Did same institution use the same book? **Yes**

Retention data folder, check graphs and look at sample sizes

* First graph sample size correct, can count by count(data$textbook); no clue where she got the sample sizes for the second graph when I’m pretty sure they should just be the same. Can hard code it by replacing the numbers. In fact, I think all the sample sizes should be the same if every student took pre, post, ret(4), and ret(16). If what I’m thinking is true, can just copy paste labels for the first 2 graphs into the 4-month retention plot.

Multilevel Models (using data = data\_box)

* section (176.1), instructor (140.1), institution (122.9)
* institution/instructor (121.2), institution/section (121.6), instructor/section (141.4)
* institution/instructor/section (122.4)

All 3 years by instructor w/ 4 textbook classifications for boxplot (Can’t figure out, data1617 not working)

Make new file with good workflow, copy things that work in a file

**yr1 more variable cleaning:**

Setwd to my own desktop

Lines 97, 98: ‘data’ -> ‘data1415’

At end, add write.csv(data1415, file = "FinalV2 1415.csv")

**yr2 more variable cleaning:**

line 1: comment to “year 2”

line 149: replacement row error (do we even need this line?)

changed to time.meet.new

Lines 169, 170: ‘data’ -> ‘data1516’

At end, add write.csv(data1516, file = "FinalV2 1516.csv")

**yr3 more variable cleaning:**

Lines 126, 127: ‘data’ -> ‘data1617’

At end, add write.csv(data1617, file = "FinalV2 1617.csv")

**Full Year Analysis Mar27(2):**

Line 200: “ls()ylab” -> “ylab”

Line 359: install.packages("dplyr")

Line 647: replacement row error for final plot (had to run only textbook.classification2 lines; commented out others so can run straight)

Line 656: need ggthemes package

Do “install.packages(“ggthemes”)”

Line 663

Line 84 only run in that section

Line 350 only run

Likes Line 633 graphic

To do: Line 640 alter graphic, re-label, sort, etc.

Add mergedfsort$Category <- factor(mergedfsort$Category , levels = c("DC","DS", "Sim", "ST", "CI")) to after question numbering (add to line 652)

Add data\_box$InstructorName\_Section\_Semester <- gsub("Section.\*$","",data\_box$InstructorName\_Section\_Semester)

to after data\_box$textbook initialization to only get instructor names and group by them; can comment out to get by name, section, and semester (Line 789 ish)

Added labels to Final Plot columns

Line 645: change to question2

Line 668: factor(question), change scale\_x\_continuous to scale\_x\_discrete

Changed line order for ISI vs SBI vs NonSBI for all:

See lines 443 swapping numbers with 448

**Tables\_visual:**

Line 368: “theme\_calc()” -> “theme\_bw()”

**Attitudes\_update.R:**

Line 301: need ggthemes package

**Chance Things**

Check what cleaning goes to what year

Final Cleaning – takes in “Further Cleaning Result .csv” for all years and returns “Final .csv” for all years

yr1/2/3 more variable cleaning – takes in “Final 14.15.csv” and returns ???

Full year takes in “more variable cleaning”

**Multilevel Models:** (Read the chapter first, 5.7)

Student Levels and Instructor Levels (how long teach, class in morning or afternoon)

Assume students in different classes are different

Students nested in classes, classes nested in instructor, instructor nested in institution, but unsure if this many nests are necessary

Run models with different nesting and check which AIC is lowest

Some RML thingies

We examine three years of cross-institutional data (pre/post tests) from a wide variety of introductory statistics courses, exploring student gains in learning, including courses that revolve around simulation-based inference. This includes comparison of student attitudes towards statistics and student retention 4-months and 16-months post course. Hierarchical models explore student-level characteristics (e.g., first statistics course, first generation, prior mathematical performance) and instructor-level characteristics (e.g., type of institution including high school and community college, years teaching, familiarity with GAISE guidelines). We find that simulation-based inference courses consistently show larger gains than non-SBI curricula, regardless of institution, year, student pre-test score and a host of other institutional and student characteristics. Further research is needed to better understand what aspects of simulation-based inference curricula are directly leading to the observed patterns in student learning.

**Poster notes:**

<https://ww2.amstat.org/meetings/jsm/2019/postertips.cfm>

Full Year Analysis Mar27.R

Separate via high school, ISCAM, etc.

ISI, SBI, one is good book other is bad, one for 5 years still good, other 20 years still bad

Recreate (640) FINAL PLOT, get it to work/reformat

Boxplot for ISCAM

Visual on retention

(951) model18 taken out insignificant vars.