

데이터 분석

R Coding

2021.01.21 김현진



What is the **Data**??

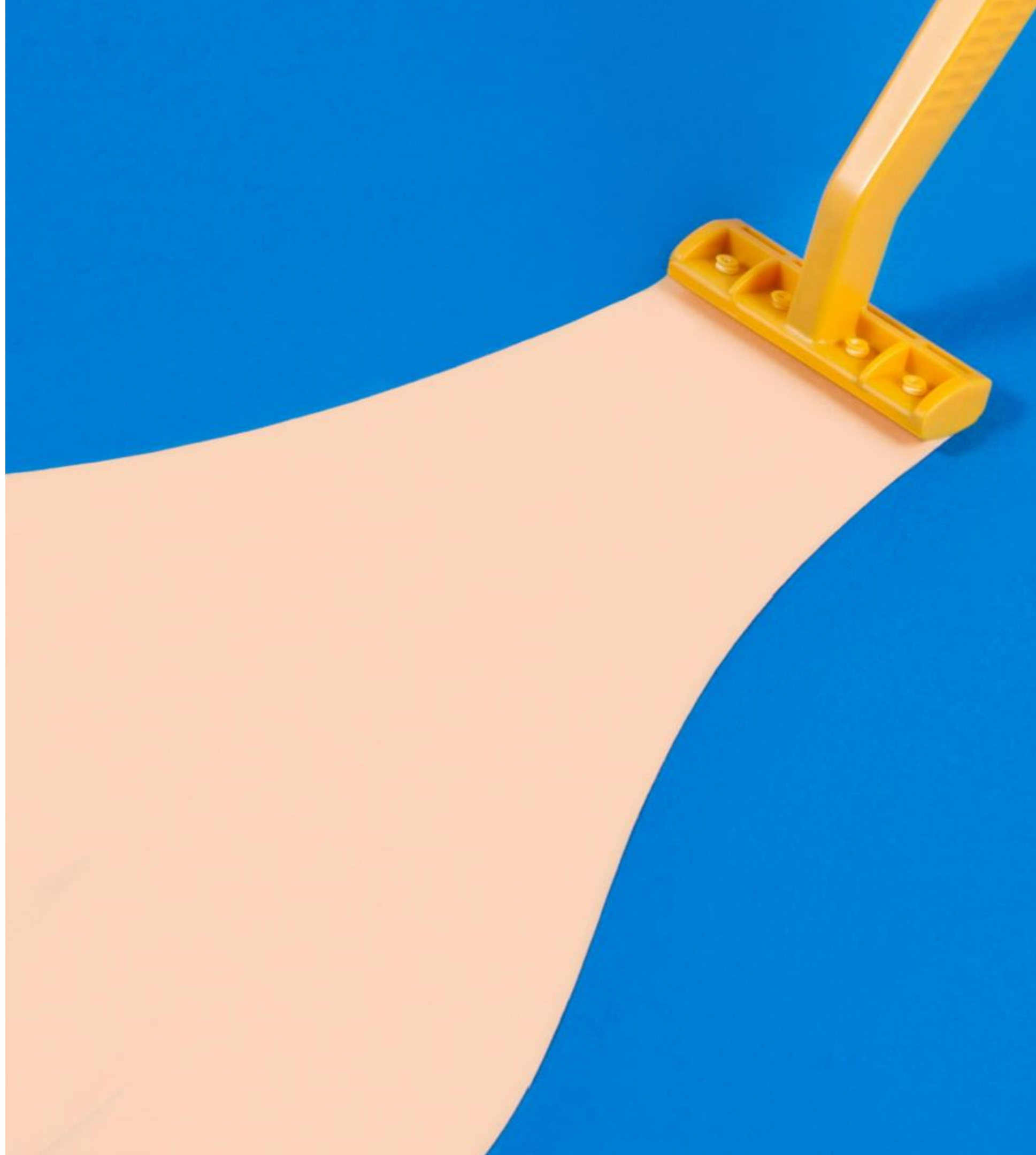
이론을 세우는데 기초가 되는 사실, 또는 자료
관찰이나 실험, 조사로 얻은 사실이나 자료
정보가 아니라 자료임에 유의하자.

(자료를 가공해 얻은것이 정보이다.)

데이터의 가치

지금 이순간에도 쌓이고 있는 데이터
데이터를 활용하여 더 빠르고 정확하게 정보를 도출.
데이터를 어떻게 가공하냐에 따라

그 데이터는 모두가 원하는 데이터가 될 수 있음.





빅데이터

" 데이터 베이스로는 처리하기 어려울 정도의
방대한 양의 데이터 "

데이터 베이스 관리 도구의 능력을 넘어서는
대량의 데이터로부터 가치를 추출하고 결과를 분석하는 기술

빅데이터의 3 요소

1

Volume (양)

일반적으로 수십 테라 바이트, 수십 페타바이트 이상의 크기를 가집니다.

Tb = 1000 Gb

2

Velocity (속도)

데이터가 생성되고, 저장되며, 시각화 되는 과정

3

Variety (다양성)

