

ggmap & word cloud

- ggmap 패키지는 ggplot2 패키지를 확장하여 지도를 그리는 패키지로 구글 맵 or 오픈스트리트 맵 등을 이용하여 시각화 가능
- ggmap 패키지 함수
 - get_map() 함수
 - gg_map() 함수
 - qmap() 함수

get_map() 함수

- RgoogleMaps 패키지의 GetMap() 함수와 같은 기능을 함수로 지도를 가져오는 함수
- 지도를 얻기 위해서 구글 맵, 오픈스트리트 맵 서버에 지역과 줌 값을 기반으로 질의하여 결과를 가져오는 역할 수행
- get_googlemap(), get_openstreetmap(), get_stamenmap() 함수 등 여러 함수를 결합하여 정의된 함수
- 형식
 - get_map (location = 위치 문자열 or c(lon = value, lat = value), zoom="auto", maptype = type ("satellite", "roadmap" 등), color = col ("color", "bw" , source = "google", "osm", "stamen" 등)

- `get_map` 함수에 의해 생성된 래스터 객체(지도)를 시각화하는 함수

- 형식

- `ggmap (get_map object, extent= "normal", "panel", "device")`
- extent option

"normal": ggplot의 시각화 처럼 회색 패널의 테두리 표현

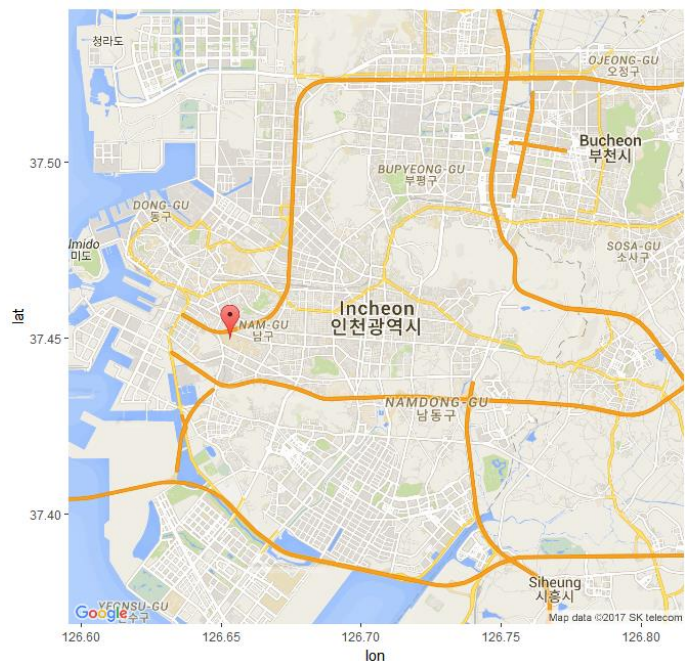
"panel": 패널의 테두리 없이 모든 패널 영역에 시각화

"device": 그래픽 디바이스 화면 전체에 시각화

ggmap 패키지 예제

■ get_googlemap + ggmap 예제

```
library(ggmap)
lon = 126.653
lat = 37.45
map = get_googlemap("Incheon", zoom=12, maptype="roadmap",
                    markers=data.frame(lon, lat))
ggmap(map)
```



