

Inequality in Access to Recreational Facilities in Seoul, South Korea

Kibin Ahn

Introduction

Participation in recreational sports has increased substantially in South Korea over the past two decades. According to the National Survey on Sports Participation, the share of individuals engaging in physical activity at least once a week rose from around 33 percent in 2000 to over 60 percent in recent years. This trend reflects growing demand for recreational facilities and active lifestyles. In addition, as of 2025, approximately 60 percent of individuals engage in physical activity at either public or private recreational facilities, highlighting the central role of such facilities in supporting active lifestyles (Ministry of Culture, Sports and Tourism, 2025).

Against this backdrop, access to recreational facilities has become an increasingly important policy issue. This is particularly relevant in Seoul, where a large share of the population is concentrated, and where disparities in access may have significant implications for equity and well-being. Focusing on Seoul as a case study, this paper examines how access to public and private recreational facilities varies across districts.

Seoul consists of 25 districts (gu), including three widely recognized affluent districts located in the southern part of the city, Gangnam-gu, Seocho-gu, and Songpa-gu, collectively referred to as the “Gangnam 3.”

Figure 1. Location of the Gangnam 3



This paper examines whether differences in fiscal capacity across districts translate into disparities in access to public and private recreational facilities. It further explores the extent to which public provision mitigates inequality, in contrast to private provision, and discusses the policy implications for improving equitable access to recreational infrastructure.

In this study, public recreational facilities refer to government-provided sports and recreational infrastructure, including various types of publicly operated athletic facilities reported in the Seoul Open Data Portal. Private recreational facilities are defined as privately operated sports and recreational facilities that are registered or reported to the government, including fitness centers, sports clubs, and other commercial recreational establishments.

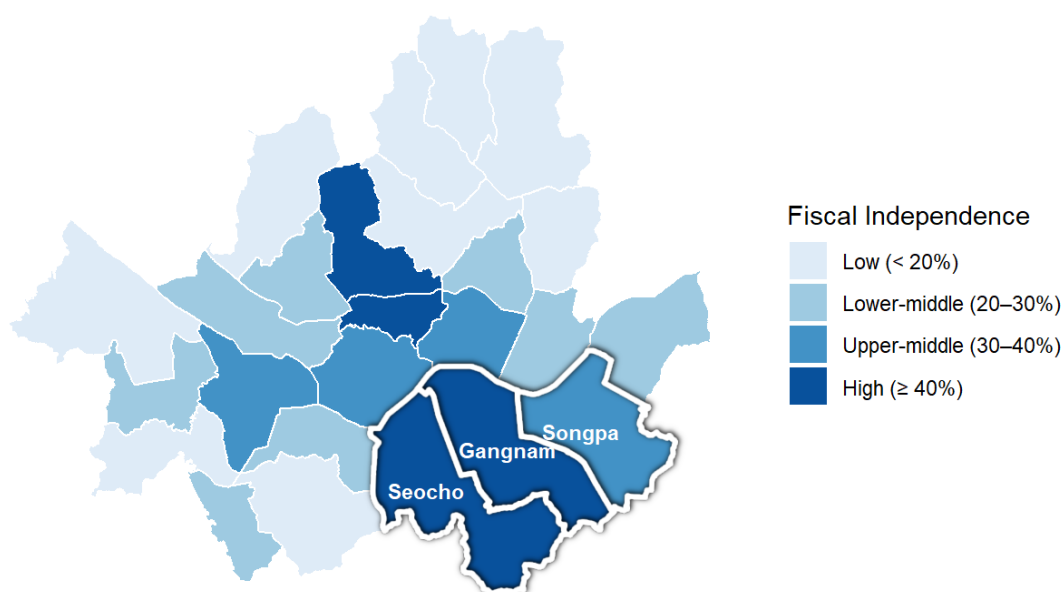
How are public recreational facilities distributed across districts?

One might expect that districts with greater fiscal capacity, that is, wealthier districts, would have better access to public recreational facilities, because they have more resources to invest in public services. To examine this, I first explore variation in fiscal capacity across districts using the fiscal independence ratio in Figure 2. Fiscal independence is calculated as the ratio of own-source revenue, including local taxes and non-tax revenue, to the general accounting budget, multiplied by 100. As shown in the map, fiscal capacity varies substantially across districts, with Gangnam 3 concentrated at the higher end of the distribution.

Figure 2. Fiscal Capacity across Districts

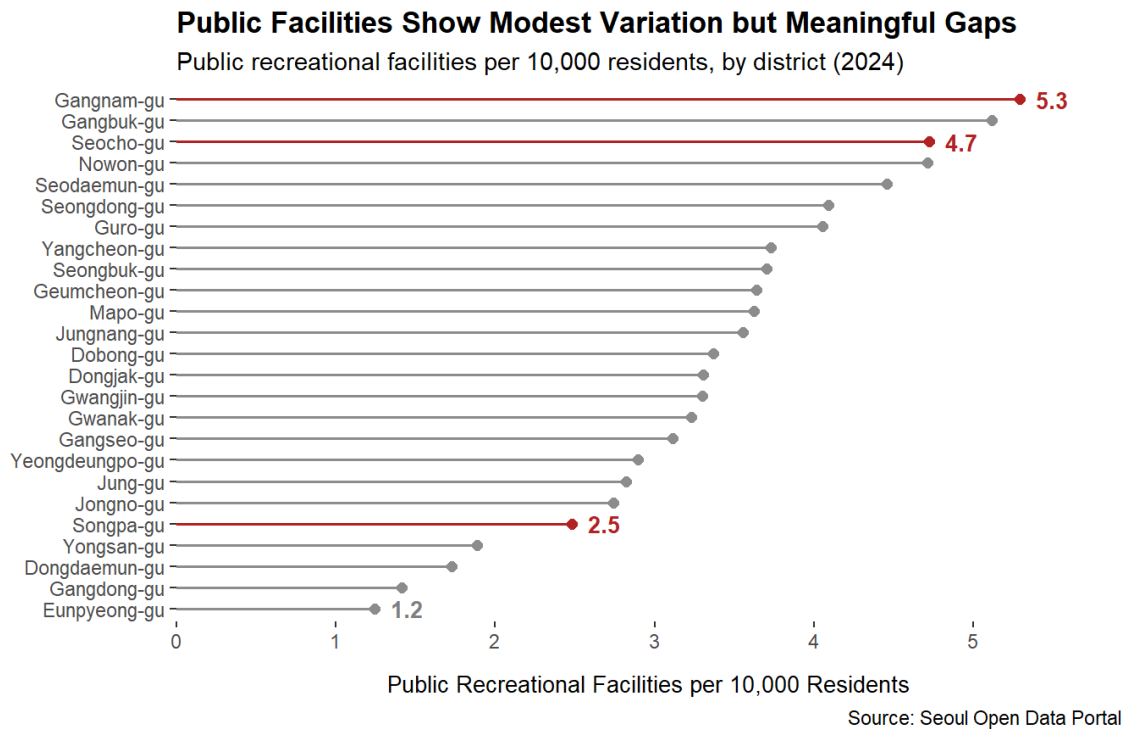
Fiscal Capacity Varies Across Districts

Fiscal independence ratio (%), by district (2024)

