What belongs together?

Using elemental groupings and Gestalt principles to help students design more insightful data visualizations

Silas Bergen

sbergen at winona dot edu

2023 US Conference on Teaching Statistics

June 2, 2023

https://bit.ly/3OM7ff6

Breakout materials

WINONA STATE UNIVERSITY

DEPARTMENT OF MATHEMATICS AND STATISTICS

College of Science and Engineering

Course context

- Week 2 of DSCI 310: data visualization course
- Just covered: grammar of graphics (how to specify a data → graph mapping)
- Not covered: any graphing software

The data: Quarterly hospital lengths of stay following an initiative to shorten lengths of stay

	Duration of stay (in hours)					
Quarter	<=24	24 -36	36 - 48	48 - 59	>=60	Unknown
Q1	12.20%	53.90%	6.30%	18.90%	6.90%	1.80%
Q2	14.30%	58.30%	4.10%	16.70%	5.90%	0.70%
Q3	19.50%	52.20%	2.80%	17.40%	7.00%	1.10%
Q4	25.40%	50.30%	2.70%	14.40%	5.60%	1.70%

The question:

Has the initiative worked?

How can we best graph these data to answer this question?

Source: https://community.storytellingwithdata.com/exercises/how-can-we-improve-this-graph
Knaflic, Cole. storytellingwithdata.com.

The data: mean income by field of occupation among MNs with master's (ACS 2016)

Sex	STEM	
	No	Yes
Female	$77.0K	$89.3K
Male	$104.1K	$113.6K

The question:

How does the income gender gap differ by field of occupation (STEM vs non-STEM)?

How can we best graph these data to answer this question?

*You must always begin with a clear sense of **what belongs together**, that is, what your readers should perceive as belonging to the same group because those units of information have something in common….*

…It's your job to make this grouping obvious to your readers. It shouldn't be up to your readers to do the work of arranging the content into meaningful groups when you can do this in advance for them.

-Show me the numbers pg. 145.

What belongs together?

Define ***elemental groupings***: an individual or small collection of data values where interest lies in comparing this individual or collection across levels of one or more grouping variable

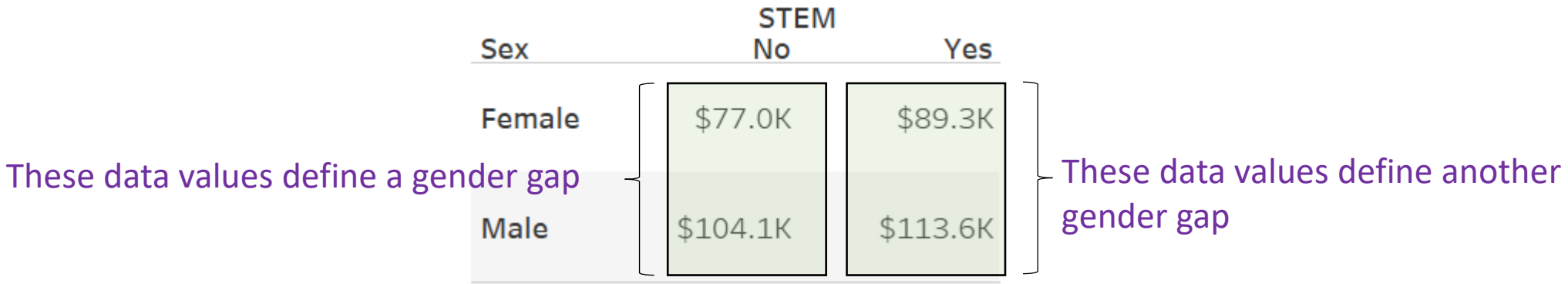
How to identify elemental groupings?

- Define your message (think: “headline” or “question”)

- Find the data values that represent the subject of the headline
  - This set of data values = one elemental grouping

- *How does the gender gap differ by field (STEM vs non-STEM)?*

- Sentence subject: *gender gap*



- ($77.0K, 104.1K) constitute one elemental grouping
- ($89.3K, $113.6K) constitute a second elemental grouping
- These groupings should be *preserved* and *emphasized* in any visualization of these data intended to address the question above

What if the question subject is unclear? Rephrase the question so it's not (from a data perspective)

Has the initiative worked?

Could become….

*Have the **short stays** increased?*

	Duration of stay (in hours)						
Quarter	<=24	24 -36	36 - 48	48 - 59	>=60	Unknown
Q1	12.20%	53.90%	6.30%	18.90%	6.90%	1.80%
Q2	14.30%	58.30%	4.10%	16.70%	5.90%	0.70%
Q3	19.50%	52.20%	2.80%	17.40%	7.00%	1.10%
Q4	25.40%	50.30%	2.70%	14.40%	5.60%	1.70%

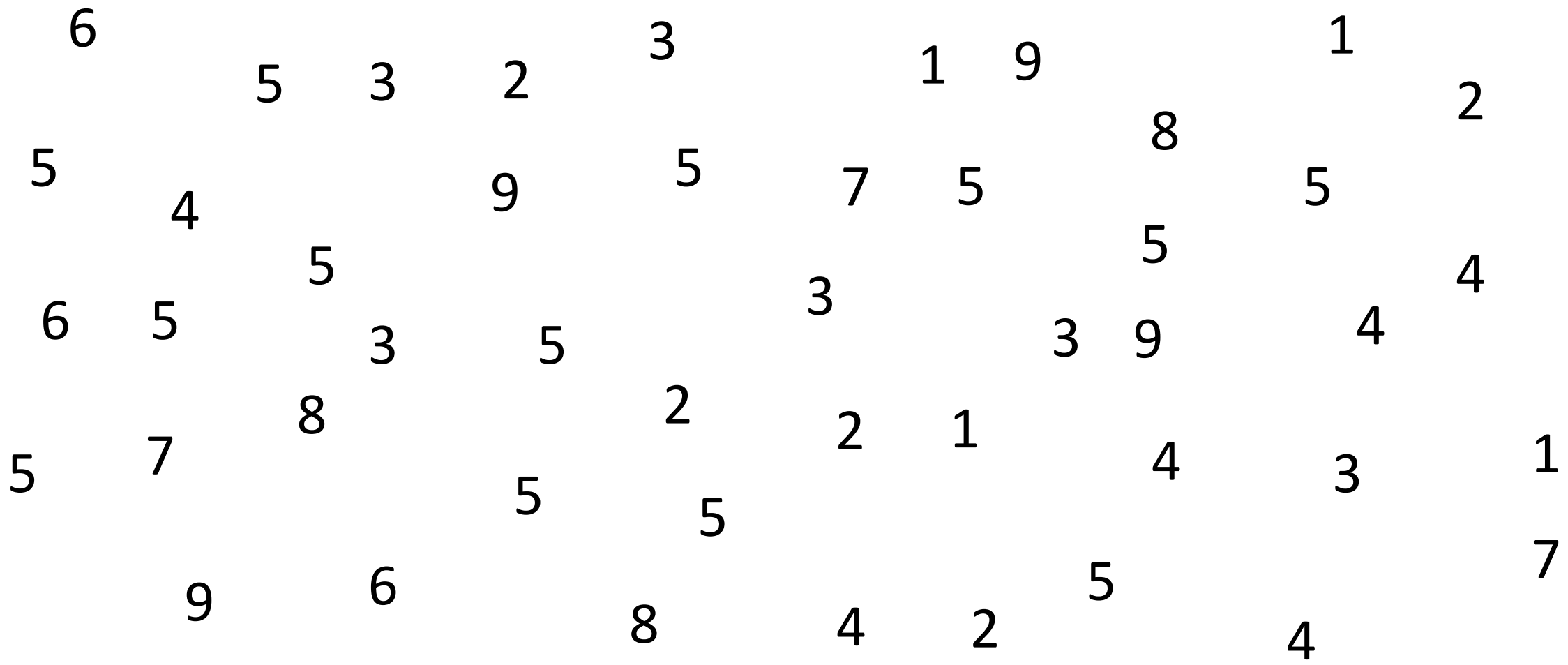
One elemental grouping

A second elemental grouping

How do we show “what belongs together”?

Gestalt principles of grouping

How many “5”s are there?



How many “5”s are there?

Inspired by Fig 6.2, *The Functional Art* by Alberto Cairo

gestalt = form or pattern

Gestalt philosophy: the whole is greater than the sum of the parts

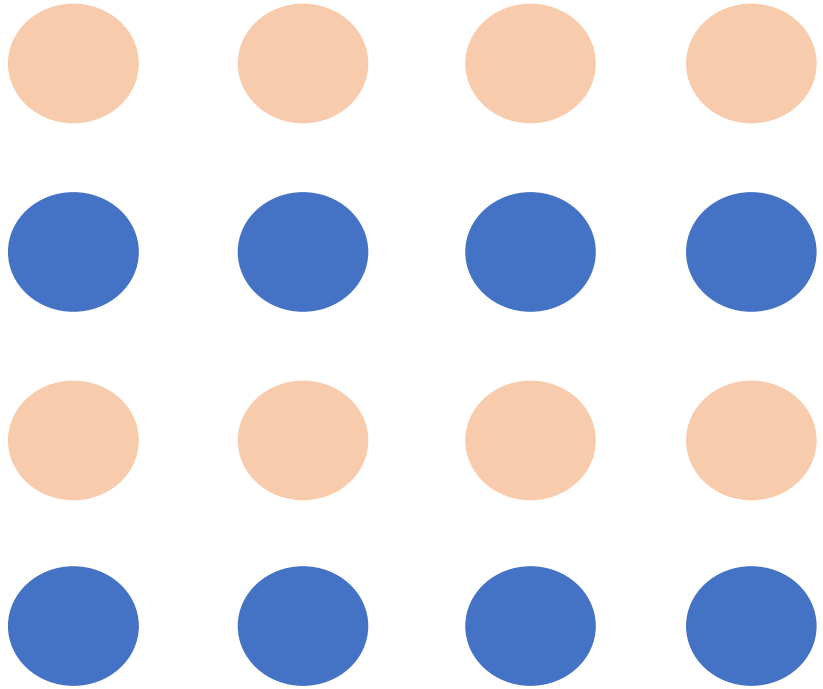
Gestalt principles: predictable ways by which we organize sensory information

