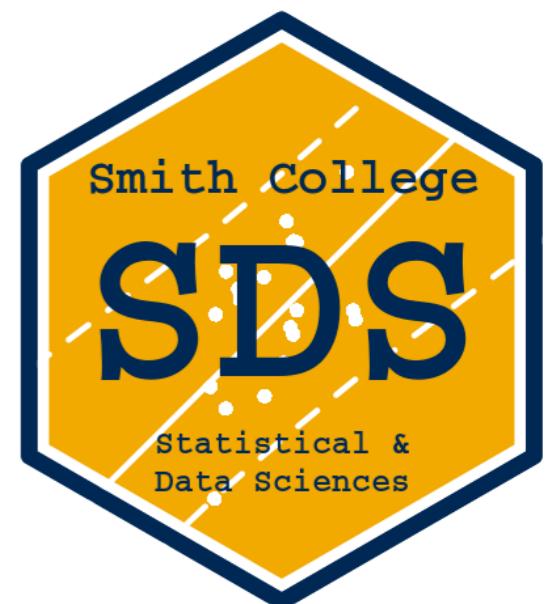


moderndive: statistical inference via the tidyverse



Albert Y. Kim
[@rudeboybert](https://twitter.com/rudeboybert)

Statistical Society of Canada
Calgary, Alberta
May 29, 2019



My Context for moderndive

My Context for moderndive

My students:

My Context for moderndive

My students:

- Undergraduate-only women's liberal arts college

My Context for moderndive

My students:

- Undergraduate-only women's liberal arts college
- Service intro stats course for all majors, all years

My Context for moderndive

My students:

- Undergraduate-only women's liberal arts college
- Service intro stats course for all majors, all years
- Calculus is a pre-req only in name

My Context for modernive

My students:

- Undergraduate-only women's liberal arts college
- Service intro stats course for all majors, all years
- Calculus is a pre-req only in name
- 13 weeks x (3 x 70min lectures + 75min lab)

My Context for moderndive

My students:

- Undergraduate-only women's liberal arts college
- Service intro stats course for all majors, all years
- Calculus is a pre-req only in name
- 13 weeks x (3 x 70min lectures + 75min lab)
- 29/40 had never coded in R prior

My Context for moderndive

My students:

- Undergraduate-only women's liberal arts college
- Service intro stats course for all majors, all years
- Calculus is a pre-req only in name
- 13 weeks x (3 x 70min lectures + 75min lab)
- 29/40 had never coded in R prior

My Context for modernRdive

My students:

- Undergraduate-only women's liberal arts college
- Service intro stats course for all majors, all years
- Calculus is a pre-req only in name
- 13 weeks x (3 x 70min lectures + 75min lab)
- 29/40 had never coded in R prior

My goals:

My Context for modernive

My students:

- Undergraduate-only women's liberal arts college
- Service intro stats course for all majors, all years
- Calculus is a pre-req only in name
- 13 weeks x (3 x 70min lectures + 75min lab)
- 29/40 had never coded in R prior

My goals:

- Goal 1: Modeling with regression

My Context for moderndive

My students:

- Undergraduate-only women's liberal arts college
- Service intro stats course for all majors, all years
- Calculus is a pre-req only in name
- 13 weeks x (3 x 70min lectures + 75min lab)
- 29/40 had never coded in R prior

My goals:

- Goal 1: Modeling with regression
- Goal 2: Sampling for inference

Getting from Point A to Point B

Getting from Point A to Point B

Point A:
Modal 1st time
stats student



Getting from Point A to Point B

Point A:
Modal 1st time
stats student



Point B:
Two goals



1. Modeling with regression
2. Sampling for inference

Getting from Point A to Point B

Point A:
Modal 1st time
stats student



Point B:
Two goals

1. Modeling with regression
2. Sampling for inference



Calculus?

😁 thru 😤

Getting from Point A to Point B

Point A:
Modal 1st time
stats student



Point B:
Two goals

1. Modeling with regression
2. Sampling for inference



Calculus?



Coding?



Getting from Point A to Point B

Point A:
Modal 1st time
stats student

via the
tidyverse

Point B:
Two goals

1. Modeling with regression
2. Sampling for inference

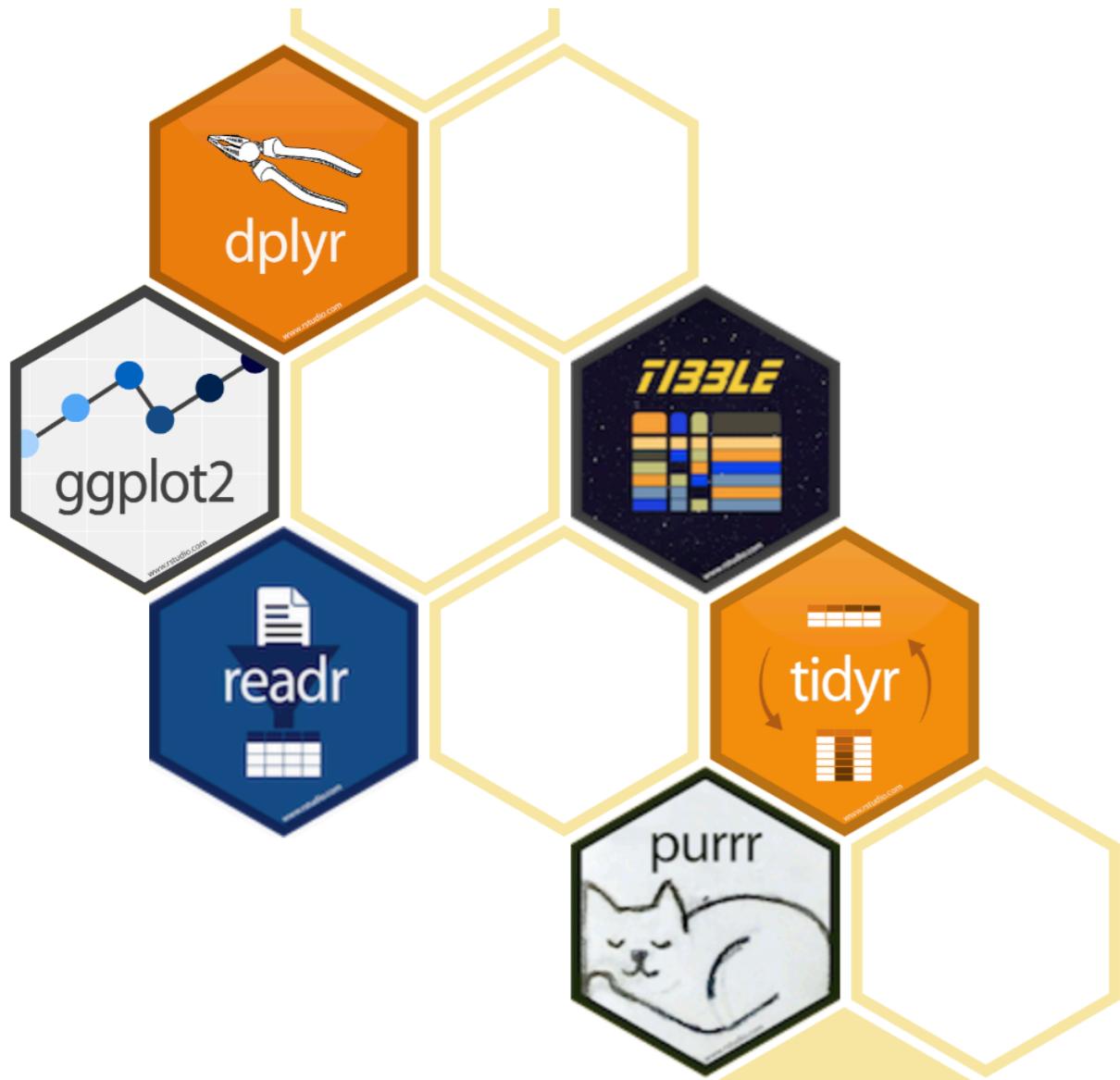


Calculus?
😁 thru 🤢

Coding?
😱 & 🤔

What is the tidyverse?

From tidyverse.org:



R packages for data science

The tidyverse is an opinionated **collection of R packages** designed for data science. All packages share an underlying design philosophy, grammar, and data structures.

Install the complete tidyverse with:

```
install.packages("tidyverse")
```

Why tidyverse for stats newbies?

Why tidyverse for stats newbies?

- IMO it's easier to learn than base R. [Others too.](#)

Why tidyverse for stats newbies?

- IMO it's easier to learn than base R. [Others too.](#)
- It scales. You leverage an entire ecosystem of online developers and support: Google & StackOverflow

Why tidyverse for stats newbies?

- IMO it's easier to learn than base R. [Others too.](#)
- It scales. You leverage an entire ecosystem of online developers and support: Google & StackOverflow
- Satisfy learning goals *while learning tools they can use beyond the classroom.*

End Deliverable of Course

End Deliverable of Course

- Think of how youths learn to play sports...

End Deliverable of Course

- Think of how youths learn to play sports...
- IMO stats newbies should learn to “*play the whole game*” in simplified form first

End Deliverable of Course

- Think of how youths learn to play sports...
- IMO stats newbies should learn to “*play the whole game*” in simplified form first
 - %>% then add layers of complexity...

End Deliverable of Course

- Think of how youths learn to play sports...
- IMO stats newbies should learn to “*play the whole game*” in simplified form first
 - %>% then add layers of complexity...
 - %>% then add more layers of complexity...

End Deliverable of Course

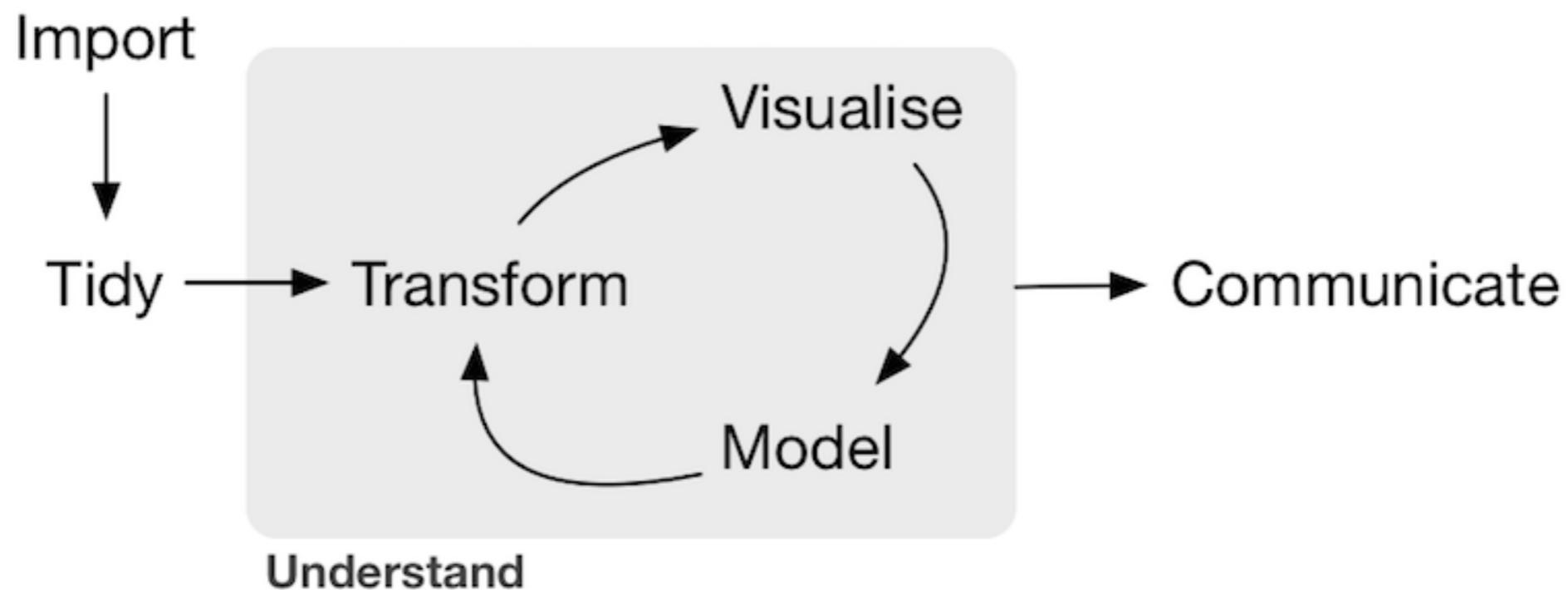
- Think of how youths learn to play sports...
- IMO stats newbies should learn to “*play the whole game*” in simplified form first
 - %>% then add layers of complexity...
 - %>% then add more layers of complexity...
 - %>% then add more layers of complexity...

