

데이터 전처리 요약 과정

1. 날씨누리 월별 지역 기상 데이터 수집 - 1990 ~ 2016

* 일일 데이터를 먼저 수집했으나 범죄데이터가 제공하지 않아 변경.

* 범죄 데이터의 지역과 날씨 데이터의 지역을 통일하기 위해 행정구역(도)를 그 지역에서 가장 많이 살고 있는 도시로 선정.

- 지역 데이터

"서울(유)" : "108", "부산(유)" : "159",
"대구(유)" : "143", "인천(유)" : "112",
"광주(유)" : "156", "대전(유)" : "133",
"울산(유)" : "152", "제주(유)" : "184",
"경기(유)" : "119", "강원(무)" : "101",
"충북(유)" : "131", "충남(무)" : "232",
"전북(유)" : "146", "전남(유)" : "168",
"경북(유)" : "138", "경남(유)" : "155"

- 지역 별 기상 데이터

최고기온 최저기온 평균기온 평균운량 일강수량 상대습도 평균평속

	년월	최고기온	최저기온	평균기온	평균운량	일강수량	상대습도	평균평속
314	2016/3	14.0	4.7	8.9	4.9	81.6	59.2	1.7
315	2016/4	19.9	10.4	14.7	5.5	263.0	66.1	1.7
316	2016/5	25.1	14.5	19.4	4.9	123.7	64.4	1.7
317	2016/6	27.0	19.2	22.5	6.8	108.8	80.0	1.5
318	2016/7	30.7	23.1	26.3	6.5	181.8	86.7	1.4
319	2016/8	32.6	23.5	27.3	3.6	110.0	77.3	1.4
320	2016/9	26.0	19.4	22.2	7.2	522.7	84.0	1.1
321	2016/10	21.3	14.4	17.4	6.3	215.8	76.0	1.4
322	2016/11	14.2	5.7	9.8	4.0	35.1	62.4	1.6
323	2016/12	10.1	1.2	5.4	3.6	136.8	54.4	1.8

-년별 지역 별 범죄 비율

	강도	살인	강간	절도	폭행
0	4.38	2.75	9.58	148.83	514.21
1	4.73	1.95	7.29	145.60	425.07
2	5.38	2.43	9.81	165.91	549.36
3	4.05	1.66	9.58	146.87	454.91
4	4.05	2.98	9.74	174.43	527.99
5	3.98	2.62	12.76	200.97	563.80
6	4.43	3.23	12.60	233.20	569.36
7	4.57	3.33	13.61	243.01	585.45
8	4.43	2.88	10.80	221.42	557.62
9	4.85	2.62	11.18	212.77	545.38

	강도	살인	강간	절도	폭행
0	4.38	2.75	9.58	148.83	514.21
1	4.73	1.95	7.29	145.60	425.07
2	5.38	2.43	9.81	165.91	549.36
3	4.05	1.66	9.58	146.87	454.91
4	4.05	2.98	9.74	174.43	527.99
5	3.98	2.62	12.76	200.97	563.80
6	4.43	3.23	12.60	233.20	569.36
7	4.57	3.33	13.61	243.01	585.45
8	4.43	2.88	10.80	221.42	557.62
9	4.85	2.62	11.18	212.77	545.38

4. 지역 별 기상 데이터와 월 별 범죄양 데이터를 통합

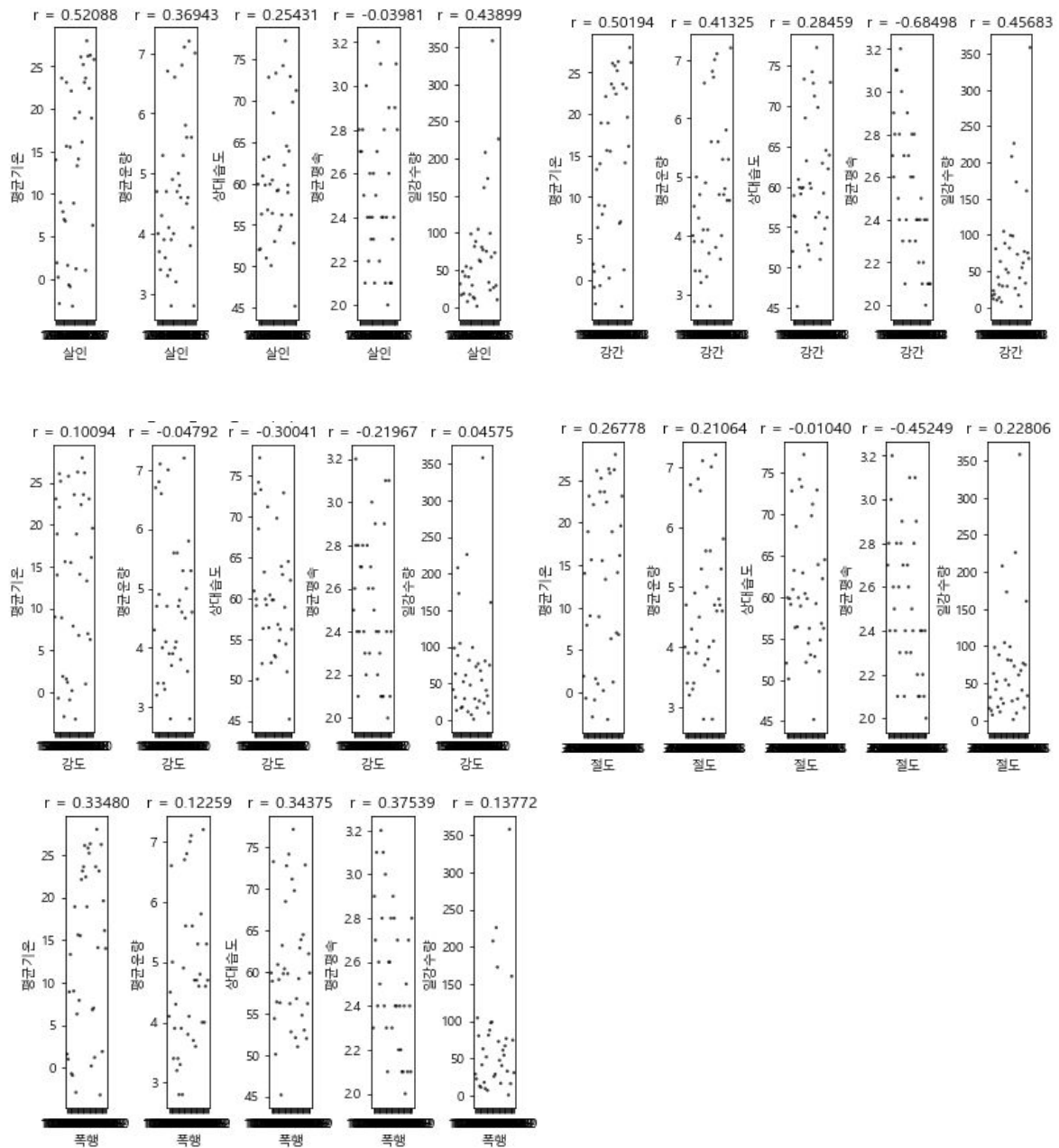
년월	최고기온	최저기온	평균기온	평균운량	일강수량	상대습도	평균평속	강도	살인	강간	절도	폭행
2016/3	13.9	2.6	7.6	4.0	67.0	61.5	1.9	3.19	1.85	10.98	234.09	422.14
2016/4	21.3	9.7	14.6	5.2	144.2	68.9	1.9	2.81	2.29	12.50	236.27	462.56
2016/5	25.7	13.6	19.2	4.6	84.3	66.8	2.1	3.70	2.31	13.98	258.58	494.22
2016/6	27.8	19.1	22.8	6.4	95.8	79.5	1.6	2.99	1.74	11.92	177.38	430.10
2016/7	31.0	23.2	26.5	6.6	251.8	82.4	1.7	3.01	2.55	15.28	212.13	480.44
2016/8	32.3	23.5	27.4	4.8	35.8	76.2	1.7	2.97	2.52	14.82	218.82	458.30
2016/9	27.3	19.0	22.6	6.5	145.4	78.7	1.4	3.22	2.42	13.86	263.61	470.68
2016/10	21.1	12.8	16.5	6.7	152.3	77.9	1.4	3.58	2.34	14.42	250.74	501.58
2016/11	13.5	4.2	8.6	4.4	37.5	68.7	1.7	2.80	1.87	10.96	240.28	421.22
2016/12	8.5	-0.4	3.8	4.8	46.8	69.4	1.8	2.36	2.16	11.98	209.73	425.00

5. 특정 지역에서 매 월의 기상과 범죄 데이터를 통해 상관관계 분석.

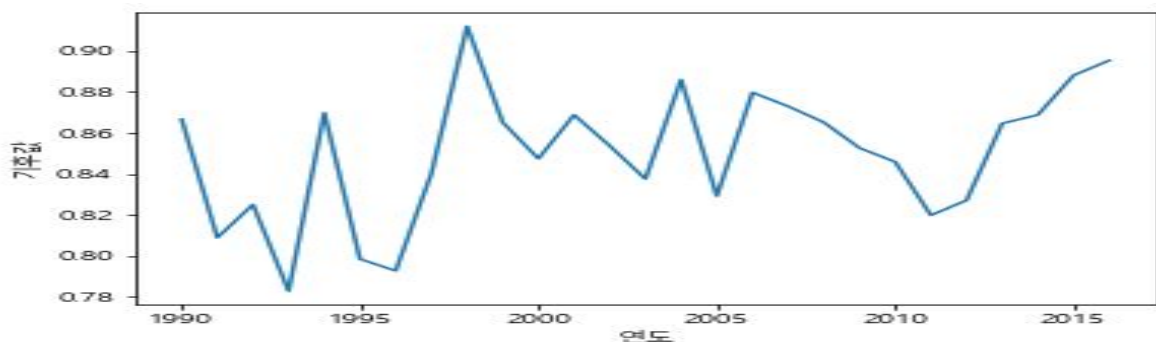
**가장 뚜렷한 정도 분석.

