# Problemset5

Sharon Allman, Diego Mamanche Castellanos & Ke-Li Chiu

11/04/2020

**Abstract**

Abstract .....

```
## Warning: package 'broom' was built under R version 3.6.2

## Warning: package 'tidyverse' was built under R version 3.6.2

## Warning: package 'tidyr' was built under R version 3.6.2

## Warning: package 'gridExtra' was built under R version 3.6.2

## Warning: package 'pastecs' was built under R version 3.6.2
```

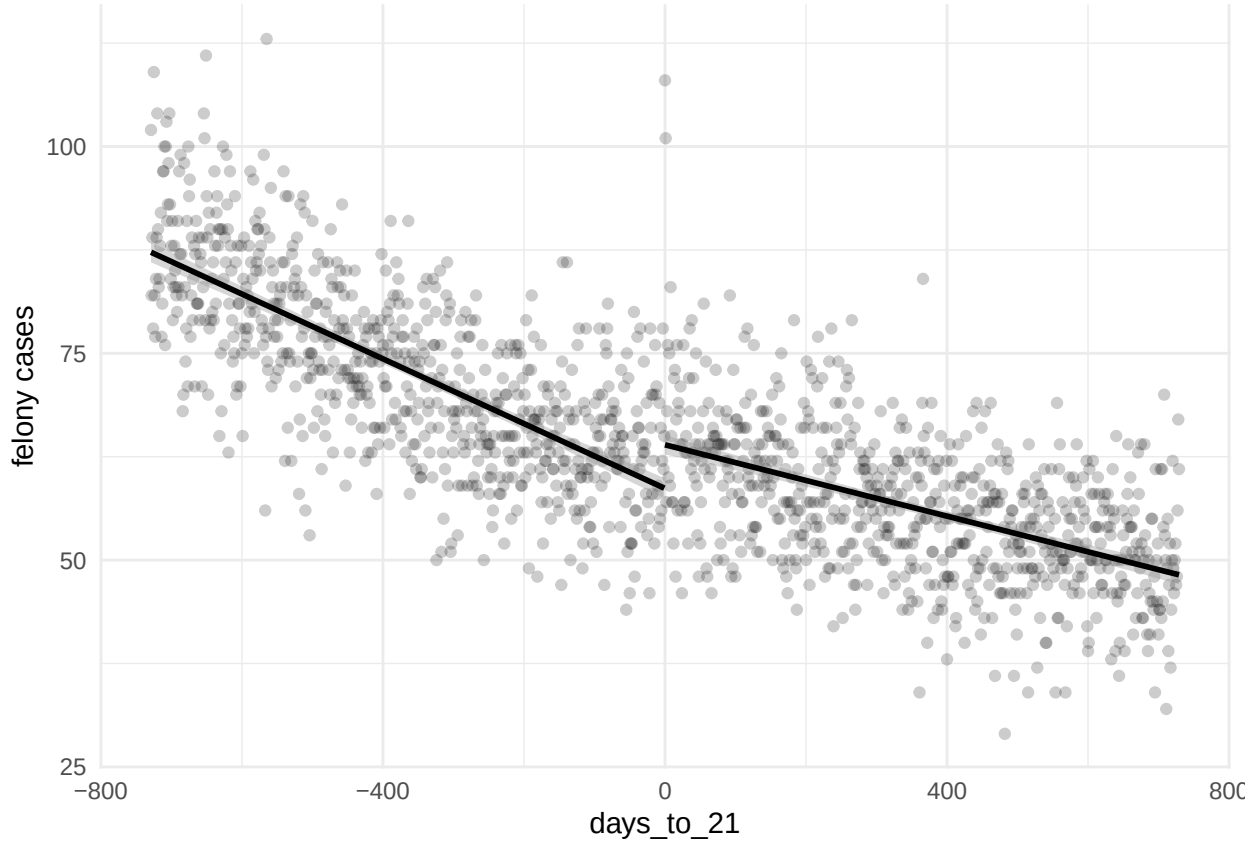
```
##   x days_to_21  all felony misdemeanor fbi_offense_miss violent murder
## 1 1      -1461 5289   1556        3064              206     686      7
## 2 2      -1460 4800   1460        2756              196     581      8
## 3 3      -1459 4464   1356        2589              214     561      8
## 4 4      -1458 4425   1382        2511              180     587     15
## 5 5      -1457 4451   1423        2535              169     596      5
## 6 6      -1456 4501   1375        2593              176     587      9
##   manslaughter rape robbery assault aggravated_assault ot_assault property
## 1            2   18     136     523                205        318     1438
## 2            1    8     113     451                183        268     1392
## 3            1    5     105     442                160        282     1329
## 4            1   10     106     455                181        274     1361
## 5            0    7     106     478                190        288     1351
## 6            0   12     110     456                171        285     1373
##   burglary larceny mv_theft stolen_prop_buy_rec_poss vandalism ill_drugs
## 1      362     576      200                      109       191       618
## 2      353     561      206                       90       182       564
## 3      331     527      172                      109       190       529
## 4      330     564      196                      105       166       498
## 5      334     546      188                       94       189       523
## 6      343     559      183                       90       198       510
##   cocaine_opio_sale_manuf mj_sale_manuf dang_non_narc_sale_manuf
## 1                      40           51                      13
## 2                      47           44                      16
## 3                      43           48                       9
## 4                      40           48                       8
## 5                      55           37                      13
```

