Concept and Purpose of Our Shiny Application: U.S. Myeloma Trends Dashboard (Deliverable 1)

URL: US_Myeloma_Trends_Dashboard
*This Shiny App was designed by collaborators Anne-Marie Feeney and Stephanie Leininger-Anderson.

This report presents an extensive epidemiological analysis of multiple myeloma incidence trends across the United States from 1999 to 2021, focusing on both age group and state-level variations. Drawing on extensive surveillance data, the report provides statistical summaries, visualizations, and formal hypothesis testing to assess disparities and trends in disease burden.

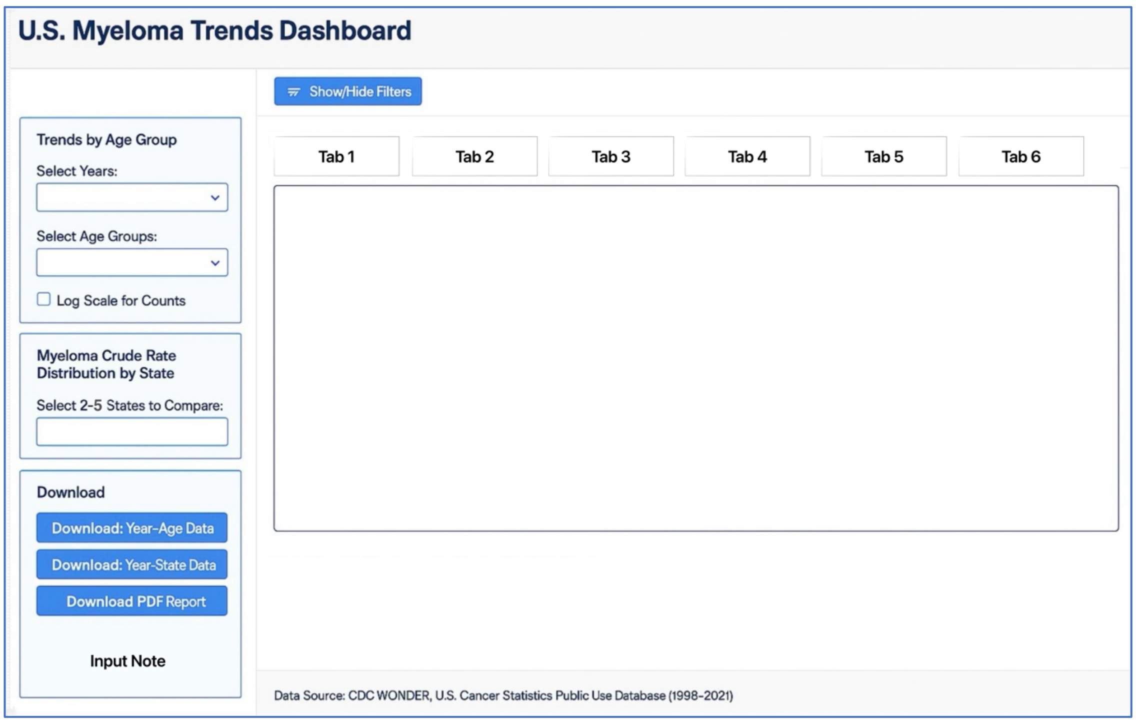
This information serves as both an analytical summary and a reproducible reference for public health professionals, epidemiologists, and policymakers seeking to understand and address disparities in multiple myeloma incidence.

Motivation and Plan for Creating the App (Deliverable 2)

The Shiny App we created answers two research questions:

1. How have myeloma rates changed over time by age group?
2. Are there regional differences in myeloma crude rates?

Below is the wireframe rendering we used to design the app:



Datasets used and its relevance to the course. (Deliverable 3)

Two datasets are used in the analysis:
1. U.S. Myeloma Incidence Grouped by Year and Age Group (URL: myeloma_data_age_year.xlsx)
2. U.S. Myeloma Incidence Grouped by Year and State (URL: myeloma_data_year_state.xlsx)

Relevance of the U.S. Myeloma Trends Dashboard to BHDS2010: Statistical Programming for Health Data Science

The U.S. Myeloma Trends Dashboard, developed using R and Shiny, serves as a comprehensive, interactive tool to explore trends in myeloma incidence across the United States. It draws from real-world public health datasets, focusing on age group, state, and year-wise trends. This application demonstrates practical competencies acquired in the BHDS2010 Spring25 S01: Statistical Programming for Health Data Science course and represents an integration of multiple course themes including data wrangling, statistical analysis, data visualization, and reproducible research.

