

# SWE404/DMT413

# BIG DATA ANALYTICS

Lecture 13: Data Visualization

Lecturer: Dr. Yang Lu

Email: luyang@xmu.edu.my

Office: A1-432

Office hour: 2pm-4pm Mon & Thur

# How Can We Acquire Information?

Listen



Taste & Smell



Touch



Look



# Efficiency

## Sound Signal

Listen



Bandwidth: **about 0.1 KB / s**

## Electronic / Light Signal

Look



**> 100 MB/s**

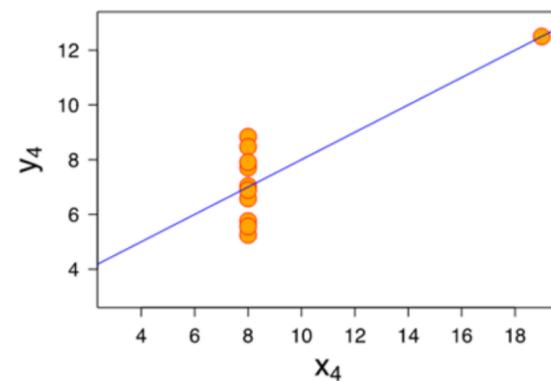
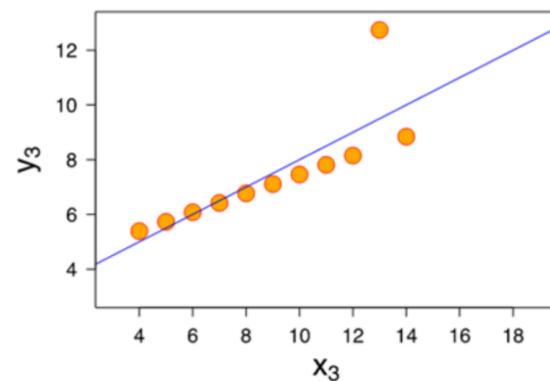
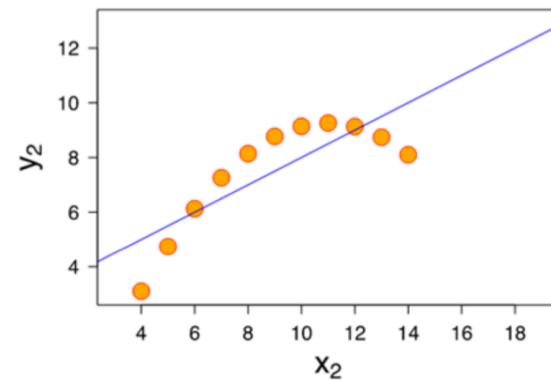
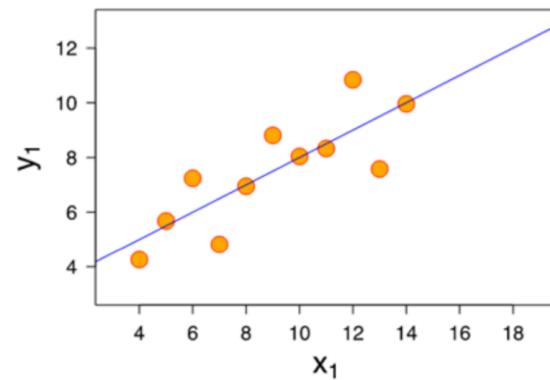
# Definition of Data Visualization

- “The action or fact of visualizing; the power or process of forming a mental picture or vision of something not actually present to the sight; a picture thus formed.”  
-- Oxford English Dictionary.
- “... finding the artificial memory that best supports our natural means of perception.”  
-- Bertin, 1983
- “The use of computer-supported, interactive, visual representations of abstract data to amplify cognition.”  
-- Cart, Mackinlay, Shneiderman, 1999

# Image vs. Number

	Set A		Set B		Set C		Set D	
	X	Y	X	Y	X	Y	X	Y
0	10	8.04	10	9.14	10	7.46	8	6.58
1	8	6.95	8	8.14	8	6.77	8	5.76
2	13	7.58	13	8.74	13	12.74	8	7.71
3	9	8.81	9	8.77	9	7.11	8	8.84
4	11	8.33	11	9.26	11	7.81	8	8.47
5	14	9.96	14	8.10	14	8.84	8	7.04
6	6	7.24	6	6.13	6	6.08	8	5.25
7	4	4.26	4	3.10	4	5.39	19	12.50
8	12	10.84	12	9.13	12	8.15	8	5.56
9	7	4.82	7	7.26	7	6.42	8	7.91
10	5	5.68	5	4.74	5	5.73	8	6.89
mean	9.00	7.50	9.00	7.50	9.00	7.50	9.00	7.50
std	3.32	2.03	3.32	2.03	3.32	2.03	3.32	2.03
corr	0.82		0.82		0.82		0.82	
lin. reg.	$y = 3.00 + 0.500x$		$y = 3.00 + 0.500x$		$y = 3.00 + 0.500x$		$y = 3.00 + 0.500x$	

# Image vs. Number



# Image vs. Words

**News illustrated**

The sudden explosion of a South Korean entertainer called Psy, has given the world Gangnam Style. It is setting the music and dance world on fire and has a set sequence. We simplify them for your perusal

**GANGNAM STYLE!!!**  
**The 5 basic steps**

**Step 1** Riding the horse  
Dress classy and dance cheesy  
Cross your hands like taking the horse reins and pulse up and down  
Do small jumps with your legs spread like you are riding a horse  
Footsteps: R L R R L R L L L

**Step 2** Lassoing the sexy lady  
Lassoing motion with your right arm  
Continue with the horse-riding movement  
Footsteps: R L R R L R L L L

**Step 3** Now everybody is looking at me  
Hands in pockets or waist and small hip side movements combined with the foot steps  
Finish this move dragging the right leg to the left leg.  
Footsteps A: L R L R L R L R  
Footsteps B: L R L R L R L R

**Step 4** Combine a few 'sexy' moves  
Now move your hips to the outside with quick movements twice  
Spread and flex your legs. Move your body up and down three times  
Footsteps C: R L R R L R L L R  
Footsteps D: R L R R L R L L R

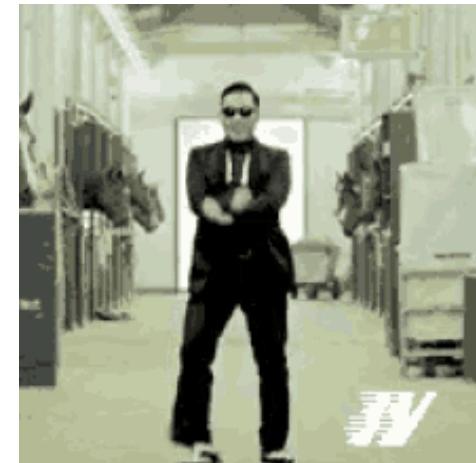
**Step 5** Finish with a cool pose  
Cross your hands over your left leg  
Spread your arms and raise your right leg (position A). Now get down quickly on your right leg and flex the left one. Now rotate your right arm and with your hand touch your chin doing a "L" shape with your thumb and index fingers (position B)

★ When to use the steps during the chorus ★

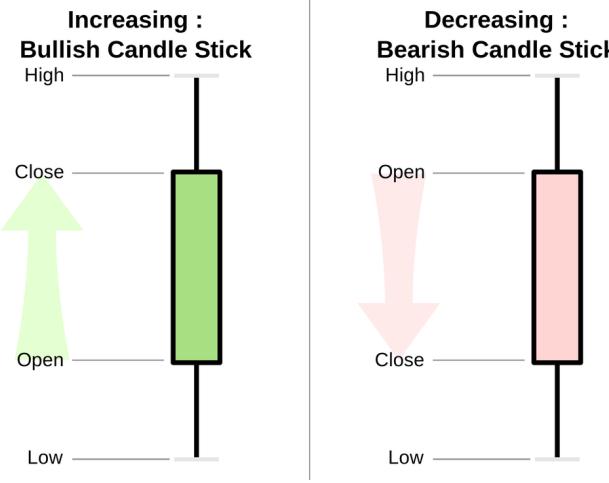
Step 1 Step 2 Step 3 or Step 1 (in the last chorus)  
Oppa is Gangnam style, ahhhhh... Gangnam style... Oh, oh oh oh, Oppa is Gangnam style...  
ahhhh... Gangnam style... Oh, oh oh oh, Oppa is Gangnam style... Eeeeh- Sexy Lady...  
Step 1 or Step 2 (in the last chorus) Step 4 Step 5 (only at the end)  
Oppa is Gangnam style, ahhhhh... Gangnam style... Oh, oh oh oh, Oppa is Gangnam style... Eeeeh- Sexy Lady oh oh oh oh. Oppa is Gangnam style.

