

1. 단순선형회귀모형
2. 전진 선택법&단순선형회귀모형
3. 후진 선택법&단순선형회귀모형
4. 이차항, 교호작용을 고려한 회귀모형
5. 전진선택법& 이차항, 교호작용을 고려한 회귀모형
6. 후진선택법& 이차항, 교호작용을 고려한 회귀모형
7. Random Forest 모형

데이터를 7:3으로 랜덤하게 분할하여 Train set, Test set으로 활용

5번모형(전진 선택법) 적합 시 회귀 계수

(Intercept)	거리_특화	평균_경사	near
68.716027	-10.582366	-21.304954	-6.707819
유동20대	지하철_승:평균_경사	거리_문화:거리_관광	거리_관광:평균_경사
12.053424	-9.507117	-5.418402	7.921523
near:유동50대			
-5.934406			

6번모형(후진 선택법) 적합 시 회귀 계수

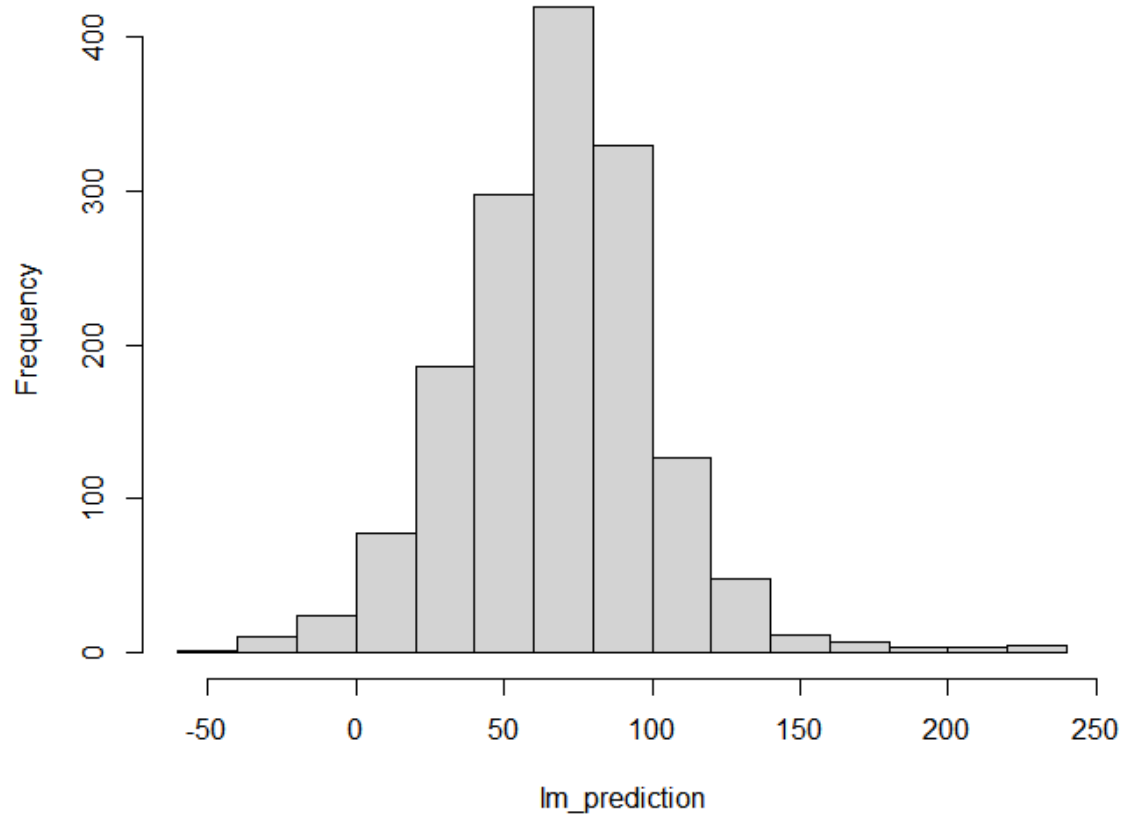
(Intercept)	거리_지하	평균_경사	near
62.285576	-11.652408	-16.347711	-9.183376
유동10대	지하철_승:거리_지하	거리_문화:거리_관광	거리_문화:거리_의료
9.095074	-9.775139	-7.096790	7.661768
거리_문화:주거_10대			
9.366974			

