Workflow for data analysis

**Step 1 — Explore the data (Descriptive Statistics)**

Goal: Get a full overview of what you have and where the gaps are.

**Actions:**

* Count frequencies of all categories:
  + type\_category, rim\_category, base\_category, size\_category, attributes
* Tabulate distributions by site (site\_name) if relevant.
* Calculate basic descriptive stats for numeric variables:
  + mean, median, range, SD for actual\_width, actual\_diameter, total\_height, etc.
* Visualizations:
  + Bar plots (we’ve done this)
  + Mosaic plots of categorical combinations (e.g. rim vs. base)
  + Histograms / density plots for continuous variables

**Outcome:** Understand sample sizes, category balance, and potential outliers.

## ****Step 2 — Assess relationships between categories****

Goal: See if certain vessel attributes tend to occur together.

**Actions:**

* Cross-tabulations (contingency tables) between:
  + type\_category × rim\_category
  + type\_category × base\_category
  + rim\_category × base\_category
* Chi-square tests for independence (to see if distributions are related)
* Cramer’s V for strength of association between categorical variables
* Multiple correspondence analysis (MCA) to visualize relationships between all categorical variables simultaneously

**Outcome:** Identify which morphological categories cluster together.

## ****Step 3 — Explore size variation and clustering****

Goal: See if vessels form size groups, and how these relate to categories.

**Actions:**

* Standardize numeric data (scale() in R)
* Perform cluster analysis (which you’ve already started):
  + hierarchical clustering (hclust)
  + k-means (to compare)
* Plot clusters and overlay type\_category or rim\_category
* Boxplots of measurements by type or size category

**Outcome:** Detect natural size groups and see if they align with traditional typology.

## ****Step 4 — Explore variation between sites (if multiple sites)****

Goal: Identify regional or contextual differences.

**Actions:**

* Compare category frequencies between site\_name groups (Chi-square)
* Use correspondence analysis (CA) to see site-category relationships
* Use ANOVA or Kruskal-Wallis to test if measurements differ between sites

**Outcome:** See whether vessel traditions differ between sites.

## ****Step 5 — Temporal analysis (if chronological data exists)****

Goal: Assess changes over time.

**Actions:**

* Group finds by time periods
* Plot changing frequencies of categories across periods
* Use logistic regression or multinomial regression if you want to model category probabilities over time

**Outcome:** See if certain vessel types emerge, peak, or disappear through time.

## ****Step 6 — Synthesis & Interpretation****

Goal: Link the statistics back to archaeological questions.

**Actions:**

* Combine results from Steps 2–5
* Interpret size groups, category co-occurrences, and spatial-temporal shifts in terms of production practices, use, and cultural traditions.