DIGITAL HEALTH

Module GPH-01: Digital Health

Responsible: Prof. Dr. Dominik Böhler

Deggendorf Institute of Technology

Faculty European Campus Rottal-Inn

Master of Global Public Health, 1st Semester

Sanjana Yadav

Pfarrkirchen, 30.01.2025

1. Introduction

This report outlines the steps taken to clean, analyze, and visualize Glomerular Filtration Rate (GFR) data using R and Shiny. The process involved importing messy data, cleaning and preparing it for analysis, and creating interactive visualizations using a Shiny app. The goal was to make the data more understandable and accessible through dynamic and customizable visualizations.

2. User Persona

The key user persona for this interactive GFR data visualization is a medical researcher or healthcare analyst specializing in nephrology or kidney disease. This user is interested in analyzing GFR trends over multiple days to track kidney function and assess the impact of different conditions or treatments.

3. Steps for Importing and Cleaning Data in RStudio

1. Installing and Loading Packages:
   The necessary R packages were installed and loaded using the "pacman" package manager to ensure all required libraries were available.

2. Importing the Dataset:
   The dataset "Sanjana messydata_.xlsx" was imported using "readxl::read_excel()". This dataset contains patient information, GFR values across multiple days, and other clinical variables.

3. Cleaning and Preprocessing the Data:
   - Renaming Columns: Column names were renamed for clarity.
   - Removing Unnecessary Columns: The 5th column, which was empty, was removed.
   - Standardizing Data Formats: Categorical values were converted to standardized formats to ensure consistency.
   - Reshaping the Data: The dataset was converted from wide to long format using "pivot_longer()" to facilitate better visualization and comparison of GFR values across different days.

4. Visualizing the Cleaned Data

Once the data was cleaned, various plots were created using "ggplot2" to analyze GFR trends:

i. Boxplot of GFR Across Days: This boxplot displays the distribution of GFR values across different days, helping to understand how GFR fluctuates over time.
ii. Violin Plot of GFR Distribution by Day: This plot combines a boxplot with a smoothed density plot, providing insight into the spread and concentration of GFR values for each day.
iii. Scatter Plot of Day 1 GFR vs. Day 5 GFR: This scatter plot compares GFR values on Day 1 and Day 5, allowing for the identification of trends over time for different racial groups.
iv. Histogram of Day 1 GFR: This histogram shows the frequency distribution of Day 1 GFR values, providing insights into whether the values are normally distributed, skewed, or have outliers.
v. Bar Chart of Diagnosis Counts: This bar chart shows the number of patients for each diagnosis category, helping to identify the most or least common diagnoses in the dataset.

vi. Facet Grid for GFR by Race: This plot separates GFR trends by race, allowing for easy comparison of how GFR values differ across racial groups.

5. Producing an Interactive Data Visualization Using Shiny

Using Shiny, an interactive data visualization app was developed to allow users to explore GFR data dynamically. The app integrates multiple visualization types:

1. Data Import & Cleaning:
The script reads an Excel file containing GFR values for multiple days and cleans the data by renaming columns, handling missing values, and restructuring it into a tidy format.

2. Interactive UI Components:
- Users can select columns dynamically using “selectInput()”.
- “colourInput()” allows users to customize colors for histograms and scatter plots.
- The UI includes a fluid page layout with a title panel, sidebar layout for user inputs, and a main panel for displaying visualizations.

3. Server Logic Implementation:
- Data loading with error handling was set up using “readxl”.
- Reactive UI elements for column selection were created based on the loaded data.
- Four types of visualizations were implemented: histogram, scatter plot, boxplot, and a data table.

4. Visualizations Implemented:
- Histogram: Shows the distribution of values in a selected column.
- Scatter Plot: Helps identify relationships between two GFR values.
- Boxplot: Highlights outliers and compares GFR trends across days.
- Data Table: Displays the dataset in an interactive format for further exploration.

[Note: #Scenarios Where the Histogram shows error :
Non-Numeric Variables: If the selected column (input$selected_column_x) is not numeric (e.g., categorical or text data), geom_histogram() will fail because it requires numeric input for the X-axis. Example: If input$selected_column_x is a column like "Gender" or "Category," which contains strings or factors, an error will occur]

5. Interactivity and Customization:
- Color pickers and sliders allow users to customize visualizations.
- Dynamic dropdowns enable column selection.
- Error handling and user feedback ensure a smooth user experience.

6. Conclusion

The interactive Shiny app enhances the workflow by enabling real-time user interaction with the data. The app supports medical decision-making by highlighting patterns in GFR changes and identifying variations in kidney function across different patient groups.

Link to the Shiny App:

https://sanjanagph-2024.shinyapps.io/digital_health/