Training the next generation of applied statisticians

Julia Sharp

Emily Griffith

Megan Higgs

CAUSE/JSDSE webinar series

Welcome from our host and moderator

Nicholas Horton (Amherst College)

CAUSE/Journal of Statistics and Data Science Education webinar series

Upcoming webinars:

- October 19, 2021 4:00-5:00pm “Looking back on 30 years of the *Journal of Statistics and Data Science Education* with some of the founders” webinar

Signup at https://www.causeweb.org/cause/webinars

Webinars are recorded and posted (with slides) at that same site

Consortium for the Advancement of Undergraduate Statistics Education

https://www.causeweb.org/cause

2021 Electronic Undergraduate Statistics Research Conference

FRIDAY, NOVEMBER 5TH, 2021

https://www.causeweb.org/usproc/eusrc/2021

Training the next generation of applied statisticians (Sharp, Griffith, and Higgs, JSDSE, 2021)

Brief Communications

Setting the Stage: Statistical Collaboration Videos for Training the Next Generation of Applied Statisticians

Julia L. Sharp, Emily H. Griffith & Megan D. Higgs

Pages 165-170 | Accepted author version posted online: 24 May 2021, Published online: 06 Jul 2021

Download citation https://doi.org/10.1080/26939169.2021.1934202

DATA SCIENCE FOR UNDERGRADUATES

Opportunities and Options

consensus report published in 2018
free download from
https://nas.edu/envisioningds

The National Academies of SCIENCES ENGINEERING MEDICINE

nas.edu/EnvisioningDS

A Central Finding

Finding 2.3 A critical task in the education of future data scientists is to instill data acumen. Key concepts involved in developing data acumen include:

- Mathematical foundations
- Computational foundations
- Statistical foundations
- Data management and curation
- Data description and visualization
- Data modeling and assessment
- Workflow and reproducibility
- **Communication and teamwork**
- Domain-specific considerations
- Ethical problem solving.

Communication and teamwork concepts

Key **communication and teamwork** concepts/skills that would be important for all students in their data science programs and critical for their success in the workforce are the following:

- Ability to understand client needs,
- Clear and comprehensive reporting,
- Conflict resolution skills,
- Well-structured technical writing without jargon, and
- Effective presentation skills.

DATA SCIENCE FOR UNDERGRADUATES

Opportunities and Options

consensus report published in 2018
free download from
https://nas.edu/envisioningds

The National Academies of SCIENCES ENGINEERING MEDICINE

nas.edu/EnvisioningDS

Training the next generation of applied statisticians (Sharp, Griffith, and Higgs, JSDSE, 2021)

Brief Communications

Setting the Stage: Statistical Collaboration Videos for Training the Next Generation of Applied Statisticians

Julia L. Sharp, Emily H. Griffith & Megan D. Higgs

Pages 165-170 | Accepted author version posted online: 24 May 2021, Published online: 06 Jul 2021

Download citation https://doi.org/10.1080/26939169.2021.1934202

Julia Sharp

Julia Sharp is an associate professor of statistics and the Director of the Graybill Statistics and Data Science Laboratory at Colorado State University. She is a widely recognized expert in statistical collaboration and recently was awarded the Outstanding Mentor Award from ASA's Section on Statistical Consulting.

When she is not working, Julia enjoys baking, hiking, and enjoying the company of family and friends.

Emily Griffith

Emily Griffith is an associate research professor of statistics at North Carolina State University. She is also a Fellow in the Office of Research Innovation working on development and strategy to further innovation in the university's data sciences initiatives.

In her free time, Emily enjoys running, cooking, and hanging out with her family.

Megan Higgs

Megan Higgs has worked as a collaborative statistician in academia and private industry, and is now working independently as Critical Inference LLC and writing posts for a blog of the same name. She volunteers as editor of the International Statistical Institute's "Statisticians React to the News" blog and serves on the ASA's Climate Change Committee. Megan loves spending time with her family and pets in Montana.

A need for resources

- Collaborative work is innate to being a statistician or data scientist.
- Real-world training is often a small part of a student's academic program.
- There is a dearth of materials meant to facilitate training in collaboration compared to materials for traditional statistical theory and methods.
- Instructors are often compensated less for this teaching (worth fewer credits or volunteer service), thus increasing value of accessible resources.

What did we need?

- Free and easily accessible online
- An engaging and active-learning resource for students
- Able to augment any course format (in person, on-line, hybrid)
- Flexible enough for use by a variety of instructors with different emphases and student backgrounds
- Promote student encounters with common real-life challenges
- Spur thought and discussion about communication, collaboration, as well as difficult statistical concepts

.... led us to want a modern set of videos and supporting materials.

