

PLAN 590

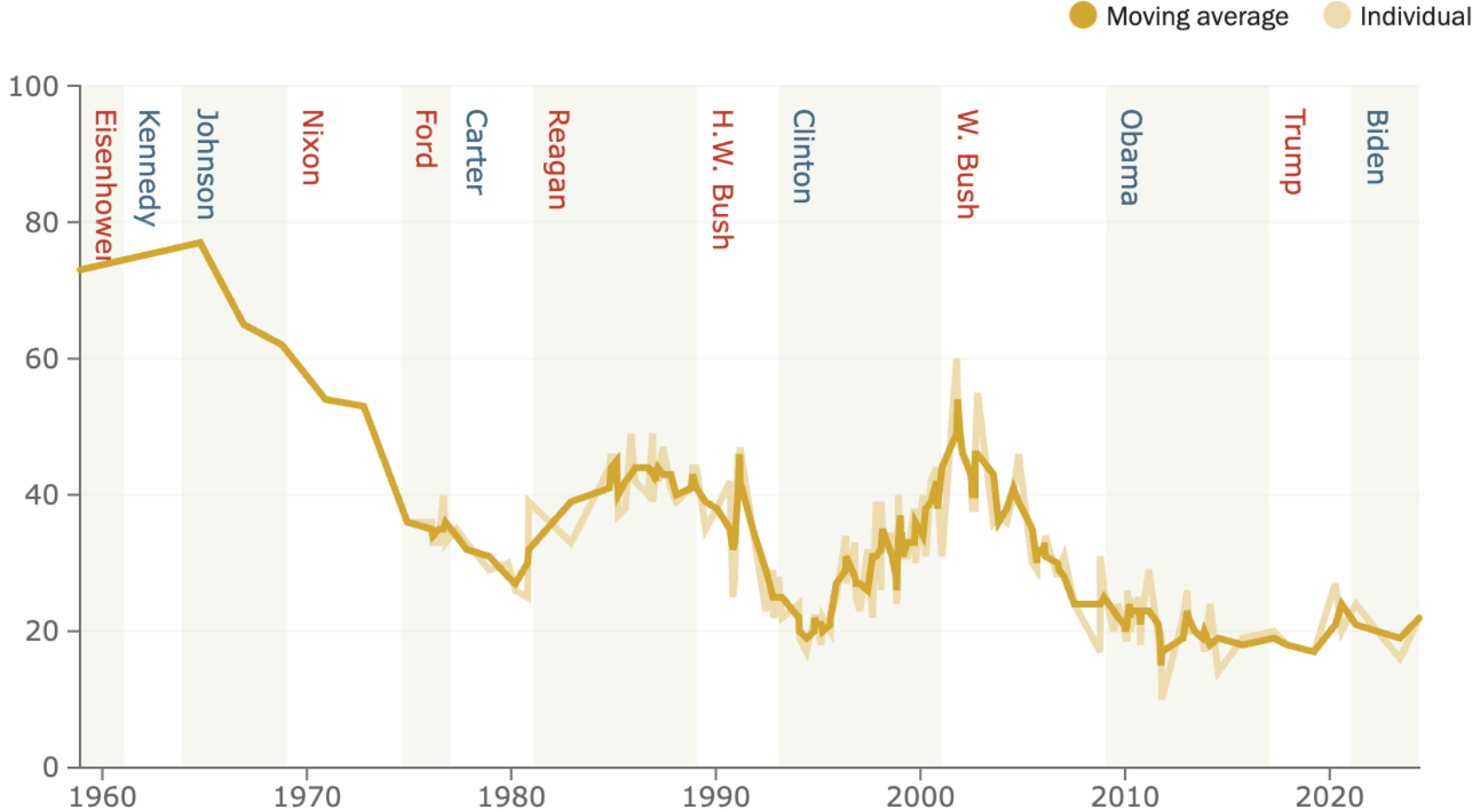
January 17, 2025

# Communicating with data

Week 1

# Public trust in government near historic lows

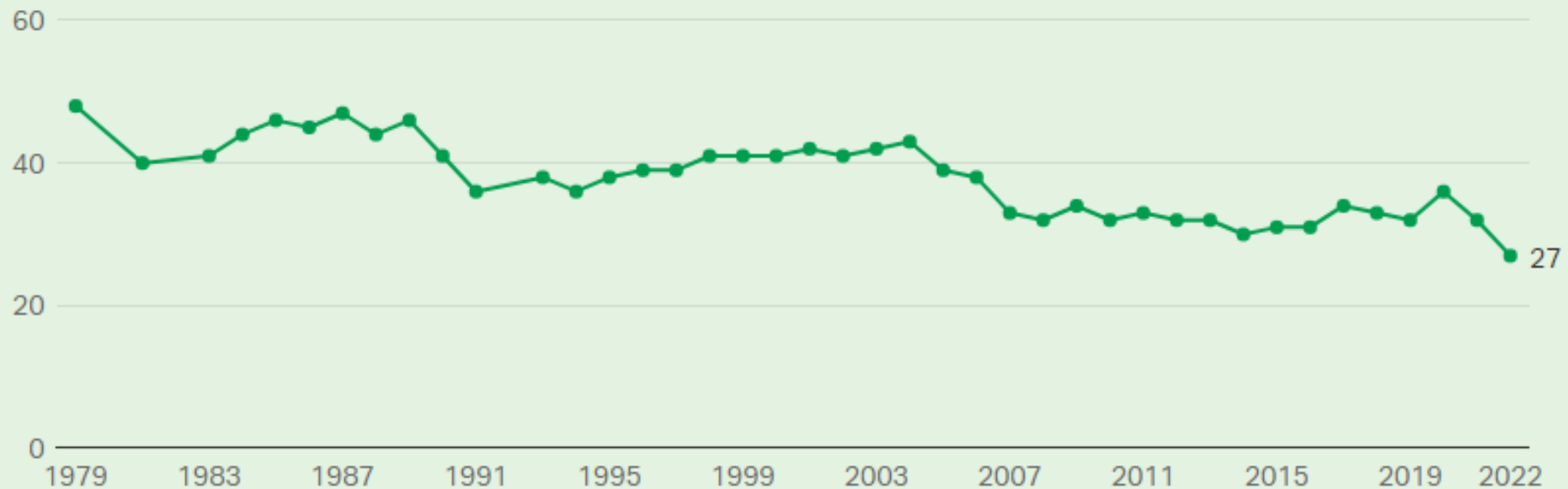
% who say they trust the government to do what is right just about always/most of the time



Sources: Pew Research Center, National Election Studies, Gallup, ABC/Washington Post, CBS/New York Times, and CNN surveys.

## Average Confidence in Major U.S. Institutions, 1979-2022

Figures are the average percentage of U.S. adults who have a great deal or quite a lot of confidence across institutions asked consistently by Gallup.



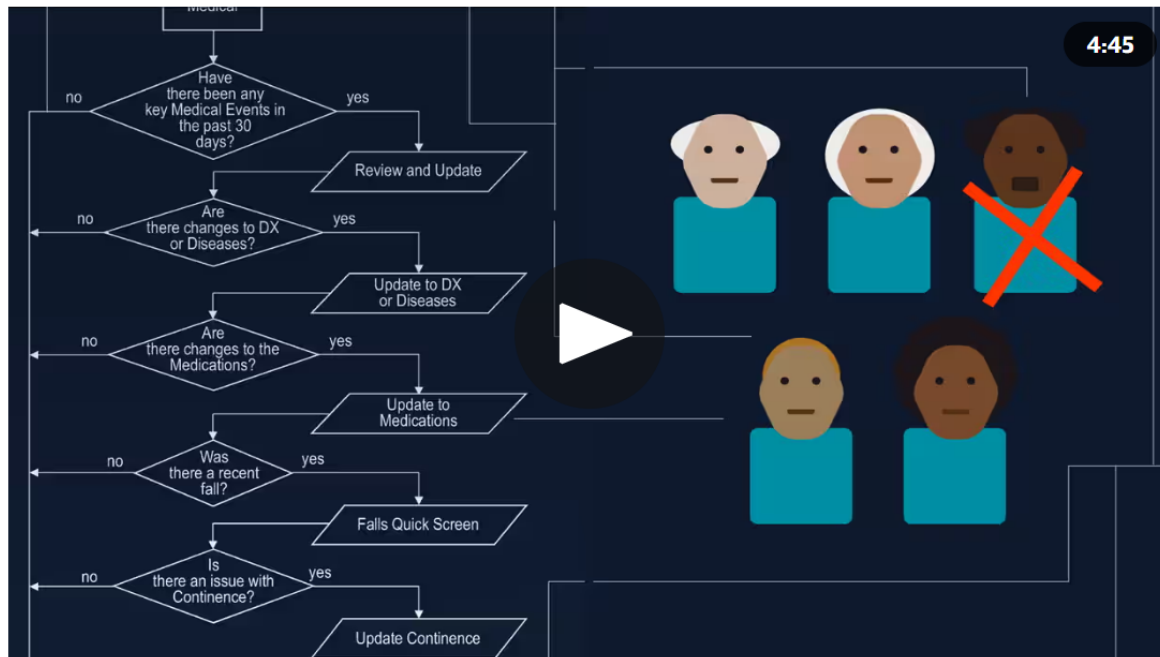
The average includes only those asked consistently by Gallup since 1979 or 1993 (when the list of institutions was expanded). The institutions asked consistently since 1979 are the church/organized religion, the military, the Supreme Court, banks, public schools, newspapers, Congress, organized labor and big business. In 1993 the list of institutions was expanded to include television news, the presidency, the police, the medical system and the criminal justice system.

[Get the data](#)

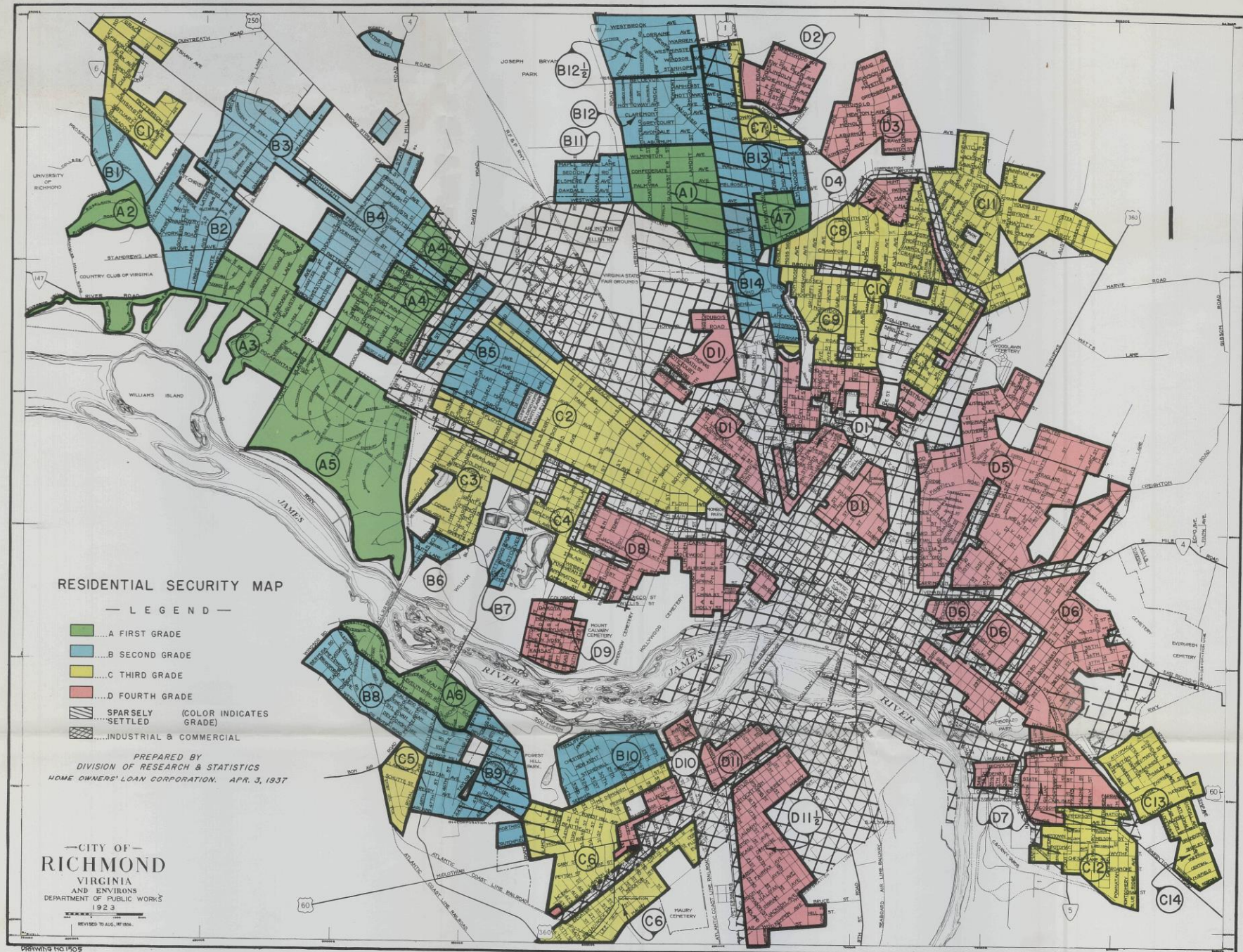
GALLUP

# What happened when a 'wildly irrational' algorithm made crucial healthcare decisions

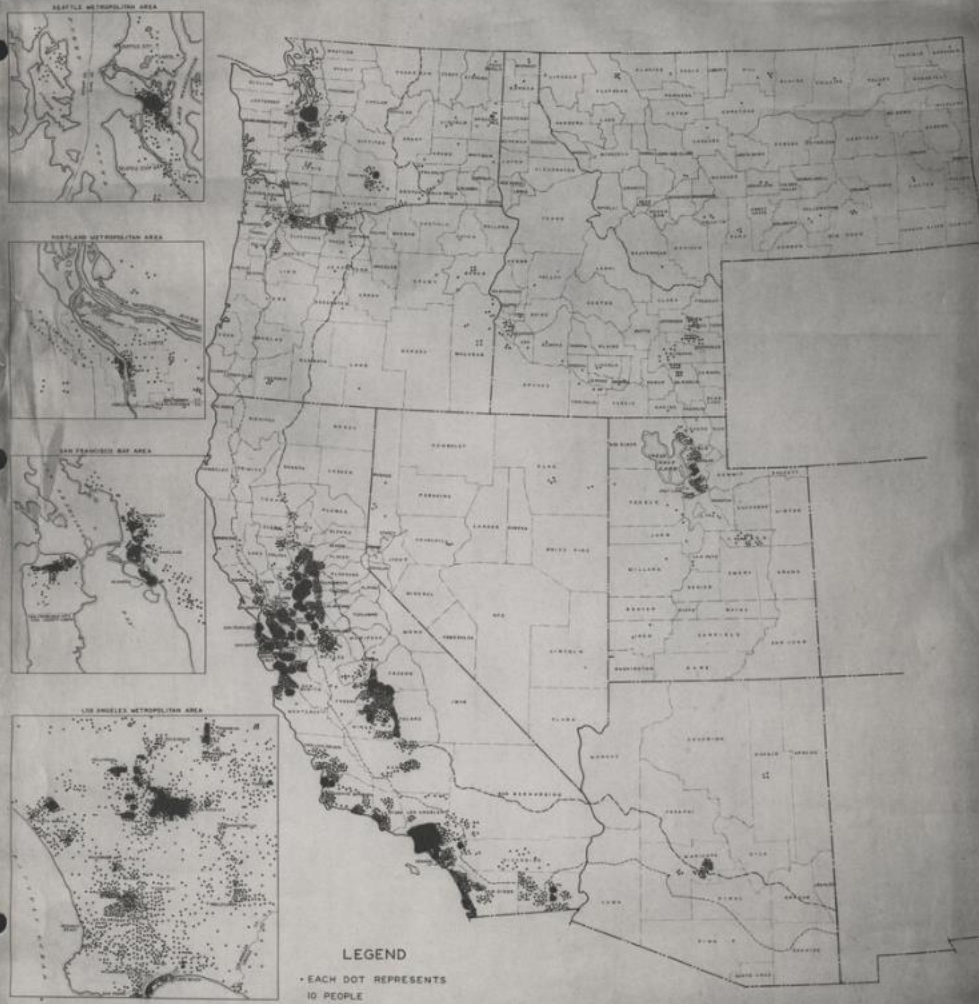
Advocates say having computer programs decide how much help vulnerable people can get is often arbitrary - and in some cases downright cruel



Advocates for people with disabilities say that deciding care algorithmically fails to consider the subtleties of individuals' situations. They worry that decisions get made in a black box with patients having no way of knowing why, thus making rulings hard to challenge.



