



4과목.빅데이터 결과 해석

Ch_02. 분석결과 해석 및 활용

SEC 01. 분석결과 해석

SEC 02. 분석결과 시각화

SEC 03. 분석결과 활용

빅데이터 분석 기사(4과목. 빅데이터 결과 해석)

CHAPTER 1. 분석 모형 평가 및 개선

CHAPTER 2. 분석 모형 개선

분석결과 해석 및 활용

분석결과 해석 및 활용 챕터는 총 3개의 작은 섹션으로 구성된다.

1. 분석결과 해석
2. 분석결과 시각화
3. 분석결과 활용

4. 분석결과 해석 및 활용 – 분석결과 해석

01 분석 모형 해석

1) 데이터 시각화(Data Visualization)

- 데이터 시각화는 주어진 데이터에 대한 이해를 돕기 위한 방법 중 하나로 데이터의 특성을 묘사하고 표현하는 방법이다.
- 그림, 색상, 그래프, 도형 등 다양한 그래픽 요소를 활용하여 데이터를 설명하는 방법이다.
- 데이터 시각화의 목적은 정보전달과 설득에 있다.
- 데이터 시각화의 유형에는 시간, 공간, 분포, 관계, 비교 시각화가 있다.

4. 분석결과 해석 및 활용 – 분석결과 해석

2) 데이터 시각화 도구

● 데이터 시각화 도구에는 파인 리포트, 차트 블록, 구글 데이터 스튜디오, 태블로, 파이썬 등이 있다.

① 파인 리포트(Fine Report) : 간단한 드래그 앤 드롭(drag & drop) 방식으로 보고서를 쉽게 만들 수 있고, 데이터 의사결정 분석 시스템을 구축할 수 있는 도구이다.

② 차트 블록(Chart Blocks) : 코딩 없이 스프레드시트와 데이터베이스를 활용하여 쉽게 데이터를 시각화 하는 도구이다.

③ 구글 데이터 스튜디오(Google Data Studio) : 구글 데이터 스튜디오는 구글 마케팅 플랫폼에 있는 웹 기반의 데이터 시각화 도구이다. 다양한 데이터를 연결, 간단한 시각화 작업, 보고서 공유가 가능하다.

④ 태블로(Tableau) : 클라우드 기반으로 서비스되며, 차트, 그래프 등 다양한 그래픽 기능을 제공하는 시각화 도구이다.

⑤ 파이썬(Python) : 다양한 그래프 라이브러리(library)를 활용한 프로그래밍을 통해 구현할 수 있는 시각화 도구이다.

⑥ R 언어 : 시각화를 위해 파이썬과 마찬가지로 패키지라는 개념을 사용해 여러 가지 시각화 도구들을 불러와 프로그래밍을 통해 구현할 수 있는 시각화 도구이다.

4. 분석결과 해석 및 활용 – 분석결과 해석

3) 데이터 시각화 분류

- 데이터 시각화는 데이터 시각화, 정보 시각화, 정보 디자인, 인포그래픽으로 분류할 수 있다.

데이터 시각화 분류	
분류	설명
데이터 시각화 (Data Visualization)	데이터 분석 결과를 쉽게 이해할 수 있도록 시각적으로 표현하는 방법
정보 시각화 (Information Visualization)	대규모 데이터를 색상, 도표, 이미지 등을 활용하여 요약적으로 표현하는 방법
정보 디자인 (Information Design)	<ul style="list-style-type: none">• 정보를 효율적으로 사용할 수 있게 조직하고 구성하는 디자인 기술 및 업무• 정보를 확인하는 대상이 명확히 이해할 수 있도록 보다 쉽게 시각화하여 표현하는 방법
인포그래픽 (Infographics)	정보를 빠르고 분명하게 표현하기 위해 정보, 자료, 지식을 그래픽을 활용하여 시각적으로 표현하는 방법

4. 분석결과 해석 및 활용 – 분석결과 해석

4) 데이터 시각화 절차(프로세스)

- 구조화 → 시각화(기본 틀) → 시각 표현(차트 완성)

구조화		시각화		시각 표현
<ul style="list-style-type: none">• 데이터 수집 및 탐색• 데이터 분류• 데이터 배열• 데이터 재배열	⇒	<ul style="list-style-type: none">• 시간 시각화• 분포 시각화• 관계 시각화• 비교 시각화• 여러 변수 비교• 공간 시각화	⇒	<ul style="list-style-type: none">• 그래픽 7요소• 그래픽 디자인 기본 원리• 인터랙션• 시각정보 디자인 7원칙

그래픽 7요소 : 명도, 색상, 질감, 형태, 위치, 방향, 크기

그래픽 디자인 기본 원리 : 정렬, 균형, 반복, 근접성, 대비 및 공간의 원칙

시각정보 디자인 7원칙 : 시각적 비교 강화, 인과 관계 표시, 다중 변수 표시, 텍스트 및 그래픽과 조화롭게 배치, 콘텐츠 질과 연관성 분명히 표현, 시간 순 아닌 공간 순으로 나열, 정량적 자료의 정량성 유지

4. 분석결과 해석 및 활용 – 분석결과 해석

02 비즈니스 기여도 평가

비즈니스 기여도 평가 기법

- 비즈니스 기여도 평가는 빅데이터 분석 모형을 실제 업무에 적용할 경우 비즈니스 상에 어떠한 개선점이 있는지 정량적으로 확인하기 위한 방법이다. 비즈니스 기여도 평가 기법에는 TCO, ROI, NPV, IRR, PP 등이 있다.

① 총소유 비용(TCO : Total Cost of Ownership)

- ▶ 하나의 장비를 획득할 때, 장비의 수명 주기 동안에 발생하게 되는 모든 연관 비용이다.
- ▶ 하나의 자산을 획득할 때, 주어진 기간 동안 모든 연관 비용을 고려하여 확인하기 위해 사용한다.

② 투자 대비 효과(ROI : Return On Investment)

- ▶ 투자로 얻을 수 있는 순 효과를 총 비용으로 나눈 값이다.
- ▶ 자본 투자에 따른 순 효과의 비율(투자 타당성)이다.

- ▶ 계산식 : $\frac{\text{순이익}}{\text{투자 비용}} * 100$

4. 분석결과 해석 및 활용 – 분석결과 해석

비즈니스 기여도 평가 기법

③ 순 현재 가치(NPV : Net Present Value)

- ▶ 투자로부터 유입되는 미래 현금의 현재 가치와 해당 투자를 위해 투입된 비용의 차액(미래 시점의 순이익 규모)이다.

④ 내부수익률(IRR : Internal Rate of Return)

- ▶ 어떤 사업에 대해 사업기간 동안의 현금 수익 흐름을 현재 가치로 환산하여 합한 값이 투자 지출과 같아지도록 할인하는 할인율이다.
- ▶ 순 현재 가치를 '0'으로 만드는 할인율(연 단위 기대 수익 규모)이다.

⑤ 투자 회수기간(PP : Payback Period)

- ▶ 누적 투자 금액과 매출 금액의 합이 같아지는 기간으로 투자에 소요되는 모든 비용을 회수하는 데 걸리는 기간을 의미한다.(흑자 전환 시점)
- ▶ 보통 연(year) 단위로 기록한다.

4. 분석결과 해석 및 활용 – 분석결과 해석

개념 체크

01 다음 중 데이터 시각화에 대한 설명으로 옳지 않은 것은?

- ① 데이터 시각화의 목적은 단순 정보 제공이다.
- ② 파인 리프트는 간단한 드래그 앤 드롭 방식으로 보고서를 쉽게 만들 수 있는 시각화 도구이다.
- ③ 데이터 시각화는 데이터 시각화, 정보 시각화, 정보 디자인, 인포그래픽으로 분류할 수 있다.
- ④ 정보 시각화는 대규모 데이터를 색상, 도표, 이미지 등을 활용하여 요약적으로 표현하는 방법이다.

데이터 시각화(Data Visualization)

- 데이터 시각화는 주어진 데이터에 대한 이해를 돕기 위한 방법 중 하나로 데이터의 특성을 묘사하고 표현하는 방법이다.
- 그림, 색상, 그래프, 도형 등 다양한 그래픽 요소를 활용하여 데이터를 설명하는 방법이다.
- 데이터 시각화의 목적은 정보전달과 설득에 있다.
- 데이터 시각화의 유형에는 시간, 공간, 분포, 관계, 비교 시각화가 있다.

02 다음 중 정보를 효율적으로 사용할 수 있게 조직하고 구성하는 디자인 기술 및 업무가 의미하는 용어는?

- ① 데이터 시각화
- ② 정보 시각화
- ③ 인포그래픽
- ④ 정보 디자인

데이터 시각화 분류 4가지

1. **데이터 시각화** : 데이터 분석 결과를 쉽게 이해할 수 있도록 시각적으로 표현하는 방법
2. **정보 시각화** : 대규모 데이터를 색상, 도표, 이미지 등을 활용하여 요약적으로 표현하는 방법
3. **정보 디자인** : 정보를 효율적으로 사용할 수 있도록 조작하고 구성하는 디자인 기술 및 업무이며, 정보를 확인하는 대상이 명확히 이해할 수 있도록 보다 쉽게 시각화하여 표현하는 방법
4. **인포그래픽** : 정보를 빠르고 분명하게 표현하기 위해 정보, 자료, 지식을 그래픽을 활용하여 시각적으로 표현하는 방법

4. 분석결과 해석 및 활용 – 분석결과 해석

03 다음 중 데이터 시각화 절차 작업 중 성격이 다른 하나 는?

- ① 그래픽 7요소
- ② 데이터 배열
- ③ 인터랙션
- ④ 시각정보 디자인 7원칙

구조화는 데이터 시각화 절차 중 '구조화'에 속하는 작업이다. 나머지는 데이터 시각화 절차 중 '시각 표현'에 속하는 작업이다.

데이터 시각화 절차(프로세스)

● 구조화 -> 시각화(기본 틀) -> 시각 표현(차트 완성)

구조화 : 데이터 수집 및 탐색, 데이터 분류, 데이터 배열, 데이터 재배열

시각화(기본 틀) : 시간, 분포, 관계, 비교, 여러 변수 비교, 공간 시각화를 할 것인지 결정

시각 표현(차트 완성) : 그래픽 7요소, 그래픽 디자인 기본 원리, 인터랙션, 시각정보 디자인 7원칙

04 다음 설명하는 내용에 알맞은 명칭은?

투자로부터 유입되는 미래 현금의 현재 가치와 해당 투자를 위해 투입된 비용의 차액으로 미래 시점의 순이익 규모를 의미한다.

- ① TCO ② NPV
- ③ ROI ④ IRR

순 현재 가치(NPV; Net Present Value)에 대한 설명이다.

총소유 비용(TCO; Total Cost of Ownership)

● 하나의 장비를 획득할 때, 장비의 수명 주기 동안에 발생하게 되는 모든 연관 비용이다.

● 하나의 자산을 획득할 때, 주어진 기간 동안 모든 연관 비용을 고려하여 확인하기 위해 사용한다.

투자 대비 효과(ROI; Return On Investment)

● 투자로 얻을 수 있는 순 효과를 총 비용으로 나눈 값이다.

● 자본 투자에 따른 순 효과의 비율(투자 타당성)이다.

● 계산식 : $\frac{\text{순이익}}{\text{투자비용}} * 100$

내부 수익률(IRR; Internal Rate of Return)

4. 분석결과 해석 및 활용 – 분석결과 해석

05 다음 중 비즈니스 기여도 평가 기법에 대한 설명으로 옳지 않은 것은?

- ① 총소유 비용은 하나의 장비를 획득할 때, 장비의 수명 주기 동안에 발생하게 되는 모든 연관 비용을 의미한다.
- ② 투자 대비 효과는 투자로 얻을 수 있는 순 효과를 총비용으로 나눈 값이다.
- ③ 순 현재 가치는 투자로부터 유입되는 미래 현금의 현재 가치와 해당 투자를 위해 투입된 비용의 차액을 의미한다.
- ④ 투자 회수 기간은 적자 전환 시점을 의미한다.

