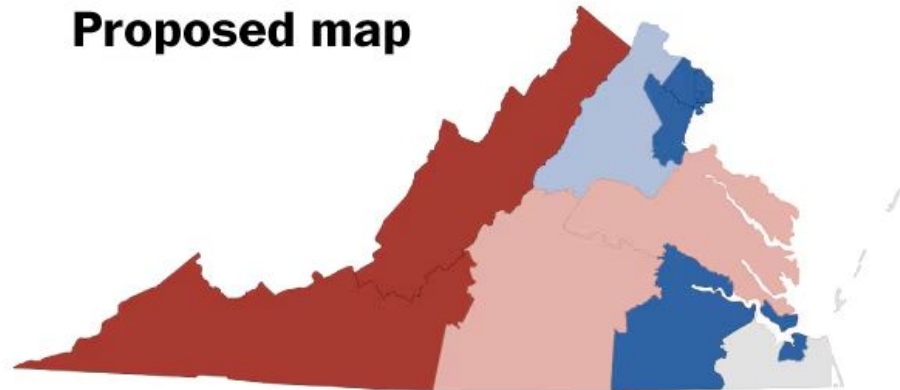


Proposed Virginia redistricting maps leave Spanberger miles away from her congressional seat

Current map



Proposed map



**how should this
affect your
geospatial
analysis?**



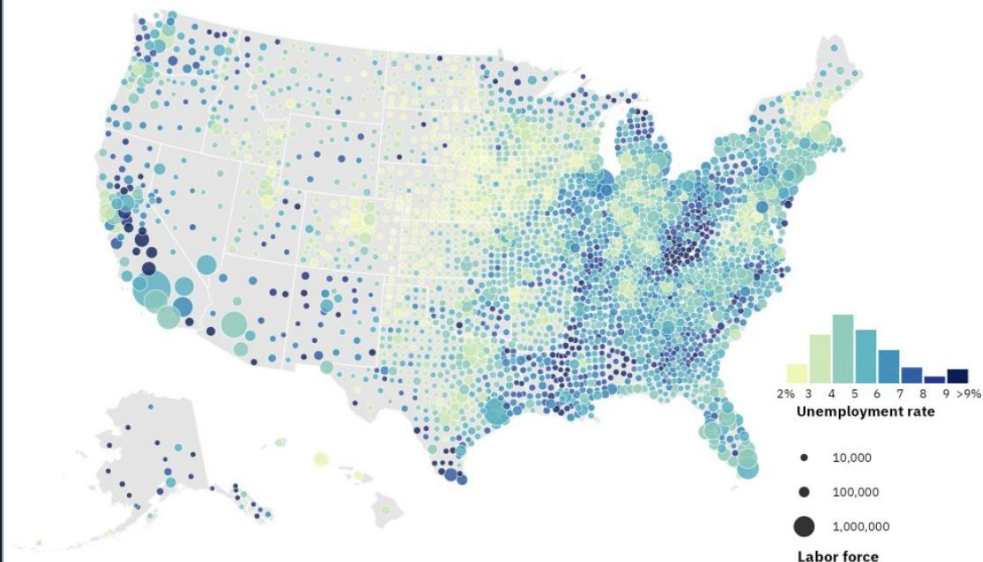
Andrew Heiss
@andrew.heiss.phd

+ Follow

Bonus update! Choropleths are still visually distorted and overrepresent land, so we can instead use points that are sized by population. I added an example at the end to show one way to do this with {sf}'s magical `st_centroid()` [#rstats](#)
www.andrewheiss.com/blog/2025/02...

US unemployment rates

2016 annual averages by county

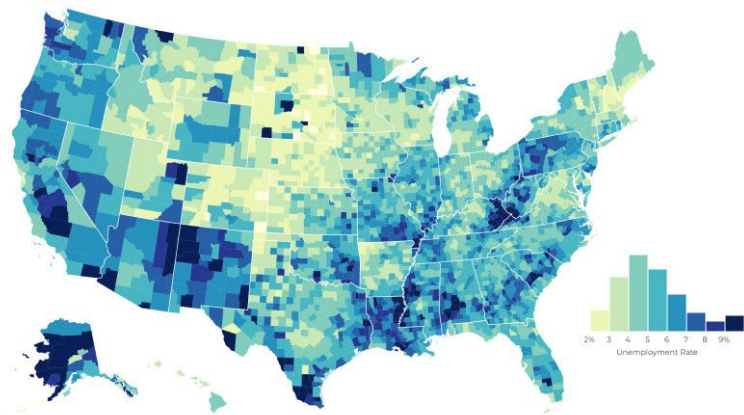


Source: US Bureau of Labor Statistics

ALT

US Unemployment Rates

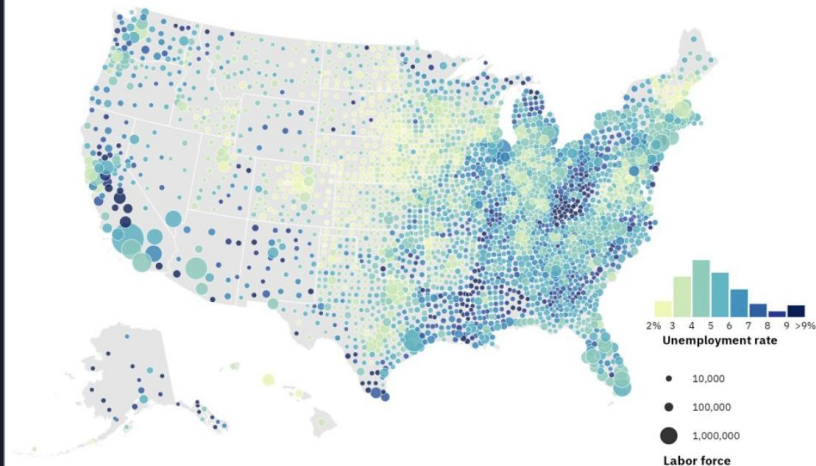
2016 Annual Averages by County



Source: US Bureau of Labor Statistics

US unemployment rates

2016 annual averages by county



Source: US Bureau of Labor Statistics

ALT

Mapping dispossession: Eviction, foreclosure and the multiple geographies of housing instability in Lexington, Kentucky

