

# 생성형AI

## Day 7

### 데이터 시각화 I



## 목차

---

1. 데이터 시각화란?
2. 시각적 인지의 기초
3. 시각화의 디자인 원칙
4. 기본 차트 유형
5. 실습과제



# 데이터 시각화란?

- 데이터를 그래픽적 요소로 변환하여 시각적으로 표현하는 과정
- 복잡한 데이터를 이해하기 쉽게 만드는 방법
- 데이터 시각화의 구성요소
- 기본요소: 점, 선, 막대, 색상, 크기, 원 등
- 시각화 도구: Excel, Tableau, Power BI, Python(matplotlib, seaborn)



# 데이터 시각화란 – 시각화 도구 선택 가이드

## Excel

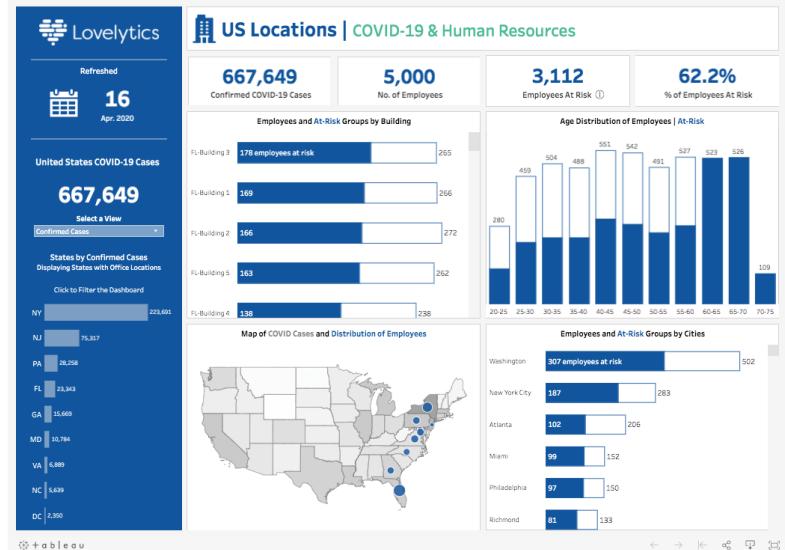
- Microsoft의 스프레드시트 프로그램으로, 데이터 관리와 기본적인 시각화 기능을 제공
- 대중적으로 사용되고 있어 접근성과 사용이 용이
- 장점
  - 사용 용이성: 직관적인 인터페이스, 드래그 앤 드롭으로 차트 생성 가능
  - 광범위한 사용: 대부분의 비즈니스 환경에서 이미 사용 중
  - 기본 통계 기능: 기본적인 통계 분석 및 피벗 테이블 기능 포함
- 단점
  - 복잡한 시각화 제한: 고급 시각화 및 대규모 데이터 처리에 한계
  - 확장성 부족: 대용량 데이터 처리 및 실시간 데이터 시각화에 제약
- 선택 이유
  - 사용이 간편하고, 대부분의 비즈니스 사용자들이 이미 익숙
  - 빠른 데이터 시각화가 필요한 경우



# 데이터 시각화란 – 시각화 도구 선택 가이드

## Tableau

- 강력한 데이터 시각화 도구로, 다양한 데이터 소스와의 통합 및 인터랙티브 시각화 기능 제공
- 다양한 데이터 소스를 결합한 시각화와 실시간 대시보드, 보고서 작성 등에 사용
- 장점
  - 인터랙티브 시각화: 대화형 대시보드와 실시간 데이터 업데이트 가능
  - 사용자 친화적: 드래그 앤 드롭 인터페이스로 쉽게 시각화 가능
  - 다양한 데이터 소스 통합: 여러 데이터베이스와 파일 포맷을 연결 가능
- 단점
  - 비용: 라이선스 비용이 높은 편
  - 복잡한 기능 학습 필요: 고급 기능 사용을 위해 추가 학습 필요
- 선택 이유
  - 복잡한 데이터 시각화 및 실시간 데이터 모니터링 필요 시
  - 대화형 시각화 및 사용자 정의 대시보드 작성 필요 시



## 데이터 시각화란 – 시각화 도구 선택 가이드

### Power BI

- Microsoft의 비즈니스 분석 서비스로, 데이터 분석 및 시각화 기능을 제공
- 기업의 KPI 모니터링 및 보고서 작성, 실시간 데이터의 시각화 등에 사용
- 장점
  - Microsoft 생태계 통합: Excel, Azure 등과 원활한 통합
  - 사용자 친화적: 친숙한 인터페이스와 쉬운 학습 곡선
  - 실시간 데이터: 실시간 데이터 피드와 자동 업데이트 기능
- 단점
  - 제한된 고급 분석 기능: 일부 고급 시각화 기능이 제한적
  - 데이터 처리 한계: 대규모 데이터 처리 시 성능 저하 가능
- 선택 이유
  - Microsoft 환경과의 통합 및 실시간 비즈니스 분석 필요 시
  - 쉬운 데이터 시각화와 빠른 배포 필요 시