Source: YouTube

HUGO A. SANCHEZ©Gulf News



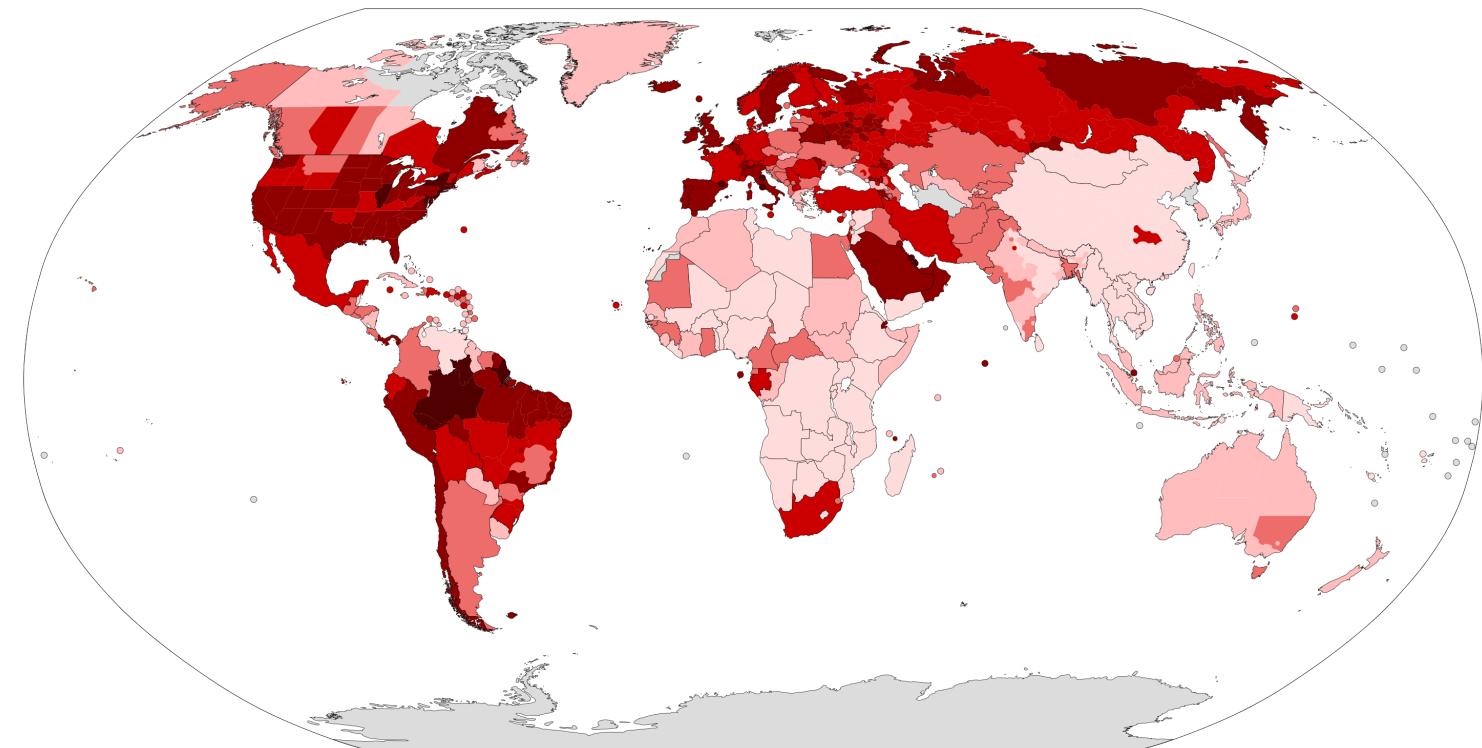
# Example: Candlestick Chart



# Example: Summarization of Airlines in United States



## Example: COVID-19 Pandemic by Country and Territory



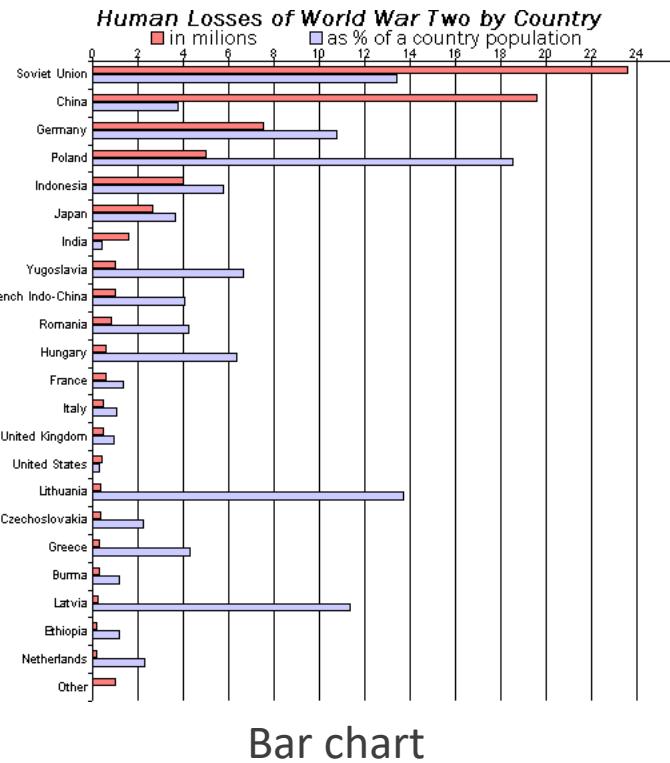
- Check this website for more COVID-19 visualization results: <https://pandemic.internationalsos.com/2019-ncov/covid-19-data-visualisation>

# Basic Diagrams for Data Visualization

- Bar chart
- Histogram
- Line chart
- Pie chart
- Scatter plot

# Bar chart

- Presents categorical data.
- The bars can be plotted vertically or horizontally.
- If the total amount of each subject is meaningful, stacked bar chart can be adopted.



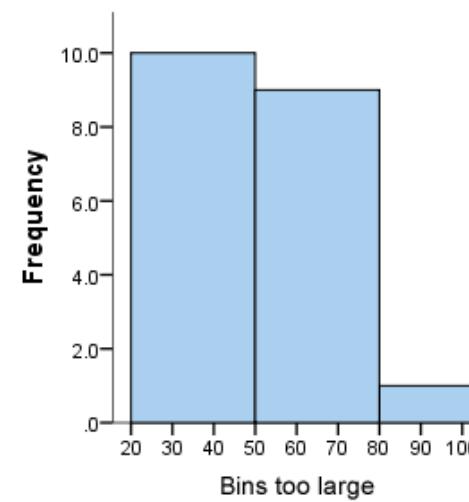
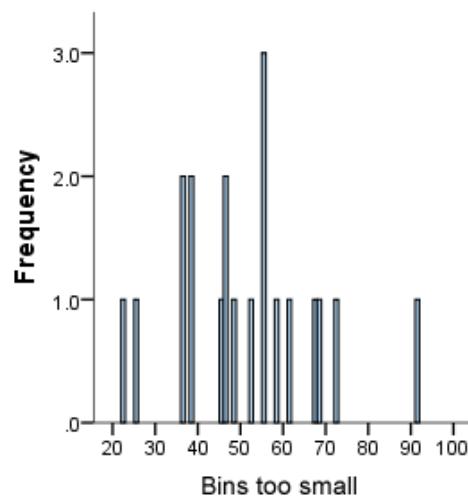
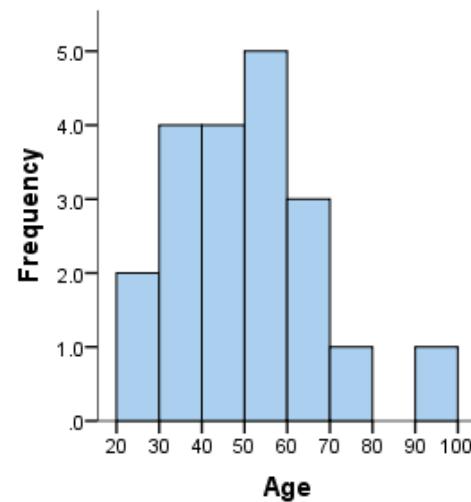
Bar chart



Stacked bar chart

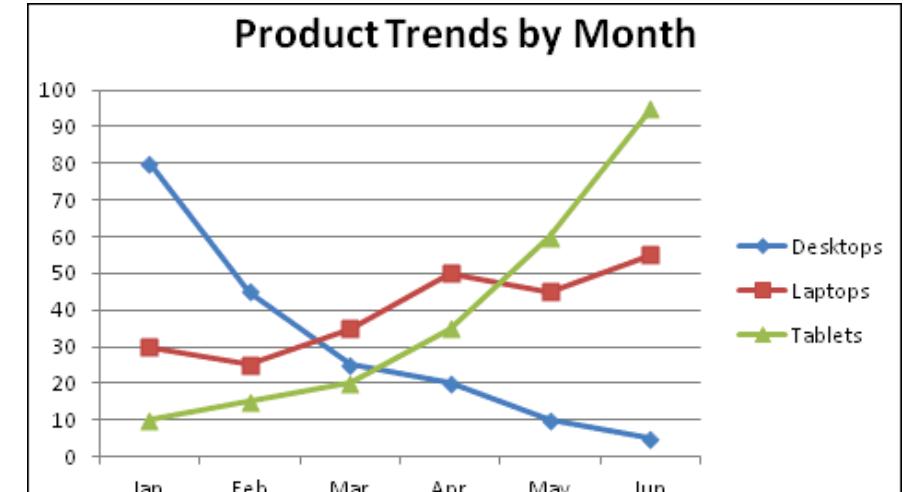
# Histogram

- Put continuous data into bins.
- The bin size is important to visualization effect.