Developed by Max Wertheimer & followers; early 20th century

Figure and ground

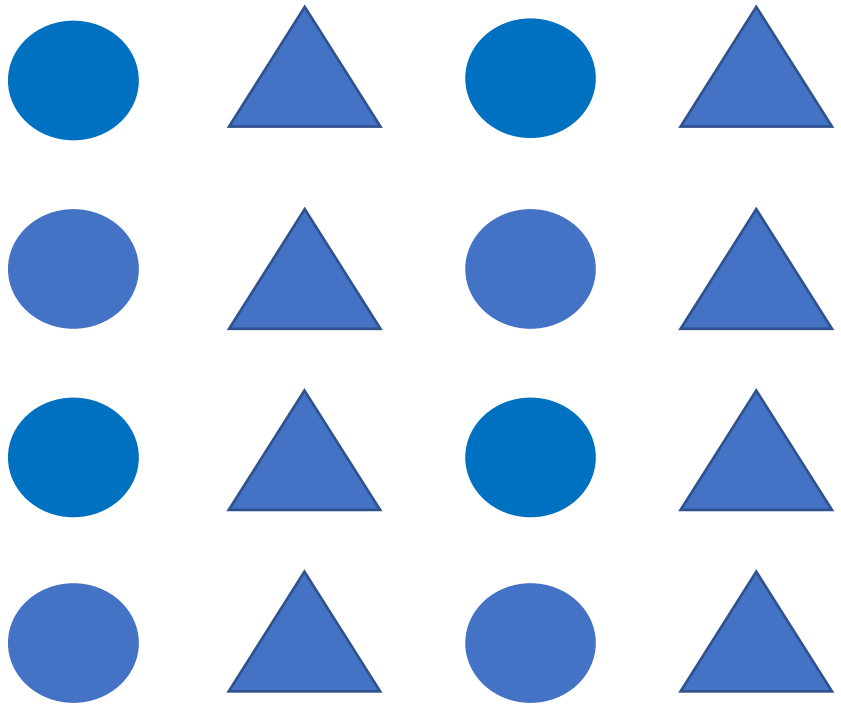
Pittsburgh Zoo & PPG Aquarium logo

Similarity



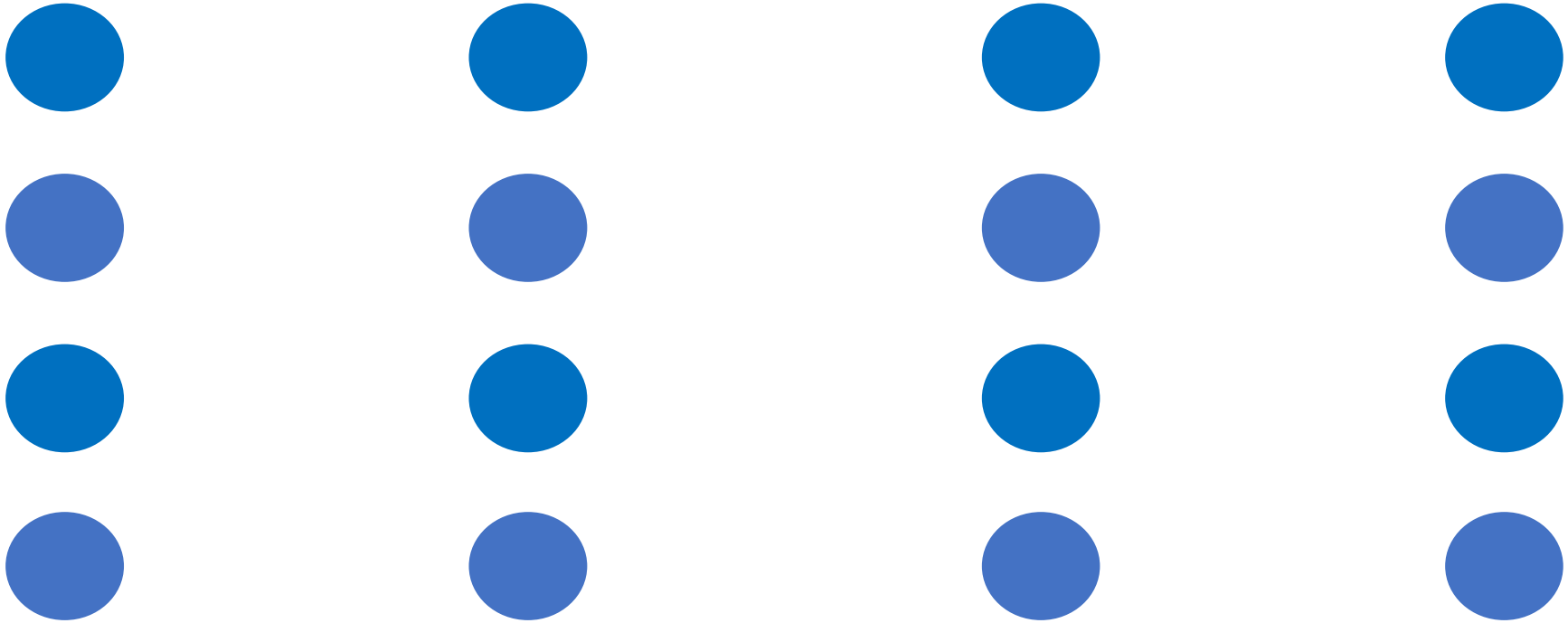
Rows or columns?

Similarity



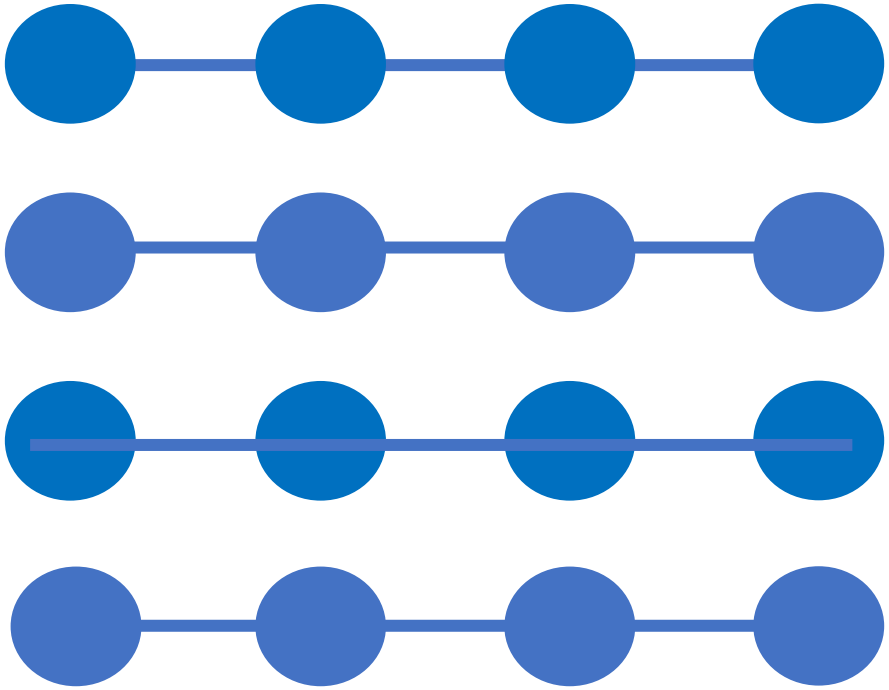
Rows or columns?

Proximity



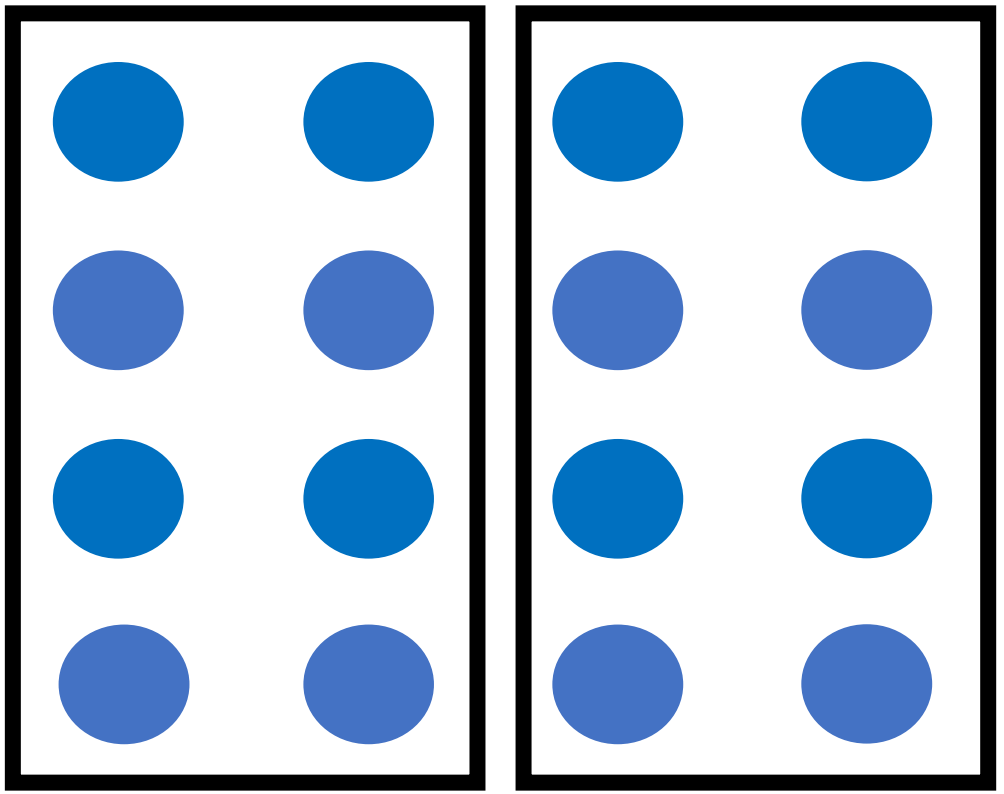
Rows or columns?

Connection



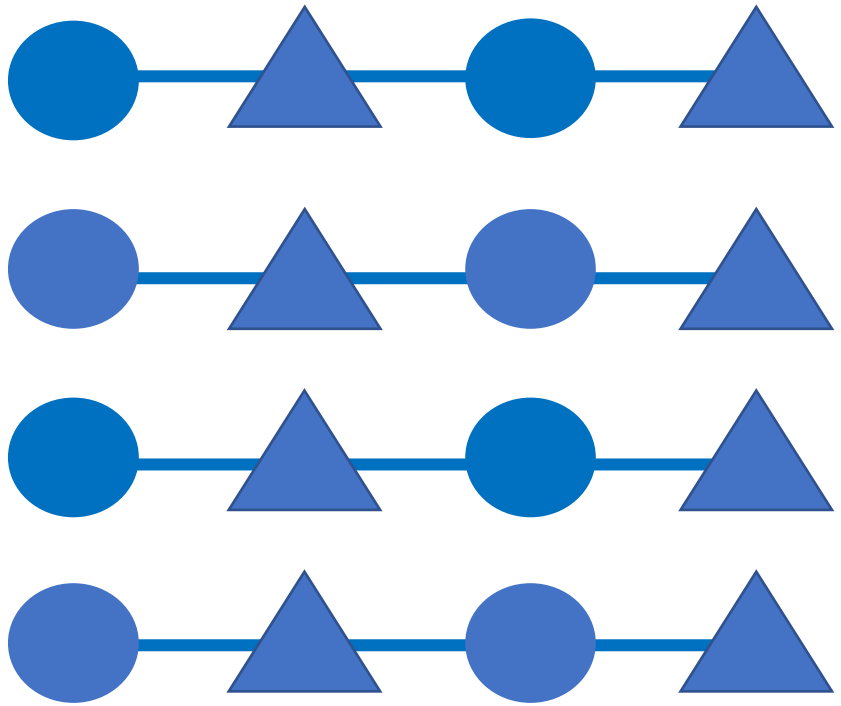
Rows or columns?

