

# Analyzing PA USDA Federal Funds Data

Stat 184 Final Document

Robert Pitsko, Zach Thomas

2025-12-18

## Table of contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Programming Paradigm and Ethical Principals . . . . .	2
<b>2</b>	<b>Data Analysis</b>	<b>2</b>
2.1	Function Code . . . . .	3
2.2	Object Code . . . . .	5
<b>3</b>	<b>Author Contributions</b>	<b>5</b>
<b>4</b>	<b>References</b>	<b>5</b>
<b>5</b>	<b>Code Appendix</b>	<b>5</b>

## 1 Introduction

When looking for data for this project we came across the U.S. Department of Agriculture (USDA) Economic Research website. From that website, we decided to explore the available Federal Funds data for the state of Pennsylvania, which ranges from years 2004 to 2010. This data captures federal funding information relating to specific counties in the state of Pennsylvania. Each row by default is a line item that contains an object code, function code, and spending total. The combined data (from 2004 to 2010) has over 5,000 rows, each representing a federal program for a given year.

We took an Exploratory Data Analysis approach as we did not yet understand the data and we did not have any hypothesis to test. In the EDA process, we cleaned and combined the data and focused on key columns like object code, function code, year, and spending. Our

goal was to explore the relationship between spending and object code over time as well as the relationship between spending and function code over time.

It is worth noting that even though this data came from the USDA Economic Research website, it includes federal spending on Pennsylvania from all government agencies not just the USDA.

## **1.1 Programming Paradigm and Ethical Principals**

### **1.1.1 Tidyverse**

We decided to use the tidyverse libraries and methods for a majority of our data wrangling and visualization as it is easy to use and efficient for small to medium data sets. Our code, shown in the appendix, follows the tidyverse style guided making it easy to read.

### **1.1.2 FAIR and CARE**

This data meets the FAIR principals as it is publicly available on the USDA Economic Research Federal Funds website. The data sets can be downloaded by anyone on the internet. The website also has clear documentation going over how the data is obtained, and explains all object and function codes. The data set and the documentation use clear and accessible language.

The data meets the CARE principals as the purpose of this data is to provide transparency for the public good. The data is inclusive and designed to minimize harm and maximize benefit through transparency.

## **2 Data Analysis**

Before we dive into the function code and object code, we wanted to examine the data set as a whole. Summary statistics provide an overview of the data's central tendencies, variability, and range, which helps us understand the scope and patterns of the data. The full summary table can be seen in Table 1.

We see that total federal spending in Pennsylvania is monotonically increasing from 2004 to 2010, growing from approximately \$102.6 billion to \$164.4 billion, an increase of about 60% over the seven-year period. The median program size remains relatively stable and small across all years, hovering between \$1.23 million and \$1.64 million, which indicates that most federal programs provide modest amounts of funding. In contrast, the mean program size is substantially higher, ranging from \$137.49 million in 2004 to \$197.38 million in 2010, suggesting that the distribution is right-skewed with a few very large programs pulling the average upward.

Table 1: Annual summary statistics for federal funds to Pennsylvania (in millions of dollars). Total Funds represents the sum across all federal programs. Min/Max/Median/Mean describe the distribution of individual program spending. Programs indicates the number of federal programs with funding each year. Note: Negative values may represent loan repayments, program adjustments, or recovered funds.

Year	Total	Min	Max	Median	Mean	Std Dev	Programs
<b>2,004</b>	102,557.1	-2.90	16,987.38	1.53	137.66	942.49	745
<b>2,005</b>	106,059.5	-5.27	17,620.88	1.50	143.91	1,011.60	737
<b>2,006</b>	110,160.2	-3.84	18,420.84	1.49	155.81	1,075.13	707
<b>2,007</b>	124,538.8	-47.71	19,239.98	1.23	163.65	1,132.60	761
<b>2,008</b>	133,843.8	-43.88	20,067.21	1.32	168.78	1,199.69	793
<b>2,009</b>	156,246.7	-15.23	21,169.65	1.62	186.01	1,293.84	840
<b>2,010</b>	164,402.9	-3.84	22,116.71	1.56	197.60	1,338.21	832

One may notice that the minimum values are all negative. These negative values typically represent loan repayments to the federal government, accounting adjustments, or recovered over payments rather than actual expenditures

## 2.1 Function Code

Figure 1 displays the temporal trends in federal spending across different functional categories from 2004 to 2010. The most striking pattern is the dominance of Income Security spending, which accounts for the majority of federal funds to Pennsylvania throughout this period. Income Security spending grew steadily from approximately \$70 billion in 2004 to nearly \$100 billion in 2010, representing programs such as Social Security, Medicare, and unemployment compensation. This category alone represents more than half of all federal spending in the state.

