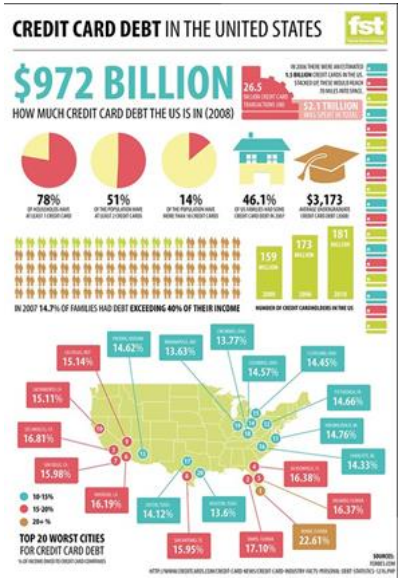
시각화

데이터 시각화

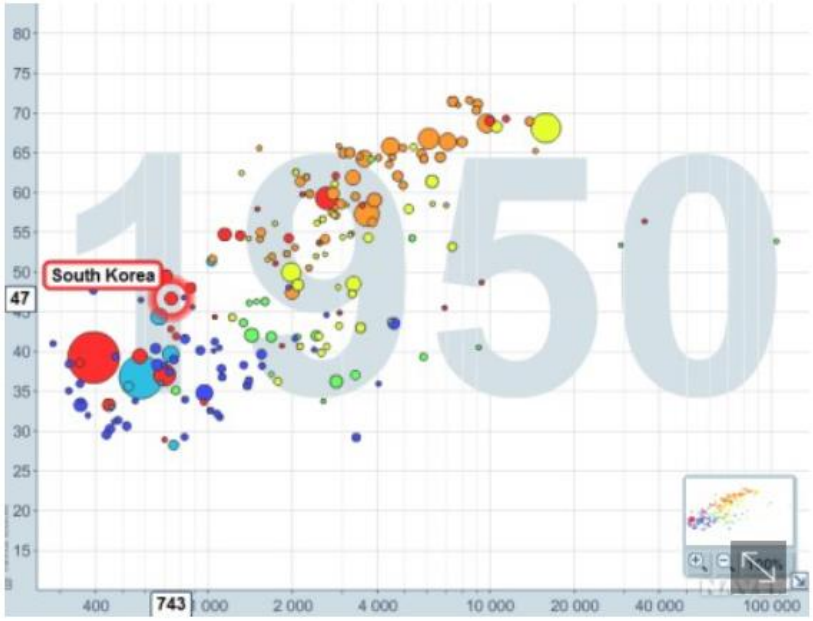
- **정의 및 목적**
  - 데이터 시각화는 데이터 분석 결과를 쉽게 이해할 수 있도록 시각적으로 표현하고 전달하는 과정을 말함
  - 데이터 시각화의 목적은 도표(graph)라는 수단을 통해 정보를 명확하고 효과적으로 전달하는 것 [Friedman, 2008]

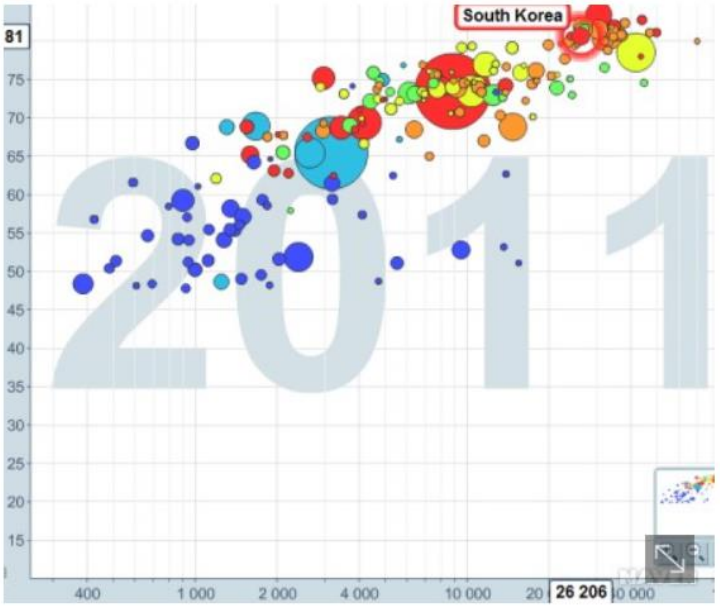
- **필요성**
  - 데이터에 대한 직관적인 분석이 가능
  - 데이터의 추세(trend)를 알 수 있음
  - 데이터 간의 관계를 한 눈에 알 수 있음



