

Lecture 2: When You Need Graphs and How We See Graphs and Merging

August 31, 2026

Course Administration

① Tutorial 1 quiz

Course Administration

- 1 Tutorial 1 quiz
- 2 Questions/issues with readings?
- 3 Make sure you're signed up for Piazza
– email me if you are not
- 4 One-page proposal is due next class in
two weeks

Course Administration

- ① Tutorial 1 quiz
- ② Questions/issues with readings?
- ③ Make sure you're signed up for Piazza
– email me if you are not
- ④ One-page proposal is due next class in two weeks

- ⑤ Be sure to check online listing for good/bad/ugly
 - Linked in pinned email on Piazza
 - I changed dates for one of you to even up
 - If you're joining the class, make sure you sign up
 - Let me know if this poses problems
 - Each week, I will always name who is next
- ⑥ Anything else?

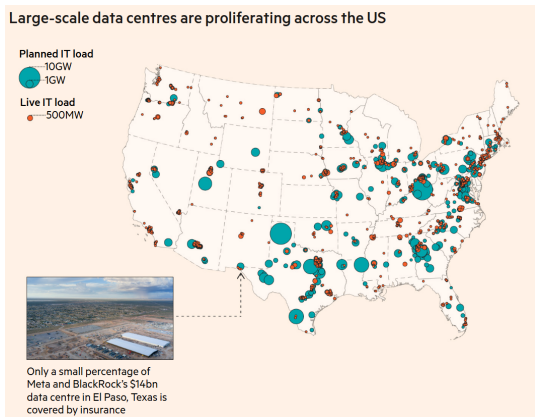
Next Week's Good Bad and Ugly

Finder, post link Wed. by noon. Find a graph relating to one of the principles we're discussing today.

Finder	Commenter
Harini	Jacob

Add yourself ASAP if you're not on the google sheet.

Today's Graphic from Jacob, No Commentor!



“The multiplying risks of financing data centres,” *Financial Times*, August 26, 2026.

[\[link\]](#)

Plan for Today

- ① Few reading
- ② Merging
- ③ R questions

1. Few:

Visual Perception and Graphical Communication

When Should You Use Tables vs. Graphs?

- Tables are for when

When Should You Use Tables vs. Graphs?

- Tables are for when
 - you care about the **actual numbers**
 - you have **very** few numbers

When Should You Use Tables vs. Graphs?

- Tables are for when
 - you care about the **actual numbers**
 - you have **very** few numbers
- Graphs are for when

When Should You Use Tables vs. Graphs?

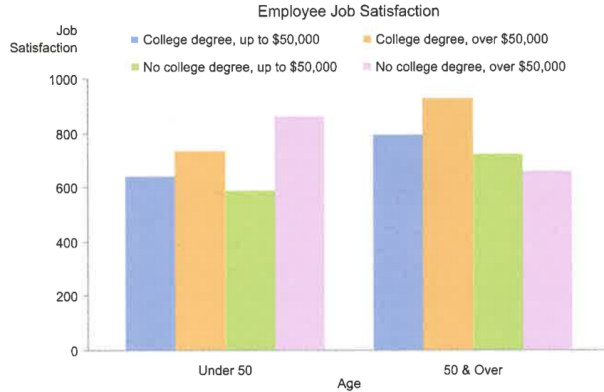
- Tables are for when
 - you care about the **actual numbers**
 - you have **very** few numbers
- Graphs are for when
 - you care about trends or general tendencies
 - you have more numbers than a table can support
 - the exact values are not critical
 - you wish to highlight a particular relationship

Starting with the Table

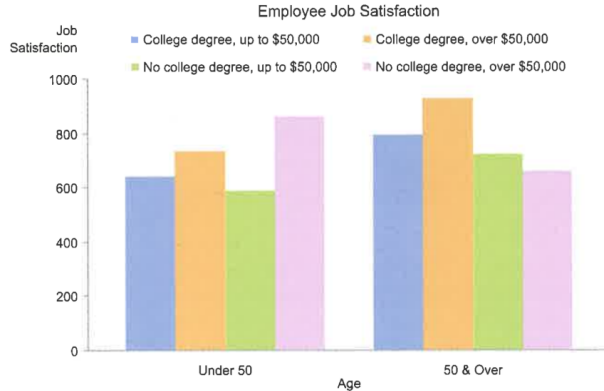
Income	College Degrees		No College Degrees	
	Under 50	50 & Over	Under 50	50 & Over
Up to \$50,000	643	793	590	724
Over \$50,000	735	928	863	662

