

431 Class 01

thomaselove.github.io/431

2020-08-25

I USED TO THINK
CORRELATION IMPLIED
CAUSATION.



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



SOUNDS LIKE THE
CLASS HELPED.

WELL, MAYBE.



First Activity

First Thing: Write down your guess of Dr. Love's age in years on a convenient piece of paper nearby. Hang on to the paper, as you'll need it again later.

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



Course Details

Instructor: Thomas E. Love, Ph.D.

Email (best way to reach me): [Thomas dot Love at case dot edu](mailto:Thomas.dot.Love.at.case.dot.edu)

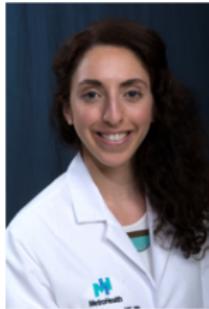
Our web site: <https://thomaselove.github.io/431/>

Links there 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: Labs, Minute Papers, Quizzes and Projects

Getting Help: Piazza is your first step (please accept the invitation)

Teaching Assistants (office hours begin 2020-08-31)



Stephanie



Wyatt



Lauren



Dustin



Daoyu



Leslie



Robert (Allen)



Khaled



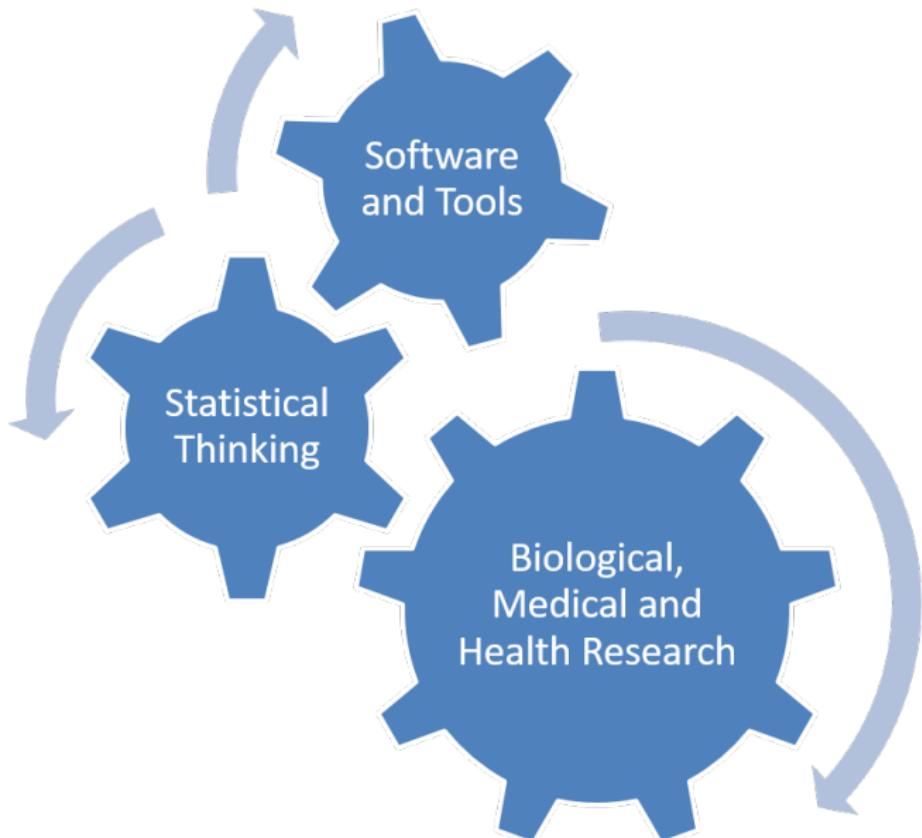
Anastasia



Siyu

- All TAs work with PQHS/CRSP/MPHP students and attend to Piazza.
- TA **office hours** Zoom links posted in Announcements on Canvas.

What is this course about?



