Teaching Soft Skills in Data Science Curriculum

Hunter Glanz, Dennis Sun, and Alex Dekhtyar

California Polytechnic State University
San Luis Obispo, California, USA

July 31, 2019

Outline

Our Aim

- Improve students' abilities to collaborate and communicate with peers or team members.

Our Aim

- Improve students' abilities to collaborate and communicate with peers or team members.
- Give students formal training on their (technical) writing skills.

Our Aim

- Improve students' abilities to collaborate and communicate with peers or team members.
- Give students formal training on their (technical) writing skills.
- Give students formal training on their (technical) oral communication skills.

Outline

Statistics at Cal Poly

- Standalone Statistics Department
- Quarter system

Admission to the Statistics Minor program requires successful completion of *one* of the following pairs of classes:

		Qtr/Yr Taken	Grade			Qtr/Yr Taken	Grade
a)	STAT 217 or 218			AND	STAT 313		
b)	STAT 251			AND	STAT 252		
c)	STAT 301			AND	STAT 302		
d)	STAT 312			AND	STAT 313[1]		
e)	STAT 321			AND	IME 326		

In addition to the preceding course pairs, you will be required to complete *four (4)*[2] of the following courses to complete the Statistics Minor:

STAT 305 or STAT 425	STAT 405	STAT 417	STAT 426
STAT 323	STAT 410	STAT 418	STAT 427
STAT 324 or STAT 334	STAT 414	STAT 419	STAT 434
STAT 330	STAT 415	STAT 421	
STAT 331	STAT 416	STAT 423	

MAJOR COURSES (95)	*Units*
STAT 150 Intro to Stat	2
MATH 141 Calculus I (B1) [1]	4
MATH 142 Calculus II (B1) [1]	4
MATH 143 Calculus III	4
MATH 206 Linear Algebra I	4
MATH 241 Calculus IV	4
STAT 301 Statistics I	4
STAT 302 Statistics II	4
STAT 305 Intro to Probability and Simulation	4
STAT 323 Design & Analysis of Experiments I	4
STAT 330 Statistical Computing with SAS	4
STAT 331 Statistical Computing with R	4
STAT 334 Applied Linear Models	4
STAT 365 Statistical Communication	2
STAT 425 Probability Theory	4
STAT 426 Estimation and Sampling Theory	4
STAT 427 Mathematical Statistics	4
STAT 461 Senior Project I	1
STAT 462 Senior Project II	2
STAT 465 Statistical Consulting	4
Statistics Electives: Select 12 units from List A: STAT 405, 414, 415, 416, 417, 418, 419, 421, 423, 434	12
Select 12 units from List B: Any 400-level STAT course (including those in List A), CSC/CPE 202, 203, CSC 236, 348, 349, 365, 369; DATA 301; IME 430; ITP 303; MATH 242, 306, 335, 336, 406, 412, 413, 414, 437, 451	12

SUPPORT COURSES (16)	
CSC/CPE 101 or CSC 235	4
MATH 248 Methods of Proof in Mathematics	4
Approved support electives [2]	8

Cross-Disciplinary Studies Minor in Data Science

Cross-Disciplinary Studies Minor in Data Science

- Cross-Disciplinary Studies Minor (CDSM):
  1. Result of partnership between 2+ target major programs.

Cross-Disciplinary Studies Minor in Data Science

- Cross-Disciplinary Studies Minor (CDSM):
  1. Result of partnership between 2+ target major programs.
  2. Students gain depth in the partner discipline.

Cross-Disciplinary Studies Minor in Data Science

- Cross-Disciplinary Studies Minor (CDSM):
  1. Result of partnership between 2+ target major programs.
  2. Students gain depth in the partner discipline.
  3. Students gain focused study in their own discipline.

Cross-Disciplinary Studies Minor in Data Science

- Cross-Disciplinary Studies Minor (CDSM):
  1. Result of partnership between 2+ target major programs.
  2. Students gain depth in the partner discipline.
  3. Students gain focused study in their own discipline.
  4. Students gain focused study in the mutual domain of the minor.

Cross-Disciplinary Studies Minor in Data Science

- Cross-Disciplinary Studies Minor (CDSM):
  1. Result of partnership between 2+ target major programs.
  2. Students gain depth in the partner discipline.
  3. Students gain focused study in their own discipline.
  4. Students gain focused study in the mutual domain of the minor.
- Our CDSM in Data Science:
  - 3-4 years old
  - Marriage of Statistics and Computer Science
  - Created by Andrew Schaffner (STAT) and Alex Dekhtyar (CS)
  - Built primarily for STAT and CS majors, but open to all.