# Line Chart

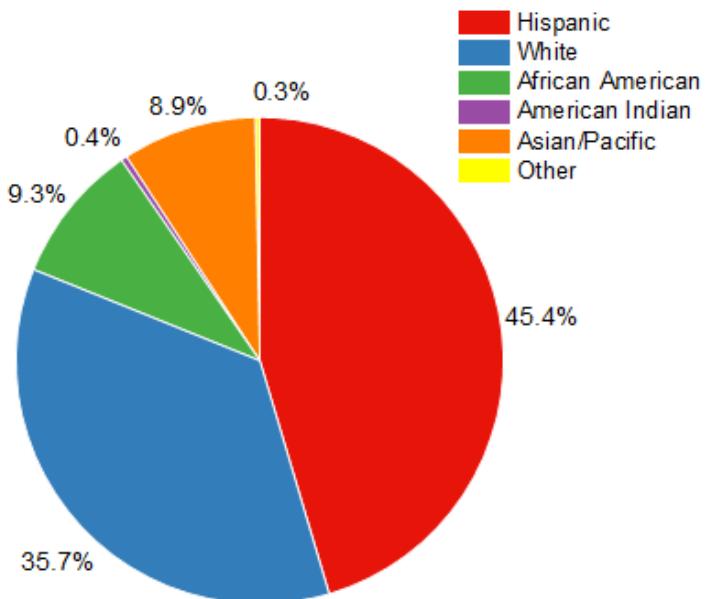
- Compared with bar chart, line chart focuses on the changes (increasing / decreasing) and trends.



TechOnTheNet.com

# Pie Chart

- Pie chart focuses on the percentage.



## Normal pie chart



## Multilevel pie chart



XIAMEN UNIVERSITY MALAYSIA  
廈門大學 馬來西亞分校



厦门大学信息学院

SCHOOL OF INFORMATICS XIAMEN UNIVERSITY

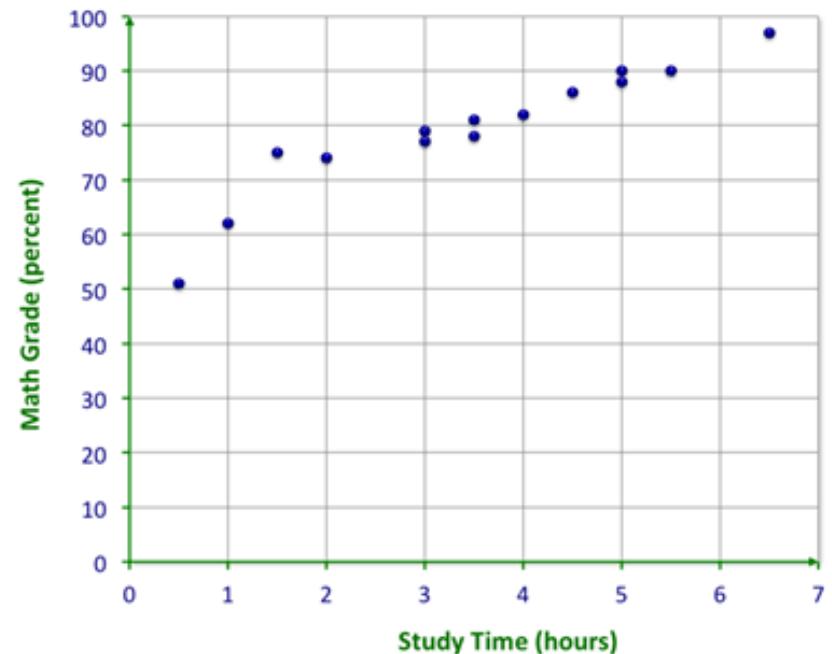


厦门大学计算机科学系  
Computer Science Department of Xiamen University

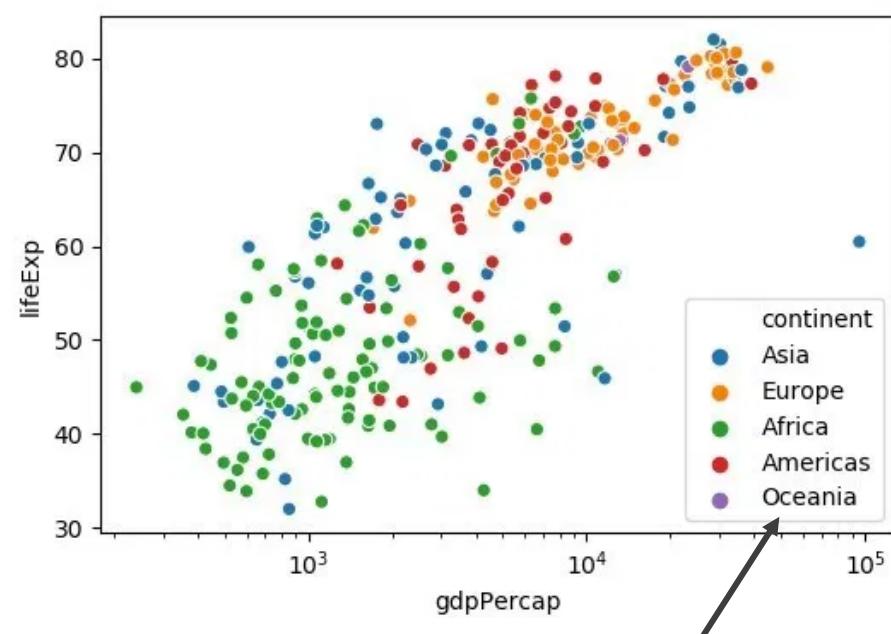
# Scatter Plot

- Display values for typically two variables for a set of data.
- How can we add another variable, e.g. age?

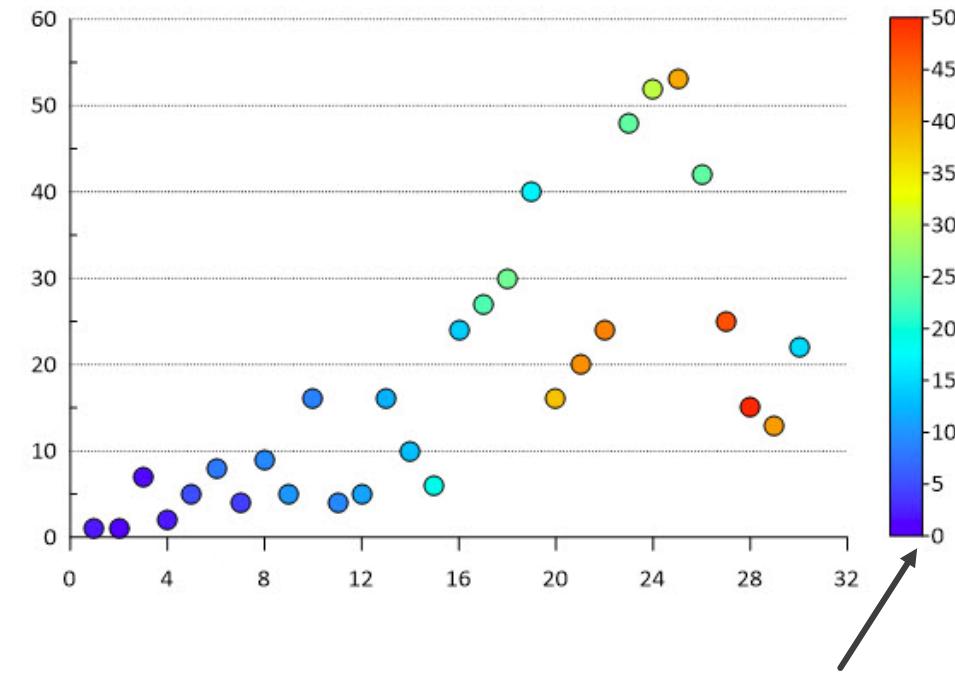
Does studying increase your grade?