Application of Statistical Programming Concepts

A core objective of BHDS2010 is to enable students to work with health data in R, from preprocessing to analysis and reporting. The Myeloma Trends Dashboard aligns closely with this goal:

- **Data Wrangling**: The app handles large datasets requiring transformation and cleaning. Techniques such as reshaping data and joining multiple data sources, and handling missing values are central to the app, mirroring skills taught through dplyr, tidyr, and base R.
- **Statistical Analysis**: The dashboard includes summary statistics and inferential procedures such as Levene's test for homogeneity of variances and ANOVA to detect differences across groups. These reflect key course content on hypothesis testing and modeling health data.
- **Visualization**: Using ggplot2 and plotly, the dashboard presents data through interactive charts (line plots, histograms, box plots, heatmaps) which enhance comprehension and user engagement. The application of advanced plotting techniques reinforces classroom lessons on visualizing statistical information effectively.

Interactive Dashboard Development Using Shiny

The Shiny framework forms a major part of the course's applied programming focus. The dashboard utilizes reactive programming, enabling dynamic updates to charts and tables based on user selections (e.g., state, year, or age group). Key Shiny components such as reactive(), observeEvent(), and dynamic UI elements like filters and sliders are employed extensively.

Additionally, the use of modular code, custom themes, and download handlers (for CSV and PDF generation) showcase an advanced grasp of Shiny development principles, which are encouraged in BHDS2010 for building scalable and professional-grade applications.

Reproducible and Transparent Reporting

A distinguishing feature of the app is the ability to generate a PDF report using RMarkdown, incorporating real-time outputs from the dashboard. This aligns with the course's emphasis on reproducible workflows, promoting clarity, transparency, and replicability in statistical analysis. The integration of rmarkdown::render() and use of parameterized reports demonstrate best practices in reporting health data findings.

Public Health Context and Relevance

The dashboard's use of real epidemiological data related to multiple myeloma directly connects to the health-focused context of the course. By enabling exploration of cancer incidence trends across regions and time, the app serves as a meaningful example of how statistical tools can inform health policy, surveillance, and resource planning. It reinforces the course's aim to bridge technical statistical programming with public health applications.

Summary

The U.S. Myeloma Trends Dashboard is a strong example of the integration of programming, statistics, and health data science in a real-world context. It not only reinforces the technical skills taught in BHDS2010, such as R programming, data visualization, and statistical modeling, but also exemplifies the course's broader goals: to enable students to apply these skills in a meaningful, health-driven environment. The dashboard serves as both a learning artifact and a potential tool for public health insights, embodying the spirit of statistical programming for health data science.

Key Features and User Interactivity (Deliverable 4)

The U.S. Myeloma Trends Dashboard is an interactive Shiny application designed to explore age, year, and state-level patterns in myeloma incidence across the United States. It emphasizes data accessibility, dynamic visualization, and customizable analytical views to support public health insights and research.

Inputs

Users interact with the dashboard through several intuitive input controls that tailor the visualizations and analyses:

- **Year Selector (years)**: Multi-select input for filtering data by specific years.
- **Age Group Selector (age_groups)**: Multi-select input to filter data by standardized age groups.
- **State Selector (state)**: Multi-select input allowing users to focus on one or more U.S. states.

