

‘It’s the Economy, Stupid!’

Joshua Rogin

github:

github.com/WhereisWald0

QUESTION:

Can we look at economic data and election outcomes and, by applying Machine Learning models, account for any relationship between them?

Pre-methodolgical principles

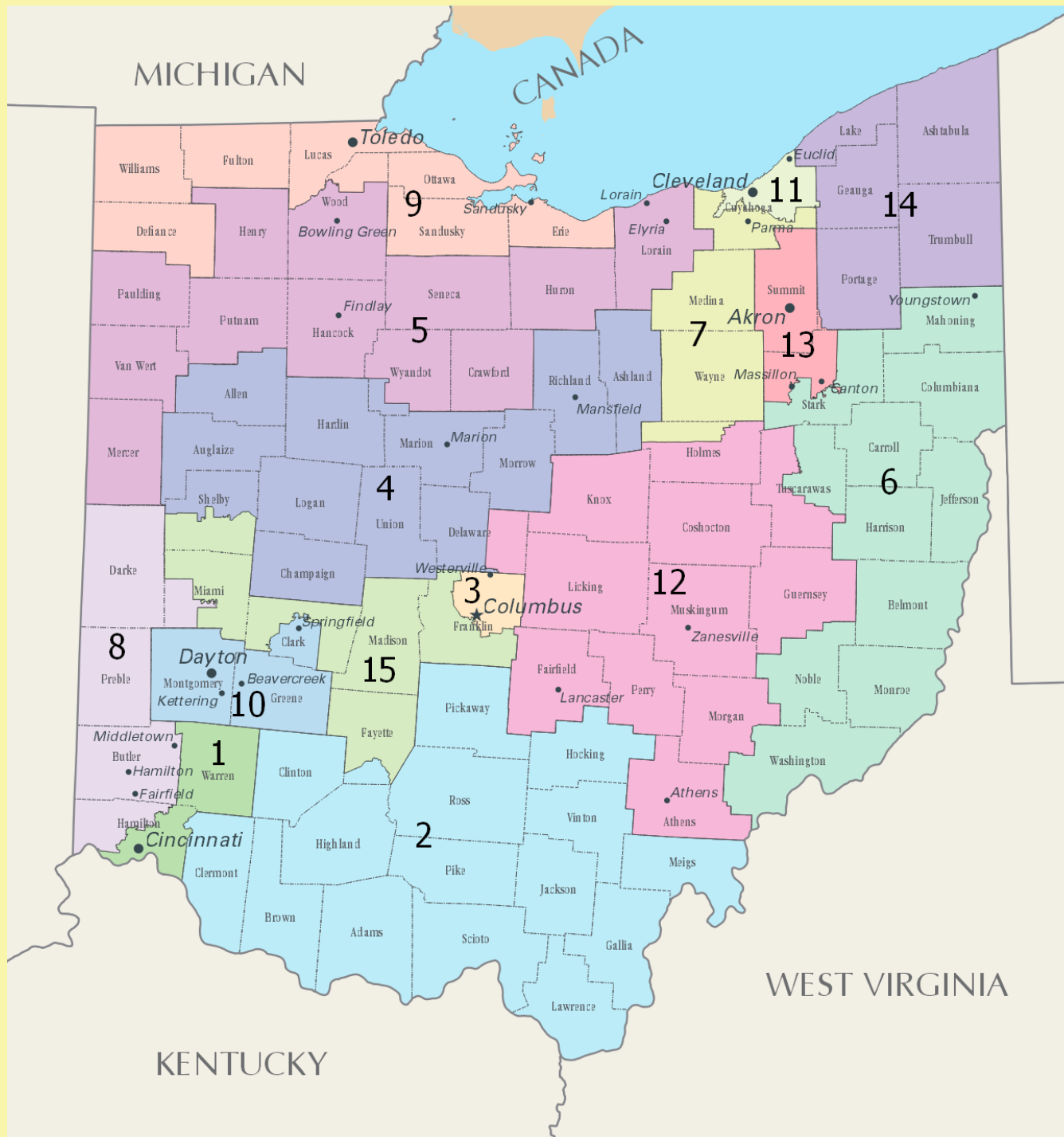
Independence - use primary sources as much as possible

Modularity - formatted data should be easily swapped or connected to new sources

Preservation - avoidance of destructive processes

Granularity - use boundaries of appropriate size.

