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

January 27, 2020

Overview

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
5. Addition to syllabus: *WSJ*'s Luis Melgar on March 30
6. One-page proposal is due next week
7. Anything else?

Next Week's Good Bad and Ugly

Finders, send link Wed. by noon.

	Finder	Commenter
1	Lindsay R	Tereese S
2	Kim W	Danielle C
3	Anna W	David N

Email me ASAP if you're not on the google sheet.

Tufte's Three Routes to a Bad Graph

1. Data are bad
2. Graphics are rotten
3. Graphics are irrelevant

Tufte's Three Routes to a Bad Graph

1. **Data are bad**
2. Graphics are rotten
3. Graphics are irrelevant

Inappropriate Data to Make This Point

Inappropriate Data to Make This Point

Heart Disease Strikes Back Across the U.S., Even in Healthy Places

Middle-aged people are increasingly dying from heart disease in cities across the country—including exercise-mad Colorado

Metro areas with the largest rate increases

1. Lexington, Ky.	27.9%	6. Beaumont, Texas	24.1
2. Atlantic City, N.J.	25.7	7. Fort Wayne, Ind.	23.9
3. Corpus Christi, Texas	25.7	**8. Greeley, Colo.**	**23.5**
4. Lincoln, Neb.	25.1	**9. Colo. Springs, Colo.**	**23.3**
5. Fort Collins, Colo.	**24.4**	10. Kennewick, Wash.	22.5

Source is Jan. 14, 2020 article here

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

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

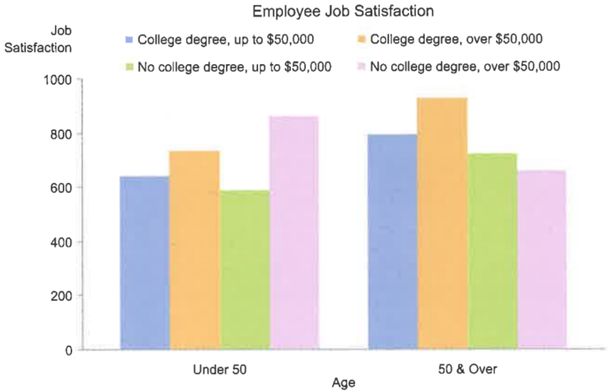
Starting with the Table

Job Satisfaction By Income, Education, and Age

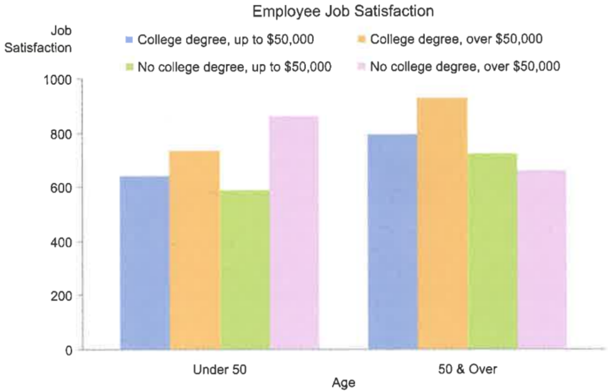
	College Degrees		No College Degrees	
Income	Under 50	50 & over	Under 50	50 & over
Up to $50,000	643	793	590	724
Over $50,000	735	928	863	662

