ISLE
A Browser-Based E-Learning Platform for Teaching Statistics & Data Analysis While Learning How Students Learn

Philipp Burckhardt
Francis R. Kovacs, Rebecca Nugent, and Ron Yurko

Carnegie Mellon University
Statistics & Data Science

CAUSE Webinar, December 11, 2018

Carnegie Mellon University

- Private university in Pittsburgh, PA
- R1 research university designation
- $\approx$ 7000 undergrads, 7000 grads
- Six undergraduate colleges (admission is college-specific) College of Fine Arts, College of Engineering, Mellon College of Science, School of Computer Science, Tepper School of Business, Dietrich College of Humanities & Social Sciences
- Economics (joint in Tepper), English, History, Information Systems, Institute for Politics and Strategy, Modern Languages, Philosophy, Psychology, Social and Decision Science, Statistics & Data Science
- Around 520 primary/additional majors; Statistics (Concentration, Economics-Statistics, Statistics and Machine Learning)
- Teach about 1/3 of the campus every semester

Dietrich General Education Curriculum

In midst of multiple-year revamp/design of new requirements

- Reasoning with Data (Intro Stat & DS with ISLE)
- Writing minis: choose 2 of 3 themes (e.g. *Writing with Data*)
- More interdisciplinary courses
- Experiential Learning and Service
- Portfolio and Self-Reflection

Instead of the final result, more of a focus on how they got there

Our goals for new course/ISLE

- Remove student reliance on cookbook approach with specific software package
- Remove computing cognitive load almost entirely
- Student-driven analyses grounded in (sort of) open-ended case studies
- Include more modern concepts (e.g., nonlinear smoothers, clustering, networks)
- Written reports, presentations (oral/poster)

Understand how/why students are analyzing data

What are the impacts of early decisions on downstream analyses?

Is this a function of student background? Interests?

What on earth are they doing? **Behavioral Data Science**

ISLE: Integrated Statistics Learning Environment

- Browser-based; Modern Web Technologies (JavaScript, React.JS)
- No need for server architecture; calculations are local
- For in-class and out-of-class use
- No coding necessary
- Adaptable; instructors can change questions during lab/class
- Instructors can write notes, automatically stored ("doc cam"); students can also take their own notes on their own computers
- Action logs are stored and reproducible; every click, every typed word, every choice, every cursive letter, every spoken word
- Instructors can summarize/visualize student responses for mediation, discussion, etc.
- Editor for authoring ISLE lessons

ISLE is used in several CMU classes; today we focus on the intro course

Statistical Widgets

ISLE has many of the standard animations/demonstrations available in similar introductory statistics platforms

- Distribution Calculators
- Confidence Intervals
- Hypothesis Testing
- Central Limit Theorem
- Conditional Probability
- ...

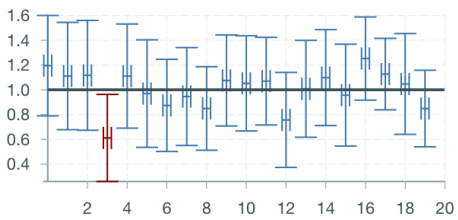
Confidence Interval Coverage for Sample Mean

$X \sim \text{Normal}(\mu, \sigma)$. Then $\bar{X} \sim \text{Normal}\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$. Our confidence interval is then $\bar{X} \pm Z_{\alpha/2} \cdot \sigma/\sqrt{n}$. For our choice of sample size (n), μ, σ, and confidence level, we will simulate 20 different samples from our normal distribution and calculate the corresponding sample means and confidence intervals.

Change parameters

Sample size (n): 30
Mean (mu): 1
Standard deviation (sigma): 1
Confidence level: 0.95
Generate

Confidence Intervals



Of the 20 confidence intervals, 19 capture the true mean **(coverage: 0.95)**.

What happens to a sampling distribution of the mean as one increases the sample size?

Your answer:

Enter your answer here...

Submit Show Solution

What is the mean of the standard normal distribution?