The remaining functional categories show much more modest spending levels, all clustering below \$25 billion annually. National Functions ranks second, displaying gradual growth from about \$15 billion to \$25 billion over the period. Defense spending shows relatively stable funding around \$10-15 billion, with a slight increase in later years. Community Development exhibits the most notable growth among the smaller categories, rising from approximately \$8 billion in 2004 to nearly \$20 billion by 2009 before declining slightly in 2010—this surge likely reflects increased federal investment in infrastructure and housing assistance during and after the 2008 financial crisis. Human Resources, Agriculture, and Natural Resources both show relatively flat trajectories at low levels throughout the period, each remaining below \$5 billion annually. The “Other” category, which captures null or unclassified function codes, represents an extremely small portion of total spending.

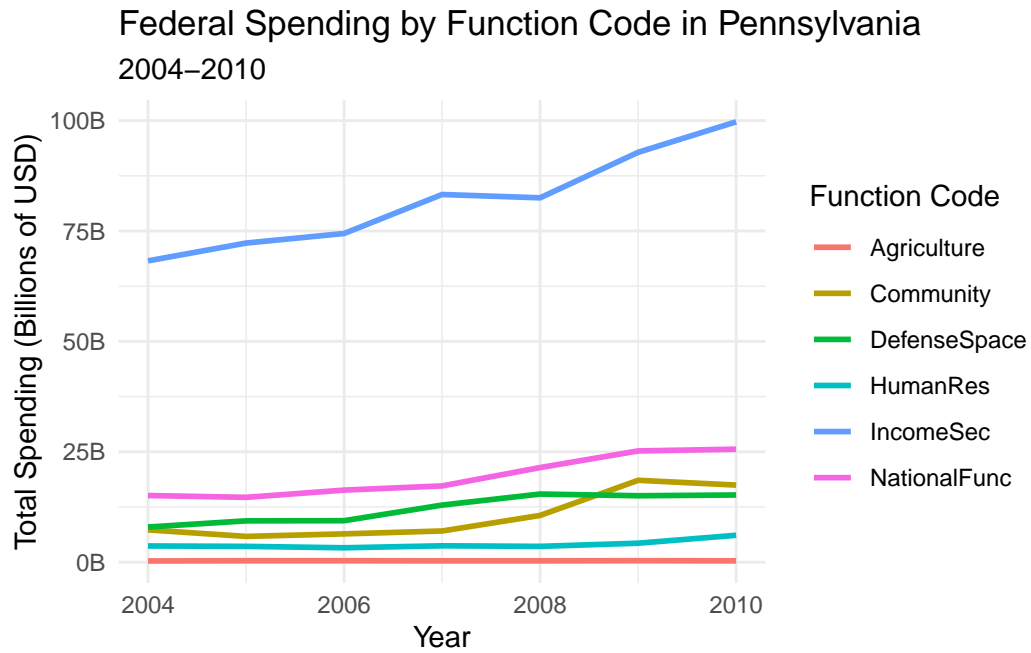


Figure 1: Spending by Function Code

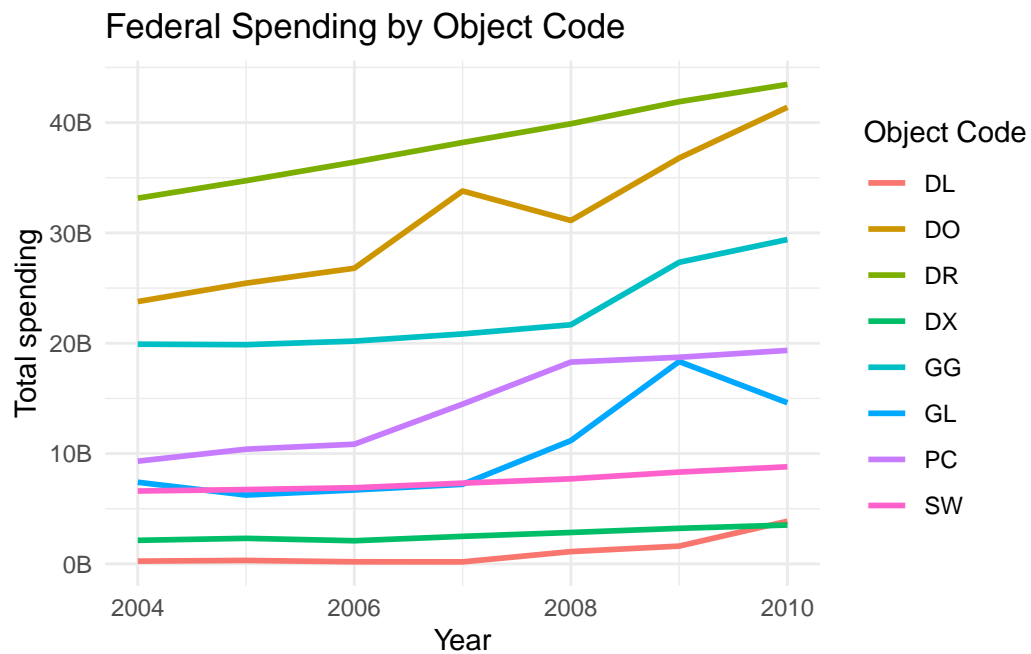


Figure 2: Federal spending in Pennsylvania by object code from 2004 to 2010.

