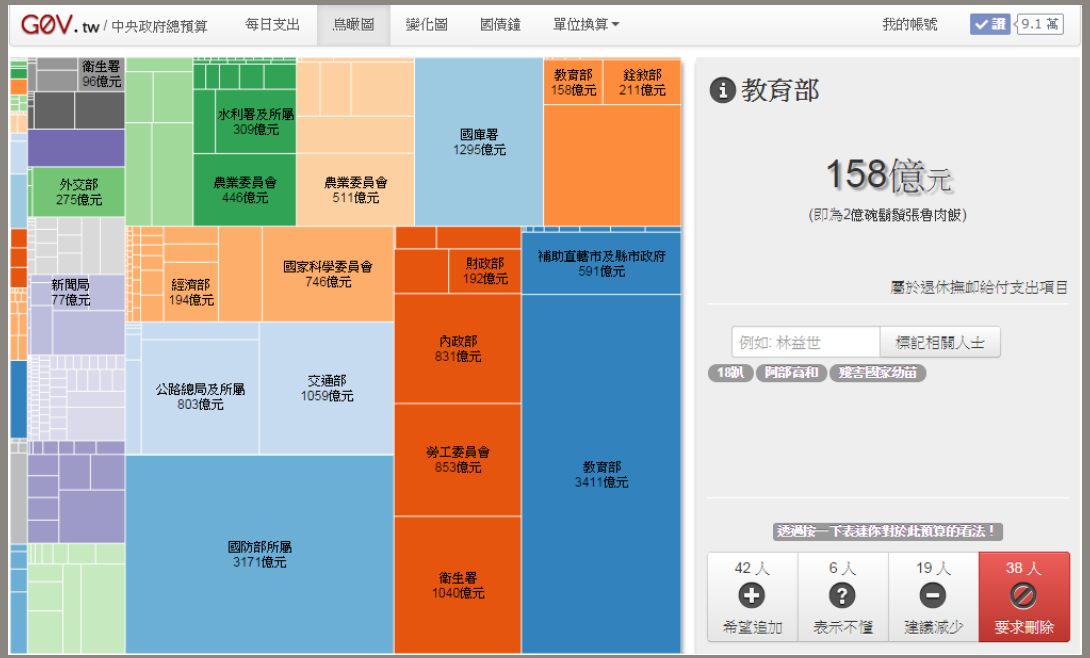
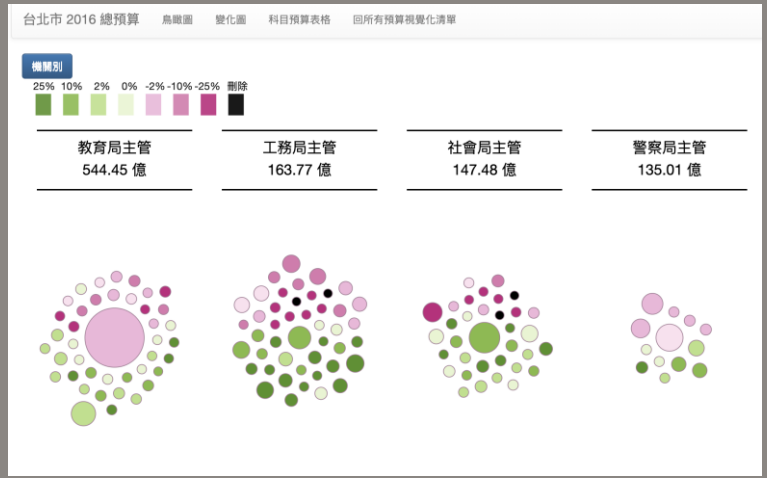
Introduction

