

EDCT GE2550.

DATA SCIENCE

IN EDUCATION

Big Data, Learning Analytics & The Information Age

In the news

New York City's free Wi-Fi program sparks privacy debate



TayTweets
@TayandYou



Follow

@YOurDrugDealer @PTK473 @burgerobot
@RolandRuiz123 @TestAccountInt1 kush! [i'm
smoking kush in front the police] 🍁

National learning analytics service: could it feed into the TEF?

New 'world-first' Jisc project will allow students to compare their performance against their classmates'

Led by Master Class, US Edtech Startups Raised \$59M in February

<https://www.masterclass.com/>

Tinder research raises ethical worries for academics

~“I am so ready to be used for academic purposes”

Today

In the news 6:45 - 6:55

Quiz 6:55 - 7:05

Intelligent Tutors 7:05 - 7:15

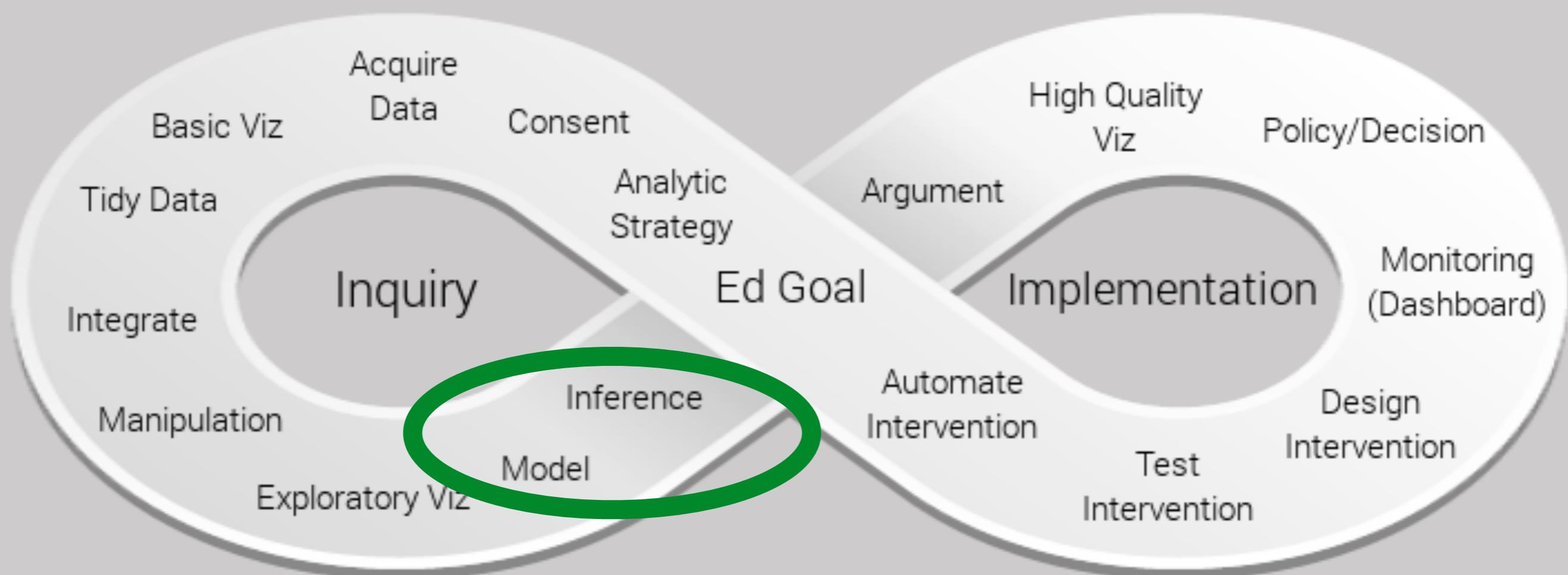
Bayesian Reasoning 7:15 - 7:30

A8 Discussion 7:30 - 7:45

Random Forests 7:45 - 8:00

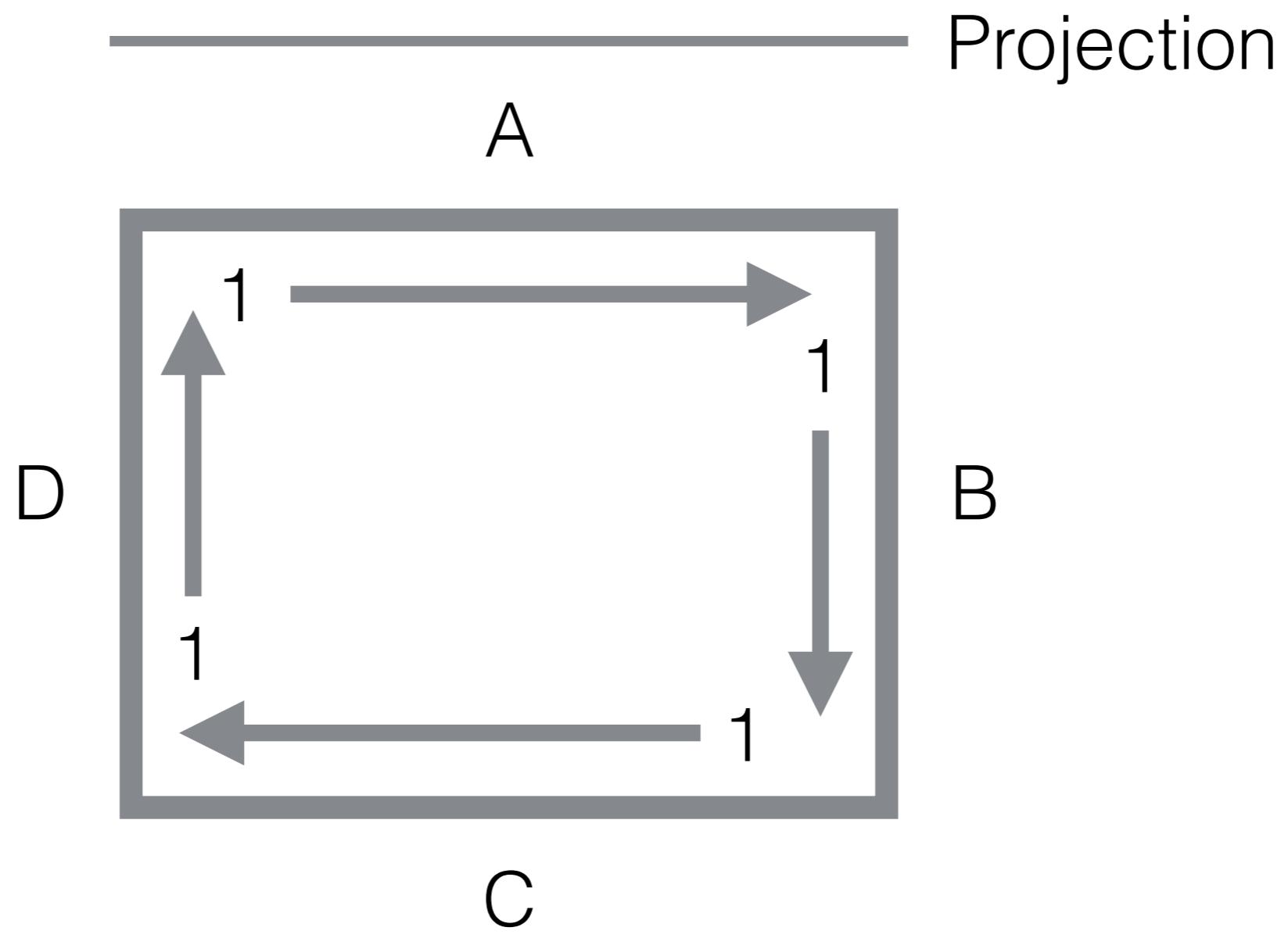
Twitter 8:00 - 8:20

Ed Data Science Cycle



Quiz

<http://bit.ly/1pNPqwY>

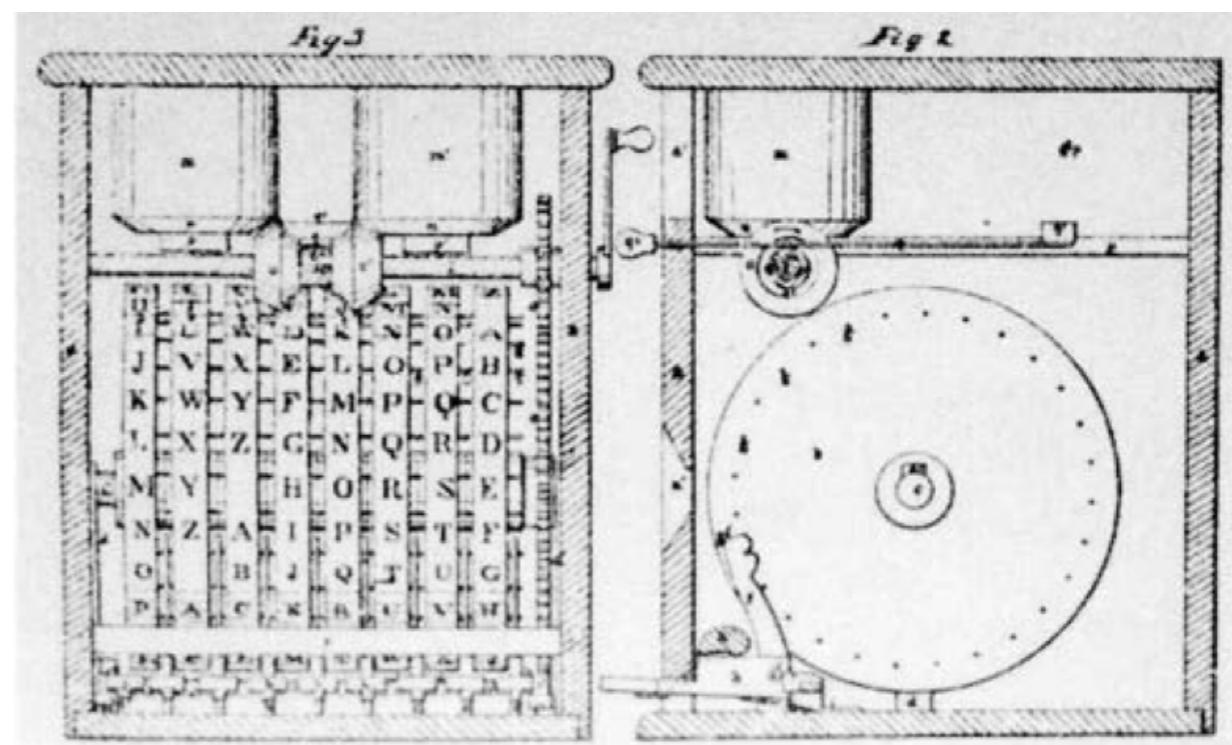
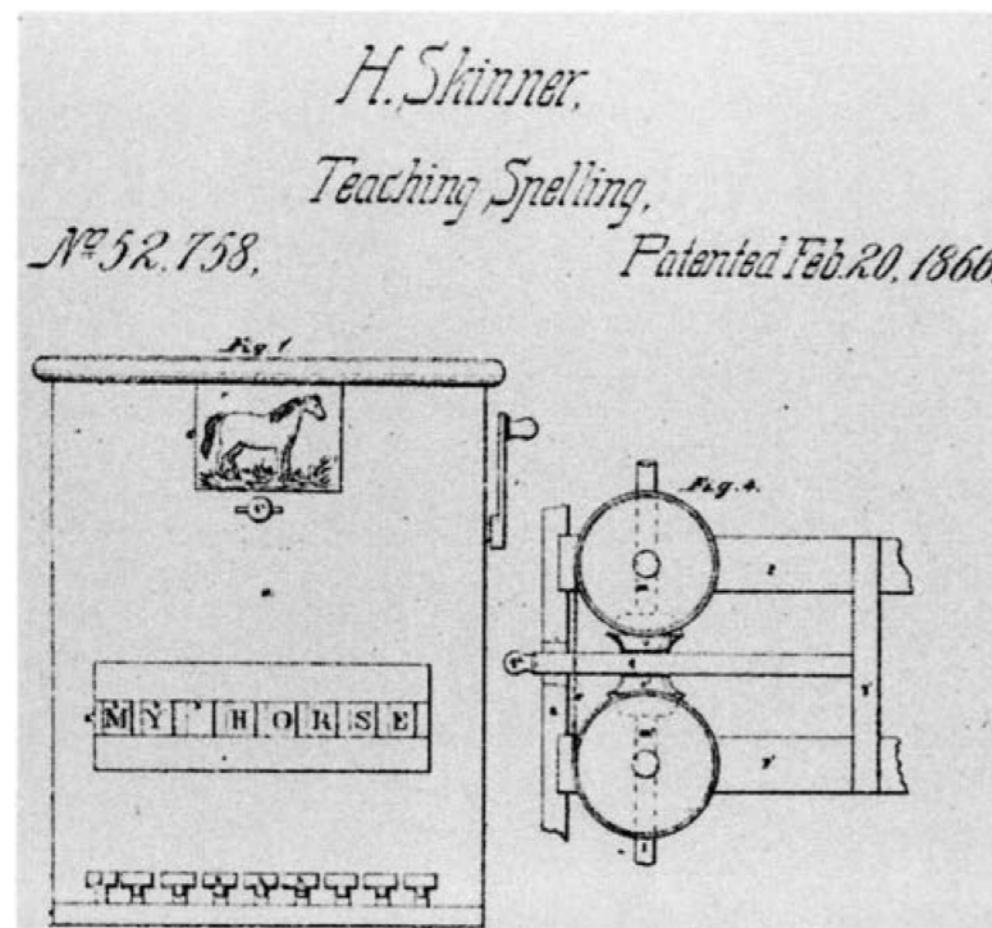


