Chicago Domestic Crimes During Pandemic

Samuel, Natalie, & Stacy

INTRODUCTION AND DATA

-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
```

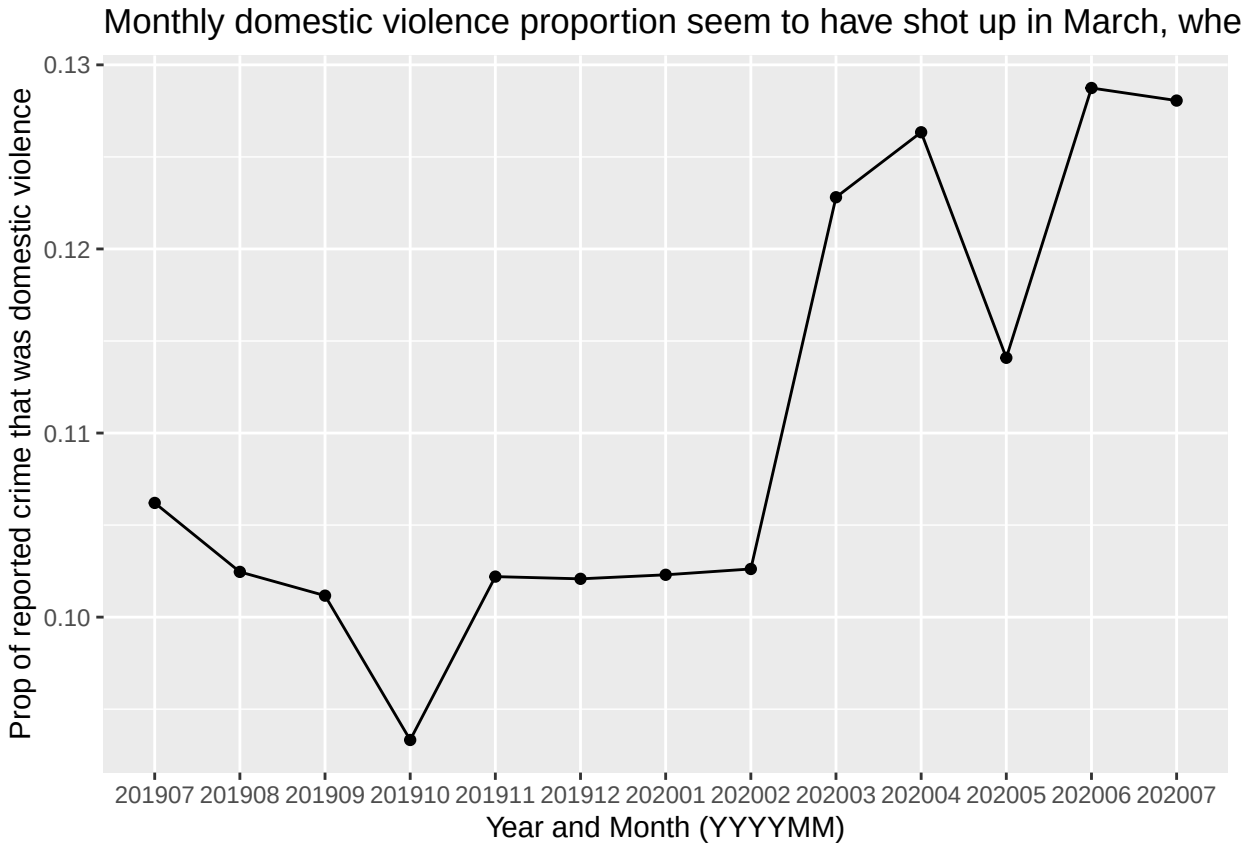
```
##   isdomviolence      n
## 1             0 206105
## 2             1  24897
```

```
## [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:

H0 : The frequency of domestic violence cases in Chicago is unrelated to the phases of the pandemic. H1 : 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 siginficance level.

Regression Analysis

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) -what you would do differently -what you would do next if you were going to continue work on the project