預算視覺化

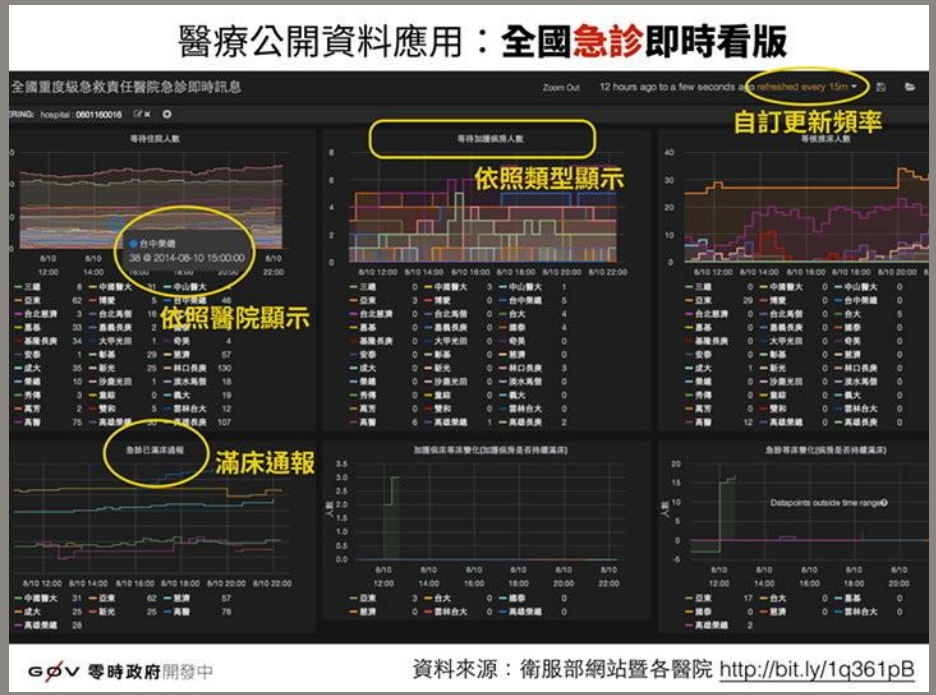
- http://budget.tonyq.org/drilldown/1





全國重度級急求責任醫院急診即時資訊儀表板 (Dashboard)