시각화 사례

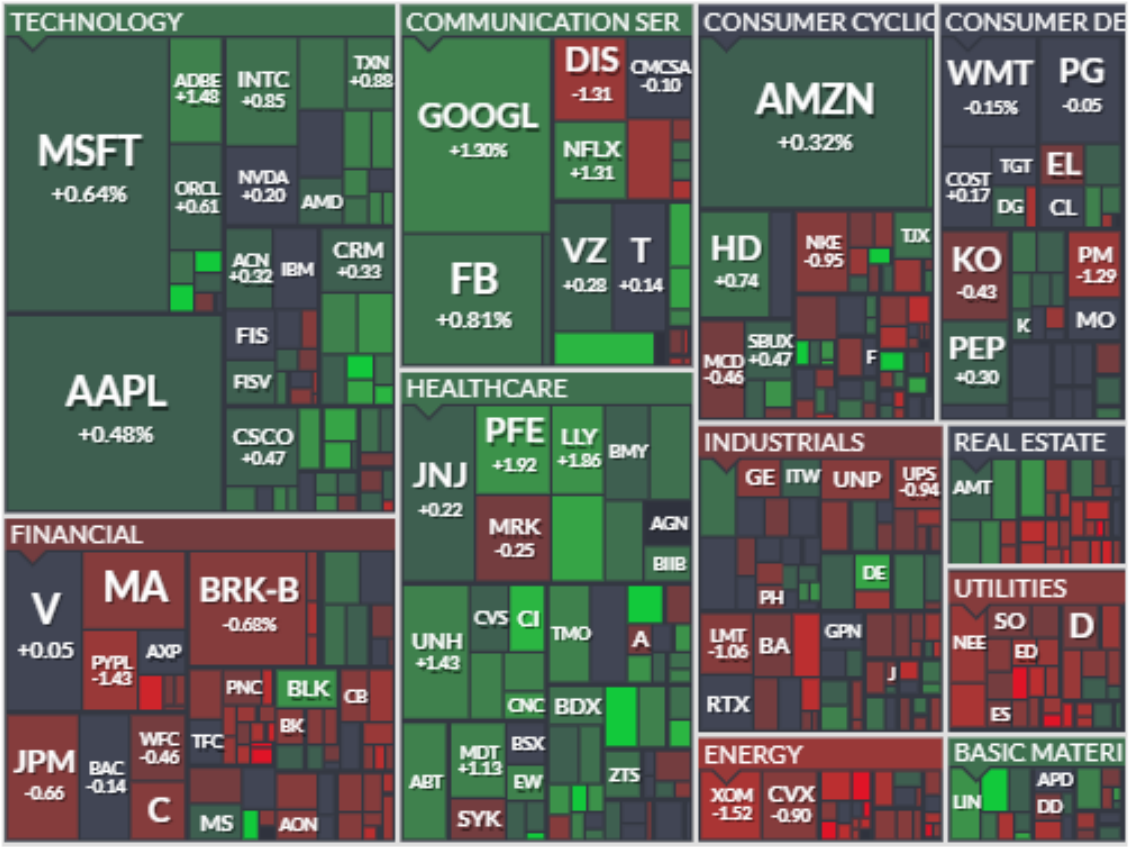
- **국가별 기대수명과 GDP(1950년 vs. 2011년)**
  - 한국의 기대수명은 47년에서 81년으로
  - GDP는 743달러에서 26,206달러로





시각화 사례2

- 미국 **S&P500** 시가총액
  - finviz.com



시각화 도구

- 시각화 도구의 특성
  - 어떤 시각화 도구를 선택하는 것이 좋을까?
  - 파이차트와 막대그래프는 어떤 데이터를 시각화하는데 좋을까?
  - 히스토그램이나 산점도는 어떤 경우에 선택하는 게 좋을까?

