

431 Class 01

Thomas E. Love

2017-08-29

This is PQHS 431 / CRSP 431 / MPH 431

I USED TO THINK
CORRELATION IMPLIED
CAUSATION.



THEN I TOOK A
STATISTICS CLASS.
NOW I DON'T.



SOUNDS LIKE THE
CLASS HELPED.
WELL, MAYBE.



Please take a copy of the survey and 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

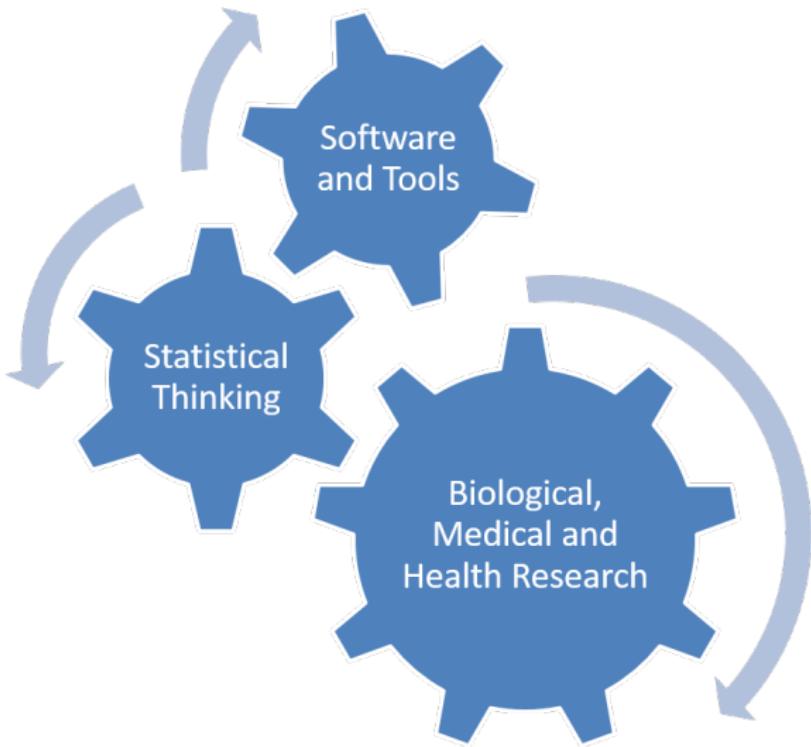
Our web site: <https://github.com/THOMASELOVE/431>

Links there to:

- Syllabus (like everything, in progress)
- Presentations (slides as PDF and R Markdown)
- Class Notes (essentially a textbook - Part A available now)
- Software Details (R and R Studio, primarily, installation and R Basics)
- Data and Code
- Assignments (there are 7), Quizzes (there are 3) and Project (after Labor Day)

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

What is this course about?



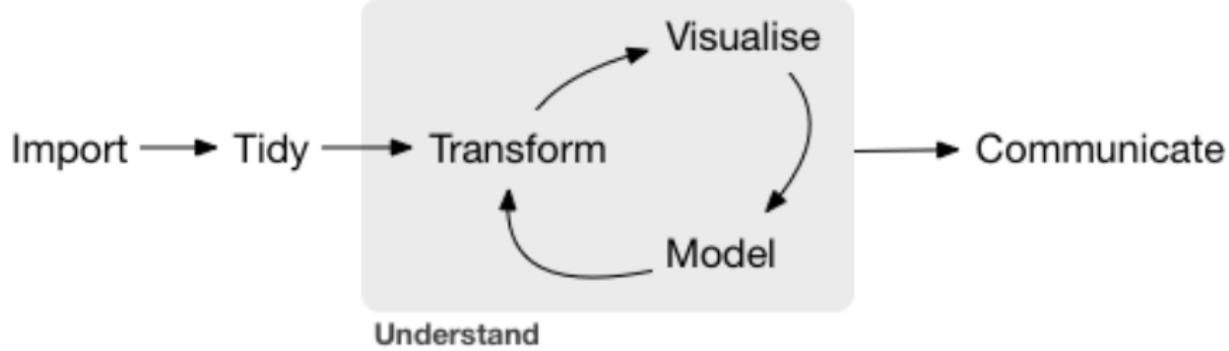
What is this course about?

