Course on Communicating Data and Statistics

Andrew Gelman

28 Oct 2021

1. Goals for students

- Communication is central to your job as a quantitative researcher. Our goal in this course is for you to improve at all aspects of statistical communication, including writing, public speaking, teaching, informal conversation and collaboration, programming, and graphics.
- Always keep in mind your **goals** and your **audience**.
- Never forget, as one of our blog commenters reminds us: your closest collaborator is you six months ago . . . and she doesn't reply to email!

2. Student responsibilities

- **Every day**, you will write an entry in your **statistics diary**; see here: `https://statmodeling.stat.columbia.edu/2015/01/07/2015-statistics-diary/`. Just set up a plain text (not Word) file and add to it each day. The diary entries can be anything. They can be short slice-of-life observations ("Today Jakey asked me what he would be the best at in ten years: piano, soccer, judo, chess, or fencing. This made me wonder: how predictable are such things? How would we gather data to measure this?"), quick research notes ("Someone emailed me a graph of birthday data from Brazil! I sent it to Aki along with a link, maybe he or a student can fit our model. Apparently Brazil has different holidays than we do in the U.S."), or things you're working on, difficult statistics problems that you might be stuck on, or have an insight about. You can write as little or as much as you want each day. The only requirement is that you write something new in it, every day. You're *not* allowed to go back a week later and fill in 7 entries at once. That would be cheating. One entry each day. Just type it in to the file.

  Submit your diary entries each week by pasting them into the corresponding assigment in Courseworks.
- **Before every class**, you will have a **homework assignment**.
- **Before every class**, you will have **readings**. These include research articles, software manuals, blog posts, various other online materials, and chapters from the five assigned books:
  - *Data Visualization: A Practical Introduction*, by Kieran Healy.
  - *Teaching Statistics: A Bag of Tricks*, by Andrew Gelman and Deborah Nolan.
  - *Judgment Under Uncertainty: Heuristics and Biases*, edited by Daniel Kahneman, Paul Slovic, and Amos Tversky.
  - *The Elements of Style*, by William Strunk and E. B. White.
  - *How to Talk So Kids Will Listen and Listen So Kids Will Talk*, by Adele Faber and Elaine Mazlish.

- *Before every class*, you will have a *jitt* (just-in-time teaching assignment); see here for background: `https://statmodeling.stat.columbia.edu/2013/09/27/setting-up-jitts-online/`. Each of your jitts will be a set of quick online items, separate from the main homework assignments, that are a mix of questions on the required reading, short exercises to get you ready for the upcoming class discussion, and feedback.

- **Each class** will involve your **active participation** in class discussion, and each class will have a **student teacher** and a **student expert**. Also, **bring your laptop computer** to class as we will sometimes be doing activities together in R.

- **At the end of the semester**, you will have to complete a **final project** in teams of 2 or 3. The topic can have a methodological or a substantive focus. Topics from previous years included Shinystan, Economic Inequality, Math at the Museum, Plotting Data for Dummies, Teaching Anova, and Imputing the National Election Study.

3. Structure of course

3.1. Topics for the 13 weeks of the course

1. Introducing yourself and telling a story
2. Principles of statistical graphics
3. Teaching
4. Making effective graphs
5. Communicating variation and uncertainty
6. Displaying fitted models
7. Giving a presentation
8. Dynamic graphics
9. Writing
10. Collaboration and the scientific community
11. Data processing and programming
12. Student projects
13. Student projects

Each week will have two classes on a common theme.

3.2. Each class

Each week we have 2 classes of 75 minutes each. Here's the plan for a typical class (although we won't follow the prototype exactly):

1. (before class begins) Share statistics stories

2. (10 min) As a class, discuss the week's readings and jitts
3. (10 min) In pairs, discuss the just-completed homeworks
4. (20 min) Conversation with special guest star, sliding into general class discussion
5. (25 min) Class discussion of several of the homework responses
6. (10 min) Discuss readings and homework for next class

Throughout the class, we'll use the blackboard to remind us of important points and keep the discussion on track.

Special guest star: Every class will feature a guest (in person or via Google hangout). The guest will not be giving a prepared talk; rather, students will have a conversation with the guest, based on their questions.

Student teacher: For each class, one of you will lead the class discussion of the week's readings and jitts.

Student expert: For each class, one of you who have some expertise in the week's topic will play the role of "intervener" and can interrupt the class at any time to bring up important points or steer us in a useful direction.

Class 1a: Introducing yourself

Readings before class: None

Homework due at beginning of class: None

In class:

1. (10 min) Slides: Display and discuss the birthdays graph (pages 26–30 of `http://www.stat.columbia.edu/~gelman/presentations/call_me_bayesian.pdf`)
2. (5 min) Blackboard: Write a statistics challenge (how to attack some real-world problem, probably something from my current research)
3. (15 min) In pairs, students discuss the challenge:
    - (2 min) Pair discussion
    - (3 min) Students think on their own
    - (5 min) Students explain their ideas to each other
    - (5 min) Further pair discussion
4. (5 min) Why I hate the Napoleon-in-Russia graph
5. (10 min) Blackboard: I introduce the course and answer questions
6. (5 min) Browser: Demonstrate jitts

7. Return to the statistics challenge
    - (5 min) Students cluster into groups of 4
    - (15 min) Class discussion
8. (5 min) Slides: Discuss readings and homework for next class
9. Give students handouts as they leave

Special guest star: You (the students in the class)!