Few, Chapter 3, Figure 3.13

Version One of a Set of Numbers



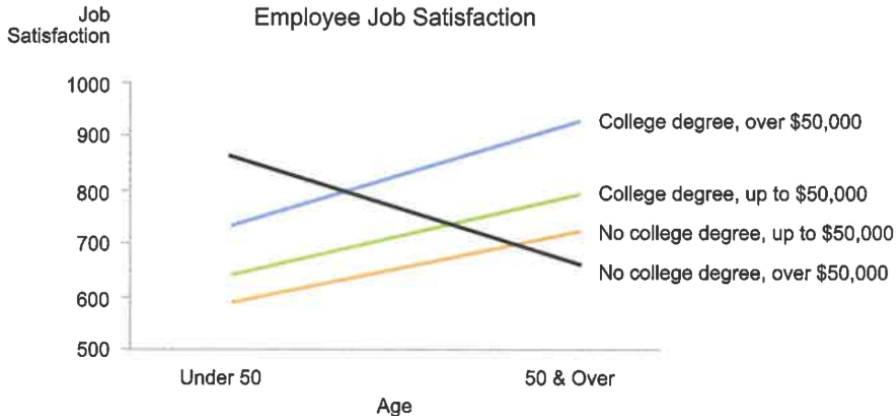
Version One of a Set of Numbers



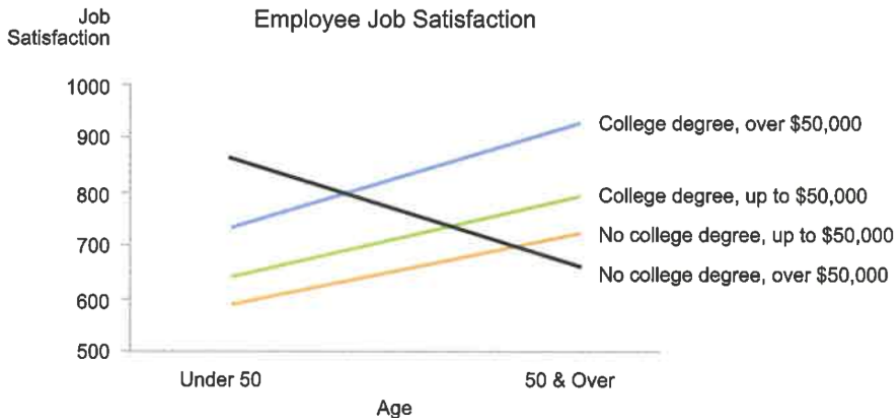
What do you think the point of this picture is?

Few, Chapter 3, Figure 3.15

Version Two of the Same Set of Numbers



Version Two of the Same Set of Numbers



And the point of this picture?

Few, Chapter 3, Figure 3.14

Choose the Graph that Leads the Reader to Your Answer

GRAPH CHOICE CHART

Does your question ask you...

about the **variability** of a group of data points? (i.e. the range of the data, the shape of the distribution, or what the center of the data is)

1. Do all high rises rise to the same height?
2. How variable are wind speeds in Germany?
3. What is the range and distribution of incomes in Dubai?

to compare **two or more groups** to decide if the groups are the same or different?

if two numeric factors are **correlated**?

1. Is the temperature inside the house correlated with the temperature outside?
2. How did electricity used by the kitchen about fluctuate during the past week?

how a **total** is proportioned into sub-groups? (Or what proportion a sub-group is of a total?)

1. What were Brazil's most significant exports in 2019?
2. What proportion of global electricity production comes from wind?
3. How do Partisans typically contribute to war?

V01 updated 3.25.18

TuVal

The Graph Choice Chart by the Maine Data Literacy Project¹, based on a work of participatory prototyping.
¹ created under a Creative Commons Attribution-ShareAlike 4.0 International license.

Do you want to compare the **variability of all data points** in each group to decide if any difference between the groups is meaningful?

1. Which of the two solar cars consistently goes the fastest?
2. Is there a meaningful difference in the heights of fertilized and unfertilized bean plants?

Are you comparing **single numbers** that summarize a group? (such as mean, median, or total...)

