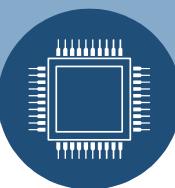


[양산인재개발평생교육원]

빅데이터 분석 결과 시각화



경남대학교 창의융합대학 교수 유현주





강의 진행 안내

- 수업 자료 공유 URL

- <https://github.com/comjoo94/>

- 수업에 필요한 강의 자료 다운로드 저장

- 소통을 위한 메일 주소: comjoo@uok.ac.kr

- comjoo@gmail.com





Contents

빅데이터 분석 결과 시각화

- 데이터 시각화
- Excel의 데이터 시각화
- 통계 기반 데이터 시각화
- 빅데이터 시각화 Power BI



데이터 시각화

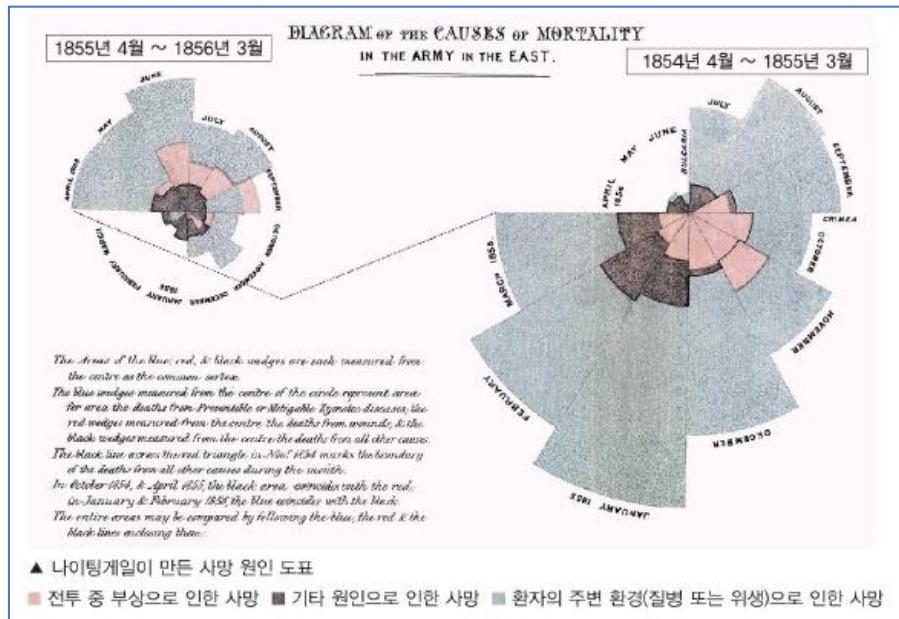




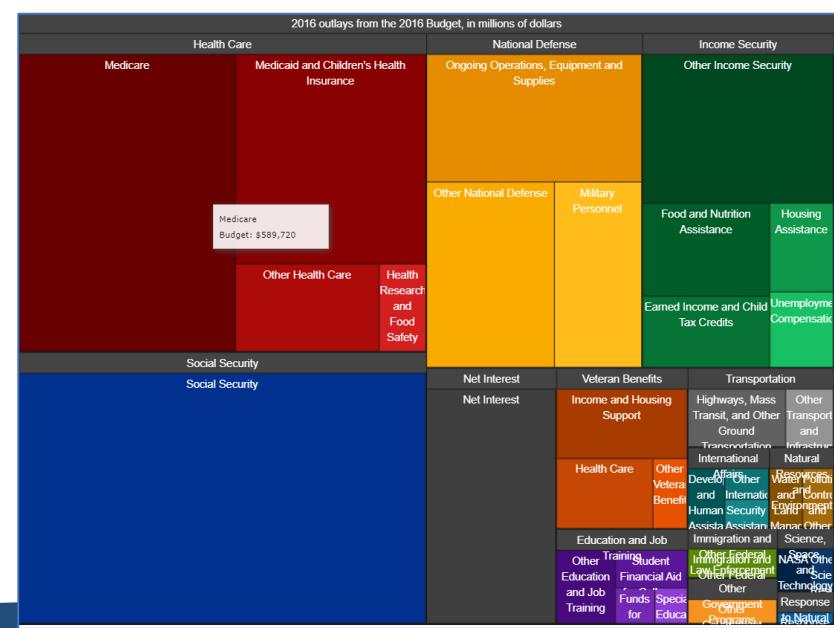
시각화

- 인간은 눈 망막 자극이 뇌에 전달될 때 정보의 의미를 파악할 수 있게 진화한 동물
- 인간은 감각 중의 대부분을 시각에 의존하는 데 일반적으로 시각이 77%, 청각이 13%, 후각이 7%, 나머지 3%는 촉각과 미각이 나눠가지는 것으로 알려져 있음
- 인간이 인간답게 활동하게 하는 가장 중요한 감각 정보기관이 시각
- 빅데이터 시대 데이터를 단순히 눈으로 볼 수 있는 한계를 넘어섬
- 데이터 시각화(Data visualization)는 매우 효과적으로 정보를 전달하는 수단이 됨

<나이팅게일의 전투 사망자 현황 시각화>



<미 예산 트리맵 >

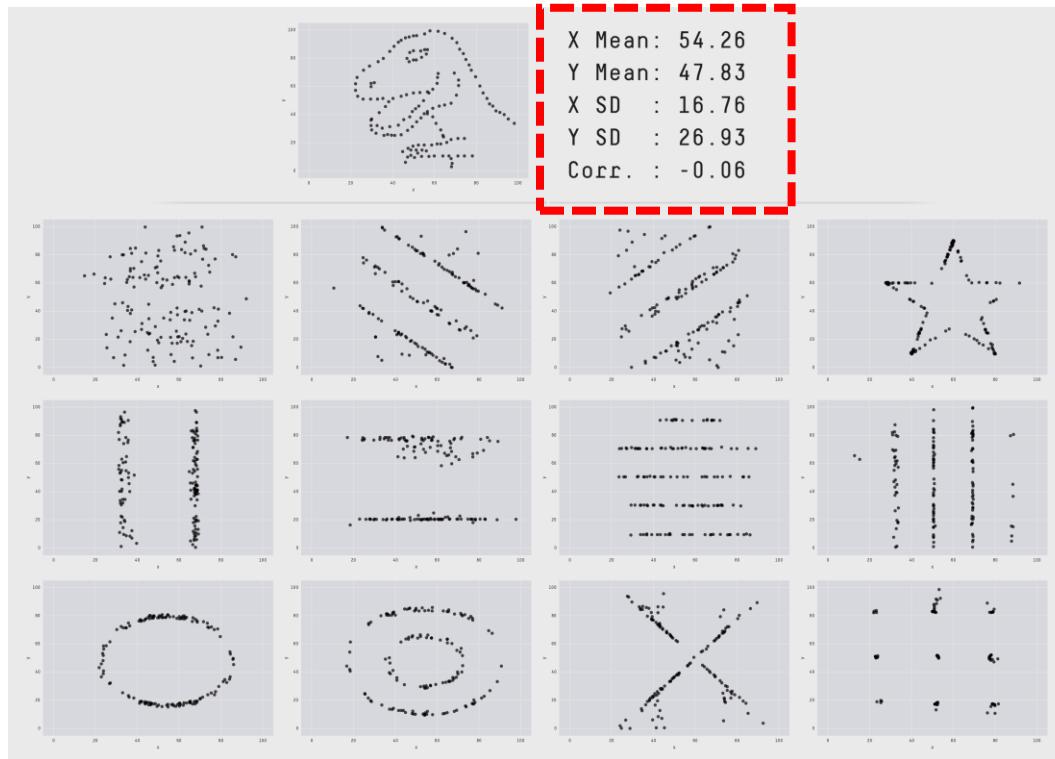




데이터 시각화의 중요성

- 'X 평균 54.26, Y 평균 47.83'으로 모두 같은 데이터
- 두 자리 소수점까지 동일한 기술 통계(평균, 표준편차, 피어슨의 상관계수)를 갖고 있음

모두 같은 데이터 인가요?

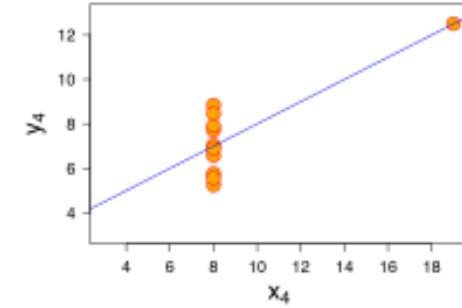
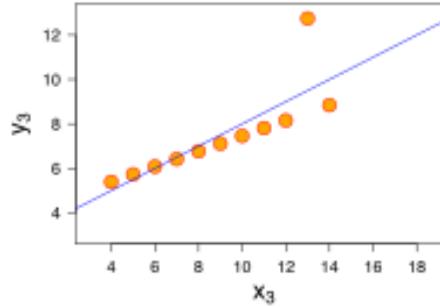
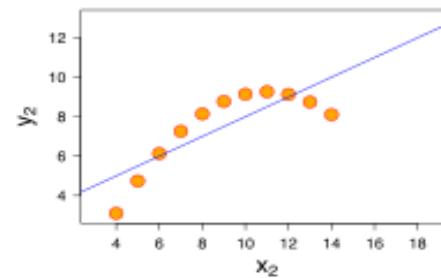
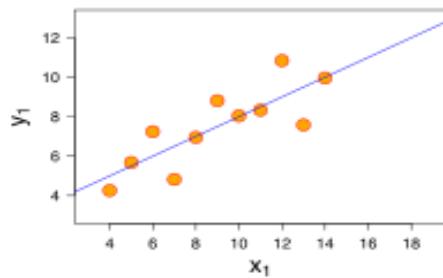




데이터 시각화의 중요성

- 데이터 시각화의 전형적인 사례로 앤스콤 4분할 그래프
- 앤스콤 4분할 그래프는 영국 프랭크 앤스콤이 데이터를 수치적으로만 확인할 때 발생할 수 있는 함정을 보여주기 위해 만든 그래프

dataset	x							y								
	count	mean	std	min	25%	50%	75%	max	count	mean	std	min	25%	50%	75%	max
I	11.0	9.0	3.316625	4.0	6.5	9.0	11.5	14.0	11.0	7.500909	2.031568	4.26	6.315	7.58	8.57	10.84
II	11.0	9.0	3.316625	4.0	6.5	9.0	11.5	14.0	11.0	7.500909	2.031657	3.10	6.695	8.14	8.95	9.26
III	11.0	9.0	3.316625	4.0	6.5	9.0	11.5	14.0	11.0	7.500000	2.030424	5.39	6.250	7.11	7.98	12.74
IV	11.0	9.0	3.316625	8.0	8.0	8.0	8.0	19.0	11.0	7.500909	2.030579	5.25	6.170	7.04	8.19	12.50





데이터 시각화

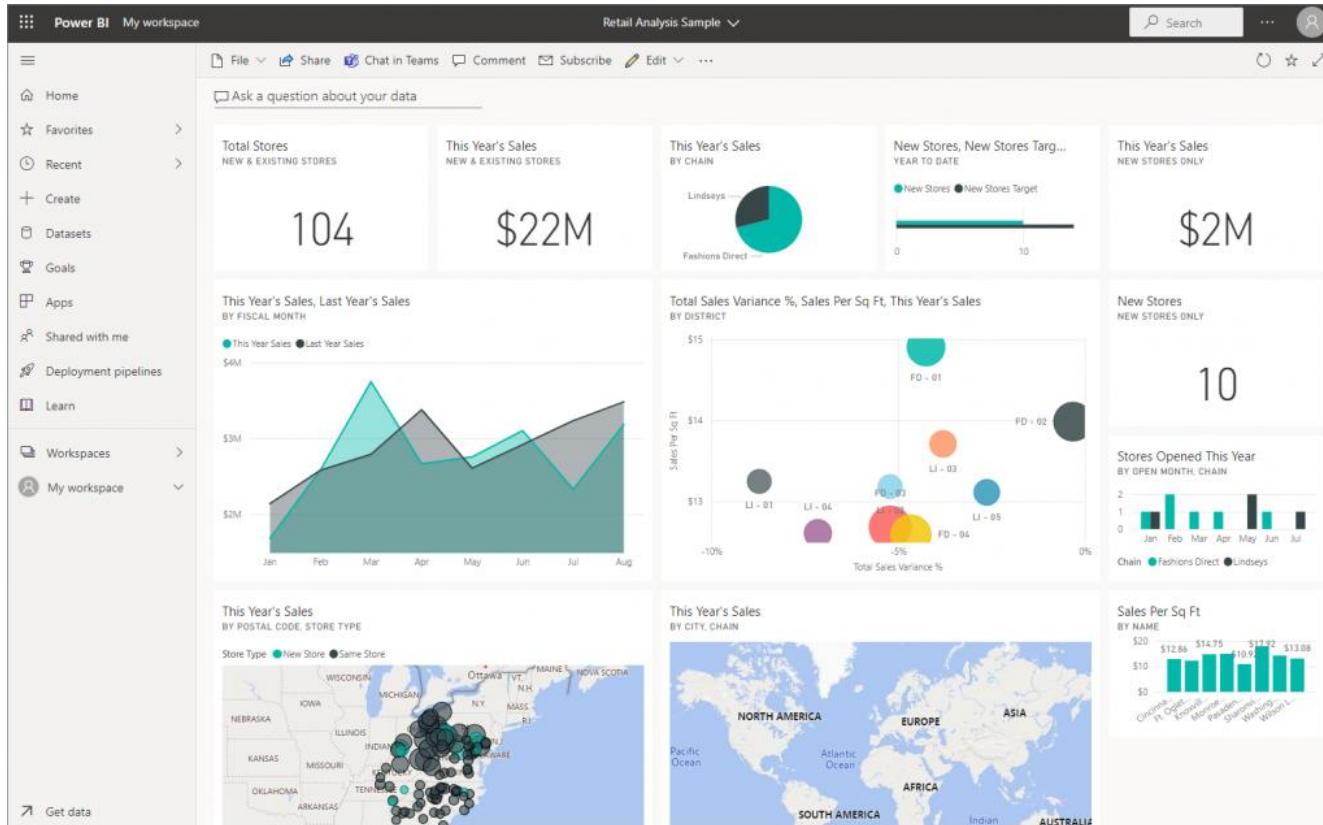
- 데이터 분석 결과를 쉽게 이해할 수 있도록 시각적으로 표현하고 전달되는 과정
- 도표(graph)라는 수단을 통해 정보를 명확하고 효과적으로 전달하는 것
- 전통적인 시각화 기술은 주로 시스템 로그나 실험 분석 결과 등에 대한 통계정보를 그래프로 보여주는 방식
- 빅데이터의 시각화 측면에서는 모든 데이터를 살펴보는 것에 제약이 있어, 시각화의 기술적인 요소와 데이터를 요약하고, 한 눈에 살펴볼 수 있도록 돋는 시각화 방법론적 요소의 중요성이 커지고 있음
- 데이터를 활용하여 수치 안에 숨겨진 인사이트를 발견
- 데이터 시각화는 정보 공유 및 커뮤니케이션을 위한 강력한 도구
- 데이터 대시보드, 대화형 보고서, 차트, 그래프, 기타 시각적 표현
- Excel의 차트 기능으로 가장 기초적인 데이터 시각화를 시작할 수 있음



데이터 시각화

- 빅데이터 시대 데이터를 단순히 눈으로 볼 수 있는 한계를 넘어섬
- 데이터 시각화(Data visualization)는 매우 효과적으로 정보를 전달하는 수단이 됨
- 빅데이터를 읽어 내기 위하여 데이터 시각화는 반드시 필요