- 어떤 시각화도구를 선택하는것이 좋을까? **[zybooks]**
  - Pie charts are better suited for low-cardinality data than high-cardinality data.
  - Bar charts are appropriate for plotting categorical data.
  - Histograms and scatter plots are well-suited for high-cardinality data.

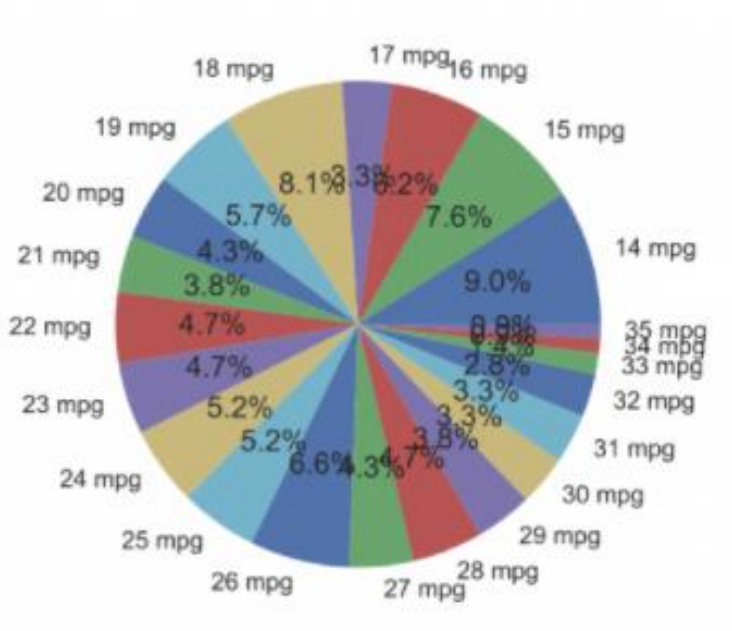
Cardinality

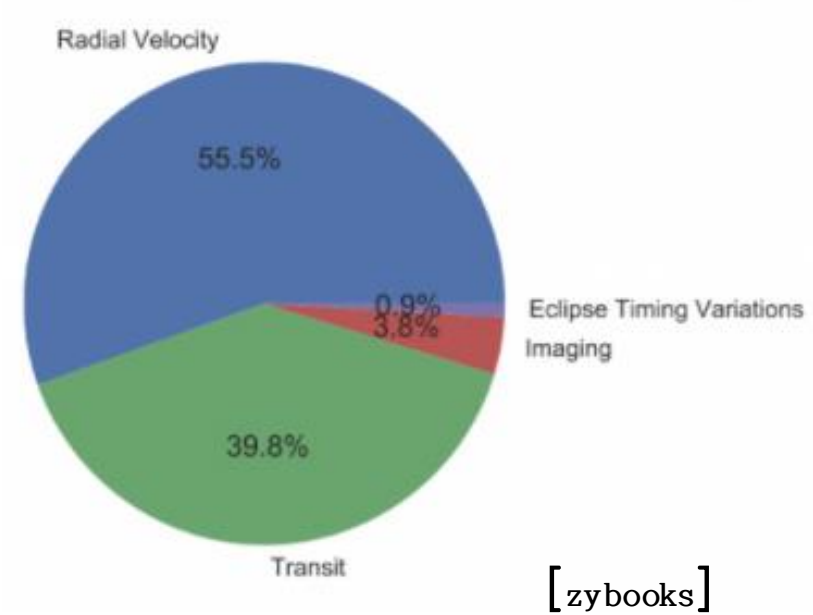
- **Cadinality의 정의**
  - Cardinality is the number of unique elements in a dataset.
  - Ex: the set of student IDs of students in a class has high cardinality, since each ID is unique, whereas the set of student ages will have lower cardinality, since many students will have the same ages
  - 즉, 상대적으로 해당 변수의 가능한 값의 집합 개수가 적을 수록 cadinality가 낮다고 할 수 있음

