

# K-means Cluster 분석을 통한 가천대학교 학생의 소비성향 파악

DATE

2018 . 11 . 16

1조 | 분석주 김기현

# 목차



01

조사설계

1.1 주제선정  
1.2 설문지 작성

02

데이터 소개

2.1 변수 설명  
2.2 데이터 탐색

03

데이터 분석

3.1 K-means cluster 분석


04

결과

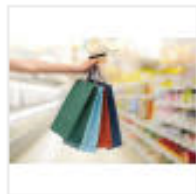


# 01

## 조사설계

- 1.1 주제 선정
  - 1.2 설문지 작성
- 

## 1.1 주제 선정



[소비도 문화다] **“시발비용”**으로 탕진하면 개꿀

머니S | 2018.07.09. | 네이버뉴스 | [🔗](#)

‘시발비용’이 젊은 세대의 새로운 소비문화로 자리  
이 신조어는 ‘스트레스를 받지 않았으면 발생하지 않았을 미중을 뜻한다. 유쾌과 비애가 외  
사에서 받은...



1인 가구 소득은 줄어들고 **“시발비용”**이 늘어나고 베타뉴스 | 2017.11.28. | [🔗](#)

더욱 씹쓸한 사실은 근로소득은 나날이 줄어가지만 물가는 오르고, ‘시발비용’이라는 신조  
어까지 생겨났다는 것이다. **시발 비용**은 비속어인 ‘시발’과 ‘비용’을 합친 단어로 ‘스트레스  
를 받지 않았으면 발생하지...

▶ 1인 가구 소득 감소 이유는?...물가와... 헤럴드경제 | 2017.11.28. | 네이버뉴스



‘**황김비용**’ ‘**쓸쓸비용**’ **“탕진잼”** 아시나요?...성인 10명 중 9명 “감정소...

캠퍼스잡앤조이 | 2018.06.08. | [🔗](#)

재미 ‘**탕진잼**’ 등 소비를 통해 심리적 안정을 얻는 감정소비 관련 신조어가 유행이다. 이는  
많은 사람들이 여전히 일상의 스트레스나 우울함 등을 소비로 분출하고 있음을 의미한다. 버  
클랜드의...



[소확행] ③ **“탕진잼”**의 유래, 소확행 체험기

헤럴드경제 | 2018.03.29. | 네이버뉴스 | [🔗](#)

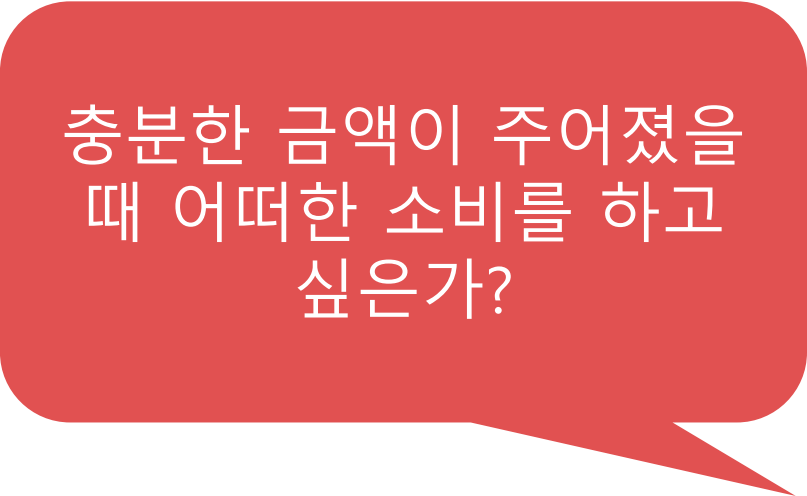
역시 ‘**탕진잼**’이다. 단기 채식을 위해 선택한 점심 샐러드 ■ 오후 1시, 광화문 샐러드가게에  
서 한끼 점심시간, 광화문 일대 식당은 주변 직장인들로 넘쳐났다. 골목으로 좀 더 들어가면  
있는 샐러드가게를 찾았다. 사실...



**“황김비용”** **“명청비용”** 우리의 소비 형태는? YTN사이언스 | 2017.07.26. | [🔗](#)

■ 이동기 / 연세대 심리학과 교수 [앵커] 혹시 ‘**황김비용**’이라는 말 들어보셨나요, 스트레  
스를 풀기 위해서 황김에 쓰는 돈을 말하는데요, 이 밖에도 ‘쓸쓸비용’ ‘명청비용’ 등 다양한  
신조어들이 등장하고 있습니다. 이런...

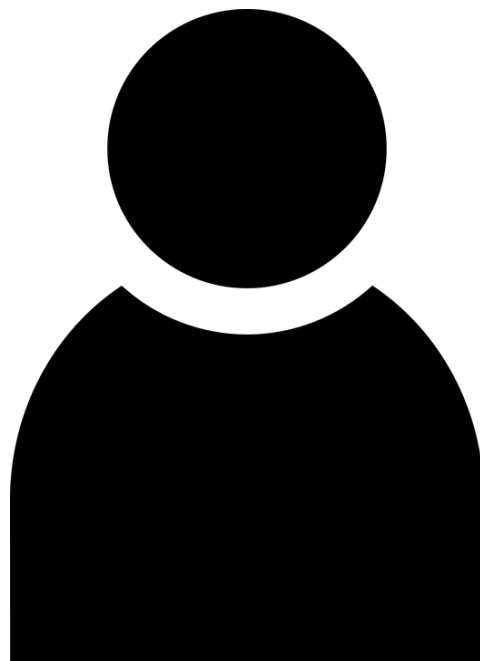
## 1.1 주제 선정



충분한 금액이 주어졌을  
때 어떠한 소비를 하고  
싶은가?



현재 자신이 원하는  
소비를 하고 있는가?



## 1.2 설문지 작성



- 조사기간 : 10월 31일 ~ 11월 6일
- 모집단의 크기 :  $N=25,000$
- 신뢰수준 : 95%
- 표본 오차 : 5%

=> 약 380개의 표본이 필요

## 1.2 설문지 작성

### 인구학적 특성 문항

1. 귀하의 성별은 무엇입니까?  
① 남자(다음 2번 문항)      ② 여자(다음 3번 문항)

### 소비 성향 관련 문항

2. 귀하는 병역의 의무를 수행하십니까?  
① 예      ② 아니요

3. 귀하의 나이는 어떻게 되십니까?  
만 \_\_\_\_\_ 세

### 소비 출처 관련 문항

4. 귀하의 학년은 무엇입니까?  
① 1학년      ② 2학년      ③ 3학년      ④ 4학년

### 소비 평가 관련 문항

5. 귀하는 현재 연애했습니까?  
① 예      ② 아니요
6. 귀하는 현재 어디서 거주 중이십니까?  
① 집      ② 기숙사(학사 포함)      ③ 자취(하숙, 쉼터 하우스 포함)

## 1.2 설문지 작성

인구학적  
특성 문항

소비 성향  
관련 문항

소비 출처  
관련 문항

소비 평가  
관련 문항

7. 이 설문지를 통해 소비성향의 변화를 알아보는 항목입니다. 첫 번째 열에는 현재의(평균적으로) 한 달 지출비를 적어 주시고, 두 번째 열에는 가상으로 한 달 지출비가 2배가 된 경우 한 달 지출비를 어떻게 사용할 것인지 적어 주십시오. (지출항목은 자유롭게 기입하되, 변화의 총합은 현재 지출액에 2배가 되게 기입해 주십시오.)

지출항목	현재 한 달 지출내역	현재 한 달 지출내역의 총합이 2배가 되었을 경우의 변화
식비 (ex 카페, 간식, 식당 등)	만원	만원
술/유흥 (ex pc방, 노래방, 담배 등)	만원	만원
패션/쇼핑 (ex 옷, 신발, 화장품 등)	만원	만원
문화/취미 (ex 영화, 공연, 취미활동 등)	만원	만원
주거/교통 (ex 집세, 렌트비, 통근비 등)	만원	만원
교육/학술 (ex 자습, 학원, 도서관 등)	만원	만원
의료/건강 (ex 진료비, 약비, 건강식품, 약국 등)	만원	만원
생활 (ex 교통비, 통신비 등 (월세 제외))	만원	만원
저축 (ex 펀드, 적금 등)	만원	만원
총합	만원	(한 달 지출내역)

(ex 현재 한 달 지출내역의 총합이 50만원일 경우, 2배가 되었을 경우의 총합이 100만원 기입)



## 1.2 설문지 작성

인구학적  
특성 문항

소비 성향  
관련 문항

소비 출처  
관련 문항

소비 평가  
관련 문항

8. 귀하의 현재 한달 지출비의 출처는 어떻게 되십니까?