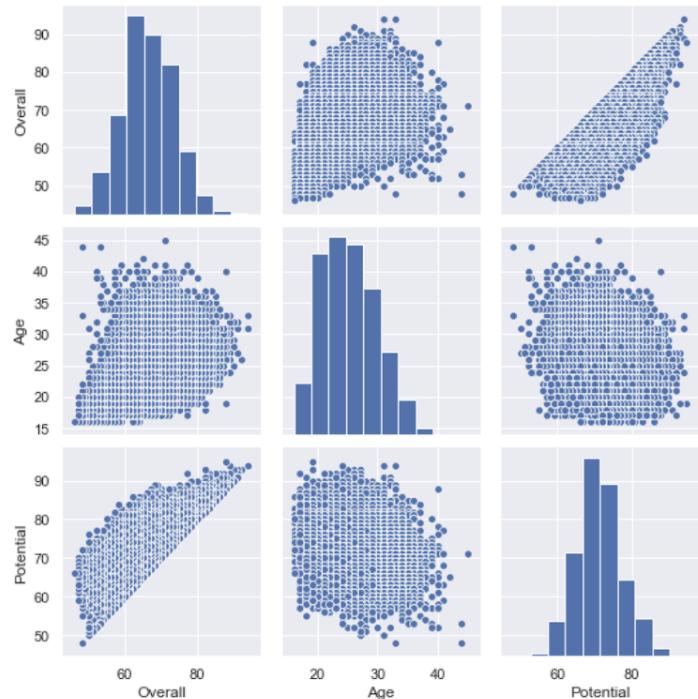


<https://learn.microsoft.com/ko-kr/power-bi/create-reports/service-dashboards>

## 데이터 시각화란 – 시각화 도구 선택 가이드

### Python (matplotlib, seaborn)

- 프로그래밍 언어인 Python을 사용한 데이터 시각화 라이브러리
- matplotlib은 기본적인 시각화 도구, seaborn은 통계적 시각화 강화
- 복잡한 데이터셋, 머신러닝 모델 결과의 시각화
- 장점
  - 고급 시각화: 복잡한 데이터 시각화 및 커스터마이징 가능
  - 오픈 소스: 무료로 사용 가능, 커뮤니티 지원
  - 통합 분석: 데이터 분석, 처리 및 시각화를 하나의 환경에서 수행
- 단점
  - 코딩 필요: 프로그래밍 지식 필요
  - 학습 곡선: 초보자에게는 학습이 다소 어려울 수 있음
- 선택 이유
  - 복잡한 시각화와 데이터 분석을 통합하여 수행할 필요 시
  - 커스터마이징이 필요한 고급 시각화 필요 시



<https://builtin.com/data-science/data-visualization-tutorial>

# 데이터 시각화란

## 시각화의 목적

- 데이터 패턴 인식: 데이터의 패턴과 트렌드를 쉽게 식별
- 통찰력 제공: 데이터에서 의미 있는 인사이트 도출
- 의사결정 지원: 데이터 기반의 전략적 의사결정에 기여
- 정보 전달: 복잡한 정보를 명확하고 간결하게 전달



표 2. 세계 교역 및 가격 증가율  
(단위: 전년대비 %)

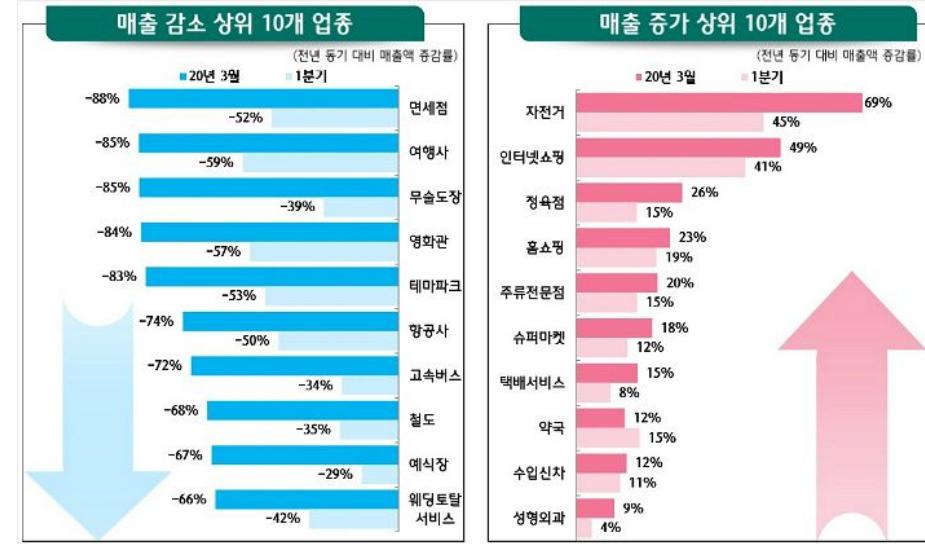
구분	2017	2018	2019	2020	
상품+서비스	교역량(volume)	5.6	3.9	0.9	<b>-8.5</b>
	교역액(value)	10.2	9.6	-1.7	<b>-10.4</b>
상품	교역량	5.6	3.8	0.3	<b>-5.1</b>
	교역액	10.8	9.5	-2.8	<b>-7.4</b>
가격	전체(상품+서비스)	4.3	5.5	-2.6	<b>-2.2</b>
	전체(상품)	4.9	5.8	-3.2	<b>-2.6</b>
제조품	0.1	2.0	0.4	<b>-3.1</b>	
석유(oil)	23.3	29.4	-10.2	<b>-32.7</b>	
기타 1차 제품	6.4	1.3	0.8	<b>6.7</b>	

주: 교역은 '상품+서비스 교역'을 의미.

