

# What is Visualization?

**Visualize:** “To form a mental vision, image, or picture of (something not visible or present to the sight, or of an abstraction); to make visible to the mind or imagination.”

**Visualization** is the use of computer graphics to create visual images which aid in the understanding of complex, often massive representations of data.

# Table vs Graph

**A table is best when:**

- **You need to look up specific values**
- **Users need precise values**
- **You need to precisely compare related values**
- **You have multiple data sets with different units of measure**

**A graph is best when:**

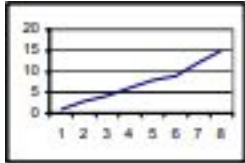
- **The message is contained in the shape of the values**
- **You want to reveal relationships among multiple values (similarities and differences)**
- **Show general trends**
- **You have large data sets**

# **Data Visualization – Common Display Types**

## **Common Display Types**

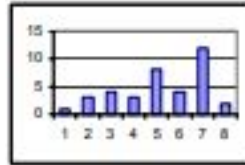
- Bar Charts**
- Line Charts**
- Pie Charts**
- Bubble Charts**
- Stacked Charts**
- Scatterplot**
- Boxplot**

# When to use which type?



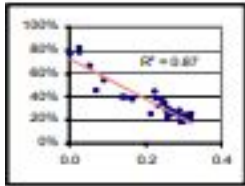
## Line Graph

- X-axis requires quantitative variable
- Variables have contiguous values



## Bar Graph

- Comparison of relative point values



## Scatter Plot

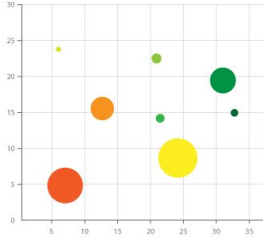
- Convey overall impression of relationship between two variables



## Pie Chart

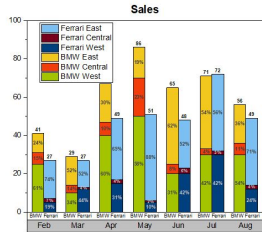
- Emphasizing differences in proportion among a few numbers

# When to use which type?



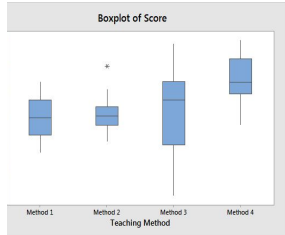
## Bubble Charts

- Primarily used to depict and show relationships between numeric variables
- Allows for the comparison between three variables rather than just two.



## Stacked Charts

- Sum of the values is as important as the individual items
- Stacked graphs are commonly used on bars, to show multiple values for individual categories, or lines, to show multiple values over time



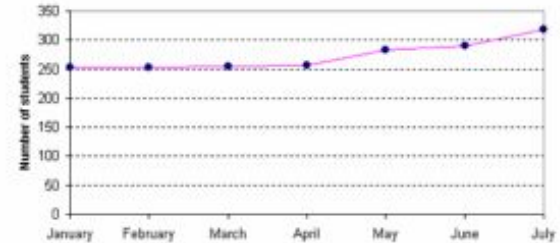
## Box Plot

- Comparisons across different categorical variables or identifying outliers, if either of those exist in a dataset

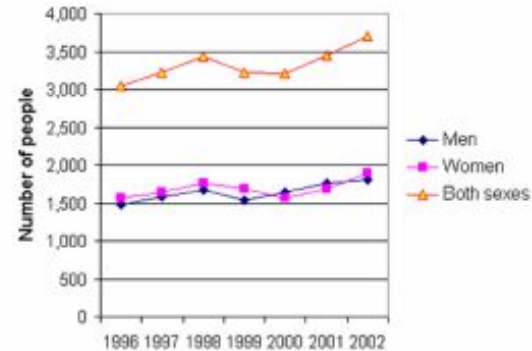
# Line Graph

Fundamental technique of data presentation

- Used to compare two variables
  - X-axis is often the control variable
  - Y-axis is the response variable
- Good at:
  - Showing specific values
  - Trends
  - Trends in groups (using multiple line graphs)



