

*# MODULE 1 : DATA EXPLORATION*



# DATA-VISUALIZATION

Interpreting visuals to uncover key takeaways



Club Informatique & Télécom  
Data Cell

# Data Visualization

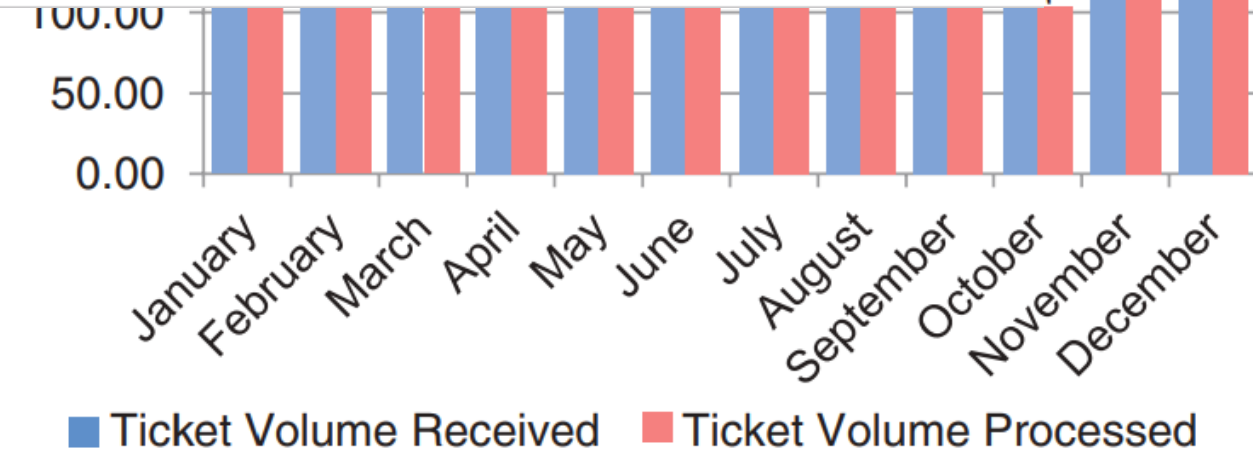
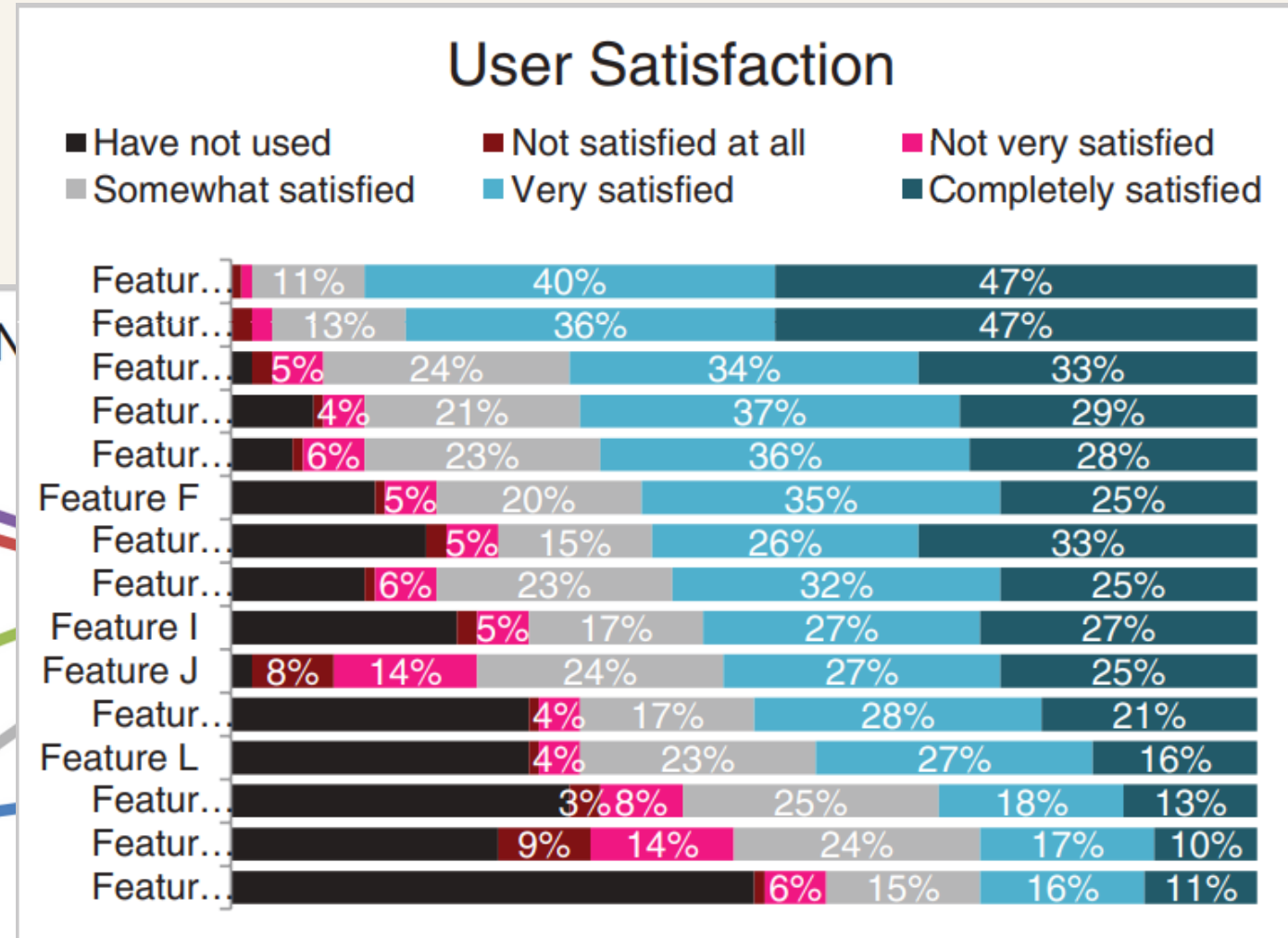
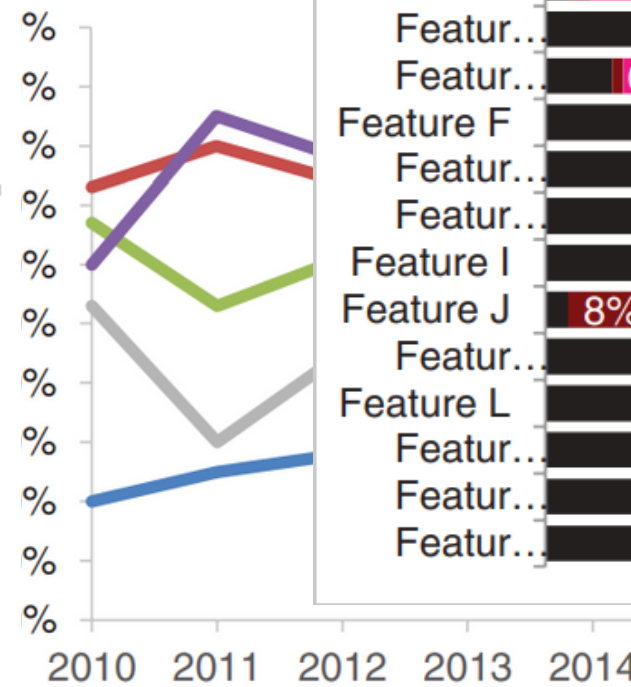
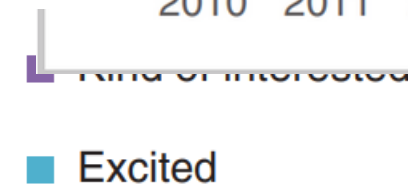
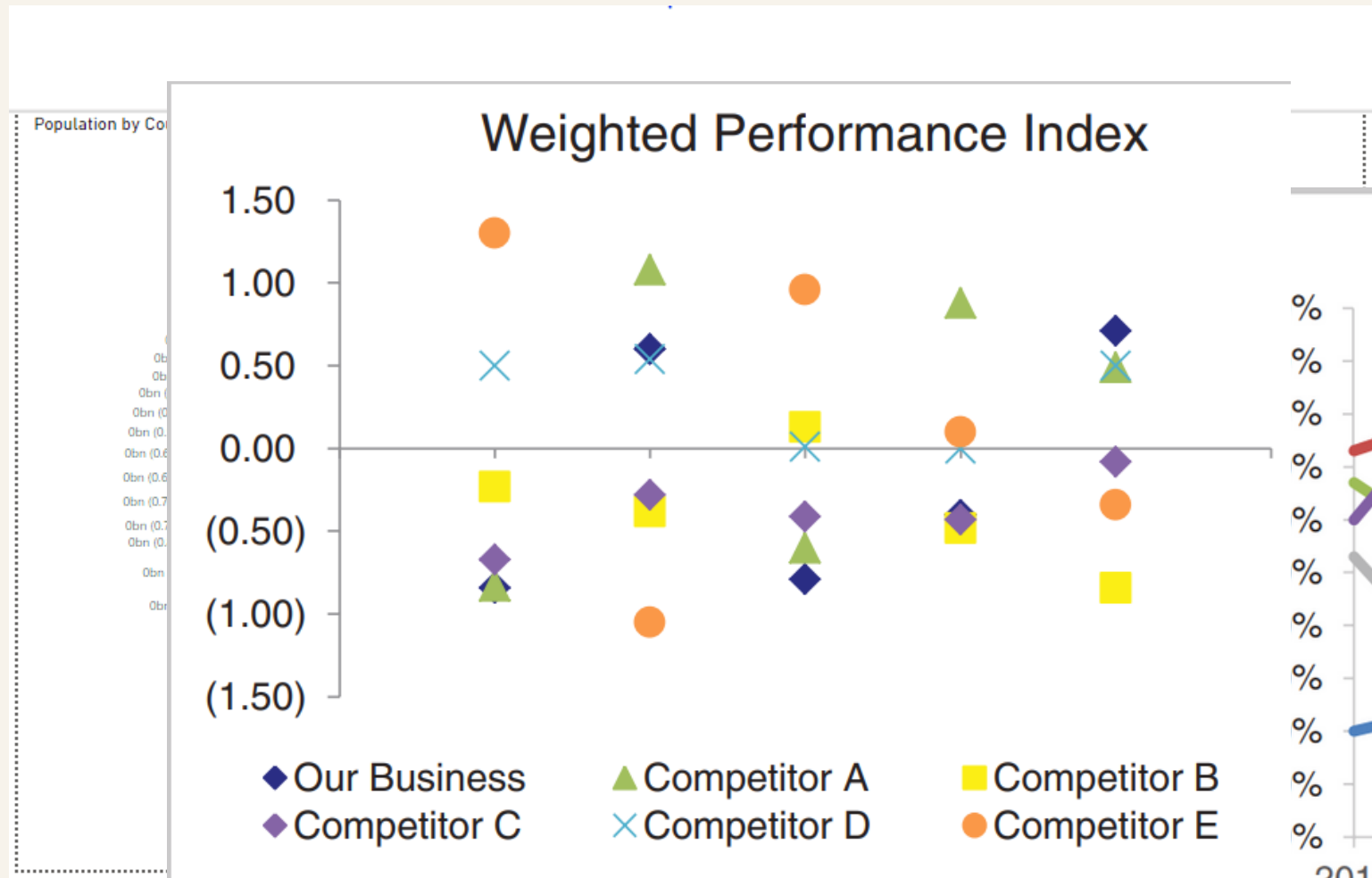
- Graphs, graphs, GRAPHS!
- What defines a good graph ?
- What are the graphs used for ?
- Main graph types
- matplotlib & seaborn main elements
- Today's lab



What's the difference between  
**Data & Information**  
?



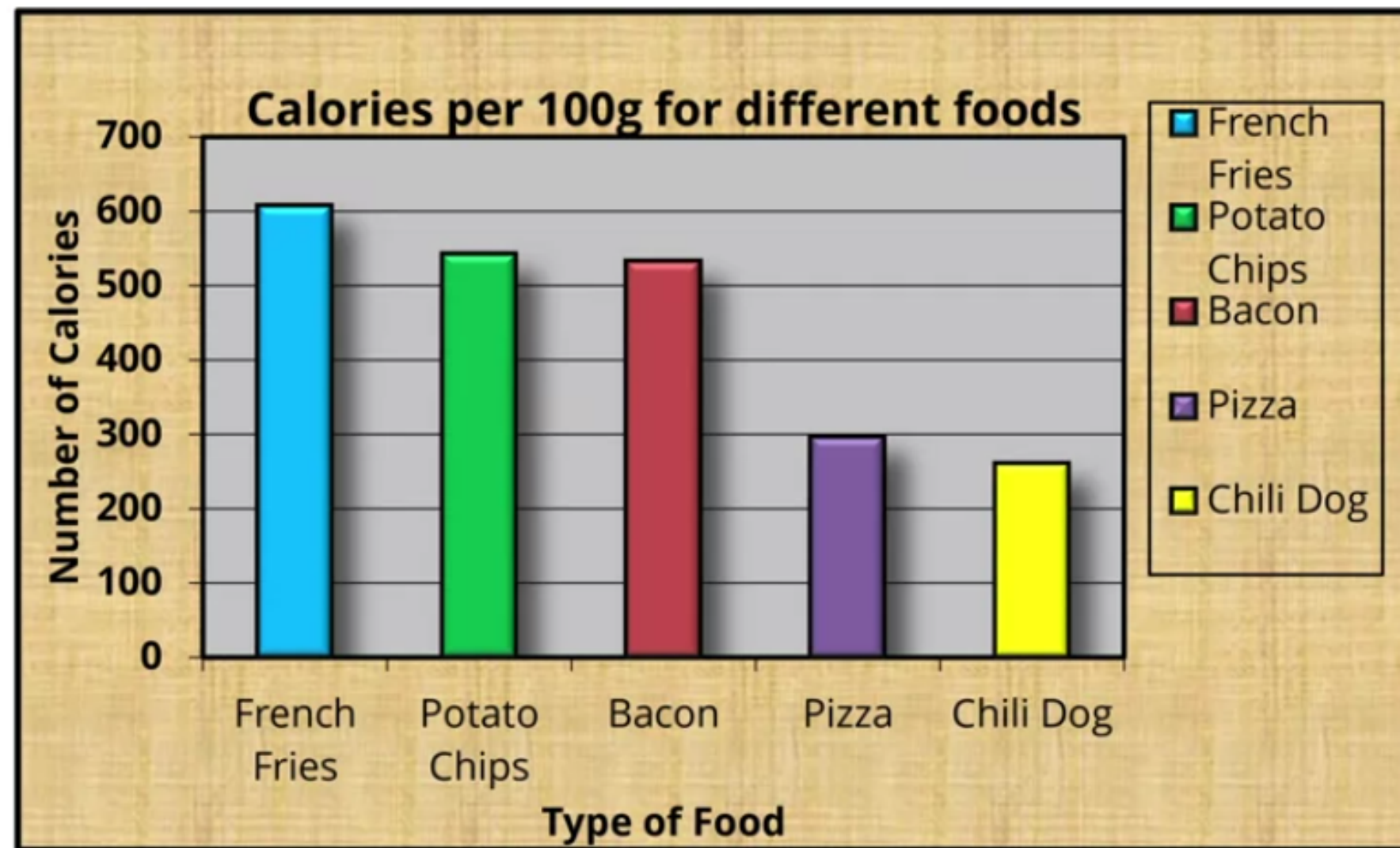
# Graphs, graphs, GRAPHS!



