Multivariate Visualization

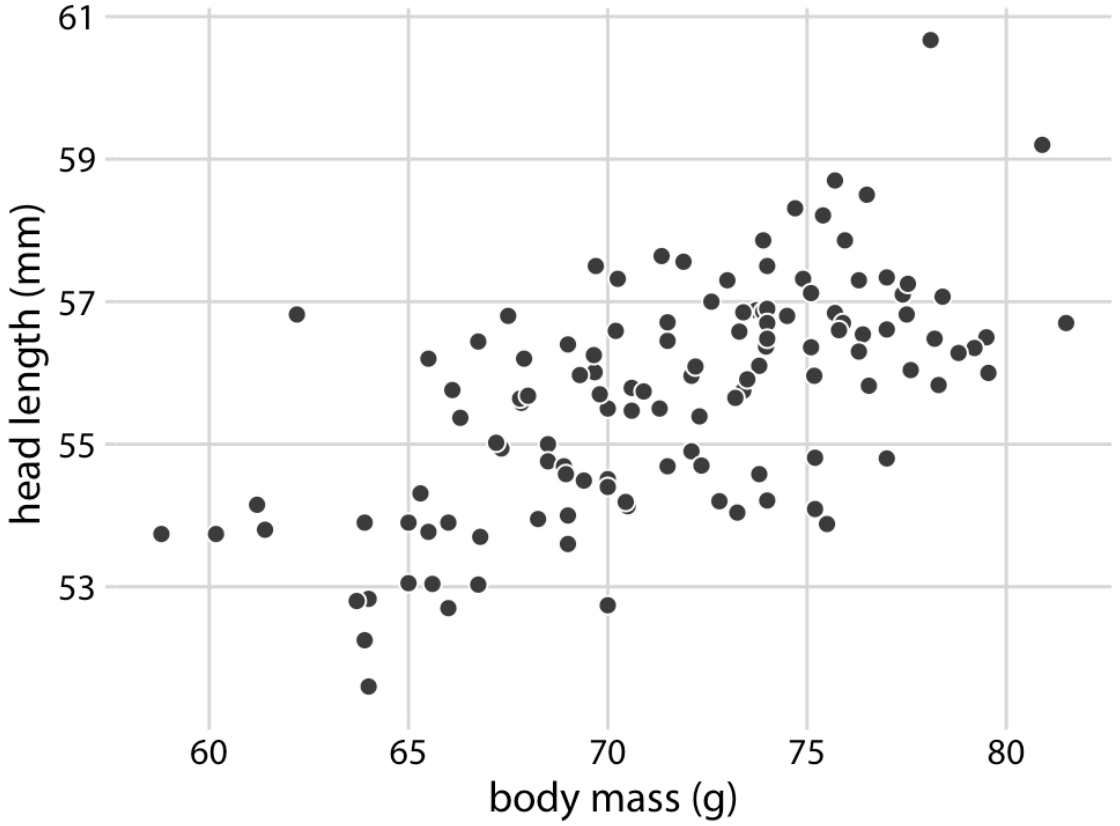
Visualizing Associations

Introduction to Associations Visualization

- **Introduction**
  - Purpose: Explore relationships among two or more quantitative variables
  - Examples: Animal measurements (height, weight, length, energy demands)
- **Visualization Techniques**
  - Scatter Plots
  - Bubble Charts
  - Scatter Plot Matrices
  - Correlograms
  - Dimension Reduction
  - Paired Data Visualizations

