



"PPT PDF버전을 제출물로 선택한 이유 :  
일련의 분석, 기획안인 만큼, 최대한 시각화하고  
간결하고 결과중심으로 전달하고자 노력하였습니다."

# 2017 1학기 기말 프로젝트

## ZARO의 한국시장 브랜드 점유율을 높이기 위한 방안

### ☰ 목차

- 1) 제5차, 7차 인체치수조사로 알아보는 한국인 남자의 신체 변화
  - 1) 우리나라 신체치수가 점점 커지고 있다는 것을 증명
- 2) OPEN API 이용 : 질병검사 신체정보 분석
  - 1) L 사이즈와 XL 사이즈 생산 증대의 필요성 증명
- 3) 네이버 Open API 를 이용하여 빅사이즈 옷에 대한 트렌트 분석
  - 1) 수요자들의 빅사이즈에 대한 구체적인 니즈 확인
- 4) Google Map 과 ggplot2 를 이용한 쇼핑몰 최적위치 탐색
- 5) SNS(twitter Open API)를 이용한 최적위치 확인 결정, 분석결과와 의사결정

학과: 회계세무학과

학년: 4학년

학번: 201031441

이름: 최관현

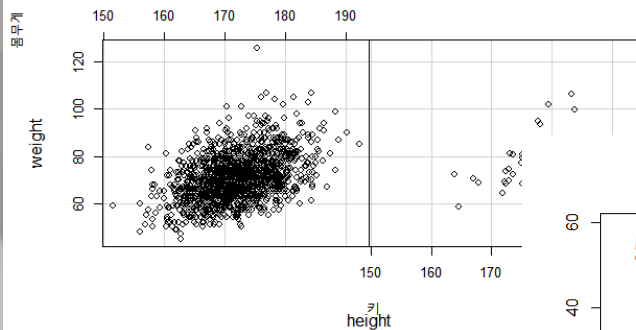
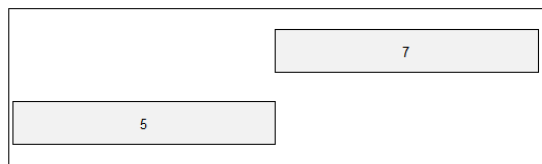


