1장: 데이터 시각화의 기본

* 데이터를 시각화 하는 이유
  + 관계와 추세를 찾을 수 있다.
  + 기억하기 수월하다.
  + 대량의 데이터를 즉시 해석할 수 있다.
* 데이터 시각화 방법
  + 전주의적 속성: 어떤 것을 보자마자 뇌에서 바로 알아차릴 수 있도록 강조하기 위한 시각화 속성

|  |  |
| --- | --- |
| 색상 | 여러가지 색상을 사용하면 전주의적 속성이 무너진다. |
| 크기 |  |
| 길이 | 막대 차트를 활용하여 빈도를 규약화 할 수 있다. |

+ 방향, 너비, 형태, 곡률, 표시 추가, 둘러싸기, 색조, 위치, 공간적 그룹핑

* 데이터 규약화 방법을 제공한다.
  + 데이터 유형

|  |  |
| --- | --- |
| 범주형 데이터 | 사물  (e.g.) 이름, 직업, 캐릭터, 성별, 제품 카테고리, 도시 등 |
| 순서형 데이터 | 분명한 순서가 있는 데이터  각 값 사이의 간격은 일정하지 않다.  (e.g.) 출생일, 시즌, 계급, 학력, 만족도, 연봉 등 |
| 정량적 데이터 | 숫자로서 측정, 분류 가능한 데이터  (e.g.) 출생일, 키, 횟수, 매출, 이익, 점수, 페이지뷰, 환자 수 등  ① 이산 데이터: 정확한 포인트로 제시, 값 사이에 아무것도 X   1. 연속 데이터: 가능한 중간 값의 무한한 수, 값 사이 값을 허용 |

* + 차트의 데이터 규약화
* 색상

\*훌륭한 데이터 시각화는 색을 사용하지 않아도 정보성이 있다.

\*색상은 의도를 갖고 사용해야 한다.

* + 색상의 사용

|  |  |  |
| --- | --- | --- |
| 순차적 | -밝은 색부터 어두운 색까지 단색 사용 |  |
| 발산형 | -중간점에서 발산하는 범위를 표시  -상반되는 2가지 측정 범위  (e.g. 긍정/부정, 이익/손실) 표현  -날씨 표현 가능 |  |
| 범주형 | -카데고리마다 다양한 색조로 구별 |  |
| 강조 | -사용자의 눈에 띄어야 하지만  경고하거나 놀래킬 필요 X |  |
| 경고 | -사용자의 주목을 끌어야 할 때 사용 |  |

* + 색각 이상(색맹)
    - 제1색맹 = 적색맹
    - 제2색맹 = 녹색맹
    - 제3색맹 = 청황색맹 (청색맹) -> 아주 드물다.
  + 파랑-오렌지 팔레트를 사용한다. (색맹 친화적, 거의 모두가 구별 가능)
  + 적색과 녹색을 함께 사용하지 않는다. (신호등 조합 피하기)
  + 적색, 녹색, 갈색, 오렌지색을 함께 사용하지 않는다. (색맹에게 갈색으로 보임)
  + 청색, 보라를 함께 사용하지 않는다. (색맹에게 보라도 청색으로 보임)
  + 분홍/적색, 회색을 함께 사용하지 않는다.
  + 회색, 갈색을 함께 사용하지 않는다.
* 일반적인 차트 유형

|  |  |  |
| --- | --- | --- |
| 막대 차트 | -척도: 길이  -카테고리를 비교하는 가장 효과적인 방법  -최대/최소 항목을 찾을 때 사용  -> 정렬 필요 |  |
| 시계열  꺾은선  차트 | -x축: 시간, y축: 척도  -장기적 변화를 보일 때 사용  -선의 높이와 경사로 추세 확인 |  |
| 산포도 | -서로 다른 두 척도 비교 가능  -2가지 변수 간의 관계를 관찰 |  |
| 점 도표 | -2개 차원에 걸쳐 값 비교 가능 |  |
| 단계 구분도 | -미리 정의된 영역 내 음영, 색 차이로 값이나 카테고리를 표현 |  |
| 기호 지도 | -특정 장소의 값을 표현 |  |
| 표 | -정확한 데이터를 표시 |  |
| 하이라이트표 | -표를 색상으로 규약화  -> 정확한 값 + 시각적 |  |
| 불릿 그래프 | -실제 값 대비 타겟의 비교 |  |