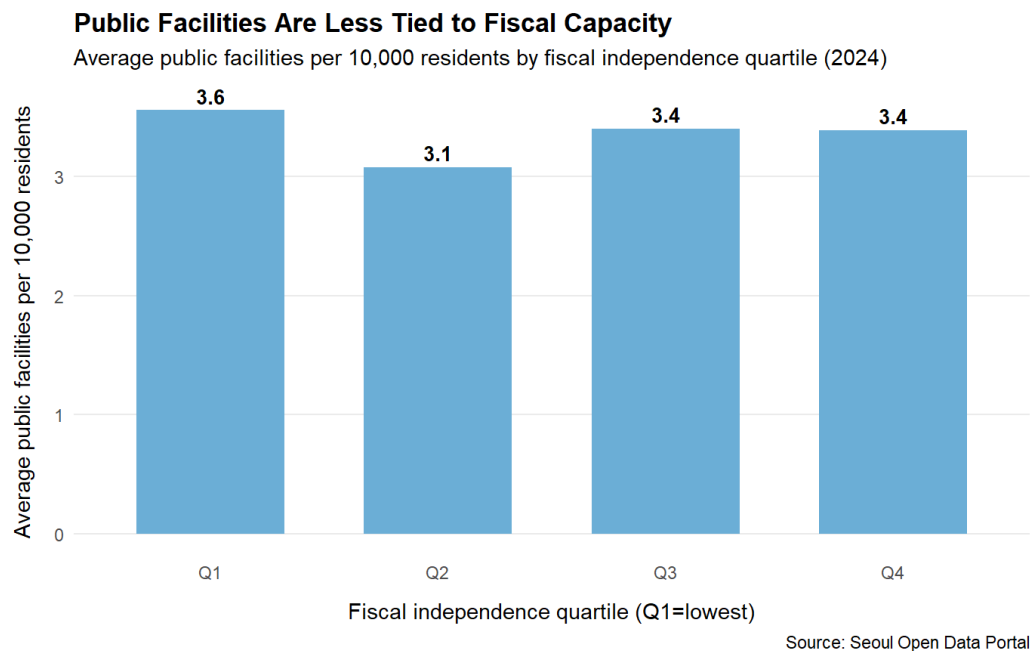


Source: Seoul Open Data Portal

Given these disparities in fiscal capacity, I next examine whether they translate into differences in access to public facilities. As shown in Figure 3, although Gangnam 3 does not consistently rank at the top, there remains a noticeable range across districts. The highest district has about 5.3 facilities per 10,000 residents, while the lowest district has about 1.2. Although the absolute range is relatively modest, the difference remains meaningful, with the highest district having more than four times as many facilities as the lowest.

Figure 3. Public Recreational Facilities by District

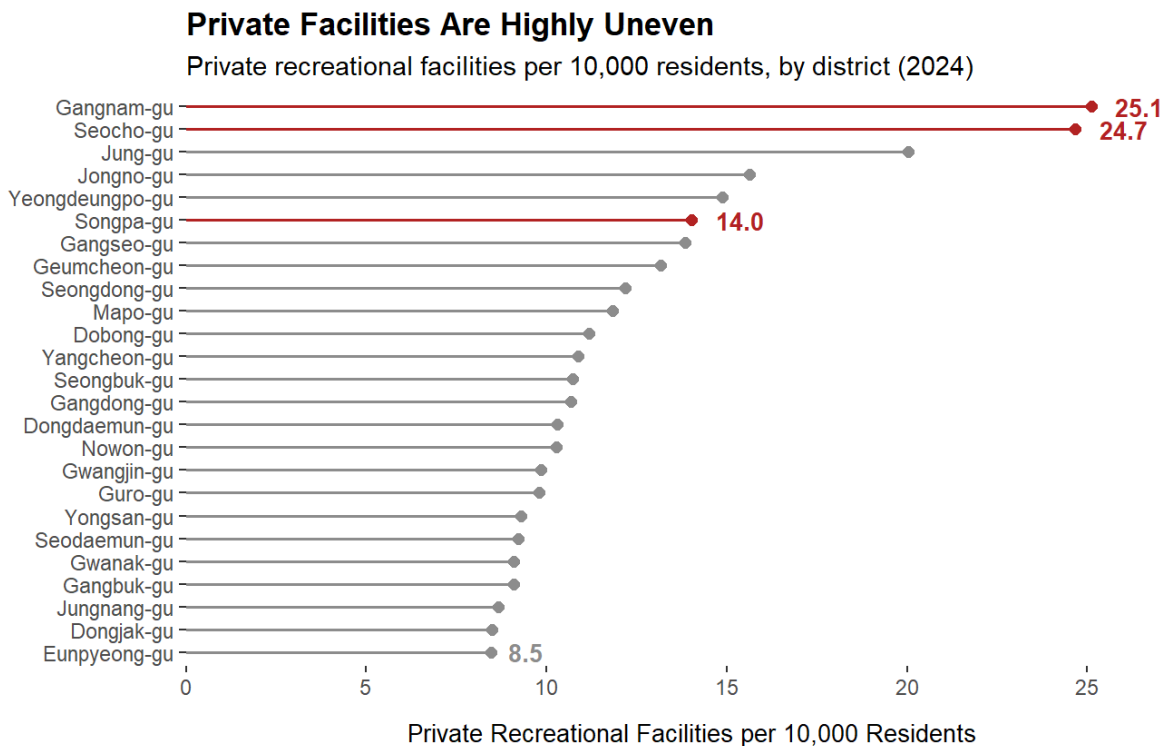
However, this variation does not appear to systematically align with fiscal capacity. As shown in Figure 4, average public facilities are broadly similar across fiscal independence quartiles, suggesting that differences in public provision are unlikely to be driven by fiscal capacity and may instead reflect redistribution policies or differences in public demand.

Figure 4. Average Public Facilities by Fiscal Independence Quartile

How are private recreational facilities distributed across districts?

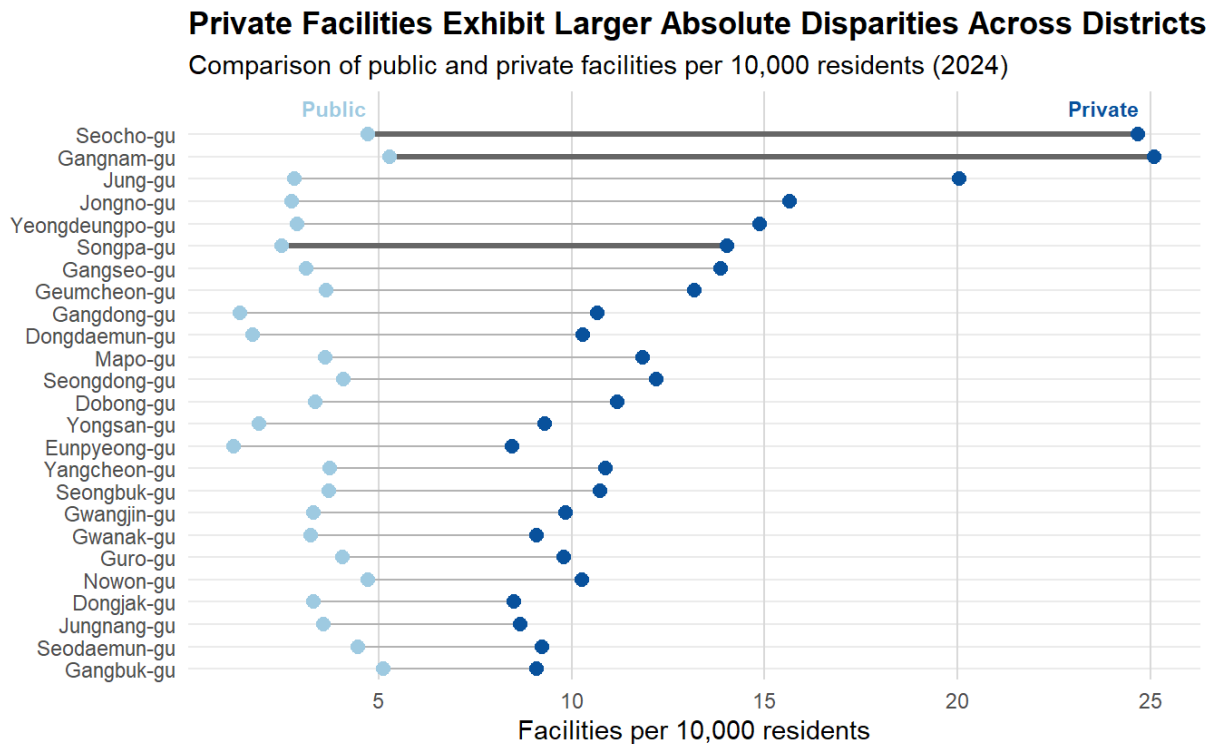
For private recreational facilities, substantial disparities are also observed. Figure 5 shows that districts such as Gangnam-gu and Seocho-gu have about 25 private recreational facilities per 10,000 residents, while at the lower end, Eunpyeong-gu has only about 8.5. The ratio between the highest and lowest districts is approximately threefold, which is slightly smaller than for public facilities. However, the absolute difference remains substantial at around 17 facilities.

