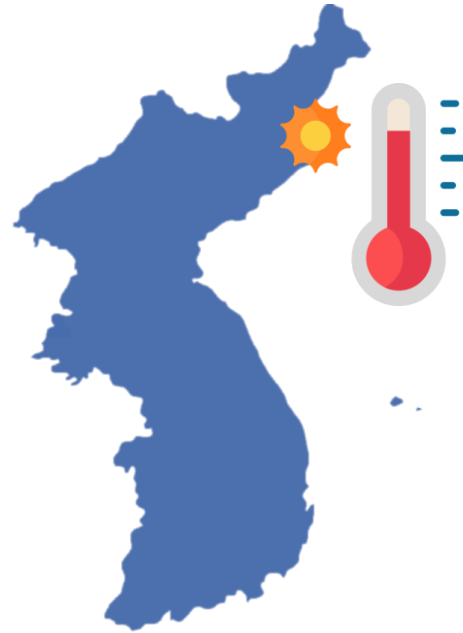


한국 기온 데이터 Functional Principal Component Analysis



1조

정영욱 최예림 오희준 배현주



01

데이터 소개 및 시각화



02

Multivariate PCA



03

Functional PCA





1

데이터 소개
및
시각화

01 데이터 소개 및 시각화

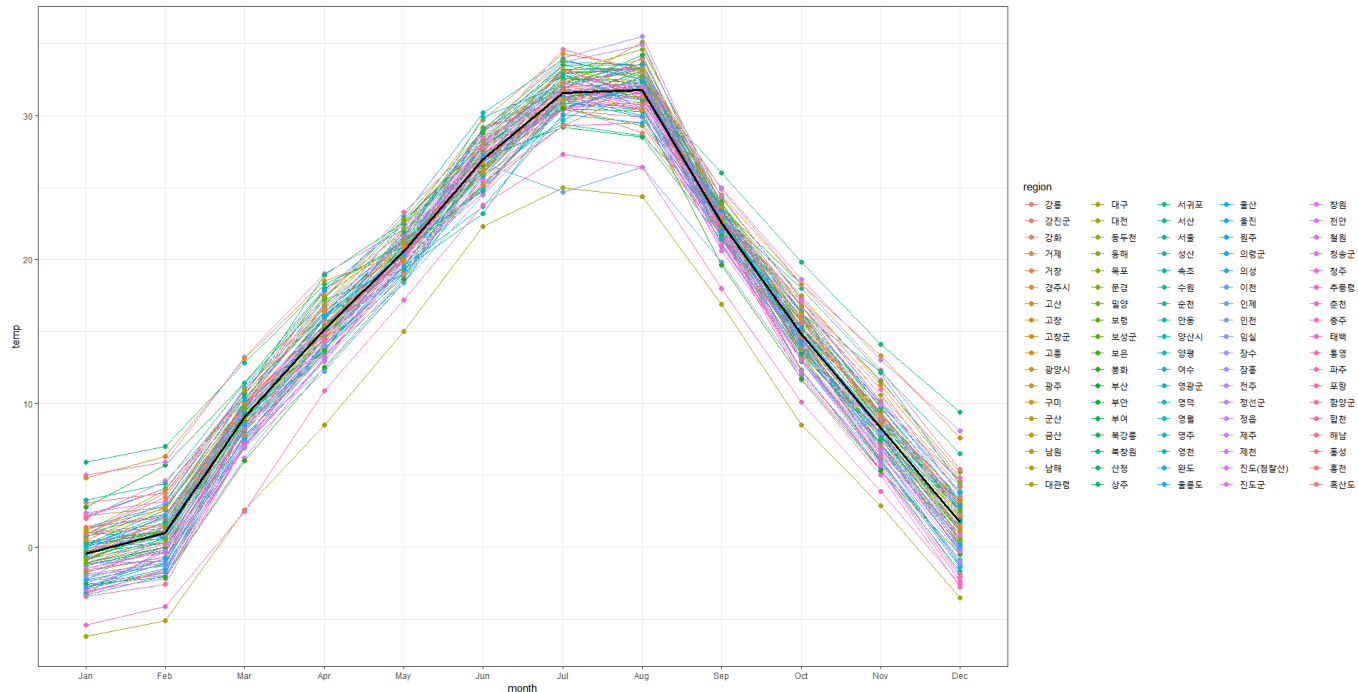
데이터 수집

1

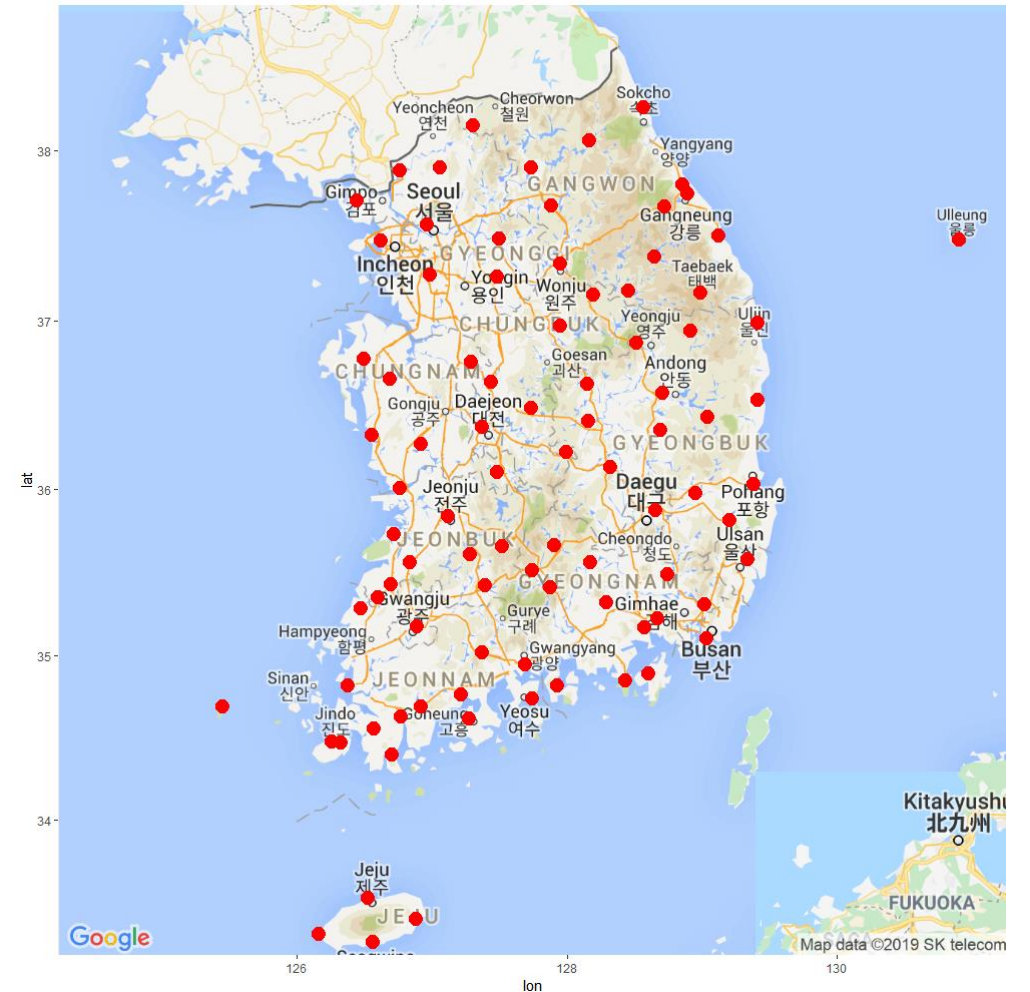
◆ 2018년 90개의 관측소에서 관측한 월별 기온 데이터 수집

데이터 출처: 기상자료개방포털
(<https://data.kma.go.kr/cmmn/main.do>)