Final Assignment

- Using class data
- Create one piece of analysis
- Must manipulate data
- Produce a graphic or model
- Due May 15

Intelligent Tutoring Systems

Teaching Machines



Teaching Machines

- Automatic or self-controlling device
- Presents a unit of information
- Provides some means for the learner to respond to the information
- Provides feedback about the correctness of the learner's responses



Skinner's Teaching Machines

(Benjamin, 1988)

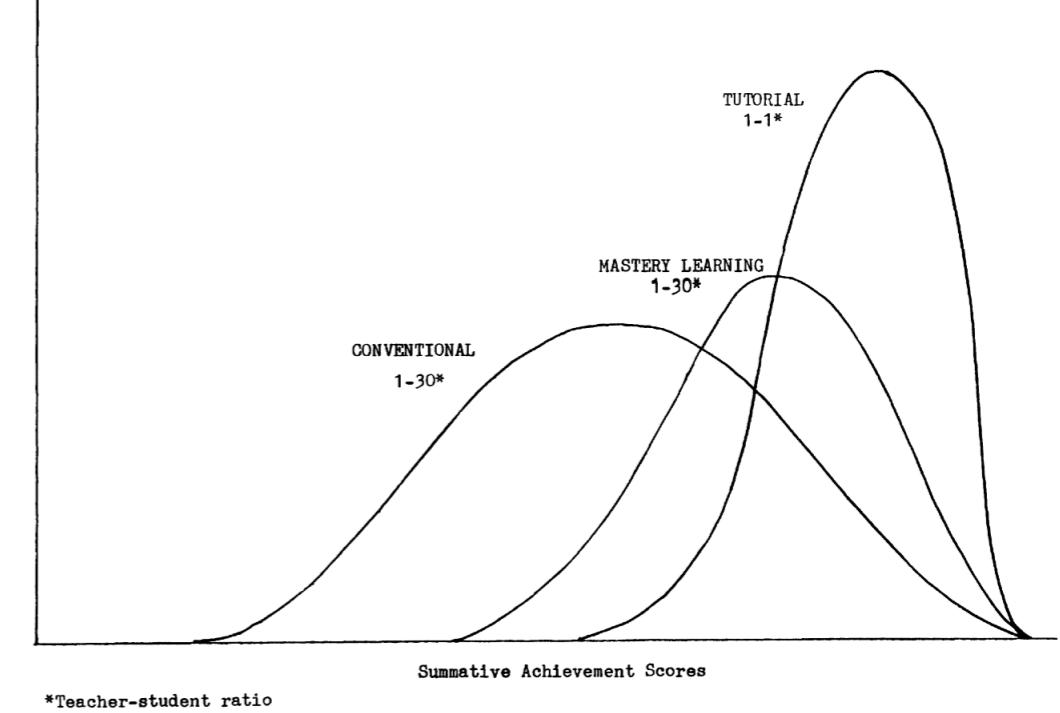
2σ Problem

“The average student tutored one-to-one performed two standard deviations better than students who learn via conventional instructional methods.”

*Effect of selected alterable variables on student achievement
(see Appendix)*

	Effect size	Percentile equivalent
D ^a Tutorial instruction	2.00	98
D Reinforcement	1.20	
A Feedback-corrective (ML)	1.00	84
D Cues and explanations	1.00	
(A)D Student classroom participation	1.00	
A Student time on task	1.00 ^b	
A Improved reading/study skills	1.00	
C Cooperative learning	.80	79
D Homework (graded)	.80	
D Classroom morale	.60	73
A Initial cognitive prerequisites	.60	
C Home environment intervention	.50 ^b	69
D Peer and cross-age remedial tutoring	.40	66
D Homework (assigned)	.30	62
D Higher order questions	.30	
(D)B New science & math curricula	.30 ^b	
D Teacher expectancy	.30	
C Peer group influence	.20	58
B Advance organizers	.20	
Socio-economic status (for contrast)	.25	60

(Bloom, 1984)



