

# 모두를 위한 R 데이터 분석 입문

2판



# Chapter 13

## 데이터 분석 사례 I



# 목차

1. 분석 대상 데이터 준비
2. 데이터 탐색
3. 기간별 분석
4. 역삼1동 상권 분석

# Section 01

## 분석 대상 데이터 준비

# 1. 분석 대상 데이터 준비

## 1. 지도 서비스의 중요성

- 공공데이터 포털 사이트(<https://www.data.go.kr/>)에서 상가 정보를 다운로드
- 서울시 자료만을 따로 추출
- 서울시 상가정보는 2015년 12월, 2016년 6월, 2016년 12월, 2017년 6월, 2017년 12월 이렇게 총 5회에 걸쳐서 조사된 것
- 파일 목록: seoul\_201512, seoul\_201606, seoul\_201612, seoul\_201706, seoul\_201712.xlsx
- 데이터가 들어 있는 파일들을 열어 필요한 열만 선택한 후, 5개 파일을 하나로 합치는 작업 준비

	A	B	C	D	E	F	G	H
1	상가업소번호	상호명	지점명	상권업종대분류코드	상권업종대분류명	상권업종중분류코드	상권업종중분류명	상권업종소분류코드
2	20316431	서민시대창짜장		Q	음식	Q02	중식	Q02A00
3	20252211	불난왕죽발보쌈		Q	음식	Q01	한식	Q01A08
4	20294848	삼근삼계탕		Q	음식	Q05	탕/오리요리	Q05A01
5	20469870	다사왕커피숍		Q	음식	Q12	커피점/카페	Q12A01
6	20468674	B.B.Q		Q	음식	Q05	탕/오리요리	Q05A08
7	20472103	삼일축염소		D	소매	D10	건강/미용식품	D10A07
8	20469164	동지호프		Q	음식	Q09	유흥주점	Q09A01
9	20076877	하늘빛건강하우스		D	소매	D10	건강/미용식품	D10A04
10	19997545	제일반찬		D	소매	D01	음/식료품소매	D01A07
11	19995020	전주집		Q	음식	Q01	한식	Q01A01
12	19994770	윤상섭참갈비		D	소매	D01	음/식료품소매	D01A01
13	20385191	오븐에빠진닭	신림점	Q	음식	Q05	탕/오리요리	Q05A08
14	20407436	바다극장		Q	음식	Q01	한식	Q01A01
15	20392880	쌈지		Q	음식	Q04	분식	Q04A01
16	20380776	이모네		Q	음식	Q01	한식	Q01A01
17	20381289	비어벨		Q	음식	Q09	유흥주점	Q09A01
18	20374459	맥주바켓2	신림역점	Q	음식	Q09	유흥주점	Q09A01
19	20296819	커피빈	코리야신림점	Q	음식	Q12	커피점/카페	Q12A01

그림 13-1 상가 정보 파일의 내용

# 1. 분석 대상 데이터 준비

## 코드 13-1

```
setwd("D:/source")

library(ggplot2)
library(ggmap)
library(readxl)

files <- c("201512","201606","201612","201706","201712")
columns <- c("상가업소번호", "상호명", "상권업종대분류명", "상권업종중  
분류명", "상권업종소분류명", "시군구명", "행정동명", "경도", "위도")
ds.total <- NULL

for (i in 1:length(files)) {
  filename <- paste("seoul_", files[i], ".xlsx", sep="")
  cat("read ", filename, "...\n")          # 읽을 파일 이름 출력

  ds <- read_excel(filename)                # 엑셀 파일 읽기
  ds <- data.frame(ds)                      # 데이터프레임으로 변환
  ds <- ds[,columns]                        # 분석에 필요한 변수만 추출
  ds$수집연월 <- rep(i, nrow(ds))          # 데이터 수집 시점
  ds.total <- rbind(ds.total, ds)          # 데이터 통합
}
head(ds.total)
```

# 1. 분석 대상 데이터 준비

```
> setwd("D:/source")
> library(ggplot2)
> library(ggmap)
> library(readxl)
> files <- c("201512", "201606", "201612", "201706", "201712")
> columns <- c("상가업소번호", "상호명", "상권업종대분류명", "상권업종중분류명",
               "상권업종소분류명", "시군구명", "행정동명", "경도", "위도")
> ds.total <- NULL
> for (i in 1:length(files)) {
+   filename <- paste("seoul_", files[i], ".xlsx", sep="")
+   cat("read ", filename, "...\n")           # 읽을 파일 이름 출력
+
+   ds <- read_excel(filename)                 # 엑셀 파일 읽기
+   ds <- data.frame(ds)                       # 데이터프레임으로 변환
+   ds <- ds[,columns]                         # 분석에 필요한 변수만 추출
+   ds$수집연월 <- rep(i, nrow(ds))           # 데이터 수집 시점
+   ds.total <- rbind(ds.total, ds)           # 데이터 통합
+ }
```

# 1. 분석 대상 데이터 준비

```
> head(ds.total)
```

	상가업소번호	상호명	상권업종대분류명	상권업종중분류명			
1	20316431	서민시대참짜장	음식	중식			
2	20252211	별난왕족발보쌈	음식	한식			
3	20294848	상근삼계탕	음식	닭/오리요리			
4	20469870	다사랑커피숍	음식	커피점/카페			
5	20468674	B.B.Q	음식	닭/오리요리			
6	20472103	삼일흑염소	소매	건강/미용식품			
	상권업종소분류명	시군구명	행정동명	경도	위도	수집연월	
1	중국음식/중국집	관악구	은천동	126.9408	37.48348	1	
2	족발/보쌈전문	관악구	은천동	126.9394	37.48395	1	
3	삼계탕전문	관악구	은천동	126.9437	37.48281	1	
4	커피전문점/카페/다방	관악구	은천동	126.9381	37.48522	1	
5	후라이드/양념치킨	관악구	은천동	126.9376	37.48615	1	
6	건강원	관악구	은천동	126.9383	37.48552	1	



# Section 02

## 데이터 탐색

## 2. 데이터 탐색

### 코드 13-2

```
str(ds.total)
unique(ds.total$수집연월)           # 수집연월
unique(ds.total$상권업종대분류명)    # 상권업종 대분류
unique(ds.total$상권업종중분류명)    # 상권업종 중분류
unique(ds.total$상권업종소분류명)    # 상권업종 소분류

# NA 포함여부 확인
sum(is.na(ds.total))

# 201712 수집 데이터만 추출
ds.201712 <- subset(ds.total, ds.total$수집연월 == 5)
dim(ds.201712)

# 업종별 점포수(대분류)
store.level_1 <- aggregate(ds.201712[,1],
                           by=list(대분류=ds.201712$상권업종대분류명),
                           FUN=length)
```

