```
library(data.table)
library(magrittr)
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

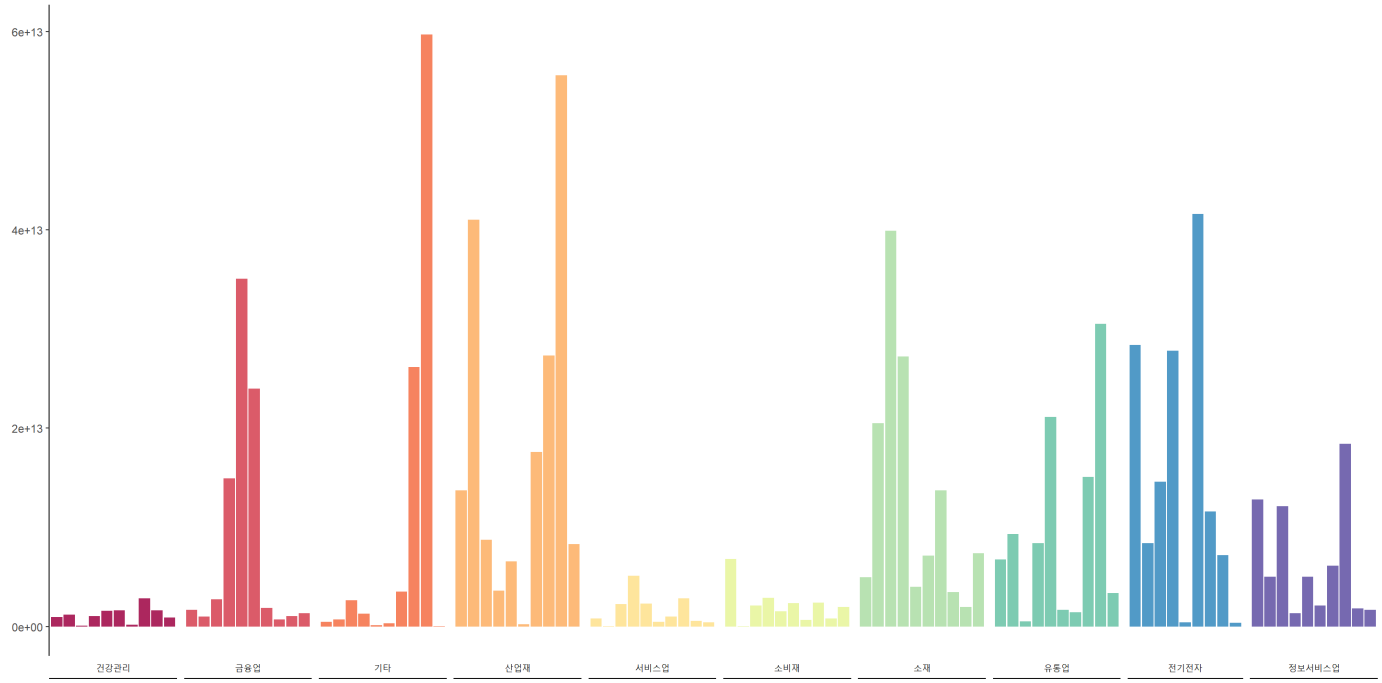
```
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.6     v dplyr   1.0.8
## v tidyr   1.2.0     v stringr 1.4.0
## v readr   2.1.2     v forcats 0.5.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::between()   masks data.table::between()
## x tidyr::extract()   masks magrittr::extract()
## x dplyr::filter()    masks stats::filter()
## x dplyr::first()     masks data.table::first()
## x dplyr::lag()       masks stats::lag()
## x dplyr::last()      masks data.table::last()
## x purrr::set_names() masks magrittr::set_names()
## x purrr::transpose() masks data.table::transpose()
```

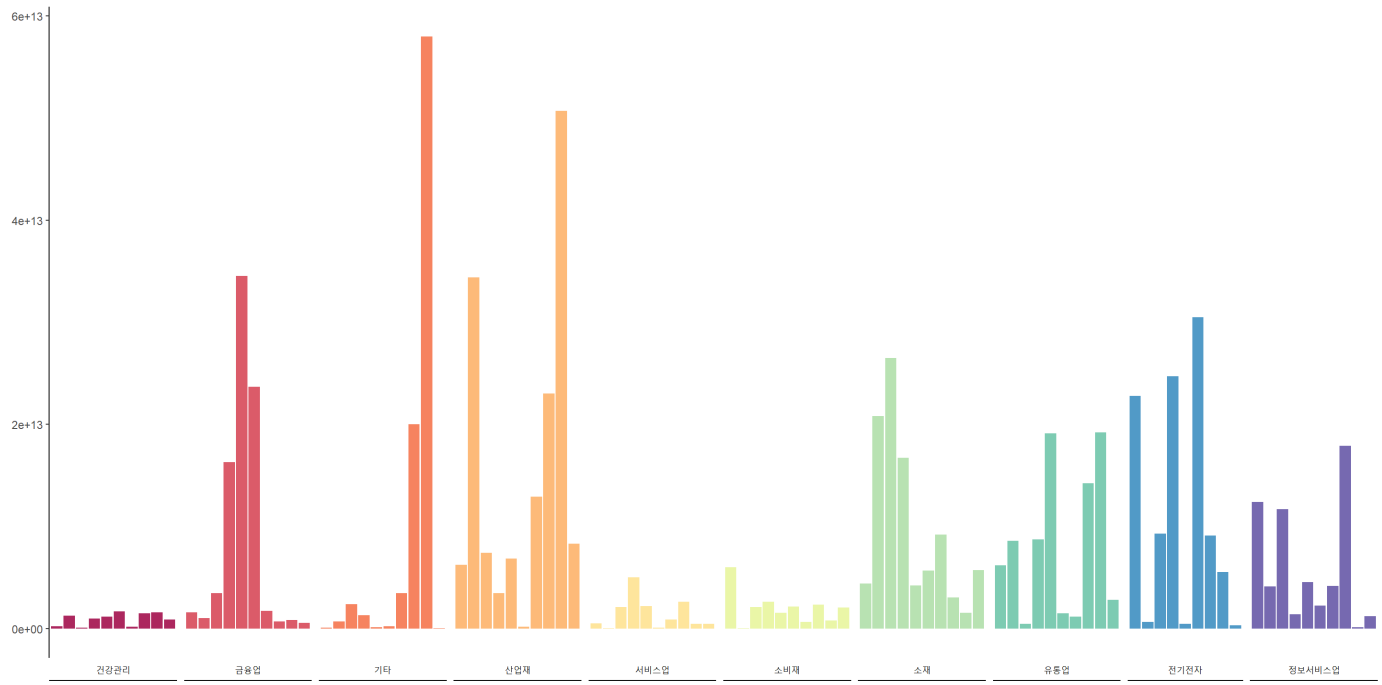
```
data<-fread('기업정보_전처리_5차_완료.csv')