End Deliverable of Course

- Think of how youths learn to play sports...
- IMO stats newbies should learn to “*play the whole game*” in simplified form first
 - %>% then add layers of complexity...
 - %>% then add more layers of complexity...
 - %>% then add more layers of complexity...
- Do this instead of learning individual components in isolation

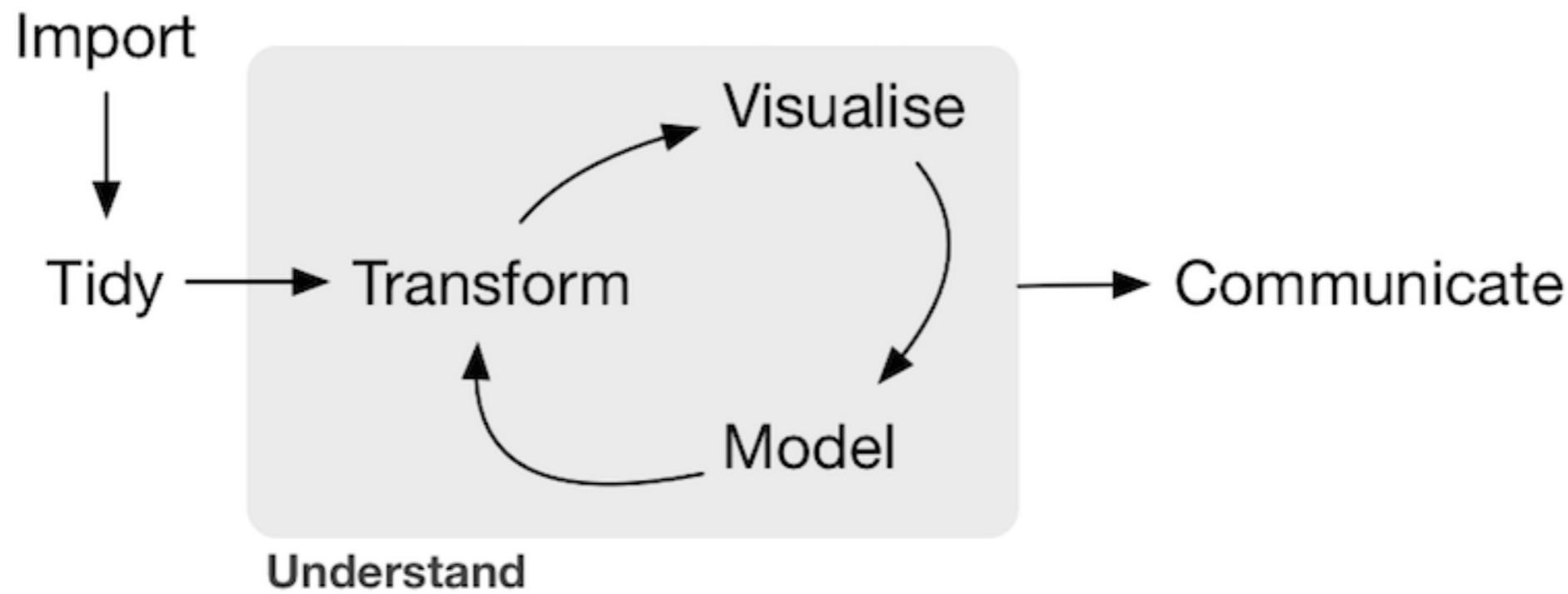
End Deliverable of Course

Final project that “plays the whole game”
of *all components* of data/science pipeline:



End Deliverable of Course

Final project that “plays the whole game”
of *all components* of data/science pipeline:



Example template given to students this semester,
based on work by students
Alexis, Andrianne, & Isabel.

The R Series

Statistical Inference via Data Science

A moderndive into R & the tidyverse



**Chester Ismay
Albert Y. Kim**

CRC Press
Taylor & Francis Group
A CHAPMAN & HALL BOOK

Fall 2019!

Development version at moderndive.netlify.com

Part I: Data Science via the tidyverse

Chapters 2 - 5

Chapter 2: Getting Started

Chapter 2: Getting Started

R: Engine



RStudio: Dashboard



Chapter 2: Getting Started

R: Engine



RStudio: Dashboard



R: A new phone



R Packages: Apps you can download



Chapter 2: Getting Started

R: Engine



RStudio: Dashboard



R: A new phone



R Packages: Apps you can download



Getting students over initial 😱 of coding

Chapter 3: Data Viz via `ggplot2`

Chapter 3: Data Viz via `ggplot2`

Often said “Intro students can’t learn ggplot”

Chapter 3: Data Viz via ggplot2

Often said “Intro students can’t learn ggplot”

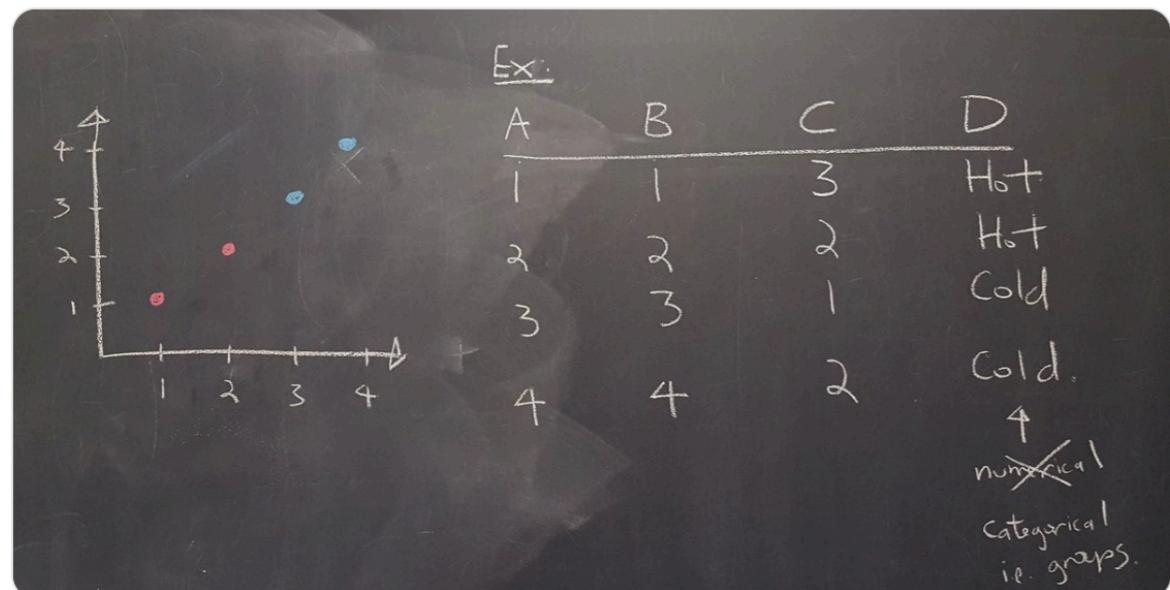


Albert Y. Kim

@rudeboybert



Intro stats & data science #chalktalk of
grammar of graphics + homage to
@katyperry today, #ggplot2 tomorrow
#rstats



11:58 AM - 11 Sep 2017 from Amherst College

5 Retweets 29 Likes



3

5

29



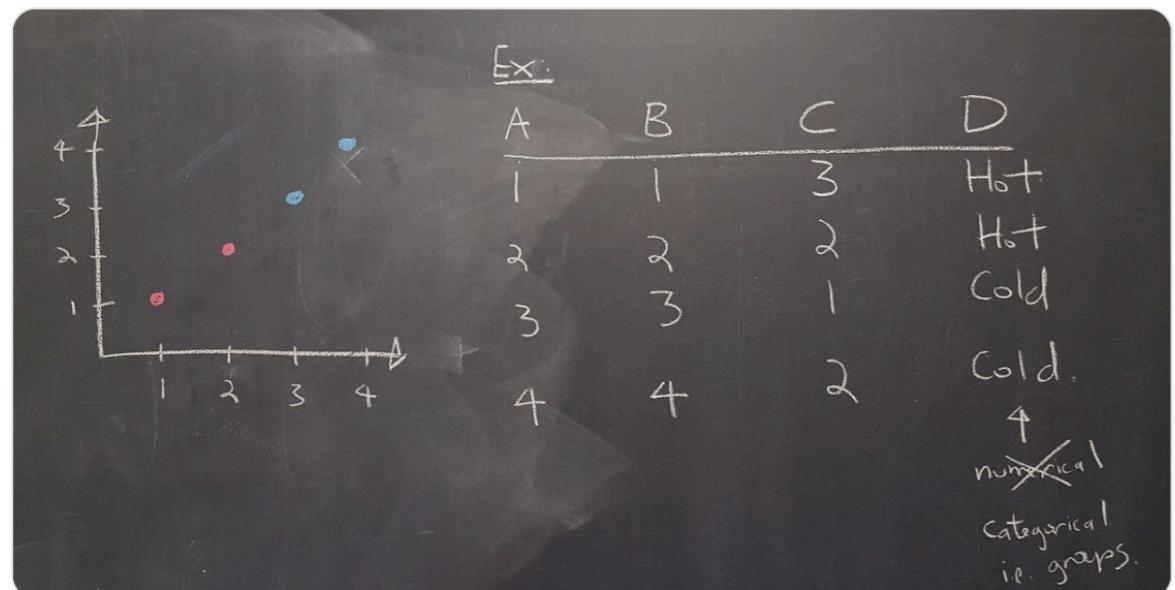
Chapter 3: Data Viz via ggplot2

Often said “Intro students can’t learn ggplot”



Albert Y. Kim
@rudeboybert

Intro stats & data science #chalktalk of grammar of graphics + homage to @katyperry today, #ggplot2 tomorrow #rstats



11:58 AM - 11 Sep 2017 from Amherst College

5 Retweets 29 Likes



3 5 29



Albert Y. Kim
@rudeboybert

#chalktalk of #GrammarOfGraphics definition of "statistical graphic" + @ModernDive's "Five Named Graphs" #5NG #ggplot2

Recall:

A statistical graphic is a mapping of data variables to aesthetic attributes of geometric objects.

Five Named Graphs 5NG

- ① Scatterplots geom_point()
- ② Line graphs geom_line()
- ③ Histograms geom_histogram()
- ④ Boxplots geom_boxplot()
- ⑤ Barplots geom_bar()

12:50 PM - 12 Sep 2017 from Amherst College

15 Retweets 61 Likes



Q 15 61

Chapter 4: Data Wrangling via `dplyr`

Chapter 5: “Tidy” Data via `tidyverse`

Chapter 4: Data Wrangling via `dplyr`

Chapter 5: “Tidy” Data via `tidyverse`

- Essential: `%>%` operator as it's needed later.

Chapter 4: Data Wrangling via `dplyr`

Chapter 5: “Tidy” Data via `tidyverse`

- Essential: `%>%` operator as it's needed later.
- Balance of how much students wrangling do vs how much you do for them?

Chapter 4: Data Wrangling via `dplyr`

Chapter 5: “Tidy” Data via `tidyverse`

- Essential: `%>%` operator as it's needed later.
- Balance of how much students wrangling do vs how much you do for them?
- To *completely* shield students from *any* data wrangling is to betray [true nature of work in our fields](#).

Chapter 4: Data Wrangling via dplyr

Chapter 5: “Tidy” Data via tidyverse

- Essential: `%>%` operator as it's needed later.
- Balance of how much students wrangling do vs how much you do for them?
- To *completely* shield students from *any* data wrangling is to betray [true nature of work in our fields](#).
- How much data [“taming”](#) is necessary?

