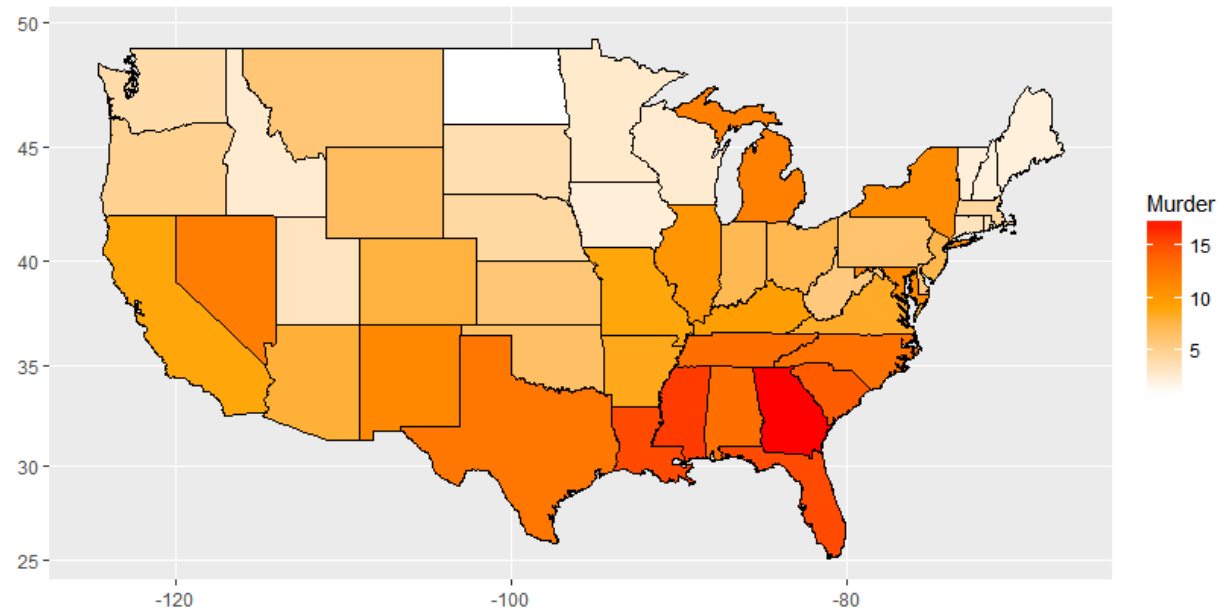


# 지도 시각화



## 단계 구분도(Choropleth Map)

- 지역별 통계치를 색깔의 차이로 표현한 지도
- 인구나 소득 같은 특성이 지역별로 얼마나 다른지 쉽게 이해할 수 있음

# 11-1. 미국 주별 강력 범죄율 단계 구분도 만들기

패키지 준비하기

```
install.packages("ggiraphExtra")
```

```
library(ggiraphExtra)
```

## 미국 주별 범죄 데이터 준비하기

```
str(USArrests)
```

```
## 'data.frame':    50 obs. of  4 variables:
## $ Murder   : num  13.2 10 8.1 8.8 9 7.9 3.3 5.9 15.4 17.4 ...
## $ Assault  : int  236 263 294 190 276 204 110 238 335 211 ...
## $ UrbanPop: int   58 48 80 50 91 78 77 72 80 60 ...
## $ Rape     : num  21.2 44.5 31 19.5 40.6 38.7 11.1 15.8 31.9 25.8 ...
```

```
head(USArrests)
```

```
##           Murder Assault UrbanPop Rape
## Alabama      13.2      236        58 21.2
## Alaska       10.0      263        48 44.5
## Arizona       8.1      294        80 31.0
## Arkansas      8.8      190        50 19.5
## California    9.0      276        91 40.6
## Colorado      7.9      204        78 38.7
```

```
library(tibble)
```

```
# 행 이름을 state 변수로 바꿔 데이터 프레임 생성
```

```
crime <- rownames_to_column(USArrests, var = "state")
```

```
# 지도 데이터와 동일하게 맞추기 위해 state의 값을 소문자로 수정
```

```
crime$state <- tolower(crime$state)
```

```
str(crime)
```

```
## 'data.frame':    50 obs. of  5 variables:  
## $ state      : chr  "alabama" "alaska" "arizona" "arkansas" ...  
## $ Murder     : num  13.2 10 8.1 8.8 9 7.9 3.3 5.9 15.4 17.4 ...  
## $ Assault    : int   236 263 294 190 276 204 110 238 335 211 ...  
## $ UrbanPop   : int    58 48 80 50 91 78 77 72 80 60 ...  
## $ Rape       : num   21.2 44.5 31 19.5 40.6 38.7 11.1 15.8 31.9 25.8 ...
```