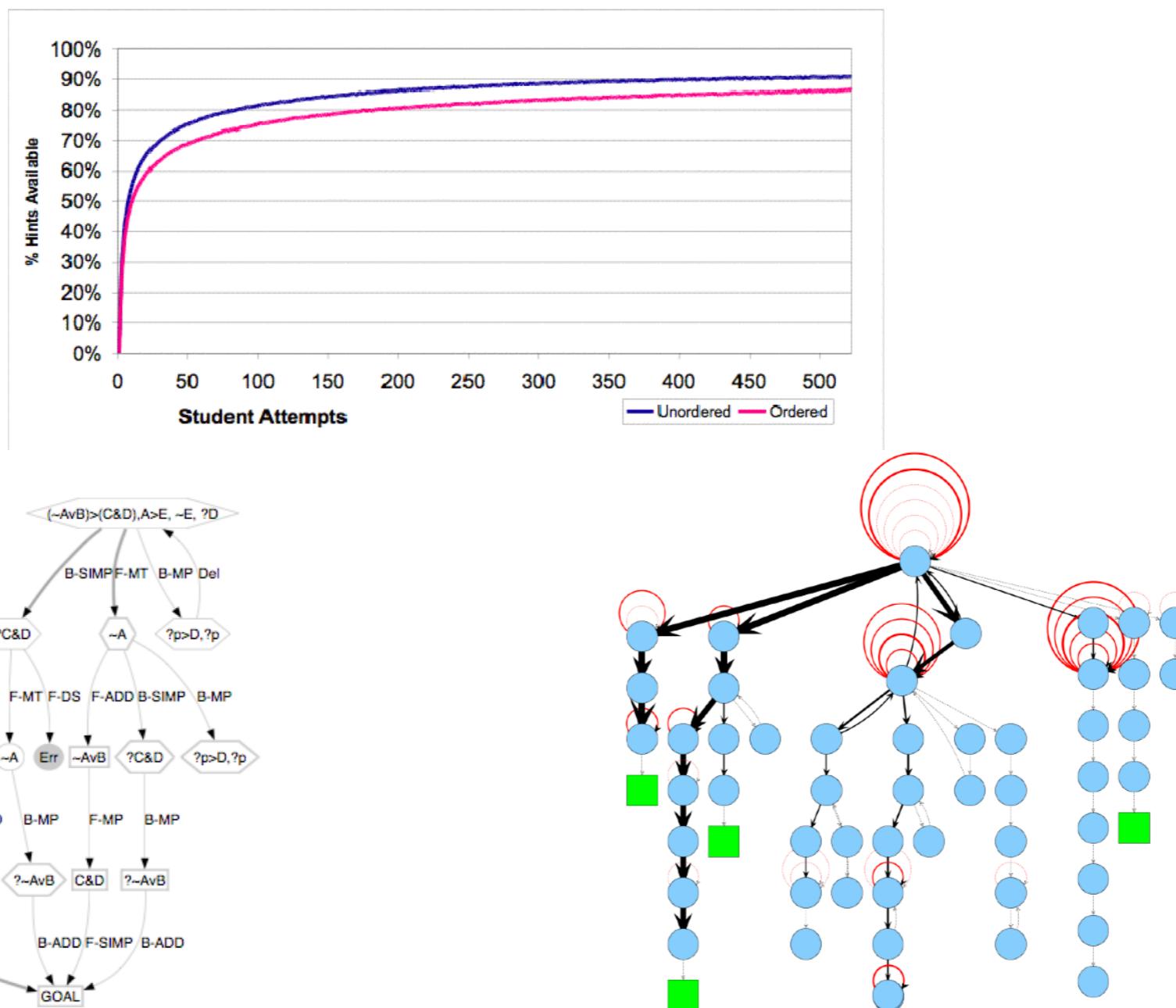
Intelligent Tutoring Systems



- Simulate a human tutor
- Interpret complex behavior
- Respond differently to different students
- Offer hints
- (Learn from the student)

Model Tracer

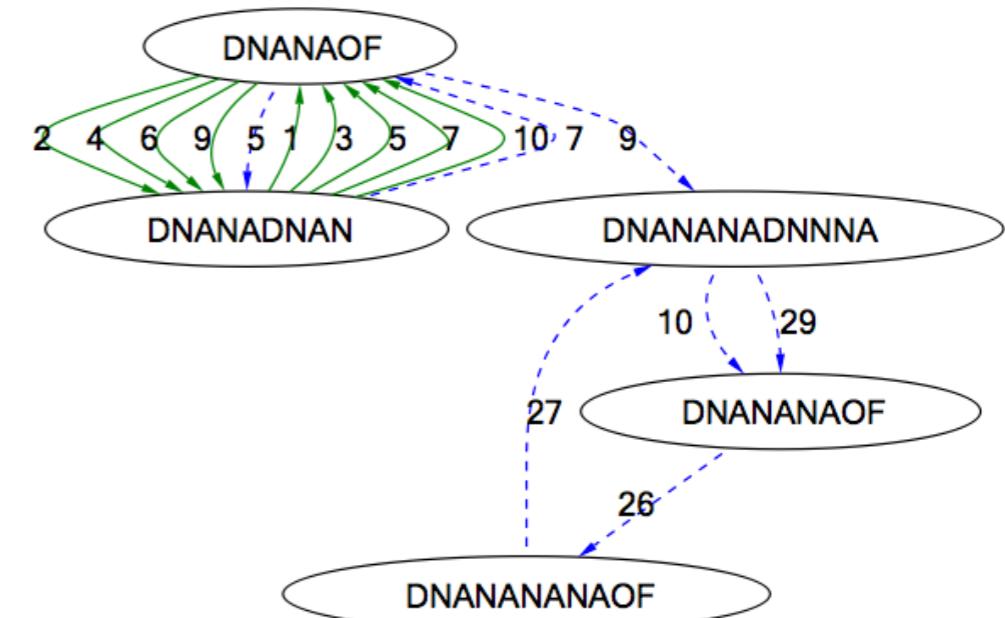
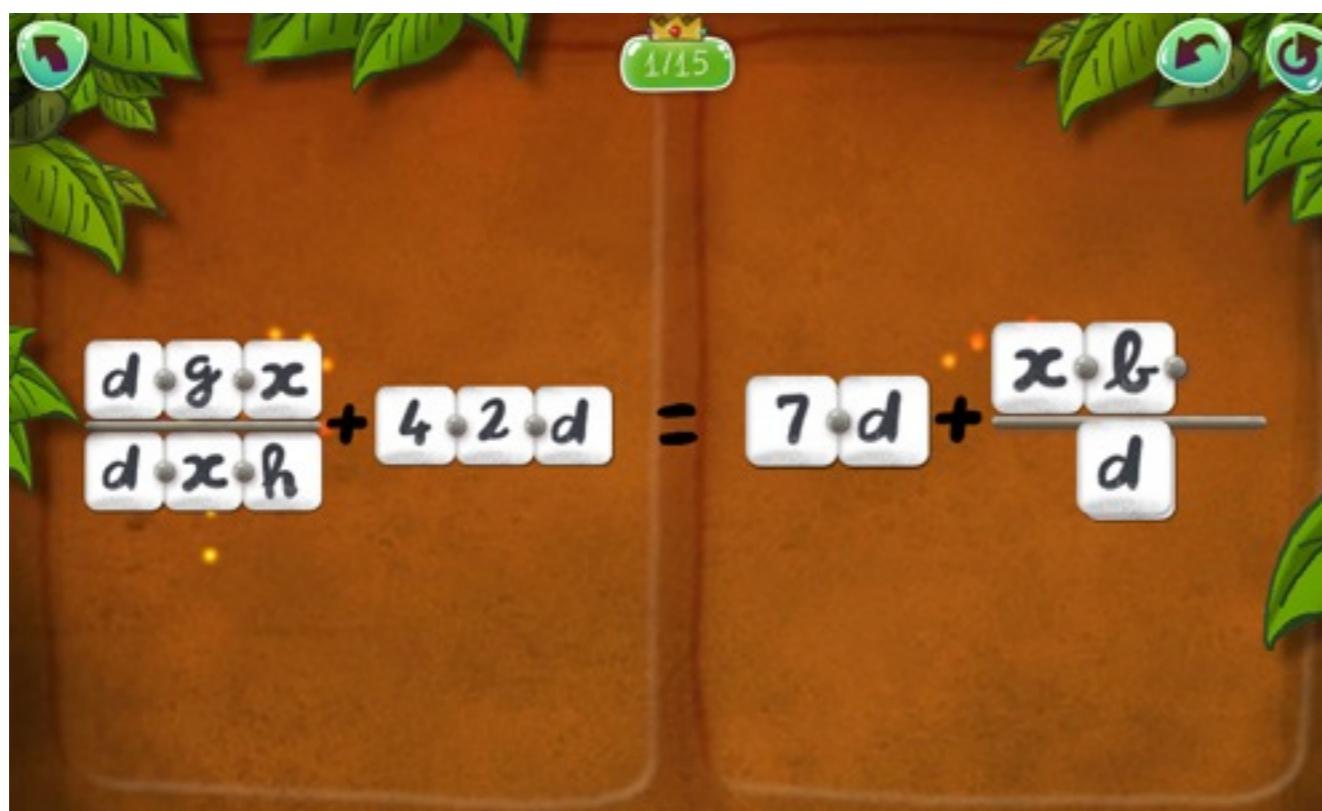
Compare models of **how** people **know**



Model Tracer 2.0

DragonBox

bringing math to life!

