

Data Visualization Basic

- ❖ 데이터시각화에 대한 기본 이해
- ❖ 데이터시각화 디자이너와의 협업과 커뮤니케이션의 중요성
- ❖ 디자이너가 알면 유용한 엑셀 함수 기본
- ❖ 실습 및 Q&A

II. 데이터시각화

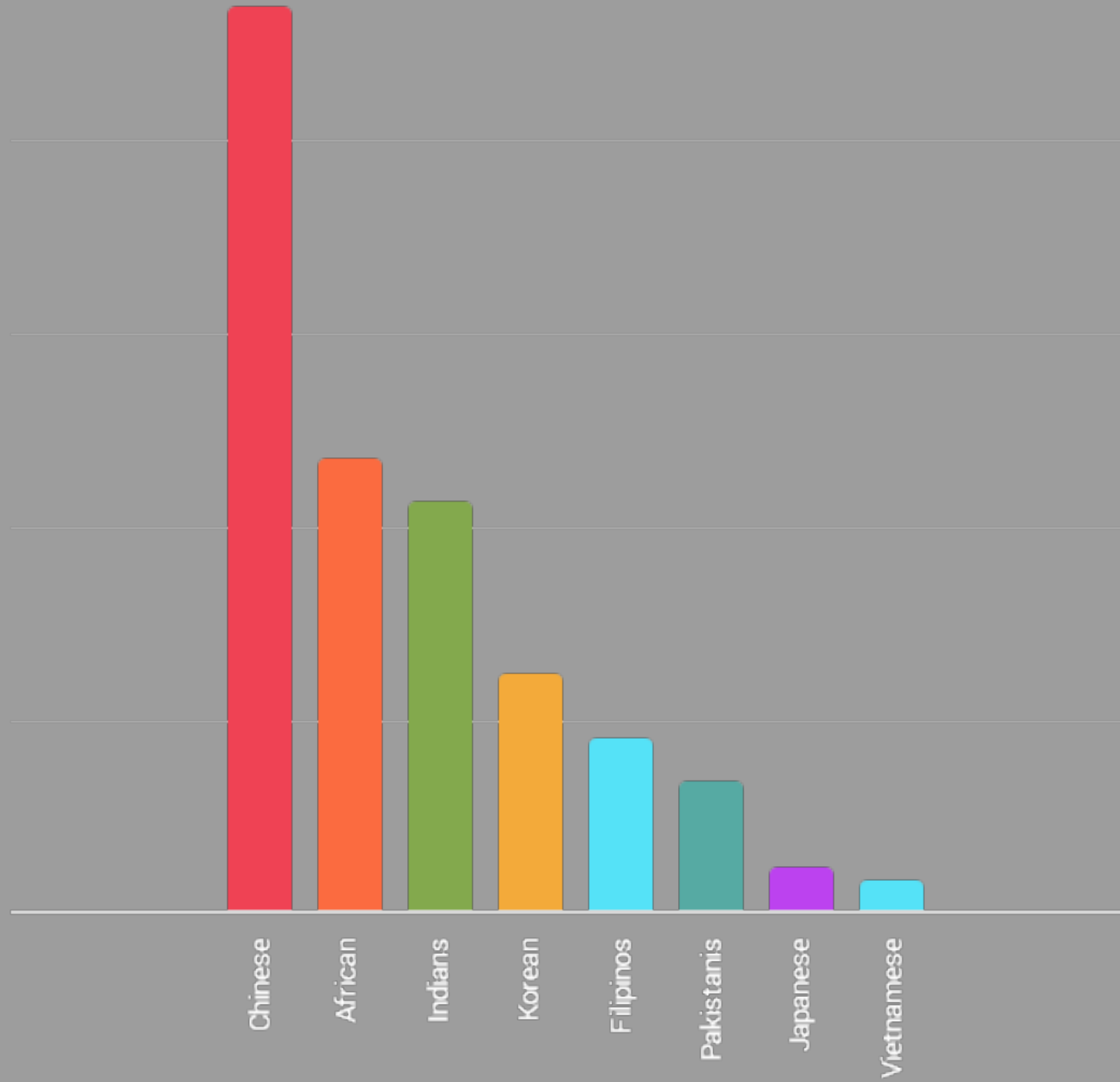
- ❖ 데이터시각화의 이해 (정의, 예제 등)
- ❖ 데이터시각화 방법론 (일러스트레이터, R, d3.js, Software)
- ❖ R을 활용한 데이터 시각화 실습 전 준비

Intro

The New York City metropolitan area is home to the largest Jewish community outside Israel. It is also home to nearly a quarter of the nation's Indian Americans and 15% of all Korean Americans and the largest Asian Indian population in the Western Hemisphere; the largest African American community of any city in the country; and including 6 Chinatowns in the city proper, comprised as of 2008 a population of 659,596 overseas Chinese, the largest outside of Asia. New York City alone, according to the 2010 Census, has now become home to more than one million Asian Americans, greater than the combined totals of San Francisco and Los Angeles. New York contains the highest total Asian population of any U.S. city proper. 6.0% of New York City is of Chinese ethnicity, with about forty percent of them living in the borough of Queens alone. Koreans make up 1.2% of the city's population, and Japanese at 0.3%. Filipinos are the largest southeast Asian ethnic group at 0.8%, followed by Vietnamese who make up only 0.2% of New York City's population. Indians are the largest South Asian group, comprising 2.4% of the city's population, and Bangladeshis and Pakistanis at 0.7% and 0.5%, respectively. / Demographics of New York, Wikipedia

http://en.wikipedia.org/wiki/Demographics_of_New_York_City

Intro



사례로 보는 데이터시각화

'불의 고리' 지진데이터 맵핑

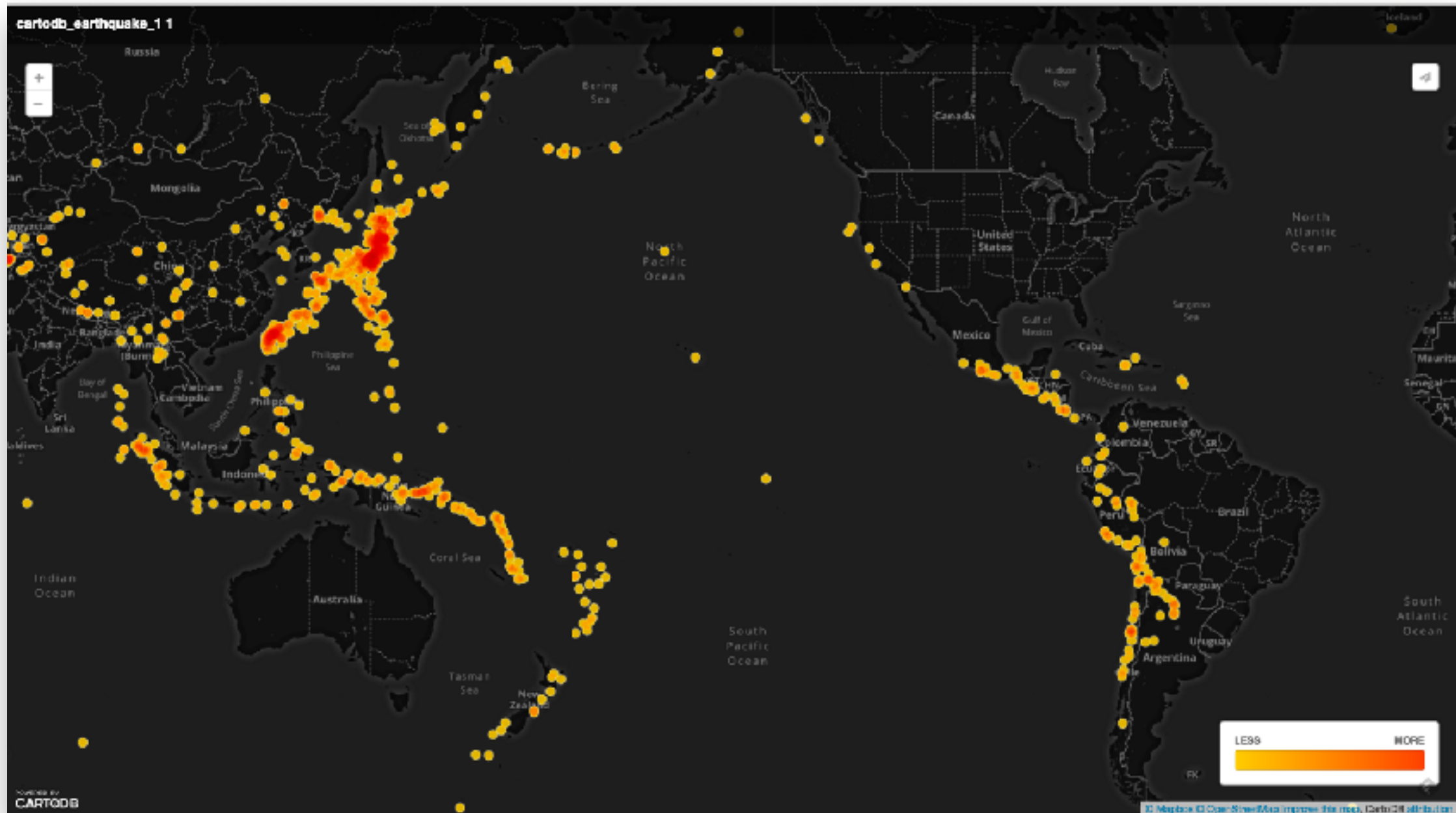
time	timetext	size	latitude	longitude	address
2014-04-01	2014년 4월 1일	5.1	36.95 N	124.5 E	충남 태안군 서격렬비도 서북서쪽 100km 해역
2004-05-29	2004년 5월 29일	5.2	36.8 N	130.2 E	경북 울진 동쪽 약 80km 해역
2003-03-30	2003년 3월 30일	5	37.8 N	123.7 E	인천광역시 백령도 서남서쪽 약 80km 해역
2016-04-19	2016년 4월 19일	5.5	32.6 N	130.7 E	일본 규슈 구마모토현 구마모토 남쪽 23km 지역
2016-04-18	2016년 4월 18일	5.8	33.0 N	131.2 E	일본 규슈 오이타현 오이타 남서쪽 46km 지역
2016-04-17	2016년 4월 17일	7.4	0.2 N	-80.1 E	에콰도르 페데르날레스 북북서쪽 14km 해역
2016-04-16	2016년 4월 16일	5.3	33.3 N	131.4 E	일본 규슈 오이타현 오이타 서북서쪽 21km 지역
2016-04-16	2016년 4월 16일	5.8	33.0 N	131.2 E	일본 규슈 오이타현 오이타 남서쪽 46km 지역
2016-04-16	2016년 4월 16일	5.8	33.0 N	131.1 E	일본 규슈 구마모토현 구마모토 동북동쪽 43km 지역
2016-04-16	2016년 4월 16일	6	32.9 N	130.9 E	일본 규슈 구마모토현 구마모토 동북동쪽 21km 지역
2015-12-26	2015년 12월 26일	6.2	36.52 N	71.14 E	아프가니스탄 카불 북동쪽 283km 지역
2015-12-18	2015년 12월 18일	6.6	15.8 N	-93.5 E	멕시코 텍시코시티 남동쪽 719km 지역
2015-12-07	2015년 12월 7일	7.2	38.26 N	72.72 E	티지키스탄 두산베 동쪽 345km 지역
2015-12-05	2015년 12월 5일	7.1	-47.74 N	85.18 E	호주 퍼스 남서쪽 3124km 해역
2015-11-26	2015년 11월 26일	6.4	-9.22 N	-71.27 E	브라질 타라우아카 남남서쪽 130km 지역
2015-11-25	2015년 11월 25일	7.5	-10.59 N	-70.97 E	페루 이베리아 서북서쪽 173km 지역
2015-11-20	2015년 11월 20일	6	25.9 N	143.4 E	일본 도쿄 남남동쪽 1142km 해역
2015-11-19	2015년 11월 19일	7	-8.7 N	158.7 E	솔로몬제도 다달리 남남서쪽 80km 해역
2015-11-15	2015년 11월 15일	5.8	31.3 N	128.8 E	일본 규슈 가고시마현 가고시마 서남서쪽 170km 해역
2015-11-14	2015년 11월 14일	7	30.9 N	128.7 E	일본 규슈 가고시마현 가고시마 서남서쪽 193km 해역
2015-11-07	2015년 11월 7일	5.8	-30.91 N	-71.54 E	칠레 산티아고 북북서쪽 295km 지역
2015-11-04	2015년 11월 4일	6.3	-8.35 N	124.9 E	동티모르 딜리 서북서쪽 77km 지역
2015-11-02	2015년 11월 2일	5.9	22.71 N	121.6 E	대만 타이둥 동쪽 47km 해역
2015-10-26	2015년 10월 26일	7.7	36.46 N	70.71 E	아프가니스탄 카불 북북동쪽 255km 지역
2015-10-21	2015년 10월 21일	5.5	37.3 N	141.2 E	일본 도쿄 북동쪽 230km 해역



