Data Visualization
Part 1

Dr. Supaporn Erjongmanee

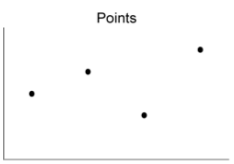
Department of Computer Engineering
Kasetsart University
fengspe@ku.ac.th

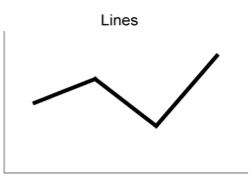
Outline

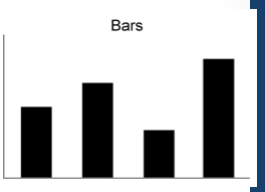
- Introduction
- Time-series
- Part-to-whole and Ranking
- Deviation

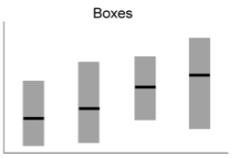
Visual Encoding

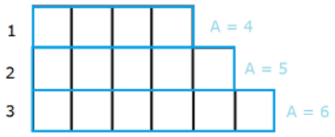
- Points
- Lines
- Bars
- Boxes
- Areas
- Colors

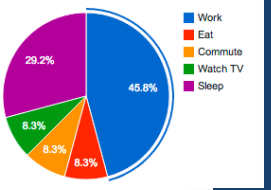








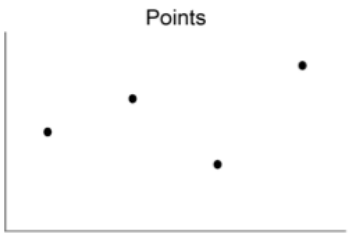




Source: [2]

Points

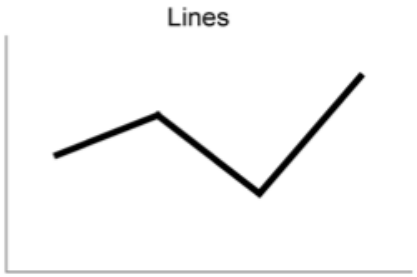
- Smallest part to represent values
- Can be many shapes: dots, squares, x's, etc.
- Stress on individual values
- Encode two quantitative values
  - Example: dot plot, scatter plot



Source: [2]

Lines

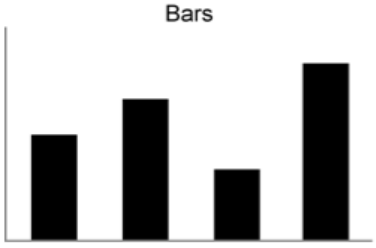
- Show movement from one value to the next
- Link individual values into series
- Good for displaying trends and patterns



Source: [2]

Bars

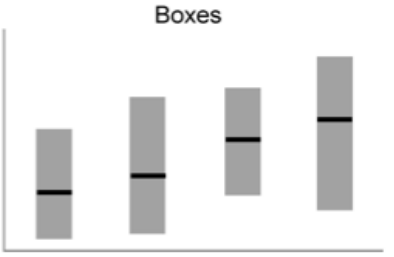
- Strongly emphasize on individual values
- Good at comparing individual values



Source: [2]

Boxes

- Both ends have values
  - Encapsulate minimum and maximum quantitative values
- Good for distribution of values
- Can add markers for 25%, 50%, 75% percentile

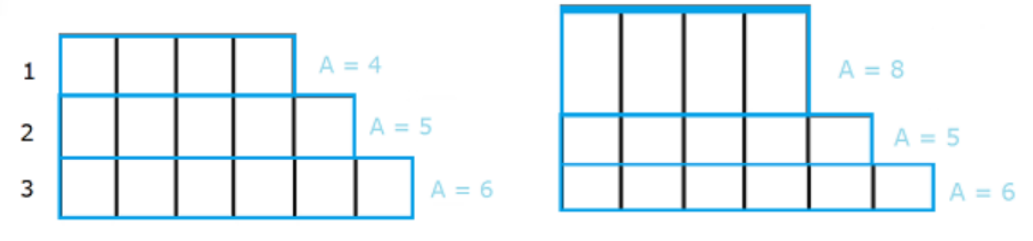


Source: [2]

7

Areas

- Present values in two dimensions
- Ratio of area must be accurately presented in its value

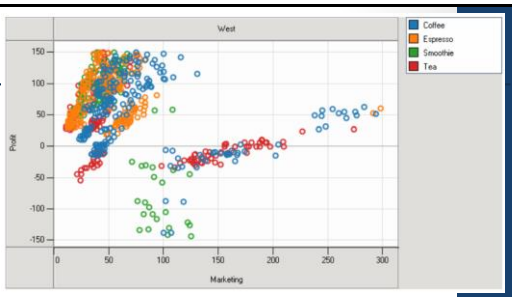


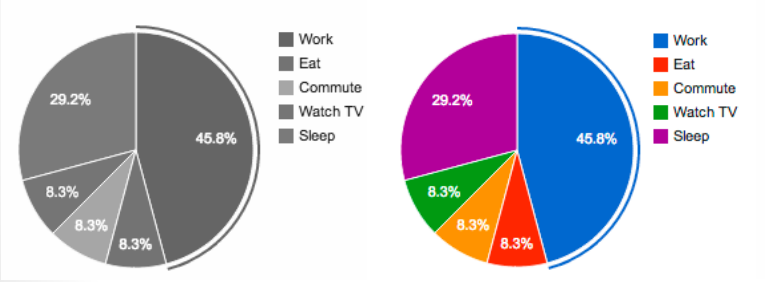
Source: http://www.statisticshowto.com/area-principle-in-statistics/

