An Overview of China's Government Fiscal Behavior

AUTHOR
张颢严

😉

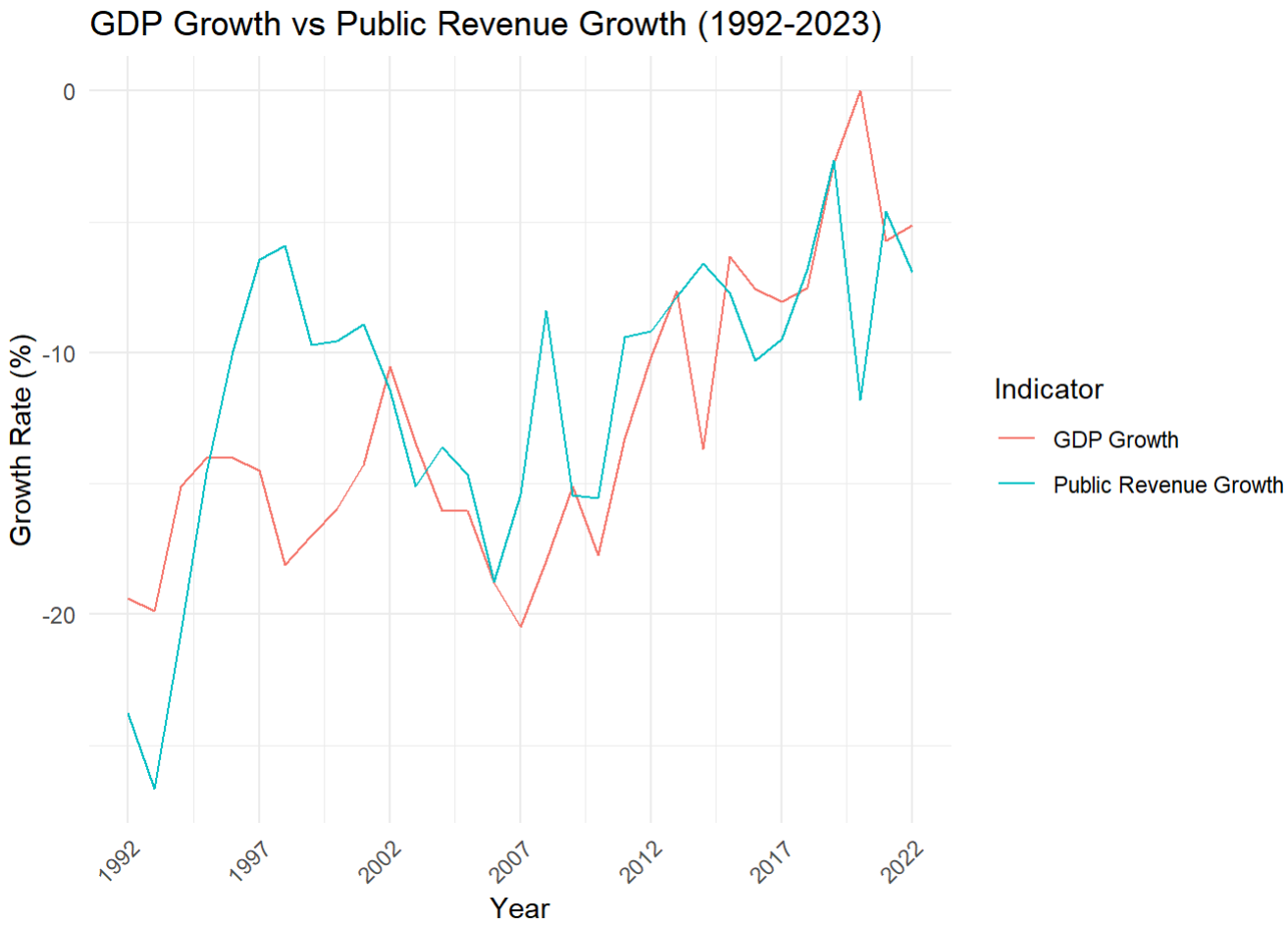
GDP和一般公共预算收入（全国）增速对比

```
# 加载必要的包
library(readxl)
library(dplyr)
```

```
载入程序包：'dplyr'

The following objects are masked from 'package:stats':

    filter, lag

The following objects are masked from 'package:base':

    intersect, setdiff, setequal, union
```

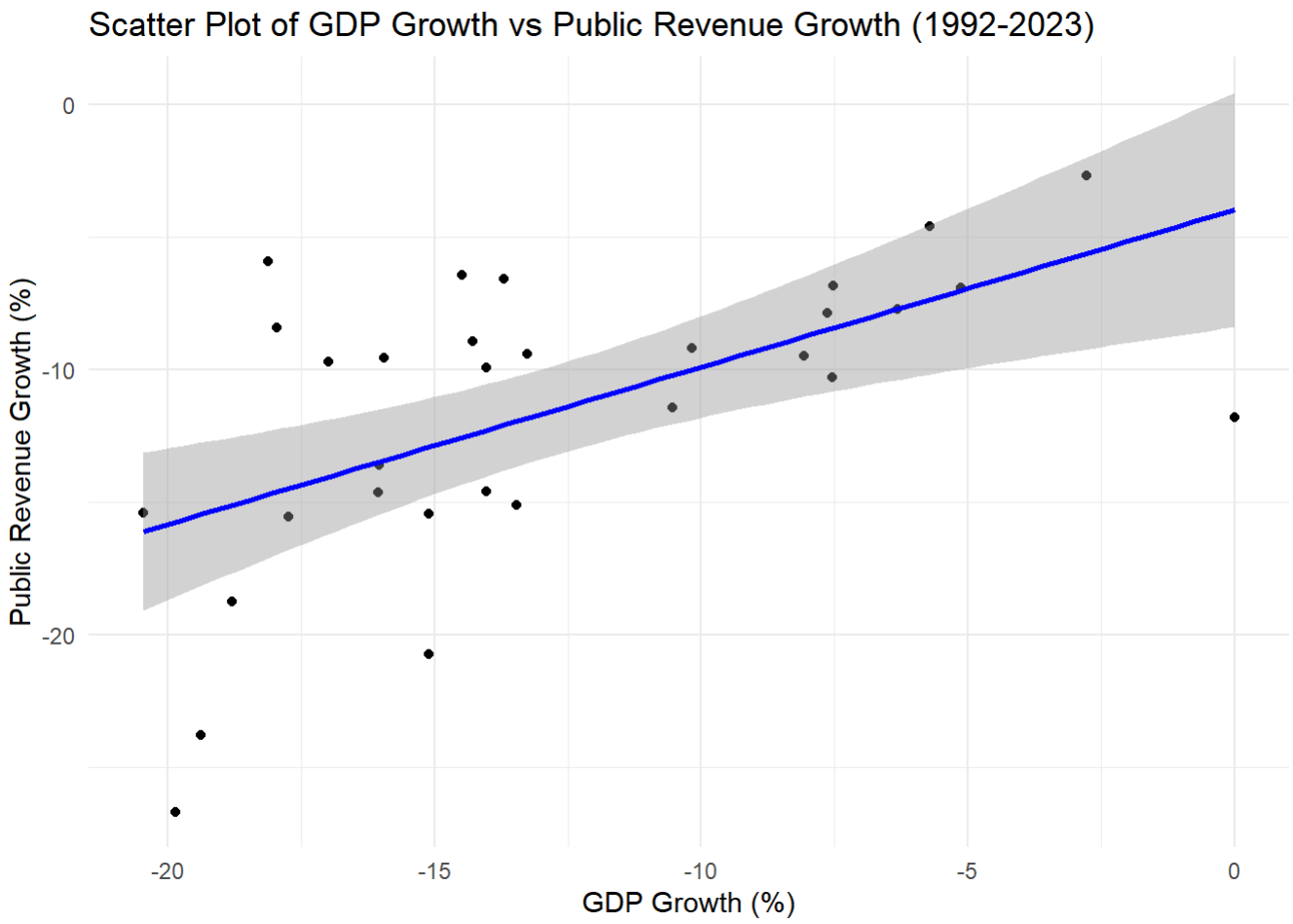
```
library(ggplot2)

# 读取Excel文件
data <- read_excel("C:\\Users\\zhy\\Desktop\\GDPvs预算收入.xlsx")
# 将中文列名转换为英文
colnames(data) <- c("Index", "Year", "GDP", "PublicRevenue")

# 清理数据，移除空行
data <- data %>% filter(!is.na(Year))

# 仅保留1992-2023年的数据
data <- data %>% filter(Year >= 1992 & Year <= 2023)

# 确保 Year 是数值型
data$Year <- as.numeric(data$Year)

# 计算GDP增速和一般公共预算收入增速
data <- data %>%
  mutate(GDP_Growth = (GDP - lag(GDP)) / lag(GDP) * 100,
         PublicRevenue_Growth = (PublicRevenue - lag(PublicRevenue)) / lag(PublicRevenue) * 10

# 移除第一行（因为增速计算会产生NA）
data <- data %>% filter(!is.na(GDP_Growth))

# 绘制折线图
ggplot(data, aes(x = Year)) +
  geom_line(aes(y = GDP_Growth, color = "GDP Growth"), group = 1) +
  geom_line(aes(y = PublicRevenue_Growth, color = "Public Revenue Growth"), group = 1) +
```

```
  labs(title = "GDP Growth vs Public Revenue Growth (1992-2023)",
       x = "Year",
       y = "Growth Rate (%)",
       color = "Indicator") +
  theme_minimal() +
  scale_x_continuous(breaks = seq(1992, 2023, by = 5),  # 每5年显示一个标签
                     labels = seq(1992, 2023, by = 5)) +  # 标签与 breaks 一致
  theme(axis.text.x = element_text(angle = 45, hjust = 1))  # 旋转横轴标签，避免重叠
```




```
# 绘制散点图
ggplot(data, aes(x = GDP_Growth, y = PublicRevenue_Growth)) +
  geom_point() +
  geom_smooth(method = "lm", col = "blue") +
  labs(title = "Scatter Plot of GDP Growth vs Public Revenue Growth (1992-2023)",
       x = "GDP Growth (%)",
       y = "Public Revenue Growth (%)") +
  theme_minimal()
```

```
`geom_smooth()` using formula = 'y ~ x'
```




```
# 计算相关系数
correlation <- cor(data$GDP_Growth, data$PublicRevenue_Growth, use = "complete.obs")
print(paste("The correlation coefficient between GDP Growth and Public Revenue Growth is:", co
```

```
[1] "The correlation coefficient between GDP Growth and Public Revenue Growth is:
0.575754186485253"
```