Few, Chapter 3, Figure 3.13

Version One of a Set of Numbers



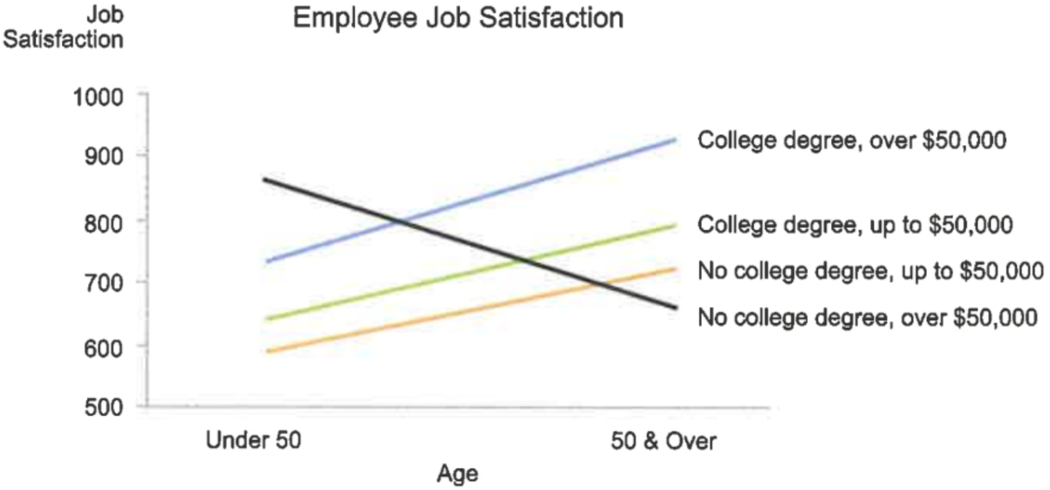
Version One of a Set of Numbers



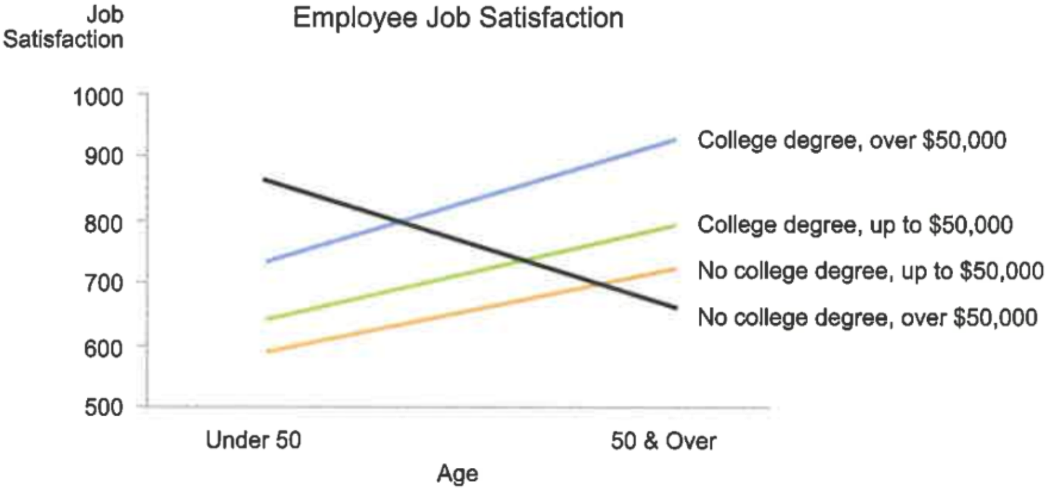
What do you think the point of this picture is?

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

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.

4892165209752058 9

Preattentive Processing

Why is this so important? Find the 5s.

4892165209752058 9

And now find the 5s.

