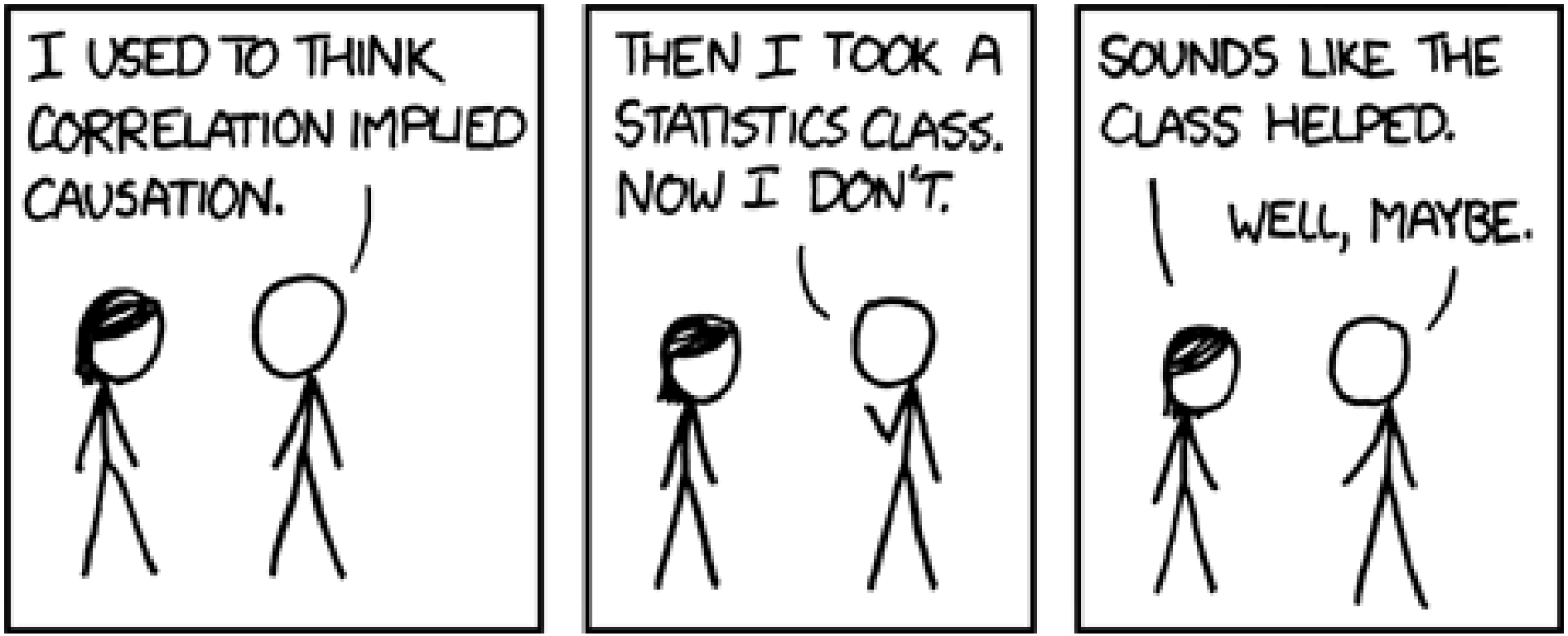
431 Class 01

Thomas E. Love

2018-08-28

This is PQHS 431 / CRSP 431 / MPHP 431



Please take a copy of (1) the **survey** and (2) an **index card** from the teaching assistants.

Wait for further instructions before writing anything down.

Instructions for the Survey

Please read these instructions carefully before writing anything down.

1. Introduce yourself to someone that you don't know.
2. Record the survey answers **for that other person**, while they record your responses.
3. Be sure to complete all 15 questions (both sides of the paper).
4. Also, write **YOUR** answer to question #4 on the index card, and keep that, for now. You'll need it later.
5. When you are finished, thank your partner and raise your hand. Someone will come to collect your survey.

Regarding Question 4, Professor Love is the large fellow standing in the front of the room.

Course Details

Instructor: Thomas E. Love, Ph.D.