자료: IMF WEO(2021. 4).

자료: IMF WEO(2021. 4).

KIEP 2021년 세계경제 전망



## 데이터 시각화란

## 시각화의 중요성

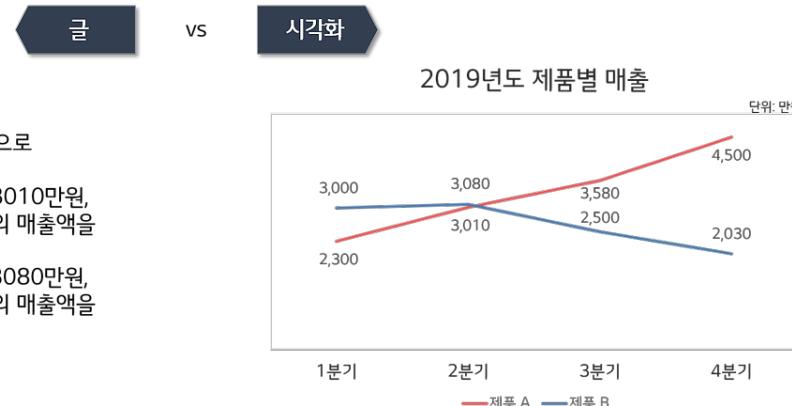
- 복잡한 데이터의 이해도 향상: 복잡한 데이터 세트를 직관적으로 이해
  - 효율적인 커뮤니케이션: 시각적 자료를 통해 정보 전달의 효율성 증대
  - 데이터 기반 의사결정: 데이터에 기반한 신뢰성 높은 결정
  - 심리적 효과: 사람들은 시각적 정보를 더 빠르고 정확하게 처리함



## 데이터 시각화란

### 시각화를 통한 데이터 이해의 예시

- 텍스트 데이터 VS 시각화 데이터
- 사례: 제품들의 분기별 매출 데이터
- 차트 종류: 선 그래프
- 분석 인사이트: 제품별 특정 분기의 매출 증/감 원인 분석



<https://blog.hsad.co.kr/2982>

# 데이터 시각화란

# 시각화의 역사

- 초기 시각화(지도 제작)
  - 항해, 탐험, 영토 관리 등 다양한 목적을 위해 지도를 제작
  - 지리적 데이터의 시각화를 통해 정보 전달과 의사결정을 지원

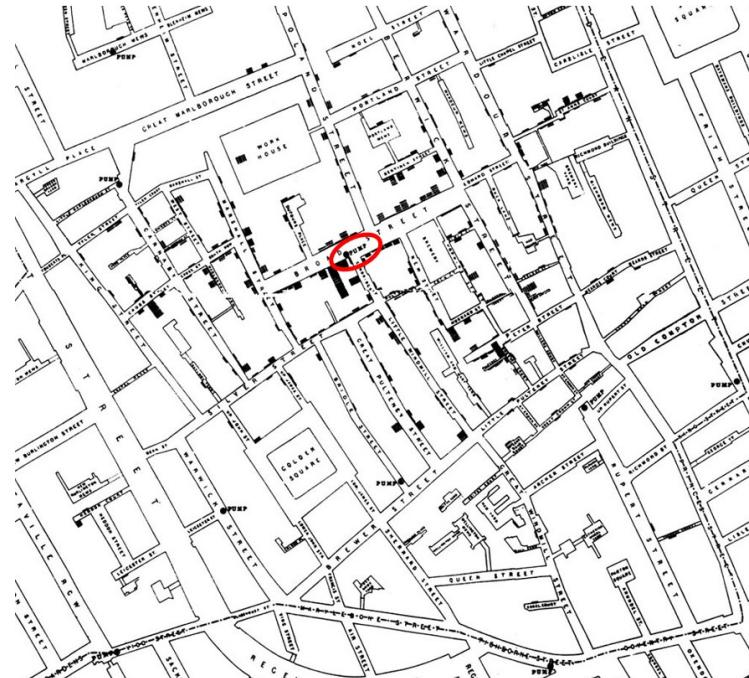


## Ptolemy의 세계지도

## 데이터 시각화란

### 시각화의 역사

- John Snow의 1854년 콜레라 지도, 런던 콜레라 발생 지역 시각화
- 1854년 런던 콜레라 유행 시기
- 사망자의 위치를 지도에 표시하여 감염 경로 추적
- 데이터 시각화를 통한 역학 조사와 공중 보건 개선
- 오염된 물 펌프가 콜레라 발병의 원인임을 발견