1. Was the total snowfall greater this winter than last winter?
2. Do cats and dogs have the same average body temperature?
3. How do the median incomes for the US and India compare?

Does it ask about how something changes through **linear TIME**?

1. Is the fuel efficiency of a car related to its weight?
2. Are smoking rates correlated with median income?
3. Given a fixed volume, how are temperature and pressure related?

1. Is sea level rising?
2. How did my weight change over the last 3 months?

Frequency Plot

FOR EACH GROUP MAKE A

MAKE BOTH

MAKE BOTH

MAKE BOTH

MAKE BOTH



- Chart is for when you know what you want to say
- Before that, do exploratory analysis

Few Chapter 5: Drawing Attention

- ① Memory
- ② Preattentive processing
 - form
 - color
 - spatial position
- ③ Applying to design
- ④ Gestalt principles of visual perception

Memory

Three types of memory

- 1 iconic memory

Memory

Three types of memory

- ① iconic memory
 - where preattentive processing works

Memory

Three types of memory

- ① iconic memory
 - where preattentive processing works
- ② working memory

Memory

Three types of memory

- ① iconic memory
 - where preattentive processing works
- ② working memory
 - what the viewer needs to keep in mind to understand your figure

Memory

Three types of memory

- ① iconic memory
 - where preattentive processing works
- ② working memory
 - what the viewer needs to keep in mind to understand your figure
- ③ long-term memory

Memory

Three types of memory

- ① iconic memory
 - where preattentive processing works
- ② working memory
 - what the viewer needs to keep in mind to understand your figure
- ③ long-term memory
 - where you want conclusion of your figure to sit

Memory

Three types of memory

- ① iconic memory
 - where preattentive processing works
- ② working memory
 - what the viewer needs to keep in mind to understand your figure
- ③ long-term memory
 - where you want conclusion of your figure to sit

Working memory

- We don't have much of it

Memory

Three types of memory

- ① iconic memory
 - where preattentive processing works
- ② working memory
 - what the viewer needs to keep in mind to understand your figure
- ③ long-term memory
 - where you want conclusion of your figure to sit

Working memory

- We don't have much of it
- People can remember 3 to 4 visual encodings for a chart
- Therefore, more than about 4 identifiers makes the graph difficult

Preattentive Processing

Why is this so important? Find the 5s.

48921652097520589

Preattentive Processing

Why is this so important? Find the 5s.

48921652097520589

And now find the 5s.

489216**5**2097**5**20**5**89

Preattentive Processing

Why is this so important? Find the 5s.

48921652097520589

And now find the 5s.

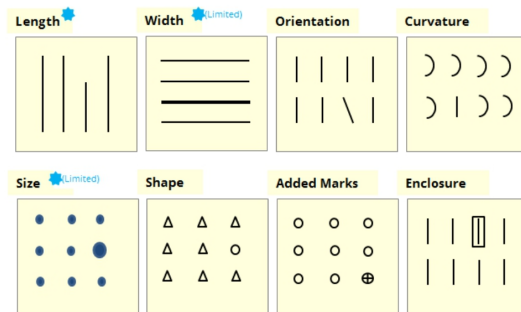
489216**5**2097**5**20**5**89

Use preattentive processing to point out what **you** think is important.

Preattentive Processing

Form
Color
Spatial Position

Form



Taken from <https://daydreamingnumbers.com/blog/preattentive-attributes-example/>

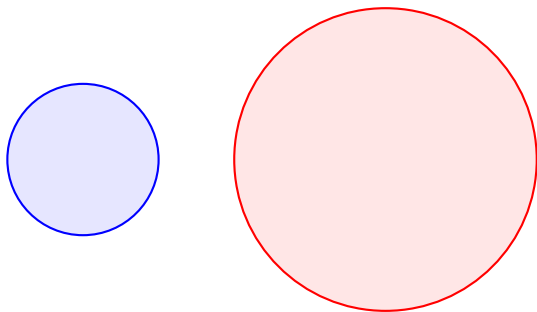
But Beware of 2-D Size

Why?

But Beware of 2-D Size

Why?