## 2.2 Object Code

Here we see our Object Spending covering our x objects that contribute to the federal funds spending. On top is retirement which matches up to our functions data showing it on top. However, we see a closer difference in the objects than we do the functions with direct payments to individuals and grants close behind. This shows us that majority of the spending in PA for agricultural related functions is salaries and retirement which does make sense for a densely populated state like PA.

## 3 Author Contributions

- **Zach Thomas:** Visualization (Object Code), Writing (Intro/EDA narrative), Data curation, QMD File Creation
- **Robert Pitsko:** Visualization (Function Code), Data curation, Writing (FAIR/CARE), QMD File Creation

## 4 References

Data Link: <https://www.ers.usda.gov/data-products/federal-funds>

## 5 Code Appendix

### 5.0.1 Get Packages

```
# Author: Zach Thomas | Reviewer: Robert Pitsko
# Style guide: Tidyverse style guide

library(tidyverse)
library(readxl)
library(here)
library(scales)
library(kableExtra)
```

### 5.0.2 Load our data frame

```

# Author: Zach Thomas | Reviewer: Robert Pitsko

files <- list.files(here("data"),
pattern = "FederalFunds_\\d{4}Data_PA\\.xls$", full.names = TRUE)

get_year <- function(path) readr::parse_number(basename(path))

combined_df <- purrr::map_dfr(files, function(f) {

read_excel(f) %>%
  mutate(Year = get_year(f))})

combined_df <- combined_df %>%
  mutate(
    StateTotal = readr::parse_number(as.character(StateTotal)),
    ObjectCode = as.character(ObjectCode)
  ) %>%
  filter(!is.na(StateTotal), !is.na(ObjectCode))

```

### 5.0.3 Create Summary Statistics Table

```

# Summary statistics by year across all programs
year_summary <- combined_df %>%
  select(StateTotal, Year) %>%
  mutate(StateTotal = as.numeric(StateTotal)) %>%
  group_by(Year) %>%
  summarize(
    total_spending = sum(StateTotal) / 1e6, # Convert to millions
    min_spending = min(StateTotal) / 1e6,
    max_spending = max(StateTotal) / 1e6,
    median_spending = median(StateTotal) / 1e6,
    mean_spending = mean(StateTotal) / 1e6,
    std_dev = sd(StateTotal) / 1e6,
    rows = n()
  )

# Create formatted table for year summary
year_summary %>%
  kbl(
    col.names = c(

```

```

    "Year",
    "Total",
    "Min",
    "Max",
    "Median",
    "Mean",
    "Std Dev",
    "Programs"
  ),
  digits = 2,
  format.args = list(big.mark = ",",)
) %>%
kable_styling(
  bootstrap_options = c("striped", "hover", "condensed"),
  full_width = FALSE,
  position = "center"
) %>%
column_spec(1, bold = TRUE)

```

#### 5.0.4 Create Spending by Function Code Graph

```

# Author: Robert Pitsko | Reviewer: Zach Thomas

function_spending_df <- combined_df %>%
  mutate(
    FunctionCodeDescription = if_else(
      is.na(FunctionCodeDescription),
      "Other/Unknown",
      as.character(FunctionCodeDescription)
    )
  ) %>%
  group_by(Year, FunctionCodeDescription) %>%
  summarise(TotalSpend = sum(StateTotal, na.rm = TRUE), .groups = "drop")

ggplot(
  function_spending_df,
  aes(
    x = Year,
    y = TotalSpend,
    color = FunctionCodeDescription,

```

```

    group = FunctionCodeDescription
  )
) +
  geom_line(linewidth = 1) +
  scale_y_continuous(labels = label_number(scale = 1e-9, suffix = "B")) +
  labs(
    title = "Federal Spending by Function Code in Pennsylvania",
    subtitle = "2004-2010",
    x = "Year",
    y = "Total Spending (Billions of USD)",
    color = "Function Code"
  ) +
  theme_minimal()

```

### 5.0.5 Create Spending by Object Code Graph

```

# Author: Zach Thomas | Reviewer: Robert Pitsko

object_spending_df <- combined_df %>%
  group_by(Year, ObjectCode) %>%
  summarise(TotalSpend = sum(StateTotal, na.rm = TRUE), .groups = "drop")

# Author: Zach Thomas | Reviewer: Robert Pitsko

ggplot(object_spending_df, aes(x = Year, y = TotalSpend, color = ObjectCode, group = ObjectCode)) +
  geom_line(linewidth = 1) +
  scale_y_continuous(labels = label_number(scale = 1e-9, suffix = "B")) +
  labs(x = "Year",
    y = "Total spending",
    title = "Federal Spending by Object Code",
    color = "Object Code") +
  theme_minimal()

```