사례로 보는 데이터시각화

'불의 고리' 지진데이터 맵핑

시각화를 통해 분명한 메시지와 인사이트를 확인할 수 있다



사례로 보는 데이터시각화

전국 고액체납자 데이터

데이터 출처 : 2015년 전국 고액체납자 명단

catobj_id - number	the_geom - geometry	address - string	age - string	cost - number	count - number	lat - number	lon - number	month - number	name - string	side - string	sigungu - string	reference - string	tax_type - string	time - string	year - number
1	124.9996, 37.5367	서울 강남구 테헤란로25길 5	30	792	2	127.53667402	124.9996207	2	강민서	서울	강남구	2012년 04월 지방소득세 등 총 2건 792	지방소득세	2017년 02월	2012
2	127.0754, 37.4874	서울 용산구 용인로154길 29-6	35	171	1	127.53948205	127.0754215	3	김민수	서울	용산구	2011년 02월 지방소득세 등 총 1건 171	지방소득세	2017년 03월	2011
3	127.0461, 37.5237	서울 강남구 도산로454	30	47	10	127.53956202	127.0461224	12	김민수	서울	강남구	2010년 12월 지방소득세 등 총 13건 47	지방소득세	2017년 02월	2010
4	127.0273, 37.5107	서울 서초구 평범로 52	31	253	4	127.51268204	127.0273223	4	김민수	서울	서초구	2012년 04월 지방소득세 등 총 5건 253	지방소득세	2017년 04월	2012
5	124.9771, 37.5382	서울 용산구 남대문로4길 40	34	141	2	127.53817234	124.9771204	3	김민수	서울	용산구	2011년 02월 지방소득세 등 총 2건 141	지방소득세	2017년 02월	2011
6	127.0947, 37.4794	서울 용산구 서대문로23길 30-5	32	75	1	127.57939944	127.0947338	5	김민수	서울	용산구	2011년 04월 지방소득세 등 총 1건 75	지방소득세	2017년 04월	2011
7	127.2207, 37.7465	경기 고양시 일산서구 일산1로 128	39	13	3	127.42648673	127.2207993	7	김민수	경기	고양시	2012년 04월 지방소득세 등 총 3건 13	지방소득세	2017년 04월	2012
8	127.7947, 37.7840	경기 고양시 일산서구 일산2로 257	38	41	4	127.79696034	127.7947943	12	김민수	경기	고양시	2012년 12월 지방소득세 등 총 5건 41	지방소득세	2017년 04월	2012
9	127.5232, 37.5091	서울 서초구 서양로27길 2	38	75	10	127.52329522	127.5232952	8	김민수	서울	서초구	2011년 04월 지방소득세 등 총 3건 75	지방소득세	2017년 04월	2011
10	124.9289, 37.5172	서울 서대문구 영등포로 4	75	16	7	127.44733293	124.9289998	11	김민수	서울	서대문구	2012년 12월 지방소득세 등 총 7건 16	지방소득세	2017년 03월	2012
11	124.9908, 37.5174	서울 중구 필동로 44	39	129	8	127.51742082	124.9907981	1	김민수	서울	중구	2011년 04월 지방소득세 등 총 3건 129	지방소득세	2017년 04월	2011
12	124.9423, 37.5437	서울 서대문구 통일로27길 20	37	42	4	127.51263673	124.9423254	7	김민수	서울	서대문구	2011년 07월 지방소득세 등 총 4건 42	지방소득세	2017년 07월	2011
13	127.0462, 37.5192	서울 강남구 선릉로130길 24	40	40	2	127.51923271	127.0464647	1	김민수	서울	강남구	2012년 04월 지방소득세 등 총 2건 40	지방소득세	2017년 04월	2012
14	127.0462, 37.5192	서울 강남구 선릉로130길 24	33	47	3	127.51923271	127.0464647	3	김민수	서울	강남구	2012년 02월 지방소득세 등 총 3건 47	지방소득세	2017년 03월	2012
15	124.9223, 37.5197	서울 은평구 세전로111-5	34	43	11	127.57563151	124.9223224	2	김민수	서울	은평구	2011년 02월 지방소득세 등 총 11건 43	지방소득세	2017년 02월	2011
16	124.9321, 37.5473	서울 서초구 개포대로 143	30	33	1	127.54749408	124.9321204	3	김민수	서울	서초구	2012년 02월 지방소득세 등 총 1건 33	지방소득세	2017년 02월	2012
17	127.1173, 37.4632	경기 남양주시 평내로 283	35	2169	9	127.6321442	127.117321	4	김민수	경기	남양주시	2012년 04월 지방소득세 등 총 9건 2169	지방소득세	2017년 04월	2012
18	124.9127, 37.4632	서울 관악구 신림2로 12-2	48	34	3	127.4522240	124.9127217	10	김민수	서울	관악구	2011년 12월 지방소득세 등 총 3건 34	지방소득세	2017년 03월	2011
19	127.4987, 34.9434	전남 순천시 상남로177	44	15	10	127.49874079	127.4987046	8	김민수	전남	순천시	2011년 04월 지방소득세 등 총 10건 15	지방소득세	2017년 04월	2011
20	124.9702, 37.4840	서울 구로구 구로로3길 41	32	44	7	127.48409172	124.970201	7	김민수	서울	구로구	2011년 04월 지방소득세 등 총 7건 44	지방소득세	2017년 04월	2011
21	124.9618, 37.5037	서울 용마루 원촌로100	39	45	1	127.50379181	124.9618905	7	김민수	서울	용마루	2011년 04월 지방소득세 등 총 1건 45	지방소득세	2017년 04월	2011
22	127.0232, 37.5045	서울 서초구 교우대로 25	34	39	3	127.50459722	127.0232492	2	김민수	서울	서초구	2012년 04월 지방소득세 등 총 3건 39	지방소득세	2017년 04월	2012
23	124.9913, 37.5007	서울 서초구 평범대로 32	36	329	3	127.50071286	124.9913277	9	김민수	서울	서초구	2011년 04월 지방소득세 등 총 3건 329	지방소득세	2017년 04월	2011
24	124.9244, 37.5117	경기 고양시 상봉로 229	36	40	4	127.2447114	124.9244224	3	김민수	경기	고양시	2012년 04월 지방소득세 등 총 4건 40	지방소득세	2017년 04월	2012
25	127.5158, 37.4944	서울 송파구 송파7로22길 15	31	95	6	127.51582022	127.5158208	5	김민수	서울	송파구	2012년 05월 지방소득세 등 총 6건 95	지방소득세	2017년 05월	2012
26	124.9444, 37.4812	서울 관악구 통일로 27	34	45	9	127.48124045	124.9444445	7	김민수	서울	관악구	2012년 04월 지방소득세 등 총 7건 45	지방소득세	2017년 04월	2012
27	124.9284, 37.5174	서울 중구 필동대로 270-30	70	112	4	127.51746487	124.9284211	7	김민수	서울	중구	2012년 04월 지방소득세 등 총 4건 112	지방소득세	2017년 04월	2012

사례로 보는 데이터시각화

전국 고액체납자 데이터 시각화

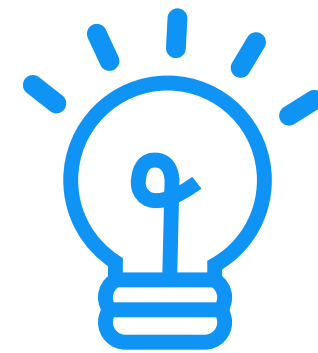
데이터 출처 : 2015년 전국 고액체납자 명단



데이터시각화의 배경

콘텐츠로서의 데이터

무엇을 느끼셨고
어떤 생각이 들었나요?



