

Week 07 • 데이터 저널리즘

Visualizing Data

Joonhwan Lee

human-computer interaction + design lab.

오늘 다룰 내용

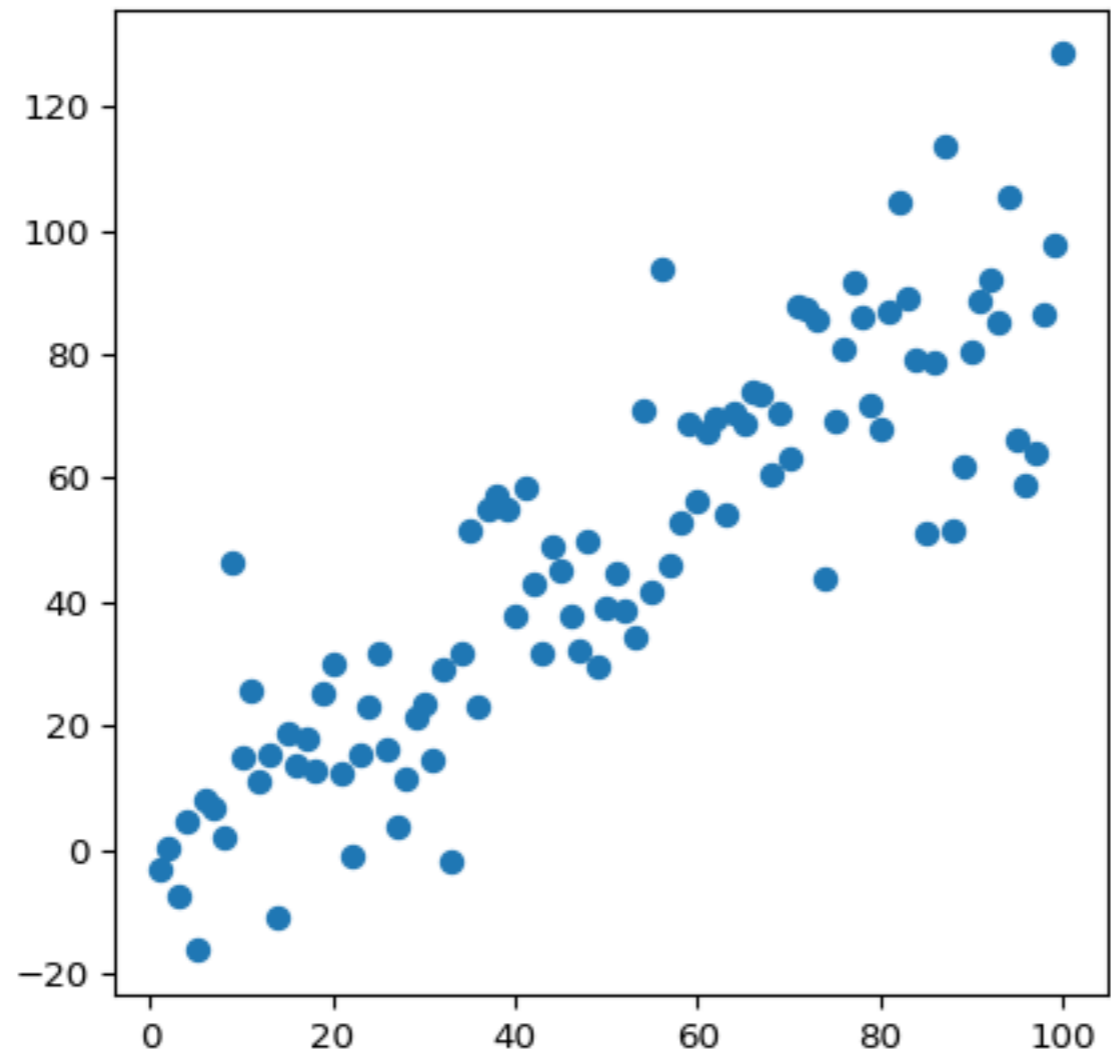
- Data Visualization

Visualizing Data

Data Visualization

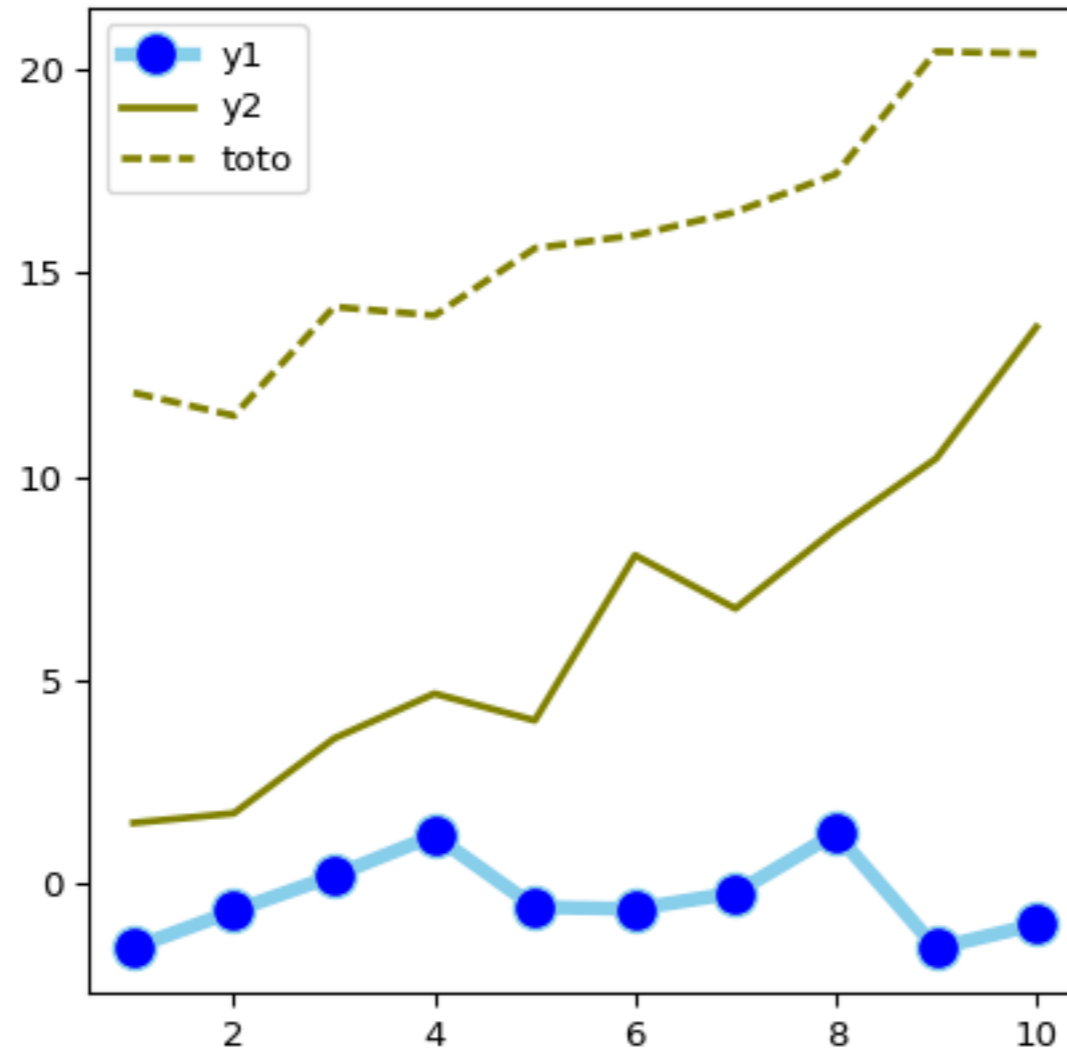
- ◆ 데이터 분석
 - ◆ Correlation
 - ◆ Pattern
 - ◆ Compare
 - ◆ Connection

Correlation



Scatterplot

Pattern



Line Graph

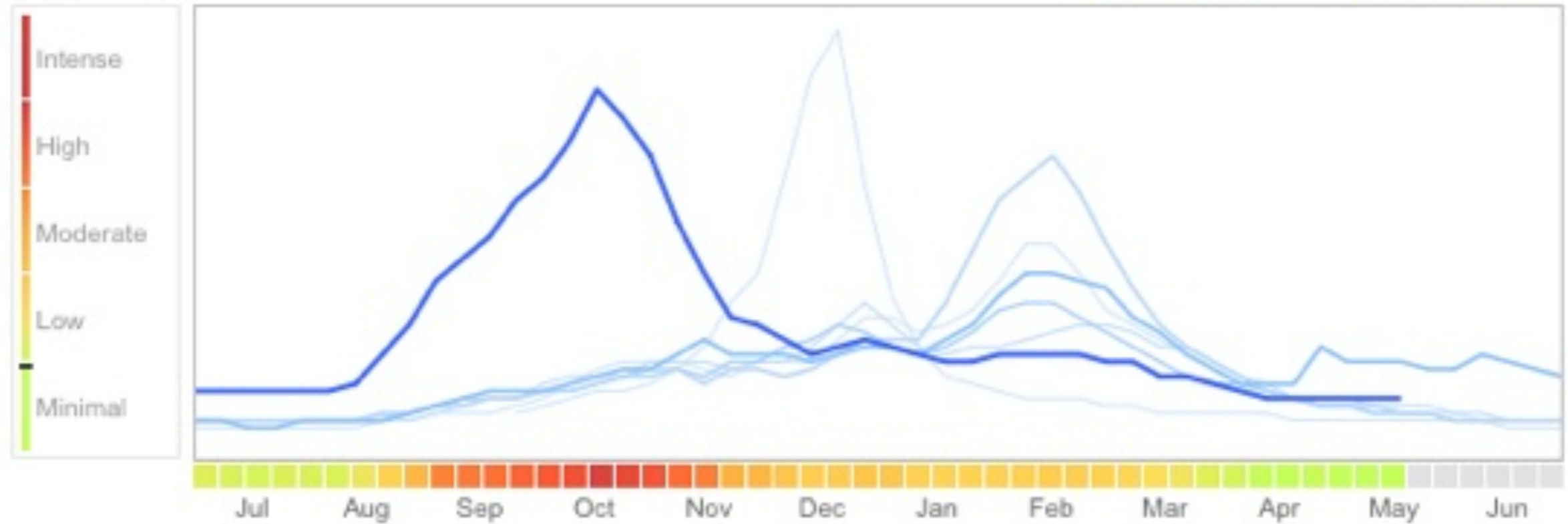
Pattern

Explore flu trends - United States

We've found that certain search terms are good indicators of flu activity. Google Flu Trends uses aggregated Google search data to estimate flu activity. [Learn more »](#)

National

● 2009-2010 ● [Past years ▼](#)



11 Ways to Visualize Changes Over Time

by Nathan Yau

◆ Line Graph

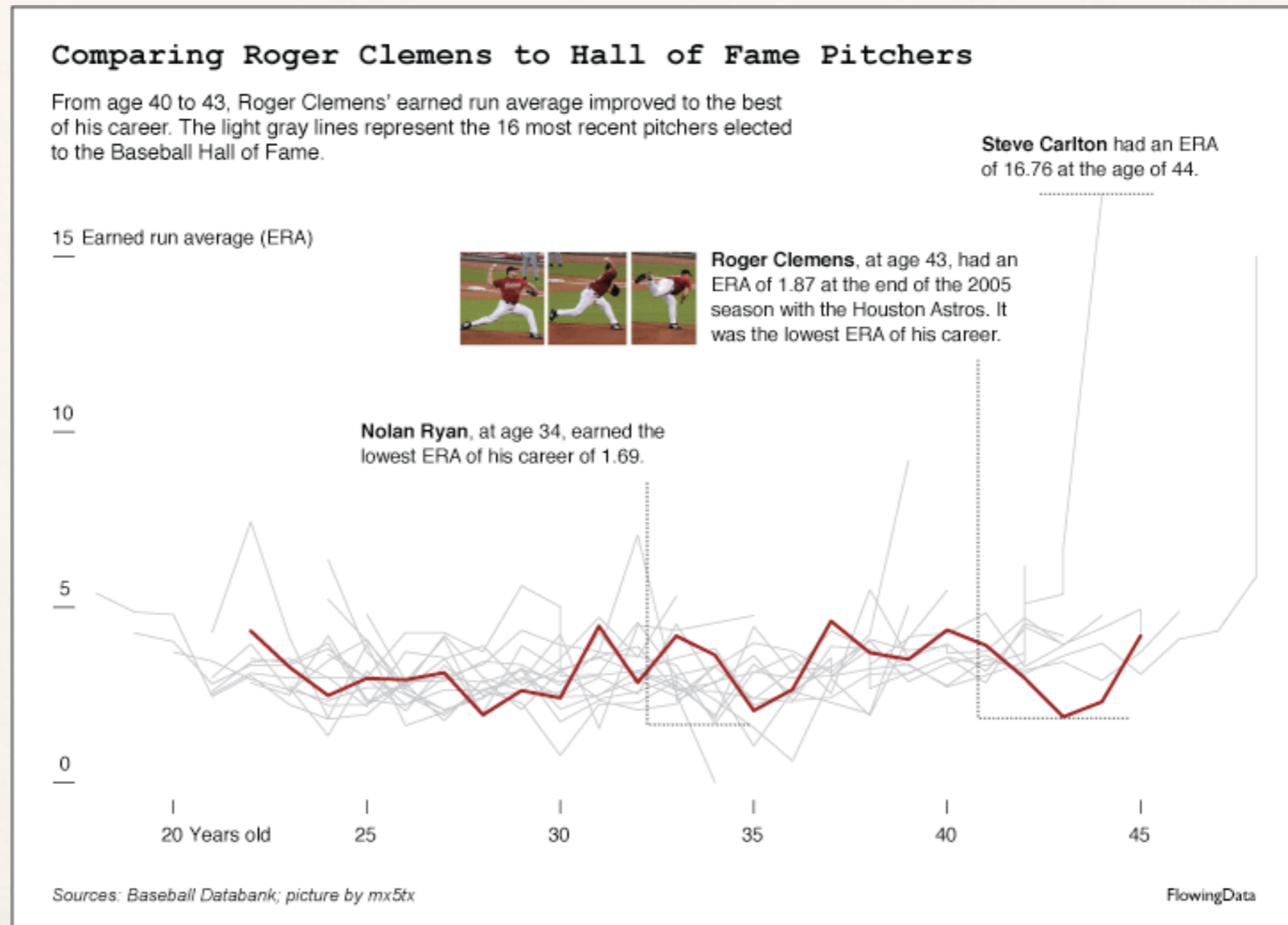


- ◆ 대부분의 time series data를 표현할 수 있음.
- ◆ 하나의 그래프에 여러 개의 series 를 그려서 비교할 수 있음.
- ◆ data point는 square, circle 등으로 표현.

