R for Everyone: Piloting an Interactive, Ethics-Aware Data Science Micro-Curriculum for Middle School

Introduction & Motivation

- Why middle school? Data literacy is now foundational; students should reason statistically and interrogate data claims.
- Project goal: design and pilot an interactive mini-sequence that moves students through **question → data → analysis → interpretation** with modern tools.

Pedagogical Framework

- **GAISE II Core:** adopt the four-component statistical problem-solving process—(I) formulate statistical investigative questions, (II) collect/consider data, (III) analyze data, (IV) interpret results. My goal is to use this as the spine of each lesson and of assessment.
- **Developmental Levels:** target **Level B** skills typical of middle school, while scaffolding from Level A if students lack prior exposure; Level C remains a future trajectory.
- **Technology Principles:** integrate accessible, web-based tools

Project Objectives & Research Questions

- **Design Objective:** produce a four-lesson, interactive learnr/bookdown micro-curriculum aligned to GAISE II.
- **Pilot Objective:** evaluate usability, engagement, and short-term learning gains in a middle-school setting.
- **Research Questions:**
    1. Do students improve at posing investigable questions and selecting appropriate displays?
    2. Can students articulate variability/limitations after activities?
    3. Does light-touch ethics embedding affect how students talk about data choices?

6) Curriculum Map

For each lesson: state Big Idea and Learning Objectives tied to each to a GAISE component.

Lesson 1 — Formulate Statistical Investigative Questions
GAISE focus: (I) Formulate. Distinguish investigable vs. vague; anticipate variability. Short RStudio on-ramp (comments only). learnr drag-and-drop; scrollytelling contrasts; output is 2 investigable questions with teacher feedback.

Lesson 2 — Collect / Consider Data
GAISE focus: (II) Collect/Consider. Data types, sources, and bias; small class survey → export CSV → `read_csv()`, `head()`, `nrow()`, `ncol()`. "Is this data?" quiz; FiveThirtyEight example. Data diary submission.

Lesson 3 — Analyze Data
GAISE focus: (III) Analyze. Descriptives and first ggplot: bar charts, basic hist/scatter; variability via sports examples. learnr coding cells and a "pick the best graph" check. Product: one labeled viz + caption noting a pattern/variability.

Lesson 4 — Interpret & Communicate
GAISE focus: (IV) Interpret. Titles/labels (labs()), ggsave(), mini R Markdown in bookdown; peer feedback; examples from data journalism. Final: **one-page mini data story** (1 plot + 1–2 sentences + title).

Ethics Micro-Prompts (embedded across lessons)

- L2: "What biases might our snack survey introduce? How could we reduce them?"
- L3: "If you tried many plots/statistics and reported only the 'best,' what's the risk?"
- L4: "Note one limitation of your data and one way to misread your graph."
  (Framed as brief discussion/written reflections per embedded-ethics guidance.)

Technology & Materials

- **Stack:** RStudio, **learnr** tutorials for directed practice; **bookdown** for a polished student-facing webbook; lightweight datasets. This aligns with GAISE's emphasis on simulation, software, and letting students see variability via technology; choose tools mindful of local constraints.
- **Access considerations:** web-first design; offline printable handouts if needed.

Pilot Study Design (Middle School)

- **Setting & Participants:** one partner middle school (two 6th–8th grade classes).
- **Design:** quasi-experimental **pre/post within-subjects** over 2–3 weeks (4 lessons).
- **Measures (aligned to GAISE):**
  - *Pre/Post Concept Inventory* on Investigable Questions, Displays, & Variability (conceptual, not computational).
  - *Performance Tasks:* Lesson-specific artifacts (questions, data diary, viz+caption, data story) scored with rubrics keyed to the four components.
  - *Ethics Reflection Check-ins:* 2–3 sentence responses to micro-prompts.
  - *Usability & Engagement:* short student survey + teacher interview.
- **Procedures:** teacher brief; deliver lessons that could fit in well with regular math blocks; capture platform analytics (e.g., learnr quiz correctness/tries).
- **Equity & Access:** if devices are limited, run demos on projector; rotate small groups; provide paper exit tickets.