

## ■ 표를 이용한 정리

- 자료구조와 분석목적
  - 구조: 분류(그룹=범주)자료 + 수치자료
  - 목적: 그룹 간 비교
- 수치자료의 특성을 나타내는 값을 그룹별로 정리
  - 기본 통계값: 표본크기, 평균, 표준편차

● 14년 지방선거 17개 광역단체장 당선자 득표율 비교

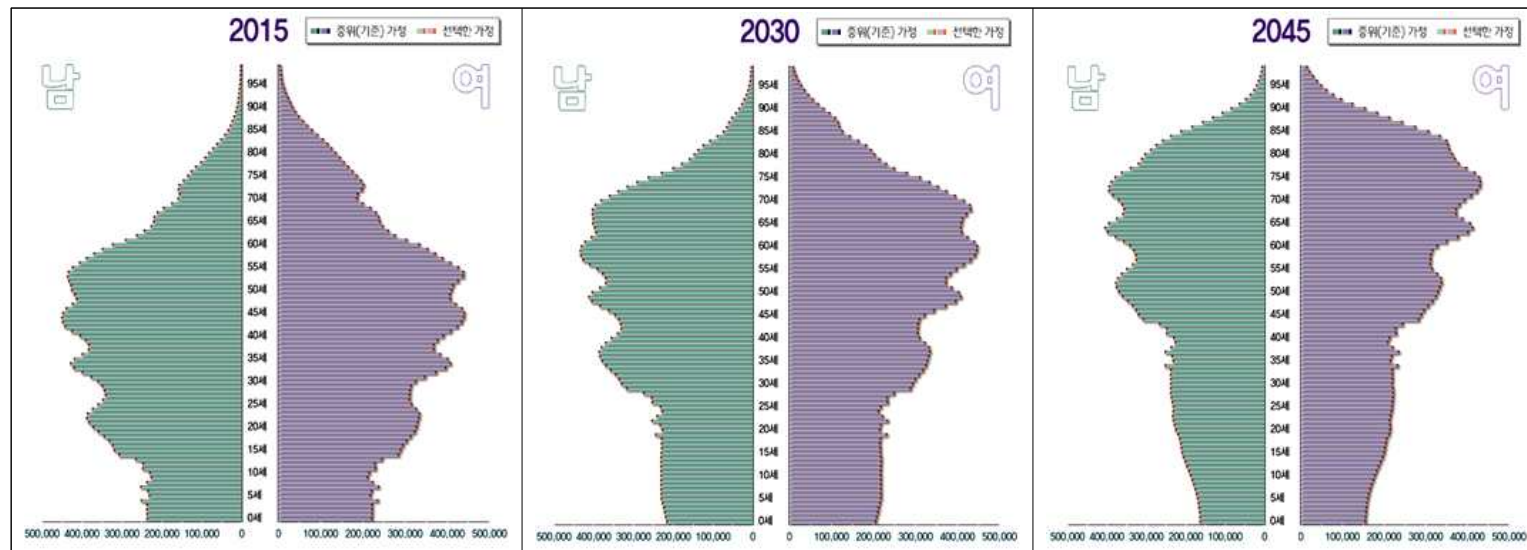
정당	B	B	A	B	A	A
득표율	49.75	49.76	49.95	50.07	50.43	50.65
정당	B	A	B	B	B	A
득표율	52.21	55.95	56.12	57.78	57.85	58.85
정당	A	A	B	A	B	
득표율	59.97	65.42	69.23	77.73	77.96	