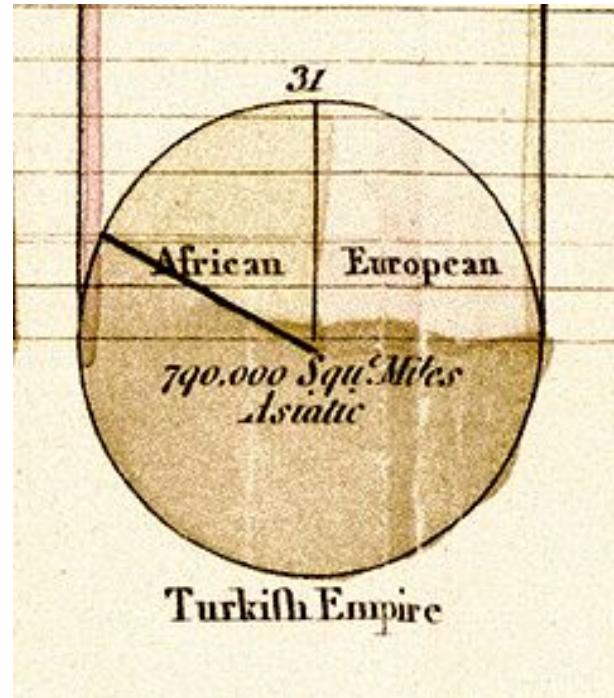
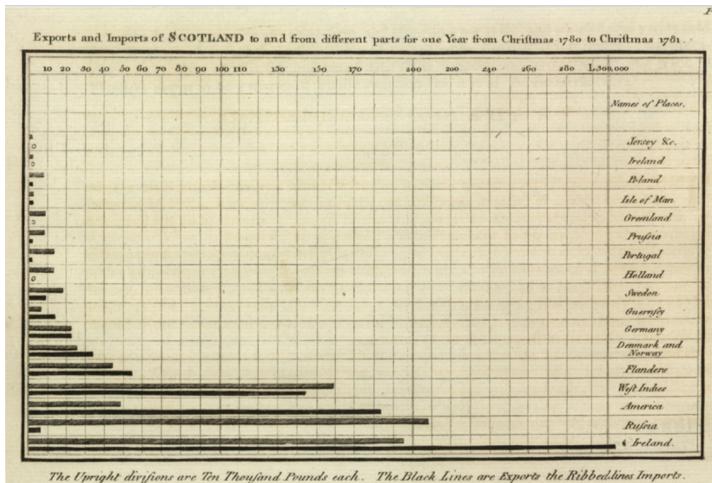


John Snow의 1854년 콜레라 지도

## 데이터 시각화란

### 시각화의 역사

- 18세기 후반, 통계 시각화의 발전
- William Playfair (현대적 차트의 창시자)
- 막대 그래프, 선 그래프, 원 그래프 등 개발
- 경제 데이터의 시각화를 통해 경제적 패턴과 트렌드를 이해

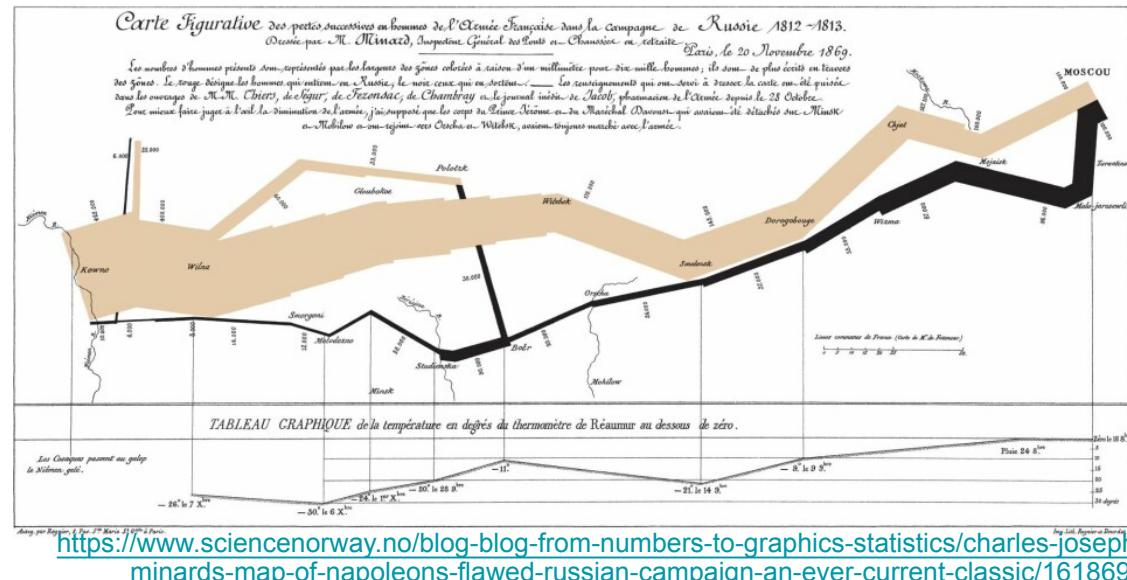


1801년 터키제국의 비율

# 데이터 시각화란

## 시각화의 역사

- 20세기 초반
- Charles Joseph Minard, *Carte Figurative des pertes successives en hommes de l'Armée Française dans la Campagne de Russie 1812 ~1813.*
- 나폴레옹의 러시아 원정을 시각화한 지도
- 병력의 수, 이동 경로, 온도 변화를 한 지도에 통합