<MS社의 시각화 도구 SW PowerBI>





데이터 시각화

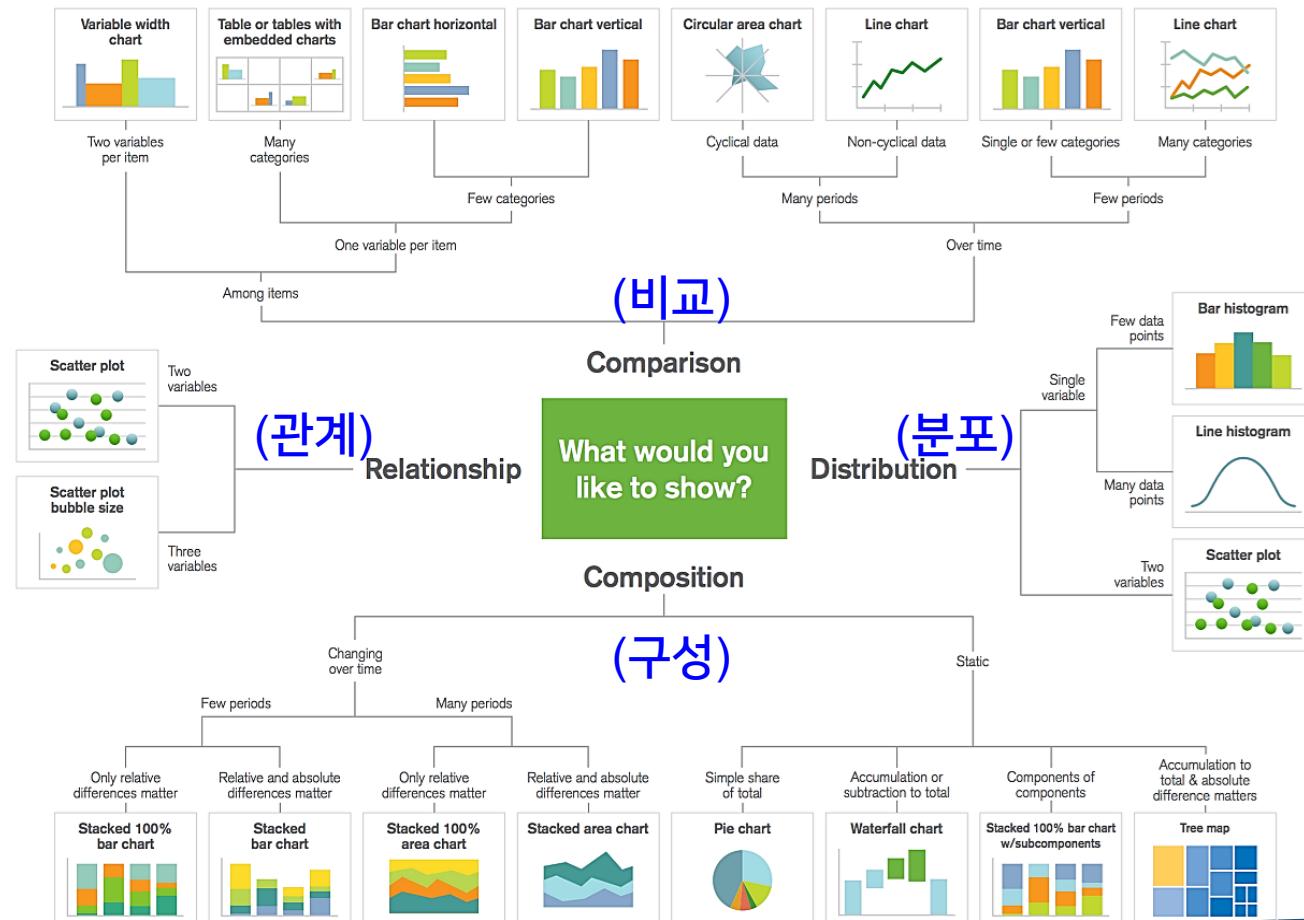
- 정보와 데이터를 그래프로 나타내는 것을 의미
- 차트, 그래프 또는 맵과 같은 시각적 요소를 사용해 데이터를 표시하는 프로세스
- 복잡하거나, 용량이 많거나, 숫자로 된 데이터를 처리하기 쉬운 시각적 표현으로 변환
- 시각적 표현을 사용해 원시 데이터에서 실행 가능한 인사이트를 추출
- 데이터 시각화를 위해 3가지 주요 구성 요소를 결합
 - 데이터 시각화의 목적
 - 스토리를 설명하는 데 도움이 되는 적절한 데이터 세트
 - 인사이트를 도출하는데 가장 적합한 시각화 방법
- 차트, 그래프, 맵과 같은 시각적 요소를 사용하여 데이터에서 추세, 이상값 및 패턴을 보고 이해할 수 있도록 해주는 접근하기 쉬운 방법



데이터 시각화 종류

● 앤드류 아벨라(Andrew V. Abela)의 차트 선택 방법

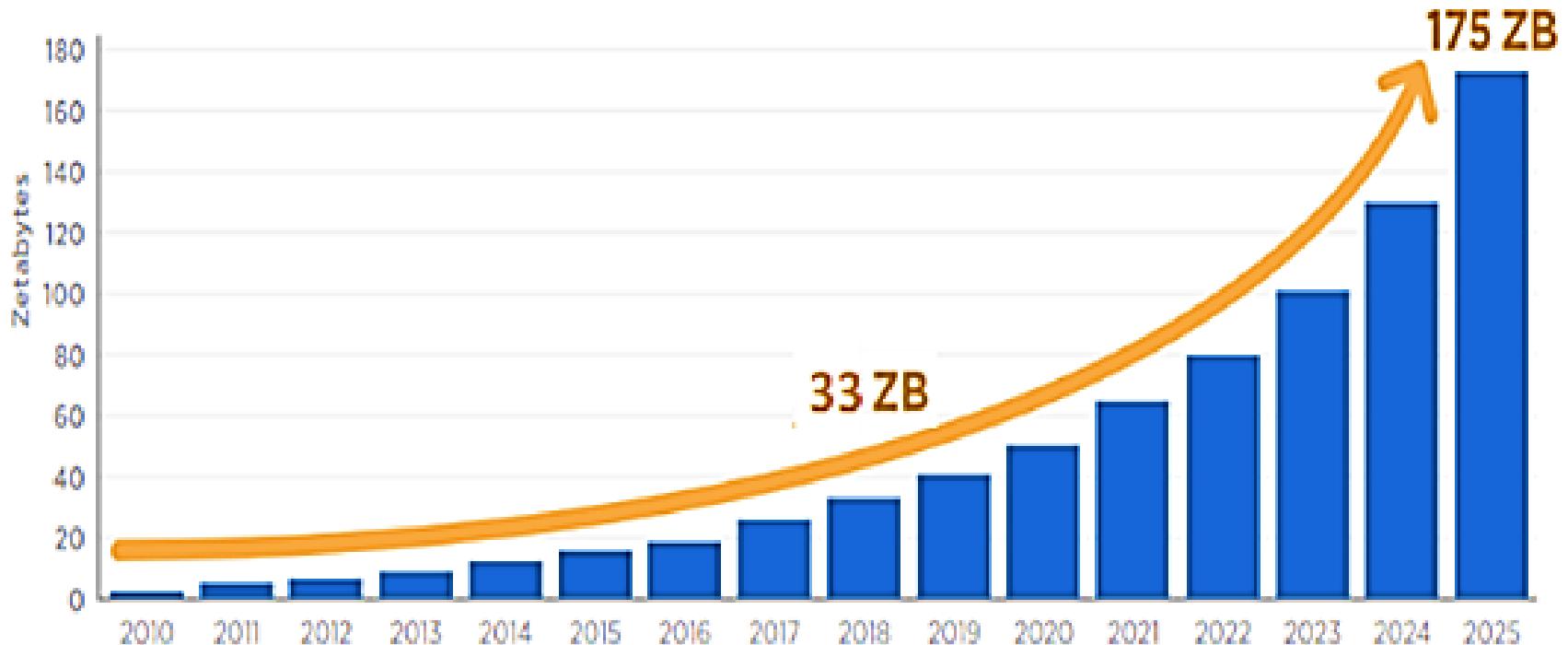
- 비교(Comparison), 구성(Composition), 분포(Distribution), 관계(Relationship)의 카테고리로 차트를 구분 (← 무엇을 보여주고 싶은지 목적에 따라)



지금은 빅데이터 시대



● 연도별 전 세계 데이터 생산량

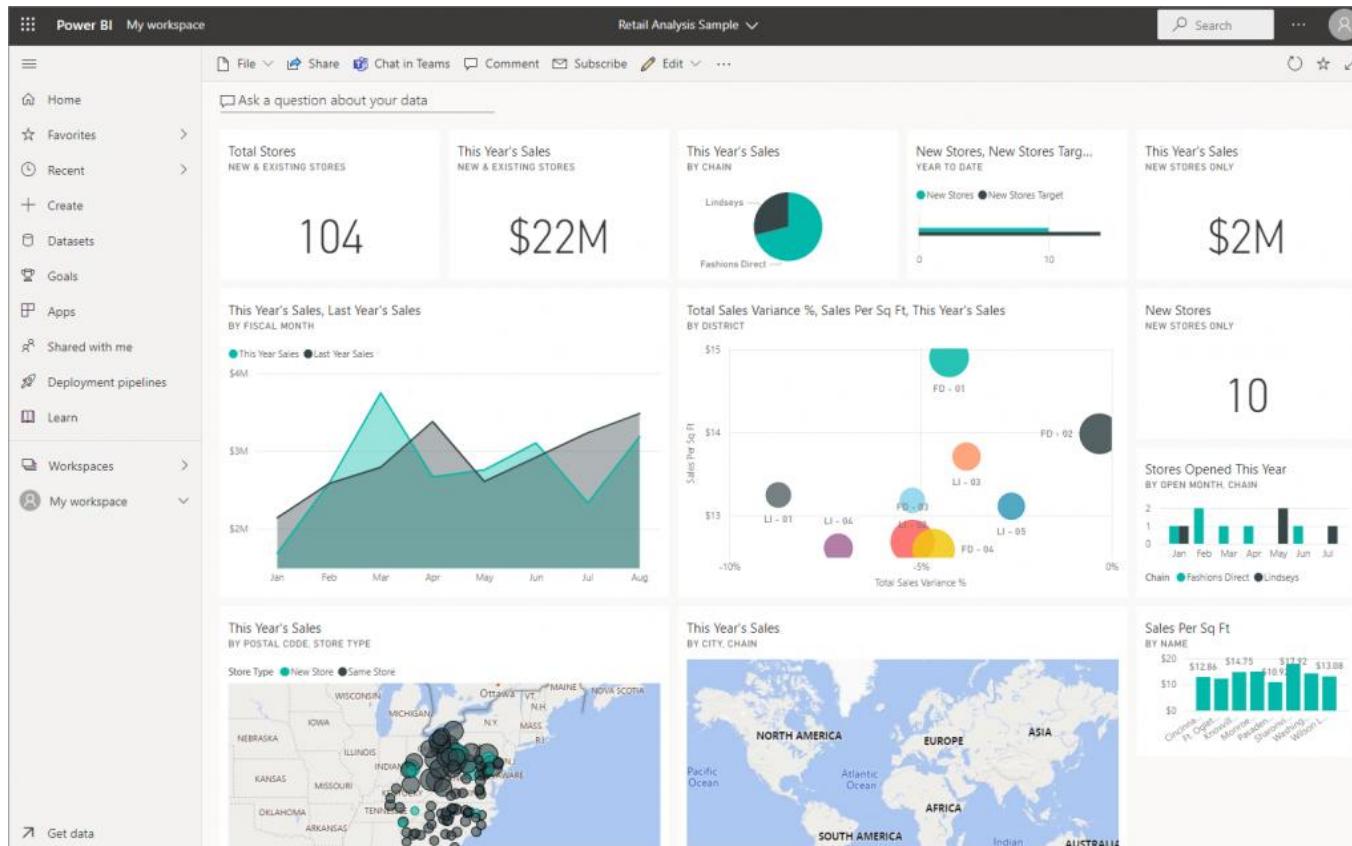


자료: IDC, IBK부자증권



빅데이터 시각화

- 빅데이터 시대 데이터를 단순히 눈으로 볼 수 있는 한계를 넘어섬
- 데이터 시각화(Data visualization)는 매우 효과적으로 정보를 전달하는 수단이 됨
- 빅데이터를 읽어 내기 위하여 데이터 시각화는 반드시 필요



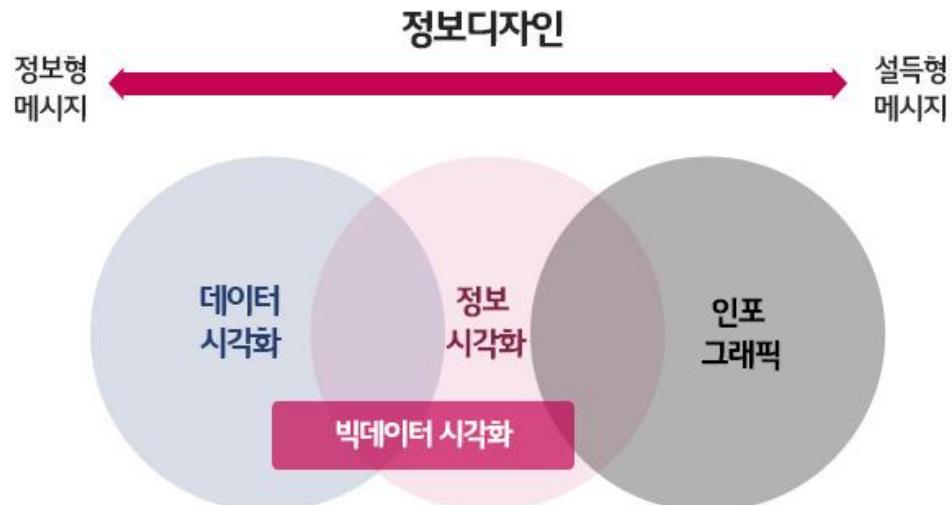


빅데이터 시각화

- 정보 시각화를 기반, 대규모 데이터(수량 및 비수량)를 색채, 통계, 이미지 등을 활용하여 시각적으로 표현하는 것을 의미함

- 정보 시각화

- 시간 시각화
- 분포 시각화
- 관계 시각화
- 비교 시각화
- 공간 시각화



- 빅데이터 시각화의 사전 작업

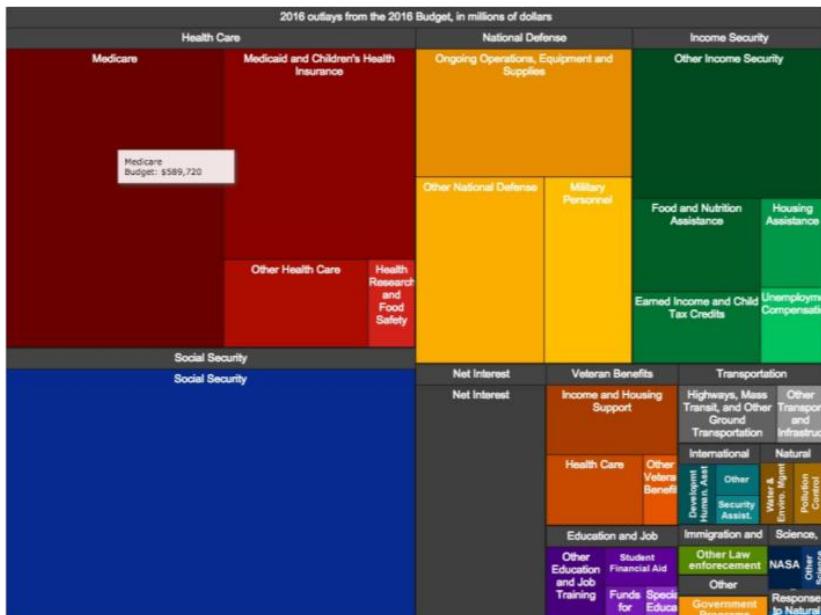
[출처:한국데이터베이스진흥원의 데이터분석전문가 가이드도서, 재구성]





데이터시각화의 매력

백악관의 2016년 미국 예산 트리맵



언제 누가 만들었나요?