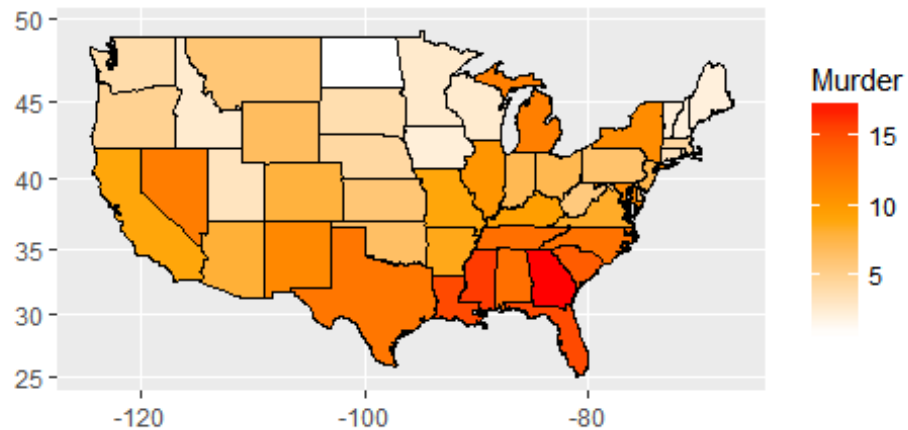
## 미국 주 지도 데이터 준비하기

```
library(ggplot2)
states_map <- map_data("state")
str(states_map)

## 'data.frame':    15537 obs. of  6 variables:
##  $ long      : num  -87.5 -87.5 -87.5 -87.5 -87.6 ...
##  $ lat       : num   30.4 30.4 30.4 30.3 30.3 ...
##  $ group     : num    1 1 1 1 1 1 1 1 1 1 ...
##  $ order     : int    1 2 3 4 5 6 7 8 9 10 ...
##  $ region    : chr   "alabama" "alabama" "alabama" "alabama" ...
##  $ subregion: chr    NA NA NA NA ...
```

## 단계 구분도 만들기

```
ggChoropleth(data = crime,           # 지도에 표현할 데이터
              aes(fill = Murder,     # 색깔로 표현할 변수
                  map_id = state),   # 지역 기준 변수
              map = states_map)      # 지도 데이터
```



## 인터랙티브 단계 구분도 만들기

```
ggChoropleth(data = crime,           # 지도에 표현할 데이터
              aes(fill = Murder,      # 색깔로 표현할 변수
                  map_id = state),    # 지역 기준 변수
              map = states_map,       # 지도 데이터
              interactive = T)        # 인터랙티브
```



## 11-2. 대한민국 시도별 인구, 결핵 환자 수 단계 구분도 만들기

### 대한민국 시도별 인구 단계 구분도 만들기

#### 패키지 준비하기

```
install.packages("stringi")  
  
install.packages("devtools")  
devtools::install_github("cardiomoon/kormaps2014")  
  
library(kormaps2014)
```

## 대한민국 시도별 인구 데이터 준비하기

```
str(changeCode(korpop1))
```

```
## 'data.frame':   17 obs. of  25 variables:
##  $ C 행정구역별_읍면동      : chr  "'11" "'21" "'22" "'23" ...
##  $ 행정구역별_읍면동      : chr  "서울특별시" "부산광역시" "대구광역시" "인천광역시" ...
##  $ 시점                    : chr  "2015" "2015" "2015" "2015" ...
##  $ 총인구_명              : chr  "9904312" "3448737" "2466052" "2890451" ...
##  $ 남자_명               : chr  "4859535" "1701347" "1228511" "1455017" ...
##  $ 여자_명               : chr  "5044777" "1747390" "1237541" "1435434" ...
##  $ 내국인_계_명          : chr  "9567196" "3404667" "2436770" "2822601" ...
##  $ 내국인_남자_명        : chr  "4694317" "1675339" "1211219" "1414793" ...
##  $ 내국인_여자_명        : chr  "4872879" "1729328" "1225551" "1407808" ...
##  $ 외국인_계_명          : chr  "337116" "44070" "29282" "67850" ...
##  $ 외국인_남자_명        : chr  "165218" "26008" "17292" "40224" ...
##  $ 외국인_여자_명        : chr  "171898" "18062" "11990" "27626" ...
##  $ 가구_계_가구          : chr  "3914820" "1348315" "937573" "1066297" ...
##  $ 일반가구_가구         : chr  "3784490" "1335900" "928528" "1045417" ...
##  $ 집단가구_가구         : chr  "2261" "686" "574" "713" ...
```

```
## $ 외국인가구_가구      : chr  "128069" "11729" "8471" "20167" ...
## $ 주택_계_호           : chr  "2793244" "1164352" "738100" "942244" ...
## $ 단독주택_호         : chr  "355039" "225697" "155801" "102914" ...
## $ 아파트_호           : chr  "1636896" "738068" "509068" "577346" ...
## $ 연립주택_호         : chr  "117235" "32120" "9381" "21589" ...
## $ 다세대주택_호       : chr  "654372" "154253" "53098" "232346" ...
## $ 비거주용_건물내_주택_호 : chr  "29702" "14214" "10752" "8049" ...
## $ 주택이외의_거처_호   : chr  "150951" "50810" "15304" "39964" ...
## $ c 행정구역별        : chr  "11" "21" "22" "23" ...
## $ code                 : chr  "11" "21" "22" "23" ...
```

