GBDi
Government Big Data Institute
สถาบันส่งเสริมการวิเคราะห์และบริหารข้อมูลขนาดใหญ่ภาครัฐ (สวข.)

depa

โครงการอบรมหลักสูตร Hand-on Data Science and Machine Learning

Data Visualization

Navavit Ponganan
Data scientist
Government Big Data Institute (GBDi)

Definition

Data visualization is the graphical representation of information and data. By using visual elements like charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data.

Data visualization beginner's guide: a definition, examples, and learning resources (https://www.tableau.com/learn/articles/data-visualization)

Visualization Goals

Communicate

explanatory

- Present data and ideas
- Explain and inform
- Provide evidence and support
- Influence and persuade

Analyze

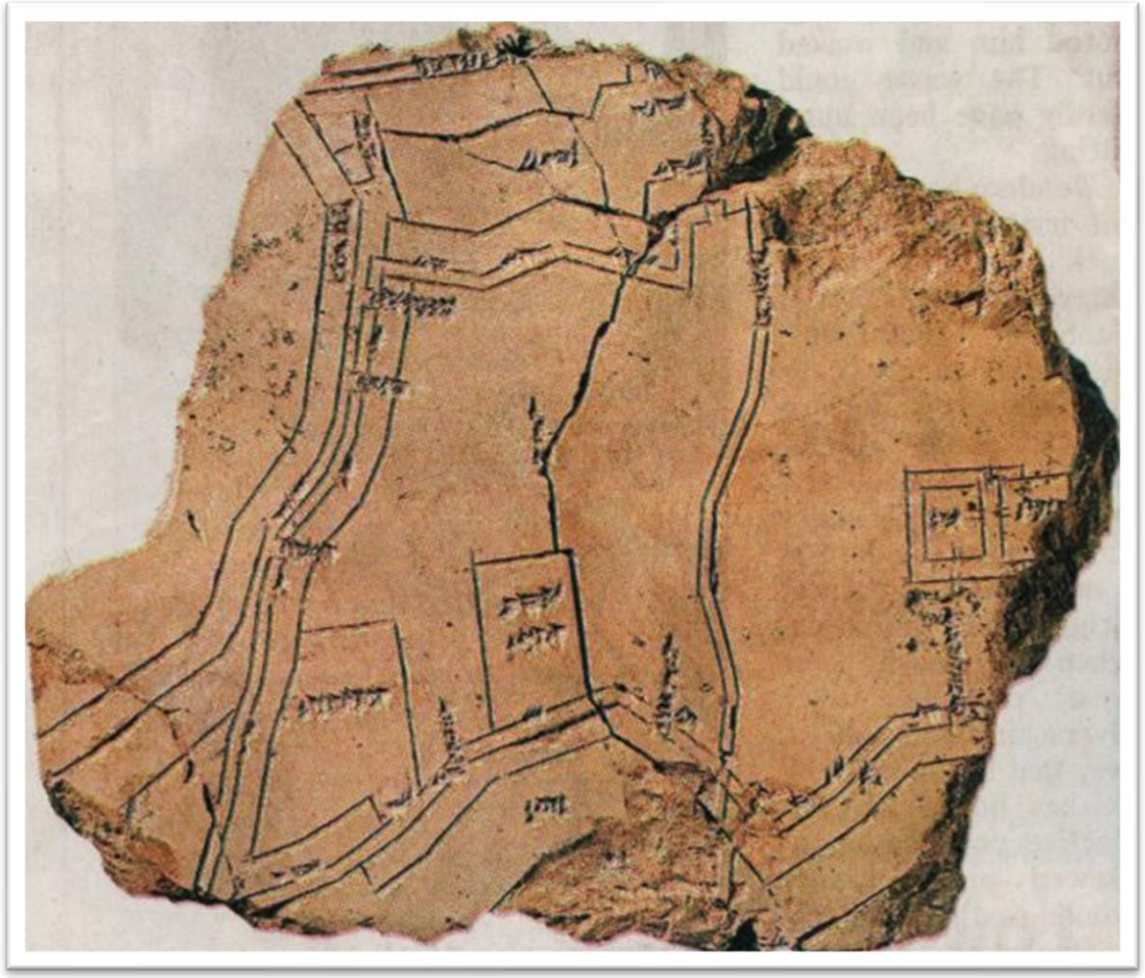
exploratory

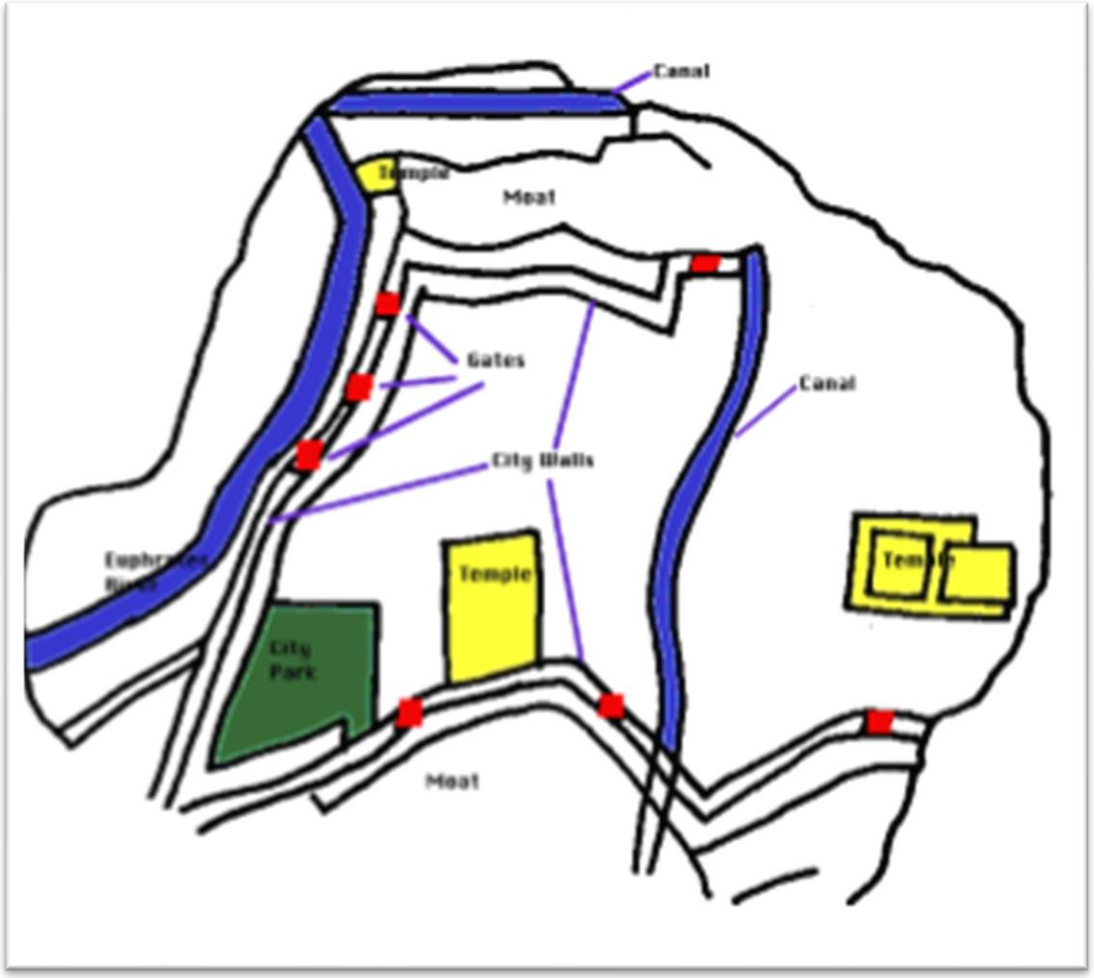
- Explore the data
- Assess a situation
- Determine how to proceed
- Decide what to do

A Wall Painting at Catal Huyuk

Ref: http://www.bbc.co.uk/radio4/science/unearthingmysteries_20021112.shtml

Clay tablet from Nippur, Babylonia

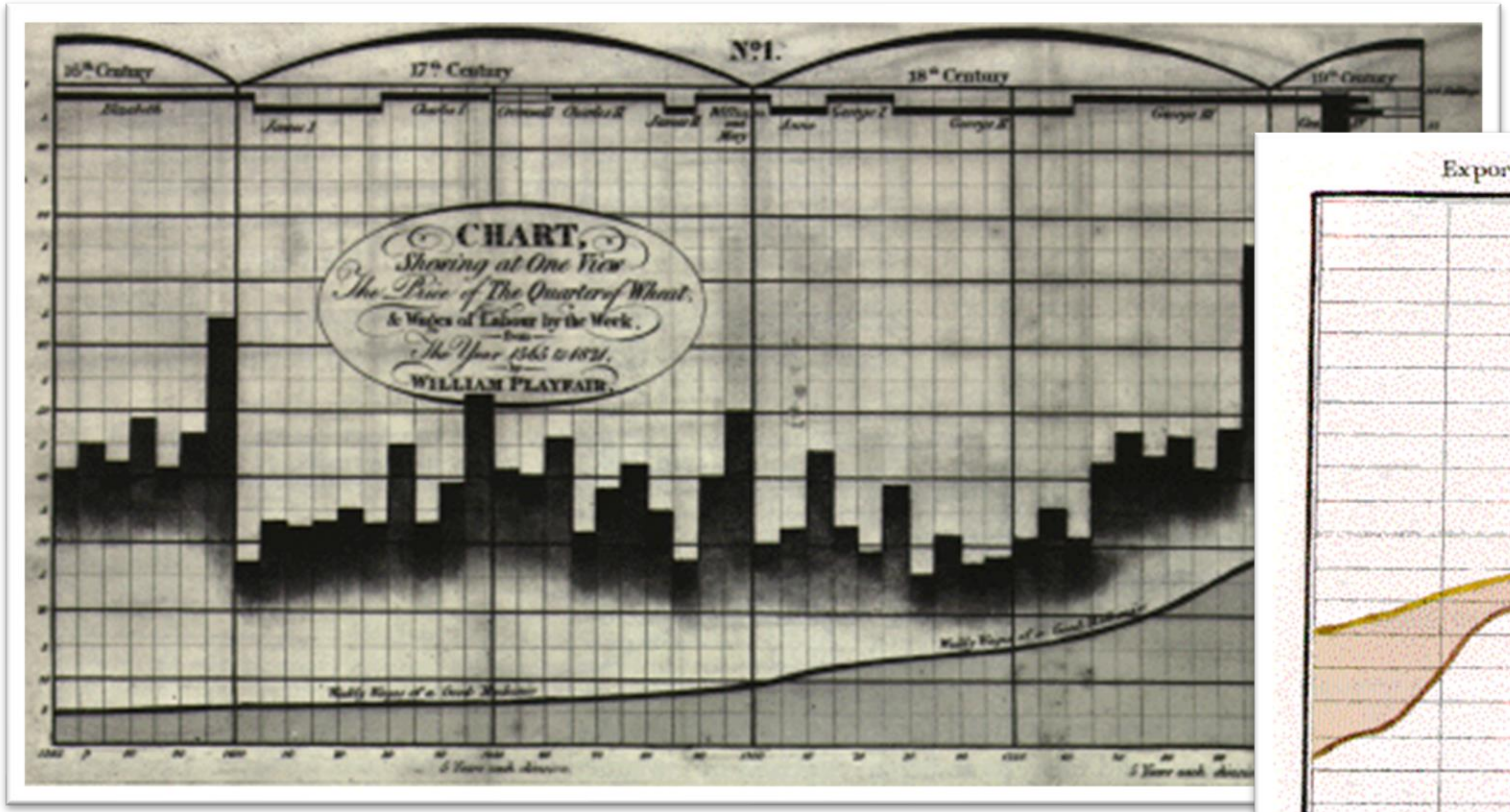


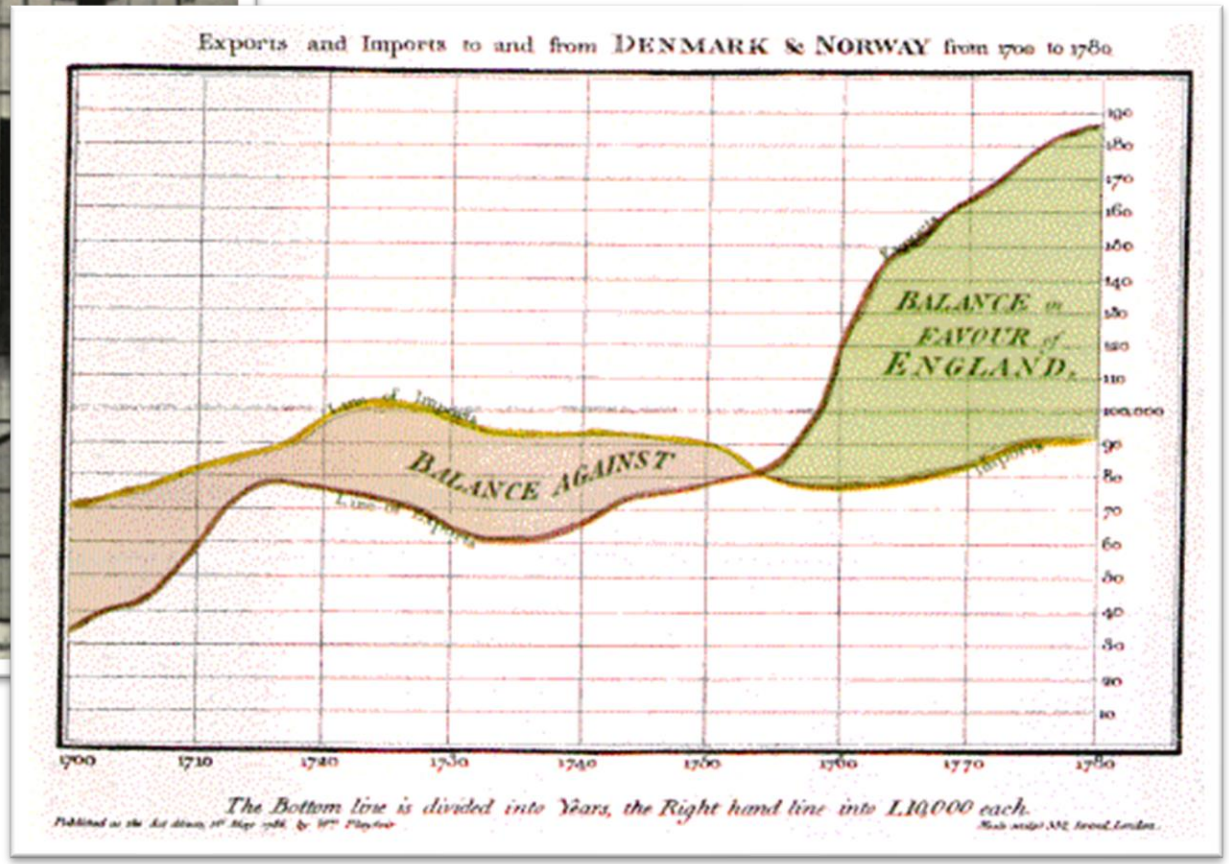


Ref: https://www.archaeology.org/issues/337-1905/features/7544-maps-iraq-babylonian-nippur-map-tablet

William Playfair: The Commercial and Political Atlas (Book)

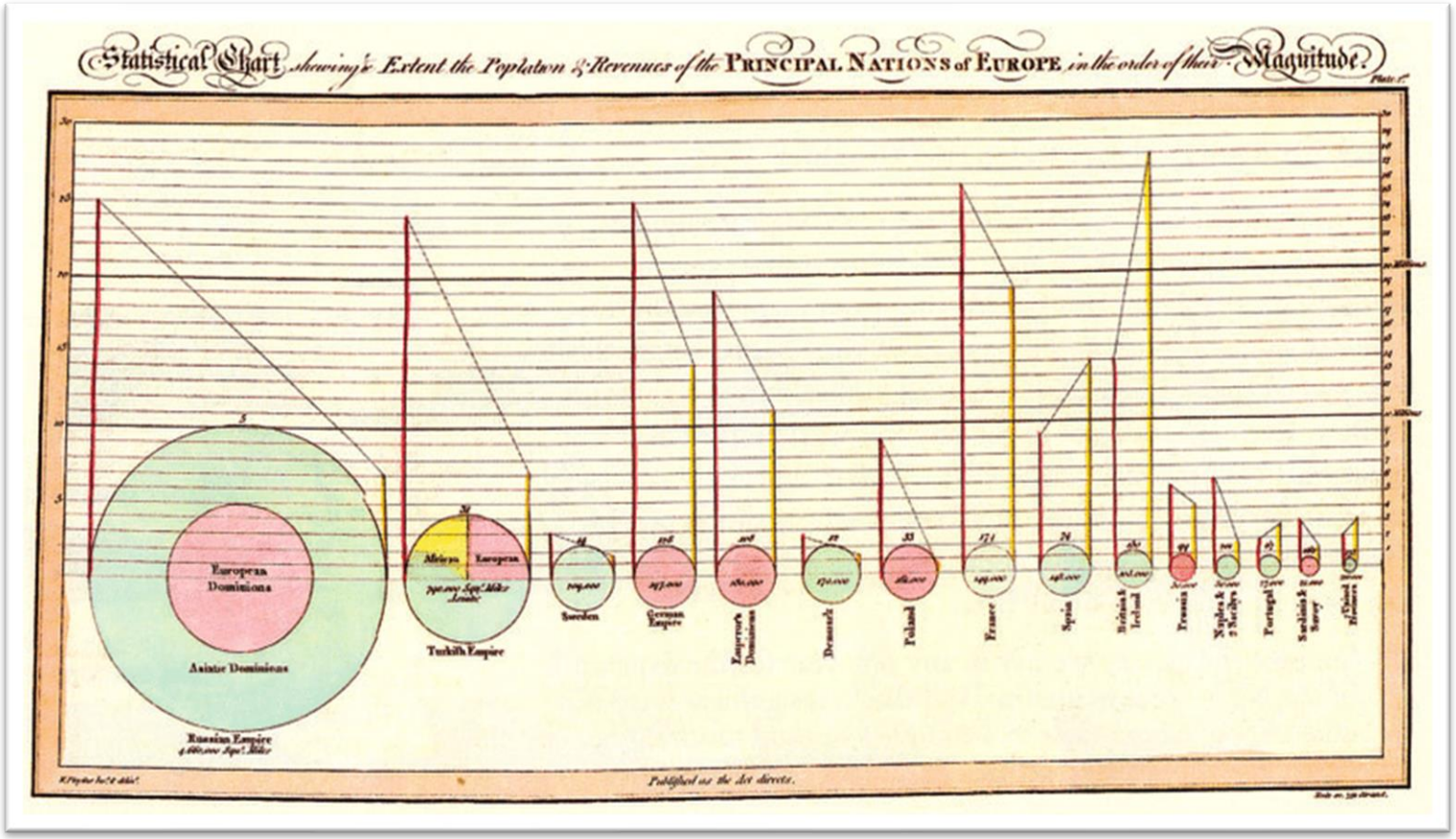
– Abstract Data Presentation

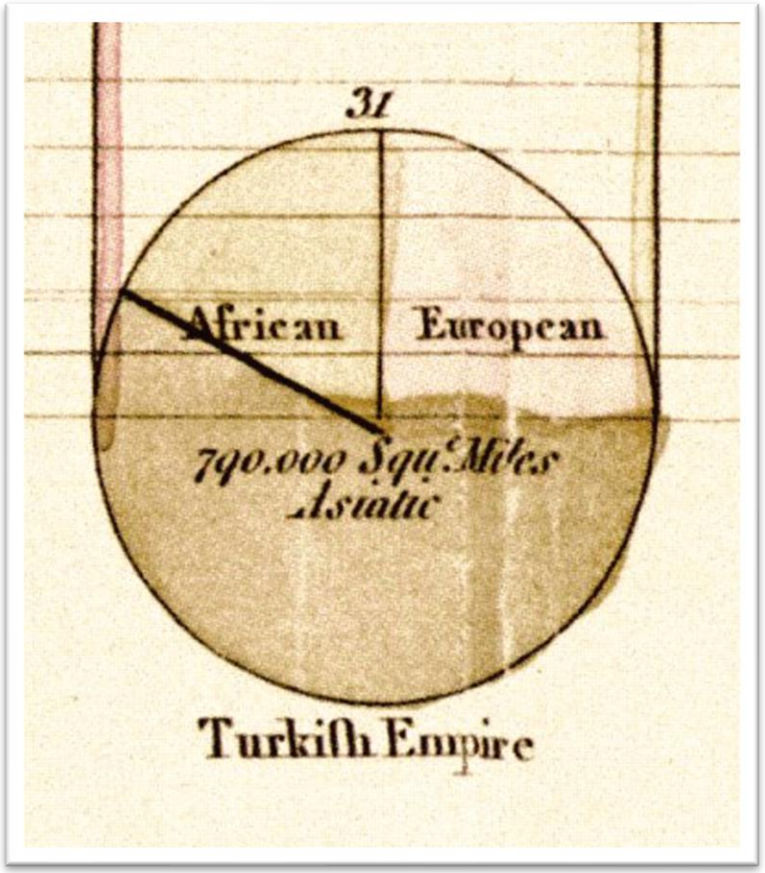




William Playfair: A Statistical Breviary (Book)

– The earliest recorded example of a Pie Chart

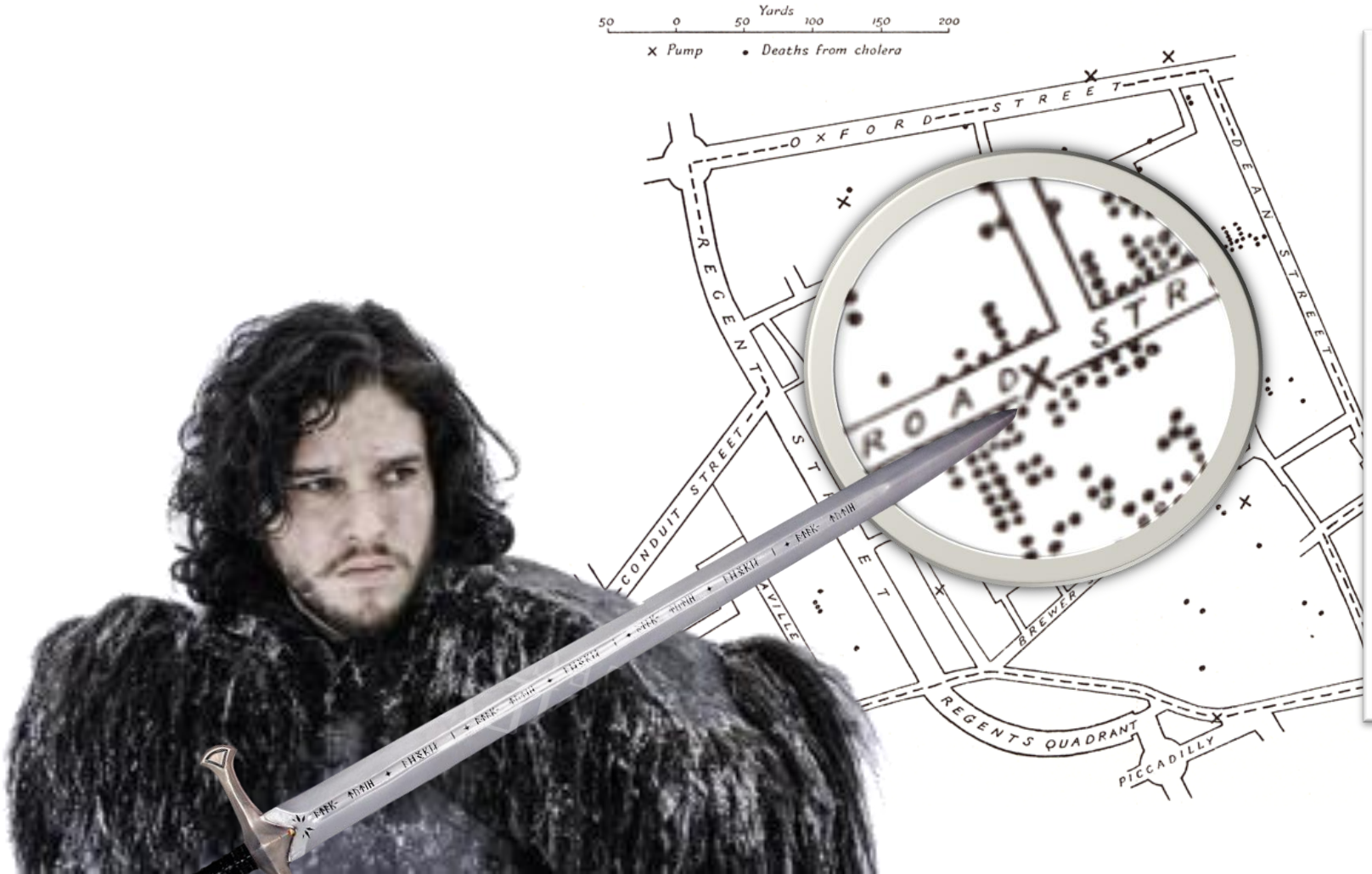




Ref: https://www.researchgate.net/figure/Pie-charts-William-Playfair-1801-From-The-Commercial-and-Political-Atlas_fig6_285544361

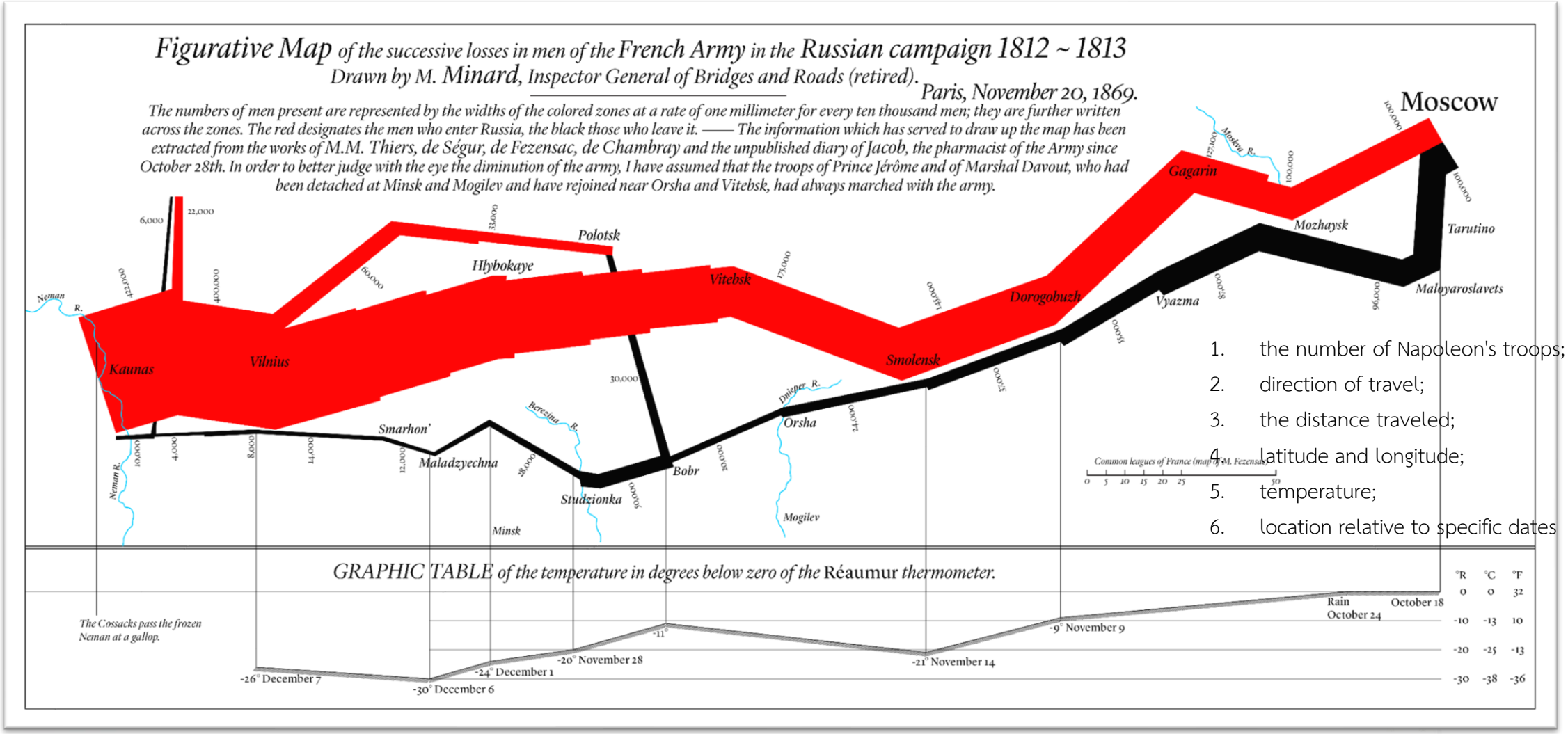
Dr.John Snow: Statistical Map Visualization

– London Cholera Epidemic



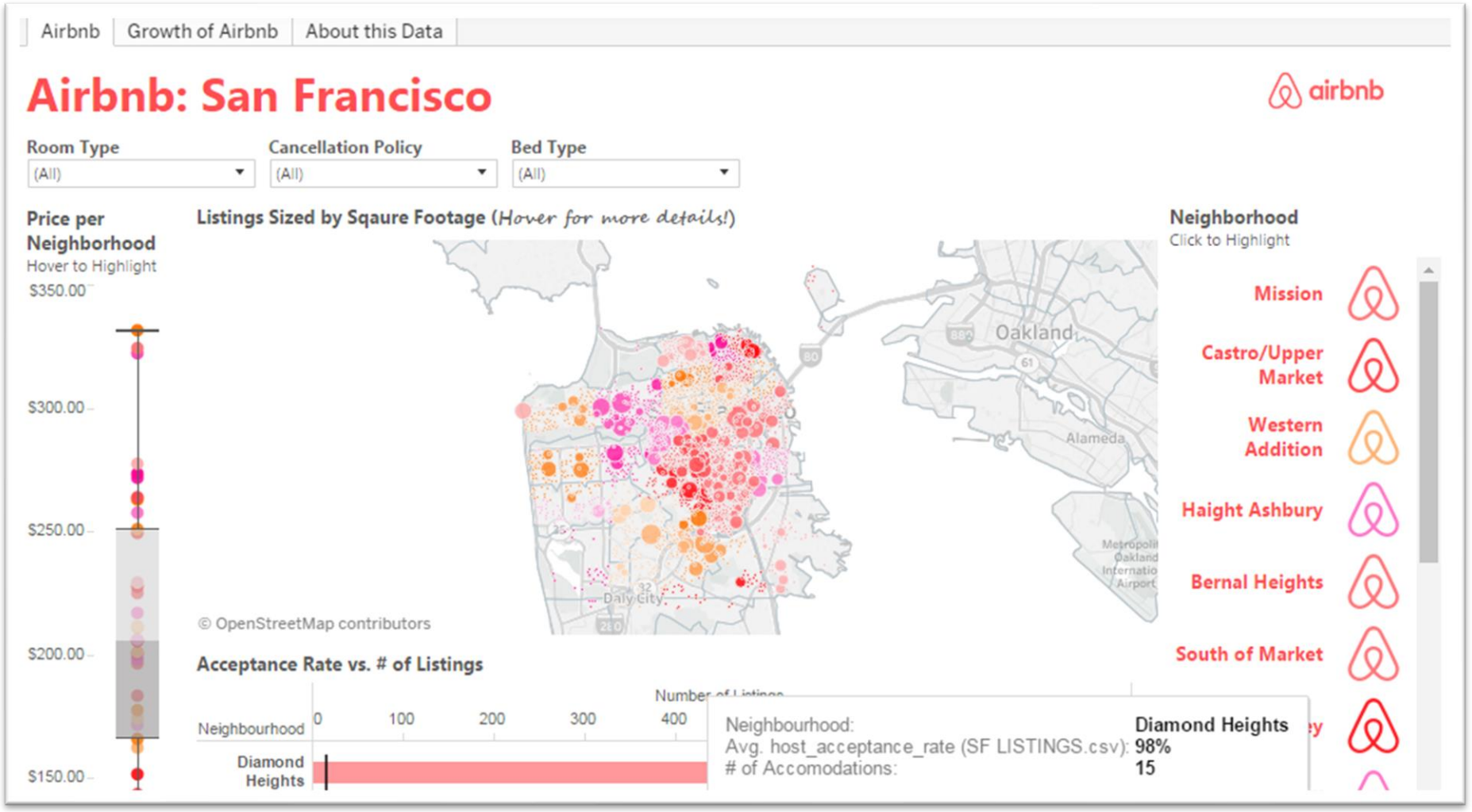
Broad Street Pump

Charles Minard: Napoleon's March



1. the number of Napoleon's troops;
2. direction of travel;
3. the distance traveled;
4. latitude and longitude;
5. temperature;
6. location relative to specific dates