투자 회수기간(PP; Payback Period)은 누적투자금액과 매출 금액의 합이 같아지는 기간으로 흑자 전환 시점을 의미한다.

06 다음 중 ROI의 계산식으로 옳은 것은?

① $\frac{\text{현재 가치}}{\text{투자 비용}} * 100$

② $\frac{\text{투자 비용}}{\text{순이익}} * 100$

③ $\frac{\text{순이익}}{\text{투자 비용}} * 100$

④ $\frac{\text{내부 수익률}}{\text{순이익}} * 100$

4. 분석결과 해석 및 활용 – 분석결과 시각화

01 시공간 시각화

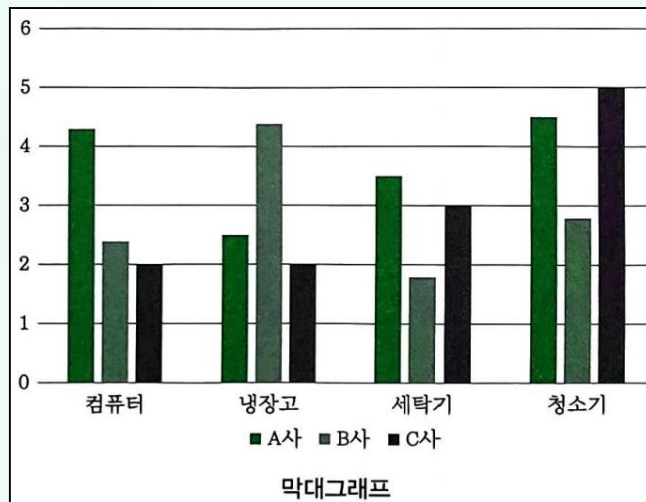
1) 시간 시각화

- 시간 시각화는 시간의 흐름에 따른 데이터의 변화를 나타낸 것이다.
- 일반적으로 x 축은 시간을, y 축은 데이터 값(value)을 나타내고, 시계열 데이터를 통한 데이터의 경향성과 흐름을 파악하는 것이 목적이다.

<시간 시각화 유형>

① 막대그래프(Bar Graph)

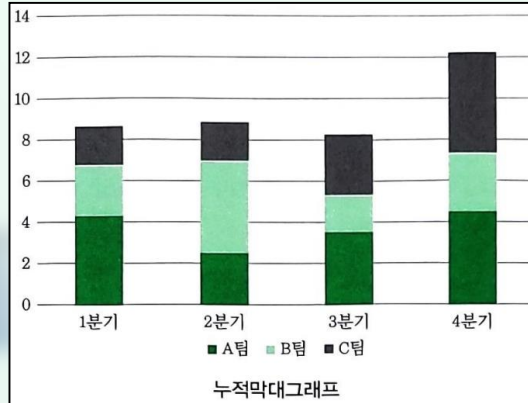
- ▶ 직사각형 형태의 막대를 사용하고, 각 막대는 특정 범주를 나타낸다.
- ▶ 막대는 서로 떨어져서 존재하고, x 축이 반드시 수량이 아니어도 된다.



4. 분석결과 해석 및 활용 – 분석결과 시각화

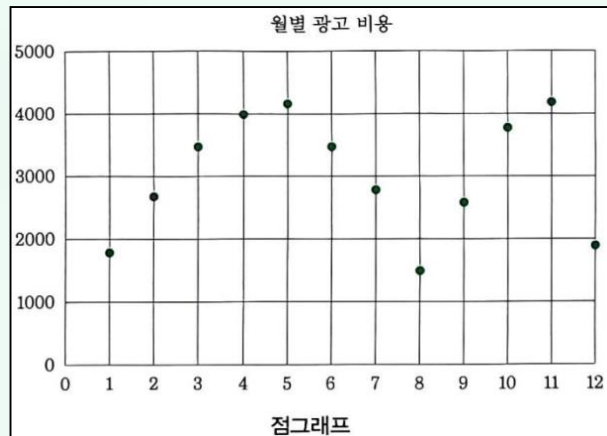
② 누적막대그래프(Stacked Bar Chart)

- ▶ 하나의 막대로 데이터의 여러 범주별 비율을 확인할 수 있는 그래프이다.



③ 점그래프(Dot Plot)

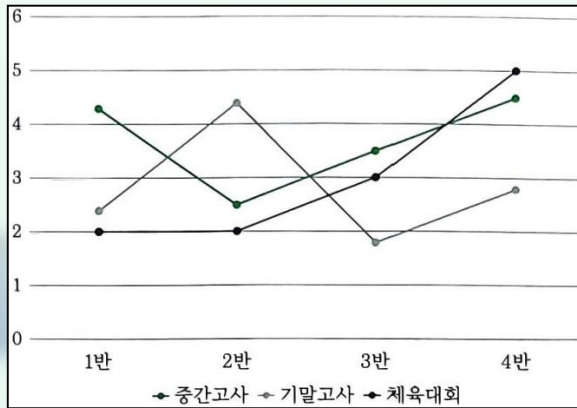
- ▶ x 축에 따른 y 축의 값을 점으로 표시한 그래프로, x 축이 시간, y 축이 데이터인 경우 시간의 흐름에 따른 데이터의 변화를 확인할 수 있다.



4. 분석결과 해석 및 활용 - 분석결과 시각화

④ 선그래프(Line Graph)

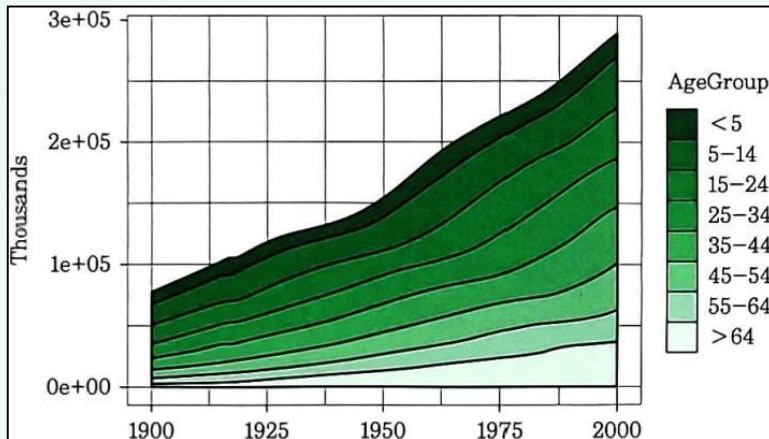
- ▶ 각 범주별 측정된 데이터들의 점을 선분으로 이어서 그린 그래프이다.



선분 : 양쪽에 끝나는 점이 있는 직선 부분

⑤ 영역차트(Area Chart)

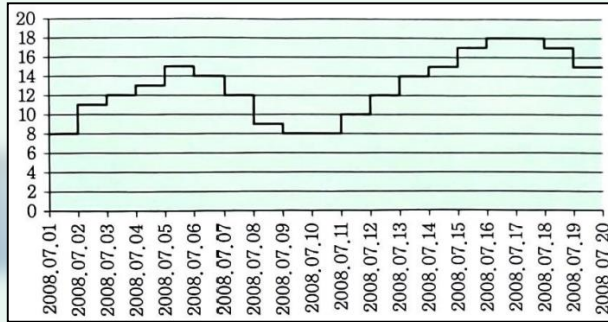
- ▶ 시간 경과에 따른 수량 변화를 표시하는 그래프로, 그래프 면적을 색으로 채워 표현하고, y축 값이 0부터 시작되는 특징이 있다.



4. 분석결과 해석 및 활용 – 분석결과 시각화

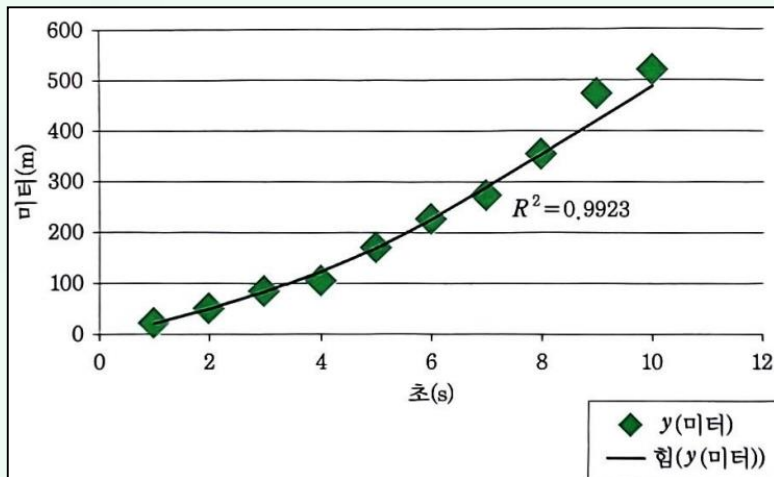
⑥ 계단식그래프(Step Line Graph)

▶ 각 범주별 측정된 데이터를 선분으로 연결하는 것이 아니라 x축과 평행한 일정한 선을 유지하고 값이 급격히 변하는 지점을 이전 데이터와 계단식으로 이어 표현해 주는 그래프이다.



⑦ 추세선(Trend Line)

▶ 일정 기간 동안 측정된 데이터들의 경향성을 보여주는 직선 또는 곡선이다.



4. 분석결과 해석 및 활용 – 분석결과 시각화

2) 공간 시각화

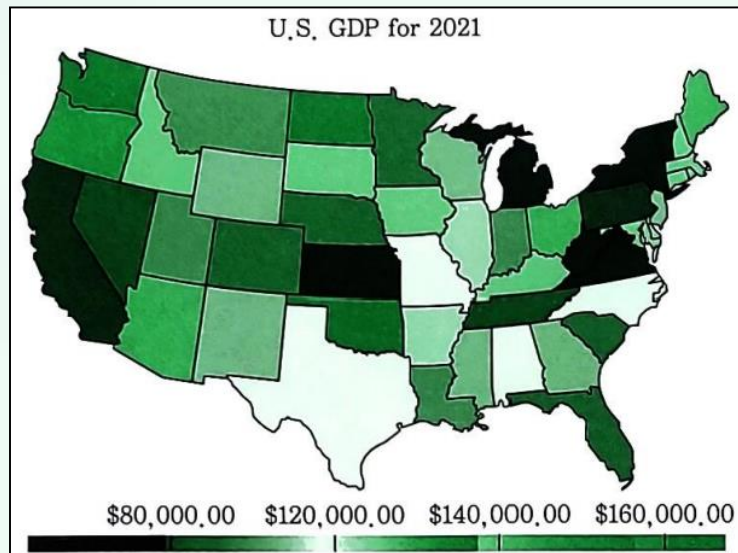
- 공간 시각화는 공간인 지도상에 데이터를 표시하는 시각화 방법이다.
- 지도 데이터를 활용하기 때문에 위도 경도 정보를 활용하여 시각화한다.

<공간 시각화 유형>

공간 시각화의 유형에는 등치지역도, 등치선도, 도트맵, 버블맵, 카토그램이 있다.