데이터시각화의 배경

콘텐츠로서의 데이터

앞서봤던 사례처럼 데이터는
그저 쌓아놓으려고 저장하는 것은 아님

전통산업에선 데이터(정보)는 서비스 대상이 아님 = **엑셀**

3차 산업이 발달되면서 데이터는 콘텐츠로 진화 = **시각화**

데이터시각화의 배경

콘텐츠로서의 데이터

5,242,880TB

인류의 탄생부터 2003년까지 쌓아 놓은 데이터의 양은 약 **5엑사바이트**

2003년 이후 이 규모의 데이터를 생성하는데 걸린 시간은 단 이틀



데이터시각화의 배경

콘텐츠로서의 데이터

기록 = 데이터

약 100년 전의 데이터



데이터시각화의 배경

콘텐츠로서의 데이터

쌓아 놓고만 있는 데이터

불과 얼마 전까지의 'DB'



데이터시각화의 배경

콘텐츠로서의 데이터

데이터에 대한 관심이 급증 <Google Ngram>

Google Books Ngram Viewer

Graph these comma-separated phrases: case-insensitive

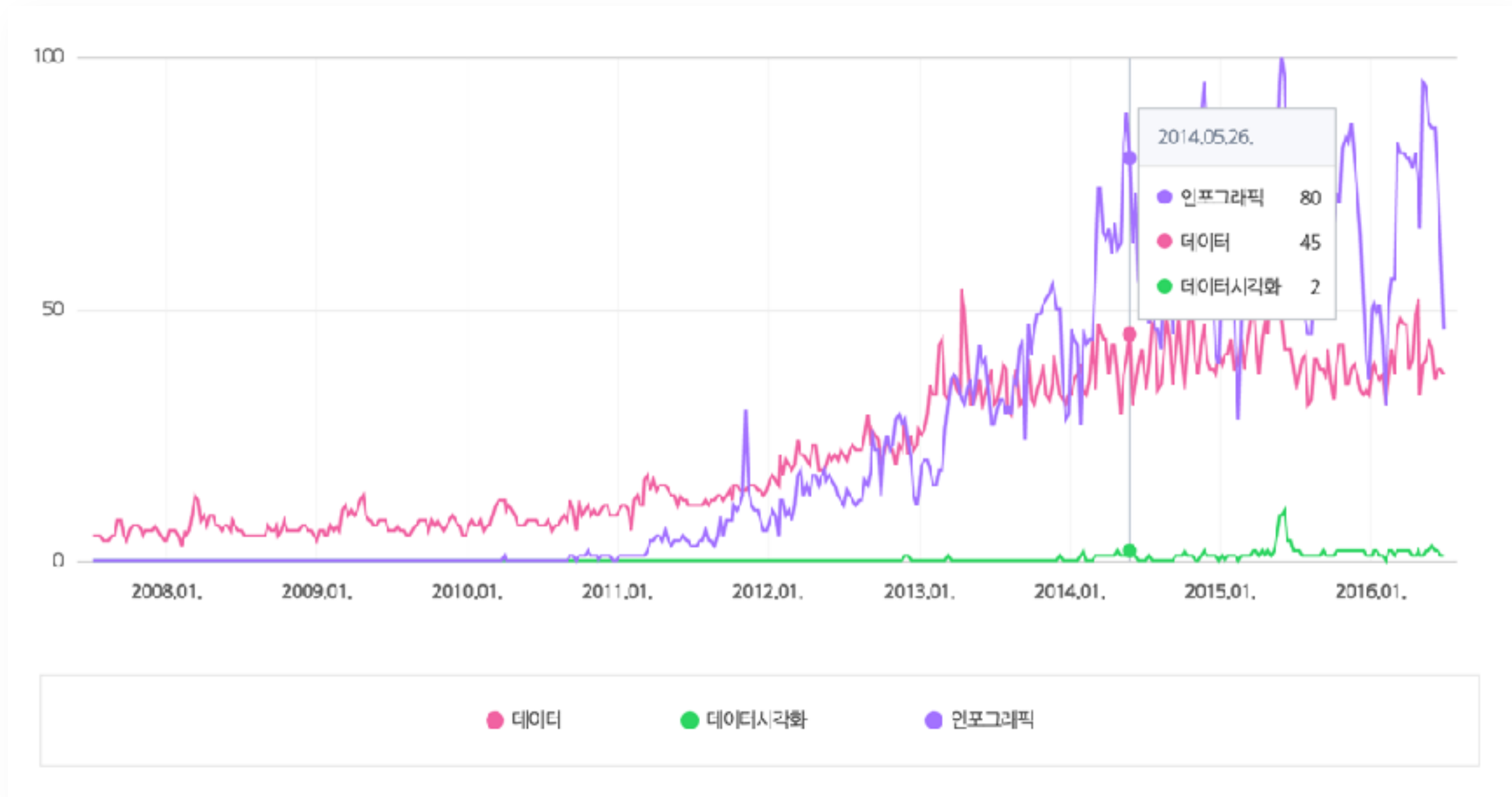
between and from the corpus with smoothing of [Search lots of books](#)



