Preparing Teachers to Help Secondary Students Evaluate Evidence

Andrew Ross, Eastern Michigan University, andrew.ross@emich.edu
Non-presenting co-authors:
Stephanie Casey, Eastern Michigan University, scasey1@emich.edu
Samantha Maddox, Jefferson High School (Georgia), sammaddox22@gmail.com
Melody Wilson, Eastern Michigan University, mwils104@emich.edu

USCOTS 2019-05-17
Friday 9:45am-11:00am

Quick survey/introductions

- Who here teaches/will teach a stats-for-teachers course?
- Who here teaches a regular stats course that has future teachers in it?
- Who here is a K-12 teacher currently?
- Who here is familiar with simulation-based inference?
- Any other thoughts?

Outline

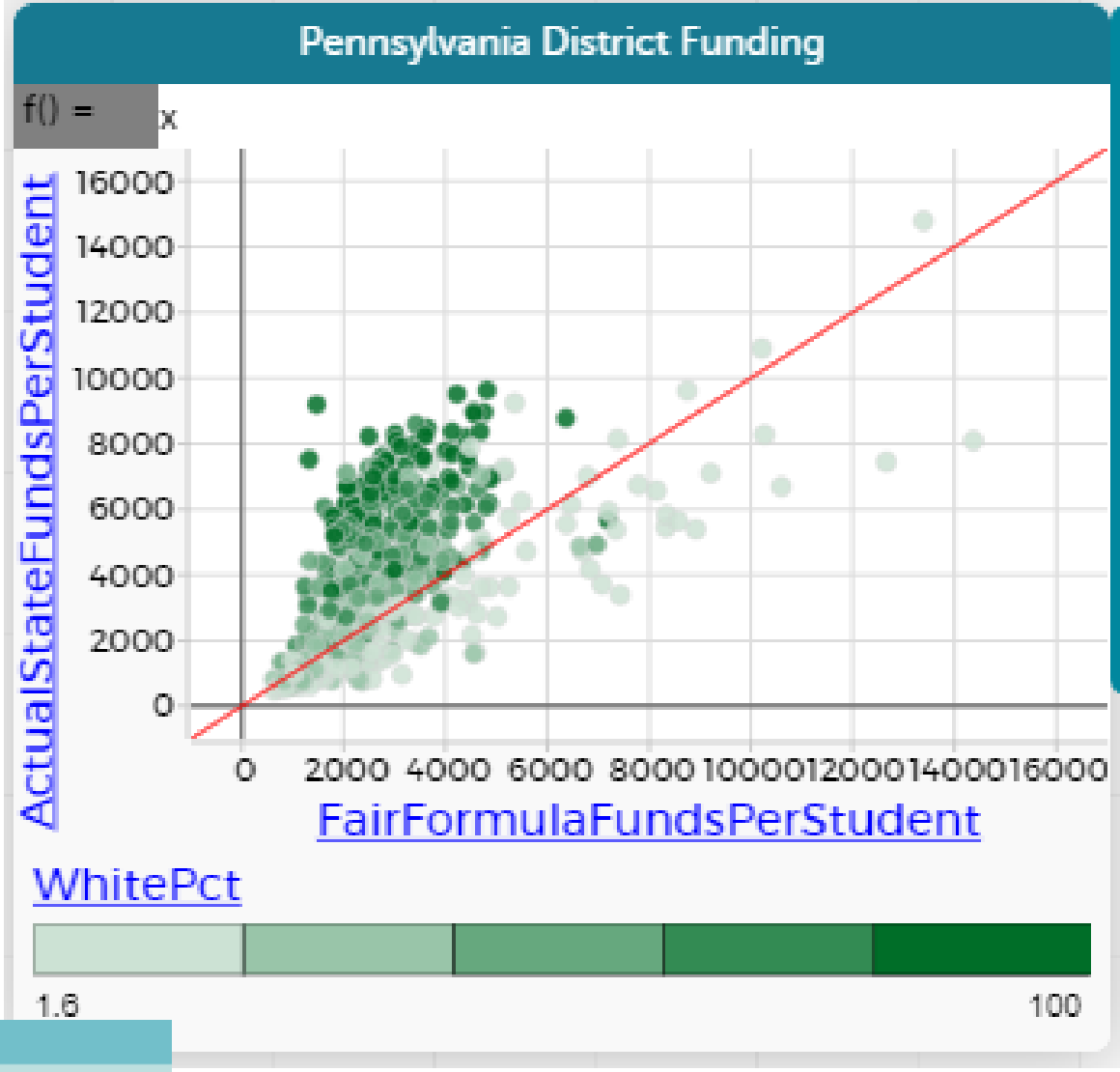
- Setting: Intro-Statistics-for-Future-Teachers undergraduate class
- Activity for a 7^{th}-grade CCSS standard
- Pedagogical thoughts and Discussion of that activity
- Activity for a high-school CCSS standard
- Pedagogical thoughts and Discussion of that activity
- Overall discussion

Setting

- For future middle- and high-school teachers
- Meant as Intro-to-statistics and Intro-to-teaching-statistics
- Could also be used in a capstone course
- Not meant as a “methods of teaching” course
- Has no prerequisites:
  - Not calculus
  - Not curriculum or assessment
  - Not teaching-methods
  - Not college stats

Setting: Stats Topics

- In Module 1, Study Design and EDA, our pre-service teachers see rich multivariate data sets and equity/social-justice topics.
- In Module 2, we focus on 1- & 2-sample inference, and include some equity topics.



PA_district_data_funding2

Cases (499 cases)

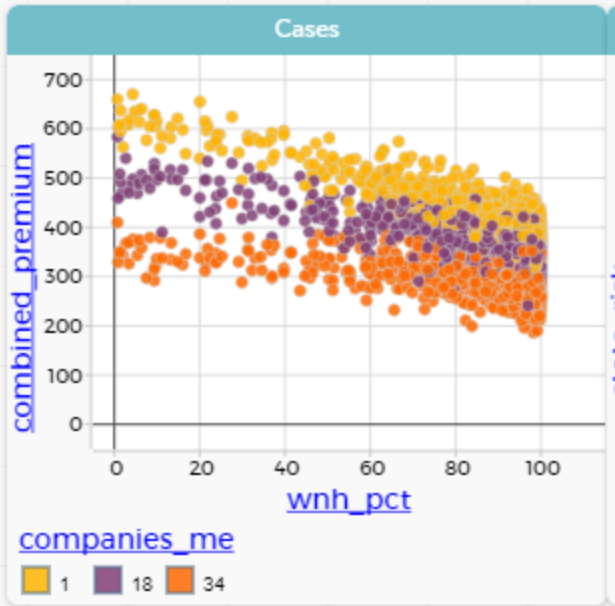
index	AUN	School District	County	MedianHHIncome	Sparsity	PovertyPct 0-99%	PovertyPct 0-184%	Student Count	White Pct	ActualStateFundsPerStudent	FairFormulaFundsPerStudent
494	12954…	Saint …	Schuylk…	40175	0.84	16.1	45.2	739	91.3	4446	3945
495	12954…	Shena…	Schuylk…	31216	0.4	38.5	53.6	1115	56.1	6570	8148
496	12954…	Schuyl…	Schuylk…	50625	0.61	10.4	24.5	1256	92.5	5040	2668
497	12954…	Tama…	Schuylk…	50536	0.72	17	31.5	2125	92.4	3288	2295
498	12954…	Tri-Vall…	Schuylk…	50242	0.87	6	35.8	918	96.8	4958	2423
499	12954…	Willia…	Schuylk…	45614	0.81	17.1	43.6	1068	94.4	6620	3254

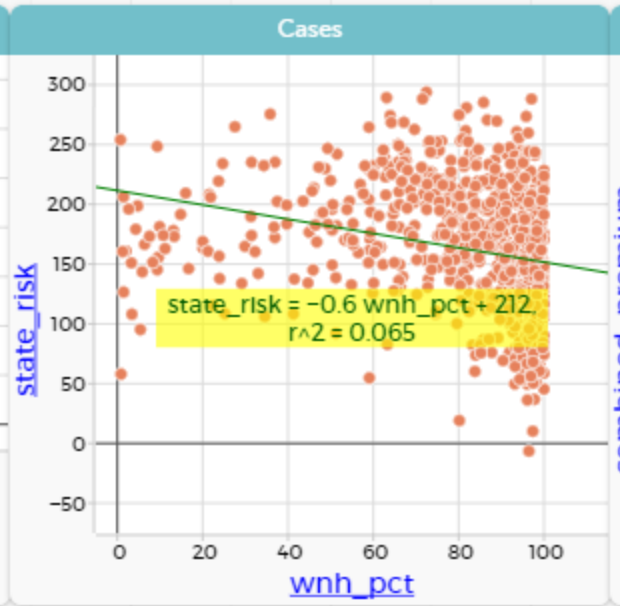
Module 3: Association, includes car insurance price patterns by %white in zipcode in Illinois

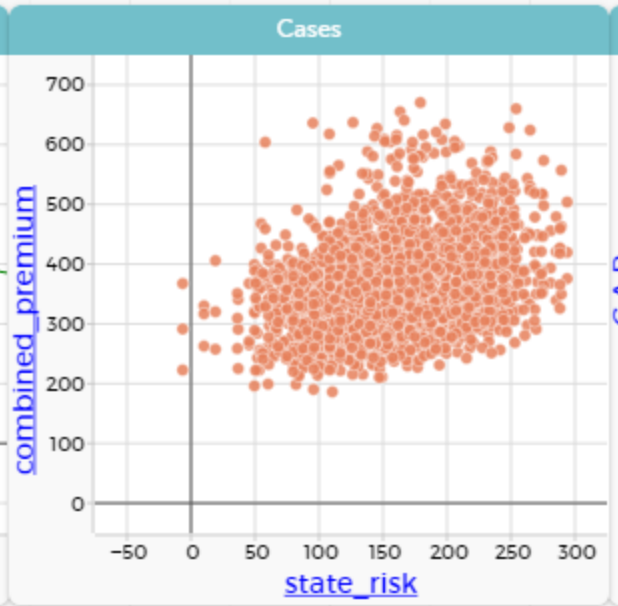
car_insurance_risk_race_3companies

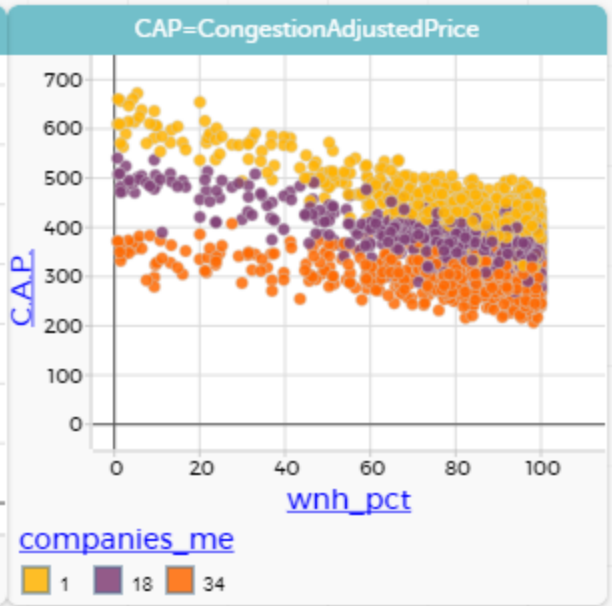
Cases (2622 cases)

index	rownum	zipcode	wnh_pct	total_pop	med_age	med_income	poverty_rate	mean_c..._minutes	aland_sqmi	awater_sqmi	lat	lon	companies_me	parent_corp_me	state_risk	combin..._premium	C.A.P.	randseed	IsArtificial
2617	2617	62298	97.88	16534	43.3	72462	4	15.22	144.21	1.45	38.31	-90.16	18	9000	243.71	375.36	336.83	1849	TRUE
2618	2618	62298	97.88	16534	43.3	72462	4	15.22	144.21	1.45	38.31	-90.16	34	15000	243.71	352.17	318.84	1849	TRUE
2619	2619	62298	97.88	16534	43.3	72462	4	15.22	144.21	1.45	38.31	-90.16	1	1000	243.71	447.48	402.87	1849	TRUE
2620	2620	62301	88.53	33024	37.8	37347	20.6	6.06	11.3	0.54	39.93	-91.39	1	1000	229.65	418.89	381.94	1849	TRUE
2621	2621	62301	88.53	33024	37.8	37347	20.6	6.06	11.3	0.54	39.93	-91.39	18	9000	229.65	402.99	371.07	1849	TRUE
2622	2622	62301	88.53	33024	37.8	37347	20.6	6.06	11.3	0.54	39.93	-91.39	34	15000	229.65	305.13	277.52	1849	TRUE









MODULE(S^2)

Mathematics Of Doing, Understanding, Learning and Educating for Secondary Schools

The Mathematics Of Doing, Understand, Learning, and Educating Secondary Schools (MODULE(S2)) project is made possible through funding from the National Science Foundation IUSE (Improving Undergraduate STEM Education) multi-institutional collaborative grant #1726707 (APLU), #1726098 (University of Arizona), #1726252 (Eastern Michigan University), #1726723 (Middle Tennessee State University), #1726744 (University of Nebraska - Lincoln), and #1726804(Utah State University).