8

Colors

- Help enhance graph
- Distinguish different components better

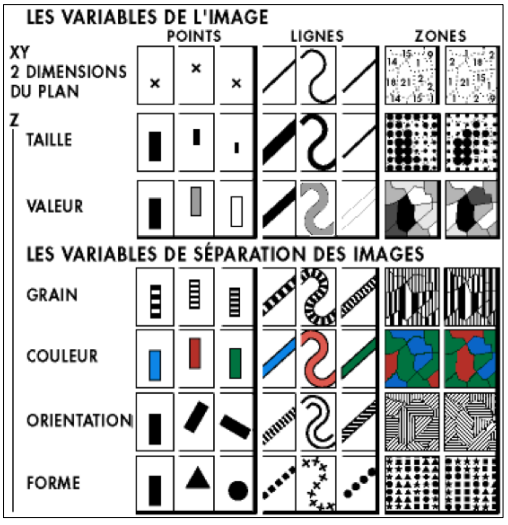




Source: https://www.perceptualedge.com/articles/b-eye/choosing_colors.pdf
Image Sources: https://www.w3.org/TR/respimg-usecases/, https://www.perceptualedge.com/articles/b-eye/choosing_colors.pdf

Visual-Encoding Options

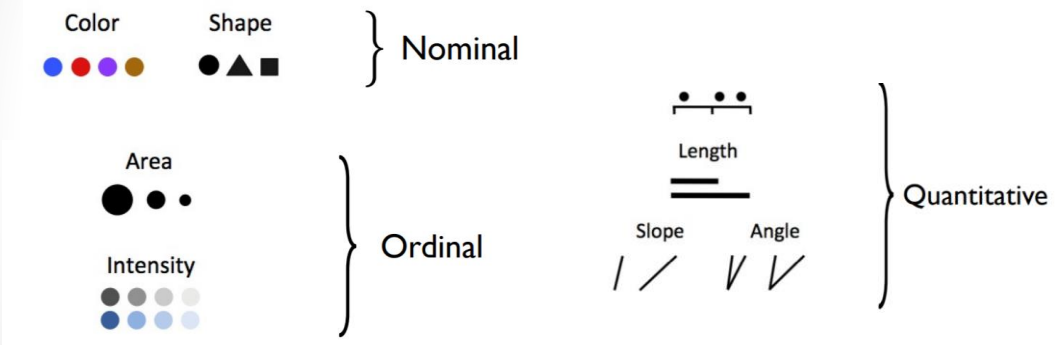
- There are 8 choices to vary!
  - Position
  - Length
  - Area
  - Volume
  - Intensity
  - Texture
  - Color
  - Orientation
  - Shape



Source: [2]

Visual-Encoding for Data Types

- For different data types



Source: [3]

Visualization

- Visual display to show relationships among attributes
  - Relationships: trend, pattern, etc.
- Graph selection and design process
  1. Choose your graph objective and what you want to convey from your graph
  2. Identify your data
     - Data types: nominal, ordinal, interval, ratio
     - Number of variables: single, double, multiple
  3. Choose graph type
  4. Add on graph components
  5. Revise until the goal objective is satisfied

Source: [1]

Common Graph Objectives

- Timeseries
- Part-to-whole and Ranking
- Deviation
- Distribution
- Correlation
- Network
- Geo-visualization

Outline

- Introduction
- Time-series
- Part-to-whole and Ranking
- Deviation

Time-Series Displays

- Commonly used
  - Line graph
  - Bar graph
- Others
  - Radar graph
  - Heatmap

15

Line Graph

- ✓ Display patterns
- ✓ Show exceptions

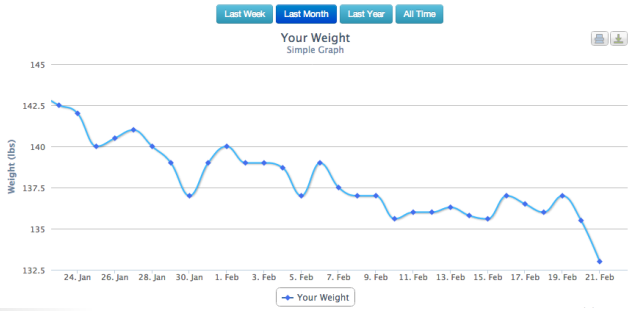
- To see change during continuous period of time → Used to display time-series
- Features
  - Length clearly display flow
  - Slope shows track of change

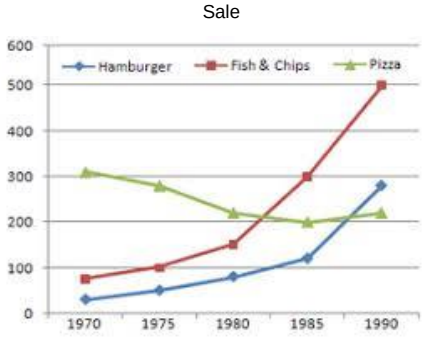
Image source: Figure 7.12, [1]

16

Line Graph (cont.)

- Commonly use with time-series to see trend & patterns



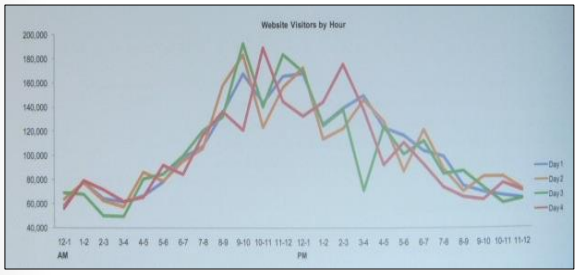


17

Image sources: http://www.weightgrapher.com/static//simple_graph.png
https://encrypted-tbn0.gstatic.com/images?q=tbn:ANd9GcQrZPoqPYVDOaW2n_Ka3Ekme8ADTL5fDlFjp4JcQXNzuZosQ-2Q

Line Graph (cont.)