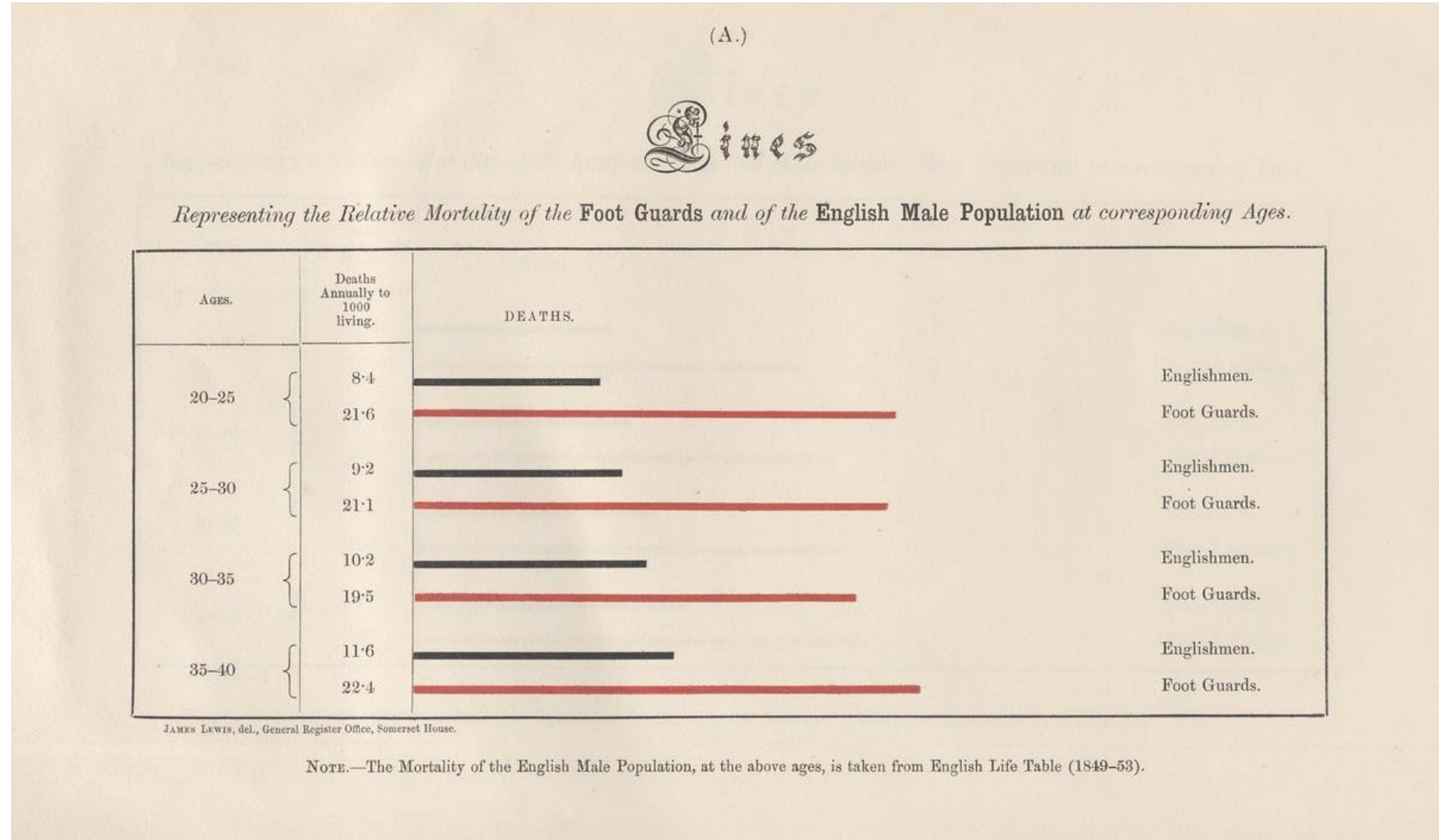
# JAPANESE POPULATION WESTERN DEFENSE COMMAND AREA: 1940



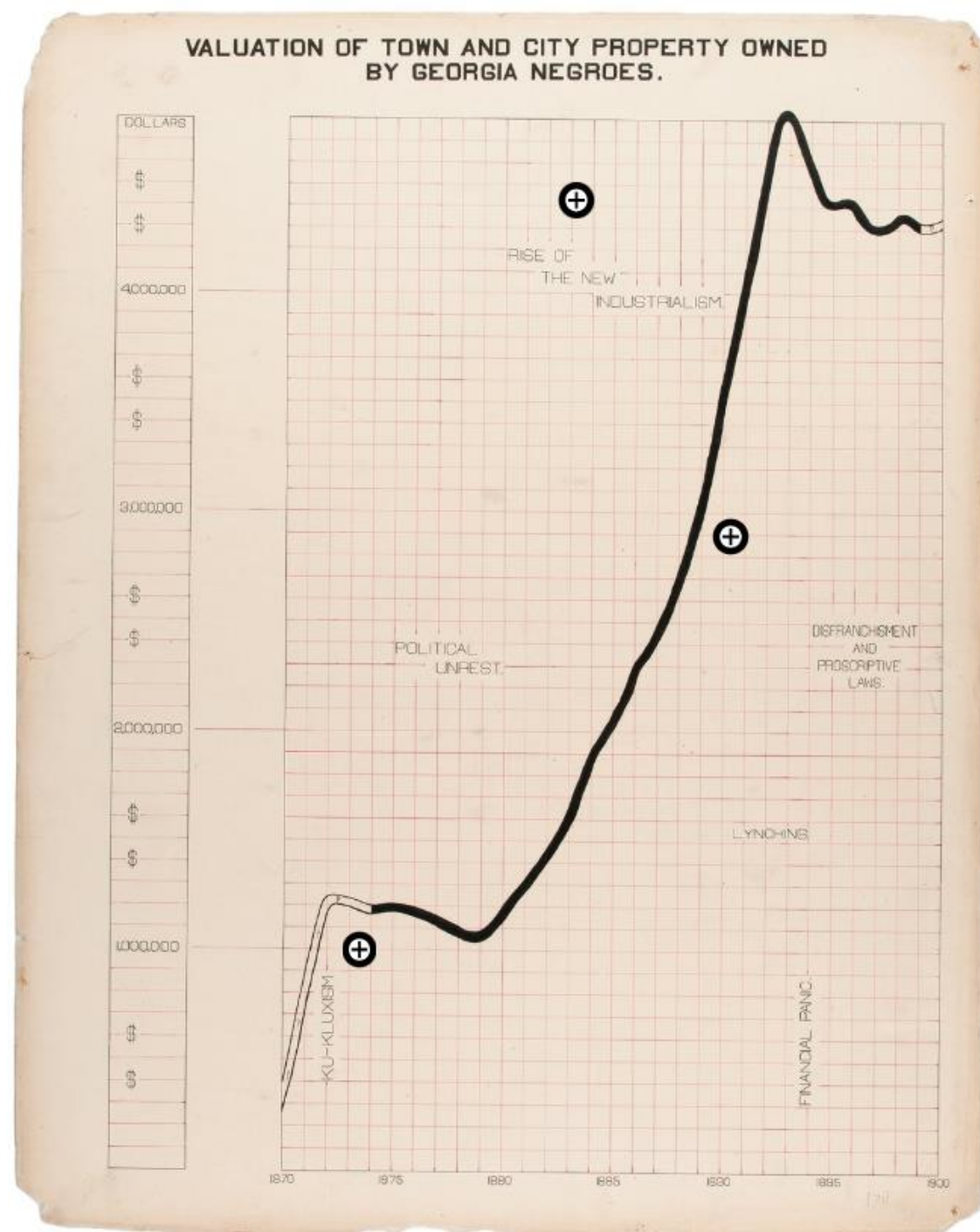
SOURCE: UNITED STATES BUREAU OF THE CENSUS

WESTERN DEFENSE COMMAND AND STUDY AREA  
WARREN G. COLEMAN, CHIEF, STATISTICAL DIVISION

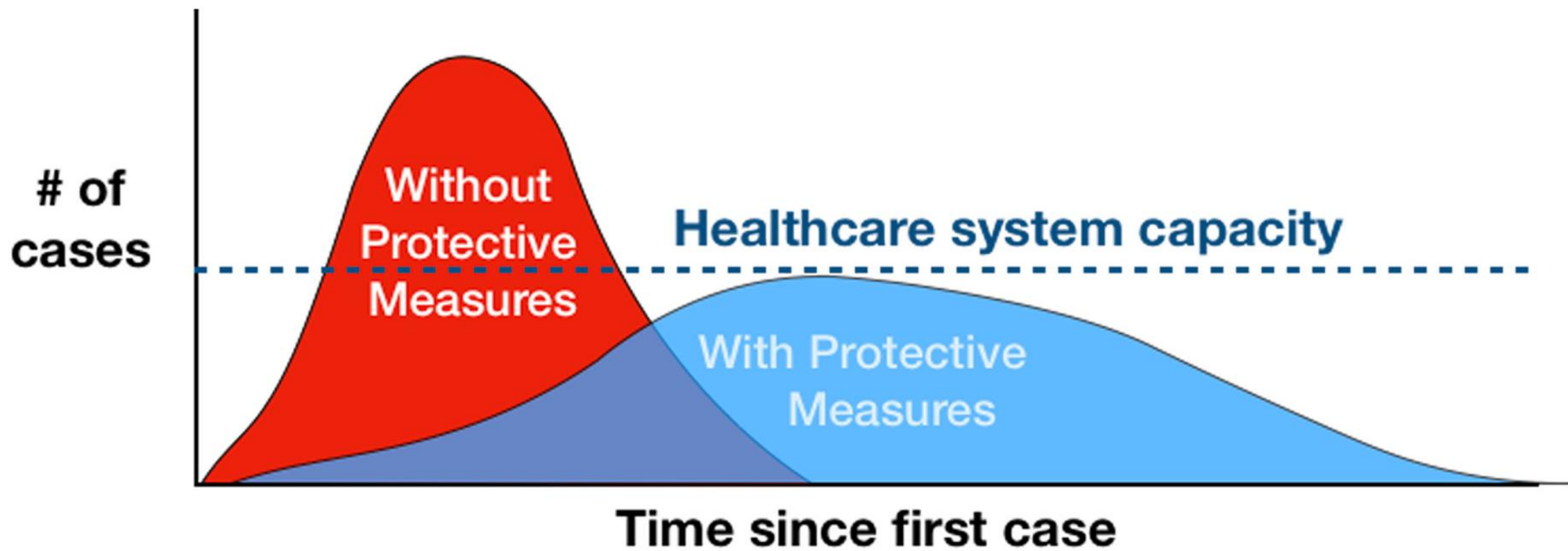
# Florence Nightingale



# W. E. B. Du Bois







*Adapted from CDC / The Economist*

# My ask of you

- Focus on decisions rather than tools
- Practice and experiment (rules will only take you so far!)

# My promise to you

- Transparency
- Responsiveness

# Schedule

## Day 1: January 17

- Setting goals for your data visualizations
- Defining your audience
- Visual Processing & Cognitive Load
- How to choose the right visualization for your data (chart bootcamp)

## Day 3: January 31

- Plain language: What it is, why it's important, and how to use it
- Writing about numbers
- What makes a compelling data story
- Mapping your data characters

## Day 2: January 24

- XRAY discussion
- Data Visualization Research: What Works
- Enough color theory to get by
- Better legends in your data visualization
- Titles, annotations, and other texts
- “Fix my chart” exercise

## Day 4: February 7

- Accessibility in data visualization
- Equity in data visualization
- Data visualization in the workplace
- Workshop time for memos

# Final Memo

Choose an analytical topic and write a memo for a data stakeholder (e.g., a colleague, client, or elected official). Your memo should include:

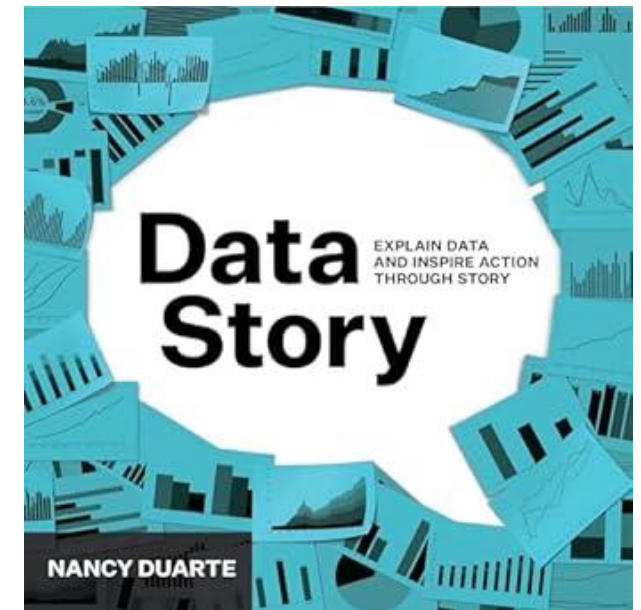
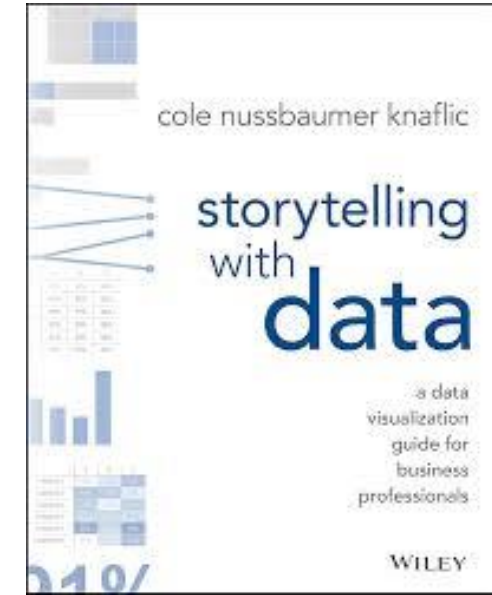
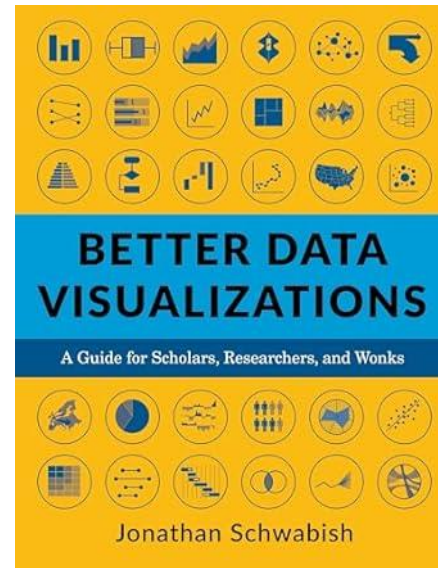
1. **The question or problem** you are addressing.
2. **Your data sources** and the approach you used to analyze the data.
3. **Your findings** and any potential limitations.

**Due Date:** February 21

**Word Limit:** No more than 2,500 words.

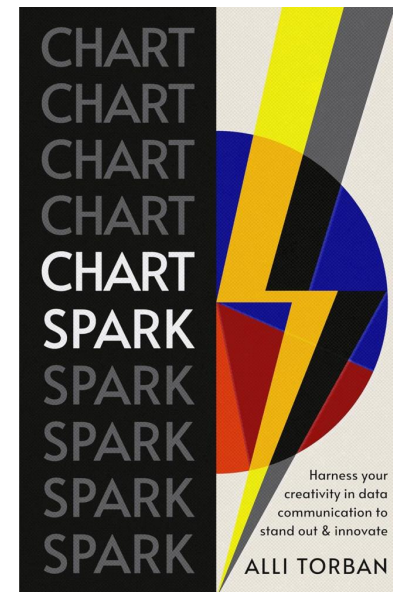
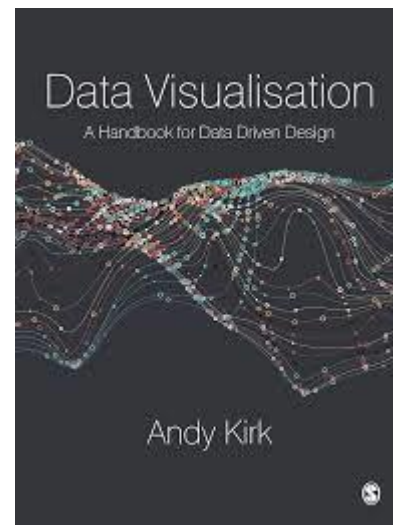
Use as many charts or visuals as needed, but only include those essential to telling your data story.

