예방의학의 관점에서 바라본 정신건강 관련 예산 추이 분석

이 사회는 과연 우리를 안아줄 준비가 되어있는가.

ChatGPT 4, 김홍표

기록 정신 취업 연세광화문정신과
병원 전문가 헬스 헬스조선
진료안내 입시 보험
불이익 진료 정리
가입 보험가입 의학
연세 최신정리
정신과
최신 건강 조선
광화문 가면 정신건강 질문 치료

치료 심리우울증 오션 세종
심리상담 상담
코로나블루
부산 골든블루 운영확대 제공
생활
블루 블랙 마음 방법 증가
치유 환자 방역 원
서비스 골든 시대 역

건강 방법 기자 정신 상상
사회적 생활 정신건강 택트기
19 거리 우울증 디지털
대구 블루 집중
병 극복 여성
기술
신조어상상코로나

Web crawling

Naver

```python
# 네이버 뉴스 크롤링: 코로나 블루, 코로나 우울
def make_naver_news_search_url(query, start_date, end_date, page):
    base_url = "https://search.naver.com/search.naver"
    params = {
        "where": "news",
        "query": query,
        "sm": "tab_opt",
        "sort": "0",
        "photo": "0",
        "field": "0",
        "pd": "3",
        "ds": start_date,
        "de": end_date,
        "start": (page - 1) * 10 + 1,
    }
    return requests.get(base_url, params=params).url


        titles.append(row["title"])
    return titles

def tokenize_titles(titles):
    kkma = Kkma()  # Kkma 형태소 분석기를 사용
    tokens = []
    for title in titles:
        tokens.extend(kkma.nouns(title))  # Kkma의 nouns 메서드를 사용해 명사를 추출
    return tokens

def create_wordcloud(tokens, top_n):
    counter = Counter(tokens)
    most_common = counter.most_common(top_n)
    wc = WordCloud(font_path='/Library/Fonts/서울한강 장체L.otf', background_color='white', width=800, height=800)
    wc.generate_from_frequencies(dict(most_common))

    plt.figure(figsize=(10, 10))
    plt.imshow(wc, interpolation='bilinear')
    plt.axis('off')
    plt.savefig('wordcloud.png')
    plt.show()

def count_related_words(tokens, related_words):
    count = 0
    for token in tokens:
        if token in related_words:
            count += 1
    return count

def main():
```

```python
def crawl_naver_news(query, start_date, end_date, max_count=1000):
    page = 1
    news_list = []

    with tqdm(desc="Crawling news", unit="page", dynamic_ncols=True, position=0) as progress_bar:
        while len(news_list) < max_count:
            url = make_naver_news_search_url(query, start_date, end_date, page)
            response = requests.get(url)
            soup = BeautifulSoup(response.text, 'html.parser')

            news_items = soup.select(".news_wrap.api_ani_send")

            if not news_items:
                break

            for item in news_items:
                title = item.select_one(".news_tit").text
                link = item.select_one(".news_tit")['href']
                press = item.select_one(".info.press").text

                news_list.append({
                    "title": title,
                    "link": link,
                    "press": press,
                })

                if len(news_list) >= max_count:
                    break

            page += 1
            progress_bar.update(1)
            time.sleep(0.45)

    return news_list[:max_count]

def read_titles_from_csv(filename):
    titles = []
    with open(filename, "r", encoding="utf-8") as csvfile:
        reader = csv.DictReader(csvfile)
        for row in reader:
```