○ 정당별 당선자 득표율 비교

정당	N	평균	표준편차	최소	최대
A당	8	58.6	9.5	50.0	77.7
B당	9	57.9	9.7	49.8	78.0

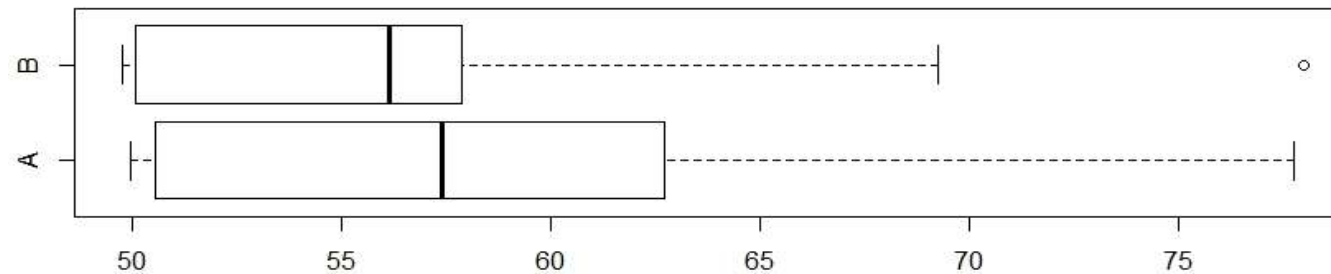
## ■ 그래프를 이용한 비교

- 점도표, 히스토그램(막대그래프)을 중첩, 병렬로 비교
  - 예】 인구추계교실(<http://kosis.kr>)



- 동일 축 상의 그룹별 상자그림 비교
- ◎ 14년 지방선거 상자그림 비교

정당	$Q_1$	$Q_2$	평균	$Q_3$
A당	50.60	57.40	58.62	61.33
B당	50.07	56.12	57.86	57.85



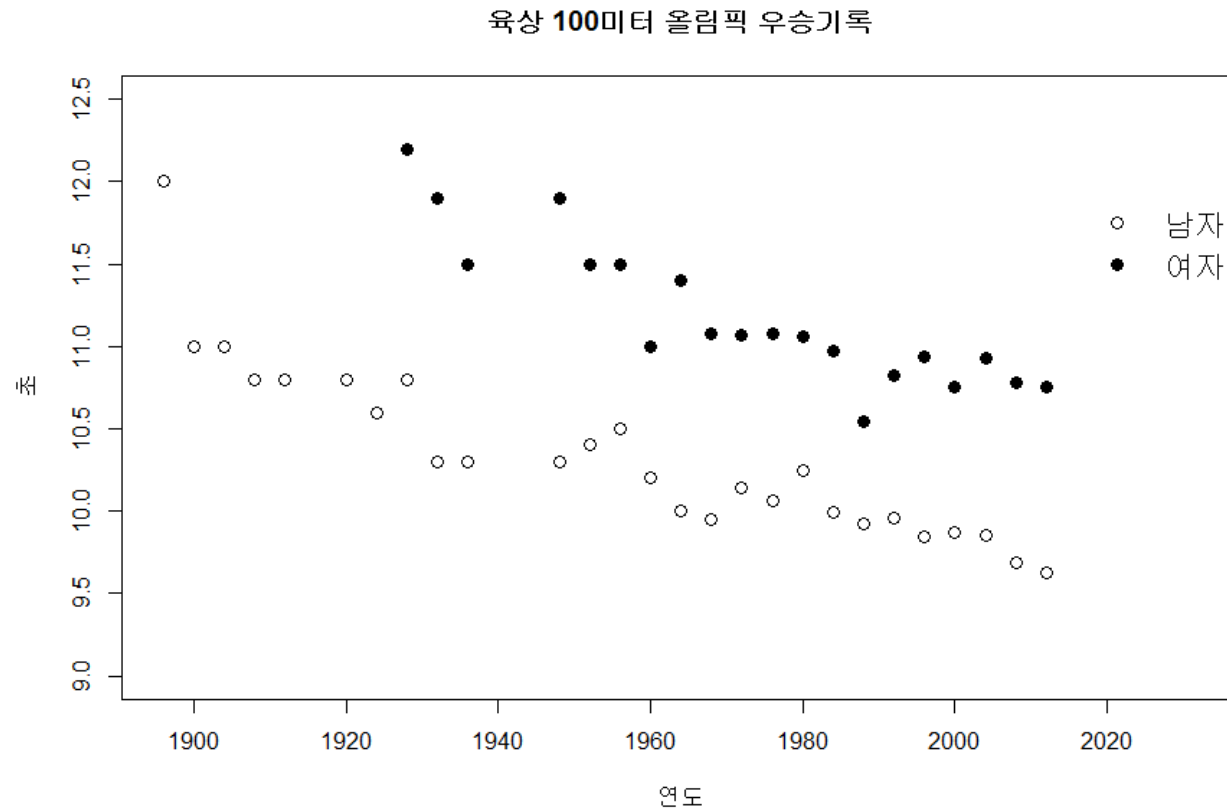
## ■ 산점도(scatter plot)