# Colored Scatter Plot



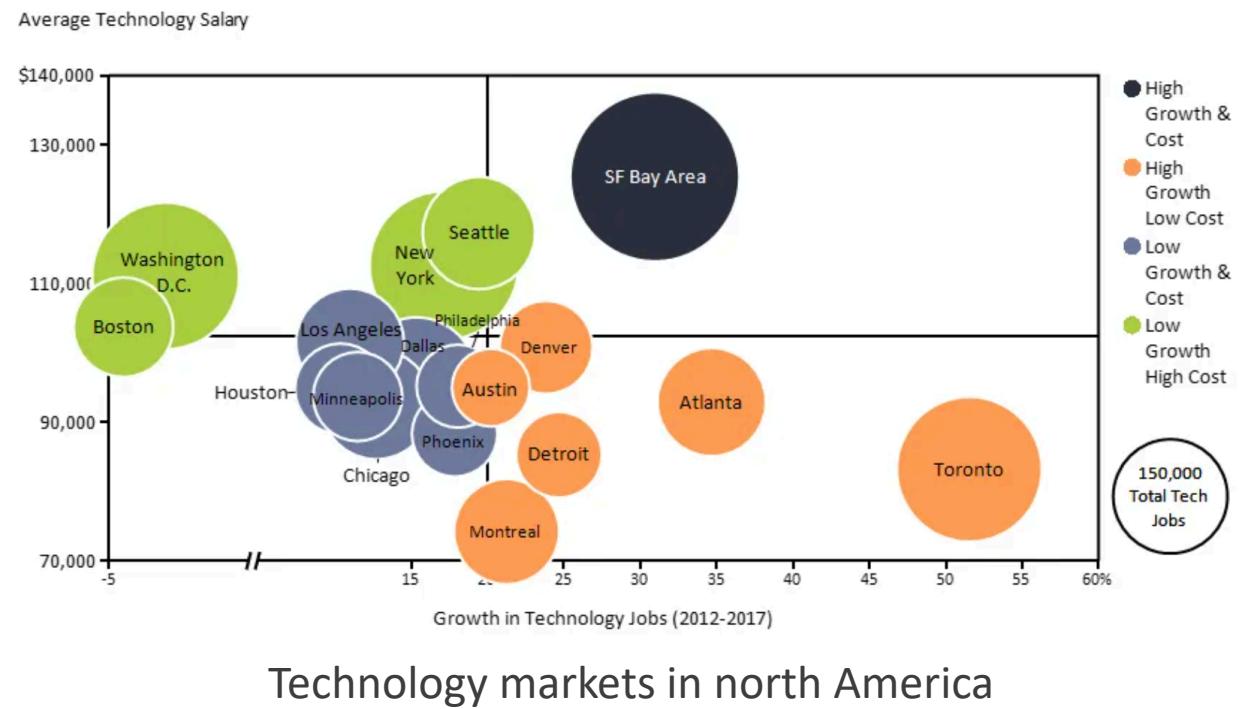
Use legend for discrete variable



Use color map for continuous variable

# Bubble Chart

- Can we add the fourth variable?
- Use the size of bubble to show the magnitude.
- For more examples:  
<https://www.mekkographics.com/resources/charts-by-type/bubble/>

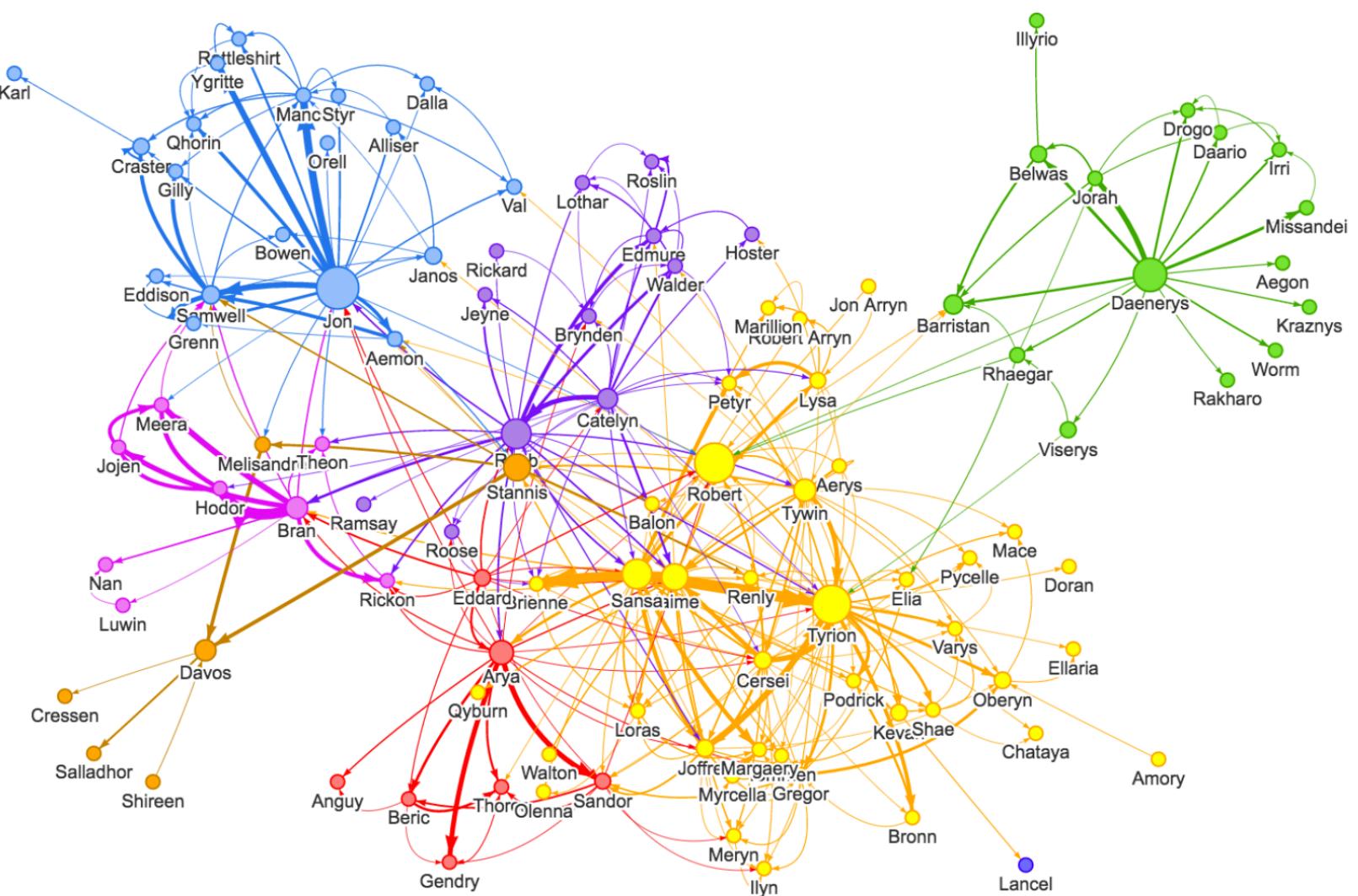


# Advanced Diagrams for Data Visualization

- Network
- Streamgraph
- Treemap
- Heatmap
- Sankey diagram
- Chord diagram
- Word Cloud