Figure 5. Private Recreational Facilities by District



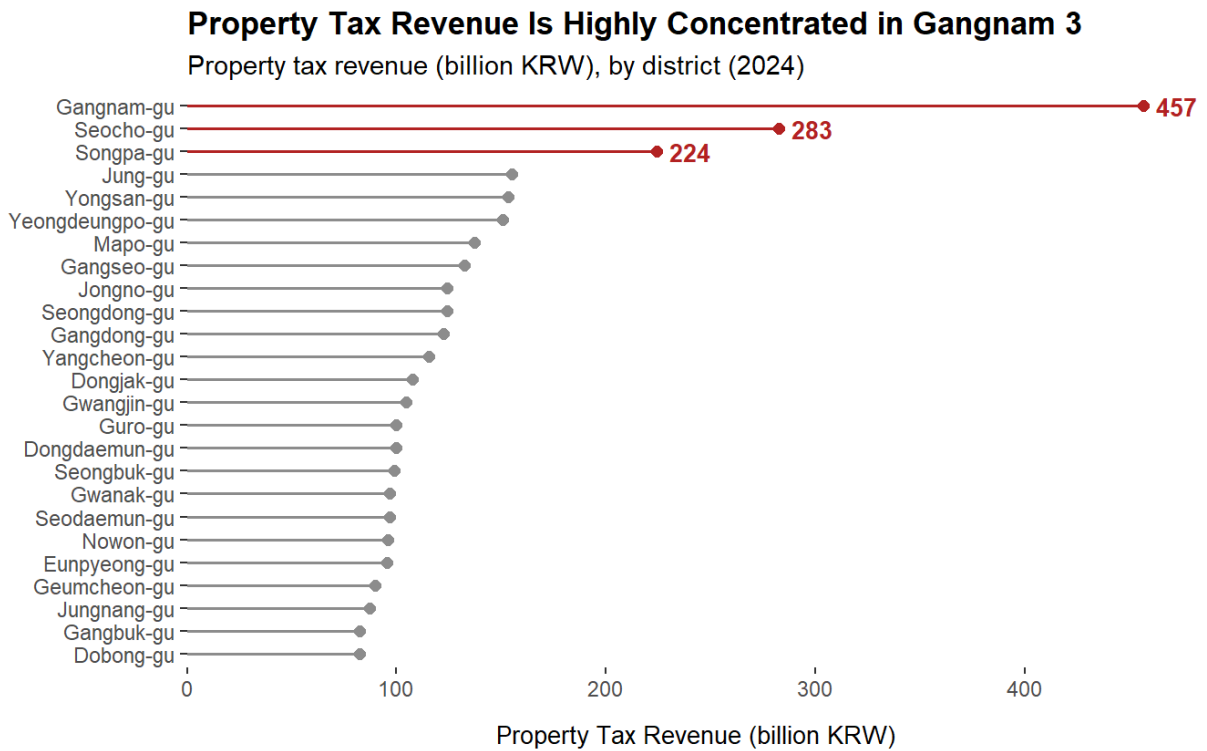
Source: Seoul Open Data Portal

Comparing public and private facilities together in Figure 6, while public facilities also show non-trivial variation across districts, private facilities exhibit larger disparities in absolute terms. In other words, although relative differences are substantial for both, public facilities exhibit a larger relative gap, whereas private facilities show much larger absolute differences.

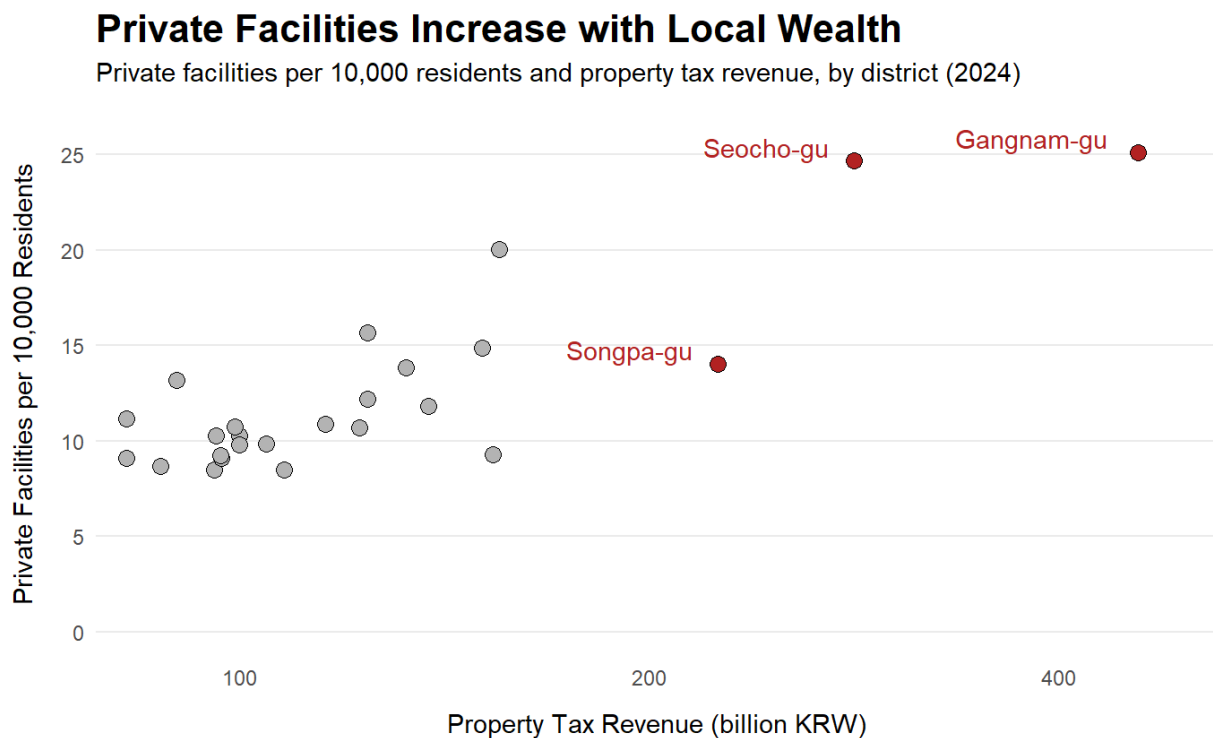
Figure 6. Public and Private Recreational Facilities by District

To better understand the drivers of these disparities, I examine how the distribution of private facilities relates to local economic conditions. Notably, Songpa-gu moves much higher in the distribution of private facilities compared to public facilities. Figure 7 illustrates that property tax revenue is highly concentrated in a few districts, particularly in Gangnam 3. In this study, property tax revenue refers to taxes levied on land and buildings, including both residential and commercial properties. As such, the level of property tax revenue reflects underlying property values and local economic activity across districts. Districts with higher property tax revenue tend to have greater economic resources and attract more private investment, creating strong incentives for private recreational facilities to cluster in these districts.

To further examine this relationship, Figure 8 plots the association between property tax revenue and private recreational facilities. The results show a clear positive relationship between local wealth and the number of private facilities. Districts with higher levels of property tax revenue tend to have more private facilities, indicating a strong positive association between local wealth and private facility access.