# 제5차, 7차 인체치수조사로 알아보는 한국인 남자의 신체변화

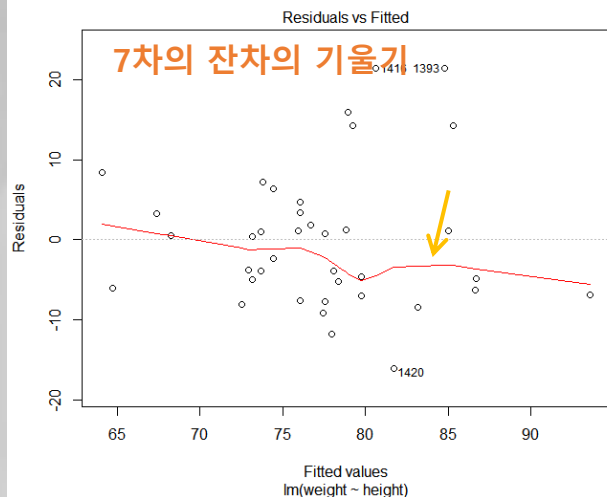
2017-1 기말 프로젝트

sizekorea 제5,7차 인체치수조사 데이터, 분석 데이터는 20 ~ 30대의 남자로 제한, 체형의 비교는 몸무게와 키로 제한

2030세대 5차와 7차 인체치수자료 비교



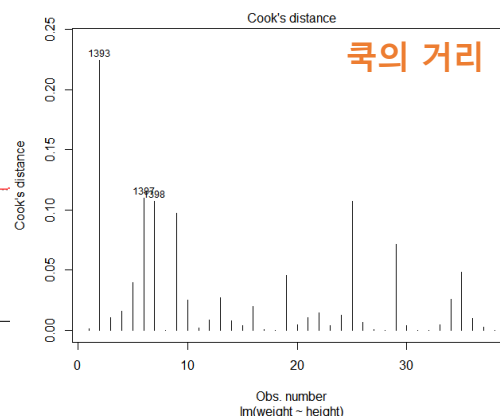
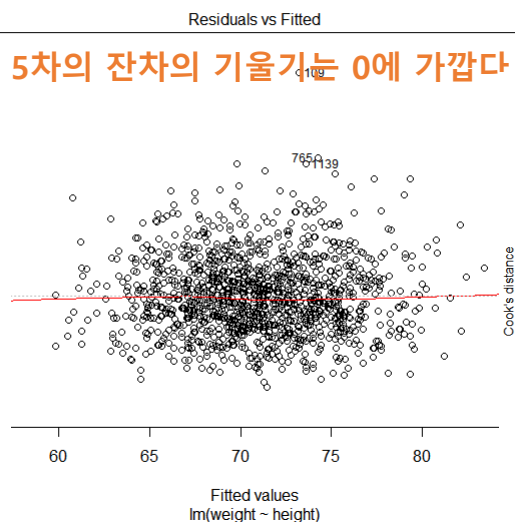
출처 sizekorea.kats.go.kr



coplot() 이용, 2030세대의 5차, 7차조사 비교 :  
제 7차 인체치수조사 자료가 부족하긴 했지만, 윗쪽의 상관관계 그래프를 보면, 5차에 비해 7차가 전반적으로 윗쪽에 있다는 것을 확인 할 수 있다.  
따라서, 이것으로 우리의 신체치수가 점점 커지고 있다는 것을 알 수 있다.

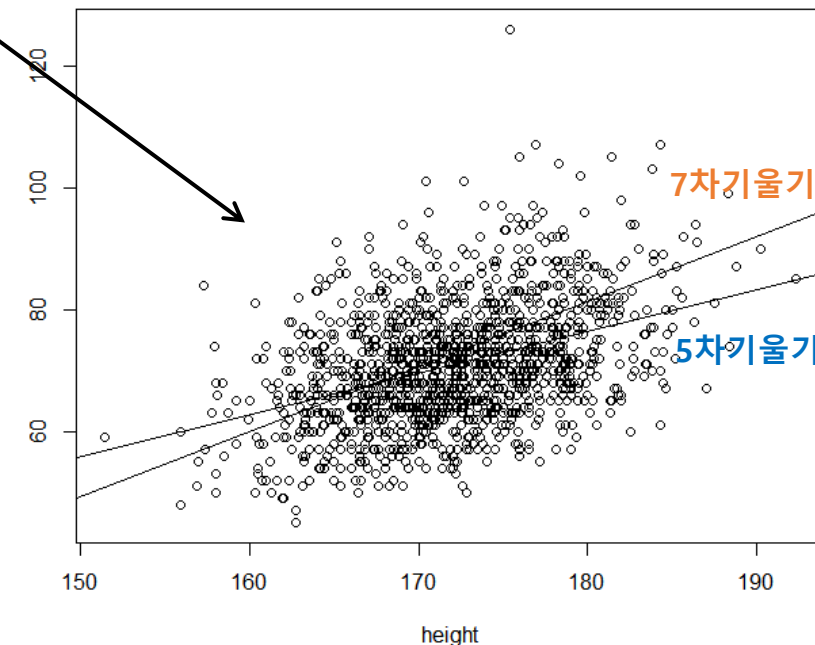
abline(coef()) 이용, 제5차, 제7차의 선형회기(lm())식을 구하여 그래프에 직선으로 나타내었다. 회기직선으로 나타내니 확실히, 기울기가 5차에 비해 높은 것으로 드러나, 우리의 신체지수가 커진 것을 검증하였다.