data %<>% mutate_at(vars(`2021당기순이익`,`2021매출액`,`2021영업이익`,
                         `2020당기순이익`,`2020매출액`,`2020영업이익`,
                         `2019당기순이익`,`2019매출액`,`2019영업이익`,
                         시가총액,상장일,총직원수,평균근속연수),
                    as.numeric)
#  mutate_if(is.numeric,function(x){return/})
```

https://stackoverflow.com/questions/38101512/the-same-width-of-the-bars-in-geom-barposition-dodge (https://stackoverflow.com/questions/38101512/the-same-width-of-the-bars-in-geom-barposition-dodge)

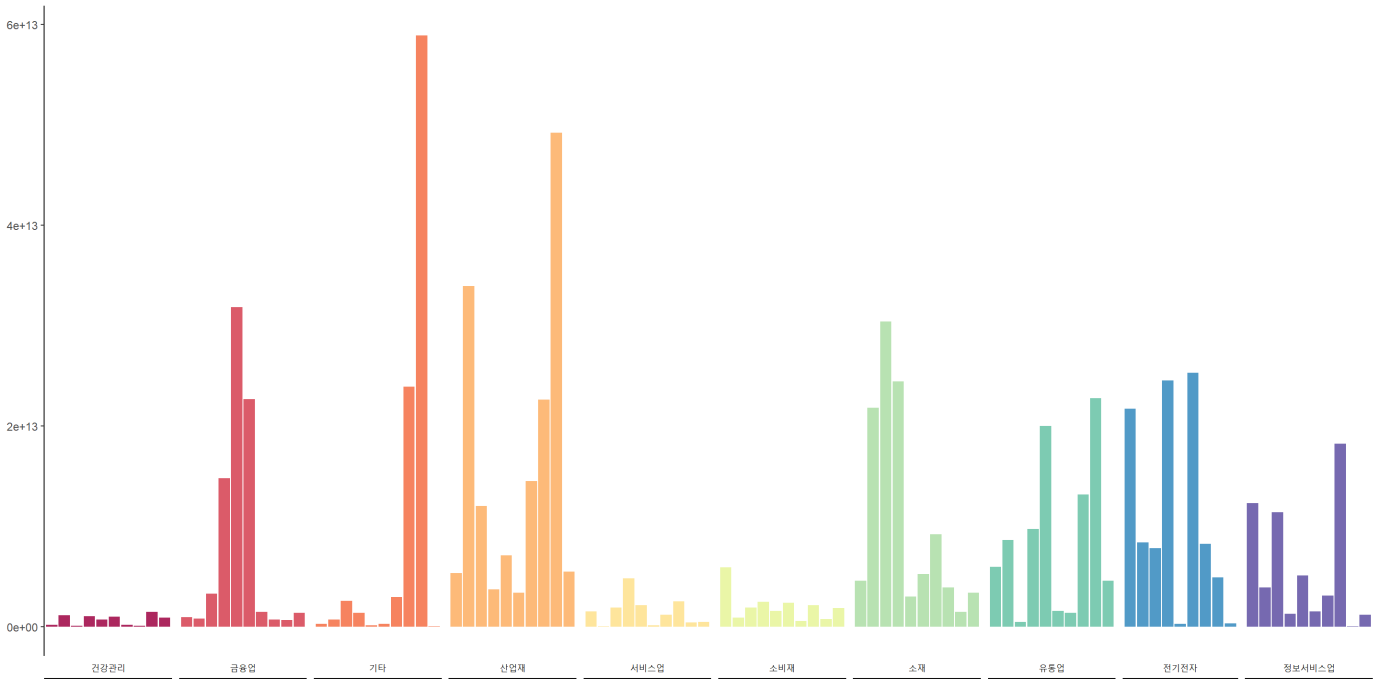
```
data %>% filter(회사명!='삼성전자') %>%
  ggplot(aes(x = 회사명,  y = `2021매출액`, fill = 대분류))  +
  geom_col(position = "dodge",alpha=0.85) +
  facet_grid(~대분류, scales = "free_x", space = "free_x", switch = "x") +
  scale_fill_brewer(palette="Spectral")+
  theme_classic()+
  theme(legend.position='none',
        axis.text.x = element_blank(),
        axis.ticks.x = element_blank(),
        axis.title.x=element_blank(),
        axis.title.y=element_blank(),
        strip.background = element_blank())
```



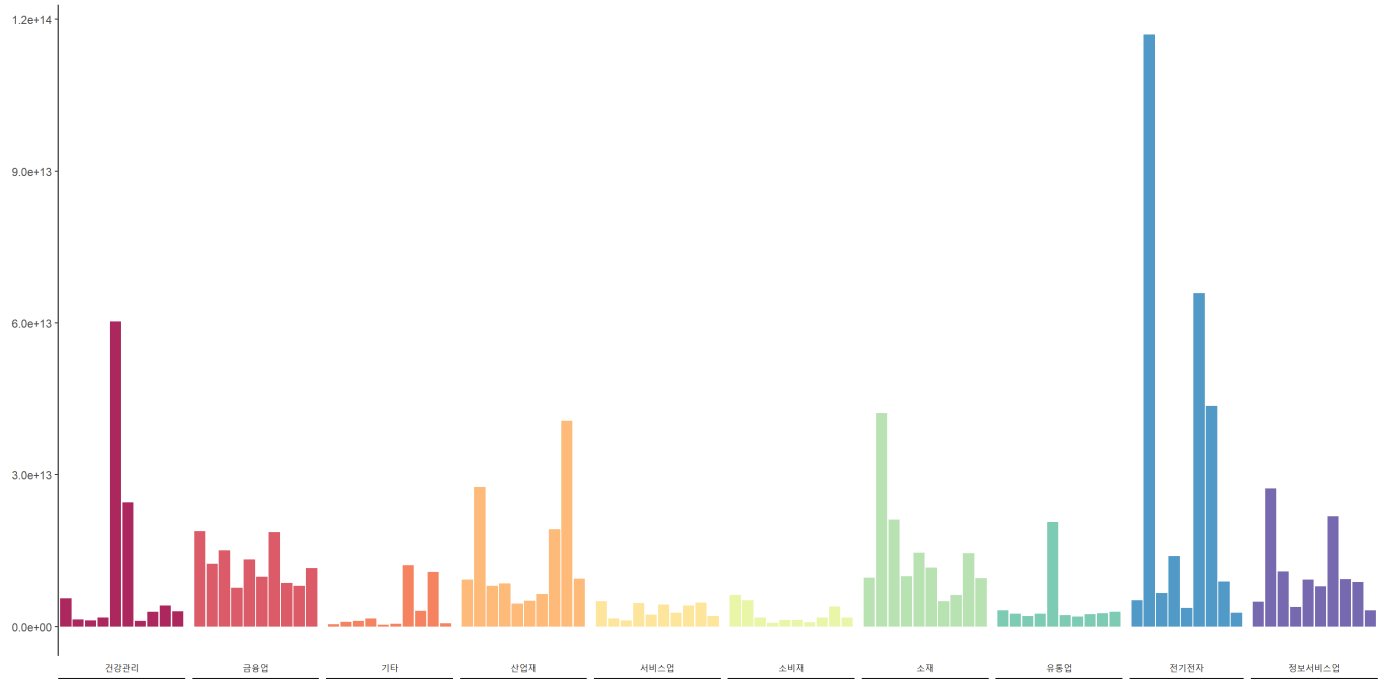