Interactive Visualization



Ref: Ref : https://public.tableau.com/en-us/s/gallery/airbnb-prices-san-francisco

173921017

238745328

701285647

GBDi

173921017

238745328

701285647

Color

GBDi

777771392

101238453

280128564

Position

GBDi

$1_7392101_7$

238_745328

Size

$_701285647$

แล้วก็มาจบที่ตรงนี้

คุณจะอ่านข้อความตรงนี้ก่อน

แล้วก็จะมาอ่านข้อความนี้

ต่อด้วยอันนี้

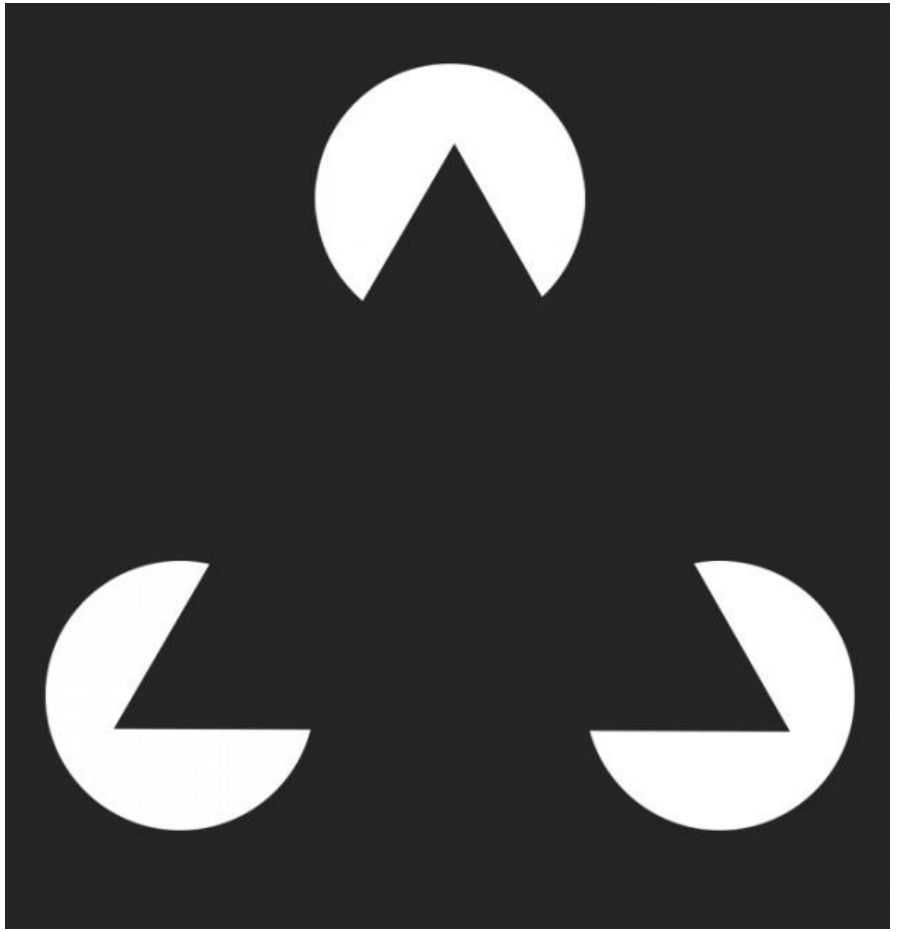
Color

Size

Enclosure

Position

Gestalt Principles of Visual Perception



Source: The Inspired Eye

Source: Gizmodo

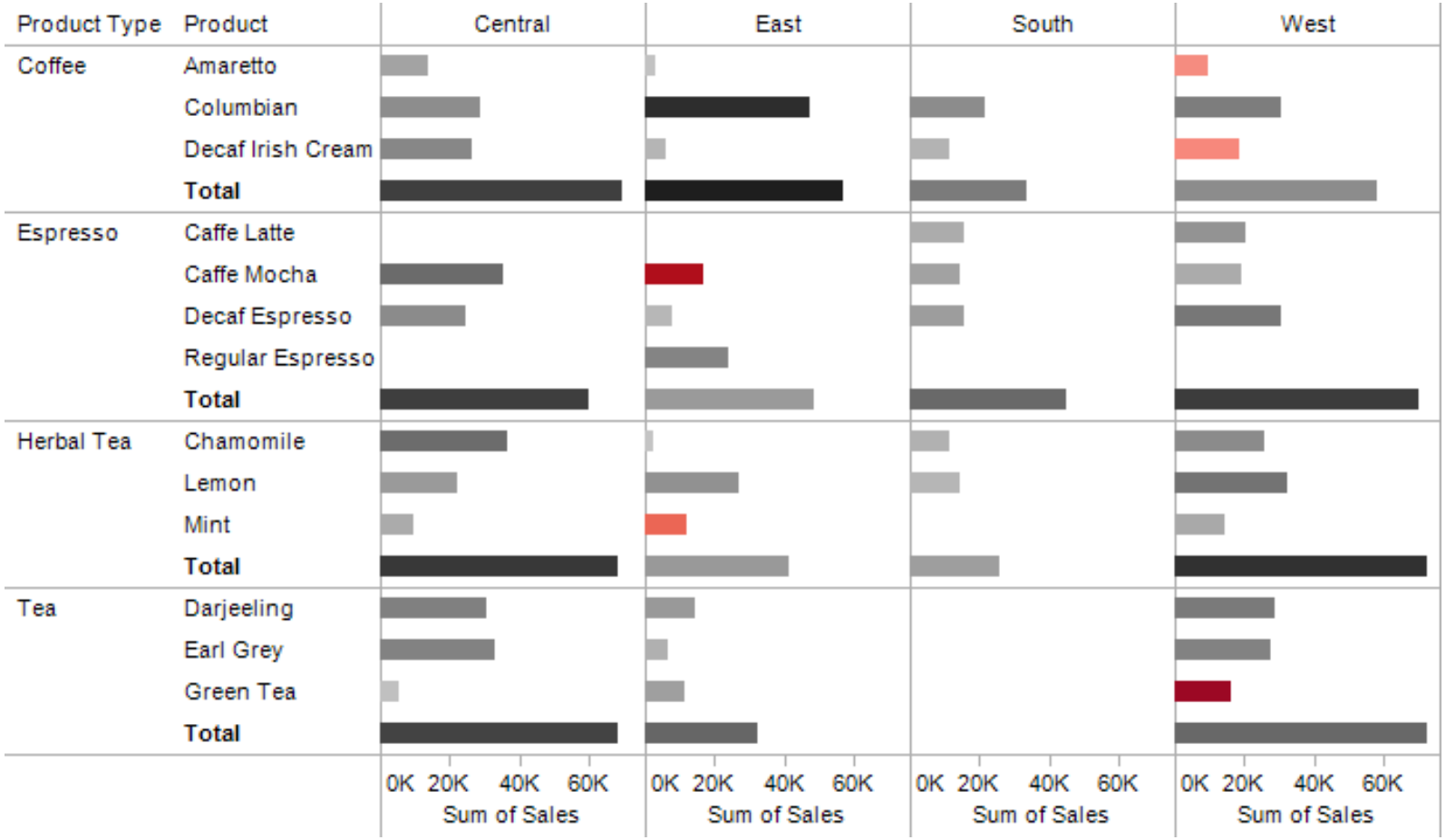
Bertin's Three Levels of Reading

Elementary: single value

		Central		East		South		West	
Product Type	Product	Sum of Profit	Sum of Sales	Sum of Profit	Sum of Sales	Sum of Profit	Sum of Sales	Sum of Profit	Sum of Sales
Coffee	Amaretto	$5,105	$14,011	$1,009	$2,993			($1,225)	$9,265
	Columbian	$8,528	$28,913	$27,253	$47,386	$8,767	$21,664	$11,253	$30,357
	Decaf Irish Cream	$9,632	$26,155	$2,727	$6,261	$2,933	$11,592	($1,305)	$18,235
	Total	$23,265	$69,080	$30,989	$56,640	$11,700	$33,256	$8,724	$57,856
Espresso	Caffe Latte					$3,872	$15,442	$7,502	$20,458
	Caffe Mocha	$14,640	$35,218	($6,230)	$16,646	$5,201	$14,163	$4,064	$18,876
	Decaf Espresso	$8,860	$24,485	$2,410	$7,722	$5,930	$15,384	$12,302	$30,578
	Regular Espresso			$10,062	$24,036				
	Total	$23,500	$59,703	$6,242	$48,405	$15,003	$44,989	$23,868	$69,911
Herbal Tea	Chamomile	$14,434	$36,570	$765	$2,194	$3,180	$11,186	$8,852	$25,632
	Lemon	$6,251	$21,978	$7,901	$27,176	$2,593	$14,497	$13,120	$32,274
	Mint	$4,069	$9,337	($2,242)	$11,992			$4,330	$14,380
	Total	$24,754	$67,885	$6,424	$41,362	$5,774	$25,683	$26,301	$72,285
Tea	Darjeeling	$10,772	$30,289	$6,497	$14,096			$11,780	$28,769
	Earl Grey	$10,331	$32,881	$3,405	$6,505			$10,425	$27,387
	Green Tea	$1,227	$5,211	$5,654	$11,571			($7,109)	$16,063
	Total	$22,330	$68,380	$15,557	$32,172			$15,097	$72,220

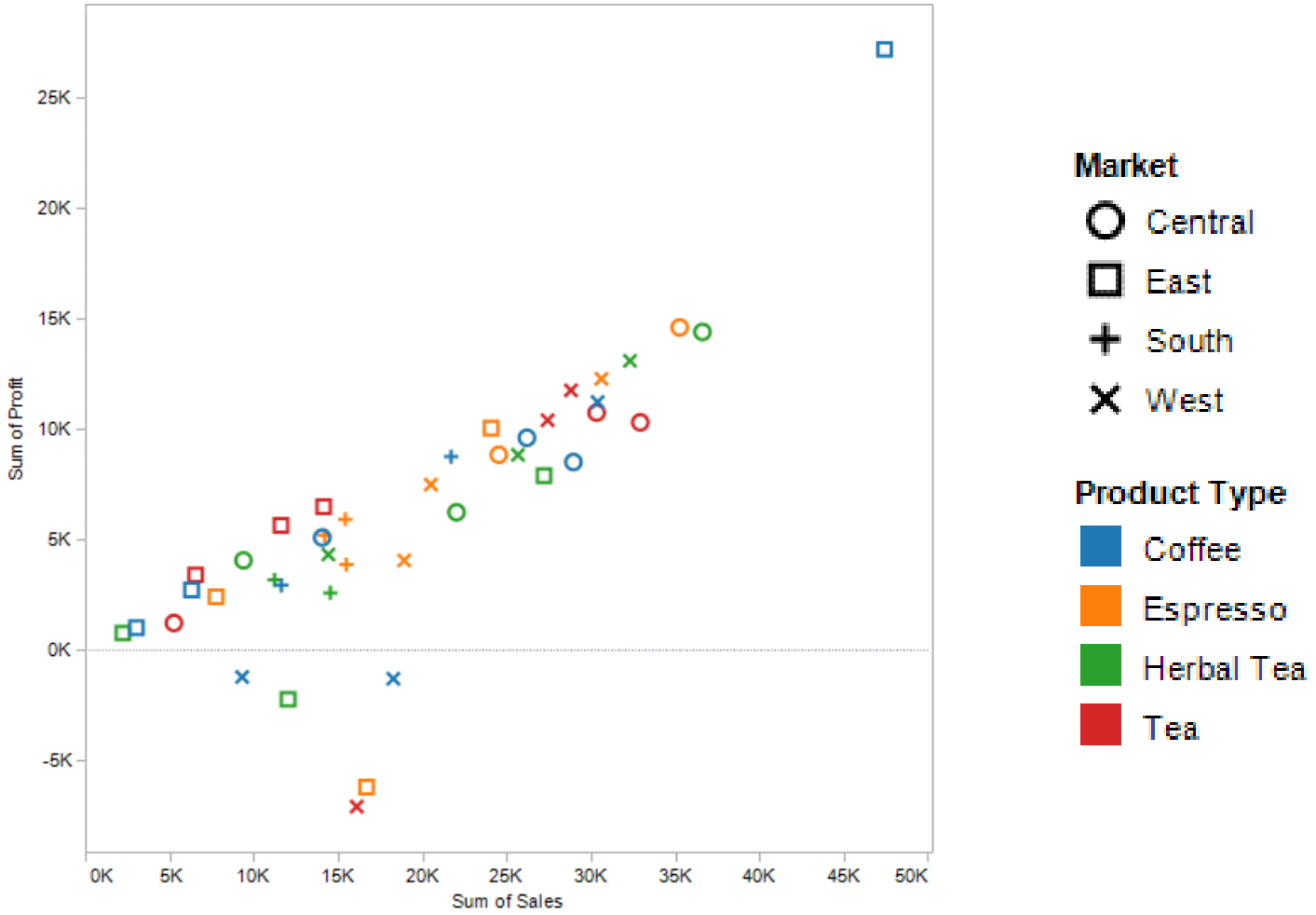
Bertin’s Three Levels of Reading

Intermediate: relationships between values



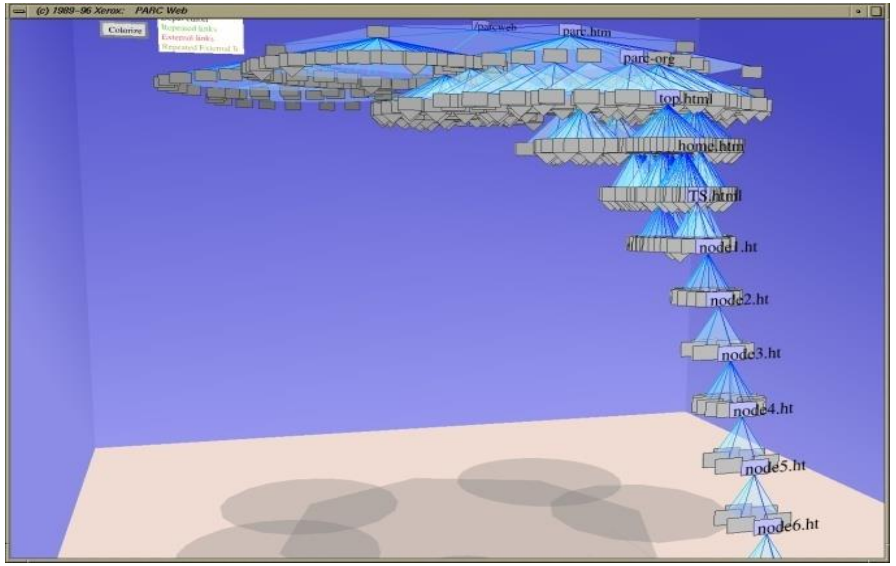
Bertin's Three Levels of Reading

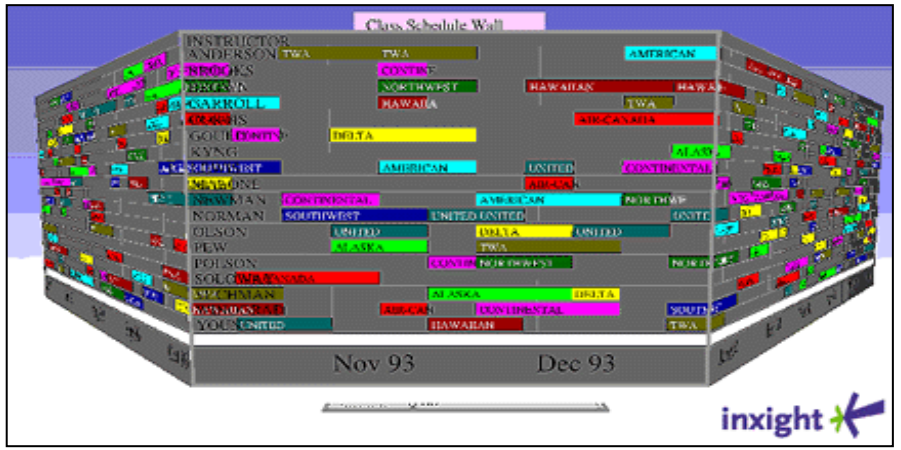
Global: relationships of the whole



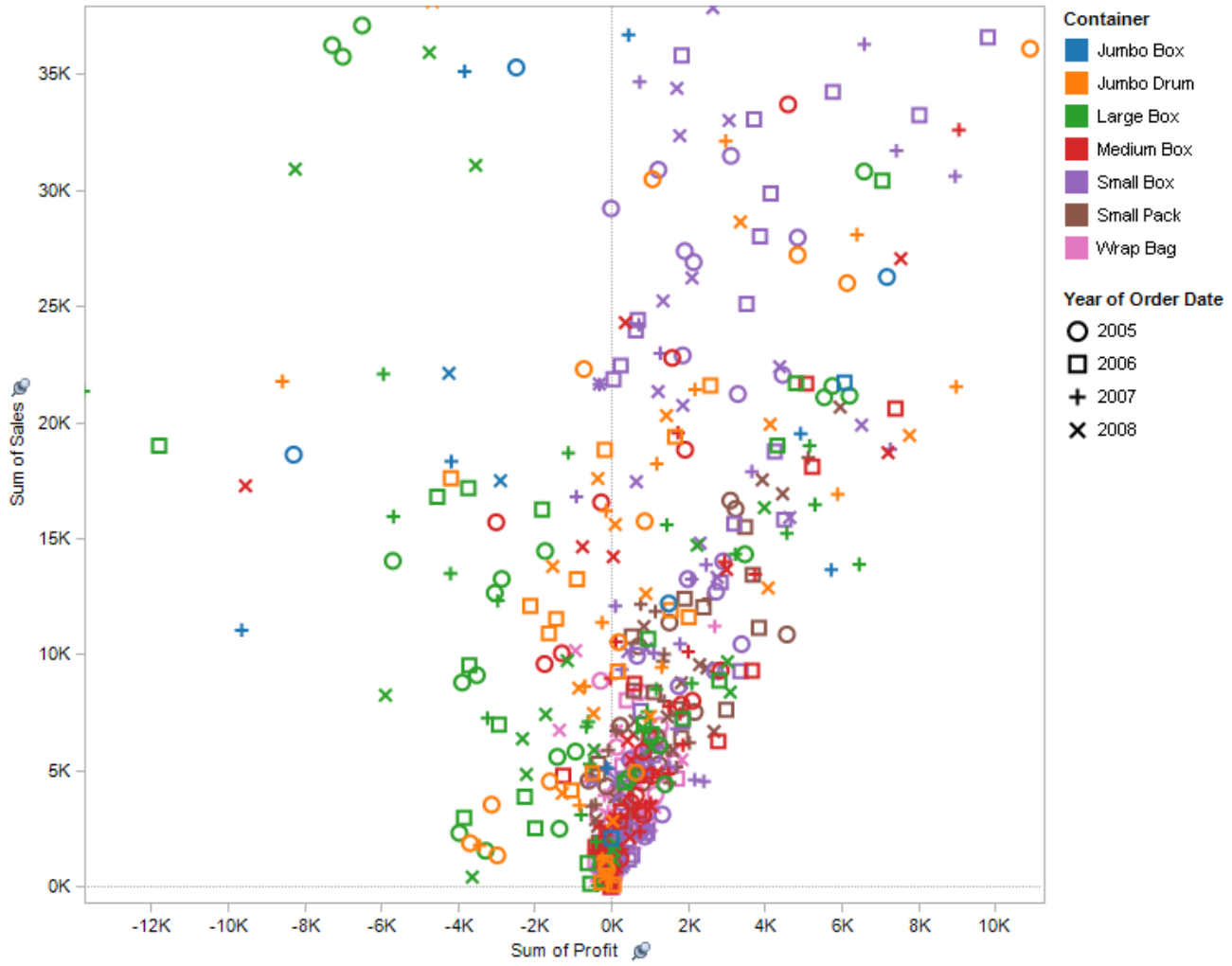
Human Perception is Limited

- Only adds a single dimension
- Creates occlusions
- Adds orientation complexities
- Easy to get lost
- Suggests a physical metaphor

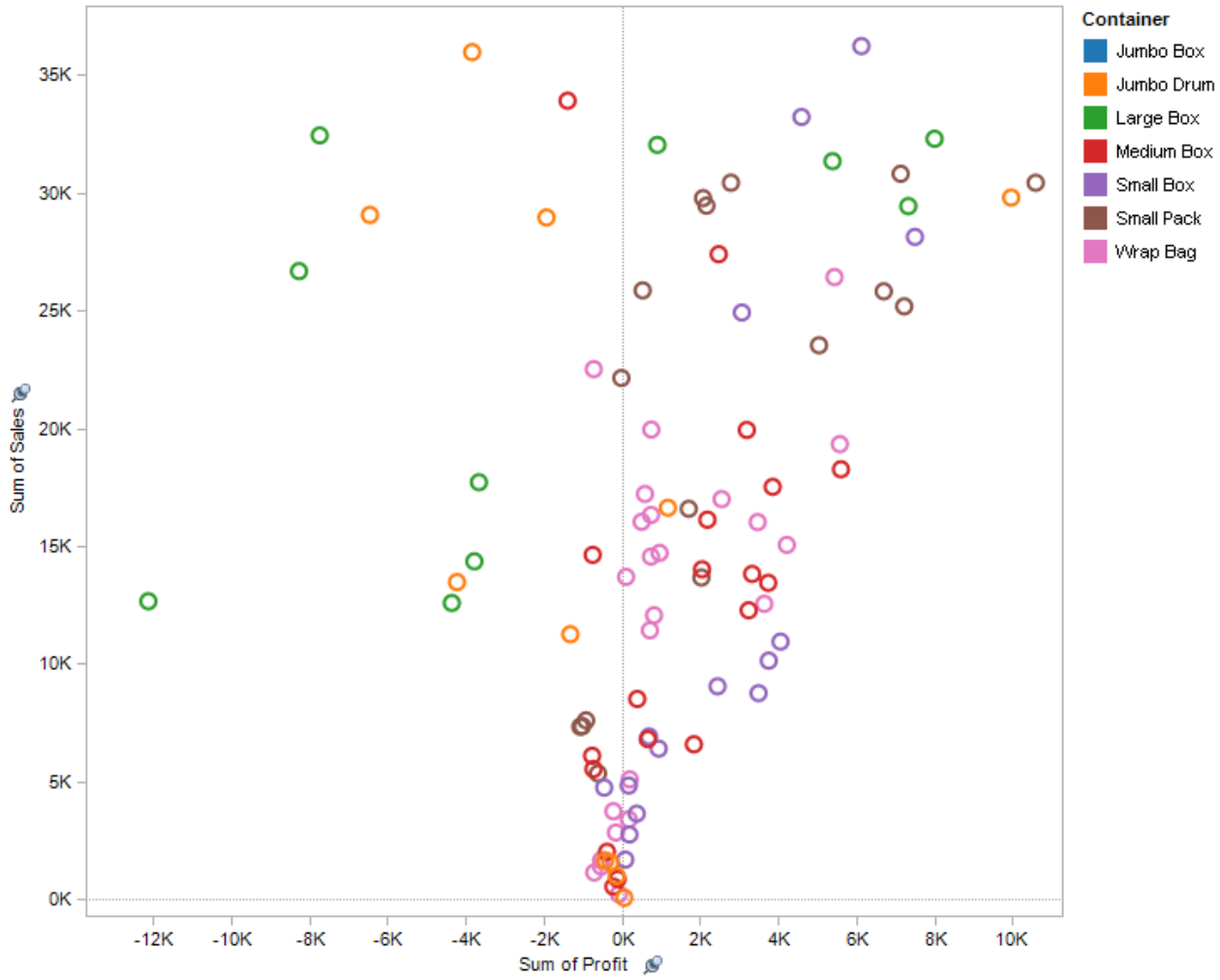




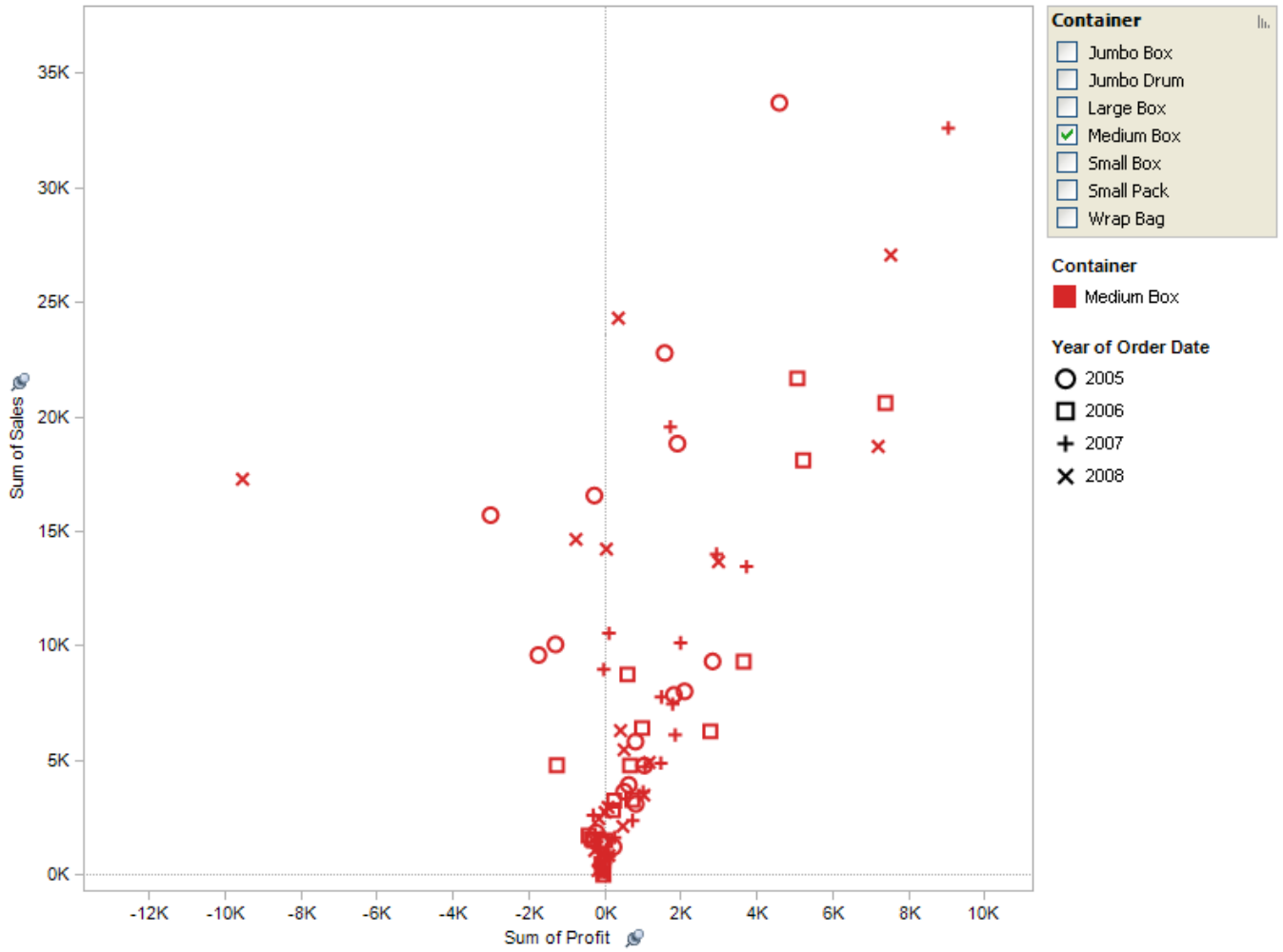
Interactivity: Too Much Data Scenario



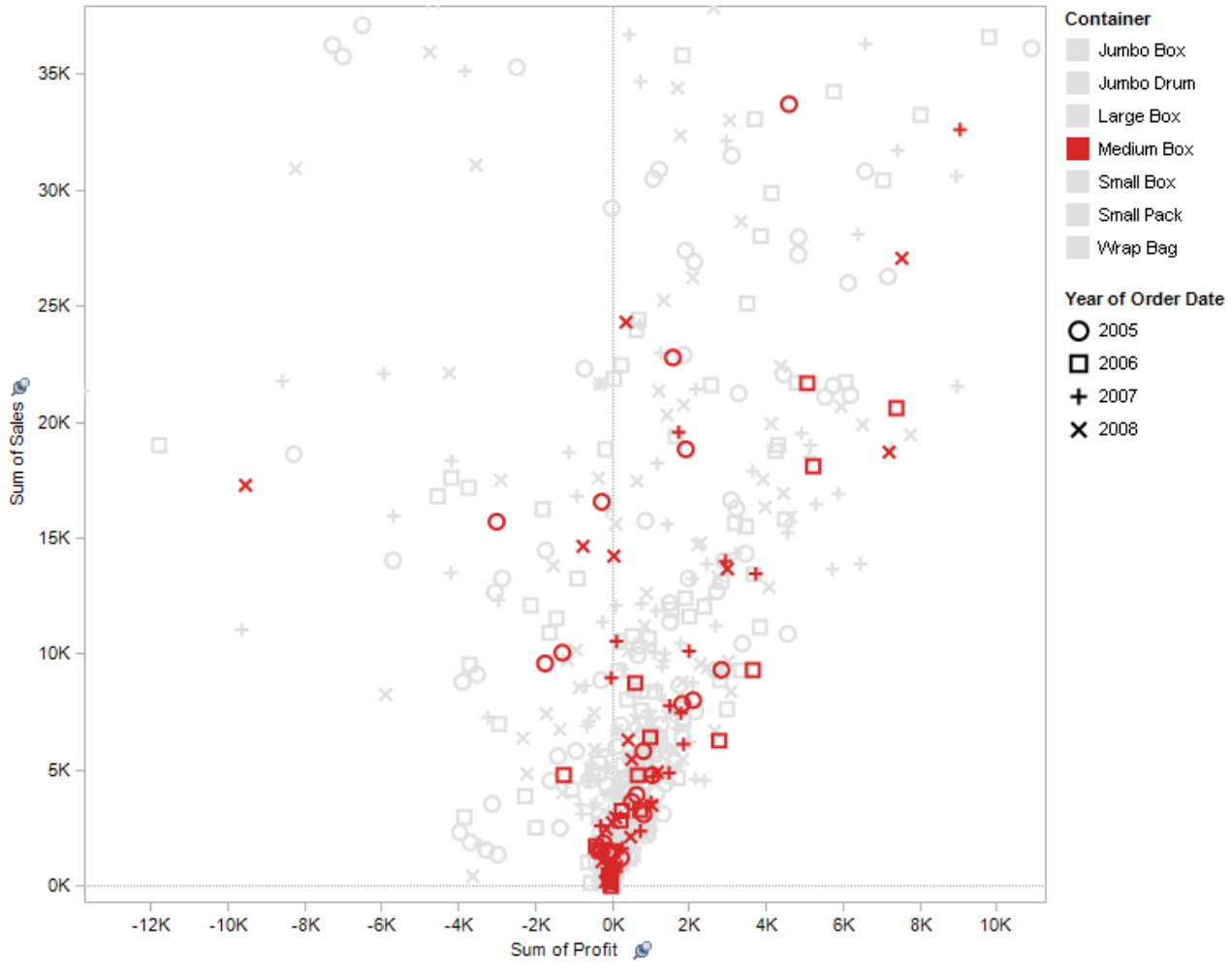
Interactivity: Aggregation



Interactivity: Filtering



Interactivity: Brushing



What do you want to tell from your data?