11 Ways to Visualize Changes Over Time

by Nathan Yau

◆ Line Graph



<http://flowingdata.com/2008/02/11/comparing-roger-clemens-to-hall-of-fame-pitchers/roger-clemens-compared-to-hall-of-fame-pitchers/>

11 Ways to Visualize Changes Over Time

by Nathan Yau

✦ Scatter Plots

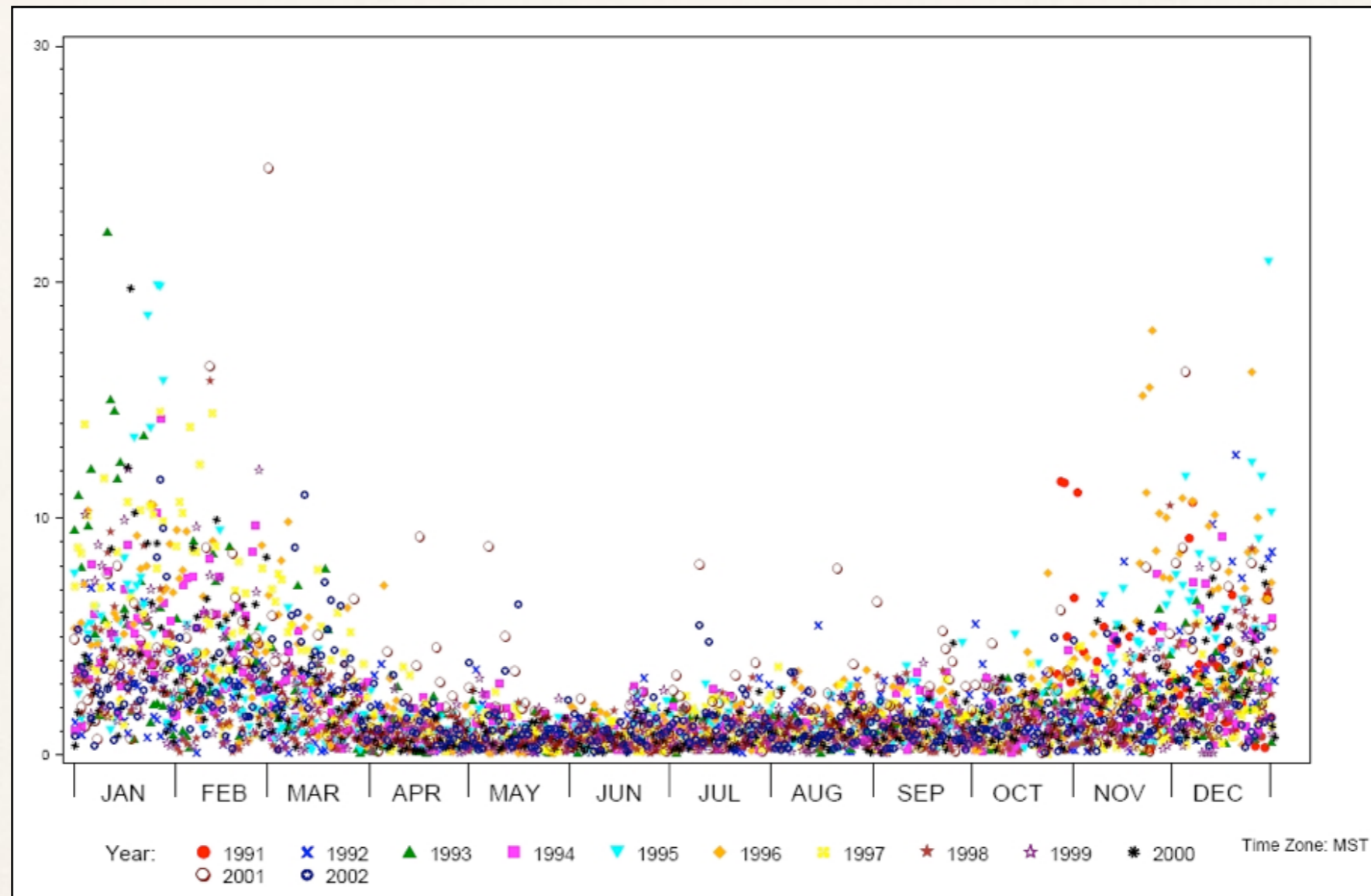


- ✦ 데이터포인트가 매우 많을 때 적합. 포인트가 적을 경우 화면에서 의미있는 패턴을 만들어 보이기가 힘들다 (gestalt theory)
- ✦ Line graph의 경우 데이터포인트가 일정한 간격으로 측정되는 등, 규칙이 있을 때 적합하지만, scatter plot은 불규칙하게 구조화된 데이터에 적합.

11 Ways to Visualize Changes Over Time

by Nathan Yau

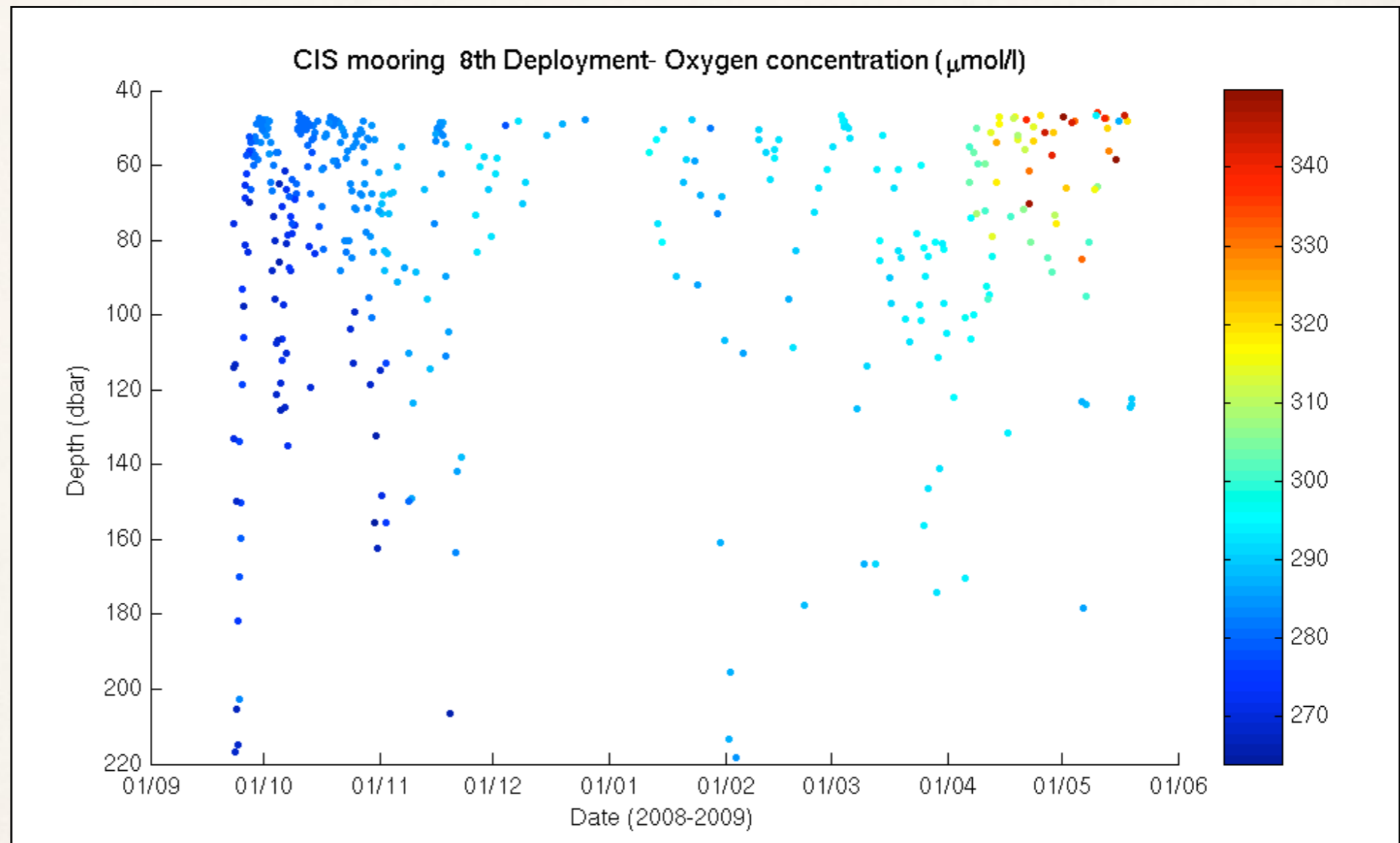
◆ Scatter Plots



11 Ways to Visualize Changes Over Time

by Nathan Yau

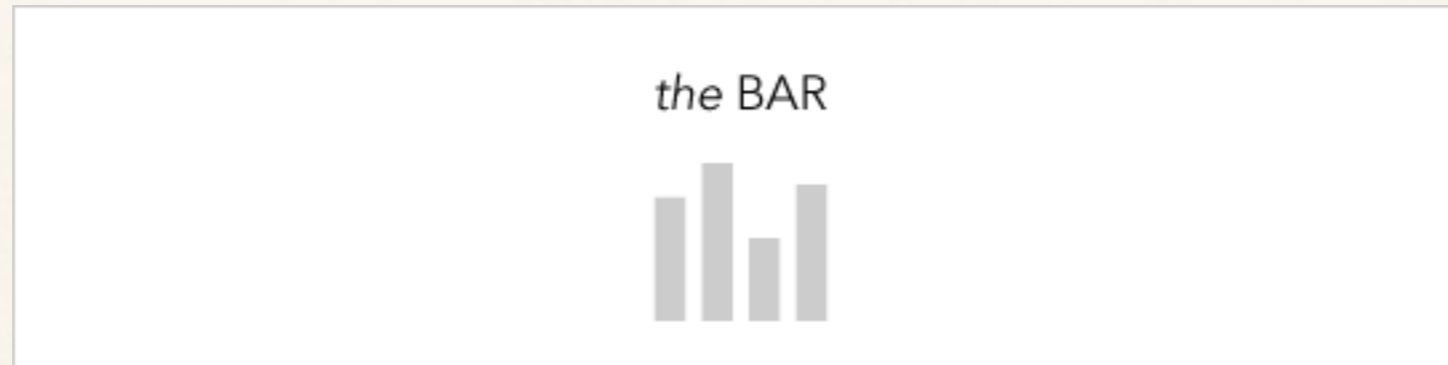
✦ Scatter Plots



11 Ways to Visualize Changes Over Time

by Nathan Yau

✦ Bar Charts

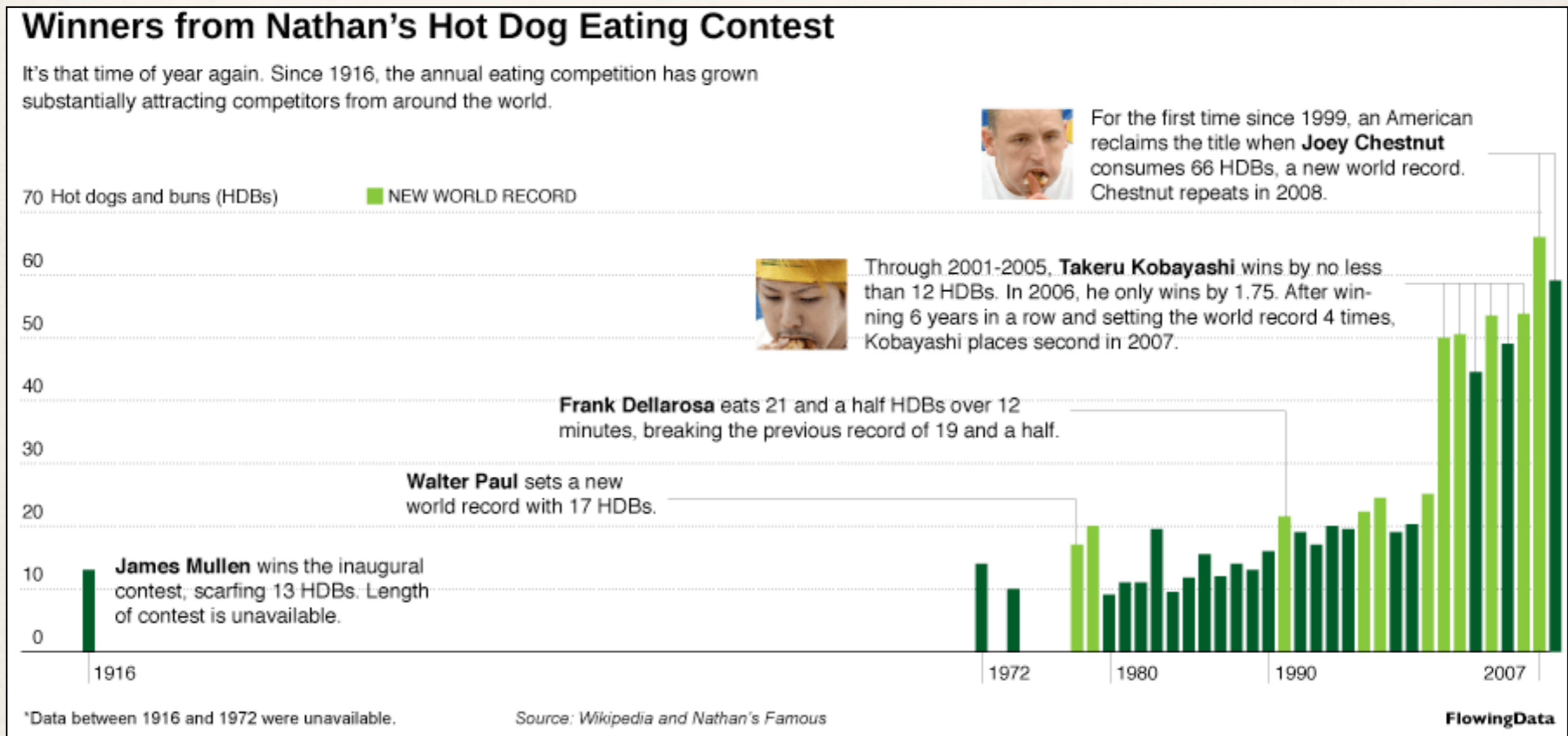


- ✦ 시간에 따라 명확하게 구분되는 데이터포인트를 사용할 때 적합한 표현 방법. (discrete vs. continuous)
- ✦ 데이터포인트가 시공간내에서 일정하게 분포되어 있을 때 사용.