- https://er.mohw.g0v.tw/#/dashboard/file/all.json

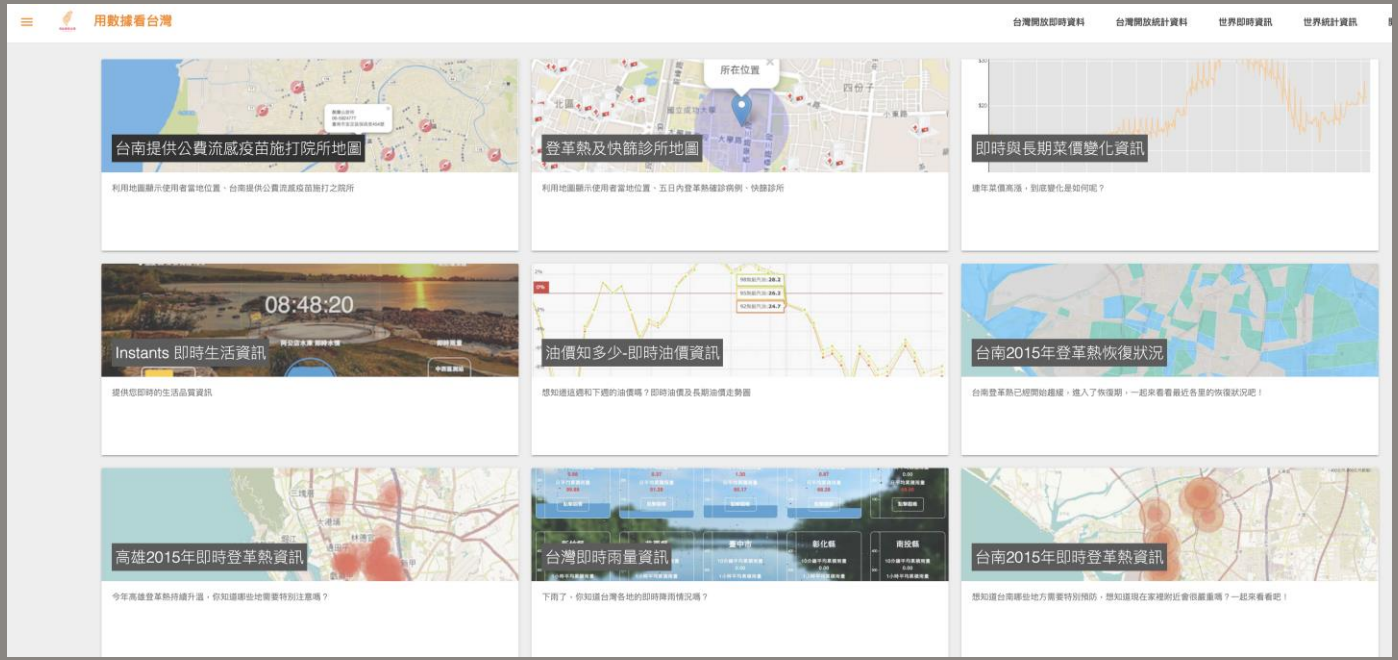


政治獻金視覺化

- http://fuyei.github.io/cf-viz/viz.htm

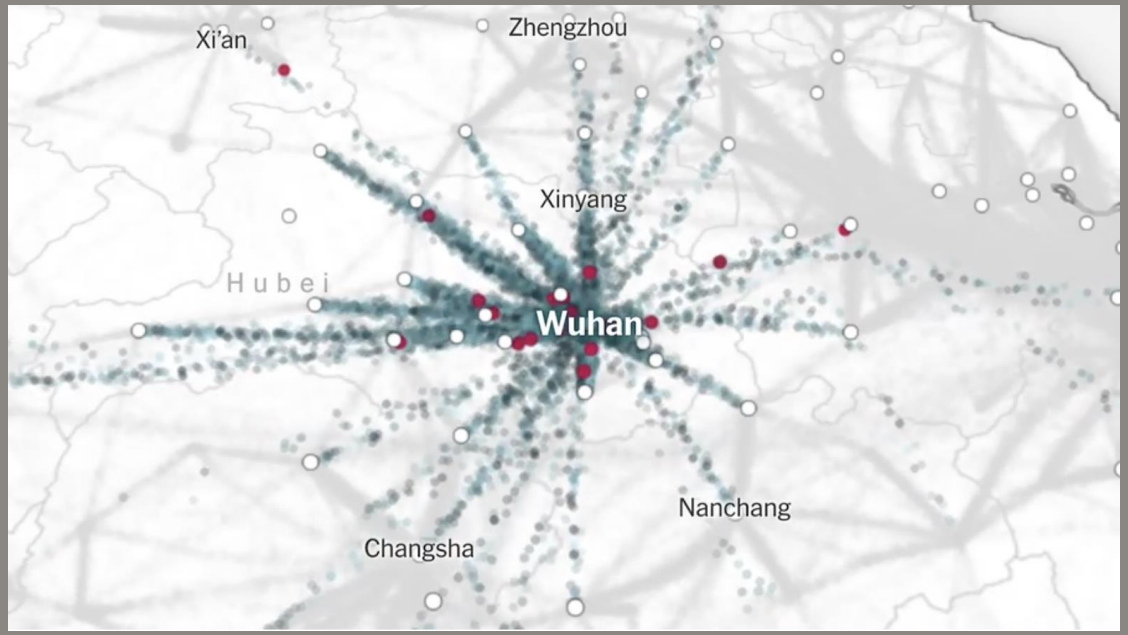
數據看台灣

- https://www.taiwanstat.com/



The New York Times

- https://www.nytimes.com/interactive/2020/12/30/us/2020-year-in-graphics.html (good at saying story with interactive visualization)

