

Clinical Visualization

Daniel Olofin

2025-09-25

Introduction

This clinical visualization project explores patient-level data from a colon cancer clinical trial to uncover patterns in treatment outcomes, survival, and prognostic factors. By applying statistical visualization techniques, the project transforms complex clinical data into interpretable graphics that highlight key insights in oncology research—bridging statistical modeling with visual storytelling to support evidence-based understanding.

Aim

The aim of this project is to develop a visualization gallery that demonstrates the power of data visualization in summarizing and interpreting clinical trial data. Using the colon dataset from the `survival` R package, the project seeks to highlight how different visualization techniques can uncover patterns, communicate trial results, and support evidence-based decision-making in oncology research.

Objectives

1. Showcase diverse visualization methods: Implement at least 5 different plots, including survival curves, subgroup analyses, and exploratory graphics, to present the colon dataset from multiple perspectives.
2. Communicate clinical insights effectively: Use plots to explain treatment effects, recurrence and survival patterns, and relationships between demographic, clinical, and outcome variables.
3. Bridge statistics and storytelling: Translate statistical outputs (e.g., hazard ratios, cumulative incidence, survival probabilities) into intuitive graphics that can be understood by both technical and non-technical audiences.
4. Highlight best practices in reproducibility: Use R (`ggplot2`, `survminer`, etc.) to build reproducible visualization pipelines for clinical trial data analysis.
5. Create a portfolio-ready deliverable: Assemble the collection into a cohesive visualization gallery (markdown/Quarto notebook or GitHub repo) to serve as a demonstration of technical, statistical, and communication skills.