Enclosure

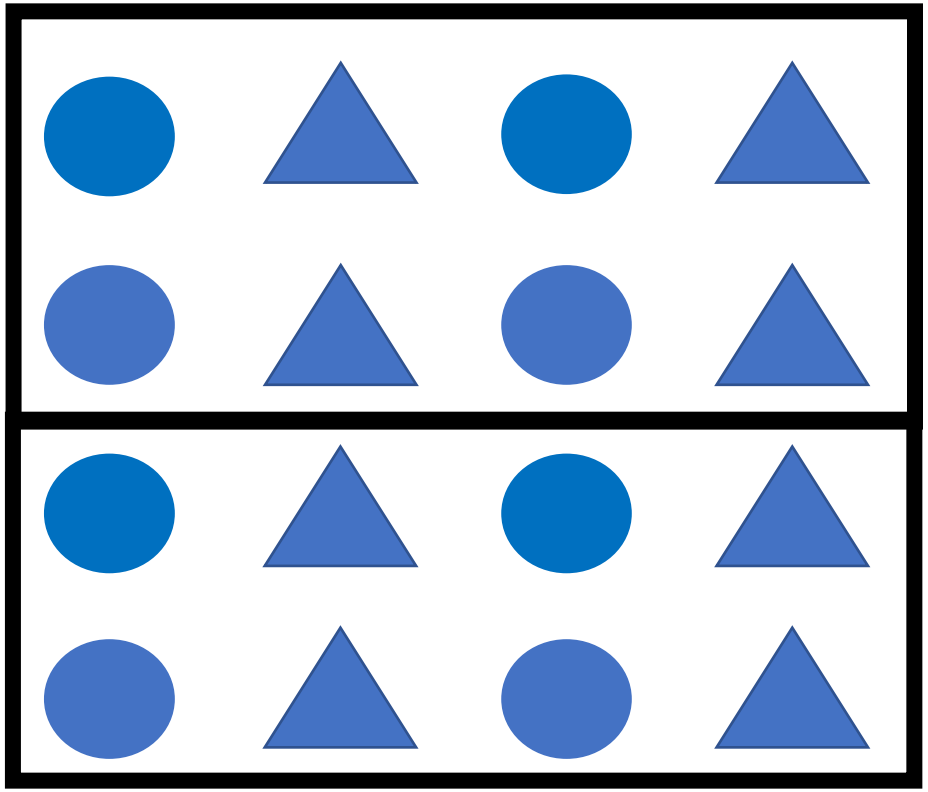


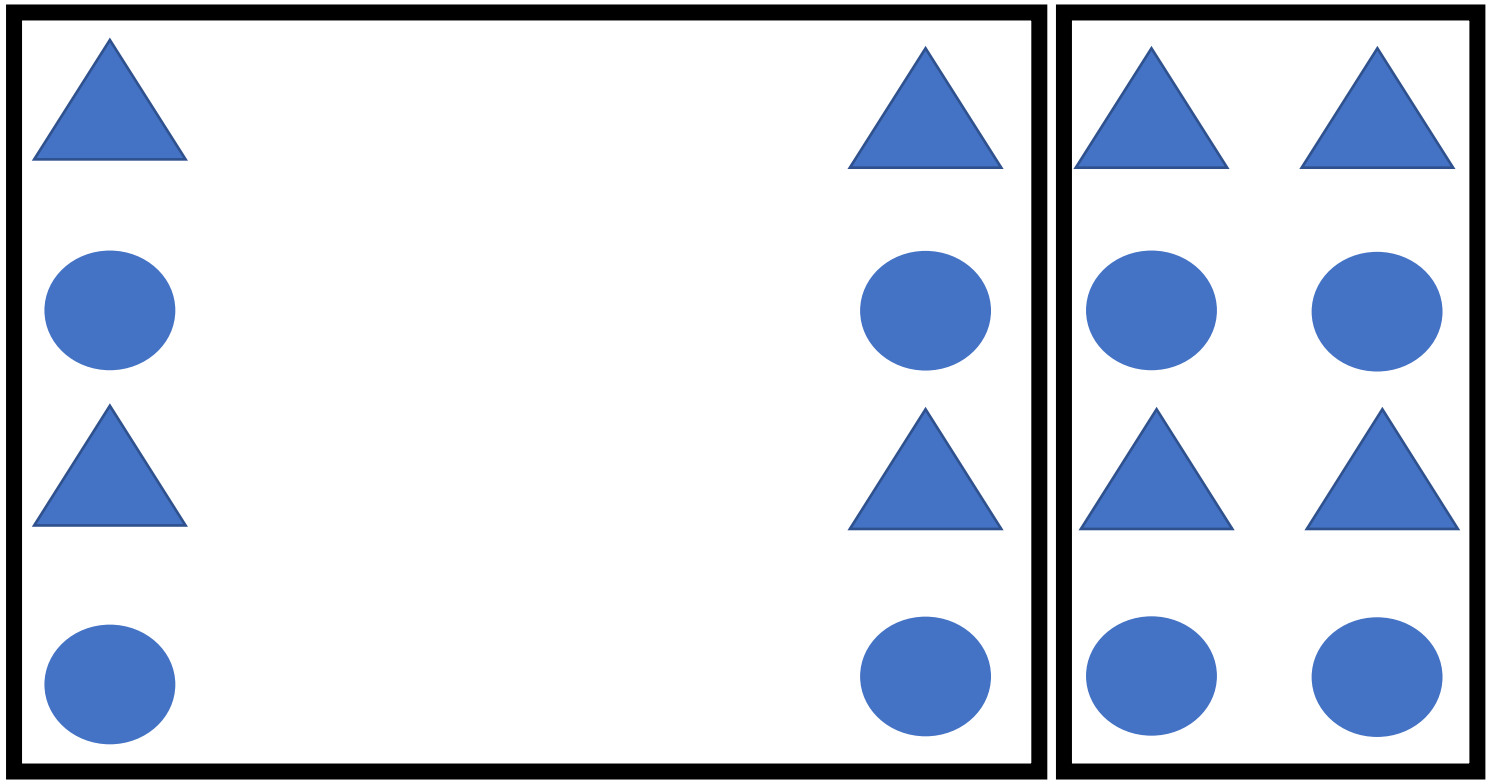
Rows or columns?

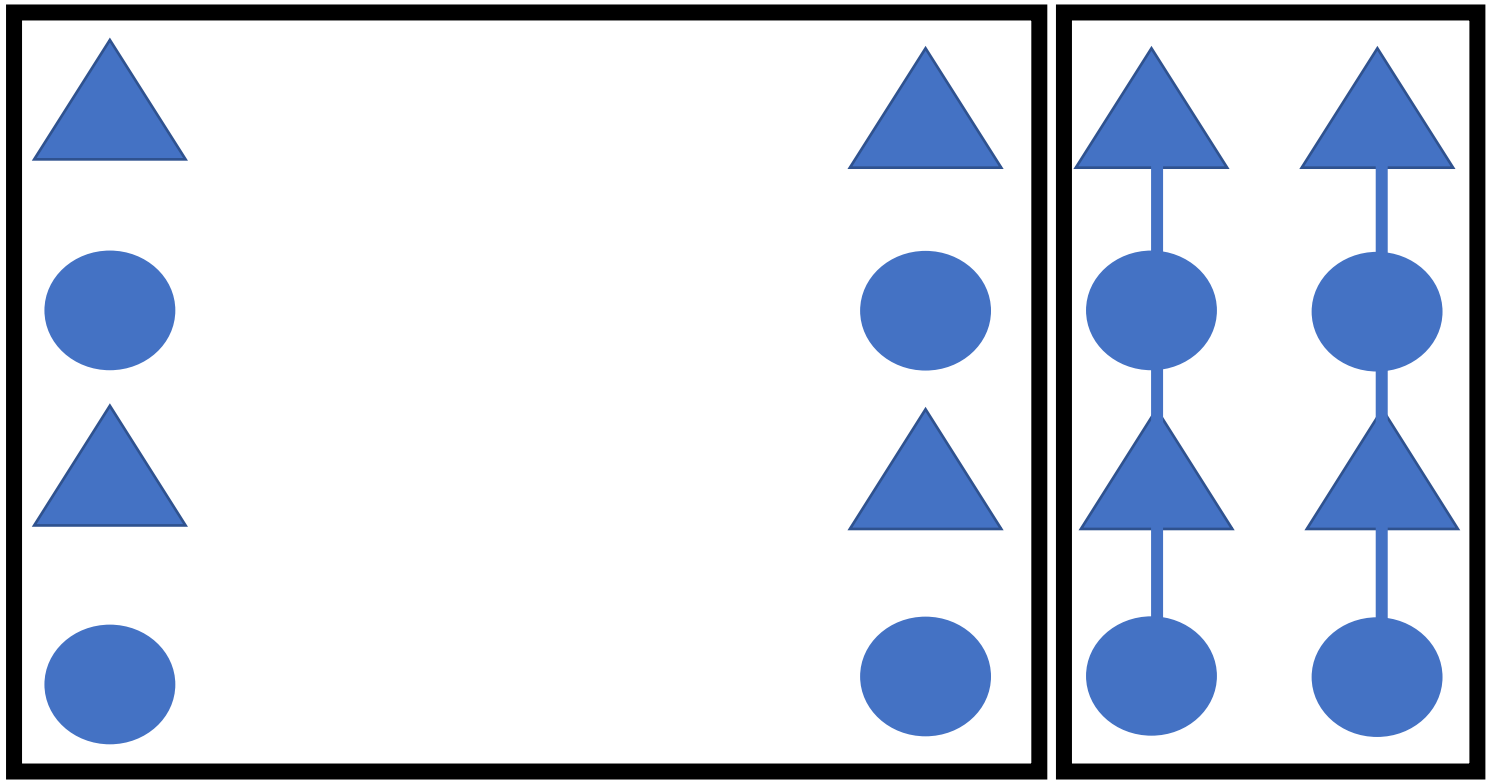
Some principles are stronger than others



Rows or columns?







Gestalt principles in hierarchy	Graphical elements that activate the principle
1. Enclosure	1. Facets
2. Connection	2. Lines
3. Proximity	3. White space
4. Similarity	4. Color/shape

Implications for practice

1. Know the graph's intent

2. Identify the elemental groupings in your data

3. Employ Gestalt principles to protect the integrity of the elemental groupings

Example: Hospital stays

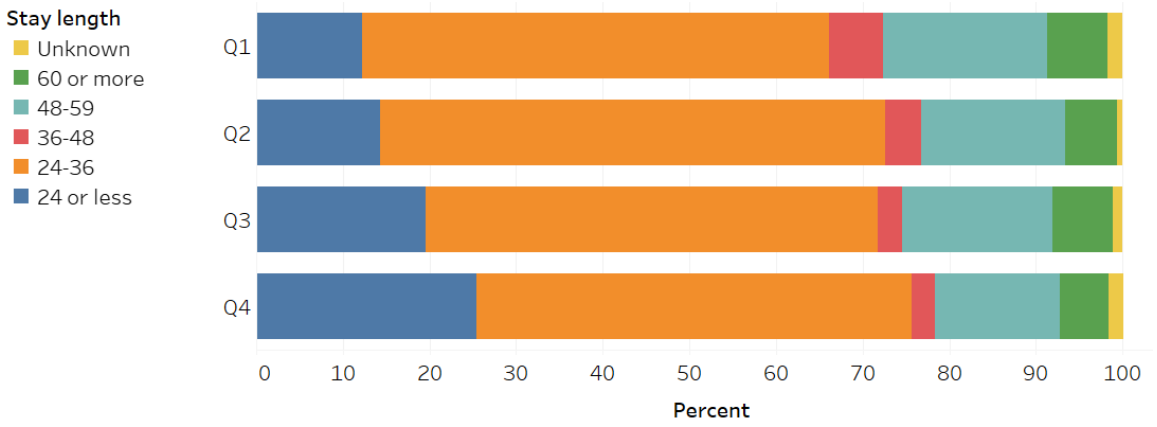
Data on quarterly duration of hospital stays following initiative to lower length of stay

	Duration of stay (in hours)					
Quarter	<=24	24 -36	36 - 48	48 - 59	>=60	Unknown
Q1	12.20%	53.90%	6.30%	18.90%	6.90%	1.80%
Q2	14.30%	58.30%	4.10%	16.70%	5.90%	0.70%
Q3	19.50%	52.20%	2.80%	17.40%	7.00%	1.10%
Q4	25.40%	50.30%	2.70%	14.40%	5.60%	1.70%

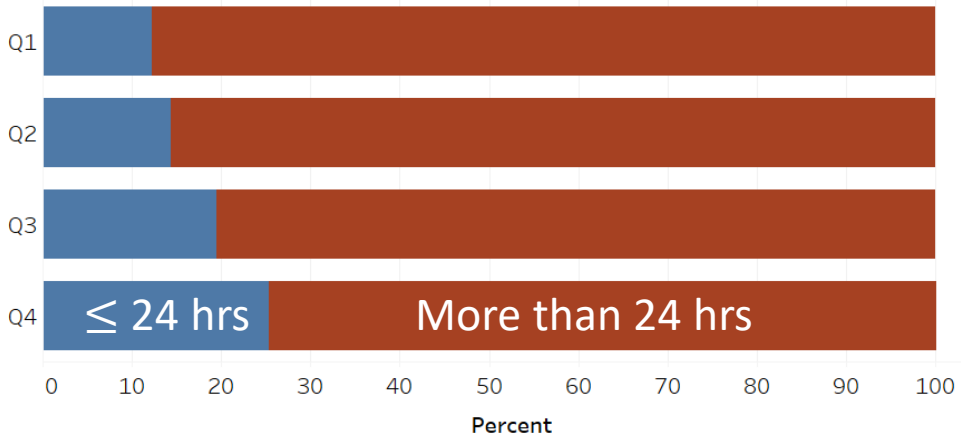
Question: Has the initiative worked?