11 Ways to Visualize Changes Over Time

by Nathan Yau

♦ Bar Charts

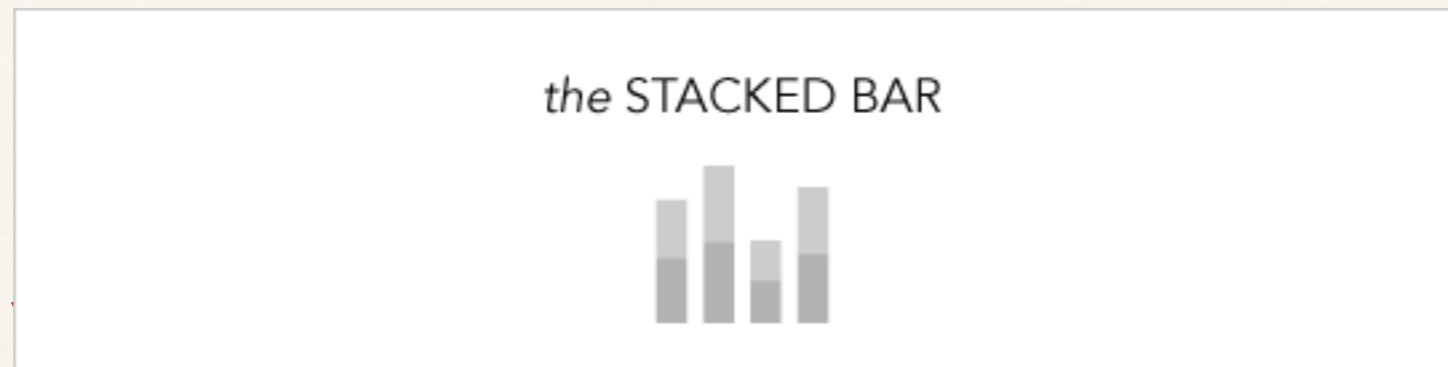


<http://flowingdata.com/2009/07/03/whos-going-to-win-nathans-hot-dog-eating-contest/hot-dogs-2/>

11 Ways to Visualize Changes Over Time

by Nathan Yau

◆ Stacked Bar

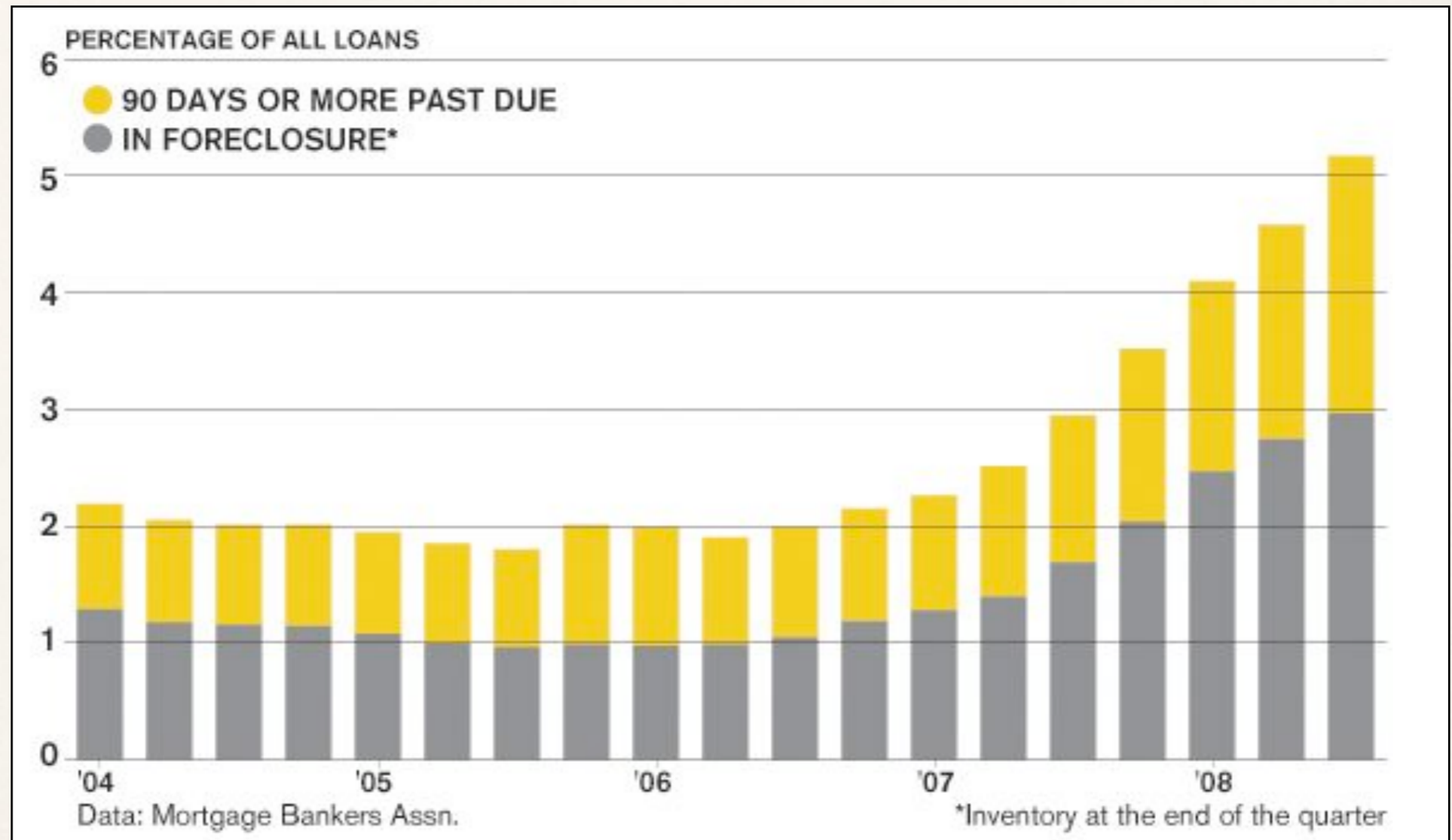


- ◆ Bar Chart와 동일한 경우에 사용하나 카테고리 여러 개 있고, 단순히 카테고리를 서로 비교하는 것이 아니라 모두를 합친 것이 의미를 가질 때 사용함.
- ◆ 만일 각각의 카테고리가 모두 합쳐졌을 때 의미가 없다면 사용하지 않음. → Bar Chart 사용.

11 Ways to Visualize Changes Over Time

by Nathan Yau

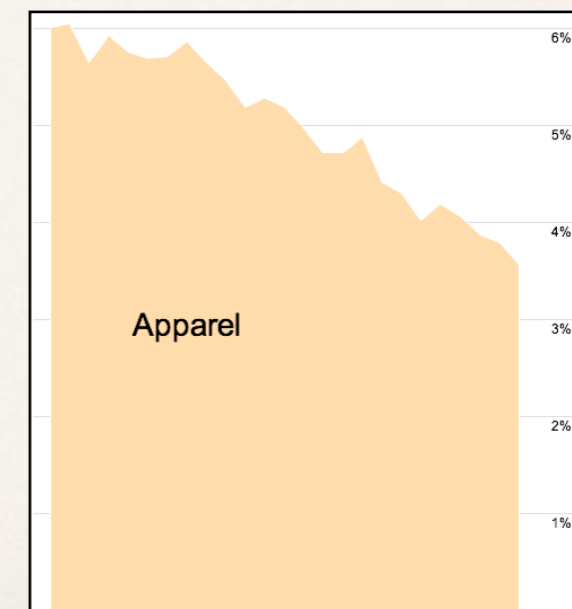
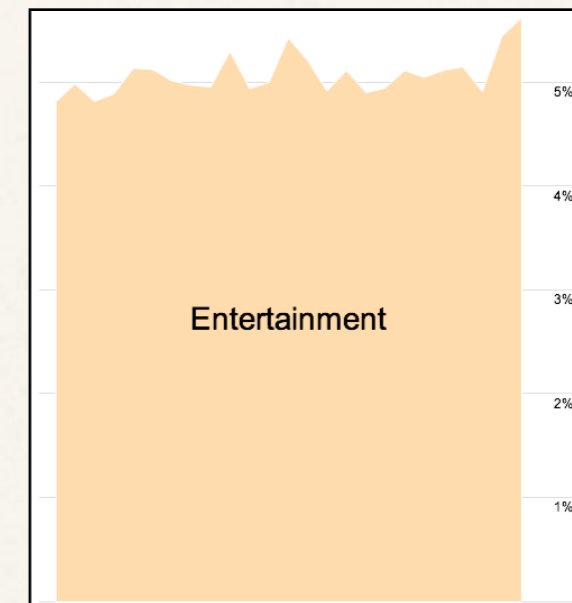
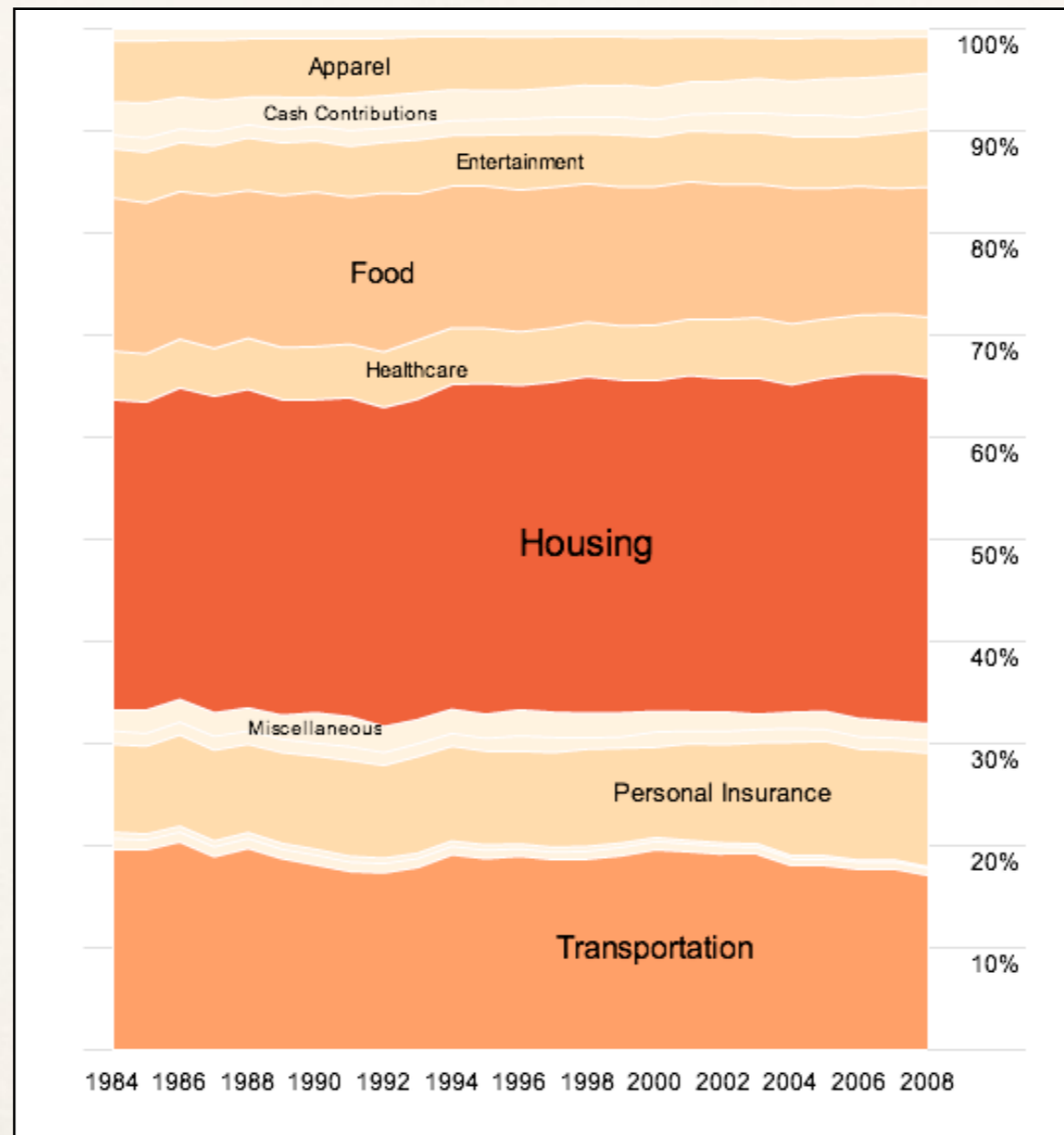
◆ Stacked Bar