Figure 7. Property Tax Revenue across Districts

Source: Seoul Open Data Portal

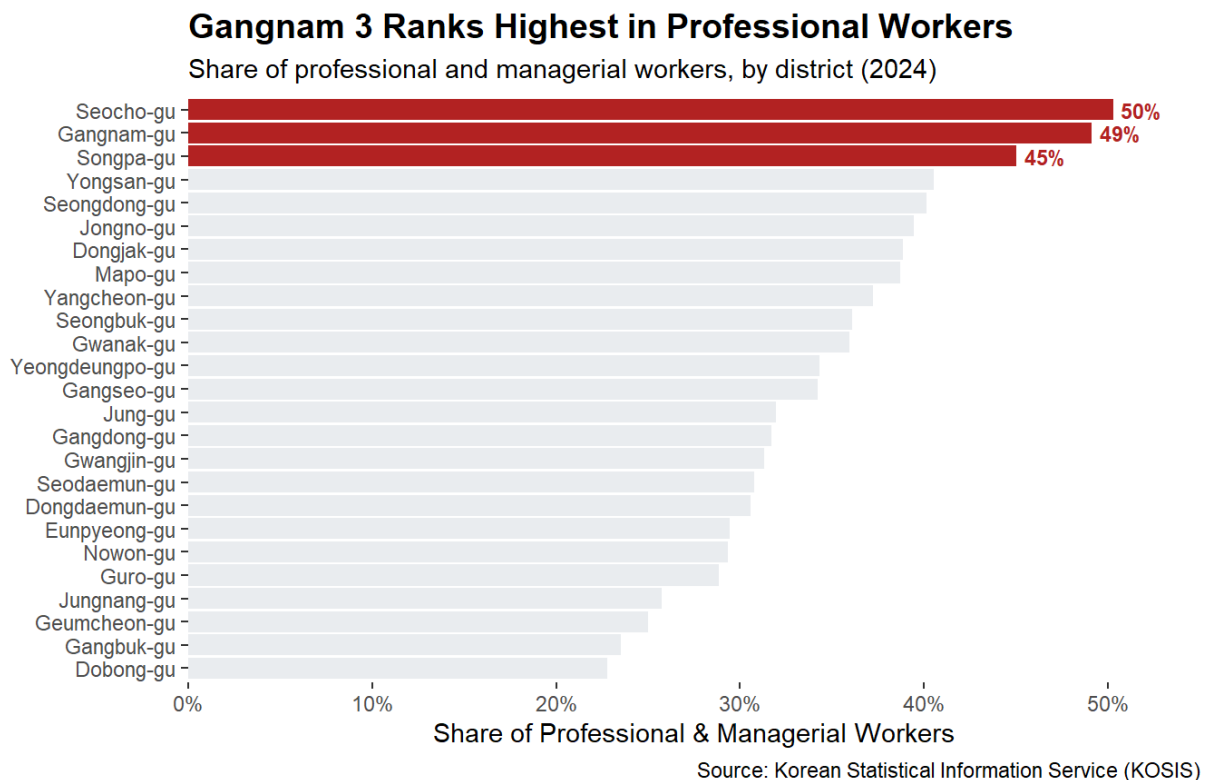
Figure 8. Scatterplot of Property Tax Revenue and Private Facilities

Note: Property tax revenue is shown on a log scale.

Source: Seoul Open Data Portal

To help explain why property tax revenue differs across districts, I examine the distribution of professional and managerial workers. According to official classifications by the Korean Ministry of Data and Statistics, professional and managerial occupations are generally associated with higher income levels and economic activity. Figure 9 shows that Gangnam 3 districts have the highest share of professional and managerial workers, accounting for approximately 45–50 percent. This concentration of high-income occupations provides a plausible explanation for the higher levels of property tax revenue observed in these districts.

Figure 9. Share of Professional and Managerial Workers by District



Policy Implications and Next Steps

Overall, access to both public and private recreational facilities varies across districts. Private recreational facilities are strongly concentrated in wealthier districts, reflecting market-driven investment patterns, while public provision, although more evenly distributed in absolute terms, does not fully offset these disparities.

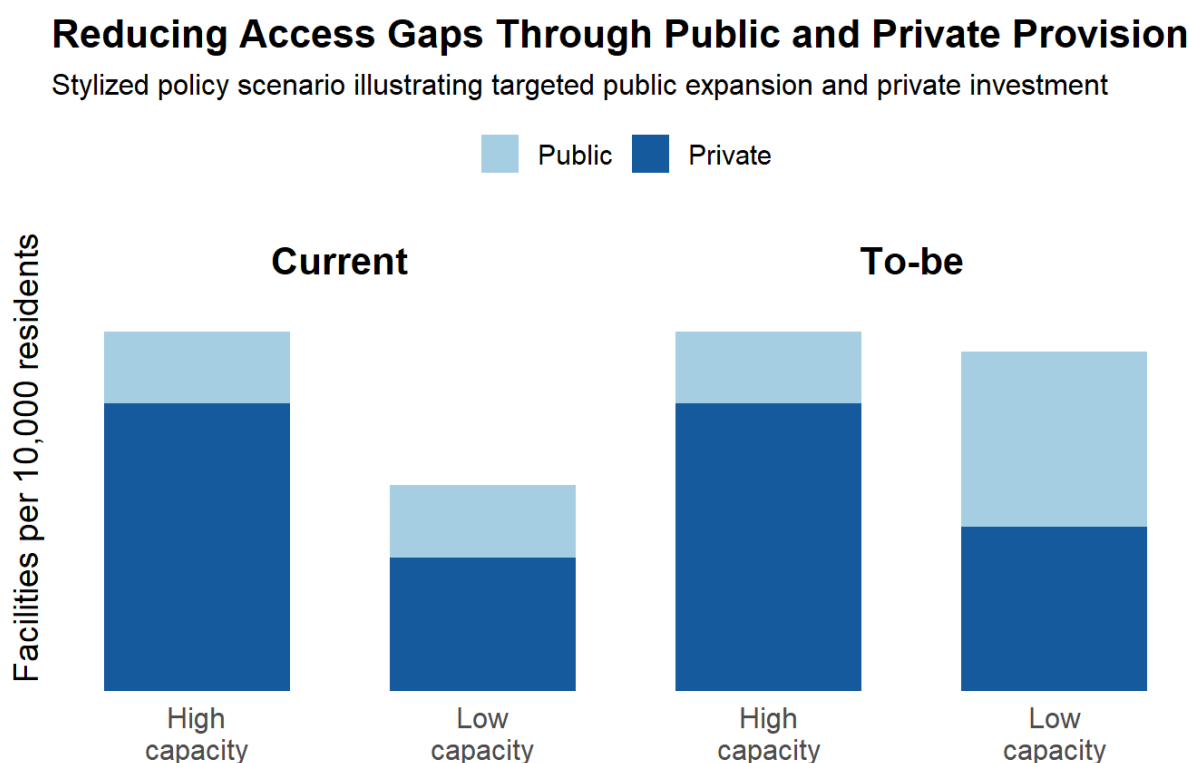
A key policy implication is that public provision should play a more targeted role in mitigating inequalities created by private markets. Public investment should be prioritized in districts that are underserved by private recreational facilities, where market provision remains limited, rather than being distributed uniformly. Such a targeted approach would help ensure more equitable

access to recreational infrastructure across districts. At the same time, policies should also consider ways to encourage private investment in underserved areas. This may include providing financial incentives, reducing regulatory barriers, or supporting public–private partnerships to expand access where market incentives are currently weak.

Figure 10 presents a stylized policy scenario comparing current and improved distributions of public and private recreational facilities across districts with different fiscal capacity. Under the current scenario, low-capacity districts have fewer total facilities, driven primarily by lower levels of private provision. In the to-be scenario, targeted expansion of public facilities in low-capacity districts, along with modest growth in private provision, reduces the gap in total access between districts. This illustrates how a combined strategy of targeted public investment and private sector incentives can effectively mitigate disparities in access.

Taken together, a balanced strategy that combines targeted public provision with incentives for private sector participation can help improve both the equity and overall availability of recreational facilities.

Figure 10. Policy Scenario: Reducing Access Gaps



References

Ministry of Culture, Sports and Tourism. *National Survey on Sports Participation* (various years). <https://www.index.go.kr/unity/potal/indicator/IndexInfo.do?cdNo=2&clasCd=2&idxCd=4252&upCd=10>