一般公共预算收入的细分 I

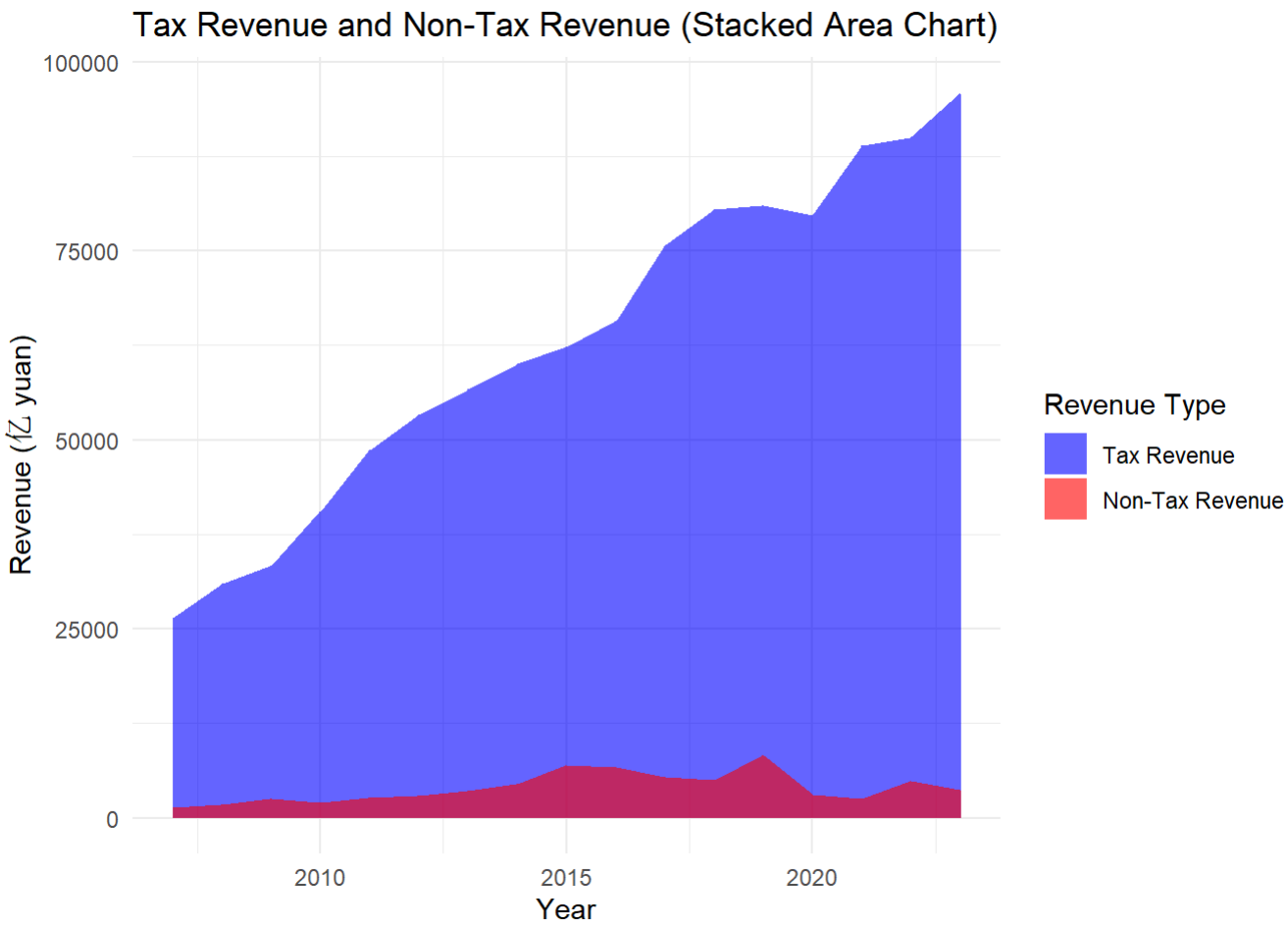
```
# 加载必要的库
library(readxl)
library(ggplot2)
library(dplyr)
library(tidyr)
# 读取 Excel 文件
df <- read_excel("C:\\Users\\zhy\\Desktop\\财政支出细分.xlsx")

# 如果列名不正确，手动设置列名
colnames(df) <- c("Index", "Year", "Total_Revenue", "Tax_Revenue", "Domestic_VAT", "Non_Tax_Re

# 检查数据类型
str(df)
```

```
tibble [135 × 8] (S3: tbl_df/tbl/data.frame)
 $ Index            : chr [1:135] "1" "2" "3" "4" ...
 $ Year             : chr [1:135] "2023" "2022" "2021" "2020" ...
 $ Total_Revenue    : num [1:135] 99567 94887 91470 82771 89309 ...
 $ Tax_Revenue      : num [1:135] 95834 89977 88946 79644 81020 ...
 $ Domestic_VAT     : num [1:135] 34588 24256 31753 28353 31160 ...
 $ Non_Tax_Revenue  : num [1:135] 3733 4910 2524 3126 8289 ...
 $ GDP              : num [1:135] 1294272 1204724 1149237 1013567 986515 ...
 $ PublicExpenditure: num [1:135] 274623 260552 245673 245679 238858 ...
```

```r
# 将 Year 列转换为数值型
df$Year <- as.numeric(df$Year)

# 将 Tax_Revenue、Non_Tax_Revenue 和 GDP 列转换为数值型
df$Tax_Revenue <- as.numeric(df$Tax_Revenue)
df$Non_Tax_Revenue <- as.numeric(df$Non_Tax_Revenue)
df$GDP <- as.numeric(df$GDP)

# 移除包含缺失值的行
df <- df %>% drop_na(Tax_Revenue, Non_Tax_Revenue, GDP)

# 再次检查数据
head(df)
```

```
# A tibble: 6 × 8
  Index  Year Total_Revenue Tax_Revenue Domestic_VAT Non_Tax_Revenue      GDP
  <chr> <dbl>         <dbl>       <dbl>        <dbl>           <dbl>    <dbl>
1 1      2023        99567.      95834.       34588.           3733. 1294272
2 2      2022        94887.      89977.       24256.           4910. 1204724
3 3      2021        91470.      88946.       31753.           2524. 1149237
4 4      2020        82771.      79644.       28353.           3126. 1013567
5 5      2019        89309.      81020.       31160.           8289.  986515.
6 6      2018        85456.      80448.       30753.           5008.  919281.
# ℹ 1 more variable: PublicExpenditure <dbl>
```

```r
# 1. 绘制税收收入和非税收入的面积堆积图
ggplot(df, aes(x = Year)) +
  geom_area(aes(y = Tax_Revenue, fill = "Tax Revenue"), alpha = 0.6) +
  geom_area(aes(y = Non_Tax_Revenue, fill = "Non-Tax Revenue"), alpha = 0.6) +
  labs(
    title = "Tax Revenue and Non-Tax Revenue (Stacked Area Chart)",
    x = "Year",
    y = "Revenue (亿 yuan)",
    fill = "Revenue Type"
  ) +
  theme_minimal() +
  scale_fill_manual(
    values = c("Tax Revenue" = "blue", "Non-Tax Revenue" = "red"),  # 确保值与 fill 映射一致
    breaks = c("Tax Revenue", "Non-Tax Revenue")  # 明确指定 breaks
  )
```

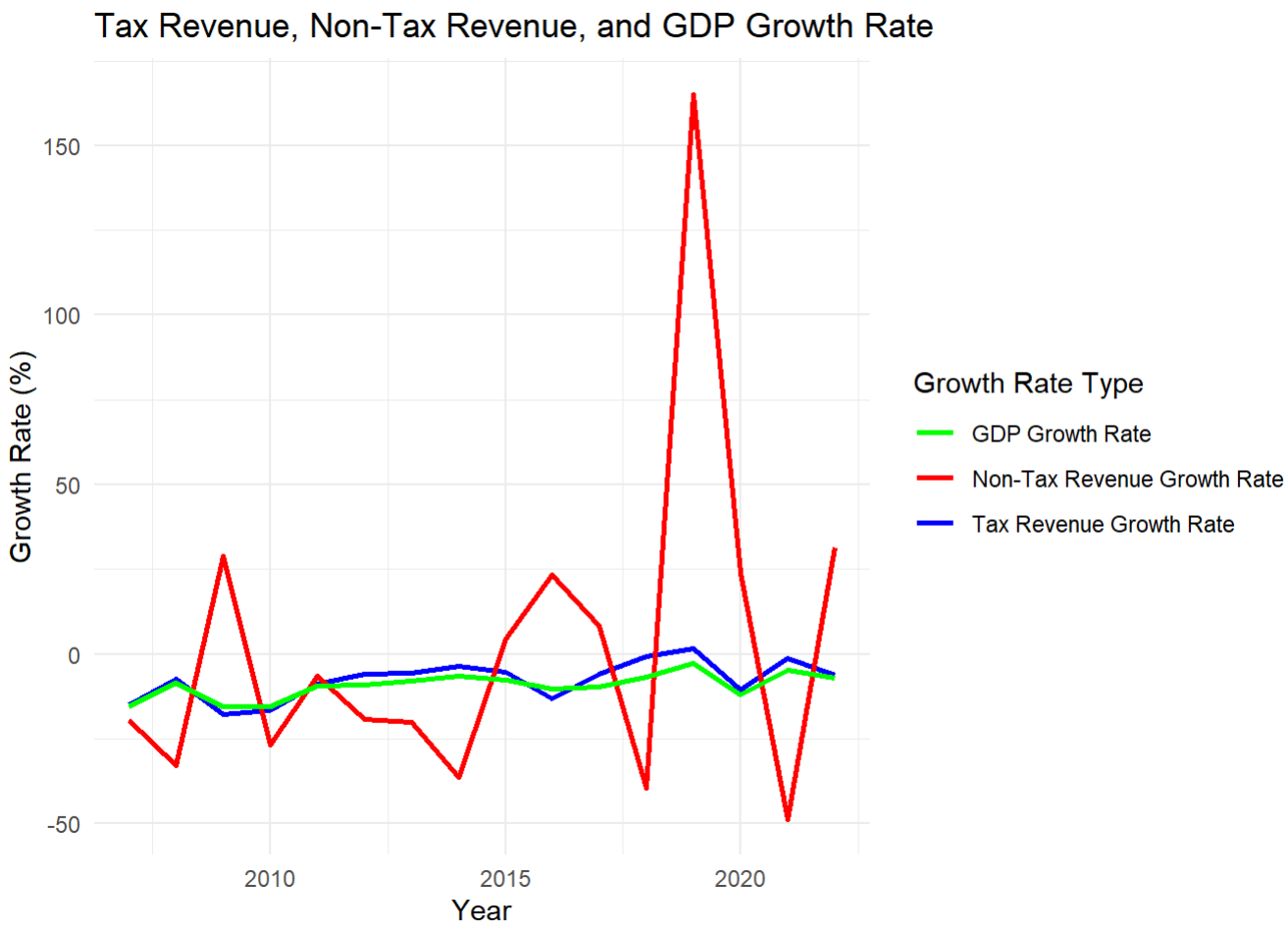



```r
# 2. 计算税收收入、非税收入和 GDP 的增速
df <- df %>%
  mutate(
    Tax_Growth_Rate = (Tax_Revenue - lag(Tax_Revenue)) / lag(Tax_Revenue) * 100,
    Non_Tax_Growth_Rate = (Non_Tax_Revenue - lag(Non_Tax_Revenue)) / lag(Non_Tax_Revenue) * 10
    GDP_Growth_Rate = (GDP - lag(GDP)) / lag(GDP) * 100
  )

# 移除增速计算中的缺失值
df <- df %>% drop_na(Tax_Growth_Rate, Non_Tax_Growth_Rate, GDP_Growth_Rate)

# 查看增速数据
head(df)
```

```
# A tibble: 6 × 11
  Index  Year Total_Revenue Tax_Revenue Domestic_VAT Non_Tax_Revenue      GDP
  <chr> <dbl>         <dbl>       <dbl>        <dbl>           <dbl>    <dbl>
1 2      2022        94887.      89977.       24256.           4910. 1204724
2 3      2021        91470.      88946.       31753.           2524. 1149237
3 4      2020        82771.      79644.       28353.           3126. 1013567
4 5      2019        89309.      81020.       31160.           8289.  986515.
5 6      2018        85456.      80448.       30753.           5008.  919281.
6 7      2017        81119.      75697.       28166.           5426.  832036.
# ℹ 4 more variables: PublicExpenditure <dbl>, Tax_Growth_Rate <dbl>,
#   Non_Tax_Growth_Rate <dbl>, GDP_Growth_Rate <dbl>
```

```r
# 绘制税收收入、非税收入和 GDP 增速的折线图
ggplot(df, aes(x = Year, group = 1)) +  # 明确指定 group = 1
  geom_line(aes(y = Tax_Growth_Rate, color = "Tax Revenue Growth Rate"), size = 1) +
  geom_line(aes(y = Non_Tax_Growth_Rate, color = "Non-Tax Revenue Growth Rate"), size = 1) +
  geom_line(aes(y = GDP_Growth_Rate, color = "GDP Growth Rate"), size = 1) +
  labs(
    title = "Tax Revenue, Non-Tax Revenue, and GDP Growth Rate",
    x = "Year",
    y = "Growth Rate (%)",
    color = "Growth Rate Type"
  ) +
  theme_minimal() +
  scale_color_manual(
    values = c(
      "Tax Revenue Growth Rate" = "blue",
      "Non-Tax Revenue Growth Rate" = "red",
      "GDP Growth Rate" = "green"  # 添加 GDP 增速的颜色
    )
  )
```

```
Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
ℹ Please use `linewidth` instead.
```