- ① Exploratory Data Analysis, Visualization
- ② Statistical Inference, Making Comparisons
- ③ 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?



Program

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

Teaching Assistants ([email 431-help@case.edu](mailto:431-help@case.edu))

431: 2017 Teaching Assistants

431-help@case.edu



Claudia Cabrera, MD
Lead TA, Section 1



Omar Alaber, MD
Lead TA, Section 2



Kara Quaid



Ruzica (Rosie) Conic, MD



Robert (Bob) Winkelman

- All TAs work with PQHS/CRSP/MPHP students and both Sections.
- TA **office hours** start on Tuesday 2017-09-05 in Wood WG-56 (Computing Lab). Details to come.

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

What will we be reading?

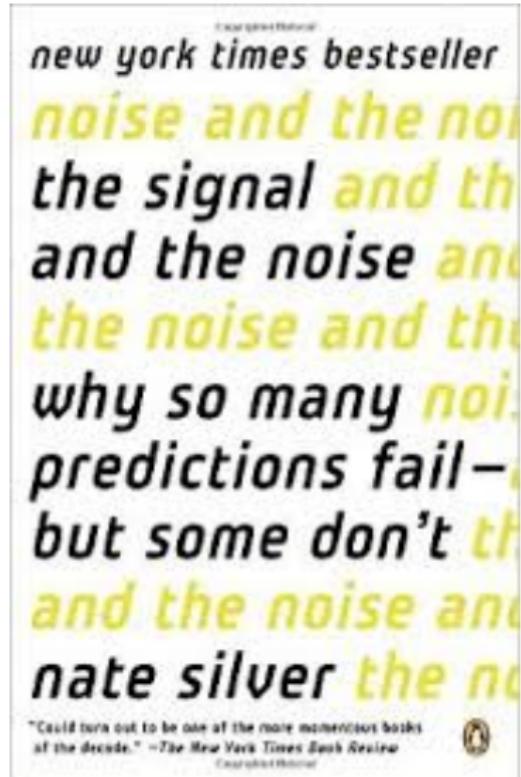
Details in the Syllabus.

Dr. Love's Class Notes

- Part A is more than 200 pages (in PDF) as it is.
- There's a Part B and Part C coming, too, before the semester is over.

Our web site: <https://github.com/THOMASELOVE/431>

What will we be BUYING and reading? (1 of 2)



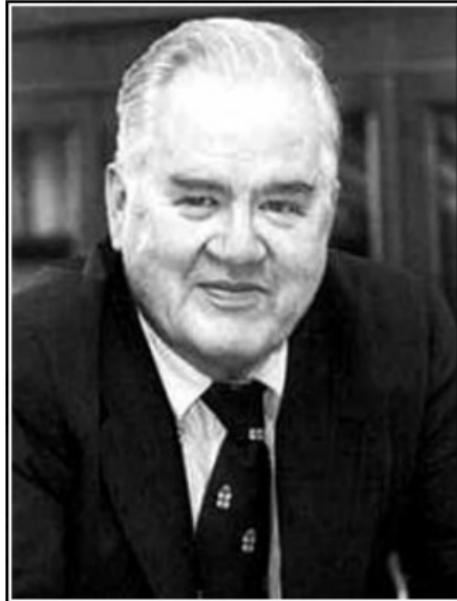
(at [Amazon](#) in various forms)

What will we be BUYING and reading? (2 of 2)



(e-book at [LeanPub](#) or for Kindle at [Amazon](#))

Great Statisticians in History



The greatest value of a picture is
when it forces us to notice what we
never expected to see.

— *John Tukey* —

AZ QUOTES

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

John **Tukey** (1915-2000)

Your First Task (Do Today, Please)

Visit <https://goo.gl/QTSGYu> and give me your preferred name and email.

