

DATALAB CHEAT SHEET

Quick Reference for Data Visualization & Analysis

GETTING STARTED

```
r

# Load the package
library(datalab)

# Load your data
data(heart_data)    # Load sample data
data <- heart_data   # Assign to 'data'

# Look at your data
head(data)          # See first few rows
```

VISUALIZATION FUNCTIONS

Line Chart

When: Show relationship between two numbers over sequence

```
r

linechart(x_variable, y_variable)
linechart(age, cholesterol)
linechart(age, cholesterol, smooth = TRUE) # Add trend line
```

Scatter Plot

When: Show relationship between two numbers

```
r
```

```
scatterplot(x, y)
scatterplot(age, cholesterol)
scatterplot(age, cholesterol, add_line = TRUE)  # Add trend line
scatterplot(age, cholesterol, color = sex)      # Color by group
scatterplot(age, chol, color = sex, size = bp)  # Color + size
```

Histogram

When: Show distribution of one number

```
r

histogram(variable)
histogram(age)
histogram(age, bins = 20)          # Custom number of bins
histogram(cholesterol, show_density = TRUE) # Add density curve
histogram(age, color = "#E63946")  # Custom color
```

Boxplot

When: Show distribution of numbers

```
r

boxplot(variable)          # Single boxplot
boxplot(age)

boxplot(variable, group = category) # Compare groups
boxplot(age, group = sex)
```

Pie Chart

When: Show proportions of categories

```
r

piechart(category)          # Single pie chart
piechart(chest_pain_type)

piechart(category, filter)   # Multiple pie charts
piechart(chest_pain_type, sex) # One chart per sex
```

Heatmap

When: Show patterns between two categories

```
r

heatmap(category1, category2)
heatmap(chest_pain_type, sex)

heatmap(cat1, cat2, color_palette = "viridis") # Different colors
heatmap(cat1, cat2, show_values = FALSE)      # Hide numbers
```

Color palettes: "blue", "red", "green", "purple", "viridis"

STATISTICS FUNCTIONS

Descriptive Statistics

When: Get numerical summaries

```
r

descriptives(variable)      # Stats for one variable
descriptives(age)

descriptives(var1, var2, var3) # Multiple variables
descriptives(age, cholesterol, resting_bp)
```

Output: Mean, SD, Median, Min, Max, N

Frequency Table

When: Count categories

```
r

frequencies(category)      # Count each category
frequencies(chest_pain_type)

frequencies(category, sort = "desc") # Sort by count
```

QUICK TIPS

Using Variables

```
r

# Assign to short names for easy typing
x = nitrogen_content
y = crop_yield
linechart(x, y)

# Reuse same letters for different variables
a = chest_pain_type
b = sex
piechart(a, b)
```

Loading Your Own Data

```
r

# CSV files
data <- read.csv("myfile.csv")

# Excel files (requires readxl)
library(readxl)
data <- read_excel("myfile.xlsx")
```

Getting Help

```
r

?linechart  # Help for linechart function
?piechart   # Help for any function
```

CHOOSING THE RIGHT FUNCTION

Your Goal	Use This
Relationship between two numbers	<code>scatterplot(x, y)</code> or <code>linechart(x, y)</code>
See trend in relationship	<code>scatterplot(x, y, add_line = TRUE)</code>
Distribution of one number	<code>histogram(var)</code> or <code>boxplot(var)</code>
Compare number across groups	<code>boxplot(var, group = category)</code>
Show category proportions	<code>piechart(category)</code>

Your Goal	Use This
Compare categories across groups	<code>piechart(category, filter)</code>
Patterns between two categories	<code>heatmap(cat1, cat2)</code>
Get exact statistics	<code>descriptives(var)</code>
Count categories	<code>frequencies(category)</code>

COMMON ERRORS & FIXES

Error Message	Problem	Fix
<code>Variable 'x' not found</code>	Typo or wrong variable	Check spelling
<code>No data frame found</code>	Data not loaded	Run <code>data <- your_data</code>
Plot looks strange	Too many categories	Try <code>frequencies()</code> first

INTERPRETATION GUIDE

Boxplot

- **Middle line** = median (middle value)
- **Box** = where middle 50% of data lives
- **Lines (whiskers)** = range of typical values
- **Dots** = unusual values (outliers)

Pie Chart

- **Bigger slice** = more frequent
- **Percentages** = proportion of total
- Compare slices within one pie, or same slices across pies

Line Chart

- **Line goes up** = positive relationship
- **Line goes down** = negative relationship
- **Line is flat** = no relationship
- **Smooth line** = overall trend

Descriptive Statistics

- **Mean** = average value
 - **SD** = how spread out (bigger = more variation)
 - **Median** = middle value
 - **N** = sample size (how many data points)
-

EXAMPLE WORKFLOW

```
r

# 1. Load and explore
library(datalab)
data(heart_data)
data <- heart_data
head(data)

# 2. Visualize distributions
histogram(age, show_density = TRUE)
boxplot(cholesterol, group = sex)

# 3. Explore relationships
scatterplot(age, cholesterol, color = sex, add_line = TRUE)

# 4. Look at categories
piechart(chest_pain_type, sex)
heatmap(chest_pain_type, sex)

# 5. Get statistics
descriptives(age, cholesterol)
frequencies(chest_pain_type)

# 6. Interpret and share your findings!
```

Need More Help? Type `vignette("datalab-intro")` for detailed examples

Questions? Contact your workshop facilitator