데이터시각화의 배경

콘텐츠로서의 데이터

데이터에 대한 관심이 급증 <네이버 데이터랩>



데이터시각화의 배경

콘텐츠로서의 데이터

결국은 쌓아 놓은 데이터를 모든 분야에서 활용하는 시대
그 중심에 데이터 시각화가 각광

시각화는 직관적이고 시각적이기 때문에
소비자 혹은 독자에게 매력적인 콘텐츠 유형

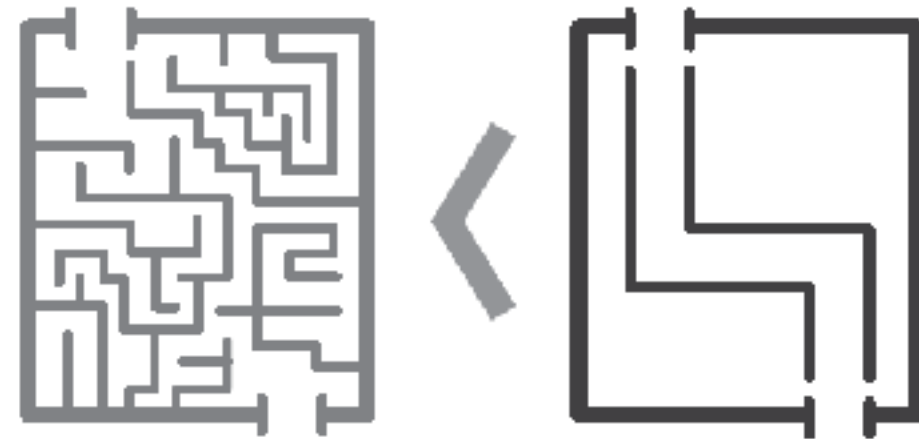
데이터시각화의 이해

데이터시각화의 우수성



흥미 유발

이미지는 글보다 **시각적인 즐거움**을 줄 수 있다
딱딱하고 재미없는 정보를 부담스러워 하는
수용자들의 접근 장벽을 낮출 수 있다



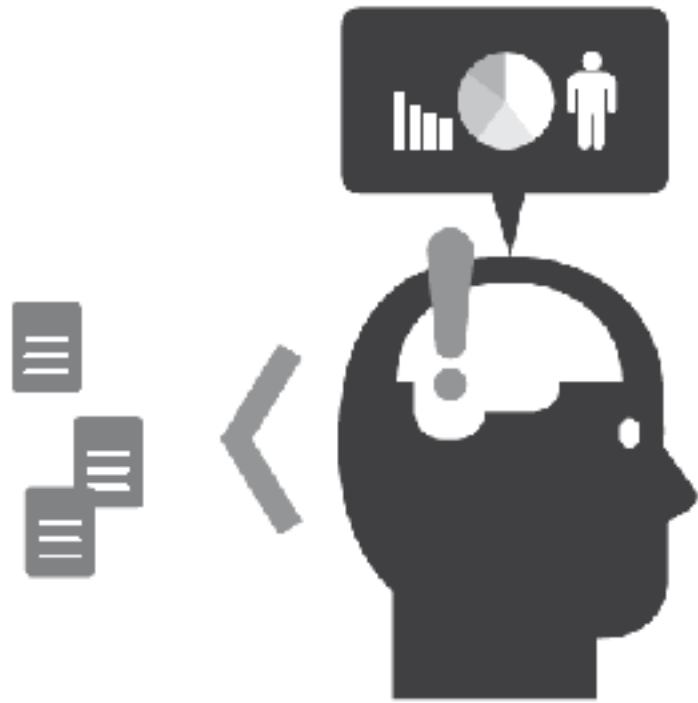
이해 쉬운 전달

보통의 사람들이 **당연히 알아야 하고 필요한 정보들이**
어렵고, 지루하게 존재하는 경우가 많다
이런 정보들은 쉬운 전달이 필요하다

출처 : 203 Infographics Lab

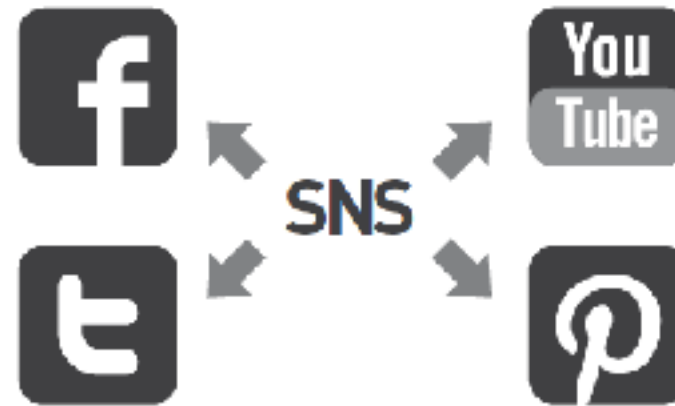
데이터시각화의 이해

데이터시각화의 우수성



오랜 기억 유지

사람은 의사소통시 **시각의존도**가 높다
이미지는 뇌에서 글보다 **더 직관적**이다
따라서 상대적으로 **오래 기억**할 수 있다



자발적 확산

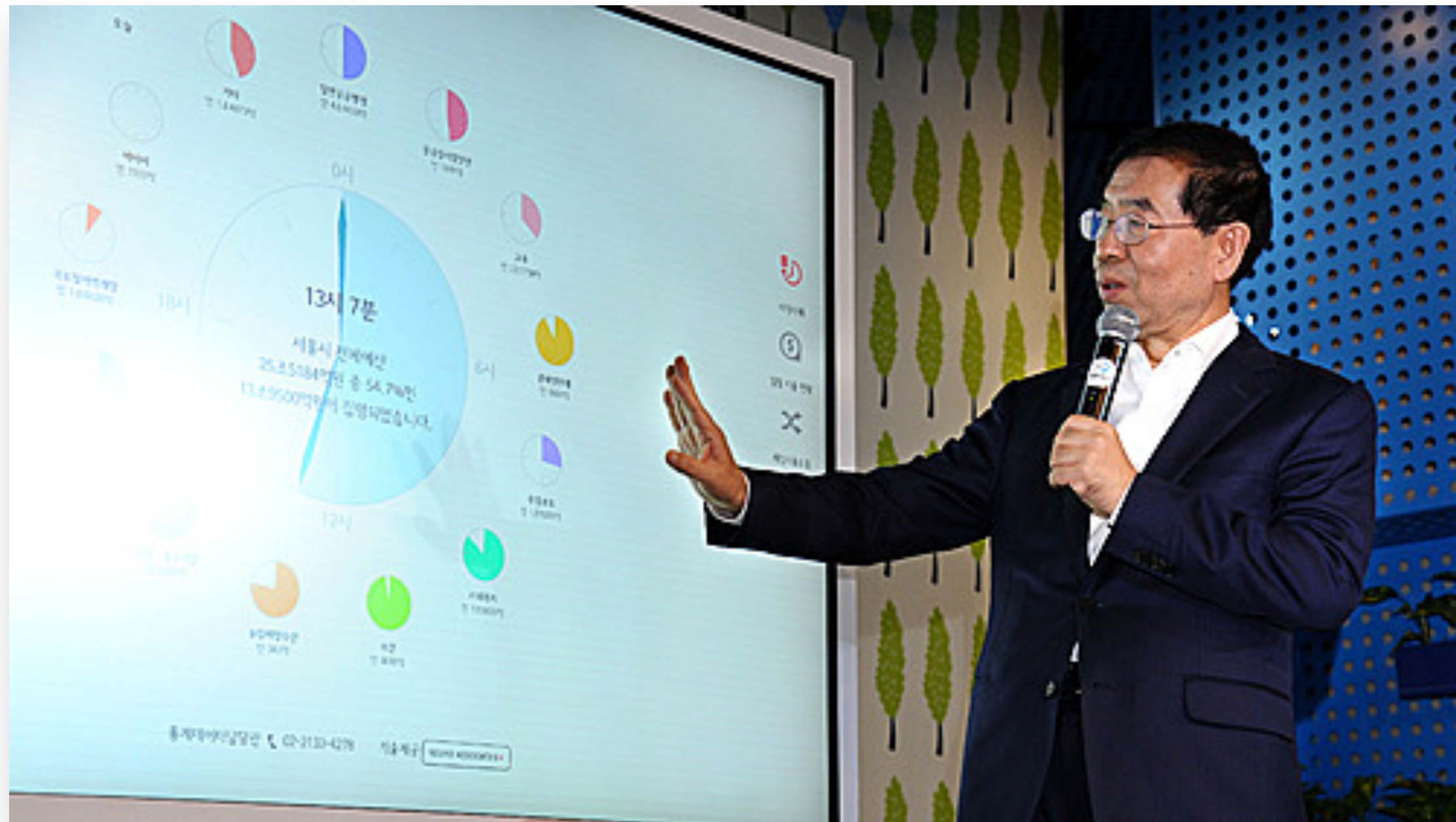
SNS를 통해 **공유**하기가 쉽다
빠른 확산이 입소문을 만든다
전 세계로 **까지 확산** 시킬 수 있다

출처 : 203 Infographics Lab

데이터시각화의 다양한 활용

데이터시각화의 확장

공공데이터를 활용한 데이터 시각화 | 서울시



데이터시각화의 다양한 활용

데이터시각화의 확장

인터랙티브 데이터시각화 탁구대



데이터시각화의 다양한 활용

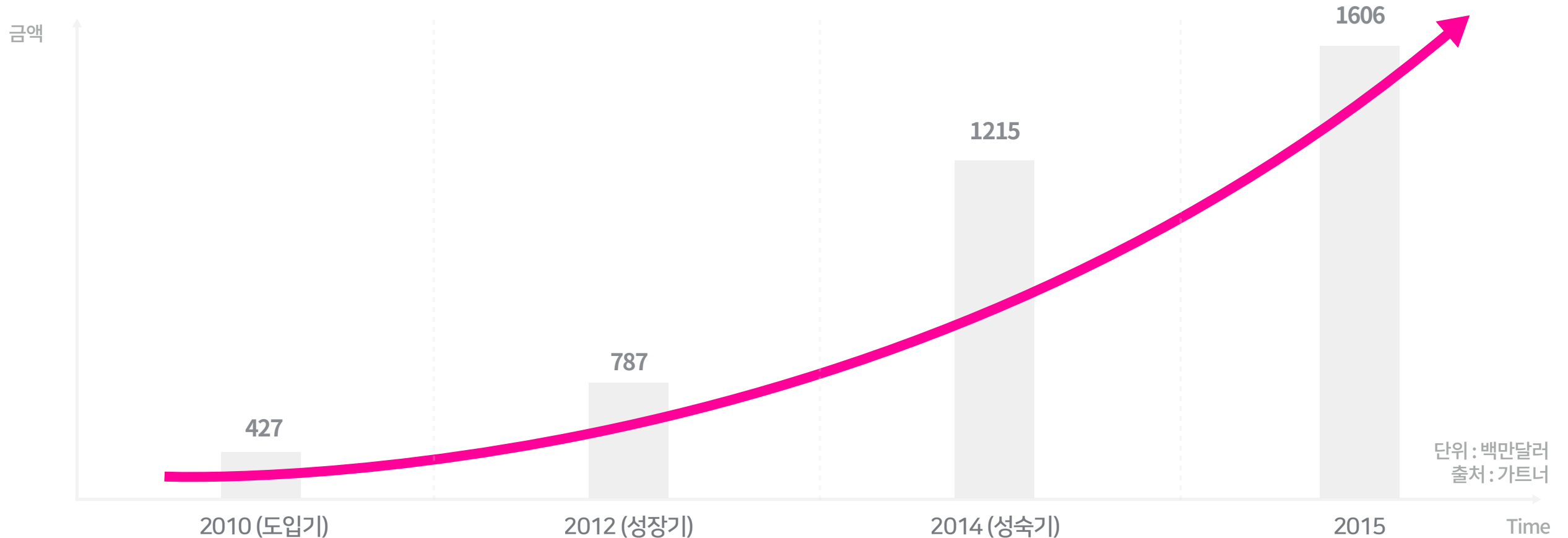
데이터시각화의 확장

데이터시각화 소프트웨어



Why Data Visualization?

데이터시각화의 매력



특성

데이터 시각화에 대한
관심 증가

개인 중심의 포트폴리오
작업 위주 시장

데이터 저널리즘과
정보시각화 교류 활발

Google chart와 같이
기본 차트형 프로그램 출현

국제기구, 해외 미디어, 대기업
등에서 시각화 콘텐츠 활용 활발

데이터 시각화 전문 미디어 등장
(NYT, VOX, FiveThirtyEight)

새로운 융합 분야 콘텐츠로
다양한 방면의 데이터와
디자인 융합 가능

국내

← 일러스트레이터를 활용한 인포그래픽(정보) 중심의 시각화

→ 폭발하는 데이터를 활용한
데이터 시각화에 대한
관심 증대

Why Data Visualization?