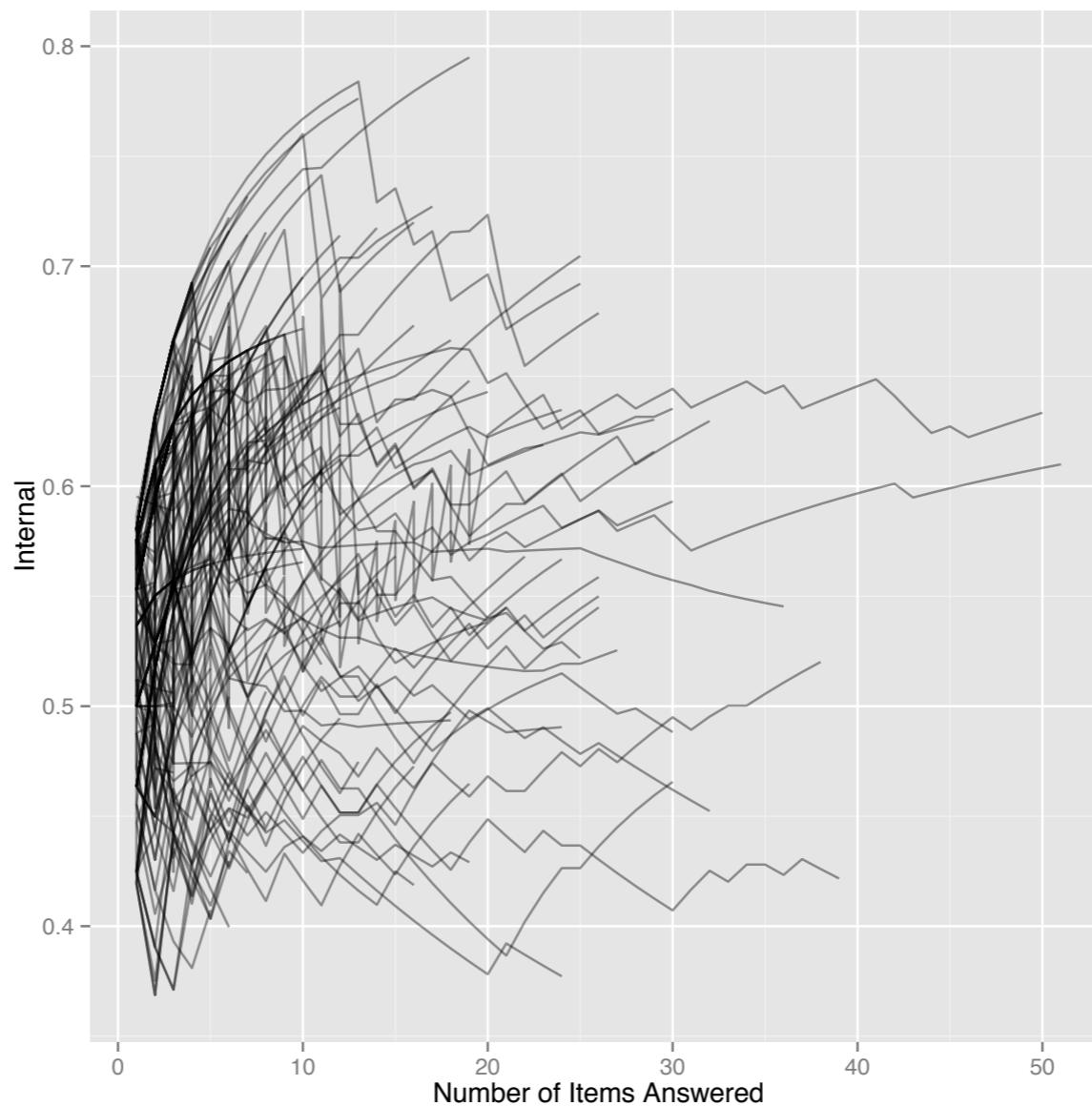


Popovic (2013)

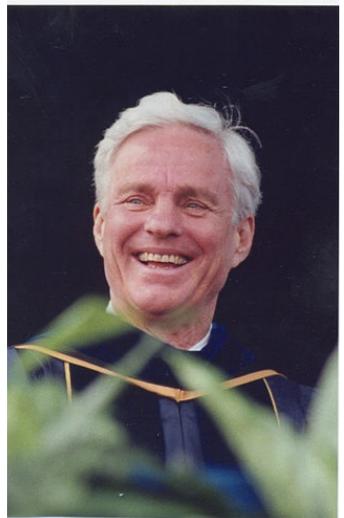
<https://www.youtube.com/watch?v=V1z68nGI3F4>

Knowledge Tracer

Measure what people **know**



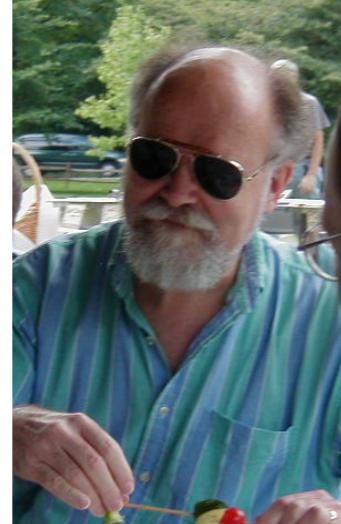
Bayesian Knowledge Tracing



Atkinson



1967



Corbett Anderson



1993



2014

What does BKT do?

