

생성형AI

Day 8

데이터 시각화 II



목차

1. 고급 차트 유형
2. 데이터의 형태에 따른 시각화
3. 인터랙티브 시각화
4. 시각화의 디자인 원칙
5. 시각화와 스토리텔링
6. 실습과제



고급 차트 유형

히트맵 (Heatmap)

- 히트맵은 행렬 형식의 데이터를 색상으로 시각화하여 데이터의 크기나 빈도 등을 파악하기 위해 사용되는 도구
- 색상이 짙어질수록 데이터 값이 크거나 빈도가 높음을 나타냄
- 예시:
 - 상관 행렬: 변수들 간의 상관 관계 시각화
 - 웹사이트 클릭: 웹페이지의 사용자 클릭 빈도를 시각화

In summer, temperatures at the North Pole can reach over 20°C

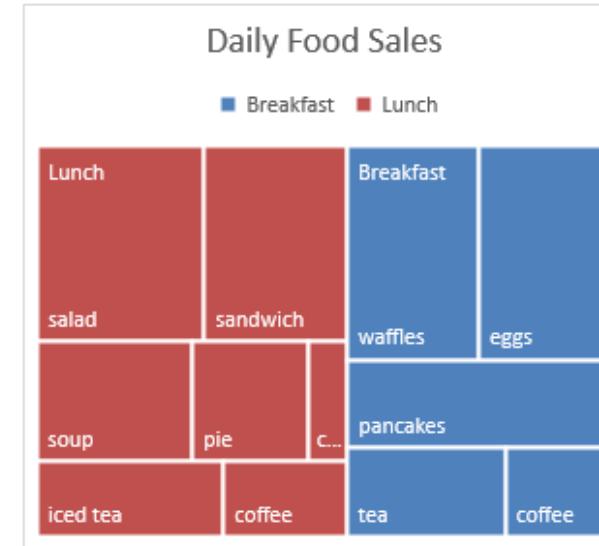
WEATHER IN DEGREES CELSIUS

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Avg. Temp (°C)	-24.2	-20.3	-12.4	-1	8.8	15	15.9	13.4	6.9	-4.1	-15.3	-23.3
Max. Temp (°C)	-19.5	-14.3	-5	4.9	14.7	21	21.5	18.9	12.1	0.2	-11.1	-19.1
Min. Temp (°C)	-28.8	-26.2	-19.8	-6.9	2	9.1	10.4	8	1.8	-8.3	-19.5	-27.4

고급 차트 유형

트리맵 (Treemap)

- 트리맵은 계층적 데이터를 직사각형으로 시각화하여 데이터의 크기와 비율을 나타냄
- 각 직사각형의 크기는 해당 데이터의 크기를 나타내며 계층 구조 내의 비율을 비교하는 데 탁월함
- 예시:
 - 파일 시스템: 디스크 사용량을 시각화하여 폴더와 파일의 크기를 비교
 - 사업 부문별 매출: 각 부문별 매출 비율을 시각화



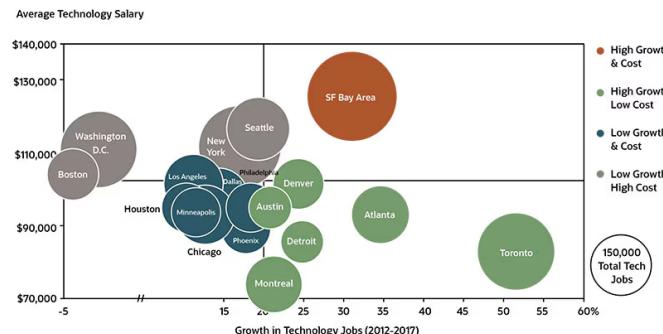
고급 차트 유형

버블차트 (Bubble chart)

- 버블 차트는 산점도의 확장된 형태로, 데이터 포인트의 크기를 추가 변수로 시각화
- 각 버블의 위치는 두 변수의 값을 나타내고, 크기는 세 번째 변수를 나타냄
- 예시:
 - GDP와 기대 수명: 각 국가의 GDP와 기대 수명을 시각화하고, 버블 크기로 인구 수를 나타냄
 - 매출과 이익률: 제품별 매출과 이익률을 시각화하고, 버블 크기로 시장 점유율을 나타냄

Technology Markets In North America

While San Francisco is still the largest market for technology jobs, Toronto is the fastest growing and it offers companies significantly lower wage costs.