Web crawling
Google

```python
query_list = ['정신과 기록', '정신과 취업', '정신과 불이익']
baseUrl = 'https://www.google.com/search?q='
base = 'https://www.google.com/'

crawled = []

for query in query_list:
    quoteUrl = quote(query)
    url = baseUrl + quoteUrl

    headers = {'User-Agent': 'Mozilla/5.0'}
    req = Request(url, headers=headers)
    webpage = urlopen(req)
    source = webpage.read()
    webpage.close()

    driver = webdriver.Safari()
    driver.get(url)
    driver.implicitly_wait(7)

    for i in range(1, 41):
        try:
            html = driver.page_source
            soup = BeautifulSoup(html, 'html.parser')
            v = soup.select('.yuRUbf')

            for j in v:
                print(j.select_one('.LC20lb.DKV0Md').text)
                print()
                crawled.append(j.select_one('.LC20lb.DKV0Md').text)

            next_url = base + soup.select_one(f'a[aria-label="Page {i+1}"]').get('href')
            driver.get(next_url)
            driver.implicitly_wait(3)

        except AttributeError:
            break

    driver.close()

print(crawled)
```

```python
## WordCloud
from konlpy.tag import Kkma
from wordcloud import WordCloud

kkma = Kkma()

### 1. 단어 추출
nouns_extend = []

for sent in crawled:
    nouns_extend.extend(kkma.nouns(sent))
print()
print(nouns_extend)


### 2. 단어 전처리
from re import match # 숫자 제외

nouns_count = {} # 단어 카운터

for noun in nouns_extend:
    if len(noun) > 1 and not(match('^[0-9]', noun)):
        nouns_count[noun] = nouns_count.get(noun, 0) + 1

print(nouns_count)


### 3. TopN Selecting
from collections import Counter # class

word_count = Counter(nouns_count)
top30_word = word_count.most_common(30)


### 4. Generate WordCloud
wc = WordCloud(font_path='/Library/Fonts/서울한강 장체L.otf',
        width=500, height=400,
        max_words=100,max_font_size=150,
        background_color='white')

wc_result = wc.generate_from_frequencies(dict(top30_word))

import matplotlib.pyplot as plt

plt.imshow(wc_result)
plt.axis('off') # 축 눈금 감추기
plt.show()
```
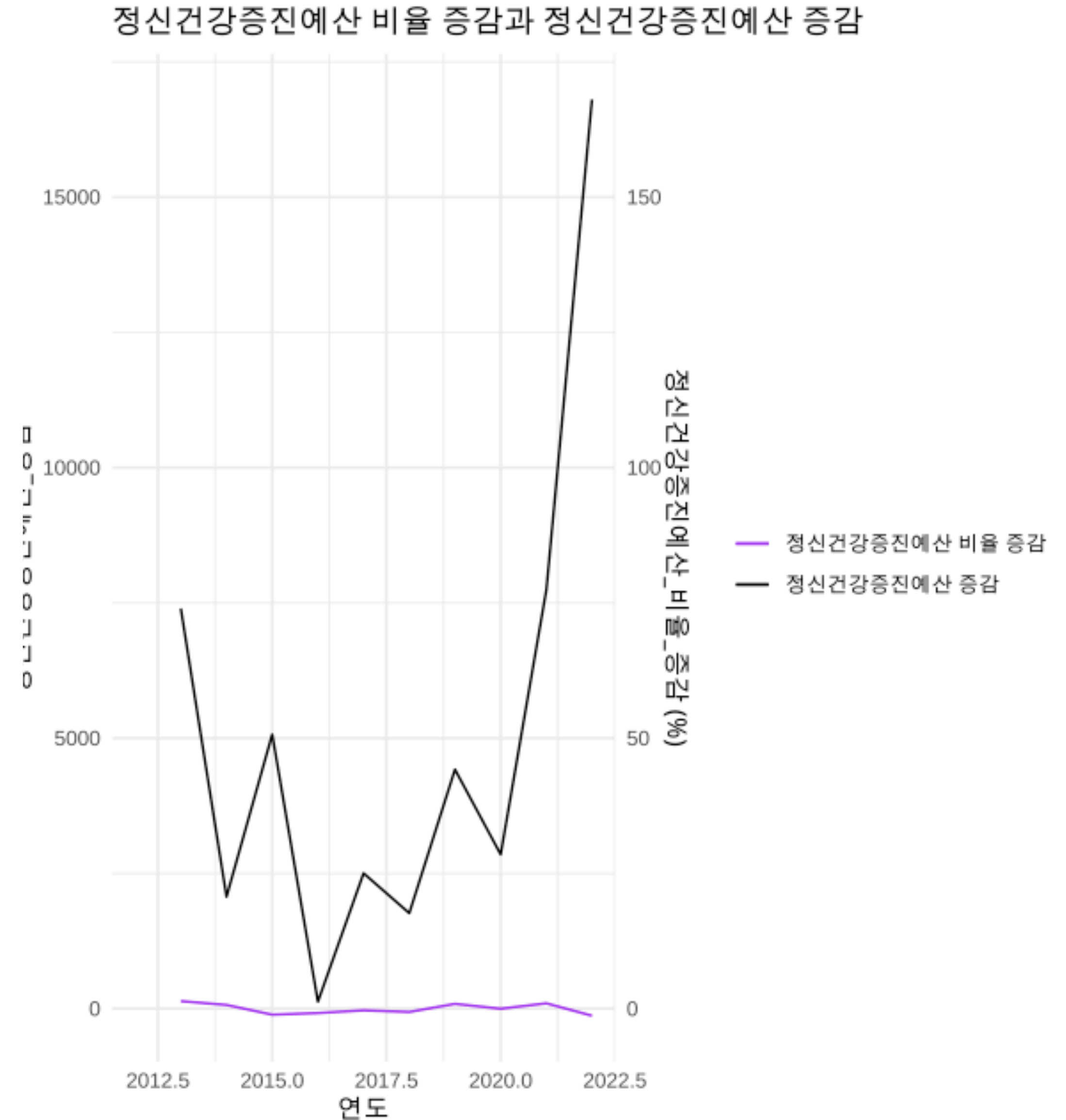
프로젝트 당위성

당신은 왜 혼자서 그렇게 피똥을 싸셨나요?

- 아직까지도 정신과 치료를 주저하게 만드는 심리가 작용.

- 코로나로 인한 정신건강 및 우울증 관심 증대.

- 우울증을 위시한 정신질환은 예방이 무척 중요함.

현실적 예산 확보문제
적정한 예산이 할당되어 있는가?

- 정신건강증진예산은 급증

- 정신건강증진예산 비율은 하락!!

정신건강증진예산 비율 증감과 정신건강증진예산 증감

R Coding
그는 어떻게 GPT를 부렸는가.

```
# 쉼표 제거
cols_to_convert <- c('people', 'mental_budget', 'per.person')
mental_budget[cols_to_convert] <-
lapply(mental_budget[cols_to_convert],
                              function(x) as.numeric(gsub(',', '', x)))
str(mental_budget)
mental_budget


cols_to_convert2 <- c('people', 'mental_cost', 'per.person')
mental_expenditure[cols_to_convert2] <-
lapply(mental_expenditure[cols_to_convert2],
                              function(x) as.numeric(gsub(',', '', x)))
str(mental_expenditure)

# Data frame: mental_balance
mental_balance <- data.frame(region = mental_budget$region,
                  people = mental_budget$people,
                  balance = mental_budget$mental_budget -
mental_expenditure$mental_cost,
                  per_person = mental_budget$per.person -
mental_expenditure$per.person)

mental_balance
```

```
# '서울시' 제거
mental_budget_temp <- mental_budget[mental_budget$region != "서울시", ]
mental_expenditure_temp <-
mental_expenditure[mental_expenditure$region != "서울시", ]
mental_balance_temp <- mental_balance[mental_balance$region !=
"서울시", ]
str(mental_budget_temp)
str(mental_expenditure_temp)
str(mental_balance_temp)

# What should I do with mental_balance
# Arranging by region
mental_budget_temp <- mental_budget_temp %>% arrange(region)
mental_expenditure_temp <- mental_expenditure_temp %>%
arrange(region)
mental_balance_temp <- mental_balance_temp %>% arrange(region)

# Cut 1,000,000
mental_budget_temp2 <- mental_budget_temp
mental_budget_temp2$mental_budget <-
round(mental_budget_temp2$mental_budget / 1000000, 3)
mental_expenditure_temp2 <- mental_expenditure_temp
mental_expenditure_temp2$mental_cost <-
round(mental_expenditure_temp2$mental_cost / 1000000, 3)
mental_balance_temp2 <- mental_balance_temp
mental_balance_temp2$balance <-
round(mental_balance_temp2$balance / 1000000, 3)
```