Story

1.Comparison

3.Composition

2.Relationship

4.Distribution

1.Comparison

Bar chart

Line chart

Bullet chart

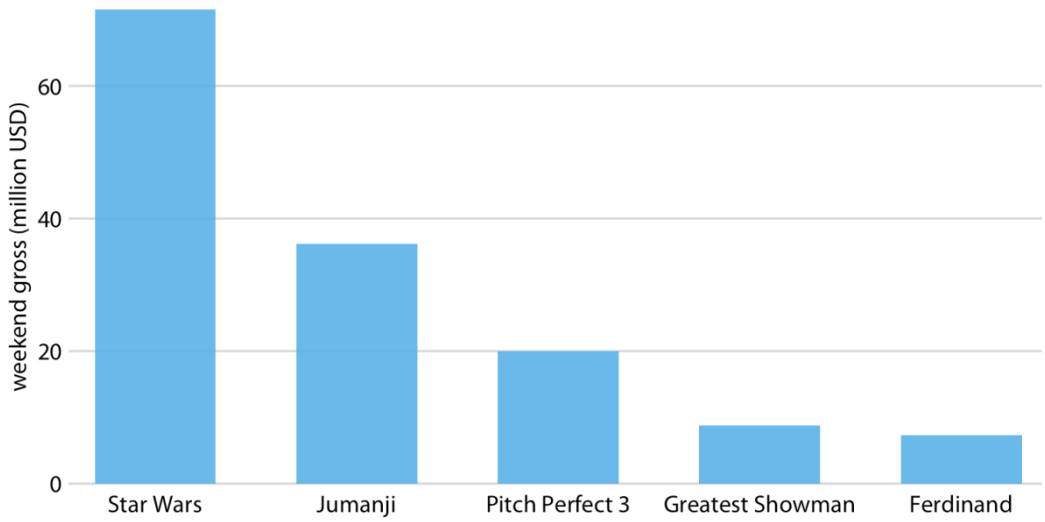
Bar chart

When to use

- Comparing data across categories

Possible extension

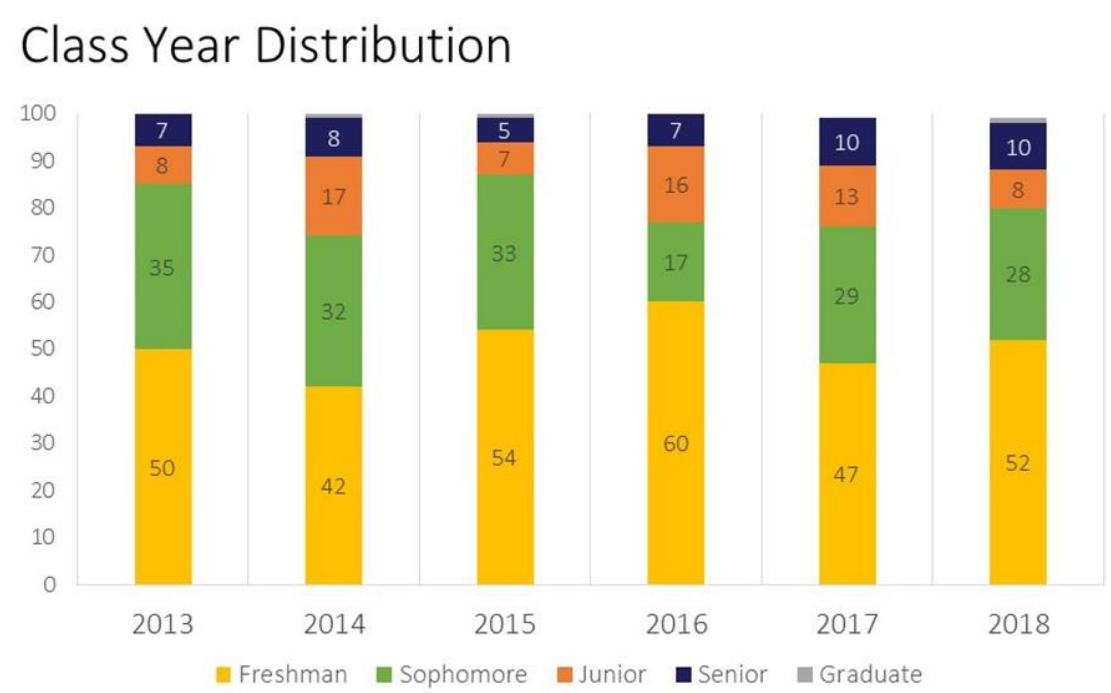
- Include multiple bar charts on a dashboard
- Add color to bars for more impact
- Use stacked bars or side-by-side bars
- Combine bar charts with maps
- Put bars on both sides of an axis

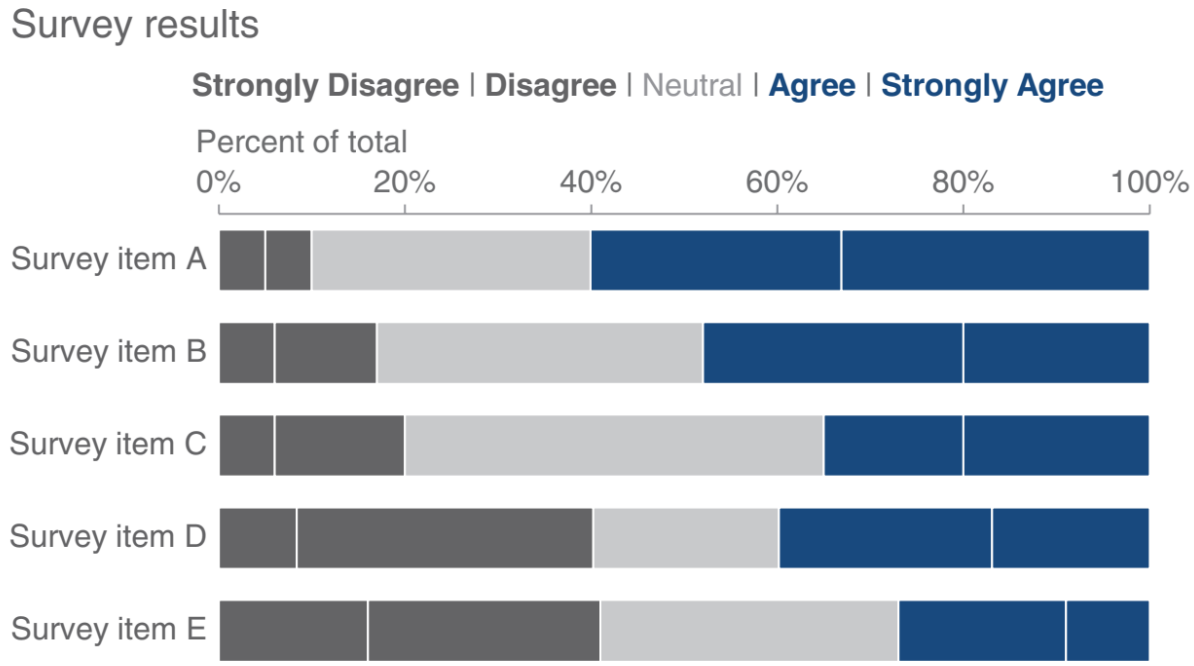


Bar chart


Transport Incident Dates
Transports
0
1
2
3
4
5
6
7
08-Sep-18
09-Sep-18
13-Sep-18
14-Sep-18
15-Sep-18
16-Sep-18
21-Sep-18
22-Sep-18
23-Sep-18
29-Sep-18
30-Sep-18
08-Oct-18
13-Oct-18
14-Oct-18
18-Oct-18
19-Oct-18
20-Oct-18
21-Oct-18
26-Oct-18
27-Oct-18
28-Oct-18
02-Nov-18
04-Nov-18
10-Nov-18
11-Nov-18
16-Nov-18
17-Nov-18
30-Nov-18
01-Dec-18
02-Dec-18
08-Dec-18
09-Dec-18
14-Dec-18

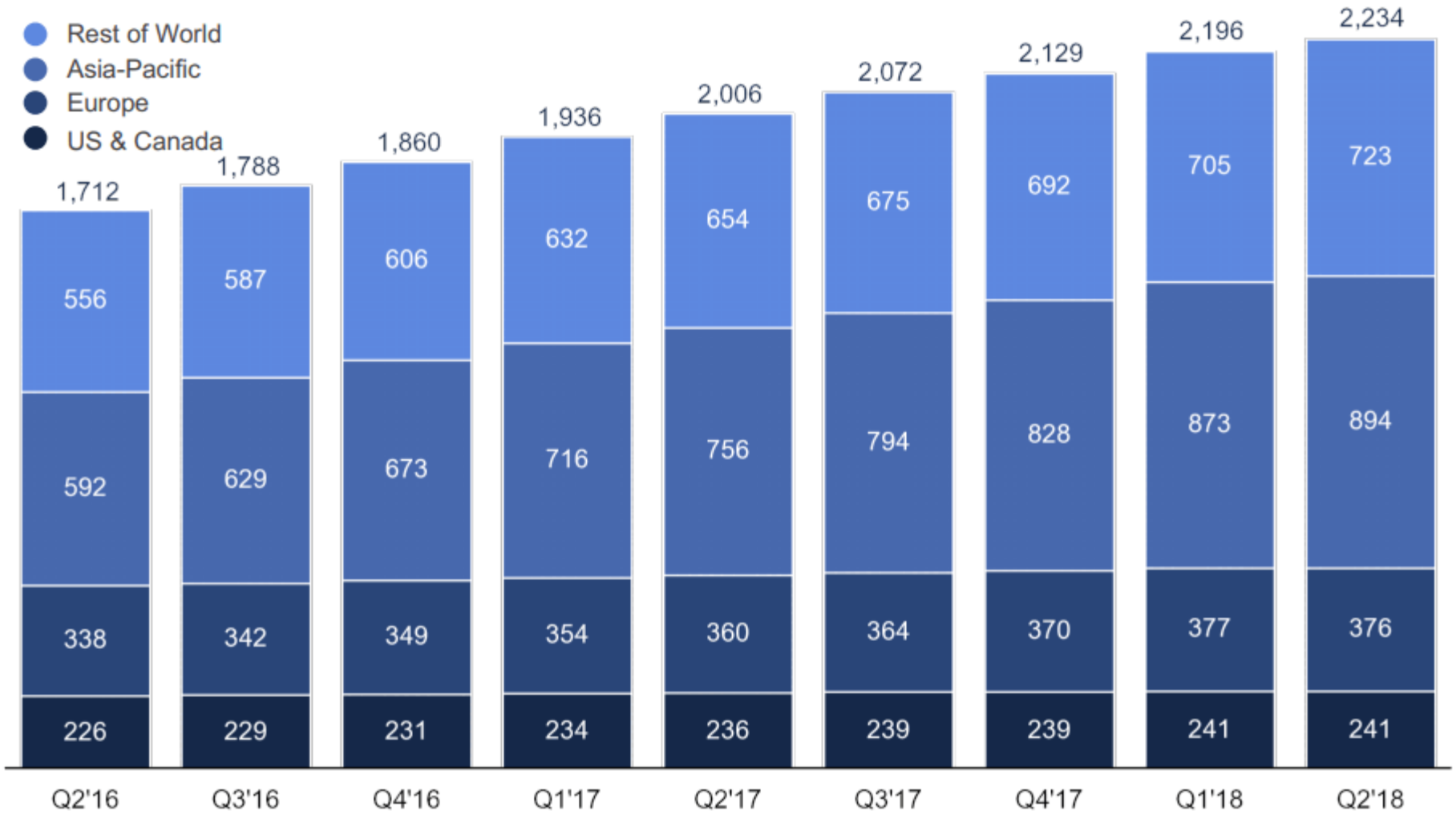
Bar chart (stacked)





Bar chart (side by side)

FACEBOOK STATISTICS 2018

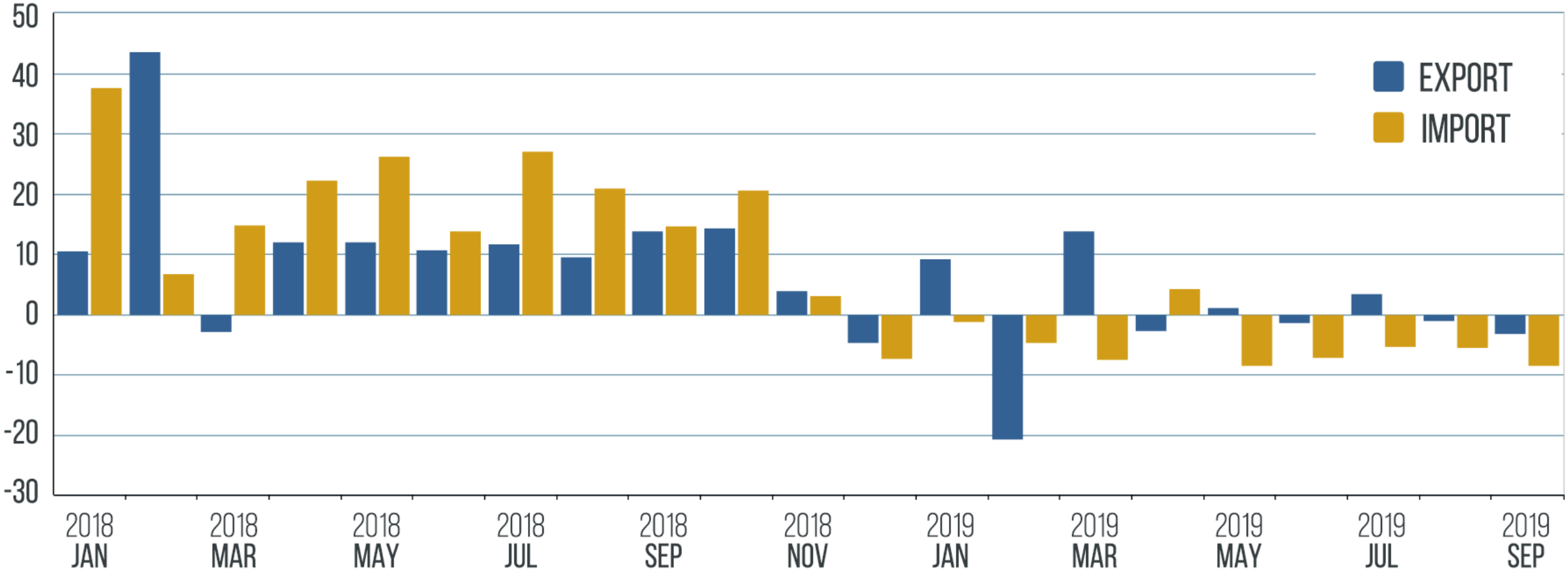


Monthly active users (MAUs), in millions

China: Export and Import Trends, Jan 2018 - Sept 2019

Data Driven

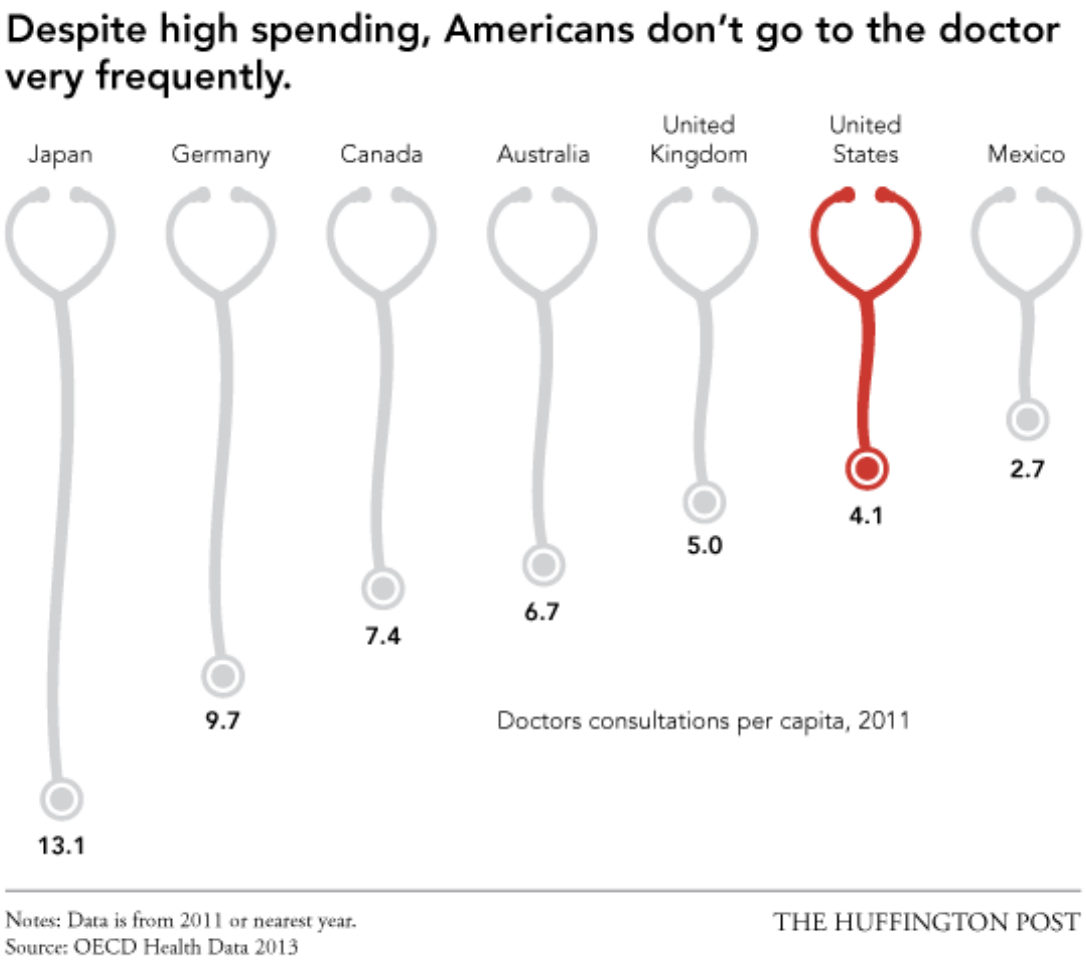
Y-o-Y % Change



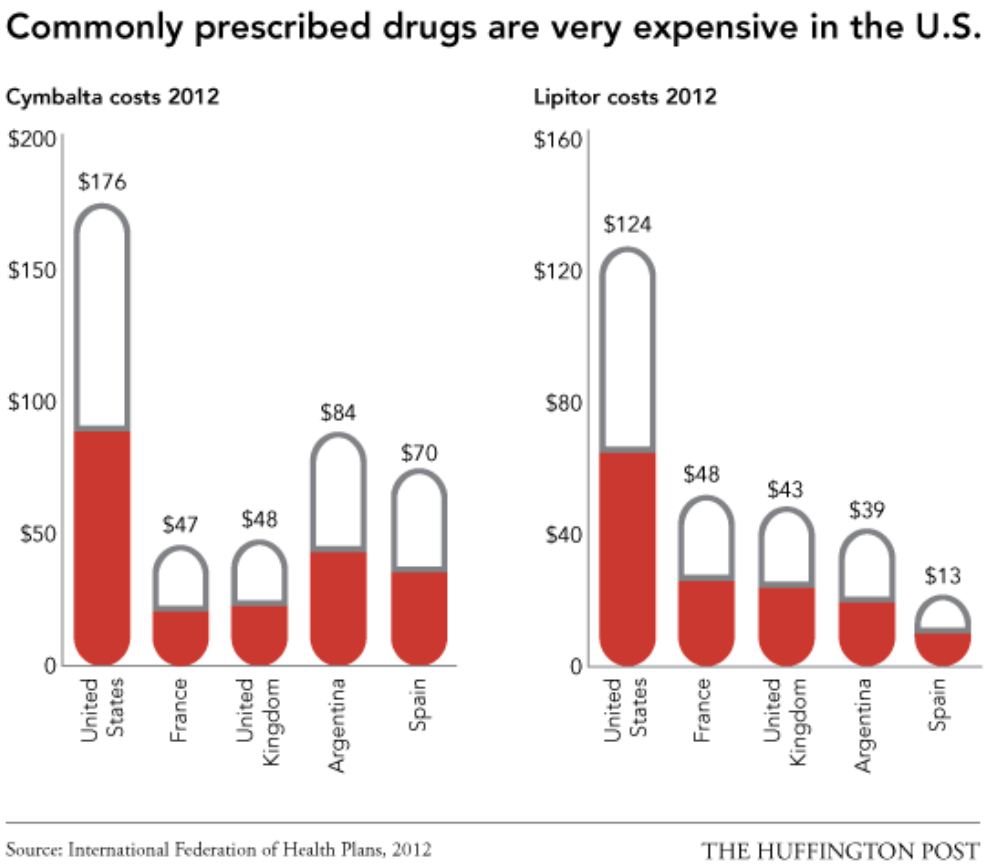
knoema

Source: General Administration of Customs of the People's Republic of China

Bar chart doesn’t have to be boring



Bar chart doesn’t have to be boring



- Missed opportunity (and can be misleading)
- Not a stacked bar chart but could be.

สถิติอุบัติเหตุ 7 วันอันตราย

สงกรานต์ 2561-2562

(วันที่ 11-17 เมษายน)

AMARIN NEWS 34 HD

2561
2562

อุบัติเหตุรวม

3,724 ครั้ง

3,338 ครั้ง

บาดเจ็บรวม

3,897 คน

3,442 คน

ผู้เสียชีวิต

418 ราย

386 ราย

จังหวัดที่มีผู้เสียชีวิตเยอะสุด

ลพบุรี-อุดรธานี 15 ราย

สาเหตุการเกิดอุบัติเหตุสูงสุด

เมาแล้วขับ

ที่มา : ศูนย์อำนวยความปลอดภัยทางถนน

เปรียบเทียบจำนวนและอัตราการว่างงาน พ.ศ. 2561 และ 2562

ผู้มีงานทำ

ก.ค. 61 38.63 ล้านคน

ก.ค. 62 37.62 ล้านคน

ลดลง

1.0 ล้านคน

ผู้ว่างงาน

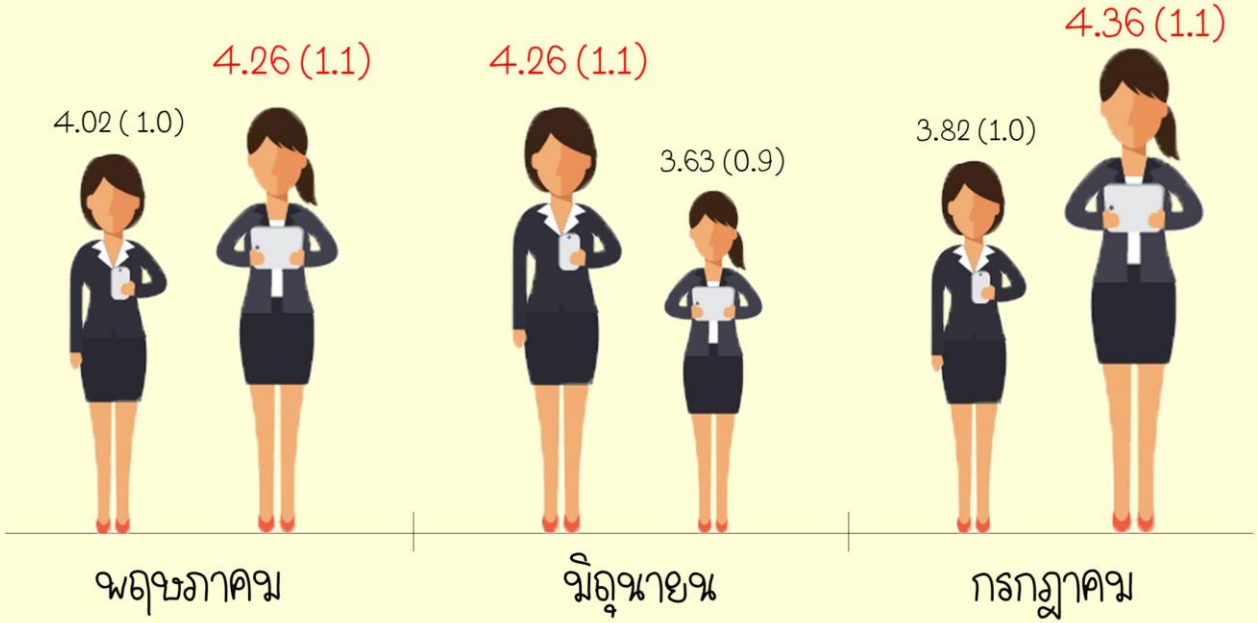
ก.ค. 61 3.82 แสนคน

ก.ค. 62 4.36 แสนคน

เพิ่มขึ้น

5.4 หมื่นคน

หน่วย : แสนคน (อัตราว่างงาน)



ที่มา : การสำรวจภาวะการทำงานของประชากรเดือนกรกฎาคม พ.ศ. 2562 สำนักงานสถิติแห่งชาติ กระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคม

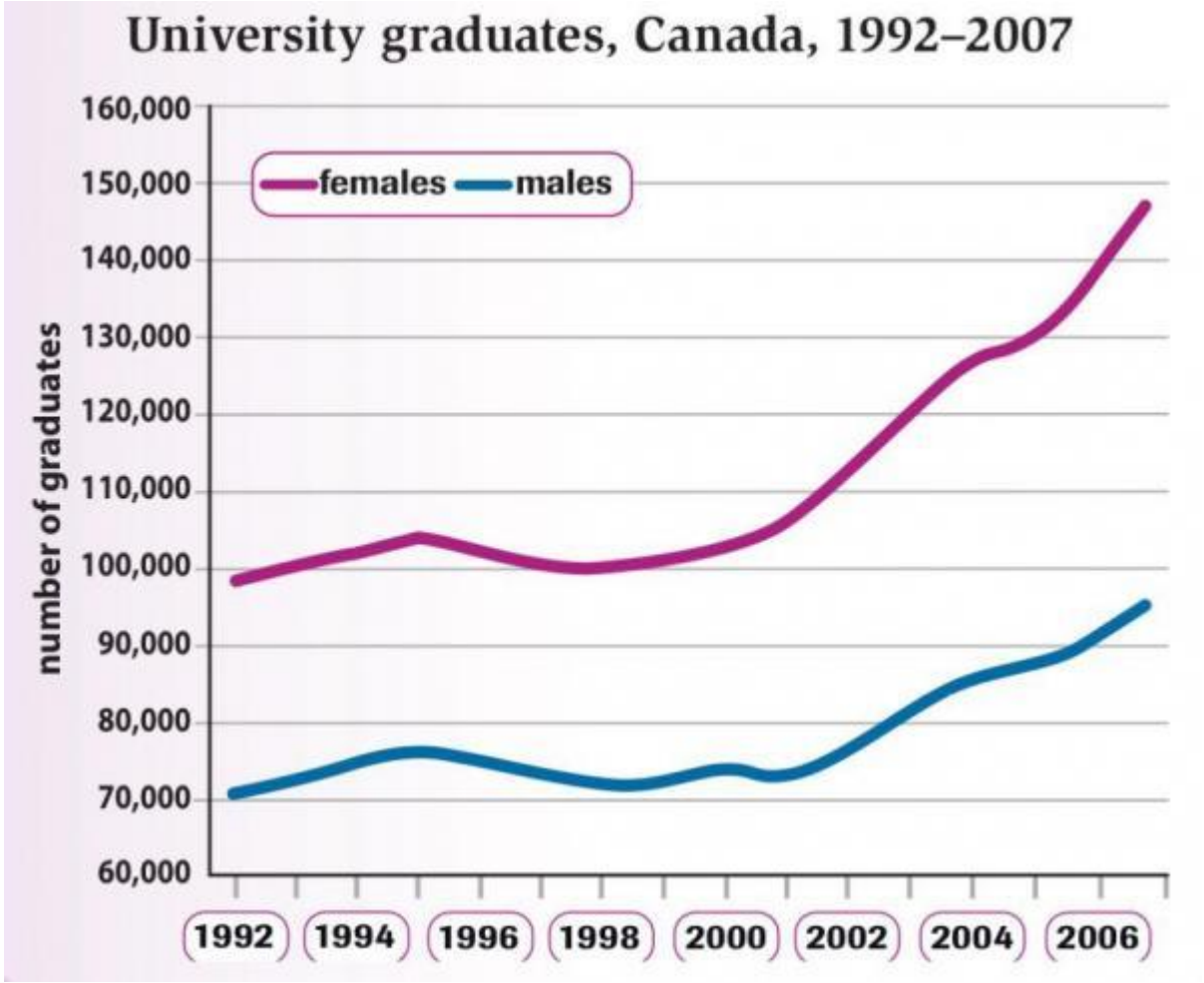
Line chart

When to use

- Viewing trends in data over time

Possible extension

- Combine a line graph with bar charts
- Shade the area under lines.



https://www.testbig.com/ielts-writing-task-i-ielts-academic-essays/graph-below-shows-number-university-graduates-canada-0