# Network

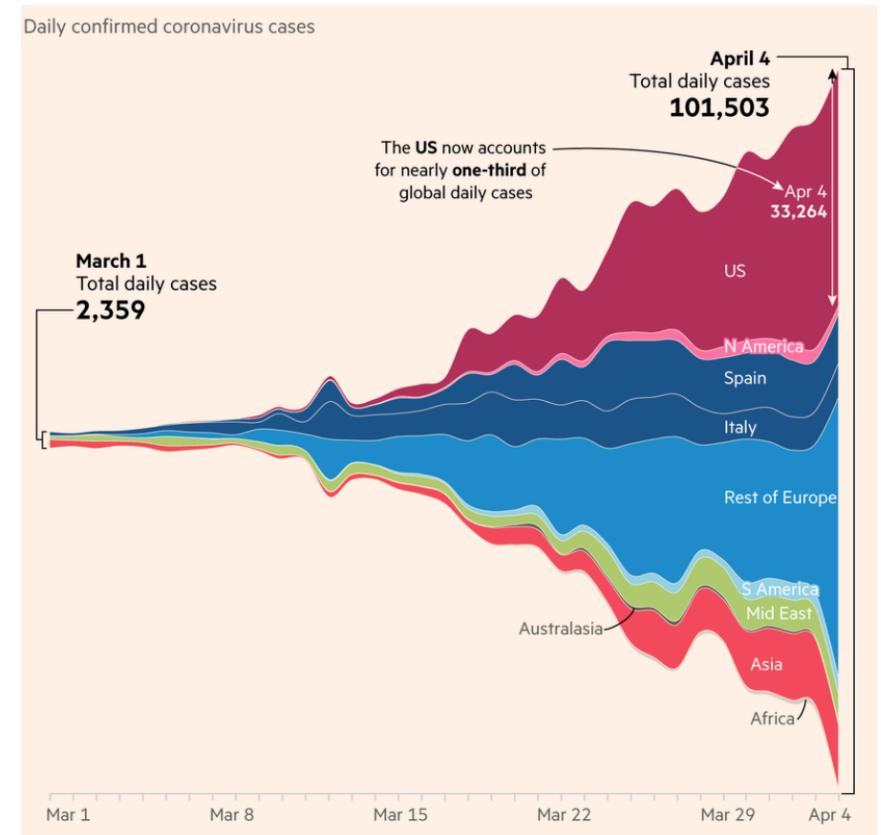
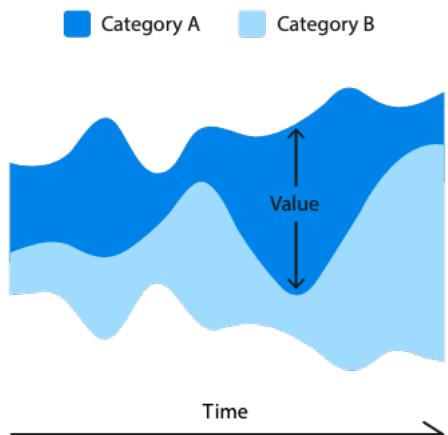
- Show the relationship between entities and clusters.
  - Additional information can be added by color, vertex size, edge weight, etc.



# Character relations in Game of Thrones

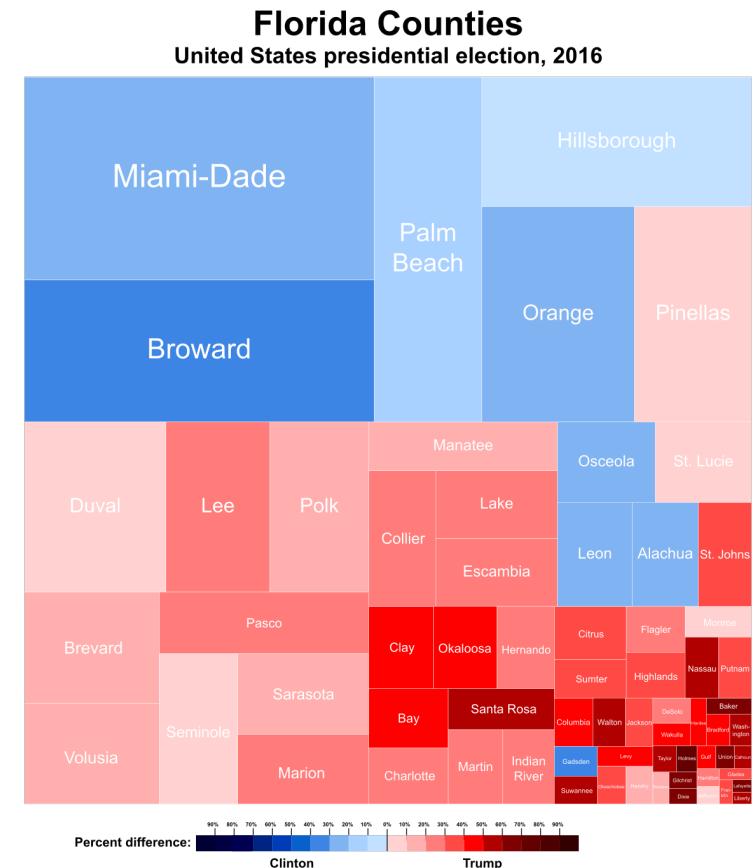
# Streamgraph

- Similar to line chart, but with more visual impact.
- The x-axis is usually time, so that the value changed and total amount along time can be focused.



# Treemap

- Similar to multilevel pie chart, but with more clear labelling, structures and color information.
- This example shows the 2016 United States presidential election results in Florida by county, on a color spectrum from Democratic blue to Republican red.



# Heatmap

■ Use color map  
to give more  
obvious visual  
cues.

**College major of wife and husband: College graduates married in the previous year**  
*Couples in which women married for the first time only; 2009-2016 American Community Survey (N=27,806)*  
*Ratio of observed to expected frequency*