* 우리의 시각 처리 체계가 배신할 때
  + 착시현상
  + 크기가 전주의적 속성인 반면에 사람들은 정확하게 크기의 차이를 구별하지 못한다.
  + 길이로는 차이를 명확히 알 수 있기 때문에 가장 효율적인 전주의적 속성 중 하나
  + 여러 전주의적 속성을 하나의 차트에 사용하면 읽기가 쉽지 않다.
  + 파이차트의 문제점: 각의 크기를 정확히 추산할 수 없다. (크기와 같은 이유)
* 모든 결정은 절충이다.
  + 선택한 차트가 현재 강조하려는 질문에 적절한 답을 제시하는지 측정할 필요

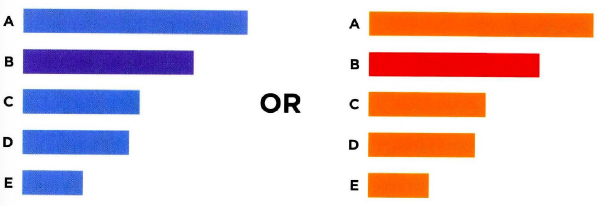
데이터 시각화를 위한 컬러 & 타이포

[1] 색의 3속성

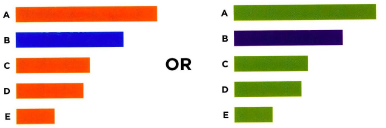
|  |  |
| --- | --- |
| 색상 |  |
| 명도 |  |
| 채도 |

[2] 색상 선택 시 주의점

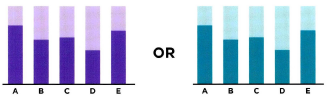
* 컬러 조합: 꼭 보여주고 싶은 부분을 색상으로 강조, 나머지 부분은 흑백 처리
* 유사 대비: 동일한 색에서 밝기를 달리하거나 색상환에서 가까운 색끼리 조합



* 보색 대비: 색상환에서 마주보고 있는 색 조합 (눈에 피로를 줄 수 있음 주의)



* 명도 대비: 대비가 심한 명도 피하기



* 명도와 채도를 이용한 컬러 표시
* 특정 주제가 연상되는 컬러 피하기
* 색의 의미

|  |  |
| --- | --- |
| 빨강 | 긍정: 적극, 열정, 힘, 운동감, 행복, 사랑, 욕망, 진취적, 외향적, 태양, 젊음, 유혹, 기쁨 |
| 부정: 마이너스, 죽음, 분노, 폭발, 흥분, 공격, 고통, 피, 혁명, 금지, 전쟁, 위험, 악마 |
| 주황 | 긍정: 활기참, 즐거움, 명랑, 친근, 따뜻, 태양, 에너지, 변화, 지혜, 존경, 사회적, 비타민, 변화 |
| 부정: 허풍, 소란스러움, 싼, 천박한, 사치, 사탄, 쉽게 변하는 |
| 노랑 | 긍정: 어린이, 행복, 즐거움, 생동감, 희망, 상상, 밝음, 지성, 귀족, 황금, 개나리, 유머 |
| 부정: 배반, 변덕, 경박, 질투, 충동적, 인색함 |
| 초록 | 긍정: 자연, 평화, 인내, 근면, 성실, 안정, 휴식, 식물, 건강, 행운, 인천, 희망, 생명, 신선, 중립, 사회적 |
| 부정: 질투, 독, 타락, 수동적, 모호한 |
| 파랑 | 긍정: 성실, 믿음, 성공, 하늘, 명상, 바다, 신선, 순수, 정의 신중, 미래, 진정 |
| 부정: 냉점, 독선적, 우울, 슬픔, 차가움, 하락, 보수, 권위, 의심 |
|  |  |
|  |
|  |  |
|  |
|  |  |
|  |
|  |  |
|  |
|  |  |
|  |
|  |  |
|  |