모형별 Test MSE 비교

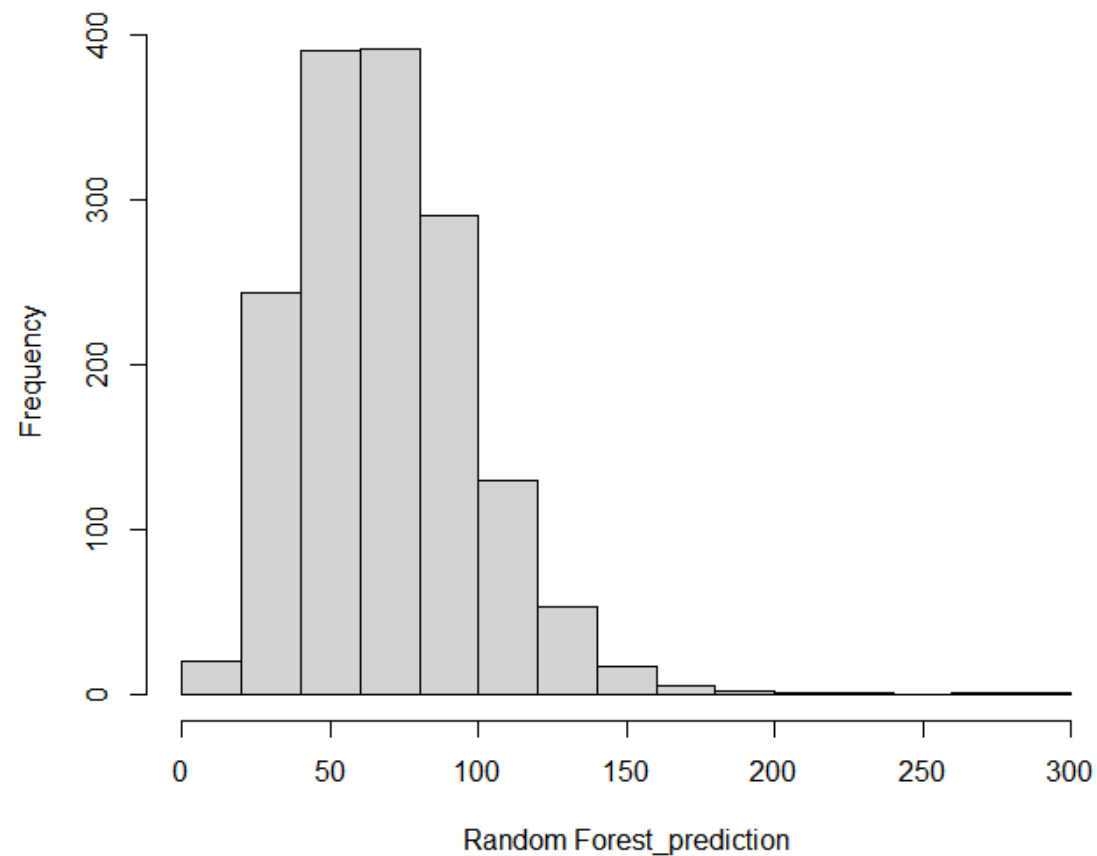
단순선형회귀	1618.166	이차선형회귀	12793.646	Tree	2178.579
단순선형회귀 전진선택법	1690.418	이차선형회귀 전진선택법	1821.916	Random Forest	1506.85
단순선형회귀 후진선택법	1821.432	이차선형회귀 후진선택법	1862.540		