- Can be used to display cycle
  - Plot multiple time-series on to same timeframe



Number website users

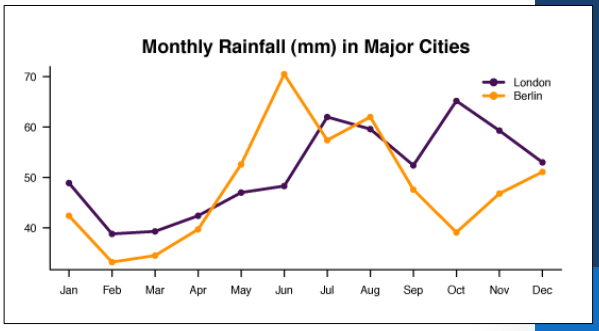
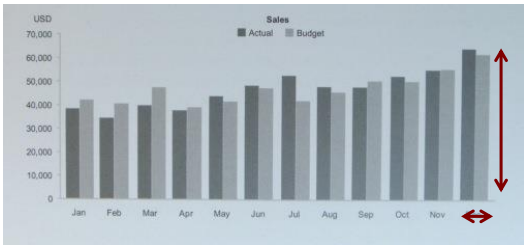


Image source: Figures 7.18, 7.19, [1]

Bar Graph

- ✓ Emphasize individual values
- ✓ Compare individual values



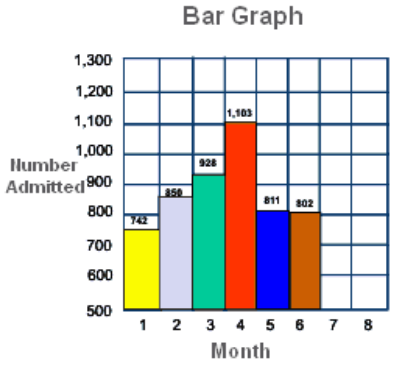
- To see or compare individual values
- Features
  - Height/length represents values
  - Width distinguishes different items

Image source: Figure 7.14, [1]

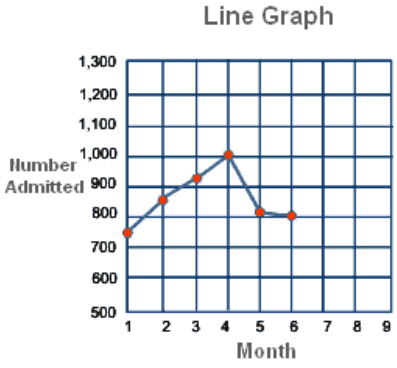
19

Example : Line vs. Bar

Emergency Department Admissions per Month



Compare month by month

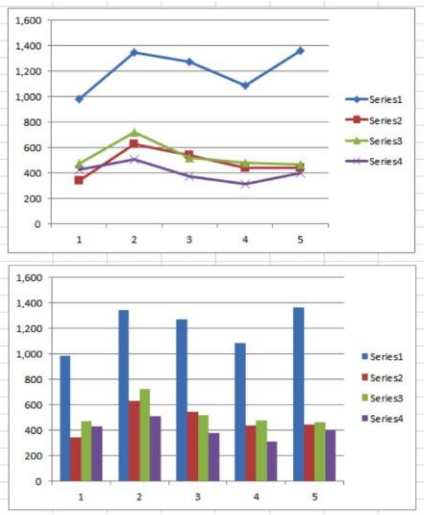


See trend of change

Image source: http://www.qualitytrainingportal.com/resources/problem_solving/images/bar_charts.gif

20

Example 3 : Line vs. Bar



- Line graph
  - See change over time in bar graph of each series more clearly
- Bar graph
  - See comparison of different series in same time more clearly

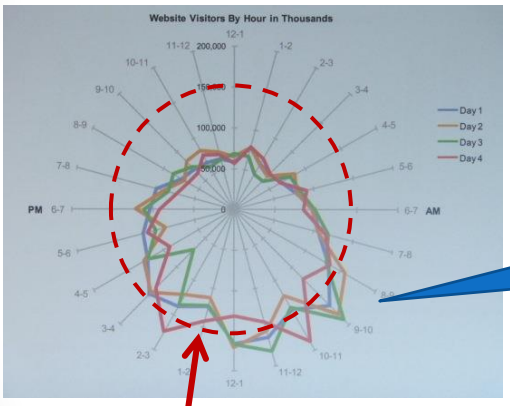
Image source: http://online-behavior.com/sites/default/files/imagecache/Content/articles/line-graph.jpg

Time-Series Displays

- Commonly used
  - Line graph
  - Bar graph
- Others
  - Radar graph
  - Heatmap

Radar Graph

✓ Compare multiple variables



- Also call spider/web/star/ polar graphs
- Plot variables around center

Almost the same pattern among 4 days
Many visitors during 9 AM-1PM

- For cycle/time-series, variables = times
- Avoid comparison with too many series
- Feature: Circular shape represents cyclical nature of time

Image source: Figure 7.18, [1]

Radar Graph (cont.)

- Example: Comparison between radar and line graphs

Maximum website users occur around 9:00-12:00

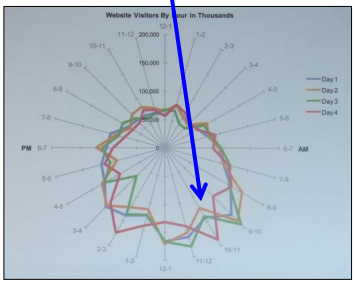
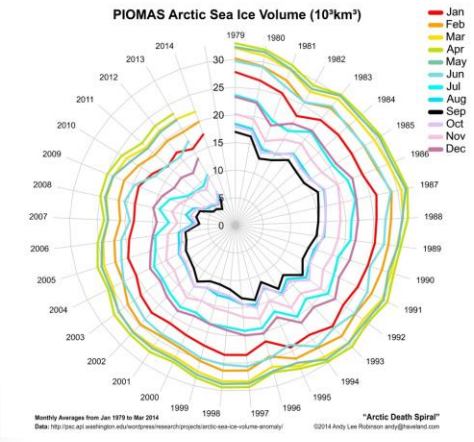


Image source: Figures 7.18, 7.19, [1]

