# The Association Between Felonies in NYC and Weather and Temporal Conditions

Brian Weinstein

April 20, 2016

STAT W4201 001
Columbia University, Spring 2016

https://github.com/BrianWeinstein/nyc-felonies

# Introduction

NYPD released incident-level felony data to the NYC Open Data portal.

Data is timestamped for every burglary, felony assault, grand larceny, grand larceny of motor vehicle, murder and non-negligent manslaughter, rape, and robbery.

- Examine the association between the daily number of felonies committed in NYC and temperature, precipitation, day of week, federal and NY holidays, and school days.
- Examine whether spikes in temperature could help explain spikes in felonies.

# Questions of Interest

Are felonies associated with temperature?

After taking temperature into account, is felonies associated with (1) precipitation, (2) school days, (3) holidays, and (4) day of week?

Although there’s no causal relationship, for a given set of these conditions, how many felonies can NYPD reasonably expect?

Are large increases in temperature (from the previous day) associated with an increase in felonies?

# Observational Study

Observational data ⇒ no causal interpretation

Observations weren’t randomly sampled, so generalization of these results to other years or cities is speculative.

# Dataset

| date | felonies | day_of_week | temp_min_degF | any_precip | is_warm | is_holiday | is_school_day |
|---|---|---|---|---|---|---|---|
| 2015-01-01 | 389 | 5 | 27.14 | 0 | 0 | 1 | 0 |
| 2015-01-02 | 257 | 6 | 35.06 | 0 | 0 | 0 | 0 |
| 2015-01-03 | 263 | 7 | 33.08 | 1 | 0 | 0 | 0 |
| 2015-01-04 | 227 | 1 | 41 | 1 | 0 | 0 | 0 |
| 2015-01-05 | 275 | 2 | 21.2 | 0 | 0 | 0 | 1 |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 2015-07-04 | 262 | 7 | 69.08 | 0 | 1 | 1 | 0 |
| 2015-07-05 | 291 | 1 | 68 | 0 | 1 | 0 | 0 |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 2015-12-31 | 209 | 5 | 42.08 | 1 | 0 | 0 | 0 |

# Dataset Schema

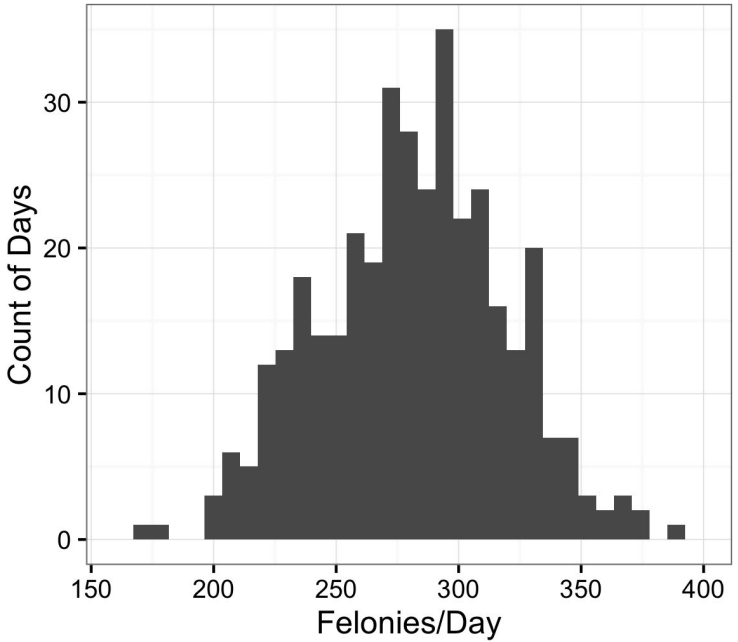
| | **date** | **felonies** | **day_of_week** | **temp_min_degF** | **any_precip** | **is_warm** | **is_holiday** | **is_school_day** |
|---|---|---|---|---|---|---|---|---|
| **Class** | Date | Integer | Factor | Numeric | Factor | Factor | Factor | Factor |
| **Description** | Date | Count of felonies | 1=Sunday, …, 7=Saturday | Minimum temperature (° F) | 1: > 0 inches of precipitation | 1: temp_min_degF ≥ 51.98 °F (median for 2015) | 1: Federal or NY Holiday | 1: Public schools are open / in session |
| **Source** | | NYC Open Data Portal: NYPD 7 Major Felony Incidents | | National Centers for Environmental Information | National Centers for Environmental Information | National Centers for Environmental Information | Federal Office of Personnel Management; New York State Department of Civil Service | github.com/ajschumacher/NYCattends/tree/master/xml |

# Assumptions (1)

Felonies are independent Bernoulli events

The sum is a Binomial random variable

Approximate this with a Normal random variable



# Assumptions (2)

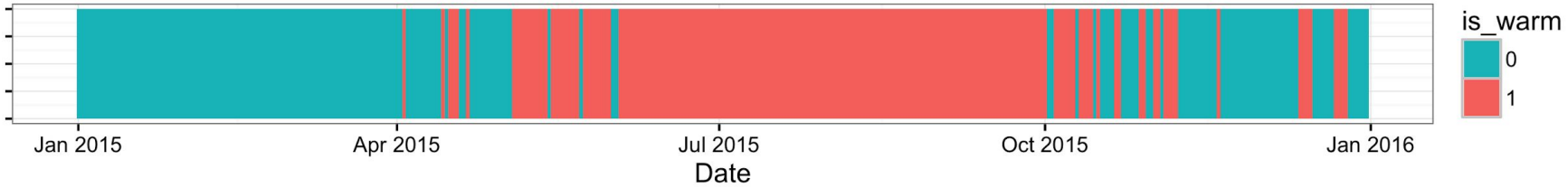
**Normality:** *felonies* is normally distributed

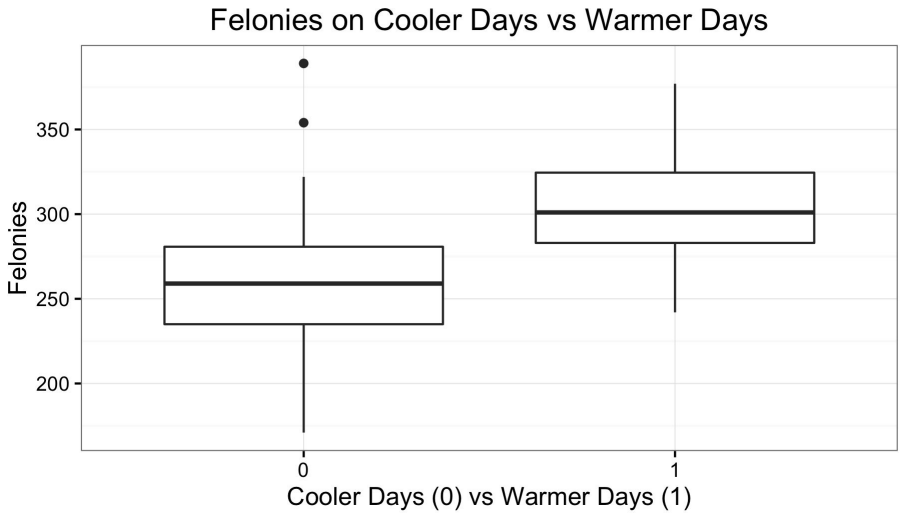
**Linearity:** *felonies* can be expressed as a linear combination of these predictors