4892165209752058 9

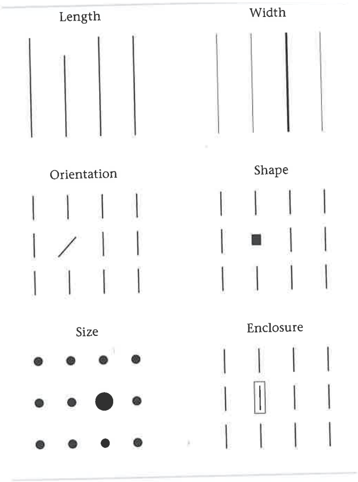
Preattentive Processing

Form

Color

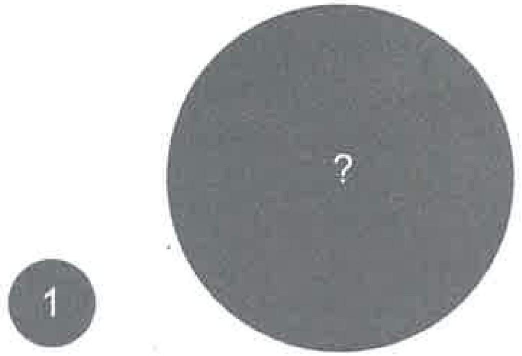
Spatial Position

Form



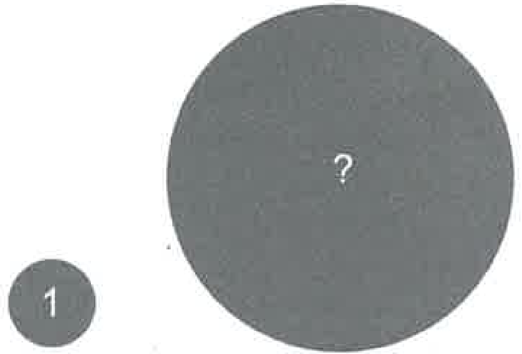
But Beware of 2-D Size

- 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

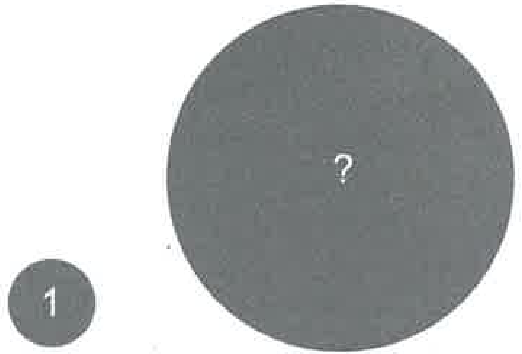
- People have a very hard time judging the relative size of 2-D objects
- Changing both length and width is a 2-D change
- Avoid unless you have a specific reason to do this – maybe you're drawing building sizes



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

But Beware of 2-D Size

- 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. Intensity
   - make it less intense: add a little gray

Color

1. Hue
   - What you think of as “color”
   - Blue, Green, etc
2. Intensity
   - make it less intense: add a little gray

Contrasting hues stand out. Intense colors stand out.

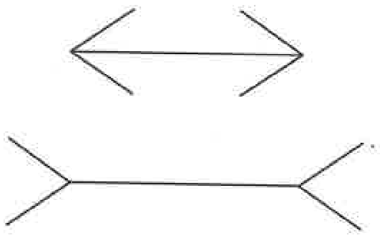
Do We Perceive Them Quantitatively?

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

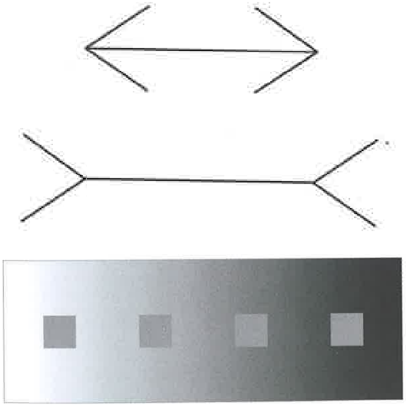
Do We Perceive Them Quantitatively?