libraries

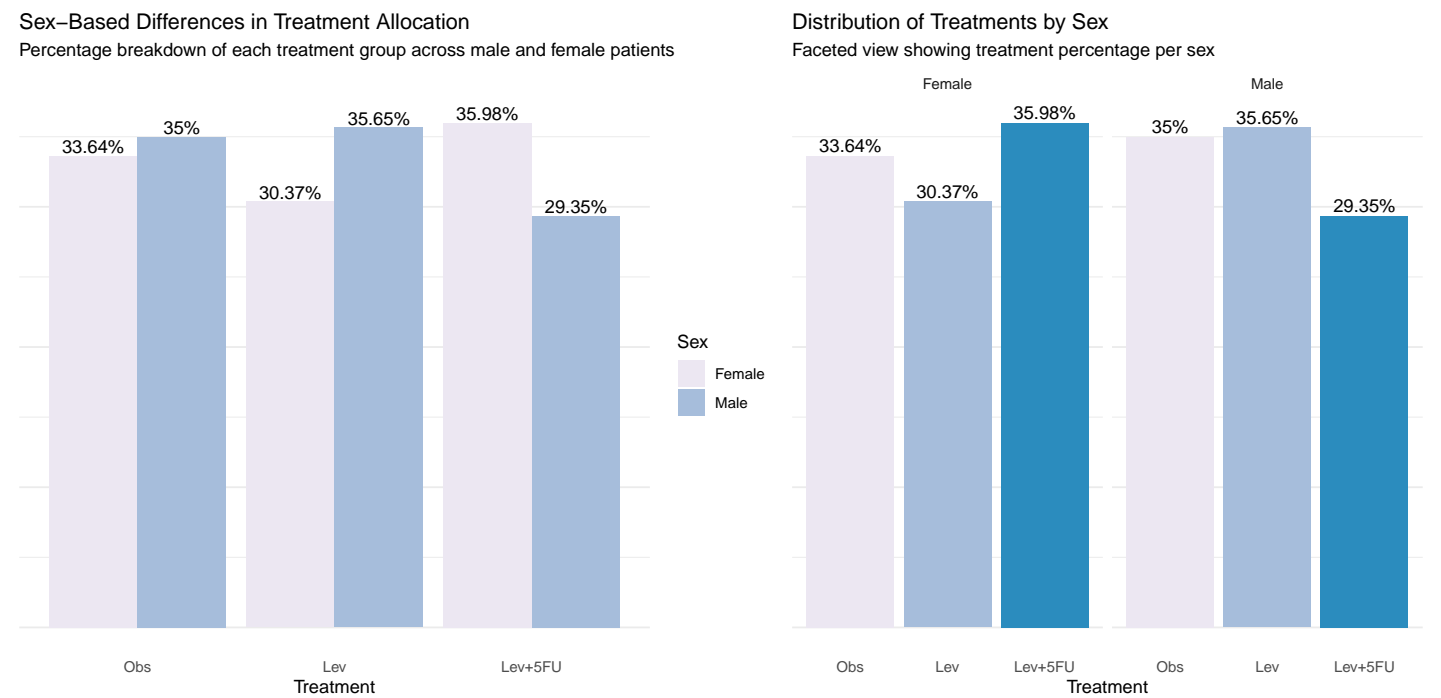
Loading of Data

Data Cleaning and Restructing

Visualization Gallery Plan

Descriptive / Baseline Characteristics:

Overview: Baseline visualization is essential in clinical trial reporting because it helps assess whether treatment arms are balanced across key demographic and clinical variables. This section focuses on sex-based distribution across treatment groups in the colon cancer dataset.



Visualization Description: The plots above summarize treatment allocation by sex using percentage-based bar charts.

- **The left panel** (“Sex-Based Differences in Treatment Allocation”) compares the proportion of male and female patients within each treatment arm: Observation (Obs), Levamisole (Lev), and Levamisole plus 5-Fluorouracil (Lev+5FU).
- **The right panel** (“Distribution of Treatments by Sex”) reverses the view, faceting by sex to show treatment allocation within males and females separately.

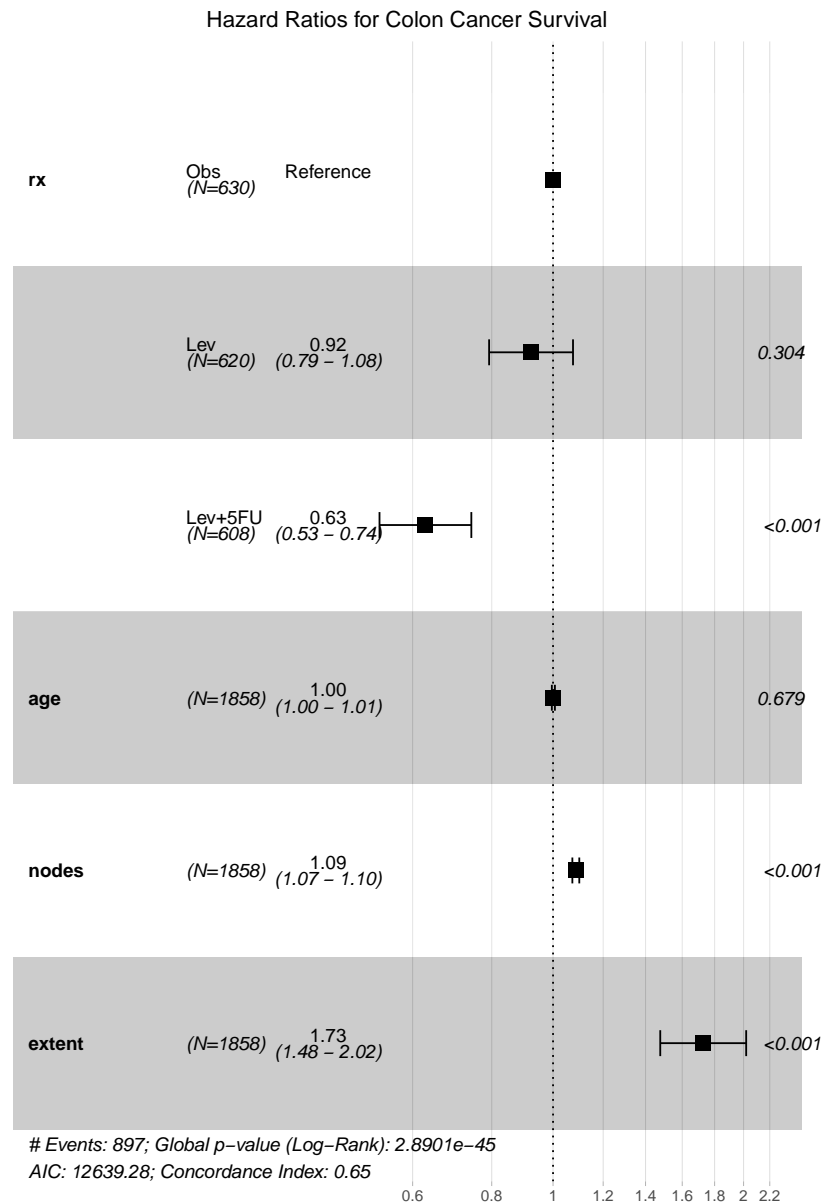
Interpretation Across treatment groups, the proportions of male and female participants appear balanced, indicating minimal sex-based bias in treatment allocation:

- **Female distribution:** ~33.5% (Obs), 29.9% (Lev), and 36.6% (Lev+5FU)
- **Male distribution:** ~34.3% (Obs), 36.6% (Lev), and 29.1% (Lev+5FU)

These small percentage variations suggest randomization was effective, maintaining equitable representation between sexes across treatment arms — a key prerequisite for unbiased inference on treatment effects.

Survival & Prognostic Analysis:

Overview: Survival analysis is a cornerstone of oncology research, allowing clinicians and statisticians to quantify treatment effects while adjusting for patient-specific prognostic factors. Here, a Cox proportional hazards model was fitted to assess how treatment type, age, nodal involvement, and disease extent influence survival among patients with resected colon cancer.



Visualization Description: The forest plot above presents hazard ratios (HRs) and their 95 % confidence intervals for each covariate. Each row represents a model term, with:

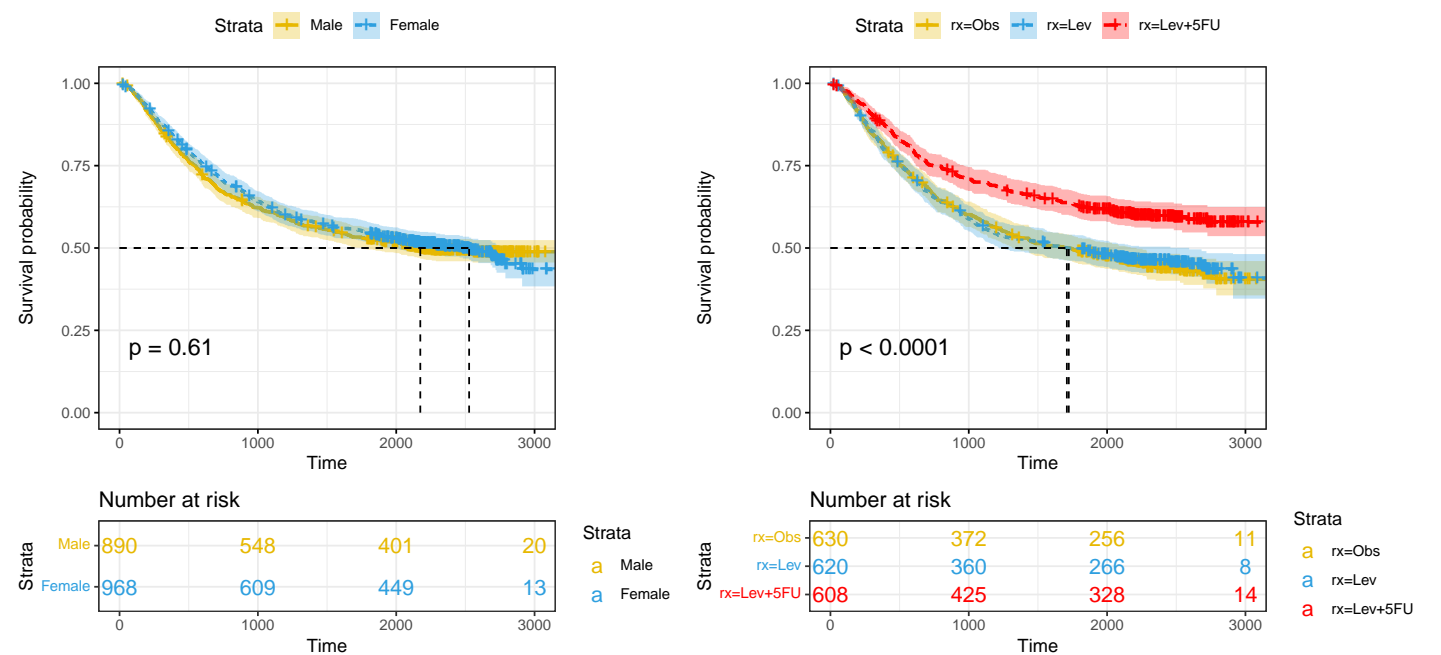
- Square markers indicating the point estimates of HRs
- Horizontal lines representing 95 % confidence intervals
- P-values displayed on the right margin to denote statistical significance
- The dashed vertical line (HR = 1) serving as the null reference (no effect)