- **Log Scale Toggle (log_scale)**: Checkbox input that allows users to apply a logarithmic transformation to the y-axis of applicable plots for better visualization of skewed data.
- **Histogram View Type (histogram_view)**: Radio buttons for choosing between:
    - **Overlay View**: All age groups in one plot
    - **Faceted View**: Separate plots by age group
- **State Comparison Picker (selected_states_box)**: Picker input enabling focused comparison between selected states in the boxplot output.
- **Sidebar Toggle (toggleSidebar)**: UI element for collapsing or expanding the control panel, improving focus on the data visualizations.

Outputs

The application generates multiple dynamic and interactive outputs in response to user inputs:

- **Line Chart by Age Group (line_age)**: Displays temporal trends in myeloma incidence across age groups.
- **Line Chart by Year (line_year)**: Aggregated trends in incidence by year for selected filters.
- **Histogram (hist_plot)**: Distribution of crude rates shown either overlaid or faceted by age group.
- **Boxplot by Age Group (box_plot)**: Visual comparison of myeloma incidence variability across age groups.
- **Summary Table (box_summary_table)**: Tabulated statistical summaries supporting the boxplot.
- **Crude Rate Heatmap (heatmap)**: A matrix-style heatmap depicting state-wise and age-wise trends.
- **Box Plot: State Comparison (state_box_plot)**: Compares distributions of crude rates among selected states.
- **Downloadable CSV Data (download_data)**: Enables users to export filtered datasets as CSV files.
- **Downloadable PDF Report (download_report)**: Generates a comprehensive report including selected plots and statistics based on user-defined filters.

Visualizations

The dashboard offers rich, interactive data visualizations including:

- Line Charts (by Age and Year)
- Overlay and Faceted Histograms
- Age Group Boxplots
- State Comparison Boxplots
- Heatmap
- Statistical Tables
- Downloadable reports and data exports

Data Sources

Visualizations and summaries are powered by two primary datasets:

- **myeloma_data_age_year.xlsx** – Incidence data stratified by age group and year
- **myeloma_data_year_state.xlsx** – Incidence data organized by year and U.S. state

Server Logic Definition (Deliverable 5)

The server component of the **U.S. Myeloma Trends Dashboard** defines how user inputs drive reactive behavior, data filtering, visualization rendering, statistical analysis, and report generation. It is the core engine that powers all interactive and analytical functionalities of the app.

Reactive Data Handling

The server logic begins by reading and preprocessing the underlying Excel datasets (myeloma_data_age_year.xlsx, myeloma_data_year_state.xlsx). These datasets are cleaned and stored as reactive objects to support dynamic filtering:

- **Reactive Subsets**: Filtered dataframes are created in response to user inputs such as selected years (input$years), age groups (input$age_groups), and states (input$state). These subsets form the basis for all plots and tables.

Reactive Visualizations

Each major plot in the dashboard is rendered reactively based on filtered data and user configuration:

- **Line Charts (output$line_age, output$line_year)**: Generated using ggplot2 or plotly, these plots display incidence trends over time stratified by age or year.
- **Histogram (output$hist_plot)**: Users can toggle between *Overlay* and *Faceted* views to examine the distribution of crude rates across age groups.
- **Boxplots (output$box_plot, output$state_box_plot)**: Boxplots summarize variability in crude rates by age or selected states. Log transformation (input$log_scale) is conditionally applied.
- **Heatmap (output$heatmap)**: A matrix-style heatmap is created to visualize spatial and demographic patterns across states and age groups.

Summary Tables & Statistics

- **output$box_summary_table**: A reactable table displays calculated statistics such as median, IQR, and counts grouped by age or state, complementing the boxplots.

Download Functionality

- **CSV Download (output$download_data)**: Enables export of the currently filtered dataset.
- **PDF Report (output$download_report)**: Dynamically generates a printable report via rmarkdown::render() that includes user-selected charts, data summaries, and comparisons. It uses the current filter inputs to tailor the report.