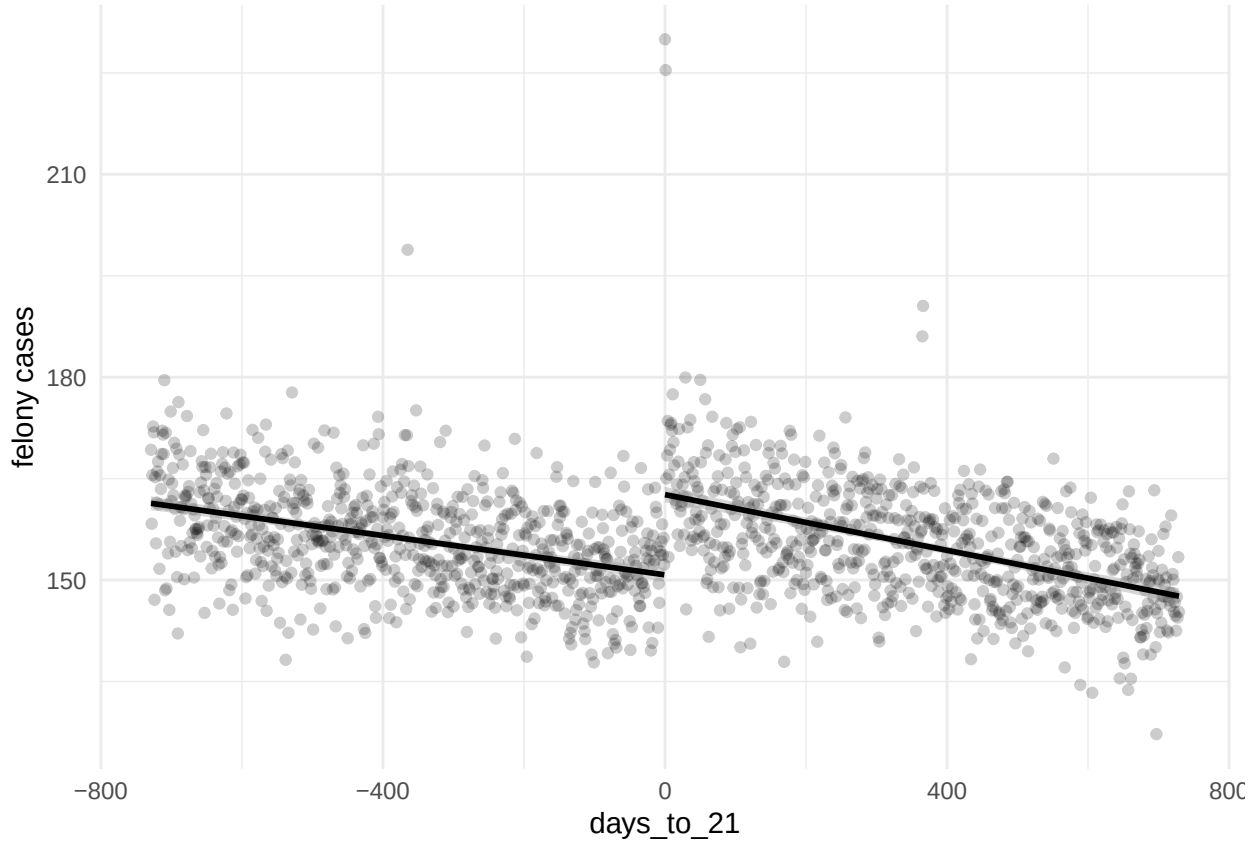
```
## 6                     51           44                      13
##   cocaine_opio_posses mj_posses dang_non_narc_posses alcohol drunk_at_risk
## 1                 129       275                  110     874             0
## 2                 123       253                   81     723             1
## 3                 115       239                   75     619             1
## 4                  86       248                   68     577             1
## 5                 112       231                   75     634             0
## 6                  98       242                   62     602             2
##   dui liquor_laws drunkeness_pc disorderly_cond vagrancy fbi_other arson
## 1 120         361           246             124       23       713     7
## 2 139         250           203             114       17       676    16
## 3  94         253           139             122       11       578     7
## 4  96         226           150              86       19       605     6
## 5  98         249           162             112       13       589     3
## 6 104         222           146             107       23       624     8
##   forgery_counterfit fraud embezelment weapons prostitution sex_offenses
## 1                 24    21           6     133           13           40
## 2                 22    23           4     129            8           43
## 3                 16    13           8     116           13           29
## 4                 16    17           3     132           11           40
## 5                 20    11          12     139           10           38
## 6                 25    29           7     118            8           36
##   bookmaking all_other_gambling aga_family_child curfew_loitering runaways
## 1          0                 3                0             331      135
## 2          0                 2                0             334       95
## 3          0                 2                0             275       99
## 4          0                 4                0             285       91
## 5          0                 3                0             261       92
## 6          1                 3                0             284      105
##   coded_on_offense county_ordinance outside_warrent traffic_violations
## 1              518             155              19                100
## 2              480             159              17                103
## 3              471             139              19                104
## 4              444             149              13                 94
## 5              419             129              16                 95
## 6              445             165              20                 94
##   hit_run_reckl_driv parole_violation dump_or_litter trespass
## 1                 85               13             11      106
## 2                 75                3              8       87
## 3                 88                9              4       71
## 4                 69               14              5       74
## 5                 57                6              3       89
## 6                 60               11              3       62
##   violate_court_order failure_to_appear cont_delinq_minor not_classified
## 1                   5                16                8            442
## 2                   5                21                2            384
## 3                  10                20                7            377
## 4                   6                18                2            353
## 5                  11                12                1            339
## 6                   4                25                1            360
##   other_reported     pop1     pop2    pop3 pop_day1 pop_day2 pop_day3
## 1            104 12877264 12486051 3175170 35280.17 34208.36 8699.095
## 2            103 12877587 12486315 3175108 35281.06 34209.08 8698.926
## 3            109 12877910 12486579 3175046 35281.95 34209.81 8698.757
```

```
## 4            102 12878234 12486843 3174984 35282.83 34210.53 8698.587
## 5             90 12878557 12487107 3174923 35283.72 34211.25 8698.418
## 6            105 12878881 12487371 3174861 35284.61 34211.98 8698.249
##   years_to_21 yrs_to_21_month yrs_to_21_fortnight post    linear   square
## 1   -4.002740         17.0137            16.99178    0 -4.002740 16.02193
## 2   -4.000000         17.0137            16.99178    0 -4.000000 16.00000
## 3   -3.997260         17.0137            16.99178    0 -3.997260 15.97809
## 4   -3.994521         17.0137            16.99178    0 -3.994521 15.95619
## 5   -3.991781         17.0137            16.99178    0 -3.991781 15.93431
## 6   -3.989041         17.0137            17.03014    0 -3.989041 15.91245
##       cubic linear_post square_post cubic_post birthday_19 birthday_19_1
## 1 -64.13160           0           0          0           0             0
## 2 -64.00000           0           0          0           0             0
## 3 -63.86858           0           0          0           0             0
## 4 -63.73735           0           0          0           0             0
## 5 -63.60629           0           0          0           0             0
## 6 -63.47541           0           0          0           0             0
##   birthday_20 birthday_20_1 birthday_21 birthday_21_1 birthday_22
## 1           0             0           0             0           0
## 2           0             0           0             0           0
## 3           0             0           0             0           0
## 4           0             0           0             0           0
## 5           0             0           0             0           0
## 6           0             0           0             0           0
##   birthday_22_1 birthday_23 birthday_23_1 ill_drugs_r
## 1             0           0             0    175.1692
## 2             0           0             0    159.8591
## 3             0           0             0    149.9350
## 4             0           0             0    141.1451
## 5             0           0             0    148.2270
## 6             0           0             0    144.5390
##   cocaine_opio_sale_manuf_r mj_sale_manuf_r dang_non_narc_sale_manuf_r
## 1                  11.33781        14.45571                   3.684789
## 2                  13.32160        12.47128                   4.535011
## 3                  12.18754        13.60469                   2.550880
## 4                  11.33696        13.60435                   2.267392
## 5                  15.58793        10.48642                   3.684419
## 6                  14.45390        12.47003                   3.684326
##   cocaine_opio_posses_r mj_posses_r dang_non_narc_posses_r    dui_r
## 1              36.56445    77.94746               31.17898 34.01344
## 2              34.86290    71.70986               22.95849 39.39791
## 3              32.59457    67.74003               21.25733 26.64252
## 4              24.37446    70.28914               19.27283 27.20870
## 5              31.74269    65.46929               21.25626 27.77485
## 6              27.77415    68.58515               17.57140 29.47461
##   liquor_laws_r drunkeness_pc_r felony_r misdemeanor_r violent_r murder_r
## 1     102.32376        69.72755 454.8596      895.6875  200.5358 2.046283
## 2      70.85955        57.53795 426.7872      805.6340  169.8379 2.338560
## 3      71.70806        39.39692 396.3776      756.8006  163.9881 2.338511
## 4      64.05381        42.51359 403.9692      733.9846  171.5846 4.384615
## 5      70.57079        45.91353 415.9450      740.9843  174.2117 1.461507
## 6      62.91696        41.37782 401.9061      757.9217  171.5773 2.630658
##   manslaughter_r   rape_r robbery_r assault_r aggravated_assault_r
## 1      0.5846524 5.261872  39.75636  152.8866             59.92687
```

```
## 2      0.2923200 2.338560  33.03216  131.8363             53.49457
## 3      0.2923139 1.461569  30.69295  129.2027             46.77022
## 4      0.2923077 2.923077  30.98461  133.0000             52.90769
## 5      0.0000000 2.046110  30.98396  139.7201             55.53728
## 6      0.0000000 3.507544  32.15248  133.2867             49.98250
##   ot_assault_r property_r burglary_r larceny_r mv_theft_r
## 1     92.95974   420.3651  105.82209  168.3799   58.46524
## 2     78.34177   406.9095  103.18897  163.9915   60.21793
## 3     82.43251   388.4851   96.75588  154.0494   50.27798
## 4     80.09230   397.8307   96.46153  164.8615   57.29230
## 5     84.18283   394.8993   97.62870  159.5966   54.95268
## 6     83.30416   401.3215  100.25729  163.3931   53.49004
##   stolen_prop_buy_rec_poss_r vandalism_r disorderly_cond_r vagrancy_r
## 1                   31.86356    55.83431          36.24845   6.723503
## 2                   26.30880    53.20225          33.32448   4.969441
## 3                   31.86221    55.53963          35.66229   3.215452
## 4                   30.69231    48.52307          25.13846   5.553846
## 5                   27.47634    55.24498          32.73777   3.799919
## 6                   26.30658    57.87447          31.27560   6.722792
##   fbi_other_r   arson_r forgery_counterfit_r  fraud_r embezelment_r
## 1    208.4286 2.0462835             7.015829 6.138850      1.753957
## 2    197.6083 4.6771205             6.431041 6.723361      1.169280
## 3    168.9574 2.0461970             4.677022 3.800080      2.338511
## 4    176.8461 1.7538460             4.676923 4.969230      0.876923
## 5    172.1656 0.8769045             5.846030 3.215316      3.507618
## 6    182.3923 2.3383625             7.307383 8.476564      2.046067
##   weapons_r prostitution_r sex_offenses_r bookmaking_r
## 1  38.87939       3.800241      11.693049    0.0000000
## 2  37.70928       2.338560      12.569761    0.0000000
## 3  33.90841       3.800080       8.477102    0.0000000
## 4  38.58461       3.215384      11.692307    0.0000000
## 5  40.62991       2.923015      11.107457    0.0000000
## 6  34.49085       2.338362      10.522631    0.2922953
##   all_other_gambling_r aga_family_child_r curfew_loitering_r runaways_r
## 1            0.8769786                  0           96.75998   39.46404
## 2            0.5846401                  0           97.63489   27.77040
## 3            0.5846277                  0           80.38631   28.93907
## 4            1.1692307                  0           83.30769   26.60000
## 5            0.8769045                  0           76.29069   26.89174
## 6            0.8768859                  0           83.01187   30.69101
##   county_ordinance_r traffic_violations_r hit_run_reckl_driv_r
## 1           45.31056             29.23262              24.84773
## 2           46.47889             30.10896              21.92400
## 3           40.63163             30.40064              25.72362
## 4           43.55384             27.47692              20.16923
## 5           37.70689             27.76864              16.66119
## 6           48.22873             27.47576              17.53772
##   dump_or_litter_r trespass_r violate_court_order_r failure_to_appear_r
## 1        3.2155883   30.98658              1.461631            4.677219
## 2        2.3385603   25.43184              1.461600            6.138721
## 3        1.1692554   20.75428              2.923139            5.846277
## 4        1.4615384   21.63077              1.753846            5.261538
## 5        0.8769045   26.01483              3.215316            3.507618
## 6        0.8768859   18.12231              1.169181            7.307383
```

```
##   cont_delinq_minor_r not_classified_r drunk_at_risk_r outside_warrent_r
## 1           2.3386097        129.20819        0.000000          21.84135
## 2           0.5846401        112.25089        1.149567          19.54264
## 3           2.0461970        110.20232        1.149590          21.84220
## 4           0.5846153        103.18461        1.149612          14.94496
## 5           0.2923015         99.09021        0.000000          18.39415
## 6           0.2922953        105.22631        2.299313          22.99313
##   parole_violation_r other_reported_r coded_on_offense_r alcohol_r
## 1          14.944083         119.5527           178.8560  249.0367
## 2           3.448702         118.4054           157.4586  207.2389
## 3          10.346306         125.3053           161.6835  177.7748
## 4          16.094567         117.2604           152.9318  165.6180
## 5           6.897806         103.4671           141.3356  180.7969
## 6          12.646224         120.7140           156.6496  174.0671
##      all_r  other_r
## 1 1561.600 516.4927
## 2 1411.163 467.3178
## 3 1321.026 440.8433
## 4 1309.141 432.9626
## 5 1310.726 412.5914
## 6 1335.773 444.2682
```