① 등치지역도(코로플레스 지도, Choropleth Map)

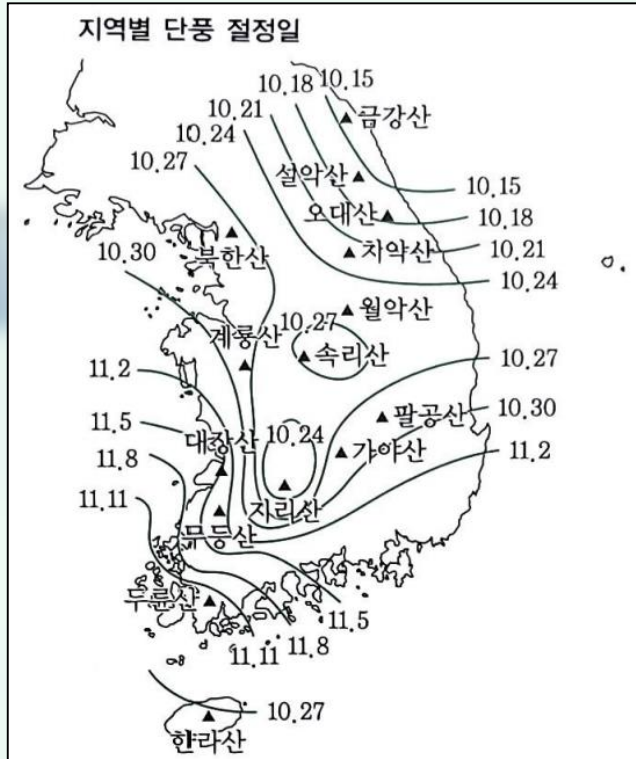
- ▶ 지역별로 색상을 구분하여 지도에 표시하는 그래프이다.
- ▶ 색상은 데이터 수치값을 기반으로 표현한다.



4. 분석결과 해석 및 활용 - 분석결과 시각화

② 등치선도(Isometric Map)

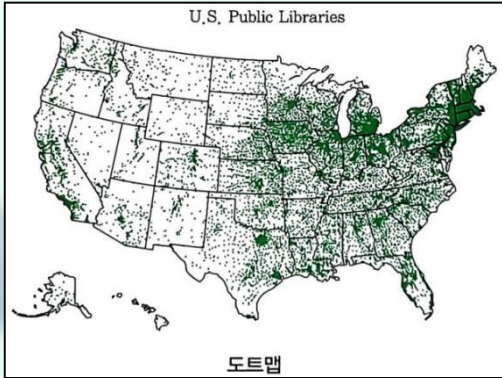
- ▶ 지도상에서 같은 값을 갖는 데이터를 선으로 이어서 표현한 그래프이다.



4. 분석결과 해석 및 활용 – 분석결과 시각화

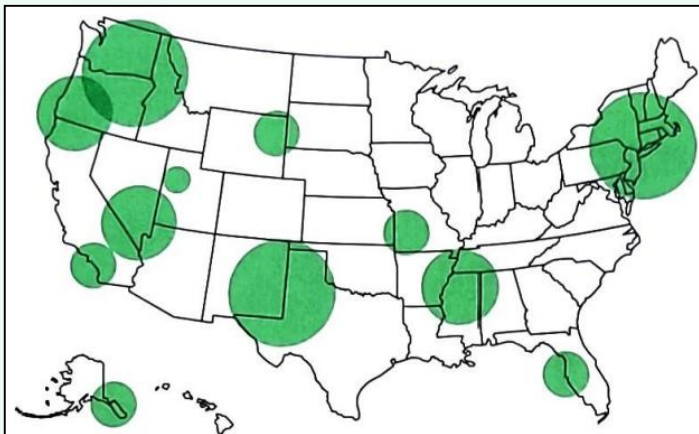
③ 도트맵(도트플롯맵, Dot Map)

- ▶ 측정된 데이터를 해당 데이터의 위도와 경도에 점으로 표시한 그래프이다.



④ 버블맵(버블플롯맵, Bubble Map)

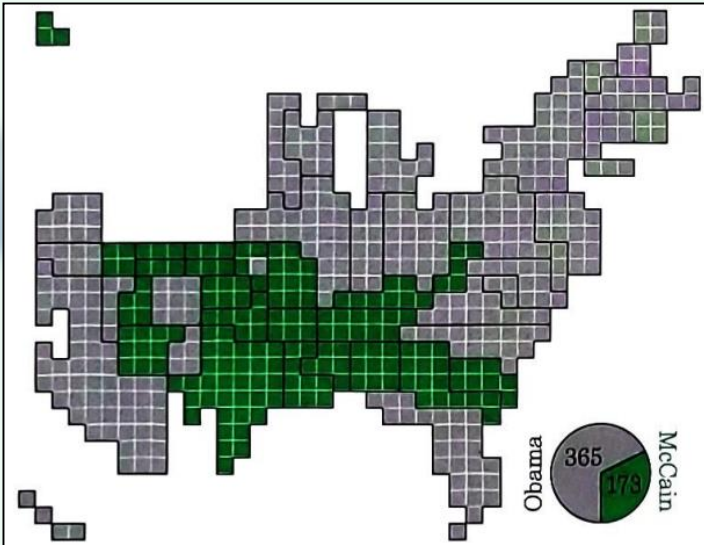
- ▶ 측정된 데이터의 위치는 위도와 경도를 기반으로 표시하고, 크기는 원형의 사이즈로 구별하여 표시하는 그래프이다.



4. 분석결과 해석 및 활용 – 분석결과 시각화

⑤ 카토그램(Catogram)

- ▶ 특정한 데이터 수치에 따라서 지도의 면적이 왜곡되는 그래프이다.
- ▶ 주로 의석 수나 선거인단 수, 인구 등의 데이터를 표현한다.



4. 분석결과 해석 및 활용 – 분석결과 시각화

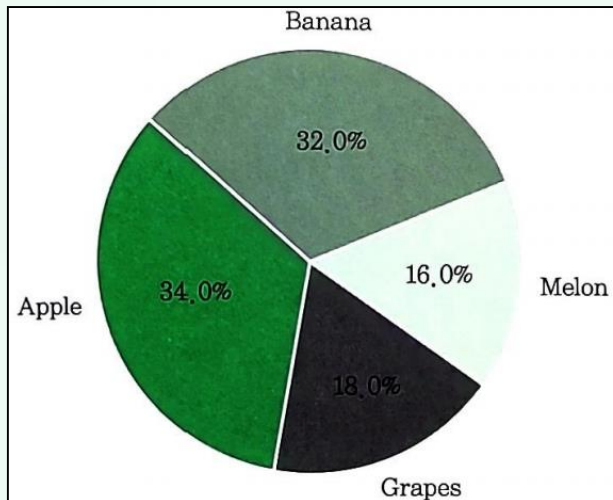
3) 분포 시각화

- 분포 시각화란 데이터의 최솟값, 최댓값, 분포를 나타내는 시각화 방법이다.
- 분포 시각화는 데이터가 차지하는 영역을 나타내는 시각화 방법으로, 시간을 기준으로 구분되는 시계열 데이터와 차이점을 갖는다.

<분포 시각화 유형>

① 파이차트(Pie Chart)

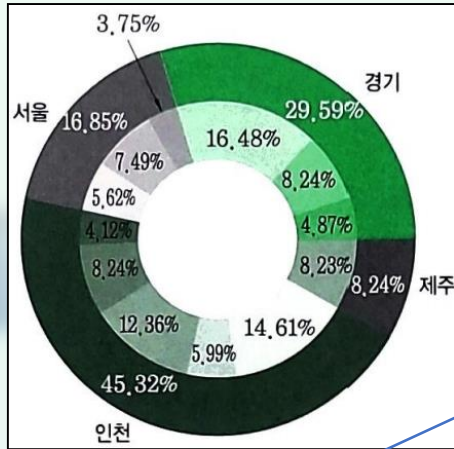
- ▶ 각 범주별 데이터가 차지하는 부분을 파이를 자른 것과 같이 비율별로 조각을 나누어 구분한 그래프이다.
- ▶ 모든 파트의 데이터 값을 합치면 데이터 값 전체의 합이 되고, 비율은 100%가 된다.



4. 분석결과 해석 및 활용 - 분석결과 시각화

② 도넛차트(Doughnut Chart)

- ▶ 파이차트와 구조가 비슷하지만, 중심 부분이 비어 있는 도넛 모양의 그래프이다.



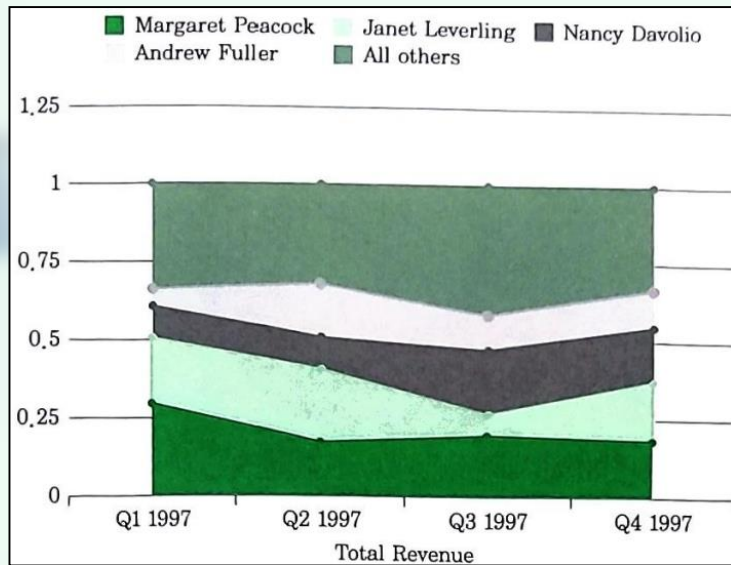
③ 트리맵(Tree Map)

- ▶ 여러 계층 구조(트리 구조)의 데이터를 표현하는 그래프이다.
- ▶ 서로 다른 크기의 사각형을 이용하여 데이터의 비율을 나타내고, 사각형을 겹쳐 배치함으로써 데이터의 대분류와 하위분류를 구분한다.

4. 분석결과 해석 및 활용 – 분석결과 시각화

④ 누적영역그래프(누적연속그래프, Stacked Area Graph)

- ▶ 여러 개의 영역 차트를 여러 층으로 쌓아서 표현한 그래프이다.
- ▶ x축은 시간을 나타내고, y축은 데이터를 나타낸다.



4. 분석결과 해석 및 활용 – 분석결과 시각화

개념 체크

01 다음 중 측정된 데이터를 해당 데이터의 위도와 경도에 점으로 표시한 그래프는?

- ① 등치지역도 ② 산점도
- ③ 도트맵 ④ 버블맵

등치지역도(코로플레스 지도, Choropleth Map)

- 지역별로 색상을 구분하여 지도에 표시하는 그래프이다.
- 색상은 데이터 수치의 값을 기반으로 표현한다.

산점도(Scatter Plot)

- 측정된 데이터들을 x축과 y축을 기반으로 점으로 표시한 그래프이다.
- 측정된 데이터의 분포를 통해 변수 간에 관계 파악이 가능하다(상관관계, 군집화, 이상값 등)

버블맵(버블플롯맵, Bubble Map)

- 측정된 데이터의 위치는 위도와 경도를 기반으로 표시하고, 크기는 원형의 사이즈로 구별하여 표시하는 그래프이다.

02 다음과 같은 특징을 갖는 시각화 방법은?

여러 개의 영역 차트를 여러 층으로 쌓아서 표현한 그래프이다.

x축은 시간을 나타내고, y축은 데이터를 나타낸다.

- ① 트리맵 ② 누적영역그래프
- ③ 카토그램 ④ 코로플레스 지도

트리맵(Tree Map)

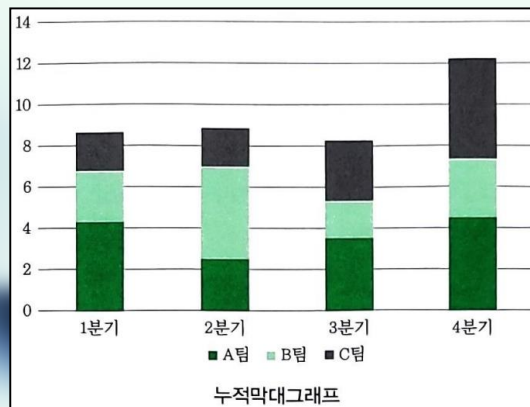
- 여러 계층 구조(트리 구조)의 데이터를 표현하는 그래프이다.
- 서로 다른 크기의 사각형을 이용하여 데이터의 비율을 나타내고, 사각형을 겹쳐 배치함으로써 데이터의 대분류와 하위분류를 구분한다.