Part II: Data Modeling via moderndive

Chapters 6, 7, & 11

Goal 1: Modeling with Regression

Goal 1: Modeling with Regression

1. Data: evals

ID	score	age	gender
1	4.7	36	female
2	4.1	36	female
3	3.9	36	female
4	4.8	36	female
5	4.6	59	male
6	4.3	59	male
7	2.8	59	male
8	4.1	51	male
9	3.4	51	male
10	4.5	40	female
11	3.8	40	female
12	4.5	40	female

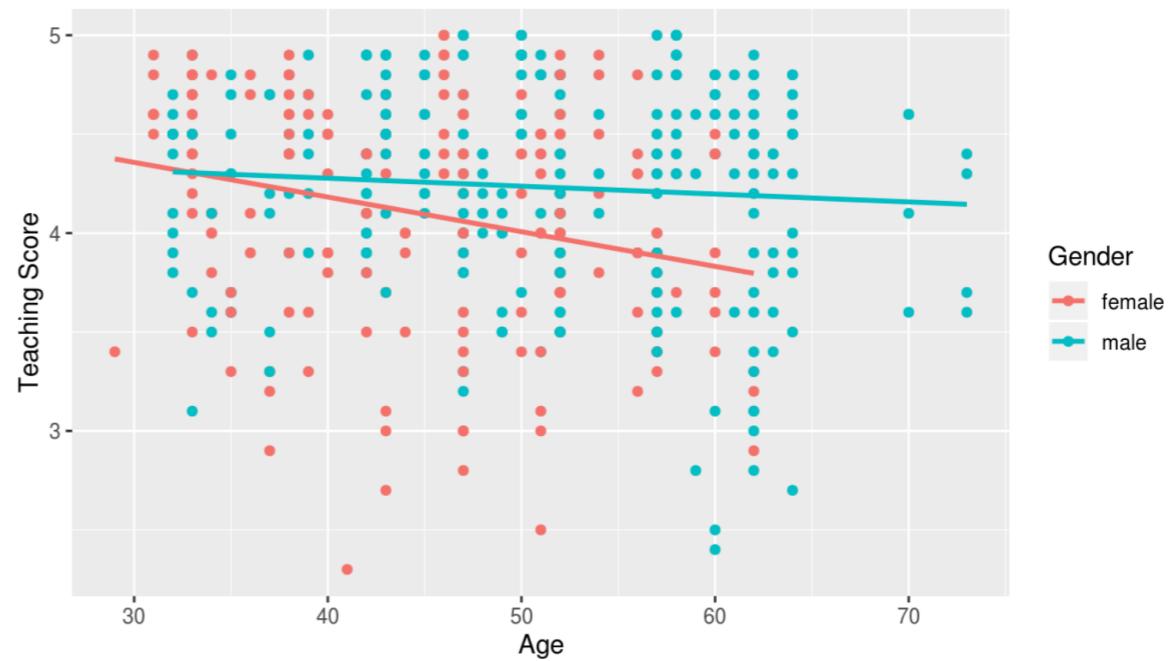
Goal 1: Modeling with Regression

1. Data: evals

	ID	score	age	gender
1	1	4.7	36	female
2	2	4.1	36	female
3	3	3.9	36	female
4	4	4.8	36	female
5	5	4.6	59	male
6	6	4.3	59	male
7	7	2.8	59	male
8	8	4.1	51	male
9	9	3.4	51	male
10	10	4.5	40	female
11	11	3.8	40	female
12	12	4.5	40	female



2. Exploratory Data Analysis

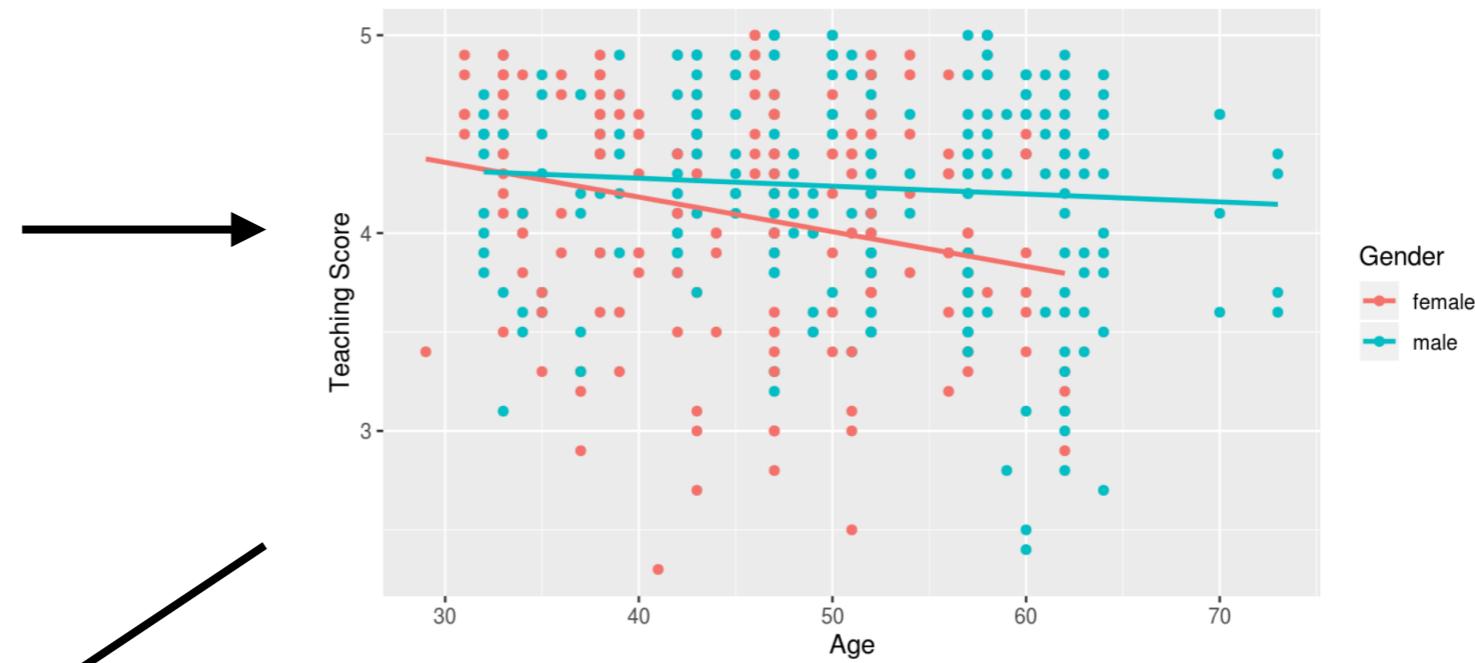


Goal 1: Modeling with Regression

1. Data: evals

	ID	score	age	gender
1	1	4.7	36	female
2	2	4.1	36	female
3	3	3.9	36	female
4	4	4.8	36	female
5	5	4.6	59	male
6	6	4.3	59	male
7	7	2.8	59	male
8	8	4.1	51	male
9	9	3.4	51	male
10	10	4.5	40	female
11	11	3.8	40	female
12	12	4.5	40	female

2. Exploratory Data Analysis



3. Regression Coeff

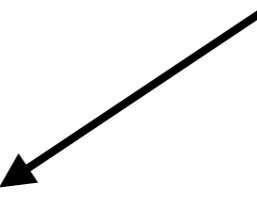
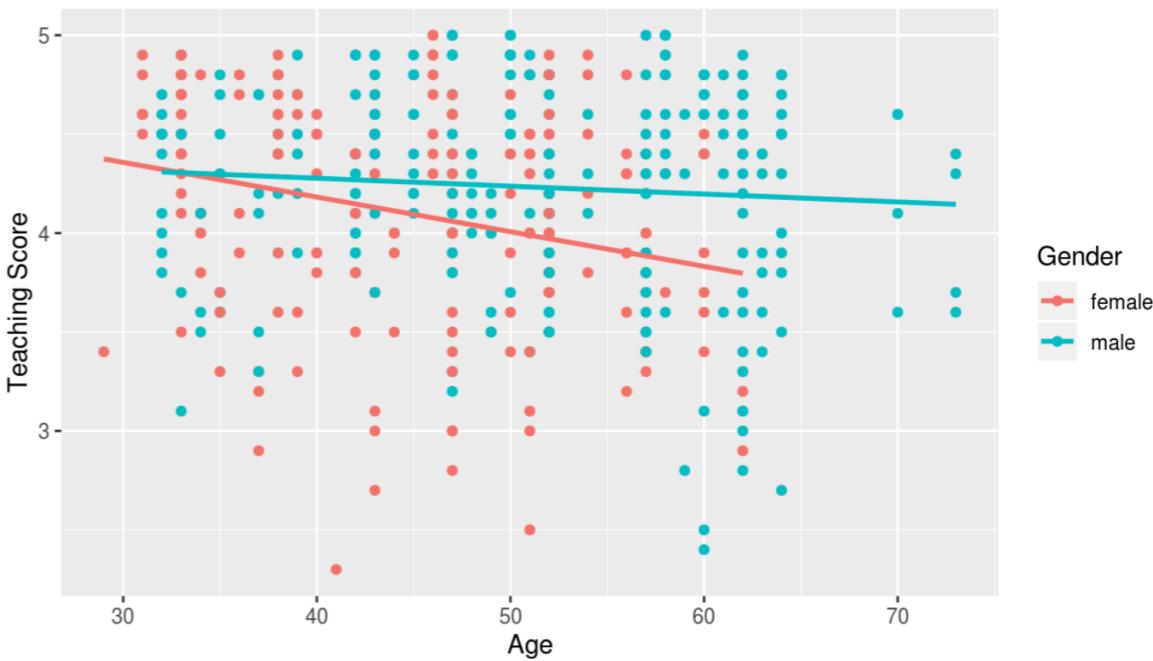
```
Console ~ / ↗
> score_model <- lm(score ~ age * gender, data = evals)
> get_regression_table(score_model)
# A tibble: 4 x 7
  term      estimate
  <chr>    <dbl>
1 intercept  4.88
2 age        -0.018
3 gendermale -0.446
4 age:gendermale  0.014
> |
```

Goal 1: Modeling with Regression

1. Data: evals

2. Exploratory Data Analysis

ID	score	age	gender
1	4.7	36	female
2	4.1	36	female
3	3.9	36	female
4	4.8	36	female
5	4.6	59	male
6	4.3	59	male
7	2.8	59	male
8	4.1	51	male
9	3.4	51	male
10	4.5	40	female
11	3.8	40	female
12	4.5	40	female



3. Regression Coeff

```
Console ~ / ↗
> score_model <- lm(score ~ age * gender, data = evals)
> get_regression_table(score_model)
# A tibble: 4 x 7
  term      estimate
  <chr>    <dbl>
1 intercept 4.88
2 age       -0.018
3 gendermale -0.446
4 age:gendermale 0.014
> |
```

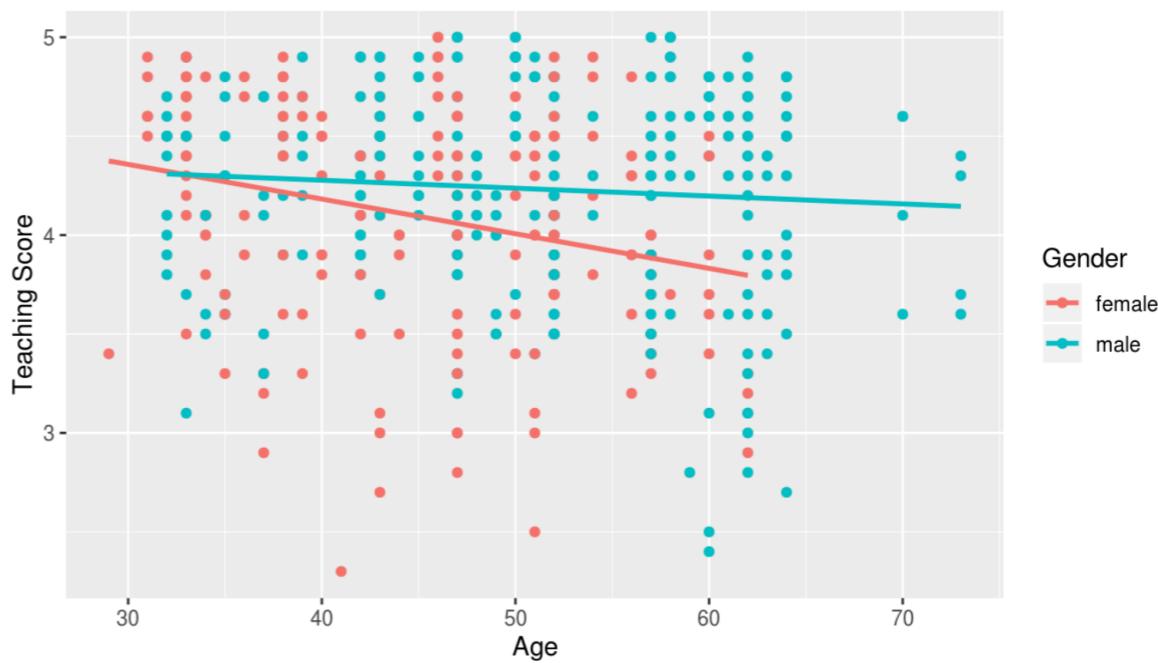
Early: Descriptive regression

Goal 1: Modeling with Regression

1. Data: evals

ID	score	age	gender
1	4.7	36	female
2	4.1	36	female
3	3.9	36	female
4	4.8	36	female
5	4.6	59	male
6	4.3	59	male
7	2.8	59	male
8	4.1	51	male
9	3.4	51	male
10	4.5	40	female
11	3.8	40	female
12	4.5	40	female

2. Exploratory Data Analysis



3. Regression Coeff

```
Console ~ / 
> score_model <- lm(score ~ age * gender, data = evals)
> get_regression_table(score_model)
# A tibble: 4 x 7
  term      estimate
  <chr>    <dbl>
1 intercept 4.88
2 age       -0.018
3 gendermale -0.446
4 age:gendermale 0.014
> |
```