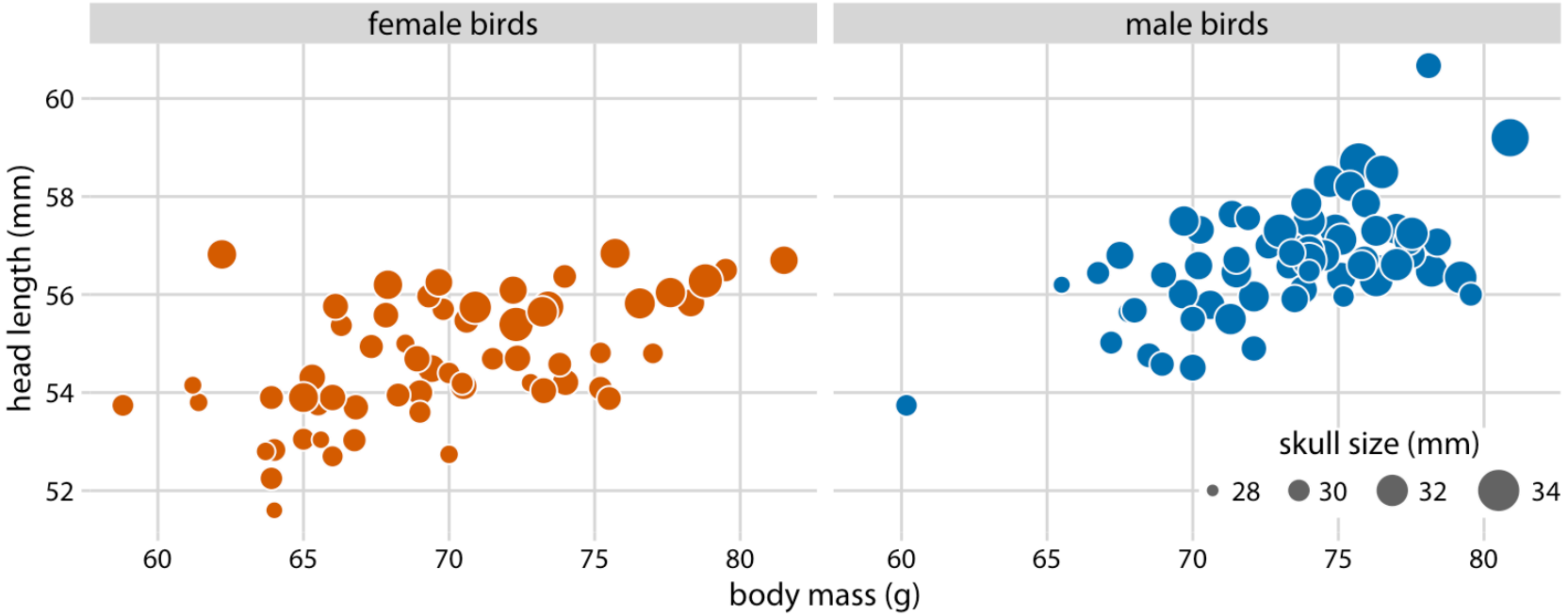
Scatter Plots

- **Basic Scatter Plot**
- **Example**: Blue jay birds dataset (head length vs. body mass)
- **Observation**: Trend of heavier birds having longer heads
- **Terminology**: Y-axis variable plotted against X-axis variable
- **Colored Scatter Plot**
- **Addition**: Points colored by bird sex
- **Observation**: Male birds tend to have longer heads at the same body mass