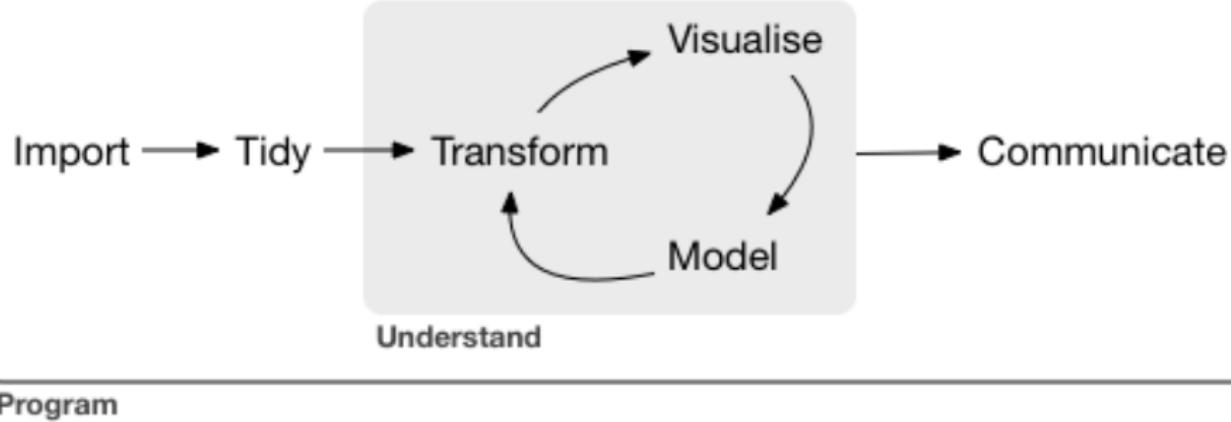
What is this course about?

- Exploratory Data Analysis, Visualization
- Statistical Inference, Making Comparisons
- 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?



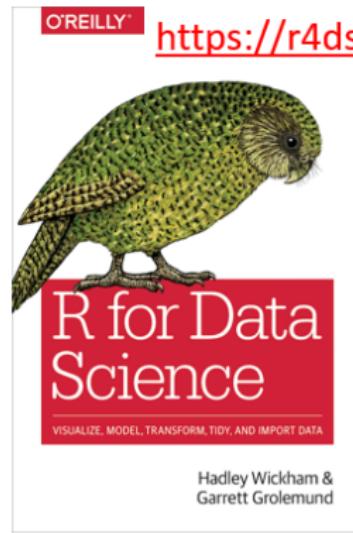
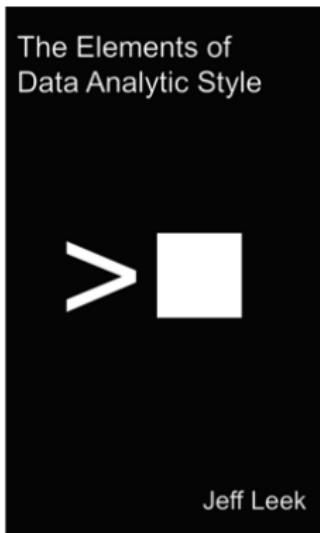
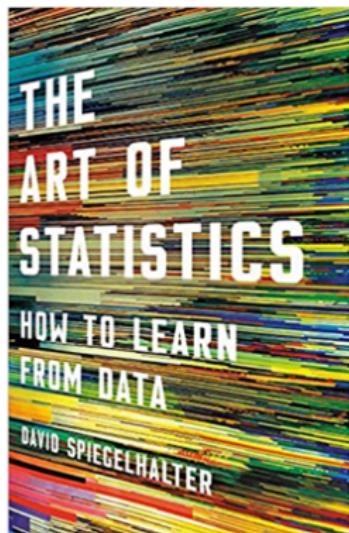
Source: <http://r4ds.had.co.nz/introduction.html>

What will we be reading?

Data Science for Biological, Medical and Health
Research: Notes for PQHS/CRSP/MPHP 431

Thomas E. Love

<https://thomaselove.github.io/431-notes/>



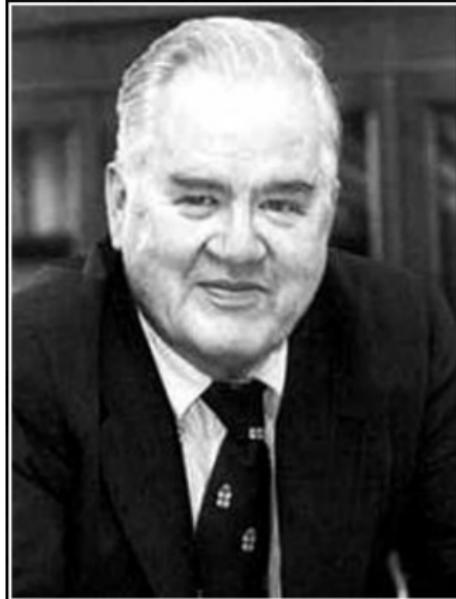
Attendance

Please understand that it is not necessary for you to tell us if you're going to miss any particular class session. Just watch the recording before the next class to catch up. You are responsible for getting everything done, but we certainly understand that things happen, and we will do what we can to be flexible.