# 시각적 인지의 기초

## 시각적 인지 정의

- 인간이 시각적 자극을 통해 정보를 처리하고 해석하는 과정
- 시각적 인지는 빠르고 효율적인 정보 처리를 가능하게 함

## 시각적 인지와 데이터 시각화의 관계

- 효율적 정보 전달: 시각적 인지를 활용하여 데이터를 효과적으로 전달
- 의사결정 지원: 시각적 인지를 통해 데이터 기반 의사결정의 신뢰성 향상

## 인간의 시각적 인지 능력

- 패턴 인식
  - 인간의 뇌는 패턴을 인식하는 데 뛰어남
  - 시각적 데이터를 통해 빠르게 패턴과 트렌드를 식별
- 정보 우선순위
  - 시각적 요소를 통해 중요한 정보를 강조
  - 크기, 색상, 위치 등의 요소를 활용하여 주의를 끌 수 있음



주: 교역은 '상품+서비스 교역'을 의미.  
자료: IMF WEO(2021. 4).

표 2. 세계 교역 및 가격 증가율

(단위: 전년대비 %)

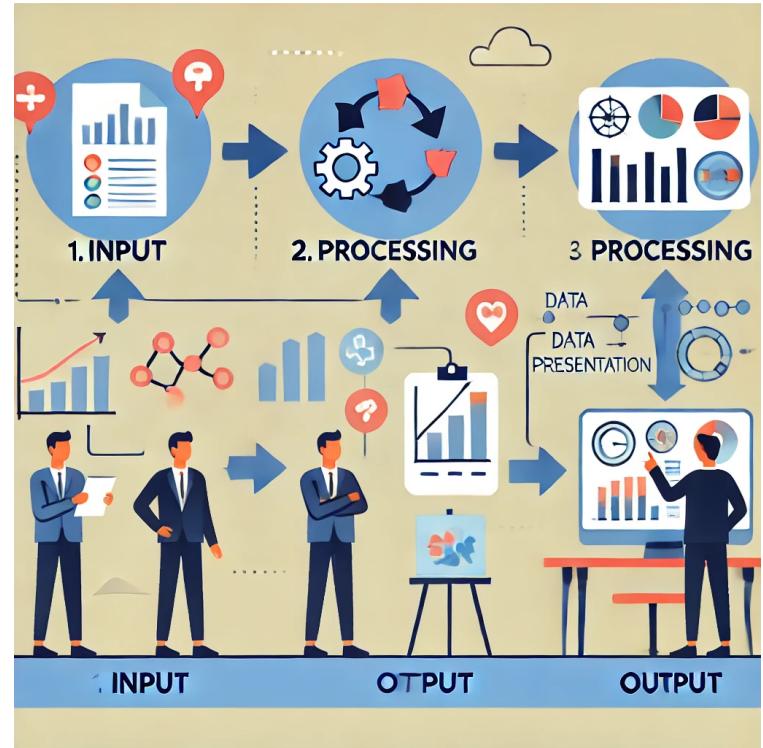
구분	2017	2018	2019	2020	
상품+서비스	교역량(volume)	5.6	3.9	0.9	-8.5
	교역액(value)	10.2	9.6	-1.7	-10.4
상품	교역량	5.6	3.8	0.3	-5.1
	교역액	10.8	9.5	-2.8	-7.4
가격	전체(상품+서비스)	4.3	5.5	-2.6	-2.2
	전체(상품)	4.9	5.8	-3.2	-2.6
제조품	제조품	0.1	2.0	0.4	-3.1
	석유(oil)	23.3	29.4	-10.2	-32.7
기타 1차 제품	6.4	1.3	0.8	6.7	

자료: IMF WEO(2021. 4).

# 시각적 인지의 기초

## 정보처리과정에서 시각적 요소의 역할

- 초기 시각 처리 단계
  - 색상, 형태, 크기 등 기본 시각적 요소를 빠르게 인식
  - 기본적인 시각적 특징을 통해 정보의 우선순위를 정함
- 중간 시각 처리 단계
  - 패턴과 그룹화 인식
  - 데이터의 연관성과 관계를 파악
  - 시각적 요소의 통합
- 통합 과정
  - 다양한 시각적 요소를 통합하여 전체적인 의미를 파악
  - 시각적 데이터를 종합적으로 해석하여 인사이트 도출