Class 1b: Telling a story

Readings before class:

1. Andrew Gelman, "Five books on statistics": `http://fivebooks.com/interviews/andrew-gelman-on-statistics`
2. Kaiser Fung, "Three hours in the life of a (glorified) 'data scientist'": `http://junkcharts.typepad.com/numbersruleyourworld/2011/11/three-hours-in-the-life-of-a-glorified-data-scientist.html`
3. Andrew Gelman, "Stan goes to the World Cup": `https://statmodeling.stat.columbia.edu/2014/07/13/stan-analyzes-world-cup-data/`
4. Andrew Gelman and Thomas Basbøll, "When do stories work? Evidence and illustration in the social sciences": `http://www.stat.columbia.edu/~gelman/research/published/storytelling.pdf`
5. Andrew Gelman, "Gladwell and Chabris, David and Goliath, and science writing as stone soup": `https://statmodeling.stat.columbia.edu/2013/10/11/gladwell-vs-chabris-david-vs-goliath/`
6. Kaiser Fung, "Figuring out what data supports the argument, and what is just window-dressing": `http://junkcharts.typepad.com/numbersruleyourworld/2015/01/figuring-out-what-data-supports-the-argument-and-what-is-just-window-dressing.html`
7. Andrew Gelman, "The persistence of the 'schools are failing' story line": `https://statmodeling.stat.columbia.edu/2014/12/05/persistence-schools-failing-story-line`
8. Thomas Basbøll, "Methodology and experience": `https://blog.cbs.dk/inframethodology/?p=3903`

Homework due at beginning of class: In the next class you will work in pairs to write up a research or work experience, following the model of Kaiser Fung's "Three hours in the life of a (glorified) 'data scientist'" and my "Stan goes to the World Cup." It is not necessary that you focus on software and data manipulation (as in Fung's article) or on statistical modeling and data analysis (as in my article). What is important is that, whatever aspect of your work you *do* describe, that you describe it as directly as possible.

In preparation for this activity, prepare half a page of notes about a research or work experience of yours. The notes do not have to be complete sentences; you just need to include enough information so that you can reconstruct the full story during the class period.

Prepare your homework assignment as a plain text or html document, print it out, and bring to class.

In class:

1. (5 min) Discuss jitts
2. (5 min) In pairs (different pairings from last time), students tell each other the stories from their notes
3. (10 min) Blackboard and projector: I demonstrate how to write a story from notes:
    - (2 min) Students pick a topic
    - (2 min) I take quick notes
    - (6 min) I write the story (type on to the computer displayed onto the screen)
4. (20 min) With guest star, discuss storytelling
5. (25 min) Each pair of students writes a story together
6. (10 min) Discuss readings and homework for next class

Special guest star: Thomas Basbøll, Writing Consultant, Copenhagen Business School

Class 2a: Graphics in statistical research and practice

Readings before class:

1. Andrew Gelman and Antony Unwin, "Infovis and statistical graphics: Different goals, different looks (with discussion)": `http://www.stat.columbia.edu/~gelman/research/published/vis14.pdf` and `http://www.stat.columbia.edu/~gelman/research/published/visreply3.pdf`
2. Kieran Healy, *Data Visualization*, prefance and chapter 2: "Get started" `https://courseworks.columbia.edu/access/content/group/QMSSG4065_001_2015_1/RStudio/G4065.html`
3. Andrew Gelman, Jennifer Hill, and Aki Vehtari, *Regression and Other Stories*, Appendix A: `https://avehtari.github.io/ROS-Examples/R_visualizaton.html`
4. Instructions on R Markdown: `http://rmarkdown.rstudio.com`

Homework due at beginning of class: Your assignment is to write two R functions:

1. An R function that creates a two-way array of independent random simulations from a normal distribution. The arguments to the function should be the number of rows, number of columns, and the standard deviation of the distribution.

2. An R function that takes a two-way array and returns two vectors corresponding to the minimum and maximum of each row. (That is, if the input to the function is a 5×10 array, the function should return two vectors of length 5.)

Document the function and demonstrate that it works in some examples and put all this in an R Markdown file. Save the output as html or pdf and upload it to Courseworks as an attachment. Also print it out and bring it to class.

Just to illustrate, here is some R markdown code. Open RStudio, go to File / New File / R Markdown to open a new R Markdown file, then copy-and-paste the code below into the text window. Then click on the Knit HTML button near the top of that text window and it should create an html document.

````
Here is a simple example of an R function:

```{r}
plus <- function(a, b, label="sum"){
  sum <- a+b
  names(sum) <- label
  return(sum)
}
```

Now the demonstration that it works:

```{r}
plus(2,3)
plus(2,3.4,"hello")
```
````

In class:

1. (5 min) Discuss jitts
2. (45 min) My presentation, "Choices in statistical graphics: my stories": `http://www.stat.columbia.edu/~gelman/presentations/vistalk_meetup_new_handout.pdf`
3. (15 min) Class discussion of principles of statistical graphics
4. (10 min) Discuss readings and homework for next class

Special guest star: me!

Class 2b: Principles of statistical graphics

Readings before class:

1. Kieran Healy, *Data Visualization*, chapters 1 and 3: "Look at data" and "Make a plot"
2. L. J. Zigerell, "R graph: plot"; `http://www.ljzigerell.com/?p=1891`
3. L. J. Zigerell, "R graph: confidence intervals": `http://www.ljzigerell.com/?p=1916`

4. Howard Wainer, "Improving data displays: Ours and the media's": http://www.stat.columbia.edu/~gelman/communication/Wainer2009.pdf

Homework due at beginning of class: Graph some data of interest to you. Your display should be a grid of at least 4 plots on the page. Accompanying your plots, write a brief description of the graph's (hypothetical) audience, your goals in making the graphs, and the particular comparisons that you want to display.

