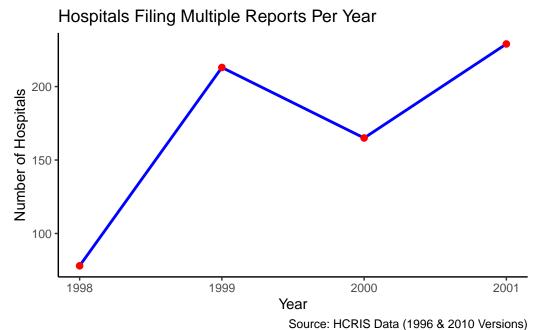# Homework 2

Research Methods, Spring 2025

Ethan Murakami

Here is a link to my repository (https://github.com/bemur3/hmwk2)

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

Table 1: Multiple reports

Hospitals Filing Multiple Reports Per Year

Source: HCRIS Data (1996 & 2010 Versions)
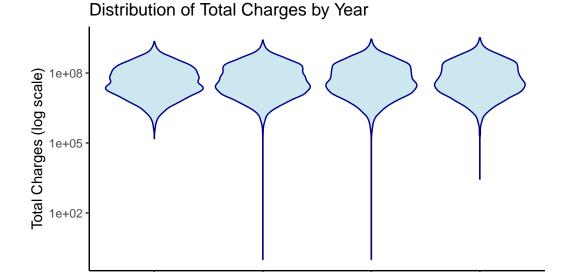
```
unique_hospital_count <- final.hcris.data %>%
  distinct(provider_number) %>%
  nrow()
```

```
cat("Number of unique hospital IDs (Medicare provider numbers):", unique_hospital_count, "\n")
```

Number of unique hospital IDs (Medicare provider numbers): 21637

```
ggplot(final.hcris.data, aes(x = factor(year), y = tot_charges)) +
  geom_violin(fill = "lightblue", color = "darkblue", alpha = 0.6) +
  scale_y_log10() + # Apply log scale to handle skewed distributions
  labs(
    title = "Distribution of Total Charges by Year",
    x = "Year",
    y = "Total Charges (log scale)",
    caption = "Source: HCRIS Data (1996 & 2010 Versions)"
  ) +
  theme_classic()
```

Warning: Removed 872 rows containing non-finite outside the scale range
(`stat_ydensity()`).

2