5차의 잔차의 기울기는 0에 가깝다



쿡의 거리

5차와 7차의 선형회기 직선 시각화



plot.lm(which=c(1,4)) 이용,  
하지만, 7차의 선형회기의 잔차의 기울기를 확인해보니 0에 가깝게 평행하지 않은 것으로 드러나, 7차에 대해서는 신뢰도가 약간 떨어지는 것으로 확인하였다.  
쿡의 거리를 확인해보니, 7차에서는 확실히 데이터의 끝이 고르지 않았다.

💡 공공데이터포털 API 사용, 분석데이터는 징병검사를 받은 1988년생(만19세) 남자의 데이터셋 10000개 ROW

Open API 운영신청, 인증키 획득

**기본정보**

서비스명: 징병검사 신체정보 조회 | 상세설명

서비스 유형: REST | API형태: 0 | 평균응답속도(초): 0

신청여부: 자동승인 | 신청유형: 온오프라인 | 활용신청: 처리상태 | 승인

발행기간: 2017-06-12 ~ 2019-06-12

**서비스정보**

연호자 (UTF-8): koy8yKa%2FNPxo51f44rnb9qsXSwbQTB60Vq5PaasqTsDqZbcK1KChSg3koc0XM2UNYHksa

MD5: k0y8yKa%2FNPxo51f44rnb9qsXSwbQTB60Vq5PaasqTsDqZbcK1KChSg3koc0XM2UNYHksa0dHfPw1km

출처 공공데이터포털 data.go.kr

```
63 library(XML)
64 searchUrl <- "http://apis.data.go.kr/1300000/jBGSSCJeongBo/list"
65 serviceKey <- "koy8yKa%2FNPxo51f44rnb9qsXSwbQTB60Vq5PaasqTsDqZbcK1KChSg3koc0XM2UNYHksa"
66
67 url <- paste(searchUrl,"?ServiceKey=",serviceKey,"&numOfRows=10000","&pageNo=1",
68             "&birth=1988",sep="")
69
70 xmlfile <- xmlParse(url)
71 xmltop <- xmlRoot(xmlfile)
72
73 DF021 <- xmlToDataFrame(getNodeSet(xmltop,"//item"),stringsAsFactors=F)
74 str(DF021)
75
76 DF02 <- DF021
77 DF02 <- DF02[,c(4,11)]
78 apply(DF02,class)
79
80 <- as.numeric(DF02$height)
81 <- as.numeric(DF02$weight)
```

