



澳門大學
UNIVERSIDADE DE MACAU
UNIVERSITY OF MACAU

Graphic Types

CISC7204: DATA SCIENCE & VISUALIZATION

Derek F. Wong

NLP²CT – Natural Language Processing &
Portuguese-Chinese Machine Translation Research Group

derekfw@um.edu.mo

E11-4010 (Ext: 4478)

Office Hours: Thu – 16:00~17:30, Fri 11:00~12:30



Natural Language Processing & Portuguese –
Chinese Machine Translation Laboratory
自然語言處理與中葡機器翻譯實驗室

Content

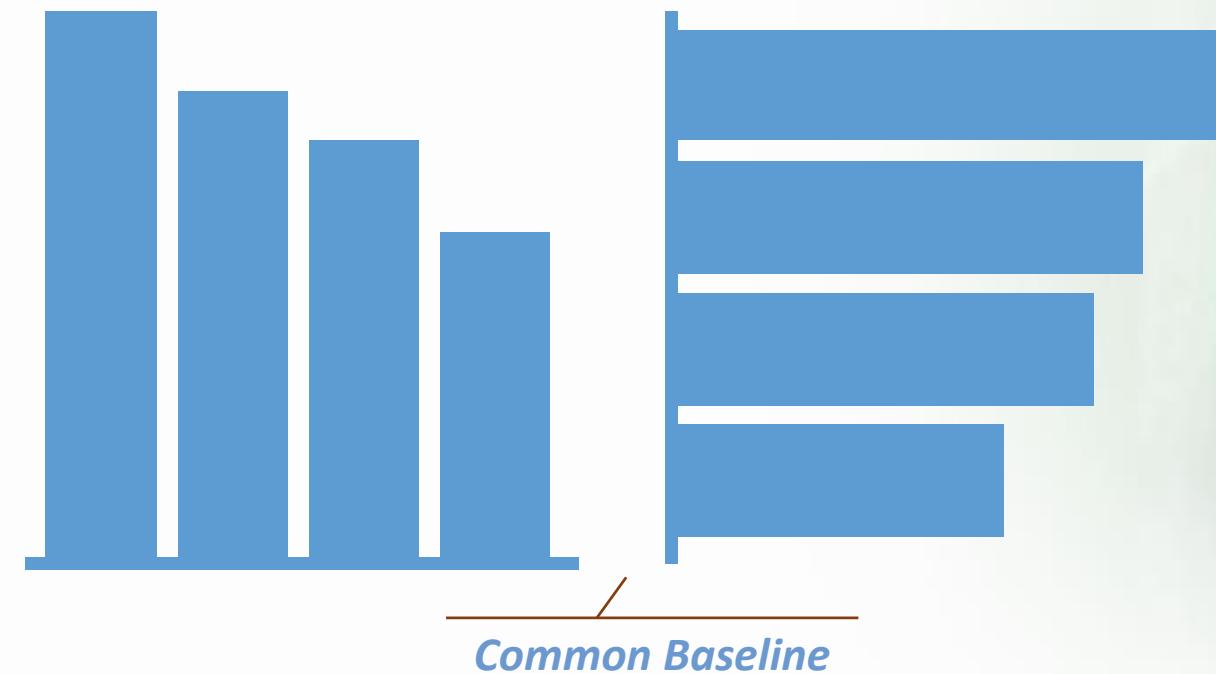
Chart Types

- Learn *basic chart types*
- How chart types *encode* data and *be used*
- Discuss *problems* with *various chart types*
- Examples of various chart types through *compare* and *contrast*
- *Choose* the right graph type

Chart Types

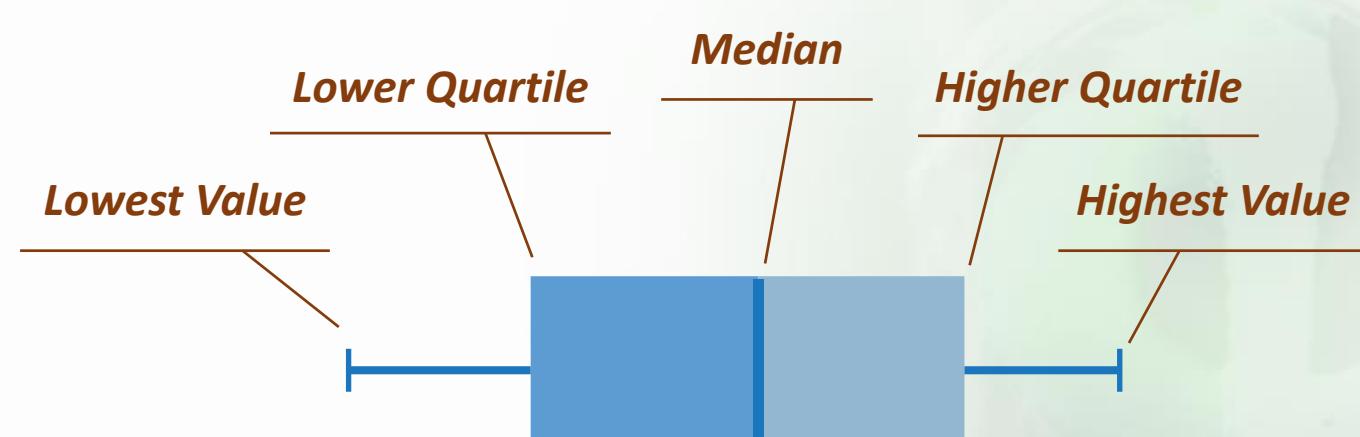
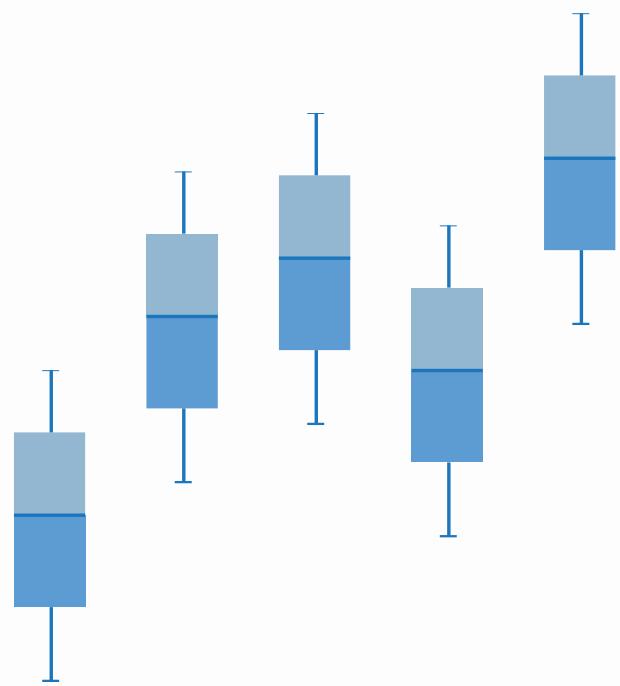
Bar Chart

Encodes data using *height/length* of bar and shows *categorical* comparisons



Box Plot

Box Plot encodes data using *position* and *height/length* to show the *distribution* of the data

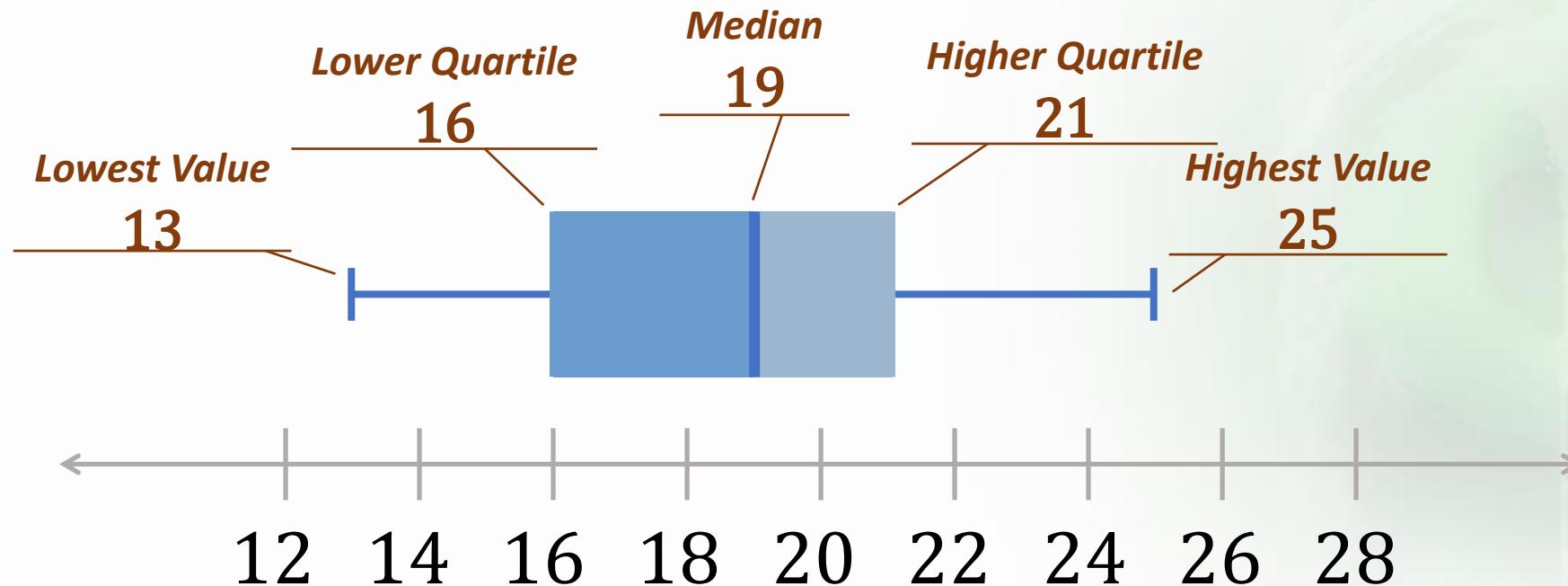


Box Plot

How to Draw?

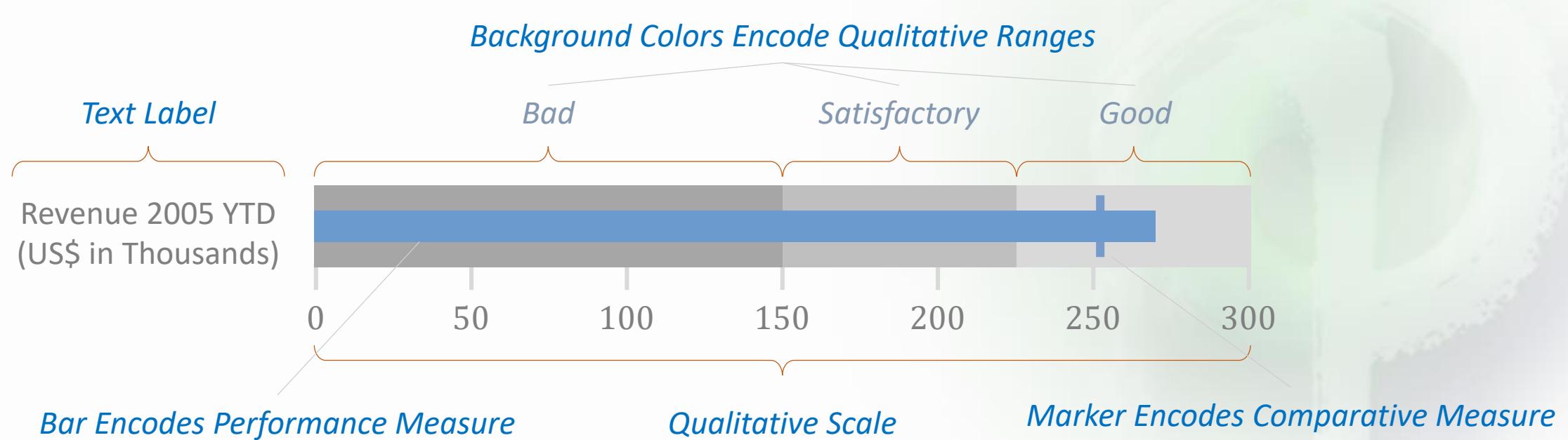
Given: 21, 25, 15, 13, 17, 19, 19, 21

Order: 13, 15, 17, 19, 19, 21, 21, 25



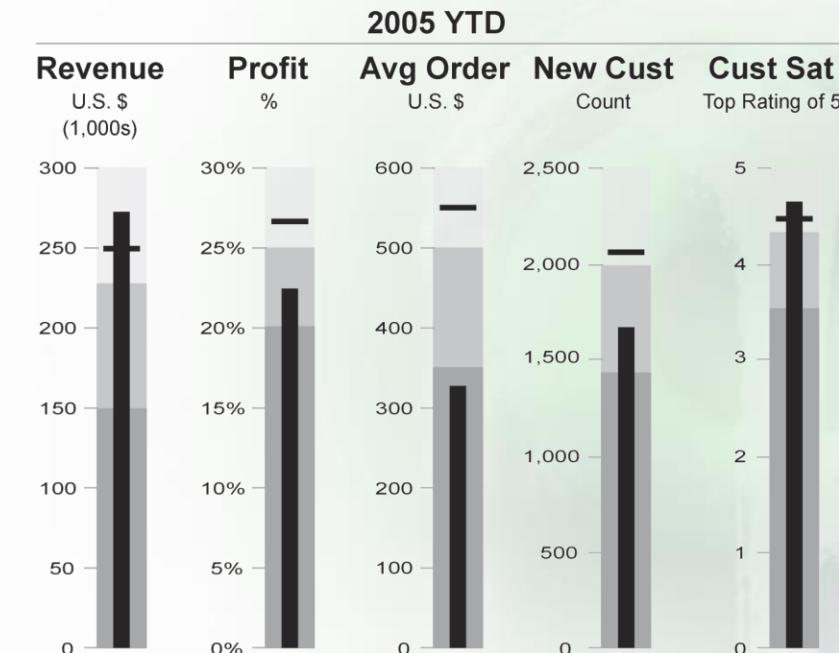
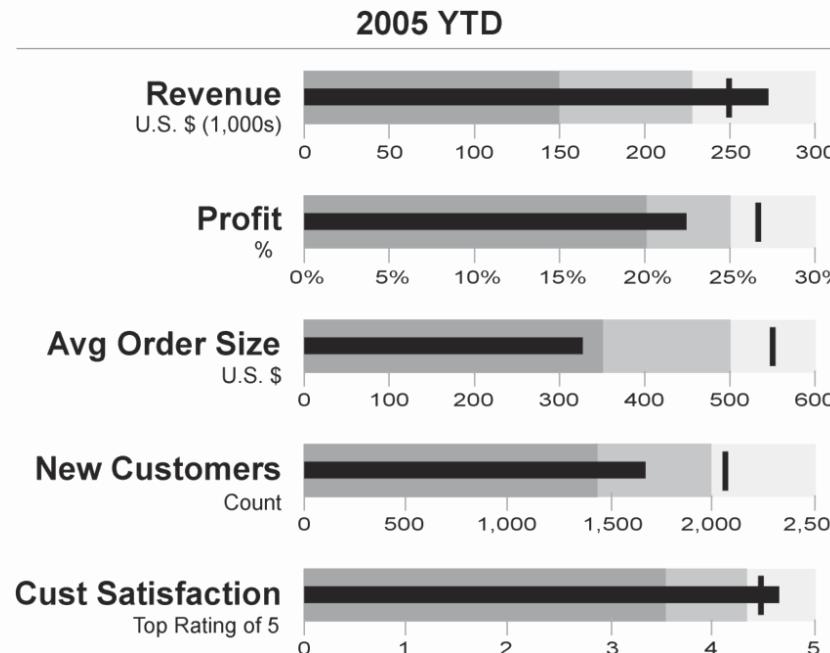
Bullet Graph

Bullet graph data using *length/height*, *position* and *color* to show actual compared to *target* and *performance bands*



Bullet Graph

Bullet graph data using *length/height*, *position* and *color* to show actual compared to *target* and *performance bands*

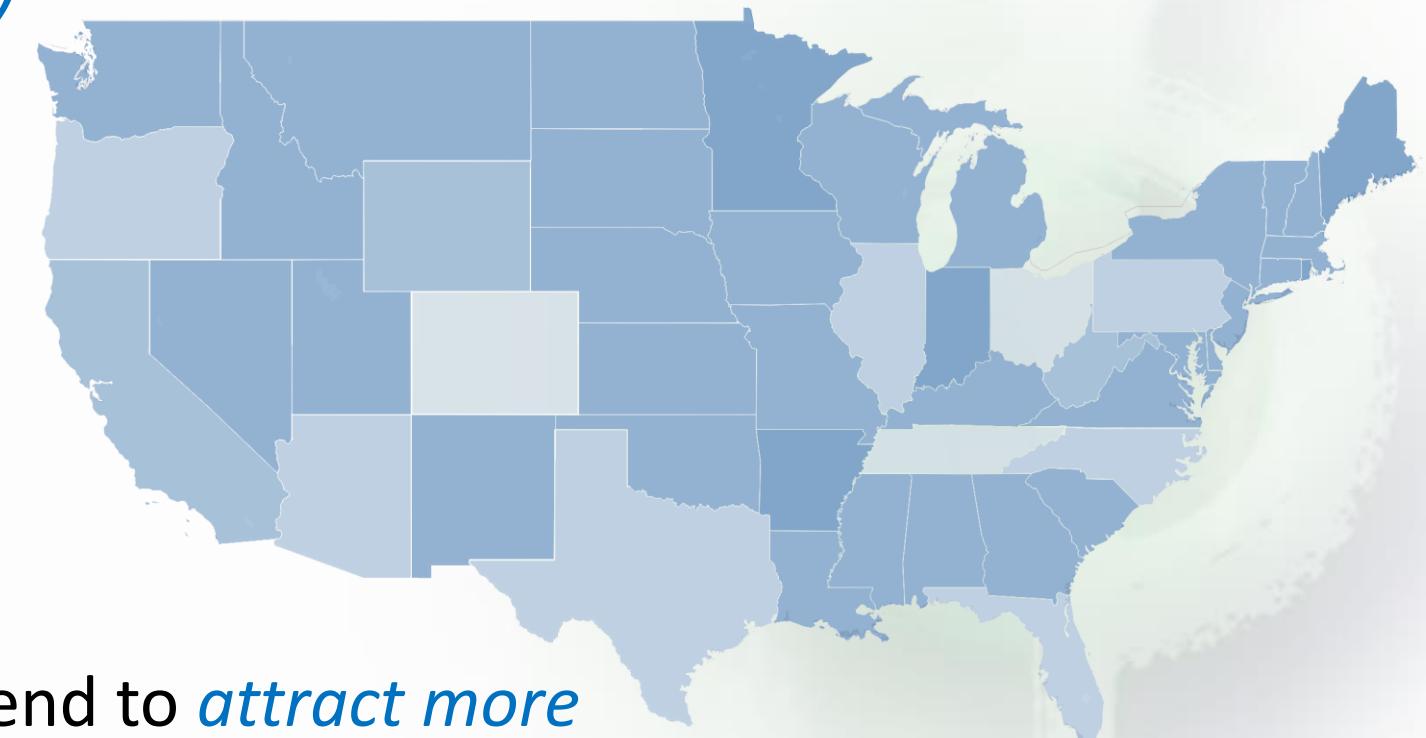


Choropleth Map (Shaded Map)

Choropleth Map encodes data using *color* and *position* to show *data geographically*

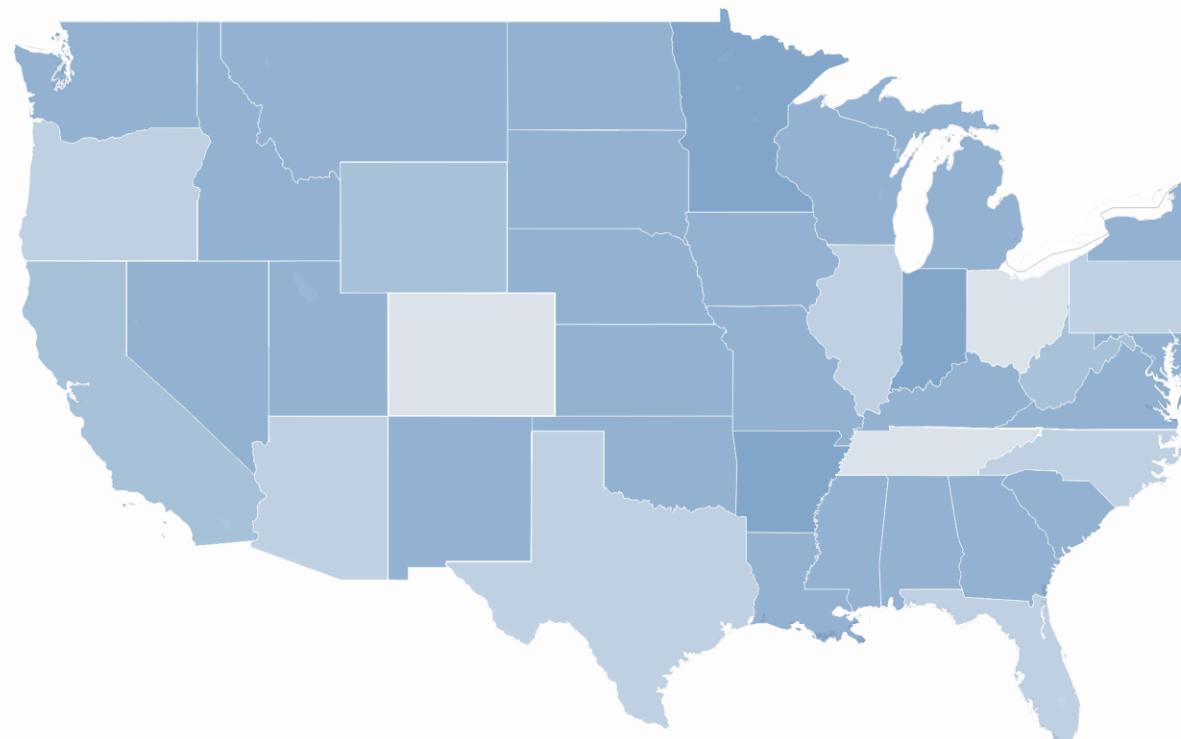
Weaknesses:

- Impression of *uniform distribution*
- *Exact values* cannot be determined
- *Larger enumeration units* tend to *attract more attention*



Choropleth Map (Shaded Map)

Choropleth Map encodes data using *color* and *position* to show *data geographically*

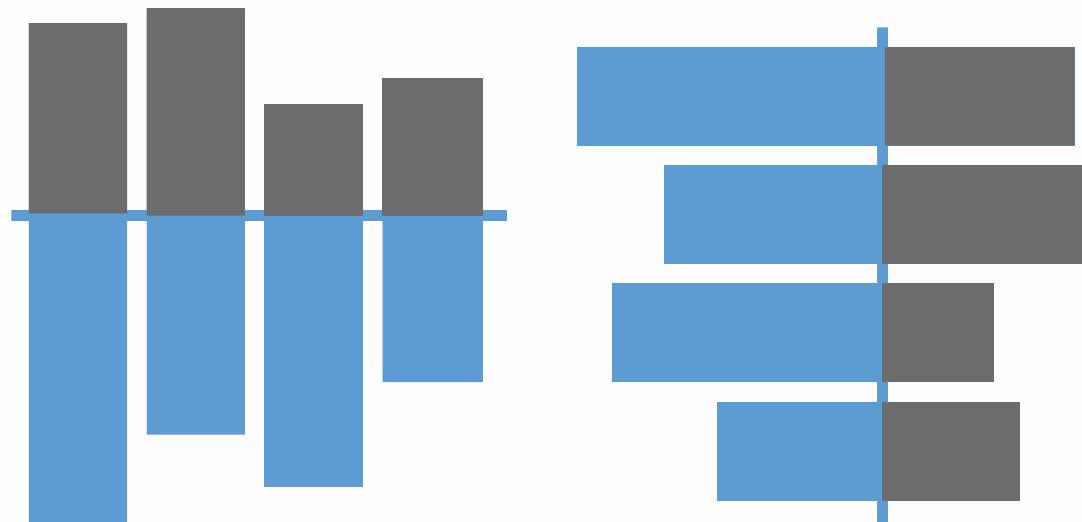


Better for *standardized data* (instead of *raw data*):

- *Density*
- *Proportion*
- *Relationships*
- *Statistical summaries*

Diverging Bar Chart

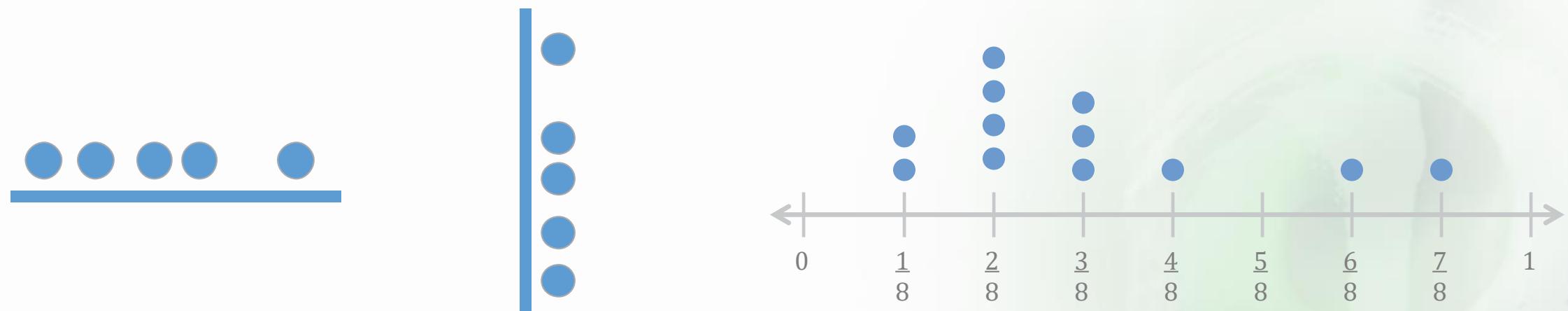
Diverging Bar Chart encodes data using *height/length* of bar diverging from a *midpoint* to show *categorical comparisons*



The **drawback** is that it is not easy to compare the values *across dimension members*

Dot Plot

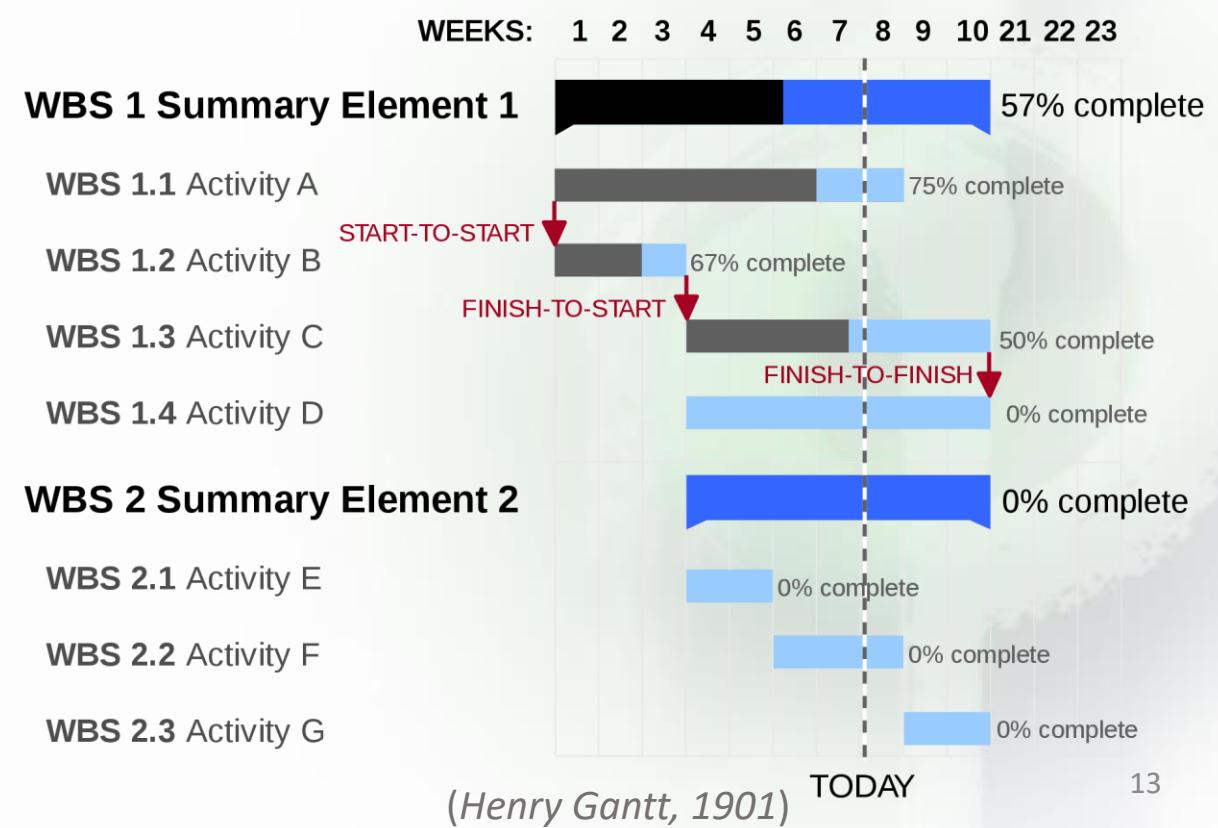
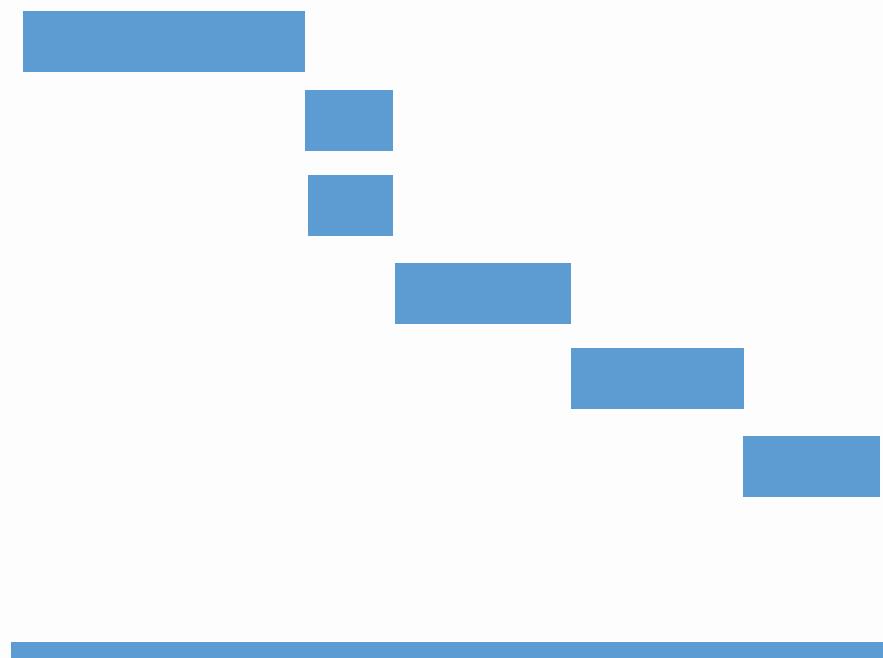
Dot Plot encodes data using *position* to show the *comparisons*



A *data* display in which each *data item* is shown as a *dot* above a *number line*

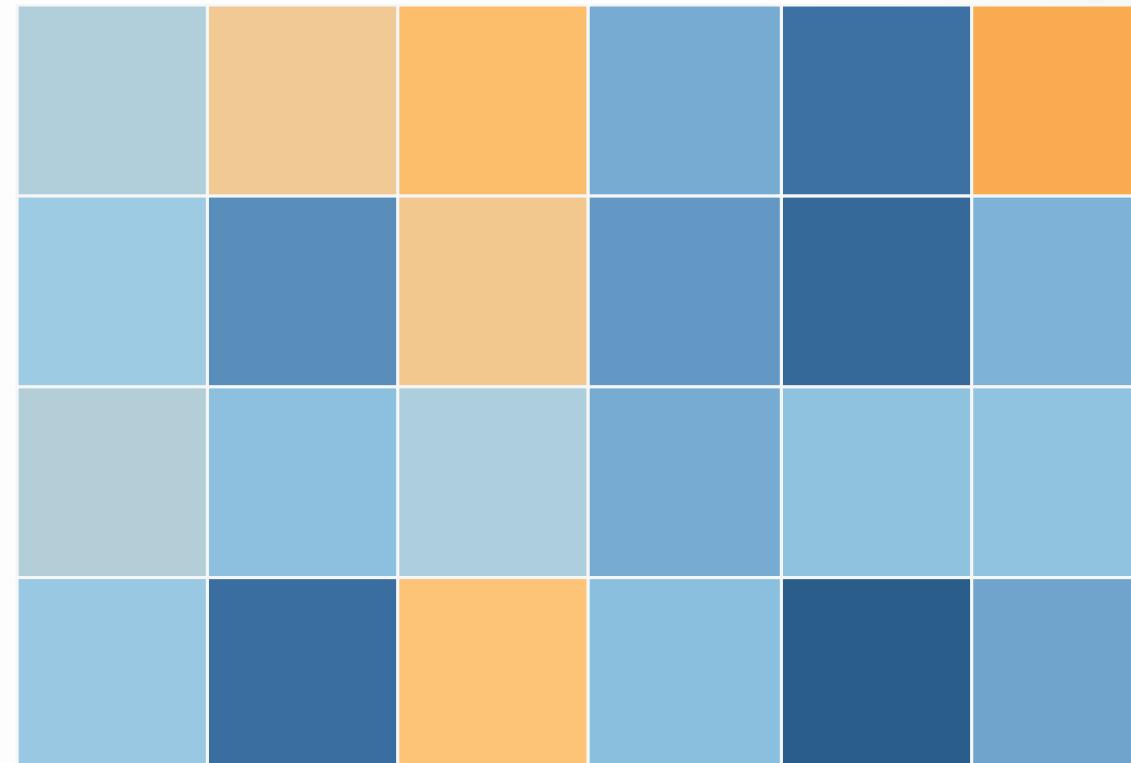
Gantt Chart

Gantt Chart encodes data using *length*, *height* and *position* to show *amount of work* completed in *segments of time*



Heat Map

Heat Map encodes a *data table* using *color* to *highlight the differences* in the table *without numbers*



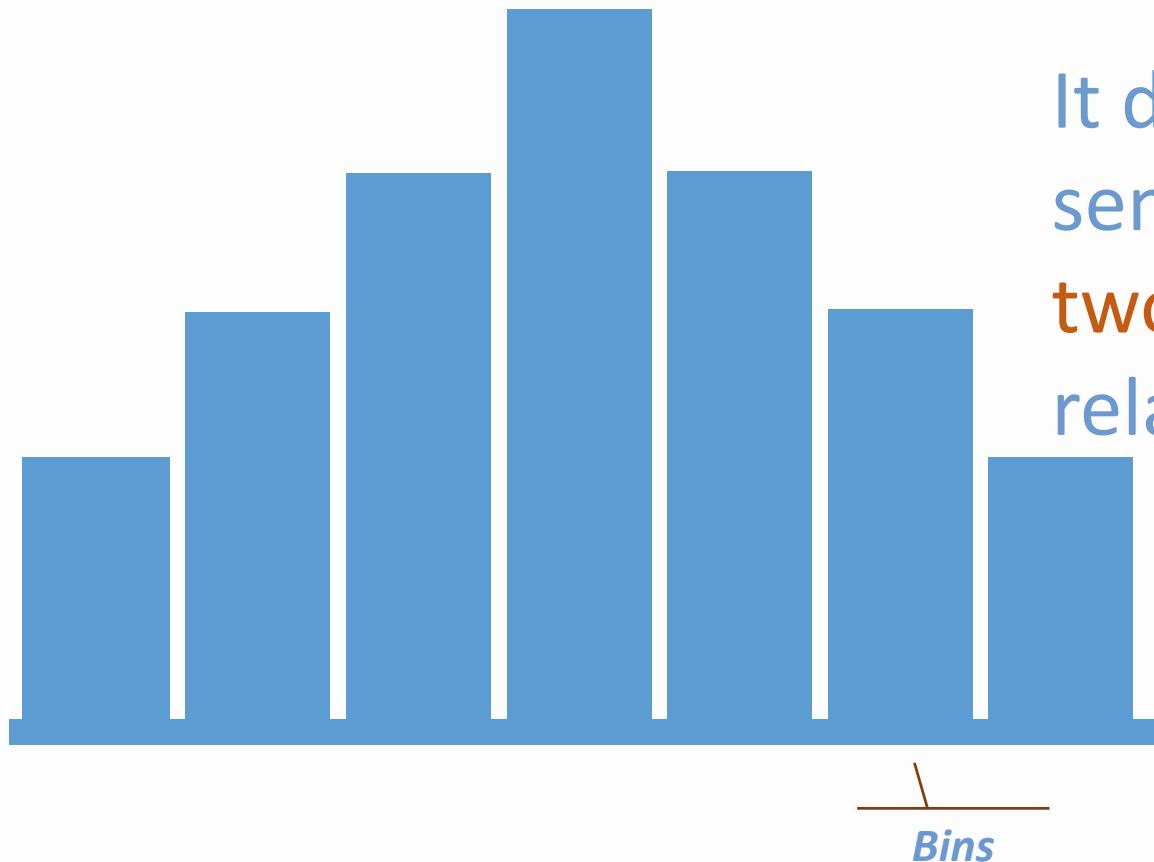
Highlight Table

Highlight table encodes a *data table* using *color* to *highlight the differences* in the table *numbers*

\$29,071	\$17,307	\$30,073
\$2,603	\$2,353	\$5,079
\$66,106	\$53,891	\$42,444
\$20,173	\$14,151	\$26,664
\$100,615	\$58,304	\$98,684
\$71,613	\$35,768	\$70,533
\$10,760	\$8,319	\$18,127
\$39,140	\$43,916	\$84,755

Histogram

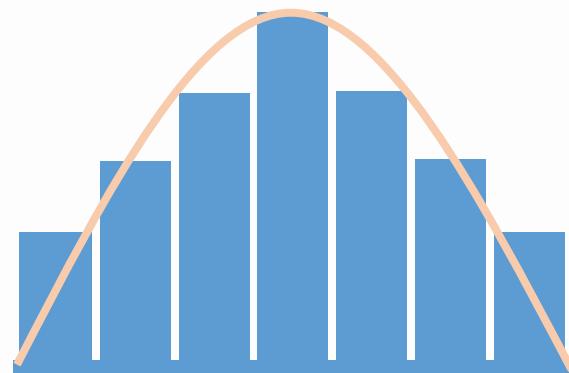
Histogram encodes data using *height* and shows a *distribution*



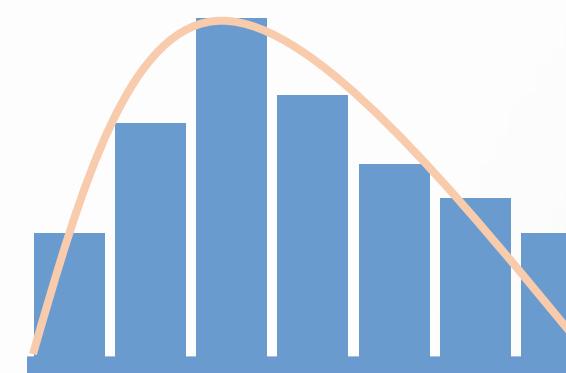
It differs from a **bar graph**, in the sense that a bar graph relates **two variables**, but a histogram relates **only one**

Histogram

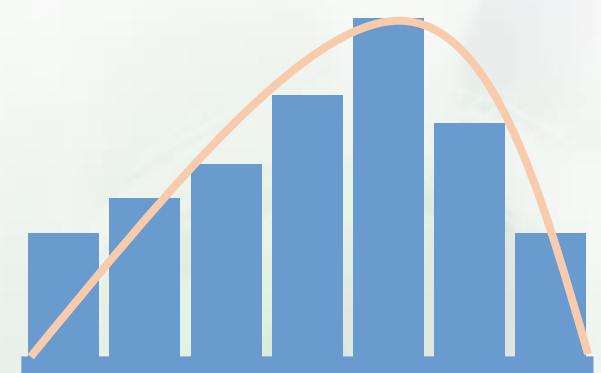
The *patterns* of histogram include:



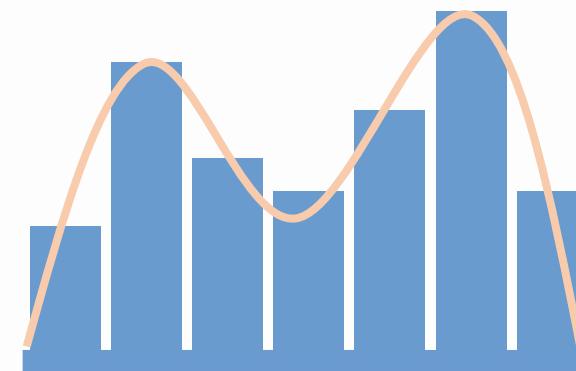
Normal Distribution



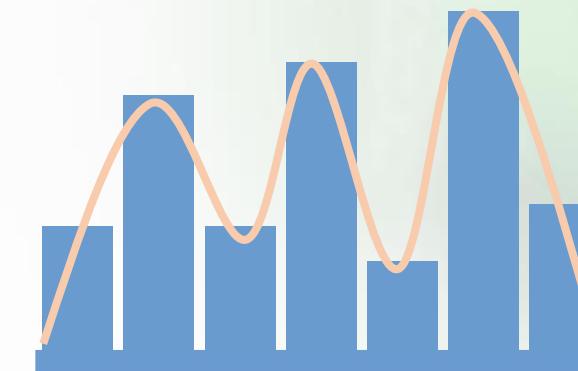
Positively Skewed



Negatively Skewed



Bimodal Distribution



Multimodal Distribution

Line Chart

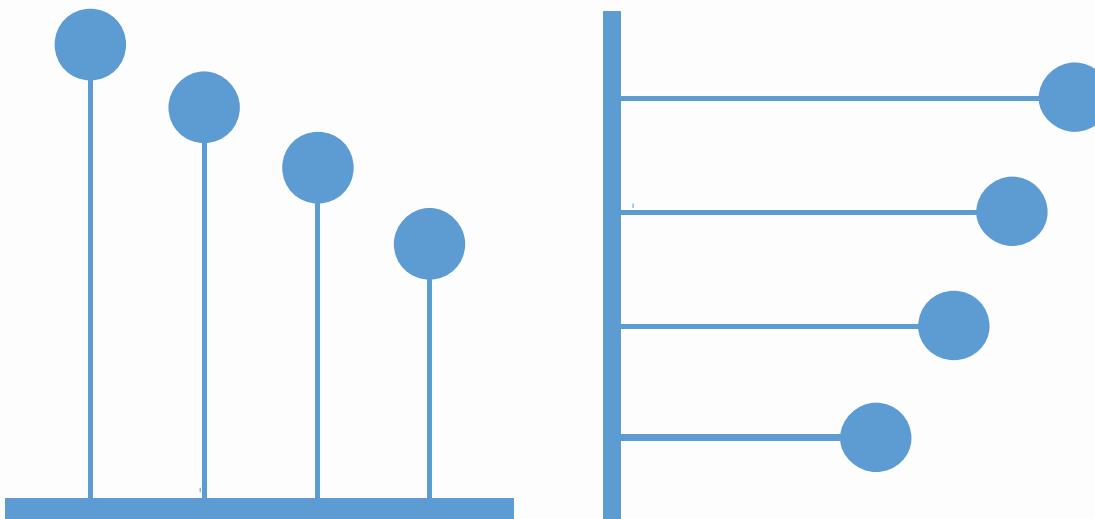
Line chart encodes data using *position* and often shows *trend over time*



Line charts show the change over time for one or more series (sales per hour). The line connects each data point in the series (shown or not)

Lollipop Chart

Lollipop chart encodes data using *height* or *length* of bar and shows *categorical comparisons*

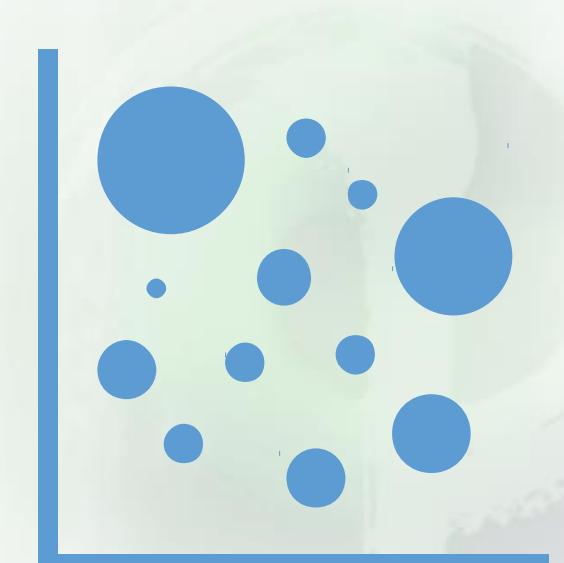
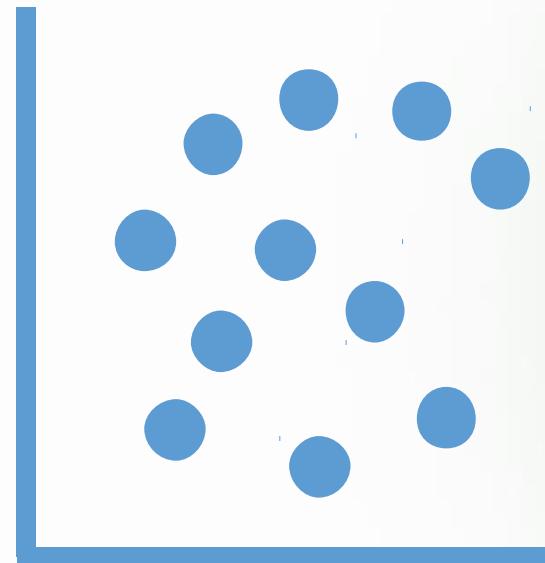


In a busy graph, Lollipop uses thin line, resulting in giving more white space between bars

The center of circle is difficult to judge, making it imprecise compared to *bar chart*

Scatter Plot

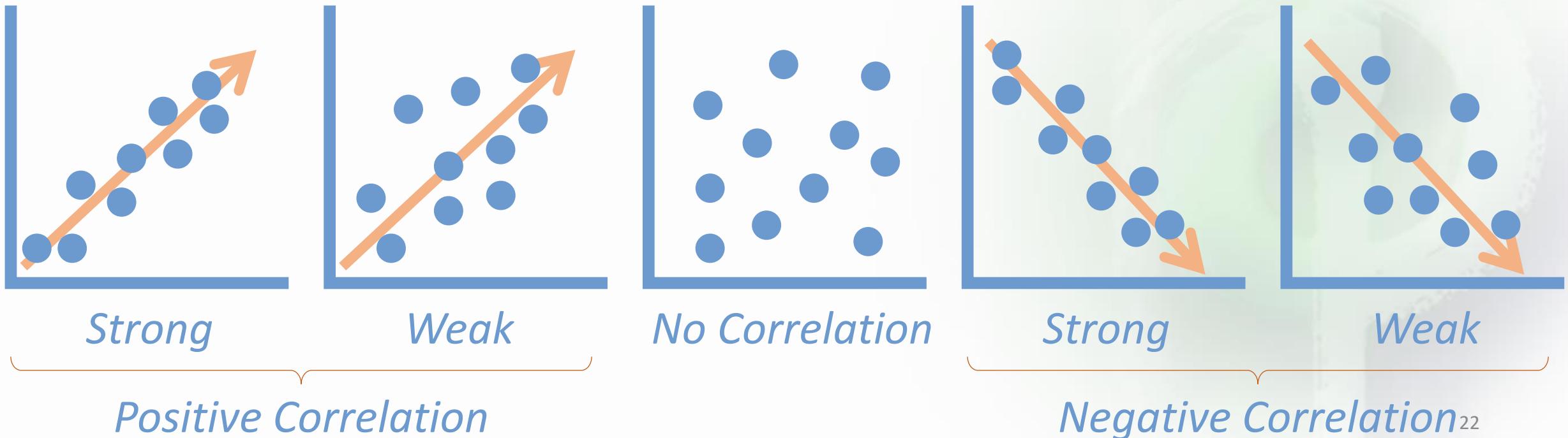
Scatter Plot encodes data using *position* to show the *relationship* between *two variables*. *Size* can also be used to show a *secondary comparison*



Scatter Plot

Correlation describes the *type of relationship* between *two data sets*, represented by *the line of best fit* comes *closest to all the points* on a scatter plot

The correlation can help analyze trends and make predictions

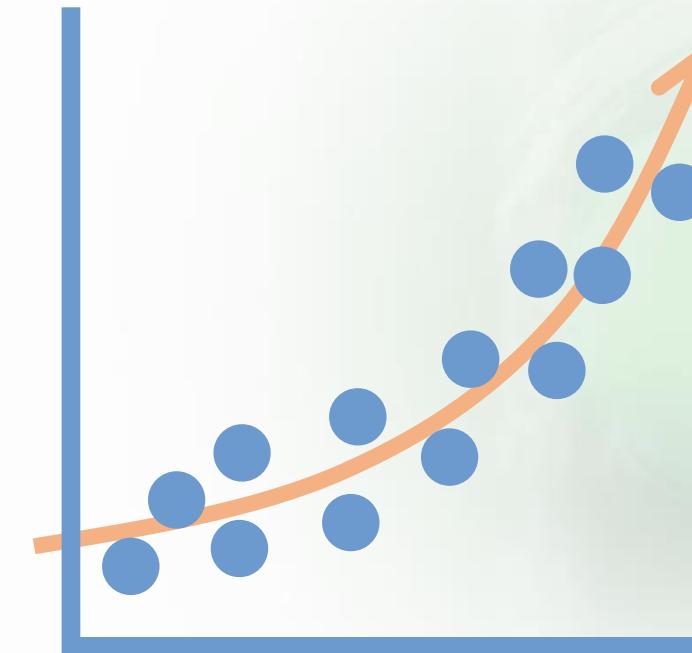
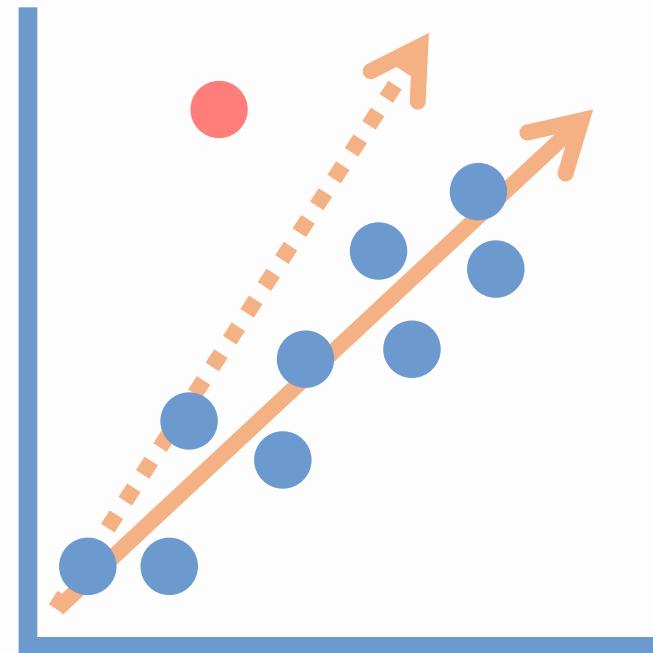


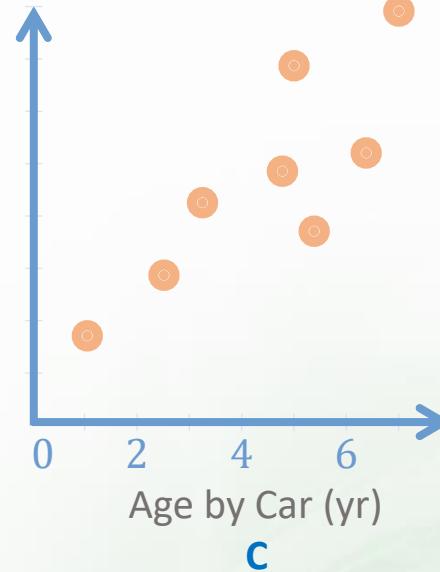
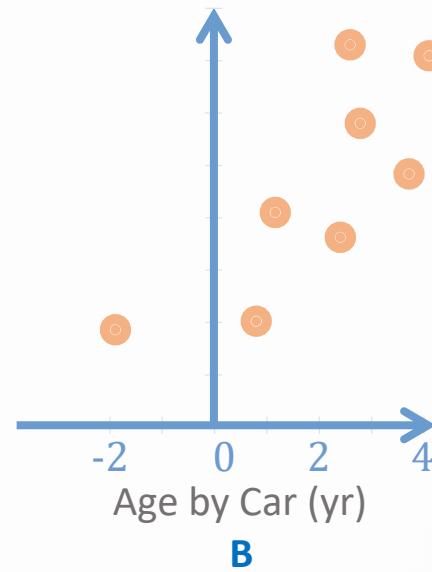
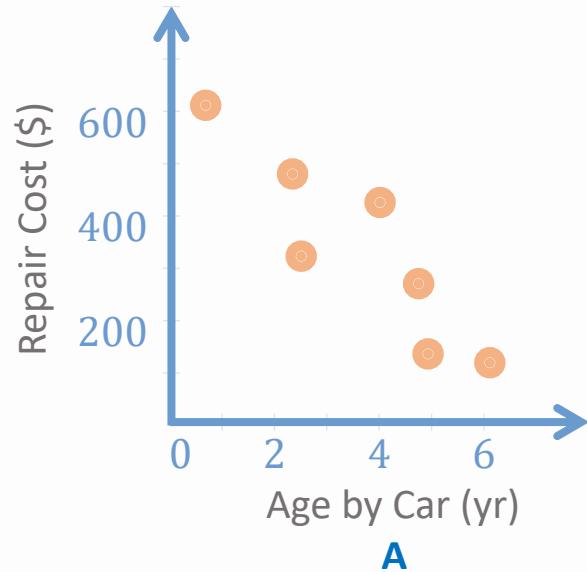
Positive Correlation

Negative Correlation ²²

Scatter Plot

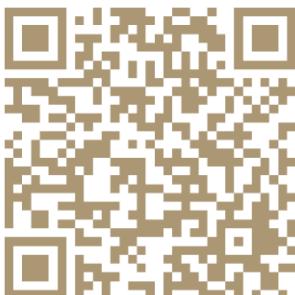
An *outlier* is a data point that is *extremely spread out* from the rest of the data or *does not follow* the general *trend of the data*





In-class Exercise

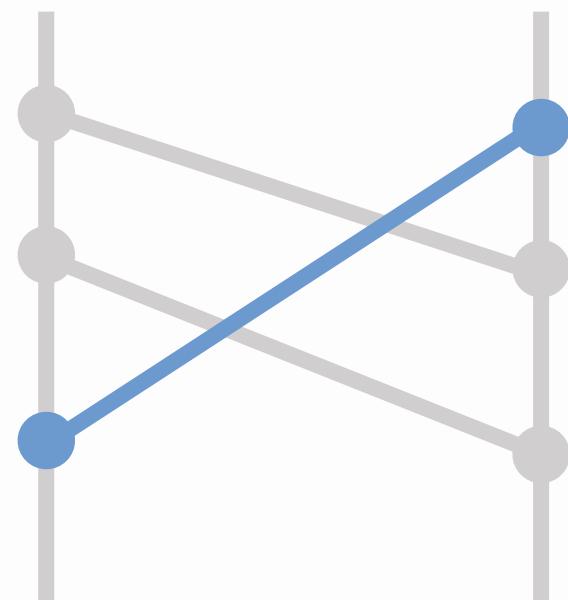
Choose the scatter plot that best represents the *relationship* between the age of a car and the amount of money spent each year on repairs, and explain!



<https://ummoodle.um.edu.mo/mod/assign/view.php?id=1368324>

Slopegraph

Scatter Plot encodes data using *position* to show *quantitative comparison* or *rank*, typically between *two time periods*

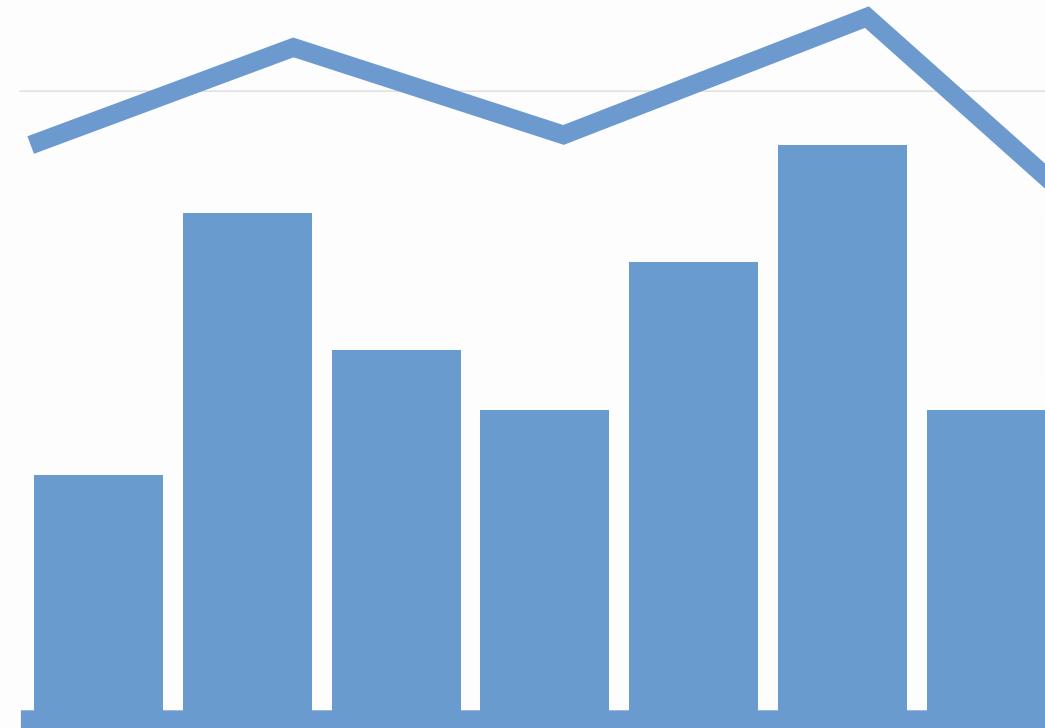


Show relative **increases** and **decreases** or **differences** across various categories between the two data points

In addition to the **absolute**, the lines that connect them give you the visual increase or decrease in **rate of change**

Sparkline/Sparkbar

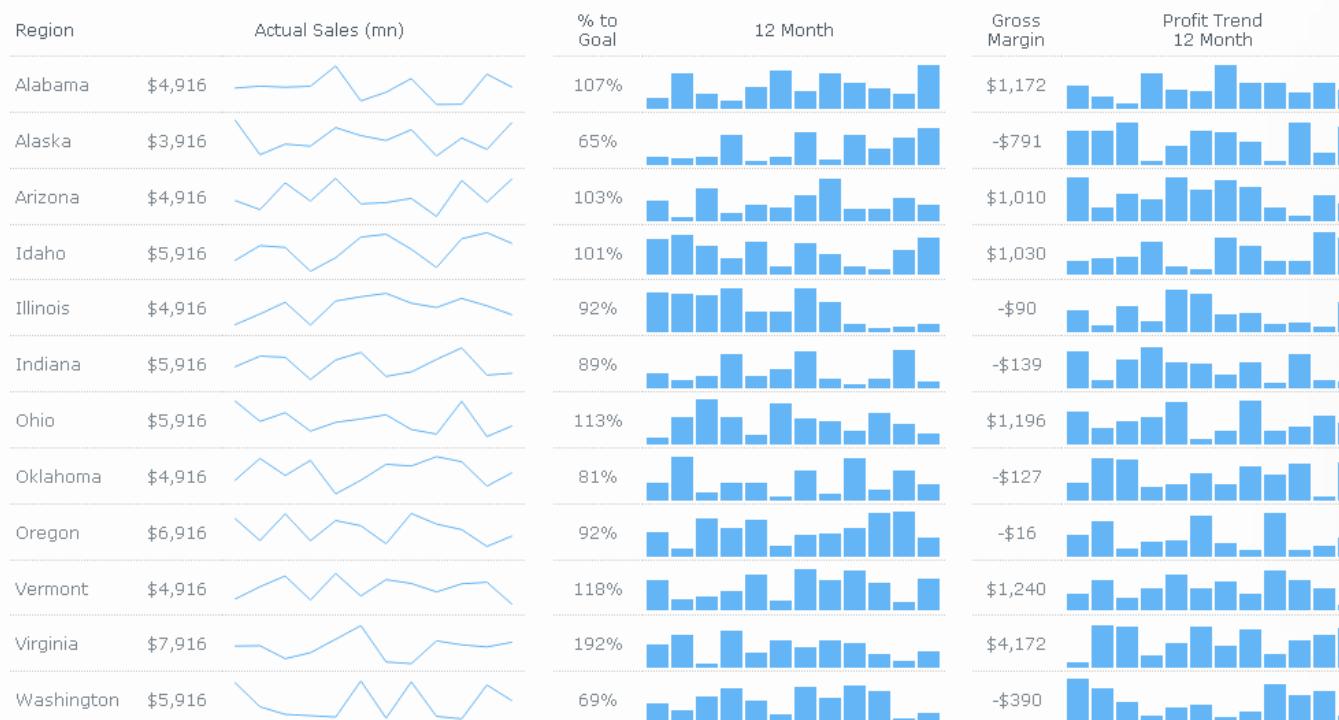
Sparkline/Sparkbar encodes data using *position* (line) or *height/length* (bar) in a *small, word-sized graphic*



Without axis or coordinates
to show the general shape
of the data

Sparkline/Sparkbar

Sparkline/Sparkbar encodes data using *position* (line) or *height/length* (bar) in a *small, word-sized graphic*