Aka:

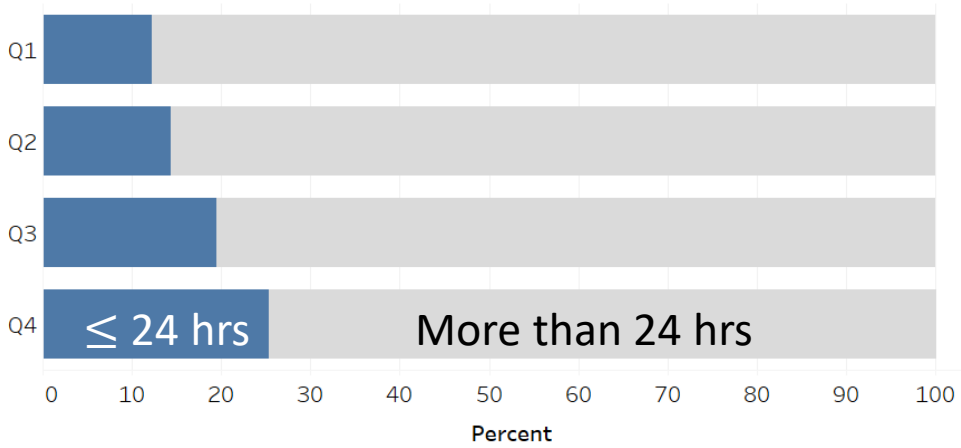
*Have the **short stays** increased?*



Graph A



Graph B



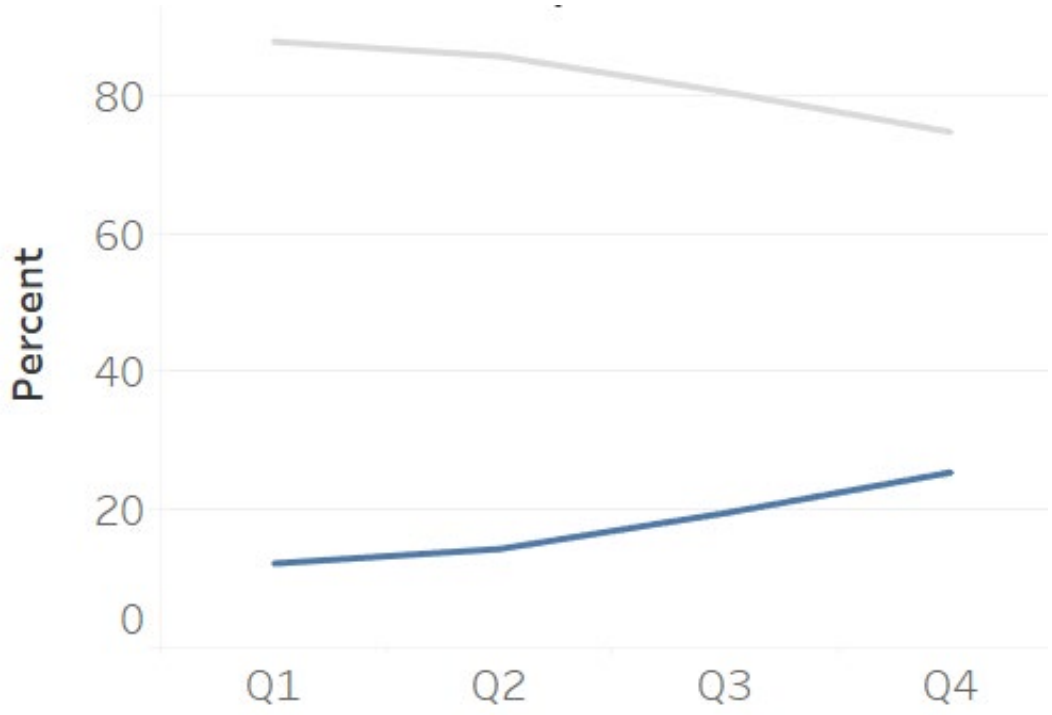
Graph C

Which graph is “best” and why?

What Gestalt principles are at play?

Similarity *and* connection:

The rate of **stays ≤ 24 hours** has increased relative to **stays > 24 hours**



Example

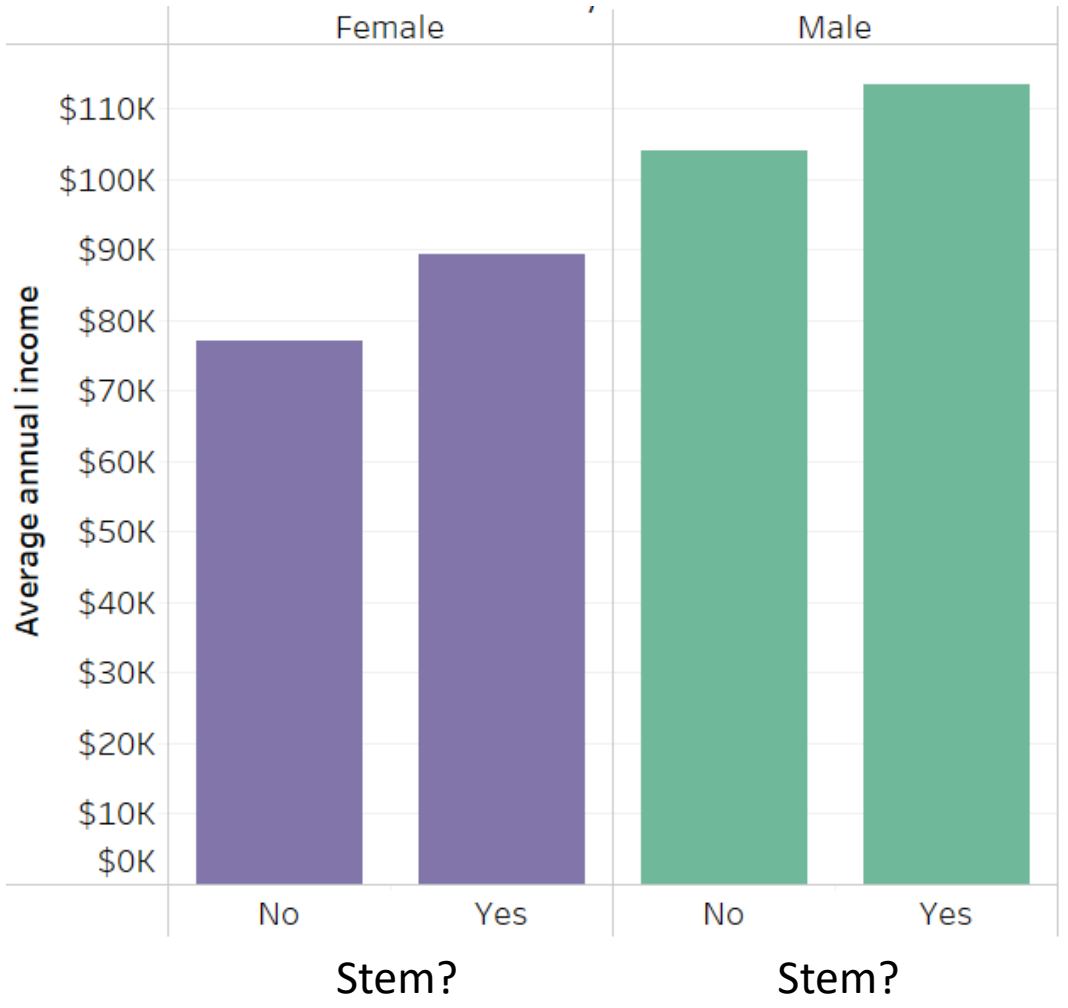
- Data: average income of Minnesotans, in 2016, among people with a master's degree working full time (source: American Community Survey)

Sex	STEM	
	No	Yes
Female	$77.0K	$89.3K
Male	$104.1K	$113.6K

- Question: *How does the gender gap differ by field (STEM vs non-STEM)?*

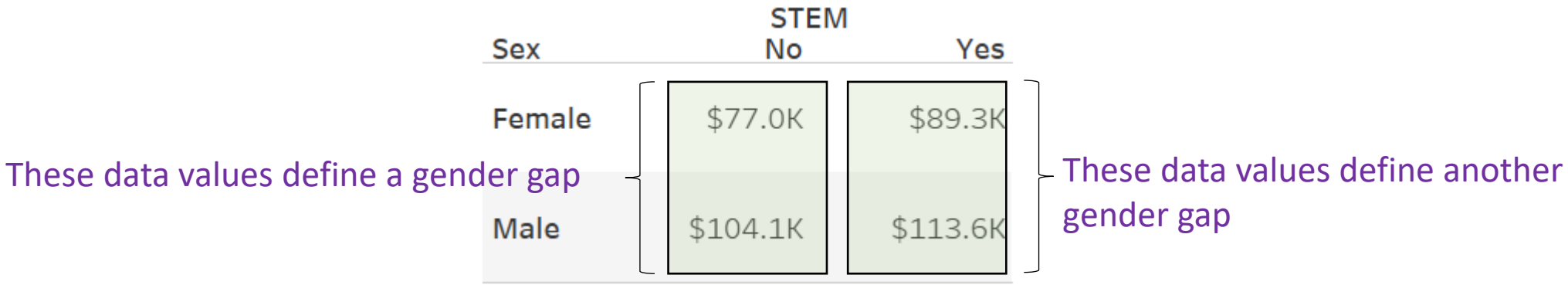
Thoughts on this graph?

Question: *How does the gender gap differ by field (STEM vs non-STEM)?*



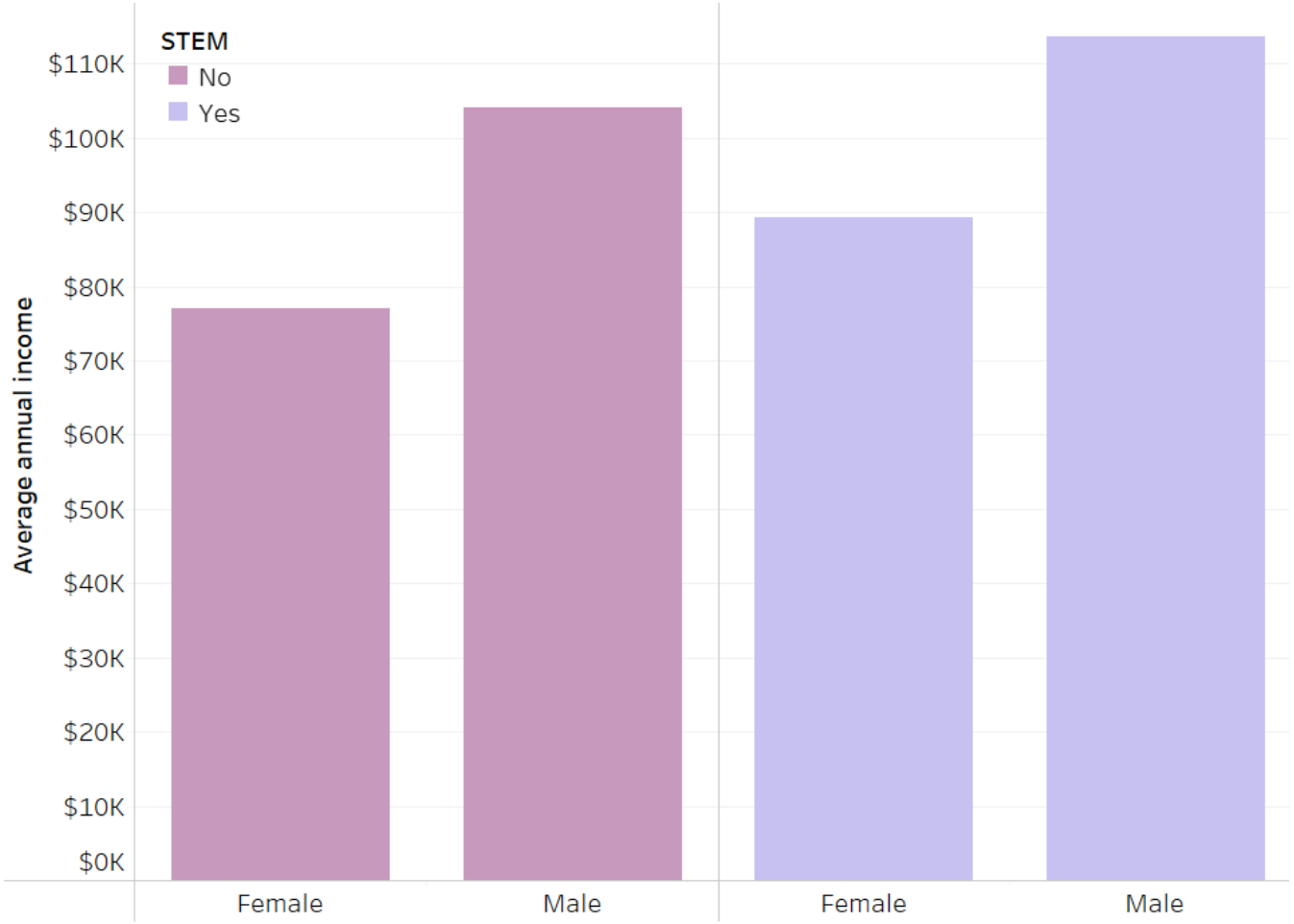
How does the gender gap differ by field (STEM vs non-STEM)?