Email (best way to reach me): Thomas.Love@case.edu

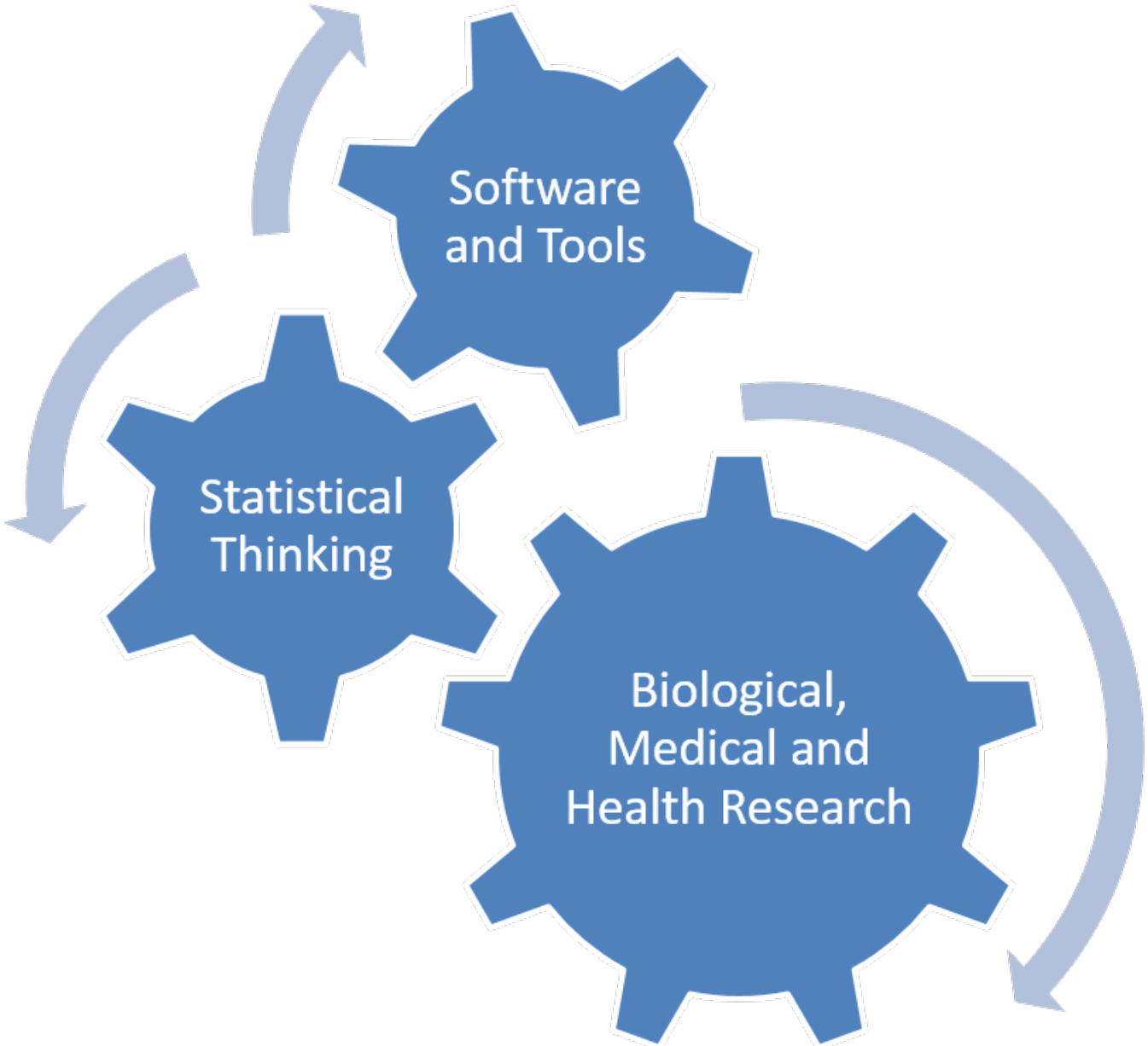
Our web site: https://github.com/THOMASELOVE/431-2018

Links there to:

- Course Syllabus
- Course Calendar
- Course Notes (essentially a textbook)
- Slides Page (in-class materials, slides as PDF and R Markdown)
- Software Details (R and R Studio, installation and R Basics)
- Data and Code
- Assignments (8), Quizzes (3) and Project (after Labor Day)

How to Get Help: 431-help@case.edu

What is this course about?



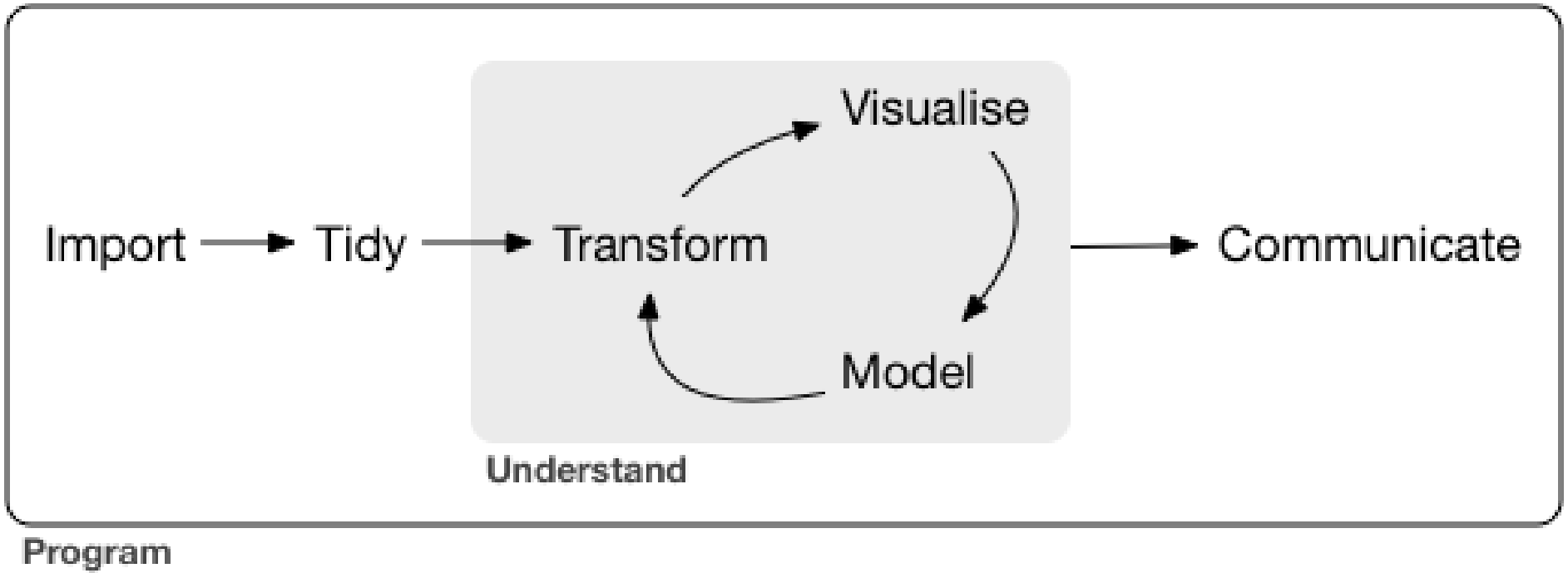
What is this course about?

a. Exploratory Data Analysis, Visualization
b. Statistical Inference, Making Comparisons
c. Linear Regression and related Models

The course is about biostatistics, replicable research, using state-of-the-art tools (R, R Studio, R Markdown), and thinking about how science is most effectively done.