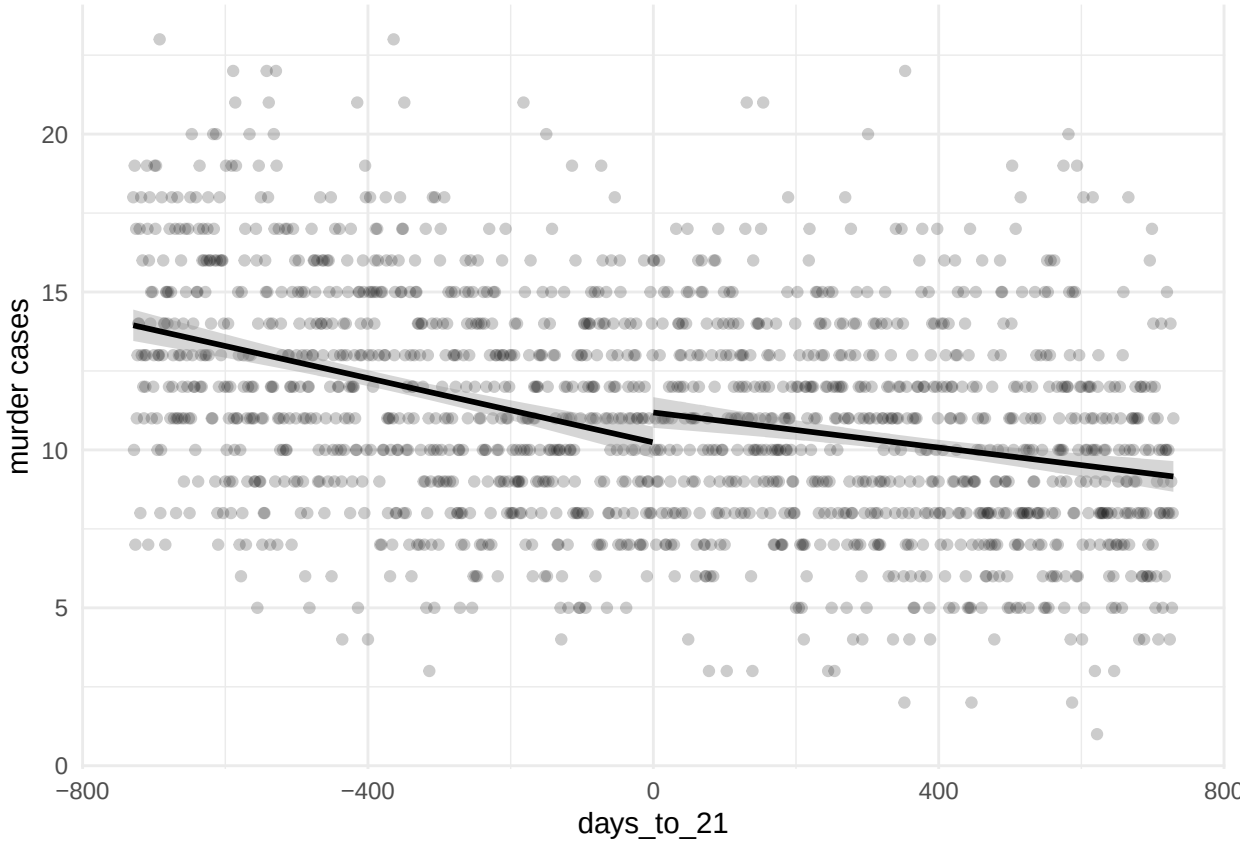


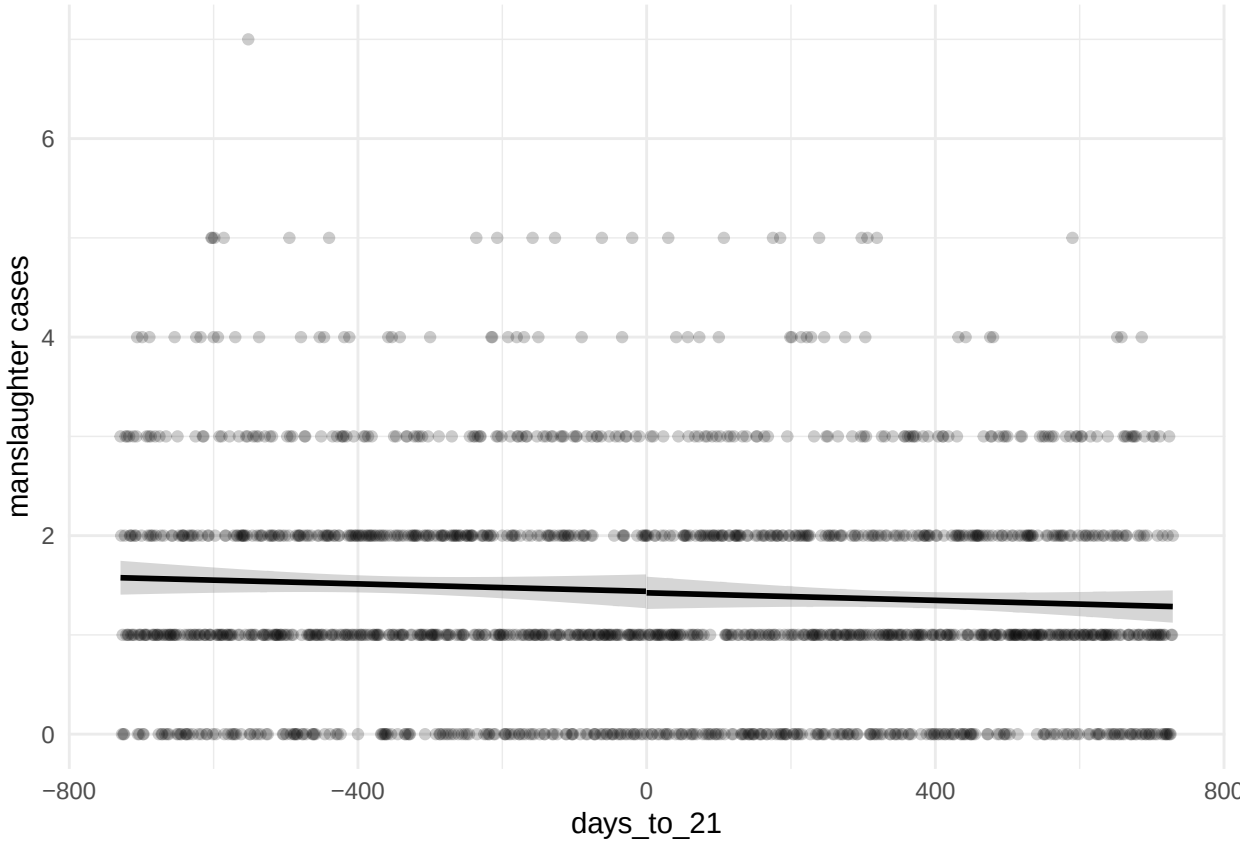


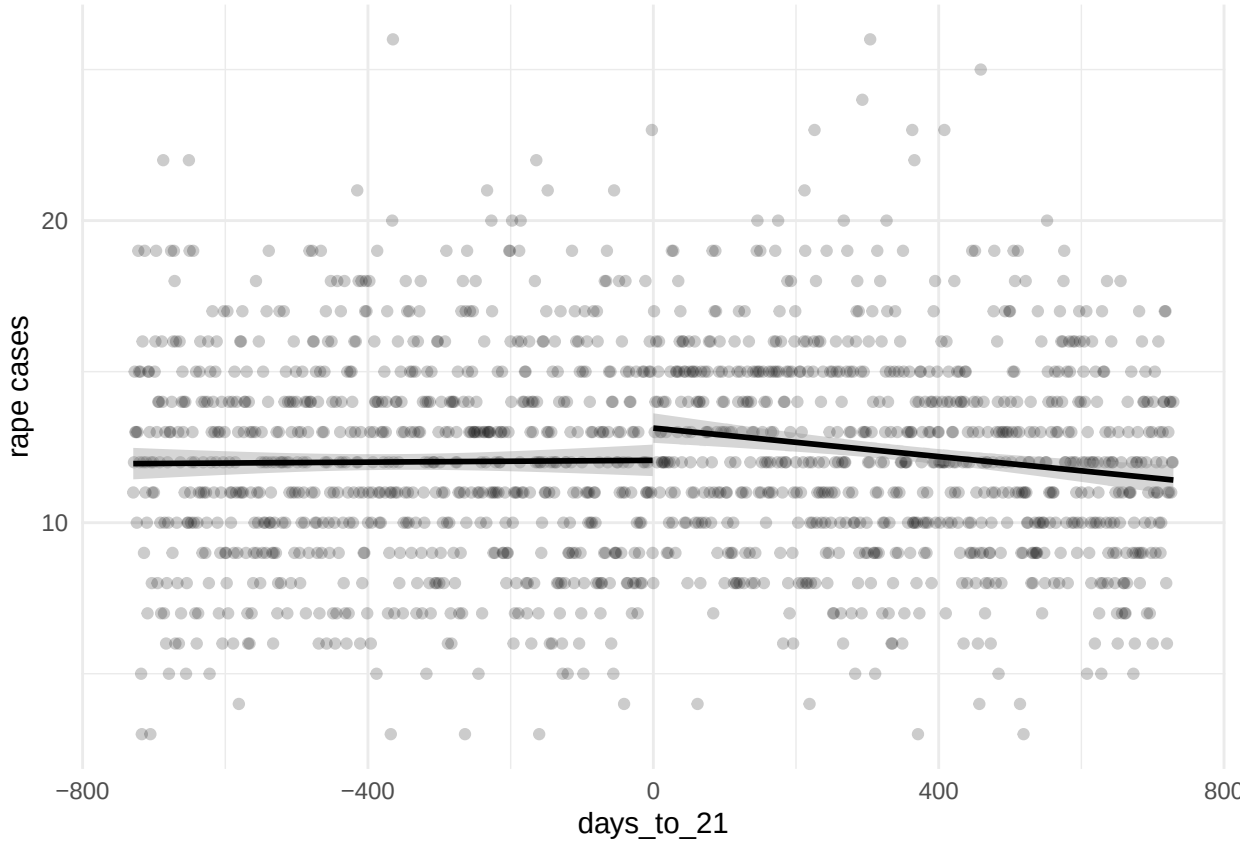

```
## # A tibble: 3 x 5
##   term         estimate std.error statistic  p.value
##   <chr>           <dbl>     <dbl>     <dbl>    <dbl>
## 1 (Intercept) 1574.       4.34       363.   0.
## 2 days_to_21    -0.0341   0.00460     -7.42 1.55e-13
## 3 under_21     -93.5      7.76       -12.1  1.09e-32
```

