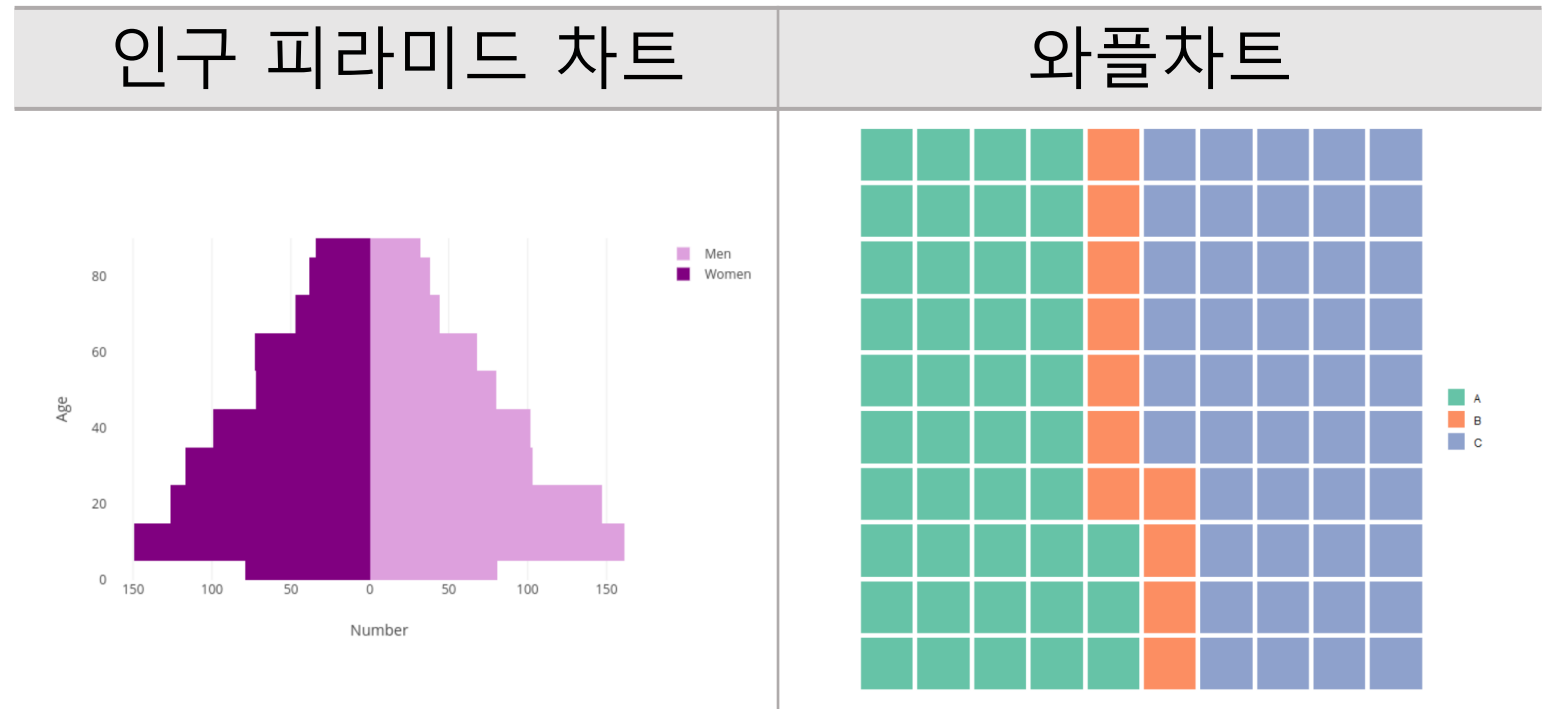


## 2.4 데이터 이해를 위한 탐색과 시각화

시간 시각화	분포 시각화	관계 시각화	비교 시각화	공간 시각화
<ul style="list-style-type: none"><li>- 막대 그래프</li><li>- 누적 막대 그래프</li><li>- 점 그래프</li></ul>	<ul style="list-style-type: none"><li>- 파이차트</li><li>- 도넛차트</li><li>- 트리맵</li><li>- 누적연속 그래프</li></ul>	<ul style="list-style-type: none"><li>- 산점도</li><li>- 버블 차트</li><li>- 히스토그램</li></ul>	<ul style="list-style-type: none"><li>- 히트맵</li><li>- 체르노프 페이스</li><li>- 스타차트</li><li>- 평행 좌표계</li><li>- 다차원 척도법</li></ul>	<ul style="list-style-type: none"><li>- 지도 매핑</li></ul>

