



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. Reading quiz



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. Reading quiz
5. 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 you didn't sign up, I signed you up
 - If date is not ok, try to switch with a
classmate



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. Reading quiz
5. 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 you didn't sign up, I signed you up
 - If date is not ok, try to switch with a
classmate
5. Additions to syllabus
 - Lecture 8, March 22: Former
students doing good: McCall Pitcher
and Kimberly Wilson
 - Lecture 9, March 29: *WashPo's*
Kate Rabinowitz 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	Sarah H.	Gabriel C.
2	Arjun B.	Anthony C.

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



When Should You Use Tables vs. Graphs?

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

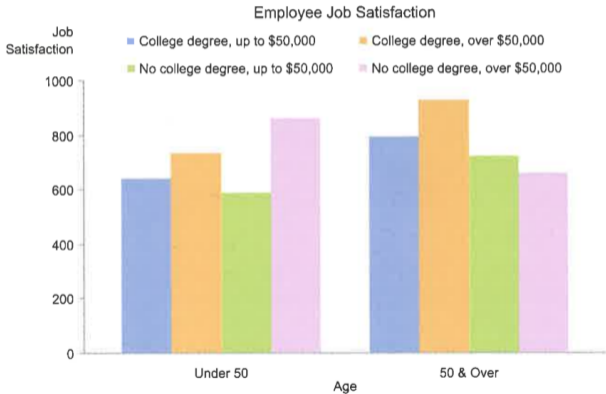
Starting with the Table

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

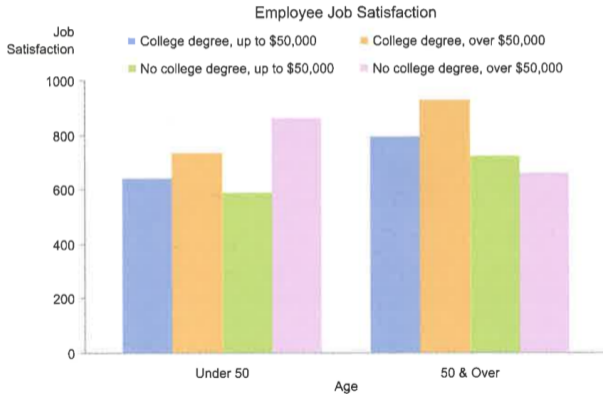
Few, Chapter 3, Figure 3.13

Version One of a Set of Numbers





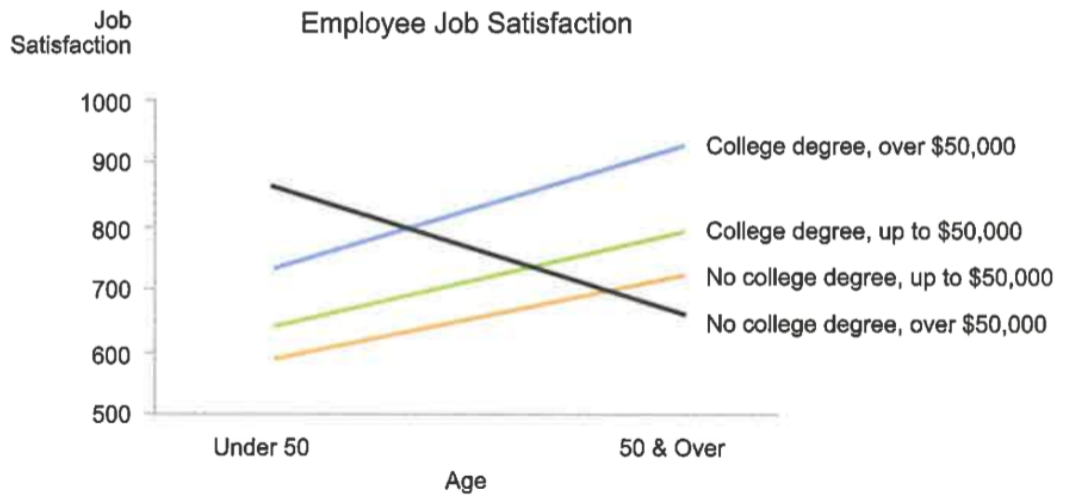
Version One of a Set of Numbers



What do you think the point of this picture is?