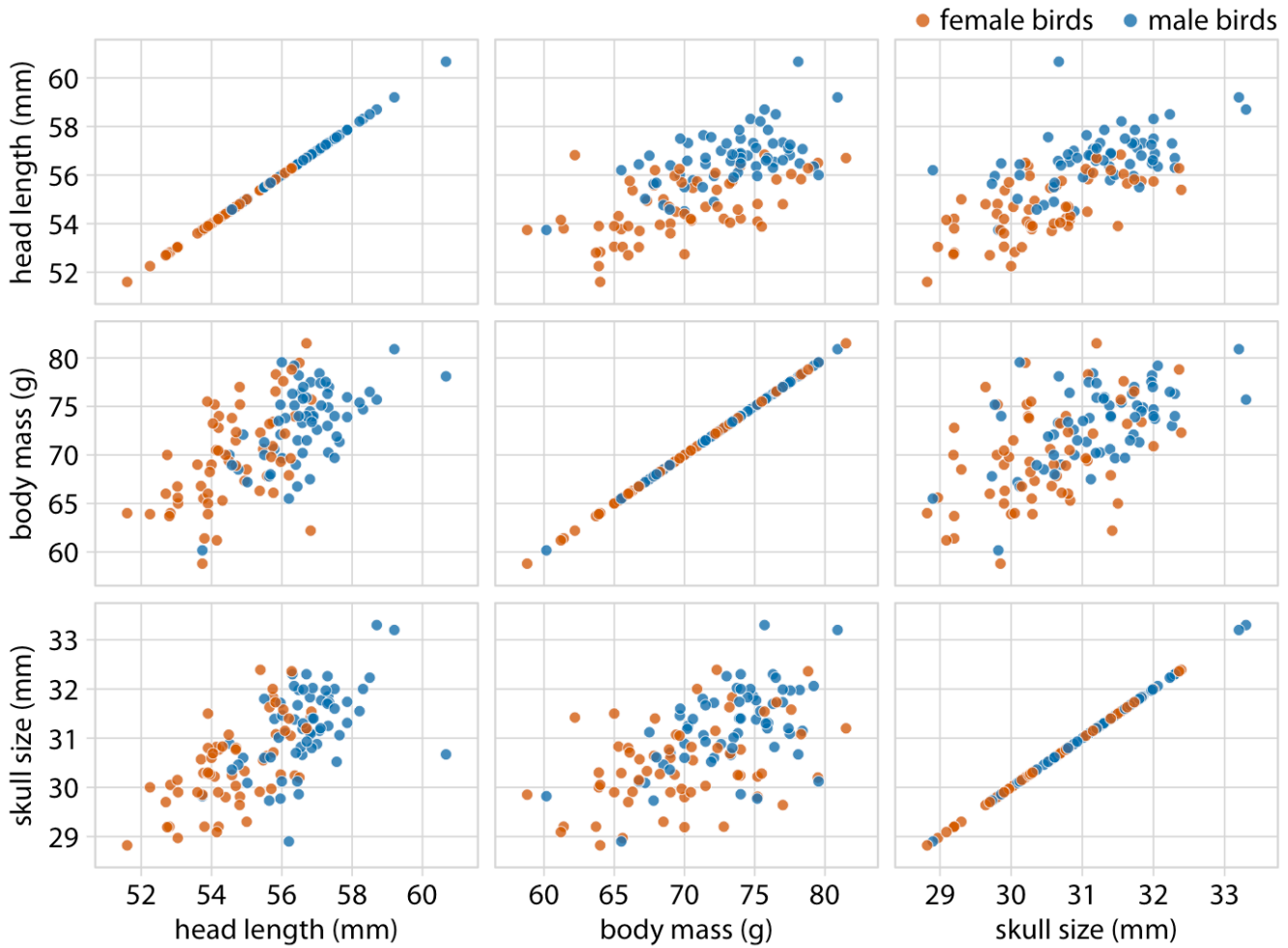
Scatter Plots (continued)

- **Bubble Chart**
- **Additional Variable**: Skull size indicated by dot size
- **Observation**: Correlation between head length and skull size, but with visual limitations



Mosaic Plots

- **Scatter Plot Matrix**
- **All-against-All Comparison**: Head length, body mass, skull size
- **Advantage**: Easier perception of correlations compared to bubble charts