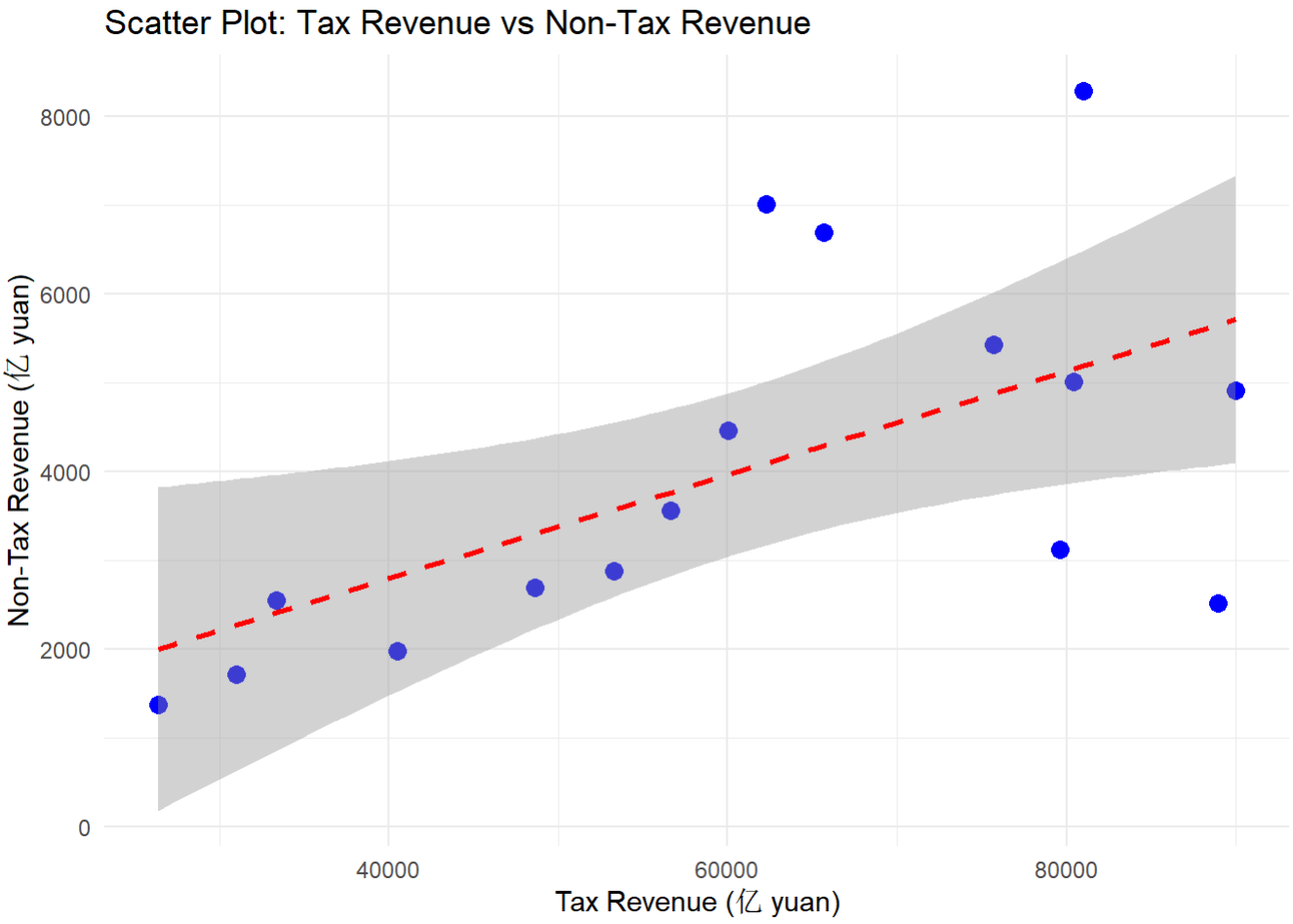



```r
# 3. 计算税收收入和非税收入的相关性
correlation <- cor(df$Tax_Revenue, df$Non_Tax_Revenue, use = "complete.obs")
print(paste("Correlation between Tax Revenue and Non-Tax Revenue:", round(correlation, 2)))
```

```
[1] "Correlation between Tax Revenue and Non-Tax Revenue: 0.59"
```

```r
# 绘制税收收入和非税收入的散点图
ggplot(df, aes(x = Tax_Revenue, y = Non_Tax_Revenue)) +
  geom_point(color = "blue", size = 3) +
  geom_smooth(method = "lm", color = "red", linetype = "dashed") +
  labs(
    title = "Scatter Plot: Tax Revenue vs Non-Tax Revenue",
    x = "Tax Revenue (亿 yuan)",
    y = "Non-Tax Revenue (亿 yuan)"
  ) +
  theme_minimal()
```

```
`geom_smooth()` using formula = 'y ~ x'
```



一般公共预决算收入细分II

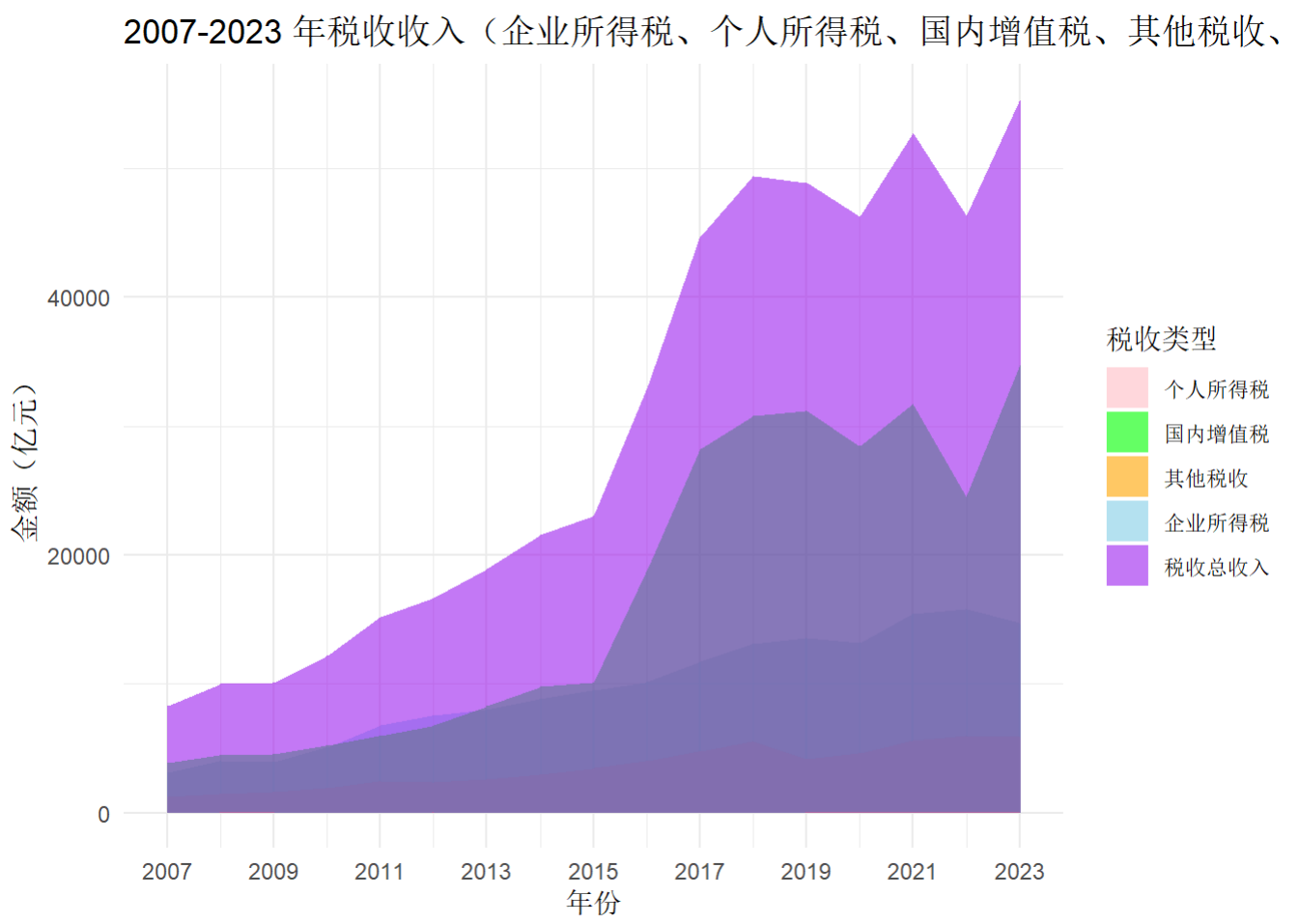
```r
# 加载必要的库
library(readxl)
library(ggplot2)
library(dplyr)
library(tidyr)

# 读取 Excel 文件
df <- read_excel("C:\\Users\\zhy\\Desktop\\细分的公共决算收入.xlsx")
```

```r
# 转换列名为英文
colnames(df) <- c("Index", "Year", "LocalRevenue", "CorporateTax", "PersonalTax", "VAT", "Othe

# 检查数据类型
str(df)
```

```
tibble [134 × 10] (S3: tbl_df/tbl/data.frame)
 $ Index             : chr [1:134] "1" "2" "3" "4" ...
 $ Year              : chr [1:134] "2023" "2022" "2021" "2020" ...
 $ LocalRevenue      : num [1:134] 220065 205704 193237 183361 175440 ...
 $ CorporateTax      : num [1:134] 14693 15828 15437 13168 13518 ...
 $ PersonalTax       : num [1:134] 5910 5969 5597 4627 4154 ...
 $ VAT               : num [1:134] 34746 24462 31767 28438 31187 ...
 $ OtherTax          : num [1:134] 14.3 25.5 19.5 22.8 39.8 ...
 $ PenaltyRevenue    : num [1:134] 3630 3687 3432 2969 2929 ...
 $ NonTaxRevenue     : num [1:134] 31927 32119 27295 25475 24100 ...
 $ AdministrativeFees: num [1:134] 3449 3576 3724 3419 3483 ...
```