MODULE(S^2) topics

Materials for:

- Abstract Algebra
- Geometry
- Modeling
- Statistics

(each a full-semester course)

Call for Piloters

- Interested in piloting stats-for-future-teachers? Or a similar course on mathematical-modeling for future teachers?
- 4-day professional development meeting in summer
- Online professional learning community during the year
- Travel/hotel costs plus $4000 stipend
- Also available (professional development in June 2020):
  - Abstract Algebra for future teachers
  - College Geometry for future teachers
- Contact: Andrew Ross, andrew.ross@emich.edu

A 7th-grade-level activity: put on the mind of a 7th grader…

Suppose that 3 medicines (A,B,C) each had a clinical trial, with random assignment of research subjects to Treatment (35 subjects) or Placebo (25 other subjects).

Afterwards, each person's blood was tested and the amount of monocytes (a type of immune system cell) was recorded, with higher amounts being better in this situation.

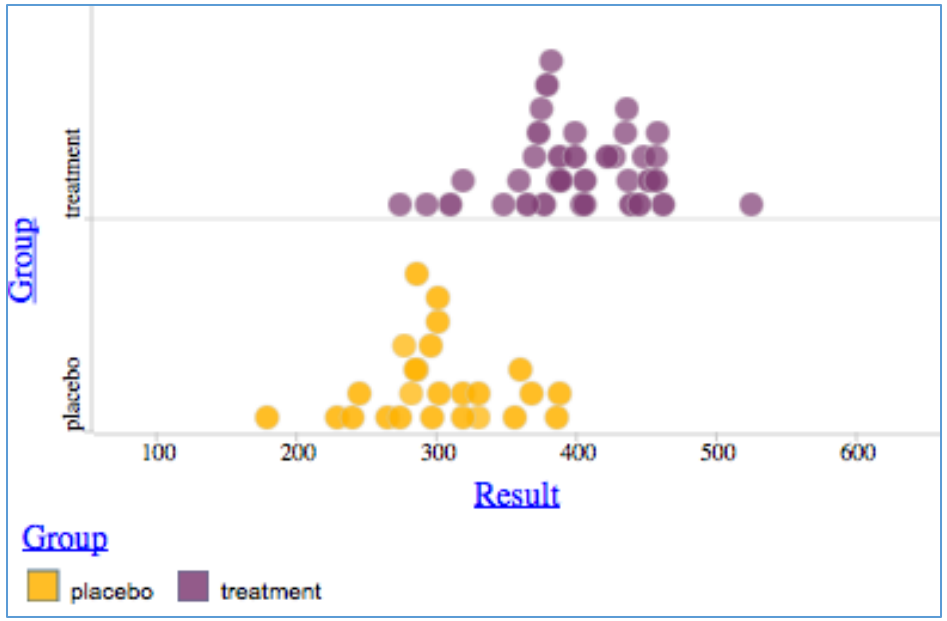
We are wondering:

1. Which study shows the most evidence of an effect: A or B?
2. Which study shows the most evidence of an effect: A or C?

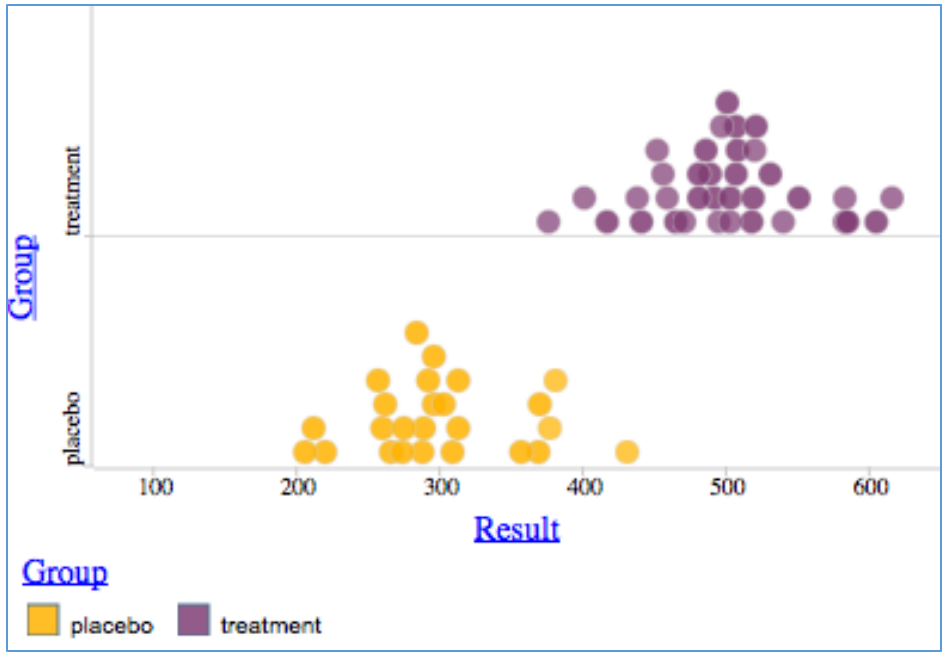
Based on Trumpower (2013) and Trumpower (2015)