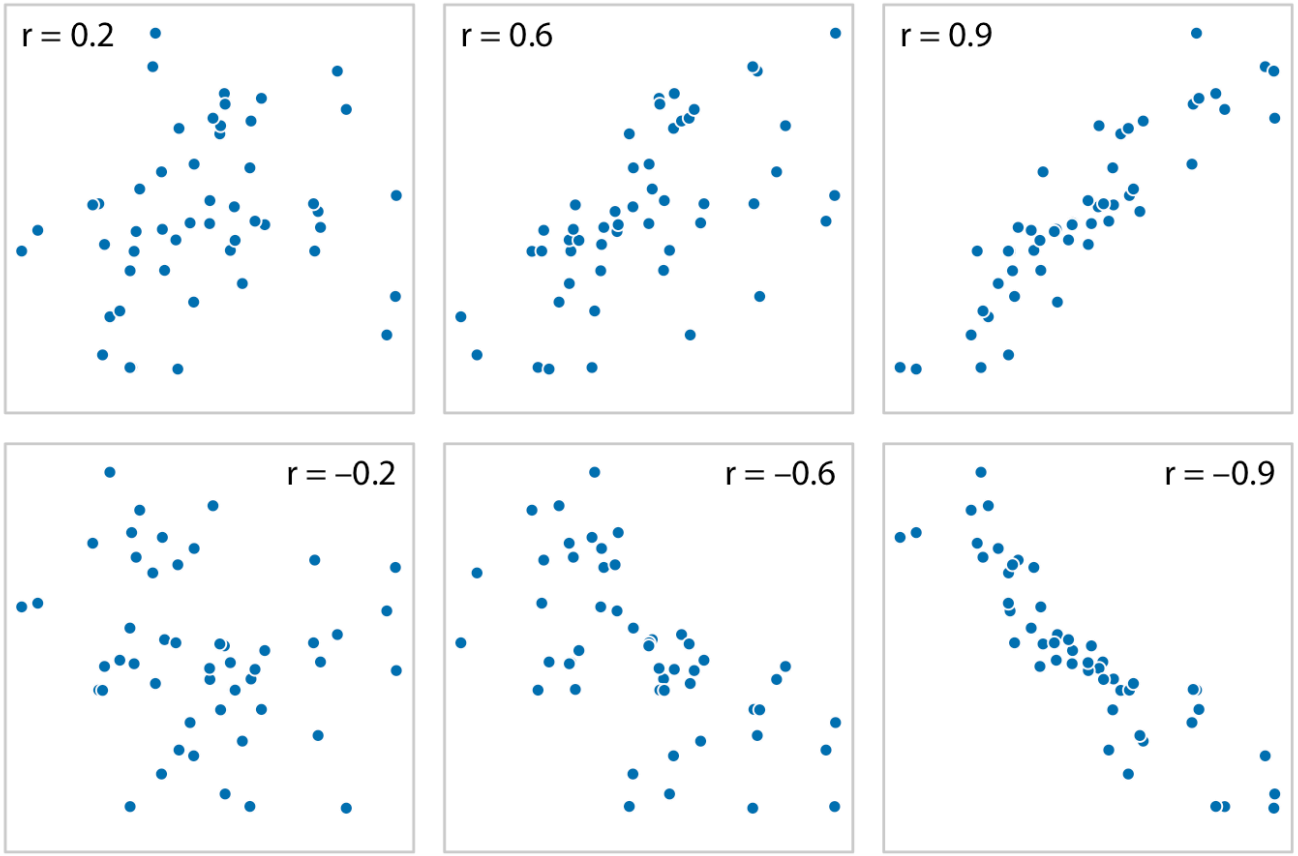


Correlograms

- **Concept**
  - **Correlation Coefficients**: Measure the strength of association (-1 to 1 range)
  - **Visualization**: Display as colored tiles or circles scaled by correlation magnitude

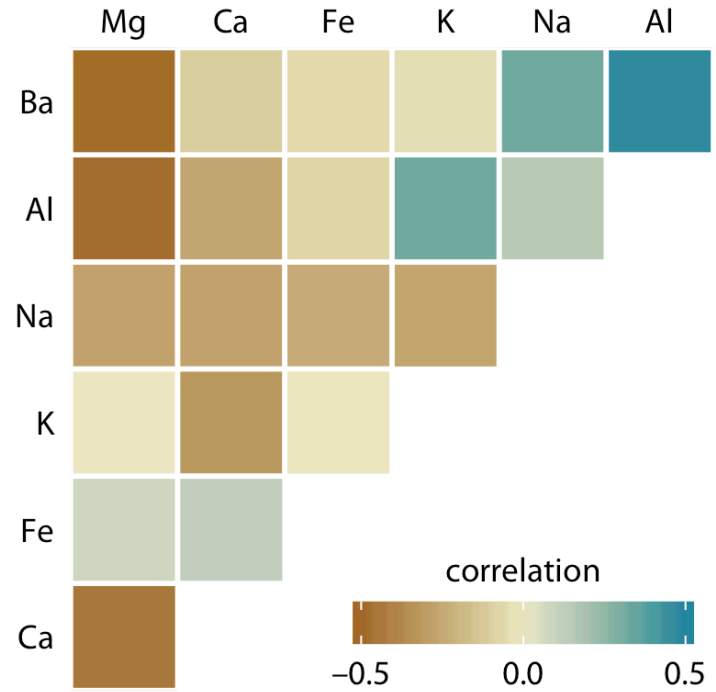
$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