Example 2 : Radar Graph

Excel: Spider chart

- Example 2: Comparison between radar and line graphs



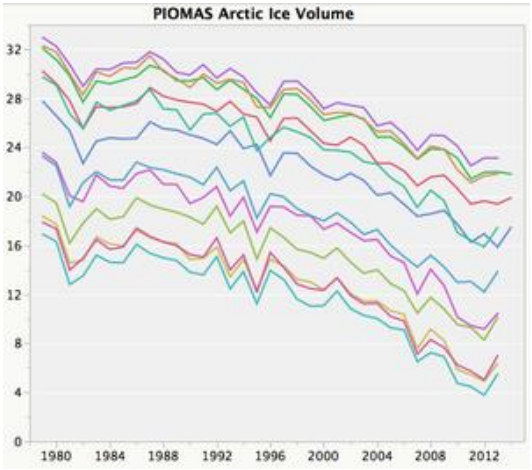
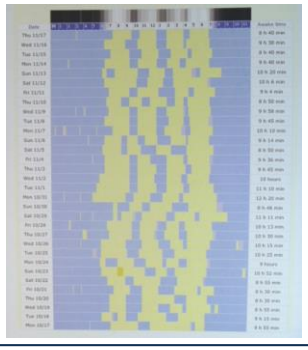


Image source: http://junkcharts.typepad.com/junk_charts/2014/04/an-overused-chart-why-it-fails-and-how-to-fix-it.html

Heatmap

- ✓ Analyze patterns of large number of series in cycles
- ✓ Show exception of multiple series

- To analyze multiple series
  - For comparison between many series, heatmap works better than radar graphs
- Feature
  - Color represents values



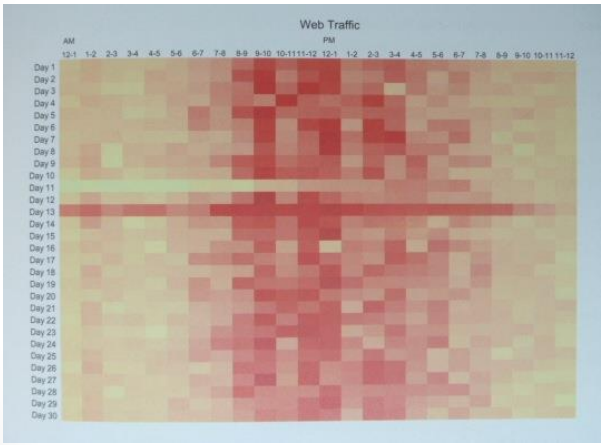
Blue = Sleep, Yellow = Awake

Sleeping patterns of a child during one month

Image source: Figure 7.21, [1]

Heatmap (cont.)

- Example 2: Multiple colors



Web Traffic

Excel:
Conditioning formatting

Image source: Figure 7.22, [1]

Outline

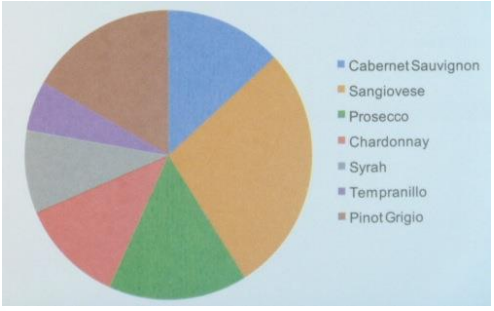
- Introduction
- Time-series
- Part-to-whole and Ranking
- Deviation

Displays for Part-to-Whole and Ranking

- Common graphs
  - Pie chart
  - Bar graph
- Other graphs
  - Donut chart
  - Stacked bar graph
  - Stacked area graph
  - Radar graph

Pie Chart

- Part-to-whole is normally displayed in pie graph
- Difficulty in
  - Comparison between slices
  - Linking between slices and legends

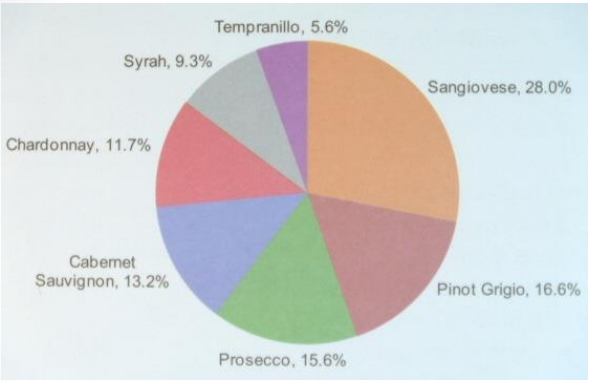


It is also difficult to identify percent amount.

How about adding numbers?

Image source: Figure 8.4, [1]

Pie Chart (cont.)



- We can also use table.

Wine	Percent
Sangiovese	28.0%
Pinot Grigio	16.6%
Prosecco	15.6%
Cabernet Sauvignon	13.2%
Chardonnay	11.7%
Syrah	9.3%
Tempranillo	5.6%
Total	100.0%

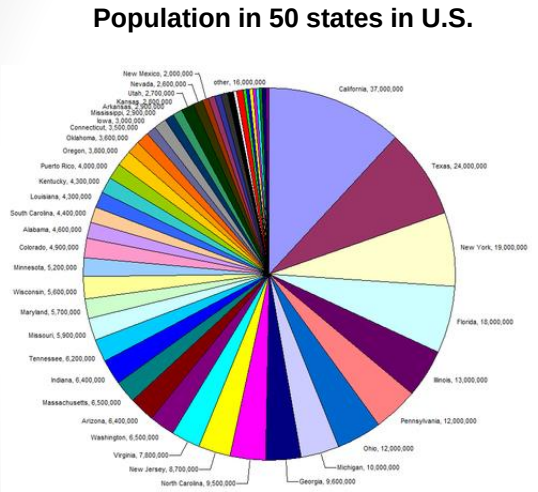
Image source: Figures 8.5, 8.6, [1]

Pie Chart (cont.)