카토그램(Catogram)

- 특정한 데이터 수치의 변화에 따라서 지도의 면적이 왜곡되는 그래프이다.
- 주로 의선 수나 선거인단 수, 인구 등의 데이터를 표현한다.

4. 분석결과 해석 및 활용 – 분석결과 시각화

03 다음과 같은 시각화 유형 그래프의 올바른 명칭은?



- ① 막대그래프
- ② 누적막대그래프
- ③ 플로팅 바 차트
- ④ 히스토그램

누적막대그래프(Stacked Bar Chart)

- 하나의 막대로 데이터의 여러 범주별 비율을 확인할 수 있는 그래프이다.

막대그래프(Bar Graph)

- 직사각형 형태의 막대를 사용하고, 각 막대는 특정 범주를 나타낸다.
- 막대는 서로 떨어져서 존재하고, x축이 반드시 수량이 아니어도 된다.

플로팅 바 차트(Floating Bar Chart)

04 다음 시각화 유형 중 시공간 시각화에 속하지 않는 것은?

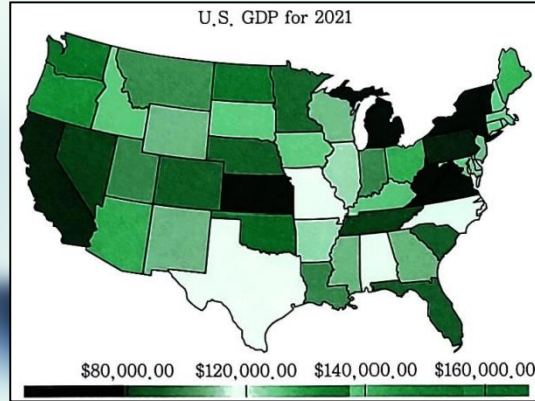
- ① 카토그램
- ② 버블맵
- ③ 산점도 행렬
- ④ 추세선

산점도 행렬(Scatter Plot Matrix)은 관계 시각화 유형에 속하며 산점도를 여러 개의 변수 조합별로 행렬의 형태로 표현한 그래프이다. 추세선(Trend Line)은 시간 시각화에 속하고, 일정 기간 동안 측정된 데이터들의 경향성을 보여주는 직선 또는 곡선이다.

카토그램과 버블맵은 공간 시간화에 속하다.

4. 분석결과 해석 및 활용 – 분석결과 시각화

05 다음과 같은 그래프의 명칭은?



- ① 히트맵 ② 등치선도
- ③ 등치지역도 ④ 카토그램

히트맵(Heat Map)은 비교 시각화 유형의 하나이다.

- 색상으로 표현할 수 있는 다양한 정보를 일정한 이미지 위에 열분포 형태로 표현한 그래프이다.
- 각 칸별 색상은 데이터 값의 크기를 나타내고, 색상이 짙을수록 데이터 값이 큰 것을 의미한다.
- 주로 웹 사이트 방문자 분석에 활용된다.

등치선도(Isometric Map)

- 지도상에서 같은 값을 갖는 데이터를 선으로 이어서 표현한 그래프이다.

06 다음에서 설명하는 시각화 도구는?

특정한 데이터 수치의 변화에 따라서 지도의 면적이 왜곡되는 그래프이다.

주로 의석 수나 선거인단 수, 인구 등의 데이터를 표현한다.

- ① 버블맵 ② 파이차트
- ③ 카토그램 ④ 버블차트

공간 시각화 도구인 카토그램에 대한 설명이다.

파이차트(Pie Chart) – 분포 시각화 유형

- 각 범주별 데이터가 차지하는 부분을 파이를 자른 것과 같이 비율별로 조각을 나누어 구분한 그래프이다.
- 모든 파트의 데이터 값을 합치면 데이터 값 전체의 합이 되고, 비율은 100%가 된다.

버블차트(Bubble Chart) – 관계 시각화 유형

- 산점도는 데이터를 점으로 표현하지만, 버블차트는 데이터의 크기를 추가적으로 고려하여 표현한 그래프이다.

4. 분석결과 해석 및 활용 – 분석결과 시각화

02 관계 시각화

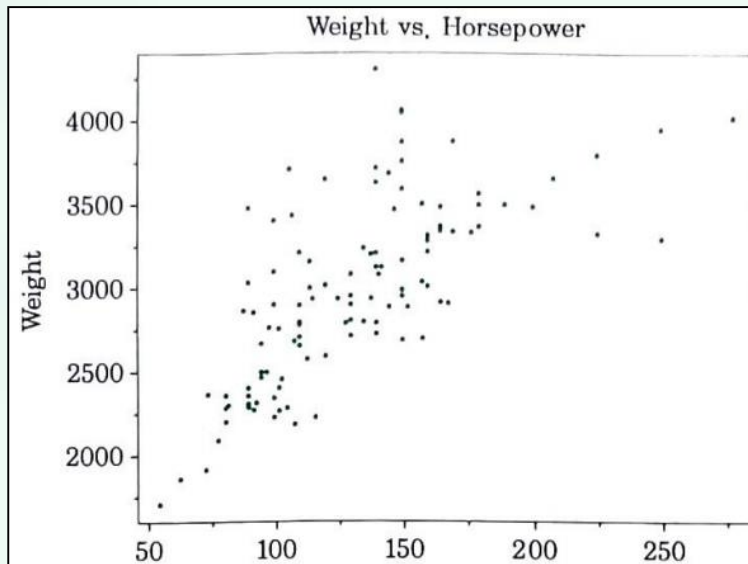
- 관계 시각화란 여러 변수들 사이에 존재하는 연관성, 분포, 패턴을 찾는 시각화 방법이다.

<관계 시각화 유형>

관계 시각화 유형에는 산점도, 산점도 행렬, 버블차트, 히스토그램, 네트워크 그래프가 있다.

① 산점도(Scatter Plot)

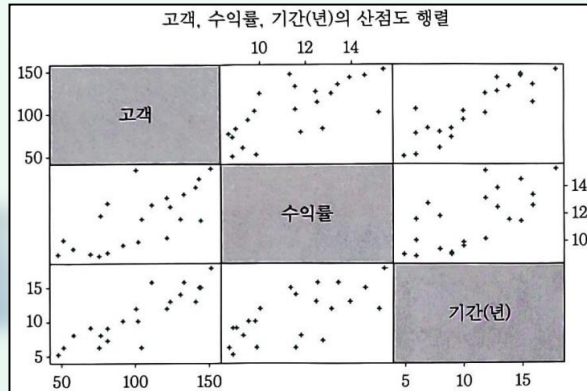
- ▶ 측정된 데이터들을 x 축과 y 축을 기반으로 점으로 표시한 그래프이다.
- ▶ 측정된 데이터의 분포를 통해 변수간의 관계 파악이 가능하다(상관관계, 군집화, 이상값 등)



4. 분석결과 해석 및 활용 – 분석결과 시각화

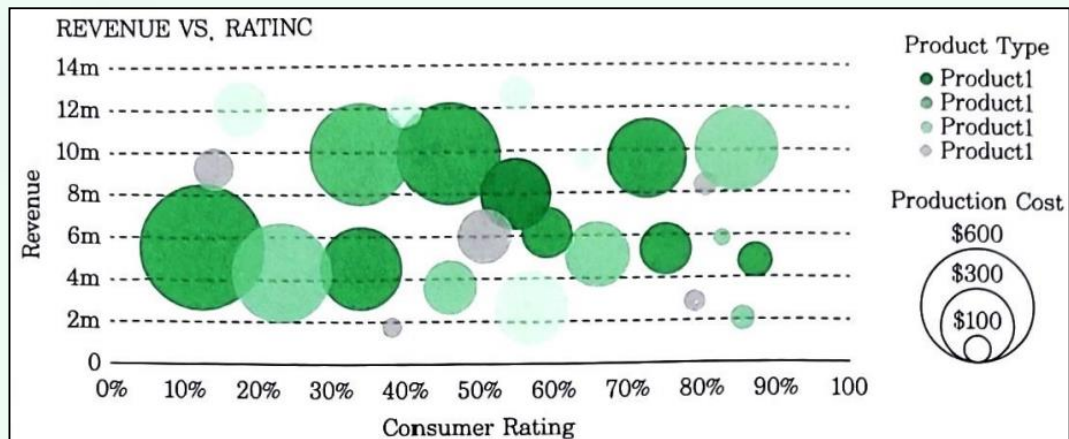
② 산점도 행렬(Scatter Plot Matrix)

▶ 산점도를 여러 개의 변수 조합별로 행렬의 형태로 표현한 그래프이다.



③ 버블차트(Bubble Chart)

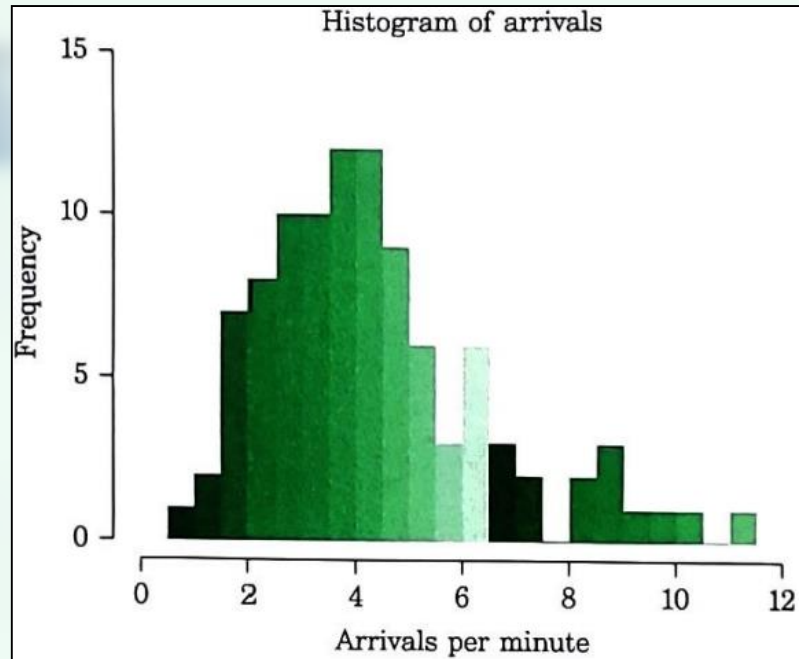
▶ 산점도는 데이터를 점으로 표현하지만, 버블차트는 데이터의 크기를 추가적으로 고려하여 표현한 그래프이다.



4. 분석결과 해석 및 활용 – 분석결과 시각화

④ 히스토그램(Histogram)

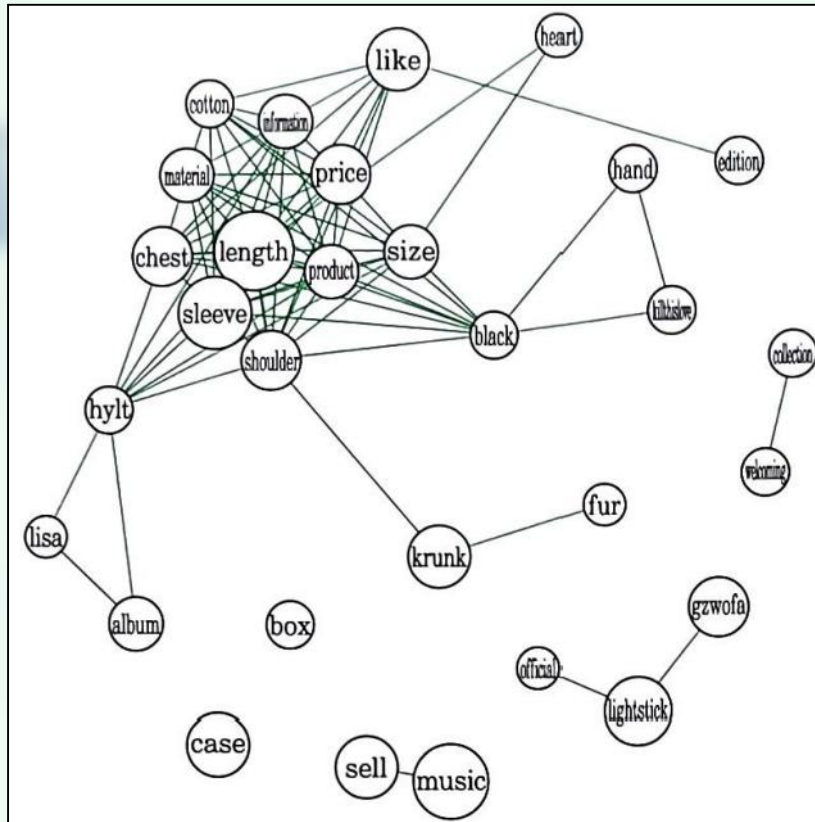
- ▶ 데이터의 분포를 서로 붙어 있는 직사각형 형태로 시각화하여 표현하는 그래프이다.
- ▶ 가로축은 반드시 수량을 나타내고, 막대의 너비는 항상 일정해야 한다.
- ▶ 이상값 확인이 가능하다(히스토그램 양쪽 끝, 고립된 막대)



4. 분석결과 해석 및 활용 - 분석결과 시각화

⑤ 네트워크 그래프(Network Graph)

- ▶ 서로 연관된 개체들 간의 관계를 표현하는 그래프이다.
- ▶ 각 개체들은 선으로 연결되며, 연결된 선의 빈도수 등을 통해 개체 간의 관계를 파악할 수 있다.



