Story 1: Infrastructure Investment and Jobs Act Funding Allocation to Population

By Tony Fraser

2024-09-04

Introduction

After the Great Depression, Franklin D. Roosevelt, a New Yorker, created many jobs through public works programs such as the Works Progress Administration (WPA), the Civilian Conservation Corps (CCC), and the Public Works Administration (PWA). Around the same time, another New Yorker, Robert Moses, began rising in NYC politics. As more money became available through the New Deal, Moses expanded his team of architects, proposed more plans, and directed significant portions of New Deal funds to New York. Eventually, Robert Moses became one of the most prolific urban planners in history and one of the most powerful figures in NYC politics. In the end, close to 15% of all New Deal money was allocated to New York City and its surrounding infrastructure.

This our American history, and this assignment, beg a similar question about the Jobs Act allocation. Let's standardize this dataset and see if they align, or if we're looking at another New York City is better scenario.

Project Plan

1. Get data
    - Dataset 1: A list of states AND provinces
    - Dataset 2: IIJA funding (Note that Delaware is missing)
    - Dataset 3: US Census data (Plus, manually find records for missing territories)
2. Clean (column names, spaces, etc.)
3. Join states to funding to population
4. Build calculated columns for expected funding, funding difference, and percentage difference
5. Persist a clean spreadsheet

6. Chart in Excel, save screen shots into ./story1_files/
7. Do the assignment – add the two charts and write the conclusion
8. Quarto publish to build final story1.pdf
9. Push all to github

Data, steps 1 - 5

We'll use python for cleaning because it works with Quarto.

```
import pandas as pd
merged_fn = './story1_files/merged_states_funding_population.csv'

funding_df = pd.read_excel('./story1_files/IIJA FUNDING AS OF MARCH 2023.xlsx')
population_df = pd.read_csv('./story1_files/NST-EST2023-ALLDATA.csv')

# We'll use this to see what's missing or incorrect.
states_list = [
    'alabama', 'alaska', 'americansamoa', 'arizona', 'arkansas', 'california',
    'colorado', 'connecticut', 'delaware', 'districtofcolumbia', 'florida',
    'georgia', 'guam', 'hawaii', 'idaho', 'illinois', 'indiana', 'iowa', 'kansas',
    'kentucky', 'louisiana', 'maine', 'maryland', 'massachusetts', 'michigan',
    'minnesota', 'mississippi', 'missouri', 'montana', 'nebraska', 'nevada',
    'newhampshire', 'newjersey', 'newmexico', 'newyork', 'northcarolina', 'northdakota',
    'northernmarianaislands', 'ohio', 'oklahoma', 'oregon', 'pennsylvania', 'puertorico',
    'rhodeisland', 'southcarolina', 'southdakota', 'tennessee', 'texas', 'utah',
    'vermont', 'virginia', 'washington', 'westvirginia', 'wisconsin', 'wyoming'
]
states_df = pd.DataFrame(states_list, columns=['state'])

funding_data = (
    pd.read_excel('./story1_files/IIJA FUNDING AS OF MARCH 2023.xlsx')
    [['State, Teritory or Tribal Nation', 'Total (Billions)']]
    .rename(
        columns={'State, Teritory or Tribal Nation':
                 'state', 'Total (Billions)': 'actual'})
    .assign(state=lambda df: df['state'].str.lower().str.replace(' ', ''))
)
population_data = (
    population_df[['NAME', 'POPESTIMATE2023']]
    .rename(columns={'NAME': 'state', 'POPESTIMATE2023': 'pop_2023'})
    .assign(state=lambda df: df['state'].str.lower().str.replace(' ', '')))
```

```python
# step 3 - do the joins
merged_df = (
    states_df
    .merge(funding_data, on='state', how='left')
    .merge(population_data, on='state', how='left')
)

# step 4 - the calculated columns
total_population = merged_df['pop_2023'].sum()
merged_df = (
    merged_df
    .assign(
        expected=lambda df: (df['pop_2023'] / total_population) * df['actual'].sum(),
        diff=lambda df: df['actual'] - df['expected'],
        percent_diff=lambda df: (df['diff'] / df['expected']) * 100
    )
    .round({'expected': 2, 'diff': 2, 'percent_diff': 2})
)
# step 5 - persist the CSV
merged_df.to_csv(merged_fn, index=False)

# data looks like this.
# state,    actual, pop_2023, expected, diff,  percent_diff
# alabama,  3.0,    5108468,  2.9,      0.1,   3.45
# alaska,   3.7,    733406,   0.42,     3.28,  788.71
# arizona,  3.5,    7431344,  4.22,     -0.72, -17.03
# arkansas, 2.8,    3067732,  1.74,     1.06,  60.78
```

All above work is only in the spirit of "showing your work." The actual assignment, the two images and the conclusion, will start on the next page.

