

431 Class 01

Thomas E. Love, Ph.D.

2023-08-29

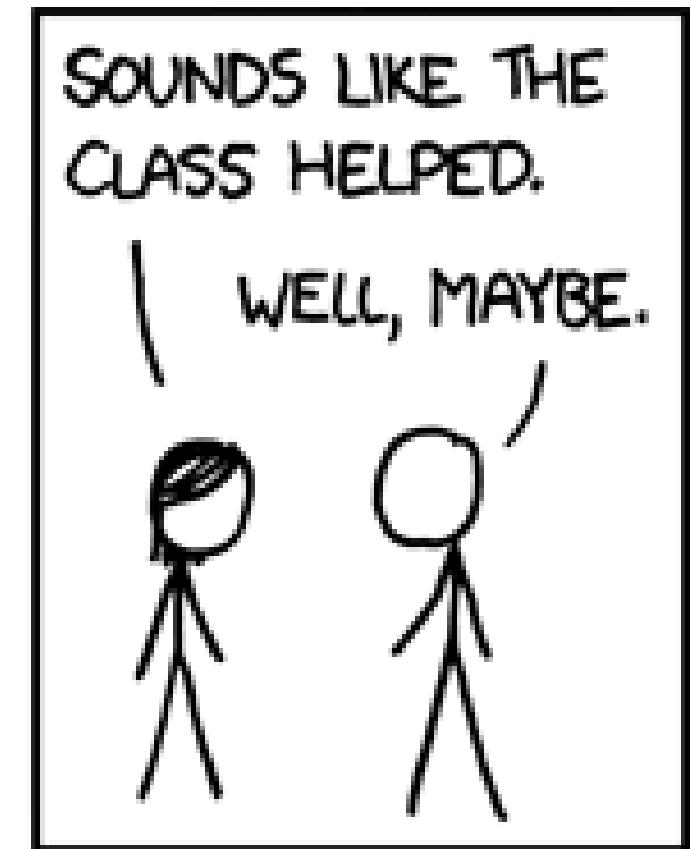
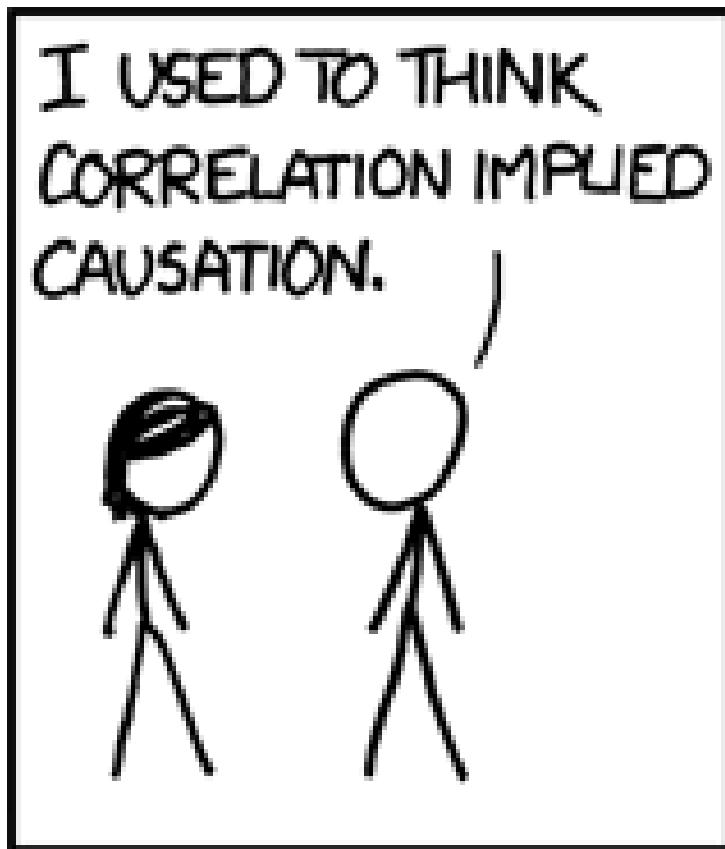
Getting To These Slides

Our web site: <https://thomaselove.github.io/431-2023/>

Visit the [Course Calendar](#) at the top of the page, which will take you to the Class 01 README page.