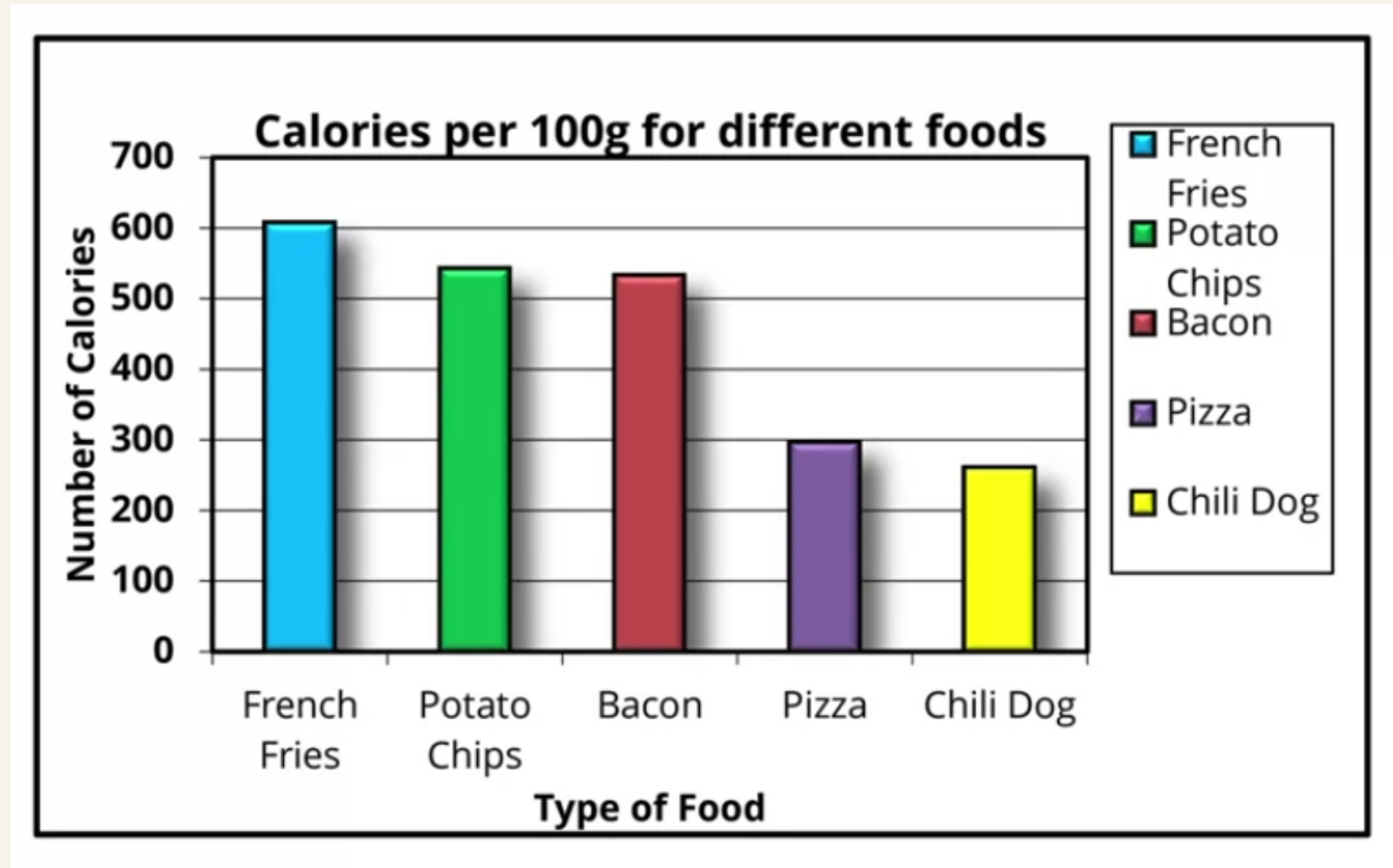
What defines  
a good graph  
?