**Constant variance** of errors for all values of Xi

**Independence** of errors

# Are felonies associated with temperature bands?




**Two-sample t-test:**

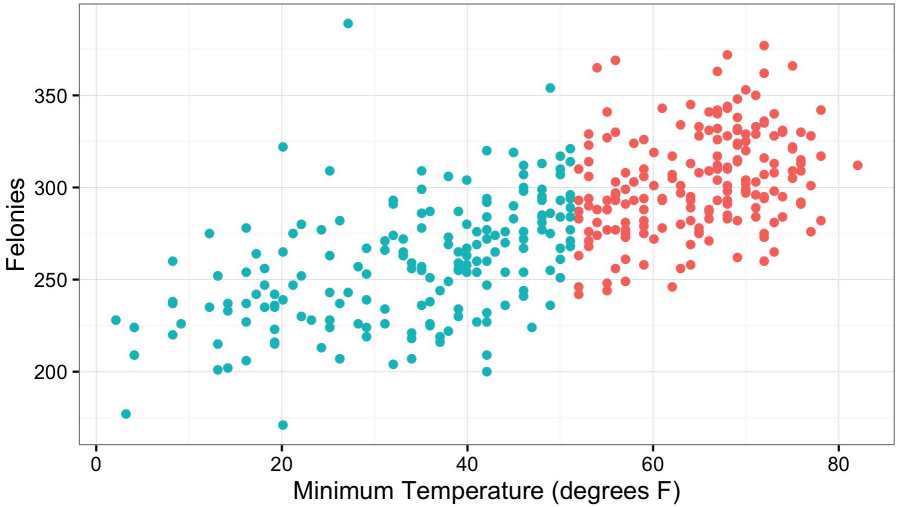
Average of 44 more felonies on warm days than on cool days

95% CI: 38 to 51 felonies

Two-sided p-value: <2e-16

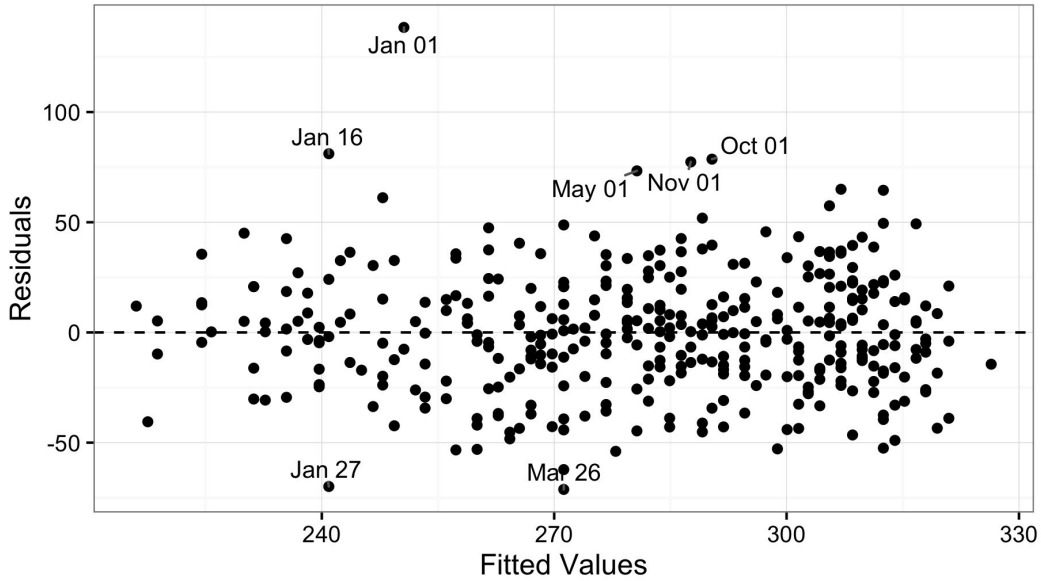
Overwhelming evidence of a difference

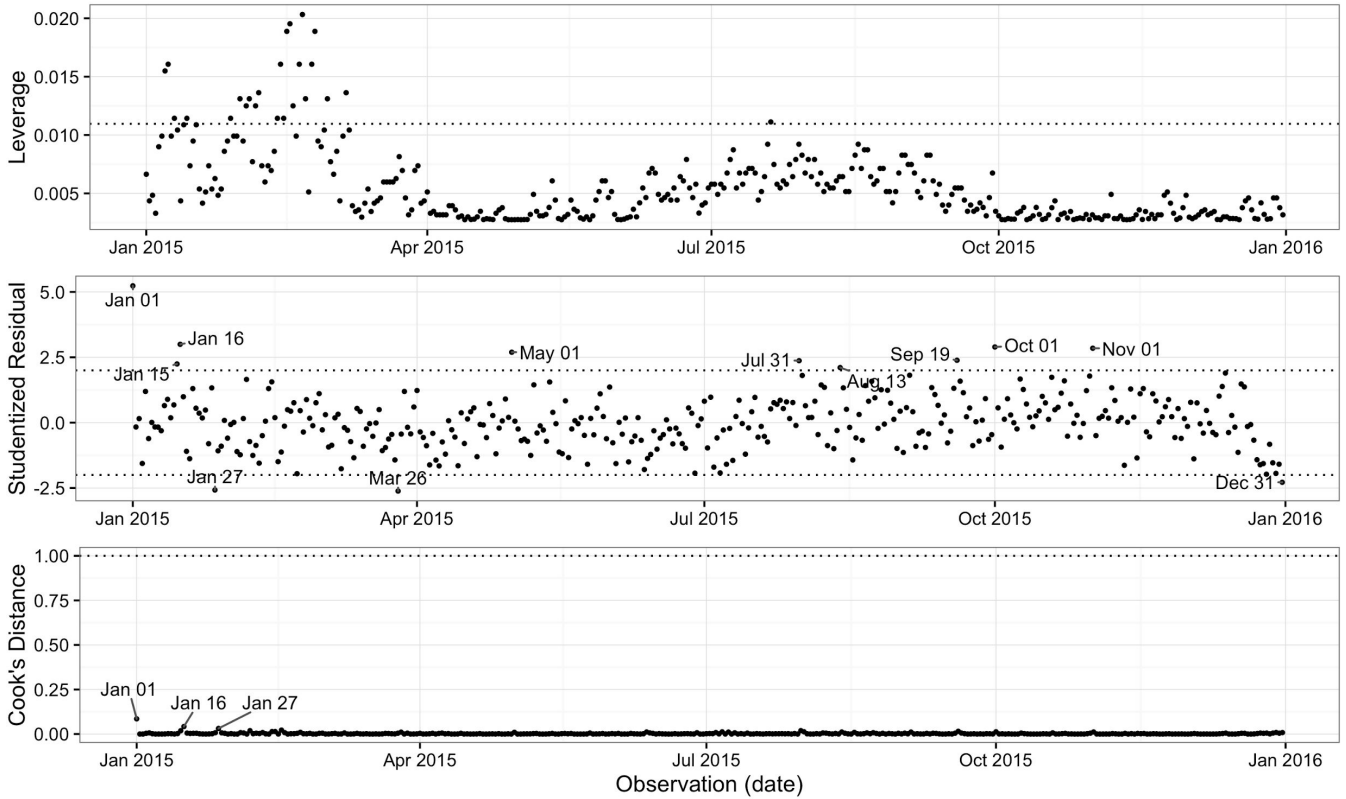
# Is there a linear association between felonies and temperature?