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

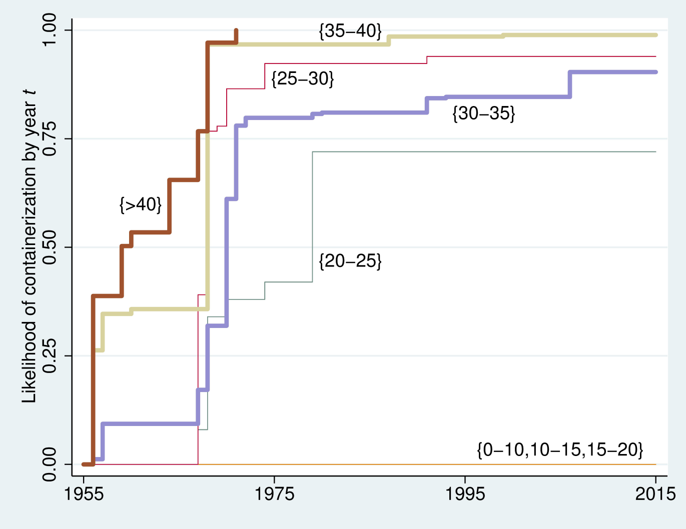
Context Matters



Context Matters



Calling Attention



Gestalt Principles of Visual Perception

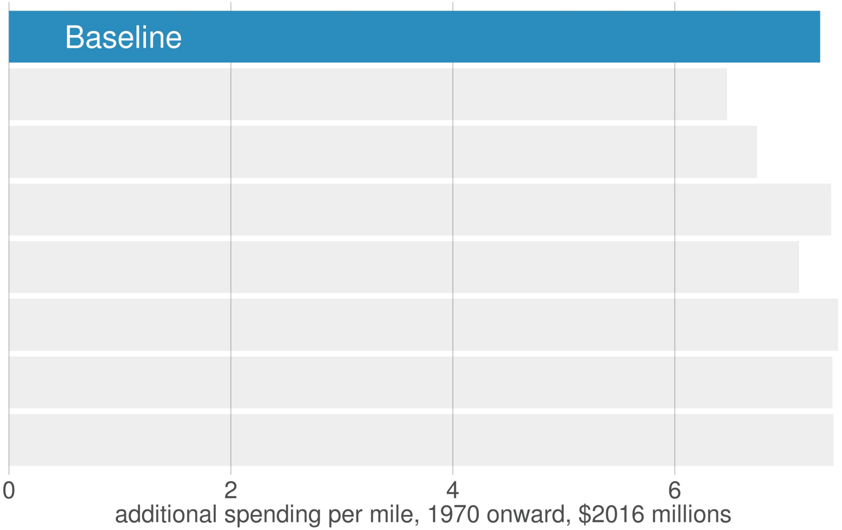
- Proximity
- Similarity
- Enclosure
- Closure
- Continuity

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

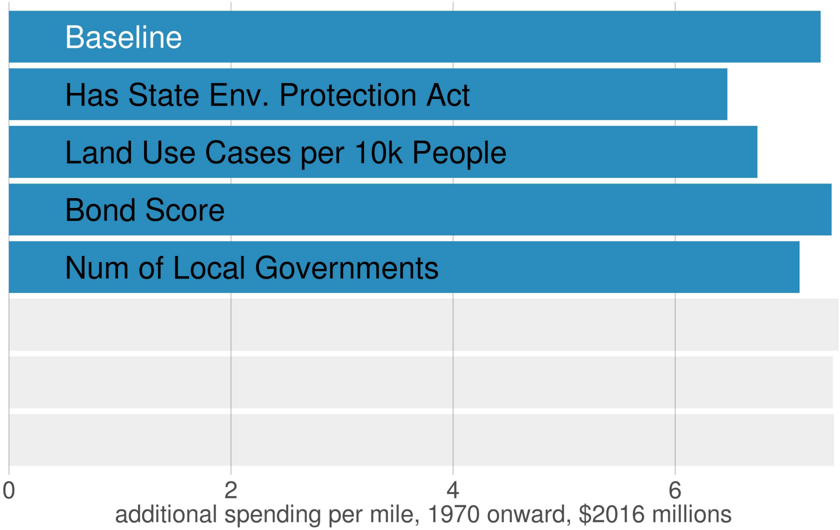
Applying These Principles

- first a set of slides that do a so-so job
- second a set of slides that do a better (but improvable) job

