

# Story1

Saloua Daouki

2025-01-31

**Research Question:** Is there a correlation between a state's dependency on federal aid and its political affiliation?

## 1. Loading the data

```
## # A tibble: 6 x 7
##   Index State Political Affiliation~1 `Dependency Score` Return on Tax Dollar~2
##   <dbl> <chr>   <chr>                                <dbl>                                <dbl>
## 1     1 New Me~ Blue                                100                                3.42
## 2     2 West V~ Red                                94.6                              2.91
## 3     3 Alaska Red                                93.5                              2.65
## 4     4 Missis~ Red                                90.5                              2.66
## 5     5 Distri~ Blue                                87.7                              1.71
## 6     6 Alabama Red                                86.4                              1.9
## # i abbreviated names: 1: `Political Affiliation`, 2: `Return on Tax Dollars`
## # i 2 more variables: `Fed Fund % of State Revenue` <dbl>, `GDP in $M` <dbl>
```

The `federal_data` contains 51 observations and 7 variables, capturing the relationship between a state's dependency on federal aid and its political affiliation. The dataset includes information on each state's `Dependency Score`, `Return on Tax Dollars`, and `Federal Funds as a Percentage of State Revenue`, alongside its GDP. The `Political Affiliation` variable categorizes states as either 'Red' (Republican-leaning) or 'Blue' (Democratic-leaning). This dataset enables an analysis of whether political affiliation correlates with reliance on federal funding.

Before diving into visualization, let's further explore the dataset and check for any missing values.

```
## tibble [51 x 7] (S3: tbl_df/tbl/data.frame)
##  $ Index          : num [1:51] 1 2 3 4 5 6 7 8 9 10 ...
##  $ State          : chr [1:51] "New Mexico" "West Virginia" "Alaska" "Mississippi" ...
##  $ Political Affiliation : chr [1:51] "Blue" "Red" "Red" "Red" ...
##  $ Dependency Score   : num [1:51] 100 94.6 93.5 90.5 87.7 86.4 84.2 80.1 79.5 78.8 ...
##  $ Return on Tax Dollars : num [1:51] 3.42 2.91 2.65 2.66 1.71 1.9 1.68 1.62 1.43 1.78 ...
##  $ Fed Fund % of State Revenue: num [1:51] 0.307 0.27 0.29 0.259 0.322 0.267 0.301 0.285 0.318 0.233
##  $ GDP in $M        : num [1:51] 130202 99511 67337 146401 174796 ...

##      Index      State      Political Affiliation Dependency Score
##  Min.   : 1.0    Length:51      Length:51          Min.   : 0.00
##  1st Qu.:13.5    Class :character  Class :character    1st Qu.: 26.10
##  Median :26.0    Mode  :character  Mode  :character    Median : 44.30
##  Mean   :26.0
##  3rd Qu.:38.5
##  Max.   :51.0
##                                Mean   : 47.30
##                                3rd Qu.: 73.75
##                                Max.   :100.00
##  Return on Tax Dollars Fed Fund % of State Revenue  GDP in $M
##  Min.   :0.460      Min.   :0.1280      Min.   : 43130
##  1st Qu.:0.770      1st Qu.:0.1755      1st Qu.: 124496
```

```
## Median :0.920      Median :0.2100      Median : 309601
## Mean   :1.195      Mean   :0.2172      Mean   : 533556
## 3rd Qu.:1.550      3rd Qu.:0.2580      3rd Qu.: 683038
## Max.   :3.420      Max.   :0.3220      Max.   :3862171

##              Index              State
##              0                  0
## Political Affiliation      Dependency Score
##              0                  0
## Return on Tax Dollars Fed Fund % of State Revenue
##              0                  0
## GDP in $M
##              0
```

To ensure data quality before visualization, we examined its structure, summary statistics, and checked for missing values. The structure of the dataset confirms that it includes numerical and categorical variables, with State and Political Affiliation being character variables, while the remaining columns are numeric.

The summary statistics indicate a wide range of dependency scores, return on tax dollars, and GDP values, suggesting significant variation across states. Additionally, the missing values check confirms that the dataset is complete, with no missing values in any of the columns.

