



서울 교통량 데이터 분석

데이터사이언스기초(03) 기말프로젝트

20205107 강지현

목차

주제 선정 이유

서울 교통량 데이터에 대한 가설 설정

연도별에 따른 서울 교통량 데이터 분석

장소별에 따른 서울 교통량 데이터 분석

월별에 따른 서울 교통량 데이터 분석

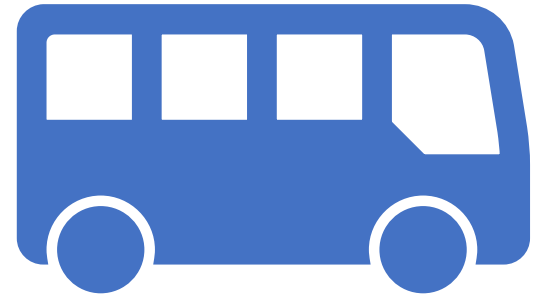
서울 교통량 데이터 연도별 예측

서울 교통량 데이터 월별 예측

검정통계량에 따른 기각역, 기각값

주제 선정 이유

- 서울에는 인구 밀도가 높고 교통 수요가 많은 도시로 각종 행사, 축제가 많아 해마다 교통 체증이 발생하여 빈번한 사건 사고가 발생했습니다. 2017년 7월 5일에는 대규모 교통 체증으로 인해 도로와 교통망이 마비된 적도 있습니다.
- 교통량 데이터를 장소와 월별, 연도별 교통량을 분석함으로써 혼잡한 구간이나 교통 체증이 발생하는 구간을 확인하는 등 교통 패턴 및 추세를 파악 하고,
도로 개선 및 확장을 위한 우선 순위를 정해서 교통 흐름을 개선하는 방안 등을 모색할 수 있을 것 같아서 주제를 선정하게 되었습니다.



서울 교통량 데이터에 대한 가설 설정

1. 귀무가설(H_0): 서울의 교통량은 연도별로 차이가 없다.
대립가설(H_1): 서울의 교통량은 연도별로 차이가 있다.
 2. 귀무가설(H_0): 서울의 교통량은 장소별로 차이가 없다.
대립가설(H_1): 서울의 교통량은 장소별로 차이가 있다.
 3. 귀무가설(H_0): 서울의 교통량은 월별에 따라 차이가 없다.
대립가설(H_1): 서울의 교통량은 월별에 따라 차이가 있다.
- 종속변수 : 교통량 (차량 수)
 - 독립변수 : 연도, 장소 (서울 내의 특정 지역, 교량 지점), 월별