Your answer = 0

Submit Get Hint

Parameters

One-Sample
Hypothesized mean (null hypothesis): 0

Let's assume that we have observed data with the following characteristics

First Sample
Enter a number :
Sample mean: 0.3
Sample standard deviation: 1.5
Sample size: 25

We conduct the following test (click on the formula to switch between the one-sided variants and the two-sided test):

$H_0 : \mu = 0 \; vs. \; H_1 : \mu \neq 0$

We calculate the following test statistic:

$$z = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

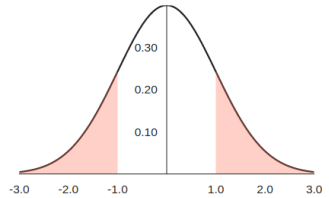
Test Result

Plugging in our values, we have:

$$z = \frac{0.3 - 0}{1.5/\sqrt{25}} = 1$$

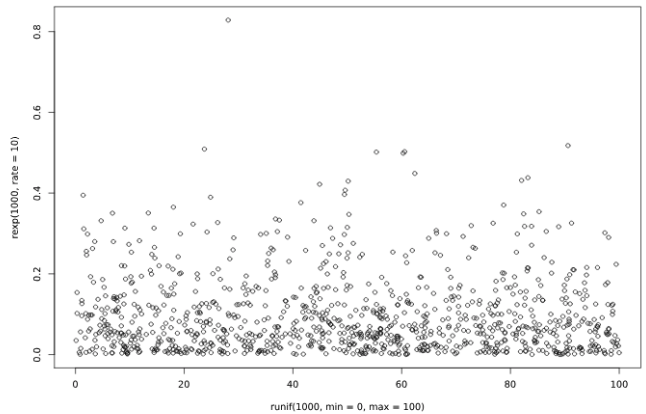
Under the null hypothesis, we calculate the p-value:

$P(|Z| > 1) = 0.317$




```
1  set.seed(1356)
2  plot(x = runif(1000, min = 0, max = 100),
3       y = rexp(1000, rate = 10))
4
```

Evaluate



Interactive Labs

- Hands-on e-learning modules that may be used both outside and **inside** the classroom
- Instructors can monitor their labs in real-time and understand how students interact with class material
- Student answers are saved and may be retrieved, e.g. when preparing for exams