The videos

- We produced 10 modern, freely available videos
- Portray scenarios of mock collaborative interactions
- Context of the videos is academic research discussions, yet can be generalized to other settings
- Supporting materials include
  - Discussion questions
  - Scripts
  - Other resources
- Designed to augment existing curricula
- Provide contexts for rich discussion, not templates for perfect interactions

General themes captured

- General communication, planning, and meeting-related issues
- Challenges with statistical inference
- Professional ethics

Most videos have more than one theme and there is overlap among videos.

General communication, planning, and other meeting-related issues

The videos and discussion questions in this theme focus on ways to:

- work with a domain expert to move expectations of the statistician away from a technician role and toward a more collaborative relationship (Scenario 1),
- negotiate the scope and timeline of work in a mutually beneficial way (Scenarios 2 and 4),
- discontinue a collaborative relationship and decline future work (Scenario 3), and
- end a meeting politely, yet firmly (Scenario 5).

Challenges with statistical concepts

- Students hear and critique conversations around technical statistical topics led by the statistician, and are encouraged to reflect on ways to discuss statistical concepts with non-statisticians.
- Common skills portrayed across these videos include listening and teaching in a respectful and professional manner.
- The videos show collaborative statisticians:
    - responding to broad and ill-defined questions and requests (Video Scenario 6),
    - discussing and explaining the difference between statistical significance and practical relevance (Video Scenario 7), and
    - encouraging scientific collaborators to expand their understanding of statistical thinking to include data visualization and exploratory descriptive summaries (Video Scenario 8).

Professional ethics

- Discussions can combine principles from the ASA's Ethical Guidelines for Statistical Practice (2018) with awareness of underlying tone or demeanor to demonstrate importance of a statistician understanding that requests for unprincipled analyses often come from a lack of understanding and not from ill-intent.
- In the ethics-related videos, scenarios include:
  - navigating requests for p-hacking (Scenario 9),
  - explaining pseudo-replication and its potential ramifications (Scenario 10), and
  - addressing inappropriate and potentially unethical analysis choices (Scenario 10).

Example (Scenario 8)

Introductory statistical training typically does not emphasize the importance of creatively plotting data before summarizing and aggregating. Statisticians may be in a position to advocate for and assist with effective
not have the necessary computing expertise. In this video, a researcher anc
identify important design information that would have otherwise been ov
thorough exploratory data analysis and considering data visualization as an

Video (3.13 minutes)