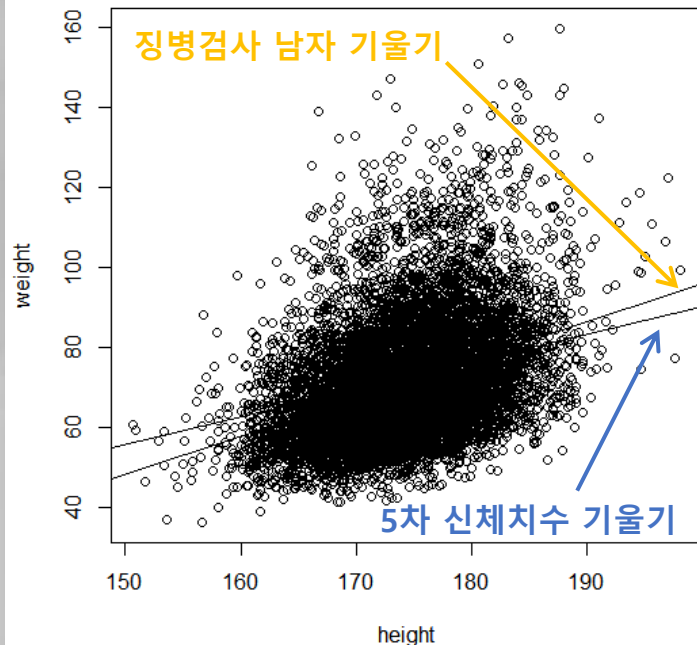
xml 형식으로 parsing하기 위해 XML package 사용

url 에 paste() 를 이용하여 요청url, ServiceKey, 요청할 데이터양을 적어 요청

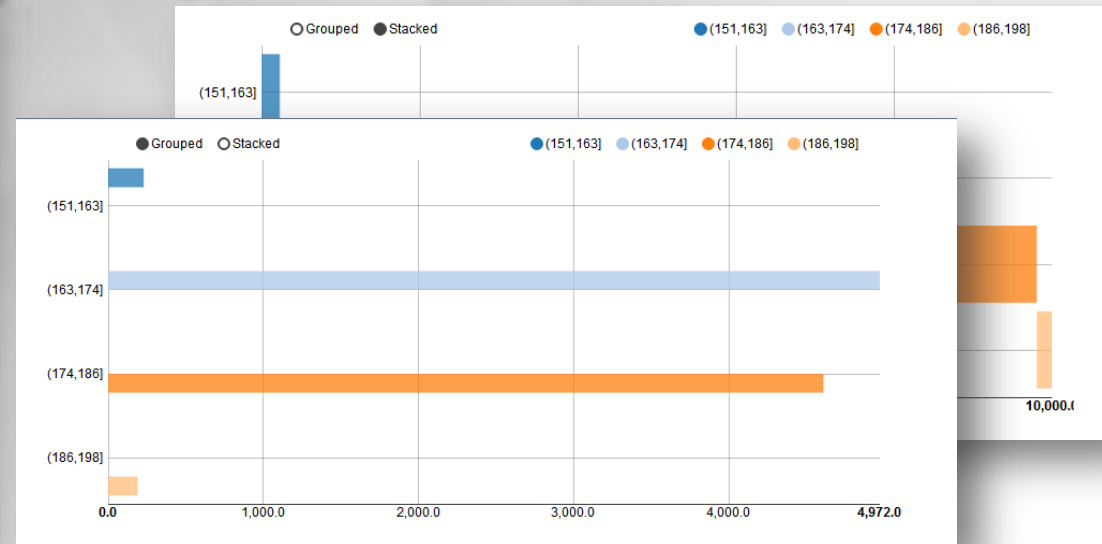
xml 데이터로 파싱한 후, xmlToDataFrame()으로 item 노드를 기준으로 data frame으로 변환

여기서 ! stringsAsFactors를 F로 두어 분석에 적합하게 해주었다

징병검사 신체정보 선형회기 그래프



abline(coef()) 이용, 1만건의 징병검사 신체정보를 이용하여 선형회기 직선을 그려보니, 역시나 5차 신체정보의 기울기보다 높은 것으로 드러나 우리나라 신체 치수가 커졌음을 확인할 수 있었다.



우리나라에는 S,M,L,XL 사이즈가 주를 이루는데, cut() 으로 키를 4등분하여 table() 로 count 한 뒤, rCharts, nPlot(), rCharts 를 이용하여 시각화 분석을 해보았다. S나 M 사이즈에 비해 L나 XL 사이즈도 거의 비슷하게 필요한 것을 알 수 있었다. 또, 30살까지 성장한다고 봤을 때, L사이즈나 XL사이즈 생산을 늘리는 것은 당연했다.