- We see two groups:
  - Purple bars on right
  - Green bars on left
- But these are <u>not</u> the elemental groupings! The geometries representing gender gap members are:
  - Far apart (~~proximity~~)
  - Different colors (~~similarity~~)
  - In different facets (~~enclosure~~)



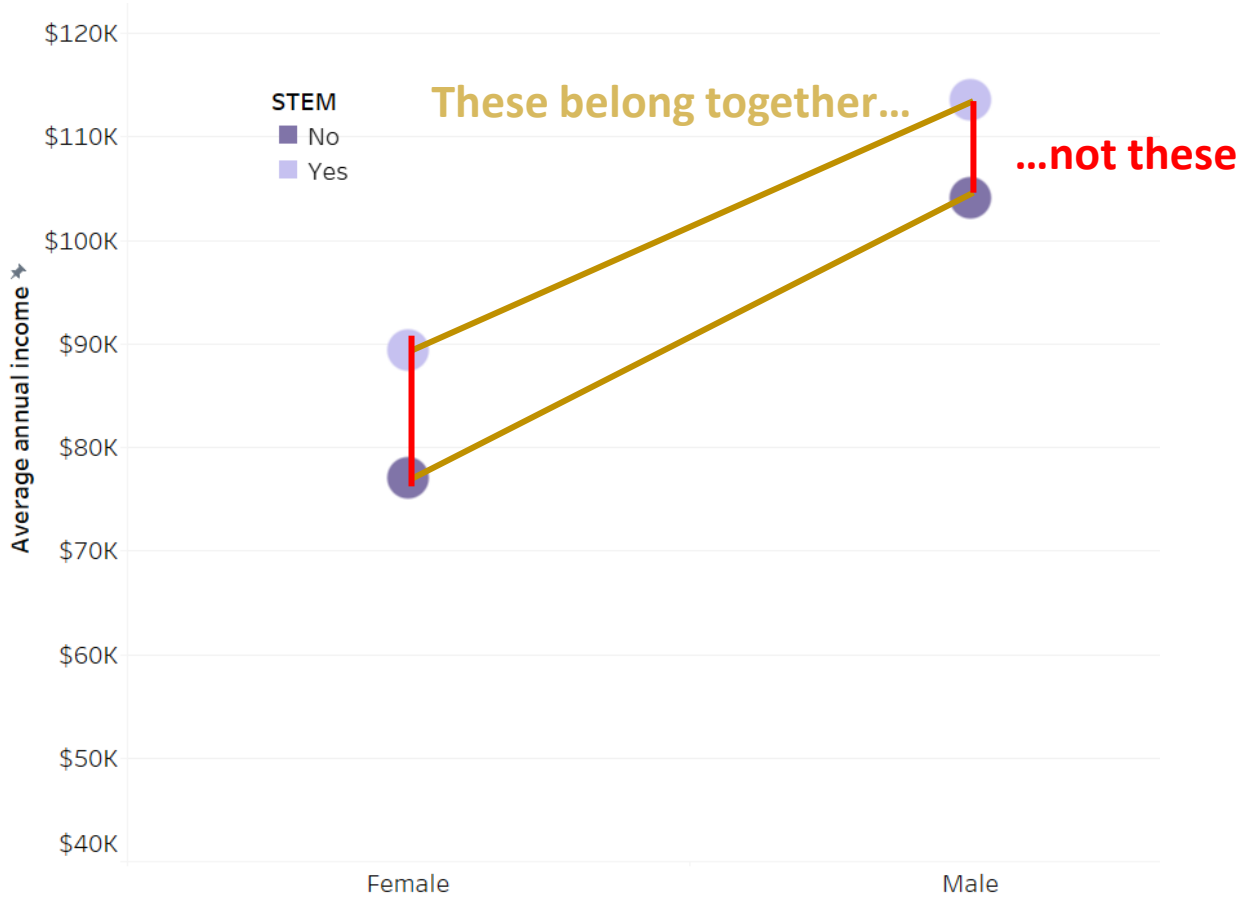
- ($77.0K, 104.1K) constitute one elemental grouping
  - They should be any/all of: *close*, *similar*, *connected*, and *enclosed* in the same region
- ($89.3K, $113.6K) constitute a second elemental grouping
  - They should be any/all of: *close*, *similar*, *connected*, and *enclosed* in the same region

Better



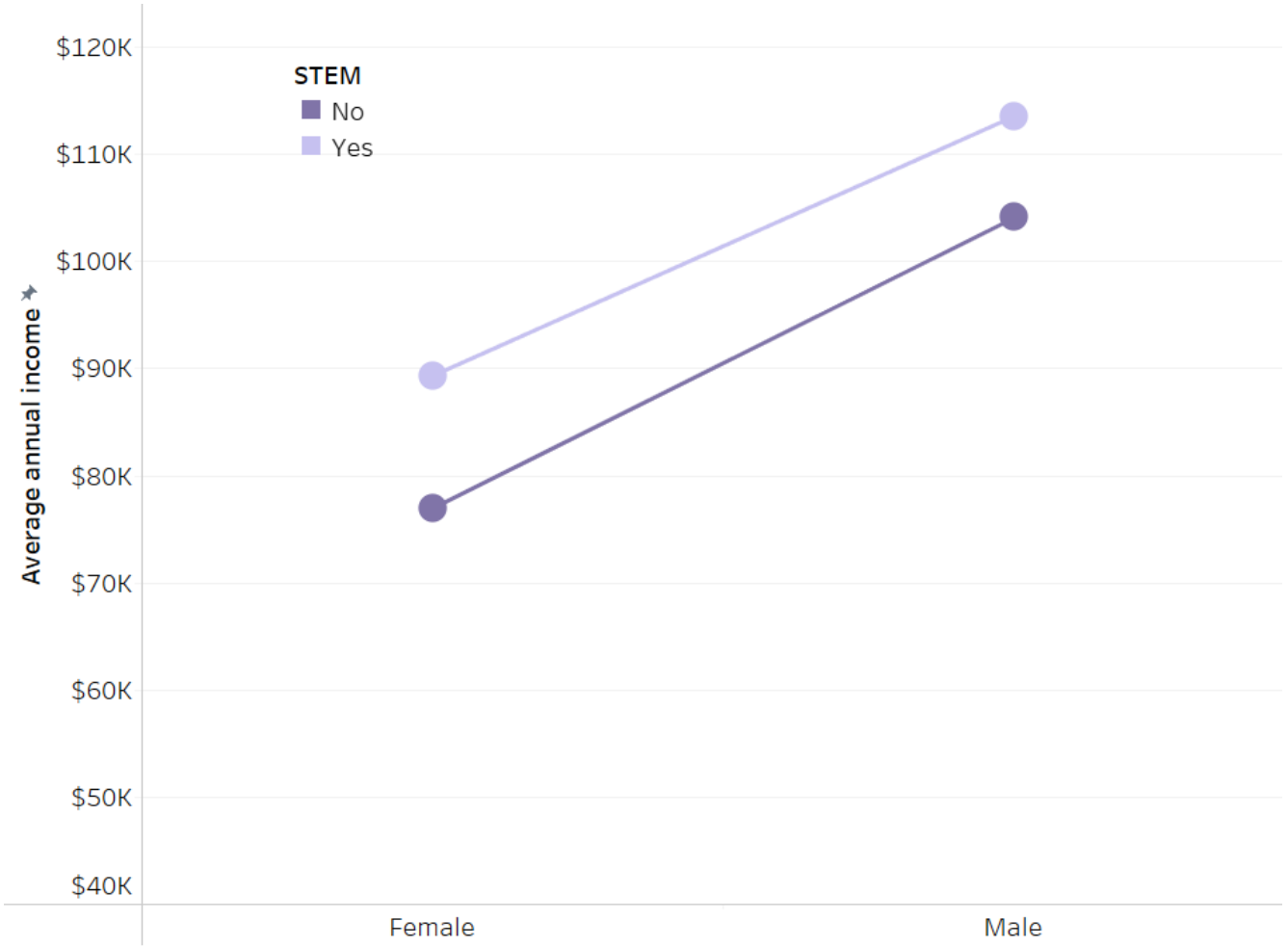
Geometries representing data points within each elemental grouping are *similar* and *proximal* (close).

Room for improvement



Geometries representing items within each elemental grouping are *similar,* but greater proximity of items <u>across</u> elemental grouping results in unclear communication of what belongs together (proximity > similarity)

So improved



Adding *connection* strengthens the sense of what belongs together

ACTIVITY TIME!!

https://bit.ly/3OM7ff6

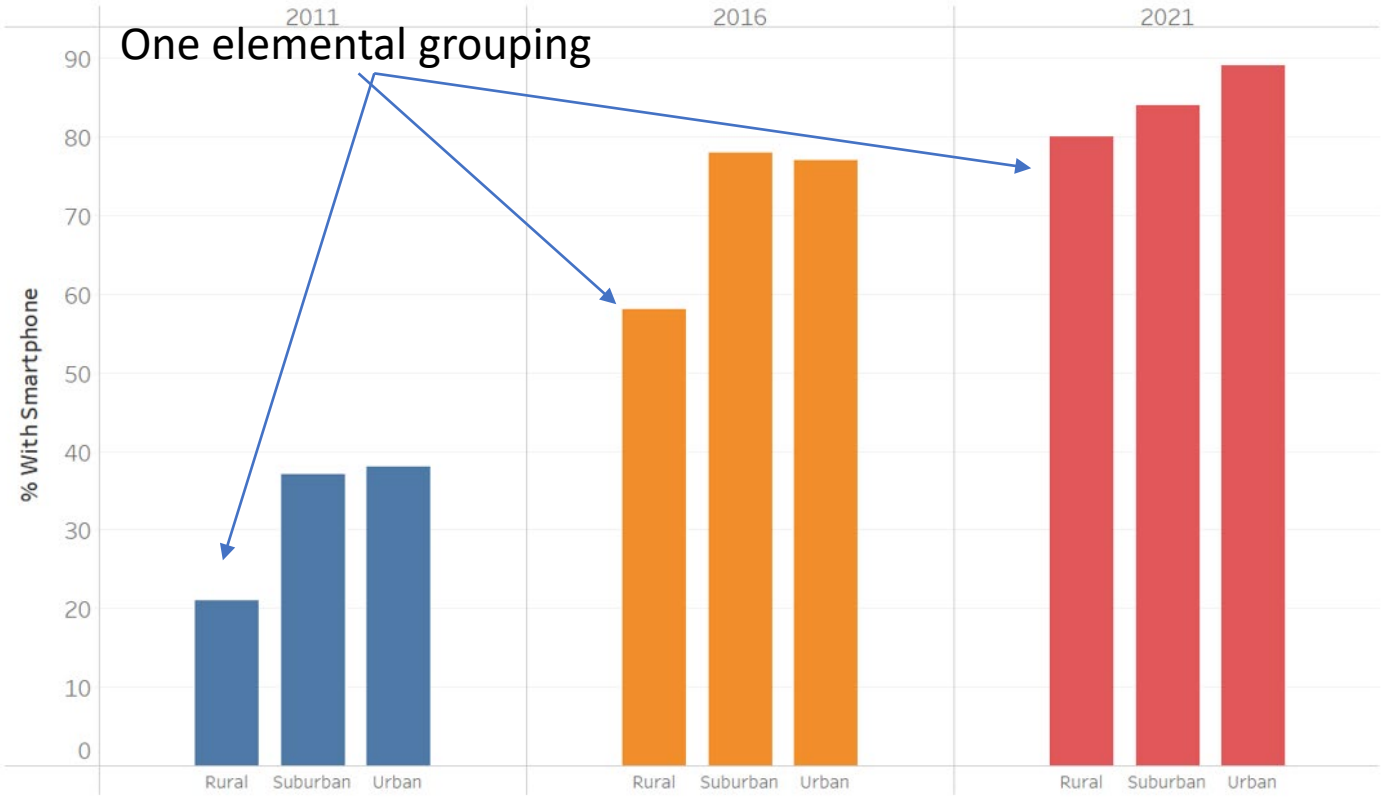
Student work and comments

Despite growth, rural Americans have consistently lower levels of smartphone ownership than urbanites and suburbanites.