# The sources for this class



Datawrapper

Blog

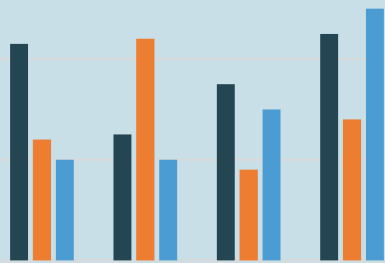


# Why visualize data?

## Discover

I want to **discover** insights about my data.

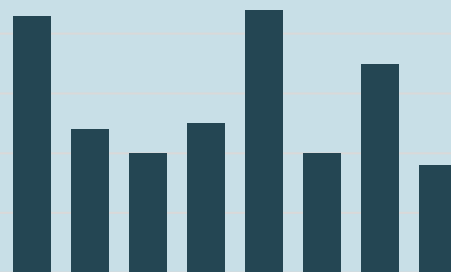
e.g. charts made during analysis



## Inform

I want to **inform** others about my data.

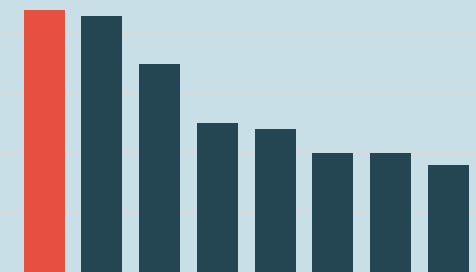
e.g. dashboards, regular reports



## Educate

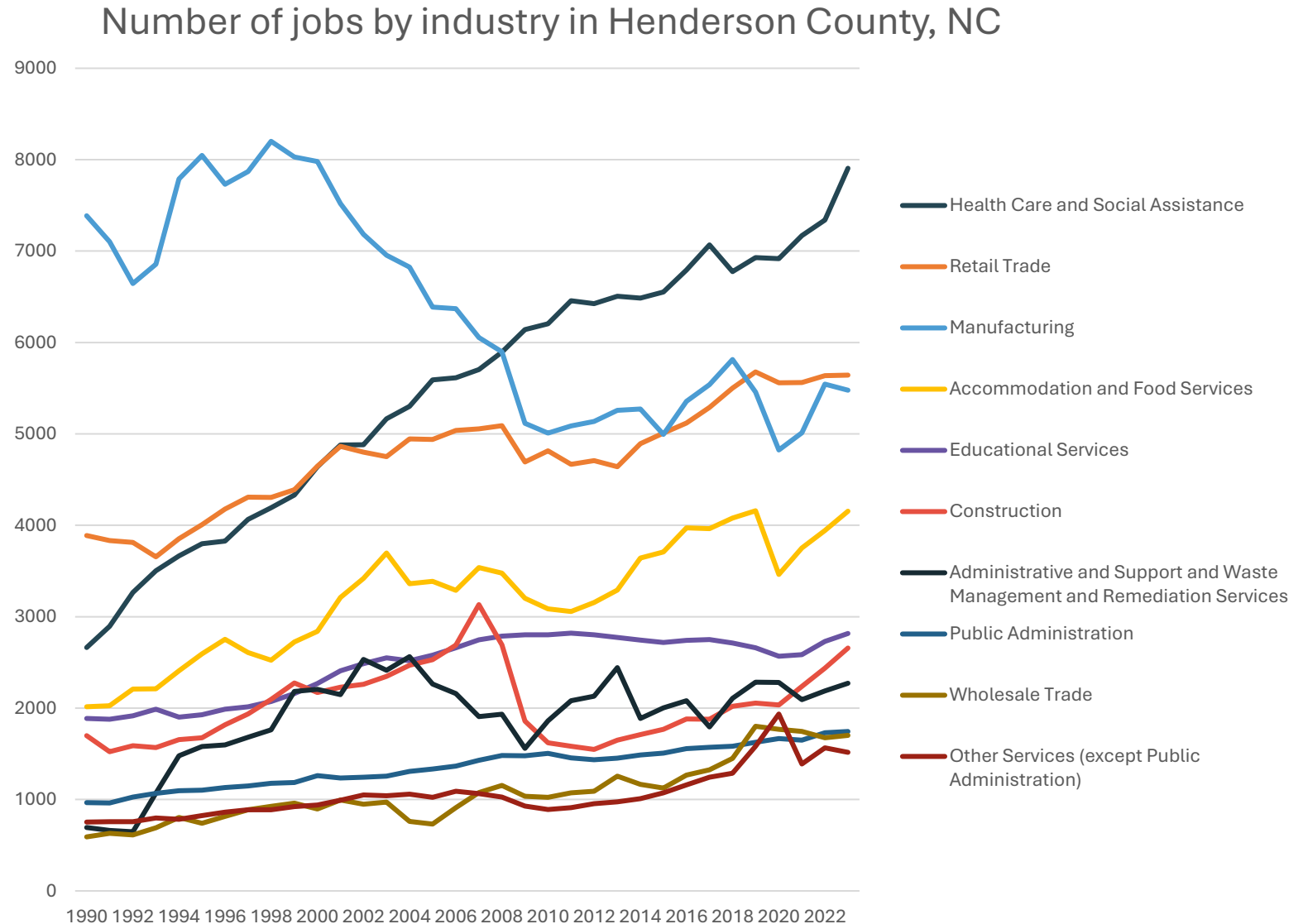
I want to **educate** others about my data.

e.g. data storytelling



# Visualizing data to discover

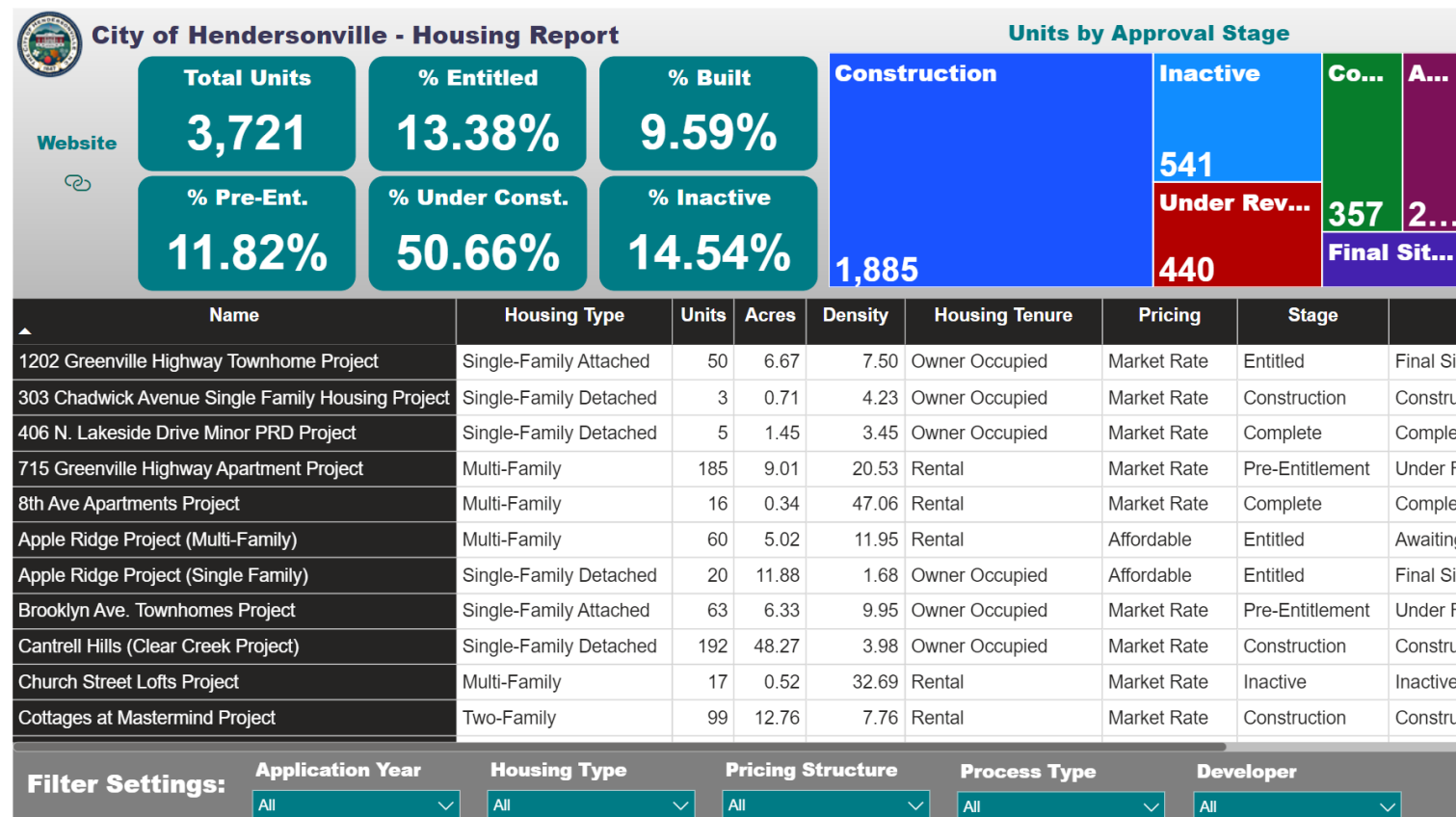
- The audience is **you**.
- Keep them **to yourself!**



Source: NC Commerce

# Visualizing data to inform

- Data is organized to allow audience to explore.

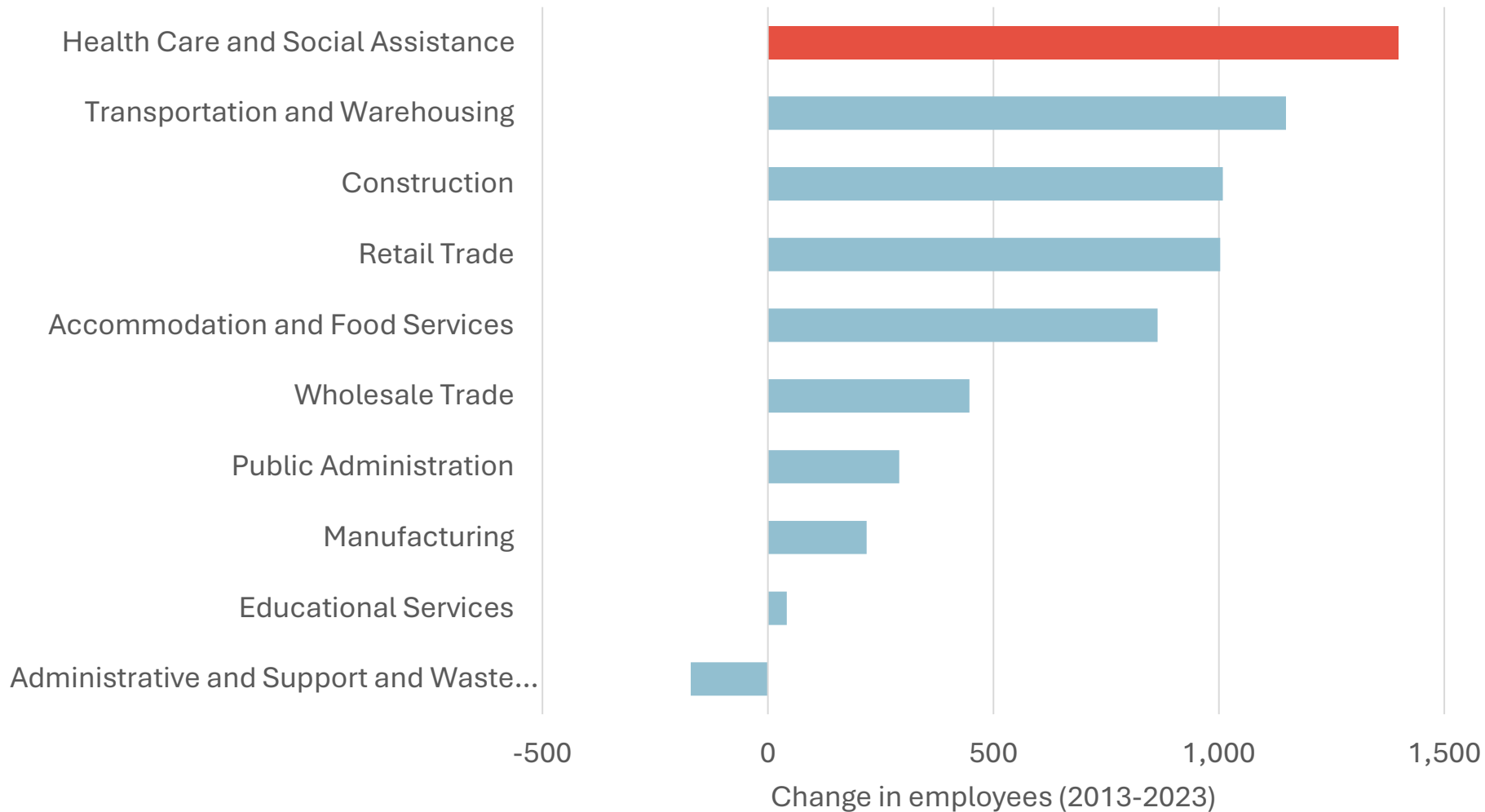




Dashboards are **not**  
data stories.

# Visualizing data to educate

**Health care** added the most jobs in Henderson County over the past ten years



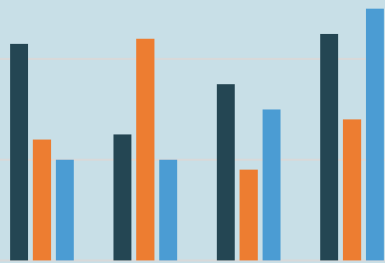
Titles includes main takeaway and clear about place and time.

Most important information is emphasized using color and ordering of information.

# Who do you visualize data for?

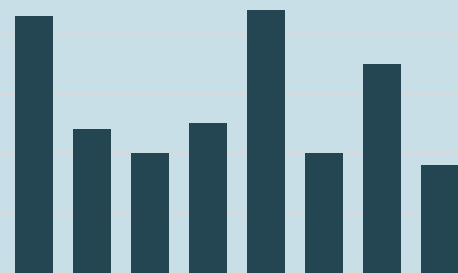
## Discover

Yourself



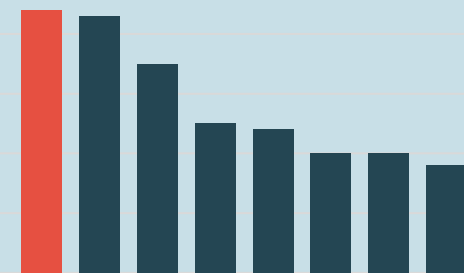
## Inform

Subject expert



## Educate

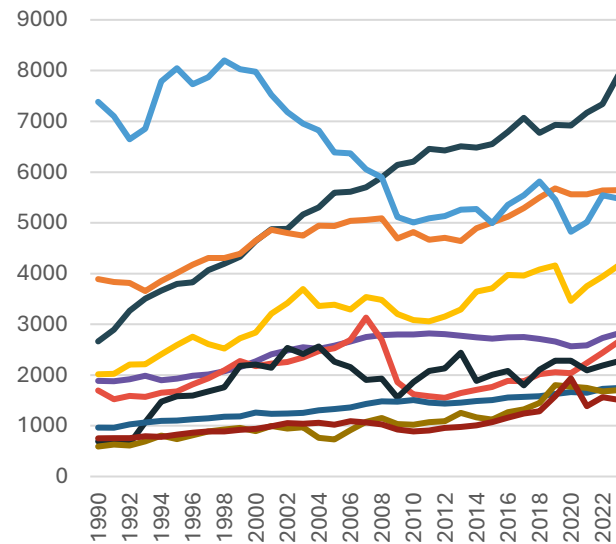
Everyone else



# Defining your audience

# Visualizing data to discover

I want to know this.



# Visualizing data to **inform**

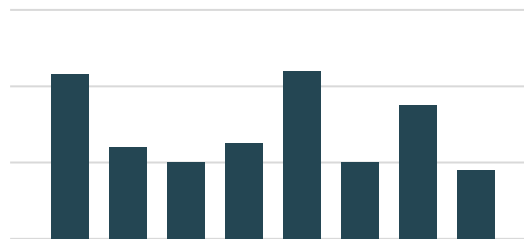
You want to know this.



I want to know this.