- To use pie chart:
  - Use compare parts to whole
    - Not to compare parts to parts
  - No part overlapping
  - Not too many sections/parts (< 5-7 sections)
  - Sort from max to min (Generally, starting from 90° and turn clockwise)
  - Clear labels
  - Sum of all parts = 100%
  - Avoid 3D, donut, or splitting pie chart

Image source: http://www.f1f9.com/blog/please-stop-using-pie-charts

Pie Chart (cont.)

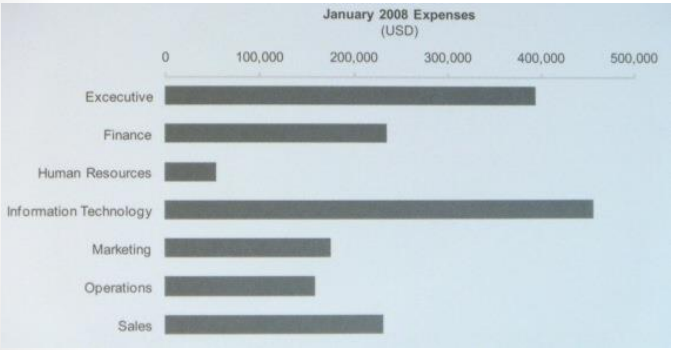


- If too many sections, make the labels clear.

Image source: https://eagereyes.org/techniques/pie-charts

Bar Graph

- Questions:
  - How much percent of expense did each department spend?
  - Which department proportionally spent the most money?



- We are looking portion, compared to whole?
- We can rank portion.

Image source: Figure 8.1, [1]

Bar Graph (cont.)

- Turn into percent of expense
- After sorting, we find that IT department spends the most money.

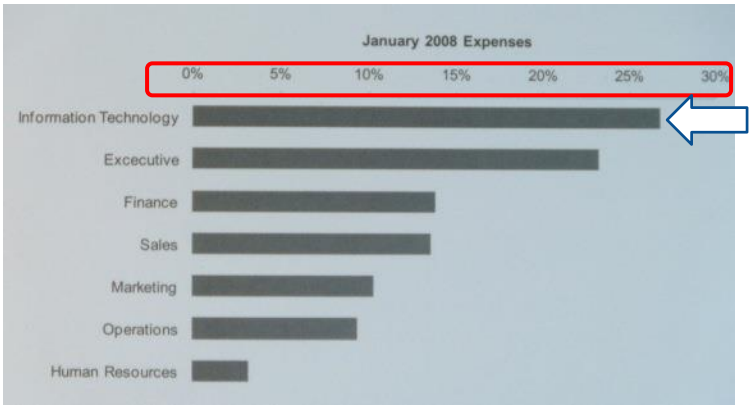
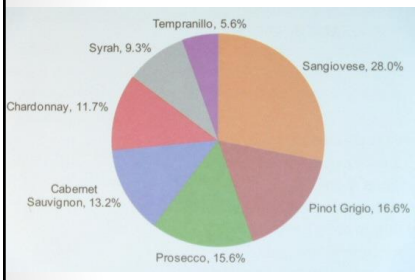


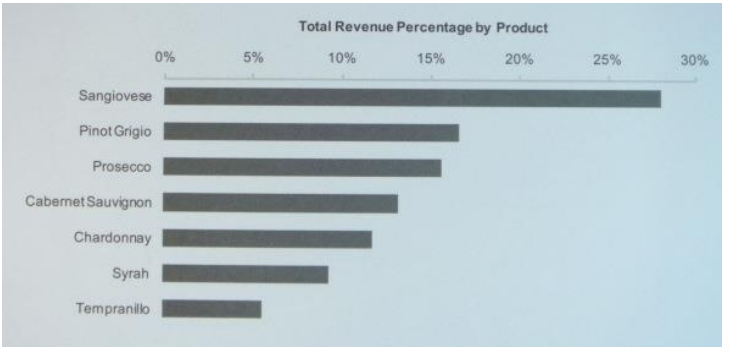
Image source: Figure 8.2, [1]

Bar Graph

- Can be used to present both part-to-whole and ranking

Same data as the previous pie graph





Drawback of bar graph? Narrow long bars.

Image source: Figure 8.7, [1]

Pie vs. Bar Graphs : Example

Image source: http://www.f1f9.com/blog/please-stop-using-pie-charts

Pie vs. Bar: Example 2

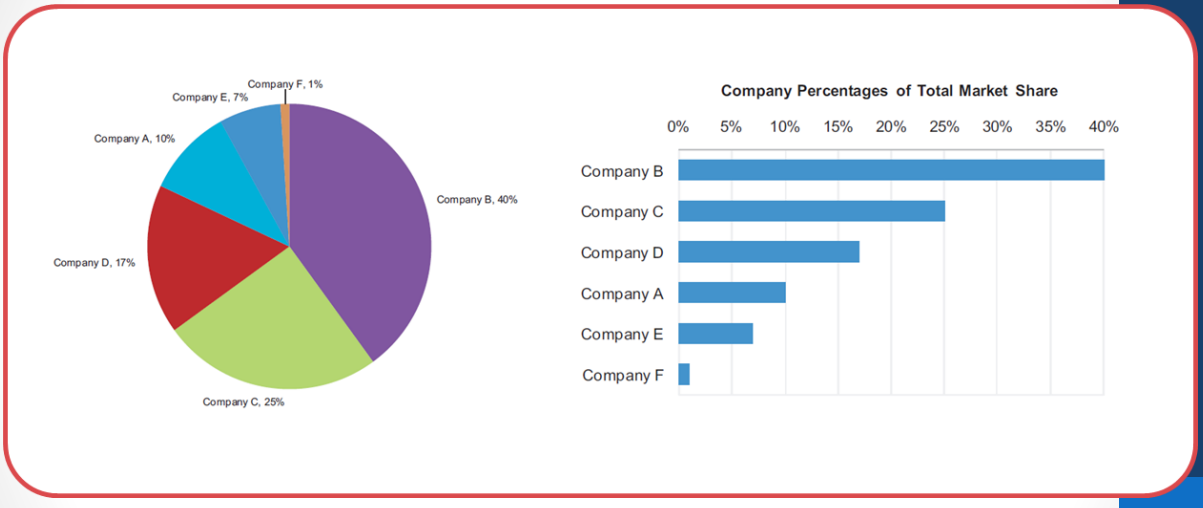


Image source: http://speakingppt.com/wp-content/uploads/2012/10/stephen-few-pie-vs-bar.png