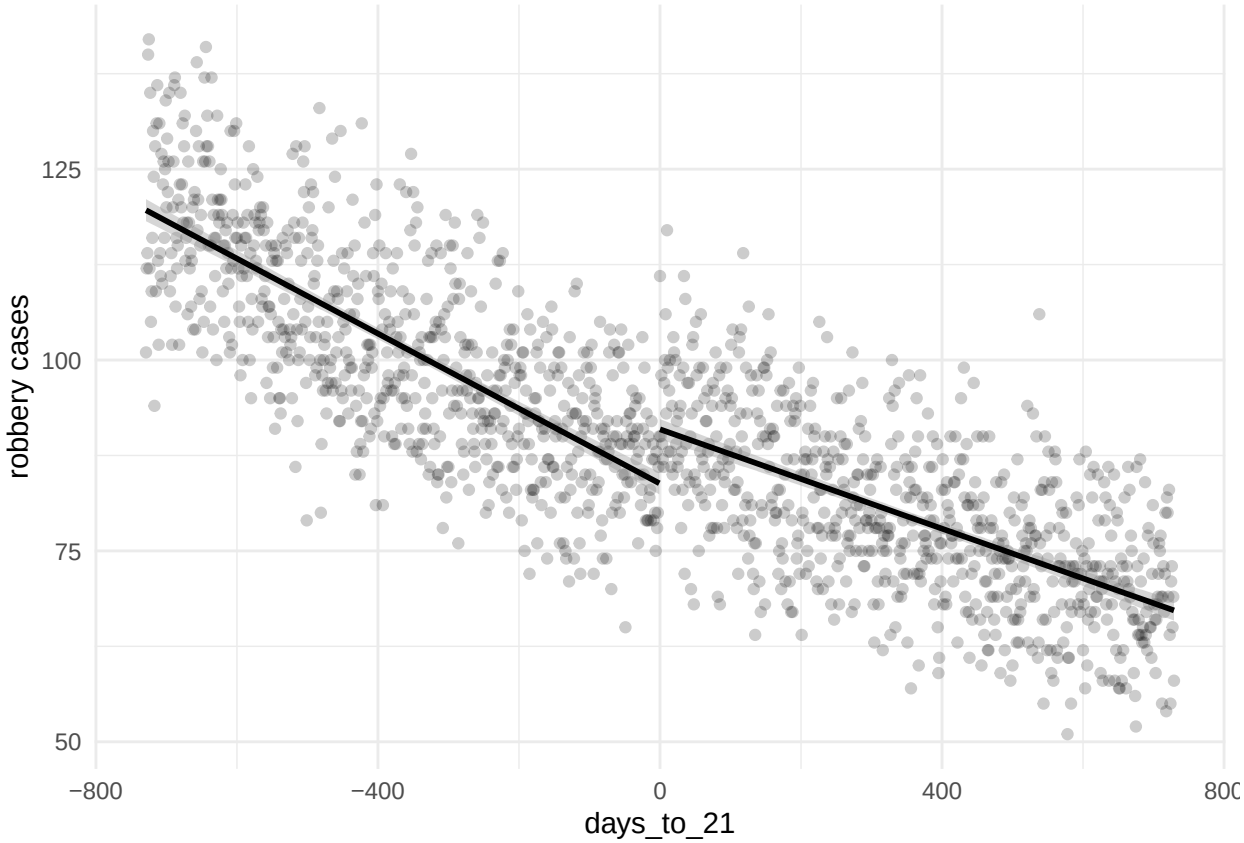
# Violent cases

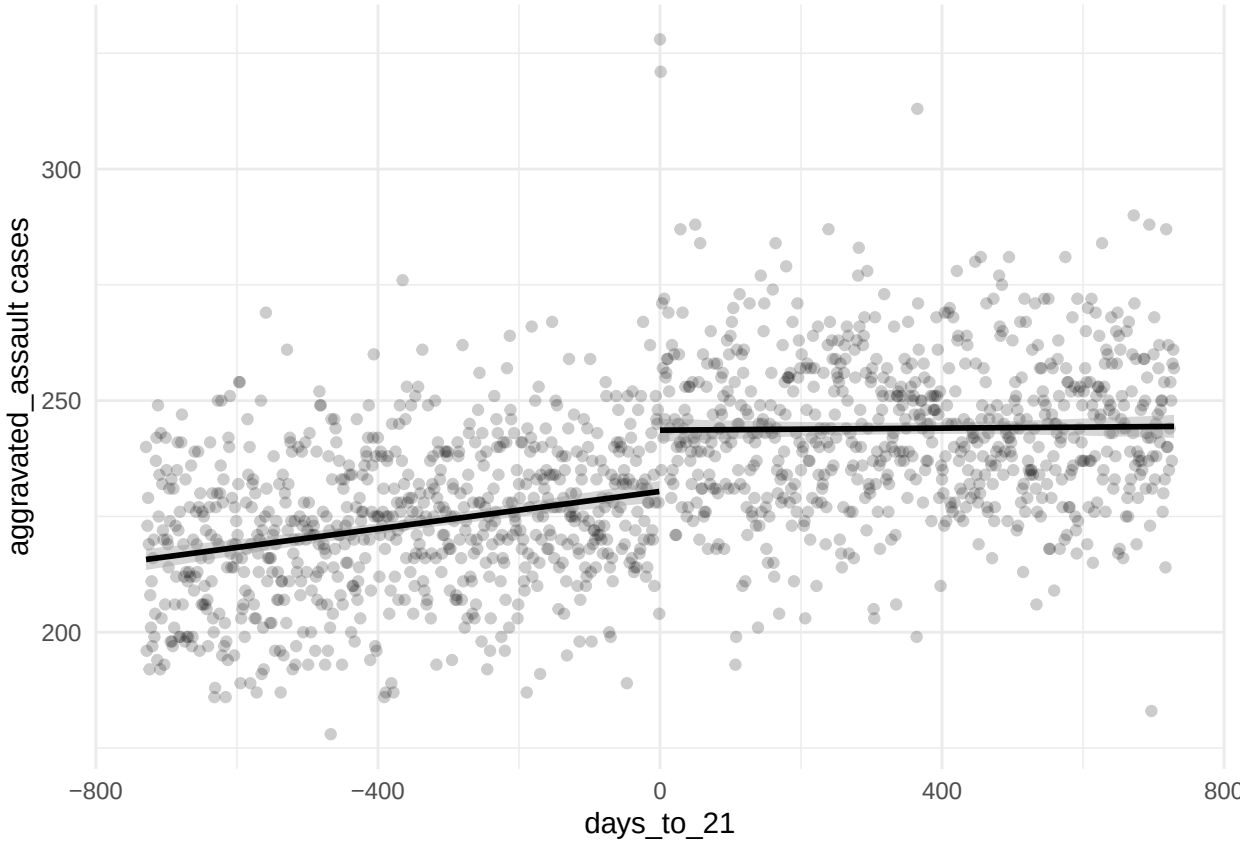
```
##   days_to_21 murder manslaughter rape robbery aggravated_assault
## 1       -729     18            3   11     101                240
## 2       -728     10            2   12     112                196
## 3       -727     19            0   15     114                223
## 4       -726      7            1   13     140                229
## 5       -725     17            0   13     142                219
## 6       -724     11            0   10     112                192
##   ot_assault under_21
## 1        215        0
## 2        218        0
## 3        204        0
## 4        210        0
## 5        206        0
## 6        186        0
```

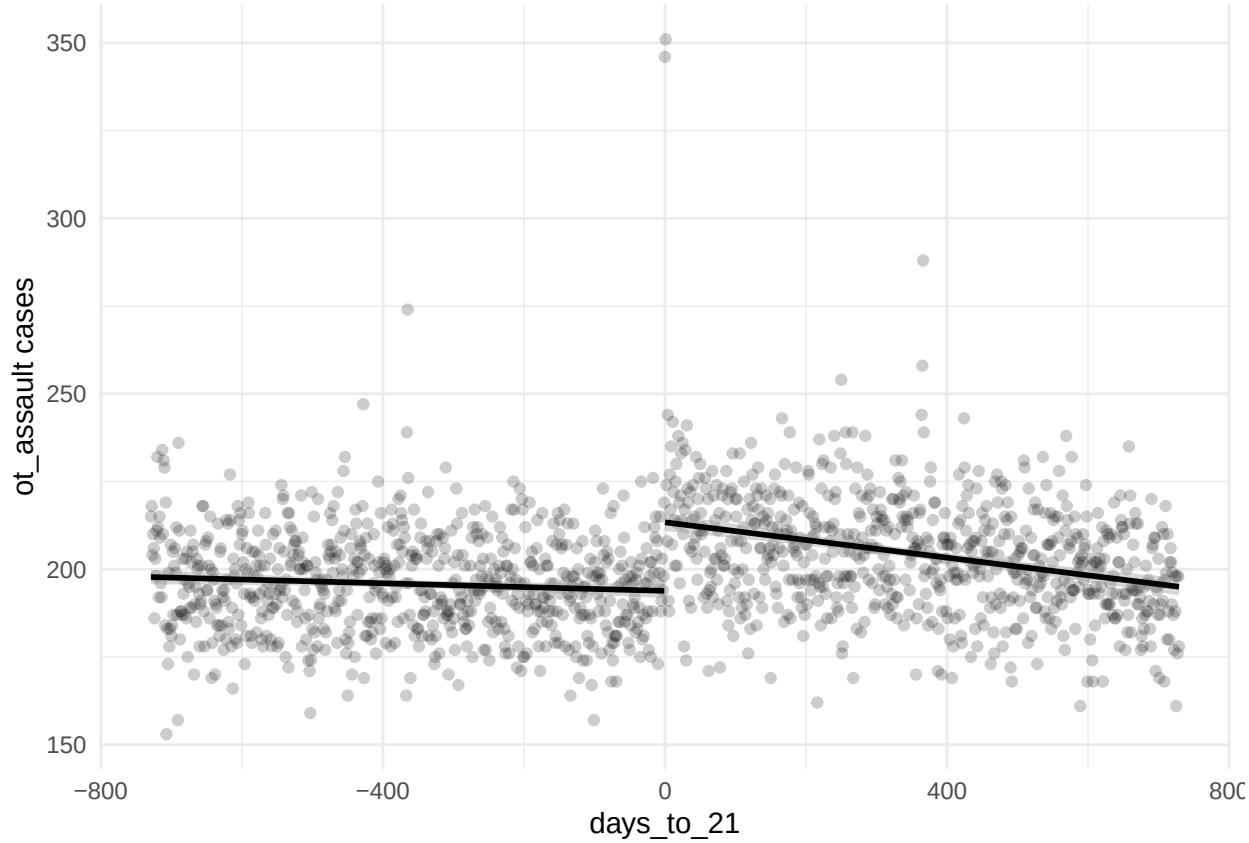

murder cases
20
15
10
5
0
−800
−400
0
400
800
days_to_21


manslaughter cases
6
4
2
0
−800
−400
0
400
800
days_to_21


20
10
rape cases
−800
−400
0
400
800
days_to_21

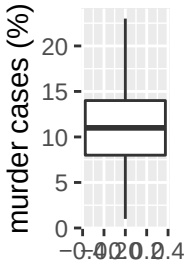

125
100
75
50
robbery cases
−800
−400
0
400
800
days_to_21

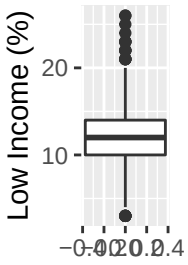

aggravated_assault cases
300
250
200
−800
−400
0
400
800
days_to_21

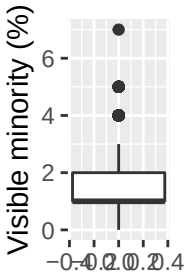



```
## # A tibble: 3 x 5
##   term          estimate std.error statistic  p.value
##   <chr>            <dbl>     <dbl>     <dbl>    <dbl>
## 1 (Intercept)  1.44      0.0671      21.4   9.20e-89
## 2 days_to_21  -0.000189  0.000143    -1.32  1.86e- 1
## 3 under_21    -0.0151    0.120       -0.126 9.00e- 1
```

```
## # A tibble: 3 x 5
##   term        estimate std.error statistic p.value
##   <chr>          <dbl>     <dbl>     <dbl>   <dbl>
## 1 (Intercept) 11.6      0.204        56.8  0
## 2 days_to_21  -0.00111  0.000433     -2.56 0.0107
## 3 under_21     1.07     0.365         2.93 0.00340
```

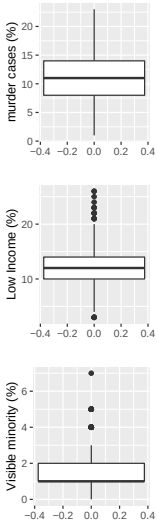



```
## TableGrob (3 x 1) "arrange": 3 grobs
##   z     cells    name           grob
## 1 1 (1-1,1-1) arrange gtable[layout]
## 2 2 (2-2,1-1) arrange gtable[layout]
## 3 3 (3-3,1-1) arrange gtable[layout]
```

| days_to_21 | murder | manslaughter | rape | robbery | aggravated_assault | ot_assault | under_21 | stats |
|---|---|---|---|---|---|---|---|---|
| 1459.00 | 1459.00 | 1459.00 | 1459.00 | 1459.00 | 1459.00 | 1459.00 | 1459.00 | nbr.val |
| 1.00 | 0.00 | 327.00 | 0.00 | 0.00 | 0.00 | 0.00 | 729.00 | nbr.null |
| 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | nbr.na |
| −729.00 | 1.00 | 0.00 | 3.00 | 51.00 | 178.00 | 153.00 | 0.00 | min |
| 729.00 | 23.00 | 7.00 | 26.00 | 142.00 | 328.00 | 351.00 | 1.00 | max |
| 1458.00 | 22.00 | 7.00 | 23.00 | 91.00 | 150.00 | 198.00 | 1.00 | range |
| 0.00 | 16243.00 | 2088.00 | 17714.00 | 131889.00 | 340741.00 | 291800.00 | 730.00 | sum |
| 0.00 | 11.00 | 1.00 | 12.00 | 89.00 | 233.00 | 199.00 | 1.00 | median |
| 0.00 | 11.13 | 1.43 | 12.14 | 90.40 | 233.54 | 200.00 | 0.50 | mean |
| 11.03 | 0.09 | 0.03 | 0.09 | 0.45 | 0.51 | 0.44 | 0.01 | SE.mean |
| 21.64 | 0.19 | 0.06 | 0.18 | 0.89 | 1.01 | 0.85 | 0.03 | CI.mean.0.95 |
| 177511.67 | 13.10 | 1.32 | 12.21 | 297.07 | 384.81 | 276.26 | 0.25 | var |
| 421.32 | 3.62 | 1.15 | 3.49 | 17.24 | 19.62 | 16.62 | 0.50 | std.dev |
| Inf | 0.33 | 0.80 | 0.29 | 0.19 | 0.08 | 0.08 | 1.00 | coef.var |