# 시각적 인지의 기초

---

## 색상의 영향력

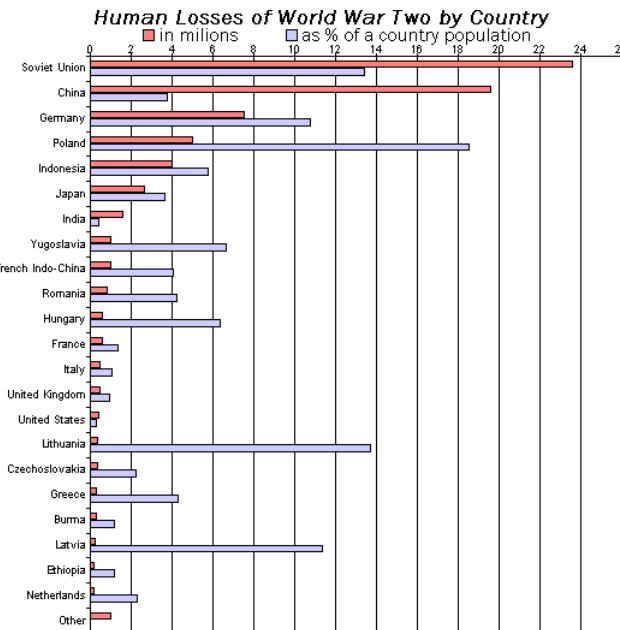
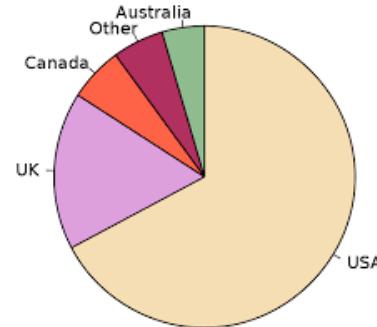
- 심리적 반응
  - 색상은 감정적 반응을 유발
  - 빨간색: 경고, 위험 / 녹색: 안전, 긍정
- 정보 전달
  - 색상을 통해 데이터의 중요한 부분을 강조
  - 예시: 차트에서 중요한 데이터 포인트를 강조하는 색상 사용



# 시각적 인지의 기초

## 형태의 영향력

- 데이터 구분과 그룹화
  - 다양한 형태를 사용하여 데이터를 구분하고 그룹화
  - 형태는 시각적으로 데이터를 구분할 수 있는 강력한 도구
- 형태를 통한 의미 전달
  - 원, 사각형, 삼각형 등의 기본 도형을 사용하여 데이터를 시각적으로 표현 가능
  - 각 도형의 특성을 통해 데이터를 직관적으로 이해 가능

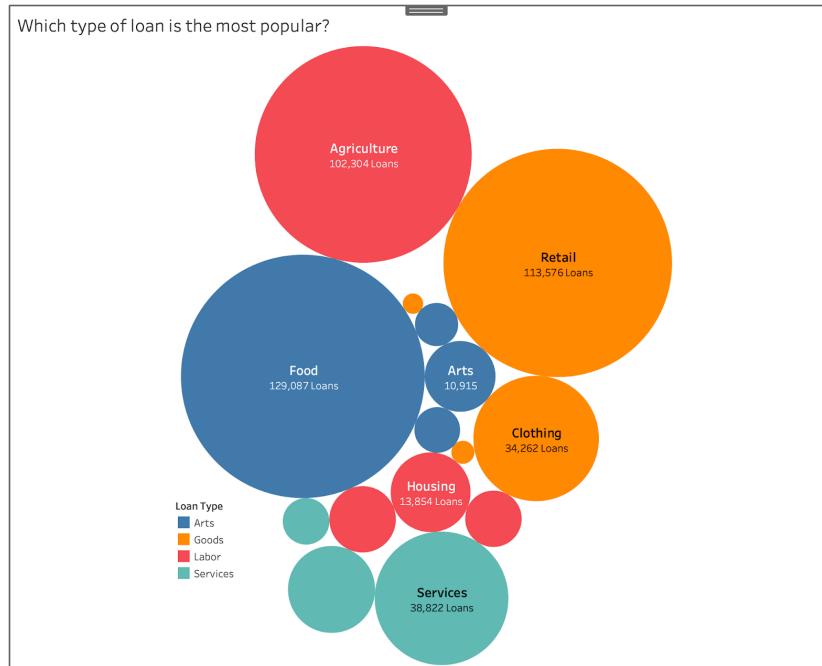


[https://ko.wikipedia.org/wiki/막대\\_그래프](https://ko.wikipedia.org/wiki/막대_그래프)

## 시각적 인지의 기초

### 크기의 중요성

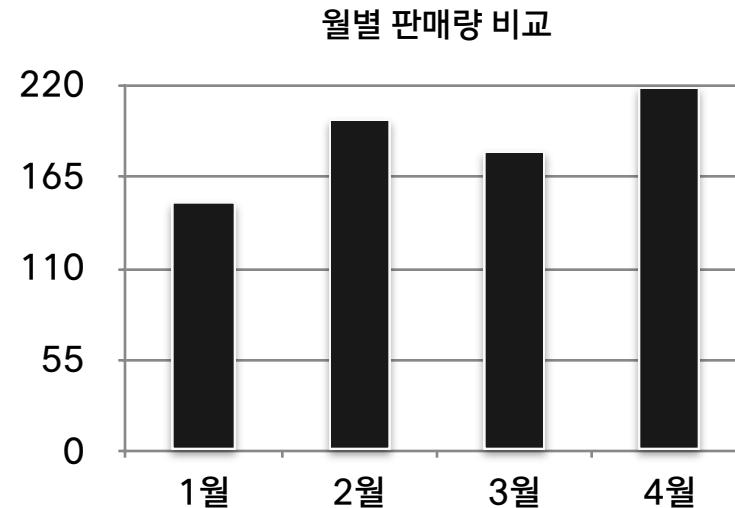
- 크기를 통해, 크기 차이의 비율을 통해 데이터의 상대적 중요성 전달 가능
- 큰 요소는 중요한 데이터, 작은 요소는 덜 중요한 데이터



