

Statistical Visualization Cheat Sheet

PRACTICAL LECTURE 3: DATA VISUALIZATION

Course: GGR2B10 / GGR02B2 with Dr JC Levin

STEP 1: Understanding Data Types

Before creating any graph, identify your data type to select the appropriate visualization method:

Data Type	Description	Example	Appropriate Graphs
Continuous Data (Quantitative)	Data that can take any value within a range	Temperature (20.6°C) Rainfall (713 mm)	Scatter plots, Line graphs, Histograms, Box plots
Discrete Data (Quantitative)	Data that can only take specific values (usually whole numbers)	Number of tourists (3500) Population count (44)	Bar graphs, Column charts, Line graphs (time series)
Categorical Data (Nominal)	Data that represents categories with no inherent order	Provinces (Western Cape) Cities (Johannesburg)	Bar graphs, Pie charts, Column charts
Ordinal Data (Categorical)	Categorical data with a natural order or ranking	Relationship strength (Weak, Moderate, Strong)	Bar graphs (ordered), Column charts (ordered)

Pro Tip: Use this table to identify your data type first, then select an appropriate graph.

Key Insight: Matching the right graph to your data type is essential for effective visualization!

STEP 2: Good Practices for All Graphs

Essential Formatting for Academic Graphs:

Regardless of graph type, these formatting practices ensure professional, consistent visualizations in academic contexts.

Layout and Labeling:

- **Axes:** Always label X and Y axes clearly with units
- **No chart title:** Use figure captions below the graph instead
- **Font color:** Ensure all text (axes, labels) is black
- **Font size:** Keep consistent (typically 10-12pt for academic work)
- **Legend:** Include only when multiple data series are present

✓ Good practices:

- Black text for all labels and axes
- No chart title (use figure caption instead)
- Clean background without gridlines
- Professional color scheme (avoid bright primary colors)
- Axes start at appropriate values (often zero)

Clean Design in Excel:

- 1 Remove gridlines: Click on gridlines and press Delete/Backspace
- 2 Change colors: Click on bars/lines/boxes, then right-click → Format Data Series
- 3 Format axes: Right-click axis → Format Axis → adjust scales and numbering
- 4 Remove border: Right-click chart area → Format Chart Area → No line
- 5 Apply consistent color scheme across all figures

✗ Poor practices:

- Missing axis labels or units
- Colored text that's hard to read when printed
- 3D effects or excessive design elements
- Inconsistent formatting across figures
- Too many decimal places in values

Academic Standard: Remember that scientific/academic graphs prioritize clarity and accuracy over visual appeal. Clean, simple designs with clear labeling are always preferred.

STEP 3: Scatter Plots

What Scatter Plots Show:

- Relationship between two continuous variables
- Direction of relationship (positive, negative, or none)
- Strength of relationship (how closely points cluster)
- Outliers

Excel Steps:

- 1 Select your data (two columns: X and Y variables)
- 2 Insert tab → Charts → Scatter
- 3 Add trendline: Right-click data series → Add Trendline
- 4 Display equation and R^2 : Check boxes in Trendline Options

Analyzing the Relationship:

R^2 (Coefficient of determination):

- Range: 0 to 1 (0% to 100%)
- Example: $R^2 = 0.75$ means 75% of variation in Y is explained by X

r (Correlation coefficient):

- Range: -1 to +1
- $r = +1$: Perfect positive correlation
- $r = -1$: Perfect negative correlation
- $r = 0$: No correlation

Calculating r from R^2 :

- Take the square root of R^2
- Look at the slope to determine if positive or negative

Interval	Relationship Level
0.80-1.000	Very strong
0.60-0.799	Strong
0.40-0.599	Moderate
0.20-0.399	Weak
0.00-0.199	Very weak

Remember: Always state whether it's a positive or negative relationship (e.g., "very strong positive relationship")

Your interpretation:

The correlation coefficient ($r = 0.95$) indicates a very strong positive relationship between temperature and rainfall...