Cal Poly CDSM in Data Science

CSC/CPE 101	Fundamentals of Computer Science	4
CSC/CPE 202	Data Structures	4
CSC/CPE 203	Project-Based Object-Oriented Programming and Design	4
CSC 348	Discrete Structures	4
or MATH 248	Methods of Proof in Mathematics	
CSC 349	Design and Analysis of Algorithms	4
CSC 365	Introduction to Database Systems	4
CSC 369	Introduction to Distributed Computing	4
CSC 466	Knowledge Discovery from Data	4
DATA 301	Introduction to Data Science	4
DATA 401	Advanced Topics in Data Science	4
DATA 451	Data Science Capstone I	2
DATA 452	Data Science Capstone II	2
MATH 143	Calculus III	4
MATH 206	Linear Algebra I	4
or MATH 244	Linear Analysis I	
STAT 302	Statistics II	4
or STAT 312	Statistical Methods for Engineers	
STAT 305	Introduction to Probability and Simulation	4
STAT 334	Applied Linear Models	4
STAT 331	Statistical Computing with R	4
STAT 419	Applied Multivariate Statistics	4
Technical Electives (CSC/STAT/DATA/MATH/PHYS)		8
Total units		**80**

Outline

Natural Opportunities for Communication Gains

Natural Opportunities for Communication Gains

- Our data science students work with both R and Python, mostly within the R Markdown/Notebook and Jupyter Notebook interfaces.

Natural Opportunities for Communication Gains

- Our data science students work with both R and Python, mostly within the R Markdown/Notebook and Jupyter Notebook interfaces.
- Many of the STAT and CS courses involve or accommodate group projects with a presentation and/or write-up component.

Natural Opportunities for Communication Gains

- Our data science students work with both R and Python, mostly within the R Markdown/Notebook and Jupyter Notebook interfaces.
- Many of the STAT and CS courses involve or accommodate group projects with a presentation and/or write-up component.
- The DATA courses were newly created with the construction of the CDSM!

Senior DATA Sequence

Senior DATA Sequence

1. DATA 401: Advanced Topics in Data Science

Senior DATA Sequence

1. DATA 401: Advanced Topics in Data Science
   - Mathematical foundations of machine learning
   - Evolving to include more on ethics and data science culture
2. DATA 451 & 452: Data Science Capstone I & II

Senior DATA Sequence

1. DATA 401: Advanced Topics in Data Science
   - Mathematical foundations of machine learning
   - Evolving to include more on ethics and data science culture
2. DATA 451 & 452: Data Science Capstone I & II
   - Client project-based sequence of two courses
   - Teams of students engage in data science consulting projects with real clients (both academic and industry) over 20 weeks

DATA 401: Advanced Topics in Data Science

DATA 401: Advanced Topics in Data Science

- Natural opportunities continued:
  - Group projects (written and oral deliverables)
  - Jupyter Notebooks
  - GitHub

DATA 401: Advanced Topics in Data Science

- Natural opportunities continued:
  - Group projects (written and oral deliverables)
  - Jupyter Notebooks
  - GitHub
- Newer opportunities
  - Peer assessment of assignments
  - Oral presentation of solutions to homework and in-class activity problems

DATA 451 & 452: Data Science Capstone I & II

DATA 451 & 452: Data Science Capstone I & II

- Teams of 3-4 students matched with a client for 20 weeks of data science consulting, e.g.:
  - Visualization and analysis of wearable monitor data (Kinesiology and Public Health)

DATA 451 & 452: Data Science Capstone I & II

- Teams of 3-4 students matched with a client for 20 weeks of data science consulting, e.g.:
  - Visualization and analysis of wearable monitor data (Kinesiology and Public Health)
  - Identification of malicious network activity (Industry)
  - Both local and remote clients

DATA 451 & 452: Data Science Capstone I & II

- Teams of 3-4 students matched with a client for 20 weeks of data science consulting, e.g.:
  - Visualization and analysis of wearable monitor data (Kinesiology and Public Health)
  - Identification of malicious network activity (Industry)
  - Both local and remote clients
- Minimum oral deliverables:
  - 30 minute meetings between team and client every other week (10 total)
  - Two project presentations: mid-project and project end

DATA 451 & 452: Data Science Capstone I & II

- Teams of 3-4 students matched with a client for 20 weeks of data science consulting, e.g.:
  - Visualization and analysis of wearable monitor data (Kinesiology and Public Health)
  - Identification of malicious network activity (Industry)
  - Both local and remote clients
- Minimum oral deliverables:
  - 30 minute meetings between team and client every other week (10 total)
  - Two project presentations: mid-project and project end
- Minimum written deliverables:
  - Commented and documented Notebook/code files
  - Two reports: mid-project and project end