	A	B	C	D	E	F
1	Order	Registered	CWRU Email	Name	Your Preferred Email	Preferred Name (Please call me...)
2	0	Teaching	Thomas.Love@case.edu	Love, Thomas	Thomas.Love@case.edu	Dr. Love, Professor Love or Tom
3	1	CRSP 431	zainab.albar@case.edu	Albar, Zainab		
4	2	CRSP 431	oghenerukeme.asagba@case.edu	Asagba, Oghenerukema		
5	3	CRSP 431	imad.bagh@case.edu	Bagh, Imad		
6	4	CRSP 431	laura.baldassari@case.edu	Baldassari,Laura Elizabeth		

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/>

Your mileage may vary, but that site's AI was 5 years too high on this black-and-white headshot of me. Do you think you did that well?



Card 1



Card 1: Eric Chong, Master Chef Canada winner, in April 2014, age 21

Card 2



Card 2: Katherine Archuleta, former U.S. OPM Director, in 2013, age 64

Card 3



Card 3: Elise Mayfield, Chef, <http://elisemayfield.com> in 2014, age 28

Card 4

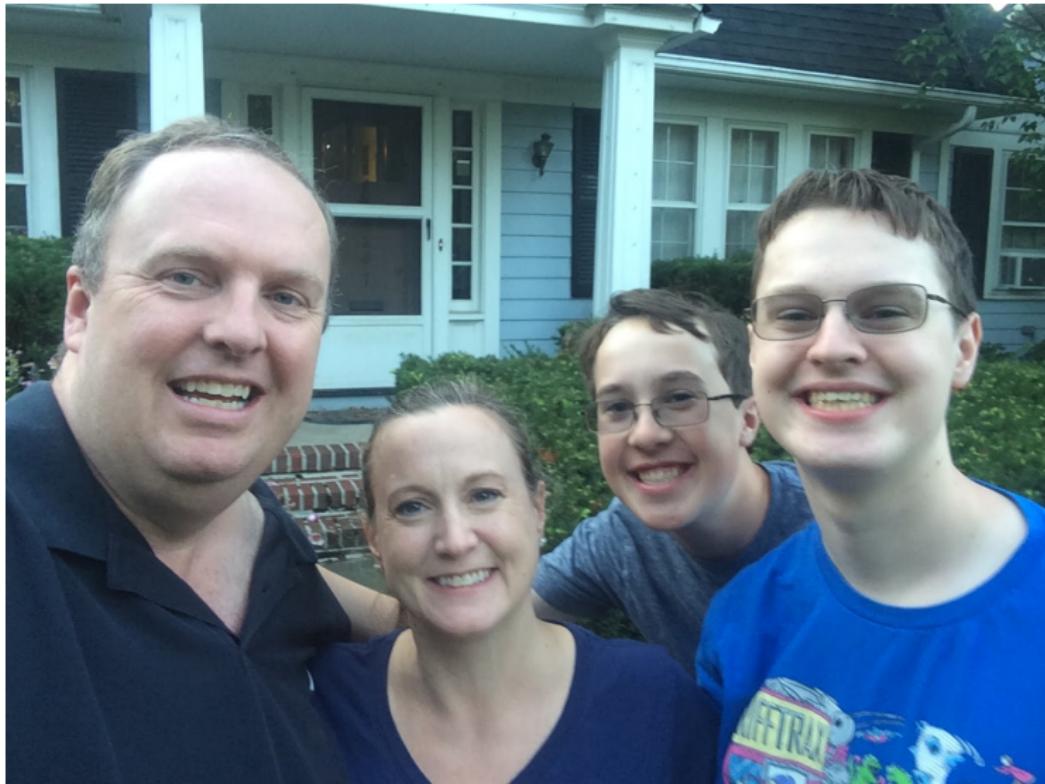


Card 4: Kevin Love, taken June 2014, age 14

No, not THAT Kevin Love



THIS Kevin Love, on the right here



He's now 18, and his brother is 15. Incidentally, I'm 50

Card 5



Card 5: Rosemary McGinn, in July 2013, age 54

Card 6



Card 6: Basketball Coach John Chaney, in 2006, age 74

Card 7



Card 7: David Storm, in August 2014 age 44

Card 8



Card 8: Writer Margo Glantz, in 2013, age 83

Card 9



Card 9: Fugitive Quade Ross Honey, in 2012, age **24**

Card 10



Card 10: Actress Bianca Lawson, in 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. <https://goo.gl/ccgUS3>