```
data %>% filter(회사명!='삼성전자') %>%
  ggplot(aes(x = 회사명,  y = `2020매출액`, fill = 대분류))  +
  geom_col(position = "dodge",alpha=0.85) +
  facet_grid(~대분류, scales = "free_x", space = "free_x", switch = "x") +
  scale_fill_brewer(palette="Spectral")+
  theme_classic()+
  theme(legend.position='none',
        axis.text.x = element_blank(),
        axis.ticks.x = element_blank(),
        axis.title.x=element_blank(),
        axis.title.y=element_blank(),
        strip.background = element_blank())
```



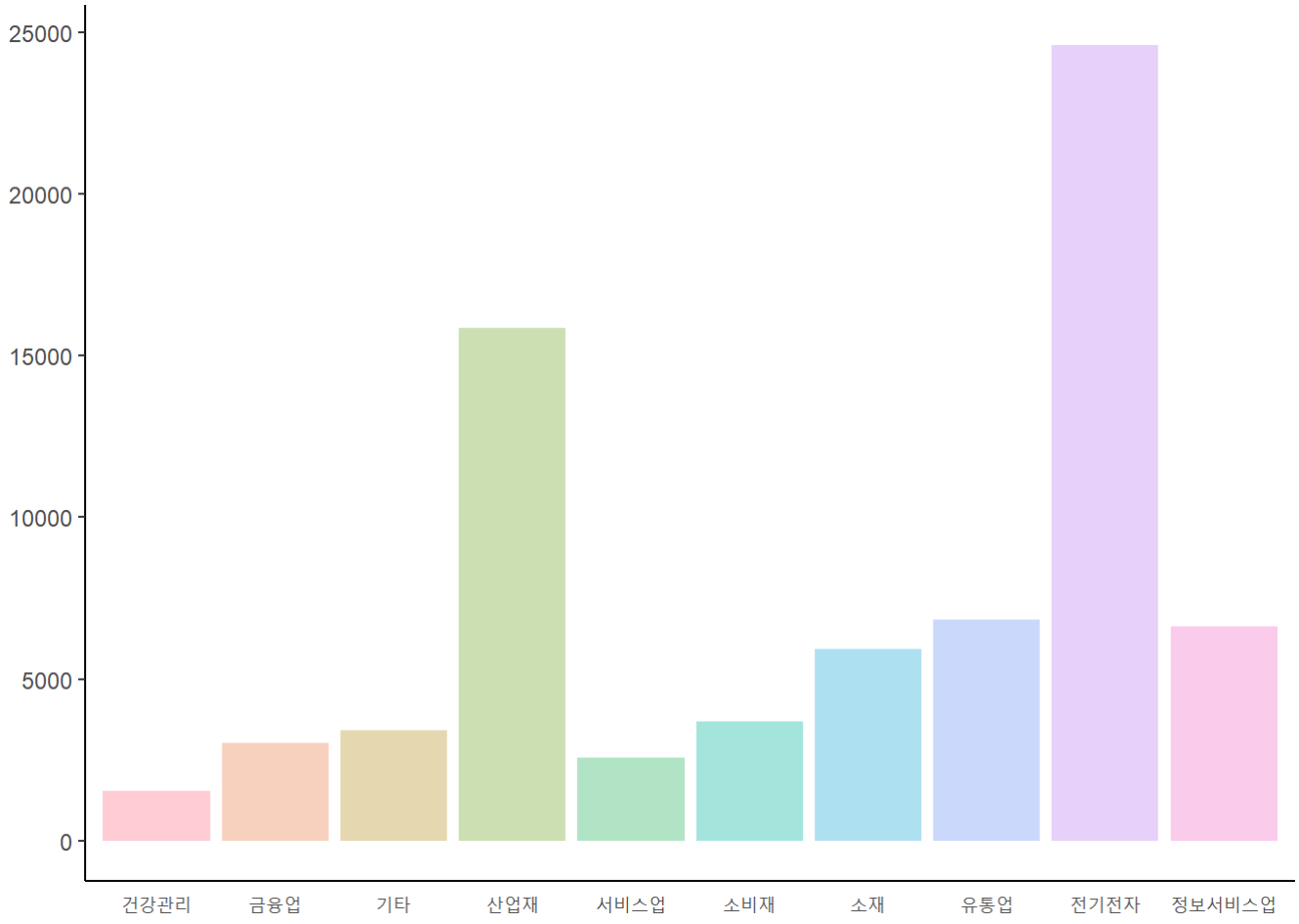
```
data %>% filter(회사명!='삼성전자') %>%
  ggplot(aes(x = 회사명,  y = `2019매출액`, fill = 대분류))  +
  geom_col(position = "dodge",alpha=0.85) +
  facet_grid(~대분류, scales = "free_x", space = "free_x", switch = "x") +
  scale_fill_brewer(palette="Spectral")+
  theme_classic()+
  theme(legend.position='none',
        axis.text.x = element_blank(),
        axis.ticks.x = element_blank(),
        axis.title.x=element_blank(),
        axis.title.y=element_blank(),
        strip.background = element_blank())
```



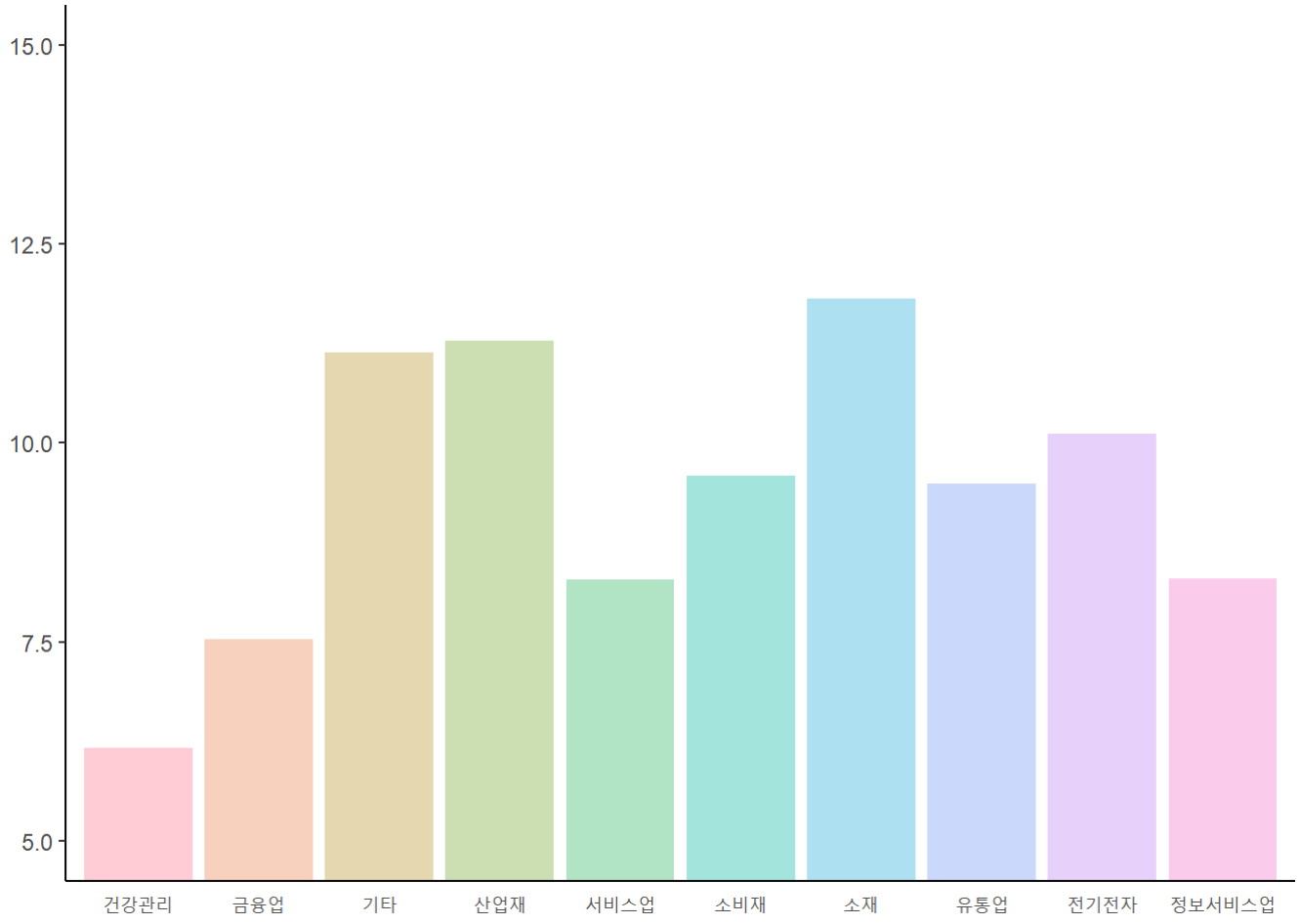