Taylor Shelton ✉

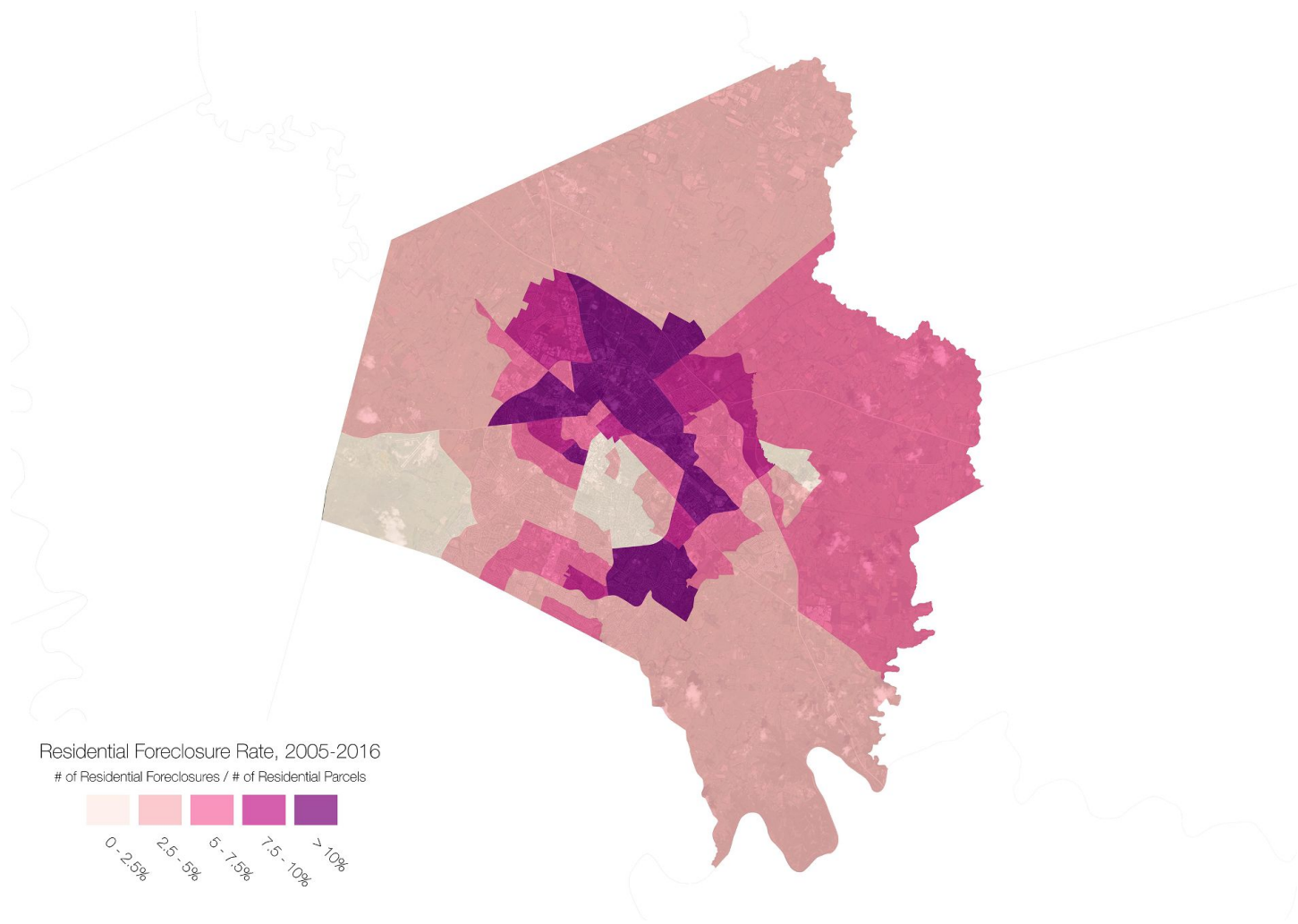
Show more ✓

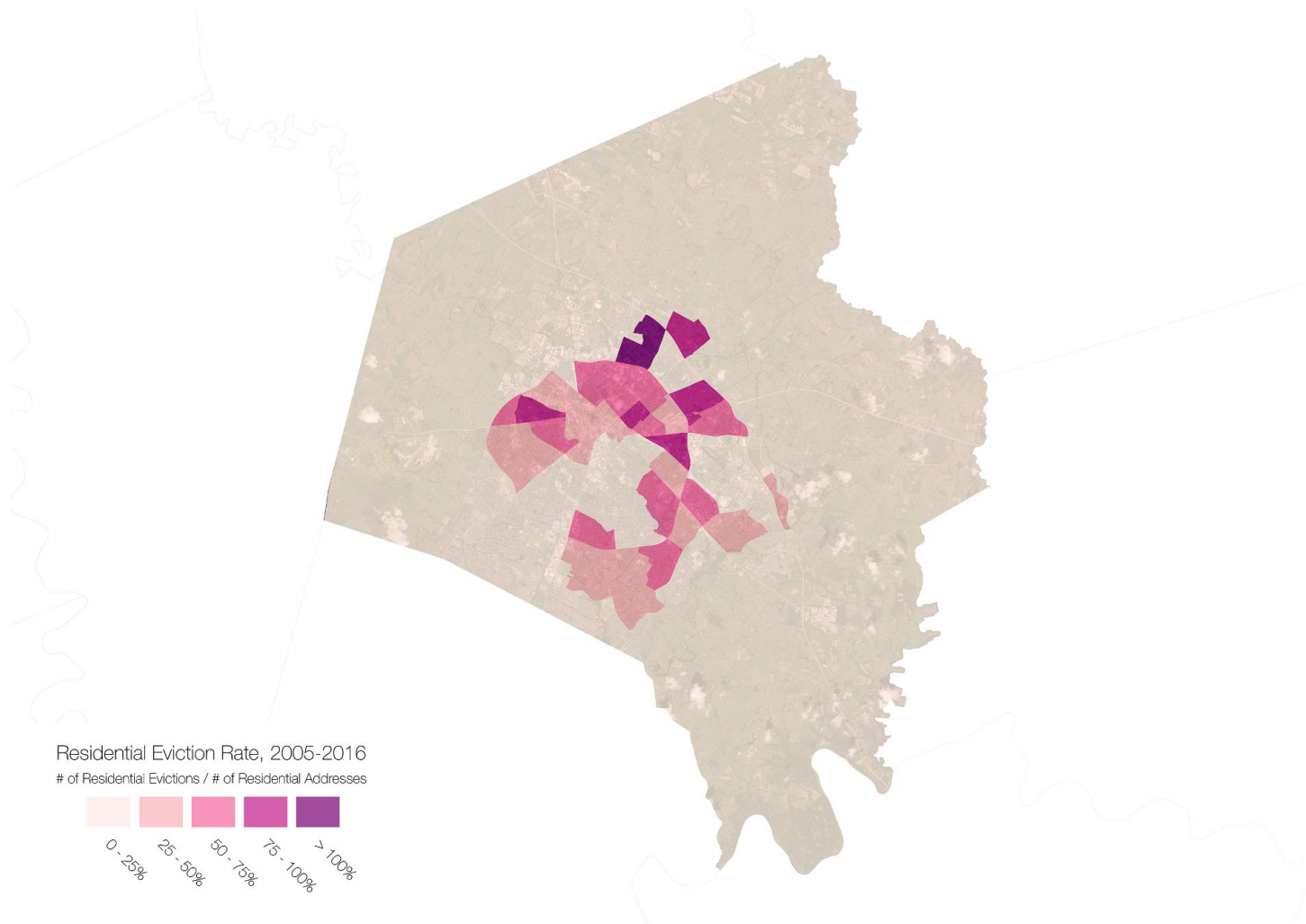