Interactivity & Reactivity

The app uses:

- observeEvent() to respond to UI toggles like input$toggleSidebar
- reactive() and reactiveValues() to ensure only relevant portions of the UI are recalculated
- Conditional UI logic to hide or display certain plots or tables depending on the user selections

Technical Stack

- **Visualization**: ggplot2, plotly
- **Data Tables**: reactable
- **PDF Reports**: rmarkdown, knitr
- **UI Enhancements**: shinyWidgets, shinyjs, and custom themes

Key Findings, Insights, and Analyses (Deliverable 6)

Age-Stratified Myeloma Trends

- **Progressive Increase with Age**: Myeloma incidence increases markedly with age, peaking in the 70–74 and 75–79 age brackets.
- **Statistical Summary**:
    - The mean number of cases ranges from 72.5 (30–34 years) to 3439.3 (70–74 years).
    - The standard deviation increases with age, indicating greater variability in older groups.
    - 85+ years group still has a high incidence (mean: 2017.8), underscoring continued risk in very old populations.
- **Visual Highlights**:
    - Line plots and histograms illustrate temporal consistency in the increasing trend with age.
    - Box plots reveal increasing spread and outliers in older groups.
    - A heatmap clearly visualizes the high concentration of cases in the 60+ age range.
- **Temporal and Geographic Patterns**
- **Stable Yearly Trends**:
    - Line charts suggest that while absolute myeloma counts vary slightly year-to-year, the age-dependent pattern remains consistent over time.
- **State-Level Comparisons**
    - Box plots of crude incidence rates across states show considerable variation, with selected states highlighted.
    - A subset of states was analyzed further using statistical tests.

Decision Framework for State Comparisons

The Myeloma Trends app allows users to compare crude incidence rates across 2 to 5 states:

Two-State Comparison

Test	Hypotheses	Application Conditions
Student's t-test	H_0: $\mu_1 = \mu_2$; H_1: $\mu_1 \neq \mu_2$	Normality ($p > \alpha$) & Equal variance ($p > \alpha$)
Welch's t-test	H_0: $\mu_1 = \mu_2$; H_1: $\mu_1 \neq \mu_2$	Normality ($p > \alpha$) & Unequal variance ($p \leq \alpha$)
Wilcoxon rank-sum	H_0: distributions equal; H_1: distributions differ	Non-normal ($p \leq \alpha$)

Three-to-Five-State Comparison

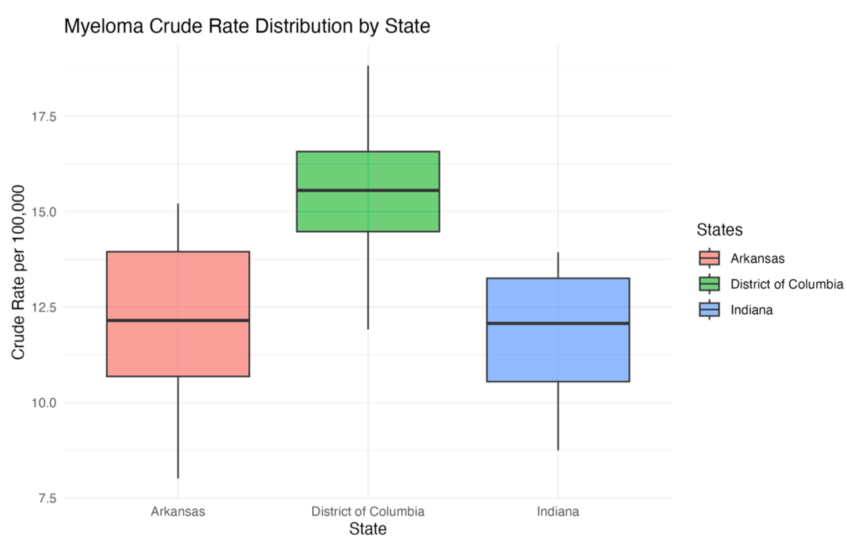
Test	Hypotheses	Application Conditions
One-way ANOVA	H_0: $\mu_1 = \ldots = \mu_k$; H_1: some μ differ	Normality ($p > \alpha$) & Equal variance ($p > \alpha$)
Kruskal–Wallis	H_0: distributions equal; H_1: at least one differs	Non-normal ($p \leq \alpha$) or Unequal variance ($p \leq \alpha$)

Tukey's HSD Post Hoc Test

After a significant one-way ANOVA ($p < \alpha$), Tukey's Honest Significant Difference (HSD) test performs pairwise comparisons between state means while controlling the family-wise error rate.