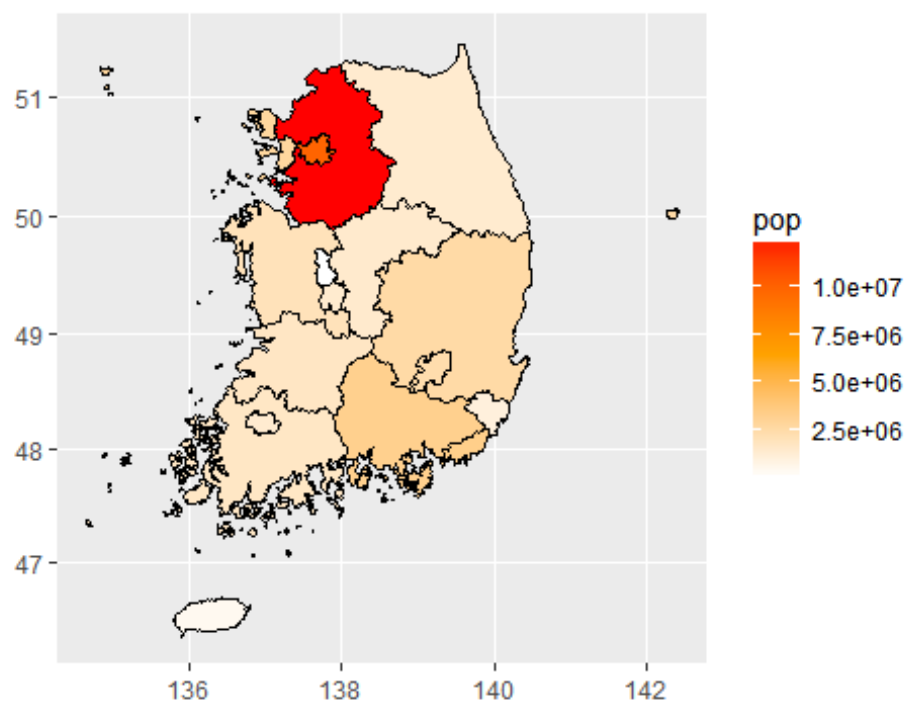
```
library(dplyr)
korp1 <- rename(korp1,
               pop = 총인구_명,
               name = 행정구역별_읍면동)

str(changeCode(korp1))

## 'data.frame':    8831 obs. of  15 variables:
## $ id          : chr  "0" "0" "0" "0" ...
## $ long        : chr  "137.774352627938" "137.779270931415" "137.780545929866" "137.814
504843261" ...
## $ lat         : chr  "50.6883045072662" "50.6899249663447" "50.6900586920365" "50.6937
941360883" ...
## $ order       : chr  "1" "2" "3" "4" ...
## $ hole        : chr  "FALSE" "FALSE" "FALSE" "FALSE" ...
## $ piece       : chr  "1" "1" "1" "1" ...
## $ group       : chr  "0.1" "0.1" "0.1" "0.1" ...
## $ SP_ID       : chr  "0" "0" "0" "0" ...
## $ SIDO_CD     : chr  "11" "11" "11" "11" ...
## $ SIDO_NM     : chr  NA NA NA NA ...
## $ BASE_YEAR   : chr  "2014" "2014" "2014" "2014" ...
## $ name        : chr  "서울특별시" "서울특별시" "서울특별시" "서울특별시" ...
## $ name1       : chr  NA NA NA NA ...
## $ region      : chr  "11" "11" "11" "11" ...
## $ code        : chr  "11" "11" "11" "11" ...
```

## 단계 구분도 만들기

```
ggChoropleth(data = korpop1,           # 지도에 표현할 데이터
              aes(fill = pop,          # 색깔로 표현할 변수
                  map_id = code,       # 지역 기준 변수
                  tooltip = name),     # 지도 위에 표시할 지역명
              map = kormap1,           # 지도 데이터
              interactive = T)         # 인터랙티브
```