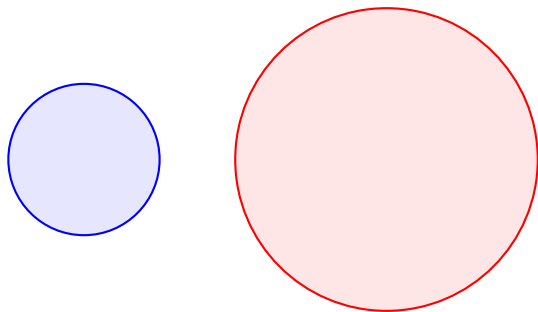
- People have a very hard time judging the relative size of 2-D objects
- How much bigger is the red circle than the blue circle?
- Changing both length and width is a 2-D change
- Avoid unless you have a specific reason to do this



But Beware of 2-D Size

Why?

- People have a very hard time judging the relative size of 2-D objects
- How much bigger is the red circle than the blue circle?
- Changing both length and width is a 2-D change
- Avoid unless you have a specific reason to do this

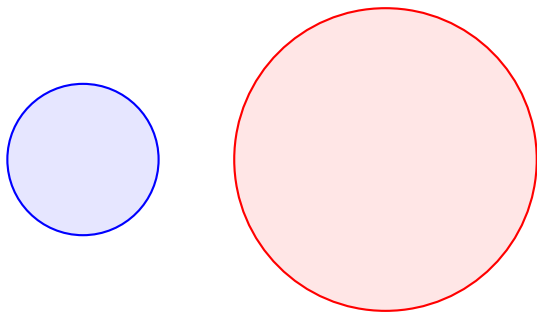


How much bigger is the small circle than the larger one?

But Beware of 2-D Size

Why?

- People have a very hard time judging the relative size of 2-D objects
- How much bigger is the red circle than the blue circle?
- Changing both length and width is a 2-D change
- Avoid unless you have a specific reason to do this



How much bigger is the small circle than the larger one? 4x

Color

① Hue

- What you think of as “color”
- Blue, Green, etc

② Saturation

- full color to white

③ Lightness

- or brightness, full color to dark

① Hue

- What you think of as “color”
- Blue, Green, etc

② Saturation

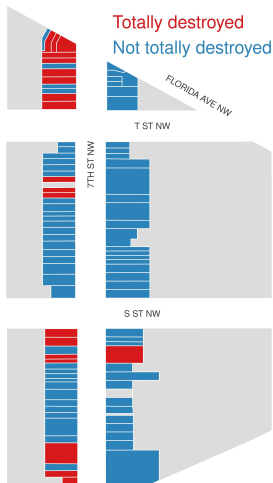
- full color to white

③ Lightness

- or brightness, full color to dark

Contrasting hues stand out. Intense colors stand out.

Using Color and Enclosure to Distinguish



- Quickly pick out two types
- Locate within larger block

Do We Perceive Them Quantitatively?

Type	Attribute
Form	Length
	Width
	Orientation
	Size
	Shape
	Enclosure
Color	Hue
	Intensity
Position	2-D Position

Do We Perceive Them Quantitatively?

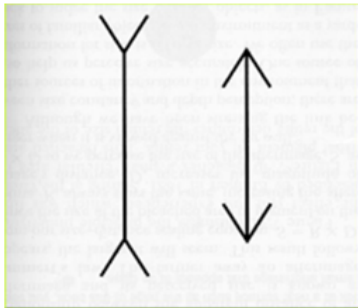
Type	Attribute	Quantitatively Perceived?
Form	Length	Yes
	Width	Yes, but limited
	Orientation	No
	Size	Yes, but limited
	Shape	No
	Enclosure	No
Color	Hue	No
	Intensity	Yes, but limited
Position	2-D Position	Yes

Do We Perceive Them Quantitatively?

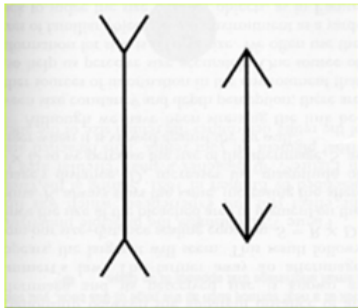
Type	Attribute	Quantitatively Perceived?
Form	Length	Yes
	Width	Yes, but limited
	Orientation	No
	Size	Yes, but limited
	Shape	No
	Enclosure	No
Color	Hue	No
	Intensity	Yes, but limited
Position	2-D Position	Yes