2016년 미국 백악관

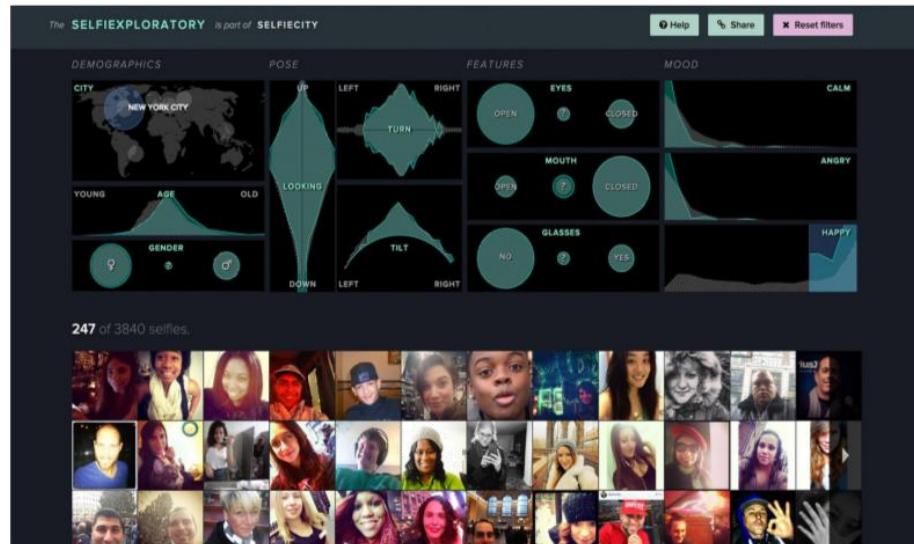
만든 목적은요?

한 해동안 정부의 예산이
어떻게 쓰였는지 알기 위해

무엇을 얻었나요?

세금, 특히 고액납세자의 세금이
어떻게 쓰였는지 분석해 인사이트 도출

OFFC의 '셀피 시티(Selfiecity)'



언제 누가 만들었나요?

2014년 뉴욕의 리서치&디자인 스튜디오 OFFC

만든 목적은요?

사회적 관행이 된 셀피 분석을 통해
전 세계 현대인의 성향 파악

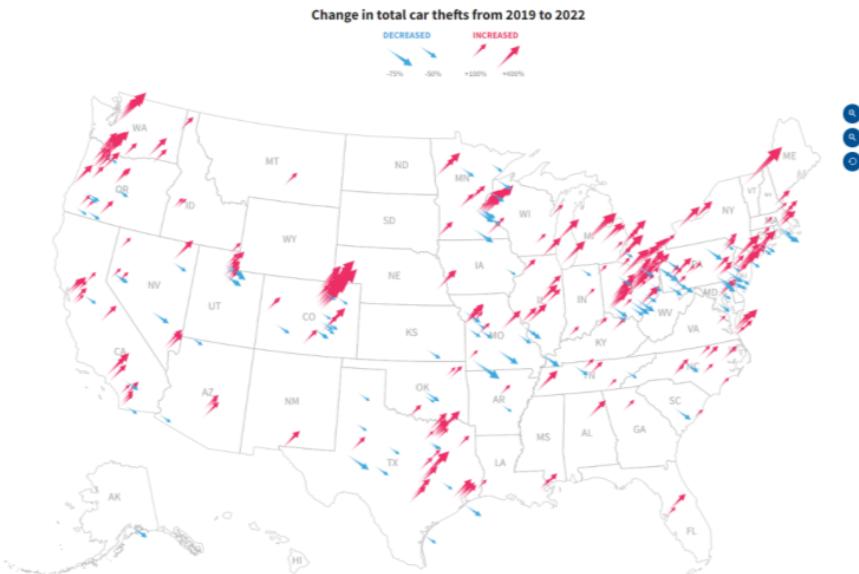
어떤 결론을 얻었나요?

전 세계인의 셀피에 있는 특정 패턴
파악을 통한 다양한 인사이트 도출

데이터시각화의 매력



USA Facts의 차량 절도율 상승 자료

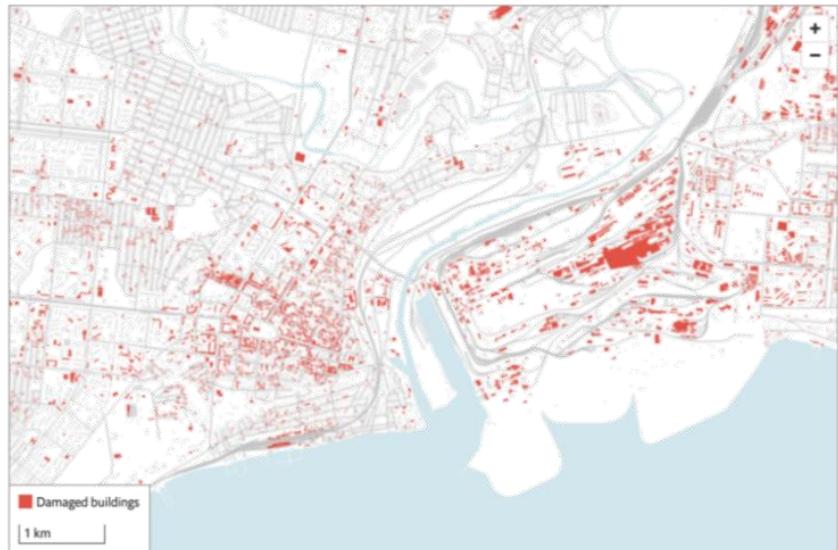


언제 누가 만들었나요?
2023년 비영리조직 USAFacts

만든 목적은요?
최근 미국 내 차량 절도율 상승 원인 파악

어떤 결론을 얻었나요?
지난 2022년 틱톡의 한 영상이 발단이 돼
특정 브랜드의 차량 절도율이 크게 상승함

이코노미스트의 우크라이나 피해도 파악



언제 누가 만들었나요?
2023년 언론사 이코노미스트

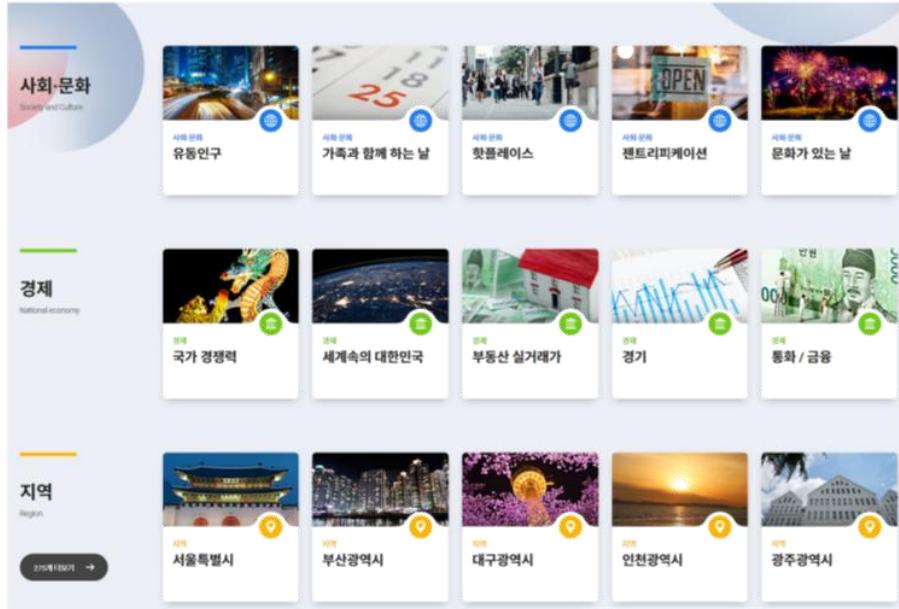
만든 목적은요?
전쟁으로 인한 우크라이나의 피해 상황 파악

어떤 결론을 얻었나요?
도시의 건축 면적 중 약 20%가 큰 피해를 입음



데이터시각화의 매력

한국데이터산업진흥원의 데이터 안심구역 Vox의 '한국인의 평균 키가 커진 이유'



누가 만들었나요?
한국데이터산업진흥원

만든 목적은요?
희소가치 높고 유용한 국내 데이터를
누구나 접할 수 있게 하기 위해

무엇을 얻었나요?

사회·문화, 경제, 지역 등 다양한 측면에서
데이터 분석을 통한 인사이트 도출



언제 누가 만들었나요?
2023년 유튜브채널 Vox

만든 목적은요?
한국인의 평균 신장율이
세계에서도 눈에 띄게 높아진 원인 파악

어떤 결론을 얻었나요?

높은 경제 성장율을 기록하며
식량공급 문제를 개선, 평균 키도 커짐



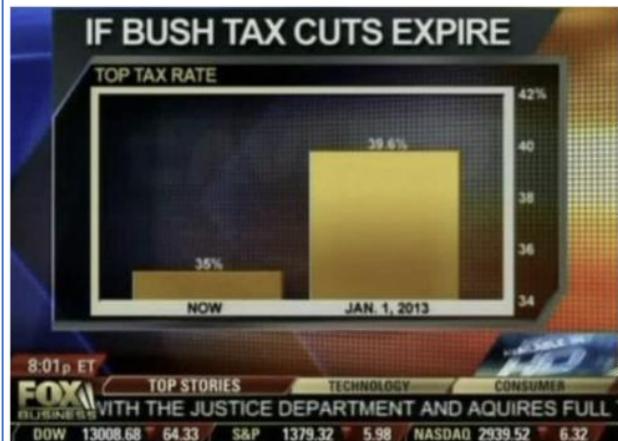
데이터시각화의 문제

- 데이터 시각화는 정보의 잘못된 표현이나 특정 사실의 과장으로 이어질 수 있는 몇 가지 문제를 갖고 있음
- 데이터의 지나친 단순화
 - 데이터를 지나치게 단순화하면 주요 정보가 손실
 - 데이터 시각화는 지나치게 단순화되어 인구 통계와 학업 성과에 영향을 미치는 기타 여러 요인을 고려하지 않았음
- 편견
 - 편견은 데이터 시각화에 부정적인 영향을 미침
 - 편견을 갖고 데이터를 선택하고 정리하여 의도치 않게 편향된 결과가 나올 수 있음
- 과장
 - 상관 없는 데이터를 시각화하여 존재하지 않는 상관 관계를 생성할 수 있음

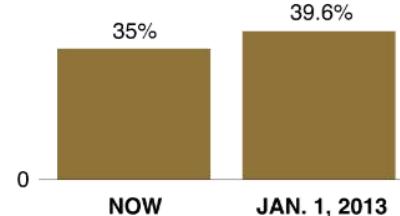


데이터시각화의 문제

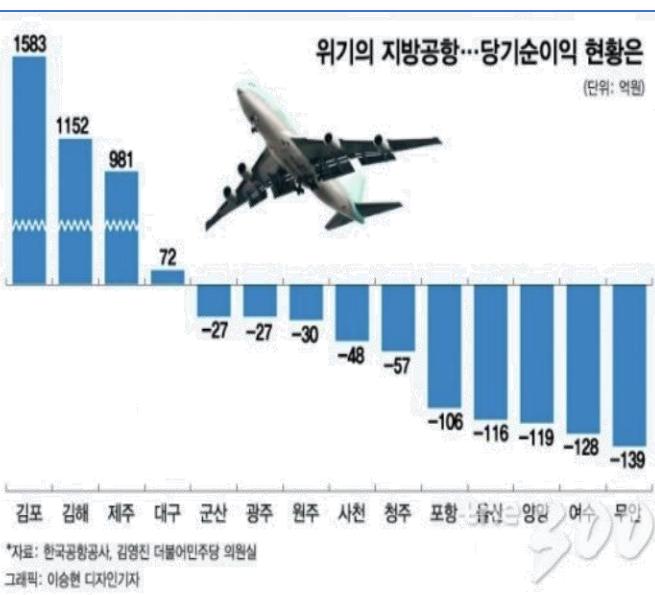
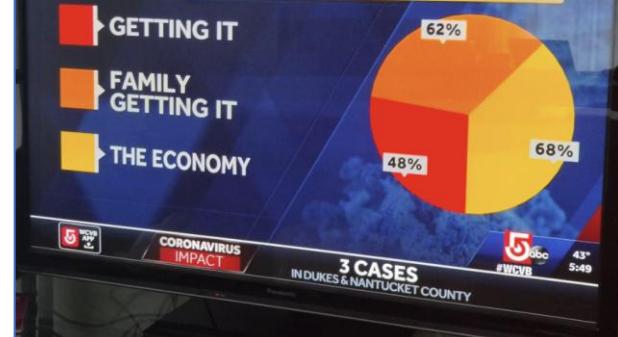
원본



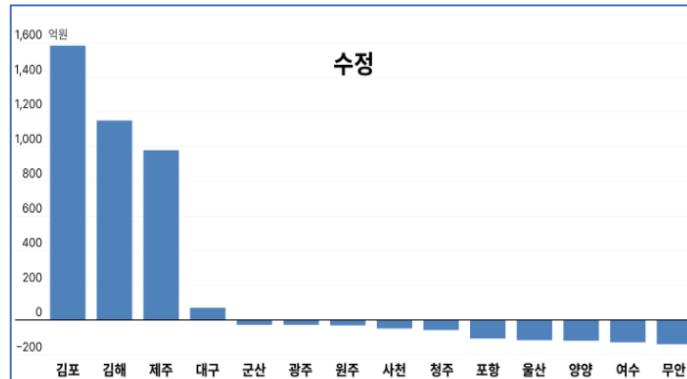
수정



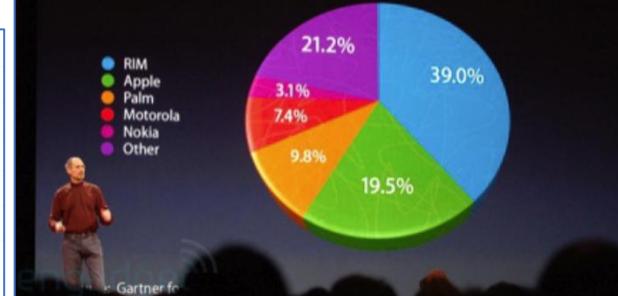
BIGGEST COVID-19 WORRIES



수정



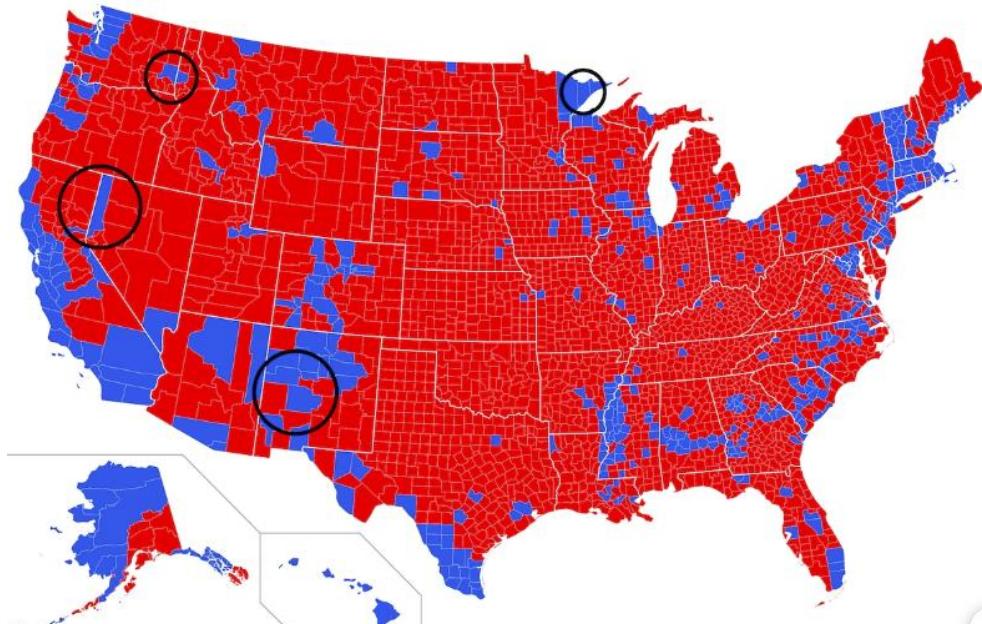
U.S. SmartPhone Marketshare



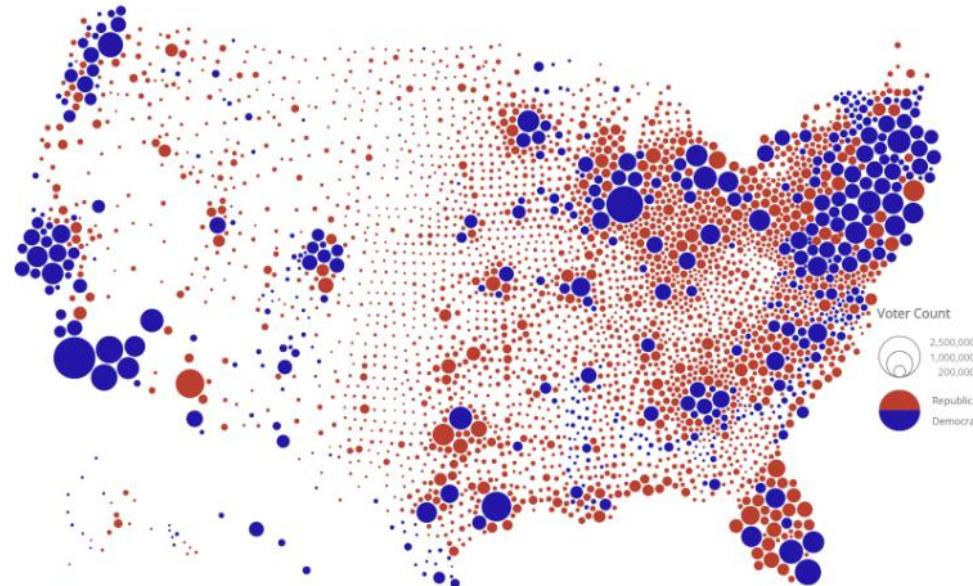


데이터시각화의 문제

- 면적과 인구 (선거 득표율)