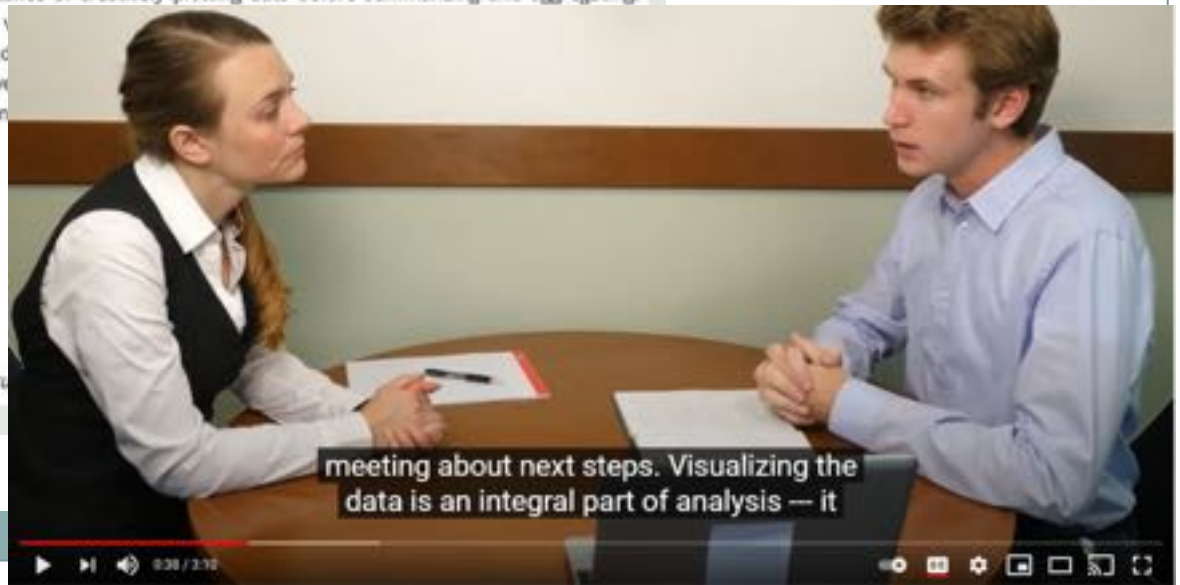
Script

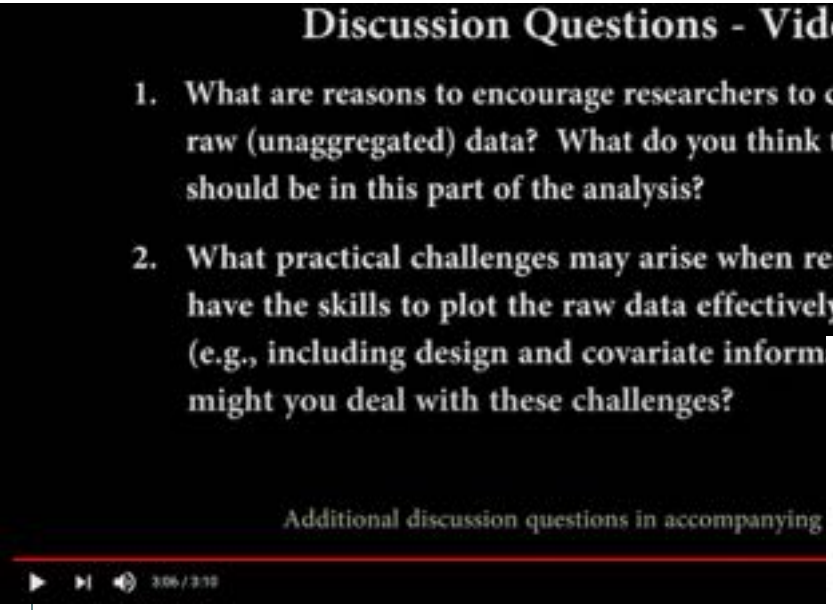
Discussion questions

Credits

To cite this video, please use the text:

- Sharp, J.L., Griffith, E., and Higgs, M. (2020). *Advocating for data visu*





Video Topic 8: Advocating for data visualization

Discussion Questions

Questions in video

1. What are reasons to encourage researchers to creatively plot the raw (unaggregated) data? What do you think the statistician's role should be in this part of the analysis?

2. What practical challenges may arise when researchers do not have the skills to plot the raw data effectively or efficiently (e.g., including design and covariate information)? How might you deal with these challenges?

Additional questions

3. What do you think the statistician did well and what could be improved in this conversation?

4. The researcher jumps to wanting to exclude data labeled as "outliers." How well do you think the statistician handled the situation? What other contexts could potential exclusion of data arise and what are strategies to communicate about the implications of excluding data?

Video Topic 8: Advocating for data visualization

Introductory training in Statistics typically does not emphasize the importance of creatively plotting data before they are summarized or aggregated. Plotting techniques are often focused on displaying results for publication, or checking assumptions of particular models. Software for making sophisticated plots is becoming more accessible and data visualization should be used as an integral part of any rigorous analysis. Statisticians are often in the position to advocate for this, to bring a different type of creativity, and may bring necessary computing skills to the collaboration. It is also a great way for the statistician to make sure they understand the design and data before any model fitting is undertaken. In this video, a researcher and statistician discuss data visualization as a first step in the analysis, and identify important design information that would have been overlooked without it.

STATISTICIAN	Good to see you again! Last meeting we discussed a possible approach and as a first step, I created some visualizations of your data.
RESEARCHER	I don't know if it's really worth the time to go through visualizations... I probably won't have space to include more than one figure when I publish. I was really looking forward to seeing the results!
STATISTICIAN	I apologize for any misunderstanding. I should have been more clear at the end of our last meeting about next steps. Visualizing the data is an integral part of analysis --- it helps us understand the information available, check for possible errors, and assess modeling assumptions.
RESEARCHER	I guess that makes sense, it's just not something we ever did in my stats classes. I do always plot averages and error bars though.
STATISTICIAN	That's very common, though software and guidelines for plotting raw data are

Link to the Videos

https://community.amstat.org/statisticalcollaborationtrainingforappliedstatisticians/home

Questions

1. What lessons did you learn from developing these resources?

Questions

2. If you had time and energy, what topics would be included in future videos?

Questions

3. What guidance do you have for instructors and curriculum developers in developing communication and collaboration skills?