**범죄 데이터가 2014 년 기준으로 최신화되어 14년도만 뽑게 됨.

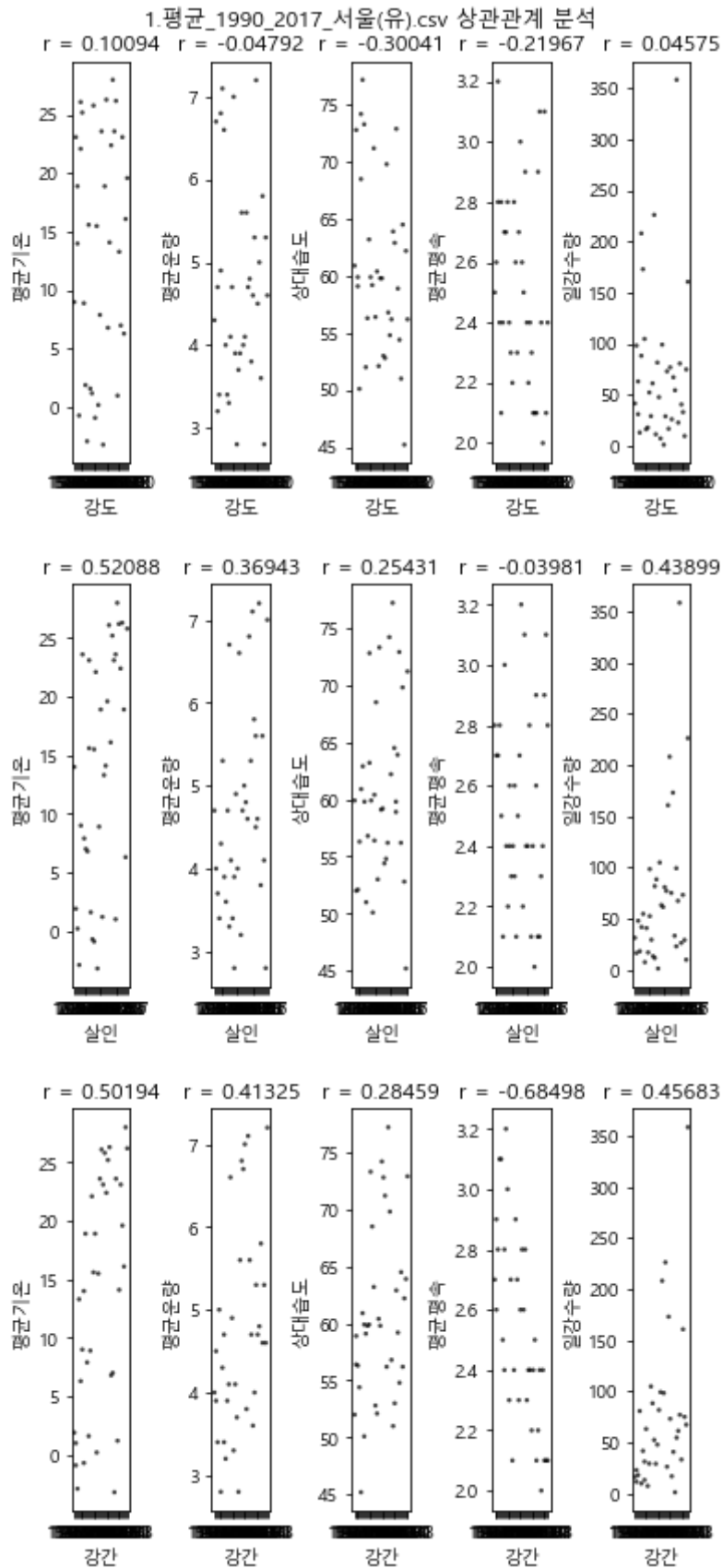
**인구가 가장 많은 서울을 기준으로 고려

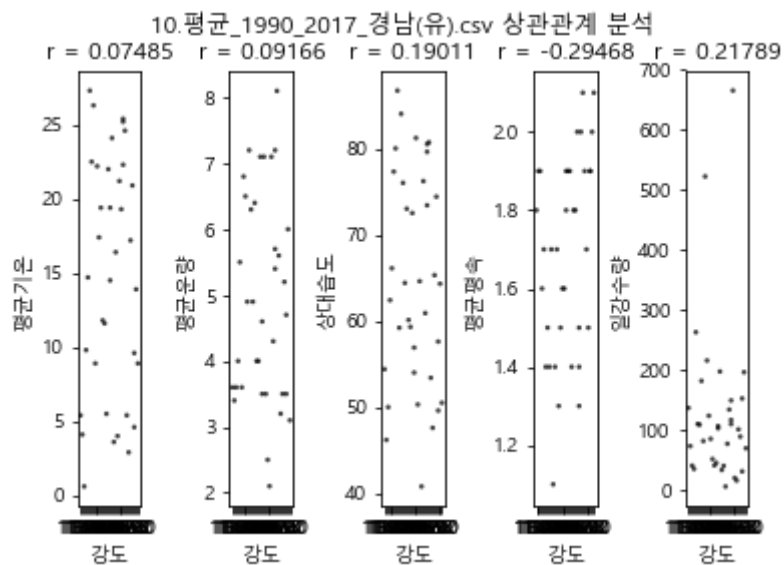
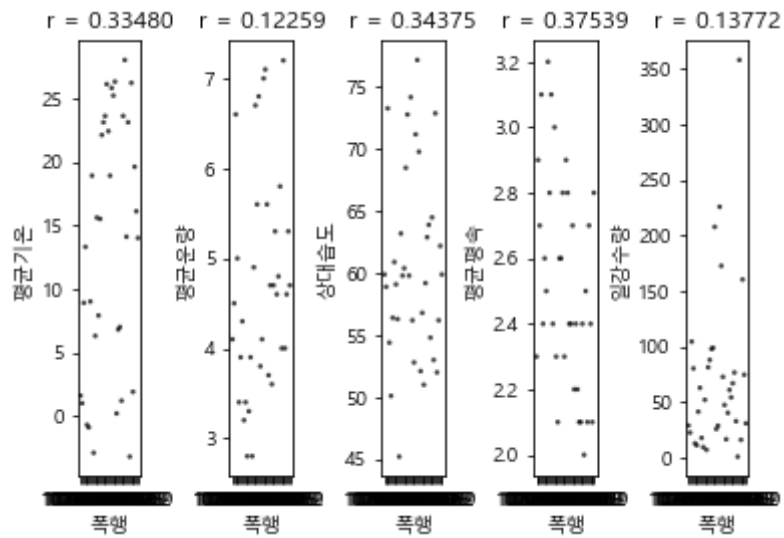
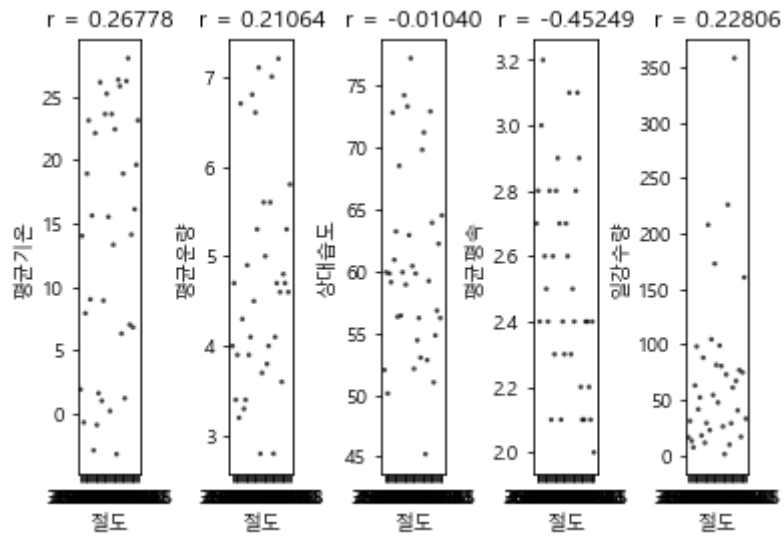


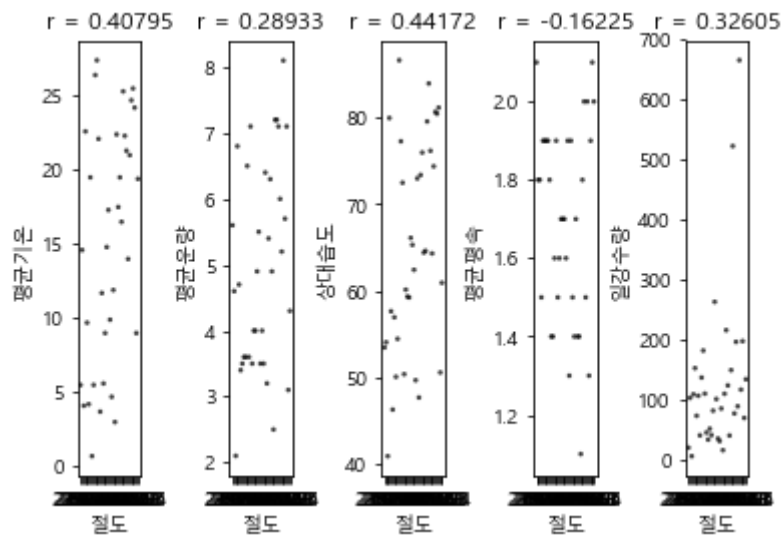
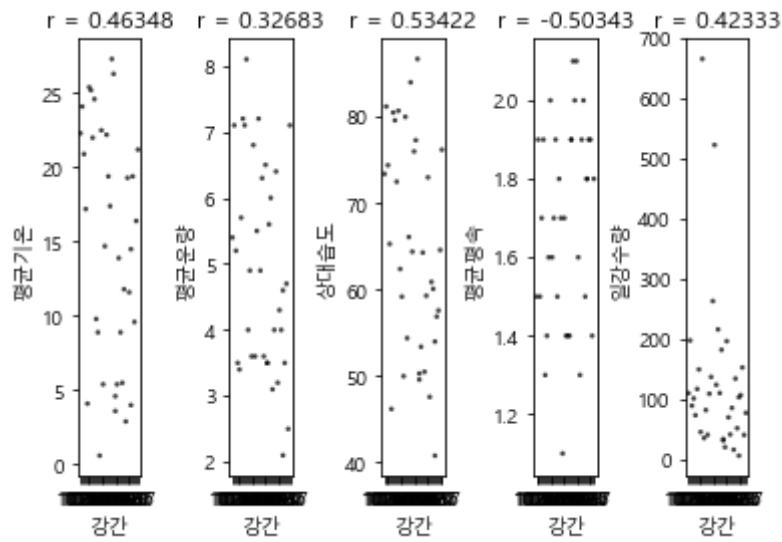
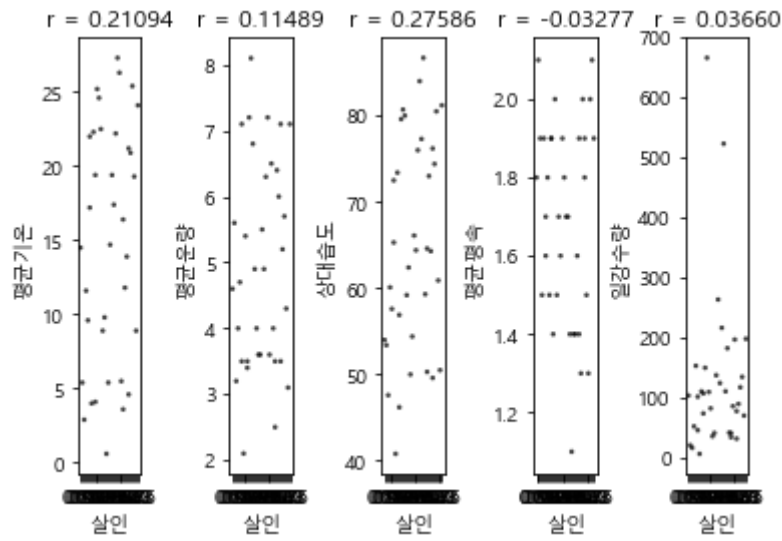
1990 - 2016 평균 온도 그래프

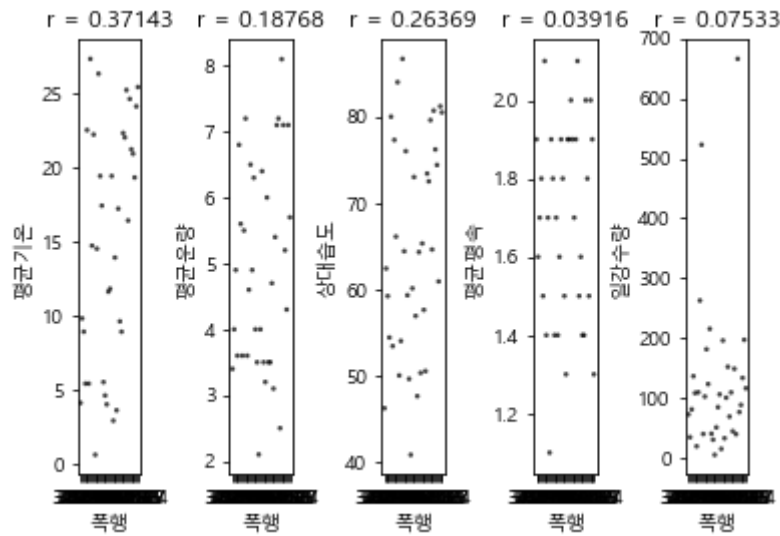


전체 상관계수 그래프

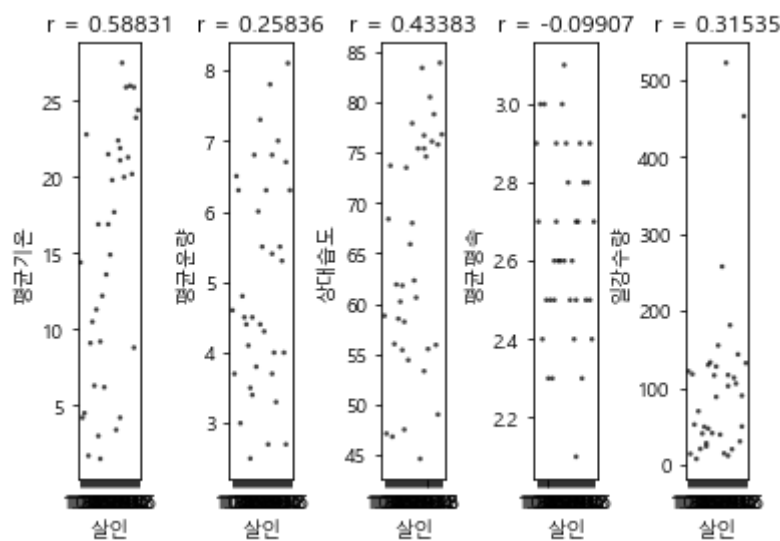
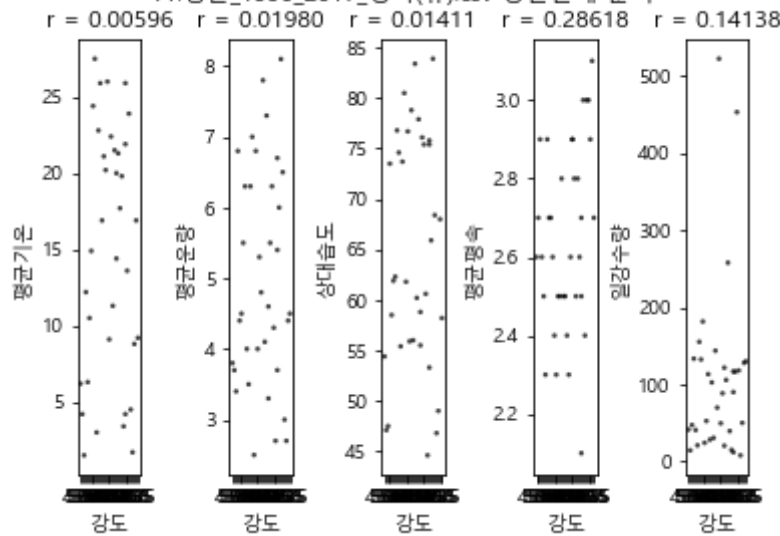


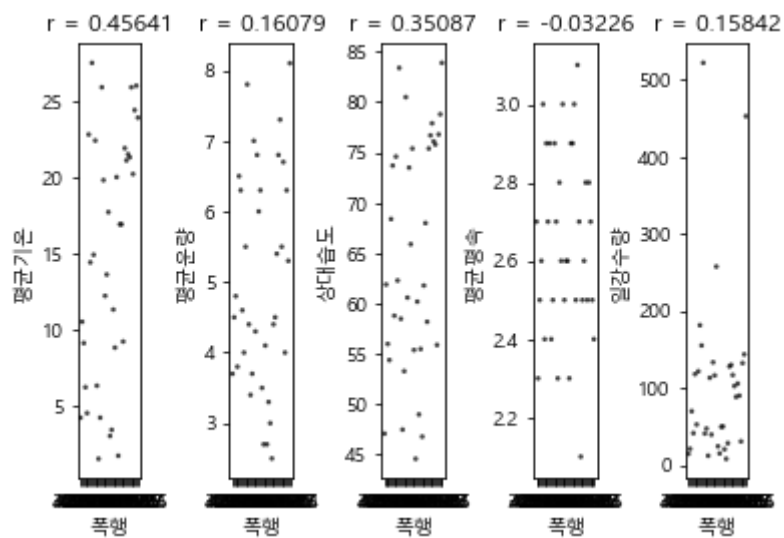
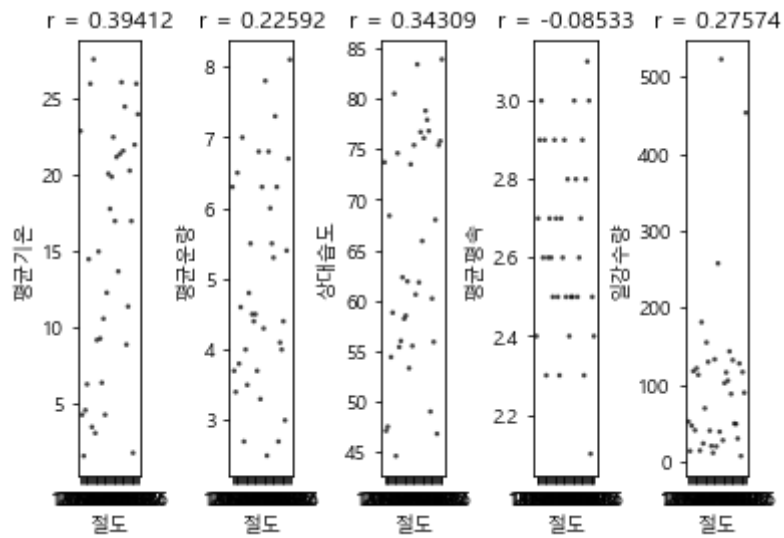
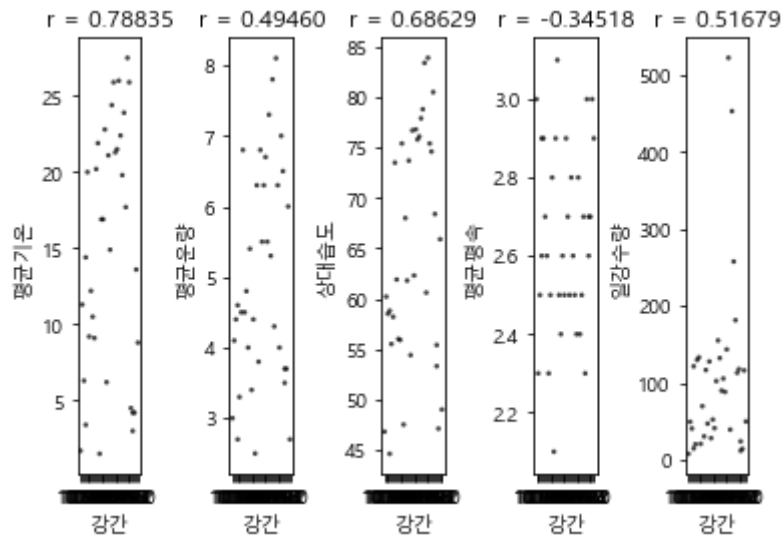