Appendix: R Code

```
rm(list=ls())

occup <- read.csv("C:/Users/kibin/Desktop/Spring 2026/PPPA 6021 Data Visualization/policy
brief/occupationalcomposition.csv")
str(occup)

library(tidyverse)

occup <- read.csv("C:/Users/kibin/Desktop/Spring 2026/PPPA 6021 Data Visualization/policy
brief/occupationalcomposition.csv") %>%
  slice(-1) %>%
  filter(str_detect(행정구역별, "^서울")) %>%
  mutate(
    행정구역별 = str_remove(행정구역별, "^서울\\s*") %>% str_trim(),
    행정구역별 = recode(행정구역별,
      "종로구"="Jongno-gu", "중구"="Jung-gu", "용산구"="Yongsan-gu", "성동구"="Seongdong-gu",
      "광진구"="Gwangjin-gu", "동대문구"="Dongdaemun-gu", "중랑구"="Jungnang-gu", "성북구"
"="Seongbuk-gu",
      "강북구"="Gangbuk-gu", "도봉구"="Dobong-gu", "노원구"="Nowon-gu", "은평구"
"="Eunpyeong-gu",
      "서대문구"="Seodaemun-gu", "마포구"="Mapo-gu", "양천구"="Yangcheon-gu", "강서구"
"="Gangseo-gu",
      "구로구"="Guro-gu", "금천구"="Geumcheon-gu", "영등포구"="Yeongdeungpo-gu", "동작구"
"="Dongjak-gu",
      "관악구"="Gwanak-gu", "서초구"="Seocho-gu", "강남구"="Gangnam-gu", "송파구"="Songpa-
gu", "강동구"="Gangdong-gu"),
    직업별 = str_trim(직업별),
    직업별 = recode(직업별,
      "계"="Total",
      "관리자, 전문가 및 관련종사자"="Professional & Managerial",
      "사무 종사자"="Clerical Workers",
      "서비스·판매 종사자"="Service & Sales Workers",
      "농림어업 숙련종사자"="Skilled Agricultural Workers",
      "기능·기계조작·조립 종사자"="Craft, Machine Operators & Assemblers",
      "단순노무 종사자"="Elementary Occupations")) %>%
  rename(employed_thousand = 3, rse = 4) %>%
  mutate(employed_thousand = na_if(employed_thousand, "-") %>% as.numeric()) %>%
  select(-rse)

occup %>% count(행정구역별) %>% arrange(행정구역별)
n_distinct(occup$행정구역별)
```

```

sum(is.na(occup$employed_thousand))
occup %>% filter(is.na(employed_thousand))

occup$employed_thousand <- ifelse(test = is.na(occup$employed_thousand)==TRUE,
                                   yes = 0,
                                   no = occup$employed_thousand)
sum(is.na(occup$employed_thousand))

library(scales)
library(ggplot2)

# proportion of professional and managerial
total_counts <- occup %>%
  filter(직업별=="Total") %>%
  select(행정구역별, total_emp=employed_thousand)

prof_share <- occup %>%
  filter(직업별=="Professional & Managerial") %>%
  left_join(total_counts, by="행정구역별") %>%
  mutate(share=employed_thousand/total_emp)

prof_share <- prof_share %>%
  mutate(group=ifelse(행정구역별 %in% c("Gangnam-gu", "Seocho-gu", "Songpa-gu"),
                      "Gangnam 3", "Others"))

two.colors <- c("Gangnam 3"="firebrick",
               "Others"="#E9ECEF")

prof_share <- prof_share %>%
  rename(district=행정구역별)

occup <- occup %>%
  rename(district=행정구역별)

# professional share
c1 <- ggplot()+
  geom_col(data=prof_share,
           mapping=aes(x=reorder(district, share), y=share, fill=group))+
  labs(title="Gangnam 3 Ranks Highest in Professional Workers",
       subtitle="Share of professional and managerial workers, by district (2024)",
       x="",
       y="Share of Professional & Managerial Workers",
       caption = "Source: Korean Statistical Information Service (KOSIS)") +
  geom_text(data = subset(prof_share, district %in% c("Seocho-gu", "Gangnam-gu", "Songpa-gu")),
           aes(x=reorder(district,share), y=share,
               label = paste0(round(share*100), "%")),
           hjust = -0.2,
           size = 3,
           color = "firebrick",
           fontface = "bold") +

```

```

scale_fill_manual(values=two.colors)+
theme(panel.background=element_blank(),
      legend.position="none",
      panel.grid.minor=element_blank(),
      plot.title=element_text(face="bold", size=14),
      plot.subtitle=element_text(size=11))+
scale_y_continuous(labels=scales::percent,
                  limits = c(0, 0.55),
                  breaks = seq(0, 0.6, by = 0.1),
                  expand=c(0,0))+

coord_flip()
c1

# property tax dataset
prop.tax <- read.csv("C:/Users/kibin/Desktop/Spring 2026/PPPA 6021 Data Visualization/policy
brief/localpropertytax.csv")
names(prop.tax)
head(prop.tax)

prop.tax <- prop.tax[!grepl("구 별\\(1\\)", prop.tax$구 별.1.), ]
head(prop.tax)
prop.tax$X2024 <- as.numeric(prop.tax$X2024)
names(prop.tax) <- c("district", "property_tax")

library(dplyr)
prop.tax <- prop.tax %>%
  mutate(district = recode(district,
                           "종로구"="Jongno-gu", "중구"="Jung-gu", "용산구"="Yongsan-gu", "성동구"
="Seongdong-gu",
                           "광진구"="Gwangjin-gu", "동대문구"="Dongdaemun-gu", "중랑구"="Jungnang-gu", "성
북구"="Seongbuk-gu",
                           "강북구"="Gangbuk-gu", "도봉구"="Dobong-gu", "노원구"="Nowon-gu", "은평구"
="Eunpyeong-gu",
                           "서대문구"="Seodaemun-gu", "마포구"="Mapo-gu", "양천구"="Yangcheon-gu", "강서
구"="Gangseo-gu",
                           "구로구"="Guro-gu", "금천구"="Geumcheon-gu", "영등포구"="Yeongdeungpo-gu", "동작구"
="Dongjak-gu",
                           "관악구"="Gwanak-gu", "서초구"="Seocho-gu", "강남구"="Gangnam-gu", "송파
구"="Songpa-gu", "강동구"="Gangdong-gu"))

names(prop.tax)
head(prop.tax)

unique(prop.tax$district)
nrow(prop.tax)

# change to billion
prop.tax <- prop.tax %>%
  mutate(property_tax_billion = property_tax / 1000)

```

```

# make it lollipop
prop.tax <- prop.tax %>%
  mutate(group=ifelse(district %in% c("Gangnam-gu", "Seocho-gu", "Songpa-gu"),
    "Gangnam 3", "Others"))

# property tax revenue by district
c3 <- ggplot()+
  geom_point(data=prop.tax,
    mapping=aes(x=property_tax_billion, y=reorder(district, property_tax_billion),
      color=group), size=2)+
  geom_segment(data=prop.tax,
    mapping=aes(x=0, xend=property_tax_billion,
      y=district, yend=reorder(district, property_tax_billion),
      color=group),
    linewidth=0.7)+
  scale_x_continuous(labels=comma,
    expand=expansion(mult=c(0,0.05)))+
  scale_color_manual(values = c("Gangnam 3" = "firebrick", "Others" = "grey55"))+
  labs(title="Property Tax Revenue Is Highly Concentrated in Gangnam 3",
    subtitle="Property tax revenue (billion KRW), by district (2024)",
    x="Property Tax Revenue (billion KRW)", y="",
    caption="Source: Seoul Open Data Portal")+
  theme(legend.position="none",
    panel.background=element_rect(fill="white", color=NA),
    plot.background=element_rect(fill="white", color=NA),
    panel.grid.major.y=element_blank(),
    panel.grid.major.x=element_blank(),
    panel.grid.minor=element_blank(),
    plot.title=element_text(face="bold"),
    axis.title.x=element_text(size=10, margin=margin(t=10)))+
  geom_text(data=subset(prop.tax, group == "Gangnam 3"),
    aes(x=property_tax_billion + 6,
      y=reorder(district, property_tax_billion),
      label=round(property_tax_billion, 0),
      color=group),
    hjust=0,
    size=3.6,
    fontface="bold")

```

c3

```

# scatter plot
# data merge
names(prop_share)
names(prop.tax)

merged_sp <- merge(prop_share, prop.tax, by="district")
names(merged_sp)
str(merged_sp)
merged_sp <- merged_sp %>%
  mutate(group=group.x) %>%
  select(-group.x, -group.y)
str(merged_sp)