① 용돈과 아르바이트

② 아르바이트만

③ 용돈만

④ 기타 (항목 8-1에 응답해 주세요.)

(8-1) 기타 항목에 대한 출처를 자세히 적어주십시오.

## 1.2 설문지 작성

인구학적  
특성 문항

소비 성향  
관련 문항

소비 출처  
관련 문항

소비 평가  
관련 문항

9. 현재 귀하의 가장 문제가 심각하다고 생각되는 지출항목은 무엇입니까? (중복선택 가능)

- ① 식비    ② 술/유류    ③ 패션/쇼핑    ④ 문화/여가    ⑤ 뷰티/미용  
⑥ 교육/학습    ⑦ 의료/건강    ⑧ 생활    ⑨ 저축    ⑩ 해당 없음

10. 귀하는 현재 계획적인 소비를 하고 있다고 생각 하십니까?

	<----->									
즉흥적	①	②	③	④	⑤	⑥	⑦	⑧	⑨	계획적

11. 귀하는 한 달 지출비가 가상으로 2배가 되었을 경우 계획적인 소비를 한다고 생각 하십니까?

	<----->									
즉흥적	①	②	③	④	⑤	⑥	⑦	⑧	⑨	계획적

12. 귀하는 현재의 한 달 지출비가 충분하십니까?

	<-----충분----->									매우 충분
불충분	①	②	③	④	⑤	⑥	⑦	⑧	⑨	

13. 귀하는 현재의 한 달 지출내역에 대하여 만족하십니까?

	<-----만족----->									매우 만족
불만족	①	②	③	④	⑤	⑥	⑦	⑧	⑨	



## 02

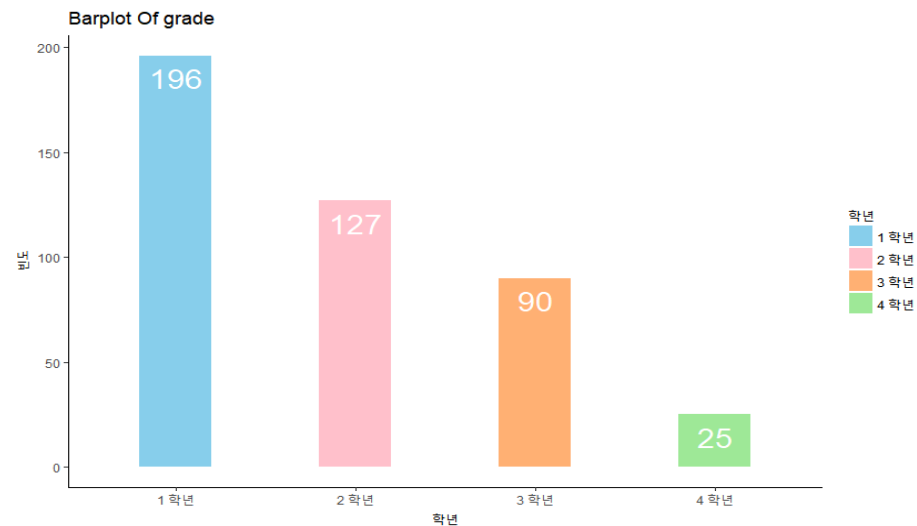
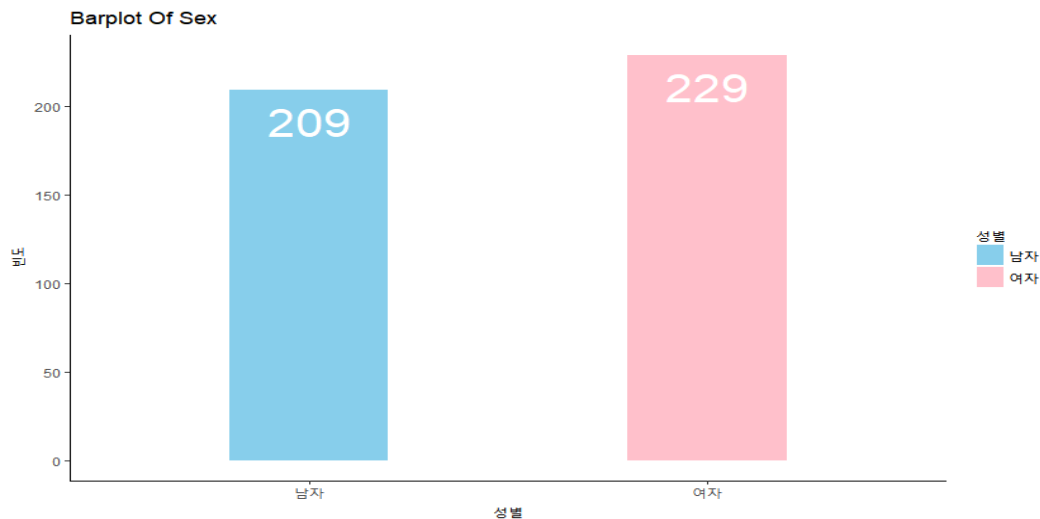
### 데이터 소개

- 2.1 변수 설명
- 2.2 데이터 탐색

## 2.1 변수 설명

성별	sex
나이	age
학년	grade
연애유무	couple
거주 유형	live
한 달 지출비	eat – total
한 달 지출내역의 2배가 되었을 때 지출비	eat2-total2
한 달 지출비의 출처	source
한 달 지출비들의 비율	eatratio-totalratio
한 달 지출비의 2배가 되었을 때 지출비의 비율	eatratio2-totalratio2
심각하다고 생각되는 지출항목	problem
계획적인 소비라고 생각하는 정도	plan
2배가 되었을 때 계획적인 소비를 할 것이라고 생각되는 정도	plan2
한 달 지출비가 충분하다고 생각되는 정도	satisfy
한 달 지출내역에 만족하는 정도	satisfy_contents

## 2.2 데이터 탐색



01

### 성비

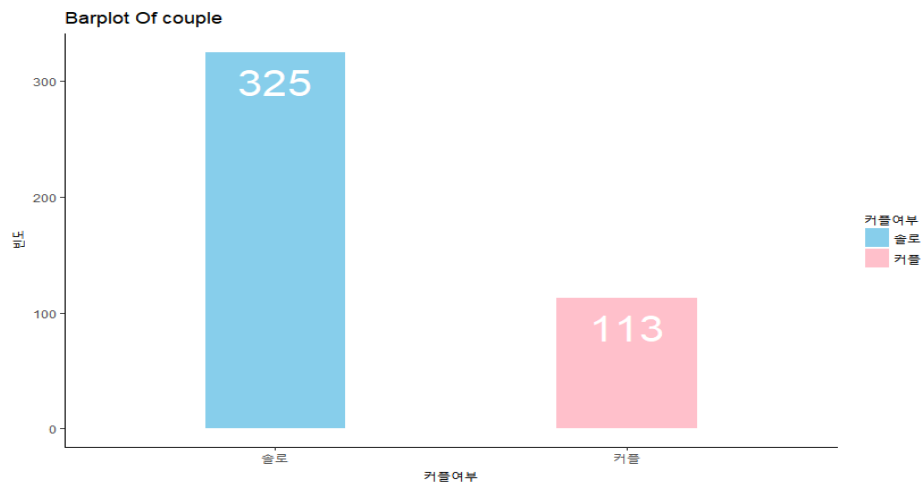
응답자 성비를 확인해 본 결과 남성이 48% 여성이 52%로 비슷한 비율을 가지고 있음을 알 수 있었다.

02

### 학년

응답자 수는 1학년, 2학년, 3학년, 4학년 순으로 나타났다.

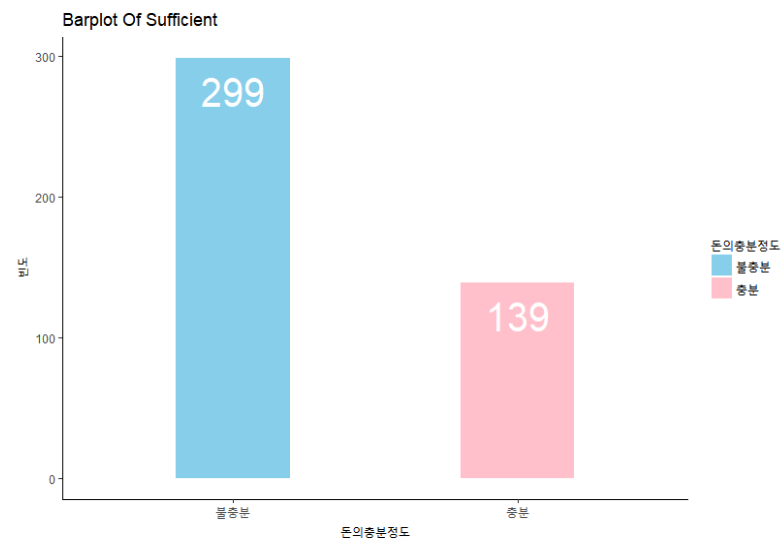
## 2.2 데이터 탐색



03

연애 유무

응답자 중 솔로가 74% 커플이 26%로 나타났다.