Few, Chapter 3, Figure 3.15

Version Two of the Same Set of Numbers





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

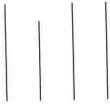
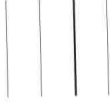
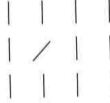
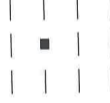

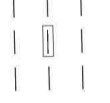
And now find the 5s.

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

Preattentive Processing

Form
Color
Spatial Position

Form

<p>Length</p> 	<p>Width</p> 
<p>Orientation</p> 	<p>Shape</p> 
<p>Size</p> 	<p>Enclosure</p> 

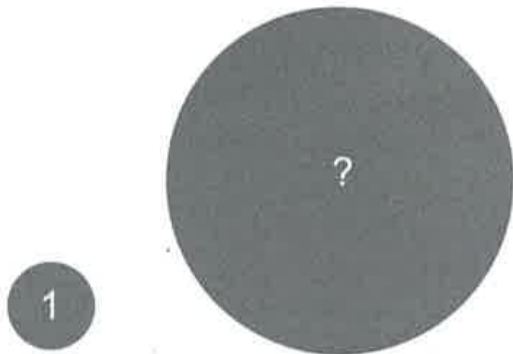
But Beware of 2-D Size

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

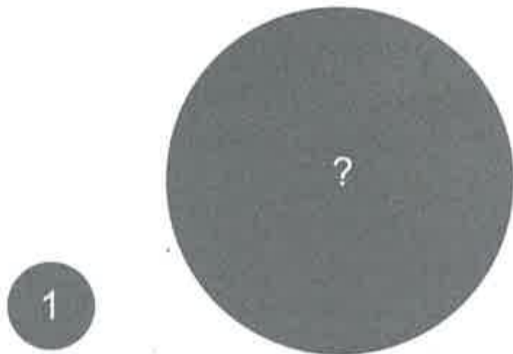


How much bigger is the small circle than the larger one?

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. 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?

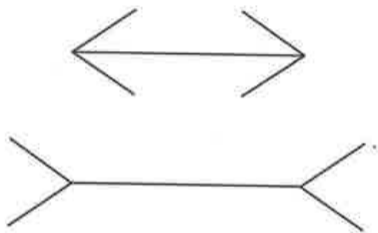
Type	Attribute
Form	Length
	Width
	Orientation
	Size
	Shape
	Enclosure
Color	Hue
	Intensity
Position	2-D Position



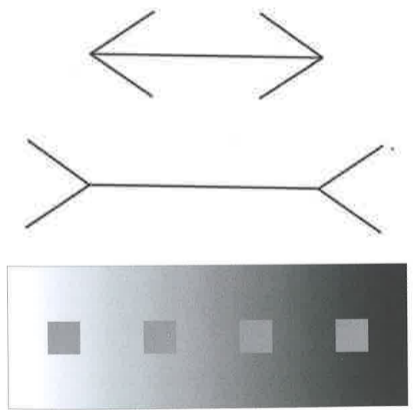
Do We Perceive Them Quantitatively?