- These Slides for Class 01 are linked at the Class 01 README.
 - We'll look at the **HTML slides** during class.
 - We also provide the Quarto code I used to build the slides.

This is PQHS / CRSP / MPH 431



Correlation

- Source: [XKCD](#)

First Activity

First Thing

Write down your guess of Dr. Love's age in years in the appropriate spot on the **convenient piece of paper** we've provided.

Hang on to the paper, as you'll need it again later.

Here's a picture, in case that's helpful.



Dr. Love

Course Details

Instructor: Thomas E. Love, Ph.D.

Email: Thomas dot Love at case dot edu

- (best way to reach me, although you won't use it much)

Our web site: <https://thomaselove.github.io/431-2023/>

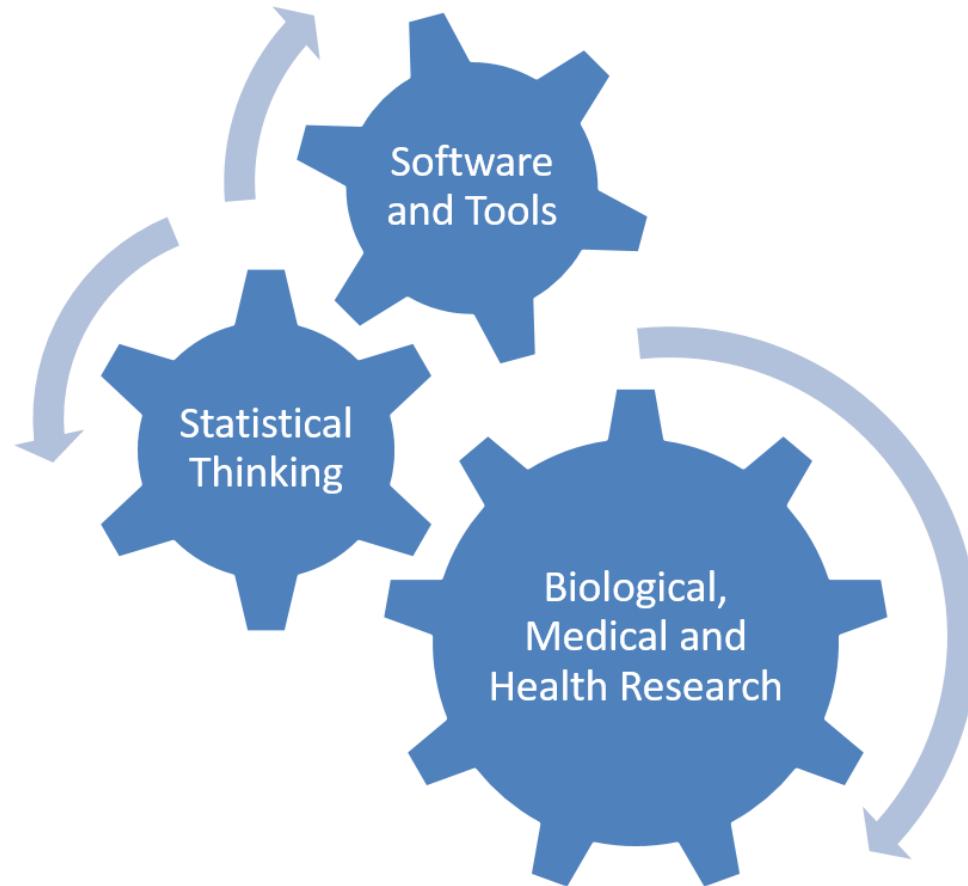
- The [Contact Us page](#) on our web site provides details on how to get help.
- If you've spent 15 minutes working on something and are stuck, don't keep working on it. **ASK FOR HELP.**

What does our web site link to?