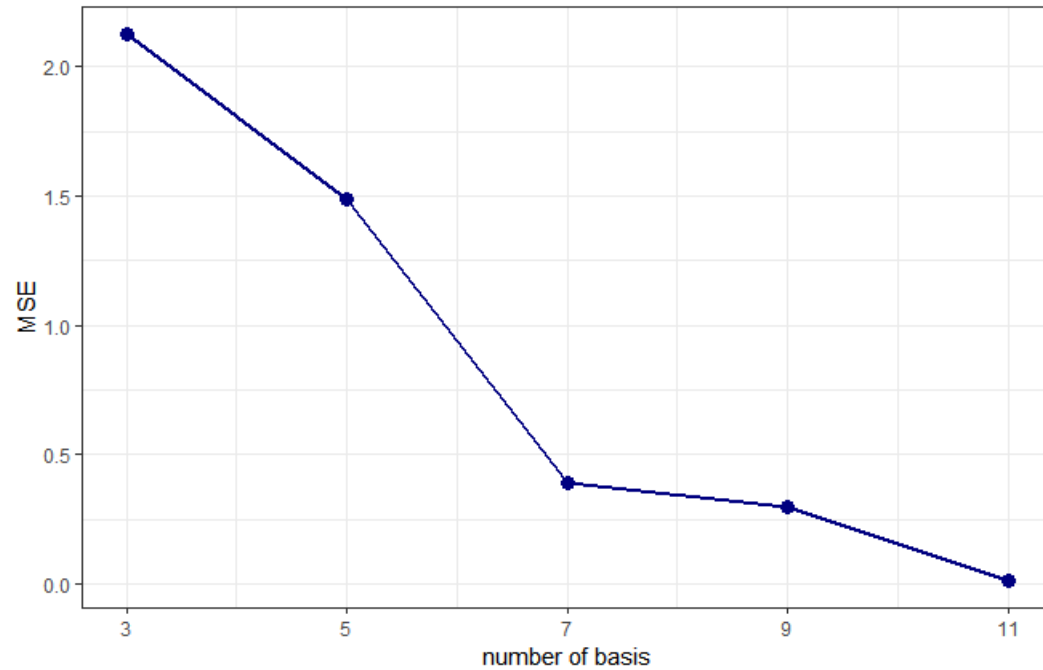
Plot of Raw Data (검은색 선은 평균값)



90개 관측소 위치



- ◆ Smooth by **Fourier basis expansion**
: MSE를 통한 basis 개수 선정 (후보: 3,5,7,9,11)



# basis	3	5	7	9	11
MSE	2.218	1.488	0.389	0.300	0.012

→ MSE의 값이 가장 크게 감소하는 지점인 7개의 basis 결정

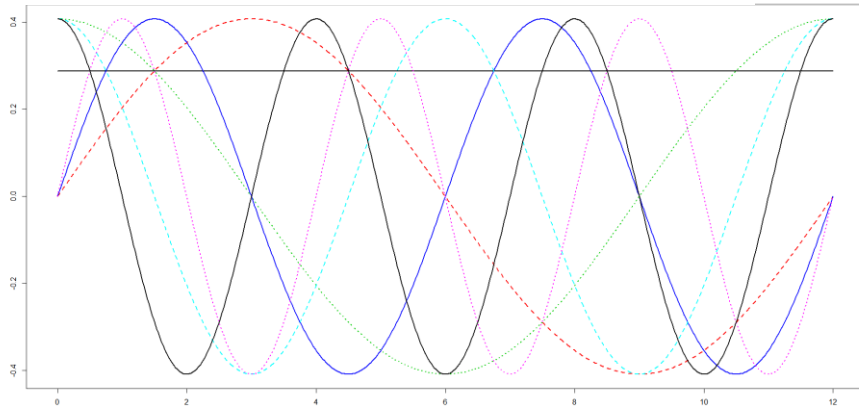
01 데이터 소개 및 시각화

Smoothing

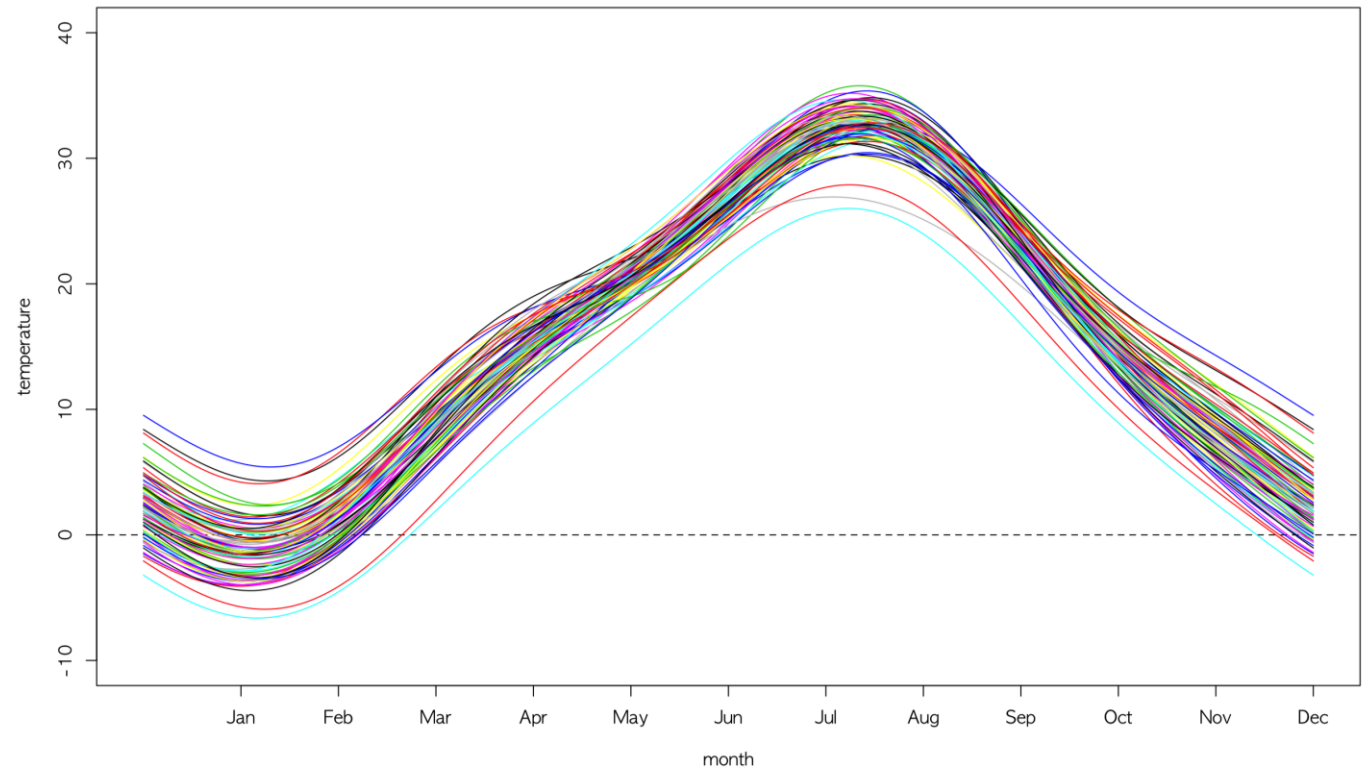
3

- ◆ Smooth by **Fourier basis expansion** (# of basis : 7)