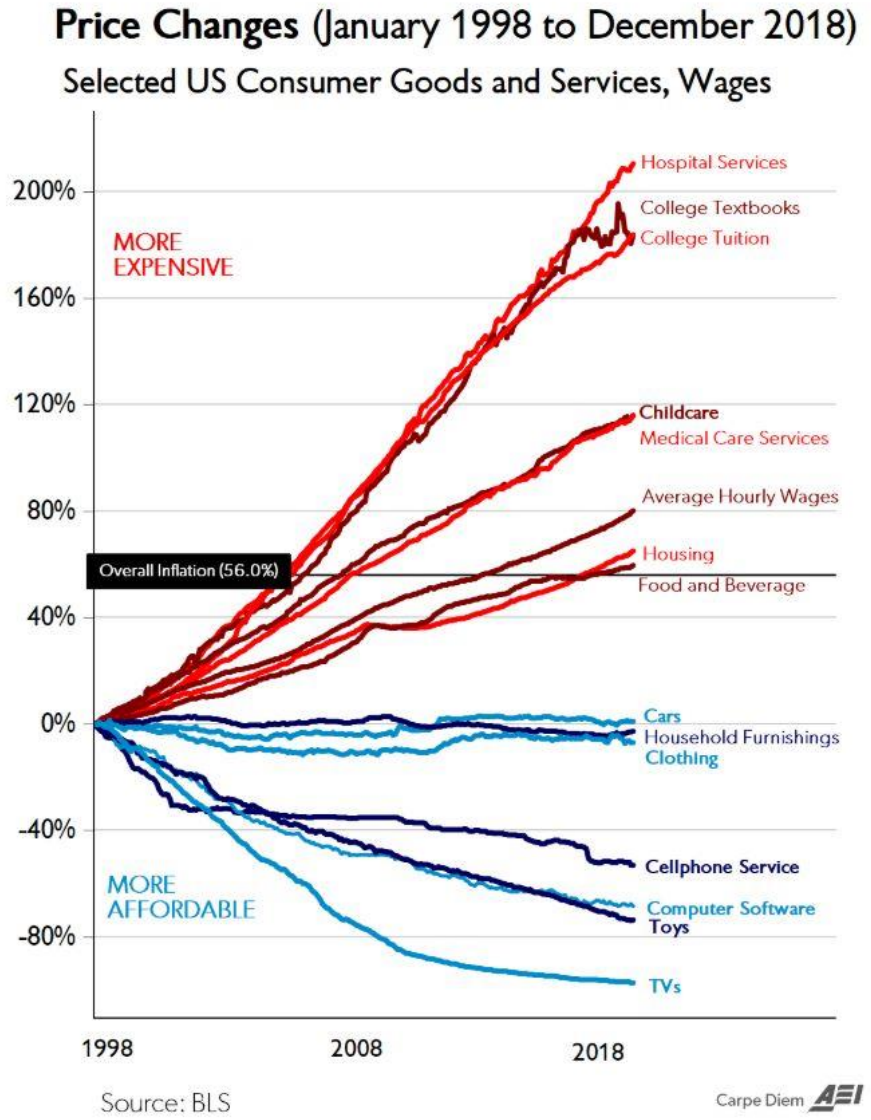
Line chart

When to use

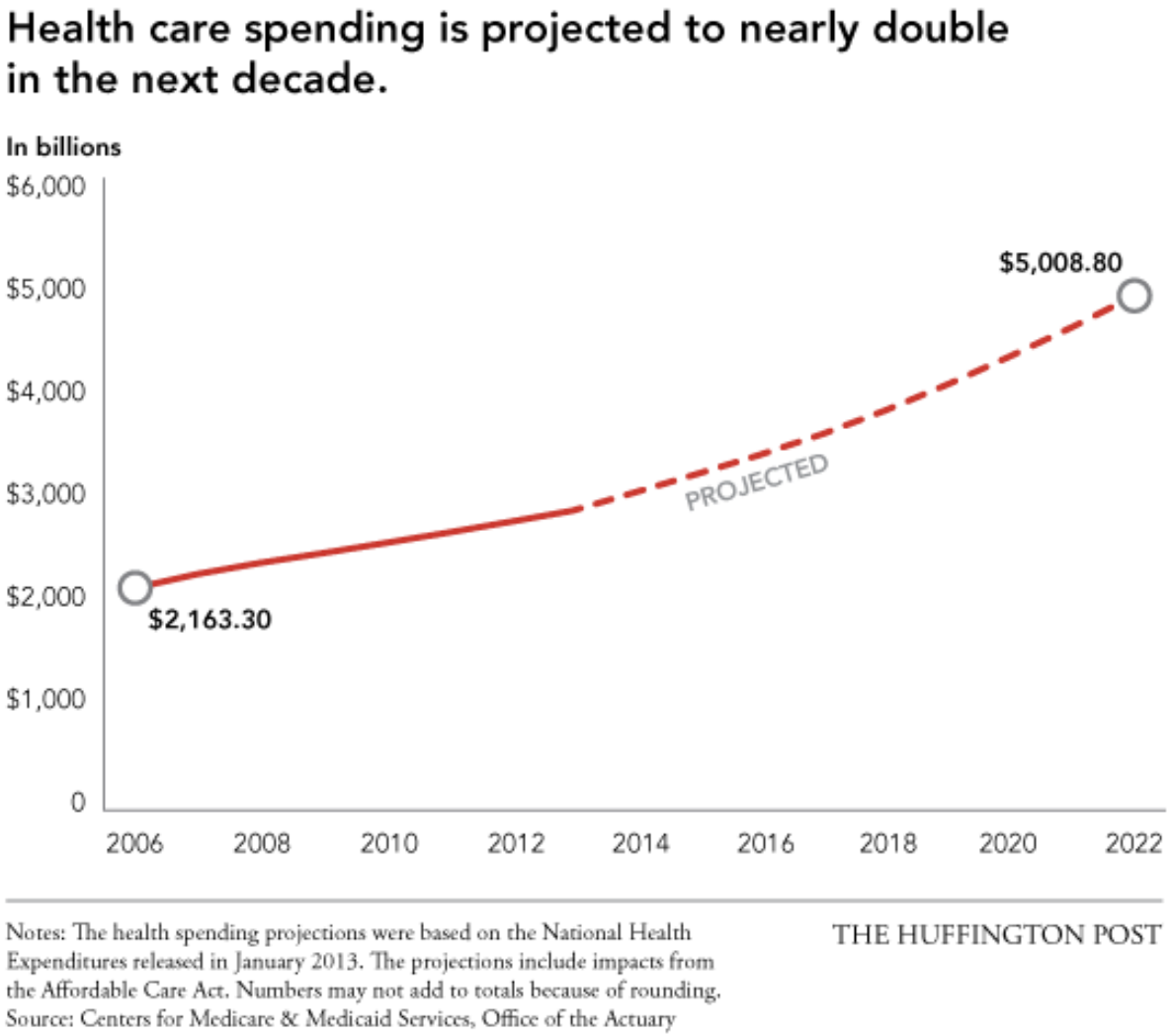
- Viewing trends in data over time

Possible extension

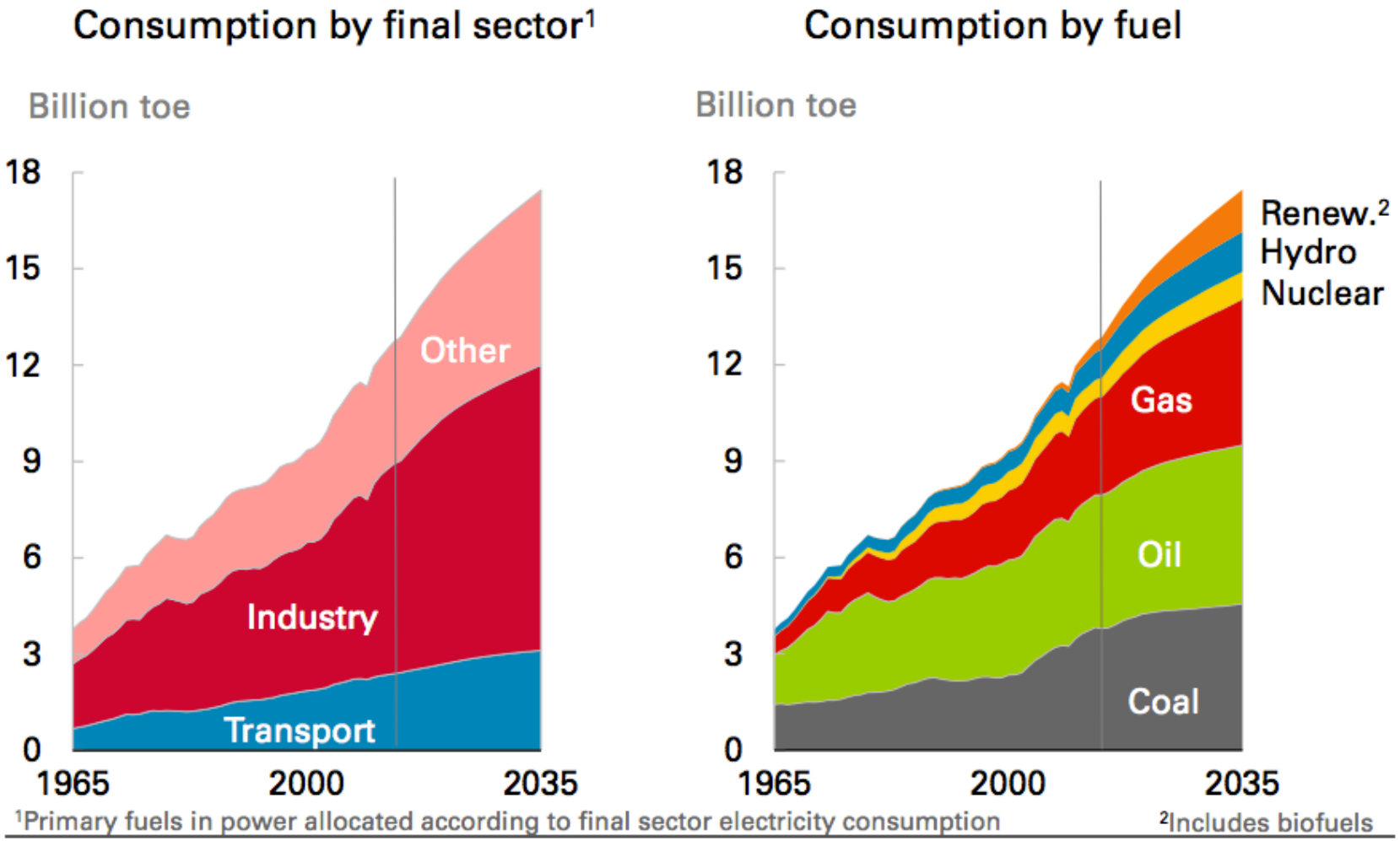
- Use different types/colors of lines to signify different data/meanings
- Combine a line graph with bar charts
- Shade the area under lines (Area chart).



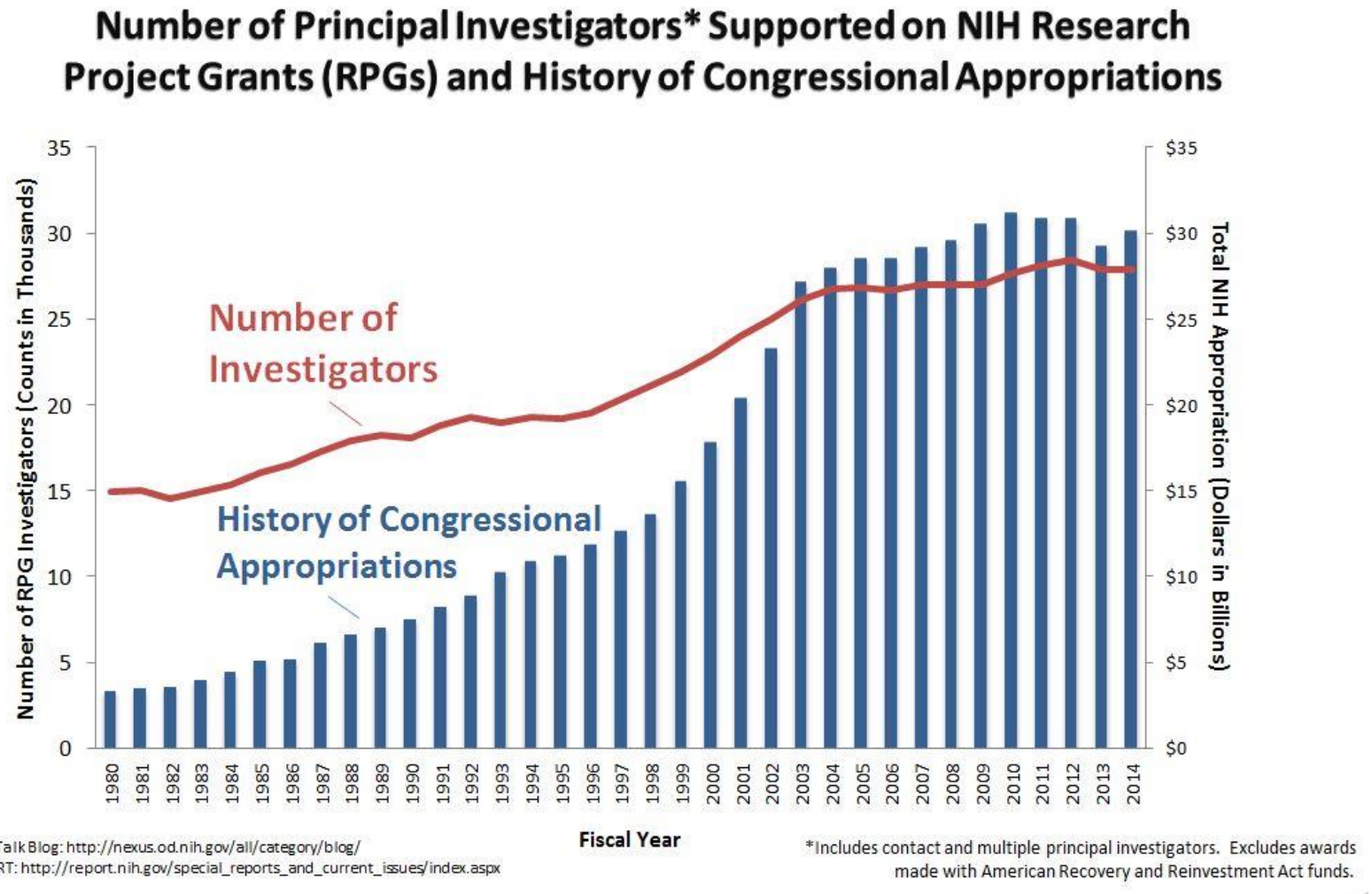
Line chart: different type of lines



Line chart (area)



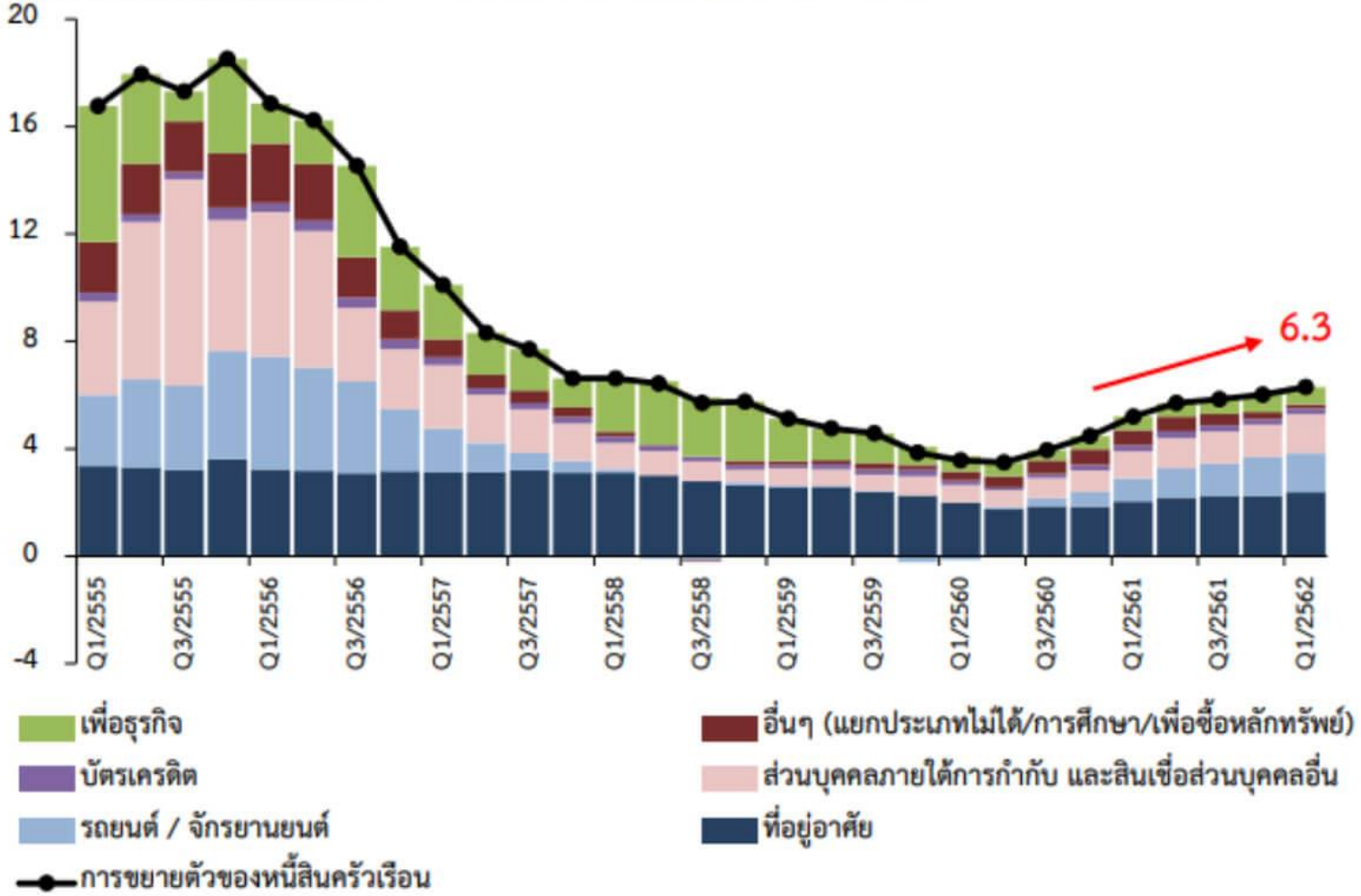
Bar chart + Line chart



https://nexus.od.nih.gov/all/2015/06/29/what-are-the-chances-of-getting-funded/

แหล่งที่มาของการขยายตัวหนี้ครัวเรือนแยกตามวัตถุประสงค์

ภาคครัวเรือนมีสัญญาณกลับมาเร่งก่อหนี้มากขึ้นในทุกหมวดสินเชื่อ โดยเฉพาะสินเชื่อหมวดรถยนต์



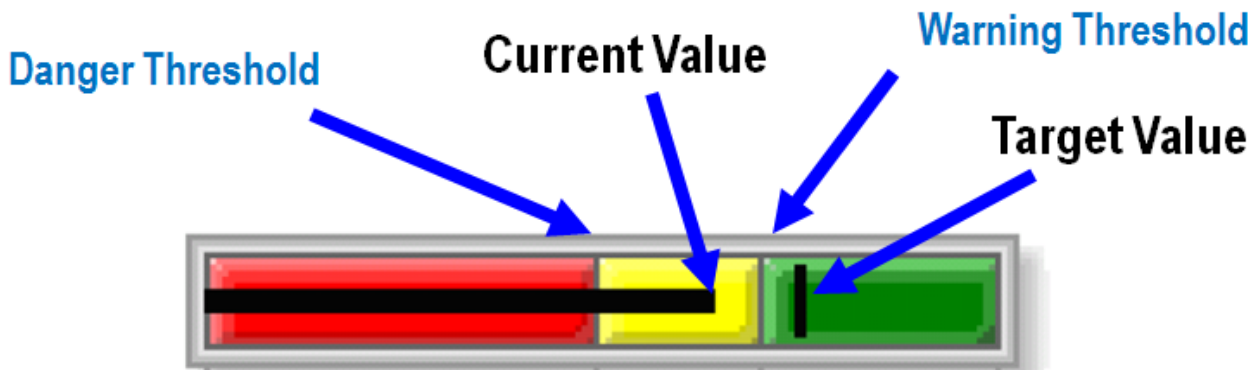
Bullet chart

When to use

- Evaluating performance of a metric against a goal

Possible extension

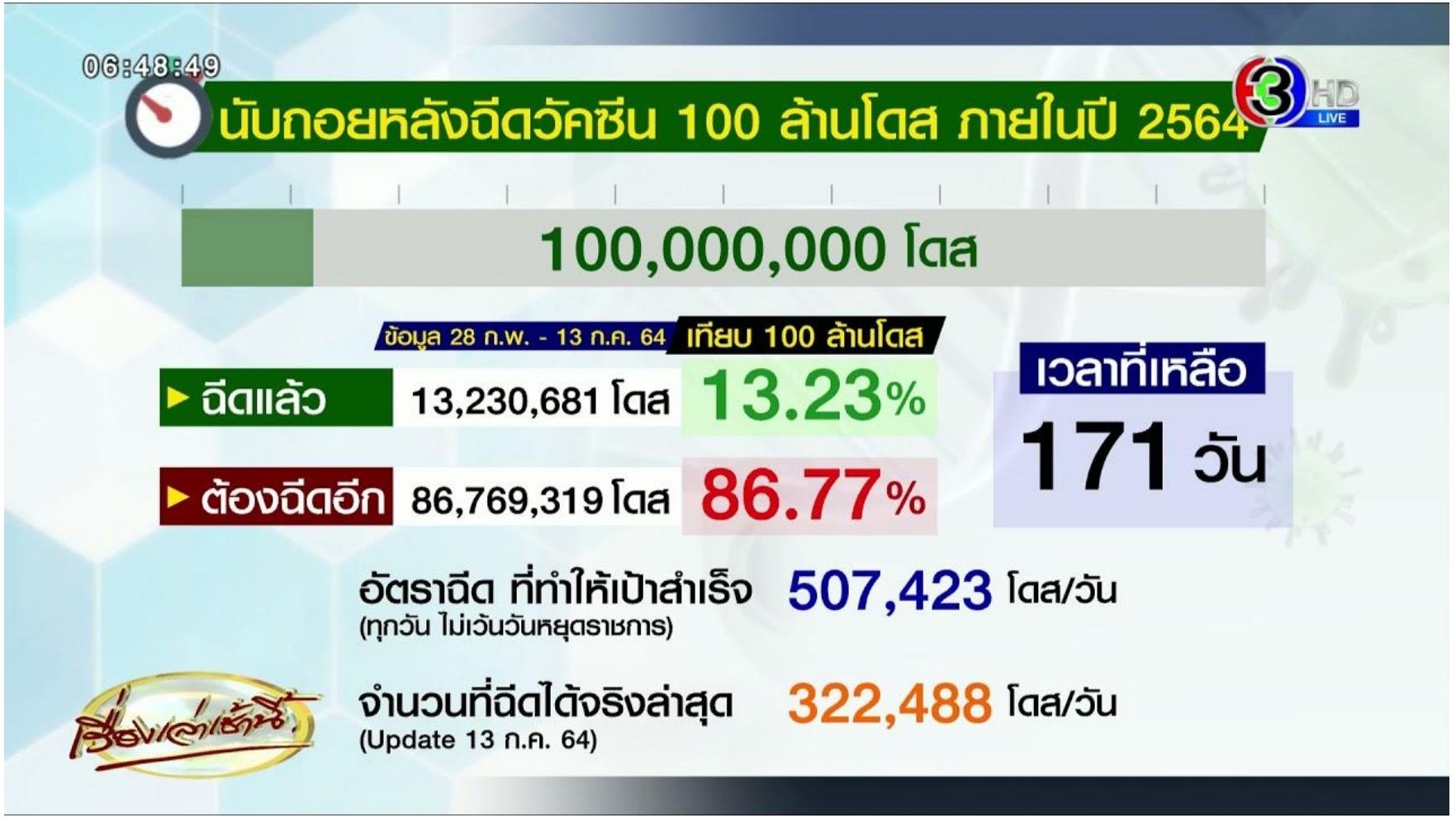
- Use color to illustrate achievement thresholds

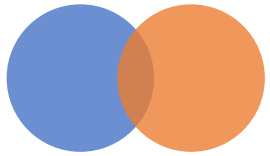


Bullet chart

Have you hit your target?

Bullet chart





2.Relationship

Scatter plot

Map

Bubble chart

Heat map

Crosstab / Highlight table

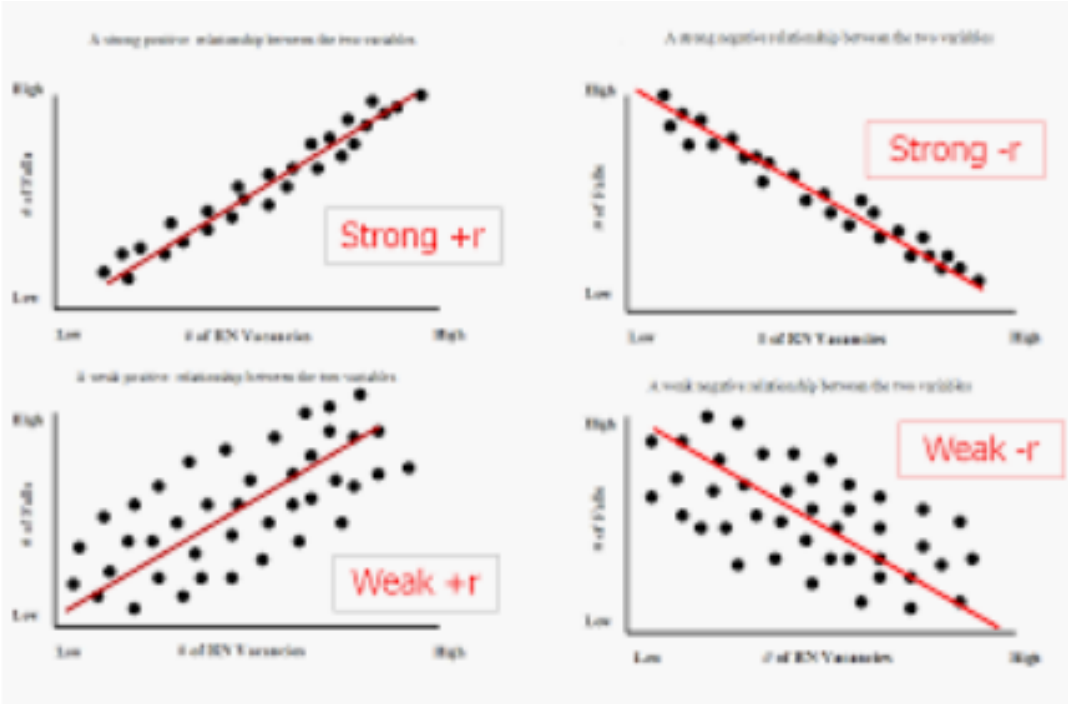
Scatter plot

When to use

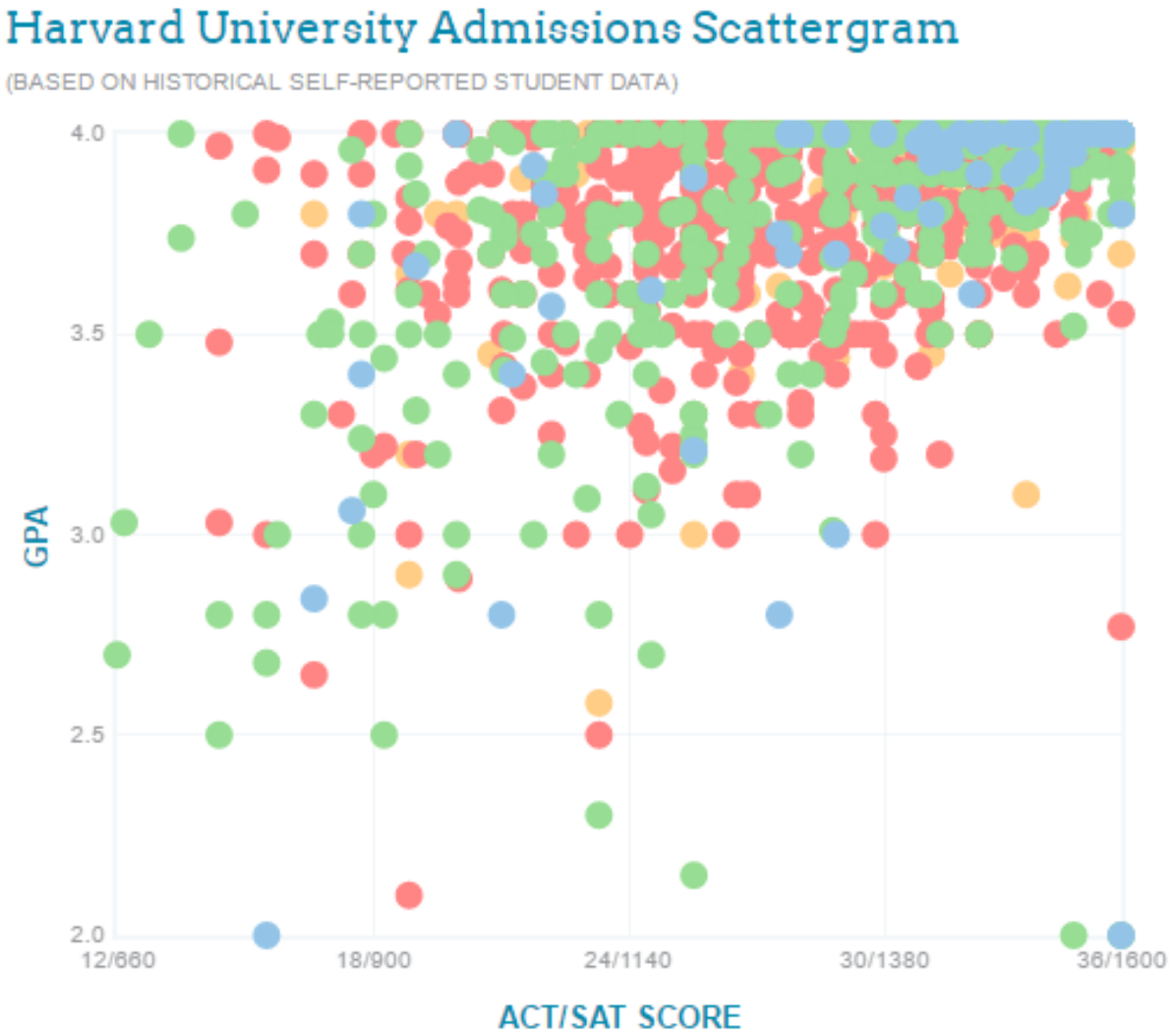
- Investigating the relationship between different variables

Possible extension

- Incorporate filters
- Use informative mark types

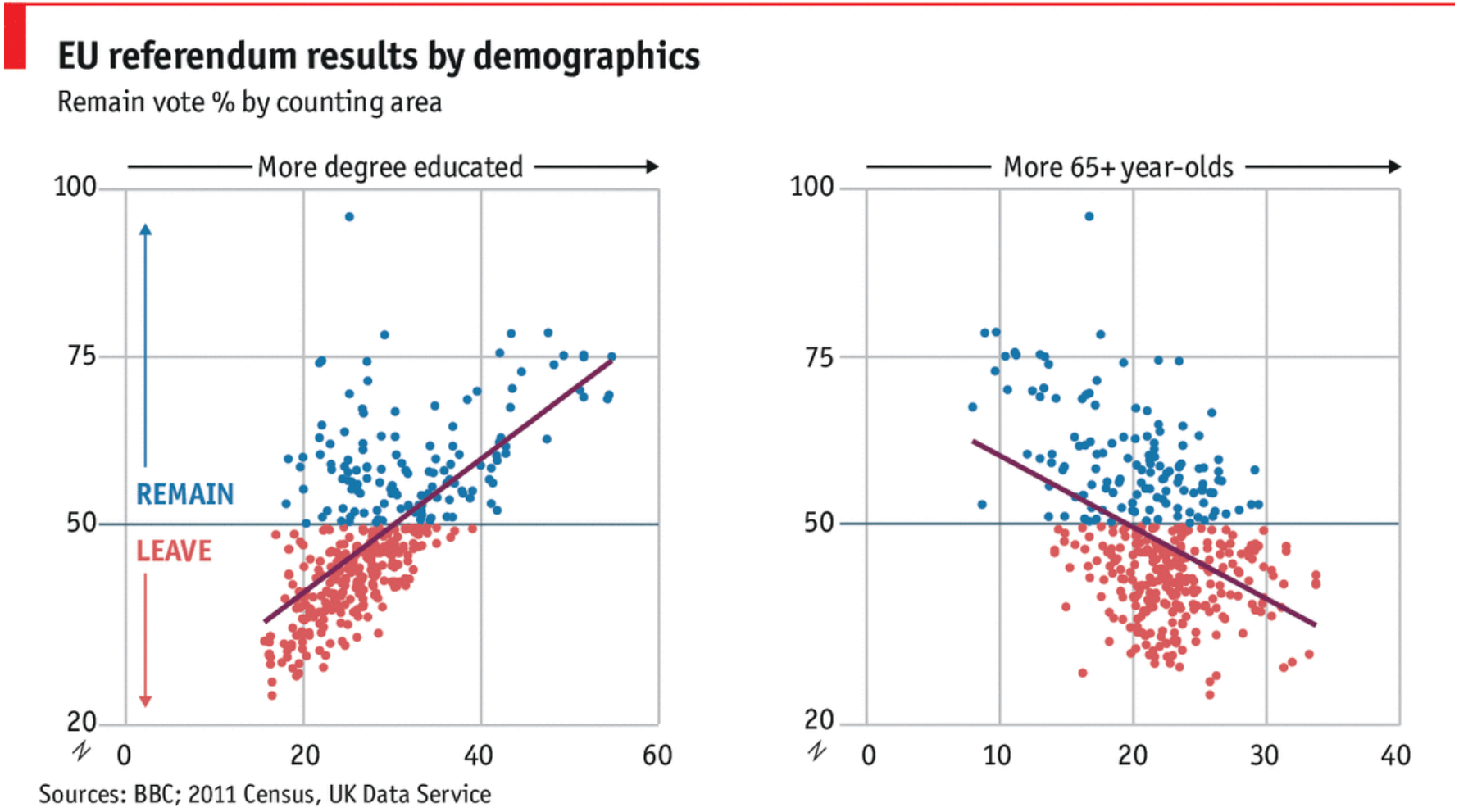


Scatter plot

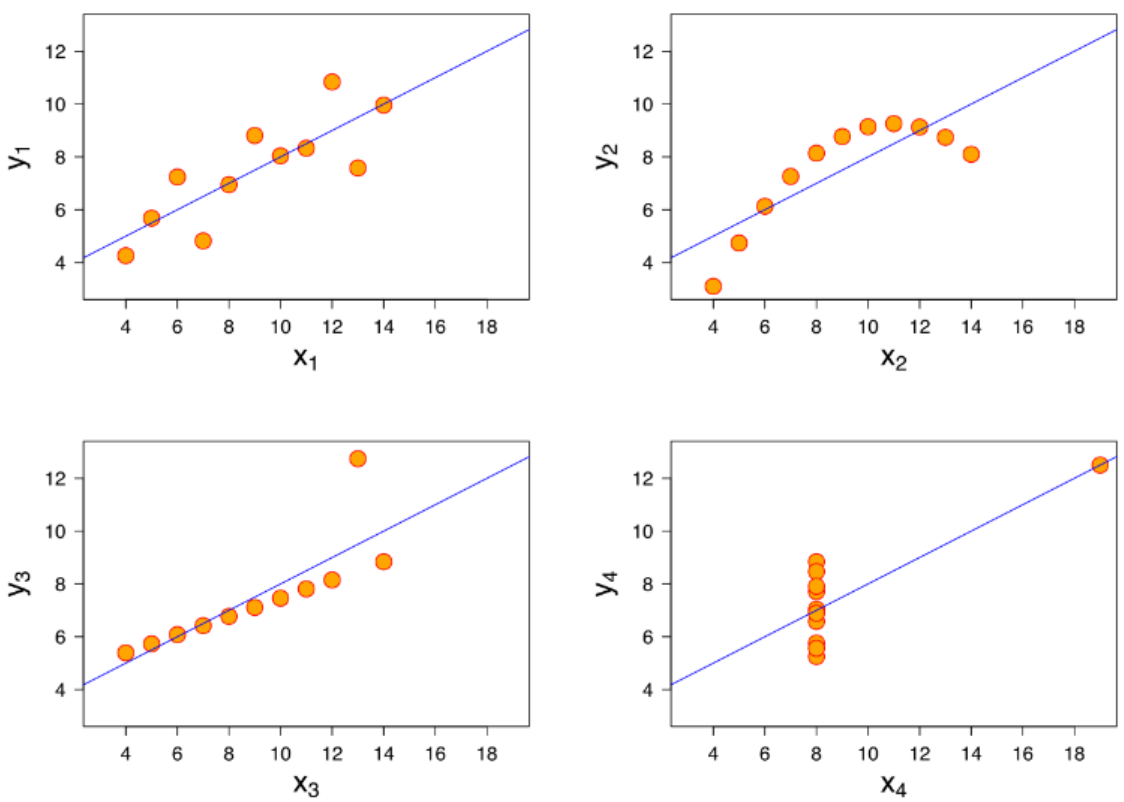


- **Green** - a successful applicant that chose to attend Harvard,
- **Blue** - an admitted applicant that chose not to attend,
- **Red** - a rejected applicant.