데이터시각화의 매력

들으면 잊는다

보면 기억한다

행동하면 이해한다

-공자-

데이터시각화의 이해

데이터시각화의 정의

데이터시각화 | Data Visualization

데이터 분석 결과를 쉽게 이해할 수 있도록 시각적으로 표현
데이터 시각화의 목적은 **차트(graph)**를 통해 정보를
명확하고 효과적으로 전달하는 것

출처 : 위키피디아

데이터시각화란?

데이터시각화의 다른 정의

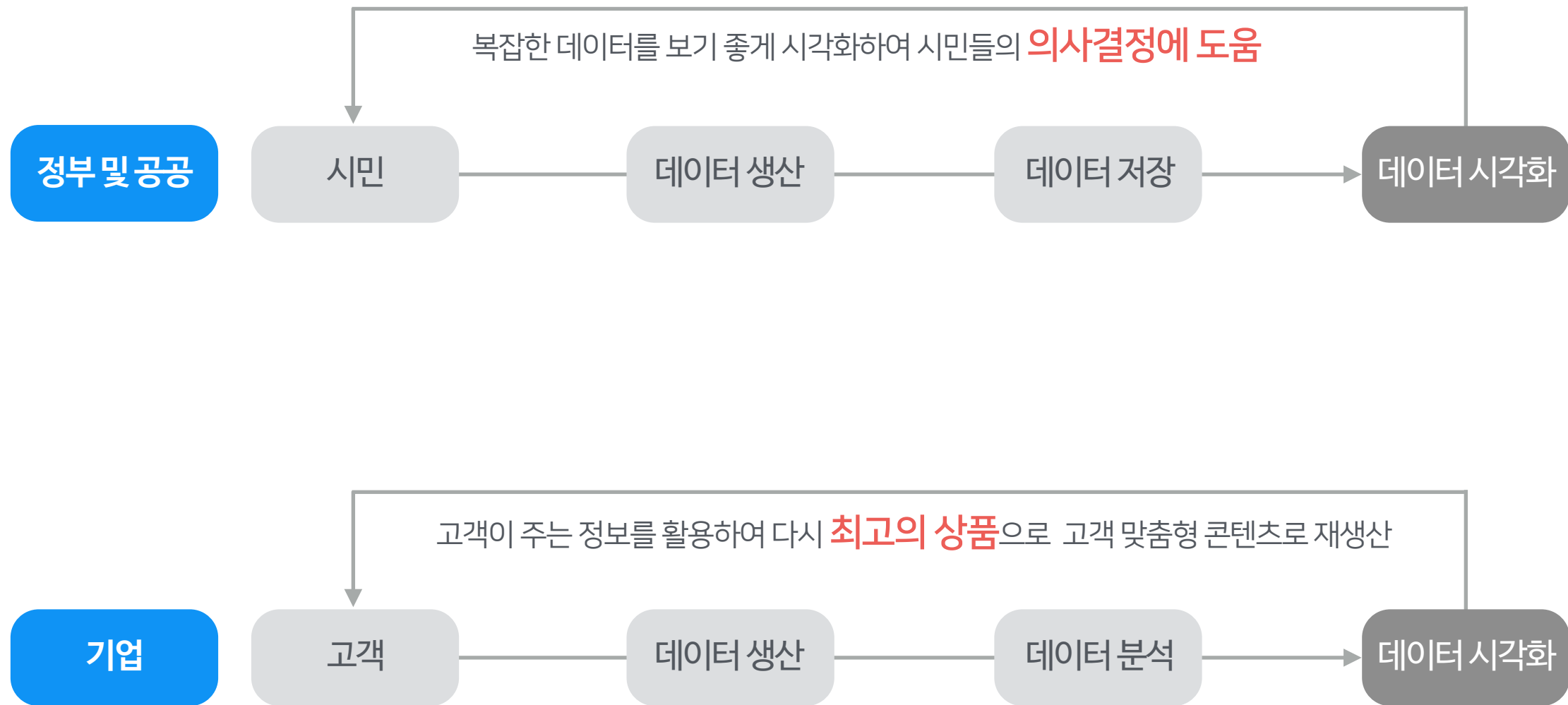
경제학자 밀턴 프리드먼이 말하는 데이터시각화

데이터 시각화는 지나치게 기능적인 측면을 강조하거나 아름답게 표현하는 데만 매달려서는 안 된다
이상적인 시각화란 단지 명확하게 의사를 전달하는 데 머물러서는 안 되고 보는 사람을 **집중하게** 하고 **참여하게** 만들어야 한다

밀턴 프리드먼 (Milton Friedman) | 경제학자

데이터시각화의 장점

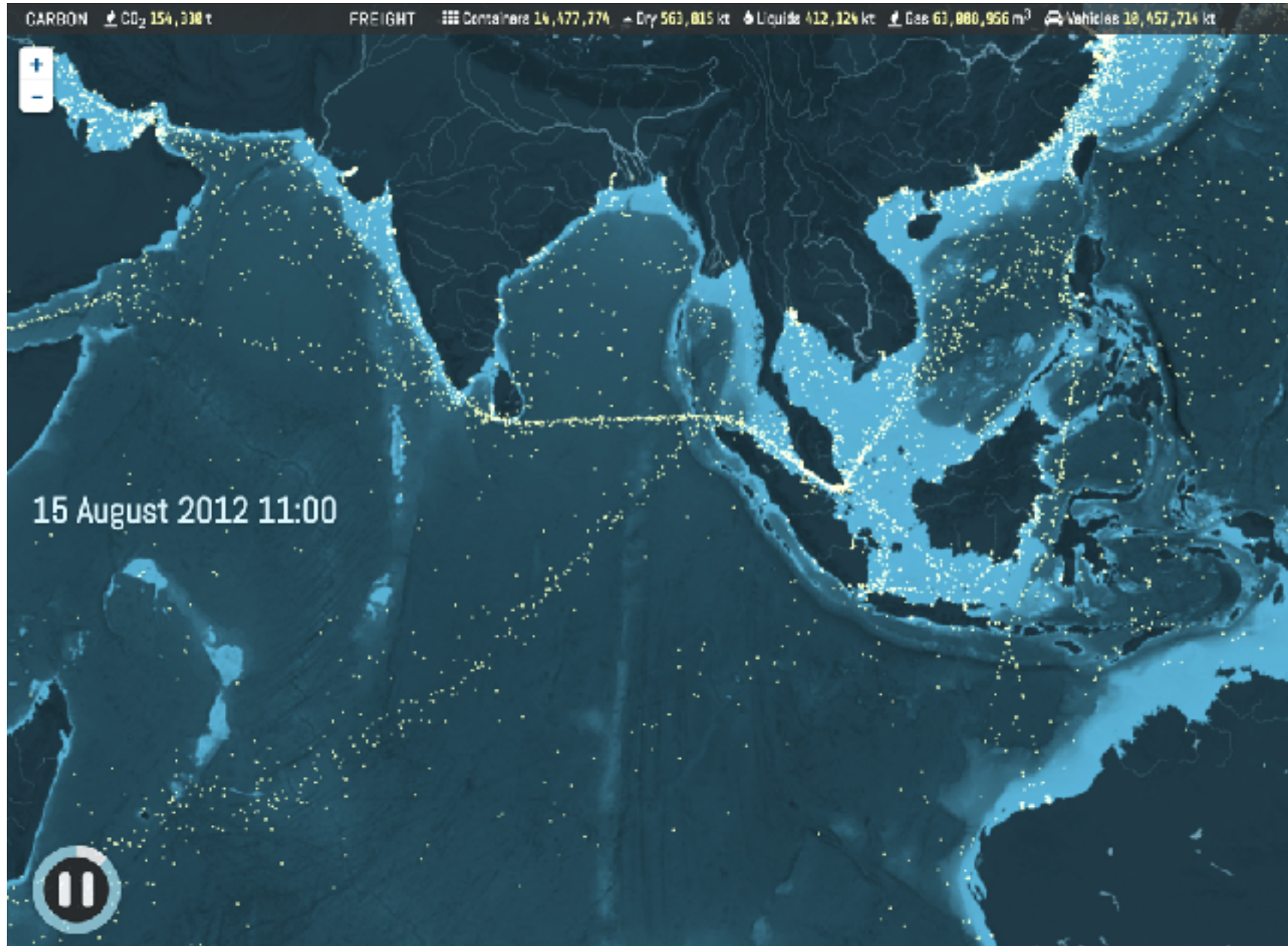
가장 많이 활용되고 있는 콘텐츠 유형 중 하나



데이터시각화란?