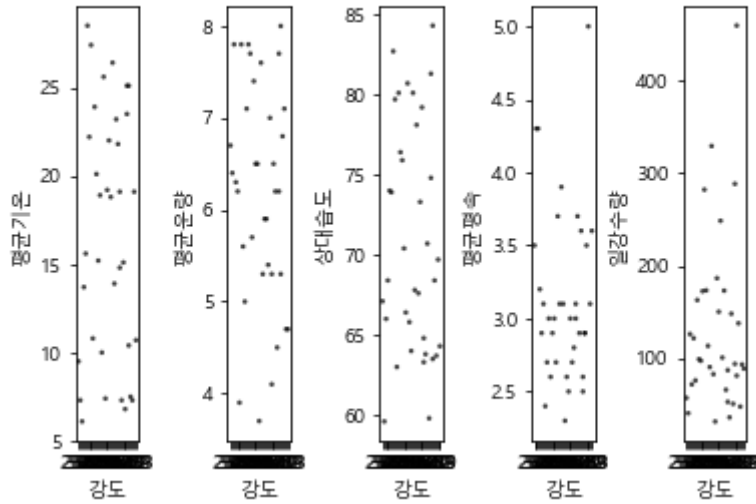
11. 평균_1990_2017_경북(유).csv 상관관계 분석



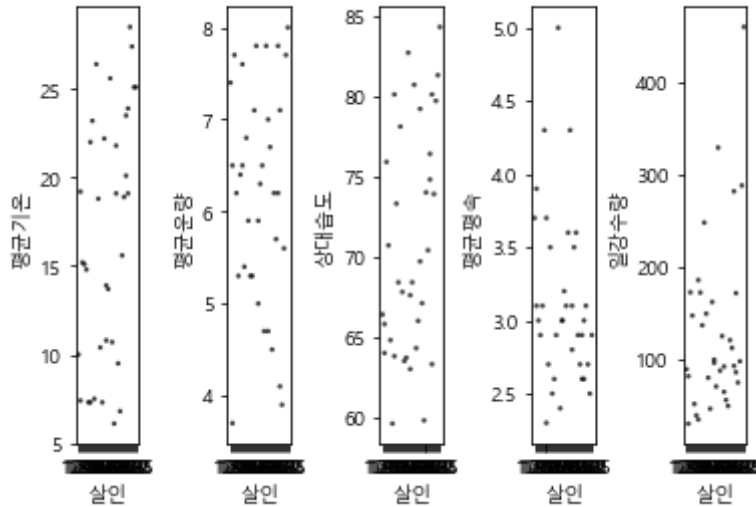


12. 평균_1990_2017_제주(유).csv 상관관계 분석

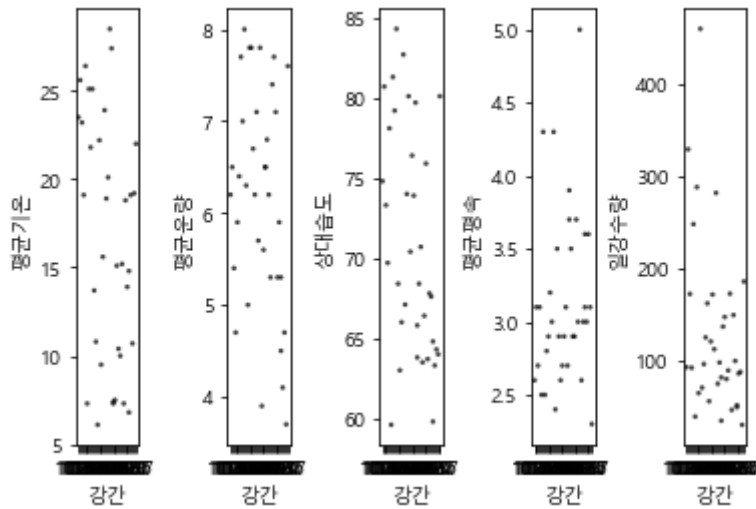
$r = -0.05906$ $r = -0.18110$ $r = -0.11639$ $r = 0.02675$ $r = 0.08006$

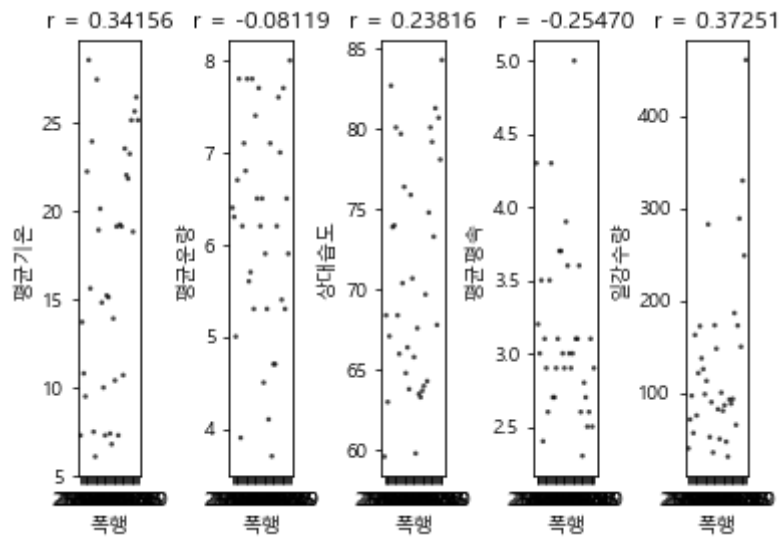
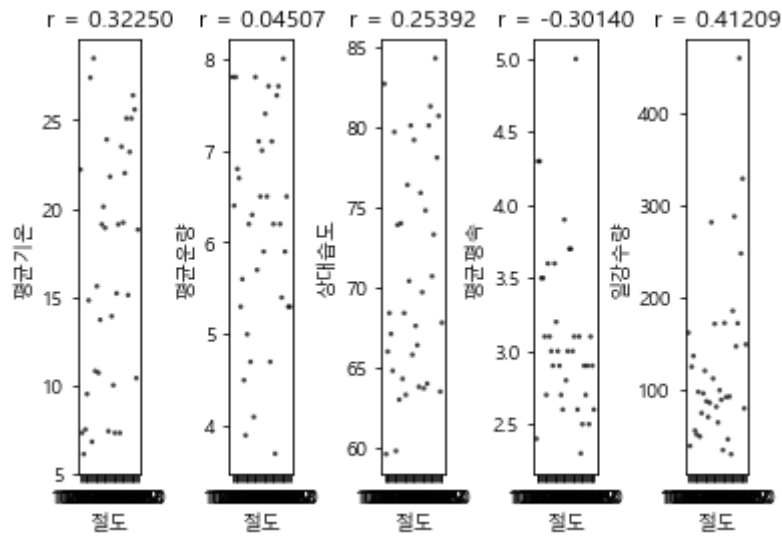


$r = 0.45343$ $r = -0.01037$ $r = 0.39613$ $r = -0.25718$ $r = 0.31303$

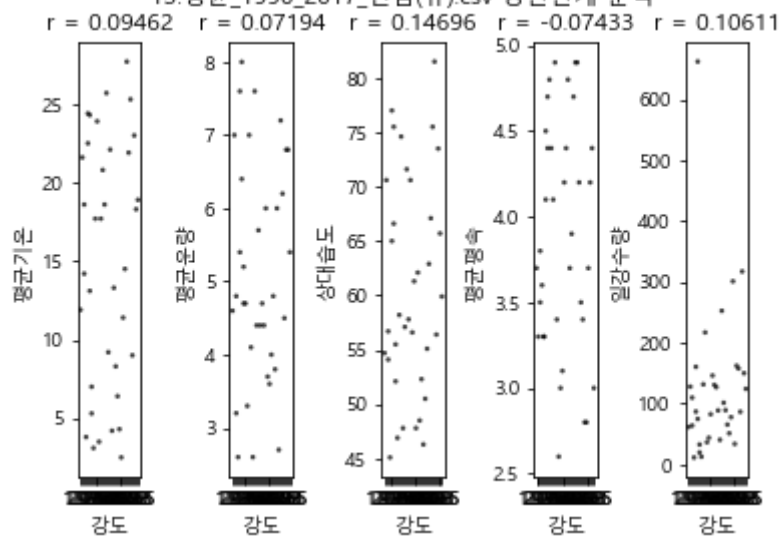


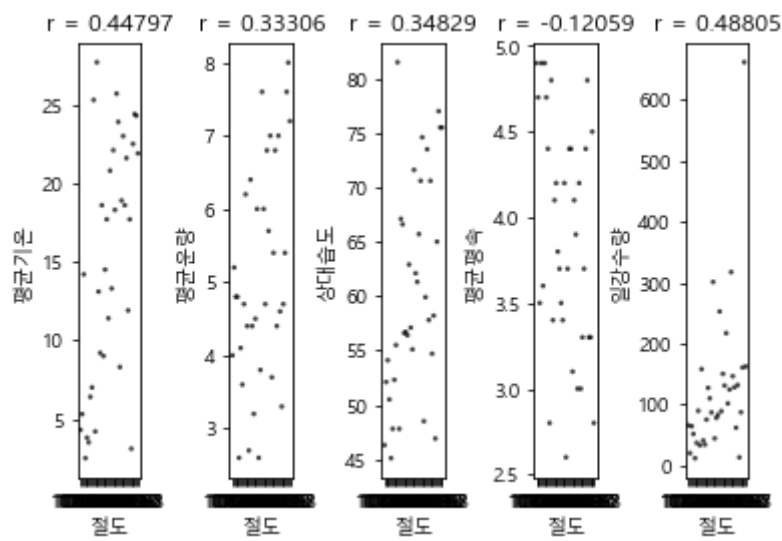
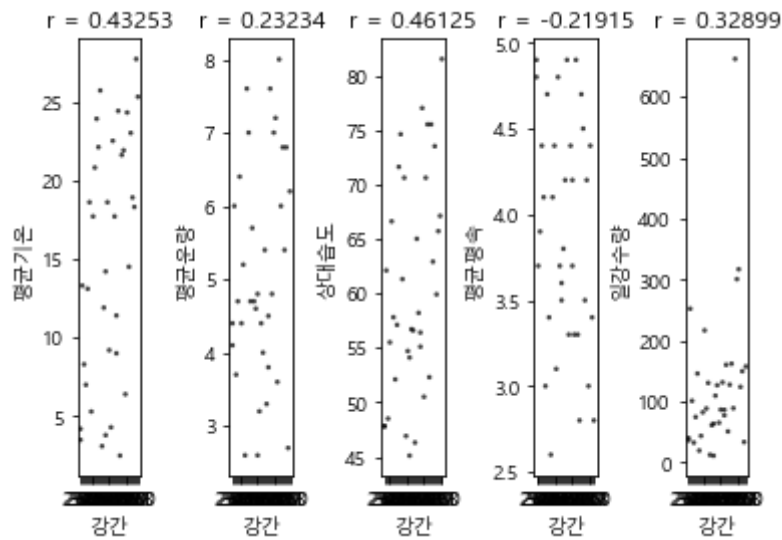
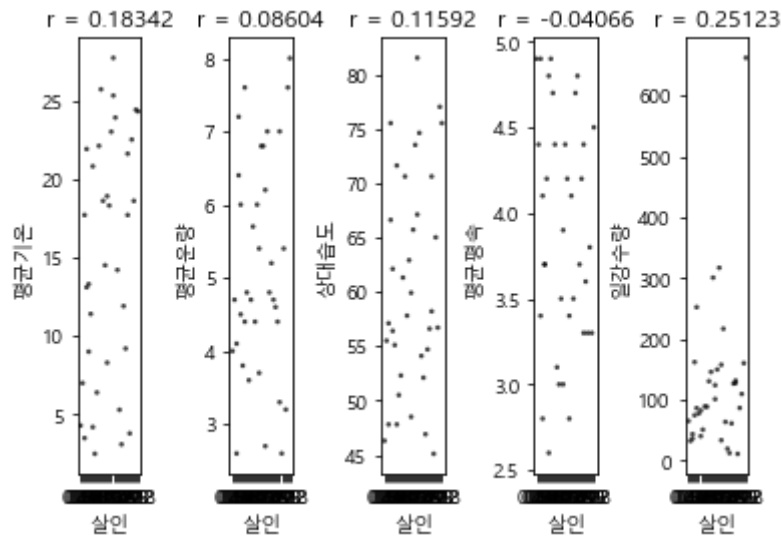
$r = 0.46214$ $r = 0.08597$ $r = 0.43776$ $r = -0.27640$ $r = 0.17570$

