

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

January 30, 2023

# Course Administration

1. Any trouble submitting tutorials?  
questions?
2. Questions/issues with readings?
3. Make sure you're signed up for Piazza  
– email me if you are not

# Course Administration

1. Any trouble submitting tutorials? questions?
2. Questions/issues with readings?
3. Make sure you're signed up for Piazza – email me if you are not

4. Be sure to check online listing for good/bad/ugly: linked at bottom of lectures page

- I moved a few of you around to even up finders and commenters
- If date is not ok, try to switch with a classmate
- If you didn't sign up, sign up and let me know

# Course Administration

1. Any trouble submitting tutorials? questions?
2. Questions/issues with readings?
3. Make sure you're signed up for Piazza – email me if you are not
4. Be sure to check online listing for good/bad/ugly: linked at bottom of lectures page
  - I moved a few of you around to even up finders and commenters
  - If date is not ok, try to switch with a classmate
  - If you didn't sign up, sign up and let me know
5. One-page proposal is due next week
6. Anything else?

## Next Week's Good Bad and Ugly

Finders, post link Wed. by noon.

	Finder	Commenter
1	Henry	Lancy

Email me ASAP if you're not on the google sheet. Link at the bottom of the lectures tab.

# Few: Visual Perception and Graphical Communication

# When Should You Use Tables vs. Graphs?

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





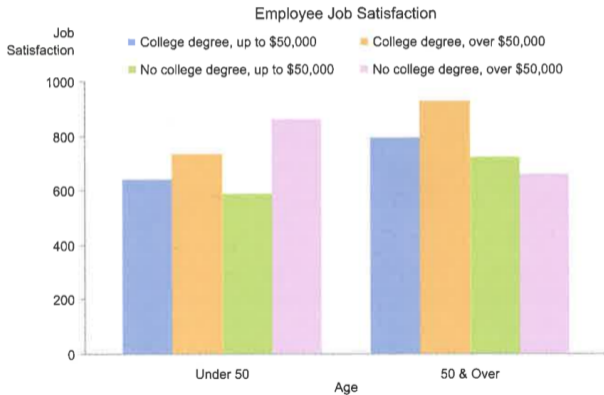
## Starting with the Table

### Job Satisfaction By Income, Education, and Age

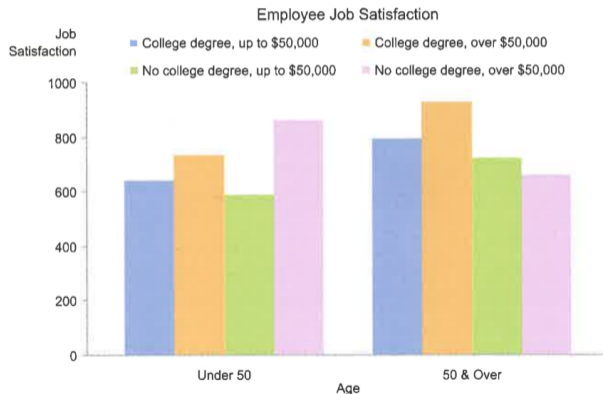
Income	College Degrees		No College Degrees	
	Under 50	50 & over	Under 50	50 & over
Up to \$50,000	643	793	590	724
Over \$50,000	735	928	863	662

Few, Chapter 3, Figure 3.13

# Version One of a Set of Numbers



## Version One of a Set of Numbers



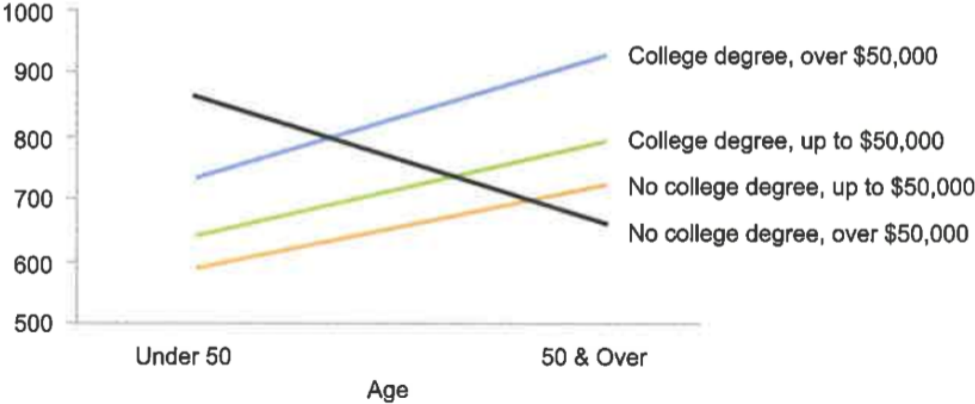
What do you think the point of this picture is?

