ANOMOLOUS PRESCRIBING DETECTION

W205 SPRING 2016 PROJECT

D. M. Burt | JUN LUO

RESEARCH PROBLEM

Prescription drugs accounted for about **9 percent** of national health expenditure in the U.S. in 2013.

A major concern is that some prescribing patterns in health care lead to **unnecessary cost** and **health outcomes burdens**.

SOLUTION

CMS has recently **changed its position** for release of physician-level prescribing data for public use.

Using this **physician-level prescription information** in concert with the **NPPES Physician Registry,** we are finding unexpected prescribing patterns among physicians, based on prescriptions and costs per person.

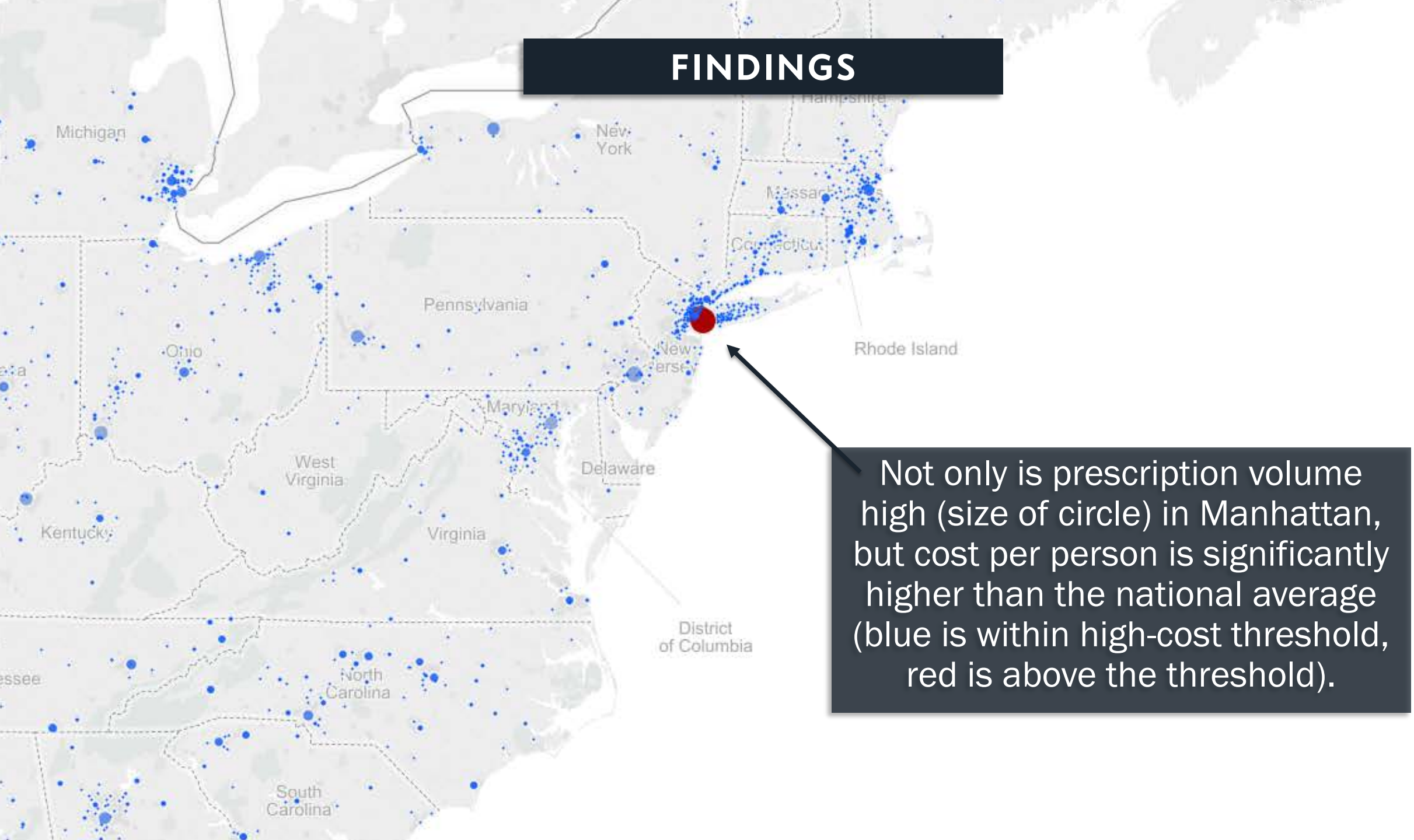
DATA SETS

- **CMS Part D Prescriber PUF, 2013**
  - Prescriptions, units, days supply, and costs by physician and drug
  - 2.7GB; 23M lines
- **CMS Part D Prescriber National Summary**
  - Prescriptions, units, and costs by drug aggregated to a national level
  - 3K lines
- **National Plan and Provider Enumeration System (NPPES)**
  - Office location and specialty (credential) by physician
  - 5.7GB; 4.8M lines

