

탐색적 자료분석 및 실습

# ‘팀 과제 2’ 1조 발표

김우진, 김주형, 남영훈, 선우지훈, 임태건

# CONTENTS

- 01 주제선정 배경
- 02 데이터 설명
- 03 시계열 분석
- 04 회귀 분석
- 05 주성분 분석
- 06 정리 및 요약

01

## 주제선정 배경

# 01 주제선정 배경



## 〈국내 여행지 소셜/커머셜 지수 순위〉

	지역명	소셜 지수
1	제주	1,219,410
2	부산	766,989
3	강화	598,383
4	여수	593,487
5	전주	416,262
	⋮	
12	김포	273,309
13	춘천	267,393
14	고성	240,771
	⋮	

	지역명	커머셜 지수
1	제주	58,336,605
2	광주(경기)	12,304,017
3	부산	8,412,801
4	구미	5,266,062
5	여수	4,443,663
	⋮	
18	이천	1,506,714
19	춘천	1,478,520
20	김포	1,455,321
	⋮	

자료 : 티티엘뉴스, 2020.08 ~ 2020.09

춘천은 국내 여행지 중 소셜 지수와 커머셜 지수 모두 높은 순위로 나타남

이를 바탕으로 춘천 관광에 대한 관심이 높다는 것을 알 수 있음

# 01 주제선정 배경

## 〈 춘천시 관광에 대한 자료조사 〉

### 세계일보

#### 교통망 좋아진 춘천...관광객 1000만 돌파

2012년, 남이섬 가장 많이 찾아

입력 2013-01-20 17:15:39, 수정 2013-01-20 19:43:15

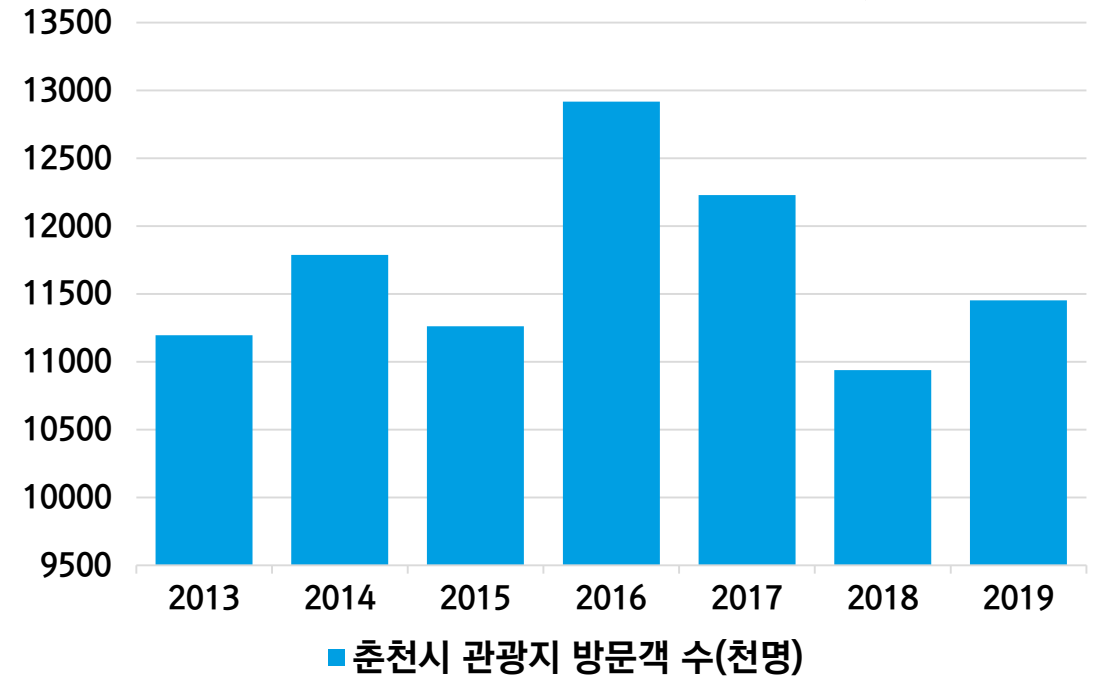
고속도로와 전철이 개통되면서 수도권과의 접근성이 획기적으로 개선된 강원도 춘천시를 찾은 관광객이 지난해 처음으로 1000만명을 넘어섰다. 춘천시는 지난해 내·외국인 관광객이 모두 1016만명으로 집계됐다고 20일 밝혔다. 2011년(865만 명)과 비교하면 17.5% 증가했다. 춘천시는 관광객이 지난해 3분기까지 745만명이었지만 4분기 들어 271만명을 기록해 1000만 명을 넘어섰다고 설명했다.

춘천시를 찾은 관광객은 2008년까지 500만명대에 정체됐지만 서울춘천고속도로(2009년 7월)와 경춘선복선전철(2010년 12월) 개통을 계기로 2009년 680만명, 2010년 737만명, 2011년 865만명으로 급증 추세를 보였다. 관광지별로는 남이섬이 259만명으로 가장 많았고 강촌 240만명, 소양강댐 140만명으로 뒤를 이었다. 외국인 관광객 증가세가 두드러져 전년(46만명)보다 50% 증가한 69만명을 기록했다.

춘천=박연직 기자



자료 : 강원통계정보, 2013 ~ 2019



2012년 처음으로 관광객 1000만명 돌파 이후 2019년까지 8년 연속 관광객 천만시대

# 01 주제선정 배경

## 〈 춘천시 관광에 대한 자료조사 〉

### 세계일보

#### 교통망 좋아진 춘천...관광객 1000만 돌파

2012년, 남이섬 가장 많이 찾아

입력 2013-01-20 17:15:39, 수정 2013-01-20 19:43:15

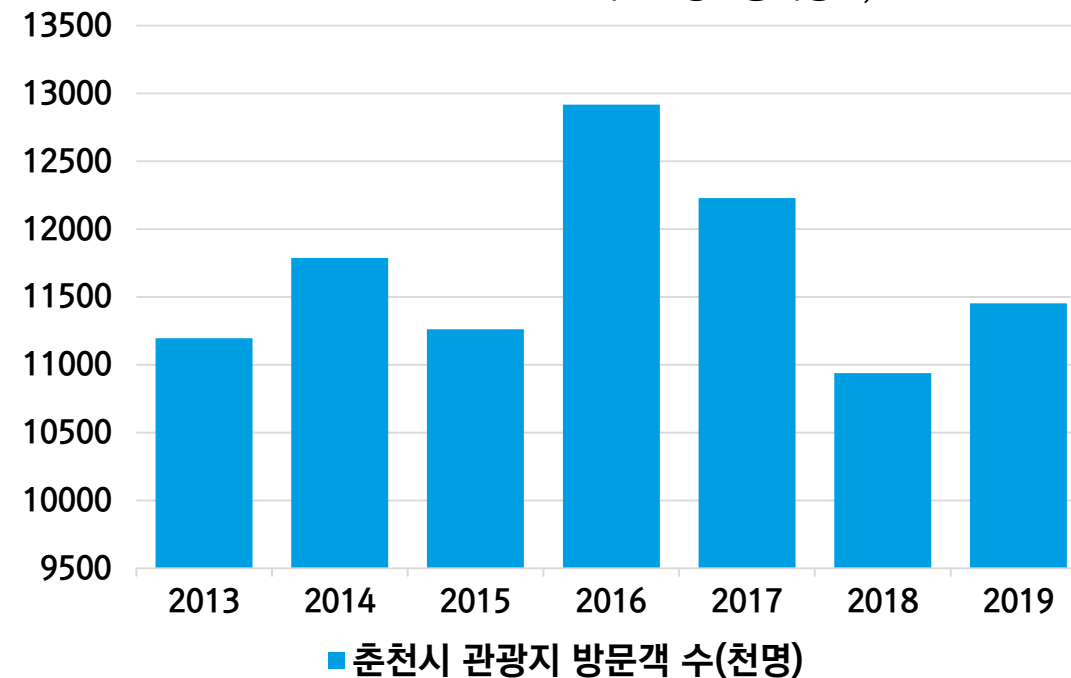
고속도로와 전철이 개통되면서 수도권과의 접근성이 획기적으로 개선된 강원도 춘천시를 찾은 관광객이 지난해 처음으로 1000만명을 넘어섰다. 춘천시는 지난해 내·외국인 관광객이 모두 1016만명으로 집계됐다고 20일 밝혔다. 2011년(865만 명)과 비교하면 17.5% 증가했다. 춘천시는 관광객이 지난해 3분기까지 745만명이었지만 4분기 들어 271만명을 기록해 1000만 명을 넘어섰다고 설명했다.

춘천시를 찾은 관광객은 2008년까지 500만명대에 정체됐지만 서울춘천고속도로(2009년 7월)와 경춘선복선전철(2010년 12월) 개통을 계기로 2009년 680만명, 2010년 737만명, 2011년 865만명으로 급증 추세를 보였다. 관광지별로는 남이섬이 259만명으로 가장 많았고 강촌 240만명, 소양강댐 140만명으로 뒤를 이었다. 외국인 관광객 증가세가 두드러져 전년(46만명)보다 50% 증가한 69만명을 기록했다.

춘천=박연직 기자



자료 : 강원통계정보, 2013 ~ 2019



2012년 처음으로 관광객 1000만명 돌파 이후 2019년까지 8년 연속 관광객 천만시대

2012년부터 관광객 증가에 큰 영향을 준 사건이 존재할 것으로 예상

# 01 주제선정 배경

〈 2012-02-28 ITX 청춘 개통 〉



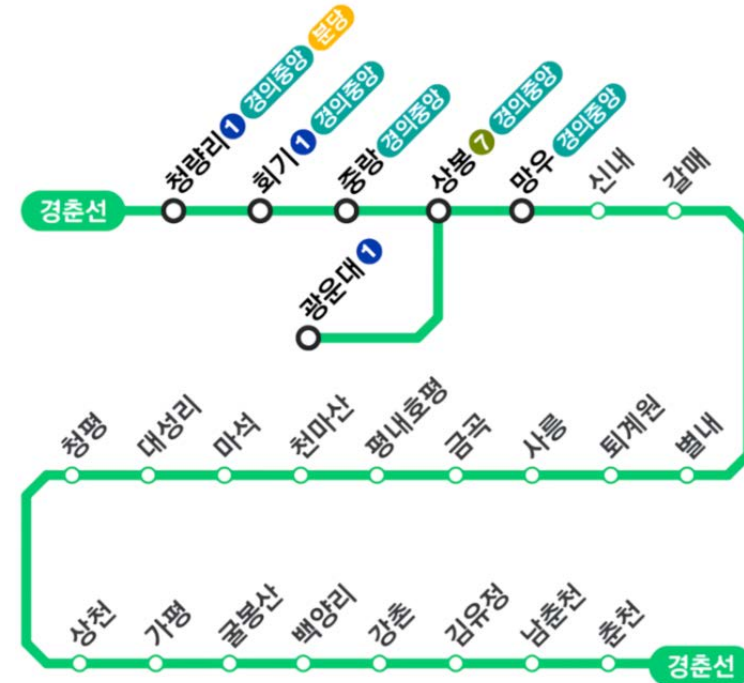
강원도민일보 2012.02.29.

'ITX-청춘' 첫 시승 축하 열기

“춘천행 편하네” ITX청춘열차가 28일 개통식과 함께 본격적인 운행에 들어간 가운데 서울 청량리역을 출발해 춘천역에 도착한 관광객들이 ITX청춘열차에서 내리...



〈 경춘선 복선전철 노선도 〉



2012년 02월 28일에 개통한 ITX 청춘의 영향으로 추정



# 01 주제선정 배경

〈 2012-02-28 ITX 청춘 개통 〉



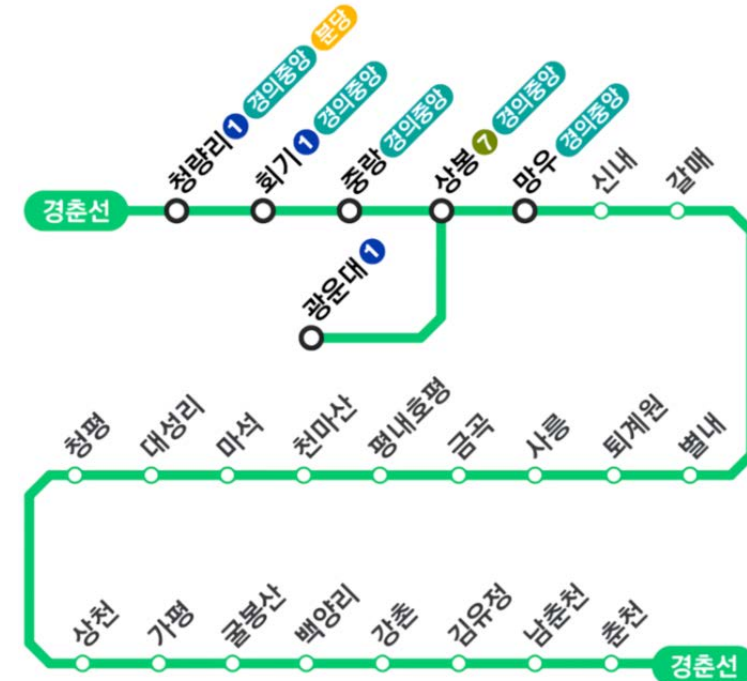
강원도민일보 2012.02.29.

'ITX-청춘' 첫 시승 축하 열기

"춘천행 편하네" ITX청춘열차가 28일 개통식과 함께 본격적인 운행에 들어간 가운데 서울 청량리역을 출발해 춘천역에 도착한 관광객들이 ITX청춘열차에서 내리...



〈 경춘선 복선전철 노선도 〉



2012년 02월 28일에 개통한 ITX 청춘의 영향으로 추정

철도 이용객 수(경춘선 전철, ITX 청춘)와 춘천 관광객 사이에 어떤 관계가 있는지 분석



02

## 데이터 설명

## 02 데이터 설명

### 1. 역별 여객수송실적 (월)

변수명	변수 설명
Year_month	날짜 변수 (“YYYYMM”)
ITX_count	해당 연월의 ‘ITX’ 하차 승객 수
GC_count	해당 연월의 ‘경춘선’ 하차 승객 수
Sum_train	해당 연월의 ITX, 경춘선 전철의 하차 승객 수 합

- ✓ 매 달 춘천 행정구역 내 역에서 하차한 승객 수에 대한 정보
- ✓ 출처 : 국토교통부 KRiC 철도통계
- ✓ 데이터 기간 : 2016년 1월 ~ 2020년 12월

### 2. 전국 주요관광지점 입장객 통계

변수명	변수 설명
Year_month	날짜 변수 (“YYYYMM”)
관광지 1, 2, ...	해당 연월의 ‘춘천시 관광지 입장객’ 수
⋮	⋮
Sum_travel	해당 연월의 관광지 입장객 수 합

- ✓ 매 달 춘천 행정구역 내 관광지 입장객 수에 대한 정보
- ✓ 출처 : 문화체육관광부 관광지식정보시스템
- ✓ 데이터 기간 : 2016년 1월 ~ 2020년 12월

## 02 데이터 설명

### 1. 역별 여객수송실적 (월)

변수명	변수 설명
Year_month	날짜 변수 ("YYYYMM")
ITX_count	해당 연월의 'ITX' 하차 승객 수
GC_count	해당 연월의 '경춘선' 하차 승객 수
Sum_train	해당 연월의 ITX, 경춘선 전철의 하차 승객 수 합

- ✓ 매 달 춘천 행정구역 내 역에서 하차한 승객 수에 대한 정보
- ✓ 출처 : 국토교통부 KRiC 철도통계
- ✓ 데이터 기간 : 2016년 1월 ~ 2020년 12월

### 2. 전국 주요관광지점 입장객 통계

변수명	변수 설명
Year_month	날짜 변수 ("YYYYMM")
관광지 1, 2, ...	해당 연월의 '춘천시 관광지 입장객' 수
:	:
Sum_travel	해당 연월의 관광지 입장객 수 합

- ✓ 매 달 춘천 행정구역 내 관광지 입장객 수에 대한 정보
- ✓ 출처 : 문화체육관광부 관광지식정보시스템
- ✓ 데이터 기간 : 2016년 1월 ~ 2020년 12월

## 02 데이터 설명

### 1. 역별 여객수송실적 (월)

변수명	변수 설명
Year_month	날짜 변수 ("YYYYMM")
ITX_count	해당 연월의 'ITX' 하차 승객 수
GC_count	해당 연월의 '경춘선' 하차 승객 수
Sum_train	해당 연월의 ITX, 경춘선 전철의 하차 승객 수 합

- ✓ 매 달 춘천 행정구역 내 역에서 하차한 승객 수에 대한 정보
- ✓ 출처 : 국토교통부 KRIc 철도통계
- ✓ 데이터 기간 : 2016년 1월 ~ 2020년 12월

### 2. 전국 주요관광지점 입장객 통계

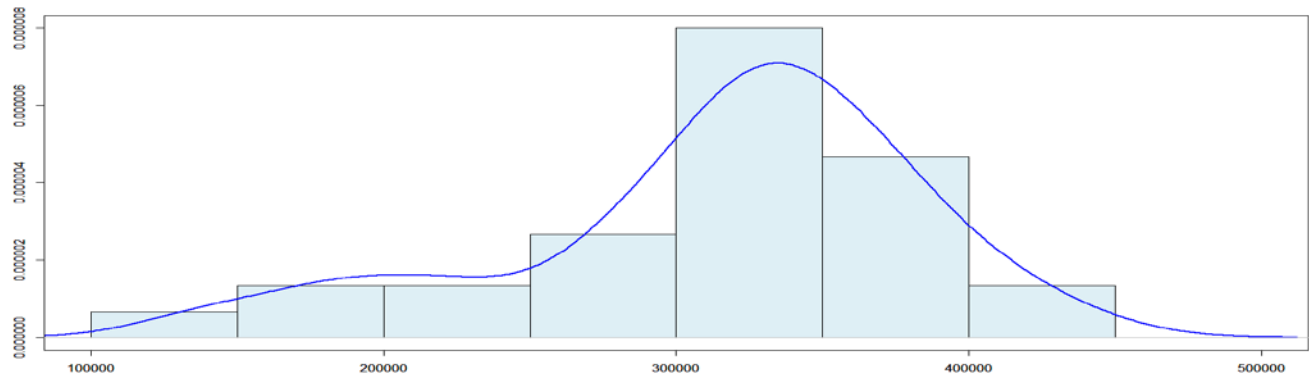
변수명	변수 설명
Year_month	날짜 변수 ("YYYYMM")
관광지 1, 2, ...	해당 연월의 '춘천시 관광지 입장객' 수
:	:
Sum_travel	해당 연월의 관광지 입장객 수 합

- ✓ 매 달 춘천 행정구역 내 관광지 입장객 수에 대한 정보
- ✓ 출처 : 문화체육관광부 관광지식정보시스템
- ✓ 데이터 기간 : 2016년 1월 ~ 2020년 12월

# 02 데이터 설명 - 여객수송 데이터

Sum\_train 변수

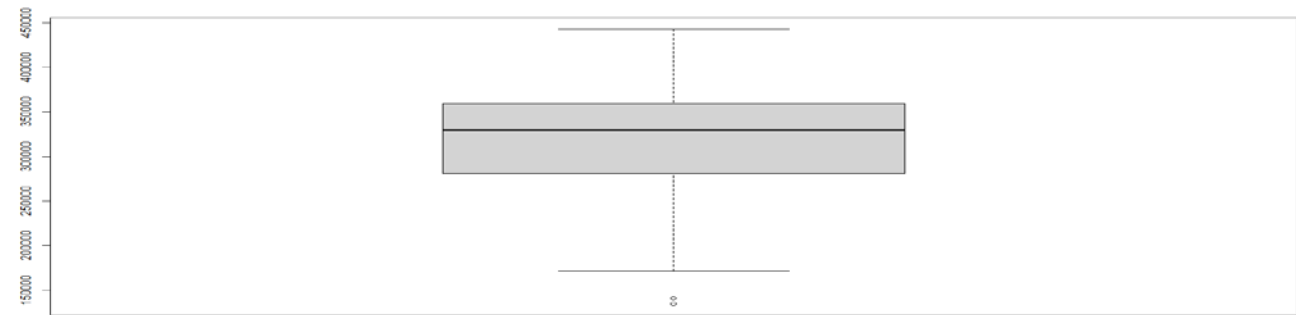
〈히스토그램과 커널곡선〉



〈줄기잎 그림〉

```
1 | 34
1 | 7889
2 | 1223
2 | 57888899
3 | 011111122233333333444
3 | 555566667778888
4 | 001224
```

〈상자그림〉

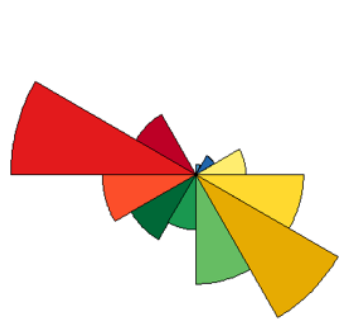


최소값	제1사분위수	제2사분위수	평균	제3사분위수	최대값
134922	281534	329734	313997	359420	442980

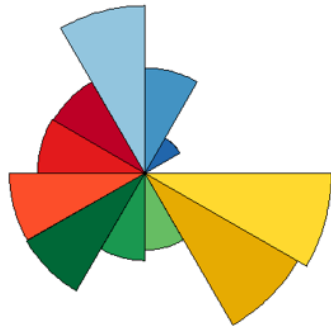
〈도수분포표〉

계급	상대도수	도수
(100000, 150000]	0.033	2
(150000, 200000]	0.067	4
(200000, 250000]	0.067	4
(250000, 300000]	0.133	8
(300000, 350000]	0.4	24
(350000, 400000]	0.233	14
(400000, 450000]	0.067	4
합계	1	60

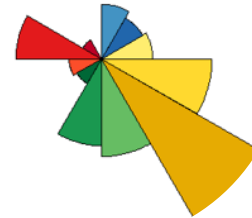
## 02 데이터 설명 - 관광지별 입장객 데이터



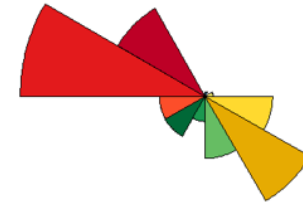
강촌레일파크



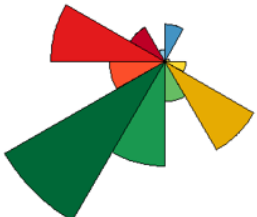
국립춘천박물관



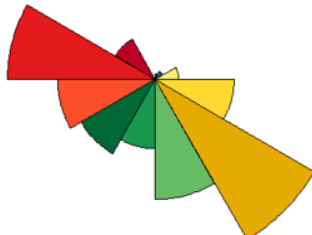
김유정문화마을



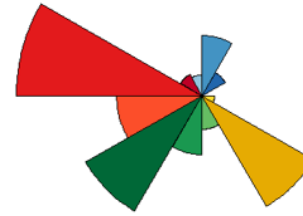
남이섬유적지



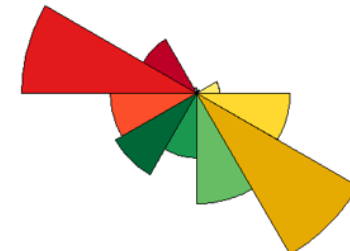
막국수체험 박물관



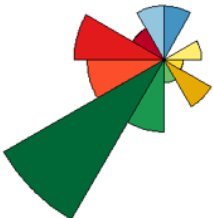
물레길



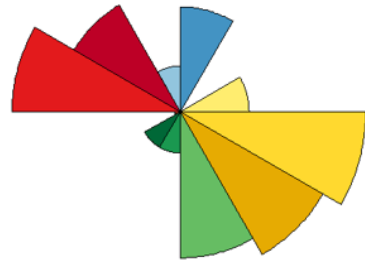
애니메이션 박물관



제이드가든



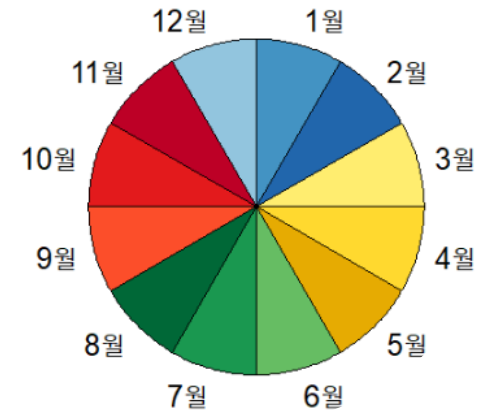
춘천에티오피아  
한국전참전기념관



춘천지구  
전적기념관

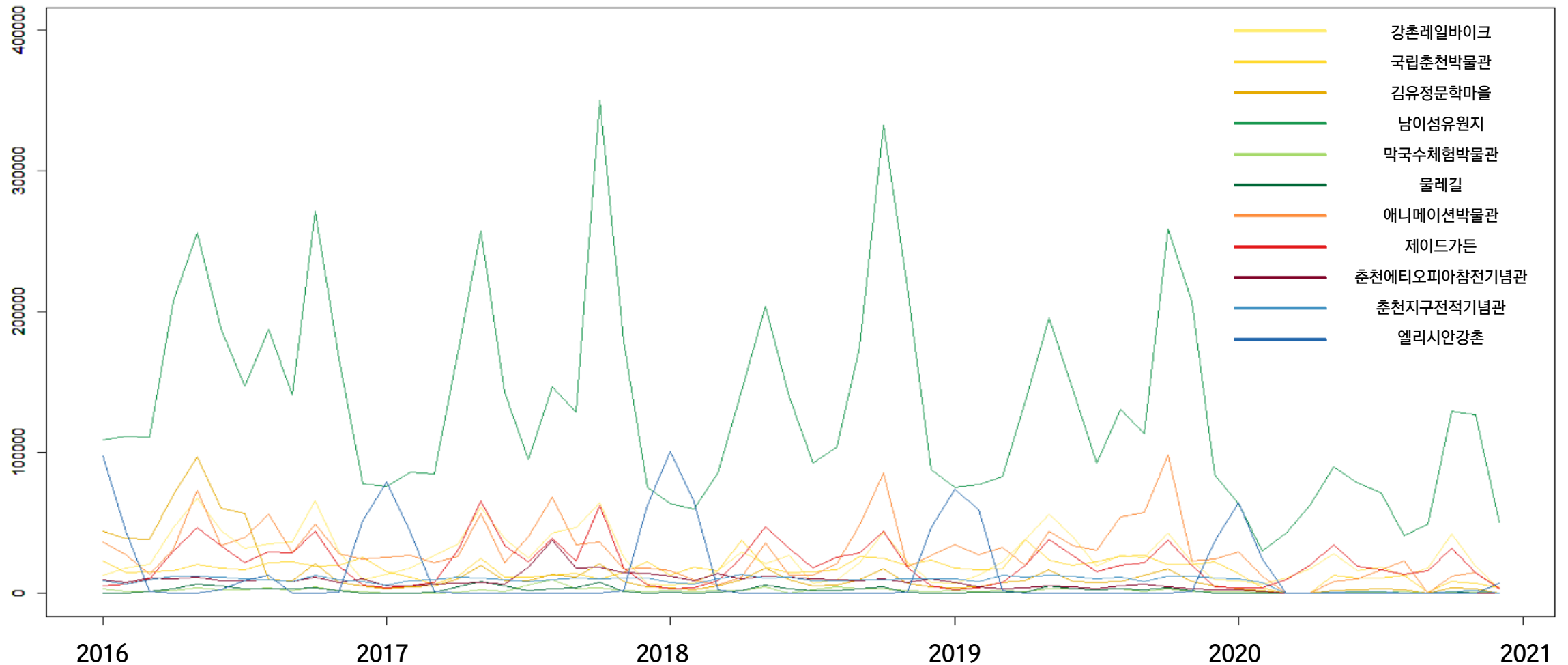


엘리시안  
강촌스키장



## 02 데이터 설명 - 관광지별 입장객 데이터

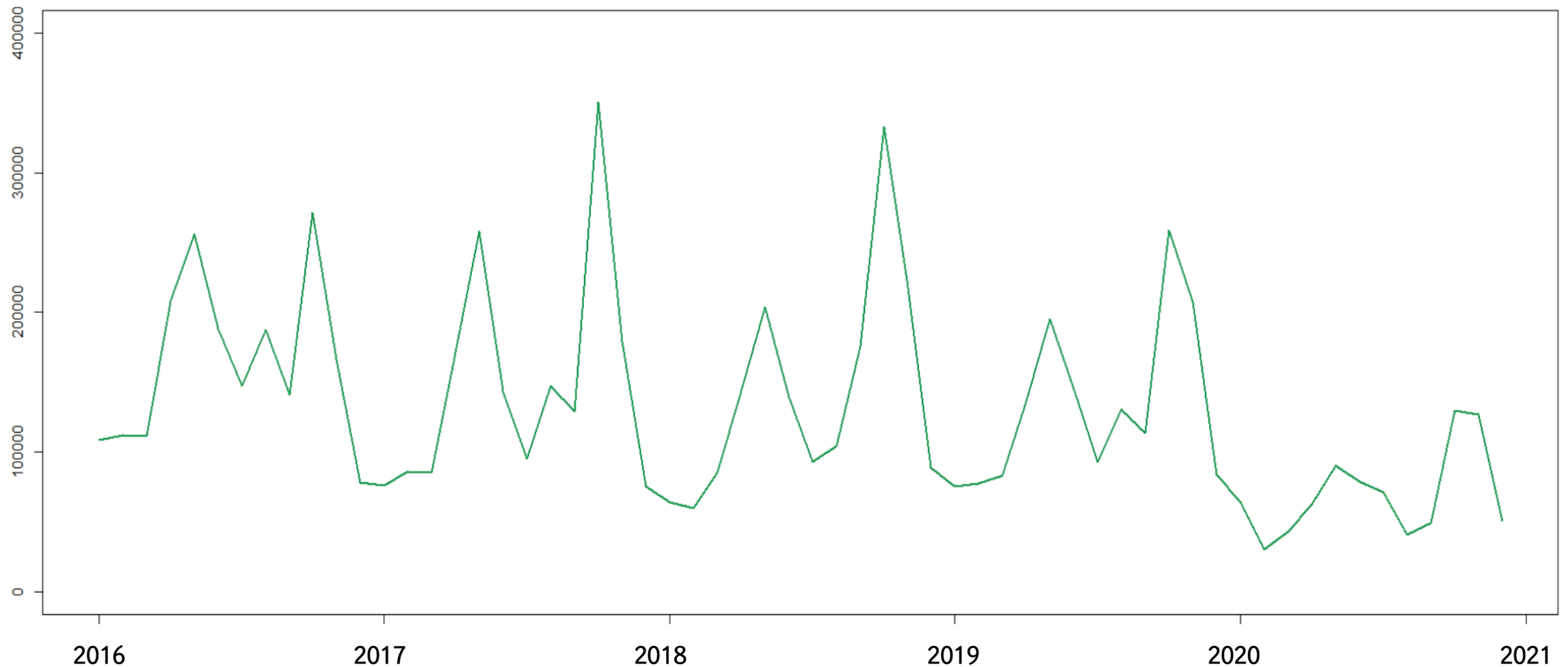
✓ 월별 관광지별 입장객 현황 (2016년 1월 ~ 2020년 12월)





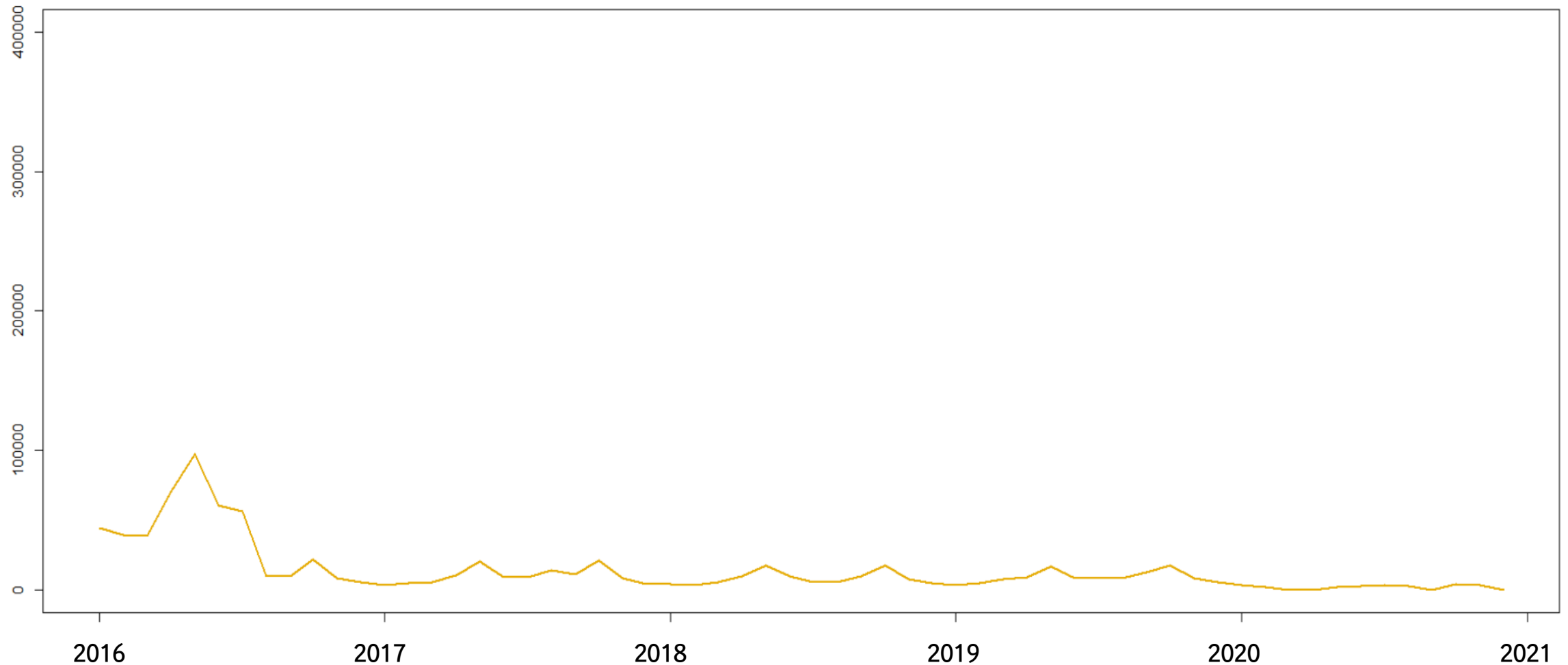
## 02 데이터 설명 - 관광지별 입장객 데이터

✓ 남이섬 유원지의 월별 입장객 현황 (2016년 1월 ~ 2020년 12월)



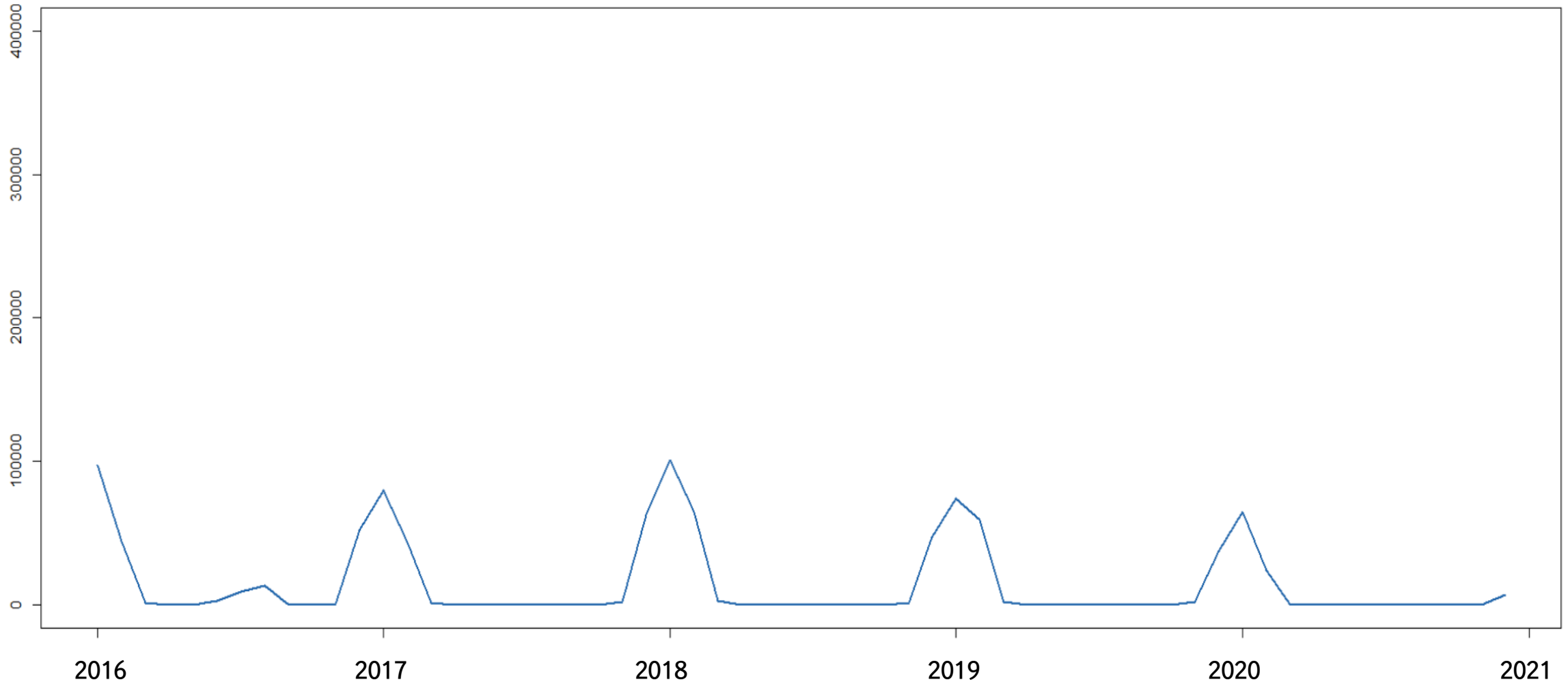
## 02 데이터 설명 - 관광지별 입장객 데이터

✓ 김유정 문학마을의 월별 입장객 현황 (2016년 1월 ~ 2020년 12월)



