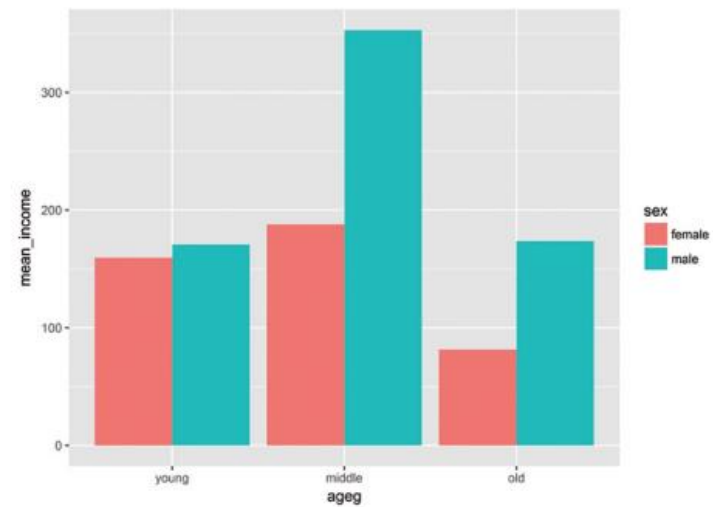
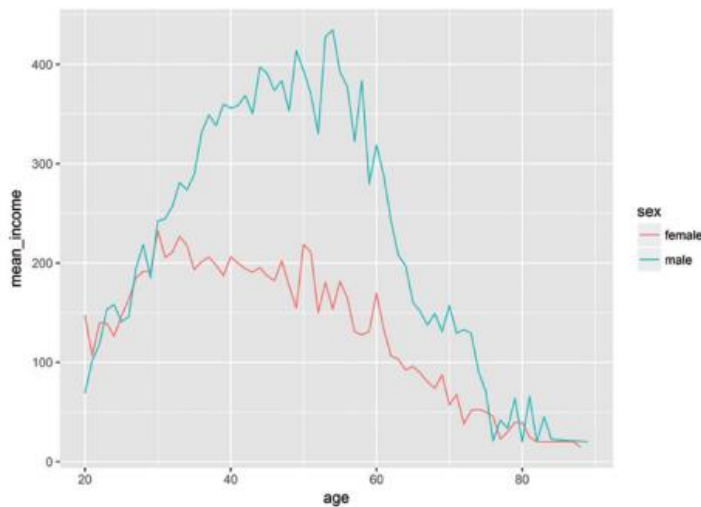


# 09. 데이터 분석 프로젝트

'한국인의 삶을 파악하라!'



## 09-1. '한국복지패널데이터' 분석 준비하기

### 한국복지패널데이터

- 한국보건사회연구원 발간
- 가구의 경제활동을 연구해 정책 지원에 반영할 목적
- 2006~2015년까지 전국에서 7000여 가구를 선정해 매년 추적 조사
- 경제활동, 생활실태, 복지욕구 등 수천 개 변수에 대한 정보로 구성

# 데이터 분석 준비하기

## 패키지 준비하기

```
install.packages("foreign") # foreign 패키지 설치
library(foreign)            # SPSS 파일 로드
library(dplyr)              # 전처리
library(ggplot2)            # 시각화
library(readxl)             # 엑셀 파일 불러오기
```

## 데이터 준비하기

```
# 데이터 불러오기
raw_welfare <- read.spss(file = "Koweps_hpc10_2015_beta1.sav",
                        to.data.frame = T)

# 복사본 만들기
welfare <- raw_welfare
```

## 데이터 검토하기

```
head(welfare)
tail(welfare)
View(welfare)
dim(welfare)
str(welfare)
summary(welfare)
```

- 대규모 데이터는 변수가 많고 변수명이 코드로 되어 있어서 전체 데이터 구조를 한눈에 파악하기 어려움
- 변수명을 쉬운 단어로 바꾼 후 분석에 사용할 변수들 각각 파악해야 함

## 변수명 바꾸기

```
welfare <- rename(welfare,  
  sex = h10_g3,          # 성별  
  birth = h10_g4,        # 태어난 연도  
  marriage = h10_g10,    # 혼인 상태  
  religion = h10_g11,    # 종교  
  income = p1002_8aq1,   # 월급  
  code_job = h10_eco9,    # 직종 코드  
  code_region = h10_reg7) # 지역 코드
```

