

# 제6부 데이터 시각화의 기본

# AGENDA

- 데이터 시각화
- 대시보드
- 데이터를 시각화 하는 이유
- 데이터 시각화 방법

# 데이터 시각화

## ▪ 정의

- 데이터를 분석하고 가공해 사용자가 시각적으로 데이터를 보게 되는 전체 과정
- 숫자가 아닌 시각적 요소로 바꾸서 받아들이는 사람이 보다 쉽고 빠르게 이해할 수 있도록 만드는 **기술이자 커뮤니케이션 도구**

## ▪ 필요성

- 의사결정 단계에서 데이터에 의존하는 경향이 점점 강해지고 있음
- 시각적 추상화 같은 작업이 없이는 데이터를 제대로 이해하기 어려움
- 단순히 숫자로 나열된 형태보다는 그림의 형태로 사물이나 현상을 파악하는 것이 더 쉬움
- 데이터 시각화 작업 없이는 데이터에 존재하는 패턴과 변칙들에 숨겨진 비효율적 요소들을 찾아내기란 거의 불가능함

## ▪ 고려사항

- 방대한 데이터를 인지할 수 있는 시각화 형태로 시각화 하기 위해서는 **많은 처리과정과 계산이 필요함**
- 데이터 시각화의 효과적인 측면을 위해서 언제 어디까지 누구를 위해 데이터를 시각해야 하는지 그 **범위와 대상을 최대한 좁혀서 구체적으로** 만들어야 함
- 차트를 보여주는 것이 목적이 되면 안되고 **어떤 활동 때문에 차트의 선이 오르내리는지를** 보여줘야 함

## ▪ 정의

- 대시보드는 상황을 모니터링하거나, 이해를 돕거나, 둘 다에 사용하는 데이터의 시각적 표시

## ▪ 고려사항

- 대시보드는 해당 데이터를 이해해야 하는 대상에 맞는 훌륭한 아이디어를 제시해야 해야함

## ▪ 대시보드 예시

- 지역별, 산업별, 신체 부위별 노동자의 보상 청구를 파악할 수 있는 인터랙티브한 도면
- 매주 월요일 아침 임원에게 이메일로 전송되는 핵심지표 **PDF** 파일
- 실시간으로 지원 센터 통계를 보여주는 대형 벽걸이 현황판
- 영업 관리자가 지역별 성과를 검토하고 전년 대비 올해의 영업 성과를 비교할 수 있는 모바일 앱
- 신규 온라인 서비스 개발을 위해 진행하는 **A/B** 테스트 결과

## 데이터를 시각화 하는 이유

- 아래 표는 1쌍으로 이루어진 4개 그룹의 숫자가 있다
- 아래 표에서 무엇이 보이는가? 숫자들 간에 패턴이나 추세의 차이를 구별할 수 있는가?

I		II		III		IV	
x	y	x	y	x	y	x	y
10	8,04	10	9,14	10	7,46	8	6,58
8	6,95	8	8,14	8	6,77	8	5,76
13	7,58	13	8,74	13	12,74	8	7,71
9	8,81	9	8,77	9	7,11	8	8,84
11	8,33	11	9,26	11	7,81	8	8,47
14	9,96	14	8,1	14	8,84	8	7,04
6	7,24	6	6,13	6	6,08	8	5,25
4	4,26	4	3,1	4	5,39	19	12,5
12	10,84	12	9,13	12	8,15	8	5,56
7	4,82	7	7,26	7	6,42	8	7,91
5	5,68	5	4,74	5	5,73	8	6,89

표: 4가지 숫자 그룹 표: 숫자의 의미를 알 수 있는가?

## 데이터를 시각화 하는 이유

- 통계적 속성을 고려하여 데이터를 확인할 수 있다.
- 아래 표는 X,Y가 각각 합계, 평균, 표준 편차가 같다.
- 같은 패턴의 데이터라고 볼 수 있는가?

	I		II		III		IV	
	x	y	x	y	x	y	x	y
	10	8,04	10	9,14	10	7,46	8	6,58
	8	6,95	8	8,14	8	6,77	8	5,76
	13	7,58	13	8,74	13	12,74	8	7,71
	9	8,81	9	8,77	9	7,11	8	8,84
	11	8,33	11	9,26	11	7,81	8	8,47
	14	9,96	14	8,1	14	8,84	8	7,04
	6	7,24	6	6,13	6	6,08	8	5,25
	4	4,26	4	3,1	4	5,39	19	12,5
	12	10,84	12	9,13	12	8,15	8	5,56
	7	4,82	7	7,26	7	6,42	8	7,91
	5	5,68	5	4,74	5	5,73	8	6,89
SUM	99,00	82,51	99,00	82,51	99,00	82,50	99,00	82,51
AVG	9,00	7,50	9,00	7,50	9,00	7,50	9,00	7,50
STDEV	3,32	2,03	3,32	2,03	3,32	2,03	3,32	2,03

표: 4가지 숫자 그룹 표: 숫자의 의미를 알 수 있는가?

## 데이터를 시각화 하는 이유

- 데이터를 차트로 옮겨 시각화해보면 통계 값으로 알 수 없었던 결과를 확인할 수 있다.
- 이 데이터는 앤스콤의 4가지 그룹이라는 숫자 세트이다. "수치 계산이 정확하며, 그래프는 대략 적"이라는 일반적인 관념에 맞섰다.

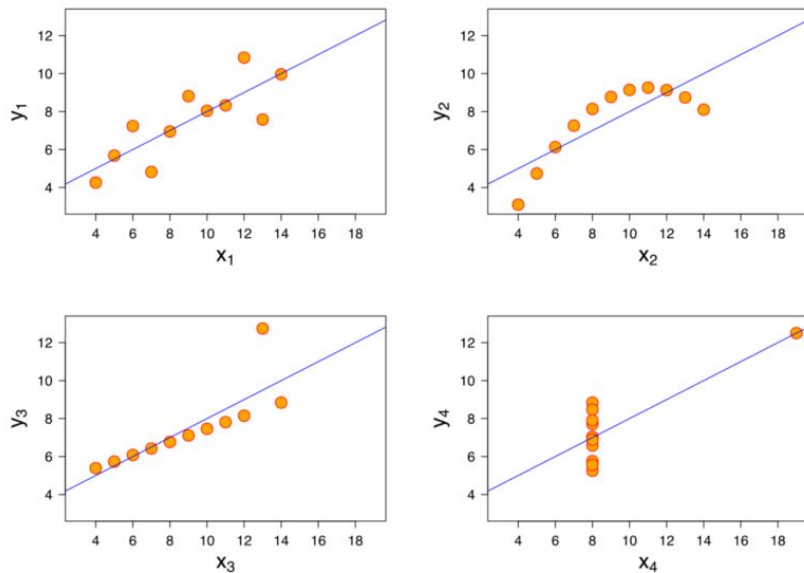


표 : 4가지 숫자 그룹 표 : 숫자의 의미를 알 수 있는가?

## 데이터를 시각화 하는 이유

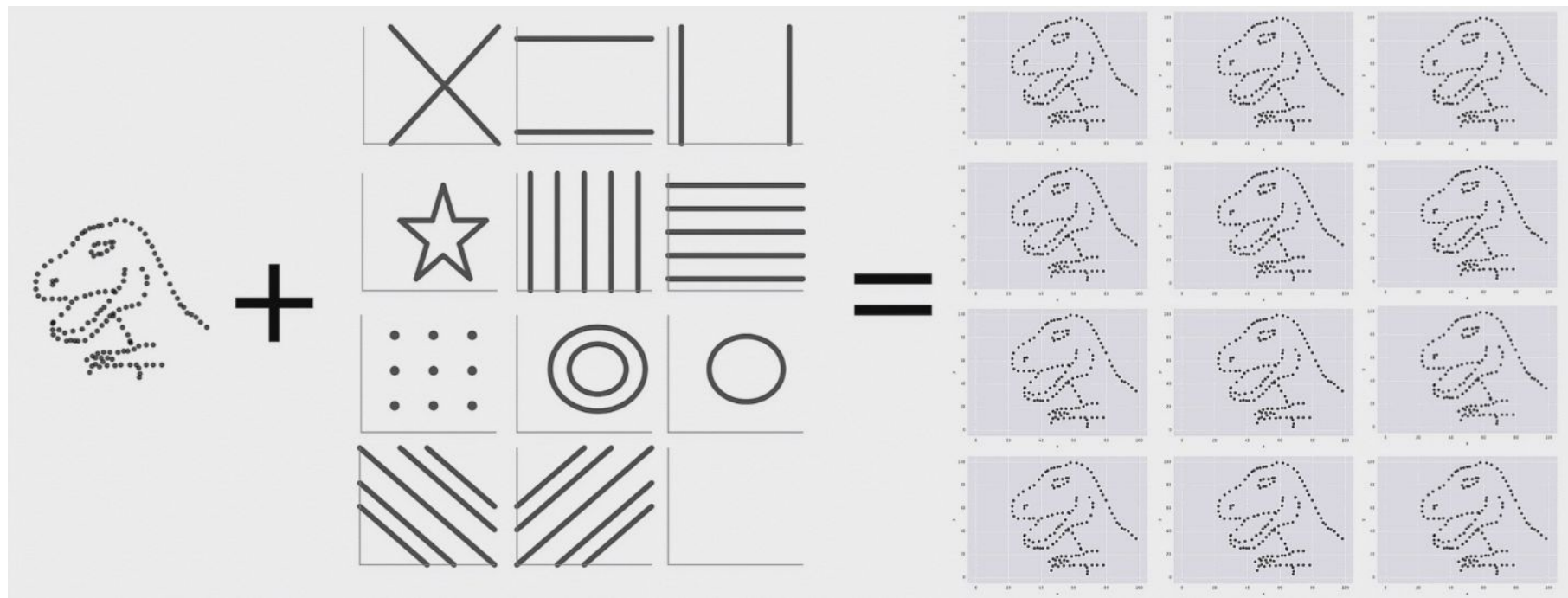
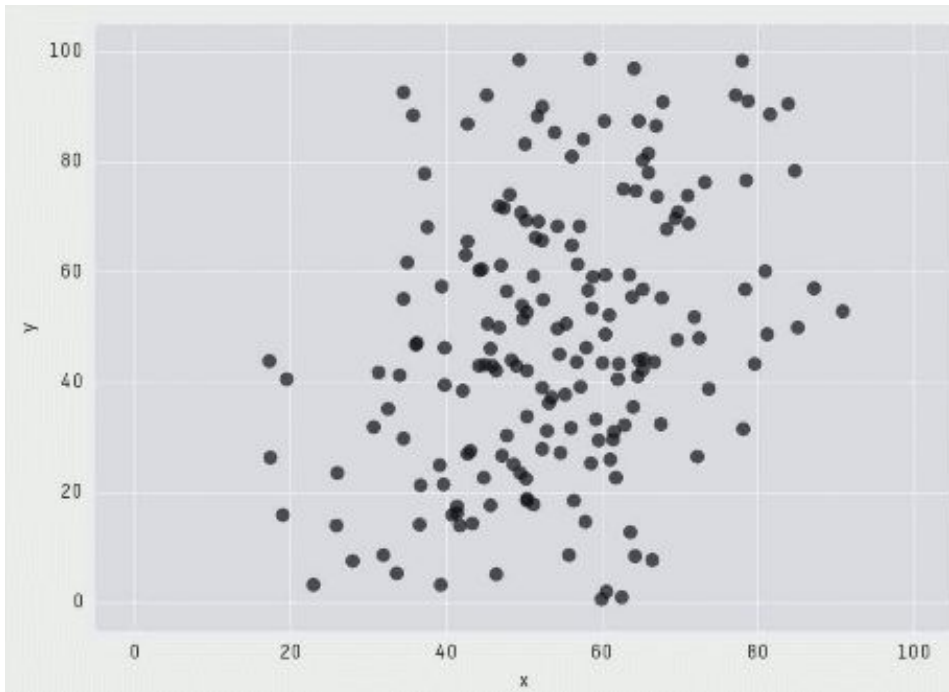


표:공룡 모양의 데이터를 다른 모양으로 변형한 데이터



## 데이터를 시각화 하는 이유

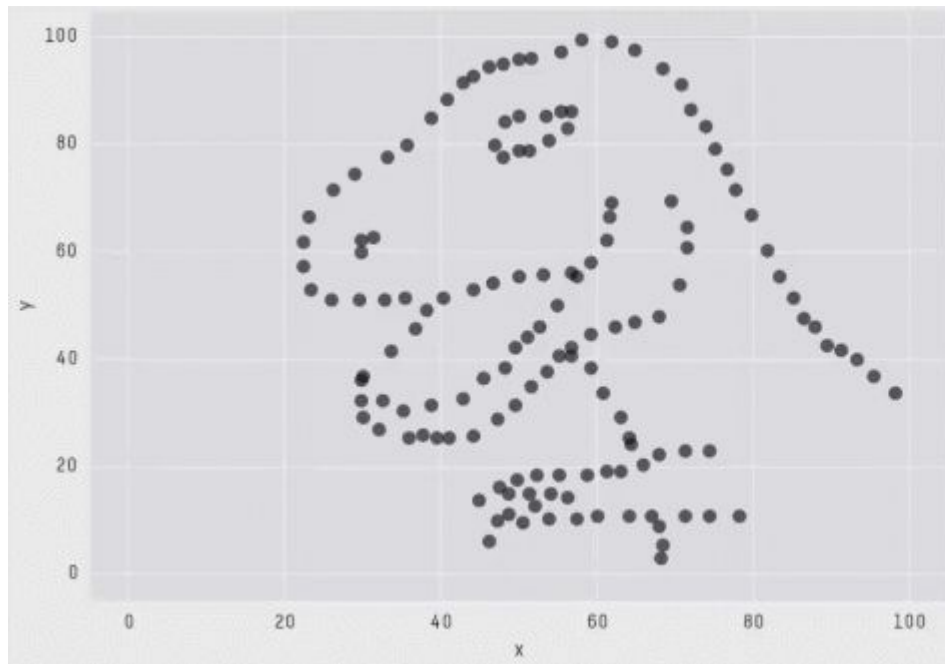
- 데이터의 패턴은 다르지만 통계 값이 같다



X Mean: 54.0236753  
Y Mean: 48.0970794  
X SD : 14.5298540  
Y SD : 24.7943127  
Corr. : +0.3280926