* 색상 계획: 주조색과 보조색을 선택하고 색상 팔레트를 만들어야 함
* 포인트 색상: 주조색의 보색 이용
* 회색의 사용: 충돌 방지 및 주조색 강조

[3] 서체

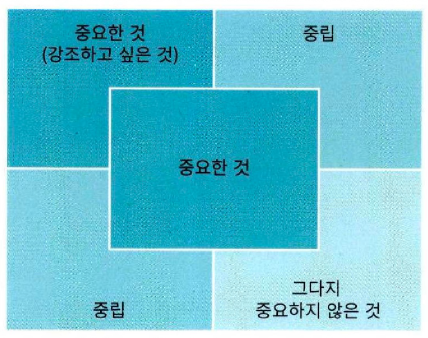
* 영문: 세리프 & 한글: 바탕체 유사
* 영문: 산세리프 & 한글: 돋움체 유사
* 탈네모틀 한글꼴은 네모틀에 비해 가독성이 떨어진다.
* 서체 사용의 단순화 (서체 패밀리: 같은 특징의 다른 두께 폰트 집합): 한 개의 폰트만으로 충분하다.
  + 유니버스 패밀리: 가장 않은 서체 패밀리를 가짐, 유니버스55가 본문에 가장 적합
  + 두께



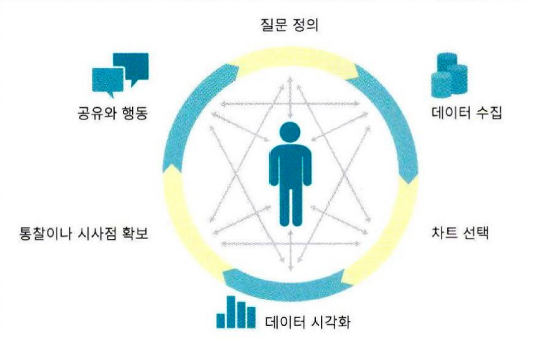
* 가독성이 좋은 폰트를 선택: 복잡한 형태 X, 필기체 X
* 대문자와 소문자를 함께 사용: 문장 제일 앞, 고유명사 제일 앞, 중요한 부분만 대문자로 표기
* 글자는 되도록 수평으로 배열
* 이탤릭(오블리크) 서체는 강조를 하는 곳에만 사용
* 글자의 두께는 두껍지도 가늘지도 않게 사용: 제목-두꺼운 글꼴, 본문: 중간 두께의 글꼴, 강조-두꺼운 글꼴
* 글자의 색은 검정색과 배경은 흰색으로 사용
* 글자의 너비(자폭) 조절: 10% 이하로 조절 해야함 (장체-자폭이 좁은 것, 평체-자폭이 넓은 것)
* 글자의 사이(자간) 조절: 한글의 자간은 살짝 줄이는 것 추천 (e.g. 윤고딕 자간: -20pt가 적정)
* 글줄의 사이(행간) 조절: 글자 크기 +3~5pt가 적당 (e.g. 윤고딕 8/12pt: 폰트 크기-8, 행간-12)
* 숫자를 표기할 때:
  + 소수점이 있는 경우: 소수점 정렬
  + 단위가 있는 경우: 우측 정렬 (단위 정렬)
  + 단위가 없는 경우: 가운데 정렬