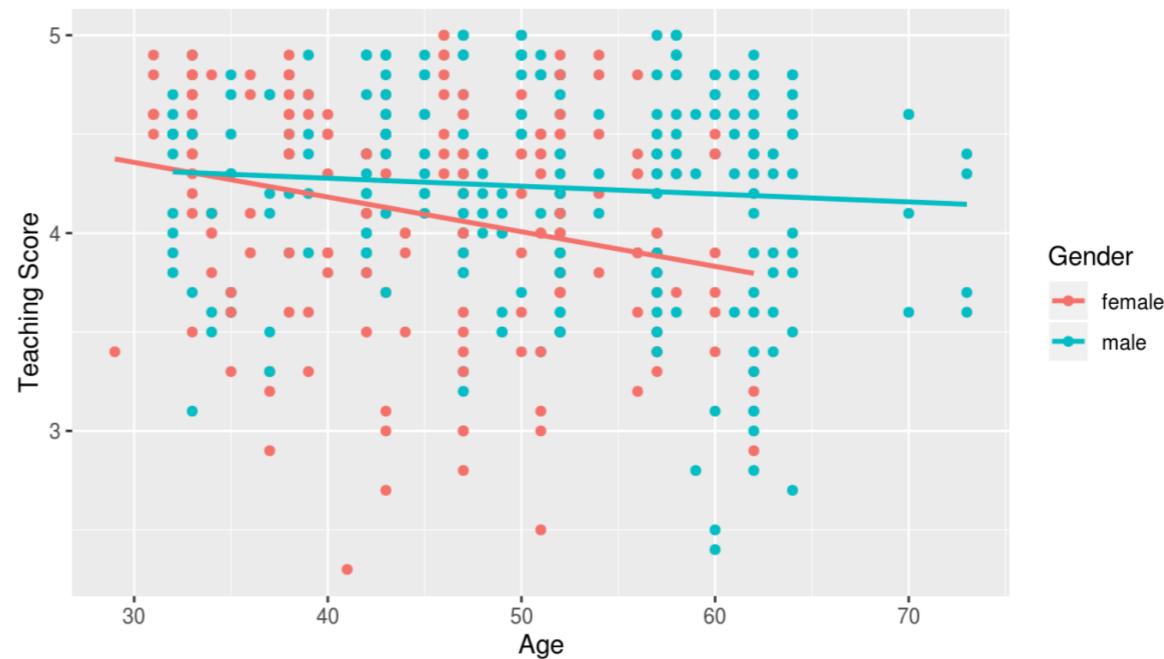
Early: Descriptive regression

Goal 1: Modeling with Regression

1. Data: evals

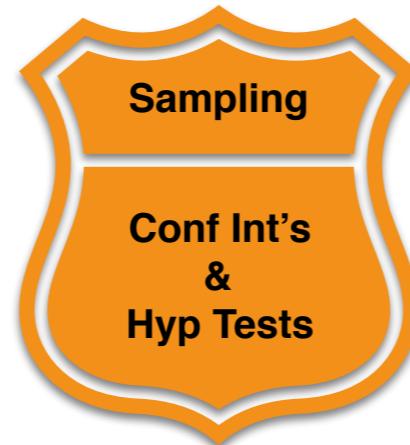
ID	score	age	gender
1	4.7	36	female
2	4.1	36	female
3	3.9	36	female
4	4.8	36	female
5	4.6	59	male
6	4.3	59	male
7	2.8	59	male
8	4.1	51	male
9	3.4	51	male
10	4.5	40	female
11	3.8	40	female
12	4.5	40	female

2. Exploratory Data Analysis



3. Regression Coeff

```
Console ~ / 
> score_model <- lm(score ~ age * gender, data = evals)
> get_regression_table(score_model)
# A tibble: 4 x 7
  term      estimate
  <chr>    <dbl>
1 intercept 4.88
2 age       -0.018
3 gendermale -0.446
4 age:gendermale 0.014
> |
```



4. Regression Table

```
Console ~ / 
> score_model <- lm(score ~ age * gender, data = evals)
> get_regression_table(score_model)
# A tibble: 4 x 7
  term      estimate std_error statistic p_value lower_ci upper_ci
  <chr>    <dbl>     <dbl>     <dbl>    <dbl>    <dbl>     <dbl>
1 intercept 4.88      0.205     23.8     0        4.48      5.29
2 age       -0.018    0.004     -3.92    0        -0.026    -0.009
3 gendermale -0.446    0.265     -1.68    0.094   -0.968    0.076
4 age:gendermale 0.014    0.006     2.45    0.015   0.003     0.024
> |
```

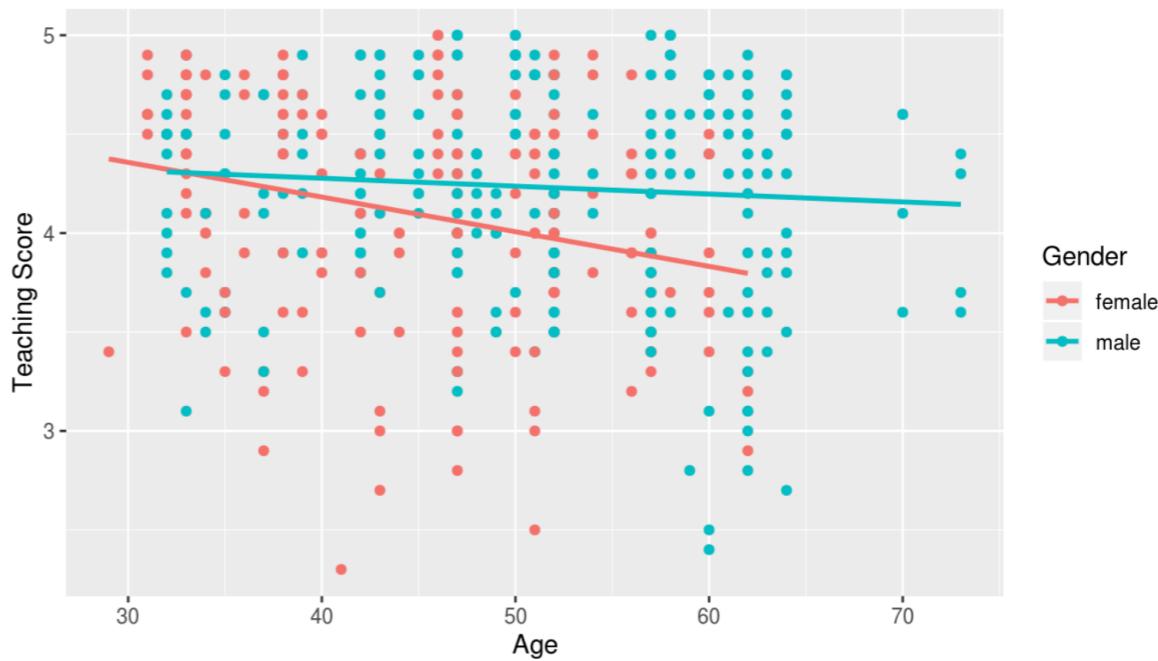
Early: Descriptive regression

Goal 1: Modeling with Regression

1. Data: evals

ID	score	age	gender
1	4.7	36	female
2	4.1	36	female
3	3.9	36	female
4	4.8	36	female
5	4.6	59	male
6	4.3	59	male
7	2.8	59	male
8	4.1	51	male
9	3.4	51	male
10	4.5	40	female
11	3.8	40	female
12	4.5	40	female

2. Exploratory Data Analysis



3. Regression Coeff

```
Console ~ / 
> score_model <- lm(score ~ age * gender, data = evals)
> get_regression_table(score_model)
# A tibble: 4 x 7
  term      estimate
  <chr>    <dbl>
1 intercept  4.88
2 age        -0.018
3 gendermale -0.446
4 age:gendermale  0.014
> |
```



```
Console ~ / 
> score_model <- lm(score ~ age * gender, data = evals)
> get_regression_table(score_model)
# A tibble: 4 x 7
  term      estimate std_error statistic p_value lower_ci upper_ci
  <chr>    <dbl>     <dbl>     <dbl>    <dbl>    <dbl>     <dbl>
1 intercept  4.88     0.205    23.8     0       4.48     5.29
2 age        -0.018   0.004    -3.92    0       -0.026   -0.009
3 gendermale -0.446   0.265    -1.68    0.094   -0.968   0.076
4 age:gendermale  0.014   0.006    2.45    0.015   0.003    0.024
> |
```

Early: Descriptive regression

Later: Inference for Regression

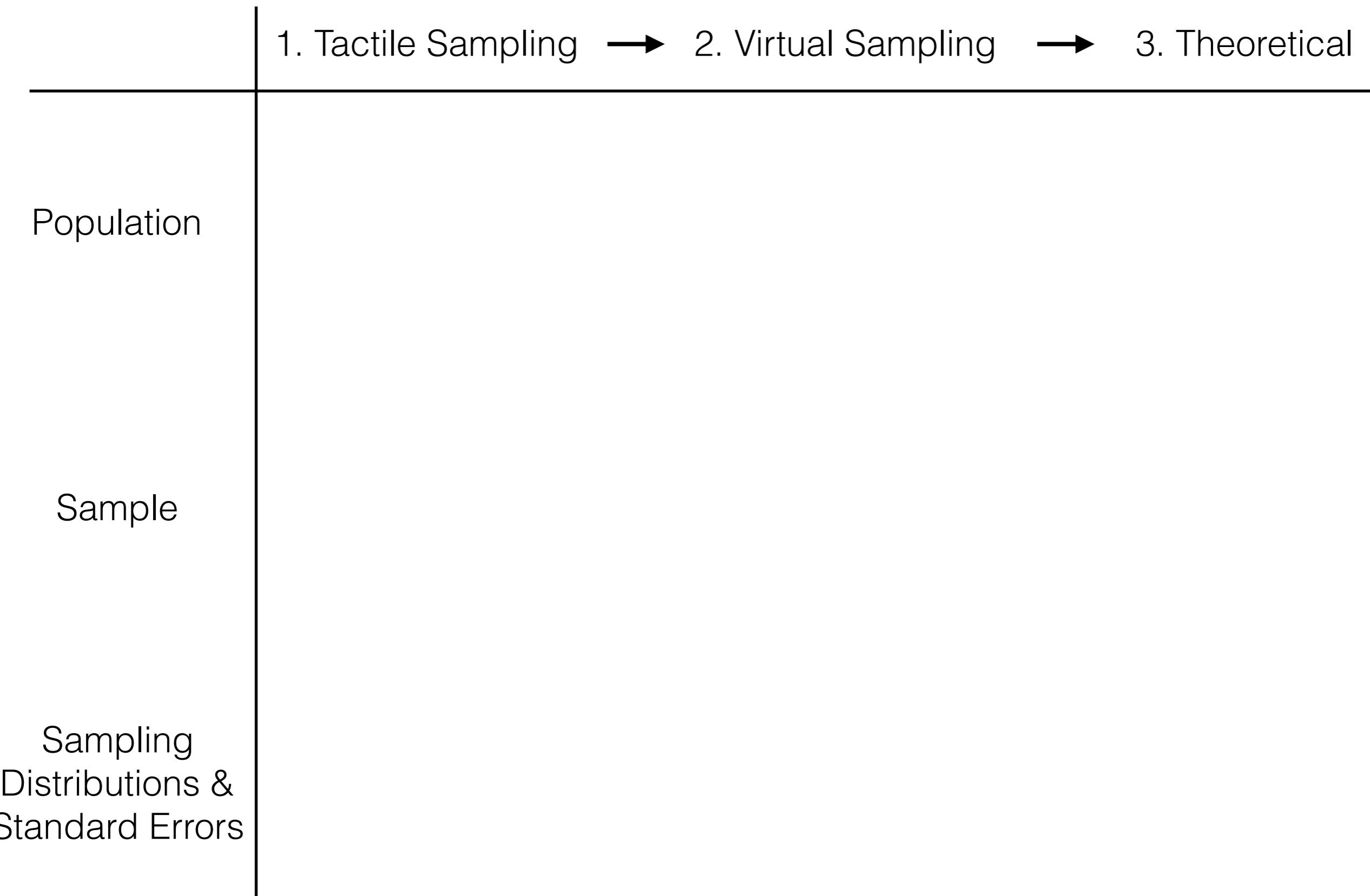
Part III: Statistical Inference via infer