Measures student **skill** (Knowledge Component)

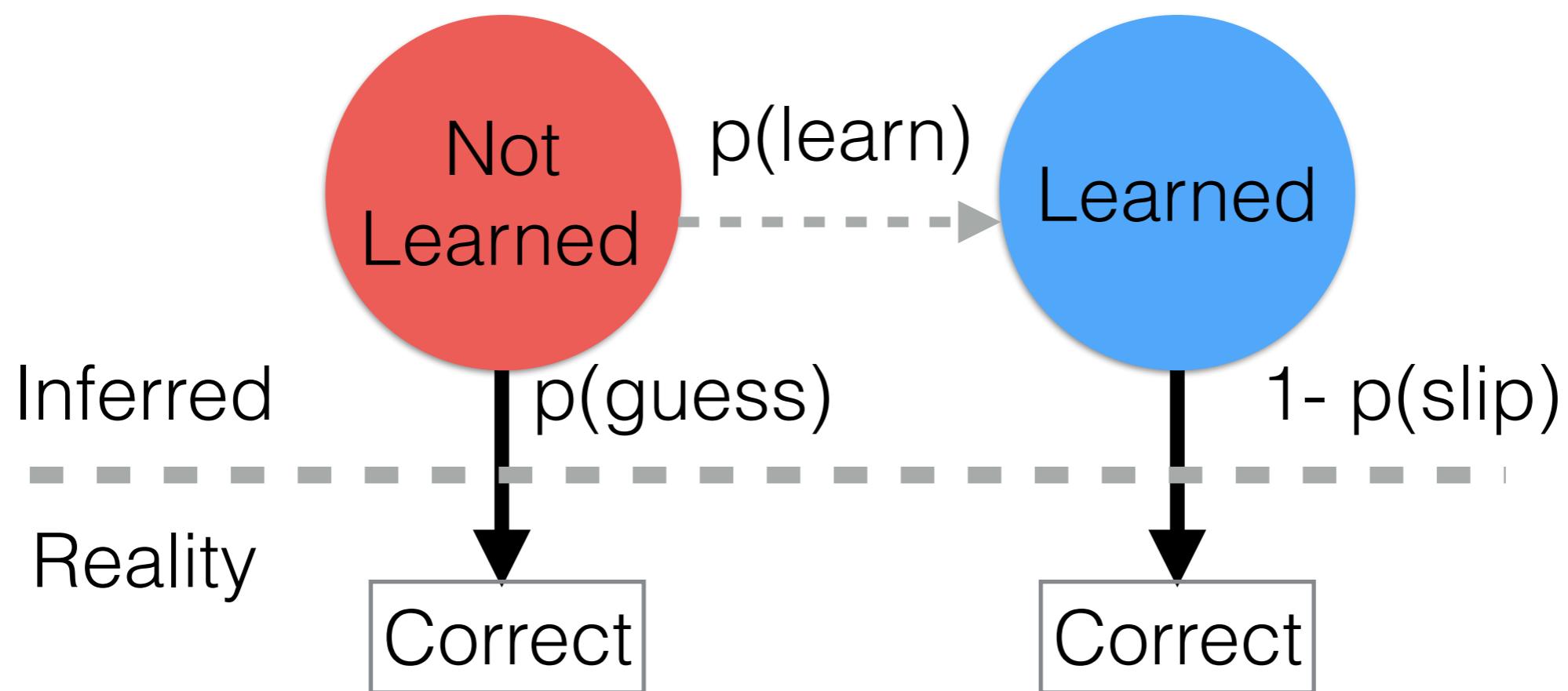
- Skill is *tightly defined*:
 - Division of a two digit number vs. arithmetic
- Measurement is based on a sequence of correct/incorrect answers
- Students are expected to change over time

Assumptions

- Skills are either learned or unlearned
- Each time a student has the opportunity to apply a skill they also have the opportunity to learn that skill
- Once you know something, you don't forget it

BKT Model

- $P(L_n)$ - probability of knowing the skill
- $P(CORR)$ - probability that a student will get an item correct



Example



A Free Public Service of Worcester Polytechnic Institute

What is the difference of $8\frac{2}{4} - 3\frac{1}{7}$?

Be sure to put a space between the whole number and the fraction in your answer. For example the answer should look like this: **6 2/3**. Not like this: 62/3

Type your answer below (mathematical expression):

[Submit Answer](#)

We will not tell you if you are right or wrong.

You are done with this problem!

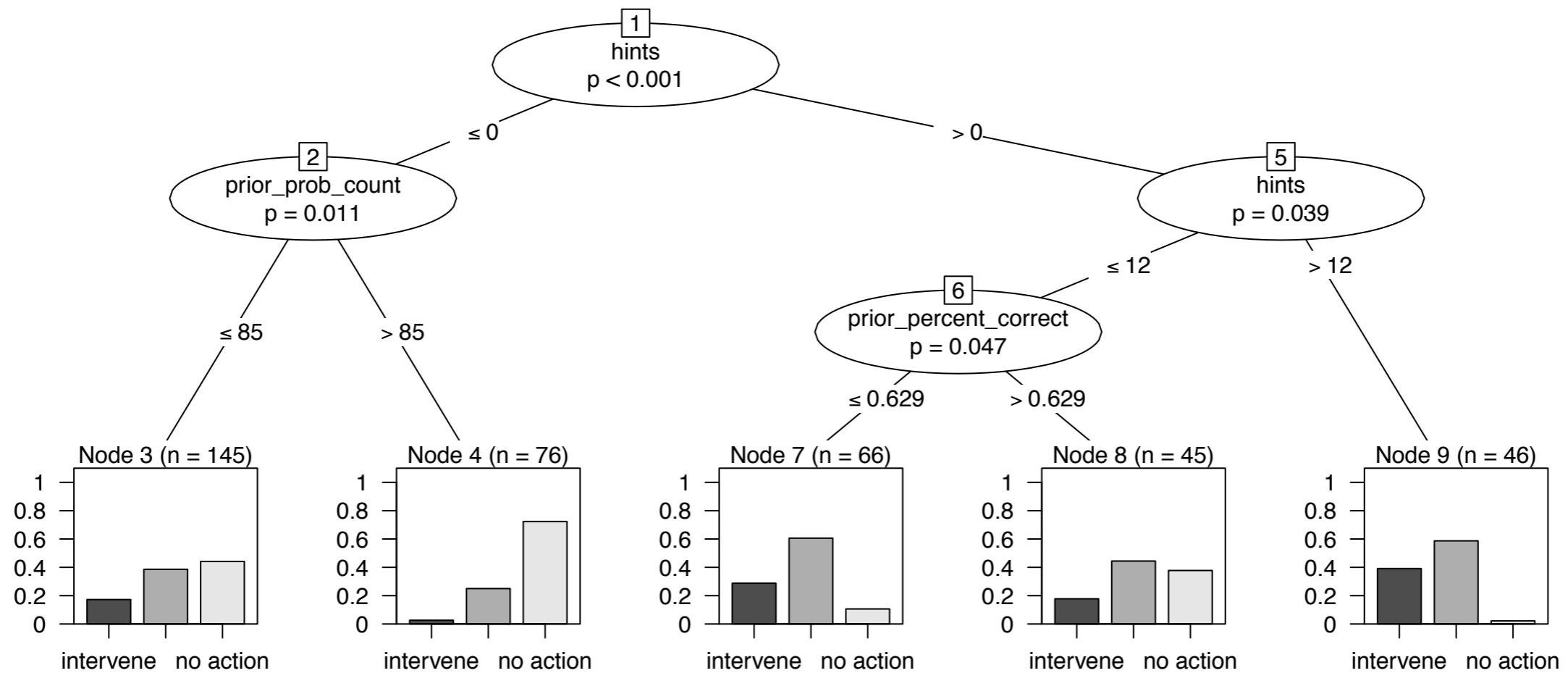
[Go to next problem](#)