```
data %>% filter(회사명!='삼성전자') %>%
  ggplot(aes(x = 회사명,  y = 시가총액, fill = 대분류))  +
  geom_col(position = "dodge",alpha=0.85) +
  facet_grid(~대분류, scales = "free_x", space = "free_x", switch = "x") +
  scale_fill_brewer(palette="Spectral")+
  theme_classic()+
  theme(legend.position='none',
        axis.text.x = element_blank(),
        axis.ticks.x = element_blank(),
        axis.title.x=element_blank(),
        axis.title.y=element_blank(),
        strip.background = element_blank())
```



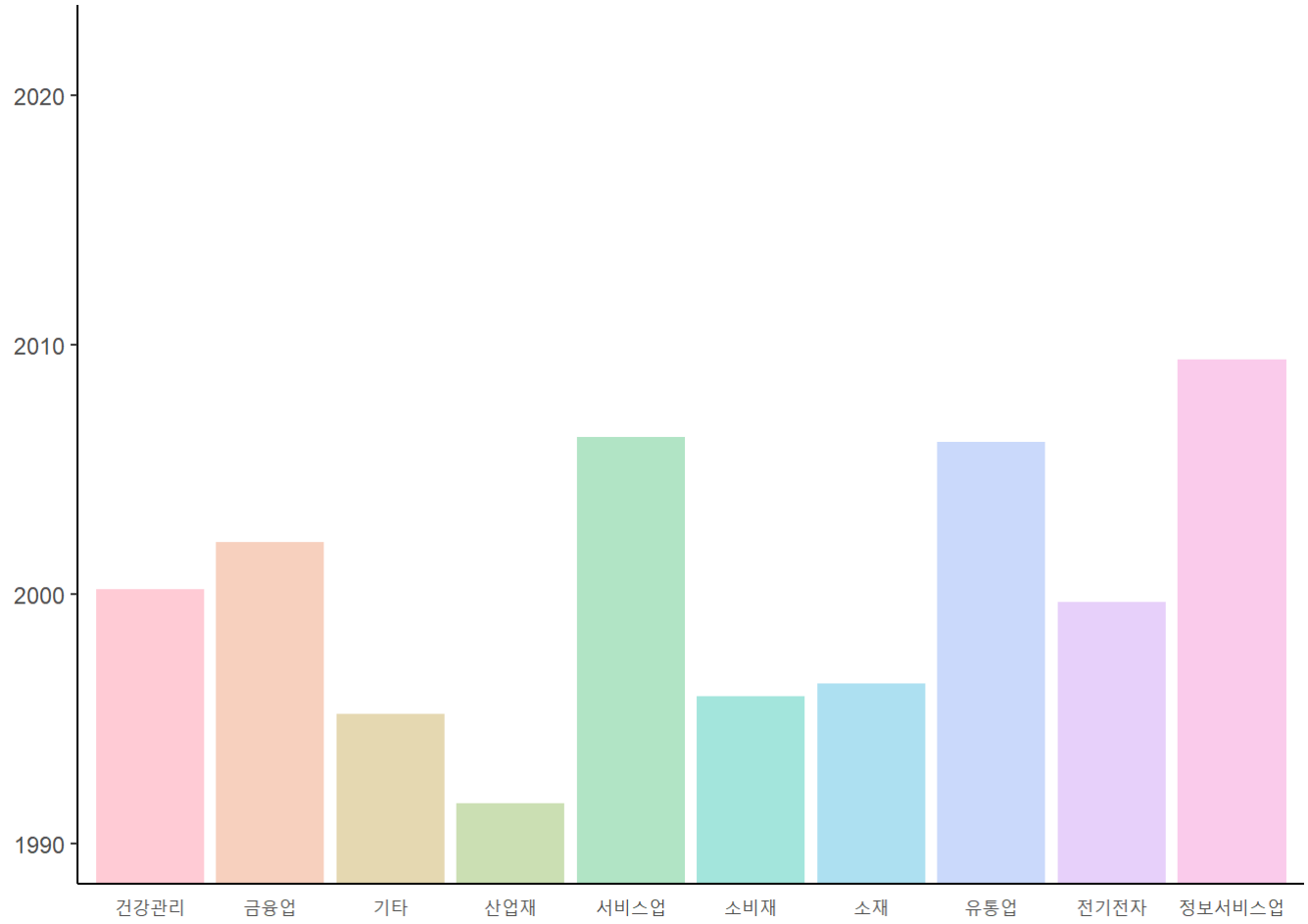

