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

```
new_data<-mydata %>%
  mutate(
    sex = ifelse(sex == 1, "Male", "Female"),
    obstruct = ifelse(obstruct == 0, "No obstruction", "Obstruction present"),
    perfor = ifelse(perfor == 0, "No", "Yes"),
    adhere = ifelse(adhere == 0, "No", "Yes"),
    status = ifelse(status == 0, "Censored",
```

```
            ifelse(status == 1, "Recurrence", "Death")),
   differ = ifelse(differ == 1, "Well-differentiated",
            ifelse(differ == 2, "Moderately differentiated", "Poorly differentiated")),
   extent = ifelse(extent == 1, "Submucosa",
            ifelse(extent == 2, "Muscle",
            ifelse(extent == 3, "Serosa", "Contiguous structures"))),
   surg = ifelse(surg == 0, "Short (<8 weeks)", "Long ( 8 weeks)"),
   node4 = ifelse(node4 == 0, "< 4 nodes", "  4 nodes"),
   etype = ifelse(etype == 1, "Recurrence", "Death")
 )


# Creating a Data Dictionary

# Relabel dictionary for colon dataset
var_labels <- c(
  id      = "Patient ID",
  study   = "Study ID",
  rx      = "Treatment Group",
  sex     = "Sex (0=Female, 1=Male)",
  age     = "Age (years)",
  obstruct = "Colon Obstruction (Yes/No)",
  perfor  = "Colon Perforation (Yes/No)",
  adhere  = "Tumor Adherence (Yes/No)",
  nodes   = "Positive Lymph Nodes",
  status  = "Event Status (0=Censored, 1=Recurrence, 2=Death)",
  differ  = "Tumor Differentiation (1=Well, 2=Moderate, 3=Poor)",
  extent  = "Extent of Spread (1=Submucosa, 4=Contiguous)",
  surg    = "Surgery to Registration (0=<8wks, 1= 8wks)",
  node4   = " 4 Positive Lymph Nodes (Yes/No)",
  time    = "Follow-up Time (days)",
  etype   = "Event Type (1=Recurrence, 2=Death)"
)
```

Visualization Gallery Plan

```
fit_1 <- survfit(Surv(time,status) ~ sex, data = mydata)


ggsurvplot(fit_1,
         pval = TRUE, conf.int = TRUE,
         risk.table = TRUE, # Add risk table
         risk.table.col = "strata", # Change risk table color by groups
         linetype = "strata", # Change line type by groups
         surv.median.line = "hv", # Specify median survival
         ggtheme = theme_bw(), # Change ggplot2 theme
         palette = c("#E7B800", "#2E9FDF"),
         legend.labs = c("Male","Female"))
```

1. Kaplan-Meier Curve

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## i The deprecated feature was likely used in the ggpubr package.
##   Please report the issue at <https://github.com/kassambara/ggpubr/issues>.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