## 2. Analysis:

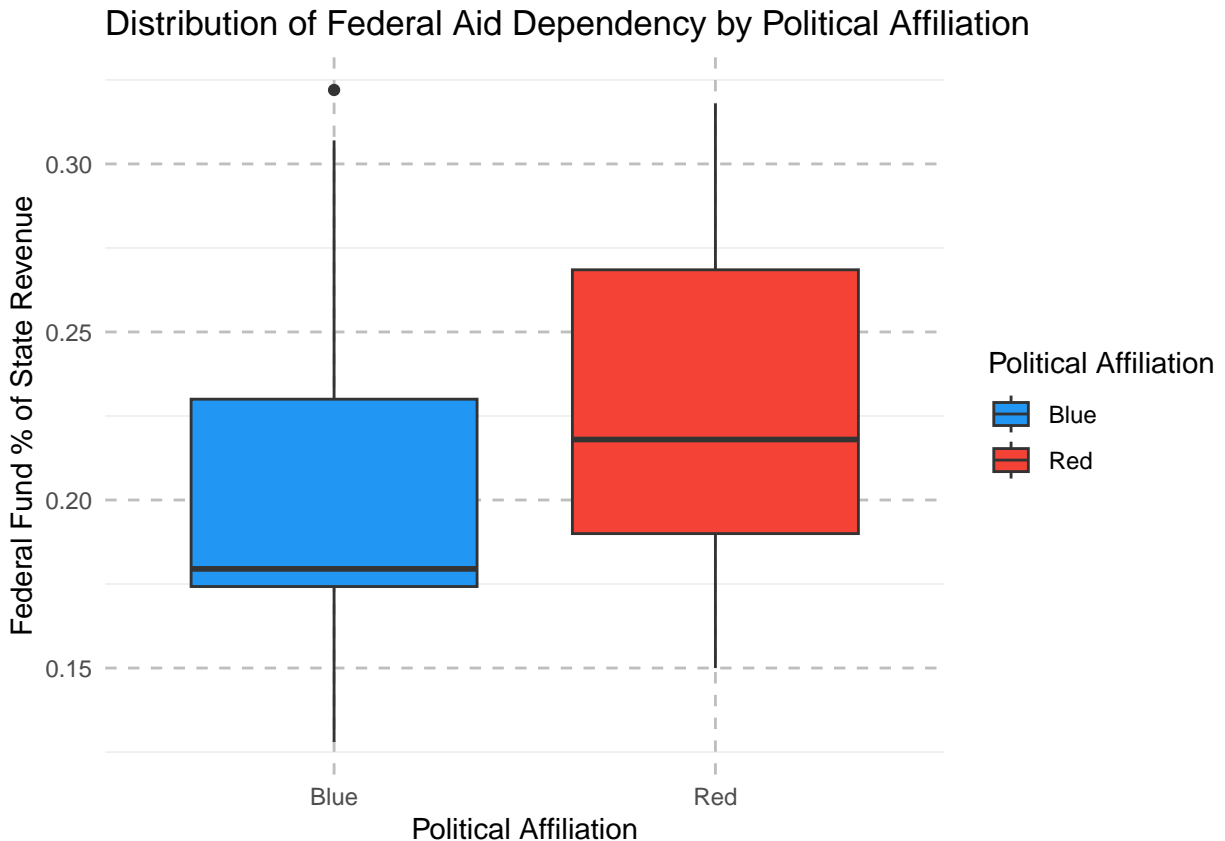
We already have the column Fed Fund % of State Revenue which represents the proportion of each state's revenue derived from federal aid. So we can calculate the average dependency by political affiliation using dplyr:

```
## # A tibble: 2 x 2
##   `Political Affiliation` Average_Fed_Fund_Percentage
##   <chr>                  <dbl>
## 1 Blue                  0.202
## 2 Red                   0.230
```

## 3. Including Plots

To better understand and effectively illustrates the relationship between federal aid dependency and political affiliation, I will create a box plot. The box plot clearly shows the distribution of federal aid dependency (or its related measure like the “Fed Fund % of State Revenue”) for both Democratic and Republican-leaning states.

```
## Warning: The `size` argument of `element_line()` is deprecated as of ggplot2 3.4.0.
## i Please use the `linewidth` argument instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```



It seems that there is an outlier in the blue boxplot below and it is hard to get the five-number summary precisely from the graph. I am going to calculate those numbers by political affiliation and check if the outlier is extreme:

```
## # A tibble: 2 x 7
##   `Political Affiliation` min_value    Q1 median_value    Q3 max_value    IQR
##   <chr>                  <dbl> <dbl>         <dbl> <dbl>    <dbl> <dbl>
## 1 Blue                  0.128 0.174         0.180 0.23     0.322 0.0558
## 2 Red                   0.15  0.19         0.218 0.268     0.318 0.0785
```

The median of the **Federal Fund % of State Revenue** for Blue states is 0.1795, while for Red states it is 0.2180. This suggests a correlation between political affiliation and federal aid dependency. The higher median for Red states indicates that, on average, Red states rely more on federal aid than Blue states.

The IQR for Blue states is 0.05575, while for Red states it's 0.0785. The larger IQR for Red states suggests that there is more variation or spread in the dependency percentages for Red states compared to Blue states. This helps explain that Red states not only tend to rely more on federal aid but also show greater variation in this reliance across the states.

```
## # A tibble: 1 x 7
##       Q1    Q3    IQR max_value upper_bound upper_bound_ext outlier_type
##   <dbl> <dbl> <dbl>    <dbl>    <dbl>         <dbl> <chr>
## 1 0.174 0.23 0.0558    0.322    0.314         0.397 Regular Outlier
```

Interestingly, the Blue states also have an extreme outlier, with a maximum value of 0.322, surpassing the upper bound of the interquartile range (0.313625). This anomaly suggests that while the trend is evident, outliers like this may warrant further investigation.

#### 4. Interpretations:

in the analysis of federal aid dependency across states, categorized by political affiliation (Red for Republican-leaning and Blue for Democratic-leaning), reveals several key insights:

- **Average Dependency:** Red states, on average, rely more on federal aid than Blue states. The median federal aid percentage for Red states is 21.8%, compared to 17.95% for Blue states.
- **Variation Across States:** Red states show more variation in federal aid reliance, indicated by a wider interquartile range (IQR). This suggests that Red states have a broader disparity in how much they depend on federal funds.
- **Outliers in Blue States:** An extreme outlier in a Blue state, with a dependency of 32.2%, stands out. This anomaly highlights the potential need for further investigation into state-specific conditions.

So, The data suggests that Red states may benefit from strategies to reduce federal aid dependency, while Blue states show more consistency in their reliance.

In summary, there appears to be a correlation between political affiliation and federal aid dependency, with Red states generally relying more on federal support. These findings offer valuable insights for addressing fiscal policies and potential disparities.