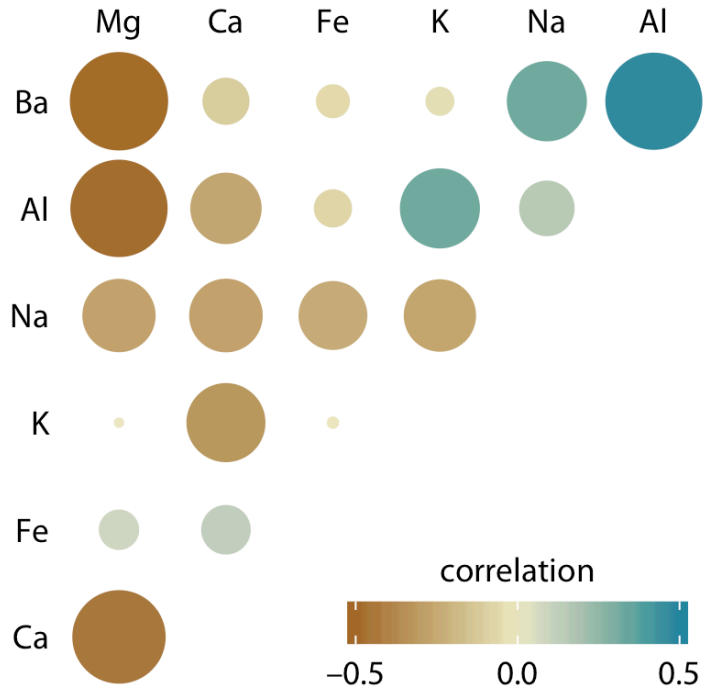
- **Example**
  - **Dataset**: Glass fragments with mineral oxide measurements
  - **Observation**: Trends in correlations (e.g., magnesium negatively correlated with other oxides)



Correlograms (continued)

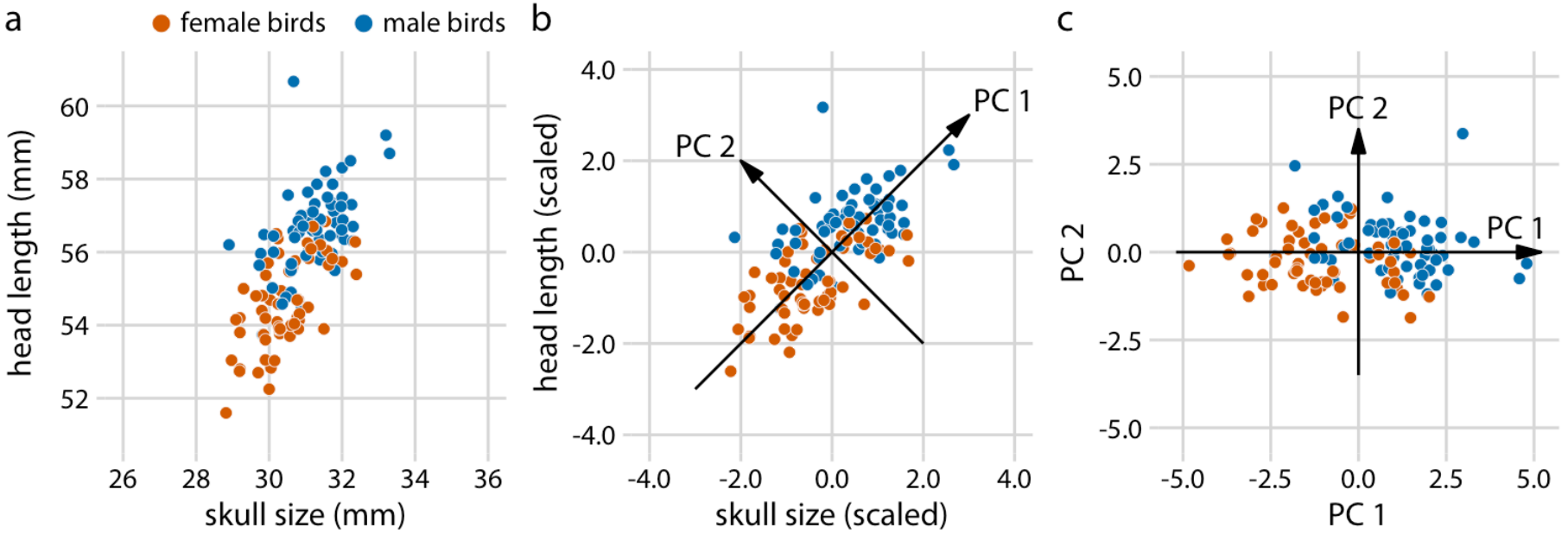
- **Improved Correlogram**
  - **Modification**: Circle size indicates correlation strength
  - **Benefit**: Low correlations are visually deemphasized