- **예제**
  - cardinality가 높다. -> Account Number (계좌번호)
  - cardinality가 보통. -> Last Name
  - cardinality가 낮다. -> Gender

파이차트

- **Pie chart**
  - Pie charts are better suited for low-cardinality data than high-cardinality data

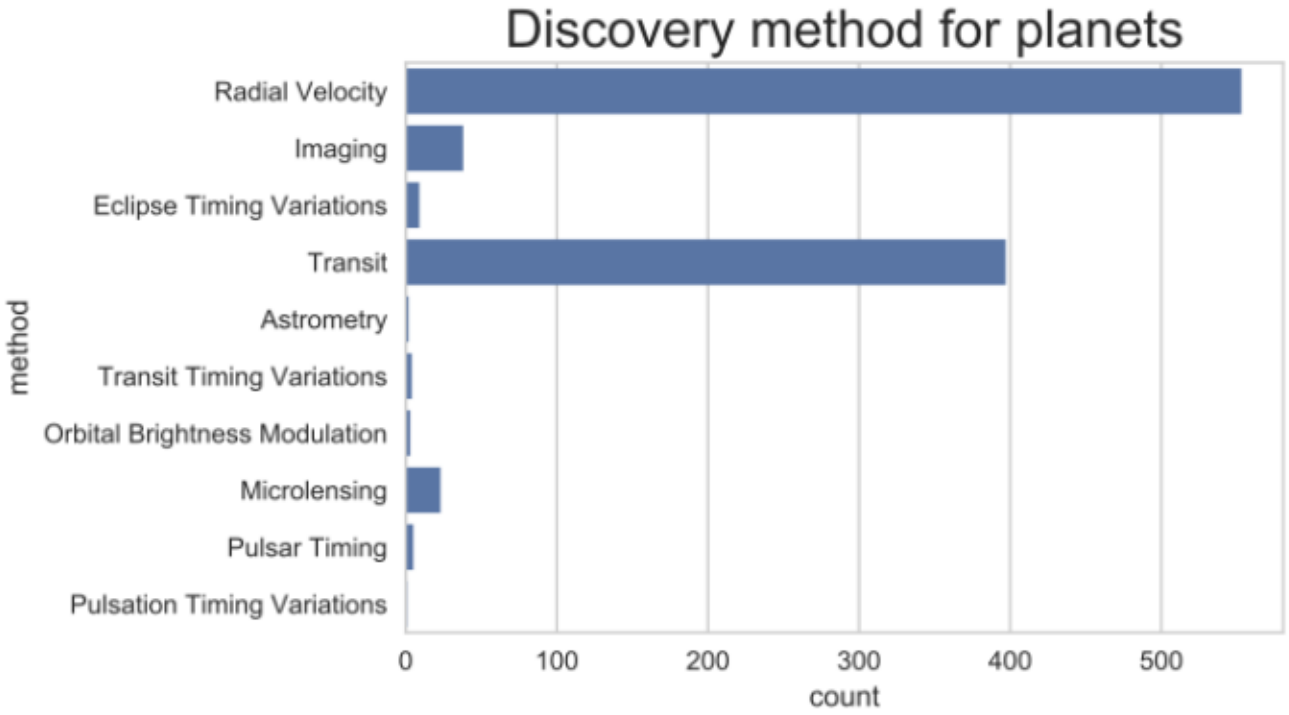




[zybooks]