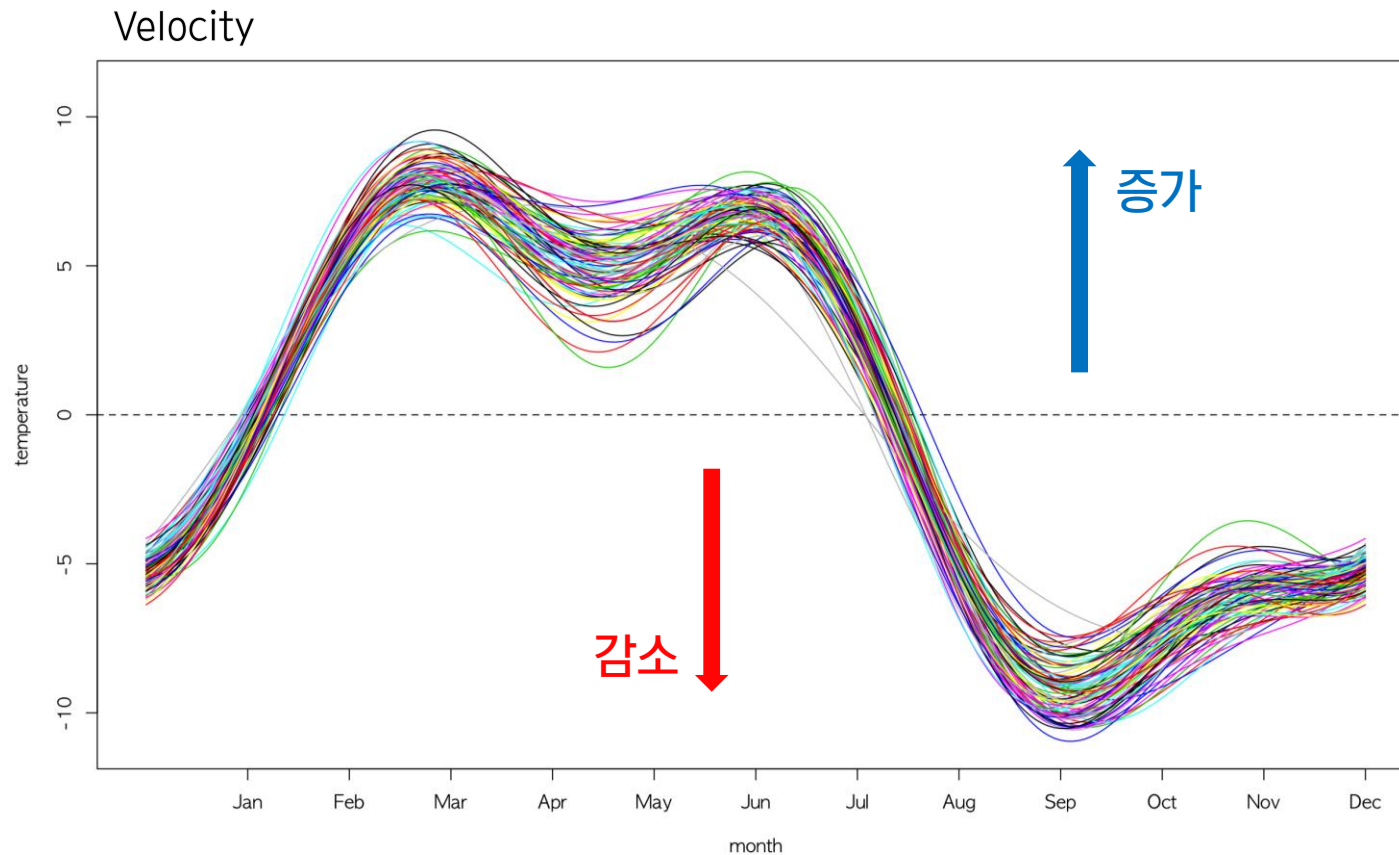
7 Fourier Basis



Smoothed Functions



◆ 1차 미분(First Derivative) Plot



2월 - 7월: 기온 **증가**

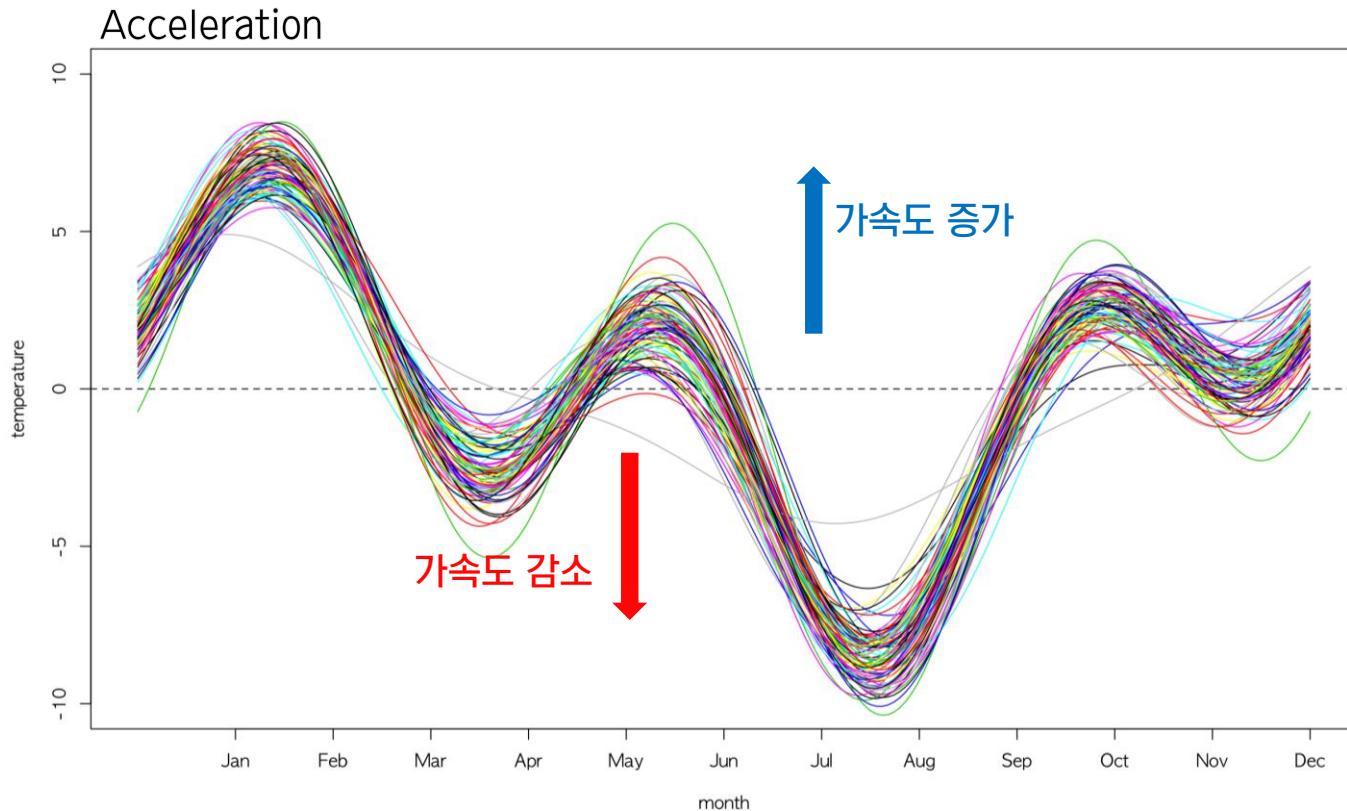
8월 - 1월: 기온 **감소**

01 데이터 소개 및 시각화

Smoothing

5

◆ 2차 미분(Second Derivative) Plot



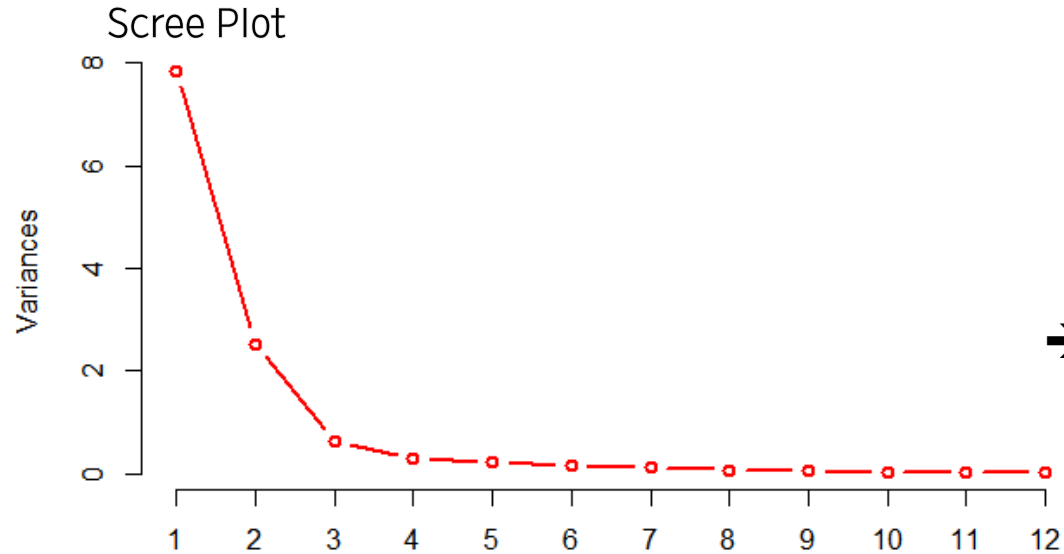
→ 1월 - 2월 / 6월 / 10월: 양의 가속도
3월 - 4월 / 7월 - 8월 / 12월: 음의 가속도