- Course [Syllabus](#)
- Course [Calendar](#), which is the final word for all deadlines, and also links to each day's [class page](#).
- Course [Notes](#) (essentially a textbook)
- [Software Details](#) (R and R Studio, installation, data and code downloads)
- [Assignments](#): 7 Labs, 2 Quizzes, 2 Projects and 10 Minute Papers

Teaching Assistants (Fall 2023)

Structure of the 431 Course



- Exploratory Data Analysis, Visualization
- Statistical Inference, Making Comparisons
- Linear Regression and related Models

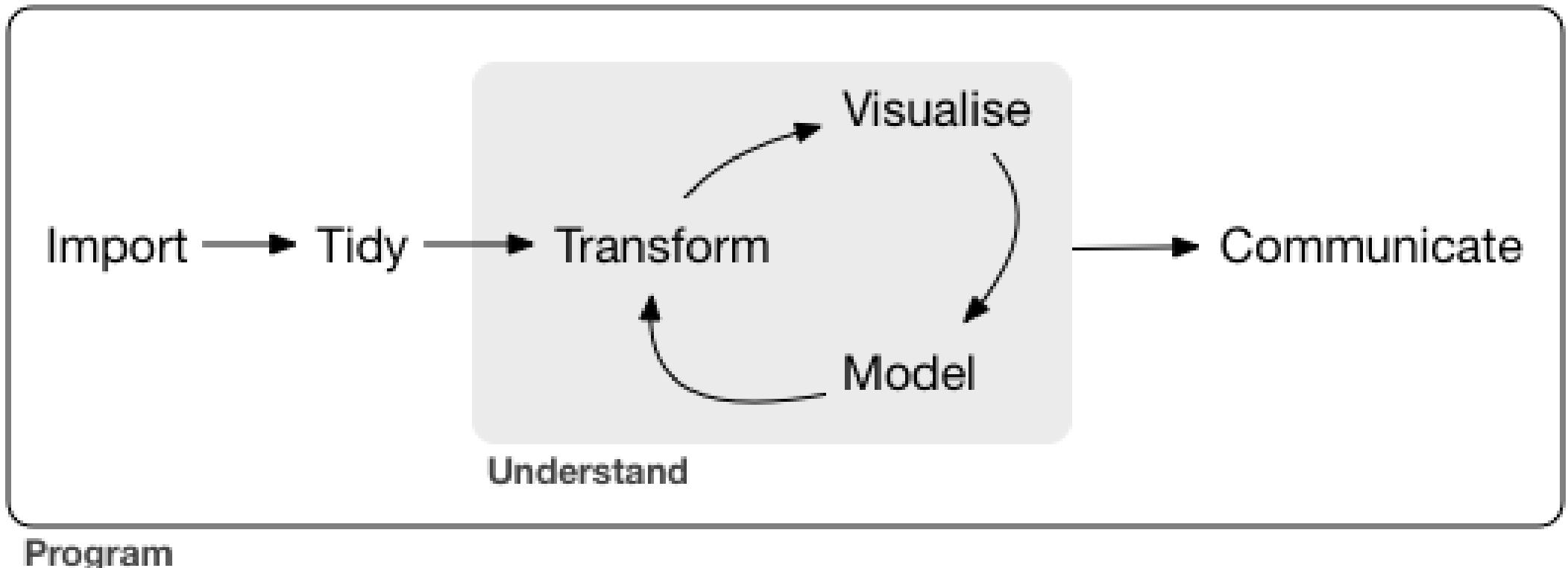
Course Philosophy (in one slide)

The course is about biostatistics, replicable research, using state-of-the-art tools (R, R Studio, Quarto), 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.

More on all of this in the [Course Syllabus](#), of course.