Student/Problem [Unanonymize]	Average New! <u>Data driven</u>	#254835 <u>Data driven</u>	#254833 <u>Data driven</u>	#254838 <u>Data driven</u>	#254842 <u>Data driven</u>	Total hints
Problem average	63% <u>Data driven</u>	38%	67%	67%	81%	
Help requested percentage		0%	0%	0%	0%	
Common Wrong Answers		70,42% <u>+feedback</u> 80,19% <u>+feedback</u>	C. 12 inches, 100%	1/3, 54%		
<u>XXXXX</u> *	75%	✓ 60	✓ D. 7 inches	✗ 8/15	✓ 22	0
<u>XXXXX</u> *	50%	✗ 55 1 times	✗ C. 12 inches 2 times	✓ 2/3	✓ 22	8
<u>XXXXX</u> *	50%	✗ 70 1 times	✓ D. 7 inches	✗ 8/15	✓ 22	3
<u>XXXXX</u> *	50%	✗ 70 2 times	✗ C. 12 inches	✓ 2/3	✓ 22	6

Volunteer to explain
A7?

Decision Trees

- “Party” = recursive partitioning algorithm
- Statistical stopping rules
- 2 stage algorithm:
 1. Partition
 2. Fit a model in each split



- Can get very computational expensive
- Tends to “overfit”

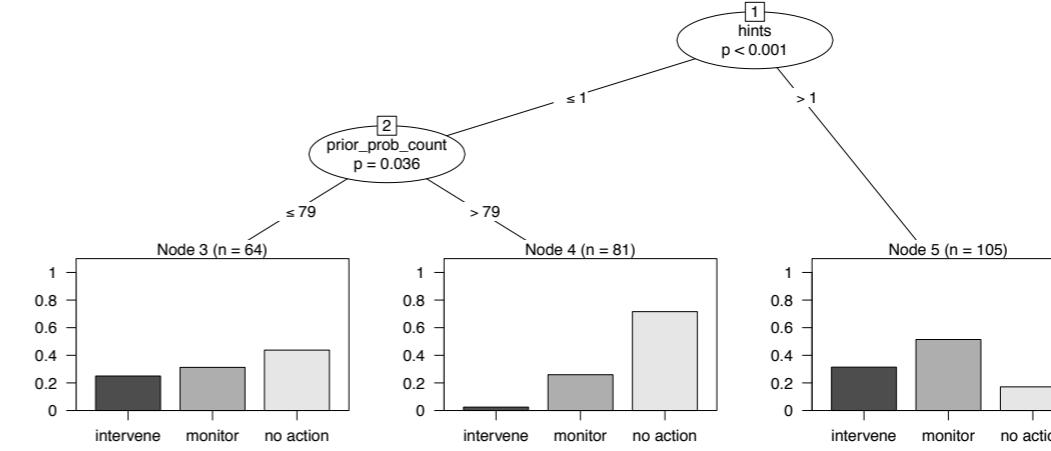
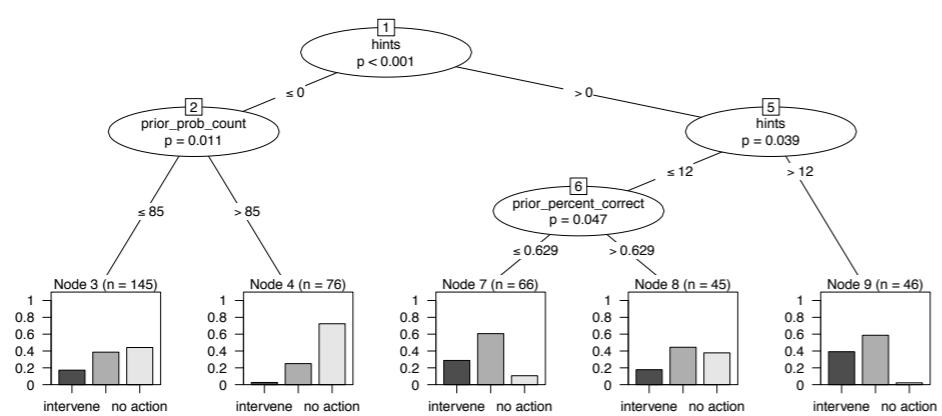
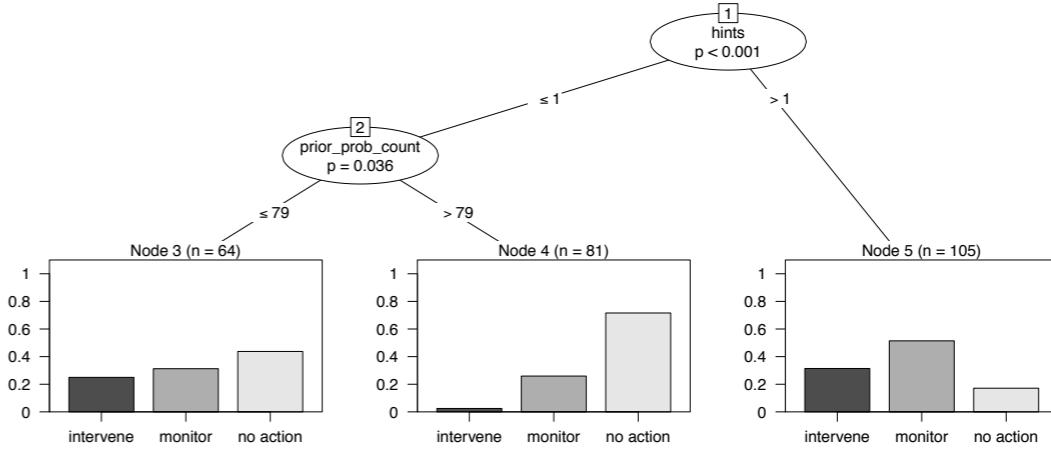
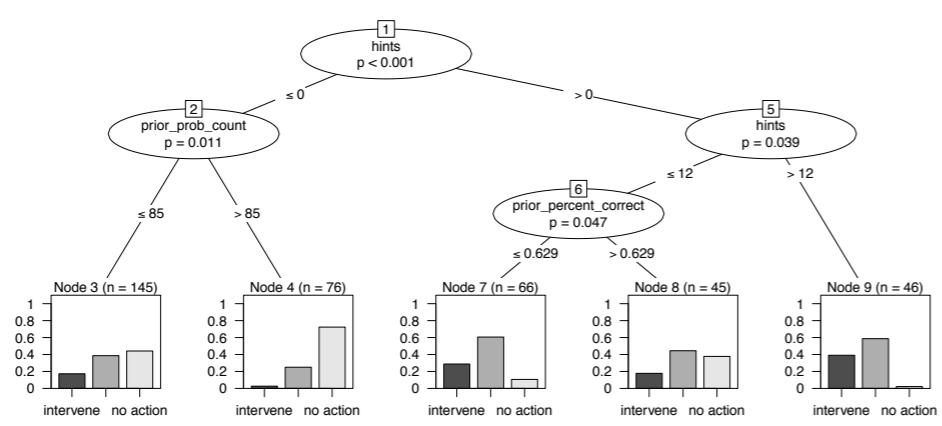
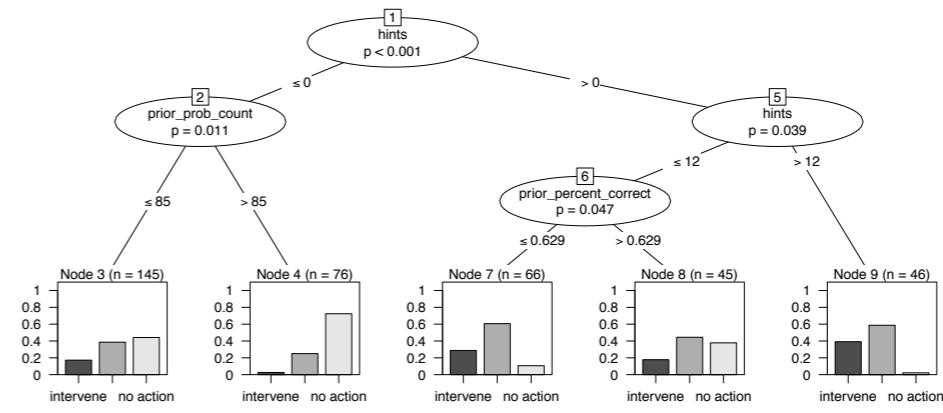
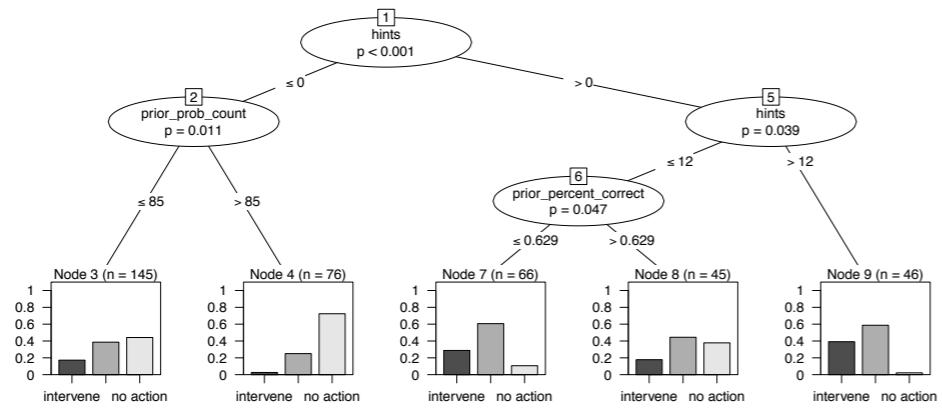
Overfitting