- It is more a course in **how** to do things (highly applied) rather than a theoretical/mathematical justification for **why** we do them. We focus here on practical work.
- It's mostly about getting you doing data science projects for biological, medical and health applications.

What is Data Science about?



Source: http://r4ds.had.co.nz/introduction.html

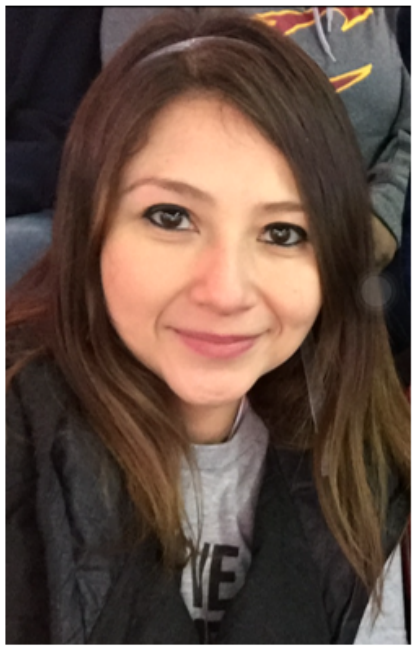
Teaching Assistants (email 431-help@case.edu)

431: 2018 Teaching Assistants

431-help@case.edu

Robert (Bob) Winkelman
Lead TA, Section 1

Satyakam Mishra
Lead TA, Section 2



Claudia Cabrera

Maher Kazimi

- All TAs work with PQHS/CRSP/MPHP students and both Sections.
- TA **office hours** start next week, with times/locations to come.

To get help at any time starting now, email 431-help@case.edu

What will we be reading?

Dr .Love's Course Notes

Data Science for Biological, Medical and Health Research: Notes for 431

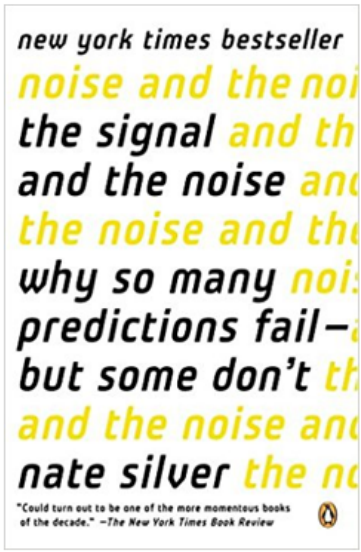
Thomas E. Love, Ph.D.

Version: 2018-08-25 14:00:06

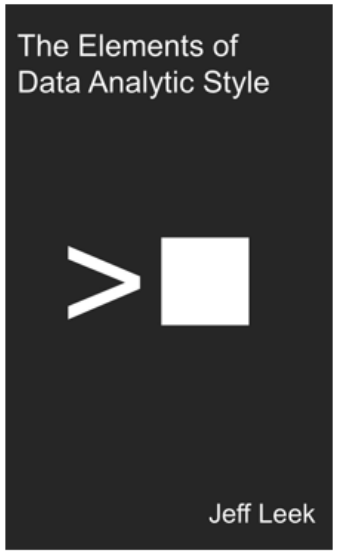
Introduction

These Notes provide a series of examples using R to work through issues that are likely to come up in PQHS/CRSP/MPHP 431.