## 데이터를 시각화 하는 이유

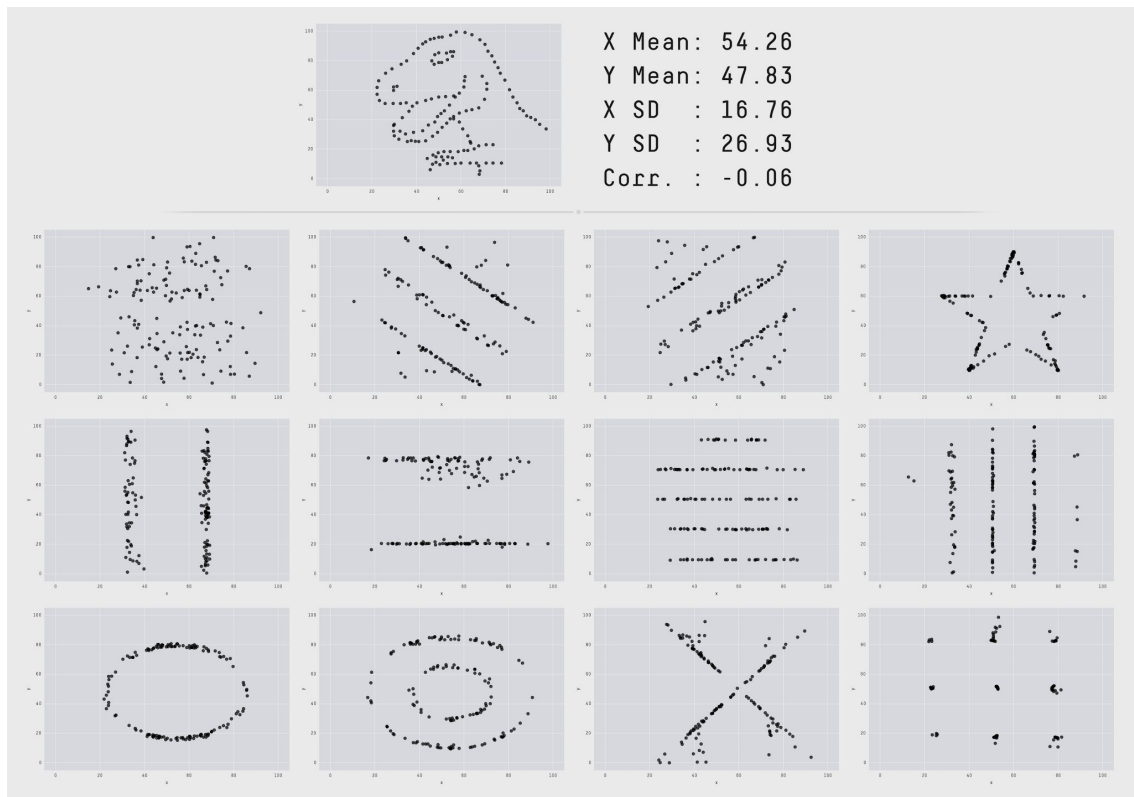
- 데이터의 패턴은 다르지만 통계 값이 같다



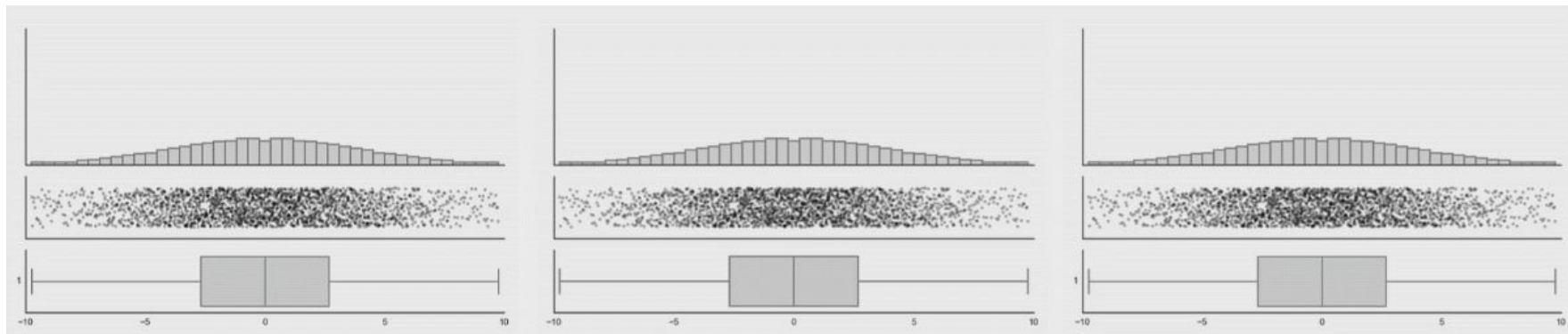
X Mean: 54.2659224  
Y Mean: 47.8313999  
X SD : 16.7649829  
Y SD : 26.9342120  
Corr. : -0.0642526

# 데이터를 시각화 하는 이유

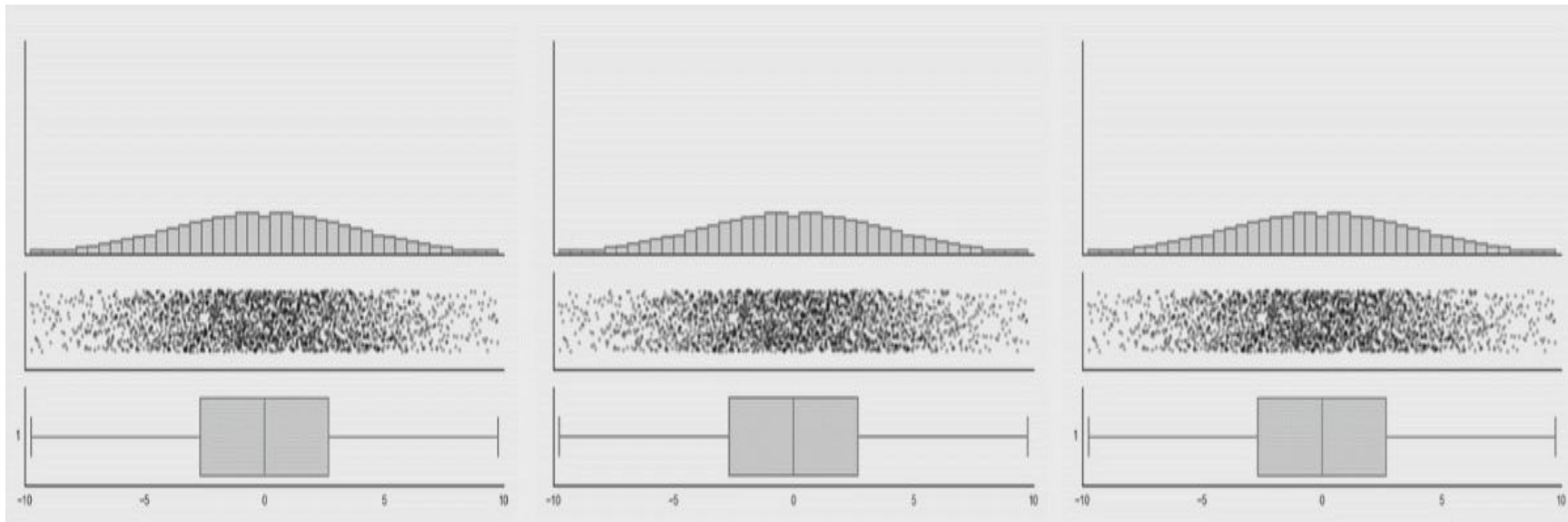
- 데이터의 패턴은 다르지만 통계 값이 같다



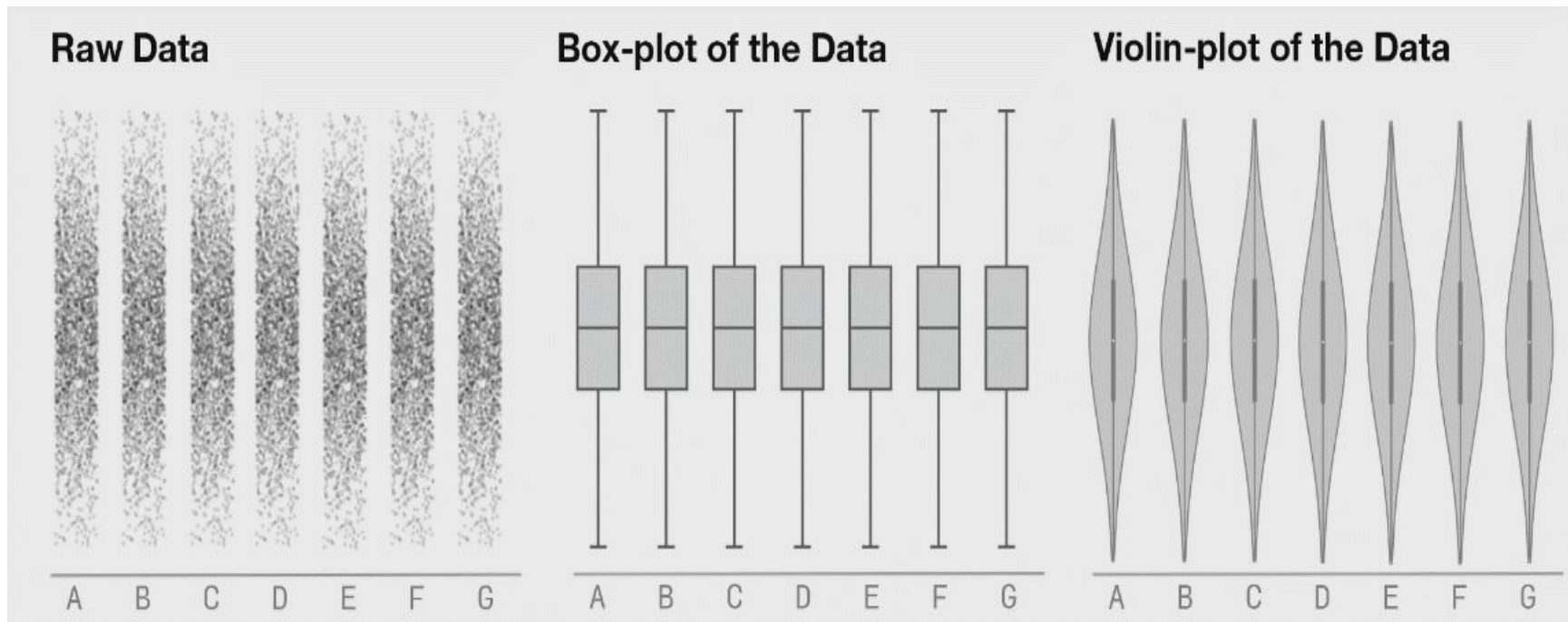
## 데이터를 시각화 하는 이유



## 데이터를 시각화 하는 이유



## 데이터를 시각화 하는 이유



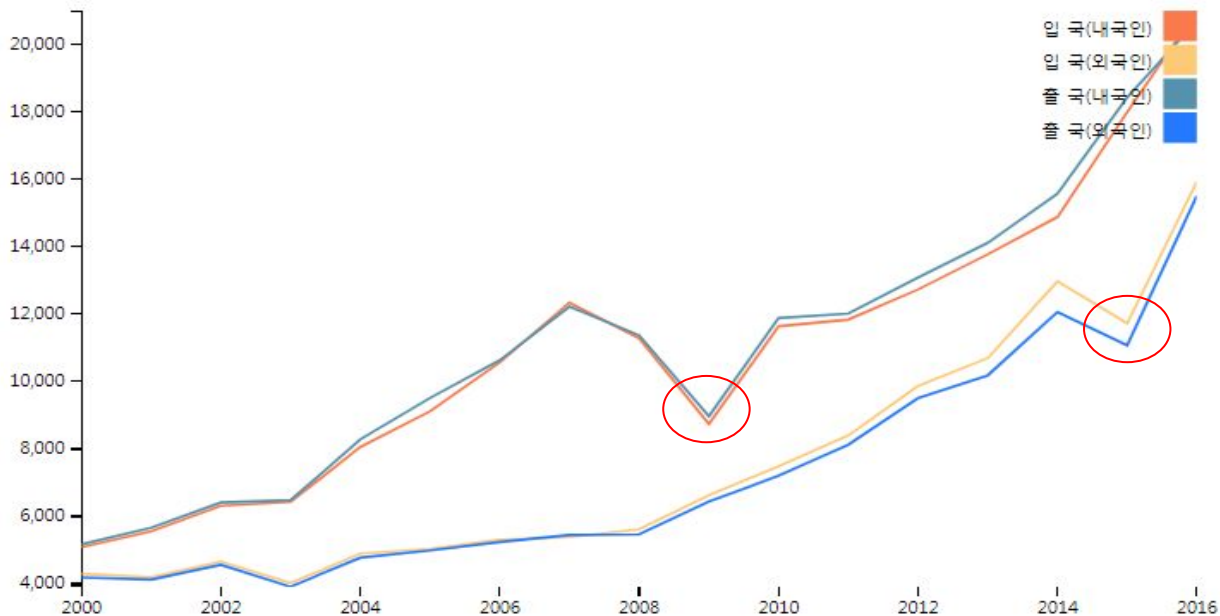
## 데이터를 시각화 하는 이유

- 연도별 입출국하는 내국인과 외국인의 추세가 어떻게 변하는지 잘 보이는가? 어떤 특징이 보이는가?

구분	입 국(내국인)	입 국(외국인)	출 국(내국인)	출 국(외국인)
2000	5080	4296	5169	4189
2001	5558	4191	5662	4123
2002	6312	4659	6410	4565
2003	6427	4020	6480	3912
2004	8057	4897	8281	4772
2005	9119	5035	9518	4990
2006	10569	5309	10635	5237
2007	12348	5382	12229	5449
2008	11291	5615	11376	5466
2009	8744	6626	8975	6439
2010	11644	7487	11889	7206
2011	11842	8403	12020	8128
2012	12735	9869	13090	9507
2013	13781	10699	14123	10187
2014	14888	12977	15578	12065
2015	17998	11730	18431	11076
2016	21012	15934	20711	15510

## 데이터를 시각화 하는 이유

- 2008년~2009년 내국인의 출입국 수 감소
- 2014년~2015년 외국인 출입국 수 감소
- 특이사항을 찾으면 결과에 대한 원인을 찾기 위해 드릴다운 분석을 해야함



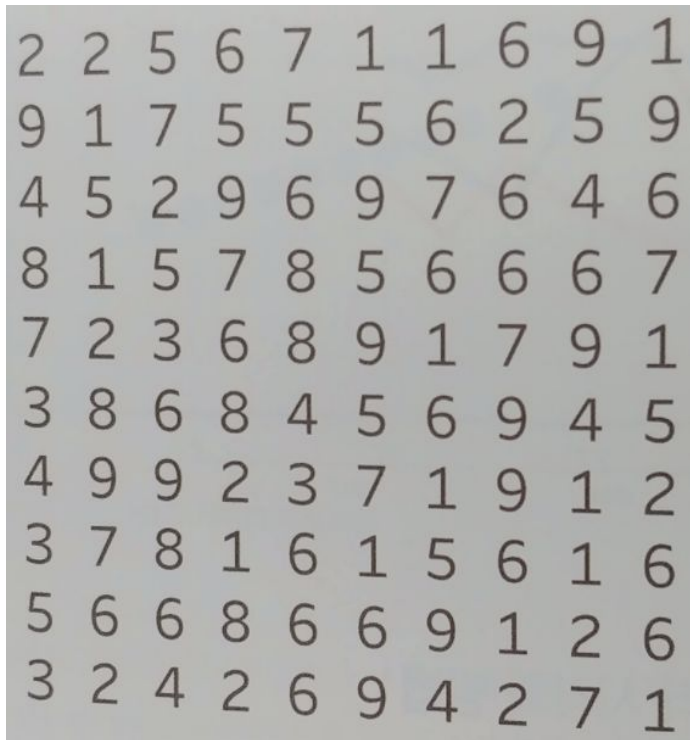


## 데이터 시각화 방법

- 시각화를 할 때는 전주의적 속성이라는 것에 집중 해야함
- 전주의적(**preattentive**) 속성
  - 사물에 주의를 주기 전에 인지할 수 있는 속성을 의미
- 사람은 어떤 것에 관심을 두기 전에 뇌는 **1밀리초**, 즉 **1,000분의 1초** 내에 처리하기 때문임
- 그래서 데이터를 시각화할 때 데이터를 캔버스 위에 표 또는 그림(차트)으로 바꿔야 함
- 이때 어떤 표 또는 그림(차트)이 데이터의 의미를 가장 잘 전달하는지 이해해야 함

## 데이터 시각화 방법 - 예시

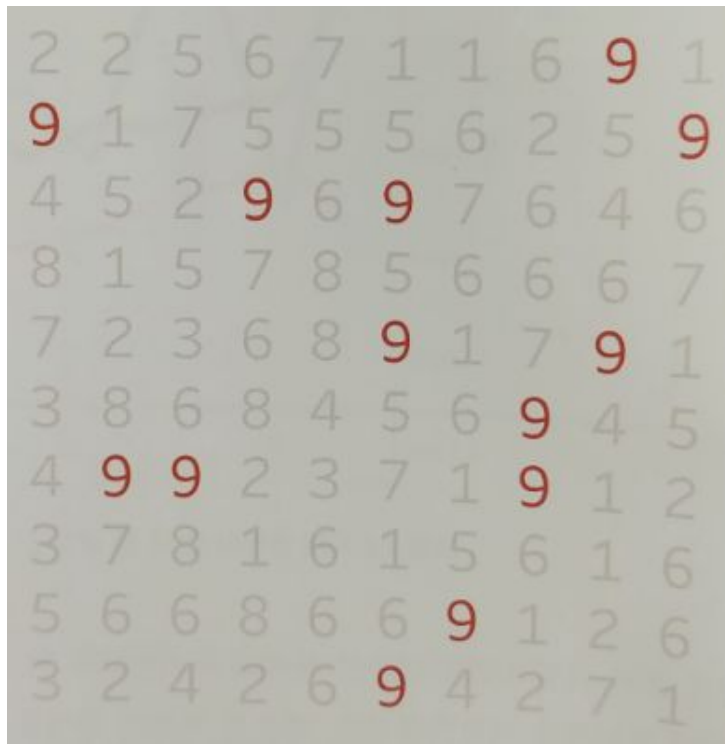
- 아래 그림에서 9가 몇 개가 있는가? 전체 숫자를 읽으며 찾아볼 수 있지만 시간이 많이 걸린다.