<출처: 워싱턴 포스트>



<출처: Fast Company>

데이터시각화를 위한 데이터 표현에 중요한 것은 **맥락!**

Excel의 데이터 시각화





데이터시각화 도구

- Excel
- PowerBI
- Tableau
- IBM Cognos Analytics
- Datawrapper
- Amplitude
-



IBM Cognos Analytics



Datawrapper



Excel 차트의 종류

● 세로 또는 가로 막대형

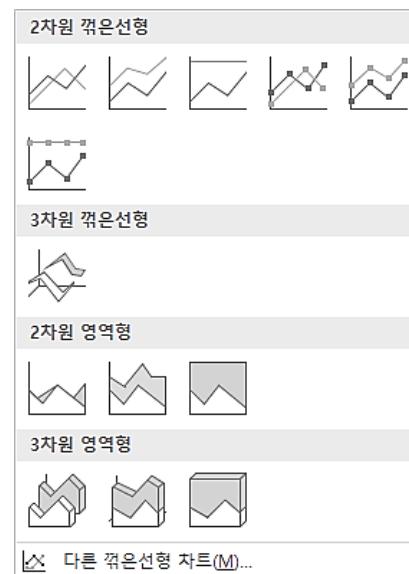
- 여러 항목 값의 크기나 상관관계를 비교하기에 적합

● 꺾은선형 또는 영역형

- 시간이나 항목에 따른 값의 변화나 추세를 확인하기에 적합

● 원형 또는 도넛형

- 전체에서 항목이 차지하는 비율이나 크기를 표현하기에 적합





Excel 차트의 종류

● 계층 구조(트리맵, 선버스트)

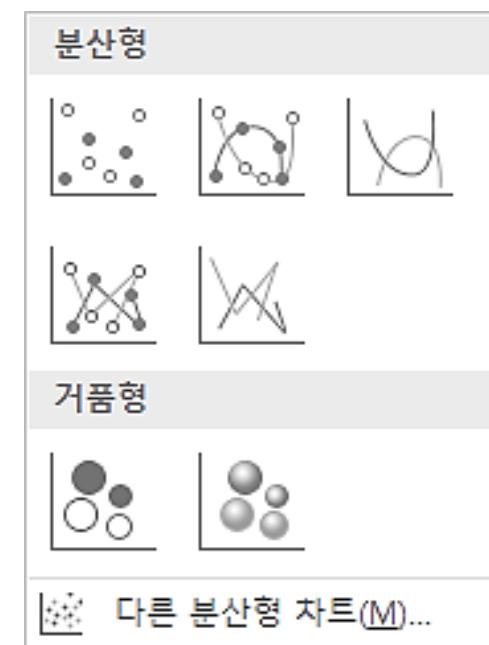
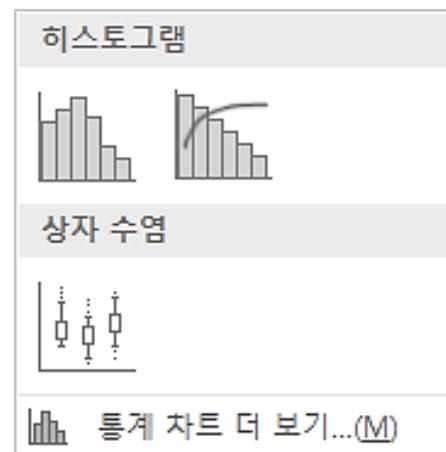
- 어떤 항목을 구성하는 하부 항목이 존재할 때 상위 항목에 속한 하위 항목을 관찰하거나 비교하기에 적합

● 통계(히스토그램, 상자 수염)

- 어떤 값(예: 평균 등)을 기준으로 데이터의 분포나 빈도를 확인하는 등 통계 분석 결과를 표현하기에 적합

● 분산형/거품형

- 2~3개 항목을 비교하거나 경향을 살피기에 적합

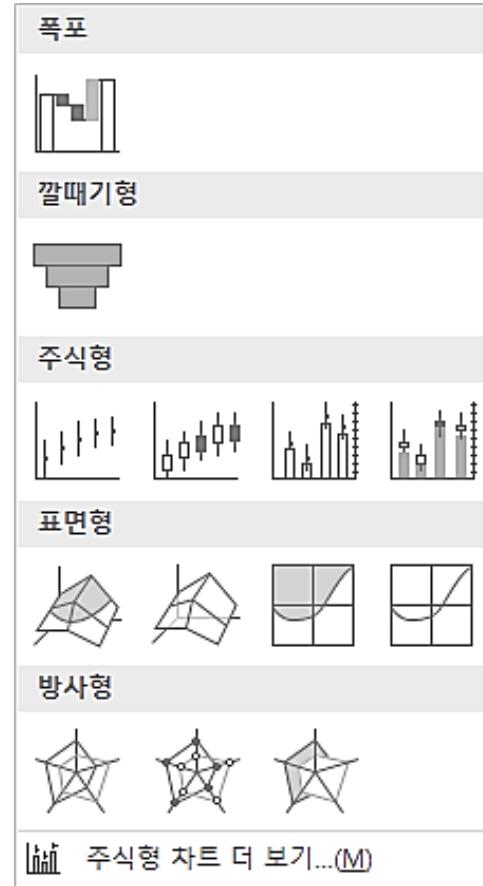




Excel 차트의 종류

● 콤보(혼합) 차트

- 보조 축을 추가하여 2개 이상의 차트로 더 광범위한 내용을 확인해야 할 때 적합



● 기타

- 폭포, 깔때기형, 주식형, 표면형, 방사형



차트 구성 요소

- [삽입]메뉴 → [차트]그룹 → [세로 또는 가로 막대형 차트]메뉴
- 가장 기본 차트인 세로막대형 차트의 구성요소
- 데이터 수치는 데이터 계열로 표현

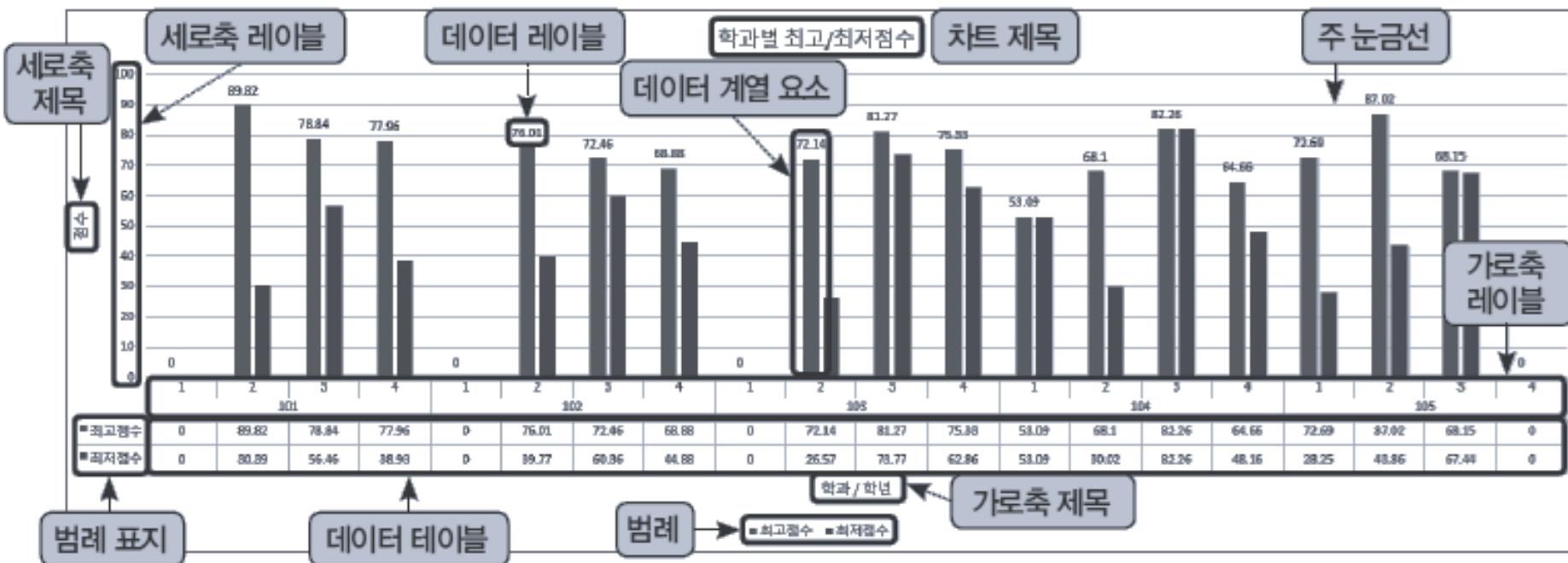




차트 작성

- [차트도구][디자인]메뉴탭 → [차트요소추가]메뉴

구성 요소 삽입/삭제 메뉴

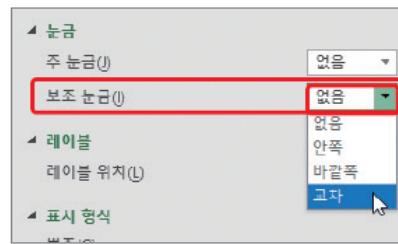
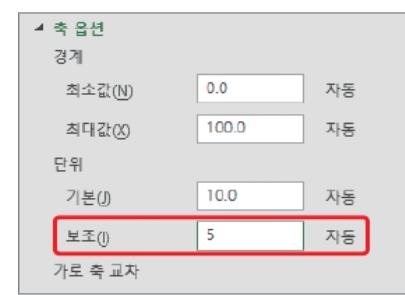
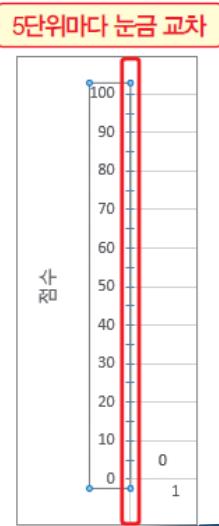
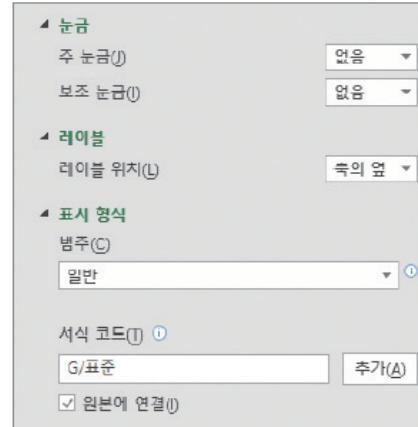
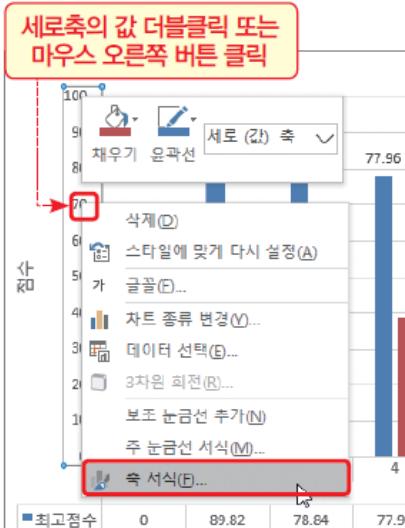
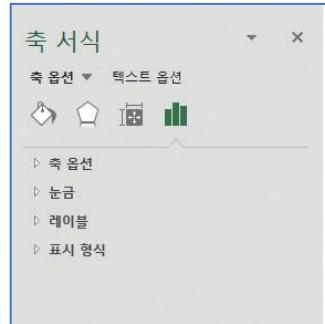
축	기본 가로(H)
축 제목(A)	기본 세로(V)
차트 제목(C)	다른 축 옵션(M)...
데이터 레이블(D)	
데이터 테이블(B)	
오차 막대(E)	
눈금선(G)	
범례(L)	
선(I)	
주세선(I)	
양선/음선(U)	

데이터 테이블(B)	없음(N)
오차 막대(E)	범례 표지 포함(W)
눈금선(G)	범례 표지 없음(L)
범례(L)	기타 데이터 표 옵션(M)...
선(I)	
주세선(I)	
양선/음선(U)	



값의 크기를 표시하는 세로 축 편집

- 세로축을 마우스 클릭 선택한 후 더블클릭 또는 단축메뉴의 서식기능 실행





데이터 레이블 추가 편집

● 데이터 값을 차트에 표시하는 데이터 레이블 추가

The screenshot illustrates the steps to add data labels to a bar chart in Microsoft Excel:

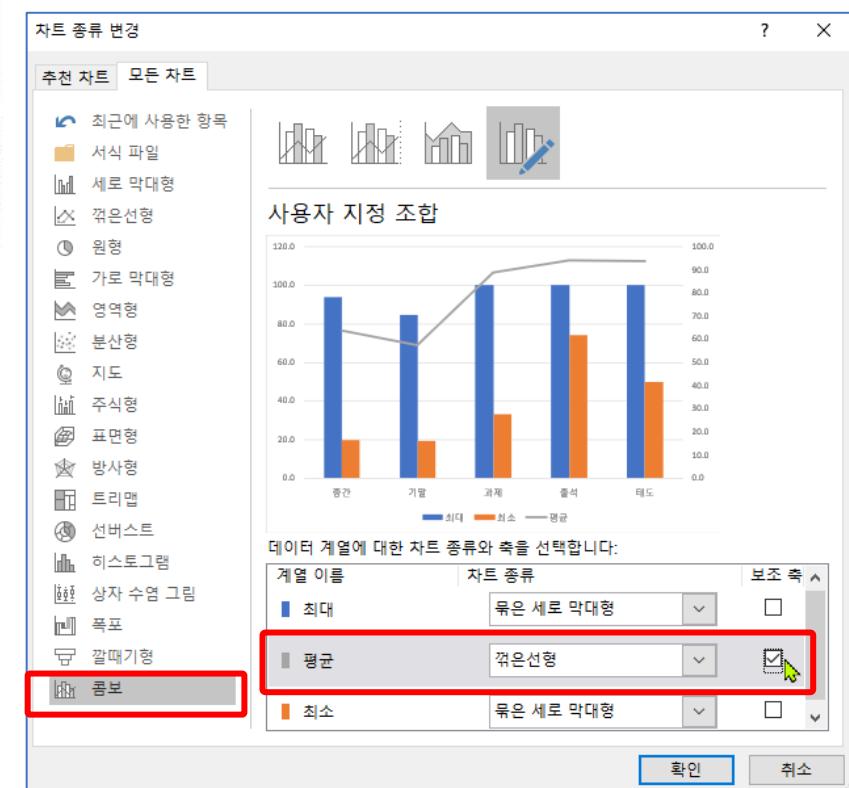
- Step 1: Open the Chart Tools ribbon.** The "Chart Tools" ribbon is open, showing the "Chart Style" tab selected. The "Add Chart Element" button (containing a plus sign) is highlighted with a red box.
- Step 2: Select "Data Labels".** A dropdown menu is open under the "Add Chart Element" button, with the "Data Labels" option highlighted with a red box.
- Step 3: Open the "Data Label Format" dialog.** The "Data Label Format" dialog is open, showing the "Text Options" tab selected. The "Text Options" tab is highlighted with a red box.
- Step 4: Set label options.** In the "Text Options" tab, the "Value" checkbox is checked, and the "Align" dropdown is set to "Center" (highlighted with a red box). The "Format Text Box" button is also highlighted with a red box.
- Step 5: Set font color.** The "Format Text Box" dialog is open, showing the "Color" tab selected. The "Solid Color" radio button is selected, and the color swatch is highlighted with a red box.

