Story 3

Matthew Tillmawitz

2025-03-07

Prepare Data

```r
health_spending <- read_csv2("state_healthcare_spending.csv")
obesity <- read_excel("adult_obesity.xlsx")
```

```r
bucketed <- obesity |>
  left_join(health_spending, by = join_by(State == Location)) |>
  mutate(quintile = ntile(`Health Spending per Capita`,5))
```

Story Plot

```r
stat_notch_lines <- function(mapping = NULL, data = NULL,
                             position = "identity", na.rm = FALSE, show.legend = NA,
                             inherit.aes = TRUE, width = 0.5, color = "black", size = 0.5, ...) {
  ggplot2::layer(
    stat = StatNotchLines,
    data = data,
    mapping = mapping,
    geom = "segment",
    position = position,
    show.legend = show.legend,
    inherit.aes = inherit.aes,
    params = list(na.rm = na.rm, width = width, color = color, size = size, ...)
  )
}

# Create the StatNotchLines class
StatNotchLines <- ggproto("StatNotchLines", Stat,
  required_aes = c("y"),

  compute_group = function(data, scales, conf = 0.95, width = 0.5) {
    n <- length(data$y)
    med <- median(data$y, na.rm = TRUE)
    iqr <- IQR(data$y, na.rm = TRUE)

    # Calculate notch using the same formula as in geom_boxplot
    se <- 1.58 * iqr / sqrt(n)
```

```r
    # Calculate notch positions
    notch_min <- med - se
    notch_max <- med + se

    # Get the x position and calculate the width of the lines
    x <- data$x[1]
    half_width <- width/2

    # Return data frame with positions for segments
    data.frame(
      x = c(x - half_width, x - half_width),
      xend = c(x + half_width, x + half_width),
      y = c(notch_min, notch_max),
      yend = c(notch_min, notch_max)
    )
  }
)

bucketed |>
  mutate(quint_level = factor(quintile)) |>
  ggplot(mapping = aes(x = quint_level, y = `Obesity %`)) +
    geom_boxplot() +
    stat_notch_lines(color = "red") +
    labs(
        x = "Per Capita Spending Quintiles",
        y = "Obesity Rate",
        title = "No Correlation Between State Health Spending and Obesity",
        caption = "Overlapping 95% confidence intervals for median values indicate no significant correl
    ) +
    theme(
      panel.background = element_rect(
        fill = "white"
      ),
      axis.line.x = element_line(),
      plot.caption = element_text(color = "darkgrey")
    ) +
    annotate(
      geom = "text",
      label = "95% Confidence\nInterval",
      x = 4.35,
      y = 0.27,
      color = "red",
      size = 2.5
    )
```
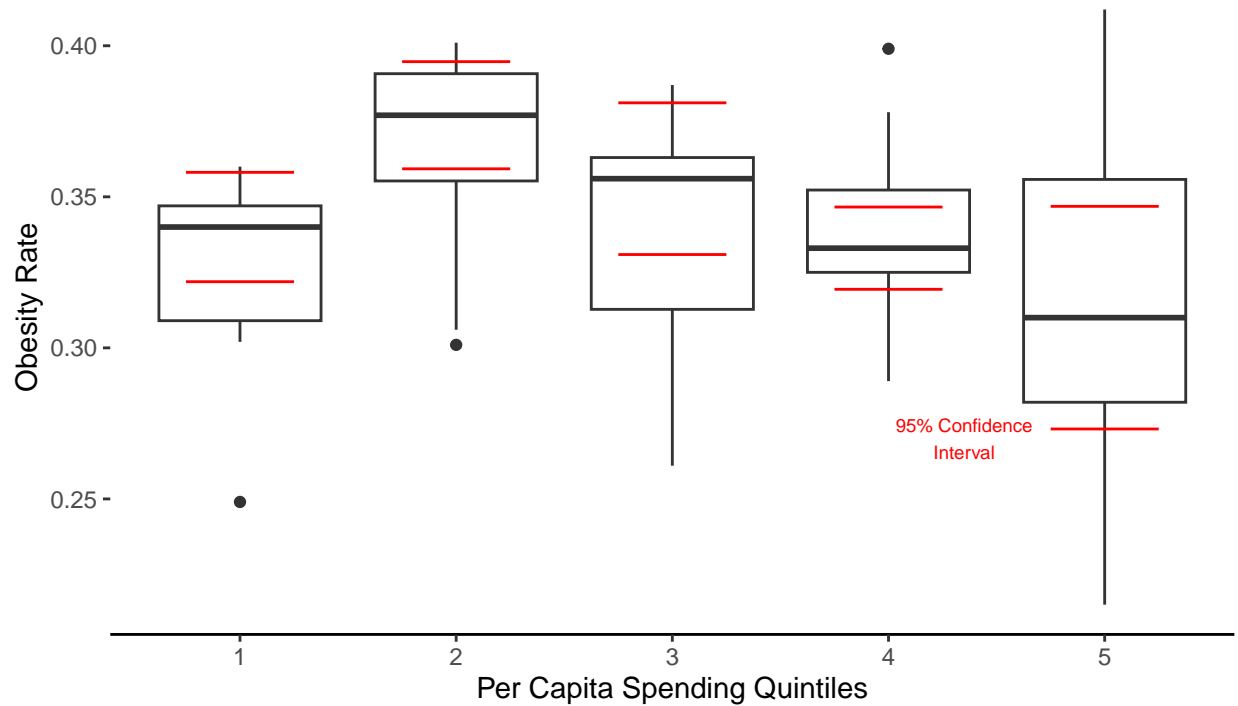
No Correlation Between State Health Spending and Obesity

Overlapping 95% confidence intervals for median values indicate no significant correlation between state per capita healthcare spending and obesity rates. State healthcare spending data comes from the Kaiser Family Foundation.

Overlapping 95% confidence intervals for the median value of each quintile indicates there is not a statistically significant correlation between per capita healthcare spending and obesity rates in states. It should be noted, however, that the confidence interval for the second quintile does not overlap with the fourth or fifth quintiles. While this may appear to indicate some level of significance, it is not unlikely to see a seemingly significant relationship in a sample of this size when there is no meaningful relationship. Given the overlap with the other two quintiles, and the significant overlap of confidence intervals among the other quintiles, we can reasonably conclude there is no relationship between the two variables.