데이터 시각화 디자인

* 대시보드 배치 전략

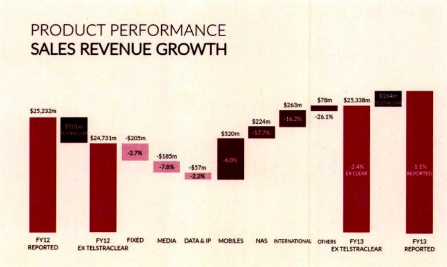
 

* 비주얼 애널리틱스 사이클

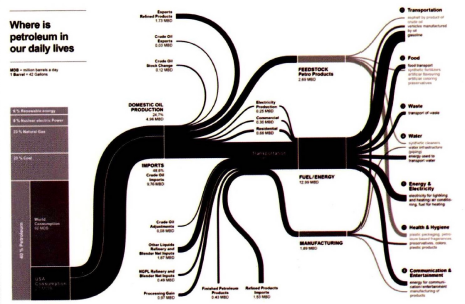


* 양 표시 그래프
  + 막대그래프
  + 롤리팝 차트
  + 도트 매트릭스 차트
  + 채워진 버블
* 비율 표시 그래프
  + 누적 막대그래프
  + 원그래프
  + 도넛 차트
  + 트리맵: 음수 표현 어려움, 이웃하지 않은 사각형 비교 어려움, 배치 제어 어려움
  + 와플 차트 (그리드 플롯)
  + 워터폴 차트: 기간이나 카테고리 간 증감을 표시

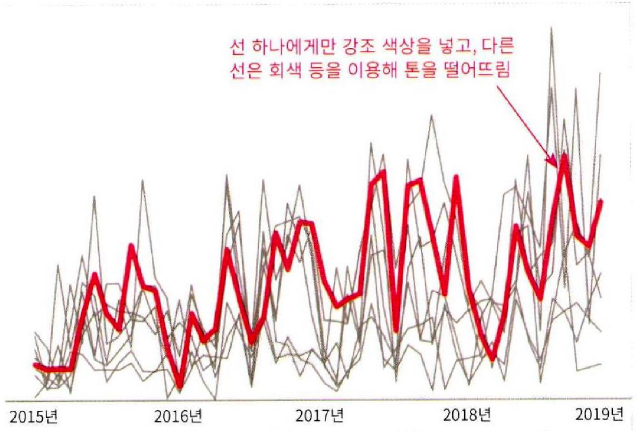
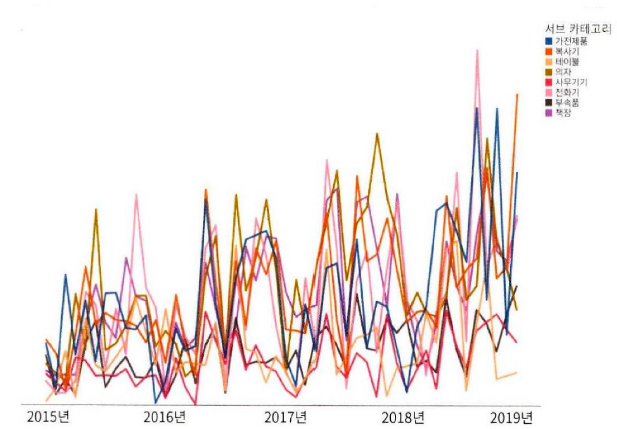
\*워터폴 차트