5월 / 11월: 0에 가까운 가속도

→ 대체적으로 1월에 기온이 빠르게 감소하고
5월에 온도가 가장 빠르게 상승한다.



2

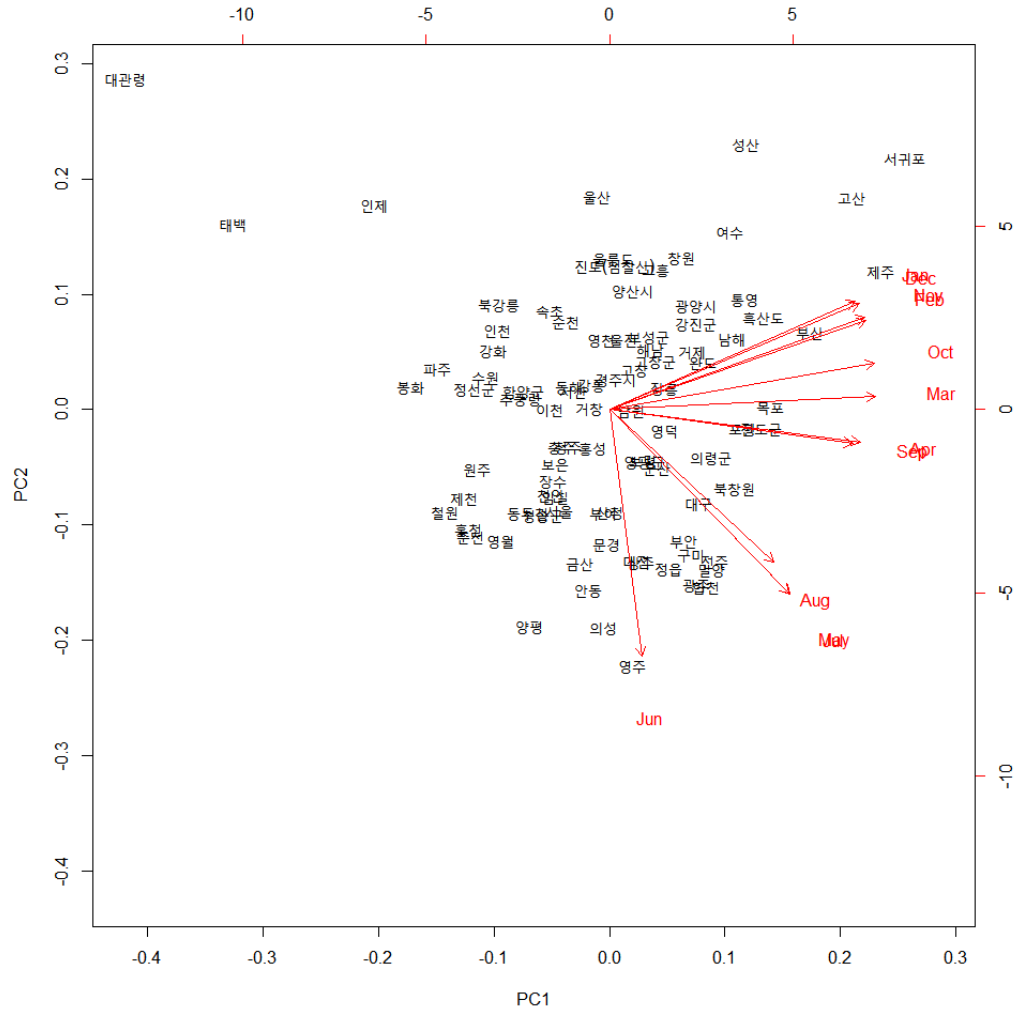
Multivariate
PCA



→ 주성분이 3개인 지점부터 분산의 변동성이 완만해짐

	PC1	PC2	PC3	...	PC11	PC12
Standard deviation	2.804071	1.592239	0.793123		0.154103	0.118509
Proportion of Variance	0.65523	0.21127	0.05242		0.00198	0.00117
Cumulative Proportion	0.65523	0.8665	0.91892		0.99883	1