What is Data Science about?



Source: Figure 1.1 in <https://r4ds.hadley.nz/intro.html>

What will we be reading?

Tools We Use in 431

- [Calendar](#) (for deadlines and what's next)
- [Campuswire](#) (for discussion / Q and A outside of class)
- [Canvas](#) (for submission of work)
- [Zoom](#) (for TA office hours and class recordings)
- [Shared Google Drive](#) (mostly Lab/Quiz/Project answer sketches and feedback)
- [Main Website](#) links to everything you'll need (see next slide)

Our Main Website

The screenshot shows a course website with a blue header bar. The header includes links for '431-2023', 'Home' (which is highlighted), 'Calendar', 'Syllabus', 'Course Notes', 'Assignments', 'Software', 'Campuswire', 'Key Links' (with a dropdown menu), 'Contact Us', and a search icon. The main content area has a title '431 Course Main Page: Fall 2023'. Below it, there's information about the author ('Thomas E. Love') and publication date ('2023-08-26'). A 'Welcome!' section states that 431 is the first part of a two-semester sequence taught by Professor Thomas Love at Case Western Reserve University. A bold statement 'Everything will be linked here' is followed by a horizontal line. To the right, a 'Key Links' dropdown menu is open, listing various resources like Canvas, Google Drive, and GitHub repositories for Project A and B.

431-2023 Home Calendar Syllabus Course Notes Assignments Software Campuswire Key Links Contact Us Q

431 Course Main Page: Fall 2023

AUTHOR
Thomas E. Love

PUBLISHED
2023-08-26

Welcome!

431 is the first part of a two-semester sequence (with 432) in applied statistics and data science taught by Professor Thomas Love in the [Department of Population and Quantitative Health Sciences](#) at [Case Western Reserve University](#).

Everything will be linked here

- Canvas (submissions & recordings)
- Shared Google Drive (CWRU)
- 431-Data page
- 431 R packages
- Minute Papers (Github)
- Labs (Github)
- Quizzes (Github)
- Project A Website
- Project B Website
- Class READMEs (Github)

The link to this page is at the bottom of every slide.

Keeping Caught Up

- If you have to miss a class, catch up before the next one.
- We'll try to record the classes using Zoom and then make them available afterwards.
 - You'll find the recordings on [Canvas](#) in the Zoom section.
 - We prefer you not to join the Zoom live, but rather watch the recording when it is posted (usually the same day.)
- Our assignments have deadlines, which are posted to the [Calendar](#), and which we expect you to meet.

Attendance Policies

1. We expect you to attend **20** of the 26 classes in person, at minimum.
2. Don't come if you are sick, please. Watch the recording instead.
3. If you're getting over an illness, but are well enough to attend class, please mask up.
4. If you will need to miss **more than two** classes in a row, or if you cannot keep up with assignments, that's when Dr. Love needs to hear with you.

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

John Tukey (1915-2000)¹

Ten Things To Do After Class 01

1. Review the [main course website](#), being sure to visit the [Course Calendar](#).
2. Read through the [Course Syllabus](#).
3. Obtain David Spiegelhalter's *The Art of Statistics: How to Learn from Data* (~\$20).
4. Complete the [Welcome to 431 survey](#).
5. [Install the software](#) you'll need.
6. Sign up for [Campuswire](#).
7. Look at the [Course Notes](#)
8. Bookmark some books, plus the Posit Cheatsheets (see README).
9. Ask us questions! Campuswire is available now. TA hours start 2023-09-05.
10. Take a look at [Lab 01](#) due 2023-09-12, our first substantial assignment.

See the [Class 01 README](#) for more details.

We'll form groups in a moment

- Shortly, we will be asking you each to join a group, containing five or six people. Join a group where you will meet at least one new person.
- Also, one member of each group will serve as recorder and will need to open a Google Form on their laptop in a moment.
 - Everyone else needs nothing except their convenient piece of paper and a pen/pencil.
- Your first task is to settle on a name for your group. Try to be a little creative.