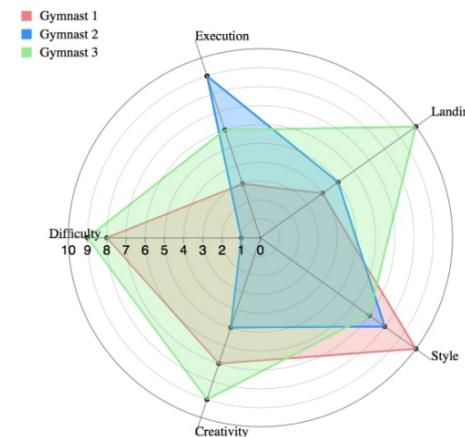


고급 차트 유형

레이더 차트 (Radar Chart)

- 여러 변수를 축으로 하여 다각형으로 시각화하는 방법
- 각 축은 하나의 변수를 나타내며, 중앙에서 각 축까지의 거리는 변수의 값
- 예시:
 - 선수의 성과 평가: 다양한 성과 지표를 시각화하여 선수의 강점과 약점 파악
 - 제품 비교: 여러 제품의 다양한 특성을 시각화하여 비교

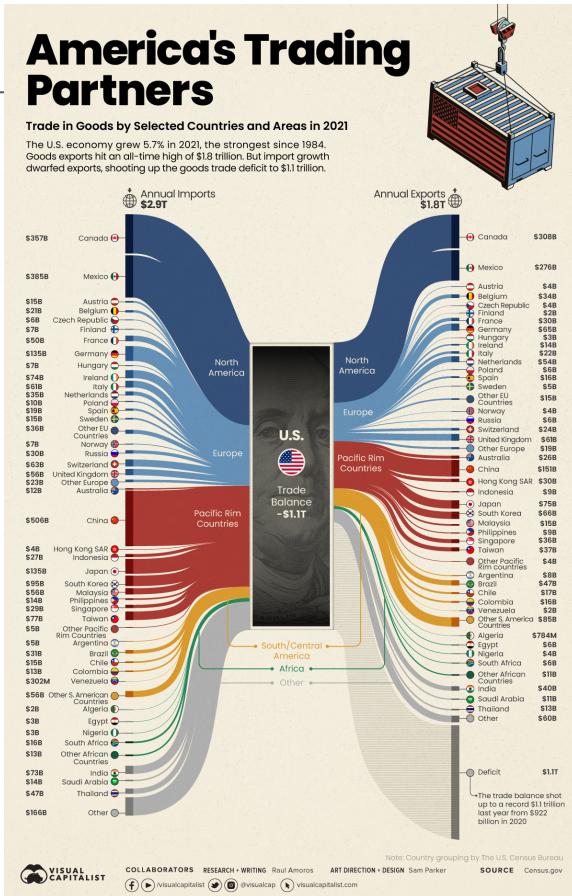
Gymnast Scoring Radar Chart



고급 차트 유형

생키 다이어그램 (Sankey Diagram)