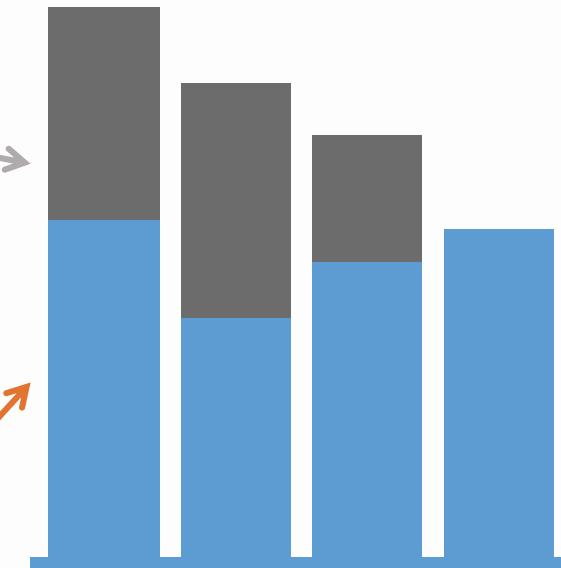
Without axis or coordinates
to show the general shape
of the data

Stacked Bar Chart

Stacked Bar Chart encodes data using *height* or *length* of bar and *color* by *segment* and shows *categorical* and *part-to-whole comparisons*

Not the stacked bars

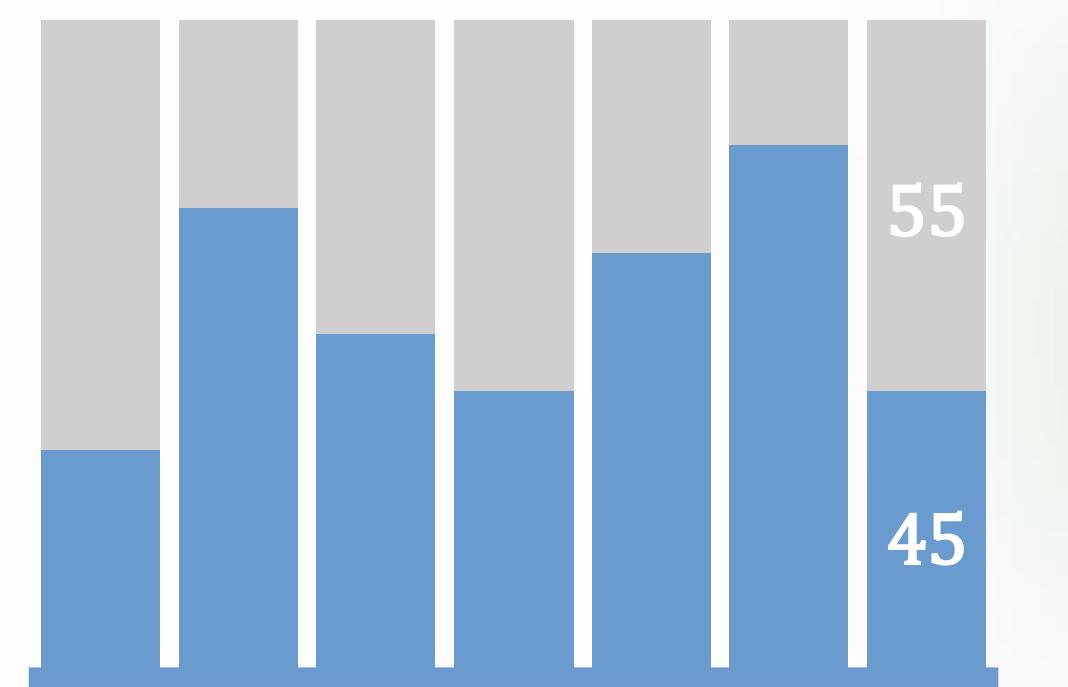
Easy to judge the lengths of the bottom bars on the baseline



Caution be careful not to slice stacked charts into too many segments

Stacked Bar Chart

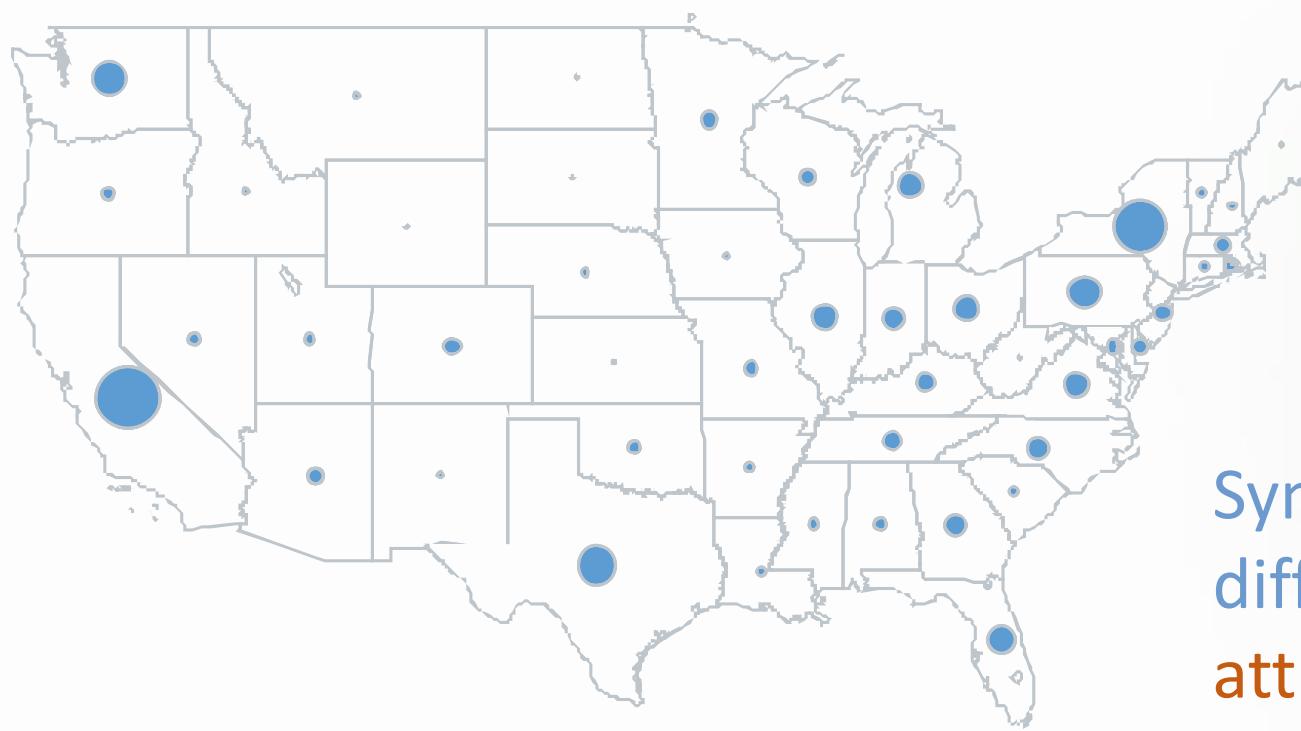
Stacked Bar Chart encodes data using *height* or *length* of bar and *color* by *segment* and shows *categorical* and *part-to-whole comparisons*



100% Stacked
bar chart

Symbol Map (Dot Map)

Symbol Map (Dot Map) encodes data using *position* to show *data geographically* and can also use *size* to show *quantitative data*



Geometric



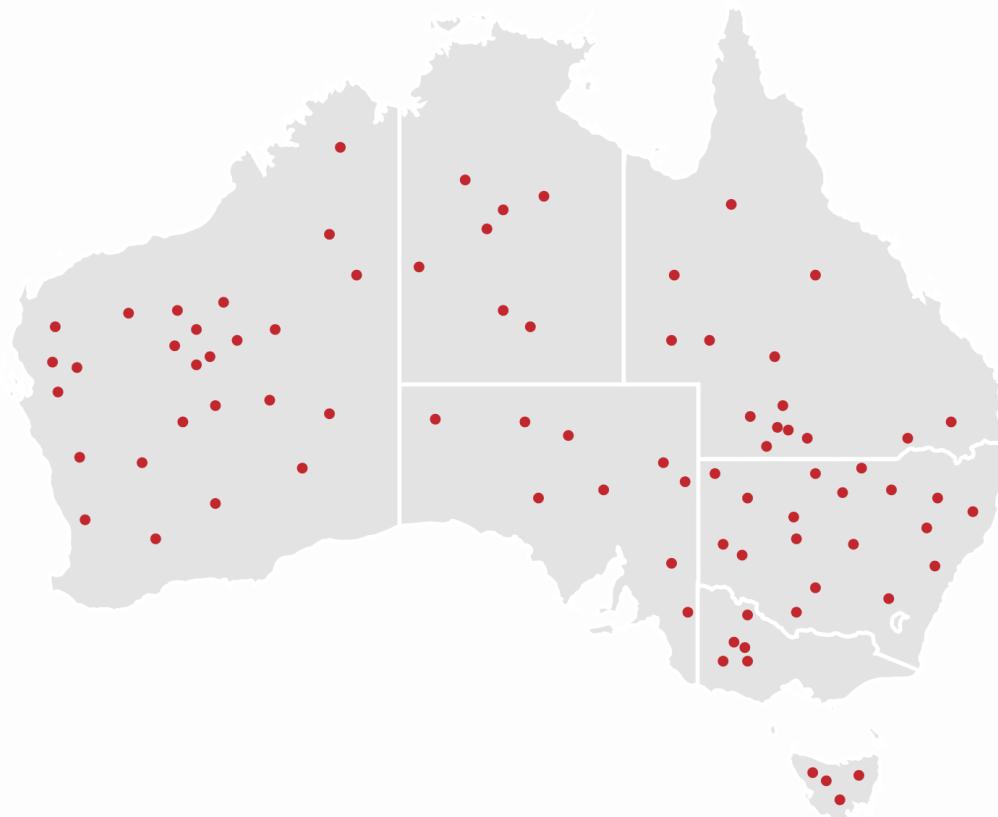
Pictographic



Symbols vary in size according to differences in magnitude of an attribute

Symbol Map (Dot Map)

Symbol Map (Dot Map) encodes data using *position* to show *data geographically* and can also use *size* to show *quantitative data*



Instead of using proportional symbols, one point represents a single count of object. Point is positioned at its exact location

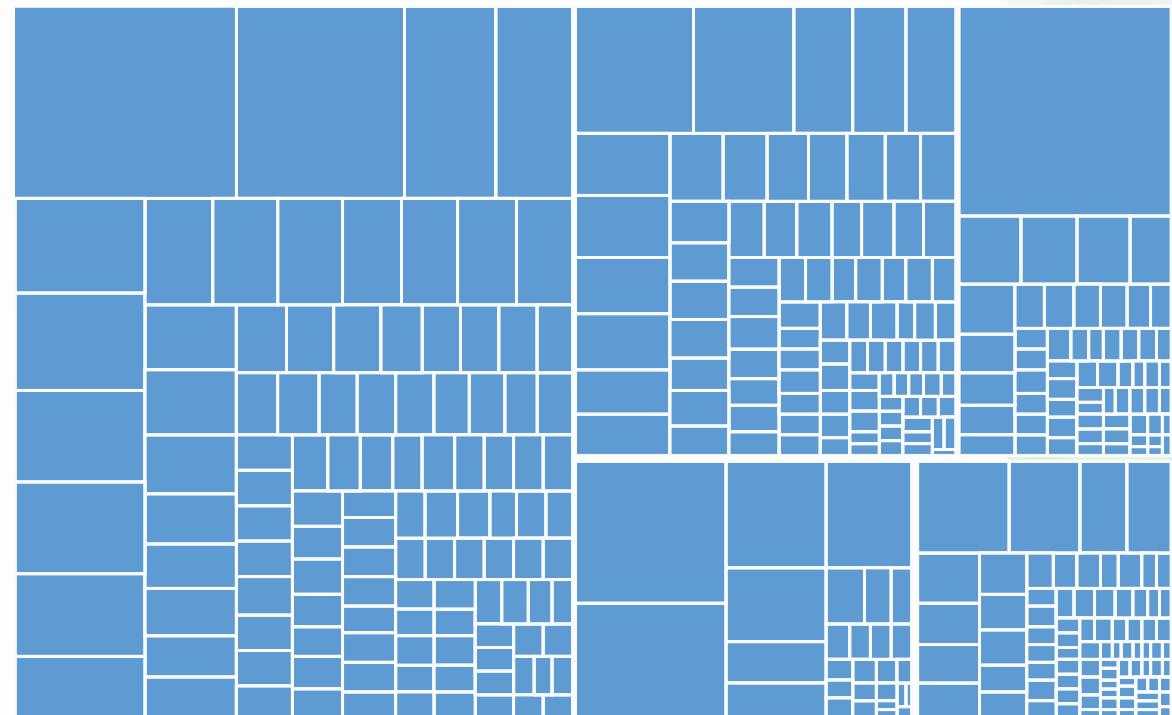
The marker does not size

Treemap

Treemap encodes data using *size* and *color* and is useful for *hierarchical data* or when there are a very *large number of categories* to compare

Show the **composition** of a whole when there are many components

Structural information is implicitly presented



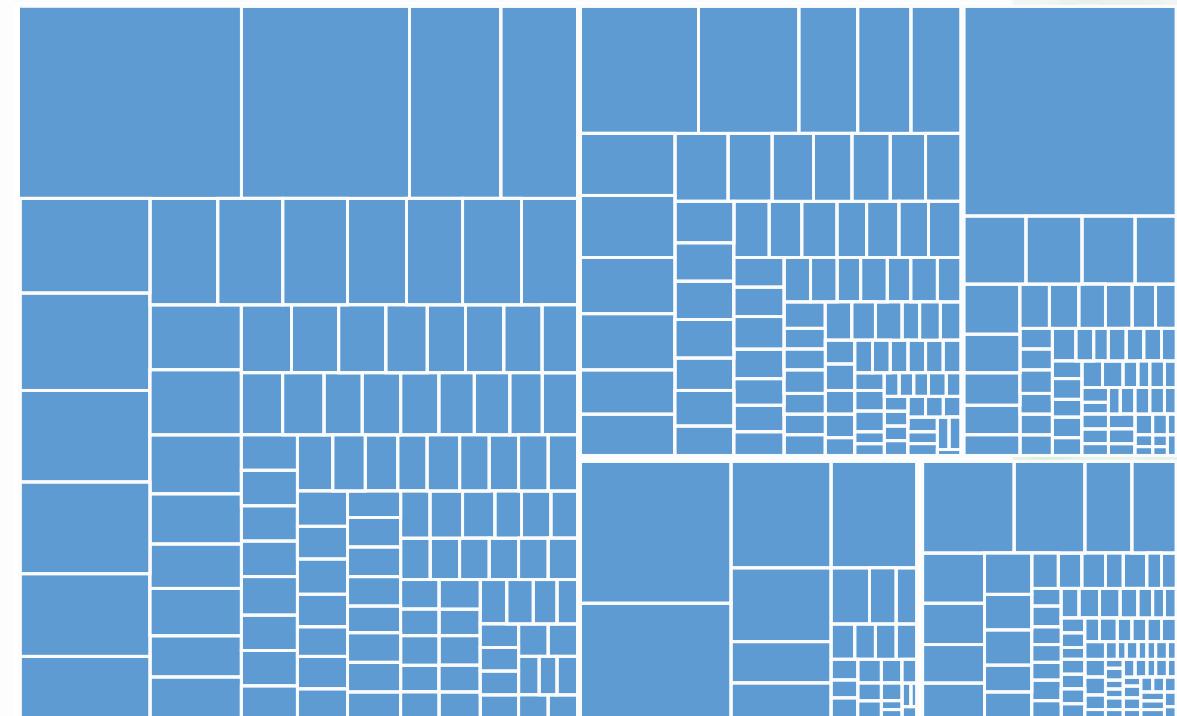
Treemap

Treemap encodes data using *size* and *color* and is useful for *hierarchical data* or when there are a very *large number of categories* to compare

Use display space **efficiently**

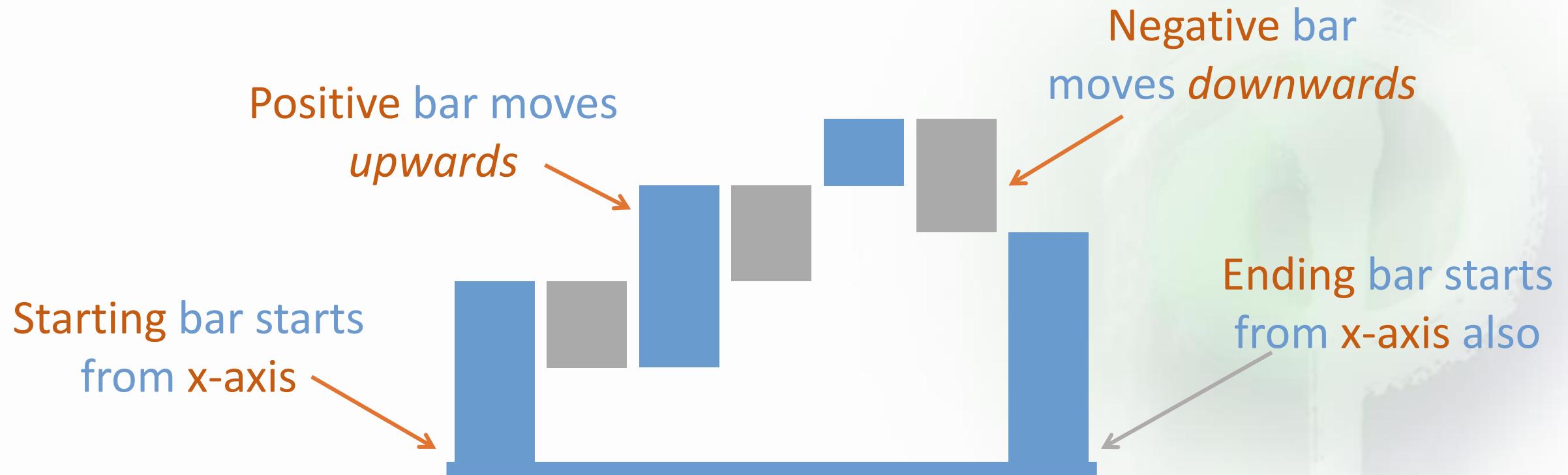
Provide an **overall (global)**
view of the entire hierarchy

Provides creative visual cues
to **communicate content**
information



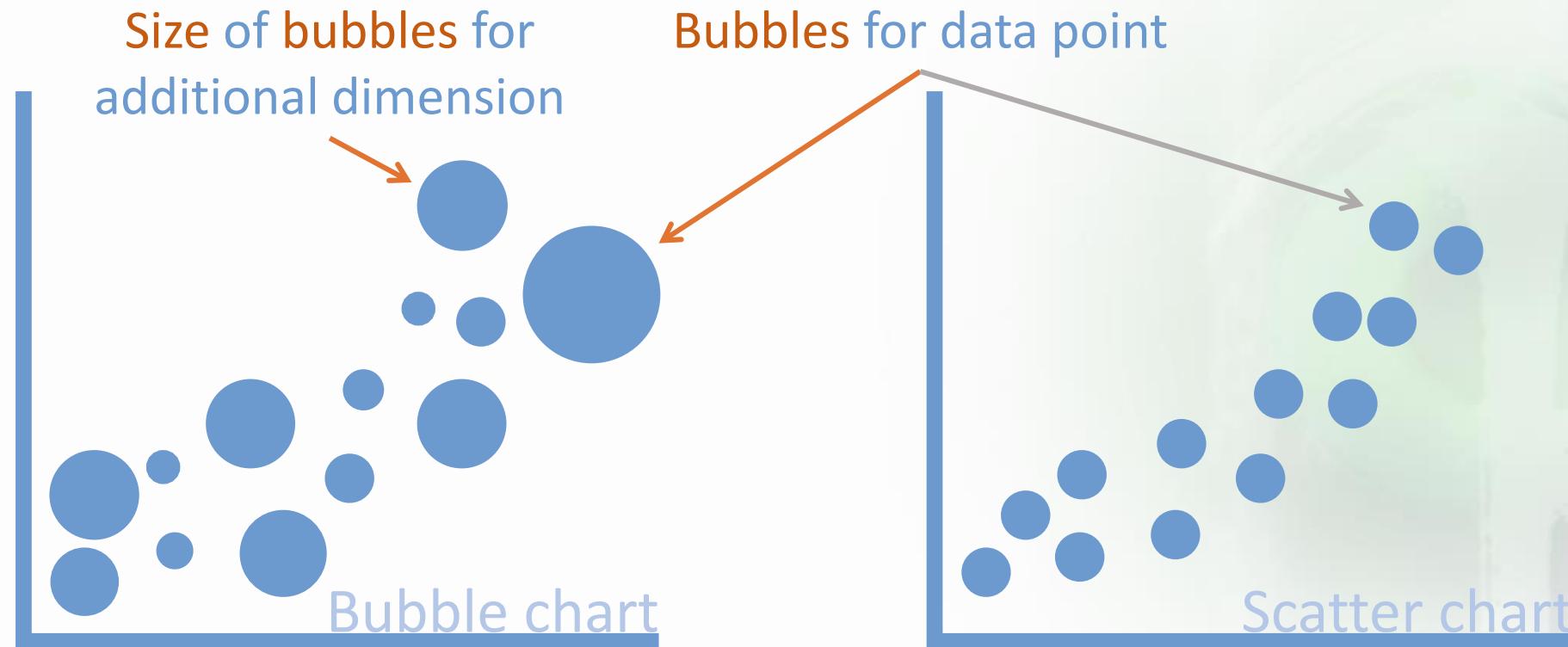
Waterfall Chart

Waterfall Chart encodes data using *height* and often *color* to show *increase* and *decrease* between *time periods* or *categories*



