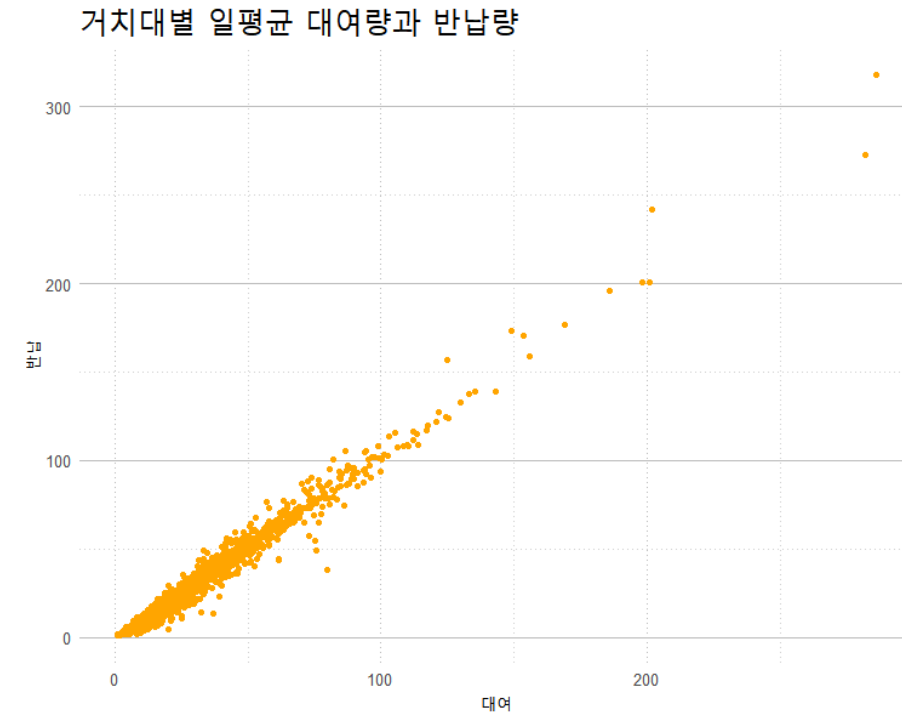


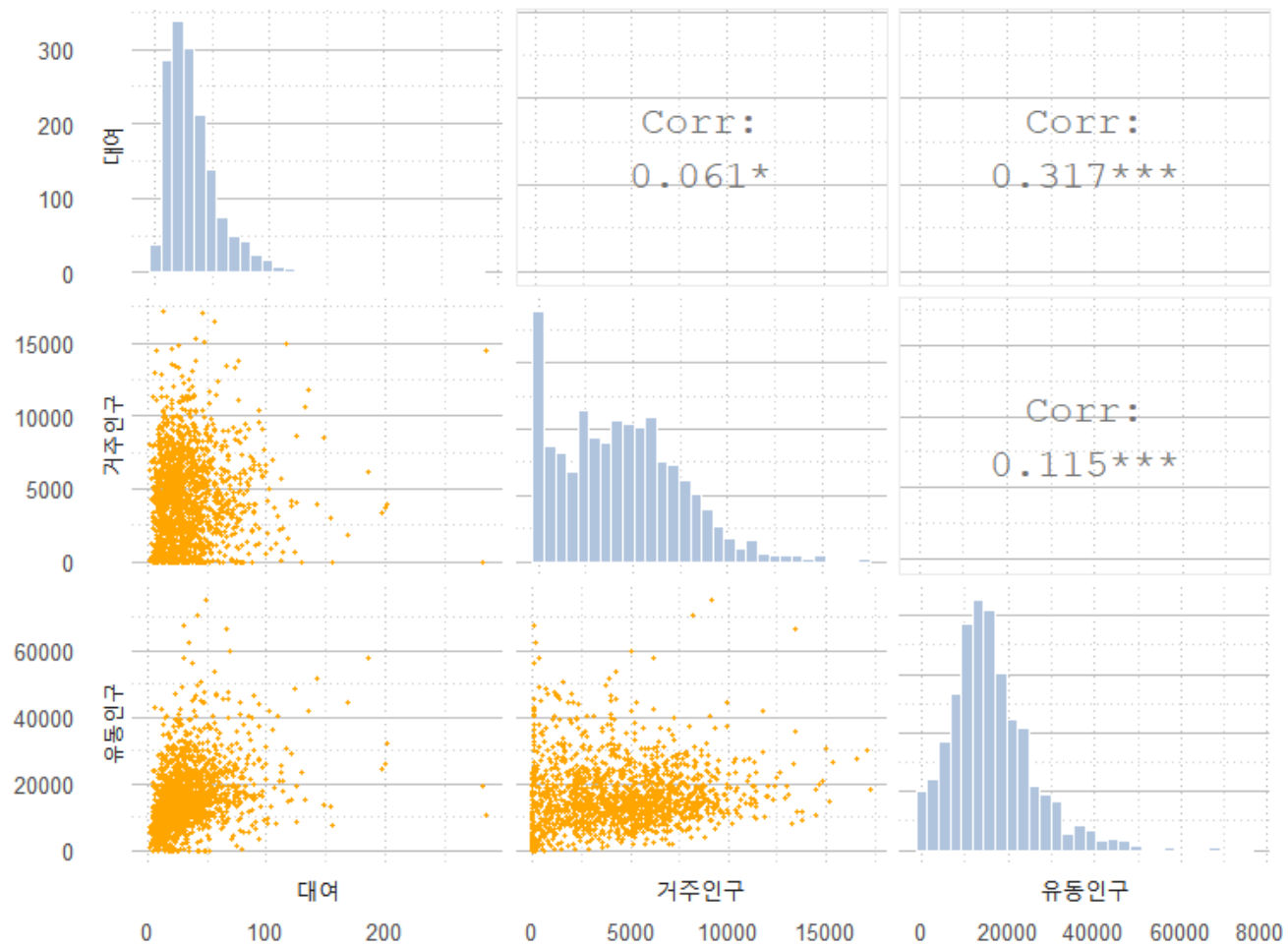
제공받은 데이터에는 14개의 거치대 주변 인구 관련 변수가 존재한다.
(6개의 변수는 연령별 거주 인구, 8개의 변수는 연령별 유동인구)

본 EDA에서는 인구 관련 변수 간에 유의미한 상관성이 존재하는지를 알아보고,
주성분분석을 진행하였다.

또한 일평균 대여량과 반납량은 비례하는 경향이 나타났기 때문에,
대여량만을 고려하여 거치대 이용량에 대한 분석을 진행하였다.