STEP 4: Line Graphs

What Line Graphs Show:

- Trends over time
- Rate of change between consecutive points
- Fluctuations, cycles, or seasonality in data
- Comparisons of multiple series over the same sequence

Excel Steps:

- 1 Select data (including time labels)
- 2 Insert tab → Charts → Line
- 3 Format axes: Right-click axis → Format Axis
- 4 Add data labels if needed: Plus (+) button near chart

Adding a Secondary Axis:

- 1 Right-click on the data series you want on secondary axis (click directly on the line/data points you want to move)
- 2 Select "Format Data Series"
- 3 Choose "Secondary Axis"
- 4 Format both axes with appropriate titles and scales

Pro Tip: Use different colors and line styles for multiple series to improve readability. For monthly data, consider starting your X-axis from January rather than defaulting to the first row.

Common uses:

- Rainfall patterns over months across different regions
- Temperature trends over time
- Tourism numbers across years

STEP 5: Bar and Column Graphs

What Bar Graphs Show:

- Comparison of values across different categories
- Relative sizes or frequencies of different groups
- Composition of data (when using stacked bars)

Excel Steps:

- 1 Select your data (categories and values)
- 2 Insert tab → Charts → Column/Bar
- 3 For horizontal bars: Choose Bar chart option
- 4 For grouped bars: Ensure data has multiple value columns

Adding Error Bars:

- 1 Select the chart
- 2 Chart Design tab → Add Chart Element → Error Bars
- 3 More Error Bars Options → Custom → Specify Value
- 4 Select ranges containing your standard error values

When to use: Use bar graphs when comparing discrete categories, especially with longer category names (horizontal bars work better)

✓ **Good practice:** Sort bars by value (not alphabetically) for better comparison when no natural order exists

✗ **Bad practice:** Using 3D bars or too many categories (more than 10-12) which makes the chart hard to read

STEP 6: Box Plots

What Box Plots Show:

- Central tendency (median)
- Spread (IQR = Q3 - Q1)
- Skewness (position of median relative to quartiles)
- Outliers (points beyond whiskers)
- Range (minimum to maximum)

Excel Steps:

- 1 Organize data in columns by category
- 2 Insert tab → Charts → Statistical → Box & Whisker
- 3 Format outliers: Right-click → Format Data Series

Interpreting Box Plots:

- **Different whisker lengths:** Different spread in the tails of the distribution
- **Non-overlapping boxes:** Strong evidence of significant differences between groups
- **Position of median line:** If closer to Q1, positive skew; if closer to Q3, negative skew

Statistical significance: When boxes don't overlap, the groups are likely significantly different!

Analyze significance:

Which regions are significantly different?

Select...



What does this tell you?

The Northern Region experiences significantly higher temperatures than...



STEP 7: Pie Charts

What Pie Charts Show:

- Proportional relationship between categories
- How individual parts contribute to a whole (100%)

Excel Steps:

- 1 Select data (categories and values)
- 2 Insert tab → Charts → Pie
- 3 Add percentage labels: Plus (+) button → Data Labels → More Options → Check "Percentage"
- 4 Explode slices: Right-click slice → Format Data Point → Series Options → adjust "Point Explosion"

Best Practices:

- Limit to 5-7 slices maximum
- Combine smaller categories into "Other"
- Start largest slice at 12 o'clock position
- Use direct labeling rather than a separate legend

Warning: Only use pie charts when categories sum to 100% and when comparing proportions is more important than exact values

⚠ STEP 8: Common Mistakes to Avoid

Mistake	Problem	Better Approach
Wrong graph for data type	Misrepresents relationships	Match graph type to data type
Missing axis labels	Unclear what's being measured	Always label X and Y axes clearly
Improper scales	Distorts visual comparison	Start bar charts at zero; use appropriate scales
Too many data series	Cluttered and confusing	Limit to 3-4 series per graph when possible
Poor color choices	Hard to distinguish elements	Use contrasting colors; consider accessibility
Interpreting correlation as causation	Incorrect conclusions	Remember correlation \neq causation

✅ STEP 9: Final Checklist

Before submitting your graphs:

- ☐ Appropriate graph type for your data
- ☐ Clear title describing what the graph shows
- ☐ Labeled axes with units
- ☐ Legend if multiple data series
- ☐ Appropriate scales that don't distort data
- ☐ Readable font sizes and colors
- ☐ Statistical measures (R^2 , trend lines) where appropriate
- ☐ Error bars included if showing statistical significance
- ☐ Simplified design (removed gridlines, unnecessary elements)
- ☐ Proper interpretation in text (what the graph tells us)

Remember: A good graph tells a clear story about your data that anyone can understand at a glance!