## 대한민국 시도별 결핵 환자 수 단계 구분도 만들기

```
str(changeCode(tbc))
```

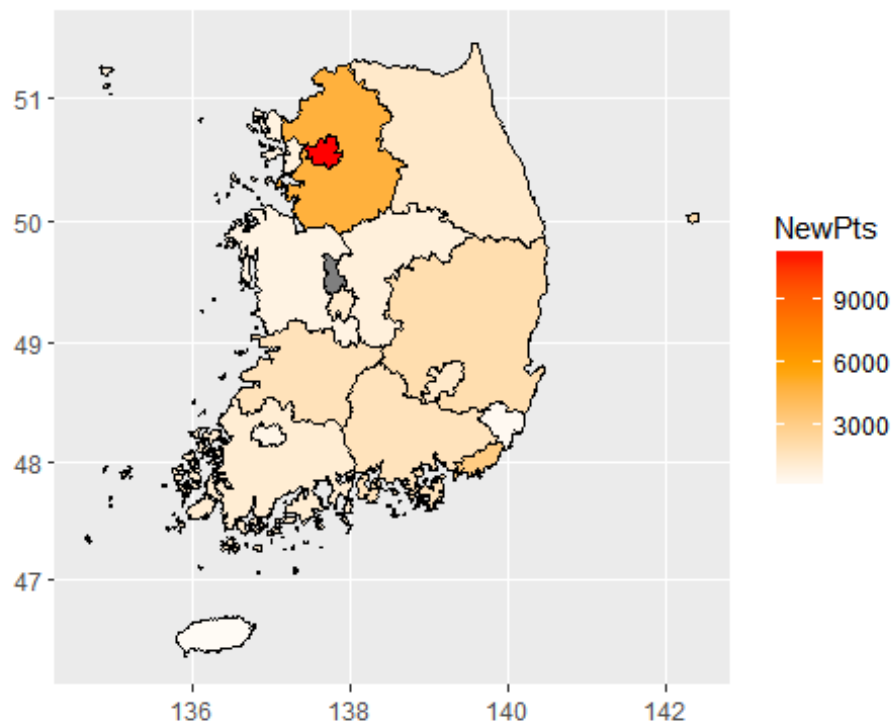
```
## 'data.frame':    255 obs. of  5 variables:
## $ name1 : chr  "강원" "경기" "경남" "경북" ...
## $ code  : chr  "32" "31" "38" "37" ...
## $ name   : chr  "강원도" "경기도" "경상남도" "경상북도" ...
## $ year   : chr  "2001" "2001" "2001" "2001" ...
## $ NewPts : chr  "1396" "4843" "1749" "2075" ...
```

```

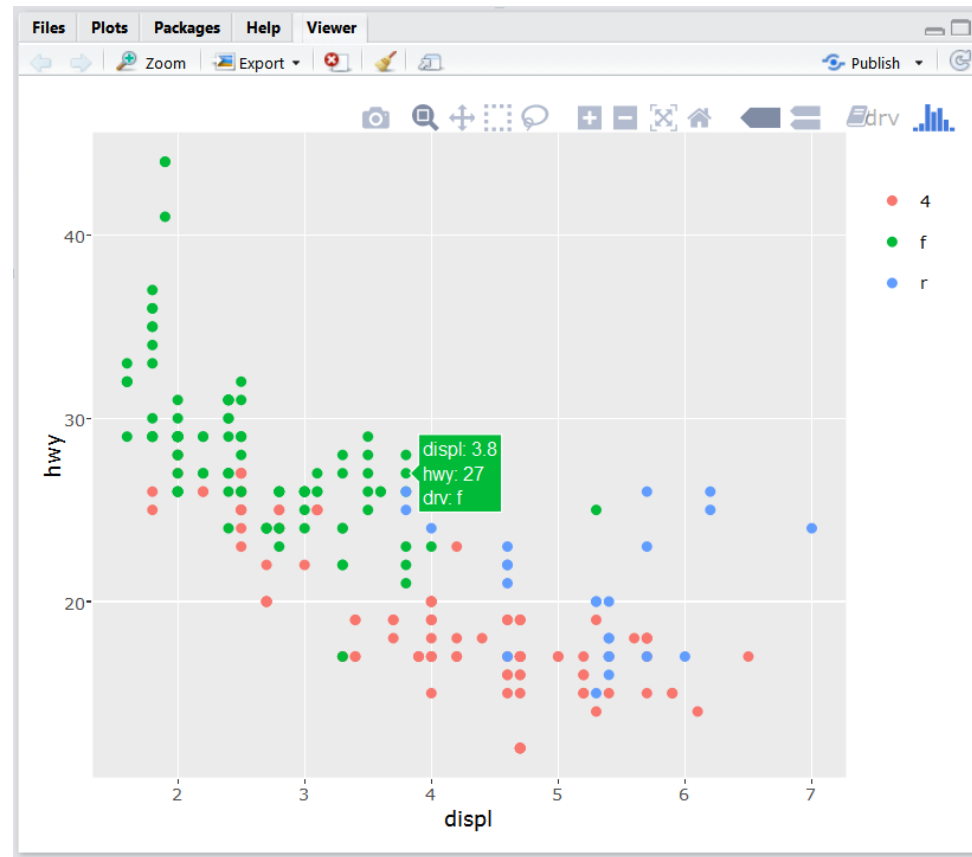
ggChoropleth(data = tbc,
  aes(fill = NewPts,
    map_id = code,
    tooltip = name),
  map = kormap1,
  interactive = T)

