Clinical Visualization

Daniel Olofin

2025-09-25

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 10 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.
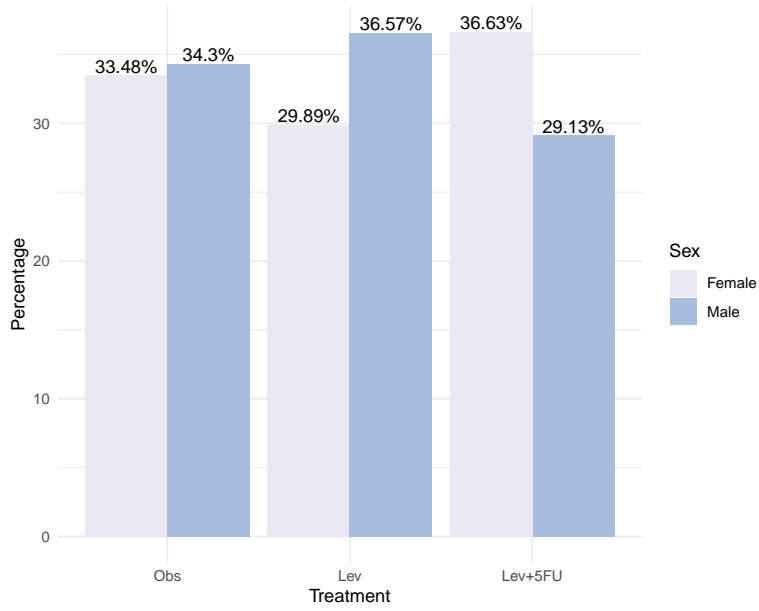
liberies

Loading of Data

Data Cleaning and Restructing

Visualization Gallery Plan
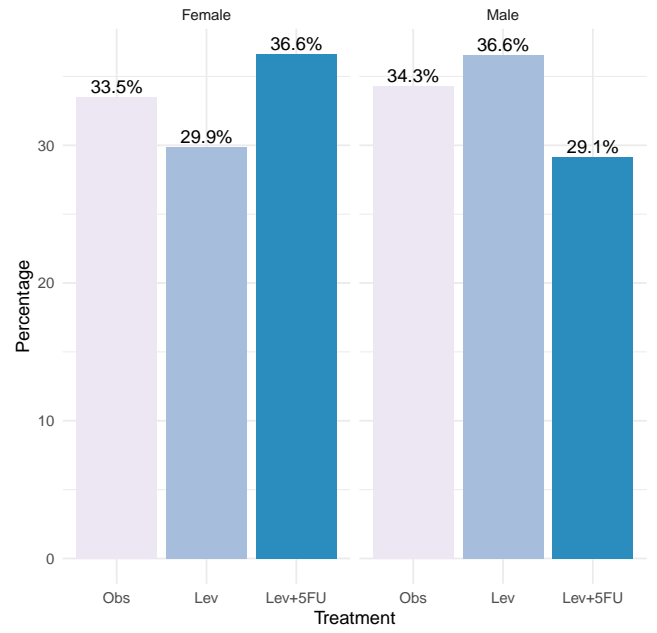
Descriptive / Baseline Characteristics