선택한 항목	설정 내용
색상	Solid Color (Solid)
선택한 항목	선택한 항목



혼합(콤보)차트

- 세로막대형과 선형을 혼합한 콤보차트 작성 가능(3차원 차트는 혼합불가)
- [차트도구][디자인]메뉴탭 → [차트종류변경]메뉴
- 작성된 세로막대형차트의 데이터 계열 중 일부를 깍은선형으로 변경

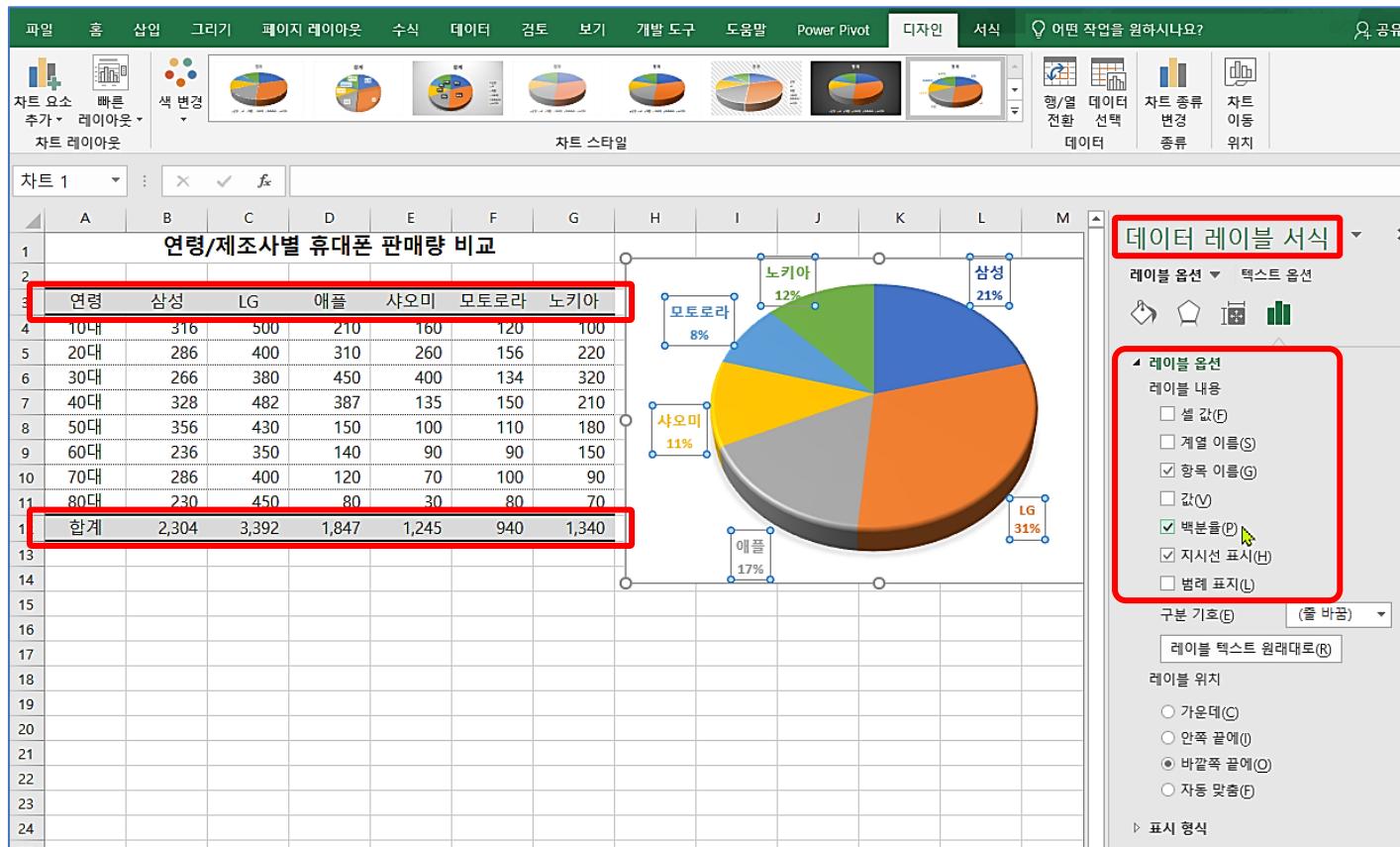


- 기본축이 아닌 보조축으로 값을 별도로 표시하기도 함

전체에서 항목이 차지하는 범위를 표시하는 원형 차트



- 원형차트는 데이터 계열이 하나여야 전체 중 차지하는 범위를 표시할 수 있음
 - 제조사별 판매량의 합계 데이터를 이용한 원형차트
 - 원형차트는 범례보다 데이터레이블을 표시하는 것이 일반적임





기술통계 차트

● 히스토그램

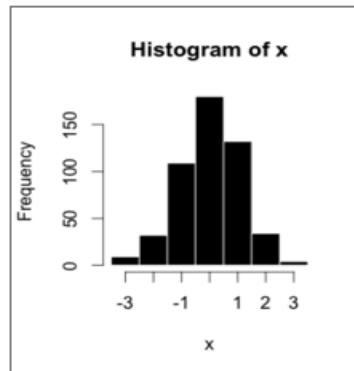
- 도수분포를 막대그래프로 시각화한 차트
- 그래프 모양(종모양, 비대칭 종모양 등)으로 데이터 분포를 파악할 수 있음

● 상자수염그림

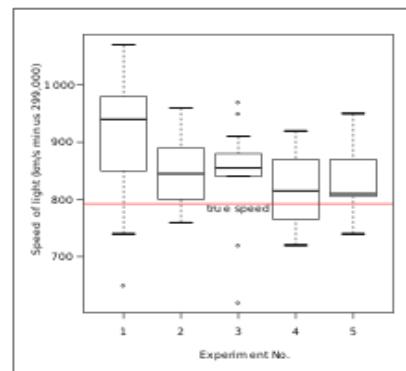
- 데이터 분포를 사분위수로 나타내며 평균, 중앙값 및 이상치를 표시할 수 있음

● 산점도

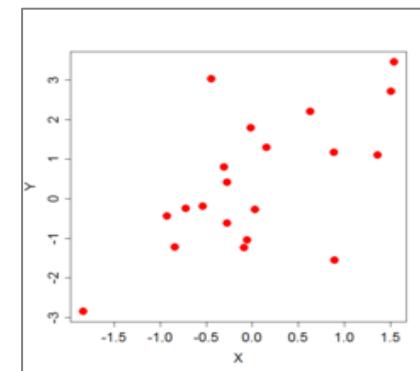
- 2개의 수치형 변수 간 관계를 시각적으로 표현
- X축, Y축의 변수를 기준으로 데이터 분포를 확인 (회귀식을 위하여 X축 독립변수, Y축 종속변수)
- Excel에서는 분산형 차트로 표현하고 선형 추세선으로 선형회귀식을 표시 가능



- 히스토그램(Histogram) -



- 상자수염그림(Box plots) -



- 산점도(Scatter plots) -



피벗 차트

● 데이터를 요약 집계한 피벗테이블로 차트를 작성

	A	B	C	D	E	F	G
1	+	주간 판매 현황					
2							
3							
4	판매일자	구매자	상품명	수량	판매가	금액	결제방법
5	02/08	이민영	MP3	2	125,000	250,000	카드
6	02/08	홍민수	디지털카메라	1	339,000	339,000	핸드폰
7	02/08	송미옥	보조배터리	4	59,000	236,000	K페이
8	02/08	최선우	태블릿	2	482,000	964,000	핸드폰
9	02/08	이민수	블랙박스	2	308,000	616,000	계좌이체
10	02/08	강수진	메모리	1	225,000	225,000	핸드폰
11	02/08	강철진	전자사전	2	225,000	450,000	계좌이체
12	02/09	김철수	MP3	1	125,000	125,000	카드
13	02/09	유민우	디지털카메라	2	339,000	678,000	계좌이체
45	02/14	○○○	○○○	○○○	○○○	○○○	○○○
50	02/14	박정우	디지털카메라	3	339,000	1,017,000	K페이
51	02/14	송철중	메모리	1	88,000	88,000	계좌이체
52	02/14	안오상	보조배터리	2	59,000	118,000	카드
53	02/14	전상철	블랙박스	1	308,000	308,000	K페이
54	02/14	정진욱	블랙박스	3	308,000	924,000	카드
55	02/14	강철진	전자사전	1	225,000	225,000	계좌이체

행 레이블	합계 : 수량	합계 : 금액
MP3	14	1,750,000
디지털카메라	17	5,763,000
메모리	9	929,000
보조배터리	9	531,000
블랙박스	14	4,312,000
전자사전	13	2,925,000
태블릿	8	3,682,000
총합계	84	19,892,000

피벗 테이블 도구

분석 디자인 어떤 작업을 원하시나요?

필드, 항목 및 집합 OLAP 도구

테이블 이동 관계 계산

피벗 차트

차트 삽입

모든 차트

최근에 사용한 항목

서식 파일

세로 막대형

꺾은선형

원형

가로 막대형

영역형

분산형

지도

주식형

표면형

방사형

트리맵

선바스트

히스토그램

상자 수염 그림

폭포

꺾여기형

콤보

사용자 지정 조합

데이터 개별에 대한 차트 종류와 측정 선택합니다.

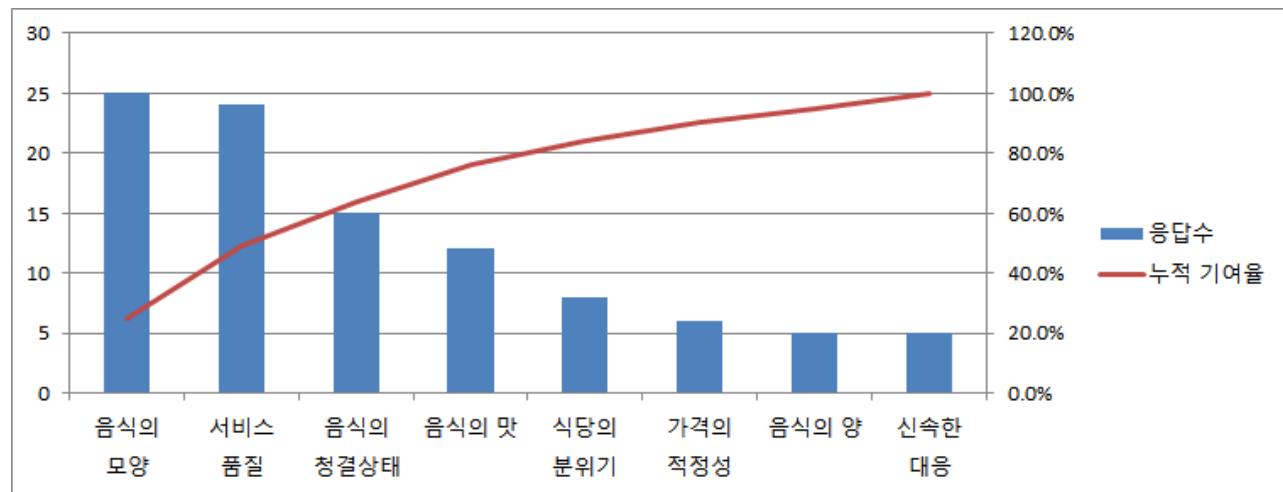
계열 이름	차트 종류	보조 측정
합계 : 수량	묶은 세로 막대형	<input type="checkbox"/>
합계 : 금액	꺾은선형	<input checked="" type="checkbox"/>

확인 취소



파레토차트

- 범주형 자료 중 기준 필드의 수준별 빈도 수를 순서대로 정렬하여 표현하는 차트
- "80/20" 규칙을 기반, 문제의 80% 정도가 20%의 원인에서 비롯되는 결과라는 의미 ("결정적 소수와 사소한 다수"라는 말로도 표현함)
- 파레토 차트는 막대 차트의 한 예시이며, 최고에서 저에서 높이 순서로 막대가 정렬되는 것이 특징임
- 주로 막대 차트에 누적 빈도 선을 추가하여 표현
- 예시



80%의 기여가 "음식의 모양"과 "서비스 품질" 또는 "음식의 청결상태" 사이에 가장 큰 비중을 차지하는 처음 2-3 가지 요인에 있음을 알 수 있음



피벗 차트로 파레토 차트 작성

- 요약된 피벗테이블의 내림차순 정렬 값 필드와 누적비율 데이터를 작성하고, 혼합형 차트로 파레토 차트 작성

The screenshot shows the Power BI PivotTable Fields pane. A context menu is open over the 'Sum' field in row 3. The menu items are: '합계 : 금액' (Sum : Amount), '정렬(S)' (Sort), and '숫자 내림차순 정렬(O)' (Sort Descending). The '정렬(S)' and '숫자 내림차순 정렬(O)' options are highlighted with red boxes.

행 레이블	합계 : 금액	합계 : 금액2
디지털카메라	5,763,000	5763000
블랙박스	4,312,000	4312000
태블릿	3,682,000	3682000
전자사전	2,925,000	2925000
MP3	1,750,000	1750000
메모리	929,000	929000
보조배터리	531,000	531000
총합계	19,892,000	19892000

행 레이블	합계 : 금액	합계 : 금액2
디지털카메라	5,763,000	28.97%
블랙박스	4,312,000	50.65%
태블릿	3,682,000	69.16%
전자사전	2,925,000	83.86%
MP3	1,750,000	92.66%
메모리	929,000	97.33%
보조배터리	531,000	100.00%
총합계	19,892,000	19892000

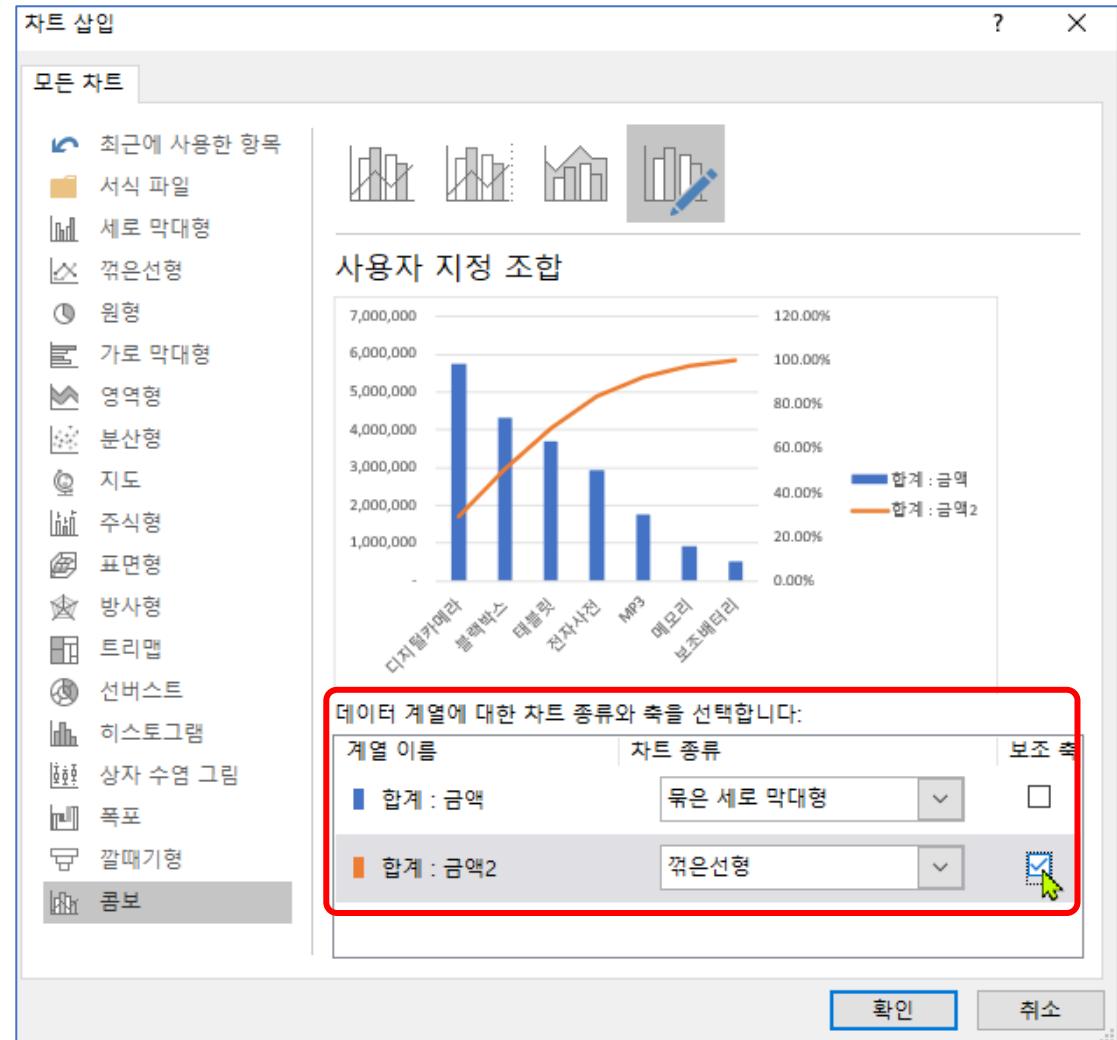
The screenshot shows the '값 필드 설정' (Value Field Settings) dialog box. It has tabs for '값 표시 기준(A)' (Value Display Criteria) and '값 표시 형식(A)' (Value Display Format). The '값 표시 형식(A)' tab is selected. Under '값 표시 형식(A)', the dropdown menu is set to '누계 비율' (Cumulative Ratio), which is highlighted with a blue selection bar. Other options in the dropdown include '상위 합계 비율' (Top Sum Ratio), '[기준값]과의 차이' (Difference from [Base Value]), '[기준값]에 대한 비율의 차이' (Difference of Ratios from [Base Value]), and '누계'. At the bottom of the dialog box are '표시 형식(N)' (Display Format), '확인' (OK), and '취소' (Cancel) buttons.