ETL / CLEANUP

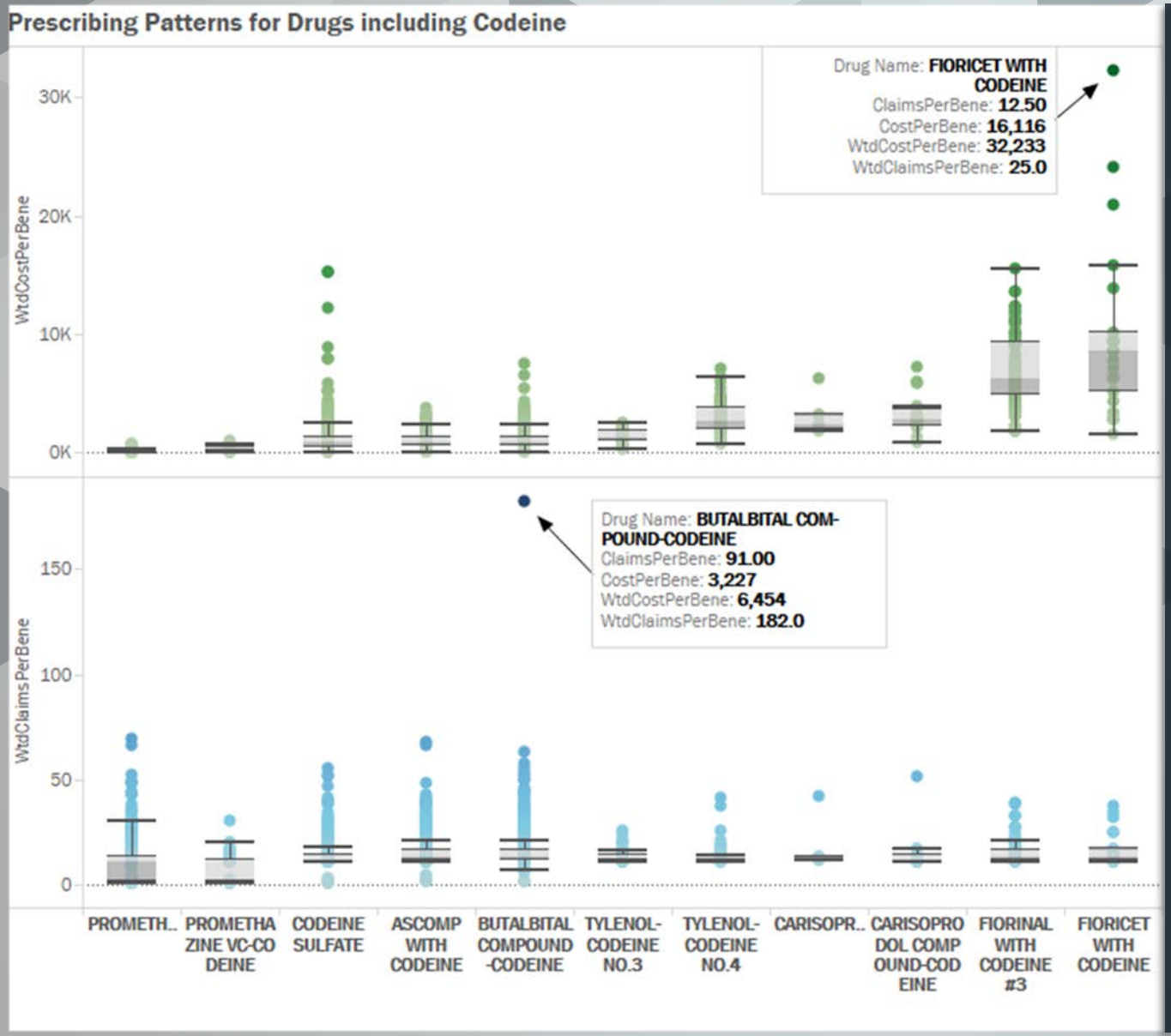
- **CMS Part D Prescriber PUF, 2013**
  - Missing values: requires imputation
  - Non-numeric values in numeric fields
- **CMS Part D Prescriber National Summary**
  - Excel: requires conversion to CSV
- **National Plan and Provider Enumeration System (NPPES)**
  - Over 200 fields: requires extraction of key attributes

FINDINGS



Not only is prescription volume high (size of circle) in Manhattan, but cost per person is significantly higher than the national average (blue is within high-cost threshold, red is above the threshold).

FINDINGS



Each dot is a physician in this data set.

Some physicians' prescriptions have a high cost per person (but few people).

Others have a large number of prescriptions per person (but not necessarily higher cost).

FINDINGS

This is an aggregate view of specialties for Oxycontin prescriptions.

High and left: many prescriptions to few people (per doc).

Low and right: few prescriptions to many people (per doc).

Colored red: very high cost per person.

A handful of medical examiners and radiologists are prescribing to relatively many people.

ROAD MAP

- **Larger data volume and velocity**
  - Implement system as a stream reader.
  - Recompute specialty aggregations as data cubes.

- **Aggregate drugs to therapeutic classes**
  - The drugs in the data set are very specific.
  - Grouping similar drugs together could help establish patterns.

APPENDIX

Using a graph database, we could find relationships between drugs if we connect them by the diseases/symptoms they treat.