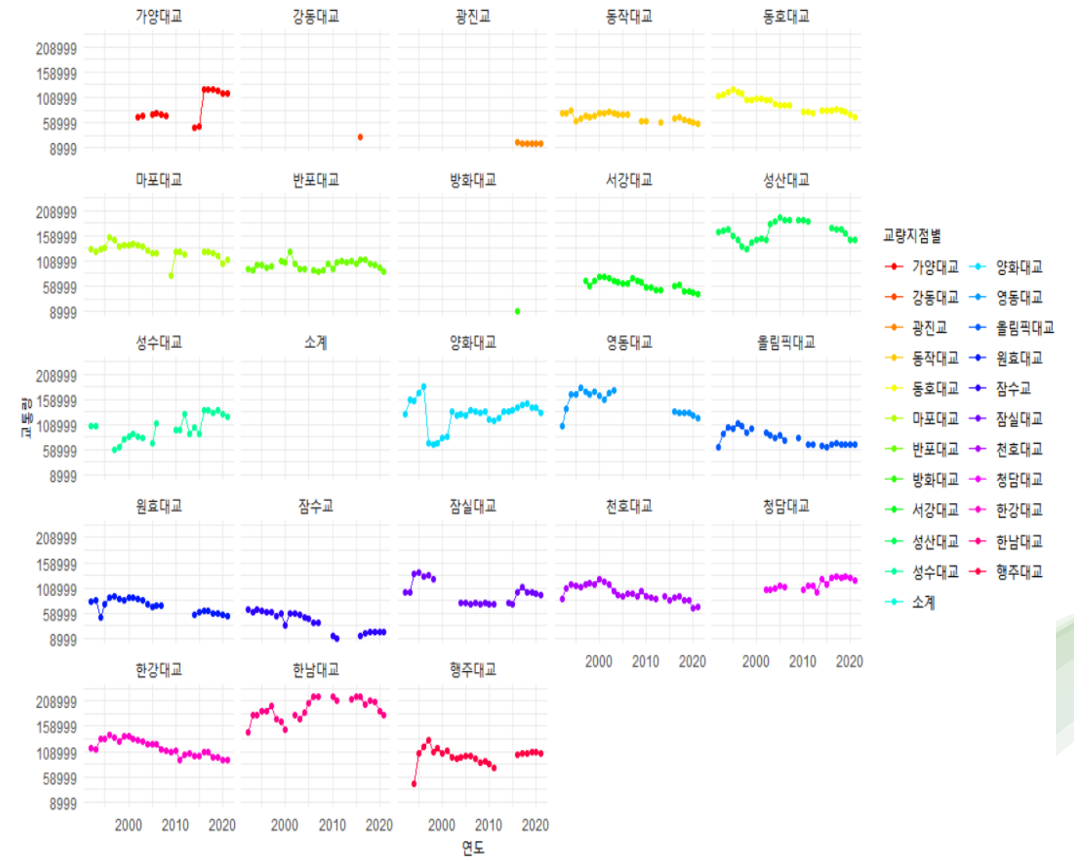
연도별 장소별에 따른 서울 교통량 데이터 분석

교량지점별로 연도에 따른 교통량의 추이를 시각화 하는 그래프로 각 교량지점별로 선 그래프와 점 그래프를 표시해서 구분해주었다.

x축은 연도 y축은 교통량으로 설정 해준 뒤 geom_line()함수를 사용하여 연도에 따른 교통량의 추이를 표현하였다. 또한, geom_point()를 활용하여 각 연도의 교통량을 나타냈다. 최종적으로

facet_wrap()을 사용하여 교량지점별에 따라 그래프를 분할하여 그래프를 그렸으며 스크립트 내 그래프 크기가 너무 작아서 x11() 함수를 사용하여 새로운 창을 띄워 그래프를 표시하였다.

