Chicago Domestic Crimes During Pandemic

Samuel, Natalie, & Stacy

INTRODUCTION AND DATA

Safer at home? Domestic Violence During the COVID-19 Pandemic

In what has been dubbed "a pandemic within a pandemic," a new public health crisis has emerged in the wake of the coronavirus - increasing rates of domestic violence, affecting those in abusive relationships around the globe. This trend is not necessarily surprising, since rates of domestic violence tend to increase whenever families spend more time together - such as during Christmas and summer vacations - according to Bristol University sociologist Marianne Hester, but still troubling nonetheless, especially in countries (like the United States) where the virus doesn't seem to be disappearing anytime soon. Without a strong support network due to social distancing and lockdown restrictions, it is much more difficult for victims of domestic violence to get help or escape (quarantined in their home), while the situation also enables the abuser by giving them greater power and control over what their partner can and can not do - suggesting that perhaps not everyone is truly "safer at home."

In order to further investigate the rate of domestic violence during COVID-19, we decided to pulled existing data from the City of Chicago Data Portal, which contains data related to reported incidents of crime (with the exception of murders where data exists for each victim) that occurred in the City of Chicago from 2001 to present. The data contained in this dataset were extracted from the Chicago Police Department's CLEAR (Citizen Law Enforcement Analysis and Reporting) system.

-introduce your general research question and your data (where it came from, how it was collected, what are the cases, what are the variables, etc.). -Define what we mean by domestic violence

METHODOLOGY

-variables used to address your research question -useful visualizations or summary statistics -introduce and justify the statistical method(s) that you believe will be useful in answering your research question.

```
## Rows: 231,002
## Columns: 21
## $ CASE.                  <chr> "JD163753", "JD212847", "JC497784", "JC459410...
## $ DATE..OF.OCCURRENCE    <chr> "02/24/2020 08:15:00 PM", "04/10/2020 10:56:0...
## $ BLOCK                  <chr> "031XX W LEXINGTON ST", "005XX W 103RD ST", "...
## $ IUCR                   <chr> "1153", "0560", "0860", "0560", "0810", "0820...
## $ PRIMARY.DESCRIPTION    <chr> "DECEPTIVE PRACTICE", "ASSAULT", "THEFT", "AS...
## $ SECONDARY.DESCRIPTION <chr> "FINANCIAL IDENTITY THEFT OVER $ 300", "SIMPL...
## $ LOCATION.DESCRIPTION   <chr> "", "RESIDENCE", "DEPARTMENT STORE", "SIDEWAL...
## $ ARREST                 <chr> "N", "N", "N", "N", "N", "N", "N", "N", "N", ...
## $ DOMESTIC               <chr> "N", "N", "N", "N", "N", "N", "N", "N", "N", ...
## $ BEAT                   <int> 1134, 2232, 1924, 122, 123, 2433, 312, 914, 3...
## $ WARD                   <int> 24, 9, 44, 4, 25, 48, 20, 11, 5, 26, 27, 37, ...
## $ FBI.CD                 <chr> "11", "08A", "06", "08A", "06", "06", "08A", ...
## $ X.COORDINATE           <int> NA, 1174583, NA, NA, NA, NA, 1180030, 1171590...
## $ Y.COORDINATE           <int> NA, 1836593, NA, NA, NA, NA, 1862317, 1887793...
```

```
## $ LATITUDE              <dbl> NA, 41.70700, NA, NA, NA, NA, 41.77747, 41.84...
## $ LONGITUDE             <dbl> NA, -87.63629, NA, NA, NA, NA, -87.61556, -87...
## $ LOCATION              <chr> "", "(41.707000821, -87.636288063)", "", "", ...
## $ MONTH                 <int> 2, 4, 11, 10, 5, 12, 5, 5, 4, 5, 4, 5, 5, 5, ...
## $ DAY                   <int> 24, 10, 3, 4, 24, 5, 7, 3, 28, 7, 25, 7, 7, 5...
## $ YEAR                  <int> 2020, 2020, 2019, 2019, 2020, 2019, 2020, 202...
## $ DATEINT               <int> 202002, 202004, 201911, 201910, 202005, 20191...
```

```
## # A tibble: 430 x 2
##    SECONDARY.DESCRIPTION               n
##    <chr>                           <int>
##  1 $500 AND UNDER                  20888
##  2 ABUSE / NEGLECT - CARE FACILITY     5
##  3 ABUSE/NEGLECT: CARE FACILITY       10
##  4 AGG CRIM SEX ABUSE FAM MEMBER      78
##  5 AGG CRIMINAL SEXUAL ABUSE          63
##  6 AGG PO HANDS ETC SERIOUS INJ        8
##  7 AGG PO HANDS NO/MIN INJURY        593
##  8 AGG PRO EMP HANDS SERIOUS INJ      16
##  9 AGG PRO.EMP: HANDGUN               26
## 10 AGG PRO.EMP: OTHER DANG WEAPON     85
## # ... with 420 more rows
```

```
##   isbeforecovid      n
## 1             0  90305
## 2             1 140697
```

```
##   isprelockdown      n
## 1             0 196383
## 2             1  34619
```

```
##   islockdown      n
## 1          0 214160
## 2          1  16842
```

```
##   isphase2      n
## 1        0 212424
## 2        1  18578
```

```
##   isphase3      n
## 1        0 218280
## 2        1  12722
```

```
##   isphase4      n
## 1        0 223458
## 2        1   7544
```