```

# 지도에 표현할 데이터  
 # 색깔로 표현할 변수  
 # 지역 기준 변수  
 # 지도 위에 표시할 지역명  
 # 지도 데이터  
 # 인터랙티브



# 인터랙티브 그래프





# 12-1. plotly 패키지로 인터랙티브 그래프 만들기

## 인터랙티브 그래프 만들기

### 패키지 준비하기

```
install.packages("plotly")
```

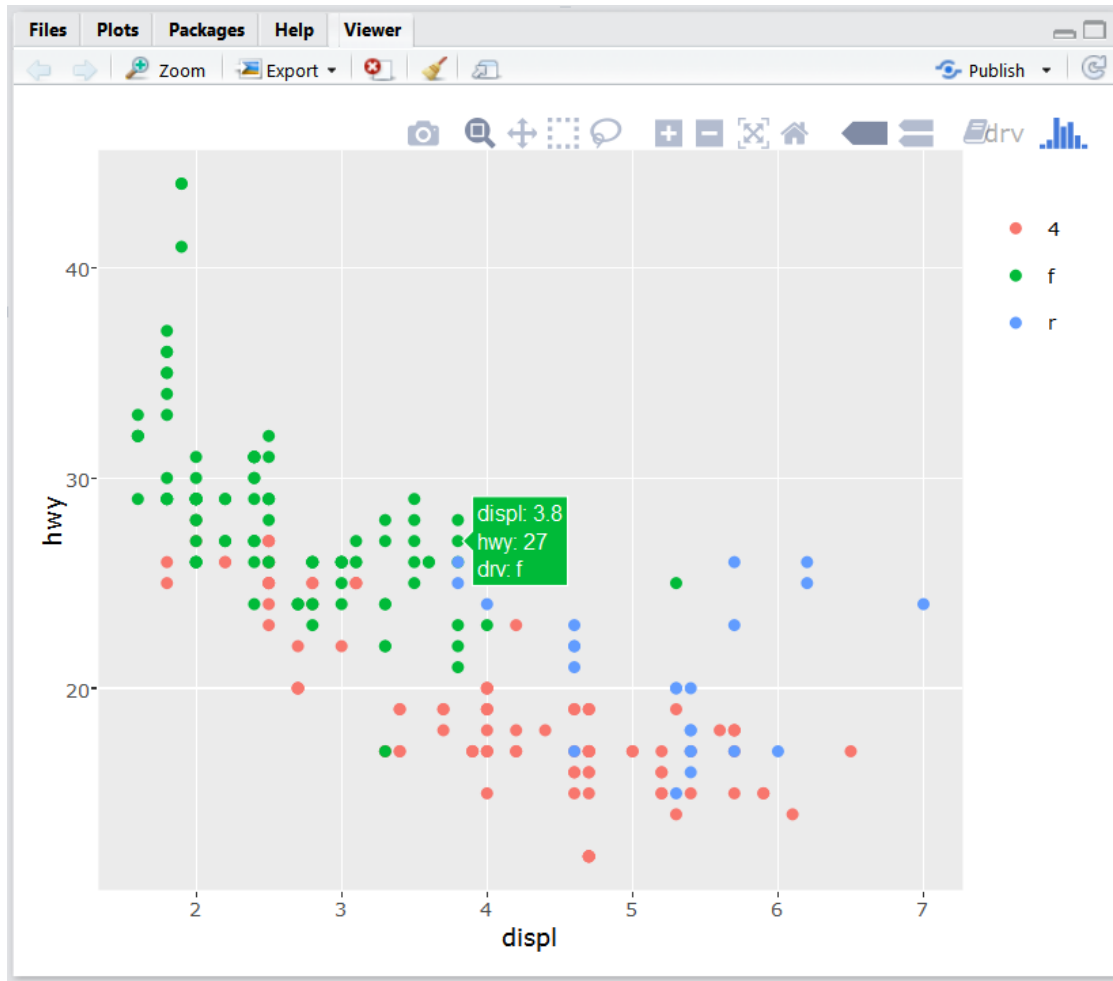
```
library(plotly)
```

## ggplot으로 그래프 만들기

```
library(ggplot2)
p <- ggplot(data = mpg, aes(x = displ, y = hwy, col = drv)) + geom_point()
```

## 인터랙티브 그래프 만들기

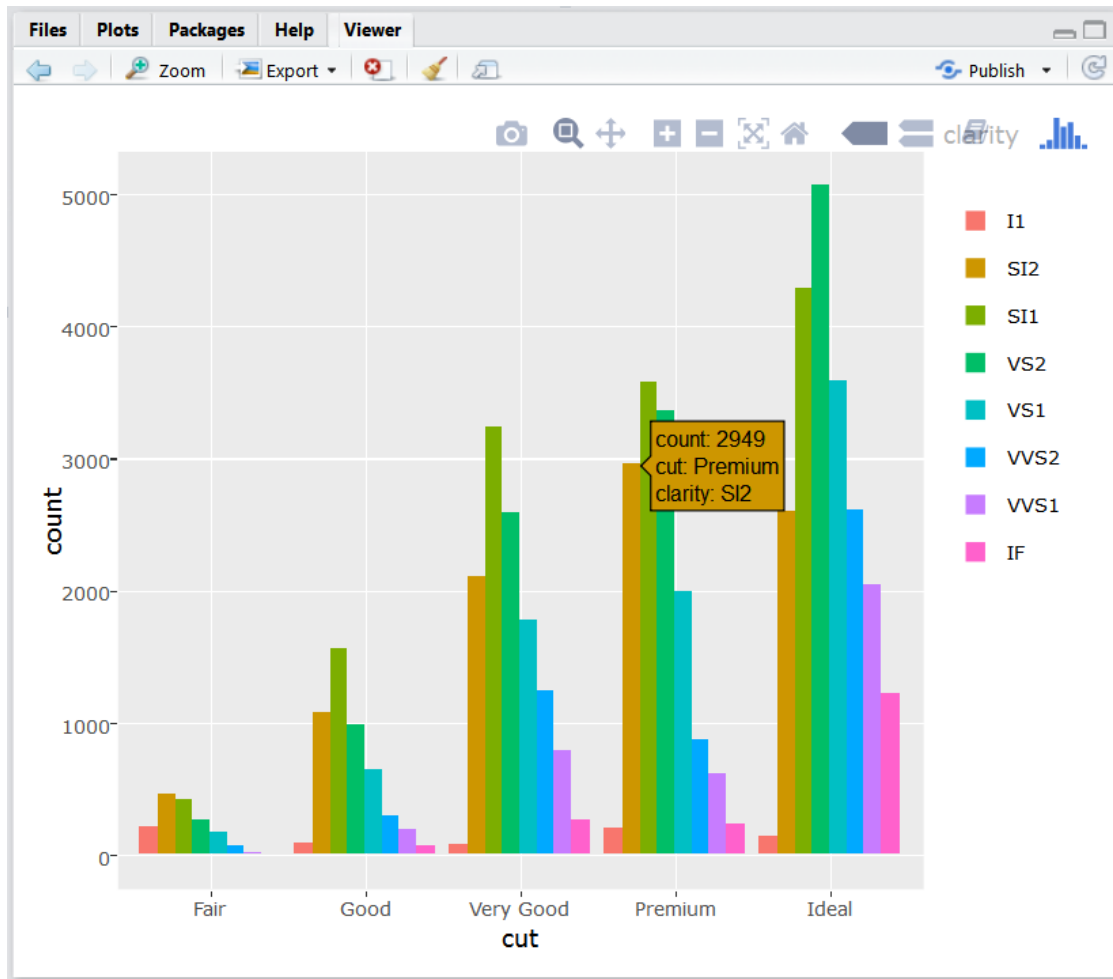
`ggplotly(p)`



## 인터랙티브 막대 그래프 만들기

```
p <- ggplot(data = diamonds, aes(x = cut, fill = clarity)) +  
  geom_bar(position = "dodge")
```

```
ggplotly(p)
```