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

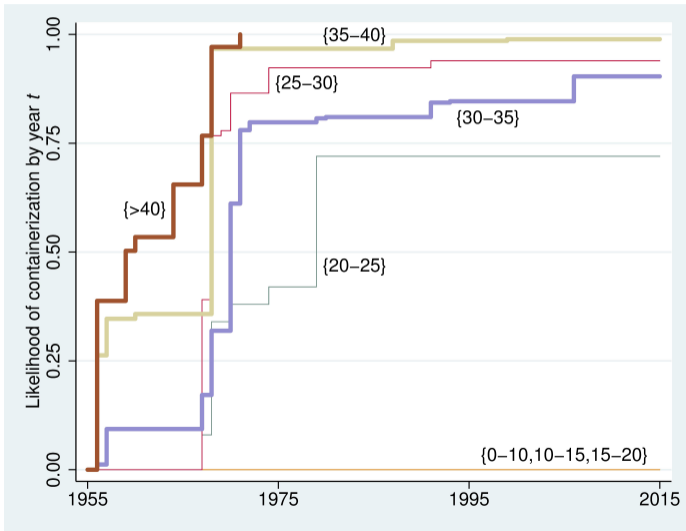
Context Matters



Context Matters



Calling Attention





Gestalt Principles of Visual Perception

- Proximity
- Similarity
- Enclosure
- Closure
- Continuity

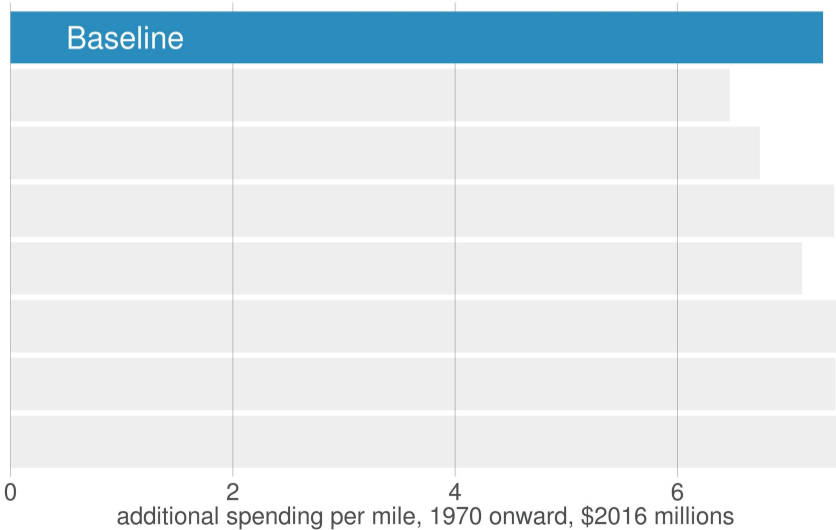
These all generate meaning, whether you intend it or not!



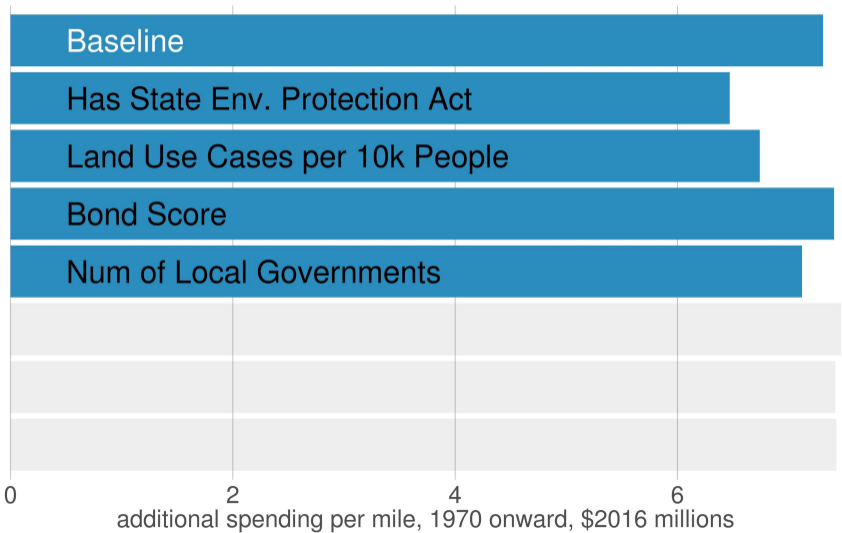
Applying These Principles

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

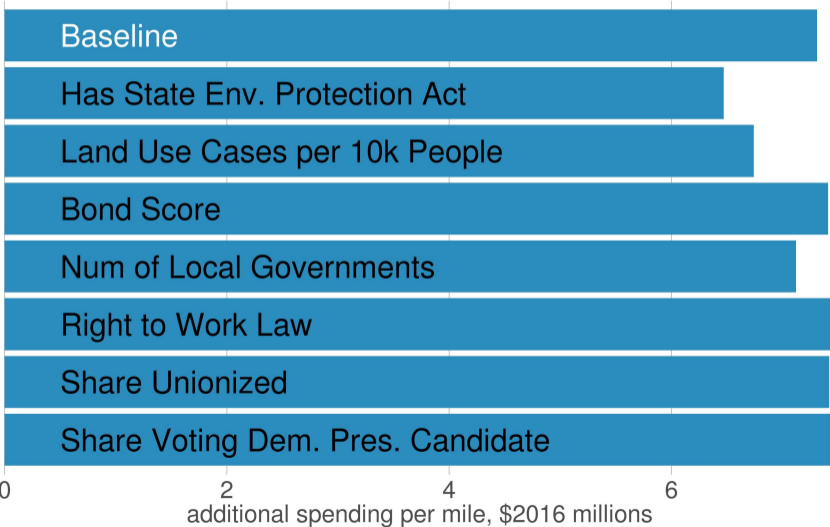
Baseline Increase of \$7.3 Million per Mile