Year	Residence	% with smartphone
2011	Rural	21
2011	Suburban	37
2011	Urban	38
2016	Rural	58
2016	Suburban	78
2016	Urban	77
2021	Rural	80
2021	Suburban	84
2021	Urban	89

One elemental grouping

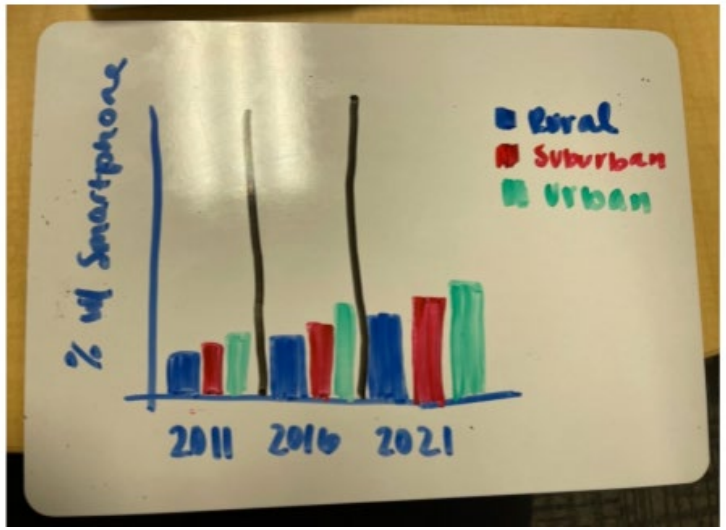
- Headline subject: ***growth***
- Elemental grouping: data points that would depict ***growth***

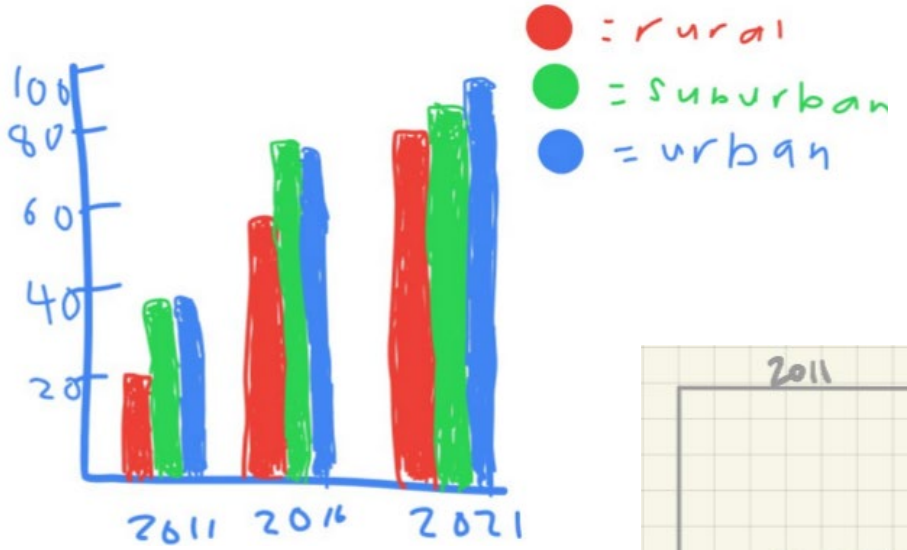


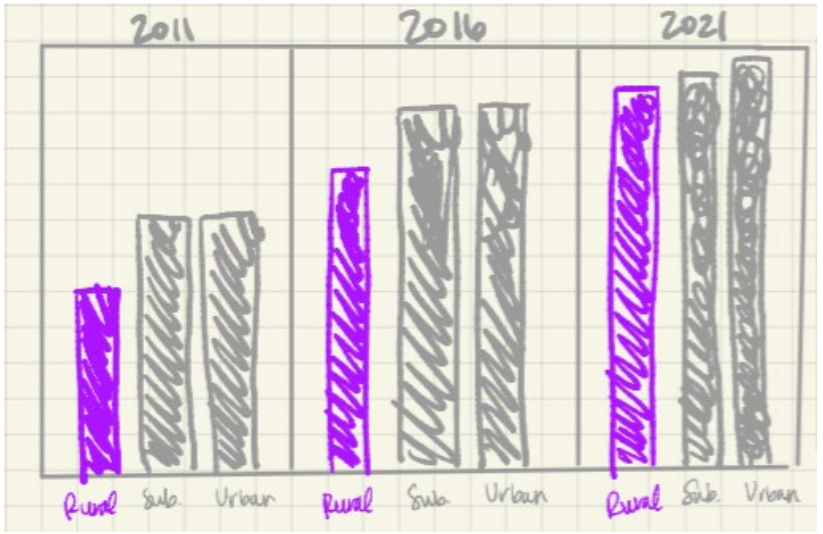
Original design: the elemental grouping members are:

- ***Dissimilar***
- ***Far apart***
- ***In separate enclosures***

- Original grouping preserved
- Color used to make the items inside the elemental grouping similar, but they are still difficult to perceive as a group because they are in separate enclosures and/or far apart.

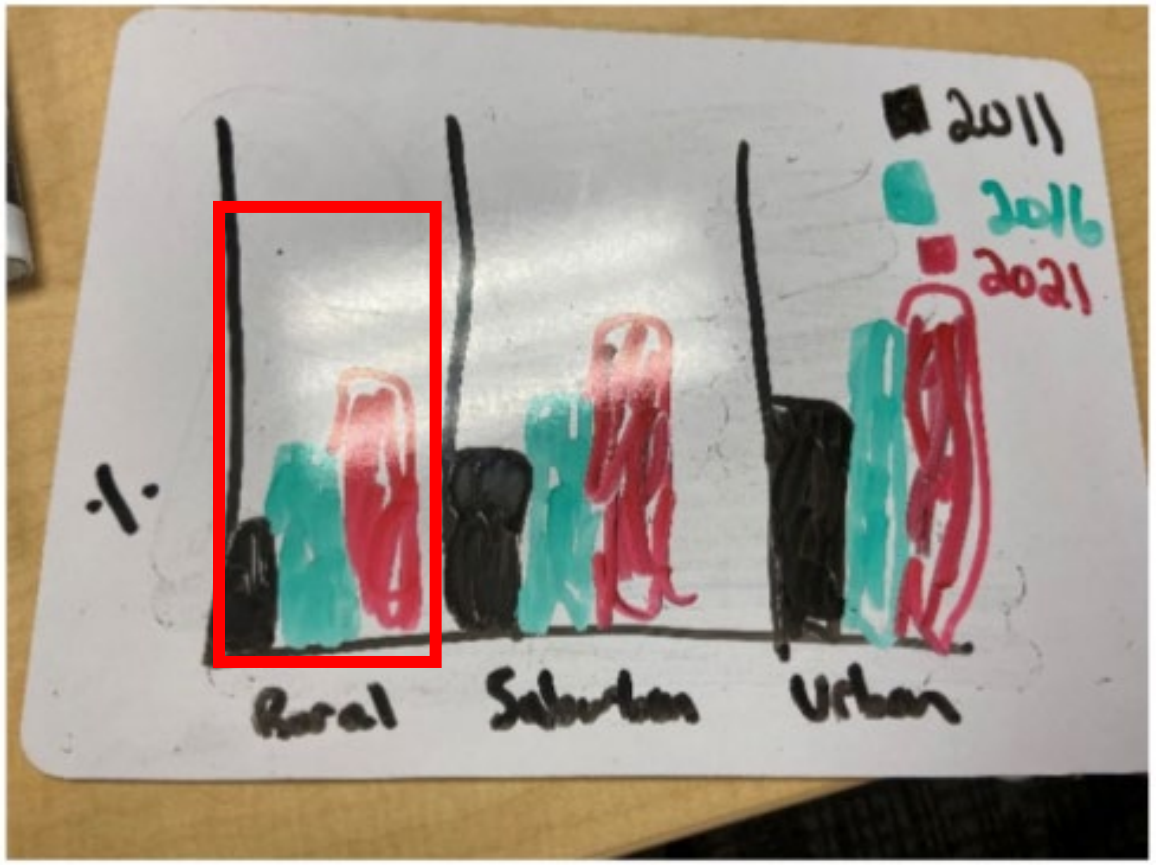


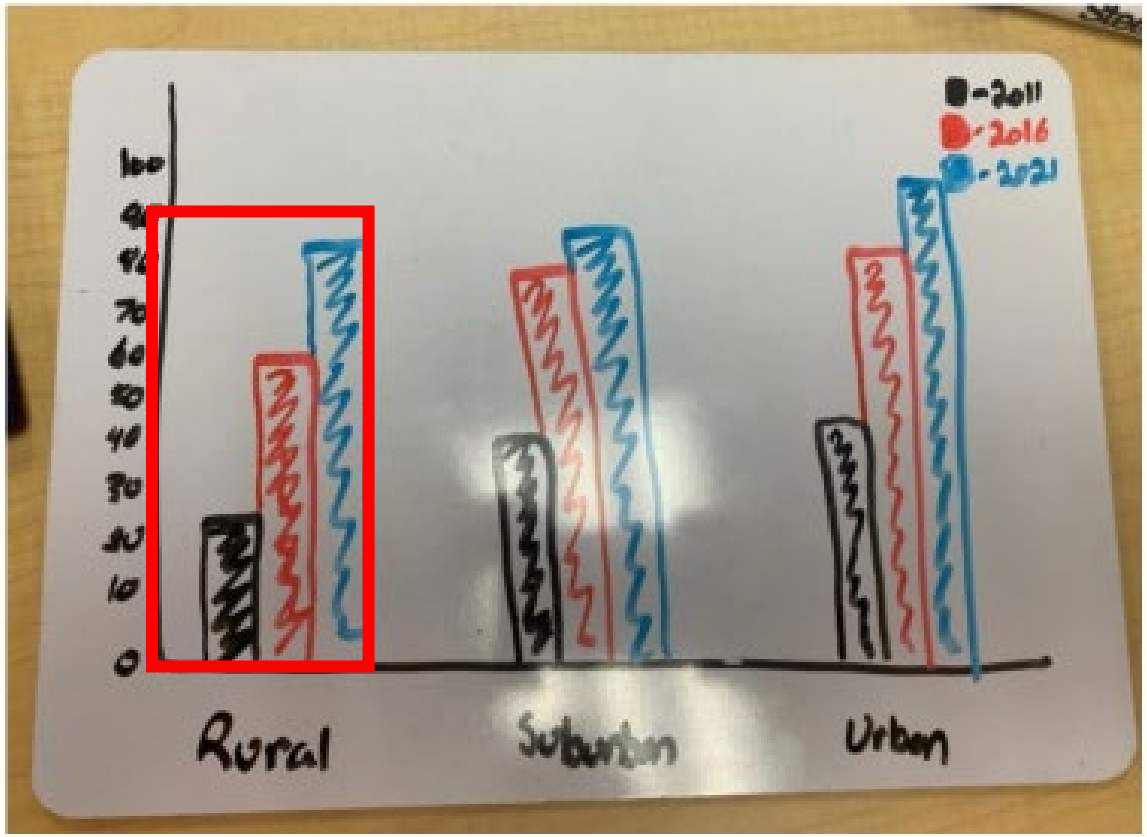




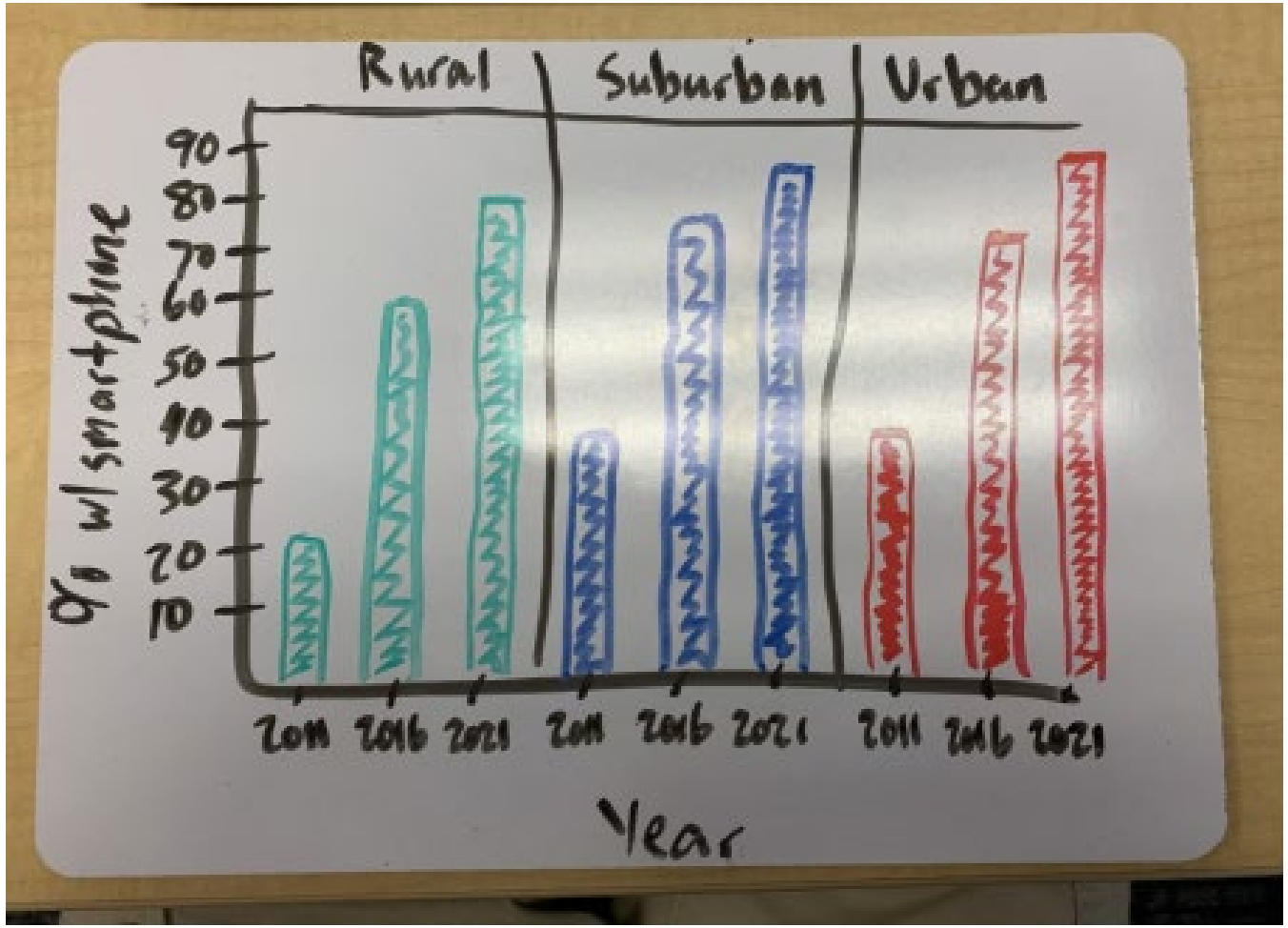
Improved enclosure / proximity

However, the items within each elemental grouping (the rural, suburban, urban triplets) are different colors (dissimilar): better not to map Year to color.