## 기본 차트 유형 및 사용법

### 막대 그래프 (Bar Chart)

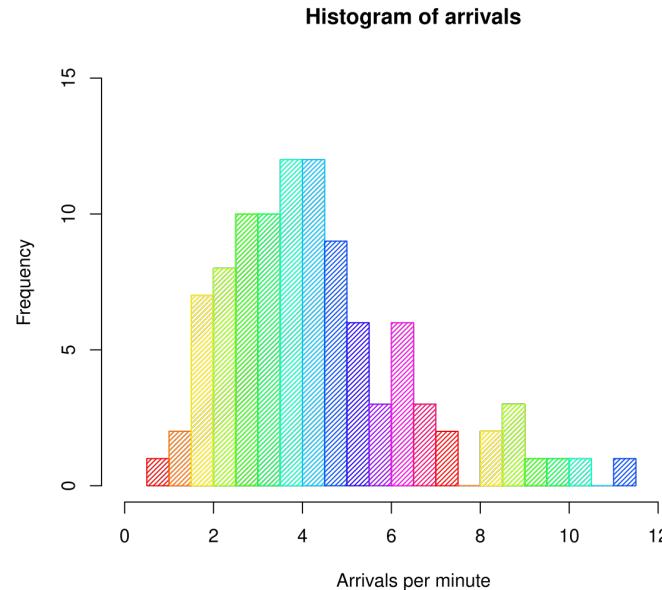
- 막대 그래프는 ‘범주형 데이터의 빈도나 값’을 나타내기 위해 사용되는 기본적인 시각화 도구
- 데이터 값은 수평 또는 수직 막대로 표현되며, 각 막대의 길이는 해당 범주의 크기나 빈도를 나타냄
- 예시
  - 비즈니스: 월별 매출 비교를 통해 성과 분석
  - 교육: 학년별 성적 비교
  - 의료: 병원별 환자 수 비교



# 기본 차트 유형 및 사용법

## 히스토그램 (Histogram)

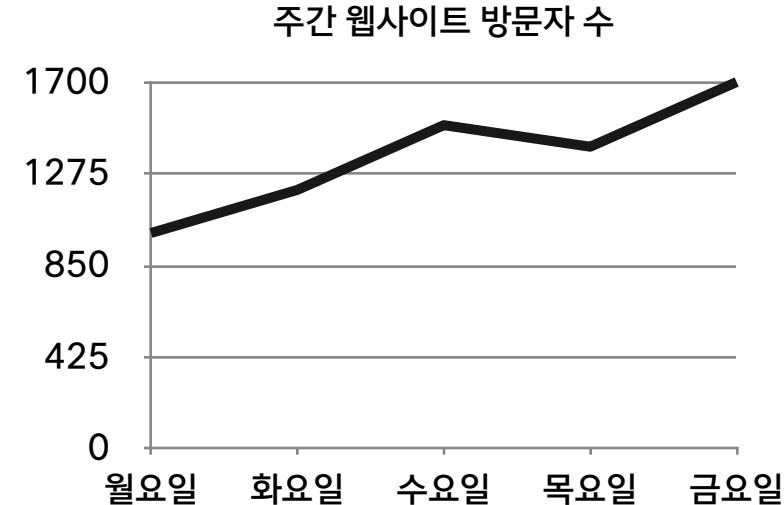
- 히스토그램은 ‘연속형 데이터의 분포’를 나타내는 데 사용
- 데이터 범위를 여러 구간(bin)으로 나누고 각 구간의 데이터 빈도를 막대로 표시
- 예시
  - 비즈니스: 제품 가격대별 판매량 분포
  - 교육: 시험 점수 분포
  - 건강: 나이대별 환자 수 분포



## 기본 차트 유형 및 사용법

### 선그래프 (Line Chart)

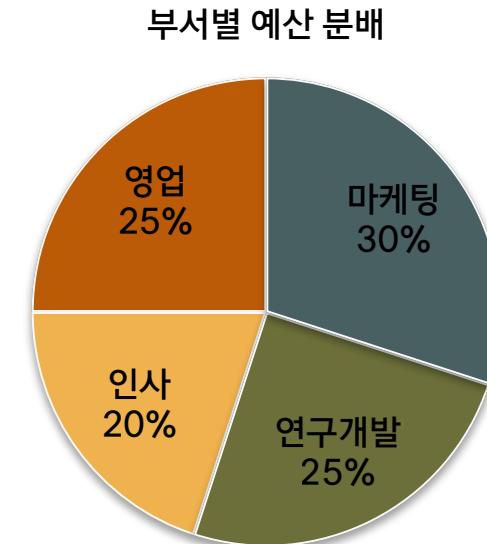
- 선 그래프는 시간에 따른 데이터 변화를 시각화하는 데 사용
- 데이터 포인트가 선으로 연결되어 시간 경과에 따른 추세를 나타냄
- 예상 시나리오
  - 비즈니스: 주간 매출 트렌드 분석
  - 교육: 학기별 학생 출석률 변화
  - 건강: 일별 혈압 수치 변화