# Data-Ink Ratio



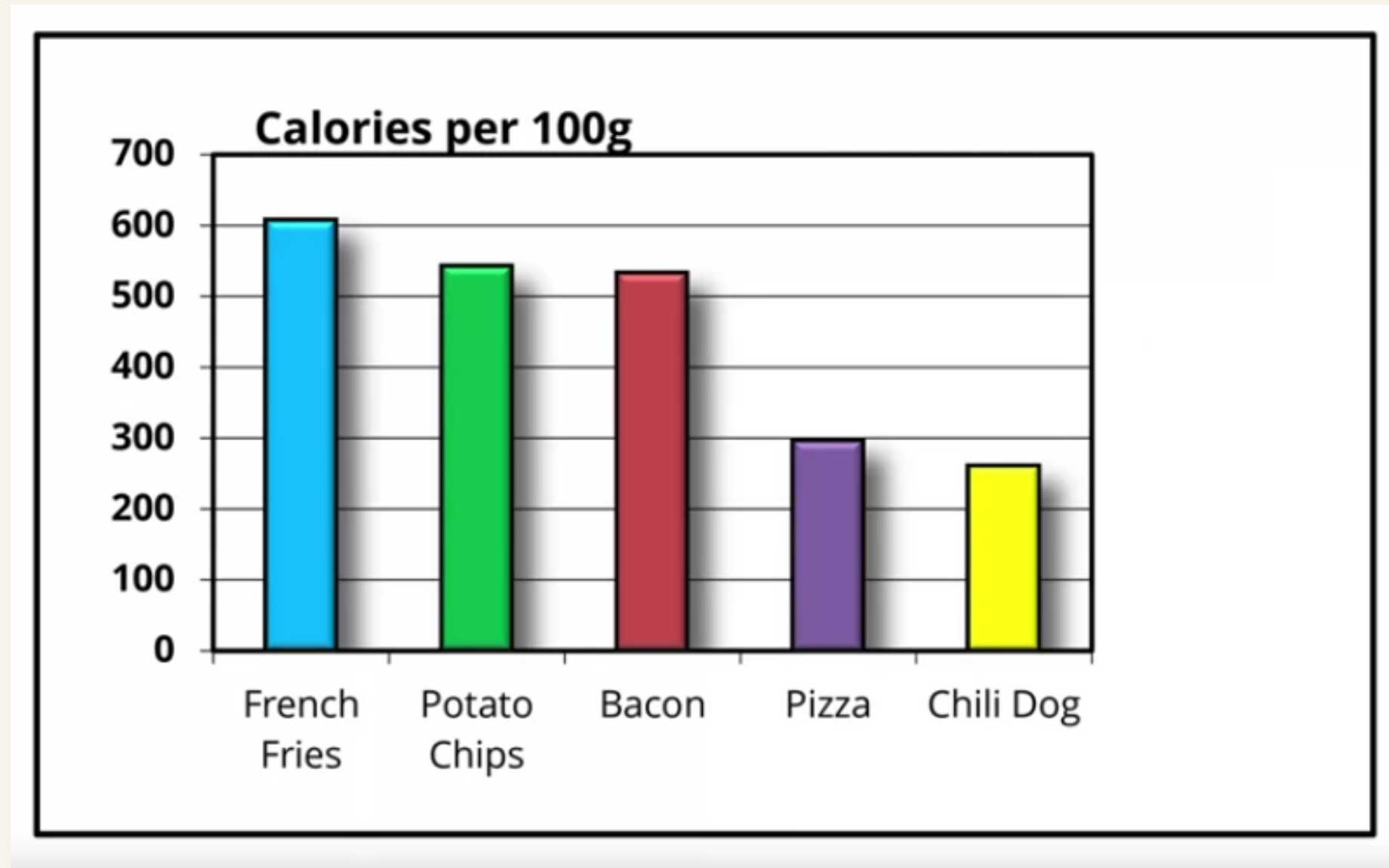
# Data-Ink Ratio



Background removed



# Data-Ink Ratio

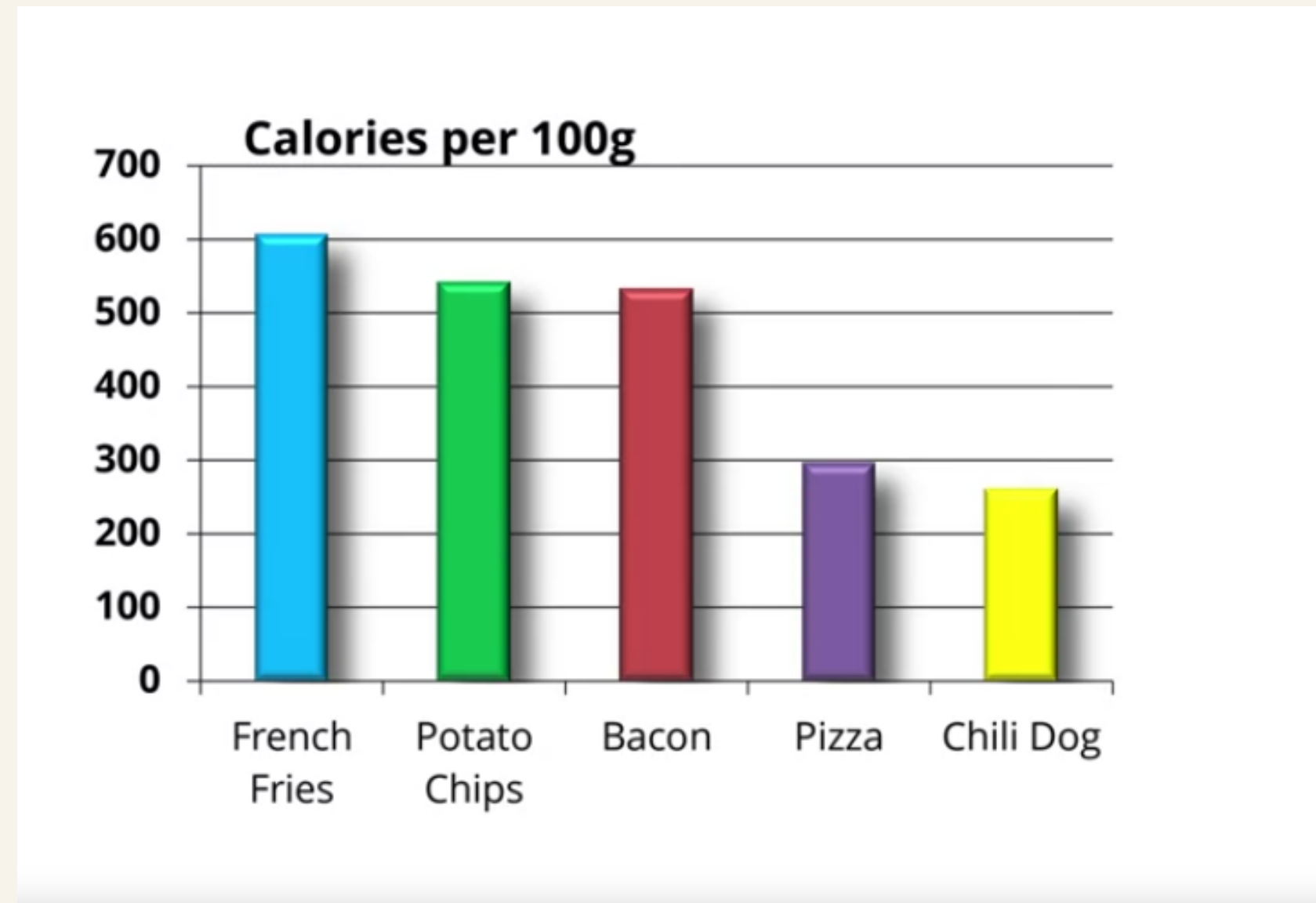


Redundant labels removed





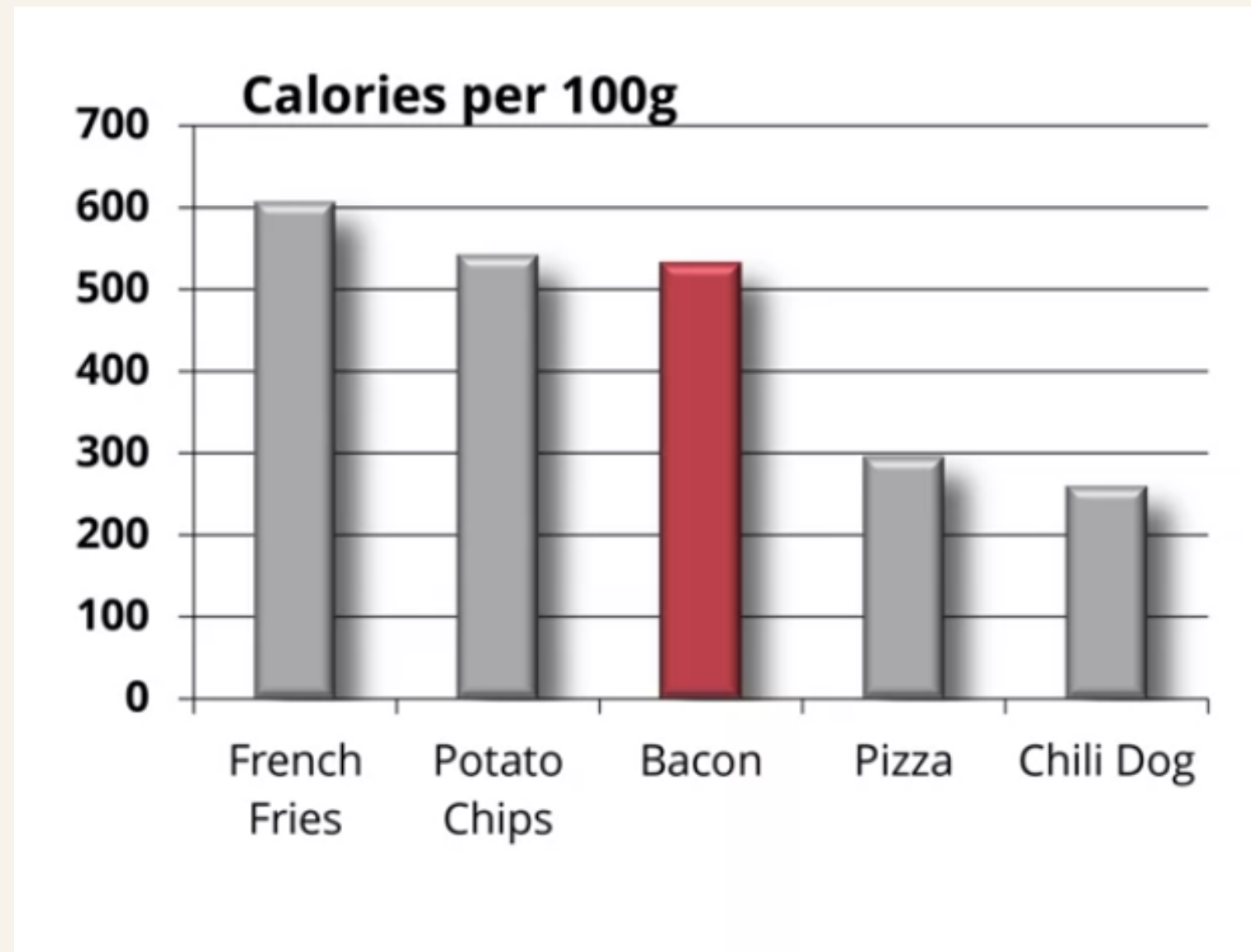
# Data-Ink Ratio



Borders removed



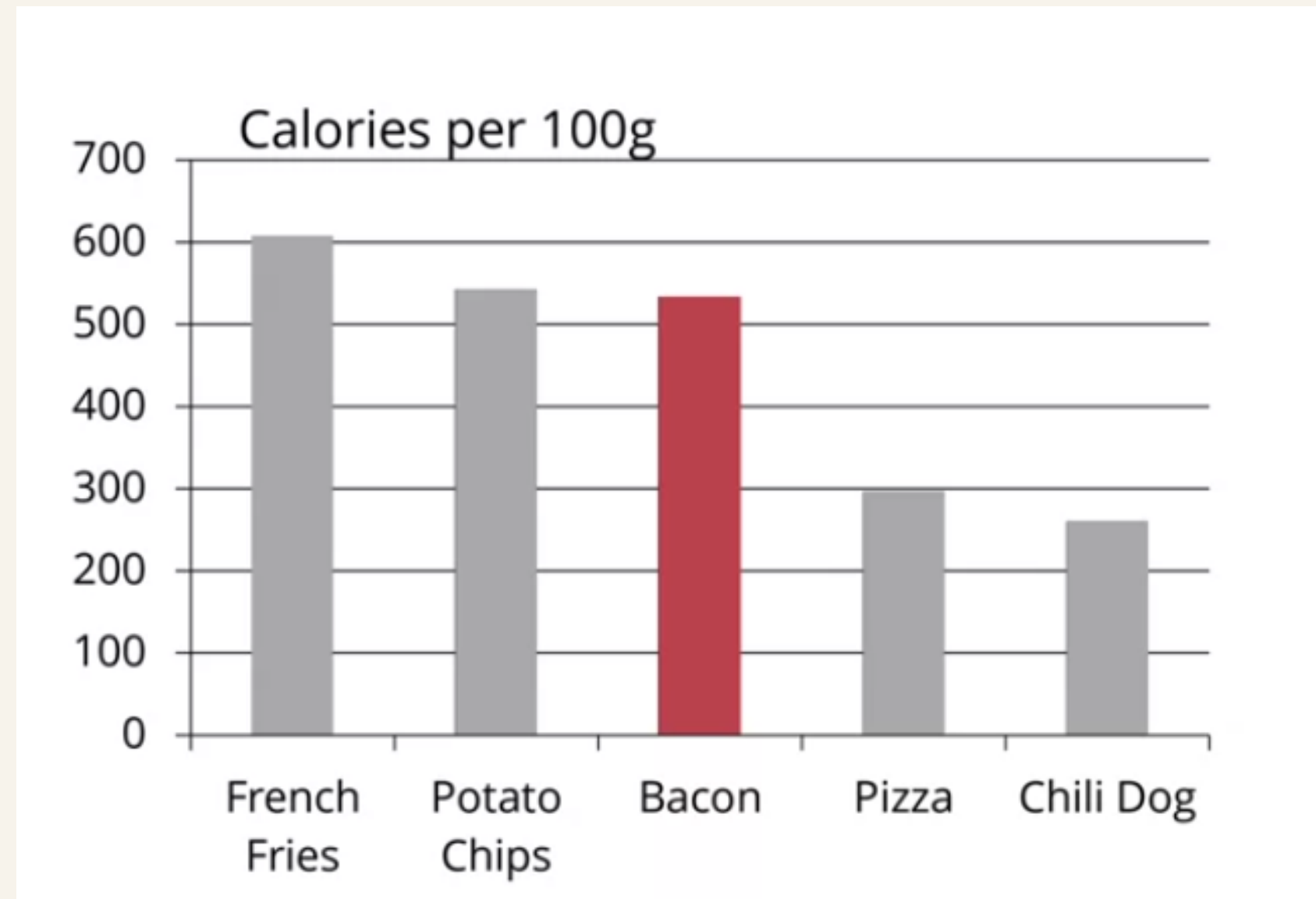
# Data-Ink Ratio



Colors reduced



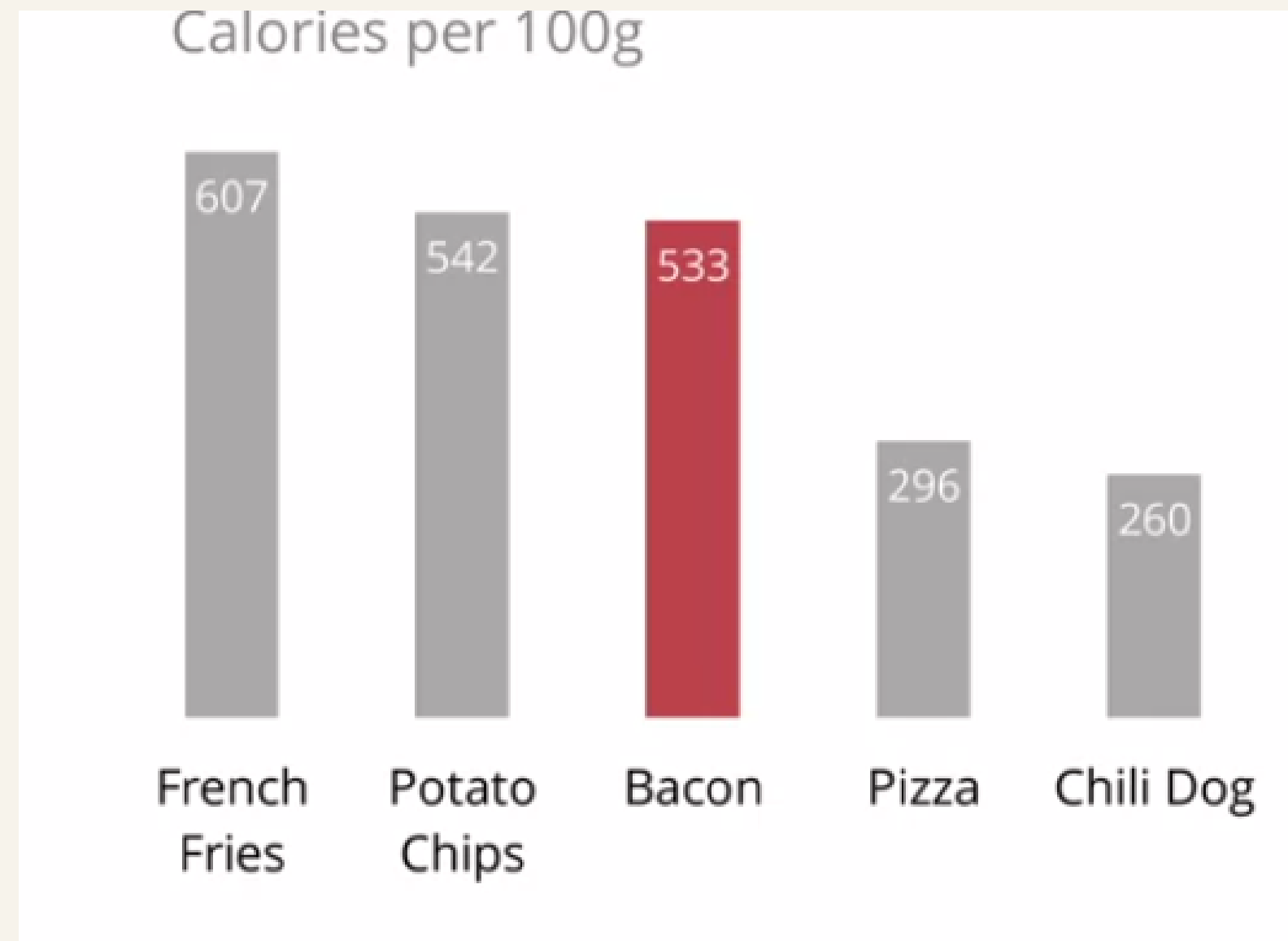
# Data-Ink Ratio



Special effects removed



# Data-Ink Ratio



Direct labels



# What are the graphs used for ?



Comparison



Distribution

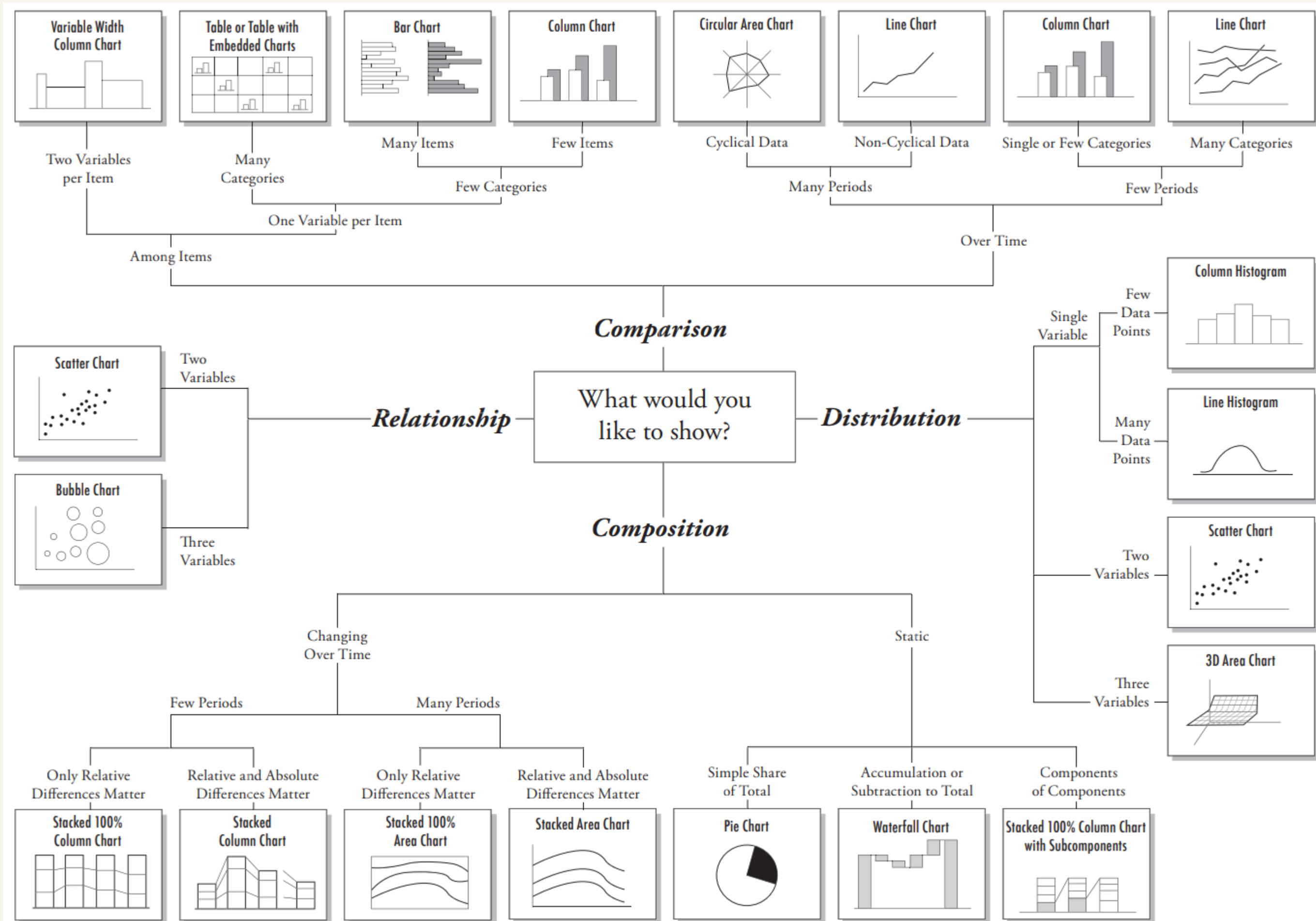


Proportion



Relationship Data over time