We rely heavily on things that we perceive entirely quantitatively

Context Matters

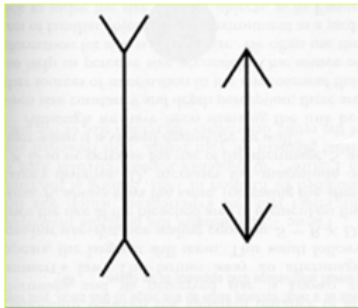


Context Matters

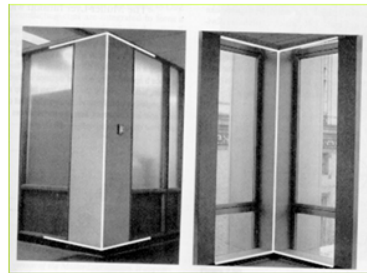


Why?

Context Matters



Why?



Thanks to [UC Irvine's Majumder](#).

Why Context Matters



Why Context Matters

Preattentive processing relies on difference.



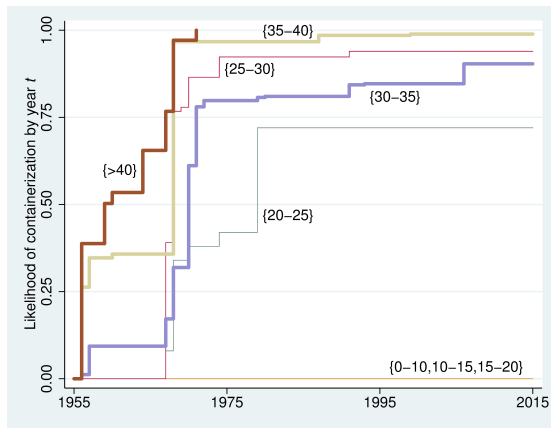
Why Context Matters

Preattentive processing relies on difference.



Too many differences obscures any one difference.

Calling Attention



Which principle do I use here?

Gestalt Principles of Visual Perception

- Proximity
- Similarity
- Enclosure
- Closure
- Continuity

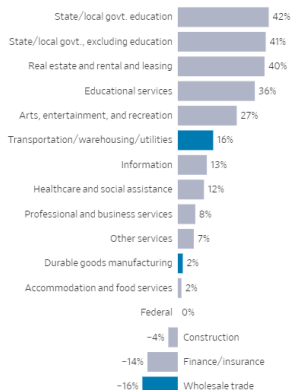
These all generate meaning, whether you intend it or not!

Applying These Principles

- WSJ graph on job openings
- My regression results
 - first a set of slides that do a so-so job
 - second a set of slides that do a better (but improvable) job

Similarity and Continuity

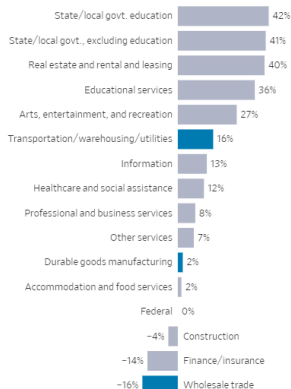
Change, 1/2018 to 11/2019



Job openings in blue-collar industries saw some of the weakest growth before the pandemic.



Change, 1/2018 to 11/2019

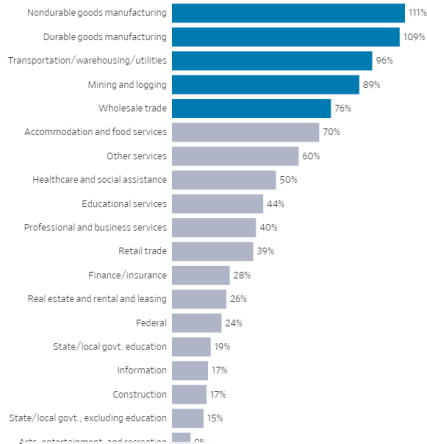


Job openings in blue-collar industries saw some of the weakest growth before the pandemic.