정형 데이터, 반정형 데이터, 비정형 데이터 등

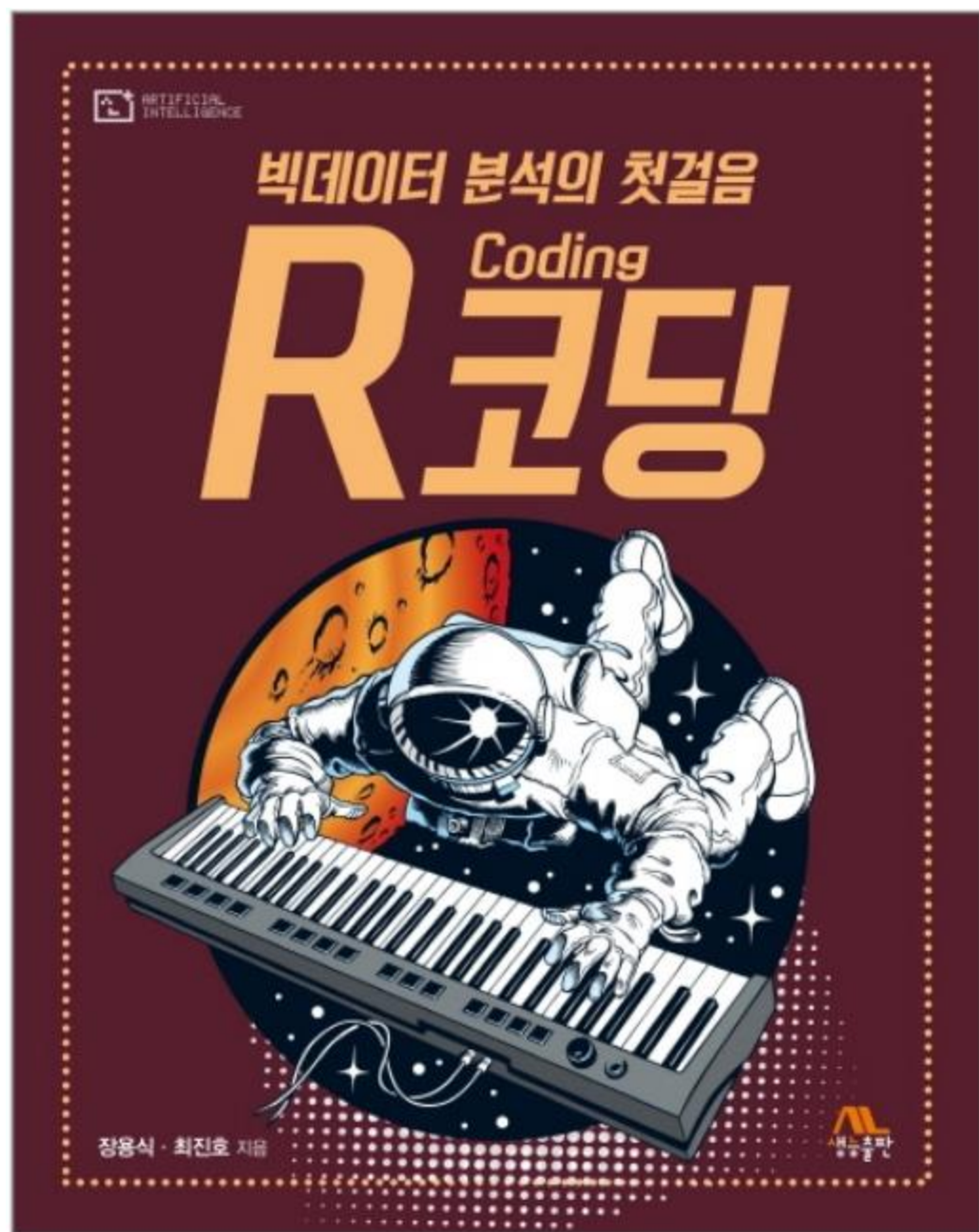
정형 : 고정된 필드에 저장된 데이터

반정형 : XML, HTML, 로그 같은 일정한 규칙

을 가진 데이터

비정형 : 사진, 오디오 데이터

R Coding



Vector (벡터)

array (배열)

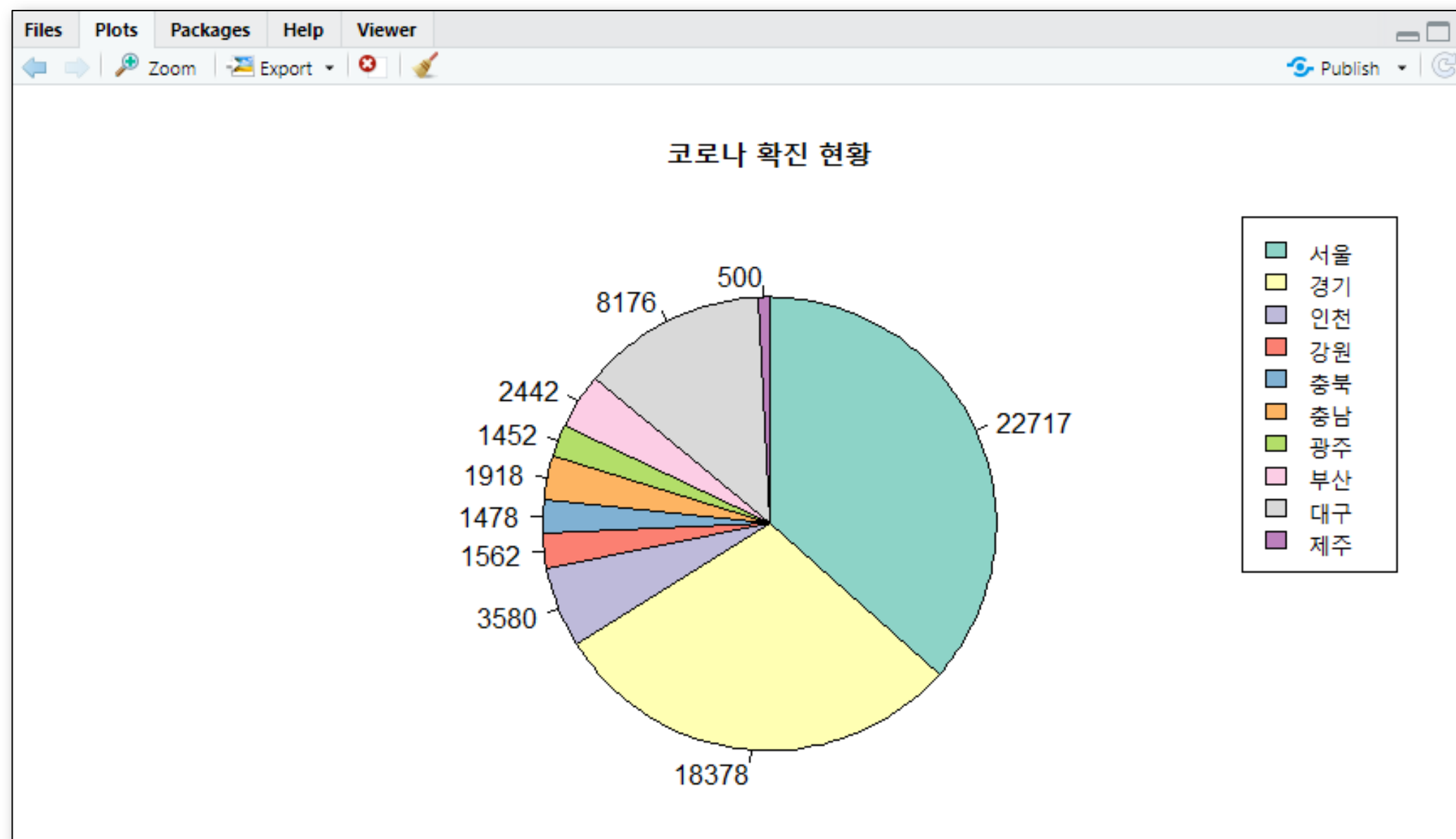
list (리스트)