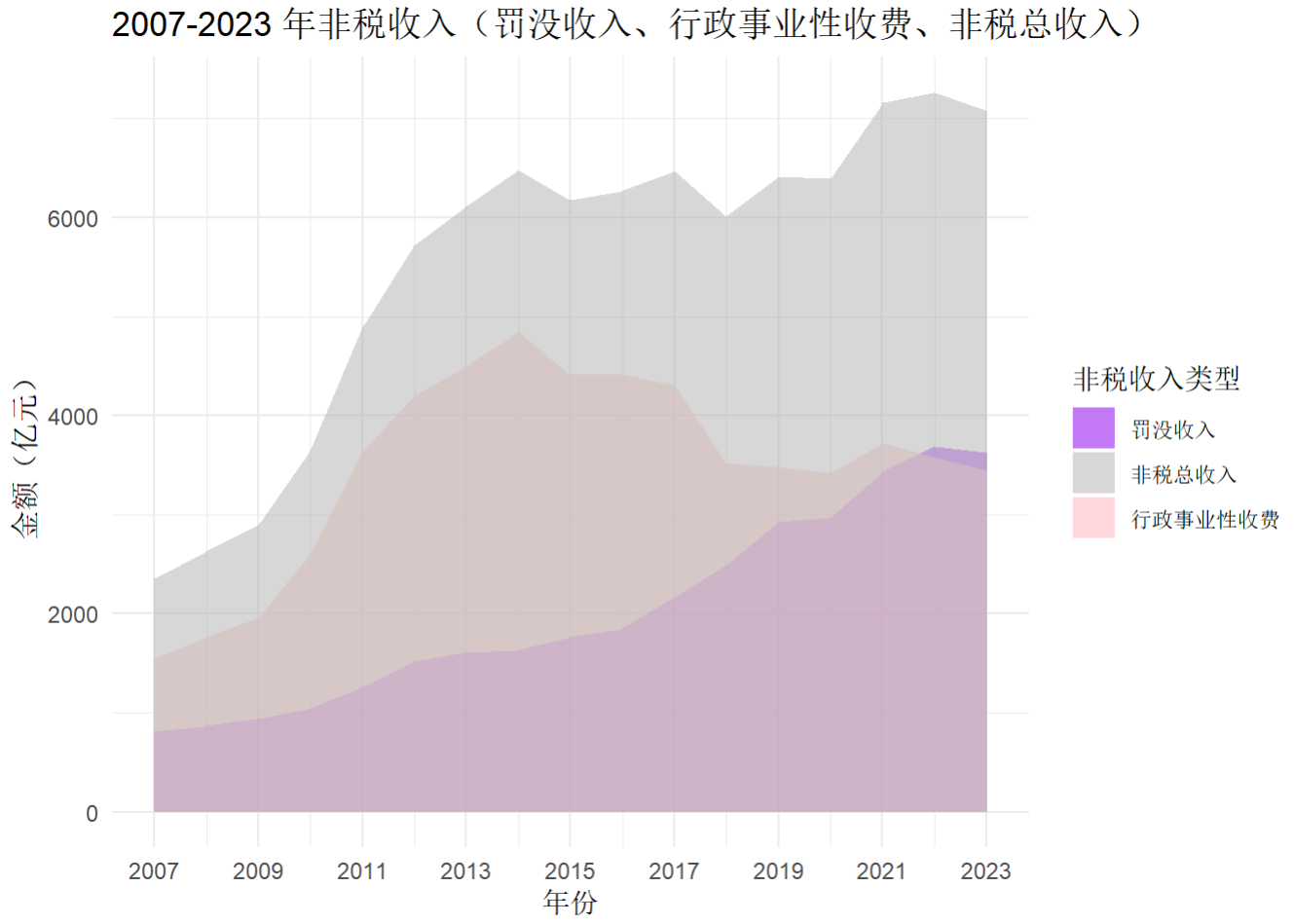
```r
# 将 Year 列转换为数值型
df$Year <- as.numeric(df$Year)

# 筛选 2007-2023 年的数据
df <- df %>% filter(Year >= 2007 & Year <= 2023)

# 移除包含缺失值的行
df <- df %>% drop_na(CorporateTax, PersonalTax, VAT, PenaltyRevenue, AdministrativeFees)

# 计算税收总收入和非税总收入
df <- df %>%
  mutate(
    TotalTaxRevenue = CorporateTax + PersonalTax + VAT + OtherTax,  # 税收总收入
    TotalNonTaxRevenue = PenaltyRevenue + AdministrativeFees        # 非税总收入
  )

# 1. 绘制税收收入面积堆积图（包含税收总收入）
ggplot(df, aes(x = Year)) +
  geom_area(aes(y = CorporateTax, fill = "企业所得税"), alpha = 0.6) +
  geom_area(aes(y = PersonalTax, fill = "个人所得税"), alpha = 0.6) +
  geom_area(aes(y = VAT, fill = "国内增值税"), alpha = 0.6) +
  geom_area(aes(y = OtherTax, fill = "其他税收"), alpha = 0.6) +
  geom_area(aes(y = TotalTaxRevenue, fill = "税收总收入"), alpha = 0.6) +  # 添加税收总收入
  labs(
    title = "2007-2023 年税收收入（企业所得税、个人所得税、国内增值税、其他税收、税收总收入）",
    x = "年份",
    y = "金额（亿元）",
    fill = "税收类型"
  ) +
  theme_minimal() +
  scale_fill_manual(
```

```
    values = c(
      "企业所得税" = "skyblue",
      "个人所得税" = "pink",
      "国内增值税" = "green",
      "其他税收" = "orange",
      "税收总收入" = "purple"  # 设置税收总收入颜色
    )
  ) +
  scale_x_continuous(breaks = seq(2007, 2023, by = 2))  # 每两年显示一个年份标签
```




```
# 2. 绘制非税收入面积堆积图（包含非税总收入）
ggplot(df, aes(x = Year)) +
  geom_area(aes(y = PenaltyRevenue, fill = "罚没收入"), alpha = 0.6) +
  geom_area(aes(y = AdministrativeFees, fill = "行政事业性收费"), alpha = 0.6) +
  geom_area(aes(y = TotalNonTaxRevenue, fill = "非税总收入"), alpha = 0.6) +  # 添加非税总收入
  labs(
    title = "2007-2023 年非税收入（罚没收入、行政事业性收费、非税总收入）",
    x = "年份",
    y = "金额（亿元）",
    fill = "非税收入类型"
  ) +
  theme_minimal() +
  scale_fill_manual(
    values = c(
      "罚没收入" = "purple",
```

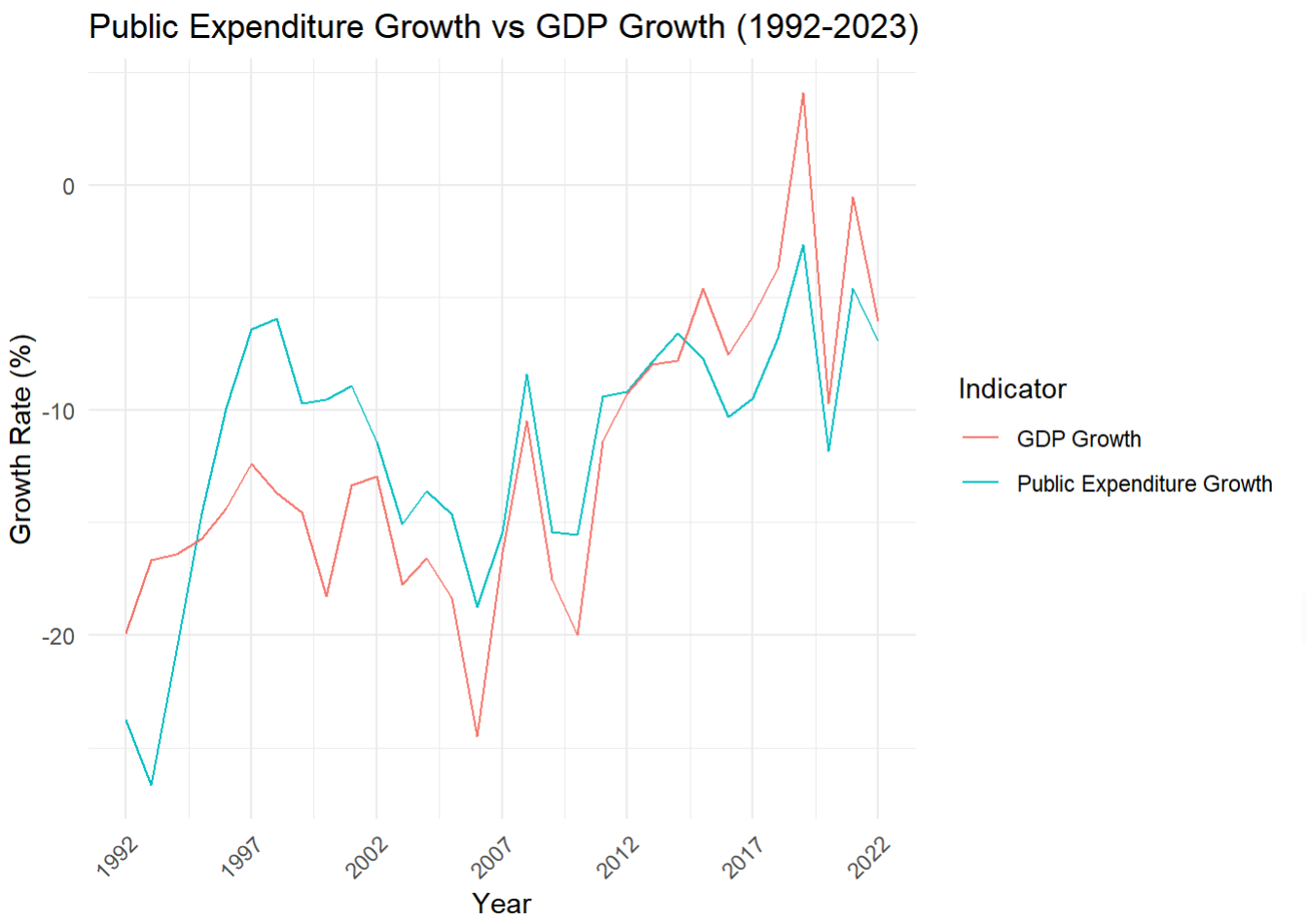
```
    "行政事业性收费" = "pink",
    "非税总收入" = "gray"  # 设置非税总收入颜色
  )
) +
scale_x_continuous(breaks = seq(2007, 2023, by = 2))  # 每两年显示一个年份标签
```