Performance Dashboard



# Visualizing data to educate

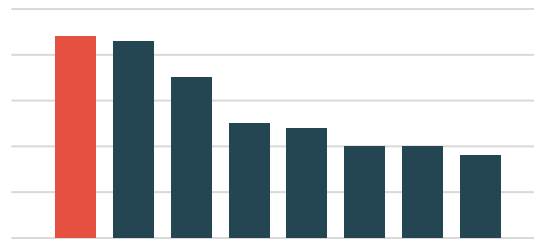
I want **you** to know this.

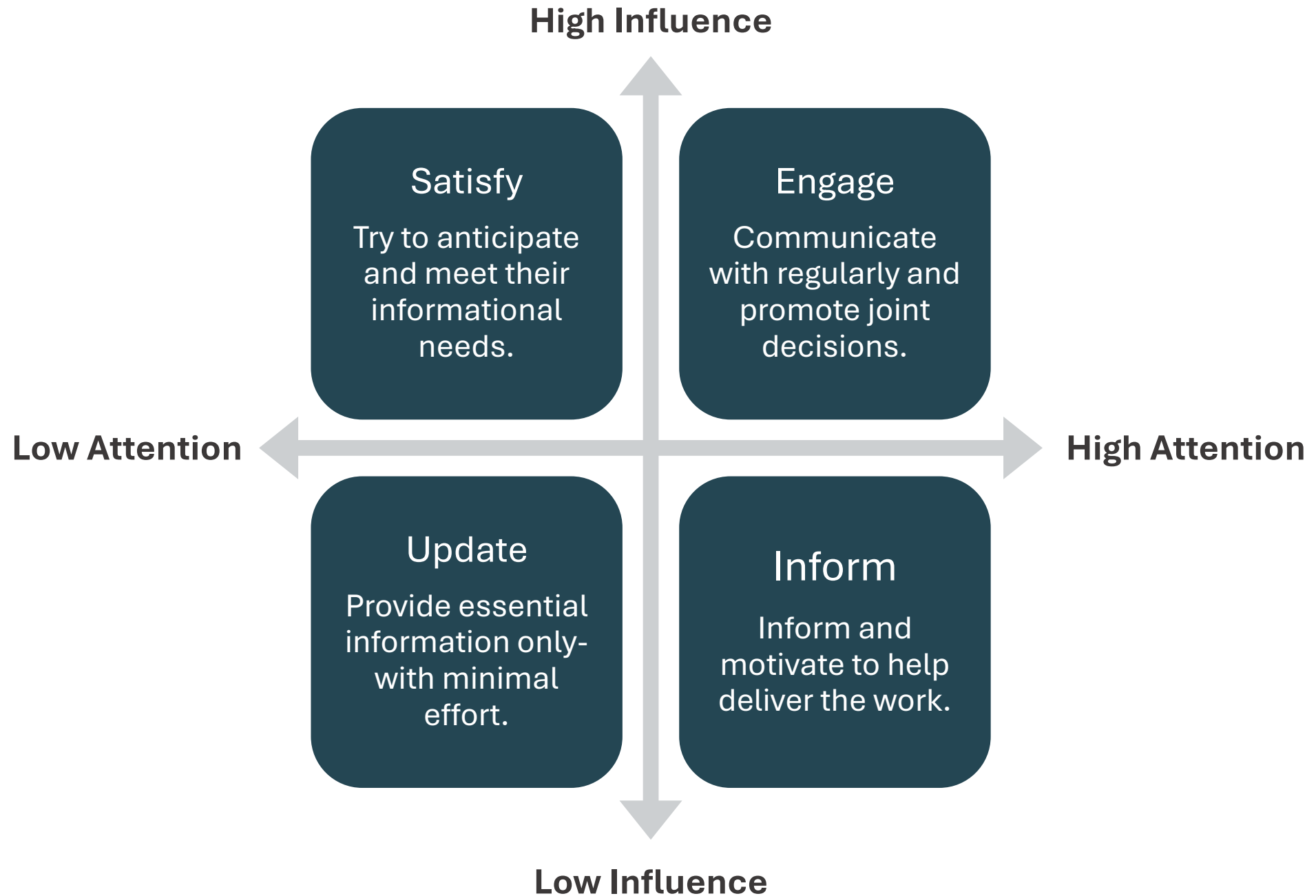


I want to know this.



Data Story







# Introductions

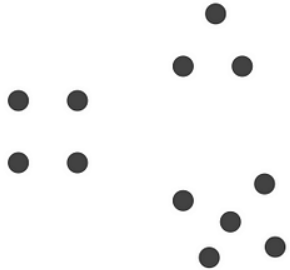
Discussion with your neighbor...

- Who have been some of data stakeholders in past jobs you've had? Thinking about your interests, who do you imagine your future data stakeholders will be?
- What do you hope to get out of this class?