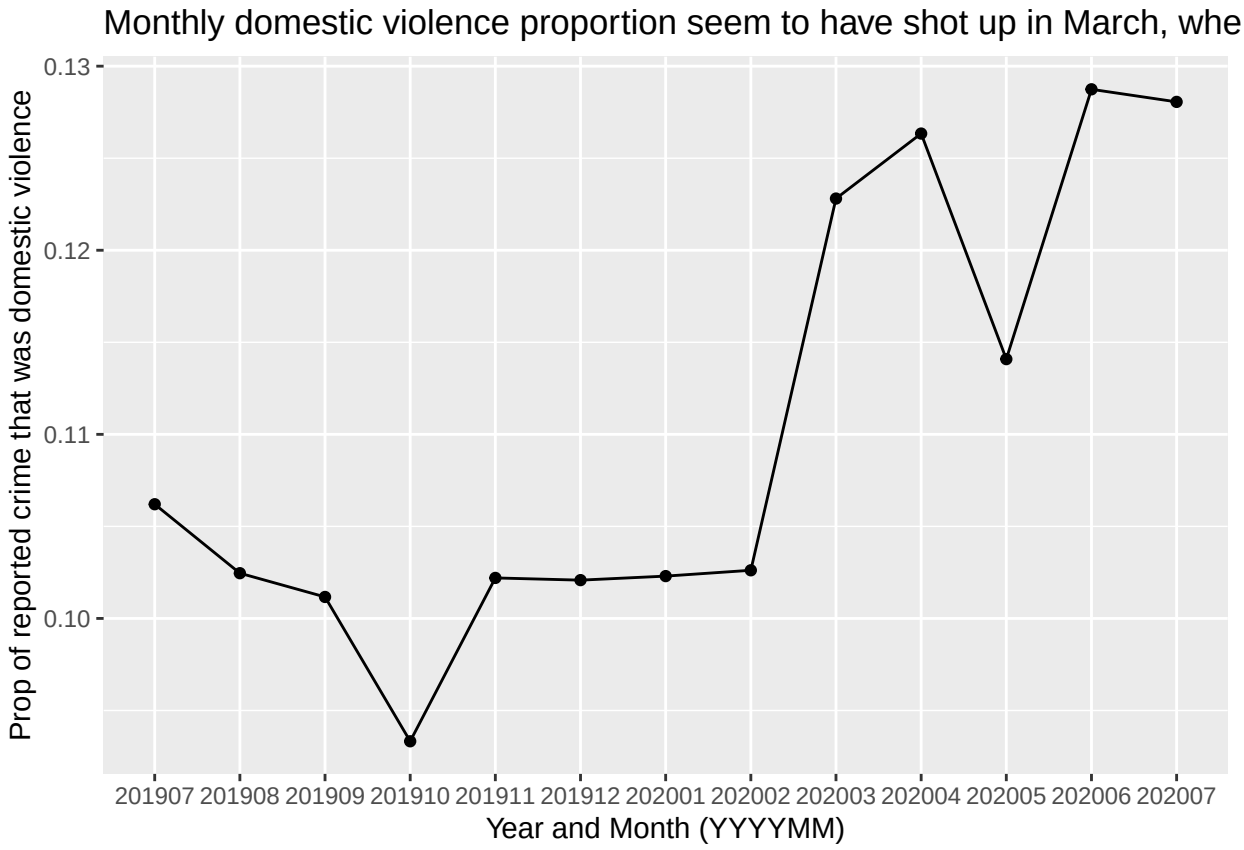
```
##   isdomviolence      n
## 1             0 206105
## 2             1  24897
```

```
## [1] isdomviolence n
## <0 rows> (or 0-length row.names)
```

```
##   isdomviolence     n
## 1             0 15502
## 2             1  1842
```