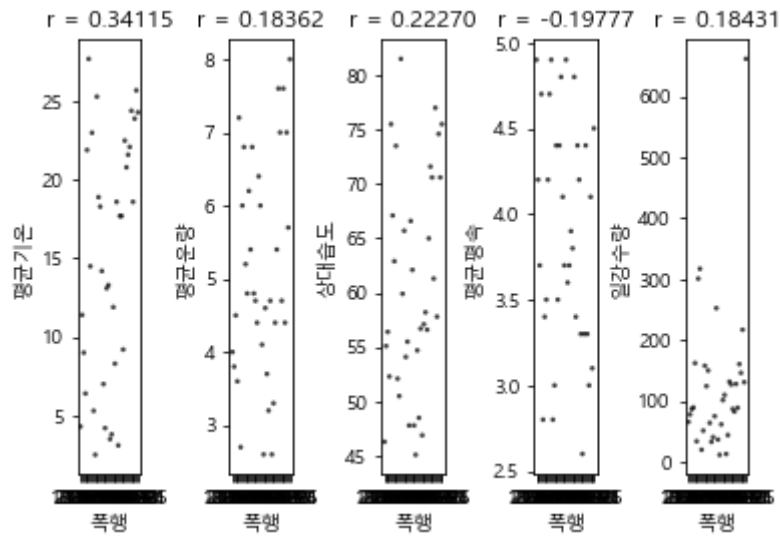




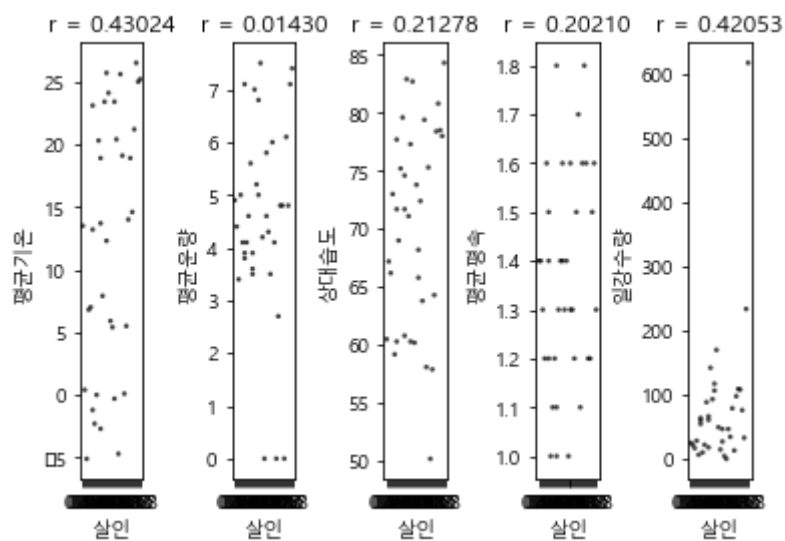
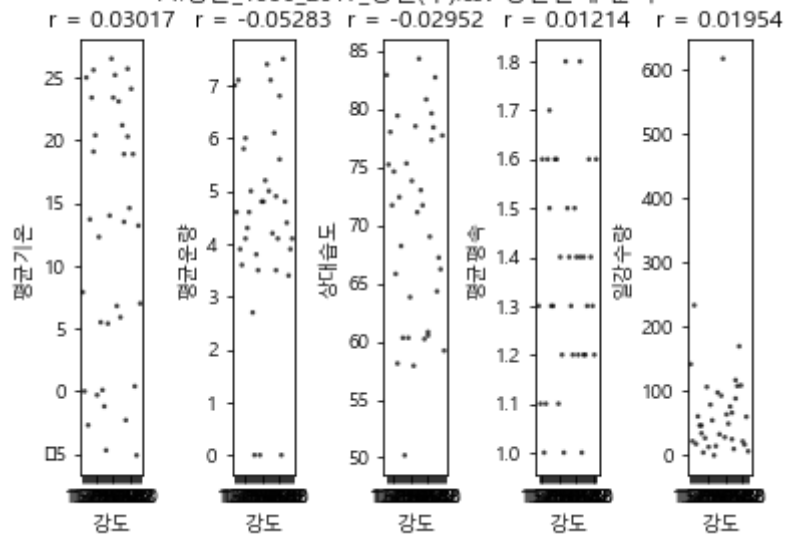
13. 평균_1990_2017_전남(유).csv 상관관계 분석

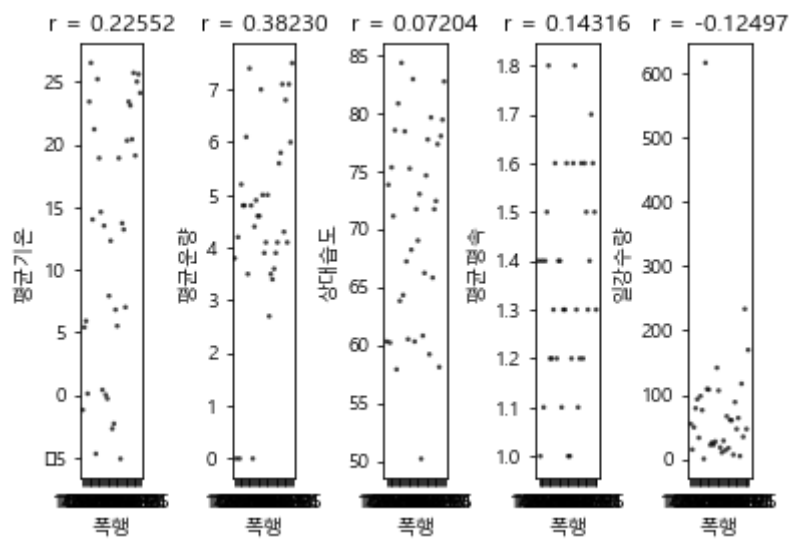
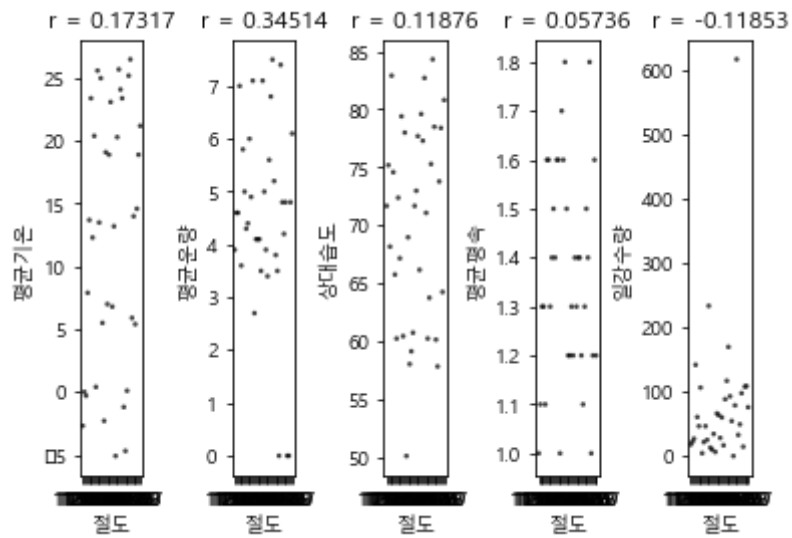
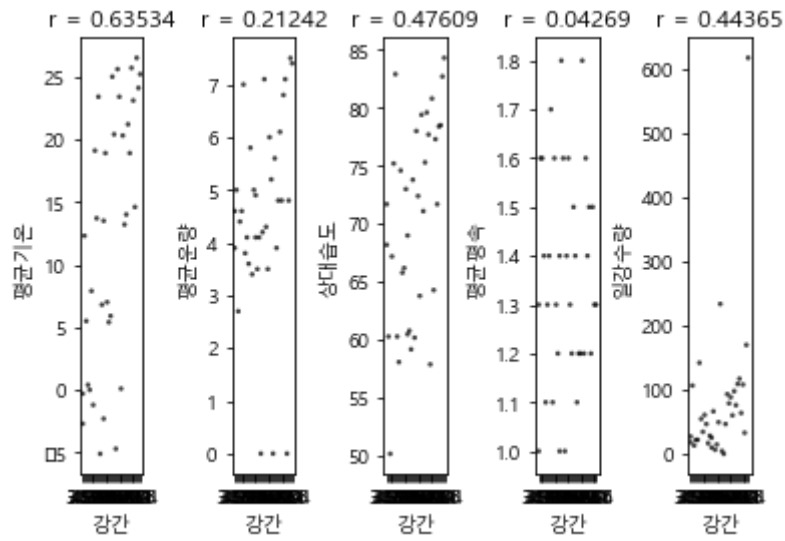






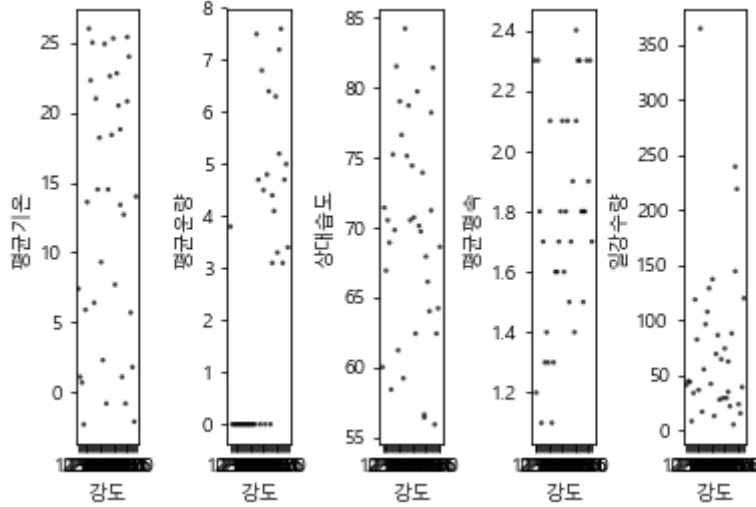
14. 평균_1990_2017_강원(무).csv 상관관계 분석



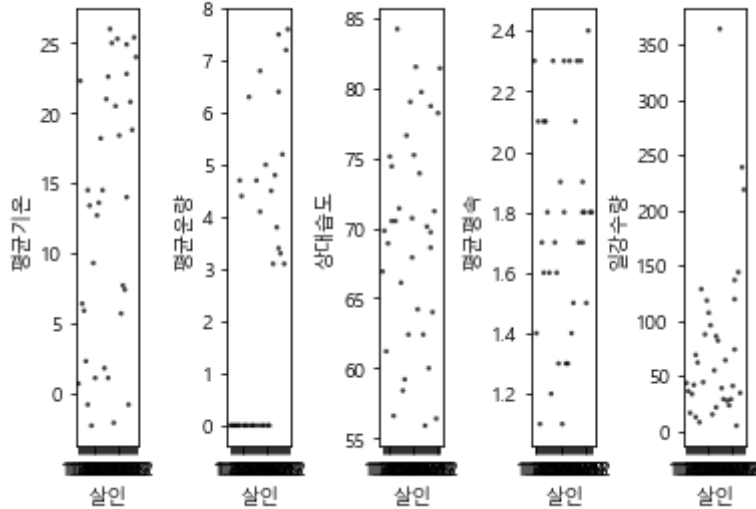


15. 평균_1990_2017_충남(무).csv 상관관계 분석

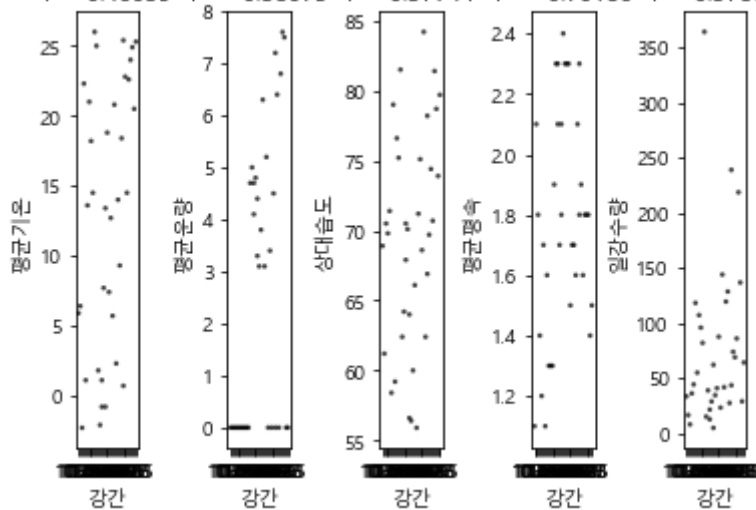
$r = -0.01727$ $r = 0.72309$ $r = -0.18809$ $r = 0.54097$ $r = -0.04234$

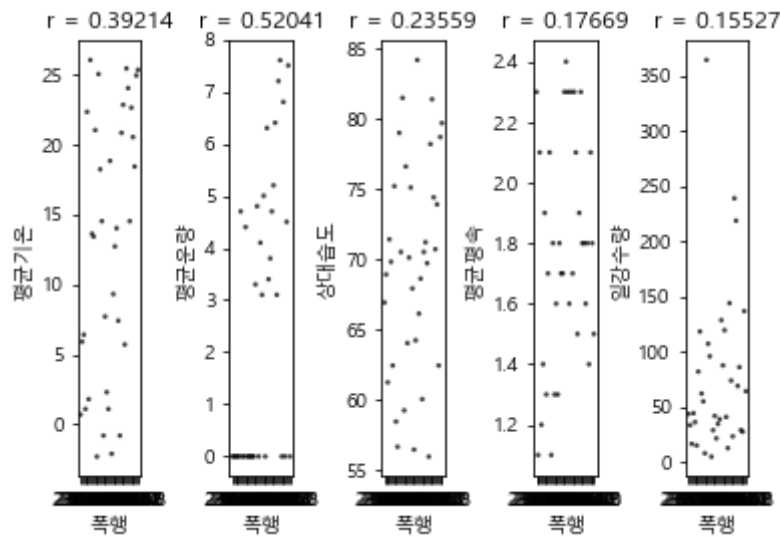
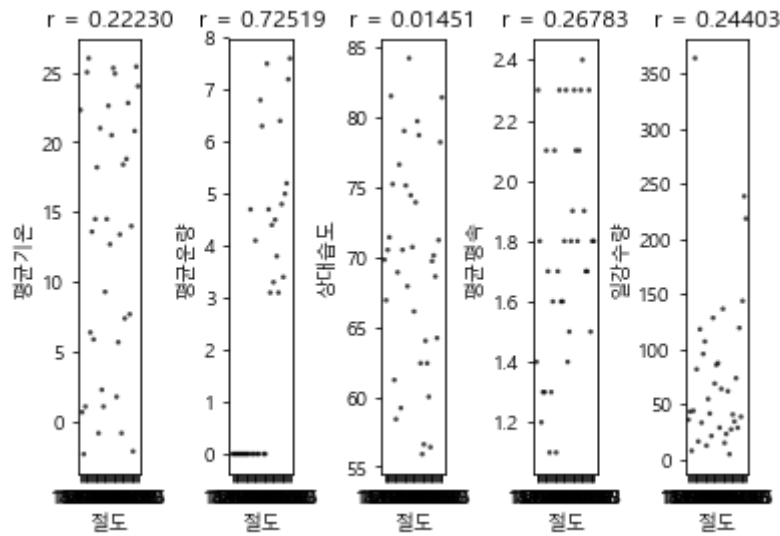