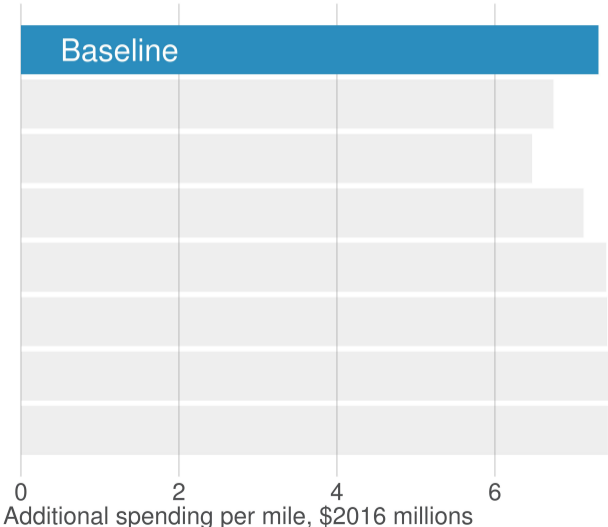
Measures of Government Quality Unrelated to Spending Increase



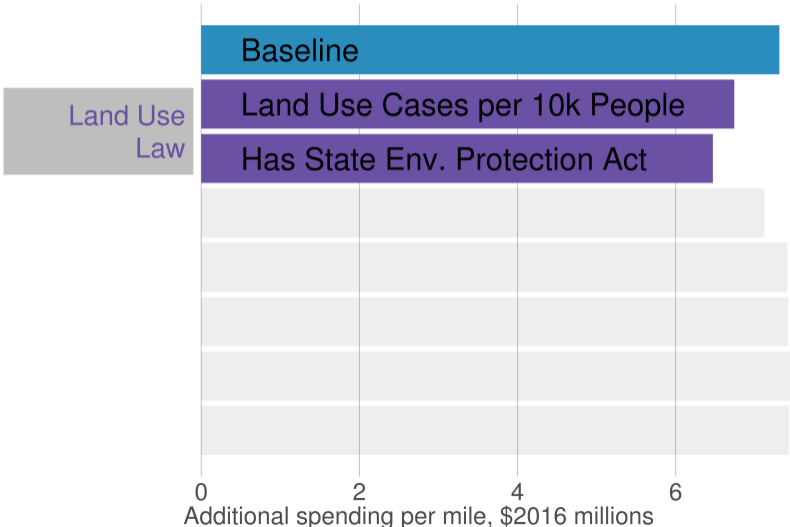
Measures of Labor Strength Unrelated to Spending Increase



