R for Everyone — An Interactive, Ethics-Aware Data Science Micro-Curriculum for Middle School

Tene Ariyo

November 04, 2025

R for Everyone — An Interactive, Ethics-Aware Data Science Micro-Curriculum for Middle School
Capstone Project — Introduction Draft
Author: Tene Ariyo **Date:** November 04, 2025

Introduction & Motivation

Data is now a native language for civic life and work. Students scroll past charts on social media, interpret dashboards in science class, and encounter numeric claims about health, climate, and finance every week. To participate well, they need guided practice with the statistical problem-solving process, posing investigable questions, collecting or considering data, analyzing patterns and variability, and interpreting results with uncertainty and audience in mind (Bargagliotti et al. 2020).

Why middle school?

We target middle school because it is the moment when many students begin owning or regularly using personal devices (e.g., iPhones, tablets, laptops), which means they already search, click, and interact with data daily. Meeting them at this stage lets us turn informal scrolling into intentional inquiry—asking good questions, collecting or considering data responsibly, and making first plots they can explain. Research also shows that data literacy outcomes are falling fastest in these grades: between 2019 and 2022, 8th-grade scores on NAEP's Data Analysis, Statistics, and Probability category declined by about **10 points**, roughly a full grade level—outpacing drops in other math areas (Drozda 2023). That reality strengthens the case for a short, high-leverage sequence that foregrounds questions → data → analysis → interpretation, aligned to GAISE II's Level B expectations (Bargagliotti et al. 2020).

Why "R for Everyone"?

R can feel like it is "for experts," but **learnr** tutorials let us hide setup complexity and foreground ideas: students run code cells, answer quick checks, and receive immediate feedback without installing anything. Paired with **bookdown**, students publish a one-page data story with a titled plot and a short takeaway—practicing not just making graphics, but communicating responsibly. The stack is lightweight, web-first, and flexible for device-constrained classrooms (demo/projector mode and paper exit tickets).

Design goals

This capstone pilots a short, four-lesson sequence that a teacher can fit into a math or science block over 2–3 weeks. Each lesson pairs one GAISE component with a concrete product:

- **L1 (Formulate):** draft two investigable questions with teacher feedback.
- **L2 (Collect/Consider):** run a tiny class survey; export to CSV; write a 2–3 sentence "data diary" about data types and bias.
- **L3 (Analyze):** make one labeled plot (bar/hist/scatter) and describe a visible pattern or variability.
- **L4 (Interpret):** produce a mini data story (title + plot + 1–2 sentence takeaway + one limitation).

Learning aims

After four lessons, students will be able to: (1) distinguish investigable vs. vague questions and anticipate variability; (2) name common data types and note at least one bias/limitation when collecting or reusing data; (3) choose and produce an appropriate first plot and describe a visible pattern; (4) write a brief interpretation that acknowledges uncertainty and at least one limitation—aligning to GAISE II's vision for statistical literacy (Bargagliotti et al. 2020).

Pedagogical Framework

GAISE II as the spine. The sequence is organized around the four components—(I) formulate statistical investigative questions, (II) collect/consider data, (III) analyze data, and (IV) interpret results (Bargagliotti et al. 2020). Level B provides the primary target, with Level A scaffolds as needed and optional extension prompts toward Level C for advanced classes. Technology is treated as integral rather than decorative: students use software to explore variability and patterns and to simulate simple ideas when appropriate.

Developmentally appropriate ramp.
- *Lesson 1* foregrounds questioning and variability with low-code interactions.
- *Lesson 2* introduces data types, sources, and bias; students run `read_csv()`, `head()`, `nrow()`, and `ncol()` on a classroom dataset and complete a brief "data diary."
- *Lesson 3* layers descriptive displays via `ggplot2` with guided cells and a "pick-the-best-graph" check.
- *Lesson 4* shifts to interpretation and audience: titles, labels, and a short, honest takeaway that calls out at least one limitation and one potential misread.

Embedded ethics micro-prompts. Each lesson includes a short reflection, e.g.: L2—"What bias might our snack survey introduce, and how could we reduce it?"; L3—"If you tried many plots and reported only the 'best,' what's the risk?"; L4—"Note one limitation of your data and one way a reader could misread your graph."

Project Objectives & Research Questions

- **Design Objective:** Produce a four-lesson, interactive micro-curriculum (learnr/bookdown) aligned to GAISE II.

- **Pilot Objective:** Evaluate usability, engagement, and short-term learning in a middle-school setting.

Research Questions
1. Do students improve at posing investigable questions and selecting appropriate displays?
2. Can students articulate variability and limitations after activities?
3. Does light-touch ethics embedding change how students talk about data choices?

