

Making Policy with Data

An Introductory Course on Policy Evaluation

Policy Briefing

Instructor: Prof Yiqing Xu
April 11

New Segment: **Policy Briefing**

Today's Topic: Stop and Frisk



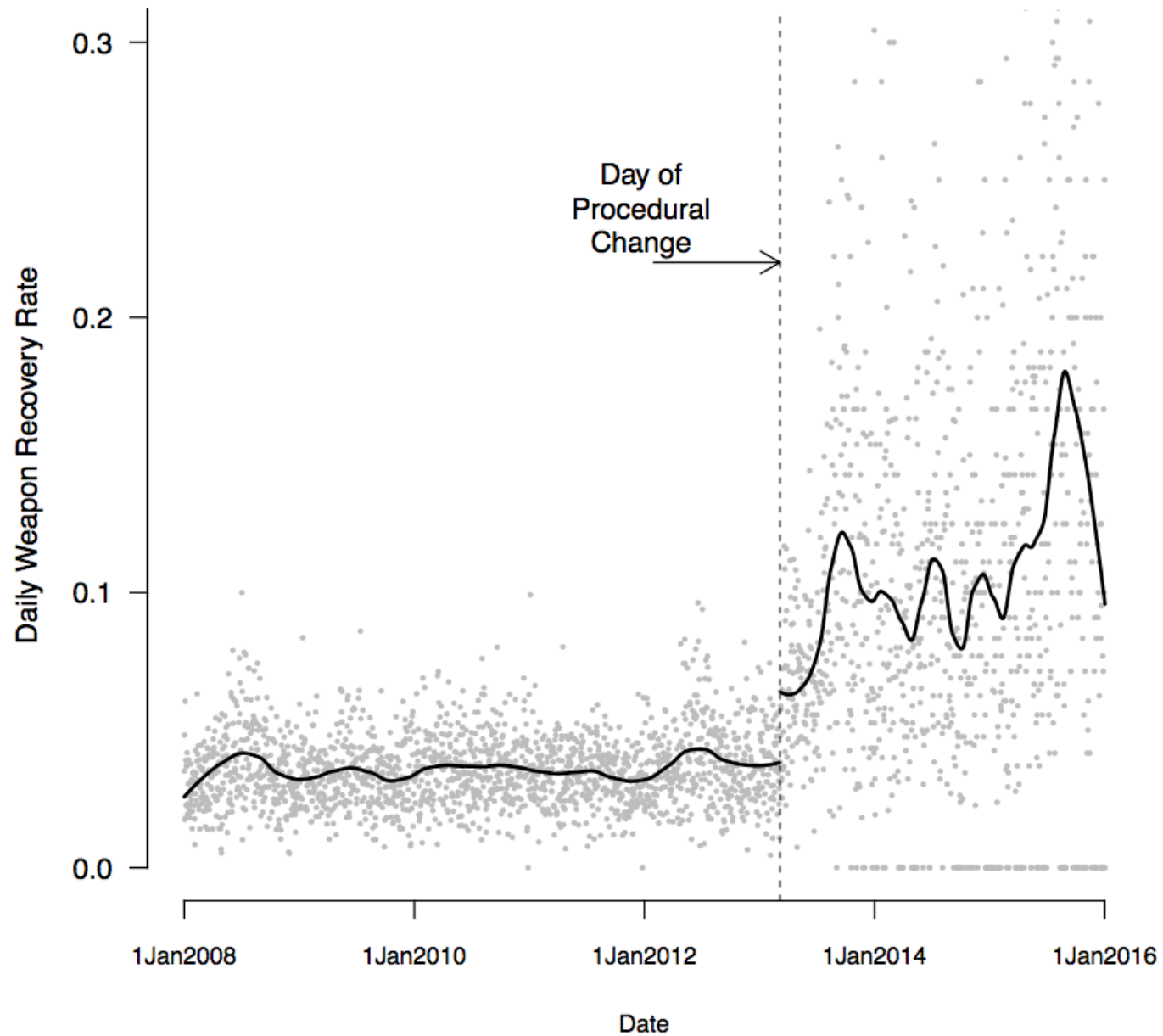
Controversies around Stop-and-Frisk

- Not working
- Unproportionally targeting minorities
- Resulting in police brutality
- More ...