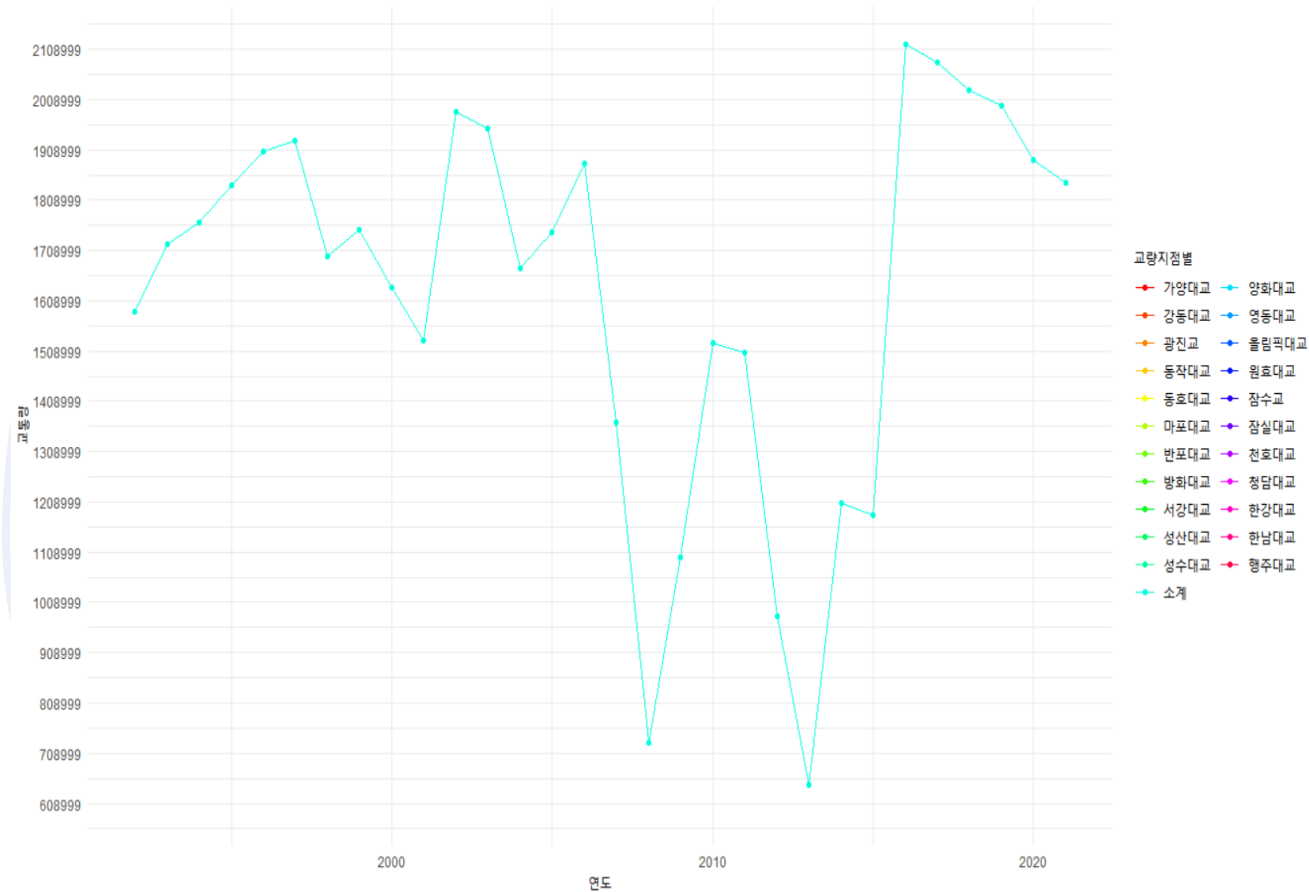
이 그래프를 보면 한남대교가 가장 교통량이 많고 잠수교 교통량이 비교적 적은 것을 확인 할 수 있다. 연도별에 따른 교통량은 한눈에 눈에 보이지 않아 각 교량지점별(장소별) 합계를 연도에 따라 그래프를 시각화 해보았다.