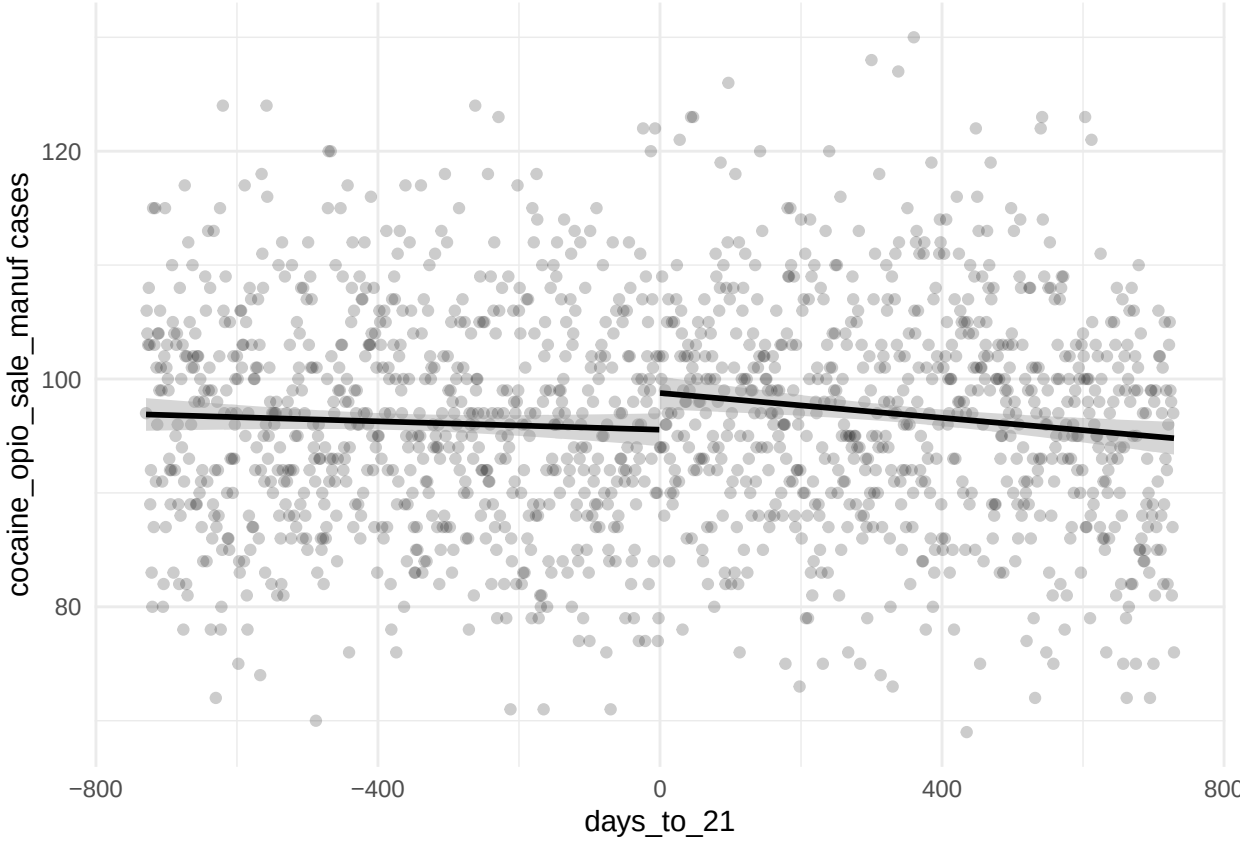
| days_to_21 | murder | manslaughter | rape | robbery | aggravated_assault | ot_assault | under_21 | stats |
|---|---|---|---|---|---|---|---|---|
| 1459.00 | 1459.00 | 1459.00 | 1459.00 | 1459.00 | 1459.00 | 1459.00 | 1459.00 | nbr.val |
| 1.00 | 0.00 | 327.00 | 0.00 | 0.00 | 0.00 | 0.00 | 729.00 | nbr.null |
| 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | nbr.na |
| −729.00 | 1.00 | 0.00 | 3.00 | 51.00 | 178.00 | 153.00 | 0.00 | min |
| 729.00 | 23.00 | 7.00 | 26.00 | 142.00 | 328.00 | 351.00 | 1.00 | max |
| 1458.00 | 22.00 | 7.00 | 23.00 | 91.00 | 150.00 | 198.00 | 1.00 | range |
| 0.00 | 16243.00 | 2088.00 | 17714.00 | 131889.00 | 340741.00 | 291800.00 | 730.00 | sum |
| 0.00 | 11.00 | 1.00 | 12.00 | 89.00 | 233.00 | 199.00 | 1.00 | median |
| 0.00 | 11.13 | 1.43 | 12.14 | 90.40 | 233.54 | 200.00 | 0.50 | mean |
| 11.03 | 0.09 | 0.03 | 0.09 | 0.45 | 0.51 | 0.44 | 0.01 | SE.mean |
| 21.64 | 0.19 | 0.06 | 0.18 | 0.89 | 1.01 | 0.85 | 0.03 | CI.mean.0.95 |
| 177511.67 | 13.10 | 1.32 | 12.21 | 297.07 | 384.81 | 276.26 | 0.25 | var |
| 421.32 | 3.62 | 1.15 | 3.49 | 17.24 | 19.62 | 16.62 | 0.50 | std.dev |
| Inf | 0.33 | 0.80 | 0.29 | 0.19 | 0.08 | 0.08 | 1.00 | coef.var |