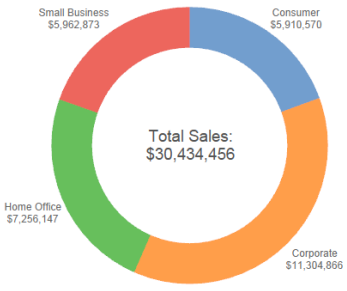
Displays for Part-to-Whole and Ranking

- Common graphs
  - Pie chart
  - Bar graph
- Other graphs
  - Donut chart
  - Stacked bar graph
  - Stacked area graph
  - Radar graph

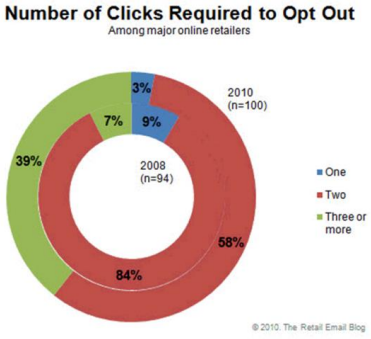
39

Donut Chart

- Similar to pie chart but
- Con: missing centric area and hard to distinguish sections



One data set

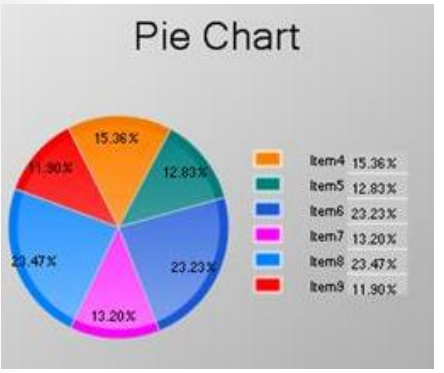


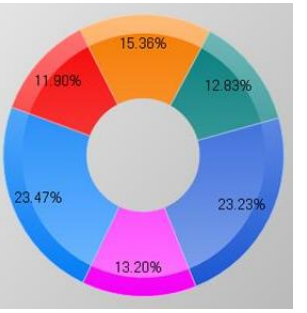
Two data sets

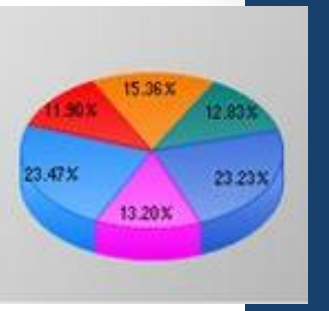
Image sources: Figures https://www.interworks.com/sites/default/files/blog/u150/donut.png
https://danjharrington.wordpress.com/2011/11/11/death-to-the-doughnut-chart/

40

Pie Chart vs. Donut (cont.)





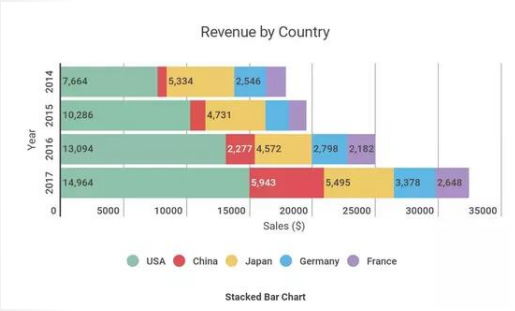


- Donut, Splitting, or 3D degrades readability
- Avoid them!

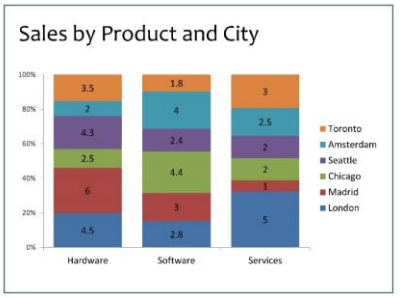
Image source: http://www.codeproject.com/KB/GDI-plus/simple_pie_chart_control/PieChart1.JPG/

Stacked Bar Graph

- Can compare with total summation or 100%



Compare with total sum
Values are actual values

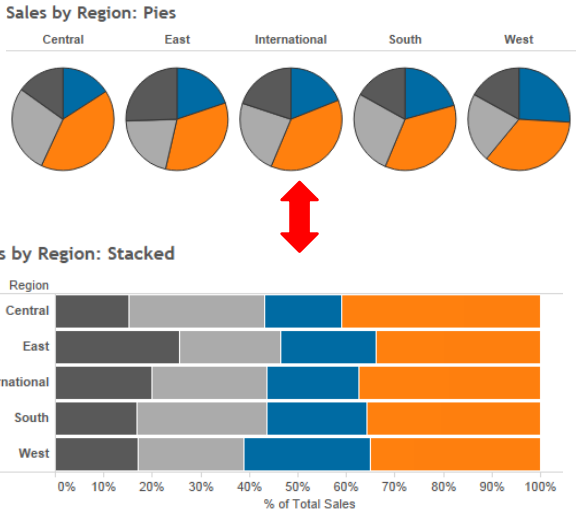


Compare with 100%
Values are in percentage

Image sources: https://qph.fs.quoracdn.net/main-qimg-d91aedca098bbf3250b5c38ad9b1a02f
http://speakingpowerpoint.files.wordpress.com/2011/05/slide4.jpg