Future-Proofing -all processes should allow for future data

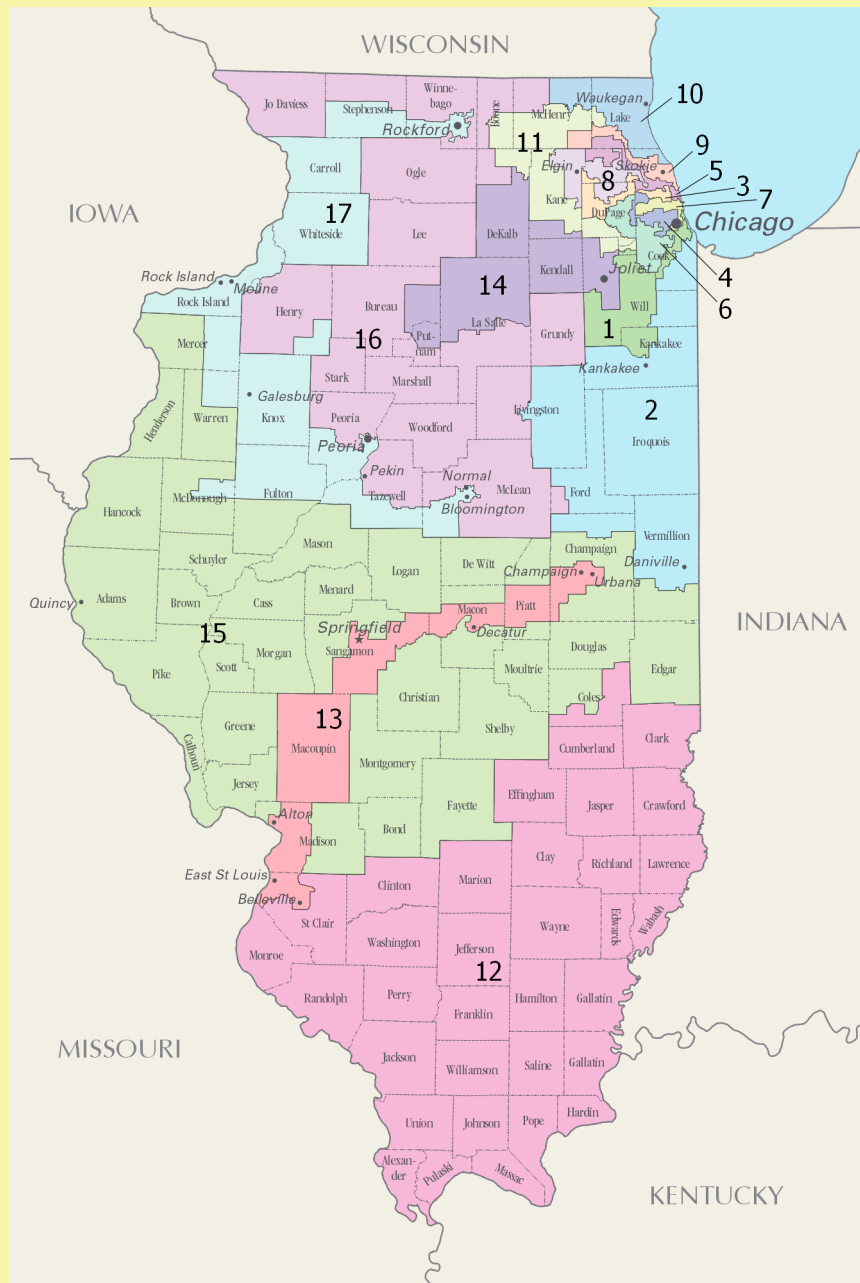