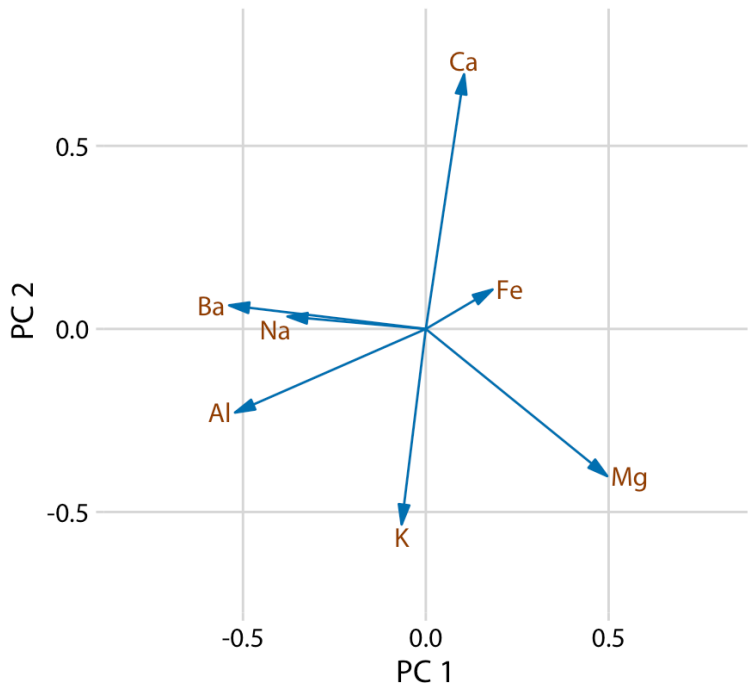
Dimension Reduction

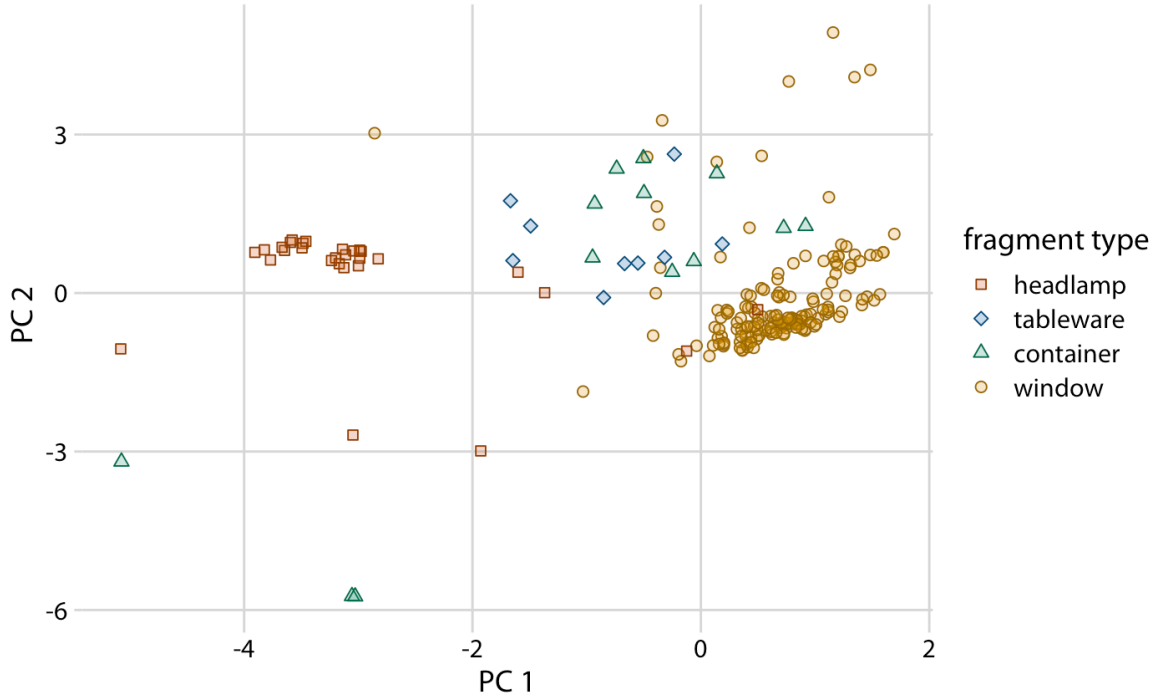
- **Concept**
  - **Goal**: Reduce dimensions while retaining key information
- **Technique: Principal Components Analysis (PCA)**
  - **Method**: Linear combinations of standardized variables
  - **Components**: Uncorrelated, ordered by variance captured