Step	Action	Interpretation
1	Calculate mean differences |mean_i – mean_j| for each pair of states	
2	Compute HSD = q_{α, k, df_error} × √(MSE/n)	
3	Compare each |mean_i – mean_j| to HSD	If > HSD → significant difference

Example Interpretation of Results



1. **Shapiro–Wilk Test (Normality):** All three states—Arkansas ($p = 0.5278$), District of Columbia ($p = 0.6644$), and Indiana ($p = 0.064$)—had p-values greater than 0.05, indicating no evidence to reject normality of their crude incidence rate distributions.

Table 4: Shapiro–Wilk Test Results by State

States	Shapiro-Wilk p-value
Arkansas	0.5278
District of Columbia	0.6644
Indiana	0.064

2. **Levene's Test (Equal Variance):** The test statistic F = 0.147 with p = 0.8638 (> 0.05) supports homogeneity of variances across the three states.

Table 5: Levene's Test Results

F.value	p.value
0.1468	0.8638

3. **One-Way ANOVA:** With F = 26.10 and p < 0.001, we reject the null hypothesis that all state means are equal. There is a statistically significant difference in mean incidence rates among at least two of the states.

Table 6: Selected Statistical Test Results

Test	F.value	p.value
One-Way ANOVA	26.0997	0 *

4. **Tukey's HSD Post-Hoc:**
   - DC vs. Arkansas: Mean difference = 3.339 (95% CI: 2.023 to 4.655), adj. p < 0.001 → significant.
   - Indiana vs. Arkansas: Mean difference = –0.237 (95% CI: –1.584 to 1.110), adj. p = 0.9066 → not significant.
   - Indiana vs. DC: Mean difference = –3.576 (95% CI: –4.923 to –2.229), adj. p < 0.001 → significant.

Tukey HSD Post-Hoc Test

Table 7: Tukey HSD Pairwise Comparisons

	Comparison	Difference	Lower CI	Upper CI	Adjusted p-value
District of Columbia-Arkansas	District of Columbia-Arkansas	3.3391908	2.022965	4.655417	0 *
Indiana-Arkansas	Indiana-Arkansas	- 0.2370176	- 1.584218	1.110183	0.9066
Indiana-District of Columbia	Indiana-District of Columbia	- 3.5762083	- 4.923409	- 2.229008	0 *

Epidemiological Interpretation

1. **Normality & Variance Assumptions**: Confirming assumptions ensures valid comparison of crude rates across states.
2. **ANOVA Finding:** There are significant geographic differences in myeloma incidence rates, suggesting regional variation in risk factors or reporting.
3. **Pairwise Differences:**
   - District of Columbia has a significantly higher crude incidence rate than Arkansas by approximately 3.3 cases per 100,000.
   - Arkansas and Indiana do not differ significantly.
   - District of Columbia's rate exceeds Indiana's by about 3.6 per 100,000.
   - These disparities may reflect differences in population demographics, healthcare access, environmental exposures, or registry completeness.

Implications

The findings underscore the importance of age as a primary risk factor and highlight geographic disparities that warrant further investigation into environmental, genetic, or healthcare access-related drivers. The significant differences in crude rates between states emphasize the need for localized intervention strategies.
The statistical decision framework also equips analysts with a repeatable method for assessing similar data, fostering better epidemiologic rigor and policy targeting.

Recommendations

- Targeted Interventions: Prioritize screening and awareness programs for adults aged 60+, especially in high-incidence states.
- Further Research: Investigate environmental or systemic contributors to state-level disparities.
- Standardize Reporting: Use the provided decision trees and statistical tests in future surveillance to ensure consistency and comparability.

Link to GitHub Repository (Deliverable 7)

git-repo: https://github.com/slanderson2025/US_Myeloma_Trends_Dashboard/tree/main