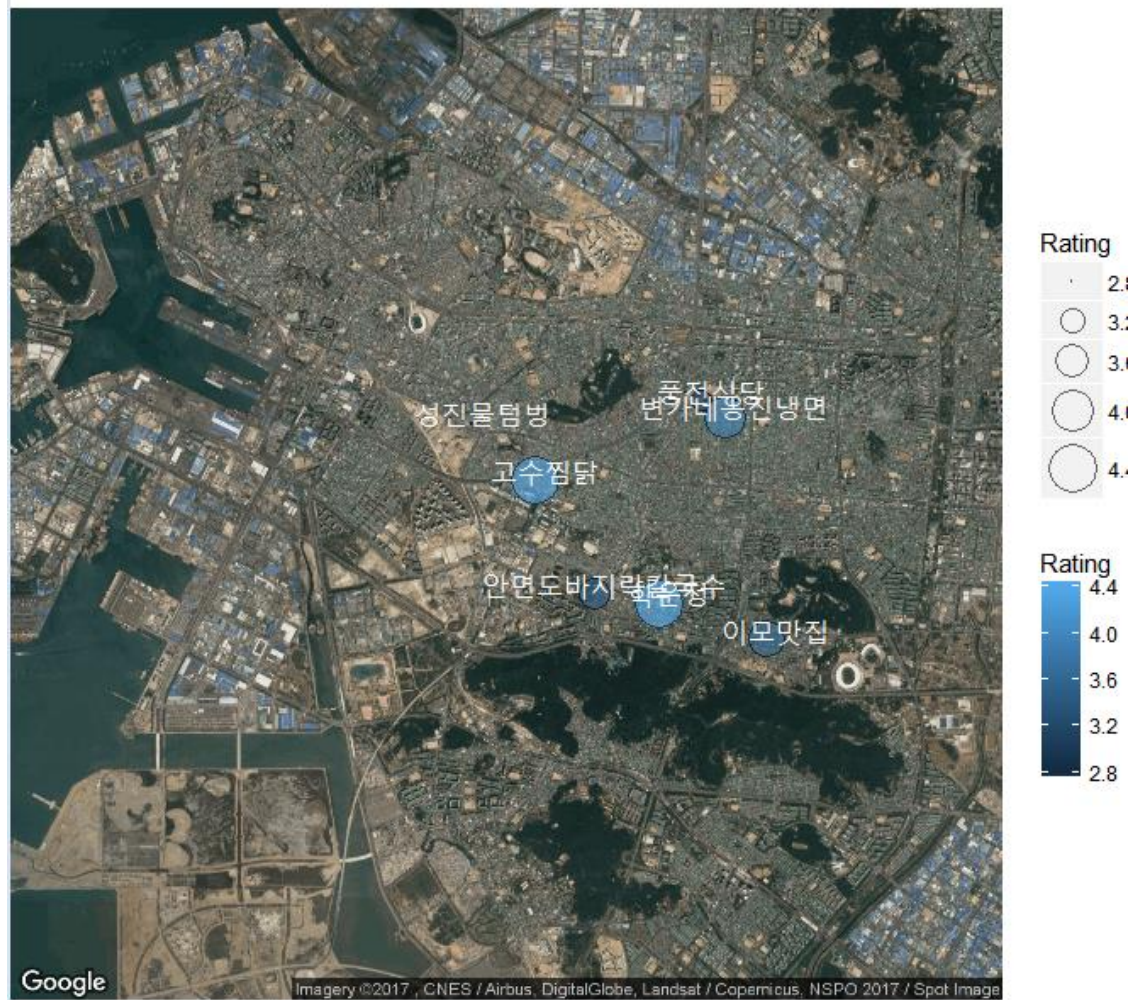
■ 인하대 근처 식당의 별점 및 위,경도 정보 이용 시각화

food.csv

Name	Lat	Lon	Rating
성진물텀벙	37.457	126.6494	2.8
고수찜닭	37.452	126.6563	4.4
안면도바지락칼국수	37.442	126.6629	3.5
학운정	37.4412	126.67	4.5
풍전식당	37.459	126.6747	3.1
변가네옹진냉면	37.4576	126.6772	4
이모맛집	37.438	126.682	3.7

```
loc = c(lon, lat)
map = get_map(loc, zoom=13, maptype='satellite')
p = ggmap(map, extent='device') +
  geom_point(
    aes(x=Lon, y=Lat, size=Rating, fill=Rating),
    alpha=0.6, pch=21, data=food) +
  scale_size(range=c(0, 10)) +
  geom_text(aes(x=Lon+0.001, y=Lat+0.001,
    label=Name), color="white", size=5, data=food)
p
```

■ 인하대 근처 식당의 별점 및 위,경도 정보 이용 시각화



- 빅데이터 연구에서 텍스트 분석의 중요성이 점차 커짐.
 - 최근 수년간 급속도로 성장한 소셜 네트워크 서비스의 영향
- 텍스트는 숫자와 달리 명목형 자료이므로 그래프로 표현하기 어려움이 따르므로 전처리 과정이 요구 됨.
 - 태그 클라우드 (tag cloud) - 단어의 노출 빈도수 계산 시각화
 - 워드 트리 (word tree) - 문장 내의 단어간 연결 구조 시각화

- 가장 대표적인 정보의 저장 단위
- 월드와이드웹 등장 이후 인류가 만들어 내는 텍스트의 양은 지속적이며 기하급수적으로 증가 추세
 - 인터넷에 존재하는 모든 텍스트를 읽기 위해서는 20만년 소요
- 문서 (documents)
 - 뉴스기사, 책, 소설
 - 이메일, 웹페이지, 블로그, 트위터
 - 태그, 코멘트
 - 컴퓨터 프로그램 소스, 로그 데이터
- 문서의 집합 (Collection of Documents)

■ 왜 텍스트를 시각화 하는가?