Few, Chapter 3, Figure 3.15

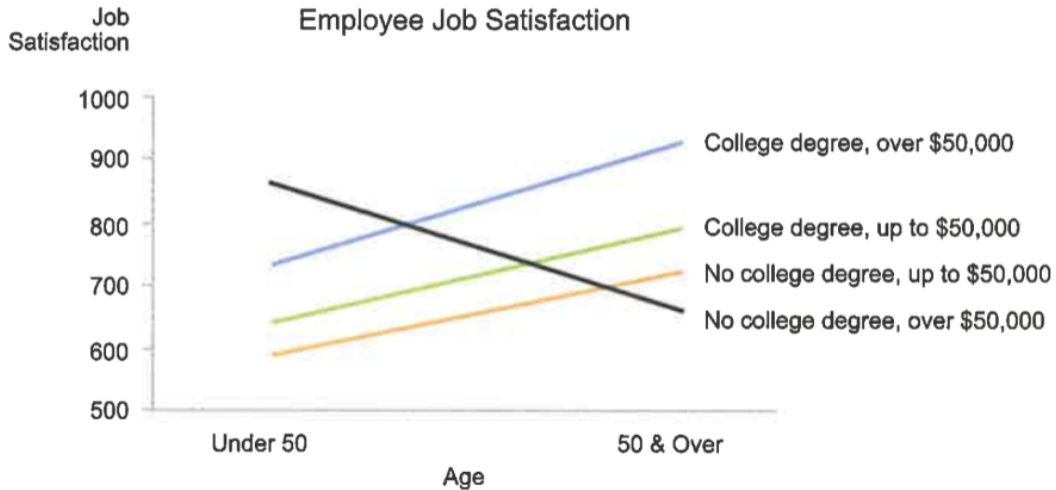
## Version Two of the Same Set of Numbers

Job Satisfaction

### Employee Job Satisfaction



## Version Two of the Same Set of Numbers



And the point of this picture?

# 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 tides rise to the same height?
2. How variable are wind speeds in Denmark?
3. What is the range and distribution of incomes in Sudan?

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 circuit 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 2015?
2. What proportion of global electricity production comes from wind?
3. How do Parisians typically commute to work?

VO.1 updated 5.29.16

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 farthest?
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**?

N

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?

T

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

Frequency Plot

MAKE EITHER

FOR EACH GROUP MAKE A

Histogram



Box Plots



Dot Plot



Bar Graph



Scatter Plot



Line Graph



Pie Chart



Stacked Bar Chart



## Few Chapter 5: Drawing Attention

1. working memory
2. preattentive processing
  - form
  - color
  - spatial position
3. applying to design
4. gestalt principles of visual perception

# Working Memory

We don't have much of it



# 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 colors as identification are distracting
- good visuals can stick in long-term memory

# 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.

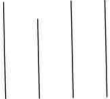
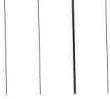
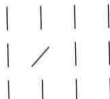
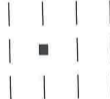

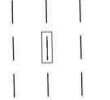
489216**5**2097**5**20**5**89

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

# Preattentive Processing

Form  
Color  
Spatial Position

# Form

Length	Width
	
Orientation	Shape
	
Size	Enclosure
	

# 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
- Changing both length and width is a 2-D change
- Avoid unless you have a specific reason to do this – maybe you're drawing building sizes





## But Beware of 2-D Size

### Why?

- People have a very hard time judging the relative size of 2-D objects
- Changing both length and width is a 2-D change
- Avoid unless you have a specific reason to do this – maybe you're drawing building sizes