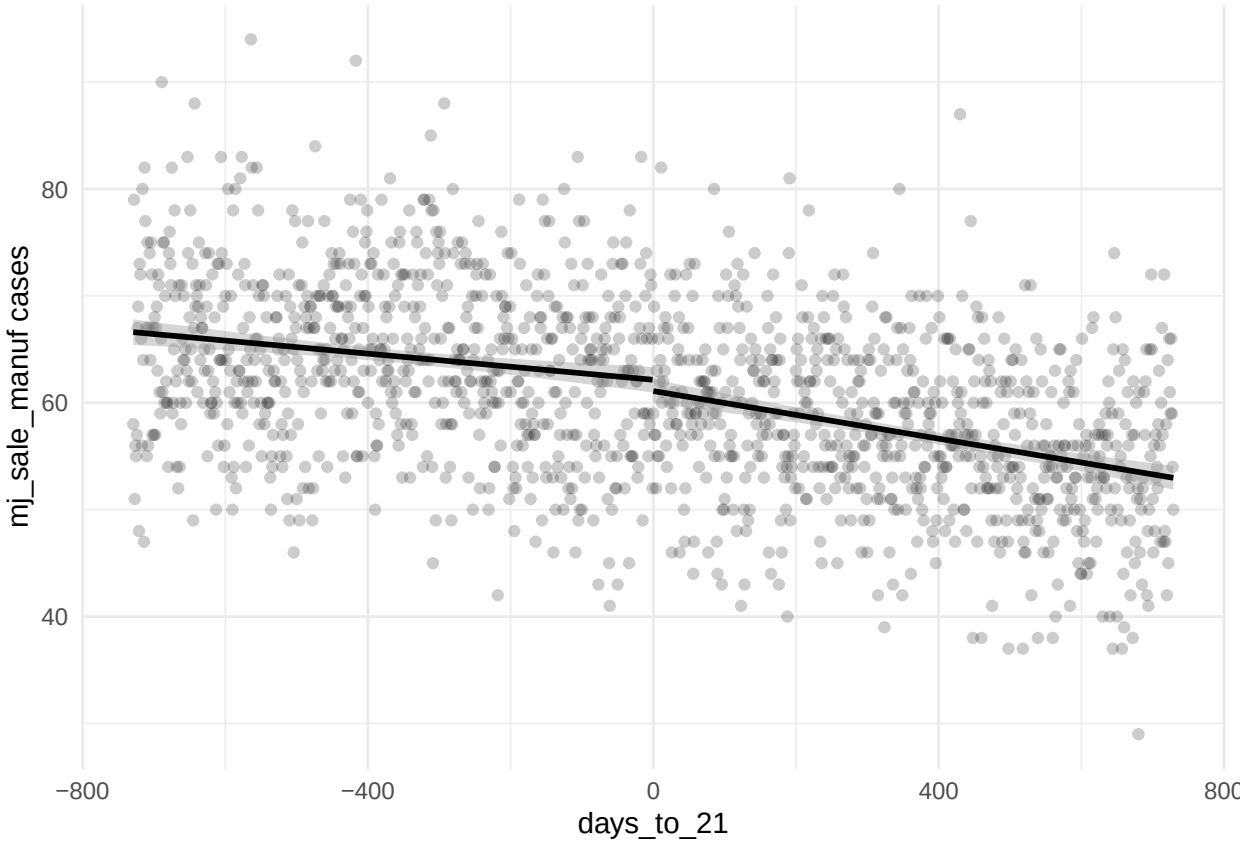

Figure 1: Descriptive statistics of the data

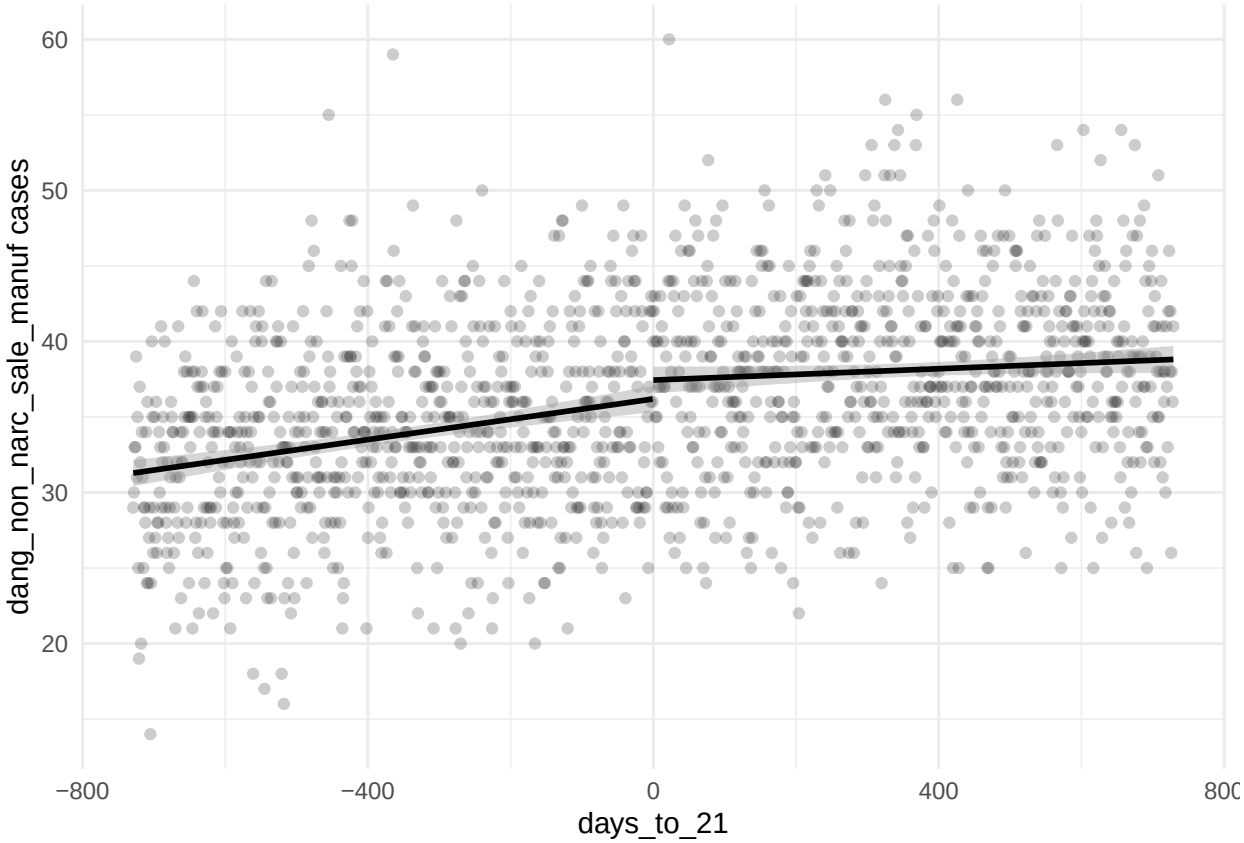
# Drugs cases

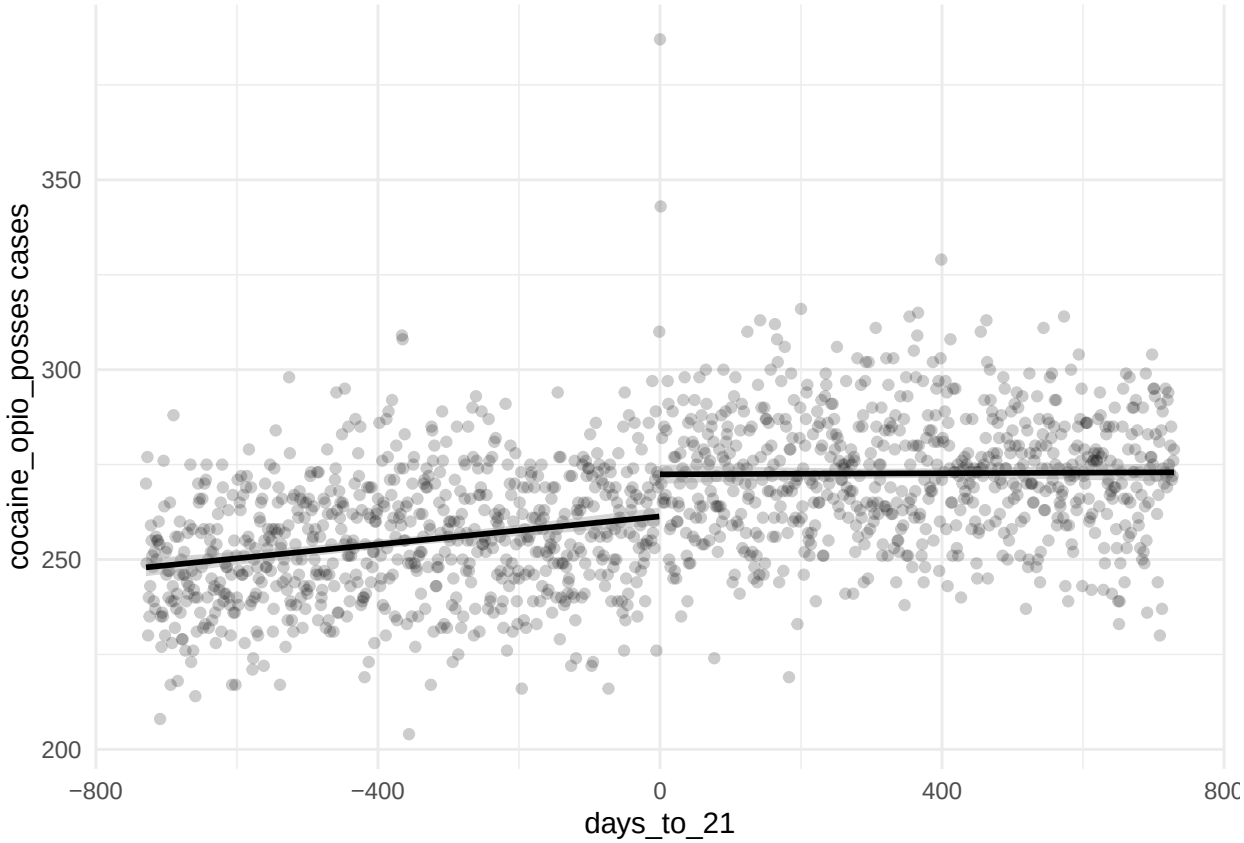
```
##   days_to_21 cocaine_opio_sale_manuf mj_sale_manuf
## 1       -729                      97           58
## 2       -728                     106           79
## 3       -727                     104           51
## 4       -726                     103           56
## 5       -725                     108           55
## 6       -724                     103           57
##   dang_non_narc_sale_manuf cocaine_opio_posses mj_posses
## 1                       29                 270       264
## 2                       30                 249       252
## 3                       33                 277       265
```

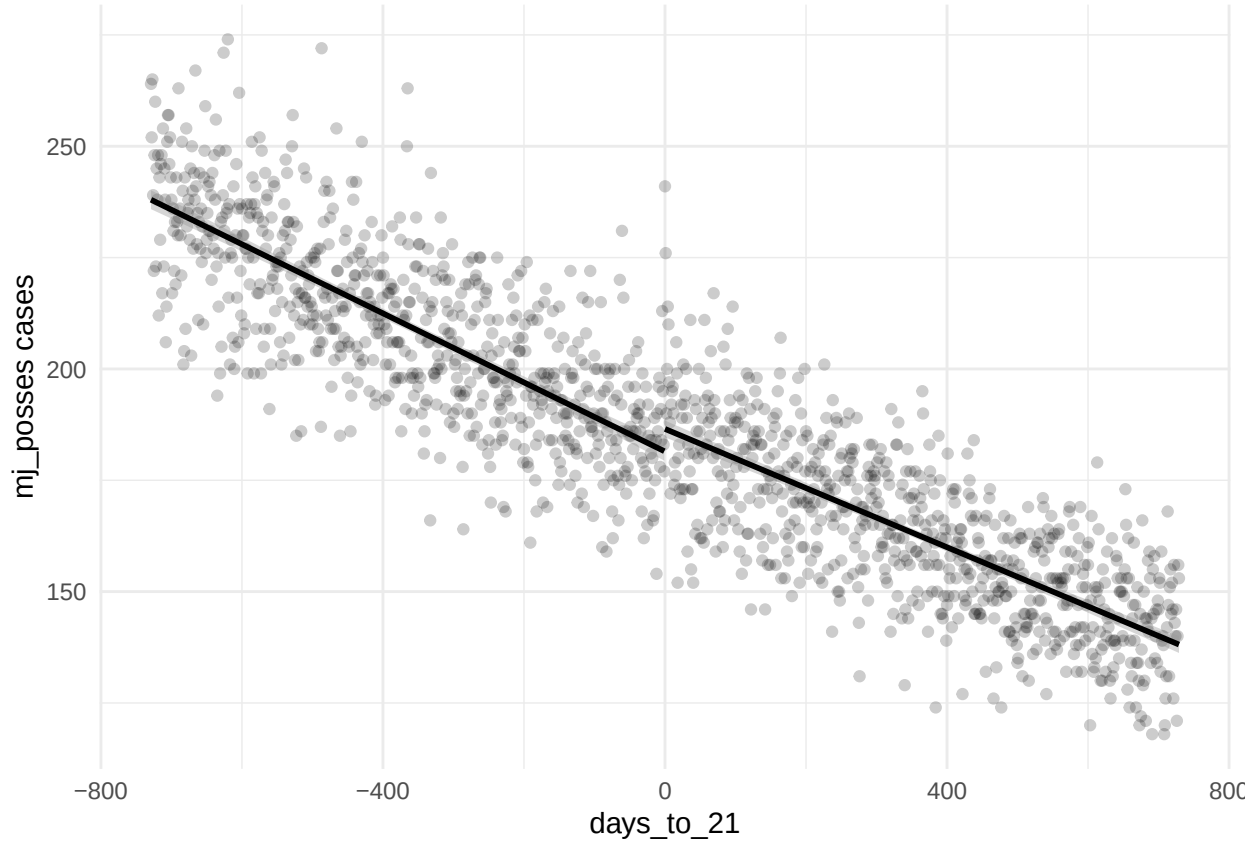
```
## 4                       33                230        239
## 5                       39                240        222
## 6                       31                235        248
##   dang_non_narc_posses under_21
## 1                  170        0
## 2                  157        0
## 3                  139        0
## 4                  151        0
## 5                  158        0
## 6                  165        0
```