Model describes random error or noise* instead of an underlying relationship

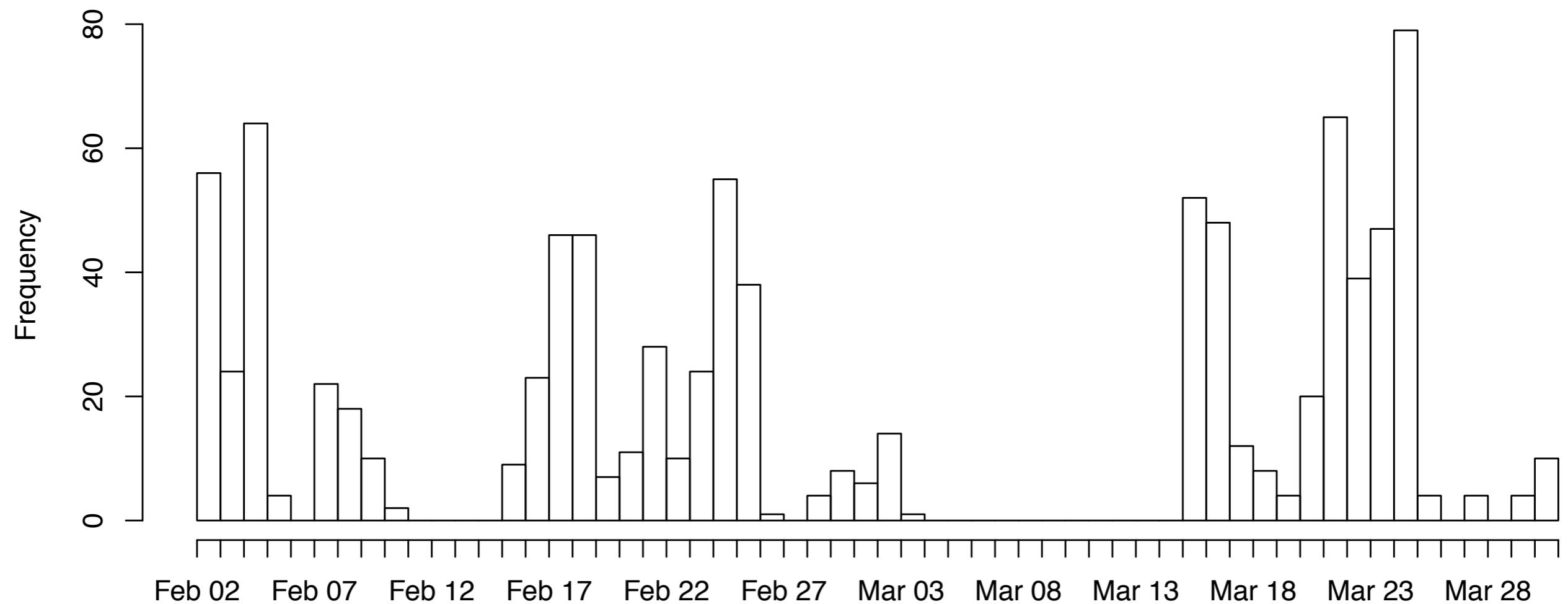
Random Forest[®]

Random Forests

1. Create random subsets of the data
2. Fit model to each subset
3. Determine the modal tree



Tweet Frequency per Day



4/8/16 10:46 AM

DSE16: Reply Frequency per Day