## ◎ 기초 그래프

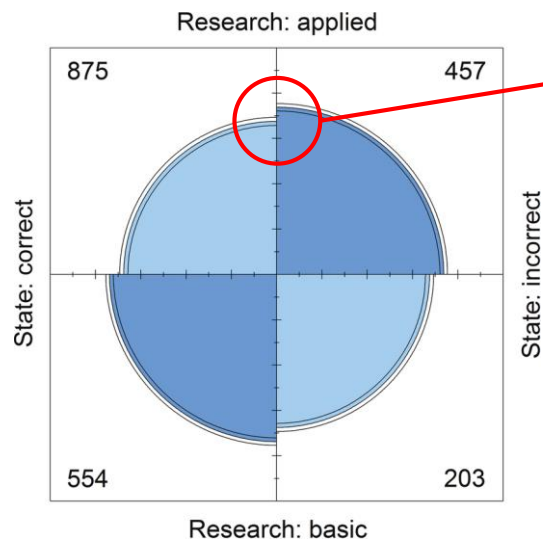
1. 산점도
2. 막대 그래프
3. 파이차트
4. 누적 영역차트
5. 히스토그램
6. 상자그림



## ◎ 범주형 자료의 시각화

1. 바 차트
2. 모자이크 차트
3. fourfoldplot()

fourfoldplot()



