

데이터 시각화(ggplot2)관련 총정리

src: 7_R/11_사례연구 참고

0) 준비 단계

```
Rm(list=ls(all.names=TRUE)) ; 기존 변수 삭제
library(패키지명) #필요 패키지 로드: dplyr, doBy, ggplot2, girdExtra, readxl, foreign 등
getwd() ; 현재 워킹 디렉토리
데이터 <- rename(데이터, NEW열이름1=열이름1, NEW열이름2=열이름2)
colnames(데이터) ; 데이터 열이름 확인      dim(데이터) ; 데이터 차원 확인
```

1) 그래프로 시각화

#fill=채우는 색, col=점, 선(테두리) 색

```
ggplot(data, aes(x= , y= , fill= ), color= ) + #열별로 색 달리하고 싶을 때 aes 안에 기재
  geom_col(position="dodge") + #position="dodge"; 분리 막대그래프 작성 가능 (누적X)
  geom_point(position="jitter", alpha=0.5) + #position="jitter"; 점 분산, alpha: 투명도(0~1)
  geom_boxplot(notch=T, width=0.5, fill="") + #notch=T; 중위수 그래프로 나타냄
  labs(title="", subtitle="", x="", y="") + #그래프 제목, 축 이름 등
  coord_cartesian(ylim=c(0, 10)) + #y축이나 x축 범위 조정 가능
  coord_flip() + #x축과 y축 바꾸기(가로 막대그래프로 변환시)
  scale_fill_manual(values=heat.colors(n)) + #색 다시 조정 가능, 팔레트 이용 가능
  scale_x_discrete(limits=c(' ', ' ')) + #x축 라벨 순서(y축이나 범례(scale_fill_discrete)도 가능)
  theme(legend.position="top", #범례위치 조정: 밖 상, 하, 좌, 우/ 그래프 안 c(0,0~1,1) / "none"
        axis.text.x=element_blank(), #x축 눈금 없애기
        axis.text.x=element_text(angle=70, vjust=0.5)) + #x축 라벨 각도 조절(겹치지 않게)
  geom_text(aes(label=y축), vjust=-0.5) + #label=y축 입력시 y축 라벨 값이 그래프에 표시
  aes(stringr::str_wrap(x축, num), y축) #x축 라벨 길이 조정: 일정 길이(num)이후 밑에 써지게
```

데이터 %>%

```
filter(!is.na(열1)&!is.na(열2)) %>% #결측치 제외, 조건 등 입력
group_by(열3) %>% #group_by, select, summarise, arrange, mutate 등 사용
summarise(변수명=함수, ... ) %>% #reorder(a, b) b값 내림차순으로 a정렬
ggplot(aes(x=, y=reorder( , ))) + #dplyr과 같이 이용시 상기 데이터를 사용하므로 data 기재X
geom_bar(stat="identity") + #y값이 있을 경우 stat="identity"기재 필요, geom_col()과 동일 결과
geom_text(aes(label=paste(y축, '%'), vjust=-0.5, size=3) #paste() 사용가능: 라벨값%, 글자 크기: 3
```