Pie vs. Stacked Bar Graph

- Example

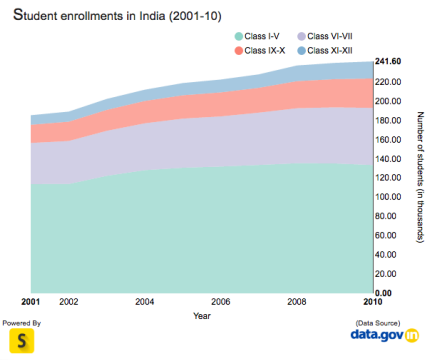


Compare with 100%

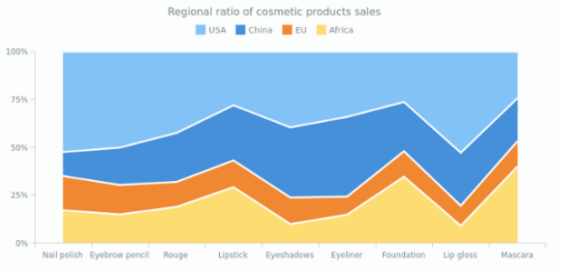
Image source: http://gravyanecdote.com/tableau/pie-or-stacked-bar/

Stacked Area Graph

- Can compare with total summation or 100%



Compare from total sum



Compare from 100%

Image sources: https://blog.socialcops.com/academy/resources/visualizing-time-series-data/
https://www.anychart.com/products/anychart/gallery/Area_Charts/

Radar Graph

- Plot of multi-variate data
- Place variables around center
  - Variables can be nominal data
- Use same scale to compare each category

Employment Candidate Review

Rating Areas	Candidates: Karen Fortou	Mike Rafun	Jack Nymbul
Experience	4.00	4.50	2.50
Communication	3.50	2.00	5.00
Friendliness	4.00	2.00	4.50
Subject matter knowledge	4.00	5.00	2.50
Presentation	3.00	1.50	2.75
Education	3.50	4.50	2.00
Average Rating	3.67	3.25	3.21

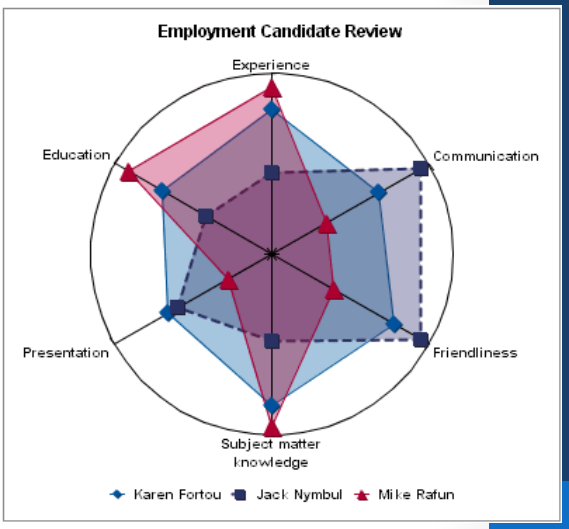


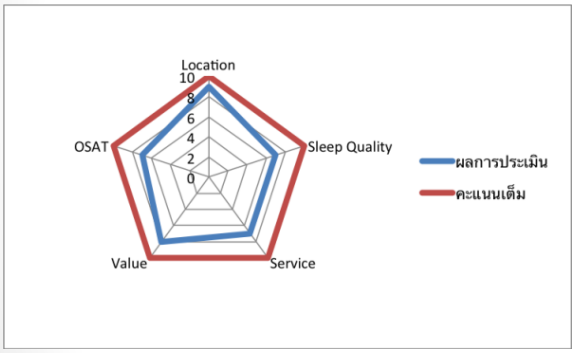
Image source: http://www.perceptualedge.com/example4.php

Example : Radar Graph

- Hotel Comparison

OSAT = overall satisfaction
Value = money-worth

โรงแรม Kinuya Hotel ทำเลสะดวก ทั้งเดินทาง จับจ่ายซื้อของ และกลับนาริตะ



โรงแรม Asakusa Hotel Wasou ห้องแคปซูล หญิง 900 บาท/คืน

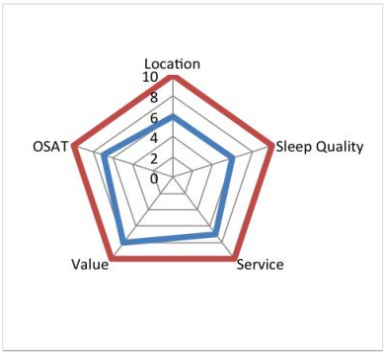


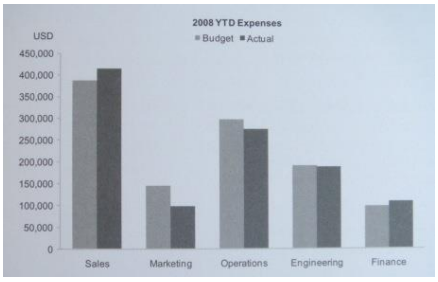
Image source: http://www.japan50.com/2014/kinuya-hotel/
http://www.japan50.com/2014/asakusa-hotel-wasou/

Outline

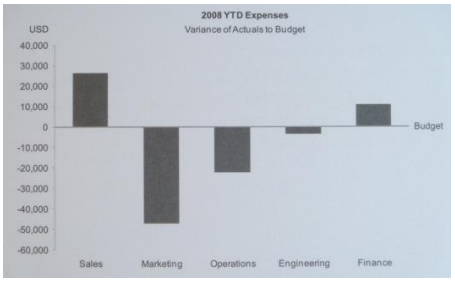
- Introduction
- Time-series
- Part-to-whole and Ranking
- Deviation

Deviation Analysis

- How much values are different from reference values
- Example: budget vs. actual expenses



Original graph



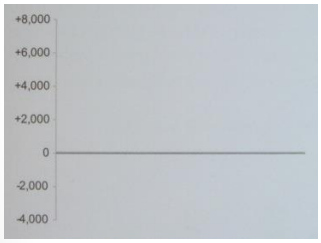
Display difference amount between budget and actual