RESULTS

-Showcase how -Provide the main results from your analysis



Chi-Square Test

```
##
##     lockdown   post    pre
##   0    31109  17599 157397
##   1     4311   2667  17919
```

```
##
##  Pearson's Chi-squared test
##
## data:  table(domvio_mut$isdomviolence, domvio_mut$chisquare_indicators)
## X-squared = 247.63, df = 2, p-value < 2.2e-16
```

Since the data (Table 1) satisfies the independent sampling assumption and is large enough (i.e. each cell $>$ 10), we will be performing a chi-square test at the $\alpha = 0.05$ significance level. We test the two hypotheses below:

H~0 : The frequency of domestic violence cases in Chicago is unrelated to the phases of the pandemic. H~1 : The frequency of domestic violence cases in Chicago is related to the phases of the pandemic.

Under the null hypothesis, our test statistic has a chi-square distribution with 2 degrees of freedom. We performed the test and obtained a chi-square value of 247.63, which corresponds to a p-value of < 0.001. Thus, at an $\alpha = 0.05$ significance level, we reject the null hypothesis; there is sufficient evidence to suggest that the frequency of domestic violence cases in Chicago is related to the phases of the pandemic.

Step Down 2 Proportion Z-Tests Since the overall Chi-square test was significant, we stepped down to identify where the differences are. We conducted three 2 proportion z-tests. To account for multiple comparisons, we will perform the Bonferroni correction and thus assess our results relative to the adjusted $\alpha = 0.05/3$ level.

```
##
##  2-sample test for equality of proportions with continuity correction
##
## data:  c(4311, 2667) out of c(35420, 20266)
## X-squared = 11.411, df = 1, p-value = 0.0007303
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  -0.01569439 -0.00408326
## sample estimates:
##    prop 1    prop 2
## 0.1217109 0.1315997
```

```
##
##  2-sample test for equality of proportions with continuity correction
##
## data:  c(2667, 17919) out of c(2667 + 17599, 17919 + 157397)
## X-squared = 166.3, df = 1, p-value < 2.2e-16
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  0.024497 0.034283
## sample estimates:
##    prop 1    prop 2
## 0.1315997 0.1022097
```

```
##
##  2-sample test for equality of proportions with continuity correction
##
## data:  c(4311, 17919) out of c(35420, 17919 + 157397)
## X-squared = 118.55, df = 1, p-value < 2.2e-16
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  0.01579582 0.02320653
## sample estimates:
##    prop 1    prop 2
## 0.1217109 0.1022097
```

```
## [1] 0.01666667
```

We found that all three pairwise difference in proportions are significant at the adjusted significance level.