$r = 0.43352$ $r = 0.62598$ $r = 0.10788$ $r = 0.11145$ $r = 0.39862$

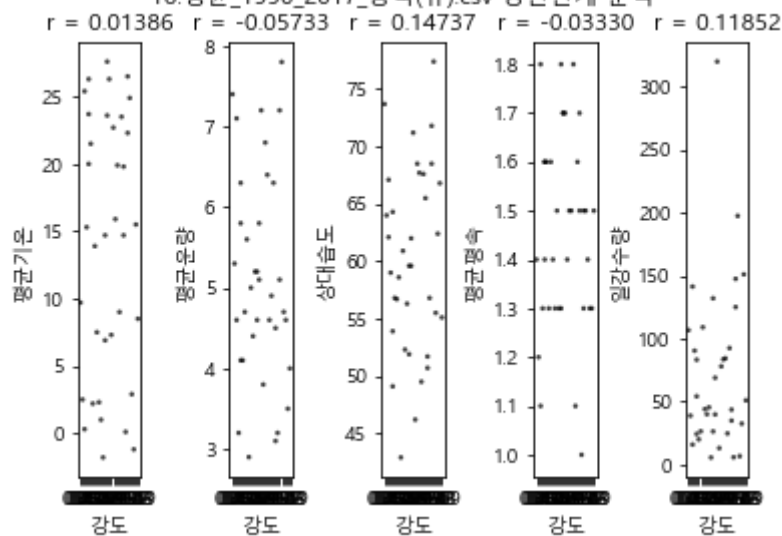


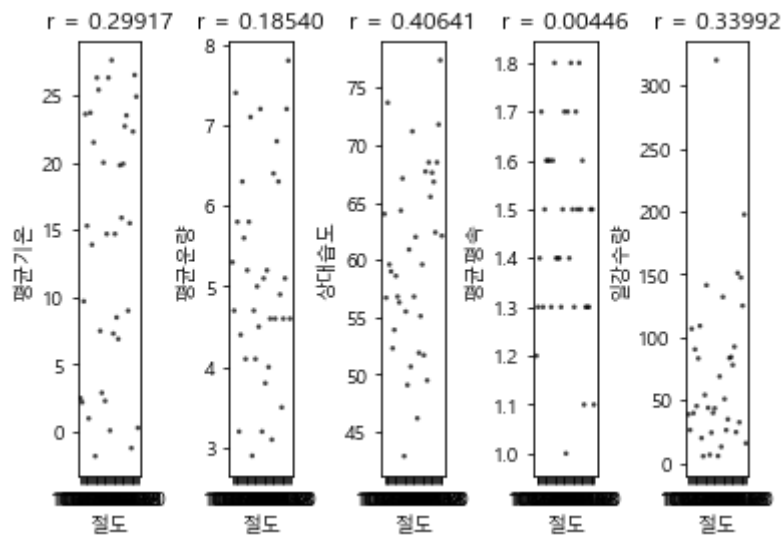
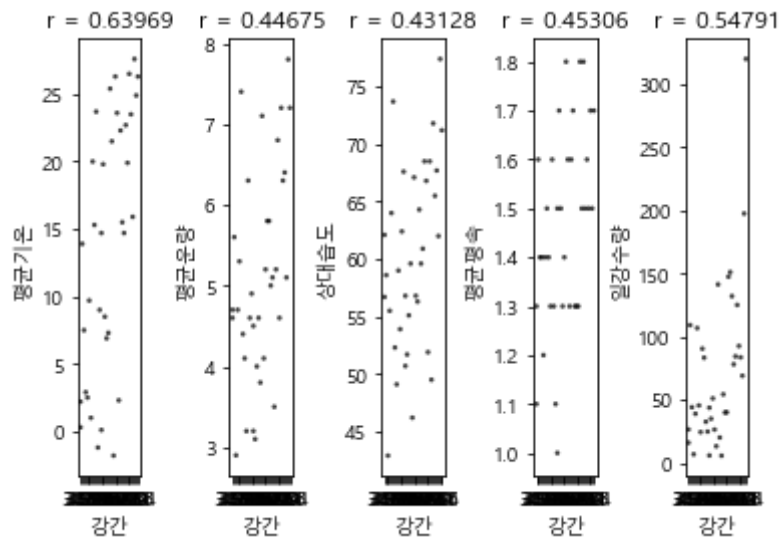
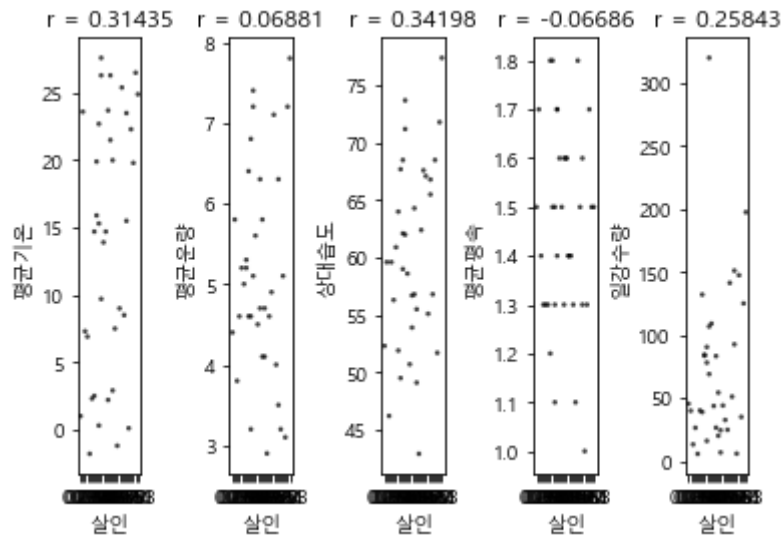
$r = 0.46059$ $r = -0.58075$ $r = 0.37744$ $r = -0.70186$ $r = 0.37802$

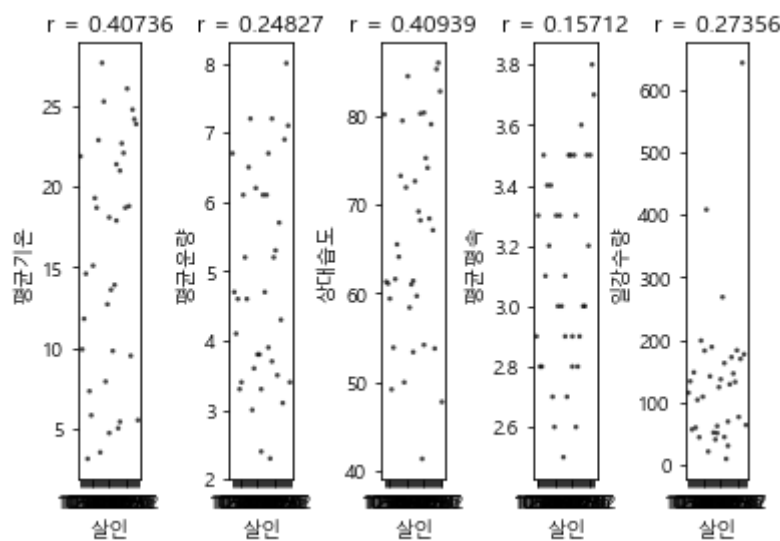
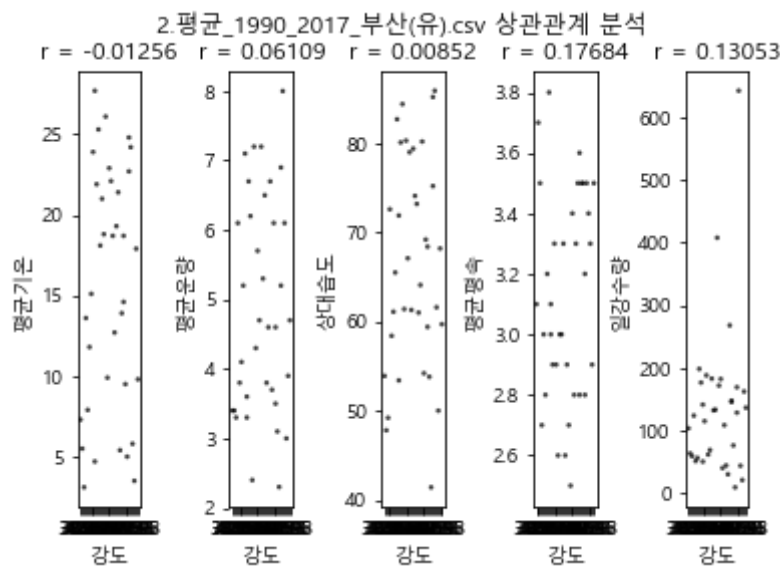
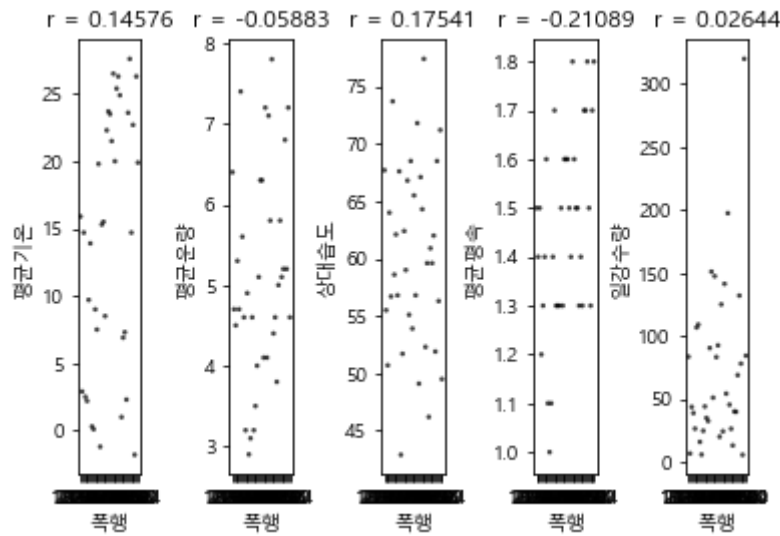


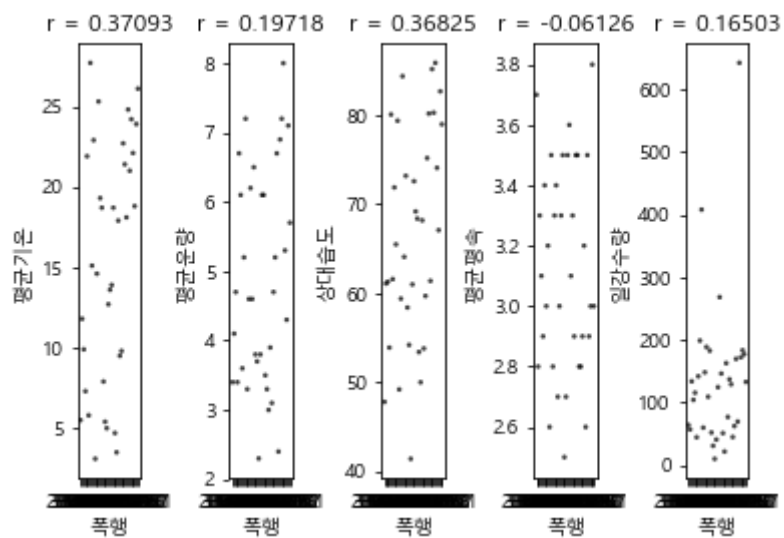
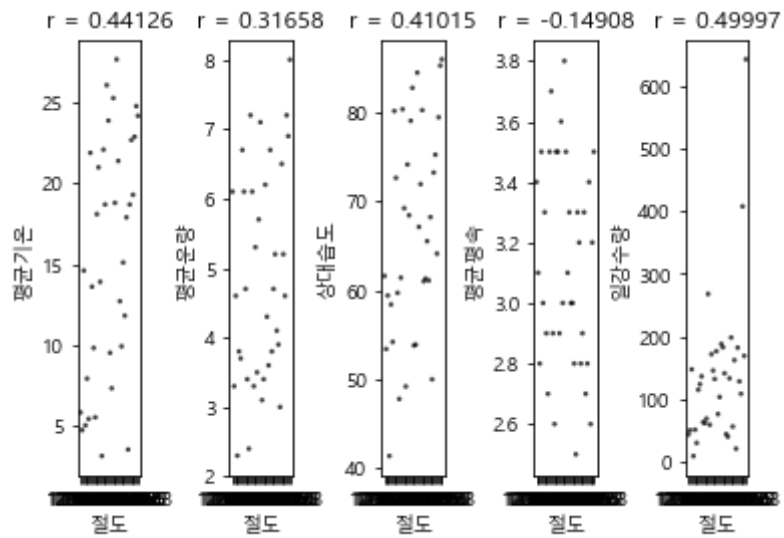
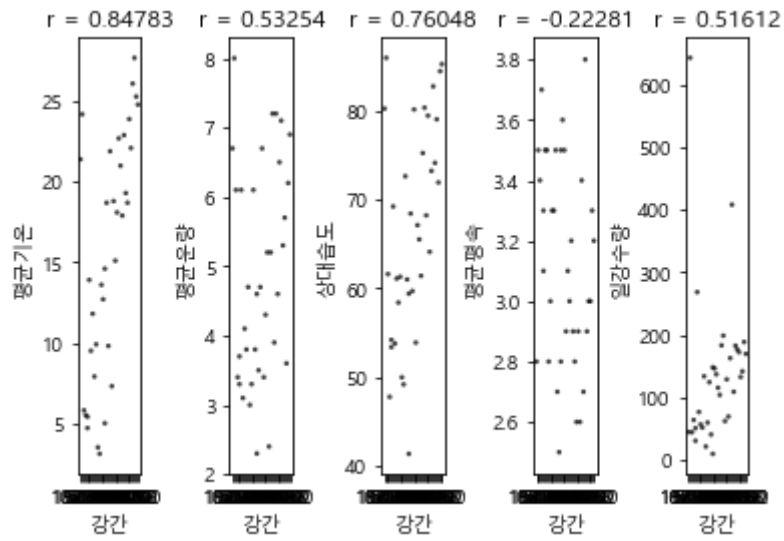