2	2	5	6	7	1	1	6	9	1
9	1	7	5	5	5	6	2	5	9
4	5	2	9	6	9	7	6	4	6
8	1	5	7	8	5	6	6	6	7
7	2	3	6	8	9	1	7	9	1
3	8	6	8	4	5	6	9	4	5
4	9	9	2	3	7	1	9	1	2
3	7	8	1	6	1	5	6	1	6
5	6	6	8	6	6	9	1	2	6
3	2	4	2	6	9	4	2	7	1

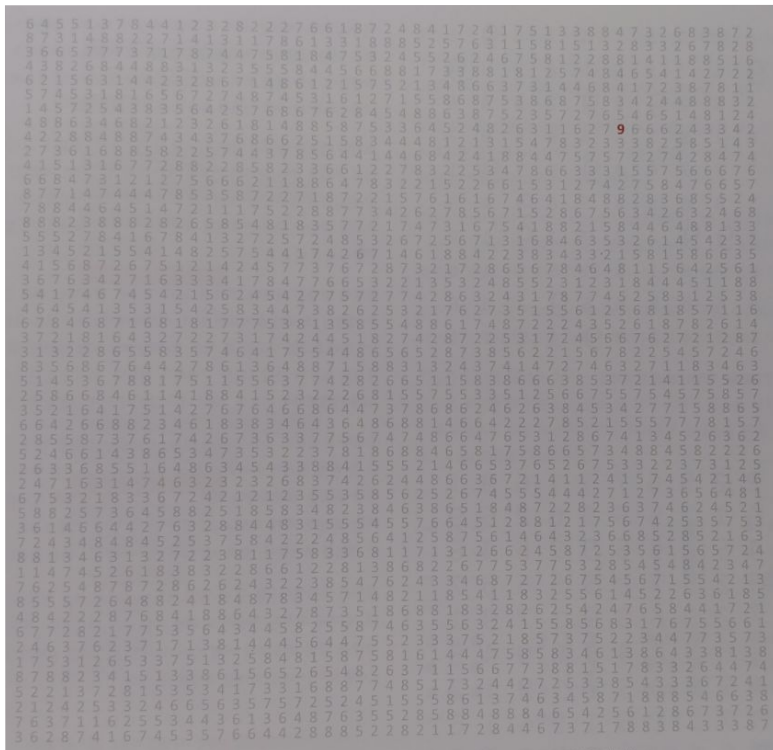
## 데이터 시각화 방법 - 예시

- 숫자의 색을 바꾸면 쉽게 원하는 숫자를 찾을 수 있다.



## 데이터 시각화 방법 - 예시

- 색을 바꾸면 2,500개의 숫자 중에서도 9가 어디에 있는지 쉽게 찾을 수 있다.



## 데이터 시각화 방법 - 예시

- 일반적으로 사람은 어떤 장면이나 차트를 관찰할 때, 이 속성을 250밀리초 내에 처리함
- 아래 그림과 같이 9의 크기를 다른 숫자와 다르게 해도 쉽게 원하는 숫자를 찾을 수 있음



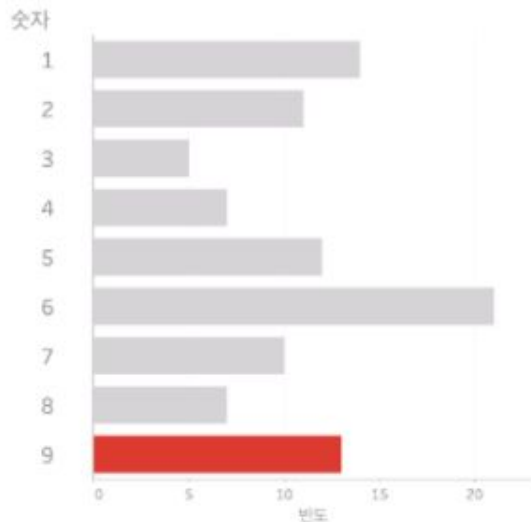
## 데이터 시각화 방법 - 예시

- 만약 모든 숫자의 빈도를 세기 위해서 숫자마다 다른 색을 입힌다면 아래 그림과 같이 **단일한 색의 전주의적 속성이 무너짐**
- 시각화에서 **1가지 색을 사용하면 1가지 카테고리**를 강조할 때 아주 효과적임
- 아래 그림과 같이 **9개의 카테고리를 다른 색으로 강조하면 구분이 어려움**

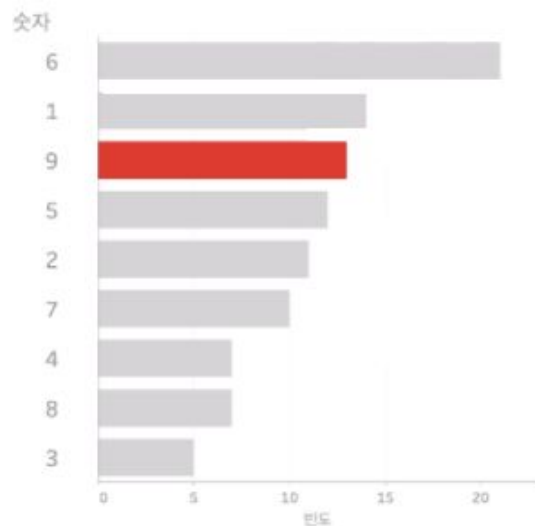
2	2	5	6	7	1	1	6	9	1
9	1	7	5	5	5	6	2	5	9
4	5	2	9	6	9	7	6	4	6
8	1	5	7	8	5	6	6	6	7
7	2	3	6	8	9	1	7	9	1
3	8	6	8	4	5	6	9	4	5
4	9	9	2	3	7	1	9	1	2
3	7	8	1	6	1	5	6	1	6
5	6	6	8	6	6	9	1	2	6
3	2	4	2	6	9	4	2	7	1

## 데이터 시각화 방법 - 예시

- 여러가지 숫자의 개수를 세기 위해서는 분류를 해야함
- 시각화는 통찰을 얻기 위해 빈도 등 분류 체계의 규약화가 핵심
- 표를 이용하지 않고 숫자마다 빈도를 파악할 수 있는 규약화를 해야함
- **가장 효과적인 방법은 길이를 사용**하는 것임
- 막대 차트를 사용하면 **빈도를 길이로 나타내어 표현**할 수 있음



9가 13개 있음



정렬을 이용해 숫자별 빈도가 높은 순서를 알 수 있음

## 데이터 시각화 방법 - 전주의적 속성

- 데이터 시각화에서 사용되는 **12가지** 전주의적 속성
- 전주의적 속성은 차트에서 데이터의 규약화 방법을 제공함
- 시각 체계가 얼마나 강력한지 알아보고 데이터를 효과적으로 표시할 때 사용할 수 있는 시각적 요소를 살펴봄



방향



길이



너비



크기



형태



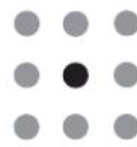
곡률



표시 추가



둘러싸기



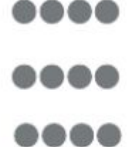
색상



색조



위치



공간적 그룹핑



## 데이터 시각화 방법 - 데이터 유형

- 시각화를 하기 위해서는 전주의적 속성과 데이터 유형을 조합해야 함
- 데이터는 범주형 데이터, 순서형 데이터, 정량적 데이터의 3가지 유형이 있음
- 범주형 데이터
  - 범주형 데이터는 사물을 나타냄
  - 이 사물은 어떤 숫자 값도 없는 상호 배타적인 레이블을 의미함
  - 명목형 데이터(nominal)라고도 함  
(예 : [남자, 여자], [국어, 수학, 과학], [성공, 실패])
- 순서형 데이터
  - 분명한 순서가 있다는 점을 제외하면 범주형 데이터와 유사함
  - 각 값 사이의 간격은 일정하지 않음  
(예 : 나이 기준 [영아, 유아, 어린이, 청소년, 청년, 장년, 노인], 성적 기준 [A, B, C, D, E, F])

## 데이터 시각화 방법 - 데이터 유형

- 정량적 데이터

- 정량적(수치) 데이터는 숫자로서 측정, 분류 가능한 데이터  
(예 : 키, 몸무게, 나이, 매출액)
- 정량적 데이터는 이산 데이터와 연속 데이터의 2가지 방법으로 표현할 수 있음
- **이산 데이터**는 미리 정의된 정확한 포인트로 제시되어 데이터 값 '사이'에는 아무것도 없음  
(예: 나이 [20,21,22,23,24,25](O) , 25.4(X))
- **연속 데이터**는 가능한 중간 값의 무한한 수가 있음  
(예: 키 [150.5, 160.44, 170.4345....])

# 데이터 시각화 방법 - 차트의 데이터 규약화

## 전주의적 속성과 3가지 데이터 유형을 조합한 결과

〈스타트렉: 더 넥스트 제너레이션〉

에피소드별 IMDB.com 평점



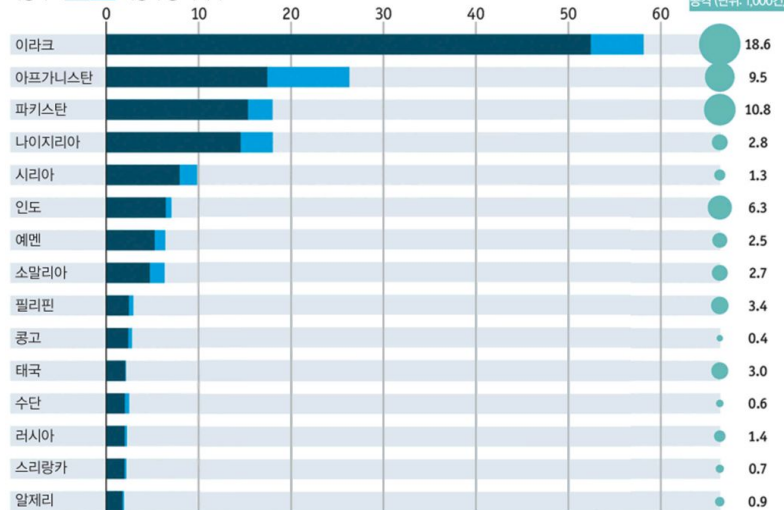
데이터	데이터 유형	규약화	참고
에피소드	범주형 데이터	위치	에피소드마다 점으로 표시된다. 각 점은 캔버스에 정해진 위치가 있다.
에피소드 번호	순서형 데이터	위치	X축은 시즌별로 에피소드의 번호를 보여준다.
시즌	순서형 데이터	색	각 시즌마다 다른 색으로 표시된다(색조).
		위치	시즌별로 차트에 별도의 섹션이 있다.
IMDB 평점	순서형 데이터	위치	평점이 좋은 에피소드는 Y축에서 더 높은 지점에 위치한다.
평균 시즌 평점	정량적 데이터	위치	패널마다 수평선으로 시즌별 에피소드의 평균 평점을 보여준다. 순서형 데이터 평점을 평균으로 처리해야 할지에 대해서는 논란이 좀 있다. 평점 처리시 매우 혼한 관행이라 허용 가능하다고 생각한다.

# 데이터 시각화 방법 - 차트의 데이터 규약화

## 끔찍한 기록

2003~2015년 테러 사망자가 가장 많은 나라(단위: 1,000명)

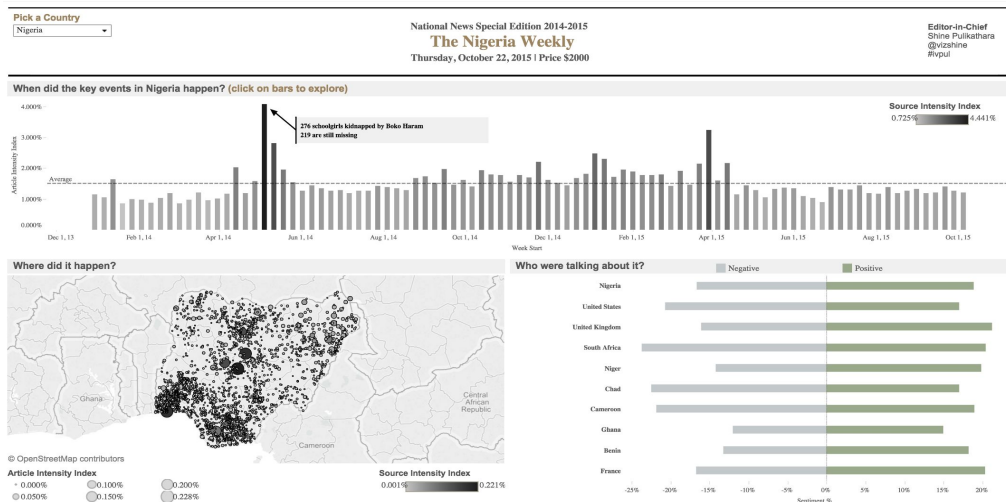
사망자 ■ 사망자 중 가해자 ■



데이터	데이터 유형	규약화	참고
나라	범주형 데이터	위치	나라마다 행별로 구분된 막대로 표시되어 있다(사망자 수 집계로 정렬).
사망자	정량적 데이터	길이	막대 길이는 사망자 수를 보여준다.
사망 유형	범주형 데이터	색	짙은 파란색은 희생자의 사망 수, 하늘색은 가해자의 사망 수를 나타낸다.
공격	정량적 데이터	크기	우측 원은 공격 건수에 따라 크기가 다르다.

# 데이터 시각화 방법 - 색상

- 색상은 데이터 시각화에서 이해해야 할 가장 중요한 요소 중 하나임에도 오용하는 경우가 많음
- 지루한 시각화에 생기를 불어넣는 목적만을 위해서 색을 사용해서는 안됨
- 훌륭한 데이터 시각화는 색을 전혀 사용하지 않아도 정보성이 있으며 충분히 아름다울 수 있음
- 색상은 의도를 갖고 사용해야 함
- 예를 들어 색으로 독자의 관심을 끌거나, 데이터 일부를 강조하거나, 다양한 카테고리를 구별할 수 있음



<그림> 나이지리아에서 발생한 주요한 사건 현황

# 데이터 시각화 방법 - 색상의 사용

- 색상은 데이터 시각화에서 **순차적**, **발산형**, **범주형**과 같은 3가지 주요 방식으로 사용해야 함
- 데이터를 **강조**하거나 중요하다는 점을 **경고**할 필요가 있을 때 아래와 같이 색상 체계별 사례를 볼 수 있음

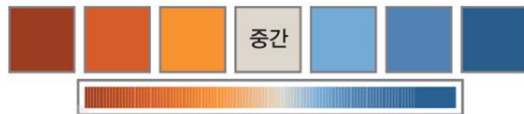
## 순차적

색상을 낮음에서 높음으로 정렬



## 발산형

중립적인 중간점이 있는 순차적인 2가지 색상



## 범주형

개별 비교를 위한 색상 대비



## 강조

원가를 강조하기 위한 색 사용



## 경고

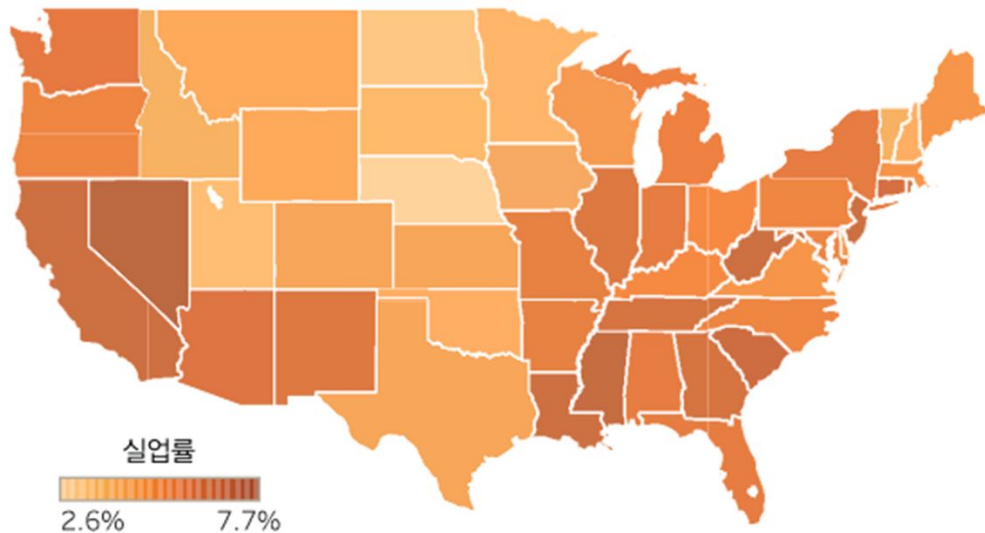
독자에게 경고하기 위한 색 사용



## 데이터 시각화 방법 - 색상의 사용

- **순차적 색상**은 밝은 색부터 어두운 색까지 단색을 사용함
- 아래 그림처럼 지역별 실업율을 규약화하는 사례라면 주황색이 어두울수록 실업률이 높고, 밝을 수록 실업률이 낮아지는 것을 표현할 수 있음