```
## Warning in geom_segment(aes(x = 0, y = max(y2), xend = max(x1), yend = max(y2)), : All aesthetics ha
## i Please consider using `annotate()` or provide this layer with data containing
##   a single row.
## All aesthetics have length 1, but the data has 2 rows.
## i Please consider using `annotate()` or provide this layer with data containing
##   a single row.
## All aesthetics have length 1, but the data has 2 rows.
## i Please consider using `annotate()` or provide this layer with data containing
##   a single row.
## All aesthetics have length 1, but the data has 2 rows.
## i Please consider using `annotate()` or provide this layer with data containing
##   a single row.
```
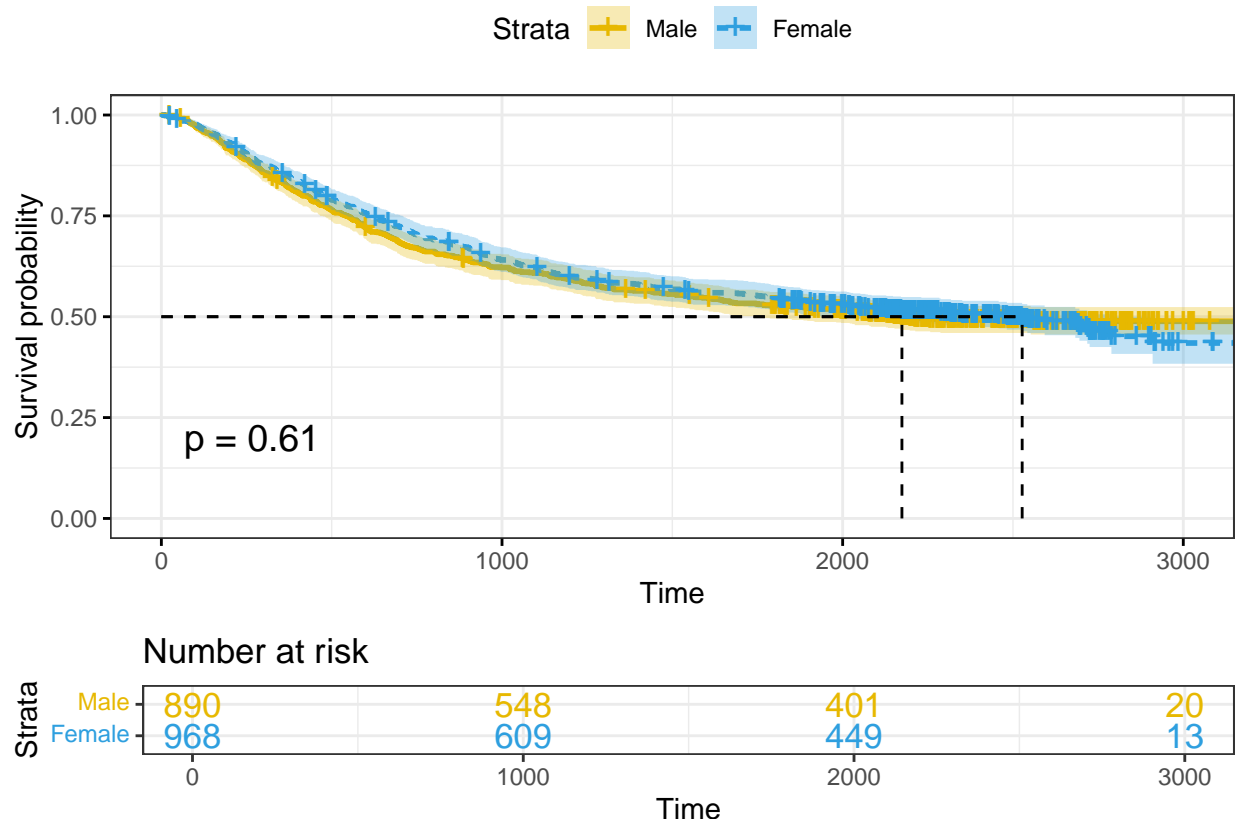


```r
fit_2 <- survfit(Surv(time, status) ~ rx, data = mydata)

ggsurvplot(fit_2,
          pval = F, conf.int = TRUE,
```
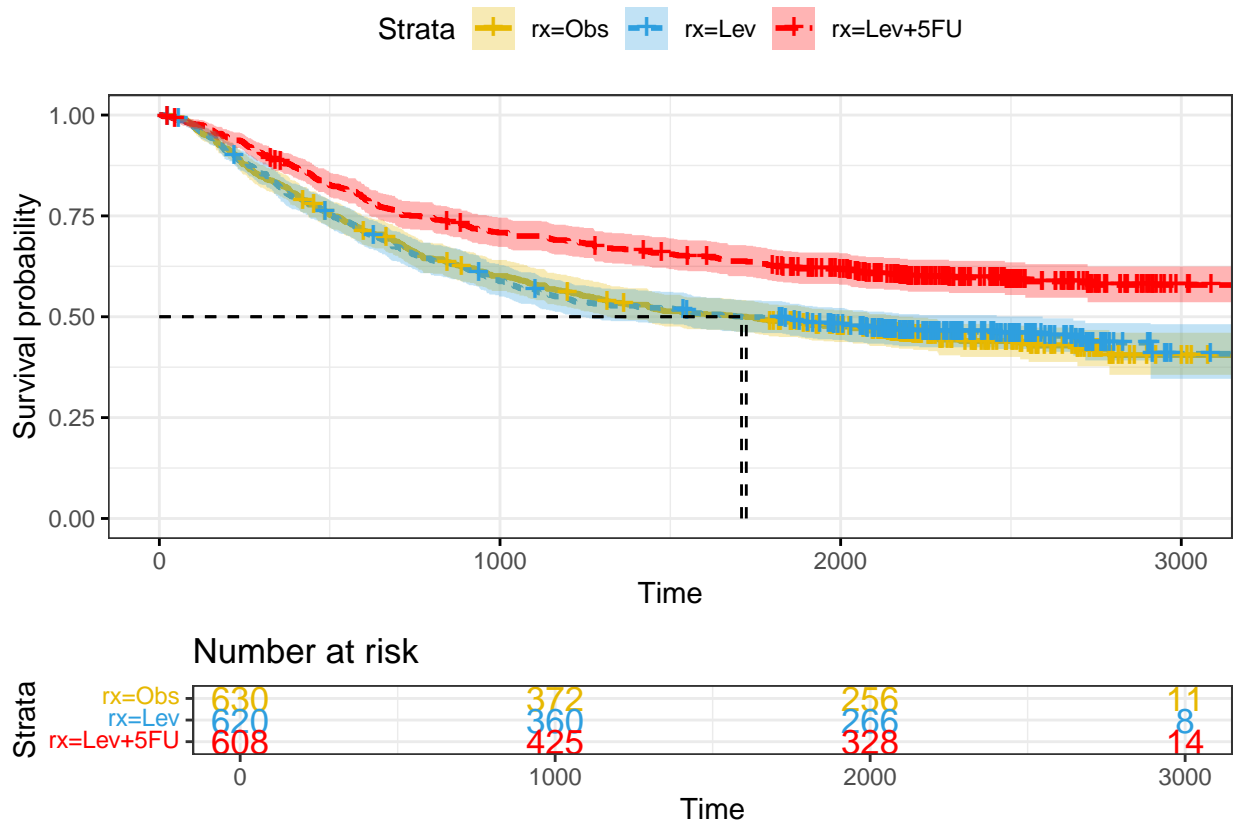
```
          risk.table = TRUE, # Add risk table
          risk.table.col = "strata", # Change risk table color by groups
          linetype = "strata", # Change line type by groups
          surv.median.line = "hv", # Specify median survival
          ggtheme = theme_bw(), # Change ggplot2 theme
          palette = c("#E7B800", "#2E9FDF", "red"))
```