4. 분석결과 해석 및 활용 – 분석결과 시각화

03 비교 시각화

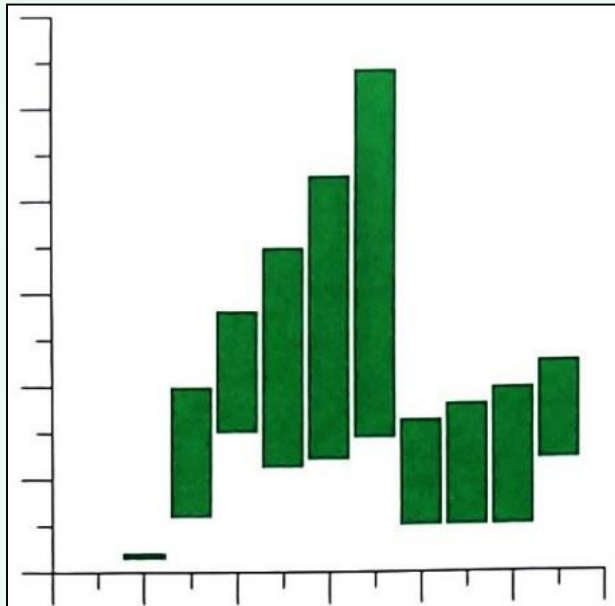
- **비교 시각화란 여러 변수의 데이터를 비교하고자 할 때 사용되는 시각화 방법이다.**

<비교 시각화 유형>

비교 시각화 유형에는 플로팅 바 차트, 히트맵, 체르노프페이스, 스타차트, 평행좌표그래프가 있다.

① 플로팅 바 차트(Floating Bar Chart)

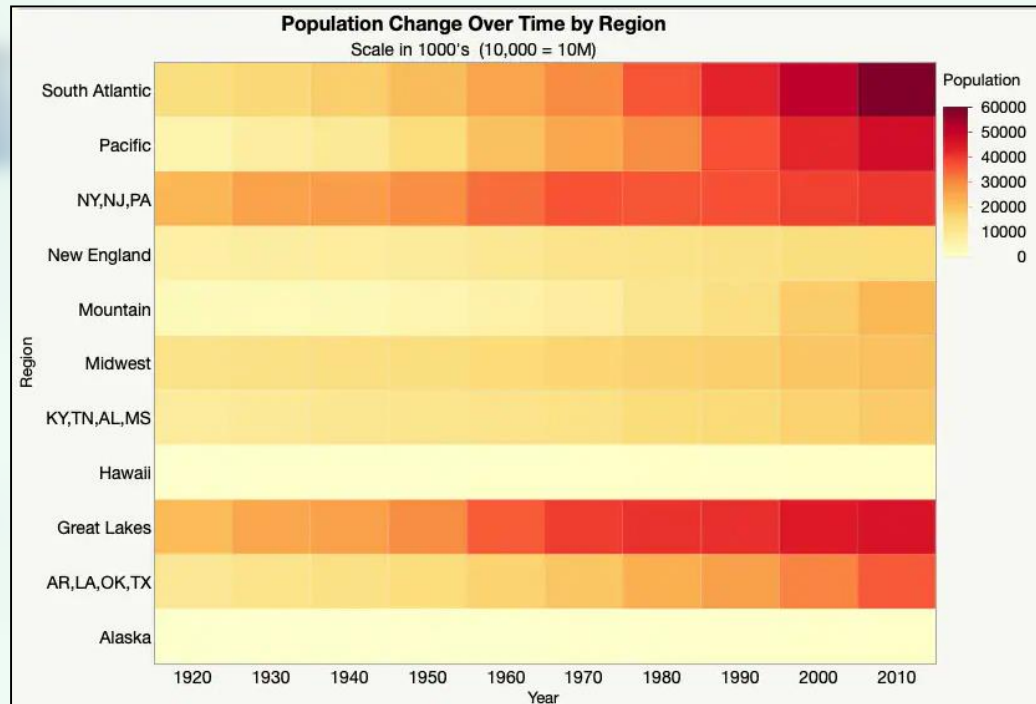
- ▶ 막대그래프를 띄워서 표현한 그래프로 순차별 데이터의 수치를 확인할 수 있다.
- ▶ 범주 내 데이터의 중복값 및 이상값 파악이 가능하다.



4. 분석결과 해석 및 활용 - 분석결과 시각화

② 히트맵(Heat Map)

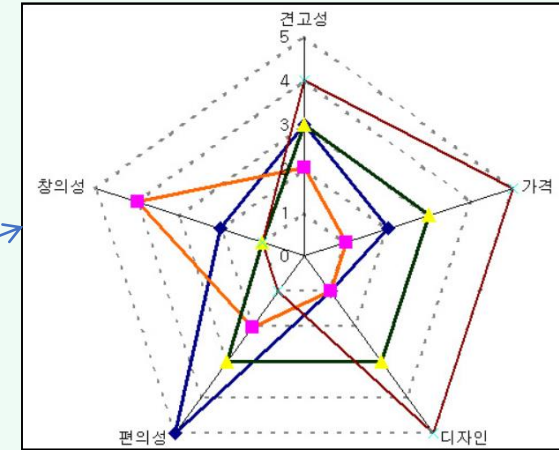
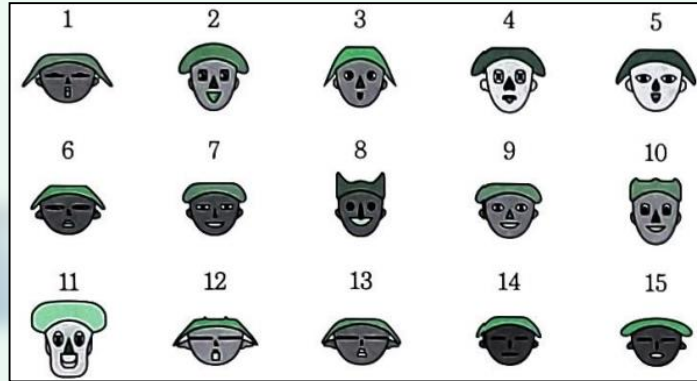
- ▶ 색상으로 표현할 수 있는 다양한 정보를 일정한 이미지 위에 열분포 형태로 표현한 그래프이다.
- ▶ 각 칸별 색상은 데이터 값의 크기를 나타내고, 색상이 짙을수록 데이터 값이 큰 것을 의미한다.
- ▶ 주로 웹사이트 방문자 분석 등 여러 방면에서 활용된다.



4. 분석결과 해석 및 활용 – 분석결과 시각화

③ 체르노프페이스(Chernoff Faces)

- ▶ 측정된 데이터를 눈, 코, 입 등 사람의 얼굴처럼 표현하는 시각화 방법이다.



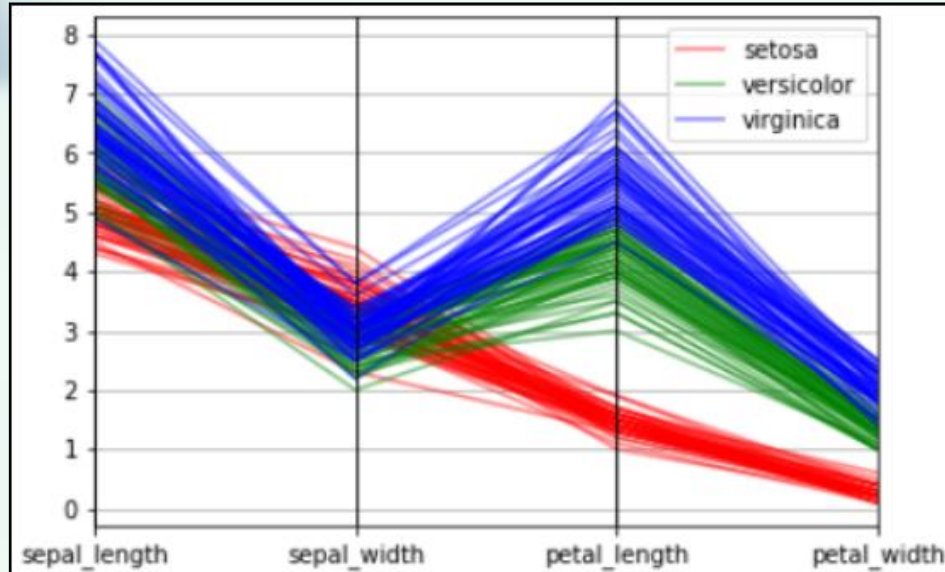
④ 스타차트(거미줄차트, 방사형차트, 레이더차트, Star Chart)

- ▶ 하나의 공간에 각각의 변수(3개 이상)를 표현하는 몇 개의 축을 그리고, 축에 표시된 해당 변수의 값을 연결하여 별 모양으로 표현한 그래프이다.
- ▶ 최솟값은 중심점이며, 최댓값은 가장 먼 끝점이 된다.
- ▶ 변수 사이의 유사점, 차이점, 이상값을 확인할 수 있다.

4. 분석결과 해석 및 활용 – 분석결과 시각화

⑤ 평행좌표그래프(Parallel Coordinates Graph)

- ▶ 여러 컬럼들을 비교하여 유사성을 찾기 위해 사용되는 그래프로 x 축은 컬럼을 나타내고, y 축은 컬럼들의 값이 표현된다.
- ▶ 유형 및 크기가 전혀 다른 데이터 값을 동일한 데이터 구간으로 표시하여 비교하는 그래프이다.
- ▶ 측정된 데이터들을 정규화하여 백분율로 표현하고, 이를 하나의 그래프 화면에 표현한다.



4. 분석결과 해석 및 활용 – 분석결과 시각화

04 인포그래픽(Infographics)

- 인포그래픽은 정보(Information)와 그래픽(Graphic)의 합성어로 정보 제공자가 전달하고자 하는 주요한 정보를 하나의 그래픽으로 표현하여 보는 사람들이 쉽고 빠르게 정보를 이해할 수 있도록 만든 시각화 방법이다.
- 인포그래픽의 목적은 정보형 메시지와 설득형 메시지를 담는 것이다.
- 정보형 메시지는 전달하는 데이터에 정보가 담겨야 하는 것을 의미하고, 설득형 메시지는 정보 제공자가 주장하고자 하는 내용을 담는 것을 의미한다.
- 인포그래픽의 종류는 지도형, 도표형, 타임라인형, 스토리텔링형, 만화형, 비교 분석형이 있다.