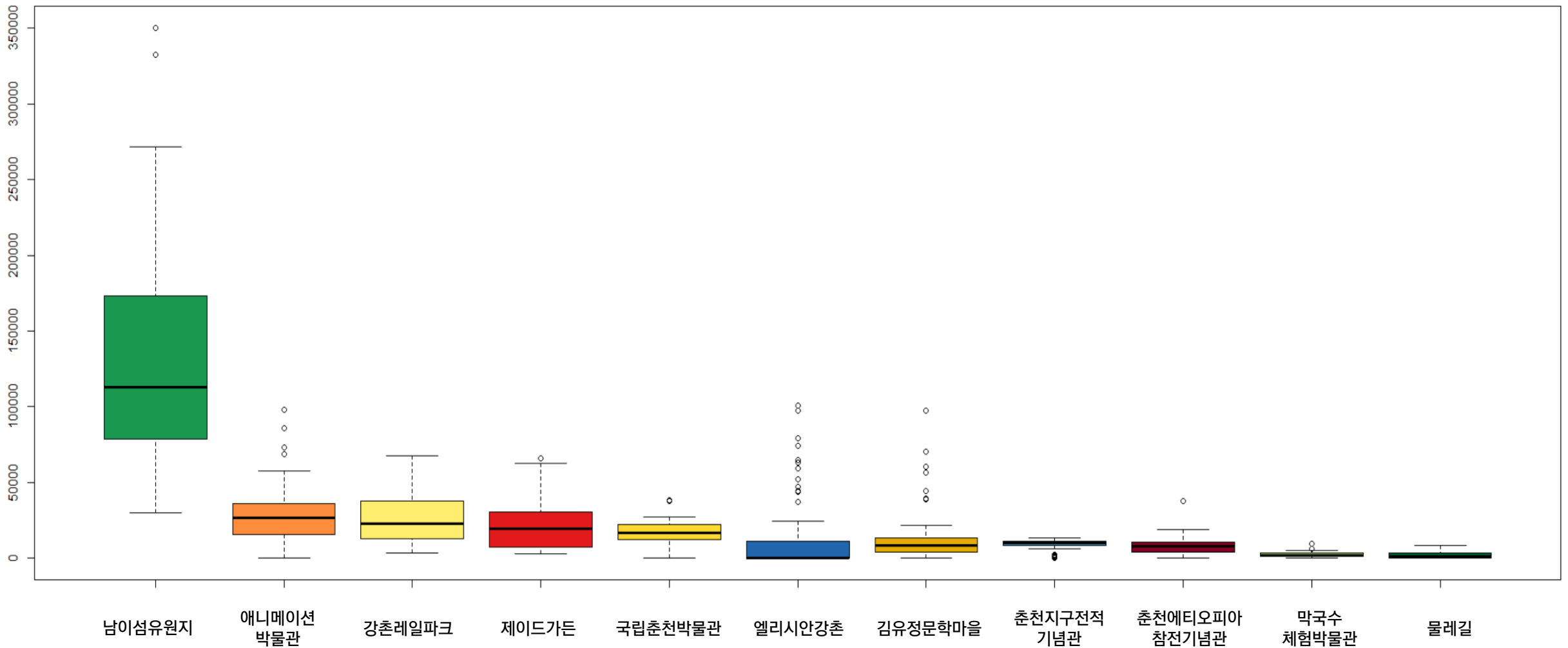
Page 10 of 10

✓ **엘리시안 강촌 스키장 의 월별 입장객 현황 (2016년 1월 ~ 2020년 12월)**



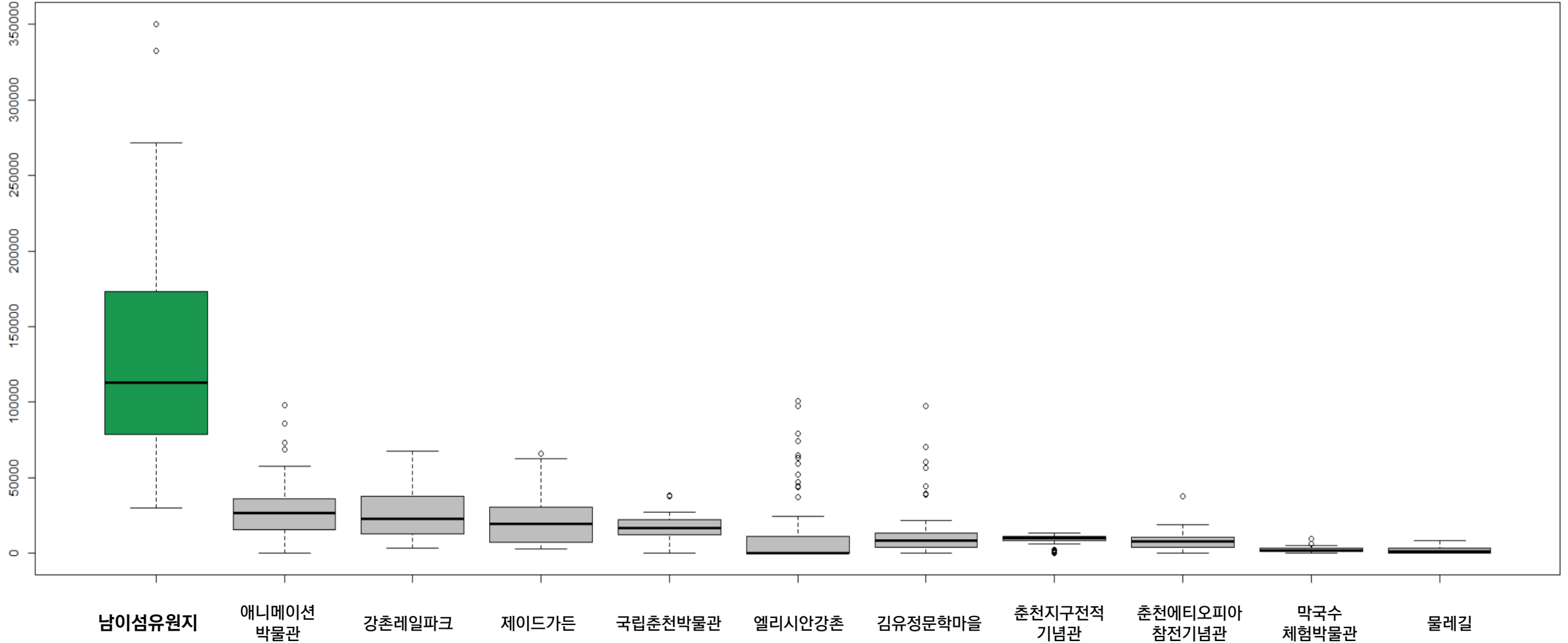
## 02 데이터 설명 - 관광지별 입장객 데이터

### ✓ 관광지별 입장객 현황 상자그림



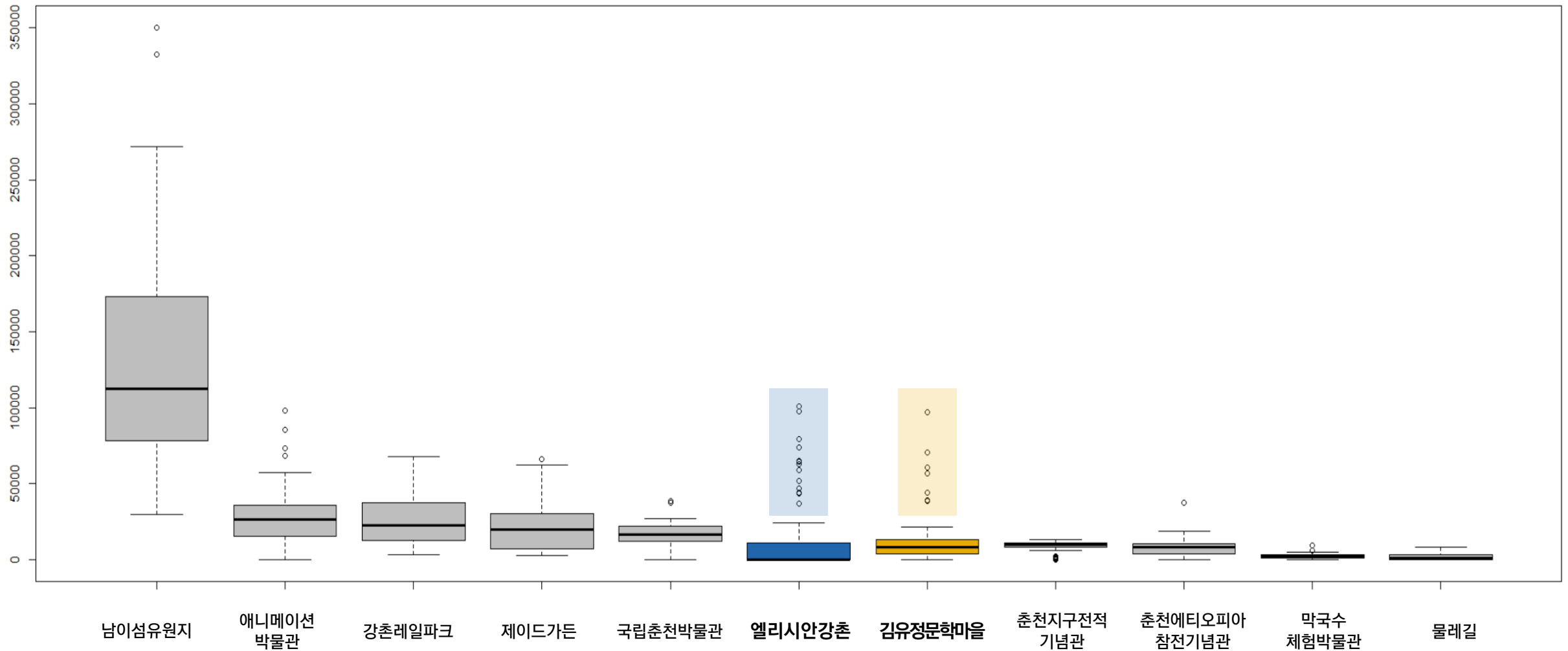
## 02 데이터 설명 - 관광지별 입장객 데이터

### ✓ 관광지별 입장객 현황 상자그림



## 02 데이터 설명 - 관광지별 입장객 데이터

### ✓ 관광지별 입장객 현황 상자그림



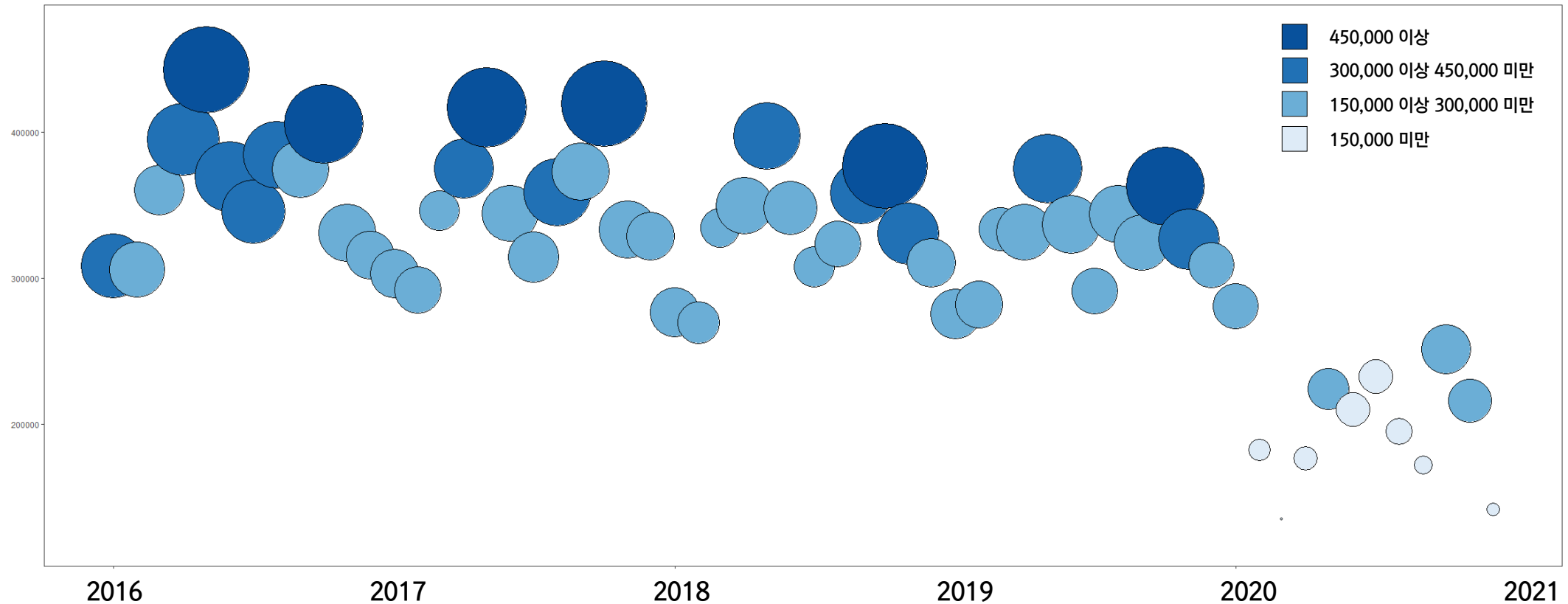
03

## 시계열 분석



### 03 시계열 분석 - 데이터 탐색

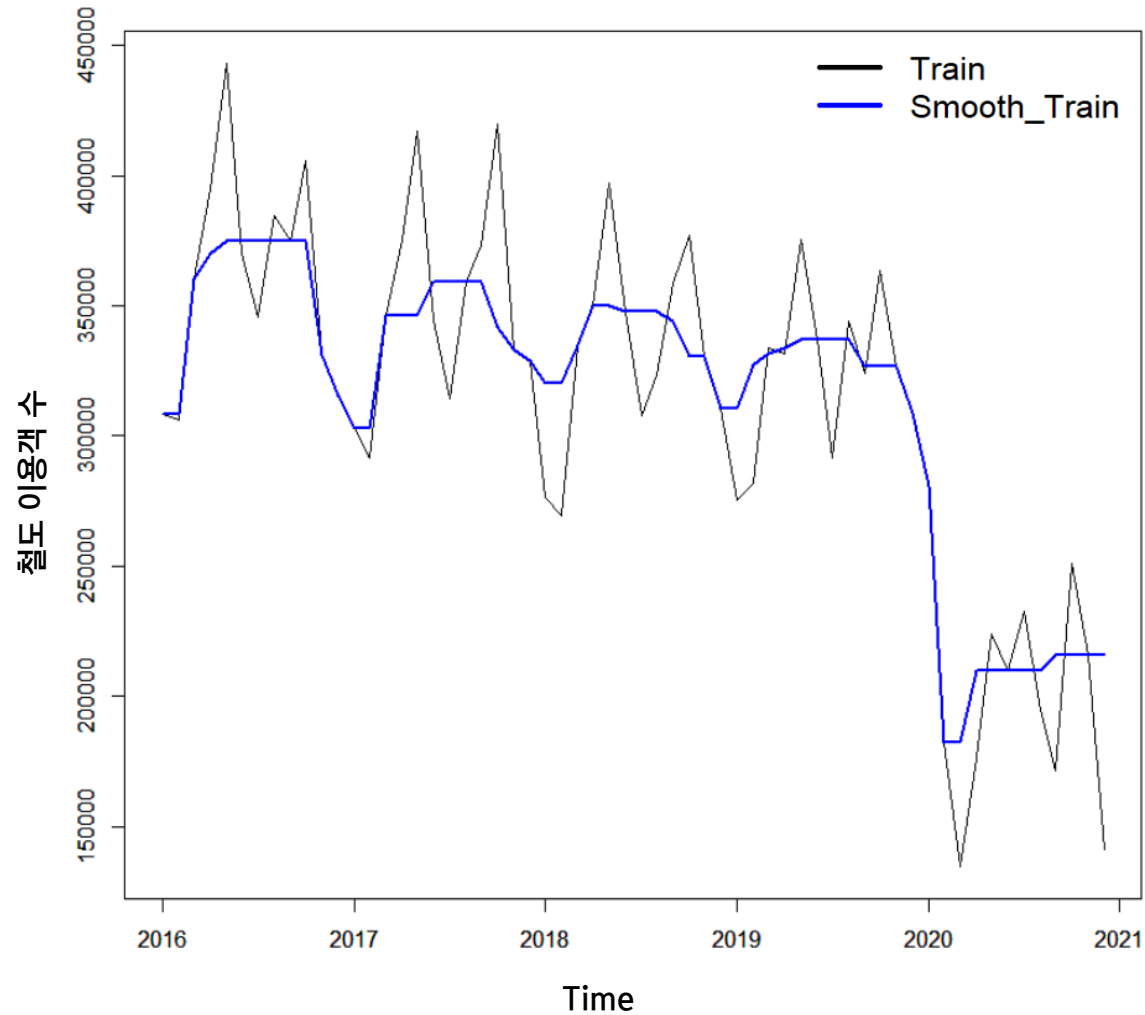
〈 월별 철도 이용객 수와 총 관광객 수에 대한 부호그림 〉



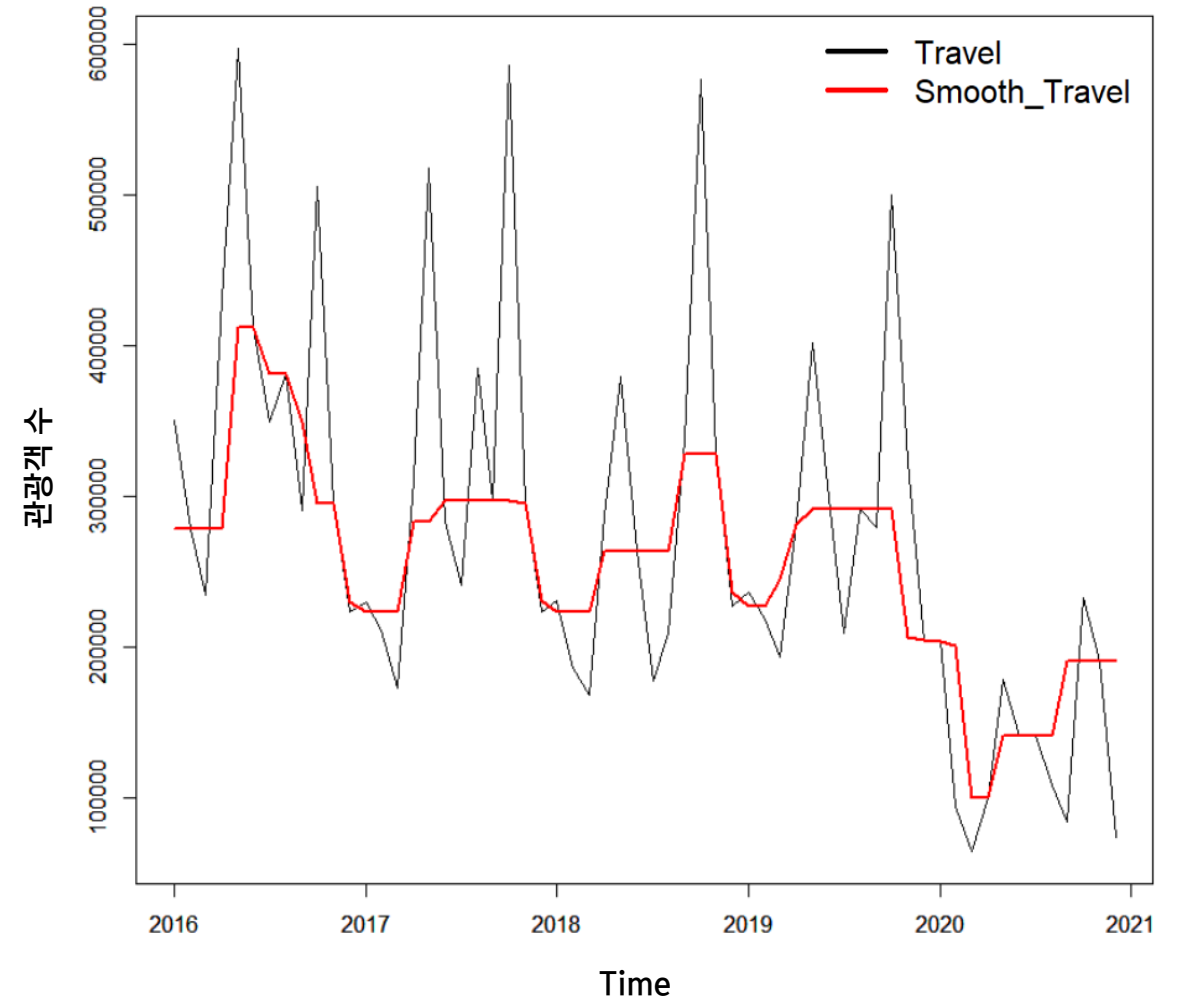
부호그림 (X=연월, Y=철도 이용객 수, Z=관광객 수) 을 그려 확인한 결과  
시간의 흐름에 따라 **하차 승객 수와 총 관광객 수가 감소하는 양상**을 확인