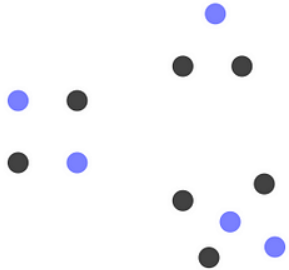
# Principles of Data Visualization

# Gestalt's Principles of Visual Processing

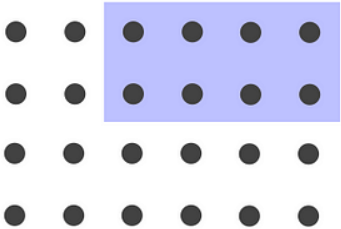
PROXIMITY



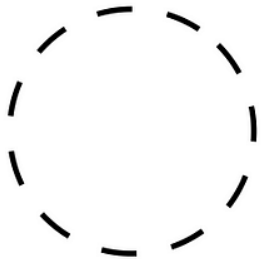
SIMILARITY



ENCLOSURE



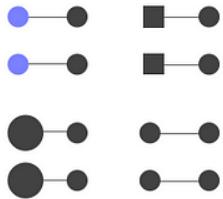
CLOSURE



CONTINUITY



CONNECTION



# Choosing the right chart

And small edits you can make to make better visuals

# Choosing the right chart for your data

Categories

Time

Relationships

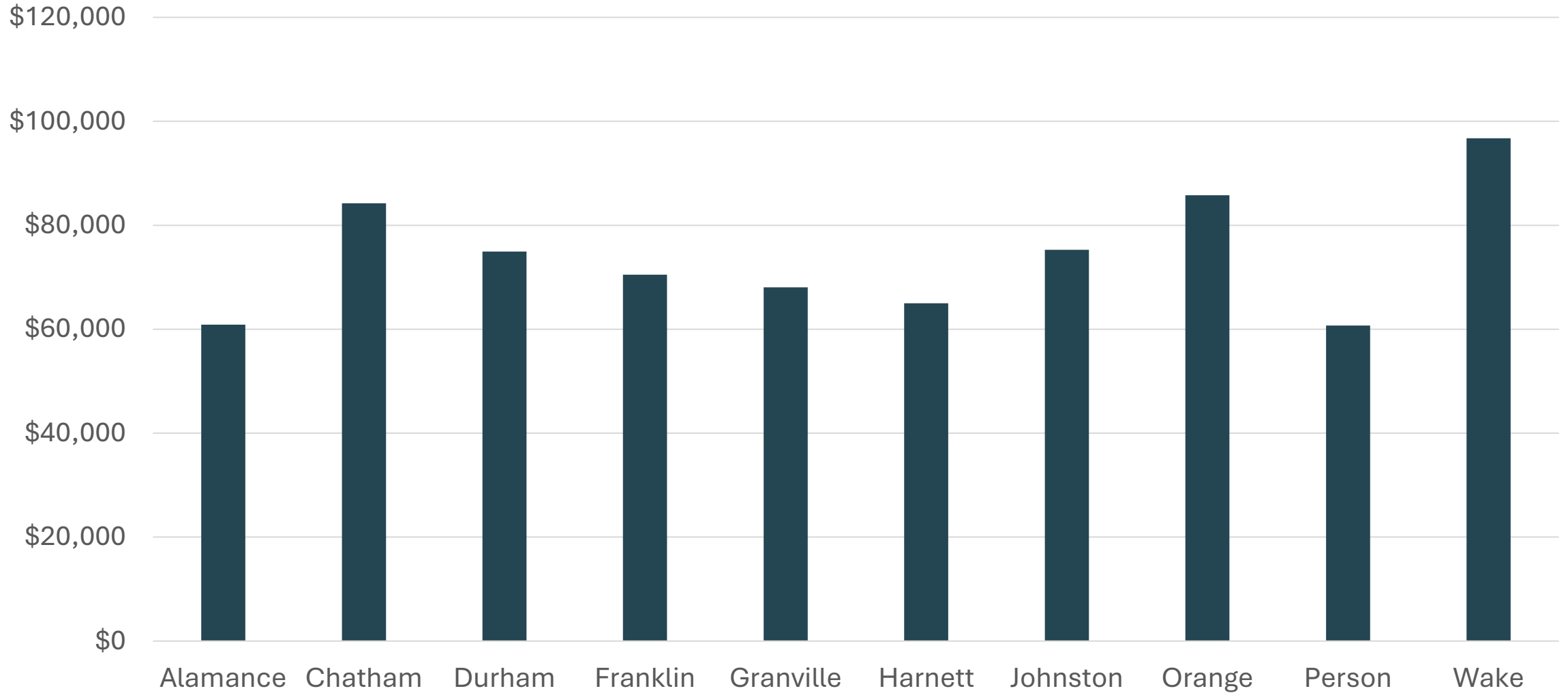
Geospatial

Qualitative

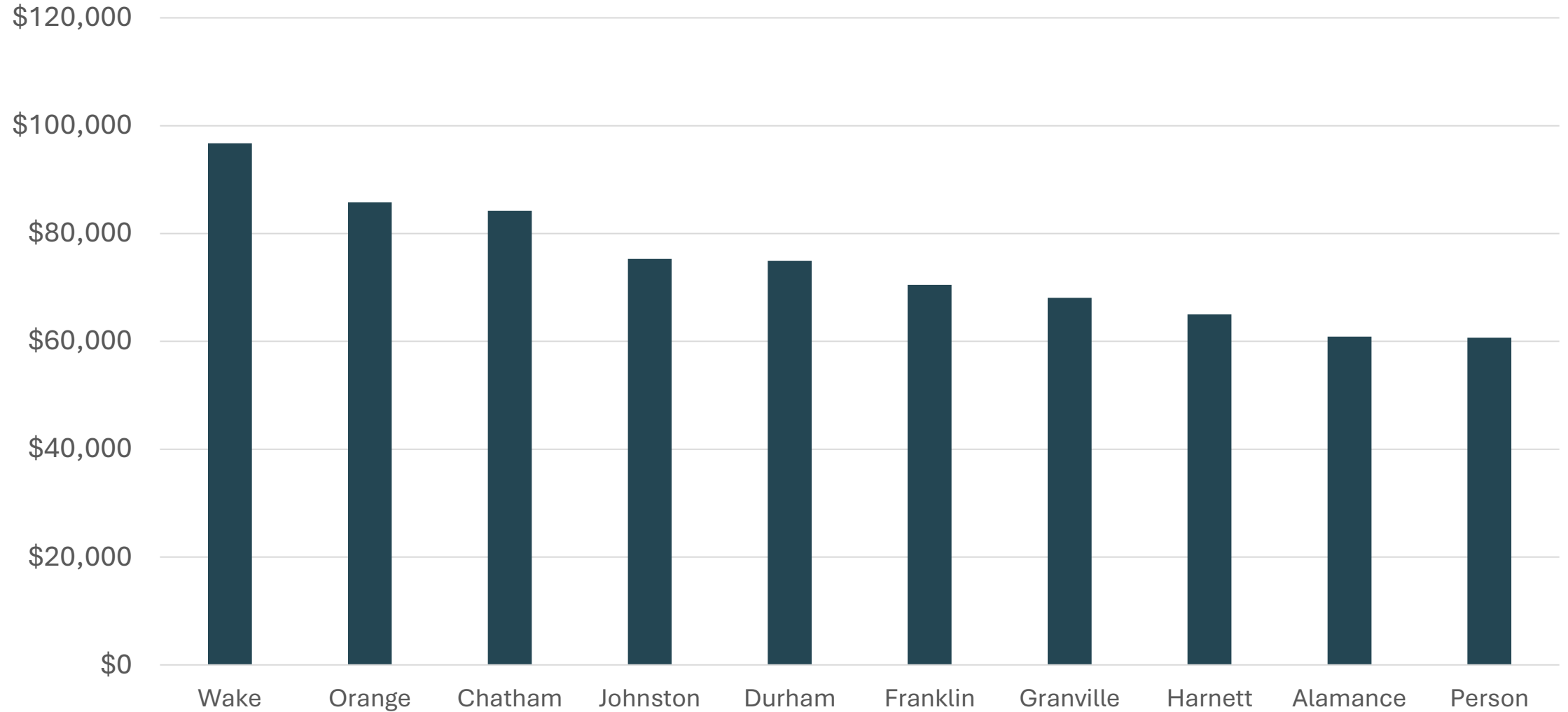
Tables

# Categories

# Median household income among North Carolina counties

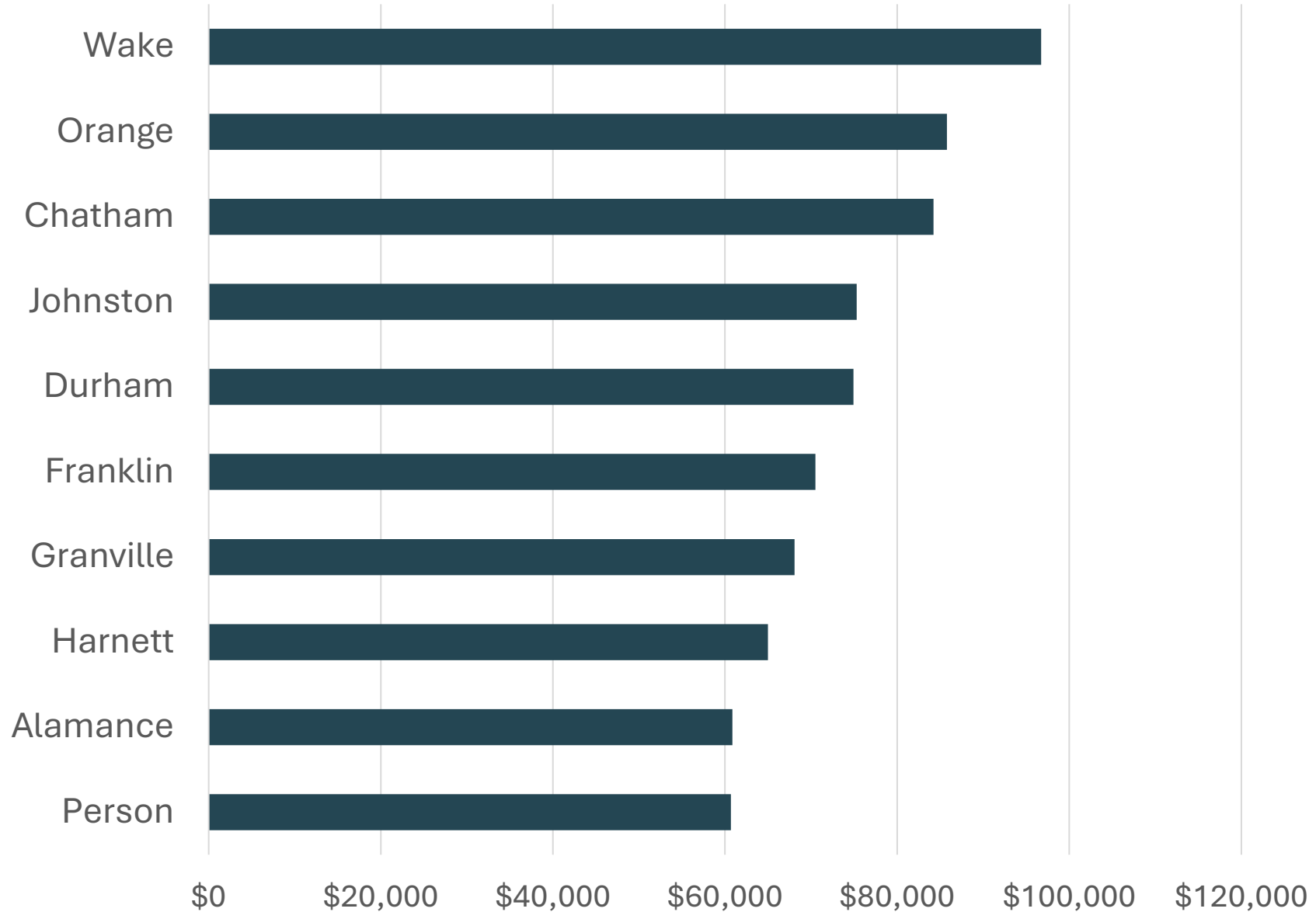


## Median household income among North Carolina counties





## Median household income among North Carolina counties

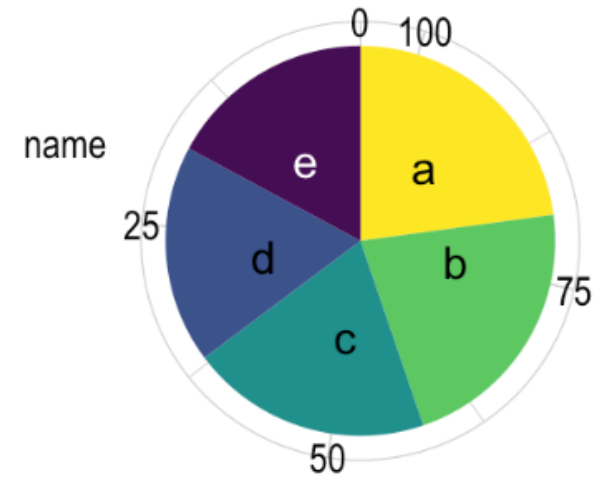
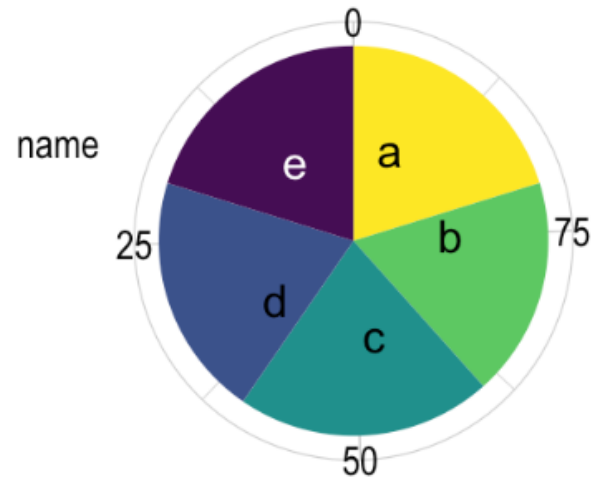
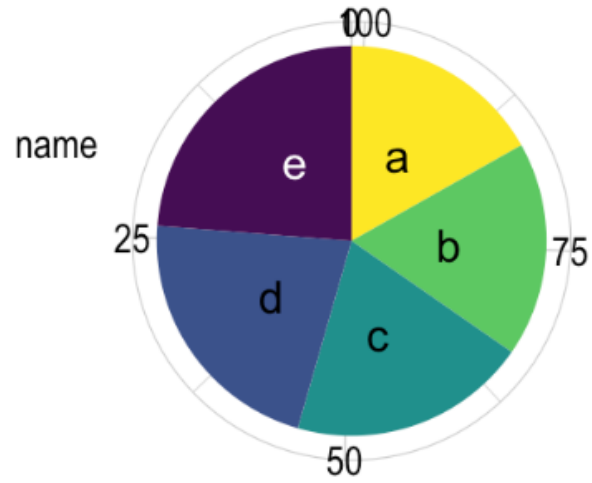


# THE ISSUE WITH PIE CHART

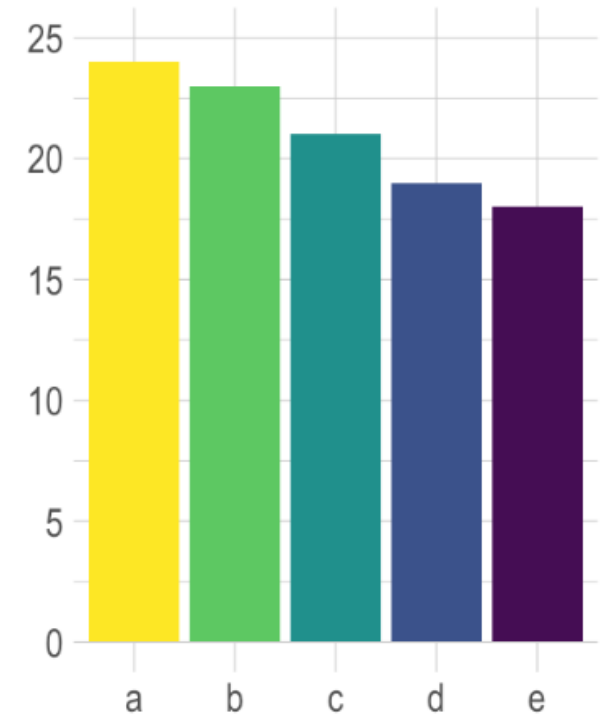
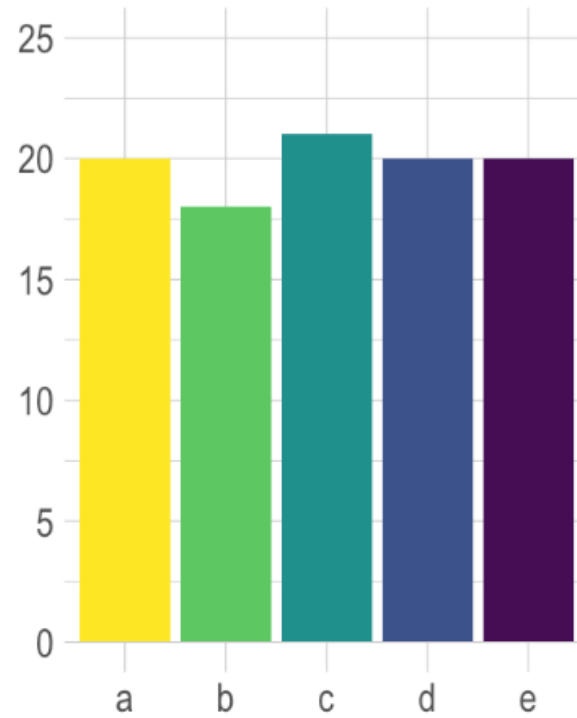
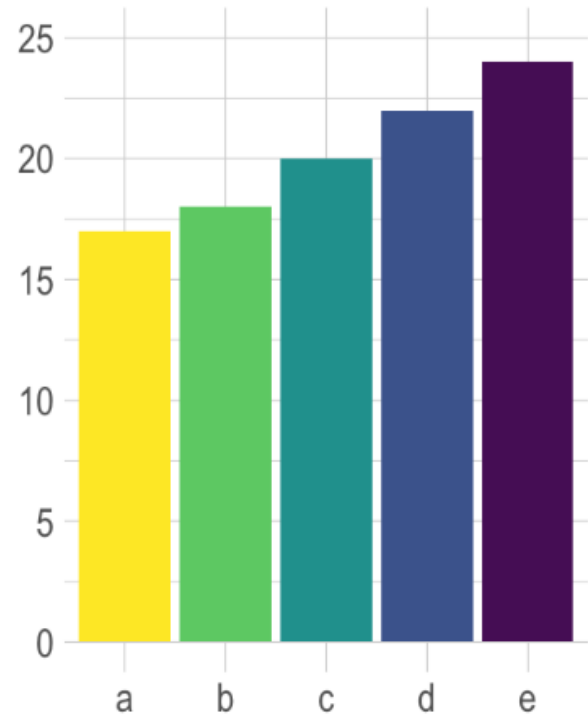
*A collection of common [dataviz caveats](#) by [Data-to-Viz.com](#)*



# Data in pie charts



# The same data in bar charts



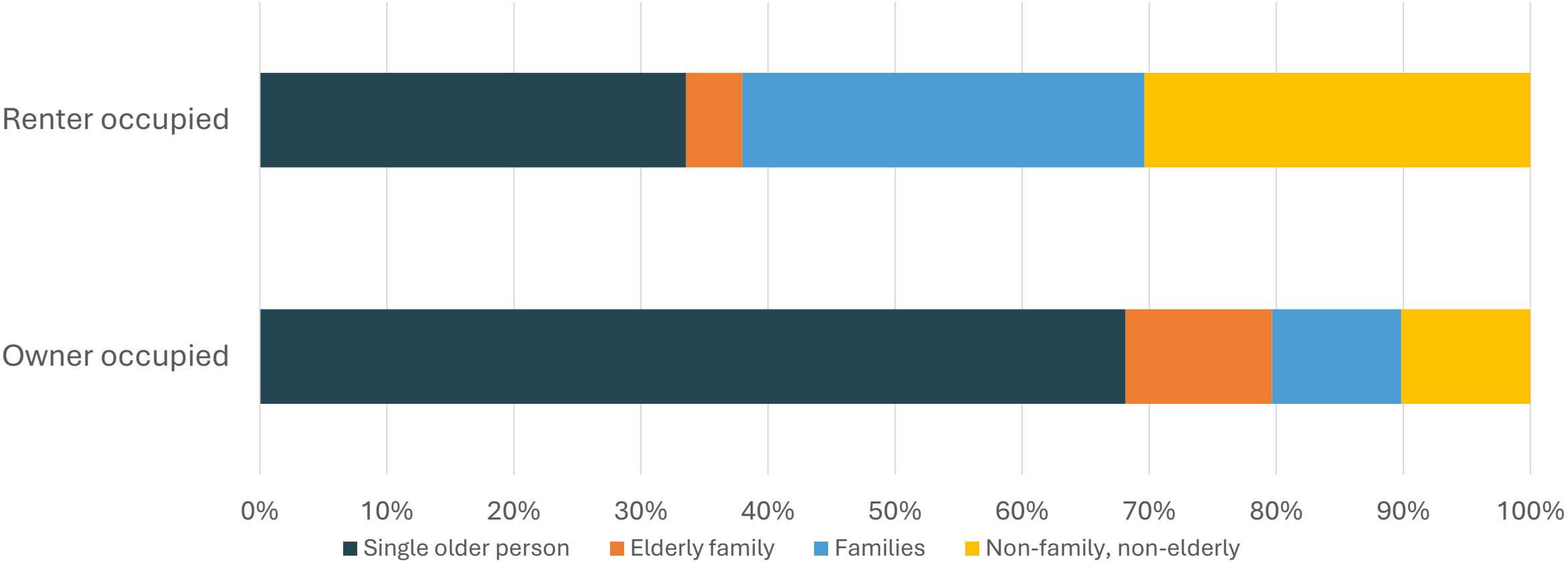
# Other ways of visualizing proportions

## Cost-burdened households in Hendersonville, NC

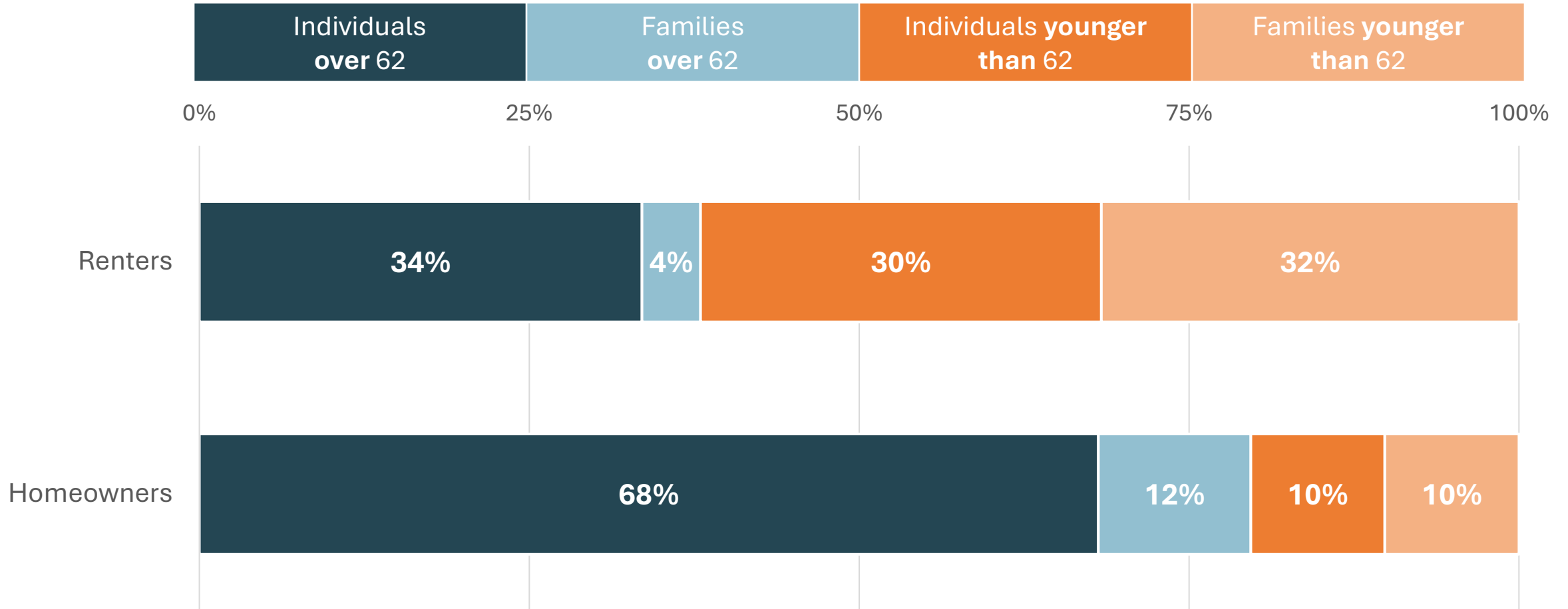
Household Type	Owner occupied	Renter occupied
Individuals over 62	68%	34%
Families over 62	12%	4%
Families under 62	10%	32%
Individuals under 62	10%	30%
<b>Total</b>	<b>100%</b>	<b>100%</b>