GDP和一般公共预算支出（全国）增速对比

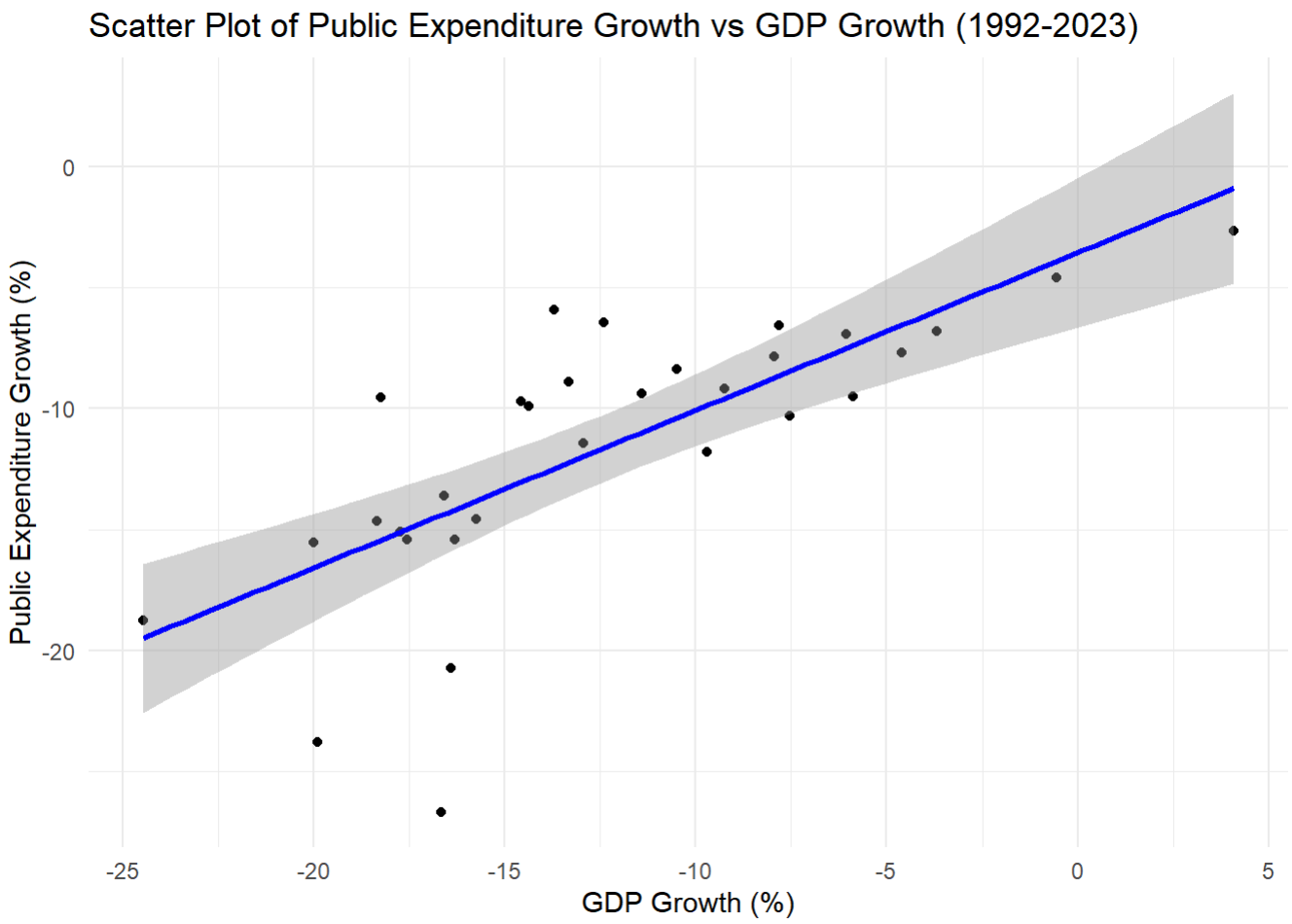
```
# 加载必要的包
library(readxl)
library(dplyr)
library(ggplot2)

# 读取Excel文件
data <- read_excel("C:\\Users\\zhy\\Desktop\\GDPvs预算支出.xlsx")

# 将中文列名转换为英文
colnames(data) <- c("Index", "Year", "PublicExpenditure", "GDP")

# 清理数据，移除空行
data <- data %>% filter(!is.na(Year))

# 仅保留1992-2023年的数据
data <- data %>% filter(Year >= 1992 & Year <= 2023)
```

```r
# 确保 Year 是数值型
data$Year <- as.numeric(data$Year)

# 计算预算支出增速和GDP增速
data <- data %>%
  mutate(PublicExpenditure_Growth = (PublicExpenditure - lag(PublicExpenditure)) / lag(PublicE
         GDP_Growth = (GDP - lag(GDP)) / lag(GDP) * 100)

# 移除第一行（因为增速计算会产生NA）
data <- data %>% filter(!is.na(PublicExpenditure_Growth))

# 绘制折线图
ggplot(data, aes(x = Year)) +
  geom_line(aes(y = PublicExpenditure_Growth, color = "Public Expenditure Growth"), group = 1)
  geom_line(aes(y = GDP_Growth, color = "GDP Growth"), group = 1) +
  labs(title = "Public Expenditure Growth vs GDP Growth (1992-2023)",
       x = "Year",
       y = "Growth Rate (%)",
       color = "Indicator") +
  theme_minimal() +
  scale_x_continuous(breaks = seq(1992, 2023, by = 5),  # 每5年显示一个标签
                     labels = seq(1992, 2023, by = 5)) +  # 标签与 breaks 一致
  theme(axis.text.x = element_text(angle = 45, hjust = 1))  # 旋转横轴标签，避免重叠
```



```r
# 绘制散点图
ggplot(data, aes(x = GDP_Growth, y = PublicExpenditure_Growth)) +
  geom_point() +
  geom_smooth(method = "lm", col = "blue") +
  labs(title = "Scatter Plot of Public Expenditure Growth vs GDP Growth (1992-2023)",
       x = "GDP Growth (%)",
       y = "Public Expenditure Growth (%)") +
  theme_minimal()
```

```
`geom_smooth()` using formula = 'y ~ x'
```




```r
# 计算相关系数
correlation <- cor(data$GDP_Growth, data$PublicExpenditure_Growth, use = "complete.obs")
print(paste("The correlation coefficient between Public Expenditure Growth and GDP Growth is:"
```

```
[1] "The correlation coefficient between Public Expenditure Growth and GDP Growth is:
0.740490116719109"
```

一般公共预决算收入细分 I

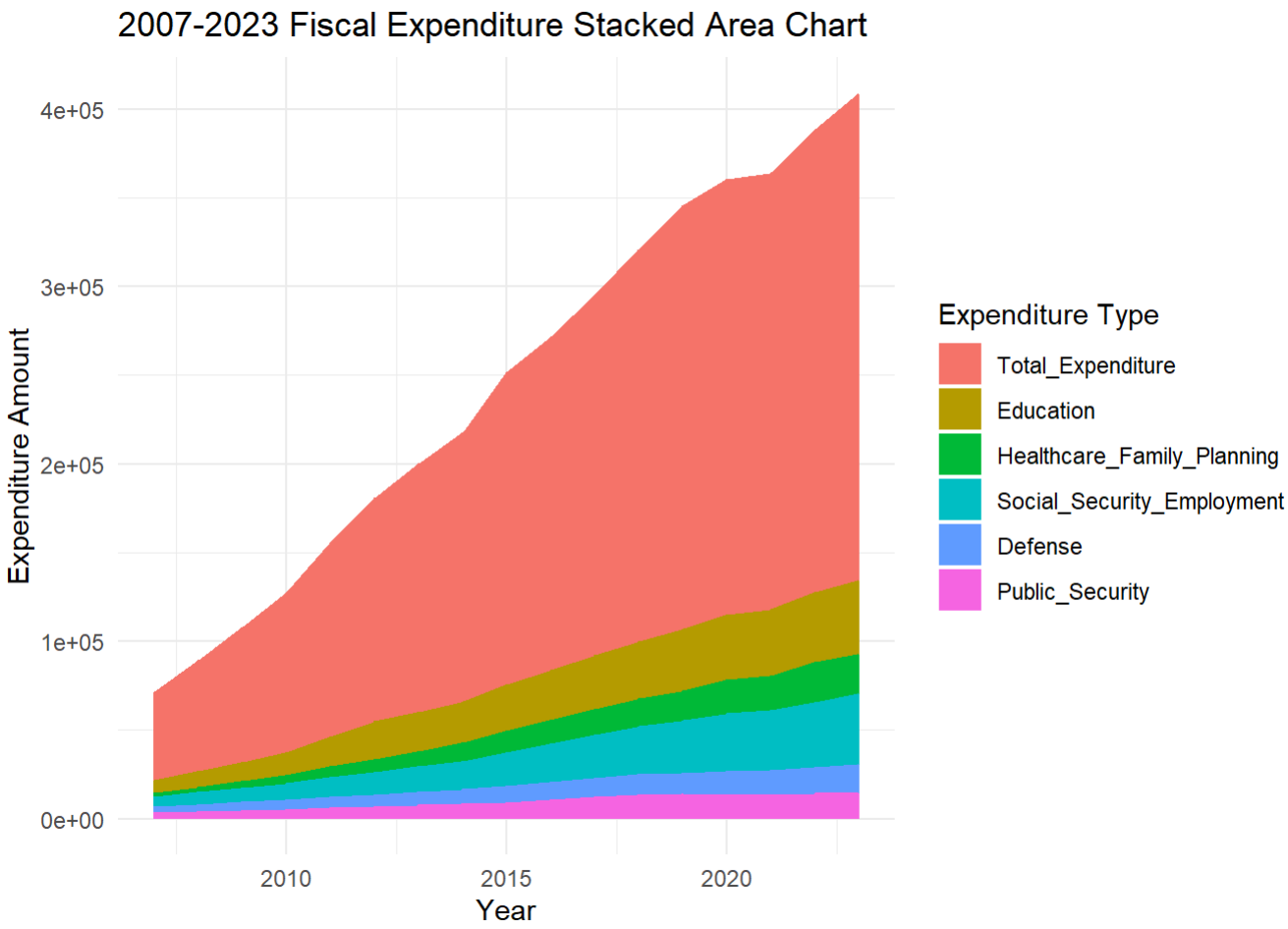
```r
# 加载必要的包
library(readxl)  # 用于读取Excel文件
```

```
library(ggplot2) # 用于绘图
library(reshape2) # 用于数据转换
```

```
载入程序包：'reshape2'

The following object is masked from 'package:tidyr':

    smiths
```

```
# 1. 读取Excel文件
data <- read_excel("C:\\Users\\zhy\\Desktop\\细分公共预算支出.xlsx")

# 2. 转换列名为英文
colnames(data) <- c("Index", "Year", "Total_Expenditure", "Public_Services", "Diplomacy", "Def

# 确保 Year 是数值型
data$Year <- as.numeric(data$Year)

# 3. 筛选2007-2023年的数据
data_filtered <- data[data$Year >= 2007 & data$Year <= 2023, ]

# 4. 选择需要的列
selected_data <- data_filtered[, c("Year", "Total_Expenditure", "Education", "Healthcare_Famil

# 5. 将数据转换为长格式
long_data <- melt(selected_data, id.vars = "Year", variable.name = "Expenditure_Type", value.n

# 6. 绘制面积堆积图
ggplot(long_data, aes(x = Year, y = Expenditure_Amount, fill = Expenditure_Type)) +
  geom_area() +
  labs(title = "2007-2023 Fiscal Expenditure Stacked Area Chart",
       x = "Year",
       y = "Expenditure Amount",
       fill = "Expenditure Type") +
  theme_minimal()
```