- 자료구조와 분석목적
  - 구조: 수치자료 + 수치자료 (+ ... + 수치자료)
    - 각각의 관측개체에 대해 두 변수의 값은  
순서쌍  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ 으로 표시
  - 목적: 수치 변수들 간의 관계를 유도
- 순서쌍 자료를 2차원 평면상에 점으로 표시한 그림

## ◎ 올림픽 100미터 우승기록

연도	우승기록		연도	우승기록		연도	우승기록		연도	우승기록	
	남자	여자		남자	여자		남자	여자		남자	여자
1896	12	-	1928	10.8	12.2	1964	10.0	11.4	1992	9.96	10.82
1900	11	-	1932	10.3	11.9	1968	9.9	11.0	1996	9.84	10.94
1904	11	-	1936	10.3	11.5	1972	10.14	11.07	2000	9.87	10.75 <sup>1)</sup>
1908	10.8	-	1948	10.3	11.9	1976	10.06	11.08	2004	9.85	10.93
1912	10.8	-	1952	10.4	11.5	1980	10.25	11.06	2008	9.69	10.78
1920	10.8	-	1956	10.5	11.5	1984	9.99	10.97	2012	9.63	10.75
1924	10.6	-	1960	10.2	11.0	1988	9.92	10.54	2016	9.81	10.71

1) 우승자인 marion L. Janes는 이후 금지약물 복용이 발각되어 기록과 메달이 박탈됨



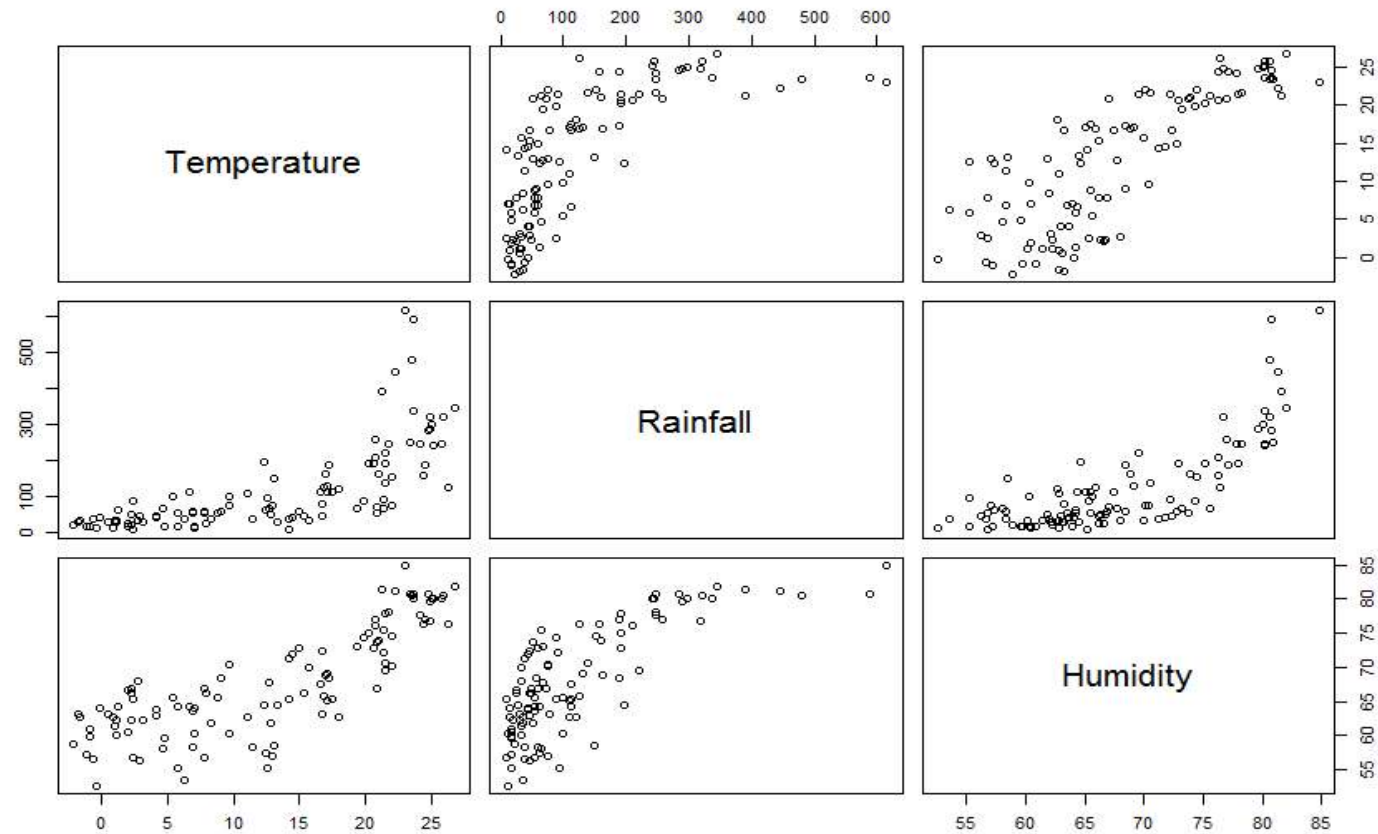
## ○ 산점도 행렬(scatter matrix)