In class:

1. (5 min) Discuss jitts
2. (5 min) Discuss readings
3. (10 min) Discuss students' homeworks
4. (20 min) With guest star, discuss principles of statistical graphics
5. (25 min) Discuss students' homeworks
6. (10 min) Discuss readings and homework for next class

Special guest star: Kieran Healy, Department of Sociology, Duke University

Class 3a: Classroom teaching

Readings before class:

1. Andrew Gelman, "Some quick disorganized tips on classroom teaching": https://statmodeling.stat.columbia.edu/2014/08/18/quick-disorganzed-tips-classroom-teaching/
2. Andrew Gelman, "How to design a mini-lecture on statistical inference?": https://statmodeling.stat.columbia.edu/2009/08/06/how_to_design_a/
3. Rachel Schutt, "Taxonomy of confusion": https://statmodeling.stat.columbia.edu/2008/09/11/taxonomy_of_con/
4. Andrew Gelman, "Building a better teacher": https://statmodeling.stat.columbia.edu/2010/03/05/building_a_bett/
5. Andrew Gelman, "Teaching Bayesian applied statistics to graduate students in political science, sociology, public health, education, economics, . . . ": http://www.stat.columbia.edu/~gelman/research/published/teachingbayes.pdf
6. Andrew Gelman, "The candy weighing demonstration, or, the unwisdom of crowds": https://statmodeling.stat.columbia.edu/2008/05/08/doing_the_candy/
7. Jessica Watkins and Eric Mazur, "Using jitt with peer instruction": http://www.stat.columbia.edu/~gelman/communication/WatkinsMazur2009.pdf
8. Catherine Crouch and Eric Mazur, "Peer instruction: Ten years of experience and results": http://www.stat.columbia.edu/~gelman/communication/CrouchMazur2001.pdf

9. Andrew Gelman and Vincent Dorie, "Setting up Jitts online": `https://statmodeling.stat.columbia.edu/2013/09/27/setting-up-jitts-online/`

10. Andrew Gelman and Deborah Nolan, *Teaching Statistics: A Bag of Tricks*, chapters 1–2: "Introduction" and "First week of class"

11. Andrew Gelman and Deborah Nolan, *Teaching Statistics: A Bag of Tricks*: choose one of chapters 3, 4, 5, or 6

Homework due at beginning of class: In pairs, prepare a 5-minute lesson (*not* a lecture). Be clear on the who is your intended audience and what would be the context for this lesson.

In class:

1. (5 min) Discuss jitts
2. (5 min) Discuss readings
3. (10 min) Discuss students' homeworks
4. (20 min) With guest star, discuss facilitating active learning
5. (30 min) Do 3 of students' prepared lessons. For each:
    - (5 min) The pair of students does the lesson for the class
    - (5 min) Class discussion and critique
6. (5 min) Discuss readings and homework for next class

Special guest star: Beth Chance, Professor of Statistics, California Polytechnic State University

Class 3b: Preparing and evaluating a class

Readings before class:

1. Andrew Gelman and Eric Loken, "Statisticians: When we teach, we don't practice what we preach": `http://www.stat.columbia.edu/~gelman/research/published/ChanceEthics2.pdf`

2. Andrew Gelman and Deborah Nolan, *Teaching Statistics: A Bag of Tricks*, chapters 12–13: "How to do it" and "Structuring an introductory statistics course"

3. Andrew Gelman and Deborah Nolan, *Teaching Statistics: A Bag of Tricks*: choose one of chapters 8, 9, 10, or 17

4. Robert Delmas, Joan Garfield, Ann Ooms, and Beth Chance, "Assessing students' conceptual understanding after a first course in statistics": `http://www.stat.columbia.edu/~gelman/communication/DelmasGarfieldOomsChance2007.pdf`

5. Randall Pruim, Nicholas Horton, and Daniel Kaplan, "Getting started with RStudio," chapter 2 of *Start Teaching With R*: `http://www.stat.columbia.edu/~gelman/communication/PruimHortonKaplan2014.pdf`

Homework due at beginning of class: Prepare to present your lesson to the class.

In class:

1. (5 min) Discuss jitts
2. (5 min) Discuss readings
3. (10 min) Discuss students' homeworks
4. (20 min) With guest star, discuss issues in evaluation of teaching effectiveness
5. (30 min) Do 3 of the prepared lessons. For each:
    - (5 min) The pair of students does the lesson for the class
    - (5 min) Class discussion and critique
6. (5 min) Discuss readings and homework for next class

Special guest star: Eric Loken, Department of Educational Psychology. University of Connecticut.

Class 4a: Graphing data: what to do

Readings before class:

1. Howard Wainer, "Graphical visions from William Playfair to John Tukey": `http://www.stat.columbia.edu/~gelman/communication/Wainer1990.pdf`
2. Howard Wainer, "How to display data badly": `http://www.stat.columbia.edu/~gelman/communication/Wainer1984.pdf`
3. Andrew Gelman, "Why tables are really much better than graphs": `http://www.stat.columbia.edu/~gelman/research/published/tables5.pdf`
4. Andrew Gelman, "A statistical graphics course and statistical graphics advice": `https://statmodeling.stat.columbia.edu/2014/03/25/statistical-graphics-course-statistical-graphics-advice/`
5. Andrew Gelman, "Tables as graphs: the Ramanujan principle": `http://www.stat.columbia.edu/~gelman/research/published/ramanujan.pdf`
6. Andrew Gelman and Jared Niemi, "Statistical graphics: making information clear — and beautiful": `http://www.stat.columbia.edu/~gelman/research/published/Niemi_GraphicsInformation.pdf`
7. Andrew Gelman, "Unalphabetize!": `https://statmodeling.stat.columbia.edu/2008/05/08/unalphabetize/`
8. Kaiser Fung, "Six quotes": `https://statmodeling.stat.columbia.edu/2014/09/14/six-quotes-kaiser-fung/`

9. Kaiser Fung, "Junk Charts trifecta checkup: The definitive guide": http://junkcharts.typepad.com/junk_charts/junk-charts-trifecta-checkup-the-definitive-guide.html

10. Andrew Gelman, "Statistical graphics for research and presentation": http://www.stat.columbia.edu/~gelman/communication/appendixB.pdf

11. Amanda Cox, "How birth year influences political views": http://www.nytimes.com/interactive/2014/07/08/upshot/how-the-year-you-were-born-influences-your-politics.html

12. Jessica Hullman, Matthew Kay, Yea-Seul Kim, and Samana Shrestha, "Imagining replications: Graphical prediction & discrete visualizations improve recall & estimation of effect uncertainty": http://www.stat.columbia.edu/~gelman/communication/HullmanKayKimShrestha2017.pdf

Homework due at beginning of class: Find a graph or set of graphs that you like, made by someone else, on a topic that interests you. Write two or three sentences explaining what the graphs say, why the topic is important, and why it is useful to display this information graphically. You should do this in the form of a one-page report including the graph and your discussion.