Scatter plot



Scatter plot: Anscombe's quartet



- All 4 sets of data shares many common statistics properties: means, variance, correlations, regression line, etc.
- the importance of graphing data.
- the effect of outliers and other influential observations on statistical properties.

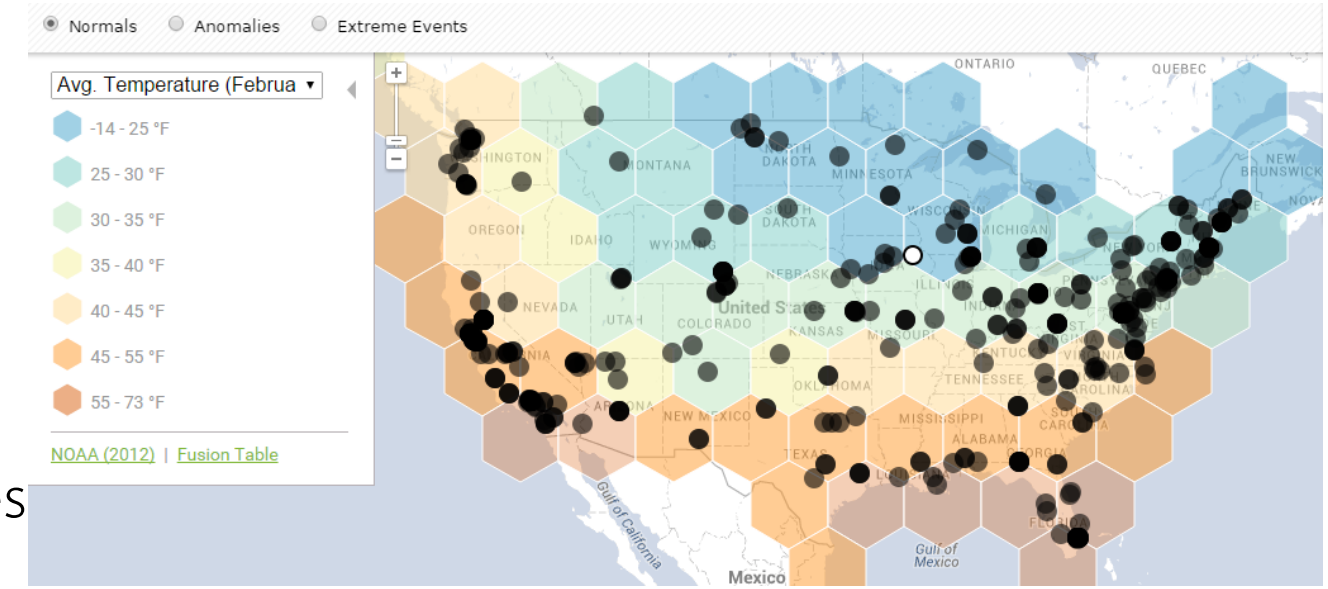
Map + Scatter plot

When to use

- Showing geocoded data

Possible extension

- Use maps as a filter for other types of charts, graphs, and tables



Climate Change news report

https://gisgeography.com/climate-change-effects-maps/

Bubble chart

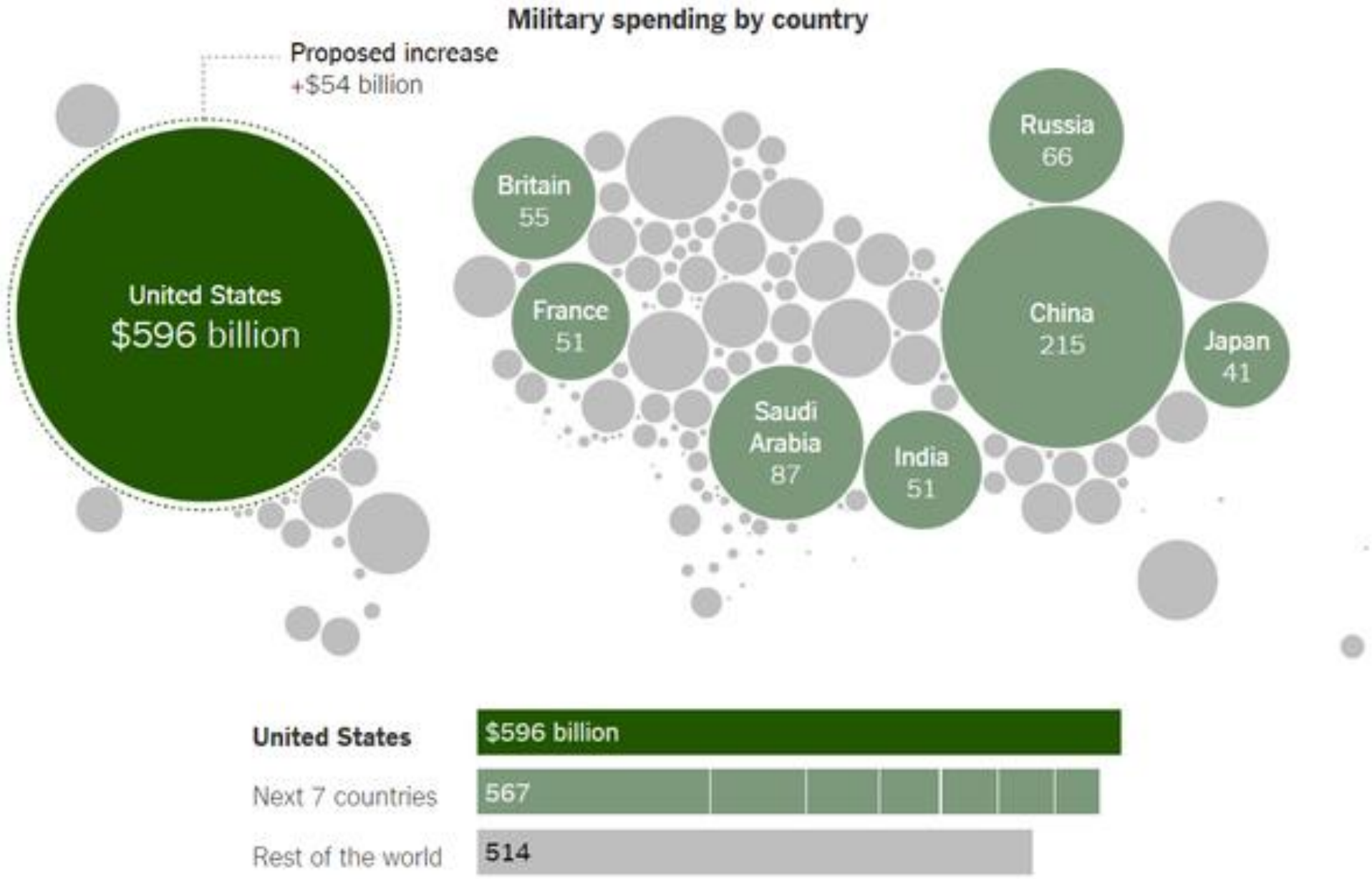
When to use

- Showing the concentration of data along two axes

Possible extension

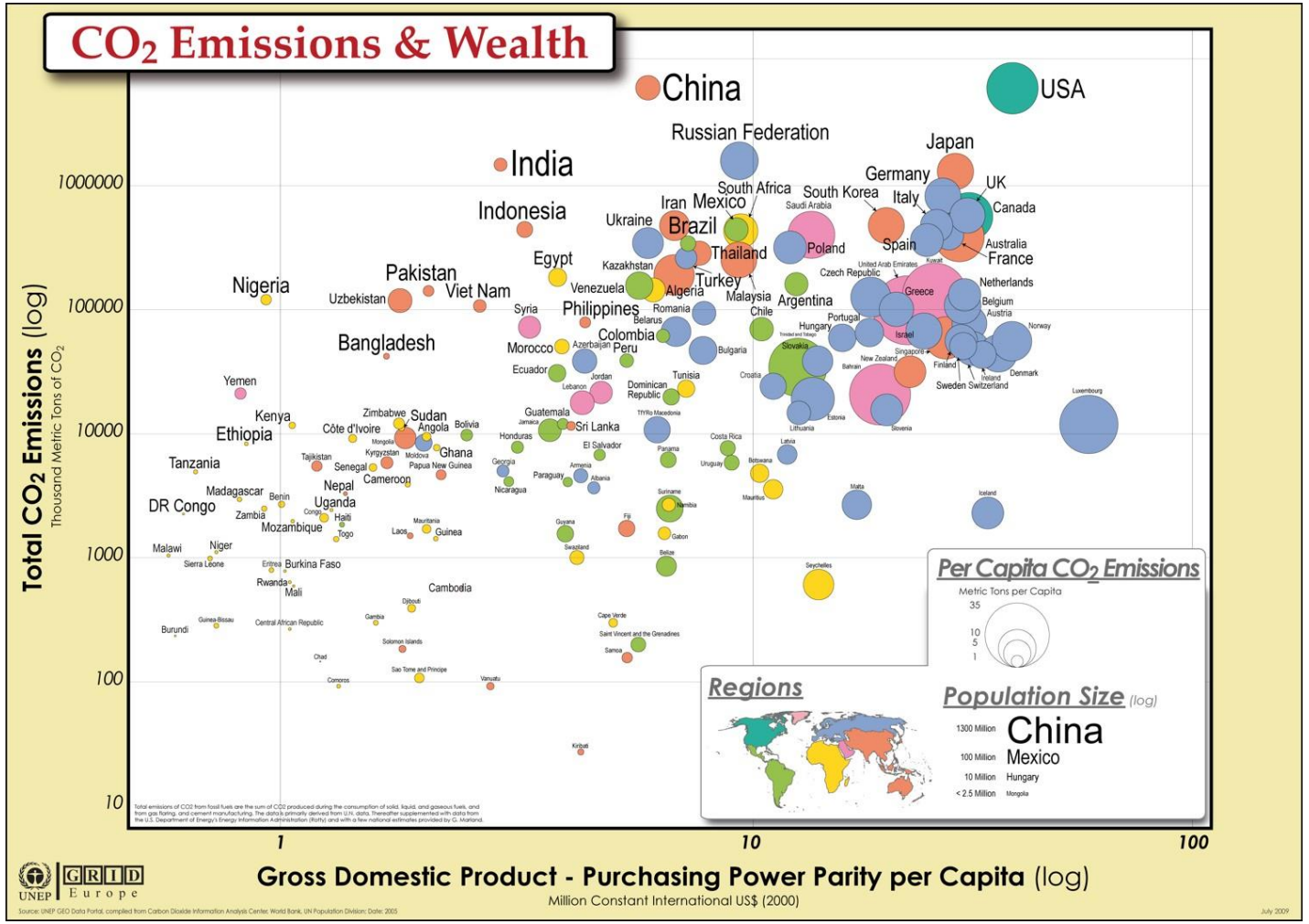
- Accentuate data on scatter plots
- Overlay on maps

http://genre.web.unc.edu/genres-in-the-disciplines/genres-in-science/

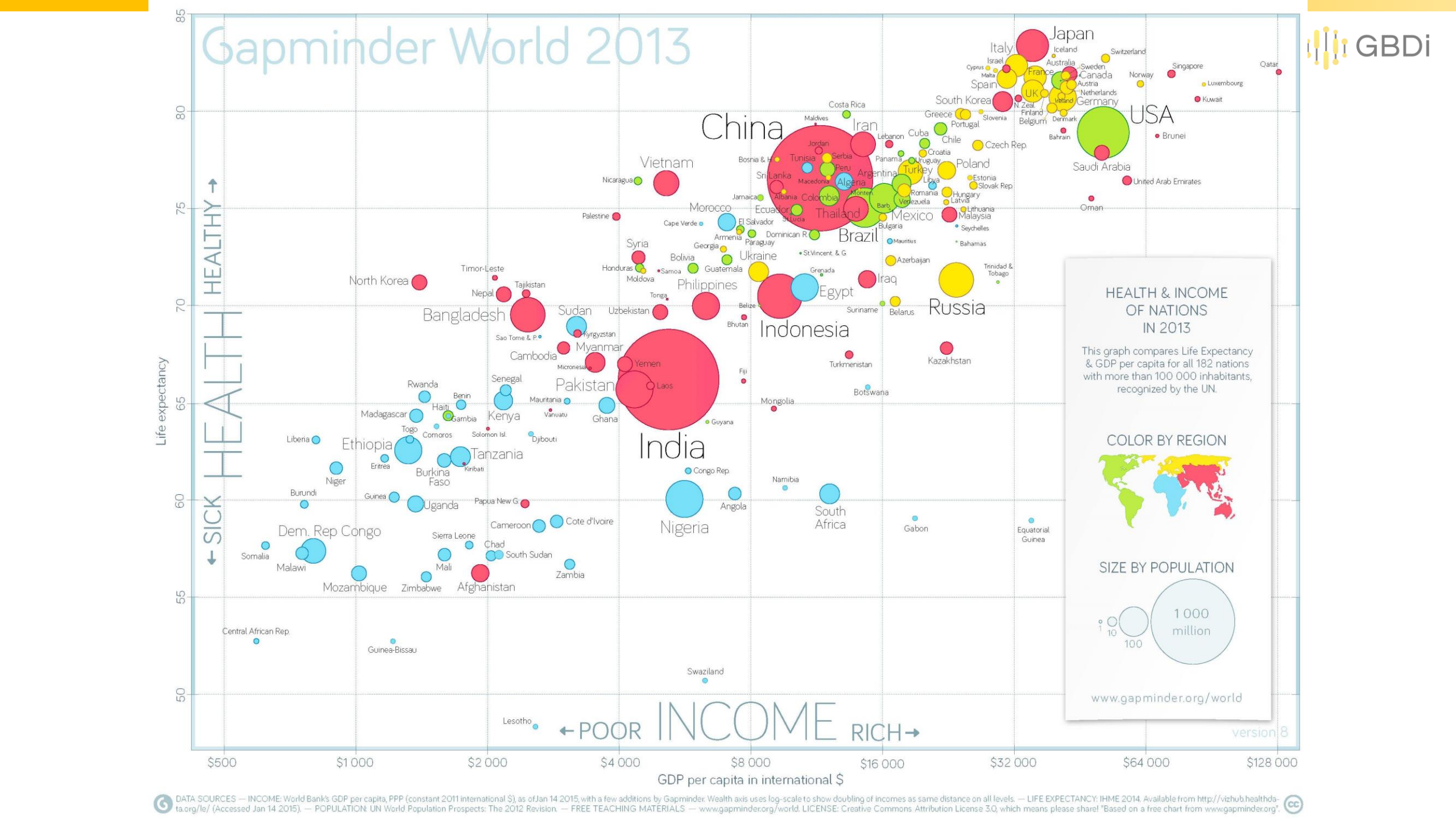


Source: NY Times

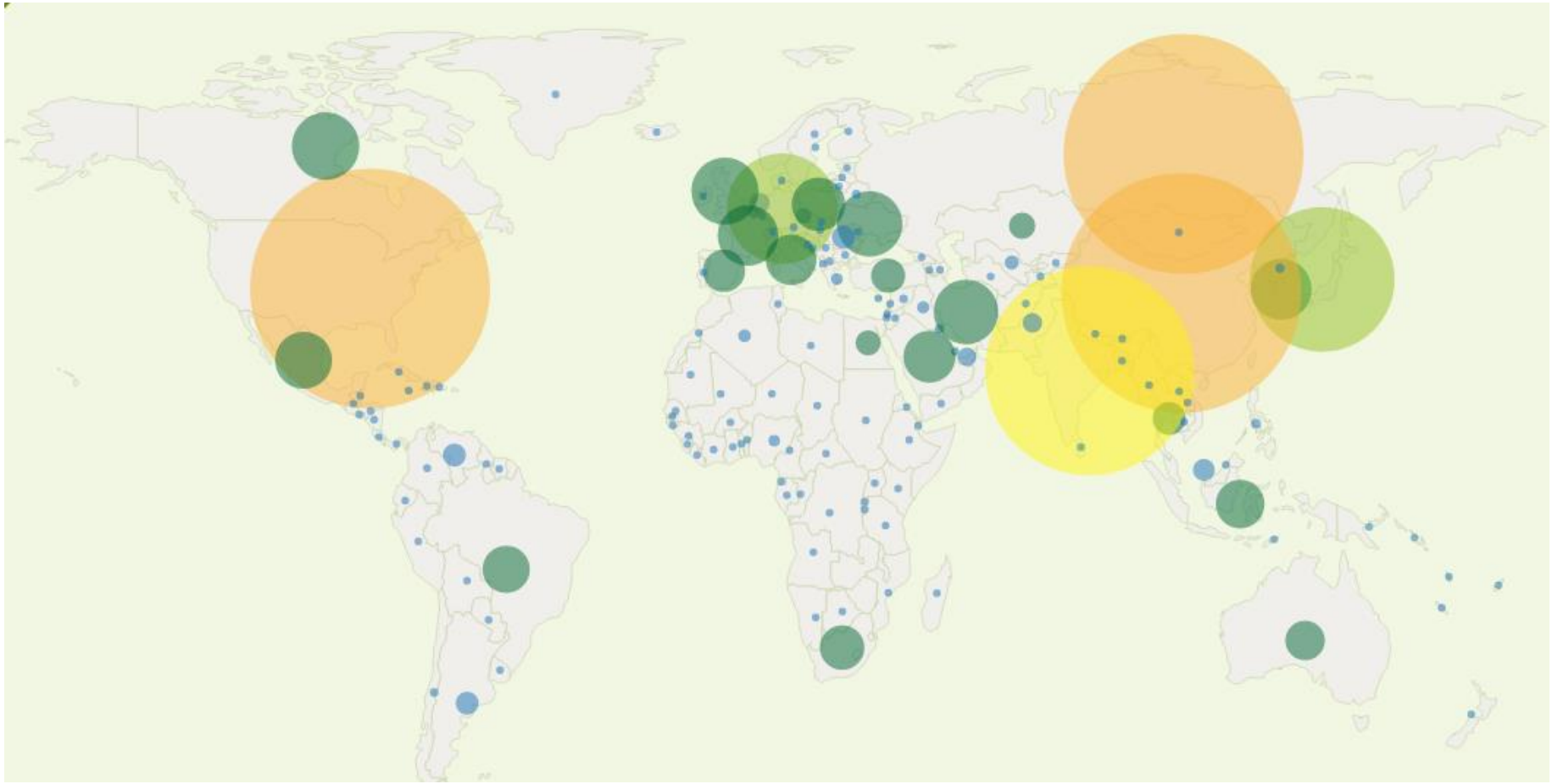
Bubble chart + Scatter plot



https://visual.ly/community/infographic/economy/co2-emissions-gdp-and-population-bubble-chartsc



Bubble chart + Map

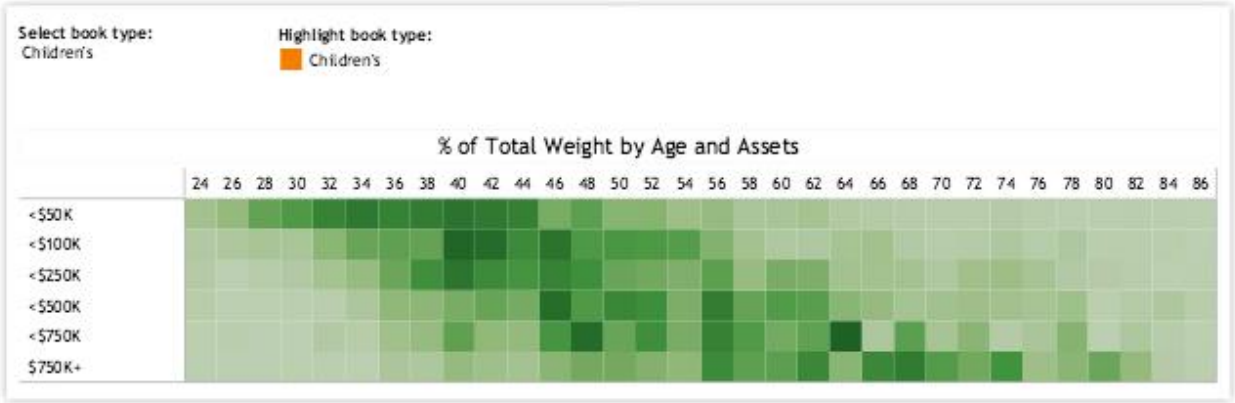


https://gisgeography.com/climate-change-effects-maps/

Heatmap

When to use

- Showing the relationship between two factors



Possible extension

- Vary the size of squares
- Using something other than squares

Crosstab / Highlight table

When to use

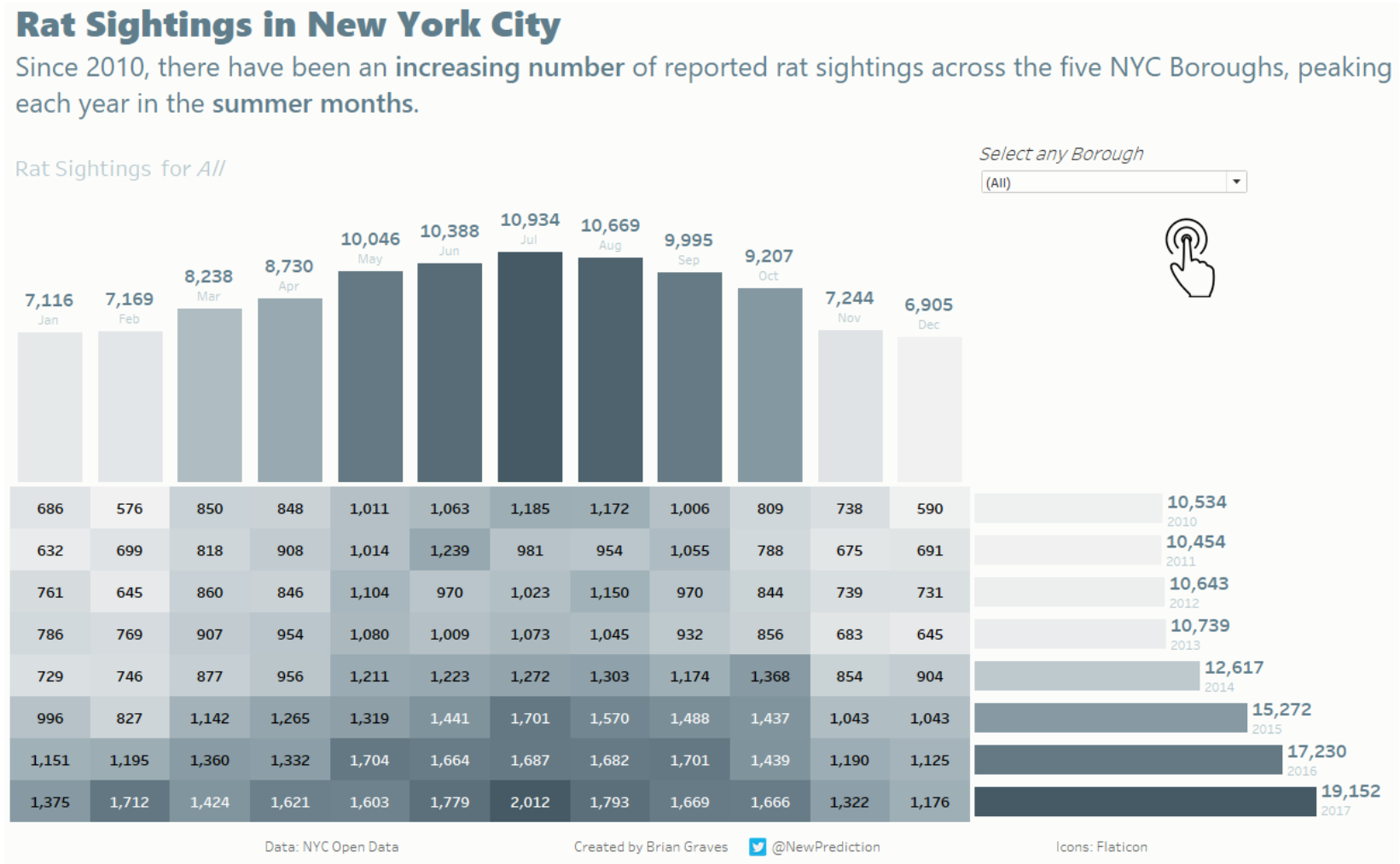
- Providing detailed information on heat maps

Possible extension

- Combine highlight tables with other chart types

ปีงบประมาณ	January	February	March	April	May	June	July	August	September	October	November	December
2549		1,180,500	305,000	7,500		13,500	10,000	25,000	150,200			
2550	10,500	116,500	1,000		5,500	52,500	1,000	1,500	28,500	2,000	32,500	230,070
2551	1,108,500	411,437	867,134	15,500	101,000	18,000	114,840	53,000	223,500	284,817	640,500	3,946,100
2552	221,500	621,752	59,200	64,000	28,400	6,105,500	86,000	1,087,000	290,500	64,500	95,000	224,899
2553	1,574,100	1,053,856	6,107,843	1,013,515	238,037	1,121,978	87,000	1,635,676	1,514,300	1,546,500	3,145,400	1,078,100
2554	912,500	1,411,624	512,853	465,641	101,500	596,500	1,694,705	203,840	725,833	16,543,300	5,955,120	10,364,730
2555	858,820	924,500	876,300	806,000	724,150	280,550	851,050	404,750	1,134,839	424,133	1,616,928	4,138,347
2556	2,203,699	1,272,799	565,800	905,400	547,651	1,220,000	321,000	982,409	946,698	31,250	2,938,380	1,138,300
2557	737,103	934,225	986,295	2,999,760	958,704	1,411,319	1,291,100	2,613,600	2,169,467	1,267,900	2,050,543	16,888,200
2558	2,645,888	1,907,457	1,054,491	627,722	376,000	1,643,867	3,533,850	16,230,631	11,154,671	952,000	2,970,500	3,027,865
2559	2,797,940	2,864,000	14,684,125	2,900,500	483,635	2,590,946	2,442,828	5,484,750	2,236,651	1,334,602	1,635,100	12,888,969
2560	2,751,941	1,558,290	4,560,788	478,278	937,405	3,218,400	1,336,045	3,019,639	2,106,749	3,051,750	3,466,650	4,965,579
2561	418,440	1,634,557	6,898,784	3,238,000	458,700	1,862,243	900,900	2,300,950	3,926,657	2,511,812	2,619,869	5,045,770
2562	1,163,000	1,736,500	2,510,200	1,718,308	1,433,568	578,959				5,264,100	6,816,792	5,958,382