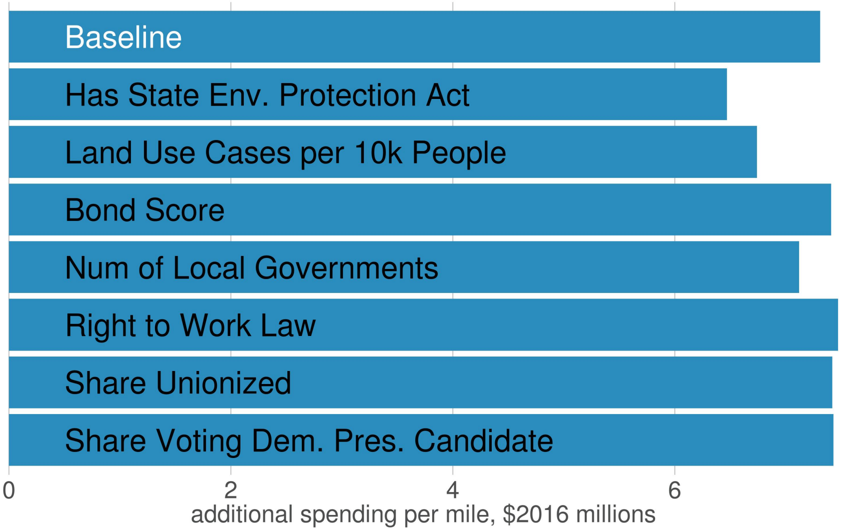
Baseline Increase of $7.3 Million per Mile



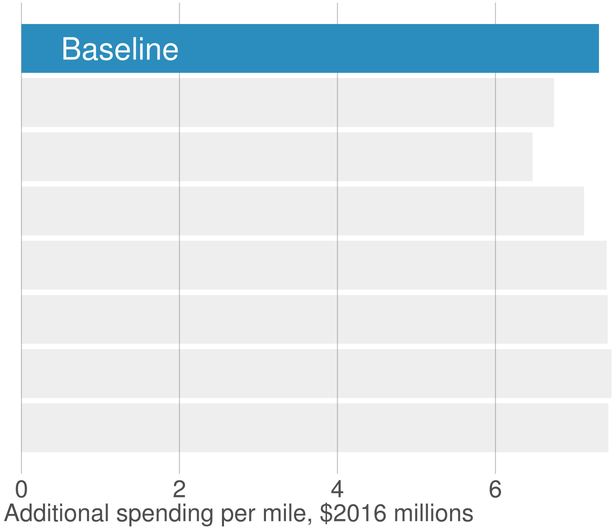
Measures of Government Quality Unrelated to Spending Increase



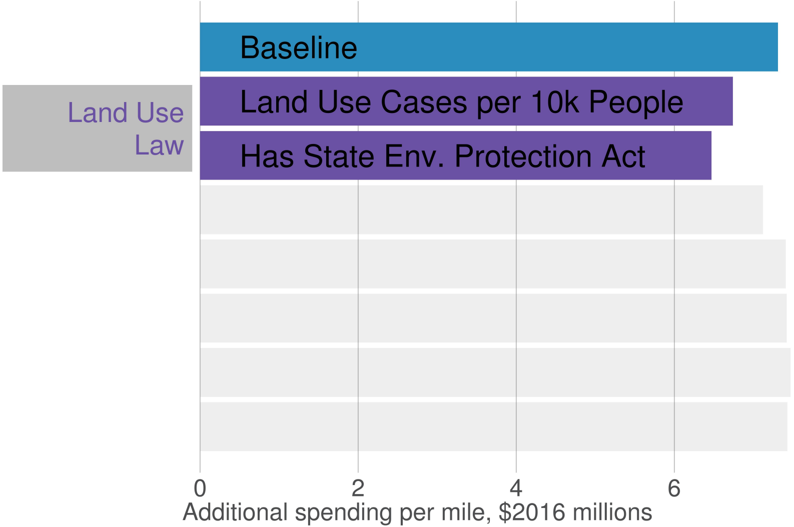
Measures of Labor Strength Unrelated to Spending Increase