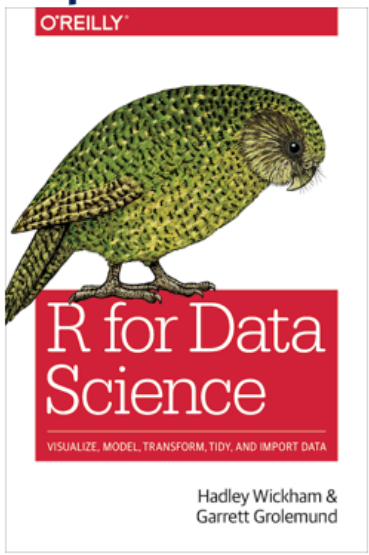
Silver



Leek



R4DS



Our web site: https://github.com/THOMASELOVE/431-2018

Great Statisticians in History

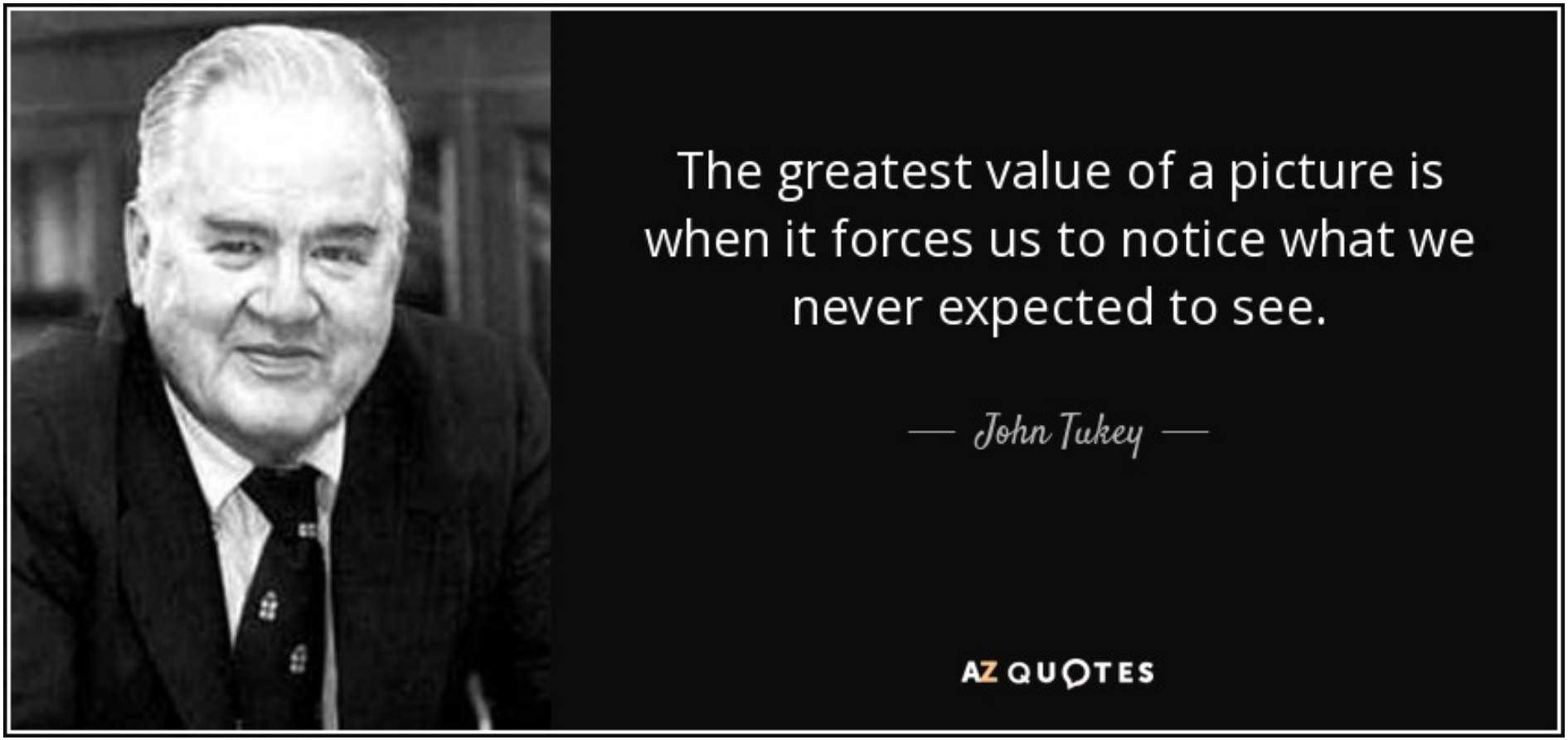


Photo Source: http://www.azquotes.com/author/14847-John_Tukey

John **Tukey** (1915-2000)

Your First Task (Do Today, Please)

Visit http://bit.ly/431-2018-names and give me your preferred name and email.

	A	B	C	D	E	F
1	If you don't find your name on this sheet, please add it to the bottom and contact Dr. Love via email.					
2	*Order*	Registered	CWRU Email	Name	Your Preferred Email	Preferred Name (Please call me...)
3	0	Teaching	Thomas.Love@case.edu	Love, Thomas	Thomas.Love@case.edu	Dr. Love, Professor Love or Tom
4	1	PQHS 431	axa1059@case.edu	Aguilar, Alicia		
5	2	CRSP 431	iob3@case.edu	Babatunde, Ife		
6	3	PQHS 431	wpb27@case.edu	Bensken, Wyatt		
7	4	CRSP 431	amb388@case.edu	Brown, Adam		

You'll need to **log in to Google via CWRU** to see the form.

Gathering Some Data: Age Guessing Activity

- You will join one of ten groups, with 4-5 students in each group.
- Your group will receive a sheet to keep track of your guesses (estimated ages.)
- Your group will then receive one of a series of cards, with a photo of a person on it.
- For each card your group receives. . .
  - estimate the age of the person on the card
  - write your (group) guess in the table on the sheet in the row corresponding to that numbered card
- Later, you will be told the true ages and will be able to compute errors.

Scientists Gather Their Own Data

If you have a little time between cards, make sure everyone in your group . . .

1. knows the name and field of everyone else in the group, and knows your group's letter.
2. writes down a new guess as to my age on their index card, now that you know me better.

So if your initial guess was that I was 18, but now you think I'm 19, your card should read 18/19.

Age Guessing Robots?

Well, yes, of course, there's a tool online to do this. More than one, in fact.

Visit https://how-old.net/

https://how-old.net

The AI's guess was... 7 years too high

6 years too low

Do you think you did that well?