Data Glimpse: A vs B

Medicine A

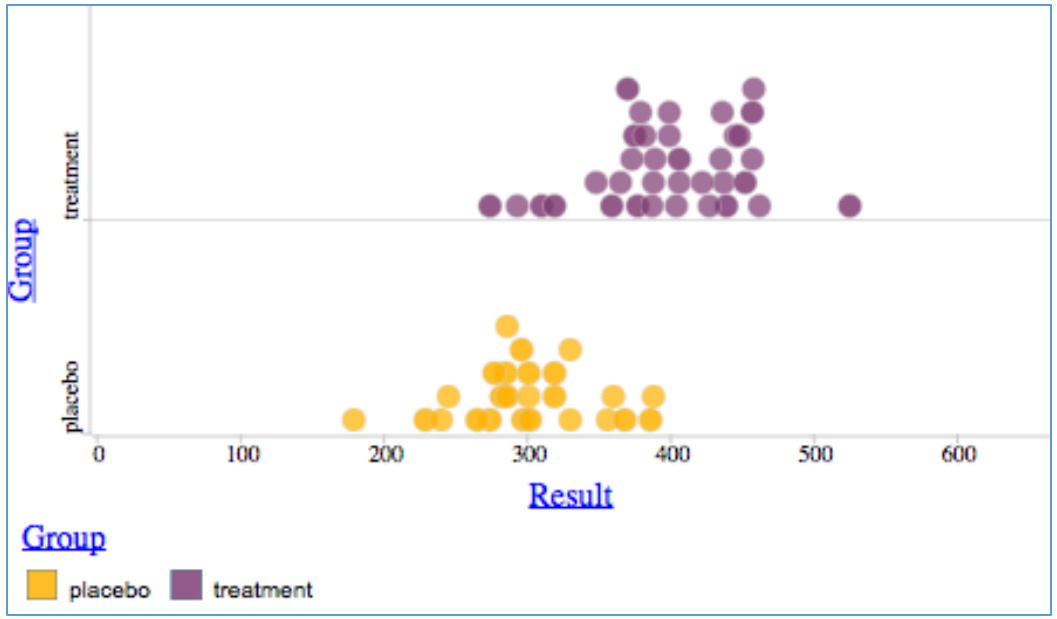


Medicine B

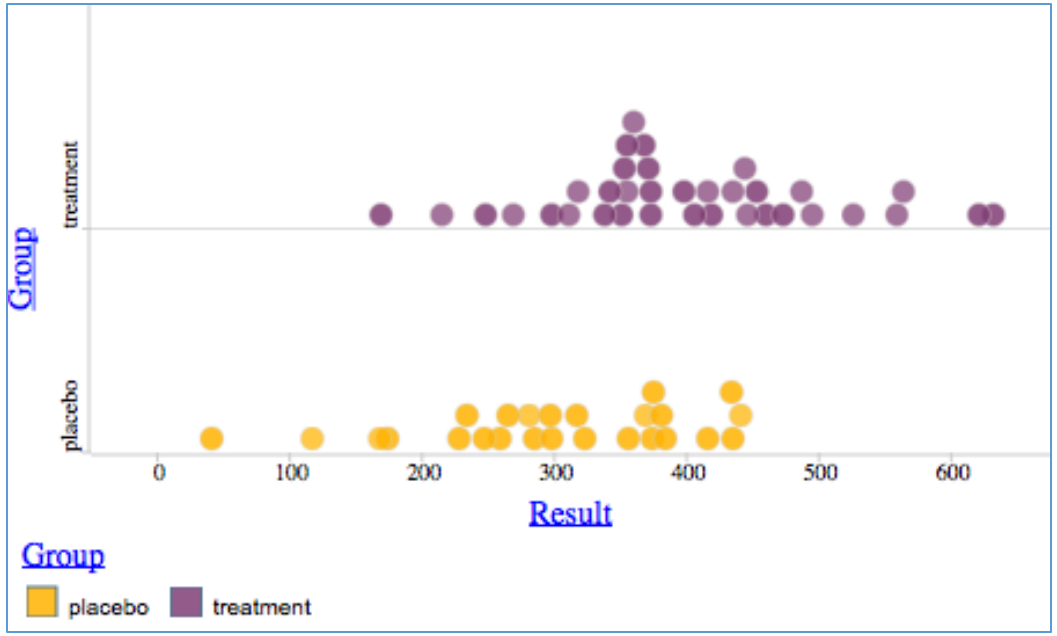


Data Glimpse: A vs C

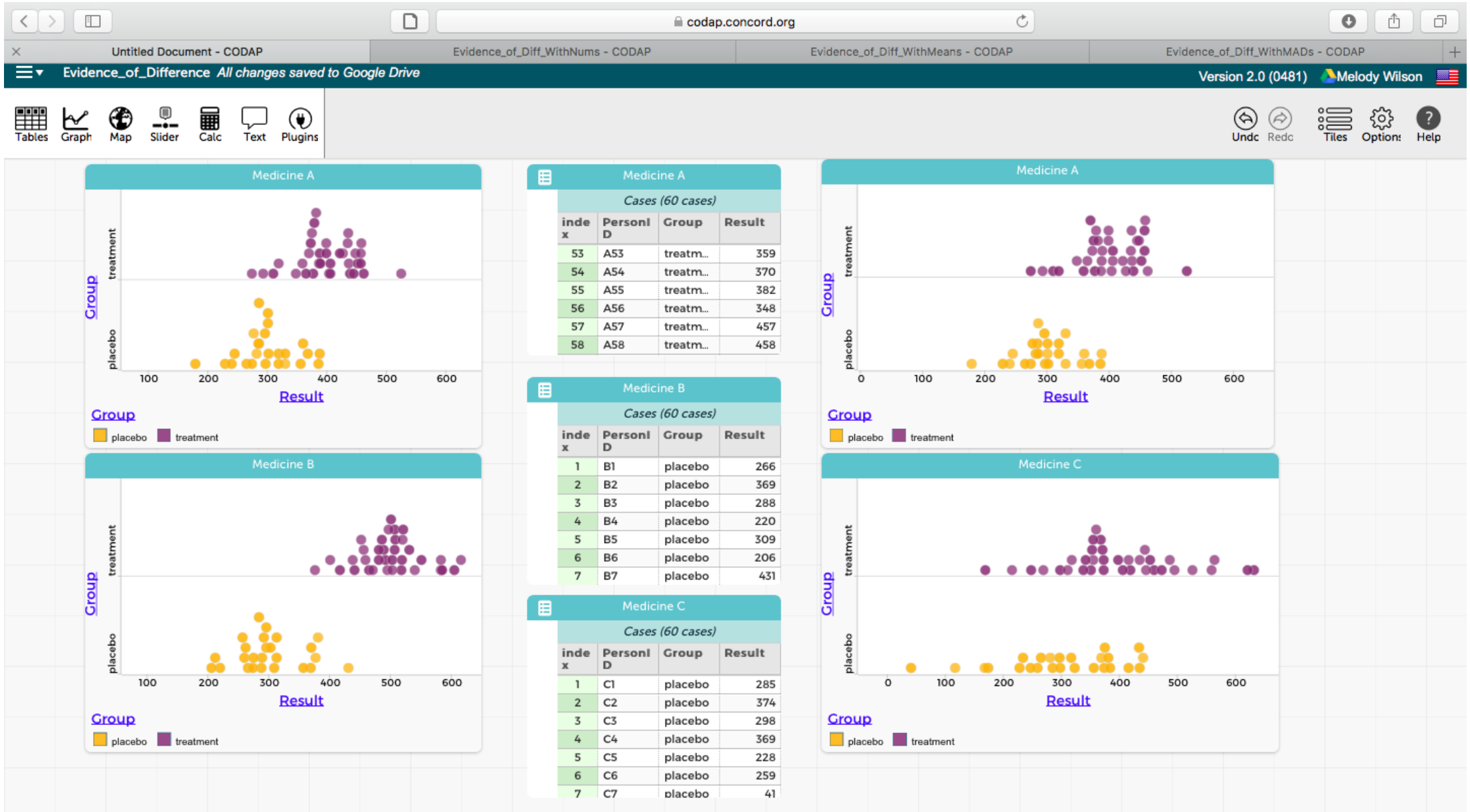
Medicine A



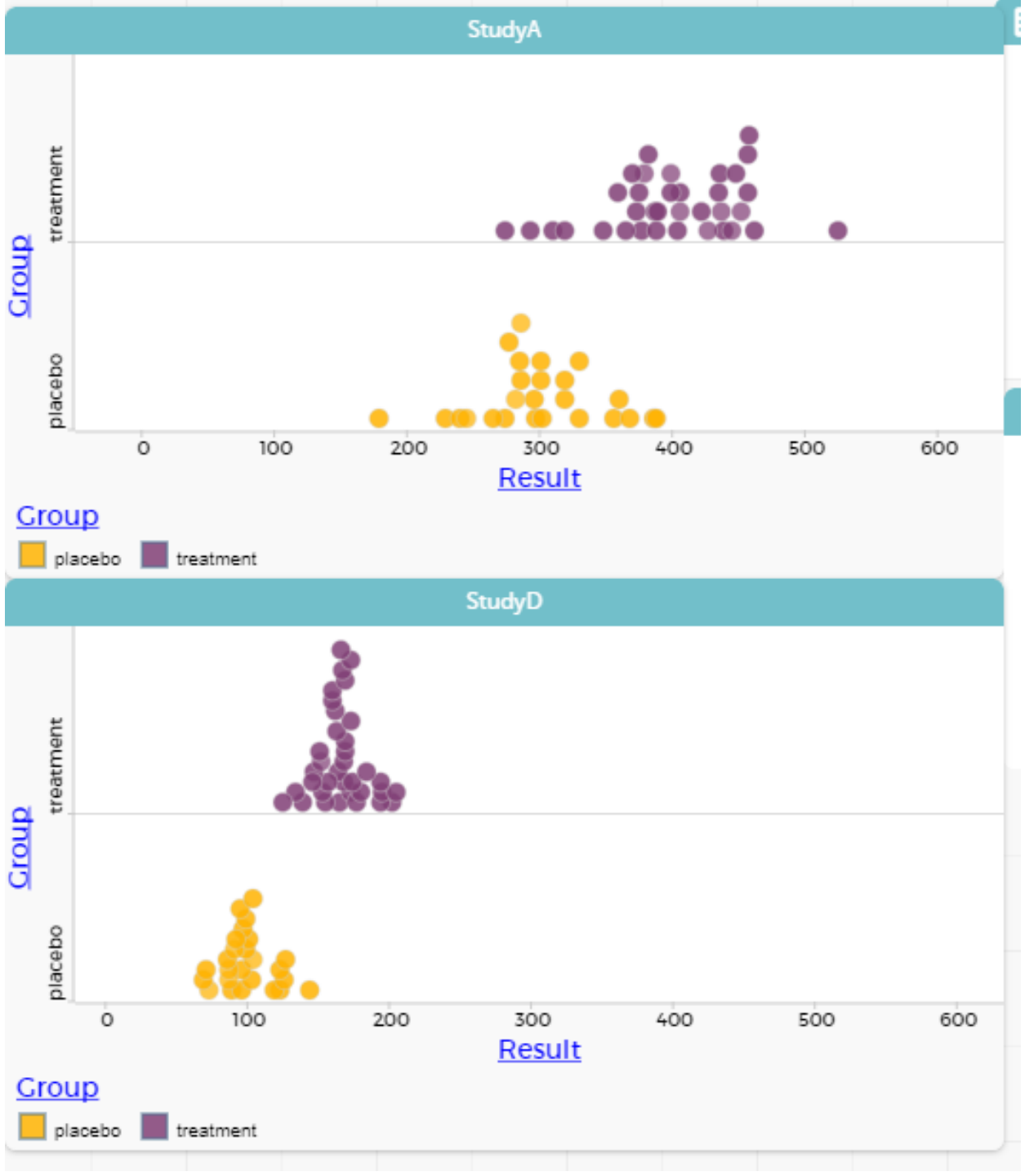
Medicine C



1. Which study shows the most evidence of an effect: A or B?
2. Which study shows the most evidence of an effect: A or C?



Consider A vs D?

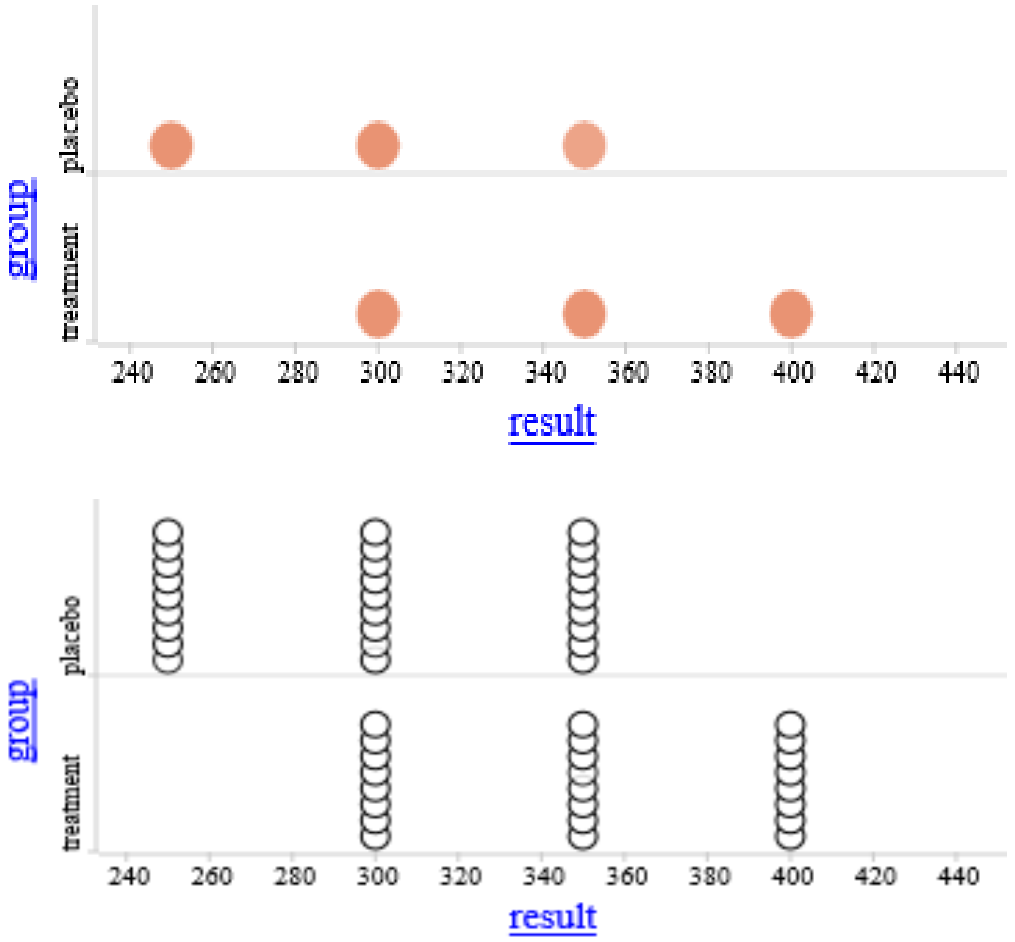


Objective way to judge?

- 3. Based on that reasoning, come up with a statistic that will give a measurement of how much evidence of an effect there is between two groups, and apply it to each of the medicines in the table. Hopefully it gives an ordering that is consistent with your ordering from the previous questions. If it doesn't, then revise your statistic.

Extending…

4. Consider these two dotplots. Which one shows more evidence of a difference between the groups? What does your formula from above say, and do you believe it? Discuss.



Knowledge of Students and Curriculum

- 5. What pre-requisite knowledge is needed for students to be able to engage this topic? In what ways does this topic require integrating different ideas (or not)?

7th grade standard

Draw informal comparative inferences about two populations.

- CCSS.MATH.CONTENT.7.SP.B.3
  Informally assess the degree of visual overlap of two numerical data distributions with similar variabilities, measuring the difference between the centers by expressing it as a multiple of a measure of variability. *For example, the mean height of players on the basketball team is 10 cm greater than the mean height of players on the soccer team, about twice the variability (mean absolute deviation) on either team; on a dot plot, the separation between the two distributions of heights is noticeable*.

Homework: Putting it all together. Thinking just of 7th-grade material…

Exercise 16-4: Suppose that one team (A) of researchers try a new curriculum for adding fractions on 4 students, and have a control group of 5 students that use the old curriculum. They give a 40-question test at the end. Another team (B) of researchers in another US state writes their own new curriculum for adding fractions and tries it on 3 students, with a control group of 6 students. They give a 15-question test at the end. Here are the test score results:

Team A

new	**27**	**35**	**29**	**31**	
control	24	33	32	29	29

Team B

new	**15**	**12**	**13**				
control	15	10	12	13	13	13	12

Which study found more evidence of a difference between the new and control groups? Explain your work/thinking.

Homework: Pedagogical knowledge

Regarding teaching 7th graders:

Exercise 16-5: Suppose that a student says "when we were comparing two groups to see if there was a difference, why did we divide the difference of the means by the spread? Why not just take the difference of the means and be done?" How would you respond, as their teacher?

Discussion so far?

- Next topic: HS activity

Intro Activity for HS-level 2-group analysis

- Want to see the relevant CCSS standard? I will show it on the next slide.
- Or, maybe you want to approach it with “beginner’s eyes”? You could avoid reading the next slide.

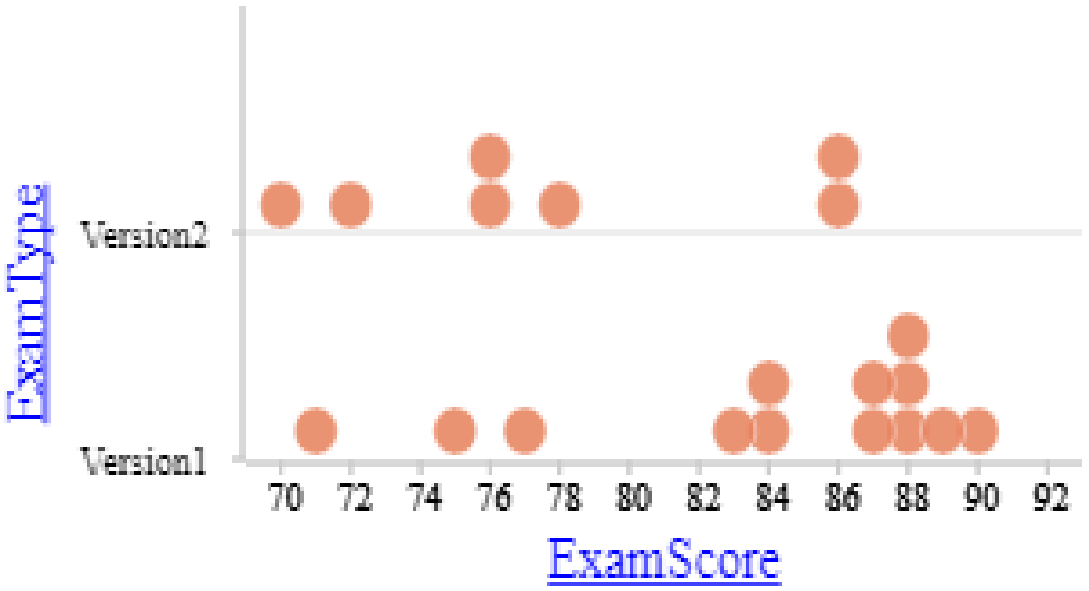
High school standard

Make inferences and justify conclusions from sample surveys, experiments, and observational studies

- CCSS.MATH.CONTENT.HSS.IC.B.5
  Use data from a randomized experiment to compare two treatments; use simulations to decide if differences between parameters are significant.