16. 평균_1990_2017_충북(유).csv 상관관계 분석



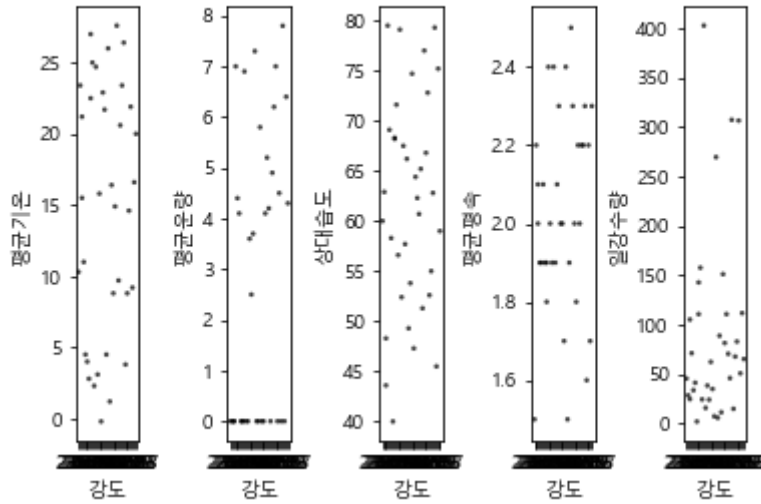




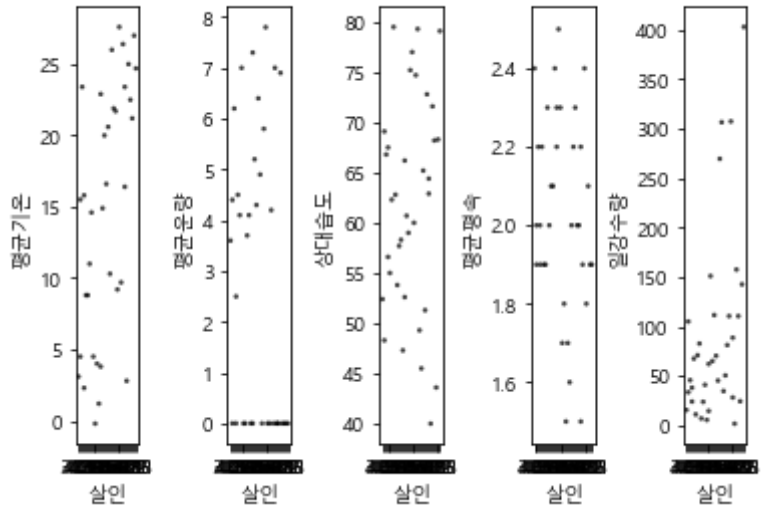


3. 평균_1990_2017_대구(유).csv 상관관계 분석

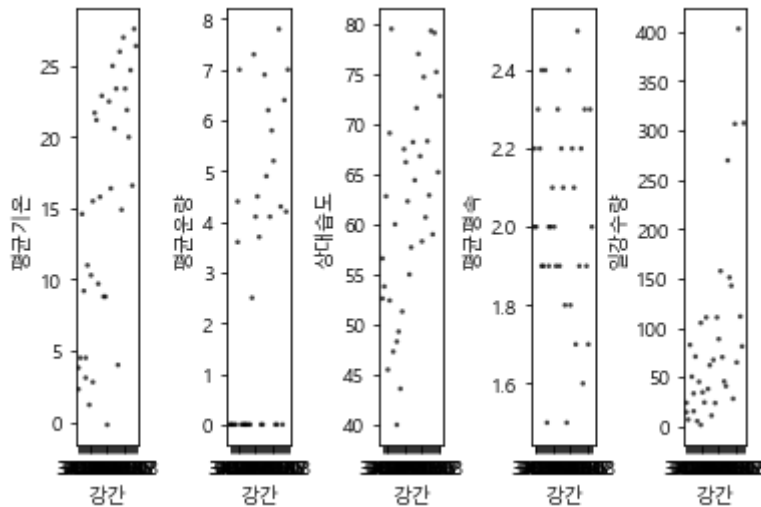
$r = 0.09120$ $r = 0.27091$ $r = 0.08431$ $r = 0.15654$ $r = 0.14278$

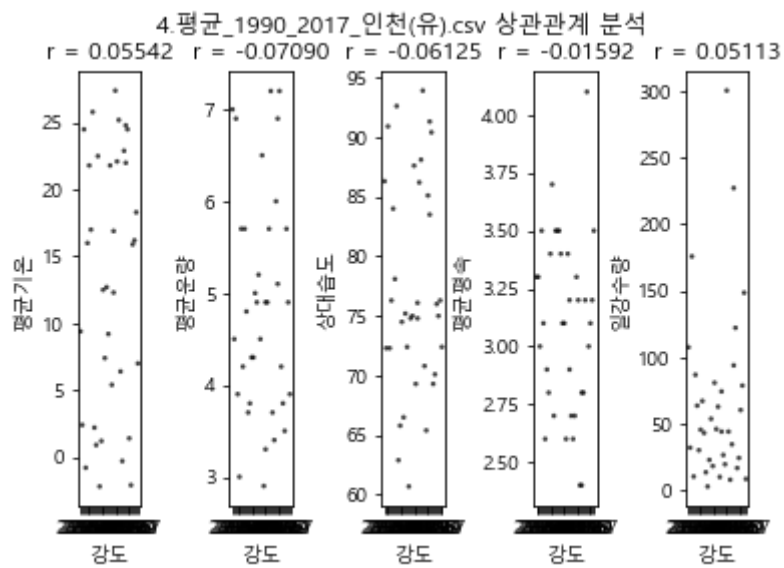
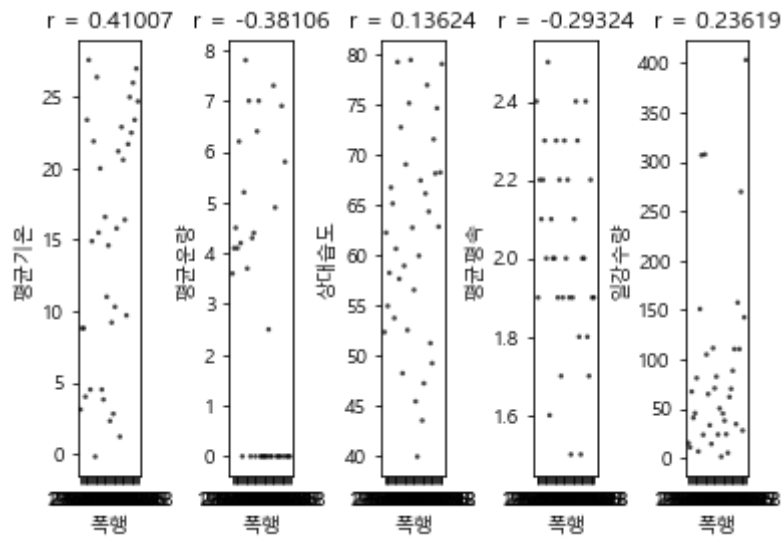
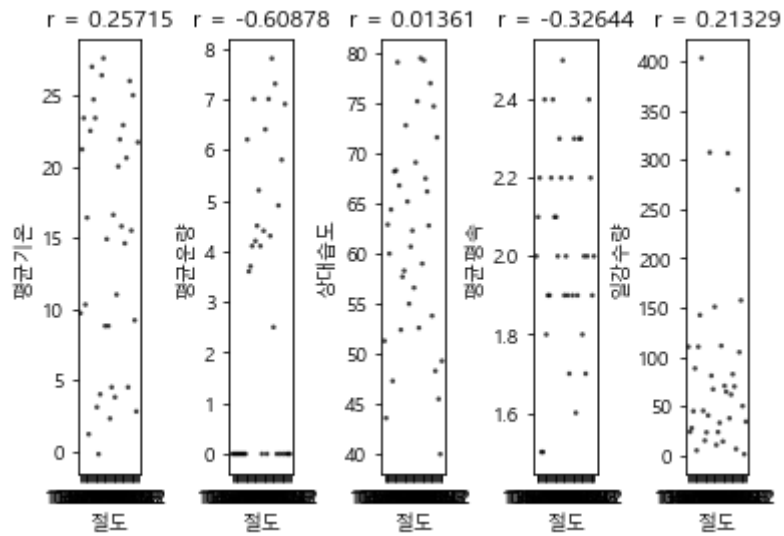


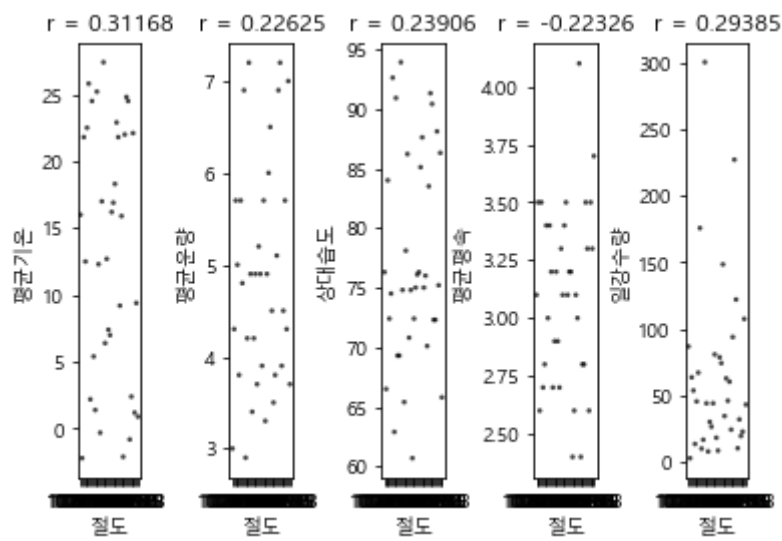
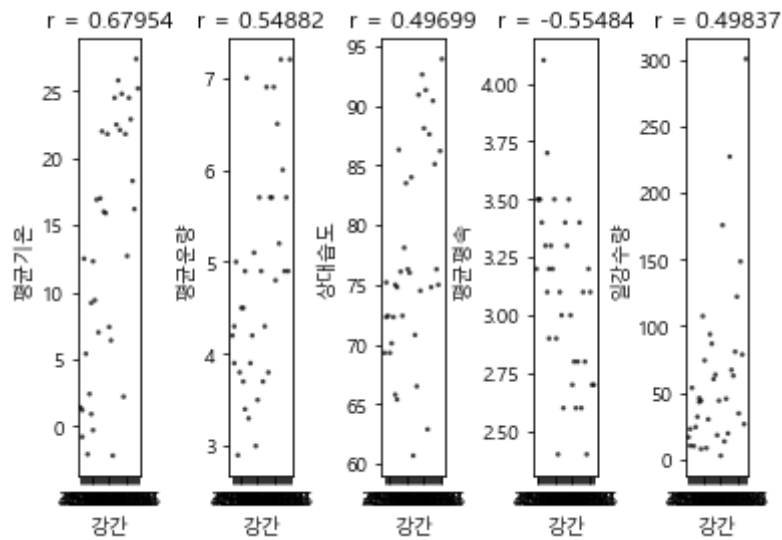
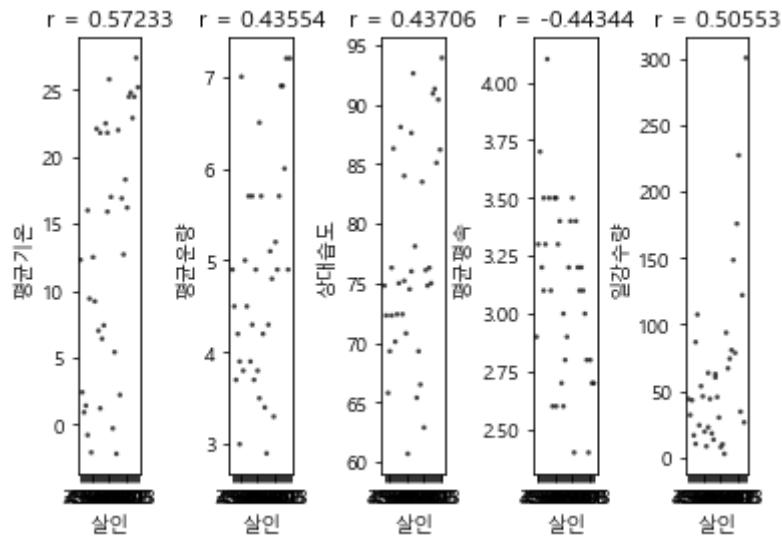
$r = 0.53240$ $r = -0.27719$ $r = 0.12853$ $r = -0.20329$ $r = 0.45516$

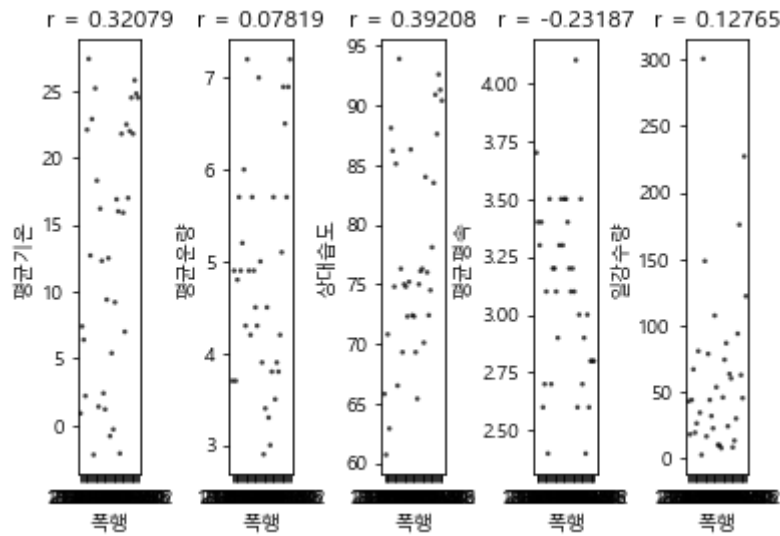


$r = 0.68538$ $r = 0.44707$ $r = 0.54511$ $r = -0.12359$ $r = 0.60728$

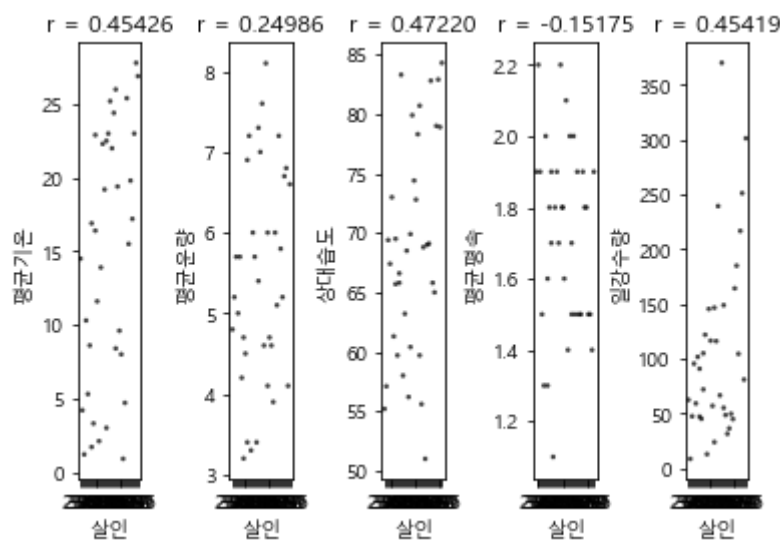
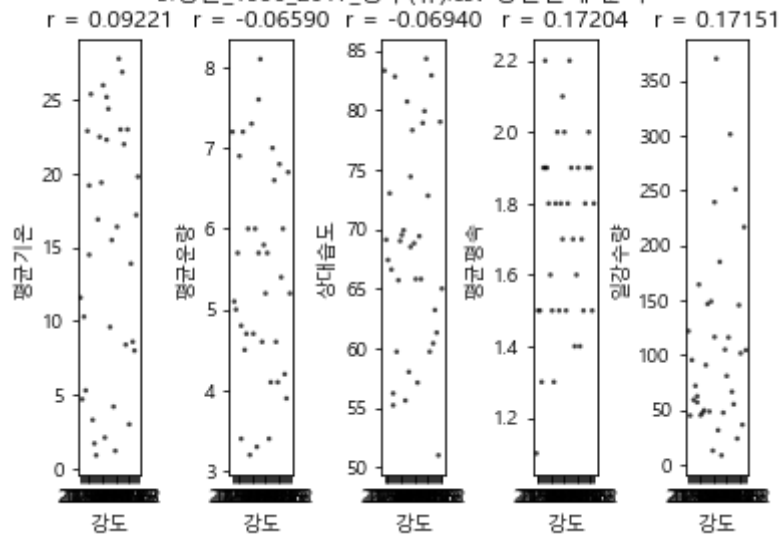