Philip N. Cohen

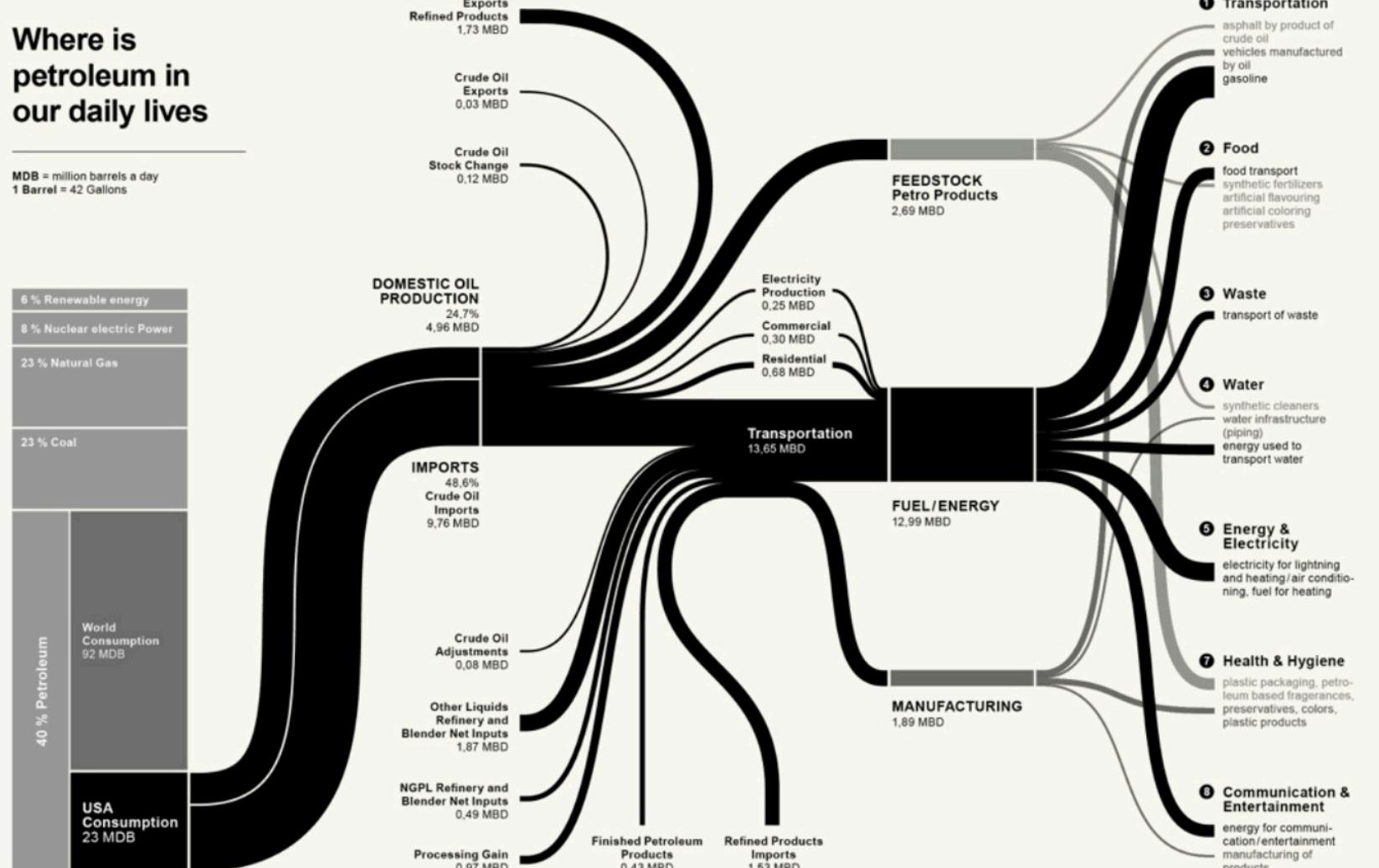
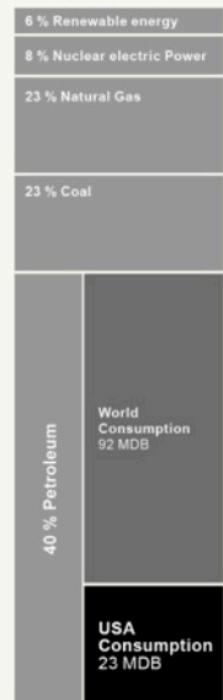
WIFE	HUSBAND																											
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
1 Agriculture	19.6	3.4	.5	.0	.4	1.0	.6	1.1	.8	.0	5.3	.4	.0	1.2	.9	.0	.8	.0	.4	.7	.9	1.0	1.5	.4	.7	1.7	.7	.5
2 Environment/NatRes	1.9	11.4	2.3	2.0	.9	.6	.4	.9	.2	.0	.0	2.4	.9	1.3	1.2	.0	.9	1.7	.0	1.2	.7	.5	.6	1.3	.6	2.3	.5	.6
3 Architecture	.2	.0	17.9	1.2	.4	1.2	.3	1.6	.0	.8	.0	.5	1.6	.8	1.5	.0	.5	2.7	.0	.3	.1	.1	.6	.8	1.0	.0	.7	.6
4 Area/Ethnic/CivStud	.7	1.7	2.2	6.9	1.3	1.5	.9	.6	.0	2.6	.0	1.0	.2	1.3	1.4	.6	.1	1.4	.0	1.0	.3	.0	.6	1.2	1.8	.6	.8	2.1
5 Communications	.7	.9	.4	1.2	2.6	.6	.8	.7	.8	.5	1.3	1.2	.9	.6	.9	1.1	.9	1.2	1.2	.7	.9	1.0	1.2	1.2	.8	.8	1.2	1.2
6 Computer/InfoSci	.4	.2	.4	.0	.8	4.7	.1	2.1	4.6	.1	.0	.3	.2	.5	.7	.4	.1	.5	.1	1.0	.4	.3	.7	.3	.4	.7	.4	.2
7 Education Admin	1.7	.9	.7	.7	.9	.8	3.3	.7	.9	.7	1.4	.8	1.3	.7	.8	1.0	1.2	.7	2.1	.6	.9	1.4	.4	.7	.9	1.0	1.0	1.3
8 Engineering	.2	.6	.9	.2	.4	1.8	.3	3.5	1.1	.5	.0	.3	.1	.6	.5	.4	.5	.2	.3	1.2	.2	.4	.9	.4	.4	.6	.4	.2
9 Engineering Techn	1.3	.0	.0	.0	.1	2.2	1.3	2.5	7.0	1.5	.0	.0	.0	1.1	.5	1.1	.3	.7	.7	.0	.2	.0	.0	.4	.8	1.4	.5	.0
10 Linguistics/Foreign	.2	1.6	1.2	1.2	.9	.9	.8	.9	.9	5.0	.0	1.9	.1	1.1	2.0	1.6	1.4	1.5	.7	.8	.7	.6	1.3	1.4	1.3	.4	.7	1.4
11 Family/Consumer Sci	.5	1.6	1.7	2.2	1.5	1.0	.5	1.0	.7	.5	2.5	1.3	2.4	.6	.8	1.5	2.5	.5	.3	.7	1.3	1.3	1.6	1.1	.7	.2	1.0	.6
12 English Language/Lit	.9	1.2	.8	1.4	1.3	1.1	.7	.7	.5	1.3	.4	2.9	1.6	.9	1.7	1.5	.6	.9	.9	1.3	.6	.8	1.3	1.3	.4	.8	1.8	
13 LiberalArts/Human	.3	.4	3.6	3.1	1.3	1.0	.8	.8	.6	1.2	.0	1.2	9.4	.7	1.6	1.3	.5	.2	1.0	1.5	.8	1.5	.3	1.1	.8	.8	.7	1.2
14 Biology and LifeSci	.9	1.2	1.2	.9	.6	.9	.5	1.0	1.0	.9	.8	.8	.6	3.0	.8	.8	1.0	.7	.8	1.8	1.1	.8	.8	1.0	.6	1.2	.7	.8
15 Math/Stats	.0	.3	.8	.0	.5	1.6	1.3	1.3	2.1	1.7	.0	1.3	.8	.6	5.7	1.7	.5	1.2	1.8	2.0	.5	.3	.8	.7	.9	1.1	.7	.5
16 Interdisc/multidisc	1.5	.5	.3	1.8	1.0	1.1	.6	.9	2.5	1.3	.0	.9	.8	1.6	.7	2.9	1.2	1.6	1.7	1.0	1.8	.5	.0	1.1	.9	.3	.8	1.2
17 PhysFit, Park/Rec	1.3	.7	.5	.4	.6	.5	1.3	.7	2.0	.8	1.6	.4	.6	1.1	.2	1.5	5.7	.5	.4	.5	.9	.8	2.5	.9	.5	.9	1.3	1.2
18 Philosophy/ReligStud	.9	1.5	1.2	1.4	.9	.7	.4	.9	1.0	3.7	.0	1.8	.7	.7	2.8	4.2	.4	8.3	1.6	.7	1.8	1.1	.0	1.0	1.2	.2	.6	1.9
19 Theology/ReligVoc	.0	.0	2.6	1.7	1.1	.3	2.4	.5	.4	1.5	.0	.4	.0	1.3	2.7	2.7	.0	4.0	31.0	.0	1.0	1.4	2.7	.3	1.7	1.5	.5	.4
20 PhysicalSci	.4	.8	.4	2.0	.4	.9	.6	.9	.5	1.9	.5	.8	.5	1.2	2.0	1.0	.6	.9	.2	7.2	.6	.6	.0	.9	.5	.8	.6	1.0
21 Psychology	.9	1.3	.8	1.0	1.0	.8	.8	.8	.4	.8	2.4	.8	.7	.8	1.0	1.4	1.0	2.0	1.2	.7	1.9	1.4	1.5	1.2	1.3	.8	1.0	1.3
22 CrimJustice/Fire	.6	.5	.5	1.5	1.0	.9	1.2	.5	.9	1.9	2.7	.2	2.3	.6	1.0	1.0	1.1	1.1	.3	.7	.5	6.3	1.1	.8	.9	1.8	1.2	.2
23 PubAff/Policy/SocWk	1.5	.6	1.1	3.9	1.0	1.1	.9	.8	1.1	1.9	2.0	1.4	1.3	.8	.3	.2	.4	.7	1.1	1.3	1.4	1.0	6.9	.9	.3	.6	1.2	.4
24 SocialSci	.4	.9	1.2	1.6	.9	.9	.5	.8	.4	.7	.3	1.2	1.1	1.0	1.1	1.4	.5	1.9	.3	1.0	1.4	.7	1.7	1.8	1.1	.4	.9	1.3
25 Fine Arts	.5	1.0	1.6	1.1	1.4	1.0	1.0	.7	.6	1.2	.5	1.8	1.3	.4	.9	.9	.8	1.0	1.1	.7	.7	.8	.6	1.0	4.4	.5	.8	.8
26 Medical/HealthSci	.8	1.1	.5	.7	.8	.7	1.1	1.0	1.0	.8	1.7	.6	.9	1.4	.8	1.0	1.6	.9	.7	.8	1.2	1.2	1.0	.7	.5	3.4	1.0	.9
27 Business	1.0	.7	.7	.5	.9	1.0	.6	1.0	1.3	1.1	.8	.7	.6	.7	.8	.5	.8	.3	.7	.6	.8	1.1	.7	.9	.6	.7	1.6	.6
28 History	.4	1.1	1.5	.5	.8	.7	1.0	.7	.2	.6	.7	2.0	.4	1.0	1.0	.8	.8	2.9	1.0	.8	1.2	.9	.7	1.6	1.0	.6	.8	2.9

# Sankey Diagram

- Show the data flow.
- The width of the arrows is proportional to the flow rate.