felonies ~ temp_min_degF

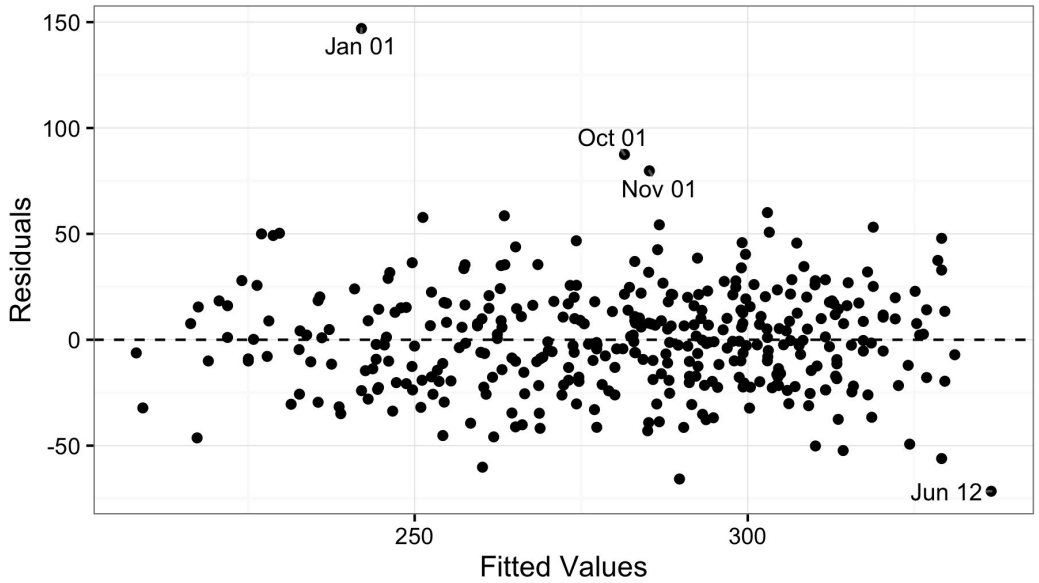
| | **Estimate** | **Pr(>\|t\|)** |
|---|---|---|
| (Intercept) | 213.1191 | < 2e-16 |
| temp_min_degF | 1.381 | < 2e-16 |


Jan 01
Jan 16
Oct 01
May 01
Nov 01
Jan 27
Mar 26
100
50
0
-50
Residuals
240
270
300
330
Fitted Values

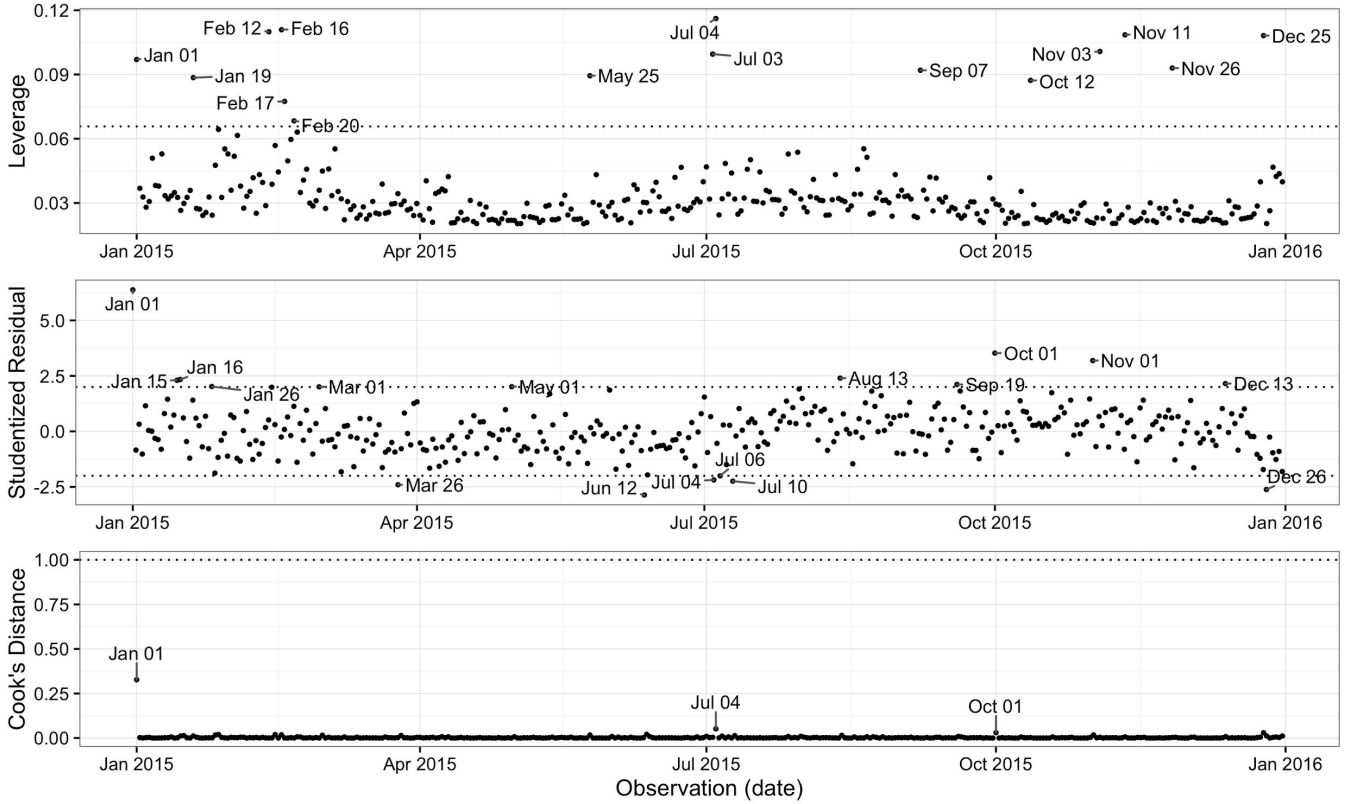

Leverage
0.020
0.015
0.010
0.005
Jan 2015
Apr 2015
Jul 2015
Oct 2015
Jan 2016
Studentized Residual
5.0
2.5
0.0
-2.5
Jan 01
Jan 16
Jan 15
Jan 27
Mar 26
May 01
Jul 31
Aug 13
Sep 19
Oct 01
Nov 01
Dec 31
Cook's Distance
1.00
0.75
0.50
0.25
0.00
Jan 01
Jan 16
Jan 27
Observation (date)