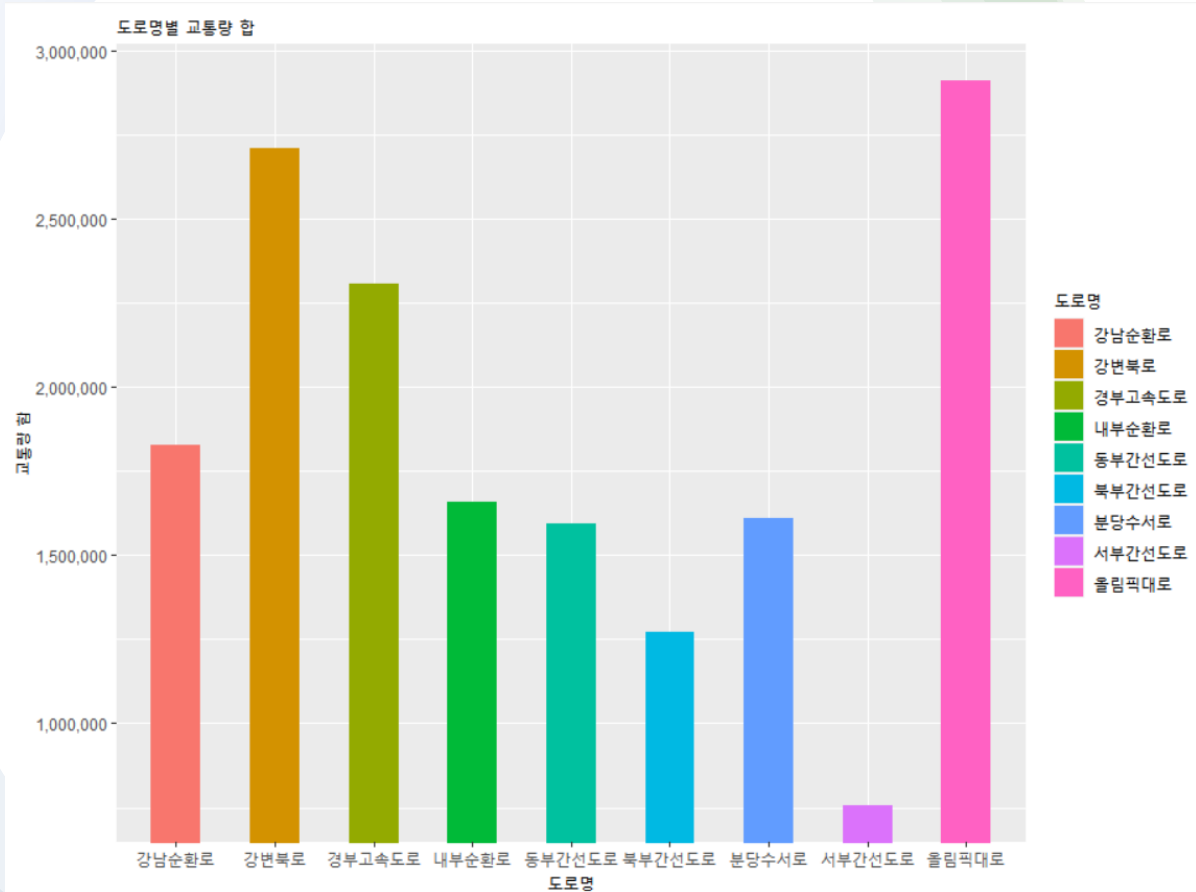
연도별에 따른 서울 교통량 데이터 분석

이 그래프는 교량지점 전체를 더한 합계를 활용하여 연도별에 따른 교통량 그래프를 그린 것이다. 이는 서울에 있는 많은 대교들의 합으로 그려낸 것으로 연도별 서울 교통량으로 볼 수 있다.

이 그래프를 보면 2015년에서 2017년 사이에 교통량이 급변하는 것을 확인할 수 있는데 이는 2017년 7월 5일 서울 강남구 신사동 일대에서 대규모 교통체증 때문이라고 예측할 수 있을 것 같다.