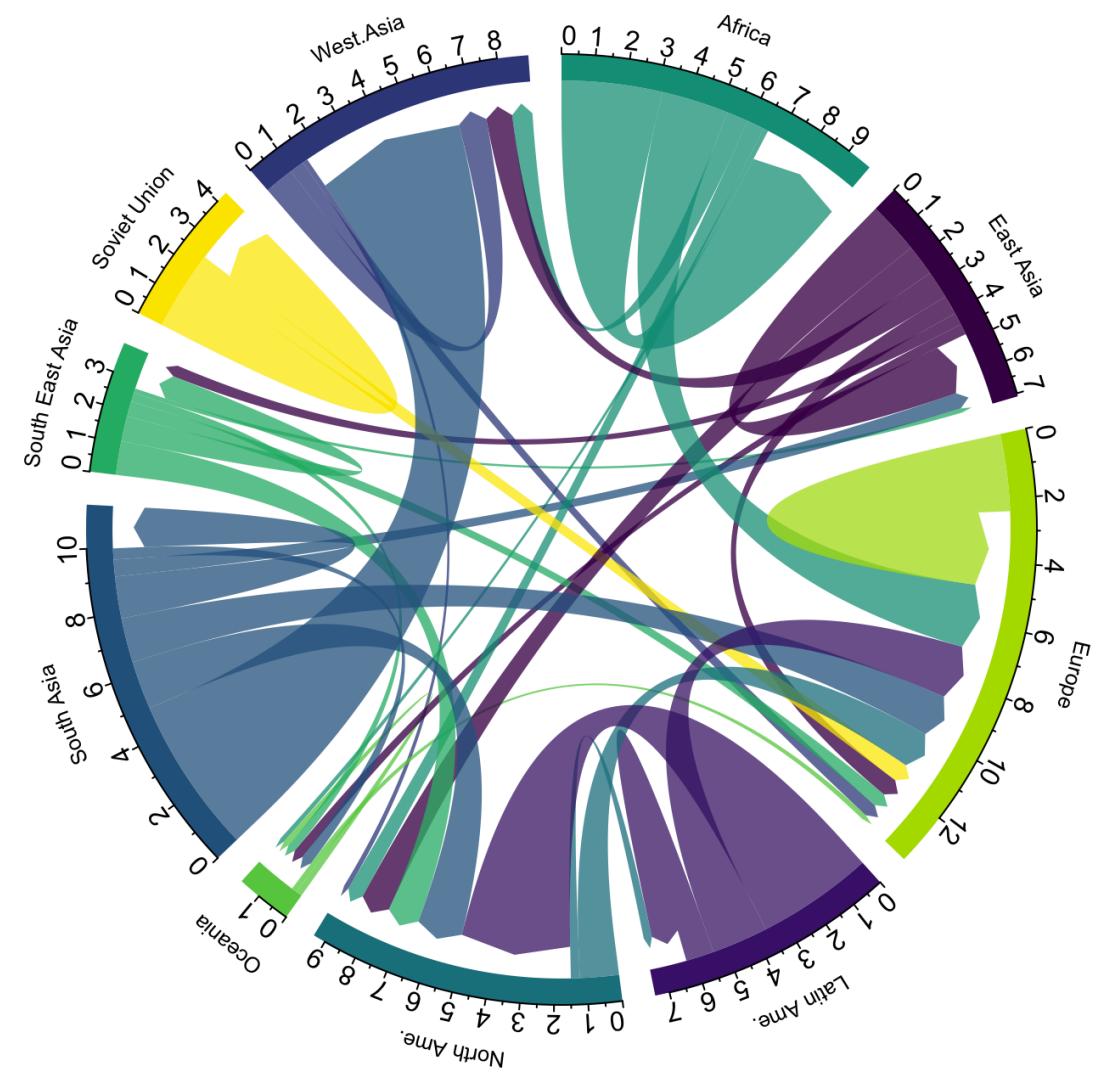
## Where is petroleum in our daily lives

MBD = million barrels a day  
1 Barrel = 42 Gallons



# Chord Diagram

- A chord diagram represents flows or connections between several entities (called nodes).
- Each entity is represented by a fragment on the outer part of the circular layout.
- Then, arcs are drawn between each entities.
  - The size of the arc is proportional to the importance of the flow.
- This example shows the number of people migrating from one country to another.



# Word Cloud

- Visually show a group of words with different frequency.

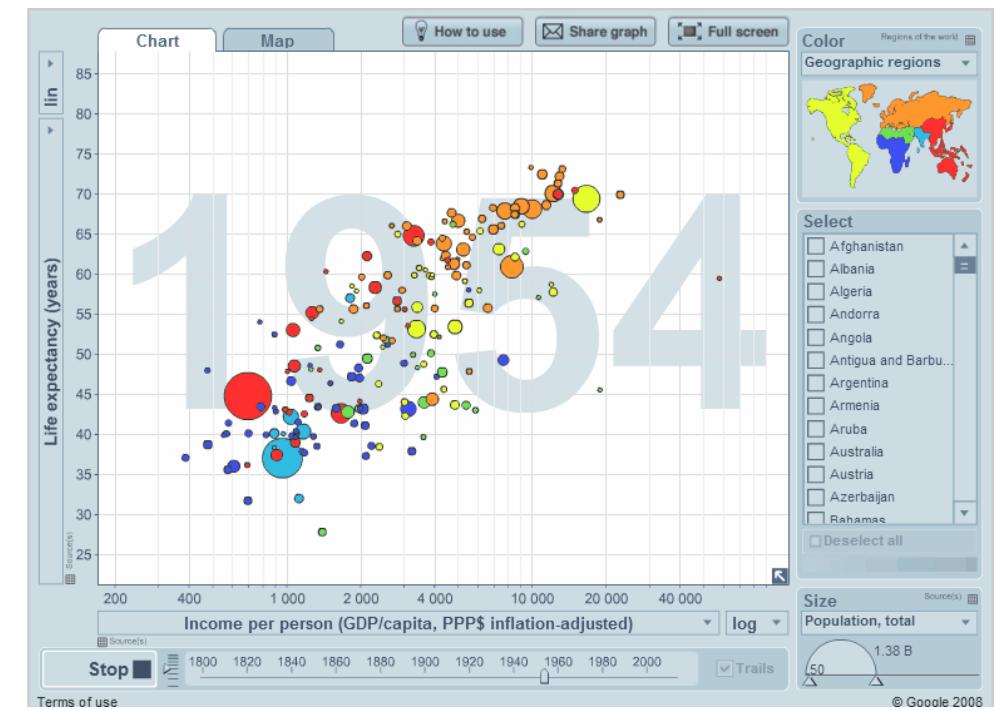
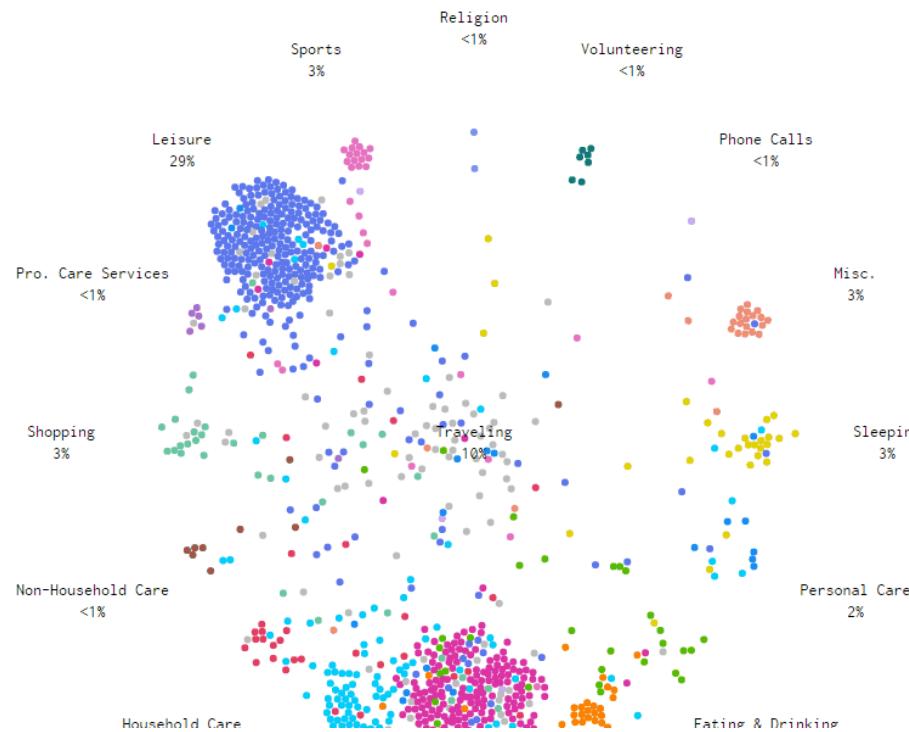