Chapters 8 - 11

Goal 2: Sampling for Inference



Goal 2: Sampling for Inference



Goal 2: Sampling for Inference

1. Tactile Sampling → 2. Virtual Sampling → 3. Theoretical

Population



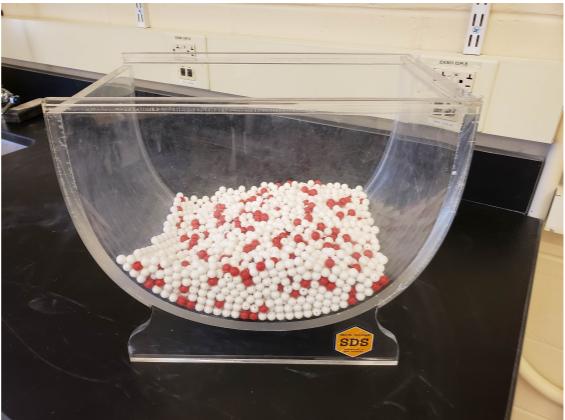
Sample

Sampling
Distributions &
Standard Errors

Goal 2: Sampling for Inference

1. Tactile Sampling → 2. Virtual Sampling → 3. Theoretical

Population



Sample



Sampling
Distributions &
Standard Errors

Goal 2: Sampling for Inference

1. Tactile Sampling → 2. Virtual Sampling → 3. Theoretical

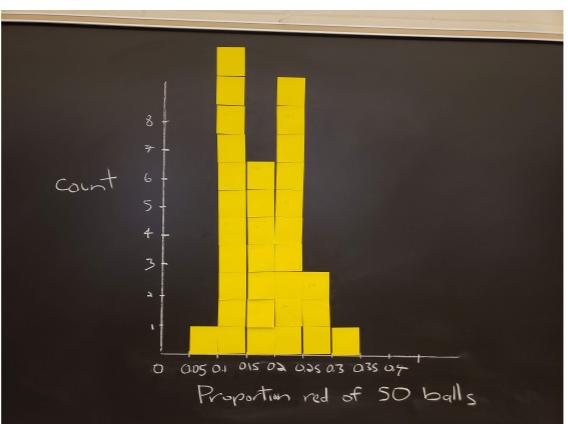
Population



Sample



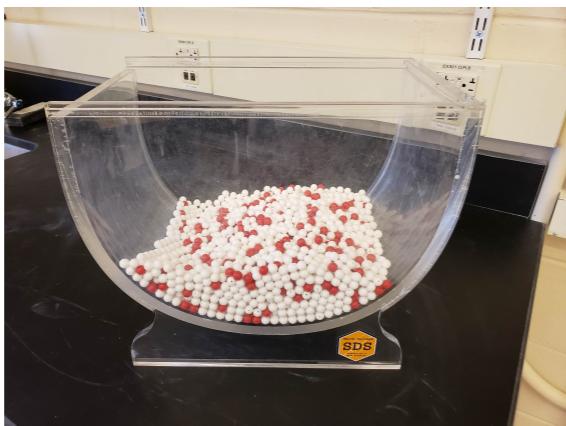
Sampling
Distributions &
Standard Errors



Goal 2: Sampling for Inference

1. Tactile Sampling → 2. Virtual Sampling → 3. Theoretical

Population

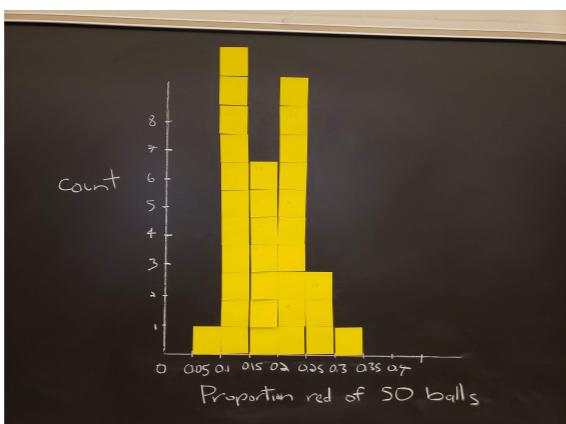


```
Console ~/ ↵
> library(moderndive)
> bowl
# A tibble: 2,400 x 2
  ball_ID color
  <int> <chr>
1     1 white
2     2 white
3     3 white
4     4 red
5     5 white
6     6 white
7     7 red
8     8 white
9     9 red
10    10 white
# ... with 2,390 more rows
> |
```

Sample



Sampling
Distributions &
Standard Errors



Goal 2: Sampling for Inference

1. Tactile Sampling → 2. Virtual Sampling → 3. Theoretical

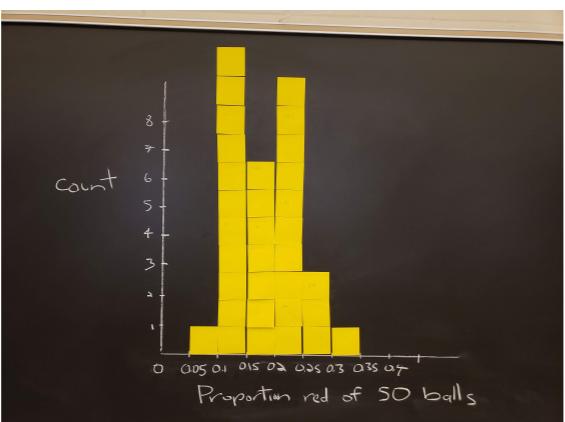
Population



Sample



Sampling
Distributions &
Standard Errors



```
Console ~/ ↵
> library(moderndive)
> bowl
# A tibble: 2,400 x 2
  ball_ID color
  <int> <chr>
1     1 white
2     2 white
3     3 white
4     4 red
5     5 white
6     6 white
7     7 red
8     8 white
9     9 red
10    10 white
# ... with 2,390 more rows
> |
```

```
Console ~/ ↵
> bowl %>%
+   rep_sample_n(size = 50, reps = 1)
# A tibble: 50 x 3
# Groups:   replicate [1]
  replicate ball_ID color
  <int> <int> <chr>
1       1     226 white
2       1    1304 red
3       1    1230 white
4       1     984 white
5       1      68 white
6       1    1965 white
7       1     431 white
8       1    1184 white
9       1    1610 red
10      1     978 white
# ... with 40 more rows
>
```

Goal 2: Sampling for Inference

1. Tactile Sampling → 2. Virtual Sampling → 3. Theoretical

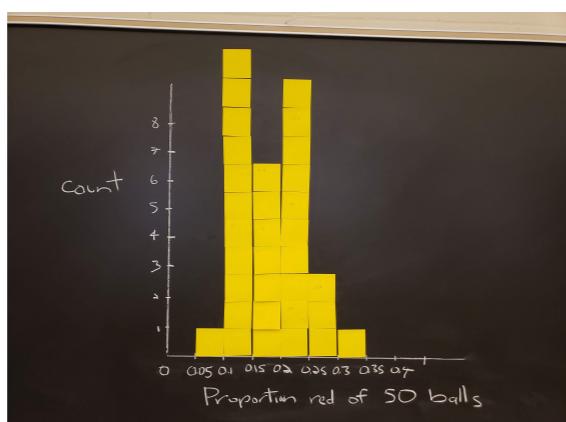
Population



Sample

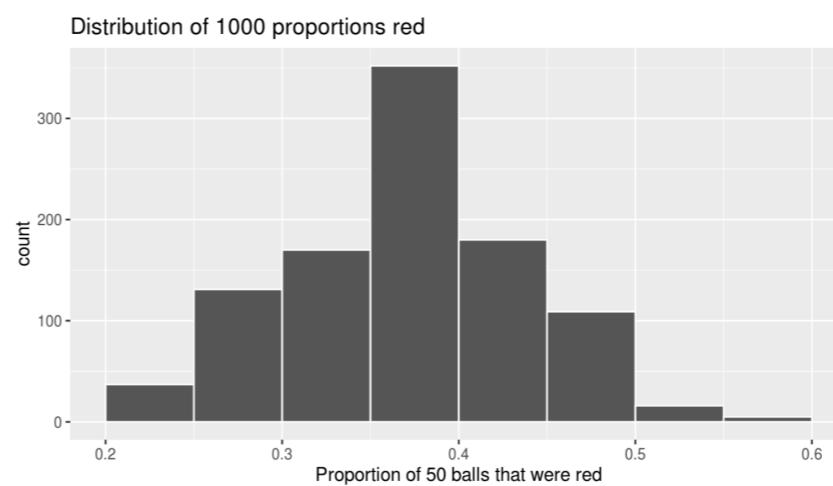


Sampling
Distributions &
Standard Errors



```
Console ~/ ↵
> library(moderndive)
> bowl
# A tibble: 2,400 x 2
  ball_ID color
  <int> <chr>
1     1 white
2     2 white
3     3 white
4     4 red
5     5 white
6     6 white
7     7 red
8     8 white
9     9 red
10    10 white
# ... with 2,390 more rows
> |
```

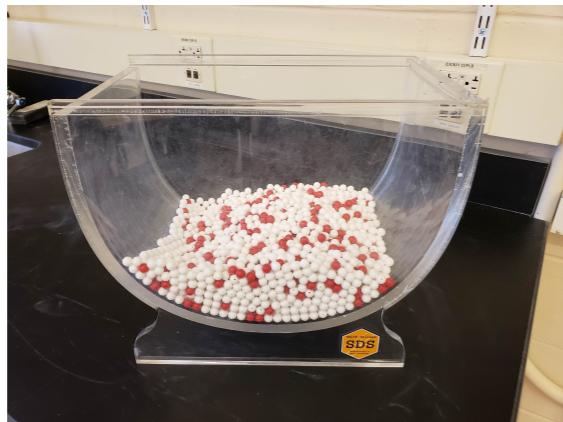
```
Console ~/ ↵
> bowl %>%
+   rep_sample_n(size = 50, reps = 1)
# A tibble: 50 x 3
# Groups:   replicate [1]
  replicate ball_ID color
  <int> <int> <chr>
1       1     1 226 white
2       1     1 1304 red
3       1     1 1230 white
4       1     1 984 white
5       1     1 68 white
6       1     1 1965 white
7       1     1 431 white
8       1     1 1184 white
9       1     1 1610 red
10      1     1 978 white
# ... with 40 more rows
> |
```



Goal 2: Sampling for Inference

1. Tactile Sampling → 2. Virtual Sampling → 3. Theoretical

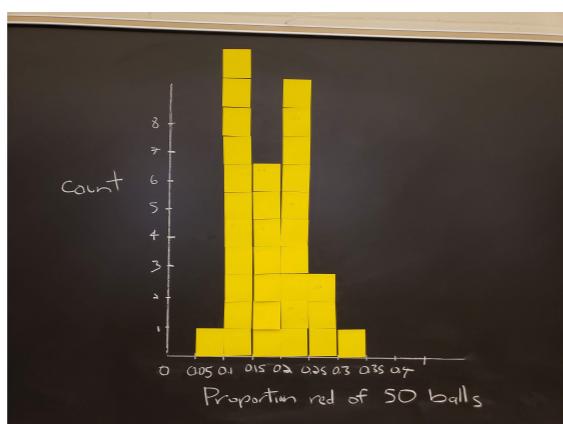
Population



Sample

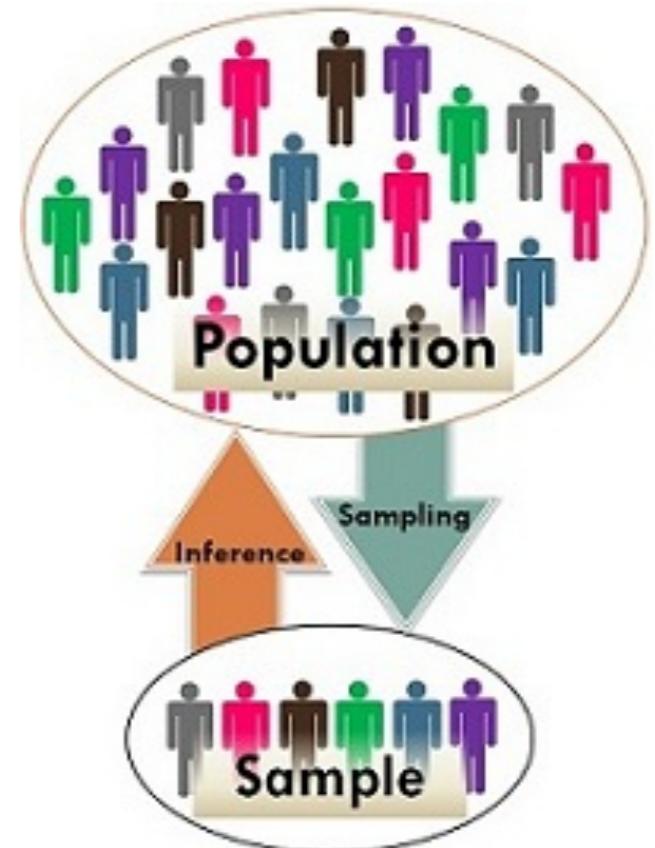
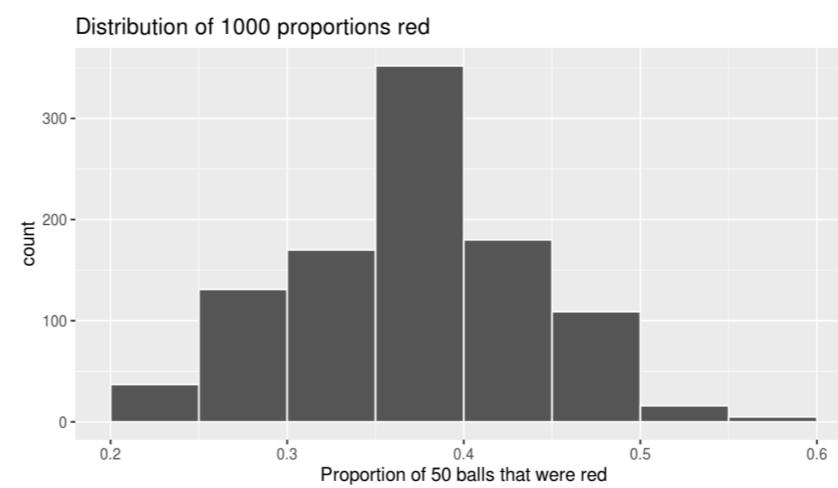


