Building an inter-disciplinary statistics data repository

Laura Taylor
Ryne VanKrevelen
Mark Weaver
Department of Mathematics and Statistics

Our Goal

- Create an inter-disciplinary repository containing data collected by students/faculty/staff at our university
- Expose introductory statistics students to the usefulness of statistics in other disciplines
- Provide instructors with data they can use for lessons, activities, quizzes, projects, etc.

Presentation Overview

- Collecting Data
- Types of Data
- Organizing the Data
- Sharing the Data
- Possible Next Steps
- What We've Learned

Collecting Data

- Created FAQ document to share
- Identified departments on campus and possible contacts
- Reached out in waves (~ one department per month per team member)

Types Of Data

- Topics: Many disciplines (STEM, humanities, arts, etc.)
- Size: anywhere from 50 or fewer to 500+ records
- Statistical uses: Useful for teaching data visualization, t-tests, ANOVA, regression, data cleaning, etc.
- Missing data included (or excluded?)!
- Students may have even helped to collect original data!

Organizing The Data

- Created a repository on Microsoft Teams
- "Staged" in development folder
- Folders for each discipline and subfolders for projects within disciplines
  - Data (raw and/or cleaned)
  - Background information
  - Other files (papers, videos, etc.)
- Next question: How can we make it easy to use?

Organizing The Data: Microsoft Teams

General Posts Files Data Catalog Data Use + Meet

Read Me!

Start Here

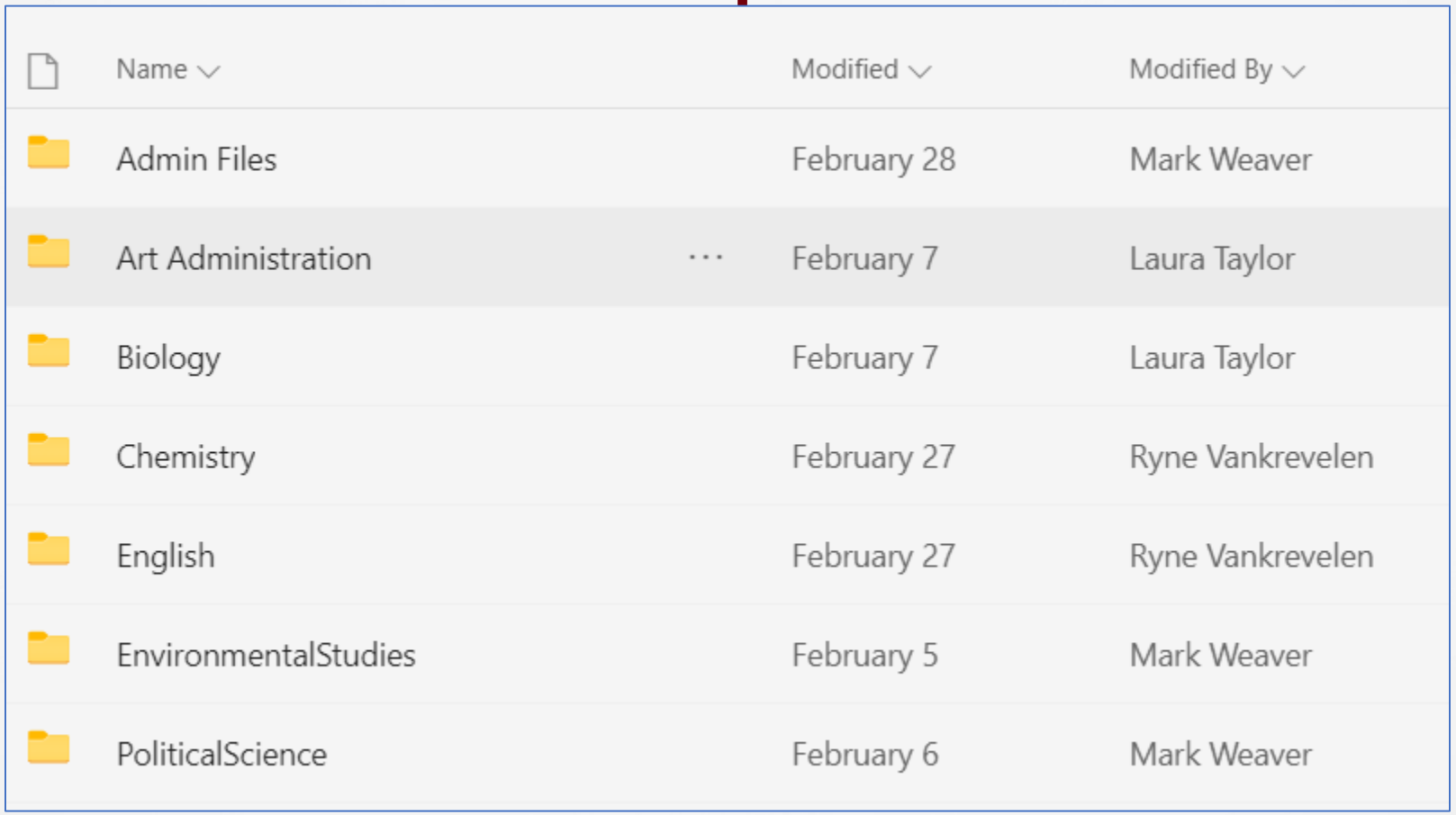
Welcome to our data repository! This resource was developed with the support of an Elon Innovates grant. In this repository, we will provide access to data contributed by faculty from many different disciplines across campus. The primary purpose of this repository is to emphasize the interdisciplinary nature of the field of Statistics and allow students to see the utility of statistics in other disciplines. To help contextualize the data, we will also include background information, web links, or academic papers, as provided by the contributing faculty.