```
col <- hcl.colors(10, palette ="Pastel1")

data %>% group_by(대분류) %>% summarise(avg=mean(총직원수)) %>%
  ggplot(aes(x = 대분류,  y =avg, fill = 대분류))  +
  geom_col(position = "dodge",alpha=0.9) +
  scale_fill_manual(values=col)+
  theme_classic()+
  theme(legend.position='none',
        axis.ticks.x = element_blank(),
        axis.title.x=element_blank(),
        axis.title.y=element_blank(),
        strip.background = element_blank())
```




```
col <- hcl.colors(10, palette ="Pastel1")

data %>% group_by(대분류) %>% summarise(avg=mean(평균근속연수)) %>%
  ggplot(aes(x = 대분류,  y =avg, fill = 대분류))  +
  geom_col(position = "dodge",alpha=0.9) +
  scale_fill_manual(values=col)+
  theme_classic()+
  theme(legend.position='none',
        axis.ticks.x = element_blank(),
        axis.title.x=element_blank(),
        axis.title.y=element_blank(),
        strip.background = element_blank())+
  coord_cartesian(ylim=c(5,15))
```



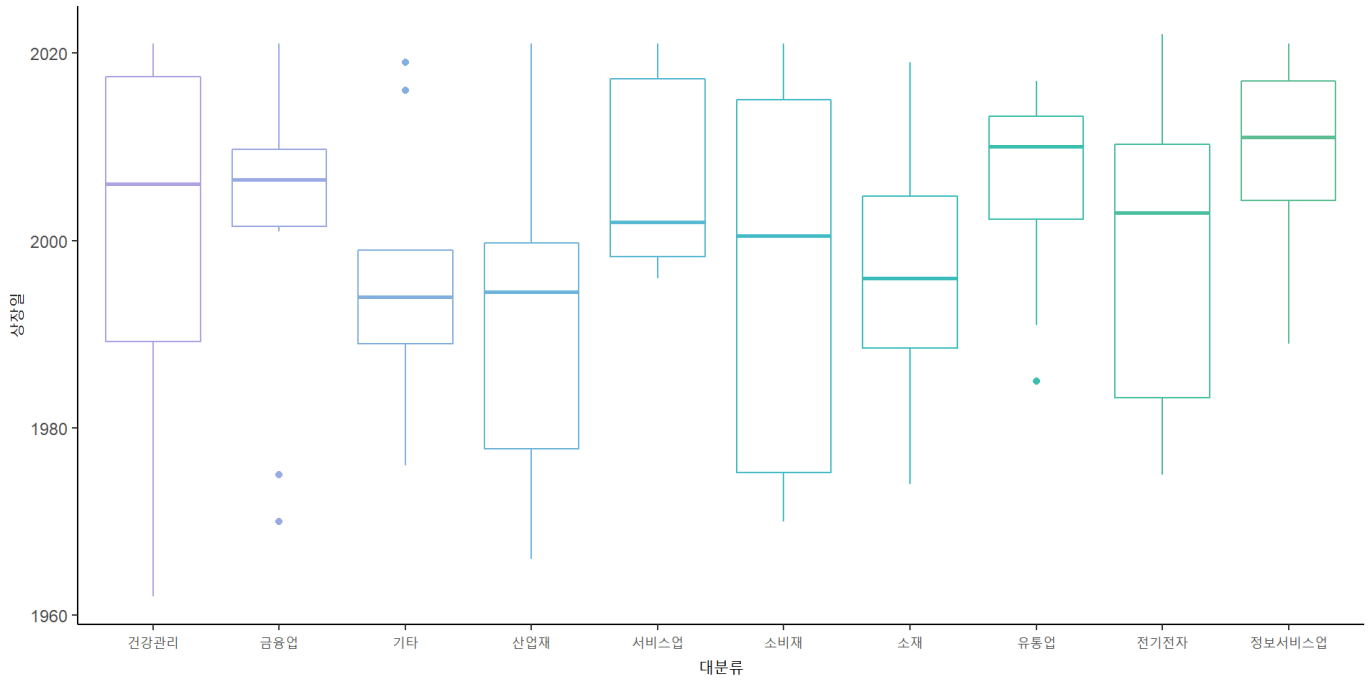

