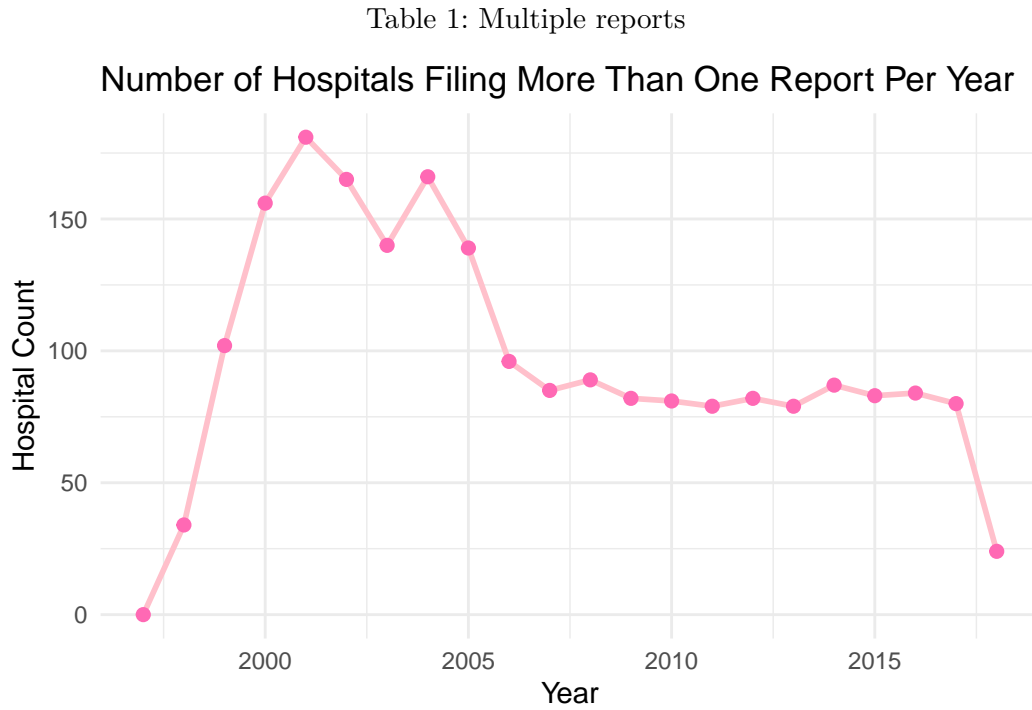# Homework 2

Sushmita Rajan

Homework 2

[Click here to view my repository](#)

1. How many hospitals filed more than one report in the same year? Show your answer as a line graph of the number of hospitals over time.

```
Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
i Please use `linewidth` instead.
```

Table 1: Multiple reports

Number of Hospitals Filing More Than One Report Per Year

2. After removing/combining multiple reports, how many unique hospital IDs (Medicare provider numbers) exist in the data?

```r
library(scales)
library(ggplot2)

# Violin Plot for Total Charges Distribution (Updated)
ggplot(final.hcris.data, aes(x = as.factor(year), y = tot_charges)) +
  # Generate the violin plot with customized fill and outline colors
  geom_violin(fill = "#8A2BE2", color = "#006400", alpha = 0.75) +  # Violet fill and dark gr

  # Apply a log scale to the y-axis and format labels with commas for readability
  scale_y_log10(
    labels = label_comma(scale = 1, accuracy = 0.1)  # Adjusting accuracy for clearer labels
  ) +

  # Add titles, axis labels, and caption with a more descriptive text
  labs(
    title = "Distribution of Total Charges Over the Years",
    x = "Year of Report",
    y = "Log of Total Charges",
    caption = "Data Source: HCRIS (Medicare/Medicaid)"
  ) +

  # Minimal theme for better aesthetics
  theme_minimal() +

  # Adjust the x-axis labels to be angled for better visibility
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1, size = 10),  # Rotate x-axis labels
    axis.title.x = element_text(size = 14),  # Increase font size for x-axis title
    axis.title.y = element_text(size = 14),  # Increase font size for y-axis title
    plot.title = element_text(hjust = 0.5, size = 16),  # Center-align and increase font size
    plot.caption = element_text(size = 10)  # Adjusting font size for the caption
  )
```

Warning in transformation$transform(x): NaNs produced

Warning in scale_y_log10(labels = label_comma(scale = 1, accuracy = 0.1)):
log-10 transformation introduced infinite values.

Warning: Removed 4760 rows containing non-finite outside the scale range

(`stat_ydensity()`).

Distribution of Total Charges Over the Years

Data Source: HCRIS (Medicare/Medicaid)