데이터시각화 사례

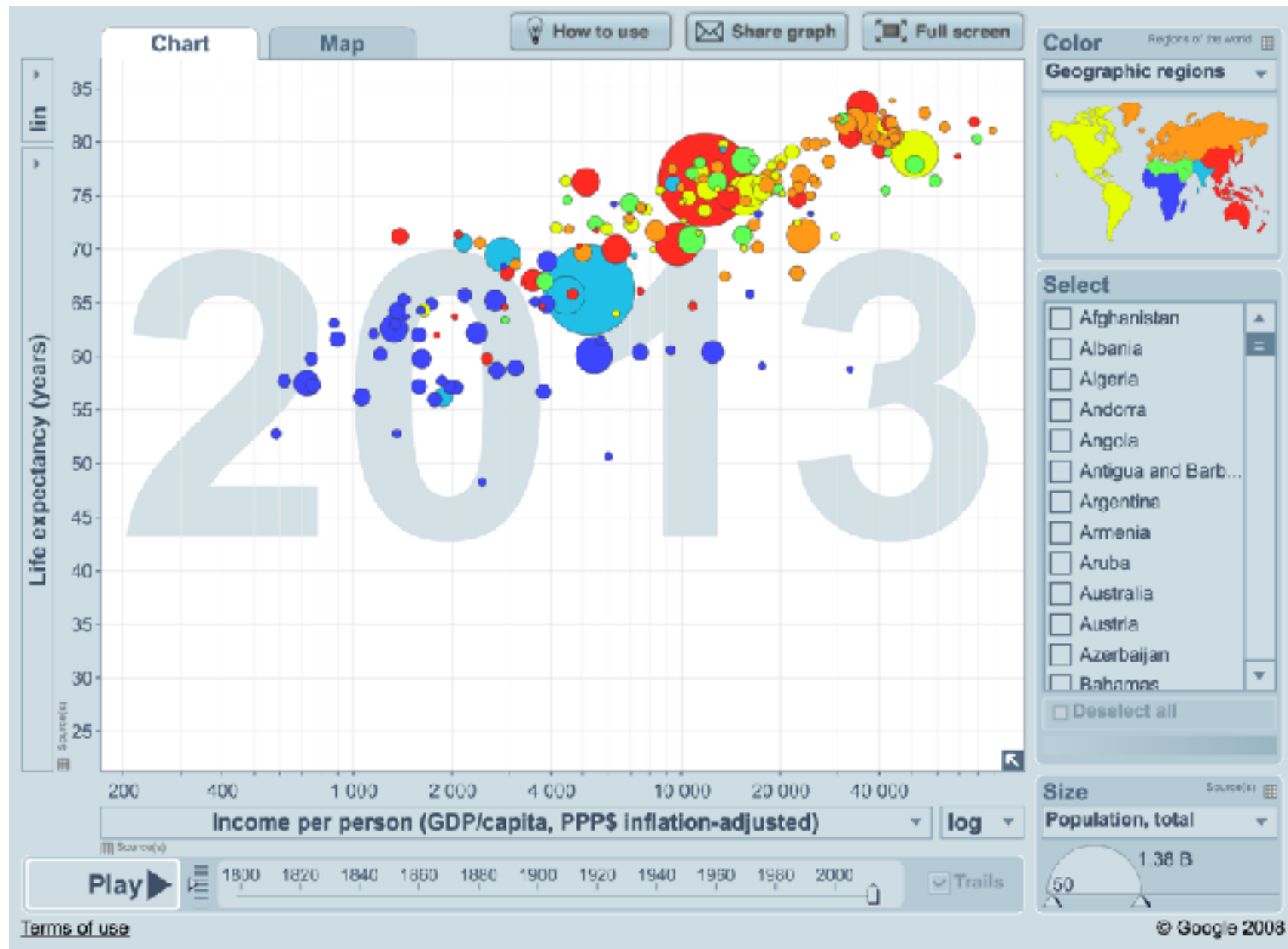
선박 운행 데이터시각화



데이터시각화란?

데이터시각화 사례

국가별 기대수명과 GDP 데이터시각화, 한스 로슬링



데이터시각화? 인포그래픽?

데이터시각화와 인포그래픽의 차이

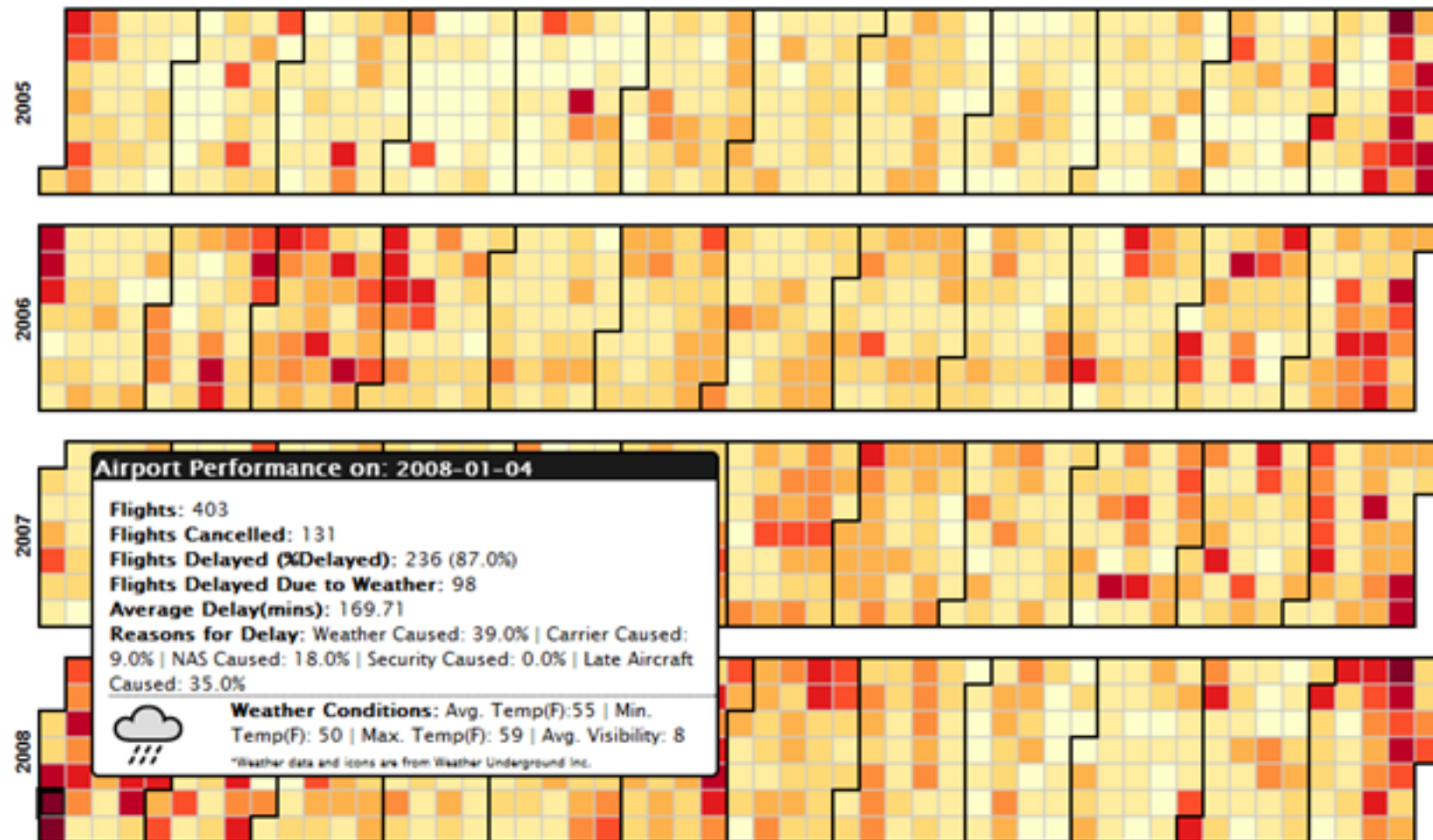
샌프란시스코 공항 출발 지연 일별 데이터 시각화

Percent of Delayed Flights Departing From **SFO** Airport (2005-2008)

For this analysis, flights departing 15 minutes or more past their scheduled time are considered as delayed flights.

***Click on a cell to get summary statistics.

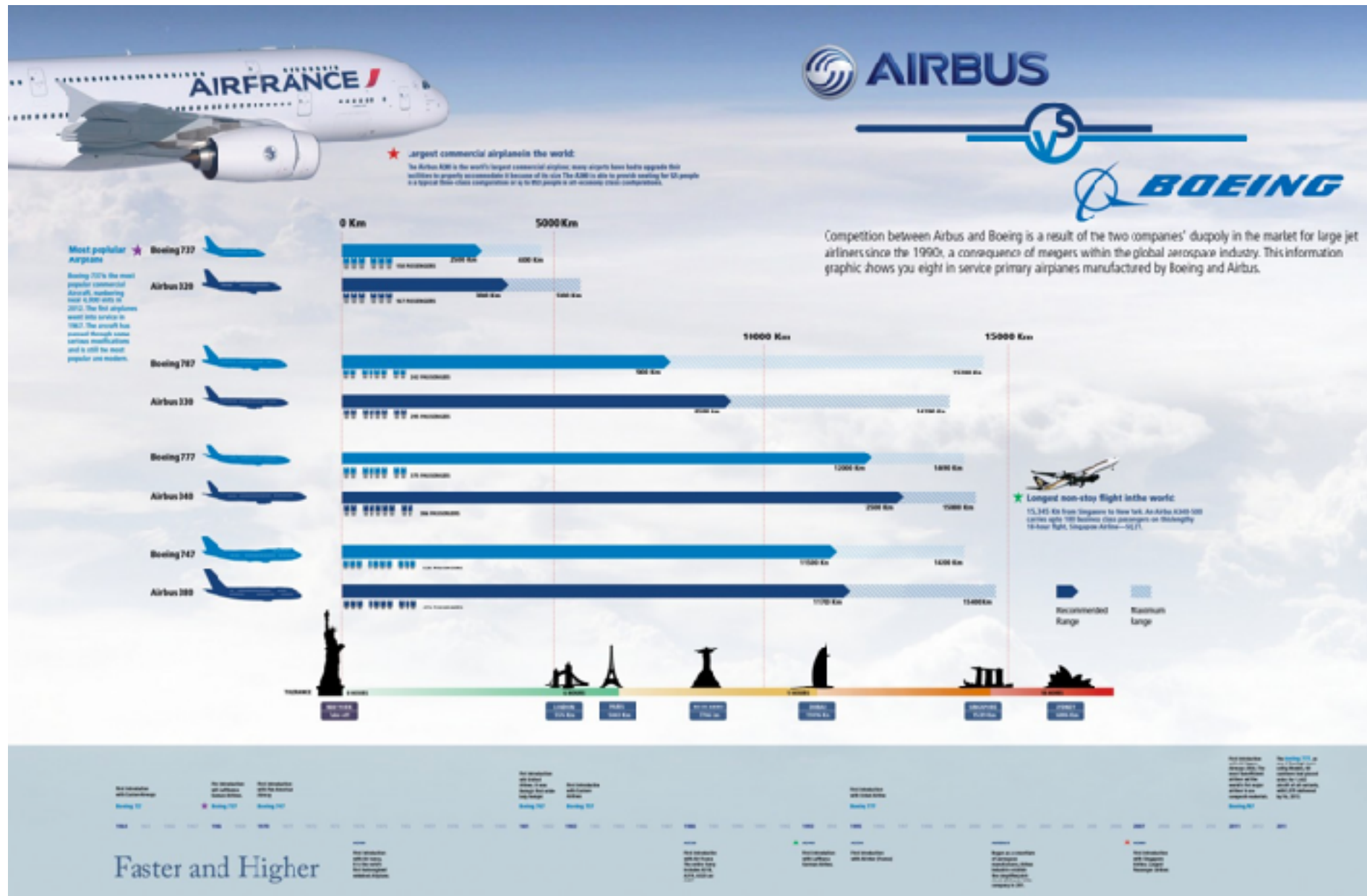
Select Airport:



데이터시각화? 인포그래픽?