11 Ways to Visualize Changes Over Time

by Nathan Yau

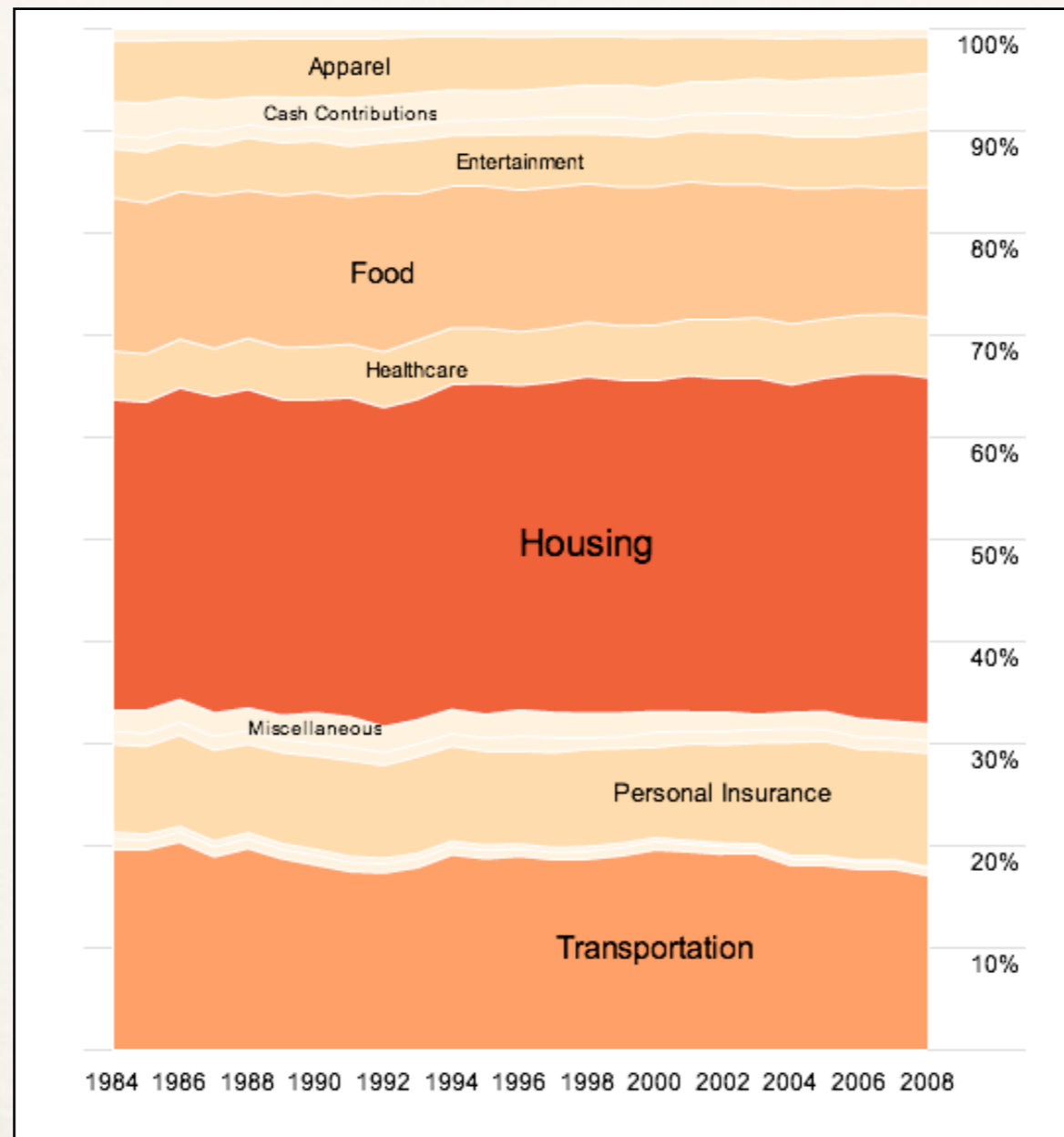
◆ Stacked Area



11 Ways to Visualize Changes Over Time

by Nathan Yau

◆ Stacked Area

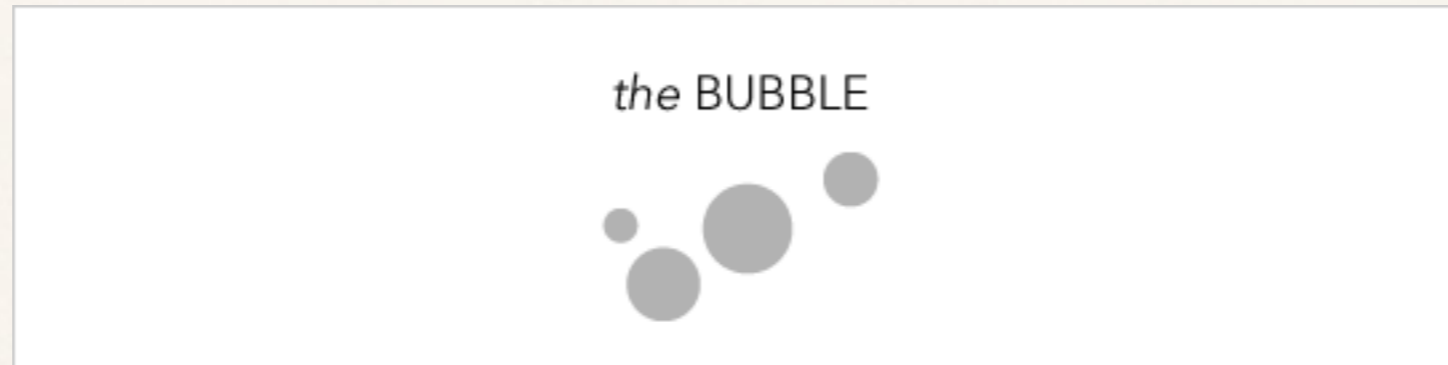


Interactive Demo

11 Ways to Visualize Changes Over Time

by Nathan Yau

✦ Bubble

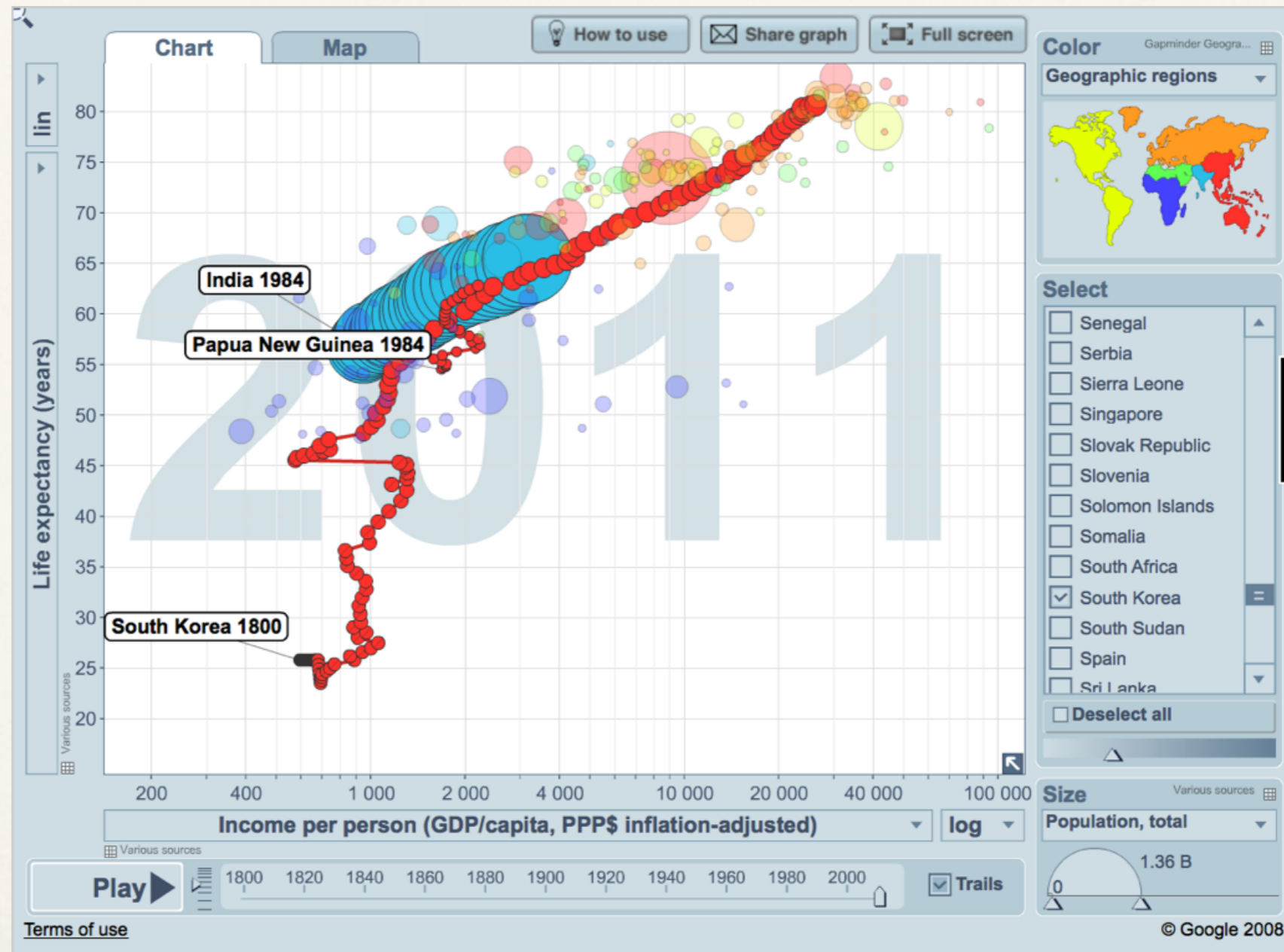


- ✦ Scatter plot과 유사하지만 scatter plot은 데이터포인트를 단순히 점으로 표현하는데 비해 bubble은 원의 크기가 데이터의 어떠한 값을 표현한다 → 시간에 따라 변화하는 두개의 값 (bubble의 x, y 위치, 크기) 을 보여줄 수 있음.
- ✦ Hans Rosling의 TED Talk (Gapminder)

11 Ways to Visualize Changes Over Time

by Nathan Yau

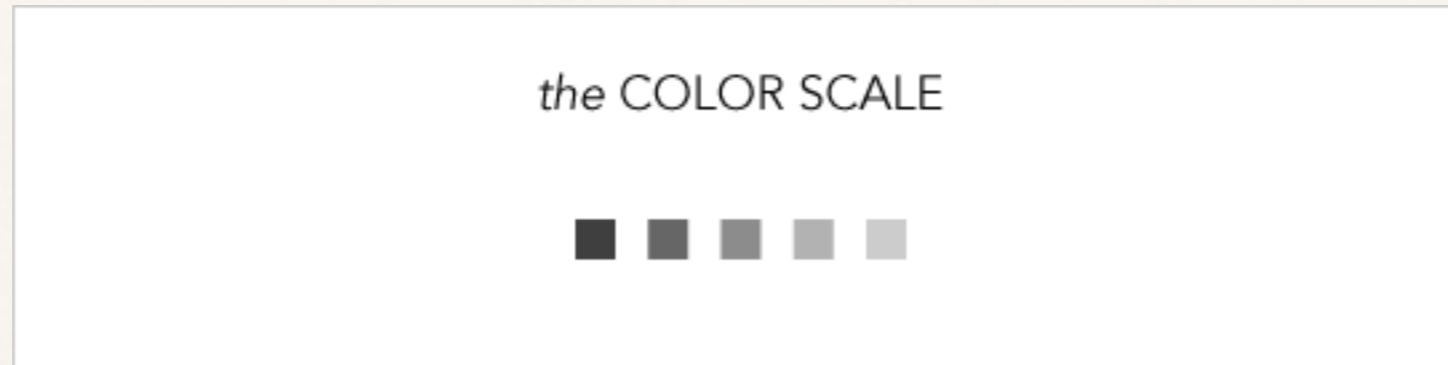
◆ Bubble



11 Ways to Visualize Changes Over Time

by Nathan Yau

◆ Color Scale

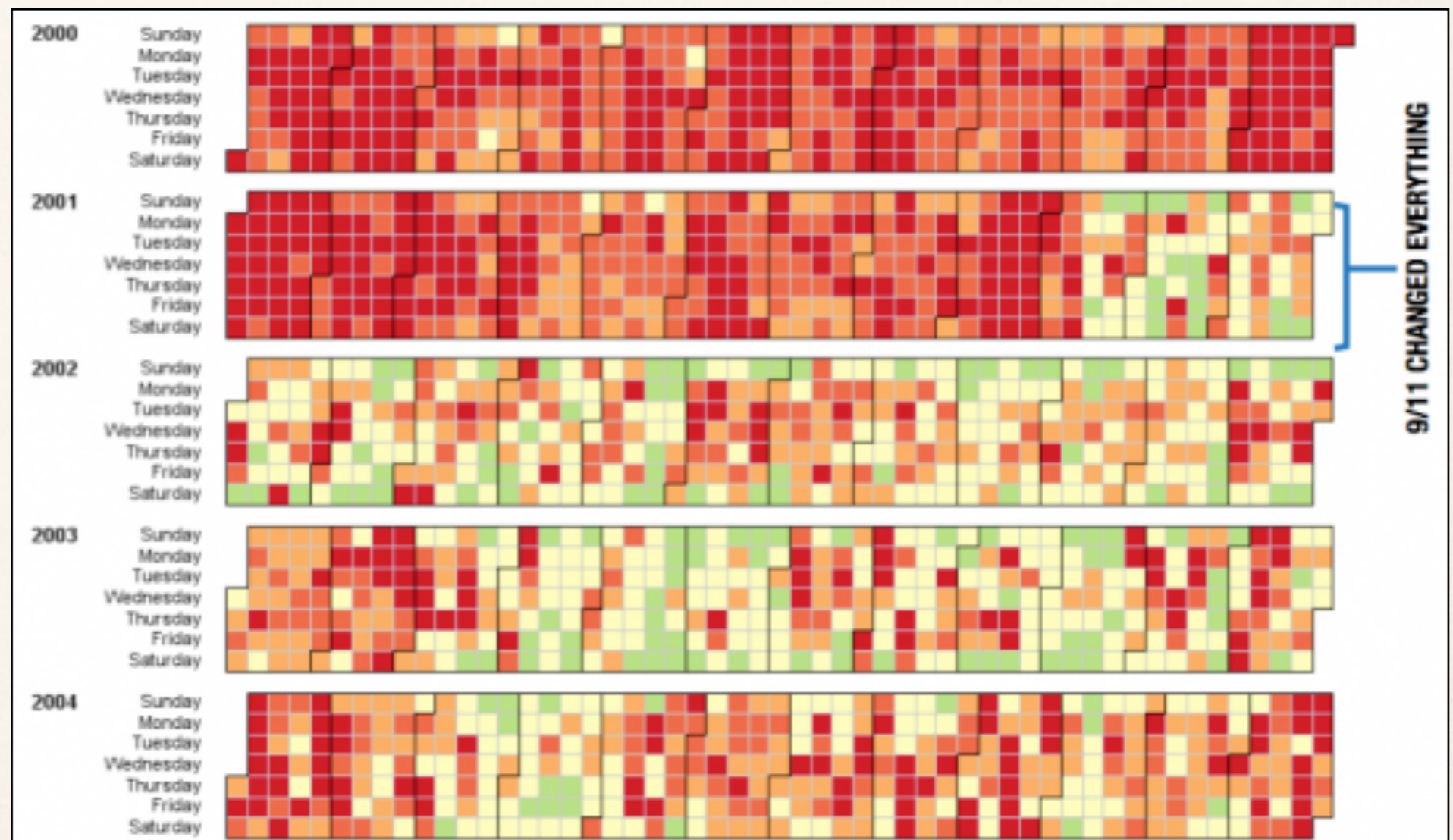


- ◆ 값의 차이를 명확하게 보여주려면 보통 bar chart와 같이 높이의 비교가 효과적이거나 화면이 부족하거나 한 경우에는 칼라를 사용하는 편이 좋다.

11 Ways to Visualize Changes Over Time

by Nathan Yau

✦ Color Scale



<http://flowingdata.com/2009/09/10/3-in-depth-views-of-flight-delays-and-cancellations/>

11 Ways to Visualize Changes Over Time

by Nathan Yau

◆ Timeline



- ◆ 어떤 사건이 언제 일어났는지를 보여주는 그래프.
- ◆ 데이터의 양이 많고, 자세한 내용을 파악하기 힘들 때 타임라인과 병행하여 그래프를 작성하면 특정 사건이 발생한 순간의 데이터를 해석하는데 도움이 됨.