Then add a second page in which you criticize the graphs and offer suggestions for improvement. You should do this in the form of a two-page report where the first page is the graph and the second page has a list of criticisms and suggestions, like this:

1. (a) Criticism: ...
   (b) Suggestion for improvement: ...
2. (a) Criticism: ...
   (b) Suggestion for improvement: ...
3. ...

In class:

1. (5 min) Discuss jitts
2. (5 min) Discuss readings
3. (10 min) In pairs, discuss the graphs of others that students brought to class
4. (20 min) With guest star, discuss telling stories using graphics
5. (10 min) With entire class, discuss the graphs made by students and the graphs of others
6. (15 min) With entire class, discuss general principles of making and interpreting graphs
7. (10 min) Discuss readings and homework for next class

Special guest star: Amanda Cox, Graphics Editor, New York Times

Class 4b: Graphing data: what works

Readings before class:

1. Andrew Gelman, "Stephen Kosslyn's principles of graphics and one more: There's no need to cram everything into a single plot": https://statmodeling.stat.columbia.edu/2012/12/06/one-more-graphics-principle-for-stephen-kosslyn-theres-no-need-to-cram-everything-into-a-single-plot/

2. Howard Wainer, Ronald Hambleton, and Kevin Meara, "Alternative displays for communicating NAEP results: A redesign and validity study": http://www.stat.columbia.edu/~gelman/communication/WainerHambletonMeara1999.pdf

3. Kaiser Fung, most recent posts from Junk Charts blog, http://junkcharts.typepad.com

4. Kieran Healy, *Data Visualization*, chapters 4 and 8: "Show the right numbers" and "Refine plots"

5. Scott Bateman, Carl Gutwin, David McDine, Regan Mandryk, Aaron Genest, and Christopher Brooks, "Useful junk? The effects of visual embellishment on comprehension and memorability of charts": http://www.stat.columbia.edu/~gelman/communication/Bateman2010.pdf

6. Andrew Gelman, "Is chartjunk really 'more useful' than plain graphs? I don't think so.": https://statmodeling.stat.columbia.edu/2010/05/17/is_chartjunk_re/

7. Jeffrey Heer and George Robertson, "Animated transitions in statistical data graphics": http://www.stat.columbia.edu/~gelman/communication/HeerRobertson2007.pdf

8. Jessica Hullman, Eytan Adar, and Priti Shah, "Benefitting InfoVis with visual difficulties": http://dl.acm.org/citation.cfm?id=2068622

Homework due at beginning of class: We will give you a data visualization challenge: some data on a particular topic that you will have to graph. You should have at least 4 plots on the page, as well as a caption of two to five sentences. Think about the many facets of information contained in the data, and explore different types of plots to illustrate them. It should all be in Markdown, thus including all source code. Also add a paragraph explaining everything. Upload the output (html or pdf) into the homework folder.

In making the graphs, push yourself a little. If you are an R beginner, make a graph that goes beyond what you know how to do. If you're an R expert, try to do something special. We will discuss each others' graphs and consider the virtues and drawbacks and tradeoffs of different choices of what to graph and how to display.

The data visualization challenge is to plot trends in deaths by drug overdose using the CDC Wonder database at https://wonder.cdc.gov/. You could get some inspiration from the following post: https://statmodeling.stat.columbia.edu/2015/11/18/first-second-and-third-order-bias-corrections-also-my-ugly-r-code-for-the-mortality-rate-graphs/ and the following article: http://www.stat.columbia.edu/~gelman/research/unpublished/age_adj_letter_4.pdf but you can do whatever you want here, there's no need to imitate the particular graphs that we made.

You can do this project individually or in pairs. If in pairs, each of you should submit the homework assignment.

In class:

1. (5 min) Discuss jitts
2. (5 min) Discuss readings
3. (10 min) Discuss students' homeworks
4. (20 min) With guest star, discuss evaluation of graphics ideas
5. (25 min) With entire class, discuss several students' homework graphs
6. (10 min) Discuss readings and homework for next class

Special guest star: Kaiser Fung, business consultant, formerly at SiriusXM, Vimeo, and American Express

Class 5a: Communicating variation and uncertainty

Readings before class:

1. Howard Wainer, "Depicting error": `http://www.stat.columbia.edu/~gelman/communication/Wainer1996.pdf`
2. Andrew Gelman, "I (almost and inadvertently) followed Dan Kahan's principles in my class today, and that was a good thing": `https://statmodeling.stat.columbia.edu/2014/11/25/almost-inadvertently-followed-dan-kahans-principles-class-today/`
3. Andrew Gelman, "The connection between varying treatment effects and the crisis of unreplicable research: A Bayesian perspective": `http://www.stat.columbia.edu/~gelman/research/published/bayes_management.pdf`
4. Amos Tversky and Daniel Kahneman, "Judgment under uncertainty: Heuristics and biases," chapter 1 in the Kahneman, Slovic, and Tversky book
5. Marc Alpert and Howard Raiffa, "A progress report on the training of probability assessors," chapter 21 in the Kahneman, Slovic, and Tversky book
6. Gerd Gigerenzer, "How to make cognitive illusions disappear: Beyond 'heuristics and biases'": `http://www.stat.columbia.edu/~gelman/communication/Gigerenzer1991.pdf`
7. Gerd Gigerenzer, "On narrow norms and vague heuristics: A reply to Kahneman and Tversky (1996)": `http://www.stat.columbia.edu/~gelman/communication/Gigerenzer1996.pdf`
8. Andrew Gelman, "I hate to get all Gerd Gigerenzer on you here, but ...": `https://statmodeling.stat.columbia.edu/2012/05/14/i-hate-to-get-all-gerd-gigerenzer-on-you-here-but/`
9. Andrew Gelman, "Battle of the Repo Man quotes: Reid Hastie's turn": `https://statmodeling.stat.columbia.edu/2012/05/22/arguments-about-stories-reid-hastie-replies/`

Homework due at beginning of class: This is an open-ended assignment: the goal is for you to think about how to communicate uncertainty in statistical inference. For example you could do a display practice: if you have a fitted model, make some graphs to depict the uncertainty of the estimation or prediction. If there is a model or a paper or some existing graphs that interest you, write a couple paragraphs to discuss issues related to uncertainty communication. If you've got some other ideas that's fine too, as long as you explore this central theme through the exercise.