Using the Principles of Proximity and Similarity

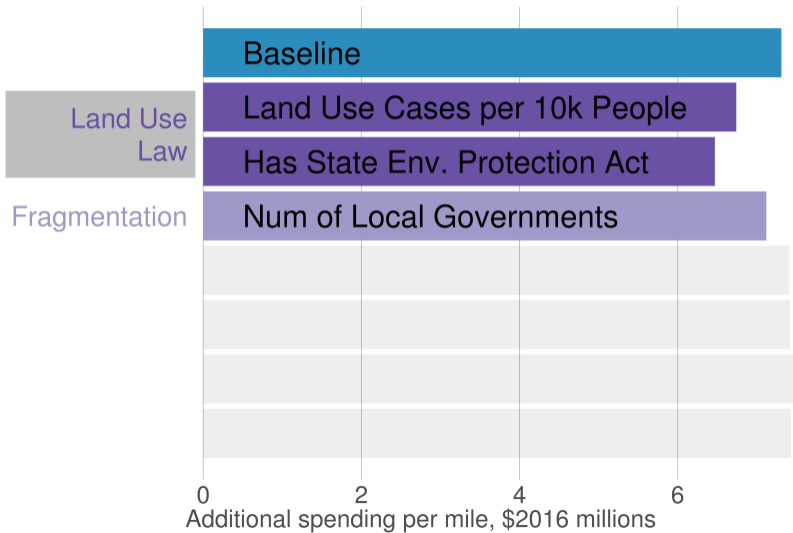


Using the Principles of Proximity and Similarity

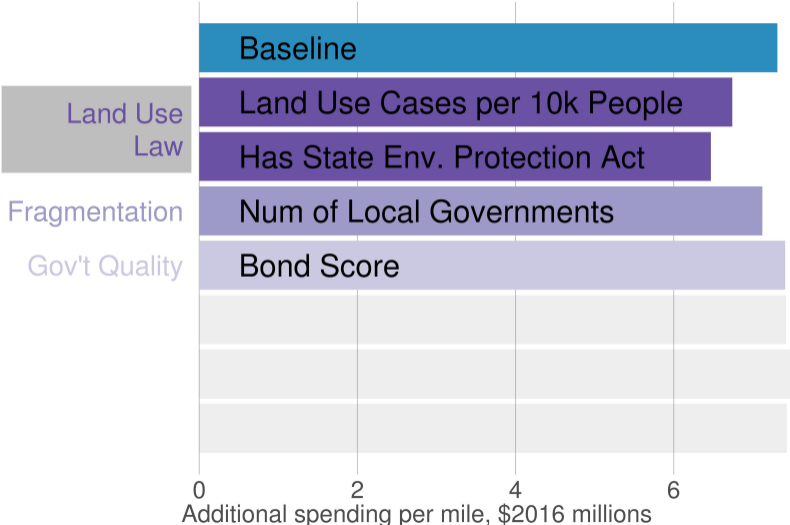




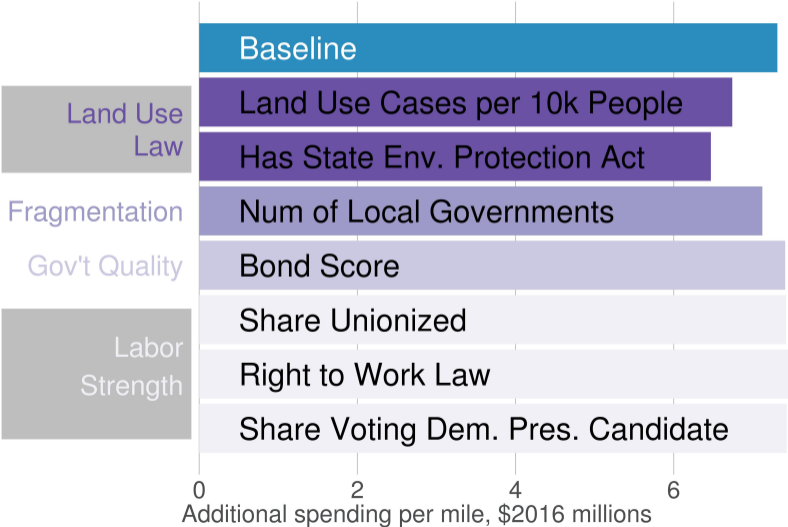
Using the Principles of Proximity and Similarity



Using the Principles of Proximity and Similarity



Using the Principles of Proximity and Similarity



R: Merging

In Your Breakout Group

Make examples

- make tiny dataframes
- to illustrate
- 1 to 1 merge
- 1 to many merge

In Your Breakout Group

Make examples

- make tiny dataframes
- to illustrate
- 1 to 1 merge
- 1 to many merge

Bonus! Why is a many to many merge problematic?

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

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!

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.



Online Lecture: Merging in R

In Class Wrap-Up

Today

- A. What is Merging?
- B. How to Merge 1:1
- C. How to Merge Many to 1
- D. Cautions with merging

A. Merging

- ▶ if you have information in more than one dataframe
- ▶ you want to combine these pieces of information
- ▶ reliably and replicably
- ▶ 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
- ▶ → 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
- ▶ → merge by neighborhood id!

Ex. 2:

- ▶ you have a dataset of student performance
- ▶ you want to add information on teacher
- ▶ → 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: Class subjects

```
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: Class attendance

```
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 d3 have?

B. Merge 1:1

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

How many rows should d3 have?

```
df3
```

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

C. Merge Many to 1

Many to 1 merge:

- ▶ this is a merge that has unique values in one dataset
- ▶ and repeat values in another

C. Merge Many to 1

Many to 1 merge:

- ▶ this is a merge that has unique values in one dataset
- ▶ and repeat values in another

Unique and repeat values:

- ▶ unique values: class in df3
- ▶ repeat values: subject in df3

df3

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

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. Frequent Problems with Merging

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

D. Frequent 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)

D. Frequent 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)

Why worry?

D. Frequent 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)

Why worry?

- ▶ bad merges yield garbage
- ▶ garbage in → garbage out

Try Today's Tutorial

- Make a .R script for whole tutorial
- Plus questions at end
- Go forth!
- I will be online till 5:20 – please stay and ask questions

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