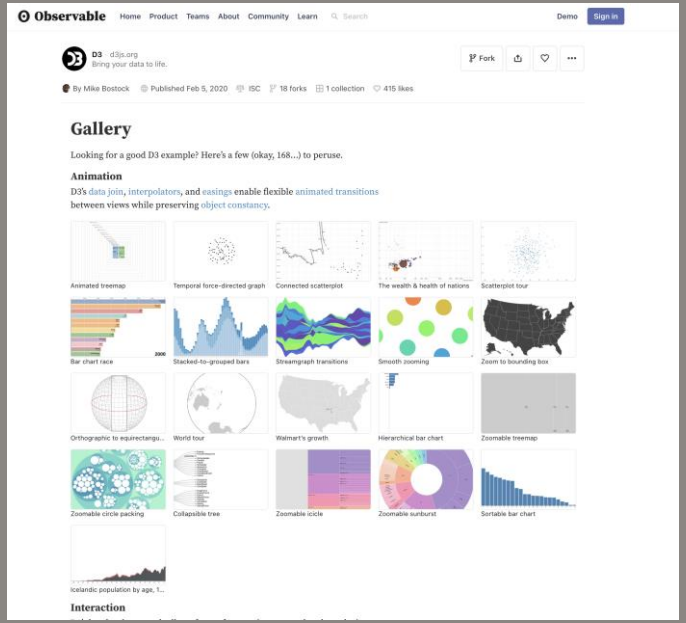


Omni.Sci

- https://www.omnisci.com/demos
  - Dashboard/analysis/interactive visualization

D3/Visualization Demonstration

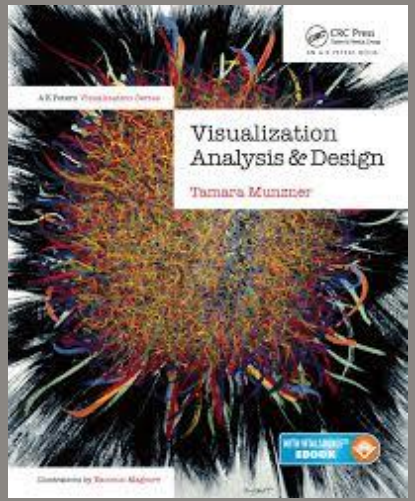
- https://observablehq.com/@d3/gallery
  - Technique demonstration



S01-01

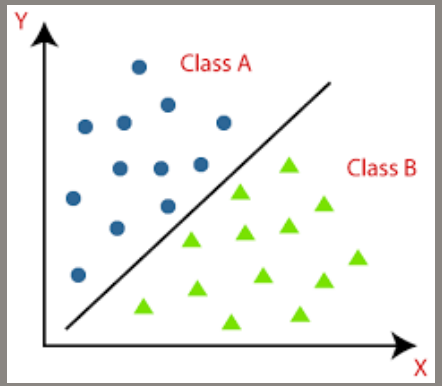
Defining Visualization (vis)

- Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.
- by



Visualization (vis) Definition & Motivation

- Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.
- Visualization is suitable when there is a need to augment human capabilities rather than replace people with computational decision-making methods
- Human in the loop needs the details & no trusted automatic solution exists
  - Does not know exactly what question to ask in advance



Machine learning: you may need to know we have class A and B in the dataset and our task it to separate A from B

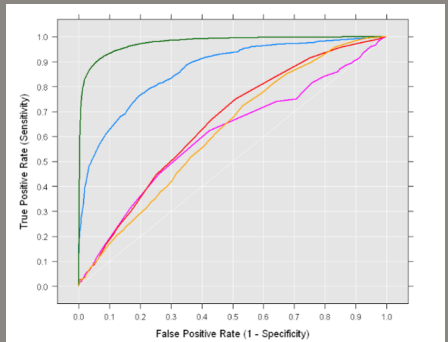
We believe twitter data is useful. We also can define some questions that can be answered by automatic solution.
The dataset might have more value and we even do not know what it is and where it is.

Visualization (vis) Definition & Motivation

- Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.
- Visualization is suitable when there is a need to augment human capabilities rather than replace people with computational decision-making methods
- Human in the loop needs the details & no trusted automatic solution exists
  - Does not know exactly what question to ask in advance
  - Exploratory data analysis (speed-up through human in-the-loop data analysis)
    - You may not have sufficient understanding about the dataset
    - You might want to ask more than 1 question
    - You can do it without vis if you have a lot of time. However, you don't.

Visualization (vis) Definition & Motivation

- Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.
- Visualization is suitable when there is a need to augment human capabilities rather than replace people with computational decision-making methods
- Human in the loop needs the details & no trusted automatic solution exists
  - Does not know exactly what question to ask in advance
  - Exploratory data analysis (speed-up through human in-the-loop data analysis)
  - Present known results to others
    - You are presenting result to others. Of course, human in the loop.