# Visualizing proportions

## Households in Hendersonville spending more than 30% of their income on housing-related expenses



# Households in Hendersonville spending more than 30% of their annual income on housing costs

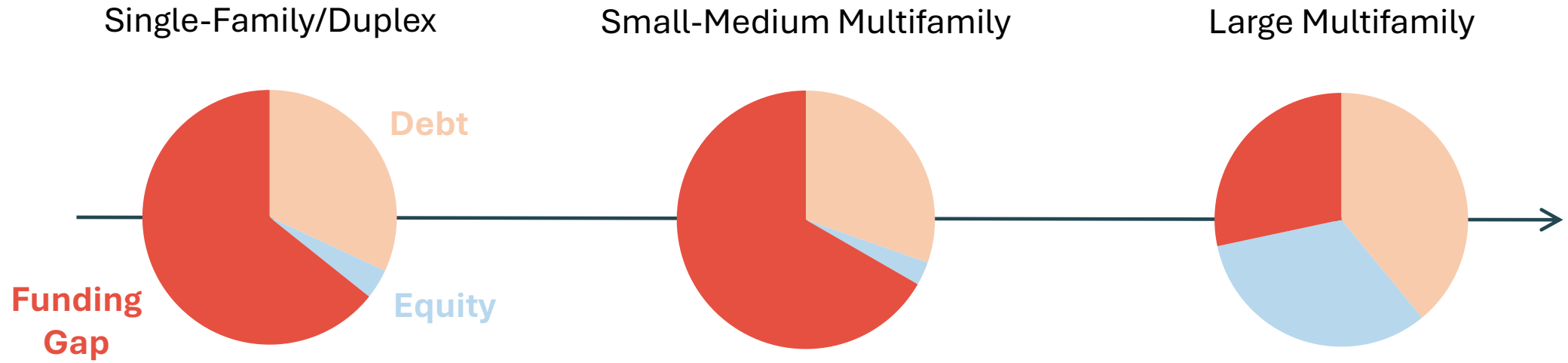


**Used shades of blue to associate households over 62.**

**And did the same thing with orange for households under 62.**

Source: HUD CHAS 2021

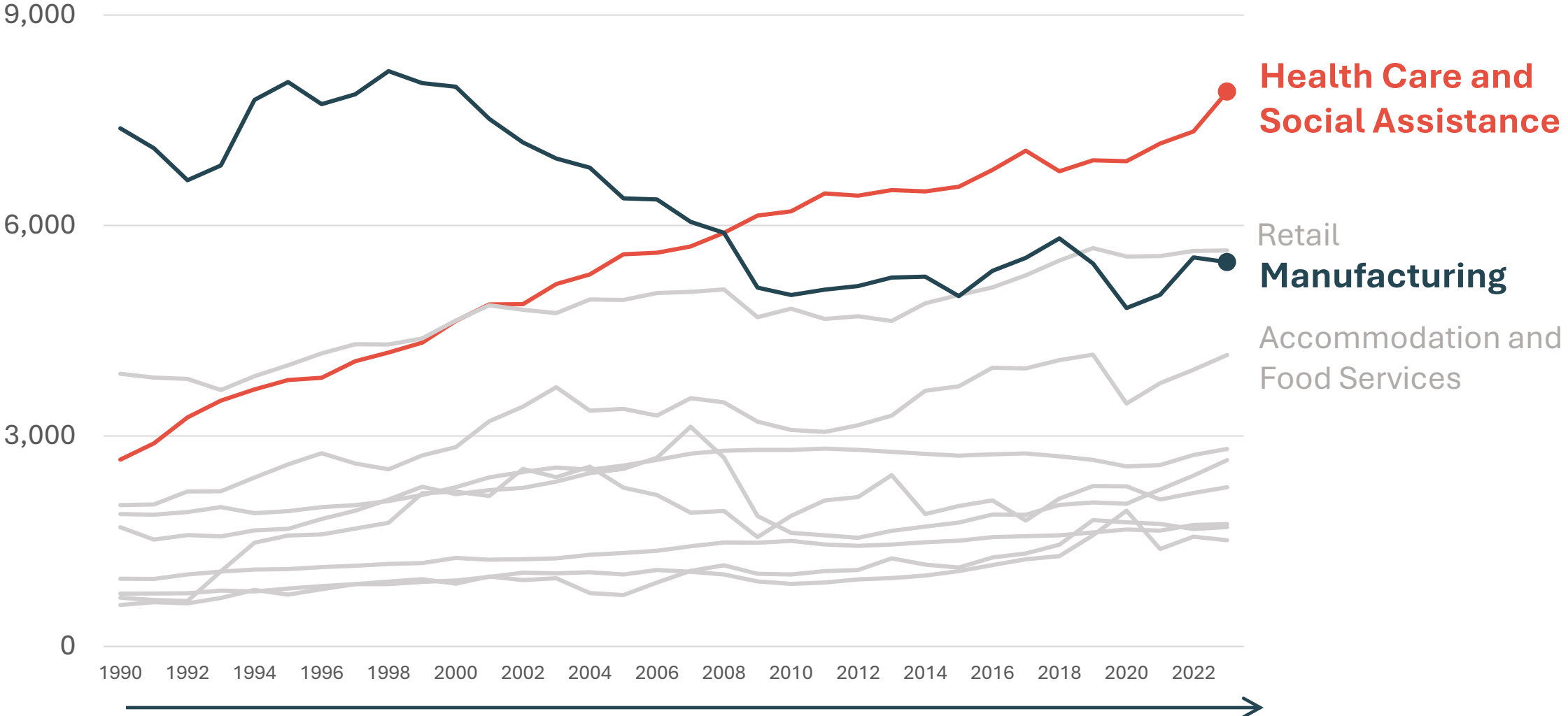
# Financial gap for housing decreases with density





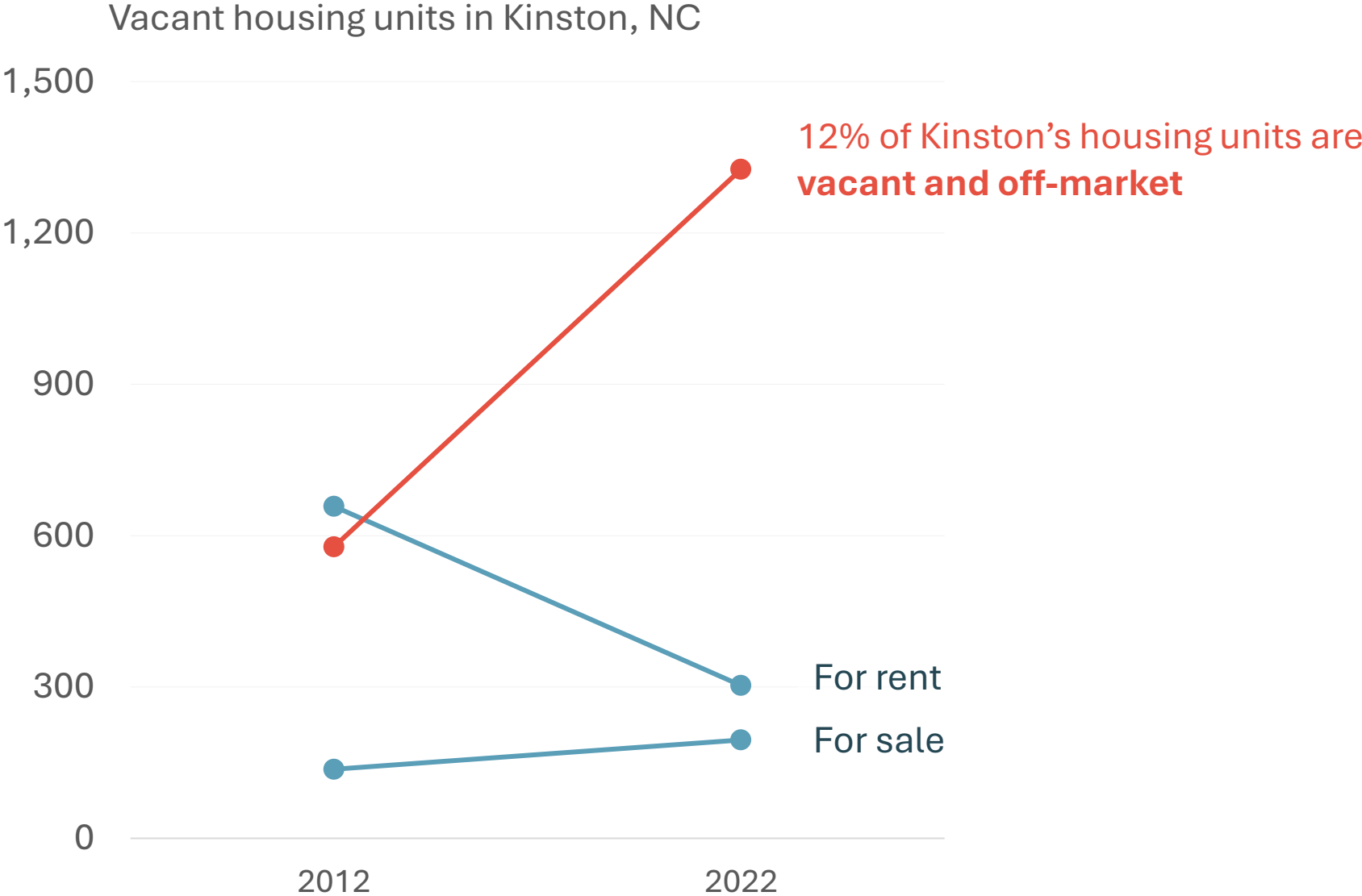
# Time

In 2008, **Health Care** surpassed **Manufacturing** to become the largest industry in Henderson County



Closed lines draw the eye from left to right.

# Slope charts

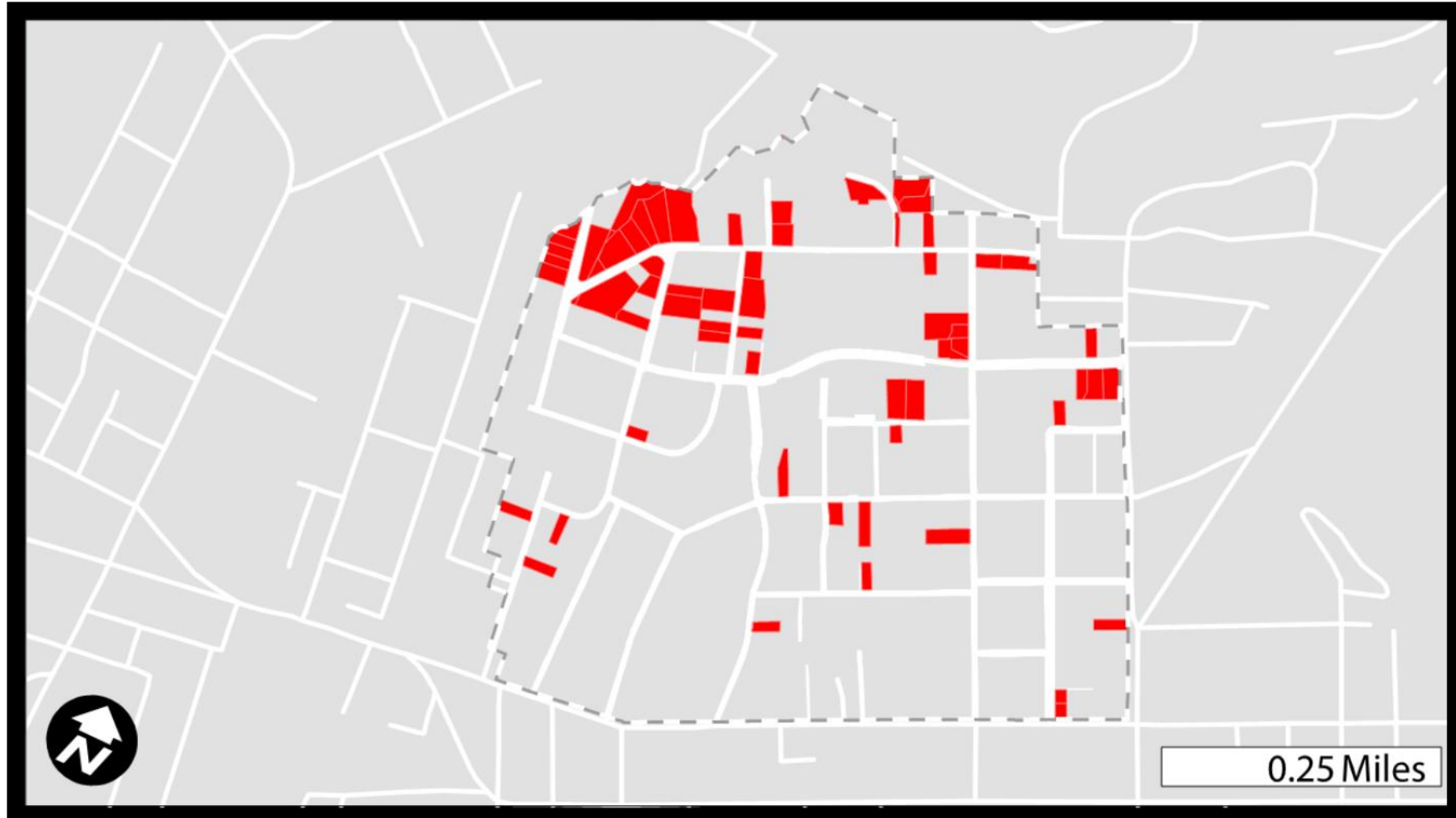


# Maps

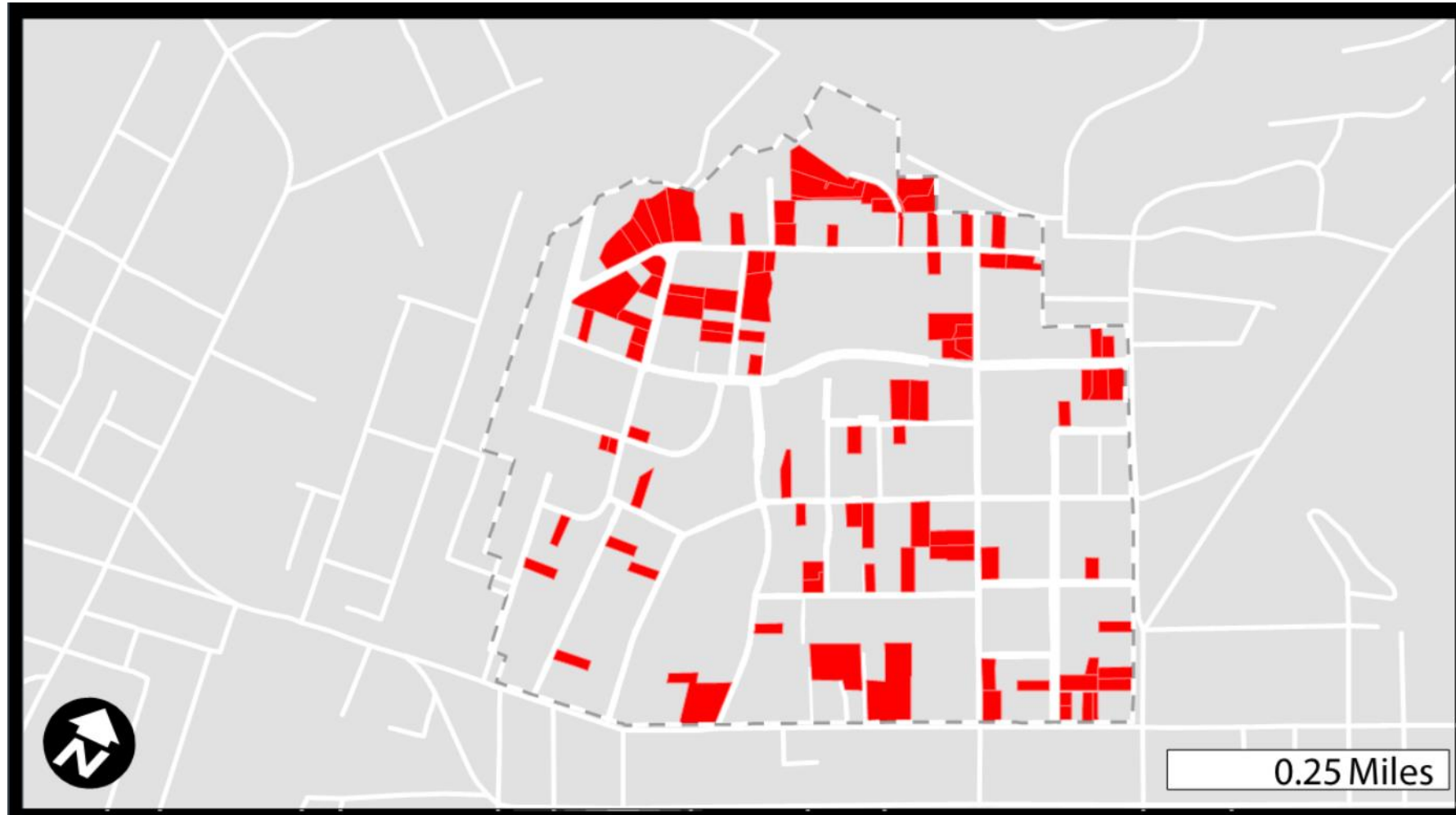
Should it be  
a map?