## 12-2. dygraphs 패키지로 인터랙티브 시계열 그래프 만들기

### 인터랙티브 시계열 그래프 만들기

#### 패키지 준비하기

```
install.packages("dygraphs")
```

```
library(dygraphs)
```

## 데이터 준비하기

```
economics <- ggplot2::economics  
head(economics)
```

```
## # A tibble: 6 x 6
```

```
##       date    pce    pop psavert uempmed unemploy  
##   <date> <dbl> <int>   <dbl>   <dbl>    <int>  
## 1 1967-07-01 507.4 198712    12.5     4.5     2944  
## 2 1967-08-01 510.5 198911    12.5     4.7     2945  
## 3 1967-09-01 516.3 199113    11.7     4.6     2958  
## 4 1967-10-01 512.9 199311    12.5     4.9     3143  
## 5 1967-11-01 518.1 199498    12.5     4.7     3066  
## 6 1967-12-01 525.8 199657    12.1     4.8     3018
```

## 시간 순서 속성을 지니는 xts 데이터 타입으로 변경

```
library(xts)
```

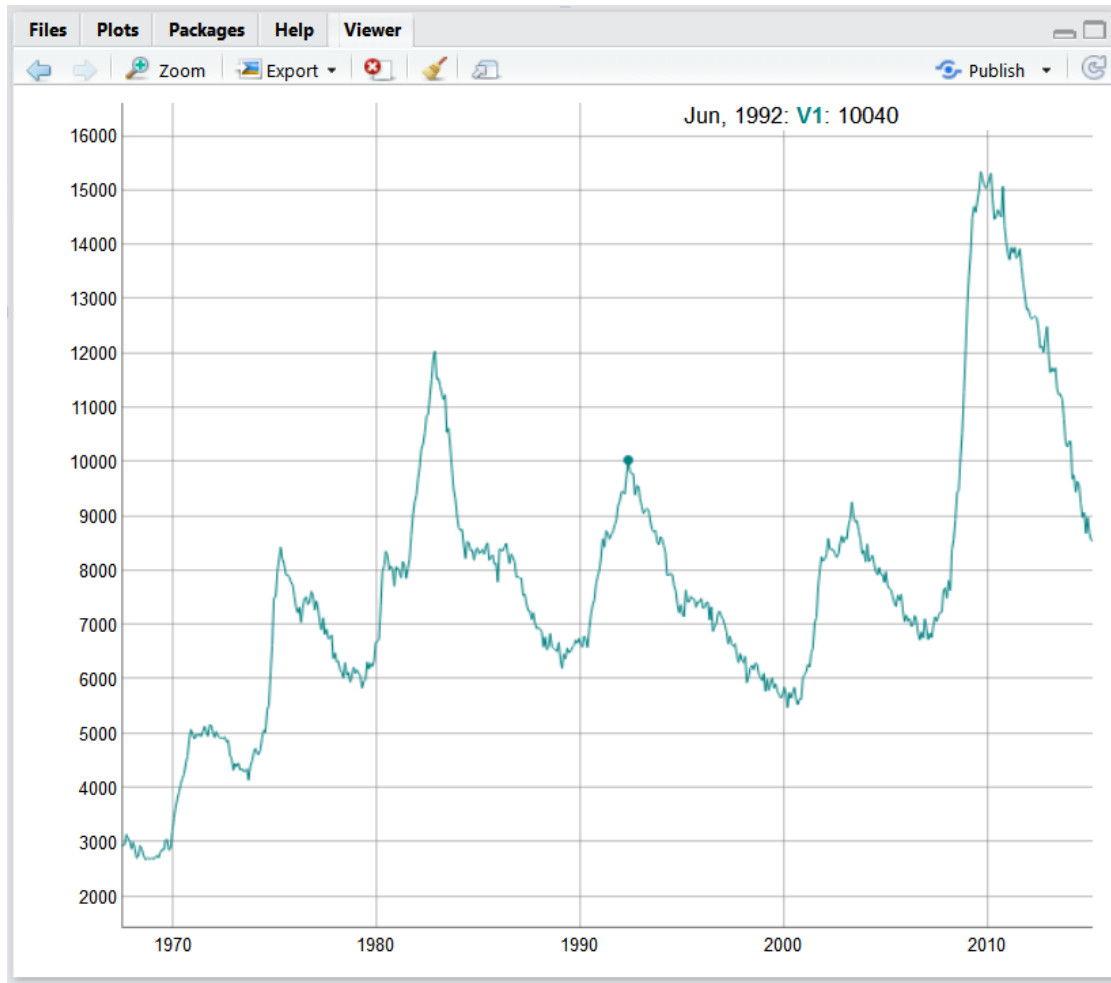
```
eco <- xts(economics$unemploy, order.by = economics$date)  
head(eco)
```

```
##           [,1]  
## 1967-07-01 2944  
## 1967-08-01 2945  
## 1967-09-01 2958  
## 1967-10-01 3143  
## 1967-11-01 3066  
## 1967-12-01 3018
```