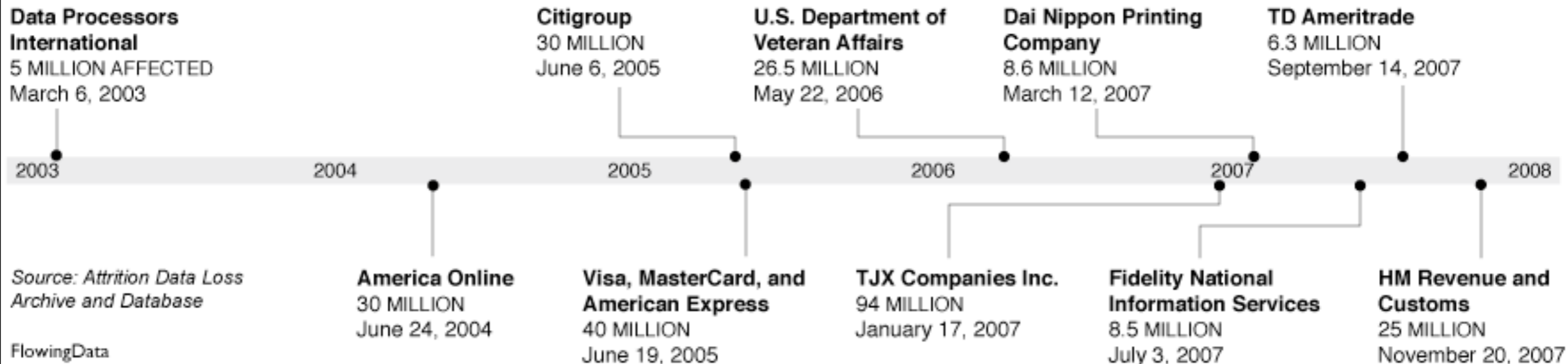
11 Ways to Visualize Changes Over Time

by Nathan Yau

◆ Timeline

10 Largest Data Breaches Since 2000

As more information goes digital, it becomes more important to protect against hackers.



11 Ways to Visualize Changes Over Time

by Nathan Yau

✦ Everything

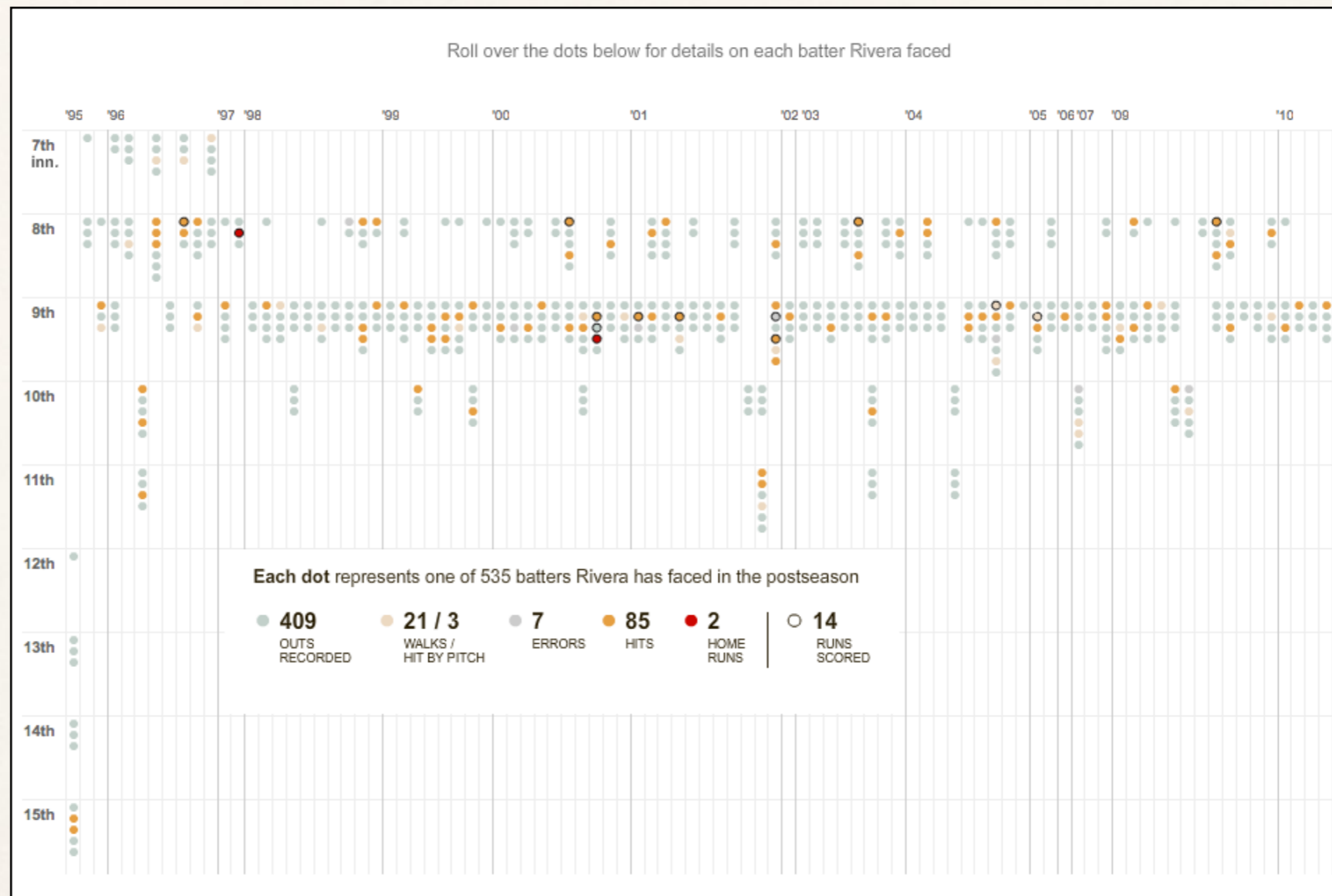


- ✦ 경우에 따라서는 모든 데이터를 보여주는 것이 도움이 되기도.
- ✦ 화면 위에 매우 많은 데이터를 표현해야 하기 때문에 산만하거나 데이터가 주는 의미를 파악하기 어려울 수도 있지만, 칼라코딩 등의 방법을 통해 데이터를 시각적으로 보기 편하게 정리할 수 있음.
- ✦ 인터랙션을 사용하여 자세한 정보(contextual information)를 추가.

11 Ways to Visualize Changes Over Time

by Nathan Yau

◆ Everything

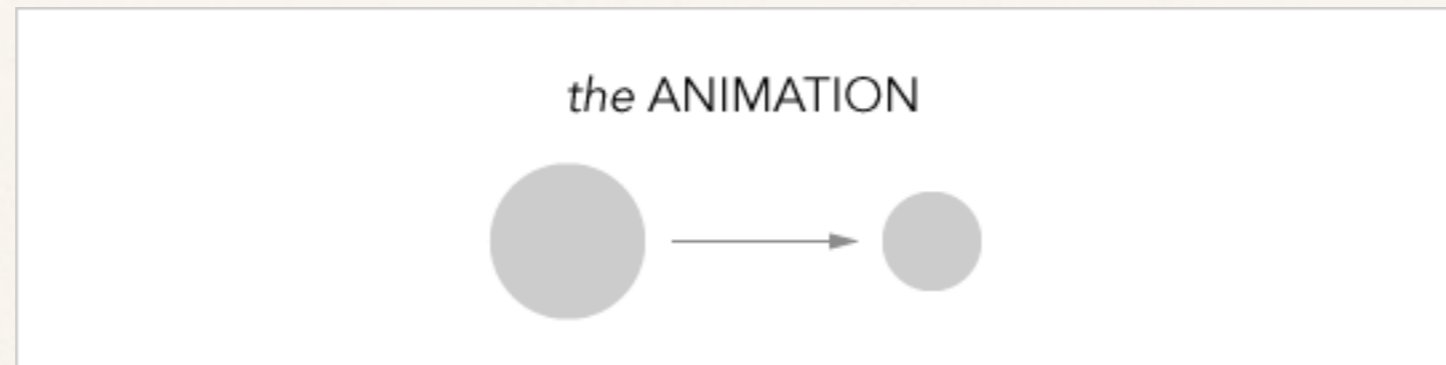


[http://www.nytimes.com/interactive/2009/11/05/sports/baseball/20091105-rivera.html?
r=0](http://www.nytimes.com/interactive/2009/11/05/sports/baseball/20091105-rivera.html?r=0)

11 Ways to Visualize Changes Over Time

by Nathan Yau

✦ Animation

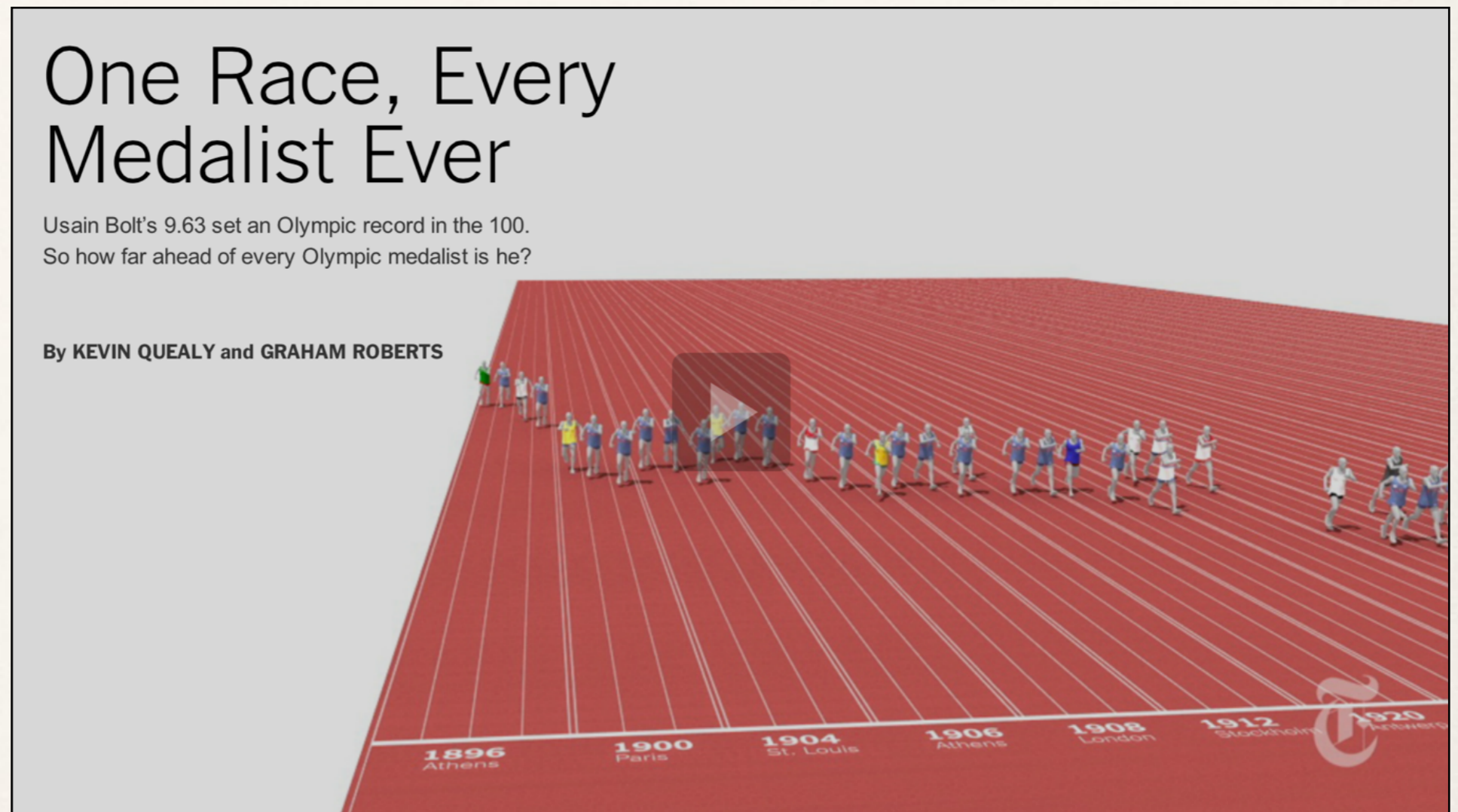


- ✦ 시간의 흐름에 따른 데이터의 변화를 애니메이션으로 표현.