Age-Guesses-in-431_2017-08-29														
File	Edit	View	Insert	Format	Data	Tools	Add-ons	Help	All changes saved in Drive					
	Print	Find	Filter	\$	%	.0	.00	123	Georgia	10	B			
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													
12														

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



#1 Age 21
AI guess 27



#2 Age 64
AI 44



#3 Age 28
AI 22



#4 Age 14
AI 19



#5 Age 54
AI 36



#6 Age 74
AI 63



#7 Age 44
AI 55



#8 Age 83
AI 79



#9 Age 24
AI 35



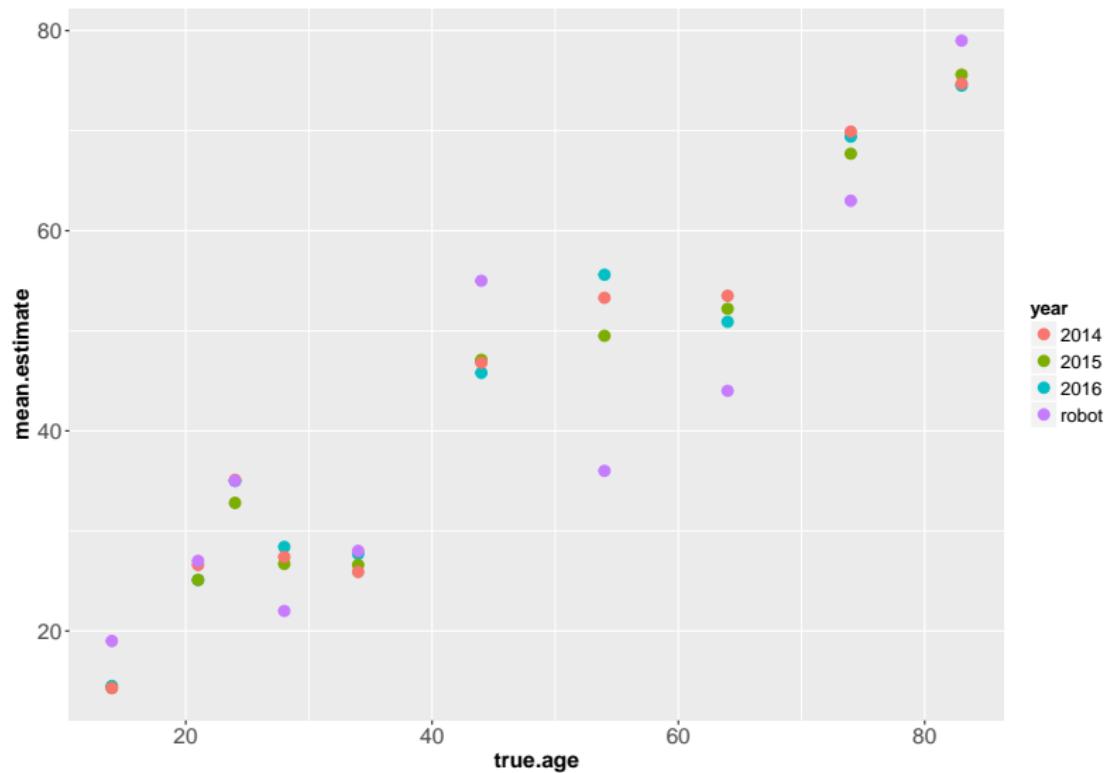
#10 Age 34
AI 28

Robot consulted on 2016-08-27. Algorithm continues to evolve.