Similarity and Continuity

Change, 1/2020 to 11/2021

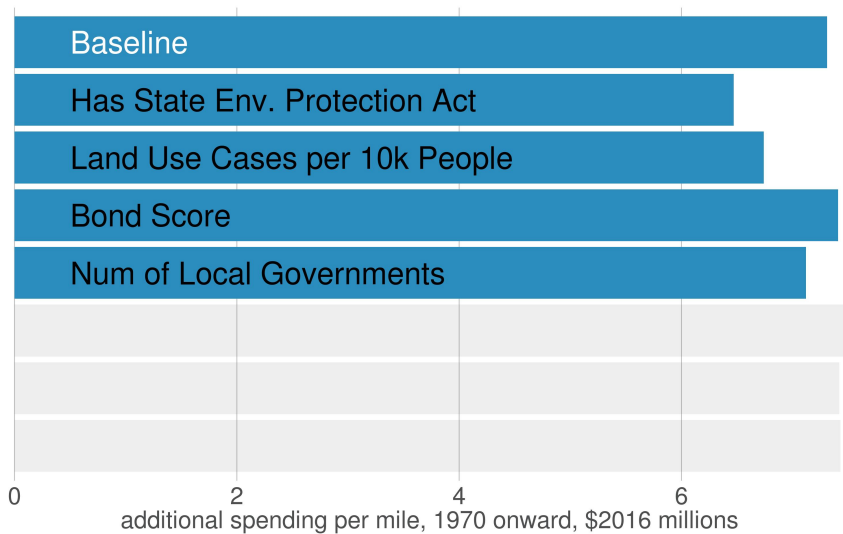


Now, blue-collar job openings are logging the biggest gains.

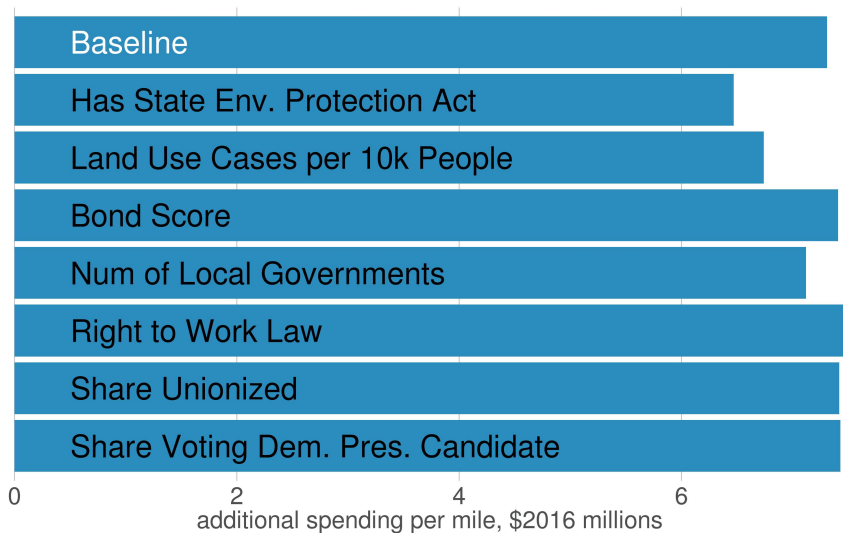
Baseline Increase of \$7.3 Million per Mile



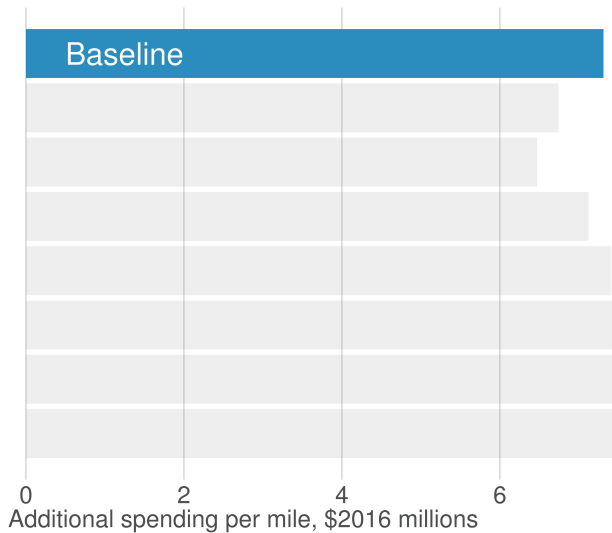
Measures of Gov't Quality Unrelated to Spending Increase