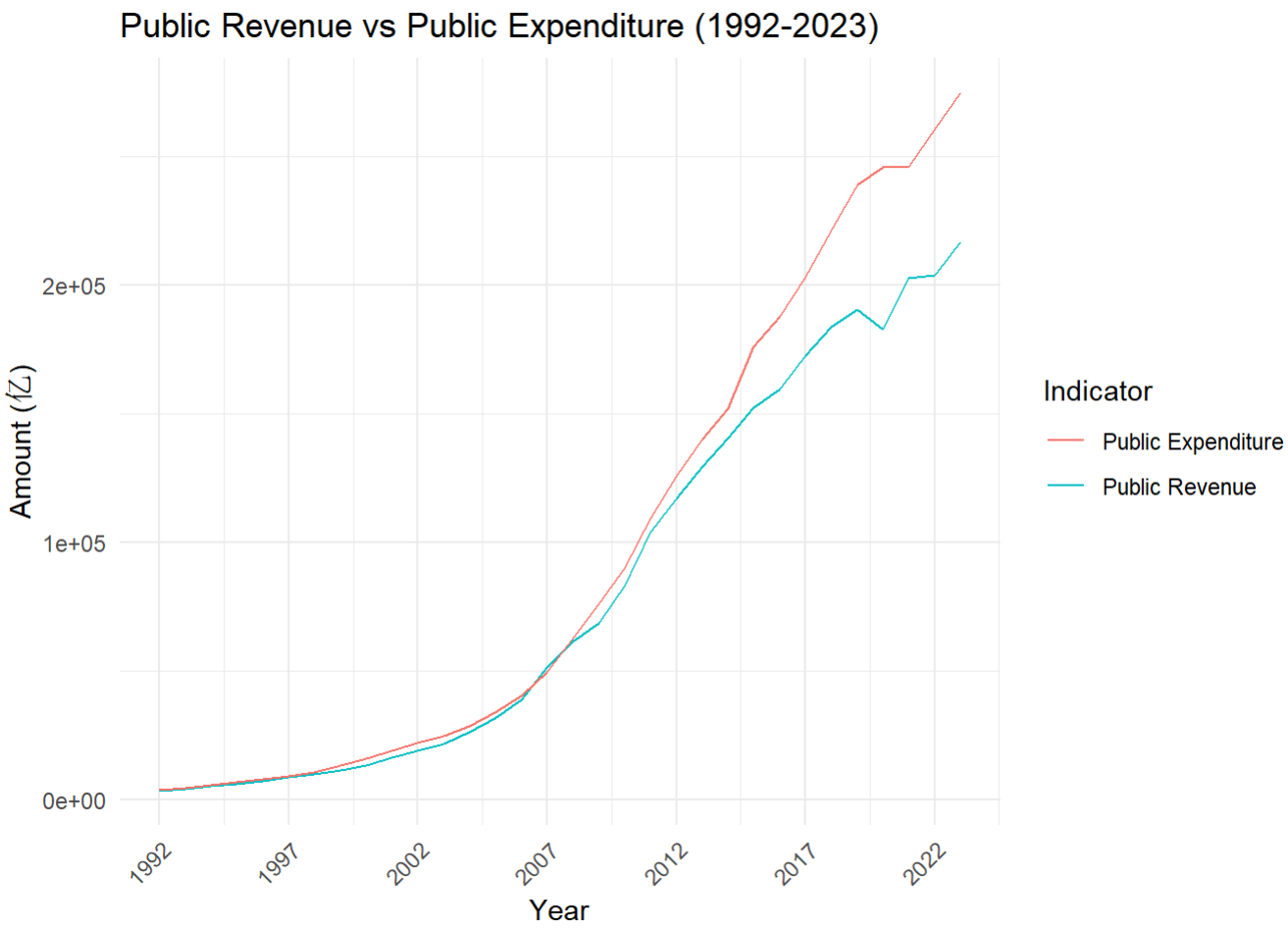
```
Warning: Removed 378 rows containing non-finite outside the scale range
(`stat_align()`).
```



财政收入vs支出

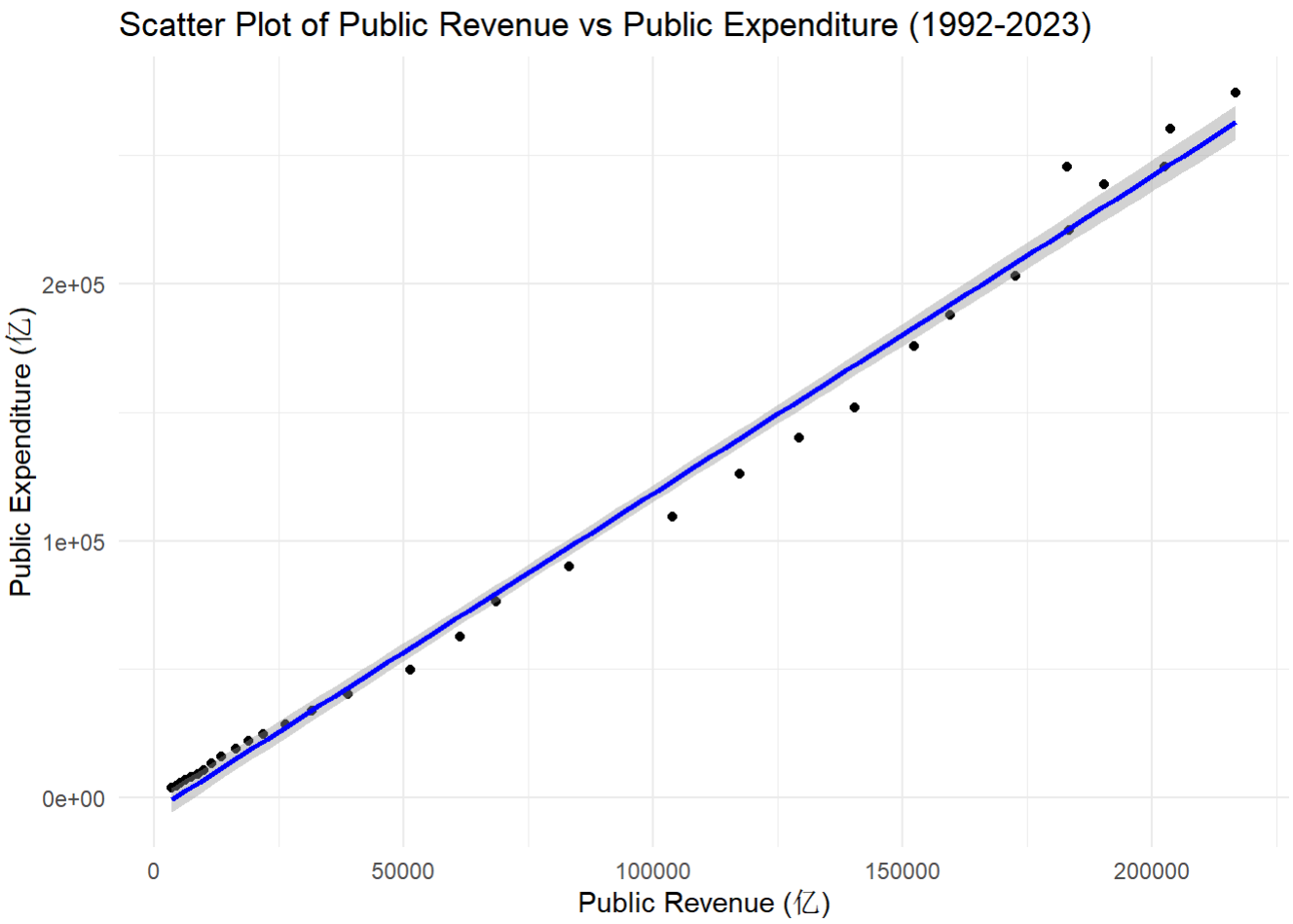
```
# 加载必要的包
library(readxl)
library(dplyr)
library(ggplot2)

# 读取Excel文件
data <- read_excel("C:\\Users\\zhy\\Desktop\\收入vs支出.xlsx")


# 将中文列名转换为英文
colnames(data) <- c("Year", "PublicRevenue", "PublicExpenditure")

# 清理数据，移除空行
data <- data %>% filter(!is.na(Year))

# 仅保留1992-2023年的数据
data <- data %>% filter(Year >= 1992 & Year <= 2023)

# 确保 Year 是数值型
data$Year <- as.numeric(data$Year)

# 绘制绝对值对比折线图
ggplot(data, aes(x = Year)) +
```

```
  geom_line(aes(y = PublicRevenue, color = "Public Revenue"), group = 1) +
  geom_line(aes(y = PublicExpenditure, color = "Public Expenditure"), group = 1) +
  labs(title = "Public Revenue vs Public Expenditure (1992-2023)",
       x = "Year",
       y = "Amount (亿)",
       color = "Indicator") +
  theme_minimal() +
  scale_x_continuous(breaks = seq(1992, 2023, by = 5),  # 每5年显示一个标签
                     labels = seq(1992, 2023, by = 5)) +  # 标签与 breaks 一致
  theme(axis.text.x = element_text(angle = 45, hjust = 1))  # 旋转横轴标签，避免重叠
```




```
# 绘制绝对值对比散点图
ggplot(data, aes(x = PublicRevenue, y = PublicExpenditure)) +
  geom_point() +
  geom_smooth(method = "lm", col = "blue") +
  labs(title = "Scatter Plot of Public Revenue vs Public Expenditure (1992-2023)",
       x = "Public Revenue (亿)",
       y = "Public Expenditure (亿)") +
  theme_minimal()
```

```
`geom_smooth()` using formula = 'y ~ x'
```