선형회귀모형 이용 시 문제점과 Random Forest 모형

중회귀모형 예측값 분포

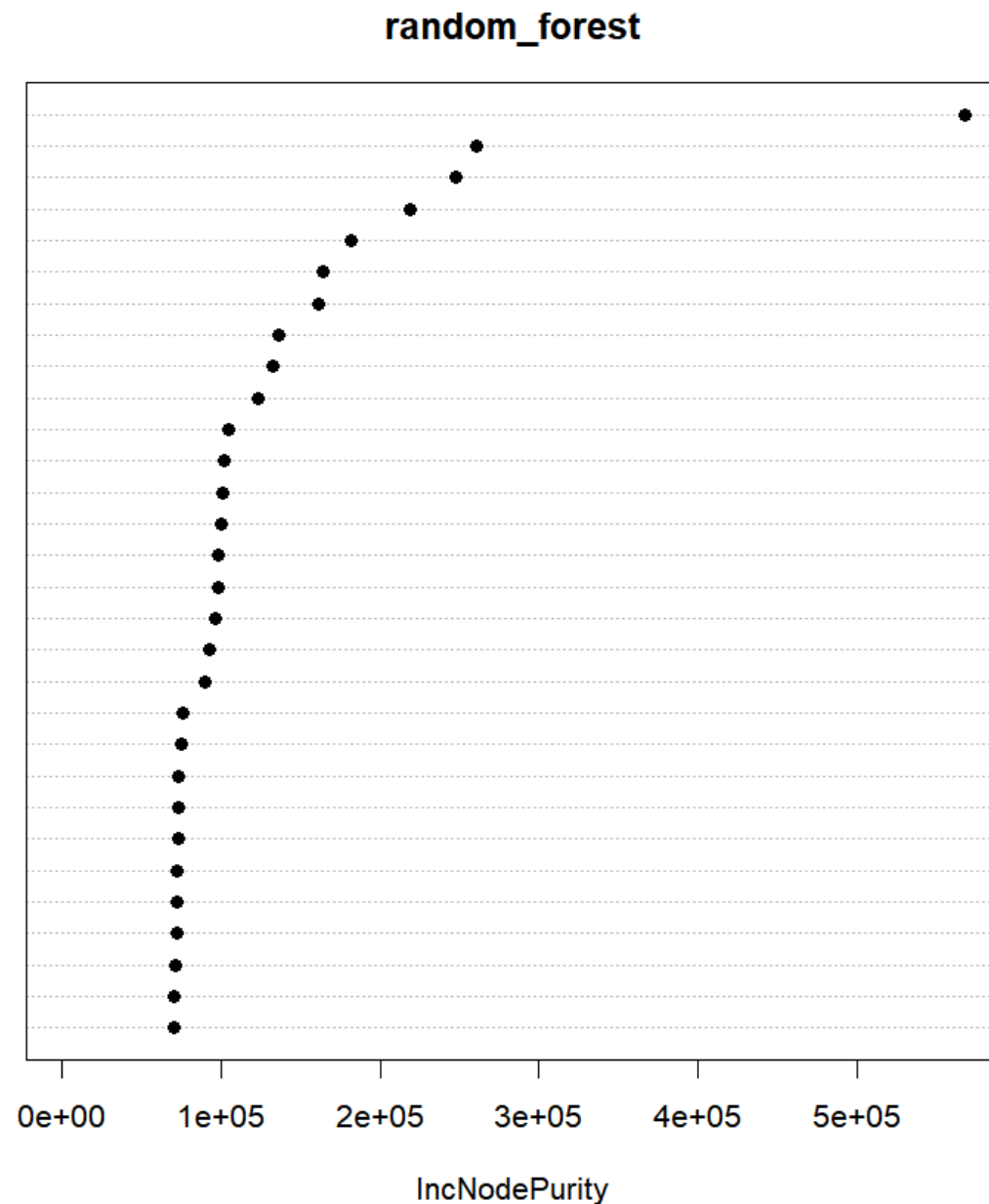


Random Forest 예측값 분포

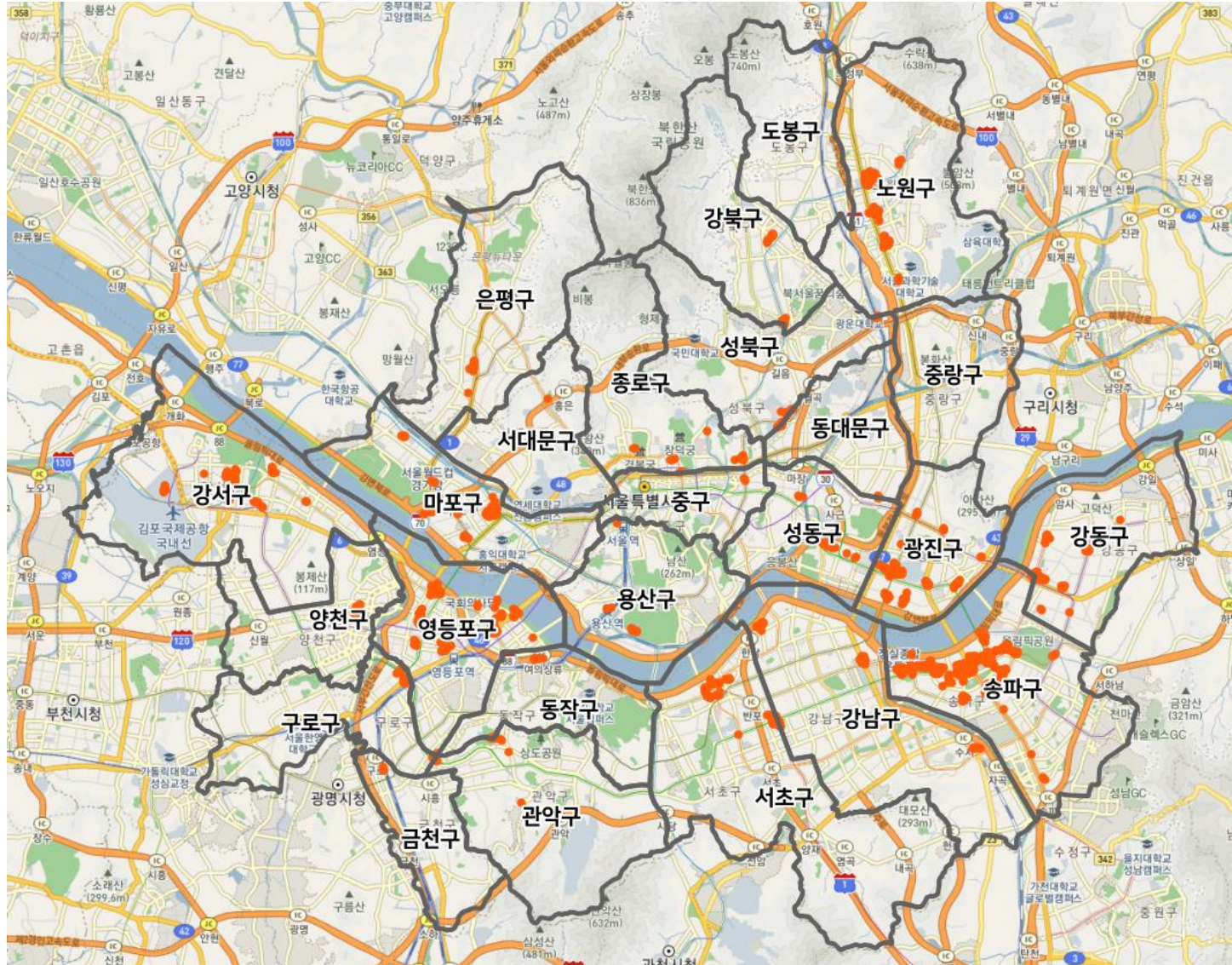