HS-level activity

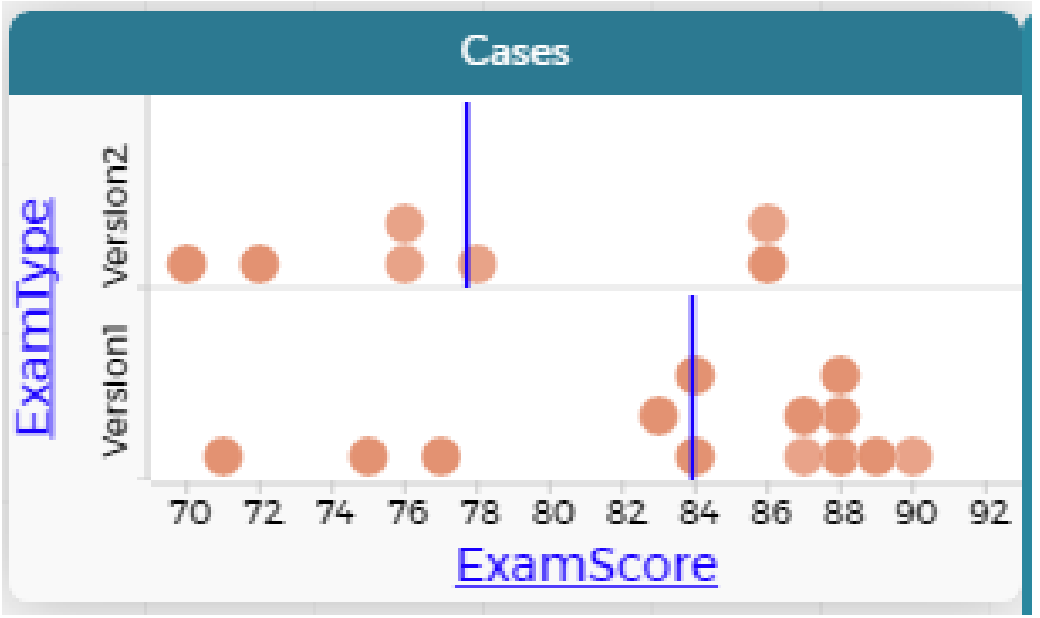
- Suppose you are teaching a high school class and decide to give 2 versions of an exam to help prevent cheating. 13 students were given version 1, and 7 students were given version 2, with the versions being assigned randomly. The resulting exam scores are shown below.
- http://bit.ly/CODAPtwoexams00



Version 1 scores: 71 77 88 88 75 87 84 87 83 88 84 90 89
Version 2 scores: 76 76 86 86 70 78 72

Do you see much evidence of a difference in the exam scores for the two versions? Explain your thoughts.

Question 1-f [Using the means], how much of a difference is there between the two groups of data shown above?



Question 1-g If you could re-run the test with new random assignment of students to test versions, and re-summarize the results, do you think it would still show the same amount of difference between the two versions? Explain.

Question 1-h Based on what we've seen so far, and your intuition, do you think there's a difference between the mean exam score of the two versions generally, not just in our sampled data? Explain.

Question 2-a If the two versions of the test really made no difference, what would the difference of the mean scores be, in the long run (if we could repeat it with many different groups of students at the same level)?

Question 2-b What kind(s) of value(s) would we expect for the difference of the means “in the short run” (for just one class, with 20 students), again if the two versions of the test really made no difference in the long run?

Here is a method that we can use to judge the evidence about whether the two groups (test versions) make any difference overall. We'll temporarily assume that the two versions of the test have no effect on scores, so it won't matter if we mix up which version each high school student got, then simulate what might happen. Each person in our class today should…

1. Start with as many index cards as there are data values, then write one data value on each index card.
2. Mix all those cards together in a random order/shuffle them.
3. Deal out 7 cards into one pile and 13 into another pile.
4. Take the mean of the values in each pile separately, then subtract those two means. (technology hints on taking the mean…)
5. Write that difference-of-means on a post-it (please write it big so it's easily visible) and contribute it to the class dot-plot.
6. Repeat from step 2, if time remains.

Question 2-d Why do we mix the cards from the 2 groups together into one pile?

Question 2-e Why do we deal out 13 cards into one pile and 7 into another?

Question 2-f What does each post-it note represent?

Question 2-g If the two exams were equally hard, does it seem plausible that we would see the difference in means that we saw in our original data? Explain your thinking.

Examining Class Answers in Writing, or Class Constructs an Answer

Question 2-g If the two exams were equally hard, does it seem plausible that we would see the difference in means that we saw in our original data? Explain your thinking.

Just having them say their answers out loud didn't work well. This next conversation took place after the instructor asked them to write down their answers, attending to the wording.

Audio Clips

- Student who almost gets it, but then gets distracted
- Student who makes a link with another math subject entirely!

Instructor: *So, what do you think? If the 2 exams were equally hard, does it seem plausible that we would see a difference of means of 6.2?*

Nicki: So, what I answered is I said that no, um, I said I mean although of course it's possible, but after looking at what we did up on the board here, no one was close other than, like, the two towards the right that were close to –

Instructor: *4.7 and 5.44?*

Nicki: Right. – were even close to having a 6 point difference, um, so based on what we did, um, no one really had that big of a difference. But at the same time, I mean, we just randomly took test scores, no one actually took a test and – you know what I mean? – so that, alone, isn't …

Instructor: *Ok.*

Question 2-h Why did we do the simulation?

Question 2-i What did the simulation tell you?

Question 2-j How does the simulation relate to the original context of the study?

Question 2-k Does our simulation assume that whether a person is assigned to version 1 or version 2 of the test has no effect on anyone else's scores? Explain.

Question 2-l What could we do to improve on our index-card simulation?

Henry: Well, you kind of have to give them like another trial run test, like the class, because you're assuming that these students get the same scores taking that other test, but we don't know if that test is harder. That's what we're trying to figure out.

Instructor: *Hmm. But we're temporarily supposing it was equally hard? Is that right?*

Henry: Yeah.

Instructor: *Ok. Is that ok, to temporarily suppose that it's equally hard?*

Stuart: That's the same – the same as –

Clair: Proof by contradiction.

Stuart: This is literally the same language we would use in Proofs class.

Clair: This is great! Proof by contradiction.

Instructor: *Proof by contradiction?*

Clair: If they were equally hard, this is what would have happened, but they're not, because that didn't happen.

Instructor: *Ok.*

Discussion so far?

- Next topic: HS pedagogy

High school standard

Make inferences and justify conclusions from sample surveys, experiments, and observational studies

- CCSS.MATH.CONTENT.HSS.IC.B.5
  Use data from a randomized experiment to compare two treatments; use simulations to decide if differences between parameters are significant.

CCSS trajectory: HT related

- 7th: Draw informal comparative inferences about two populations.
- CCSS.MATH.CONTENT.7.SP.B.3 Informally assess the degree of visual overlap of two numerical data distributions with similar variabilities, measuring the difference between the centers by expressing it as a multiple of a measure of variability.
- CCSS.MATH.CONTENT.7.SP.B.4 Use measures of center and measures of variability for numerical data from random samples to draw informal comparative inferences about two populations.

- HS: CCSS.MATH.CONTENT.HSS.IC.B.5 Use data from a randomized experiment to compare two treatments; use simulations to decide if differences between parameters are significant.

- How are these the same/related?
- How are they different?
- What is a HS student expected to do that a 7th grade student is not?

CCSS trajectory: CI-related

- Use data from a random sample to draw inferences about a population with an unknown characteristic of interest. Generate multiple samples (or simulated samples) of the same size to gauge the variation in estimates or predictions (CCSS.7.SP.A.2)
- Use data from a sample survey to estimate a population mean or proportion; develop a margin of error through the use of simulation models for random sampling. (CCSS.HSS.IC.B.4)

- How are these the same/related?
- How are they different?
- What is a HS student expected to do that a 7th grade student is not?

Recalling 7th Grade Standards

Exercise 2-4 Here are some Common Core State Standards from 7th grade:

> *7.SP.2*: Use data from a random sample to draw inferences about a population with an unknown characteristic of interest. Generate multiple samples (or simulated samples) of the same size to gauge the variation in estimates or predictions.
>
> *7.SP.4*: Use measures of center and measures of variability for numerical data from random samples to draw informal comparative inferences about two populations.

(a) Which one of them is more applicable to the two-exam-versions situation? Explain why.

(b) In what way(s) does even that standard that you chose not entirely apply to the two-exam-versions situation?

(c) Regardless of that mismatch in part (b), do the best you can to apply that standard to the two-exam-versions situation, by doing appropriate computations and writing a concluding sentence or two. Hint: this does not require a simulation.

(d) Does your analysis in part (c) align fairly well with the results of your class simulation, or are they somewhat different? Quite different? If they are different, speculate on why.

Difficulties from Case/Jacobbe paper

- Student Difficulty #1: Distinguishing between a sample and a sampling distribution of the mean
- Student Difficulty #2: Transitioning from a sample level to a sampling distribution of the mean level
- Student Difficulty #3: Distinguishing between simulation and replication

- Will zoom in on #2 next.

Prompt

Task 2: Oil and Blood Pressure (Tabor et al., 2012, p. 245)

In a study reported in the *New England Journal of Medicine* (Knapp & Fitzgerald, 1989), researchers investigated whether fish oil can help reduce blood pressure. Fourteen males with high blood pressure were recruited and randomly assigned to one of two treatments. The first treatment was a four-week diet that included fish oil, and the second was a four-week diet that included regular oil. At the end of the four weeks, each volunteer's blood pressure was measured again and the reduction in diastolic blood pressure was recorded. The results of this study are shown below. Note that a negative value means that the subject blood pressure increased.

Fish oil	8	12	10	14	2	0	0
Regular oil	-6	0	1	2	-3	-4	2

What Makes for a Good Teacher Response?

1. Asking questions is often better than making statements.
2. It can be helpful to give an example of the student's reasoning but taken to extremes (either numerical extremes or context extremes where the answer should be clear).
3. Simply telling the student the right way to think is often not as helpful as one might hope—they need to see why their reasoning isn't correct.
4. High quality responses to student work:
   a) Move students toward the student learning objective;
   b) Draw on and are consistent with the student thinking presented and research on students' mathematical development; and
   c) Leave space for student's future thinking (not just teacher's thinking).

AP Stats student response

Eva:	I mean I would probably—you would put it on the graph, wouldn't you? You put like each—like this (referring to graph of sample data, R2) … you would put that on the graph. You would just graph what you got a bunch of times.
Interviewer:	So you would have a bunch of graphs that look like this?
Eva:	No, it would be the same graph.
Interviewer:	The same graph. So like what would… Each dot on the graph—what would each dot be?
Eva:	Each dot would be the improvement score.
Interviewer:	Improvement score for a single person?
Eva:	Yeah.

Question 10-g: What difficulty is Eva having?

Question 10-h: What would you say or do to correct or deepen Eva's thinking?

Pedagogical Aspects

- Written Simulation of Teaching Practice: preparing to have fruitful discussions
- Only a part of the method in the “5 Practices for Orchestrating Productive Mathematics Discussion” book

Written Simulation of Teaching Practice

Inference Module – Simulation of Practice Written Assignment

Designing a Simulation for Comparing Two Treatments

Students in your high school math class are learning about comparing two treatments from a randomized experiment.

CCSS.Math.Content.HSS.IC.B.5

> Use data from a randomized experiment to compare two treatments; use simulations to decide if differences between parameters are significant.

You are going to center a discussion-based lesson around the following task:

Does increasing the amount of calcium in our diet reduce blood pressure?

Examination of a large sample of people revealed a relationship between calcium intake and blood pressure. The relationship was strongest for black men. Such observational studies do not establish causation. Researchers therefore designed a randomized comparative experiment.

Written Simulation of Teaching Practice

Designing a Simulation for Comparing Two Treatments (cont.)