+ Add to Mendeley  Share  Cite

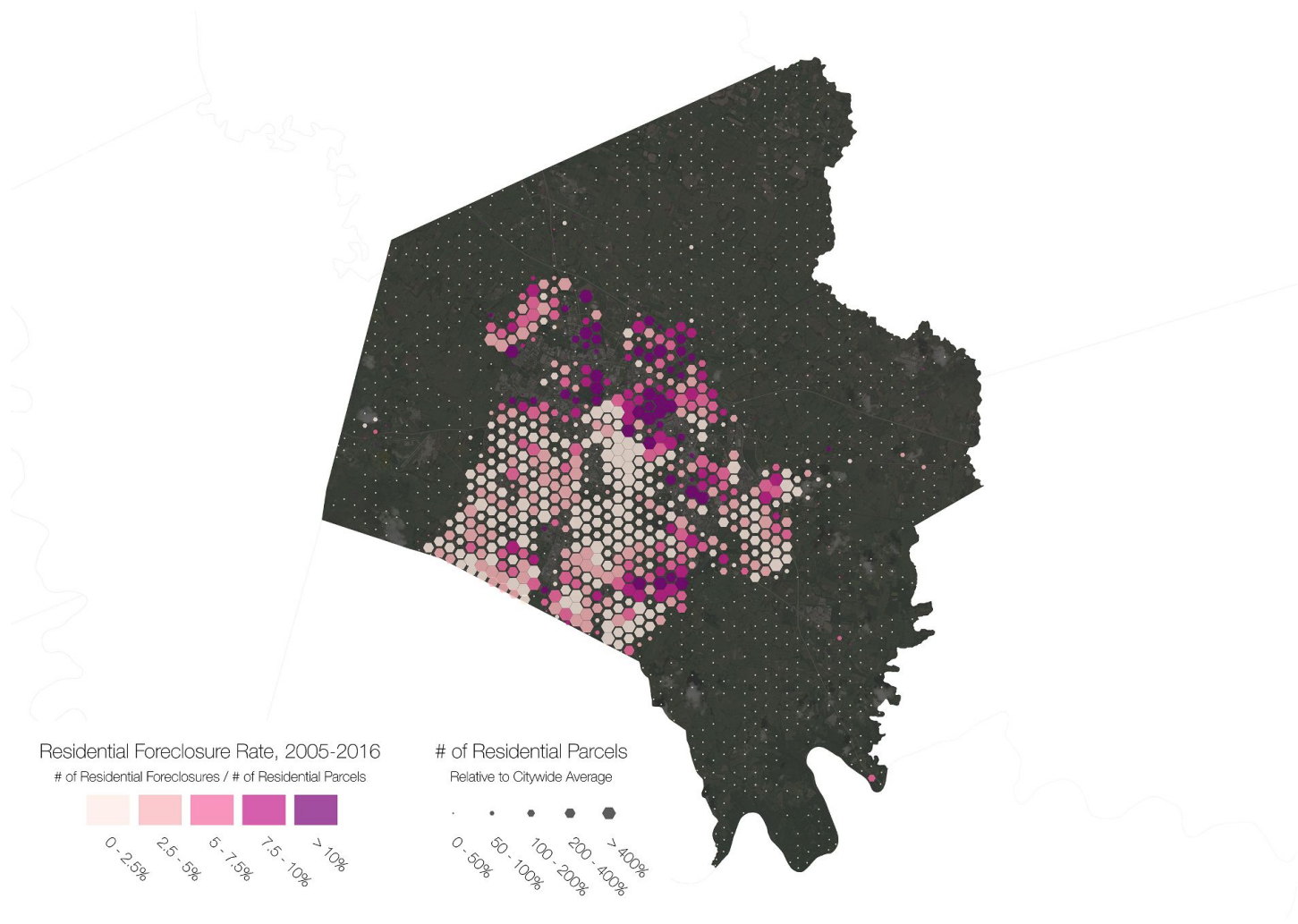
<https://doi.org/10.1016/j.geoforum.2018.09.028> ↗