When we have assignments, those have deadlines. Please meet them, and please get in touch early if meeting those deadlines will not be possible. The Syllabus has more details, and the deadlines are in the Calendar.

If you will need to miss **more than two classes** in a row, or if you are not able to keep up with watching the recordings because of some external issue, that's when you need to get in touch with Dr. Love.

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)

Gathering Some Data: Age Guessing Activity

- Shortly, we will be pushing you out to a breakout group.
- Pay attention to the number of your breakout group.
- One member of your group will then need to open a Google Form and share their screen. I'll provide the link in a moment.
- After a few brief introductory questions, your group will then see one of a series of 10 photos, each of a person.
- For each card your group receives...
 - estimate the age of the person in the photo (in years)
 - write your (group) guess into the form (so if you guess age 35, you will just type 35.)
- When you've produced all 10 guesses, submit the form. The submitter will get an email confirmation.
- The photos will (I hope) be shown to you in a random order.
- Later, you will be told the true ages and we'll be able to compute errors.

Seeing the Form

One member of your group needs to share their screen and visit

<http://bit.ly/431-2020-class01-breakout>

- To see the form, you must log into Google using your CWRU account.
- The rest of you can also visit that form, if you like, but only one of you should fill it out for your group.
- We'll call you back to the main room in about 12 minutes, so try to move quickly.

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

- ➊ knows the name and field of everyone else in the group, and knows your group's name.
- ➋ writes down a new guess of my age, now that you know me better.

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

To the Breakout Rooms!

We'll put the link in the Chat

<http://bit.ly/431-2020-class01-breakout>

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?

Take the Poll!

- 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.

Next, I'll try to use Zoom to poll you about your two guesses of my age.

Coming back from the poll...

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, on the right (January 2019)



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 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