장소별에 따른 서울 교통량 데이터 분석

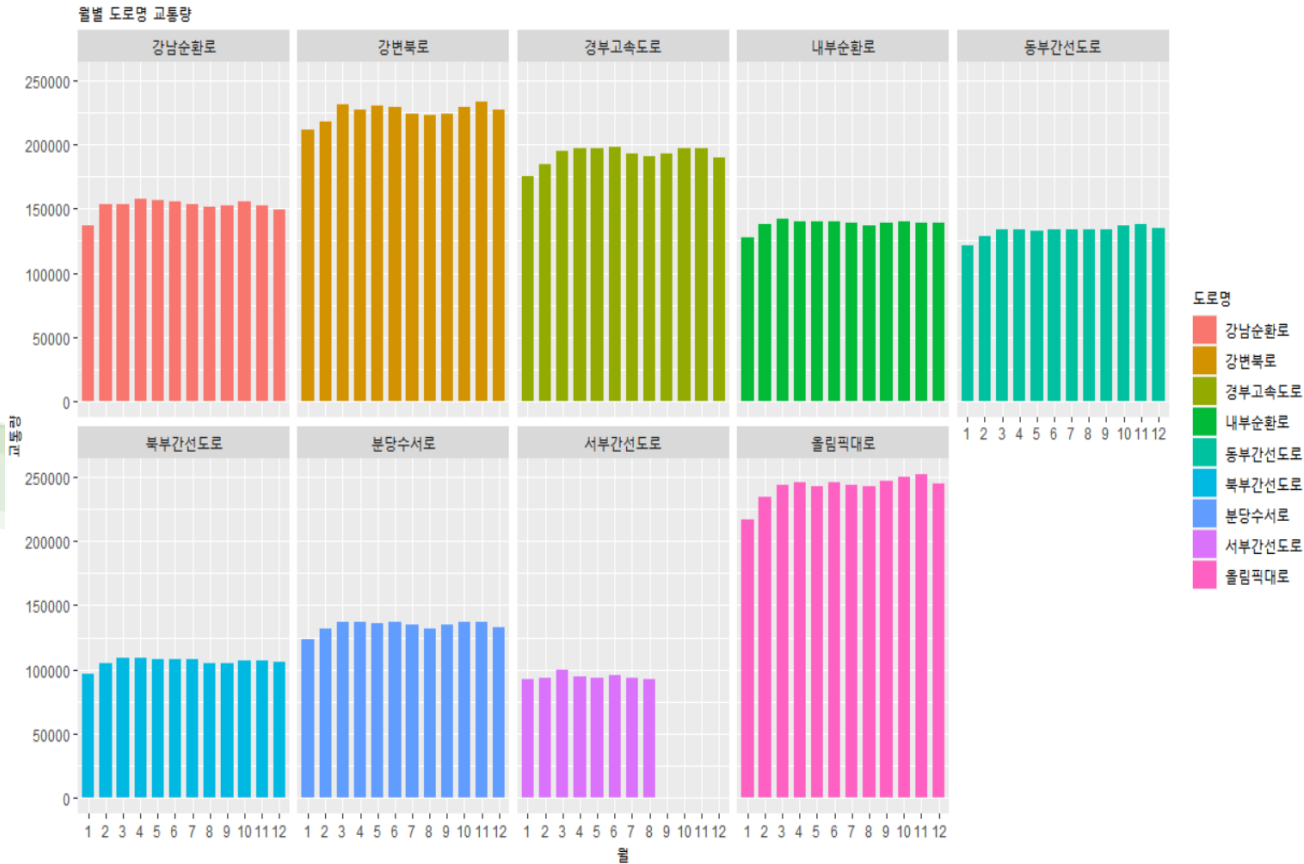


이 그래프는 각 도로명 별 교통량의 합을 그래프를 나타낸 것으로 가장 교통량이 많은 도로는 올림픽 대로인 것을 알 수 있다.

서부간선도로와 북부간선도로는 다른 도로에 비해 비교적 교통량이 적은 편이고 가장 높은 올림픽대로 교통량은 강남순환로의 교통량의 2배정도 되어 보이는 것을 확인할 수 있다.

이 그래프 결과를 보면 올림픽대로를 가장 우선적으로 도로 개선 및 확장을 해서 교통 흐름을 개선해야 할 것 처럼 보인다.