## 인터랙티브 시계열 그래프 만들기

# 그래프 생성

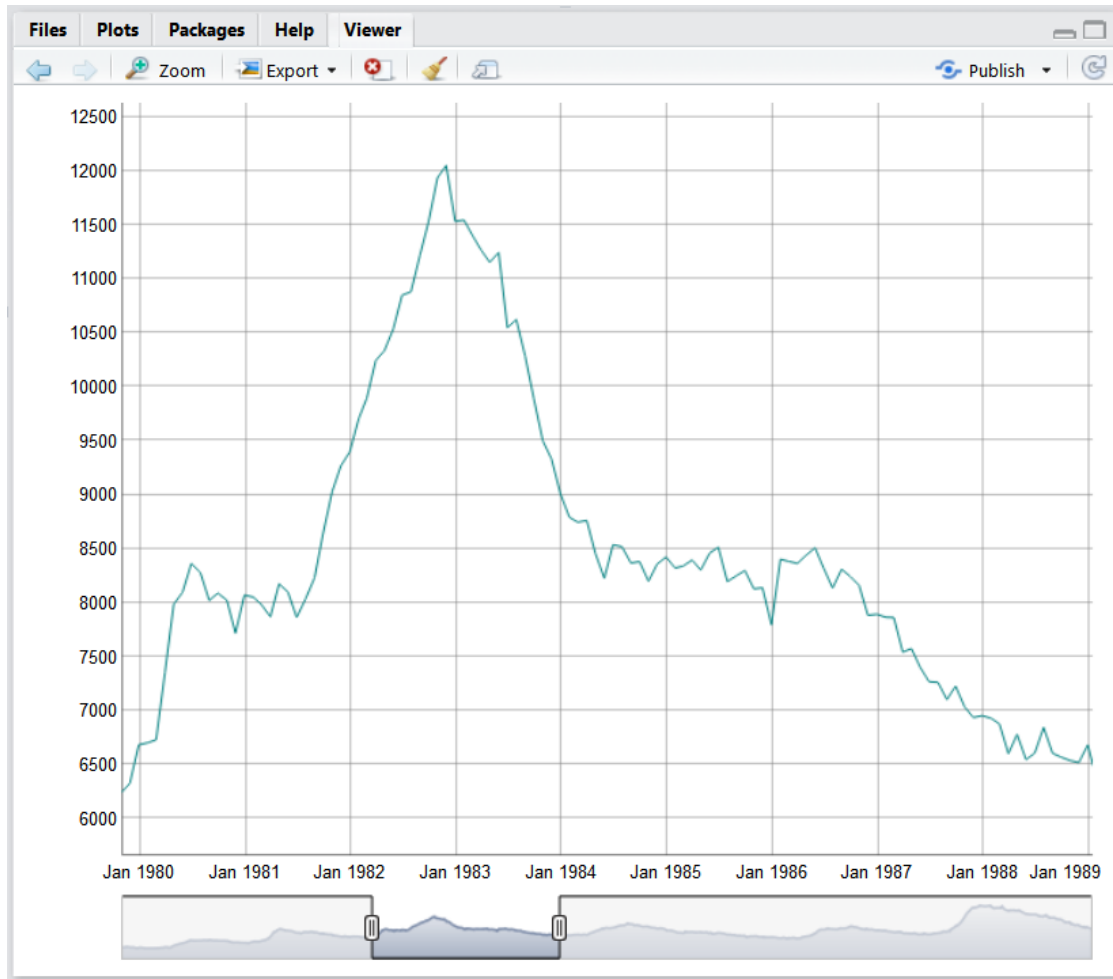
`dygraph(eco)`





## 날짜 범위 선택 기능

```
dygraph(eco) %>% dyRangeSelector()
```



## 여러 값 표현하기

*# 저축률*

```
eco_a <- xts(economics$psavert, order.by = economics$date)
```

*# 실업자 수*

```
eco_b <- xts(economics$unemploy/1000, order.by = economics$date)
```

## 합치기

```
eco2 <- cbind(eco_a, eco_b) # 데이터 결합  
colnames(eco2) <- c("psavert", "unemploy") # 변수명 바꾸기  
head(eco2)
```

```
##           psavert unemploy  
## 1967-07-01     12.5     2.944  
## 1967-08-01     12.5     2.945  
## 1967-09-01     11.7     2.958  
## 1967-10-01     12.5     3.143  
## 1967-11-01     12.5     3.066  
## 1967-12-01     12.1     3.018
```

## 그래프 만들기

```
dygraph(eco2) %>% dyRangeSelector()
```



# R Markdown으로 데이터 분석 보고서 만들기

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

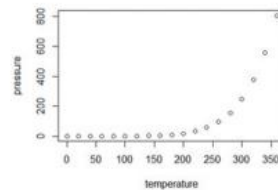
When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
summary(cars)

##      speed          dist
##  Min.   : 4.0    Min.   :  2.00
## 1st Qu.:12.0    1st Qu.: 26.00
##  Median:15.0    Median: 36.00
##   Mean :15.4     Mean : 42.98
## 3rd Qu.:19.0    3rd Qu.: 56.00
##   Max. :25.0     Max.   :120.00
```

## Including Plots

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

# 14-1. 신뢰할 수 있는 데이터 분석 보고서 만들기

## 신뢰할 수 있는 데이터 분석 보고서

- 코드와 결과물이 설명 글과 함께 어우러진 보고서
  - 독자가 분석 과정을 명확히 이해할 수 있음
  - 직접 코드를 실행하면서 동일한 결과가 도출되는지 확인 가능
  - 코드를 자신의 분석 작업에 활용 가능
- 재현성(Reproducibility)을 갖춘 보고서
  - 동일한 분석 과정을 거쳤을 때 동일한 분석 결과가 반복되어 나옴

## R 마크다운(R Markdown)

- 마크다운 문법을 이용해 R 데이터 분석 보고서를 생성하는 기능
- HTML, 워드, PDF 등 다양한 포맷으로 저장 가능, 별도 문서 작성 소프트웨어 없이 R로 보고서 작성 가능

## 마크다운 문법

## R 마크다운이란?

마크다운 문법으로 코드를 작성해 [HTML](<https://ko.wikipedia.org/wiki/HTML>)로 변환할 수 있습니다. R 마크다운을 이용하면 분석 과정과 결과를 자세하게 기술할 수 있기 때문에 **\*\*재현 가능성\*\***을 갖춘 **\*데이터 분석 보고서\***를 만들 수 있습니다.

### R 마크다운 문법 예제

아래와 같이 작성하면 코드와 함께 ``summary(cars)``를 실행한 결과물이 출력됩니다.