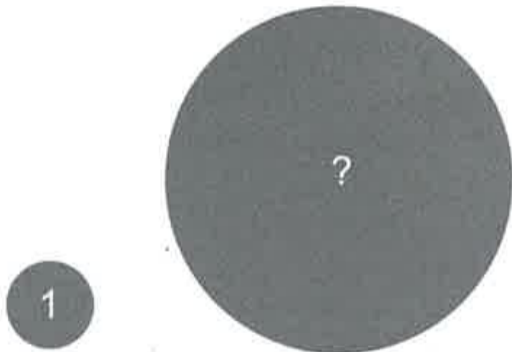


How much bigger is the small circle than the larger one?

## But Beware of 2-D Size

### Why?

- People have a very hard time judging the relative size of 2-D objects
- Changing both length and width is a 2-D change
- Avoid unless you have a specific reason to do this – maybe you're drawing building sizes



How much bigger is the small circle than the larger one? 16x

# 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

# Color

## 1. Hue

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

## 2. Saturation

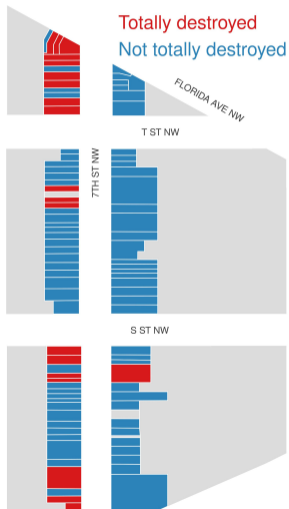
- full color to white

## 3. Lightness

- or brightness, full color to dark

Contrasting hues stand out. Intense colors stand out.

## Using Color and Enclosure to Distinguish



- Quickly pick out two types
- Locate within larger block

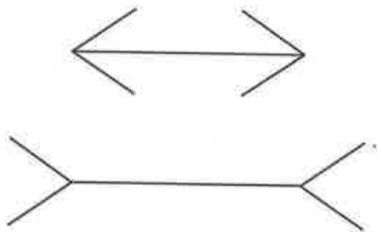
## Do We Perceive Them Quantitatively?

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

# Context Matters

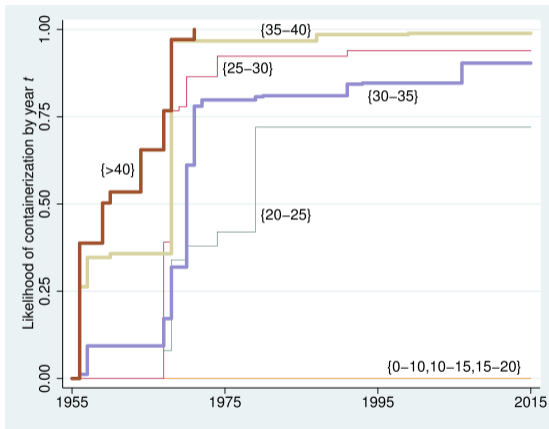




# Context Matters



# Calling Attention



Which principle do I use here?

# Gestalt Principles of Visual Perception

- Proximity
- Similarity
- Enclosure
- Closure
- Continuity

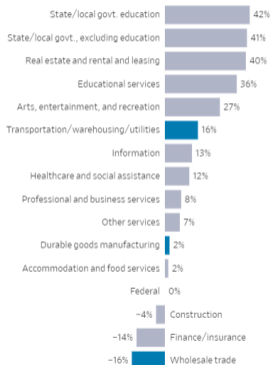
These all generate meaning, whether you intend it or not!

## 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

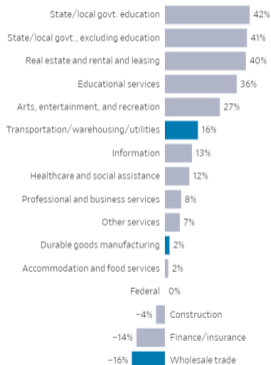


Job openings in blue-collar industries saw some of the weakest growth before the pandemic.