Card 1

Card 1

Eric Chong
Master Chef Canada winner
Photo date: April 2014

Age 21

Card 2

Card 2

Katherine Archuleta
Former U.S. OPM Director
Photo date: 2013

Age 64

Card 3

Card 3

Elise Mayfield
Chef, Actor, Baker
Photo date: 2014

Age 28

Card 4

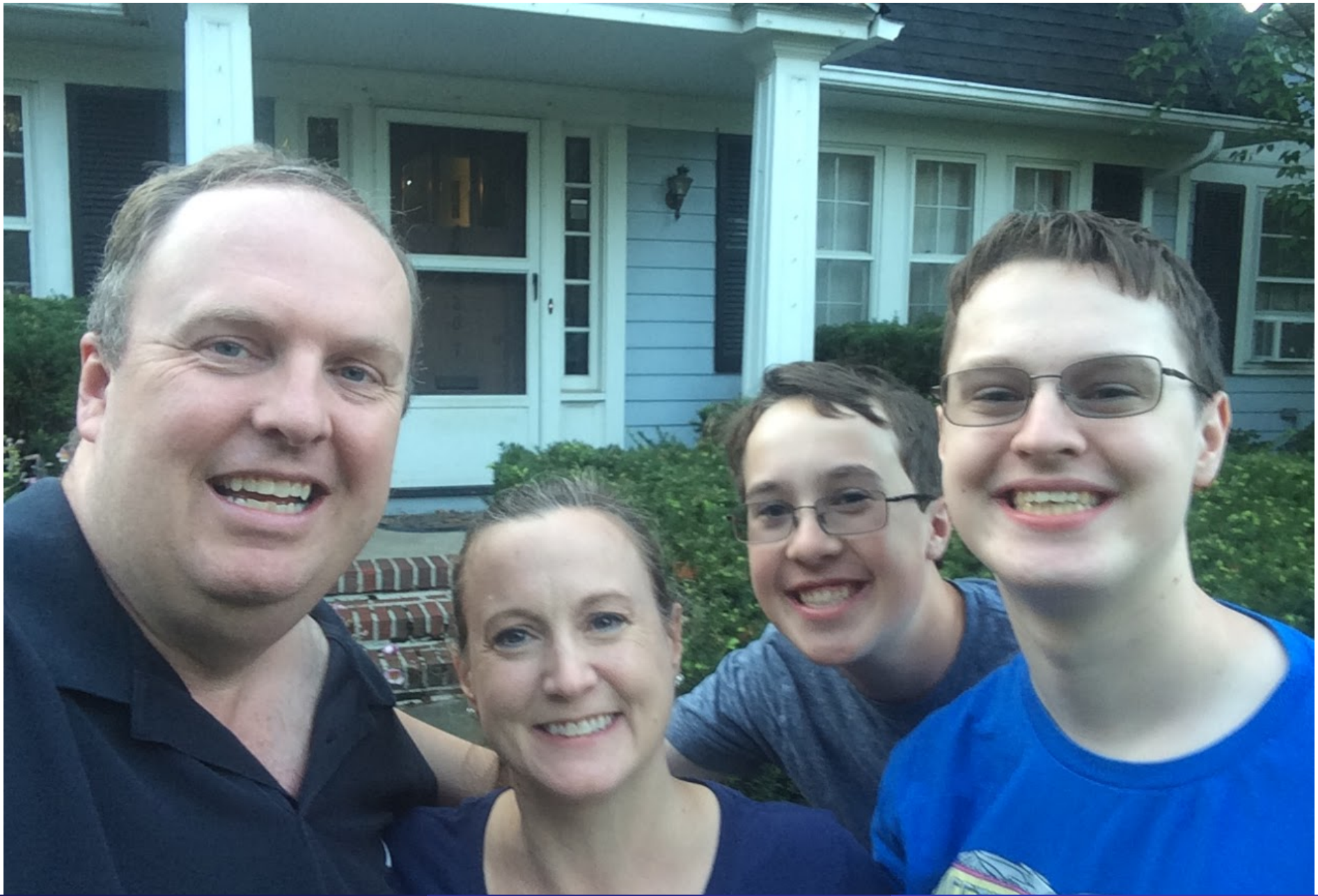
Card 4

Kevin Love
(then) High School Student
Photo date: June 2014

Age 14

No, not THAT Kevin Love

THIS Kevin Love, on the right here in 2017



Card 5

Card 5

Rosemary McGinn

Photo date: July 2013

Age 54

Card 6

Card 6

John Chaney
Basketball Coach
Photo date: 2006

Age 74

Card 7

Card 7

David Storm

Photo date: August 2014

Age 44

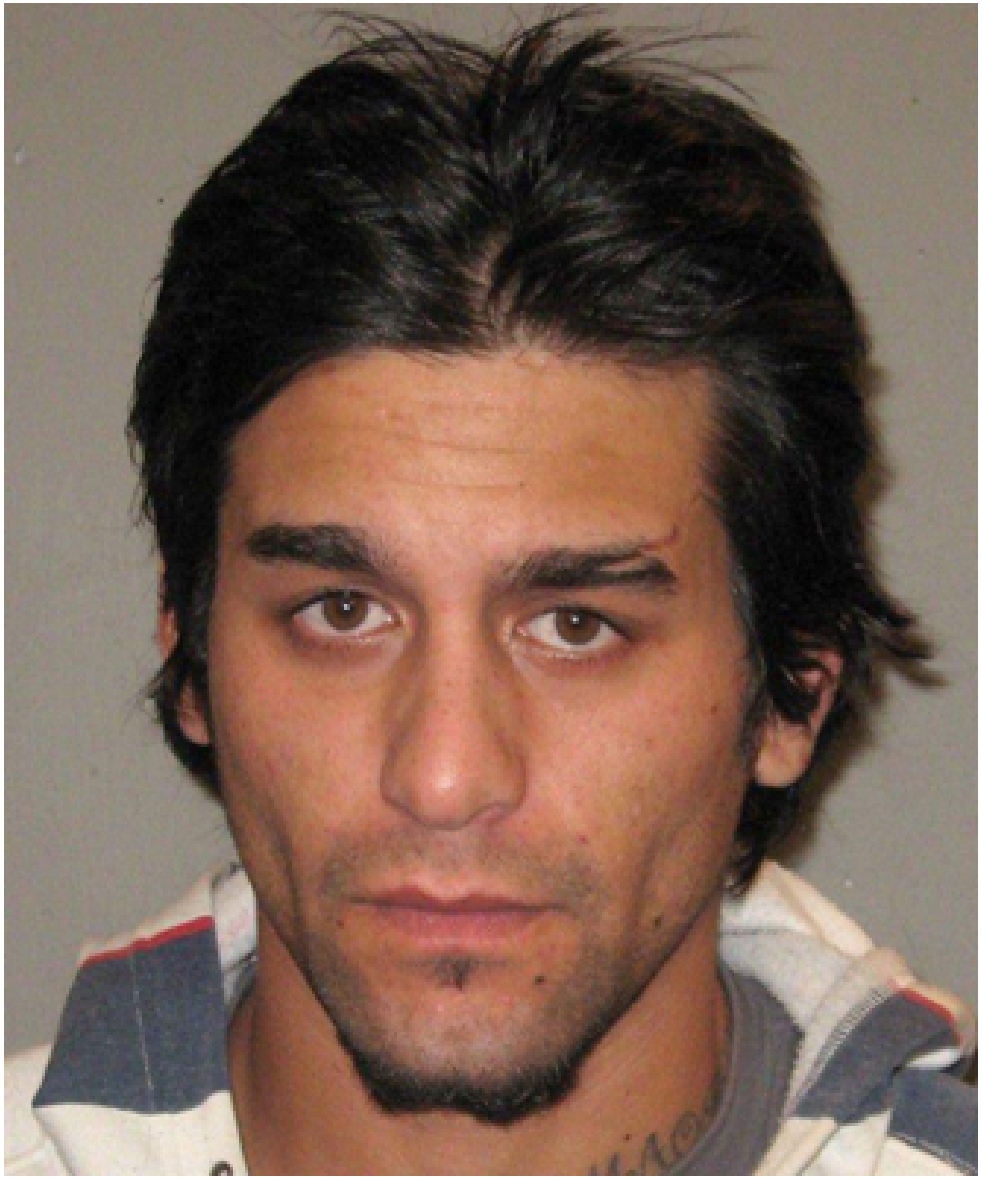
Card 8

Card 8

Margo Glantz
Writer
Photo date: 2013

Age 83

Card 9



Card 9

Quade Ross Honey
Fugitive
Photo date: 2012

Age 24

Card 10

Card 10

Bianca Lawson
Actress
Photo date: 2013

Age 34

So, How did we do?

#1 Age **21** #2 Age **64** #3 Age **28** #4 Age **14** #5 Age **54**

#6 Age **74** #7 Age **44** #8 Age **83** #9 Age **24** #10 Age **34**

Collecting the Results

We'll collect some key results in a Google sheet, that you should be able to reach when logged into CWRU for Google.

The sheet is at http://bit.ly/431-2018-day1-ageguess

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	group	group size	females	card 1	card 2	card 3	card 4	card 5	card 6	card 7	card 8	card 9	card 10
2	A												
3	B												
4	C												
5	D												
6	E												
7	F												
8	G												
9	H												
10	I												
11	J												

And how did the AI at https://how-old.net do?





#1 Age **21**	#2 Age **64**	#3 Age **28**	#4 Age **14**	#5 Age **54**
AI guess 27	**AI 44**	**AI 22**	**AI 19**	**AI 36**
#6 Age **74**	#7 Age **44**	#8 Age **83**	#9 Age **24**	#10 Age **34**
AI 63	**AI 55**	**AI 79**	**AI 35**	**AI 28**