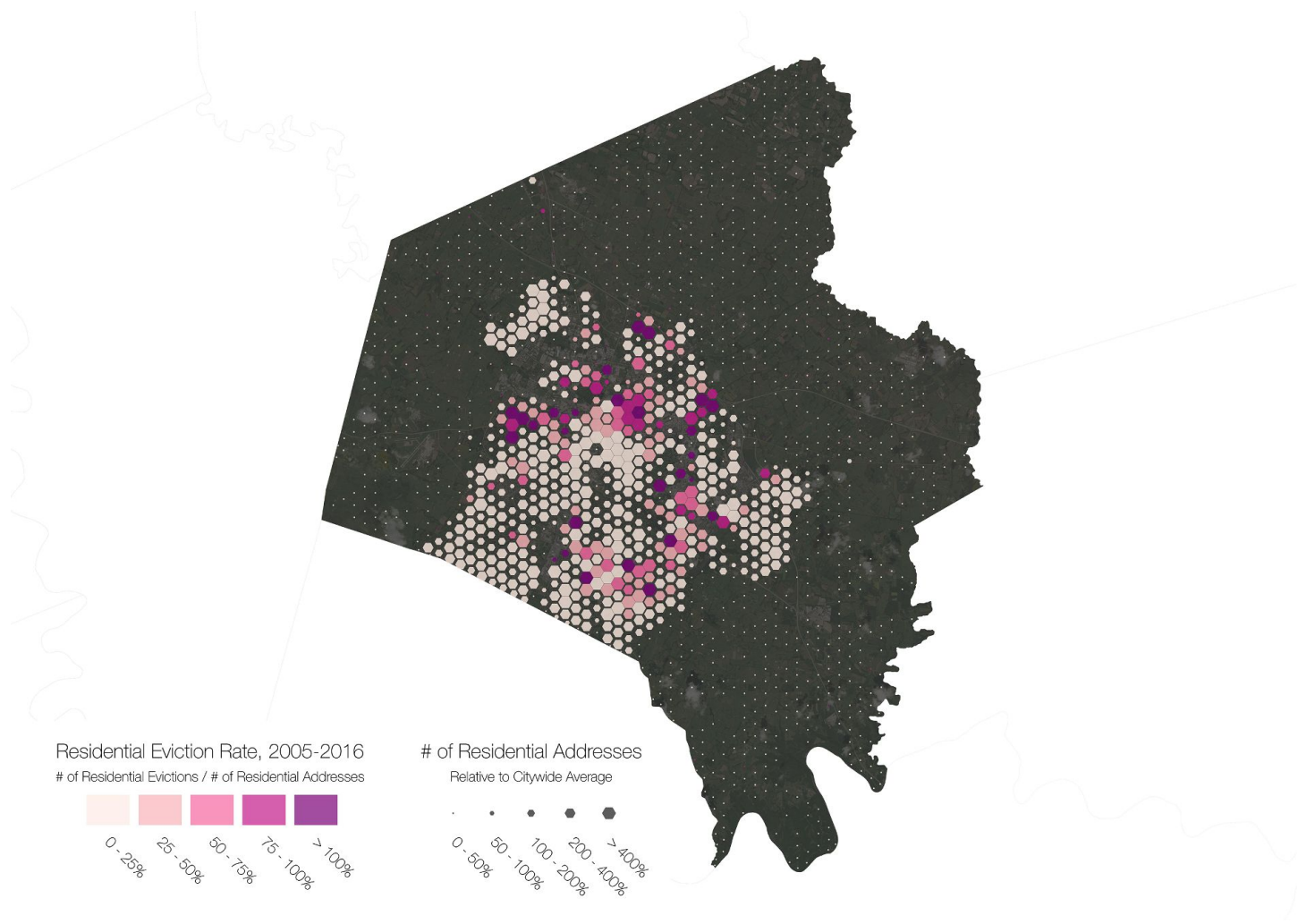
Get rights and content ↗

● Full text access

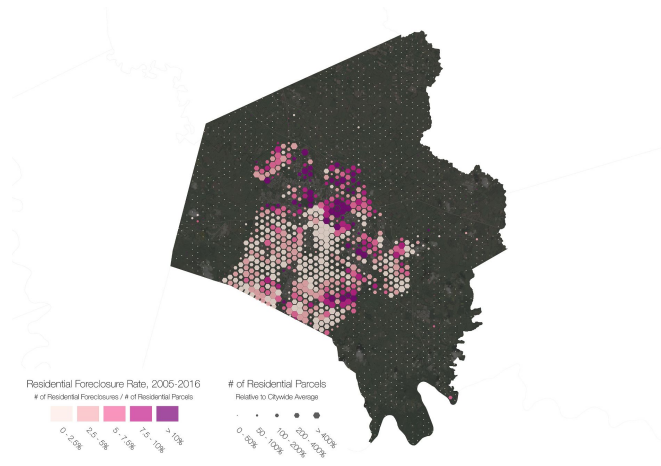
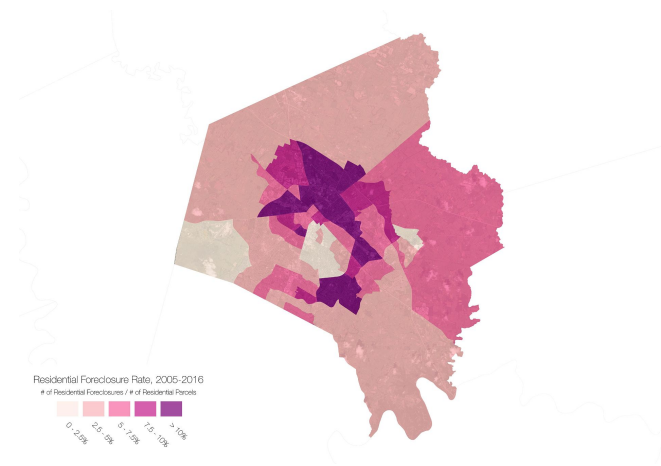




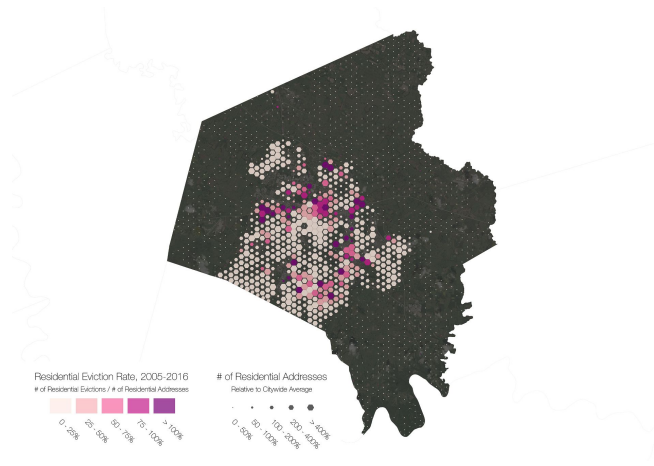
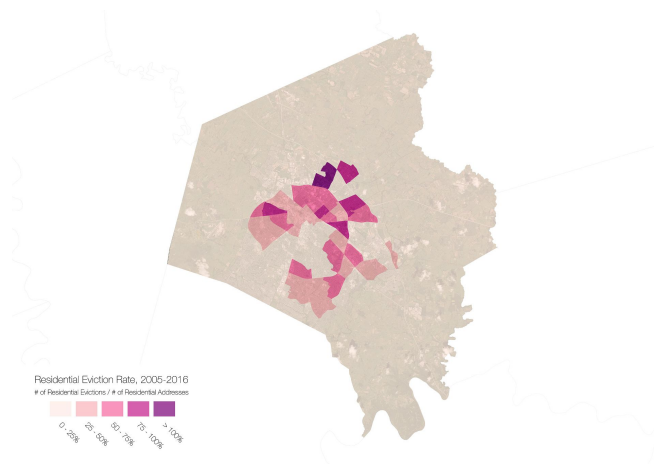




FORECLOSURES



EVICTIONS



Abstract

Housing is increasingly recognized as playing a crucial role in shaping any number of other social processes and domains. But fundamental inequalities in housing mean that the benefits of quality housing accrue unequally to different groups of people. In order to explore how housing works not only to reflect, but also produce, socio-spatial inequality, this paper explores the geographies and temporalities of housing dispossession in Lexington, Kentucky by drawing on an extensive dataset of all evictions and foreclosures in the city from 2005 to 2016. The paper demonstrates that while mortgage foreclosure has tended to dominate discussions of housing dispossession since the financial crisis of 2007–2008, residential evictions are both far more widespread and much more consistent over time. The paper also concludes that while foreclosure rates are much more strongly correlated with variables of racial and class segregation, evictions are more spatially concentrated across multiple scales. The paper demonstrates that many of those places experiencing the most acute forms of housing dispossession are not gentrifying inner-city neighborhoods, but rather persistently or increasingly impoverished inner-ring suburbs. Ultimately, this paper argues that these spatialities are the result not of individual failings, but rather of powerful actors in the local housing market acting in conjunction with broader forces of racialized capitalist urbanization.