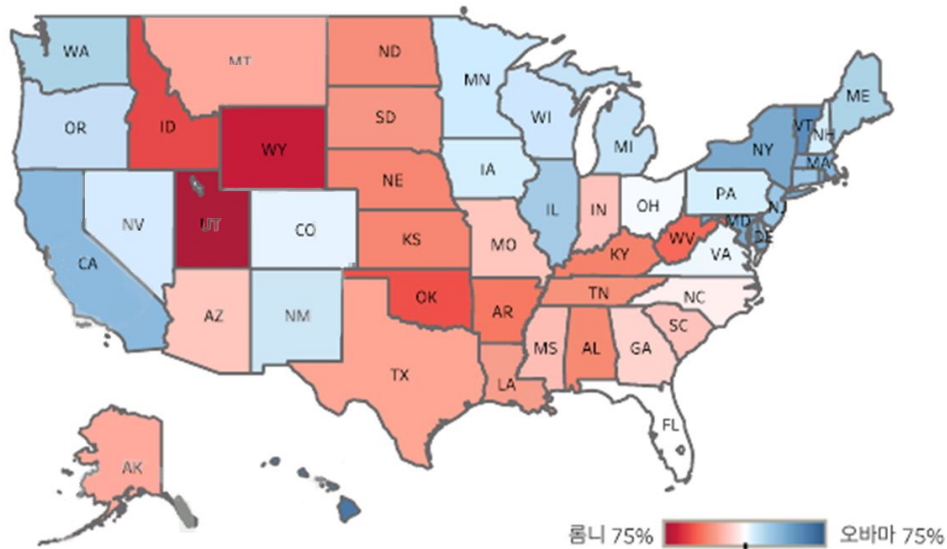
주별 실업률



<그림> 순차적 색 체계를 사용한 주별 실업률

## 데이터 시각화 방법 - 색상의 사용

- **발산형 색상**은 중간점에서 발생하는 범위를 표시할 때 사용
- 발산형 색상을 순차적으로 색상 체계와 같은 방식으로 사용할 수도 있지만, **상반되는 2가지 측정 범위를 나타내거나 2가지 카테고리 중에서 1가지 측정 범위를 규약화**하는 데도 쓰임
- 아래 그림은 주별로 유권자가 민주당이나 공화당중 어느 정당에 투표했는지를 보여주는 그래프

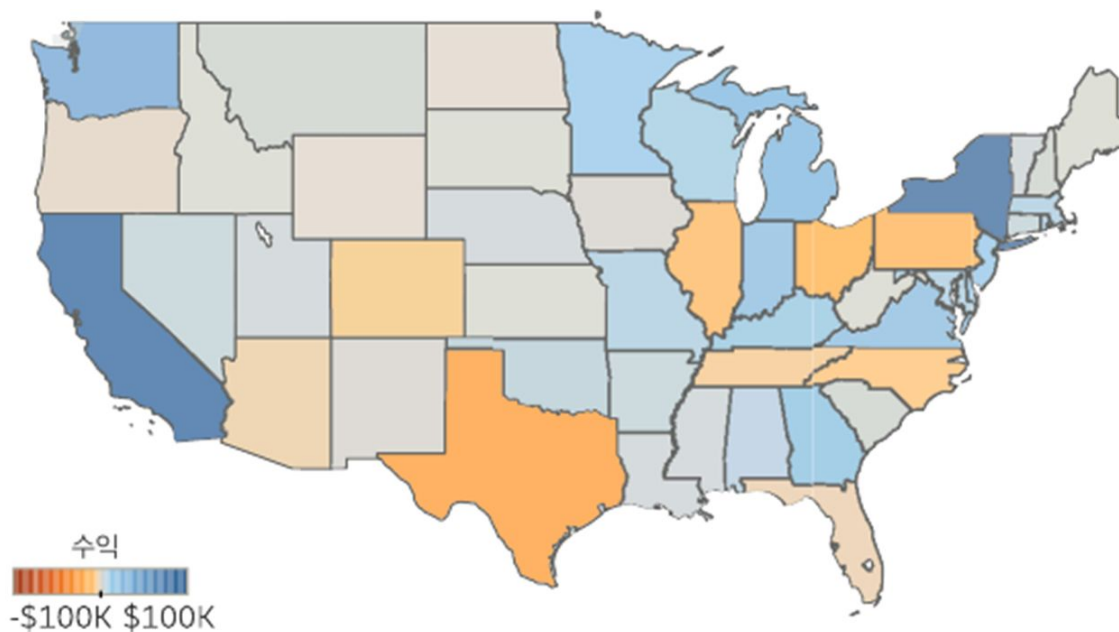


<그림> 각 주별 투표자 감정 등급



## 데이터 시각화 방법 - 색상의 사용

- **발산형 색상**으로 날씨도 보여줄 수 있는데 파란색은 낮은 기온, 빨간색은 높은 기온을 나타냄
- 중간점은 평균을 나타내거나, 혹은 긍정이나 부정의 수치를 나타낼 경우는 0이 되기도 함
- 아래 그림은 주별 수익의 사례를 보여주는데 이익은 파란색, 손실은 오렌지색으로 표시함

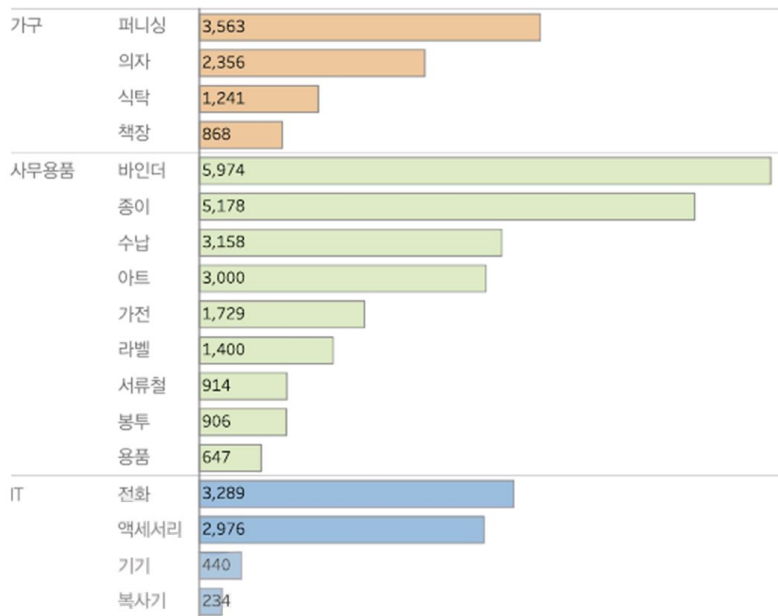


<그림> 발산형 색상 체계를 사용한 주별 수익

## 데이터 시각화 방법 - 색상의 사용

- 범주형 색상은 카테고리(범주)마다 다양한 색조로 구별함
- 예를들어 아래 그림과 같이 가구, 사무용품, IT 관련 카테고리를 정할 수 있음

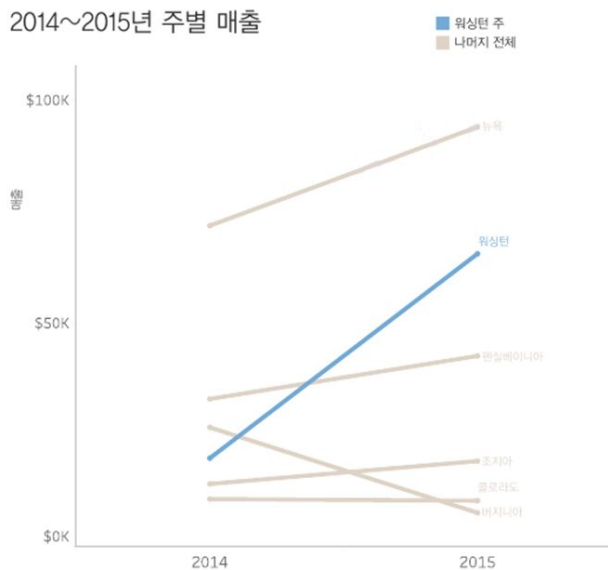
카테고리별 수량



<그림> 발산형 색상 체계를 사용한 주별 수익

## 데이터 시각화 방법 - 색상의 사용

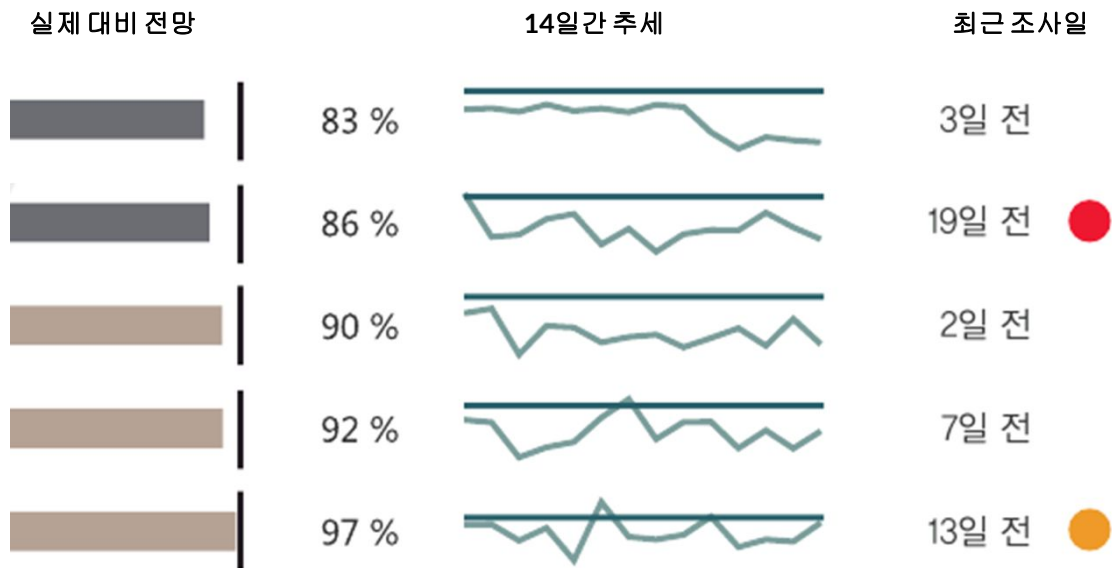
- 사용자의 눈에 띄어야 하지만, 경고하거나 놀라킬 필요는 없다면 강조색을 사용
- 강조는 특정 데이터 포인트, 표 안의 텍스트, 꺾은선 차트의 특정선, 막대 차트의 특정 막대를 강조할 때처럼 여러 방식으로 사용할 수 있음
- 아래 그림은 주 하나를 파란색으로 강조한 경사 차트를 나타냄



<그림> 단색으로 워싱턴 주를 강조하며 2014년~2015년 주별 매출을 나타낸 경사 차트

## 데이터 시각화 방법 - 색상의 사용

- 경고색은 사용자의 주목을 끌어야 할 때 사용
- 이 경우는 종종 밝고 경종을 울리는 색을 사용하는 편이 가장 좋은데, 그러면 아래 그림과 같이 차트를 보는 사람의 주의를 금방 끌 것임

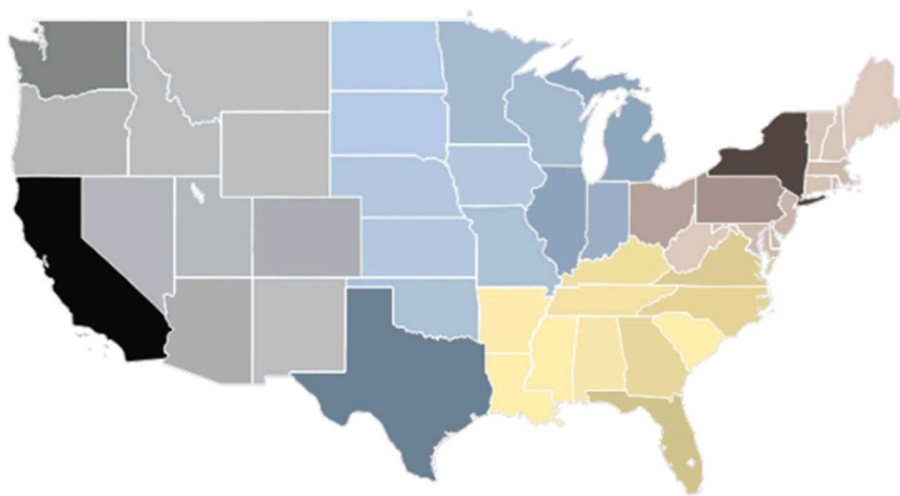


<그림> 빨간색, 오렌지 색 원은 뭔가 주목해야 한다는 것을 경고함

## 데이터 시각화 방법 - 색상의 사용

- 범주형 색상과 순차적 색상을 조합한 색 체계도 있을 수 있음
- 이 경우는 카테고리마다 나타내는 척도에 따라 더 어둡거나 밝은 색조로 구별함
- 아래 그림은 범주형 색상(회색, 노랑, 갈색, 파랑)을 사용하면서 동시에 순차적인 색을 함께 나타내, 지역의 척도를 규약화하는 4개의 지역을 지도 사례를 보여줌
- 음영이 더 어두운 주에서 매출이 높다고 짐작할 수 있음

지역별 매출



<그림> 4가지 지역을 범주형 색상으로 나타내고, 순차적 색상으로 매출을 나타냄

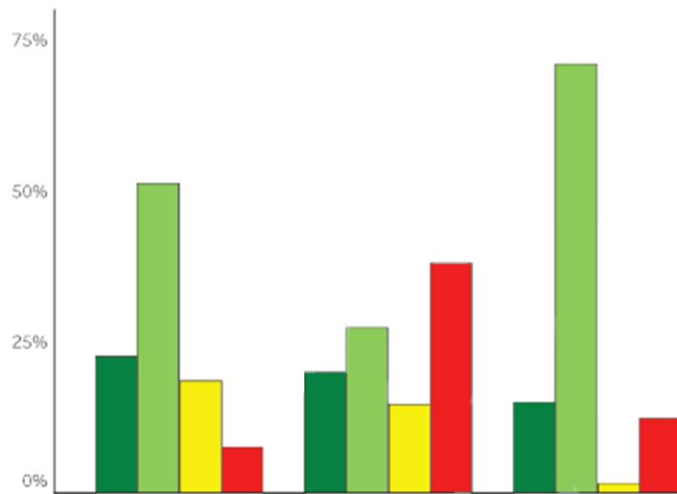
## 데이터 시각화 방법 - 색각 이상

- 연구(Birch 1993) 결과, 남성 중 8%, 여성은 0.4%의 색각 이상(CVD, Color Vision Deficiency)이 있다.
- 색각 이상은 모든 색을 보는 데 필요한 눈 안의 3개 추상체 유형 중 하나가 부족하면 생기는 장애
- **CVD로 고생하는 사람은 실제로 색을 볼 수는 있지만, 다른 사람과 같은 방식으로 색을 구별할 수 없음**
- 어떤 추상체가 부족한지에 따라 CVD가 있는 사람은 스펙트럼을 보는 방식 때문에 특정 색들을 구별하기 아주 어려울 수 있음
- CVD는 대체로 유전이며, 수치에서 볼 수 있듯 주로 남성이 많이 겪음
- **남성 중 8%는 작은 숫자처럼 보이지만, 큰 규모의 기업 또는 일반 대중을 위한 시각화라면 디자이너는 CVD를 잘 이해하고 염두에 두고 디자인해야 함**
- CVD의 3가지 종류
  - 제1색맹 : 긴 파장의 추상체가 부족(적색맹)
  - 제2색맹 : 중간 파장의 추상체가 부족(녹색맹)
  - 제3색맹 : 짧은 파장의 추상체가 부족(청색색맹)