Sex–Based Differences in Treatment Allocation
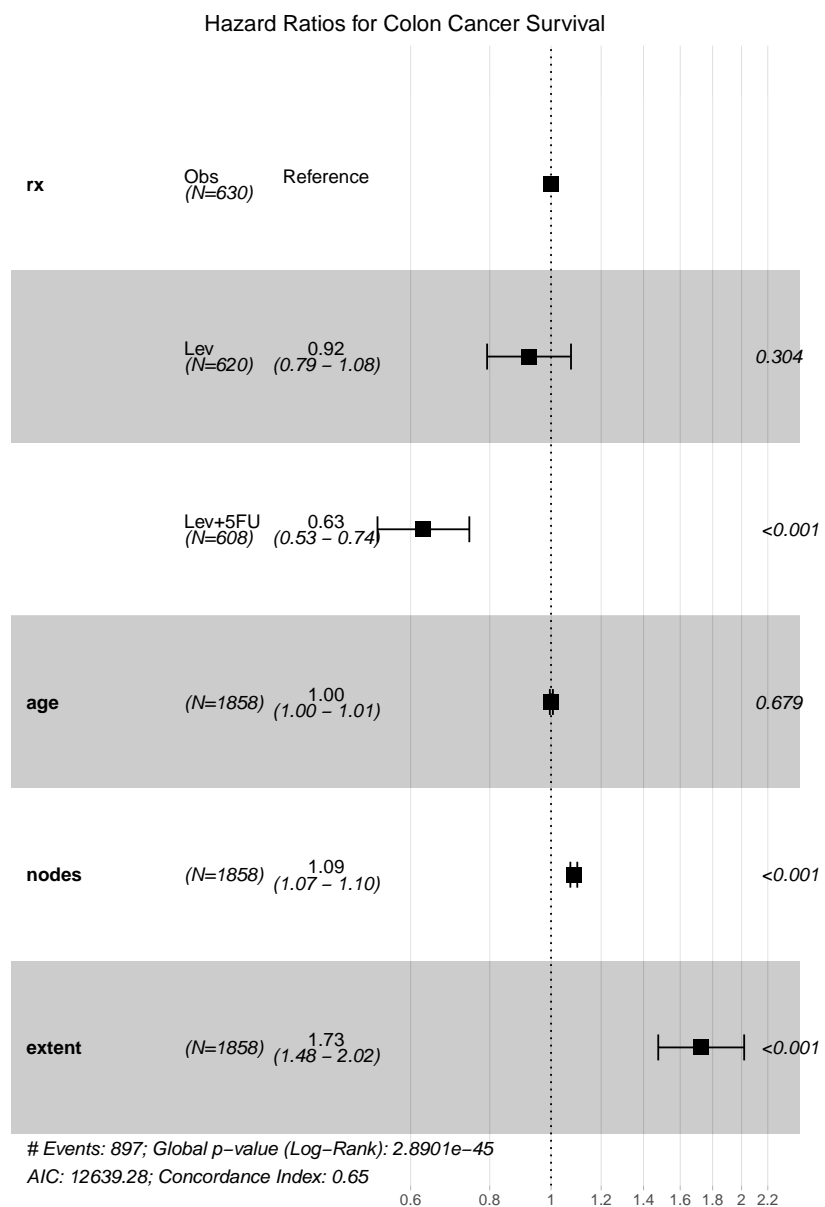Percentage breakdown of each treatment group across male and female patients

Distribution of Treatments by Sex
Faceted view showing treatment percentage per sex


Survival & Prognostic Analysis

Hazard Ratios for Colon Cancer Survival

rx	Obs (N=630)	Reference		
	Lev (N=620)	0.92 (0.79 – 1.08)		0.304
	Lev+5FU (N=608)	0.63 (0.53 – 0.74)		<0.001
age	(N=1858)	1.00 (1.00 – 1.01)		0.679
nodes	(N=1858)	1.09 (1.07 – 1.10)		<0.001
extent	(N=1858)	1.73 (1.48 – 2.02)		<0.001

Events: 897; Global p–value (Log–Rank): 2.8901e–45
AIC: 12639.28; Concordance Index: 0.65
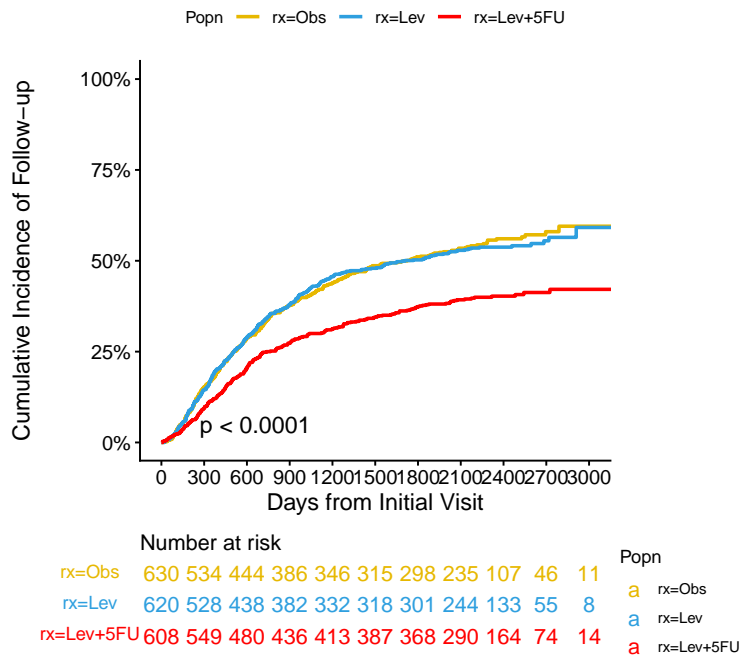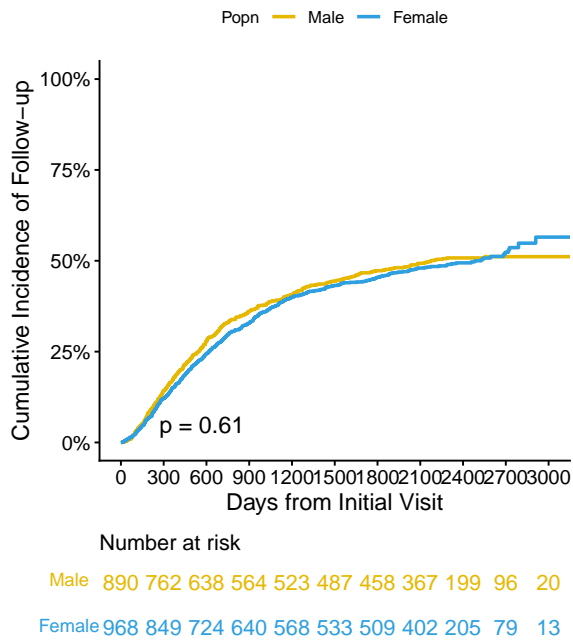
0.6   0.8   1   1.2   1.4 1.6 1.8 2 2.2