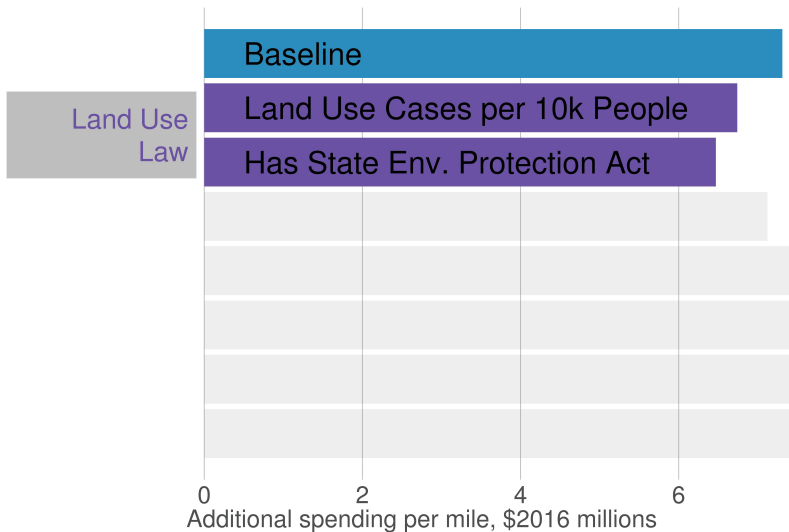
Labor Strength Unrelated to Spending Increase



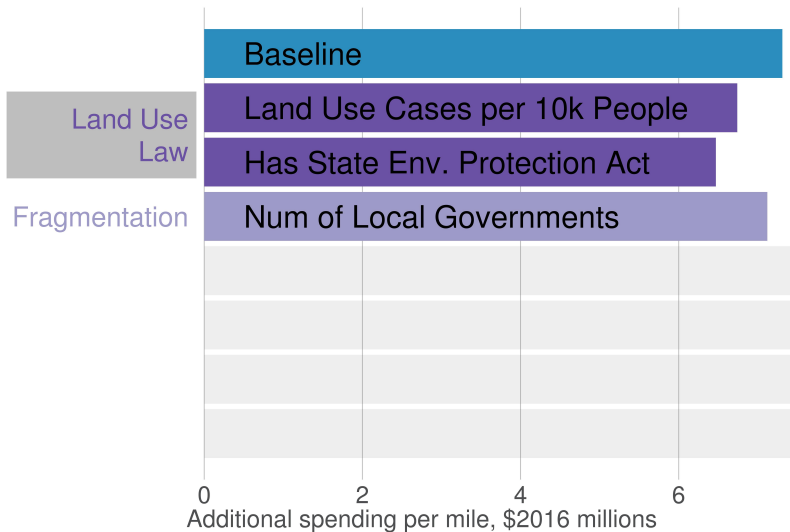
Using the Principles of Proximity and Similarity



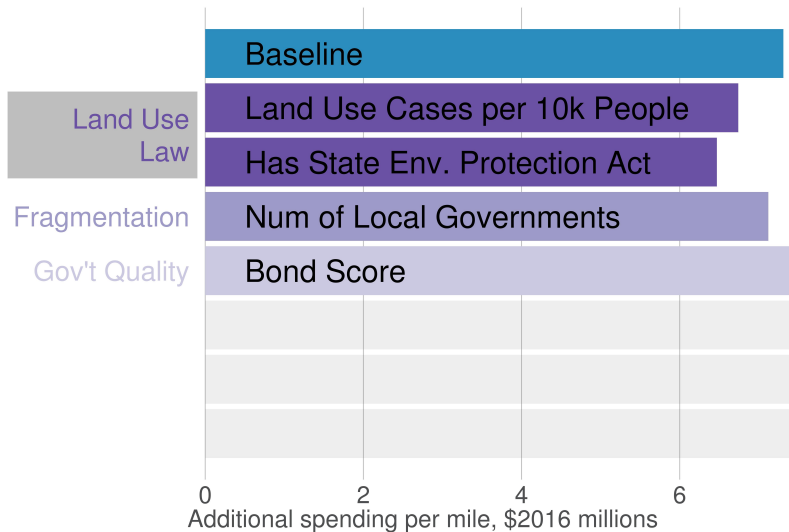
Using the Principles of Proximity and Similarity