# Main graph types



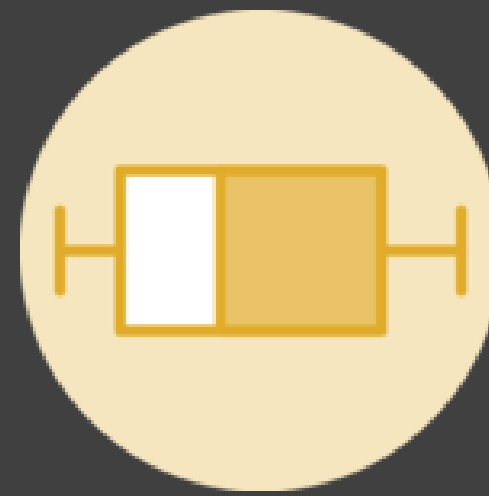
BARPLOT



HISTOGRAM



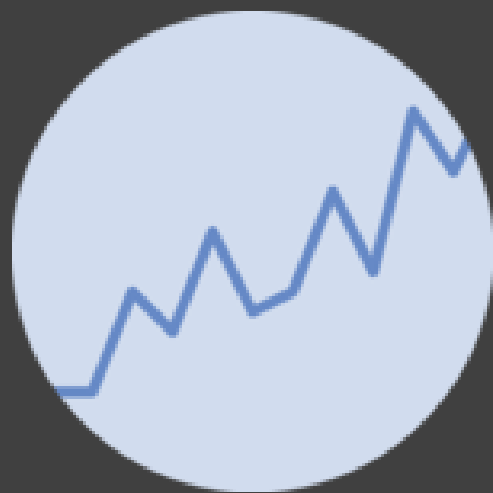
DENSITY PLOT



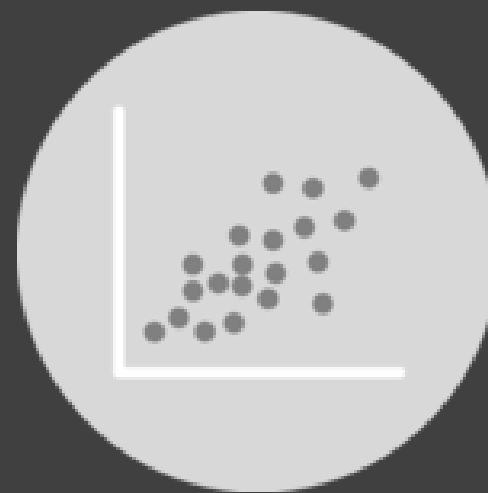
BOXPLOT



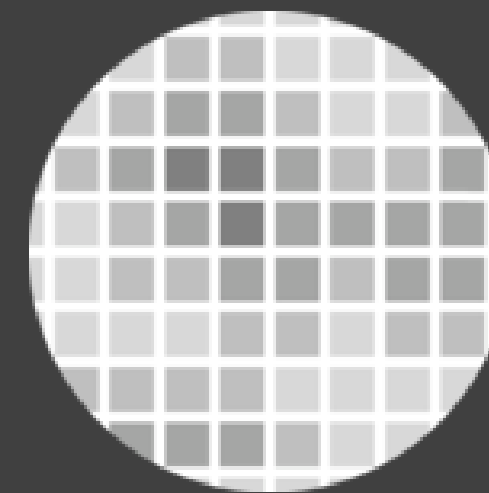
VIOLIN



LINE CHART



SCATTERPLOT



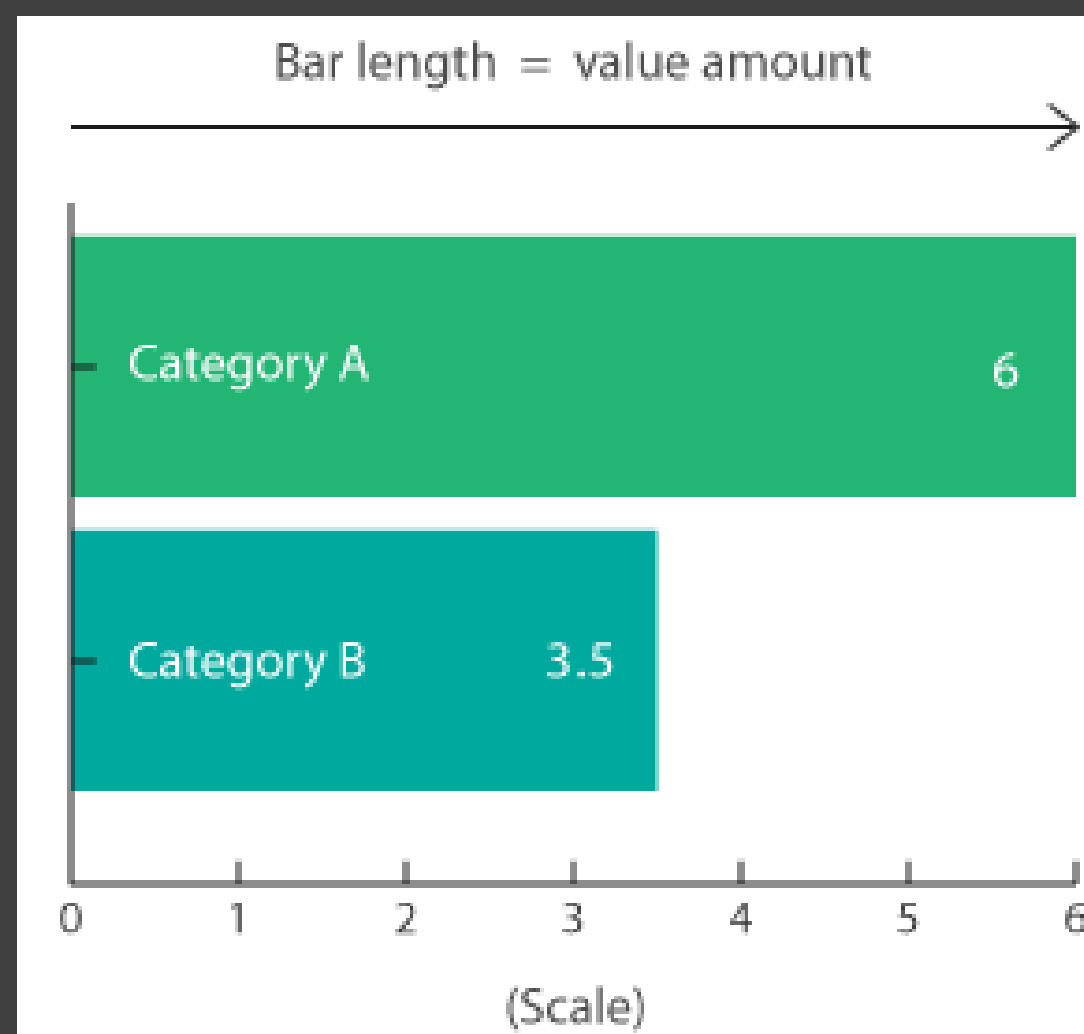
HEATMAP





# BARPLOT

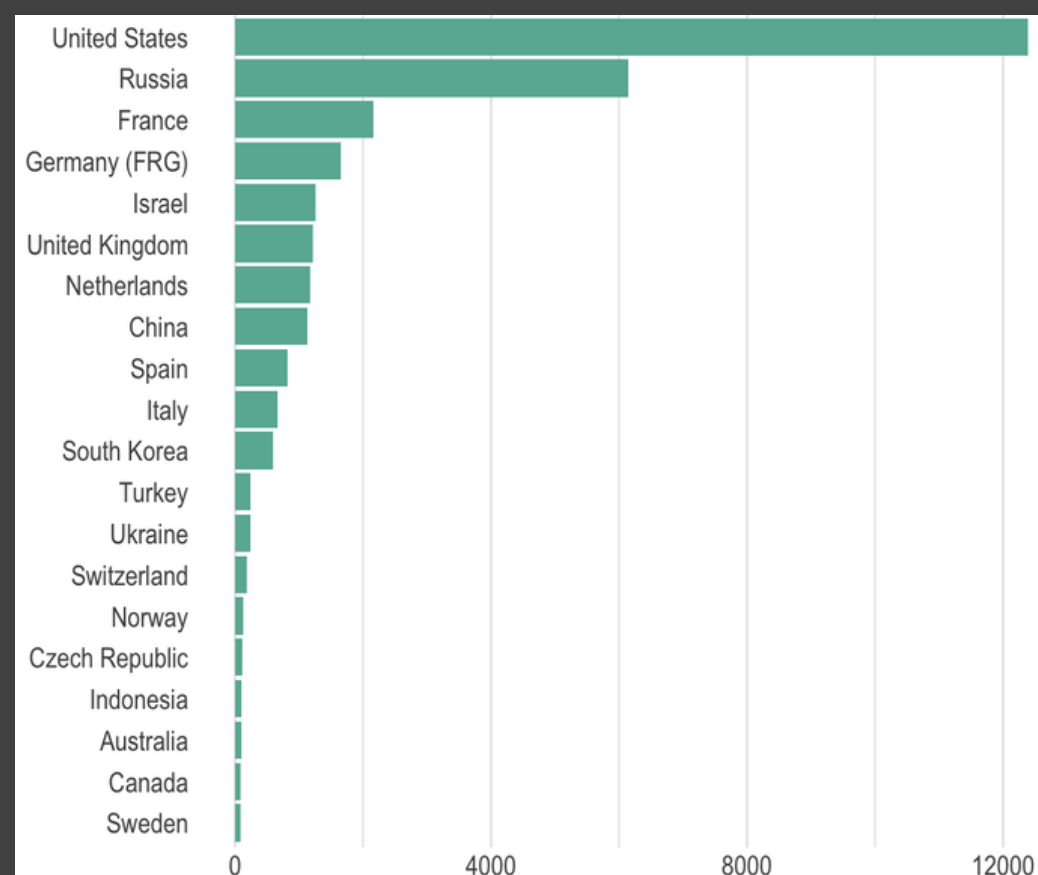
Relationship between **Numeric** & **Categorical** variables



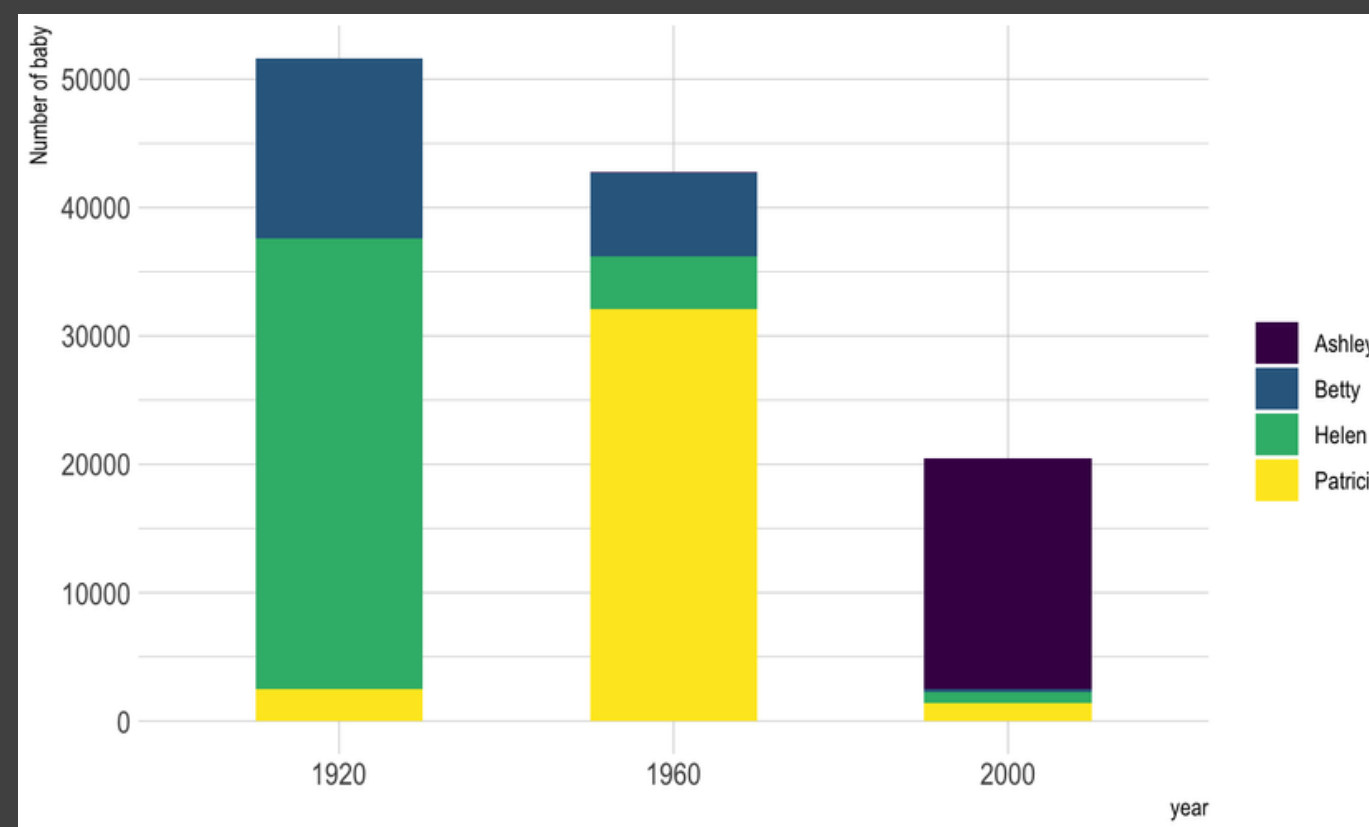




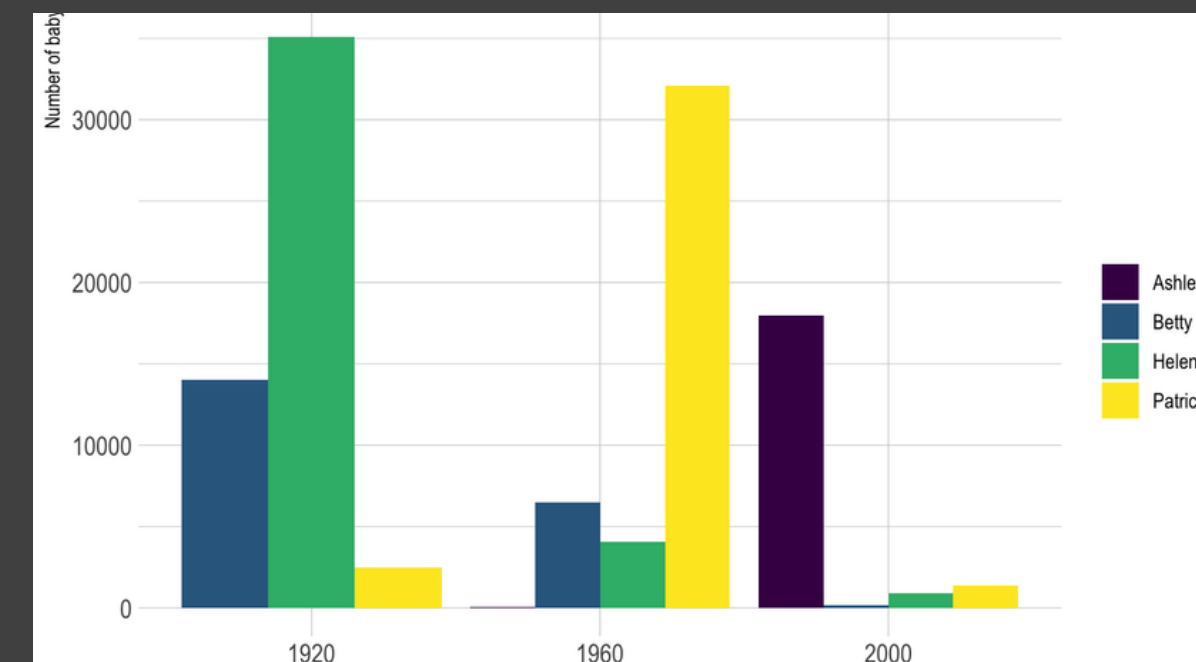
# Bar plot



Vertical bar plot



Stacked bar plot



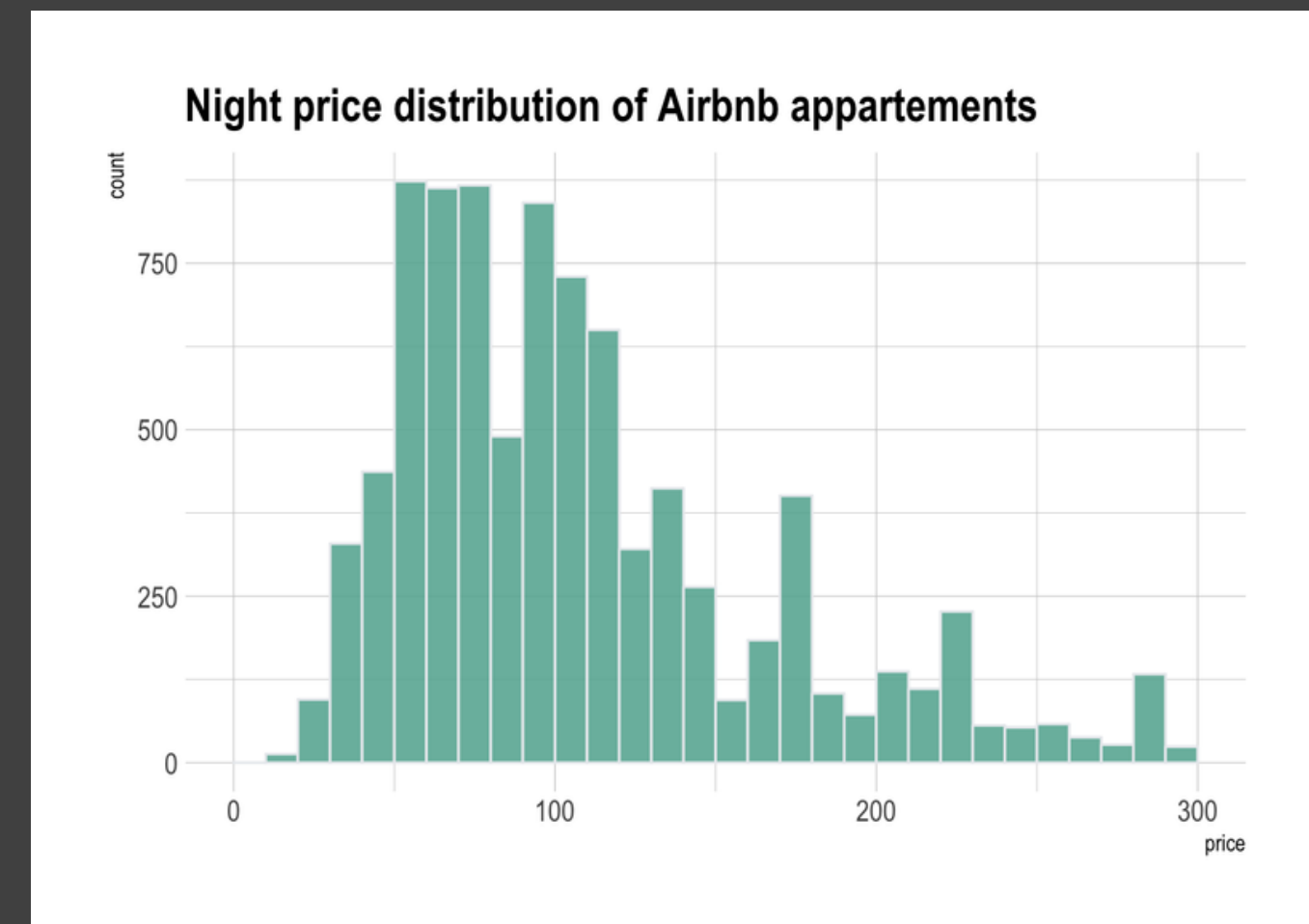
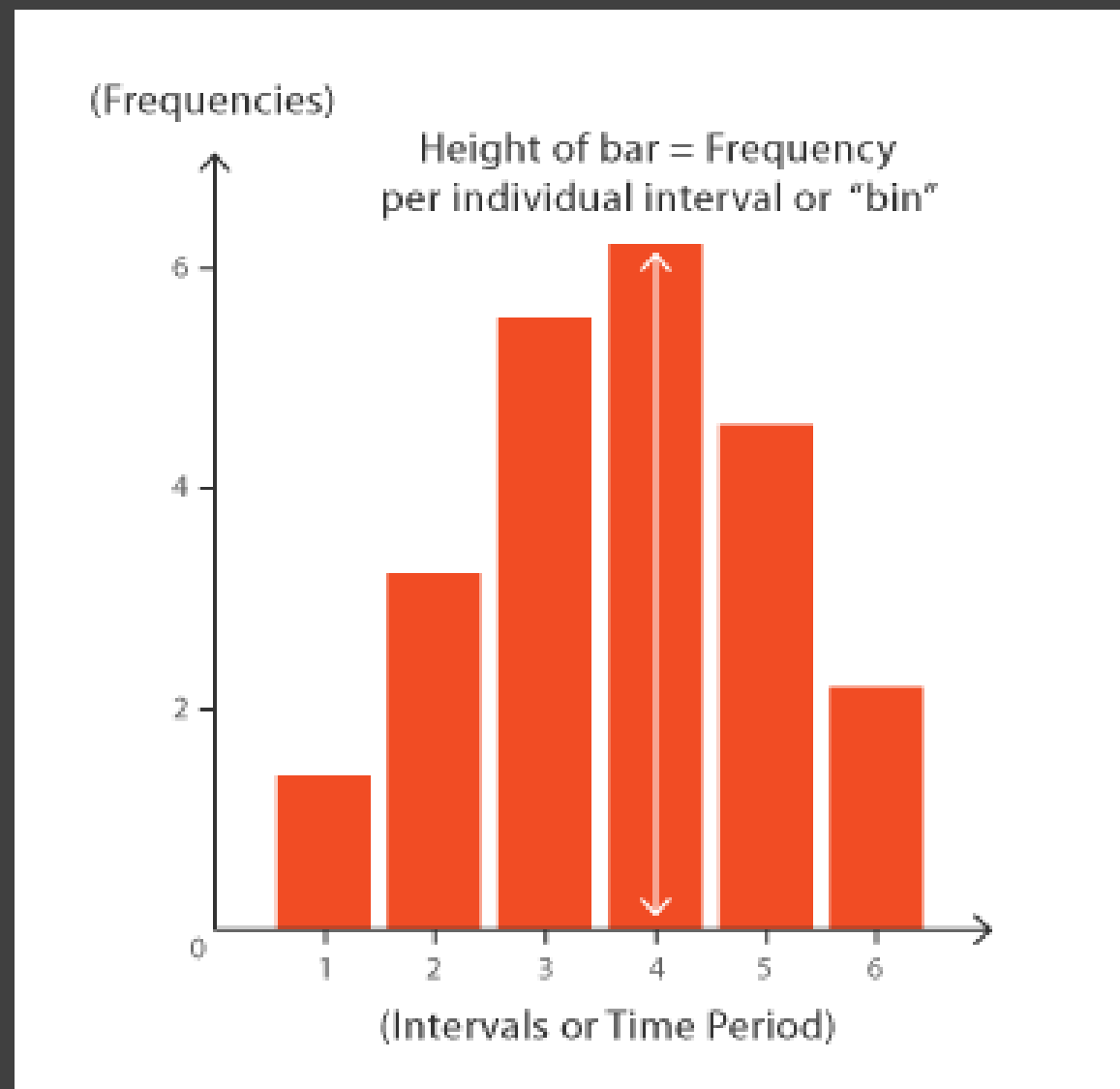
Grouped bar plot