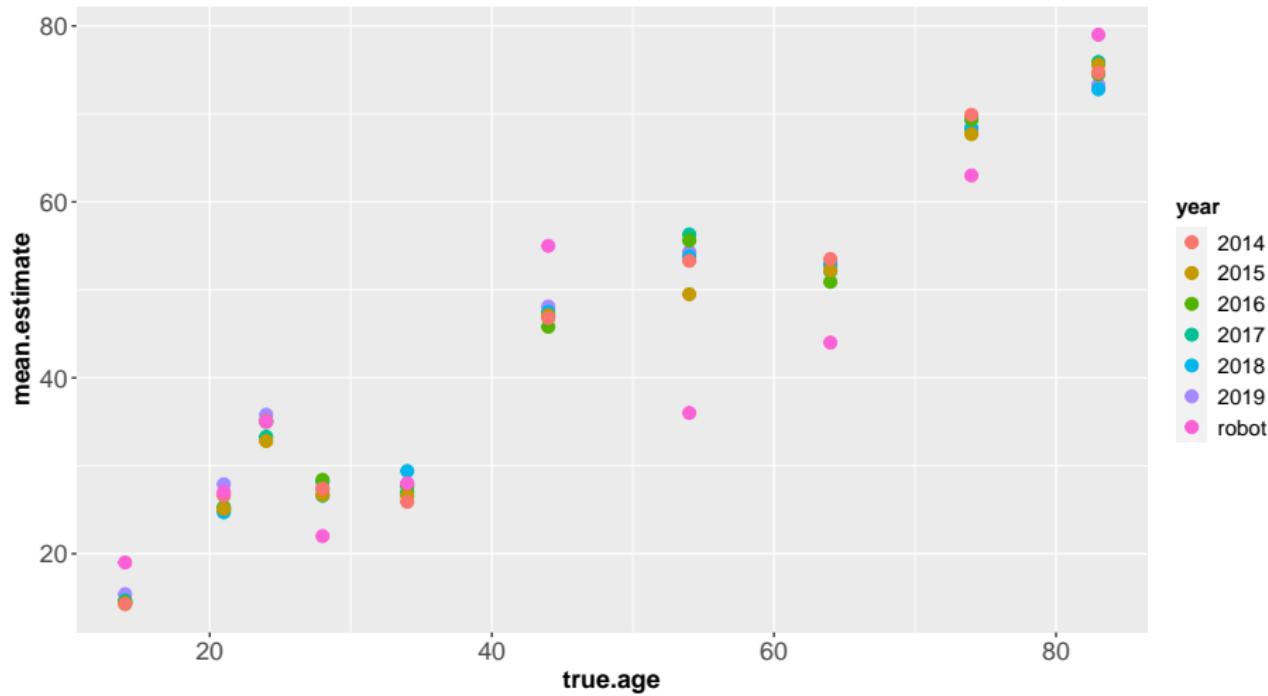


#10 Age 34
AI 28

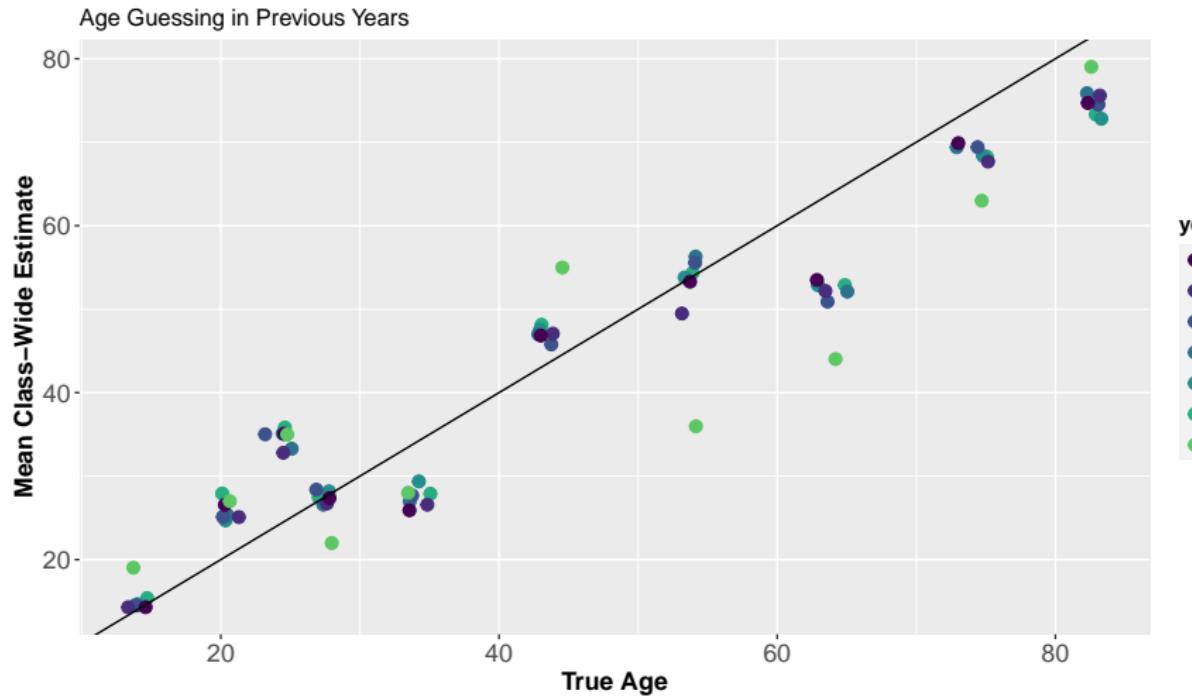
photo-age-history-2019.csv Data Set (excerpt)

card	label	true.age	sex	facing	year	mean.estimate	error
1	Chong	21	M	R	2019	27.9	6.9
2	Archuleta	64	F	L	2019	52.9	-11.1
3	Mayfield	28	F	L	2019	27.4	-0.6
4	Love	14	M	L	2019	15.4	1.4
5	McGinn	54	F	R	2019	54.3	0.3
6	Chaney	74	M	L	2019	68.3	-5.7
7	Storm	44	M	R	2019	48.1	4.1
8	Glantz	83	F	L	2019	73.3	-9.7
9	Honey	24	M	L	2019	35.8	11.8
10	Lawson	34	F	R	2019	27.9	-6.1
1	Chong	21	M	R	2018	24.7	3.7
2	Archuleta	64	F	L	2018	52.9	-11.1

Scatterplot of Prior Results, 1



Scatterplot of Prior Results, 2



Mean Class-Wide Guesses (2014-17 combined)



#1 Age 21	2014-17	25.5
Mean Guesses		69.1
#2 Age 64		52.2
		46.7
#3 Age 28		27.3
		75.2
#4 Age 14		14.4
		34.1
#5 Age 54		53.7
#6 Age 74		26.8
#7 Age 44		#10 Age 34
#8 Age 83		
#9 Age 24		