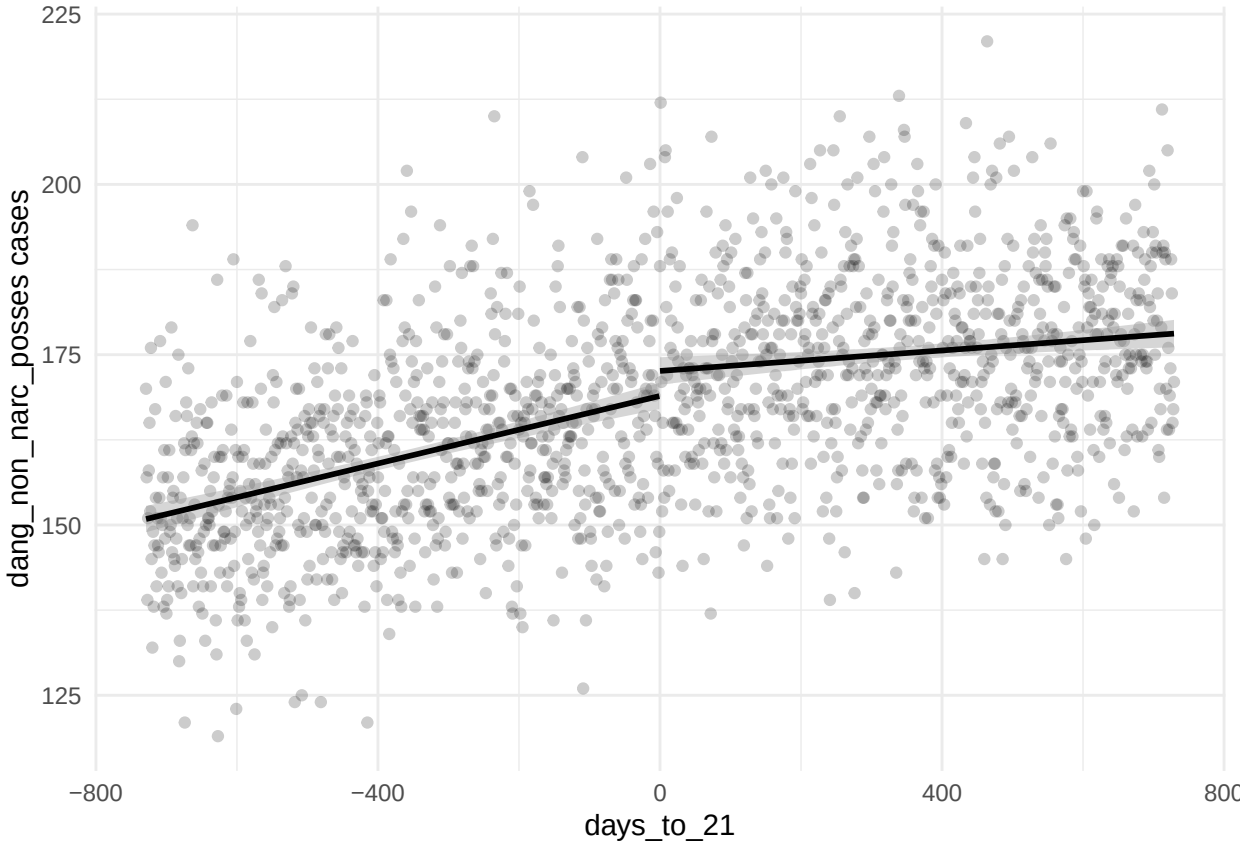



mj_sale_manuf cases
80
60
40
−800
−400
0
400
800
days_to_21


60
50
40
30
20
dang_non_narc_sale_manuf cases
−800
−400
0
400
800
days_to_21

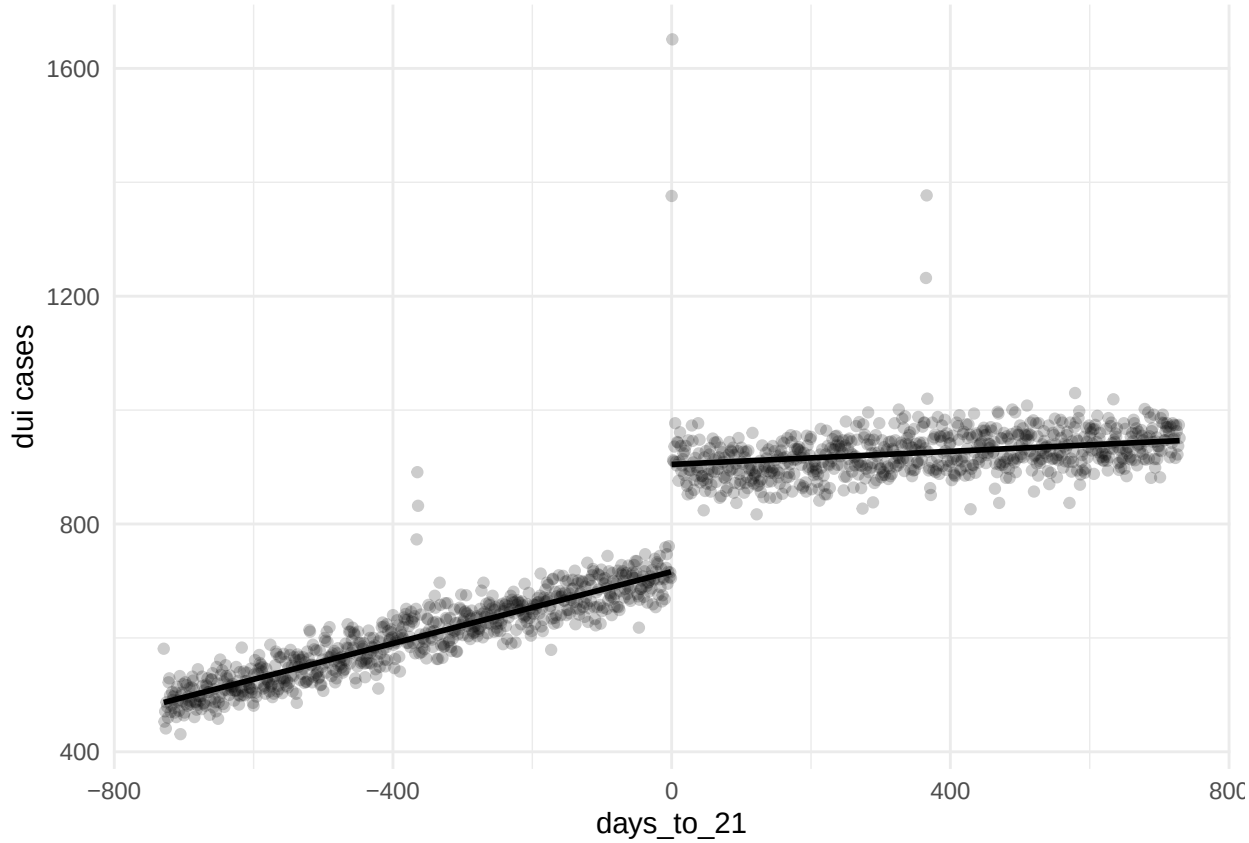

cocaine_opio_posses cases
350
300
250
200
−800
−400
0
400
800
days_to_21

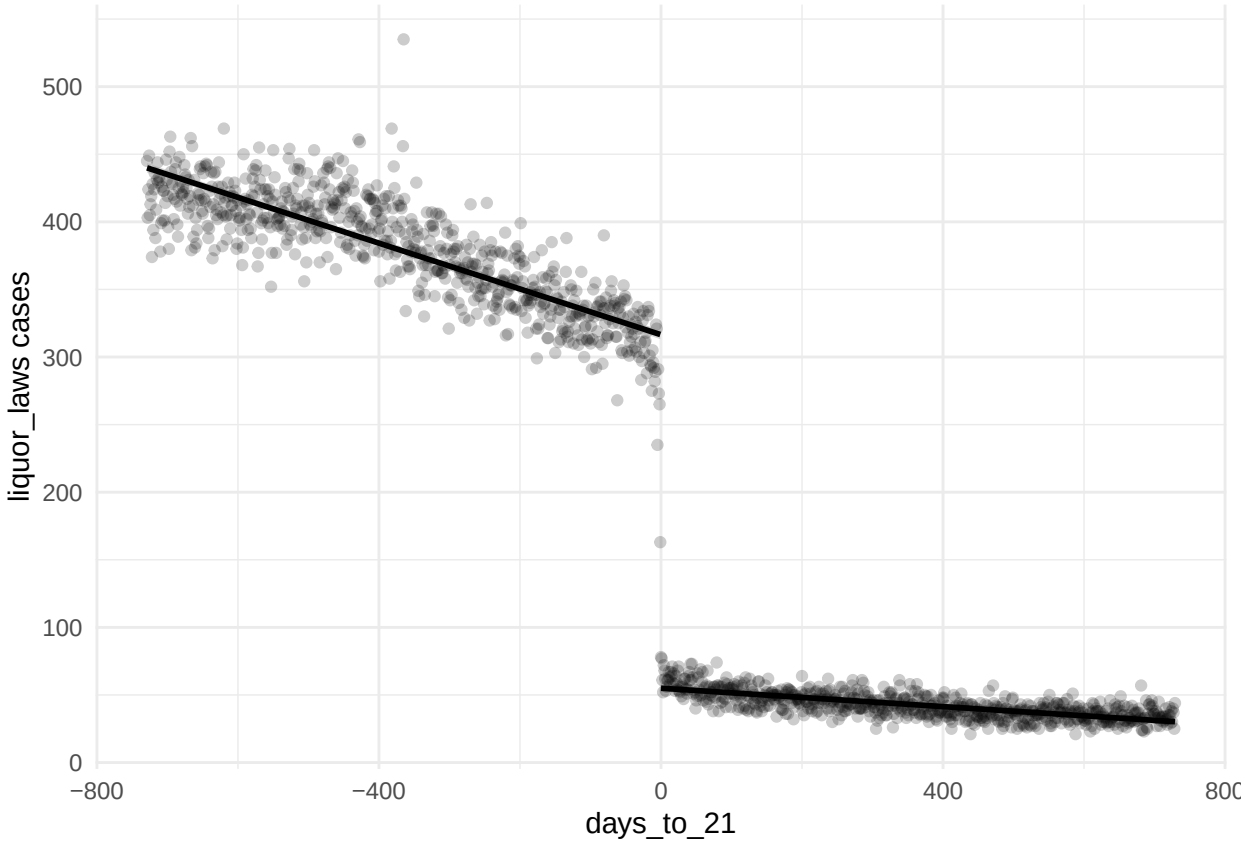

250
200
150
mj_posses cases
−800
−400
0
400
800
days_to_21

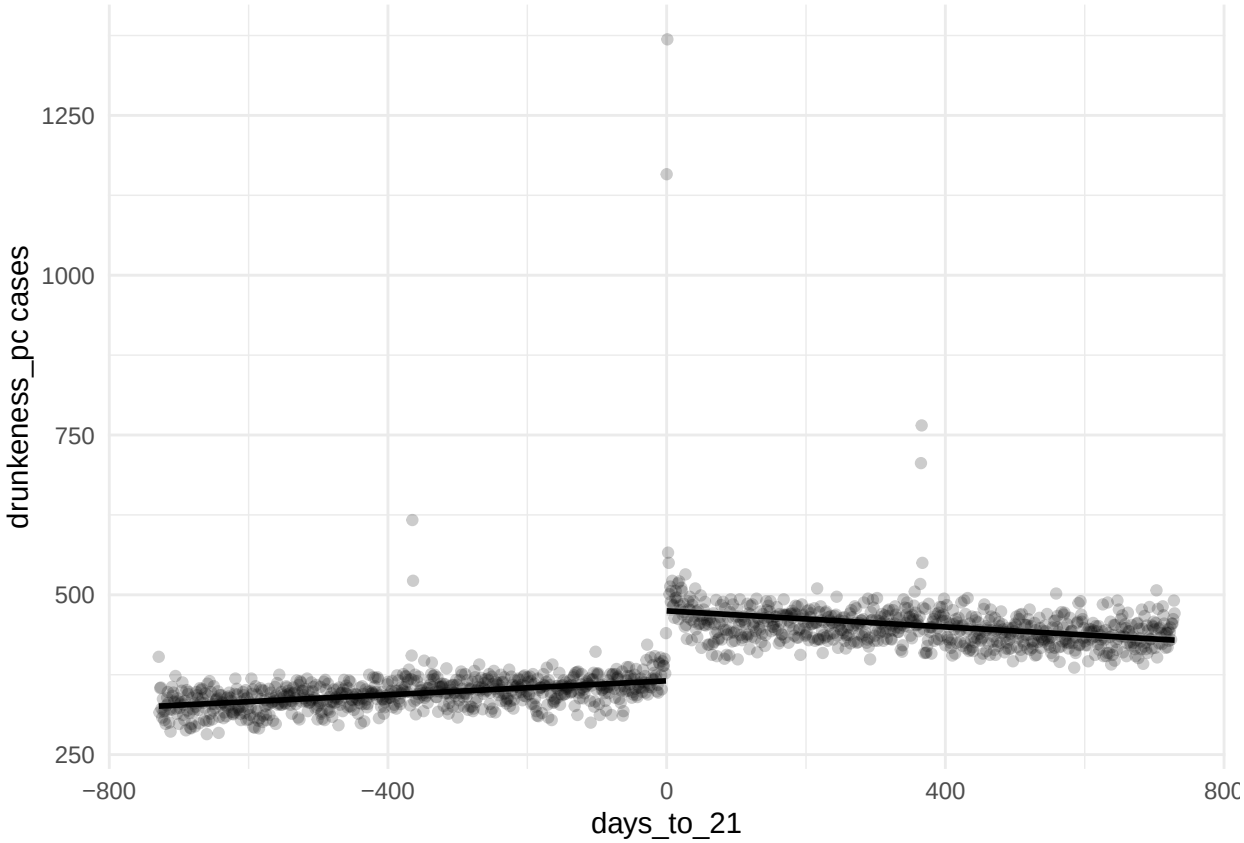


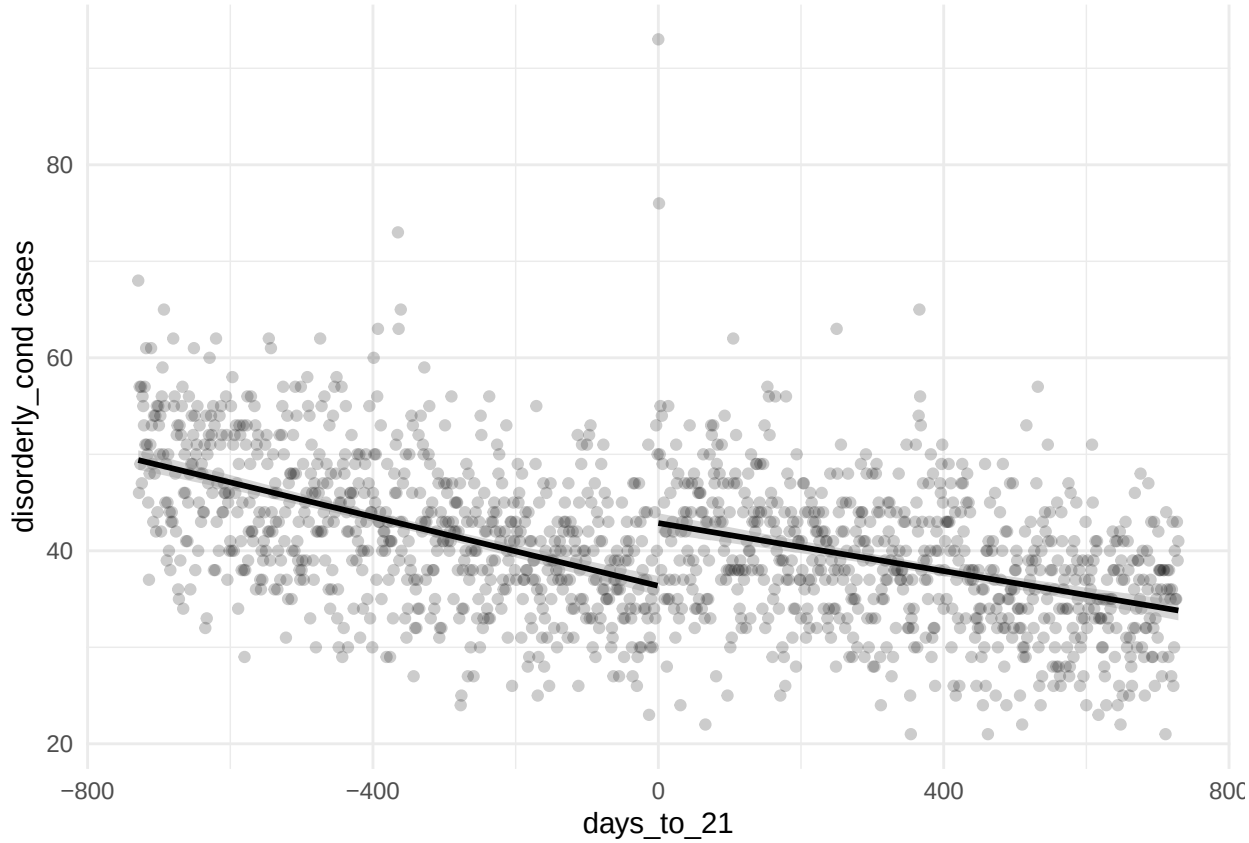

## Alcohol Cases

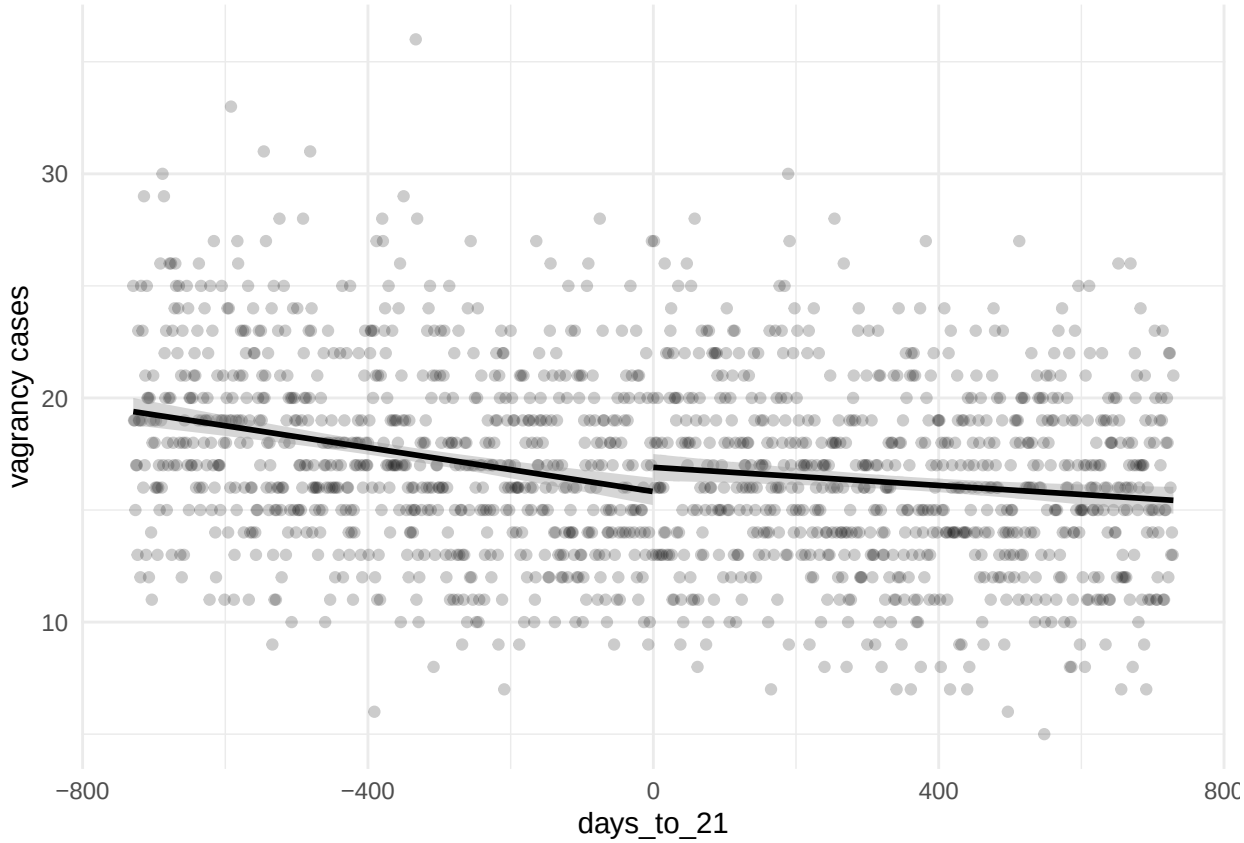
```
##   days_to_21 dui liquor_laws drunkeness_pc disorderly_cond vagrancy
## 1       -729 581         445           403              68       25
## 2       -728 453         403           