# HISTOGRAM

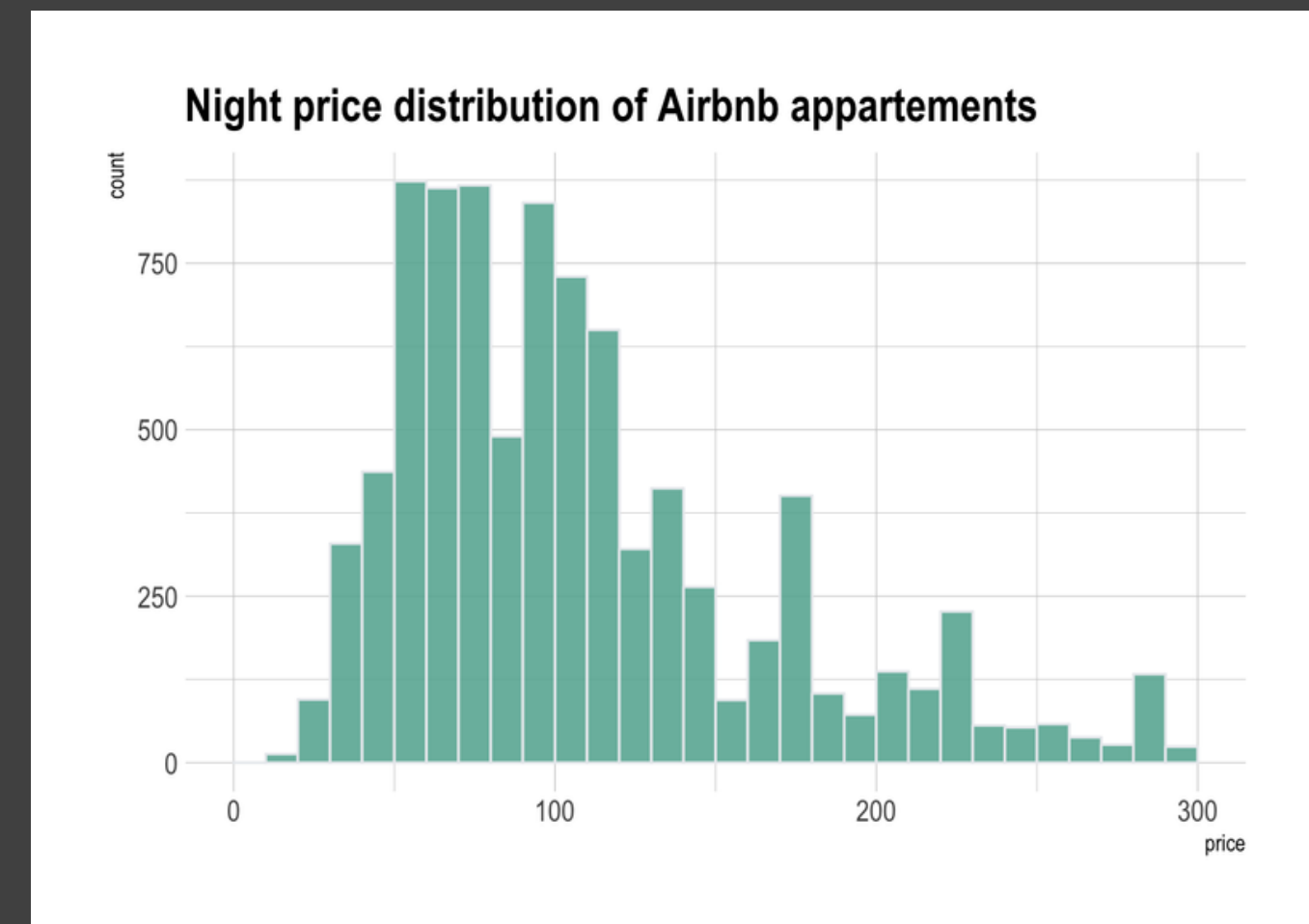
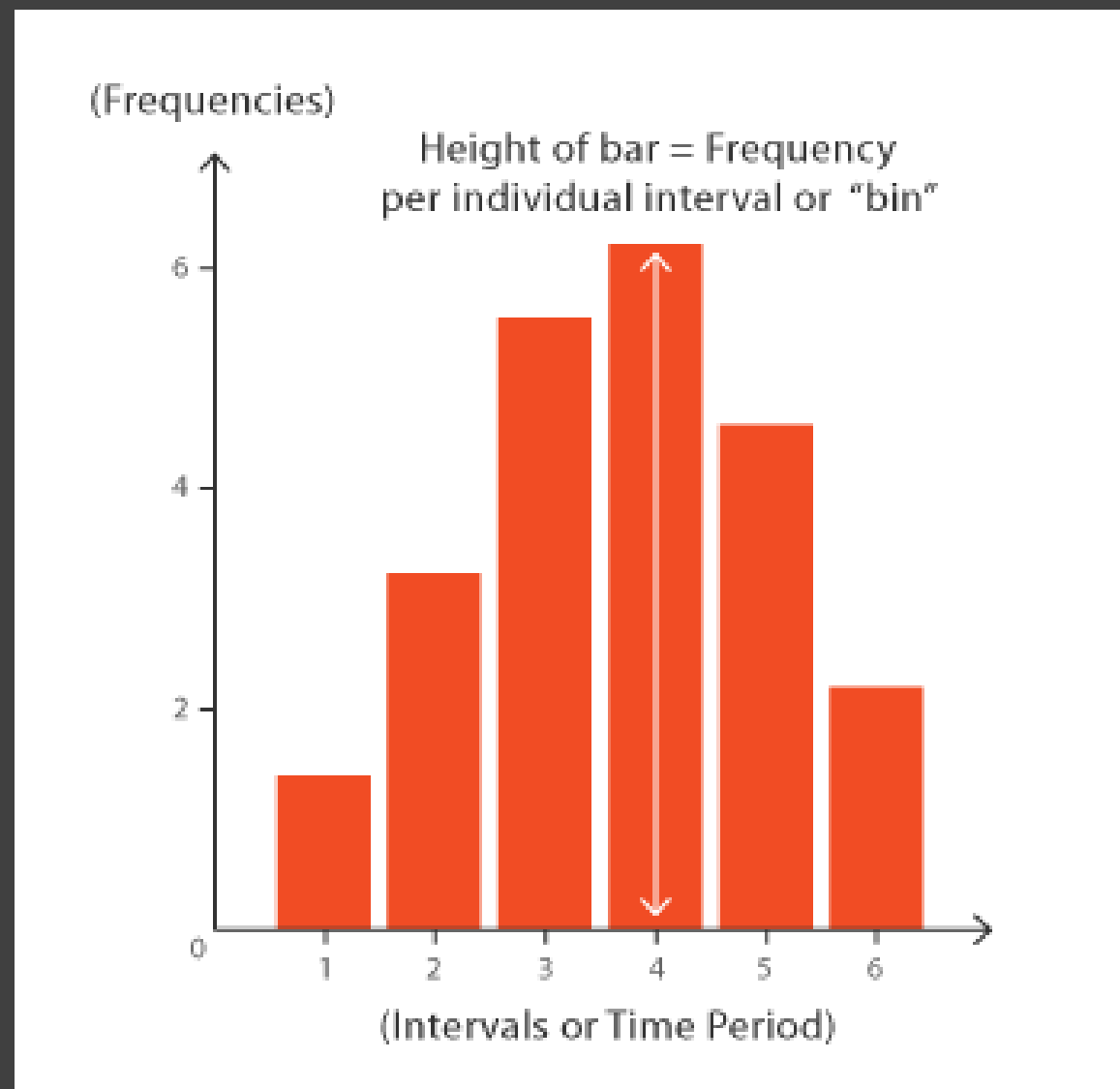
Relationship between **Numeric** variables, displays frequency





# HISTOGRAM

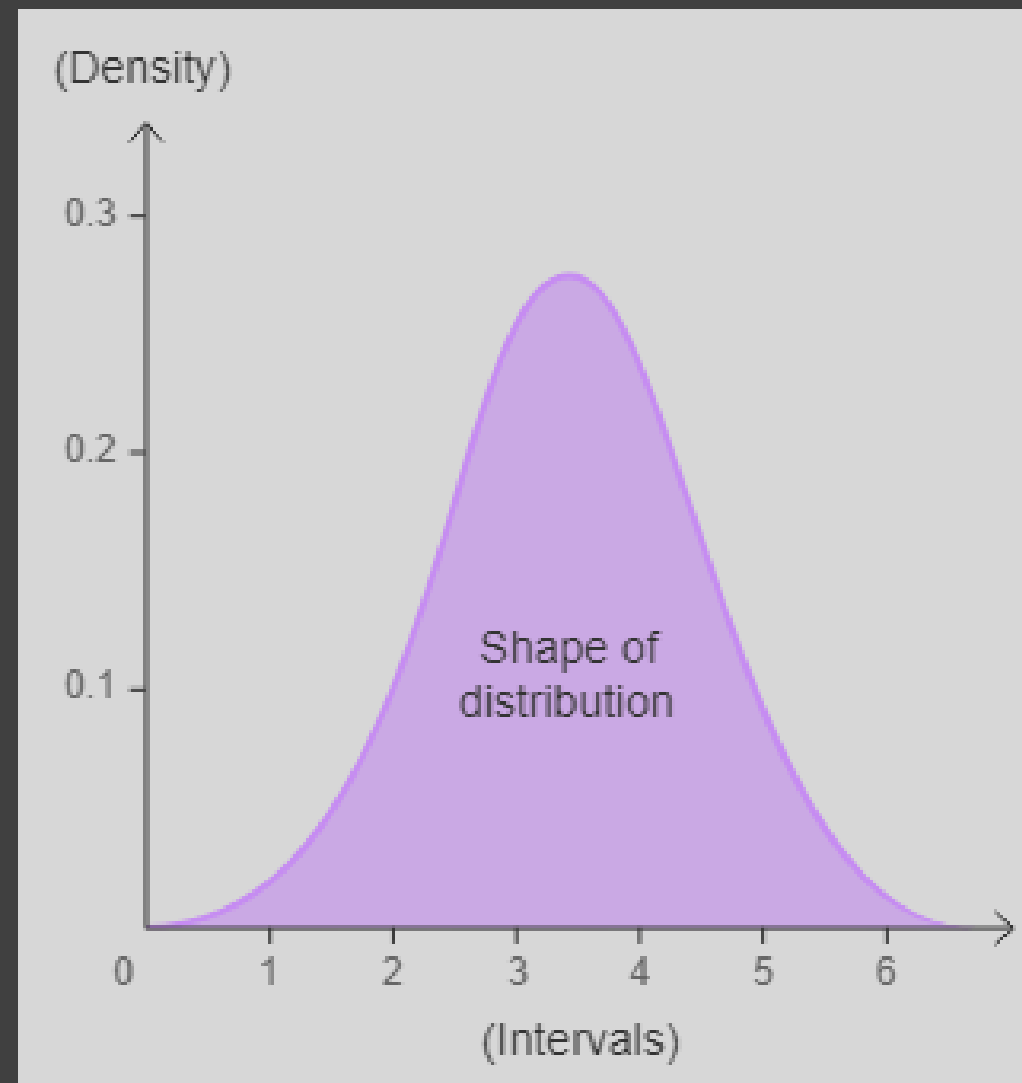
Relationship between **Numeric** variables, displays frequency





