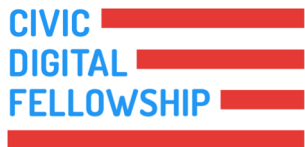# REVENUE-BASED COUNTY BUSINESS PATTERNS

Economy-Wide Statistics Division
Supervised by Javier Miranda and Anne Russell

ANDY GREEN
Georgetown University
Data Science / Public Policy

DRISHAAN JAIN
University of Michigan
Computer Science

CIVIC DIGITAL FELLOWSHIP

United States Census Bureau

PROJECT GOAL AND BENEFITS

- Background on CBP
    - County Business Patterns (CBP) is an annual series that provides subnational economic data by industry
- Goal
    - Create a proof-of-concept for an extension of CBP, by adding revenue data to the existing product
- Benefits
    - Census Bureau
    - Other government agencies
   - Private businesses, researchers

CHALLENGES

- Missing data
    - SU: 1.1M / 5.8M establishments (20%)
    - MU: 53K / 191K firms (27%)
- Unreliable data / outliers
- Structural differences between revenue measures across data sources for certain industries

APPROACH

- Outlier detection
  - Revenue quality flags
  - Additional outlier detection rules
- Imputation algorithms
  - ASE-based approach
  - Regression model
- Industry deep-dive comparison of Economic Census vs. Administrative revenue
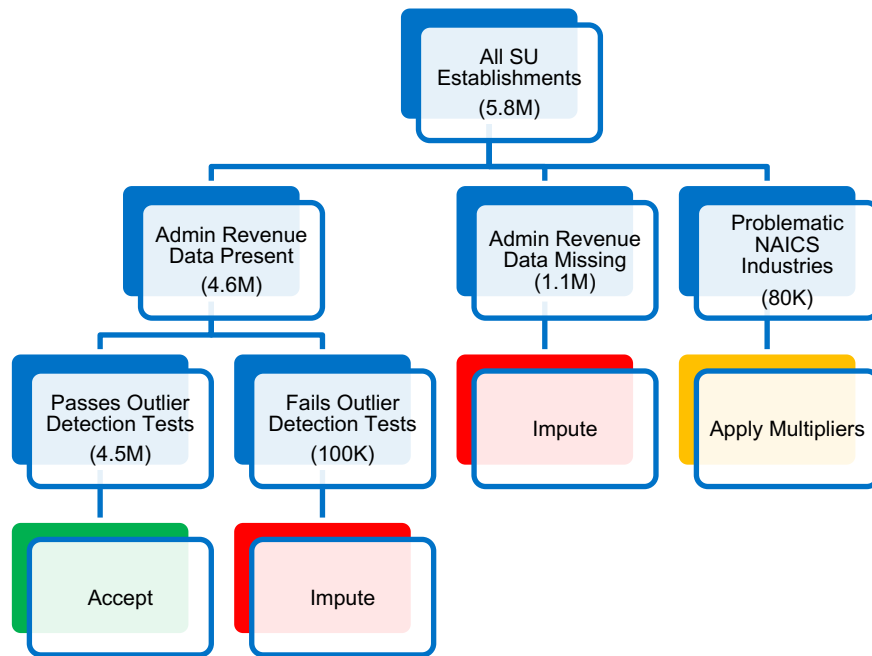
RESULTS

Comparison of Administrative Revenue Data vs.

Economic Census Revenue Data

	Before	After
Total Revenue % Difference (Admin vs. EC)	40%	1.5%
Average Establishment-Level $ Difference (Admin vs. EC)	$681K	$25K

NEXT STEPS + IMPROVEMENTS

Multi-Unit Establishments

- Disaggregate revenue from firm to establishment level
- Refine revenue for large firms with unlinked variables
- Create aggregate firm value tabulations

Industry Adjustment

- Create MU adjustment ratios based on existing framework
- Generate specific adjustment ratios based on establishment size

Overall Improvements

- Implement more advanced imputation approaches: regression, CART models
- Special processing/refinement for certain industries and geographies