The subjects were 21 healthy black men who volunteered to take part in the experiment. They were randomly assigned to two groups: 10 of the men received a calcium supplement for 12 weeks, while the control group of 11 men received a placebo pill that looked identical. The experiment was double blind. The response variable is the decrease in systolic (top number) blood pressure for a subject after 12 weeks, in millimeters of mercury. An increase appears as a negative number. Here are the data:

Group 1 (calcium)	**7**	**-4**	**18**	**17**	**-3**	**-5**	**1**	**10**	**11**	**-2**	
Group 2 (placebo)	-1	12	-1	-3	3	-5	5	2	-11	-1	-3

Use a statistical approach to determine whether there is convincing evidence that a calcium supplement, on average, reduces blood pressure more than a placebo. Then describe any limitations you need to consider when inferring about these treatments.

Source: Starnes, D.S., Tabor, J., Yates, D.S., & Moore, D.S. (2014). The Practice of Statistics (5th edition). New York: W.H. Freeman and Company.

Written Simulation of Teaching Practice

Designing a Simulation for Comparing Two Treatments (cont.)

Provide responses to the following prompts:

1. Provide your solution to the task, including all work.
2. Anticipate how students might approach the task. Include:
   a) An alternate solution to the task (This solution must utilize a simulation if your original solution did not).
   b) An incorrect response to the task, and why it is incorrect
   c) Another incorrect response to the task, and why it is incorrect
   d) At least one response that is essentially correct but imprecise (for example, a response that is imprecise in its use of mathematical language)
3. Describe how you will introduce the task to the class, including familiarizing students with the context for the dataset used in the task.

Written Simulation of Teaching Practice

Designing a Simulation for Comparing Two Treatments (cont.)

4. Write a plan regarding how you will conduct a productive whole class discussion following individual student work time on the task. The discussion should allow you to elicit student thinking about their solutions and move the class toward successfully designing a simulation to compare two treatments. Your plan for the discussion should include both a list of discussion questions (exactly what you would say, do, and/or ask your students to do to further investigate the ideas as the lesson continues) and descriptions of the ways you anticipate that student might respond to these questions. Be sure to present a clear picture of how you envision the lesson will continue, including both student and teacher statements.
5. Give a variant on the initial task (for example, alter the numbers in the two treatment groups) that you could pose to the class that will deepen students' current thinking about using data from a randomized experiment to compare two treatments using simulations (CCSS.Math.Content.HSS.IC.B.5).
6. Solve the variant task.

The Statistics MODULE(S^2)

- Module 1: Study Design and Exploratory Data Analysis
- Module 2: Statistical Inference
- Module 3: Statistical Association

The Statistics MODULE(S^2)

- Module 1: Study Design and Exploratory Data Analysis
- Module 2: Statistical Inference
  - Hypothesis Test via simulation
  - Confidence Interval via simulation
  - HT via formulas
- Module 3: Statistical Association
  - Regression
  - Contingency Tables/Chi-Squared analysis

The Stats Writing Team

- Prof. Stephanie Casey -- former AP Stats teacher, now math-ed faculty
  - the PI for our writing team
- Prof. Andrew Ross -- operations research, stats (no math-ed background)
- Samantha Maddox -- current HS teacher (AP Stats & non-AP stats)
- Melody Wilson -- graduate assistant, former math teacher, studying Social Foundations of Education

Equity and Social Justice

- Trying to model equitable teaching practices (Aguirre, Mayfield-Ingram, Martin, 2013)
- Using statistics to discover issues:
  - Unfair economic mobility for white and black boys growing up rich
  - Tennessee STAR class-size study
  - Different car insurance costs in zip codes according to racial composition
- Using statistics to address issues:
  - Encouraging community-based projects with their future classes
- A large focus of the AMTE standards for teacher preparation

What should Pre-Service Teachers take?

The “Statistical Education of Teachers” document from the ASA calls for HS teachers to take 3 stats courses:

1. An introductory course that emphasizes a modern data-analytic approach to statistical thinking, a simulation-based introduction to inference using appropriate technologies, and an introduction to formal inference (confidence intervals and tests of significance)
2. A second course in statistical methods that builds on the first course and includes both randomization and classical procedures for comparing two parameters based on both independent and dependent samples (small and large), the basic principles of the design and analysis of sample surveys and experiments, inference in the simple linear regression model, and tests of independence/homogeneity for categorical data
3. A statistical modeling course based on multiple regression techniques, including both categorical and numerical explanatory variables, exponential and power models (through data transformations), models for analyzing designed experiments [ANOVA], and logistic regression models

Green/underline=included in our MODULES

The "Statistical Education of Teachers" document from the ASA calls for HS teachers to take 3 stats courses: Our materials include the green topics:

1. An introductory course that emphasizes a modern data-analytic approach to statistical thinking, a simulation-based introduction to inference using appropriate technologies, and an introduction to formal inference (confidence intervals and tests of significance)
2. A second course in statistical methods that builds on the first course and includes both randomization and classical procedures for comparing two parameters based on both independent and dependent samples (small and large), the basic principles of the design and analysis of sample surveys and experiments, inference in the simple linear regression model, and tests of independence/homogeneity for categorical data
3. A statistical modeling course based on multiple regression techniques, including both categorical and numerical explanatory variables, exponential and power models (through data transformations), models for analyzing designed experiments [ANOVA], and logistic regression models

What Did We Leave Out (vs. Stat 101)

- Probability topics:
  - Combinations and permutations
  - Binomial Distribution
  - Bayes' Theorem (medical testing, false-positives, etc)
- Statistics topics:
  - Student's t distribution
  - HT, CI for a proportion or difference-of-proportions (treated as means but with data values of 0's and 1's)
  - Formula-based CI
  - Simpson's Paradox
- Technology
  - Graphing calculators
  - Spreadsheets

Not a “Methods of Teaching Math” course

Only addressing methods implicitly (by trying to model them.) For example:

- Balancing group work, individual work, whole-class discussions, & teacher explanations
- Ways to have a whole-class discussion (e.g. class debate on a scenario where class must choose plan A vs. B; have students present work on behalf of their groups, etc.)
- Choosing group-worthy tasks
- Choosing relevant tasks; letting students’ interests lead the choice of tasks
- Assessment philosophy and practices
- Highlighting valuable strengths of students who are not traditionally considered “smart at math”
- Attending to which student voices are being heard
- Positioning students as doers and thinkers, not just absorbers of information
- Pushing procedure-focused students to justify their answers
- Bridging everyday language to discipline-specific vocabulary
- Addressing student struggles productively
- Classroom management

MODULE(S^2)

Mathematics Of Doing, Understanding, Learning and Educating for Secondary Schools

The Mathematics Of Doing, Understand, Learning, and Educating Secondary Schools (MODULE(S2)) project is made possible through funding from the National Science Foundation IUSE (Improving Undergraduate STEM Education) multi-institutional collaborative grant #1726707 (APLU), #1726098 (University of Arizona), #1726252 (Eastern Michigan University), #1726723 (Middle Tennessee State University), #1726744 (University of Nebraska - Lincoln), and #1726804(Utah State University).

MODULE(S^2) topics

Materials for:

- Abstract Algebra
- Geometry
- Modeling
- Statistics

(each a full-semester course)

Call for Piloters

- Interested in piloting stats-for-future-teachers? Or a similar course on mathematical-modeling for future teachers?
- 4-day professional development meeting in summer
- Online professional learning community during the year
- Travel/hotel costs plus $4000 stipend
- Also available (professional development in June 2020):
  - Abstract Algebra for future teachers
  - College Geometry for future teachers
- Contact: Andrew Ross, andrew.ross@emich.edu

Funding

- Part of the Mathematics Teacher Education Partnership, MTE-P
  - Also working on: active learning in precalc-to-calc2, attracting people to the profession, retention in first year of teaching, and equity/social justice
- The Mathematics Of Doing, Understanding, Learning, and Educating for Secondary Schools (MODULES) project is partially supported by funding from a collaborative grant of the National Science Foundation under Grant Nos. DUE-1726707, 1726804, 1726252, 1726723, 1726744, and 1726098. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Wrapup: Connections

- 7th grade: "difference between the centers … as a multiple of a measure of variability" is essentially "Cohen's d", a measure of effect size.
- Jessica Utt's 5-minute talk last night referred to it.
- "The New Statistics" (G. Cumming) advocates for focusing on effect sizes and not on p-values and statistical significance
- Related to today's keynote "Moving to a world beyond $p < .05$"
- How will moves away from $p<0.05$ affect CCSS (mostly at the HS level)?

Wrapup: Divergences

- Simulation-Based Inference: HS standard
  - If using percentile to get p-value, does not use SD at all!
- Formula-based inference:
  - 7th grade standard calls for essentially “Cohen’s d”
    - (xbar1 - xbar2)/SD(control group) usually
    - Sample size has no influence on this statistic
  - AP Stats/formula-based-Stat-101:
    - t=(xbar1 - xbar2 – H0)/SE
    - Larger sample gives us more power
- So, how does 7th grade standard feed into HS standard? Discuss.

Wrapup

- Other thoughts?

Birds of a Feather lunch discussion today

Teaching Secondary Teachers to Teach Statistics

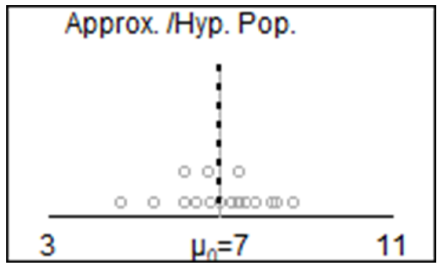
We will talk about preparing K-12 teachers (mostly pre-service, but also in-service) to teach statistics. We can discuss

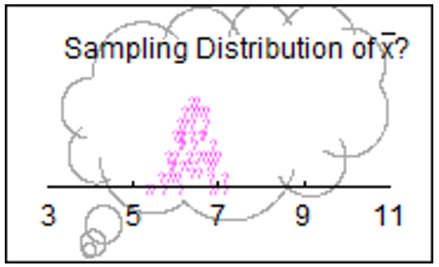
- Materials we use
- How we structure our courses
- How we structure our programs
- the Statistical Education of Teachers (SET) document
- Teachers are (still) skipping or barely skimming the statistics standards in the CCSS-M, largely due to the lack of true stats questions on standardized tests.
- Anything else important

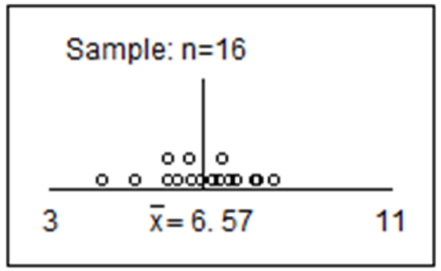
Appendices Start Here

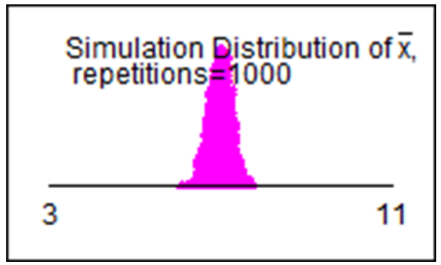
Case/Jacobbe Framework

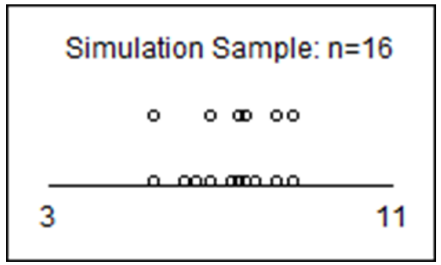
- Scrambled-order graphs:

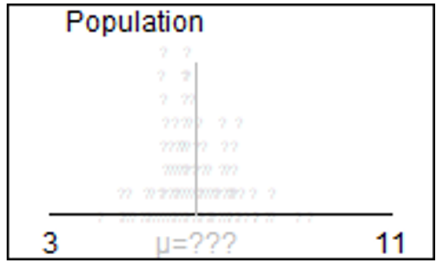








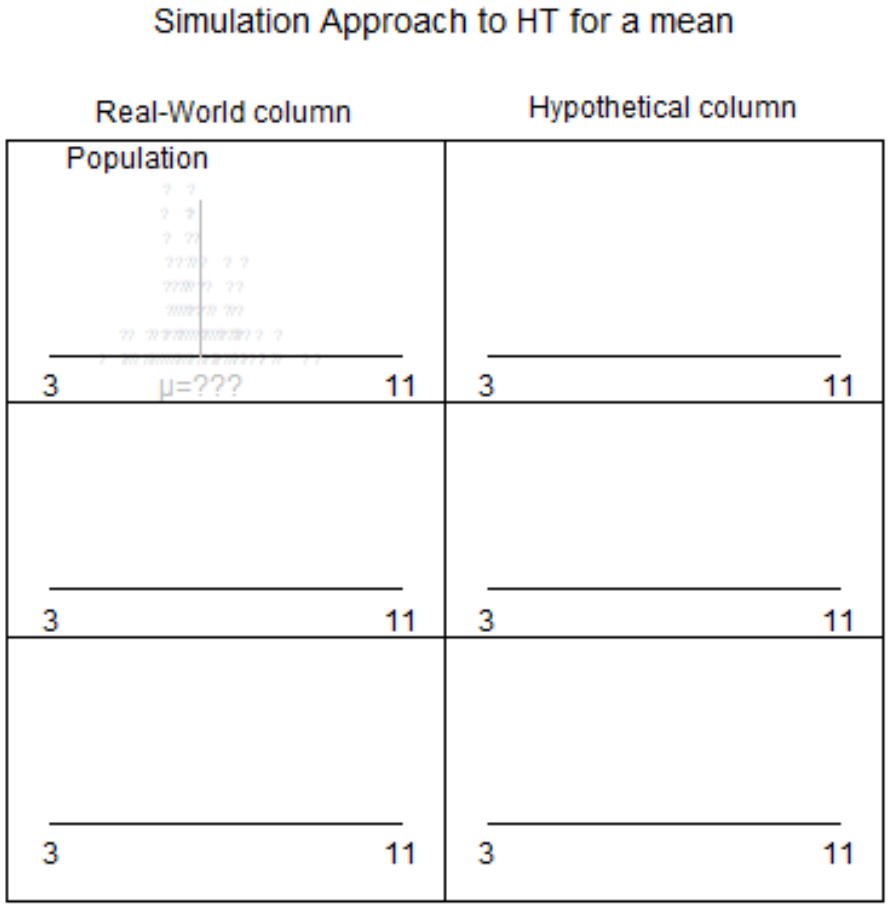




Real-World	Hypothetical
R1	H1
R2	H2
R3	H3

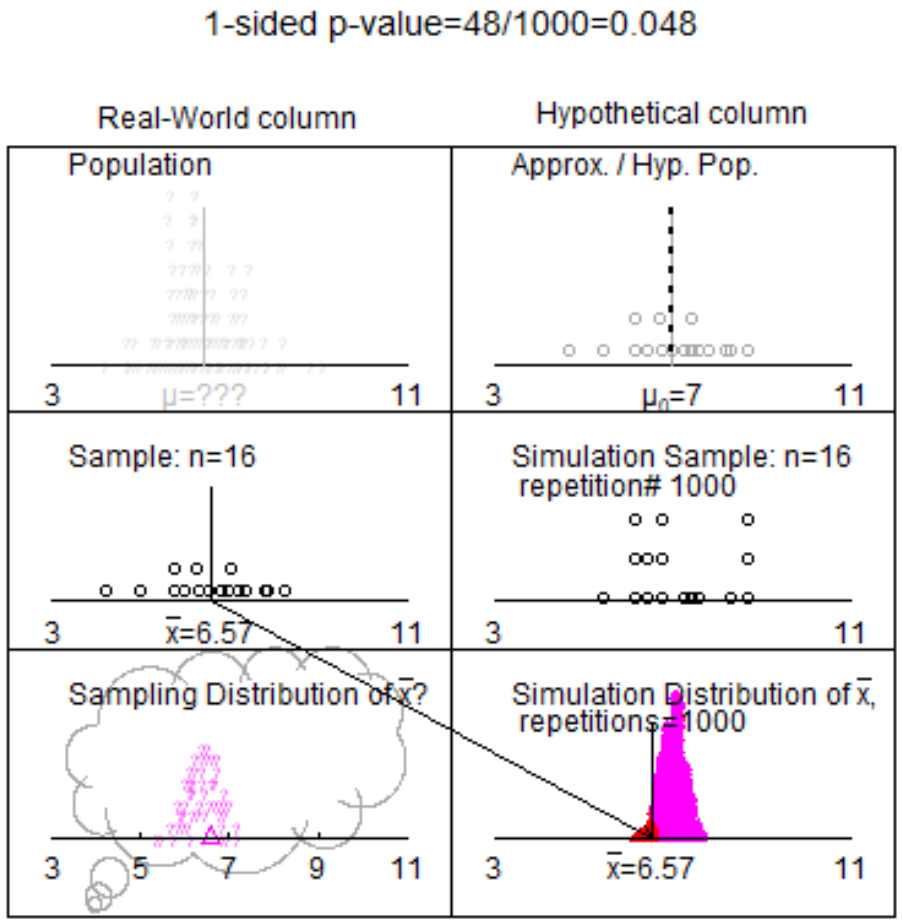
Case/Jacobbe Framework

- Animation at http://bit.ly/animht1m



Case/Jacobbe Framework

- Animation at http://bit.ly/animht1m



Module 1: Study Design and Exploratory Data Analysis

- **Lesson 1: What is Statistics?** (Length: ~90 minutes)

  Activity 1: The Statistical Investigation Cycle

  - Confounding in first example
  - What Is Statistics?
  - The Statistical Investigation Cycle

  Activity 2: Asking Good Statistical Questions

  - Mathematical vs Statistical questions
  - Sources of Variation in Data Values
  - Variation in Summaries of Data
  - Student Struggles with Sampling Variation
  - Statistical Habits of Mind

Module 1: Study Design and Exploratory Data Analysis

- **Lesson 2: Study Design** (Length: ~90 minutes)
  - Activity 3: Random Sampling & Common Forms of Bias
  - Activity 4: Types of Studies
    - Random Assignment, experimental vs observational
    - Confounding
    - Additional Activity: Purpose of Random Assignment

Module 1: Study Design and Exploratory Data Analysis

- **Lesson 3: Equity Conversations and Data Visualizations**
  - Activity 5: Preparing to have Courageous Conversations about Equity
    - Gutstein's 3-C framework: Community knowledge, Critical knowledge, Classical knowledge
    - Classroom norms for discussions
  - Activity 6: Modern Multivariate Data Visualizations
    - Healthy Foods Scatterplot
    - Class Mobility and Race: animated data plot

Module 1: Study Design and Exploratory Data Analysis

- **Lesson 4: Multivariate Exploratory Data Analysis with CODAP**
  - Activity 7: Finding A Story in the Data (I)
    - Graphing Quantitative Variables: SOCS, scatterplots
  - Activity 8: Finding A Story in the Data (II)
    - Graphing categorical variables
    - Plots with 1 Categorical and 1 Quantitative Variable

Module 1: Study Design and Exploratory Data Analysis

- **Lesson 5: Teaching Statistics for Social Justice**
  - Activity 9: Advocating for Change with Existing Data Sets
    - Pennsylvania school district fair-funding data
  - Activity 10: Student Projects and Social Justice
    - Counting Trucks, Measuring Air Pollution
    - Using Projects in the Classroom

Module 1: Study Design and Exploratory Data Analysis

- **Lesson 6: Interpreting Graphs**
  - Activity 11: Students' Graph Comprehension
    - 3 frameworks
    - Case-to-aggregate views of data
  - Activity 12: Displaying Aggregated Data
    - Detroit Western High: Combining schools, lengthening commutes
    - Students using Dotplots, Histograms, Box & Whisker

Module 1: Study Design and Exploratory Data Analysis

- **Lesson 7: Interpreting and Responding to Student Thinking**
  - Activity 13: Student Thinking on Univariate Categorical Graphs
    - Joy using CODAP to explore Vehicles data set
    - Trina using CODAP to explore Vehicles data set
  - Activity 14: Responding to Student Thinking
    - Comparing animal-weight data in 2 groups
    - Test Results boxplot: which exam did the class do the best on?

Module 1: Study Design and Exploratory Data Analysis

- **Lesson 8: Characteristics of Distributions**
  - Activity 15: Measures of Center & Spread
    - 3 views of Mean
    - MAD, SD
    - Learning Trajectory for Center and Spread
  - Activity 16: Wrapping Up Shape, Center, and Spread
    - Mean vs Median in skewed data
    - Judging Evidence of a Difference
    - Signal and Noise

Module 2: Statistical Inference

- **Lesson 1: Simulation-based Hypothesis Test: Comparing 2 Groups (means)**
  - Activity 1: Is There a Difference?
  - Activity 2: Do the Shuffle
- **Lesson 2: Random Assignment Simulations with Technology**
  - Activity 3: Technology for Simulations: Lock5stat.com/StatKey
  - Activity 4: Effect of Class Size on Students: the TN STAR study

Pedagogical Aspects

- Video Simulation of Teaching Practice: videorecord yourself as if you were responding to a student's thinking

What Makes for a Good Teacher Response?

1. Asking questions is often better than making statements.
2. It can be helpful to give an example of the student's reasoning but taken to extremes (either numerical extremes or context extremes where the answer should be clear).
3. Simply telling the student the right way to think is often not as helpful as one might hope—they need to see why their reasoning isn't correct.
4. High quality responses to student work:
   a) Move students toward the student learning objective;
   b) Draw on and are consistent with the student thinking presented and research on students' mathematical development; and
   c) Leave space for student's future thinking (not just teacher's thinking).

Video Simulation of Teaching Practice

Inference Module – Simulation of Practice Video Assignment

Responding to Student Statements about Confidence Intervals

Students in your eleventh-grade math class are learning about simulation-based confidence intervals to meet the following learning standard:

CCSS.Math.Content.HSS.IC.B.4

> Use data from a sample survey to estimate the population mean or proportion; develop a margin of error through the use of simulation models for random sampling.

You taught them how to make and interpret a simulation-based confidence interval. Inspired by the activities in Lessons 3 and 4 about how much sleep college students get on average, you decided to have your students practice the process for making a simulation-based confidence interval for a mean in the context of amount of sleep high school students get on average. They have selected a simple random sample of 50 students from the school and collected data on how much sleep they got last night. You gather this data from them at the start of the class period then work together as a class using the procedure for a simulation-based confidence interval for a mean to produce the following 95% confidence interval: (6.39, 6.91)

Video Simulation of Teaching Practice

Responding to Student Statements about Confidence Intervals (cont.)

You pose the following question to the class for discussion:

"What does this confidence interval mean?"

Julio responds, "It means, like, 95% of the kids at our school get between 6.39 and 6.91 hours of sleep a night."

Record a video of yourself where you include:

1. Your interpretation of Julio's response, including what is worthwhile or reasonable in his thinking.
2. Your response to Julio (exactly what you would say and/or write) as part of this class discussion. Your response should help the student revise any incorrect statements and move the student forward in their understanding of how to interpret confidence intervals.

Video Simulation of Teaching Practice

Responding to Student Statements about Confidence Intervals (cont.)

Following the class discussion about how to interpret this confidence interval, Luciana speaks up:

"But we only asked like 50 kids and we have a really big school. So I think we need to ask more kids because that will get us a bigger confidence interval."

Record a video of yourself where you include:

1. Your interpretation of Luciana's statement, including what is worthwhile or reasonable in her thinking.
2. Your response to Luciana (exactly what you would say and/or write) as part of this class discussion. Your response should help the student revise any incorrect statements and move the student forward in their understanding of confidence intervals.
3. A detailed description of what you would do with the class in the time remaining in that class session (approximately 15 minutes). This would fill the time from the end of your response to Luciana to when the class session ends.

Abstract

Secondary students are expected to learn how to informally evaluate evidence of a difference between two groups. At the 7th grade level, they look at the "difference between the centers by expressing it as a multiple of a measure of variability" and "Use measures of center and measures of variability for numerical data from random samples to draw informal comparative inferences about two populations". At the high school level, they "Use data from a randomized experiment to compare two treatments; use simulations to decide if differences between parameters are significant" (Common Core State Standards for Mathematics). In this presentation, we will share new teacher education curriculum materials from the MODULE(S^2) project which prepare teachers to teach students how to evaluate evidence of a difference between two groups. The materials include both content and pedagogical aspects of teaching the topic.

Session description (just for reference while writing)

- We will have participants work through some of the activities that are in our curriculum. For example, we will do an open-ended activity where participants put in order 6 data sets according to how much evidence of a difference between groups they show, then develop a formula to quantify the evidence.
- We will also do an activity about using simulation to decide if differences between two groups are significant, and discuss common developing student conceptions for the topic.

A 7th-grade-level activity

- Based on Trumpower (2013) and Trumpower (2015)
- Suppose that 6 medicines (A,B,C,D,E,F) each had a clinical trial, with random assignment of research subjects to Treatment (3 subjects) or Placebo (3 other subjects).
- Afterwards, each person's blood was tested and the amount of monocytes (a type of immune system cell) was recorded, with higher amounts being better in this situation.
- We are wondering: which study shows the most evidence of an effect? Which study shows the least evidence of an effect? What's the overall order, from least evidence to most evidence?