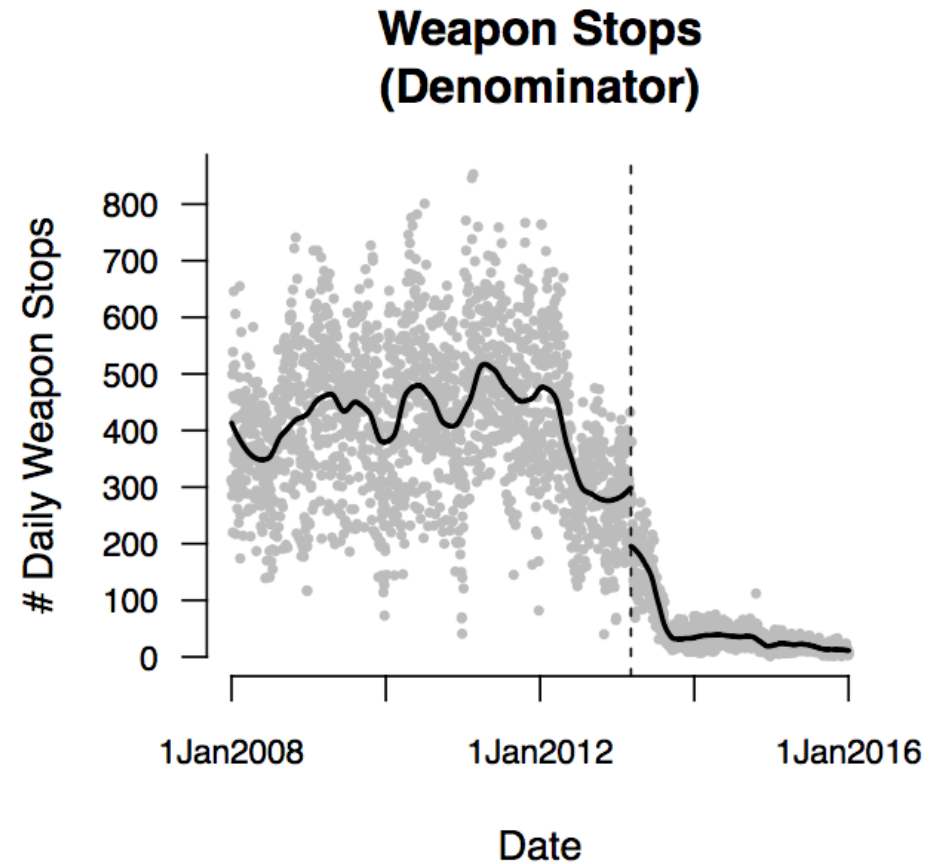
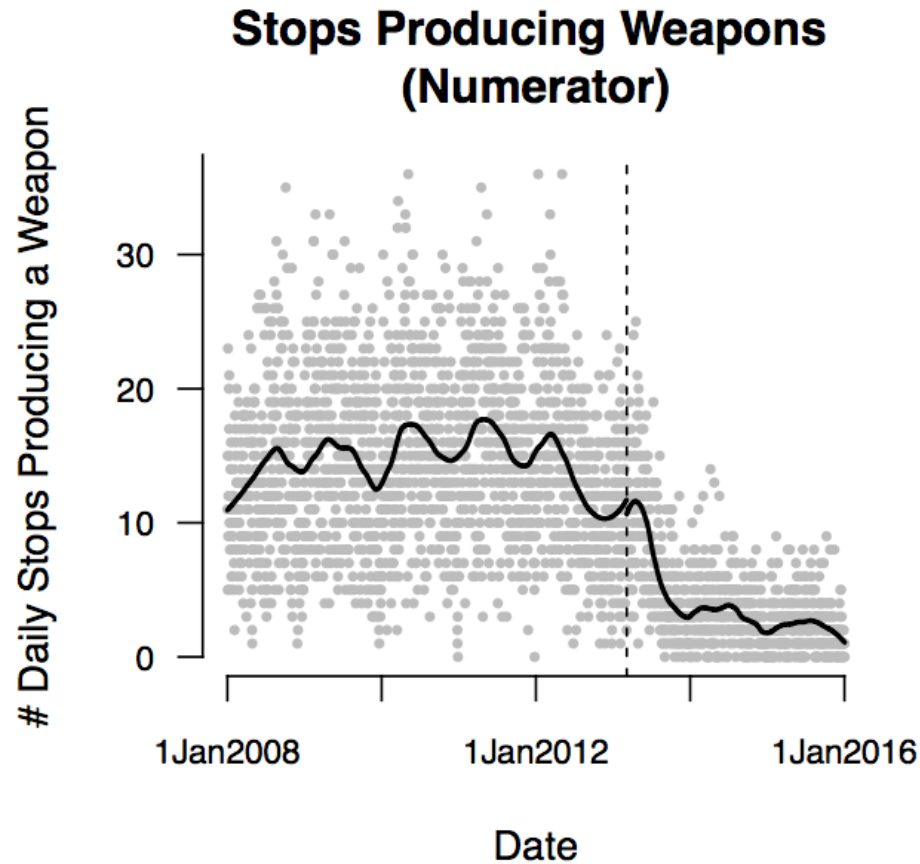
Background