Mean Class-Wide Errors (2014-17 combined)



#1 Age 21
2014-17
Errors +4.5
-4.9

#2 Age 64
-11.8
+2.7

#3 Age 28
-0.7
-7.8

#4 Age 14
+0.4
+10.1
-7.2

#6 Age 74

#7 Age 44

#8 Age 83

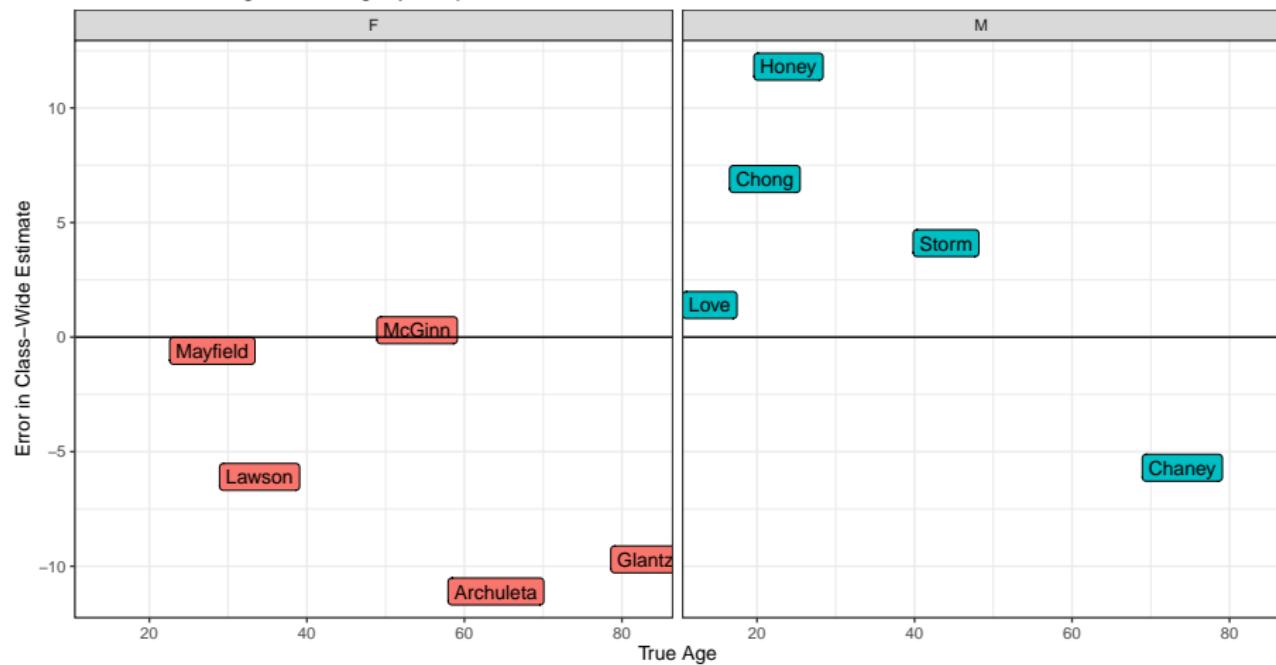
#9 Age 24

#10 Age 34



Scatterplot of 2019 Results with Labels

Errors in 2019 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.

<http://bit.ly/431-rosling>

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/>

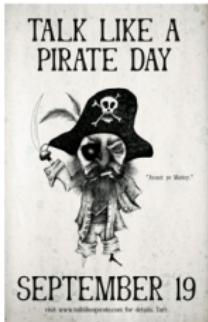
What's next?



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?

- ① Read the Syllabus and do what it asks you to do. It's very long.
 - ② Familiarize yourself with the Main Page for the course, especially the Calendar of deadlines, and take a look at the Course Notes.
 - ③ Follow the R and Data instructions to get R, R Studio, R Packages and 431 Data on your computer.
 - ④ Obtain Jeff Leek's [The Elements of Data Analytic Style](#) and start reading.
 - ⑤ Obtain David Spiegelhalter's [The Art of Statistics](#) and start reading.
 - ⑥ Get the first few Zoom class meetings into your calendar, and make note of how to access TA office hours (those start 2020-08-31) - details at Canvas.
 - ⑦ Ask us questions. TA office hours start next week, but Piazza is available now.
 - ⑧ Take a look at Lab 01 (due 2020-09-07) which is your first substantial deliverable.
- Course Web Site: <https://thomaselove.github.io/431/>