data frame (데이터 프레임)

...

차트 프로그래밍

다양한 차트를 통해 데이터를 분석하고 결과를 전달한다.



파이 차트

R 코드 파이 차트

```
library(RColorBrewer)
library(graphics)
city <- c("서울", "경기", "인천", "강원", "충북", "충남", "광주", "부산", "대구", "제주")
COVID <- c(22717, 18378, 3580, 1562, 1478, 1918, 1452, 2442, 8176, 500)
color <- brewer.pal(10, name='Set3')
pie(COVID, labels=COVID, col=color, clockwise = TRUE, main = "코로나 확진 현황")
legend("topright", legend = city, fill = color, border = "black")
|
```

graphics 패키지 : 차트 생성을 위한 패키지

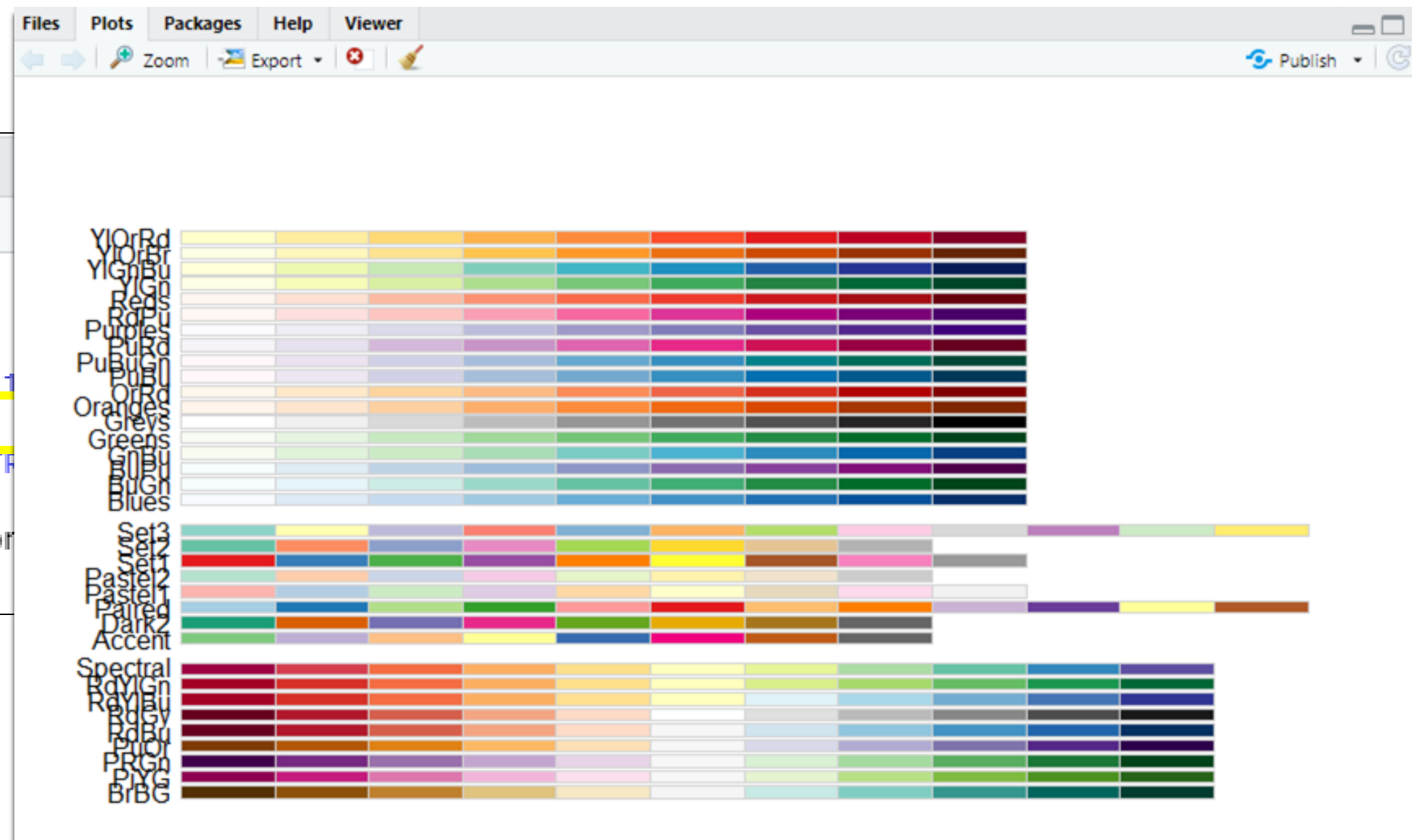
RColorBrewer 패키지 : 팔레트 색을 넣기 위한 패키지

R 코드 파이 차트

```
library(RColorBrewer)
library(graphics)
city <- c("서울", "경기", "인천", "강원", "충북", "충남", "광주", "부산", "대구", "제주")
COVID <- c(22717, 18378, 3580, 1562, 1478, 1918, 1452, 2442, 8176, 500)
color <- brewer.pal(10, name="Set3")
pie(COVID, labels=COVID, col=color, clockwise = TRUE, main = "코로나 확진 현황")
legend("topright", legend = city, fill = color, border = "black")
|
```