종류	설명	사용 예시
지도형	특정 국가 혹은 지역의 지도를 바탕으로 정보를 표현하는 방식	→ 지역별 투표율
도표형	다양한 도표와 그래프를 사용하여 정보를 표현하는 방식	→ 신규 서비스 사용자 가입률과 기업 인지도 변화
타임라인형	특정 주제와 관련된 히스토리를 타임라인(시간 순서) 형식으로 표현하는 방식	→ 기업의 발전 과정
스토리텔링형	특정 사건 혹은 주제에 대한 정보를 이야기를 들려주듯이 표현하는 방식	→ 추석 차례상 차리는 방법
만화형	캐릭터 또는 애니메이션 작업을 통해 정보를 표현하는 방식	→ 올바른 손씻기 방법
비교분석형	두 가지 이상의 비교 집단에 대한 정보를 비교하여 표현하는 방식	→ 가전 업체 A, B 기업의 노트북 비교

4. 분석결과 해석 및 활용 - 분석결과 시각화

개념 체크

01 다음 중 유형 및 크기가 전혀 다른 데이터 값을 동일한 데이터 구간으로 표시하여 비교하는 그래프로 측정된 데이터들을 정규화하여 백분율로 표현하고, 이를 하나의 그래프 화면에 표현하는 그래프의 명칭은?

- ① 스타차트 ② 히트맵
③ 체르노프페이스 ④ 평행 좌표 그래프

스타차트(거미줄차트, 방사형차트, 레이더차트, Star Chart)

- 하나의 공간에 각각의 변수(3개 이상)를 표현하는 몇 개의 축을 그리고, 축에 표시된 해당 변수의 값을 연결하여 별 모양으로 표현한 그래프이다.
- 최솟값은 중심점이며, 최댓값은 가장 먼 끝점이 된다.
- 변수 사이의 유사점, 차이점, 이상값을 확인할 수 있다.

히트맵(Heat Map)

- 색상으로 표현할 수 있는 다양한 정보를 일정한 이미지 위에 열분포 형태로 표현한 그래프이다.
- 각 칸별 색상은 데이터 값의 크기를 나타내고, 색상이 짙을수록 데이터 값이 큰 것을 의미한다.

02 다음 중 시각화 유형 및 그래프가 잘못 짝지어진 것은?

- ① 시간 시각화 - 추세선
- ② 공간 시각화 - 등치선도
- ③ 분포 시각화 - 카토그램
- ④ 비교 시각화 - 평행좌표그래프

카토그램은 공간 시각화에 속한다.

공간 시각화의 유형

등치지역도, 등치선도, 도트맵, 버블맵, 카토그램

분포 시각화의 유형

파이차트, 도넛차트, 트리맵, 누적영역그래프

관계 시각화의 유형

산점도, 산점도 행렬, 버블차트, 히스토그램, 네트워크 그래프

비교 시각화의 유형

플로팅 바 차트, 히트맵, 체르노프페이스, 스타차트, 평행좌표 그래프

4. 분석결과 해석 및 활용 – 분석결과 시각화

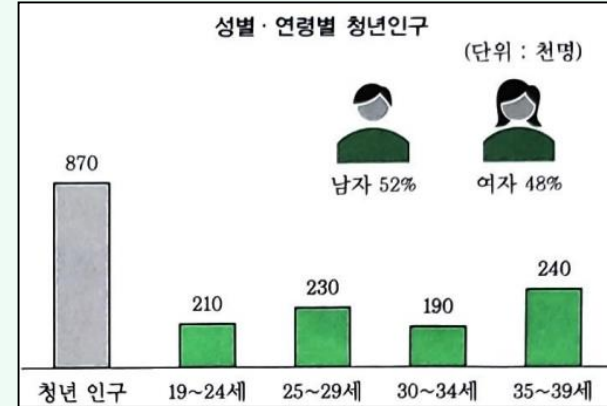
03 다음 중 하나의 공간에 각각의 변수를 표현하는 몇 개의 축을 그리고, 축에 표시된 해당 변수의 값을 연결하여 표현한 그래프의 올바른 명칭은?

- ① 버블차트 ② 체르노프페이스
- ③ 스타차트 ④ 히트맵

버블차트

● 산점도는 데이터를 점으로 표현하지만, 버블차트는 데이터의 크기를 추가적으로 고려하여 표현한 그래프이다.

04 다음과 같은 형태의 시각화 유형에 해당하는 것은?



- ① 플로팅 바 차트 ② 히스토그램
- ③ 막대그래프 ④ 인포그래픽

주어진 그림은 정보와 그래픽을 함께 담고 있는 인포그래픽 (Infographics)이다.

인포그래픽(Infographics)

- 인포그래픽은 정보(Information)와 그래픽(Graphic)의 합성어로 정보 제공자가 전달하고자 하는 주요한 정보를 하나의 그래픽으로 표현하여 보는 사람들이 쉽고 빠르게 정보를 이해할 수 있도록 만든 시각화 방법이다.
- 인포그래픽의 목적은 정보형 메시지와 설득형 메시지를 담는 것이다.

4. 분석결과 해석 및 활용 – 분석결과 시각화

05 다음 중 산점도와 비슷한 시각화 유형은?

- ① 버블차트 ② 파이차트
- ③ 히트맵 ④ 등치선도

06 다음 중 인포그래픽에 대한 설명으로 옳지 않은 것은?

- ① 빅데이터는 많은 양의 데이터를 표현해야 하므로 결과가 복잡하고 이해가 어려울 수 있다.
- ② 다양한 정보를 그래픽화하여 나타낸 시각화 방법이다.
- ③ 정보를 온라인상에 쉽고 빠르게 공유할 수 있다.
- ④ 도표, 글에 비해 정보 이용자가 쉽게 이해할 수 있고, 오래 기억할 수 있다.

인포그래픽은 복잡한 데이터를 정보와 그래픽을 활용하여 간단하게 표현하는 시각화 기법이다.

4. 분석결과 해석 및 활용 – 분석결과 활용

- 데이터 분석이 완료되면 주어진 절차에 맞게 분석된 결과를 활용하게 된다.

[분석 결과 활용 단계]

분석 모형 전개 → 분석 결과 활용 시나리오 개발 → 분석 모형 모니터링 → 분석 모형 리모델링

01 분석 모형 전개(Deployment)

- 분석 모형 전개는 분석한 데이터를 기반으로 각 데이터가 필요한 과정별로 데이터를 배치하는 것을 의미한다.
- 분석 모형 개발 및 운영 프로세스는 분석 목적 정의, 가설 검토, 데이터 준비 및 처리, 모델링 및 분석, 정확도 및 성능평가, 운영 순이다.

과정	설명
분석 목적 정의	분석을 위해 정확한 분석 목적이 무엇인지 명확히 하는 단계
가설 검토	분석 목적을 정의한 후 문제 해결을 위한 가설 검토 및 수립
데이터 준비 및 처리	분석에 대상이 되는 데이터를 수집하고, 분석을 위한 전처리 진행
모델링 및 분석	분석 목적에 맞는 모델 선택 및 분석 모형 도출
정확도 및 성능평가	도출된 분석 모형을 바탕으로 분석 모형의 정확도 및 성능을 평가
운영	분석 목적에 맞게 분석 모형 활용 및 운영

4. 분석결과 해석 및 활용 – 분석결과 활용

02 분석 결과활용 시나리오 개발

1) 분석 결과 시나리오 개발

; 분석 결과 시나리오 개발은 분석된 데이터 결과를 기반으로 인사이트를 도출하고, 의사결정을 할 수 있는 방법을 개발하는 것이다.

① 인사이트 발굴 및 확장

▶ 분석을 통해 인사이트를 발굴하고 이를 확장시켜 적용할 수 있는 범위를 확인한다.

② 의사결정 방법 선택

▶ 도출된 인사이트를 적용할 시점 및 방법을 검토하여 의사결정 방법을 선택한다.

③ 시각화 도구 선택

▶ 분석 결과를 효율적으로 시각화하여 표현할 수 있는 적합한 시각화 도구를 선택한다.

▶ R.Studio에서 제공하는 샤이니(Shiny)를 활용하여 모델링 결과를 배포할 수 있다.

데이터 인사이트는 데이터에 있는 패턴 및 통계를 깊이 있게 분석하고 얻을 수 있다. 조직은 이러한 패턴을 해석하여 인사이트를 고객의 요구 사항 예측에 적용하고 의사 결정의 효율성을 높일 수 있다.

샤이니(Shiny) : shiny는 R로 웹 어플리케이션을 만들 수 있게 해주는 패키지이다.

4. 분석결과 해석 및 활용 – 분석결과 활용

2) 분석 결과 활용 시 고려사항

- 분석 모형 최종 평가 시에는 데이터 학습 때 사용되지 않았던 데이터를 사용한다.
- 평가 지표를 활용하여 분석 모형의 성능을 확인한다.
- 분석 모형 개발과 지속적인 피드백으로 분석 모형의 성능을 향상시킨다.

4. 분석결과 해석 및 활용 – 분석결과 활용

03 분석 모형 모니터링

- 분석 모형 결과는 담당자의 의사결정 자료로 활용되므로 주기적으로 분석 모형을 모니터링하며 모형의 성능을 유지시켜야 한다.
- 분석 모형의 성능지표는 일별, 주별 등 일정 주기로 계산해서 DB에 저장한다.
- 분석 모형 성능은 측정 지표는 모형의 유형(회귀, 예측)에 따라 다르다.
- 분석 모형이 시스템에 적용된 이후에도 주기적인 모니터링을 통해 추가적인 학습이 이루어진다. 최신 데이터를 활용한 업데이트 혹은 모델 성능이 떨어졌을 때 새로운 모델을 구축하는 방법이 이에 해당한다.

4. 분석결과 해석 및 활용 – 분석결과 활용

04 분석 모형 리모델링

- 분석 모형을 모니터링 한 결과 분석 모형의 성능이 크게 떨어지거나 성능의 저하가 지속되는 경우에는 분석 모형 리모델링을 수행한다.

[분석 모형 리모델링 절차]

기존 분석 모형 분석 -> 데이터 수집 및 전처리 -> 분석 모형 구축 -> 최종 분석 모형 선정 및 반영

① 기존 분석 모형 분석

- ▶ 모형에 적합한 평가 지표를 활용하여 기존 모형의 성능을 분석한다.

② 데이터 수집 및 전처리

- ▶ 기존 분석 모형의 내용을 참고하여 추가적으로 데이터를 수집하고, 전처리 한다.

③ 분석 모형 구축

- ▶ 수집된 데이터를 기반으로 분석 모형을 새롭게 구축한다.

④ 최종 모형 선정 및 반영

- ▶ 구축된 모형을 기반으로 최종 모형을 선정하고, 이를 실제 실무 운영에 반영한다.

4. 분석결과 해석 및 활용 – 분석결과 활용

개념 체크

01 다음 중 분석 모형 개발 및 운영 프로세스에 대한 설명으로 옳지 않은 것은?

- ① 분석 목적 정의는 분석 목적에 맞는 모델 선택 및 분석 모형을 도출하는 단계이다.
- ② 가설 검토는 분석 목적을 정의한 후 문제 해결을 위한 가설을 검토하고 수립하는 단계이다.
- ③ 데이터 준비 및 처리는 분석에 대상이 되는 데이터를 수집하고, 분석을 위한 전처리를 진행하는 단계이다.
- ④ 정확도 및 성능평가는 도출된 분석 모형을 바탕으로 분석 모형의 정확도 및 성능을 평가하는 단계이다.

분석 모형 개발 및 운영 프로세스에서 분석 목적 정의는 분석을 위해 정확한 분석 목적이 무엇인지 명확히 하는 단계이다. 분석 목적에 맞는 모델 선택 및 분석 모형을 도출하는 단계는 모델링 및 분석에 대한 설명이다.

분석 모형 개발 및 운영 프로세스

1. 분석 목적 정의 : 분석을 위해 정확한 분석 목적이 무엇인지 명확히 하는 단계

02 다음 중 분석 모형 리모델링에 대한 설명으로 옳지 않은 것은?