11 Ways to Visualize Changes Over Time

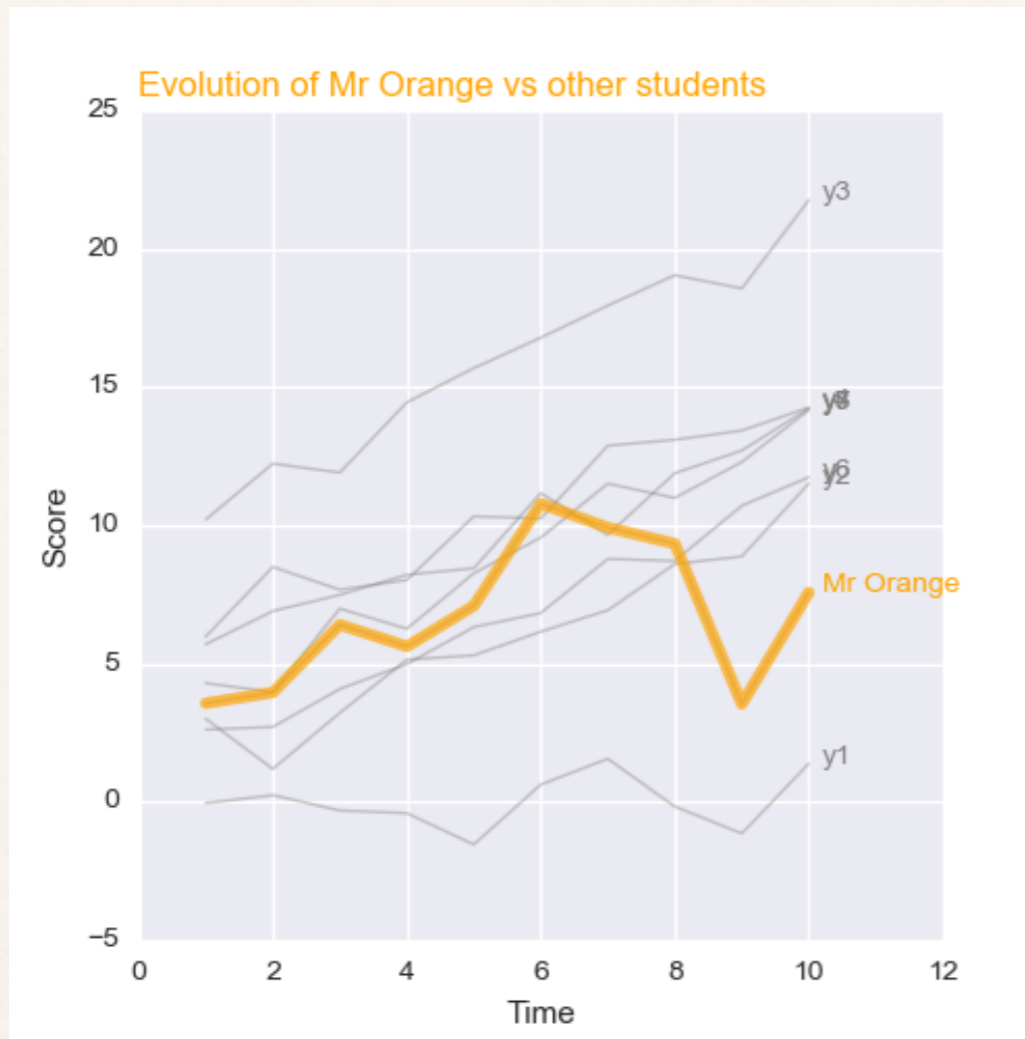
by Nathan Yau

✦ Animation

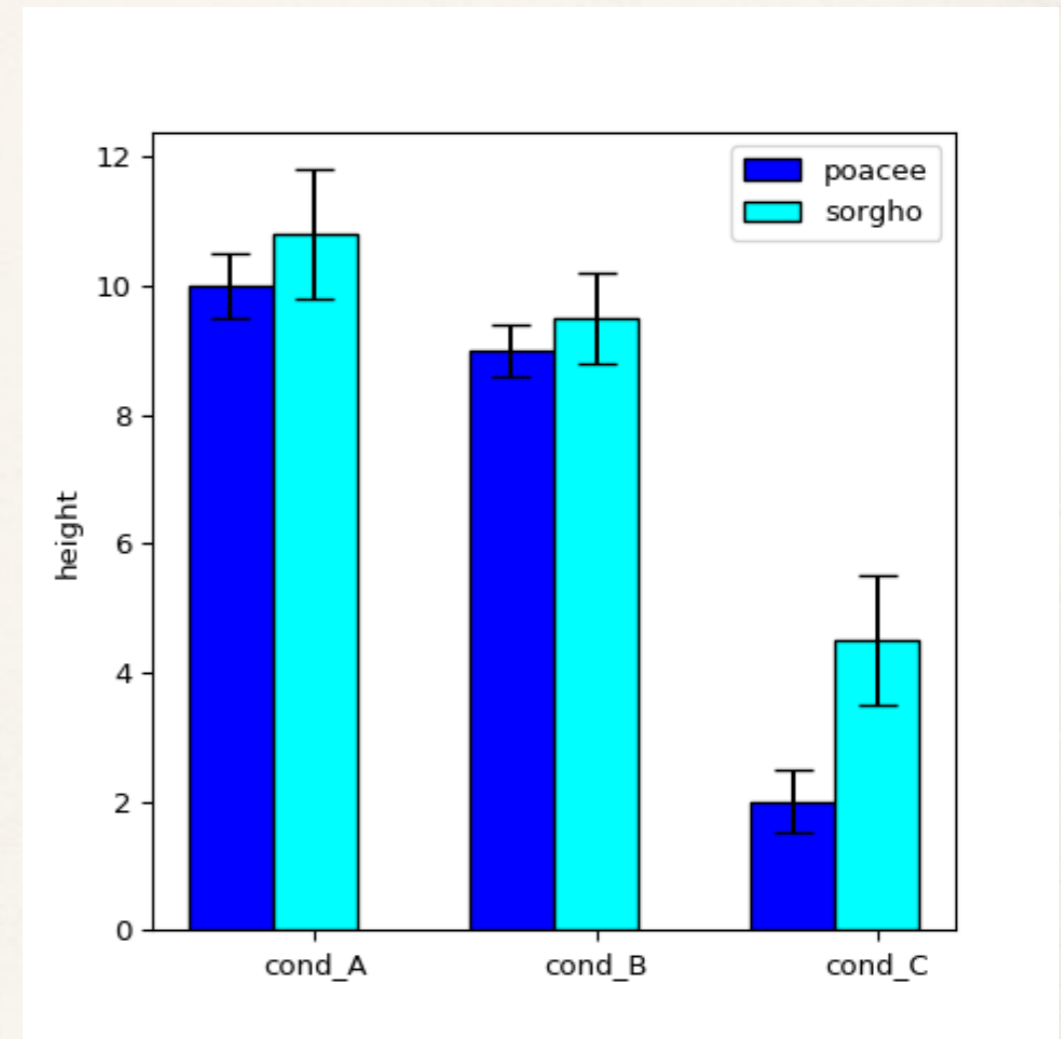


http://www.nytimes.com/interactive/2012/08/05/sports/olympics/the-100-meter-dash-one-race-every-medalist-ever.html?_r=0

Compare



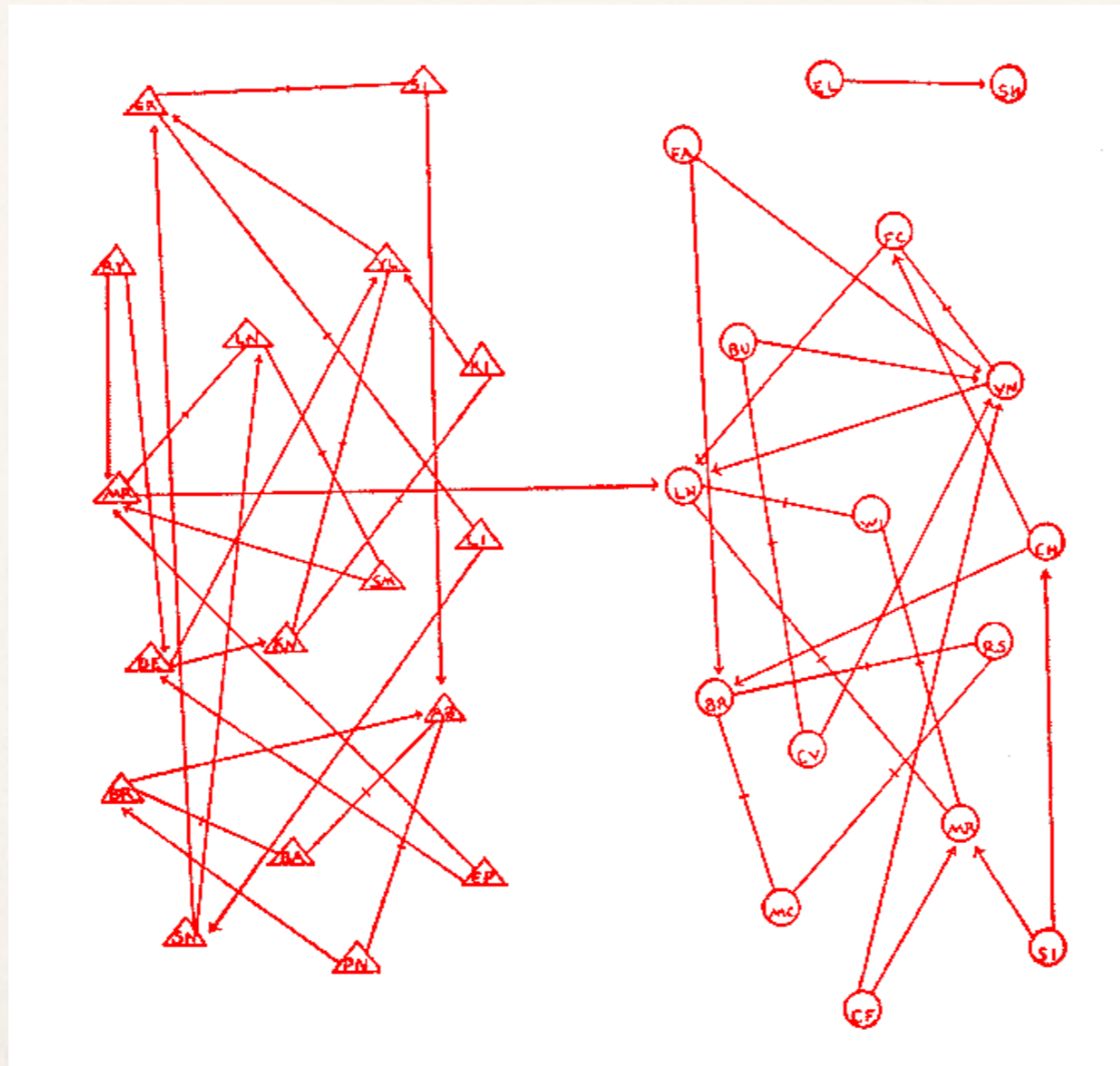
Line Graph



Bar Graph

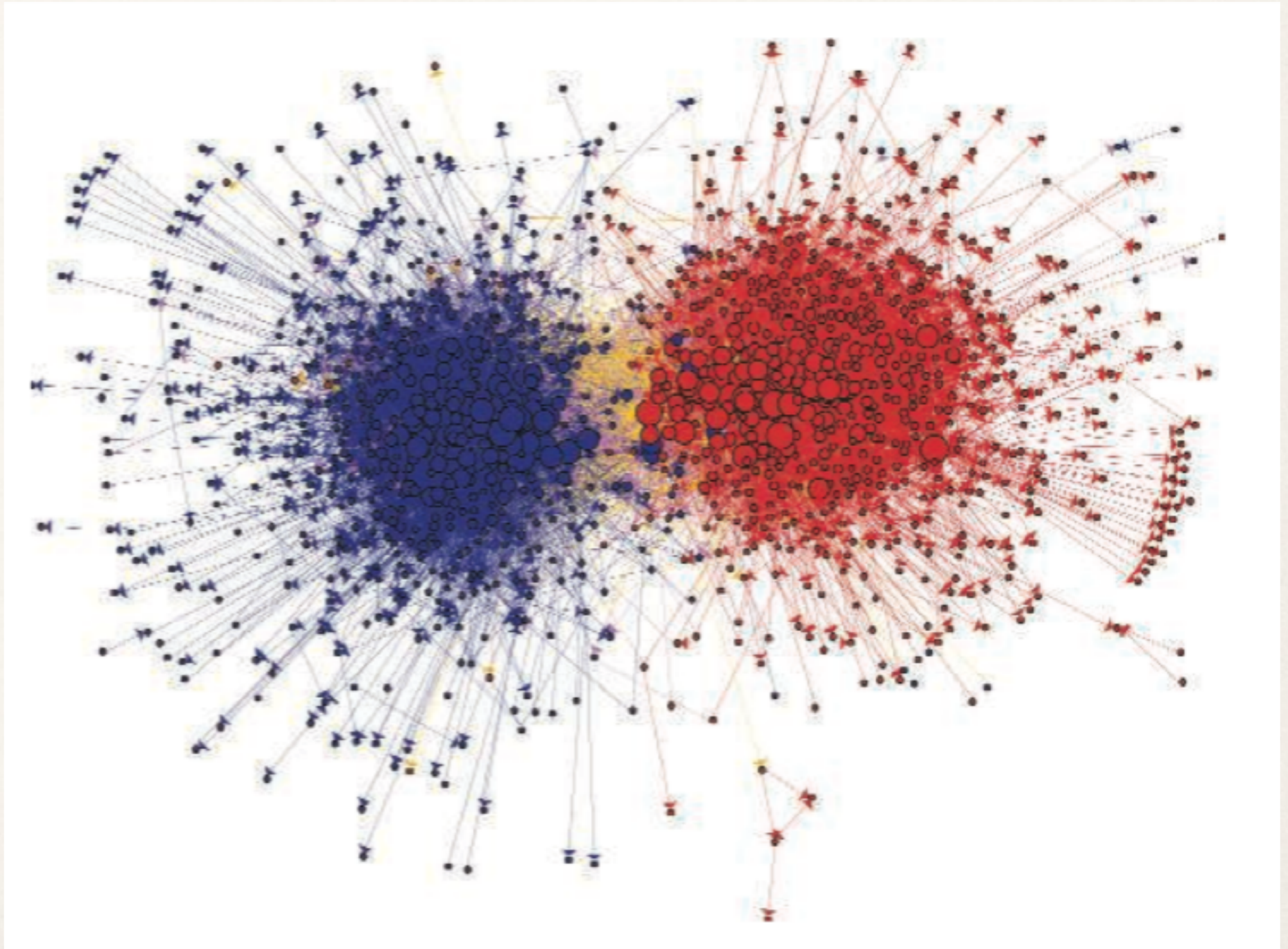
Connection

- ◆ 야코프 모레노(Jacob Moreno)가 1934년 network graph의 원형을 제시



초등학교 4학년 교우관계 조사

Connection: Social Network 시각화



미국 정치 블로거의 소셜 네트워크 연결망 구조의 시각화 (Lazer et al. 2009)

Connection: Social Network 시각화

- ✦ **Vertices** (nodes) connected by **Edges** (links)

	A	B	C
A	0	1	1
B	1	0	0
C	1	0	0

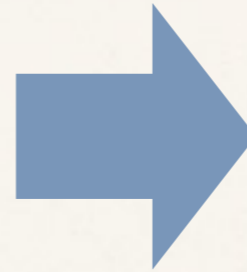
adjacency matrix

Connection: Social Network 시각화

- ✦ **Vertices** (nodes) connected by **Edges** (links)

	A	B	C
A	0	1	1
B	1	0	0
C	1	0	0

adjacency matrix



A: B, C

B: A

C: A

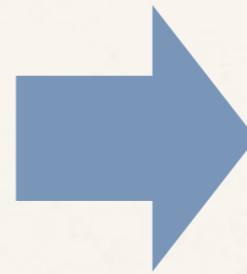
adjacency list

Connection: Social Network 시각화

- ✦ **Vertices** (nodes) connected by **Edges** (links)