R 코드 파이 차트

```
library(RColorBrewer)
library(graphics)
city <- c("서울", "경기", "인천", "강원", "충북",
COVID <- c(22717, 18378, 3580, 1562, 1478, 1918, 1
color <- brewer.pal(10, name='Set3')
pie(COVID, labels=COVID, col=color, clockwise = T)
legend("topright", legend = city, fill = color, bor
```



R 코드 파이 차트

```
library(RColorBrewer)
library(graphics)
city <- c("서울", "경기", "인천", "강원", "충북", "충남", "광주", "부산", "대구", "제주")
COVID <- c(22717, 18378, 3580, 1562, 1478, 1918, 1452, 2442, 8176, 500)
color <- brewer.pal(10.name='Set3')
pie(COVID, labels=COVID, col=color, clockwise = TRUE, main = "코로나 확진 현황")
legend("topright", legend = city, fill = color, border = "black")
```

pie() : 파이 차트 생성 함수

legend() : 범례 생성 함수

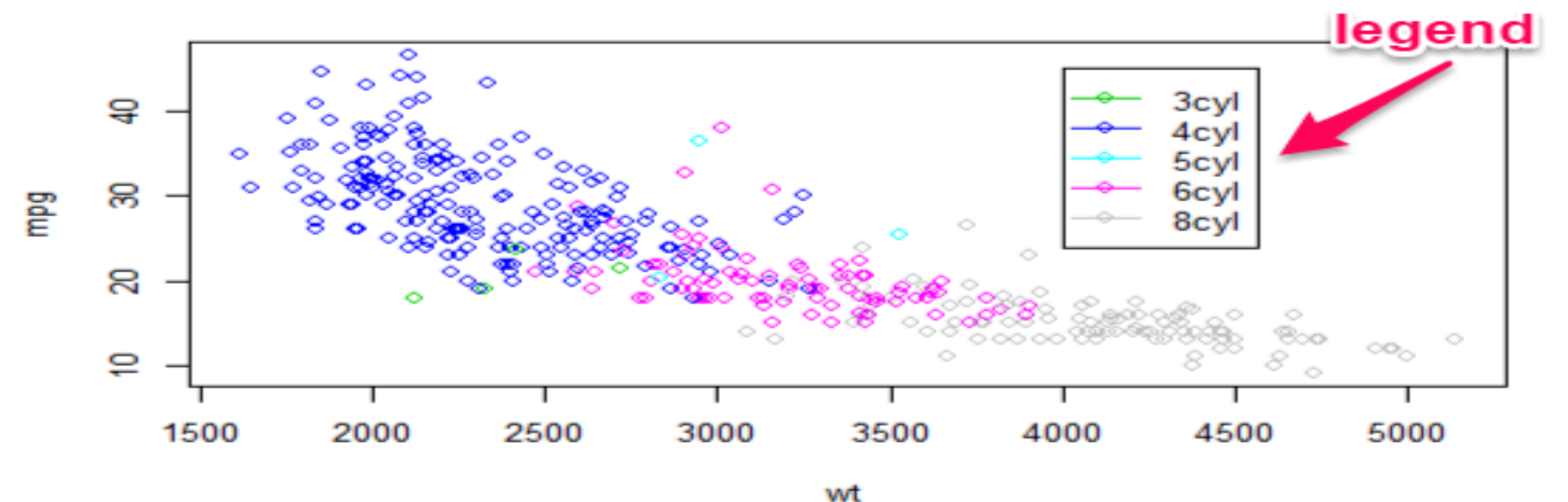
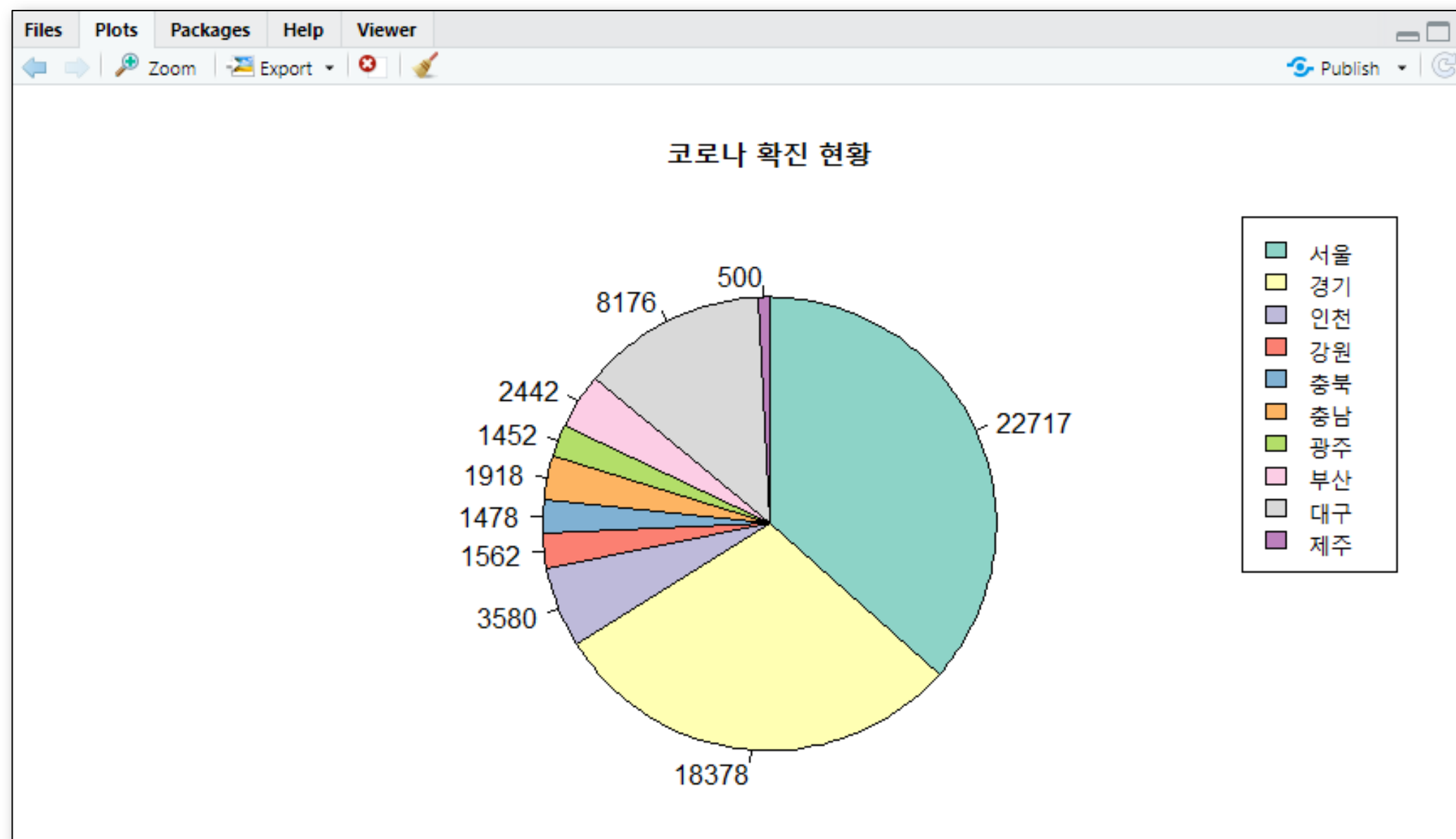