피벗 차트로 파레토 차트 작성

- 요약 집계된 피벗테이블의 계산 필드 내림차순 정렬 후 피벗 차트 작성

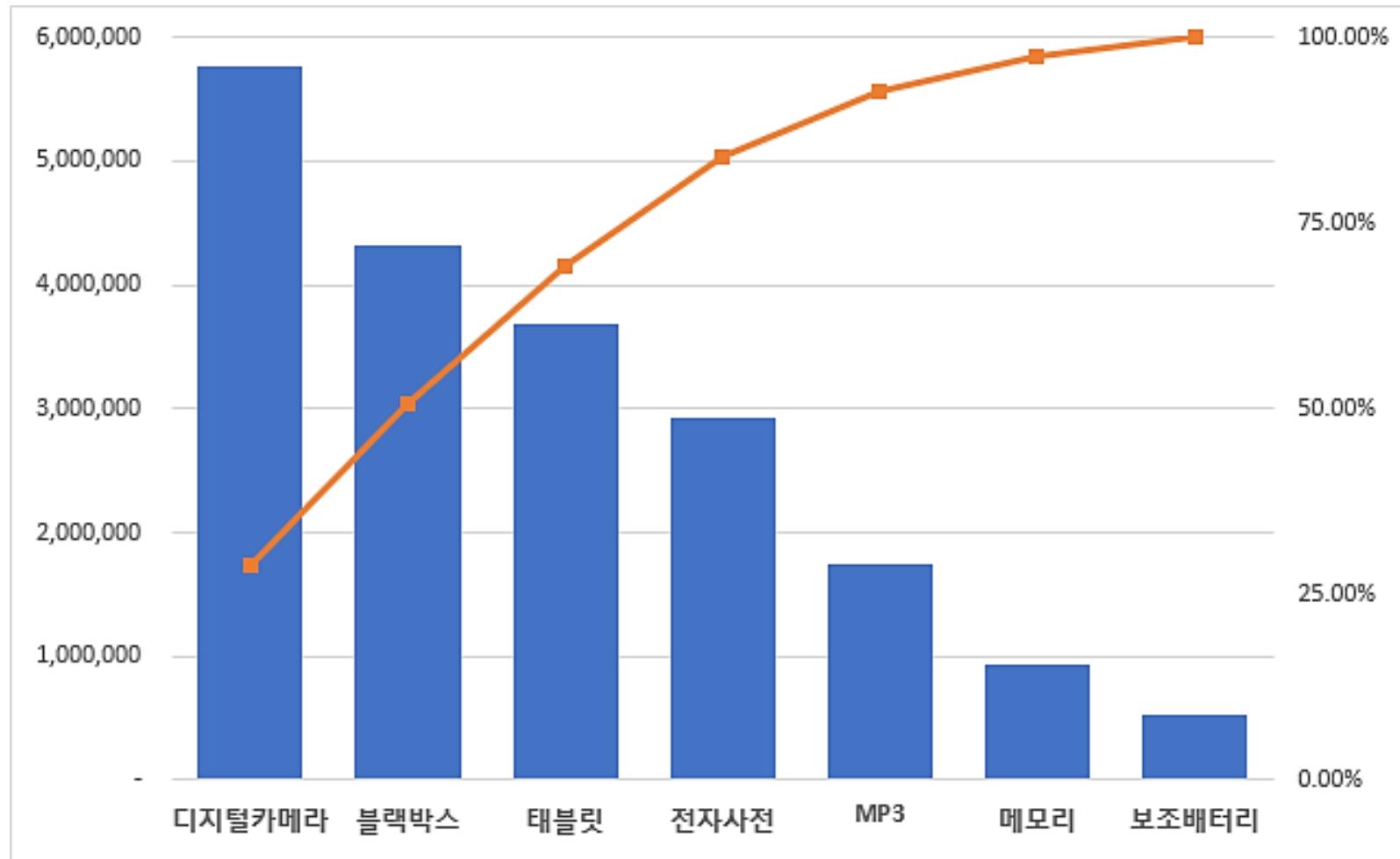
행 레이블	합계 : 금액	합계 : 금액2
디지털카메라	5,763,000	28.97%
블랙박스	4,312,000	50.65%
태블릿	3,682,000	69.16%
전자사전	2,925,000	83.86%
MP3	1,750,000	92.66%
메모리	929,000	97.33%
보조배터리	531,000	100.00%
총합계	19,892,000	





파레토 차트

● 파레토 차트 작성 결과



통계 기반 데이터 시각화





데이터 시각화: 변량별 차트 선택

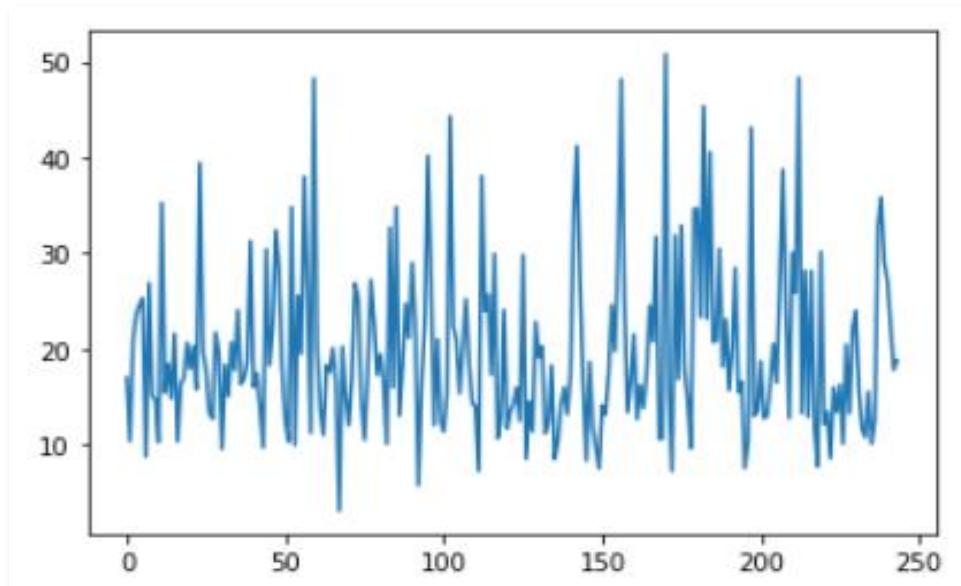
- 변수(변량): 데이터를 구성하는 단위, 변하는 양이라는 의미를 포함
- 일변수, 이변수, 다변수에 따라 적절한 차트를 선택

변수 개수	변수 형태	추천 그래프
일변수	범주형	막대 그래프(Bar Chart)
		원 그래프(Pie Chart)
	연속형	박스 그래프(Box Plot)
		바이올린 그래프(Violin Plot)
		히스토그램(Histogram)
		커널밀도곡선(Kernel Density Curve)
이변수이상	범주형	히트맵(Heatmap)
	연속형	산점도(Scatter)
		선 그래프(Line Chart)



선그래프

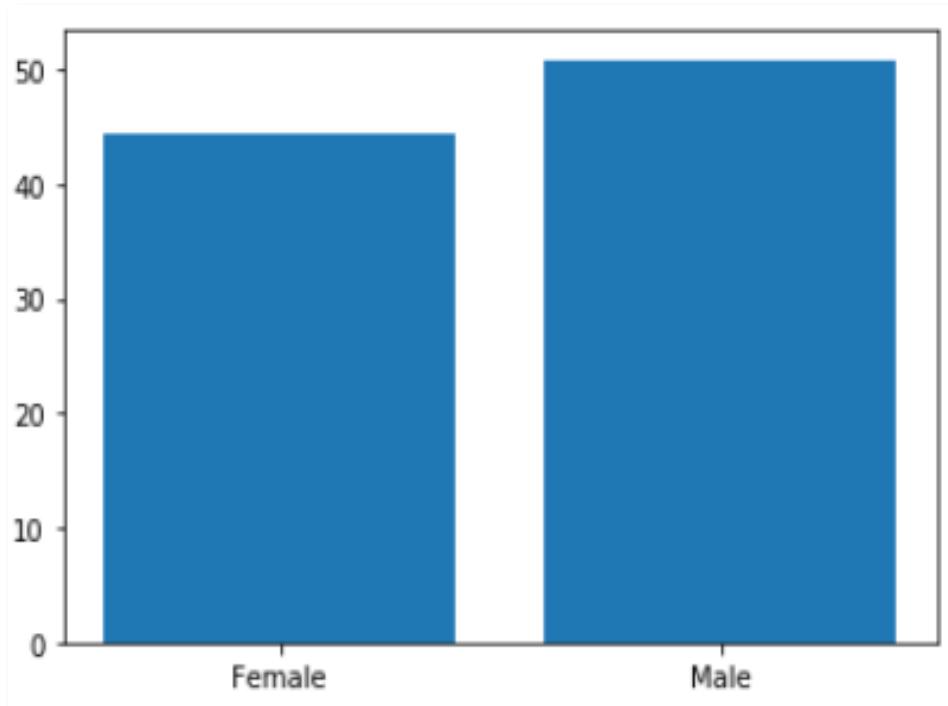
- 선그래프(line plot)는 ‘연속하는 데이터 값’들을 직선이나 곡선으로 연결하여 데이터 값 사이 관계를 나타냄
- 선 그래프에 입력할 x, y 축 데이터를 선택
- 선 그래프는 x축에 시간 변수를 지정하여 시간 변화량을 나타내는 시계열 데이터를 표현하는 그래프로도 많이 사용됨





막대그래프

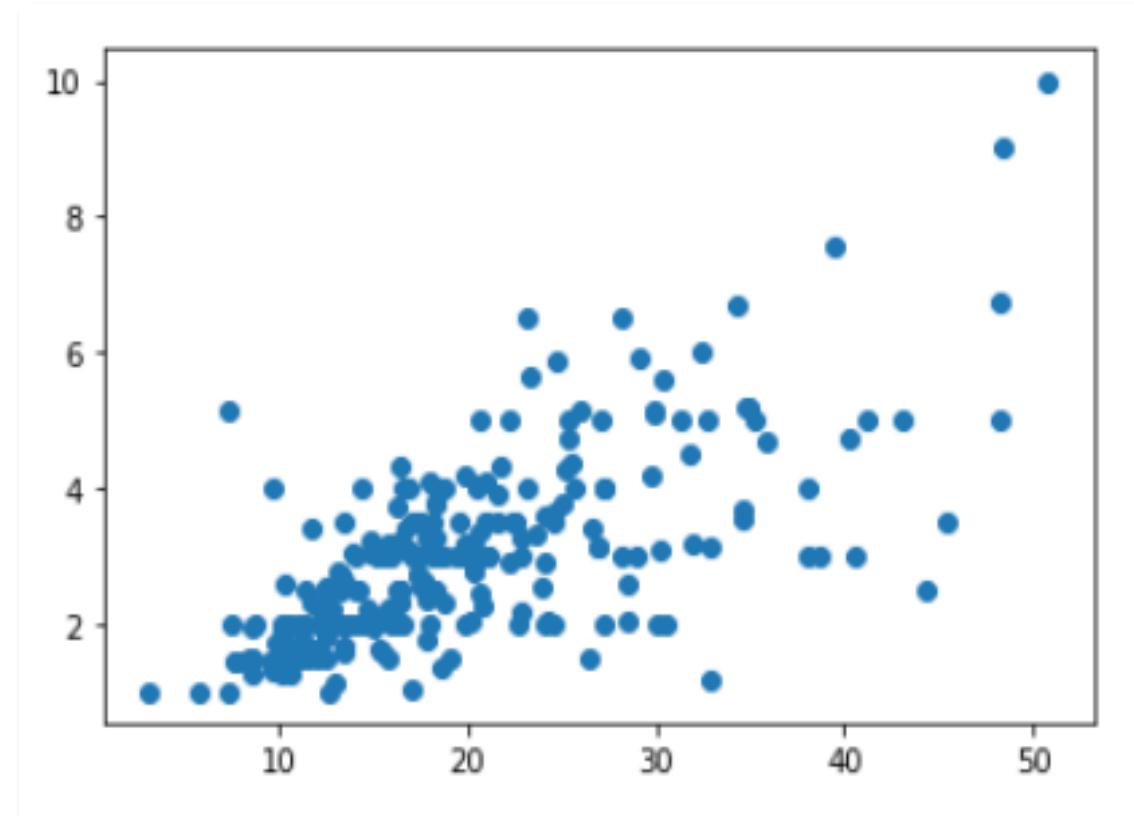
- 막대그래프(Bar plot)는 ‘집단 별 차이를 표현’할 때 사용
- 막대그래프는 데이터 값의 크기에 비례하여 높이를 갖는 직사각형 막대로 표현하며 막대 높이의 상대적 길이 차이를 통해 값의 크기를 시각화 할 수 있음
- 막대 그래프는 x축 데이터, y축 데이터를 지정