- ① 분석 모형의 성능이 크게 떨어지거나 성능의 저하가 지속되는 경우에는 분석 모형 리모델링을 수행한다.
- ② 분석 모형 리모델링 절차는 기존 분석 모형 분석, 데이터 수집 및 전처리, 분석 모형 구축, 최종 분석 모형 선정 및 반영 순이다.
- ③ 기존 분석 모형 분석 단계에서 모형에 적합한 평가 지표 를 활용하여 기존 모형의 성능을 분석한다.

- ④ 데이터 수집 및 전처리 단계에서 분석 모형을 새롭게 구축한다.

데이터 수집 및 전처리 단계에서는 기존 분석 모형의 내용을 참고하여 추가적으로 데이터를 수집하고 전처리 한다. 분석 모형을 새롭게 구축하는 것은 분석 모형 구축 단계에서 진행된다.

분석 모형 리모델링

- 분석 모형을 모니터링 한 결과 분석 모형의 성능이 크게 떨어지거나 성능의 저하가 지속되는 경우에는 분석 모형

4. 분석결과 해석 및 활용 – 분석결과 활용

03 다음 중 분석 모형 최종 평가 시에 사용 가능한 데이터 는?

- ① 데이터 학습에 한 번 사용된 데이터
- ② 데이터 학습에 사용된 적 없는 데이터
- ③ 데이터 검증에 여러 번 사용된 데이터
- ④ 데이터 검증에 한 번 사용된 데이터

분석 모형 최종 평가 시에는 데이터 학습 때 사용되지 않았던 데이터를 사용한다.

04 다음 중 분석 모형 개발 및 운영 프로세스에 포함되지 않는 과정은?

- ① 모델 리모델링
- ② 분석 목적 정의
- ③ 가설 검토
- ④ 모델링 및 분석

분석 모형 개발 및 운영 프로세스 과정은 분석 목적 정의, 가설 검토, 데이터 준비 및 처리, 모델링 및 분석, 정확도 및 성능 평가, 운영 순이다.

4. 분석결과 해석 및 활용 – 분석결과 활용

05 다음 분석 모형 리모델링 과정 중 모형에 적합한 평가
지표를 활용하여 기존 모형의 성능을 분석하는 과정은?

- ① 데이터 수집 및 전처리
- ② 기존 분석 모형 분석
- ③ 분석 모형 구축
- ④ 최종 모형 선정 및 반영

4. 분석결과 해석 및 활용 – 예상 문제

예상 문제

1. 다음 중 데이터 시각화에 대한 설명으로 틀린 것은?

- ① 데이터 시각화는 주어진 데이터에 대한 이해를 돕기 위한 방법 중 하나로 데이터의 특성을 묘사하고 표현하는 방법이다.
- ② 데이터 시각화는 그림, 색상, 그래프, 도형 등 다양한 그래픽 요소를 활용하여 데이터를 설명하는 방법이다.
- ③ 데이터 시각화의 목적은 정보전달과 설득에 있다.
- ④ 데이터 시각화의 유형에는 시간, 공간, 분포, 관계, 비율 시각화가 있다.

데이터 시각화 유형에는 시간, 공간, 분포, 관계, 비교 시각화가 있다. 비율 시각화는 없다.

2. 데이터 시각화 절차에서 시각표현에 속하지 않는 것은?

- ① 여러 변수 비교
- ② 그래픽 7요소
- ③ 인터랙션
- ④ 그래픽 디자인 기본 원리

데이터 시각화 절차에서 시각표현에 속하는 것은 그래픽 요소, 그래픽 디자인 기본 원리, 인터랙션, 시각정보 디자인

3. 다음 중 데이터 시각화 도구와 이에 대한 설명이 옳지 않은 것은?

- ① 파인 리포트는 간단한 드래그 앤 드롭 방식으로 보고서를 쉽게 만들 수 있고, 데이터 의사결정 분석 시스템을 구축할 수 있는 도구이다.
- ② 구글 데이터 스튜디오는 코딩 없이 스프레드시트와 데이터베이스를 활용하여 쉽게 데이터를 시각화하는 도구이다.
- ③ 태블로는 클라우드 기반으로 서비스되며, 차트, 그래프 등 다양한 그래픽 기능을 제공하는 시각화 도구이다.
- ④ 파이썬은 다양한 그래프 라이브러리(library)를 활용한 프로그래밍을 통해 구현할 수 있는 시각화 도구이다.

데이터 시각화 도구 6가지

- 1. 파인 리포트(Fine Report) : 간단한 드래그 앤 드롭 방식으로 보고서를 쉽게 만들 수 있고, 데이터 의사결정 분석 시스템을 구축할 수 있는 도구이다.
- 2. 차트 블록(Chart Blocks) : 코딩 없이 스프레드시트와 데이터베이스를 활용하여 쉽게 데이터를 시각화하는 도구

4. 분석결과 해석 및 활용 – 예상 문제

5. 다음 중 데이터 시각화 절차로 옳은 것은?

- ① 구조화 → 구체화 → 시각화
- ② 구조화 → 시각화 → 시각표현
- ③ 구체화 → 시각표현 → 시각화
- ④ 구체화 → 시각화 → 시각표현

6. 다음 중 데이터 시각화 분류에 대한 설명으로 옳지 않은 것은?

- ① 정보 시각화는 대규모 데이터를 색상, 도표, 이미지 등을 활용하여 요약적으로 표현하는 방법이다.
- ② 데이터 시각화는 데이터 분석 결과를 쉽게 이해할 수 있도록 시각적으로 표현하는 방법이다.
- ③ 정보 디자인은 정보를 제공하는 대상이 정보를 명확하게 파악하기 위해 시각화하는 방법이다.
- ④ 인포그래픽은 정보를 빠르고 분명하게 표현하기 위해 정보, 자료, 지식을 그래픽을 활용하여 시각적으로 표현 하는 방법이다.

정보 디자인은 정보를 확인하는 대상이 정보를 명확히 이해

7. 데이터 시각화 절차 중 시각화에 속하지 않는 작업은?

- ① 공공 시각화 ② 분포 시각화
- ③ 비교 시각화 ④ 시간 시각화

데이터 시각화 절차 중 시각화 작업에 속하는 작업은 시간 시각화, 분포 시각화, 관계 시각화, 비교 시각화, 공간 시각화, 여러 변수 비교가 있다.

8. 다음에 설명하는 알맞은 명칭은?

투자로 얻을 수 있는 순 효과를 총 비용으로 나눈 값을 나타낸다.

자본 투자에 따른 순 효과의 비율을 의미한다.

$$\text{계산식} : \frac{\text{순이익}}{\text{투자비용}} * 100$$

- ① TCO ② NPV
- ③ ROI ④ IRR

총소유 비용(TCO; Total Cost of Ownership)

● 하나의 장비를 획득할 때, 장비의 수명 주기 동안에 발생하게 되는 모든 연관 비용이다.

● 하나의 자산을 획득할 때, 주어진 기간 동안 모든 연관

4. 분석결과 해석 및 활용 - 예상 문제

9. 다음 중 비즈니스 기여도 평가 기법에 속하지 않는 것은?

- ① ROI ② PP
- ③ TCO ④ RR

PP(Payback Period) : 투자 회수 기간을 의미하는 비즈니스 기여도 평가 기법이다. 흑자 전환점을 의미한다.

RR(Relative Risk) : 상대위험도는 위험인자에 노출되었을 때, A 집단의 사건 발생 확률과 B 집단의 사건 발생 확률에 대한 확률이다.

10. 다음 중 시간 시각화 유형에 속하지 않는 것은?

- ① 막대그래프 ② 카토그램
- ③ 계단식그래프 ④ 영역차트

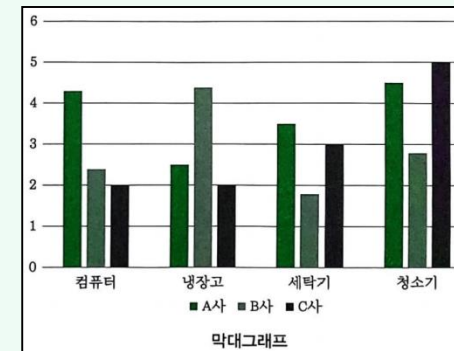
시간 시각화 유형에는 막대그래프, 누적막대그래프, 점그래프, 선그래프, 영역차트, 계단식그래프, 추세선이 있다. 카토그램은 공간 시각화 유형에 속한다.

11. 다음 중 공간 시각화 유형에 속하지 않는 것은?

- ① 등치지역도 ② 도트맵
- ③ 버블맵 ④ 파이차트

공간 시각화 유형에는 등치지역도, 등치선도, 도트맵, 버블맵, 카토그램이 있다. 파이차트는 분포 시각화 유형에 속한다.

12. 다음 그래프에 대한 설명으로 옳지 않은 것은?



- ① 하나의 막대로 데이터의 여러 범주별 비율을 확인할 수 있다.
- ② 직사각형 형태의 막대를 사용하고, 각 막대는 특정 범주를 나타낸다.
- ③ 막대는 서로 떨어져서 분포한다.
- ④ 그래프의 축이 반드시 수량이 아니어도 된다.

4. 분석결과 해석 및 활용 - 예상 문제

13. 다음에서 설명하는 시각화 도구는?

특정한 데이터 수치의 변화에 따라서 지도의 면적이 왜곡되는 그래프이다.

주로 의석 수나 선거인단 수, 인구 등의 데이터를 표현한다.

- ① 버블맵 ② 파이차트
- ③ 카토그램 ④ 버블차트

왜곡 이라는 단어가 나오면 공간 시각화 도구에 속하는 카토그램에 대한 설명인 것을 기억하자.

14. 다음 중 관계 시각화 도구에 속하는 것은?

㉠ 산점도	㉡ 히트맵
㉢ 버블차트	㉣ 스타차트
㉤ 평행좌표그래프	㉥ 산점도 행렬
㉦ 히스토그램	㉧ 누적막대그래프

- ① ㉠, ㉡, ㉢, ㉥ ② ㉡, ㉢, ㉦, ㉧
- ③ ㉠, ㉢, ㉥, ㉦ ④ ㉠, ㉢, ㉤, ㉦

관계 시각화의 유형

산점도, 산점도 행렬, 버블차트, 히스토그램, 네트워크 그래프

15. 다음 설명하는 시각화 도구는?

색상으로 표현할 수 있는 다양한 정보를 일정한 이미지 위에 열분포 형태로 표현한 그래프이다.

각 칸별 색상은 데이터 값의 크기를 나타내고, 색상이 짙을수록 데이터 값이 큰 것을 의미한다.

- ① 산점도 ② 카토그램
- ③ 히스토그램 ④ 히트맵

16. 다음 중 비교 시각화에 속하지 않는 것은?

- ① 플로팅 바 차트 ② 버블맵
- ③ 체르노프페이스 ④ 스타차트

버블맵은 공간 시각화 유형에 속한다.

4. 분석결과 해석 및 활용 – 예상 문제

17. 다음 중 인포그래픽에 대한 설명으로 옳지 않은 것은?

- ① 인포그래픽의 목적은 정보형 메시지와 공유형 메시지를 담는 것이다.
- ② 인포그래픽은 정보(Information)와 그래픽(Graphic)의 합성어이다.
- ③ 인포그래픽은 정보 제공자가 전달하고자 하는 주요한 정보를 하나의 그래픽으로 표현하여 보는 사람들이 쉽고 빠르게 정보를 이해할 수 있도록 만든 시각화 방법이다.
- ④ 정보형 메시지는 전달하는 데이터에 정보가 담겨야 하는 것을 의미한다.

인포그래픽(Infographics)

- 인포그래픽은 정보(Information)와 그래픽(Graphic)의 합성어로 정보 제공자가 전달하고자 하는 주요한 정보를 하나의 그래픽으로 표현하여 보는 사람들이 쉽고 빠르게 정보를 이해할 수 있도록 만든 시각화 방법이다.
- 인포그래픽의 목적은 정보형 메시지와 설득형 메시지를 담는 것이다.
- 정보형 메시지는 전달하는 데이터의 정보가 담겨야 하는

19. 다음 중 분석 결과 활용 시 고려사항이 아닌 것은?