**different
boundaries can
suggest different
conclusions about
the same data**

**so can different
data classification
schemes**

**before moving on,
let's discuss how
these kinds of
maps are made**

shapefile

- geospatial feature data
- consists of multiple files
- drag and drop into ArcGIS Pro, instantly render

spreadsheet

- tabular attribute data
- single file (.csv, .xlsx)
- need to plot, either as XY data or through geocoding

Name	Date modified	Type	Size
 Admin.dbf	3/18/2011 2:36 PM	DBF File	592 KB
 Admin.prj	3/18/2011 2:36 PM	PRJ File	1 KB
 Admin.sbn	3/18/2011 2:36 PM	SBN File	26 KB
 Admin.sbx	3/18/2011 2:36 PM	SBX File	2 KB
 Admin.shp	3/18/2011 2:36 PM	SHP File	7,350 KB

shapefile

- geospatial feature data
- consists of multiple files
- drag and drop into ArcGIS Pro, instantly render

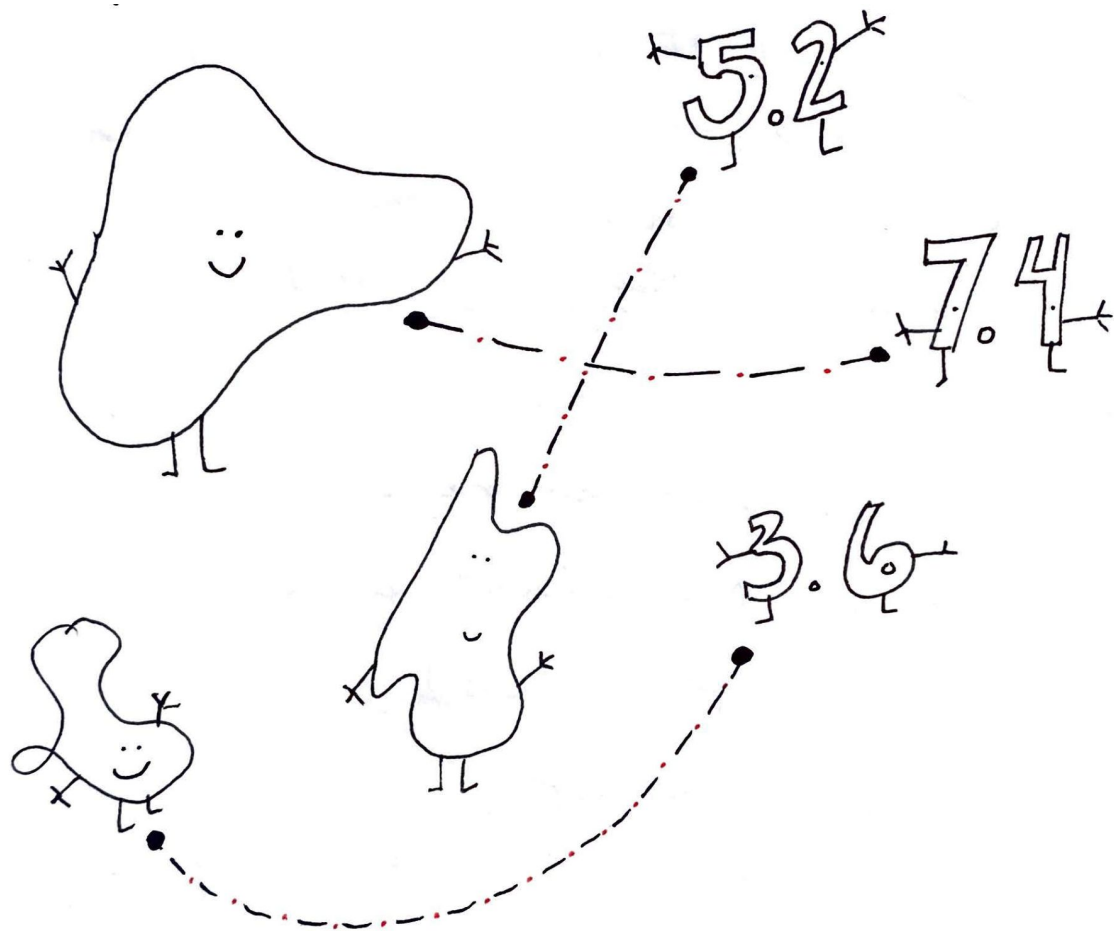
spreadsheet

- tabular attribute data
- single file (.csv, .xlsx)
- need to plot, either as XY data or through geocoding