### 03 시계열 분석 - 데이터 탐색

〈 월별 철도 이용객 수 〉

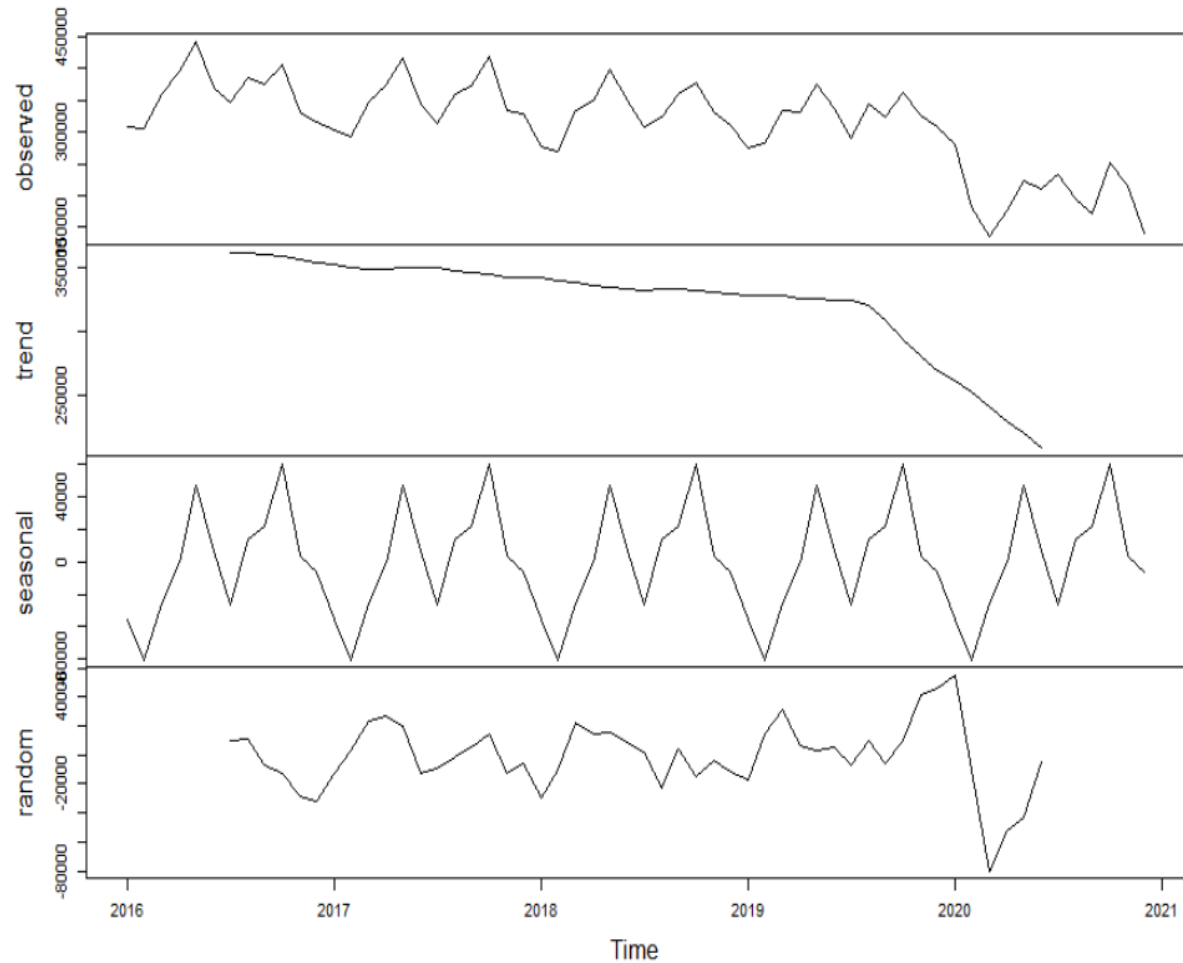


〈 월별 총 관광객 수 〉

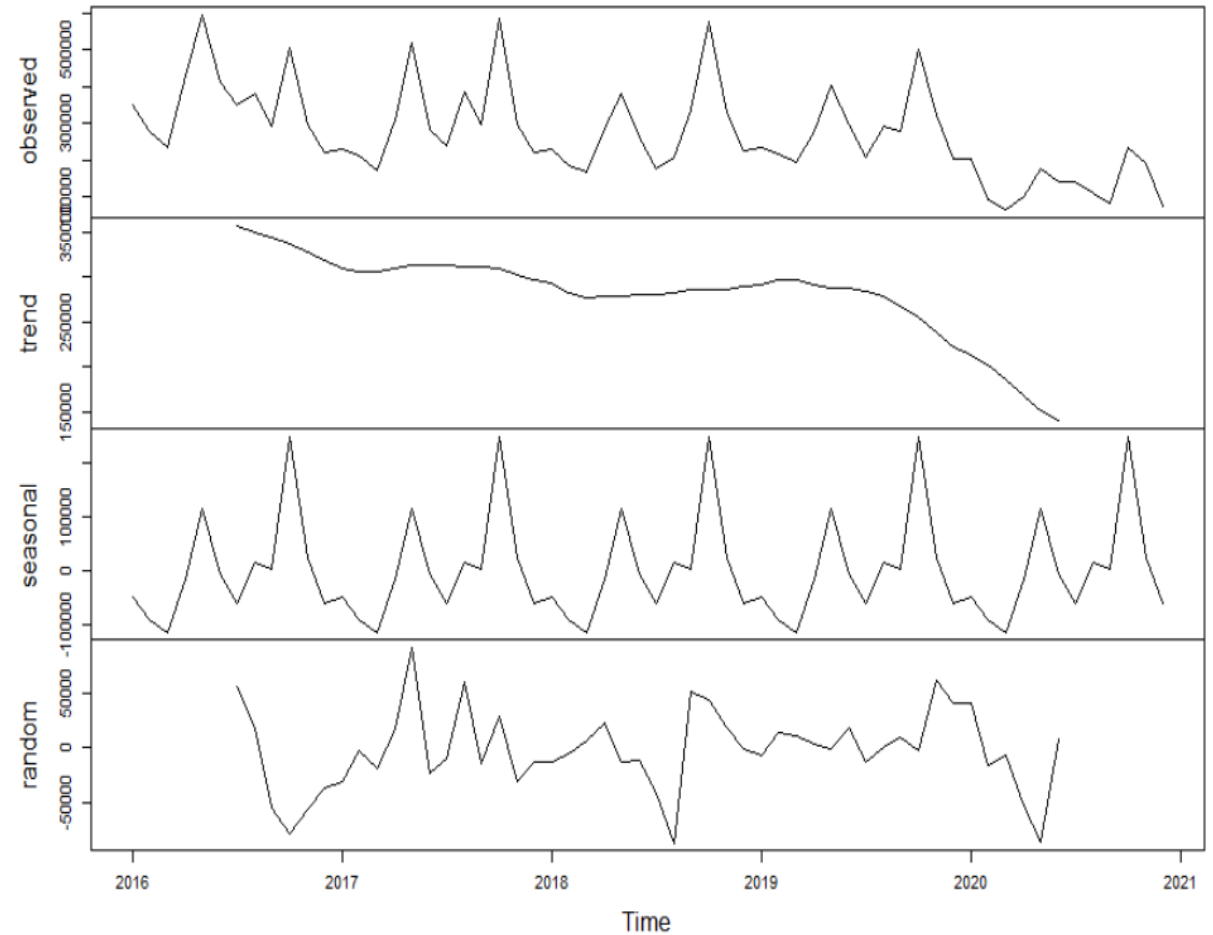


### 03 시계열 분석 - 시계열 분해

〈 월별 철도 이용객 수 시계열 분해 〉

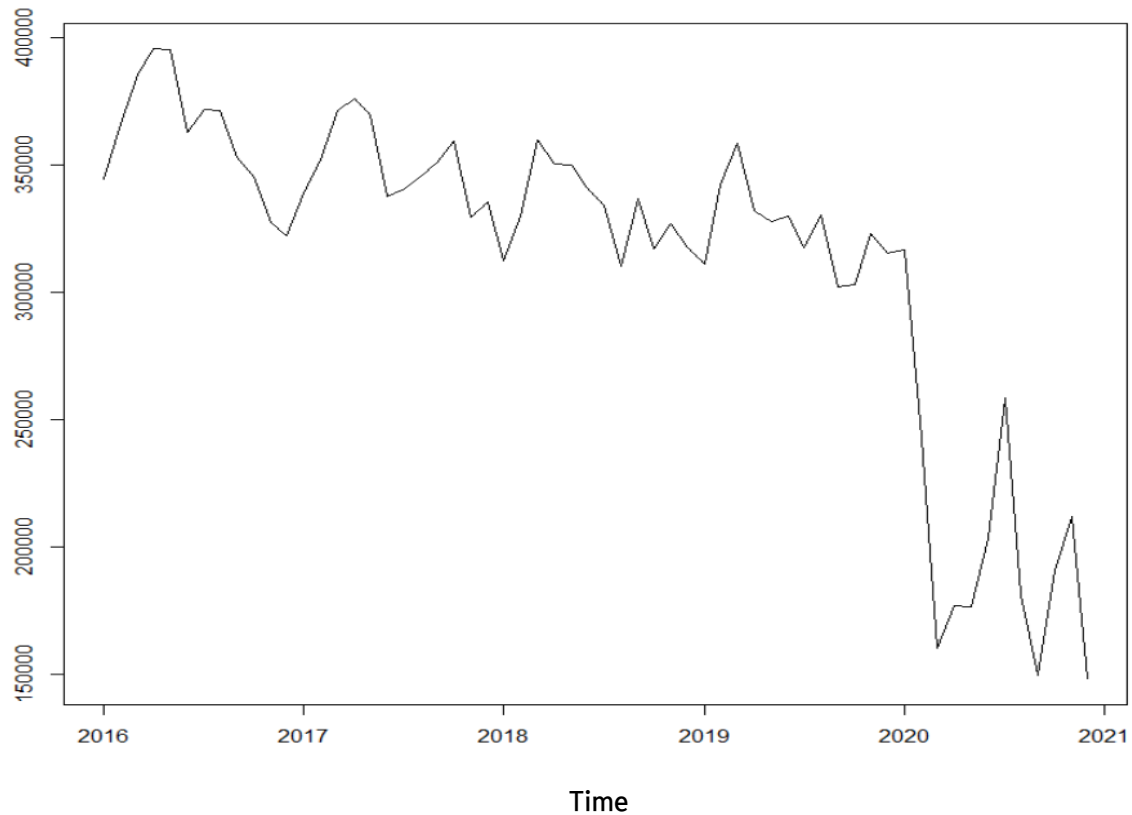


〈 월별 총 관광객 수 시계열 분해 〉

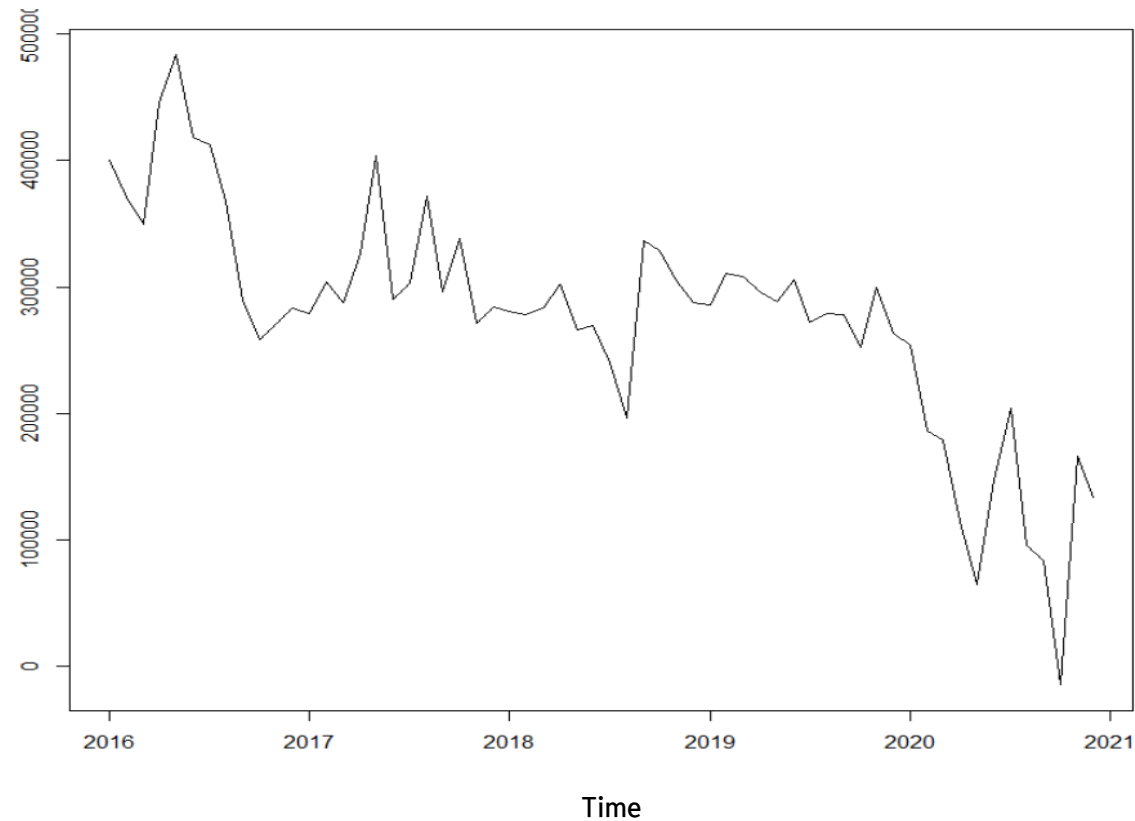


### 03 시계열 분석 - 시계열 분해

〈계절 요인을 제거한 월별 철도 이용객 수〉

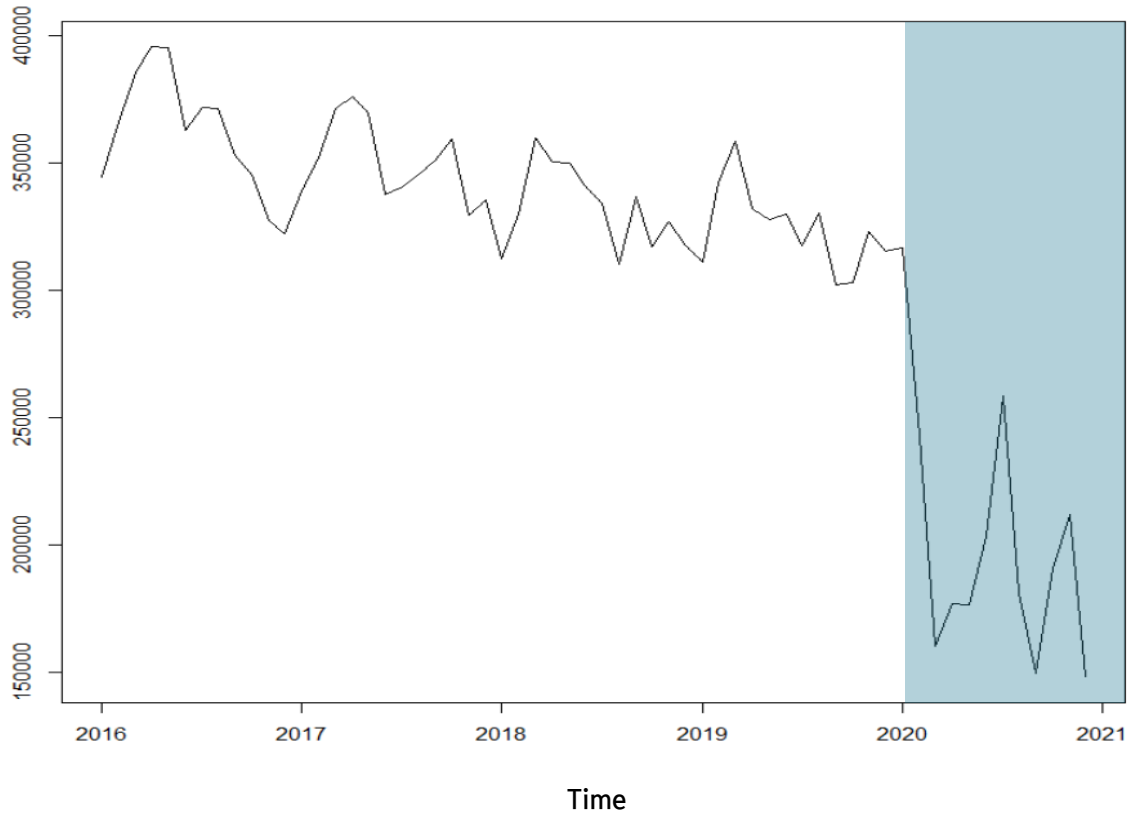


〈계절 요인을 제거한 월별 총 관광객 수〉

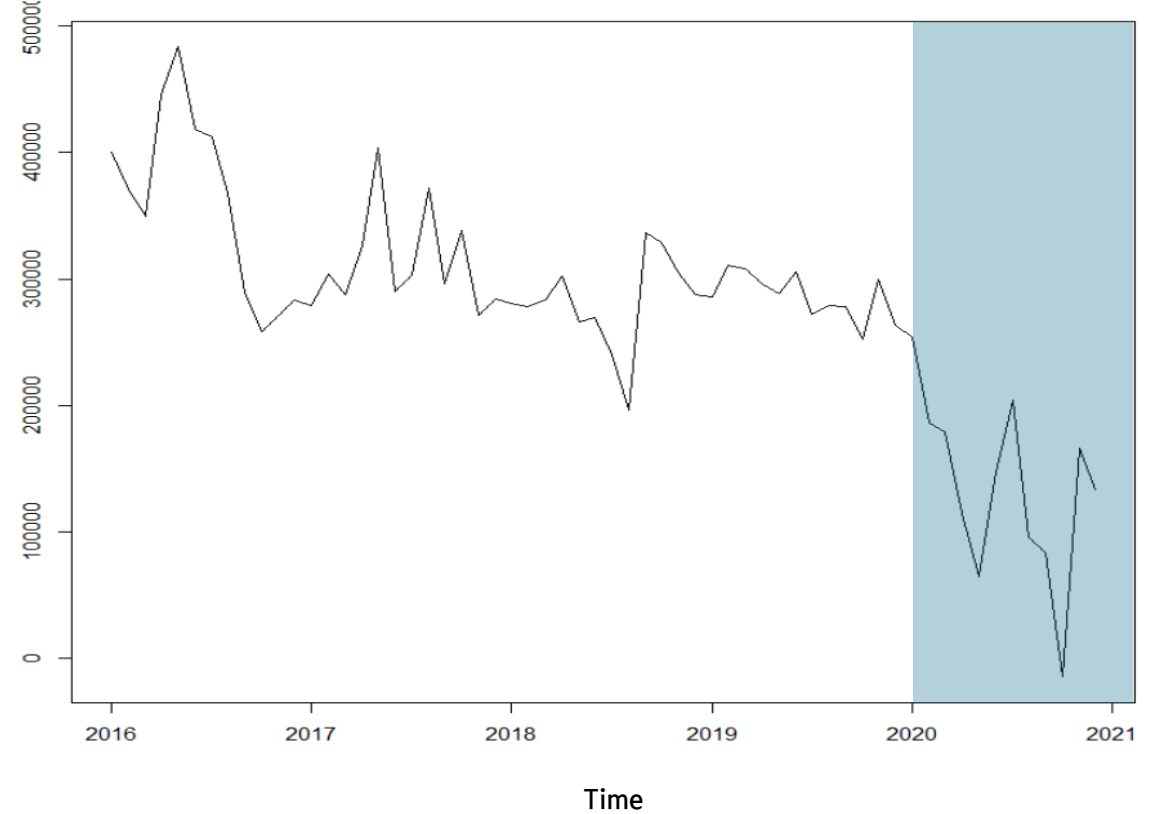


### 03 시계열 분석 - 시계열 분해

〈계절 요인을 제거한 월별 철도 이용객 수〉



〈계절 요인을 제거한 월별 총 관광객 수〉

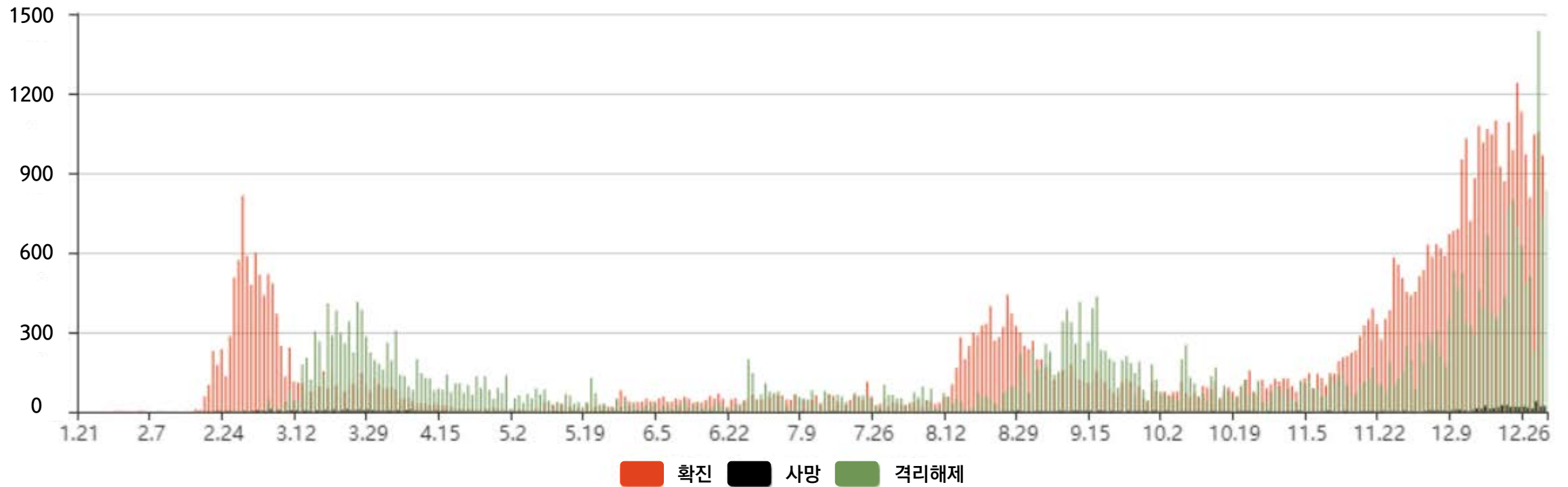


계절성을 제거하였음에도 2020년을 기점으로 시계열 분포에 큰 변화가 식별  
해당 원인으로 **코로나 19의 영향으로 추측**

### 03 시계열 분석 - 코로나 19 이후 정의

#### 대한민국 코로나 19(COVID-19) 추이

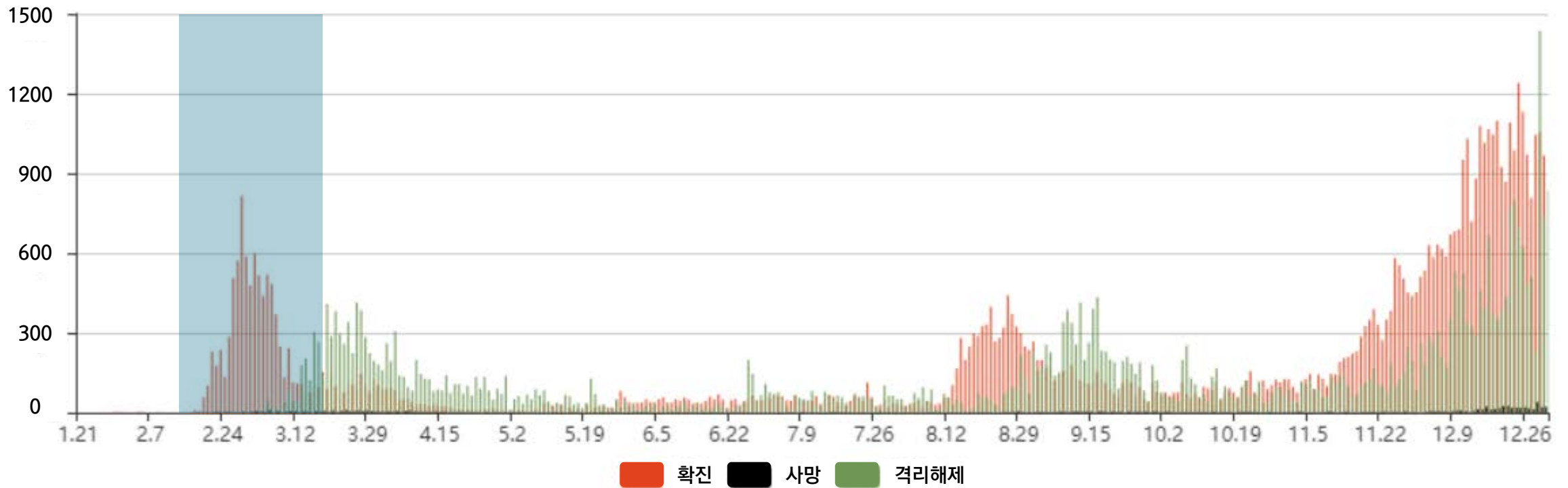
총 확진자 : 138,898명. 사망:1,946명, 격리해제 : 128,761명  
질병관리청 2021-05-28 00:00 집계 기준



### 03 시계열 분석 - 코로나 19 이후 정의

#### 대한민국 코로나 19(COVID-19) 추이

총 확진자 : 138,898명. 사망:1,946명, 격리해제 : 128,761명  
질병관리청 2021-05-28 00:00 집계 기준

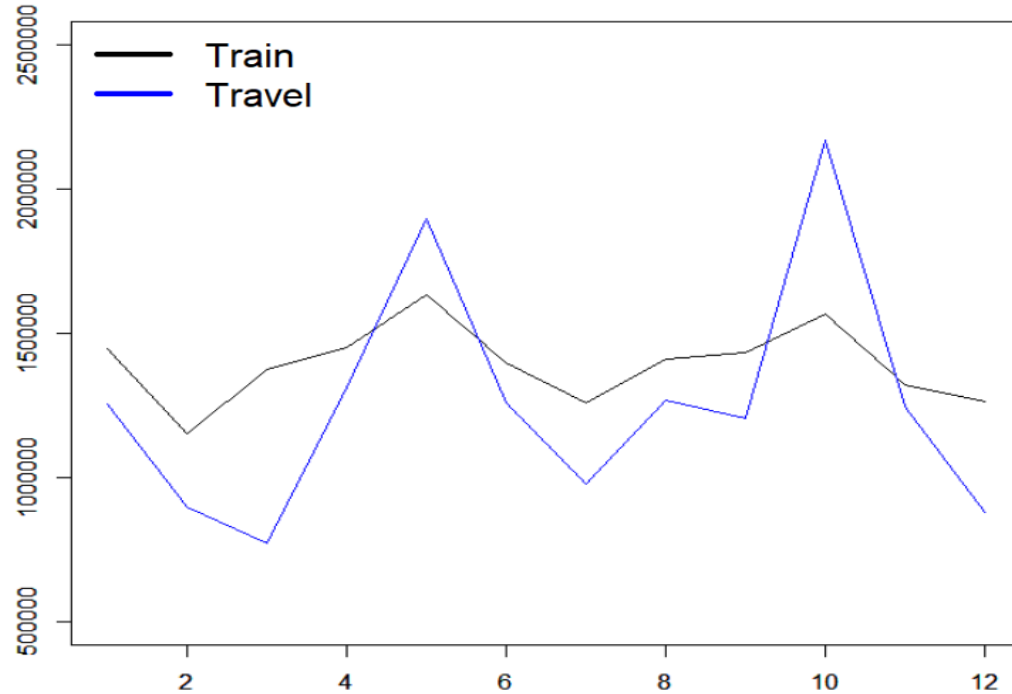


코로나 19가 본격적으로 확산되기 시작한 ‘2월’을 기준으로  
코로나 19 확산 전/후 데이터로 나눠서 시계열 분석을 진행

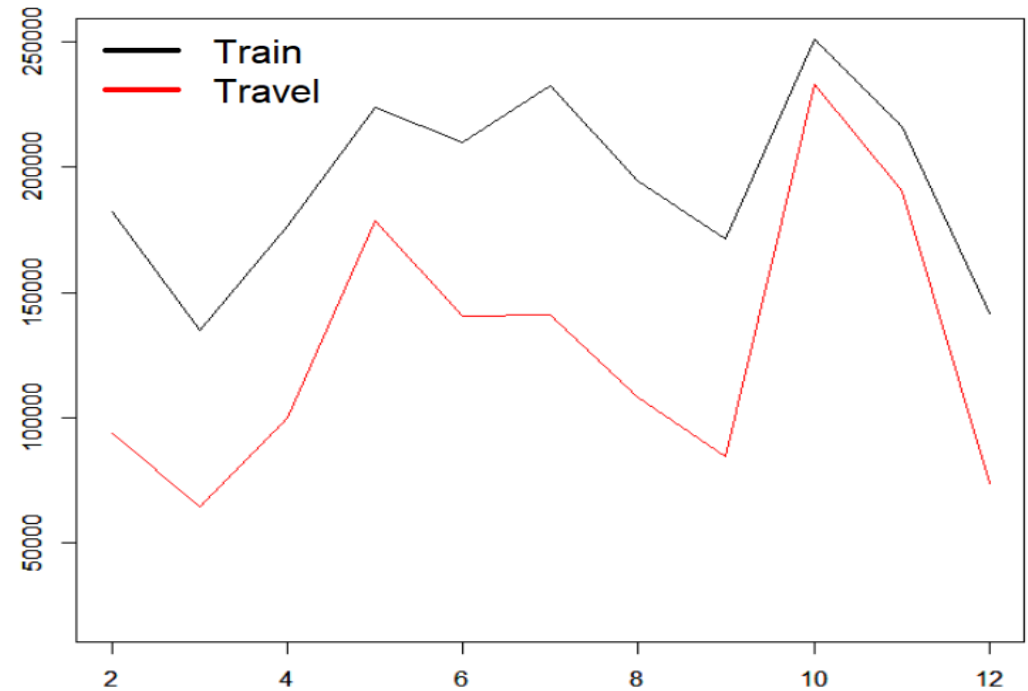


### 03 시계열 분석 - 데이터 비교

〈 코로나 19 이전 월별 철도 이용객 수 및 총 관광객 수 〉



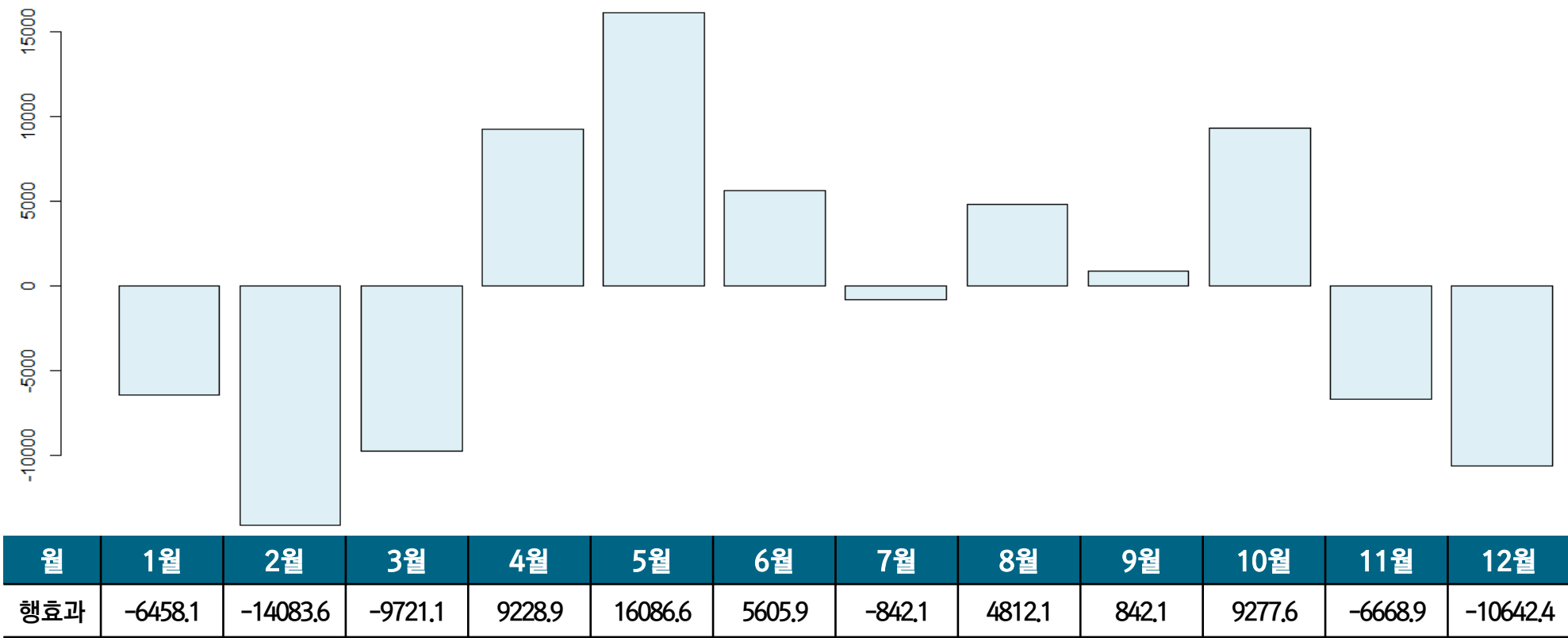
〈 코로나 19 이후 월별 철도 이용객 수 및 총 관광객 수 〉



코로나 19 전/후로 나눠서 월별 철도 이용객 및 총 관광객 수의 합을 확인한 결과  
각각의 시점 모두 비슷한 증감을 나타내는 것을 확인

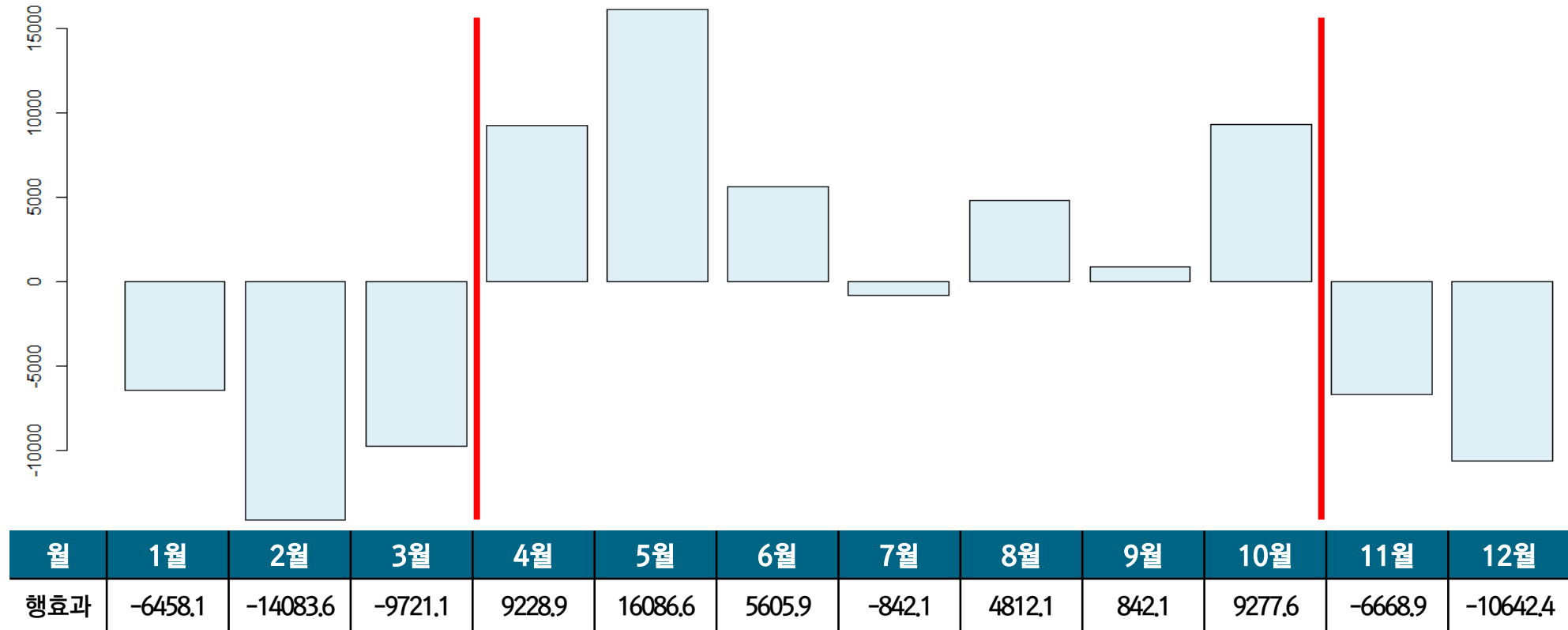
### 03 시계열 분석 - 중간값 다듬기

〈 코로나 19 이전 Row Effects 〉



### 03 시계열 분석 - 중간값 다듬기

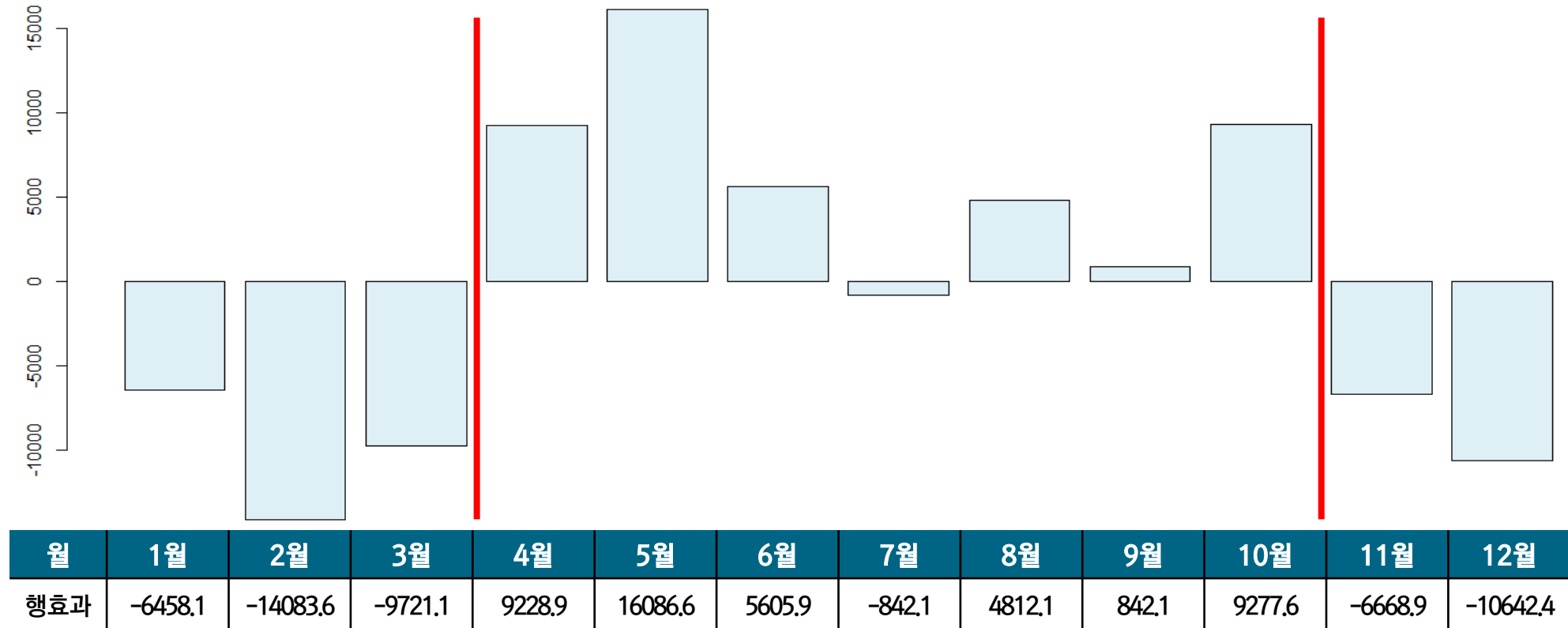
〈 코로나 19 이전 Row Effects 〉



11월 부터 3월까지(겨울)의 행효과는 음수 값으로, 7월을 제외한 나머지 시점은 양수 값으로 나옴

### 03 시계열 분석 - 중간값 다듬기

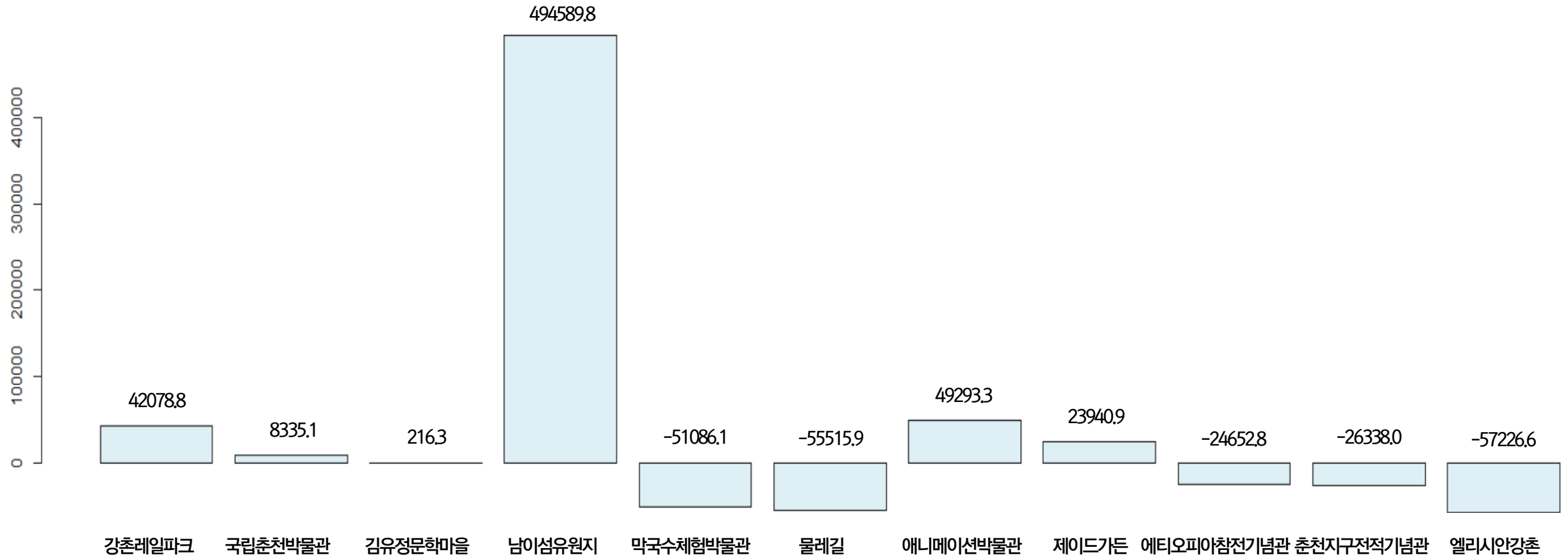
〈 코로나 19 이전 Row Effects 〉



11월 부터 3월까지(겨울)의 행효과는 음수 값으로, 7월을 제외한 나머지 시점은 양수 값으로 나옴  
야외 활동을 하기 용이한 봄, 여름, 가을에 관광객이 많다는 것을 알 수 있음

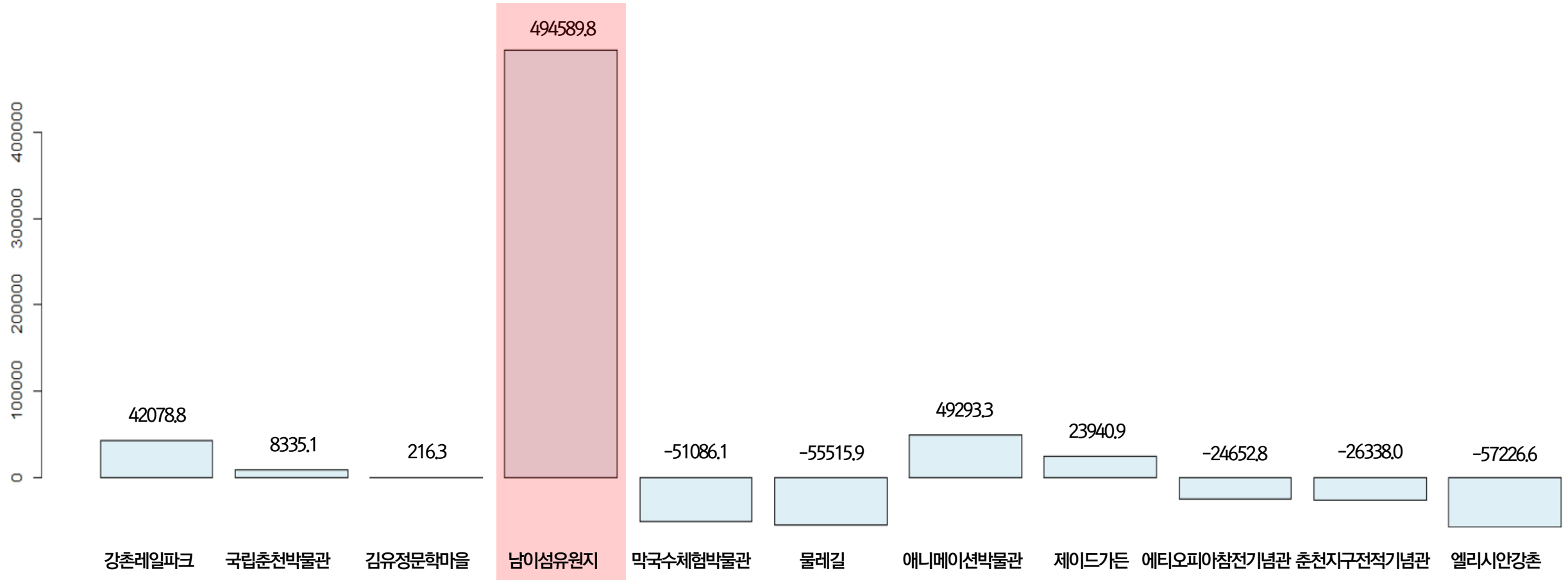
### 03 시계열 분석 - 중간값 다듬기

#### 〈 코로나 19 이전 Column Effects 〉



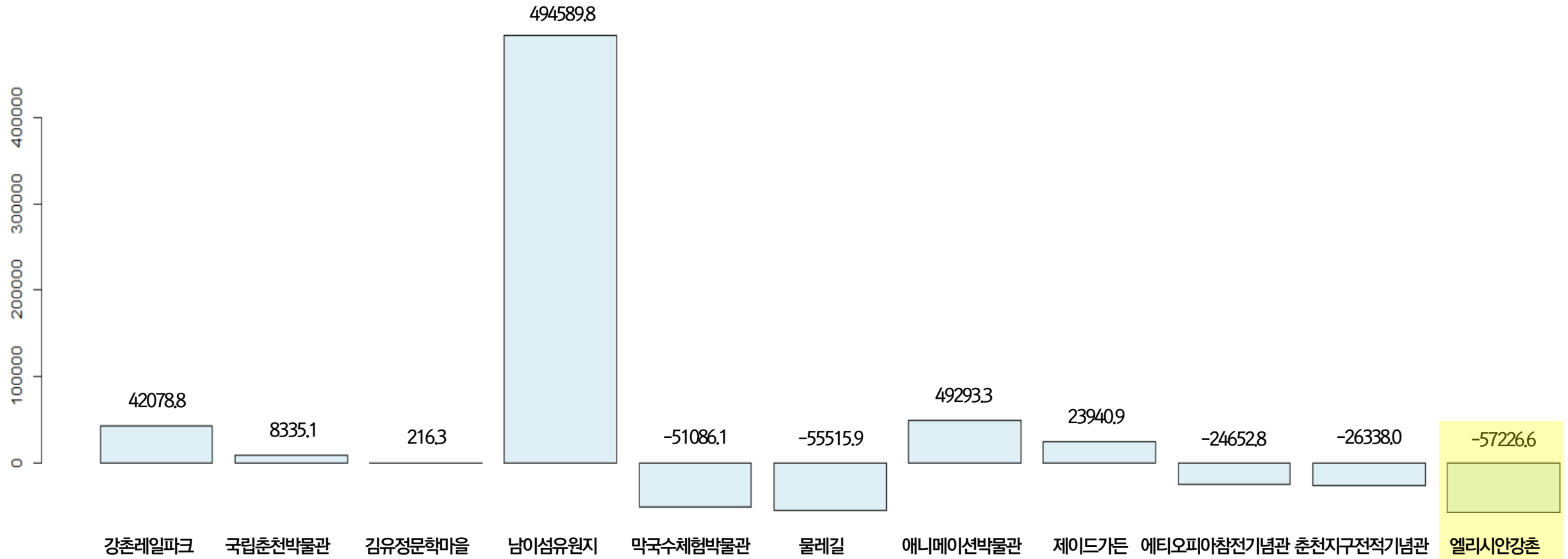
### 03 시계열 분석 - 중간값 다듬기

#### 〈 코로나 19 이전 Column Effects 〉



### 03 시계열 분석 - 중간값 다듬기

#### 〈 코로나 19 이전 Column Effects 〉





## 03 시계열 분석 - 중간값 다듬기

### 〈 코로나 19 이전 Residuals 〉

	남이섬유원지	강촌레일파크	제이드가든	김유정문학마을	물레길	애니메이션박물관	국립춘천박물관	춘천지구전적 기념관	막국수체험박물관	에티오피아 참전 기념관	엘리시안강촌
01월	-166099.2	-53030.2	-64043.4	-648.8	-3521.6	33196.2	14994.4	10840.5	153.6	3075.2	413543.1
02월	-211275.8	-38150.8	-56488.9	-216.2	4556.9	-9166.2	1224.9	4968	3246.1	-995.2	217049.6
03월	-186602.2	-12505.2	-44910.4	216.2	2398.4	-31086.8	-9359.6	12099.5	153.6	3372.2	7622.1
04월	89284.8	31588.8	7416.6	23928.2	-8468.6	-35662.8	20667.4	-366.5	-17136.4	-18979.8	-18503.9
05월	336522	82031	92385.9	69084.5	-129.4	78315.5	-2079.4	-9000.2	-16455.1	-20311.5	-25249.6
06월	48787.8	37793.8	31879.6	15745.2	1021.4	-18865.8	-15180.6	-366.5	-12353.4	-15132.8	-12204.9
07월	-131866.2	-23208.2	-12035.4	13477.2	129.4	9166.2	-6387.6	366.5	153.6	892.2	844.1
08월	3202.5	1547.5	18815.4	-33540	-3175.9	80589	-1224.9	-4571.8	2223.4	16037	-844.1
09월	-3202.5	21801.5	11561.4	-23716	2558.1	53610	14964.1	-1984.8	-5366.6	0	-9989.1
10월	642883	99924	88837.9	1743.5	-129.4	144896.5	-8642.4	-3101.2	-8941.1	-6005.5	-18552.6
11월	216689.5	-1547.5	-7416.6	-27781	2216.1	-21620	7839.1	9691.2	-153.6	0	2953.9
12월	-224343	-64064	-57595.1	-35832.5	-129.4	-10902.5	31511.6	11430.8	1483.9	7317.5	200440.4

## 03 시계열 분석 - 중간값 다듬기

### 〈 코로나 19 이전 Residuals 〉

	남이섬유원지	강촌레일파크	제이드가든	김유정문학마을	물레길	애니메이션박물관	국립춘천박물관	춘천지구전적 기념관	막국수체험박물관	에티오피아 참전 기념관	엘리시안강촌
01월	-166099.2	-53030.2	-64043.4	-648.8	-3521.6	33196.2	14994.4	10840.5	153.6	3075.2	413543.1
02월	-211275.8	-38150.8	-56488.9	-216.2	4556.9	-9166.2	1224.9	4968	3246.1	-995.2	217049.6
03월	-186602.2	-12505.2	-44910.4	216.2	2398.4	-31086.8	-9359.6	12099.5	153.6	3372.2	7622.1
04월	89284.8	31588.8	7416.6	23928.2	-8468.6	-35662.8	20667.4	-366.5	-17136.4	-18979.8	-18503.9
05월	336522	82031	92385.9	69084.5	-129.4	78315.5	-2079.4	-9000.2	-16455.1	-20311.5	-25249.6
06월	48787.8	37793.8	31879.6	15745.2	1021.4	-18865.8	-15180.6	-366.5	-12353.4	-15132.8	-12204.9
07월	-131866.2	-23208.2	-12035.4	13477.2	129.4	9166.2	-6387.6	366.5	153.6	892.2	844.1
08월	3202.5	1547.5	18815.4	-33540	-3175.9	80589	-1224.9	-4571.8	2223.4	16037	-844.1
09월	-3202.5	21801.5	11561.4	-23716	2558.1	53610	14964.1	-1984.8	-5366.6	0	-9989.1
10월	642883	99924	88837.9	1743.5	-129.4	144896.5	-8642.4	-3101.2	-8941.1	-6005.5	-18552.6
11월	216689.5	-1547.5	-7416.6	-27781	2216.1	-21620	7839.1	9691.2	-153.6	0	2953.9
12월	-224343	-64064	-57595.1	-35832.5	-129.4	-10902.5	31511.6	11430.8	1483.9	7317.5	200440.4

## 03 시계열 분석 - 중간값 다듬기

### 〈 코로나 19 이전 Residuals 〉

	남이섬유원지	강촌레일파크	제이드가든	김유정문학마을	물레길	애니메이션박물관	국립춘천박물관	춘천지구전적 기념관	막국수체험박물관	에티오피아 참전 기념관	엘리시안강촌
01월	-166099.2	-53030.2	-64043.4	-648.8	-3521.6	33196.2	14994.4	10840.5	153.6	3075.2	413543.1
02월	-211275.8	-38150.8	-56488.9	-216.2	4556.9	-9166.2	1224.9	4968	3246.1	-995.2	217049.6
03월	-186602.2	-12505.2	-44910.4	216.2	2398.4	-31086.8	-9359.6	12099.5	153.6	3372.2	7622.1
04월	89284.8	31588.8	7416.6	23928.2	-8468.6	-35662.8	20667.4	-366.5	-17136.4	-18979.8	-18503.9
05월	336522	82031	92385.9	69084.5	-129.4	78315.5	-2079.4	-9000.2	-16455.1	-20311.5	-25249.6
06월	48787.8	37793.8	31879.6	15745.2	1021.4	-18865.8	-15180.6	-366.5	-12353.4	-15132.8	-12204.9
07월	-131866.2	-23208.2	-12035.4	13477.2	129.4	9166.2	-6387.6	366.5	153.6	892.2	844.1
08월	3202.5	1547.5	18815.4	-33540	-3175.9	80589	-1224.9	-4571.8	2223.4	16037	-844.1
09월	-3202.5	21801.5	11561.4	-23716	2558.1	53610	14964.1	-1984.8	-5366.6	0	-9989.1
10월	642883	99924	88837.9	1743.5	-129.4	144896.5	-8642.4	-3101.2	-8941.1	-6005.5	-18552.6
11월	216689.5	-1547.5	-7416.6	-27781	2216.1	-21620	7839.1	9691.2	-153.6	0	2953.9
12월	-224343	-64064	-57595.1	-35832.5	-129.4	-10902.5	31511.6	11430.8	1483.9	7317.5	200440.4