R Coding
그는 어떻게 GPT를 부렸는가.

```r
# Calculate means and medians
mean_budget <- mean(mental_budget_temp2$per.person)
mean_expenditure <- mean(mental_expenditure_temp2$per.person)
mean_balance <- mean(mental_balance_temp2$per_person)

median_budget <- median(mental_budget_temp2$per.person)
median_expenditure <- median(mental_expenditure_temp2$per.person)
median_balance <- median(mental_balance_temp2$per_person)

# Create data frame
mean_median_2021 <- data.frame(
  value = c(mean_budget, mean_expenditure, mean_balance,
         median_budget, median_expenditure, median_balance),
  row.names = c("mean_budget", "mean_expenditure", "mean_balance",
         "median_budget", "median_expenditure", "median_balance")
)
arrange(mean_median_2021)


# Print data frame
mean_median_2021

# csv 파일로 저장
write.csv(mean_median_2021, file = "/Users/f___yo_/Documents/
SemiProject/csv/2021_mean_median.csv", row.names = T)
```

```r
# Visualization
# 지역명의 공백 제거
mental_budget_temp2$region <-
trimws(mental_budget_temp2$region)
mental_expenditure_temp2$region <-
trimws(mental_expenditure_temp2$region)

# combined_data 생성
combined_data <- inner_join(mental_budget_temp2,
mental_expenditure_temp2, by = "region") %>%
  select(region, mental_budget, mental_cost)
combined_data

# 컬럼명 변경
colnames(combined_data)[colnames(combined_data) ==
"mental_budget.x"] <- "mental_budget"
colnames(combined_data)[colnames(combined_data) ==
"mental_cost.y"] <- "mental_cost"

# 결측치 확인
missing_rows <- is.na(combined_data$mental_budget) |
is.na(combined_data$mental_cost)
combined_data[missing_rows, ]
```
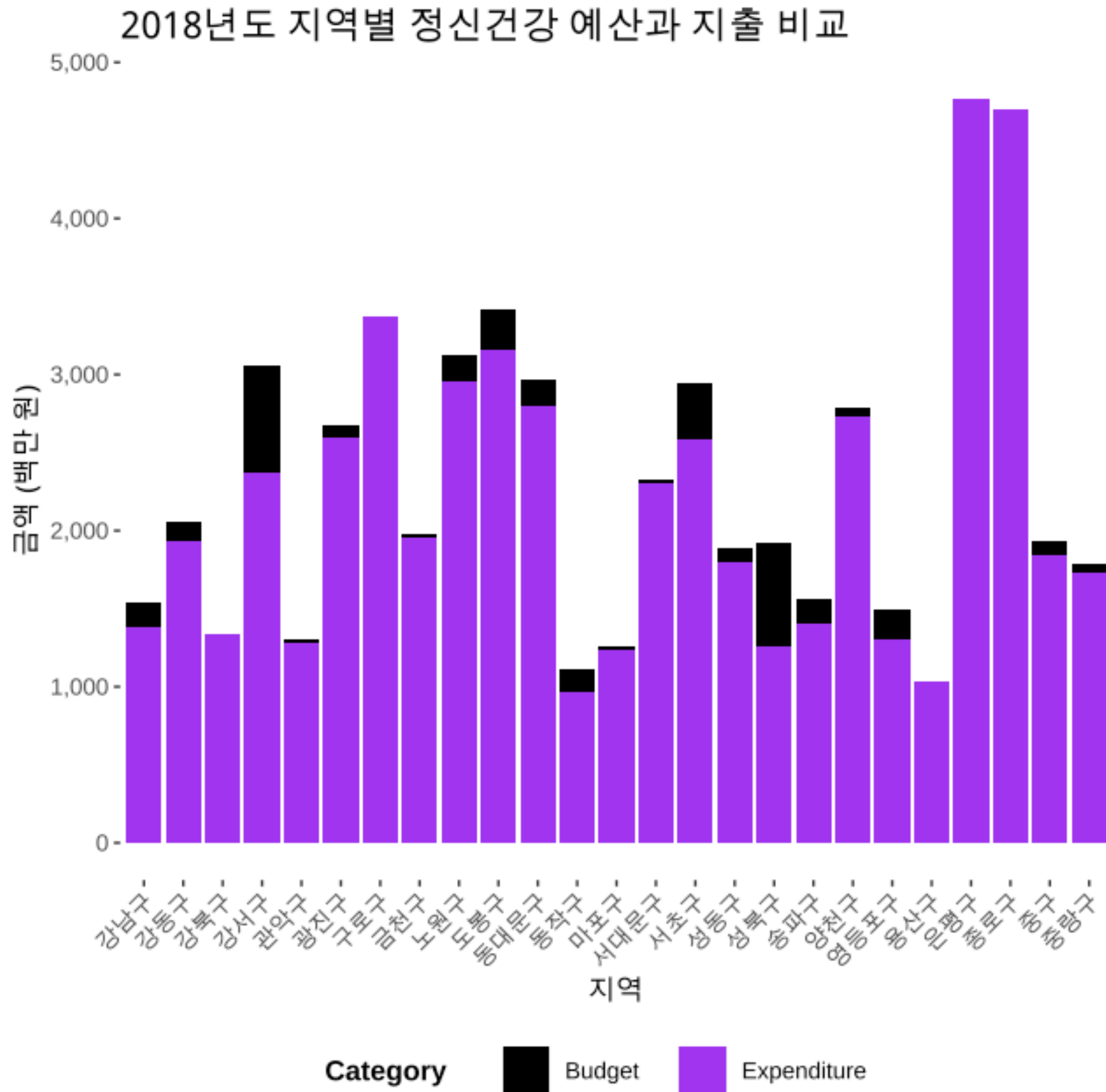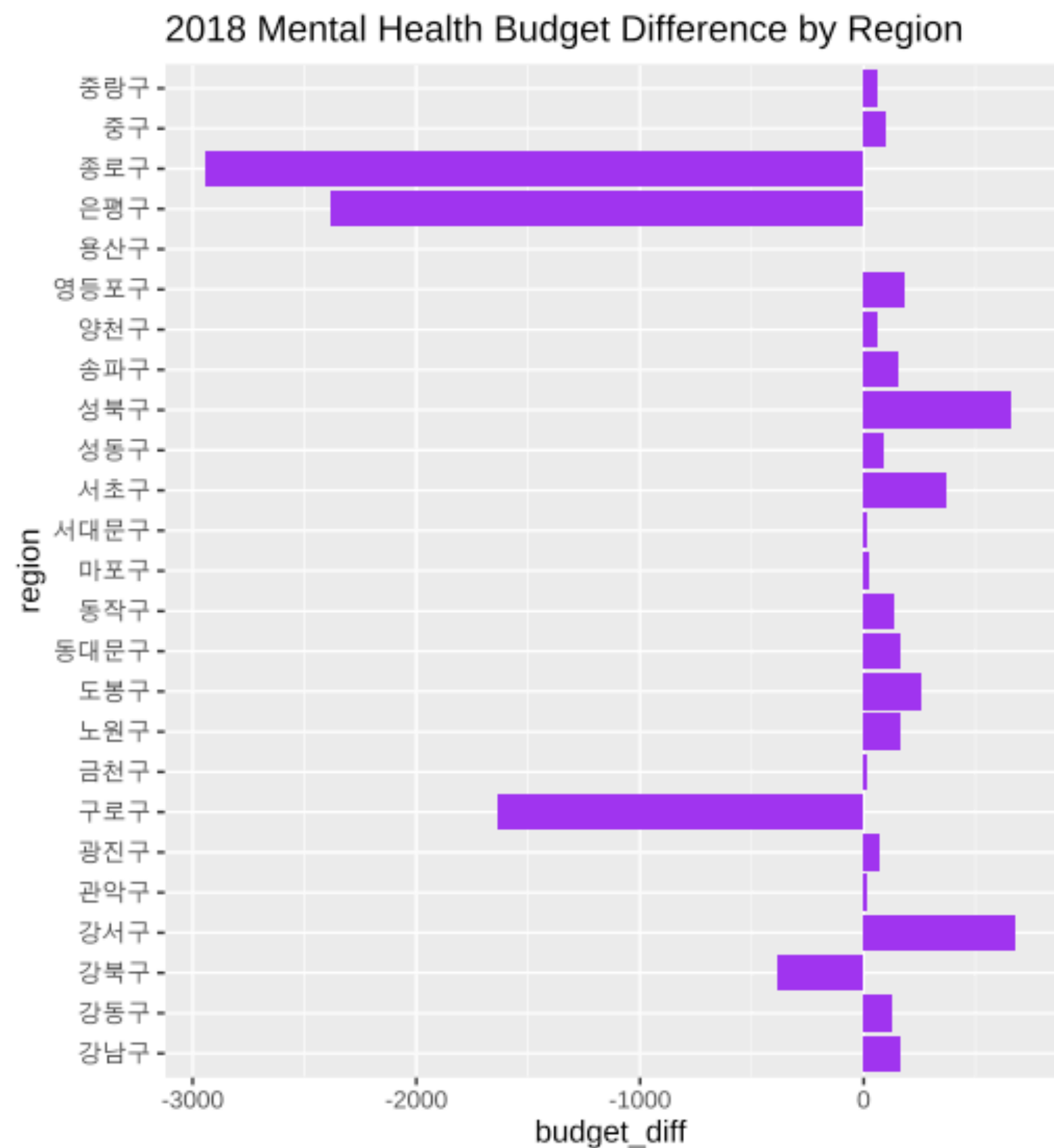
R Coding
그는 어떻게 **GPT**를 부렸는가**.**

```r
# 잔액 표 만들기 & CSV 저장
combined_data$budget_diff <- combined_data$mental_budget -
combined_data$mental_cost
combined_data <- combined_data %>% arrange(budget_diff)
combined_data %>%
  select(region, mental_budget, mental_cost, budget_diff) %>%
  arrange(combined_data$region) %>%
  write.csv(file = "/Users/f___yo_/Documents/SemiProject/csv/
mental_balance_2021.csv", row.names = FALSE)
```

```r
# 막대 그래프
ggplot(data = combined_data, aes(x = region)) +