각 PC의 분산 설명력



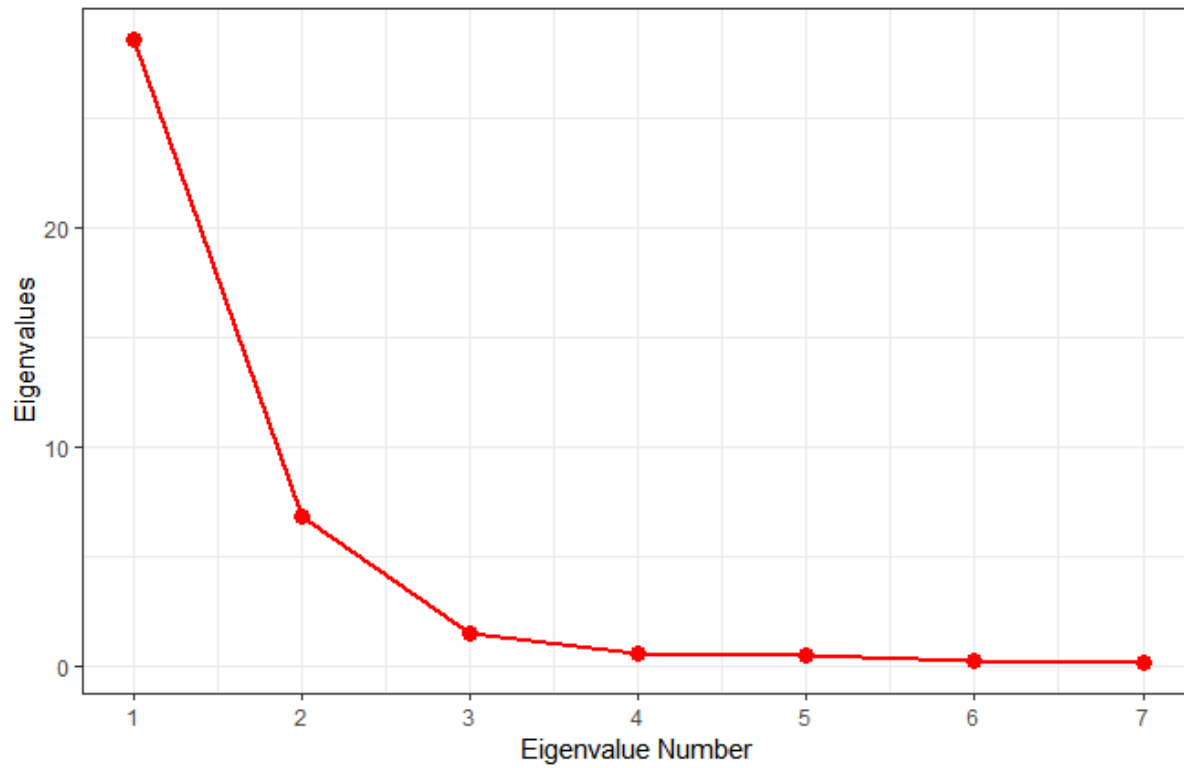
◆ 제 1 주성분에 대하여 4 계절이 모두 강한 양의 상관

◆ 제 2 주성분에 대하여
‘여름’의 계절이 강한 음의 상관관계를,
‘겨울’의 계절이 강한 양의 상관관계를 보임

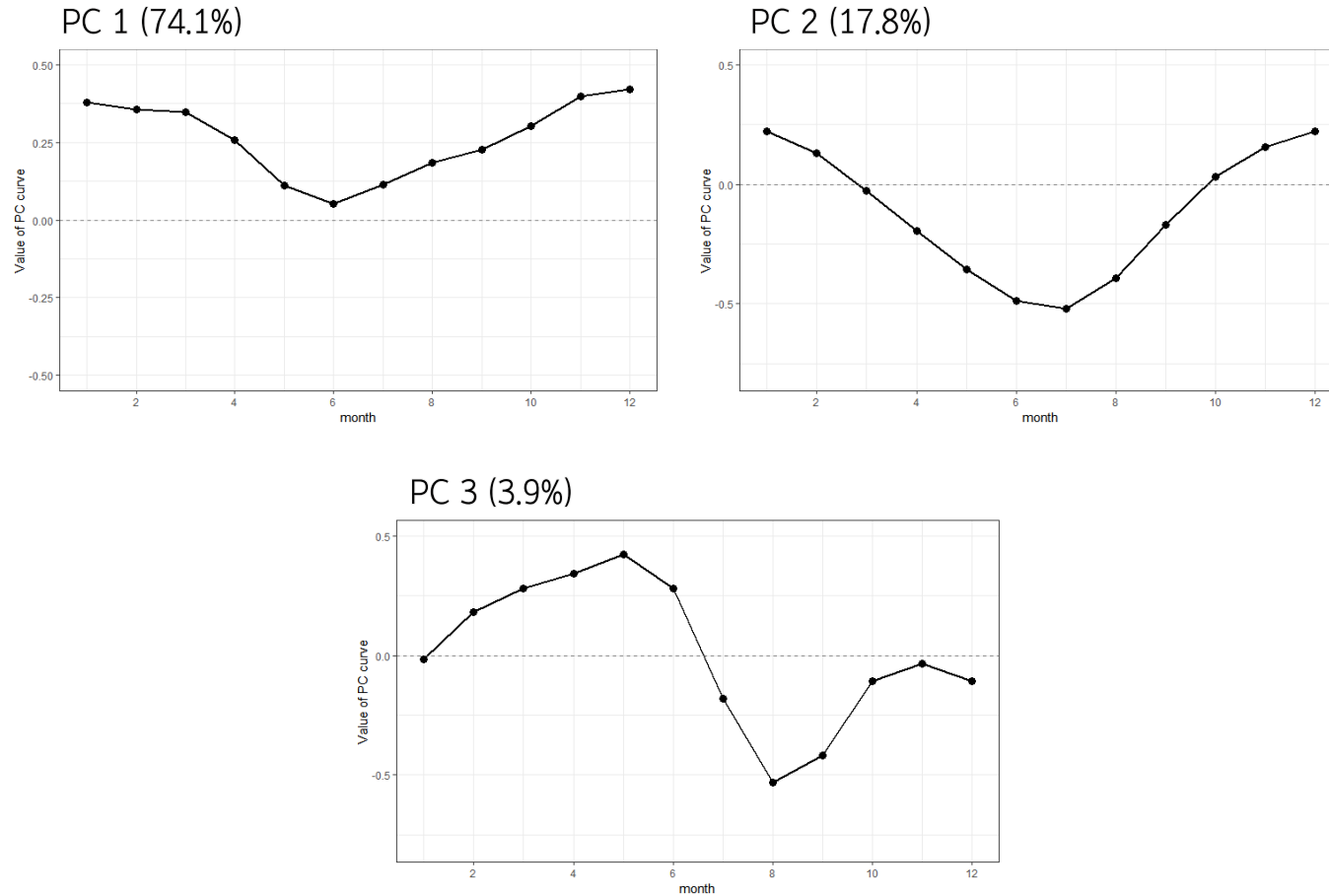


3

Functional
PCA

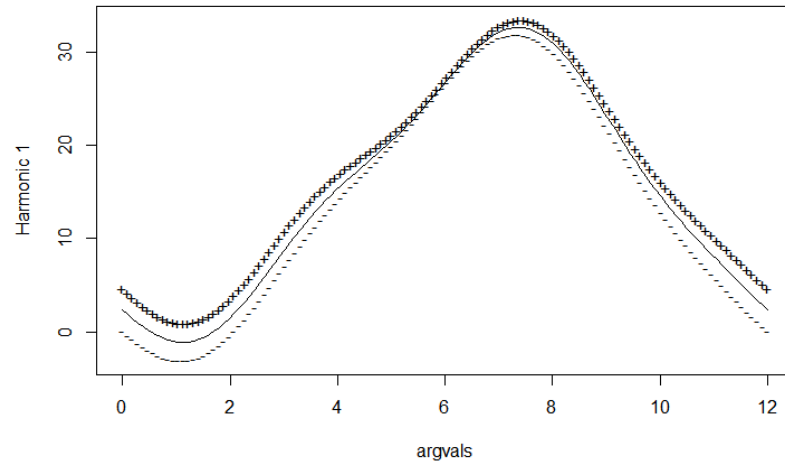


- ◆ basis의 개수에 따라 PC의 개수도 7개로 설정
→ 적절한 PC의 개수는 3개로 판단

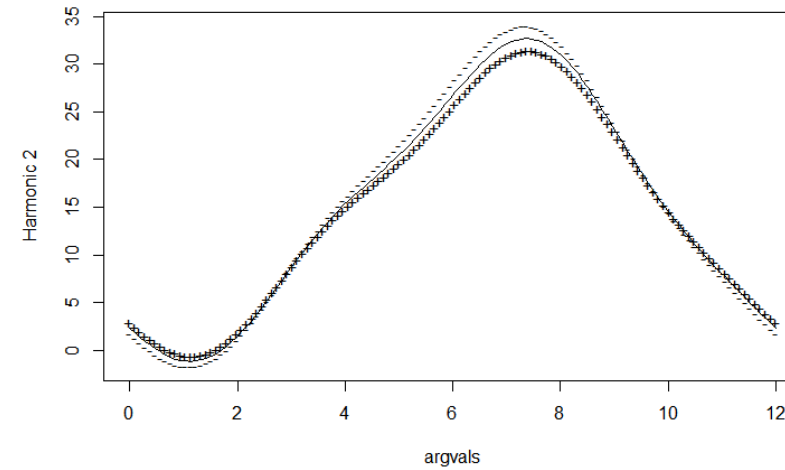


- ◆ centered된 데이터에 PCA 적용
- ◆ 3개의 PC로 전체 변동의 96% 설명
- ◆ 첫 세 개의 PC의 eigenfunction
 - 관측소 간 가장 큰 변동은 겨울에 발생
 - 첫 번째 eigenfunction이 모두 양수이기 때문에 두 번째 eigenfunction에 음수 부분 존재