- To locate data for your classes or research projects, you can first search keywords in the "**Data Catalog**" tab at the top of this page to help identify files that might have the type of data that you need.
- You can access data files in the "**Files**" tab at the top of this page.

Organizing The Data: Keyword-Searchable Data Catalog

Subject	Subfolder	File Name	Statistic Keywords	Context provided?
EnvironmentalStudies	Elon Soil	ElonSoil	ANOVA; ttest	Yes
EnvironmentalStudies	Elon Ponds	ElonPondsF2020	ANOVA; ttest; SLR; MLR	Some
EnvironmentalStudies	Elon Ponds	ElonPondsS2021	ANOVA; ttest; SLR; MLR	Some
EnvironmentalStudies	Elon Soil	SoilEcology	ANOVA; 2-way ANOVA; ttest	No
EnvironmentalStudies	Elon Soil	SoilEcology14	ANOVA; 2-way ANOVA; ttest	No
PoliticalScience	Climate Change Activism	Sparks_Study1	SLR; MLR; chi-square; t-test	Yes
PoliticalScience	Climate Change Activism	Sparks_Study2	SLR; MLR; chi-square; t-test	Yes
Art Administration	Museum Data	Museum Data	data cleaning; SLR, MLR, chi-square test; t-te	Some
English		COVID_Writing	Categorical graphing; Free response data	Yes
Chemistry	Pennies	PennyData	SLR; two sample ttest; outliers; graphing	Yes
Biology	Lizards Running	Sprint_Data_STS.xlsx	ANOVA; SLR; MLR; chi-square; t-test	Yes

Organizing The Data: File Structure by Discipline



Name	Modified	Modified By
Admin Files	February 28	Mark Weaver
Art Administration	February 7	Laura Taylor
Biology	February 7	Laura Taylor
Chemistry	February 27	Ryne Vankrevelen
English	February 27	Ryne Vankrevelen
EnvironmentalStudies	February 5	Mark Weaver
PoliticalScience	February 6	Mark Weaver

Organizing The Data, with Context!

Name	Modified	Modified By
FellowsPaper1	February 7	Laura Taylor
Lizard Sprint Speed Study.pptx	February 7	Laura Taylor
lizardrunningvideo1.MOV	February 7	
lizardrunningvideo2.MOV	February 7	
lizardrunningvideo3.MOV	February 7	
README.docx	February 7	
Sprint_data_STS.xlsx	February 27	

Sharing The Data: Tracking Data Use

Please let us know how you used these data in your classes or projects

Name	Course Used in	Data File Used	Briefly, how did you use these data?	Any feedback or suggestions for these data?
	STS2120	Lizards Running	I used two variables, sex and max_velocity, in a two-independent samples hypothesis test question on a unit exam	This is a great data set with all types of uses in STS2120!
	STS2320	Sparks study 1	I'm using these data as part of an activity to introduce MLR models, tests, and assumptions	Lots of possibilities with these data! Worked really well for MLR!

Possible Next Steps

- Create way for students to share data from their research at our annual undergraduate research forum
- Involve student researchers to ‘clean’ data, create example analyses, and help manage repository
- Supplement with other relevant open source data
- Add example assignments

Questions, Suggestions, Etc.?

Want to chat about what we've learned?
Have ideas to share with us?
Curious about how to implement this at your institution?

Visit our poster session on Tuesday, May 24th from 2:40 – 3:20 pm ET or reach out to any of us via email!

Laura Taylor
ltaylor18@elon.edu

Ryne VanKrevelen
rvankrevelen@elon.edu

Mark Weaver
mweaver11@elon.edu