United States House Districts

-435 seats

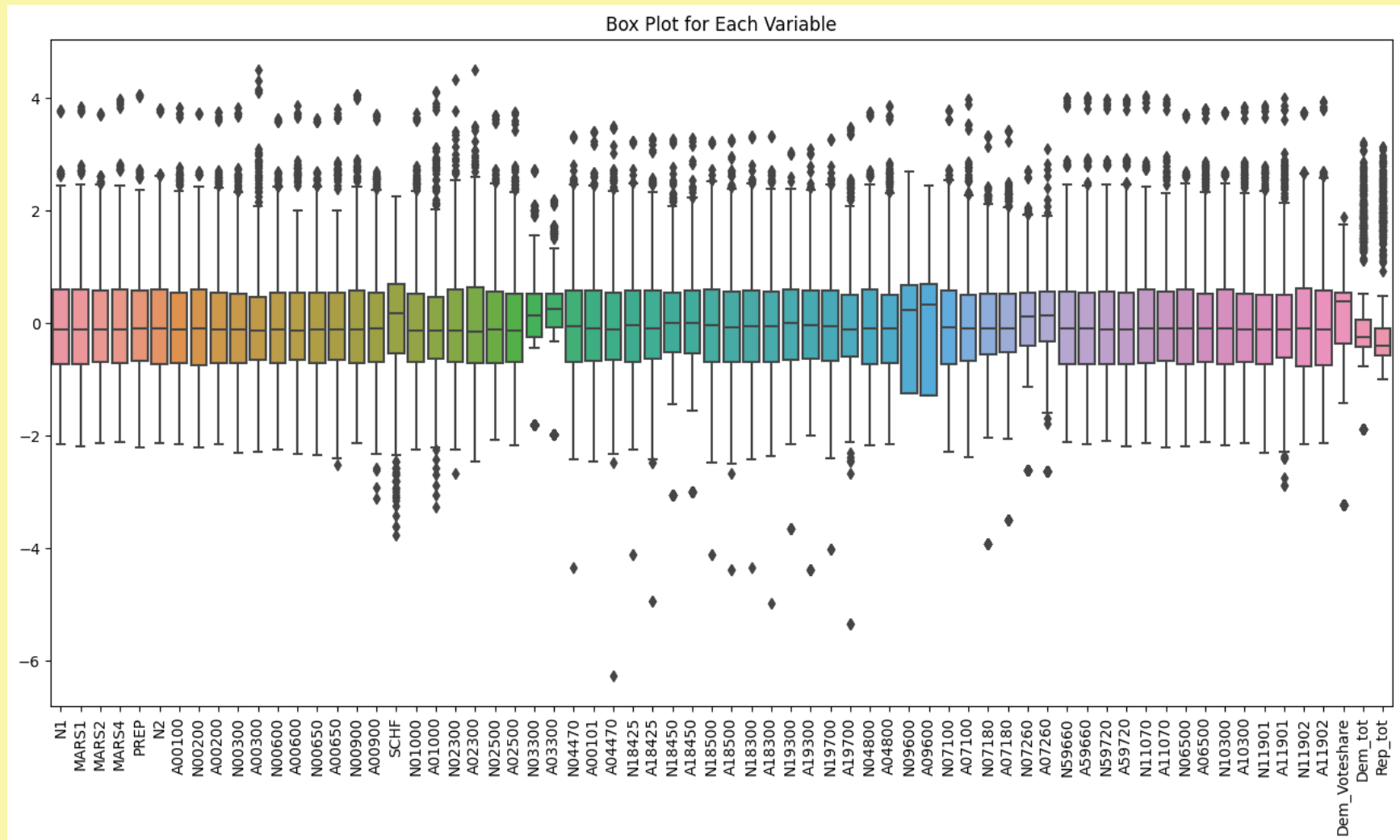
-Elections every two years