1. Kaplan-Meier Curve Treatment & Subgroup Comparison

Kaplan–Meier Curves with Compact Risk Tables
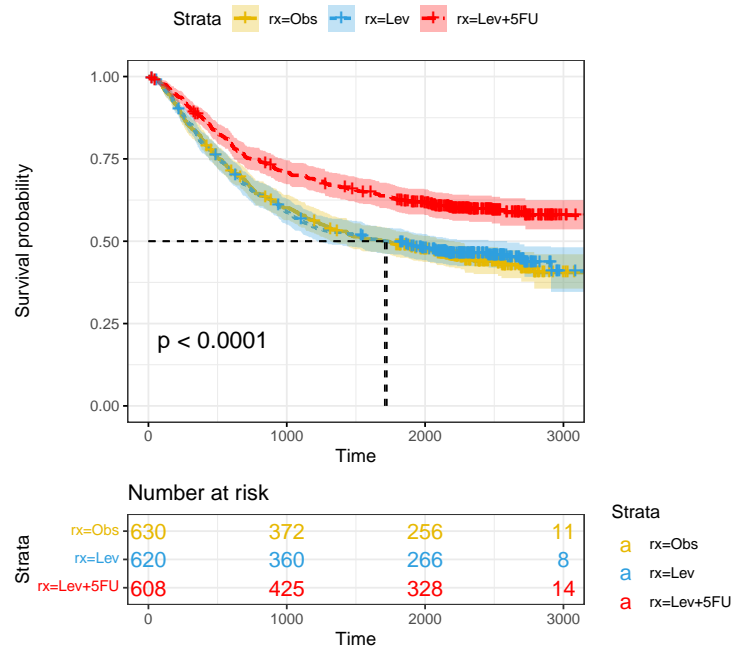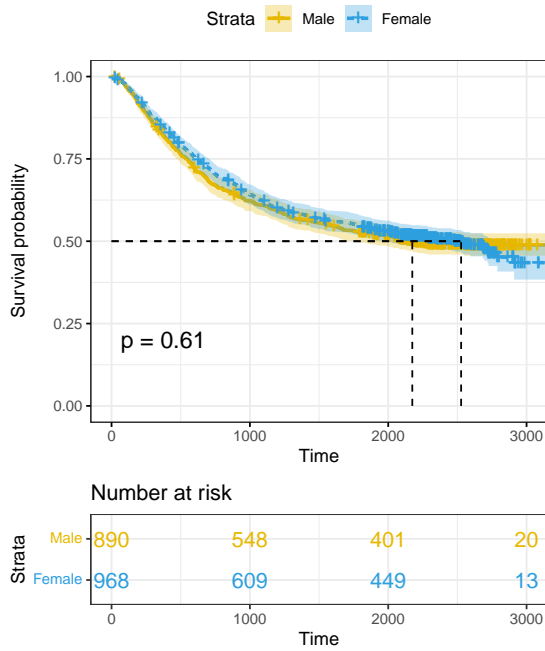
Cumulative Incidence Curve

Kaplan–Meier Curves with Compact Risk Tables

Variable Importance from Cox Model
Based on absolute magnitude of regression coefficients