Students participating in sporting activities



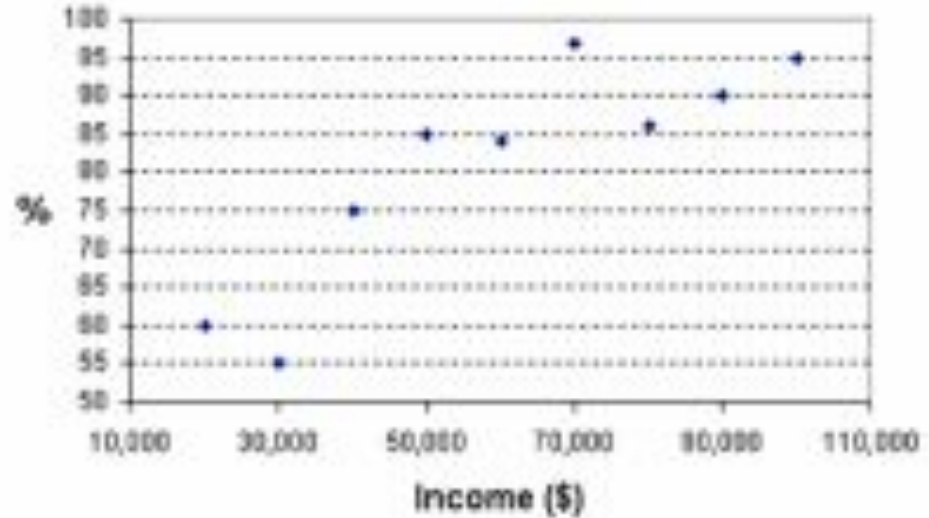
Mobile Phone use

# Time Line Graph - shows dynamics of measurements



# Scatter Plot

- Used to present measurements of two variables
- Effective if a relationship exists between the two variable

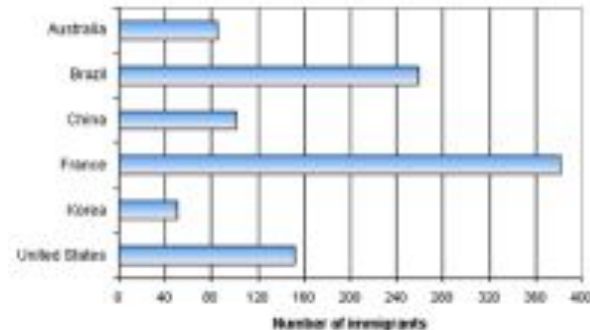
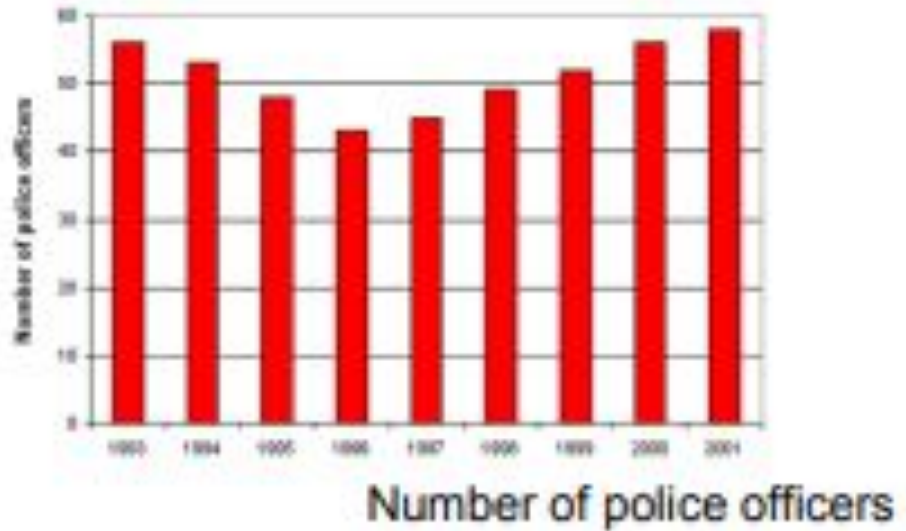