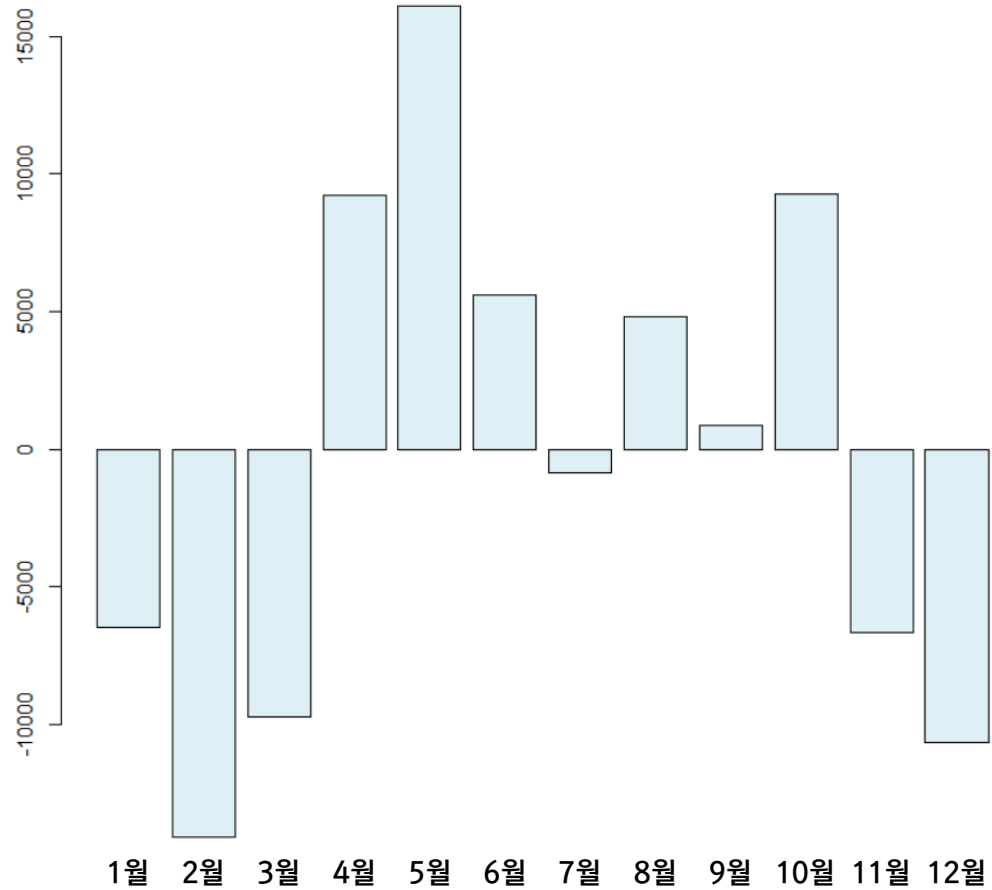
## 03 시계열 분석 - 중간값 다듬기

### 〈 코로나 19 이전 Residuals 〉

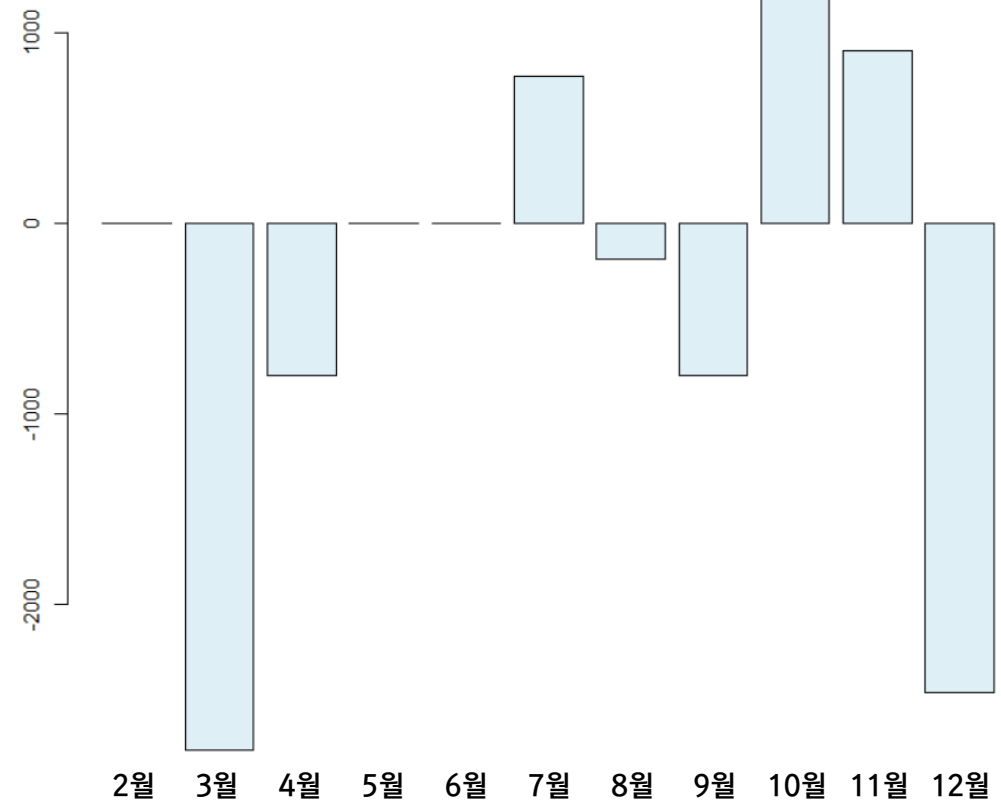
	남이섬유원지	강촌레일파크	제이드가든	김유정문학마을	물레길	애니메이션박물관	국립춘천박물관	춘천지구전적 기념관	막국수체험박물관	에티오피아 참전 기념관	엘리시안강촌
01월	-166099.2	-53030.2	-64043.4	-648.8	-3521.6	33196.2	14994.4	10840.5	153.6	3075.2	413543.1
02월	-211275.8	-38150.8	-56488.9	-216.2	4556.9	-9166.2	1224.9	4968	3246.1	-995.2	217049.6
03월	-186602.2	-12505.2	-44910.4	216.2	2398.4	-31086.8	-9359.6	12099.5	153.6	3372.2	7622.1
04월	89284.8	31588.8	7416.6	23928.2	-8468.6	-35662.8	20667.4	-366.5	-17136.4	-18979.8	-18503.9
05월	336522	82031	92385.9	69084.5	-129.4	78315.5	-2079.4	-9000.2	-16455.1	-20311.5	-25249.6
06월	48787.8	37793.8	31879.6	15745.2	1021.4	-18865.8	-15180.6	-366.5	-12353.4	-15132.8	-12204.9
07월	-131866.2	-23208.2	-12035.4	13477.2	129.4	9166.2	-6387.6	366.5	153.6	892.2	844.1
08월	3202.5	1547.5	18815.4	-33540	-3175.9	80589	-1224.9	-4571.8	2223.4	16037	-844.1
09월	-3202.5	21801.5	11561.4	-23716	2558.1	53610	14964.1	-1984.8	-5366.6	0	-9989.1
10월	642883	99924	88837.9	1743.5	-129.4	144896.5	-8642.4	-3101.2	-8941.1	-6005.5	-18552.6
11월	216689.5	-1547.5	-7416.6	-27781	2216.1	-21620	7839.1	9691.2	-153.6	0	2953.9
12월	-224343	-64064	-57595.1	-35832.5	-129.4	-10902.5	31511.6	11430.8	1483.9	7317.5	200440.4

### 03 시계열 분석 - 중간값 다듬기

〈 코로나 19 이전 Row Effects 〉

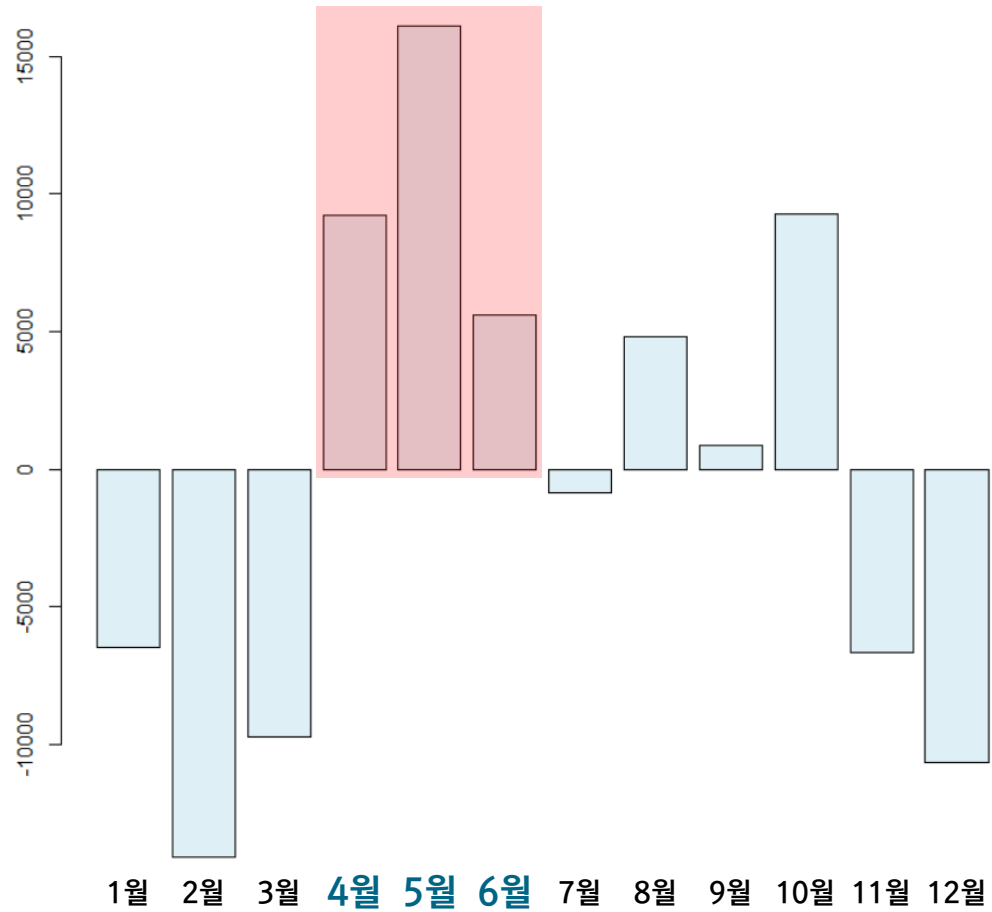


〈 코로나 19 이후 Row Effects 〉

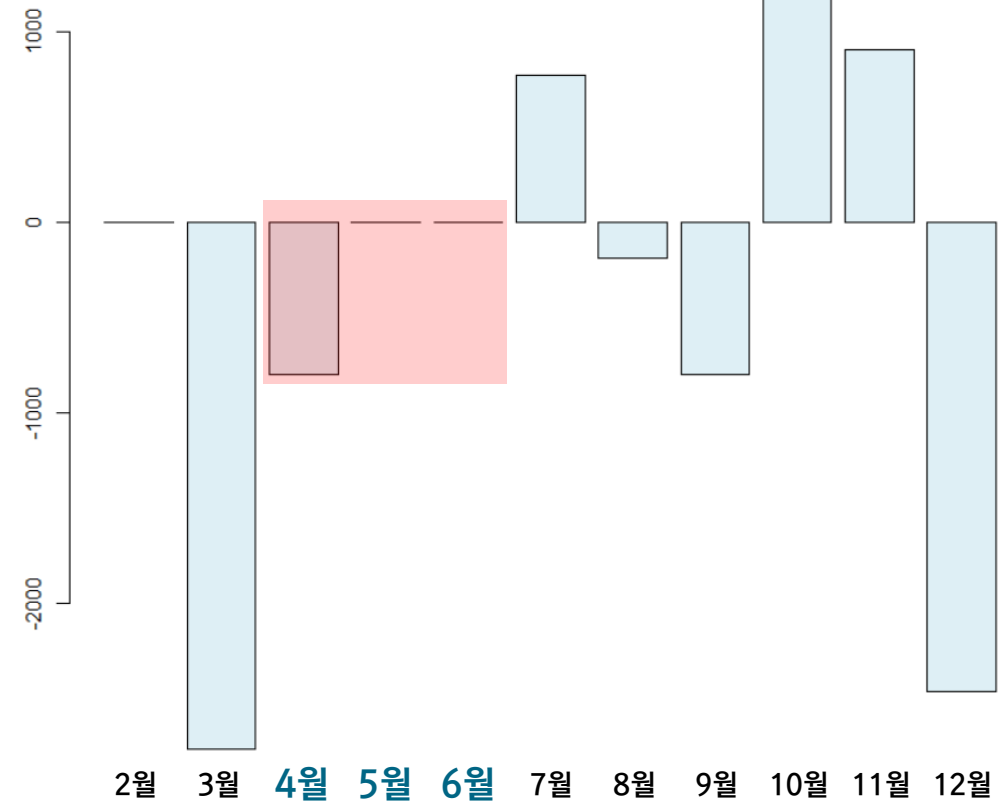


### 03 시계열 분석 - 중간값 다듬기

〈 코로나 19 이전 Row Effects 〉



〈 코로나 19 이후 Row Effects 〉



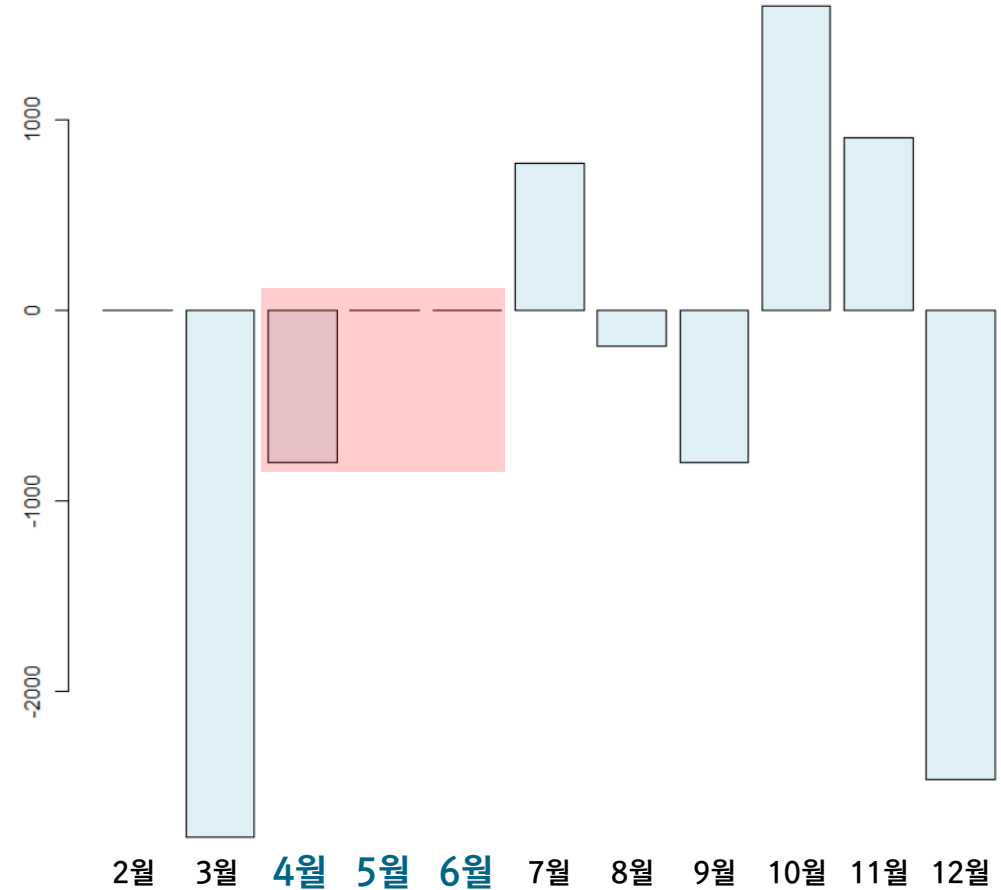
## 03 시계열 분석 - 중간값 다듬기

### 한국 코로나19 일일 확진자 추세

자료: 질병관리청, 단위: 명

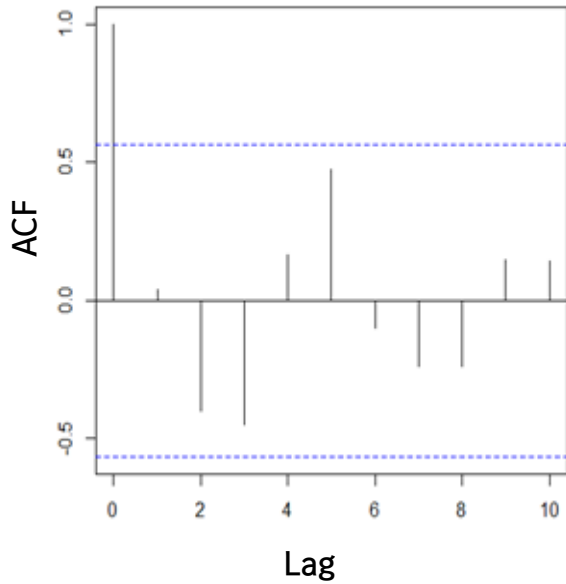


### 〈 코로나 19 이후 Row Effects 〉

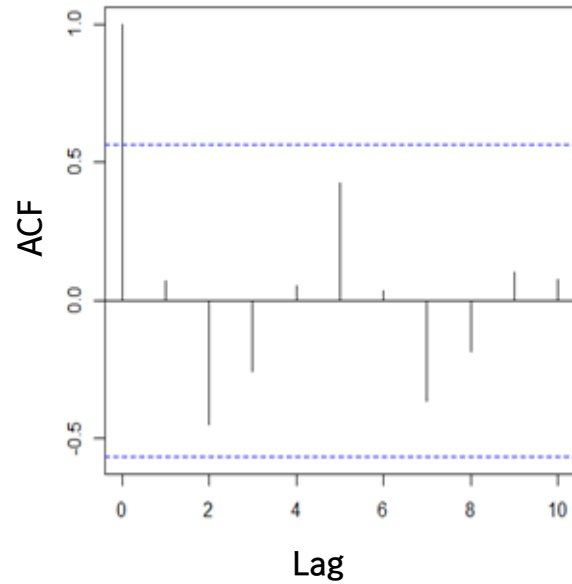


### 03 시계열 분석 - 상관관계

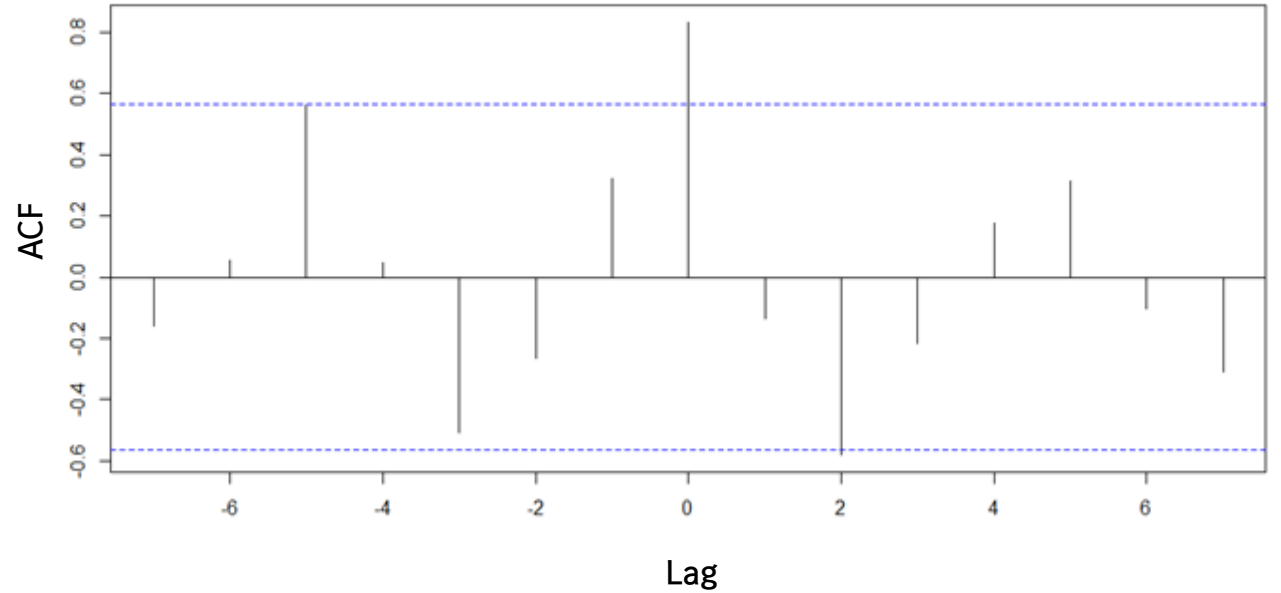
〈 철도 이용객 수 ACF 〉



〈 관광객 수 ACF 〉



〈 철도 이용객수 & 관광객 수 CCF 〉



코로나 19 이전 두 변수의 ACF와 CCF를 확인한 결과  
같은 시점에서 매우 강한 양의 상관관계가 있다고 판단

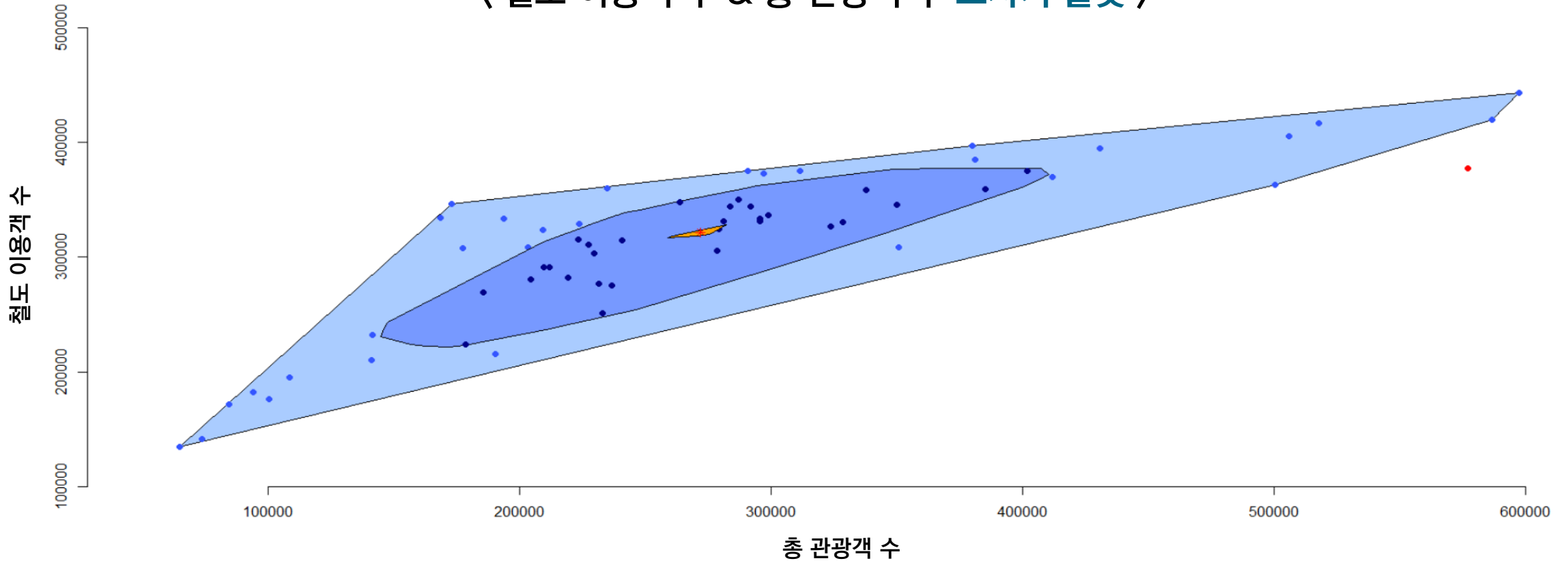


04

## 회귀 분석

## 04 회귀 분석 - 데이터 탐색

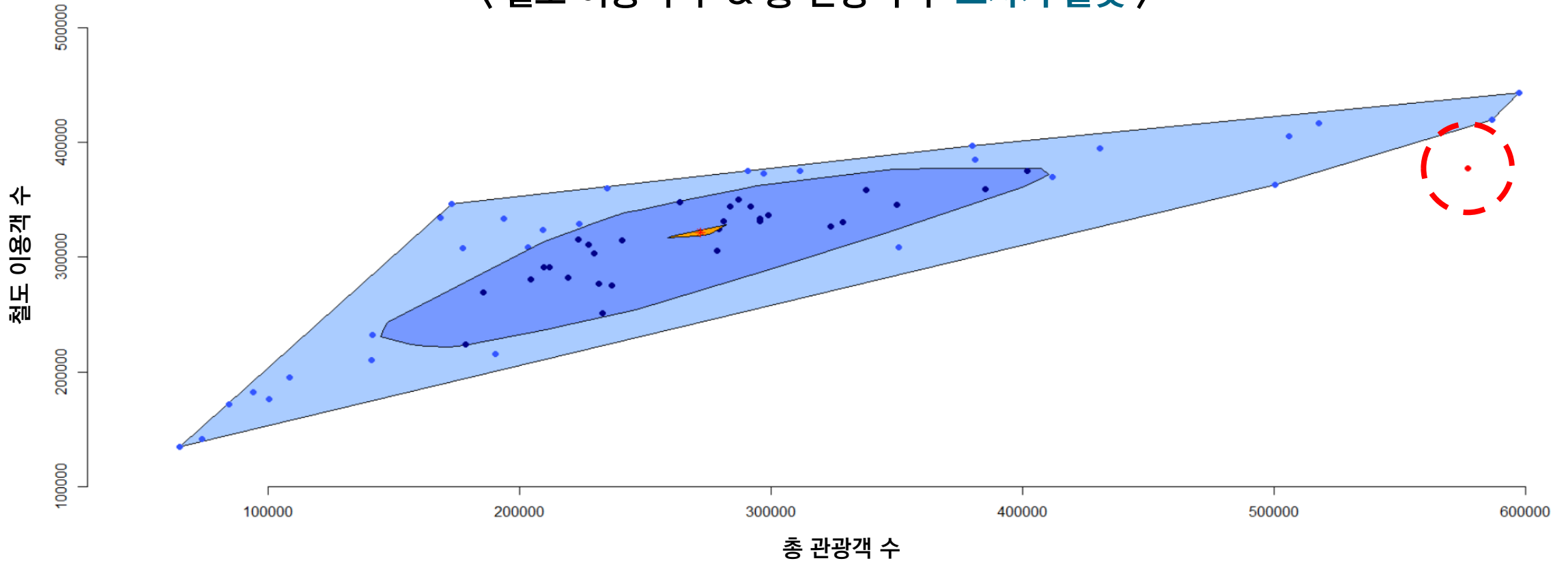
〈 철도 이용객 수 & 총 관광객 수 **보자기 플롯** 〉



보자기 플롯(Y=철도 이용객 수, X=총 관광객 수)를 그려 확인한 결과  
총 관광객 수가 증가할 때, 철도 이용객 수도 함께 증가하는 양상을 확인