차트 프로그래밍

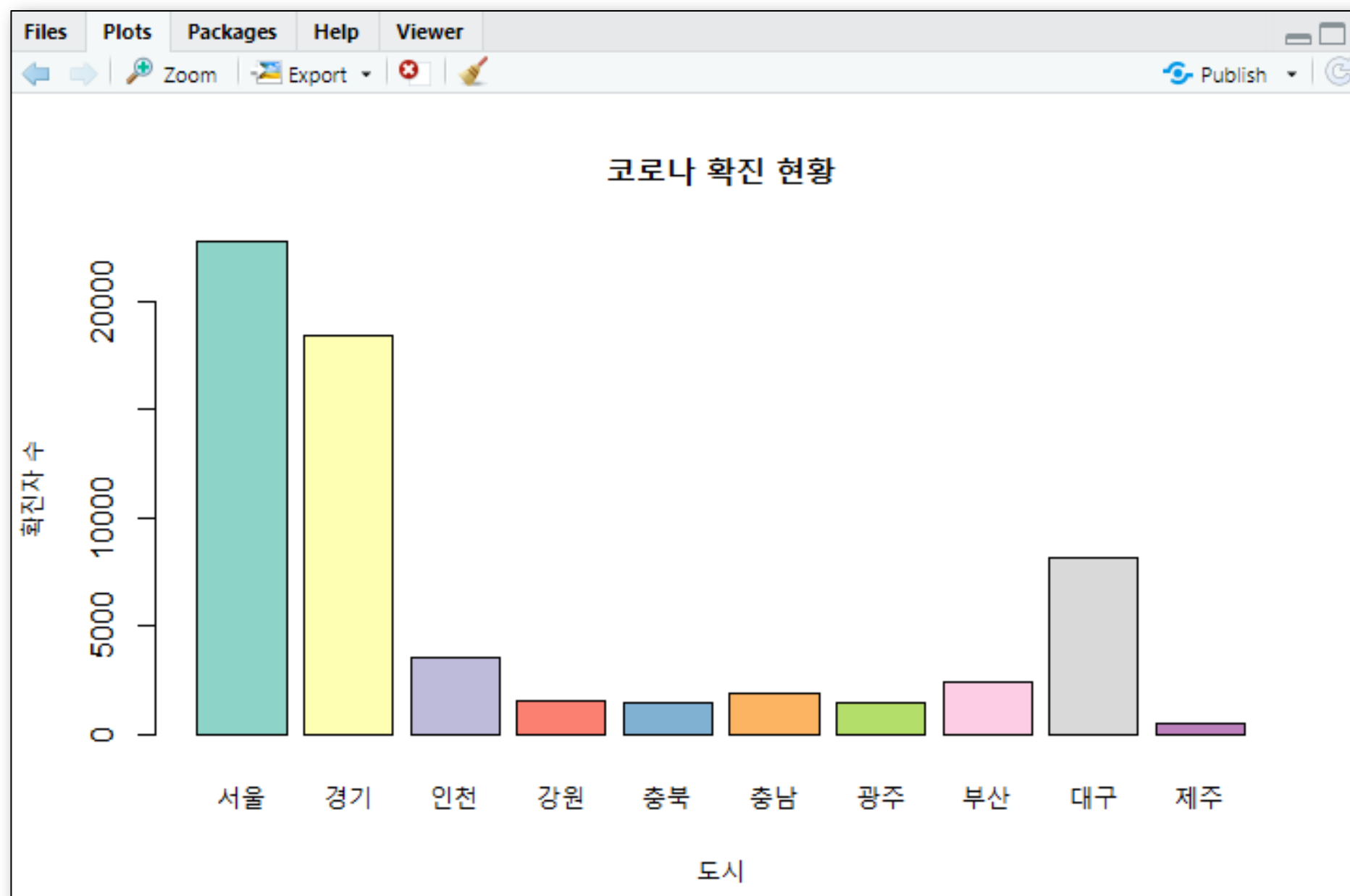
다양한 차트를 통해 데이터를 분석하고 결과를 전달한다.