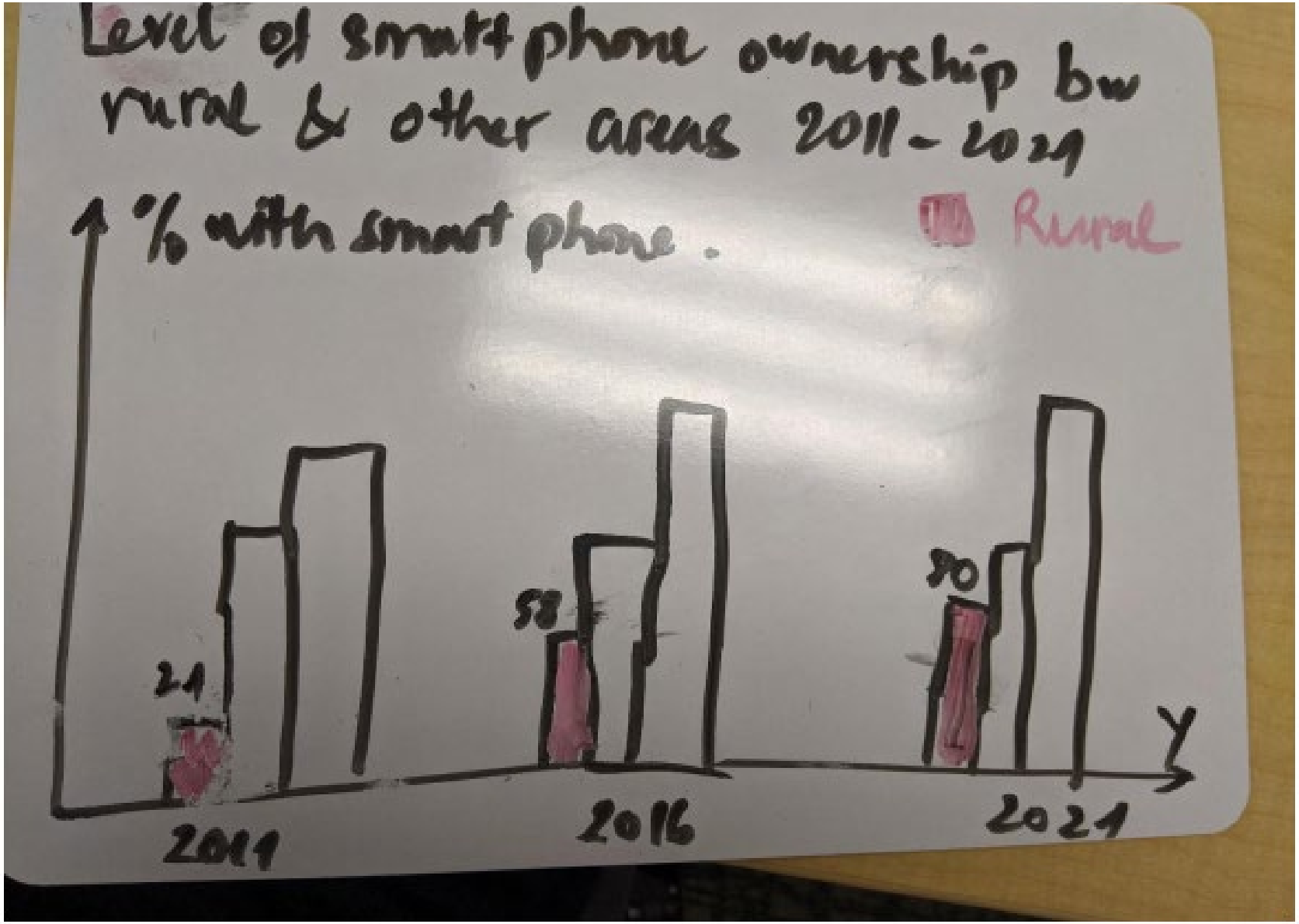




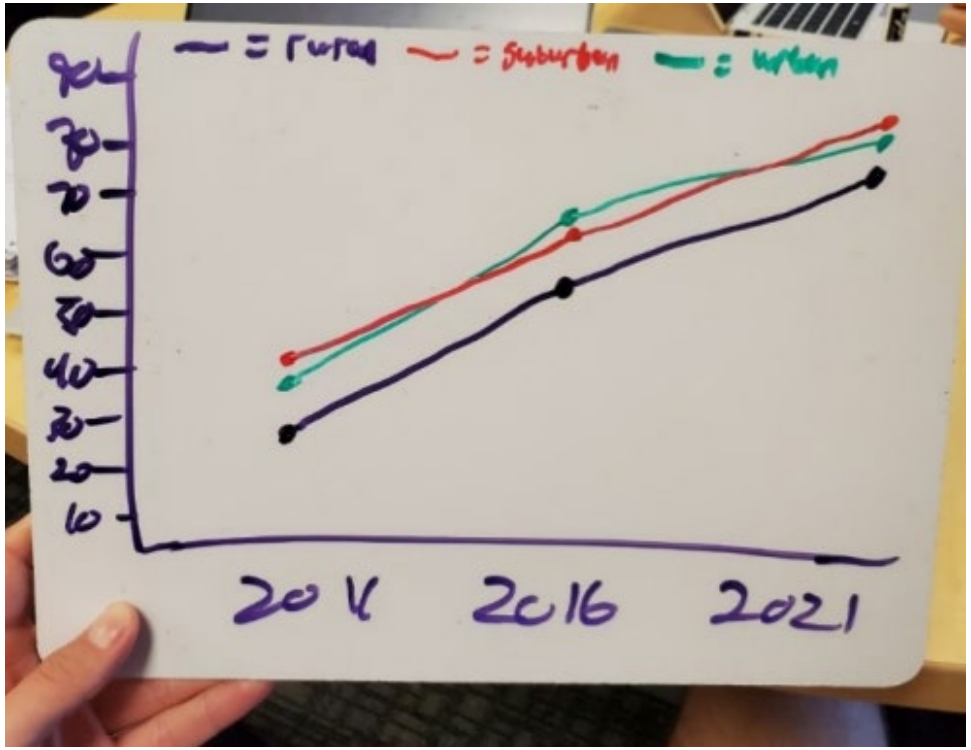
Members of elemental groupings are now similar and in same enclosure

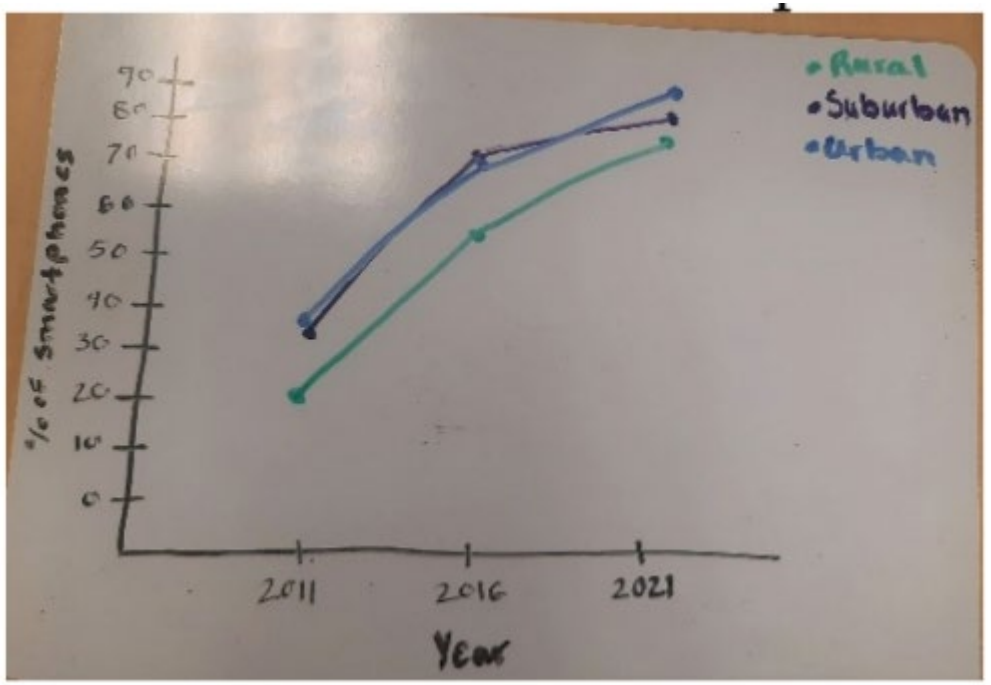


Boxes removed, “rural” set apart from the other two regions
Suburban/urban now clearly “other” from rural (but need labeling somehow)