## 04 회귀 분석 - 데이터 탐색

〈 철도 이용객 수 & 총 관광객 수 **보자기 플롯** 〉

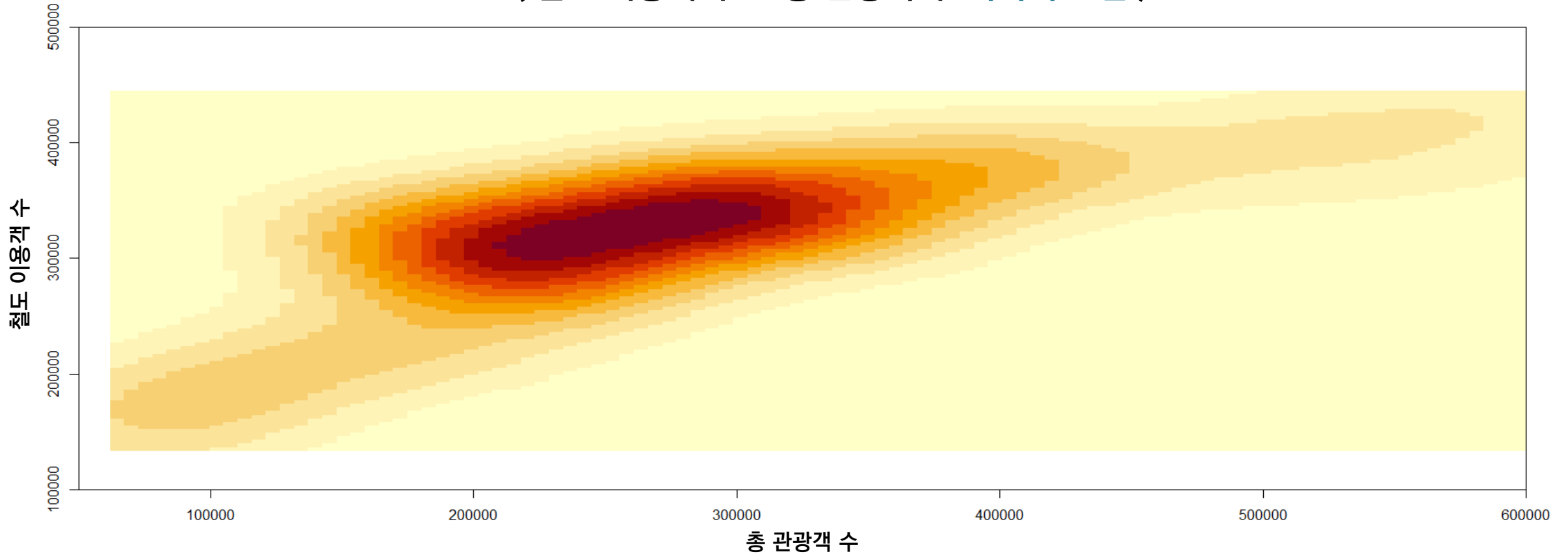


식별된 하나의 이상점에 대해 파악한 결과

2018년 10월의 자료이며 이때의 **철도 이용객 수 : 377,197명 / 총 관광객 수 : 577,155명**

## 04 회귀 분석 - 데이터 탐색

〈 철도 이용객 수 & 총 관광객 수 이미지 그림 〉

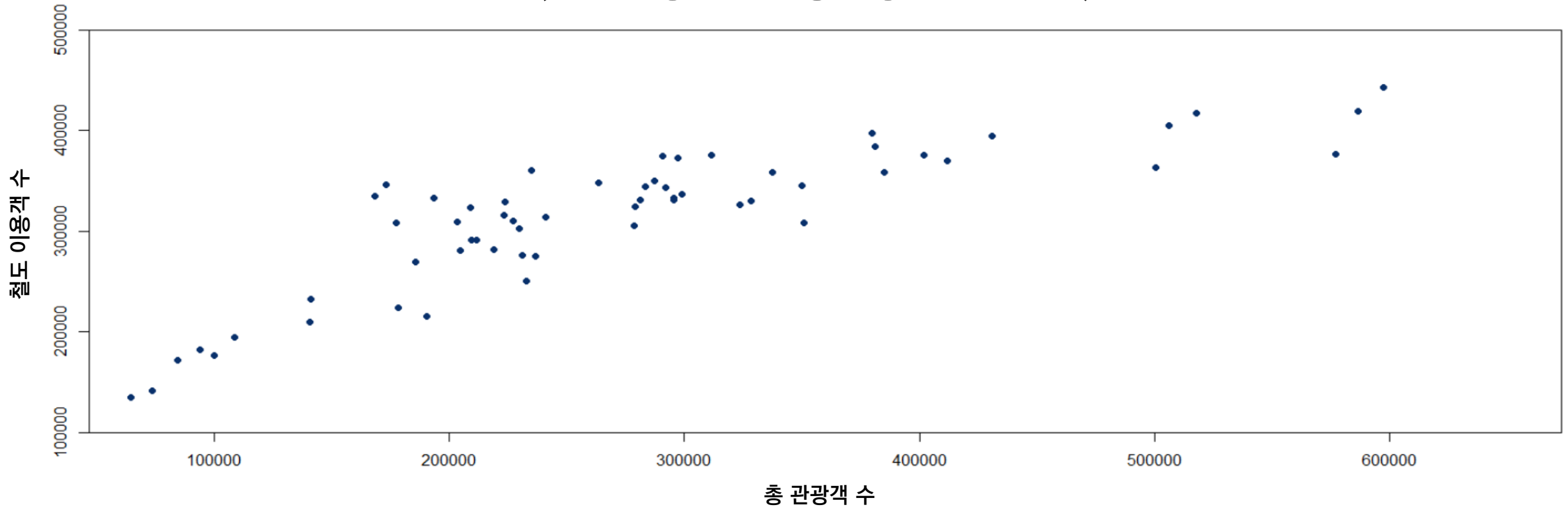


좌표 (250000 , 320000) 을 중심으로 대부분의 자료가 밀집

해당 좌표에 **하나의 봉우리가 존재**하는 것을 확인

## 04 회귀 분석 - 조각별 선형회귀

〈 철도 이용객 수 & 총 관광객 수 산점도 〉

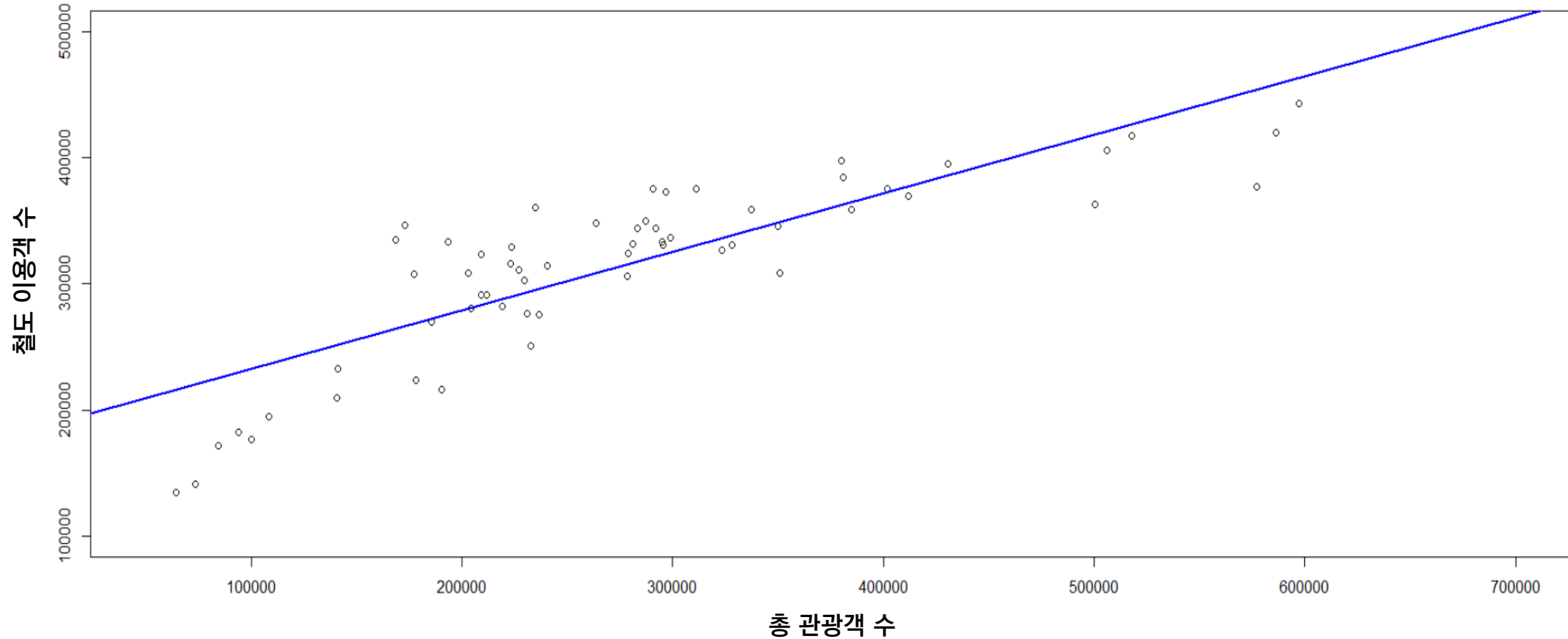


앞서 보자기 플롯과 이미지 그림을 통해 두 변수 사이에 선형관계가 있을 것이라 결론

통계적 기법(Y=철도 이용객 수, X=총 관광객 수)을 통해 수치적으로 어떤 관계가 있는지를 확인

## 04 회귀 분석 - 조각별 선형회귀

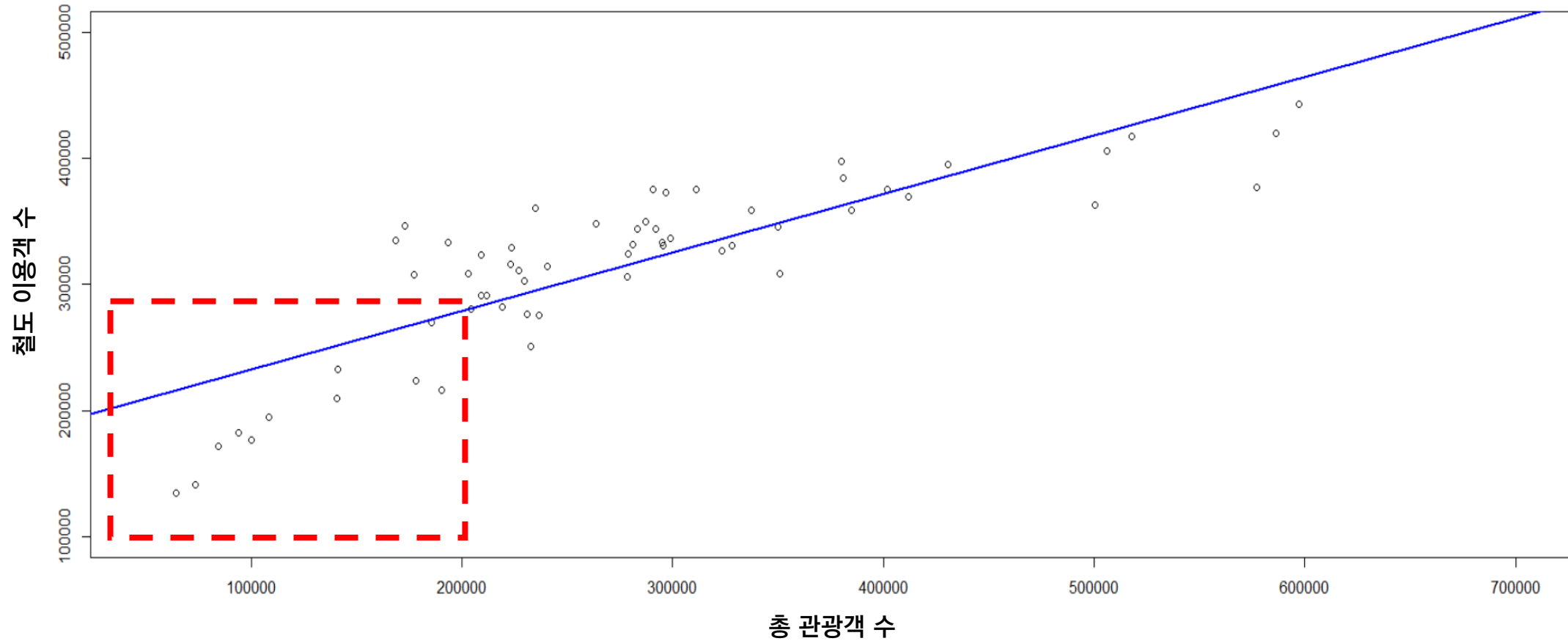
〈 철도 이용객 수 & 총 관광객 수 LS 회귀분석 〉



최소제곱법(Least Squared) 방법을 이용해 회귀분석한 결과를 회귀직선으로 표현

## 04 회귀 분석 - 조각별 선형회귀

〈 철도 이용객 수 & 총 관광객 수 LS 회귀분석 〉

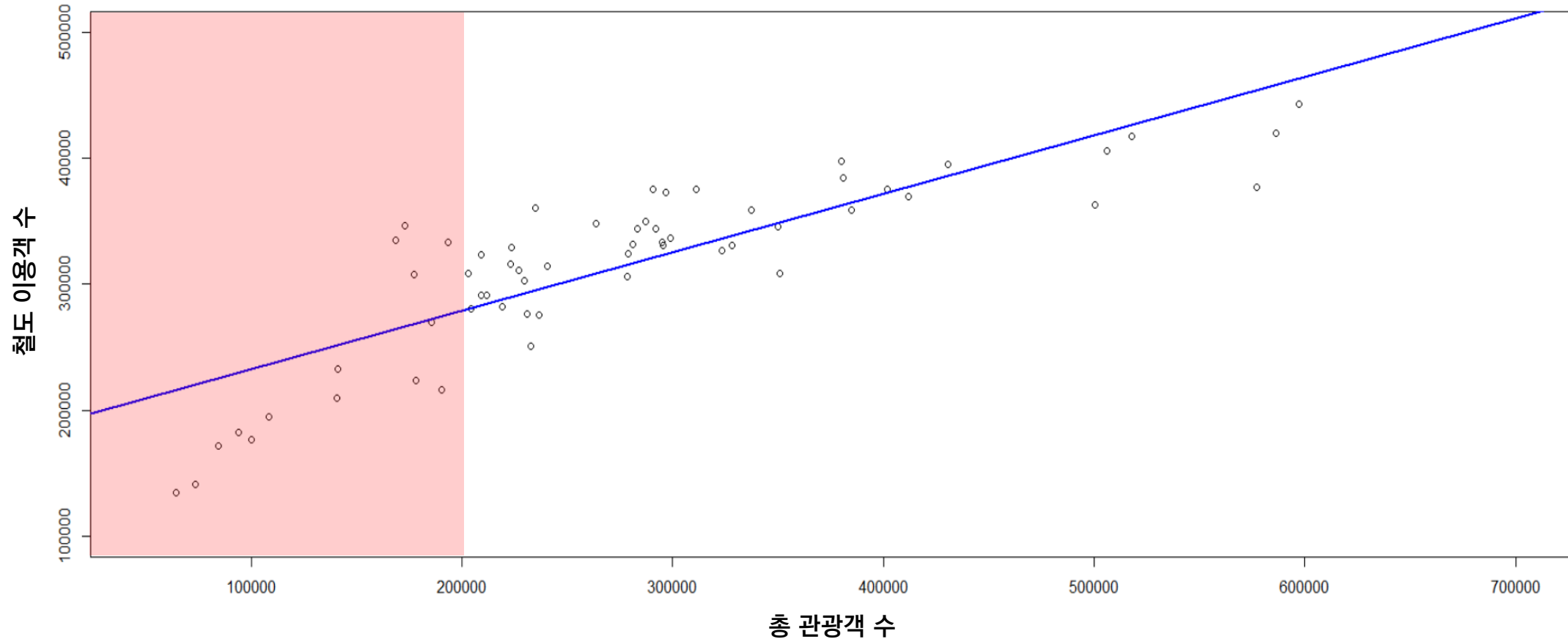


총 관광객 수가 200,000명보다 적은 경우,

최소 제곱법을 이용한 회귀 직선이 자료의 변동을 설명하지 못하는 부분을 식별

## 04 회귀 분석 - 조각별 선형회귀

〈 철도 이용객 수 & 총 관광객 수 LS 회귀분석 〉

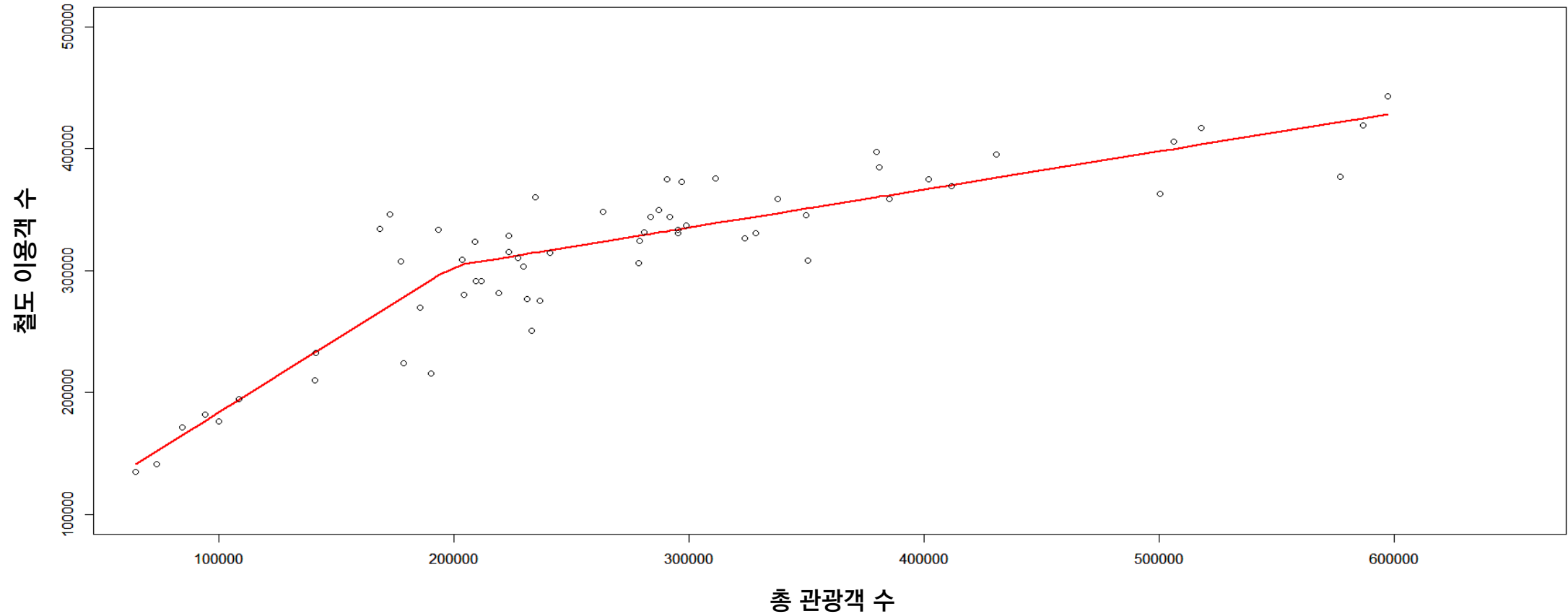


총 관광객 수가 200,000명보다 적은 경우의 해당 연도를 확인한 결과  
2017년은 1개, 2018년은 3개, 2019년은 1개, 그리고 2020년은 10개 식별



## 04 회귀 분석 - 조각별 선형회귀

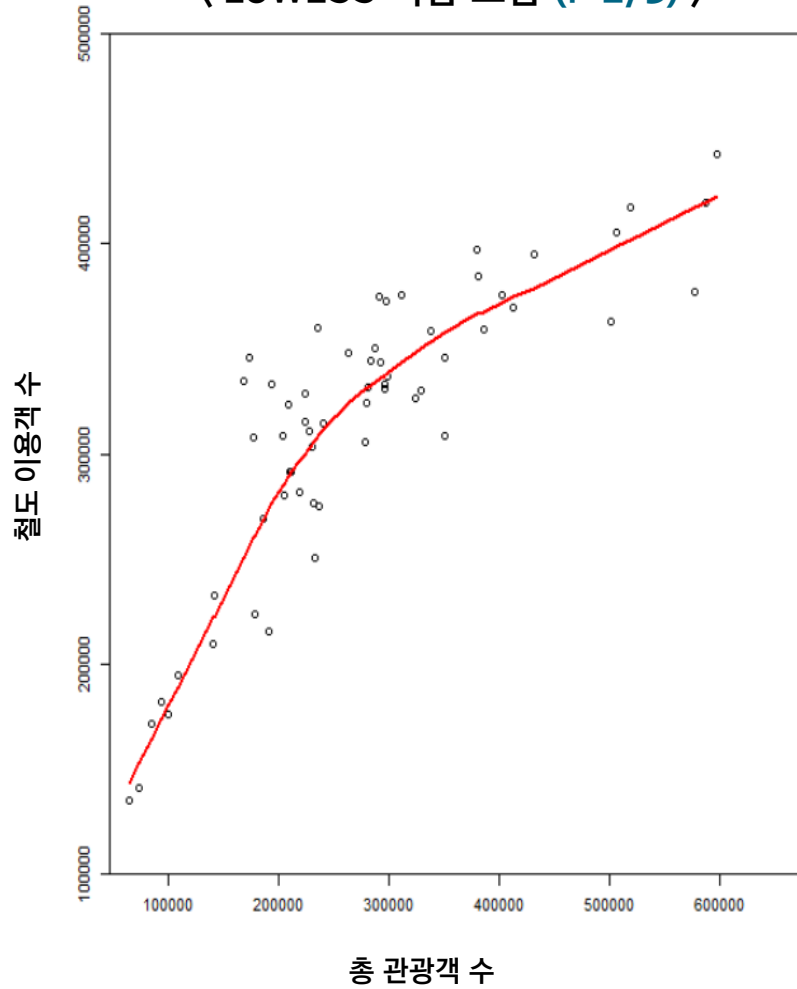
〈 철도 이용객 수 & 총 관광객 수 조각별 선형회귀 〉



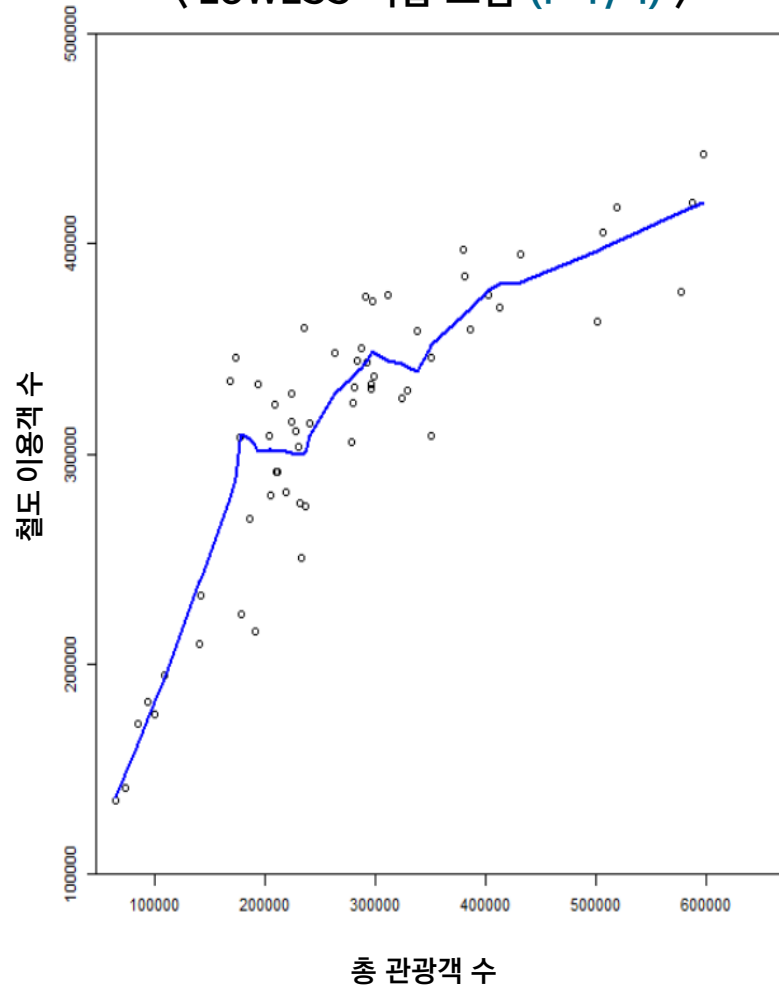
총 관광객 수 **200,000명**을 기준으로 조각별 선형회귀를 적합한 결과  
LS 방법을 이용했을 때보다 자료의 변동을 설명하기에 용이