## 2. 데이터 탐색

```
store.level_1
names(store.level_1)[2] = c("count")
ggplot(store.level_1, aes(x=대분류, y=count)) +
  geom_bar(stat="identity", width=0.7, fill="steelblue") +
  ggtitle("업종별 점포수") +
  theme(plot.title = element_text(color="black", size=14, face="bold"))
```

# 구별 점포수

```
store.region <- aggregate(ds.201712[,1],
                          by=list(구이름=ds.201712$시군구명),
                          FUN=length)
store.region
names(store.region)[2] = c("count")
names(store.region)[2] = c("count")
ggplot(store.region, aes(x=구이름, y=count)) +
  geom_bar(stat="identity", width=0.7, fill="steelblue") +
  ggtitle("구별 점포수") +
  theme(plot.title = element_text(color="black", size=14, face="bold"),
        axis.text.x = element_text(angle = 45))
store.region.loc <- aggregate(ds.201712[,c("경도","위도")],
                             by=list(구이름=ds.201712$시군구명),
                             FUN=mean)
```

## 2. 데이터 탐색

```
# 지도 위에 구별 점포수 표시
store.region <- data.frame(store.region, store.region.loc[,2:3])
register_google(key='AlzaSyCK....E9urxjSpPOA')
cen <- c(mean(store.region$경도), mean(store.region$위도))
map <- get_googlemap(center=cen, # 마커 없는 지도 가져오기
                     maptype="roadmap",
                     size=c(640,640),
                     zoom=11)
gmap <- ggmap(map) # 지도를 저장
gmap+geom_point(data=store.region,
               aes(x=경도,y=위도,size=count),
               alpha=0.5, col="red") +
scale_size_continuous(range = c(1, 15))+ # 원의 크기 조절
geom_text(data=store.region, # 지도위에 텍스트 표시
          aes(x=경도,y=위도), # 텍스트 위치 (= 구의 좌표)
          size=3, # 텍스트 크기
          label=store.region$구이름) # 텍스트 내용
```

## 2. 데이터 탐색

```
# 점포수가 많은 상위 10개 동 확인
store.dong <- aggregate(ds.201712[,1],
                        by=list(동 이름=ds.201712$행정동명),
                        FUN=length)
store.dong
names(store.dong)[2] = c("count")
store.dong <- store.dong[order(store.dong$count,decreasing=T),]
dong.top10 <- store.dong[1:10,]
dong.top10
ggplot(dong.top10, aes(x=reorder(동 이름, -count), y=count)) +
  geom_bar(stat="identity", width=0.7, fill="steelblue") +
  ggtitle("점포수 많은 상위 10개 동") +
  theme(plot.title = element_text(color="black", size=14, face="bold"),
        axis.text.x = element_text(angle = 45))
```

## 2. 데이터 탐색

### 1. 데이터 기본 정보 확인 및 분석 대상 데이터 추출

```
> str(ds.total)
'data.frame':1941432 obs. of 10 variables:
 $ 상가업소번호    : num  20316431 20252211 20294848 20469870 20468674 ...
 $ 상호명         : chr  "서민시대참짜장" "별난왕족발보쌈" "상근삼계탕" "다사랑커피숍" ...
 $ 상권업종대분류명: chr  "음식" "음식" "음식" "음식" ...
 $ 상권업종중분류명: chr  "중식" "한식" "닭/오리요리" "커피점/카페" ...
 $ 상권업종소분류명: chr  "중국음식/중국집" "족발/보쌈전문" "삼계탕전문" "커피전문점/카페/다방" ...
 $ 시군구명       : chr  "관악구" "관악구" "관악구" "관악구" ...
 $ 행정동명       : chr  "은천동" "은천동" "은천동" "은천동" ...
 $ 경도           : num  127 127 127 127 127 ...
 $ 위도           : num  37.5 37.5 37.5 37.5 37.5 ...
 $ 수집연월       : int  1 1 1 1 1 1 1 1 1 1 ...

> unique(ds.total$수집연월)                                # 수집연월
[1] 1 2 3 4 5
```

## 2. 데이터 탐색

```
> unique(ds.total$상권업종대분류명)           # 상권업종 대분류
[1] "음식"           "소매"           "생활서비스"
[4] "부동산"         "관광/여가/오락" "의료"
[7] "학문/교육"      "숙박"           "스포츠"
> unique(ds.total$상권업종중분류명)           # 상권업종 중분류
[1] "중식"           "한식"
[3] "닭/오리요리"    "커피점/카페"
[5] "건강/미용식품"  "유흥주점"
...(중간 생략)
[97] "비수용복지시설" "특수교육기관"
[99] "기타교육기관"    "실내운동시설"
[101] "실외운동시설"    "운영관리시설"
```

## 2. 데이터 탐색

```
> unique(ds.total$상권업종소분류명)           # 상권업종 소분류
[1] "중국음식/중국집"           "족발/보쌈전문"
[3] "삼계탕전문"                 "커피전문점/카페/다방"
[5] "후라이드/양념치킨"         "건강원"
[7] "호프/맥주"                  "건강식품판매"
[9] "반찬가게"                   "한식/백반/한정식"
...(중간생략)
[765] "축구장"                     "행글라이더"
[767] "실내테니스장"               "패러글라이딩"
[769] "게이트볼장"                 "프로스포츠경기운영"
[771] "야구장"                     "프로골프"

> # NA 포함여부 확인
> sum(is.na(ds.total))
[1] 0

> # 201712 수집 데이터만 추출
> ds.201712 <- subset(ds.total, ds.total$수집연월== 5)
> dim(ds.201712)
[1] 321436    10
```