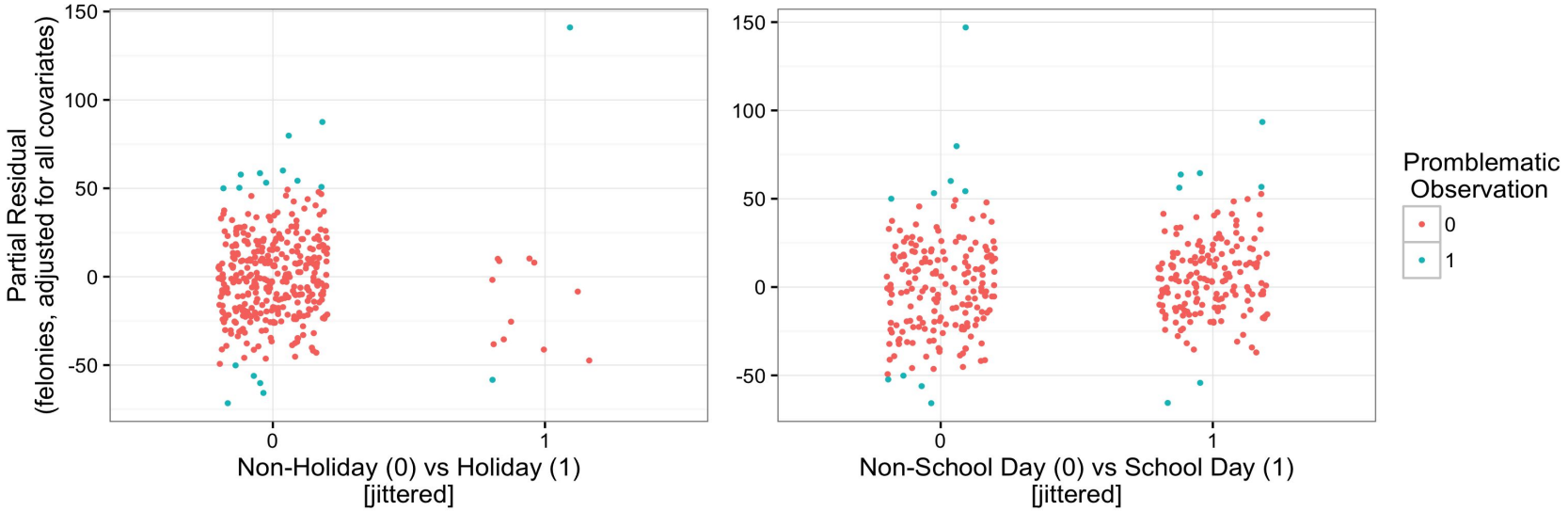
# Are other variables associated with felonies, after taking temperature into account?

felonies ~ temp_min_degF + any_precip + temp_min_degF * any_precip + is_holiday + is_school_day + day_of_week



| | With all observations | |
|---|---|---|
| | **Estimate** | **Pr(>\|t\|)** |
| (Intercept) | 210.7121 | < 2e-16 |
| temp_min_degF | 1.381 | < 2e-16 |
| any_precip1 | -21.0104 | 0.015 |
| is_holiday1 | -6.0063 | 0.4349 |
| is_school_day1 | 5.9484 | 0.1133 |
| day_of_week2 | 0.0687 | 0.9905 |
| day_of_week3 | -3.2969 | 0.5636 |
| day_of_week4 | -3.4695 | 0.5444 |
| day_of_week5 | -0.1765 | 0.9752 |
| day_of_week6 | 19.0105 | 0.0009 |
| day_of_week7 | 14.2132 | 0.0049 |
| temp_min_degF:any_precip1 | 0.1567 | 0.3434 |


Leverage
0.12
0.09
0.06
0.03
Feb 12
Feb 16
Jan 01
Jan 19
Feb 17
Feb 20
Jul 04
Jul 03
May 25
Sep 07
Nov 03
Oct 12
Nov 11
Nov 26
Dec 25
Jan 2015
Apr 2015
Jul 2015
Oct 2015
Jan 2016
Studentized Residual
5.0
2.5
0.0
-2.5
Jan 01
Jan 15
Jan 16
Jan 26
Mar 01
May 01
Aug 13
Sep 19
Oct 01
Nov 01
Dec 13
Mar 26
Jun 12
Jul 04
Jul 06
Jul 10
Dec 26
Jan 2015
Apr 2015
Jul 2015
Oct 2015
Jan 2016
Cook's Distance
1.00
0.75
0.50
0.25
0.00
Jan 01
Jul 04
Oct 01
Jan 2015
Apr 2015
Jul 2015
Oct 2015
Jan 2016
Observation (date)


Partial Residual
(felonies, adjusted for all covariates)
150
100
50
0
-50
0
1
Non-Holiday (0) vs Holiday (1)
[jittered]
150
100
50
0
-50
0
1
Non-School Day (0) vs School Day (1)
[jittered]
Promblematic
Observation
0
1

| | With all observations | | Excluding the 17 problematic observations | |
|---|---|---|---|---|
| | **Estimate** | **Pr(>\|t\|)** | **Estimate** | **Pr(>\|t\|)** |
| (Intercept) | 210.7121 | < 2e-16 | 202.2984 | < 2e-16 |
| temp_min_degF | 1.381 | < 2e-16 | 1.477 | < 2e-16 |
| any_precip1 | -21.0104 | 0.015 | -20.4116 | 0.0049 |
| is_holiday1 | -6.0063 | 0.4349 | -13.2689 | 0.0528 |
| is_school_day1 | 5.9484 | 0.1133 | 6.6914 | 0.0331 |
| day_of_week2 | 0.0687 | 0.9905 | 4.0126 | 0.4055 |
| day_of_week3 | -3.2969 | 0.5636 | 0.1073 | 0.9819 |
| day_of_week4 | -3.4695 | 0.5444 | -0.1882 | 0.9683 |
| day_of_week5 | -0.1765 | 0.9752 | -2.7734 | 0.5639 |
| day_of_week6 | 19.0105 | 0.0009 | 22.9916 | 2.41E-06 |
| day_of_week7 | 14.2132 | 0.0049 | 19.5004 | 5.45E-06 |
| temp_min_degF:any_precip1 | 0.1567 | 0.3434 | 0.138 | 0.3161 |

# Interaction between temperature and precipitation isn't significant.

felonies ~ temp_min_degF + any_precip + is_holiday + is_school_day + day_of_week

| | With all observations | | Excluding 15 problematic observations | |
|---|---|---|---|---|
| | **Estimate** | **Pr(>\|t\|)** | **Estimate** | **Pr(>\|t\|)** |
| (Intercept) | 208.705 | < 2e-16 | 202.3947 | < 2e-16 |
| temp_min_degF | 1.4222 | < 2e-16 | 1.4917 | < 2e-16 |
| any_precip1 | -13.367 | 0.00001 | -12.6094 | 0.000001 |
| is_holiday1 | -5.6495 | 0.4620 | -13.425 | 0.0529 |
| is_school_day1 | 6.0688 | 0.1059 | 6.7206 | 0.0346 |
| day_of_week2 | -0.274 | 0.9619 | 3.7736 | 0.4366 |
| day_of_week3 | -3.3049 | 0.5626 | -1.1931 | 0.8029 |
| day_of_week4 | -3.533 | 0.5369 | -1.2456 | 0.7950 |
| day_of_week5 | -0.1034 | 0.9855 | -3.7071 | 0.4457 |
| day_of_week6 | 18.8008 | 0.0011 | 21.8025 | 0.00001 |
| day_of_week7 | 14.0292 | 0.0055 | 18.2067 | 0.00002 |

# Questions of Interest

Are large increases in temperature (from the previous day) associated with an increase in felonies?

Is there an association after accounting for the other variables?

