

## Pedagogical and Technical Framework for R Exercise Series

### 1 Purpose and Design Rationale

This [example exercise](#) was designed specifically for criminology students with no prior programming experience to introduce them to data analysis and statistical reasoning using R. The design process focused on selecting criminologically relevant scenarios, ensuring accessibility for beginners, and embedding best practices in instructional design. Each step, from data simulation to exercise scaffolding, was informed by evidence-based frameworks (e.g., Bloom's Taxonomy, PRIMM, and scaffolding theory) to promote both technical skill development and statistical thinking.

### 2 Development Process and Instructional Design

This exercise was created by mapping key criminological concepts to R programming tasks, generating synthetic datasets that reflect real-world crime data, and structuring the activity with clear instructions, worked examples, and opportunities for independent practice. The design is grounded in evidence-based frameworks:

- **Bloom's Taxonomy** (Krathwohl, 2002) is a hierarchical model that classifies learning objectives by cognitive complexity, from basic recall to higher-order skills like analysis and creation. This ensures students progress from foundational knowledge to independent problem-solving.
- **PRIMM** (Sentance et al., 2019) is a phased approach to programming instruction: Predict, Run, Investigate, Modify, and Make. It guides students from understanding existing code to creating their own solutions.
- **Scaffolding** (Wood et al., 1976) refers to the instructional supports provided to students as they learn new concepts. These supports are gradually removed as students gain confidence and competence, fostering independence.

Because reporting is just as important as analysis, I plan to use R to help students structure their findings clearly and correctly. Tools such as *apaTables* and *flextable* allow us to format results in APA style directly from code, making the reporting process more transparent and reproducible. I intend to use these tools to teach students how to interpret test statistics and report their results according to academic standards—preparing them for thesis writing and publication.

Automated feedback and metadata are included for instructional tracking and continuous improvement. This template can be adapted for new topics, datasets, or skill levels by modifying the context, adjusting scaffolding, or integrating advanced R techniques.

### 3 Exercise Progression

	Exercise	Bloom Level	Scaffolding	PRIMM Phase
1	Data structure	Understand	Full support	Run
2	Creating variables with appropriate types	Apply	Full support	Modify
3	Identifying measurement levels	Understand & Apply	Full support	Investigate
4	Assign number of burglaries	Remember	Full support	Run
5	Load crime dataset	Apply	Full support	Modify
6	Calculate average offenses	Apply	Full support	Run
7	Generate frequency tables	Apply	Full support	Run, Modify
8	Describe offender age	Apply	Full support	Run
9	Filter violent crimes	Apply	Partial support	Modify
10	Crime rate calculation	Apply	Partial support	Modify
11	Crime rates national average	Apply & Analyze	Partial support	Modify
12	Scatterplot unemployment	Analyze	Partial support	Modify
13	Filtering crime data with dplyr	Apply	Partial support	Modify
14	Selecting columns with dplyr	Apply	Partial support	Modify
15	Creating new columns with mutate	Create	Minimal support	Make

### 4 Reproducibility and Synthetic Data

All exercises use synthetic datasets that simulate real criminological data (e.g., district-level crime rates, demographics, police response times). These datasets are generated using the [crimsyndata package](#), which ensures privacy, reproducibility, and ethical compliance. The package allows instructors to tailor datasets to specific teaching goals and updates them as needed. Students learn APA-style reporting using packages like [apaTables](#), [sjPlot](#), or [papaja](#), focusing on interpretation rather than manual formatting. This approach enables safe sharing and publication of results and supports differentiated instruction as the course evolves.

## 5 Using crimsyndata in R

The crimsyndata package is available on GitHub.

This package provides carefully designed synthetic datasets that simulate real criminological data patterns while ensuring privacy and ethical compliance. Students can use these datasets for practice without concerns about confidentiality or data protection requirements.

To use the [crimsyndata package](#):

1. First install the package from GitHub:
  - Install the remotes package if not already available
  - Use `remotes::install_github("KKural/crimsyndata")` to install from the repository
2. Load the package with `library(crimsyndata)`
3. Access a variety of criminologically relevant datasets such as:
  - **crime\_stats\_ghent**: Multi-year crime data including demographic variables by district, crime types, and population statistics
  - **fear\_of\_crime\_survey**: Individual-level survey responses on perceived safety with demographic information
  - **neighborhood\_index**: Contextual variables for neighbourhoods including socioeconomic indicators
  - **police\_effort\_index**: Law enforcement resource allocation and effectiveness metrics by district

Each dataset is designed to exhibit realistic criminological patterns while being completely synthetic, supporting teaching concepts such as crime rate calculation, demographic analysis, and spatial patterns. An important feature of the [crimsyndata package](#) is its flexibility – new datasets can be generated to match specific learning objectives or course requirements. Additional datasets can be created based on the lesson goals, statistical concepts, or analytical techniques in the curriculum. This ensures that examples are always relevant to the specific criminological concepts being taught.