We'll be guessing some ages

- I will display a series of 10 photographs, each of a person.
- For each photo your group will ...
 - estimate the age of the person in the photo (in years)
 - have the recorder type your (group) guess into the form
(so if you guess age 35, you will just type 35.)
- When you've produced guesses for each of the 10 photos, submit the form. The recorder will get an email confirmation.
- Later, we'll reveal the true ages and compute errors.

OK. Let's form the groups.

- Remember, your group should have FIVE OR SIX people, with at least one person you don't already know.

1. Select a group name.
2. Select a recorder, who should visit the link below after logging into Google via CWRU.

<https://bit.ly/431-2023-class01-breakout>

3. Make sure everyone knows everyone else's name, as well as your group's name.

Here come the photos

- We'll give you a little more time for the first two photos than the other eight.

Remember to have the recorder fill out the form at

<https://bit.ly/431-2023-class01-breakout> although it may help all of you to keep track on paper, as well.

Photo 1



Photo 2



Making Progress

- Your group's guess for each photo should be in the form at <https://bit.ly/431-2023-class01-breakout>.
 - You might also want to keep track on your convenient piece of paper, so that when I tell you the ages later, you'll be in a position to see how your group did.
- In spare time between photos, please make the effort to learn **something** about each of the other people in your group beyond their name: perhaps what field they are in, or where they come from.

Photo 3



Photo 4



Photo 5



Photo 6



Photo 7



Photo 8



Photo 9



Photo 10



Now, guess My Age again

1. You should have an initial guess of my age written down from the start of the session.
2. Now, make a second guess of my age based on what you know about me now, and write that down next to the initial guess.

So if you guessed 18 initially, but now think I'm 19, you should write 18/19. If you still think I'm 18, write 18/18. Make it easy for us to understand your guesses of my age on the **convenient piece of paper**. Don't guess my age as a group - just write down your own guess.

Age Guessing Robots?

Well, Microsoft used to have a tool online at how-old.net to do this. There are some related robots that still do the job online, although most people are unwilling to use them.

<https://how-old.net>



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



6 years too low

Do you think you did that well?

OK. Back to the photos!

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, at left (Aug 2023)



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

How did the AI do in August 2016?