Crosstab + Bar chart



3.Composition

Pie chart

Tree map

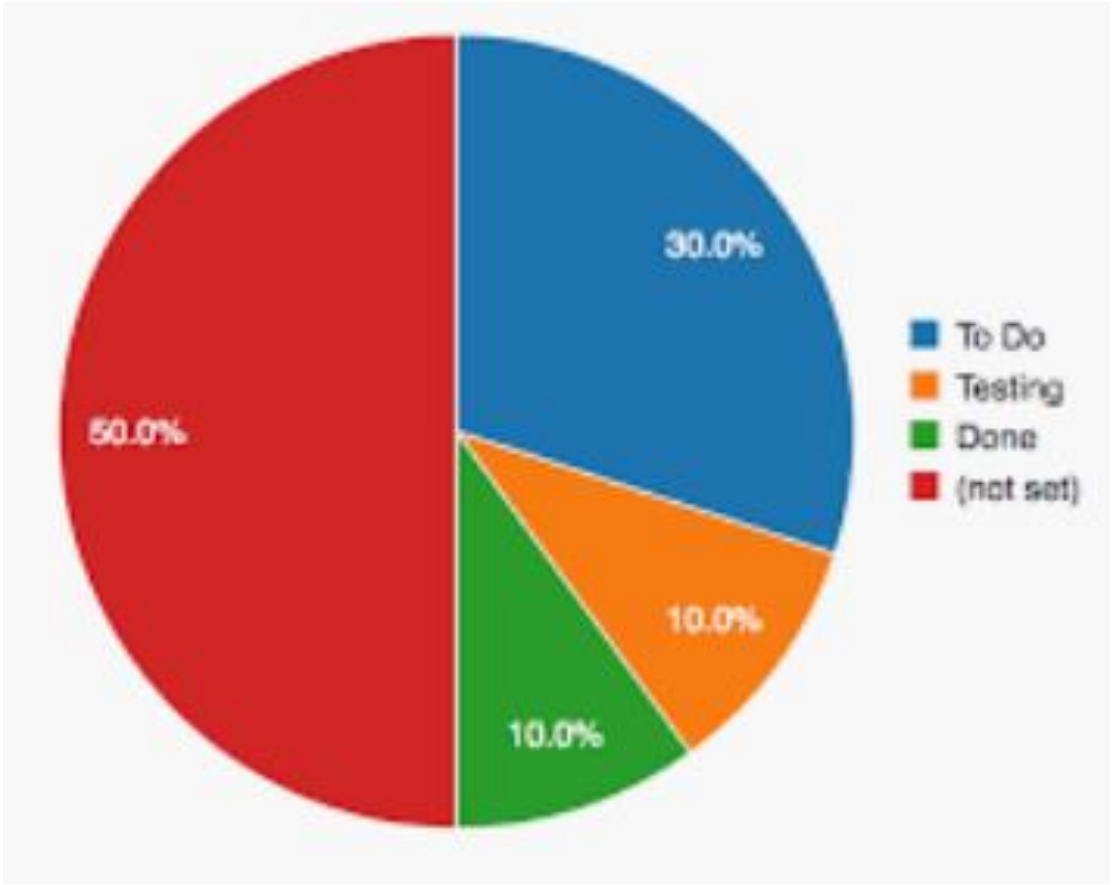
Pie chart

When to use

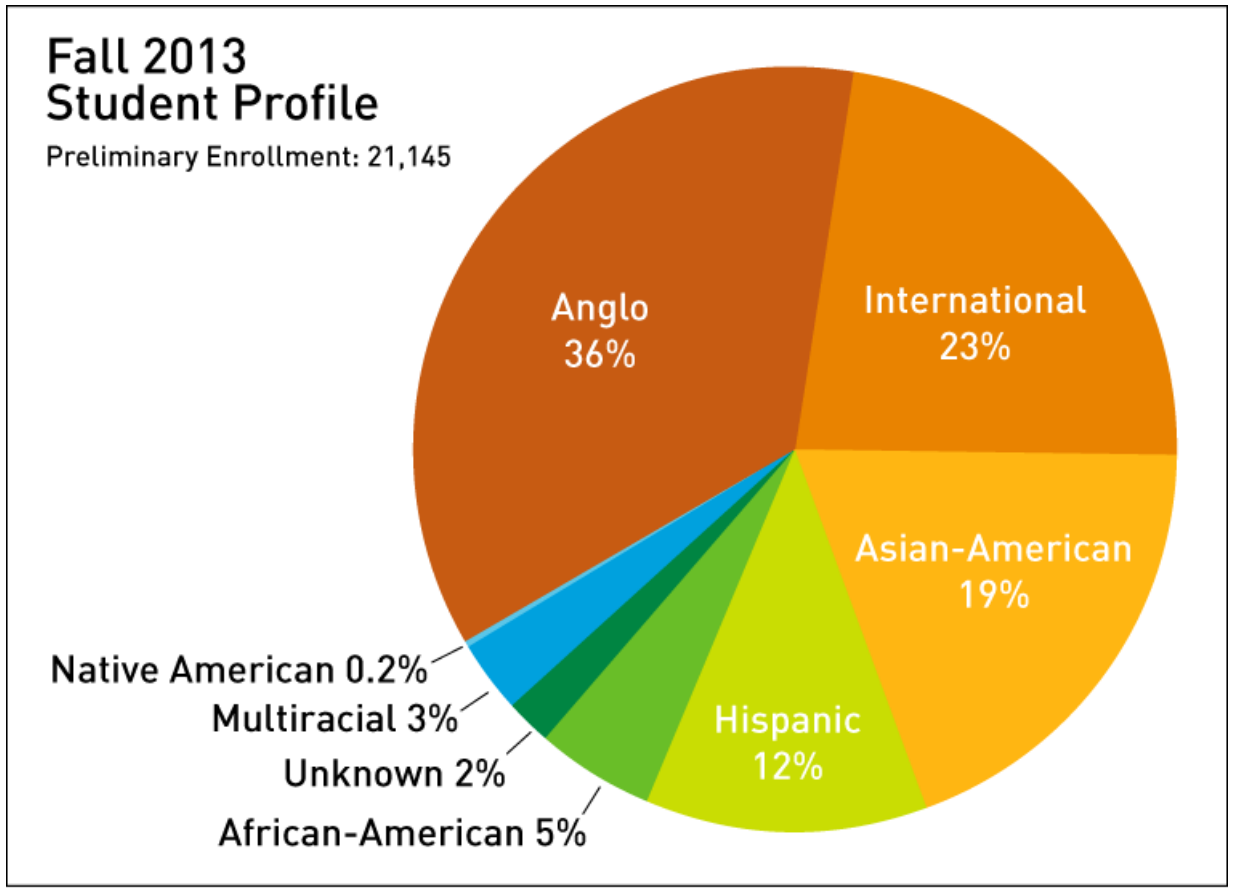
- Showing proportions

Possible extension

- Limit pie wedges to six
- Overlay pies on maps



Pie chart



- UT Dallas' diversity breakdown
- Total of 100.2 !

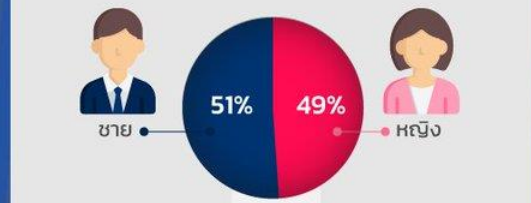
https://www.utdallas.edu/news/2013/9/25-26431_US-News-Ranks-UT-Dallas-as-One-of-Nations-Most-Div_story-wide.html

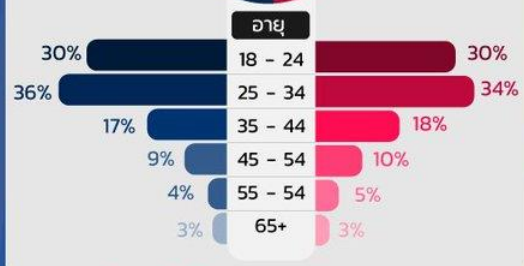
สถิติของประชากร facebook ในไทย

ลงทุนแมน

ประชากร Facebook ในไทยทั้งหมดมีประมาณ 50 ล้านบัญชี

เพศ



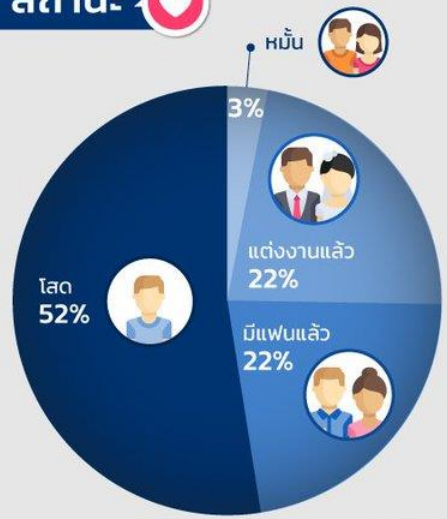


"กลุ่มคนวัยทำงาน" เล่น Facebook มากที่สุด

การศึกษา

ส่วนใหญ่จบการศึกษา "ปริญญาตรี"

สถานะ



ครึ่งหนึ่งของคนเล่น Facebook มีสถานะ "โสด"

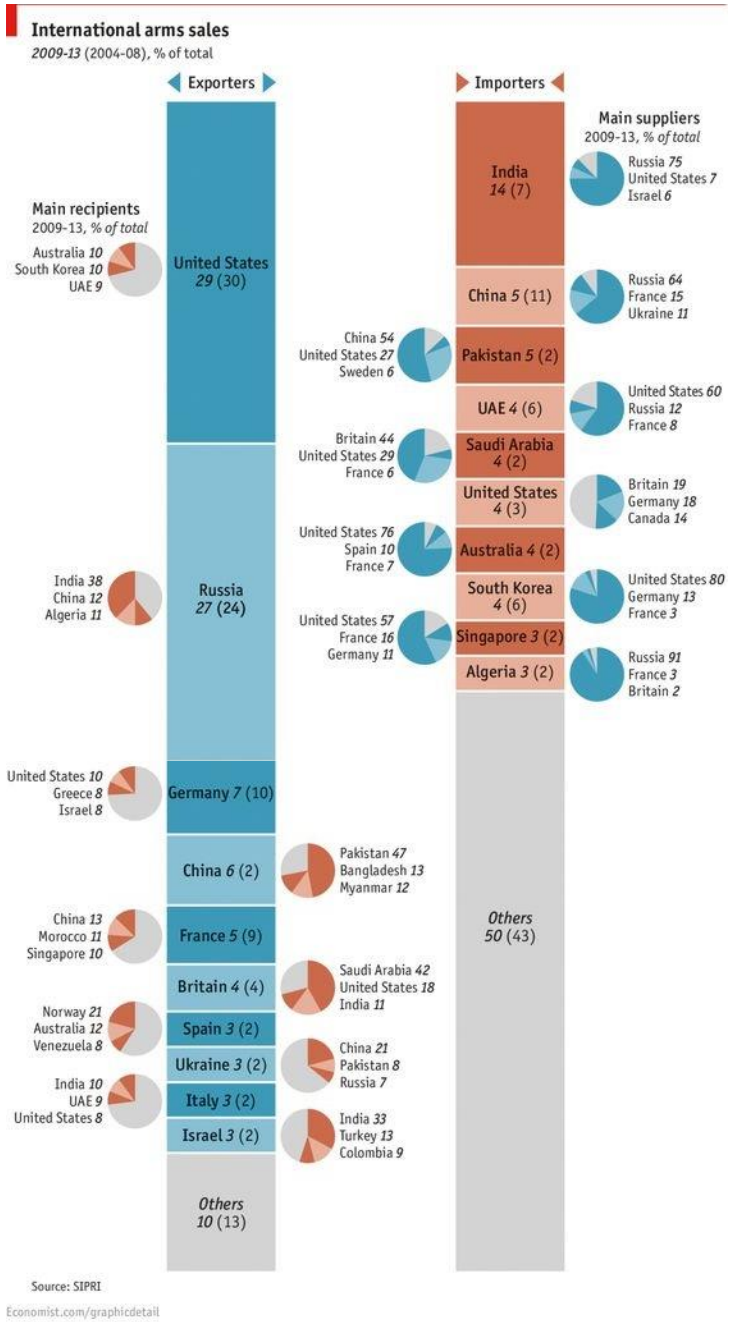
อาชีพ

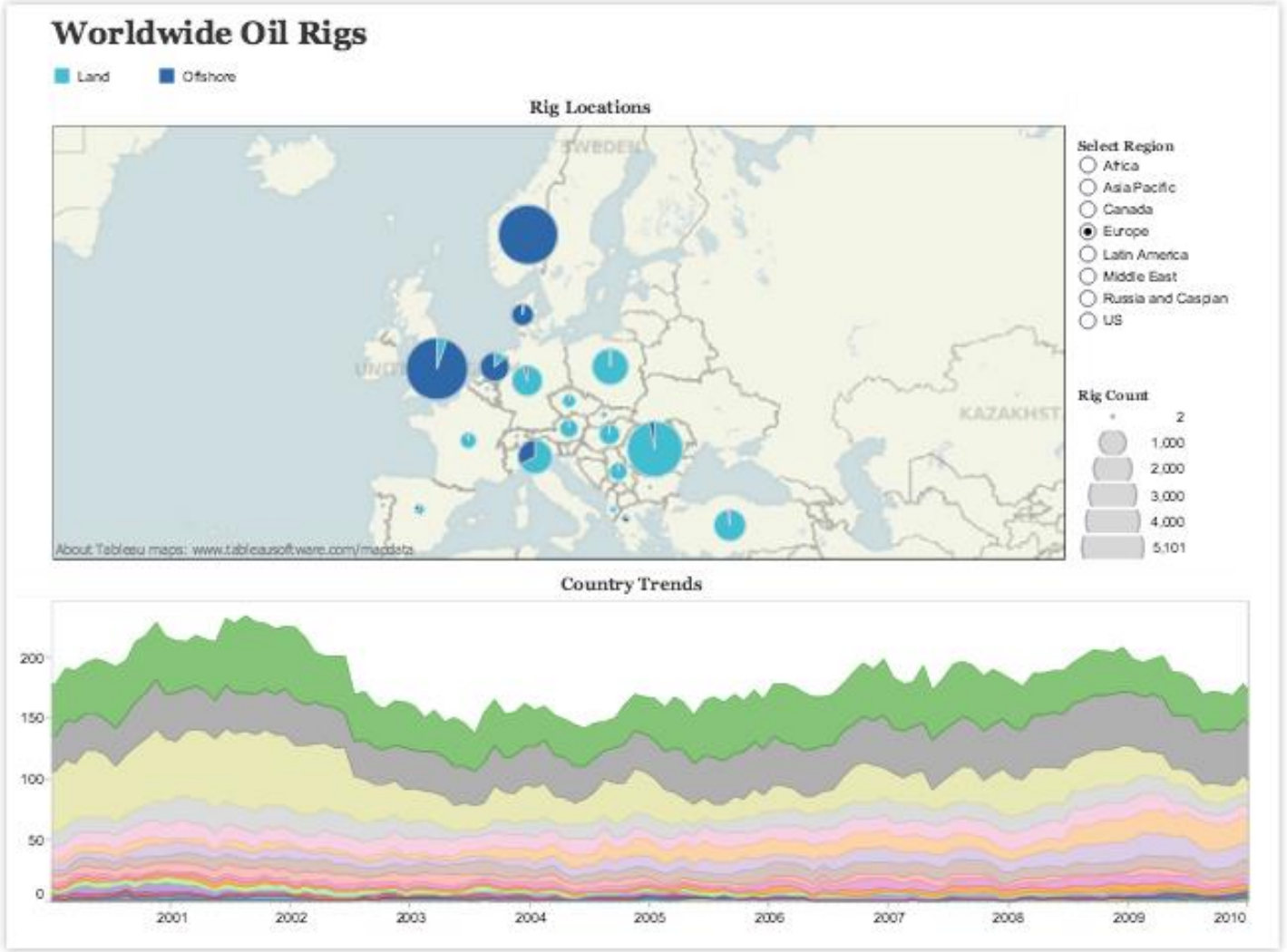
10 อันดับแรก

อาชีพยอดนิยม คือ "ผู้บริหาร"

หมายเหตุ : ข้อมูลจาก Facebook Audience Insights ณ วันที่ 26 ตุลาคม 2561

GBDi


International arms sales
2009-13 (2004-08), % of total
Exporters
Importers
Main recipients
2009-13, % of total
Main suppliers
2009-13, % of total
United States 29 (30)
Russia 27 (24)
Germany 7 (10)
China 6 (2)
France 5 (9)
Britain 4 (4)
Spain 3 (2)
Ukraine 3 (2)
Italy 3 (2)
Israel 3 (2)
Others 10 (13)
India 14 (7)
China 5 (11)
Pakistan 5 (2)
UAE 4 (6)
Saudi Arabia 4 (2)
United States 4 (3)
Australia 4 (2)
South Korea 4 (6)
Singapore 3 (2)
Algeria 3 (2)
Others 50 (43)
Australia 10
South Korea 10
UAE 9
India 38
China 12
Algeria 11
United States 10
Greece 8
Israel 8
China 13
Morocco 11
Singapore 10
Norway 21
Australia 12
Venezuela 8
India 10
UAE 9
United States 8
China 54
United States 27
Sweden 6
Britain 44
United States 29
France 6
United States 76
Spain 10
France 7
United States 57
France 16
Germany 11
Pakistan 47
Bangladesh 13
Myanmar 12
Saudi Arabia 42
United States 18
India 11
China 21
Pakistan 8
Russia 7
India 33
Turkey 13
Colombia 9
Russia 75
United States 7
Israel 6
Russia 64
France 15
Ukraine 11
United States 60
Russia 12
France 8
Britain 19
Germany 18
Canada 14
United States 80
Germany 13
France 3
Russia 91
France 3
Britain 2
Source: SIPRI
Economist.com/graphicdetail


Worldwide Oil Rigs
Land
Offshore
Rig Locations
Select Region
Africa
Asia Pacific
Canada
Europe
Latin America
Middle East
Russia and Caspian
US
Rig Count
2
1,000
2,000
3,000
4,000
5,101
About Tableau maps: www.tableausoftware.com/mapdata
Country Trends
200
150
100
50
0
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010

Treemap

When to use

- Showing hierarchical data as a proportion of a whole

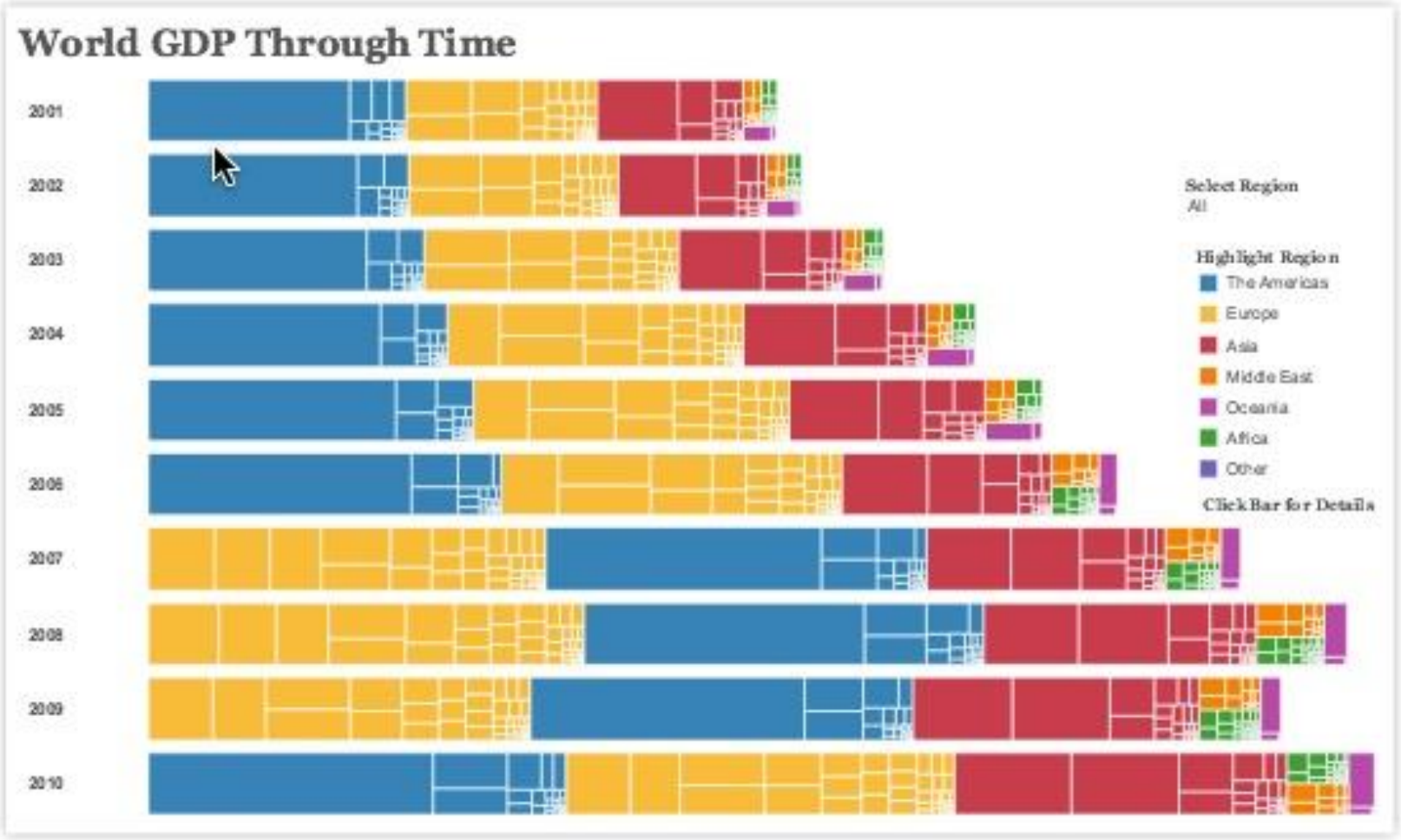
Possible extension

- Colouring the rectangles by a category
- Combining tree maps with bar charts

Treemap + Coloring

https://www.researchgate.net/figure/Treemap-visualizing-the-distribution-of-the-2969-publications-citing-Rittel-and-Webber_fig1_333183021

Treemap + Bar chart



4.Distribution

Histogram chart

Box plot

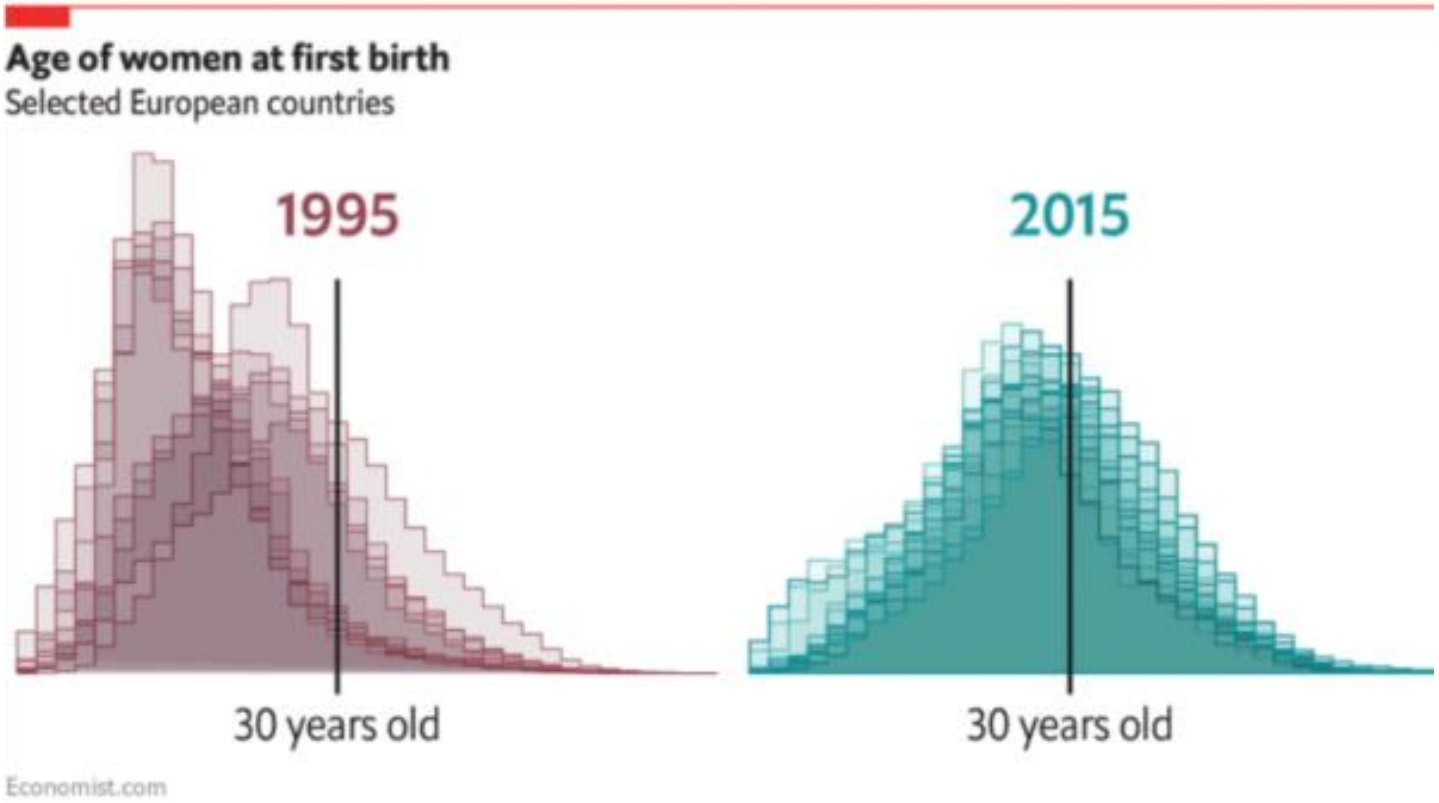
Histogram chart

When to use

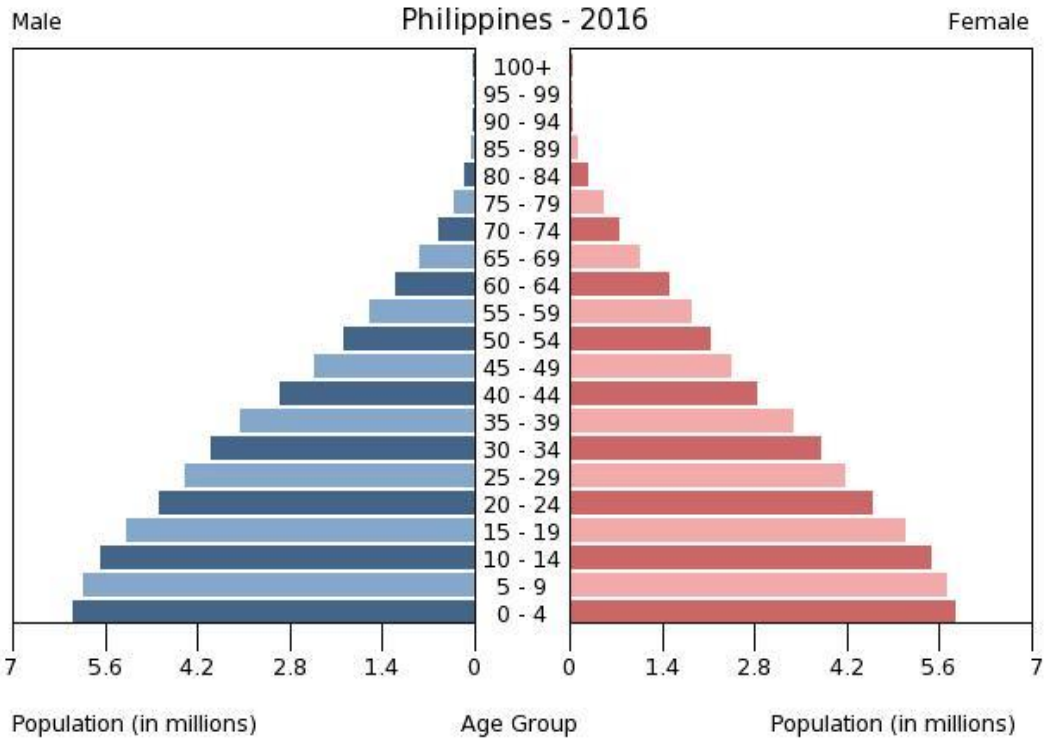
- Understanding the distribution of your data