-Numerous candidates

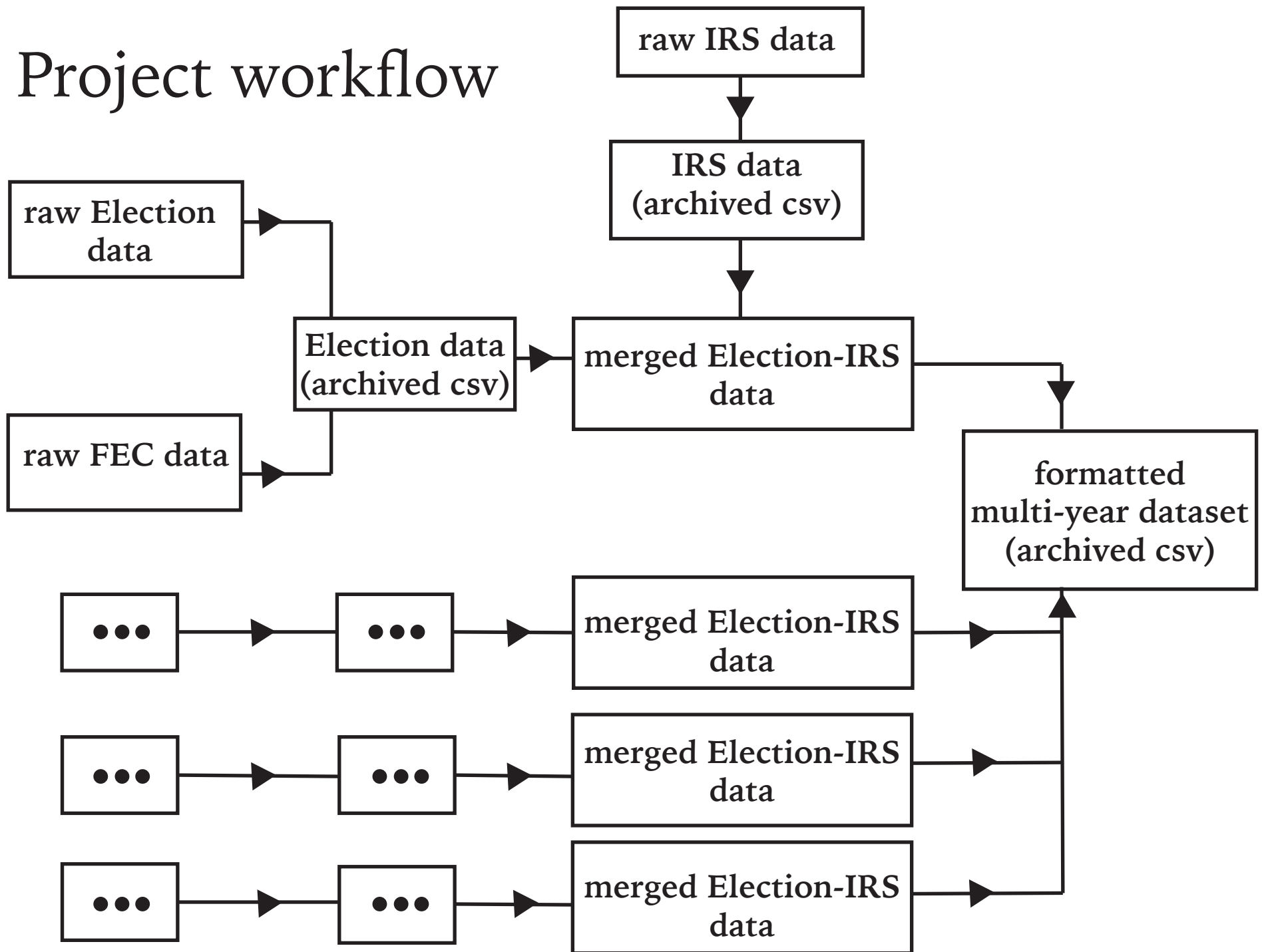


IRS data

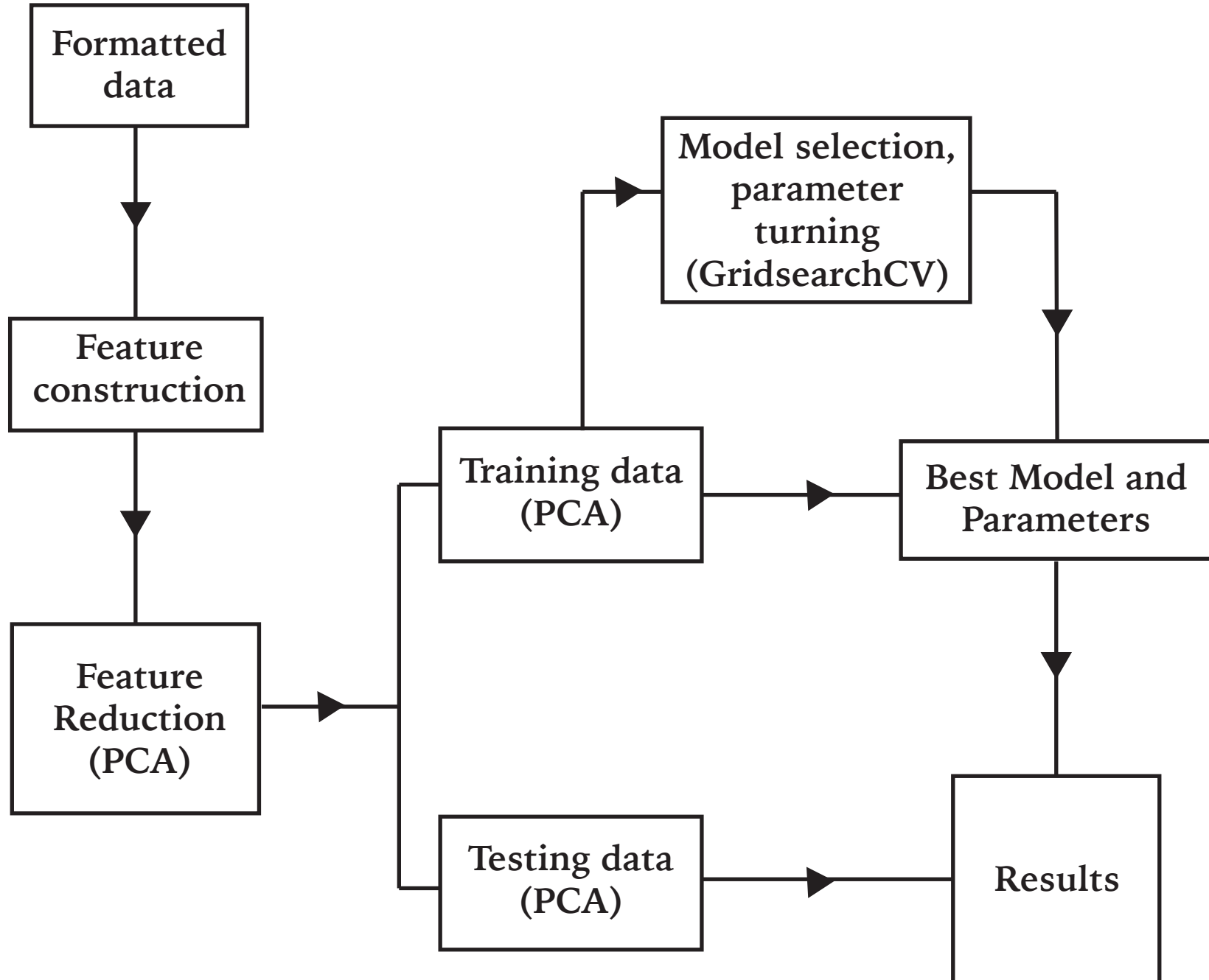
- data shape - 3155 rows, 150 columns (dense)
- lots of outliers (reflects economic reality w/in US)