# Similarity and Continuity

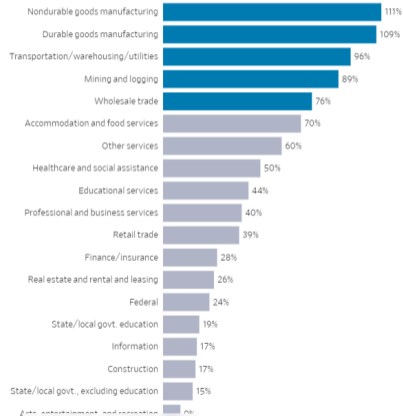
Change, 1/2018 to 11/2019



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

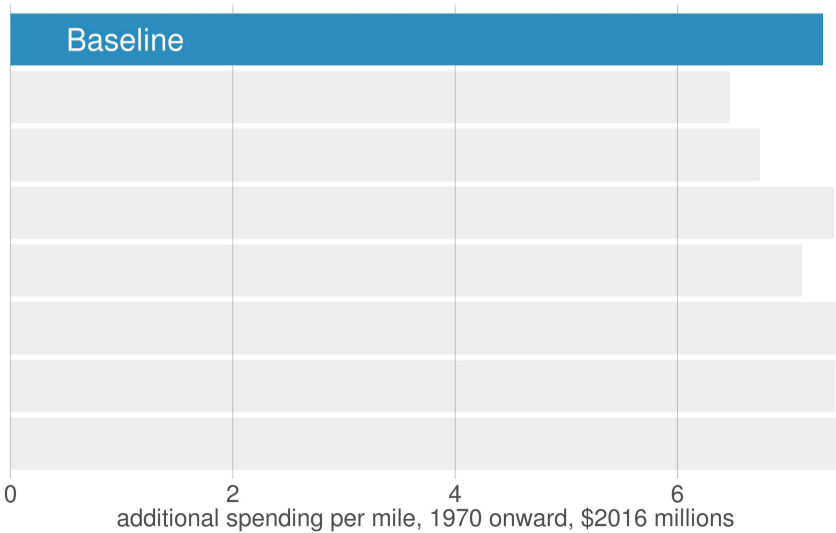


Change, 1/2020 to 11/2021

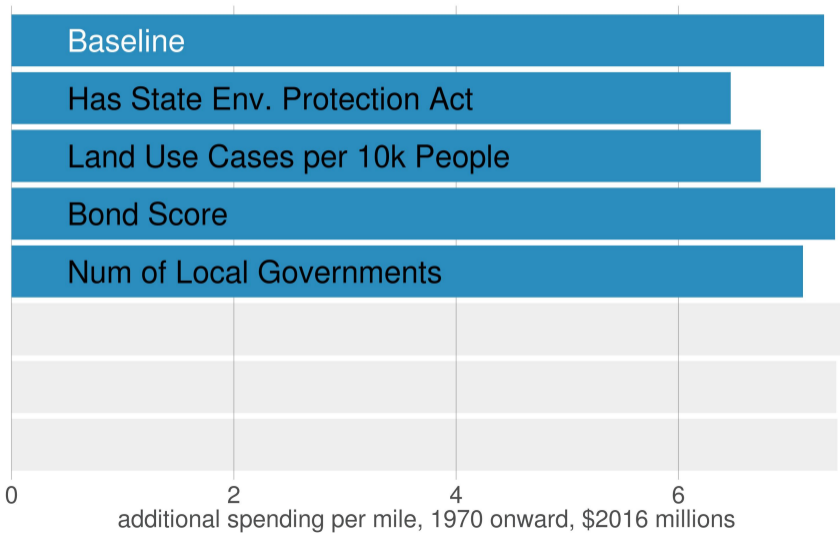


Now, blue-collar job openings are logging the biggest gains.