# Dataset 2

| date | felonies | felonies_diff | day_of_week | temp_min_degF | temp_min_degF_diff | temp_jump | any_precip | is_holiday | is_school_day |
|---|---|---|---|---|---|---|---|---|---|
| 2015-01-01 | 389 | 133 | 5 | 27.14 | 0 | 0 | 0 | 1 | 0 |
| 2015-01-02 | 257 | -132 | 6 | 35.06 | 7.92 | 0 | 0 | 0 | 0 |
| 2015-01-03 | 263 | 6 | 7 | 33.08 | -1.98 | 0 | 1 | 0 | 0 |
| 2015-01-04 | 227 | -36 | 1 | 41 | 7.92 | 0 | 1 | 0 | 0 |
| 2015-01-05 | 275 | 48 | 2 | 21.2 | -19.8 | 0 | 0 | 0 | 1 |
| ... | ... | ... | ... | ... | | ... | ... | ... | ... |
| 2015-12-23 | 246 | -22 | 4 | 51.98 | -1.08 | 0 | 1 | 0 | 1 |
| 2015-12-24 | 256 | 10 | 5 | 62.96 | 10.98 | 1 | 1 | 0 | 0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 2015-12-31 | 209 | -13 | 5 | 42.08 | 4.14 | 0 | 1 | 0 | 0 |

| | | |
|---|---|---|
| **felonies_diff** | Integer | Difference in felonies from the previous day |
| **teim_min_degF_diff** | Numeric | Difference in temperature from the previous day |
| **temp_jump** | Factor | 1: temp_min_degF_diff > 8<br>(90th percentile of increases in 2015) |

# Assumptions

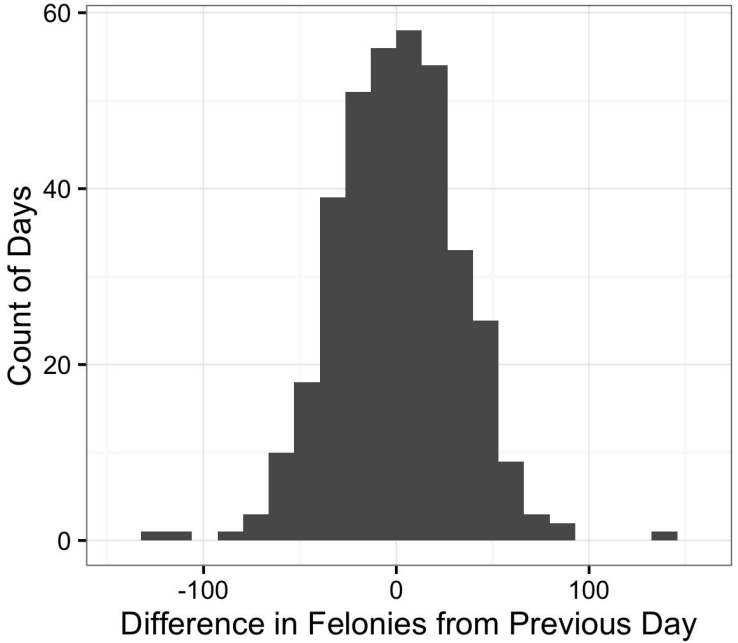
The *difference in felonies* from the previous day follows a normal distribution.

Normality

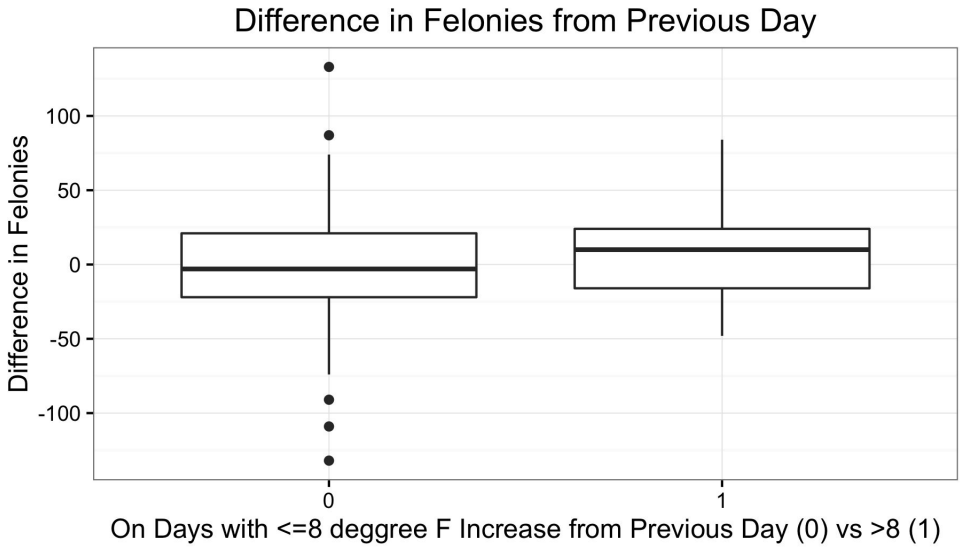
Linearity

Constant variance

Independence



# Is the difference in felonies different on days with larger temperature increases?




**Two-sample t-test:**

Average of 8.3 additional felonies on days after > 8 °F change as compared to ≤ 8 °F change

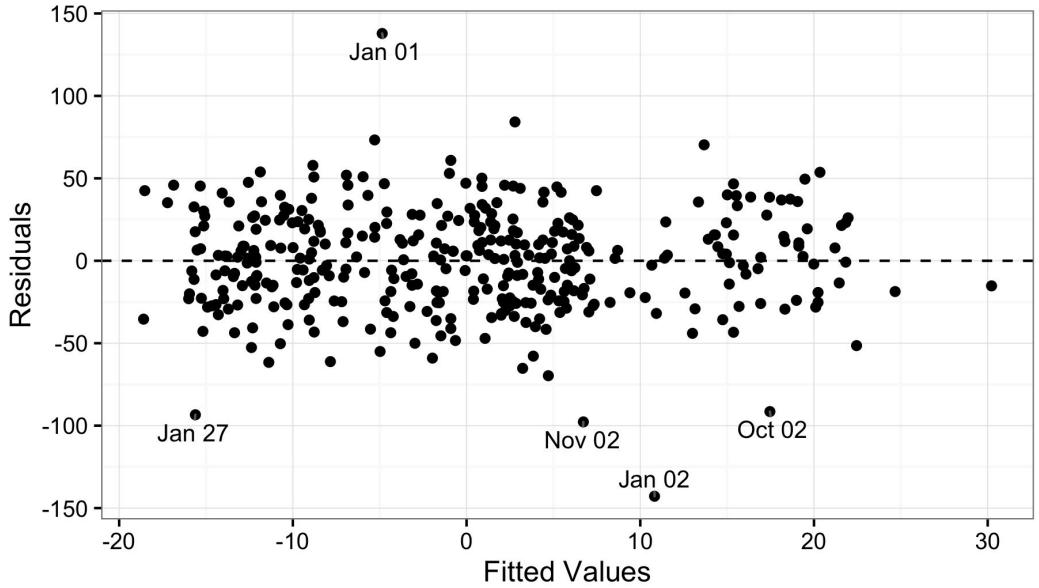
95% CI: -2.7 to 19.4 additional felonies