```
col <- hcl.colors(10, palette ="Pastel1")

data %>% group_by(대분류) %>% summarise(avg=mean(상장일)) %>%
  ggplot(aes(x = 대분류,  y =avg, fill = 대분류))  +
  geom_col(position = "dodge",alpha=0.9) +
  scale_fill_manual(values=col)+
  theme_classic()+
  theme(legend.position='none',
        axis.ticks.x = element_blank(),
        axis.title.x=element_blank(),
        axis.title.y=element_blank(),
        strip.background = element_blank())+
  coord_cartesian(ylim=c(1990,2022))
```



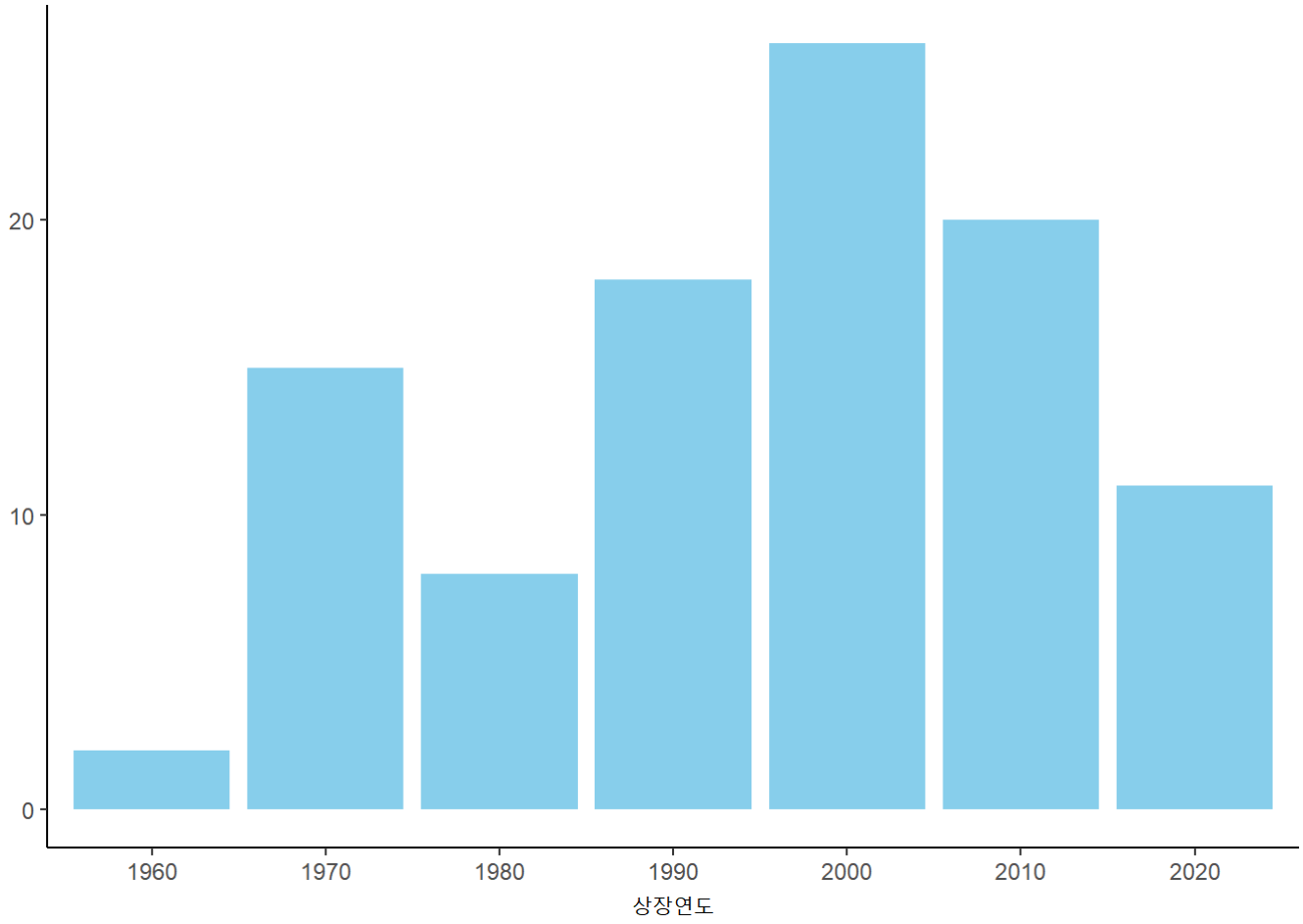
상장연도

```
data %<>% mutate('상장연도'=상장일-상장일%%10) %>%
  mutate_at(vars(상장연도),as.factor)
```