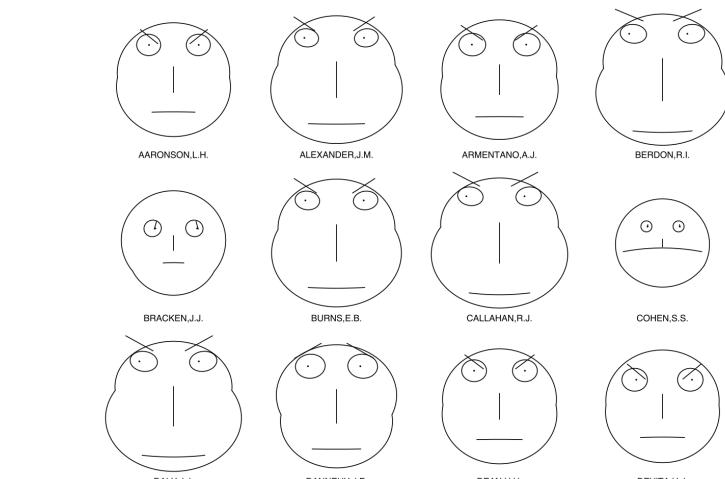
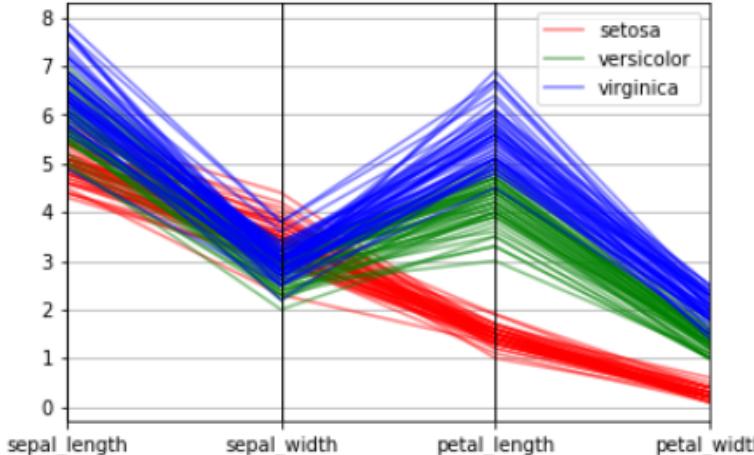
- Sankey 다이어그램은 데이터의 흐름과 비율을 시각화하는 방법
- 흐름의 두께가 데이터의 크기를 나타냅니다
- 예시:
 - 에너지 소비 흐름: 에너지의 생산부터 소비까지의 흐름 시각화
 - 재정 흐름: 자금의 유입과 유출 시각화



데이터의 형태에 따른 시각화

다변량 데이터 시각화

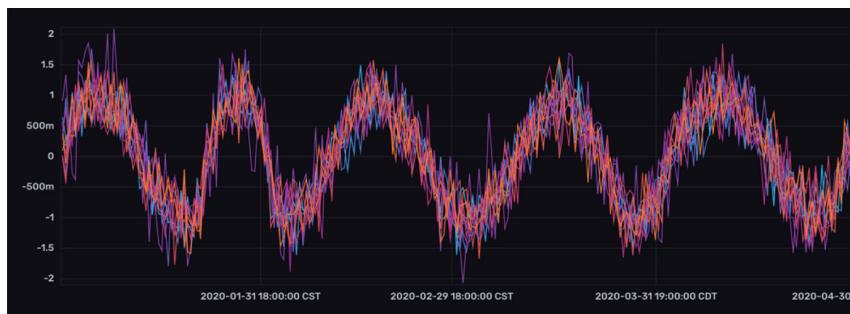
- 여러 변수 간의 관계와 패턴을 파악하는 시각화 방법
- 데이터 세트를 이해하고 해석하는 데 이점
- 너무 많은 변수를 한 그래프에 표시하면 복잡해져서 해석이 어려울 수 있음
- 산점도, 평행좌표, 히트맵, 체르노프 페이스



