

The Association Between Felonies in NYC and Weather and Temporal Conditions

Brian Weinstein

Columbia University

STAT W4201: Advanced Data Analysis

May 2, 2016

Abstract

Background: The New York City Police Department (NYPD) recently released incident-level felony data to the New York City Open Data portal. The dataset includes timestamped information for all felonies committed in NYC.

We first examine the association between the daily number of felonies committed in NYC in 2015 and temperature, presence of precipitation, day of week, federal and New York (NY) holidays, and school days. Second, we examine the association between large increases in temperature ($> 8^{\circ}\text{F}$ from the previous day) and increases in the number of felonies.

Methods and Results: We initially test for a difference between the number of felonies on warmer and cooler days ($\geq 51.98^{\circ}\text{F}$ and $< 51.98^{\circ}\text{F}$, respectively — the NYC 2015 median), finding overwhelming evidence of a difference, with there being 44 more felonies, on average, on warmer days than on cooler day (95% CI 38 to 51 felonies; two sided p-value < 0.000001 from a two-sample t-test).

After accounting for presence of precipitation, holidays, school days, and day of week, the data provides overwhelming evidence that for every 1°F increase in temperature there are, on average, 1.4 additional felonies per day (95% CI 1.3 to 1.6 felonies; two-sided p-value < 0.000001 for a test that the linear regression coefficient is 0).

We next find that, after accounting for day of week, there is little evidence to suggest that increases in felonies from the previous day are associated with large increases in temperature (two-sided p-value 0.1001 for a test that the “difference in differences” linear regression coefficient is 0).

Conclusions: There is a clear association between warmer temperatures and an increased number of felonies. There is no association between increases in the number of felonies and large increases in temperature.

1 Introduction

1.1 describing the problem of interest

Questions of interest:

- asdf

1.2 the data set

1.2.1 Data sources

- NYPD 7 Major Felony Incidents: data.cityofnewyork.us/d/hyij-8hr7
- NYPD Motor Vehicle Collisions: data.cityofnewyork.us/d/h9gi-nx95
- National Centers for Environmental Information (weather conditions and temperature data): ncdc.noaa.gov/cdo-web/search
- New York State Holidays: cs.ny.gov/attendance/leave/2015_legal_holidays.cfm
- Federal Holidays: opm.gov/policy-data-oversight/snow-dismissal-procedures/federal-holidays/#url=2015
- School Attendance: github.com/ajschumacher/NYCattdends/tree/master/xml (The NYC Department of Education also publishes this data to the NYC Open Data portal, but historical data is only retained there for the current school year. Instead Ill scrape the daily XML snapshots stored in the linked Github repository.)

1.2.2 Data cleaning

1.2.3 EDA

1.3 the organization of the entire report

2 statistical model and statistical analysis

2.1 model setup: the statistical model employed and assumptions

2.2 statistical analysis / inference

3 model checking (and model improvement if the originally proposed model is not appropriate)

4 Conclusion