- **Question:**
Can protocols improve stop-and-frisk practices?
- March 5, 2013, NYPD Chief of Patrol James P. Hall mandated all patrol units to **photocopy** and **submit narrative descriptions of the reasons** they stopped suspects to supervisors after each shift
- Research by Jonathan Mummolo based on over 3 million records of police citizen interactions (*Journal of Politics* forthcoming)

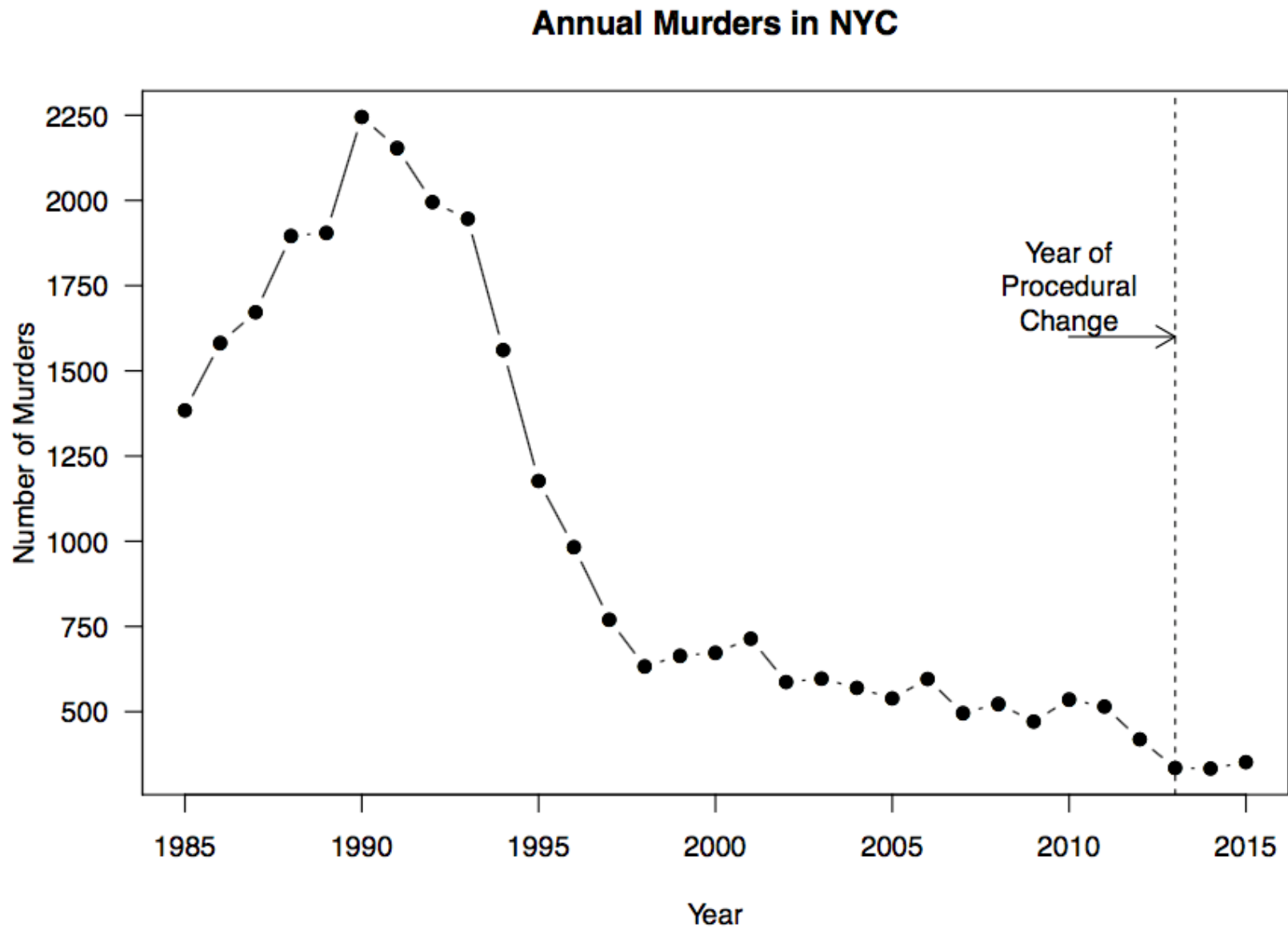
Hit Rate Increases Markedly When Procedural Reform Introduced