Random Forest 모형 이용 시, 변수별 중요도

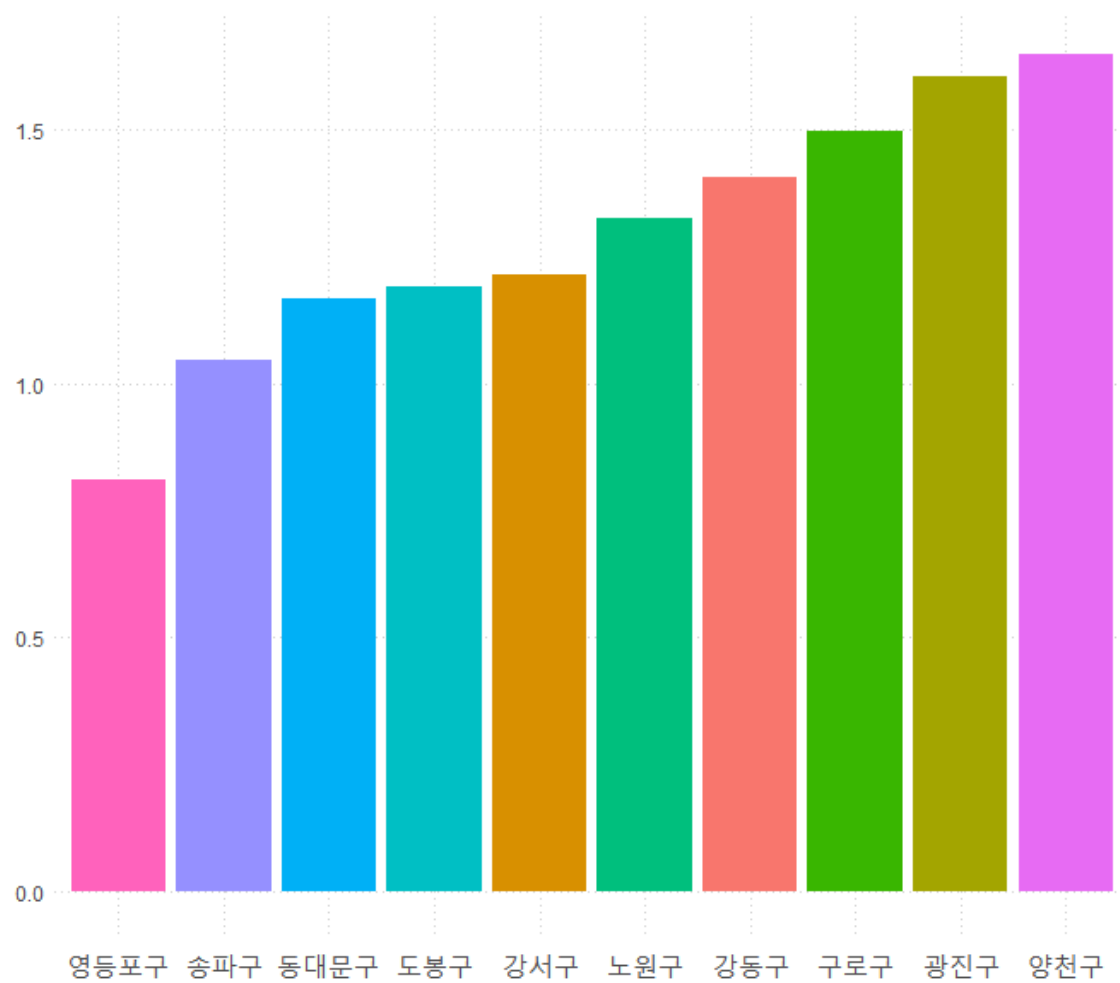
평균_경사
 지하철_승
 유동20대
 거리_지하
 거리_관광
 버스_승객
 유동30대
 거리_하천
 거리_자전
 유동10대
 거리_특화
 거리_공원
 거리_문화
 거리_영화
 주거_50대
 거리_상업
 주거_40대
 거리_대학
 유동40대
 거리_체육
 유동10대미
 거리_공공
 거리_교통
 거리_의료
 유동70대이
 주거_20대
 유동60대
 유동50대
 거리_주차
 near