## 2. 데이터 탐색

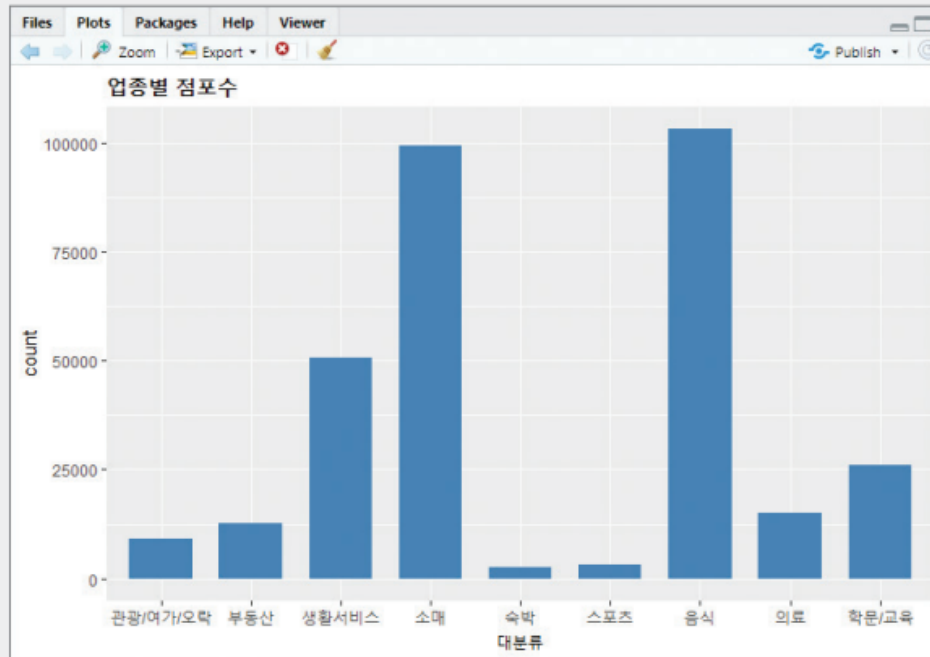
### 2. 업종별 점포수 그래프의 작성

```
> # 업종별 점포수(대분류)
> store.level_1 <- aggregate(ds.201712[,1],
+                             by=list(대분류=ds.201712$상권업종대분류명),
+                             FUN=length)
> store.level_1
```

	대분류	x
1	관광/여가/오락	8968
2	부동산	12677
3	생활서비스	50550
4	소매	99494
5	숙박	2582
6	스포츠	3112
7	음식	103086
8	의료	15065
9	학문/교육	25902

## 2. 데이터 탐색

```
> names(store.level_1)[2] = c("count")  
> ggplot(store.level_1, aes(x=대분류, y=count)) +  
+   geom_bar(stat="identity", width=0.7, fill="steelblue") +  
+   ggtitle("업종별 점포수") +  
+   theme(plot.title = element_text(color="black", size=14, face="bold"))
```



## 2. 데이터 탐색

### 3. 구별 점포수 그래프의 작성

```
> # 구별 점포수
> store.region <- aggregate(ds.201712[,1],
+                             by=list(구이름=ds.201712$시군구명),
+                             FUN=length)
> store.region
```

	구이름	x
1	강남구	29394
2	강동구	13658
3	강북구	8223
4	강서구	12354
5	관악구	13616
6	광진구	14147
7	구로구	12140
8	금천구	9467
9	노원구	11271
10	도봉구	8174

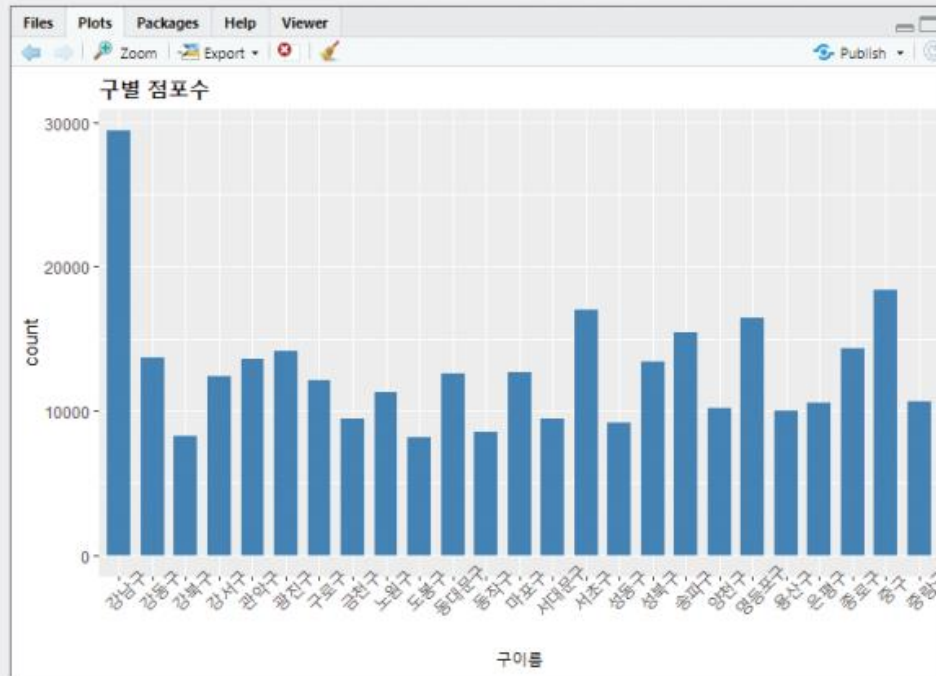
## 2. 데이터 탐색

11	동대문구	12581
12	동작구	8543
13	마포구	12701
14	서대문구	9449
15	서초구	17038
16	성동구	9138
17	성북구	13447
18	송파구	15456
19	양천구	10231
20	영등포구	16471
21	용산구	10021
22	은평구	10581
23	종로구	14318
24	중구	18358
25	중랑구	10659

```
> names(store.region)[2] = c("count")
```

## 2. 데이터 탐색

```
> names(store.region)[2] = c("count")
> ggplot(store.region, aes(x=구이름, y=count)) +
+   geom_bar(stat="identity", width=0.7, fill="steelblue") +
+   ggtitle("구별 점포수") +
+   theme(plot.title = element_text(color="black", size=14, face="bold"),
+         axis.text.x = element_text(angle = 45))
```



