

3. { Data 전처리 실습 }

(R Study group 자료)

+) 지난 내용 요약

group_by : 특정 기준으로 데이터를 구분 및 인식

summarise : 원하는 function (or verb)로 요약

+ **n ()** # 빈도 계산

arrange : 변수를 기준으로 정렬

+ **desc()** (= -변수) #내림차순 정렬

ifelse : 변수의 조건이 참, 거짓일때의 값을 출력

ifelse(변수 조건, 참일 때 값, 거짓일 때 값)

(R Study group 자료)

0) lol_data를 가지고 실습해보기

준비하기

```
library(tidyverse)
```

```
lol_df <- read.csv("lol_data.csv",  
                  stringsAsFactor = F,  
                  encoding = "euc-kr") # encoding = " " 인코딩설정
```

```
head(lol_df)
```

```
summary(lol_df)
```

```
str(lol_df) # 자료형태 및 구조 확인하기
```

(R Study group 자료)

0) lol_data를 가지고 실습해보기

Q. 랭크, 자유랭크 게임 중 챔피언별 평균 평점이 높은 챔피언을 나열 하시오.

```
lol_df %>%  
  filter(game_type %in% c("자유랭크", "랭크")) %>%  
  group_by(champ) %>%  
  summarise(value = mean(mean_value)) %>%  
  arrange(desc(value))
```

(R Study group 자료)

0) lol_data를 가지고 실습해보기

Q. 챔피언별 승점은 ? (승 : +1, 패 : -1)

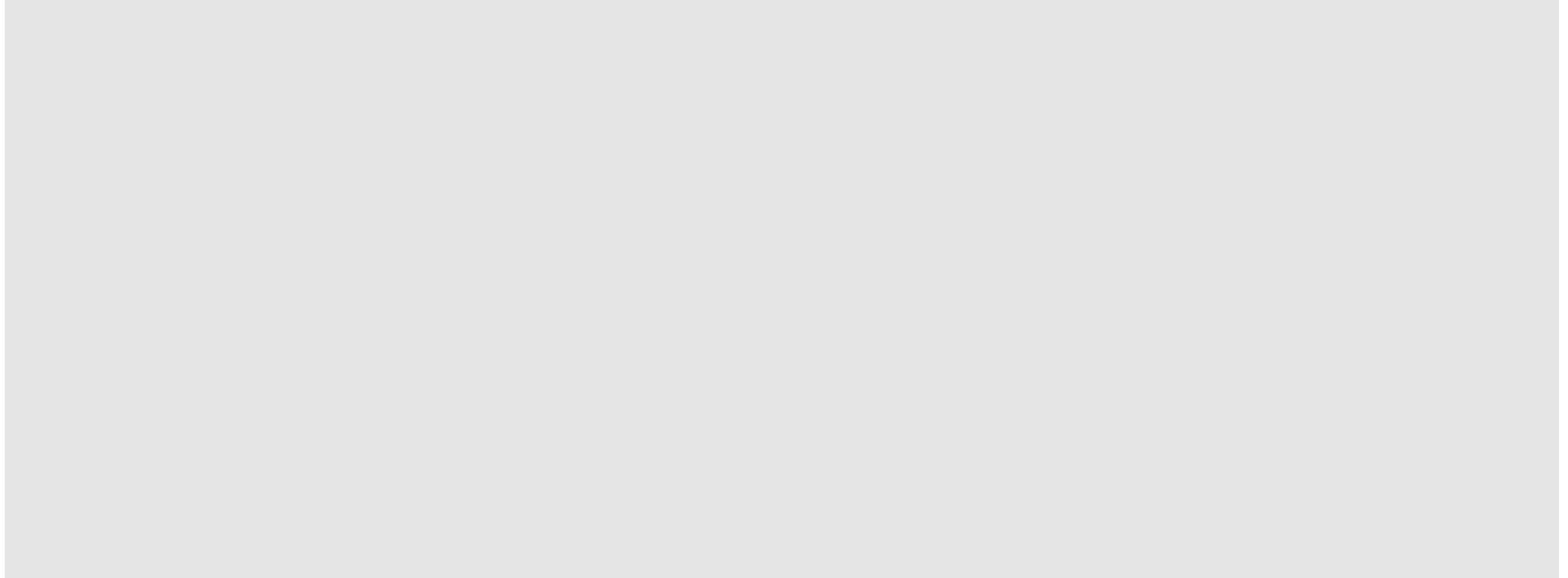
```
lol_df %>%  
  group_by(champ) %>%  
  mutate(win_loss_nm = ifelse(win_loss == "승", 1, -1)) %>%  
  summarise(sum = sum(win_loss_nm)) %>%  
  arrange(sum)
```

기본구조 mutate(이름 = 계산식) # 새로이 변수를 생성해줄
 ifelse(변수 조건, 참일 경우 값, 거짓일 경우 값)

(R Study group 자료)

0) lol_data_2를 가지고 실습해보기

+. 궁금증을 가지고 데이터를 바라보기 !!



(R Study group 자료)

4. { Data 시각화 }

(R Study group 자료)

1) Data Visualization (데이터 시각화)

: 더 직관적으로 보기 위함.

변수의 개수 1 : bar plot, histogram, box plot, density ... etc.

주로 빈도 분포를 나타낼 때

변수의 개수 2 : dat(point) plot, line plot, bar plot, box plot, pie chart ... etc.

주로 빈도 분포를 나타낼 때

변수의 개수 3 이상 : 특정 조건을 부여하여 시각화
ex) 색, 모양, 색상의 진하기, 화면 분할 ... etc.

(R Study group 자료)