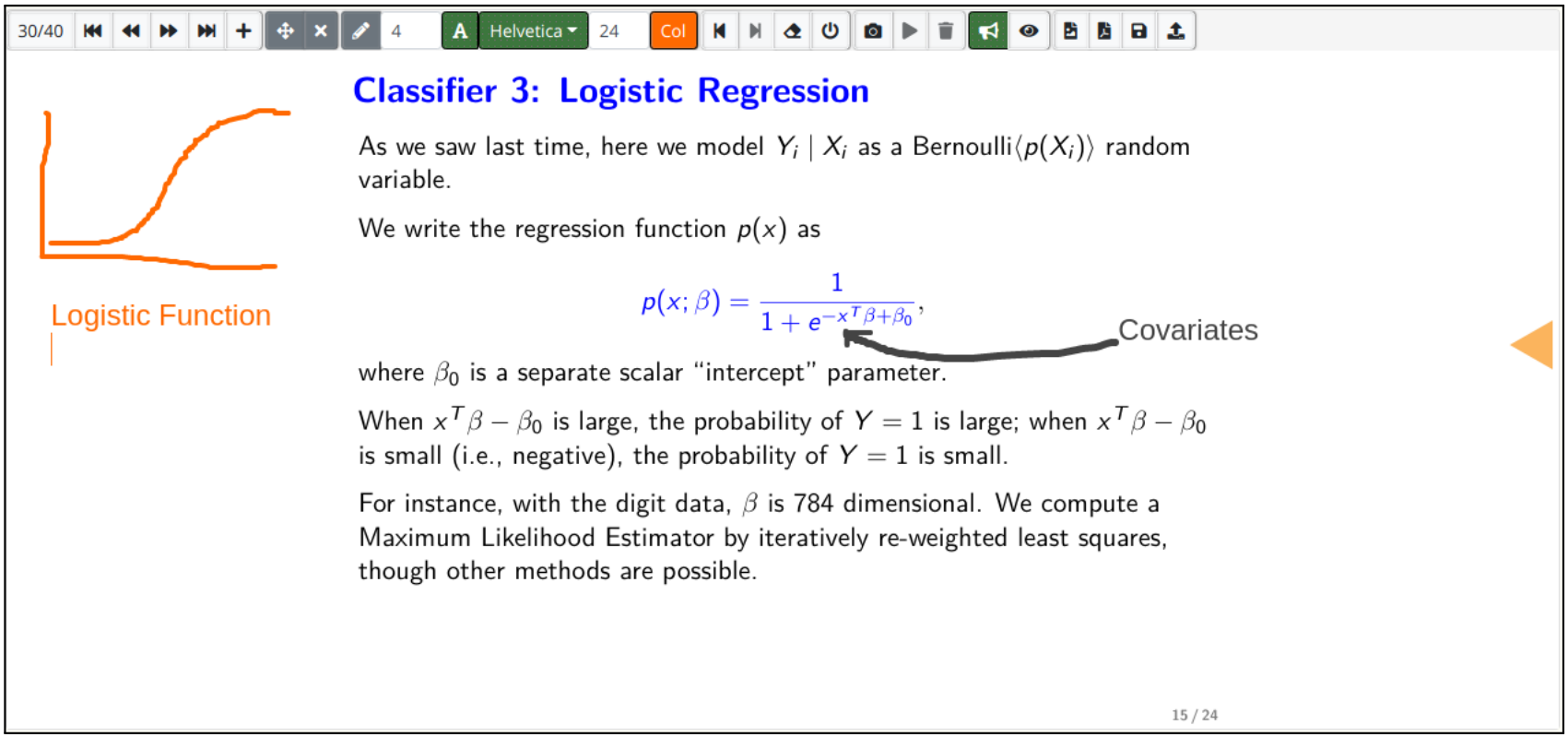
Instructor and Student Views

Beyond Intro Statistics: Analyzing Texts, Networks, Images

Real-Time Communication

- Chat Communication
- Peer-Review
- ...

Annotate Slides with the ISLE Sketchpad

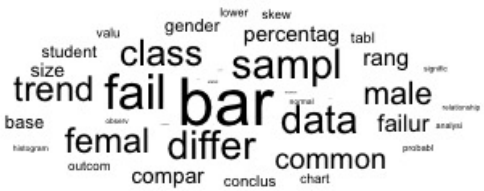


- Instructors can draw on lecture slides and upload the annotated slides with a single click
- Annotations are transmitted in real-time to students
- All other ISLE components can be overlaid on top of the slides

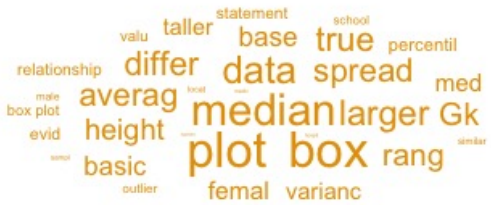
Data Explorer

Text Mining Student Answers Reveals Variation in Word Usage Depending on the Visualization