# Baseline Increase of \$7.3 Million per Mile

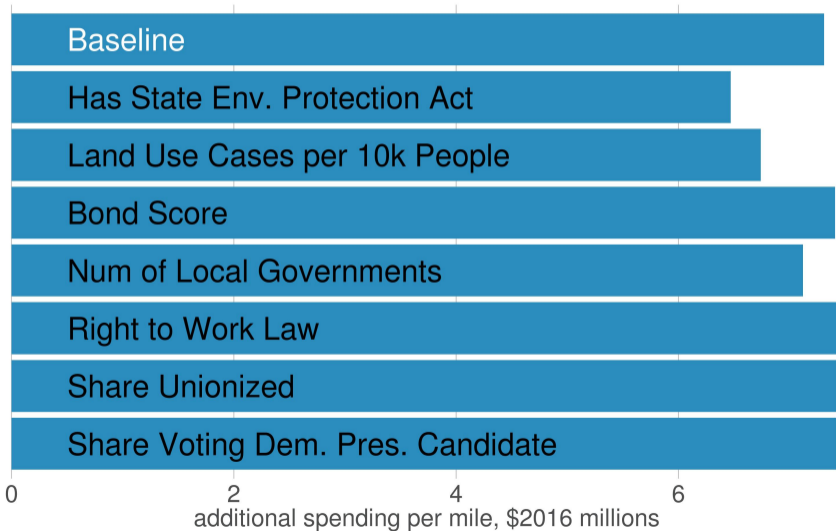


# Measures of Government Quality Unrelated to Spending Increase

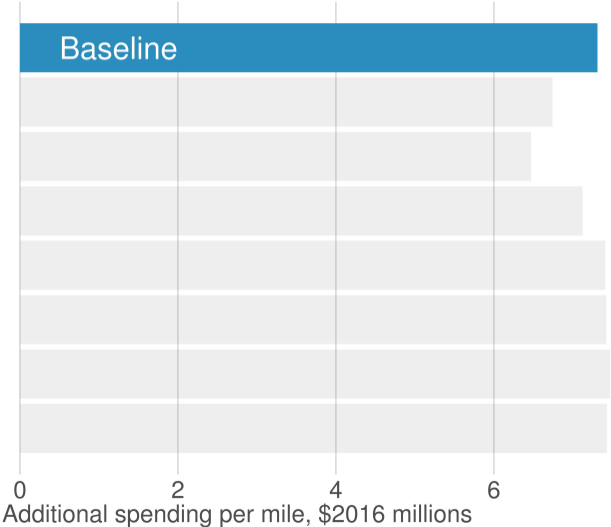




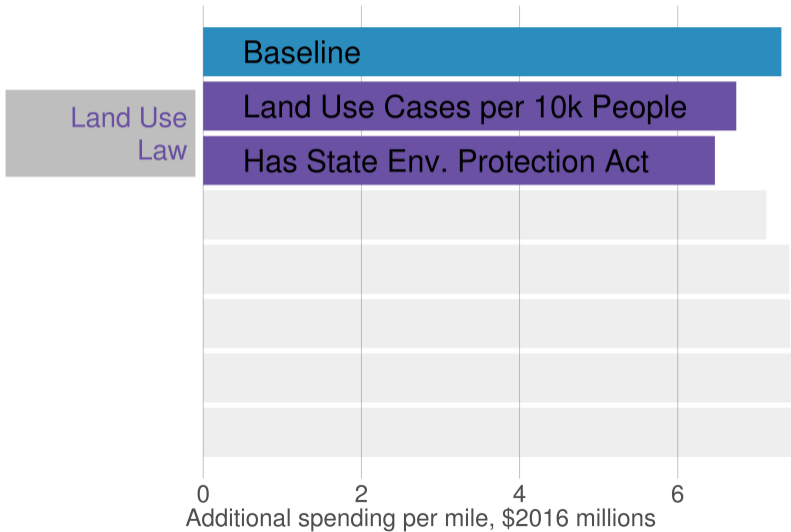
# Measures of Labor Strength Unrelated to Spending Increase