Bubble Chart

Bubble Chart encodes data using *size of circle* to show *comparisons* which is *difficult* for making *precise quantitative* comparisons



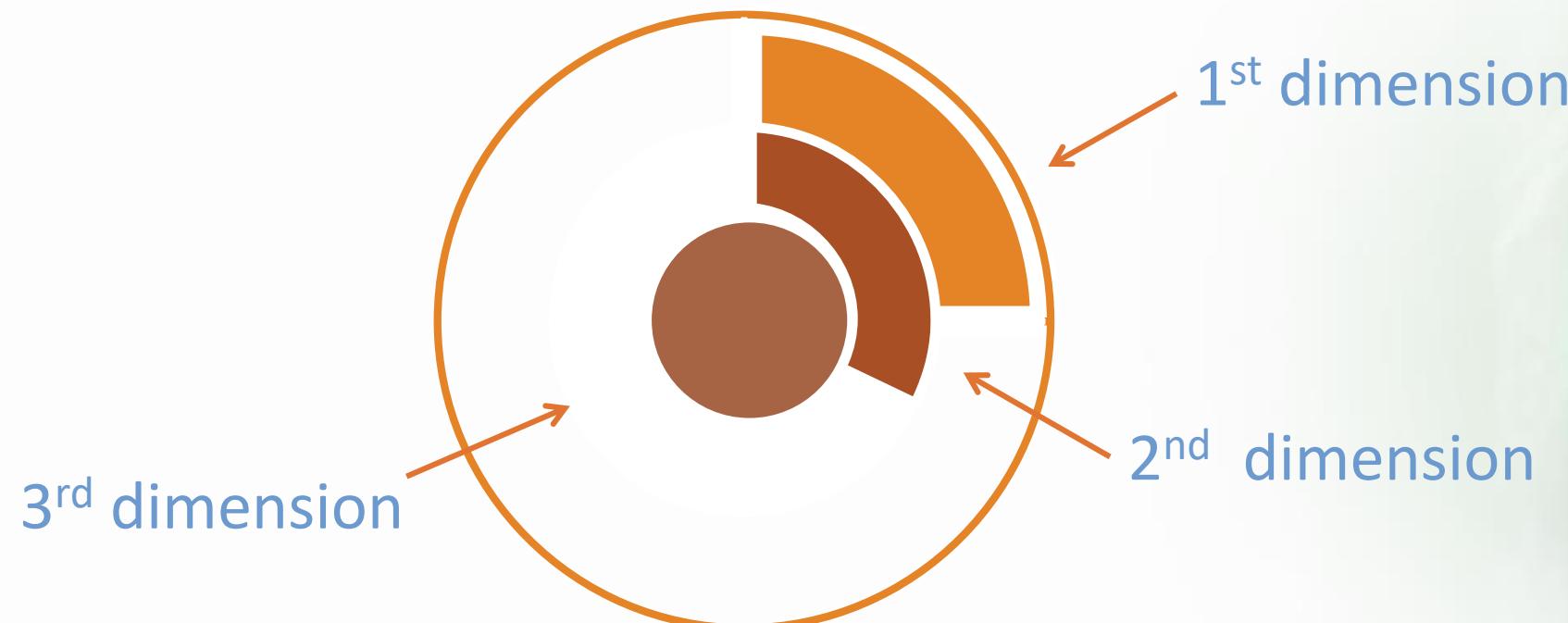
Bubble Chart

Bubble Chart encodes data using *size of circle* to show *comparisons* which is *difficult* for making *precise quantitative* comparisons



Concentric Circles

Concentric Circles encodes data using *arc* and *area* to show comparisons but *problematic* for *many reasons*



*This chart type is not recommended

Donut Chart

Donut Chart encodes data using *arc* and *area* to show a *part-to-whole comparison* but *problematic* for many reasons



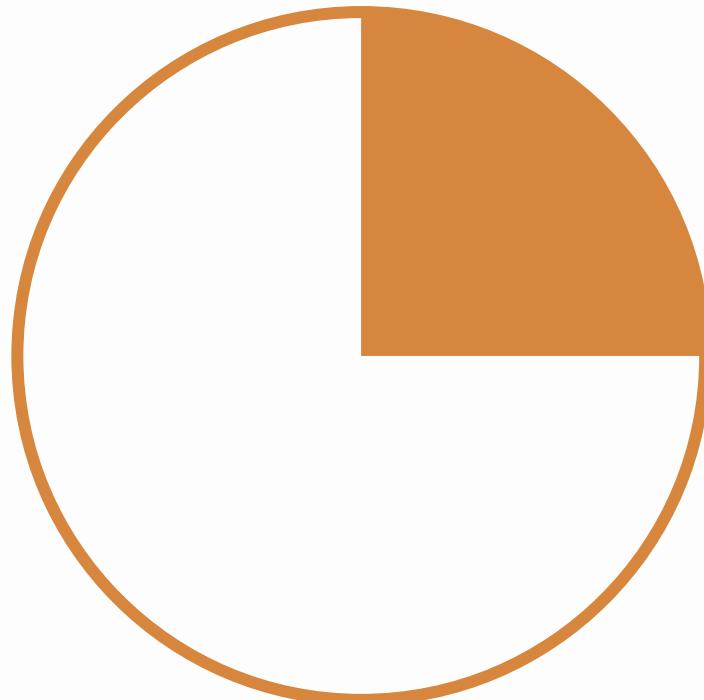
Display single series of data
in a **two-dimensional circle**

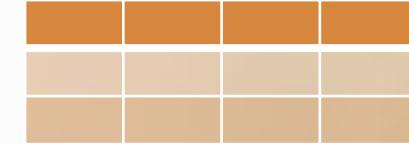
Further extend to
hierarchical data

*This chart type is not recommended

Pie Chart

Pie Chart encodes data using *angle*, *area* and *arc* to show a *part-to-whole comparison* but *problematic* for many reasons



- A visual way instead of table 
- Useful for nominal and ordinal data
- Suitable for proportional data
- Good for 6 categories or less

*This chart type is not recommended

Word Cloud

Word Cloud encodes data using *size* of word to show *comparisons* which is *difficult* for making *precise quantitative* comparisons



Least More
Few **Most**
Many Some

*This chart type is not recommended



<https://ummoodle.um.edu.mo/mod/assign/view.php?id=1368309>

In-class Exercise

*Draw a Box Plot for the given data, by clearly stating the **upper and lower quartiles**, as well as the **mean** and **range***

15, 83, 75, 12, 19, 74, 21

Pie Chart

Not Recommend!!

Save the Pies for Dessert

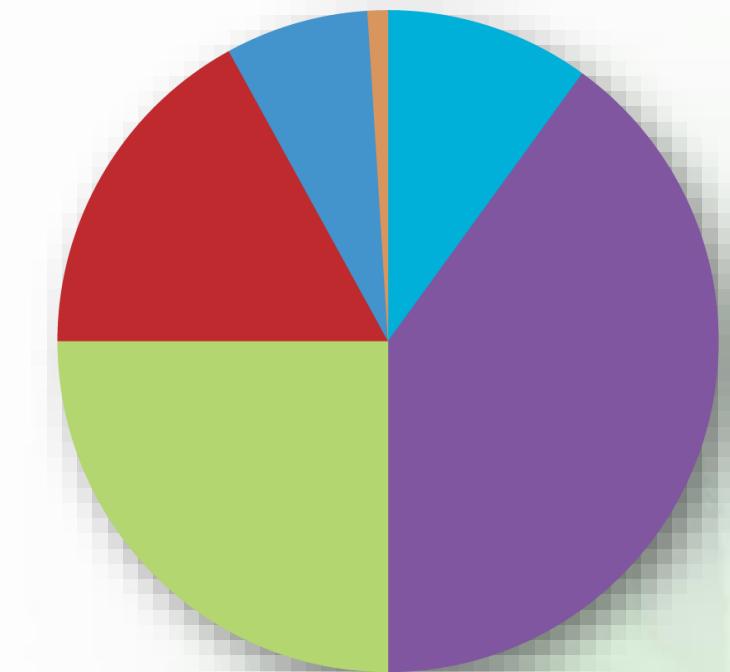
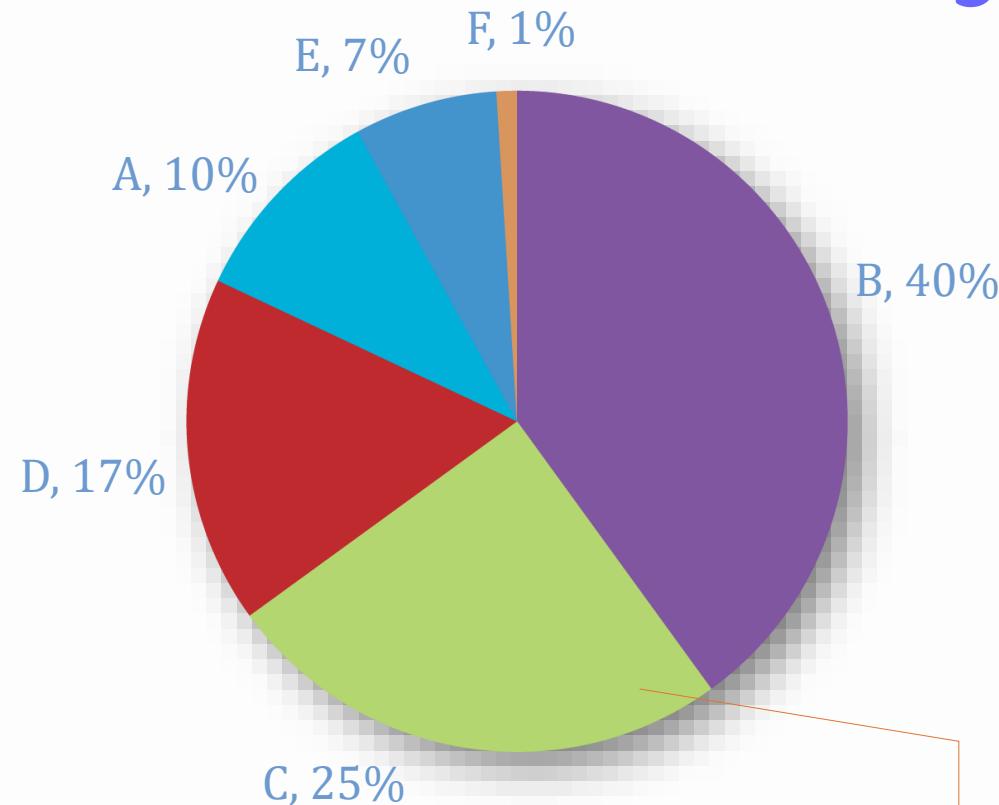
Stephen Few



- It shows part-to-whole relationship
- Good for proportional data
- Easy to judge magnitude of slice when 0%, 25%, 50%, 75%, 100%

Save the Pies for Dessert

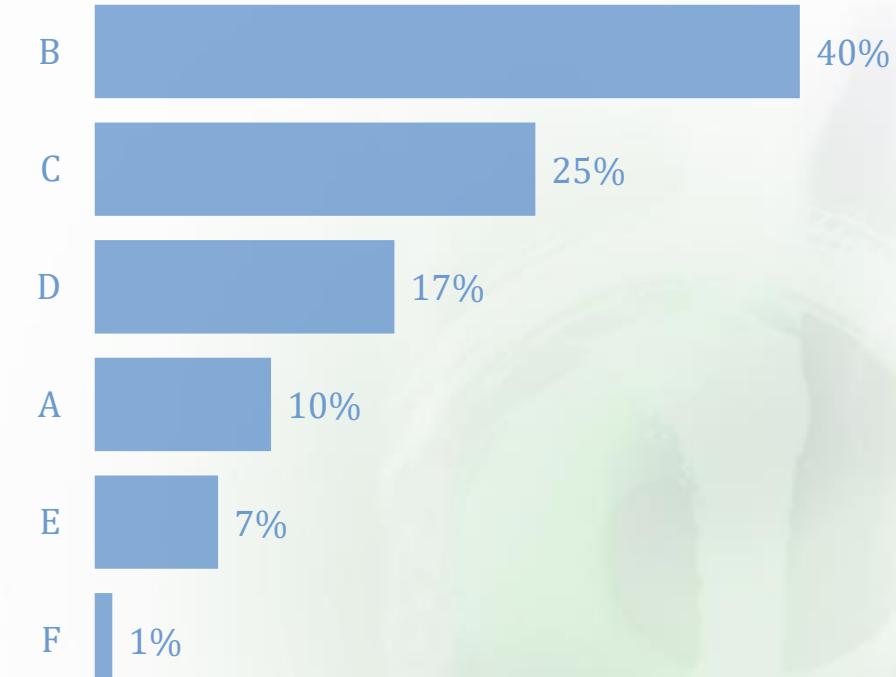
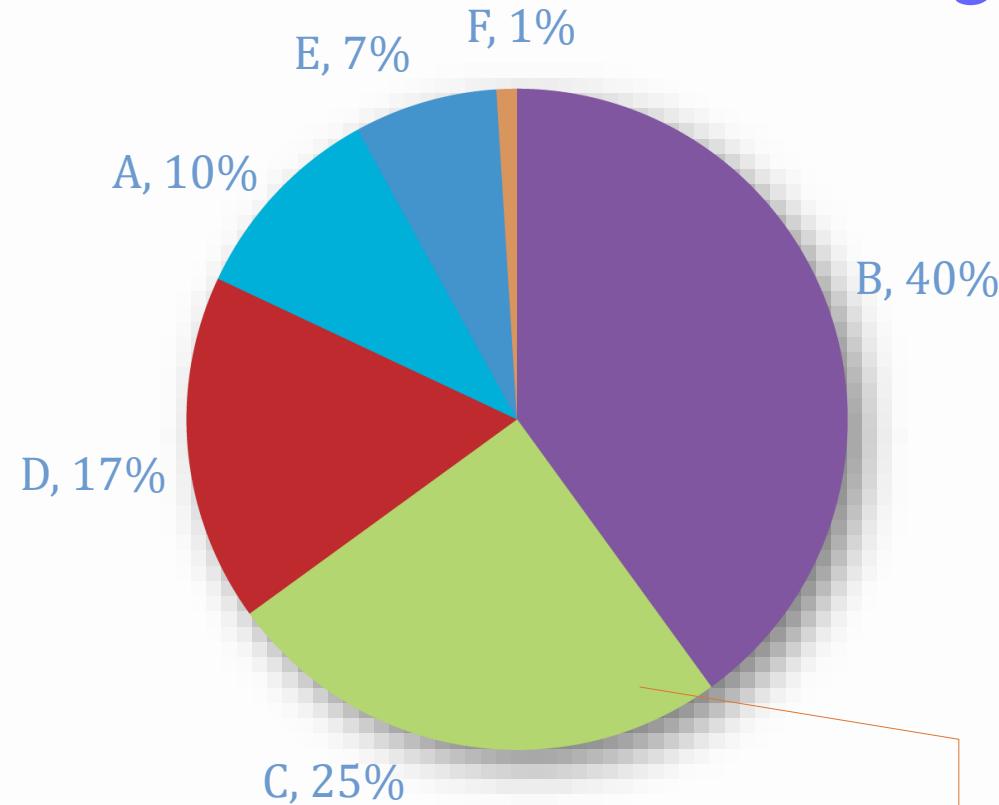
Stephen Few



What is the *percentage* of Green slice?

Save the Pies for Dessert

Stephen Few



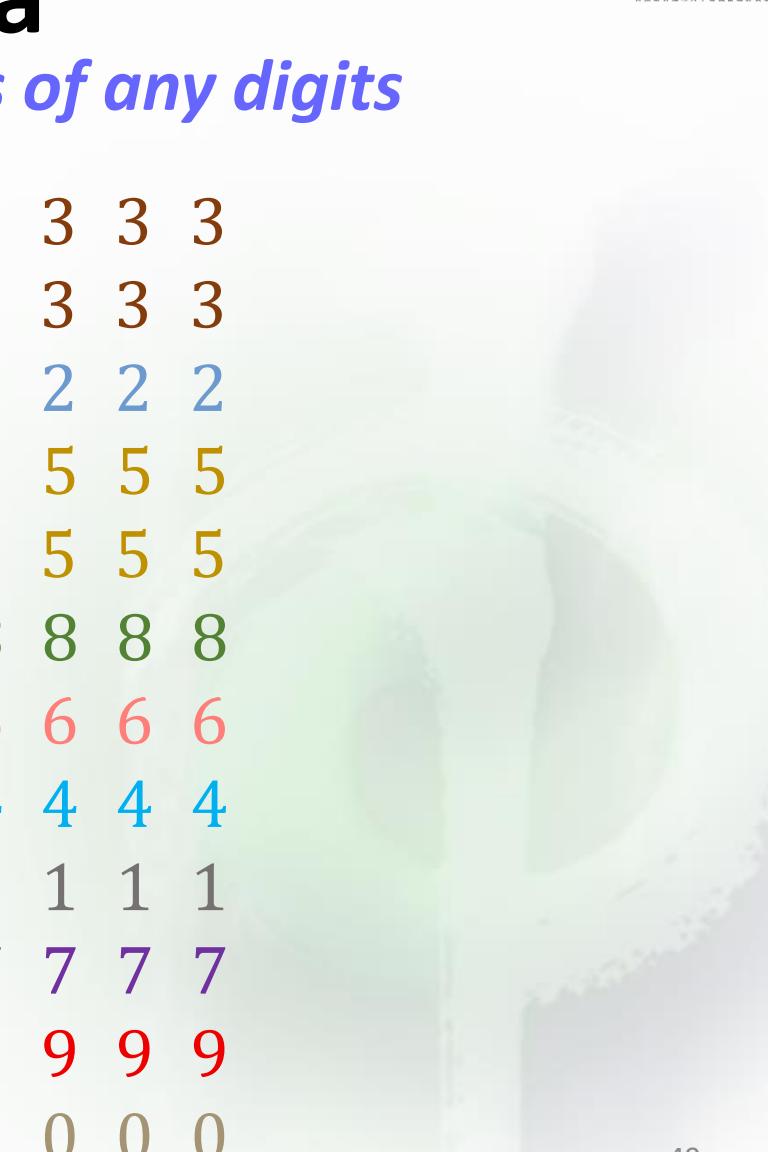
What is the *percentage* of Green slice?