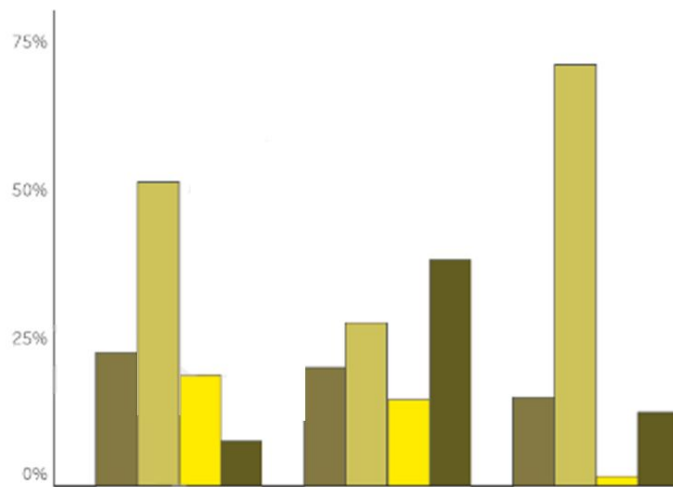
## 데이터 시각화 방법 - 색각 이상

- 색을 잘못 선택하면 **CVD**가 있는 사람에게 어떤 혼란을 일으킬 수 있는지 사례를 살펴본다
- 신호등 색을 사용하는 막대 차트와 제1색맹 시뮬레이션을 보면 우측 적색과 녹색 막대는 제1색맹이 있는 사람이 구별하기 아주 어려움을 알 수 있음

신호등 색



제1색맹(적색맹) 시뮬레이션

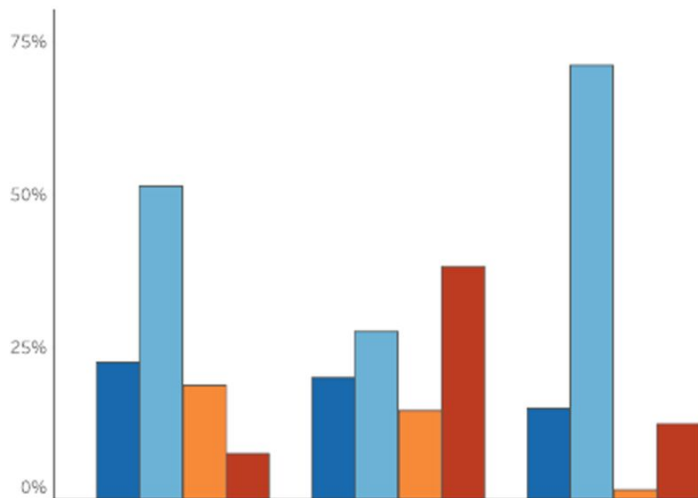


<그림> 녹색과 빨간색이 비슷한 색으로 보임

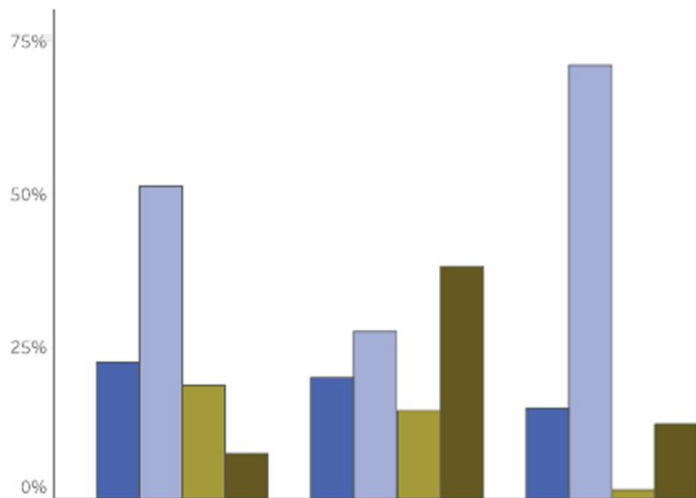
# 데이터 시각화 방법 - 색각 이상

- 밑의 그림은 색맹 친화적인 파랑, 오렌지 팔레트를 사용한 막대 차트와 제1색맹 시뮬레이션 결과

색맹 친화적인 파란색과 오렌지색



제1색맹(적색맹) 시뮬레이션



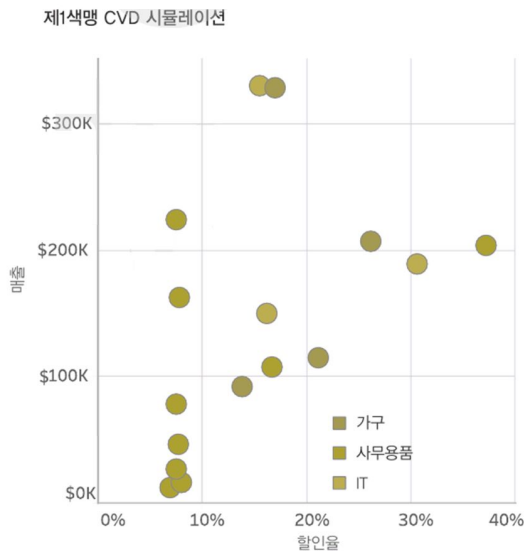
<그림> 파랑색과 오렌지색을 이용하면 색각 이상자도 색 구분이 가능



# 데이터 시각화 방법 - 색각 이상

- 적색과 녹색 외에 중요한 문제점

- 차트에서 색상을 선택시 좋은 것은 **녹색**, 나쁜 것은 **적색**을 흔히 사용함
- CVD**가 있는 사람을 위해 색을 구별할 때의 문제는 단순히 적색이니 녹색보다 더 복잡하다는 점을 이해해야함
- 적색, 녹색, 갈색, 오렌지색 모두 CVD가 심한 사람에게는 갈색**으로 보이기에 주의가 필요함
- 아래 그림은 갈색, 오렌지색, 녹색을 3개 카테고리에 대해 함께 사용한 산포도임
- 제1색맹 시뮬레이션에 적용할 때 산포도의 점들은 모두 유사한 색처럼 보임

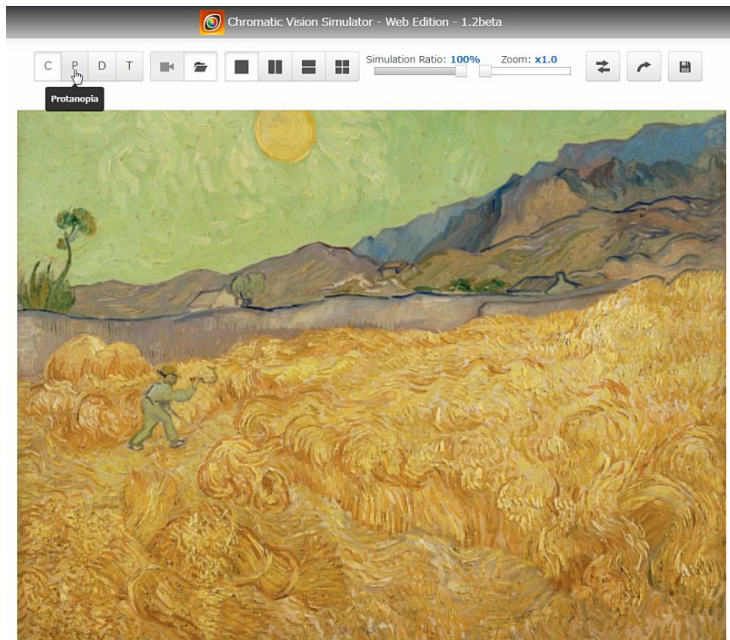


## 데이터 시각화 방법 - 색각 이상

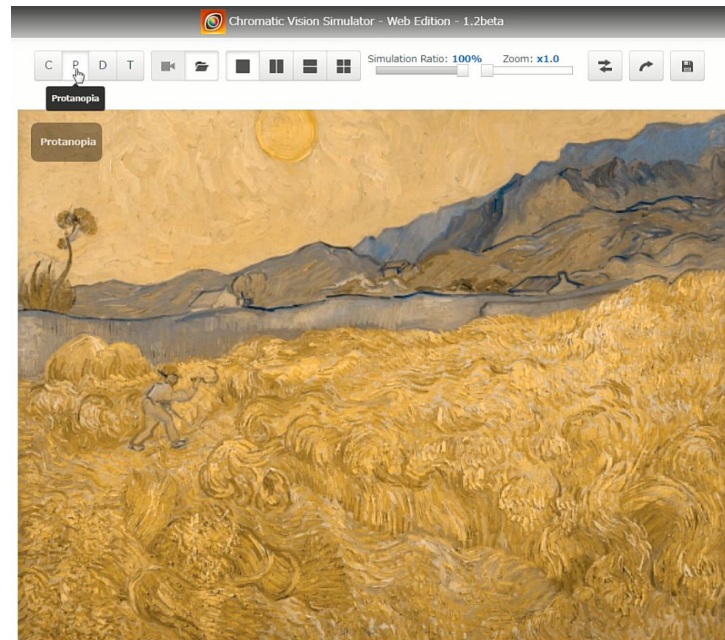
- 적색과 녹색 외에 중요한 문제점
  - 시각화를 디자인 할 때 이런 문제를 잘 이해해야 함
  - 데이터를 규약화 하려고 색상을 사용하는데, 사용자가 색상을 구별해서 시각화를 이해하게 만들어야 한다면, 색맹 친화적인 팔레트를 사용할지 검토가 필요
- CVD 유형 시뮬레이션 도구
  - 크로매틱 비전 시뮬레이터
  - 노커피 비전 시뮬레이터

# 데이터 시각화 방법 - 색각 이상

- 크로매틱 비전 시뮬레이터(<https://asada.website/webCVS/>)



<그림> 일반 색



<그림> 제1색맹 색

# 데이터 시각화 방법 - 색각 이상

- 노커피 비전 시뮬레이터 (<https://chrome.google.com/webstore/detail/nocoffee/jieeggmbnhckmgdhdmgdckeigabjfbddl>)

chrome 웹 스토어

품 > 확장 프로그램 > NoCoffee



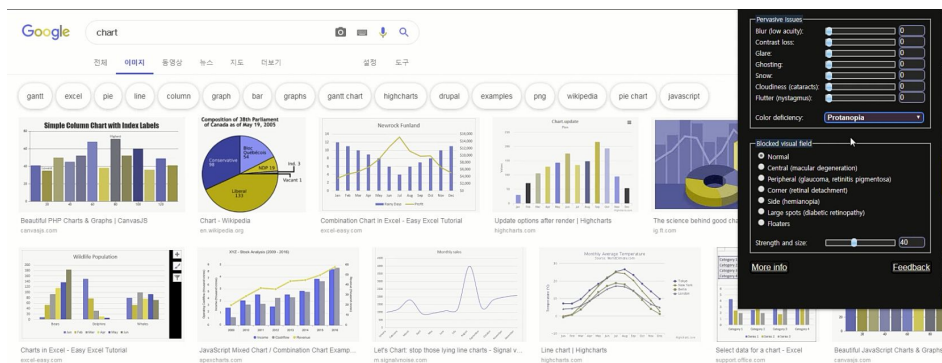
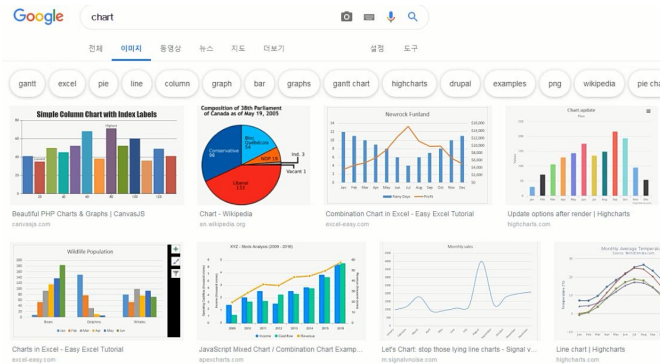
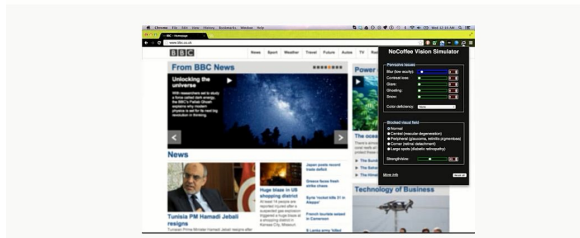
NoCoffee

제공자: Aaron Leventhal

★★★★☆ 40 | 접근성 | 사용자 18,463명

Chrome에서 사용

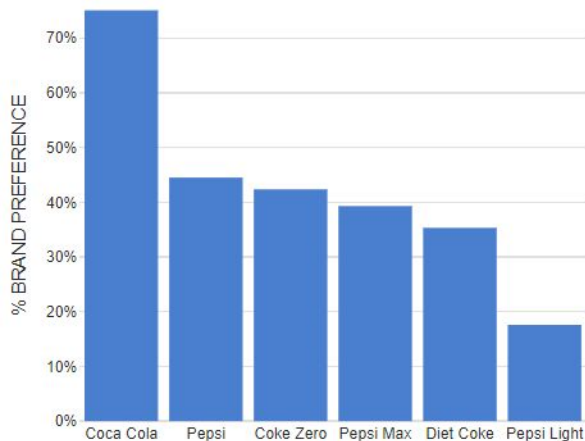
개요 리뷰 지원 관련 프로그램



# 데이터 시각화 방법 - 일반적인 차트 유형

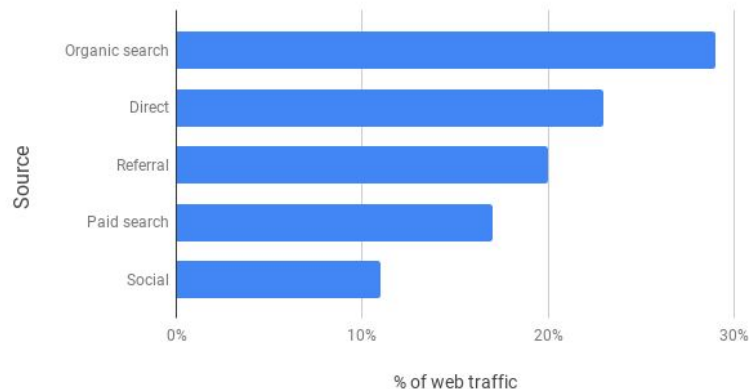
## 막대차트

- 막대 차트는 길이로 척도를 나타냄
- 사람은 공통적인 기준선으로부터 길이가 아주 조금만 차이더라도 빨리 인지할 수 있음
- 막대 차트는 카테고리를 비교하는 가장 효과적인 방법으로 데이터 시각화에 널리 사용됨
- 막대는 수평이나 수직 방향으로 사용할 수 있음
- 막대 차트는 최대/최소 항목을 찾을 때 가장 흔히 사용하므로, 정렬해서 비교하면 매우 유용함