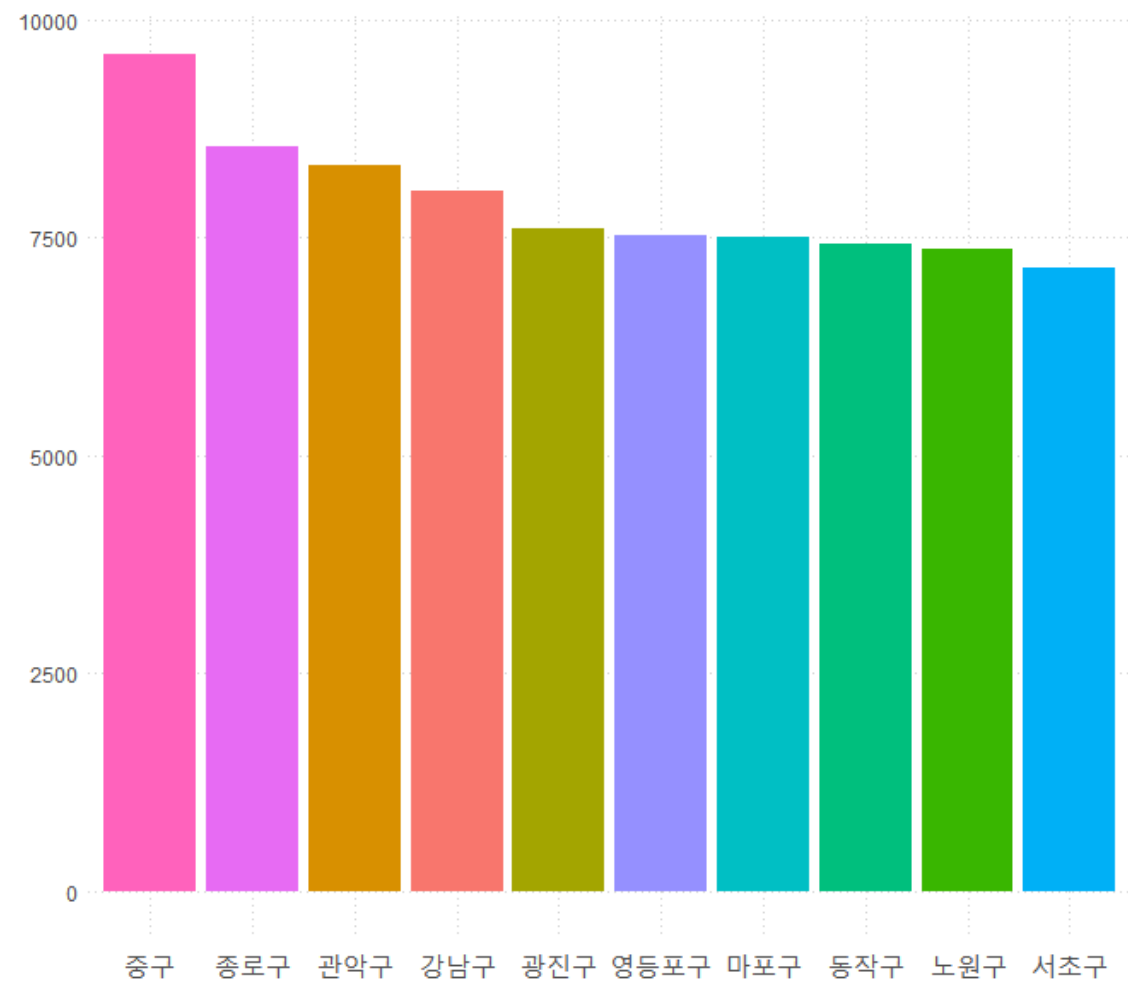
총 이용량 예측값 중 상위 5% (122.67명 이상)