파이 차트

차트 프로그래밍

다양한 차트를 통해 데이터를 분석하고 결과를 전달한다.



바 차트

R 코드 바 차트

```
library(RColorBrewer)
library(graphics)
city <- c("서울", "경기", "인천", "강원", "충북", "충남", "광주", "부산", "대구", "제주")
COVID <- c(22717, 18378, 3580, 1562, 1478, 1918, 1452, 2442, 8176, 500)
color <- brewer.pal(10,name='Set3')
barplot(COVID,names.arg =city,main ="코로나 확진 현황",col=color,xlab="도시",ylab ="확진자 수")
```

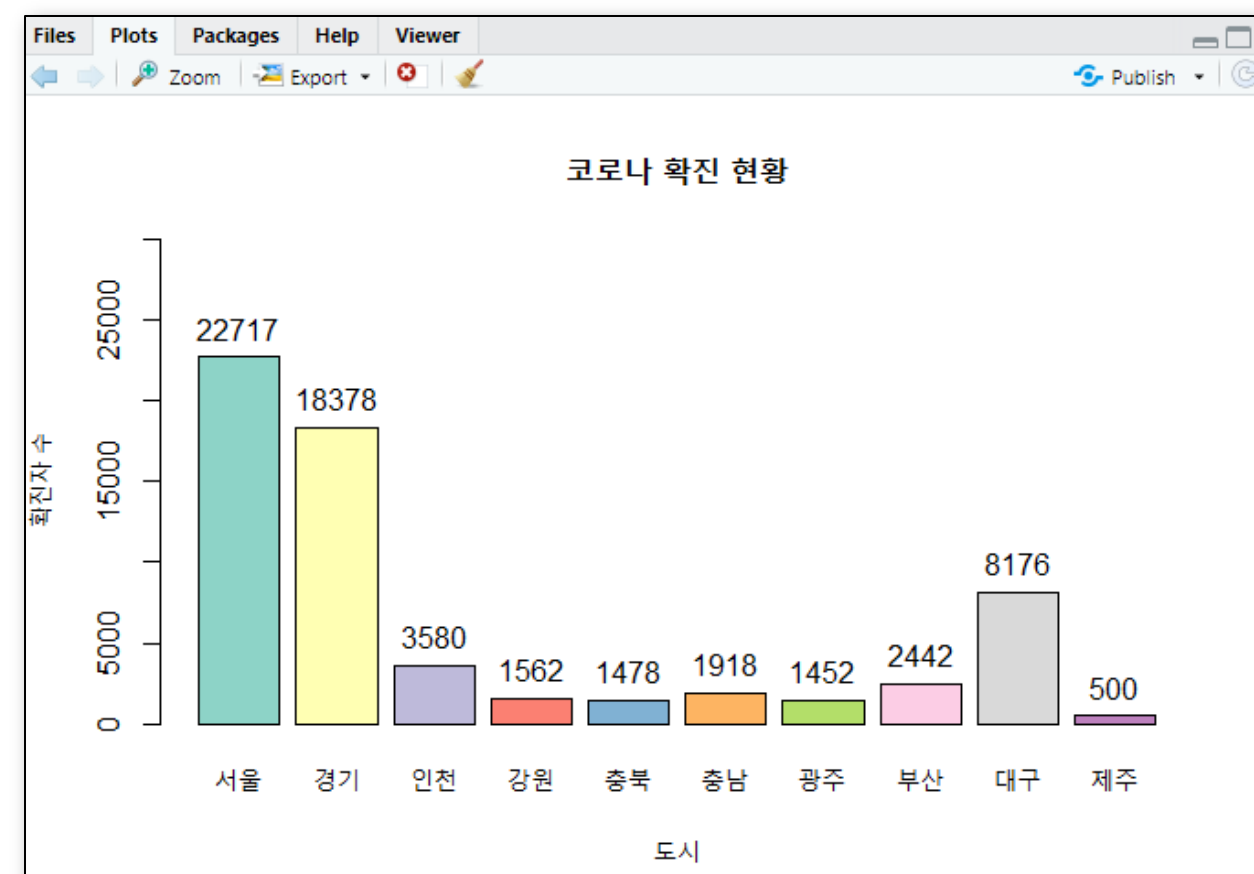
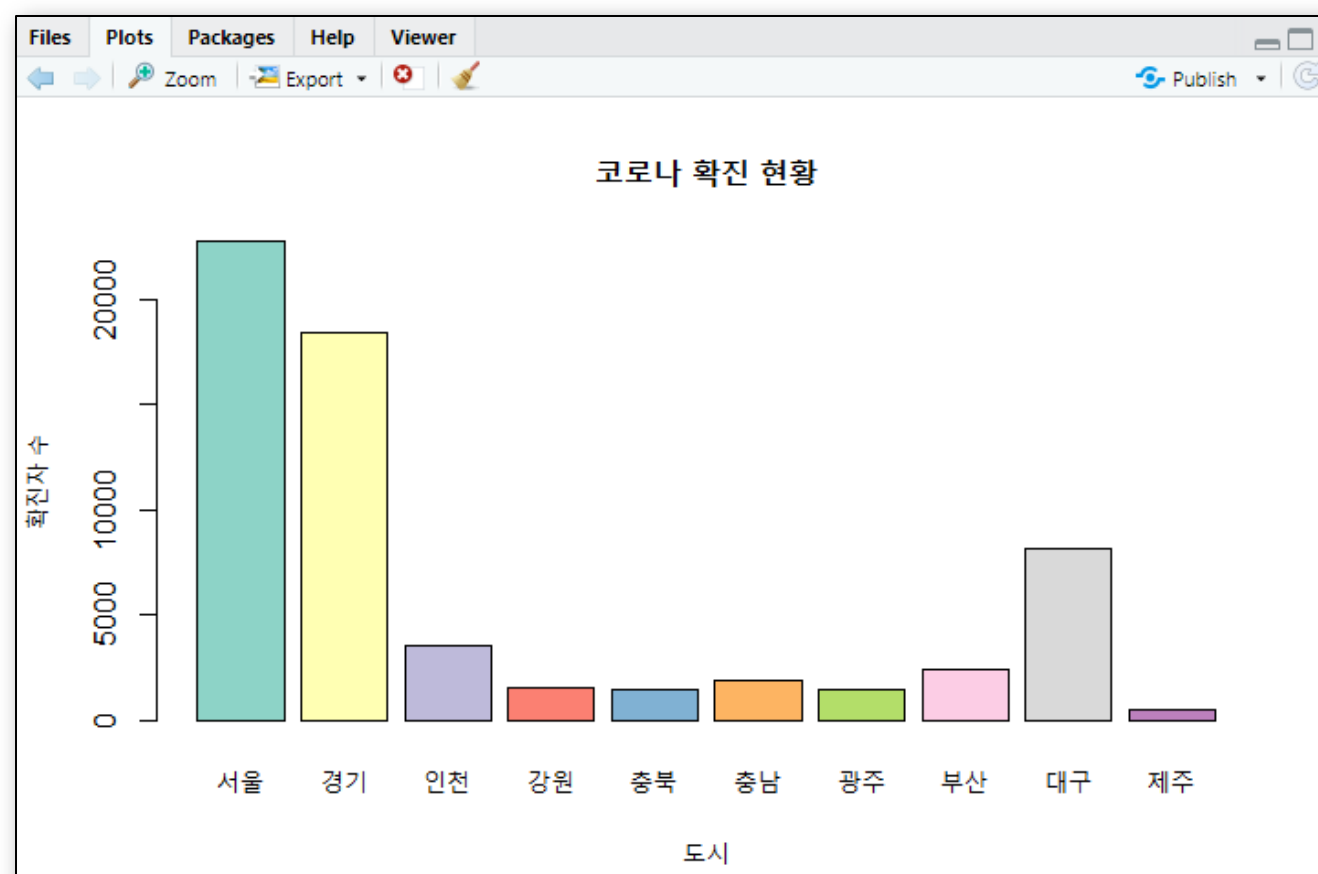
barplot() : 파이 차트 생성 함수

```
R* x
Source on Save
library(RColorBrewer)
library(graphics)
city <- c("서울", "경기", "인천", "강원", "충북", "충남", "광주", "부산", "대구", "제주")
COVID <- c(22717, 18378, 3580, 1562, 1478, 1918, 1452, 2442, 8176, 500)
color <- brewer.pal(10,name='Set3')
bar <- barplot(COVID,names.arg =city,main ="코로나 확진 현황",col=color,xlab="도시",ylab ="확진자 수",ylim=c(0,30000))
text(x=bar, y=COVID, labels=COVID,pos = 3)
```