수직 바 차트

Web traffic sources

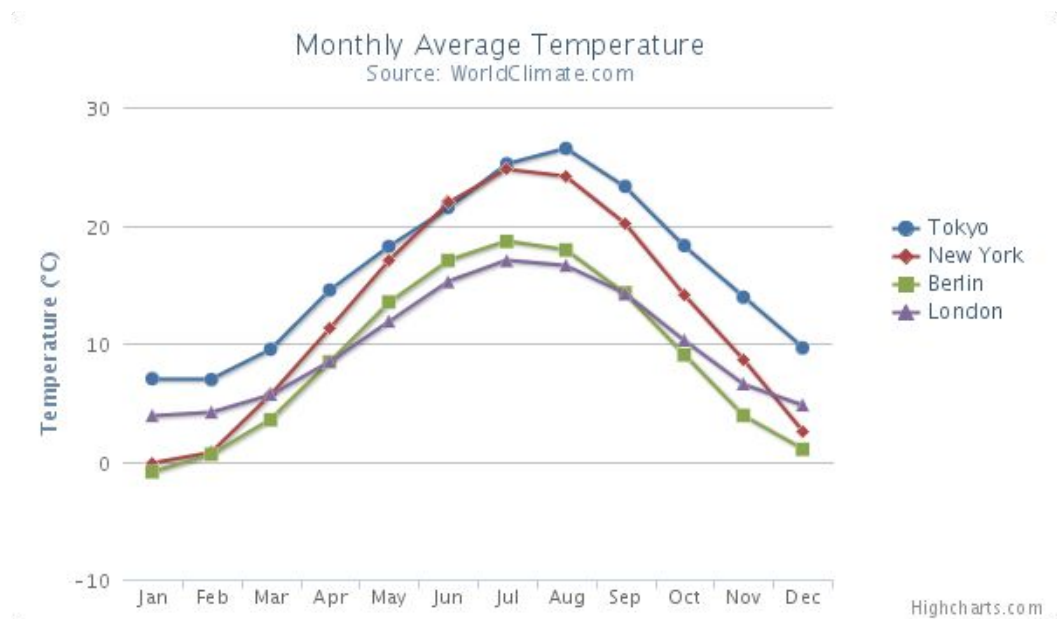


수평 바 차트

# 데이터 시각화 방법 - 일반적인 차트 유형

- 시계열 꺾은선 차트

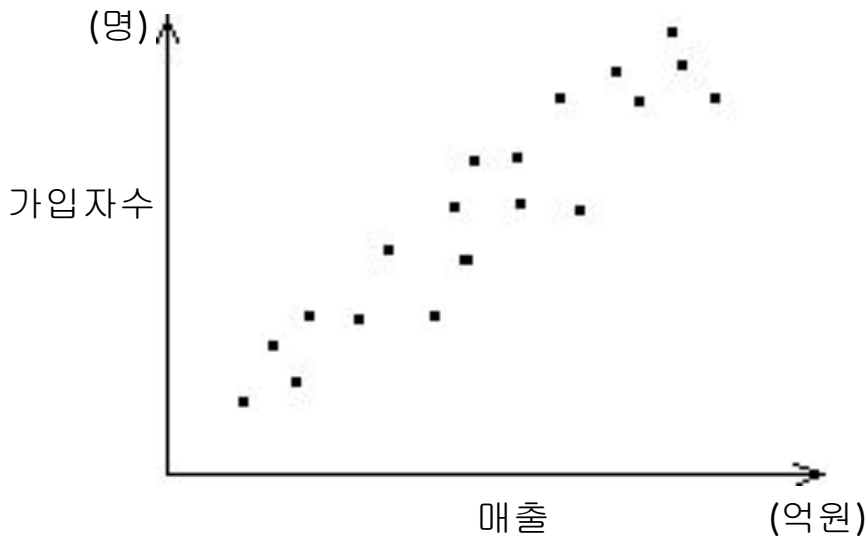
- 꺾은선 차트는 일반적으로 **장기적인 변화를 볼 때 이용**
- 수평인 **X**축은 시간을 나타내며, 수직인 **Y**축에는 척도를 표시
- 선의 높이와 경사로 추세를 볼 수 있음



## 데이터 시각화 방법 - 일반적인 차트 유형

- 산포도

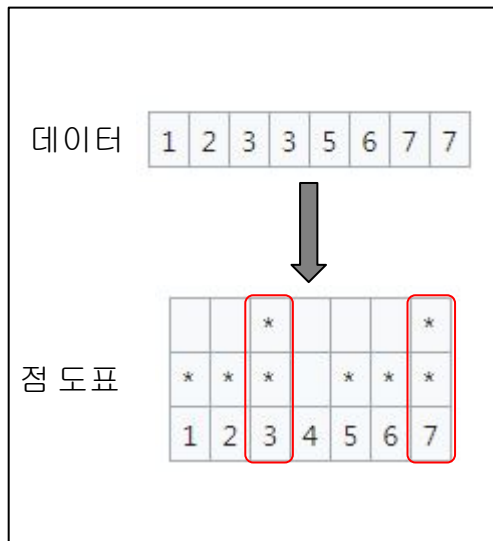
- 서로 다른 두 척도를 비교할 수 있음
- 각 척도는 수평 축과 수직 축상의 위치를 써서 규약화 함
- 산포도는 2가지 변수 간의 관계를 관찰할 때 유용함



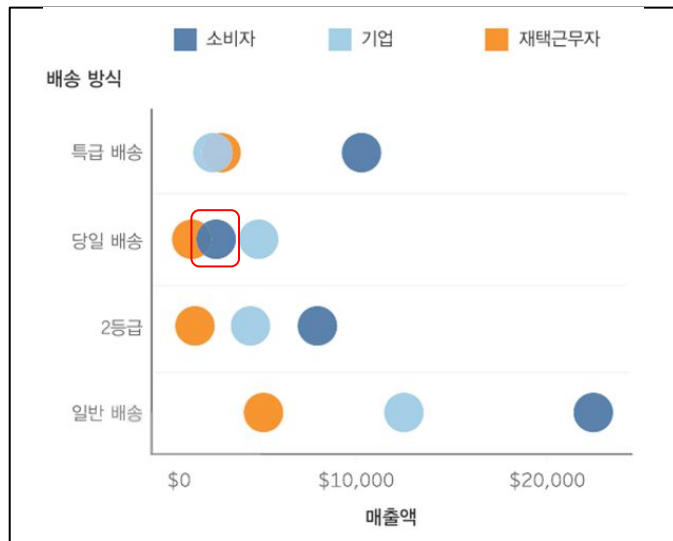
# 데이터 시각화 방법 - 일반적인 차트 유형

## · 점도표(Dot Plot)

- 점도표는 통계학에서 **데이터들의 분포를 점으로 나타내주는 도표** 또는 그러한 도표로 나타내는 방법
- 2개 차원에 걸쳐 값을 비교 가능함

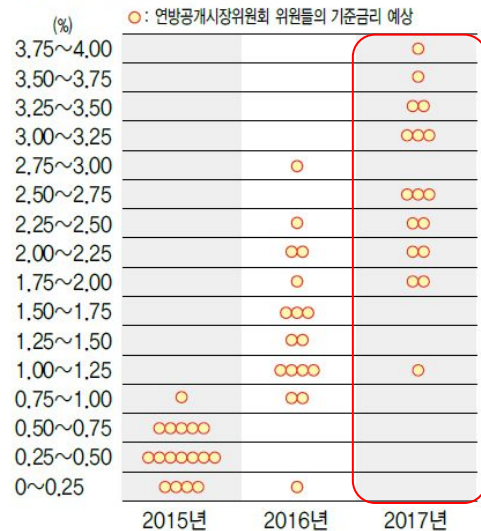


<그림1> 데이터를 점도표로 변환



<그림2> 소비자 대상 매출은 당일 배송 방식에서 가장 적음

## 9월 미 연방준비제도 점도표



\*예상 금리 구간은 0.25%포인트로 간략화함. 자료: 미국 연방준비제도

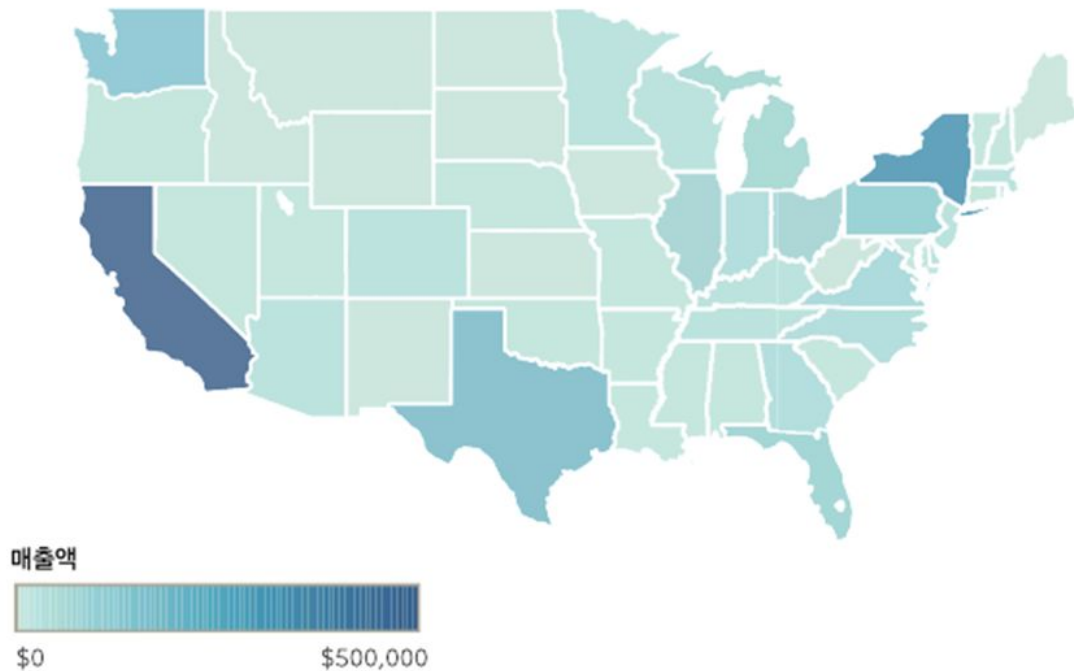
<그림3> 2017년 금리2.5~3.25 예상



## 데이터 시각화 방법 - 일반적인 차트 유형

- 단계 구분도

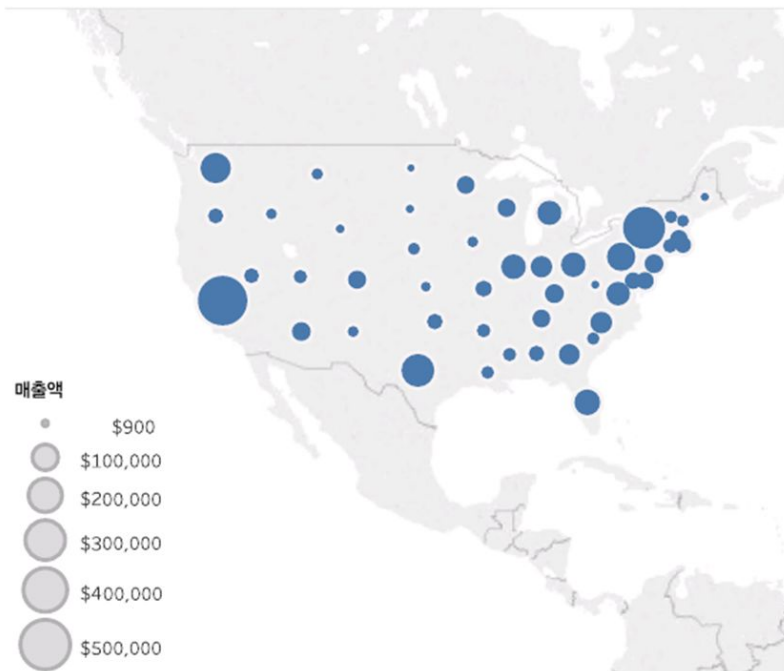
- 단계 구분도는 미리 정의된 영역 내에서 음영이나 색의 차이로 그 영역의 값이나 카테고리를 나타냄



## 데이터 시각화 방법 - 일반적인 차트 유형

- 기호 지도

- 기호지도는 특정 장소의 값을 보여줌
- 규모가 큰 지역의 중앙 지점일 수도 있고, 정확한 위/경도 척도로 결정한 특정 위치 일수도 있음



## 데이터 시각화 방법 - 일반적인 차트 유형

- 표
  - 정확한 데이터를 알아야 하는 경우에 적절한 방법
  - 대부분의 대시보드에서 표는 요약 차트와 함께 세부 데이터를 보여줌

		Central	East	South	West
Furniture	Bookcases	(\$22,399)	(\$14,607)	\$3,599	(\$175)
	Chairs & Chairmats	\$51,534	\$54,181	\$14,988	\$28,947
	Office Furnishings	\$30,852	\$20,645	\$21,865	\$27,066
	Tables	(\$27,545)	(\$65,504)	\$887	(\$6,902)
Office Supplies	Appliances	\$49,445	\$17,556	\$17,732	\$12,425
	Binders and Binder Access..	\$90,004	\$94,617	\$61,612	\$61,180
	Envelopes	\$12,794	\$11,162	\$9,715	\$14,512
	Labels	\$4,733	\$2,253	\$4,067	\$2,624
	Paper	\$19,509	\$10,723	\$8,307	\$6,725
	Pens & Art Supplies	\$1,927	\$2,578	\$1,620	\$1,440
	Rubber Bands	\$173	(\$167)	\$57	(\$165)
	Scissors, Rulers and Trimm..	(\$4,406)	(\$1,229)	(\$1,865)	(\$299)
	Storage & Organization	(\$11,358)	\$630	\$12,072	\$5,320
Technology	Computer Peripherals	\$27,675	\$23,452	\$25,461	\$17,700
	Copiers and Fax	\$44,542	\$41,350	\$22,403	\$59,066
	Office Machines	\$108,471	\$55,305	\$155,488	(\$11,551)
	Telephones and Communic..	\$105,941	\$64,906	\$64,500	\$81,604

## 데이터 시각화 방법 - 일반적인 차트 유형

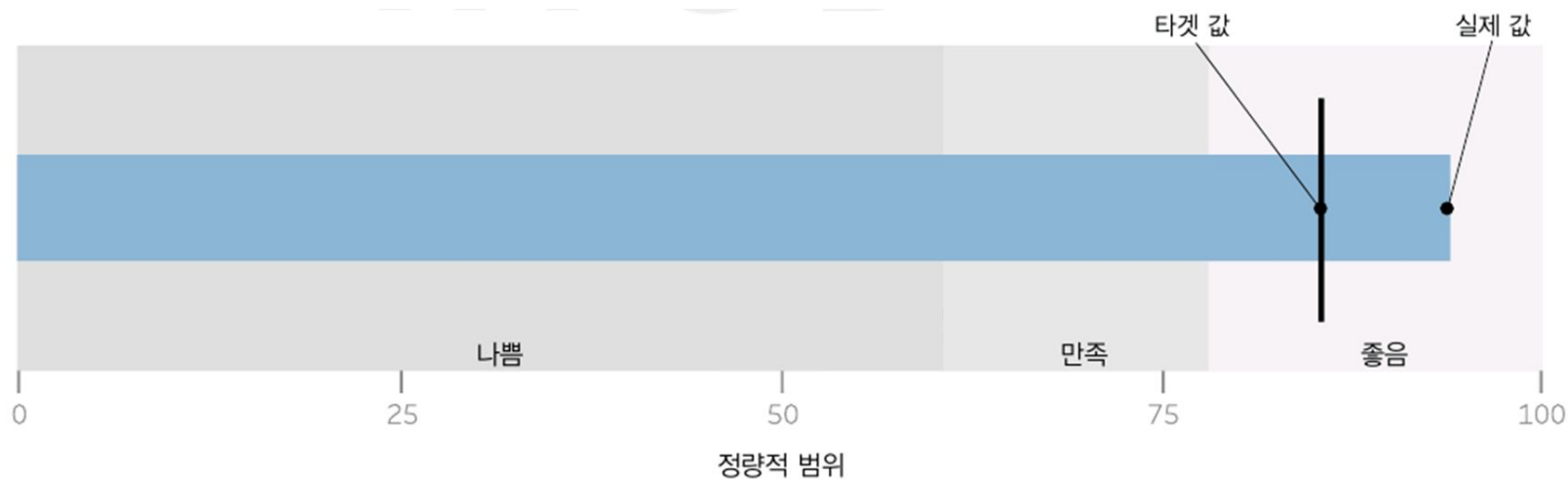
- 하이라이트 표
  - 표를 색상으로 규약화하면 모든 값을 정확히 찾을 수 있으면서도, 시각적으로 보기 좋게 바꿀 수 있음

		Central	East	South	West
Furniture	Bookcases	(\$22,399)	(\$14,607)	\$3,599	(\$175)
	Chairs & Chairmats	\$51,534	\$54,181	\$14,988	\$28,947
	Office Furnishings	\$30,852	\$20,645	\$21,865	\$27,066
	Tables	(\$27,545)	(\$65,504)	\$887	(\$6,902)
Office Supplies	Appliances	\$49,445	\$17,556	\$17,732	\$12,425
	Binders and Binder Access..	\$90,004	\$94,617	\$61,612	\$61,180
	Envelopes	\$12,794	\$11,162	\$9,715	\$14,512
	Labels	\$4,733	\$2,253	\$4,067	\$2,624
	Paper	\$19,509	\$10,723	\$8,307	\$6,725
	Pens & Art Supplies	\$1,927	\$2,578	\$1,620	\$1,440
	Rubber Bands	\$173	(\$167)	\$57	(\$165)
	Scissors, Rulers and Trimm..	(\$4,406)	(\$1,229)	(\$1,865)	(\$299)
	Storage & Organization	(\$11,358)	\$630	\$12,072	\$5,320
Technology	Computer Peripherals	\$27,675	\$23,452	\$25,461	\$17,700
	Copiers and Fax	\$44,542	\$41,350	\$22,403	\$59,066
	Office Machines	\$108,471	\$55,305	\$155,488	(\$11,551)
	Telephones and Communic..	\$105,941	\$64,906	\$64,500	\$81,604