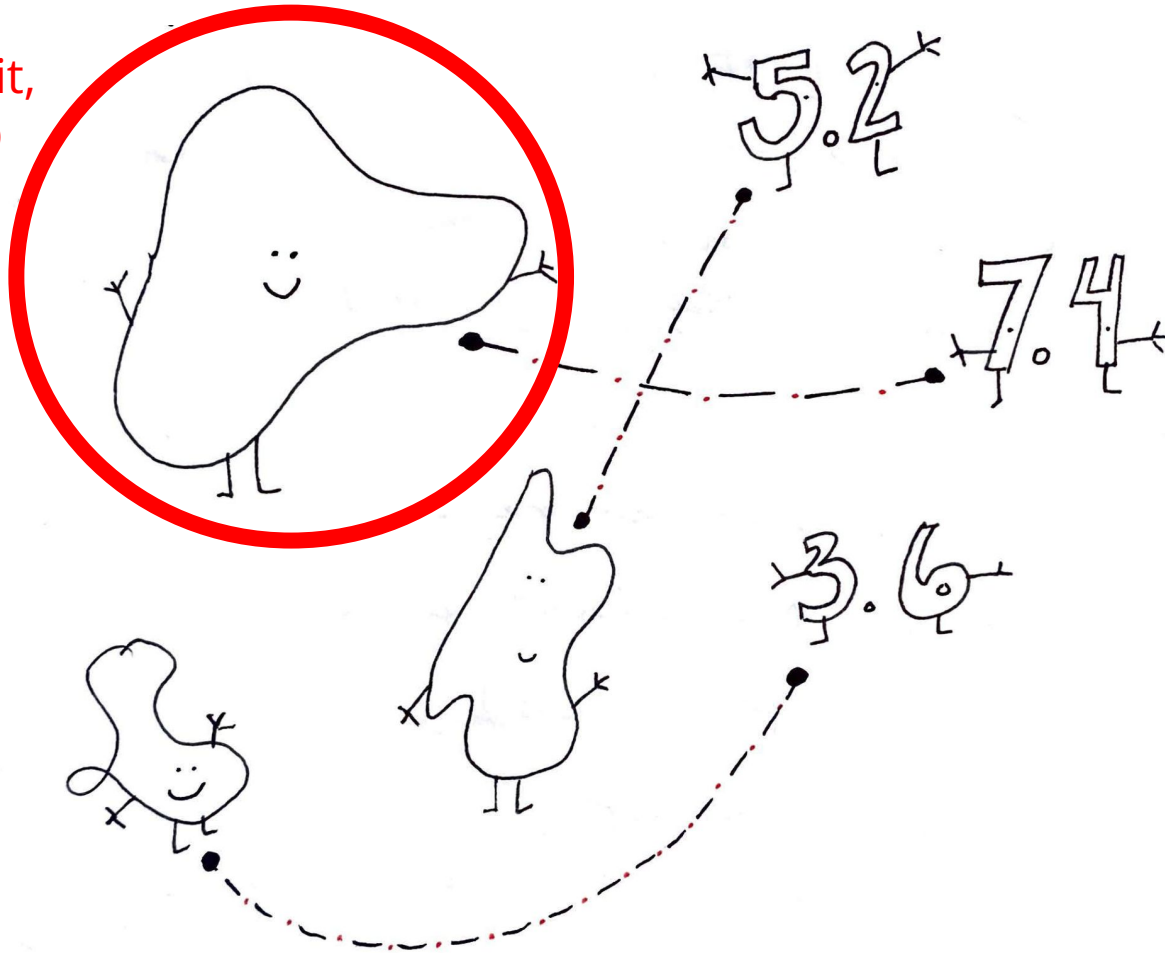
A	B	C	D	E	F	G	H	I
id	name	host_id	host_profile	host_name	neighbourhood	neighbourhood	latitude	longitude
2384	Hyde Park	2613	146250631	Rebecca		Hyde Park	41.7879	-87.5878
7126	Tiny Studio	17928	146250685	Sarah		West Town	41.90166	-87.6802
10945	The Biddle	33004	146250720	Elizabeth		Lincoln Park	41.91196	-87.6398
12140	Lincoln Park	46734	146250756	Shar And Robert		Lincoln Park	41.92357	-87.6495
71930	Rest, Relax	334241	146251125	Michael And Veronica		West Town	41.89615	-87.6793
94450	Comfy Gar	504470	146252016	Mark		Edgewater	41.979601	-87.6652
145659	Trendy Ro	683529	146252605	Joe		North Center	41.94342	-87.6812
145690	4BR/2BR	683529	146252605	Joe		Lincoln Park	41.91289	-87.6386
189821	Best in Ch	899757	146253272	Meighan		Logan Square	41.92921	-87.7022