Data Glimpse

	Placebo:		Treatment:	
Medicine	PersonID	Result	PersonID	Result
A	A1	299	A4	304
	A2	300	A5	305
	A3	301	A6	306

- http://bit.ly/PlaceboTreatment_00
- Or https://codap.concord.org/releases/latest/static/dg/en/cert/index.html#shared=43840

Untitled Document

Tables Graph Map Slider Calc Text Plugins

medicineA

Cases (6 cases)

index	PersonID	group	result
1	A1	placebo	299
2	A2	placebo	300
3	A3	placebo	301
4	A4	treatme...	304
5	A5	treatme...	305
6	A6	treatme...	306

medicineB

Cases (6 cases)

index	PersonID	group	result
1	B1	placebo	250
2	B2	placebo	300
3	B3	placebo	350
4	B4	treatme...	350
5	B5	treatme...	400
6	B6	treatme...	450

medicineC

Cases (6 cases)

index	PersonID	group	result
1	C1	placebo	250
2	C2	placebo	300
3	C3	placebo	350
4	C4	treatme...	255
5	C5	treatme...	305
6	C6	treatme...	355

Cases (6 cases)

index	PersonID	group	result
1	D1	placebo	250
2	D2	placebo	300
3	D3	placebo	350
4	D4	treatme...	300
5	D5	treatme...	350
6	D6	treatme...	400

medicineE

Cases (6 cases)

Text

These graphs are currently in letter-order (A,B,C,D,E,F) but that's not necessarily the same order as evidence-of-a-difference. Re-order them to show least evidence at the top, most evidence at the bottom.

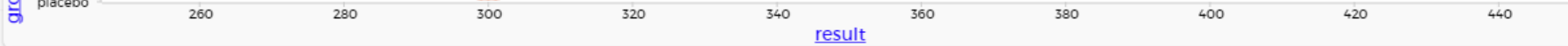
medicineA

medicineB

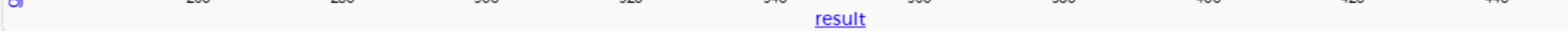
medicineC

medicineD

medicineE



medicineF



- Note that the x-axes for the dotplots are all the same, which is important—do not change them. Right now they are just in order by the medicine's letter (A,B,C,D,E,F). Using just your intuition for guidance, re-order the dotplots in CODAP so they are in evidence-of-effect order, with the least evidence along the top and the most evidence at the bottom . Discuss with your neighbor what your reasoning is. Record your thinking and your results here

7^{th}-grade activity: Discussion

- Exercise 16-5: Suppose that a student says “when we were comparing two groups to see if there was a difference, why did we divide the difference of the means by the spread? Why not just take the difference of the means and be done?” How would you respond, as their teacher?

Thoughts on the old 7th-grade activity

- Need better understanding of sampling variability of the means—the means aren't always what the data shows
- Related: need more understanding of signal-vs-noise
- Instead of all 6 at once, focus on pairs of studies: some pairs have different difference-of-means, others have same difference-of-means but more or less spread.
- More data in the graphs: 30 points each? 100?
- Aspect ratio of the graphs, not so long-and-skinny
- More experience using a measure of spread, or just tell them what spread value to use for each—having them decide what to use, and calculating it for each little data set, took too much time. They weren't ready to just eyeball it.
- Knowing z-scores might be helpful
- Motivation to come up with a formula: if different people in the class have different opinions, we want an objective way to decide.

7th-grade standard and Progressions document

Another 7th grade standard

Draw informal comparative inferences about two populations.

- CCSS.MATH.CONTENT.7.SP.B.4
  Use measures of center and measures of variability for numerical data from random samples to draw informal comparative inferences about two populations. *For example, decide whether the words in a chapter of a seventh-grade science book are generally longer than the words in a chapter of a fourth-grade science book*.

Data Glimpse: A vs B

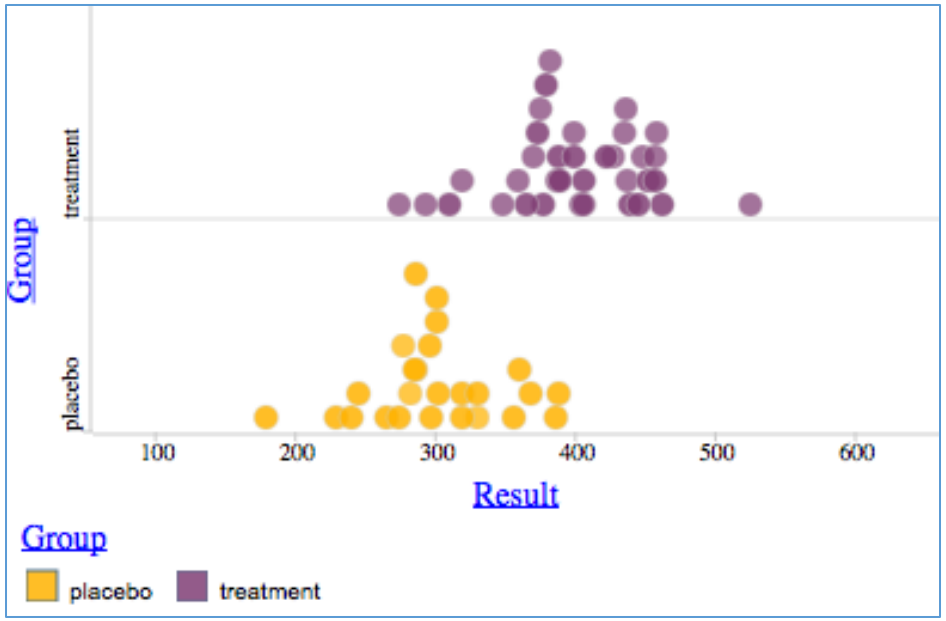
Visual overlap only:
https://codap.concord.org/releases/latest/static/dg/en/cert/index.html#shared=103797

With means:
https://codap.concord.org/releases/latest/static/dg/en/cert/index.html#shared=103802

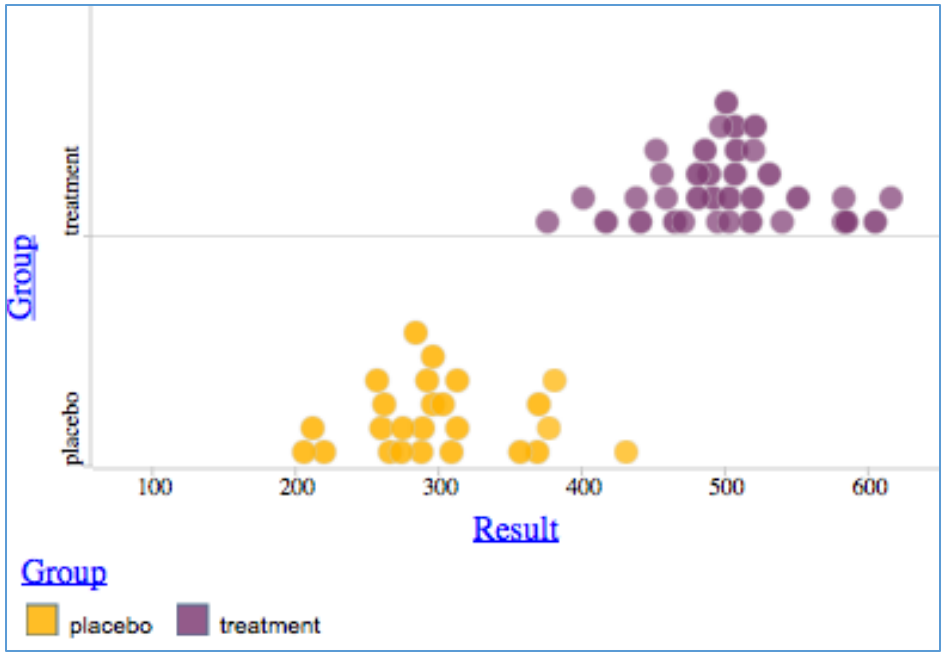
With MADs:
https://codap.concord.org/releases/latest/static/dg/en/cert/index.html#shared=103803

With means and MADs:
https://codap.concord.org/releases/latest/static/dg/en/cert/index.html#shared=103799

Medicine A



Medicine B



Data Glimpse: A vs C

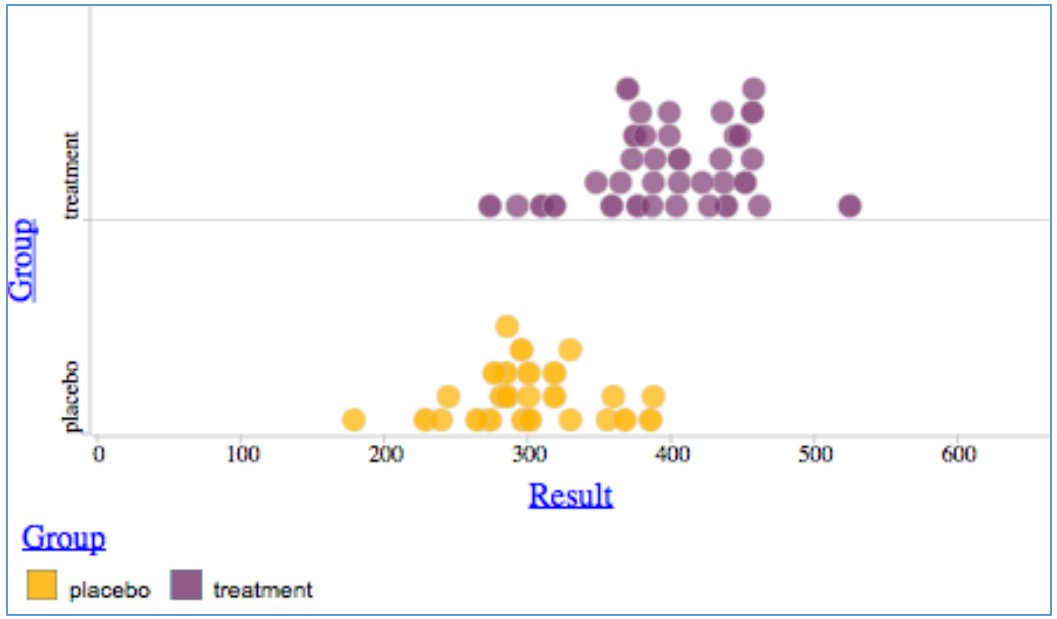
Visual overlap only:
https://codap.concord.org/releases/latest/static/dg/en/cert/index.html#shared=103797

With means:
https://codap.concord.org/releases/latest/static/dg/en/cert/index.html#shared=103802

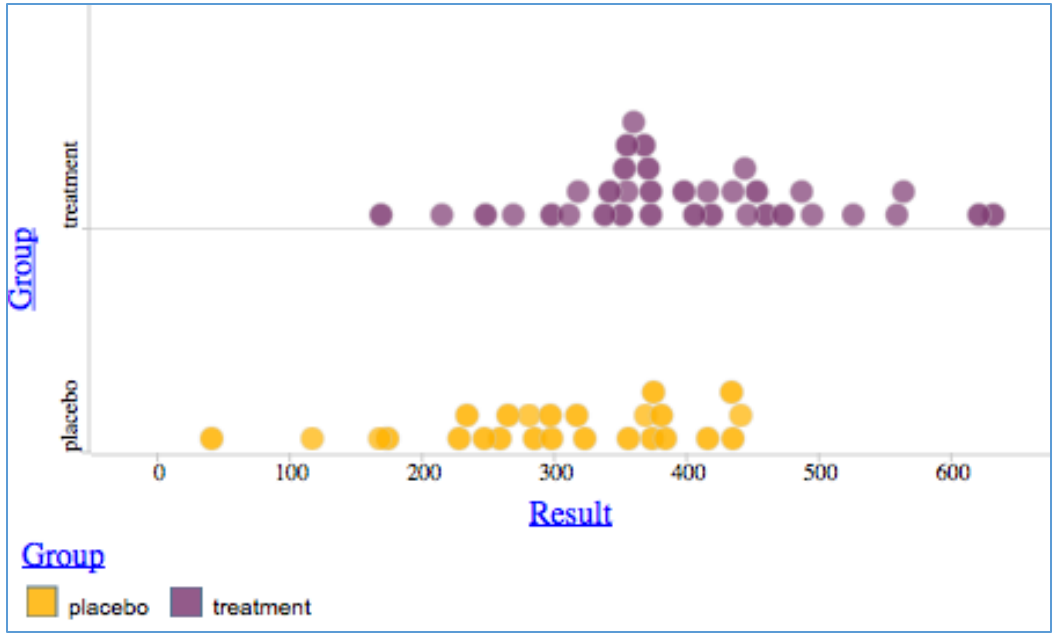
With MADs:
https://codap.concord.org/releases/latest/static/dg/en/cert/index.html#shared=103803