## 2. 데이터 탐색

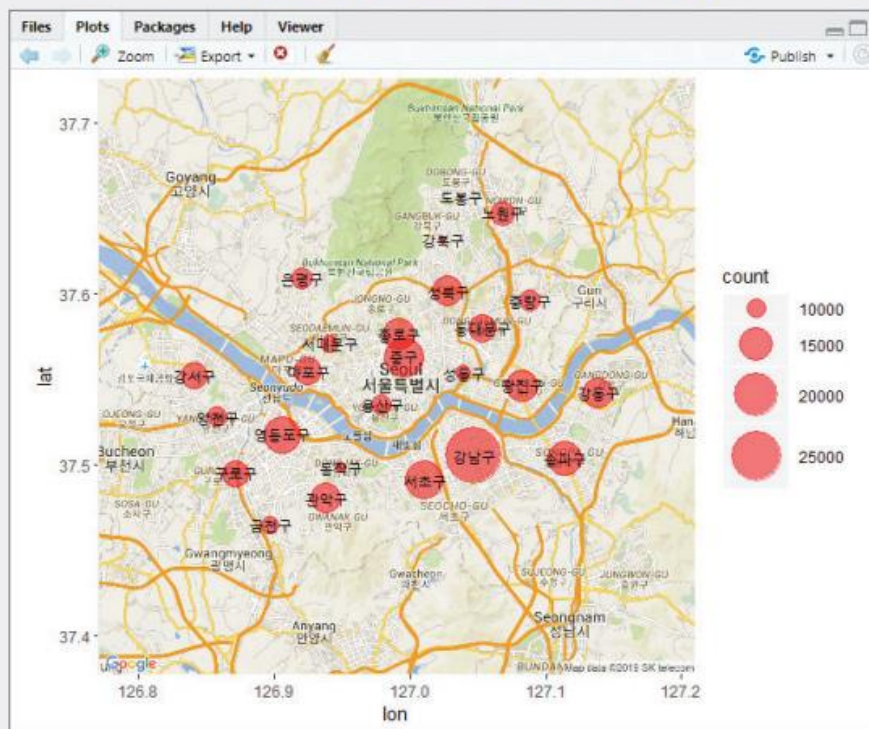
### 4. 지도 위에 구별 점포수 표시

```
> store.region.loc <- aggregate(ds.201712[,c("경도","위도")],  
+                               by=list(구이름=ds.201712$시군구명),  
+                               FUN=mean)  
  
> store.region <- data.frame(store.region, store.region.loc[,2:3])  
  
> register_google(key='AIzaSyCK....E9urxjSpPOA')  
  
> cen <- c(mean(store.region$경도),mean(store.region$위도))  
  
> map <- get_googlemap(center=cen,                                # 마커 없는 지도 가져오기  
+                       maptype="roadmap",  
+                       size=c(640,640),  
+                       zoom=11)  
  
Source : https://maps.googleapis.com/maps/api/staticmap?center=37.551938,126.989693&zoom=11&size=640x640&scale=2&maptype=roadmap&key=xxx-DrgE2qfrE9urxjSpPOA  
  
> gmap <- ggmap(map)                                             # 지도를 저장
```

## 2. 데이터 탐색

```
> gmap+geom_point(data=store.region,  
+                 aes(x=경도,y=위도,size=count),  
+                 alpha=0.5, col="red") +  
+   scale_size_continuous(range = c(1, 15))+  
+   geom_text(data=store.region,  
+            aes(x=경도,y=위도),  
+            size=3,  
+            label=store.region$구이름)
```

# 원의 크기 조절  
# 지도 위에 텍스트 표시  
# 텍스트 위치 (= 구의 좌표)  
# 텍스트 크기  
# 텍스트 내용



## 2. 데이터 탐색

### 5. 점포수가 많은 상위 10개 동 확인

```
> store.dong <- aggregate(ds.201712[,1],  
+                          by=list(동이름=ds.201712$행정동명),  
+                          FUN=length)  
  
> store.dong  
      동이름      x  
1      가락1동  131  
2      가락2동  633  
3      가락본동  956  
...(중간 생략)  
422     휘경2동  400  
423     흑석동  627  
  
> names(store.dong)[2] = c("count")  
> store.dong <- store.dong[order(store.dong$count,decreasing=T),]  
> dong.top10 <- store.dong[1:10,]  
> dong.top10
```

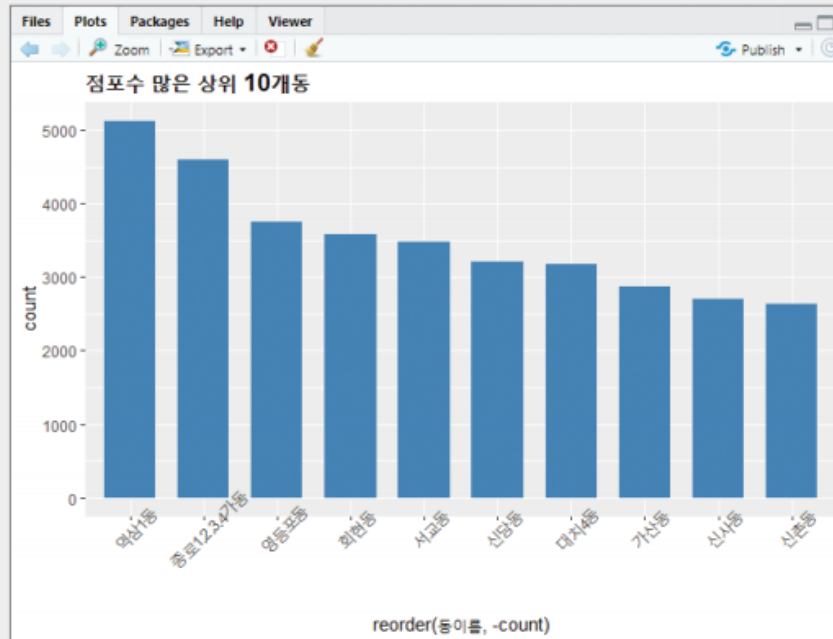


## 2. 데이터 탐색

	동 이름	count
278	역삼1동	5120
353	종로1.2.3.4가동	4601
285	영등포동	3750
418	회현동	3579
188	서교동	3488
238	신당동	3205
74	대치4동	3171
5	가산동	2874
245	신사동	2696
261	신촌동	2632

## 2. 데이터 탐색

```
> ggplot(dong.top10, aes(x=reorder(동이름, -count), y=count)) +  
+   geom_bar(stat="identity", width=0.7, fill="steelblue") +  
+   ggtitle("점포수 많은 상위 10개동") +  
+   theme(plot.title = element_text(color="black", size=14, face="bold"),  
+         axis.text.x = element_text(angle = 45))
```