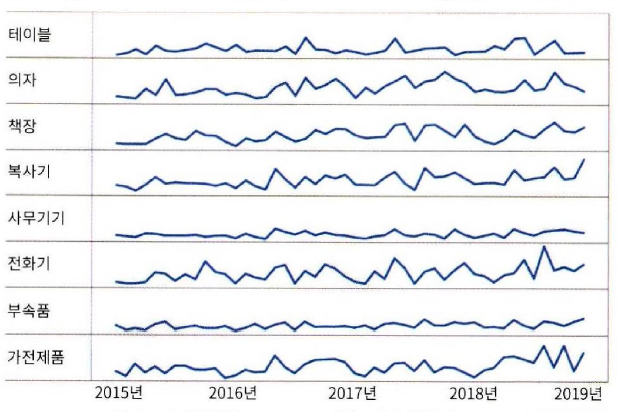
* 흐름 표시하기
  + 산키 다이어그램



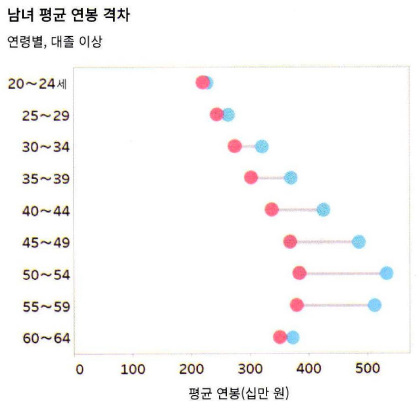
* 시간 추이 표시하기
  + 선 그래프
* 스파게티 차트



* 스파이크 차트



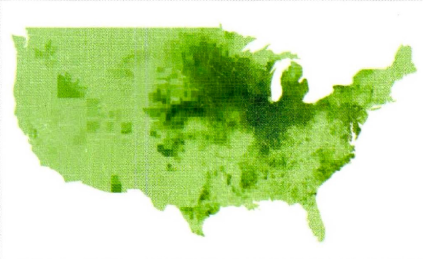
* 이중 축 사용
  + 기울기 차트: 처음과 끝만 표시
  + 히트맵
* 분포 표시하기
  + 히스토그램
  + 도트 플롯 (덤벨 차트)



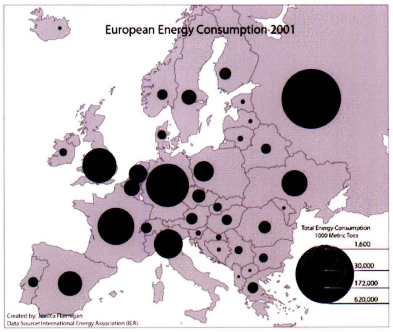
* + 박스 플롯
  + 버터플라이 차트



* 순위 표시하기
  + 범프 차트: 순위만 표시하여 연결한 차트 (실제 값 X)
* 관계성 표시하기
  + 산포도
  + 버블 차트
* 차이 표시하기
  + 분기한 막대그래프: 기준선이 존재하는 막대차트
  + 분기한 색상선 그래프: 기준선이 존재하는 영역차트
* 지리공간 표시하기
  + 등치맵



* + 심벌 크기 (기호 차트)



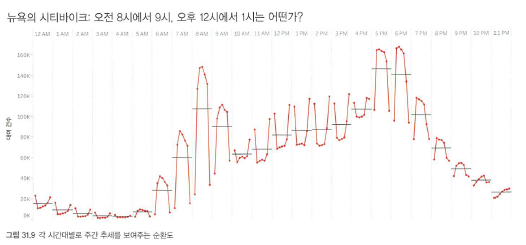
* + 지도 위에 점으로 표시
  + 지도 위에 흐름 표시

3장: 실무 적용 사례

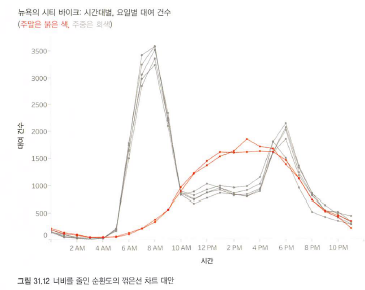
* 대시보드: 데이터와 사용자 간에 개인적인 관계를 만드는 것
* 버블차트를 피하라.
  + 정확한 정량 비교에 유용하지 않다.
  + 계층적 데이터 또는 카테고리가 너무 많은 경우에는 트리맵이 가장 좋다.
* 대시보드의 지루함 해소 & 대시보드로 사람들을 유인할 방법
  + 개인에게 데이터가 얼마나 자신과 관련되는지를 보여주기
* 개인화된 대시보드: 개인의 입력에 따라 변하는 대시보드
  + 인구통계 대시보드의 개인화
    - 사용자가 개인화된 설정을 적용할 수 있는 파라미터 제시
    - 파라미터 설정을 취해 대시보드에서 강조