Boosting Oral Communication

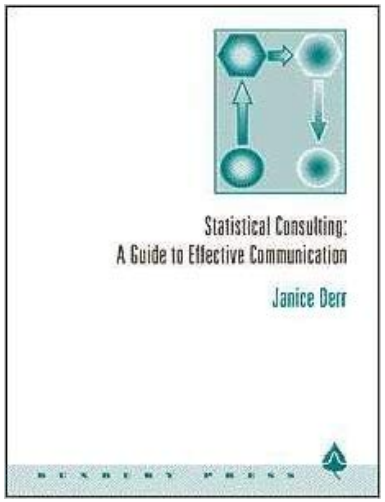
- Borrowing from STAT 465 (our consulting course) and beyond!

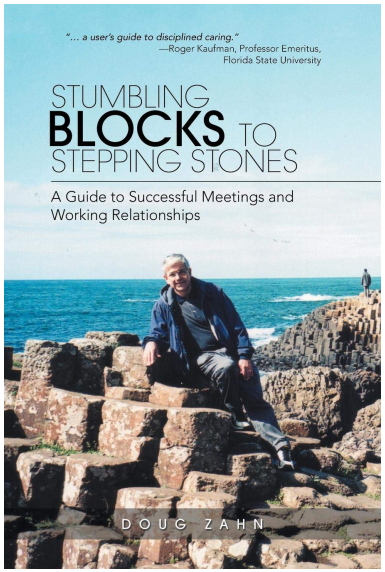
ESSENTIAL COLLABORATION SKILLS: THE ASCCR FRAME FOR COLLABORATION

A PREPRINT

Eric A. Vance*
Department of Applied Mathematics
Laboratory for Interdisciplinary Statistical Analysis (LISA)
University of Colorado Boulder
Boulder, CO 80309-0526
Eric.Vance@Colorado.EDU

Heather S. Smith
Department of Statistics
Cal Poly San Luis Obispo
San Luis Obispo, CA 93407
hsmith@calpoly.edu



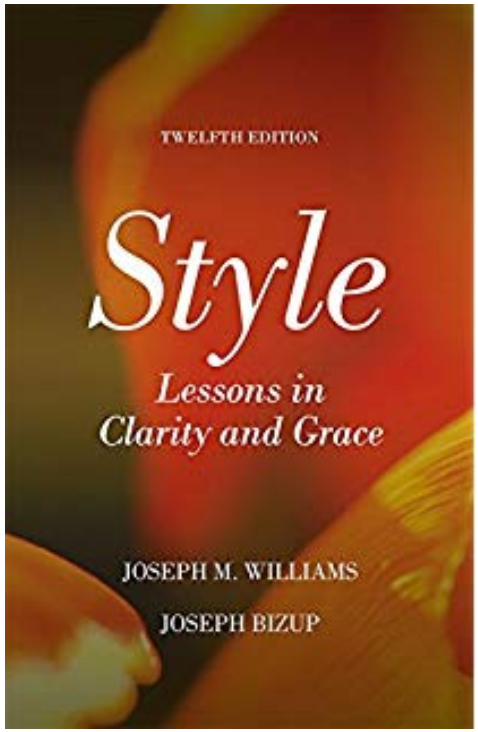


Boosting Written Communication

- Applying lessons in style and grace to their own reports from DATA 401 and eventually DATA 451!

Boosting Written Communication

- Applying lessons in style and grace to their own reports from DATA 401 and eventually DATA 451!
- Oral presentations of their revisions accompanied by discussions of the lessons themselves.



Conclusion

- Improve students' abilities to collaborate and communicate with peers or team members.

Conclusion

- Improve students' abilities to collaborate and communicate with peers or team members.
- Give students formal training on their (technical) writing skills.

Conclusion

- Improve students' abilities to collaborate and communicate with peers or team members.
- Give students formal training on their (technical) writing skills.
- Give students formal training on their (technical) oral communication skills.

Thank You! Questions?

The Section on Statistics and Data Science Education invites you to contribute feedback on this presentation.

You can use the JSM app or a paper form!

The feedback will be used to pick the Ron Wasserstein Best Contributed Paper Award winner

- Slides at https://github.com/hglanz/jsm2019-datasciencetalk
- Data Science Education blog (by Jo Hardin, Nick Horton, and me) https://teachdatascience.com/