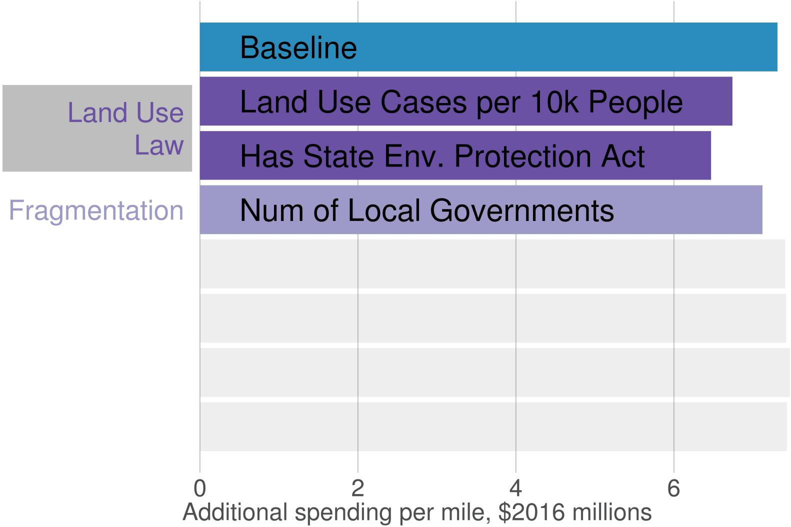
Using the Principles of Proximity and Similarity



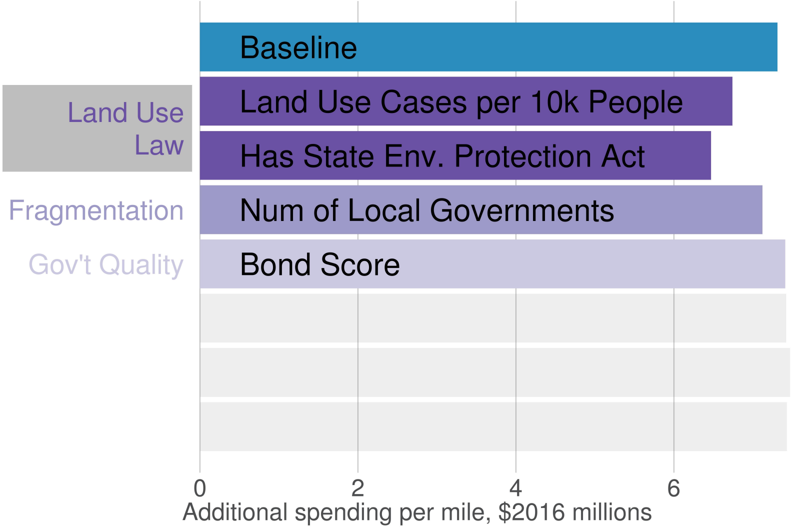
Using the Principles of Proximity and Similarity



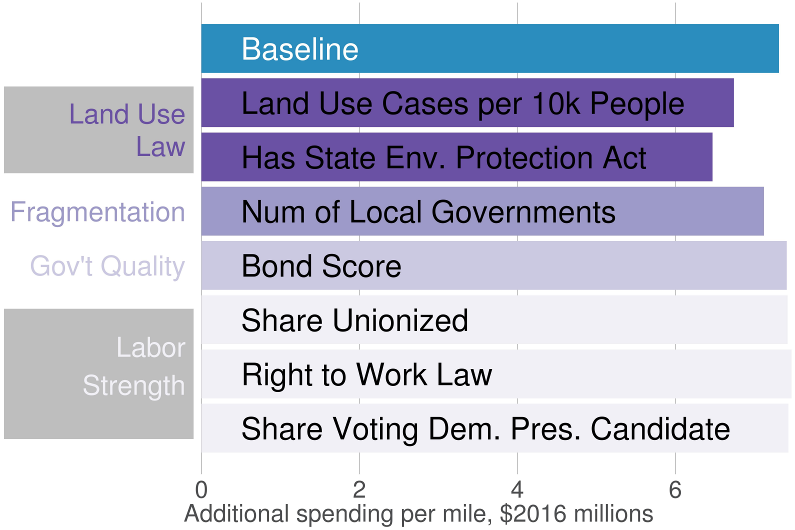
Using the Principles of Proximity and Similarity