  geom_bar(aes(y = mental_budget, fill = "Budget"), stat = "identity",
position = "dodge") +
  geom_bar(aes(y = mental_cost, fill = "Expenditure"), stat = "identity",
position = "dodge") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  labs(title = "2021년도 지역별 정신건강 예산과 지출 비교", x = "지역", y =
"금액 (백만 원)") +
    scale_fill_manual(values = c("Budget" = "black", "Expenditure" =
"purple"), name = "Category") +
  theme(plot.title = element_text(size = 16, face = "bold")) +
  theme(axis.title = element_text(size = 12, face = "bold")) +
  theme(axis.text = element_text(size = 10)) +
  theme(legend.title = element_text(size = 12, face = "bold")) +
  theme(legend.text = element_text(size = 10)) +
  theme(panel.background = element_blank()) +
  theme(legend.position = "bottom") +
  scale_y_continuous(labels = scales::comma) +
  labs(fill = "") +
  theme(legend.key.size = unit(0.75, "cm")) +
  theme(plot.margin = margin(1, 1, 1, 1, "cm"))
```
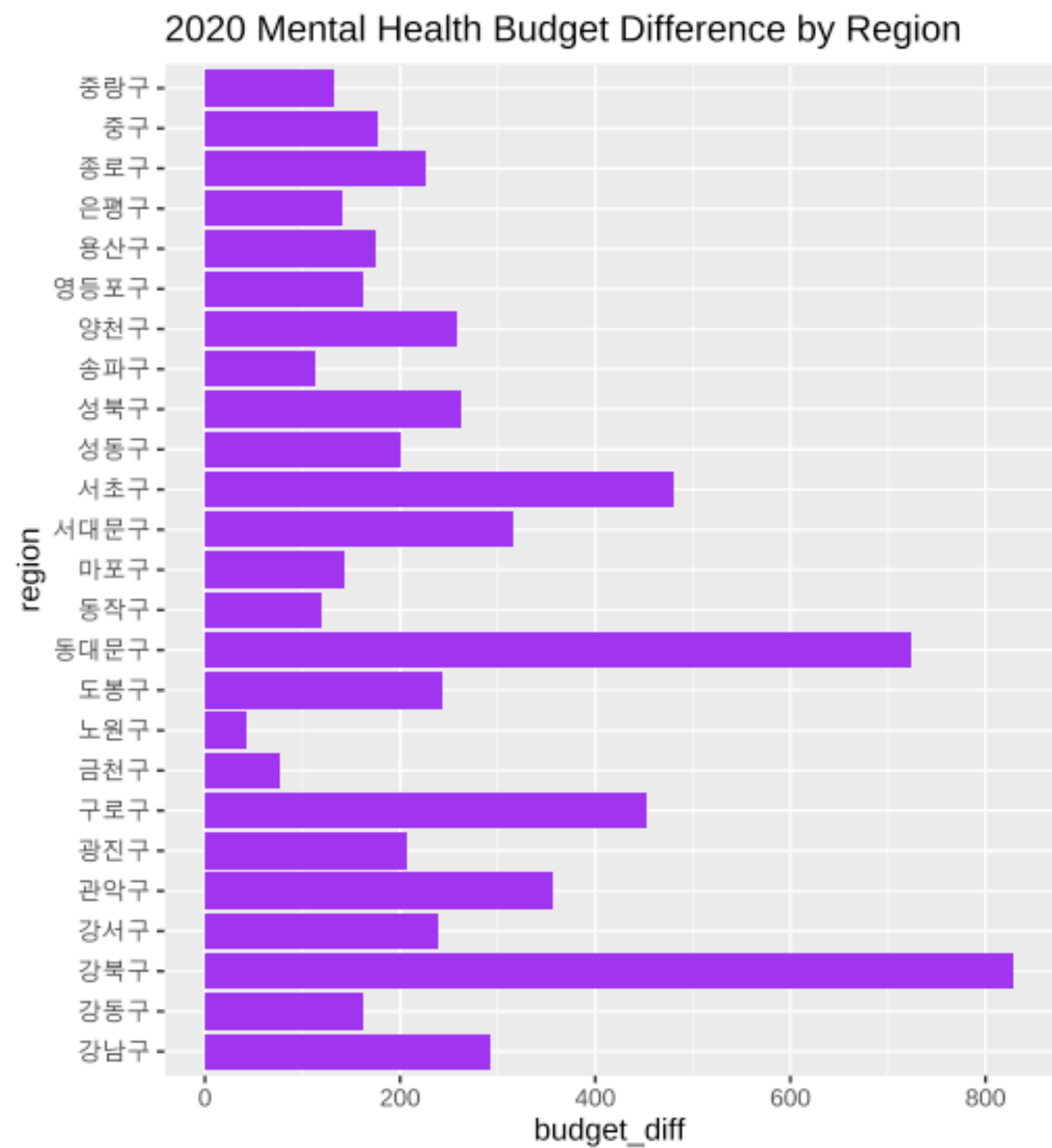
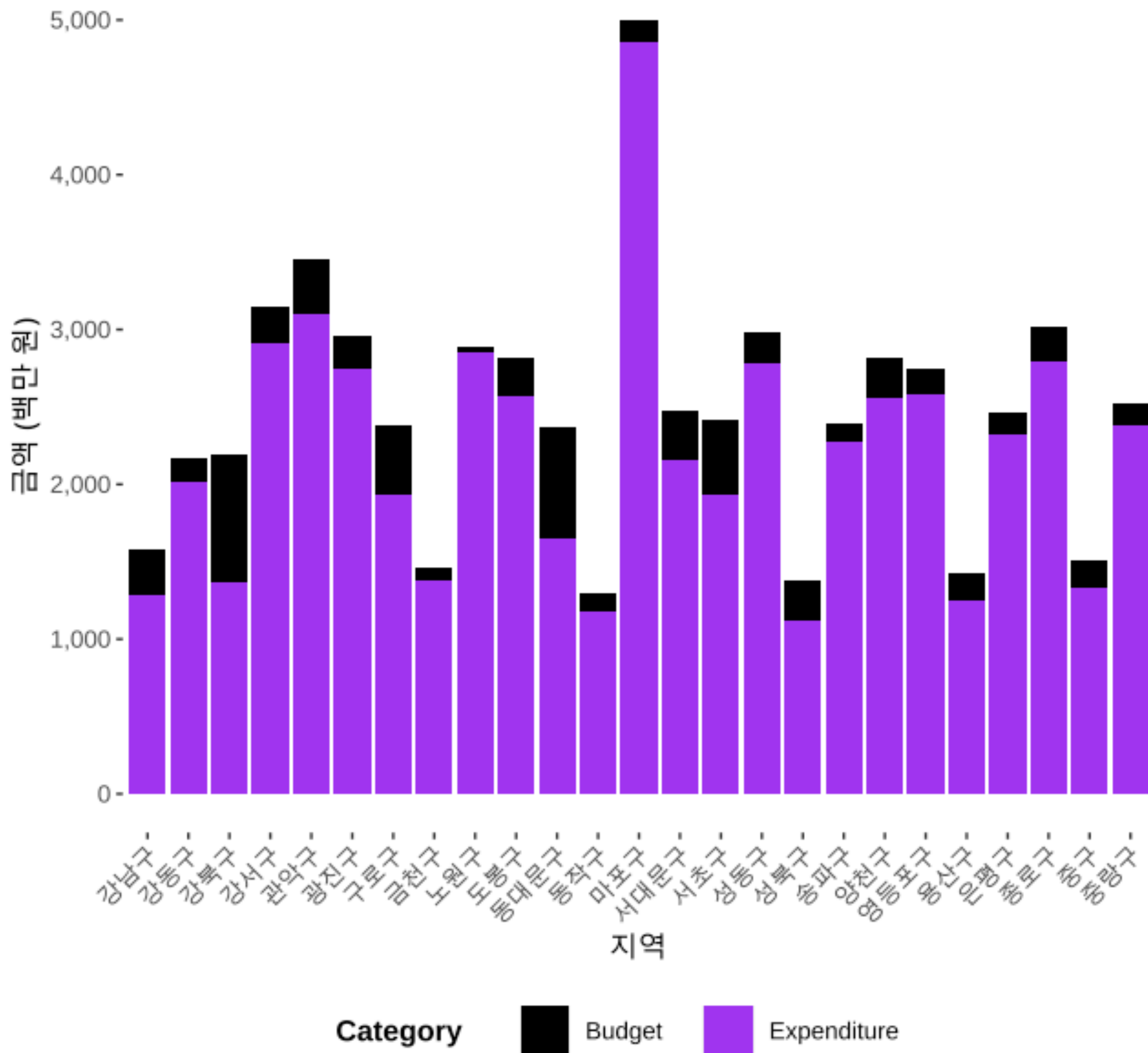# 2018년도 현황