```

```

# outliers
merged_sp$log_tax <- log(merged_sp$property_tax_billion)

lab_etc <-merged_sp %>%
  filter(district %in% c("Seocho-gu", "Songpa-gu"))

lab_gn <-merged_sp %>%
  filter(district=="Gangnam-gu")

# scatter plot
library(ggplot2)

# fiscal independence
fis.ind <- read.csv("C:/Users/kibin/Desktop/Spring 2026/PPPA 6021 Data Visualization/policy
brief/fiscalindependence.csv")
names(fis.ind)
head(fis.ind)

# label to English
fis.ind2 <- fis.ind[!(fis.ind$구분별.1. %in% c("서울시 평균", "본청", "자치구 평균")), ]
fis.ind2$X2024 <- as.numeric(fis.ind2$X2024)
head(fis.ind2)
names(fis.ind2) <- c("district", "fiscal_index")
head(fis.ind2)
fis.ind2 <- fis.ind2 %>%
  mutate(district=recode(district,
    "종로구"="Jongno-gu", "중구"="Jung-gu", "용산구"="Yongsan-gu", "성동구"="Seongdong-gu",
    "광진구"="Gwangjin-gu", "동대문구"="Dongdaemun-gu", "중랑구"="Jungnang-gu", "성북구"="Seongbuk-gu",
    "강북구"="Gangbuk-gu", "도봉구"="Dobong-gu", "노원구"="Nowon-gu", "은평구"="Eunpyeong-gu",
    "서대문구"="Seodaemun-gu", "마포구"="Mapo-gu", "양천구"="Yangcheon-gu", "강서구"="Gangseo-gu",
    "구로구"="Guro-gu", "금천구"="Geumcheon-gu", "영등포구"="Yeongdeungpo-gu", "동작구"="Dongjak-gu",
    "관악구"="Gwanak-gu", "서초구"="Seocho-gu", "강남구"="Gangnam-gu", "송파구"="Songpa-gu", "강동구"="Gangdong-gu"))

# map
library(sf)
seoul.map <- st_read("C:/Users/kibin/Desktop/Spring 2026/PPPA 6021 Data Visualization/policy
brief/data/gu_seoul.shp", options = "ENCODING=CP949")
head(seoul.map)

names(seoul.map)[names(seoul.map) == "gu"] <- "district"
seoul.map$district <- trimws(seoul.map$district)
fis.ind2$district <- trimws(fis.ind2$district)

```

```

map.data <- dplyr::left_join(seoul.map, fis.ind2, by = "district")
head(map.data)
sum(is.na(map.data$fiscal_index))

map.data <- map.data %>%
  mutate(district_eng = recode(district,
                                "종로구" = "Jongno-gu", "중구" = "Jung-gu",
                                "용산구" = "Yongsan-gu", "성동구" = "Seongdong-gu",
                                "광진구" = "Gwangjin-gu", "동대문구" = "Dongdaemun-gu",
                                "중랑구" = "Jungnang-gu", "성북구" = "Seongbuk-gu",
                                "강북구" = "Gangbuk-gu", "도봉구" = "Dobong-gu",
                                "노원구" = "Nowon-gu", "은평구" = "Eunpyeong-gu",
                                "서대문구" = "Seodaemun-gu", "마포구" = "Mapo-gu",
                                "양천구" = "Yangcheon-gu", "강서구" = "Gangseo-gu",
                                "구로구" = "Guro-gu", "금천구" = "Geumcheon-gu",
                                "영등포구" = "Yeongdeungpo-gu", "동작구" = "Dongjak-gu",
                                "관악구" = "Gwanak-gu", "서초구" = "Seocho-gu",
                                "강남구" = "Gangnam-gu", "송파구" = "Songpa-gu",
                                "강동구" = "Gangdong-gu"))

head(map.data[, c("district", "fiscal_index")])
summary(map.data$fiscal_index)
sum(is.na(map.data$fiscal_index))
class(map.data$fiscal_index)

names(fis.ind) <- c("district", "fiscal_index")

fis.ind2 <- fis.ind[!(fis.ind$district %in% c("서울시 평균", "본청", "자치구 평균")), ]
fis.ind2$district <- trimws(fis.ind2$district)
fis.ind2$fiscal_index <- as.numeric(fis.ind2$fiscal_index)
map.data <- dplyr::left_join(seoul.map, fis.ind2, by = "district")

head(map.data[, c("district", "fiscal_index")])
sum(is.na(map.data$fiscal_index))

seoul.map$district <- as.character(seoul.map$district)
fis.ind2$district <- as.character(fis.ind2$district)

seoul.map$district <- trimws(seoul.map$district)
fis.ind2$district <- trimws(fis.ind2$district)

# map
map.data <- map.data %>%
  mutate(
    fiscal_group = case_when(
      fiscal_index < 20 ~ "Low",
      fiscal_index < 30 ~ "Lower-middle",

```

```

    fiscal_index < 40 ~ "Upper-middle",
    TRUE ~ "High"))
unique(map.data$fiscal_group)
map.data$fiscal_group <- factor(map.data$fiscal_group,
                               levels = c("Low", "Lower-middle", "Upper-middle", "High"))

map.data <- map.data %>%
  mutate(district_eng = recode(district,
                                "강남구" = "Gangnam",
                                "서초구" = "Seocho",
                                "송파구" = "Songpa"))

gangnam3 <- subset(map.data, district %in% c("강남구", "서초구", "송파구"))
gangnam3$label <- c("Seocho", "Gangnam", "Songpa")

ind.color <- c("Low" = "#deebf7",
               "Lower-middle" = "#9ecae1",
               "Upper-middle" = "#4292c6",
               "High" = "#08519c")

library(ggfx)

c5 <- ggplot()+
  geom_sf(data=map.data,
          mapping=aes(fill = fiscal_group), color = "white", linewidth = 0.4)+
  with_shadow(
    geom_sf(data=subset(map.data, district %in% c("강남구", "서초구", "송파구")),
            fill=NA, color="white", linewidth=1.1),
    sigma=3, x_offset=1.2, y_offset=-1.2,
    color="black", alpha=0.22)+
  geom_sf_text(data=gangnam3,
              mapping=aes(label = label),
              color = "white", size = 3, fontface = "bold")+
  theme_minimal()+
  scale_fill_manual(values=ind.color, name = "Fiscal Independence",
                    labels = c("Low" = "Low (< 20%)", "Lower-middle" = "Lower-middle (20–30%)",
                               "Upper-middle" = "Upper-middle (30–40%)", "High" = "High (≥ 40%)"))+
  labs(title="Fiscal Capacity Varies Across Districts",
       subtitle="Fiscal independence ratio (%), by district (2024)",
       fill="Fiscal Independence",
       caption="Source: Seoul Open Data Portal")+
  theme(plot.title=element_text(face="bold", size=16, lineheight=1, margin=margin(b=5)),
        plot.subtitle=element_text(margin=margin(b=10)),
        plot.background=element_rect(fill = "white", color = NA),
        panel.grid=element_blank(),
        axis.text=element_blank(),
        axis.title=element_blank(),
        plot.caption=element_text(size = 9, hjust = 1))

```

c5


```

# where is Gangnam 3
map.data <- map.data %>%
  mutate(gangnam_group = if_else(
    district %in% c("강남구", "서초구", "송파구"),
    "Gangnam 3", "Others"))

gangnam3_label <- map.data %>%
  filter(district %in% c("강남구", "서초구", "송파구")) %>%
  mutate(label=recode(district,
    "강남구" = "Gangnam",
    "서초구" = "Seocho",
    "송파구" = "Songpa"))

c11 <- ggplot(map.data)+
  geom_sf(aes(fill = gangnam_group,
    color = "white",
    linewidth = 0.4))+
  with_shadow(geom_sf(data=map.data %>% filter(gangnam_group=="Gangnam 3"),
    fill="firebrick",
    color="white",
    linewidth=0.7),
    x_offset=2, y_offset=2,
    sigma=3, color="grey35")+
  geom_sf_text(data=gangnam3_label,
    aes(label=label),
    color="white",
    size=3.5,
    fontface="bold")+
  scale_fill_manual(values=c("Gangnam 3"="firebrick",
    "Others"="grey85")) +
  theme_void() +
  theme(legend.position = "none",
    plot.title = element_text(size = 18, face = "bold"),
    plot.subtitle = element_text(size = 13))
c11

# dataset
pop <- read.csv("C:/Users/kibin/Desktop/Spring 2026/PPPA 6021 Data Visualization/policy
brief/registeredpopulation.csv")
names(pop)
head(pop)

pop_clean <- pop %>%
  slice(-(1:3)) %>%
  transmute(district=동별.2.,
    population=as.numeric(X2024.1)+as.numeric(X2024.2))

pop_clean <- pop_clean %>%
  mutate(district=recode(district,
    "종로구"="Jongno-gu", "중구"="Jung-gu", "용산구"="Yongsan-gu", "성동구