## 데이터 시각화 방법 - 일반적인 차트 유형

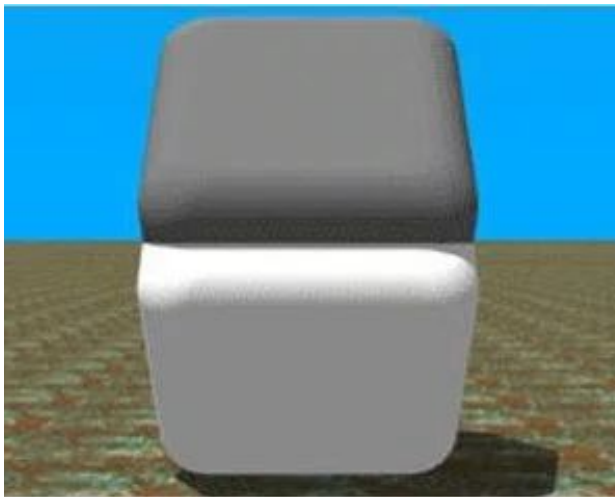
- 불릿 그래프

- 불릿 그래프는 실제 값 대비 타겟의 비교를 보여주는 가장 좋은 방법에 속함
- 파란 막대는 실제 값이며, 검은 선은 타겟 값을 나타내며, 회색 음영 영역은 성과의 범위임

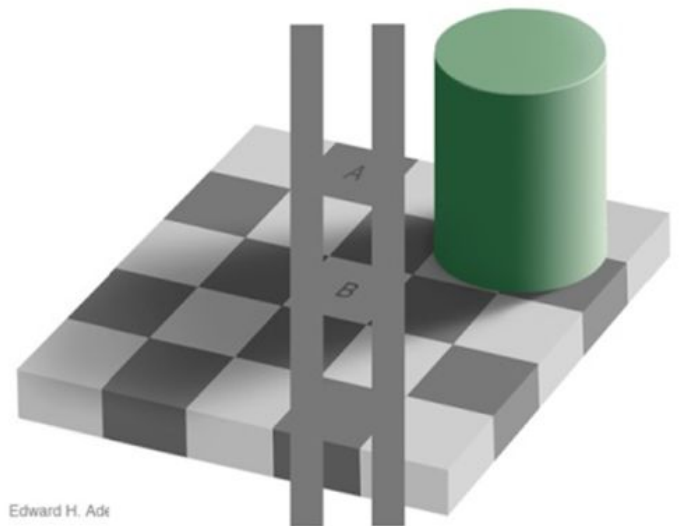


## 데이터 시각화 방법 - 시각 처리의 오류

- 전주의적 속성으로 시각 체계의 힘을 우리에게 유리하게 이용할 수 있음
- 한편으로 시각 체계는 쉽게 혼란에 빠지기도 함
- 이미지의 모호함으로 때문에 재미있는 착시현상이 일어나지만, 데이터 시각화 관점에서는 문제가 될 수 있음



<그림1> 상단의 회색이 하단의 회색보다 어두운가?

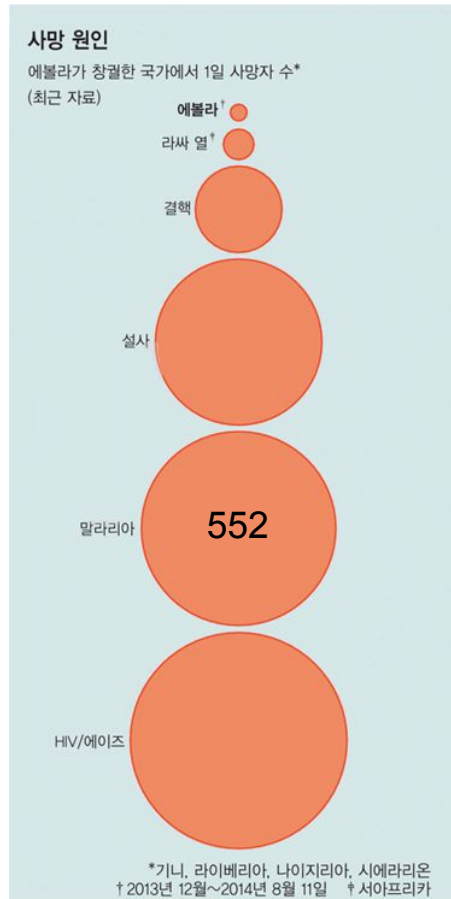


Edward H. Adelson

<그림2> A와 B는 기울어져 있는가?

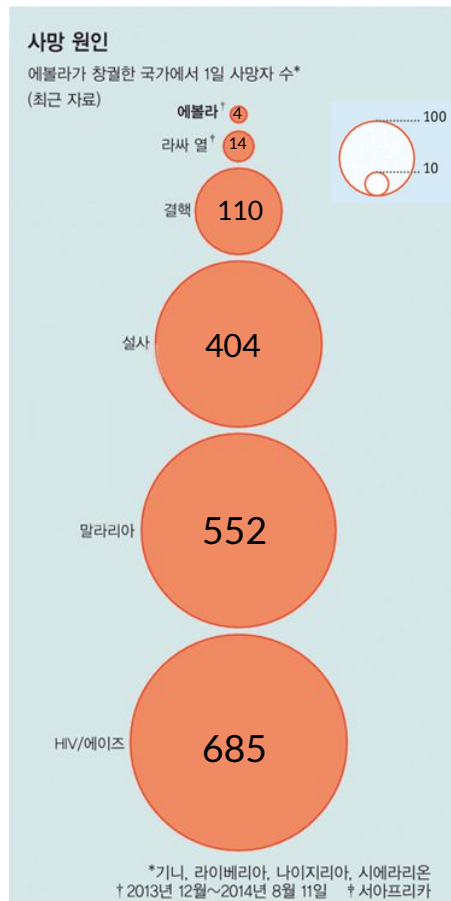
## 데이터 시각화 방법 - 시각 처리의 오류

- 데이터를 시각화할 때, 대부분은 가능한 한 **짧은 시간** 안에 가장 **정확히 해석**할 수 있는 **측정의 가치를 전달**하려함
- 이러한 목적에 적합한 전주의적 속성은 따로 있음



## 데이터 시각화 방법 - 시각 처리의 오류

- 앞의 차트와 같이 표의 크기로는 실제 사망자 수를 판단 하기 어려움
- 대부분의 사람들은 원의 크기를 실제보다 작다고 착각함
- 핵심은 크기가 전주의적인 속성인 반면에 사람들은 정확하게 그 크기의 차이를 전혀 구별하지 못함

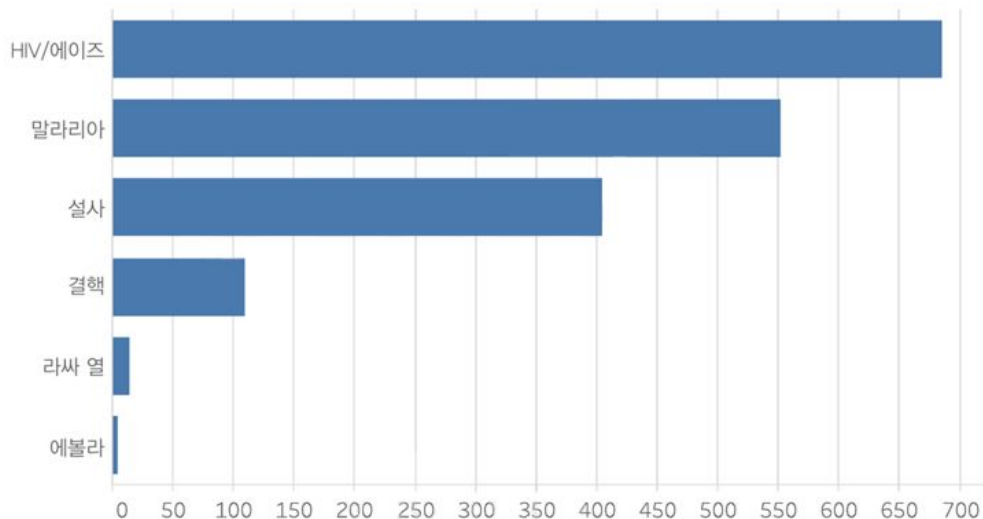




## 데이터 시각화 방법 - 시각 처리의 오류

- 막대 차트에서는 정량적 변수인 1일 사망자 수를 길이로 규약화 한 것임
- 질병 별 차이를 명확하게 확인할 수 있음
- 이러한 이유로 인해 막대차트에서 보여주는 길이는 시각화를 처리하기에 효율적인 전주의적 속성임

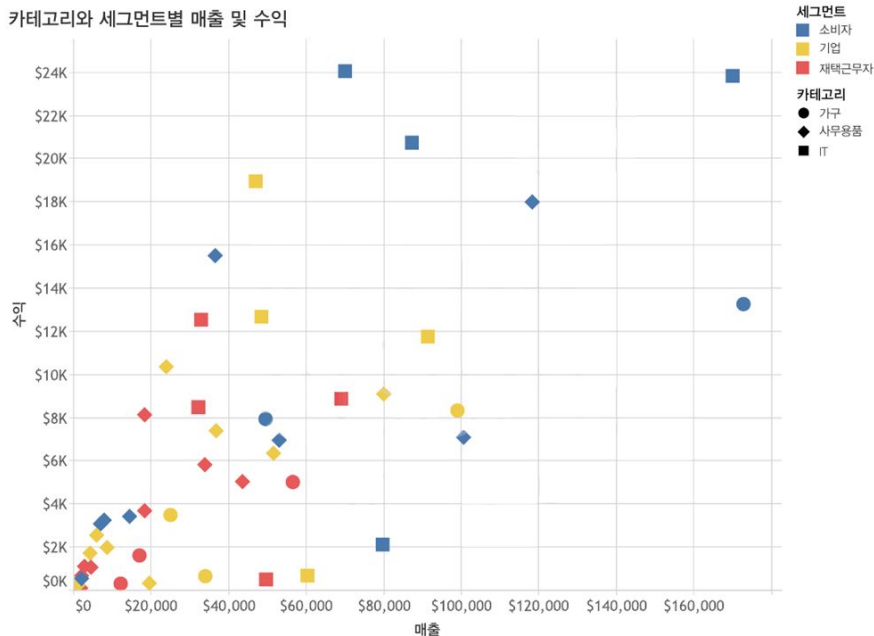
에볼라가 발생한 국가의 1일 사망자수



\*기니, 라이베리아, 나이지리아, 시에라리온  
(출처: 세계보건기구, 미국 질병관리 예방센터, <이코노미스트>)

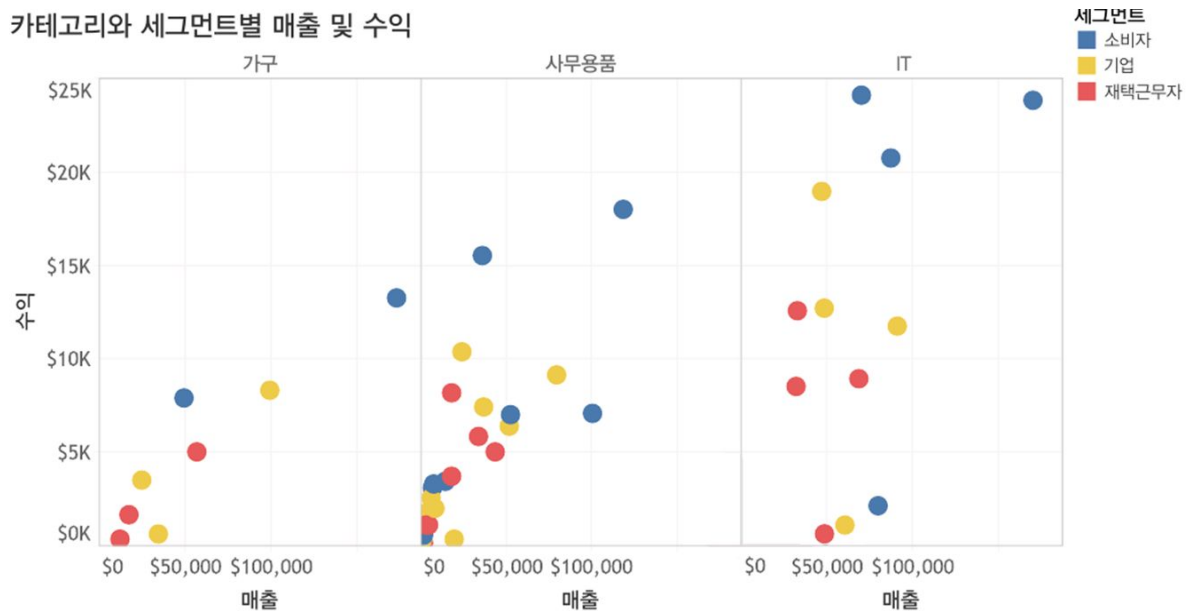
# 데이터 시각화 방법 - 시각 처리의 오류

- 여러가지 전주의적 속성을 차트 하나에 함께 사용하면 문제로 이어질 수 있음
- 아래 그림은 가상 판매업체의 매출액과 수익에 대한 산포도임
- 색상은 다양한 세그먼트를 보여주고, 형태는 제품의 카테고리를 나타냄
- 어떤 카테고리의 수익이 평균적으로 가장 많았는가?



## 데이터 시각화 방법 - 시각 처리의 오류

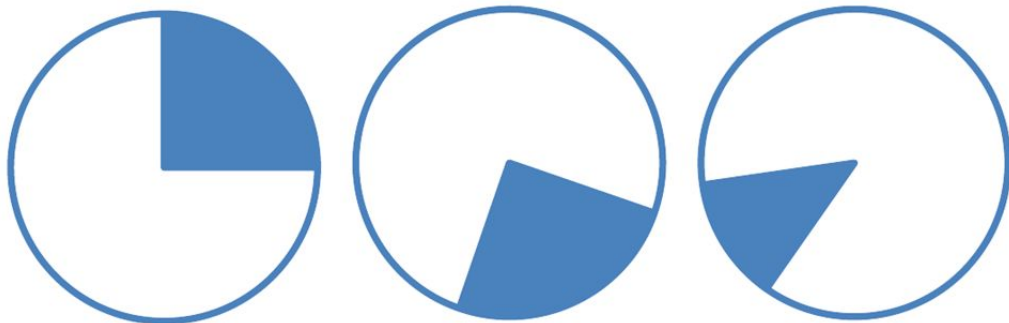
- 카테고리별로 단일 산포도를 3개 패널로 그리면 이전 차트보다 결과가 눈에 띄게 보임
- IT 제품이 매출대비 수익이 큰 것을 명확히 확인할 수 있음



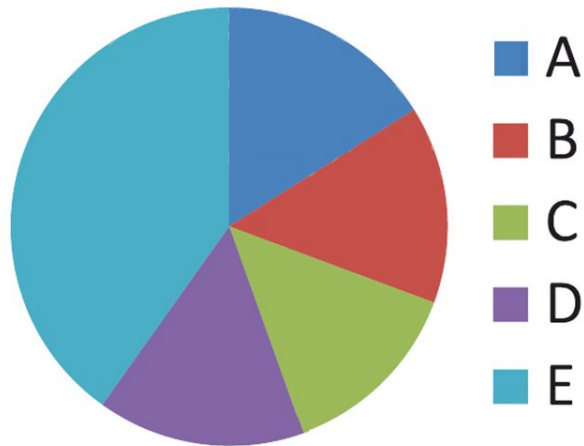
<그림> 카테고리별로 차트가 하나인 매출과 수익

## 데이터 시각화 방법 - 시각 처리의 오류

- 파이 차트의 사용을 권장하지 않는 이유
  - 아래 그림과 같이 파이차트는 비슷한 크기가 있을 때 구분이 어려움



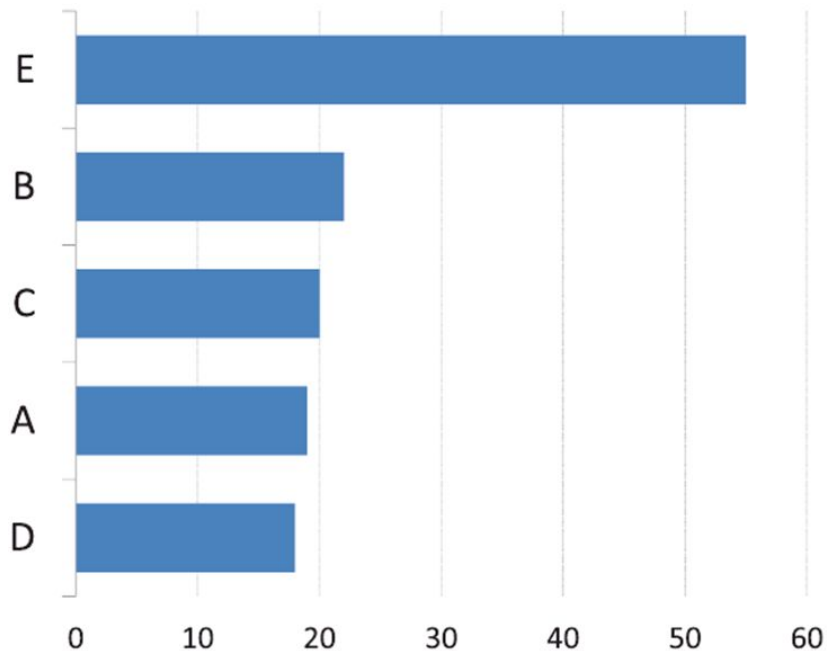
<그림1> 파란 영역은 몇 비중이 몇 퍼센트 인가?