How Pie Chart Represents Data

5	2	8	3	6	1	9	3	6	2	5	3	7	4	3	8	3
8	5	8	9	6	2	1	4	4	3	9	3	6	5	2	4	9
1	0	2	7	5	2	8	3	6	1	6	2	9	3	8	3	8
5	8	4	7	2	0	3	7	3	5	4	7	1	8	2	0	1
2	5	3	6	4	3	9	1	0	8	9	5	7	3	4	5	3
2	7	5	2	8	3	6	1	6	2	9	3	8	3	8	5	8
4	7	2	0	3	7	3	5	4	7	1	8	2	0	1	9	6
2	1	4	4	3	9	3	6	5	2	4	9	1	0	2	7	5
2	8	3	6	1	6	2	9	3	8	3	8	5	8	4	7	2
0	3	7	3	5	4	7	1	8	2	0	1	2	5	3	6	4
3	9	1	0	8	9	5	7	3	4	5	3	2	7	5	2	8
3	6	1	6	2	4	6	2	7	5	9	1	5	2	6	3	6

How Pie Chart Represents Data

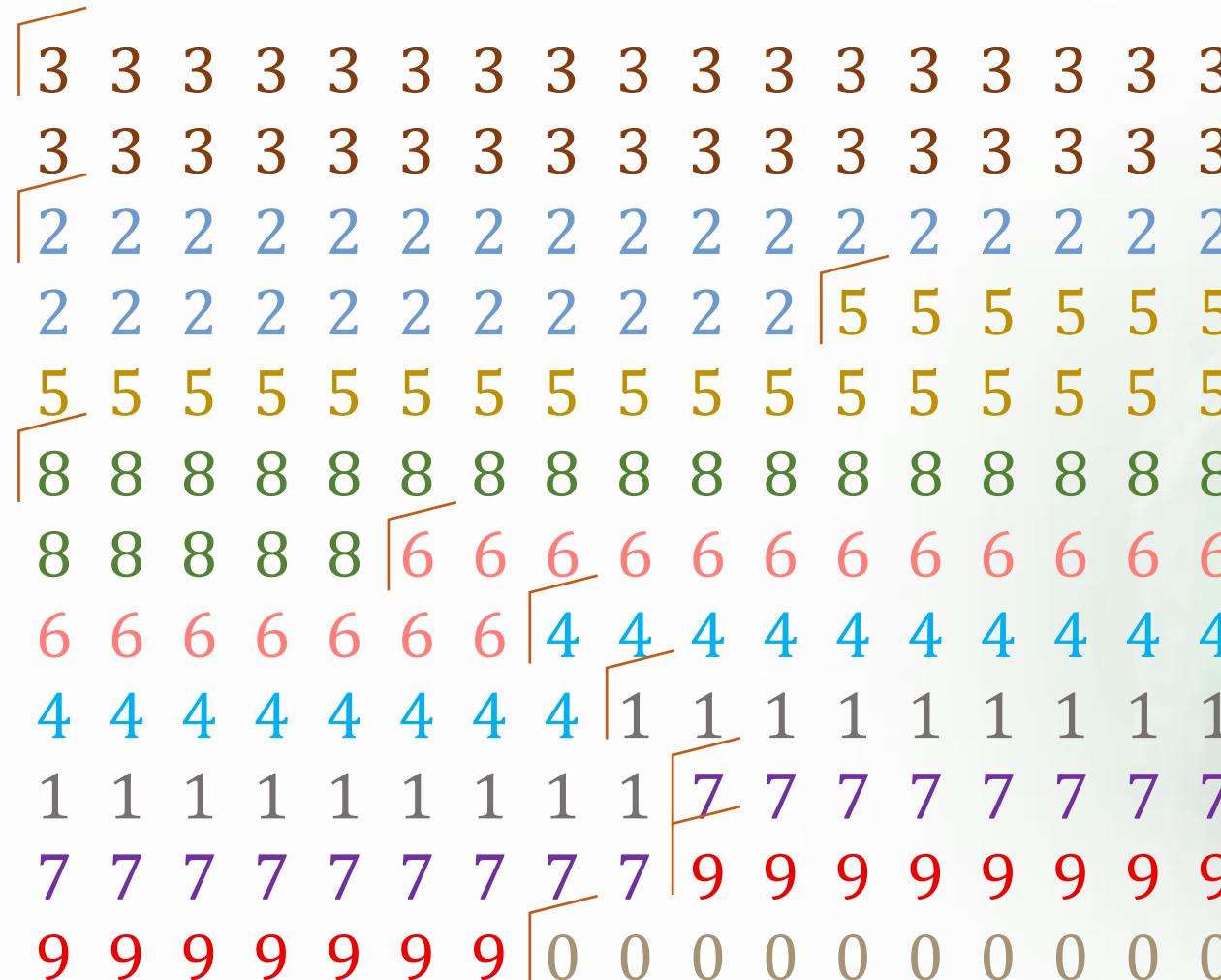
Try to quickly compare the totals of any digits



3
3
2
2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 5 5 5 5 5 5
5
8
8 8 8 8 8 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6
6 6 6 6 6 6 6 4 4 4 4 4 4 4 4 4 4 4 4 4 4
4 4 4 4 4 4 4 4 1 1 1 1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 7 7 7 7 7 7 7 7 7 7 7 7
7 7 7 7 7 7 7 7 7 9 9 9 9 9 9 9 9 9 9 9 9
9 9 9 9 9 9 9 0 0 0 0 0 0 0 0 0 0 0 0 0 0

How Pie Chart Represents Data

Baseline moves based on previous Digits

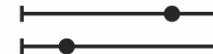


Rankings of Visual Attributes

Recap

④ Magnitude Channels: Ordered Attributes

Position on common scale



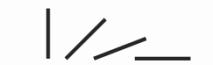
Position on unaligned scale



Length (1D size)



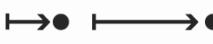
Tilt/angle



Area (2D size)



Depth (3D position)



Color luminance



Color saturation



Curvature



Volume (3D size)



Angle, Arc and Area

④ Identity Channels: Categorical Attributes

Spatial region



Color hue



Motion



Shape



▲ Most Effective ▼ Least Effective

How Bar Chart Represents Data

Easier comparison

3 3

2 2

5 5

8 8

6 6

4 4

1 1

7 7

9 9

0 0 0 0 0 0 0 0 0

How Bar Chart Represents Data

Easier comparison, even without color encoding!

3 3

2 2

5 5

8 8

6 6

4 4

1 1

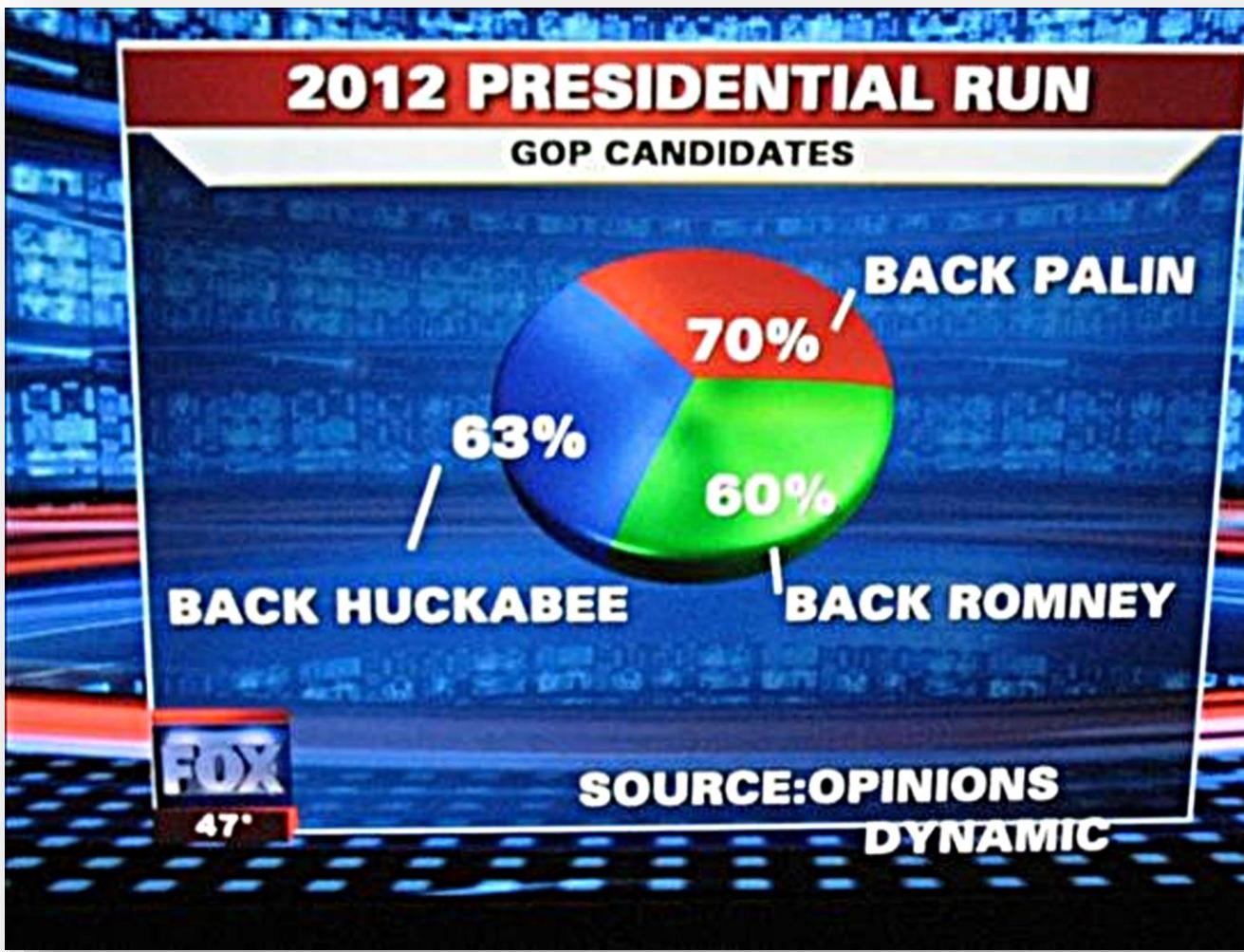
7 7

9 9

0 0 0 0 0 0 0 0 0

Remove
to improve
the **pie chart** edition

Worst Pie Chart Ever Done

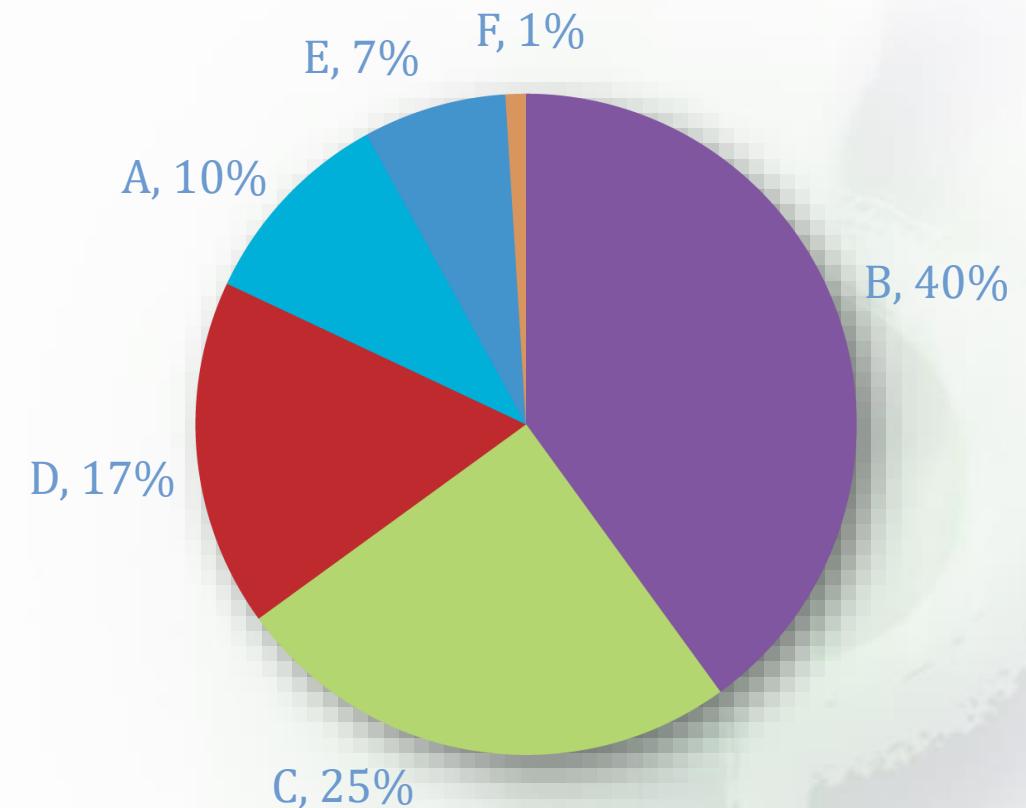


General Rules for Pie Charts

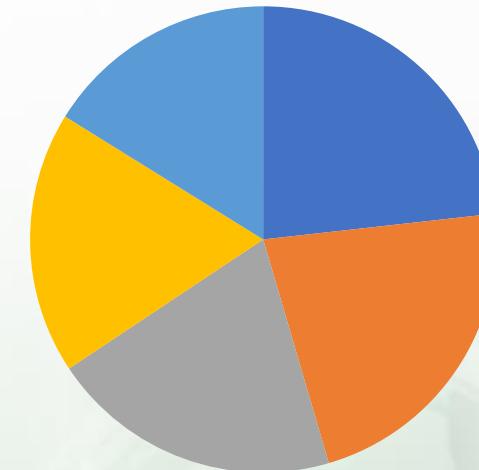
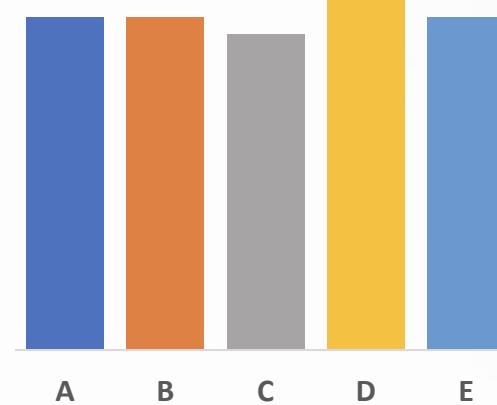
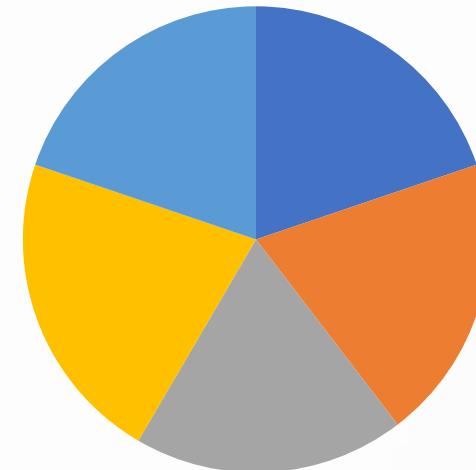
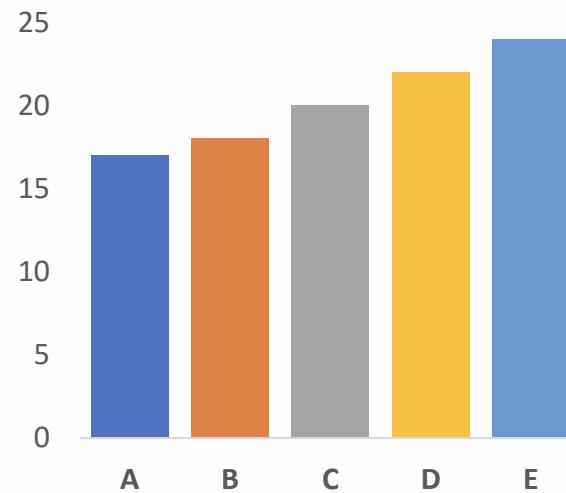
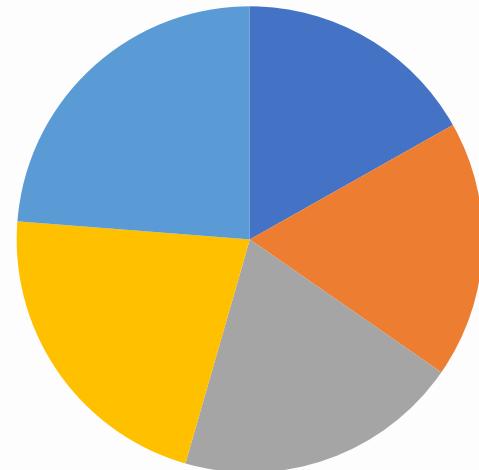
Don't Use Pie Charts

If you *must break Rule* then:

- Make sure it *adds up 100%*
- Only a *few categories*
- Start at *noon* and *move clockwise*
- *Largest to Smallest* values
- Add *Labels* for %
- *Avoid 3D*
- Keep it *Simple*



Can You Find the Differences?

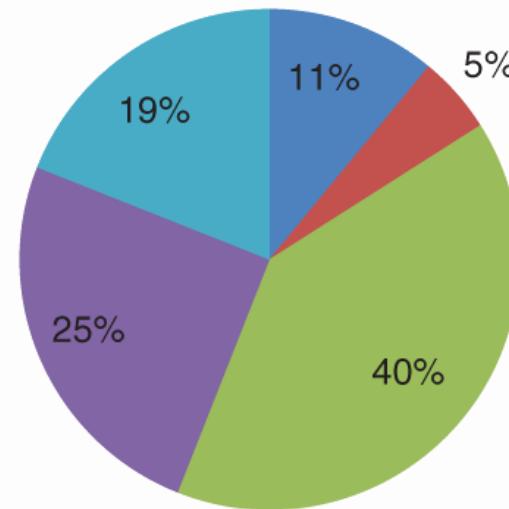


Alternatives to Pie Charts

Survey: Summer Learning Program on Science

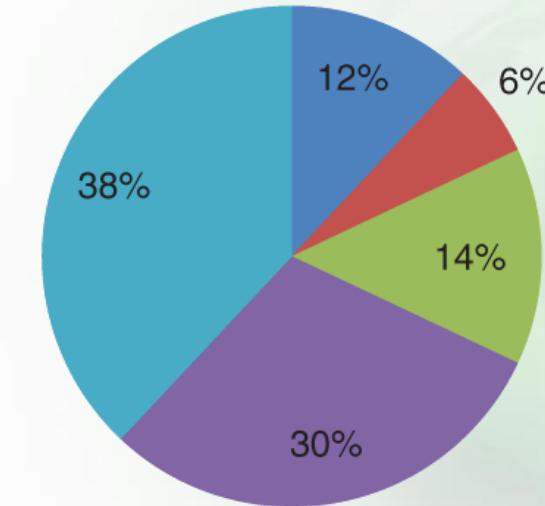
PRE: How do you feel about doing science?

■ Bored ■ Not great ■ OK ■ Kind of interested ■ Excited



POST: How do you feel about doing science?

■ Bored ■ Not great ■ OK ■ Kind of interested ■ Excited



Alternatives to Pie Charts

Survey: Summer Learning Program on Science

Pilot program was a success

After the pilot program,

68%

of kids expressed interest towards science,
compared to 44% going into the program.

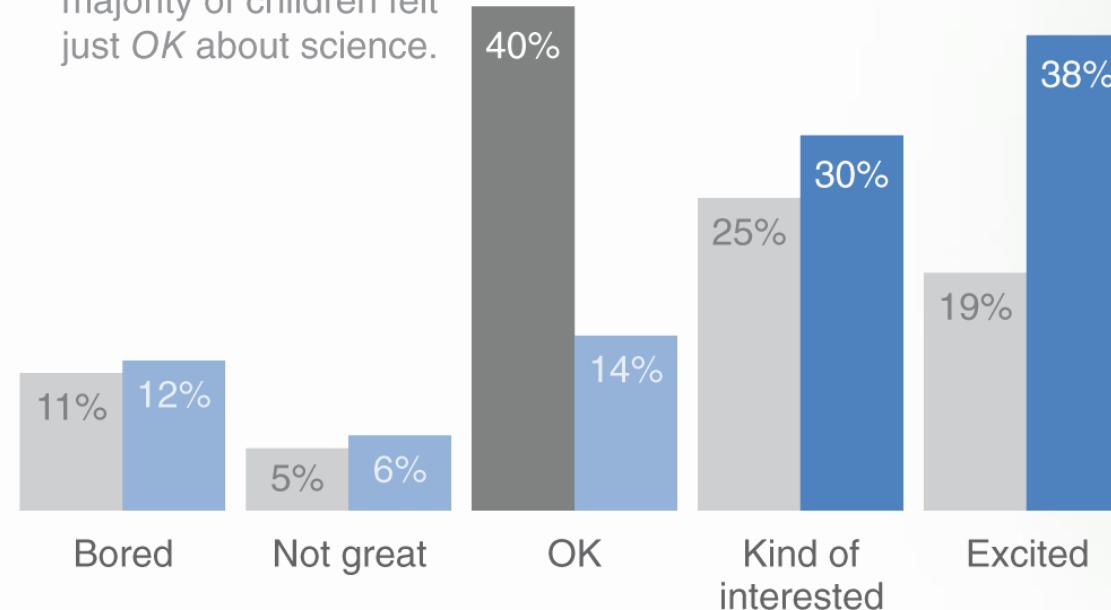
Alternatives to Pie Charts

Survey: Summer Learning Program on Science

Pilot program was a success

How do you feel about science?

BEFORE program, the majority of children felt just *OK* about science.