# Data Visualization with Animation

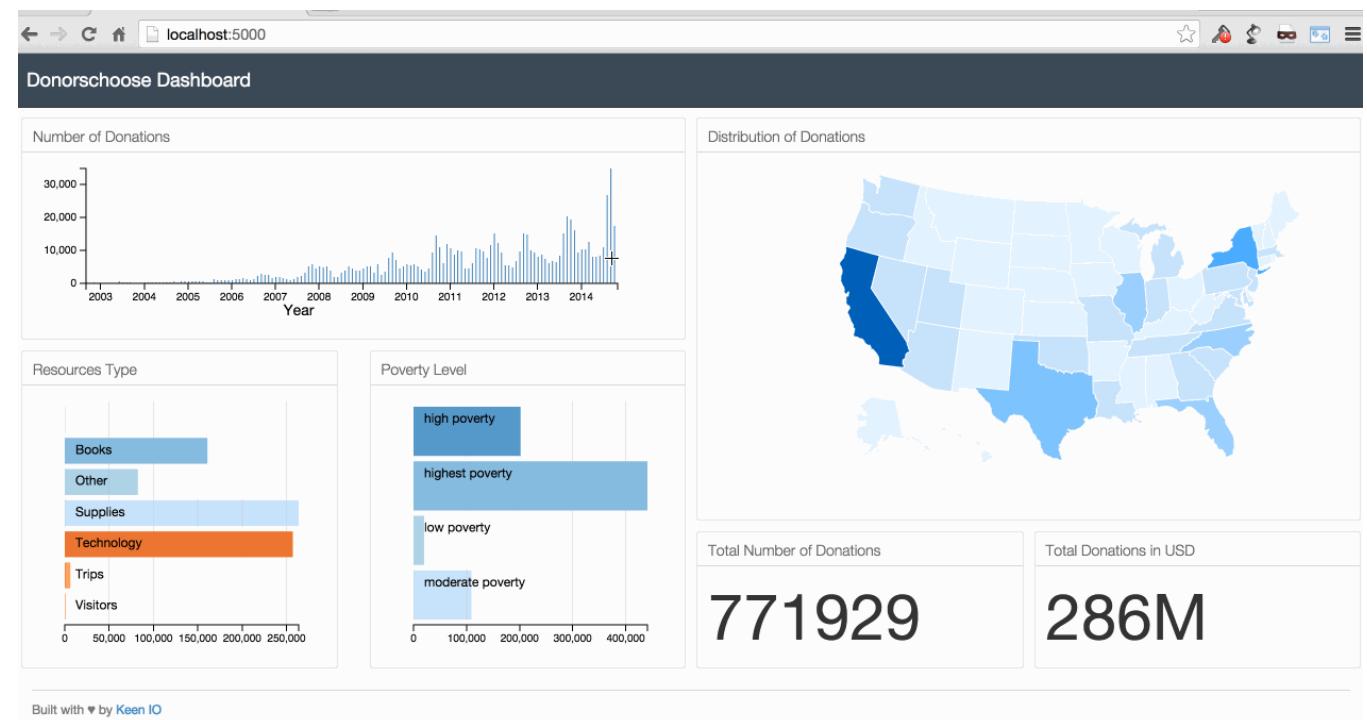
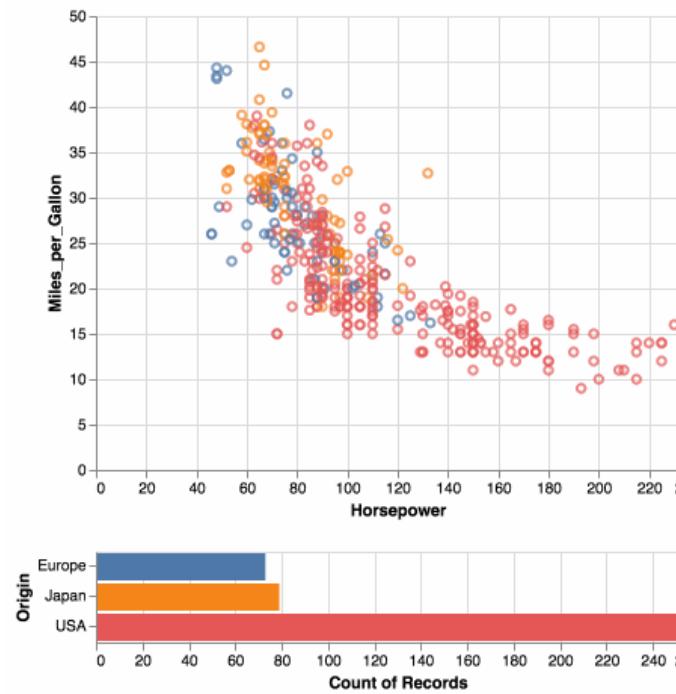
3:36pm

SLOW MEDIUM **FAST**

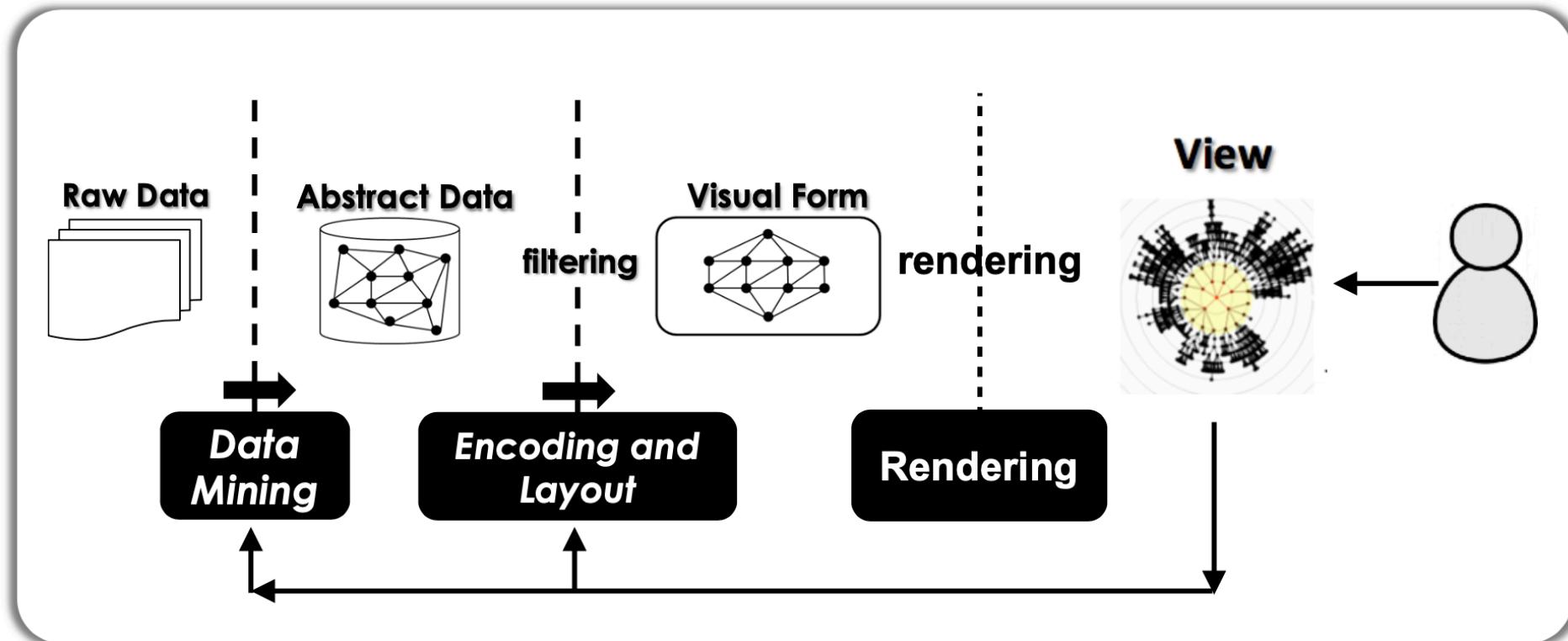
Coffee break? Again, at the top of the hour, you see a shift in activity.



# Interactive Data Visualization



# Data Visualization Workflow



# Data Visualization Tools

- Tableau: an interactive data visualization software.
  - Check more examples on [Tableau Gallery](#).
- D3.js: a JavaScript library for producing dynamic, interactive data visualizations in web browsers.
  - Check more examples on [the official site](#).
- Python with data visualization packages, e.g. matplotlib, seaborn, plotly, ggplot and altair.
  - For network, we can use iGraph and Networkx packages.

# Conclusion

- The same data can be represented in a variety of forms.
- It depends on the focus of an analyst's view point.

# Conclusion

Making data into art  
Information is beautiful

# Conclusion

After this lecture, you should know:

- Why do we need data visualization.
- What are the basic data visualization charts.
- Which chart should be adopted for different cases.

# Assignment 5

- Assignment 5 is released. The deadline is **18:00, 13th July.**

# Thank you!

- Any question?
- Don't hesitate to send email to me for asking questions and discussion. ☺