Regression Analysis

```
## # A tibble: 6 x 5
##   term                       estimate std.error statistic  p.value
##   <chr>                         <dbl>     <dbl>     <dbl>    <dbl>
## 1 (Intercept)                 -2.18     0.00883   -247.   0.
## 2 pandemic_timelockdown        0.303    0.0244      12.4  1.88e-35
## 3 pandemic_timephase 2         0.111    0.0249       4.48 7.47e- 6
## 4 pandemic_timephase 3         0.309    0.0276      11.2  3.38e-29
## 5 pandemic_timephase 4         0.272    0.0355       7.68 1.61e-14
## 6 pandemic_timepre-lockdown    0.0472   0.0196       2.41 1.60e- 2
```

Our logistic regression model corresponding to the probability of success that a reported crime for a given time period will be one of domestic violence is as follows:

$\hat{\beta}_0$ + $\hat{\beta}_1$(pandemic_time == lockdown) + $\hat{\beta}_2$(pandemic_time == phase 2) + $\hat{\beta}_3$(pandemic_time == phase 3) + $\hat{\beta}_4$(pandemic_time == phase 4) + $\hat{\beta}_5$(pandemic_time == pre-lockdown)

At the $\alpha = 0.05$ significance level, the true β coefficients corresponding to the logit of the probability that a case reported during that time period was related to domestic violence for each of our dummy variables (all relative to the baseline/reference category of before the COVID-19 pandemic) were statistically significant. We were specifically interested in β_1, which corresponds to the logit of the probability of success (that a reported case was related to domestic violence) during lockdown, as compared to before the pandeimc began, and whether there was a relationship between the probability of a reported crime being related to domestic violence during lockdown, when compared to before the pandemic.

H_0: $\beta_1 = 0$ (There is no relationship between our predictor and the probability of success that a reported case is related to domestic violence, while holding all other variables constant. The true population parameter β_1 is equal to 0).

H_1: (There is a relationship between our predictor and the probability of success that a reported case is related to domestic violence, while holding all other variables constant. The true population parameter β_1 is not equal to 0).

Under the null hypothesis, our test statistic follows a standard normal distribution. The value of our test statistic is equal to approximately 12.426, which corresponds to a p-value of less than 0.001. Thus, at the α = 0.05, we reject our null hypothesis; we have sufficient evidence to suggest that the true value of β_1 is not equal to 0, such that this predictor tells us soemthing about the probability of success of our outcome, while holding all other variables constant.

We see from the model output above that the estimated $\hat{\beta}_1$ coefficent, corresponding to the effect of being in the "lockdown" phase of time in quarantine - relative to the time period before COVID-19 - on the probability of success that a reported case will be related to domestic violence, is approximately 0.303. This corresponds to an odds ratio of exp(0.303), which is approximately 1.354. Therefore, we would expect crimes reported during the lockdown phase of the stay-at-home order in the city of Chicago to have 1.354 times the odds of being related to domestic violence, relative to cases that were reported before the COVID-19 pandemic began, while holding all other variables constant.

```
## # A tibble: 231,002 x 4
##    isdomviolence pandemic_time .fitted pred_prob
##            <dbl> <chr>           <dbl>     <dbl>
##  1             0 pre-lockdown    -2.14     0.106
##  2             0 lockdown        -1.88     0.132
##  3             0 before_covid    -2.18     0.101
##  4             0 before_covid    -2.18     0.101
##  5             0 phase 2         -2.07     0.112
##  6             0 before_covid    -2.18     0.101
##  7             0 phase 2         -2.07     0.112
##  8             0 phase 2         -2.07     0.112
##  9             0 lockdown        -1.88     0.132
## 10             0 phase 2         -2.07     0.112
## # ... with 230,992 more rows
```

DISCUSSION

-summary of what you have learned about your research question along -statistical arguments supporting your conclusions -critique your own methods and provide suggestions for improving your analysis (Issues pertaining to the reliability and validity of your data and appropriateness of the statistical analysis) Make

sure to mention possible confounders we did not include in our model, perhaps also the fact that we were just looking at Chicago, vs. the US as a whole or even multiple countires. -what you would do differently -what you would do next if you were going to continue work on the project Are rates of domestic violence during lockdown different in different parts of the world and/or US?