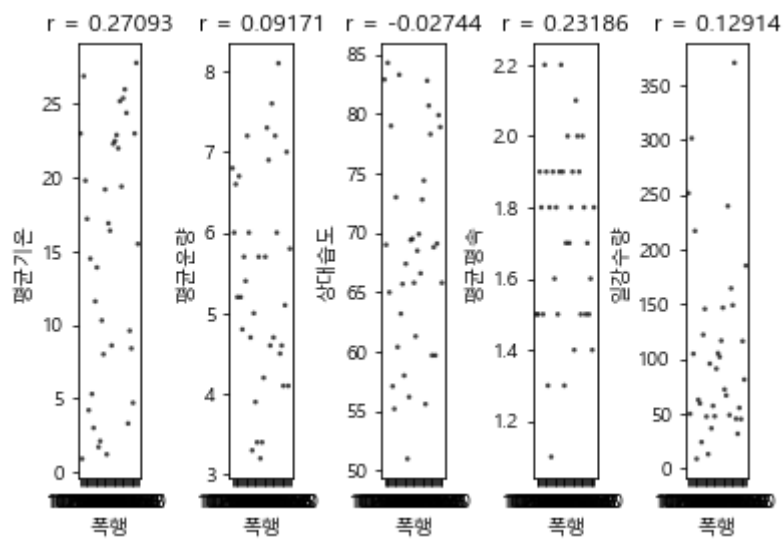
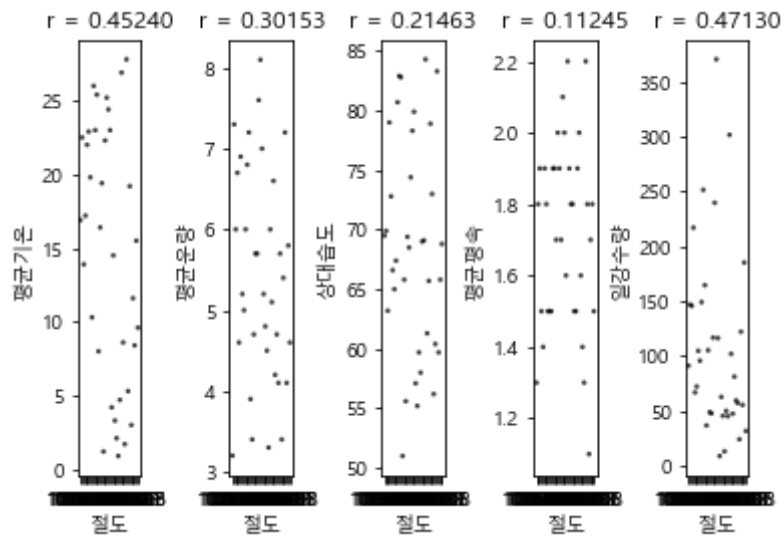
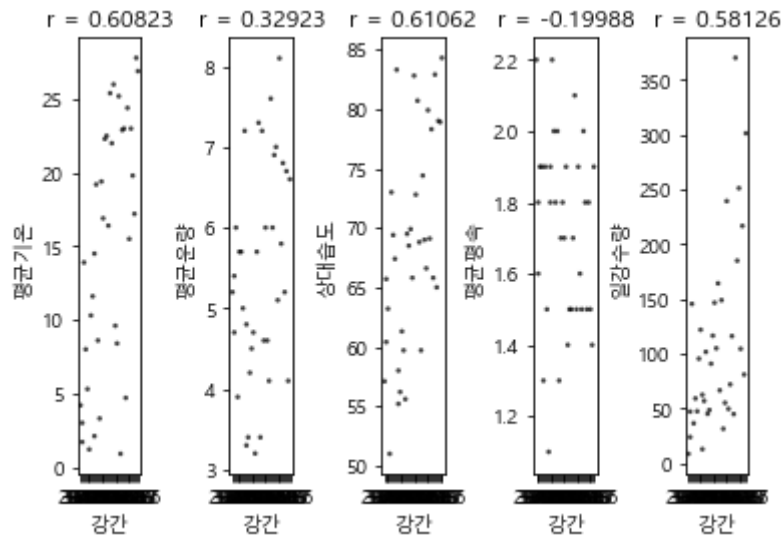






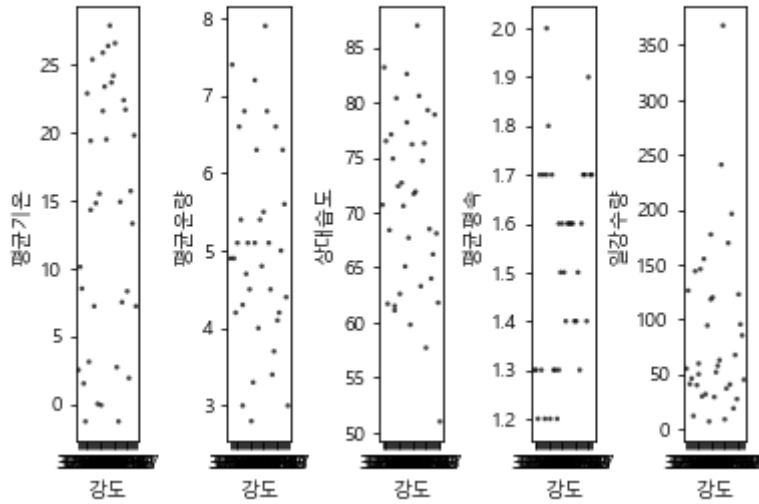
5. 평균_1990_2017_광주(유).csv 상관관계 분석



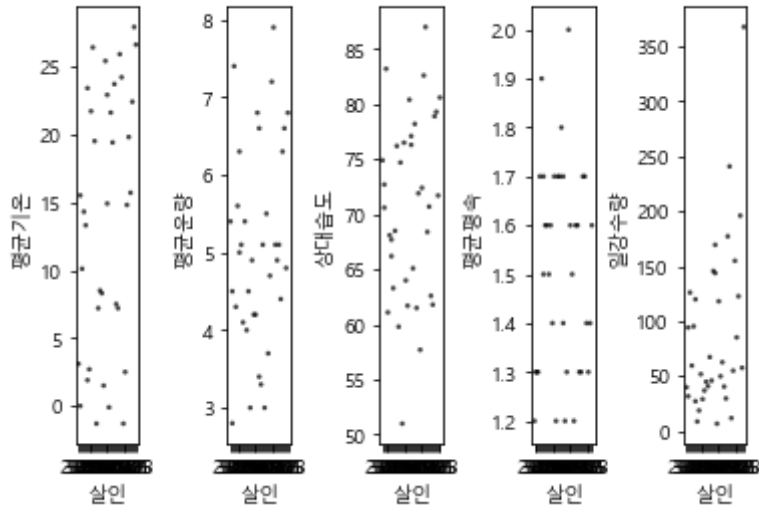


6. 평균_1990_2017_대전(유).csv 상관관계 분석

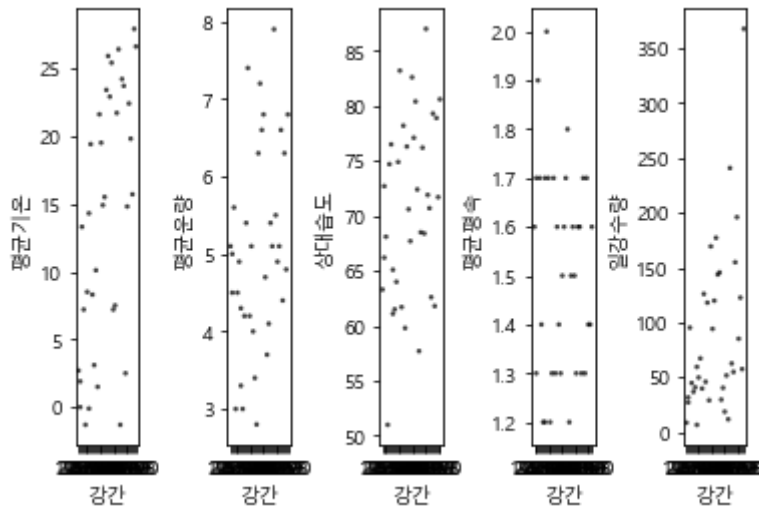
$r = 0.08697$ $r = -0.14834$ $r = -0.27388$ $r = 0.33104$ $r = 0.03235$

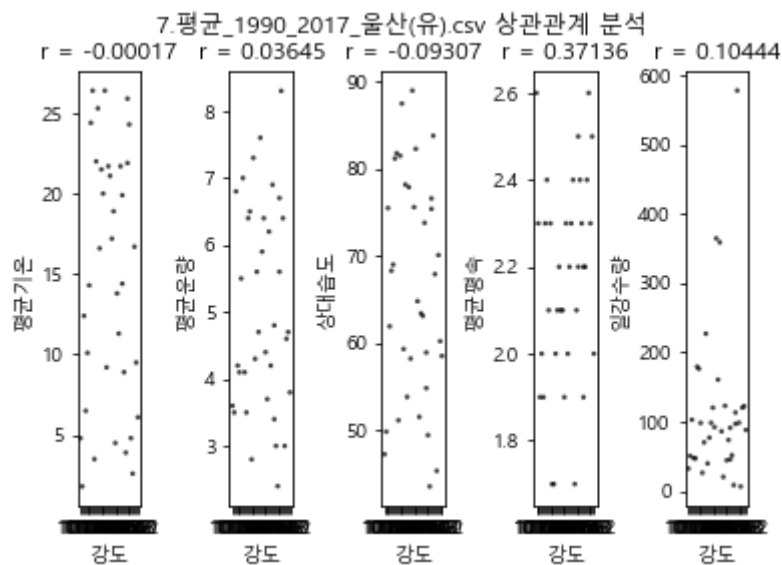
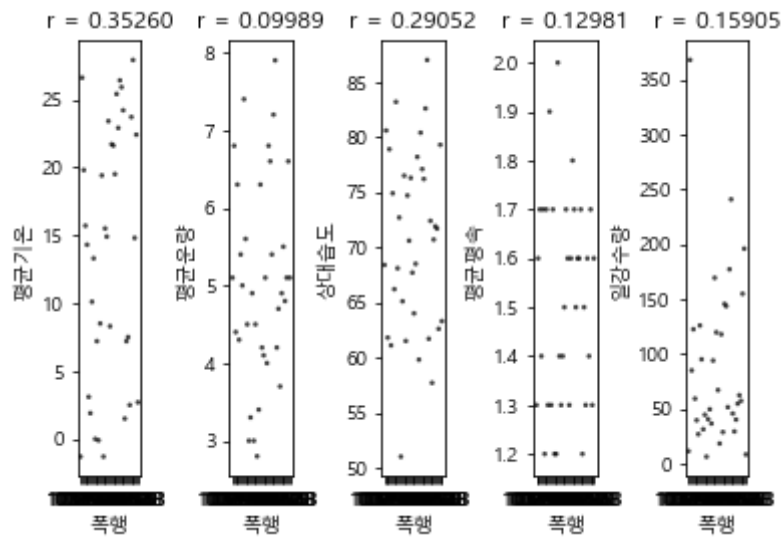
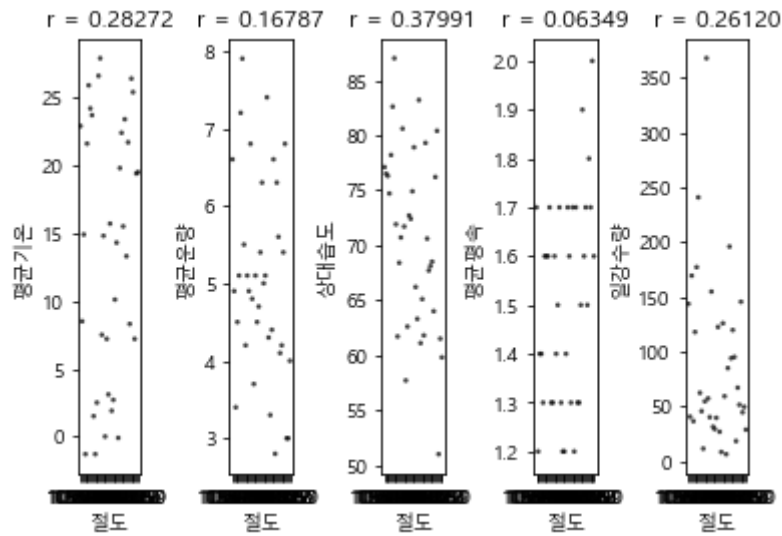