# Section 03

## 기간별 분석

### 3. 기간별 분석

#### 코드 13-3

```
# 업종별 점포수의 변화
store.change <- aggregate(ds.total[,1],
                          by=list(연월=ds.total$수집연월,
                                  업종대분류=ds.total$상권업종대분류명),
                          FUN=length)
head(store.change)
names(store.change)[3] <- c("count")
ggplot(store.change, aes(x=연월, y=count, colour=업종대분류, group=업
종대분류)) +
  geom_line( ) +
  geom_point(size=6, shape=19, alpha=0.5) +
  ggtitle("업종별 점포수 변화 (대분류)") +
  ylab("점포수") +
  scale_x_continuous(breaks=1:5,
                    labels=files) +
  theme(plot.title = element_text(color="black", size=14, face="bold"))
```

### 3. 기간별 분석

```
# 점포수 변화가 큰 상위 10개 업종
store.tmp <- aggregate(ds.total[,1],
  by=list(연월=ds.total$수집연월,
    업종소분류=ds.total$상권업종소분류명),
  FUN=length)
names(store.tmp)[3] <- c("count")
store.201512 <- store.tmp[store.tmp$연월==1,]
names(store.201512)[3] <- c("cnt_2015")
store.201712 <- store.tmp[store.tmp$연월==5,]
names(store.201712)[3] <- c("cnt_2017")
store.diff <- merge(store.201512[,2:3], store.201712[,2:3])
store.diff$diff <- abs(store.diff$cnt_2015-store.diff$cnt_2017)
store.diff <- store.diff[order(by=store.diff$diff, decreasing=T),]
top10 <- store.diff[1:10,1] # 업종명
top10
store.change <- subset(store.tmp, store.tmp$업종소분류 %in% top10)
ggplot(store.change, aes(x=연월, y=count, colour=업종소분류, group=업
종소분류)) +
  geom_line() +
  geom_point(size=6, shape=19, alpha=0.5) +
  ggtitle("점포수 변화 Top 10 업종(소분류)") +
  ylab("점포수") +
  scale_x_continuous(breaks=1:5,
    labels=files) +
  theme(plot.title = element_text(color="black", size=14, face="bold"))
```

### 3. 기간별 분석

```
# 구별 점포수의 변화
store.gu <- aggregate(ds.total[,1],
                      by=list(연월=ds.total$수집연월,
                              구이름=ds.total$시군구명),
                      FUN=length)
names(store.gu)[3] <- c("count")
ggplot(store.gu, aes(x=연월, y=count, colour=구이름, group=구이름)) +
  geom_line( ) +
  geom_point(size=6, shape=19, alpha=0.5) +
  ggtitle("구별 점포수 변화 (대분류)") +
  ylab("점포수") +
  scale_x_continuous(breaks=1:5,
                    labels=files) +
  theme(plot.title = element_text(color="black", size=14, face="bold"))

# 점포수 변화가 큰 상위 10개 동
store.tmp <- aggregate(ds.total[,1],
                      by=list(연월=ds.total$수집연월,
                              동이름=ds.total$행정동명),
                      FUN=length)
names(store.tmp)[3] <- c("count")
store.dong.201512 <- store.tmp[store.tmp$연월==1,]
```

### 3. 기간별 분석

```
names(store.dong.201512)[3] <- c("cnt_2015")
store.dong.201712 <- store.tmp[store.tmp$연월==5,]
names(store.dong.201712)[3] <- c("cnt_2017")
store.diff <- merge(store.dong.201512[,2:3], store.dong.201712[,2:3])
store.diff$diff <- abs(store.diff$cnt_2015-store.diff$cnt_2017)
store.diff <- store.diff[order(by=store.diff$diff, decreasing=T),]
top10 <- store.diff[1:10,1] # 동명
top10
store.change <- subset(store.tmp, store.tmp$동이름 %in% top10)
ggplot(store.change, aes(x=연월, y=count, colour=동이름, group=동이름))
+
  geom_line( ) +
  geom_point(size=6, shape=19, alpha=0.5) +
  ggtitle("점포수 변화 Top 10 동") +
  ylab("점포수") +
  scale_x_continuous(breaks=1:5,
                     labels=files) +
  theme(plot.title = element_text(color="black", size=14, face="bold"))
```

### 3. 기간별 분석

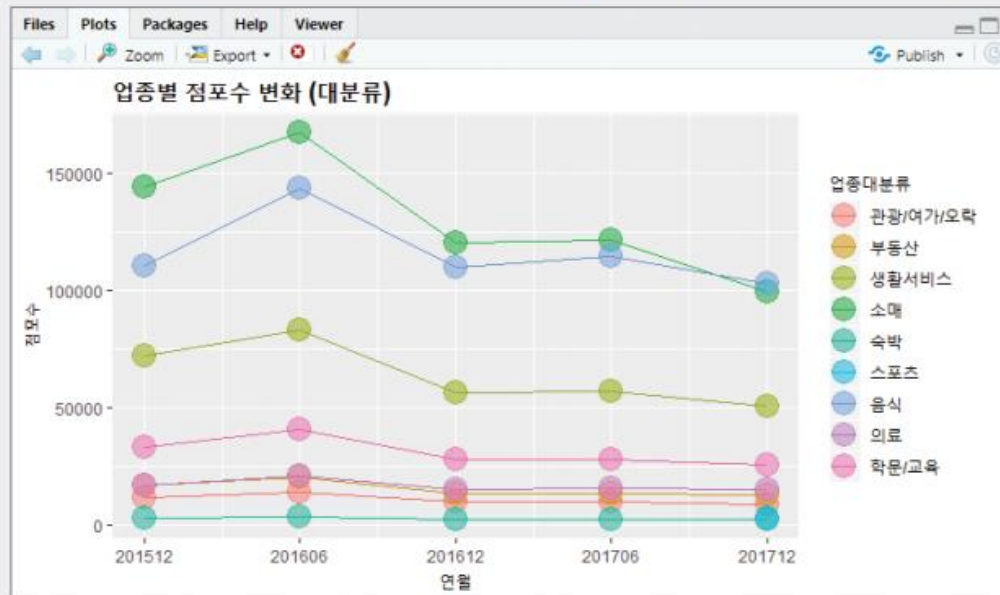
#### 1. 업종별 점포수의 변화

```
> store.change <- aggregate(ds.total[,1],
+                             by=list(연월=ds.total$수집연월,
+                                     업종대분류=ds.total$상권업종대분류명),
+                             FUN=length)
> head(store.change)
  연월  업종대분류    x
1   1  1  관광/여가/오락 11846
2   2  2  관광/여가/오락 14235
3   3  3  관광/여가/오락  9870
4   4  4  관광/여가/오락  9804
5   5  5  관광/여가/오락  8968
6   6  1      부동산 16843
> names(store.change)[3] <- c("count")
```