공공데이터포털 API 사용, 분석데이터는 질병검사를 받은 1988년생(만19세) 남자의 데이터셋 10000개 ROW

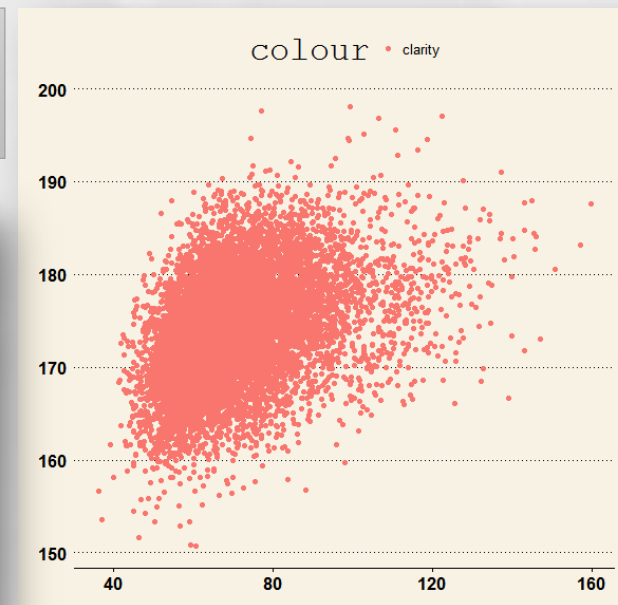
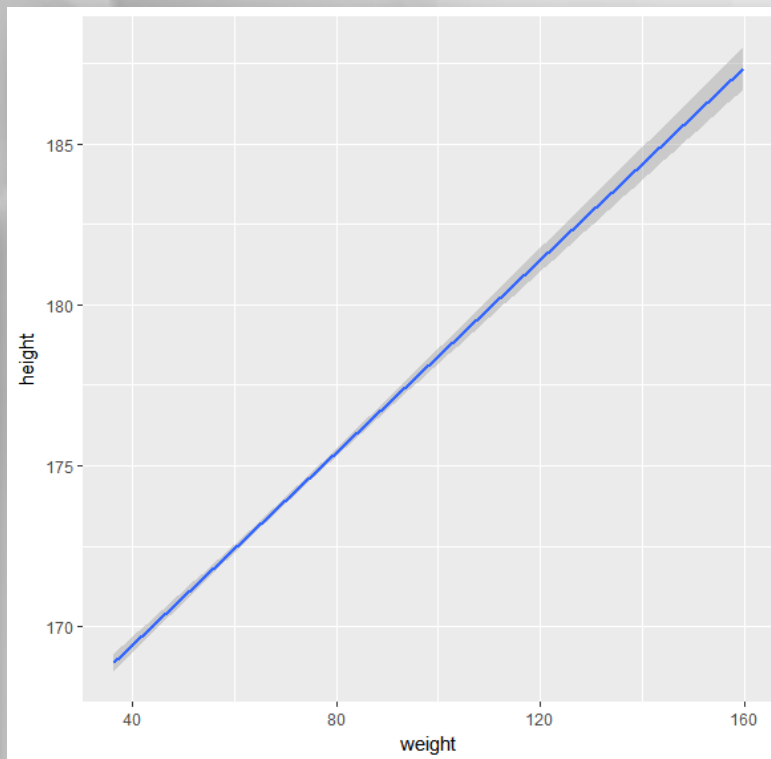
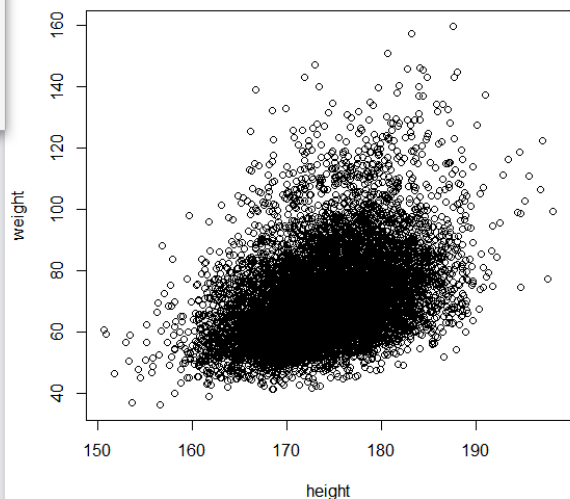
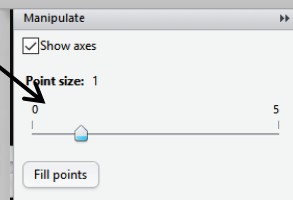
manipulate 패키지를 이용한 시각화

```
112 library(manipulate)
113 manipulate(
114   plot( weight ~ height, data=DF02
115     , axes=axes
116     , cex=cex
117     , pch=if(pch) 19 else 1)
118   , axes=checkbox(TRUE, "Show axes")
119   , cex=slider(0,5,initial=1,step=0.1,label="Point size")
120   , pch=button("Fill points")
121 )
```