Interpretation: The Cox model revealed that Levamisole plus 5-Fluorouracil (Lev+5FU) significantly improved survival compared to observation, reducing the risk of death by 37% (HR = 0.63, p < 0.001). In contrast, Levamisole alone showed no survival benefit (HR = 0.92, p = 0.304). Age had no significant impact on outcomes, while both lymph node involvement (HR = 1.09, p < 0.001) and disease extent (HR = 1.73, p < 0.001) were strong predictors of poorer survival. Overall, combination therapy markedly enhanced prognosis, whereas advanced disease features worsened outcomes.

Kaplan-Meier Curve Treatment & Subgroup Comparison

Overview: This analysis explores survival outcomes among patients with resected colon cancer using Kaplan–Meier survival curves. The objective is to assess whether survival differs significantly across key subgroups — specifically sex (male vs. female) and treatment type (Observation, Levamisole, Levamisole + 5-Fluorouracil). The analysis provides a non-parametric visualization of time-to-event data, allowing for comparison of survival probabilities and patterns across groups.

Kaplan–Meier Curves with Compact Risk Tables



Visualization Description:

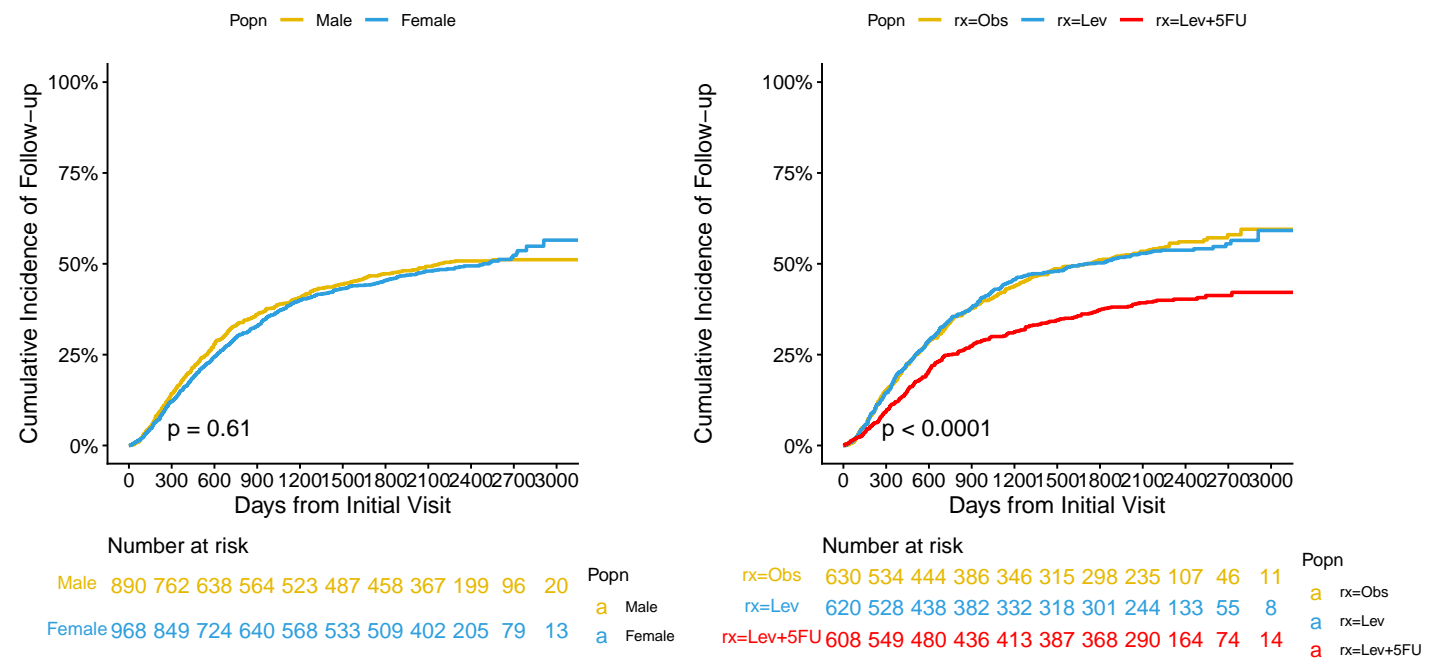
- Left panel** illustrates Kaplan–Meier curves stratified by sex, with survival probability plotted against time and compact risk tables indicating the number of subjects at risk at each time point. Male and female survival trajectories overlap closely, reflecting minimal visual difference between the two groups.
- Right panel** presents Kaplan–Meier curves by treatment type. Three groups are compared — Observation (Obs), Levamisole (Lev), and Levamisole + 5-Fluorouracil (Lev+5FU). Each curve displays the cumulative probability of survival over time, with shaded confidence bands and a corresponding risk table. Clear separation is observed between treatment groups, particularly the superior survival of the Lev+5FU group.