```
## Warning in geom_segment(aes(x = 0, y = max(y2), xend = max(x1), yend = max(y2)), : All aesthetics hav
## i Please consider using `annotate()` or provide this layer with data containing
##   a single row.
## All aesthetics have length 1, but the data has 2 rows.
## i Please consider using `annotate()` or provide this layer with data containing
##   a single row.
## All aesthetics have length 1, but the data has 2 rows.
## i Please consider using `annotate()` or provide this layer with data containing
##   a single row.
## All aesthetics have length 1, but the data has 2 rows.
## i Please consider using `annotate()` or provide this layer with data containing
##   a single row.
```



```
ggsurvplot(fit_1,
          data = mydata,
          fun = "event",
          censor = FALSE,
```

```
             pval = TRUE,       # set to FALSE if don't want p-value
             pval.coord = c(270, 0.05),   # can delete this line if don't want p-value
             risk.table = TRUE,
             legend.title = "Popn",
             legend.labs = c("Male", "Female"),
             palette = c("#E7B800", "#2E9FDF"),
             break.time.by = 300,
             xlab = "Days from Initial Visit",
             ylab = "Cumulative Incidence of Follow-up",
             axes.offset = TRUE,
             ylim = c(0, 1),
             tables.theme = theme_void(),
             tables.height = 0.15,
             tables.col = "strata",
             surv.scale = "percent")
```
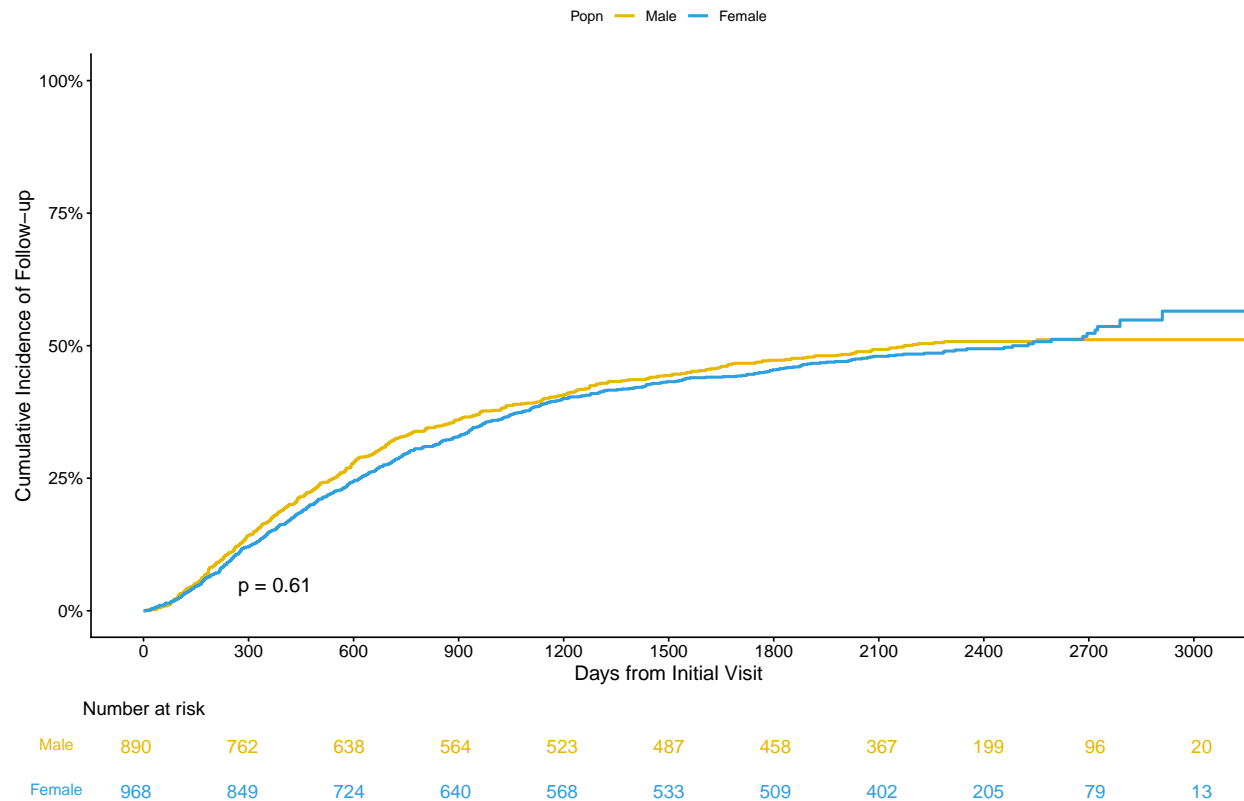
```
## Warning: ggtheme is not a valid theme.
## Please use `theme()` to construct themes.
```

```
## Ignoring unknown labels:
## * fill : "Popn"
## * linetype : "1"
```