Dimension Reduction (continued)

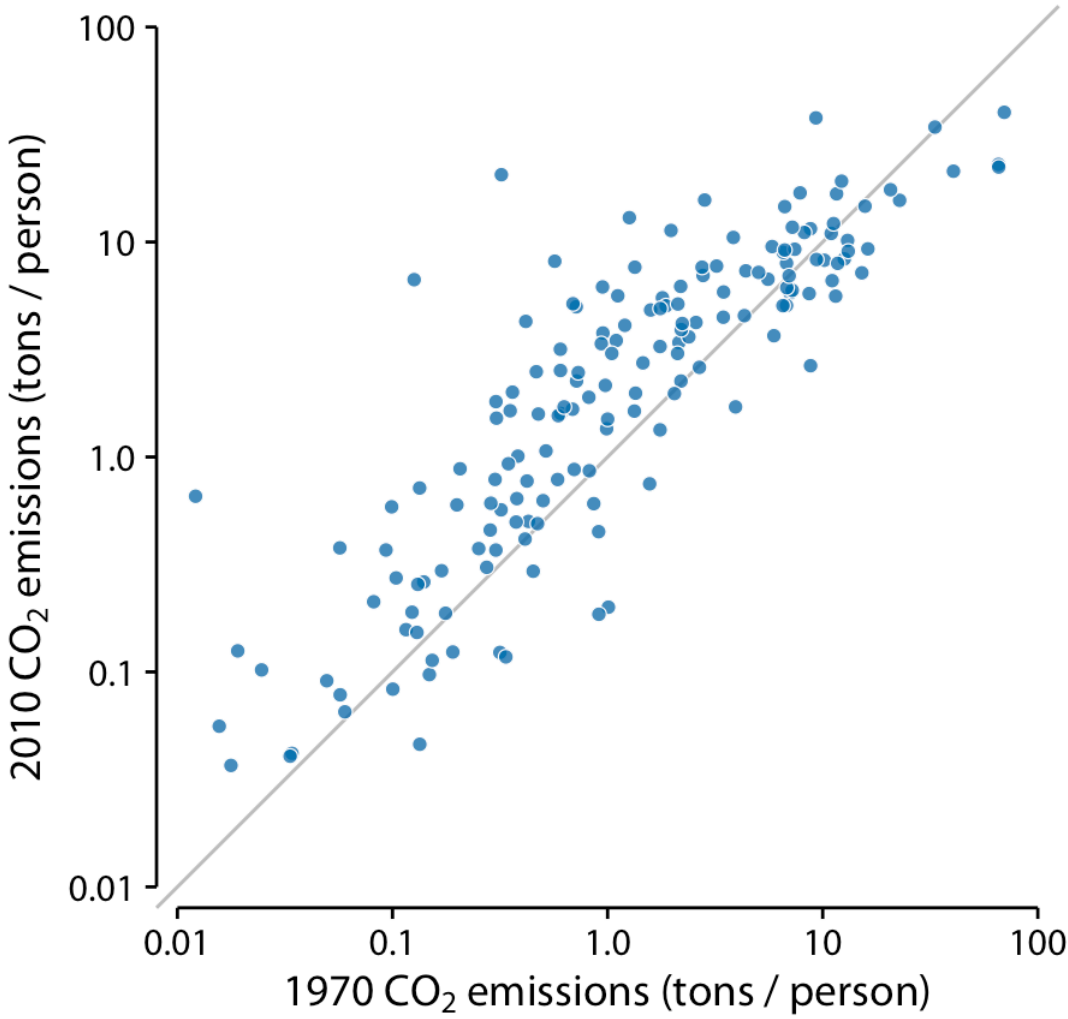
- **Example: Forensic Glass Dataset**
  - **PC Composition**: Variables contributing to first two PCs
  - **PC Space Projection**: Clustering of glass fragment types