대다수 예산 부족, 몇몇은 심지어 적자
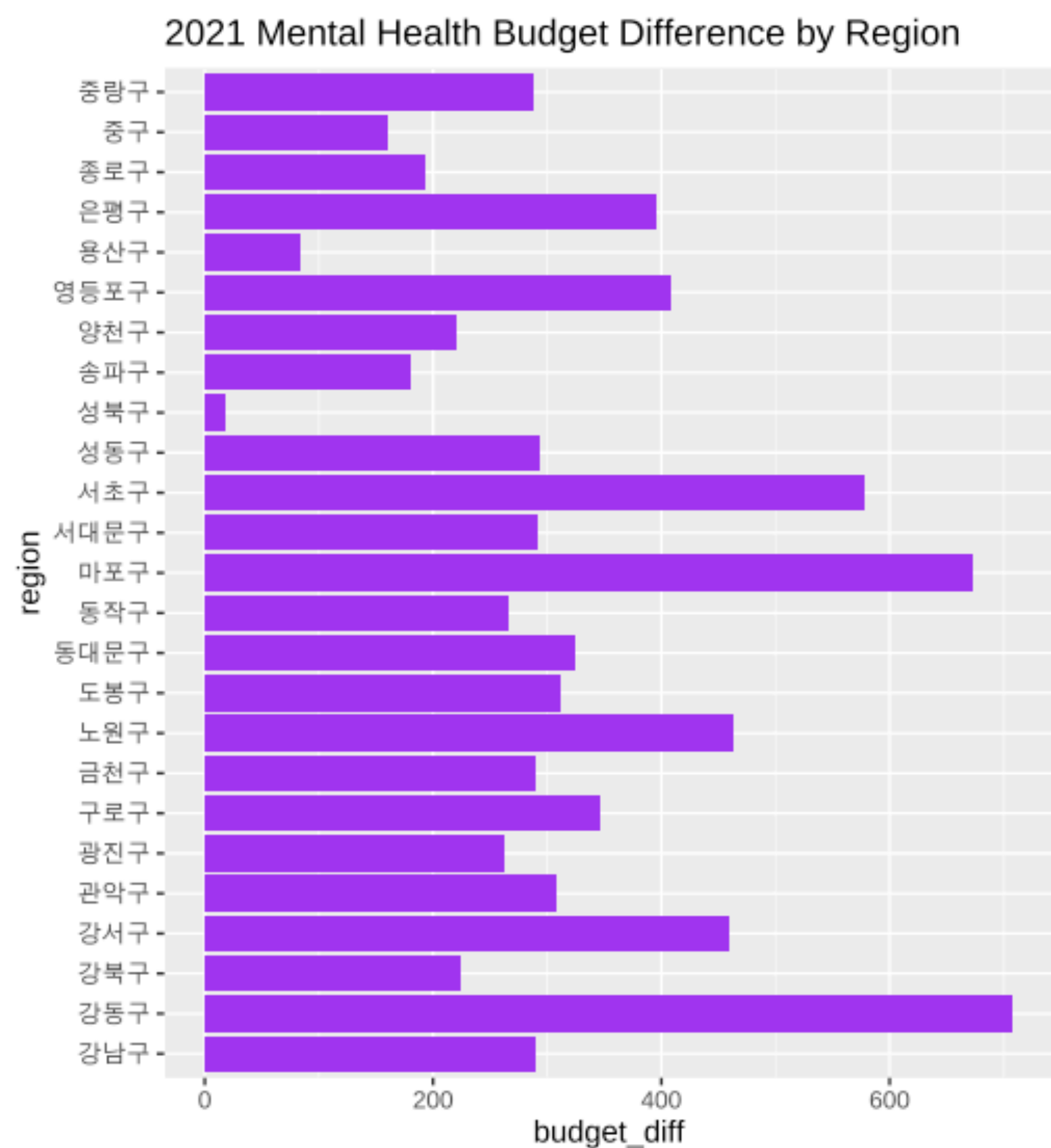
2020년도 현황

금천구, 노원구, 동작구 적자 겨우 면함

2020 Mental Health Budget Difference by Region

2020년도 지역별 정신건강 예산과 지출 비교