Maps are for  
**spatial stories.**

# INVESTOR-OWNERSHIP IN 2000

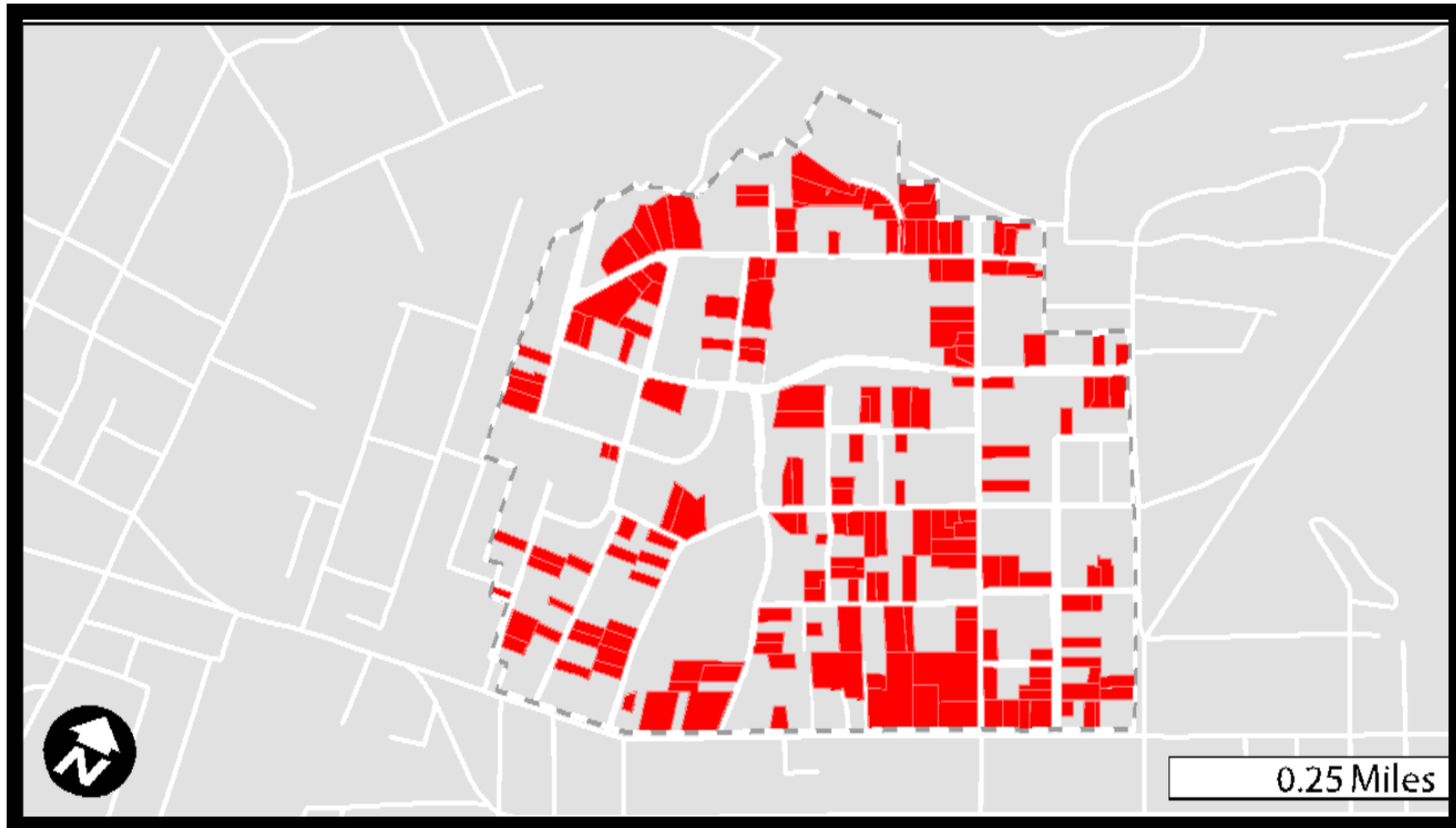


# INVESTOR-OWNERSHIP IN 2005



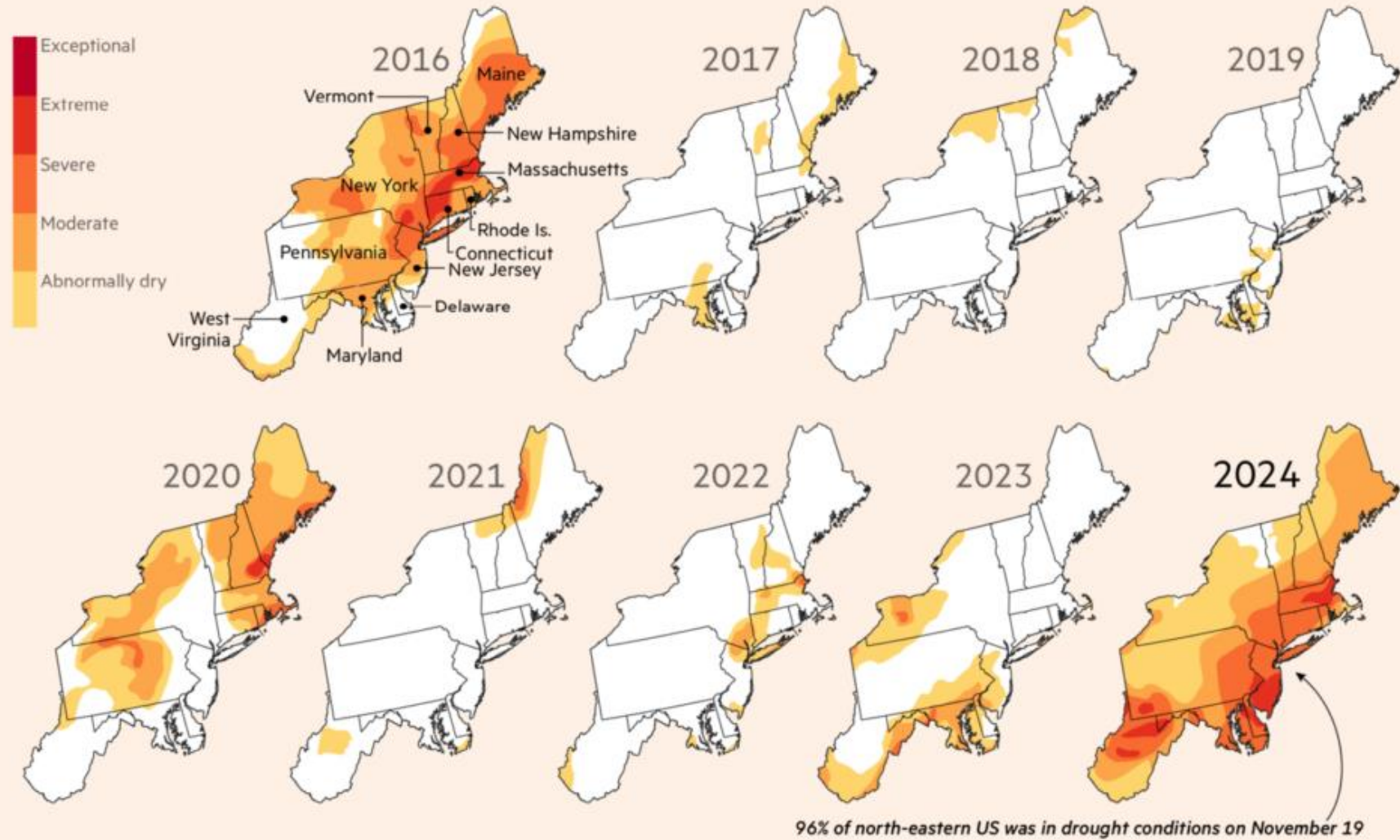


# INVESTOR-OWNERSHIP IN 2011



## North-eastern US suffers worst November drought since 2016

Drought intensity for November in north-east US



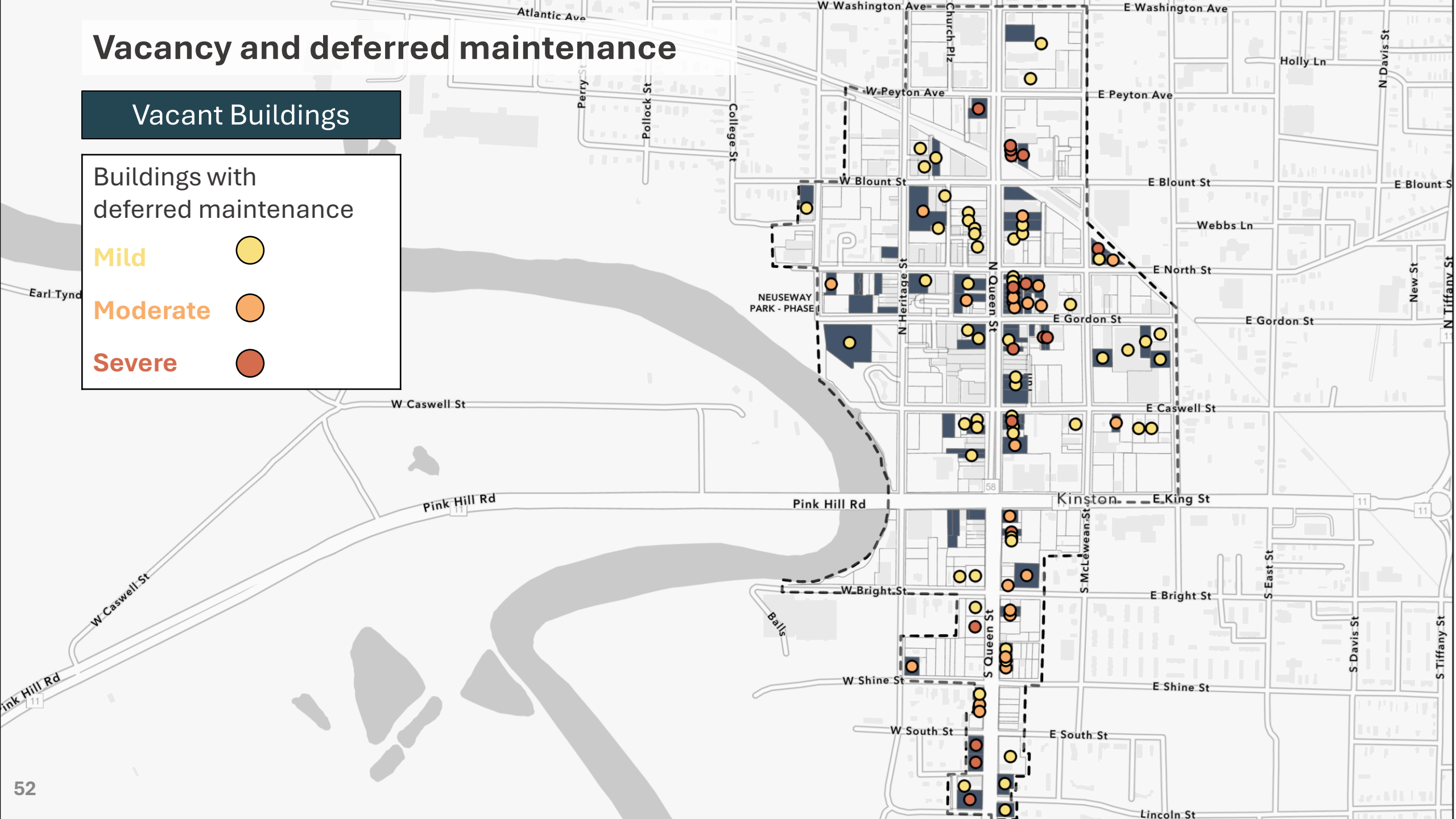
Does your basemap  
**add or distract**  
from your story?

# Vacancy and deferred maintenance

## Vacant Buildings

Buildings with deferred maintenance

- Mild 
- Moderate 
- Severe 



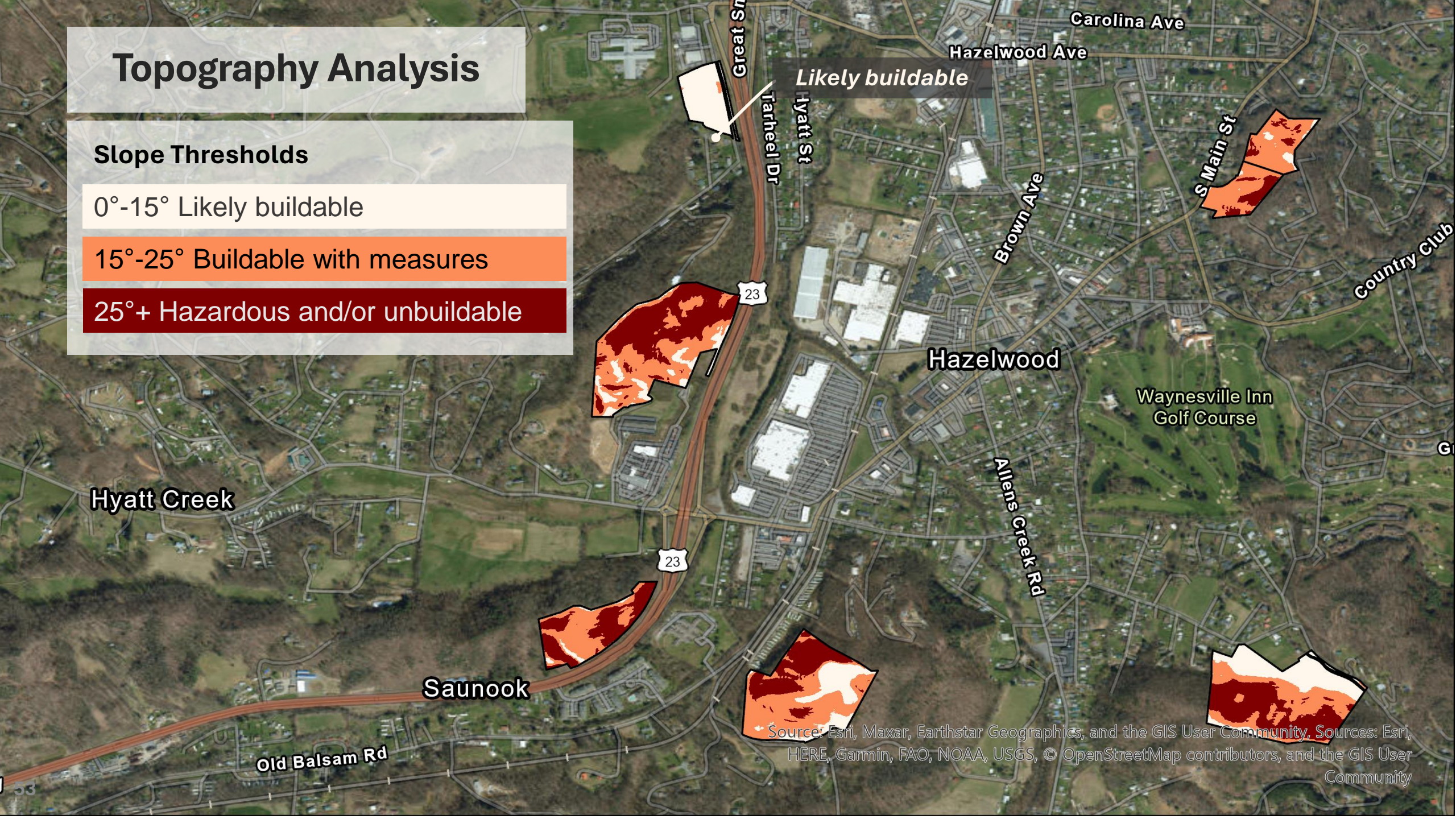
# Topography Analysis

## Slope Thresholds

0°-15° Likely buildable

15°-25° Buildable with measures

25°+ Hazardous and/or unbuildable



Source: Esri, Maxar, Earthstar Geographics, and the GIS User Community, Sources: Esri, HERE, Garmin, FAO, NOAA, USGS, © OpenStreetMap contributors, and the GIS User Community

## Clear Creekside Drive

*Estimated 30 units impacted by Helene*

*100- and 500-Year  
Flood Zones*

*Estimated flooding from  
Hurricane Helene*

Homes in this neighborhood sell for \$300-\$350K and are affordable to households making between 60% and 80% of AMI.