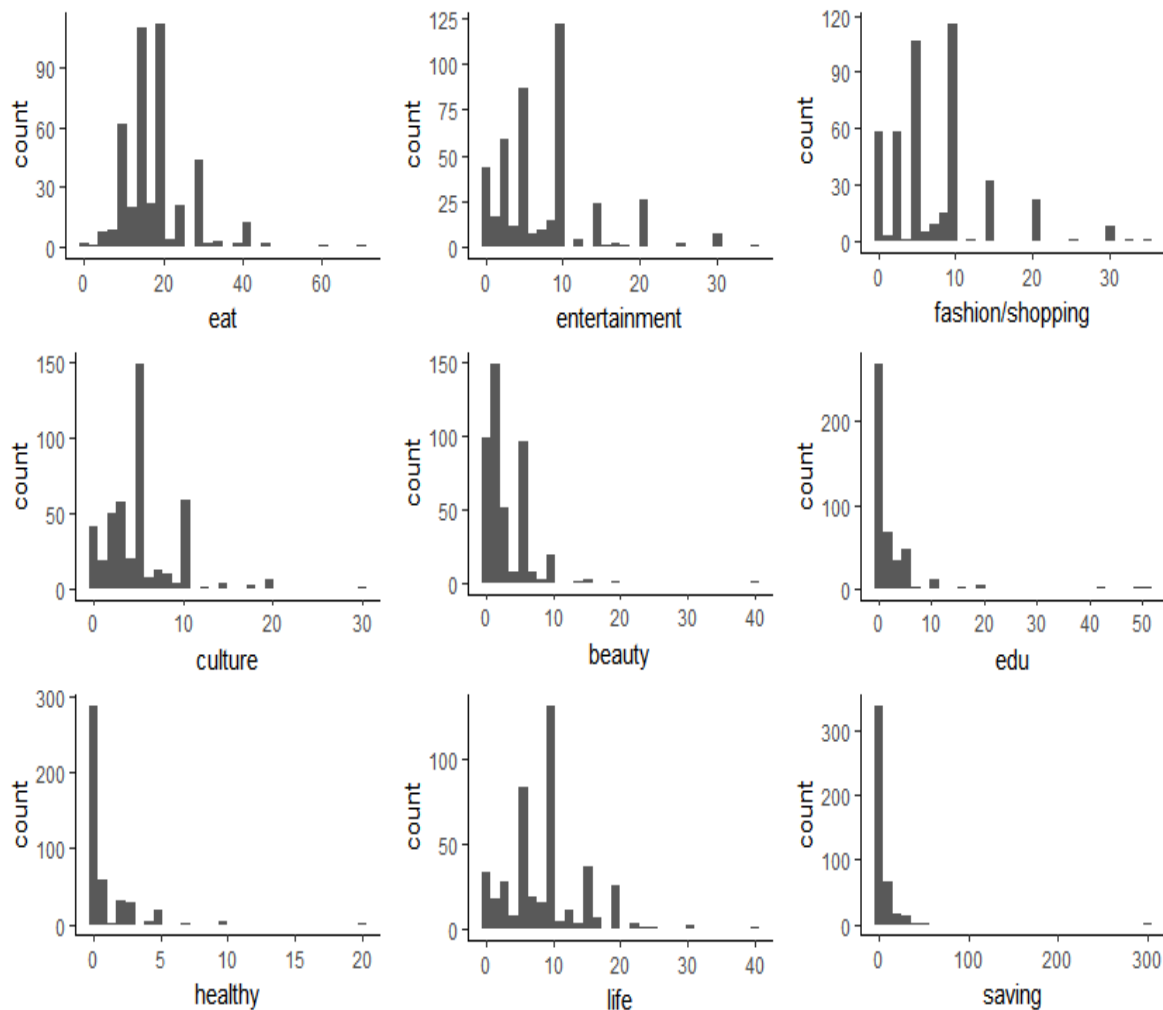


04

한 달 지출비 충분도

한 달 지출비가 불충분하다고 생각하는 응답자가 68% 충분하다고 생각하는 응답자가 32%로 나타났다.

## 2.2 데이터 탐색



05

한 달 지출비 분포와 요약

한 달 지출내역을 요약 해본 결과  
최소값 16, 1사분위수 44, 중위수  
55, 평균 58.35, 3사분위수 68 최대  
값 390으로 나타났다.