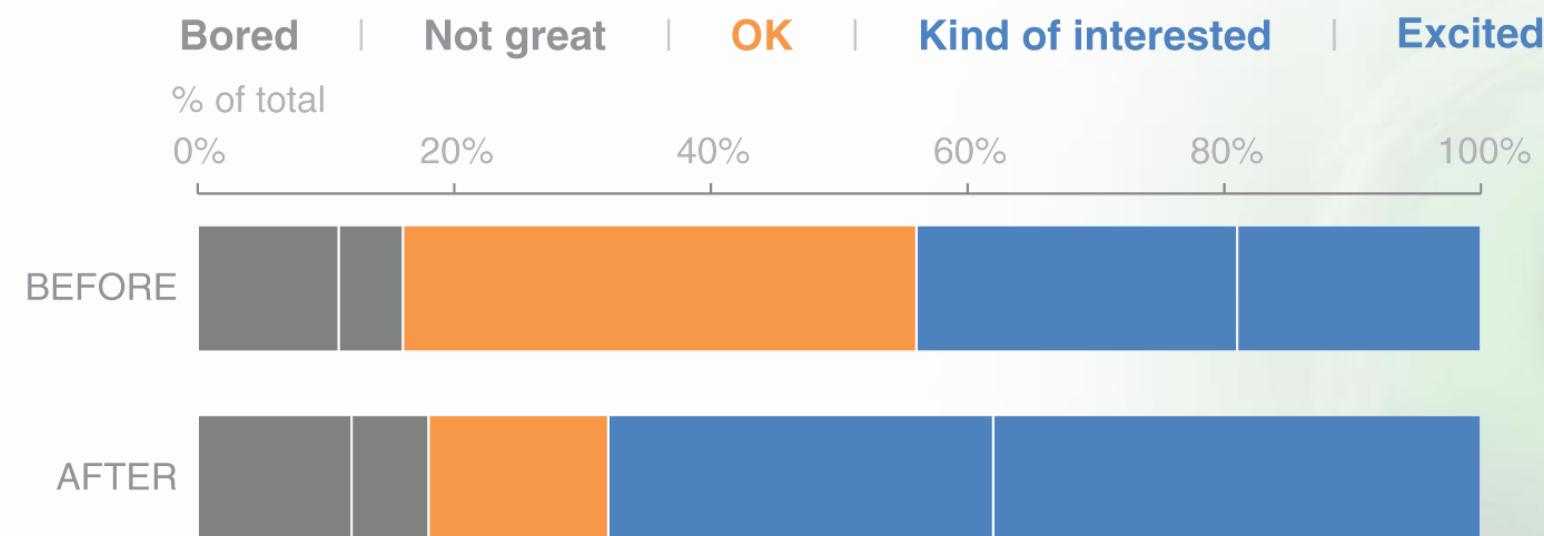
AFTER program, more children were *Kind of interested* & *Excited* about science.

Alternatives to Pie Charts

Survey: Summer Learning Program on Science

Pilot program was a success

How do you feel about science?



BEFORE program, the majority of children (40%) felt just *OK* about science.

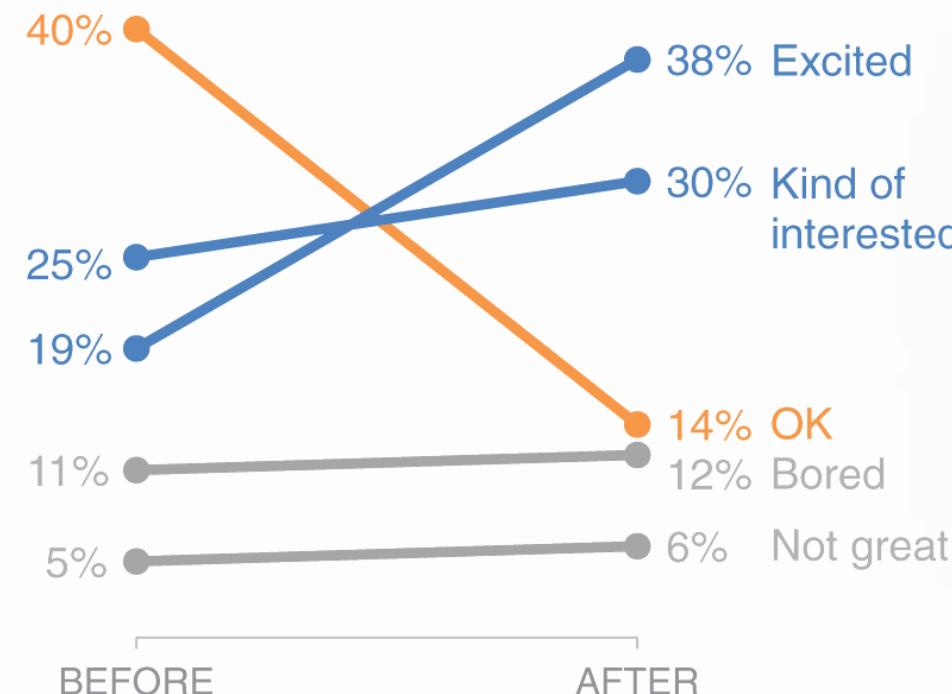
AFTER program, more children were *Kind of interested* (30%) & *Excited* (38%) about science.

Alternatives to Pie Charts

Survey: Summer Learning Program on Science

Pilot program was a success

How do you feel about science?



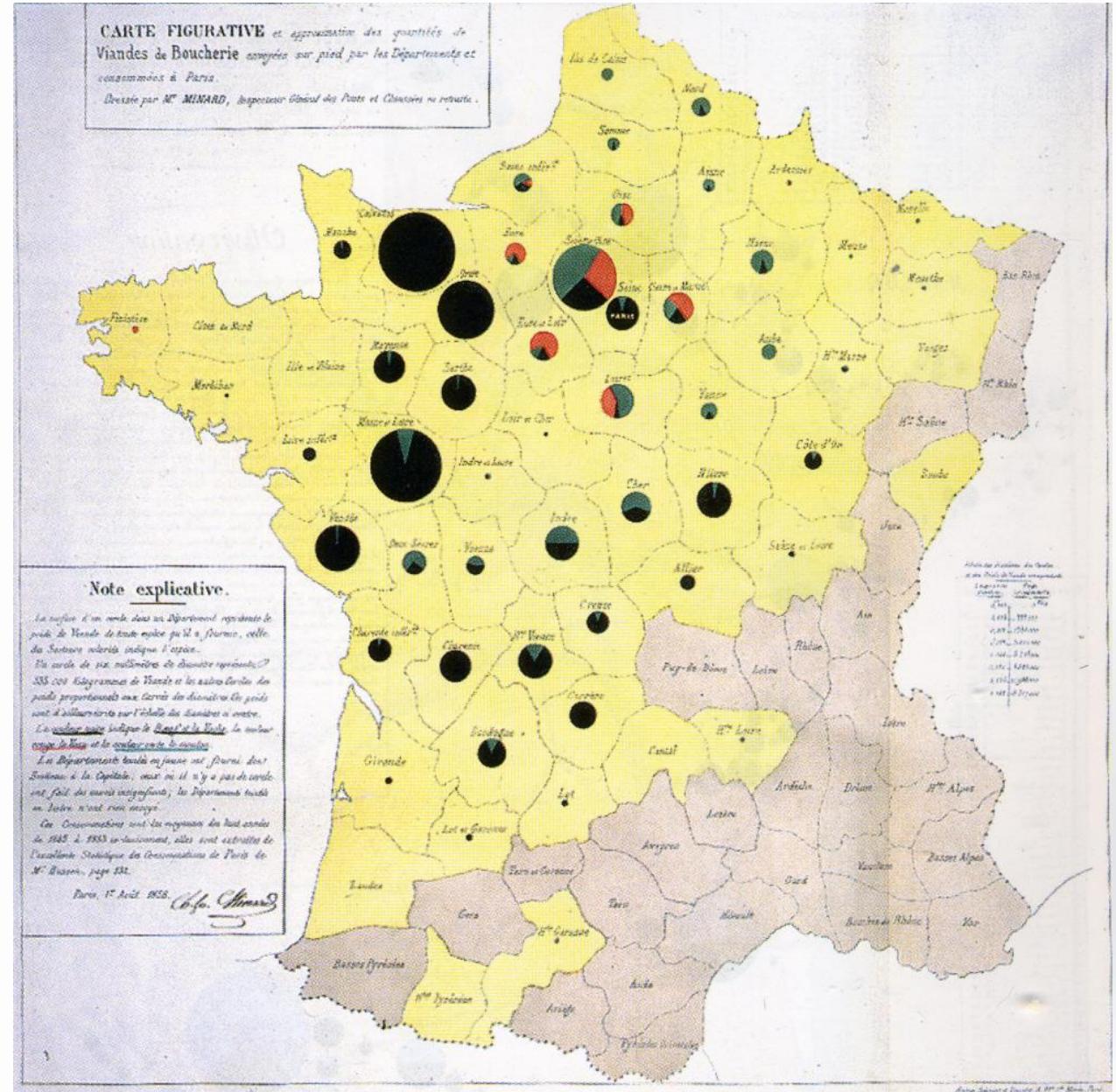
BEFORE program, the majority of children felt just *OK* about science.

AFTER program, more children were *Kind of interested* & *Excited* about science.

Pie Chart on a Map

This is acceptable that a map relies on 2D position, there is no common baseline

No better way to visualize a part-to-whole relationship or categorical comparison



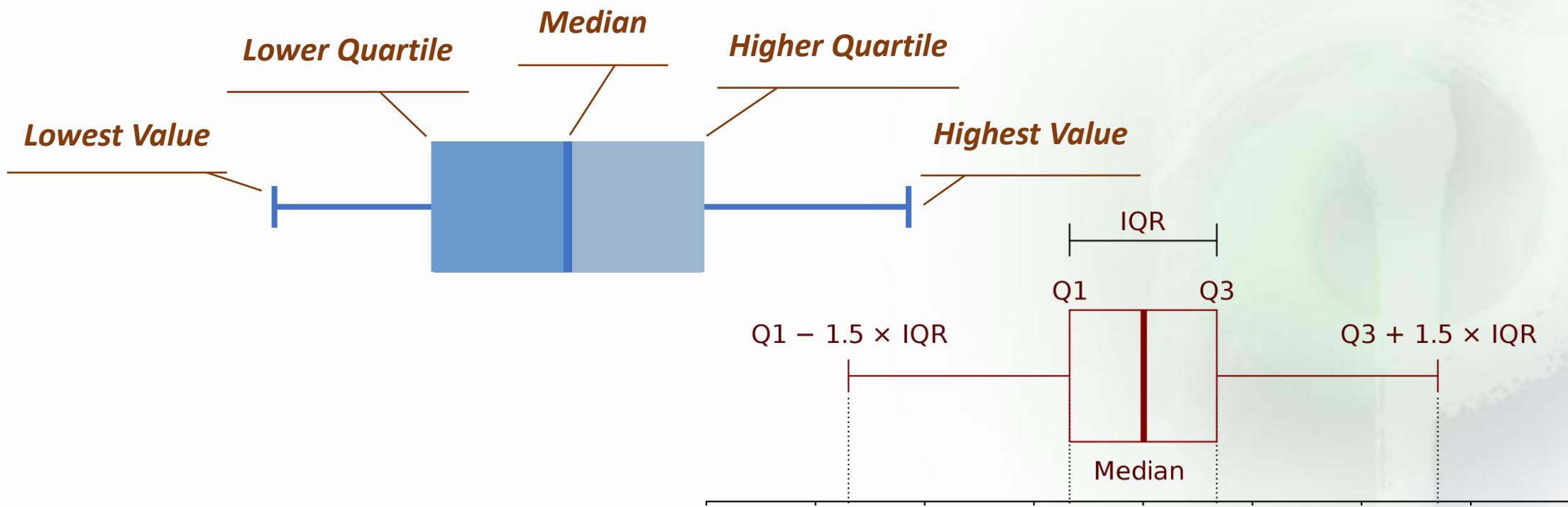
by Charles Joseph Minard, 1858

More Graphs

Box Plot

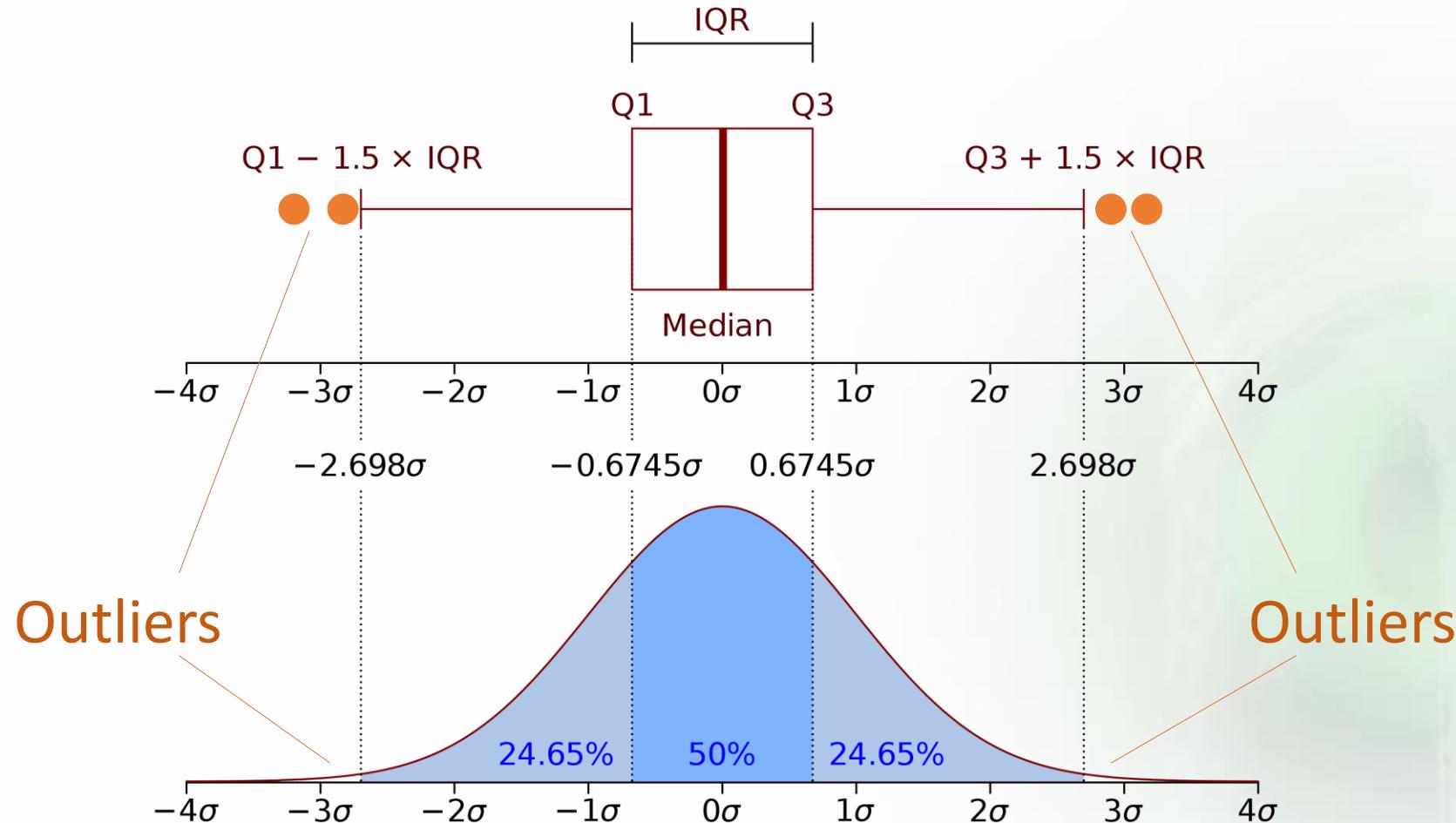
Invented by John Tukey

The *strength* of Box Plot is able to show the *distribution* of *all data* whatever the *size of data*



Box Plot

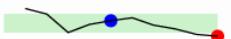
Invented by John Tukey



Sparklines

Invented by Edward Tufte

Small, high-resolution graphics embedded in a *context of words, numbers, images*. Sparklines are *data-intense, design-simple, word-sized* graphics

simple, wordlike graphics, so
parkline:  (the stock
of a more detailed introduction.

You can combine bars and lines: .

Data can be hard to grasp – visualising it can make comprehension faster. Sparklines (tiny charts in text, like this:

123  789) are a useful tool but creating them for the web has always required code. Removing the need for code makes it more accessible. If you can use type, you can use Spark.

Mauricio Pochettino has lead Spurs on their best run 8TH  2ND in 24 years of the Premier League

Alibaba stock is at 5 yr high 93.89  152.11 as of July 2017

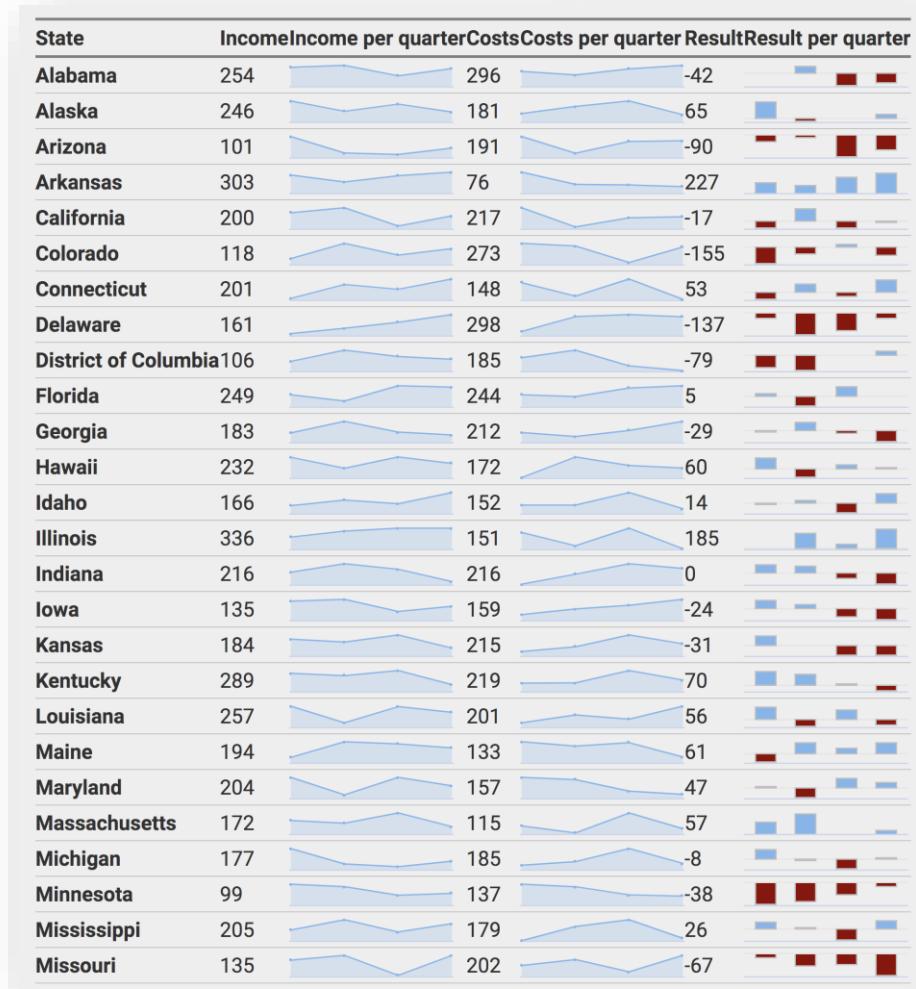
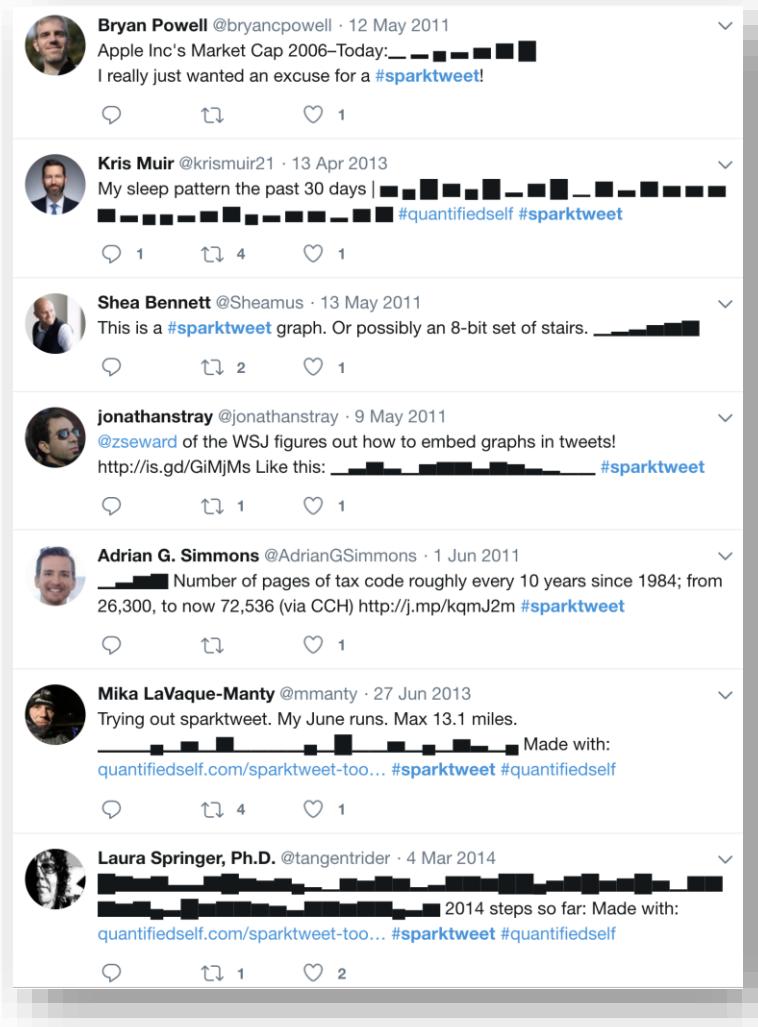
The FTSE100 Brexit bounce 5562  7501 continues one year on from the vote last summer