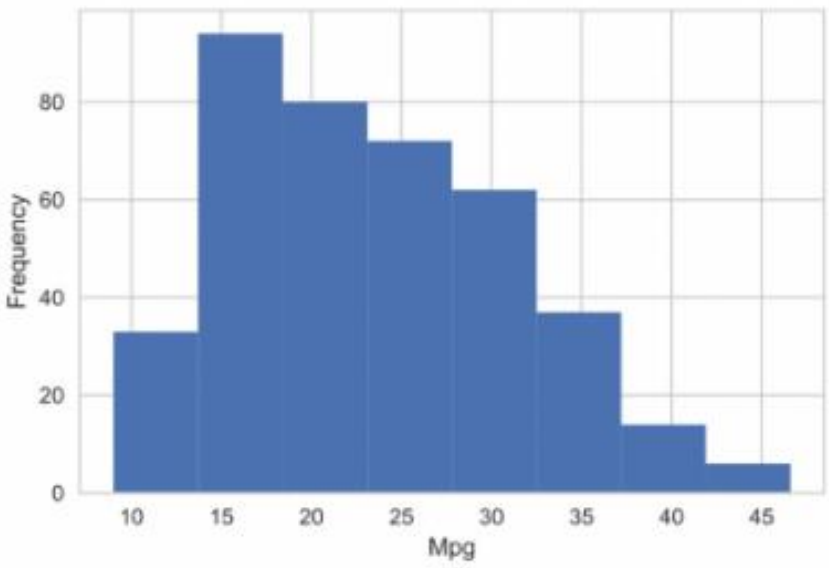
막대 그래프

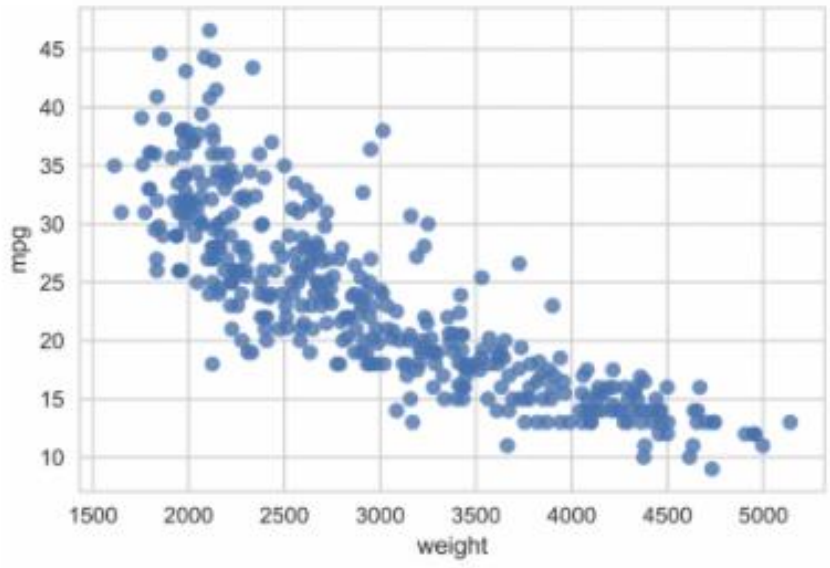
- **Bar chart**
  - Bar charts are appropriate for plotting categorical data



히스토그램 & 산점도

- **Histogram & Scatter plot**
  - Histograms and scatter plots are well-suited for high-cardinality data.