Min	1st Qu	Median	Mean	3rd.Qu	Max
16	44	55	58.35	68	390



# 03

## 데이터 분석

### 3.1 K-means cluster 분석





STORY  
**01**



---

# Why K-means Cluster Analysis ?



## 3.1 Why K-means Cluster Analysis ?



### 목적

- Clustering의 목적은 비슷한 개체끼리 한 그룹으로 묶는 것
- 군집 간 분산 최대화
- 군집 내 분산 최소화



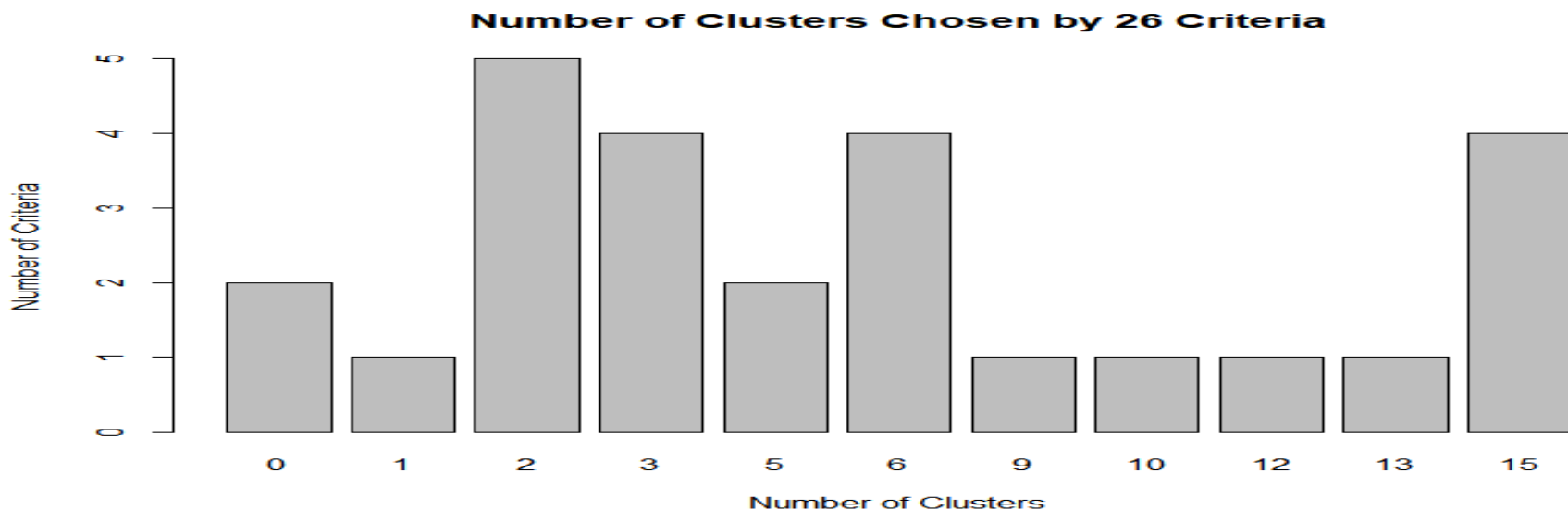
STORY  
02



현재 한 달 지출비를  
이용한 K-means Cluster 분  
석



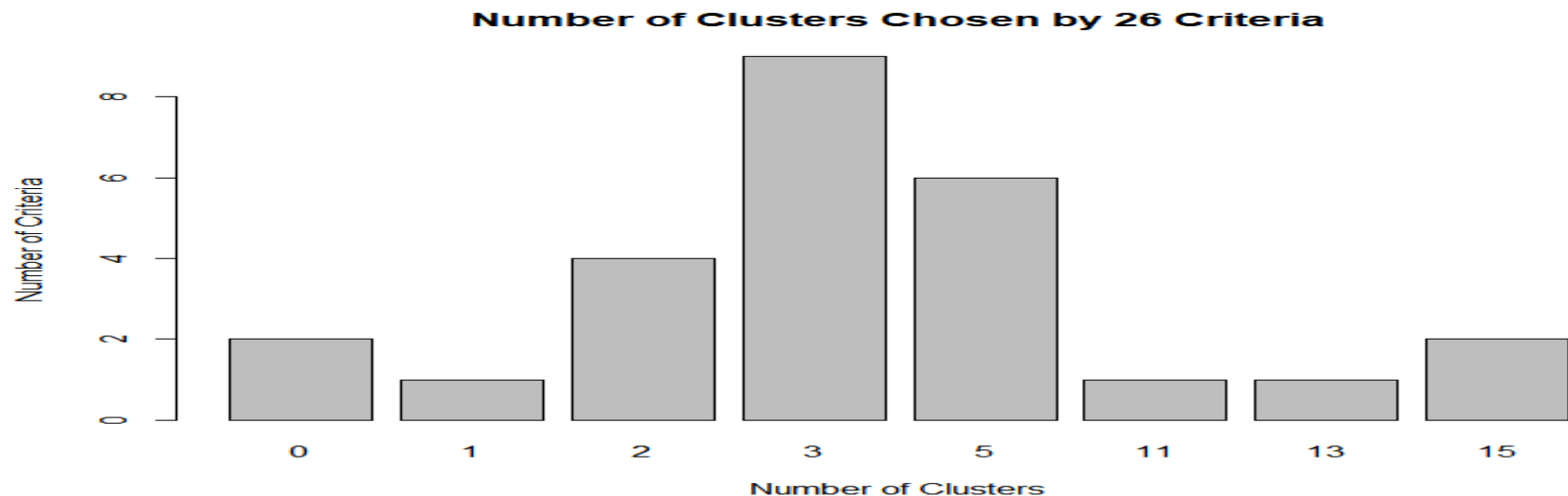
## 3.2 현재 한 달 지출비를 이용한 K-means Cluster분석



### Cluster들의 중심점

	eat	entertain ment	fashion/s hopping	culture	beauty	edu	health	life	saving
1	0.4669	0.1252	0.1003	0.0840	0.0432	0.0218	0.0151	0.1250	0.0210
2	0.2631	0.1575	0.1598	0.0973	0.0572	0.0295	0.0142	0.1885	0.0324
3	0.2528	0.0862	0.0921	0.0641	0.0383	0.0338	0.0113	0.1522	0.2688

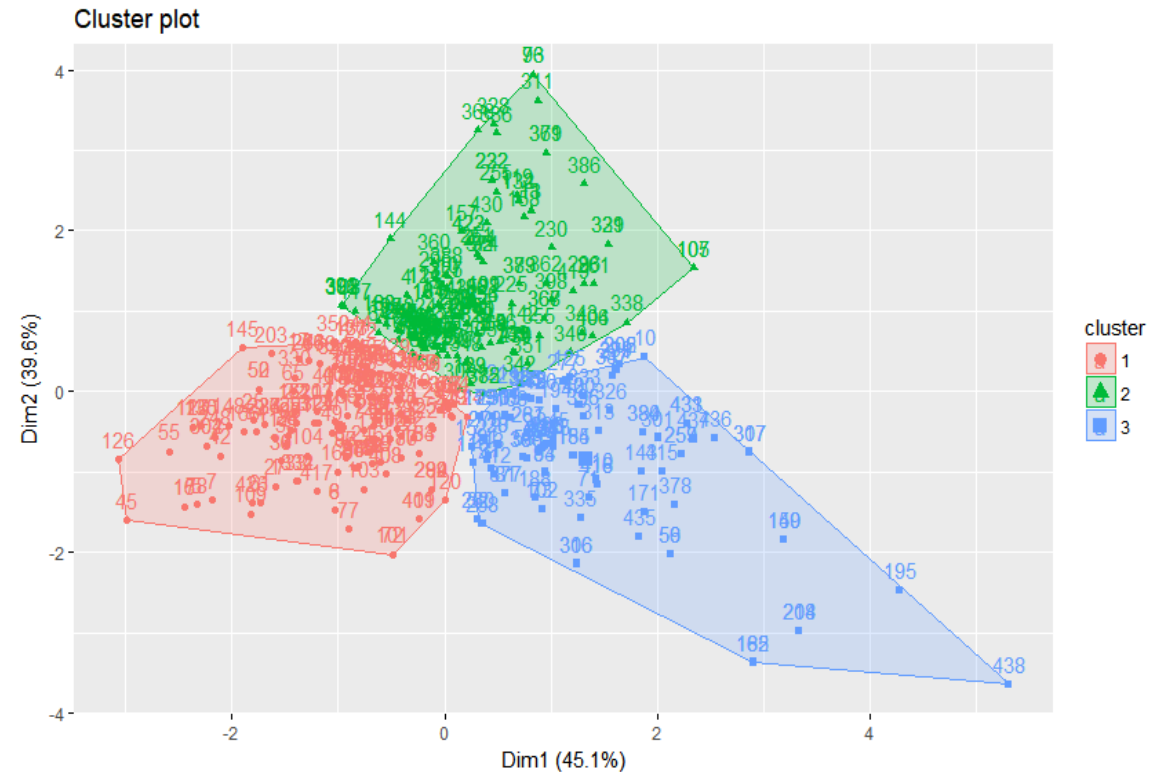
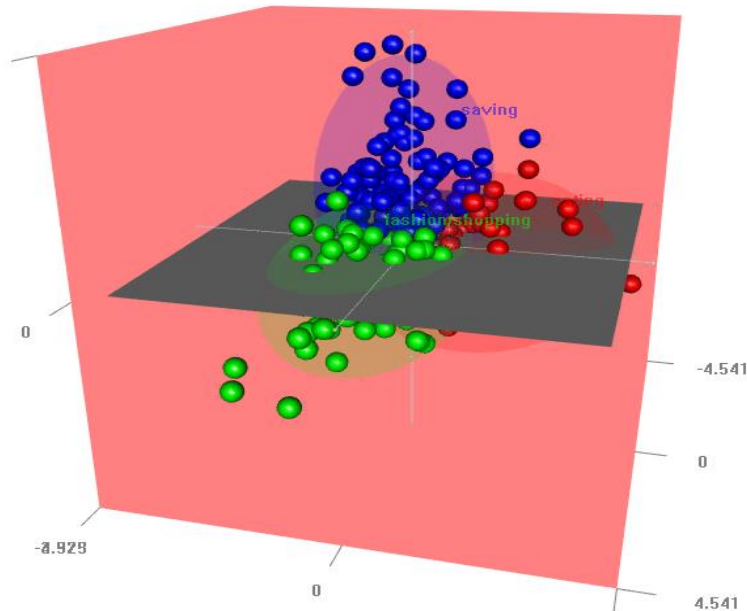
## 3.2 현재 한 달 지출비를 이용한 K-means Cluster분석



### Cluster들의 중심점

	<i>eat</i>	<i>fashion/shopping</i>	<i>saving</i>
1	0.4641	0.0850	0.0206
2	0.2698	0.1774	0.0275
3	0.2444	0.0896	0.2531