Why? Denominator drops faster than numerator!



Murder rate keeps declining



New Segment: **Buzzword**

Big Data

“Big Data”



Dan Ariely ✓

January 6, 2013 · 🌐

 Follow



Big data is like : everyone talks about it, nobody really knows how to do it, everyone thinks everyone else is doing it, so everyone claims they are doing it...

What's Not Big Data?

How big is big?

- GDP growth rate of all countries in the world?
 - Information of all NBA players?
 - All English Wikipedia pages?
-
- We are not dealing with Big Data in this course at all.
 - Many successful data science applications do not involve big data.
 - These days your laptop can easily deal with millions of observations if the tasks are simple.

So What Is Big Data?

Big Data is a set of approaches to store, structure, and process massive amount of data

It usually involves the following techniques

- Data collection (e.g. fast web scraping)
- Data storage (often uses relational databases, e.g., MySQL, MongoDB)
- Data processing (mostly, **prediction**, often requires fast algorithms to deal with high-dimensional data)

The key is whether a method is **scalable**.

Housekeeping

- Charlie's office hour
- Teammate
- Reading: Angrist and Pischke, Chapter 1. pp. 1-33
(Quiz: **next Thursday**)
- Homework (due **next Tuesday**)
 - R workflow
 - The turnout problem

R Workflow

- Coding in R
 - Set working directory
 - Load libraries, if necessary
 - Load data
 - Work on the problem
 - (Optional) Save results or workspace
- Write-up
 - Insert the R code into a RMarkdown file
 - Write comments and/or explanations
 - Knitr to DOC/PDF/HTML

(Check out [R Markdwon Cheatsheet](#))

R is Object-based

- Character, integer, numeric, logical; vector
- Matrix
- **Data frame**
(with columns as “variables” and rows as observations)
- List
- Function

Turnout Problem

- Turnout: how many people show up in an election

Turnout = Number of people who voted / All voters

- VAP vs VEP
 - VAP: voting age population
 - VEP: voting eligible population
- How come the data discrepancy?
- *demo*