```
col <- hcl.colors(10, palette ="Cold")
data %>% ggplot(aes(대분류,상장일,color=대분류))+
  geom_boxplot()+
  theme_classic()+
  scale_color_manual(values = col)+
  theme(legend.position='none')
```

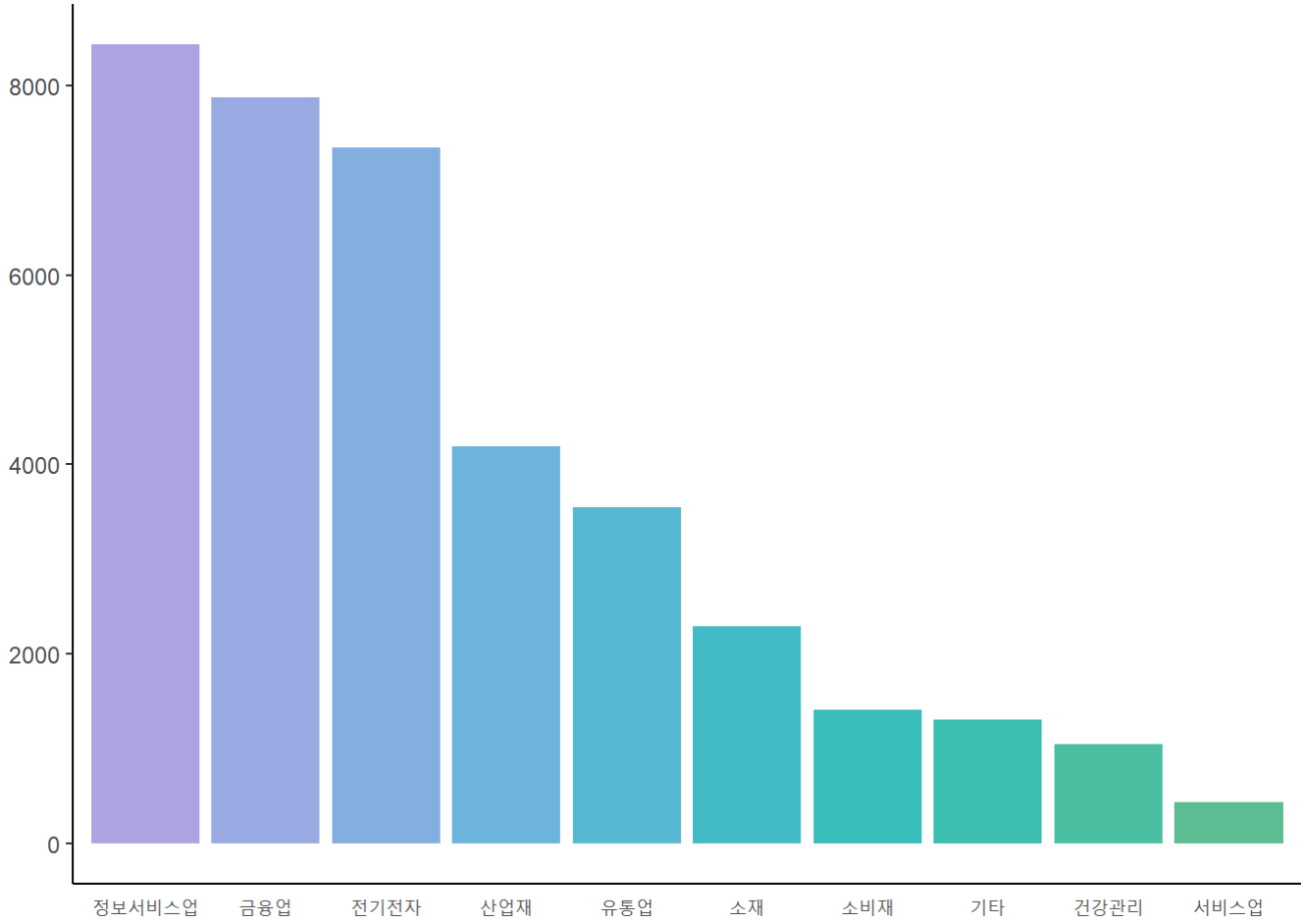



```
data %>% group_by(상장연도) %>% summarise(num=n()) %>%
  ggplot(aes(x=상장연도,y=num,fill=상장연도))+
  geom_bar(stat='identity')+
  scale_fill_manual(values=rep("skyblue",7))+
  theme_classic()+
  theme(axis.title.y=element_blank(),
        legend.position='none')
```



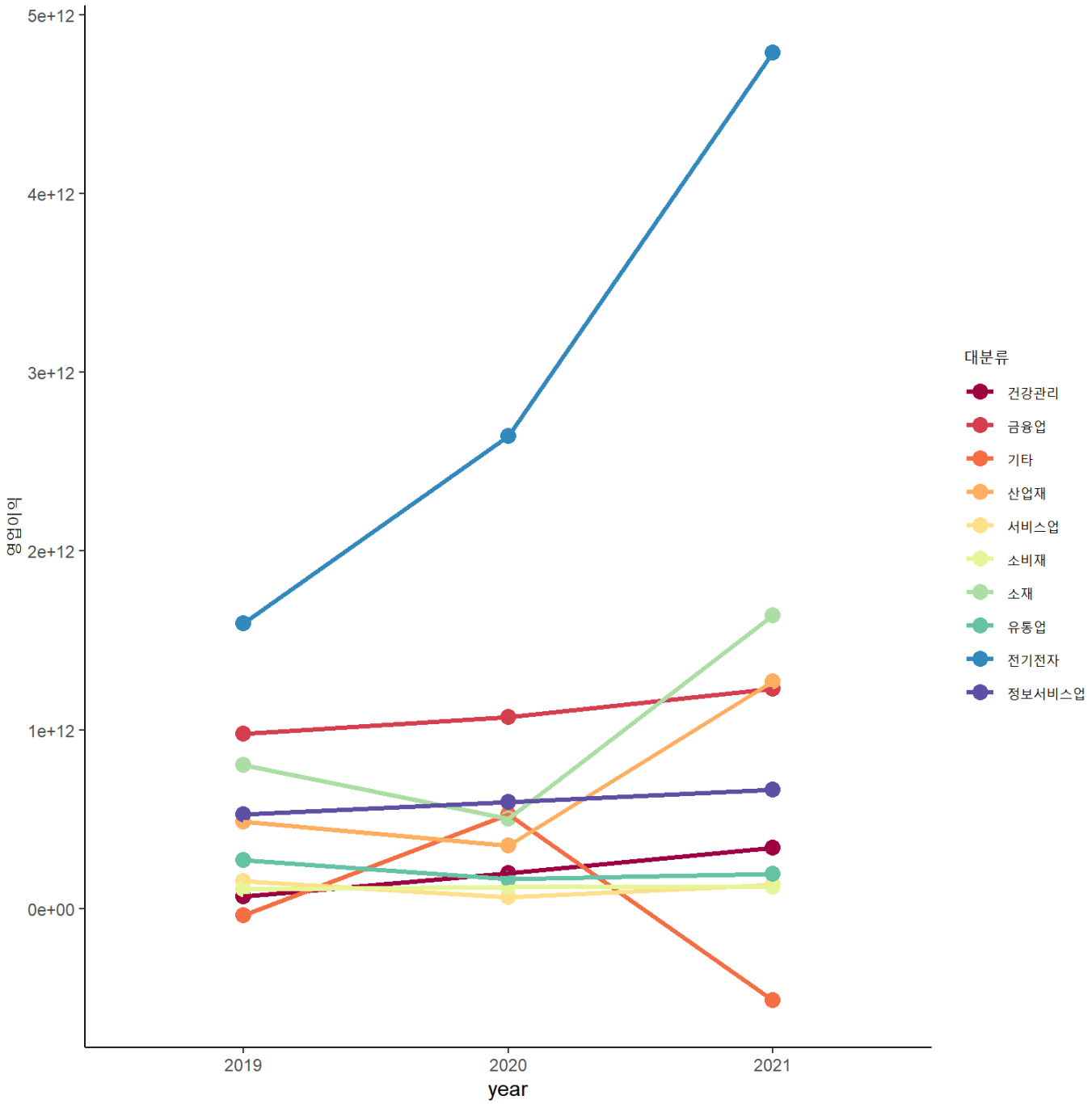
기사 수