text() : 차트에 문자를 추가하는 함수

차트 프로그래밍

다양한 차트를 통해 데이터를 분석하고 결과를 전달한다.



바 차트

텍스트 마이닝

" 비정형 텍스트에서 의미 있는 정보를 추론하는 기술 "

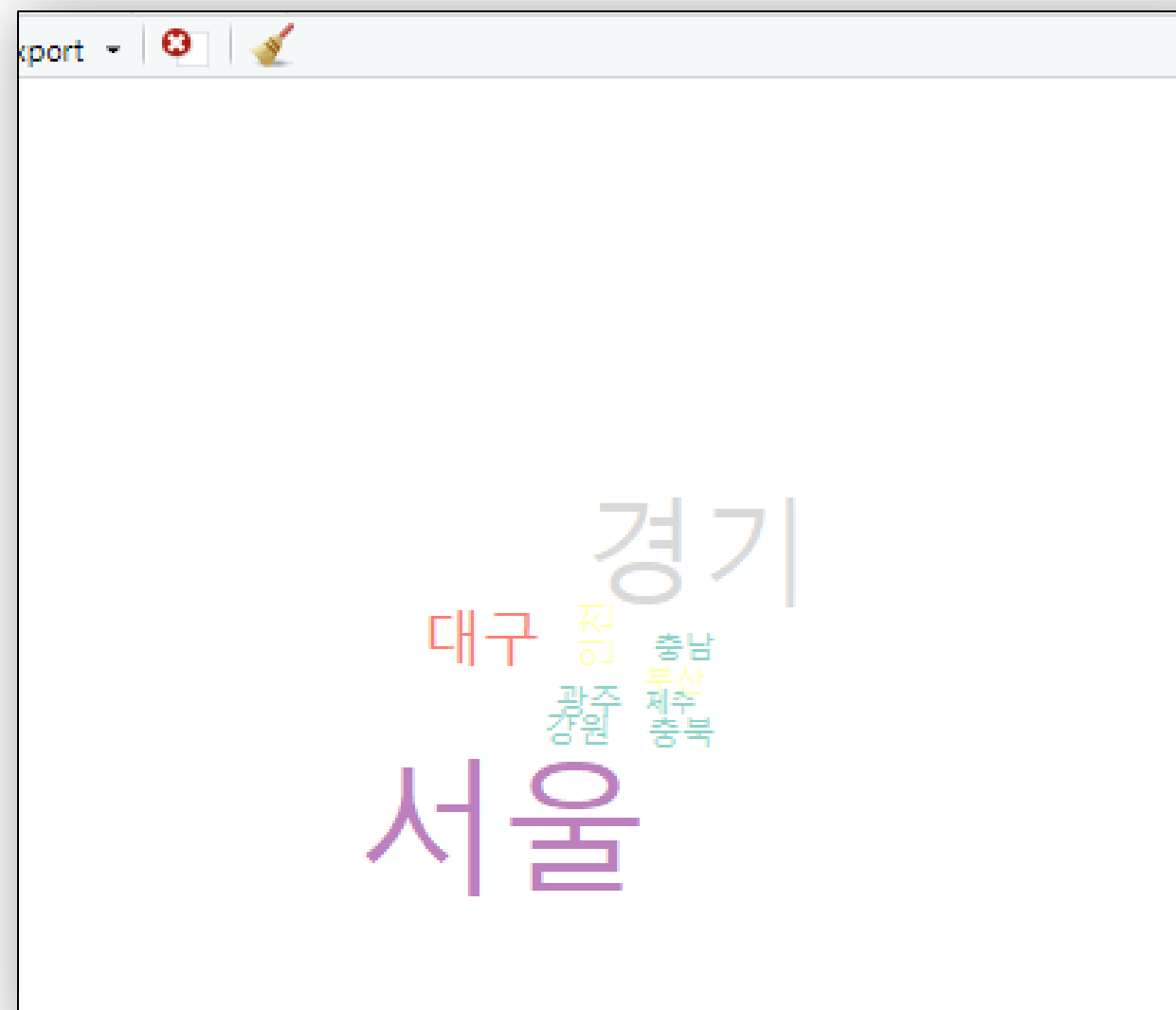
정보를 추출하고 문서를 요약할 때 주로 활용하는 기술로 자주 활용되는 단어일 수록 폰트의 크기를 크게 배치하여 강조