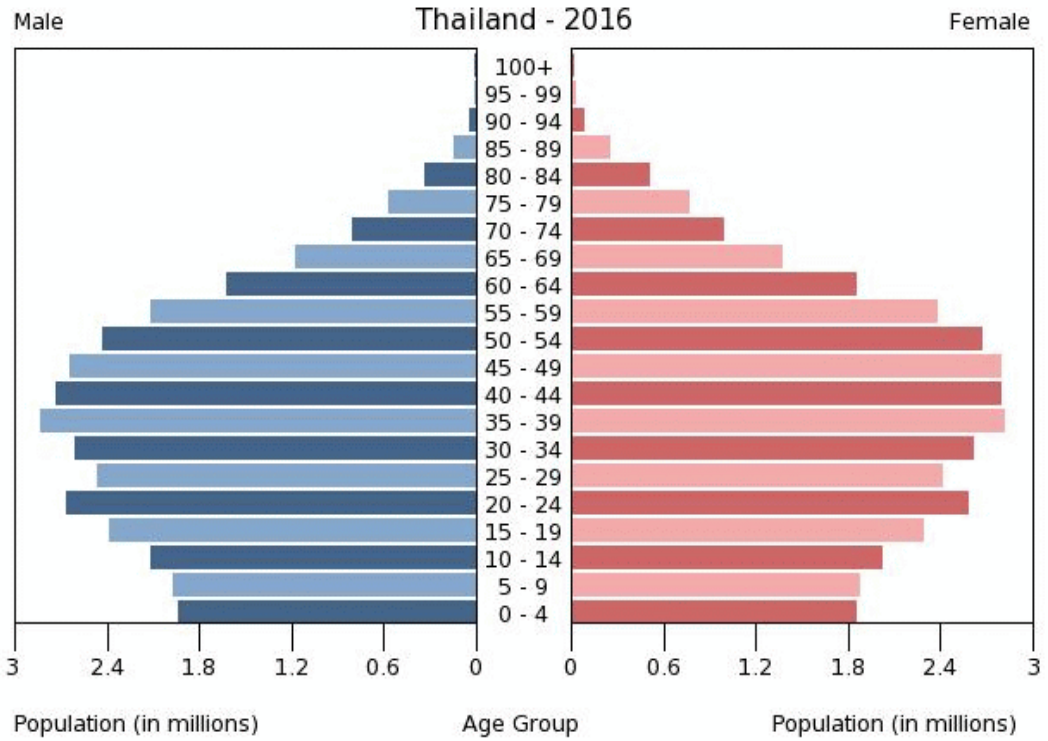
Possible extension

- Test different groupings of data
- Add a filter



Histogram chart





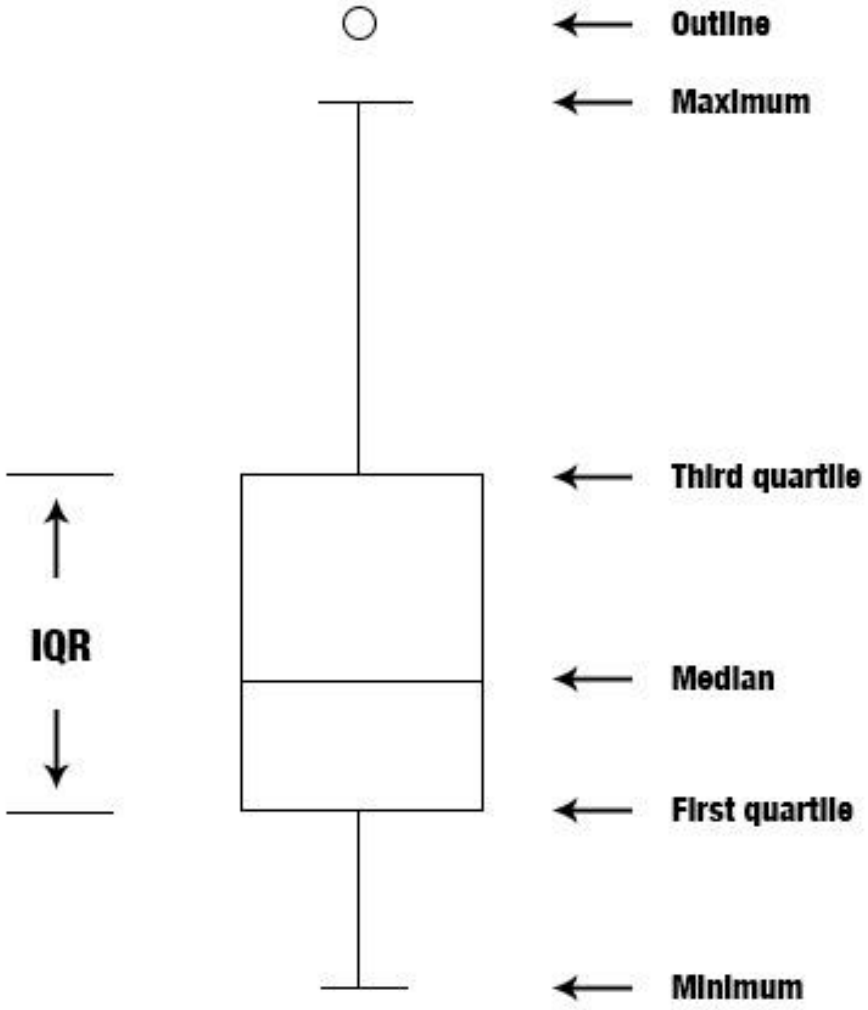
Box plot

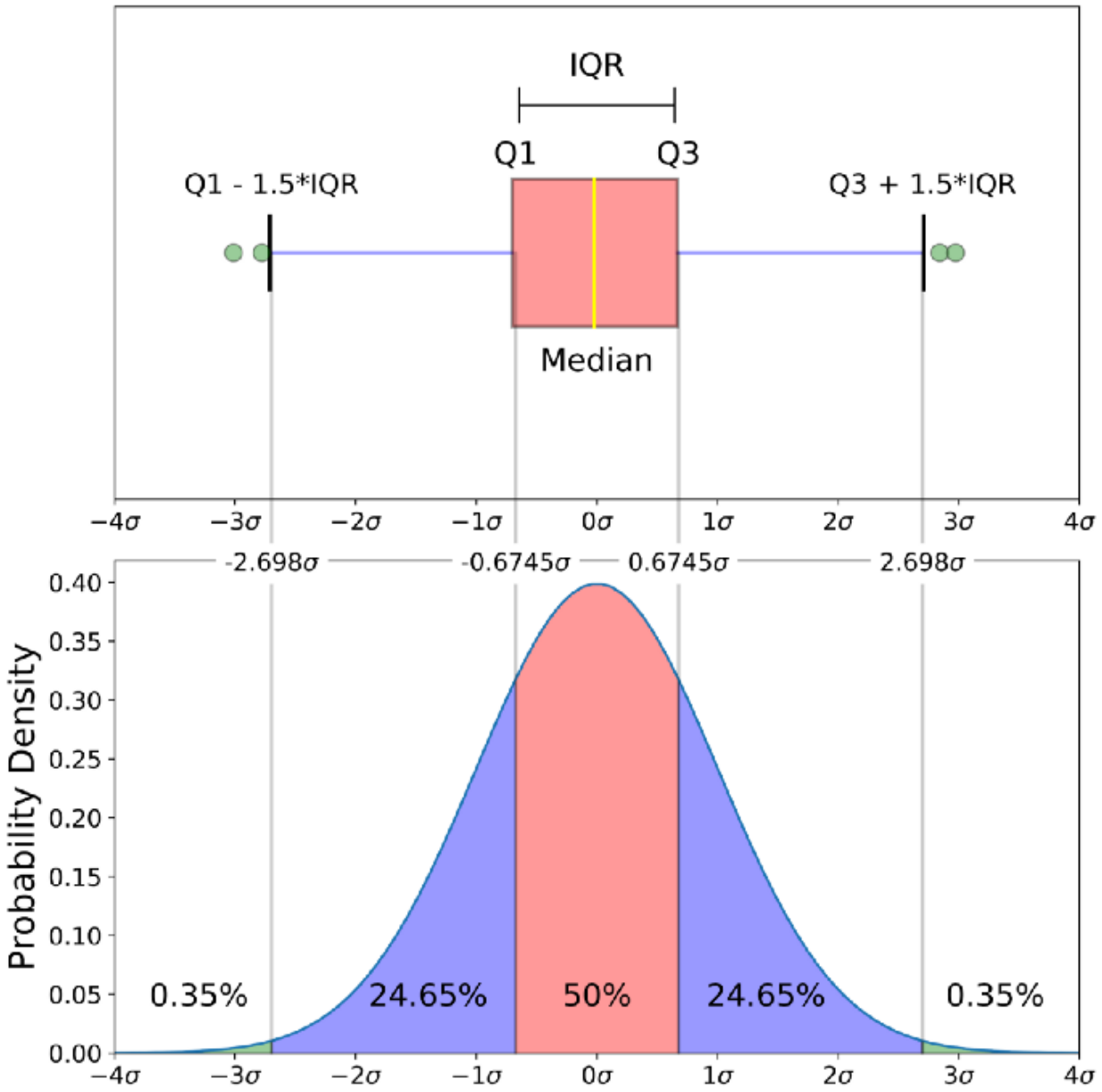
When to use

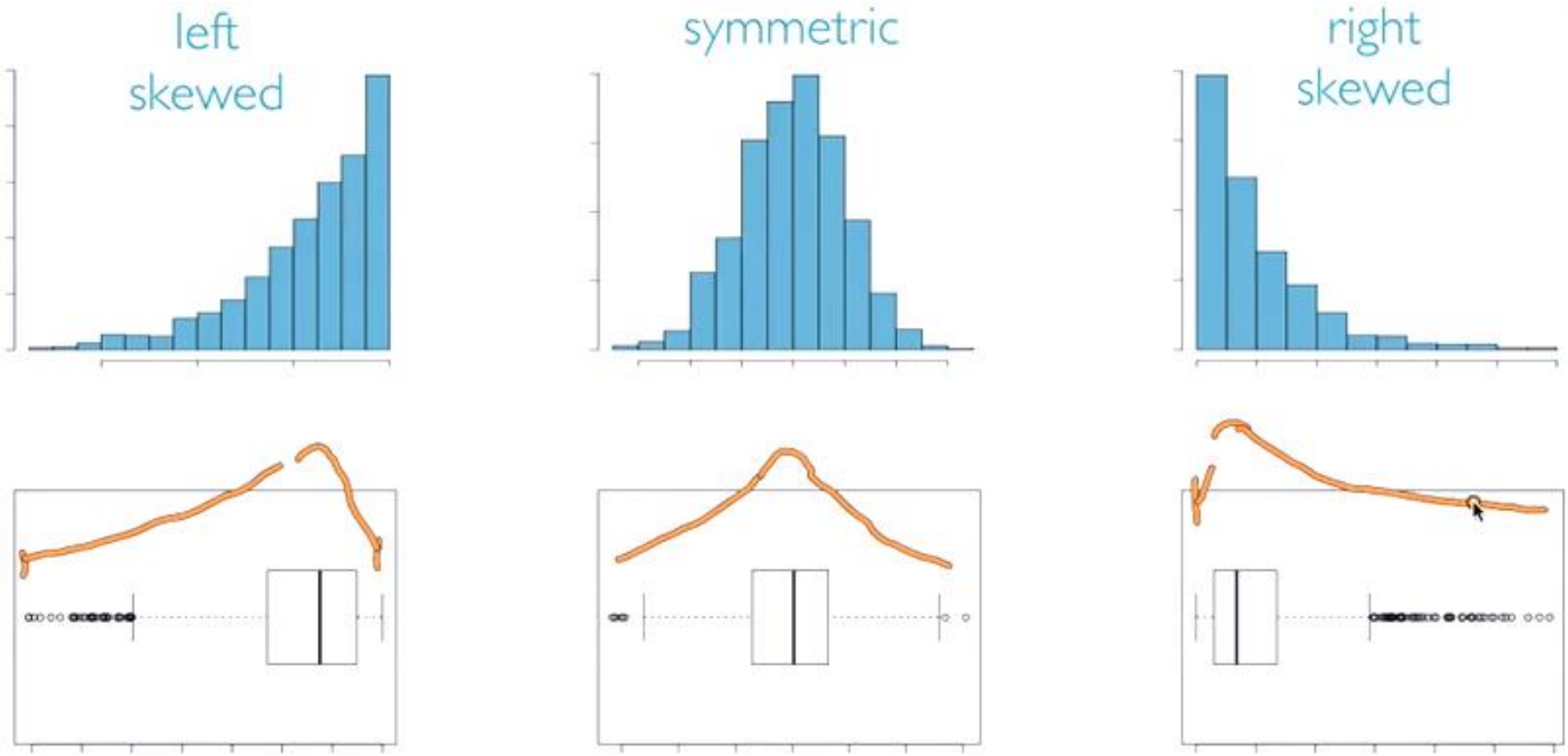
- Showing the distribution of a set of data

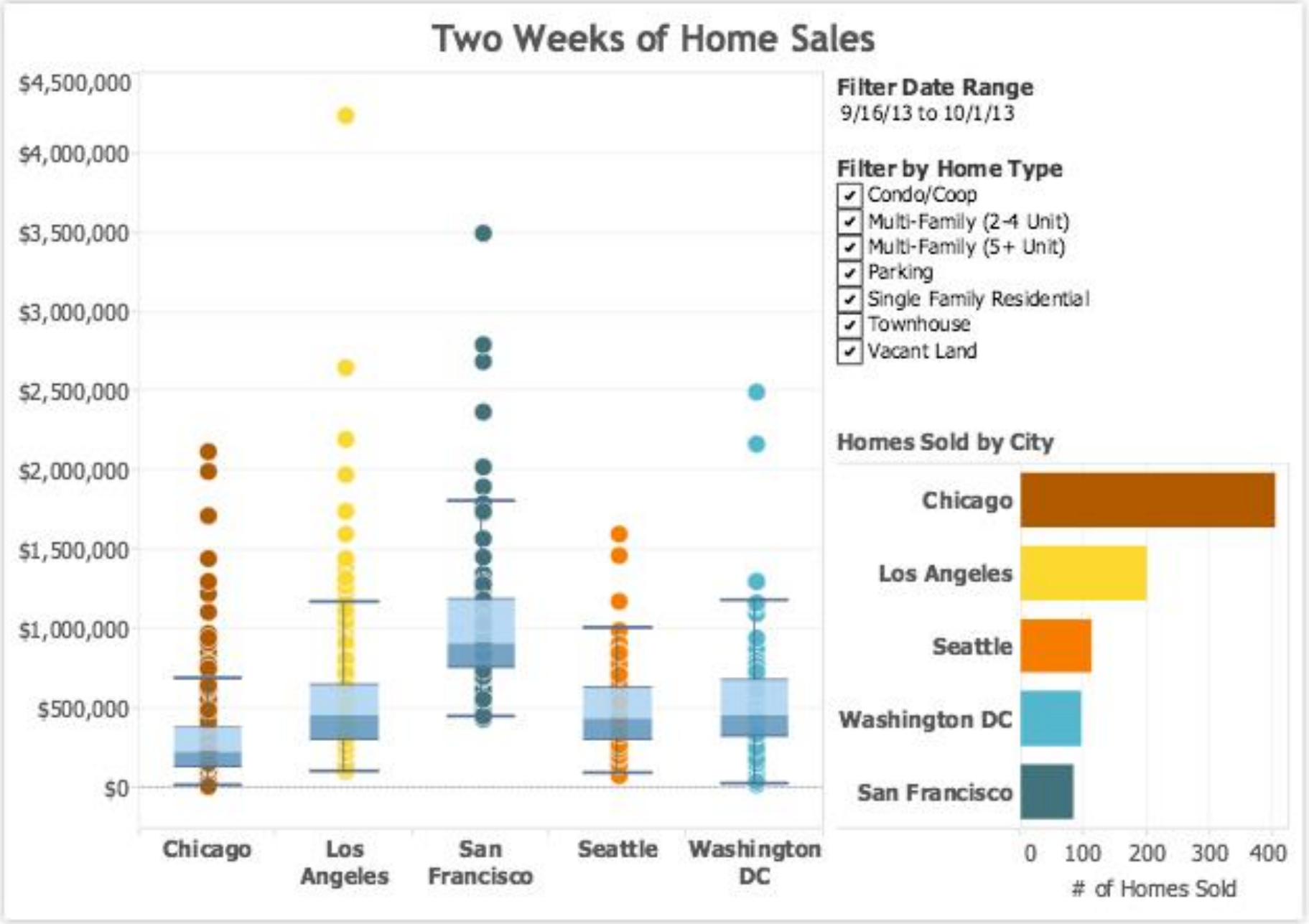
Possible extension

- Hiding the points within the box
- Comparing boxplots across categorical dimensions




IQR
Q1
Q3
Q1 - 1.5*IQR
Q3 + 1.5*IQR
Median
−4σ
−3σ
−2σ
−1σ
0σ
1σ
2σ
3σ
4σ
-2.698σ
-0.6745σ
0.6745σ
2.698σ
Probability Density
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0.35%
24.65%
50%
24.65%
0.35%


left
skewed
symmetric
right
skewed


Two Weeks of Home Sales
$4,500,000
$4,000,000
$3,500,000
$3,000,000
$2,500,000
$2,000,000
$1,500,000
$1,000,000
$500,000
$0
Chicago
Los Angeles
San Francisco
Seattle
Washington DC
Filter Date Range
9/16/13 to 10/1/13
Filter by Home Type
Condo/Coop
Multi-Family (2-4 Unit)
Multi-Family (5+ Unit)
Parking
Single Family Residential
Townhouse
Vacant Land
Homes Sold by City
Chicago
Los Angeles
Seattle
Washington DC
San Francisco
0
100
200
300
400
# of Homes Sold

What do you want to tell from your data?

Story

1.Comparison

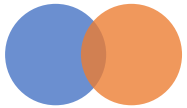
Bar chart, Line chart

Bullet chart

3.Composition

Pie chart

Tree map



2.Relationship

Scatter plot, Map

Bubble chart, Heat map

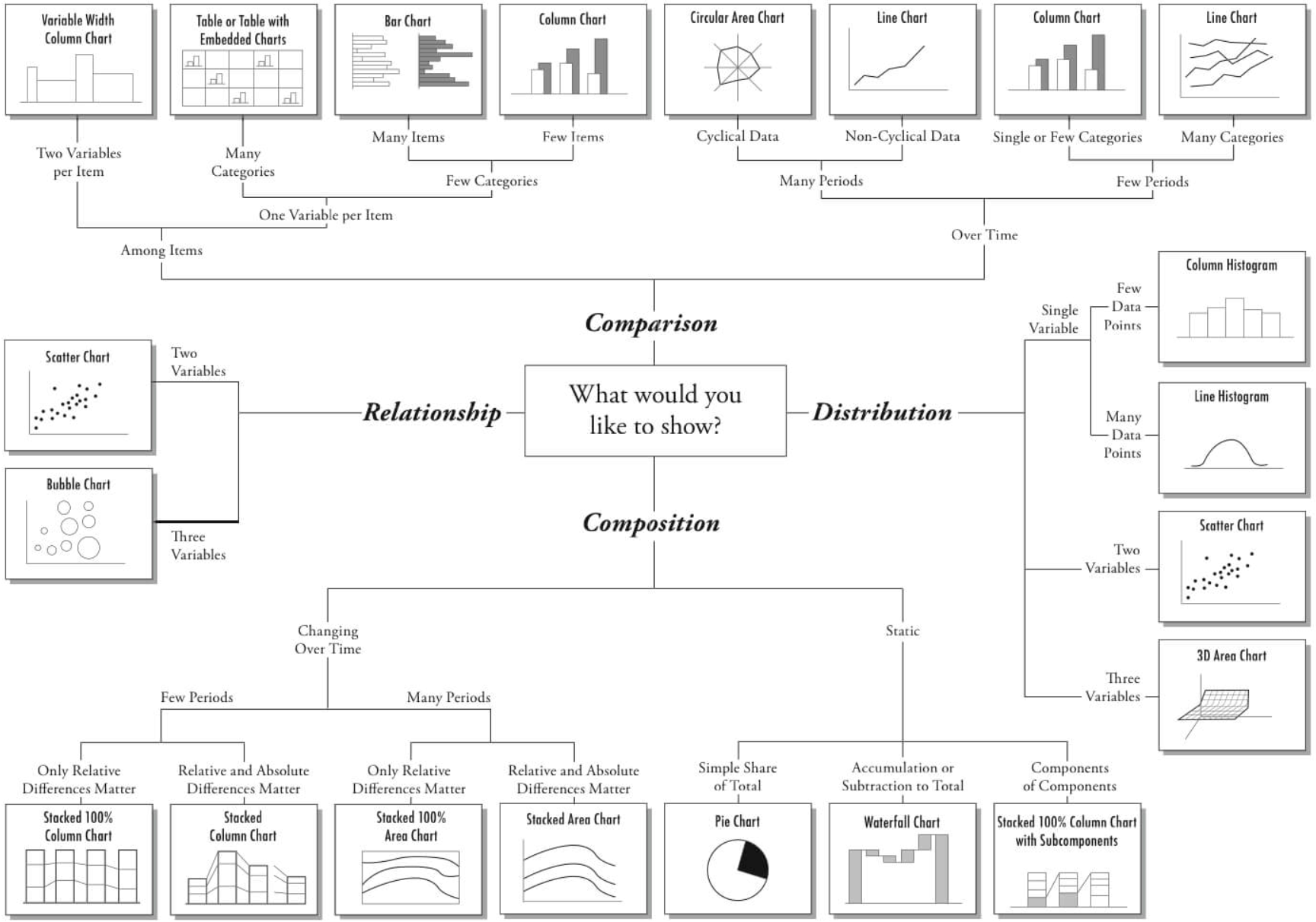
Crosstab / Highlight table

4.Distribution

Histogram chart

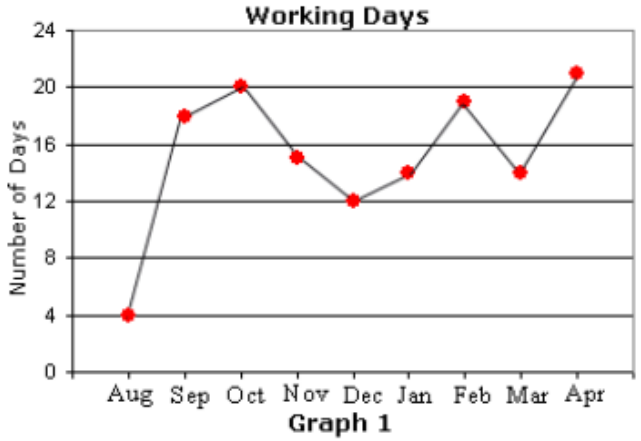
Box plot

Chart Suggestions—A Thought-Starter



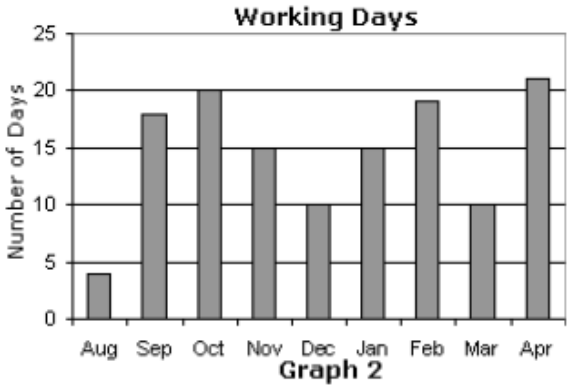
Month	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr
Number of Days	4	18	20	15	12	14	19	14	21

Relationship >



Graph 1

< Comparison



Graph 2

Distribution >

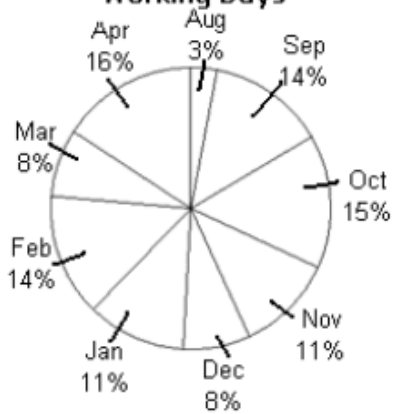
Working Days

Month	Number of School Days
August	●
September	● ● ● ●
October	● ● ● ● ●
November	● ● ● ●
December	● ● ●
January	● ● ● ●
February	● ● ● ● ●
March	● ● ●
April	● ● ● ● ●

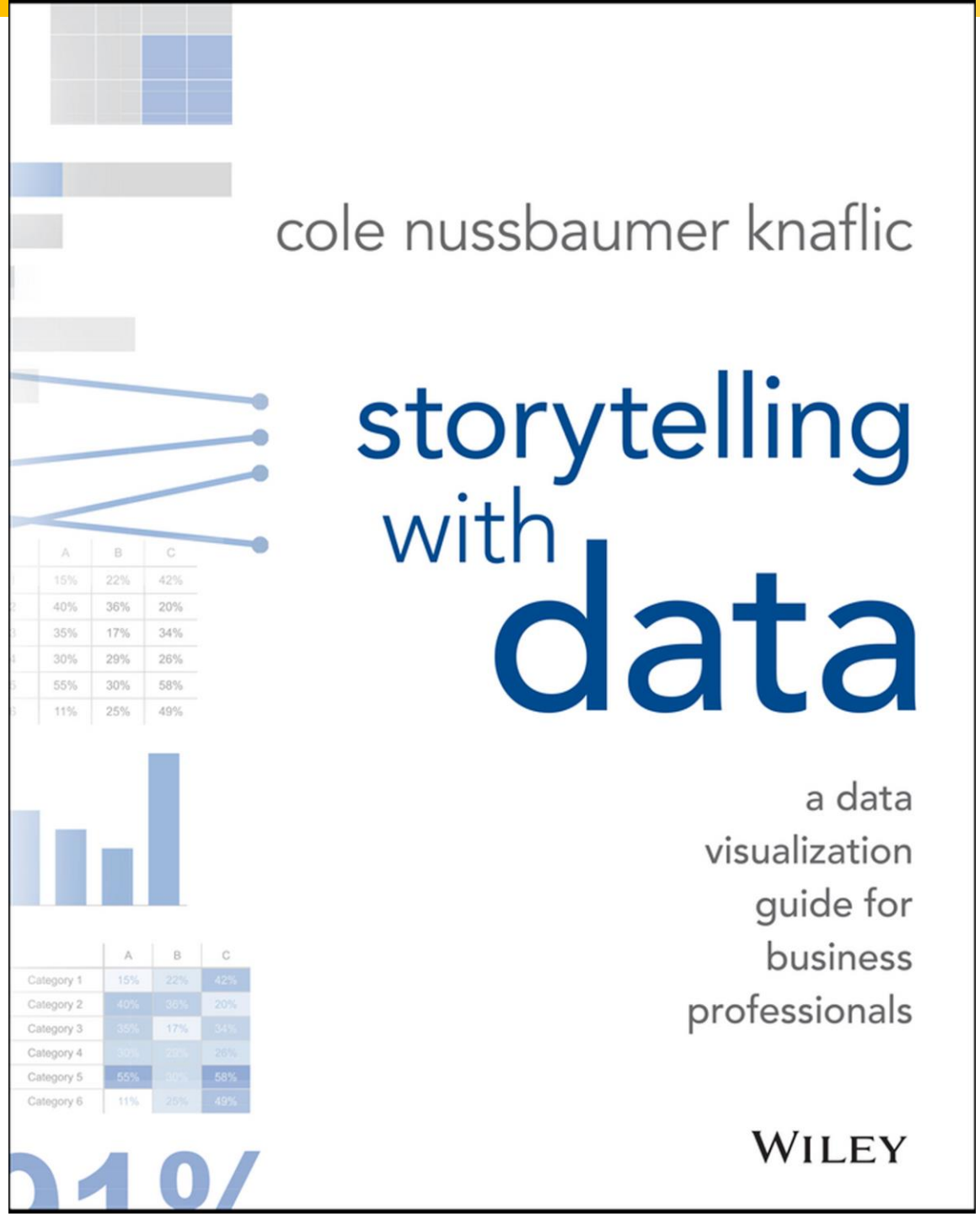
● = 4 days

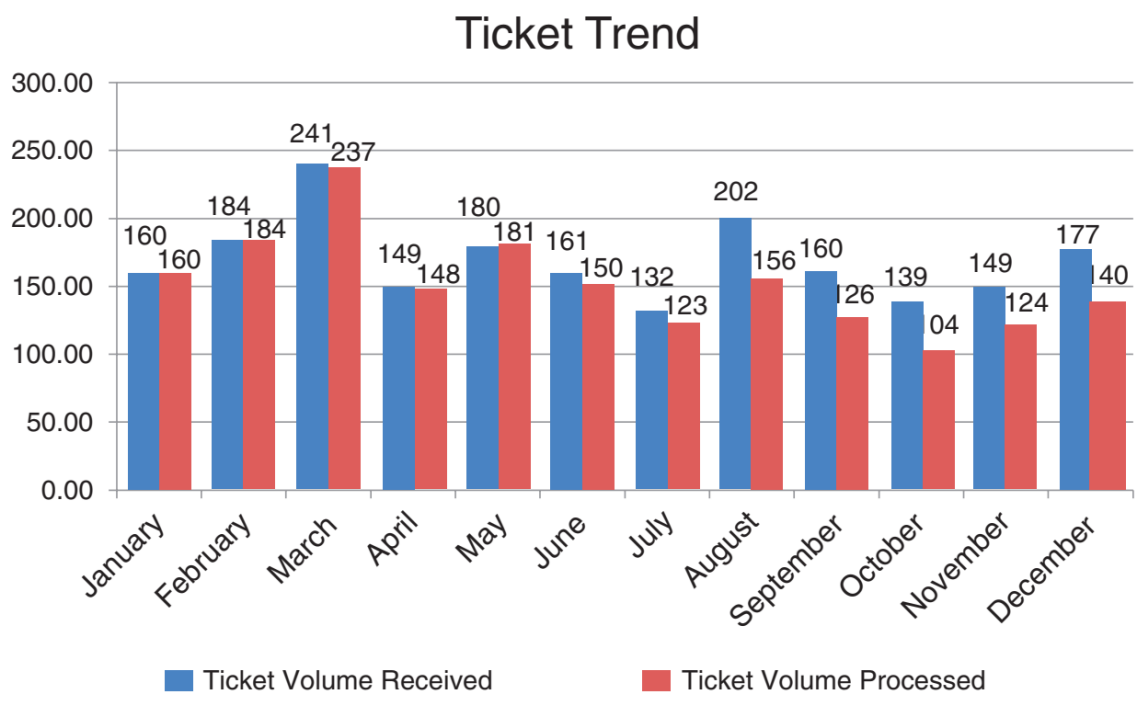
Graph 3

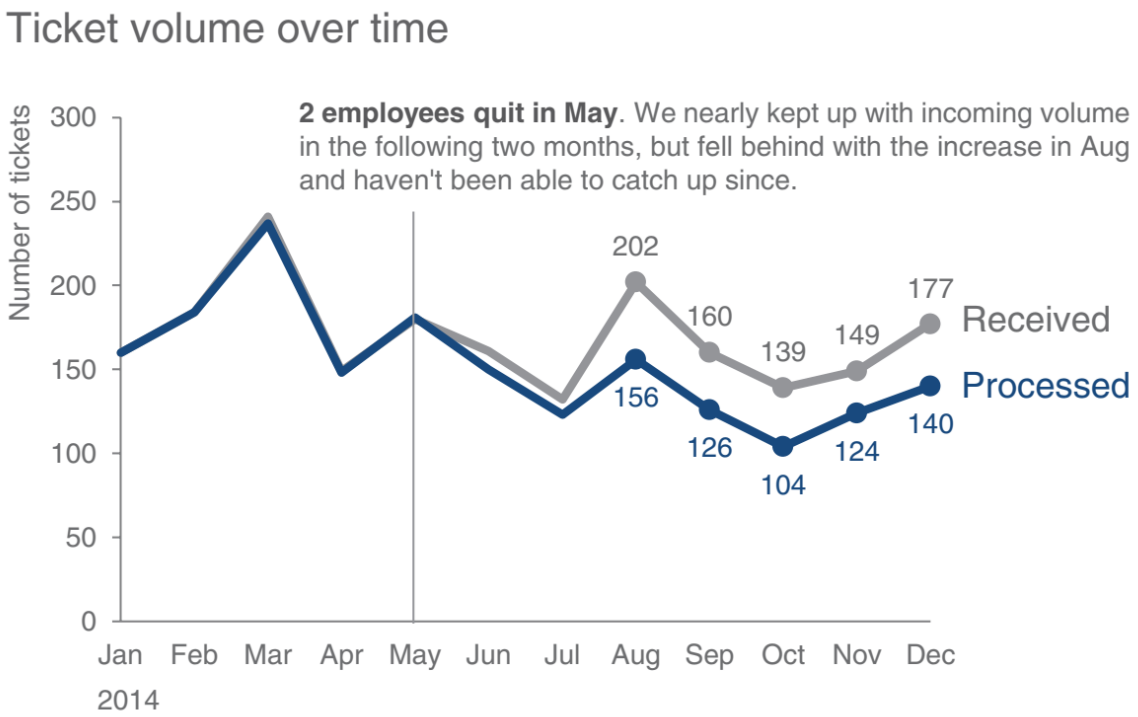
< Composition



Graph 4


cole nussbaumer knaflic
storytelling with data
a data visualization guide for business professionals
WILEY

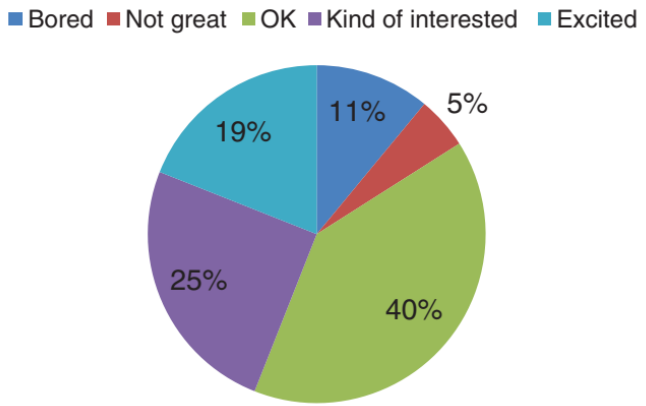




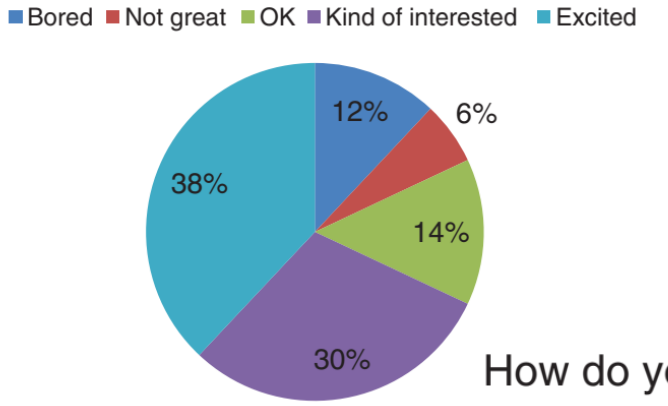
Data source: XYZ Dashboard, as of 12/31/2014 | A detailed analysis on tickets processed per person and time to resolve issues was undertaken to inform this request and can be provided if needed.

Survey Results

PRE: How do you feel about doing science?