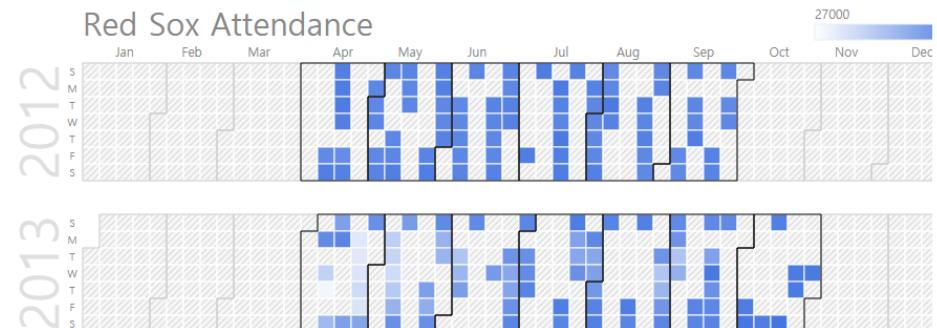
데이터의 형태에 따른 시각화

시계열 데이터 시각화

- 시간에 따른 변화를 분석하여 추세, 계절성, 주기성을 파악
- 선그래프, 히트맵, 캘린더 차트



<https://www.ciokorea.com/t/544/bi/199544>

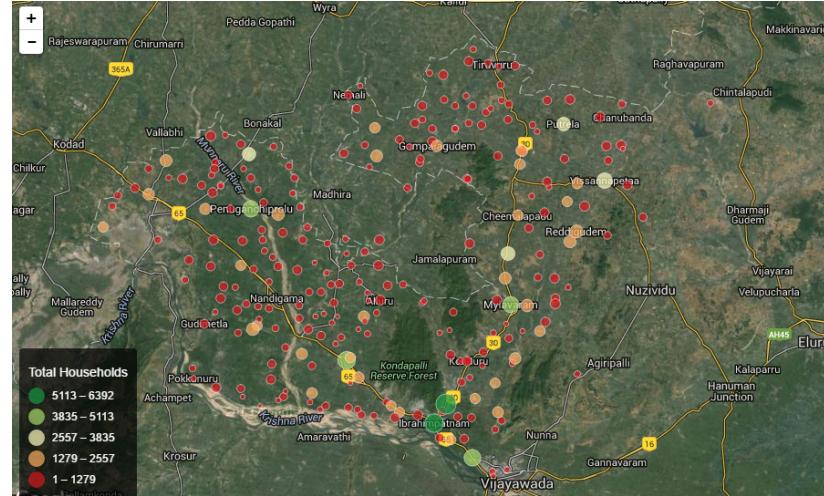
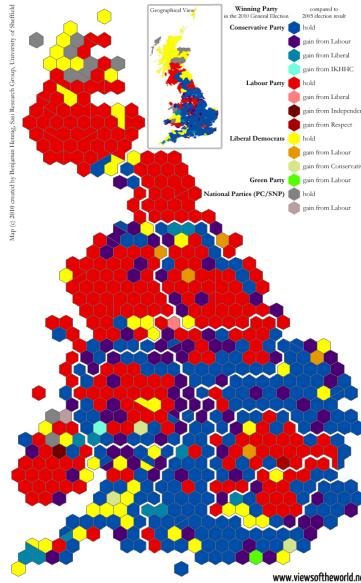


<https://developers.google.com/chart/interactive/docs/gallery/calendar?hl=ko>

데이터의 형태에 따른 시각화

지리 데이터 시각화

- 공간적인 패턴과 관계를 파악하고, 특정 위치의 데이터를 분석
- 지도(map)



<https://humansofdata.atlan.com/2016/10/7-techniques-to-visualize-geospatial-data/>

<https://blog.socialcops.com/wp-content/uploads/2016/10/Geospatial-3.jpg>

인터랙티브 시각화

인터랙티브 시각화?

- 인터랙티브 시각화는 사용자와 상호작용할 수 있는 데이터 시각화 방법
- 사용자가 직접 그래프나 차트를 클릭하거나, 확대 축소하거나, 데이터를 필터링 가능
- 사용자는 상호작용을 통해 데이터를 더 깊이 탐색하고, 다양한 관점에서 데이터를 분석 가능
- 이해 증진: 데이터를 다양한 방식으로 탐색할 수 있어, 복잡한 데이터를 더 쉽게 이해할 수 있게 도와줌
- 사용자 참여: 상호작용 요소는 사용자의 참여를 유도함
- 효율적 탐색: 데이터 필터링, 확대, 축소, 세부 정보 확인이 가능해 효율적인 데이터 탐색 가능