### 3. 기간별 분석

```
> ggplot(store.change, aes(x=연월, y=count, colour=업종대분류, group=업종대분류)) +  
+   geom_line() +  
+   geom_point(size=6, shape=19, alpha=0.5) +  
+   ggtitle("업종별 점포수 변화 (대분류)") +  
+   ylab("점포수") +  
+   scale_x_continuous(breaks=1:5,  
+                       labels=files) +  
+   theme(plot.title = element_text(color="black", size=14, face="bold"))
```



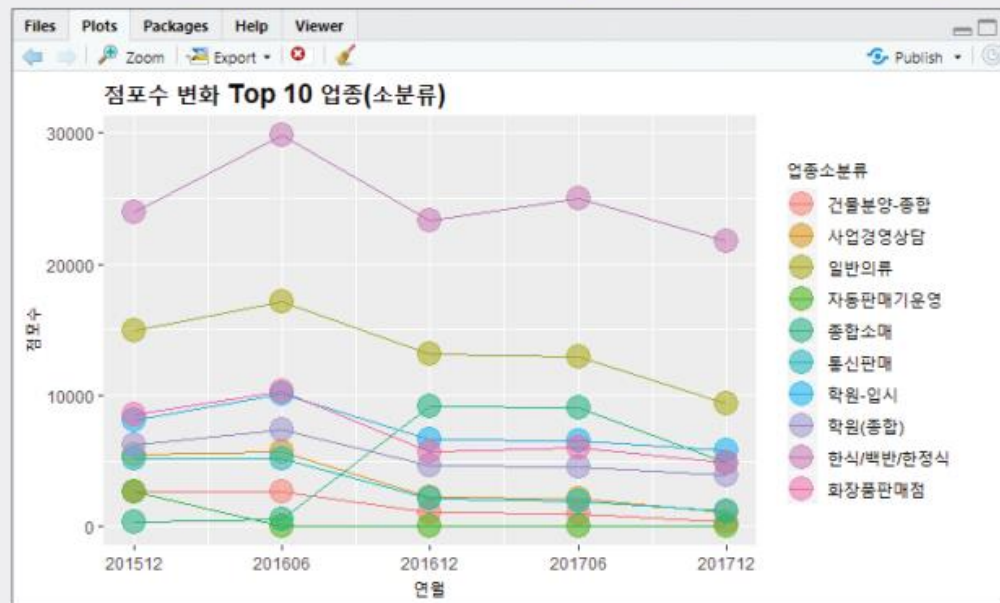
### 3. 기간별 분석

#### 2. 점포수 변화가 큰 상위 10개 업종

```
> store.tmp <- aggregate(ds.total[,1],
+                         by=list(연월=ds.total$수집연월,
+                                업종소분류=ds.total$상권업종소분류명),
+                         FUN=length)
> names(store.tmp)[3] <- c("count")
> store.201512 <- store.tmp[store.tmp$연월==1,]
> names(store.201512)[3] <- c("cnt_2015")
> store.201712 <- store.tmp[store.tmp$연월==5,]
> names(store.201712)[3] <- c("cnt_2017")
> store.diff <- merge(store.201512[,2:3], store.201712[,2:3])
> store.diff$diff <- abs(store.diff$cnt_2015-store.diff$cnt_2017)
> store.diff <- store.diff[order(by=store.diff$diff, decreasing=T),]
> top10 <- store.diff[1:10,1]      # 업종명
> top10
[1] "역삼1동"      "서교동"      "논현1동"
[4] "서초3동"      "명동"        "신사동"
[7] "종로1.2.3.4가동" "삼성2동"     "삼성1동"
[10] "압구정동"
```

### 3. 기간별 분석

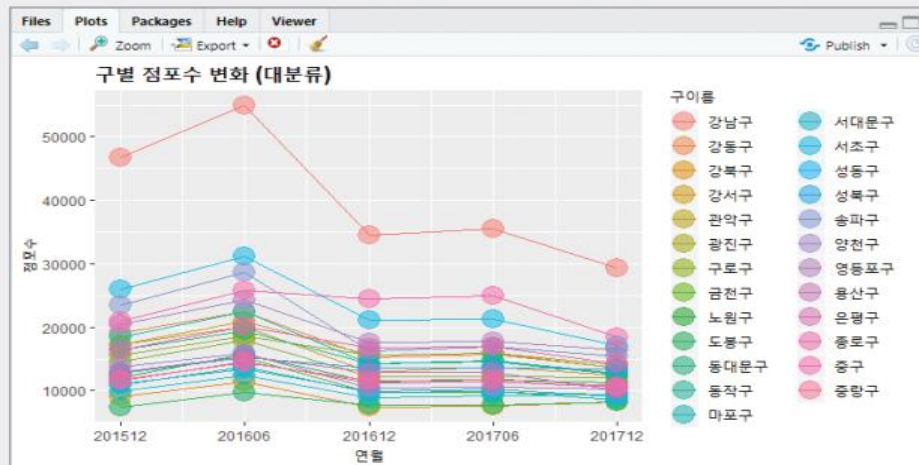
```
> store.change <- subset(store.tmp, store.tmp$업종소분류 %in% top10)
> ggplot(store.change, aes(x=연월, y=count, colour=업종소분류, group=업종소분류)) +
+   geom_line() +
+   geom_point(size=6, shape=19, alpha=0.5) +
+   ggtitle("점포수 변화 Top 10 업종(소분류)") +
+   ylab("점포수") +
+   scale_x_continuous(breaks=1:5,
+                       labels=files) +
+   theme(plot.title = element_text(color="black", size=14, face="bold"))
```