Project workflow



Machine Learning Pipeline



Final Results

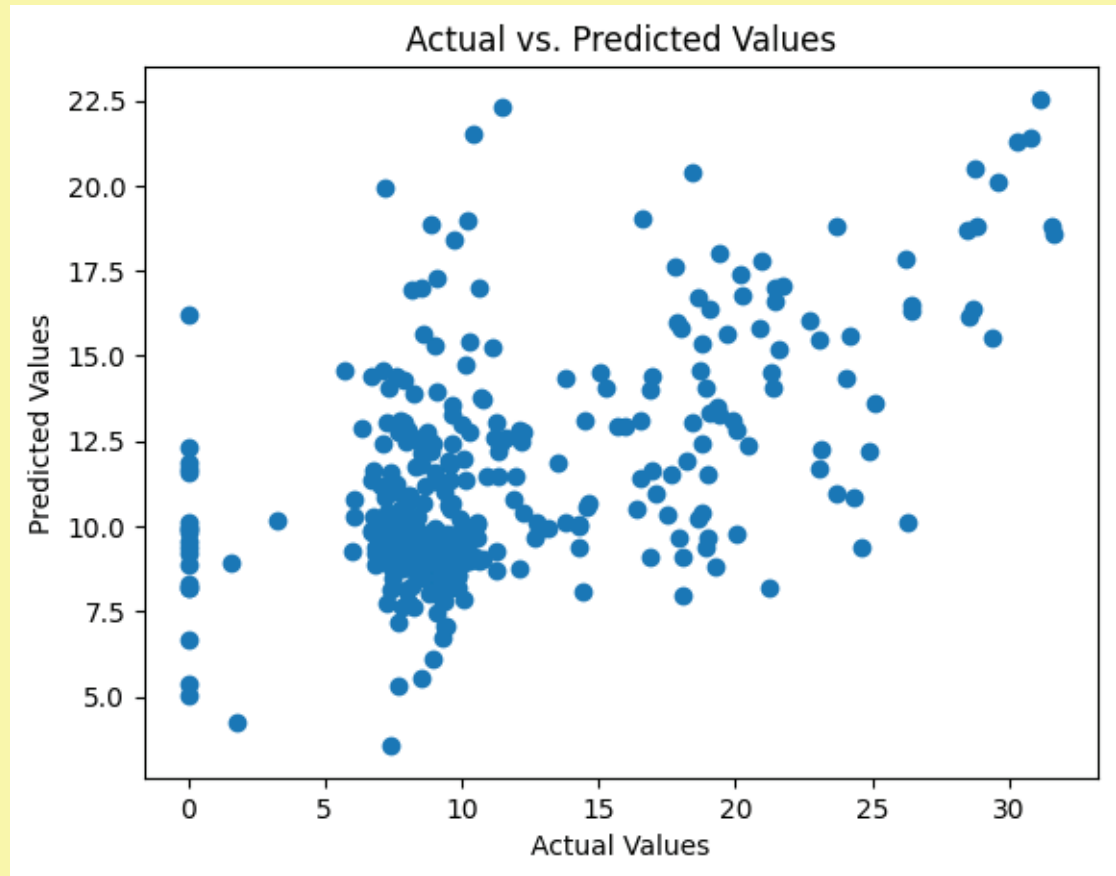
After Grid-
searchCV:

best model - RandomForest

Gives Us:

Mean Squared Error: 26.77

R-squared: 0.319



concl./ thank you

Next Steps:

- Extract and process additional States data
- Divide IRS data by income tax brackets (could potentially add 7x as many features)