With means and MADs:
https://codap.concord.org/releases/latest/static/dg/en/cert/index.html#shared=103799

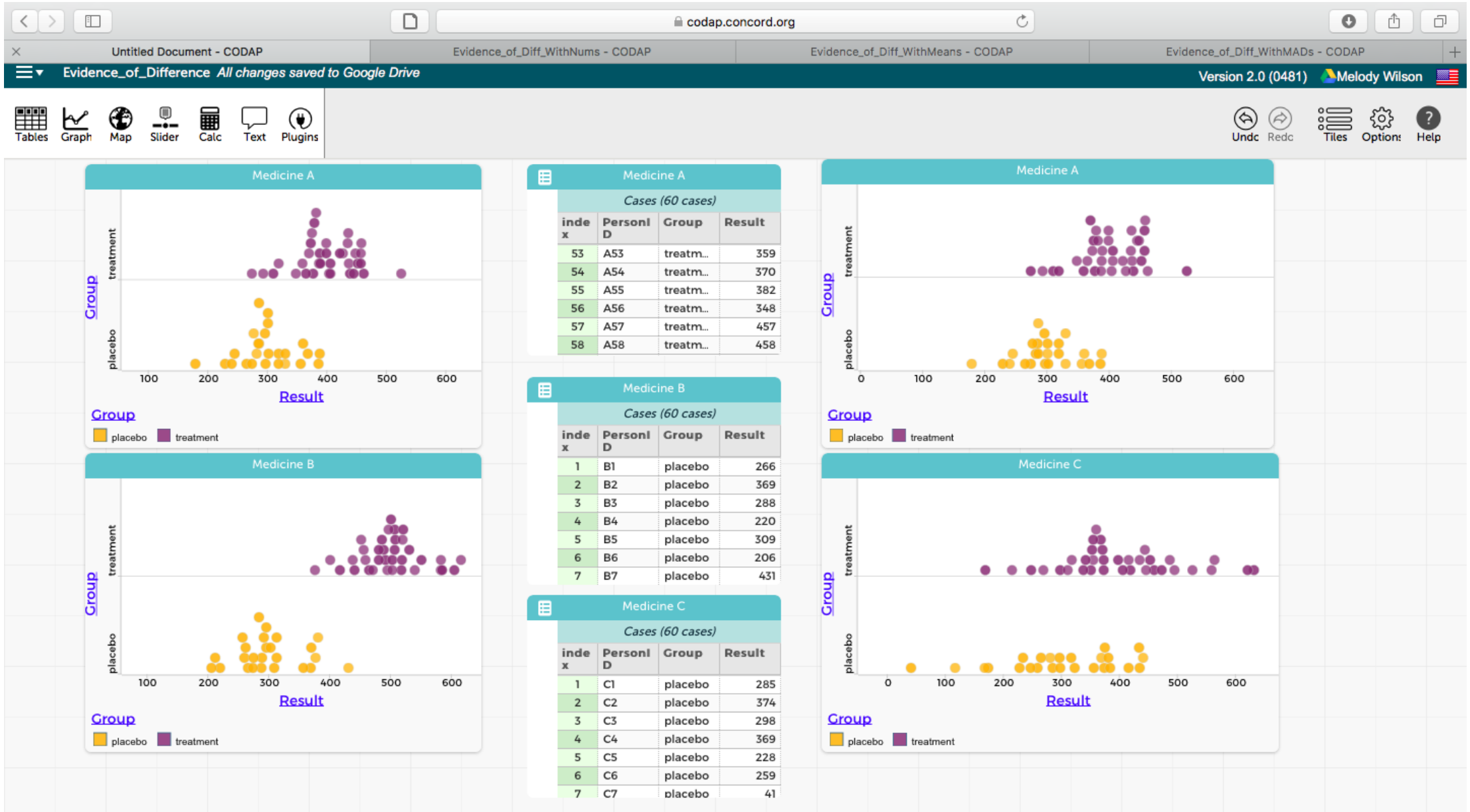
Medicine A



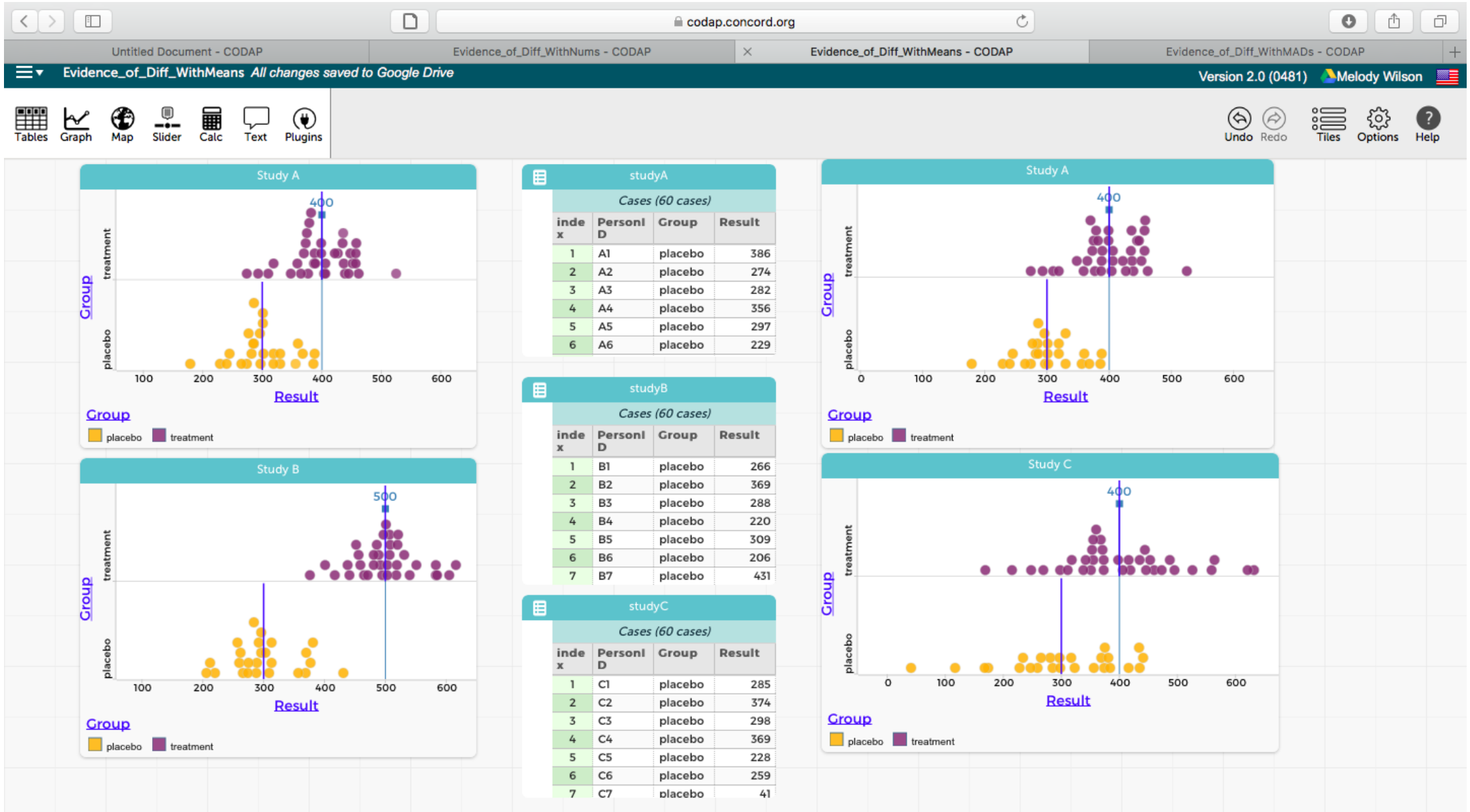
Medicine C



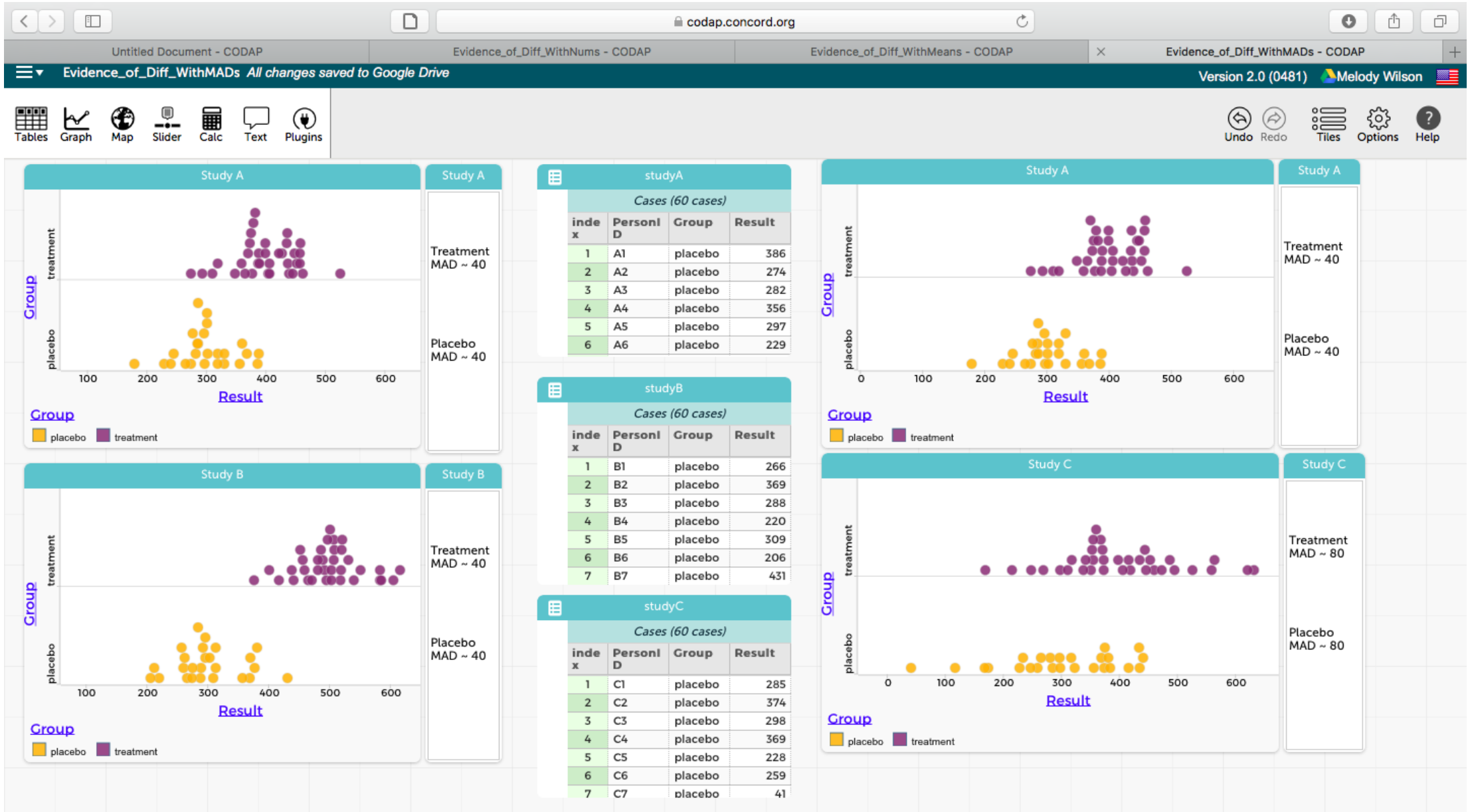
1. Which study shows the most evidence of an effect: A or B?
2. Which study shows the most evidence of an effect: A or C?



1. Which study shows the most evidence of an effect: A or B?
2. Which study shows the most evidence of an effect: A or C?



1. Which study shows the most evidence of an effect: A or B?
2. Which study shows the most evidence of an effect: A or C?



1. Which study shows the most evidence of an effect: A or B?
2. Which study shows the most evidence of an effect: A or C?