```
```{r}
summary(cars)
```
```

별도로 이미지 파일을 붙여 넣는 작업을 하지 않아도 코드 아래에 그래프를 삽입할 수 있습니다.

```
```{r}
library(ggplot2)
qplot(data = mpg, x = drv, fill = drv)
```
```

# HTML 출력 결과



# 양식 적용하기

## 마크다운 문법

문자 앞뒤에 \*특수 문자\*를 넣으면 기울임체가 됩니다.

문자 앞뒤에 \*\*특수 문자\*\*를 넣으면 강조체가 됩니다.

문자 앞뒤에 ~~특수 문자~~를 넣으면 취소선을 만듭니다.

문자 앞뒤에 [특수 문자](<http://www.google.com/search?q=special+character>)를 넣으면 하이퍼링크를 만듭니다.

# 1 단계 제목

## 2 단계 제목

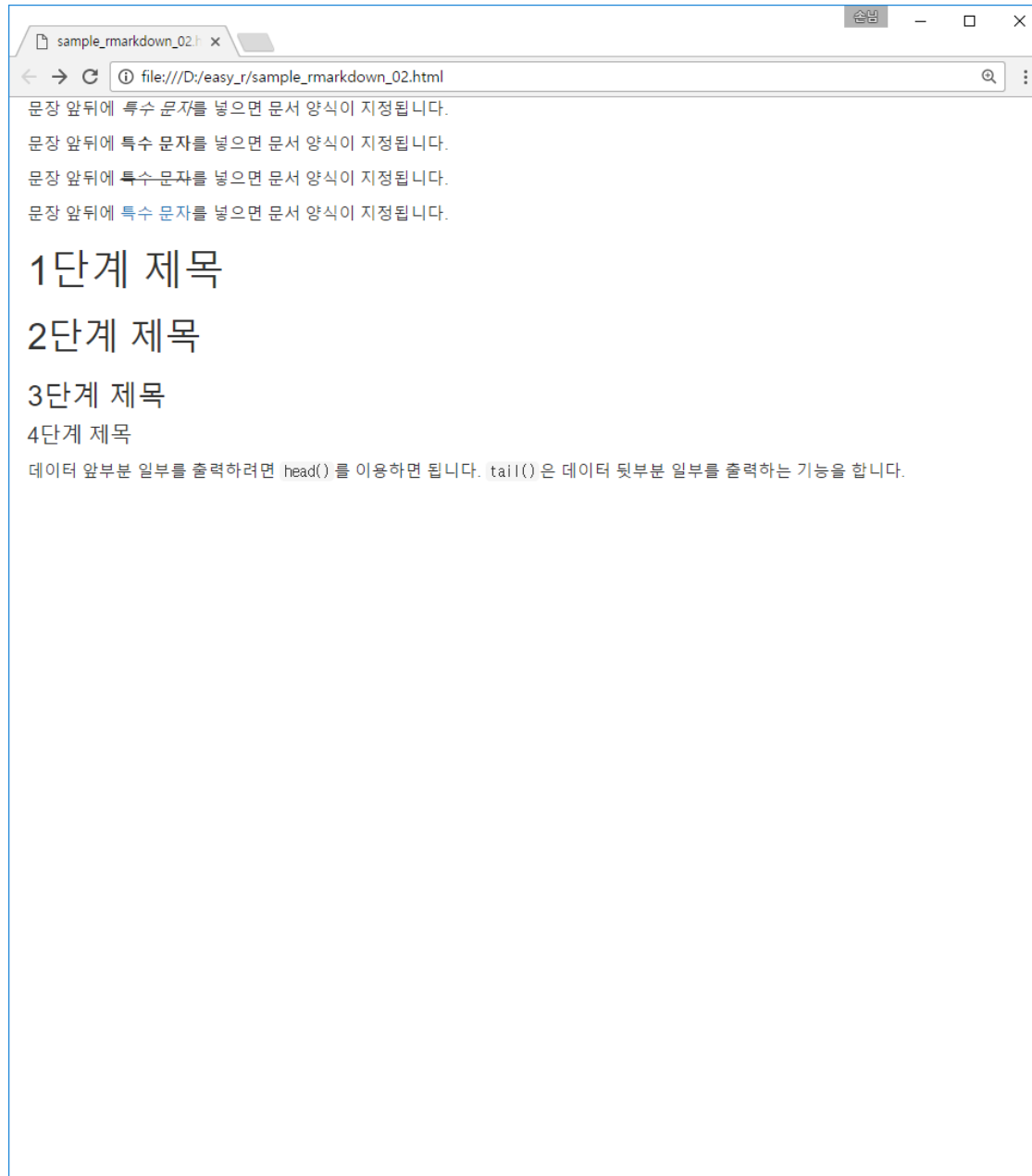
### 3 단계 제목

#### 4 단계 제목

데이터 앞부분 일부를 출력하려면 `head()`를 이용하면 됩니다. `tail()`은 데이터 뒷부분 일부를 출력하는 기능을 합니다.



## HTML 출력 결과



## 코드와 실행 결과

코드 청크(Code chunk) 안에 코드 입력(단축키 [Ctrl + Alt + I])

```
```{r}
summary(cars)
library(ggplot2)
qplot(data = mpg, x = drv, fill = drv)
```
```

HTML 출력 결과