데이터 : 질병검사 신체정보 : manipulate 패키지는 설정창을 띄울 수 있고, 이 설정창에서 마음대로 조작을 할 수 있는 기능들을 많이 제공한다. Show axes를 체크 해제하면 x축 y축을 지울 수 있으며, point를 검게 채울 수도 있으며, point size를 키울 수도 있다.

```
library(ggplot2)
library(ggthemes)
ggplot(data=DF02,aes(x=weight,y=height,colour='clarity')) + geom_point() + theme_wsj()
ggplot(data=DF02,aes(x=weight,y=height)) + geom_smooth(method="lm")
```

ggplot() 을 이용하여 데이터를 시각화 해보았다. 단순 R에서 제공하는 plot() 보다는 훨씬 많은 테마들과 기능들을 제공한다.



geom\_smooth(method="lm")으로 쉽게 회귀직선을 그릴 수 있다





## 💡 NAVER Open API 사용, 검색어 "빅사이즈" 로 요즘 빅사이즈 옷에 대한 트렌드 파악

```
130 searchUrl2 <- "https://openapi.naver.com/v1/search/blog.xml"
131
132 client_id <- "2grDuh3XEgKvN6DFHhSD"
133 client_secret <- "QFEn14fgJg"
134
135 query <- URLencode(iconv("빅 사이즈", "euc-kr", "UTF-8"))
136 url2 <- paste(searchUrl2, "?query=", query, "&display=100", sep="")
137 doc <- getURL(url2,
138               httpheader=c('Content-Type'="application/xml"
139                             , 'X-Naver-Client-Id'=client_id
140                             , 'X-Naver-Client-Secret'=client_secret))
141
142 doc2 <- htmlParse(doc, encoding="UTF-8")
143 text <- xpathApply(doc2, "//item/description", xmlValue)
```

요청주소 url 로 client\_id 와 client\_secret 을 담아서 GET 방식으로 요청하였다

100개의 결과를 이용하였고, html 형식을 파싱해주어 UTF-8 로 변환시켜준다

wordcloud(), 워드클라우드로 검색어를 정제하여 트렌드를 파악하였다

### "남자빅사이즈" 검색



```
171 palette <- brewer.pal(9, "Set1")
172 wordcloud(names(wordcount), freq=wordcount,
173           scale=c(8,1), rot.per=0, min.freq=1,
174           random.order=F, random.color=T, colors=palette)
```

**워드클라우드** 분석결과로 미루어보아, 여름계절용 빅사이즈 옷이 최근 사람들이 흥미가 많으며, 빅사이즈임에도 사람들이 다양한 코디, 디자인을 원하고 있으며, 전문적인 매장을 원하고 있다는 것을 알 수 있었다.

또, 남녀 할 것 없이 빅사이즈에 대한 니즈가 크고, 사이즈가 맞는지 안맞는지가 중요한 만큼, 매장을 찾아보는 사람들이 많았다. 오프라인 매장에 대한 니즈를 파악할 수 있었다.

또, "남자빅사이즈"로 검색을 해보니, 자켓, 반바지, 니트, 조끼 등 다양한 상품 바운더리에서 수요가 있다는 것을 확인 할 수 있었다.