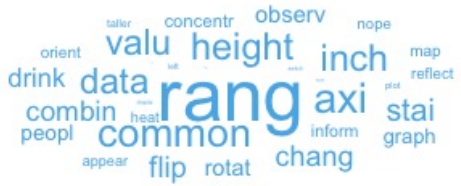
barchart



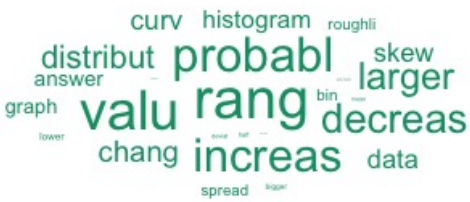
boxplot



heatmap

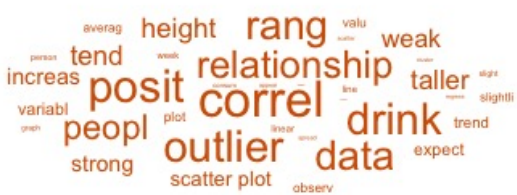


histogram

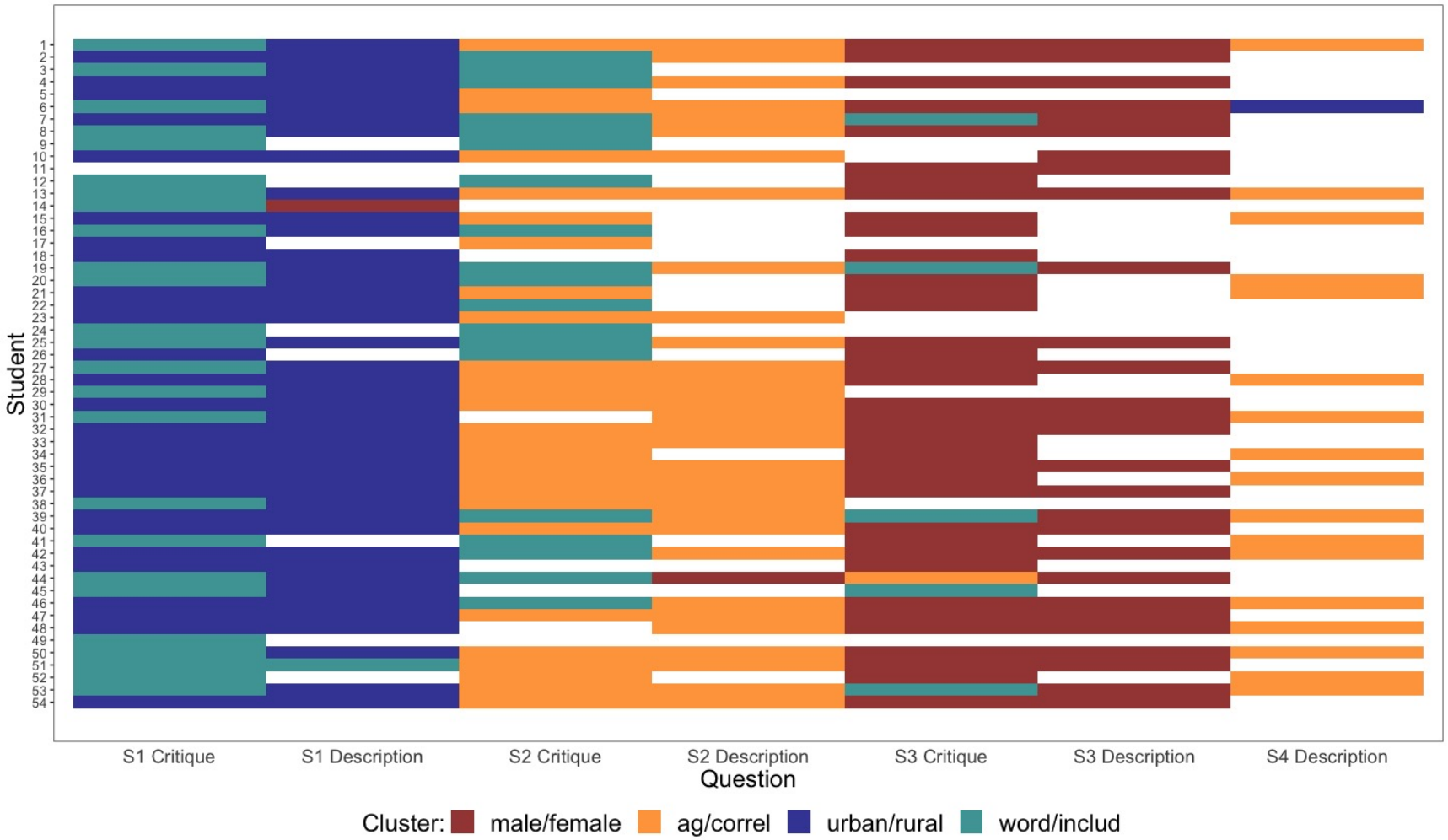


mosaic

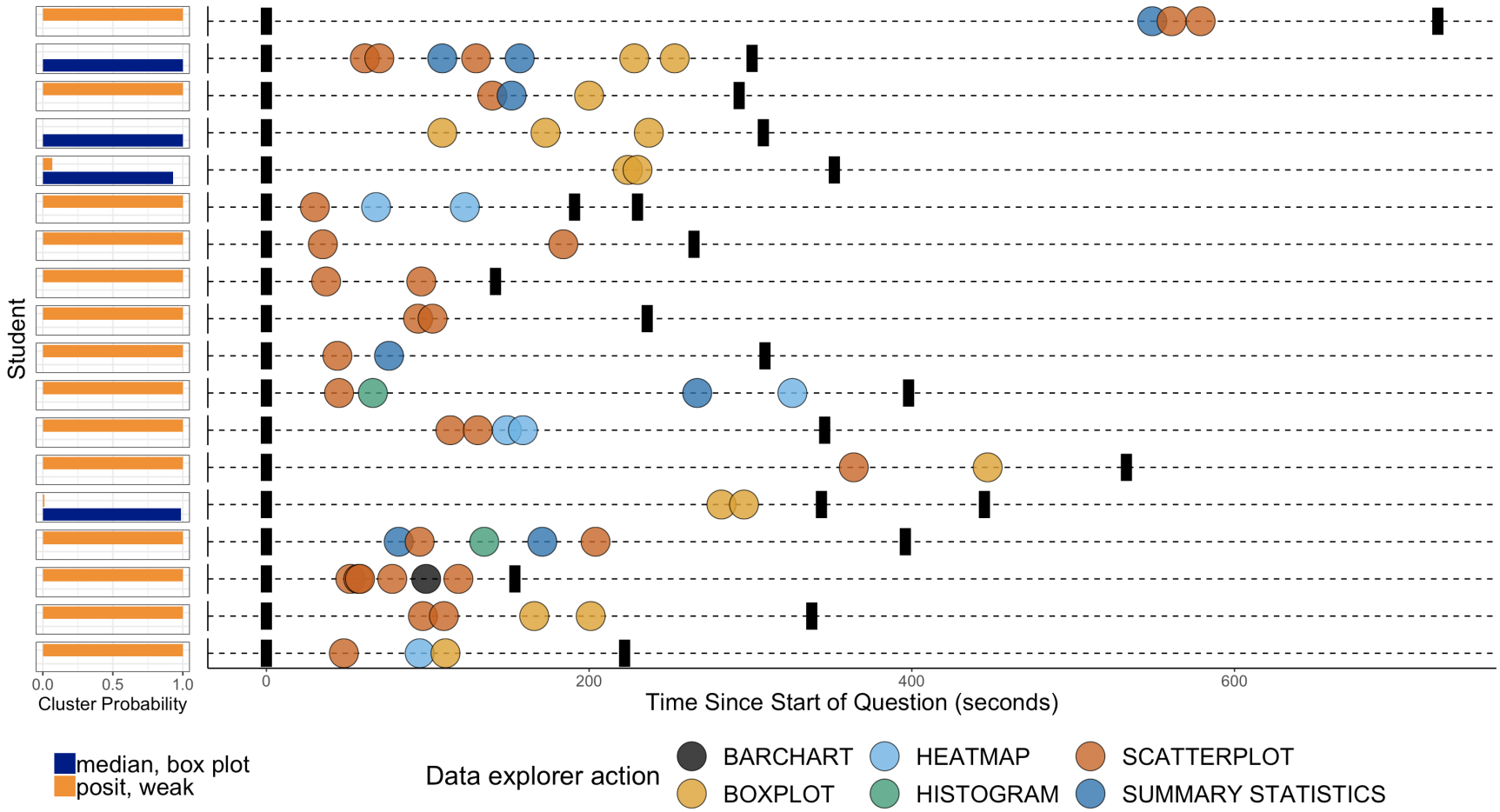
scatterplot



Cluster Student Answers to Identify Unique Responses



Timeline of Student Actions Reveals Effect of Visualization Choice on the Text of Student's Answer



Contact Information

ISLE Project Page

`http://stat.cmu.edu/isle/`

Presenters

- Philipp Burckhardt, `pgb@andrew.cmu.edu`
- Francis R. Kovacs, `fkovacs@andrew.cmu.edu`
- Rebecca Nugent, `rnugent@stat.cmu.edu`
- Ronald Yurko, `ryurko@andrew.cmu.edu`