- ① 분석 모형 최종 평가 시에는 데이터 학습 때 사용되지 않았던 데이터를 사용한다.
- ② 분석 모형 평가 시에는 모형의 유형에 관계없이 동일한 평가지표를 활용한다.
- ③ 평가지표를 활용하여 분석 모형의 성능을 확인한다.
- ④ 분석 모형 개발과 지속적인 피드백으로 분석 모형의 성능을 향상시킨다.

분석 모형 평가 시에는 모형의 유형에 따라 각기 다른 평가 지표를 사용한다.

분석 결과 활용 시 고려사항

- 분석 모형 최종 평가 시에는 데이터 학습 때 사용되지 않았던 데이터를 사용한다.
- 평가 지표를 활용하여 분석 모형의 성능을 확인한다.
- 분석 모형 개발과 지속적인 피드백으로 분석 모형의 성능을 향상시킨다.

20. 다음 중 분석 모형 모니터링에 대한 설명으로 옳지 않은 것은?

4. 분석결과 해석 및 활용 – 마무리 문제

1. 다음 중 회귀 모형 평가 지표에 대한 수식이 잘못된 것은?

① MAE : $\frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$

② MSE : $\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$

③ RMSE : $\sum_{i=1}^n (y_i - \bar{y}_i)^2$

④ MAPE : $\frac{100}{n} * \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|$

RMSE(평균제곱근오차, Root Mean Squared of Error)

수식 $\sum_{i=1}^n (y_i - \bar{y}_i)^2$ 는 전체 제곱합(SST)에 대한 수식이다.

● 평균제곱오차(MSE; Mean Squared Error) 제곱근을 씌운 값이다.

● MSE는 값이 커지는 경향이 있으므로 제곱근을 씌운 RMSE를 실무에서 일반적으로 사용한다.

● 수식 : $\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$

회귀 모형 평가 지표

1. 평균절대오차(MAE; Mean Absolute Error)

● 모형의 실제값과 예측값 차이에 절댓값을 취하여 그 합을

나눈 값이 평균절대오차이다.

3. 다음과 같은 결정계수 데이터가 나왔을 때, 가장 정확도가 높은 모형은?

A 모형 : 0.75

B 모형 : 0.88

C 모형 : 0.95

D 모형 : 0.99

① A 모형 ② B 모형

③ C 모형 ④ D 모형

결정계수는 0~1 사이의 범위를 갖고, 데이터가 1에 가까울수록 모형의 정확도가 높다고 할 수 있다.

4. 다음 중 결정계수(R^2) 연산에 포함되지 않는 것은?

① SST ② SSE

③ SSR ④ STS

결정계수 연산에는 SST(전체 제곱합), SSR(회귀 제곱합), SSE(오차 제곱합)이 사용된다.

4. 분석결과 해석 및 활용 - 마무리 문제

5. 다음과 같은 혼동행렬에서 빈칸에 알맞은 명칭은?

		예측 범주값	
		Predicted Positive	Predicted Negative
실제 범주값	Actual Positive	㉠	㉡
	Actual Negative	㉢	㉣

㉠

㉡

㉢

㉣

- ① TP FN FP TN
- ② TP FP TN FN
- ③ FP TP FN TN
- ④ FP TN TP FN

㉠은 True Positive(TP), ㉡은 False Negative(FN)

㉢은 False Positive(FP), ㉣은 True Negative(TN)

6. 다음 중 혼동행렬을 활용한 모형의 평가 지표 연산이 잘못된 것은?

① Recall : $\frac{TP}{TP+FN}$

② Specificity : $\frac{FP}{TN+FP}$

다음 혼동행렬을 보고 물음에 답하시오.(7~9)

실제값 \ 예측값	Positive	Negative
Positive	30	40
Negative	70	10

7. 주어진 혼동행렬에서 정밀도는 얼마인가?

- ① 5/10 ② 3/10 ③ 3/7 ④ 4/5

정밀도(Precision)는 예측 긍정 범주 중 긍정의 비율을

의미한다. 공식은 $\frac{TP}{TP+FP}$ 이며, 따라서 주어진 혼동행렬에서

정밀도를 구하고자 할 경우 $\frac{30}{30+70} = \frac{30}{100} = \frac{3}{10}$ 이 된다.

8. 주어진 혼동행렬에서 오차비율은 얼마인가?

- ① 4/11 ② 7/10 ③ 11/15 ④ 4/15

오차 비율(Error Rate)은 전체 범주 중 잘못 예측한 비율을

의미한다. 공식은 $\frac{FP+FN}{TP+TN+FP+FN} = \frac{70+40}{30+70+40+10} = \frac{110}{150} =$

$\frac{11}{15}$ 이 된다.

4. 분석결과 해석 및 활용 – 마무리 문제

9. 주어진 혼동행렬에서 F1-Score는 얼마인가?

- ① 6/17 ② 3/11 ③ 3/5 ④ 4/7

F1-Score의 공식은 $2 * \frac{Precision(정밀도) * Recall(재현율)}{Precision(정밀도) + Recall(재현율)}$ 과 같다.

따라서 정밀도와 재현율을 먼저 연산한다. 7번 문제 에서 정밀도는 3/10으로 확인되었으니 재현율(Recall)만 추가적으로 연산을 하면 된다. 재현율은 참 긍정률(TP Rate) = 재현율(Recall) = 민감도(Sensitivity) 같으며, 의미는 실제 긍정 범주 중 긍정의 비율이다.

수식은 $\frac{TP}{TP+FN}$ 가 되기 때문에 $\frac{30}{30+40} = \frac{30}{70} = \frac{3}{7}$ 이 된다.

이를 F1-Score 공식에 대입하면 $2 * \left(\frac{\frac{3}{10} * \frac{3}{7}}{\frac{3}{10} + \frac{3}{7}} \right) = 2 * \left(\frac{\frac{9}{70}}{\frac{51}{70}} \right) = 2 * \left(\frac{9}{51} \right)$

$= 2 * \left(\frac{630}{3570} \right) = 2 * \left(\frac{3}{17} \right) = \left(\frac{6}{17} \right)$ 이 된다.

10. 다음 중 ROC 곡선에 대한 설명으로 옳지 않은 것은?

- ① ROC 곡선은 가로축(x)을 혼동행렬의 거짓 긍정률(FP Rate)로 두고, 세로축(y)을 참 긍정률(TP Rate)로 두어

11. 다음과 같은 특징을 갖는 검증 방법은?

데이터를 K개로 나누어 (K-1)개는 학습에, 나머지 하나는 검증에 사용하는 방법이다.

검증 절차 : 테스트 데이터를 제외한 데이터를 무작위로 중복되지 않는 K개의 데이터로 분할 → K-1개의 데이터를 학습 데이터로 사용하고, 나머지 1개 데이터를 검증 데이터로 사용 검증 데이터를 바꾸며 K번 반복해서 분할된 데이터가 한 번씩 검증 데이터로 사용한다.

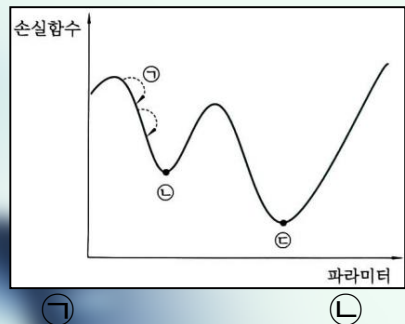
- ① LOOCV
② K-means clustering
③ K-fold cross validation
④ Bootstrap

LOOCV(Leave-One-Out-Cross Validation)

- N개의 데이터 중 1개만 평가 데이터로 사용하고, 나머지 N-1개는 훈련 데이터로 사용하는 과정을 N번 반복하는 방법이다.
- 많은 데이터를 훈련 데이터로 활용할 수 있지만 계산량

4. 분석결과 해석 및 활용 – 마무리 문제

13. 다음과 같은 경사하강법 그래프에서 빈 칸의 알맞은 명칭은?



- | | | |
|--------------|---------------|---------------|
| ① 지역 최솟값 | 전역 최솟값 | 학습률 |
| ② 학습률 | 전역 최솟값 | 지역 최솟값 |
| ③ 전역 최솟값 | 학습률 | 지역 최솟값 |
| ④ 학습률 | 지역 최솟값 | 전역 최솟값 |

14. 다음 설명하는 매개변수 최적화 기법은?

AdaGrad에서 최적값에 도달하기 전에 학습률이 0에 가까워지는 상황을 방지하기 위해 개선된 방법이다.

- ① Momentum
- ② Adam
- ③ **RMSProp**
- ④ Nesterov momentum

15. 데이터 시각화 절차에서 시각표현에 속하지 않는 것은?

- ① **여러 변수 비교**
- ② 그래픽 7요소
- ③ 인터랙션
- ④ 그래픽 디자인 기본 원리

데이터 시각화 절차에서 시각표현에 속하는 것은 그래픽 7요소, 그래픽 디자인 기본 원리, 인터랙션, 시각정보 디자인 7원칙이다. 여러 변수 비교는 시각화 단계에 포함 된다.

16. 다음 설명하는 비즈니스 평가 기법은?

어떤 사업에 대해 사업기간 동안의 현금 수익 흐름을 현재 가치로 환산하여 합한 값이 투자지출과 같아지도록 할인하는 할인율이다.

순 현재 가치를 '0'으로 만드는 할인율을 의미한다.

- | | |
|--------------|-------|
| ① IRR | ② ROI |
| ③ NPV | ④ TCO |

투자 대비 효과(ROI; Return On Investment)

- 투자로 얻을 수 있는 순 효과를 총 비용으로 나눈 값이다.
- 자본 투자에 따른 순 효과의 비율(투자 타당성)이다.

4. 분석결과 해석 및 활용 – 마무리 문제

17. 다음 중 비즈니스 평가 기법에 대한 설명으로 옳지 않은 것은?

① 비즈니스 기여도 평가는 빅데이터 분석 모형을 실제 업무에 적용할 경우 비즈니스 상에 어떠한 개선점이 있는지 정량적으로 확인하기 위한 방법이다.

② TCO는 투자로 얻을 수 있는 순 효과를 총 비용으로 나눈 값을 의미한다.

③ 비즈니스 기여도 평가 기법에는 TCO, ROI, NPV, IRR, PP 등이 있다.

④ PP는 누적투자금액과 매출금액의 합이 같아지는 기간을 의미한다.

TCO는 총소유 비용이며, 투자로 얻을 수 있는 순 효과를 총비용으로 나눈 값은 투자 대비 효과(ROI)에 대한 설명이다.

18. 다음과 같은 그래프의 명칭은?



19. 다음의 특징을 갖는 시각화 도구는?

여러 계층 구조의 데이터를 표현하는 그래프이다.

서로 다른 크기의 사각형을 이용하여 데이터의 비율을 나타내고, 사각형을 겹쳐 배치함으로써 데이터의 대분류와 하위 분류를 구분한다.

① 스타차트 ② 트리맵

③ 버블맵 ④ 산점도

스타차트(거미줄차트, 방사형차트, 레이더차트, Star Chart)

● 하나의 공간에 각각의 변수(3개 이상)를 표현하는 몇 개의 축을 그리고, 축에 표시된 해당 변수의 값을 연결하여 별 모양으로 표현한 그래프이다.

● 최솟값은 중심점이고, 최댓값은 가장 끝점이 된다.

● 변수 사이의 유사점, 차이점, 이상값을 확인할 수 있다.

버블맵(버블플롯맵, Bubble Map)

● 측정된 데이터의 위치는 위도와 경도를 기반으로 표시하고, 크기는 원형의 사이즈로 구별하여 표시하는 그래프이다.

산점도(Scatter Plot)



감사합니다.