In class:

1. (5 min) Discuss jitts
2. (5 min) Discuss readings
3. (10 min) In pairs, discuss homework assignment
4. (20 min) With guest star, discuss the psychology of decision making under uncertainty
5. (25 min) Discuss homework assignment with class as a whole
6. (10 min) Discuss readings and homework for next class

Special guest star: Megan Higgs, statistical consultant, formerly at Montana State University

Class 5b: Statistical reporting

Readings before class:

1. Ellen Langer, "The illusion of control," chapter 16 in the Kahneman, Slovic, and Tversky book
2. Andrew Gelman, "The illusion of the illusion of control," `https://statmodeling.stat.columbia.edu/2015/03/10/the-illusion-of-the-illusion-of-control/`
3. Andrew Gelman, "Ellen Langer: expert on, and victim of, the illusion of control," `https://statmodeling.stat.columbia.edu/2015/03/09/ellen-langer-expert-victim-illusion-control/`
4. Justin Horton, "Figure of fun": `http://streathambrixtonchess.blogspot.com.es/2013/02/figure-of-fun.html`
5. Andrew Gelman, "Reuters 1, New York Times 0": `https://statmodeling.stat.columbia.edu/2009/12/30/reuters_1_new_y/`
6. Andrew Gelman, "'Tragedy of the science communication commons'": `https://statmodeling.stat.columbia.edu/2013/04/30/tragedy-of-the-science-communication-commons/`
7. Andrew Gelman, "Common sense and statistics": `https://statmodeling.stat.columbia.edu/2014/12/25/common-sense-statistics/`
8. Andrew Gelman, "Pollster Doug Schoen illustrates a general point: There's more than one way to present survey results": `https://statmodeling.stat.columbia.edu/2011/10/27/occupy-polls/`

Homework due at beginning of class: Prepare some ideas for your final project.

In class:

1. (5 min) Discuss jitts
2. (5 min) Discuss readings
3. (10 min) Discuss project ideas
4. (20 min) With guest star, discuss statistical reporting in the news media
5. (25 min) Discuss possibilities for student projects
    - Students talk in pairs
    - Discuss the mechanics of forming pairs
    - Instructor shares some project ideas
        - Graphing fitted regression models and data—coefplot, zoom and click
        - Math at the museum
        - The virtual database query
        - Projects related to writing and teaching?
    - Other project ideas from students
6. (10 min) Discuss readings and homework for next class

Special guest star: Blake McShane, Department of Marketing, Northwestern University

Class 6a: Understanding fitted models

Readings before class:

1. Gerd Gigerenzer, Wolfgang Gaissmaier, Elke Kurz-Milcke, Lisa Schwartz, and Steven Woloshin, "Helping doctors and patients make sense of health statistics": `http://www.stat.columbia.edu/~gelman/communication/GigerenzerGaissmaierKurzSchwartzWoloshin2008.pdf`
2. Antony Unwin, Chris Volinsky, and Sylvia Winkler, "Parallel coordinates for exploratory modelling analysis": `http://www.stat.columbia.edu/~gelman/research/published/UnwinVolinskyWinkler2003.pdf`
3. Kenny Shirley and Andrew Gelman, "Hierarchical models for estimating state and demographic trends in U.S. death penalty public opinion": `http://www.stat.columbia.edu/~gelman/research/published/A12052_Shirley.pdf`

Homework due at beginning of class: Make a graph in R displaying some data and a model fit to the data. The example should be of some inherent interest. It doesn't have to be "exciting" or "newsworthy"—it could, for example, be something you're working on in the lab or something you just happen to care about—but it shouldn't just be something out of a textbook.

In class:

1. (5 min) Discuss jitts
2. (5 min) Discuss readings
3. (10 min) In pairs, discuss the graphs made by the students
4. (20 min) With guest star, discuss large scale data applications
5. (25 min) With entire class, discuss the graphs made by the students
6. (10 min) Discuss readings and homework for next class

Special guest star: Aki Vehtari, Department of Computer Science, Aalto University

Class 6b: Displaying fitted models

Readings before class:

1. Andrew Gelman, Cristian Pasarica, and Rahul Dodhia, "Let's practice what we preach: turning tables into graphs": `http://www.stat.columbia.edu/~gelman/research/published/dodhia.pdf`
2. Yair Ghitza and Andrew Gelman, "Deep interactions with MRP": `http://www.stat.columbia.edu/~gelman/research/published/misterp.pdf`
3. Andrew Gelman, "College football, voting, and the law of large numbers": `https://statmodeling.stat.columbia.edu/2012/10/25/college-football-voting-and-the-law-of-large-numbers/`

Homework due at beginning of class: Fit some simple logistic regression models to the speed-dating data; see here: `https://statmodeling.stat.columbia.edu/2008/01/21/the_speeddating_1/`. You can do this one in pairs; a logical pairing would be one person who is good with data and one person who is good with graphics. Display the inferences from your fitted models.

In class:

1. (5 min) Discuss jitts
2. (5 min) Discuss readings
3. (10 min) Discuss students' homeworks
4. (20 min) With guest star, discuss graphics for applying research to the real world
5. (10 min) Discuss student homeworks
6. (15 min) Discuss projects
7. (5 min) Discuss readings and homework for next class
8. (5 min) My mini-presentation on my problem with the football and voting research

Special guest star: Yair Ghitza, Senior Scientist, Catalist

Class 7a: Giving a presentation

Readings before class:

1. Howard Wainer, "Improving data displays: Ours and the media's": `http://www.stat.columbia.edu/~gelman/communication/Wainer2009.pdf`
2. Andrew Gelman, "W. Bradford Paley's talk and, more generally, what should I do to encourage audience participation when I speak?": `https://statmodeling.stat.columbia.edu/2009/04/21/w_bradford_pale/`
3. Andrew Gelman, "'A gift to the audience rather than a plea for attention': Brad Paley's tips on encouraging seminar participants to ask so many damn questions you have to tell them to shut the heck up already so you can hear the rest of the damn talk": `https://statmodeling.stat.columbia.edu/2009/04/24/a_gift_to_the_a/`

Homework due at beginning of class: Prepare a 5-minute research presentation. The topic could be technical or non-technical but it should be research (ideally, your research), not standard classroom material.