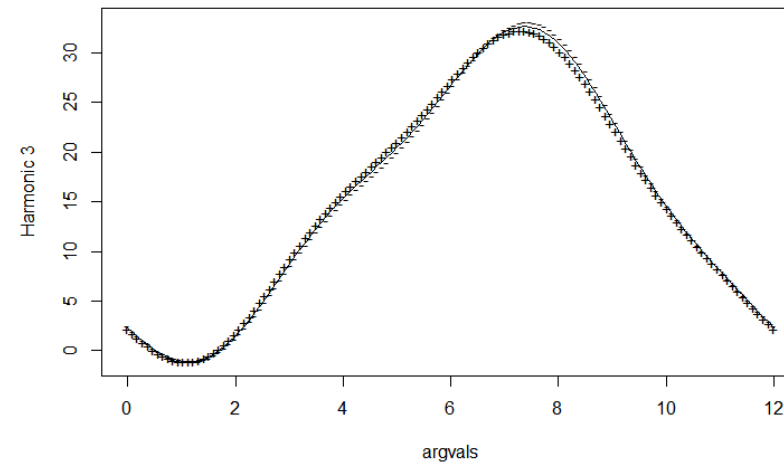
PCA function 1 (Percentage of variability 74.1)

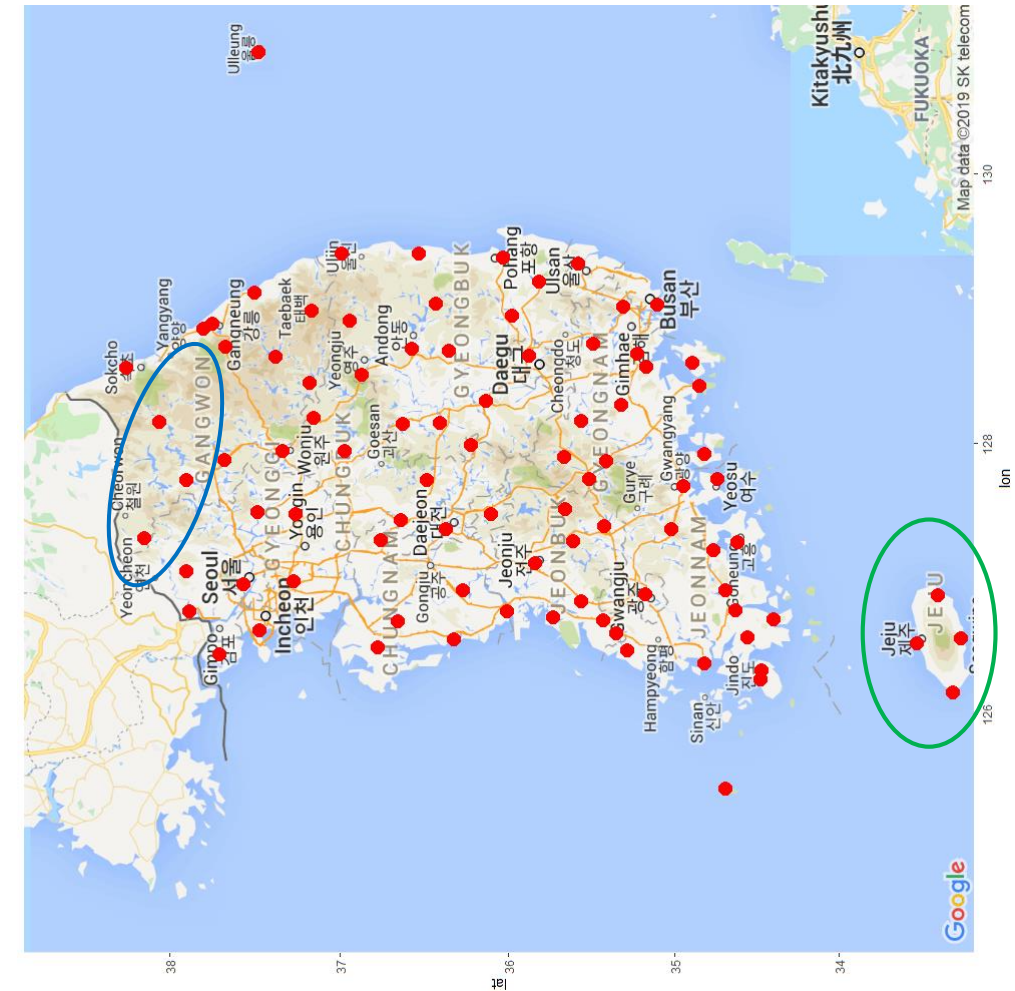
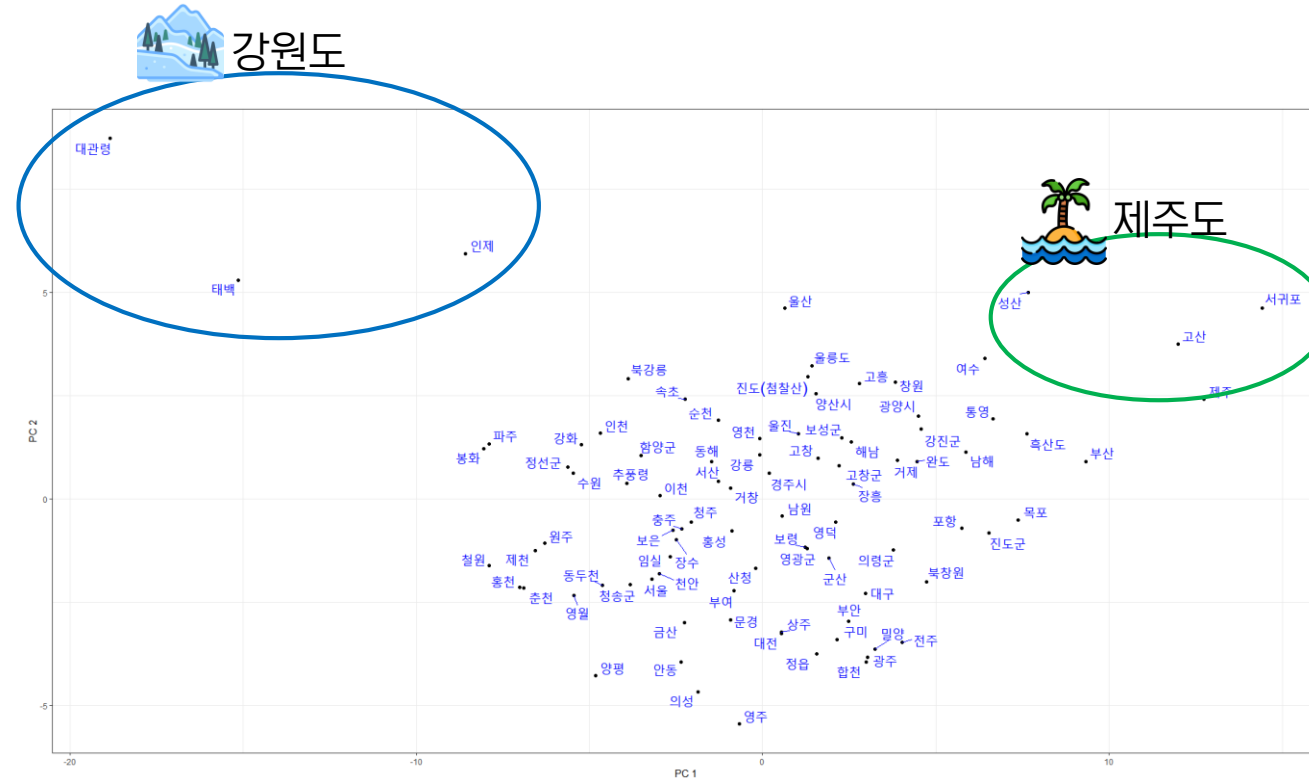