Car ownership by household income

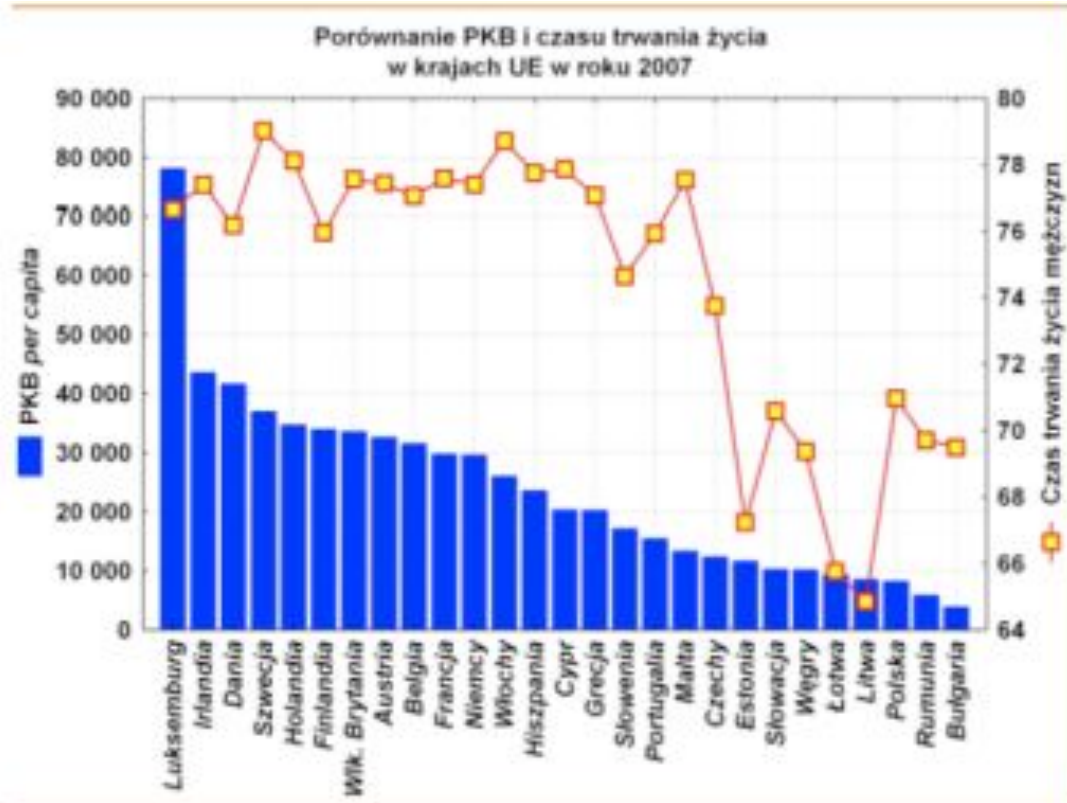


# Bar Graph

- Presents categorical variables
- Height of bar indicates value
- Double bar graph allows comparison
- Note spacing between bars – Can be horizontal (when would you use this?)

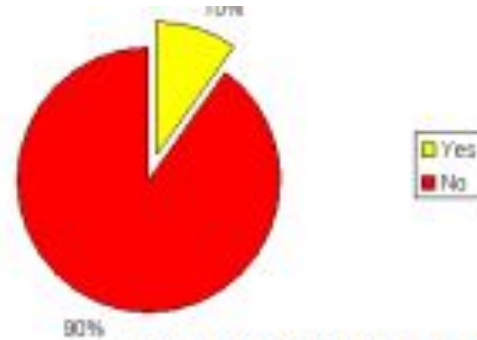


# Integrating various graphs

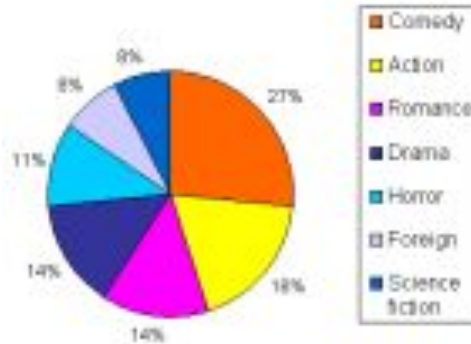


# Pie Chart

- Pie chart summarises a set of categorical/nominal data
- But use with care: too many segments are harder to compare than in a bar chart



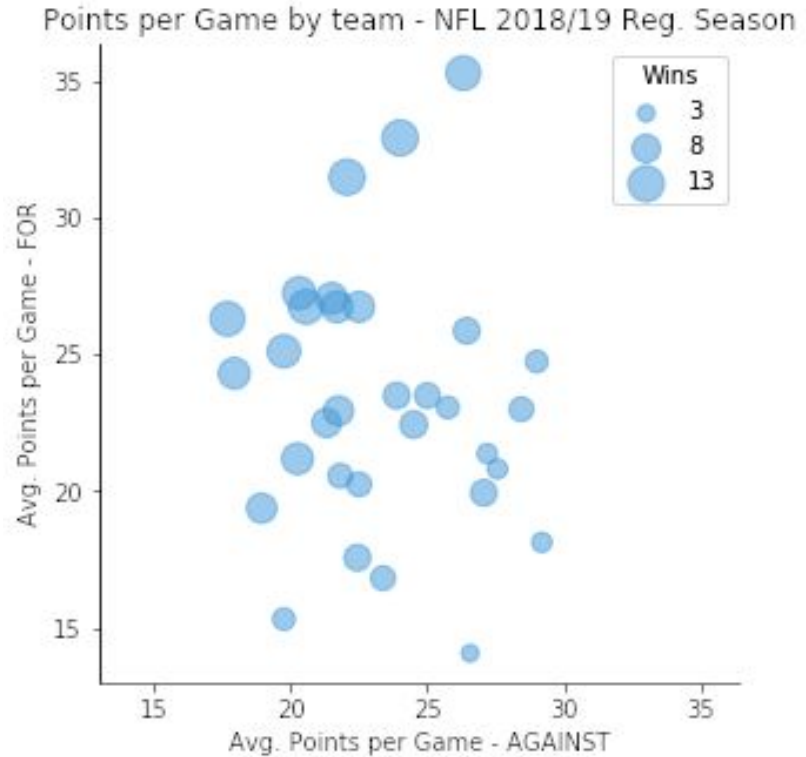
Should we have a long lecture?