산점도

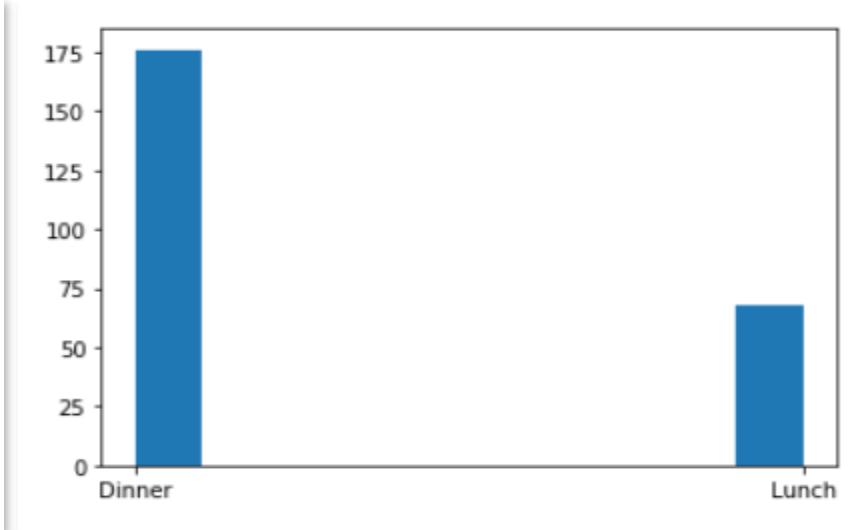
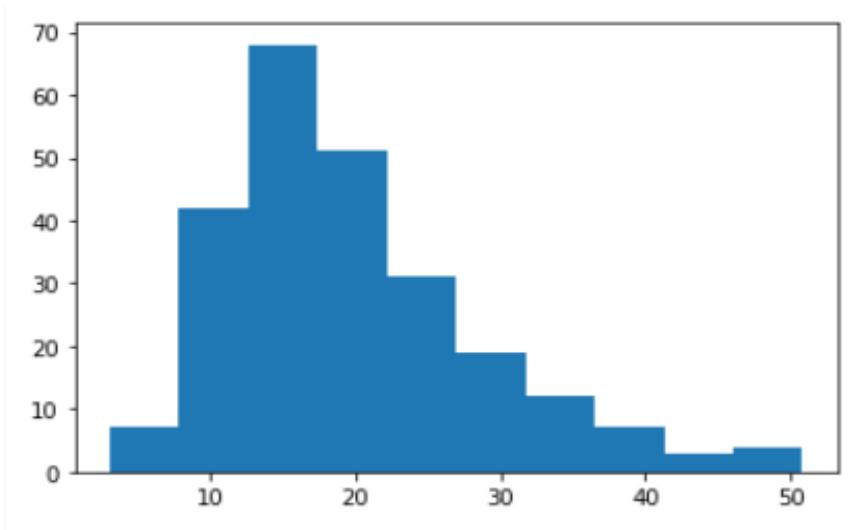
- 산점도(Scatter plot)는 서로 다른 ‘**두 연속형 변수**’ 사이 관계를 나타내며 연속형 변수의 상관성을 확인





히스토그램

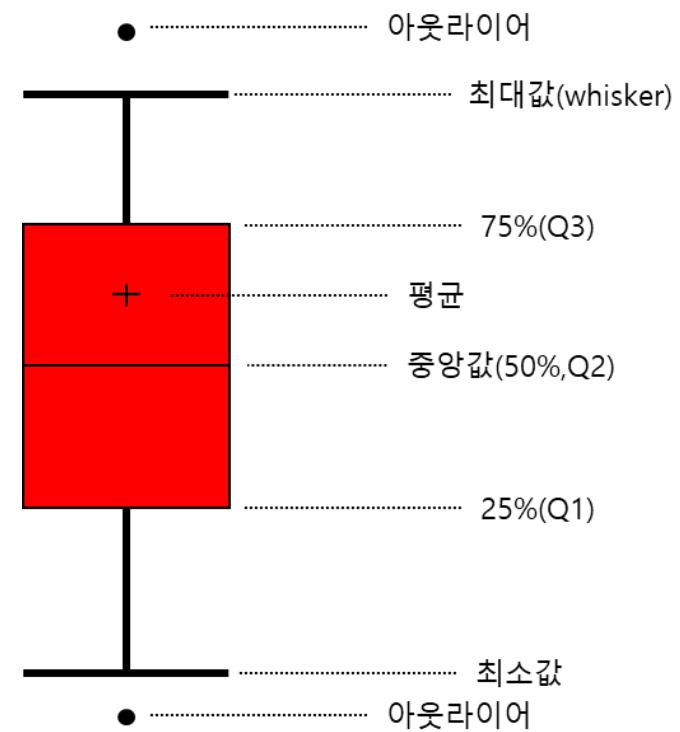
- 히스토그램(Histogram)은 ‘**변수가 하나인 데이터의 빈도수**’를 그래프로 표현
- X축을 등간격으로 나누고 각 구간에 속하는 데이터 값의 빈도를 y축에 표시
- 구간을 나누는 간격의 크기에 따라 빈도와 히스토그램 모양이 변할 수 있음





박스플롯(상자수염그림)

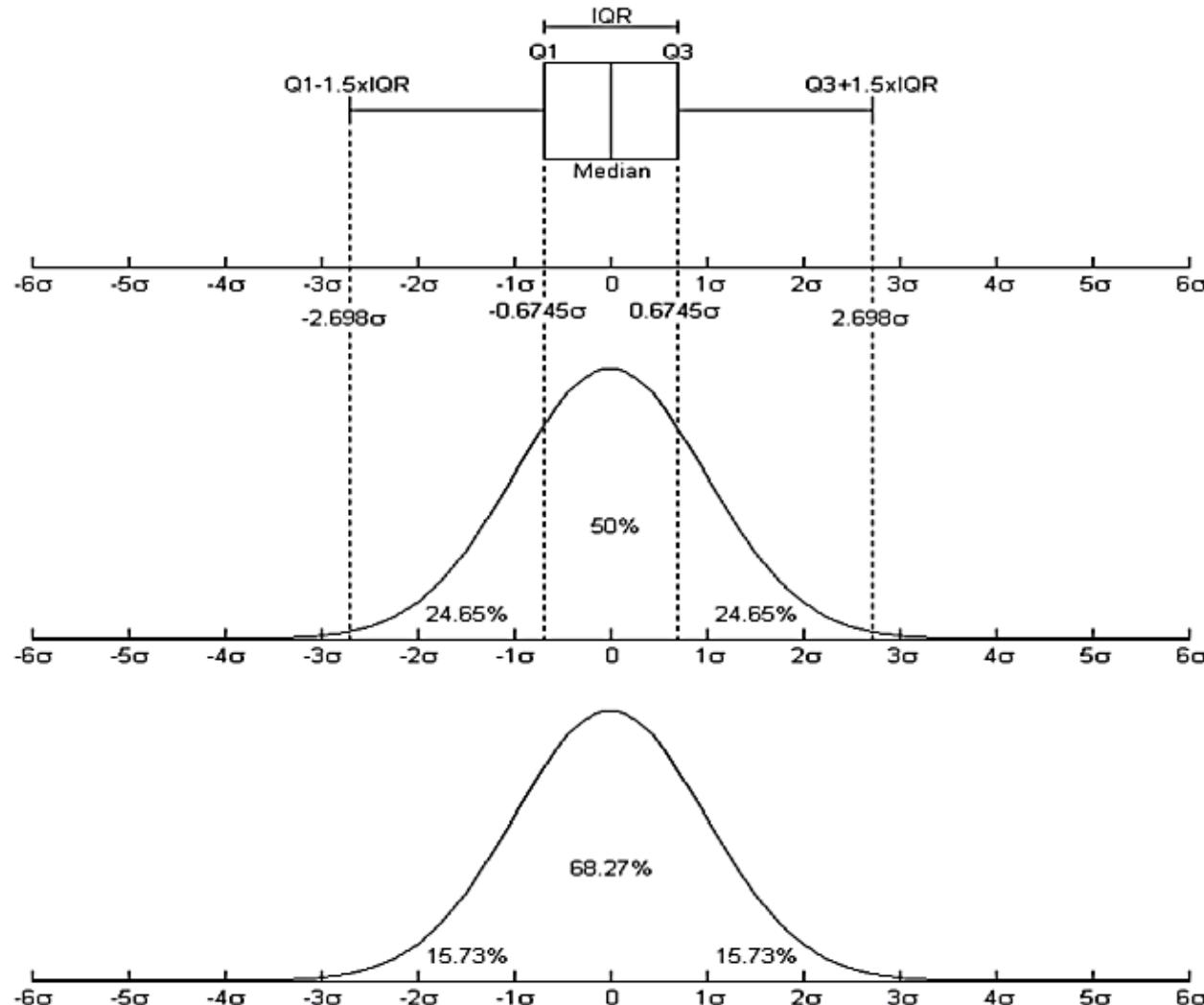
- 박스플롯의 정확한 명칭은 상자 수염 그림(box-and-whisker plot)
- 존 터키(John W. Tukey)가 데이터 분포를 표현하기 위한 시각적 방법, 탐색적 그래프
 - 최소값 : 제 1사분위에서 $1.5 \times \text{IQR}$ (Inter Quartile Range)을 뺀 위치
 - 제 1사분위(Q1) : 25% 위치 값
 - 제 2사분위(Q2) : 50%의 위치로 중앙값(median) 값
 - 제 3사분위(Q3) : 75%의 위치를 의미
 - 최대값 : 제 3사분위에서 $1.5 \times \text{IQR}$ 을 더한 위치
- 박스 플롯을 통한 이상치 제거후 시각화 필요
 - 박스플롯으로 텁에서 연속데이터 만을 살펴보면 최대값을 넘는 이상치들이 존재
 - 이상치 제거 전처리를 수행하고 데이터를 다시 시각화 필요





박스플롯과 정규분포의 연관성

- 박스 플롯과 데이터 분포와의 연관성을 기반으로 이상치를 확인할 수 있음



$$Q^1 = \frac{1}{4} (n - 1)th \text{ value}$$

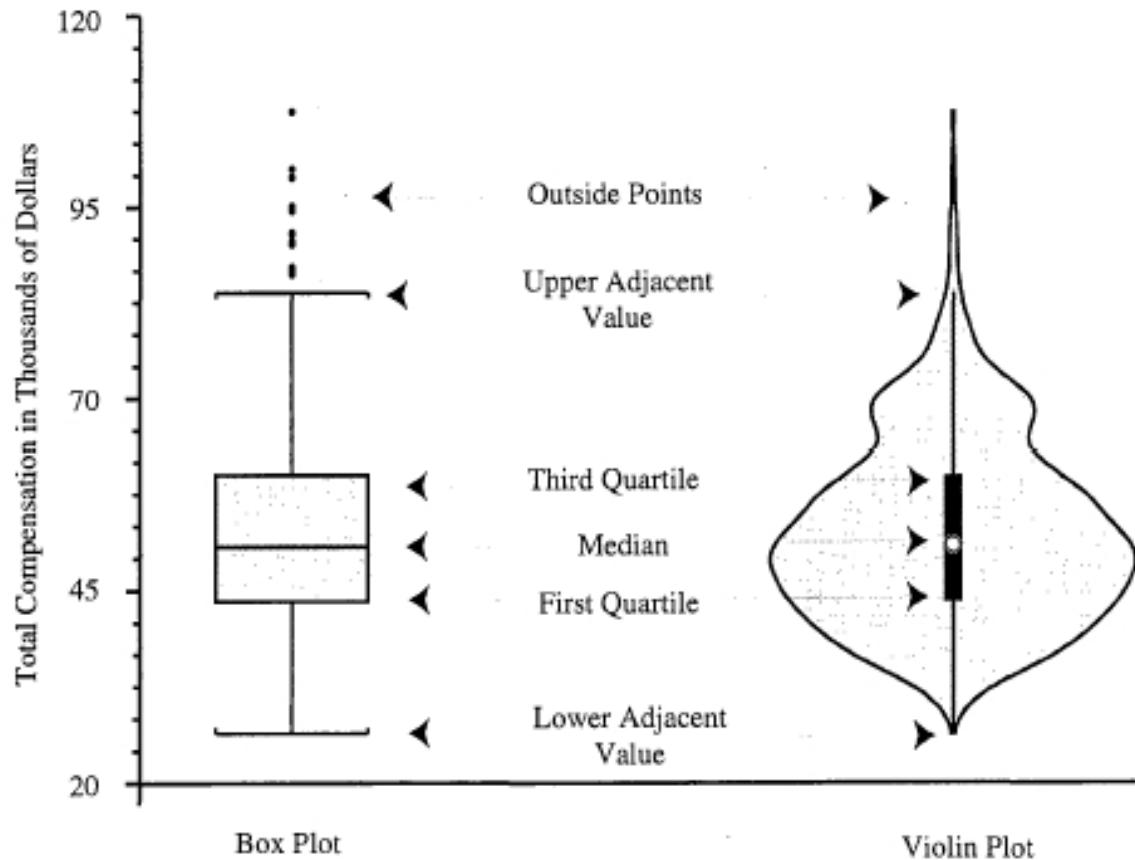
$$Q^2 = \frac{2}{4} (n - 1)th \text{ value}$$

$$Q^3 = \frac{3}{4} (n - 1)th \text{ value}$$



바이올린플롯

- 일변량, 연속형 데이터의 분포를 설명하기 위해 사용되는 그래프
- 카테고리값에 따른 각 분포의 실제 데이터 또는 전체 형상을 보여준다는 장점



빅데이터 시각화 Power BI



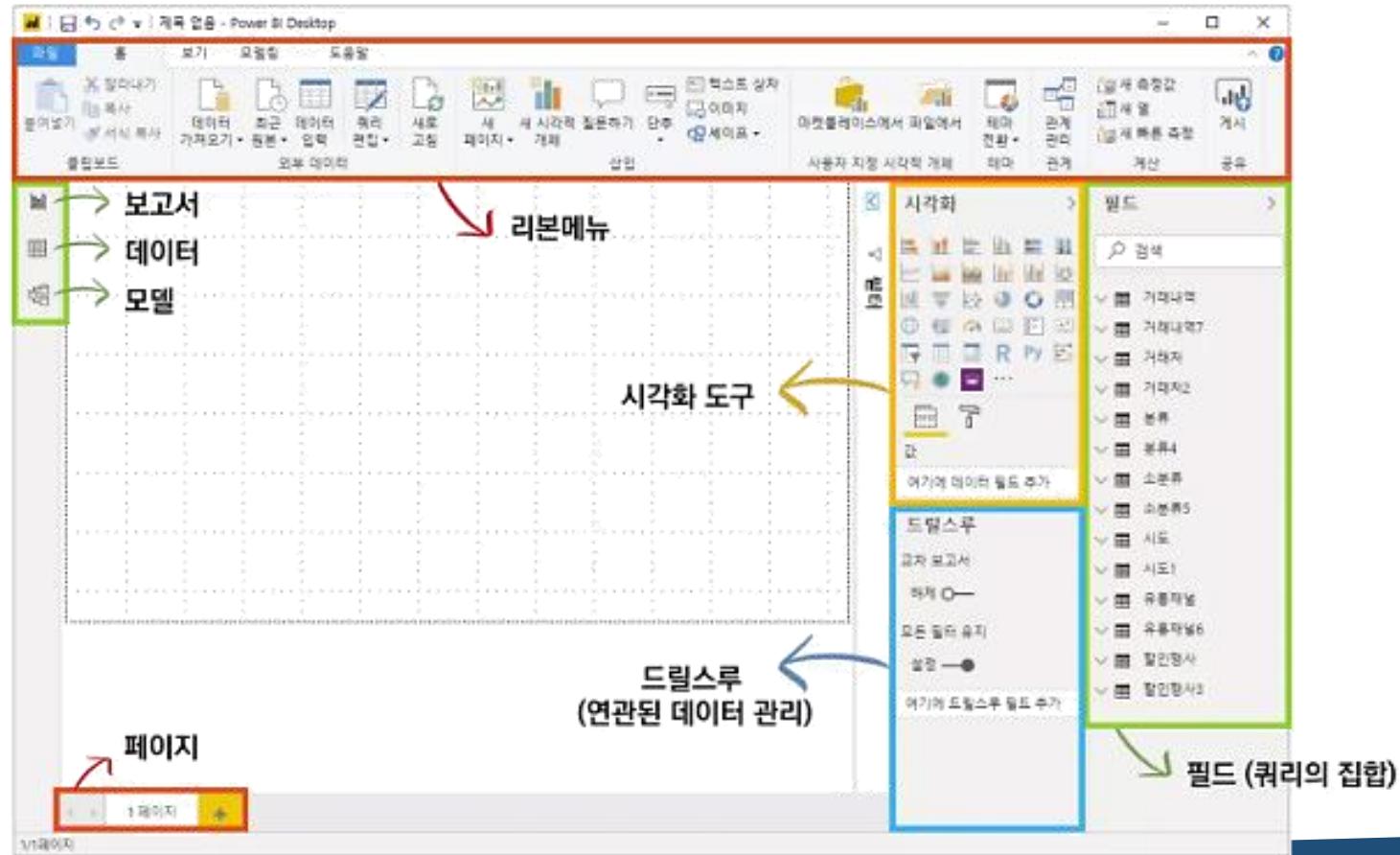


PowerBI (Microsoft社)