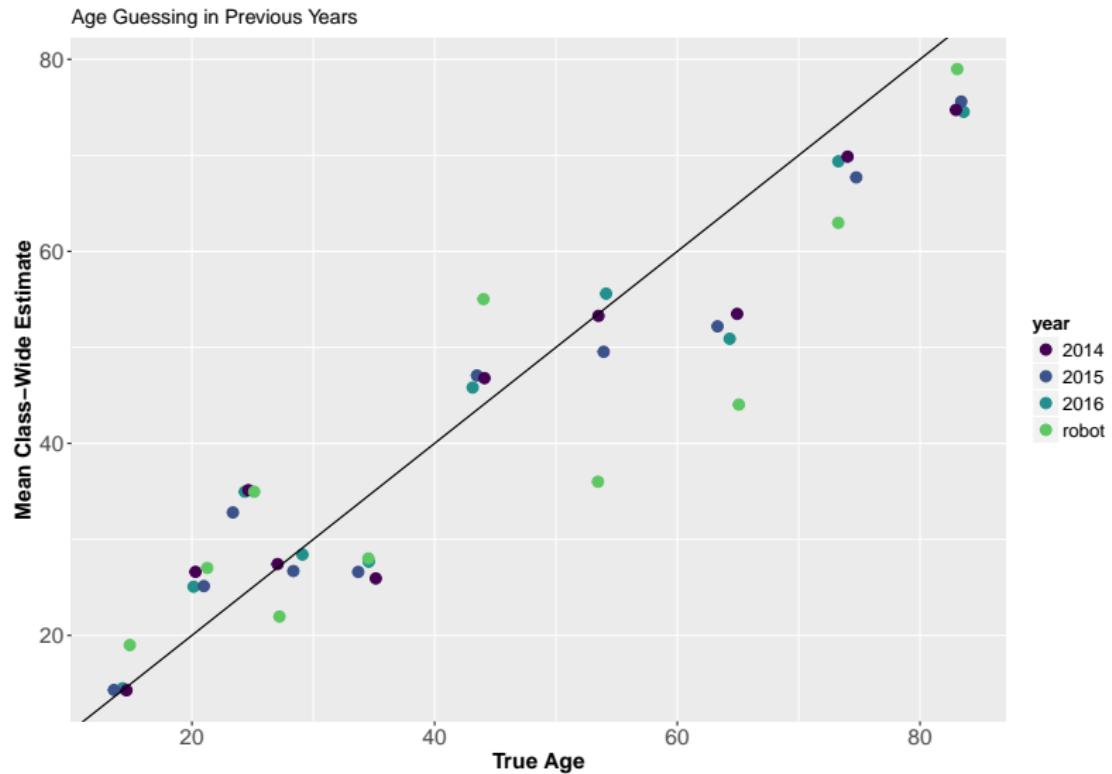
ageguesshistory.csv Data Set (excerpt)

card	label	true.age	sex	facing	year	mean.estimate	error
1	Chong	21	M	R	2016	25.1	4.1
2	Archuleta	64	F	L	2016	50.9	-13.1
3	Mayfield	28	F	L	2016	28.4	0.4
4	Love	14	M	L	2016	14.5	0.5
5	McGinn	54	F	R	2016	55.6	1.6
6	Chaney	74	M	L	2016	69.4	-4.6
7	Storm	44	M	R	2016	45.8	1.8
8	Glantz	83	F	L	2016	74.5	-8.5
9	Honey	24	M	L	2016	35.0	11.0
10	Lawson	34	F	R	2016	27.7	-6.3
1	Chong	21	M	R	2015	25.1	4.1
2	Archuleta	64	F	L	2015	52.2	-11.8

Scatterplot of Prior Results, 1



Scatterplot of Prior Results, 2



Mean Class-Wide Guesses (2014-16 combined)



#1 Age 21	#2 Age 64	#3 Age 28	#4 Age 14	#5 Age 54
2014-16	25.6	52.2	27.5	52.8
Mean Guesses	69.0	46.6	74.9	34.3
#6 Age 74	#7 Age 44	#8 Age 83	#9 Age 24	#10 Age 34



Mean Class-Wide Errors (2014-16 combined)