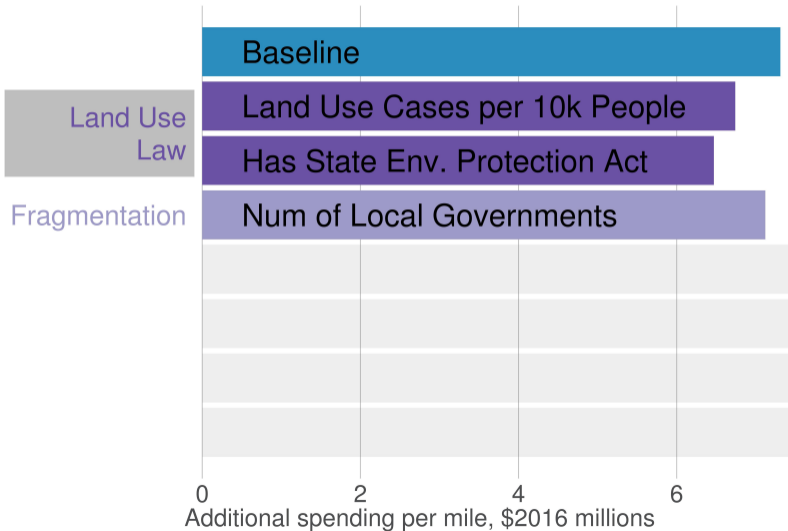
# Using the Principles of Proximity and Similarity



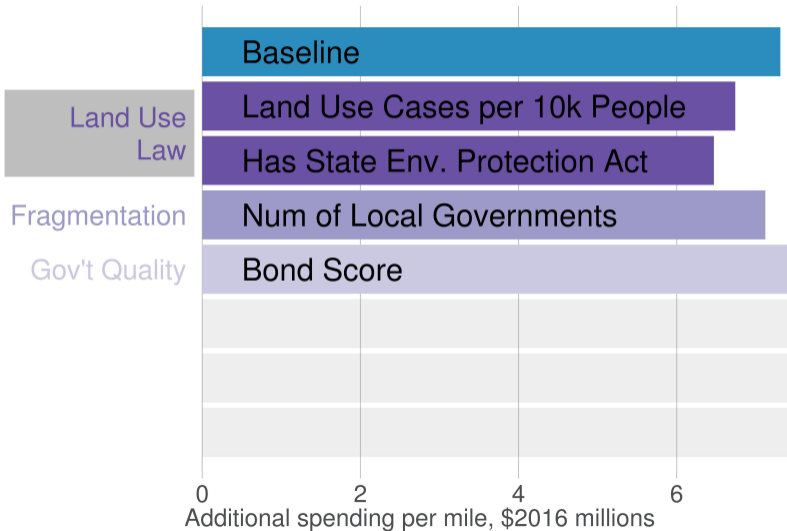
# Using the Principles of Proximity and Similarity



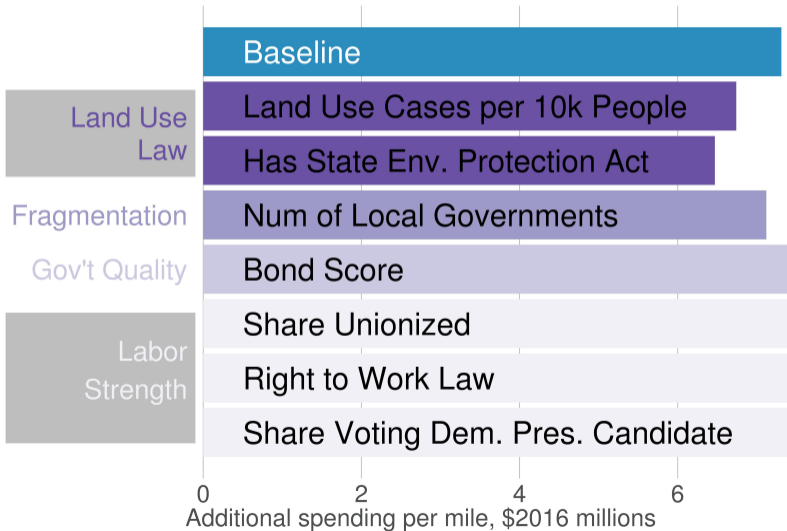
# Using the Principles of Proximity and Similarity



# Using the Principles of Proximity and Similarity



# Using the Principles of Proximity and Similarity





# Why Do You Need to Know How to Merge?

If you want to say anything about something in more than one dataset.



## 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

## 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

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

How many rows 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

How many rows should it have?

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



## 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. Probably something is wrong with A.

# Merging in R



## Next Lecture

- Turn in PS 2
- Read Few Chapter 9 and Chapter 10, pages 210-217 (on bars)
- Read Chang, Chapter 3
- Read two linked examples from *WSJ*
- Turn in policy brief proposal