Sampling
Distributions &
Standard Errors



```
Console ~/ ↵
> library(moderndive)
> bowl
# A tibble: 2,400 x 2
  ball_ID color
  <int> <chr>
1     1 white
2     2 white
3     3 white
4     4 red
5     5 white
6     6 white
7     7 red
8     8 white
9     9 red
10    10 white
# ... with 2,390 more rows
> |
```

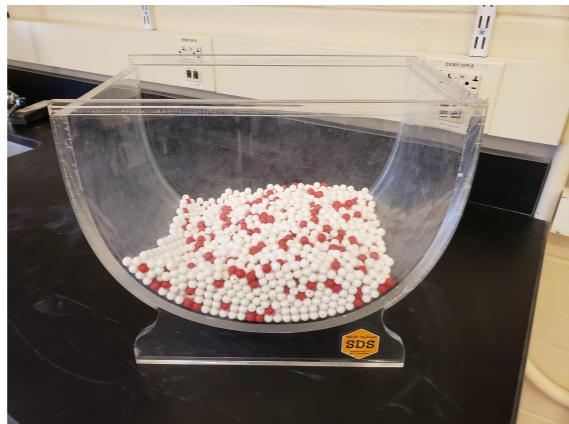
```
Console ~/ ↵
> bowl %>%
+   rep_sample_n(size = 50, reps = 1)
# A tibble: 50 x 3
# Groups:   replicate [1]
  replicate ball_ID color
  <int> <int> <chr>
1       1     1  white
2       1     1  red
3       1     1  white
4       1     1  white
5       1     1  white
6       1     1  white
7       1     1  white
8       1     1  white
9       1     1  red
10      1     1  white
# ... with 40 more rows
> |
```



Goal 2: Sampling for Inference

1. Tactile Sampling → 2. Virtual Sampling → 3. Theoretical

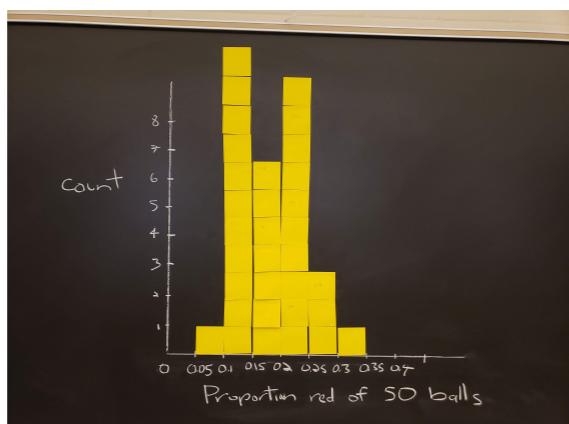
Population



Sample

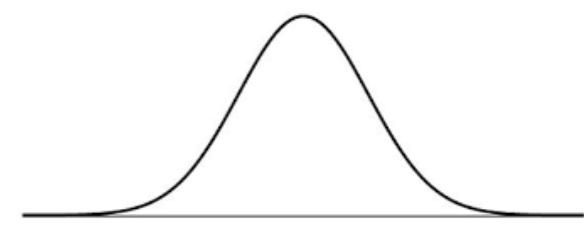
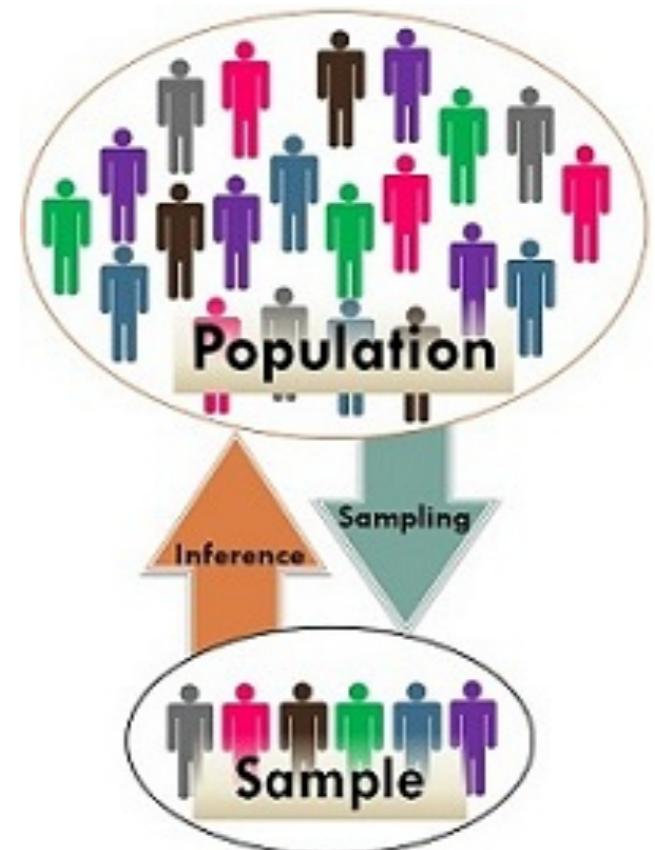
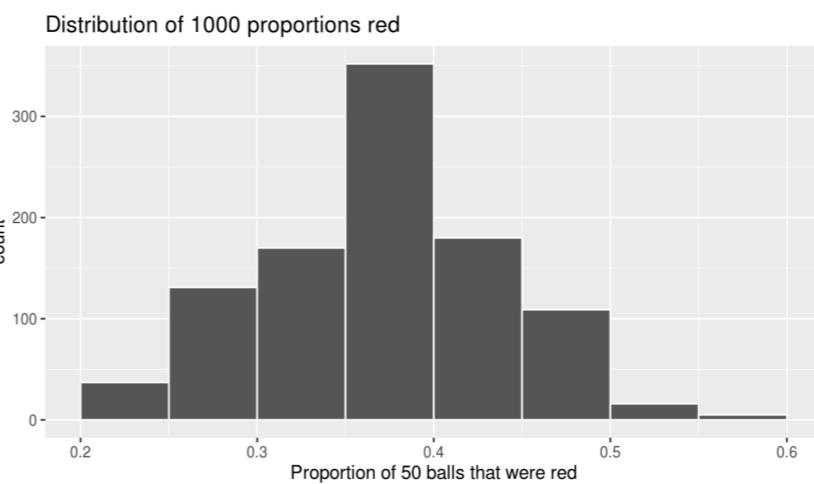


Sampling
Distributions &
Standard Errors



```
Console ~/ ↵
> library(moderndive)
> bowl
# A tibble: 2,400 x 2
  ball_ID color
  <int> <chr>
1     1 white
2     2 white
3     3 white
4     4 red
5     5 white
6     6 white
7     7 red
8     8 white
9     9 red
10    10 white
# ... with 2,390 more rows
> |
```

```
Console ~/ ↵
> bowl %>%
+   rep_sample_n(size = 50, reps = 1)
# A tibble: 50 x 3
# Groups: replicate [1]
  replicate ball_ID color
  <int> <int> <chr>
1       1     1  white
2       1     1  red
3       1     1  white
4       1     1  white
5       1     1  white
6       1     1  white
7       1     1  white
8       1     1  white
9       1     1  red
10      1     1  white
# ... with 40 more rows
> |
```



$$SE = \sqrt{\frac{p(1-p)}{n}}$$

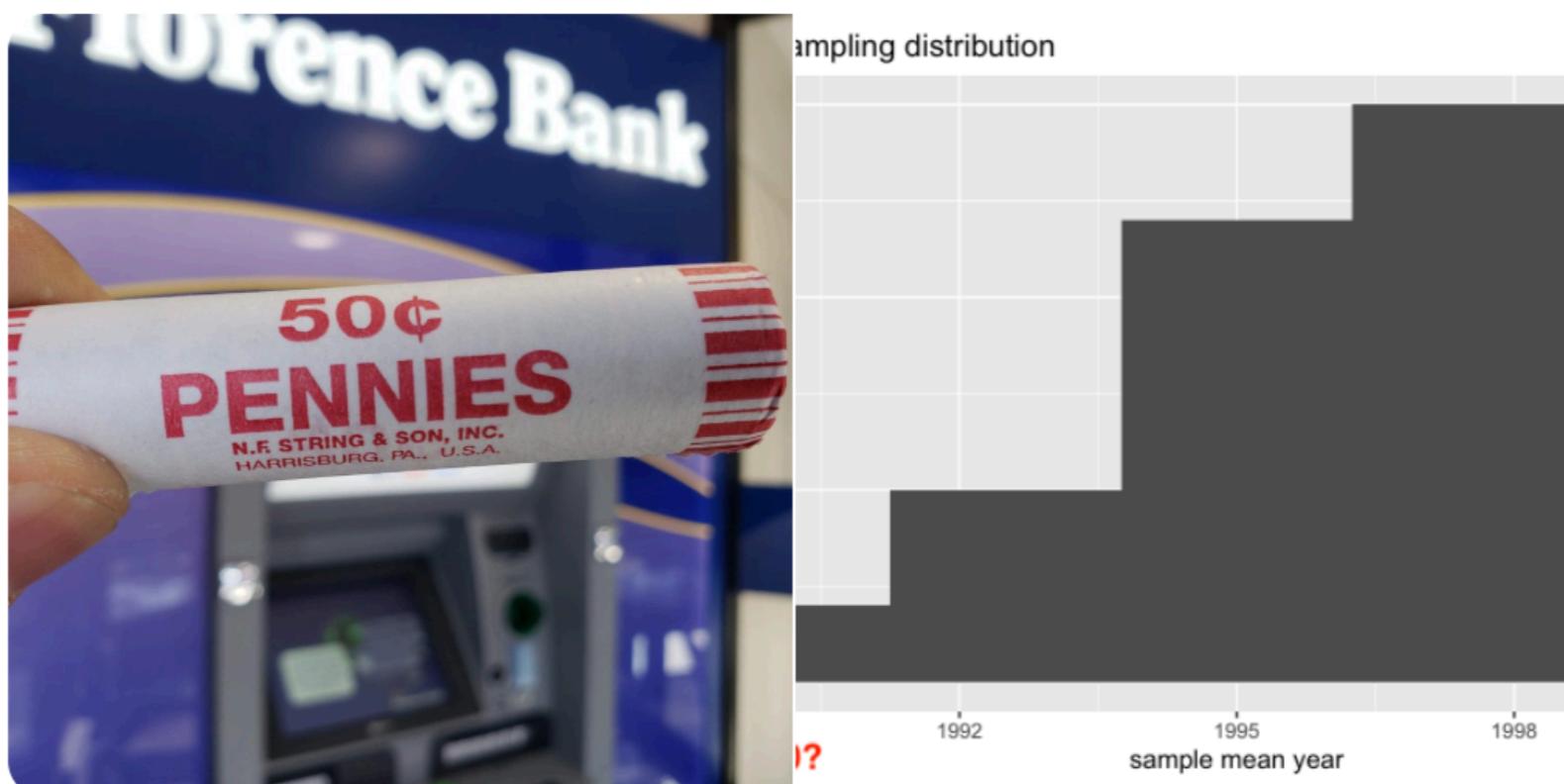
Chap 9: Confidence Intervals

Click me!!!



ModernDive @ModernDive · Mar 27

Hey intro stats profs! Do you teach statistical inference w/ the bootstrap method? Do you get Q's like "Why do we resample WITH replacement?" or "How many samples are there?" If so, consider doing "tactile resampling" first, THEN %>% do "virtual resampling" the [@moderndive way!](#)



2



7



15



[Show this thread](#)

Chap 10: Hypothesis Testing via infer

Click me!!!



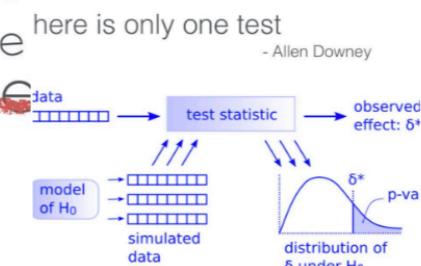
Albert Y. Kim

@rudeboybert

Replies to @AmeliaMN @djnavarro and 3 others

Indeed! Per [@crite](#): "the `infer` package makes statistical inference tidy & transparent!"
github.com/rudeboybert/JS ...

