Final Project Title

Brian Weinstein

Columbia University
STAT W4201: Advanced Data Analysis

May 2, 2016

Abstract

An abstract briefly summarizing your work.

1 Introduction

1.1 describing the problem of interest

Questions of interest:

- are the number of felonies associated with temp? (either min or max temp)
- is felonies associated with presence of precipitation?
- does the association between precip and felonies depend on temperature?
- is felonies associated with windspeed?
- does the association between windspeed and felonies depend on temperature?
- after taking temp into account, is felonies associated with accidents?
- is felonies associated with holidays?
- is felonies associated with school days?
- is felonies associated with day of week?

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/NYCattends/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