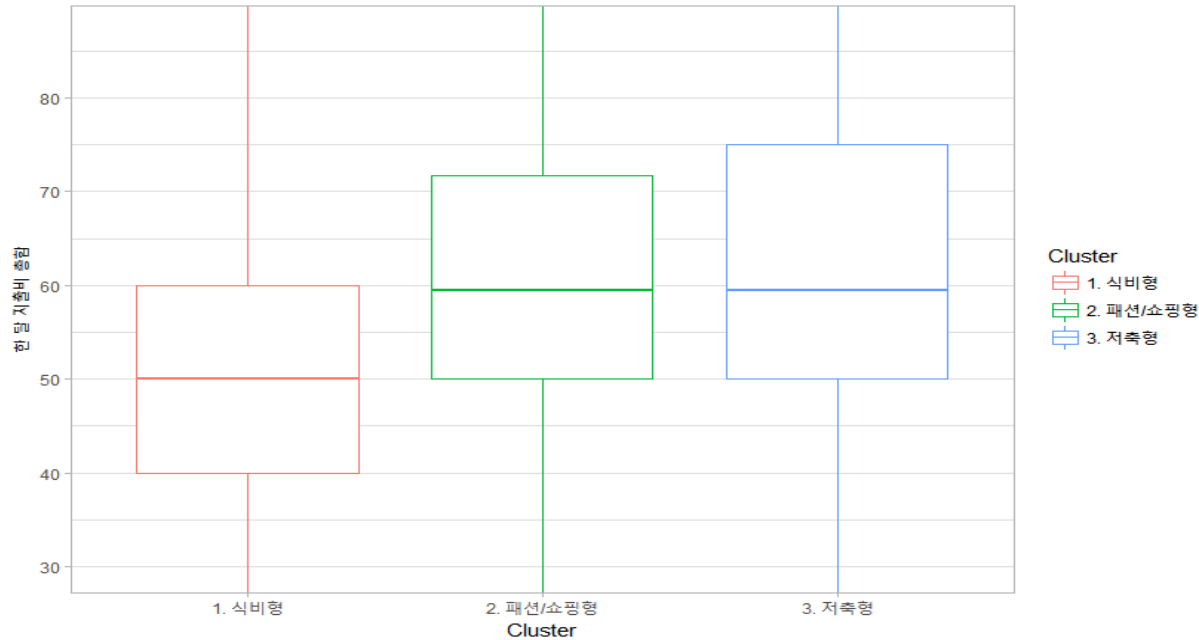
## 3.2 현재 한 달 지출비를 이용한 K-means Cluster분석



### Cluster의 요약

Cluster	Cluster 정의	표본 수
1	식비형	161(37%)
2	패션/쇼핑형	190(43%)
3	저축형	87(20%)

## 3.2 현재 한 달 지출비를 이용한 K-means Cluster분석



- 평균 한 달 지출비 : 58.35 만원
- 식비형 그룹 : 49.69 만원
- 패션/쇼핑형 그룹 : 61.33 만원
- 저축형 그룹 : 67.85 만원

### Cluster 한 달 지출비 평균 비교 (ANOVA 분석)

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Cluster	2	21596	10798.0000	16.0900	1.91e-07 ***
Residual	435	291883	671		



STORY  
**03**



**한 달 지출비가 2배가 되었을  
때 K-means Cluster 분석**





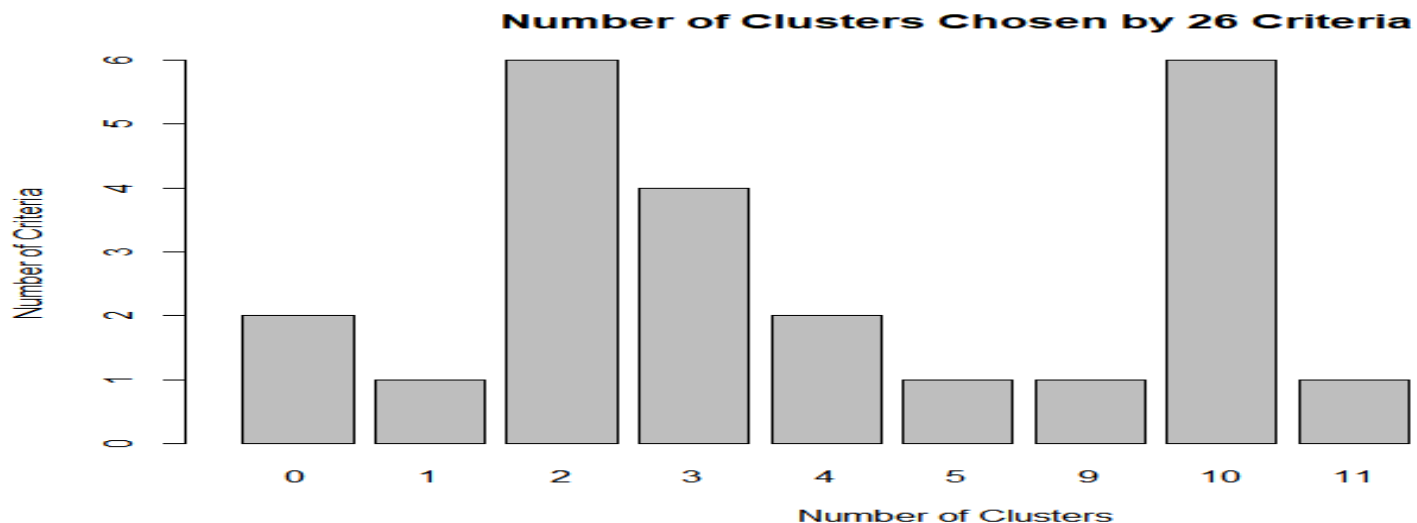
### 3.3 한 달 지출비가 2배가 되었을 때 K-means Cluster분석



#### Cluster들의 중심점

	eat2	entertain ment2	fashion/sho pping2	culture 2	beauty2	edu2	health2	life2	saving2
1	0.3393	0.1502	0.1259	0.0958	0.0533	0.0308	0.0137	0.1132	0.0772
2	0.2045	0.0907	0.2499	0.1221	0.0856	0.0407	0.0184	0.0943	0.0934
3	0.2047	0.0875	0.1129	0.0700	0.0434	0.0251	0.0124	0.0918	0.3518

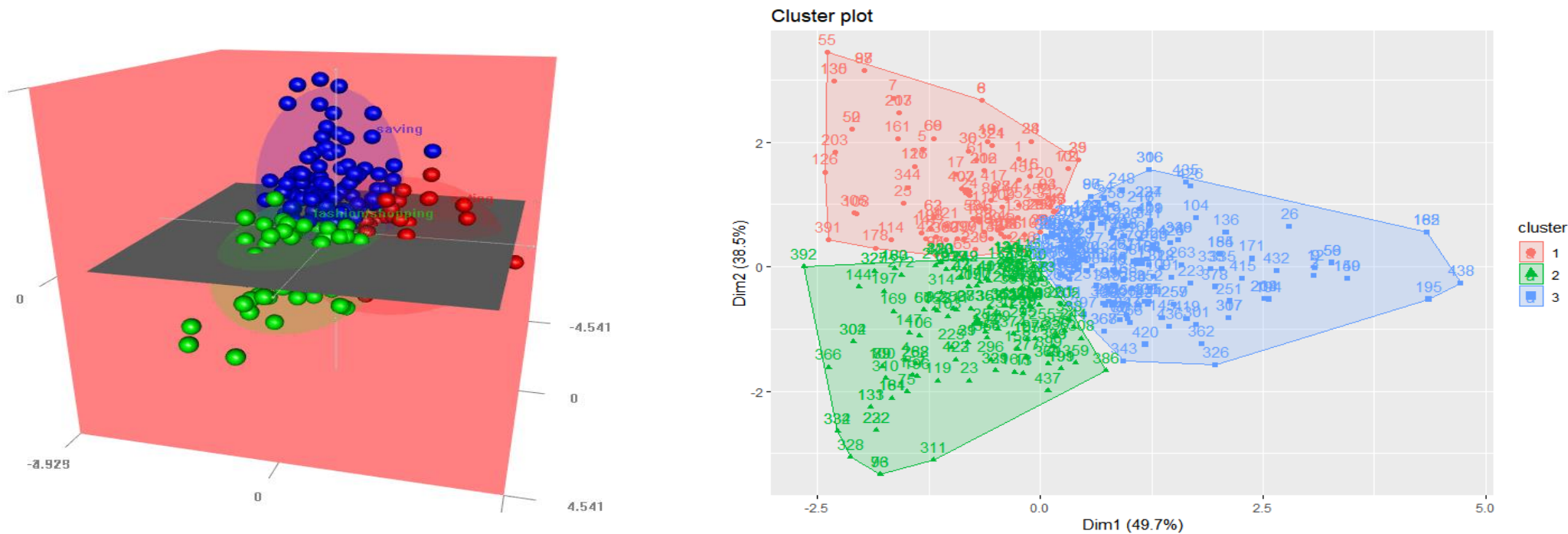
### 3.3 한 달 지출비가 2배가 되었을 때 K-means Cluster분석



#### Cluster들의 중심점

	eat2	fashion/shopping2	saving2
1	0.3815	0.1284	0.0687
2	0.2083	0.2229	0.0959
3	0.2048	0.1097	0.3508

### 3.3 한 달 지출비가 2배가 되었을 때 K-means Cluster분석



#### Cluster의 요약

Cluster	Cluster 정의	표본수
1	식비형	120(27%)
2	패션/쇼핑형	188(43%)
3	저축형	130(29%)