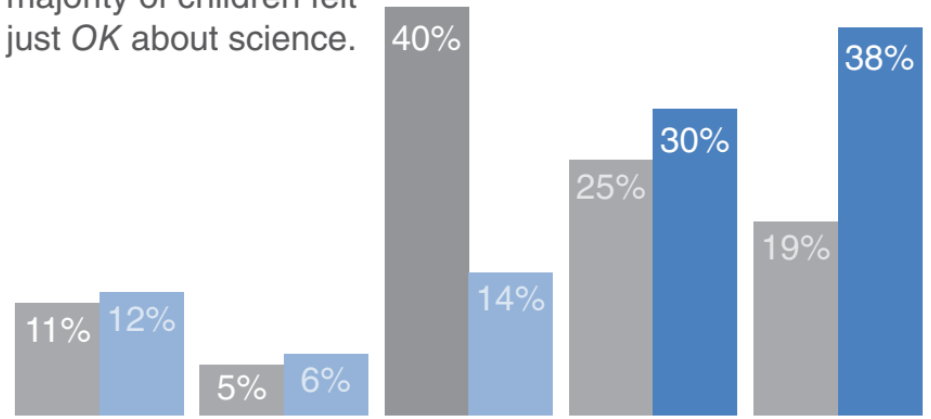


POST: How do you feel about doing science?



How do you feel about science?

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

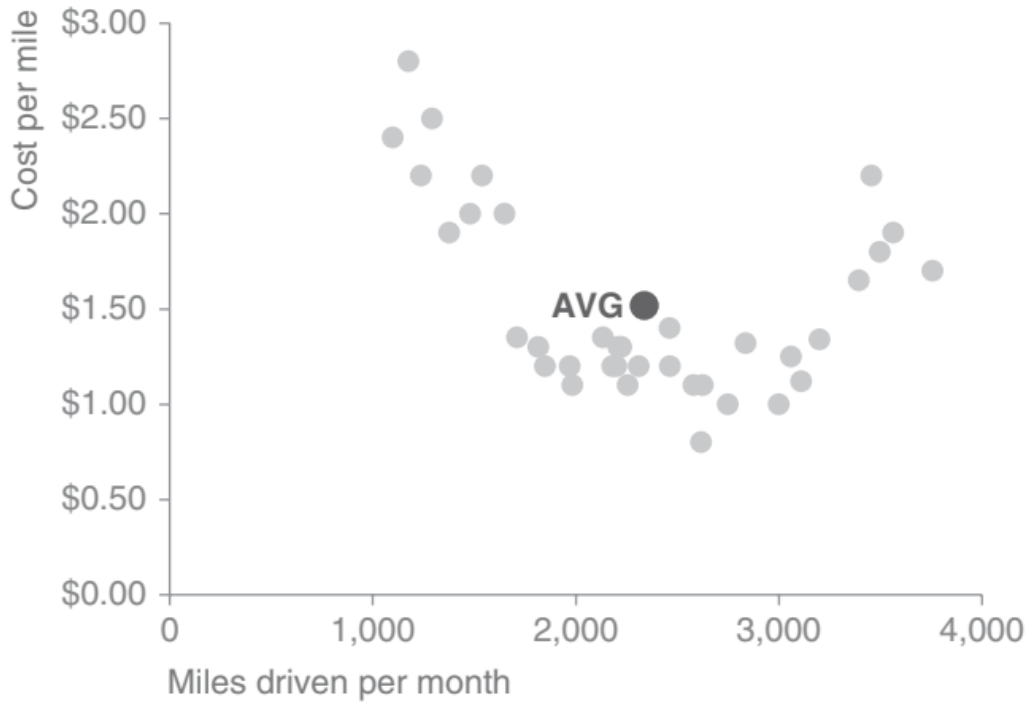


AFTER program, more children were *Kind of interested* & *Excited* about science.

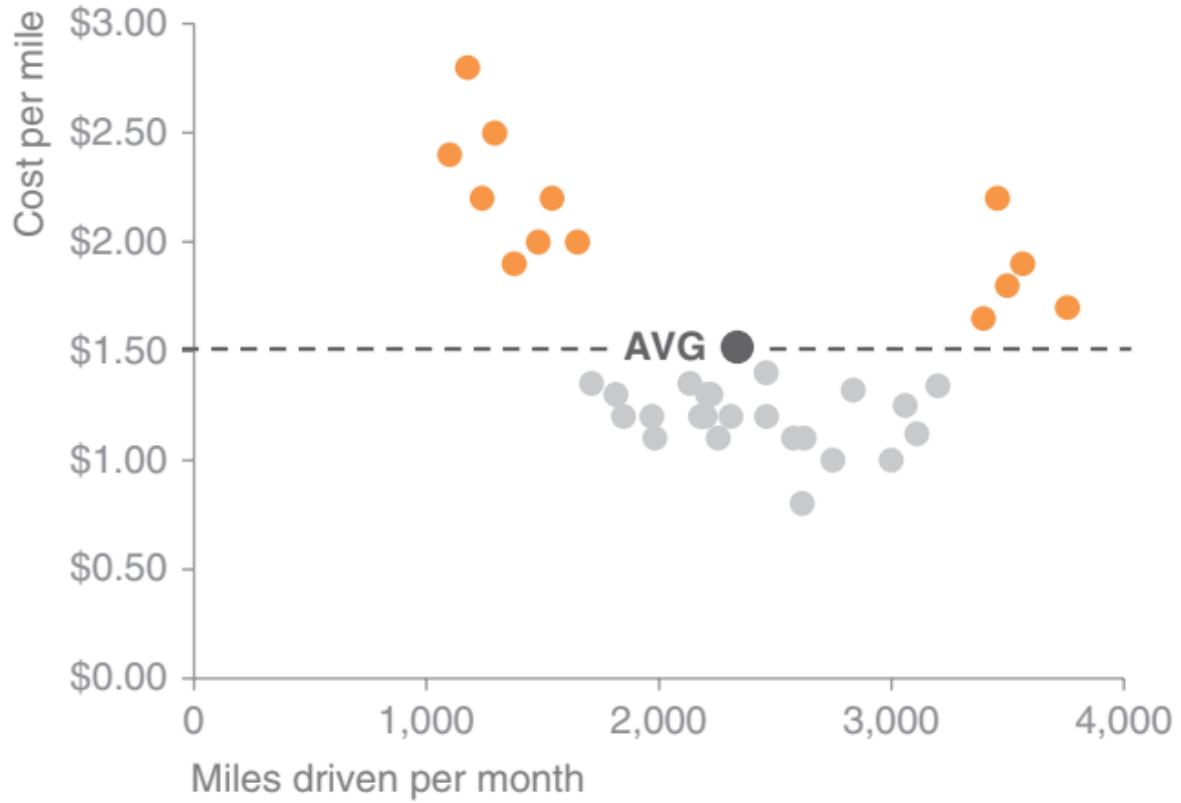
Based on survey of 100 students conducted before and after pilot program (100% response rate on both surveys).

GBDi

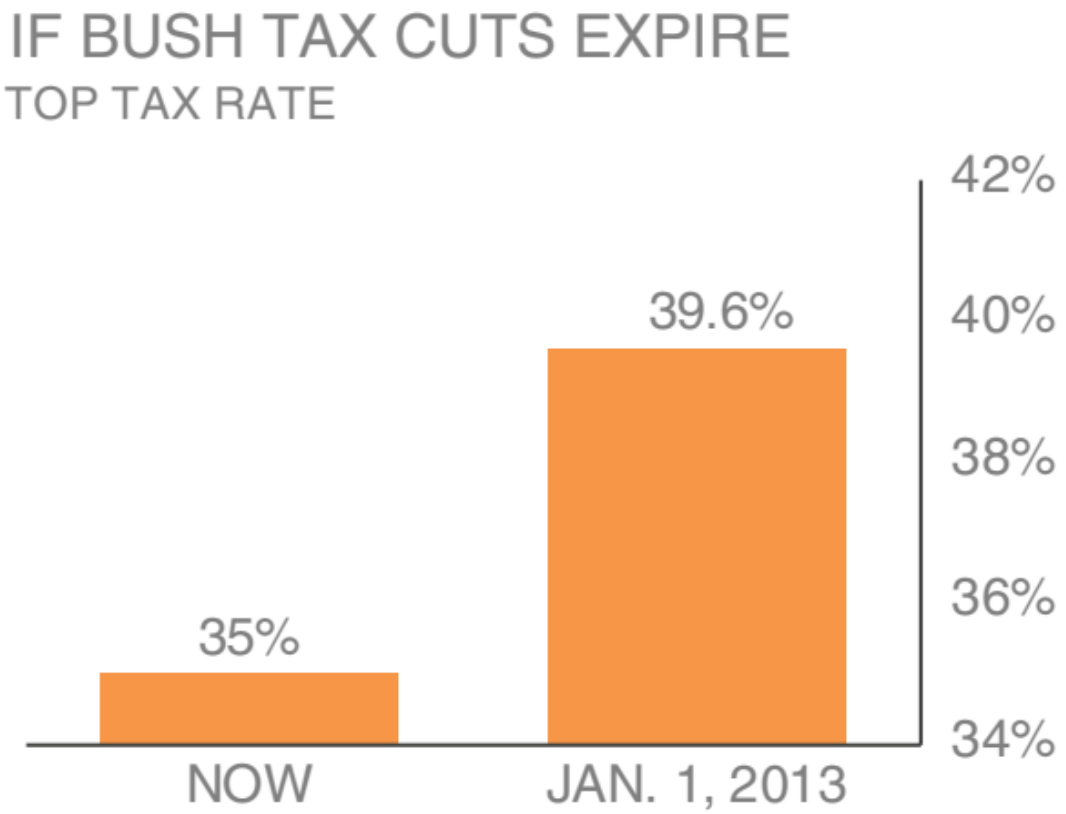
Cost per mile by miles driven



Cost per mile by miles driven



Non-zero baseline: as originally graphed



Zero baseline: as it should be graphed

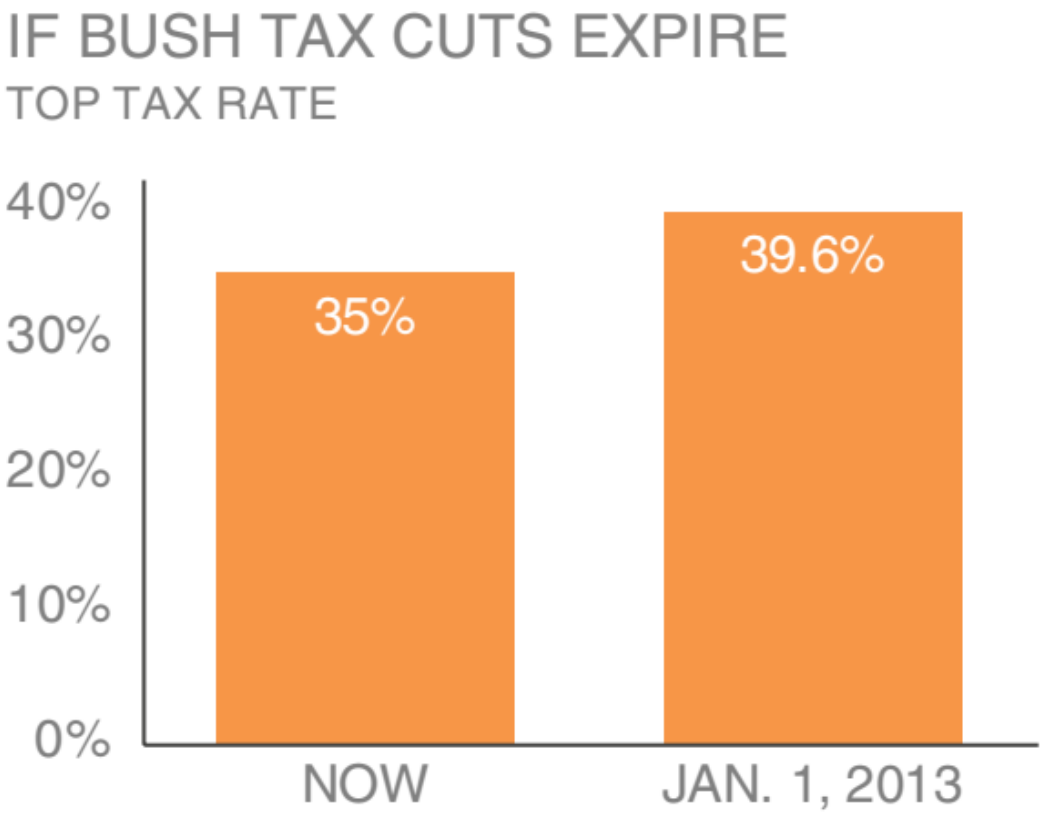
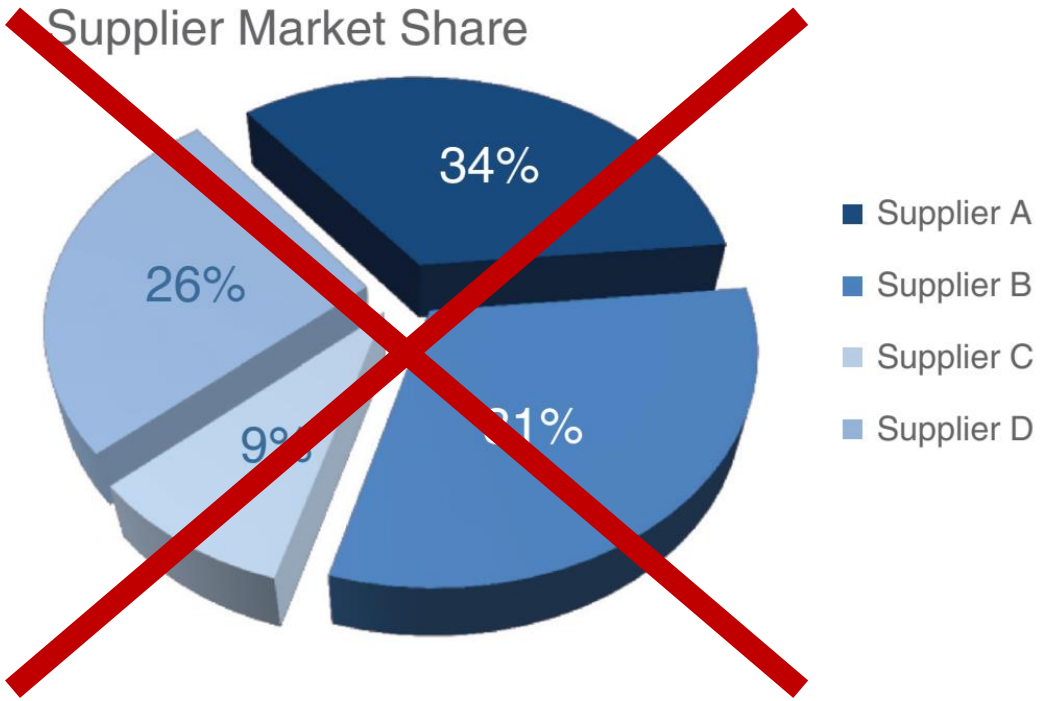
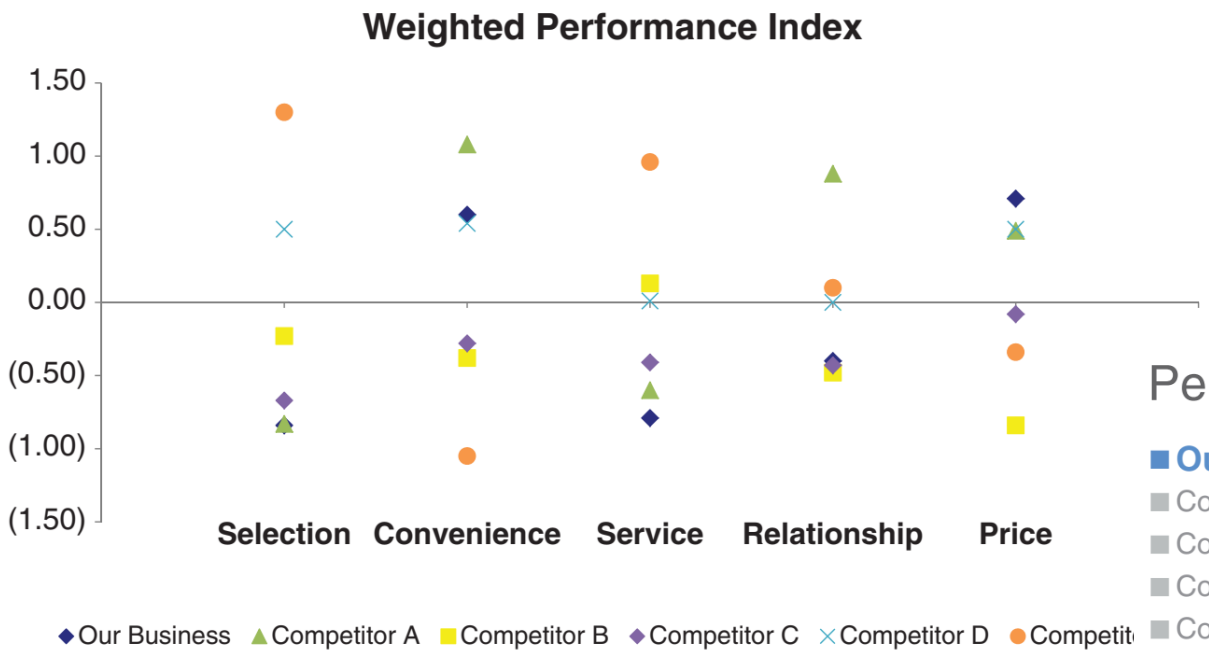


FIGURE 2.13 Bar charts must have a zero baseline



Supplier Market Share

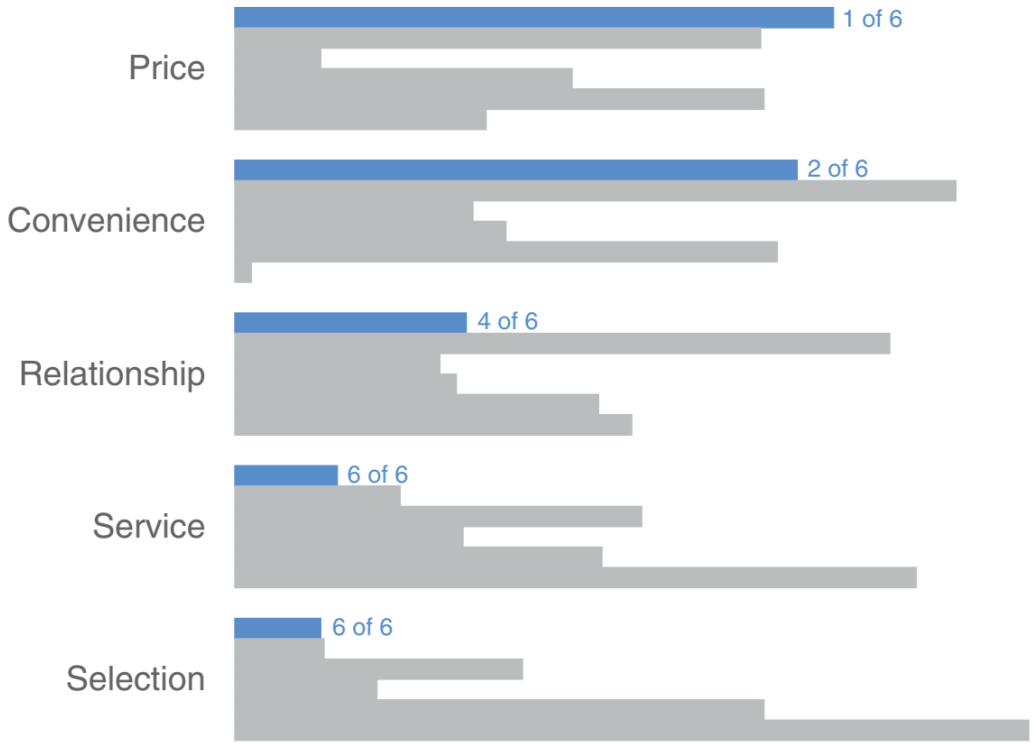
Supplier A	34%
Supplier B	31%
Supplier D	26%
Supplier C	9%
Total	100%

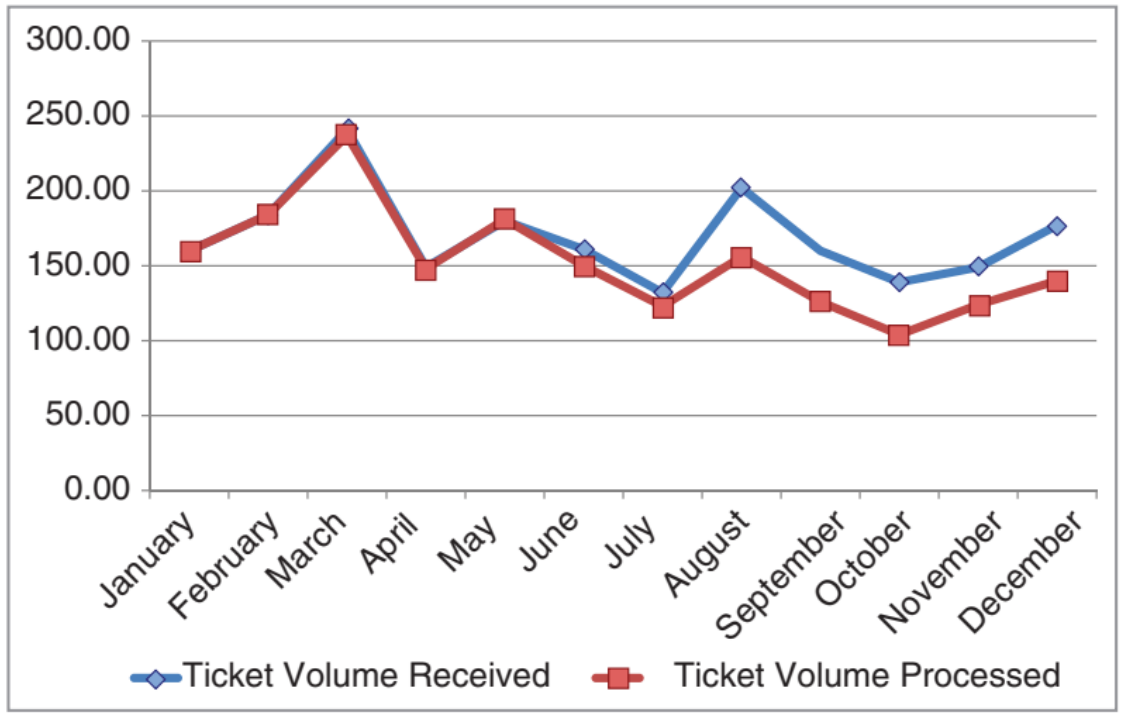


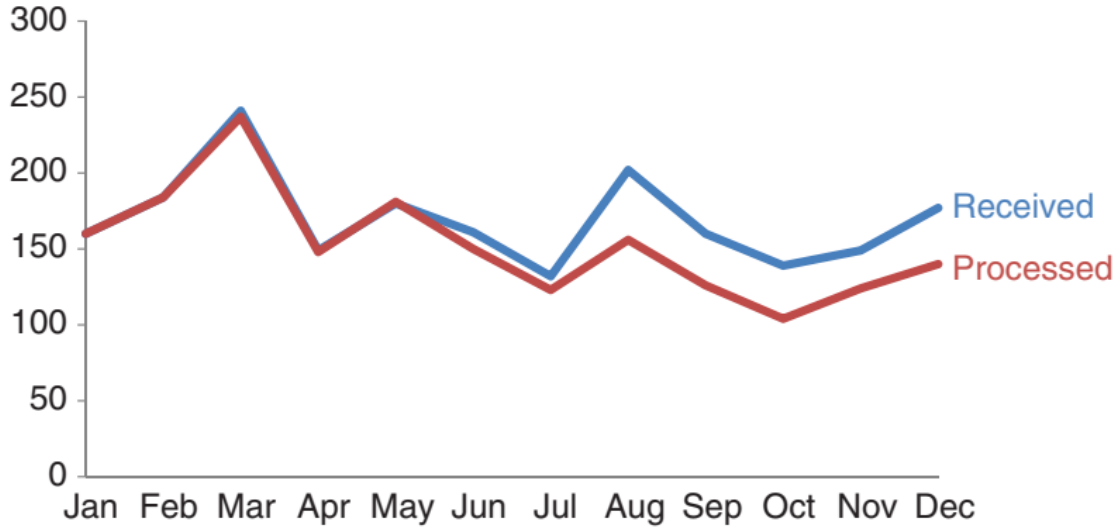
Performance overview

- **Our business**
- Competitor A
- Competitor B
- Competitor C
- Competitor D
- Competitor E

Weighted performance index | relative rank




300.00
250.00
200.00
150.00
100.00
50.00
0.00
January
February
March
April
May
June
July
August
September
October
November
December
Ticket Volume Received
Ticket Volume Processed


300
250
200
150
100
50
0
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Received
Processed

Country Level Sales Rank Top 5 Drugs

Rainbow distribution in color indicates sales rank in given country from #1 (red) to #10 or higher (dark purple)

Country	A	B	C	D	E
AUS	1	2	3	6	7
BRA	1	3	4	5	6
CAN	2	3	6	12	8
CHI	1	2	8	4	7
FRA	3	2	4	8	10
GER	3	1	6	5	4
IND	4	1	8	10	5
ITA	2	4	10	9	8
MEX	1	5	4	6	3
RUS	4	3	7	9	12
SPA	2	3	4	5	11
TUR	7	2	3	4	8
UK	1	2	3	6	7
US	1	2	4	3	5

Top 5 drugs: country-level sales rank

RANK 1 2 3 4 5+

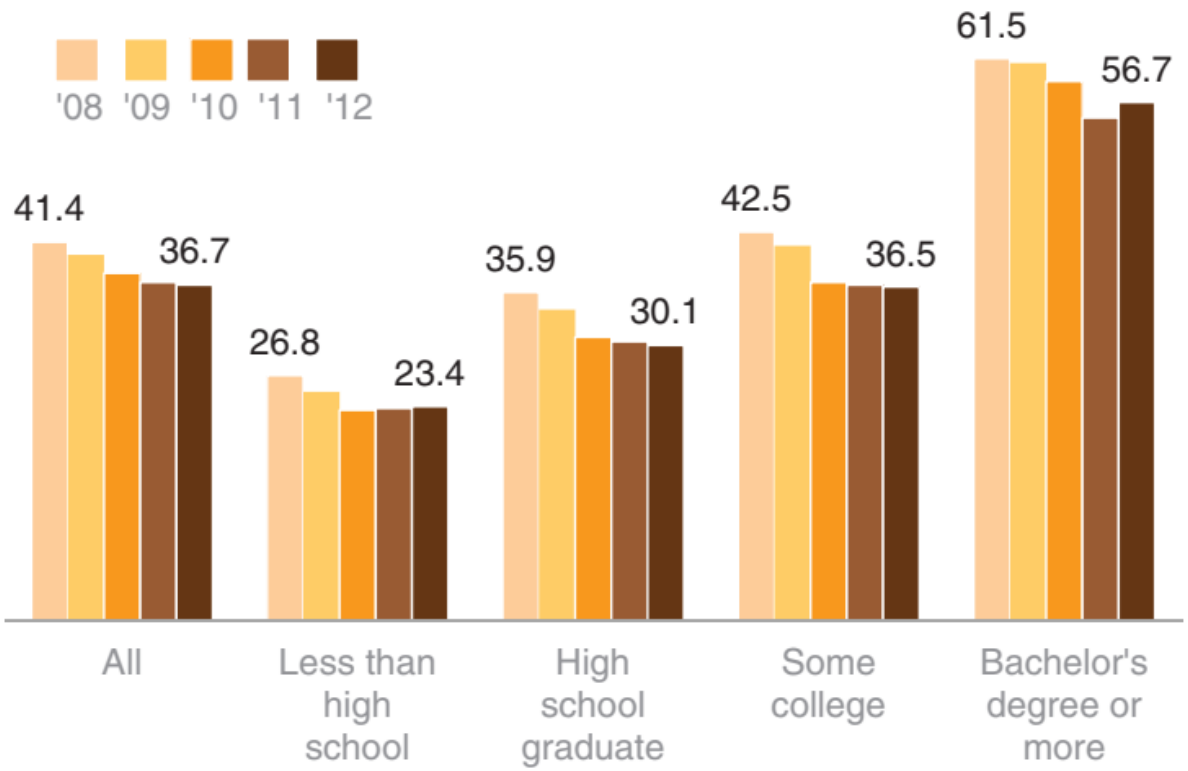
COUNTRY | DRUG

	A	B	C	D	E
Australia	1	2	3	6	7
Brazil	1	3	4	5	6
Canada	2	3	6	12	8
China	1	2	8	4	7
France	3	2	4	8	10
Germany	3	1	6	5	4
India	4	1	8	10	5
Italy	2	4	10	9	8
Mexico	1	5	4	6	3
Russia	4	3	7	9	12
Spain	2	3	4	5	11
Turkey	7	2	3	4	8
United Kingdom	1	2	3	6	7
United States	1	2	4	3	5

GBDi

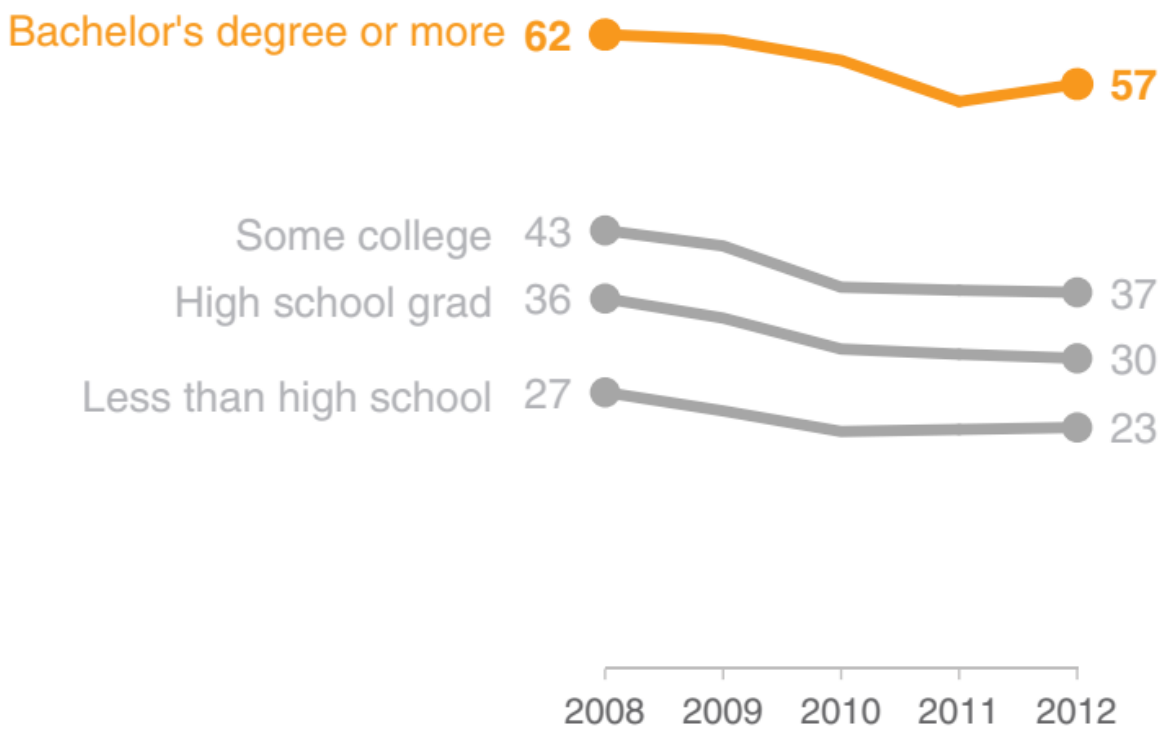
New Marriage Rate by Education

Number of newly married adults per 1,000 marriage eligible adults



New marriage rate by education

Number of newly married adults per 1,000 marriage eligible adults



Follow us on

GBDi

govbigdata

Twitter

Blockdit

GBDi

gbdi.depa.or.th

YouTube

Government Big Data Institute (GBDi)

Line Official

@gbdi