<그림2> 3번째로 큰 영역은 어떤 카테고리인가?

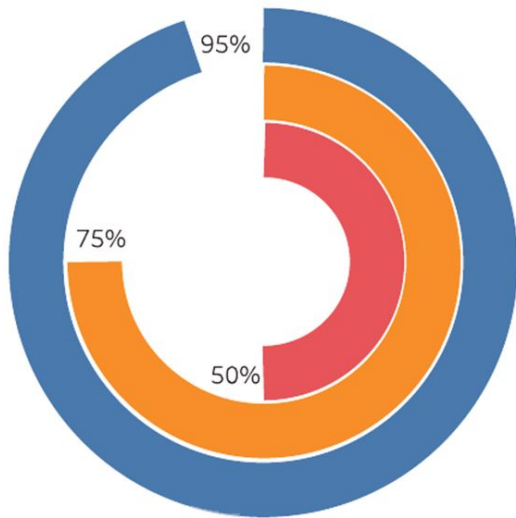
## 데이터 시각화 방법 - 시각 처리의 오류

- 앞 장의 <그림2>에 나오는 카테고리의 크기를 바 차트로 타나내면 아래와 같이 명확하게 구분이 가능함
- 즉 길이는 너무나 효과적인 시각적 속성이라 아주 작은 차이도 쉽게 알아차릴 수 있음



## 데이터 시각화 방법 - 시각 처리의 오류

- 파이 차트의 단점을 보완한 원형 차트를 이용하기 위해 아래와 같이 도넛 차트를 활용할 수 있음
- 파이 차트보다는 각 카테고리별 차이가 명확해 보임
- 하지만 최선의 선택인지 확인이 필요함



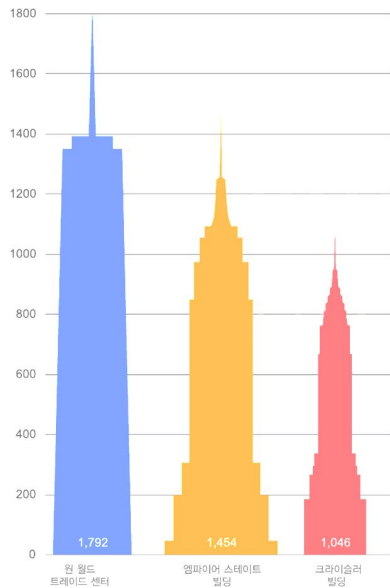
<그림> 중심이 있는 도넛 차트(방사형 막대 차트)

예시

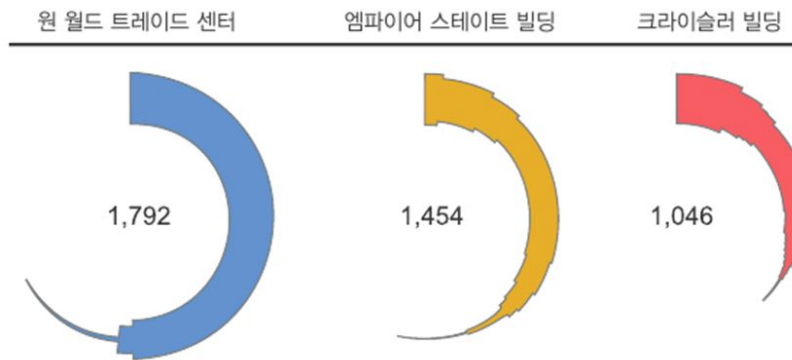
<https://public.tableau.com/profile/michael.daddona#!/vizhome/StackedDoughnutCharts/StackedDoughnutCharts>

# 데이터 시각화 방법 - 시각 처리의 오류

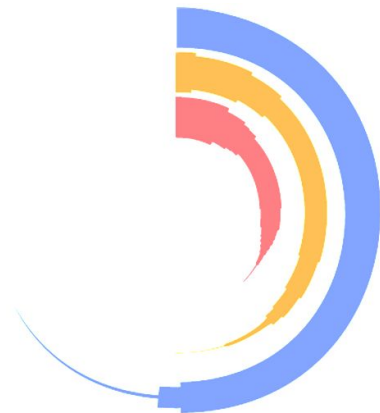
- 3개 빌딩의 높이를 비교 할 때 어떤 차트가 가장 직관적으로 확인할 수 있는가?



<그림1> 빌딩 모양의 바 차트



<그림2> 반원으로 왜곡된 3개 고층 빌딩



<그림3> 시작점을 일치 시킨 도넛 차트

예시

<https://www.tableau.com/about/blog/2017/2/viz-whiz-donut-charts-65653>

## 데이터 시각화 방법 - 데이터에 따른 시각화 선택

- 목적과 데이터에 따라서 시각화를 통해 강조하려는 점이 있을 것임
- 다양한 시각화 방법 중 적절한 시각화 하는 방법을 선택해야 함

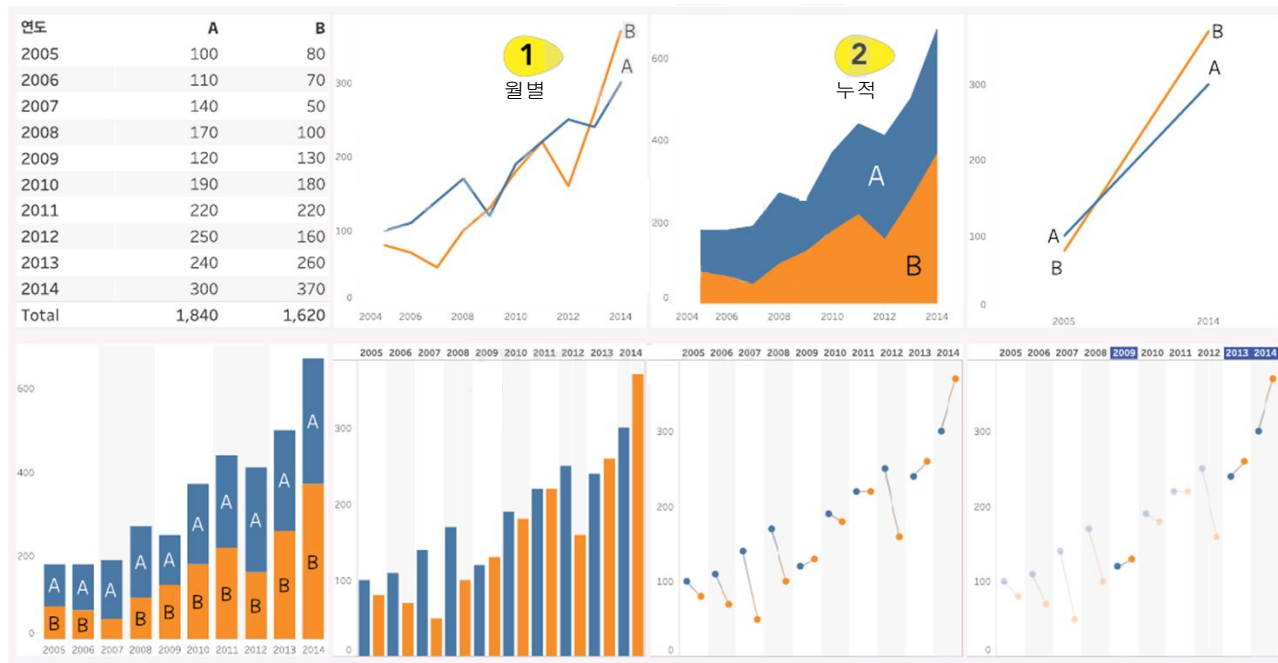
문제 : 아래와 같이 **A** 제품, **B** 제품의 **10**년간 매출은 어떻게 시각화 하는 것이 좋은가?(당해, 누적)

	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	합계
A	100	110	140	170	120	190	220	250	240	300	1,840
B	80	70	50	100	130	180	220	160	260	370	1,620



# 데이터 시각화 방법 - 데이터에 따른 시각화 선택

	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	합계
A	100	110	140	170	120	190	220	250	240	300	1,840
B	80	70	50	100	130	180	220	160	260	370	1,620



# 데이터 시각화 방법 - 차트 유형(1/4)



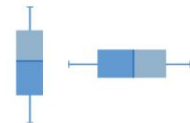
## 영역 차트(Area Chart)

위치와 높이로 데이터를 규약화하여 장기적 동향이나 규모를 표시



## 막대 차트(Bar Chart)

막대의 높이/길이를 데이터로 규약화하며 카테고리 비교를 표시



## 박스 플롯(Box Plot)

위치와 높이/길이를 데이터로 규약화해 데이터 분포를 표시



## 볼릿 그래프(Bullet Graph)

길이/높이, 위치와 색으로 데이터를 규약화해 타겟이나 성과 대역에 비교한 실제를 표시



## 단계구분도(Choropleth Map) / 음영 처리된 지도(Shaded Map)

색과 위치로 데이터를 규약화해 데이터를 지리적으로 표시



## 확산형 막대 차트(Diverging Bar Chart)

중간 지점으로부터 확산되는 막대의 높이/길이를 데이터로 규약화해 카테고리 비교를 표시



## 점 도표(Dot Plot)

위치로 데이터를 규약화해 비교를 표시

## 데이터 시각화 방법 - 차트 유형(2/4)



### 지터 점 도표(Dot Plot with Jitter) / 지터플롯(JitterPlot)

위치로 데이터를 규약화해 비교를 표시하지만, 무작위로 점을 찍어서 점이 중첩되지 않음



### 간트 차트(Gantt Chart)

길이와 위치로 데이터를 규약화해 시간의 세그먼트 내에 완료된 작업량을 표시



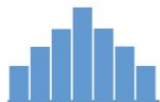
### 히트맵(Heat Map)

색으로 데이터 표를 규약화해 숫자 없이 표의 차이를 강조

\$29,071	\$17,307	\$36,079
\$1,653	\$2,853	\$5,079
\$1,853	\$17,307	\$10,079
\$20,379	\$14,551	\$25,664
\$100,615	\$10,004	\$10,004
\$71,653	\$35,769	\$70,533
\$10,760	\$8,239	\$18,127
\$98,340	\$13,125	\$10,769

### 하이라이트 표(Highlight Table)

색으로 데이터 표를 규약화해 표의 숫자 차이를 강조



### 히스토그램(Histogram)

높이로 데이터를 규약화하고 분포를 표시



### 꺾은선 차트(Line Chart)

위치로 데이터를 규약화하며, 장기적 추세를 표시할 때도 쓰임



### 롤리팝 차트(Lollipop Chart)

막대의 높이/길이로 데이터를 규약화하고 카테고리 비교를 표시

## 데이터 시각화 방법 - 차트 유형(3/4)



### 산포도(Scatter Plot)

위치로 데이터를 규약화해 2가지 변수의 관계를 표시. 크기로도 2차 비교 가능



### 경사 차트(Slopegraph)

위치로 데이터를 규약화해 일반적으로 2개 기간 간의 정량 비교나 순위를 표시



### 스파크라인(Sparkline) / 스파크 막대(Sparkbar)

작은 단어 크기인 그래픽의 위치(선)나 높이/길이(막대)로 데이터를 규약화



### 누적 막대 차트(Stacked Bar Chart) \*주의: 많은 세그먼트 사용 지양

막대의 높이/길이와 세그먼트별 색으로 데이터를 규약화하고 카테고리 부분에서 전체로의 비교를 표시



### 심볼 지도(Symbol Map) / 점 지도(Dot Map)

위치로 데이터를 규약화해 지리적으로 데이터를 표시하고, 크기로도 정량 데이터를 표시할 수 있음



### 트리맵(Treemap)

크기와 색으로 데이터를 규약화하여 계층적 데이터나 대량의 카테고리를 비교할 때 유용



### 워터폴 차트(Waterfall Chart)

높이와 종종 색으로 데이터를 규약화해, 기간이나 카테고리 간 증감을 표시

## 데이터 시각화 방법 - 차트 유형(4/4)



### 버블 차트(Bubble Chart)

원 크기로 데이터를 규약화해 비교를 표시하는데, 정확한 정량 비교는 어려움



### 중심있는 원(Concentric Circles) / 방사형 막대 차트(Radial Bar Chart)

위치로 데이터를 규약화해 일반적으로 2개 기간 간의 정량 비교나 순위를 표시



### 도넛 차트(Donut Chart)

아치와 영역으로 데이터를 규약화해 부분에서 전체로의 비교를 표시하지만, 여러 이유로 문제가 있음



### 파이 차트(Pie Chart)

각도, 영역, 아치로 데이터를 규약화해 부분에서 전체로의 비교를 표시하지만, 여러 이유로 문제가 있음

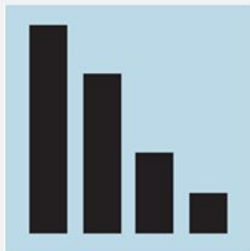


### 워드 클라우드(Word Cloud)

단어 크기로 데이터를 규약화해 비교를 표시하는데, 정확한 정량 비교는 어려움

## 데이터 시각화 방법 - 목적에 맞는 차트 유형

### 비 교



막대 차트

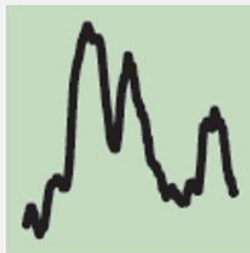


그룹 막대 차트



버블 차트

### 추 이



선 차트



영역 차트



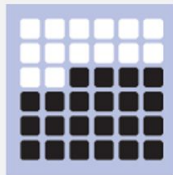
타임라인 차트

# 데이터 시각화 방법 - 목적에 맞는 차트 유형

## 비 중



파이 차트



와플 차트

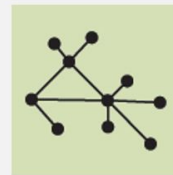


트리맵

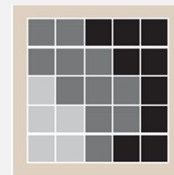
## 관 계



산점도



네트워크



XY 히트맵

## 지 도



Filed map



Symbol map



Flow map

### - tableau

- 시각화 가이드

- : [https://onlinehelp.tableau.com/current/pro/desktop/en-us/dataview\\_examples.htm](https://onlinehelp.tableau.com/current/pro/desktop/en-us/dataview_examples.htm)

- 무료 교육 동영상

- : <https://www.tableau.com/ko-kr/learn/training>

- 실습용 데이터

- : [https://public.tableau.com/s/resources?qt-overview\\_resources=1#qt-overview\\_resources](https://public.tableau.com/s/resources?qt-overview_resources=1#qt-overview_resources)