Visualization (vis) Definition & Motivation

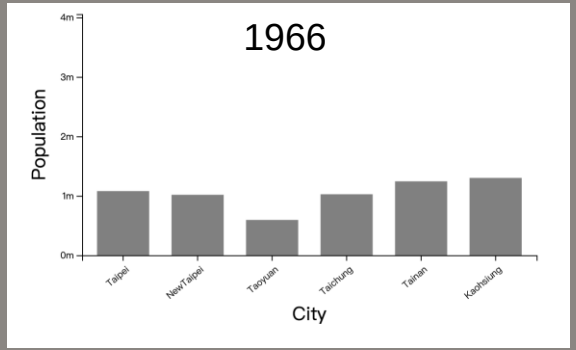
- Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.
- Visualization is suitable when there is a need to augment human capabilities rather than replace people with computational decision-making methods
- Human in the loop needs the details & no trusted automatic solution exists
  - Does not know exactly what question to ask in advance
  - Exploratory data analysis (speed-up through human in-the-loop data analysis)
  - Present known results to others
  - Stepping stone towards automation
    - Before model creation to provide understanding
    - During algorithm creation to refine, debug, set parameters
    - Before or during deployment to build trust and monitor

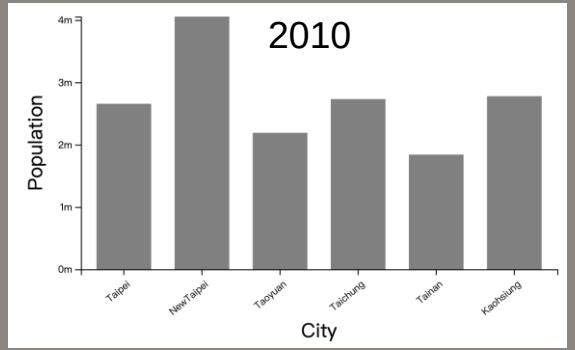
Why use an external representation?

- Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.
  - External representation: replace cognition with perception
  - Free cognitive resource to do some thing more interesting
  - Visualization is not to limit your thinking. Visualization is to save your cognitive resource to think something at higher level.

Population

	A	B	C	D	E	F	G
1	City	1956	1966	1980	1990	2000	2010
2	Taipei	737029	1202952	2267584	2760475	2624257	2655515
3	NewTaipei	668093	1143288	2364521	3065779	3722082	4054467
4	Taoyuan	411575	660293	1067951	1377934	1808833	2190342
5	Taichung	784475	1113291	1634820	2057857	2499527	2731056
6	Tainan	992411	1333735	1541402	1695110	1846379	1840257
7	Kaohsiung	899828	1445669	2225021	2512858	2756775	2777384





People usually deal with much larger and complicated datasets

Why depend on vision?

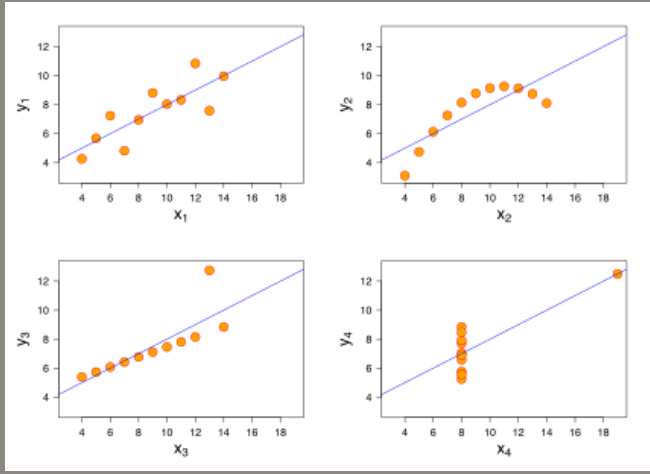
- Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

- Human visual system is high-bandwidth channel to brain
  - Overview possible due to background processing
    - Subjective experience of seeing everything simultaneously
    - Significant processing occurs in parallel and pre-attentively
- Sound: lower bandwidth and different semantics
  - Overview not supported
    - Subjective experience of sequence stream
- Touch/haptics: impoverished record/replay capacity
  - Only very low-bandwidth communication thus far
- Taste, smell: no viable record/replay devices

Why represent all the data?

- Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.
- Why are we representing all of the data? (why not just show a portion of the data)
  - Summaries lose information, details matter
    - Confirm expected and find unexpected pattern
    - Assess validity of statistical model

Property	Value
Mean of x	9
Sample variance of x : s_x^2	11
Mean of y	7.50
Sample variance of y : s_y^2	4.125
Correlation between x and y	0.816
Linear regression line	$y = 3.00 + 0.500x$
Coefficient of determination of the linear regression : R^2	0.67



Things might be even worse if your data is bigger

Why represent all the data?

