# EPPS 6356: Data Visualization – Assignment 1

## 1. Anscombe’s Quartet (1973)

This section revisits Anscombe’s (1973) classic example demonstrating the importance of visualization in data analysis. The four datasets known as Anscombe’s Quartet share nearly identical summary statistics and regression lines, yet differ greatly in their distributions and relationships when plotted.

The analysis reveals that while all four datasets have similar means, variances, correlations, and regression coefficients, they exhibit distinct structures: (1) a linear relationship, (2) a curved relationship, (3) a linear trend distorted by an outlier, and (4) a nearly constant x-value with a single high-leverage point. These discrepancies highlight the danger of relying solely on numerical summaries without visual inspection. To address such issues, analysts should visualize data before modeling, use diagnostic plots to check assumptions, and apply robust regression techniques when outliers or nonlinearity are detected.

Generated Visualizations:

• anscombe\_base\_panels.png

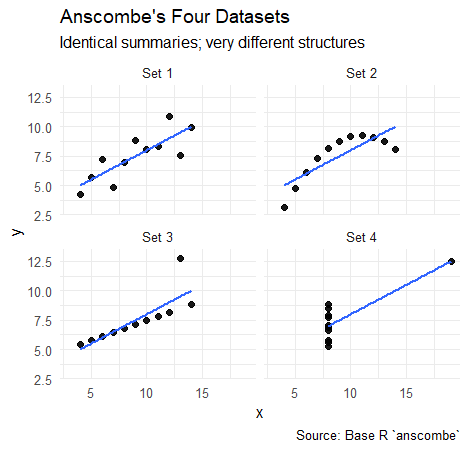
• anscombe\_faceted.png

• anscombe\_resid\_vs\_fitted.png (optional diagnostic)

## 2. Fall.R Color Customization Exercise

The Fall.R exercise emphasizes color design and aesthetic principles in visualization. By replacing the default 'Fall' palette with a self-selected 'Winter' palette, students demonstrate control over color theory and accessibility. The new color scheme incorporates cool hues deep blues and soft whites to evoke a winter theme while maintaining clear contrast and readability.

Generated Visualization: winter\_plot\_from\_FallR.png



## Conclusion

This assignment demonstrates that visualization is not merely decorative but diagnostic. Anscombe’s example shows that identical statistics can mask very different data patterns; the Fall.R exercise reinforces thoughtful aesthetic design; and the chart critique builds critical literacy in evaluating visual communication.