## 04 회귀 분석 - 국소가중회귀곡선

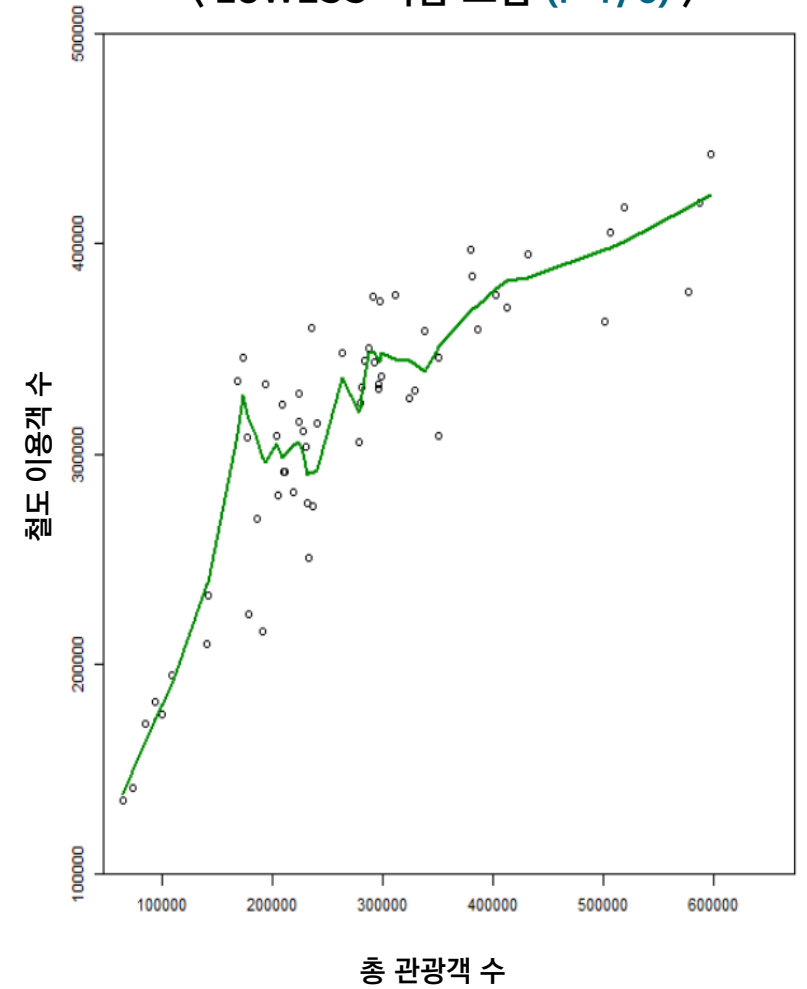
〈 LOWESS 적합 그림 ( $f=2/3$ ) 〉



〈 LOWESS 적합 그림 ( $f=1/4$ ) 〉

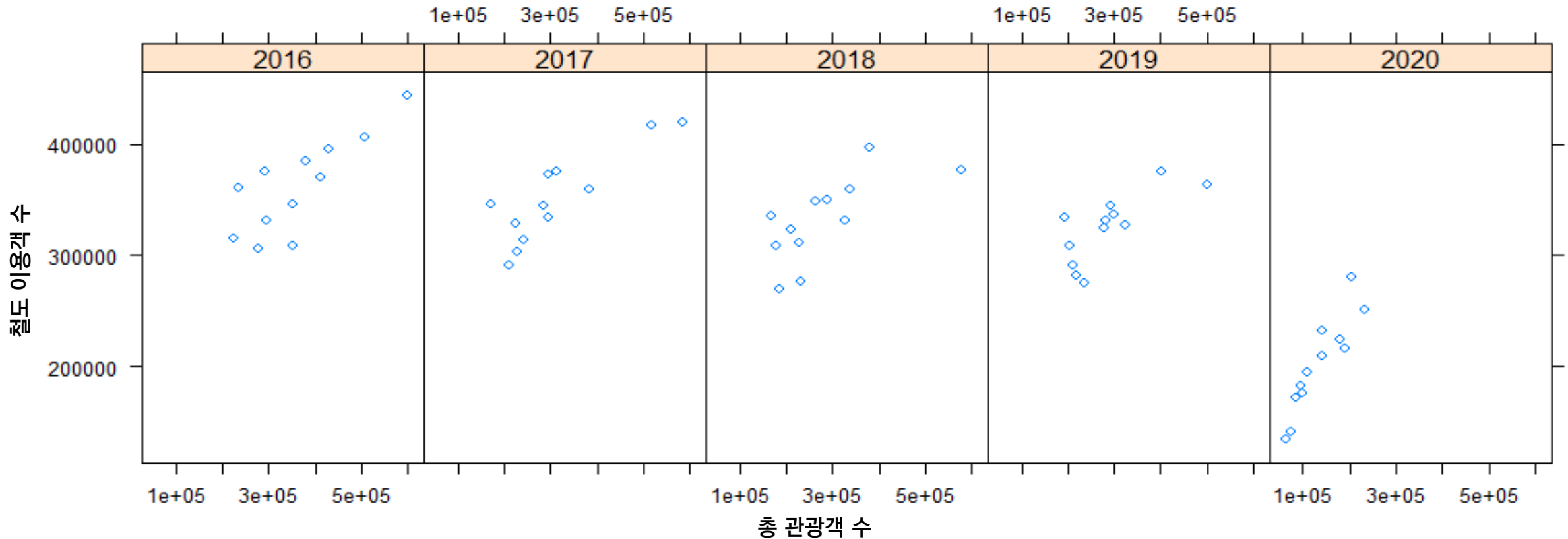


〈 LOWESS 적합 그림 ( $f=1/6$ ) 〉



## 04 회귀 분석 - 조건부 플롯

〈 철도 이용객 수 & 총 관광객 수 & year\_month 조건부 플롯 〉

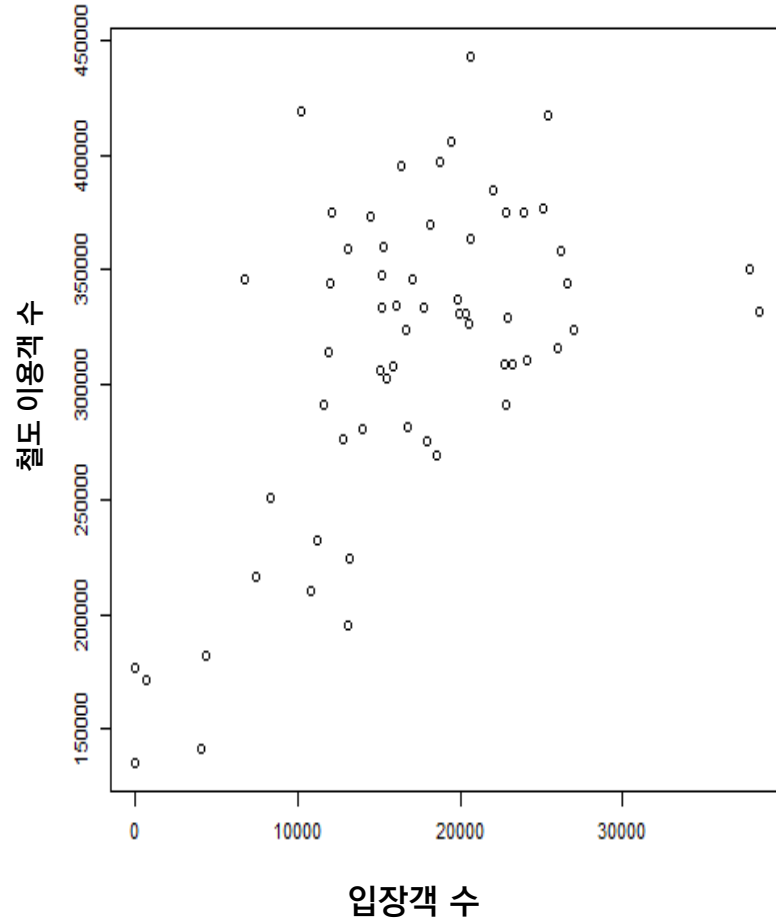


**2020년** 철도 이용객 수와 총 관광객 수의 분포가 다른 연도에 비해 작게 나타남

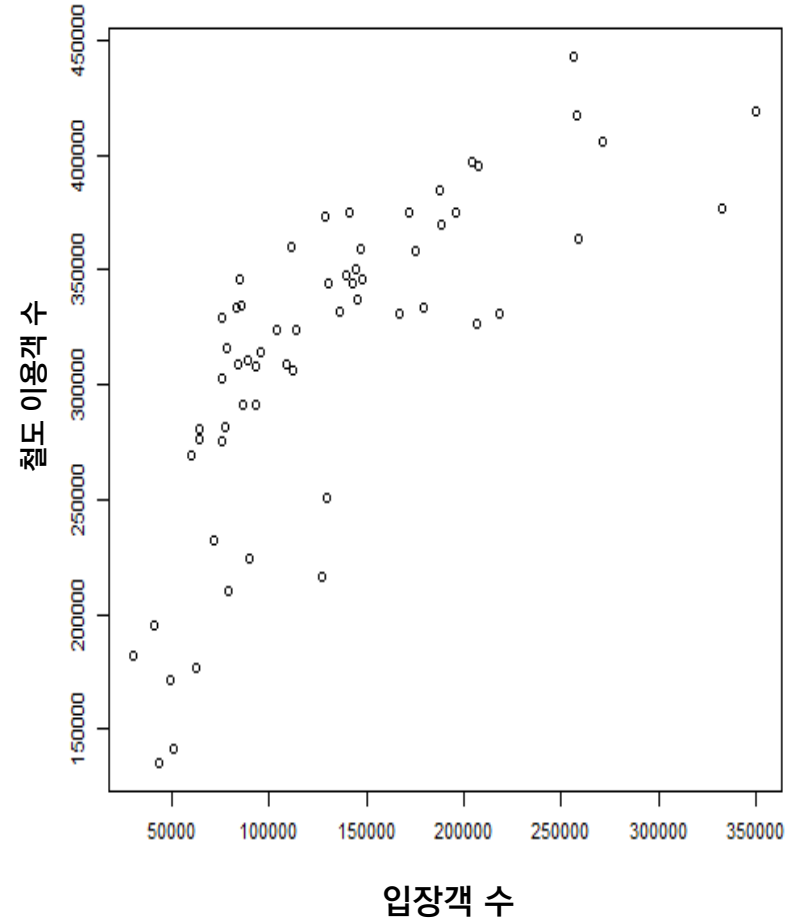
이는 **코로나 19 확산으로 인한 영향**으로 판단됨

## 04 회귀 분석 - 산점도 해석

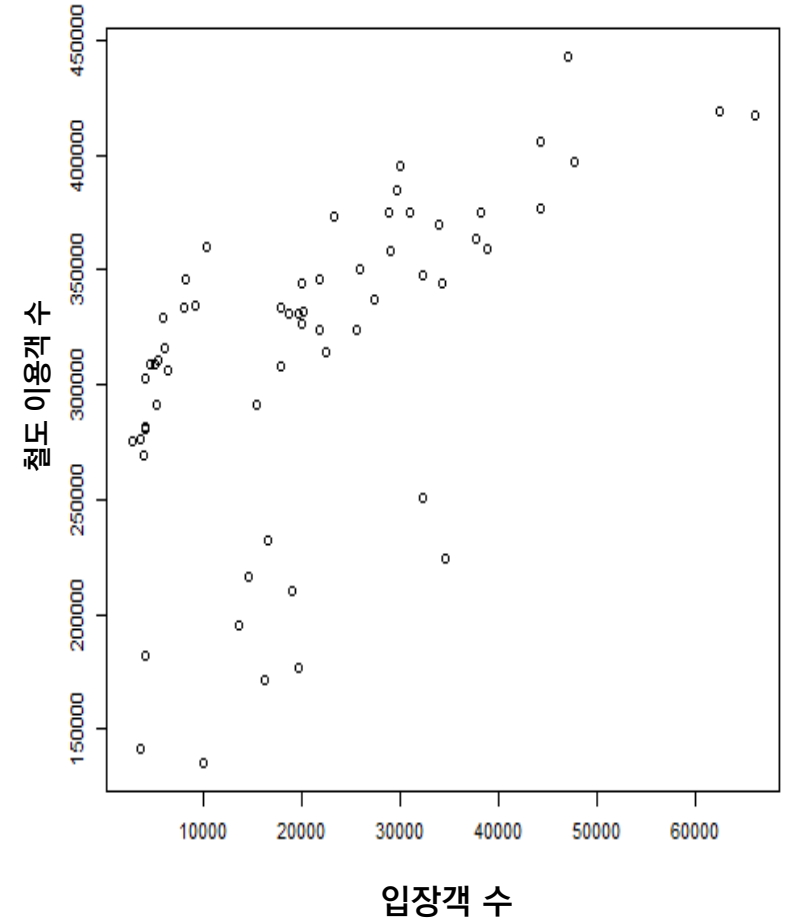
〈 국립춘천박물관 〉



〈 남이섬 유원지 〉



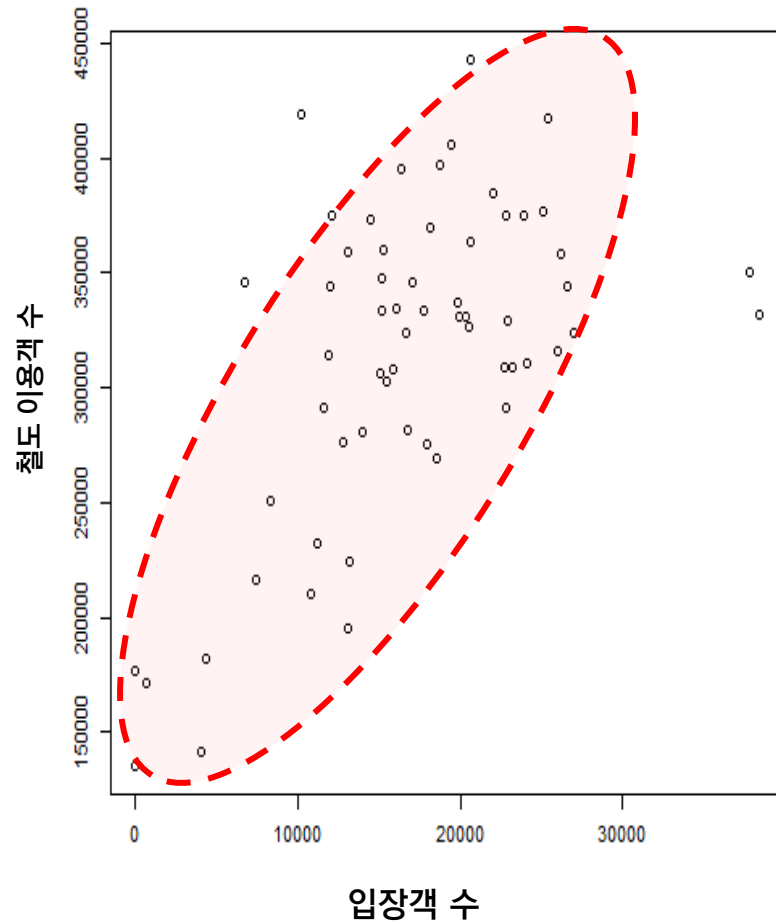
〈 제이드 가든 〉



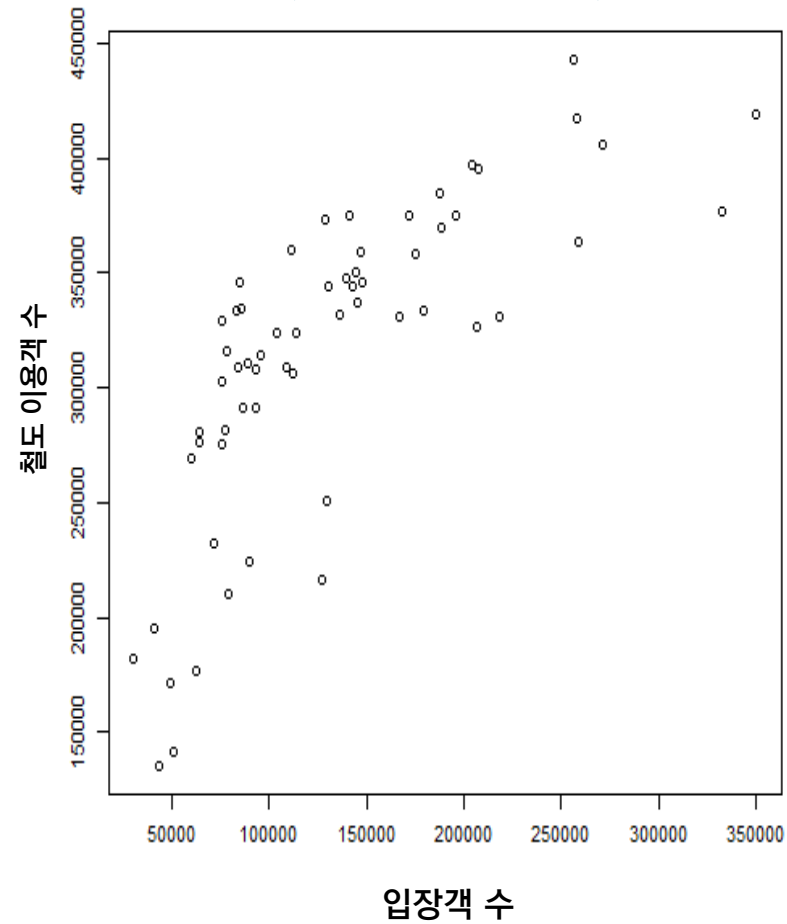
각 관광지별 입장객 수와 철도 이용객 수 비교

## 04 회귀 분석 - 산점도 해석

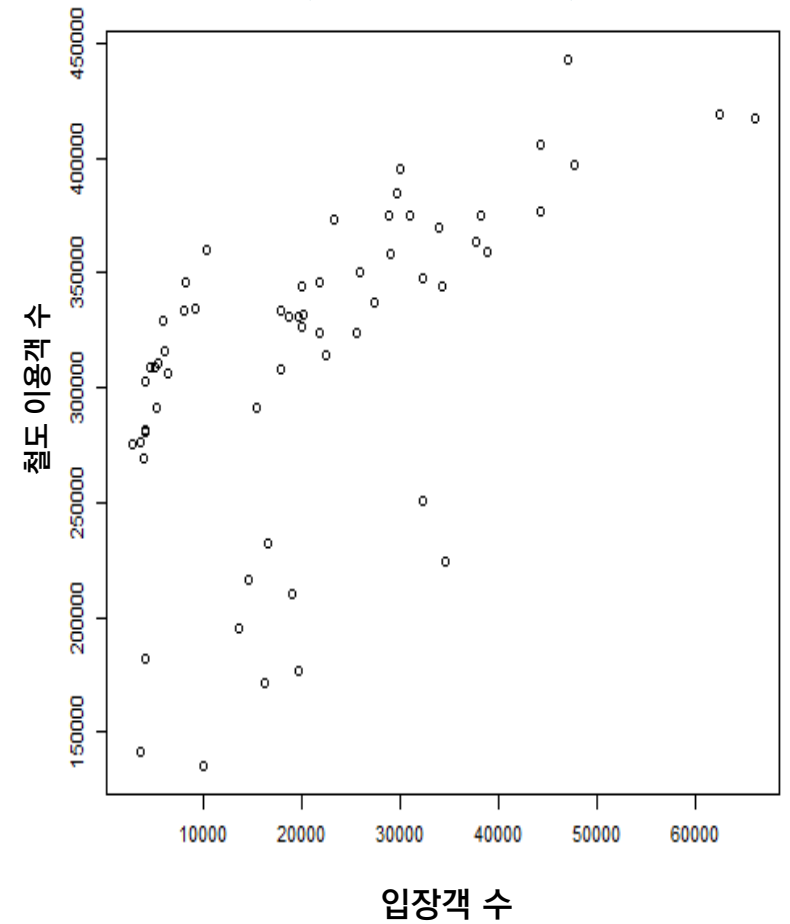
〈 국립춘천박물관 〉



〈 남이섬 유원지 〉

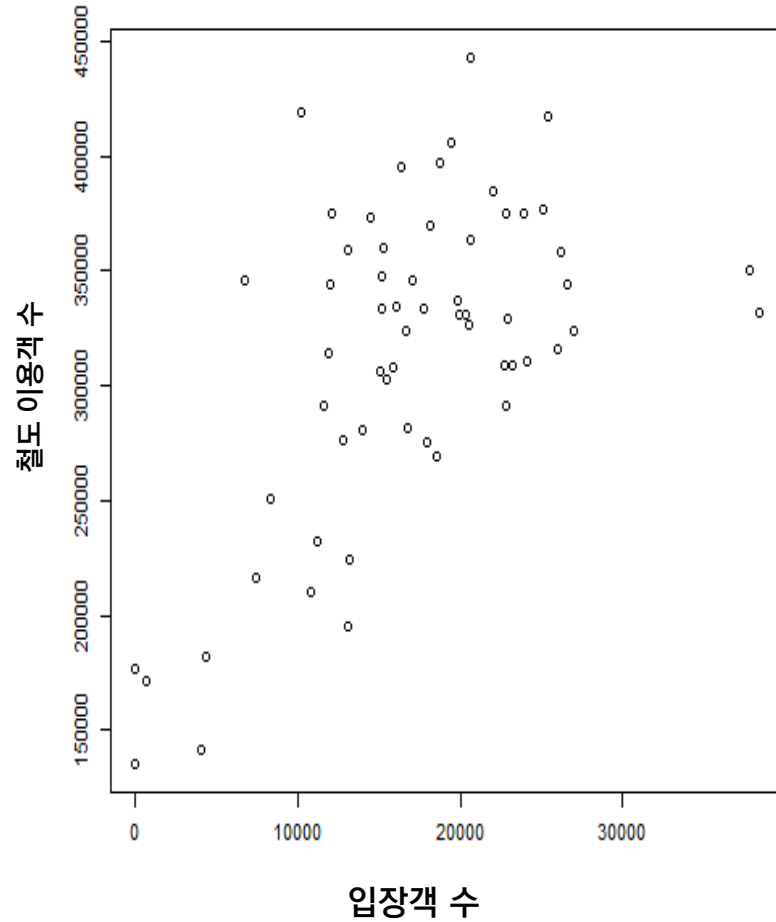


〈 제이드 가든 〉

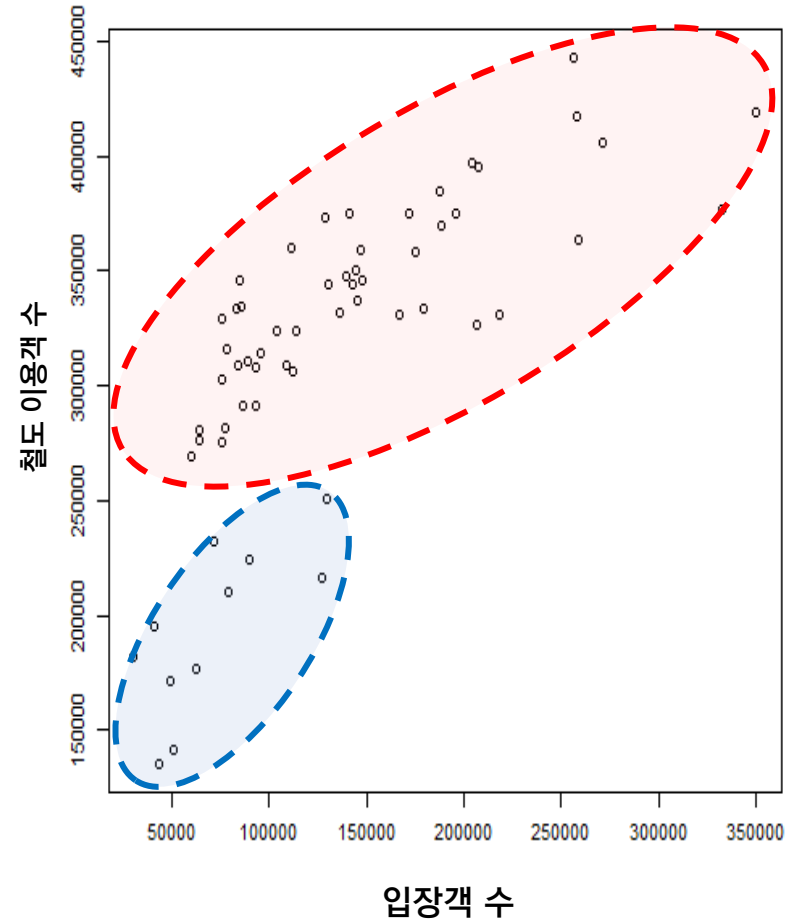


## 04 회귀 분석 - 산점도 해석

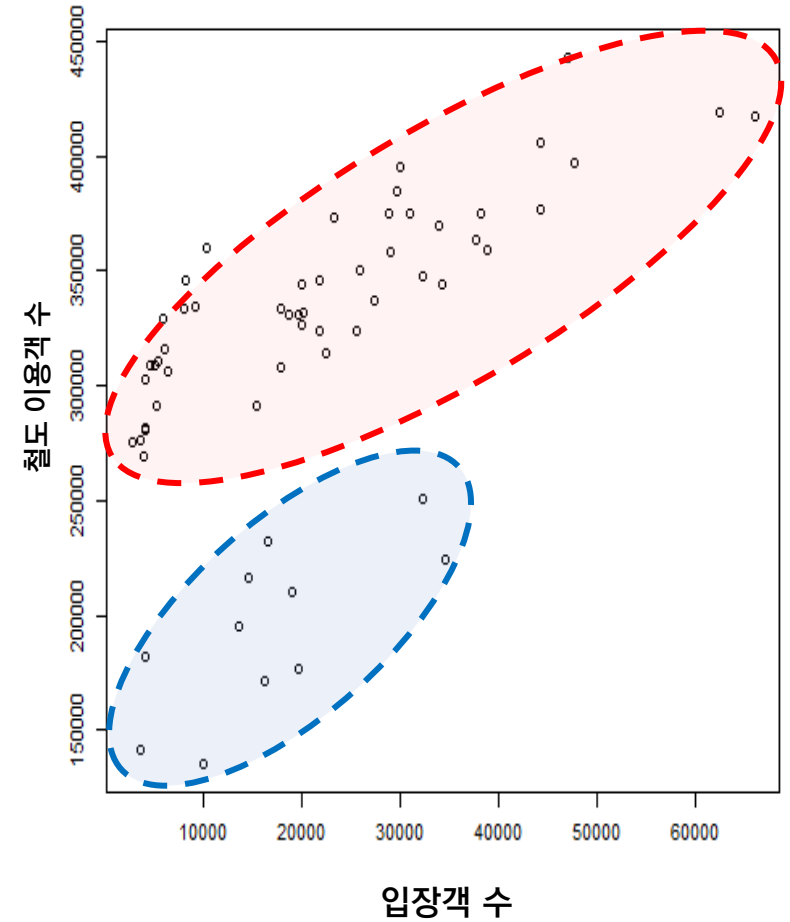
〈 국립춘천박물관 〉



〈 남이섬 유원지 〉

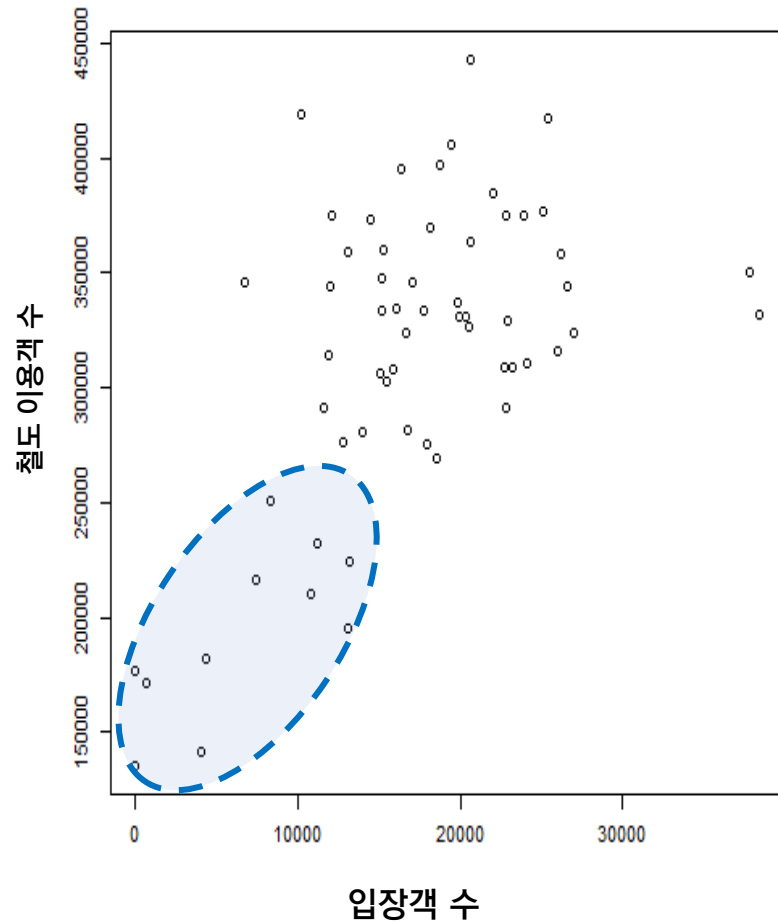


〈 제이드 가든 〉

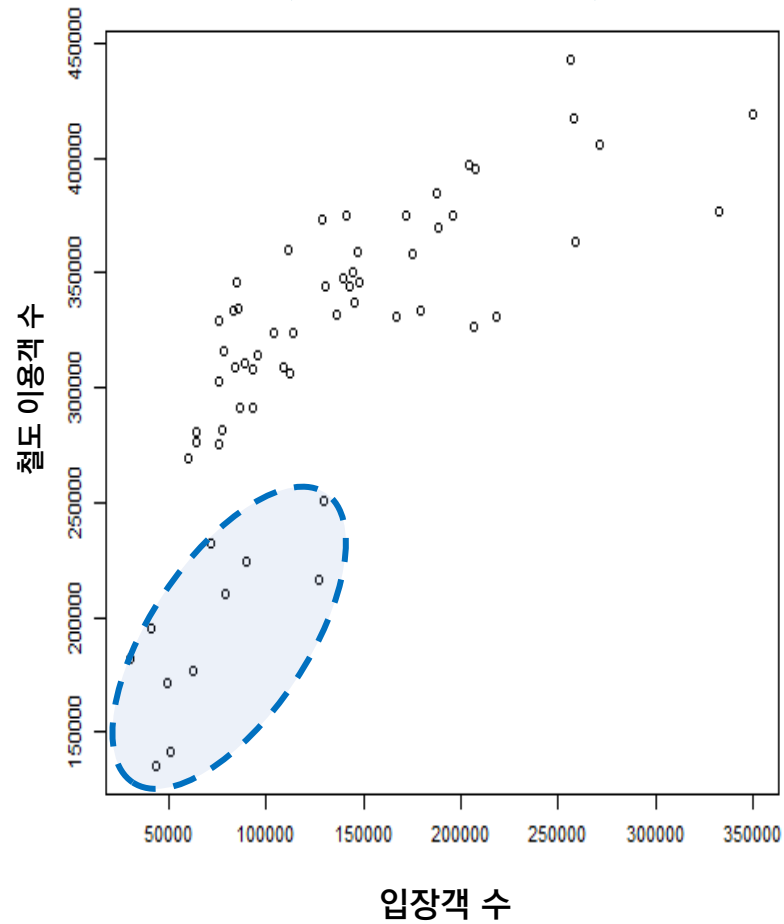


## 04 회귀 분석 - 산점도 해석

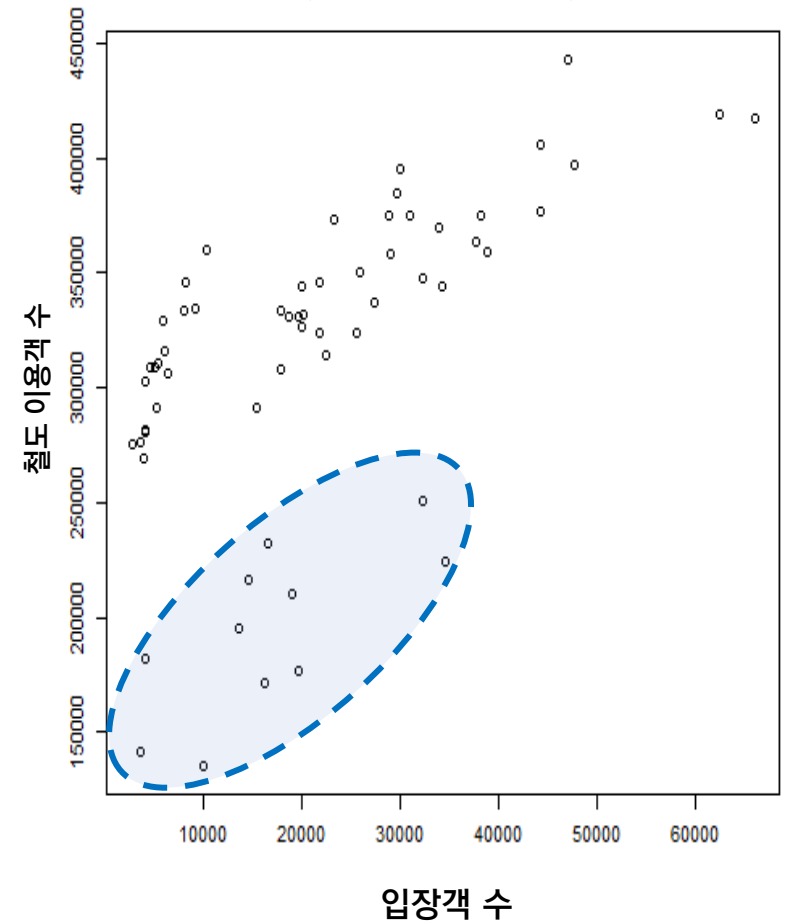
〈 국립춘천박물관 〉



〈 남이섬 유원지 〉

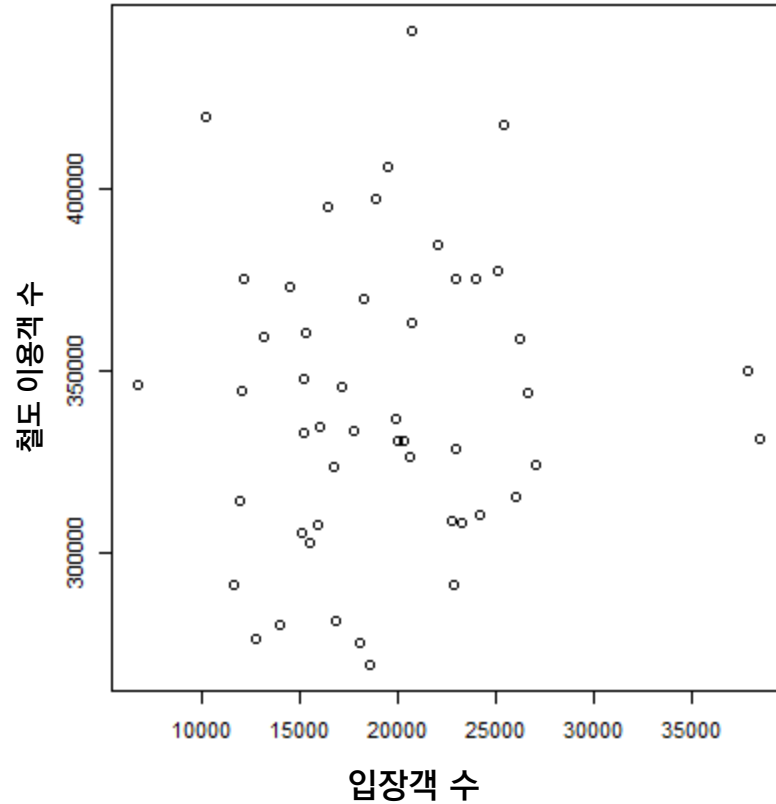


〈 제이드 가든 〉

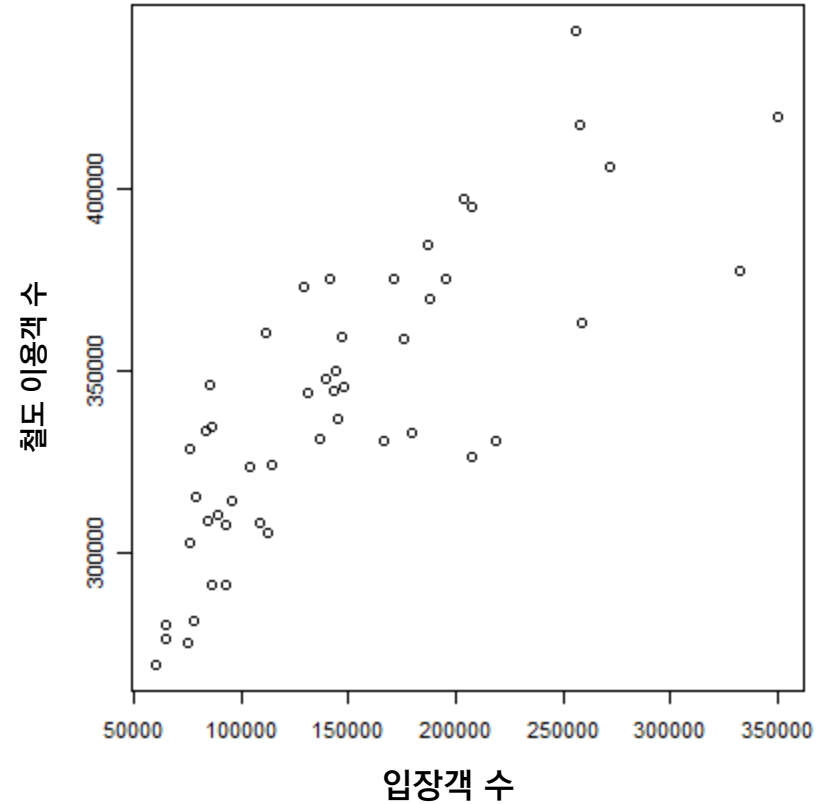


## 04 회귀 분석 - 산점도 해석

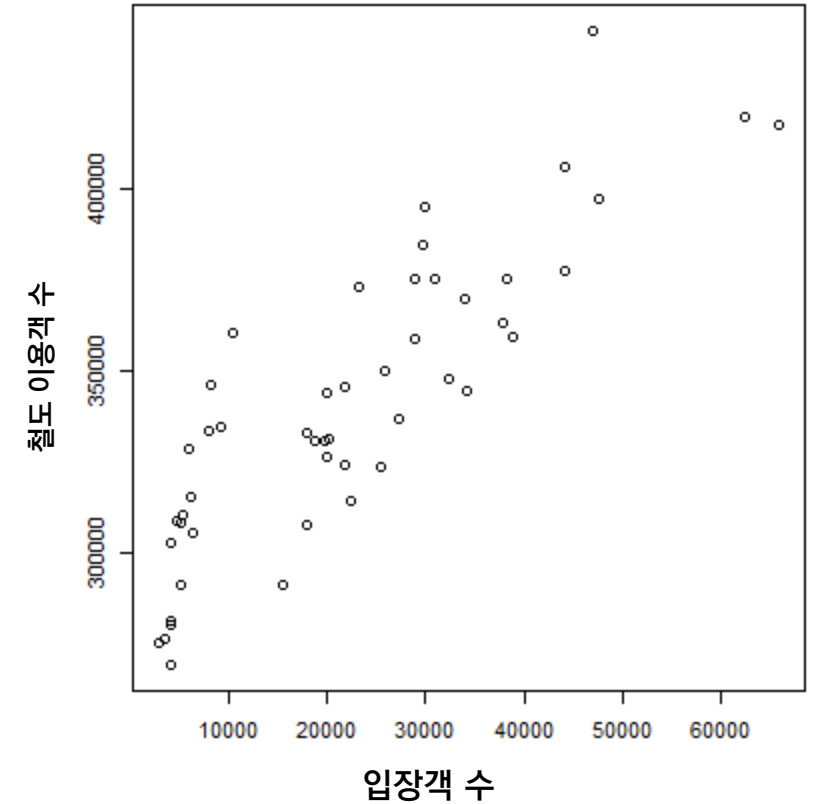
〈 국립춘천박물관 〉



〈 남이섬 유원지 〉



〈 제이드 가든 〉

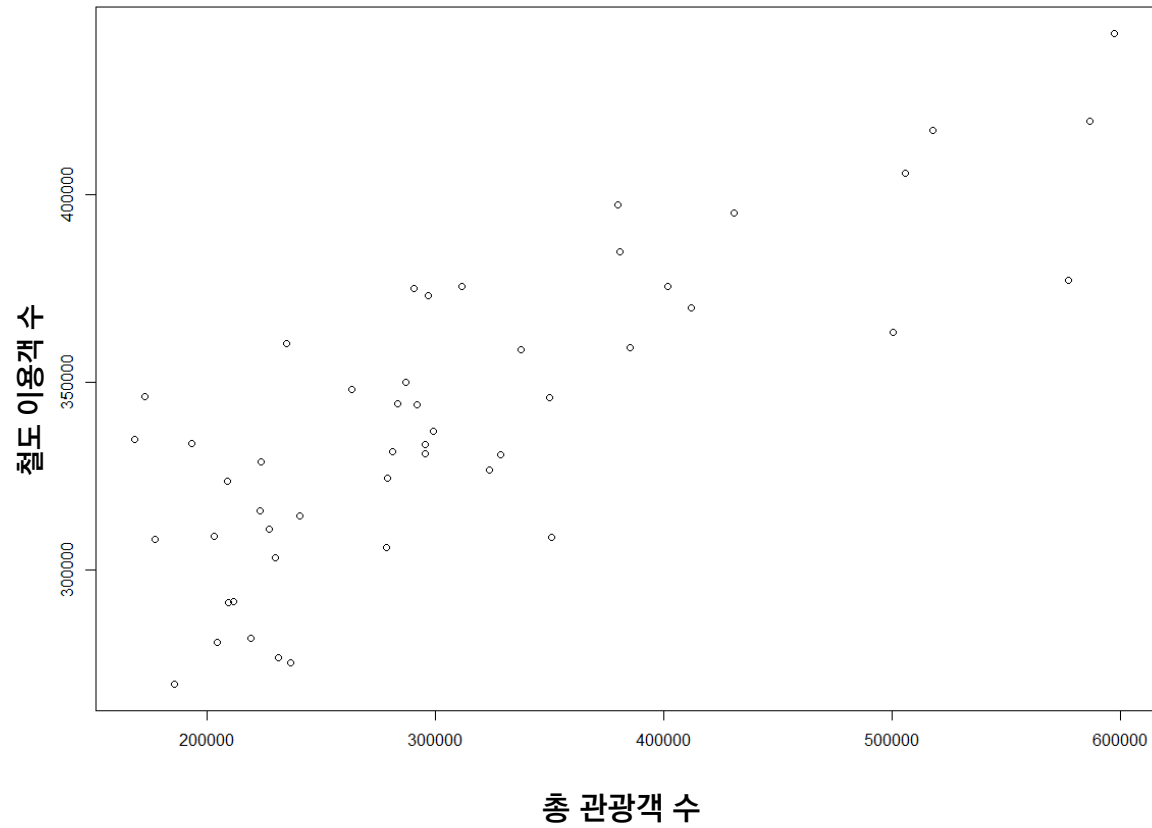


코로나 19 확산 이후 데이터를 제거한 뒤 각 관광지별 입장객 수와 철도 이용객 수 비교



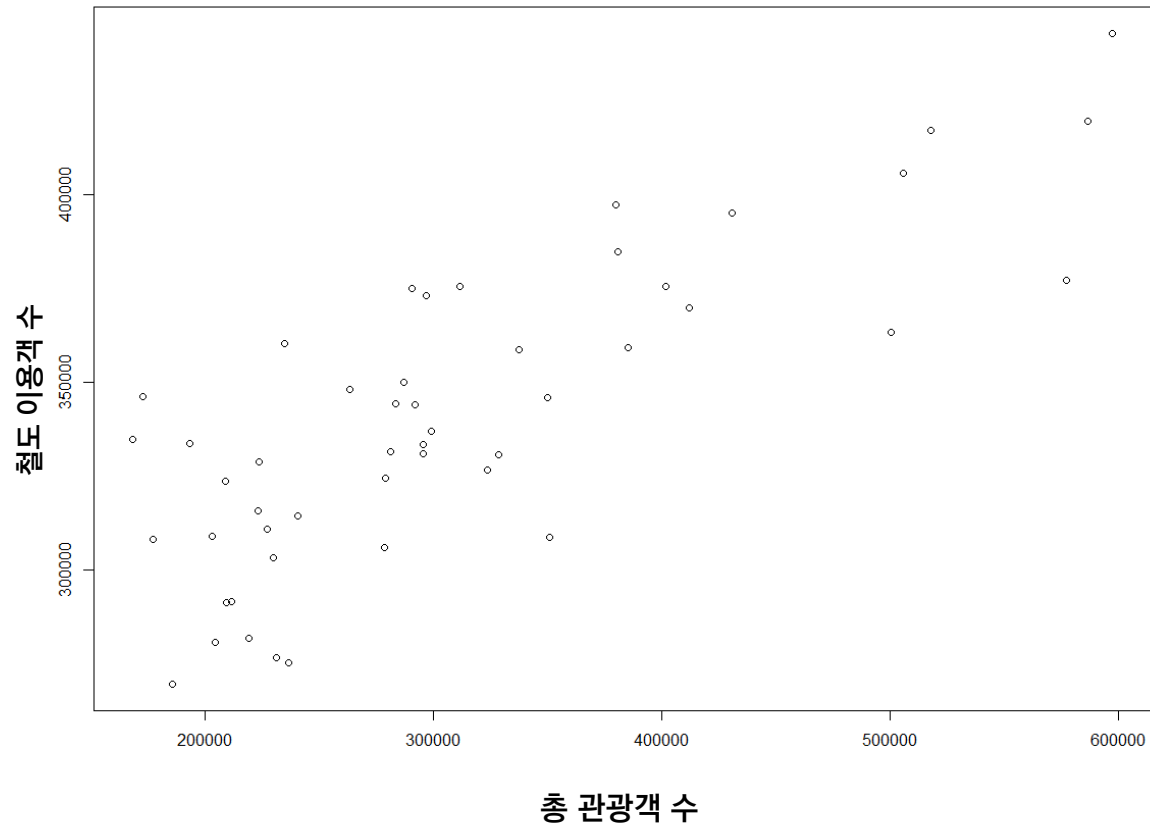
## 04 회귀 분석 - 단순선형회귀분석

〈 코로나 19 이전 총 관광객 수 & 철도 이용객 수 〉

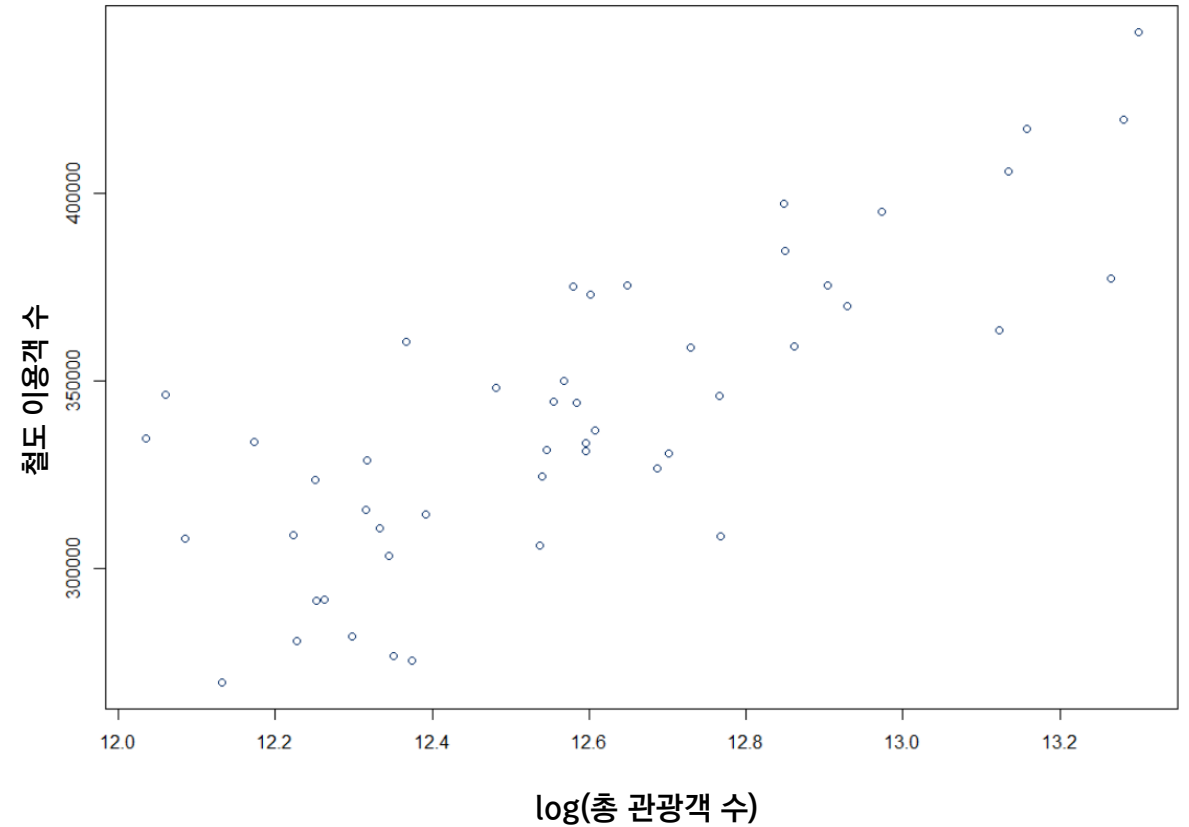


## 04 회귀 분석 - 단순선형회귀분석

〈 코로나 19 이전 총 관광객 수 & 철도 이용객 수 〉

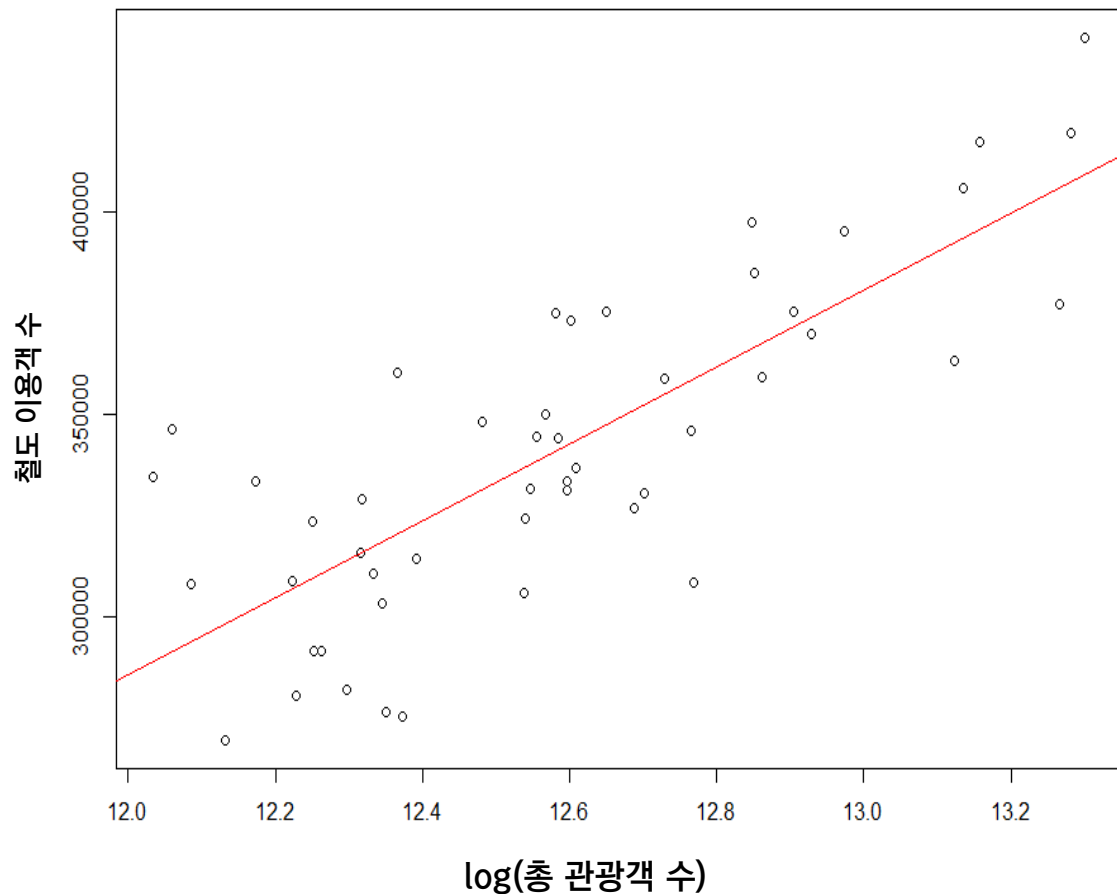


〈 코로나 19 이전 log(총 관광객 수) & 철도 이용객 수 〉



## 04 회귀 분석 - 단순선형회귀분석

〈 철도 이용객 수 ~ log(총 관광객 수) 단순 선형회귀 〉



Call:

```
lm(formula = sum_train ~ log(sum_travel), data = wd_before)
```

Residuals:

Min	1Q	Median	3Q	Max
-49994	-18296	-151	16595	54759

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-850914	135321	-6.288	9.85e-08 ***
log(sum_travel)	94725	10751	8.811	1.62e-11 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

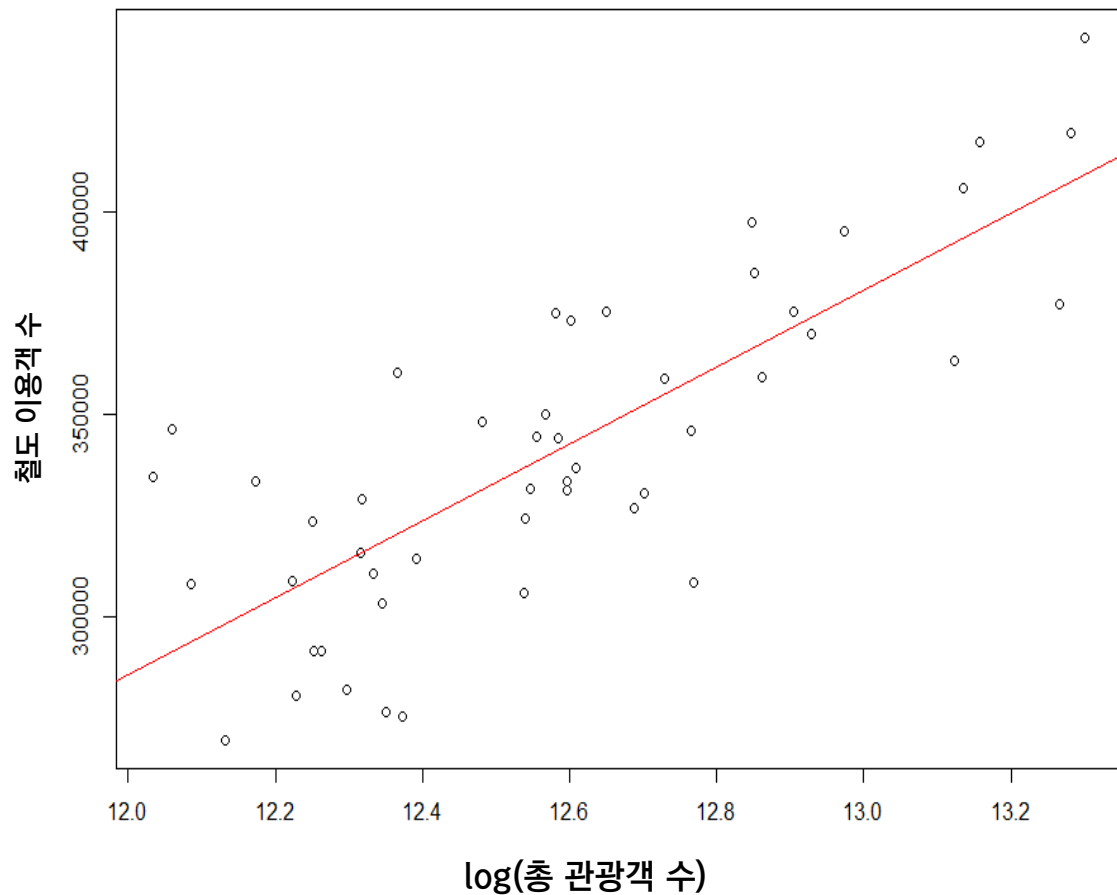
Residual standard error : 25030 on 47 degrees of freedom

Multiple R-squared : 0.6229, Adjusted R-squared: 0.6148

F-statistic : 77.63 on 1 and 47 DF, p-value: 1.617e-11

## 04 회귀 분석 - 단순선형회귀분석

〈 철도 이용객 수 ~ log(총 관광객 수) 단순 선형회귀 〉



Call:

```
lm(formula = sum_train ~ log(sum_travel), data = wd_before)
```

Residuals:

Min	1Q	Median	3Q	Max
-49994	-18296	-151	16595	54759

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-850914	135321	-6.288	9.85e-08 ***
log(sum_travel)	94725	10751	8.811	1.62e-11 ***

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error : 25030 on 47 degrees of freedom

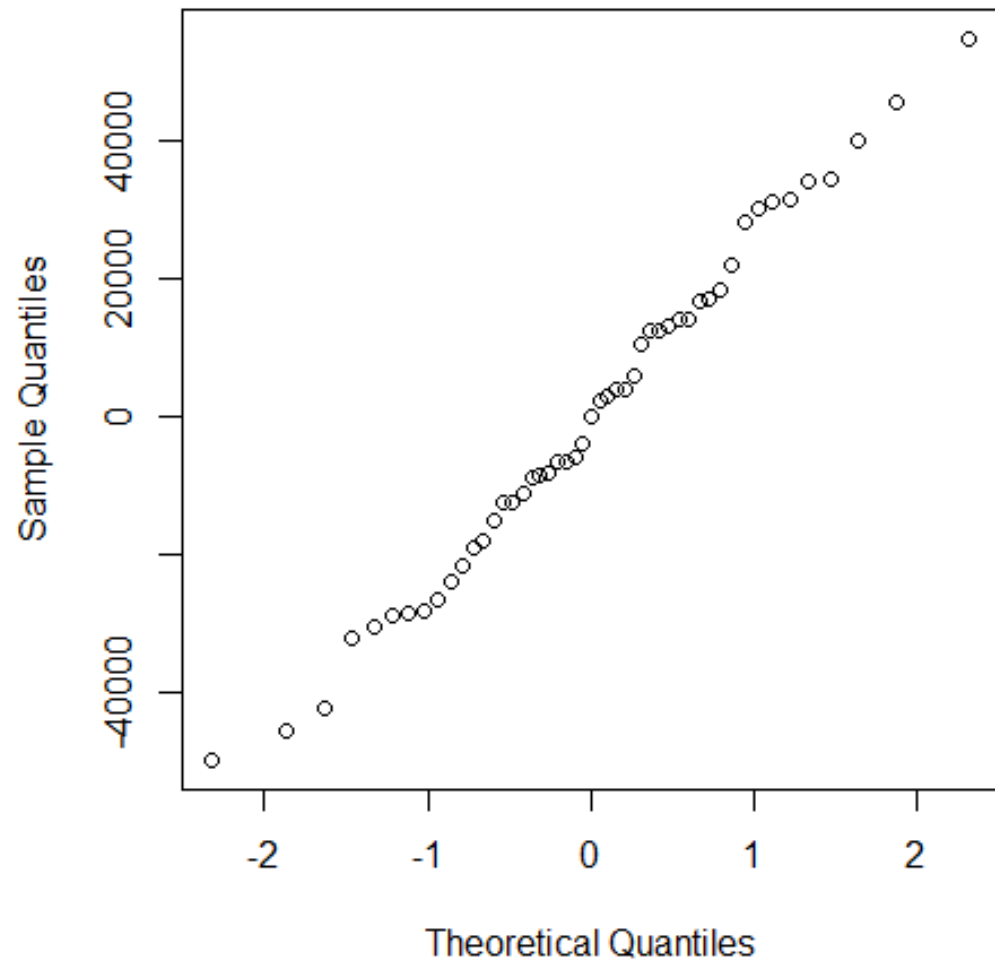
Multiple R-squared : 0.6229, Adjusted R-squared: 0.6148

F-statistic : 77.63 on 1 and 47 DF, p-value: 1.617e-11

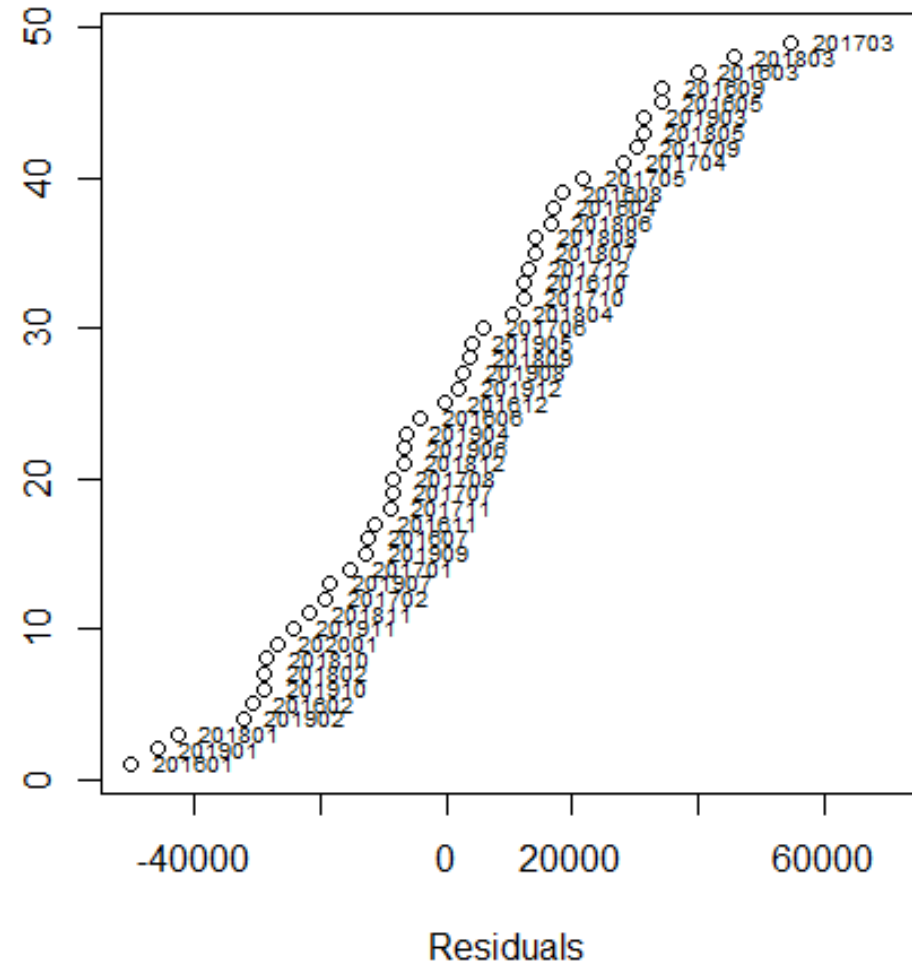
$$\hat{y} = -850914 + 94725 * \log(x)$$

## 04 회귀 분석 - 단순선형회귀분석

〈 Normal Q-Q plot 〉



〈 잔차그림 〉



## 04 회귀 분석

### 〈 LS방법을 사용한 다중선형회귀분석 〉

Call:

```
lm(formula = sum_train ~ ., data = wd_before[, c(-1, -2, -3, -16)])
```

Residuals:

Min	1Q	Median	3Q	Max
-28049	-11195	-5292	11614	38827

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	2.568e+05	2.187e+04	11.745	4.8e-14 ***
강촌레일파크	4.234e-01	3.287e-01	1.288	0.20575
국립춘천박물관	2.066e-01	4.549e-01	0.454	0.65231
김유정문학마을	4.182e-01	1.538e-01	2.718	0.00993 **
남이섬유원지	4.681e-02	8.683e-02	0.539	0.59304
막국수체험박물관	-5.643e+00	2.792e+00	-2.021	0.05056 .
물레길	1.120e+00	4.332e+00	0.258	0.79746
애니메이션.박물관	8.128e-02	1.927e-01	0.422	0.67554
제이드가든	1.038e+00	8.190e-01	1.268	0.21290
춘천에티오피아한국전참전기념관	1.618e+00	7.083e-01	2.285	0.02817 *
춘천지구전적기념관	2.748e+00	2.017e+00	1.362	0.18140
엘리시안강촌스키장	-1.919e-01	1.192e-01	-1.610	0.11596

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 16850 on 37 degrees of freedom