- 이해 (Understanding)
 - > 문서의 핵심 요지를 파악하기 위해서
- 그룹/분류 (Grouping)
 - > 전체를 조망하기 위한 군집을 만들거나 분류하기 위해
- 비교 (Compare)
 - > 문서들을 비교하거나 문서의 집합이 어떻게 시간에 따라 변화해왔는지 파악하기 위해서
- 연관성 (Correlate)
 - > 텍스트에 나타나는 패턴을 다른 데이터와 비교하기 위해서

■ 텍스트 시각화 방법

- 텍스트 데이터는 숫자와 달리 명목형 데이터 => 표현이 어려움
- 몇 가지 전처리 과정을 거쳐서 시각화 할 수 있는 형태로 변환
 - 단어의 빈도수 계산
 - 문장 내에서 단어들 사이의 관계나 구조를 표현

■ 태그 클라우드 (Tag Cloud)

- 문서에 등장하는 단어의 빈도수를 분석하여 많이 사용된 단어를 강조하여 보여주는 방법
 - 태그의 빈도수 측정
 - 가장 많이 사용한 태그의 글자 크기를 키워서 강조

BY CATHERINE RAMPPELL SEPTEMBER 9, 2009 11:41 PM



INHA UNIVERSITY

■ Wordcloud in R

```
library(tm)
library(SnowballC)
library(wordcloud)

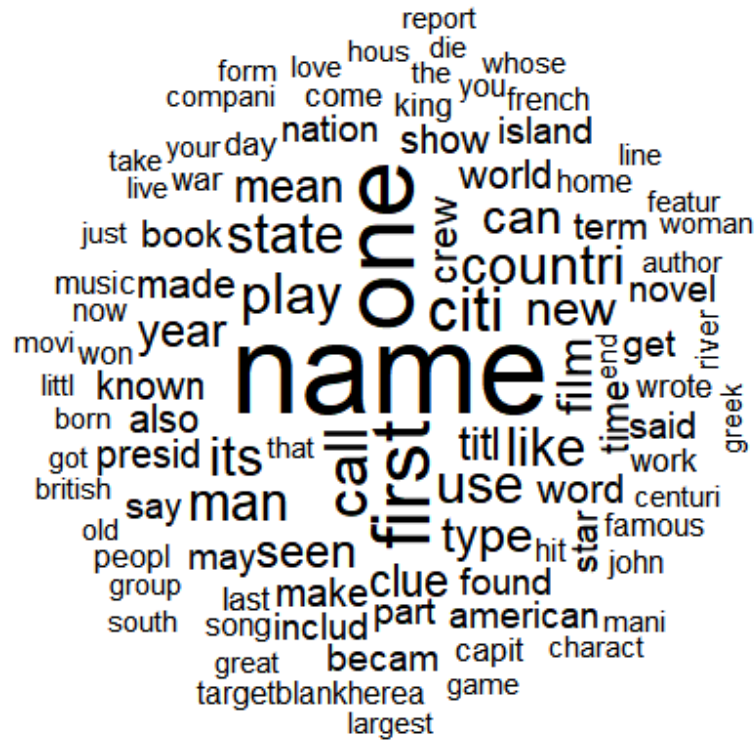
jeopQ <- read.csv('JEOPARDY_CSV.csv', stringsAsFactors = FALSE)

### Data Cleaning
jeopCorpus <- Corpus(VectorSource(jeopQ$Question))
jeopCorpus <- tm_map(jeopCorpus, removePunctuation) ## . , ? ! ()
jeopCorpus <- tm_map(jeopCorpus, removeWords,
                     c("the", "this", "The", "This", stopwords('english')))
                     ## stopwords("en")
jeopCorpus <- tm_map(jeopCorpus, stemDocument) ## went -> go

wordcloud(jeopCorpus, max.words = 100, random.order = FALSE)
```

<https://www.r-bloggers.com/building-wordclouds-in-r/>

■ Wordcloud in R



<https://www.r-bloggers.com/building-wordclouds-in-r/>

References

- W. N. Venables, D. M. Smith and the R Core Team, An Introduction to R. (Notes on R: A Programming Environment for Data Analysis and Graphics)
- R 그래픽스, 자유아카데미
- Manuals for ff and ffbase packages
- 장원철, (2016). 데이터 시각화와 탐색적 자료분석, 서울대학교 통계연구소
- ggplot2 documents: <http://ggplot.yhathq.com/docs/>
- R을 활용한 데이터 시각화, 인사이트