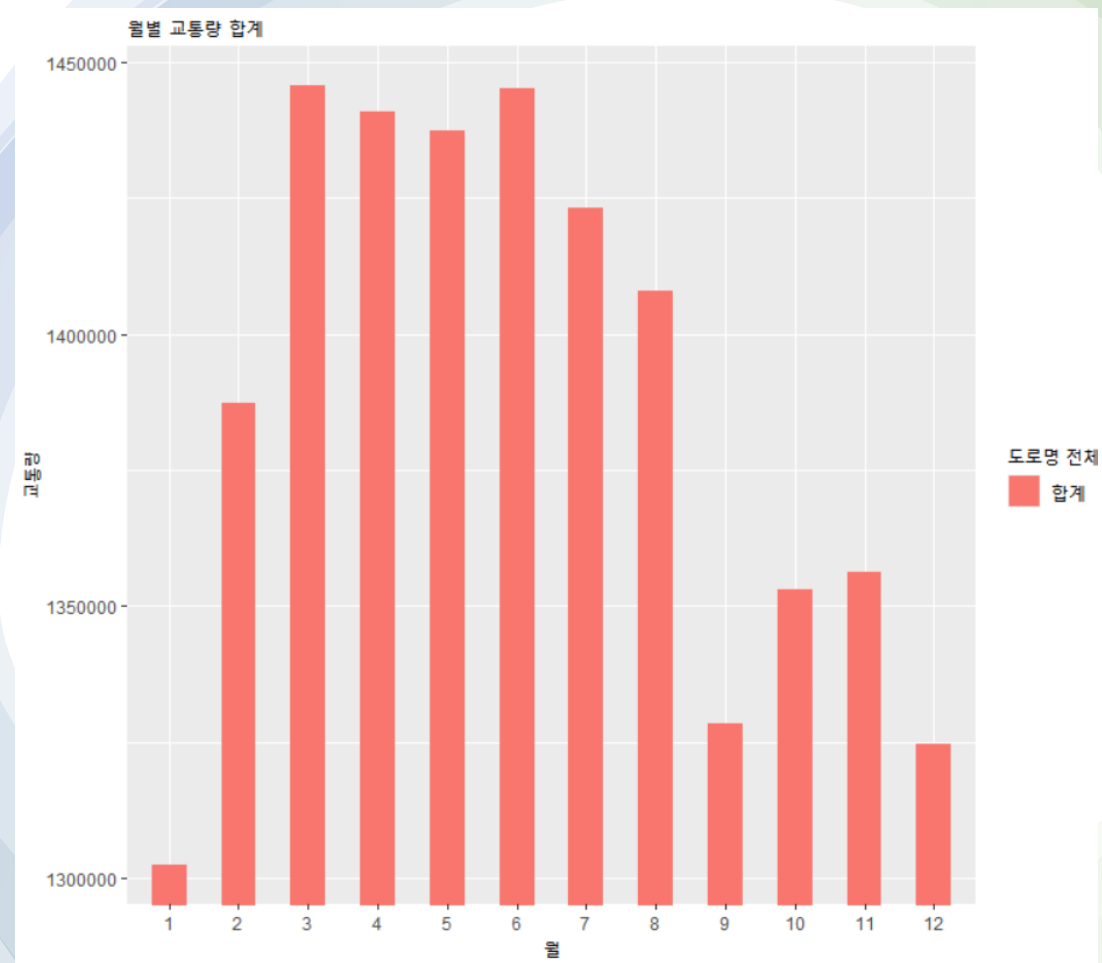
월별에 따른 서울 교통량 데이터 분석



이 그래프는 월별에 따라 각 도로명을 분할하여 그래프를 그린 것이다. 그래프를 각 도로명으로 분할해서 표현하다 보니 앞에서 본 장소별에 따른 서울 교통량 데이터 분석과 유사하게 월별에 따른 데이터보다 장소(도로명)에 따른 데이터 시각화가 더 잘 나타난 것으로 보인다.