## 3. 기간별 분석

### 3. 구별 점포수의 변화

```
> store.gu <- aggregate(ds.total[,1],  
+                         by=list(연월=ds.total$수집연월,  
+                                 구이름=ds.total$시군구명),  
+                         FUN=length)  
> names(store.gu)[3] <- c("count")  
> ggplot(store.gu, aes(x=연월, y=count, colour=구이름, group=구이름)) +  
+   geom_line() +  
+   geom_point(size=6, shape=19, alpha=0.5) +  
+   ggtitle("구별 점포수 변화 (대분류)") +  
+   ylab("점포수") +  
+   scale_x_continuous(breaks=1:5,  
+                       labels=files) +  
+   theme(plot.title = element_text(color="black", size=14, face="bold"))
```



### 3. 기간별 분석

#### 4. 점포수 변화가 큰 상위 10개 동 확인

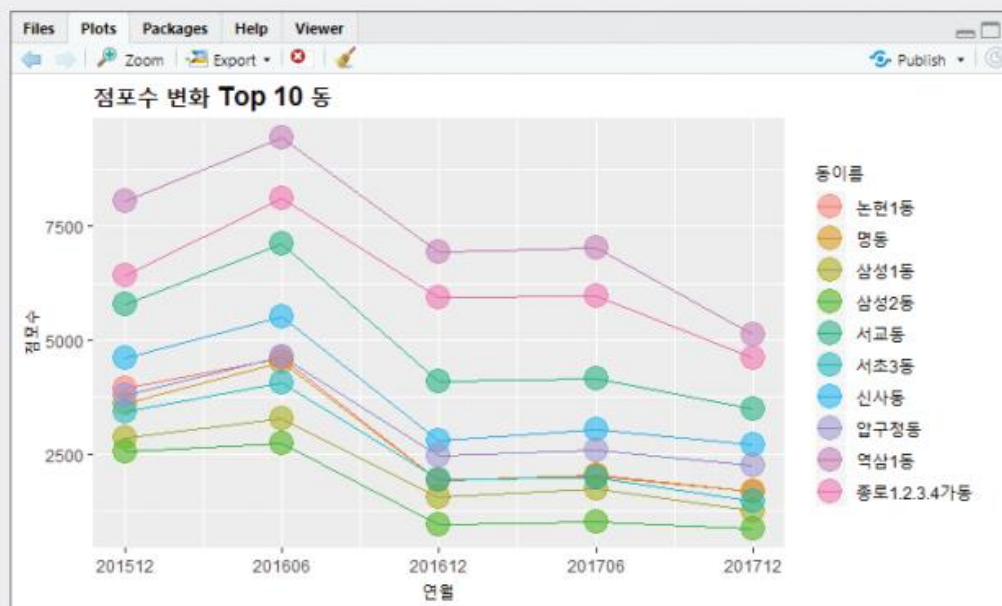
```
> store.tmp <- aggregate(ds.total[,1],
+                        by=list(연월=ds.total$수집연월,
+                               동이름=ds.total$행정동명),
+                        FUN=length)
> names(store.tmp)[3] <- c("count")
> store.dong.201512 <- store.tmp[store.tmp$연월==1,]
> names(store.dong.201512)[3] <- c("cnt_2015")
> store.dong.201712 <- store.tmp[store.tmp$연월==5,]
> names(store.dong.201712)[3] <- c("cnt_2017")

> store.diff <- merge(store.dong.201512[,2:3], store.dong.201712[,2:3])
> store.diff$diff <- abs(store.diff$cnt_2015-store.diff$cnt_2017)
> store.diff <- store.diff[order(by=store.diff$diff, decreasing=T),]
> top10 <- store.diff[1:10,1]                # 동명
> top10
[1] "역삼1동"      "서교동"      "논현1동"
[4] "서초3동"      "명동"        "신사동"
[7] "종로1.2.3.4가동" "삼성2동"     "삼성1동"
[10] "압구정동"
```



### 3. 기간별 분석

```
> store.change <- subset(store.tmp, store.tmp$동이름 %in% top10)
> ggplot(store.change, aes(x=연월, y=count, colour=동이름, group=동이름)) +
+   geom_line() +
+   geom_point(size=6, shape=19, alpha=0.5) +
+   ggtitle("점포수 변화 Top 10 동") +
+   ylab("점포수") +
+   scale_x_continuous(breaks=1:5,
+                       labels=files) +
+   theme(plot.title = element_text(color="black", size=14, face="bold"))
```



# Section 04

## 역삼1동 상권 분석

## 4. 역삼1동 상권 분석

### 코드 13-4

```
ds.yeoksam <- subset(ds.total, ds.total$수집연월==5 &
  ds.total$행정동명=="역삼1동")
# 점포를 업종별로 구분하여 지도에 표시
cen <- c(mean(ds.yeoksam$경도),mean(ds.yeoksam$위도))
map <- get_googlemap(center=cen, # 마커 없는 지도 가져오기
  maptype="roadmap",
  size=c(640,640),
  zoom=15)
gmap <- ggmap(map) # 지도를 저장
gmap+geom_point(data = ds.yeoksam,
  aes(x=경도,y=위도,color=상권업종대분류명),size=2,alpha=0.7) +
  labs(x = "Longitude", y = "Latitude",
  title="역삼1동 업종별 점포", color = "업종")

# 커피 점포만 지도에 표시
ds.yeoksam2 <- subset(ds.yeoksam, ds.yeoksam$상권업종소분류명==
  "커피전문점/카페/다방")
gmap+geom_point(data = ds.yeoksam2,
  aes(x=경도,y=위도), size=2, alpha=0.5, col="red") +
  labs(x = "Longitude", y = "Latitude",
  title="역삼1동 커피점" )
```