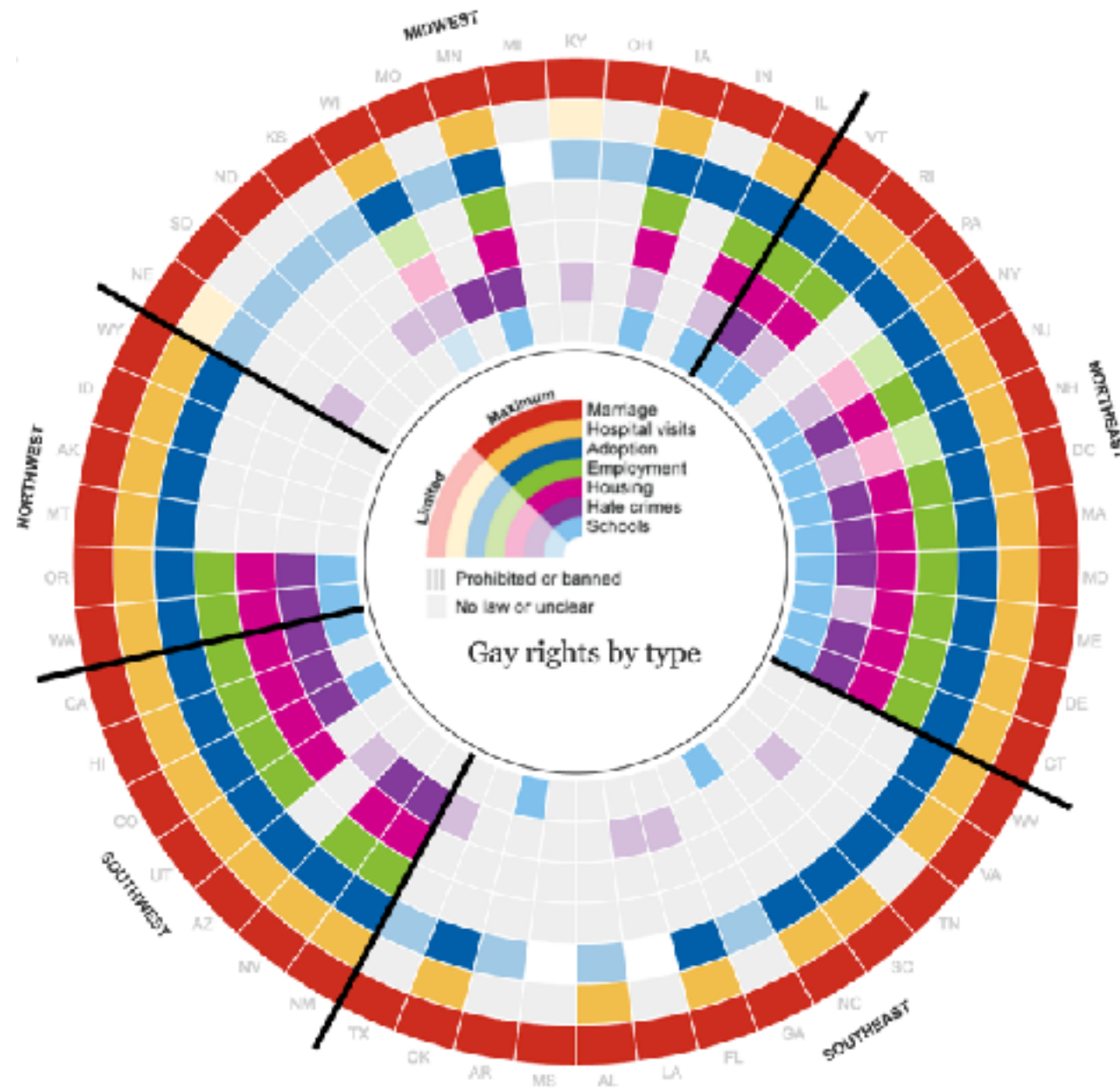
데이터시각화와 인포그래픽의 차이

Boeing vs. Airbus



데이터시각화와 미디어

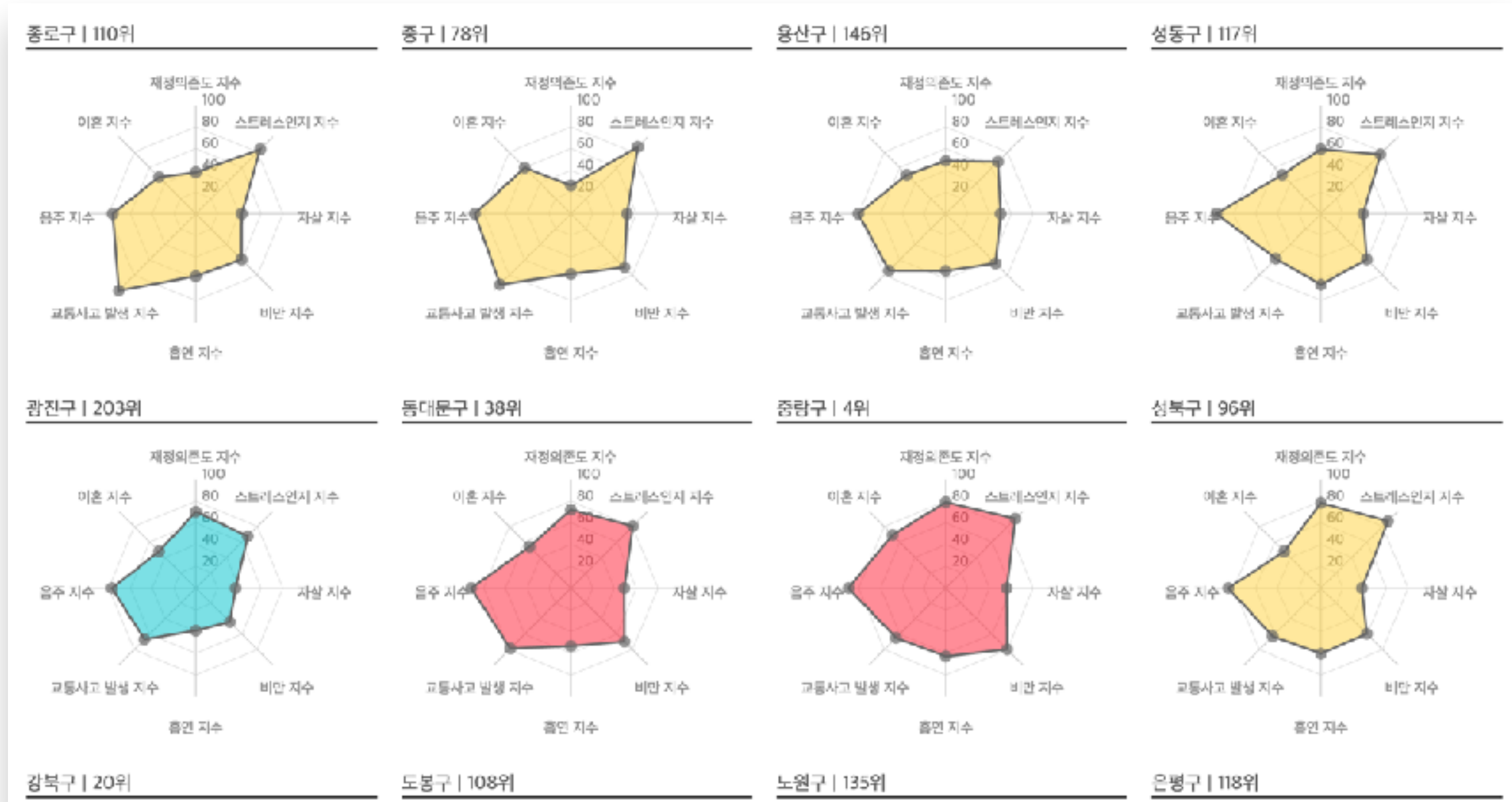
미디어에서 데이터시각화 활용 사례



The Guardian | Gay rights in the US, state by state

데이터시각화와 미디어

미디어에서 데이터시각화 활용 사례

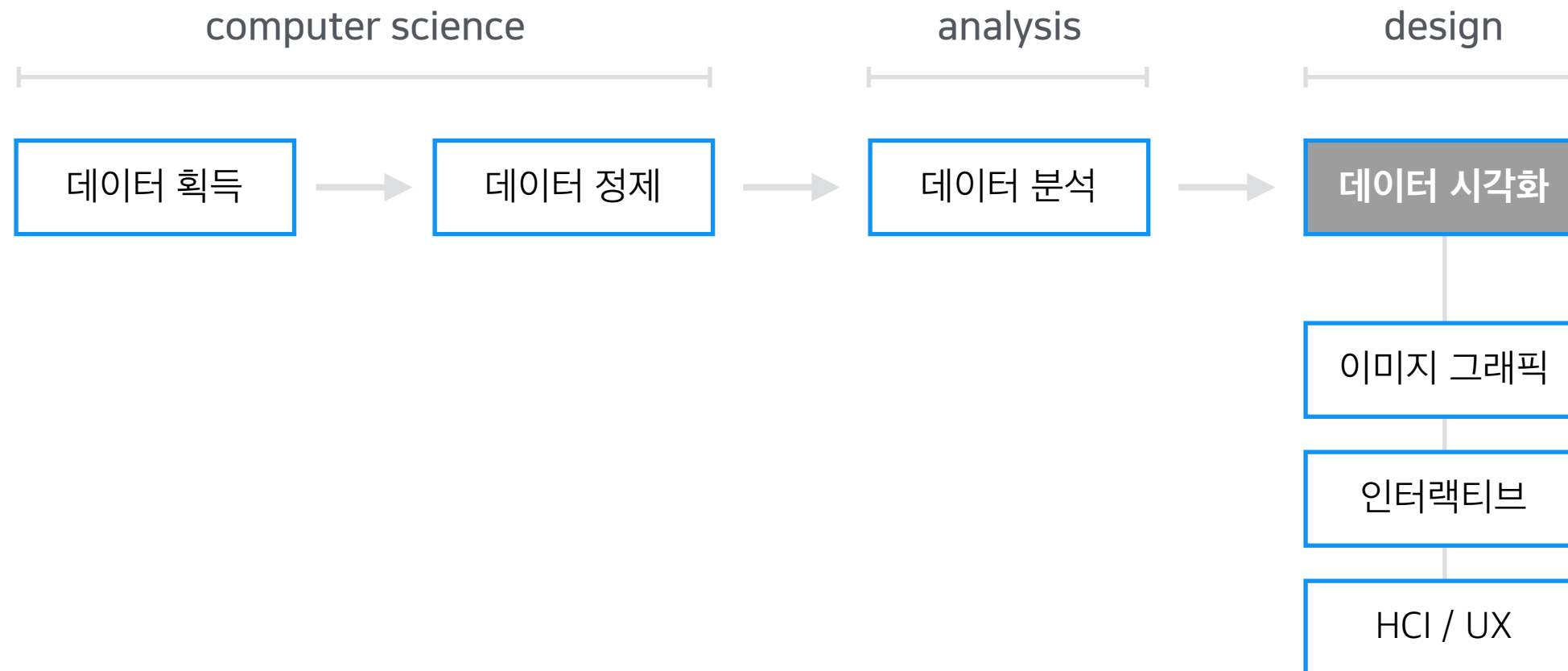


SBS 데이터저널리즘팀 | 한국에서 가장 살기 힘든 지역... '힘내라 도시'

데이터시각화와 방법론 (Process)

데이터시각화 구현의 보편적 프로세스

보편적인 데이터시각화 프로세스



데이터시각화 및 저널리즘 팀

국내 데이터저널리즘팀 현황 (2016. 7월 현재)

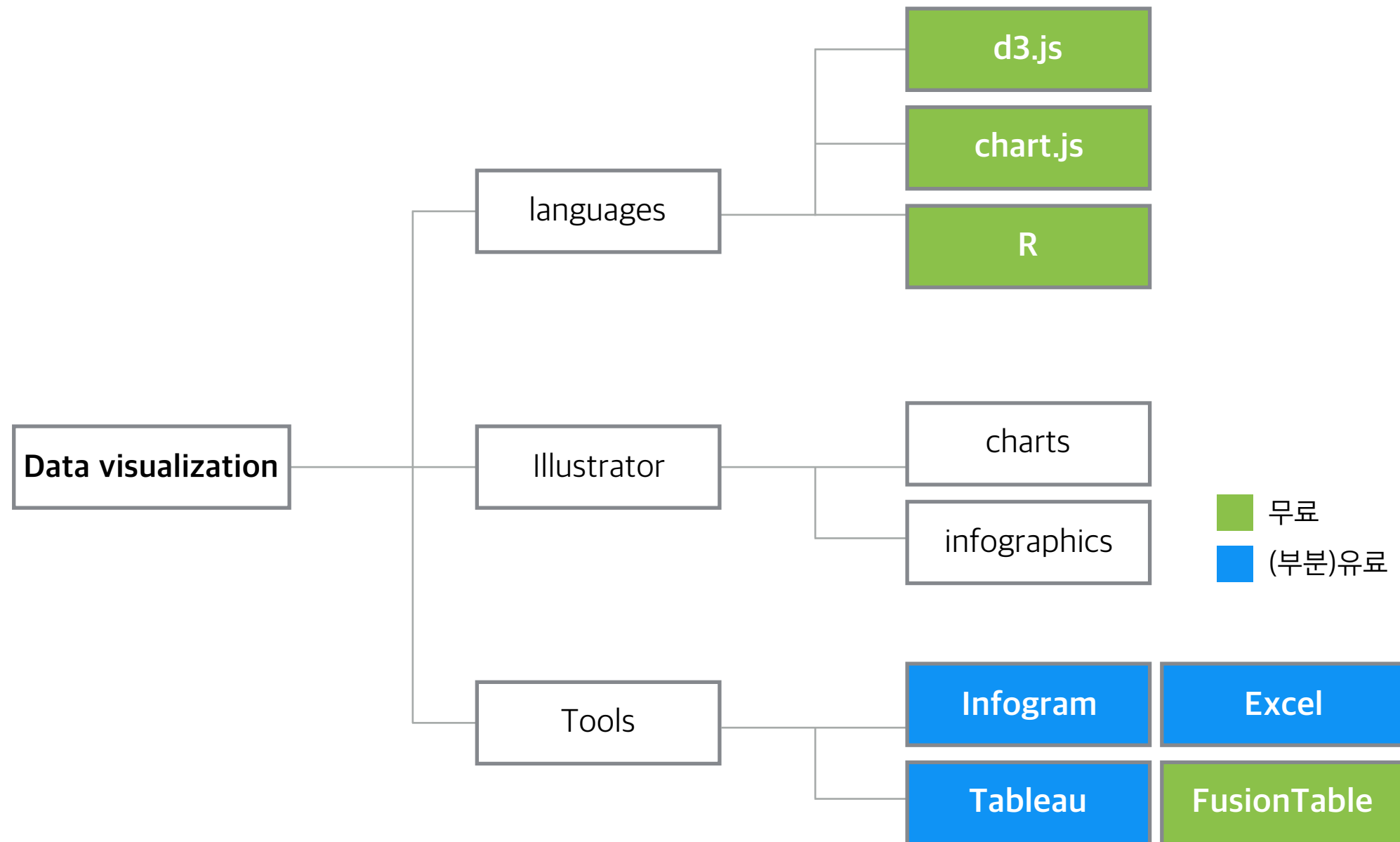
국내 데이터저널리즘팀 현황

미디어	구성	시각화	비고
SBS 데이터저널리즘팀 (마부작침)	기자(2), 분석(1), 개발겸디자이너(1), 인턴(1)	시각화 전문가 채용	
뉴스타파 데이터팀	에디터(1), 기자(4)		
KBS 데이터저널리즘팀 (데이터랩)	기자(2), 개발(1), 리서치(2), 디자인(1)	시각화 전문가 채용	
YTN 데이터저널리즘팀	기자(2), 분석(1)		
한겨레 데이터	기자(1)		Closed
중앙일보 데이터저널리즘팀			

데이터시각화 방법론 (tech)

데이터시각화 구현의 기술적 접근 방법

데이터시각화는 몇 가지 전문적인 언어와 툴을 활용



데이터시각화 방법론 (Programming, 1/3)

Long Data 구조는 프로그래밍 시각화의 재료

데이터 정제 -> 데이터 프로그래밍 -> 완성

2014년 3/4분기 인구이동(동별), 서울통계