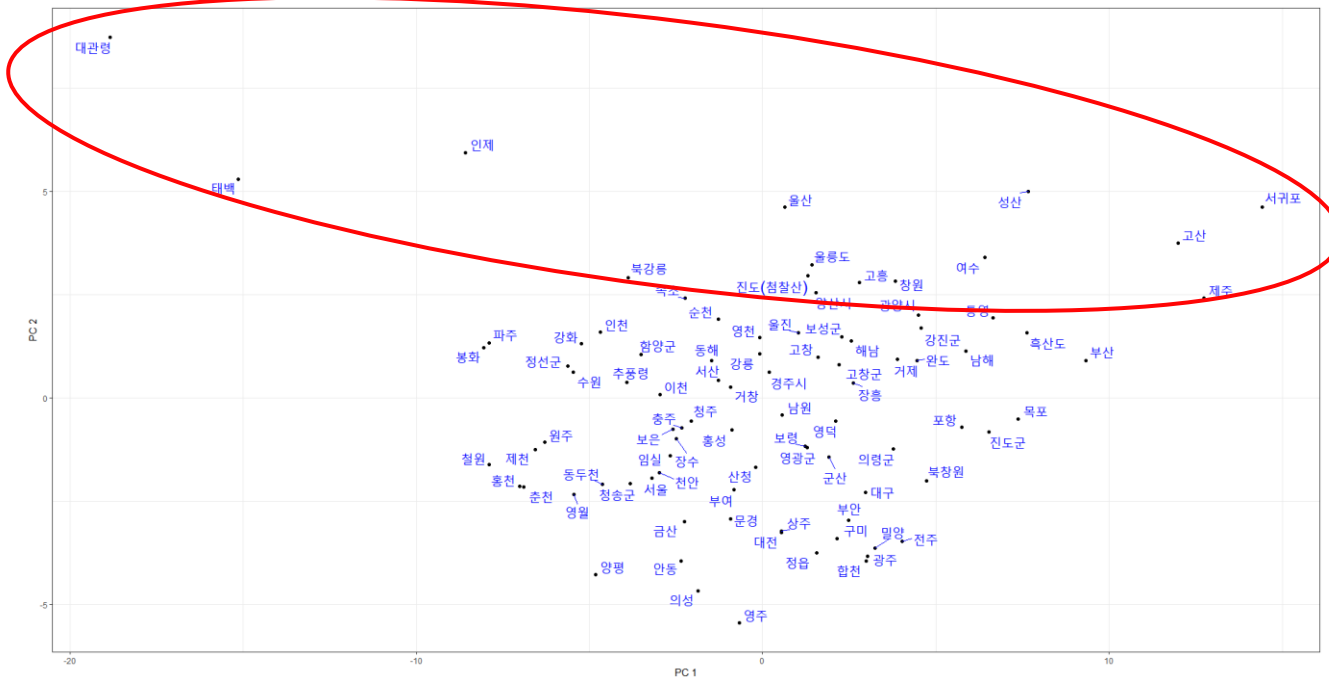
PCA function 2 (Percentage of variability 17.8)



PCA function 3 (Percentage of variability 3.9)