\*막대차트의 대안: 롤리팝 차트

* 사람들이 대시보드를 사용하지 않는다면 그것은 의미가 없는 정보이기 떄문이다. (대시보드의 문제 X)
* 대시보드를 꾸미는 것이 처음에 관심을 끌지는 모르지만, 데이터를 이해하는 데 방해만 될 가능성이 크다.
* 사람들의 참여를 원한다면 데이터를 의미 있게 만들어야 한다. (-> 개인화된 대시보드)
* 시간의 시각화
  + 시간 기반 질문
    - 기간의 처음과 비교해 오늘의 추이는 어떤가?
    - 데이터에 순환적인 패턴이 있는가?
    - 2개의 시간 차원 간에 걸친 추세를 어떻게 찾을까?
    - 장기에 걸친, 수치가 아닌 순위는 어떻게 관찰할까?
    - 동시에 발생하지 않은 값들은 어떻게 비교할까?
    - 각기 다른 시기에 발생한 사건의 성장세는 어떻게 비교할까?
    - 프로세스의 병목에 어떻게 집중할 수 있을까?
  + 시간을 보는 다양한 방법
    - 연속적 시간: 꺾은선 차트 (X축-시간(좌->우) / Y축-정량적 척도)
    - 범주형 시간(e.g. 요일): 막대차트 (그룹핑해서 추이 표현)
  + [1] 초기 데이터와 최근 데이터의 비교
    - 시계열에서 처음과 마지막 시점을 제외하고 모든 포인트 제거 -> 상승/하락 표현 가능
    - 여러 선의 중첩이 일어날 경우 -> 겹치지 말고 레이블마다 가로로 떼어서 그리기
  + [2] 순환적인 패턴
    - 시간 순서대로 꺾은선 그래프 -> 시간이 잘 안보이는 문제
    - 시간으로 그룹핑하여 막대 차트 -> 요일 별 차이 관측 실패
    - 순환도: 막대차트의 장점 + 시계열 꺾은선 차트의 장점 -> 시간대 별, 요일 별 둘다 확인 가능



* + 순환도는 가로로 길어야 한다.
  + 순환도는 많은 공간을 차지한다.
  + 공간이 문제라면 요일마다 시간 순서 꺾은선 차트를 겹쳐서 그릴 수 있다.



* + [3] 2개의 시간 차원 간에 걸친 추세
    - 시간의 흐름에 따른 꺾은선 차트: 정점과 저점을 보긴 쉽지만, 나머지의 정보를 파악하기 어렵다.
    - 하나의 시간 차원에 따른 막대 차트: 남은 하나의 시간에 대한 정보 확인 불가.
    - 두 시간 차원 축을 갖는 히트맵(하이라이트 표): 두 가지 시간 정보 다 확인 가능.

-> 히트맵에 레이블을 추가하면 표를 시각화 하는 효과

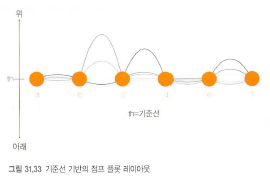
* + [4] 장기에 걸친, 수치가 아닌 순위 관찰
    - 범프 차트: 해당 시점의 값이 아닌 순위를 표시하는 꺾은선 차트 -> 순위가 중요한 정보일 때 사용

(실제값을 알 수 없다는 문제 -> 각 포인트마다 레이블에 실제 값 표기로 해결, 색이 많으면 과하다)

* + [5] 다른 시기에 발생한 사건의 성장세 비교
    - 막대 차트: 인기 순위를 비교할 수는 있지만 성장세 비교 불가능
    - 꺾은선 차트: 성장세를 보여줄 수는 있지만 특정 기준선에 얼마나 빨리 도달했는지 비교 불가
    - 인덱스 차트: X축을 발표 이래 경과일 (인덱스)로 해 시작점을 일치시키면 비교 가능
  + [6] 프로젝트나 이벤트 기간
    - 간트 차트: 시작 시점과 지속 기간 표현



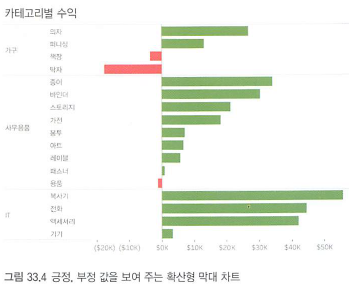
* + - 점프 플롯: 과업 사이의 지속기간을 곡선으로 표현 (여러 시퀀스 표현 가능 -> 막대보다 좋음)