```

```

"="Seongdong-gu",
      "광진구"="Gwangjin-gu", "동대문구"="Dongdaemun-gu", "중랑구"="Jungnang-gu", "성
북구"="Seongbuk-gu",
      "강북구"="Gangbuk-gu", "도봉구"="Dobong-gu", "노원구"="Nowon-gu", "은평구
"="Eunpyeong-gu",
      "서대문구"="Seodaemun-gu", "마포구"="Mapo-gu", "양천구"="Yangcheon-gu", "강서구
"="Gangseo-gu",
      "구로구"="Guro-gu", "금천구"="Geumcheon-gu", "영등포구"="Yeongdeungpo-gu", "동
작구"="Dongjak-gu",
      "관악구"="Gwanak-gu", "서초구"="Seocho-gu", "강남구"="Gangnam-gu", "송파구
"="Songpa-gu", "강동구"="Gangdong-gu"))
head(pop_clean)
str(pop_clean)

# public facilities data
fac <- read.csv("C:/Users/kibin/Desktop/Spring 2026/PPPA 6021 Data Visualization/policy
brief/publicsportsfacility.csv")
names(fac)
head(fac)

fac_clean <- fac %>%
  slice(-(1:3)) %>%
  transmute(district=자치구별.2.,
            numfac=as.numeric(X2024))

fac_clean <- fac_clean %>%
  mutate(district=recode(district,
      "종로구"="Jongno-gu", "중구"="Jung-gu", "용산구"="Yongsan-gu", "성동구
"="Seongdong-gu",
      "광진구"="Gwangjin-gu", "동대문구"="Dongdaemun-gu", "중랑구"="Jungnang-gu", "성
북구"="Seongbuk-gu",
      "강북구"="Gangbuk-gu", "도봉구"="Dobong-gu", "노원구"="Nowon-gu", "은평구
"="Eunpyeong-gu",
      "서대문구"="Seodaemun-gu", "마포구"="Mapo-gu", "양천구"="Yangcheon-gu", "강서구
"="Gangseo-gu",
      "구로구"="Guro-gu", "금천구"="Geumcheon-gu", "영등포구"="Yeongdeungpo-gu", "동
작구"="Dongjak-gu",
      "관악구"="Gwanak-gu", "서초구"="Seocho-gu", "강남구"="Gangnam-gu", "송파구
"="Songpa-gu", "강동구"="Gangdong-gu"))

names(fac_clean)
head(fac_clean)

# public facilities per 10,000 residents
popfac <- merge(pop_clean, fac_clean, by="district")

```

```

popfac$facperc <- (popfac$numfac/popfac$population)*10000
head(popfac)

# public facilities per 10,000 residents
popfac <- popfac %>%
  mutate(group=ifelse(district %in% c("Gangnam-gu", "Seocho-gu", "Songpa-gu"),
    "Gangnam 3", "Others"))

c6 <- ggplot()+
  geom_point(data=popfac,
    mapping=aes(x=facperc, y=reorder(district,facperc),
      color=group), size=2)+
  geom_segment(data=popfac,
    mapping=aes(x=0,xend=facperc,
      y=district, yend=reorder(district,facperc),
      color=group),
    linewidth=0.7)+
  scale_x_continuous(labels=comma,
    expand=expansion(mult=c(0,0.12)))+
  scale_color_manual(values = c("Gangnam 3" = "firebrick", "Others" = "grey55"))+
  labs(title="Public Facilities Show Modest Variation but Meaningful Gaps",
    subtitle="Public recreational facilities per 10,000 residents, by district (2024)",
    x="Public Recreational Facilities per 10,000 Residents", y="",
    caption="Source: Seoul Open Data Portal")+
  theme(legend.position="none",
    panel.background=element_rect(fill="white",color=NA),
    plot.background=element_rect(fill="white",color=NA),
    panel.grid.major.y=element_blank(),
    panel.grid.major.x=element_blank(),
    panel.grid.minor=element_blank(),
    plot.title=element_text(face="bold"),
    axis.title.x=element_text(size=10,margin=margin(t=10)))+
  geom_text(data=subset(popfac, group == "Gangnam 3"),
    aes(x=facperc,
      y=reorder(district, facperc),
      label=round(facperc, 1),
      color=group),
    hjust=-0.5,
    size=3.6,
    fontface="bold")+
  geom_text(data=subset(popfac, district=="Eunpyeong-gu"),
    aes(x=facperc,
      y=reorder(district,facperc),
      label=round(facperc,1),
      color="grey55"),
    hjust=-0.5,
    size=3.6,
    fontface="bold")

c6

# private facilities
prifac <- read.csv("C:/Users/kibin/Desktop/Spring 2026/PPPA 6021 Data Visualization/policy
brief/registeredprivatefacility.csv")

```

```

names(prifac)
head(prifac)
str(prifac)

prifac_clean <- prifac %>%
  slice(-(1:2)) %>%
  transmute(district=자치구별.2.,
            numfac=as.numeric(X2024))
head(prifac_clean)

prifac_clean <- prifac_clean %>%
  mutate(district=recode(district,"종로구"="Jongno-gu","중구"="Jung-gu","용산구"="Yongsan-gu","성동구"="Seongdong-gu",
                        "광진구"="Gwangjin-gu","동대문구"="Dongdaemun-gu","중랑구"="Jungnang-gu","성북구"="Seongbuk-gu",
                        "강북구"="Gangbuk-gu","도봉구"="Dobong-gu","노원구"="Nowon-gu","은평구"="Eunpyeong-gu",
                        "서대문구"="Seodaemun-gu","마포구"="Mapo-gu","양천구"="Yangcheon-gu","강서구"="Gangseo-gu",
                        "구로구"="Guro-gu","금천구"="Geumcheon-gu","영등포구"="Yeongdeungpo-gu","동작구"="Dongjak-gu",
                        "관악구"="Gwanak-gu","서초구"="Seocho-gu","강남구"="Gangnam-gu","송파구"="Songpa-gu","강동구"="Gangdong-gu"))
head(prifac_clean)

# private facilities per 10,000 residents
popprifac <- merge(pop_clean, prifac_clean, by="district")
popprifac$facperc <- (popprifac$numfac/popprifac$population)*10000
head(popprifac)

# private facilities per 10,000 residents
popprifac <- popprifac %>%
  mutate(group=ifelse(district %in% c("Gangnam-gu","Seocho-gu","Songpa-gu"),
                      "Gangnam 3","Others"))

c8 <- ggplot()+
  geom_point(data=popprifac,
            mapping=aes(x=facperc, y=reorder(district,facperc),
                      color=group),
            size=2)+
  geom_segment(data=popprifac,
            mapping=aes(x=0,xend=facperc,
                      y=district, yend=reorder(district,facperc),
                      color=group),
            linewidth=0.7)+
  scale_x_continuous(labels=comma,
                    limits=c(0,26), breaks=seq(0,26,by = 5),

```

```

      expand=expansion(mult=c(0,0.08)))+
scale_color_manual(values = c("Gangnam 3" = "firebrick", "Others" = "grey55"))+
labs(title="Private Facilities Are Highly Uneven",
      subtitle="Private recreational facilities per 10,000 residents, by district (2024)",
      x="Private Recreational Facilities per 10,000 Residents", y="",
      caption="Source: Seoul Open Data Portal")+
theme(legend.position="none",
      panel.background=element_rect(fill="white",color=NA),
      plot.background=element_rect(fill="white",color=NA),
      panel.grid.major.y=element_blank(),
      panel.grid.major.x=element_blank(),
      panel.grid.minor=element_blank(),
      plot.title=element_text(face="bold"),
      axis.title.x=element_text(size=10,margin=margin(t=10)))+
geom_text(data=subset(popprifac, group == "Gangnam 3"),
          aes(x=facperc,
              y=reorder(district, facperc),
              label=sprintf("%.1f",facperc),
              color=group),
          hjust=-0.5,
          size=3.6,
          fontface="bold")+
geom_text(data=subset(popprifac, district=="Eunpyeong-gu"),
          aes(x=facperc,
              y=reorder(district,facperc),
              label=sprintf("%.1f",facperc)),
          color="grey55",
          hjust=-0.5,
          size=3.6,
          fontface="bold")
c8

```

```

# scatter plot (figure 9)
merged_sp3 <- merge(prop.tax, popprifac, by="district")
head(merged_sp3)
merged_sp3 <-merged_sp3 %>%
  mutate(group=group.x) %>%
  select(-group.x, -group.y)
merged_sp3$log_tax <- log(merged_sp3$property_tax_billion)
str(merged_sp3)

c9 <-ggplot()+
  geom_point(data=merged_sp3,
             mapping=aes(x=property_tax_billion,y=facperc,fill=group),
             size=3, shape=21, color="black", stroke=0.5)+
  geom_text(data=subset(merged_sp3,group=="Gangnam 3"),
            mapping=aes(x=property_tax_billion,y=facperc,label=district),
            color="firebrick",
            size=3.8,
            hjust=1.2,
            vjust=-0.2)+
  scale_x_log10(labels=comma, expand=expansion(mult=c(0.03,0.08)),
               breaks=c(75,100,200,400))+