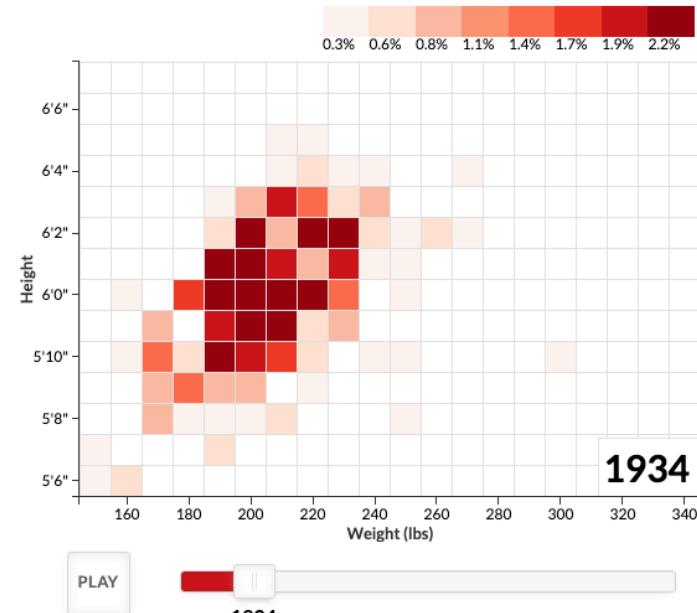
인터랙티브 시각화 주의점

- 과도한 상호작용: 너무 많은 상호작용 요소는 사용자를 혼란스럽게 함
- 데이터 정확성: 실시간 데이터 분석 시 데이터의 정확성을 유지 필수
- 성능 최적화: 대량의 데이터를 시각화할 때는 성능을 고려

인터렉티브 시각화

인터렉티브 요소

- 슬라이서 (Slicer)
 - 데이터의 특정 범위를 선택할 수 있는 슬라이드 바 형태의 컨트롤러
 - 예시: 시간 범위를 선택하여 해당 기간의 데이터를 시각화
- 드롭다운 메뉴 (Dropdown Menu)
 - 여러 옵션 중 하나를 선택할 수 있는 메뉴
 - 특정 카테고리나 지역을 선택하여 해당 데이터를 시각화
- 버튼 (Buttons)
 - 특정 동작을 실행할 수 있는 버튼
 - 예시: 특정 필터를 적용하거나, 원래 상태로 리셋
- 다중 축 (Multiple Axes)
 - 두 개 이상의 축을 사용하여 여러 변수를 동시에 시각화
 - 예시: 매출과 이익률을 동일 그래프에 시각화
- 연결된 시각화 (Linked Visualizations)
 - 하나의 시각화에서 선택한 데이터가 다른 시각화에 반영
 - 예시: 지도에서 특정 지역을 클릭하면 해당 지역의 상세 데이터가 표에 나타남.



<https://noahveltman.com/nflplayers/>

인터렉티브 시각화

Choose an area
UK - Principal

64,596,752 people in 2014

age 64

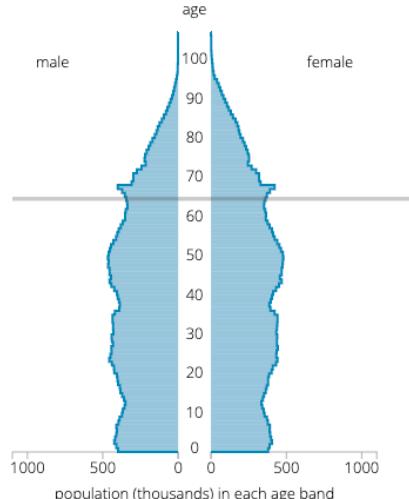
344,106 males

48.7%

362,424 females

51.3%

706,530 people (1.1% of total population)



Choose an area
UK - Low fertility

64,596,752 people in 2014

age 64

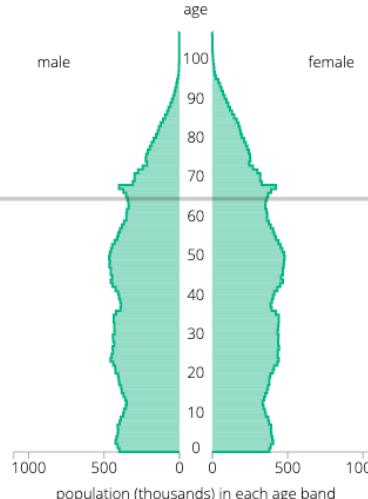
344,106 males

48.7%

362,424 females

51.3%

706,530 people (1.1% of total population)



◀ back a year ▶ play ▶ forward a year

2014 2019 2024 2029 2034 2039

Outlines unlocked Overlap off number Clear selection

시각화의 디자인 원칙

- 목적성
 - 시각화의 목적을 명확히 하는 것
 - 시각화는 특정 질문에 답하거나, 특정 인사이트를 도출하거나, 데이터를 더 쉽게 이해할 수 있도록 돋는 역할