## 구글맵을 사용하여 요즘 '핫' 한 쇼핑스트리트에서 프로모션을 진행하라!

빅사이즈 옷을 프로모션할 '핫' 한 쇼핑스트리트를 지도상에 표시해보자

```
234 library(ggmap)
235 library(ggplot2)
236 addr <- c("서울특별시 중구 명동",
237          "서울특별시 강남구 영동대로 513",
238          "서울특별시 영등포구 국제금융로 10",
239          "서울특별시 광진구 아차산로 200",
240          "서울특별시 영등포구 영중로 15",
241          "서울특별시 강남구 신사동",
242          "서울특별시 서초구 반포동",
243          "서울특별시 중구 을지로 281")
244 name <- c("명동쇼핑스트리트", "IFC몰", "커먼그라운드", "영등포타임스퀘어",
245          "가로수길", "파미에스테이션", "동대문디자인플라자")
246 gc <- geocode(enc2utf8(addr))
247 DF05 <- cbind(name, gc)
248 DF05
249 center <- c(mean(gc$lon), mean(gc$lat))
250 namecolor <- DF05[,1]
251 color <- as.integer(namecolor)*10
252
253 map <- get_googlemap(center=center, maptype="roadmap", zoom=12, marker=gc)
254 V05 <- ggmap(map, extent = "device") + geom_point(data=DF05, aes(x=DF05$lon, y=DF05$lat),
255          size=2, alpha=0.7, color=color)
256 V051 <- V05 + geom_text(data=DF05, aes(x=DF05$lon, y=DF05$lat, label=DF05$name), size=5)
257 V051
```

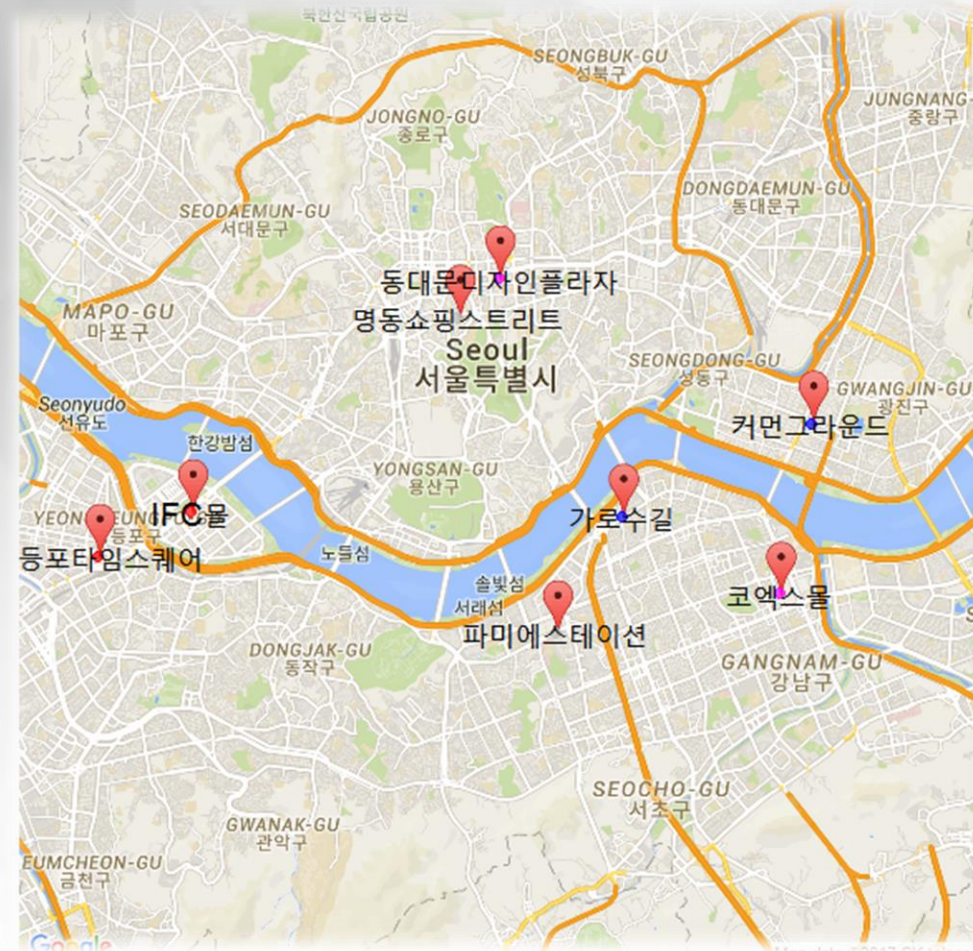
geocode(enc2utf8(addr)) 로 위도와 경도를 가져온다  
center 변수를 주어 지도를 표시할 곳을 잡아주었다