# DENSITY PLOT

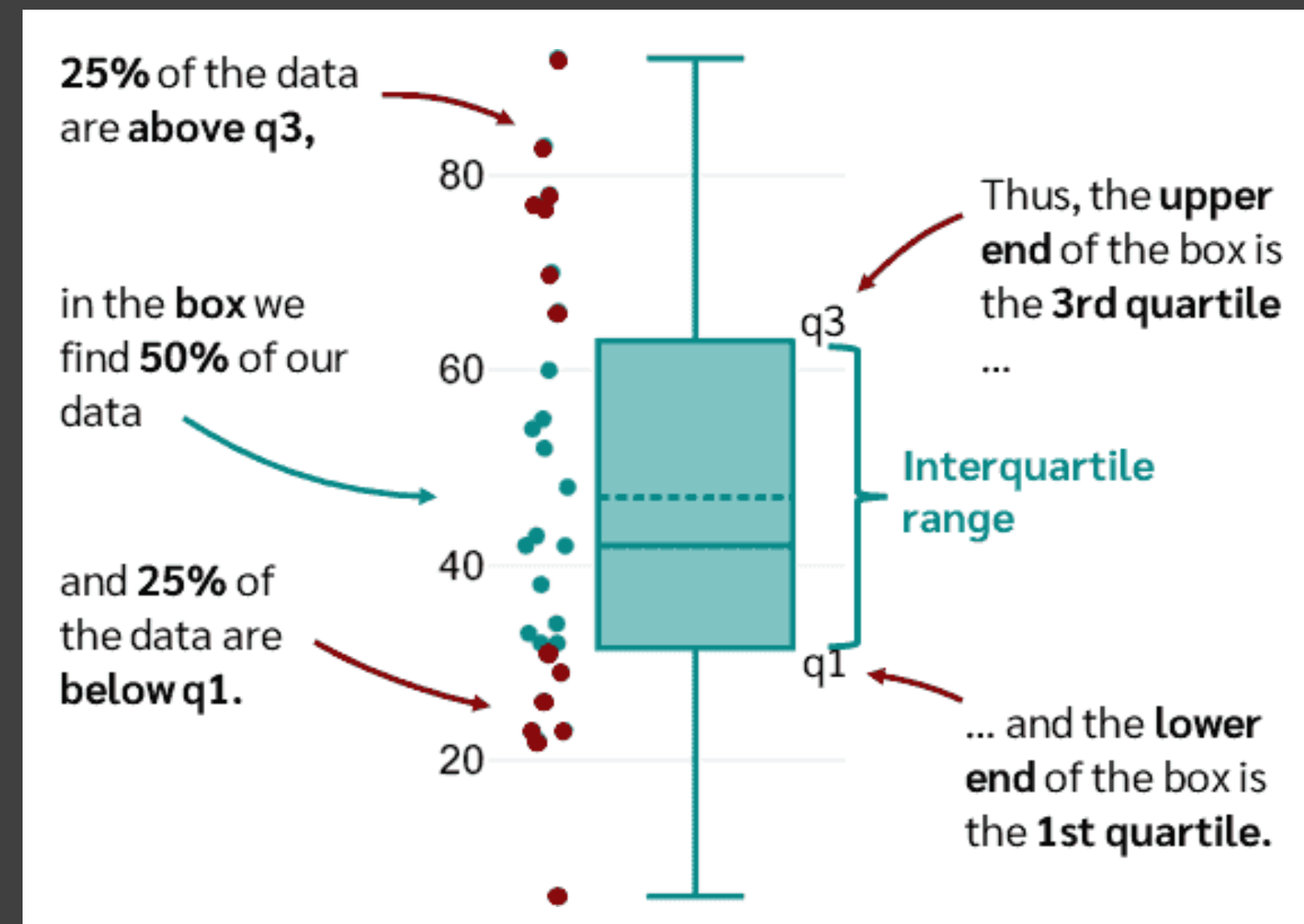
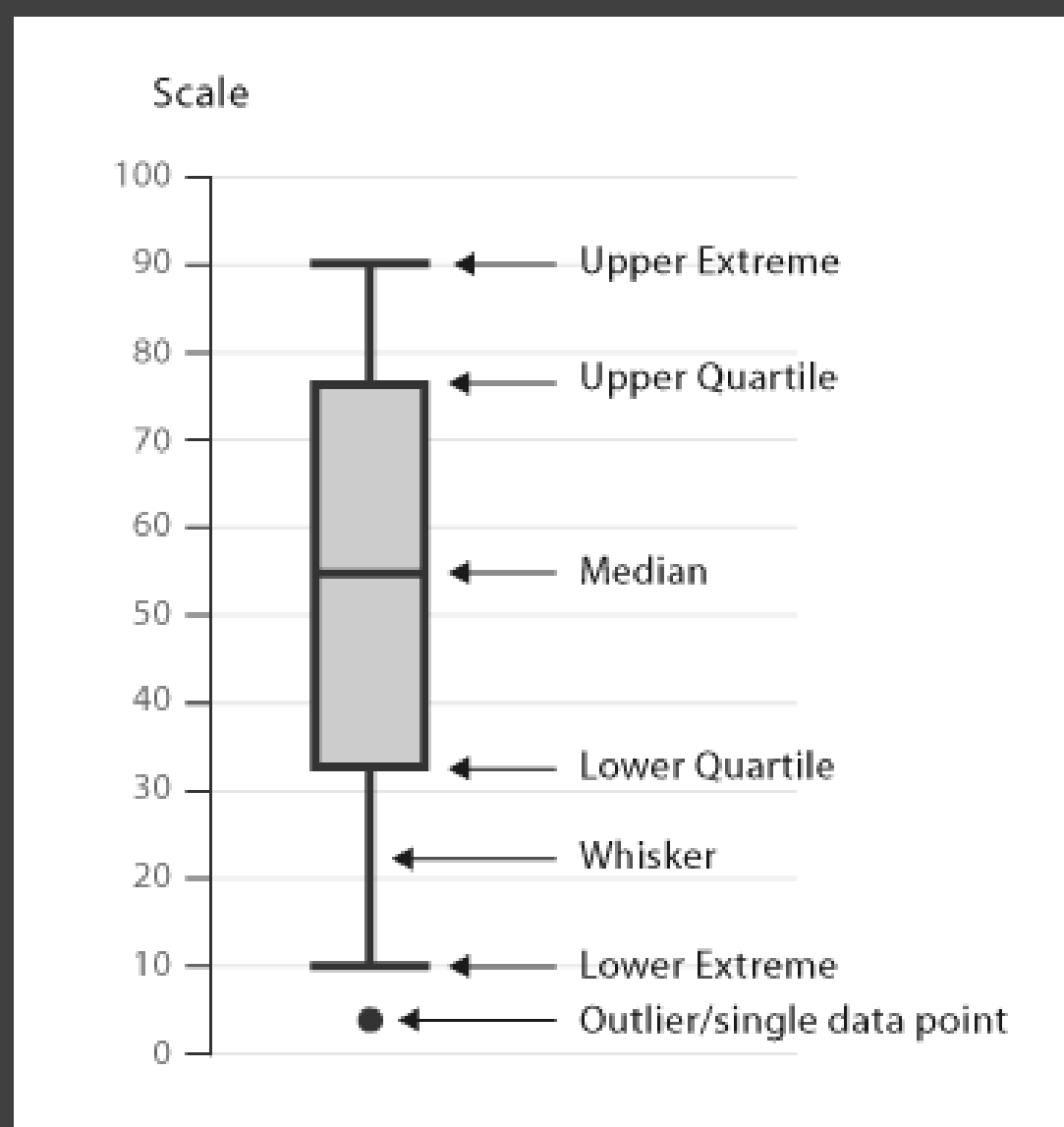
Study the **distribution** of one or a few variables





# BOX PLOT

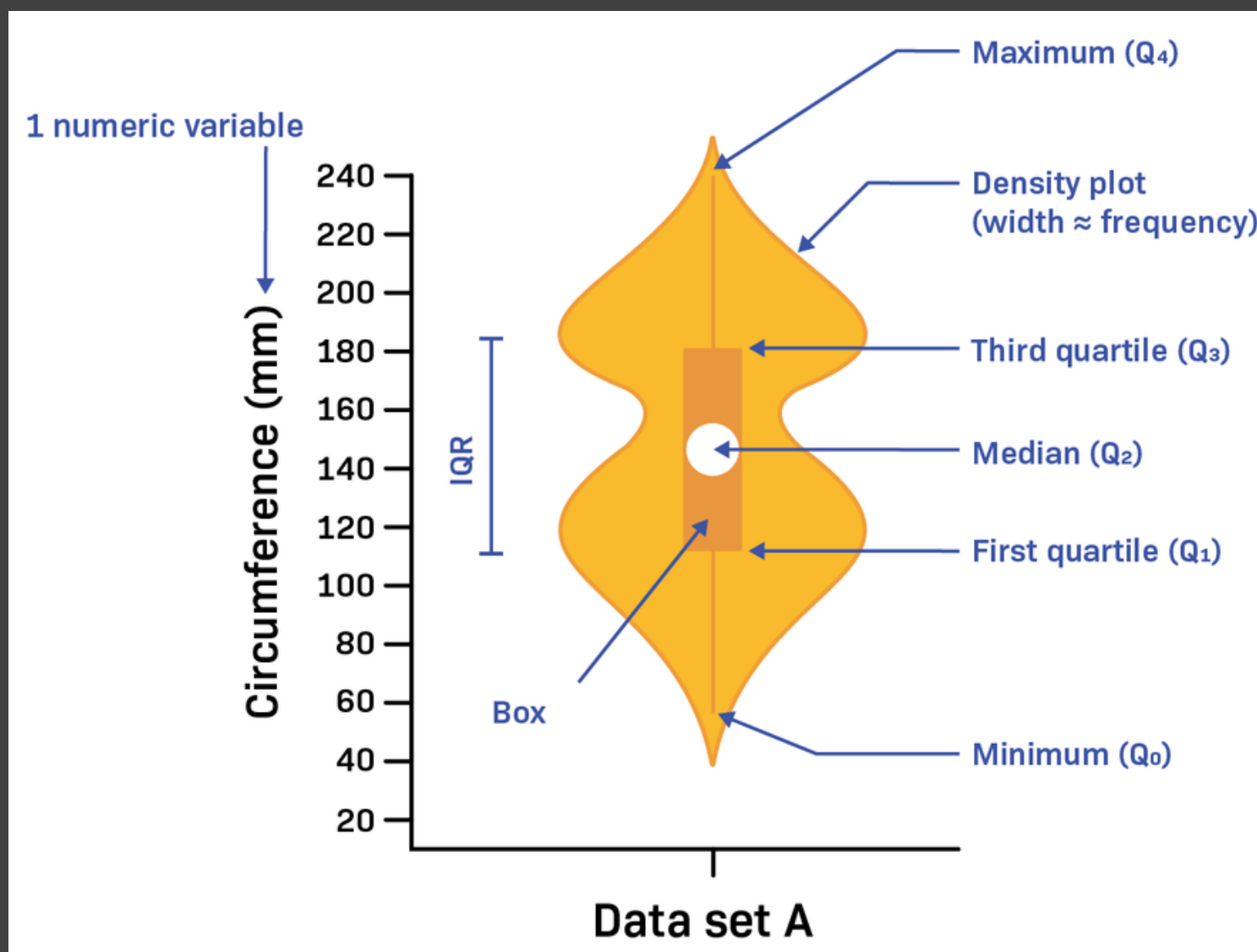
Display data distribution through quartiles





# VIOLIN PLOT

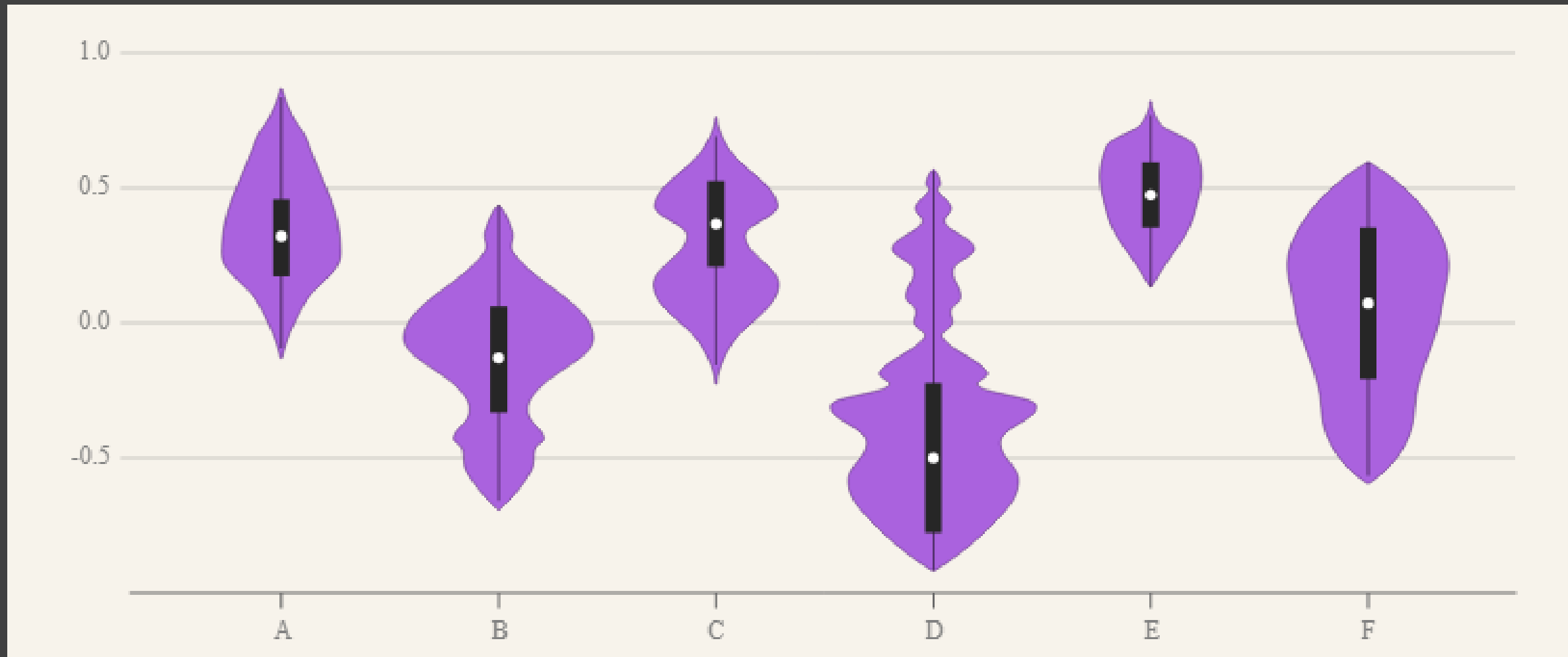
Visualize the distribution of a numeric variable for one or several groups





# VIOLIN PLOT

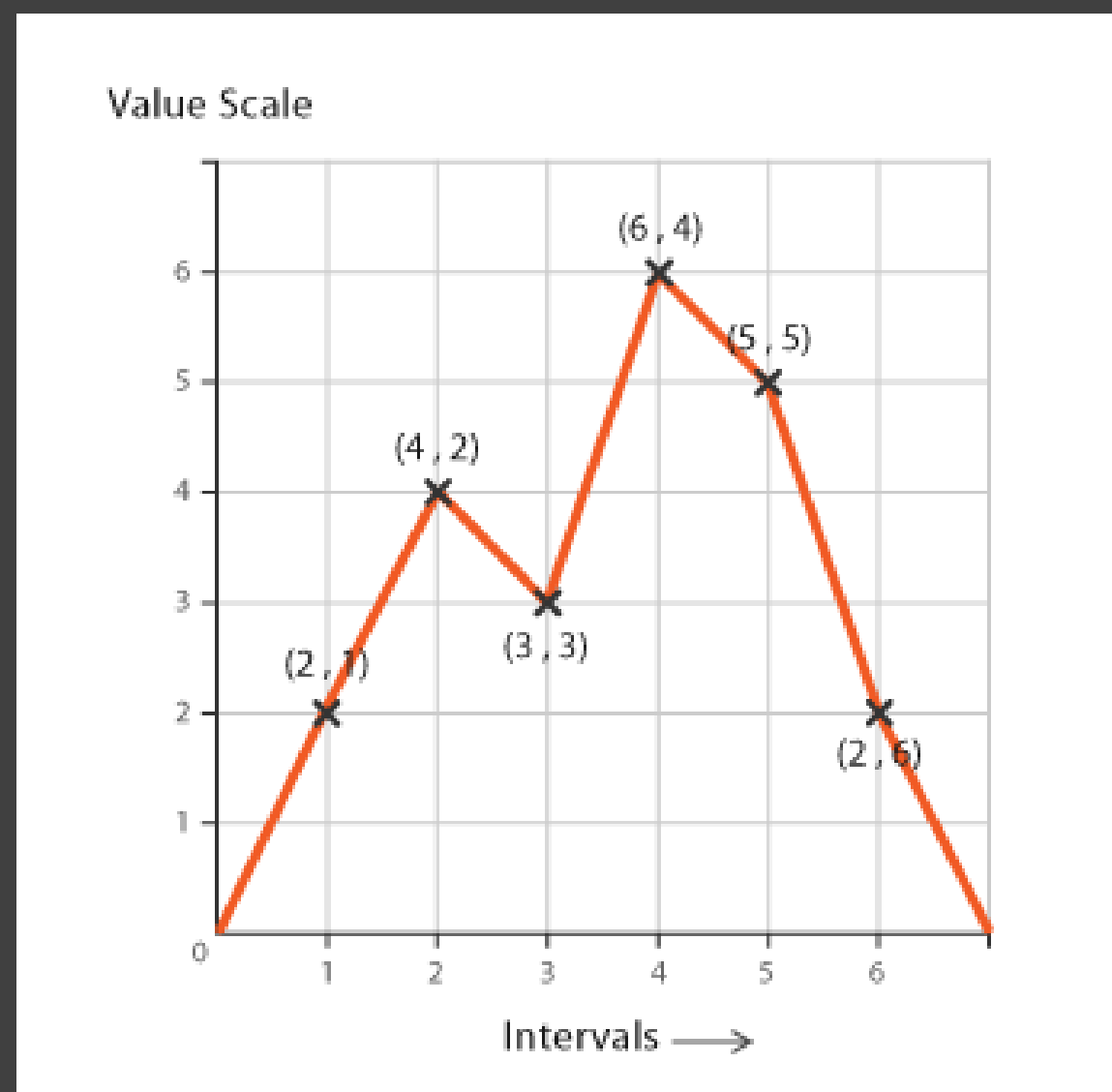
Visualize the distribution of a numeric variable for one or **several** groups





