DOCUMENTATION OF THE DGH EXAM

ALHAGIE LAMIN DRAMMEH

MATNUMBER: 2409243

This dataset contains variables that help analyze student performance and behavior, including categorical variables like sex, employed, and student (useful for grouping data by gender, employment status, or enrollment type), and numerical variables like age, sleep, courses, studying, and gpa (ideal for plotting relationships). Scatter plots can compare study hours (studying) against GPA to see if more studying improves grades, or sleep hours (sleep) against GPA to check if rest impacts academic success. Adding color-coding by employment status or gender can reveal subgroup trends, such as whether working students perform differently than non-working ones.

For deeper insights, trend lines can highlight correlations (e.g., does GPA rise with more sleep?), while faceting by gender or enrollment type can compare patterns across groups. For example, plotting courses vs. GPA while coloring by employment status could show if employed students struggle with heavier course loads. These visualizations help educators and policymakers identify factors influencing student success, such as whether part-time work harms grades or if sleep deprivation lowers performance. By mixing variables creatively, scatter plots turn raw data into actionable insights.

**Visualization Context and Implementation**  
This Shiny application analyzes relationships between student academic performance and lifestyle factors through interactive scatter plots. It helps identify correlations between study habits, sleep patterns, and GPA while accounting for demographic variables like gender and employment status. Designed for education researchers, academic advisors, and administrators, the visualization features customizable axes, color-coded categorical variables, dynamic point sizing, and a clean minimalist interface. The technical implementation uses a single app.R file containing three key components: data preparation (importing and cleaning the Excel dataset), a fluidPage UI with sidebar controls, and server logic generating reactive visualizations.

**Data and Repository Structure**  
The analysis uses student\_data.xlsx (originally DGH.xlsx) containing academic (GPA, courses, studying), demographic (age, sex), and lifestyle (sleep, employment status) variables. The project follows a standardized repository structure with app.R for the main application code, the dataset file, renv.lock for dependency management, and README.md for documentation. Both GitHub repository and Shiny application ( <https://alhagielamindrammeh.shinyapps.io/EXAM3594/>) links are provided for easy access and deployment.

**Dependency Management**  
The project utilizes renv for reproducible environment management, with dependencies including shiny (v1.7.4), readxl (v1.4.1), ggplot2 (v3.4.0), and dplyr (v1.1.0) as recorded in the renv.lock file. Developers can initialize the environment using renv::init() and restore exact package versions with renv::restore(). Key deployment considerations include using relative file paths, standardized dataset naming (student\_data.xlsx), and compatibility testing with R 4.2.1.

**Deployment and Reproducibility**  
The documentation ensures seamless deployment by maintaining all file paths relative to the project root and providing complete version specifications. The renv.lock file serves as a critical component for environment reproducibility, allowing collaborators to exactly recreate the development setup. This approach guarantees consistent results across different systems while maintaining accessibility for both technical and non-technical stakeholders involved in educational data analysis.