#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

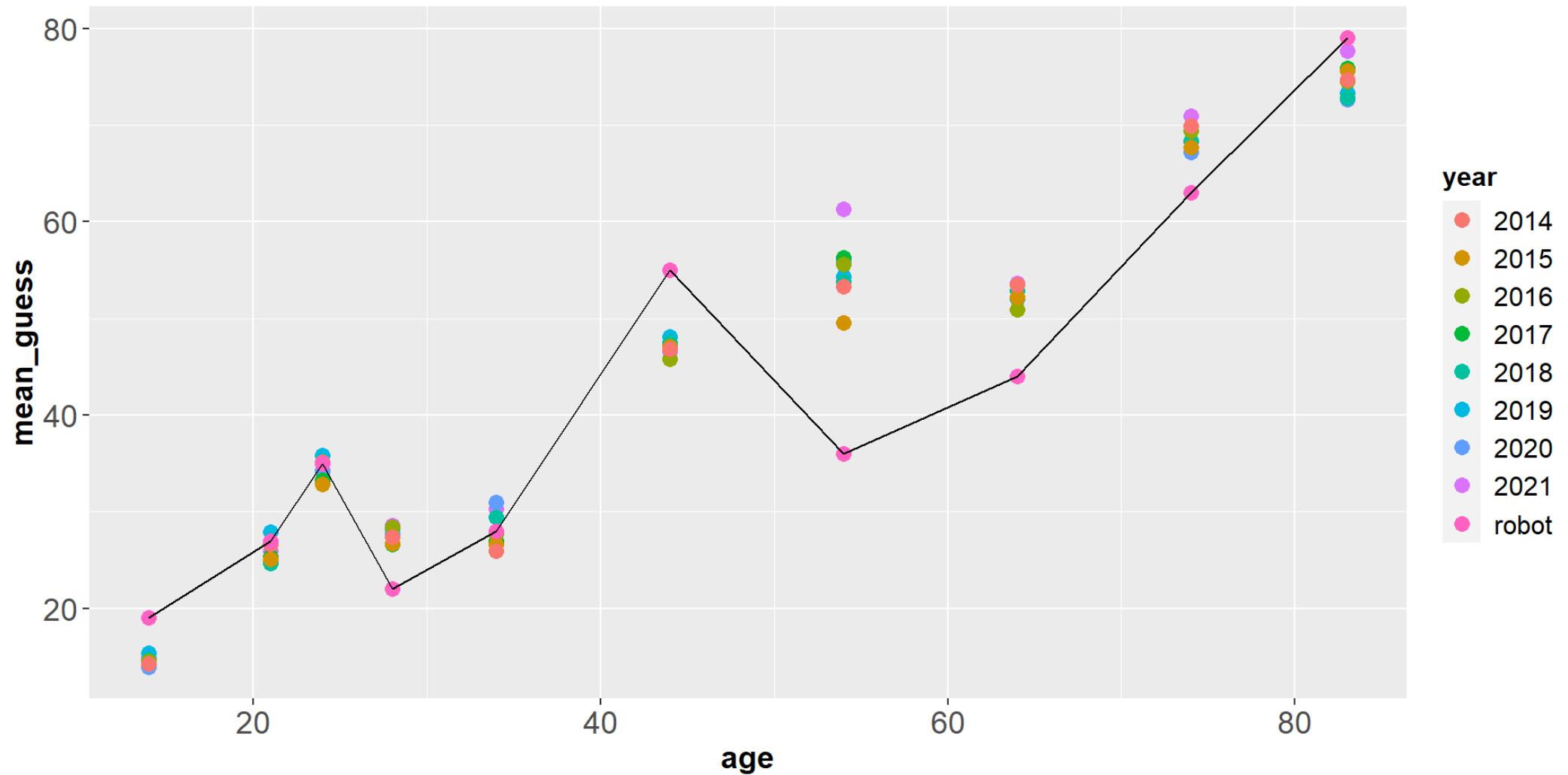
#10 Age **34**
AI 28



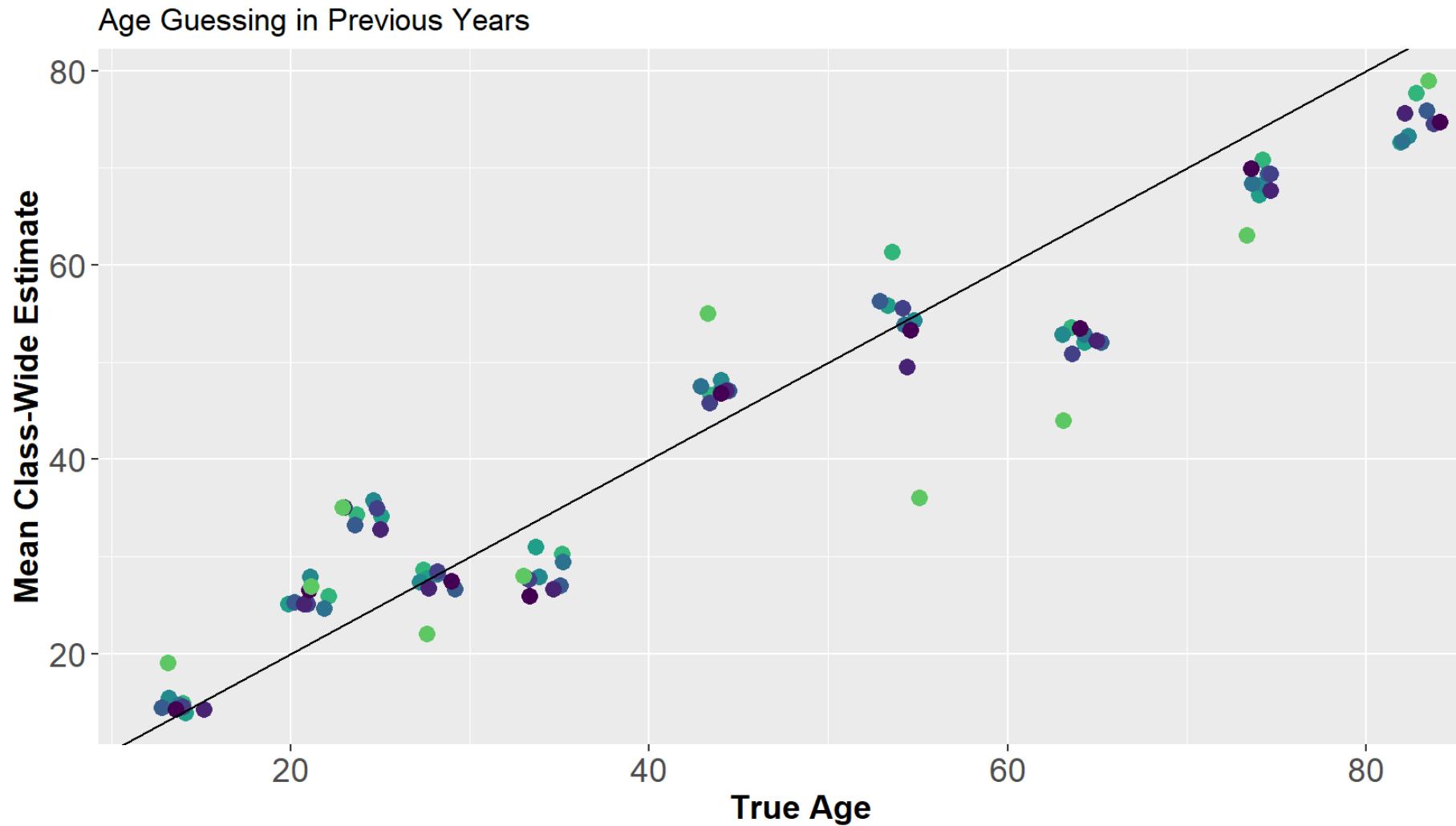
Some Data from Prior Years

label	year	age	mean_guess	error
Chong	2021	21	25.9	4.9
Archuleta	2021	64	53.6	-10.4
Mayfield	2021	28	28.6	0.6
Love	2021	14	14.9	0.9
Chong	2020	21	25.1	4.1
Chong	2019	21	27.9	6.9
Chong	2018	21	24.7	3.7

Scatterplot of Prior Results, 1



Scatterplot of Prior Results, 2



Mean Class-Wide Guesses (2014-21)



	#1 Age 21 2014-2021 Mean Guesses	#2 Age 64 25.7	#3 Age 28 52.5	#4 Age 14 27.6	#5 Age 54 14.6
	#6 Age 74	#7 Age 44	#8 Age 83	#9 Age 24	#10 Age 34



Mean Class-Wide Errors (2014-21)