# LINE CHART

Display the evolution of one or several numeric variables

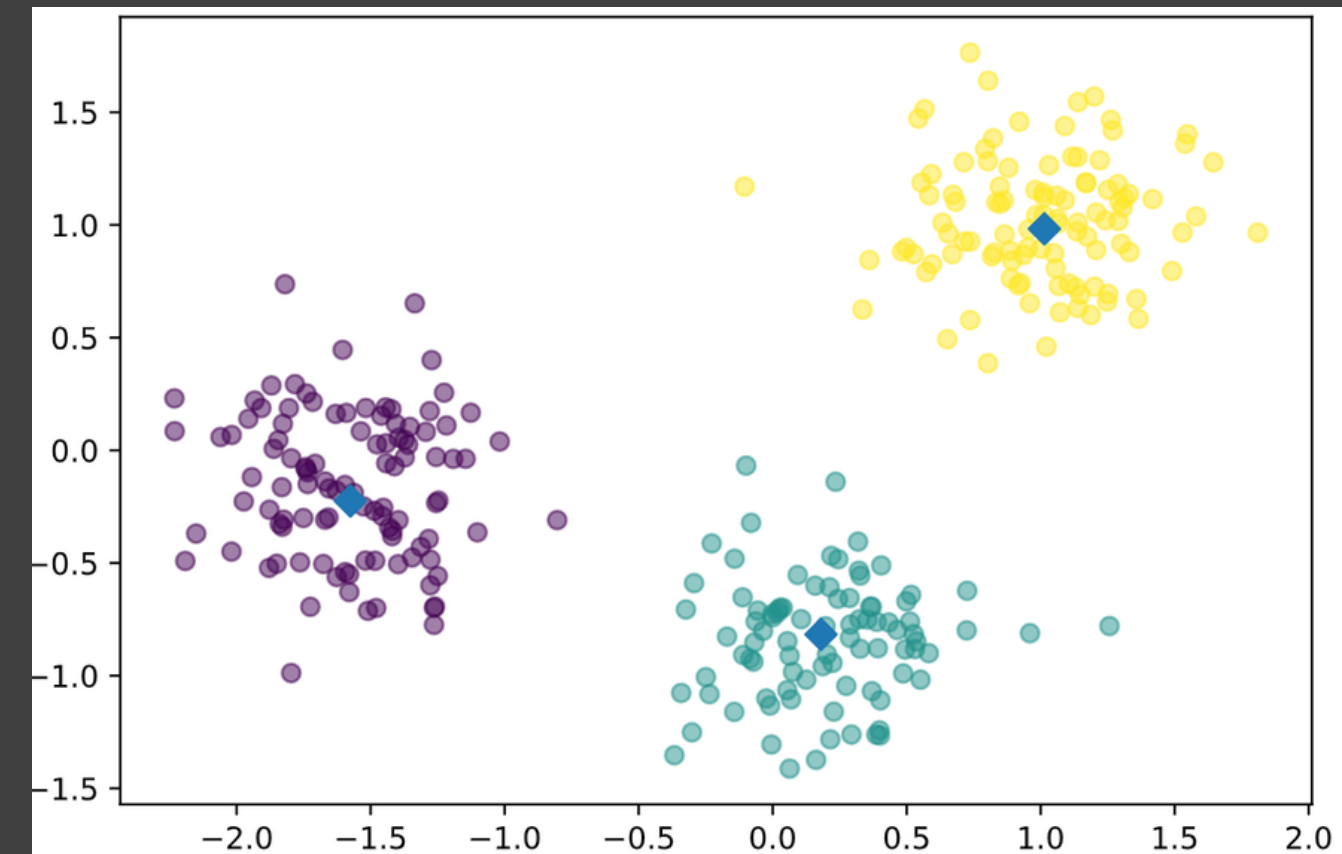
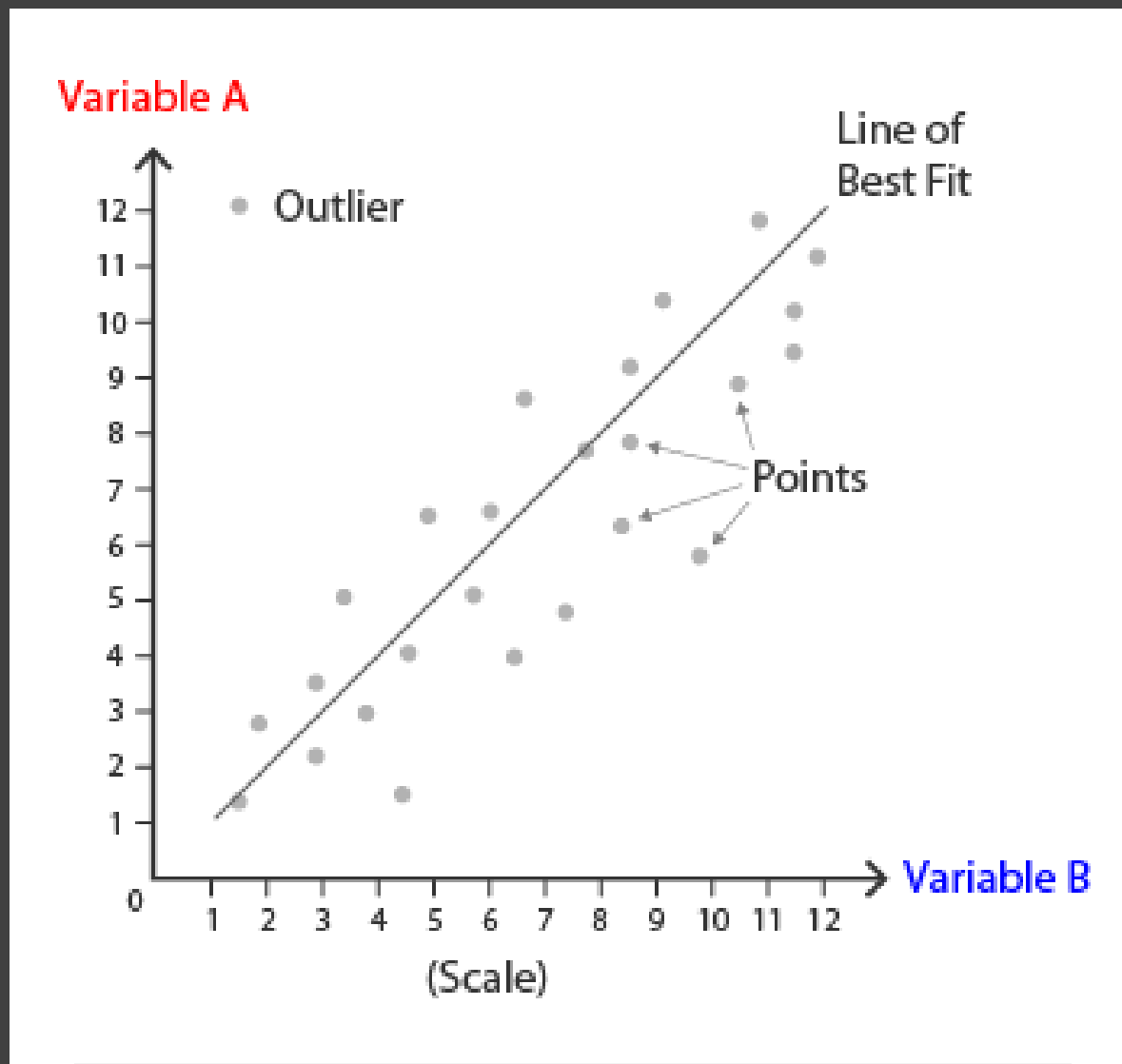






# SCATTER PLOT

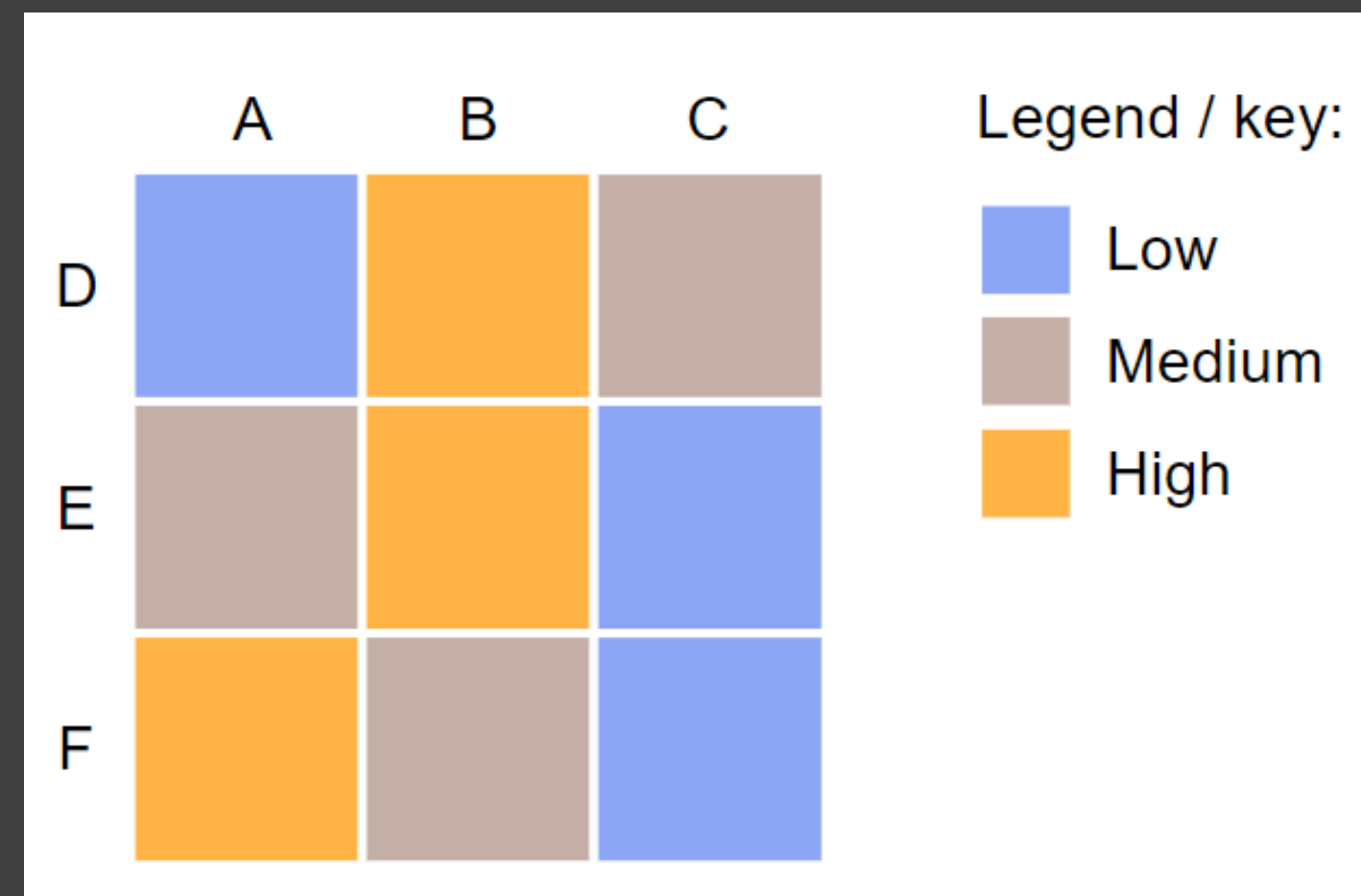
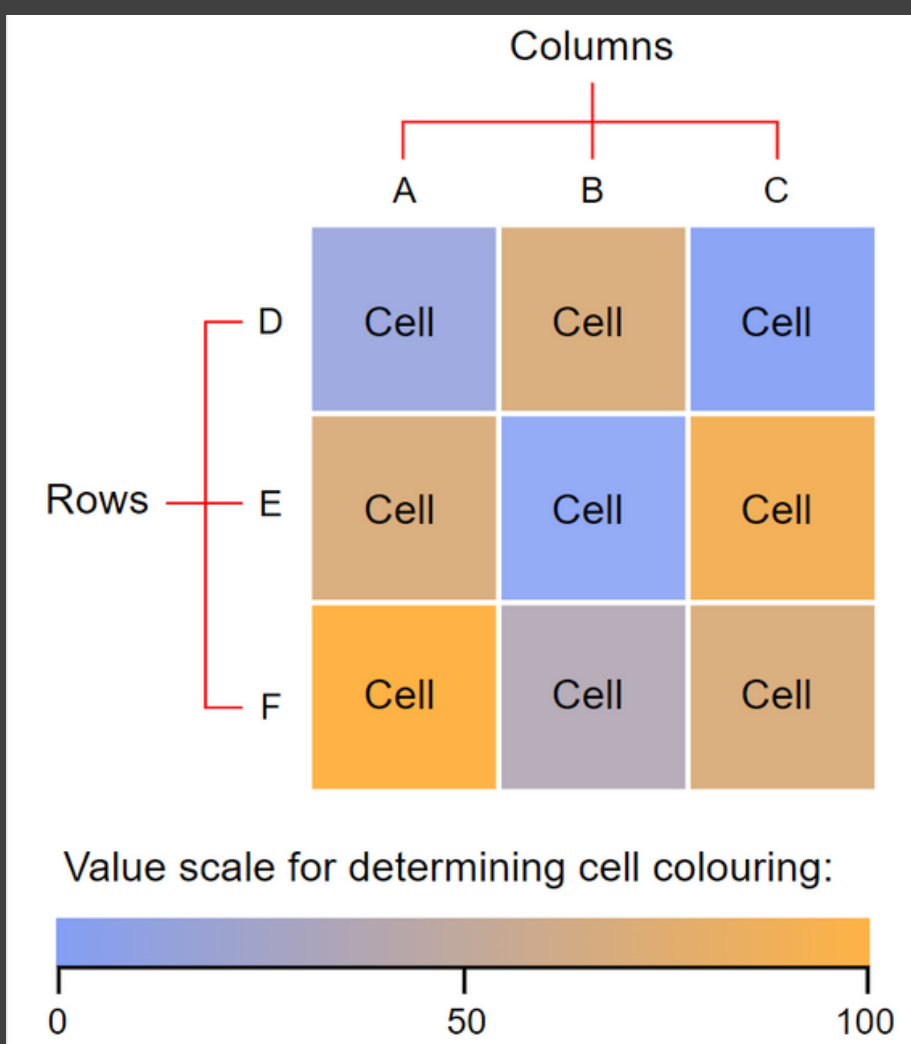
Display the relationship between 2 numeric variables