* 막다른 길에 다다른 대시보드에 주의하자.
  + 대시보드는 끝이 없다. 계속 진화해야 한다.
  + 성장하는 대시보드를 만드는 절차
    - 1. KPI의 검토: 대시보드가 표현하고 있는 이 타겟이 적절한지 검토
    - 2. 데이터 활용의 추적: 활용도가 낮은 대시보드의 삭제 또는 수정이 필요
    - 3. 사용자와의 대화
    - 4. 초심자용 대시보드 구축: 기본적인 대시보드를 제공한 후 각자 수요에 맞춰 개인화 하도록
* 적색과 녹색의 유혹
  + 색각 이상/색맹/CVD -> 적색과 녹색을 함께 사용한 시각화는 문제 -> 색맹 친화적인 시각화 필요
  + 색맹 친화적인 시각화
    - 1. 색맹 친화적 팔레트: 청색/오렌지색, 청색/적색, 청색/갈색



* + - 2. 적색과 녹색을 함께 사용해야 한다면
      * 색 없이도 긍정 부정을 잘 구분할 수 있는 경우



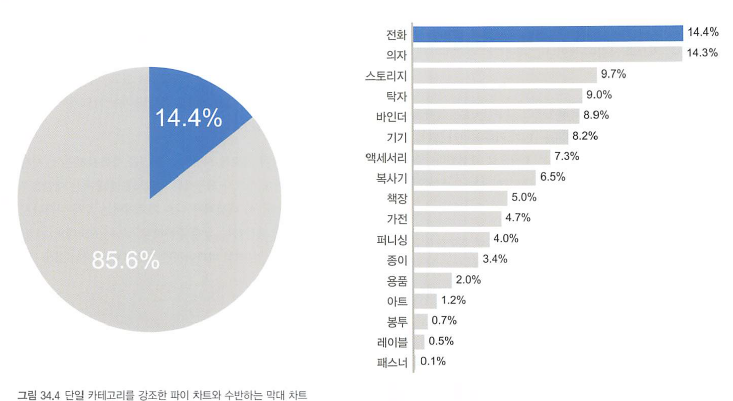
* + - * 아이콘 (e.g. 방향 있는 화살표) 또는 레이블을 함께 사용



* + - * 색을 전환하는 옵션 제공
      * 밝은 녹색과 어두운 적색 사용
      * 밝고 파란 녹색과 어두운 적색 사용 (녹색에 청색 섞기)
* 파이와 도넛의 달콤한 유혹
  + 정확한 정량 비교를 하는 규약화 방법
    - 1. 기준선으로부터의 길이나 높이 (막대 차트)
    - 2. 위치를 사용해 비교하는 방법 (점 도표)

\*정량 비교를 할 때 각도, 아치, 원 등을 사용하는 것 좋지 X -> 파이 차트, 도넛 차트, 버블 차트 별로!

* 파이 차트: 카테고리가 많을 때 색이 너무 많고 레이블과 오락가락 해야 하는 문제
  + 해결방안: 하나의 카테고리만 강조 & 막대 차트 추가



* 도넛 차트: 최대 목표가 100%가 아니거나 그 이상을 표현하지 못하는 문제
  + 해결방안1: 불릿 차트, 기준선이 있는 막대 차트
  + 해결방안2: 도넛 차트에 막대 차트 또는 점 추가





* 아득한 구름과 물거품
  + 워드 클라우드의 문제: 언급량의 비교 불가, 순위 파악 어려움
  + 버블 차트의 문제: 가장 큰 원을 제외한 나머지 원의 데이터를 이해하는데 오래걸림
  + 해결방안
    - 막대 차트: 데이터 전달력이 좋지만, 지루하다는 단점

-> 한 카테고리만 색으로 강조 나머지는 회색

* + - 롤리팝 차트: 막대 차트의 지루함 해소
    - 설득력 있는 제목 추가
* 미지의 세계로의 여정
  + 모든 대시보드는 완벽하지 않다.
  + 먼저 개괄하고 줌과 필터를 제공한 다음 온디맨드로 세부사항을 제시한다.
  + 대시보드에 부가가치를 더할 수 있는 방법
    - 끊임없이 질문하기
    - 거침없는 시각적 탐구: 텅 빈 백지에 자유롭게 데이터 흩뿌리기-> 새로운 통찰을 찾을 때까지