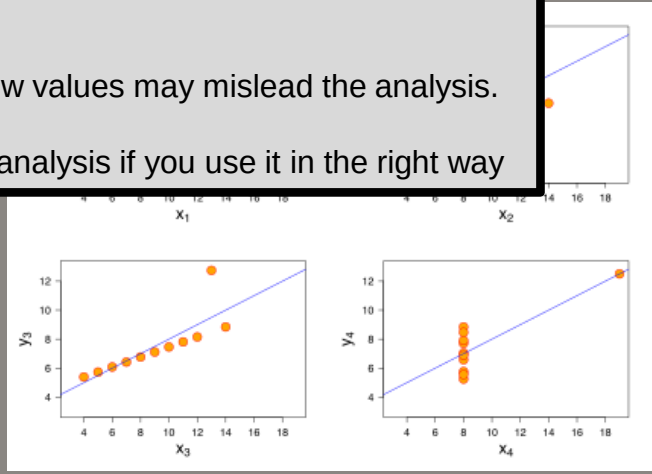
- Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.
- Why are we representing all of the data? (why not just show a portion of the data)

I do not say that never use statistics to visualize or analyze the data.

I say that understating the data ONLY from a portion of the dataset or a few values may mislead the analysis.

However, statistical values is powerful in the loop of data visualization or analysis if you use it in the right way

Sample variance of x : s_x^2	11
Mean of y	7.50
Sample variance of y : s_y^2	4.125
Correlation between x and y	0.816
Linear regression line	$y = 3.00 + 0.500x$
Coefficient of determination of the linear regression : R^2	0.67



Things might be even worse if your data is bigger

Why represent all the data?

- Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.
- Why are we representing all of the data? (why not just show a portion of the data)
  - Summaries lose information, details matter
    - Confirm expected and find unexpected pattern
    - Assess validity of statistical model

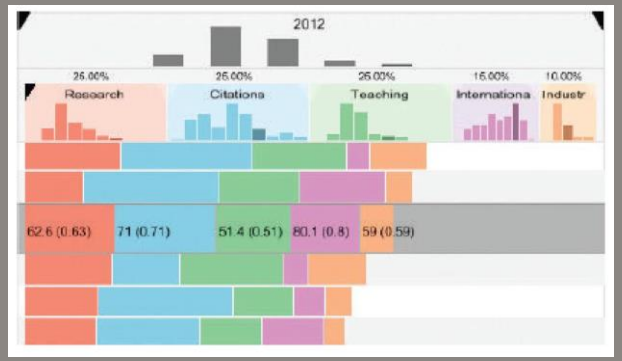
Same state, different graphs, CHI2017
https://www.youtube.com/watch?v=DbJyPELmhJc&ab_channel=AutodeskResearch

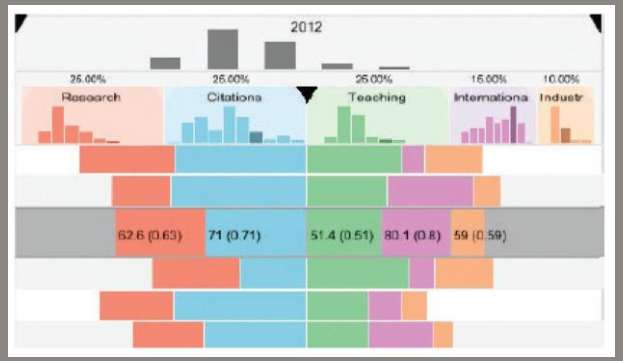
Why focus on tasks and effectiveness?

- Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

- Effectiveness requires match between data/task and representation
  - Set of representation is huge
  - Many are ineffective mismatch for specific data/task combo
  - Increases chance of finding good solutions if you understand full space of possibilities

The best research/teaching university? best the university?





Why focus on tasks and effectiveness?

- Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

- Effectiveness requires match between data/task and representation
  - Set of representation is huge
  - Many are ineffective mismatch for specific data/task combo
  - Increases chance of finding good solutions if you understand full space of possibilities
- What counts as effective?
  - Novel: enable entirely new kinds of analysis
  - *Faster: speed up existing workflows*

Why focus on tasks and effectiveness?