## 데이터 분석 절차

- 1단계. 변수 검토 및 전처리
- 2단계. 변수 간 관계 분석

### 1단계 변수 검토 및 전처리

전처리

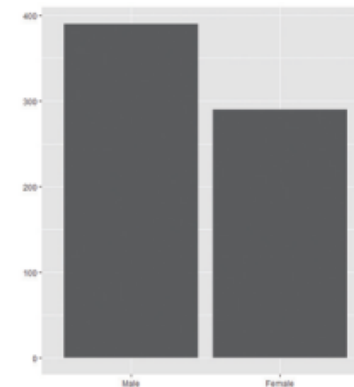
| sex | income |   | sex | income |
|-----|--------|---|-----|--------|
| 2   | 270    |   | 2   | 270    |
| 3   | 210    | → | 1   | 350    |
| 1   | 350    |   | 1   | 430    |
| 2   | 0      |   | 2   | 320    |
| 1   | 430    |   |     |        |
| 2   | 320    |   |     |        |

### 2단계 변수 간 관계 분석

요약표 만들기

| sex | income |
|-----|--------|
| 1   | 390    |
| 2   | 295    |

그래프 만들기



## 09-2. 성별에 따른 월급 차이

### - "성별에 따라 월급이 다를까?"

#### 분석 절차

##### 1. 변수 검토 및 전처리

- 성별
- 월급

##### 2. 변수 간 관계 분석

- 성별 월급 평균표 만들기
- 그래프 만들기

# 성별에 따른 월급 차이 분석하기

## 1. 성별 월급 평균표 만들기

```
sex_income <- welfare %>%  
  filter(!is.na(income)) %>%  
  group_by(sex) %>%  
  summarise(mean_income = mean(income))
```

```
sex_income
```

```
## # A tibble: 2 x 2  
##       sex mean_income  
##   <chr>      <dbl>  
## 1 female    163.2471  
## 2  male     312.2932
```



## 09-3. 나이와 월급의 관계

### - "몇 살 때 월급을 가장 많이 받을까?"

#### 분석 절차

##### 1. 변수 검토 및 전처리

- 나이
- 월급

##### 2. 변수 간 관계 분석

- 나이에 따른 월급 평균표 만들기
- 그래프 만들기

# 나이와 월급의 관계 분석하기

## 1. 나이에 따른 월급 평균표 만들기

```
age_income <- welfare %>%  
  filter(!is.na(income)) %>%  
  group_by(age) %>%  
  summarise(mean_income = mean(income))
```

```
head(age_income)
```

```
## # A tibble: 6 x 2  
##   age mean_income  
##   <dbl>      <dbl>  
## 1    20    121.3000  
## 2    21    105.5185  
## 3    22    130.0923  
## 4    23    141.7157  
## 5    24    134.0877  
## 6    25    144.6559
```

## 09-4. 연령대에 따른 월급 차이

### - "어떤 연령대의 월급이 가장 많을까?"

#### 분석 절차

##### 1. 변수 검토 및 전처리

- 연령대
- 월급

##### 2. 변수 간 관계 분석

- 연령대별 월급 평균표 만들기
- 그래프 만들기

## 09-5. 연령대 및 성별 월급 차이

### - "성별 월급 차이는 연령대별로 다를까?"

#### 분석 절차

##### 1. 변수 검토 및 전처리

- 연령대
- 성별
- 월급

##### 2. 변수 간 관계 분석

- 연령대 및 성별 월급 평균표 만들기
- 그래프 만들기

## 09-6. 직업별 월급 차이

### - "어떤 직업이 월급을 가장 많이 받을까?"

#### 분석 절차

##### 1. 변수 검토 및 전처리

- 직업
- 월급

##### 2. 변수 간 관계 분석

- 직업별 월급 평균표 만들기
- 그래프 만들기

## 09-7. 성별 직업 빈도

### - "성별로 어떤 직업이 가장 많을까?"

#### 분석 절차

##### 1. 변수 검토 및 전처리

- 성별
- 직업

##### 2. 변수 간 관계 분석

- 성별 직업 빈도표 만들기
- 그래프 만들기

## 09-8. 종교 유무에 따른 이혼율

### - "종교가 있는 사람들이 이혼을 덜 할까?"

#### 분석 절차

##### 1. 변수 검토 및 전처리

- 종교
- 혼인 상태

##### 2. 변수 간 관계 분석

- 종교 유무에 따른 이혼율 표 만들기
- 그래프 만들기

## 09-9. 지역별 연령대 비율

### - "노년층이 많은 지역은 어디일까?"

#### 분석 절차

##### 1. 변수 검토 및 전처리

- 지역
- 연령대

##### 2. 변수 간 관계 분석

- 지역별 연령대 비율표 만들기
- 그래프 만들기