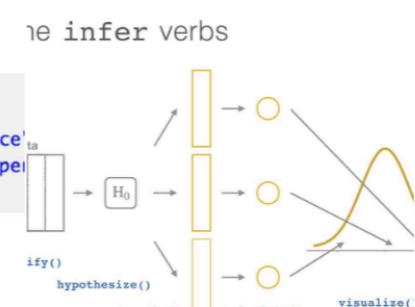
~~infer makes p-value easier to compute.~~
tidy and transparent.



statistic
inference
here is only one test
- Allen Downey

```
.test(gss$party, gss$space)
```

```
gss %>%  
  specify(space ~ party) %>%  
  hypothesize(null = "independence")  
  generate(reps = 1000, type = "perm")  
  calculate(stat = "Chisq")
```



8:39 AM - 21 May 2019

1 Retweet 9 Likes



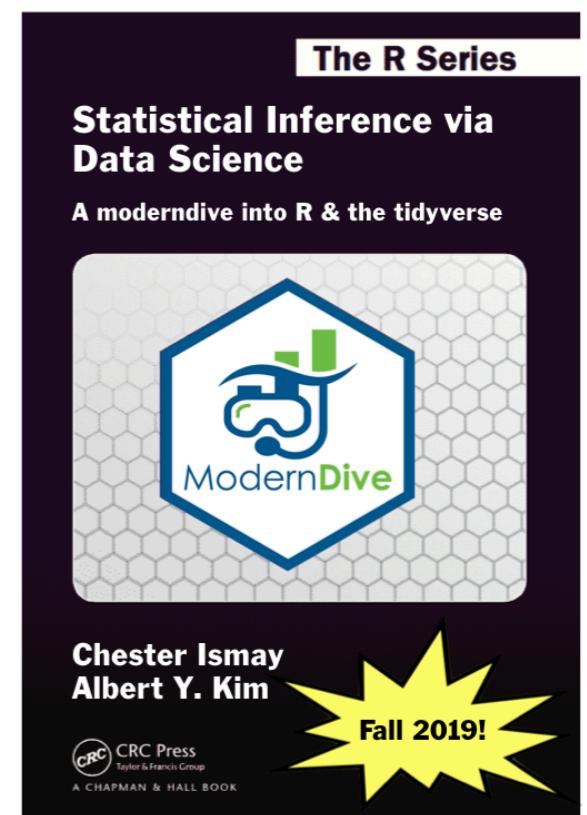
Conclusion

Timeline



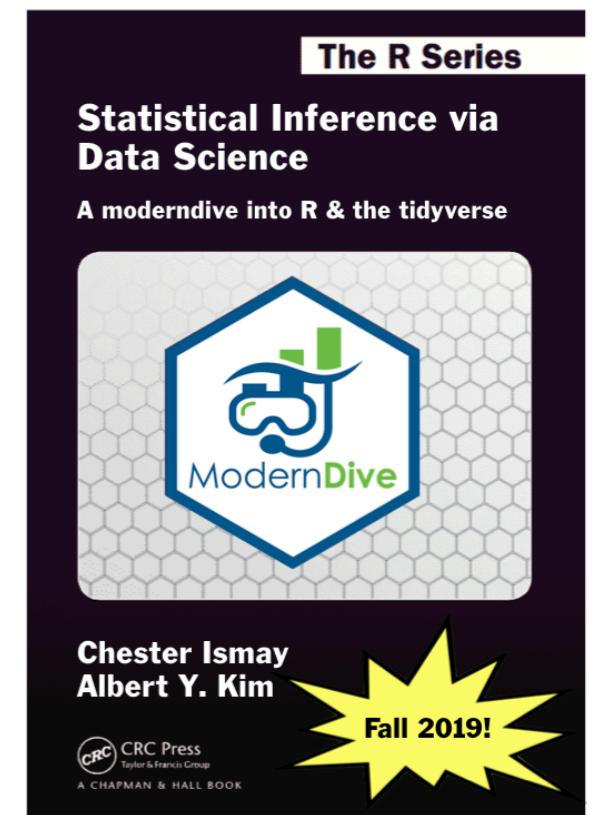
Timeline

- **Now:** Development version on moderndive.netlify.com being edited.



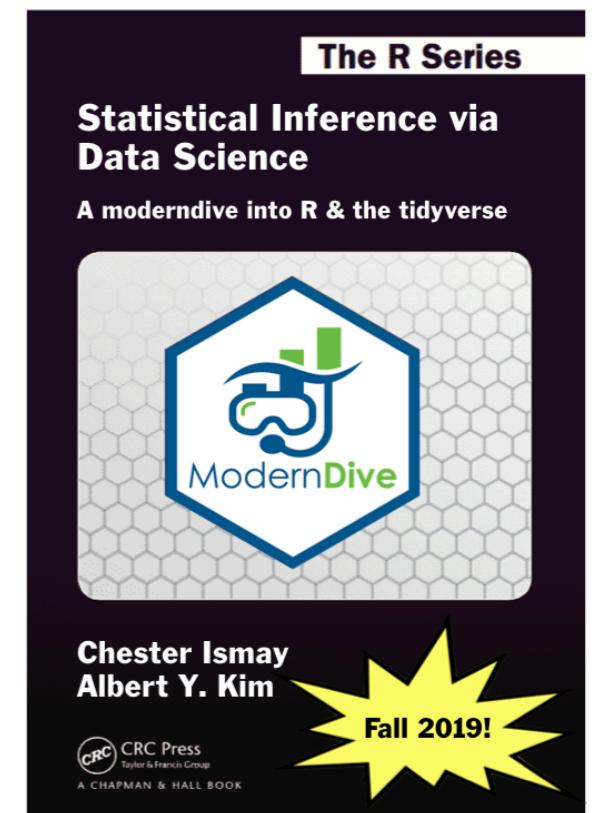
Timeline

- **Now:** Development version on moderndive.netlify.com being edited.
- 🚧 Ch9-11 on CI, HT, & inference for regression 🚧



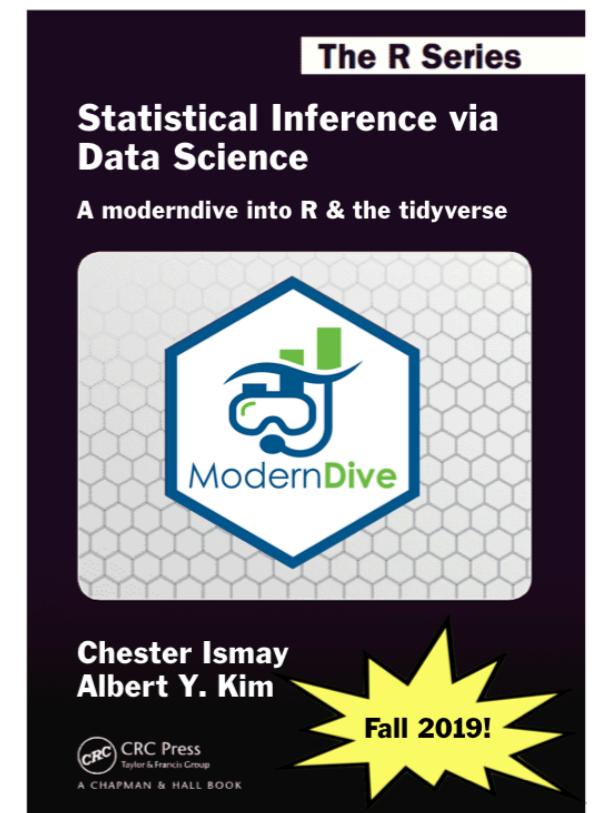
Timeline

- **Now:** Development version on moderndive.netlify.com being edited.
- 🚧 Ch9-11 on CI, HT, & inference for regression 🚧
- **Late-June:** Preview of print edition available on moderndive.com



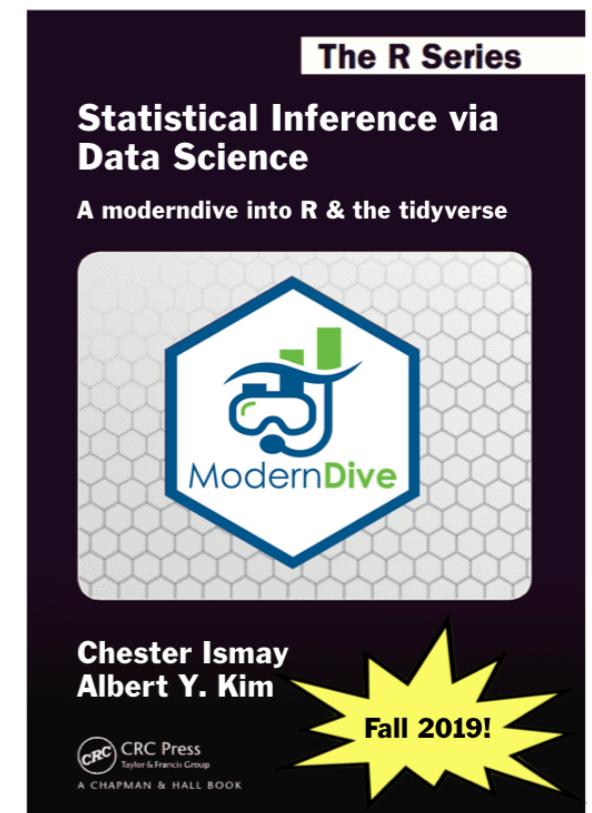
Timeline

- **Now:** Development version on moderndive.netlify.com being edited.
- 🚧 Ch9-11 on CI, HT, & inference for regression 🚧
- **Late-June:** Preview of print edition available on moderndive.com
- **Late-July:** Posting labs/problems sets & example final project samples



Timeline

- **Now:** Development version on moderndive.netlify.com being edited.
- 🚧 Ch9-11 on CI, HT, & inference for regression 🚧
- **Late-June:** Preview of print edition available on moderndive.com
- **Late-July:** Posting labs/problems sets & example final project samples
- **Fall 2019:** Print edition available!



Resources

Resources

- Two versions of moderndive

Resources

- Two versions of moderndive
 1. Development (being edited):
moderndive.netlify.com

Resources

- Two versions of moderndive
 1. Development (being edited):
moderndive.netlify.com
 2. Latest release (updated x2 yearly):
moderndive.com

Resources

- Two versions of moderndive
 1. Development (being edited):
moderndive.netlify.com
 2. Latest release (updated x2 yearly):
moderndive.com
- On GitHub at github.com/moderndive/

Resources

- Two versions of moderndive
 1. Development (being edited):
moderndive.netlify.com
 2. Latest release (updated x2 yearly):
moderndive.com
- On GitHub at github.com/moderndive/
 1. bookdown source code for book

Resources

- Two versions of moderndive
 1. Development (being edited):
moderndive.netlify.com
 2. Latest release (updated x2 yearly):
moderndive.com
- On GitHub at github.com/moderndive/
 1. **bookdown** source code for book
 2. **moderndive** package source code

Resources

- Two versions of moderndive
 1. Development (being edited):
moderndive.netlify.com
 2. Latest release (updated x2 yearly):
moderndive.com
- On GitHub at github.com/moderndive/
 1. bookdown source code for book
 2. **moderndive** package source code
- Course [webpage](#) from Spring 2019

Resources

- Two versions of moderndive
 1. Development (being edited):
moderndive.netlify.com
 2. Latest release (updated x2 yearly):
moderndive.com
- On GitHub at github.com/moderndive/
 1. bookdown source code for book
 2. **moderndive** package source code
- Course [webpage](#) from Spring 2019
- moderndive mailing list: eepurl.com/cBkItf

Thank you!

Why tidyverse in general?

Why tidyverse in general?

From [tidy tools manifesto](#):

Why tidyverse in general?

From [tidy tools manifesto](#):

1. Reuse existing data structures

Why tidyverse in general?

From [tidy tools manifesto](#):

1. Reuse existing data structures
2. Compose simple functions with the pipe

Why tidyverse in general?

From [tidy tools manifesto](#):

1. Reuse existing data structures
2. Compose simple functions with the pipe
3. Embrace functional programming

Why tidyverse in general?

From [tidy tools manifesto](#):

1. Reuse existing data structures
2. Compose simple functions with the pipe
3. Embrace functional programming
4. Design for humans

Why tidyverse in general?

From [tidy tools manifesto](#): Say what?

1. Reuse existing data structures
2. Compose simple functions with the pipe
3. Embrace functional programming
4. Design for humans

Why tidyverse in general?

From [tidy tools manifesto](#): Say what?

- 1. Reuse existing data structures
 - 2. Compose simple functions with the pipe
 - 3. Embrace functional programming
 - 4. Design for humans
-
- 1. Don't reinvent the wheel!

Why tidyverse in general?

From [tidy tools manifesto](#): Say what?