- 3개 이상의 수치변수에 대해 두 변수씩 쌍으로 조합하여 산점도를 행렬형태로 표시

### ● 기상자료들 간의 관계

- 2002년 1월부터 2010년 12월까지 우리나라 전체의 월간 평균기온(Temperature), 강수량(Rainfall), 평균습도(Humidity)





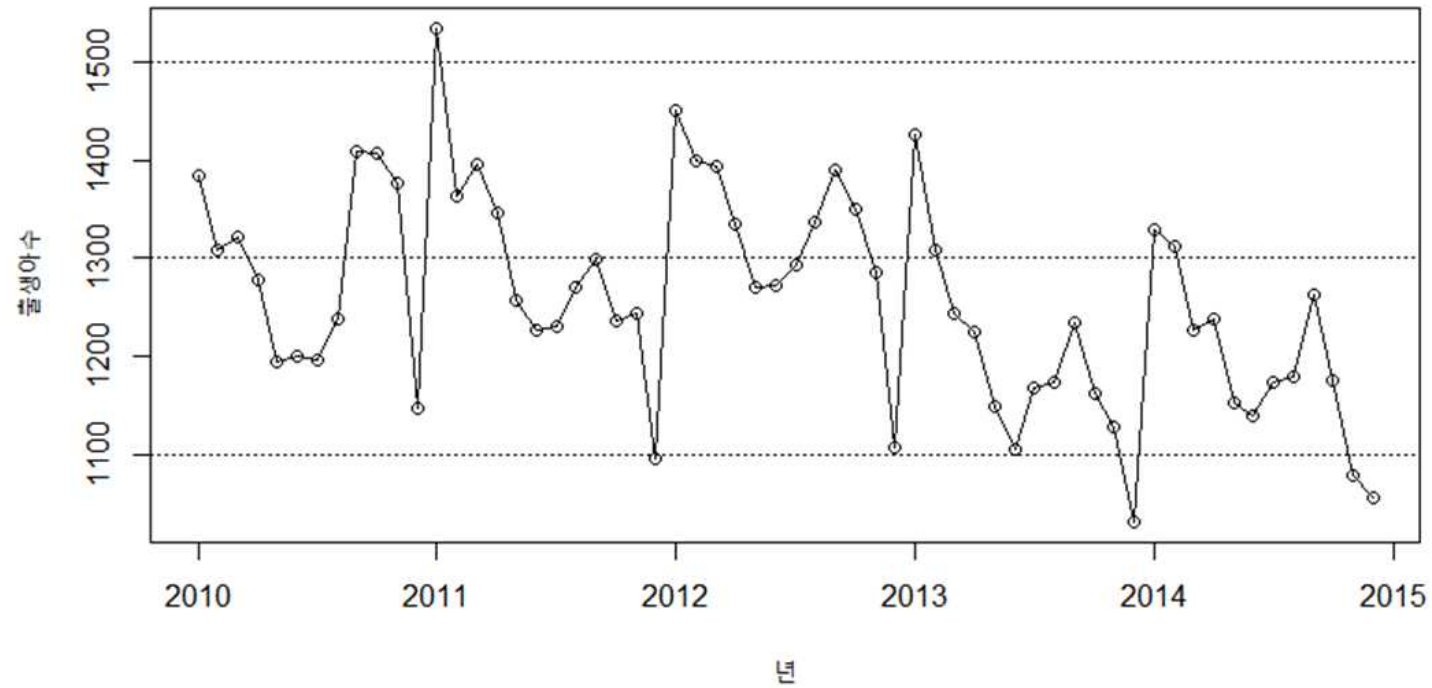
## ○ 시계열그림(Time series plot)