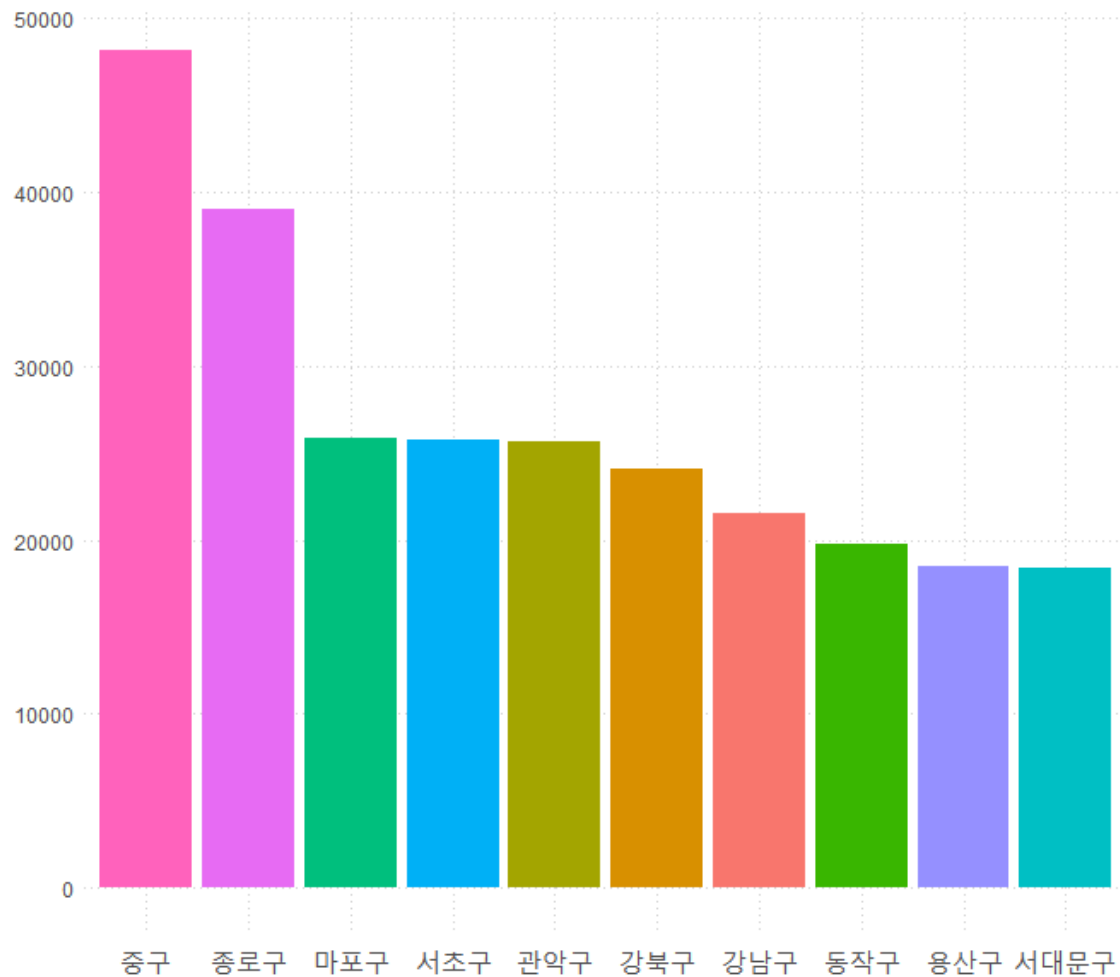
일정 반경 내의 평균 경사도



일정 반경 내의 30대 이하 유동인구 수



일정 반경 내의 평균 대중교통 이용자 수



이전에 진행했던 기초통계분석과
직관에 어느정도 들어맞는 결과가
도출되었다고 할 수 있다.