Interpretation: The log-rank test indicates no significant difference in survival by sex (p = 0.61), confirming that male and female patients have comparable outcomes. However, treatment effects are highly significant (p < 0.0001). Patients treated with Lev+5FU experience markedly better survival than those in the observation or Levamisole-only groups. This demonstrates the therapeutic advantage of combination chemotherapy, which reduces mortality risk and prolongs survival.

Cumulative Incidence Curve:

Overview: The cumulative incidence analysis examines the probability of recurrence or death over time among colon cancer patients, accounting for competing risks. This visualization provides a complementary perspective to the Kaplan–Meier survival curves by focusing on event occurrence rather than event-free survival, thereby highlighting how treatment and demographic factors influence the likelihood of experiencing an event during follow-up.

Kaplan–Meier Curves with Compact Risk Tables



Visualization Description:

- **Left panel** displays cumulative incidence curves stratified by sex, illustrating the probability of experiencing a clinical event (e.g., recurrence or death) over time. Both male and female curves closely overlap, indicating similar progression patterns throughout the study period.
- **Right panel** presents cumulative incidence by treatment type—Observation (Obs), Levamisole (Lev), and Levamisole + 5-Fluorouracil (Lev+5FU). The red curve (Lev+5FU) rises more gradually than the others, showing a lower cumulative incidence of events, while the observation group exhibits the steepest curve, reflecting higher event occurrence rates.