In class:

1. (5 min) Discuss jitts
2. (5 min) Discuss readings
3. (10 min) In pairs, discuss students' homeworks
4. (20 min) With guest star, class discussion on presentations and communication
5. (10 min) Do 2 of the students' prepared presentations in front of the whole class. For each:
    - (5 min) Student gives presentation to class
    - (5 min) Class discussion of what worked and what didn't work in the presentation
6. (10 min) In groups of 3, softly do research presentations to each other:
    - (5 min) Student A presents to students B and C
    - (5 min) B and C give feedback to A
7. (5 min) Discuss readings and homework for next class

Special guest star: Tamara Broderick, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology

Class 7b: Giving a presentation

Readings before class:

1. Adele Faber and Elaine Mazlish, *How to Talk So Kids Will Listen and Listen So Kids Will Talk*
2. Andrew Gelman, "Quick tips on giving research presentations": `https://statmodeling.stat.columbia.edu/2014/12/01/quick-tips-giving-research-presentations/`
3. Andrew Gelman, "Present each others' posters": `https://statmodeling.stat.columbia.edu/2009/06/05/present_each_ot/`

Homework due at beginning of class: If you gave your presentation in class 7a, write two short paragraphs saying what you learned from the experience. If you have not given your presentation yet, spend a few minutes improving it, in light of what you have learned from watching the other students present.

In class:

1. (5 min) Discuss jitts
2. (5 min) Discuss readings
3. (10 min) In pairs, discuss students' homeworks
4. (20 min) With guest start, class discussion on presentations and communicatio
5. (10 min) Do 1 of the students' prepared presentations in front of the whole class:
    - (5 min) Student gives presentation to class
    - (5 min) Class discussion of what worked and what didn't work in the presentation
6. (20 min) In groups of 3, softly do research presentations to each other:
    - (5 min) B presents to A and C
    - (5 min) A and C give feedback
    - (5 min) C presents
    - (5 min) A and B give feedback
7. (5 min) Discuss readings and homework for next class

Special guest star: Mariel Finucane, Mathematica Policy Research

Class 8a: Dynamic graphics

Readings before class:

1. Nikola Sander, Guy Abel, and Ramon Bauer, "The global flow of people": `http://www.global-migration.info`

2. Laura and Martin Wattenberg, "Name Voyager": http://www.babynamewizard.com/voyager

3. Andrew Gelman, "Interactive demonstrations for linear and Gaussian process regressions": https://statmodeling.stat.columbia.edu/2015/03/07/interactive-demonstrations-linear-gaussian-process-regressions/

4. Bret Victor, "Up and down the ladder of abstraction": http://worrydream.com/LadderOfAbstraction/

5. Mike Bostock, "mbostock's blocks": http://bl.ocks.org/mbostock

Homework due at beginning of class:

1. In pairs, design a dynamic graphic. Your assignment due for this class is as follows: Make some static plots and write a plan of what you want to do in the dynamic graph. Put this together as a short (approximately two-page) document. We will discuss in class, then next time you will have to make the dynamic version.

2. Make a dynamic graphic in Shiny and get it to run. It doesn't have to be related to part 1 of this assignment, it can be anything. Just get it working so that you will be prepared to do something interesting for your next assignment.

In class:

1. (5 min) Discuss jitts

2. (5 min) Discuss readings

3. (10 min) In pairs, discuss students' homeworks

4. (20 min) With guest star, discuss dynamic graphics

5. (25 min) Discuss students' homeworks in pairs and in class

6. (10 min) Discuss readings and homework for next class

Special guest star: Michael Malecki, Product Manager, Graphics, at Crunch.io

Class 8b: Dynamic graphics

Readings before class:

1. RStudio, "Teach yourself Shiny": http://shiny.rstudio.com/tutorial

2. RStudio, "Shiny": http://shiny.rstudio.com

3. RStudio, "Shiny gallery": http://shiny.rstudio.com/gallery

4. Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer, "Data-driven documents": http://idl.cs.washington.edu/papers/d3/

5. Arvind Satyanarayan, Kanit Wongsuphasawat, and Jeffrey Heer, "Declarative interaction design for data visualization": http://www.stat.columbia.edu/~gelman/communication/Satyanarayan2014.pdf

6. Simon Potter, "gridSVG": https://sjp.co.nz/projects/gridsvg/

7. Hadley Wickham, "Tidy data": http://vita.had.co.nz/papers/tidy-data.pdf

Homework due at beginning of class: In pairs, make a dynamic graphic.