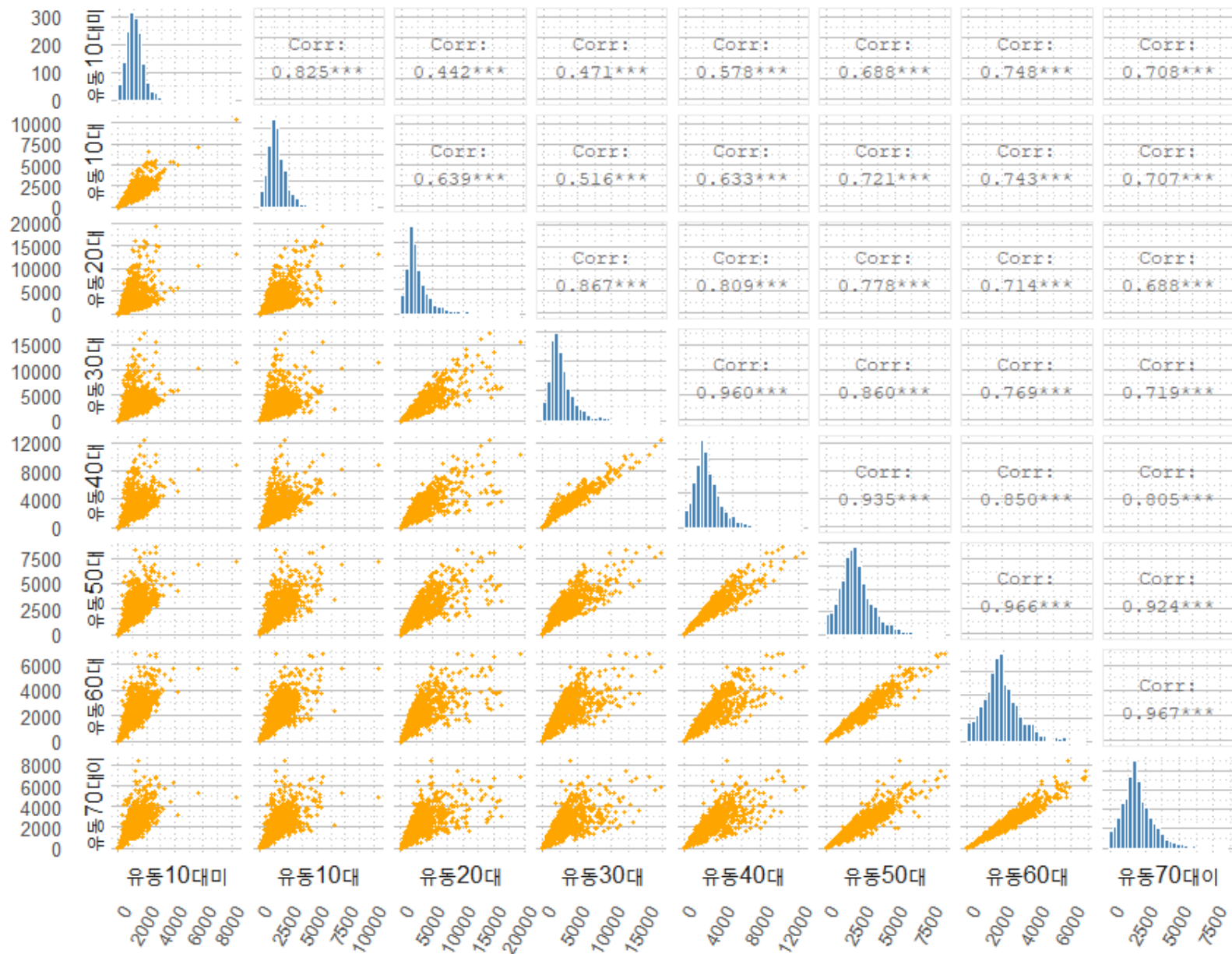
거치대 이용량과 주변 거주인구, 유동인구



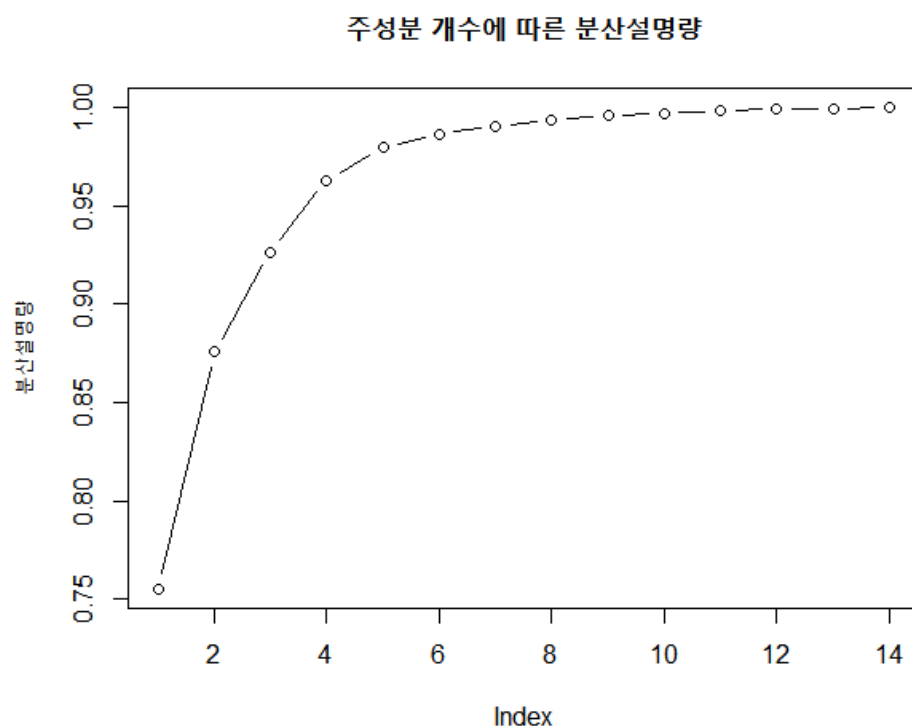
거치대 이용량은 특히
거치대 근방 유동인구와 연관성이 높다.