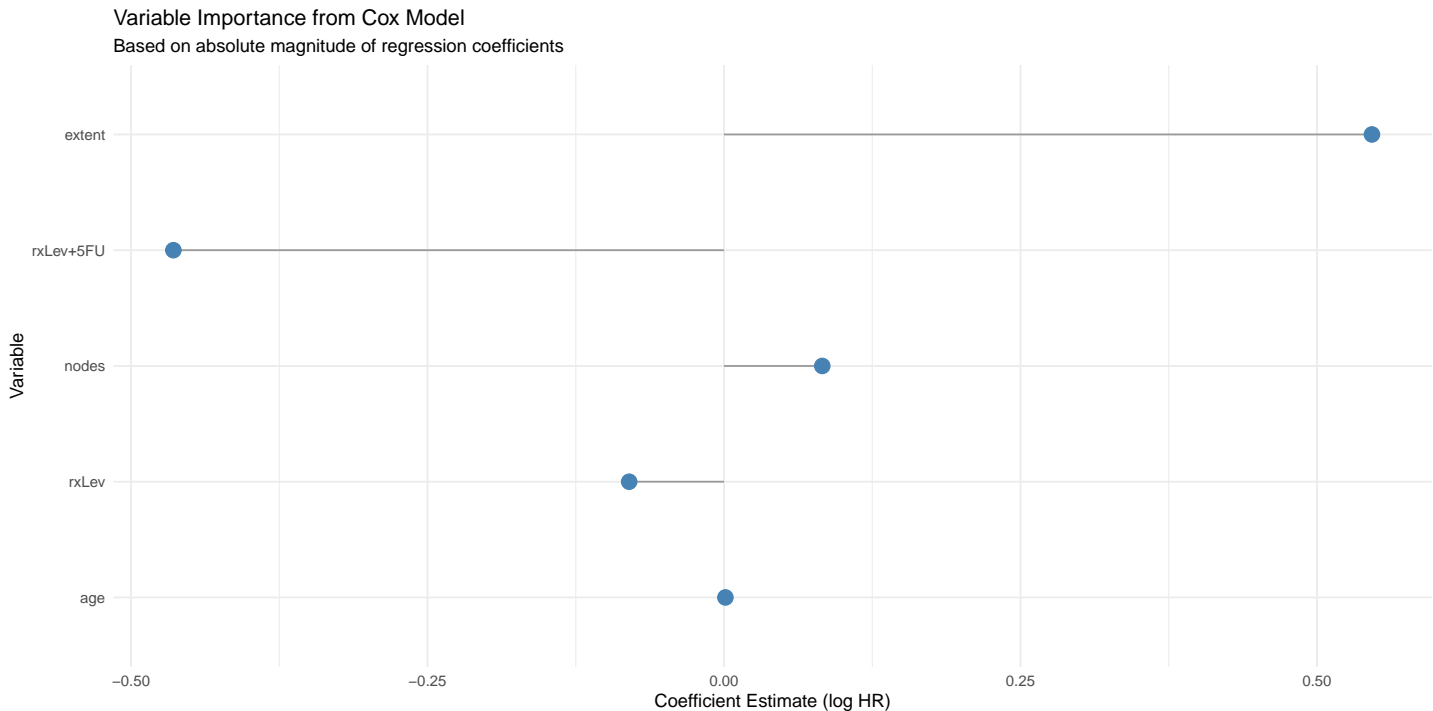
Risk tables below each plot summarize the number of patients still under observation at each follow-up time.

Interpretation: The cumulative incidence curves confirm that sex has no significant effect on event probability (p = 0.61), consistent with survival analyses. In contrast, treatment type shows a marked difference (p < 0.0001). Patients receiving Lev+5FU have a substantially lower incidence of recurrence or death compared to those on Levamisole alone or under observation.

This pattern reinforces the protective effect of combination therapy, demonstrating that Lev+5FU effectively delays or prevents adverse events over time. Overall, treatment regimen—not sex—is the major determinant of clinical outcome in this cohort.

Predictive & Exploratory Visuals:

Overview: This section explores the relative importance of key prognostic variables in predicting survival among colon cancer patients using the Cox proportional hazards model. By examining the magnitude and direction of regression coefficients, the visualization identifies which variables exert the strongest influence on patient outcomes and treatment efficacy.



Visualization Description: The plot displays the log hazard ratio (log HR) estimates for selected predictors — treatment (Lev and Lev+5FU), age, number of positive lymph nodes, and extent of disease. Each point represents a variable’s estimated coefficient, while its position along the x-axis indicates the direction and strength of association with survival:

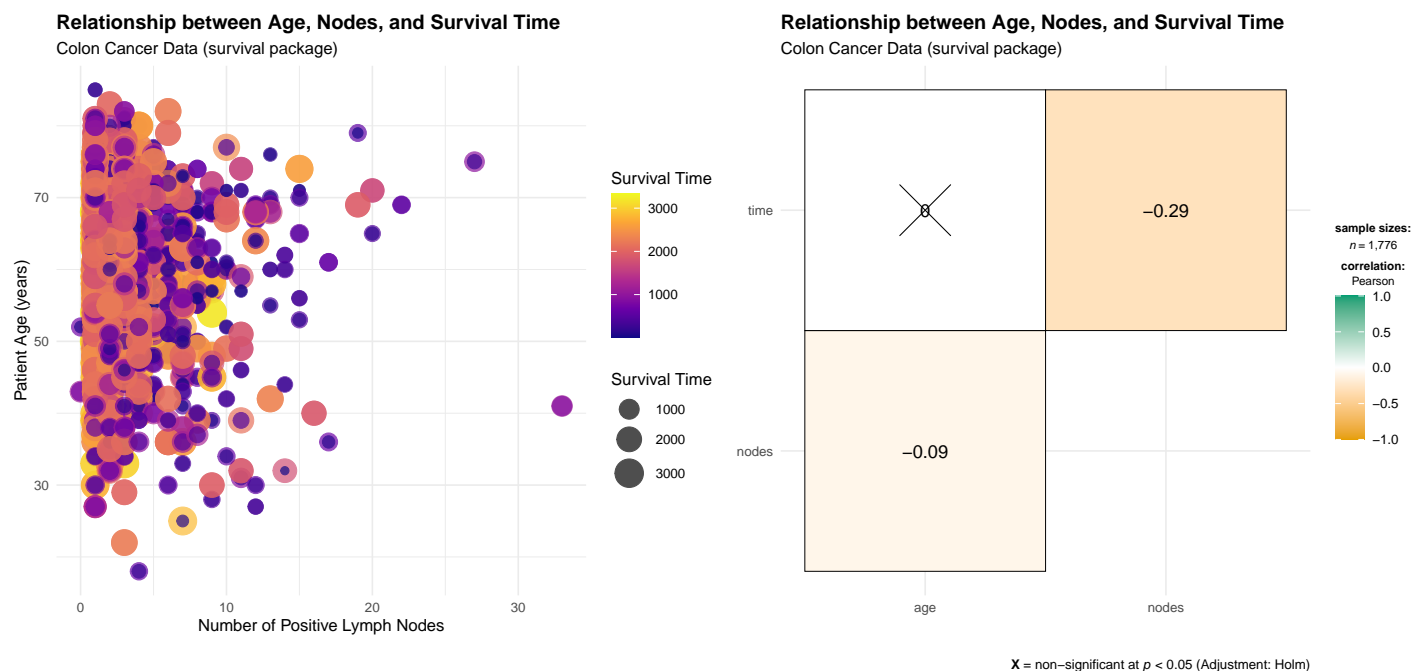
- Negative values imply a protective effect (reduced hazard),
- Positive values indicate higher risk of death or recurrence.

The visualization provides a comparative view of variable impact, emphasizing the most influential predictors in the Cox regression model.

Interpretion: The analysis reveals that disease extent and lymph node involvement are the strongest predictors of poor survival, while Lev+5FU treatment significantly improves outcomes. In contrast, Levamisole alone and age have minimal predictive influence, confirming that disease burden drives prognosis, and combination therapy offers the greatest survival benefit.

Correlation and Distribution of Key Prognostic Variables

Overview: This visualization explores how patient age, number of positive lymph nodes, and survival time interact in the colon cancer dataset. It provides both a graphical and statistical perspective to identify whether demographic or pathological features are associated with survival duration.



Visualization Description:

- **Left panel** shows a bubble plot, where each point represents a patient. The x-axis shows the number of positive lymph nodes, the y-axis represents age, and the color and size of each bubble correspond to survival time — warmer colors and larger circles indicate longer survival.
- **Right panel** presents a correlation heatmap quantifying relationships among variables (age, nodes, and survival time). Correlation coefficients are shown with color shading, where darker tones denote stronger associations.

Interpretation: The results reveal no strong correlations among the three variables. Survival time shows a moderate negative correlation with lymph node count ($r = -0.29$), indicating that patients with more positive nodes tend to have shorter survival. The correlation between age and survival time ($r = 0$) is negligible, suggesting that age alone is not a major determinant of survival in this cohort.

Conclusion

This visualization project effectively demonstrates how statistical graphics can enhance the interpretation of clinical trial data. Through survival curves, cumulative incidence plots, and prognostic visualizations, key treatment and patient-level patterns were revealed—particularly the survival benefit of combination therapy (Lev+5FU) and the prognostic impact of disease extent and lymph node involvement. Overall, the project highlights the importance of data visualization as a powerful tool for communicating complex biomedical insights and supporting evidence-based decision-making in oncology research.

Citation

Olofin, D. (2025). *Clinical Visualization Gallery: Exploring Colon Cancer Survival and Prognostic Patterns Using R*. Retrieved from GitHub Repository and full report available at mywebsite.