	A	B	C
A	0	1	1
B	1	0	0
C	1	0	0

adjacency matrix

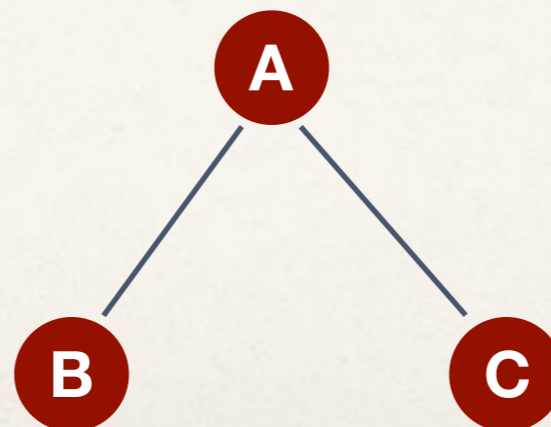


A: B, C

B: A

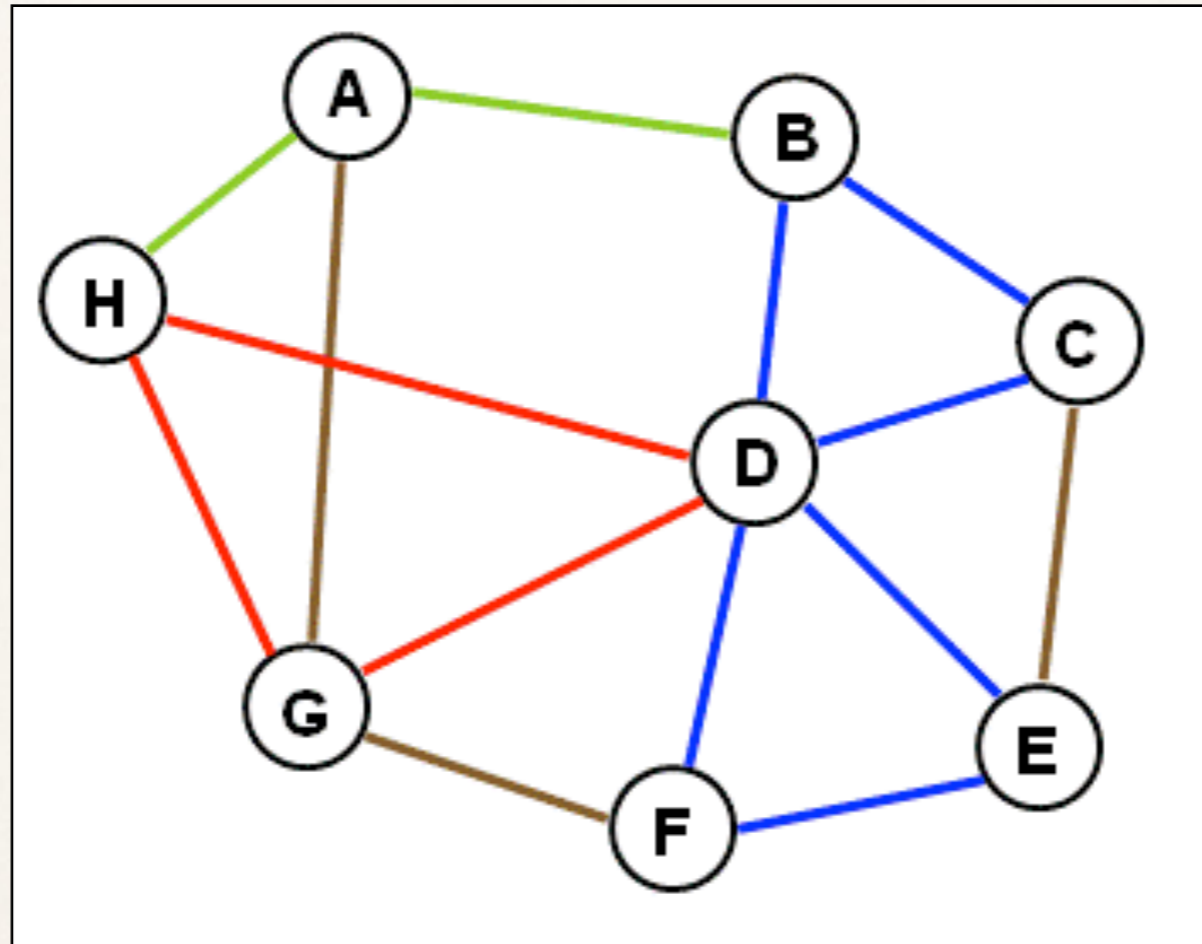
C: A

adjacency list



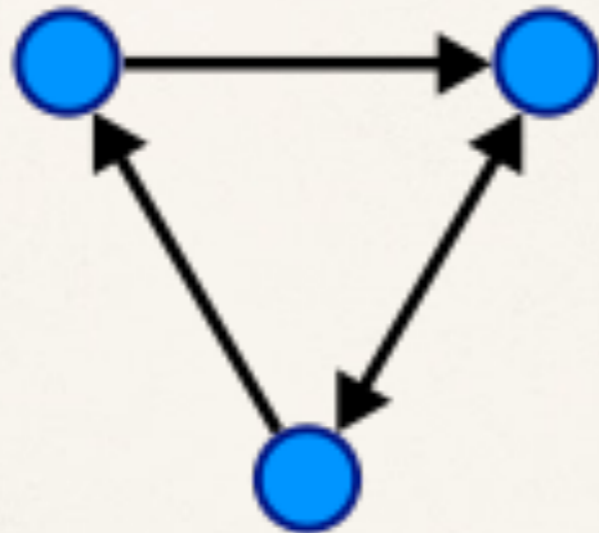
Connection: Social Network 시각화

- Graph는 cycle을 가질 수 있다.
 - cycle은 closed walk 또는 simple path 를 지칭

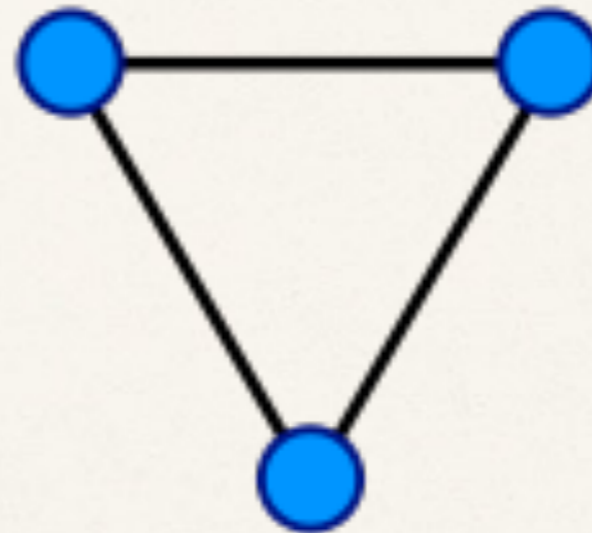


Connection: Social Network 시각화

- Graph의 edge는 directed 또는 undirected



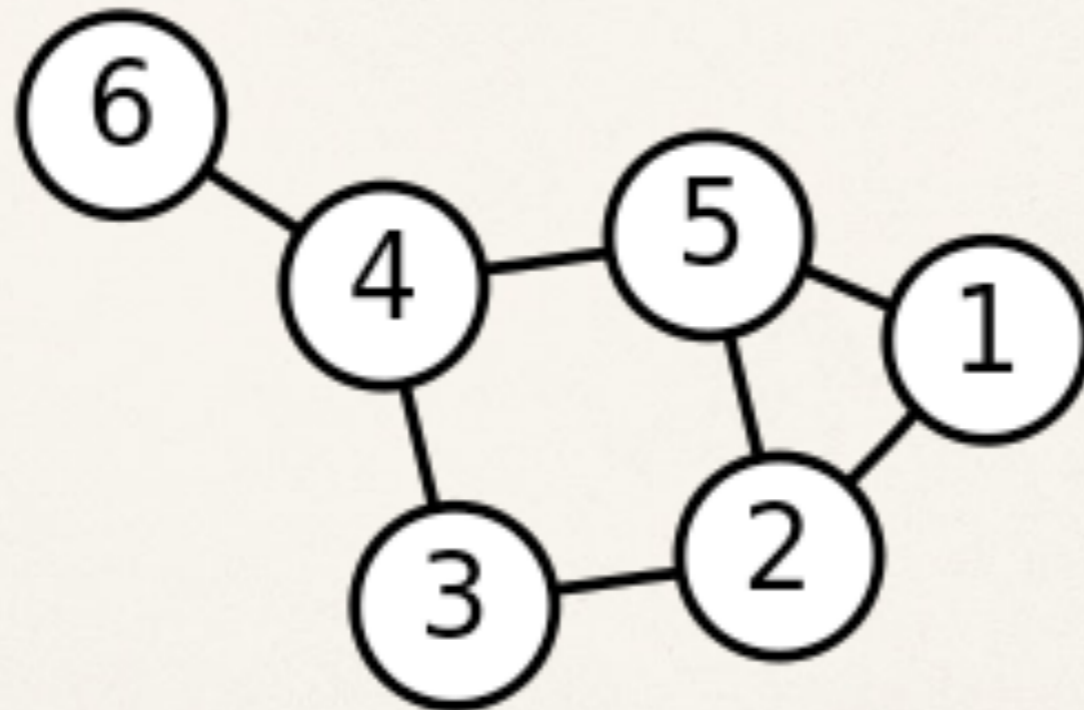
directed



undirected

Connection: Social Network 시각화

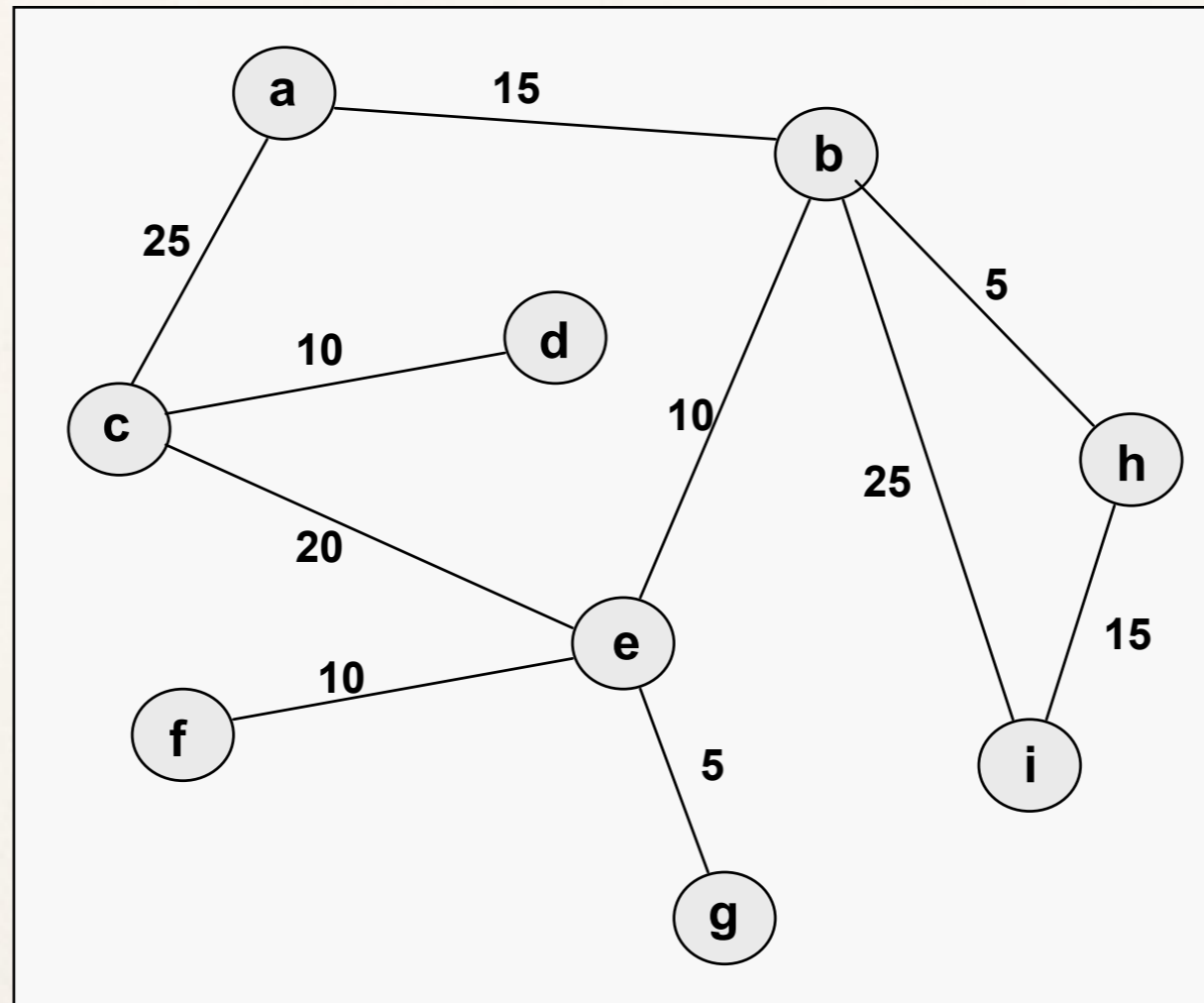
- ✦ degree of vertex
 - ✦ node(vertex)에 연결된 edge의 숫자
 - ✦ in-degree & out-degree (directed graph)