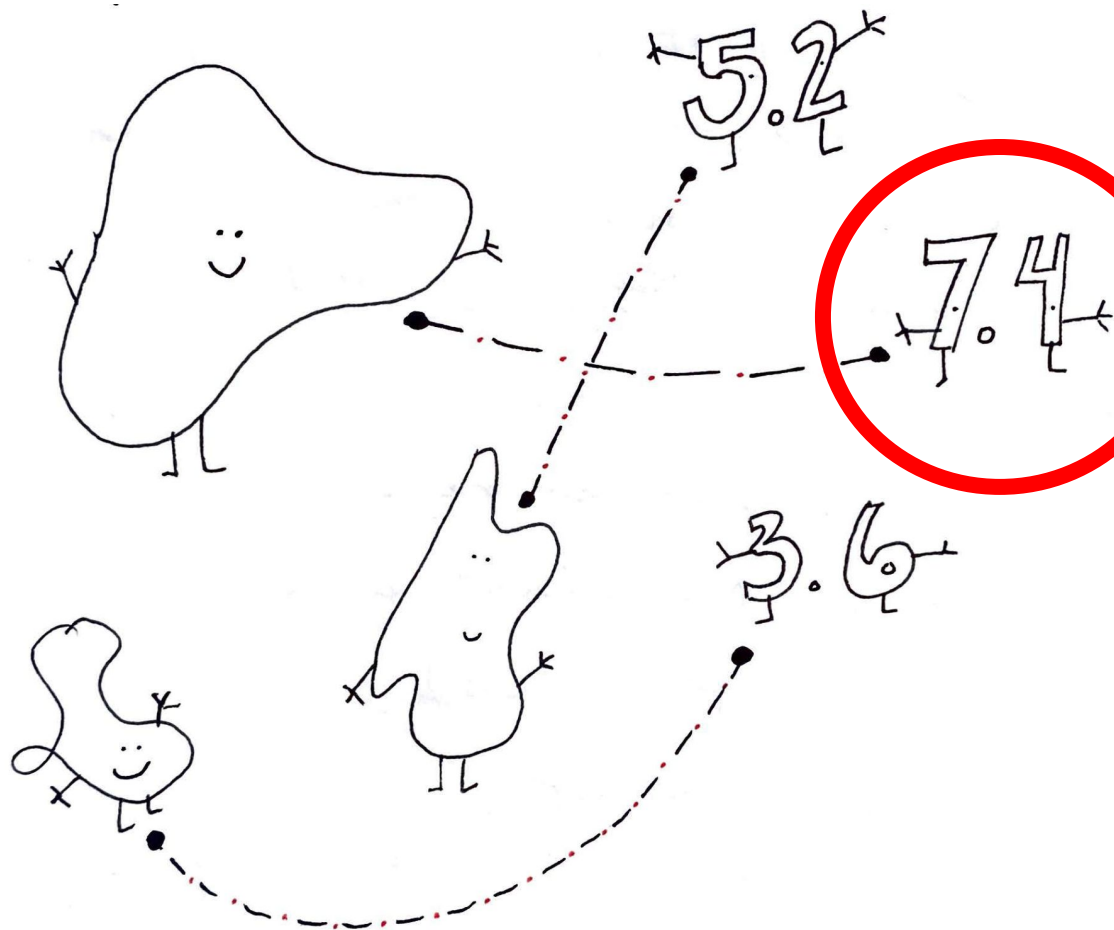
**sometimes, we
want to connect
tabular data to
spatial data**

this is called a
table join

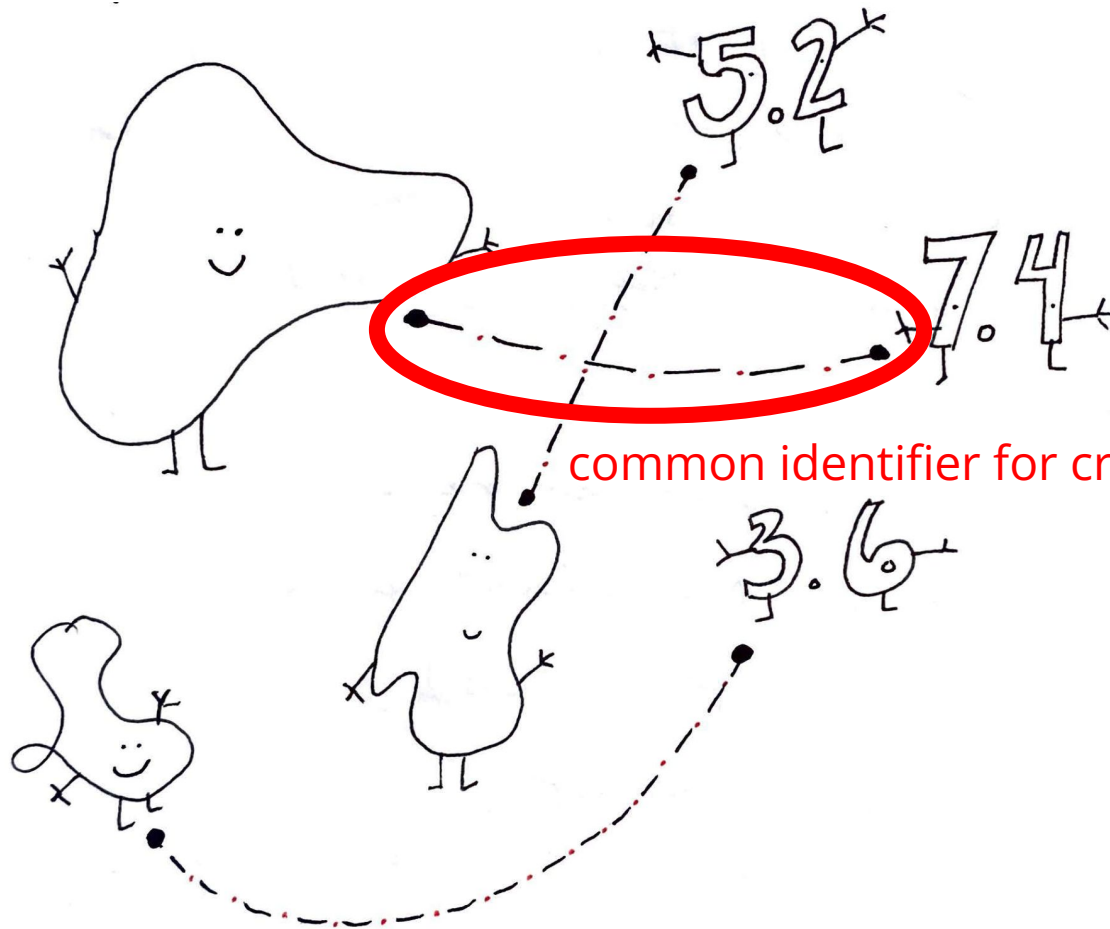


areal unit,
like a zip
code or
census
tract





attribute from
a table that
we want
to join with
the feature



**this process is
common for
making**
choropleth maps