# HEATMAP

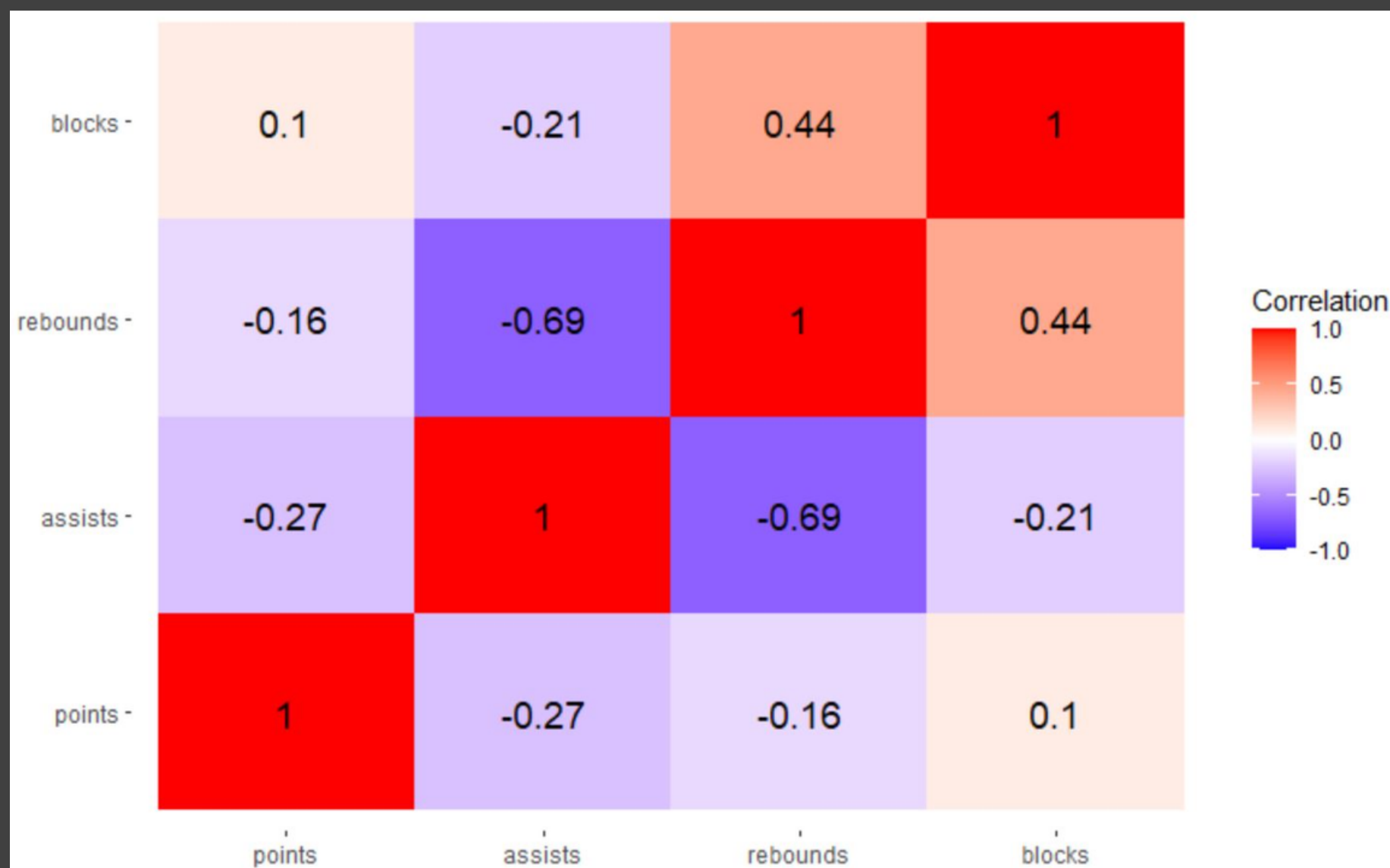
Cross-examining multivariate data



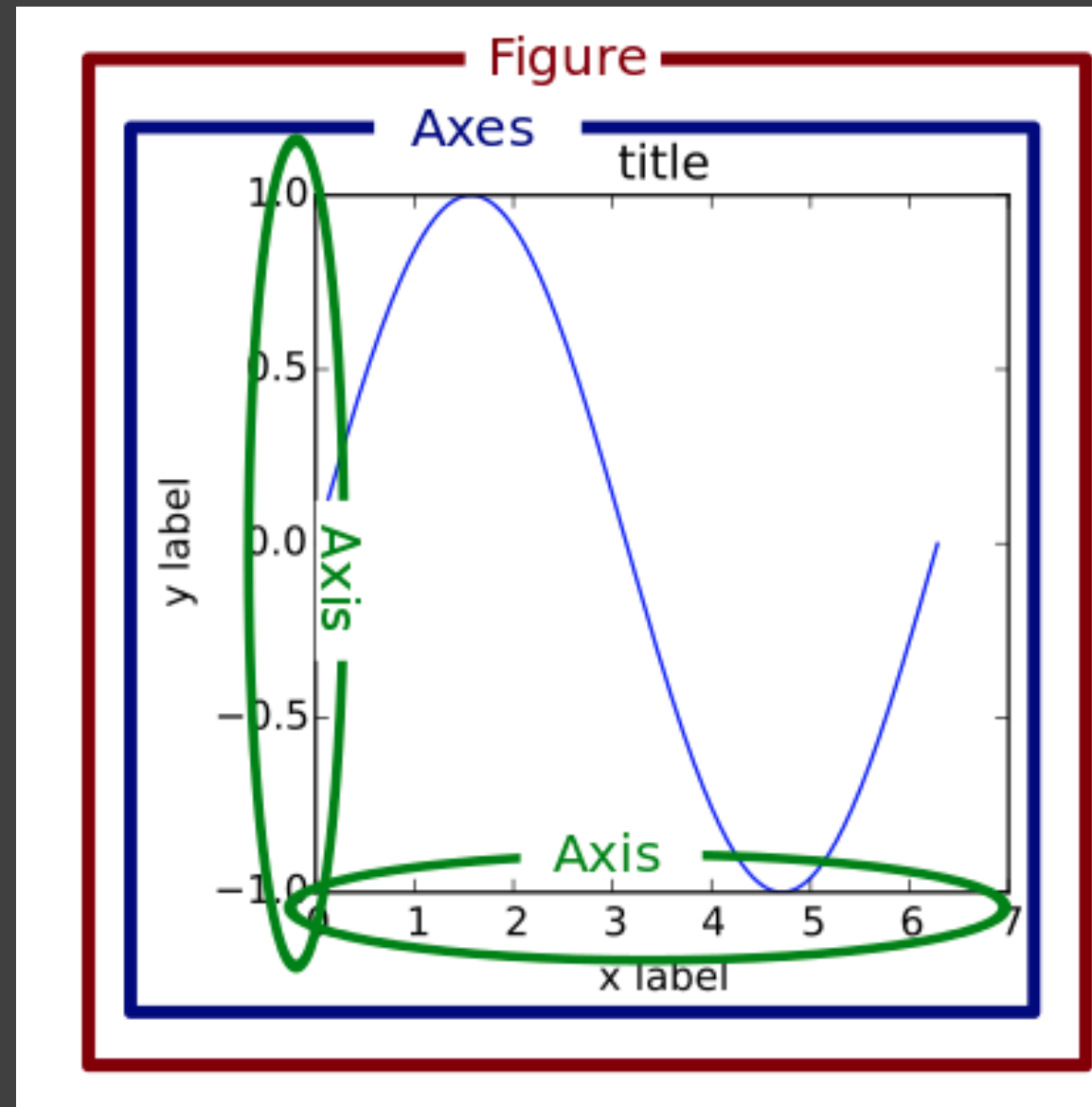


# HEATMAP

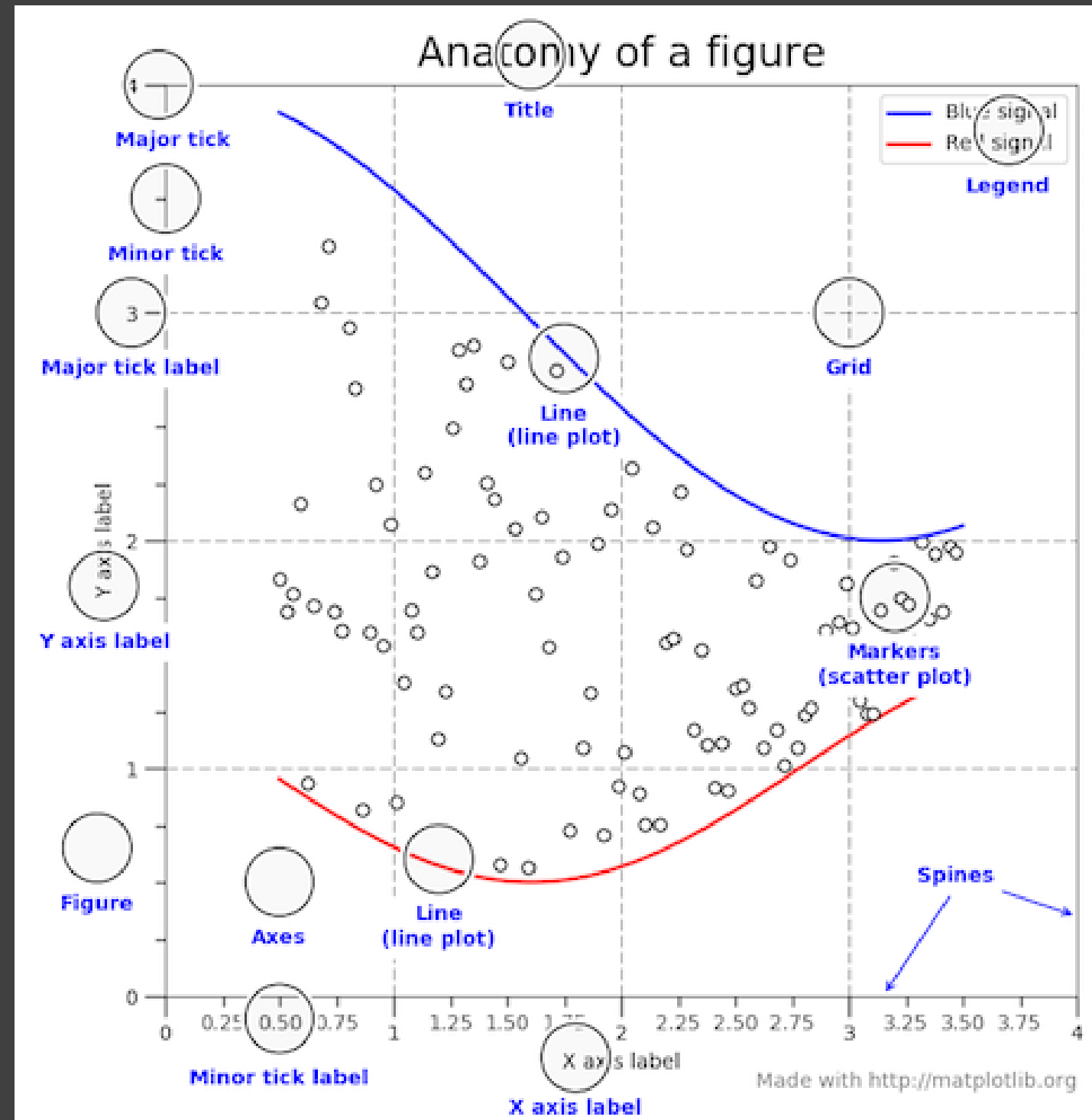
Cross-examining multivariate data



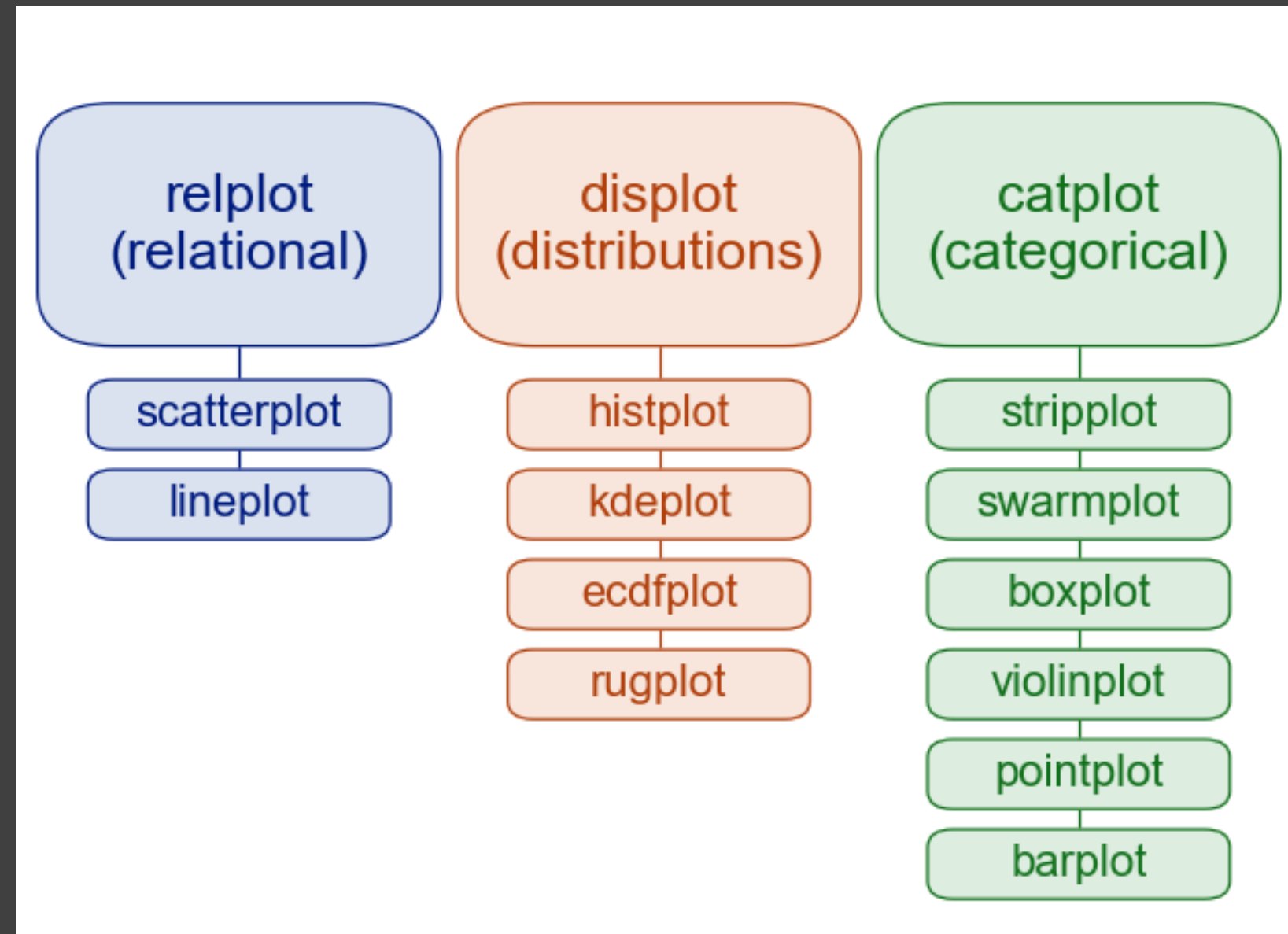
# Understanding *matplotlib*



# Understanding *matplotlib*



# Understanding *seaborn*



# Time to practice !



