R for Everyone: Piloting an Interactive, Ethics-Aware Data Science Mini-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?

Curriculum Map

For each lesson: Big Idea and Learning Objectives tied to GAISE are stated. Embedded Ethics Micro-Prompts in each Lesson

Lesson 1: Formulate Statistical Investigative Questions

- *Graded in-platform:*
    - Drag-and-drop: "statistical vs. non-statistical" (auto-graded).
    - Short item: pick the better investigable question (MCQ).
- *Offline reflection (no input collected):* "Write down two investigable questions you're curious about this week."
- *Exit ticket (auto):* 1 MCQ distinguishing vague vs. investigable.

Lesson 2: Collect / Consider Data

- *Graded in-platform:*
    - Import a tiny CSV; pass checks for `read_csv()`, `head()`, `nrow()`/`ncol()`.
    - "Is this data?" image/audio quiz (auto-graded).
- *Offline reflection (no input collected):* "What biases might our snack survey introduce? How could we reduce them?"
- *Exit ticket (auto):* Identify a potential bias in a short scenario (MCQ).

Lesson 3: Analyze Data

- *Graded in-platform:*
    - Create one plot; auto-checks for title present, axes labeled, appropriate geom for data type, and basic caption keyword (e.g., "higher/lower").
    - "Which graph best answers this question?" (MCQ).
- *Offline reflection (no input collected):* "What surprised you about the distribution or pattern you saw?"// "If you tried many plots/statistics and reported only the 'best,' what's the risk?"
- *Exit ticket (auto):* Choose the correct display for a given question (MCQ).

Lesson 4: Interpret & Communicate

- *Graded in-platform:*
    - Add labels with `labs()`; pass a simple `ggsave()` check.
    - Identify an overclaim vs. fair interpretation (MCQ).
- *Offline reflection (no input collected):* "State one limitation of your data and one way a viewer might misread your graph."
- *Exit ticket (auto):* Pick the fair interpretation for a shown plot (MCQ).

Technology & Materials

- 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

- **Setting & Participants:** one partner middle school (two 6th–8th grade classes).
- **Design:** quasi-experimental **pre/post within-subjects** over 2–3 weeks.
- **Measures (aligned to GAISE):**
    - *Pre/Post Concept Inventory* on Investigable Questions, Displays, & Variability (conceptual, not computational).
    - *In-Lesson Auto Checks (replace rubrics):* Hidden tests for required elements (e.g., imported data, correct geom for data type, title/labels present, basic caption keyword). Pass/fail and quiz correctness/tries logged automatically.

    - *Ethics Reflection Check-ins:* 2-3 sentence responses to micro-prompts submitted to teacher.
    - *Usability & Engagement:* short student survey and teacher interview.
- **Procedures:** teacher brief; deliver lessons that could fit in well with regular math blocks; Platform **auto-captures analytics** (quiz correctness, attempts, pass/fail checks, timestamps (iffy on how I will get this to teacher)
- **Equity & Access:** if devices are limited, run demos on projector; rotate small groups; provide paper.