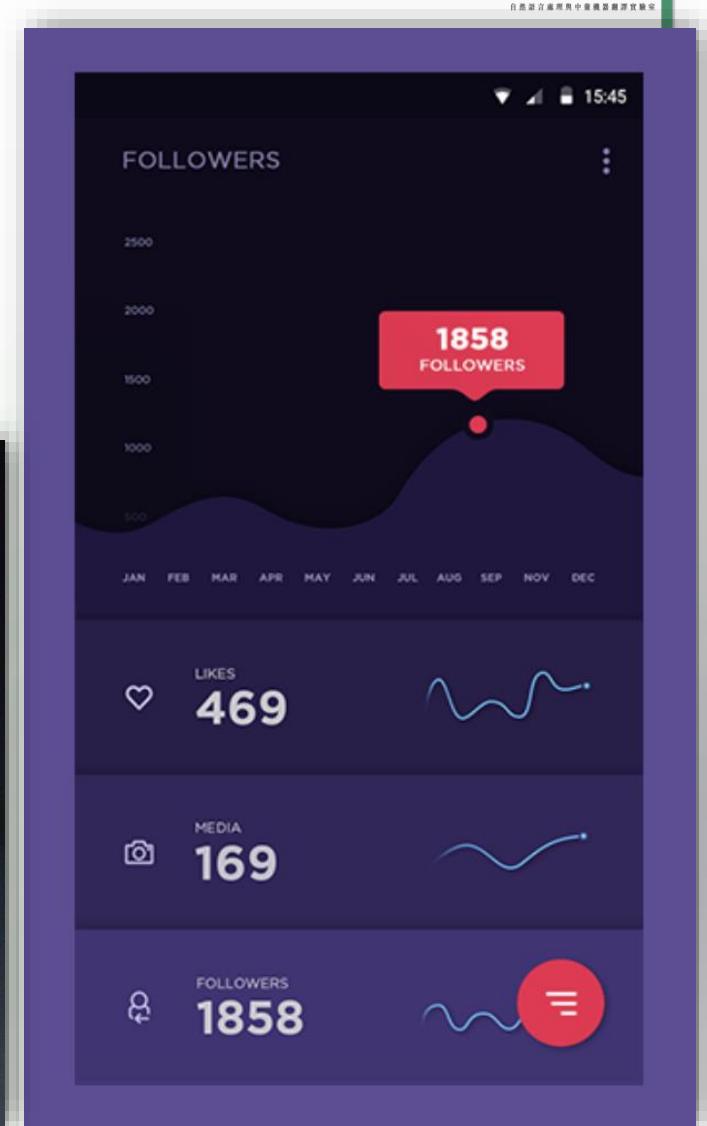
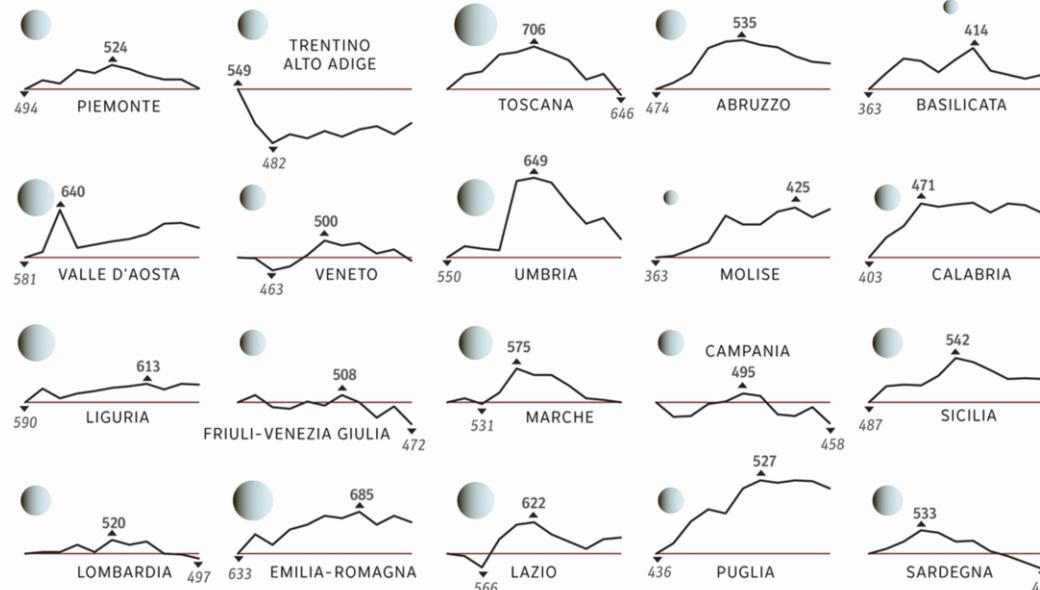
AtF Spark

sparklines have a limited reserving power, like the volume size of sparklines reclaims the overall labels and scaling of conventional static displays. Most of our examples have, however, depicted contextual methods for quantifying sparklines: the gray normal limits and the red encoding to link data points to exact numbers  GLUCOSE 6.6, global scale band labels for sparkline clusters; and, probably best of all, sparkline with an implicit data-scaling box formed by numbers that label key data points (such as beginning/end/bounce) 1.1025  1.1907 | 0783 | 2858

Sparkline Variants



Sparkline Variants



Small Multiples

Invented by Edward Tufte

A *small multiple* is a series of *similar graphs* or *charts* using the *same scale* and *axes*, allowing them to be *easily compared*



Regional Support for Same-Sex Marriage

% favoring same-sex marriage, 2003-2014



Note: Regional breakdowns are based on the U.S. Census regions and divisions, with three exceptions. Maryland, Delaware and D.C. are grouped in the mid-Atlantic with New York, New Jersey and Pennsylvania, instead of in the South Atlantic. The census divisions of East South Central and West South Central are combined into a single South Central designation.

PEW RESEARCH CENTER

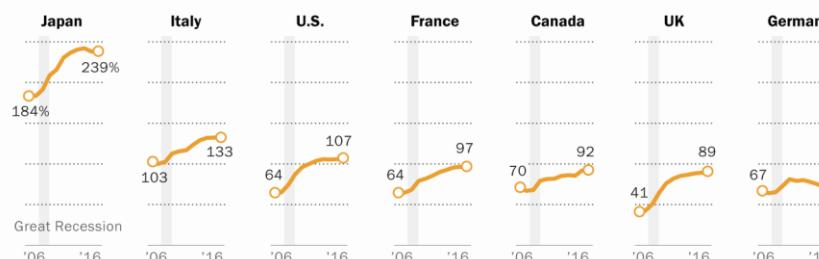
Small Multiples

Invented by Edward Tufte

A *small multiple* is a series of *similar graphs* or *charts* using the *same scale* and *axes*, allowing them to be *easily compared*

After Great Recession, debt increased substantially in most G-7 economies

Total gross debt as a share of GDP in the Group of Seven nations



Note: Gross debt represents total liabilities of all levels and units of government — national, state/provincial and local — less liabilities held by other levels or units of government, unless otherwise noted by source.

Source: The International Monetary Fund, World Economic Outlook, accessed Sept 7, 2017.

PEW RESEARCH CENTER

Shifting partisan views of government performance

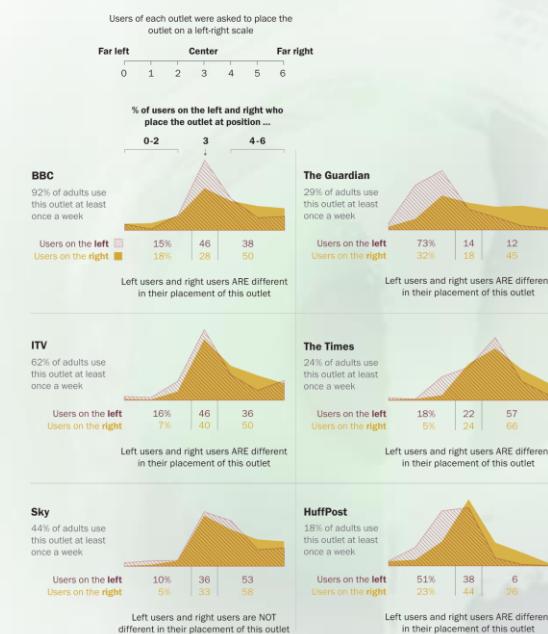
% who say the federal government is doing a very/somewhat good job ...



Source: Survey of U.S. adults conducted Nov. 29-Dec. 4, 2017.

PEW RESEARCH CENTER

Ideological placement of news outlets in the UK



Note: Some outlets are not included, because their audience sample sizes are too small to analyze. Left and right users' outlet placements are considered different if the percentage of left and right users that place the outlet on the left (from 0 to 2), on the right (from 4 to 6), or both are significantly different.

Source: Survey of eight Western European countries conducted Oct. 30-Dec. 20, 2017.
"In Western Europe, Public Attitudes Toward News Media More Divided by Populist Views Than Left-Right Ideology."

PEW RESEARCH CENTER

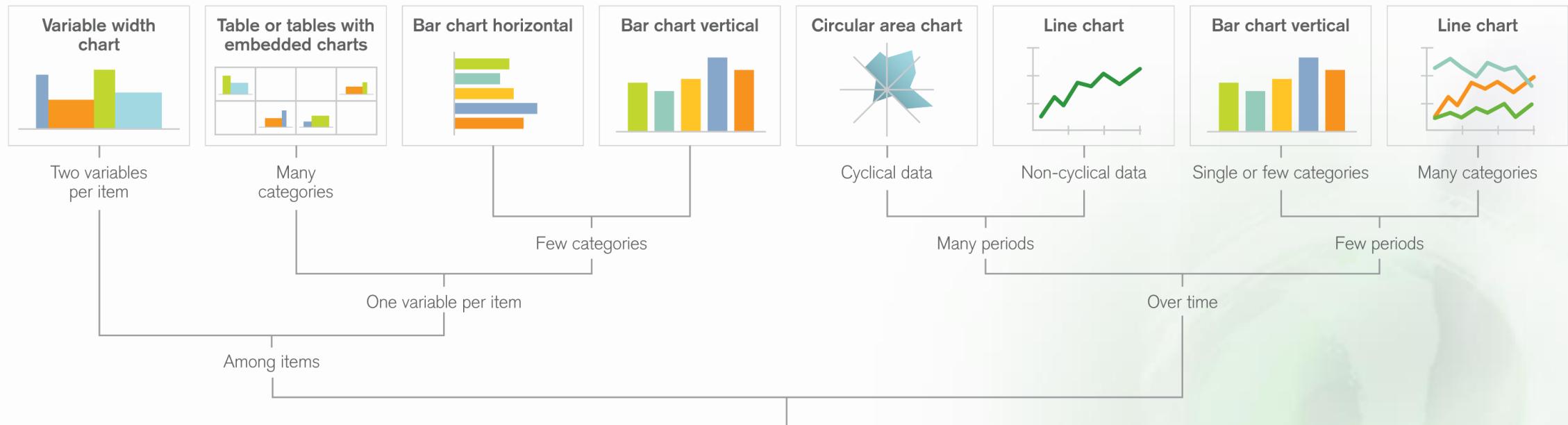
Choose the Right Charts

Four Visual Data Analysis

There are *four basic methods* of data analysis that can help to *turn your data into knowledge*:

- Comparison
- Composition
- Distribution
- Relationship

Comparison Visualizations



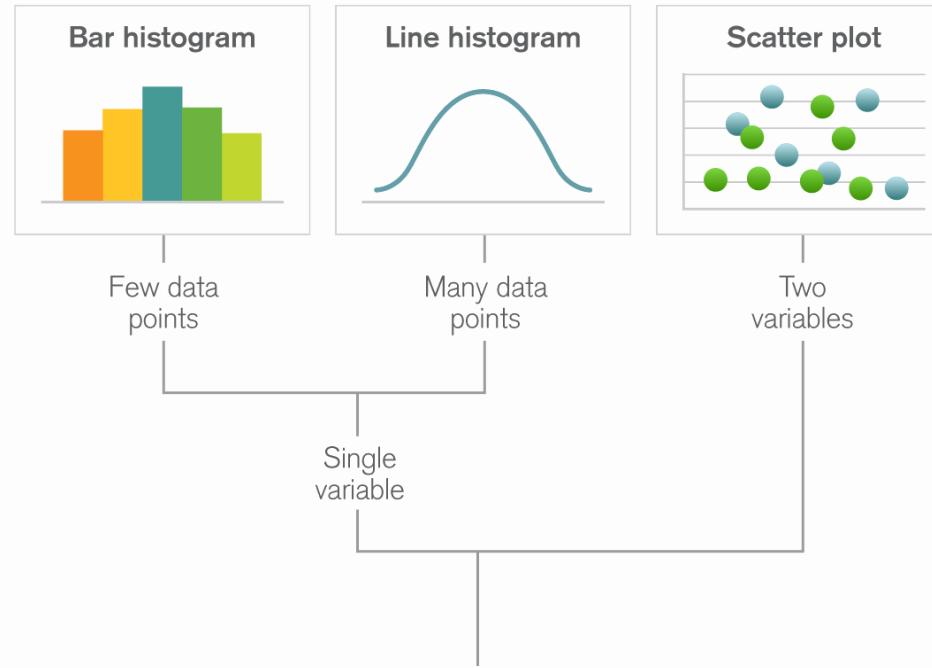
To *compare* the *magnitude of values* to each other and can be used to easily find the *lowest* and *highest* values in the data, the *current values* versus *old* to see if the values are *increasing* or *decreasing*

Composition Visualizations



To see how a *part* compares to the *whole* and how a *total value* can be divided into *shares*, shows the *relative value*, but some charts can also be used to show the *absolute* difference

Distribution Visualizations

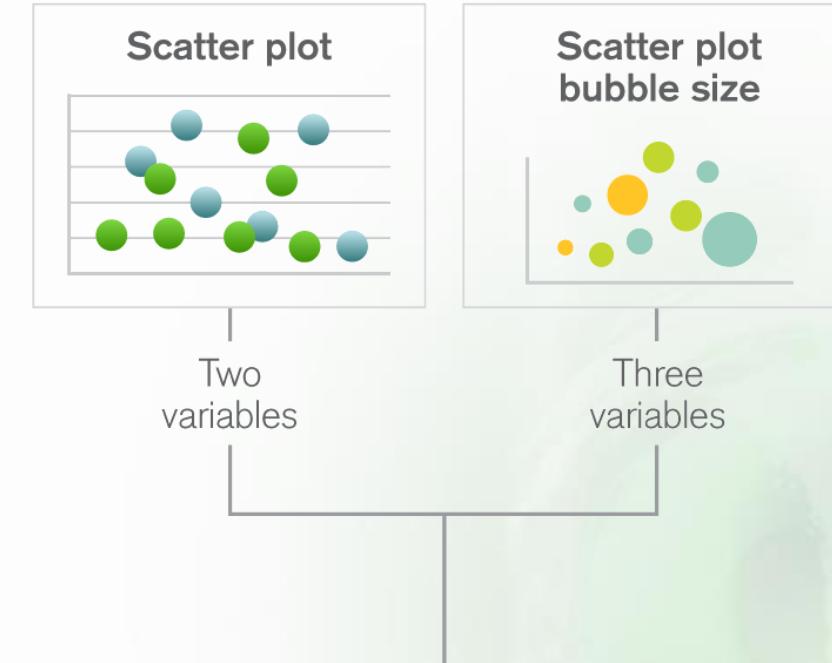


To see how quantitative values are distributed along an axis from lowest to highest

To look at the *shape of the data* a user can identify characteristics such as the *range of values*, *central tendency*, *shape* and *outliers*, can be used for *prediction* and *making decision*

Relationship Visualizations

Question: is there a **correlation** between **advertising spend** and **sales** for our products

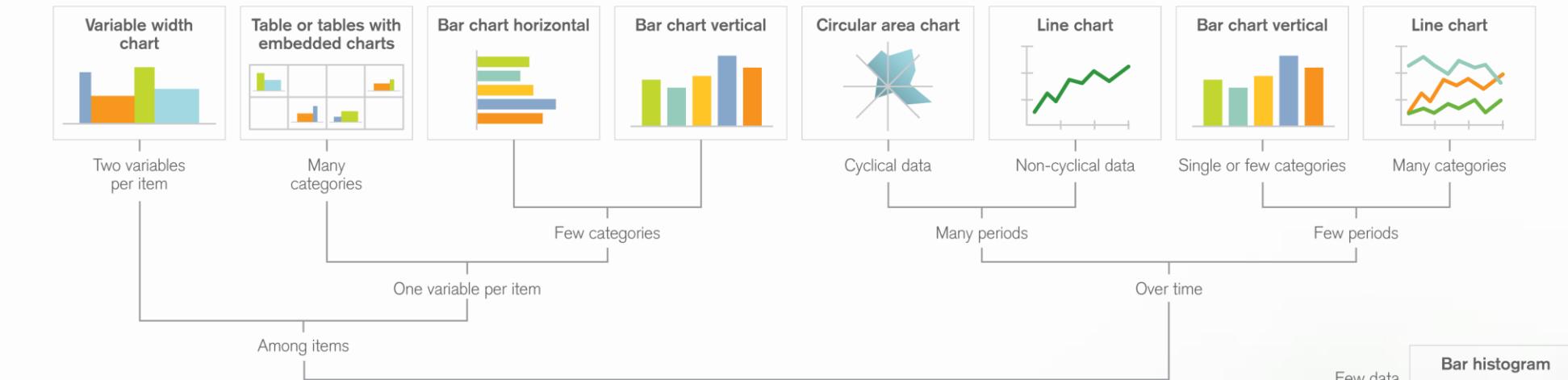


To see the *relationship* between the data and can be used to find *correlations*, *outliers* and *clusters* of data

Picking The Right Chart Type

For different methods of analysis there are *different types of charts* that work best

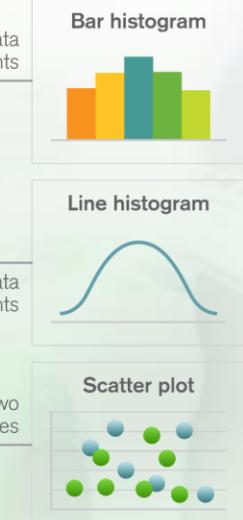
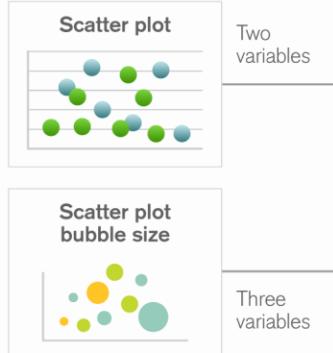
- How *many variables* do you want to show in a single chart? One, two, three, many?
- How *many items* (data points) will you display for each variable? Only a few or many?
- Will you display values *over a period of time*, or among items or groups?



Comparison

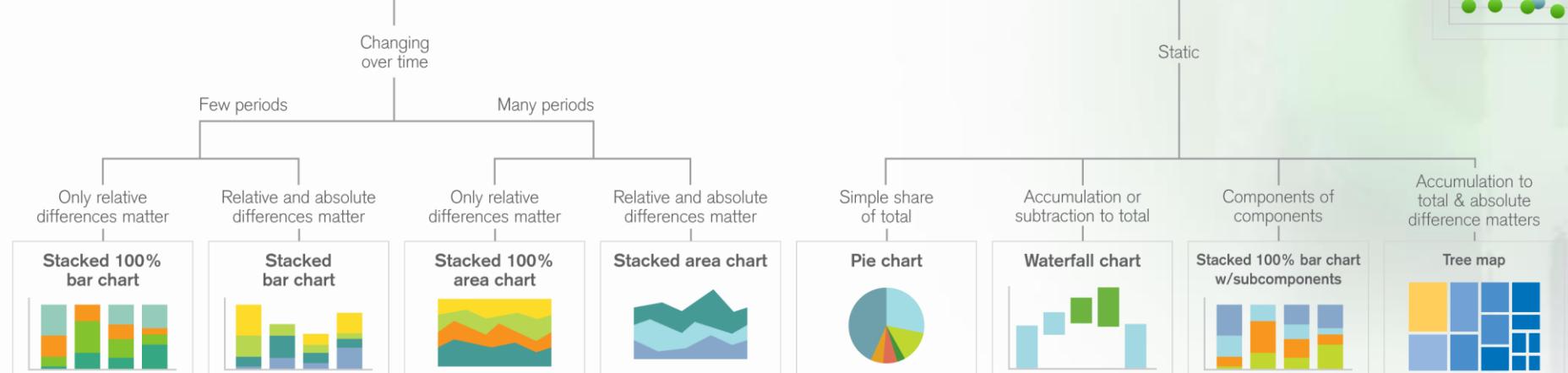
What would you like to show?

Relationship



Distribution

Composition



References

- Few, S., & Edge, P. (2007). *Save the pies for dessert*. Visual Business Intelligence Newsletter, 1-14.

Acknowledgements

Some of the materials are adapted from:

- Jeffrey A. Shaffer
- Qlik.com