One-sided p-value: 0.0695

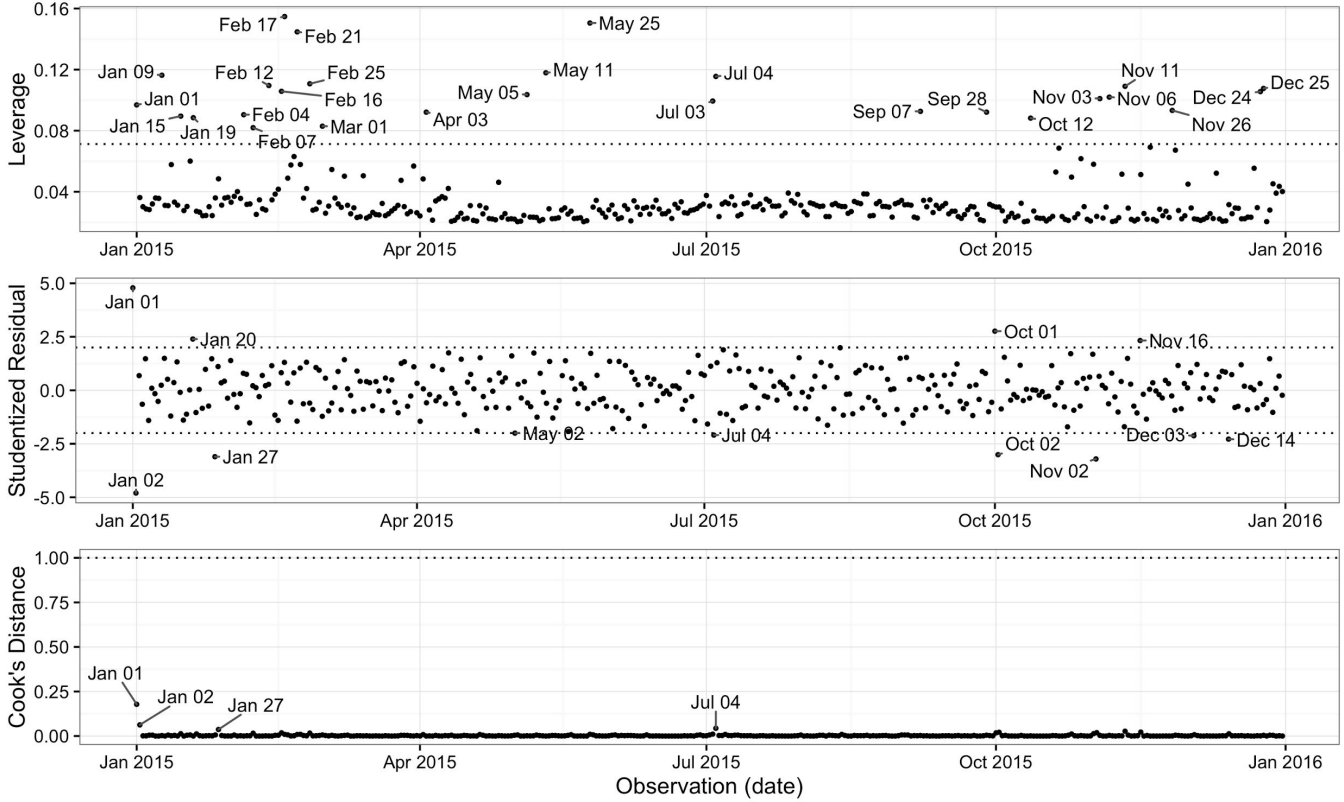
Suggestive, but inconclusive evidence

# Are other variables associated with increases in felonies, after taking large temperature increases into account?

felonies_diff ~ temp_jump + temp_min_degF + temp_jump * temp_min_degF + any_precip + is_holiday + is_school_day + day_of_week



| | With all observations | |
|---|---|---|
| | **Estimate** | **Pr(>\|t\|)** |
| (Intercept) | -19.0988 | 0.0044 |
| temp_jump1 | 21.7376 | 0.1854 |
| temp_min_degF | 0.1232 | 0.2068 |
| any_precip1 | -1.9053 | 0.6089 |
| is_holiday1 | 0.9505 | 0.9195 |
| is_school_day1 | 6.9512 | 0.1318 |
| day_of_week2 | 12.2219 | 0.0817 |
| day_of_week3 | 2.913 | 0.6765 |
| day_of_week4 | 11.0533 | 0.1151 |
| day_of_week5 | 9.9466 | 0.1529 |
| day_of_week6 | 25.5985 | 0.0003 |
| day_of_week7 | 1.7965 | 0.7702 |
| temp_jump1:temp_min_degF | -0.2825 | 0.4187 |


Leverage
0.16
0.12
0.08
0.04
Feb 17
Feb 21
May 25
Jan 09
Feb 12
Feb 25
May 11
Jul 04
Nov 11
Dec 25
Jan 01
Feb 16
May 05
Nov 03
Nov 06
Dec 24
Feb 04
Apr 03
Jul 03
Sep 07
Sep 28
Jan 15
Jan 19
Mar 01
Oct 12
Nov 26
Feb 07
Jan 2015
Apr 2015
Jul 2015
Oct 2015
Jan 2016
Studentized Residual
5.0
2.5
0.0
-2.5
-5.0
Jan 01
Jan 20
Oct 01
Nov 16
May 02
Jul 04
Dec 03
Dec 14
Jan 27
Oct 02
Jan 02
Nov 02
Jan 2015
Apr 2015
Jul 2015
Oct 2015
Jan 2016
Cook's Distance
1.00
0.75
0.50
0.25
0.00
Jan 01
Jan 02
Jan 27
Jul 04
Jan 2015
Apr 2015
Jul 2015
Oct 2015
Jan 2016
Observation (date)