## 기본 차트 유형 및 사용법

### 파이차트 (Pie Chart)

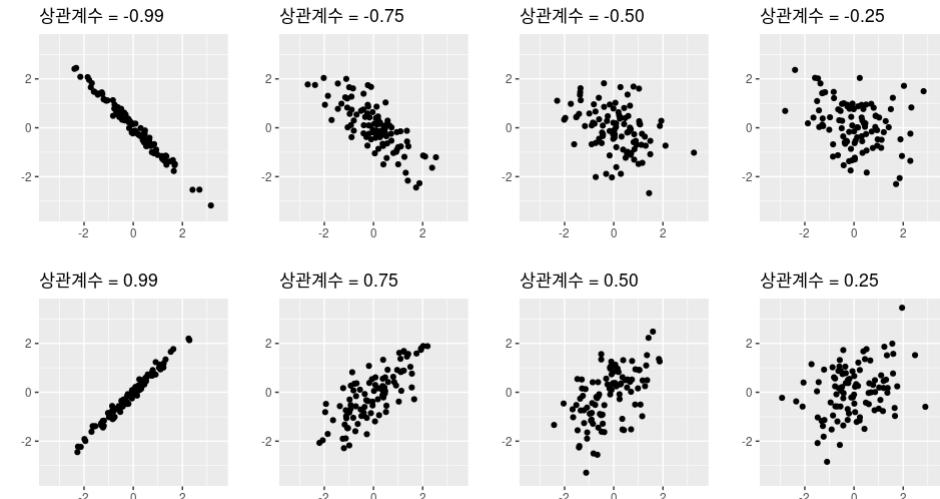
- 파이 차트는 전체에서 각 부분의 비율을 시각화하는 데 사용
- 원형 차트로, 각 조각의 크기는 전체에서 해당 항목의 비율을 나타냄
- 예상 시나리오
  - 비즈니스: 예산 분배, 시장 점유율 비교
  - 교육: 학과별 학생 비율
  - 의료: 병원 내 각 부서별 환자 비율



## 기본 차트 유형 및 사용법

### 산점도 (Scatter Plot)

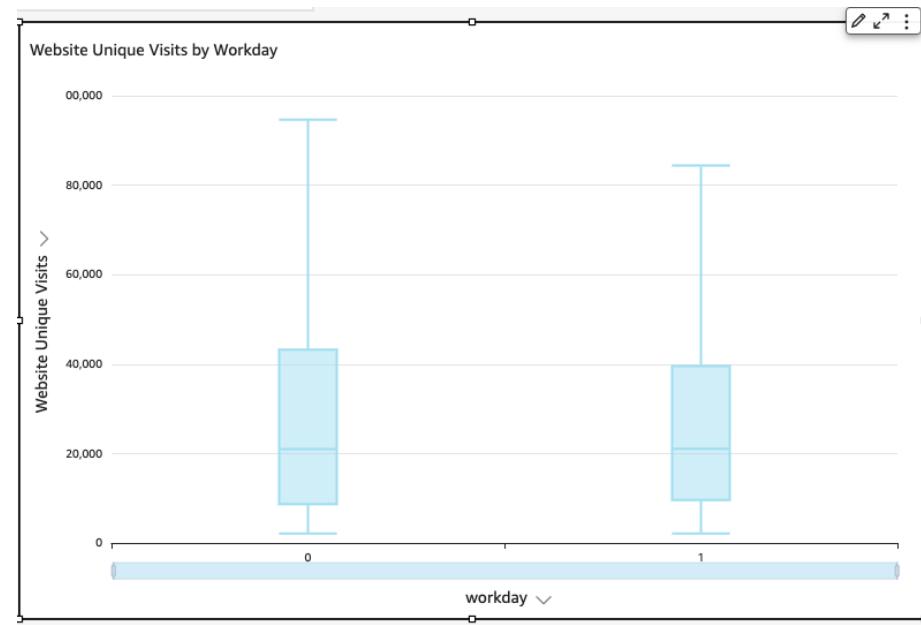
- 산점도는 두 변수 간의 관계를 시각화하는 데 사용
- 각 데이터 포인트는 두 변수 값의 교차점에 위치
- 예상 시나리오
  - 비즈니스: 광고비와 매출 간의 상관관계 분석
  - 교육: 공부 시간과 성적 간의 관계
  - 건강: 체중과 혈압 간의 관계



## 기본 차트 유형 및 사용법

### 박스플롯 (Box Plot)

- 박스플롯은 데이터의 분포와 이상치를 시각화하는 데 사용
- 중앙값, 사분위수, 최소값, 최대값 및 이상치를 표시
- 예상 시나리오
  - 비즈니스: 직원 연봉 분포 분석
  - 교육: 시험 점수 분포와 이상치 분석
  - 건강: 체질량지수(BMI) 분포



[https://docs.aws.amazon.com/ko\\_kr/quicksight/latest/user/box-plots.html](https://docs.aws.amazon.com/ko_kr/quicksight/latest/user/box-plots.html)

## 이론 코드

---

<https://colab.research.google.com/drive/1MM6jVxwVZbzYVgfWgUdDzew-o9J4ANHD?usp=sharing>

## 실습 과제

---

[https://colab.research.google.com/drive/14TwUEkryfyxI-X2kA5ELGj8picE8TfA2?  
usp=sharing](https://colab.research.google.com/drive/14TwUEkryfyxI-X2kA5ELGj8picE8TfA2?usp=sharing)

# 실습 진행