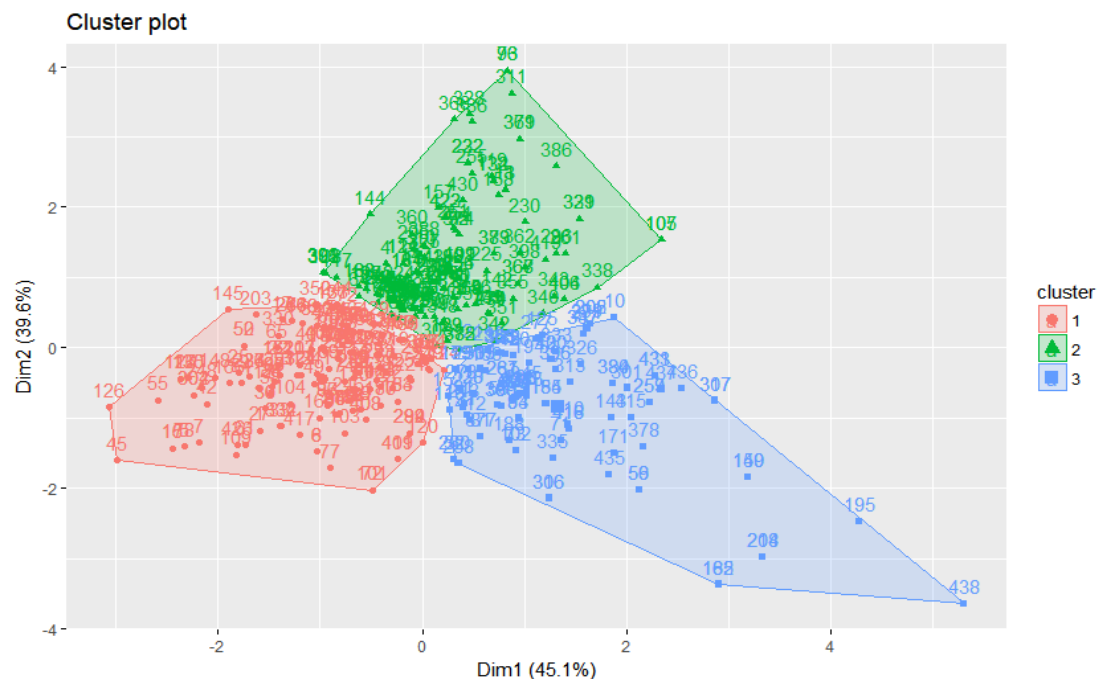
STORY  
**04**



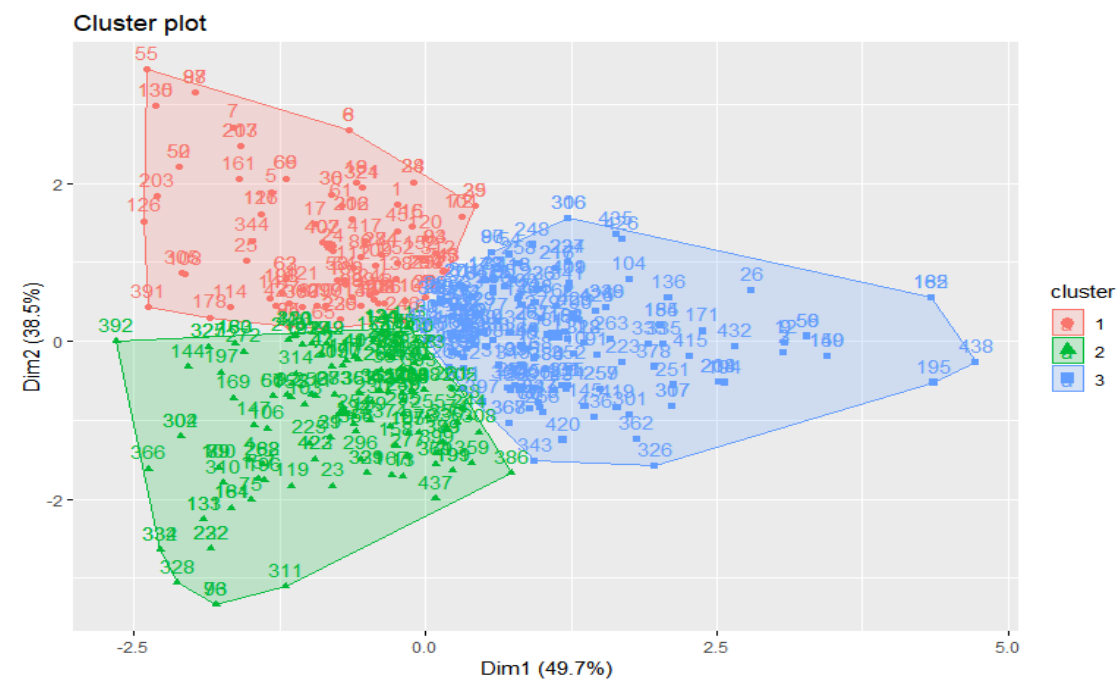
한 달 지출비의 Cluster 분석과  
2배가 된 후의 Cluster 분석 비교



### 3.4 한 달 지출비가 2배가 되기 전 Cluster분석과 2배가 된 후 Cluster 분석 비교



<한 달 지출비 2배 전>

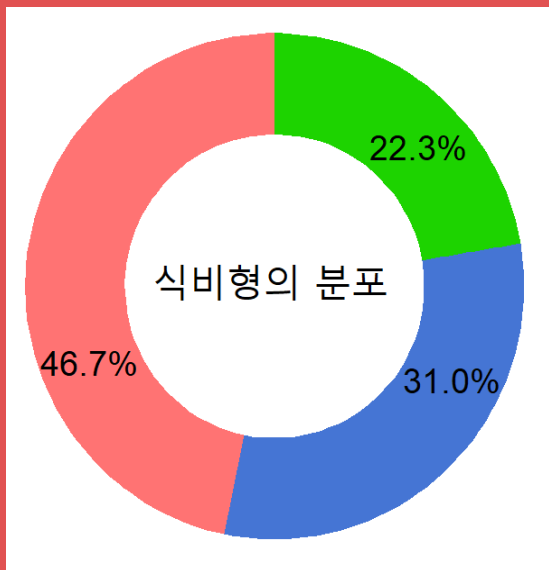


<한 달 지출비 2배 후>

## Cluster의 비교

	한 달 지출비 2배 전	한 달 지출비 2배 후
식비형	161(37%)	120(27%)
패션/쇼핑형	190(43%)	188(43%)
저축형	87(20%)	130(30%)

### 3.4 한 달 지출비가 2배가 되기 전 cluster와 2배가 된 후 cluster 분석 비교



161개의 표본 중 55.3%, 약 85개  
(패션 쇼핑형 35개+저축형50개)의  
데이터가 이동

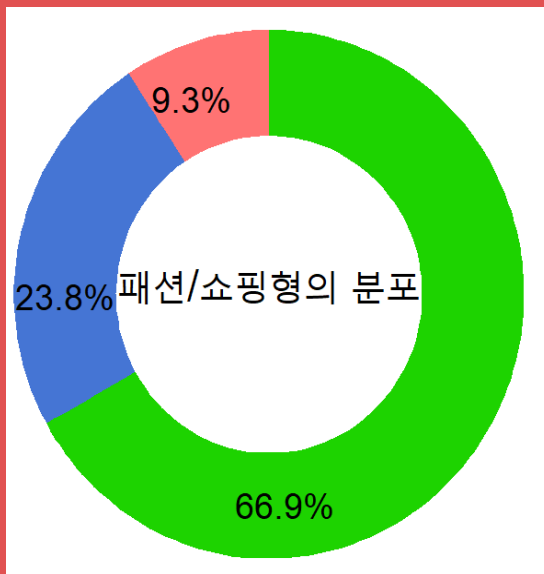
패션/쇼핑형의 분포

저축형의 분포

● 식비형 ● 패션/쇼핑형 ● 저축형

### 3.4 한 달 지출비가 2배가 되기 전 cluster와 2배가 된 후 cluster 분석 비교

식비형의 분포



190개의 표본 중 33.1%, 약 63개(식비형 18개+저축형45개)의 데이터가 이동

저축형의 분포

● 식비형 ● 패션/쇼핑형 ● 저축형

### 3.4 한 달 지출비가 2배가 되기 전 cluster와 2배가 된 후 cluster 분석 비교

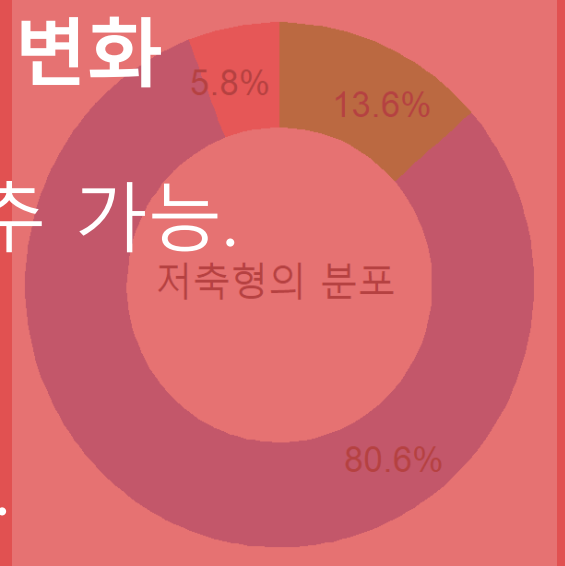
#### Cluster 비교를 통한 소비 성향의 변화

- 2배가 된 후 자신이 원하는 소비를 하고 있다고 유추 가능.