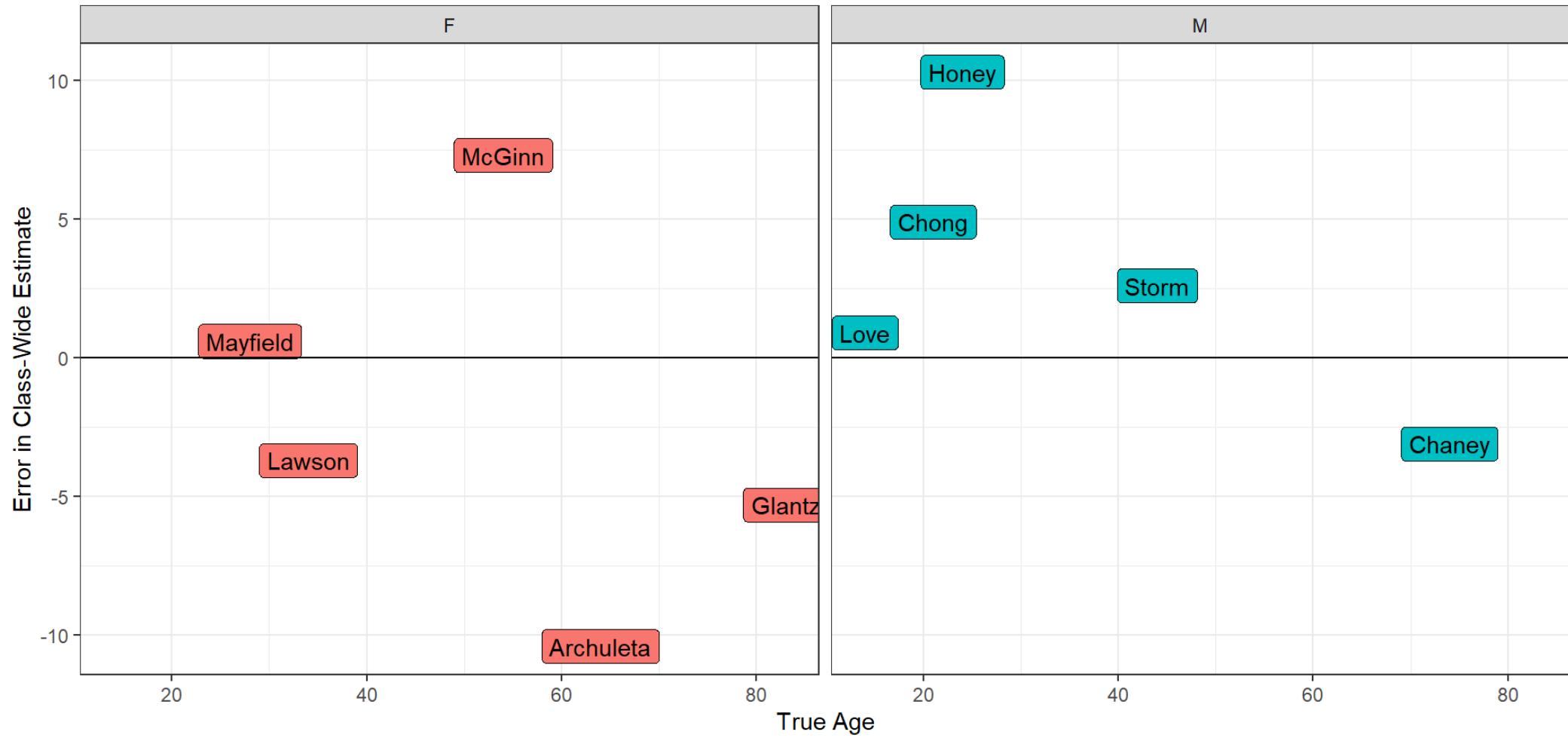


	#1 Age 21 2014-21 Mean Error	+4.7	#2 Age 64 -11.5	+3.0	#3 Age 28 -0.4	-8.4	#4 Age 14 +0.6	+10.5	#5 Age 54 1.0	-5.8
	#6 Age 74		#7 Age 44		#8 Age 83		#9 Age 24		#10 Age 34	



2021 Results: Labeled Scatterplot

Errors in 2021 Age Guessing, by Subject's Sex



Hans Rosling “The Joy of Stats”

200 countries over 200 years using 120,000 numbers, in less than 5 minutes.

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/>

Thanks for coming!

See you Thursday at 1 PM right here.