gu	dong	code	in	out
종로구	사직동	1	1665	1872
종로구	삼청동	1	431	538
종로구	부암동	1	1668	1802
종로구	평창동	1	2534	2559
종로구	무악동	1	1249	1256
종로구	교남동	1	849	3676
종로구	가회동	1	671	845

데이터시각화 방법론 (Programming, 2/3)

Long Data 구조는 프로그래밍 시각화의 재료

데이터 정제 -> 데이터 프로그래밍 -> 완성

```
var xScale = d3.scale.linear()
    .domain([0, (d3.max(data, function(d) { return d.in; }))]
    .range([padding, w - padding * 2]);

var yScale = d3.scale.linear()
    .domain([0, (d3.max(data, function(d) { return d.out; }))]
    .range([h - padding, padding]);

//Define X axis
var xAxis = d3.svg.axis()
    .scale(xScale)
    .orient("bottom")
    .ticks(10);

//Define Y axis
var yAxis = d3.svg.axis()
    .scale(yScale)
    .orient("left")
    .ticks(10);
```

데이터시각화 방법론 (Programming, 3/3)

Long Data 구조는 프로그래밍 시각화의 재료

데이터 정제 -> 데이터 프로그래밍 -> 완성

서울시의 동별 전입·전출 현황

2014년 3/4분기의 서울시 동별 전입·전출 현황입니다. 마우스를 올리면 자세한 정보를 보실 수 있습니다

데이터: 2014년 3/4분기 연구이동(동별), 서울통계