In class:

1. (5 min) Discuss jitts
2. (5 min) Discuss readings
3. (10 min) In pairs, discuss students' homeworks
4. (20 min) With guest star, discuss the intersection between statistical modeling and graphics
5. (25 min) With entire class, discuss students' dynamic graphics projects
6. (10 min) Discuss readings and homework for next class

Special guest star: Jessica Hullman, Department of Computer Science & Engineering and School of Journalism, Northwestern University

Class 9a: Collaboration and consulting

Readings before class:

1. "Good resources for learning Git and GitHub": https://help.github.com/articles/good-resources-for-learning-git-and-github/
2. "Github cheat sheet": https://github.com/tiimgreen/github-cheat-sheet
3. Andrew Gelman, "Is there too much coauthorship in economics (and science more generally)? Or too little?": https://statmodeling.stat.columbia.edu/2013/06/25/is-there-too-much-coauthorship-in-economics-and-science-more-generally-or-too-little/
4. David Rindskopf, "Rules for statistical consulting": https://statmodeling.stat.columbia.edu/2008/01/25/rindskopfs_rule/
5. Andrew Gelman, "What is expected of a consultant": https://statmodeling.stat.columbia.edu/2012/11/29/what-is-expected-of-a-consultant/
6. Andrew Gelman, "My consulting policy": https://statmodeling.stat.columbia.edu/2007/12/27/my_consulting_p/
7. Andrew Gelman, "Who's holding the pen?, The split screen, and other ideas for one-on-one instruction": https://statmodeling.stat.columbia.edu/2010/12/10/tutoring_studen/

8. Andrew Gelman, "Statistics for cigarette sellers": http://www.stat.columbia.edu/~gelman/research/published/ChanceEthics4.pdf

9. Donald Rubin, "The ethics of consulting for the tobacco industry": http://www.stat.columbia.edu/~gelman/communication/Rubin2002.pdf

Homework due at beginning of class: Progress report on projects.

In class:

1. (5 min) Discuss jitts
2. (5 min) Discuss readings
3. (10 min) As a class, discuss students' homeworks
4. (20 min) With guest star, discussion of statistical consulting and collaboration
5. (25 min) In pairs and with entire class, discuss project plans
6. (10 min) Discuss readings and homework for next class

Special guest star: David Rindskopf, Department of Educational Psychology, City University of New York

Class 9b: Communication and its impact on science

Readings before class:

1. Gerd Gigerenzer, "Mindless statistics": http://www.stat.columbia.edu/~gelman/communication/Gigerenzer2004.pdf
2. Martha Smith, "Another mixed bag: Gigerenzer's mindless statistics": https://web.archive.org/web/20150710073533/http://www.ma.utexas.edu/blogs/mks/2015/02/03/another-mixed-bag-gigerenzers-mindless-statistics/
3. Andrew Gelman, "Childhood intervention and earnings": http://www.symposium-magazine.com/childhood-intervention-and-earnings/
4. Andrew Gelman, "The failure of null hypothesis significance testing when studying incremental changes, and what to do about it": http://www.stat.columbia.edu/~gelman/research/published/incrementalism_3.pdf
5. Andrew Gelman, "Post-publication peer review: How it (sometimes) really works": https://statmodeling.stat.columbia.edu/2013/09/01/post-publication-peer-review-how-it-sometimes-really-works/
6. Andrew Gelman, "How to fix the tabloids? Toward replicable social science research": https://statmodeling.stat.columbia.edu/2013/05/31/how-to-fix-the-tabloids-toward-replicable-social-science-research/

7. Andrew Gelman and Eric Loken, "The garden of forking paths: Why multiple comparisons can be a problem, even when there is no 'fishing expedition' or 'p-hacking' and the research hypothesis was posited ahead of time": http://www.stat.columbia.edu/~gelman/research/unpublished/p_hacking.pdf

8. Katherine Button, John Ioannidis, Claire Mokrysz, Brian Nosek, Jonathan Flint, Emma Robinson, and Marcus Munafo, "Power failure: why small sample size undermines the reliability of neuroscience": http://www.stat.columbia.edu/~gelman/communication/Button2013.pdf

9. Andrew Gelman and John Carlin, "Some natural solutions to the p-value communication problem—and why they won't work": http://www.stat.columbia.edu/~gelman/research/published/jasa_signif_2.pdf

10. Andrew Gelman, "Replication controversies": https://statmodeling.stat.columbia.edu/2014/11/19/24265/

11. Andrew Gelman, "Confirmationist and falsificationist paradigms of science": https://statmodeling.stat.columbia.edu/2014/09/05/confirmationist-falsificationist-paradigms-science/

Homework due at beginning of class: Prepare something on quantitative research on early childhood intervention.

In class:

1. (5 min) Discuss jitts
2. (5 min) Discuss readings
3. (10 min) In pairs, discuss students' homeworks
4. (20 min) With guest star, discussion of criticism and replication in empirical research
5. (25 min) Class discussion of crisis of criticism in the context of early childhood intervention
6. (10 min) Discuss readings and homework for next class

Special guest star: Beth Tipton, Department of Statistics and Institute for Policy Research, Northwestern University

Class 10a: Writing for a technical audience

Readings before class:

1. James Lloyd-Smith, Sebastian Schreiber, Ekkehard Kopp, and Wayne Getz, "Superspreading and the effect of individual variation on disease emergence": http://www.stat.columbia.edu/~gelman/communication/LloydSmithSchreiberKoppGetz2005.pdf

2. Andrew Gelman, "How to read (in quantitative social science). And by implication, how to write.": https://statmodeling.stat.columbia.edu/2014/12/02/read-quantitative-social-science-implication-write/

3. Andrew Gelman, "'Patchwriting' is a Wegmanesque abomination but maybe there's something similar that could be helpful?": https://statmodeling.stat.columbia.edu/2014/11/12/patchwriting-wegmanesque-abomination-maybe-theres-something-similar-helpful/
4. Andrew Gelman, "No, I didn't say that!": https://statmodeling.stat.columbia.edu/2014/09/30/didnt-say/
5. Andrew Gelman, "Advice for writing research articles": https://statmodeling.stat.columbia.edu/2009/07/30/advice_on_writi/
6. Andrew Gelman, "A tale of two discussion papers": https://statmodeling.stat.columbia.edu/2013/05/09/a-tale-of-two-discussion-papers/

Homework due at beginning of class: Write a two-page technical article. "Technical" does not necessarily mean theoretical or mathematical; rather, it implies that you are writing for a specialized audience. It's fine for this to be something that you need to write for some other purpose: not another class, but for example something you are doing at work would be fine.