오즈비에 대한 99% 신뢰구간

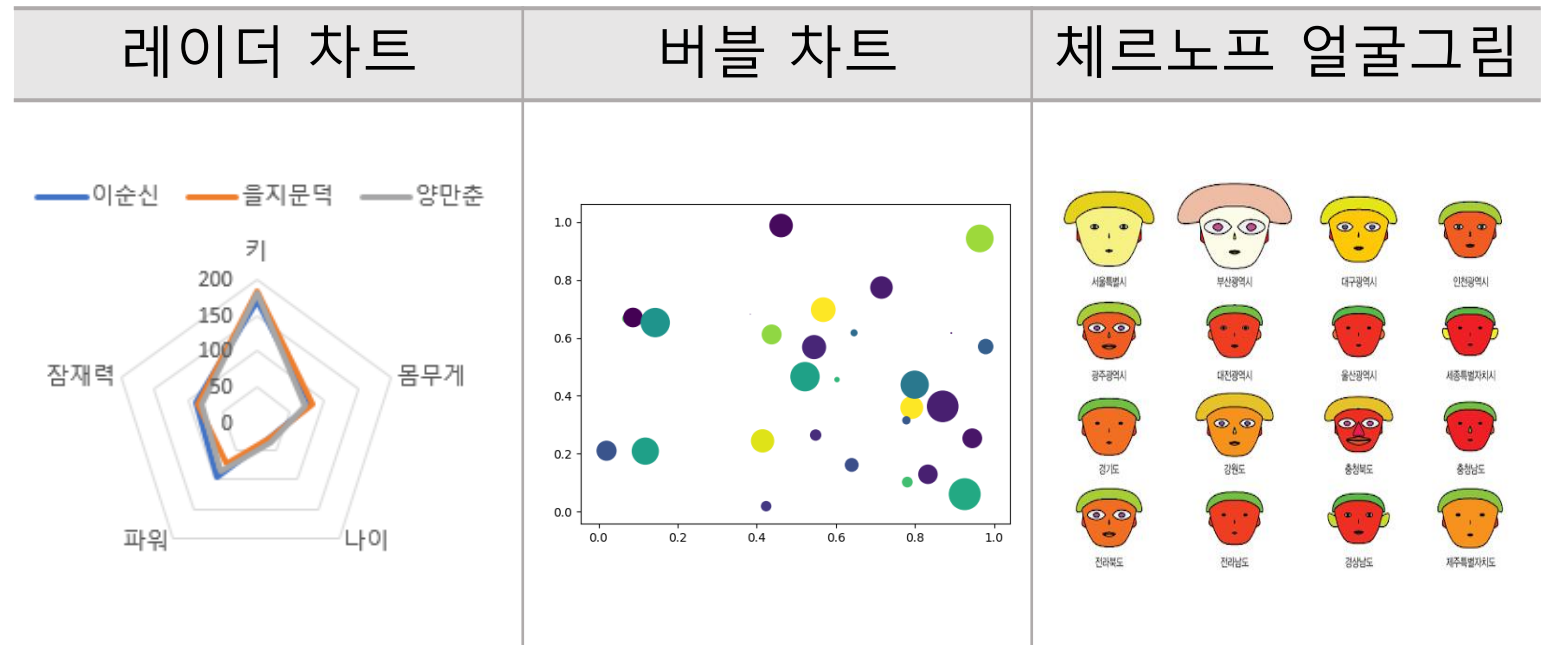
### Odds ratio calculation

$$OR = \frac{a / b}{c / d} = \frac{ad}{bc} \quad \text{where}$$

		Cancer	
		✓	✗
Exposure	✓	a	b
	✗	c	d

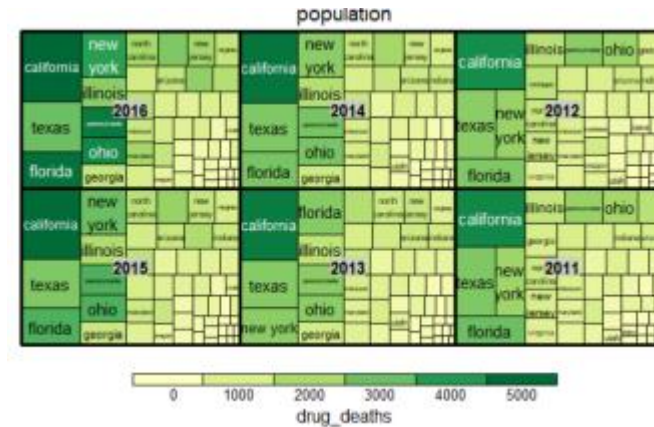
## ◎ 다변량 자료의 시각화

1. 다변량 산점도
2. 레이더 차트
3. 버블차트
4. 체르노프 얼굴그림



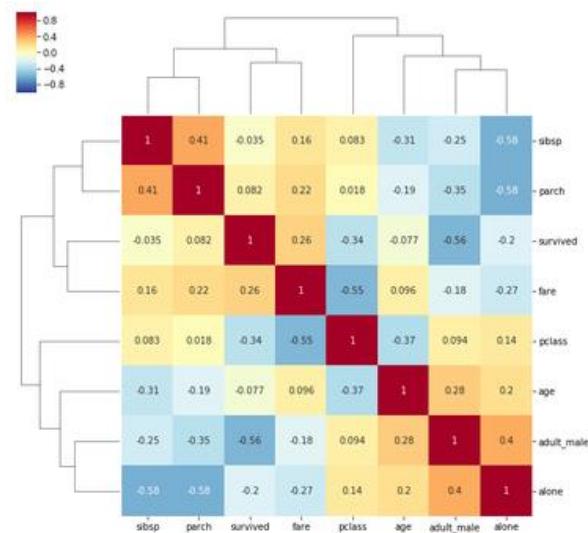
## ◎ 트리맵

위계적인 자료로 공간을 채우는 시각화  
크기와 색의 진하기로 변수의 크기 표현



## ◎ 히트맵

숫자로 구성된 행렬 자료 시각화



## ◎ 상관관계 분석

○ 모수적 방법 - 피어슨 상관계수

$$r_{XY} = \frac{\sum_i^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_i^n (X_i - \bar{X})^2} \sqrt{\sum_i^n (Y_i - \bar{Y})^2}}$$

○ 비모수적 방법 - 스피어만 상관계수

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$