Curriculum Map (Overview)

- **Lesson 1 — Formulate Statistical Investigative Questions**
  Distinguish investigable vs. vague; anticipate variability; produce two investigable questions.
- **Lesson 2 — Collect / Consider Data**
  Data types, sources, and bias; tiny class survey → CSV → `readr`; "data diary."
- **Lesson 3 — Analyze Data**
  Descriptives and first `ggplot2`: bar, histogram, scatter; caption about a visible pattern/variability.
- **Lesson 4 — Interpret & Communicate**
  Titles/labels, `ggsave()`, mini R Markdown; peer feedback; one-page data story.

Technology & Materials

- **Stack: learnr** for directed practice; **bookdown** for a shareable webbook.
- **Access:** Web-first design; projector-friendly demos; printable exit tickets if devices are scarce.
- **Datasets:**Classroom micro-surveys and small public datasets (sports or school-relevant).

Pilot Study Design (Middle School)

- **Setting & Participants:** One partner middle school (two 6th–8th grade classes).
- **Design:** Quasi-experimental pre/post within-subjects across 2–3 weeks (4 lessons).
- **Measures (aligned to GAISE):**
  - Pre/Post concept inventory: investigable questions, displays, and variability (conceptual, not computational).
  - Performance tasks: lesson artifacts scored with rubrics keyed to the four components.
  - Ethics reflection check-ins: 2–3 sentences per micro-prompt.
  - Usability & engagement: short student survey + teacher interview.
- **Equity & Access:** If devices are limited, run projector demos; rotate small groups; use paper exit tickets.

Pilot Study Design (Alternate: Expert Review)

- **Setting & Participants:** 4–8 middle-school teachers and/or math-ed faculty who review a 1–2 page outline and a published Lesson 1 asynchronously (10-minute Google Form rubric).
- **Design:** Single-round expert validation in lieu of a student pilot; reviewers may choose a 2–3 minute Quick Pass (feasibility + willingness to try) or the full 8–10 minute rubric; responses are anonymous unless consent to quote is given.
- **Measures** (aligned to GAISE):

  - Likert (1–5): task clarity; cognitive load; age fit (6–8); guidance toward appropriate graph choice (sets up Lesson 3); variability/limitations talk (sets up RQ2); ethics micro-prompt length/tone; one-period feasibility; logistics/equity (limited devices); overall willingness to try.
  - Open prompts: "Where will students stumble?" "What would you cut or simplify for a single period?" "How would you adapt if devices are limited?" "Any accessibility/language supports needed?" (plus optional consent to quote anonymously).
  - Usability & engagement: quick overall rating and "would you pilot this next term?" checkbox.

- **Equity & Access**: Explicitly solicit device-scarcity workarounds (projector/rotation/paper exit tickets), timing adjustments for shorter periods, and suggestions for multilingual and accessibility supports; incorporate edits into teacher notes and student materials.

Bibliography

Bargagliotti, Anna, Christine Franklin, Pip Arnold, Rob Gould, Sheri Johnson, Leticia Perez, and Denise A. Spangler. 2020. *Pre-k–12 Guidelines for Assessment and Instruction in Statistics Education II (GAISE II): A Framework for Statistics and Data Science Education.* Alexandria, VA: American Statistical Association. https://www.amstat.org/docs/default-source/amstat-documents/gaiseiiprek-12_full.pdf.

Drozda, Zarek. 2023. "Data Science Is Vital to Student Success. So Why Are Outcomes Going Down?" Data Science 4 Everyone. https://d2c47c0f-cda1-420e-ae77-f1279b4ac7ea.usrfiles.com/ugd/d2c47c_067273b2bef041c7ae08cab6c7a3be8c.pdf.

Education, Stanford Graduate School of. 2021. "How to Teach Data Science in k–12 Schools? Stanford-Led Team Launches 'Big Ideas'." October 13, 2021. https://ed.stanford.edu/news/how-teach-data-science-k-12-schools-stanford-led-team-launches-big-ideas.

FiveThirtyEight. 2023. "FiveThirtyEight Data Repository." 2023. https://github.com/fivethirtyeight/data.

qmd-lab. 2025. "Closeread: A Quarto Format for Scrollytelling." 2025. https://closeread.dev/guide/.

Schloerke, Barret, Garrick Aden-Buie, RStudio, and PBC Posit Software. 2025. *Learnr: Interactive Tutorials for r.* https://rstudio.github.io/learnr/.

Xie, Yihui. 2024. *Bookdown: Authoring Books and Technical Documents with r Markdown.* https://bookdown.org/yihui/bookdown/.

YouCubed. n.d. "Data Science k–10 Big Ideas." https://www.youcubed.org/data-big-ideas/.