In class:

1. (5 min) Discuss jitts
2. (5 min) Discuss readings
3. (10 min) In pairs, discuss students' homeworks
4. (20 min) With guest star, discuss writing for a specialized audience
5. (20 min) In pairs, read and discuss each others' technical articles:
    - (10 min) Read each others' articles
    - (5 min) Discuss student A's article
    - (5 min) Discuss student B's article
6. (10 min) With entire class, go through and discuss some of the students' technical articles
7. (5 min) Discuss readings and homework for next class

Special guest star: Leslie McCall, Department of Sociology, City University of New York

Class 10b: Writing for a non-technical audience

Readings before class:

1. Andrew Gelman and David Rothschild, "We need to move beyond election-focused folling": http://www.slate.com/articles/technology/future_tense/2017/09/what_is_the_future_of_polling.html
2. William Strunk and E. B. White, *The Elements of Style*

3. Geoffrey Pullum, "50 years of stupid grammar advice": http://www.stat.columbia.edu/~gelman/communication/Pullum2009.pdf

4. Andrew Gelman, "Politics and the English language, 2014 edition": http://www.washingtonpost.com/blogs/monkey-cage/wp/2014/08/29/politics-and-the-english-language-2014-edition/

5. Andrew Gelman and Kaiser Fung, "Freakonomics: What went wrong?": http://www.stat.columbia.edu/~gelman/research/published/freakwww.pdf

6. Andrew Gelman, "A kaleidoscope of responses to Dubner's criticisms of our criticisms of Freakonomics": https://statmodeling.stat.columbia.edu/2012/03/20/a-kaleidoscope-of-responses-to-dubners-criticisms-of-our-criticisms-of-freaknomics/

7. Dean Baker, "Influencing the debate from outside the mainstream: Keep it simple": http://www.stat.columbia.edu/~gelman/communication/Baker2014.pdf

8. Thomas Basbøll, "The good, the bad, and the popular": http://secondlanguage.blogspot.dk/2009/09/gail-hornstein-takes-up-familiar-theme.html

9. Andrew Gelman, "'How many zombies do you know?': Using indirect survey methods to measure alien attacks and outbreaks of the undead": http://www.stat.columbia.edu/~gelman/research/published/zombies.pdf

Homework due at beginning of class: Write a two-page non-technical article about a statistical topic. (By "statistical," I don't mean that it has to be statistical theory or a description of a statistical method; it could be, for example, some quantitative social science.)

In class:

1. (5 min) Discuss jitts
2. (5 min) Discuss readings
3. (10 min) In pairs, discuss students' homeworks
4. (20 min) With guest star, discuss challenges of writing for a general audience
5. (20 min) In pairs, read and discuss each others' non-technical articles:
    - (10 min) Read each others' articles
    - (5 min) Discuss student A's article
    - (5 min) Discuss student B's article
6. (10 min) With entire class, go through and discuss some of the students' non-technical articles
7. (5 min) Discuss readings and homework for next class

Special guest star: Lizzie Wolkovich, Department of Forest & Conservation Sciences, University of British Columbia

Class 11a: Data processing

Readings before class:

1. Andrew Gelman, "What is Russia's GDP per capita?": https://statmodeling.stat.columbia.edu/2009/03/16/what_is_russias/
2. Andrew Gelman, "Statistics is the *least* important part of data science": https://statmodeling.stat.columbia.edu/2013/11/14/statistics-least-important-part-data-science/
3. Andrew Gelman, "More on 'data science' and 'statistics'": https://statmodeling.stat.columbia.edu/2013/11/19/22182/
4. Hadley Wickham, "Data science: how is it different to statistics?": http://bulletin.imstat.org/2014/09/data-science-how-is-it-different-to-statistics%E2%80%89/
5. Yair Ghitza and Andrew Gelman, "The Great Society, Reagan's revolution, and generations of presidential voting": http://www.stat.columbia.edu/~gelman/research/unpublished/cohort_voting_20140605.pdf

Homework due at beginning of class: Data analysis assignment on some topic related to current research.

In class:

1. (5 min) Discuss jitts
2. (5 min) Discuss readings
3. (10 min) In pairs, discuss students' homeworks
4. (20 min) With guest star, discuss challenges in working with big data
5. (25 min) In class, discuss the students' data analysis
6. (10 min) Discuss readings and homework for next class

Special guest star: Shira Mitchell, Statistician, OpenLabs R&D

Class 11b: Programming

Readings before class:

1. Jeff Atwood, *Coding Horror Stories*: http://blog.codinghorror.com/a-pragmatic-quick-reference/
2. Bob Carpenter, "Modeling as software development," section 9 of "Bayesian Workflow": http://www.stat.columbia.edu/~gelman/research/unpublished/Bayesian_Workflow_article.pdf

Homework due at beginning of class: Your task is to scrape a dataset (or more) from the internet, then graph it. For the scraping part, some instructions are at `http://www.stat.columbia.edu/~gelman/communication/webscraping_in.md` (which you can "knit" in R using R Markdown) or `http://www.stat.columbia.edu/~gelman/communication/webscraping_in_python.py` (which you can open with a text editor or Xcode). In choosing which webpages to scrape, you may stick with the ones given in the tutorials, but scrape at least two datasets other than the ones shown in the examples, and preferably merging all information for graphing. A more interesting option is to scrape some page with data interesting to you. For example, ESPN has tons of data ripe for scraping for whichever sport interests you most. For the graphing part, keep in mind all the good plotting and communication practices we've learned so far.

Present your work (including source code) in `.html` or `.pdf` format.

In class:

1. (5 min) Discuss jitts
2. (5 min) Discuss readings
3. (15 min) Discuss students' scraping experiences
4. (45 min) With guest star, discuss programming
5. (5 min) Discuss upcoming presentations of student projects

Special guest star: Mitzi Morris, Institute for Social and Economic Research and Policy, Columbia University

Weeks 12–14: Student projects

Readings before class:

1. One page for each pair's project

Homework due at beginning of class: Prepare your presentation, along with a one-page summary to be read by everyone beforehand and a one-page handout to be given out during your presentation

In class: We will divide the total amount of time by the number of pairs of students doing projects, and then do the presentations one at a time. We follow each presentation with a brief class discussion of the project and of more general issues in statistical communication.

Special guest star: You!