1. 라이브러리 설치

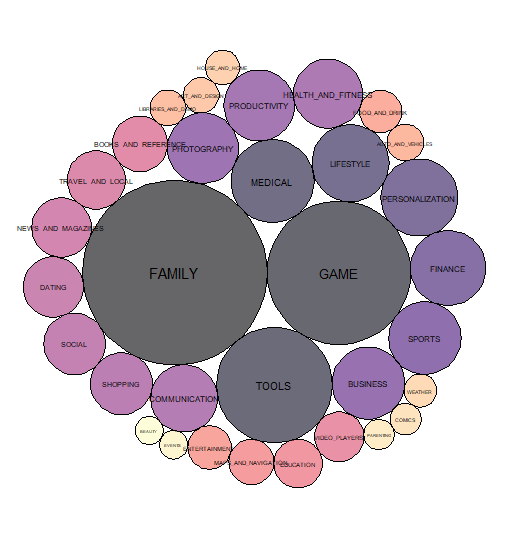
|  |
| --- |
| # 엑셀 파일 불러오기  library(readxl)  # 데이터 전처리  library(dplyr)  # 그래프  library(tidyverse)  library(packcircles)  library(ggplot2)  library(viridis) |

1. 데이터 전처리

|  |
| --- |
| # 데이터셋 가져오기  google <- read\_excel("googleplaystore1.xlsx")  # 데이터 전처리  google$Size <- ifelse(google$Size == 'Varies with device', NA , google$Size)  google$Rating <- ifelse(google$Rating == 'NaN', NA , google$Rating)  google$Reviews <- as.numeric(google$Reviews)  google <- na.omit(google)  # Size 의 ‘M’ 지우기  google$Size <- gsub('M',"" ,google$Size)  # Installs 의 ‘+’ 지우기  google$Installs <- ifelse( google$Installs == '1+', 1,  ifelse(google$Installs == '10+', 10 ,  ifelse(google$Installs == '100+',100,  ifelse(google$Installs == '1,000+',1000,  ifelse(google$Installs == '10,000+',10000,  ifelse(google$Installs=='100,000+',100000,  ifelse(google$Installs=='5+',5,  ifelse(google$Installs=='50+',50,  ifelse(google$Installs=='500+',500,  ifelse(google$Installs =='5,000+',5000,  ifelse(google$Installs=='50,000+',50000,  ifelse(google$Installs=='1,000,000+',1000000,  ifelse(google$Installs=='500,000+',500000,  ifelse(google$Installs=='5,000,000+',5000000,  ifelse(google$Installs=='10,000,000+',10000000,  ifelse(google$Installs=='50,000,000+',50000000, ifelse(google$Installs=='100,000,000+',100000000, ifelse(google$Installs=='500,000,000+',500000000, ifelse(google$Installs=='1,000,000,000+',1000000000,NA))))))))))))))))))) |

1. 장르별 빈도수 구하기

|  |
| --- |
| # 장르별 빈도수 데이터 생성  category <- google %>%  filter(!is.na(Category)) %>%  group\_by(Category) %>%  summarise(n = n()) %>%  arrange(desc(n))  # 그래프  # 1. 그래프 데이터 만들기  data <- data.frame(group=category$Category, value=category$n)  # 2. 레이아웃을 생성하고, sizetype은 값에 비례할 항목에 대한 선호도에 따라 면적 또는 반지름이 된다.  packing <- circleProgressiveLayout(data$value, sizetype='area')  data <- cbind(data, packing)  dat.gg <- circleLayoutVertices(packing, npoints=50)  # 3. 색상 지정하고 그래프 보여주기  ggplot() +  geom\_polygon(data = dat.gg, aes(x, y, group = id, fill=as.factor(id)), colour = "black", alpha = 0.6) +  scale\_fill\_manual(values = magma(nrow(data))) +  geom\_text(data = data, aes(x, y, size=value, label = group)) +  scale\_size\_continuous(range = c(1,4)) +  theme\_void() +  theme(legend.position="none") +  coord\_equal() |



1. 장르별 평균 다운로드 수와 리뷰 수

|  |
| --- |
| # 장르별 평균 다운로드 수 데이터 생성  Category\_install <- google2 %>%  filter(!is.na(Installs) & !is.na(Category) & !is.na(Type)) %>%  group\_by(Category, Type) %>%  summarise(mean\_install = mean(Installs)) %>%  arrange(desc(mean\_install))  # 평균 다운로드 수 그래프 생성  result <- ggplot(data = Category\_install, aes( x = reorder(Category\_install$Category,-Category\_install$mean\_install) , y=Category\_install$mean\_install)) + geom\_col() + coord\_flip()  result + labs(x="장르",y="다운로드",title = "장르별 다운로드 수")  # 장르별 평균 리뷰 수 데이터 생성  Category\_review <- google2 %>%  filter(!is.na(Category) & !is.na(Type) & !is.na(Reviews)) %>%  group\_by(Category , Type) %>%  summarise(mean\_review = mean(Reviews)) %>%  arrange(desc(mean\_review))  # 평균 리뷰 수 그래프 생성  result1 <- ggplot(data = Category\_review, aes( x = reorder(Category\_review$Category,-Category\_review$mean\_review) , y=Category\_review$mean\_review)) + geom\_col() + coord\_flip()  result1 + labs(x="장르",y="리뷰 수",title = "장르별 리뷰뷰 수")  텍스트이(가) 표시된 사진  자동 생성된 설명텍스트, 스크린샷이(가) 표시된 사진  자동 생성된 설명 |

1. 장르별 유료 어플과 무료 어플 평균 다운로드 수

|  |
| --- |
| # 데이터는 Category\_install 사용  mycolor <- ifelse(Category\_install$Type == 'Paid', "type1", "type2")  # 그래프 생성  ggplot(data = Category\_install, aes(x=Category\_install$Type, y=Category\_install$mean\_install)) +  geom\_segment( aes(x=Category\_install$Category, xend=Category\_install$Category, y=1, yend=Category\_install$mean\_install, color=mycolor), size=1, alpha=0.9) +  theme\_light() +  theme(  axis.text.x = element\_text(angle = 90),  legend.position = "none",  panel.border = element\_blank(),  ) +  xlab("") +  ylab("Value of Y") |

1. 장르별 평균 다운로드 수와 사이즈 수 비교

|  |
| --- |
| # 장르별 평균 다운로드 수와 평균 사이즈 데이터 설정  Category\_install\_size <- google %>%  filter(!is.na(Size) & !is.na(Category) & !is.na(Installs)) %>%  group\_by(Category) %>%  summarise(mean\_install = mean(Installs) , mean\_size = mean(Size)) %>%  arrange(desc(mean\_size))  # 그래프  result3 <- ggplot(data=Category\_install\_size, aes( x = reorder(Category\_install\_size$Category,Category\_install\_size$mean\_install) , y= Category\_install\_size$mean\_size)) + geom\_col( ) + coord\_flip()  result3 + labs(x="Genre",y="Size And Install",title = "Size by genre")  스크린샷, 텍스트이(가) 표시된 사진  자동 생성된 설명 |