Favourite movie genres

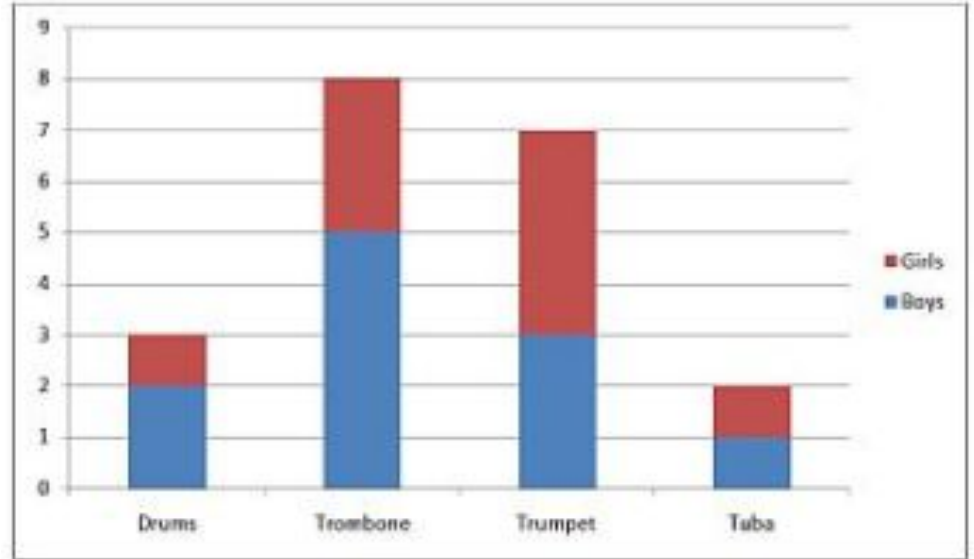
# Bubble Chart

- an extension of the scatter plot used to look at relationships between three numeric variables.
- Each dot in a bubble chart corresponds with a single data point, and the variables' values for each point are indicated by horizontal position, vertical position, and dot size.



# Stacked Chart

- to break down and compare parts of a whole.
- Each bar in the chart represents a whole, and segments in the bar represent different parts or categories of that whole

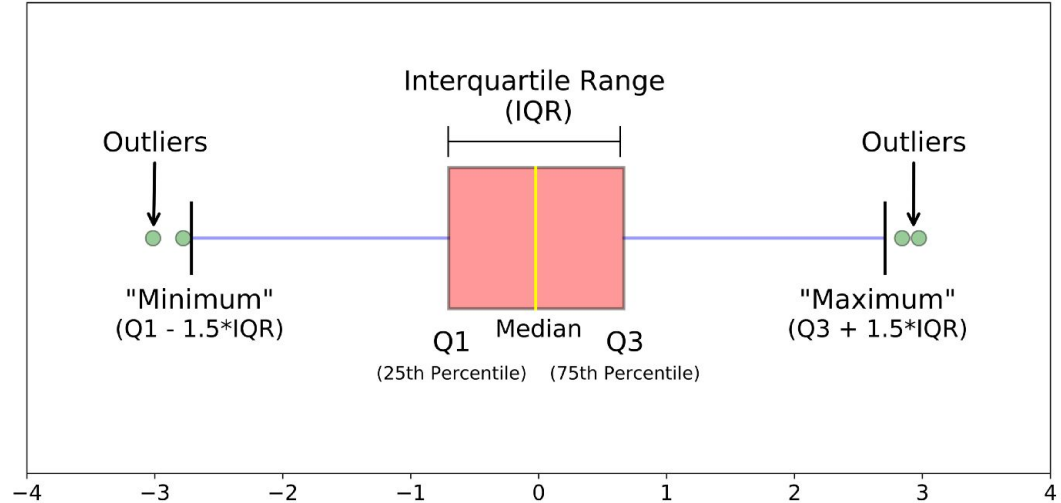


# Boxplot

- displaying the distribution of data based on a five number summary

- “minimum”,
- first quartile (Q1),
- median,
- third quartile (Q3),
- “maximum”

- outliers and what their values are



# Elements of a good graph

- Title
- Error Bars
- Axis Label
- Date