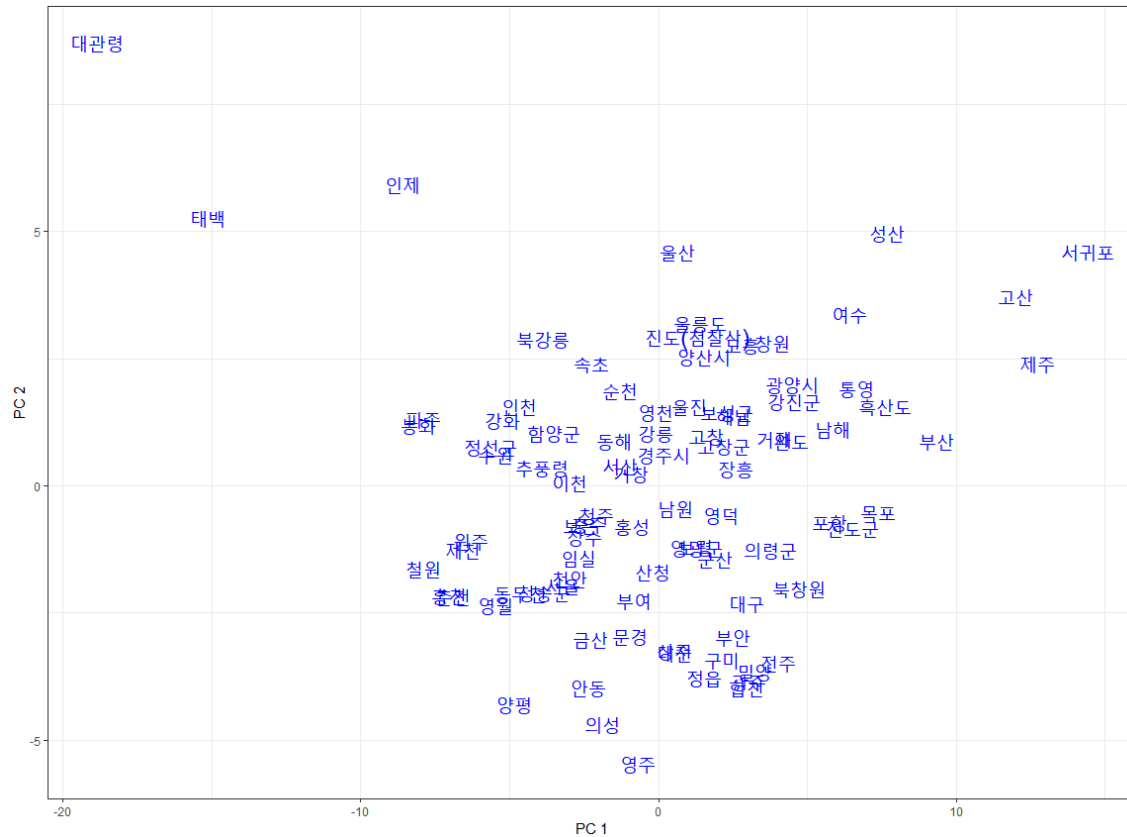


→ 여름과 겨울의 기온차가 작은 지역이 PC2의 값이 큼

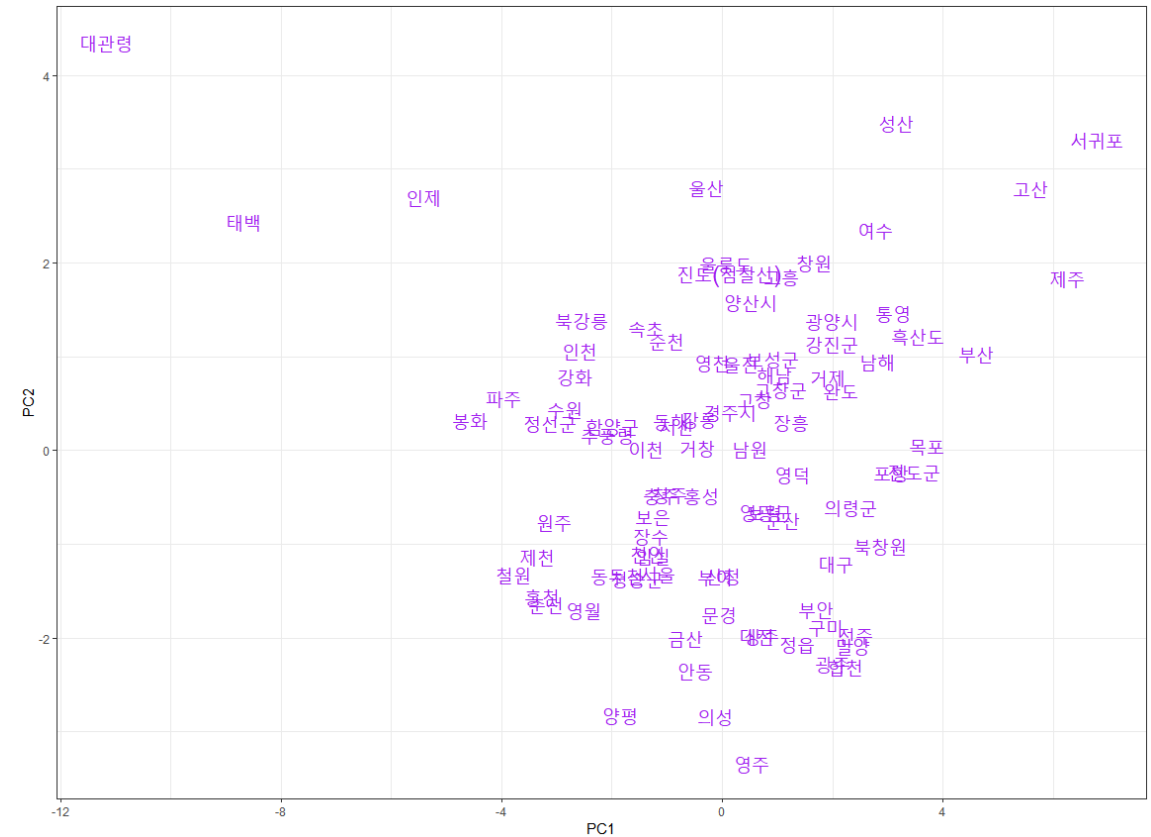
Region	diff
서귀포	26.1
고산	26.7
제주	27.3
성산	28.3
인제	28.4
강릉	28.6
창원	28.9
속초	29.5
울산	29.5
고흥	29.6
강진군	29.7
북강릉	29.8
울진	29.8
여수	29.8
진도(침찰산)	29.8
동해	30
흑산도	30.1
양산시	30.3
부산	30.4
해남	30.4
거제	30.4
대관령	30.6

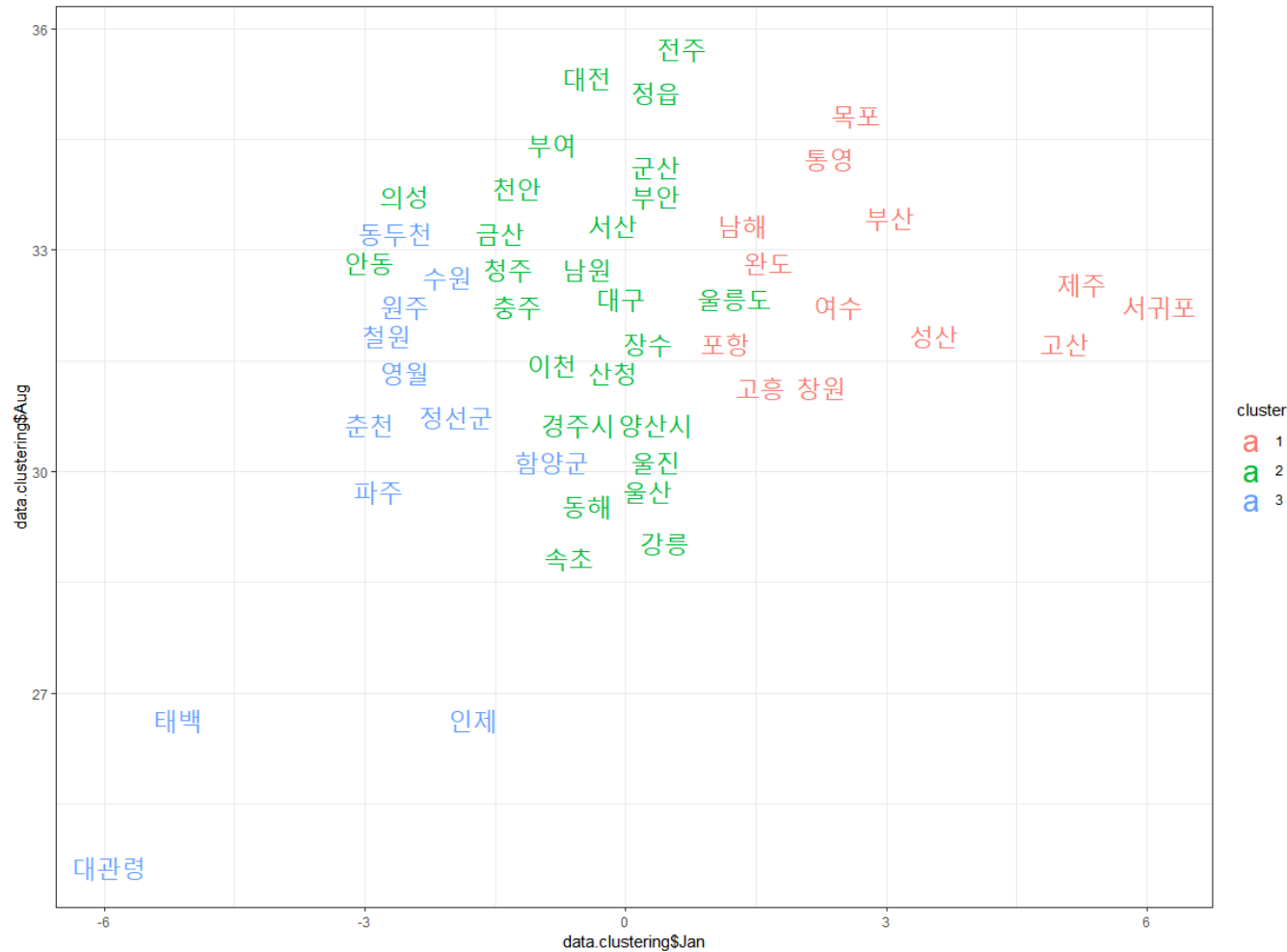
8월 기온 - 1월 기온

Functional PCA



Multivariate PCA





◆ 3개의 군집으로 k-means 군집분석