```
col <- hcl.colors(10, palette ="Cold")

data %>% group_by(대분류) %>% summarise(sum=sum(제목)) %>%
  ggplot(aes(x = reorder(대분류,-sum),  y =sum, fill = reorder(대분류,-sum)))  +
  geom_col(position = "dodge") +
  scale_fill_manual(values=col)+
  theme_classic()+
  theme(legend.position='none',
        axis.ticks.x = element_blank(),
        axis.title.x=element_blank(),
        axis.title.y=element_blank(),
        strip.background = element_blank())
```

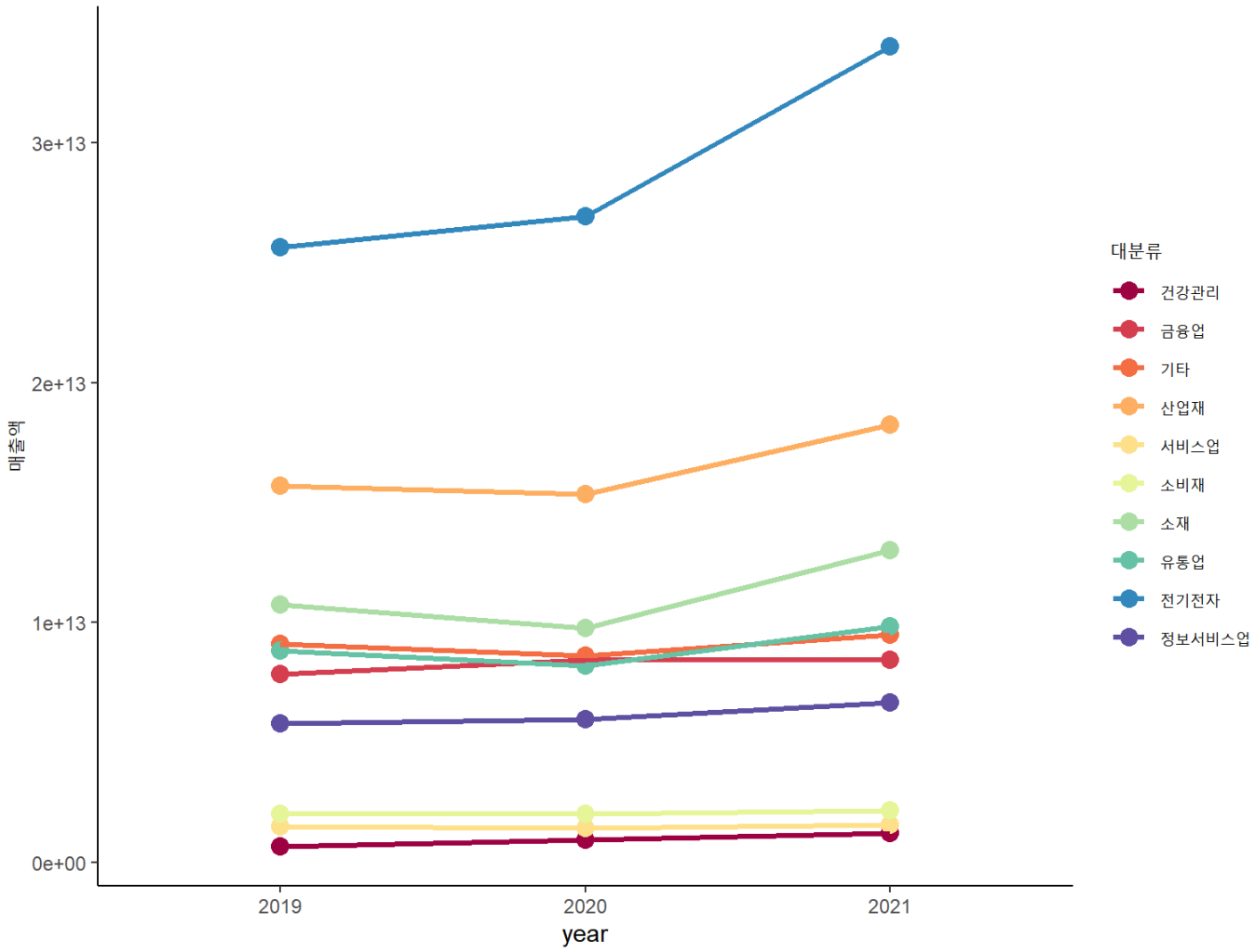


영업이익 시각화

```
data %>%
  select(회사명,대분류,`2021영업이익`,`2020영업이익`,`2019영업이익`) %>%
  group_by(대분류) %>%
  summarise_if(is.numeric,mean) %>%
  rename('2021'='2021영업이익',
         '2020'='2020영업이익',
         '2019'='2019영업이익') %>%
  gather('2021','2020','2019',key='year',value='benefit') %>%
  ggplot(aes(x=year,y=benefit,color=대분류))+
  geom_point(size=3.5)+
  geom_line(mapping=aes(group=대분류), size=1.2)+
    scale_color_brewer(palette="Spectral")+
    theme_classic()+ylab("영업이익")
```




```
data %>%
  select(회사명,대분류,`2021매출액`,`2020매출액`,`2019매출액`) %>%
  group_by(대분류) %>%
  summarise_if(is.numeric,mean) %>%
  rename('2021'='2021매출액',
         '2020'='2020매출액',
         '2019'='2019매출액') %>%
  gather('2021','2020','2019',key='year',value='benefit') %>%
  ggplot(aes(x=year,y=benefit,color=대분류))+
  geom_point(size=3.5)+
  geom_line(mapping=aes(group=대분류), size=1.2)+
    scale_color_brewer(palette="Spectral")+
    theme_classic()+ylab("매출액")
```


3e+13
2e+13
1e+13
0e+00
매출액
2019
2020
2021
year
대분류
건강관리
금융업
기타
산업재
서비스업
소비재
소재
유통업
전기전자
정보서비스업