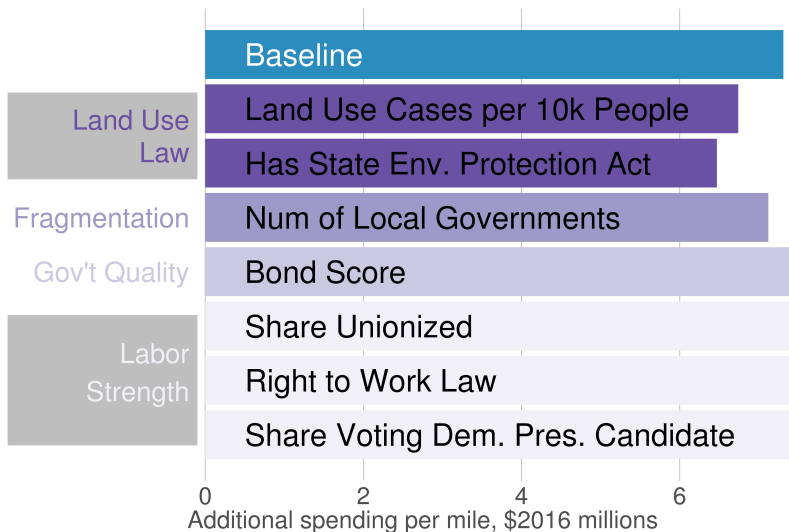
Using the Principles of Proximity and Similarity



Using the Principles of Proximity and Similarity



Using the Principles of Proximity and Similarity



2. R: Merging

Why Do You Need to Know How to Merge?

If you want to say something about data in more than one dataset, you must merge!

What is a Merge?

You want to put together

Dataset A – One obs/ID

ID	Income
A	50
B	100

Dataset B – One obs/ID

ID	Pool
A	TRUE
B	FALSE

How many rows should the merged dataset have?

What is a Merge?

You want to put together

Dataset A – One obs/ID

ID	Income
A	50
B	100

Dataset B – One obs/ID

ID	Pool
A	TRUE
B	FALSE

How many rows should the merged dataset have?

Into

ID	Income	Pool
A	50	TRUE
B	100	FALSE

This is a 1 to 1 merge.

What is a Many to 1 Merge?

You want to put together

Dataset A – One obs/ID

ID	Income
A	50
B	100

Dataset B – many obs/ID

ID	Pool	Year
A	TRUE	2020
B	FALSE	2020
A	TRUE	2021
B	TRUE	2021

What is a Many to 1 Merge?

You want to put together

Dataset A – One obs/ID

ID	Income
A	50
B	100

Dataset B – many obs/ID

ID	Pool	Year
A	TRUE	2020
B	FALSE	2020
A	TRUE	2021
B	TRUE	2021

How many rows and columns should it have?

What is a Many to 1 Merge?

You want to put together

Dataset A – One obs/ID

ID	Income
A	50
B	100

Dataset B – many obs/ID

ID	Pool	Year
A	TRUE	2020
B	FALSE	2020
A	TRUE	2021
B	TRUE	2021

ID	Pool	Year	Income
A	TRUE	2020	50
B	FALSE	2020	100
A	TRUE	2021	50
B	TRUE	2021	100

How many rows and columns should it have?

What is a Many to Many Merge?

A mess!

What is a Many to Many Merge?

A mess!

Dataset A

ID	Income
A	50
A	60
B	100

Dataset B

ID	Pool	Year
A	TRUE	2020
B	FALSE	2020
A	TRUE	2021
B	TRUE	2021

What is a Many to Many Merge?

A mess!

Dataset A

ID	Income
A	50
A	60
B	100

Dataset B

ID	Pool	Year
A	TRUE	2020
B	FALSE	2020
A	TRUE	2021
B	TRUE	2021

There is no logical path to merge A and B.

What is a Many to Many Merge?

A mess!

Dataset A

ID	Income
A	50
A	60
B	100

Dataset B

ID	Pool	Year
A	TRUE	2020
B	FALSE	2020
A	TRUE	2021
B	TRUE	2021

There is no logical path to merge A and B. Very probably something is wrong with A.

Merging in R

Trying a Few R Features

- Time permitting, work through a merge example
- Work on tutorial
- Make a .R script for whole tutorial
- Plus questions at end
- I will be here till 5:20 – please stay and ask questions

Next Lecture

- Be ready for quiz on Tutorial 2
- Read Few: Chapter 6, pages 87-95; Chapter 9; and Chapter 10, pages 210-217 (on bars)
- Read Chang Chapter 3
- Read two linked articles from *WSJ* – let me know if you need a pdf
- Turn in policy brief proposal