Image one: Expected disbursements versus actual disbursements

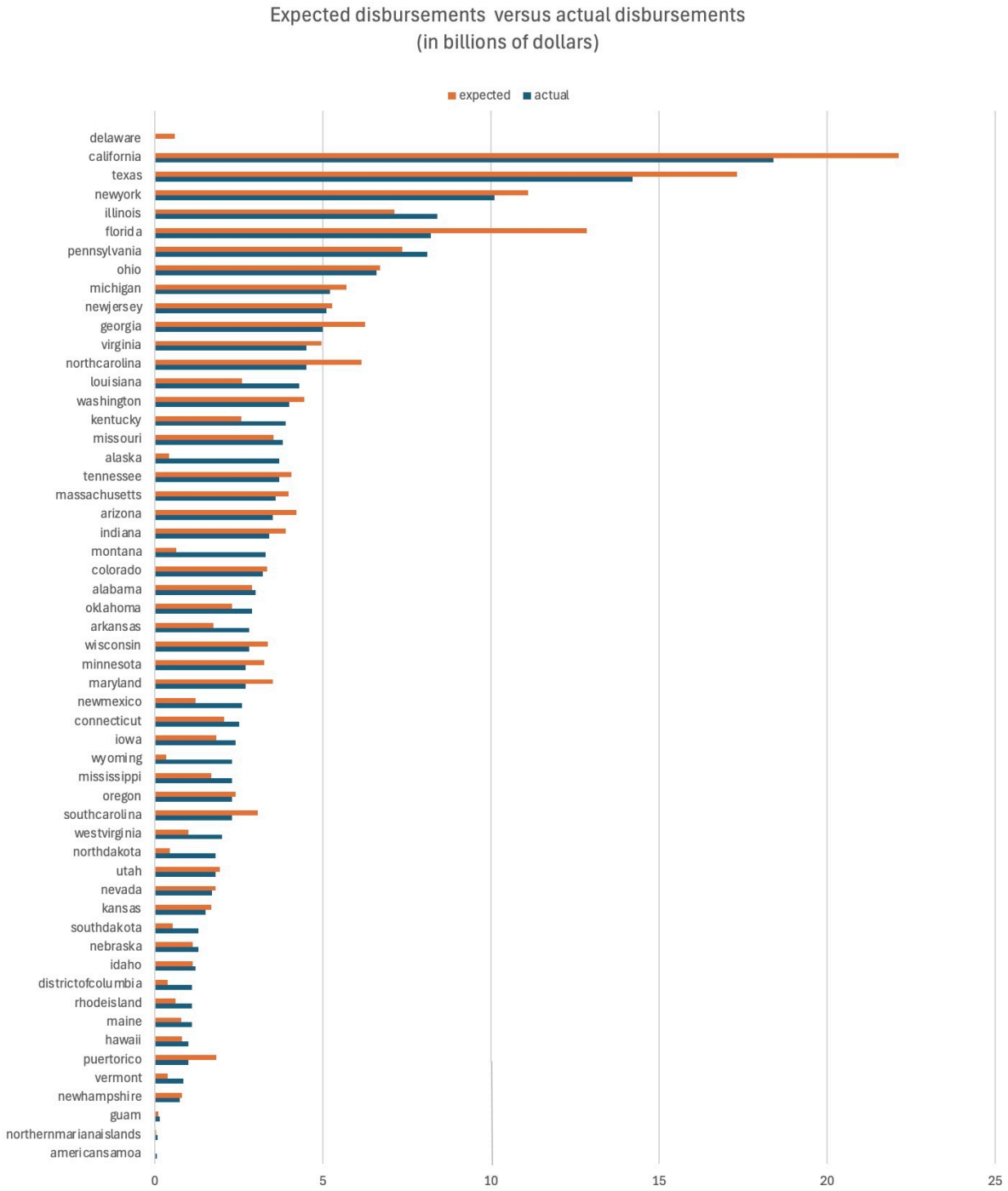
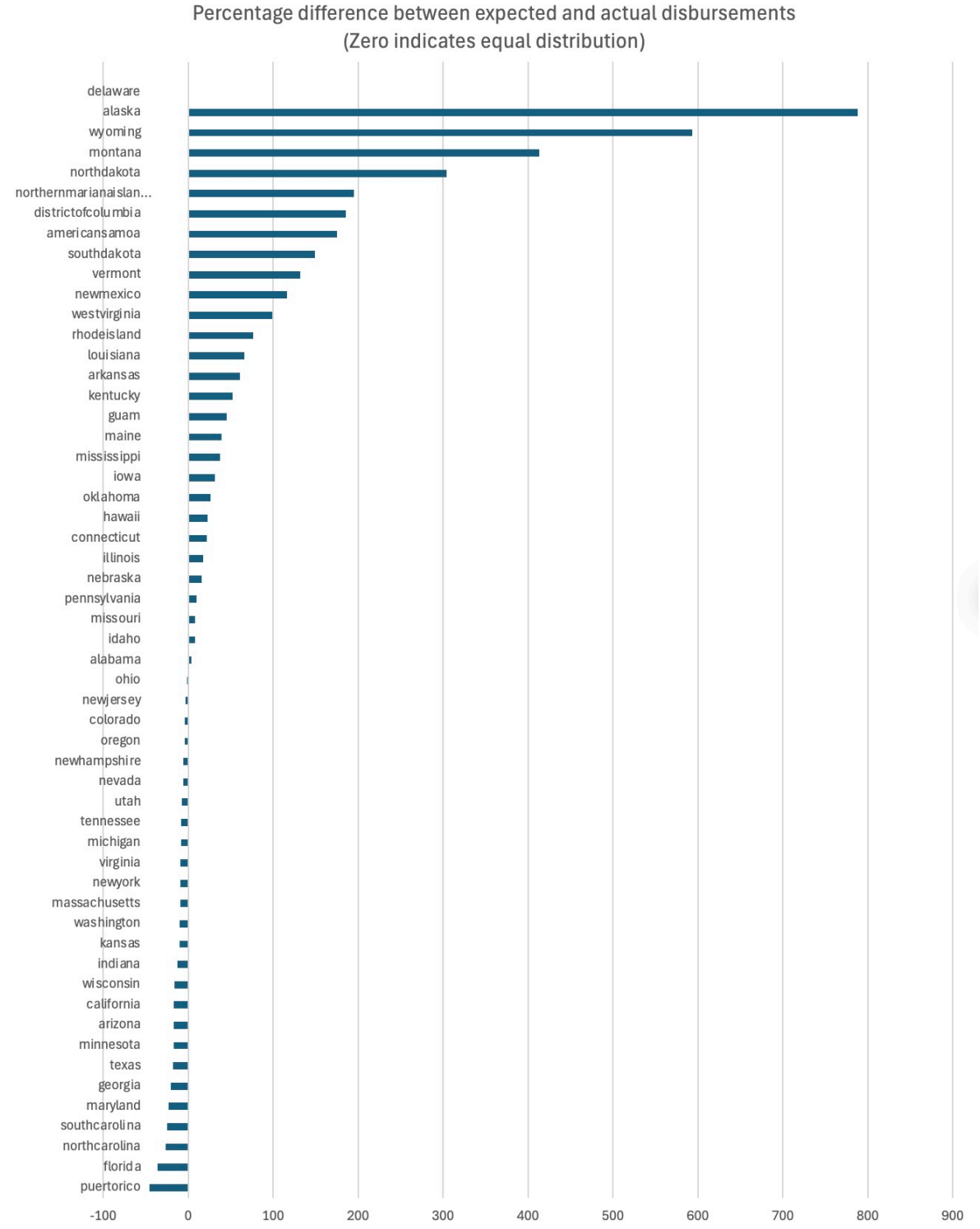


Image two: Percentage difference between expected and actual disbursements



Conclusion

Neither of these charts demonstrates a strictly proportional relationship between state population size and money distributed for IIJA projects. For example, Alaska, with a relatively small population, receives significantly more funding than expected, potentially for reasons like infrastructure and energy projects that could impact gas prices. Several traditionally Republican-leaning states are at the top of the spending list, while Democratic-leaning states like New York and California show slight underspending relative to population. Although the reasons for this distribution are unclear without further review of individual projects, the current pattern could suggest decisions aligned with specific political or economic strategies of the Democratic administration.