Multiple R-squared: 0.8654, Adjusted R-squared: 0.8254

F-statistic: 21.63 on 11 and 37 DF, p-value: 7.396e-13

## 04 회귀 분석

### 〈 LS방법을 사용한 다중선형회귀분석 〉

Call:

```
lm(formula = sum_train ~ ., data = wd_before[, c(-1, -2, -3, -16)])
```

Residuals:

Min	1Q	Median	3Q	Max
-28049	-11195	-5292	11614	38827

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	2.568e+05	2.187e+04	11.745	4.8e-14 ***
강촌레일파크	4.234e-01	3.287e-01	1.288	0.20575 .
국립춘천박물관	2.066e-01	4.549e-01	0.454	0.65231
김유정문학마을	4.182e-01	1.538e-01	2.718	0.00993 **
남이섬유원지	4.681e-02	8.683e-02	0.539	0.59304
막국수체험박물관	-5.643e+00	2.792e+00	-2.021	0.05056 .
물레길	1.120e+00	4.332e+00	0.258	0.79746
애니메이션.박물관	8.128e-02	1.927e-01	0.422	0.67554
제이드가든	1.038e+00	8.190e-01	1.268	0.21290
춘천에티오피아한국전참전기념관	1.618e+00	7.083e-01	2.285	0.02817 *
춘천지구전적기념관	2.748e+00	2.017e+00	1.362	0.18140
엘리시안강촌스키장	-1.919e-01	1.192e-01	-1.610	0.11596

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

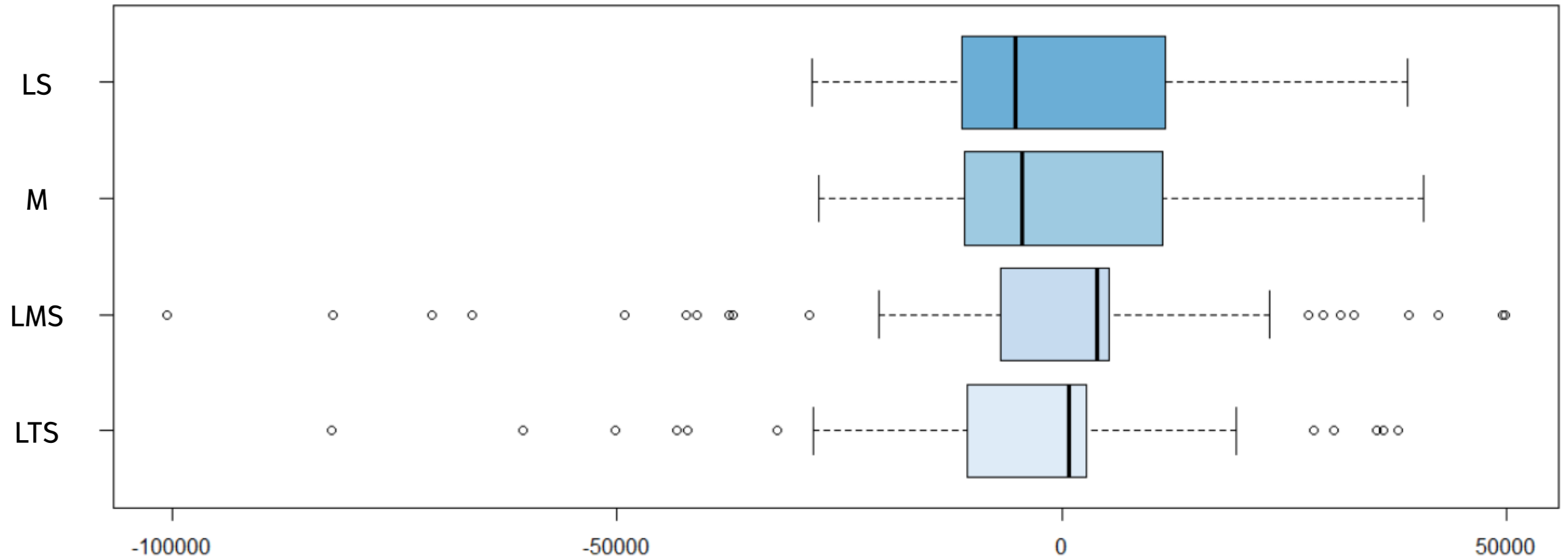
Residual standard error: 16850 on 37 degrees of freedom

Multiple R-squared: 0.8654, Adjusted R-squared: 0.8254

F-statistic: 21.63 on 11 and 37 DF, p-value: 7.396e-13

## 04 회귀 분석

〈 회귀분석 방법별 잔차분포 〉

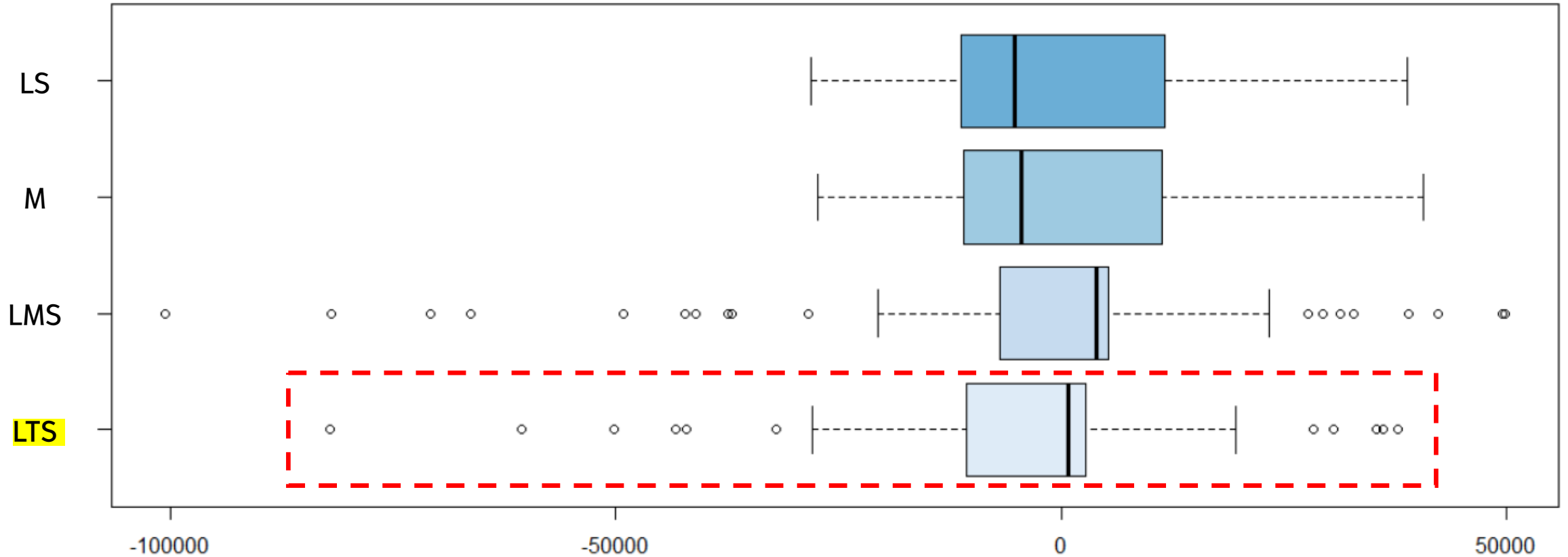


회귀분석 방법에 따른 잔차분포를 비교한 결과 **분포에 영향을 주는 이상치**가 있다고 판단



## 04 회귀 분석

〈 회귀분석 방법별 잔차분포 〉



LTS 회귀분석 방법에서 나타난 잔차를 이용해 이상치 제거 후 다중선형회귀분석 실시

## 04 회귀 분석

### < 모든 독립변수에 대한 다중선행회귀분석 >

Call:

```
lm(formula = sum_train ~ .,  
    data = wd_before[(lm_tr.lts$residuals) < 25000 & -30000 < (lm_tr.lts$residuals),  
                      c(-1, -2, -3, -16)])
```

Residuals:

Min	1Q	Median	3Q	Max
-17129	-5249	-859	7531	15748

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	2.148e+05	1.594e+04	13.476	3.08e-13 ***
강촌레일파크	5.472e-01	2.703e-01	2.024	0.053338 .
국립춘천박물관	2.670e-02	3.206e-01	0.083	0.934255
김유정문학마을	2.285e-01	1.055e-01	2.166	0.039652 *
남이섬유원지	7.214e-02	6.498e-02	1.110	0.277122
막국수체험박물관	-1.619e+01	2.433e+00	-6.657	4.63e-07 ***
물레길	-9.938e-02	2.937e+00	-0.034	0.973264
애니메이션.박물관	9.558e-01	2.105e-01	4.540	0.000113 ***
제이드가든	7.245e-01	6.240e-01	1.161	0.256226
춘천에티오피아한국전참전기념관	3.244e+00	5.263e-01	6.163	1.62e-06 ***
춘천지구전적기념관	6.038e+00	1.509e+00	4.001	0.000466 ***
엘리시안강촌스키장	-1.657e-01	8.399e-02	-1.973	0.059175 .

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10160 on 26 degrees of freedom

Multiple R-squared: 0.9568, Adjusted R-squared: 0.9385

F-statistic: 52.34 on 11 and 26 DF, p-value: 6.997e-15

## 04 회귀 분석

### < 모든 독립변수에 대한 다중선행회귀분석 >

Call:

```
lm(formula = sum_train ~ .,  
    data = wd_before[(lm_tr.lts$residuals) < 25000 & -30000 < (lm_tr.lts$residuals),  
    c(-1, -2, -3, -16)])
```

Residuals:

Min	1Q	Median	3Q	Max
-17129	-5249	-859	7531	15748

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	2.148e+05	1.594e+04	13.476	3.08e-13 ***
강촌레일파크	5.472e-01	2.703e-01	2.024	0.053338 .
국립춘천박물관	2.670e-02	3.206e-01	0.083	0.934255
김유정문학마을	2.285e-01	1.055e-01	2.166	0.039652 *
남이섬유원지	7.214e-02	6.498e-02	1.110	0.277122
막국수체험박물관	-1.619e+01	2.433e+00	-6.657	4.63e-07 ***
물레길	-9.938e-02	2.937e+00	-0.034	0.973264
애니메이션.박물관	9.558e-01	2.105e-01	4.540	0.000113 ***
제이드가든	7.245e-01	6.240e-01	1.161	0.256226
춘천에티오피아한국전참전기념관	3.244e+00	5.263e-01	6.163	1.62e-06 ***
춘천지구전적기념관	6.038e+00	1.509e+00	4.001	0.000466 ***
엘리시안강촌스키장	-1.657e-01	8.399e-02	-1.973	0.059175 .

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10160 on 26 degrees of freedom

Multiple R-squared: 0.9568, Adjusted R-squared: 0.9385

F-statistic: 52.34 on 11 and 26 DF, p-value: 6.997e-15

## 04 회귀 분석

### < 모든 독립변수에 대한 다중선형회귀분석 >

Call:

```
lm(formula = sum_train ~ .,  
    data = wd_before[(lm_tr.lts$residuals) < 25000 & -30000 < (lm_tr.lts$residuals),  
                      c(-1, -2, -3, -16)])
```

Residuals:

Min	1Q	Median	3Q	Max
-17129	-5249	-859	7531	15748

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	2.148e+05	1.594e+04	13.476	3.08e-13 ***
강촌레일파크	5.472e-01	2.703e-01	2.024	0.053338 .
국립춘천박물관	2.670e-02	3.206e-01	0.083	0.934255
김유정문학마을	2.285e-01	1.055e-01	2.166	0.039652 *
남이섬유원지	7.214e-02	6.498e-02	1.110	0.277122
막국수체험박물관	-1.619e+01	2.433e+00	-6.657	4.63e-07 ***
<b>물레길</b>	<b>-9.938e-02</b>	<b>2.937e+00</b>	<b>-0.034</b>	<b>0.973264</b>
애니메이션.박물관	9.558e-01	2.105e-01	4.540	0.000113 ***
제이드가든	7.245e-01	6.240e-01	1.161	0.256226
춘천에티오피아한국전참전기념관	3.244e+00	5.263e-01	6.163	1.62e-06 ***
춘천지구전적기념관	6.038e+00	1.509e+00	4.001	0.000466 ***
엘리시안강촌스키장	-1.657e-01	8.399e-02	-1.973	0.059175 .

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10160 on 26 degrees of freedom

Multiple R-squared: 0.9568, Adjusted R-squared: 0.9385

F-statistic: 52.34 on 11 and 26 DF, p-value: 6.997e-15

## 04 회귀 분석

### 〈 ‘물레길’ 을 제외한 독립변수에 대한 다중선형회귀분석 〉

Call:

```
lm(formula = sum_train ~ .,  
    data = wd_before[(lm_tr.lts$residuals) < a & -b < (lm_tr.lts$residuals), c(-1, -2, -3, -9, -16)])
```

Residuals:

Min	1Q	Median	3Q	Max
-17174	-5331	-844	7558	15726

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	2.147e+05	1.539e+04	13.949	7.36e-14	***
강촌레일파크	5.478e-01	2.645e-01	2.071	0.048030	*
국립춘천박물관	3.075e-02	2.919e-01	0.105	0.916873	
김유정문학마을	2.276e-01	9.995e-02	2.277	0.030952	*
남이섬유원지	7.320e-02	5.580e-02	1.312	0.200591	
막국수체험박물관	-1.619e+01	2.381e+00	-6.800	2.65e-07	***
애니메이션.박물관	9.553e-01	2.061e-01	4.636	8.09e-05	***
제이드가든	7.067e-01	3.296e-01	2.144	0.041174	*
춘천에티오피아한국전참전기념관	3.248e+00	4.982e-01	6.520	5.44e-07	***
춘천지구전적기념관	6.033e+00	1.473e+00	4.095	0.000344	***
엘리시안강촌스키장	-1.657e-01	8.240e-02	-2.011	0.054451	.

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 9973 on 27 degrees of freedom

Multiple R-squared: 0.9568, Adjusted R-squared: 0.9408

F-statistic: 59.79 on 10 and 27 DF, p-value: 8.731e-16

## 04 회귀 분석

### < ‘물레길’ 을 제외한 독립변수에 대한 다중선형회귀분석 >

Call:

```
lm(formula = sum_train ~ .,  
    data = wd_before[(lm_tr.lts$residuals) < a & -b < (lm_tr.lts$residuals), c(-1, -2, -3, -9, -16)])
```

Residuals:

Min	1Q	Median	3Q	Max
-17174	-5331	-844	7558	15726

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	2.147e+05	1.539e+04	13.949	7.36e-14 ***
강촌레일파크	5.478e-01	2.645e-01	2.071	0.048030 *
국립춘천박물관	3.075e-02	2.919e-01	0.105	0.916873
김유정문학마을	2.276e-01	9.995e-02	2.277	0.030952 *
남이섬유원지	7.320e-02	5.580e-02	1.312	0.200591
막국수체험박물관	-1.619e+01	2.381e+00	-6.800	2.65e-07 ***
애니메이션.박물관	9.553e-01	2.061e-01	4.636	8.09e-05 ***
제이드가든	7.067e-01	3.296e-01	2.144	0.041174 *
춘천에티오피아한국전참전기념관	3.248e+00	4.982e-01	6.520	5.44e-07 ***
춘천지구전적기념관	6.033e+00	1.473e+00	4.095	0.000344 ***
엘리시안강촌스키장	-1.657e-01	8.240e-02	-2.011	0.054451 .

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 9973 on 27 degrees of freedom

Multiple R-squared: 0.9568, Adjusted R-squared: 0.9408

F-statistic: 59.79 on 10 and 27 DF, p-value: 8.731e-16

## 04 회귀 분석

### 〈 최종 다중선형회귀모형 〉

Call:  
lm(formula = sum\_train ~ .,  
data = wd\_before[(lm\_tr.lts\$residuals) < a & -b < (lm\_tr.lts\$residuals), c(-1, -2, -3, -5, -9, -16)])

Residuals:

Min	1Q	Median	3Q	Max
-17156	-5389	-869	7545	15720

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	2.148e+05	1.512e+04	14.208	2.51e-14 ***
강촌레일파크	5.420e-01	2.540e-01	2.134	0.0417 *
김유정문학마을	2.273e-01	9.815e-02	2.316	0.0281 *
남이섬유원지	7.309e-02	5.479e-02	1.334	0.1930
막국수체험박물관	-1.617e+01	2.330e+00	-6.938	1.53e-07 ***
애니메이션 박물관	9.555e-01	2.024e-01	4.721	5.94e-05 ***
제이드가든	7.120e-01	3.198e-01	2.227	0.0342 *
춘천에티오피아한국전참전기념관	3.237e+00	4.778e-01	6.775	2.34e-07 ***
춘천지구전적기념관	6.096e+00	1.325e+00	4.601	8.24e-05 ***
엘리시안강촌스키장	-1.642e-01	7.981e-02	-2.058	0.0490 *

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9796 on 28 degrees of freedom  
Multiple R-squared: 0.9568, Adjusted R-squared: 0.9429  
F-statistic: 68.86 on 9 and 28 DF, p-value: < 2.2e-16

Y : 열차 이용객 수 (Sum\_train)  
X<sub>1</sub> : 강촌레일파크  
X<sub>2</sub> : 국립춘천 박물관  
X<sub>3</sub> : 김유정문학마을  
X<sub>4</sub> : 남이섬 유원지  
X<sub>5</sub> : 막국수체험 박물관  
X<sub>6</sub> : 물레길  
X<sub>7</sub> : 애니메이션 박물관  
X<sub>8</sub> : 제이드가든  
X<sub>9</sub> : 춘천 에티오피아 한국전 참전기념관  
X<sub>10</sub> : 춘천지구전적기념관  
X<sub>11</sub> : 엘리시안강촌 스키장

## 04 회귀 분석

### 〈 최종 다중선형회귀모형 〉

Call:  
lm(formula = sum\_train ~ .,  
data = wd\_before[(lm\_tr.lts\$residuals) < a & -b < (lm\_tr.lts\$residuals), c(-1, -2, -3, -5, -9, -16)])

Residuals:

Min	1Q	Median	3Q	Max
-17156	-5389	-869	7545	15720

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	2.148e+05	1.512e+04	14.208	2.51e-14 ***
강촌레일파크	5.420e-01	2.540e-01	2.134	0.0417 *
김유정문학마을	2.273e-01	9.815e-02	2.316	0.0281 *
남이섬유원지	7.309e-02	5.479e-02	1.334	0.1930
막국수체험박물관	-1.617e+01	2.330e+00	-6.938	1.53e-07 ***
애니메이션 박물관	9.555e-01	2.024e-01	4.721	5.94e-05 ***
제이드가든	7.120e-01	3.198e-01	2.227	0.0342 *
춘천에티오피아한국전참전기념관	3.237e+00	4.778e-01	6.775	2.34e-07 ***
춘천지구전적기념관	6.096e+00	1.325e+00	4.601	8.24e-05 ***
엘리시안강촌스키장	-1.642e-01	7.981e-02	-2.058	0.0490 *

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9796 on 28 degrees of freedom  
Multiple R-squared: 0.9568, Adjusted R-squared: 0.9429  
F-statistic: 68.86 on 9 and 28 DF, p-value: < 2.2e-16

Y : 열차 이용객 수 (Sum\_train)

X<sub>1</sub> : 강촌레일파크

~~X<sub>2</sub> : 국립춘천 박물관~~

X<sub>3</sub> : 김유정문학마을

X<sub>4</sub> : 남이섬 유원지

X<sub>5</sub> : 막국수체험 박물관

~~X<sub>6</sub> : 물레길~~

X<sub>7</sub> : 애니메이션 박물관

X<sub>8</sub> : 제이드가든

X<sub>9</sub> : 춘천 에티오피아 한국전 참전기념관

X<sub>10</sub> : 춘천지구전적기념관

X<sub>11</sub> : 엘리시안강촌 스키장



## 04 회귀 분석

### 〈 최종 다중선형회귀모형 〉

Call:  
lm(formula = sum\_train ~ .,  
data = wd\_before[(lm\_tr.lts\$residuals) < a & -b < (lm\_tr.lts\$residuals), c(-1, -2, -3, -5, -9, -16)])

Residuals:

Min	1Q	Median	3Q	Max
-17156	-5389	-869	7545	15720

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	2.148e+05	1.512e+04	14.208	2.51e-14 ***
강촌레일파크	5.420e-01	2.540e-01	2.134	0.0417 *
김유정문학마을	2.273e-01	9.815e-02	2.316	0.0281 *
남이섬유원지	7.309e-02	5.479e-02	1.334	0.1930
막국수체험박물관	-1.617e+01	2.330e+00	-6.938	1.53e-07 ***
애니메이션 박물관	9.555e-01	2.024e-01	4.721	5.94e-05 ***
제이드가든	7.120e-01	3.198e-01	2.227	0.0342 *
춘천에티오피아한국전참전기념관	3.237e+00	4.778e-01	6.775	2.34e-07 ***
춘천지구전적기념관	6.096e+00	1.325e+00	4.601	8.24e-05 ***
엘리시안강촌스키장	-1.642e-01	7.981e-02	-2.058	0.0490 *

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9796 on 28 degrees of freedom  
Multiple R-squared: 0.9568, Adjusted R-squared: 0.9429  
F-statistic: 68.86 on 9 and 28 DF, p-value: < 2.2e-16

Y : 열차 이용객 수 (Sum\_train)

X<sub>1</sub> : 강촌레일파크

~~X<sub>2</sub> : 국립춘천 박물관~~

X<sub>3</sub> : 김유정문학마을

X<sub>4</sub> : 남이섬 유원지

X<sub>5</sub> : 막국수체험 박물관

~~X<sub>6</sub> : 물레길~~

X<sub>7</sub> : 애니메이션 박물관

X<sub>8</sub> : 제이드가든

X<sub>9</sub> : 춘천 에티오피아 한국전 참전기념관

X<sub>10</sub> : 춘천지구전적기념관

X<sub>11</sub> : 엘리시안강촌 스키장

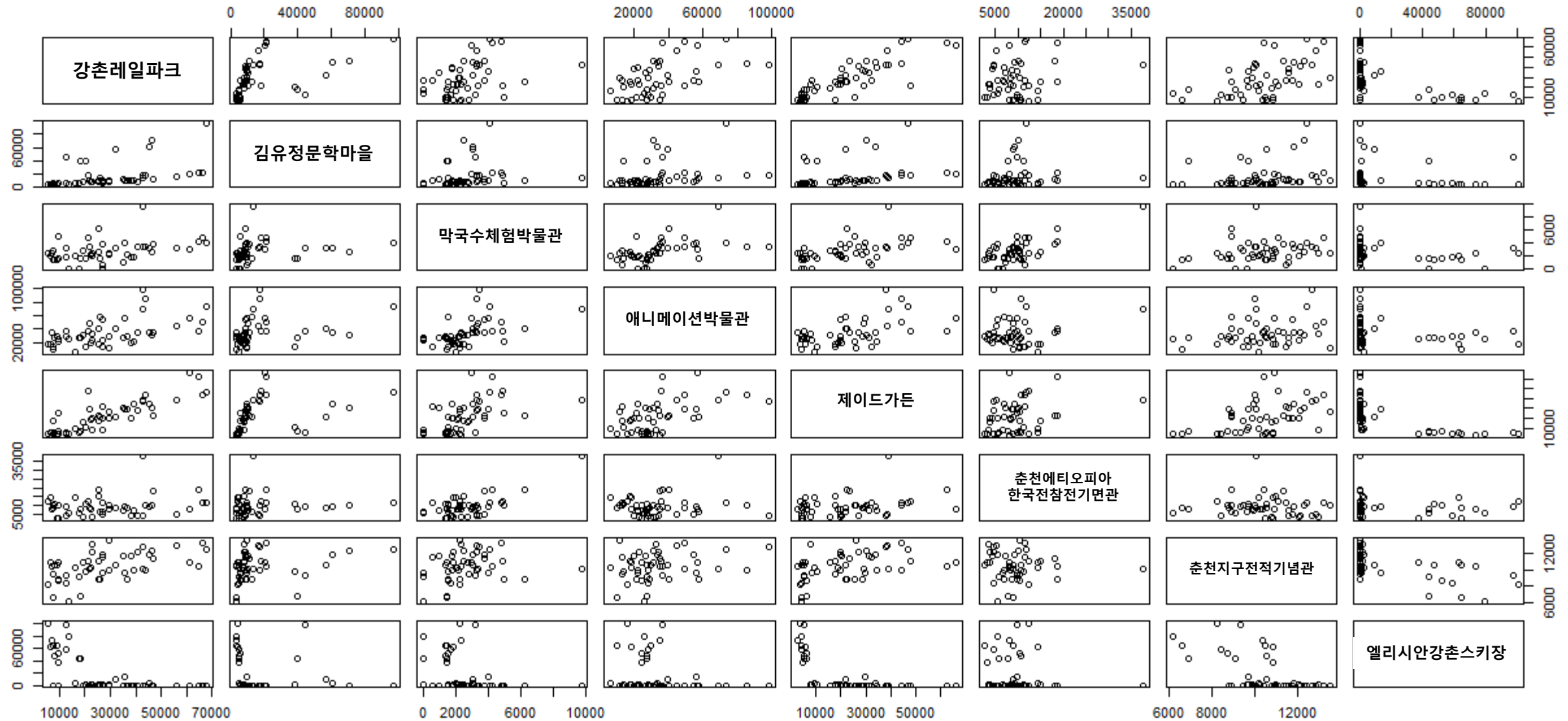
$R^2$	$adj R^2$	$F$	$df$	$P - Value$
0.9568	0.9429	68.86	28	2.2e-16

05

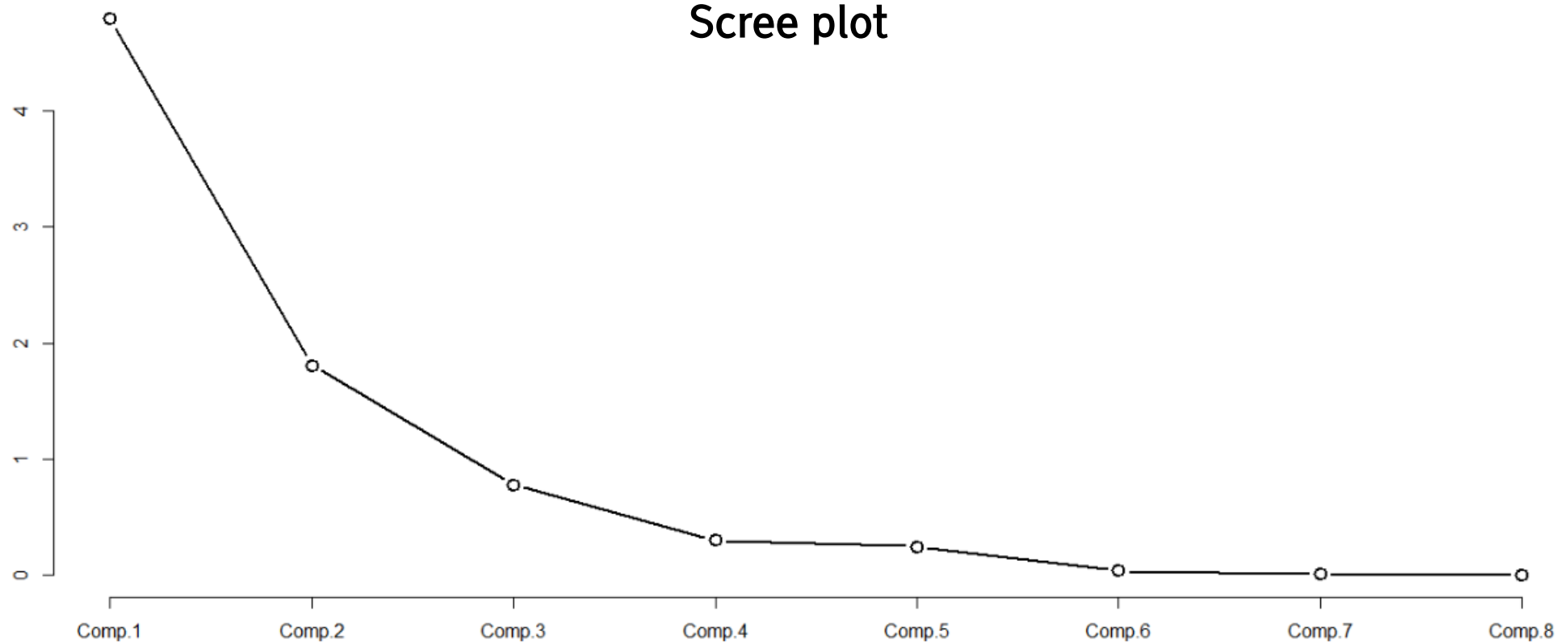
## 주성분 분석

## 05 주성분 분석

### 산점도 행렬



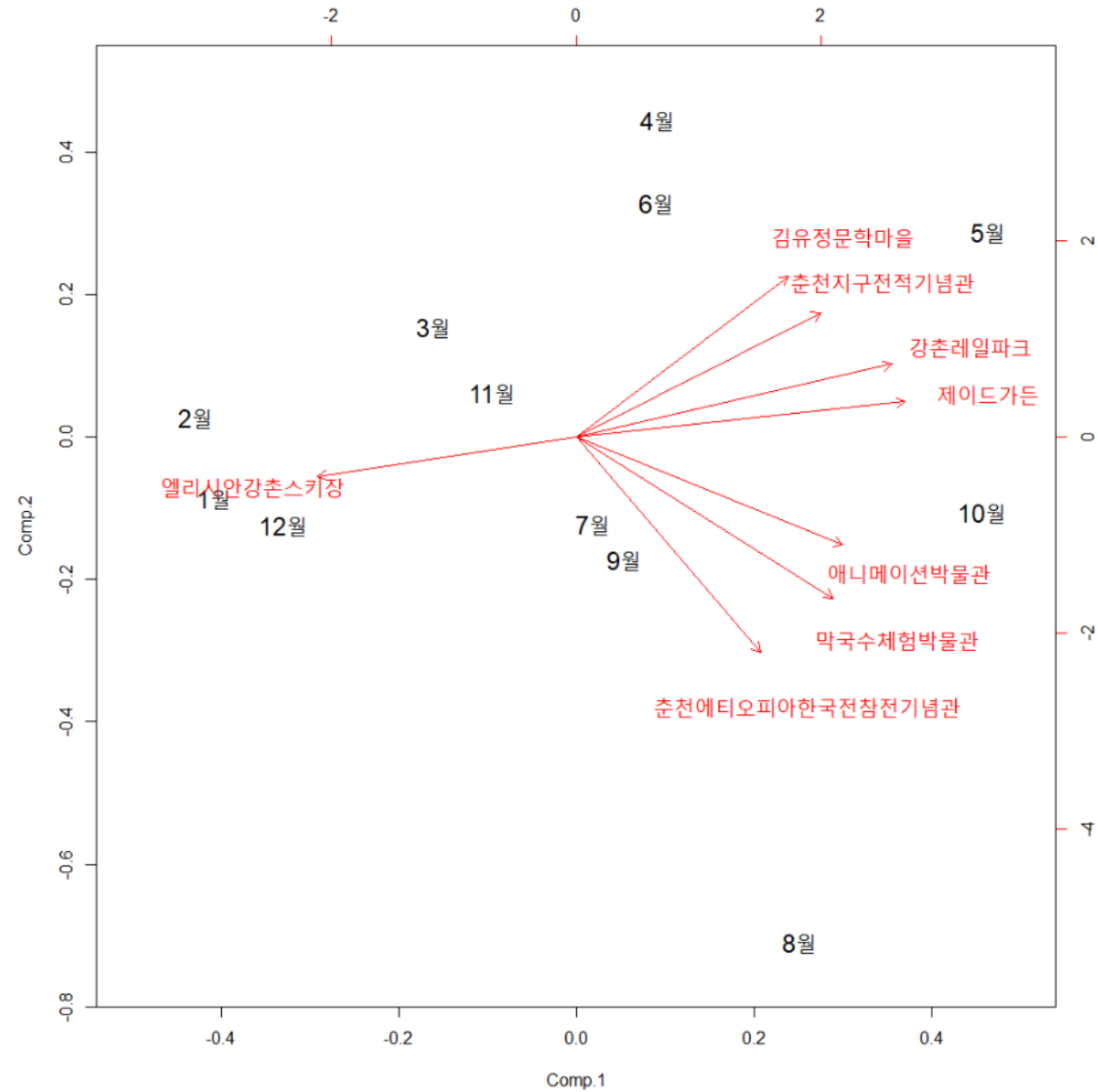
## 05 주성분 분석



Summary(PCA)	Comp.1	Comp.2	Comp.3	Comp.4	...	Comp.8
Standard deviation	2.1889	1.3452	0.8852	0.5496	...	0.0940
Proportion of Variance	0.5989	0.2262	0.0979	0.0378		0.0011
Cumulative Proportion	0.5989	0.8251	0.9231	0.9608		1.0000