- Power BI: 비즈니스를 위한 데이터 분석과 시각화 도구
- 표/테이블 형식의 데이터를 시각화
- 데이터를 대시보드 형태로 시각화해주기 때문에 다양한 데이터를 더욱 쉽고 빠르게 분석



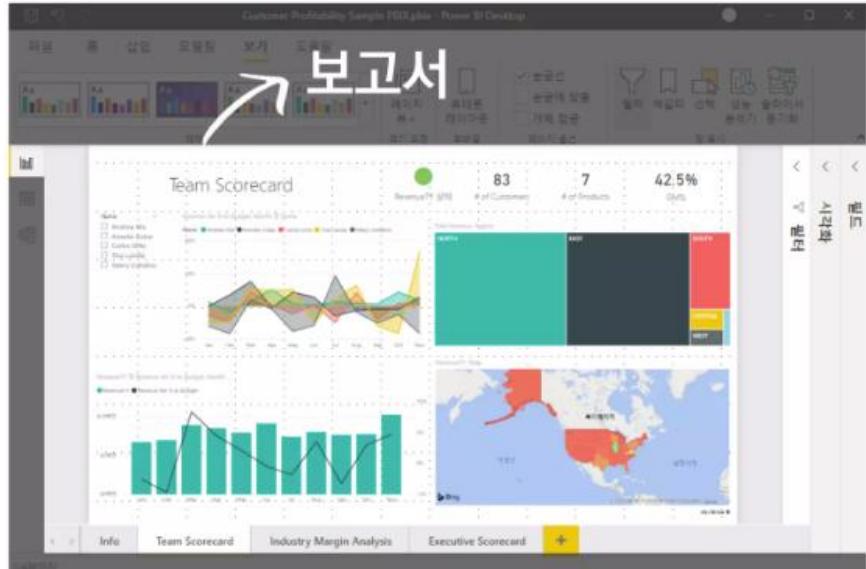
Power BI





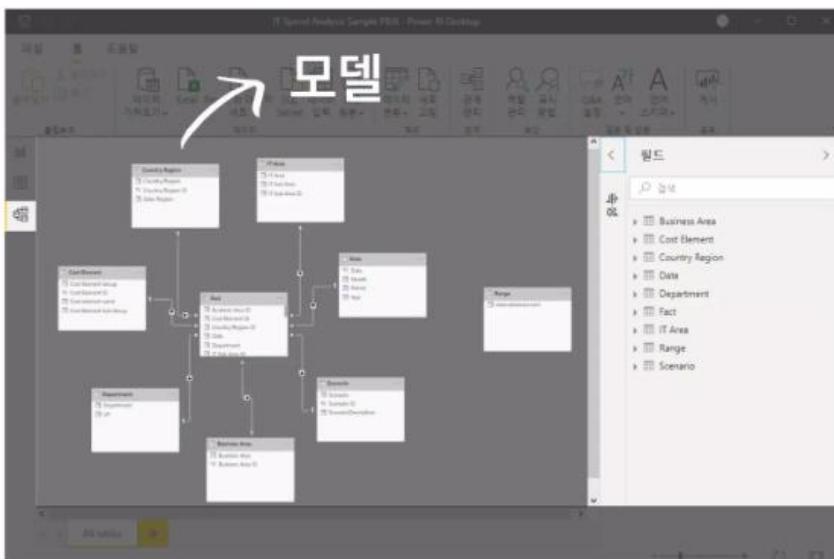
PowerBI

● Power BI의 화면 구성



데이터

This screenshot shows the Power BI Desktop interface with a data view. A large table is displayed with columns for Date, Value, Department, Cost Element ID, Country/Region ID, Business Area ID, IT Sub Area ID, and Scenario ID. The table contains numerous rows of data. To the right of the table is a "필드" (Fields) pane showing a hierarchical list of fields categorized by type, such as Actual, Amount, Business Area ID, and Date.





● 데이터 시각화 준비

- 시각화를 위한 테이블 준비(측정값 계산, 날짜 테이블 사전 생성 등)
- 모델링이 잘 되었는지 확인(연결되어 있지 않은 경우 테이블 관계 설정)





● 보고서 작성 도구

시작화 >

시작화 도구

- ✓ 데이터 시작화에 적합한 다양한 차트 선택

필드

- ✓ 차트에 입력할 데이터 필드를 드래그하여 추가
- ✓ 차트 별 필요한 항목 지정 (x축, y축, 최소값, 최대값 등)

서식

- ✓ 차트 서식 설정
- ✓ 제목, 데이터 색, 테두리 등 세부 설정



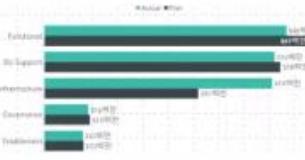
PowerBI

● 기본 차트 종류

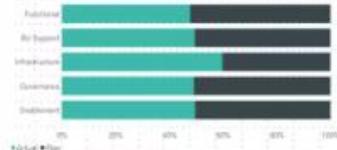
가로 막대형 차트



묶은 가로 막대형 차트



100% 누적 세로 막대형 차트



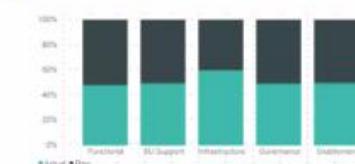
누적 세로 막대형 차트



묶은 세로 막대형 차트



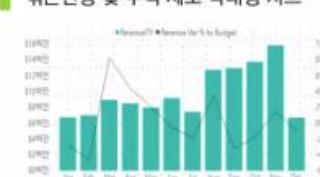
100% 누적 세로 막대형 차트



꺾은선형 차트



꺾은선형 및 누적 세로 막대형 차트

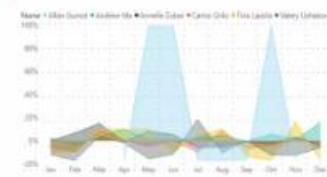


꺾은선형 및 묶은 세로 막대형 차트

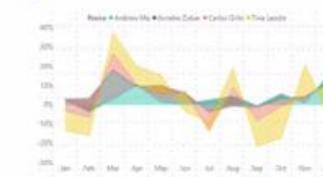


▶ 영역형 차트

영역형 차트



누적 영역형 차트

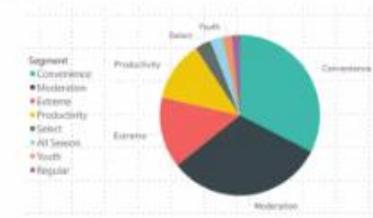




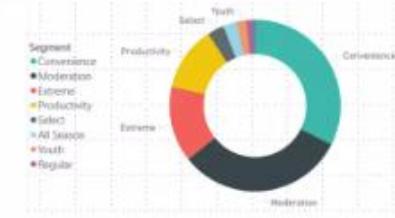
PowerBI

● 기타 차트 종류

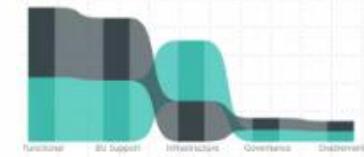
원형 차트



도넛형 차트



리본 차트



폭포 차트



▶ 카드형 차트

카드



여러 행 카드



Tree Map



맵





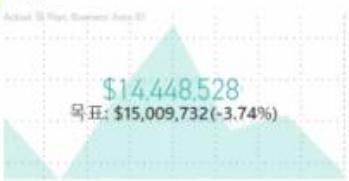
PowerBI

● 추가 차트 종류 및 기타 도구

등치 지역도



KPI



행렬

	Actual	Actual/Plan	목표	Actual/Plan	상태
Labor	\$136,644,019	\$139,850,862	●		
Degr & Amort	\$59,487,299	\$65,494,401	●		
Hardware & Software	\$48,779,413	\$49,330,396	●		
Other	\$14,547,130	\$14,397,095	●		
Shared Services	\$437,086	\$1,047,990	●		
CAPEX	(\$2,369,253)	(\$1,621,811)	●		
Administrative			(\$25,212,487)		
합계	\$257,325,697	\$243,285,467	●		

계기



테이블

Cost Element Group	Cost element name	VR	Department/Actual
Degr & Amort	Software	Ami Beck	\$12,14,101
Hardware & Software	Software Maintenance	Linda Lee	\$9,06,303
Degr & Amort	Software	Ann Swanson	\$1,00,000
Degr & Amort	Software	Phil Clark	\$30,16,771
Other	Other Miscellaneous Expenses	Ann Sigg	400
Labor	Consulting	Nora Jones	\$12,33,040
Degr & Amort	Software	Tom Brown	\$13,18,121
Labor	Professional Services	Linda Lee	\$3,76,996
Degr & Amort	Software	Linda Lee	\$12,14,091
Labor	Professional Services	Ann Sigg	\$10,048,833
Labor	Outsourcing	Phil Clark	\$11,18,500
합계			\$217,626,407

Power BI 시각적 개체

사용자 지정 시각적 개체를 생성하고, 공유 할 수 있도록 하는 도구로, 다른 이용자가 제작하여 공유한 다양한 시각적 개체 활용 가능
[홈] > 삽입 > 다른 시각적 개체 > AppSource에서

Power BI 시각적 개체

마켓플레이스 > 내 조직

추가 기능이 개인 및 문서 환경에 액세스할 수 있습니다. 추가 기능을 사용하는 것은 개인 사용 권한, 사용 조건 및 개인 정보 보호 등과 함께 동의하는 것입니다.

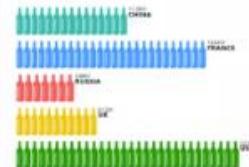
검색: 주천 항목:

범주: 주천 항목:

- Power KPI** A powerful KPI indicator with multi-line chart and labels for current date, value and variances. ★★★★☆
- Enlighten Aquarium** Make dashboards fun with this award-winning Aquarium visualization. ★★★★☆
- Radar Chart** Multiple measures plotted over a categorical axis. Useful to compare attributes. ★★★★☆
- Sankey Chart** Flow diagram where the width of the series is proportional to the quantity of the flow. ★★★★☆



different sites
famous content
media digital
content friend
networking like
information people
share



슬라이서

원하는 정보만 차트에서 볼 수 있도록 차트 내용을 슬라이싱(필터링) 하는 도구

Sales Region

- Africa & Asia
- Aus and NZ
- Canada
- Europe
- Latin America
- USA

Month

- Jan
- Feb
- Mar
- Apr
- May
- Jun
- Jul

고용일자

2000-11-06 2006-02-04

질문 및 답변

해당 데이터와 관련하여 질문 자동 생성 및 답변 제시하는 도구



● DAX 함수

- DAX(Data Analysis Expressions) 식에서 하나 이상의 값을 계산하고 반환하는데 사용할 수 있는 함수의 모음
- DAX는 모델에 있는 데이터를 활용하여 새로운 정보를 생성할 수 있음
- 시각화 과정에서 계산이 필요할 때 사용

▶ Dax 함수 사용 방법

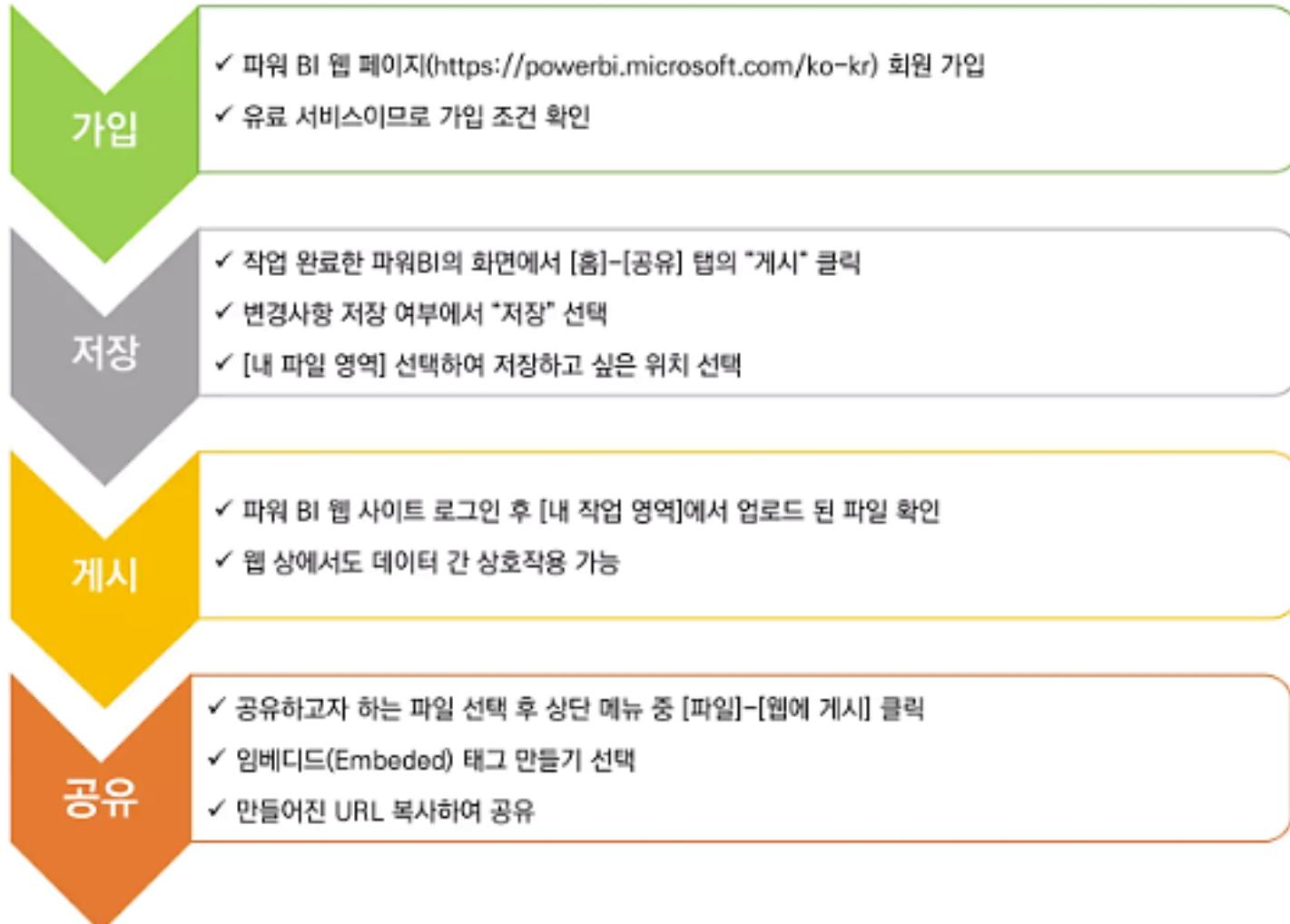
The screenshot shows the Power BI Desktop interface. In the ribbon, the 'Data' tab is selected. On the far left, under 'New Calculated Column', there is a red box around the 'New Column' button. Below the ribbon, a data grid displays several rows of sales data. A red box highlights the 'New Column' button in the ribbon. To the right of the data grid, a 'Fields' pane is open, showing a list of columns: '판매ID', '날짜', '이벤트번호', '제품코드', '거래처코드', '제품코드', '원가', '단가', '수량', '매출금액', and '매출점가'. The '거래처코드' column is currently selected, indicated by a yellow box.

- 데이터 탭 → 새 열 OR 열 삽입 원하는 필드에서 마우스 오른쪽 버튼 → 새 열
- 열 이름 지정 후 데이터 계산 함수 작성
예시) 매출 합계=CALCULATE(SUM("거래내역"[매출금액]))
- 참고) 총 합계, 총 개수, 총 분산, 평균값 등과 같은 요약 그룹화는 '새 측정값'에서 함수 작성하는 것이 편리



PowerBI

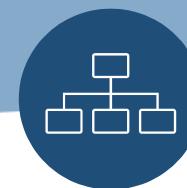
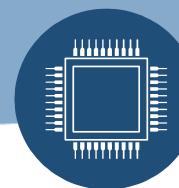
● 게시 및 공유 방법





Thank you!

See you next time.



담당교수 : 유 현 주
comjoo@uok.ac.kr