# Iteratively remove insignificant variables

Interaction between the temperature jump indicator and temperature

Temperature

Precipitation indicator

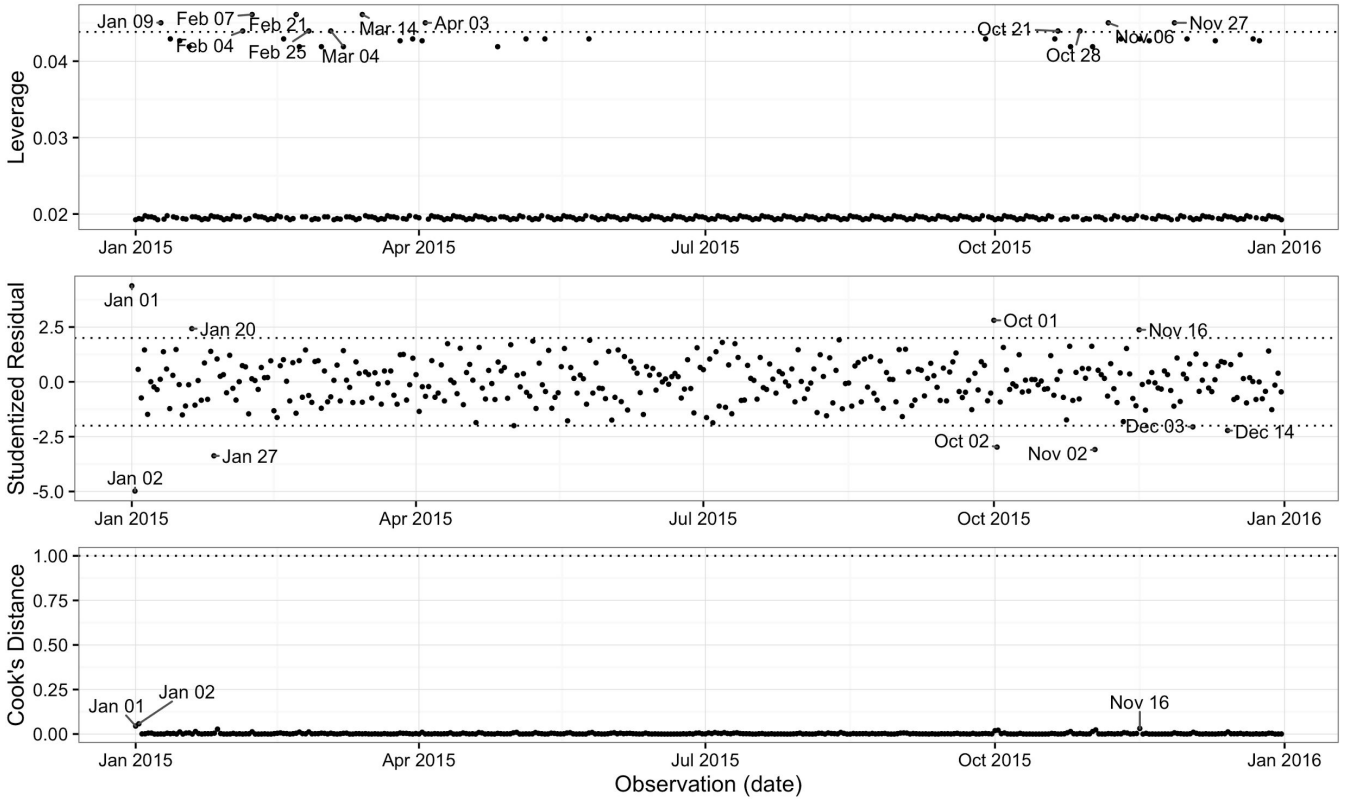
Holiday indicator

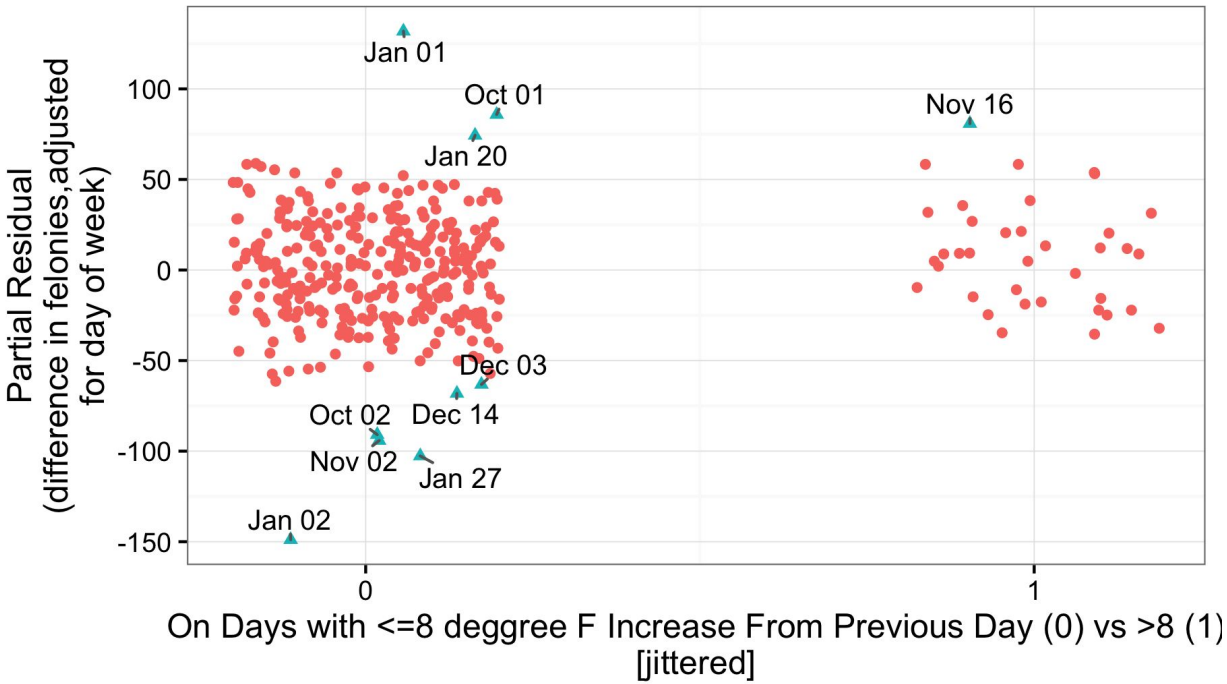
School day indicator

# Change in felonies is explained by day of week (and large temperature increases)

felonies_diff ~ temp_jump + day_of_week

| | With all observations | |
|---|---|---|
| | **Estimate** | **Pr(>\|t\|)** |
| (Intercept) | -13.3383 | 0.0025 |
| temp_jump1 | 8.9417 | 0.1001 |
| day_of_week2 | 16.4796 | 0.0074 |
| day_of_week3 | 6.9989 | 0.253 |
| day_of_week4 | 16.1516 | 0.0086 |
| day_of_week5 | 14.4958 | 0.0177 |
| day_of_week6 | 30.2082 | 0.000001 |
| day_of_week7 | 1.7455 | 0.7759 |


Leverage
0.04
0.03
0.02
Jan 09
Feb 07
Feb 21
Feb 04
Feb 25
Mar 04
Mar 14
Apr 03
Oct 21
Oct 28
Nov 06
Nov 27
Jan 2015
Apr 2015
Jul 2015
Oct 2015
Jan 2016
Studentized Residual
2.5
0.0
-2.5
-5.0
Jan 01
Jan 20
Jan 27
Jan 02
Oct 01
Nov 16
Oct 02
Nov 02
Dec 03
Dec 14
Jan 2015
Apr 2015
Jul 2015
Oct 2015
Jan 2016
Cook's Distance
1.00
0.75
0.50
0.25
0.00
Jan 01
Jan 02
Nov 16
Jan 2015
Apr 2015
Jul 2015
Oct 2015
Jan 2016
Observation (date)


Jan 01
Oct 01
Jan 20
Nov 16
Dec 03
Oct 02
Dec 14
Nov 02
Jan 27
Jan 02
100
50
0
-50
-100
-150
0
1
Partial Residual
(difference in felonies,adjusted
for day of week)
On Days with <=8 deggree F Increase From Previous Day (0) vs >8 (1)
[jittered]

| | With all observations | | Excluding 10 problematic observations | |
|---|---|---|---|---|
| | **Estimate** | **Pr(>\|t\|)** | **Estimate** | **Pr(>\|t\|)** |
| (Intercept) | -13.3383 | 0.0025 | -12.9815 | 0.0007 |
| temp_jump1 | 8.9417 | 0.1001 | 6.2912 | 0.1856 |
| day_of_week2 | 16.4796 | 0.0074 | 18.2375 | 0.0007 |
| day_of_week3 | 6.9989 | 0.253 | 7.5266 | 0.1585 |
| day_of_week4 | 16.1516 | 0.0086 | 16.0497 | 0.0025 |
| day_of_week5 | 14.4958 | 0.0177 | 11.3666 | 0.0335 |
| day_of_week6 | 30.2082 | 0.000001 | 34.8582 | 2.2e-10 |
| day_of_week7 | 1.7455 | 0.7759 | 1.5416 | 0.7707 |