- 간결성
 - 간결하고, 복잡하지 않아야 함
 - 불필요한 요소를 제거하여 핵심 정보를 간결하게 전달

- 명확성
 - 데이터를 명확히 전달해야 함
 - 적절한 레이블, 축, 제목을 사용하여 정보를 명확히 전달

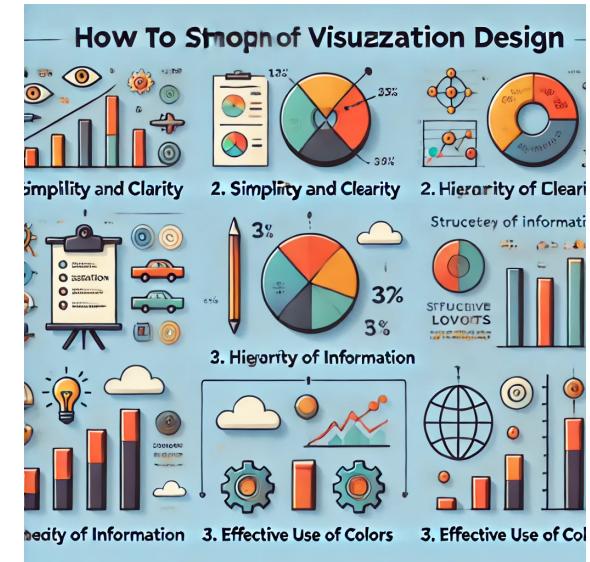
- 정보의 계층구조
 - 데이터의 중요도에 따라 시각적으로 구분하여 정보를 계층적으로 전달 해야함
 - 중요한 정보는 더 강조하고, 덜 중요한 정보는 덜 강조

- 일관성 (Consistency)

- 시각화 스타일, 색상, 레이블 등을 일관되게 유지하여 쉽게 인식하고 이해할 수 있도록 해야함

- 접근성 (Accessibility)

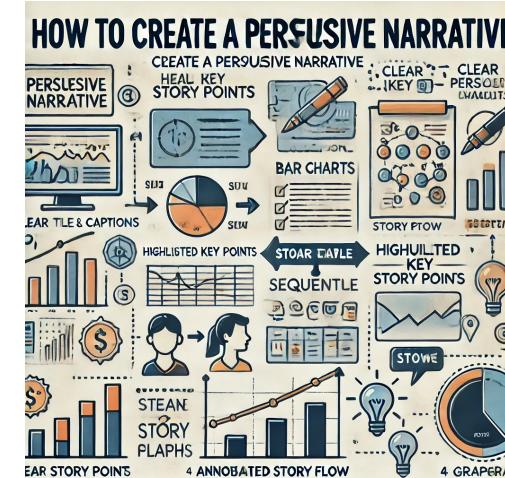
- 모든 사용자가 쉽게 이해할 수 있어야함
- 색맹 사용자나 시각장애인을 고려하여 색상과 레이블 선택



시각화와 스토리텔링

시각화 요소를 통한 설득력 있는 이야기 구성

- 다양한 시각적 요소를 효과적으로 사용하면 데이터를 더 명확하게 전달하고, 청중의 관심을 끌며, 중요한 메시지를 강조하는 데 도움이 됨
- 시각적 강조: 중요한 데이터를 강조하여 청중의 주의를 끌고 핵심 메시지를 전달
 - 색상 사용, 크기 변화, 텍스트 추가
- 애니메이션과 전환 효과: 애니메이션을 사용하여 데이터 변화와 흐름을 더 명확하게 전달
 - 애니메이션, 전환효과
- 비주얼 히스토리: 시각적 요소를 통해 데이터의 역사적 변화를 보여줌
 - 타임라인, 과거 데이터와 비교