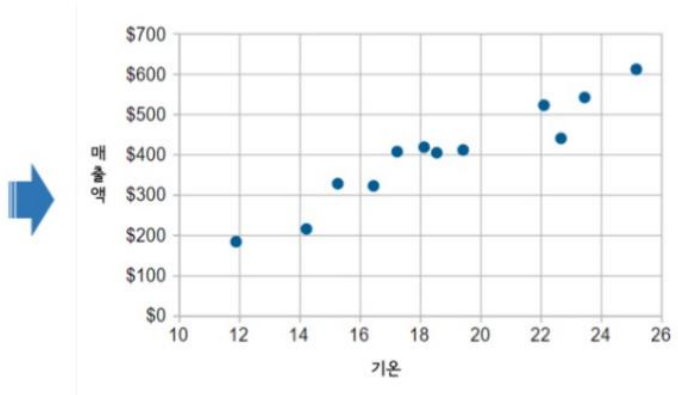




산점도

- 산점도scatter plot
  - 산점도는 서로 다른 두 변수 사이의 관계를 표현
  - 이때 각 변수는 연속되는 값을 가짐
  - 일반적으로 정수형(int64) 또는 실수형(float64)값

아이스크림 매출과 기온	
기온	아이스크림 매출액
14.2°	$215
16.4°	$325
11.9°	$185
15.2°	$332
18.5°	$406
22.1°	$522
19.4°	$412
25.1°	$614
23.4°	$544
18.1°	$421
22.6°	$445
17.2°	$408



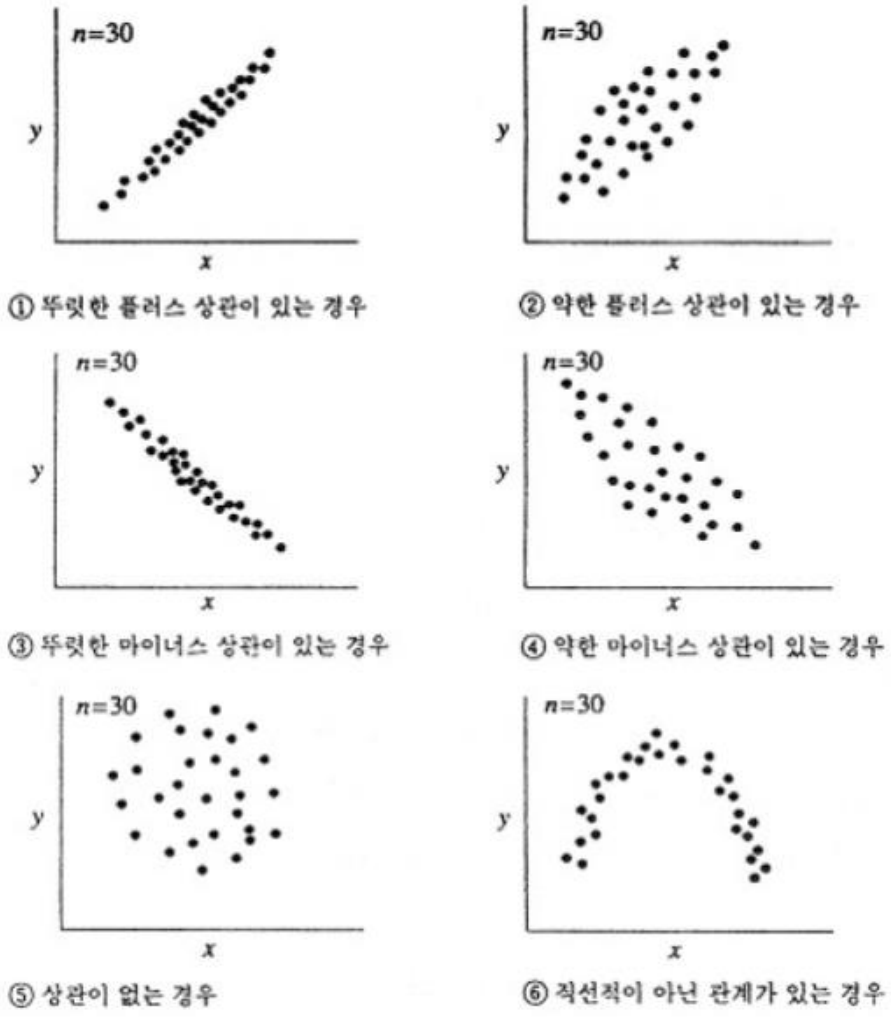
산점도 코딩

- 아이스크림 매출 예제
  - 기온의 열 Label은 ‘temperature’
  - 아이스크림 매출의 열 Label은 ‘sales’
  - 기온과 아이스크림 매출이 icecream 데이터프레임으로 주어지면
  - 기온과 아이스크림 매출에 대한 산점도 코딩은
  - icecream.plot(kind=‘scatter’, x=‘temperature’, y=‘sales’)

    plt.show()