- 1. Reuse existing data structures
 - 2. Compose simple functions with the pipe
 - 3. Embrace functional programming
 - 4. Design for humans
-
- 1. Don't reinvent the wheel!
 - 2. Breakdown large tasks into steps using `%>%` "then"

Why tidyverse in general?

From [tidy tools manifesto](#): Say what?

- 1. Reuse existing data structures
 - 2. Compose simple functions with the pipe
 - 3. Embrace functional programming
 - 4. Design for humans
- 1. Don't reinvent the wheel!
 - 2. Breakdown large tasks into steps using `%>%` "then"
 - 3. What is the [goal](#) of your code?

Why tidyverse in general?

From [tidy tools manifesto](#): Say what?

- 1. Reuse existing data structures
 - 2. Compose simple functions with the pipe
 - 3. Embrace functional programming
 - 4. Design for humans
-
- 1. Don't reinvent the wheel!
 - 2. Breakdown large tasks into steps using `%>%` "then"
 - 3. What is the [goal](#) of your code?
 - 4. Make code understandable to humans

Also model selection!

Also model selection!

- Is there a way to teach
- ✓ model selection
 - ✓ model complexity vs parsimony
 - ✓ occam's razor

To intro stats students? 

YES! Via data viz  & EDA !

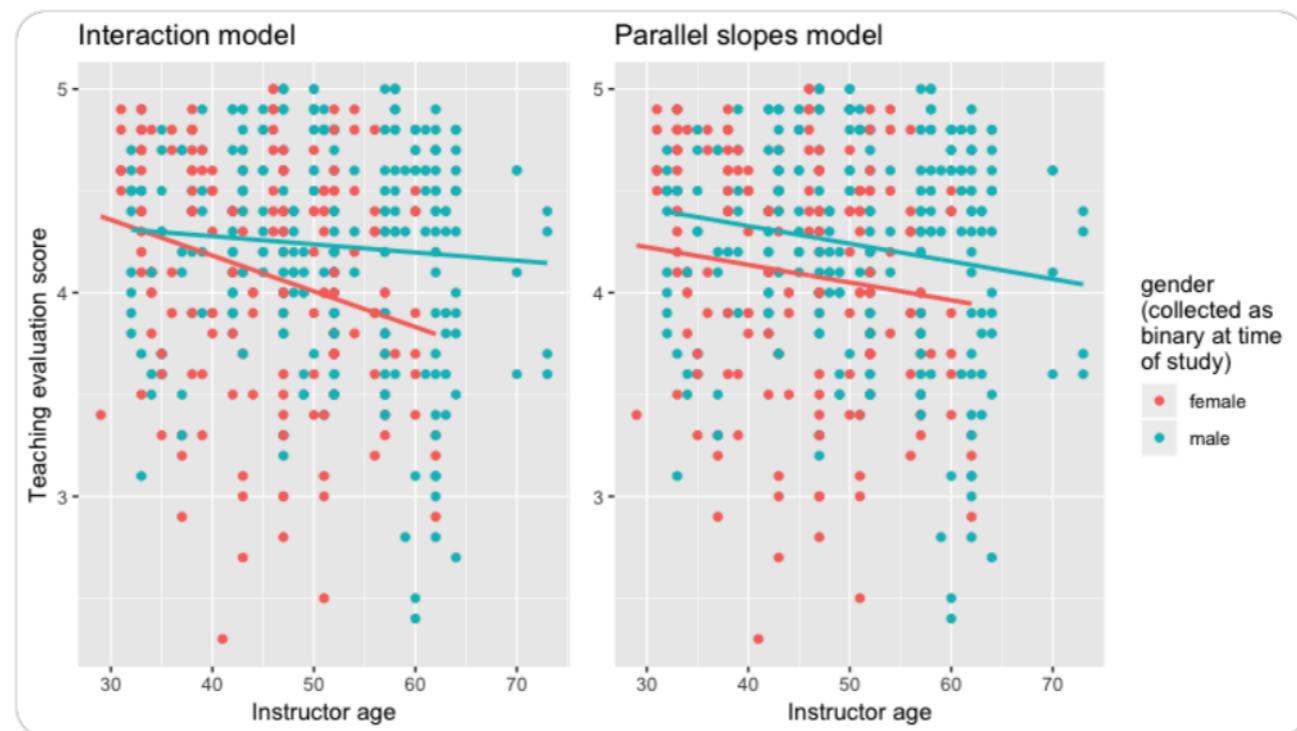
Also model selection!

- Is there a way to teach
- ✓ model selection
 - ✓ model complexity vs parsimony
 - ✓ occam's razor

To intro stats students? 

YES! Via data viz  & EDA !

First show a case study where
"interaction model" >>> "parallel slopes
model"! 1/4

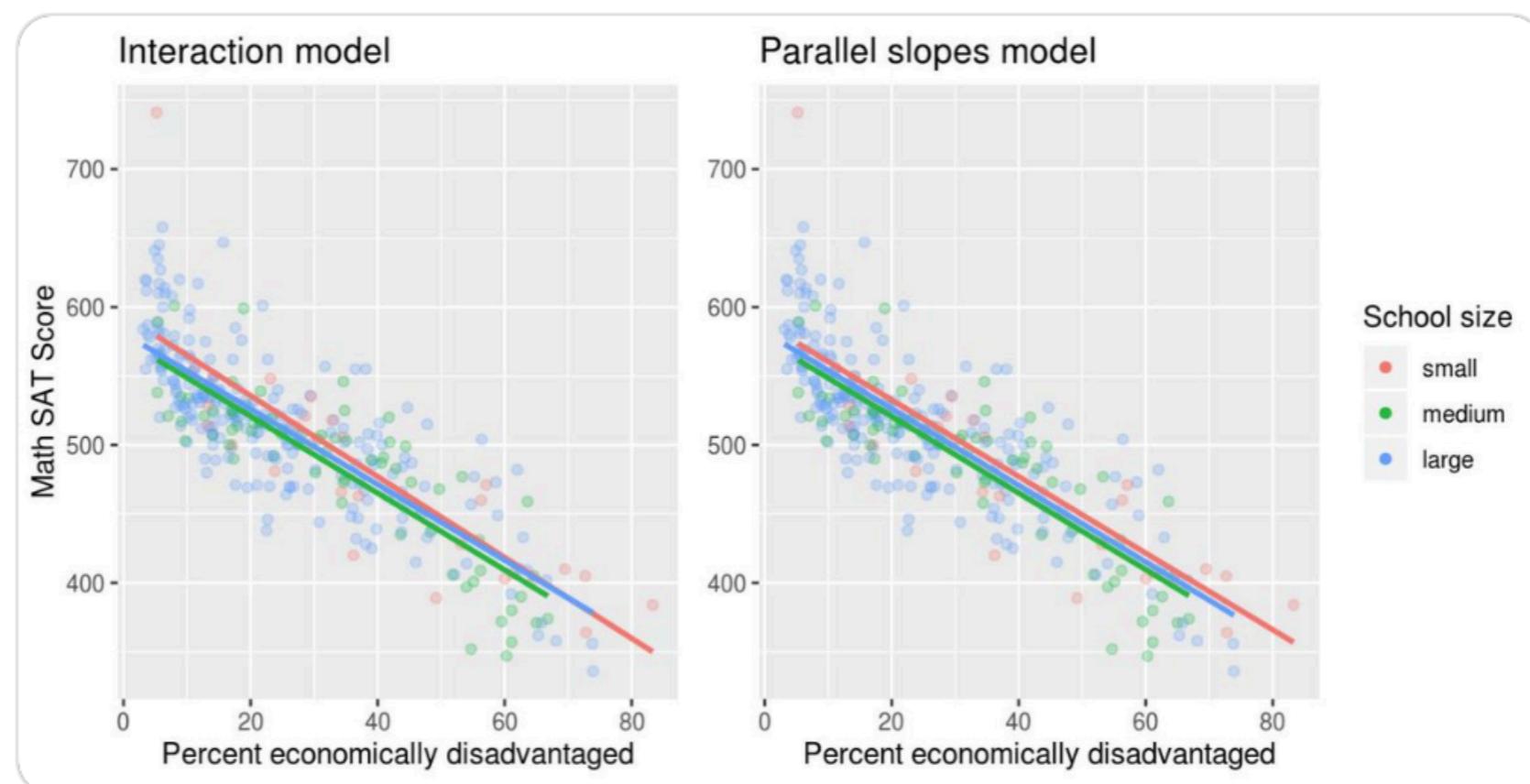


More model selection



ModernDive @ModernDive · Apr 19

Next show a case study where "interaction model" vs "parallel slopes model" is
"I dunno?!? They look kinda the same to me?!?" 2/4



1

1

1

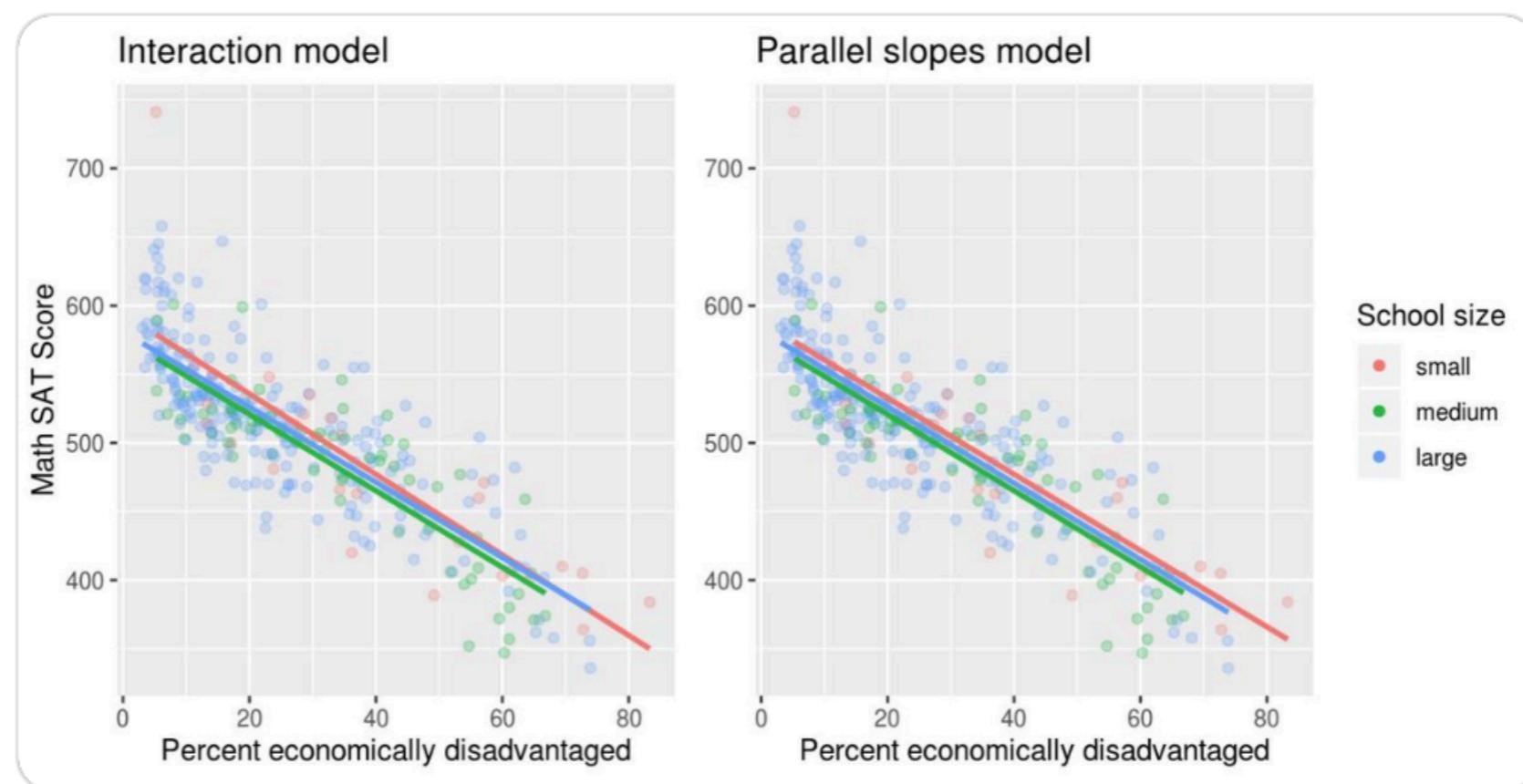
1

More model selection



ModernDive @ModernDive · Apr 19

Next show a case study where "interaction model" vs "parallel slopes model" is "I dunno?!? They look kinda the same to me?!?" 2/4



1

1

1

1



ModernDive @ModernDive · Apr 19

Then ask them to consider:

"Is the extra 😬 COMPLEXITY 😬 of the interaction model warranted?"

or

"Should I favor the 😊 SIMPLER 😊 parallel slopes model?" 3/4