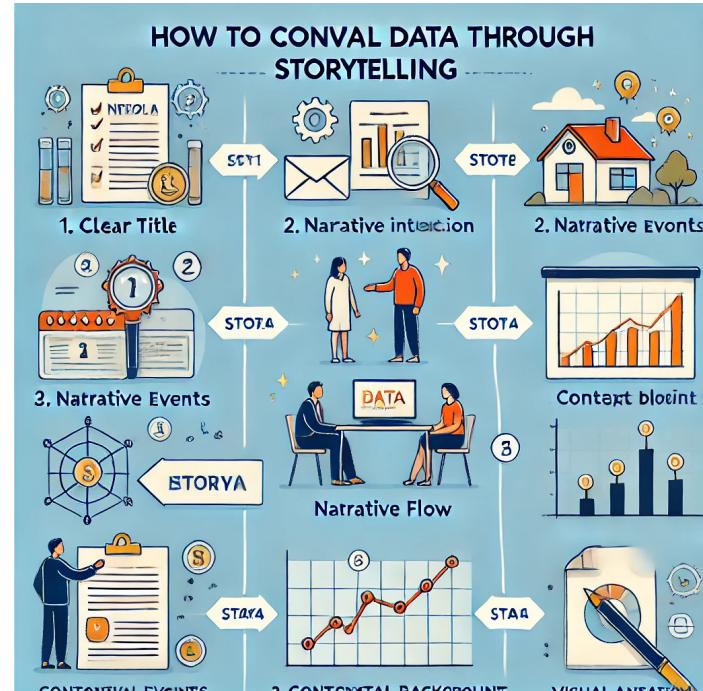
시각화와 스토리텔링

이야기를 통한 데이터 전달

- 데이터의 흐름을 스토리의 형태로 구성
- 스토리텔링을 통해 복잡한 데이터를 쉽게 이해 가능
- 스토리 형식은 데이터를 더 오래 기억에 남게함

스토리라인 구성

- 데이터의 흐름과 주요 포인트를 자연스럽게 연결
 - 도입: 데이터의 배경과 목적을 소개
 - 전개: 데이터를 단계별로 설명하며, 주요 인사이트를 도출
 - 클라이맥스: 데이터에서 가장 중요한 부분을 강조하여 전달
 - 결말: 데이터의 요약과 결론을 제시, 필요한 경우 행동 요구