식비형의 분포

패션/쇼핑형의 분포

- 2배가 되었을 때 저축형으로 이동하는 표본이 많음.



- 원래 저축형이었던 표본들은 대부분 2배가 된 후에도 저축형에 속하는 표본이 많음.

87개의 표본 중 19.4%, 약 17개 (식비형 5개+소비/패션형 12개)의 데이터가 이동

● 식비형 ● 패션/쇼핑형 ● 저축형





STORY  
05

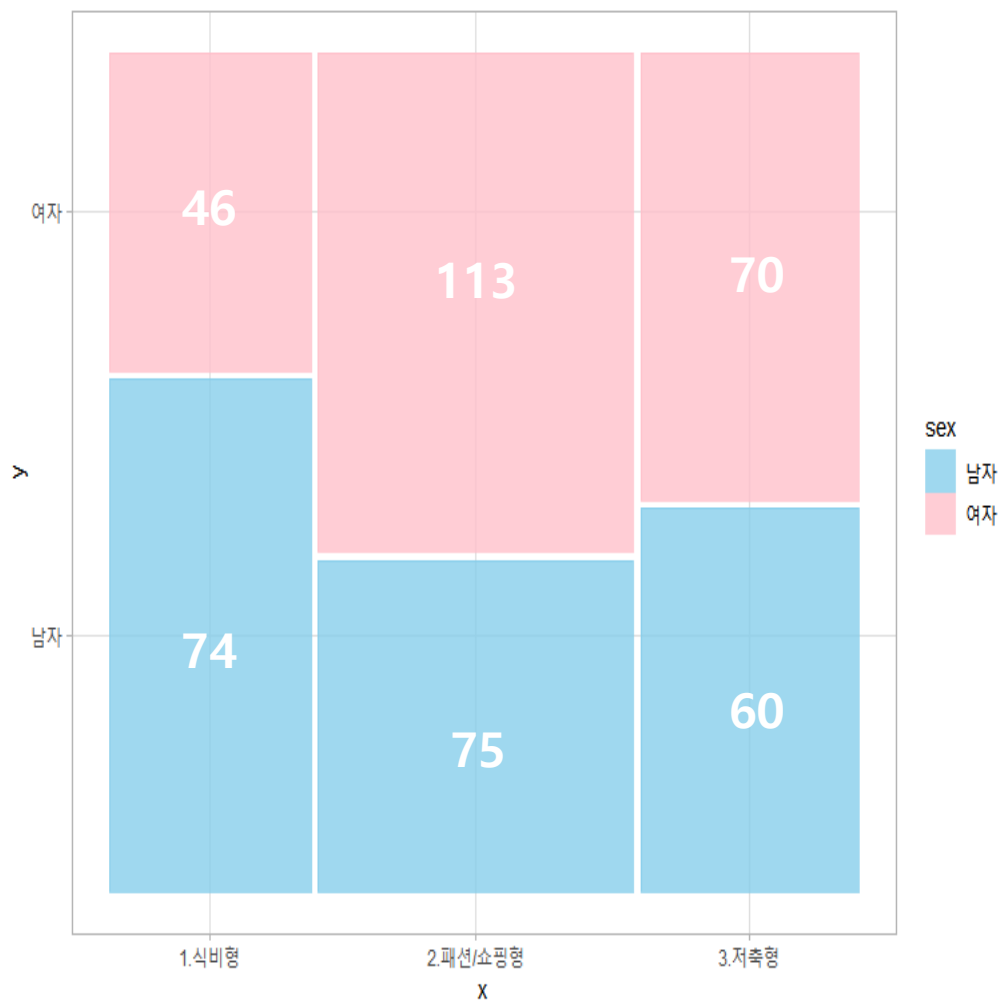


한 달 지출비가 2배가 되었을  
때 Cluster들의 특성 파악



### 3.5 한 달 지출비가 2배가 되었을 때 Cluster들의 특성 파악

① 성별에 따라 cluster들 간의 차이가 있는지 카이제곱검정(독립성 검정)



#### Pearson's Chi-squared test

X-squared = 14.1

df = 2

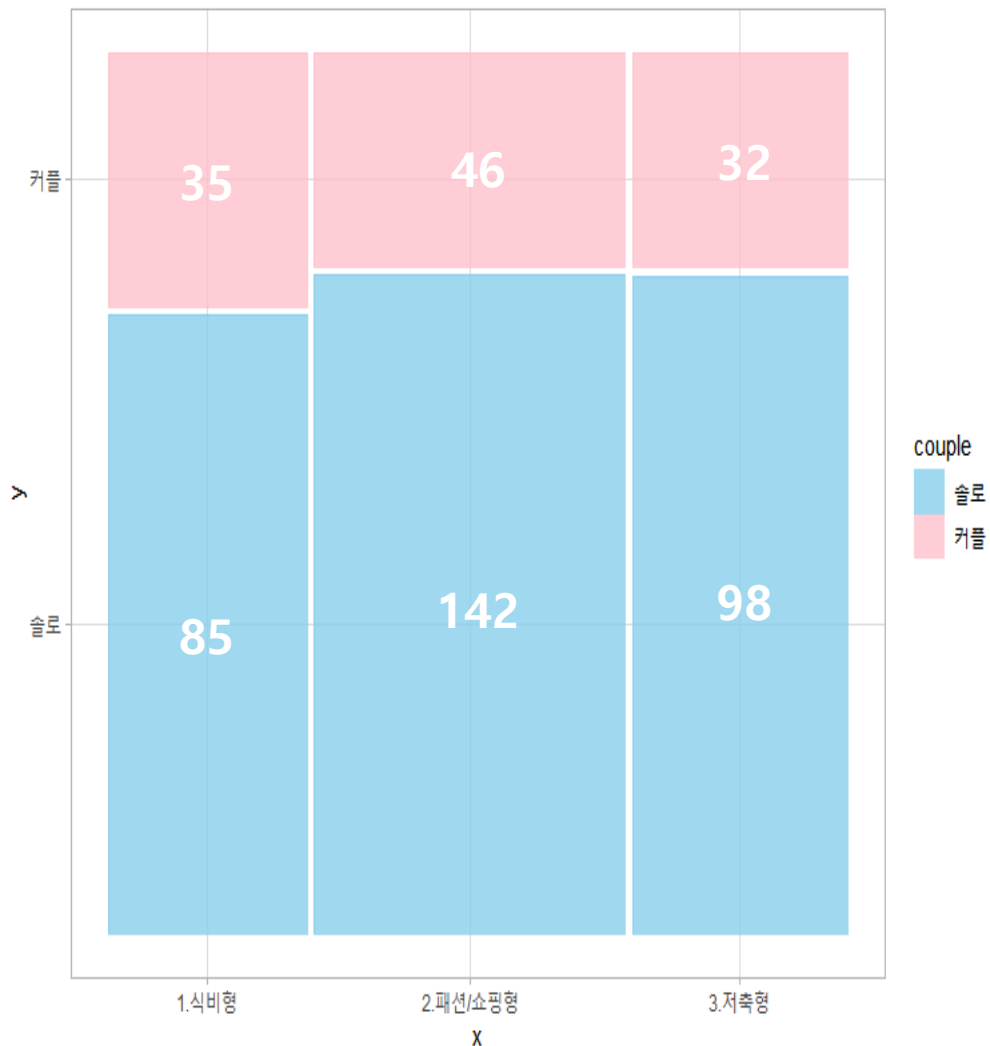
p-value = 0.0008676

#### Cluster와 성별 교차표

	남	여
식비형	74(61%)	46(39%)
패션/쇼핑형	75(40%)	113(60%)
저축형	60(46%)	70(54%)