Reference = budget

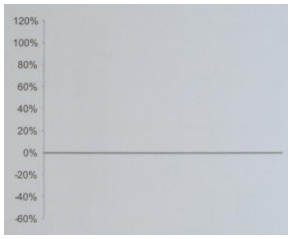
Image source: Figures 9.1, 9.2 [1]

Reference Line

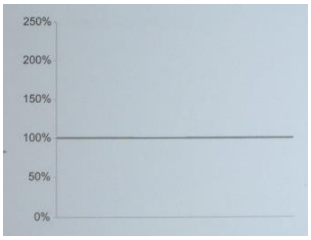
- Required feature in deviation
- Commonly used values
  - Zero
  - Zero or hundred percent
  - Exact reference value



Zero



0%



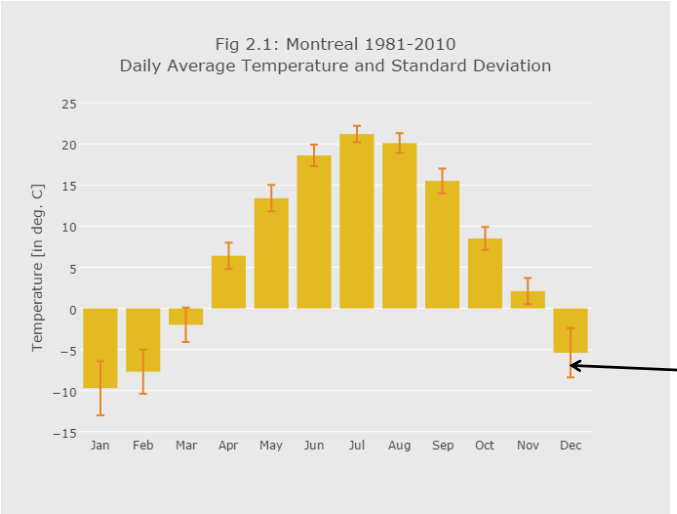
100%

Image source: Figure 9.3 [1]

Deviation Displays

- Bar graph
- Line graph

Example 1 : Bar Graph



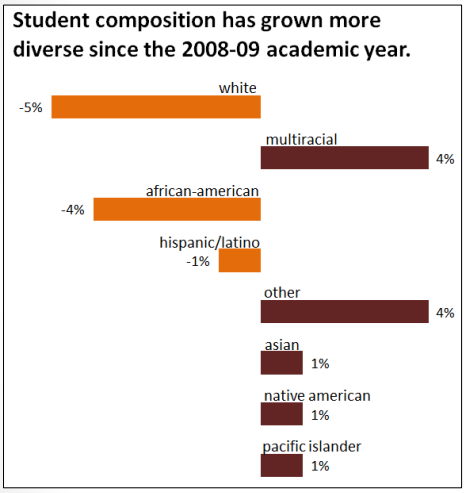
- Compare average temperature in 12 months

Reference = 0°C

Added confidence interval

Image source: https://plot.ly/python/bar-charts-tutorial/

Example 2 : Bar Graph



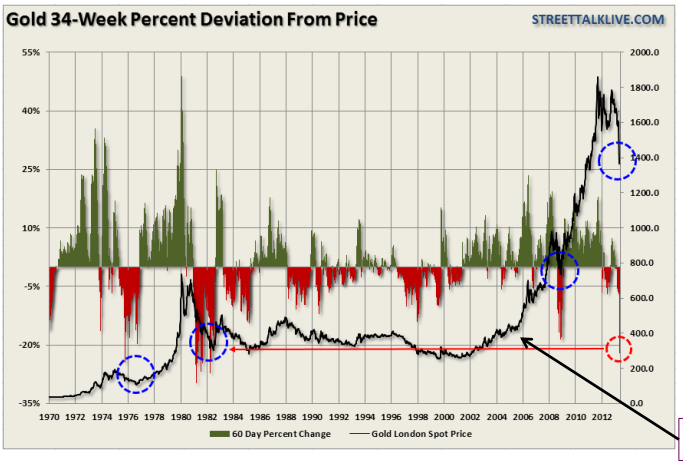
- Compare change in student composition
- Sort from maximum to minimum **absolute change**
  - Put majority change first
- No line plotted in graph

Reference = 0%

Image source: http://betterevaluation.org/evaluation-options/split_axis_bar_graph/

Line Graph

- Example : Change of gold price between 1970-2013



Reference: 34-week moving average

Price increase

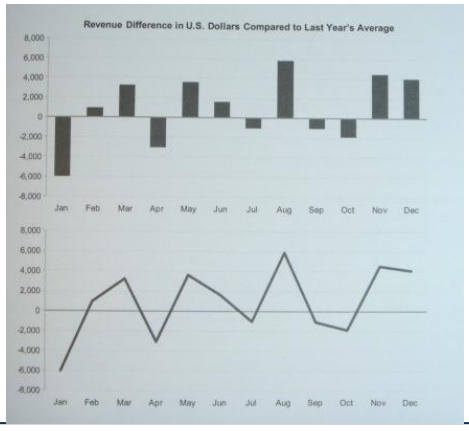
Price decrease

Actual gold price

Image source: http://www.pragcap.com/the-gold-crash-what-its-not-telling-us/

Bar vs. Line Graph

- Bar graph focuses on exact values
- Line graph focuses on change in deviation
- Example 1:



Reference: Average revenue from last year

Compared: Current month's revenue

Bar graph

Line graph

Image source: Figure 9.6, [1]

Reference

1. Now You See It: Simple Visualization Techniques for Quantitative Analysis, Stephen Few, Analytics Press, 2009
2. Multidimensional Visualization, Maneesh Agrawala, 2011
3. http://saravanan-thirumuruganathan.github.io/cse5334Spring2015/slides/02_IntroToViz/2_IntroToViz_final.pdf