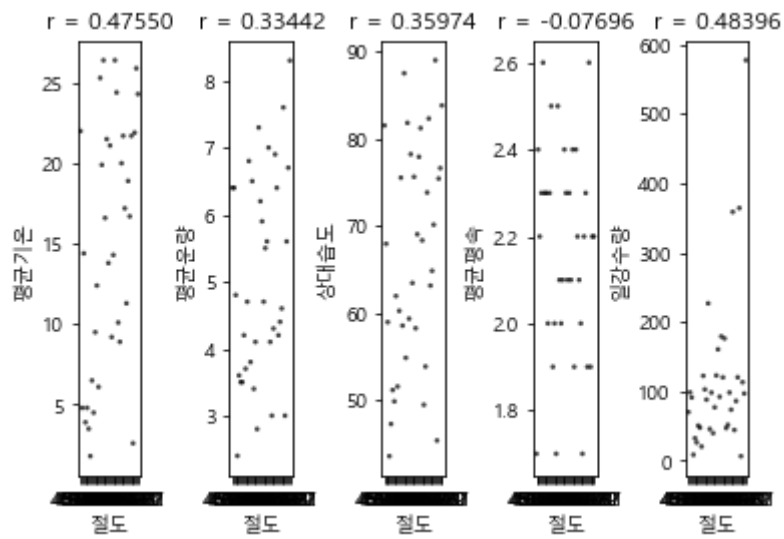
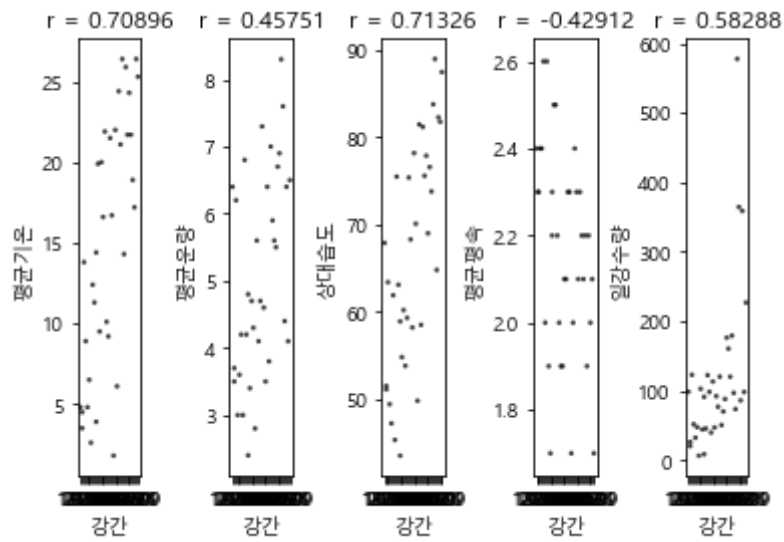
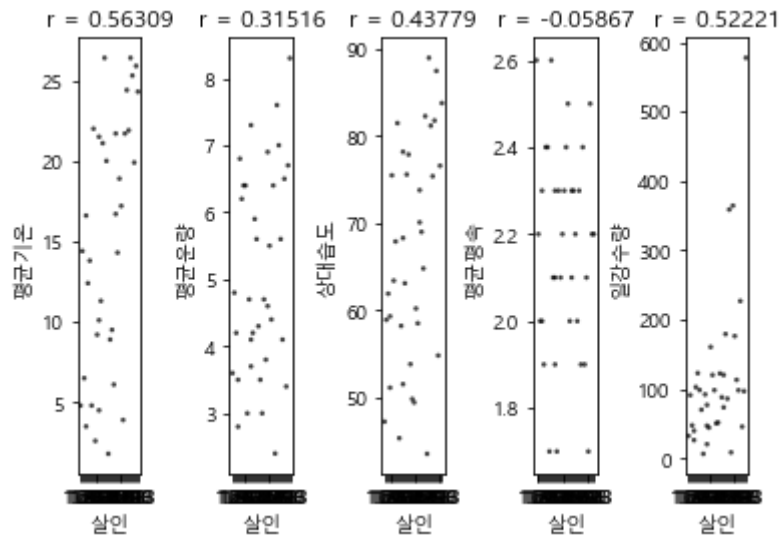
$r = 0.32977$ $r = 0.23508$ $r = 0.17298$ $r = -0.03682$ $r = 0.44465$

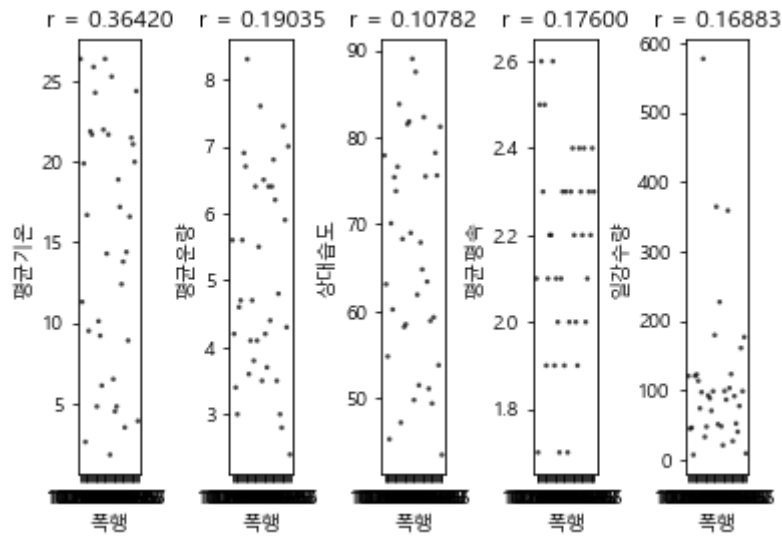


$r = 0.57532$ $r = 0.36082$ $r = 0.30832$ $r = -0.06853$ $r = 0.51947$

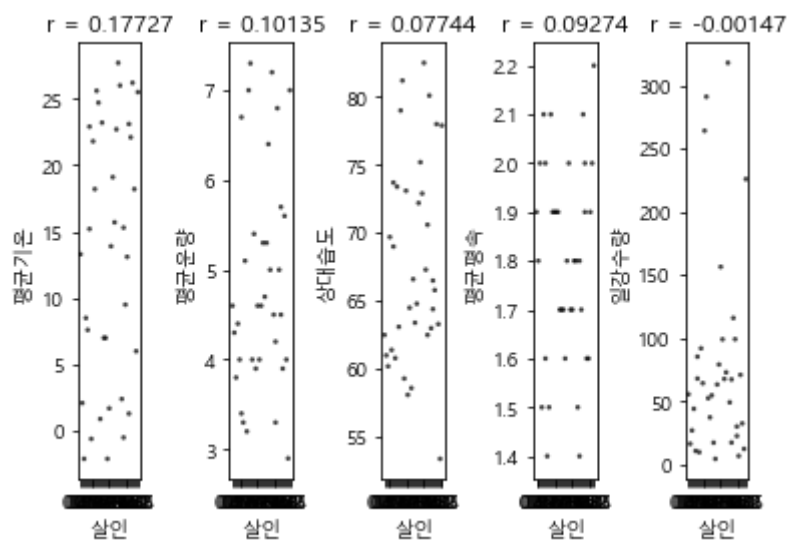
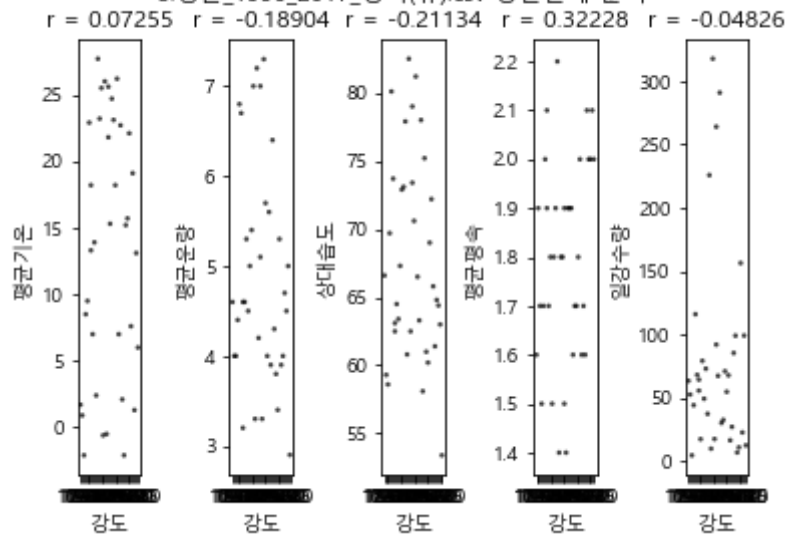


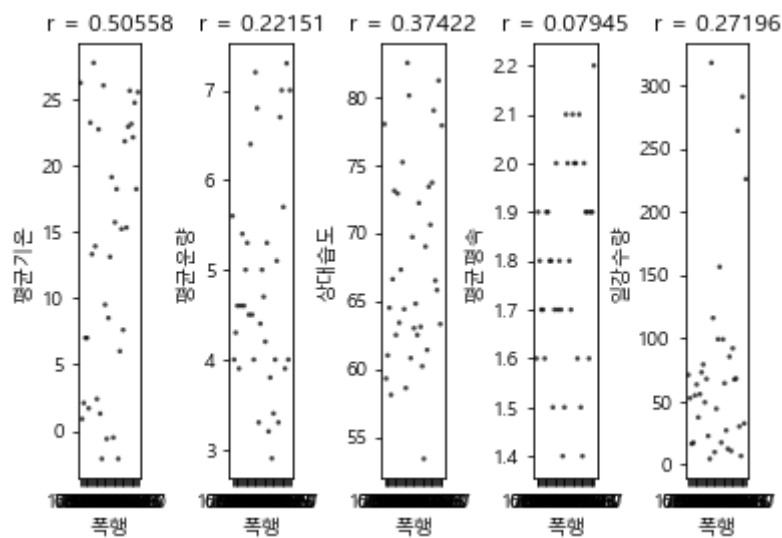
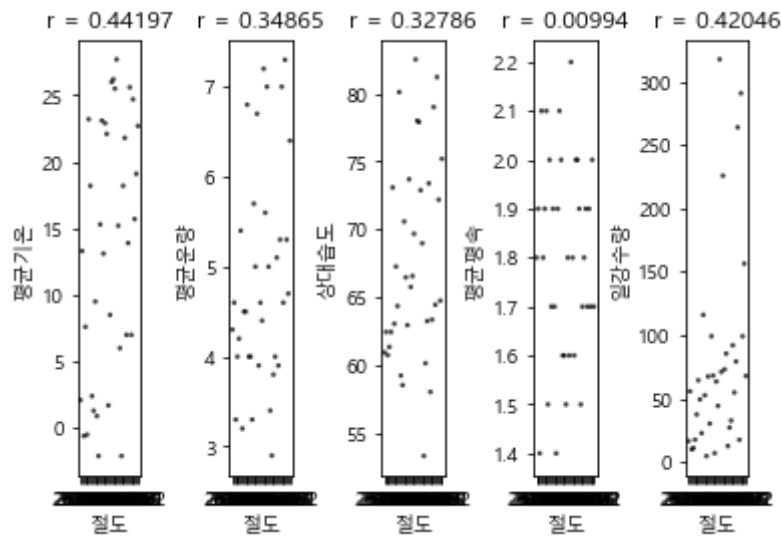
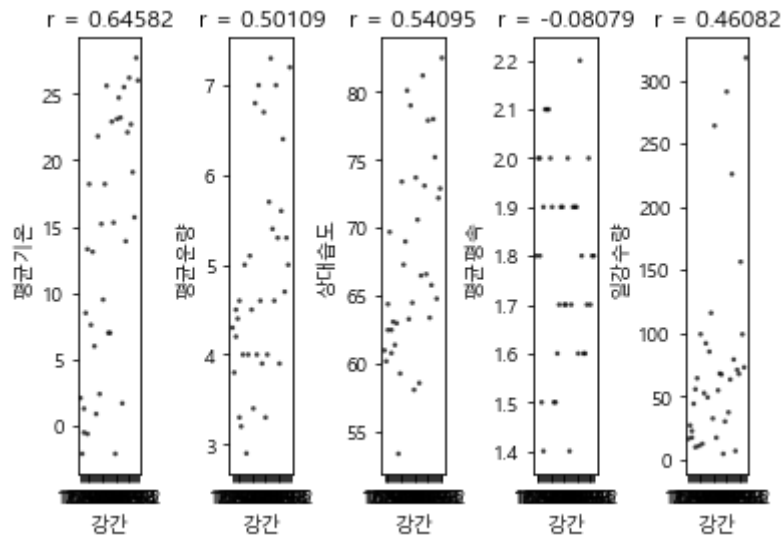






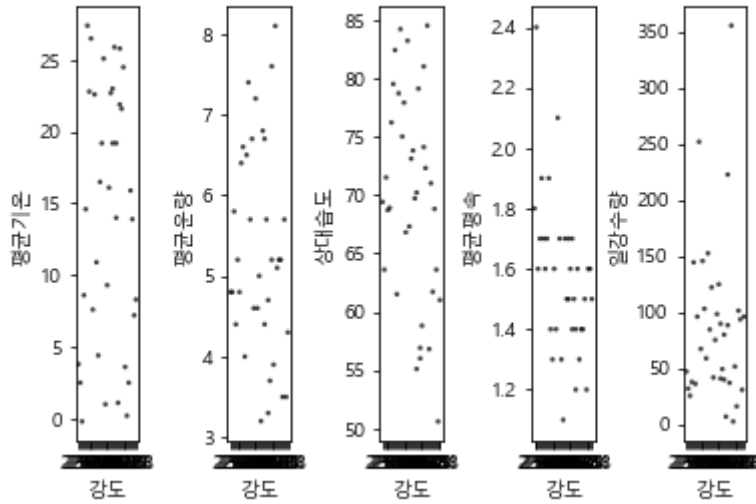
8. 평균_1990_2017_경기(유).csv 상관관계 분석



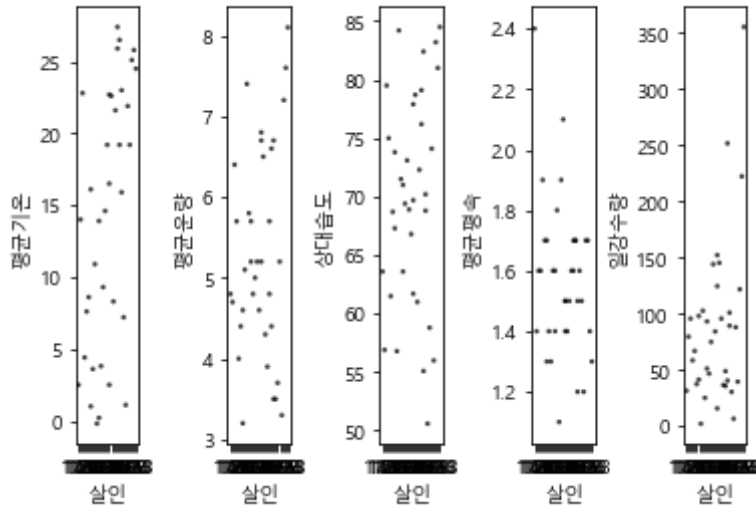


9. 평균_1990_2017_전북(유).csv 상관관계 분석

$r = -0.00920$ $r = -0.15795$ $r = -0.32575$ $r = -0.49910$ $r = 0.00875$



$r = 0.54644$ $r = 0.24875$ $r = 0.21840$ $r = -0.30965$ $r = 0.46328$



$r = 0.76428$ $r = 0.40704$ $r = 0.64904$ $r = 0.15614$ $r = 0.55659$