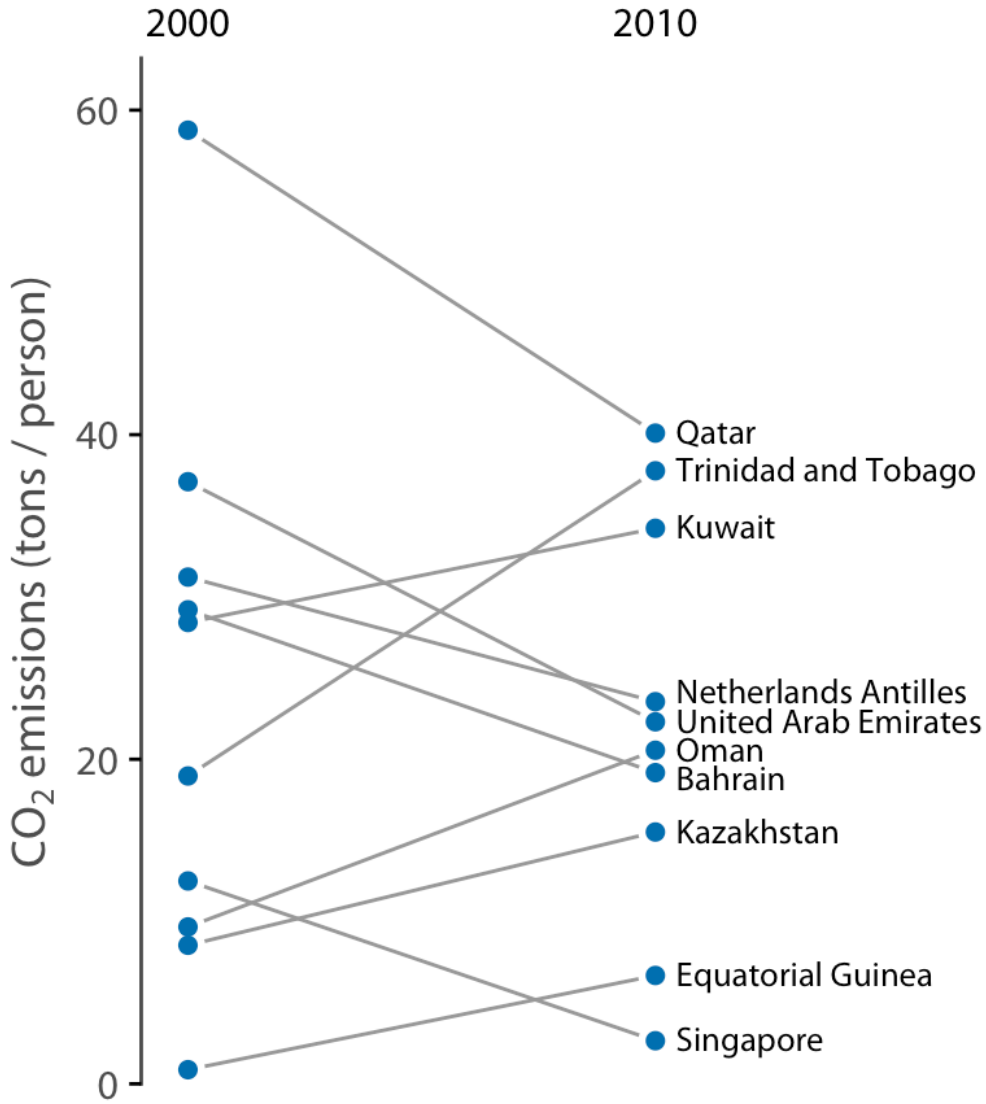
Paired Data

- **Concept**
- **Data Type**: Two or more measurements of the same quantity under different conditions
- **Visualization Techniques**
- **Scatter Plot with Diagonal Line**
  - **Example**: CO2 emissions per person (1970 vs. 2010)
  - **Observation**: Majority of countries show increased emissions



Paired Data (continued)

- **Slopegraph**
  - **Use Case**: Small number of observations, identity of individual cases
  - **Example**: CO2 emissions (2000 vs. 2010 for top 10 countries)



Paired Data (continued)

- **Extended Slopegraph**
  - **Comparison**: More than two time points (e.g., 2000, 2005, 2010)
  - **Observation**: Highlight changes over multiple intervals