## 6 Platform Integration

A multi-platform approach ensures students have access to resources, feedback, and collaborative tools throughout the course:

- **Dodona**: Delivers auto-graded exercises, instant feedback, and a structured learning path. Students can submit code, receive hints, and track their progress, fostering self-paced learning and accountability.
- **Ufora**: Integrates quizzes, hosts supplementary resources, and provides a space for announcements and discussions. Ufora supports blended learning by connecting in-class and online activities.

- **GitHub:** Serves as a repository for datasets, sample scripts, and templates. Students can explore real code, contribute to open resources, and access extension exercises for further practice and exploration.

## 7 Onboarding: R & Dodona

To ensure all students start with a solid foundation, the course begins with a short onboarding quiz. This quiz introduces basic R syntax, variable assignment, and the Dodona environment. It is designed as a low-stakes, formative assessment to help students identify gaps in their understanding and become comfortable with submitting code and interpreting feedback before tackling the main exercises.

## 8 Interactive Learning with swirl

swirl offers a complementary approach to Dodona, providing interactive R tutorials that run directly within the R console. I've developed a [prototype swirl course specifically for criminology students](#).

### 8.1 The Bachelor Criminology swirl Course Prototype

This initial prototype demonstrates how swirl can address common challenges criminology students face when learning statistics and coding by:

1. Using the same synthetic crime datasets as the main exercises for consistency
2. Providing immediate, contextual feedback on code and answers
3. Offering step-by-step guidance through data exploration and analysis
4. Creating a safe environment for experimentation without fear of mistakes

### 8.2 Current Prototype Lessons:

- **Lesson 1: Basics of R** - Introduces fundamental R functions like `str()`, `colnames()`, `head()`, and basic data exploration using criminological datasets
- **Lesson 2: T Test and Chi Square** - Guides students through conducting and interpreting these essential statistical tests with crime data examples

### 8.3 Installation & Usage:

- Students install the [swirl package](#) and download the course with `swirl::install_course_github("KKKural", "Data_Analysis_Bachelor_Criminology")`
- Starting with `swirl()` in the console, they select the criminology course
- Each lesson provides context-specific hints and error correction

This prototype demonstrates the concept and can be expanded into a comprehensive curriculum based on specific course requirements. Additional lessons can be developed to cover:

- More advanced data manipulation techniques for crime data
- Statistical models commonly used in criminology (regression, ANOVA)
- Specialized visualization approaches for spatial crime patterns

- Customized modules that align with specific course topics or departmental needs

The interactive console-based format complements the Dodona platform, giving students multiple paths to mastery while addressing statistical anxiety. The prototype has already shown that students appreciate the immediate feedback and self-paced structure, which builds confidence before they tackle more complex, independent assignments.

## 9 Metadata for Instructional Tracking

Each exercise includes metadata in the form of comments (Bloom Level, Scaffolding Level, PRIMM Phase, [see here for example](#)). This metadata enables instructors to track the alignment of exercises with pedagogical goals, monitor the progression of cognitive demand, and ensure a balanced mix of support and independence. While invisible to students during grading, this information is valuable for continuous course improvement and instructional research.

## 10 Student Support & Continuous Improvement

Student questions and feedback are actively encouraged throughout the course. The teaching team monitors queries submitted via Dodona and Ufora, providing timely responses and clarifications. Frequently asked questions and common challenges are tracked and used to refine exercise instructions, update datasets, and improve the overall learning experience. This iterative process ensures the course remains clear, effective, and responsive to student needs.

## 11 Supporting Software Transitions in Statistical Education

With proficiency in both SPSS and R, I am well-positioned to support the department's software transition initiatives. My dual expertise allows me to understand both the technical challenges students face when learning a new statistical environment and the conceptual connections between familiar operations and new syntax.

By leveraging my experience with both platforms, I can create a smoother learning path for students and help reduce technical barriers to statistical learning, allowing them to focus on criminological applications rather than software mechanics.

## 12 References

- Krathwohl, D. R. (2002). A Revision of Bloom's Taxonomy: An Overview. *Theory Into Practice*, 41(4), 212–218. [https://doi.org/10.1207/s15430421tip4104\\_2](https://doi.org/10.1207/s15430421tip4104_2)
- Sentance, S., Waite, J., & Kallia, M. (2019). Teaching computer programming with PRIMM: A sociocultural perspective. *Computer Science Education*, 29(2–3), 136–176. <https://doi.org/10.1080/08993408.2019.1608781>
- Wood, D., Bruner, J. S., & Ross, G. (1976). The Role of Tutoring in Problem Solving. *Journal of Child Psychology and Psychiatry*, 17(2), 89–100. <https://doi.org/10.1111/j.1469-7610.1976.tb00381.x>