```
## Warning: ggtheme is not a valid theme.
## Please use `theme()` to construct themes.
```

```
## Ignoring unknown labels:
## * fill : "Popn"
## * linetype : "1"
## Ignoring unknown labels:
## * fill : "Popn"
## * linetype : "1"
## Ignoring unknown labels:
## * fill : "Popn"
## * linetype : "1"
```

Number at risk

Male	890	762	638	564	523	487	458	367	199	96	20
Female	968	849	724	640	568	533	509	402	205	79	13

```r
ggsurvplot(fit_2,
           data = mydata,
           fun = "event",
           censor = FALSE,
           pval = TRUE,    # set to FALSE if don't want p-value
           pval.coord = c(270, 0.05),  # can delete this line if don't want p-value
           risk.table = TRUE,
           legend.title = "Popn",
           palette = c("#E7B800", "#2E9FDF", "red"),
           break.time.by = 300,
           xlab = "Days from Initial Visit",
           ylab = "Cumulative Incidence of Follow-up",
           axes.offset = TRUE,
           ylim = c(0, 1),
           tables.theme = theme_void(),
           tables.height = 0.15,
           tables.col = "strata",
           surv.scale = "percent")
```

```
## Warning: ggtheme is not a valid theme.
## Please use `theme()` to construct themes.

## Ignoring unknown labels:
## * fill : "Popn"
## * linetype : "1"

## Warning: ggtheme is not a valid theme.
```

```
## Please use `theme()` to construct themes.

## Ignoring unknown labels:
## * fill : "Popn"
## * linetype : "1"
## Ignoring unknown labels:
## * fill : "Popn"
## * linetype : "1"
## Ignoring unknown labels:
## * fill : "Popn"
## * linetype : "1"
```
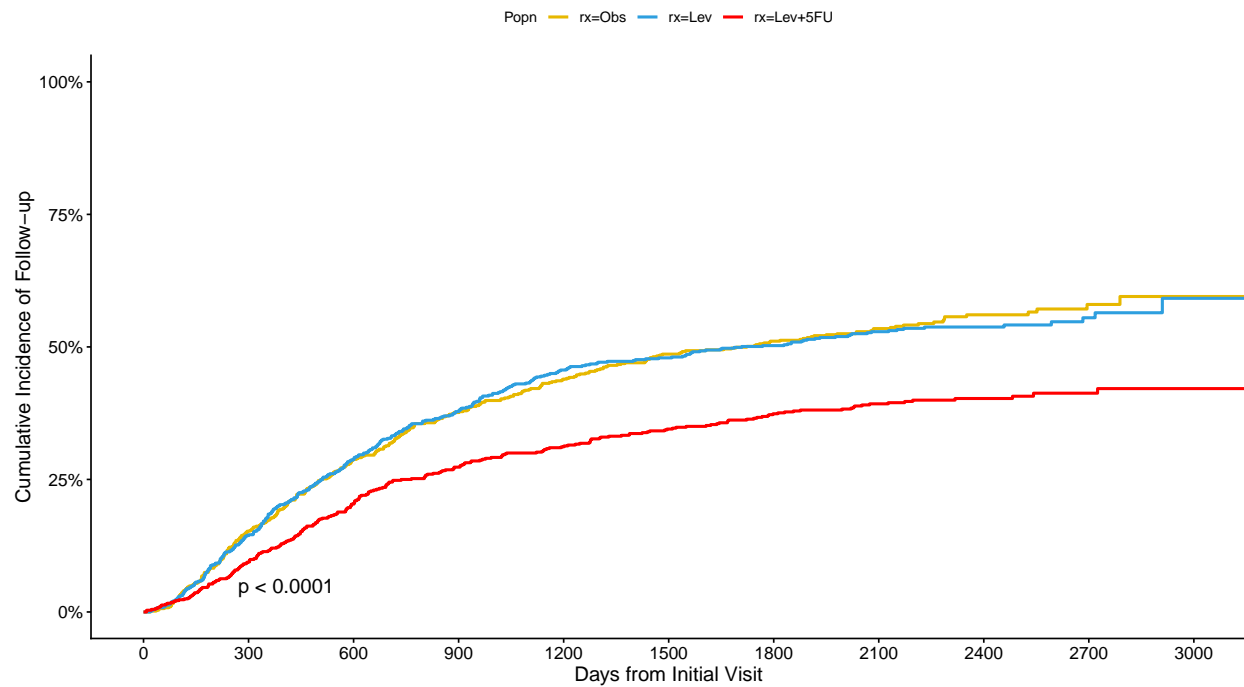
Popn — rx=Obs — rx=Lev — rx=Lev+5FU