월별에 따른 서울 교통량 데이터 분석

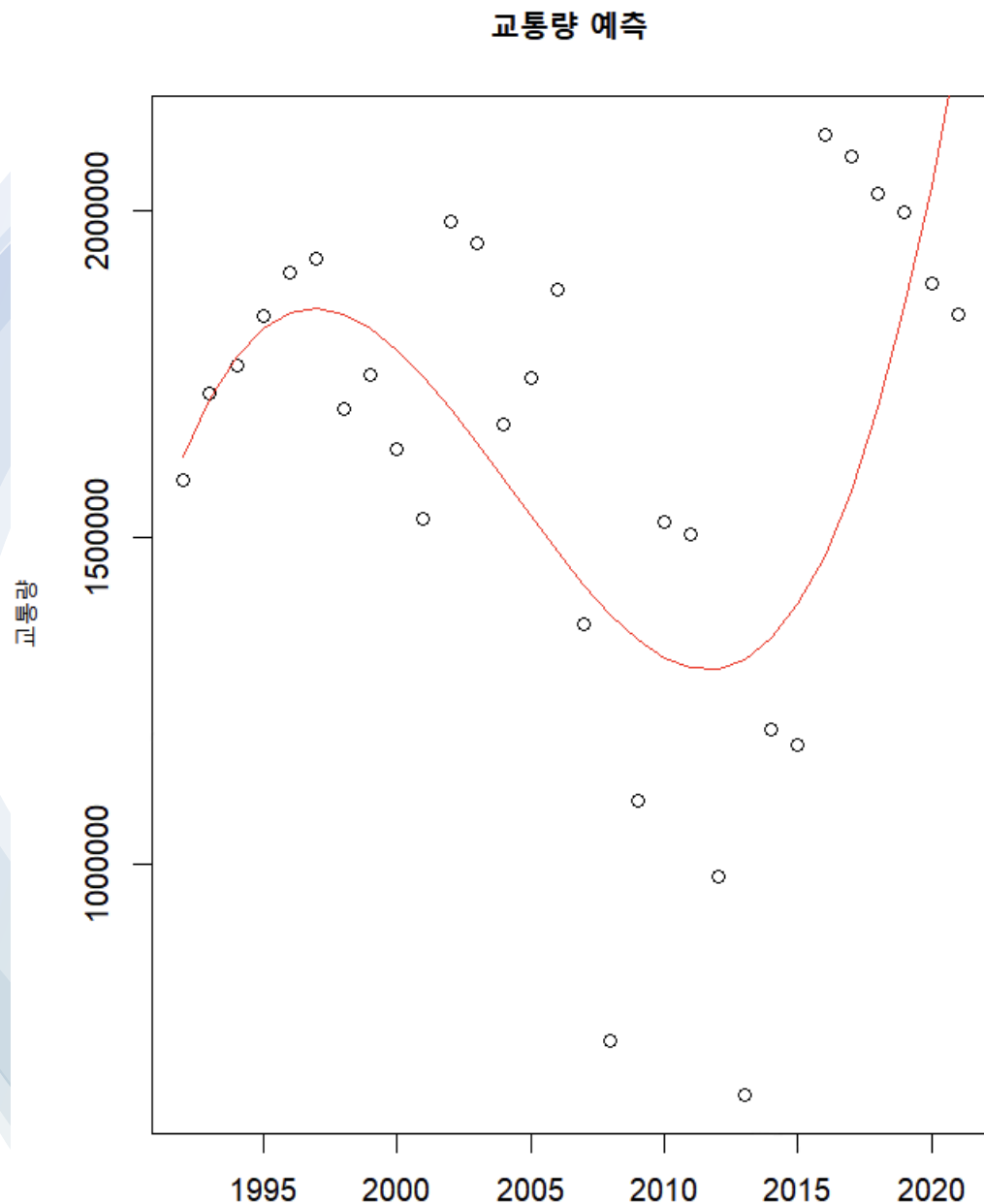
- 이 그래프는 각 도로명을 합해서 월별로 그래프를 그린 것이다.
- 그래프를 보면 3월 6월인
- 봄, 여름에 교통량이 가장 높은 것으로 나타나고 겨울인 12월 1월에는 교통량이 눈에 띄게 낮은 것을 확인할 수 있다.