2021년도 현황

성북구 적자 겨우 면함

2021 Mental Health Budget Difference by Region
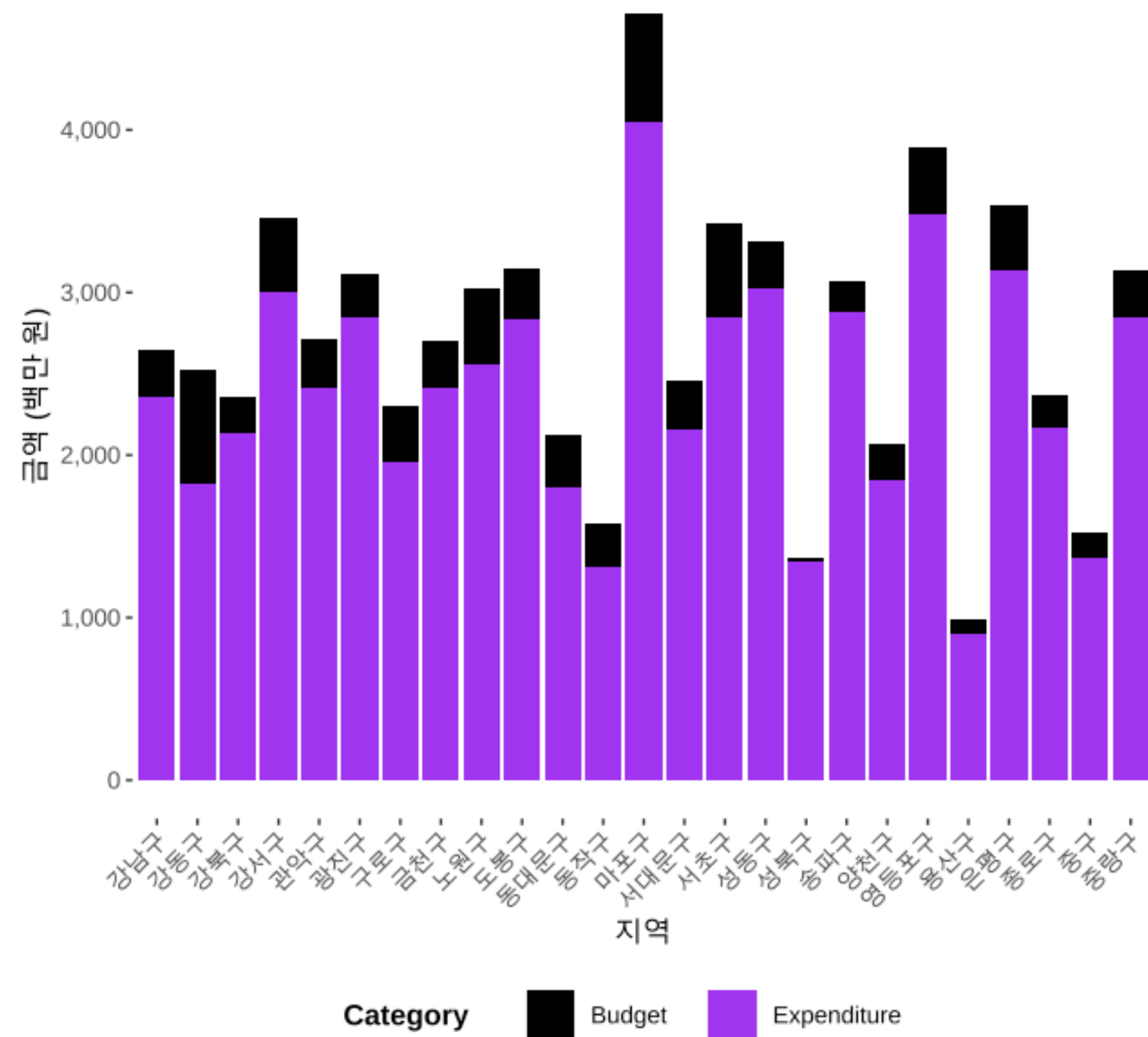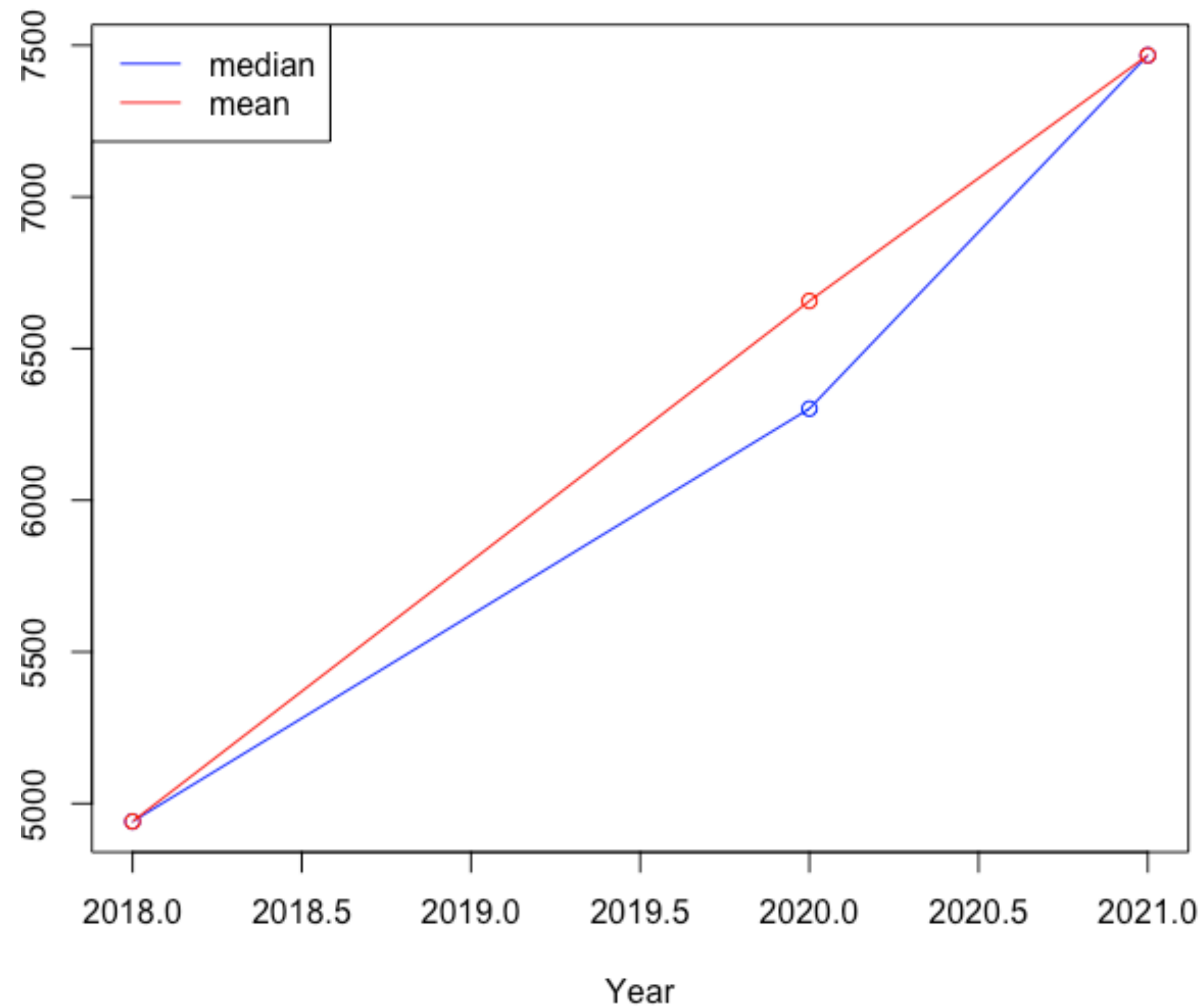