시각화와 스토리텔링

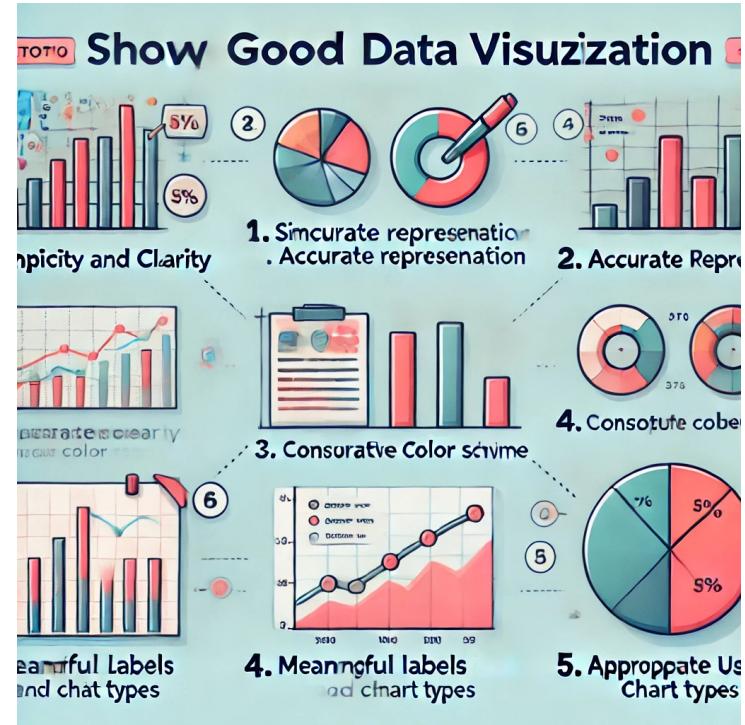
청중을 고려한 시각화 설계

- 시각화를 만들기 전에 청중이 누구인지, 그들의 배경과 지식 수준이 어떤지 파악하는 것이 중요
 - 데모그래픽: 청중의 나이, 성별, 직업, 교육 수준 등
 - 지식 수준: 청중의 데이터 분석 및 시각화에 대한 이해 수준
 - 관심사: 청중이 어떤 정보에 관심이 있는지, 어떤 결론을 도출하고자 하는지
- 맞춤형 시각화
 - 간단한 시각화: 복잡한 데이터를 간단한 차트나 그래프로 표현하여, 비전문가도 쉽게 이해할 수 있도록 함
 - 상세한 시각화: 전문가나 데이터 분석가를 위해 복잡한 데이터 관계와 분석 결과를 상세히 시각화
 - 유형별 맞춤화: 청중이 선호하는 시각화 유형을 사용 (예: 경영진은 대시보드, 개발자는 상세 그래프)
 - 경영진에게는 요약된 KPI 대시보드와 주요 성과 지표를 기술팀에게는 데이터의 세부 분석 결과와 상관 관계 그래프

시각화와 스토리텔링

좋은 시각화

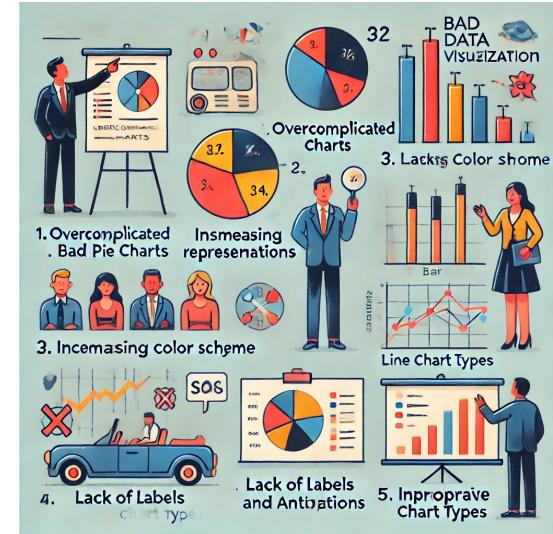
- 명확성 (Clarity)
 - 데이터를 명확하고 일관되게 전달
 - 적절한 레이블, 축, 제목을 사용
- 간결성 (Simplicity)
 - 불필요한 요소를 제거하여 핵심 정보를 강조
 - 그래프에 불필요한 3D 효과나 장식을 제거
- 정확성 (Accuracy)
 - 데이터의 정확한 값을 표시하고, 왜곡 없이 데이터를 전달
 - 적절한 축 범위를 설정하여 데이터의 실제 변화를 정확히 반영
- 적합한 차트 유형 선택
 - 데이터의 특성과 전달하려는 메시지에 적합한 차트 유형을 선택
 - 예를 들어 시간에 따른 변화를 시각화할 때는 선 그래프를 사용



시각화와 스토리텔링

나쁜 시각화

- 혼란스러움 (Confusion)
 - 데이터가 명확하지 않고, 불필요한 요소로 인해 혼란을 초래
 - 3D 효과나 과도한 색상 사용으로 데이터를 이해하기 어렵게 만드는 경우
- 불필요한 복잡성 (Unnecessary Complexity)
 - 지나치게 복잡한 그래프는 데이터를 이해하기 어려움
 - 하나의 그래프에 너무 많은 데이터 세트를 포함하는 경우
- 왜곡된 데이터 (Distorted Data)
 - 데이터 축을 잘못 설정, 특정 데이터를 과장해 왜곡된 메시지를 전달해 오해를 만듦
 - 축의 범위를 조정하여 데이터의 변화를 과장하는 경우
- 잘못된 차트 유형 (Inappropriate Chart Type)
 - 데이터의 특성에 맞지 않는 차트 유형을 선택하여 데이터에 대한 오해를 부름
 - 시간에 따른 변화를 파이 차트로 표현하여 잘못된 인식을 주는 경우



이론 실습

colab: <https://colab.research.google.com/drive/1IQXNxrvQMnk-kXZQx0m-D6xxGZhqB6D3?usp=sharing>

실습 과제

1. 고급 차트 유형에서 제시한 5개 시각화 방법 사용해보기

- 히트맵
- 트리맵
- 버블차트
- 레이더 차트
- 생키 다이어그램

2. 시각화한 차트를 인터랙티브하게 만들어보기

3. 지리 정보 데이터 시각화 해보기

- geopandas
- folium

실습 진행