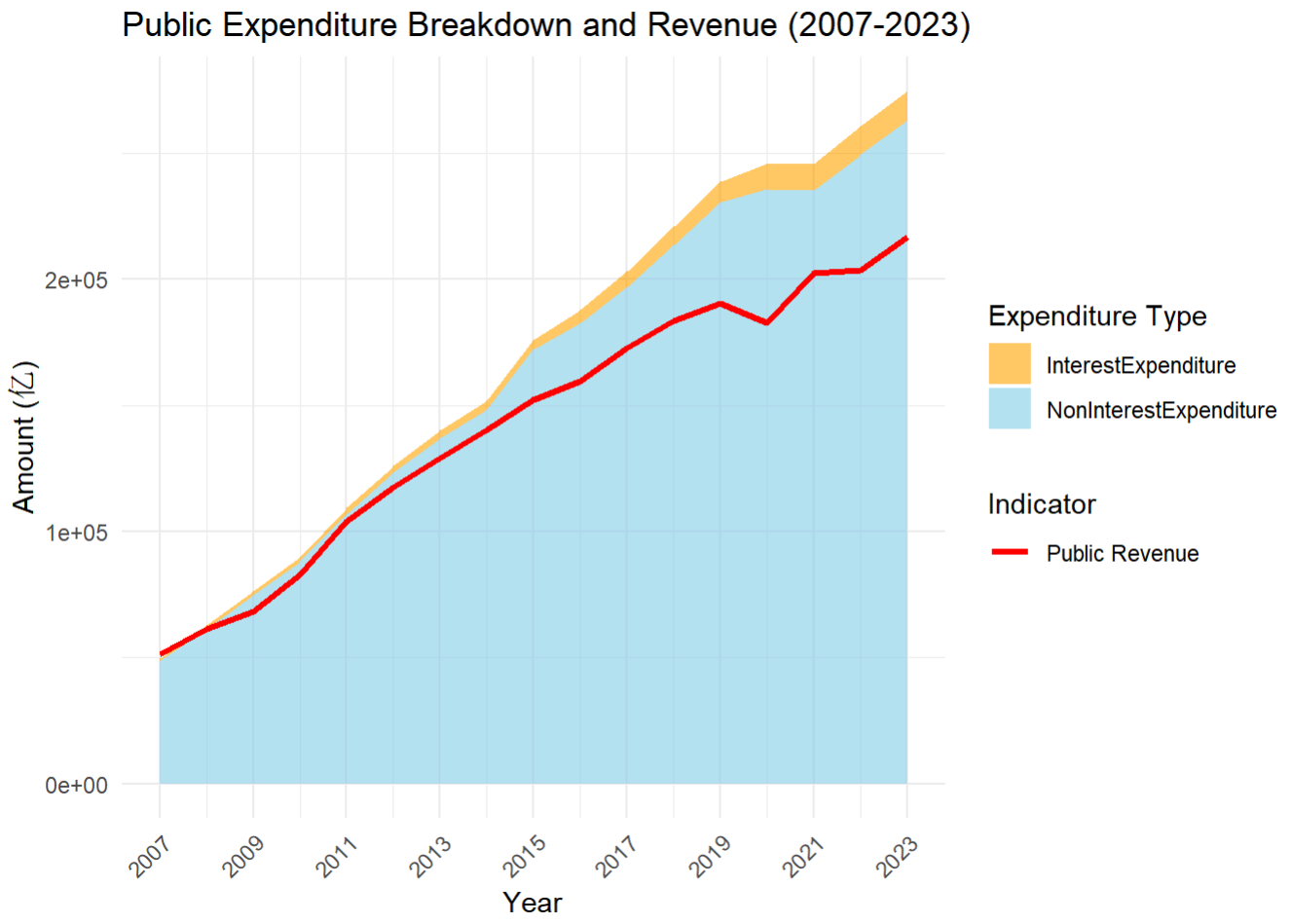

```
# 计算绝对值相关系数
correlation <- cor(data$PublicRevenue, data$PublicExpenditure, use = "complete.obs")
print(paste("The correlation coefficient between Public Revenue and Public Expenditure is:", c
```

```
[1] "The correlation coefficient between Public Revenue and Public Expenditure is:
0.995592991105381"
```

财政收支平衡

```
# 加载必要的包
library(readxl)
library(dplyr)
library(ggplot2)
library(tidyr)

# 读取Excel文件
data <- read_excel("C:\\Users\\zhy\\Desktop\\表格.xlsx")

# 将中文列名转换为英文
colnames(data) <- c("Year", "GDP", "PublicRevenue", "PublicExpenditure", "DebtInterestPayment"

# 清理数据，移除空行
data <- data %>% filter(!is.na(Year))
```

```r
# 仅保留2007年以后的数据
data <- data %>% filter(Year >= 2007)

# 确保 Year 是数值型
data$Year <- as.numeric(data$Year)

# 拆分一般公共预算支出为非利息支出和利息支出
data <- data %>%
  mutate(NonInterestExpenditure = PublicExpenditure - DebtInterestPayment,
         InterestExpenditure = DebtInterestPayment)

# 将数据转换为长格式，便于绘图
data_long <- data %>%
  pivot_longer(cols = c(NonInterestExpenditure, InterestExpenditure),
               names_to = "ExpenditureType",
               values_to = "ExpenditureValue")

# 绘制面积堆积图和折线图
ggplot() +
  # 绘制非利息支出和利息支出的面积堆积图
  geom_area(data = data_long, aes(x = Year, y = ExpenditureValue, fill = ExpenditureType), alp
  # 绘制一般公共预算收入的折线图
  geom_line(data = data, aes(x = Year, y = PublicRevenue, color = "Public Revenue"), size = 1.
  # 设置标题和标签
  labs(title = "Public Expenditure Breakdown and Revenue (2007-2023)",
       x = "Year",
       y = "Amount (亿)",
       fill = "Expenditure Type",
       color = "Indicator") +
  # 设置横轴标注间隔
  scale_x_continuous(breaks = seq(2007, 2023, by = 2)) +
  # 设置颜色和图例
  scale_fill_manual(values = c("NonInterestExpenditure" = "skyblue", "InterestExpenditure" = "
  scale_color_manual(values = c("Public Revenue" = "red")) +
  # 设置主题
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



一般公共预算收入细分

```
# 加载必要的包
library(readxl)
library(dplyr)
library(gridExtra)
```

```
载入程序包：'gridExtra'

The following object is masked from 'package:dplyr':

    combine
```

```
# 读取Excel文件
data <- read_excel("C:\\Users\\zhy\\Desktop\\细分公共预算支出.xlsx")

# 将中文列名转换为英文
colnames(data) <- c("Index", "Year", "TotalExpenditure", "GeneralPublicServices", "Diplomacy",
                    "Education", "ScienceTechnology", "CultureSportsMedia", "SocialSecurityEmp
                    "EnergyConservationEnvironmentalProtection", "UrbanRuralCommunityAffairs",
                    "Transportation", "ResourceExplorationElectricityInformation", "CommerceSe
                    "AidOtherRegions", "HousingSecurity", "GrainMaterialReserves", "DebtIntere
                    "PublicRevenue")
```

```r
# 清理数据，移除空行
data <- data %>% filter(!is.na(Year))

# 仅保留2007-2023年的数据
data <- data %>% filter(Year >= 2007 & Year <= 2023)

# 确保 Year 是数值型
data$Year <- as.numeric(data$Year)

# 将所有支出列转换为数值型
data <- data %>%
  mutate(across(TotalExpenditure:OtherExpenditure, as.numeric))

# 计算各项支出的 CAGR
growth_rates <- data %>%
  summarise(across(TotalExpenditure:OtherExpenditure,
                   ~ (last(.) / first(.))^(1 / (n() - 1)) - 1,
                   .names = "CAGR_{col}"))

# 计算每项支出的标准差
std_dev <- data %>%
  summarise(across(TotalExpenditure:OtherExpenditure,
                   ~ sd((. - lag(.)) / lag(.) * 100, na.rm = TRUE),
                   .names = "StdDev_{col}"))

# 标记异常值（增速超过平均值 ± 2倍标准差）
anomalies <- data %>%
  mutate(across(TotalExpenditure:OtherExpenditure,
                ~ ifelse(abs((. - lag(.)) / lag(.) * 100 - mean((. - lag(.)) / lag(.) * 100, n
                         2 * sd((. - lag(.)) / lag(.) * 100, na.rm = TRUE), "异常", "正常"),
                .names = "Anomaly_{col}"))

# 汇总异常值总数
anomaly_counts <- anomalies %>%
  summarise(across(starts_with("Anomaly"), ~ sum(. == "异常", na.rm = TRUE)))

# 输出结果
result <- data.frame(
  Category = gsub("CAGR_", "", colnames(growth_rates)),  # 去掉 "CAGR_" 前缀
  AverageGrowthRate = as.numeric(-growth_rates[1, ]),
  StandardDeviation = as.numeric(std_dev[1, ]),
  Anomalies = as.numeric(anomaly_counts[1, ])
)



# 将 result 转换为表格图形对象
table_grob <- gridExtra::tableGrob(result, theme = gridExtra::ttheme_minimal(base_size = 10))

# 使用 ggplot2 绘制表格
p <- ggplot() +
  annotation_custom(table_grob, xmin = -Inf, xmax = Inf, ymin = -Inf, ymax = Inf) +
  theme_void()  # 使用空白主题
```