Connection: Social Network 시각화

- ◆ Weighted Graph

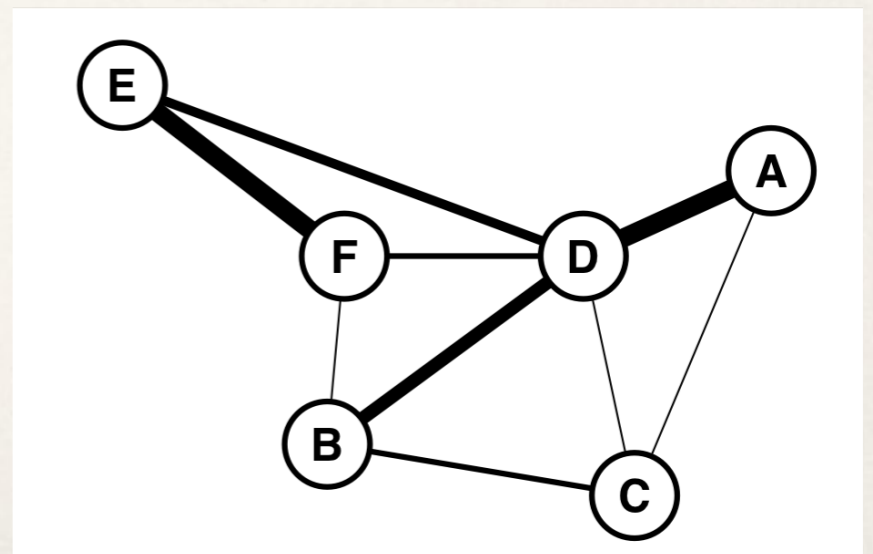
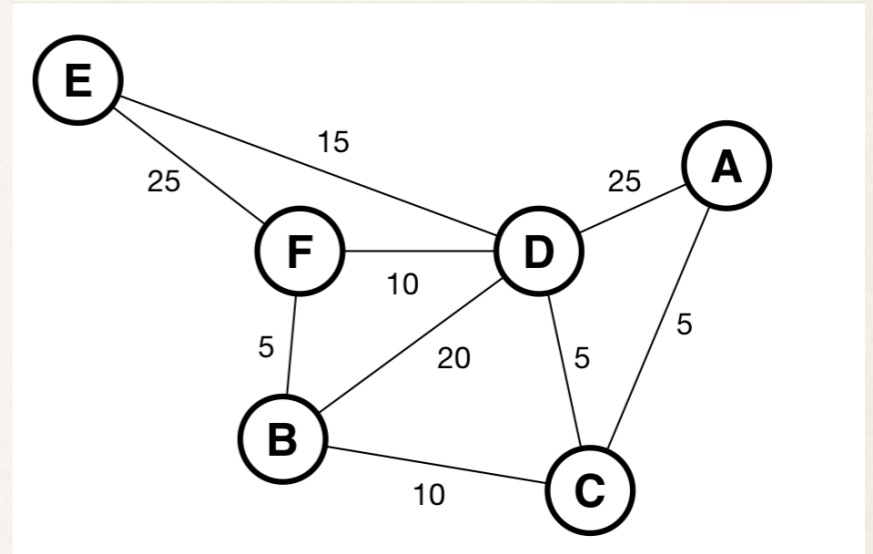
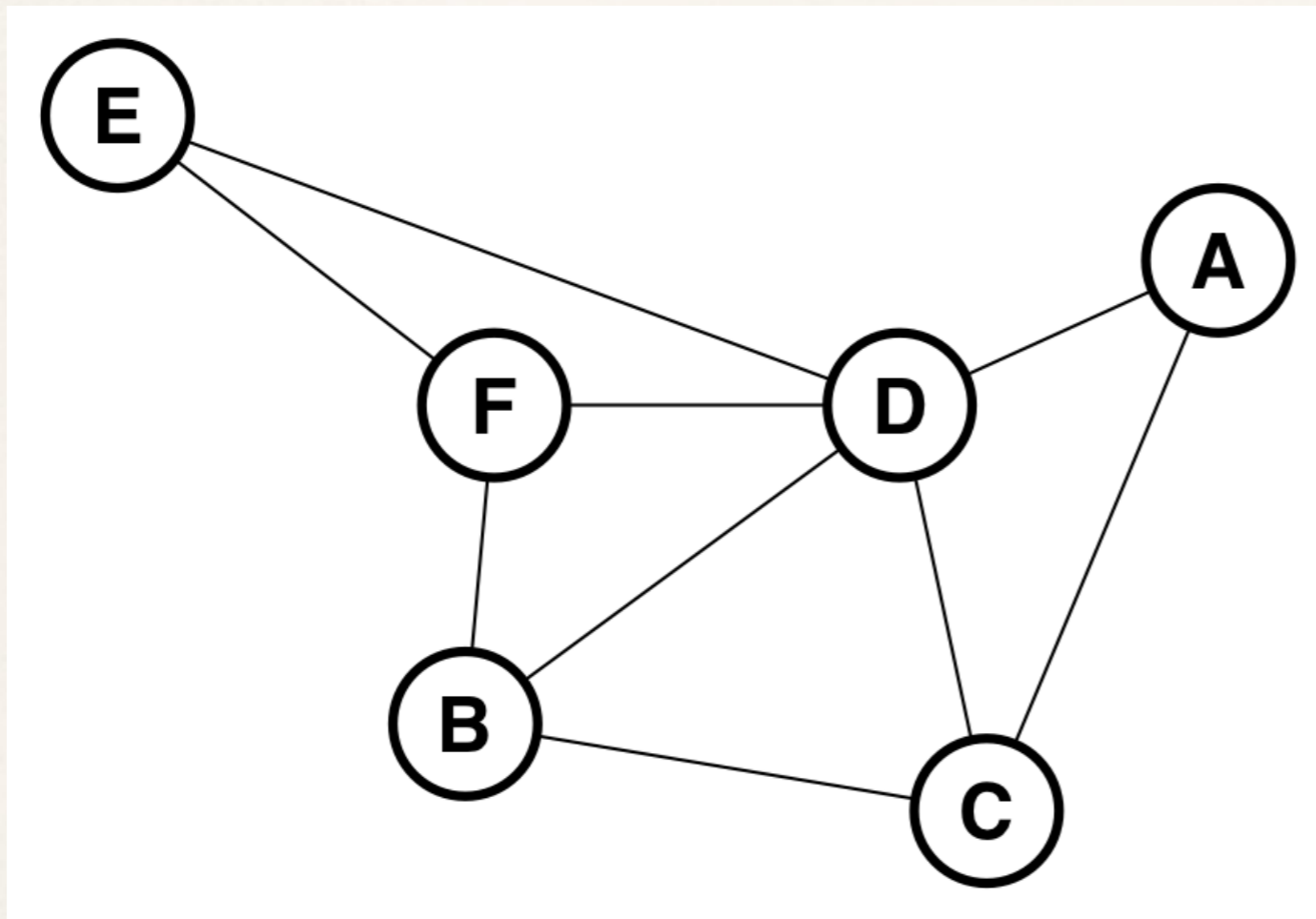
- ◆ 각각의 edge는 값을 가진다 (nominal, ordinal, quantitative)



Connection: Social Network 시각화

- ◆ Weighted Graph

- ◆ 각각의 edge는 값을 가진다 (nominal, ordinal, quantitative)



정당

- 한나라
- 민주
- 자유선진
- 민노
- 기타

의원 소모임

- 국민통합포럼
- 여의포럼
- 민본21
- 동행
- 허심회
- 함께내일로
- 선초회
- 아테네**
- 비례친목모임
- 이목회
- 일초회**
- 엔빅스모임
- 개혁과 미래
- 더좋은민주주의연구소**
- 다시민주주의
- 민주연대
- 국민과함께하는국회...
- 민주시니어
- 10인회**

2010년

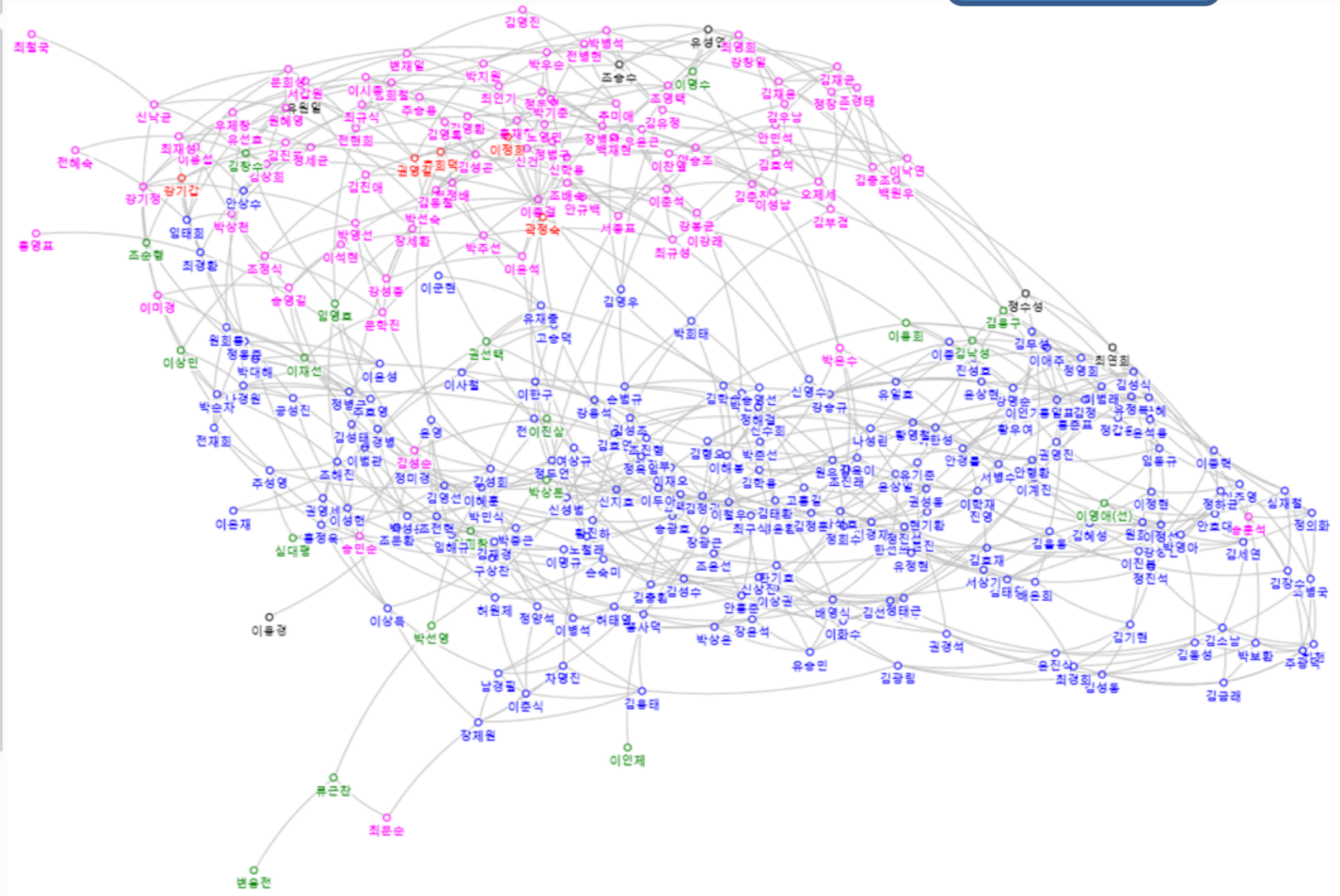
의원 선수

- 초선
- 재선
- 3선
- 4선
- 원로

지역

의원 검색

- 홍사덕**
- 홍정욱
- 홍준표



정당



의원 소모임

- 국민통합포럼
- 여의포럼
- 민본21
- 동행
- 허심회
- 함께내일로
- 선초회
- 아테네**
- 비례친목모임
- 이목회
- 일초회**
- 엔빅스모임
- 개혁과 미래
- 더좋은민주주의연구소**
- 다시민주주의
- 민주연대
- 국민과함께하는국회..
- 민주시니어
- 10인회**

의원 선수



지역

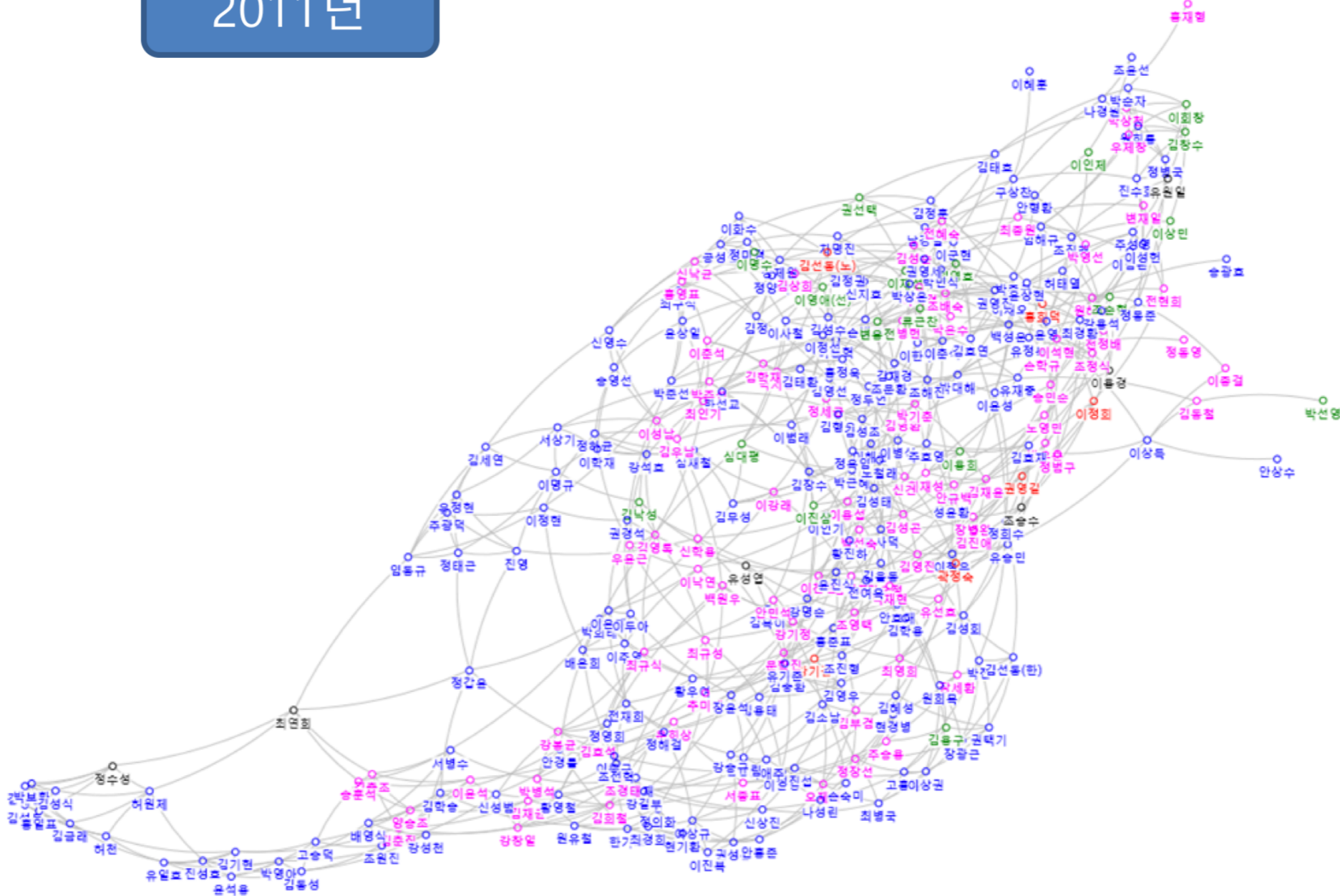


의원 검색

홍

- 홍사덕**
- 홍정욱
- 홍준표

2011년



Visualizing Data with Python

- ◆ matplotlib (<https://matplotlib.org/examples/>)
- ◆ Seaborn (<http://seaborn.pydata.org/examples/index.html>)
- ◆ ggplot (<http://ggplot.yhathq.com/>)
- ◆ ggplot (<https://plotnine.readthedocs.io/en/stable/>)
- ◆ bokeh* (<http://bokeh.pydata.org/en/latest/docs/gallery.html>)
- ◆ plot.ly* (<https://plot.ly/>)

* interactive visualization

Visualizing Data with Python

- ◆ Some resources:

- ◆ <https://python-graph-gallery.com/>

- ◆ <https://blog.modeanalytics.com/python-data-visualization-libraries/>

- ◆ <https://codeburst.io/overview-of-python-data-visualization-tools-e32e1f716d10>

Questions?