서울 교통량 데이터 연도별 예측

- 이는 연도별에 따른 서울 교통량 데이터를 예측한 그래프이다. 실제 데이터를 가지고 종속변수는 교통량 독립변수는 연도로 설정한 뒤 lm함수를 가지고 3차 모델을 구축해서 모델 요약 통계를 출력하고 데이터의 산점도를 그렸다. 또한 lines함수를 사용하여 3차 모델 예측선을 그린 뒤 2025년에 대한 예측을 수행하였다.

- 최종 예측 값은 predicted value 변수에 저장했는데 예측 결과는 3498685로 예상된다.



서울 교통량 데이터 월별 예측

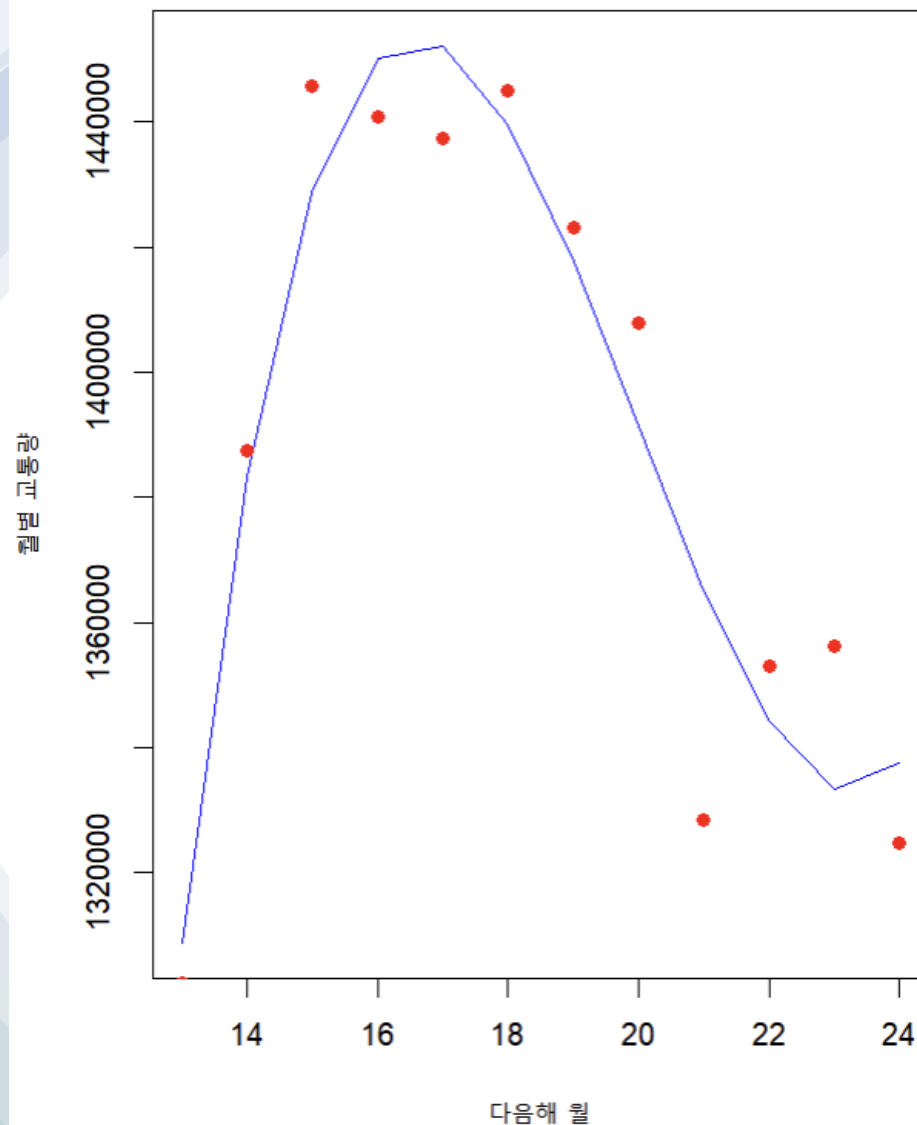
- 이는 월별에 따른 서울 교통량 데이터를 예측한 그래프이다. 종속변수는 교통량 독립변수는 월 단위로 설정한 뒤 1월부터 12월 까지의 실제 데이터는 빨간 점으로 표시하고 다음 해 1월부터 12월까지의 14월부터 24월로 설정하여 파란선으로 그래프를 나타냈다.

각 월별 예측 값을 predicted_data_3변수에 저장했다.

예측 결과값은 다음과 같다.

	Month	Predicted_3	total_sum
1	13	1308783	1302387
2	14	1383338	1387351
3	15	1428801	1445674
4	16	1450075	1440773
5	17	1452058	1437362
6	18	1439653	1445056
7	19	1417759	1423202
8	20	1391276	1407783
9	21	1365105	1328293
10	22	1344148	1352899
11	23	1333303	1356231
12	24	1337472	1324760

다음해 월별 교통량 예측 그래프



검정통계량에 따른 기각역, 기각값

1. 귀무가설(H_0): 서울의 교통량은 연도별로 차이가 없다.

대립가설(H_1): 서울의 교통량은 연도별로 차이가 있다.

H_0 기각, “서울의 교통량은 연도별로 차이가 없다.”를 기각 시키고

H_1 를 채택한다. 이는 회귀관계가 성립된다는 것으로 볼 수 있다.

2. 귀무가설(H_0): 서울의 교통량은 장소별로 차이가 없다.

대립가설(H_1): 서울의 교통량은 장소별로 차이가 있다.

H_0 기각, “서울의 교통량은 장소별로 차이가 없다.”를 기각 시키고

H_1 를 채택한다. 서울 장소별에 따라 데이터 분석을 한 결과 올림픽대로가 다른 곳에 비해 눈에 띄게 높다는 것을 볼 수 있었다. 이는 회귀관계가 성립된다는 것을 의미한다.

3. 귀무가설(H_0): 서울의 교통량은 월별로 따라 차이가 없다.

대립가설(H_1): 서울의 교통량은 월별로 따라 차이가 있다.

H_0 기각, “서울의 교통량은 월별로 차이가 없다.”를 기각 시키고

H_1 를 채택한다. 서울 월별로 따라 데이터 분석을 한 결과 봄과 여름에 교통량이 가장 높은 것으로 나타났고 12월 1월인 겨울에는 교통량이 현저하게 낮은 것을 확인할 수 있었다. 이는 회귀관계가 성립된다는 것을 의미한다.