get\_googlemap() 으로 지도를 불러온다  
색상을 다르게 표시하기 위해 color 변수를 만들었다

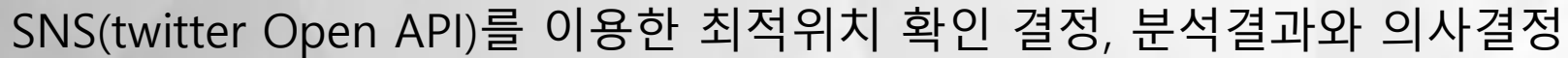
ggmap() 으로 지도를 표시해주고 ggplot2 패키지를 이용하여 꾸며주었다

geom\_point 로 각 점의 색상을 다르게 표시해주었고 (변수 color)  
geom\_text 를 이용하여 각 점에 이름을 붙여주었다

Google Map 을 활용한 시각화, 점 색깔 다르게 표시, 라벨링 (ggplot2 적용)  
요즘 '핫' 한 쇼핑스트리트를 꼽아 지도상에 표시하고 오프라인 매장 런칭을  
고려하는 것으로 한다







💡 twitter Open API 를 이용한 워드클라우드, barplot() 으로 확인

```
270 setup_twitter_oauth(  
271   consumer_key = "NHY572m33si1HywHV31TS2L4D",  
272   consumer_secret = "CVLbp6eEq7ta2Fy3ic6pJ5L5ELrTAaDDS6LthysAeyj988n7NW",  
273   access_token = "412876876-SWNM20E3Miv1khu6eY4fdpMq7R599a8UmXPLheUw",  
274   access_secret = "JY0ymGg3TpOORYQMIHZ6uLxW5Km4CsT86DDFWuiTaOjfk"  
275 )  
276 ? searchTwitter()  
277 keyword <- enc2utf8("커먼 그라운드")  
278 tweets <- searchTwitter(keyword,n=1000, resultType="recent", retryOnRateLimit=100)
```



## 최종 의사결정

ZARO 그룹은, 작은 사이즈옷은 재고가 남았지만, 큰 사이즈옷이 잘 팔려서 재고가 부족할 정도였다. 그 이유를 찾아보았는데, sizekorea에서 제공하는 제5차, 제7차 인체치수조사를 조사하여보니, 회기분석을 통해 그동안 우리나라 사람들의 체형이 커졌음을 알 수 있었다. 또, 공공데이터포털의 징병검사 신체정보를 받아서 회기분석 및 시각화를 해보니, 확실히 신체 체격이 커졌음을 검증할 수 있었다. 따라서, 앞으로의 ZARO의 사이즈 생산에는 S사이즈 생산을 줄이고, XL사이즈 생산을 늘려야할 것이다

더 나아가 ZARO는 더 큰 빅사이즈 생산, XXL사이즈 생산을 고려해야할 것이다. 이것은 네이터 오픈API 검색결과를 분석하여 시각화하여 보니, 빅사이즈 시장에서도 일반사이즈와 다를 것 없이, 디자인의 다양성, 전문성 등의 고객들의 니즈를 파악할 수 있었고, 무엇보다도 빅 사이즈매장을 위하여 직접 입어볼 수 있는 환경이 필요함을 알 수 있었다.

그렇게 하여 오프라인매장 런칭을 고려해야하는데, 마땅한 장소를 선정하여 구글맵에 시각화하여 보았다. 장소들은 쇼핑스트리트로 유명한 지역을 표시한 것이다. 그 중, 건대입구역의 '커먼그라운드'에 대해 트위터API 를 통해 재 확인 결정하였다. 최종적으로 ZARO 그룹은 S 사이즈 생산 감소, XL사이즈 생산증가, XXL사이즈 기획, 오프라인 매장인 '커먼그라운드'에서의 매장오픈을 준비하는 것으로 의사결정 짓기로 하였다.