2021년도 지역별 정신건강 예산과 지출 비교

예산 추세
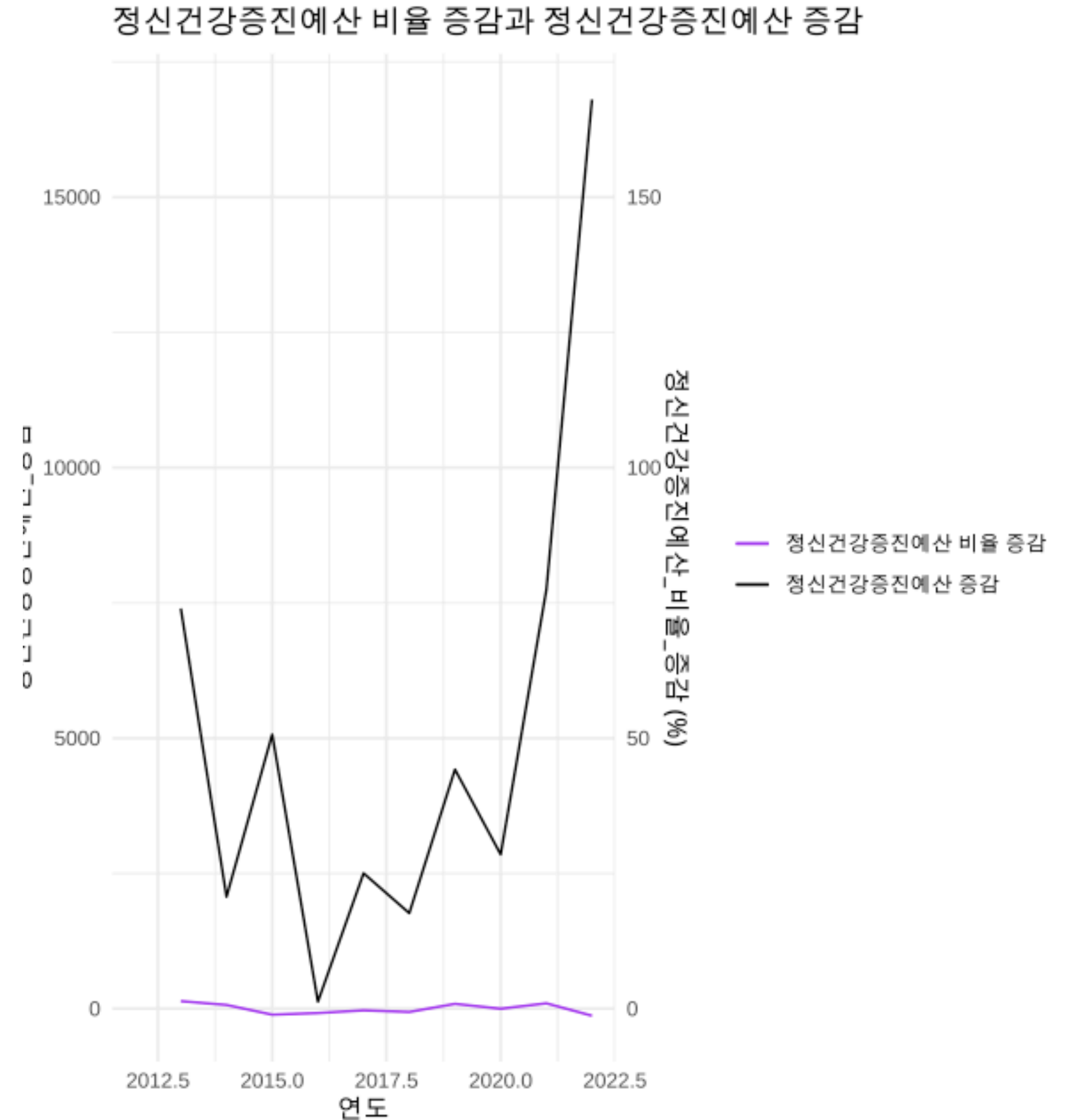
- 예산은 증가추세에 있음.

결론

우리사회는 입으로만 상처를 보듬는다

- 증가하는 예산만큼 지출도 늘어나는 추이

- 이는 아직까지 정신건강증진예산이 부족함을 나타내는 증거

- 다음 그림과 같은 정신건강증진예산 비율을 탈피할 필요

정신건강증진예산 비율 증감과 정신건강증진예산 증감

Library

Python

- import requests

- from bs4 import BeautifulSoup

- import time

- import csv

- from tqdm import tqdm

- from collections import Counter

- from konlpy.tag import Kkma

- from wordcloud import WordCloud

- import matplotlib.pyplot as plt

- from urllib.request import urlopen

- from urllib.request import Request

- from urllib.parse import quote

- from bs4 import BeautifulSoup

- from selenium import webdriver

Library

R

- library(scales)

- library(sysfonts)

- library(showtext)

- library(tidyr)

- library(tidyverse)

- library(readxl)

- library(ggplot2)

- library(scales)

- library(dplyr)

- library(readr)

- library(readxl)

- library(cowplot)

참고문헌 및 출처

논문

- Van Zoonen, K., Buntrock, C., Ebert, D. D., Smit, F., Reynolds, C. F., Beekman, A. T., & Cuijpers, P. (2014). Preventing the onset of major depressive disorder: a meta-analytic review of psychological interventions. International Journal of Epidemiology, 43(2), 318-329.

- Stice, E., Shaw, H., Bohon, C., Marti, C. N., & Rohde, P. (2009). A meta-analytic review of depression prevention programs for children and adolescents: factors that predict magnitude of intervention effects. Journal of Consulting and Clinical Psychology, 77(3), 486-503.

- Muñoz, R. F., Cuijpers, P., Smit, F., Barrera, A. Z., & Leykin, Y. (2010). Prevention of major depression. Annual Review of Clinical Psychology, 6, 181-212.

- Yook, S. P., & Kim, Z. S. (2007). A clinical study on the Korean version of Beck Depression Inventory: comparative study of patient and non-patient. Korean Journal of Clinical Psychology, 26(1), 101-124.

데이처 출처

- 서울시정신건강통계: https://seoulmentalhealth.kr

- 네이버: https://www.naver.com

- 구글: https://www.google.com

의의 및 한계

의의

- 정신건강증진예산 증감의 추이를 비교분석함

- 예산확보의 중요성 역설

- 프롬프트 엔지니어로서의 가능성 발견

한계

- 공공데이터포털의 업무태만으로 인한 2018년~2021년까지의 정신건강의료기관 실진자수 데이터 확보 실패

- 서울시정신통계의 업무태만으로 인한 2019년 데이터 소실

Q&A

감사