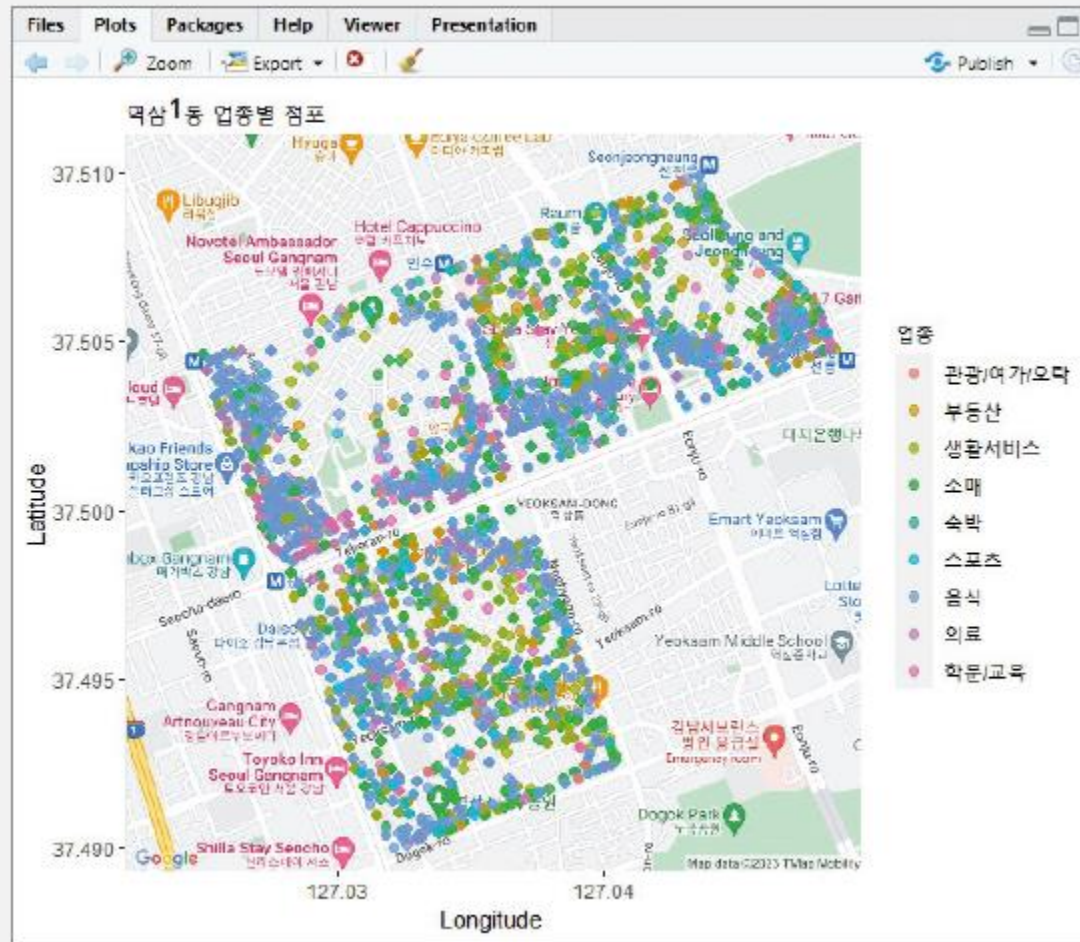
## 4. 역삼1동 상권 분석

```
> ds.yeoksam <- subset(ds.total, ds.total$수집연월==5 &  
+                        ds.total$행정동명=="역삼1동")
```

### 1. 점포를 업종별로 구분하여 지도에 표시

```
cen <- c(mean(ds.yeoksam$경도),mean(ds.yeoksam$위도))  
  
> map <- get_googlemap(center=cen,                # 마커 없는 지도 가져오기  
+                      maptype="roadmap",  
+                      size=c(640,640),  
+                      zoom=15)  
Source : https://maps.googleapis.com/maps/api/staticmap?center=37.500  
22,127.035814&zoom=15&size=640x640&scale=2&maptype=roadmap&key=xxx-  
DrgE2qfrE9urxjSpP0A  
  
> gmap <- ggmap(map)                             # 지도를 저장  
> gmap+geom_point(data = ds.yeoksam,  
+                 aes(x=경도,y=위도,color=상권업종대분류명),size=2,alpha=0.7) +  
+                 labs(x = "Longitude", y = "Latitude",  
+                     title="역삼1동 업종별 점포", color = "업종")
```

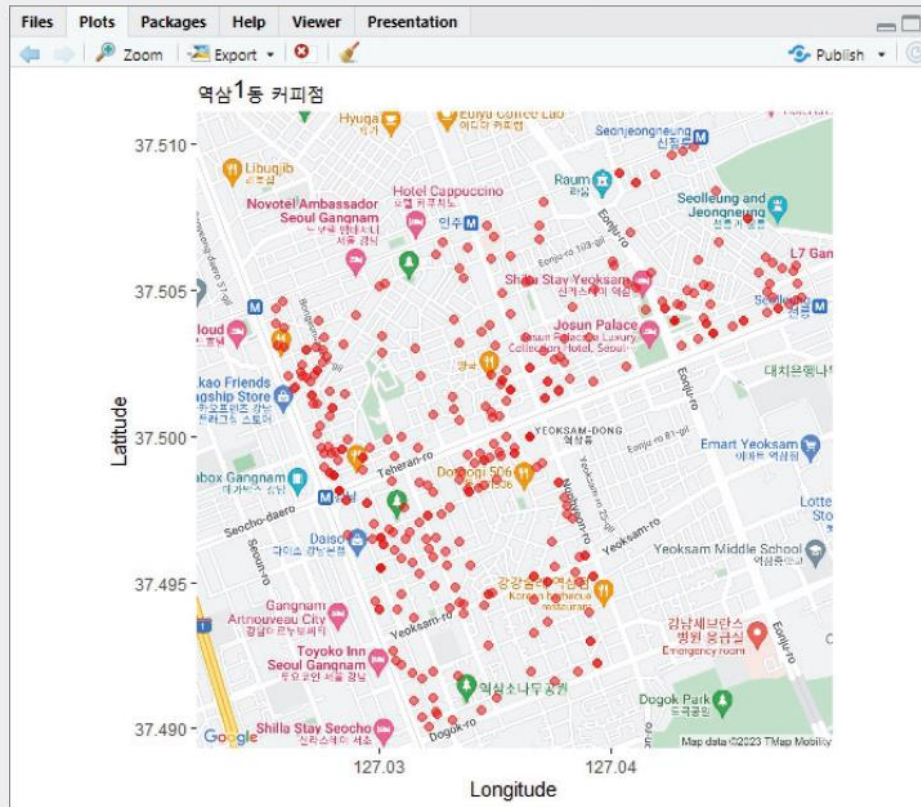
## 4. 역삼1동 상권 분석



## 4. 역삼1동 상권 분석

### 2. 커피 점포만 지도에 표시

```
> ds.yeoksam2 <- subset(ds.yeoksam, ds.yeoksam$상권업종소분류명==  
+ "커피전문점/카페/다방")  
> gmap+geom_point(data = ds.yeoksam2,  
+ aes(x=경도,y=위도), size=2, alpha=0.5, col="red") +  
+ labs(x = "Longitude", y = "Latitude",  
+ title="역삼1동 커피점" )
```



# Thank you!