Meeting Notes

June 16

Agenda Draft (75 min–1:30 to 2:45 ET):

Target size of breakout groups: 5-7?

Breakout group roles: - note taker - speaker (for large group) - "includer"–make sure everyone shares - timekeeper / keep conversation moving to include multiple prompts - etc

- (1:30-1:40pm; 10 min) Opening remarks about EDA (and IDA initial data analysis)–each sharing:
  - why we're interested in the topic
  - our current perspectives (individually)
  - which courses/student populations are front of mind
- (1:40-1:55pm; 15 min) small group discussion prompt(s)–e.g.,
  - what brought you here? why the interest in EDA breakout?
  - what should a thorough EDA include?
  - where/how should students learn about EDA?
  - how to prioritize EDA among existing topics (what can/should we give up)?
- (1:55-2:05pm; 10 min) large group discussion based on observations from small groups
- (2:05-2:15pm; 10 min) small groups review & discuss various guidance for EDA–e.g.,
  - Velleman & Hoaglin (2012). Exploratory data analysis. In *APA handbook of research methods in psychology* <doi.org/10.1037/13621-003>
  - Shan (2020). An extensive step by step guide to exploratory data analysis. *Towards data science.*
  - Wickham & Grolemund (2020). *R for Data Science* <r4ds.had.co.nz>
- (10 min) large group compare & contrast observations from small group discussions. Any revisions or new insights about what a thorough EDA should include?
- (10 min) small group discussions prompt(s)–e.g.,
  - Data science has helped to shift attention toward a wider view of data (volume, variety, etc). Which (if any) priorities for EDA need to adapt, and which are unchanged?
  - Do we need to rethink where and how EDA should be addressed in Statistics & Data Science courses
- (15 min) large group discussion based on observations from small groups

Other prep & next steps:

- Remarks about origin of EDA–what was happening in Tukey's day?
- each prepare to lead a summary on the various EDA resources:
  - Velleman & Hoaglin (2012). Exploratory data analysis. In *APA handbook of research methods in psychology* <doi.org/10.1037/13621-003>

- Shan (2020). An extensive step by step guide to exploratory data analysis. *Towards data science.*
- [LL] Wickham & Grolemund (2020). *R for Data Science* <r4ds.had.co.nz>

(optional if needed to fill time) - think about how we emphasize & assess EDA in our own courses if needed to share at the end

March 25

Author Order

Christopher Desjardins, Lisa Lendway, Matthew Beckman

I want to suggest alphabetical author order, and I feel like there's a running joke at USCOTS that Allan Rossman is always trying to get alphabetical by **first** name to catch on.

Affiliations

- Christopher Desjardins–Saint Michael's College
- Lisa Lendway–Macalester College
- Matthew Beckman–Penn State University

Title:

- Reframing EDA to keep up with the times (Matt sort of likes this one)
- Closing the gap between real & perceived value of EDA
- Make EDA Great Again! (n.b. 90% chance of hats if we choose this, though it makes Matt uncomfortable that it would appear on his CV)

(not the title): Is Exploratory Data Analysis Dying When We Need it Most?

This 1-2-paragraph description should include a condensed version of what you provided in your proposal:

- Overview of topic
- Specific goals of session
- Indication of how session will be interactive

Session Description (194 of 250 words):

Exploratory Data Analysis (EDA) perhaps suffers a wider gap between real & perceived value than any contribution the data analyst offers. The popularity of Data Science has ushered in a rising tide of open-data initiatives and elegant software solutions readily accessible to non-specialists, thereby democratizing data analysis for all manner of subjects. Initially, some novices may feel overwhelmed and others may risk spurious results utilizing opaque methods with fancy names, yet all can be empowered by careful EDA based on lucid methods powered by curiosity, creativity, and critical thinking to build fruitful intuition and responsible insights. However, EDA too often seems a cursory obligation to be minimized.

The session will include several focus group style conversations in small groups reflecting on issues including the evolution of contemporary EDA, top priorities & best practices for a careful EDA to be reinforced with our students, and other topics. As a result, the session aims to highlight areas of consensus and debate as we (1) recast/affirm goals and considerations for EDA in a world laden with found data, big data, etc. (2) rethink where/how EDA should be addressed in data analysis courses at all levels.