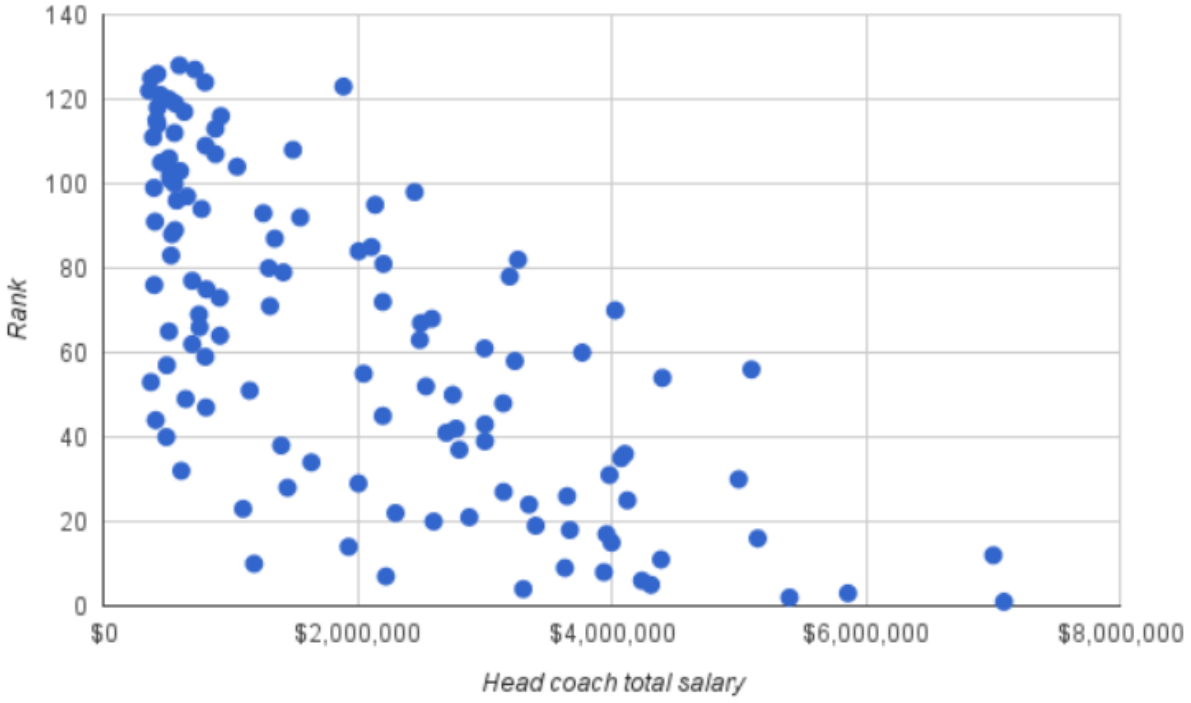
산점도 해석법

- 산점도scatter plot
  - 1. 뚜렷한 +상관관계
  - 2. 약한 +상관관계
  - 3. 뚜렷한 –상관관계
  - 4.약한 –상관관계
  - 5. 상관관계 없음
  - 6. 비선현 상관관계



산점도 해석 사례

- 대학 축구 랭킹과 수석코치연봉의 상관관계는?
  - 연봉이 높을 수록 랭킹이 높은가?



참고도서

➢ reference

[1] 파이썬을 활용한 데이터길들이기

- 프로그래밍인사이트

[2] 모두의 데이터분석

- 길벗

[3] 파이썬머신러닝 판다스데이터분석

- 정보문화사

End