- 시계열자료: 시간에 따라 수집된 자료
  - 자료 간 종속성이 존재할 수 있어 관측순서가 중요
- 시계열그림:  $x$  축에 관측시점,  $y$  축에 관측된 값을 표시한 산점도
  - 예】 올림픽 육상 100m 우승기록 그림
  - 순서 파악을 쉽게하기 위해 일반적으로 관측 순으로 관측값들을 선으로 연결

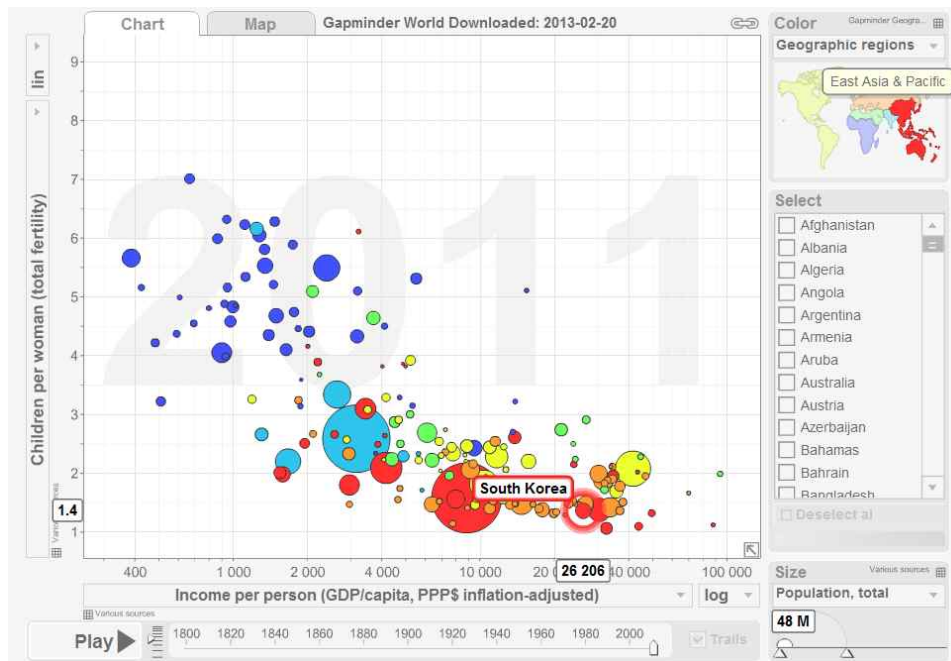
## ◎ 월별 하루 평균출생아수

- 통계청에서 발표하는 『월간 인구동향』 통계 중 2010년 1월부터 2014년 12월까지 자료

연도 \ 월	1월	2월	3월	4월	5월	6월	7월	8월	9월	10월	11월	12월
2010	1385	1309	1322	1278	1194	1201	1197	1238	1410	1407	1377	1146
2011	1535	1363	1395	1346	1257	1226	1231	1270	1299	1237	1244	1096
2012	1451	1400	1394	1336	1271	1272	1294	1337	1391	1351	1286	1107
2013	1427	1309	1243	1225	1149	1105	1167	1174	1235	1163	1128	1031
2014	1330	1313	1226	1239	1153	1139	1174	1179	1263	1176	1079	1055



- Gapminder
  - [www.ted.com](http://www.ted.com) : statistics => Hans Rosling



이미지 출처: <https://www.gapminder.org/news/upgrade-notes-gapminder-world-offline-version-0-0-7/>

- 정리
  - 그룹별 수치자료 정리(표): 표본크기, 평균, 표준편차
  - 그룹별 수치자료 비교: 히스토그램(막대그래프), 상자그림
  - 산점도: 산점도 행렬, 시계열그림