```r
# 保存为图片
ggsave("result_table.png", p, width = 12, height = 8, dpi = 300)
```

政府专项债券发行vs城镇固投

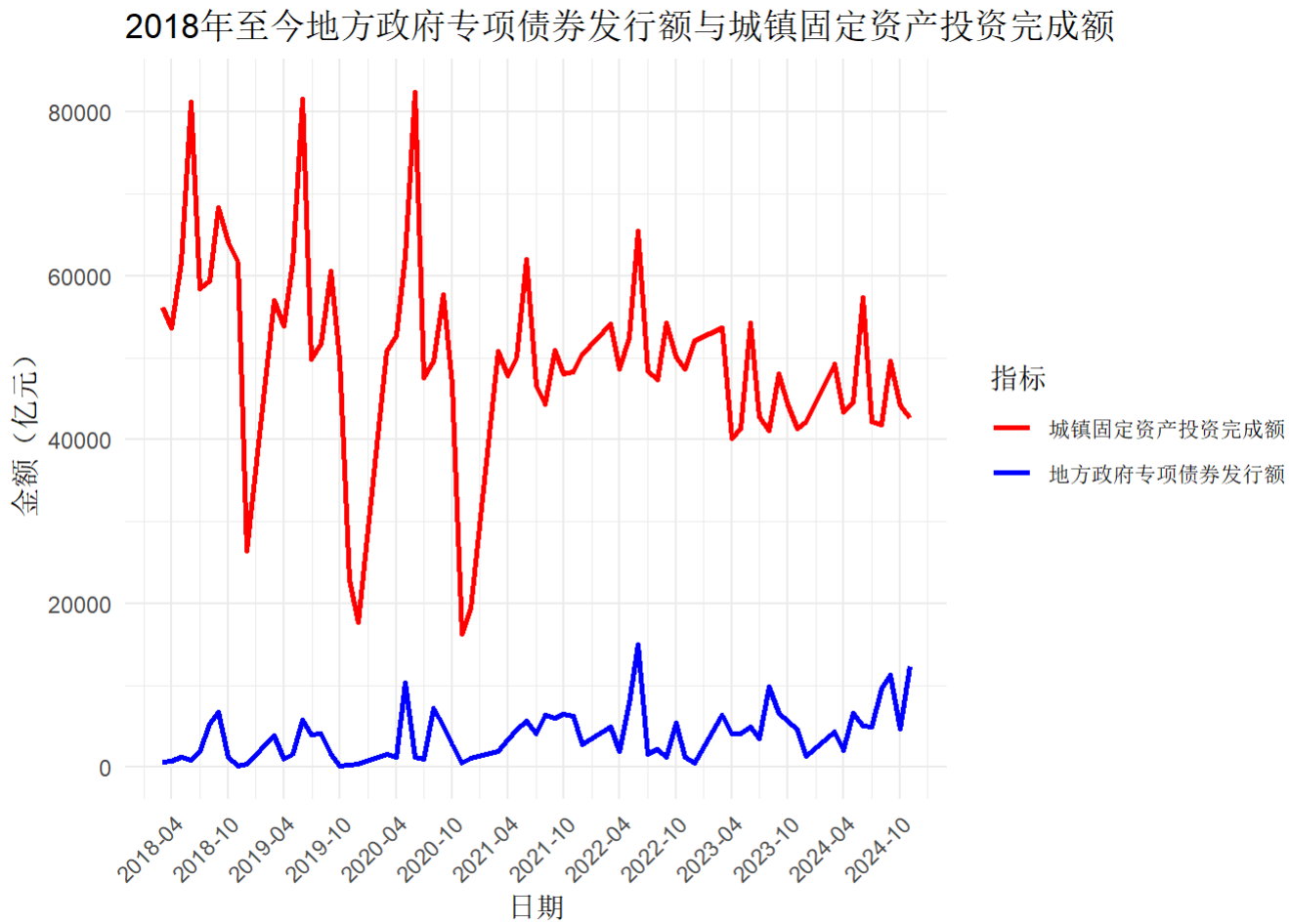
```r
# 加载必要的库
library(readxl)
library(ggplot2)
library(dplyr)
library(tidyr)

df <- read_excel("C:\\Users\\zhy\\Desktop\\债券发行额vs固投.xlsx")

# 转换列名为英文
colnames(df) <- c("Index", "Date", "LocalGovBond", "FixedAssetInvestment")

# 检查数据类型
str(df)
```

```
tibble [437 × 4] (S3: tbl_df/tbl/data.frame)
 $ Index               : chr [1:437] "1" "2" "3" "4" ...
 $ Date                : chr [1:437] "2024-11" "2024-10" "2024-09" "2024-08" ...
 $ LocalGovBond        : num [1:437] 12307 4659 11219 9558 4943 ...
 $ FixedAssetInvestment: num [1:437] 42617 44244 49593 41774 42220 ...
```

```r
# 将 Date 列转换为日期格式
df$Date <- as.Date(paste0(df$Date, "-01"), format = "%Y-%m-%d")

# 将 LocalGovBond 和 FixedAssetInvestment 列转换为数值型
df$LocalGovBond <- as.numeric(df$LocalGovBond)
df$FixedAssetInvestment <- as.numeric(df$FixedAssetInvestment)

# 筛选 2018 年至今的数据
df <- df %>% filter(Date >= as.Date("2018-01-01"))

# 移除包含缺失值的行
df <- df %>% drop_na(LocalGovBond, FixedAssetInvestment)

# 再次检查数据
head(df)
```

```
# A tibble: 6 × 4
  Index Date       LocalGovBond FixedAssetInvestment
  <chr> <date>            <dbl>                <dbl>
1 1     2024-11-01        12307                42617
2 2     2024-10-01         4659                44244
3 3     2024-09-01        11219                49593
4 4     2024-08-01         9558                41774
5 5     2024-07-01         4943                42220
6 6     2024-06-01         5082                57385
```

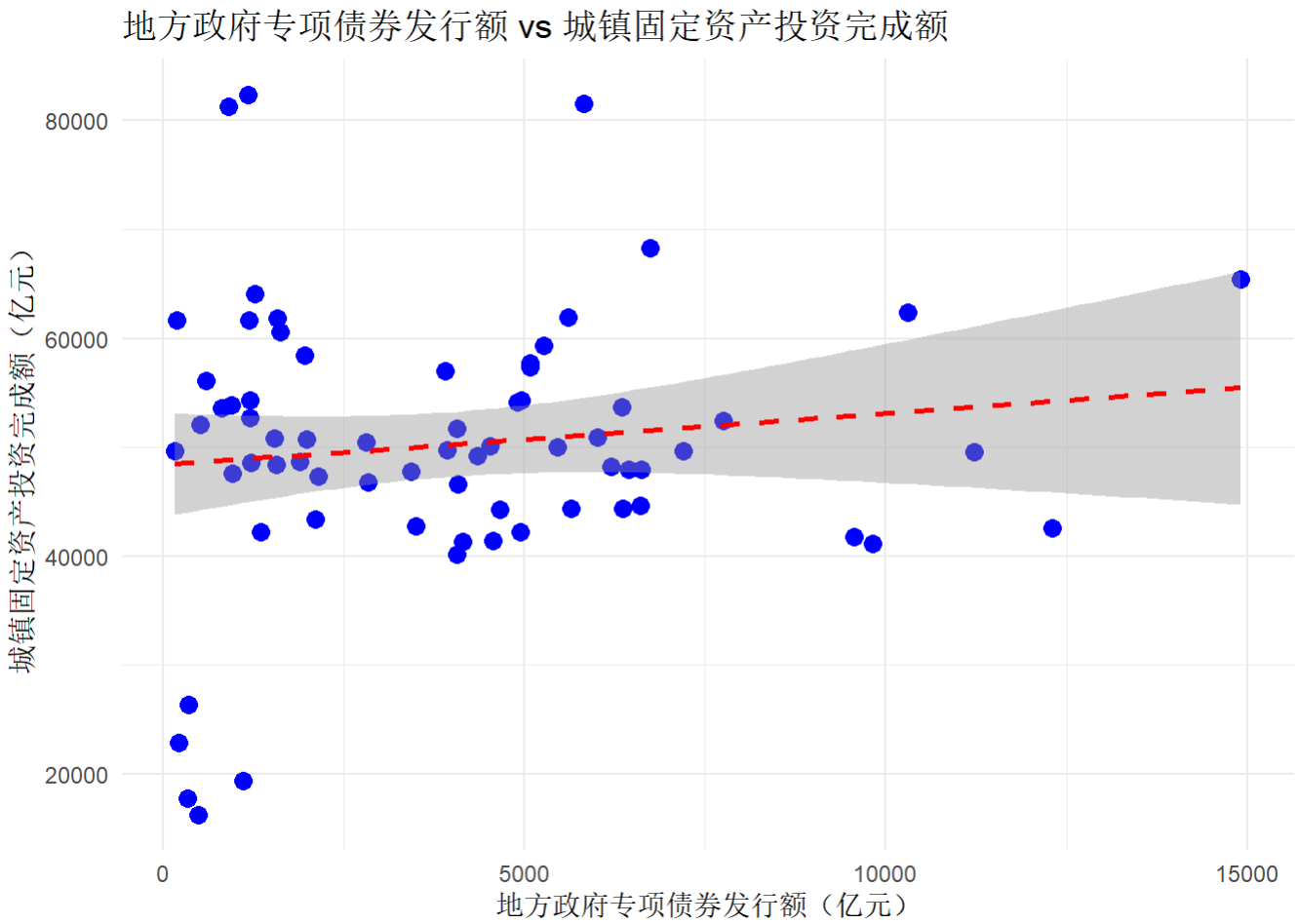
```r
# 绘制折线图
ggplot(df, aes(x = Date)) +
  geom_line(aes(y = LocalGovBond, color = "地方政府专项债券发行额"), size = 1) +
  geom_line(aes(y = FixedAssetInvestment, color = "城镇固定资产投资完成额"), size = 1) +
  labs(
    title = "2018年至今地方政府专项债券发行额与城镇固定资产投资完成额",
    x = "日期",
    y = "金额（亿元）",
    color = "指标"
  ) +
  theme_minimal() +
  scale_color_manual(
    values = c("地方政府专项债券发行额" = "blue", "城镇固定资产投资完成额" = "red")
  ) +
  scale_x_date(date_labels = "%Y-%m", date_breaks = "6 months") +  # 每6个月显示一个日期标签
  theme(axis.text.x = element_text(angle = 45, hjust = 1))  # 旋转 x 轴标签
```




```r
# 1. 绘制散点图
ggplot(df, aes(x = LocalGovBond, y = FixedAssetInvestment)) +
  geom_point(color = "blue", size = 3) +  # 绘制散点
  geom_smooth(method = "lm", color = "red", linetype = "dashed") +  # 添加线性拟合线
  labs(
    title = "地方政府专项债券发行额 vs 城镇固定资产投资完成额",
    x = "地方政府专项债券发行额（亿元）",
    y = "城镇固定资产投资完成额（亿元）"
```

```
  ) +
  theme_minimal()
```

```
`geom_smooth()` using formula = 'y ~ x'
```




```
# 2. 计算相关性
correlation <- cor(df$LocalGovBond, df$FixedAssetInvestment, use = "complete.obs")
print(paste("地方政府专项债券发行额与城镇固定资产投资完成额的相关性为:", round(correlation, 2)))
```

```
[1] "地方政府专项债券发行额与城镇固定资产投资完成额的相关性为: 0.12"
```