또한 거주인구와 유동인구 사이의 유의미한
상관계수가 나타나는 것을 확인할 수 있다.

연령대별 거치대 주변 유동인구

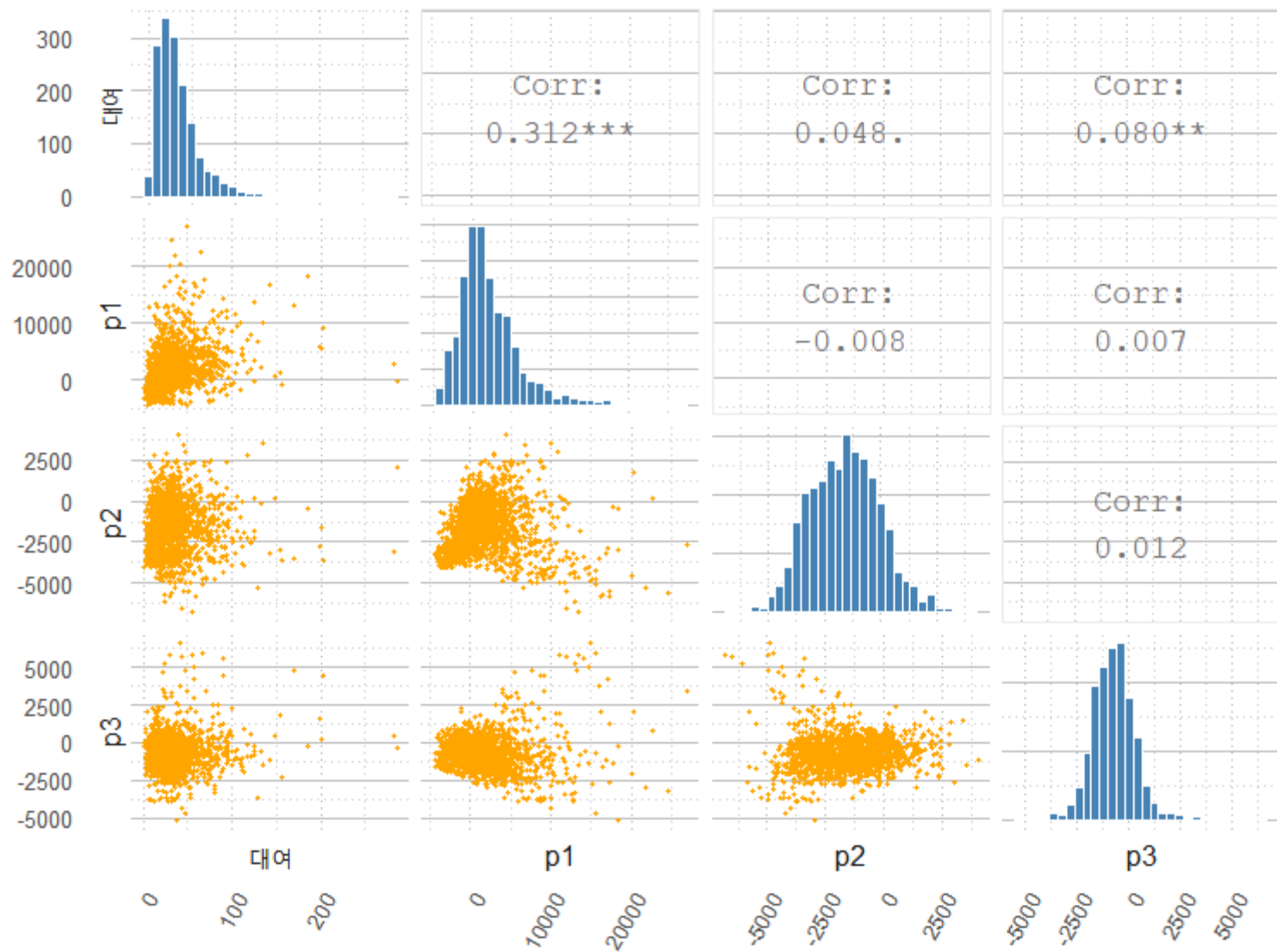


거치대 인근 연령별 거주인구와 유동인구에 대한 변수들은 총 14개이고, 변수간 유의미한 상관성이 나타났다. 따라서 차원 축소와 모형 적합에서 발생할 수 있는 다중 공선성문제 해결을 위해, 주성분 분석을 진행하였다.



	p1	p2	p3
주거_10대	0.001562428	0.44897248	0.153052998
주거_20대	0.013927169	0.22716698	0.132587908
주거_30대	0.008570597	0.32425066	0.127905861
주거_40대	0.003180565	0.33753733	0.112455051
주거_50대	0.006590821	0.27364106	0.089785164
주거_60대_	0.006423153	0.24629932	0.077172217
유동10대미	0.097001911	0.21870242	0.001647691
유동10대	0.166276984	0.23166273	0.243022322
유동20대	0.568731959	-0.29989212	0.720149075
유동30대	0.514765366	-0.18701508	-0.325108915
유동40대	0.394049276	0.02871575	-0.381651821
유동50대	0.305814190	0.18568409	-0.217034407
유동60대	0.245056901	0.24815685	-0.140764377
유동70대이	0.255516069	0.26251260	-0.141501008

인구 관련 변수들에 대한 주성분