```

```

scale_y_continuous(labels=comma,
                    breaks=seq(0,26,by=5),
                    limits=c(0,26))+
scale_fill_manual(values=c("Gangnam 3" = "firebrick","Others" = "gray70"))+
theme_minimal()+
theme(plot.title=element_text(face="bold",
                               size=16, lineheight=1, margin=margin(b=5)),
      plot.subtitle=element_text(margin=margin(b=10)),
      panel.grid.major.x=element_blank(),
      panel.grid.minor=element_blank(),
      legend.position="none",
      axis.title.x=element_text(margin=margin(t=10)),
      axis.title.y=element_text(margin=margin(r=10)))+
labs(title="Private Facilities Increase with Local Wealth",
     subtitle="Private facilities per 10,000 residents and property tax revenue, by district (2024)",
     x="Property Tax Revenue (billion KRW)",
     y="Private Facilities per 10,000 Residents",
     caption="Note: Property tax revenue is shown on a log scale.
Source: Seoul Open Data Portal")
c9

```

```

# for dumbbell chart
fac_all <-merge(popfac[,c("district","facperc")],
               popprifac[,c("district","facperc")],
               by="district",
               suffixes=c("_public","_private"))

head(fac_all)
names(fac_all) <- c("district","public","private")
head(fac_all)
top_y <-length(unique(fac_all$district))

# plot a dumbbell chart
c10 <- ggplot(data=fac_all)+
  geom_segment(aes(x=public, xend=private,
                  y=district, yend=district),
              color="grey70")+
  geom_segment(data=subset(fac_all, district %in% c("Gangnam-gu","Seocho-gu","Songpa-gu")),
              aes(x=public, xend=private,
                  y=district, yend=district),
              color="grey40",linewidth=1.2)+
  geom_point(mapping=aes(x=public,
                        y=reorder(district,private-public)),
             color="#9ecae1", size=2.5)+
  geom_point(mapping=aes(x=private,
                        y=reorder(district,private-public)),
             color="#08519c", size=2.5)+
  labs(title="Private Facilities Exhibit Larger Absolute Disparities Across Districts",
       subtitle="Comparison of public and private facilities per 10,000 residents (2024)",
       x="Facilities per 10,000 residents", y="",
       caption="Source: Seoul Open Data Portal")+
  scale_y_discrete(expand = expansion(mult = c(0.02, 0.08)))+
  theme_minimal()+
  theme(panel.grid.major.x=element_line(color="grey85"),

```

```

    panel.grid.minor=element_blank(),
    legend.position="none",
    plot.title=element_text(face="bold"),
    axis.ticks.x=element_blank())+
  annotate("text", x=3.0, y=26.2,
    label="Public", color="#9ecae1",
    fontface="bold", size=3, hjust=0)+
  annotate("text", x=24.7, y=26.2,
    label="Private", color="#08519c",
    fontface="bold", size=3, hjust=1)
c10

# use quartile
fis.ind2 <- fis.ind2 %>%
  mutate(district=recode(district,
    "종로구"="Jongno-gu", "중구"="Jung-gu", "용산구"="Yongsan-gu", "성동구"
    "Seongdong-gu",
    "광진구"="Gwangjin-gu", "동대문구"="Dongdaemun-gu", "중랑구"="Jungnang-gu", "성
    북구"="Seongbuk-gu",
    "강북구"="Gangbuk-gu", "도봉구"="Dobong-gu", "노원구"="Nowon-gu", "은평구"
    "Eunpyeong-gu",
    "서대문구"="Seodaemun-gu", "마포구"="Mapo-gu", "양천구"="Yangcheon-gu", "강서구"
    "Gangseo-gu",
    "구로구"="Guro-gu", "금천구"="Geumcheon-gu", "영등포구"="Yeongdeungpo-gu", "동
    작구"="Dongjak-gu",
    "관악구"="Gwanak-gu", "서초구"="Seocho-gu", "강남구"="Gangnam-gu", "송파구"
    "Songpa-gu", "강동구"="Gangdong-gu"))
setdiff(popfac$district, fis.ind2$district)
setdiff(fis.ind2$district, popfac$district)

merged_public_fiscal <- merge(fis.ind2, popfac, by="district")
head(merged_public_fiscal)

fi_public_quartile <- merged_public_fiscal %>%
  mutate(fiscal_q=ntile(fiscal_index, 4),
    fiscal_q=factor(fiscal_q, levels = 1:4,
    labels=c("Q1", "Q2", "Q3", "Q4")))%>%
  group_by(fiscal_q) %>%
  summarize(avg_public=mean(facperc, na.rm = TRUE),
    n = n(), .groups = "drop")

c11 <- ggplot(fi_public_quartile, aes(x=fiscal_q, y=avg_public)) +
  geom_col(fill="#6baed6", width=0.65) +
  geom_text(aes(label=round(avg_public, 1)), vjust=-0.4, size=3.5, fontface="bold") +
  labs(title="Public Facilities Are Less Tied to Fiscal Capacity",
    subtitle="Average public facilities per 10,000 residents by fiscal independence quartile (2024)",
    x="Fiscal independence quartile (Q1=lowest)",
    y="Average public facilities per 10,000 residents",
    caption="Source: Seoul Open Data Portal")+

```

```

theme_minimal() +
theme(panel.grid.major.x=element_blank(),
      panel.grid.minor=element_blank(),
      plot.title=element_text(face="bold"),
      plot.subtitle=element_text(margin=margin(b = 10)),
      axis.title.x=element_text(margin=margin(t = 10)),
      axis.title.y=element_text(margin=margin(r = 10)))
c11

# policy recommendation
scenario <- tribble(
  ~time,      ~fiscal_capacity, ~public, ~private,
  "Current", "High fiscal capacity", 3.5, 14,
  "Current", "Low fiscal capacity",  3.5, 6.5,
  "To-be",   "High fiscal capacity", 3.5, 14,
  "To-be",   "Low fiscal capacity",  8.5, 8)

scenario_long <- scenario %>%
  mutate(bar_label=factor(fiscal_capacity,
                          levels=c("High fiscal capacity", "Low fiscal capacity")),
         x_group=case_when(
           time == "Current" & fiscal_capacity == "High fiscal capacity" ~ 1,
           time == "Current" & fiscal_capacity == "Low fiscal capacity" ~ 2,
           time == "To-be"    & fiscal_capacity == "High fiscal capacity" ~ 3,
           time == "To-be"    & fiscal_capacity == "Low fiscal capacity" ~ 4)) %>%
  pivot_longer(cols=c(public, private), names_to="type", values_to="value") %>%
  mutate(type=str_to_title(type),
         type=factor(type, levels=c("Public", "Private")))

c12 <- ggplot(data=scenario_long, aes(x=x_group, y=value, fill=type)) +
  geom_col(width=0.65) +
  annotate("text", x=1.5, y=21, label="Current", size=5.5, fontface="bold") +
  annotate("text", x=3.5, y=21, label="To-be", size=5.5, fontface="bold") +
  scale_x_continuous(breaks=c(1, 2, 3, 4),
                    labels=c("High\ncapacity", "Low\ncapacity",
                              "High\ncapacity", "Low\ncapacity")) +
  scale_y_continuous(limits=c(0, 22), expand=expansion(mult=c(0, 0.03)), breaks=NULL) +
  scale_fill_manual(values=c("Public"="#A6CEE3", "Private"="#155A9C")) +
  labs(
    title="Reducing Access Gaps Through Public and Private Provision",
    subtitle="Stylized policy scenario illustrating targeted public expansion and private investment",
    x=NULL, y="Facilities per 10,000 residents", fill = NULL) +
  theme_minimal(base_size=14) +
  theme(panel.grid.major.x=element_blank(),
        panel.grid.minor=element_blank(),
        axis.text.x=element_text(size=12),
        axis.title.y=element_text(size=14),
        plot.title=element_text(face="bold", size=16),
        plot.subtitle=element_text(size=12),
        legend.position="top",
        plot.caption=element_text(hjust=1, size=10),
        plot.margin=margin(10, 15, 10, 10))
c12

```