### 3.5 한 달 지출비가 2배가 되었을 때 Cluster들의 특성 파악

② 연애 여부에 따라 cluster들 간의 차이가 있는지 카이제곱검정(독립성 검정)



### Pearson's Chi-squared test

X-squared = 0.098

df = 2

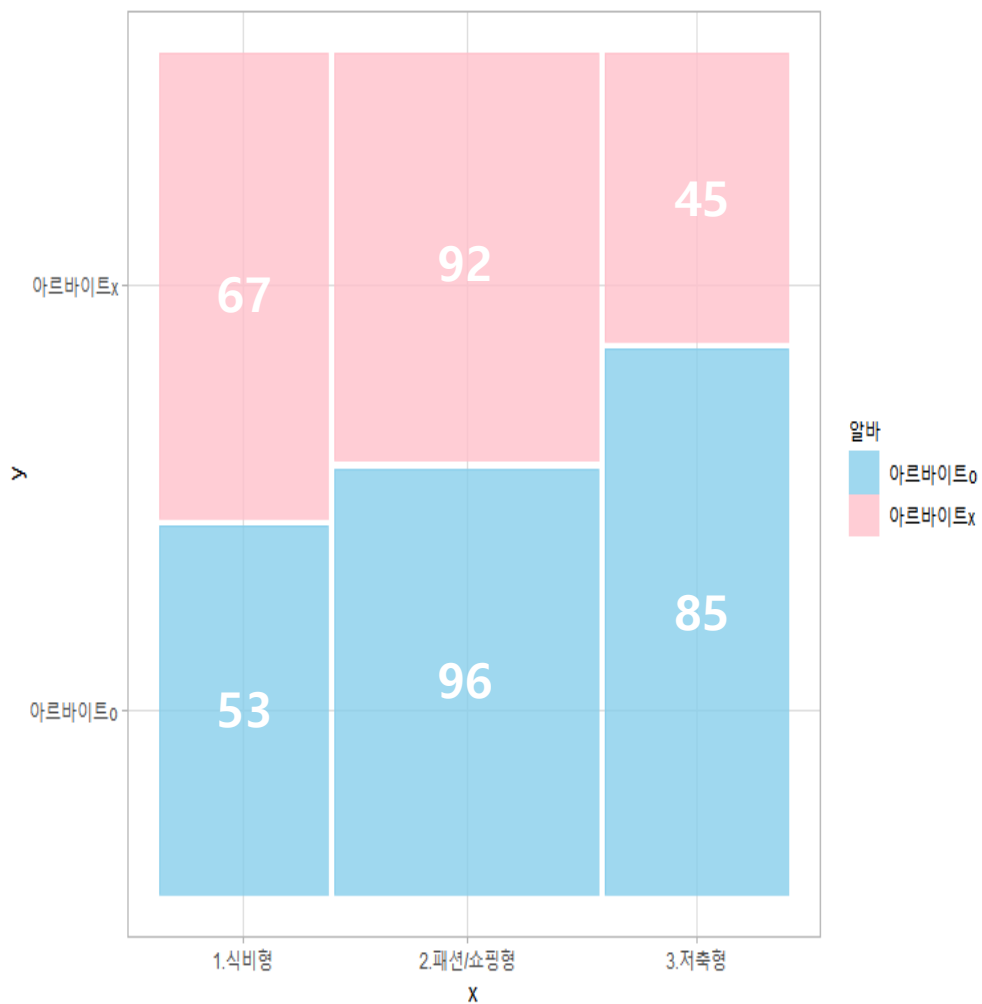
p-value = 0.6126

### Cluster와 연애여부 교차표

	커플	솔로
식비형	35(29%)	85(71%)
패션/쇼핑형	46(24%)	142(76%)
저축형	32(25%)	98(75%)

## 3.5 한 달 지출비가 2배가 되었을 때 Cluster들의 특성 파악

③ 아르바이트 여부에 따라 클러스터들 간의 차이가 있는지 카이제곱 검정(독립성 검정)



## Pearson's Chi-squared test

X-squared = 12.028

df = 2

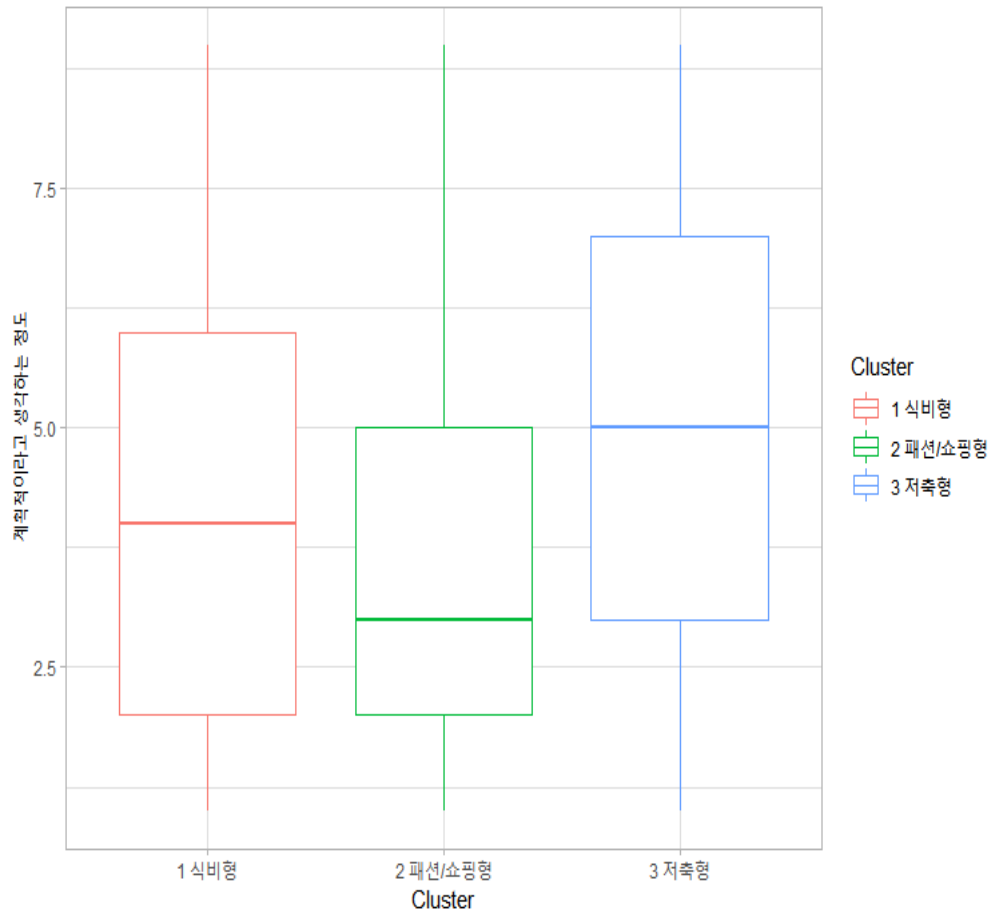
p-value = 0.002445

## Cluster와 연애여부 교차표

	아르바이트 O	아르바이트 X
식비형	53(44%)	67(56%)
패션/쇼핑형	96(51%)	92(49%)
저축형	85(65%)	45(35%)

### 3.5 한 달 지출비가 2배가 되었을 때 Cluster들의 특성 파악

④ cluster들 간의 '자신이 계획적인 소비를 하고 있다고 생각하는 점수(9점 척도)' 평균 비교(ANOVA검정)



- 전체 평균 : 4.3 점
- 식비형 그룹 : 4.2 점
- 패션/쇼핑형 그룹 : 3.9 점
- 저축형 그룹 : 4.93 점

#### ANOVA Analysis

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Cluster	2	75	37.2500	6.8280	0.0012**
Residual	435	2373.1	5.46		






04

결과





	식비형 	패션/쇼핑형 	저축형 
표본 수	120 (27%)	188 (43%)	130 (29%)
성별	남성 다수	여성 다수	차이 없음
계획적 소비	4.2	3.9	4.93
소비내역 만족도	차이 없음 (4.59/4.68/4.87)		
아르바이트 여부	44%	51%	65%
기타	술/유흥 지출비율이 타 cluster에 비해 높 음	문화/여가 지출비율이 타 cluster에 비해 높음	-

# THANK YOU

감사합니다