# Legend

## Top Owners

- Gandalf
- Gimli
- Aragorn
- Shire Inc
- Frodo
- Samwise
- Legolas





**Aragorn**

**Legolas**

**Gandalf**

**Frodo**

**Samwise**

**Shire Inc**

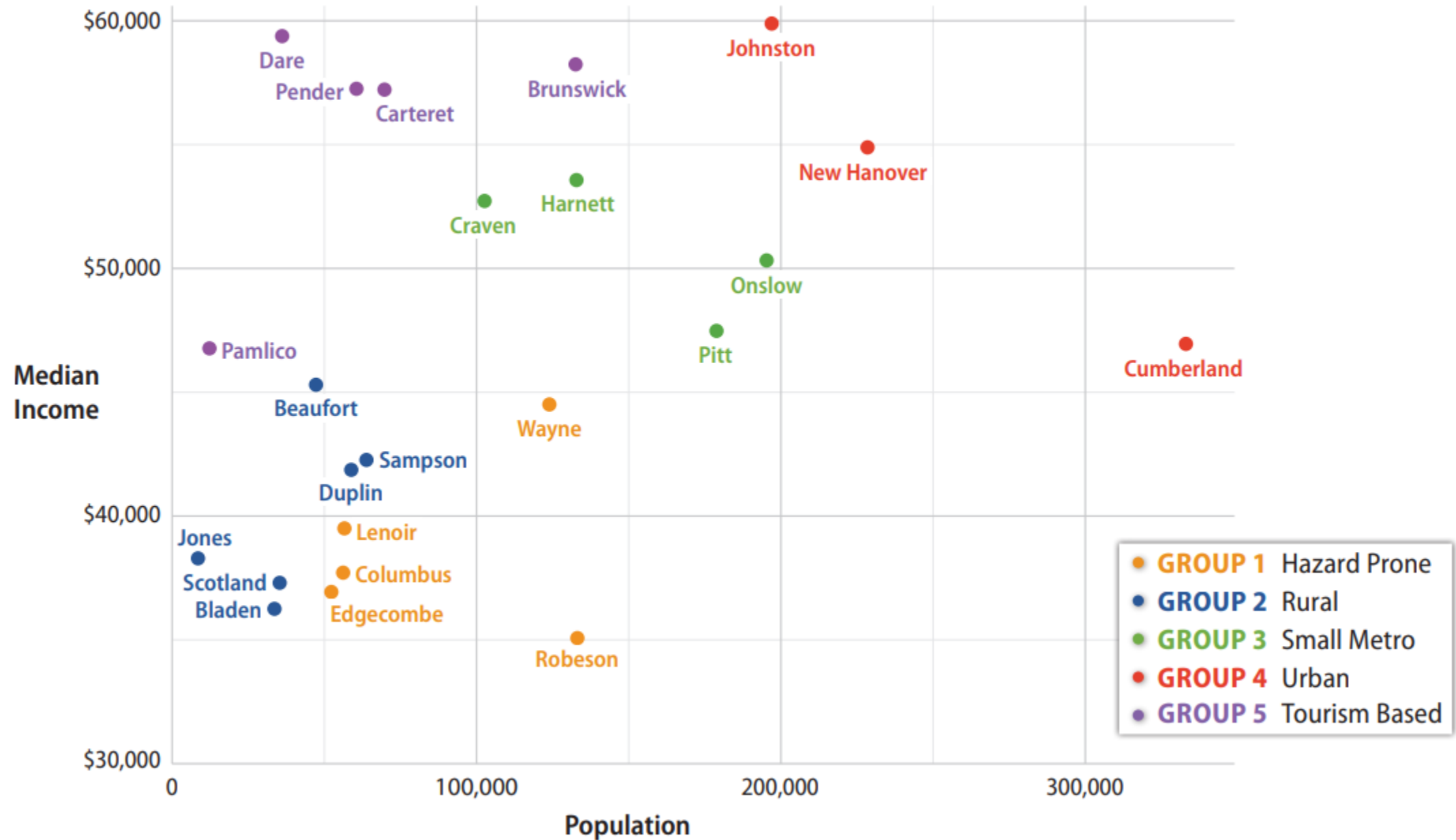
**Gimli**

**Gandalf**



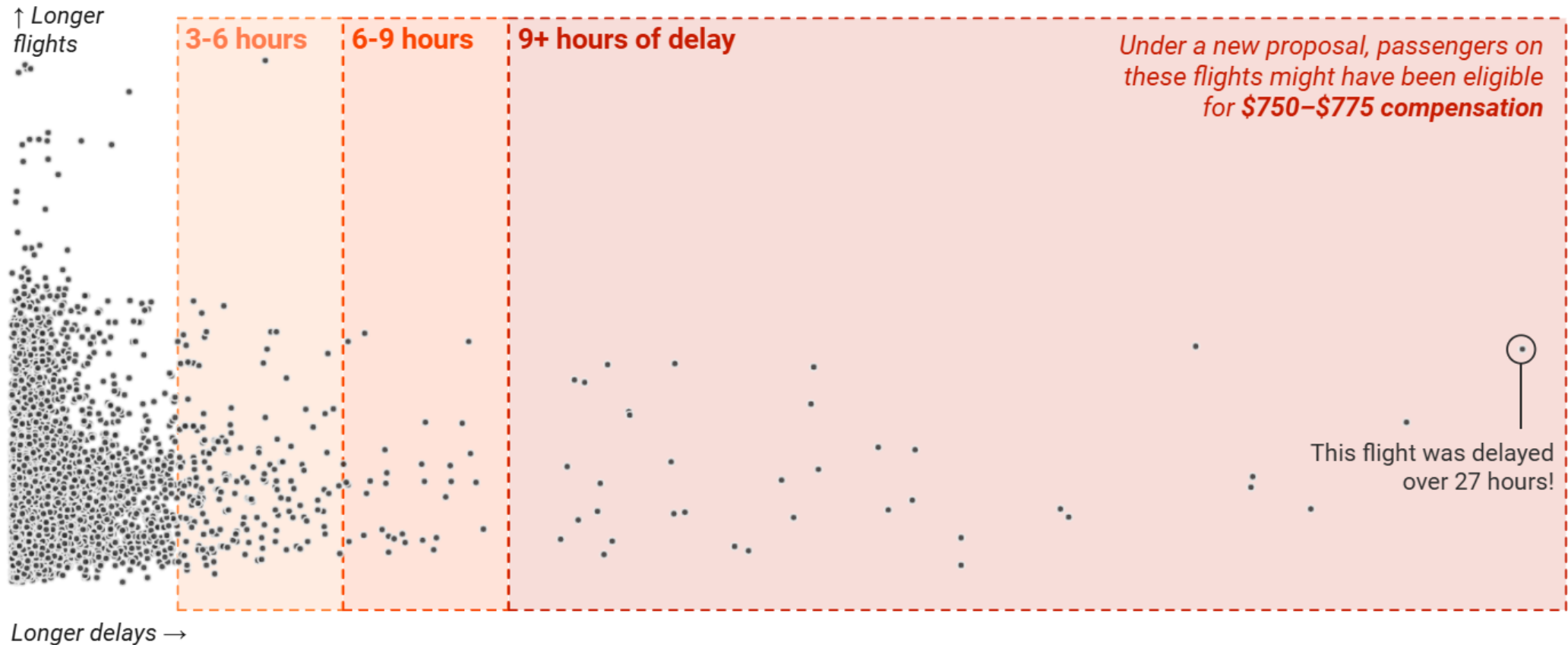
# Relationships

Figure B.1. Population and Median Income by Group



# How long are flight delays during the winter holiday season?

Each dot represents one delayed U.S. domestic flight. **3.4%** were delayed for 3-6 hours, **0.6%** for 6-9 hours and **0.5%** for more than 9 hours. Does not include delays due to weather.



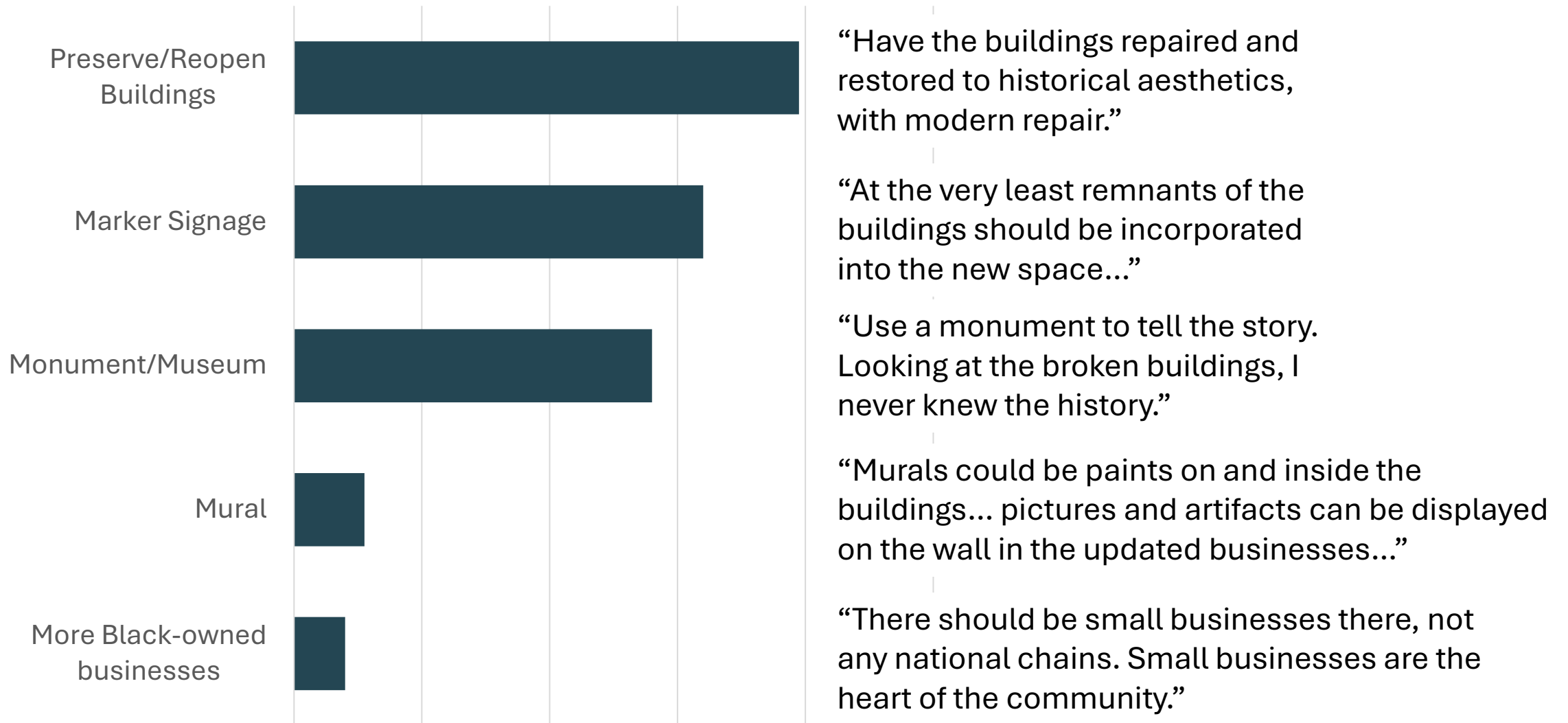
All 5,908 U.S. domestic flights that were delayed due to the carrier between December 19, 2022 and January 1, 2023.

Chart: Michael Do Thoi • Source: [US DOT: Bureau of Transportation Statistics via Kaggle](#)

# Qualitative Data



Do you have suggestions on how the history of the Railroad Corner could be preserved or incorporated into future redevelopment?



# Tables

# When to use each

He further suggests when to use which format:

## Use Tables When

- The display will be used to look up individual values
- It will be used to compare individual values
- Precise values are required
- Quantitative values include more than one unit of measure
- Both detail and summary values are included

## Use Graphs When

- The display will be used to reveal relationships among whole sets of values
- The message is contained in the shape of the values (e.g., patterns, trends, exceptions)

Adapted from:

Few, Stephen. (2012). *Show Me the Numbers: Designing Tables and Graphs to Enlighten*.(4)57



# The **average** table: not ideal

Sources	Scenario 1	Scenario 2	Scenario 3
Equity	\$10.9M	\$12.1M	\$14.3M
Debt	\$10.7M	\$10.7M	\$10.7M
Gap	\$11.5M	\$12.7M	\$20M
Total	\$33.1M	\$35.4M	\$45M

# Better


Numbers simplified by putting  
“millions” in the column name.

<b>Sources</b> <i>(millions of \$)</i>	<b>Scenario 1</b>	<b>Scenario 2</b>	<b>Scenario 3</b>
Equity	\$10.9	\$12.1	\$14.3
Debt	\$10.7	\$10.7	\$10.7
Gap	\$11.5	\$12.7	\$20.0
<b>Total</b>	<b>\$33.1</b>	<b>\$35.4</b>	<b>\$45.0</b>

Totals & headers  
separated from  
main table

Text is left justified – Numbers are right justified

# Better still?

		Sources <i>(millions of \$)</i>			
		Equity	Debt	Gap	Total
	Scenario 1	\$10.9	\$10.7	\$11.5	\$33.1
	Scenario 2	\$12.1	\$10.7	\$12.7	\$35.4
	Scenario 3	\$14.3	\$10.7	\$20.0	\$45.0

Readers generally have an easier time comparing data vertically than horizontally

# Use gray to differentiate rows in long tables

<b>Land Use</b>	<b>Parcels Count</b>	<b>Land Tax Value Dollars</b>	<b>Building Tax Value Dollars</b>	<b>Building Sq Footage Square Feet</b>
Residential	322	2,857,713	10,714,447	405,641
Undeveloped	218	2,191,274	125,987	-
Open Space	20	293,400	7,277	-
Parking	20	361,353	81,065	-
Vacant commercial	17	248,019	603,073	40,760
Religious	16	579,344	4,222,738	97,169
Industrial	11	336,163	359,784	48,987
Retail	9	230,096	394,625	43,229
Office	6	139,636	568,079	22,839
Institutional	3	107,628	474,416	10,424
<b>Total</b>	<b>642</b>	<b>\$7,344,626</b>	<b>\$17,551,491</b>	<b>669,049</b>

# But not too much gray

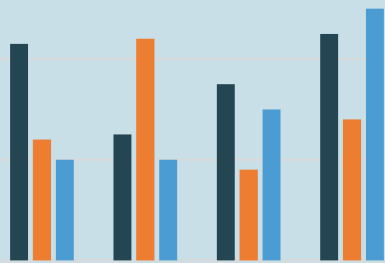
Land Use	Parcels Number	Land Tax Value Dollars	Building Tax Value Dollars	Building Sq Footage Square Feet
Residential	322	2,857,713	10,714,447	405,641
Undeveloped	218	2,191,274	125,987	-
Open Space	20	293,400	7,277	-
Parking	20	361,353	81,065	-
Vacant commercial	17	248,019	603,073	40,760
Religious	16	579,344	4,222,738	97,169
Industrial	11	336,163	359,784	48,987
Retail	9	230,096	394,625	43,229
Office	6	139,636	568,079	22,839
Institutional	3	107,628	474,416	10,424
<b>Total</b>	<b>642</b>	<b>\$7,344,626</b>	<b>\$17,551,491</b>	<b>669,049</b>

# Recap: Why visualize data?

## Discover

I want to **discover** insights about my data.

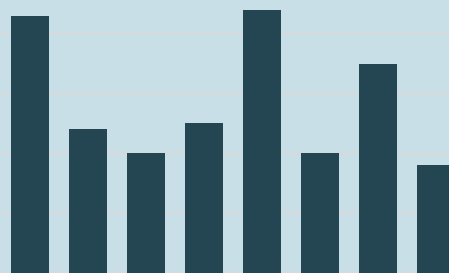
e.g. charts made during analysis



## Inform

I want to **inform** others about my data.

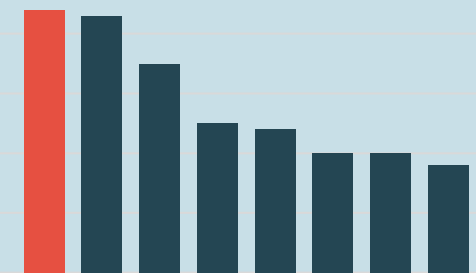
e.g. dashboards, regular reports



## Educate

I want to **educate** others about my data.

e.g. data storytelling



# For next time: Alli Torban's XRAY Practice

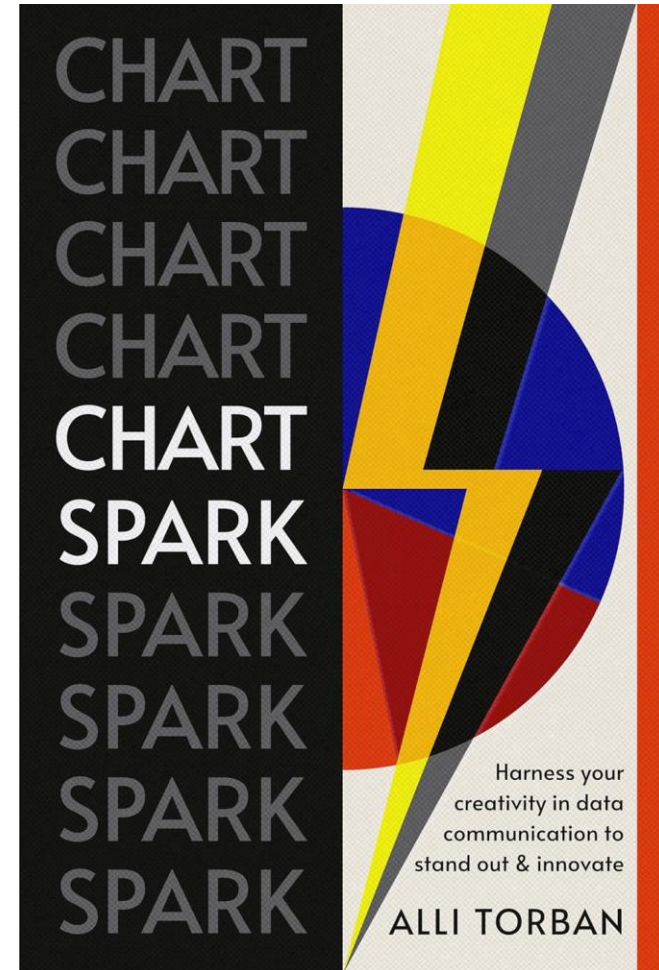
XRAY Exercise

**X** for Excites: What excites you?

**R** for Rules: What rules does it follow?

**A** for Anarchy: How is it breaking the rules?

**Y** for You: How can you use it in the future?



# Measles