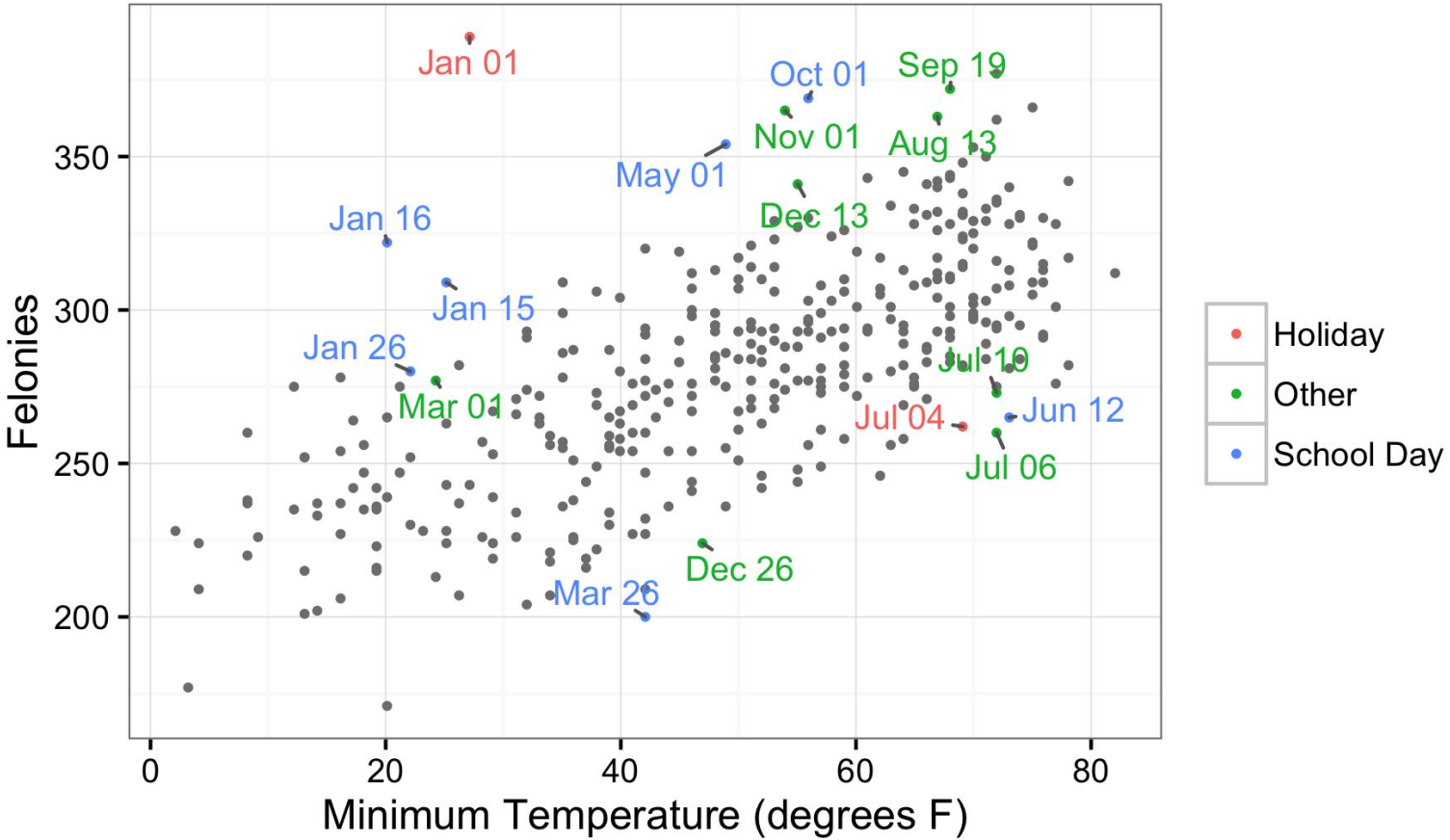
# Conclusions

After accounting for precipitation, holidays, school days, and day of week, for every 1°F increase in temperature, the NYPD can expect ~1.4 additional felonies per day (95% CI: 1.3–1.6 felonies; two-sided p-value < 2e-16).

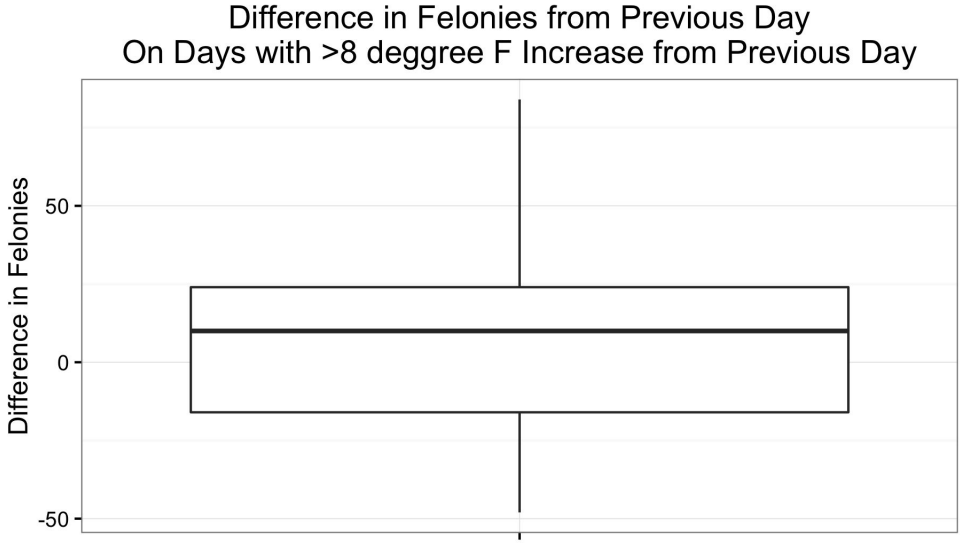
After accounting for day of week, a large jump in temperature (>8°F) from the previous day is not a significant indicator of the number of felonies relative to the previous day (two-sided p-value 0.1856).

Scope of inference:

- Observational data ⇒ no causal interpretation
- Observations weren't randomly sampled, so generalization of these results to other years or cities is speculative.


Jan 01
Oct 01
Sep 19
Nov 01
Aug 13
May 01
Jan 16
Dec 13
Jan 15
Jan 26
Jul 10
Mar 01
Jul 04
Jun 12
Jul 06
Dec 26
Mar 26
Felonies
350
300
250
200
0
20
40
60
80
Minimum Temperature (degrees F)
Holiday
Other
School Day

# Are increases in felonies associated with large increases in temperature?




**Paired/one-sample t-test:**

Average of 7.34 additional felonies on days with >8 °F increase

95% CI: -2.51 to 17.22 additional felonies

One-sided p-value: 0.0697

Suggestive, but inconclusive evidence