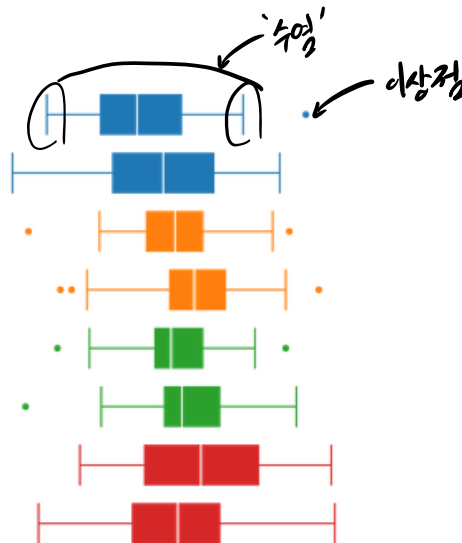


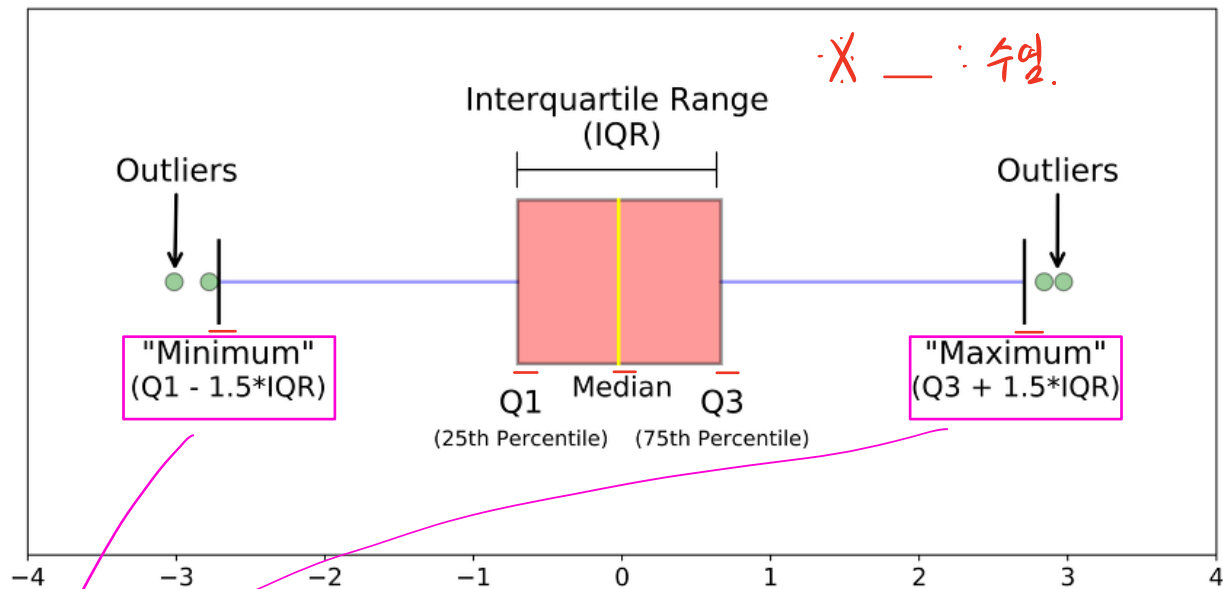
박스 플롯(box plot) 또는 박스앤위스커 플롯(상자수염플롯; box-and-whisker plot)은 데이터의 대략적인 분포와 개별적인 이상치들을 동시에 보여줄 수 있으며 서로 다른 데이터 뭉치를 쉽게 **비교**할 수 있도록 도와주는 시각화 기법으로 가장 널리 쓰이는 시각화 형태 중 하나이다.

통계학자 존 튜키(John Tukey)가 1977년 저서 "탐색적 데이터 분석(Exploratory Data Analysis)"에서 처음 제시한 것으로 알려져 있으나, 매리 엘리너 스피어 Mary Eleanor Spear가 1952년에 저서 <Charting Statistics>에서 "범위 막대(range bar)"라는 이름으로 **먼저 제안한 바 있다**. 범위 막대에서는 수염의 양 끝이 최대값과 최소값을 나타내고, 상자는 사분위수범위를 나타내며, 상자 안에 중앙값을 표현하는 방식으로 "5가지 요약 수치"를 표현한다. 튜키에 비해 25년 앞서 거의 동일한 형태를 제안하였으나 상대적으로 잘 알려져 있지 않다.

데이터 뭉치 사이에 관련성이 있다면 같은 색상을 써서 시각적으로 짝지어주면 좋다. 데이터 뭉치 사이에 어떠한 순서(order)가 있는게 아니라면 명도나 채도 차이를 보다는 색조(hue) 차이를 주어야 불필요한 오해를 줄일 수 있다.



※ 데이터가 적을 때는 평균보다 중앙값이 대표값으로 신뢰성이 높다.
적은 수의 데이터 분포를 볼 때는, Boxplot을 사용하는 것이
좋다.



수업의 maximum과 minimum은 실제 원본 데이터의 maximum과 minimum이 아닐 수 있다. 박스 플롯을 통해서 봤을 때, 해당 수업의 maximum과 minimum이 정상적인 maximum과 minimum이 되어야 한다는 것을 암시해준다.

5) 상자그림(Box plot): 다섯 숫자 요약을 통해 그림으로 표현 (최소값, Q1, Q2, Q3, 최대값)

- 사분위수범위(IQR) : $Q3 - Q1$
- 안울타리(inner fence) : $Q1 - 1.5 \times IQR$ 또는 $Q3 + 1.5 \times IQR$
- 바깥울타리(outer fence) : $Q1 - 3 \times IQR$ 또는 $Q3 + 3 \times IQR$
- 보통이상점(mild outlier) : 안쪽 울타리와 바깥 울타리 사이에 있는 자료
- 극단이상점(extreme outlier) : 바깥울타리 밖의 자료

타자들의연봉데이터(단위:백만원)

630	400	180	550	162	500	270	200
192	200	200	80	310	200	135	300
160	70	220	350	700	88	100	250
400	400	70	160	185	170	85	202
80	72	100	350	85	500	140	100

다섯 숫자(70,100,196,330,700)

사분위수범위:230

안울타리(0,675)

바깥울타리(0,1020)