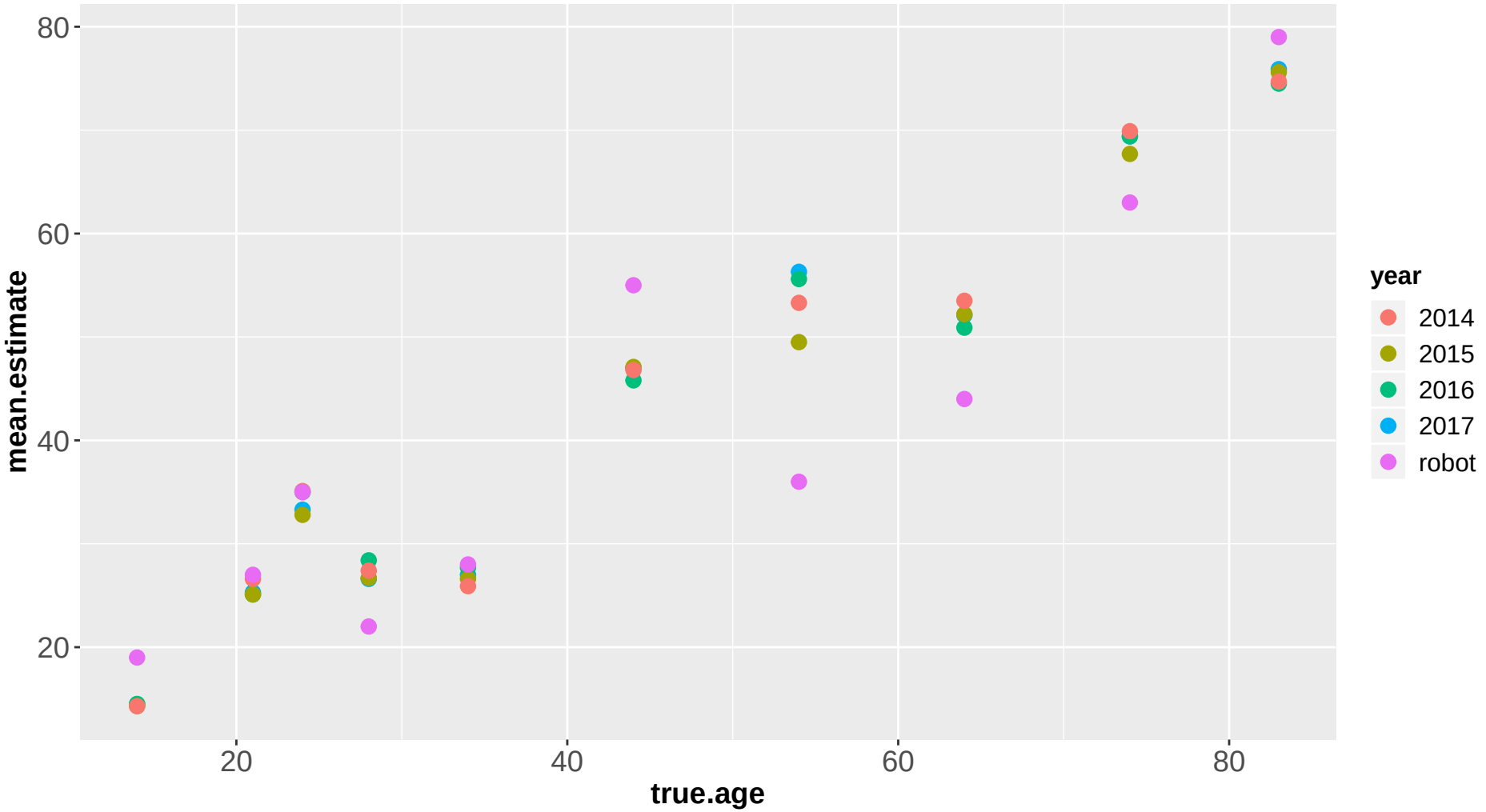




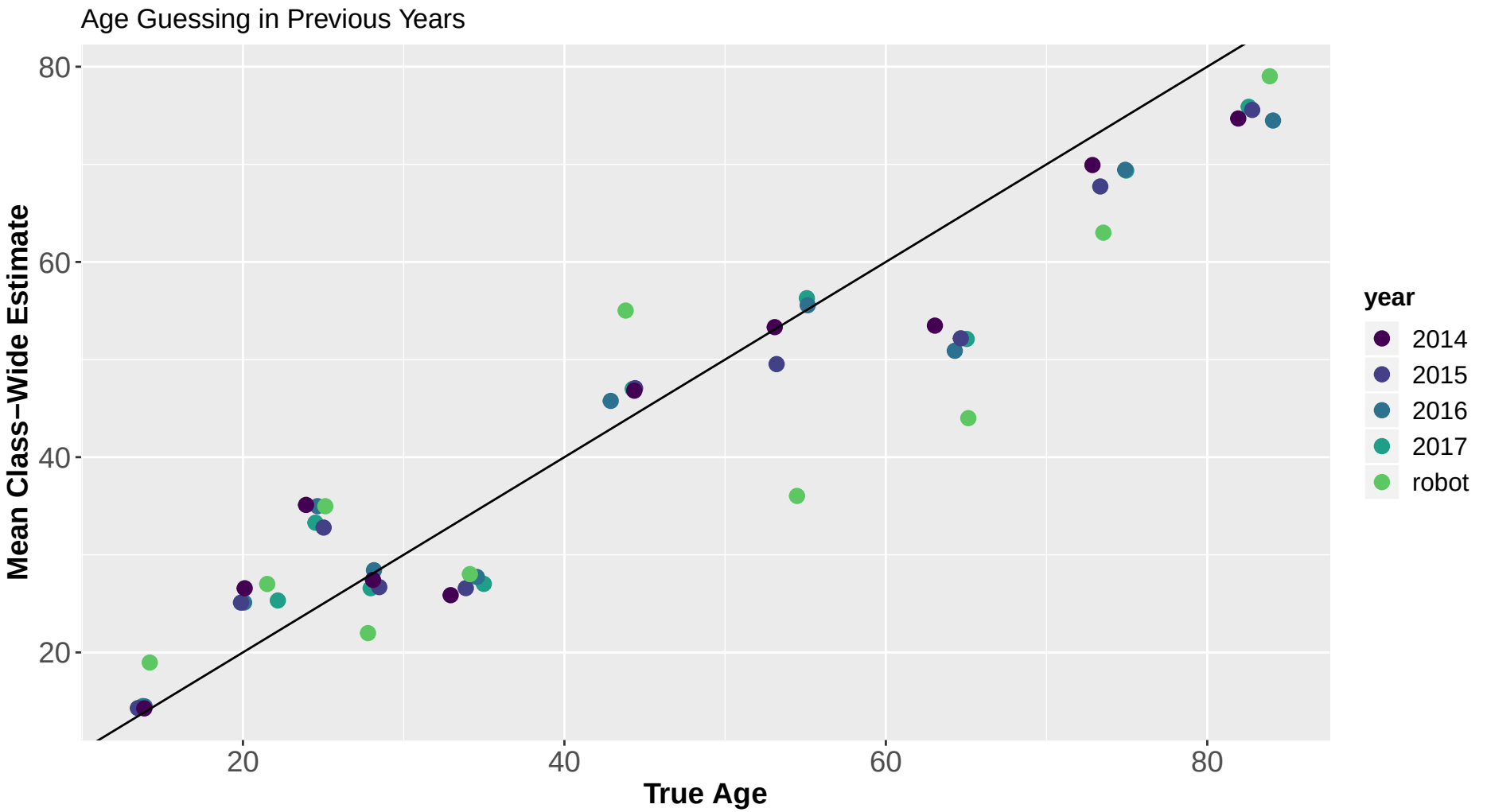
ageguesshistory.csv Data Set (excerpt)

card	label	true.age	sex	facing	year	mean.estimate	error
1	Chong	21	M	R	2017	25.3	4.3
2	Archuleta	64	F	L	2017	52.1	-11.9
3	Mayfield	28	F	L	2017	26.6	-1.4
4	Love	14	M	L	2017	14.5	0.5
5	McGinn	54	F	R	2017	56.3	2.3
6	Chaney	74	M	L	2017	69.4	-4.6
7	Storm	44	M	R	2017	47.0	3.0
8	Glantz	83	F	L	2017	75.9	-7.1
9	Honey	24	M	L	2017	33.3	9.3
10	Lawson	34	F	R	2017	27.0	-7.0
1	Chong	21	M	R	2016	25.1	4.1
2	Archuleta	64	F	L	2016	50.9	-13.1

Scatterplot of Prior Results, 1



Scatterplot of Prior Results, 2



Mean Class-Wide Guesses (2014-17 combined)