" word cloud "

워드 클라우드 프로그래밍

단어를 분류하여 빈도를 한눈에 보기 쉽게 한다.



워드 클라우드

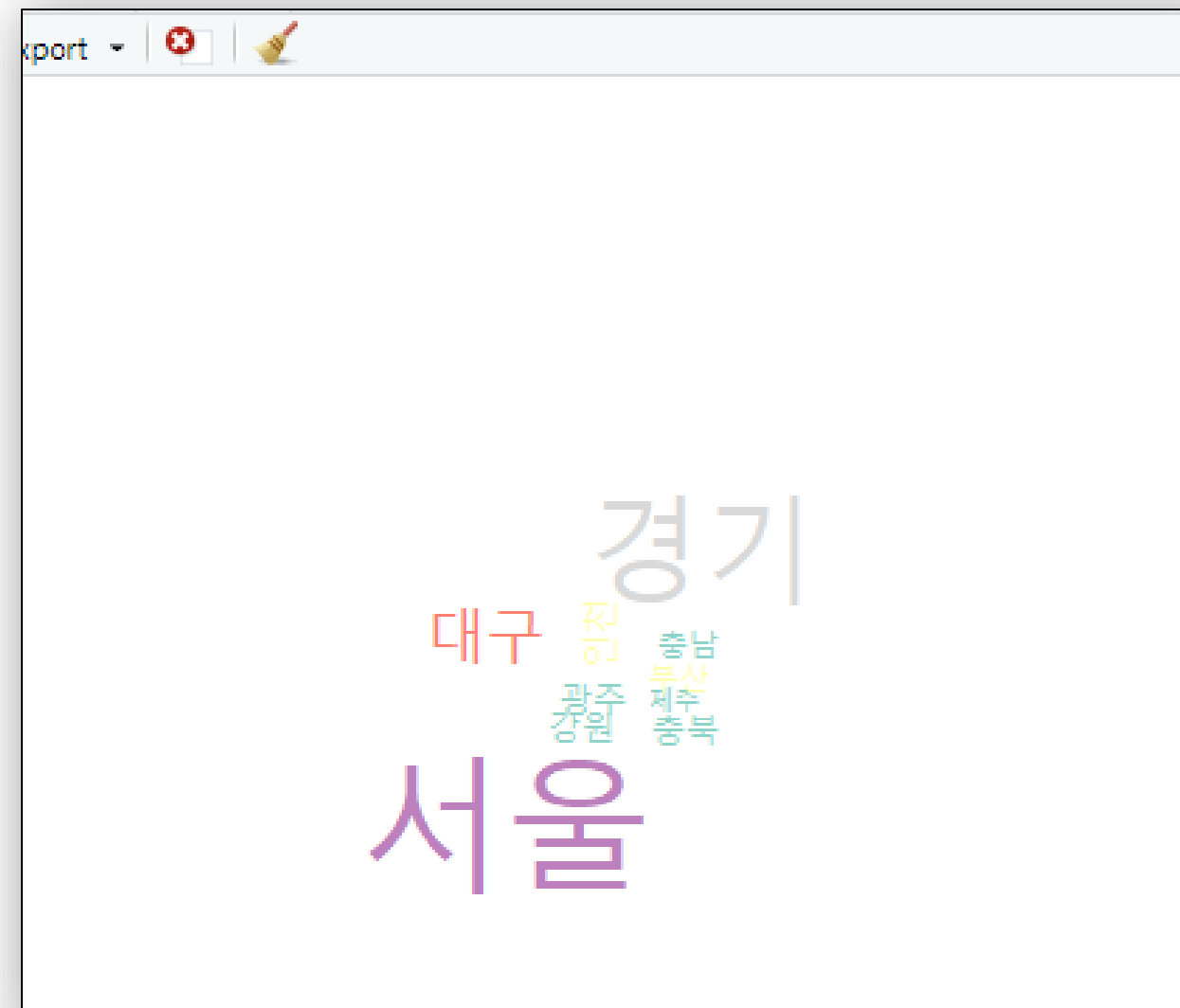
R 코드 워크 클라우드

```
R* x
Source on Save
library(RColorBrewer)
library(wordcloud)
DAY <- c("서울", "경기", "인천", "강원", "충북", "충남", "광주", "부산", "대구", "세주")
COVID <- c(22717, 18378, 3580, 1562, 1478, 1018, 1452, 2442, 8176, 500)
wordcloud(city, freq = COVID, colors = rainbow(10))
```

wordcloud 패키지 : 단어를 분류하여 빈도를 한눈에 보기 쉽게 하기 위한 패키지

워드 클라우드 프로그래밍

단어를 분류하여 빈도를 한눈에 보기 쉽게 한다.



워드 클라우드

THANK
YOU