## 05 주성분 분석

Biplot

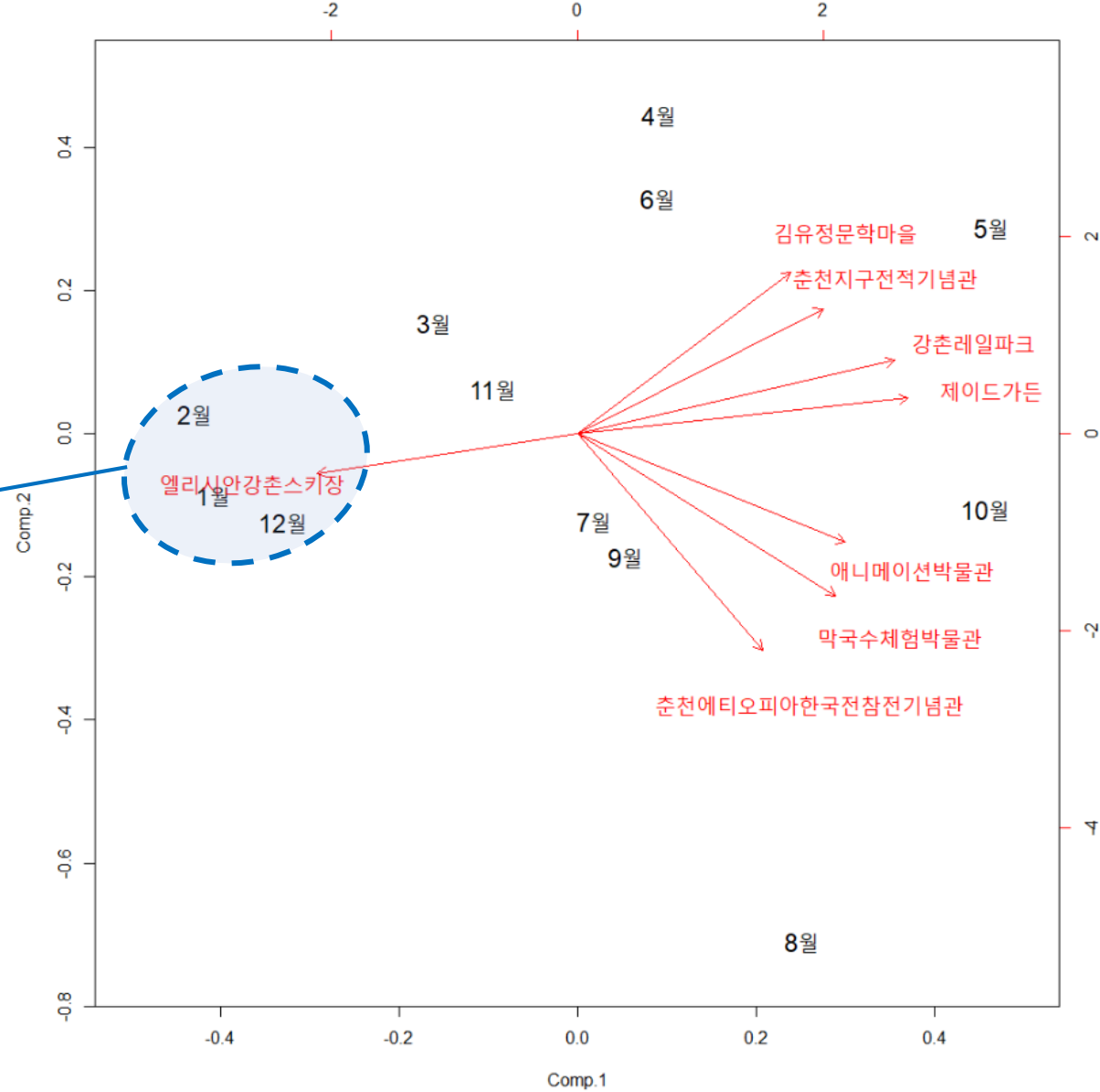


## 05 주성분 분석

### 동계 관광지

- 12월, 1월, 2월
- 동계 스포츠 관광객 ↑

Biplot

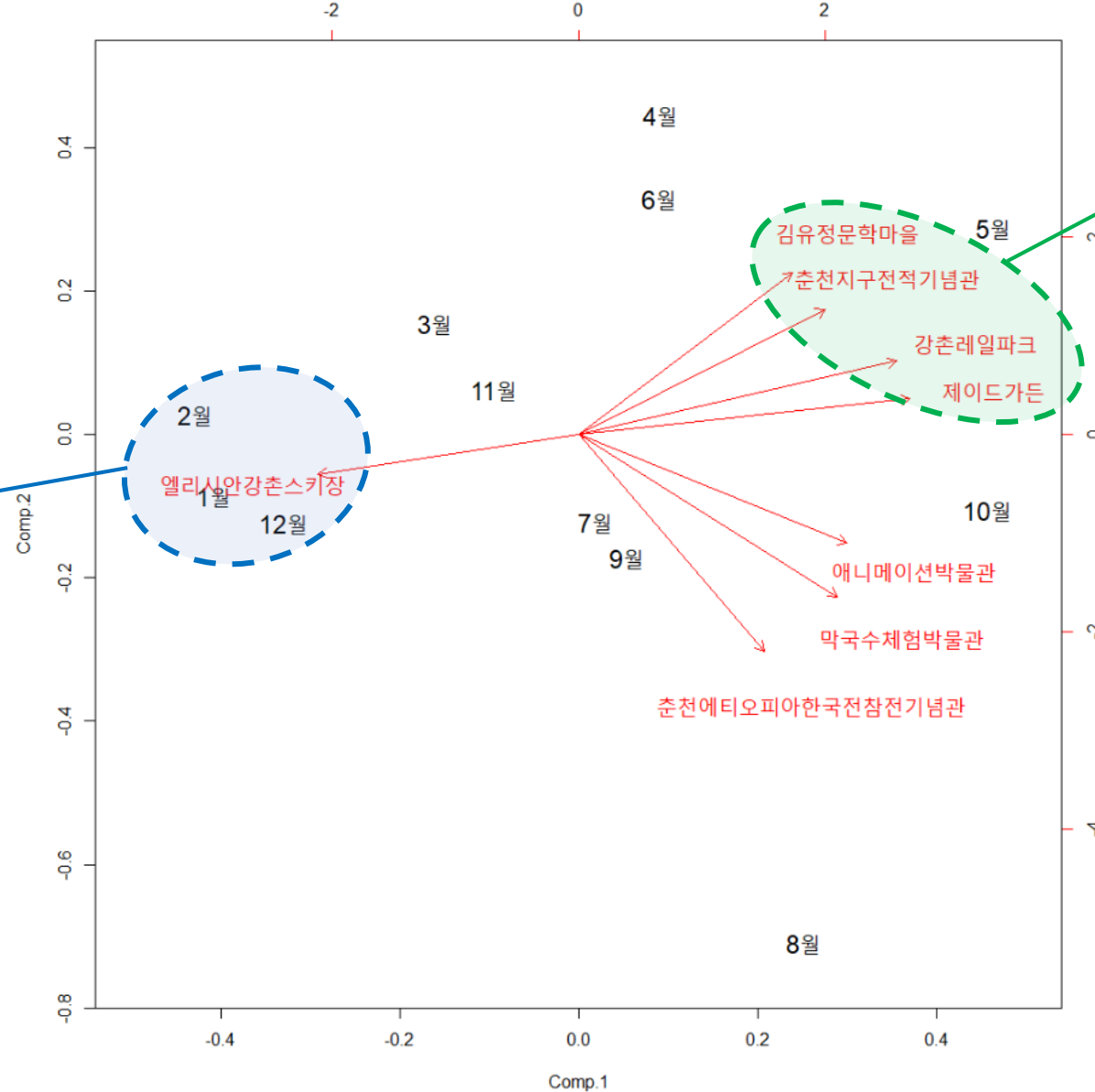


## 05 주성분 분석

Biplot

### 동계 관광지

- 12월, 1월, 2월
- 동계 스포츠 관광객 ↑

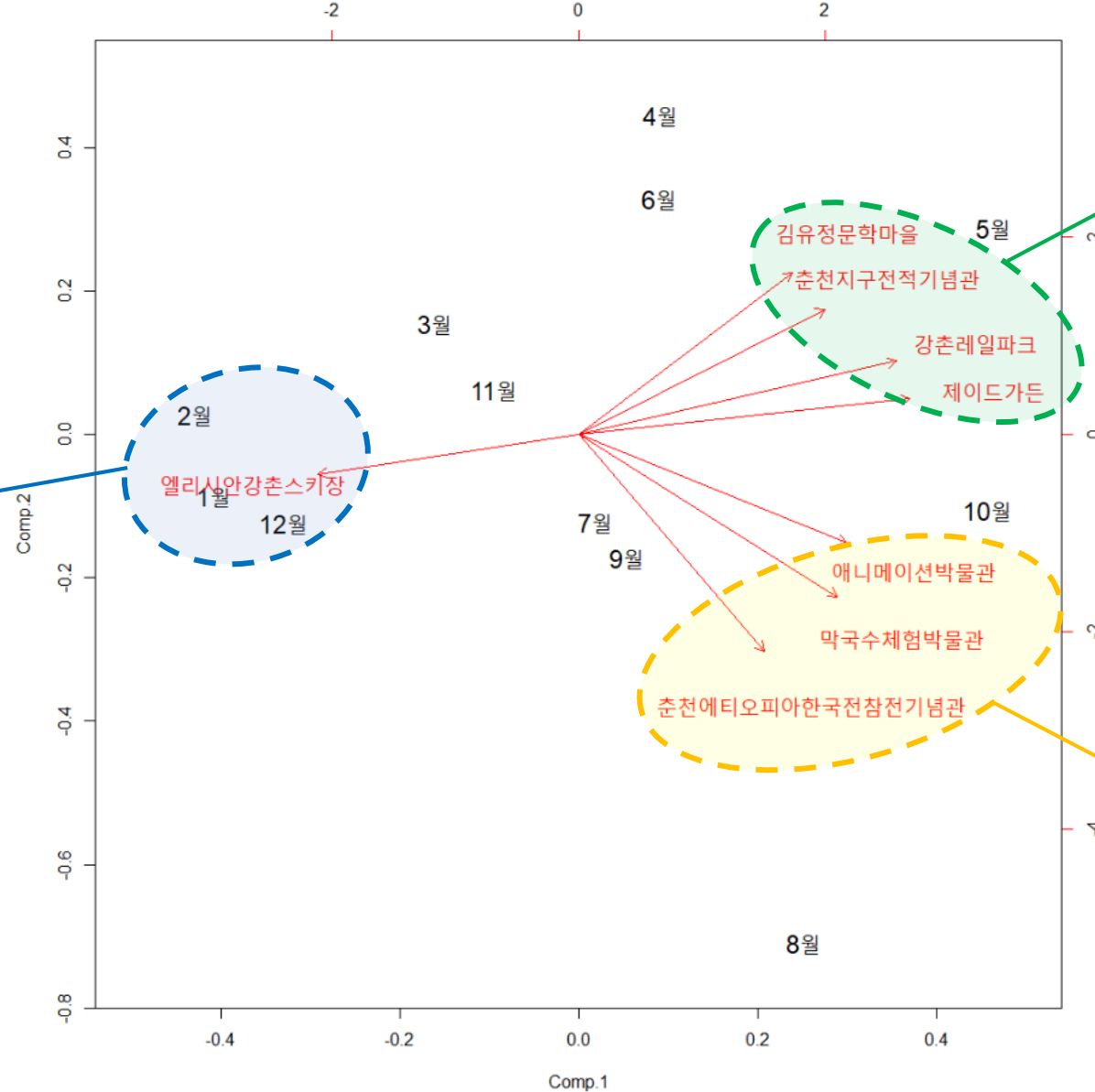


### 실외 관광지

- 5월, 4월, 6월
- 야외활동 하기 좋은 기온

## 05 주성분 분석

Biplot



### 동계 관광지

- 12월, 1월, 2월
- 동계 스포츠 관광객 ↑

### 실외 관광지

- 5월, 4월, 6월
- 야외활동 하기 좋은 기온

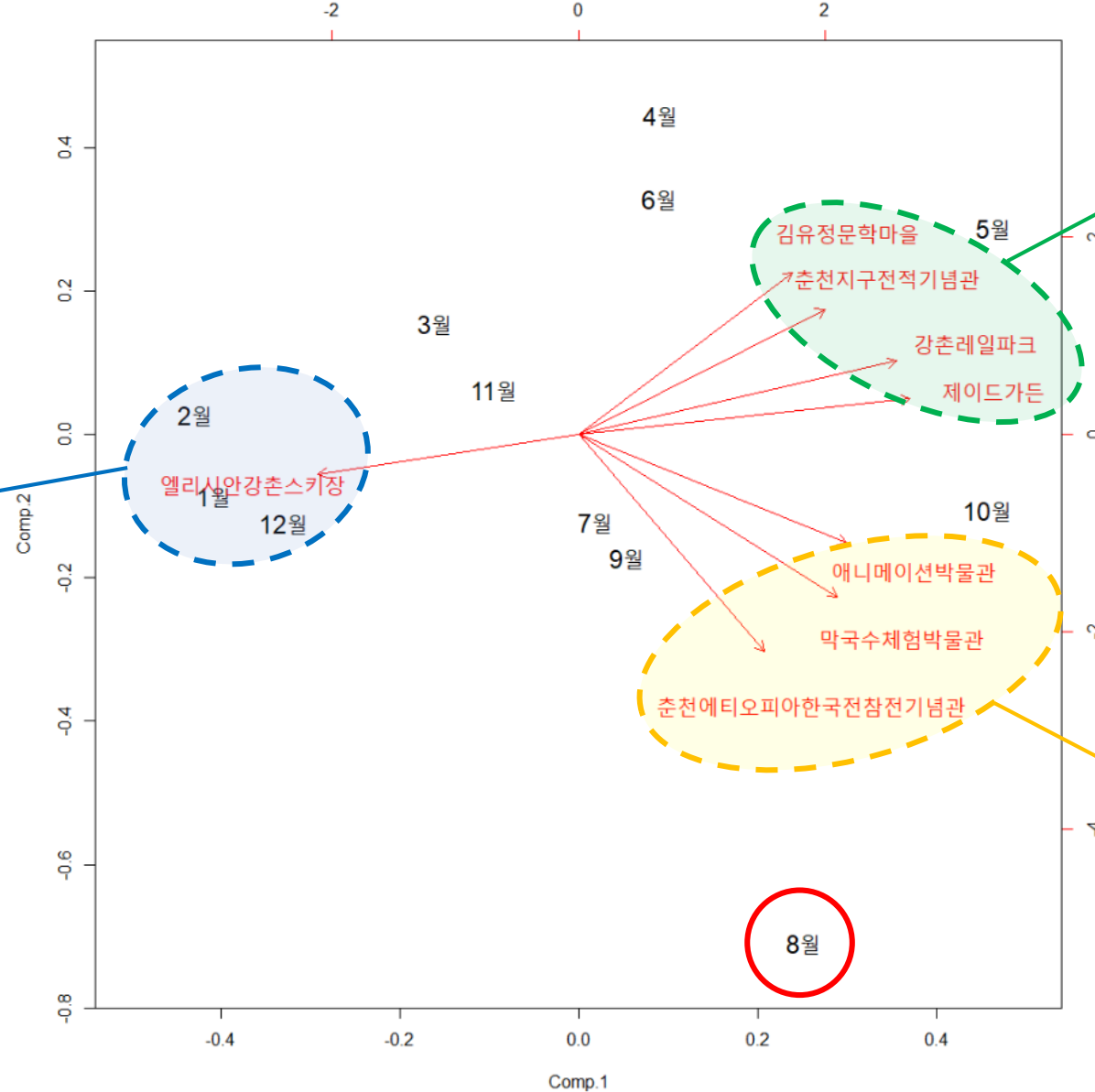
### 실내 관광지

- 10월
- 비교적 선선하여 실내 선호



## 05 주성분 분석

Biplot



### 동계 관광지

- 12월, 1월, 2월
- 동계 스포츠 관광객 ↑

### 실외 관광지

- 5월, 4월, 6월
- 야외활동 하기 좋은 기온

### 실내 관광지

- 10월
- 비교적 선선하여 실내 선호

06

정리 및 요약

# “철도 이용객 수와 춘천 관광객 수는 어떤 관계가 있는가?”

### 1. 역별 여객수송실적 (월)

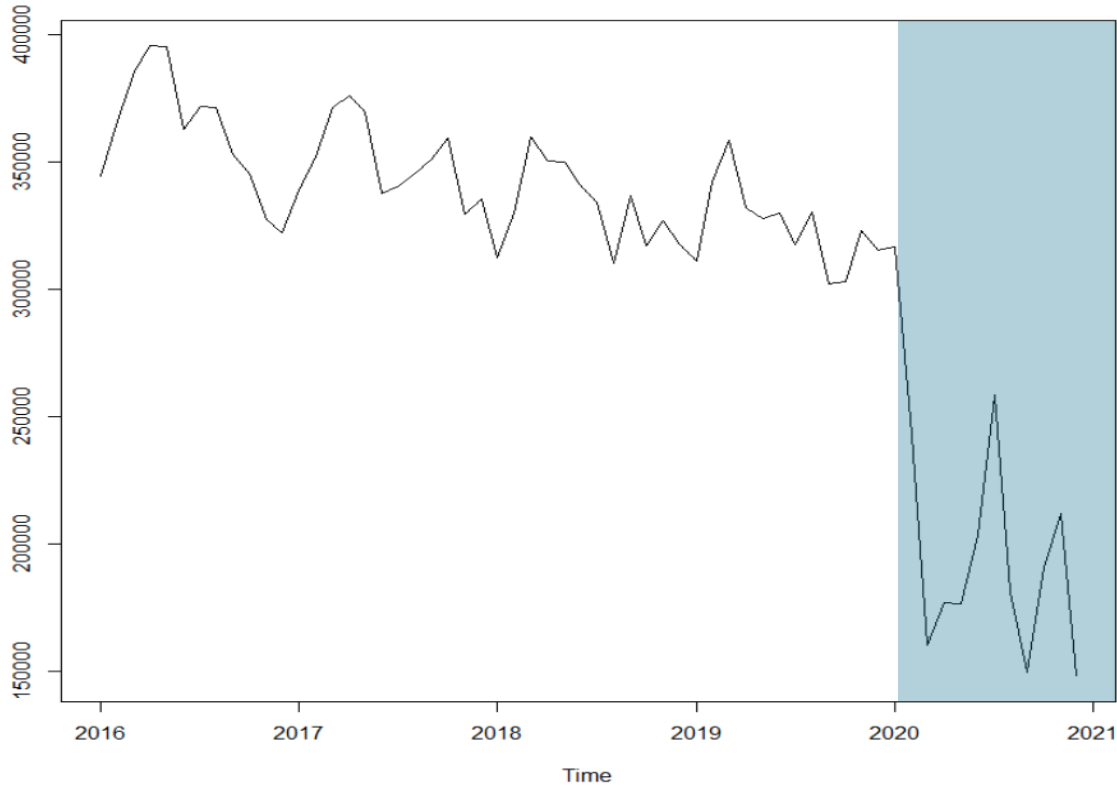
변수명	변수 설명
Year_month	날짜 변수 (“YYYYMM”)
ITX_count	해당 연월의 ‘ITX’ 하차 승객 수
GC_count	해당 연월의 ‘경춘선’ 하차 승객 수
Sum_train	해당 연월의 ITX, 경춘선 전철의 하차 승객 수 합

### 2. 전국 주요관광지점 입장객 통계

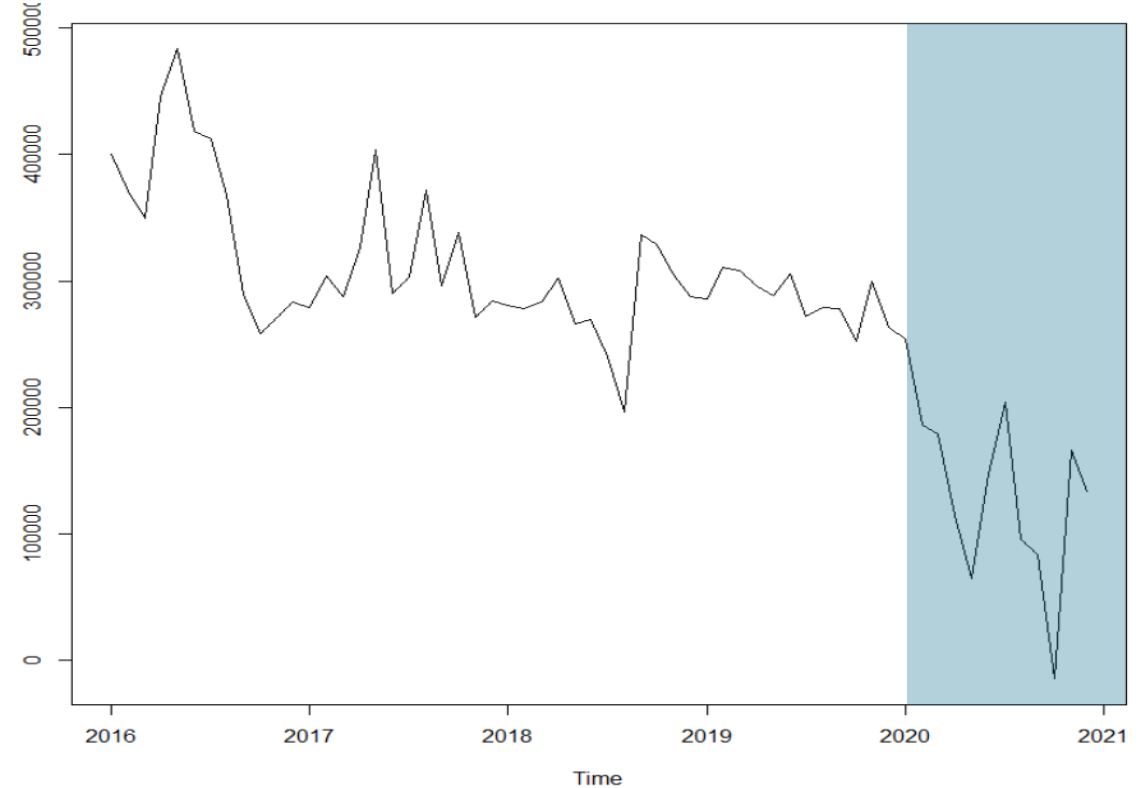
변수명	변수 설명
Year_month	날짜 변수 (“YYYYMM”)
관광지 1, 2, ...	해당 연월의 ‘춘천시 관광지 입장객’ 수
⋮	⋮
Sum_travel	해당 연월의 관광지 입장객 수 합

## 06 정리 및 요약 - 시계열 분석

〈계절 요인을 제거한 월별 철도 이용객 수〉



〈계절 요인을 제거한 월별 총 관광객 수〉

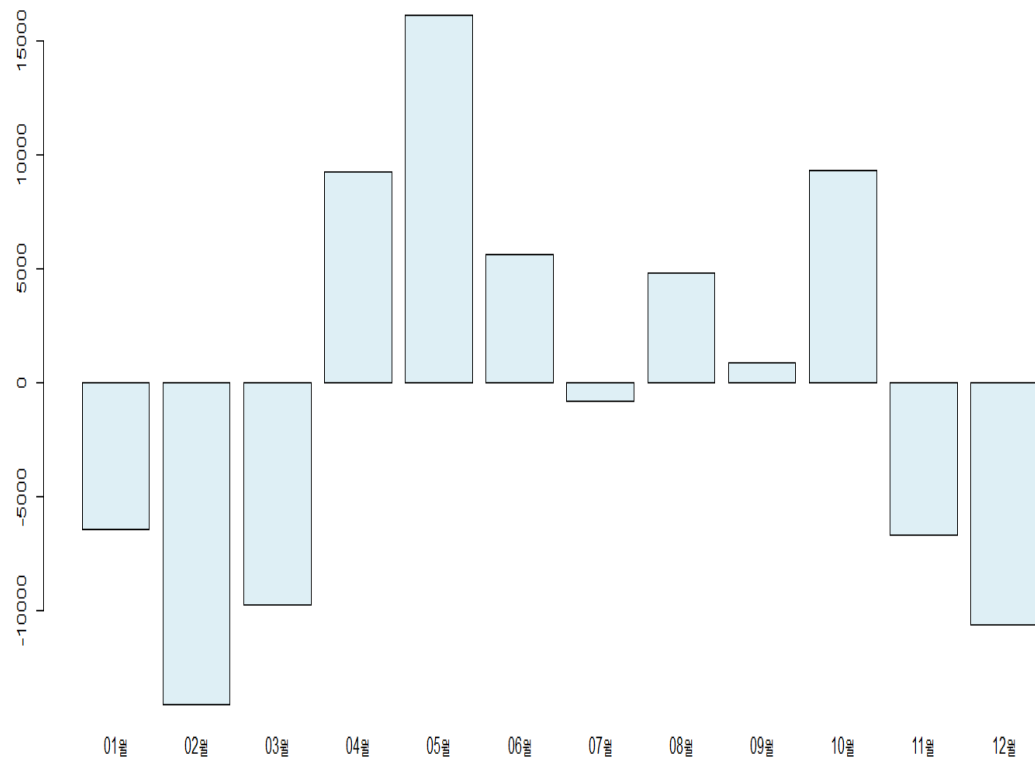


계절성을 제거하였음에도 2020년을 기점으로 시계열 분포에 큰 변화가 식별

**코로나 19의 영향으로 추측**

## 06 정리 및 요약 - 시계열 분석

〈 코로나 19 이전 Row Effects 〉



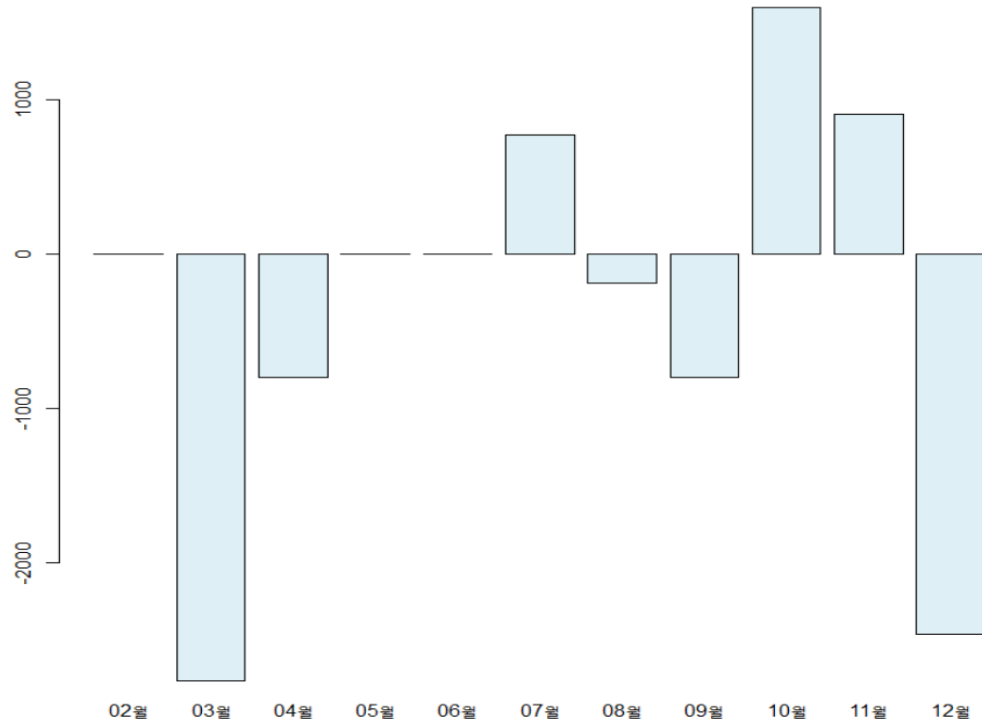
〈 코로나 19 이전 Residuals 〉

	남아섬유원지	강촌레일파크	제이드가든	김유정 문학마을	물레길	애니메이션 박물관	국립춘천 박물관	춘천지구 전적 기념관	막국수체험 박물관	에티오피아 참전기념관	엘리시안강촌
01월	-166099.2	-53030.2	-64043.4	-648.8	-3521.6	33196.2	14994.4	10840.5	153.6	3075.2	413543.1
02월	-211275.8	-38150.8	-56488.9	-216.2	4556.9	-9166.2	1224.9	4968	3246.1	-995.2	217049.6
03월	-186602.2	-12505.2	-44910.4	216.2	2398.4	-31086.8	-9359.6	12099.5	153.6	3372.2	7622.1
04월	89284.8	31588.8	7416.6	23928.2	-8468.6	-35662.8	20667.4	-366.5	-17136.4	-18979.8	-18503.9
05월	336522	82031	92385.9	69084.5	-129.4	78315.5	-2079.4	-9000.2	-16455.1	-20311.5	-25249.6
06월	48787.8	37793.8	31879.6	15745.2	1021.4	-18865.8	-15180.6	-366.5	-12353.4	-15132.8	-12204.9
07월	-131866.2	-23208.2	-12035.4	13477.2	129.4	9166.2	-6387.6	366.5	153.6	892.2	844.1
08월	3202.5	1547.5	18815.4	-33540	-3175.9	80589	-1224.9	-4571.8	2223.4	16037	-844.1
09월	-3202.5	21801.5	11561.4	-23716	2558.1	53610	14964.1	-1984.8	-5366.6	0	-9989.1
10월	642883	99924	88837.9	1743.5	-129.4	144896.5	-8642.4	-3101.2	-8941.1	-6005.5	-18552.6
11월	216689.5	-1547.5	-7416.6	-27781	2216.1	-21620	7839.1	9691.2	-153.6	0	2953.9
12월	-224343	-64064	-57595.1	-35832.5	-129.4	-10902.5	31511.6	11430.8	1483.9	7317.5	200440.4

월별 기온의 변화에 따라 관광지의 특성이 나뉨

## 06 정리 및 요약 - 시계열 분석

〈 코로나 19 이후 Row Effects 〉



한국 코로나19 일일 확진자 추세

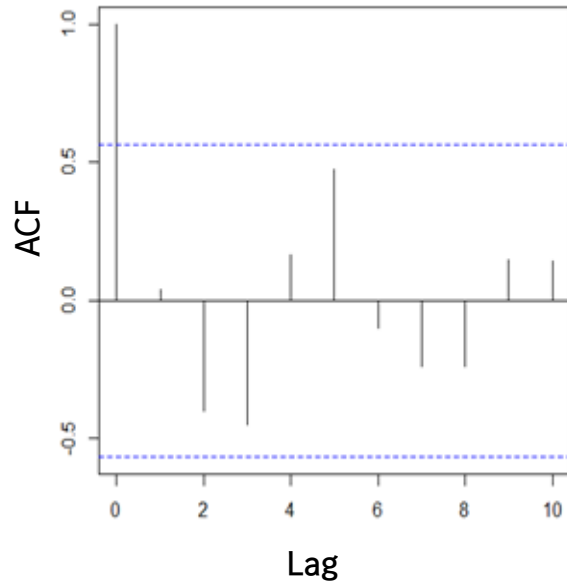
자료: 질병관리청, 단위: 명



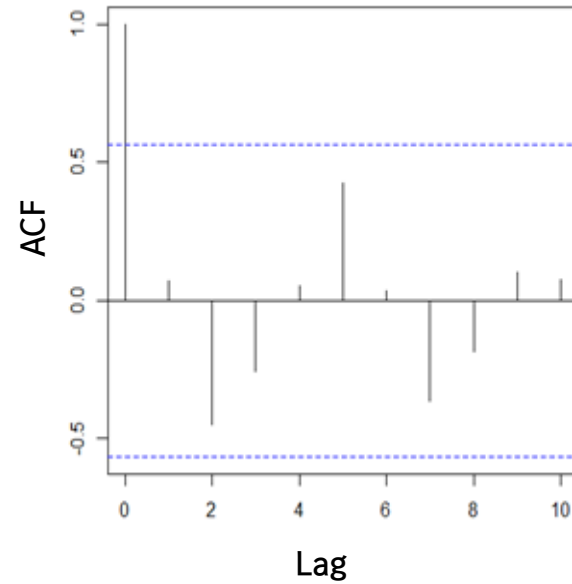
코로나 19 이후 행 효과는 **코로나 19 확진자 수(외생변수)**에 영향을 받음

## 06 정리 및 요약 - 시계열 분석

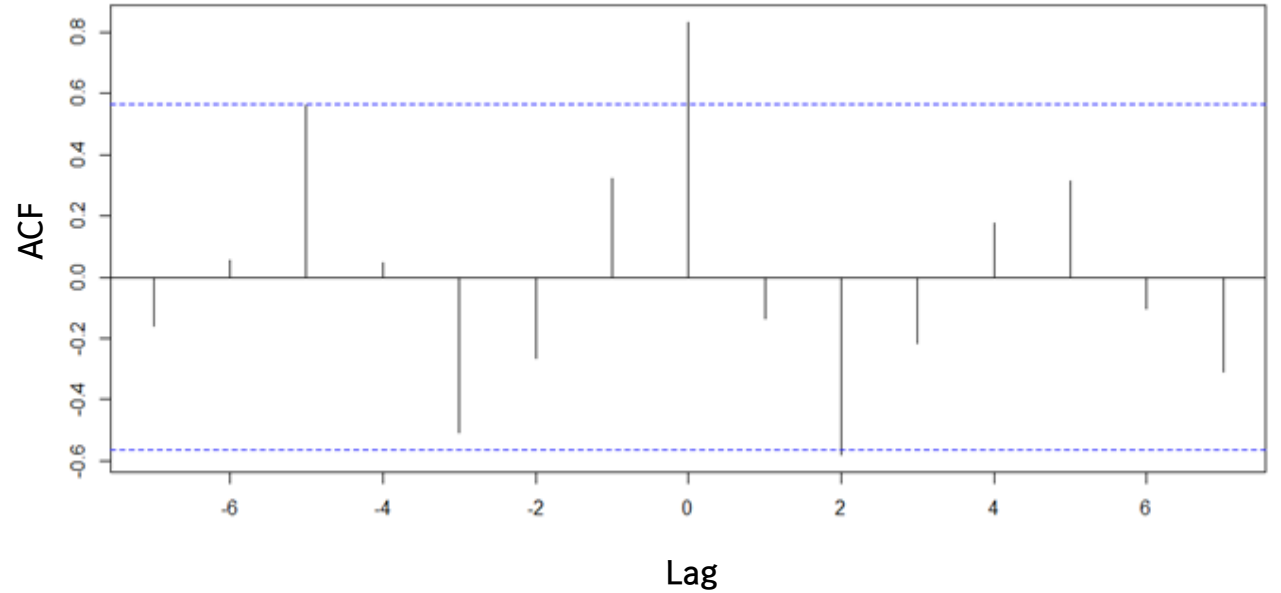
〈 철도 이용객 수 ACF 〉



〈 관광객 수 ACF 〉



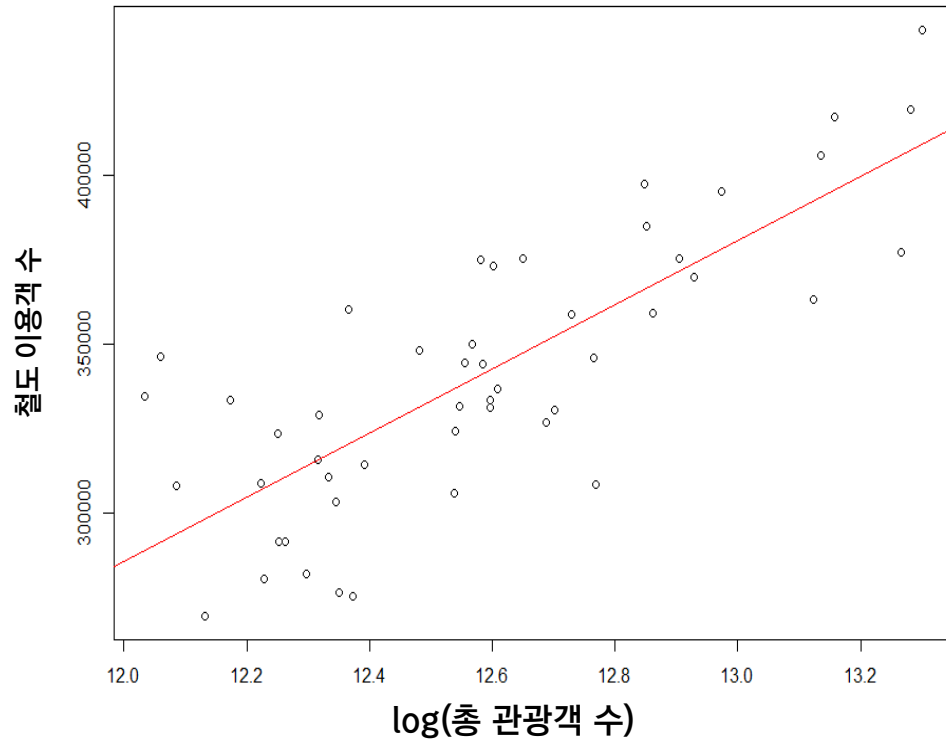
〈 철도 이용객수 & 관광객 수 CCF 〉



코로나 19 이전 두 변수의 ACF와 CCF를 확인한 결과  
같은 시점에서 매우 강한 양의 상관관계가 있다고 판단

## 06 정리 및 요약 - 회귀 분석

### 〈 철도 이용객 수 ~ log(총 관광객 수) 단순 선형회귀 〉



Call:

```
lm(formula = sum_train ~ log(sum_travel), data = wd_before)
```

Residuals:

Min	1Q	Median	3Q	Max
-49994	-18296	-151	16595	54759

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-850914	135321	-6.288	9.85e-08 ***
$\log(\text{sum\_travel})$	94725	10751	8.811	1.62e-11 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 25030 on 47 degrees of freedom

Multiple R-squared: 0.6229, Adjusted R-squared: 0.6148

F-statistic: 77.63 on 1 and 47 DF, p-value: 1.617e-11

$$\hat{y} = -850914 + 94725 * \log(x)$$

코로나 19 이후 데이터를 제외한 후 단순선형회귀분석 결과

$\log(\text{총 관광객 수})$ 가 증가함에 따라 철도 이용객 수 또한 증가



## 06 정리 및 요약 - 회귀 분석

### 〈 최종 다중선형회귀모형 〉

Call:

```
lm(formula = sum_train ~ .,  
    data = wd_before[(lm_tr.lts$residuals) < a & -b < (lm_tr.lts$residuals), c(-1, -2, -3, -5, -9, -16)])
```

Residuals:

Min	1Q	Median	3Q	Max
-17156	-5389	-869	7545	15720

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	2.148e+05	1.512e+04	14.208	2.51e-14 ***
강촌레일파크	5.420e-01	2.540e-01	2.134	0.0417 *
김유정문학마을	2.273e-01	9.815e-02	2.316	0.0281 *
남이섬유원지	7.309e-02	5.479e-02	1.334	0.1930
막국수체험박물관	-1.617e+01	2.330e+00	-6.938	1.53e-07 ***
애니메이션.박물관	9.555e-01	2.024e-01	4.721	5.94e-05 ***
제이드가든	7.120e-01	3.198e-01	2.227	0.0342 *
춘천에티오피아한국전참전기념관	3.237e+00	4.778e-01	6.775	2.34e-07 ***
춘천지구전적기념관	6.096e+00	1.325e+00	4.601	8.24e-05 ***
엘리시안강촌스키장	-1.642e-01	7.981e-02	-2.058	0.0490 *

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9796 on 28 degrees of freedom

Multiple R-squared: 0.9568, Adjusted R-squared: 0.9429

F-statistic: 68.86 on 9 and 28 DF, p-value: < 2.2e-16

$Y$  : 열차 이용객 수 (Sum\_train)

$X_1$  : 강촌레일파크

~~$X_2$  : 국립춘천 박물관~~

$X_3$  : 김유정문학마을

$X_4$  : 남이섬 유원지

$X_5$  : 막국수체험 박물관

~~$X_6$  : 물레길~~

$X_7$  : 애니메이션 박물관

$X_8$  : 제이드가든

$X_9$  : 춘천 에티오피아 한국전 참전기념관

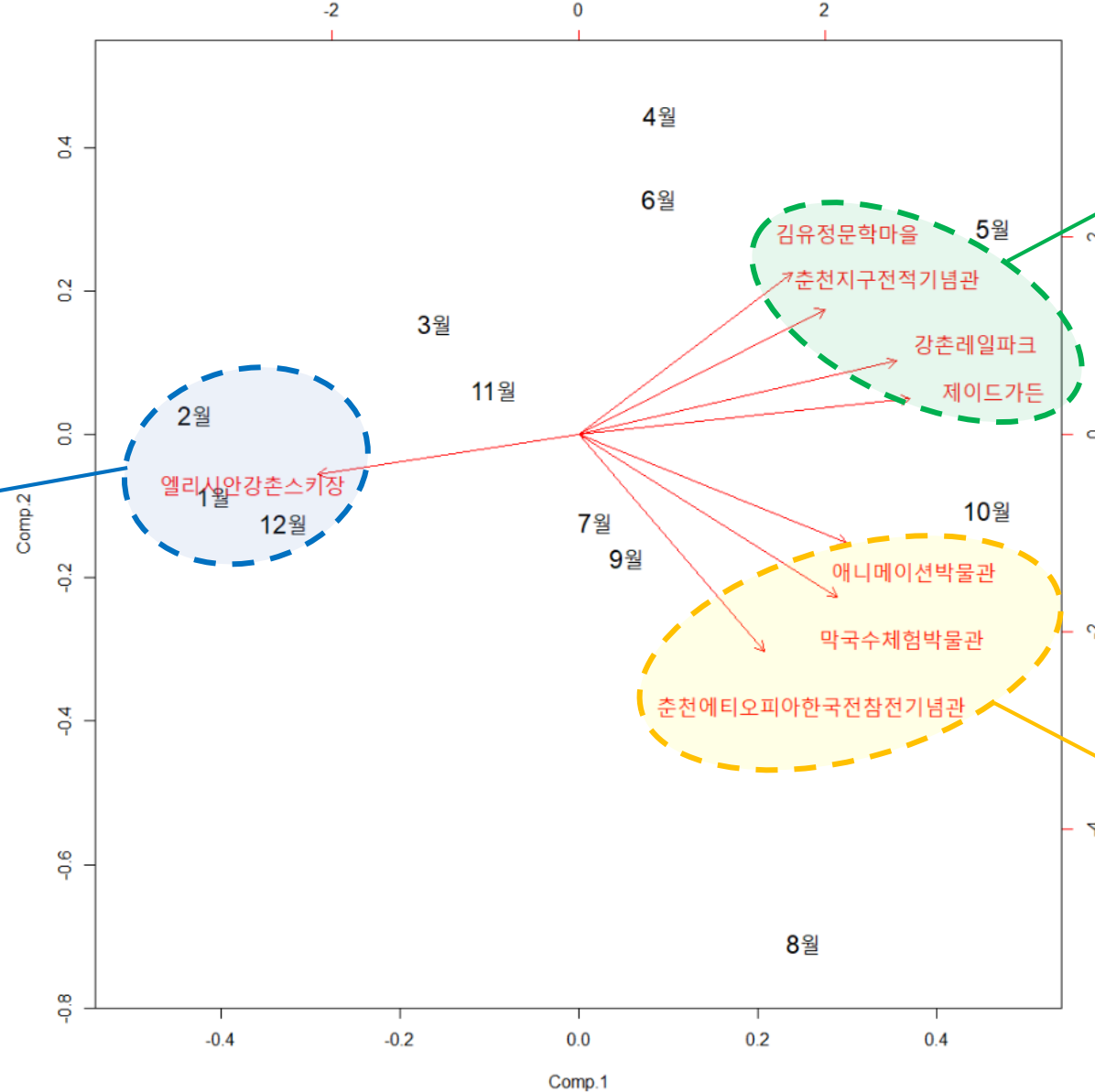
$X_{10}$  : 춘천지구전적기념관

$X_{11}$  : 엘리시안강촌 스키장

## 06 정리 및 요약 - 주성분 분석

### 동계 관광지

- 12월, 1월, 2월
- 동계 스포츠 관광객 ↑



### 실외 관광지

- 5월, 4월, 6월
- 야외활동 하기 좋은 기온

### 실내 관광지

- 10월
- 비교적 선선하여 실내 선호

**THANK  
YOU**