주성분 분석 이전과 이후의 모형 비교

Call:

```
lm(formula = 대여 ~ . - 반납, data = age_notna)
```

Coefficients:

(Intercept)	주거_10대	주거_20대	주거_30대	주거_40대	주거_50대
21.0579943	0.0031348	0.0066307	0.0048798	-0.0025872	-0.0057382
주거_60대_	유동10대미	유동10대	유동20대	유동30대	유동40대
-0.0076451	0.0006841	0.0073543	-0.0004906	0.0093756	-0.0176699
유동50대	유동60대	유동70대이			
0.0196638	-0.0192449	0.0075538			

$$\text{adj } R^2 = 0.135$$

> vif(fit2)

주거_10대	주거_20대	주거_30대	주거_40대	주거_50대	주거_60대_	유동10대미
6.422413	11.380615	10.252485	17.005026	24.501002	9.101681	5.980942
유동10대	유동20대	유동30대	유동40대	유동50대	유동60대	유동70대이
11.062720	14.668422	63.660398	82.765698	73.682511	69.188789	22.246297

Call:

```
lm(formula = 대여 ~ p1 + p2 + p3, data = pca_r)
```

Coefficients:

(Intercept)	p1	p2	p3
3.224e+01	2.128e-03	8.395e-04	1.724e-03

$$\text{adj } R^2 = 0.1043$$

> vif(fit1)

p1	p2	p3
1.000109	1.000215	1.000195