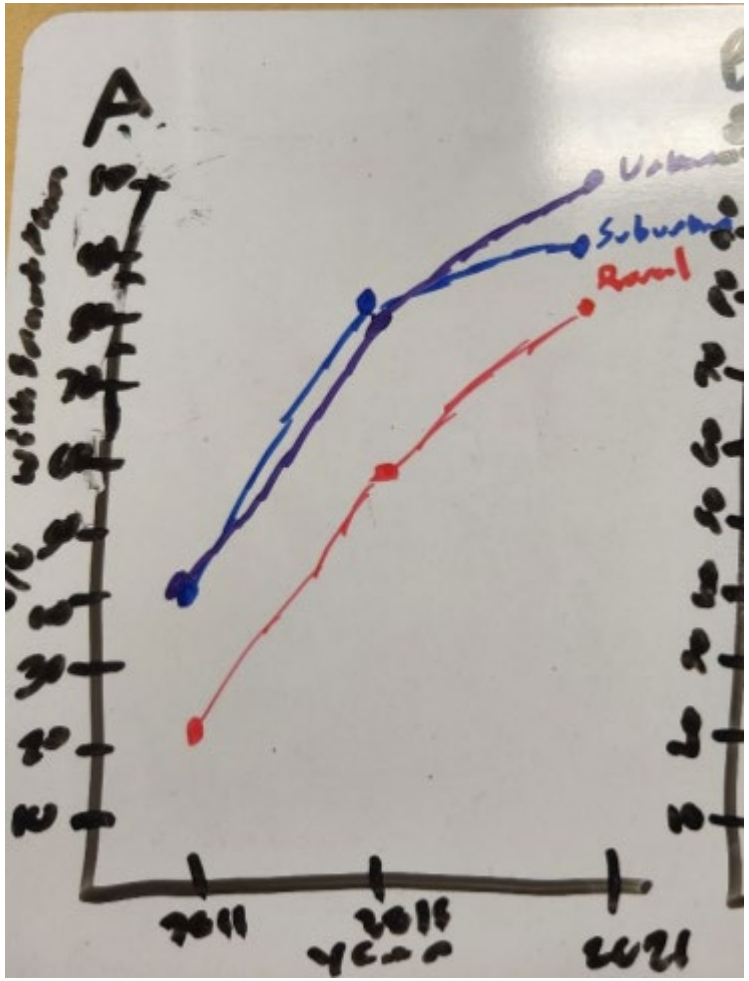


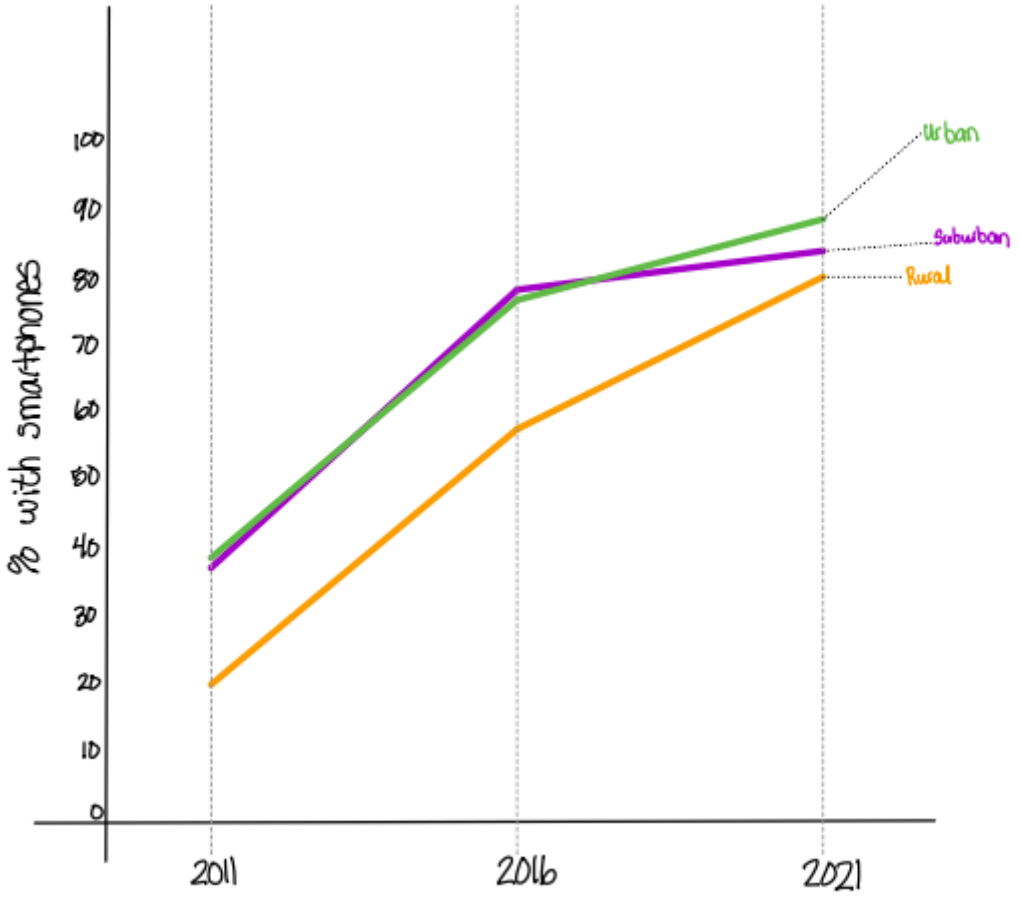
Added connection to enhance the sense of “what belongs together” – growth by region now much more apparent, lack of enclosure makes it easy to compare the trends



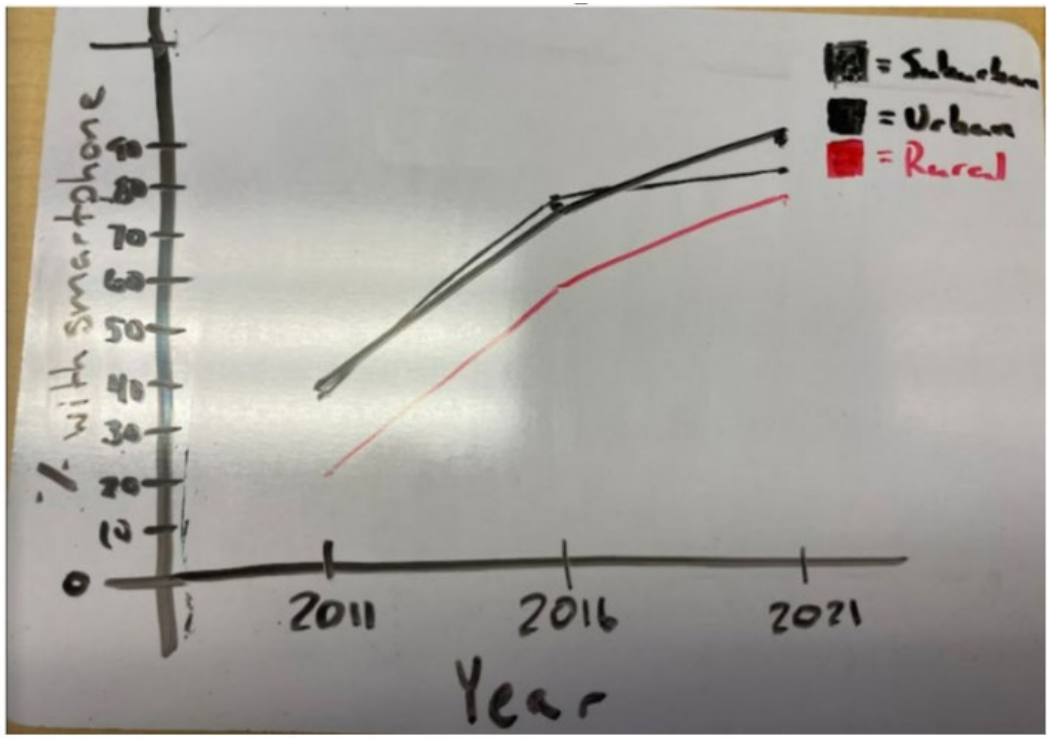


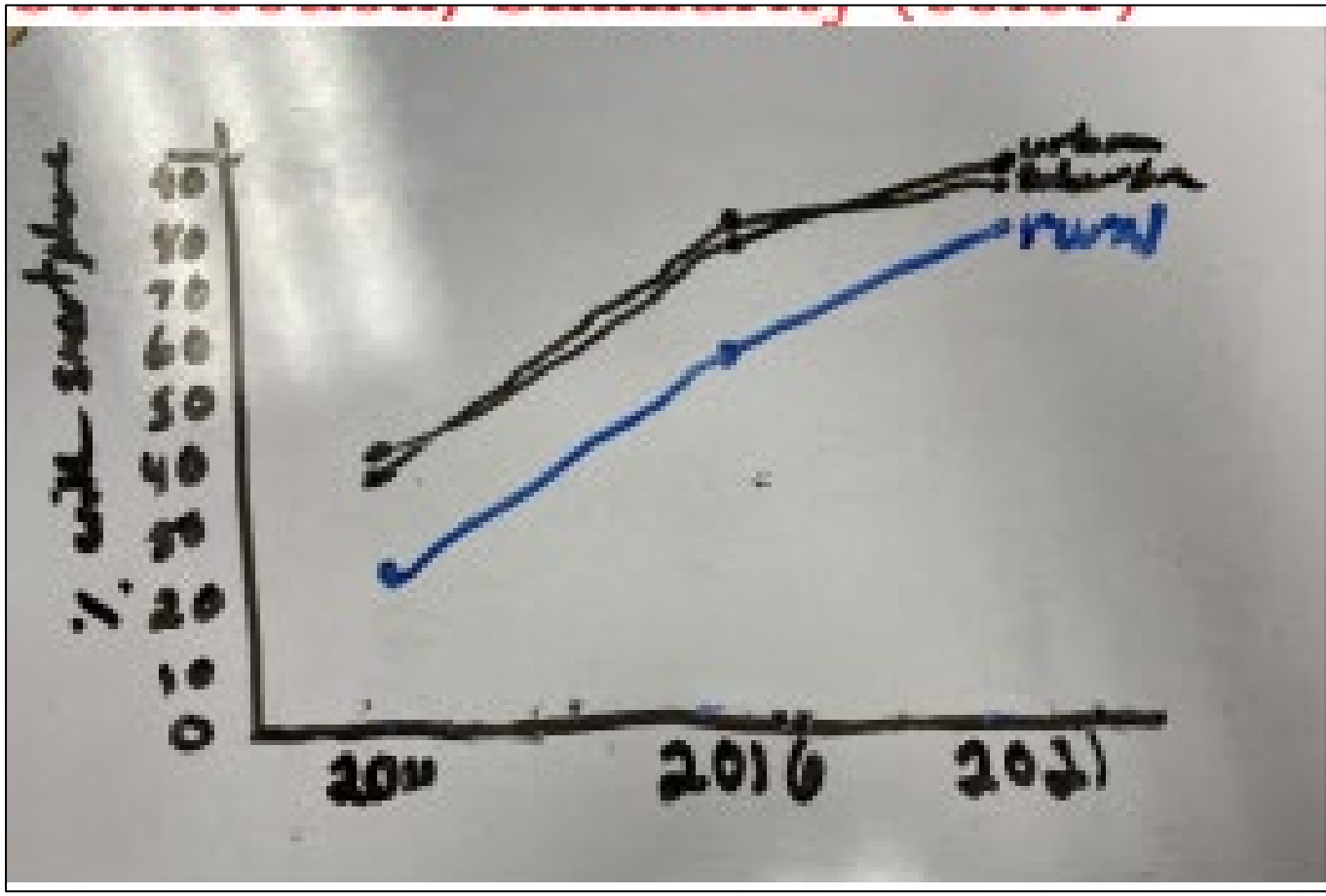
Enhanced, color-coded line-end labeling allows doing away with legend





Emphasizing “rural” enhances the clarity of what viewers should focus on (though the legend doesn’t work when 2 colors are identical)



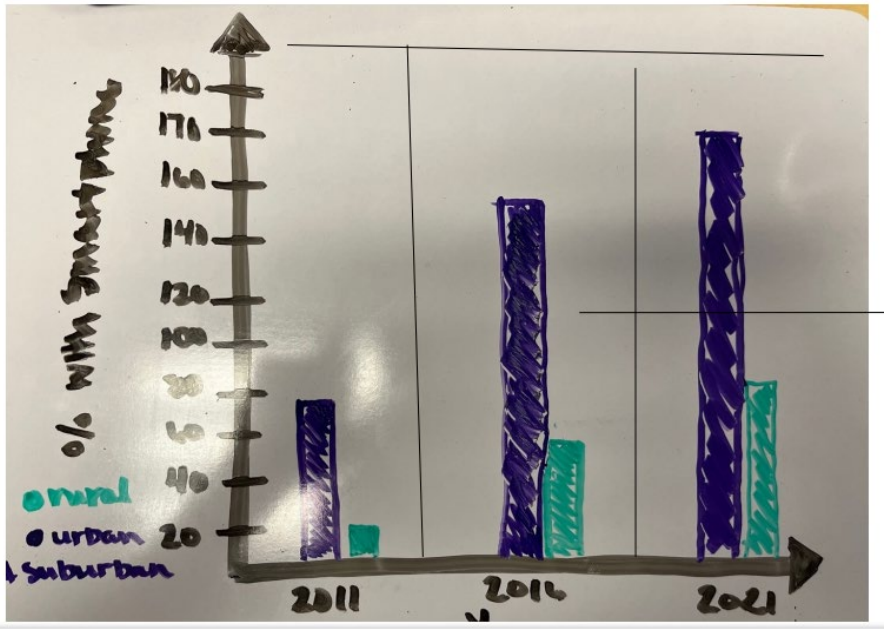


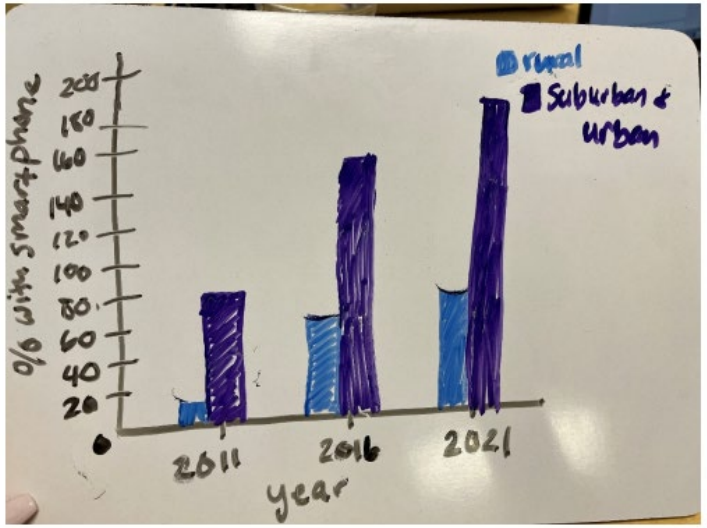
CAREFUL WITH COMBINING!!

Year	Residence	% with smartphone
2011	Rural	21
2011	Suburban	37
2011	Urban	38
2016	Rural	58
2016	Suburban	78
2016	Urban	77
2021	Rural	80
2021	Suburban	84
2021	Urban	89

No combination of these %’s is required to sum to 100%.

Careful with combining into “other”! 2016 and 21 exceed 100%.





Awesome sources for your data viz class

- Stephen Few. 2012. *Show Me the Numbers: Designing Tables and Graphs to Enlighten*. Analytics Press; Second edition, Burlingame, CA.

- Cole Knaflic. *Storytelling with data* (2015) and *Storytelling with data: let's practice!* (2020) Also the storytellingwithdata.com blog.

- https://www.makeovermonday.co.uk/