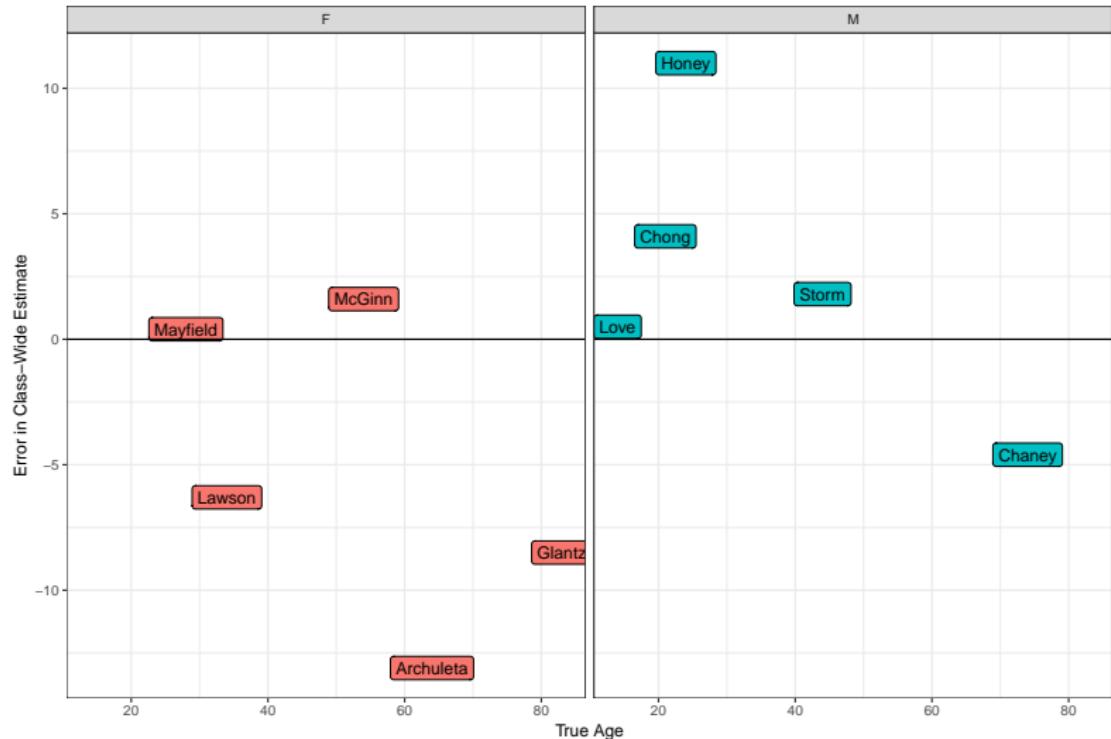


#1 Age 21	#2 Age 64	#3 Age 28	#4 Age 14	#5 Age 54
2014-16	+4.6	-11.8	-0.5	-1.2
Errors	-5.0	+2.6	-8.1	+10.3
#6 Age 74	#7 Age 44	#8 Age 83	#9 Age 24	#10 Age 34



Scatterplot of Prior Results, 3

Errors in 2016 Age Guessing, by Subject's Sex



Hans Rosling and “The Joy of Stats”

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

<https://www.youtube.com/watch?v=jbkSRLYSoj>

And if you liked that ...

- The 20 minute version (from 2007):
<https://www.youtube.com/watch?v=RUwS1uAdUcl>
- The full documentary from the BBC:
<https://www.gapminder.org/videos/the-joy-of-stats/>
- Video playlist from Gapminder: <https://www.gapminder.org/videos/>

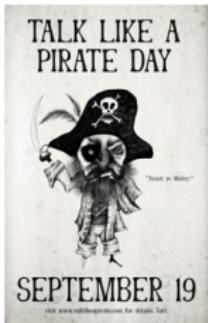
What's next?



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



R Packages



R Markdown

from R Studio

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?

- ➊ Turn in your index card with your two guesses of my age, please.
 - ➋ Visit <https://goo.gl/QTSGYu> and give us your preferred name and email.
 - ➌ Follow the [software installation instructions](#) to get R, R Studio, R Packages and 431 Data on your computer.
 - ➍ Obtain Jeff Leek's [The Elements of Data Analytic Style](#).
 - ➎ Obtain Nate Silver's [The Signal and the Noise](#).
 - ➏ Read the syllabus and look at the rest of the website. Make sure you view the Course Notes.
 - ➐ Ask us questions. TA office hours start next week, but email is available now.
-
- Course Web Site: <https://github.com/thomaselove/431>
 - Want help? Email 431-help@case.edu