Using the Principles of Proximity and Similarity



Using the Principles of Proximity and Similarity



R

Today

A. What is Merging?

B. How to Merge 1:1

C. How to Merge Many to 1

D. Cautions with merging

A. Merging

- suppose you have information in more than one dataframe
- you want to combine these pieces of information
- this is an **enormous** advantage of statistical software

Examples of When You Need to Merge

Ex. 1:

- you have a dataset on crimes, with addresses
- you want to add the neighborhood median income
- $\rightarrow$ merge by neighborhood id!

Examples of When You Need to Merge

Ex. 1:

- you have a dataset on crimes, with addresses
- you want to add the neighborhood median income
- $\rightarrow$ merge by neighborhood id!

Ex. 2:

- you have a dataset of student performance
- you want to add information on teacher
- $\rightarrow$ merge by teacher id!

Merging Command Overview

```
merge(x = data.frame.1,
      y = data.frame.2,
      by = "varname",
      all = TRUE)
```

Merging Command Overview

```
merge(x = data.frame.1,
      y = data.frame.2,
      by = "varname",
      all = TRUE)
```

Now a very simple example

Sample dataframe 1

```
df1 <- data.frame(class = c(1,2,3),
                  subject = c("basics","basics","graphs"))
df1
```

```
##   class subject
## 1     1  basics
## 2     2  basics
## 3     3  graphs
```

Sample dataframe 2

```r
df2 <- data.frame(class = c(1,2,3),
                  attendance = c(33,45,26))
df2
```

```
##   class attendance
## 1     1         33
## 2     2         45
## 3     3         26
```

B. Merge 1:1

```
df3 <- merge(x = df1, y = df2, by = "class", all = TRUE)
```

How many rows should this have?

B. Merge 1:1

```r
df3 <- merge(x = df1, y = df2, by = "class", all = TRUE)
```

How many rows should this have?

```r
df3
```

```
##   class subject attendance
## 1     1  basics         33
## 2     2  basics         45
## 3     3  graphs         26
```

C. Merge m:1

- this is a merge that has unique values in one dataset
- and repeat values in another
- for us, repeat values are in `subject`

Dataset to merge in

```
df4 <- data.frame(subject = c("basics","graphs"),
                  difficulty = c("easy","hard"))
df4
```

```
##   subject difficulty
## 1  basics       easy
## 2  graphs       hard
```

Merging in

```
df5 <- merge(x = df3, y = df4, by = "subject", all = TRUE)
```

How many rows should this have?

Merging in

```
df5 <- merge(x = df3, y = df4, by = "subject", all = TRUE)
```

How many rows should this have?

```
df5
```

```
##   subject class attendance difficulty
## 1  basics     1         33       easy
## 2  basics     2         45       easy
## 3  graphs     3         26       hard
```

D. Problems with Merging

- you want to merge 1:1 but one dataframe has repeat values

D. Problems with Merging

- you want to merge 1:1 but one dataframe has repeat values
- you want to merge 1:1 but the merge doesn't work as expected (see tutorial)
- bad merges cause big problems

Try Today's Tutorial

- Make a .R script for whole tutorial
- Plus questions at end
- Go forth!

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