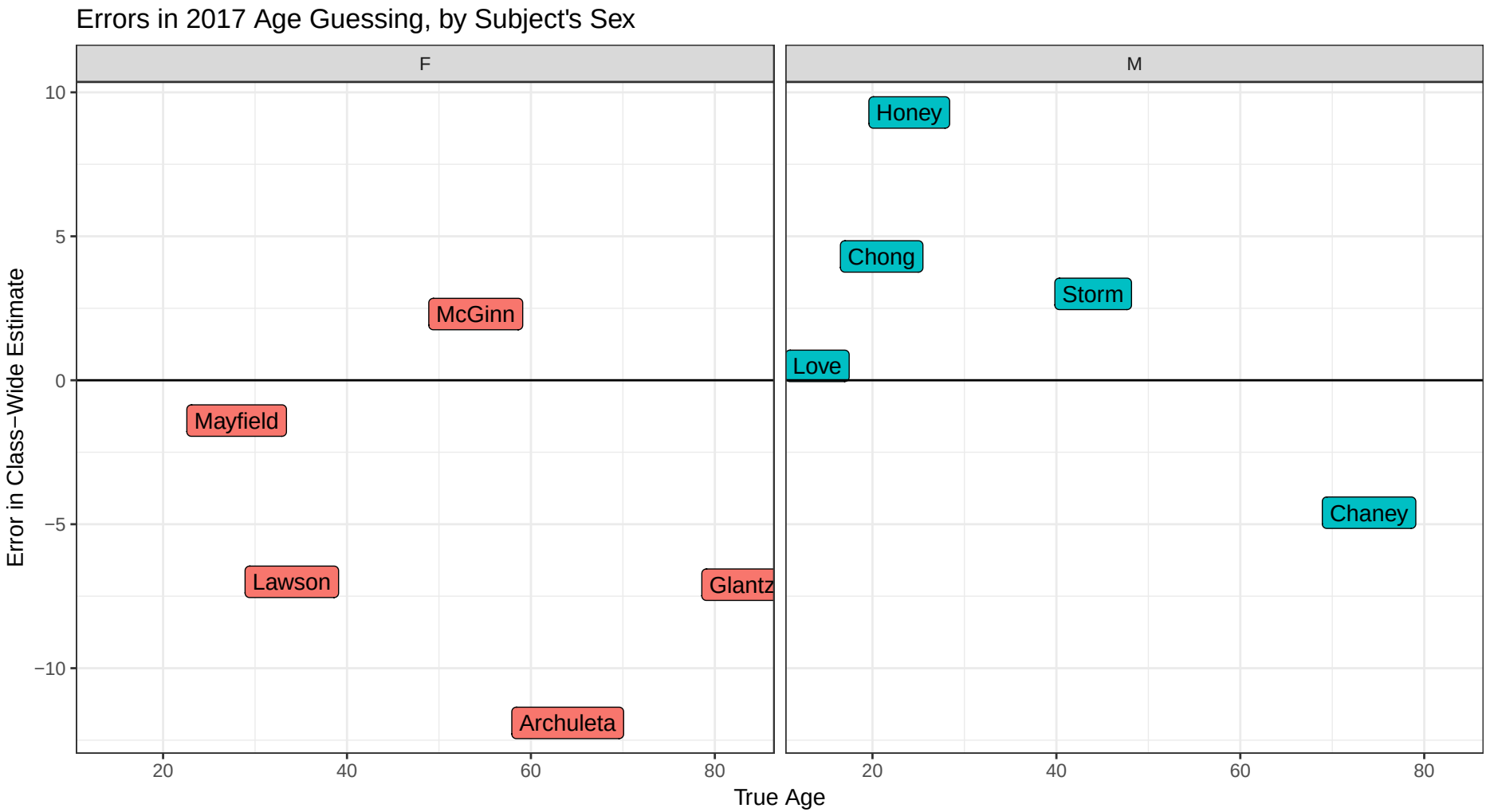
	#1 Age **21**	#2 Age **64**	#3 Age **28**	#4 Age **14**	#5 Age **54**
2014-17	25.5	52.2	27.3	14.4	53.7
Mean Guesses	69.1	46.7	75.2	34.1	26.8

#6 Age **74**	#7 Age **44**	#8 Age **83**	#9 Age **24**	#10 Age **34**

Mean Class-Wide Errors (2014-17 combined)

	#1 Age **21**	#2 Age **64**	#3 Age **28**	#4 Age **14**	#5 Age **54**
2014-17 Errors	+4.5	-11.8	-0.7	+0.4	-0.3
	-4.9	+2.7	-7.8	+10.1	-7.2
	#6 Age **74**	#7 Age **44**	#8 Age **83**	#9 Age **24**	#10 Age **34**

Scatterplot of 2017 Results with Labels



Hans Rosling and "The Joy of Stats"

200 countries over 200 years using 120,000 numbers, in about 4 minutes.

http://bit.ly/431-rosling

And if you liked that . . .

- The 20 minute version (from 2007): https://www.youtube.com/watch?v=RUwS1uAdUcI
- The full documentary from the BBC: https://www.gapminder.org/videos/the-joy-of-stats/
- Video playlist from Gapminder: https://www.gapminder.org/videos/

What's next?

RStudio

RStudio makes R easier to use. It includes a code editor, debugging & visualization tools.

R Packages

R Markdown

from RStudio

Analyze. Share. Reproduce.

Your data tells a story. Tell it with R Markdown. Turn your analyses into high quality documents, reports, presentations and dashboards.

What's next?

0. Turn in your index card with your two guesses of my age, please.
1. Visit http://bit.ly/431-2018-names : provide preferred name, email.
2. Follow the software instructions to get R, R Studio, R Packages and 431 Data on your computer.
3. Obtain Jeff Leek's The Elements of Data Analytic Style.
4. Obtain Nate Silver's The Signal and the Noise.
5. Read the syllabus and look at the rest of the website. Make sure you view the Course Notes.
6. Ask us questions. TA office hours start next week, but email is available now.

- Course Web Site: https://github.com/THOMASELOVE/431-2018
- Want help? Email `431-help@case.edu`