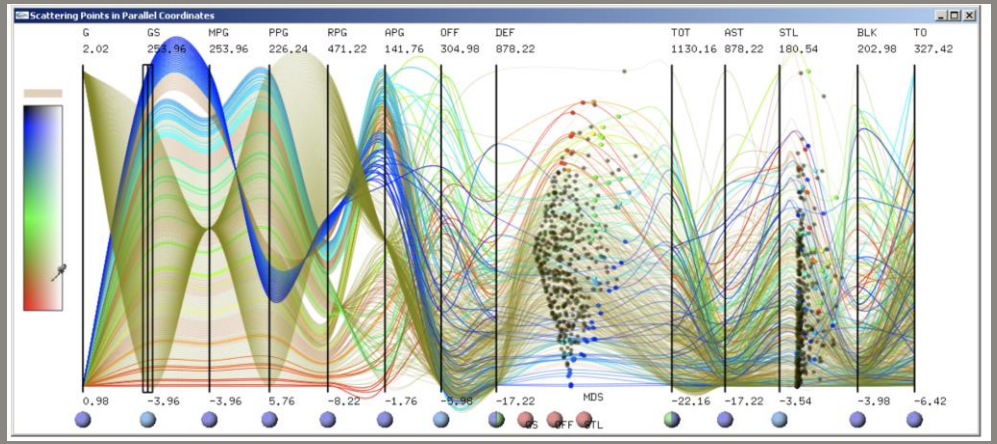
- Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

- Effectiveness requires match between data/task and representation
  - Set of representation is huge
  - Many are ineffective mismatch for specific data/task combo
  - Increases chance of finding good solutions if you understand full space of possibilities
- What counts as effective?
  - Novel: enable entirely new kinds of analysis
  - *Faster: speed up existing workflows*
- How to validate effectiveness
  - Many aspects, so many methods, must pick appropriate one for your context

What resources limitations are we faced with?

- Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

- Computational limits
  - Processing time
  - System memory



What resources limitations are we faced with?

- Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

- Computational limits
  - Processing time
  - System memory
- Human limits
  - Human attention and memory
  - How human attend to something? How much information human can remember?

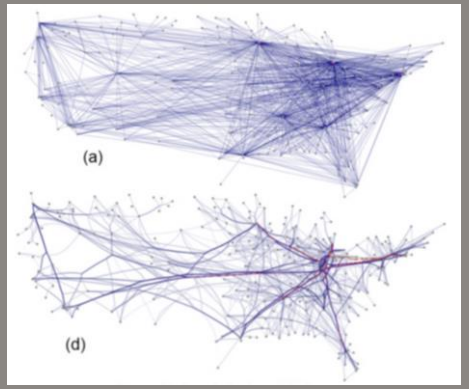
What resources limitations are we faced with?

- Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

- Computational limits
  - Processing time
  - System memory
- Human limits
  - Human attention and memory
- Display limits
  - Pixels are precious resource, the most constrained resource
    - It is not hard to have 25millions data points. But we do not have 25 million pixel on every display.

What resources limitations are we faced with?

- Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.



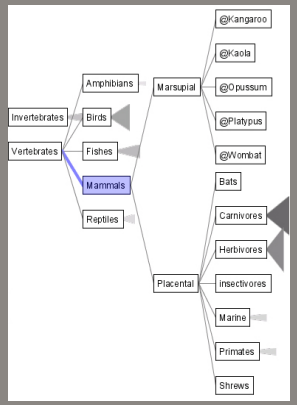
- Computational limits
  - Processing time
  - System memory
- Human limits
  - Human attention and memory
- Display limits
  - Pixels are precious resource, the most constrained resource
    - It is not hard to have 25millions data points. But we do not have 25 million pixel on every display.
  - Information density: ratio of space used to encode info vs unused white space
    - Tradeoff between clutter and wasting space, find sweet spot between dense and sparse

S01-02

Why analyze?

- Imposes structure on huge design space
  - Scaffold to help you think systematically about choices
  - Analyzing existing as stepping stone to designing new
  - Most possibilities ineffective for particular task/data combination

SpaceTree



TreeJuxtaposer