316              46       19
## 3       -727 471         424           354              57       19
## 4       -726 441         449           355              49       15
## 5       -725 487         405           322              57       17
## 6       -724 479         413           308              47       17
##   under_21
## 1        0
## 2        0
## 3        0
## 4        0
## 5        0
## 6        0
```


1600
1200
800
400
dui cases
−800
−400
0
400
800
days_to_21


500
400
300
200
100
0
liquor_laws cases
−800
−400
0
400
800
days_to_21


drunkeness_pc cases
1250
1000
750
500
250
−800
−400
0
400
800
days_to_21

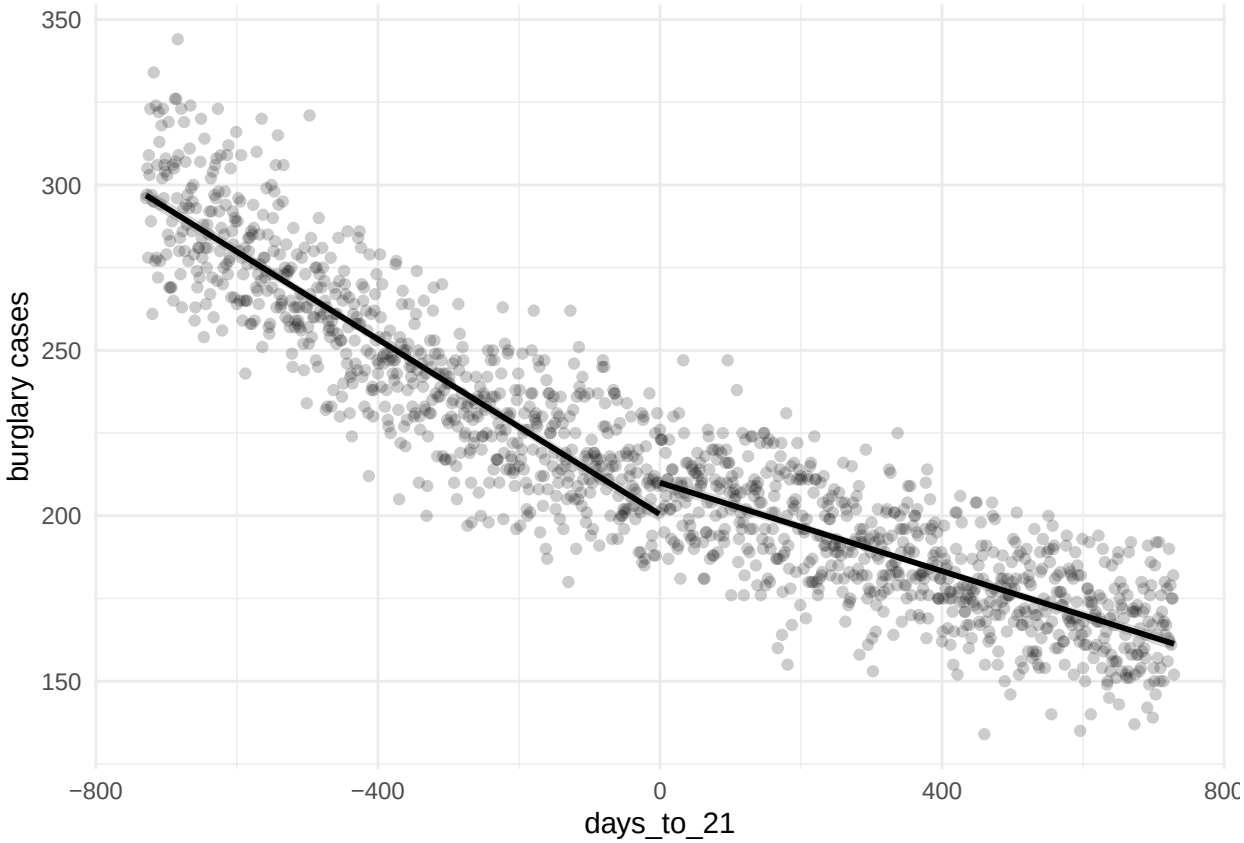

80
60
40
20
disorderly_cond cases
−800
−400
0
400
800
days_to_21

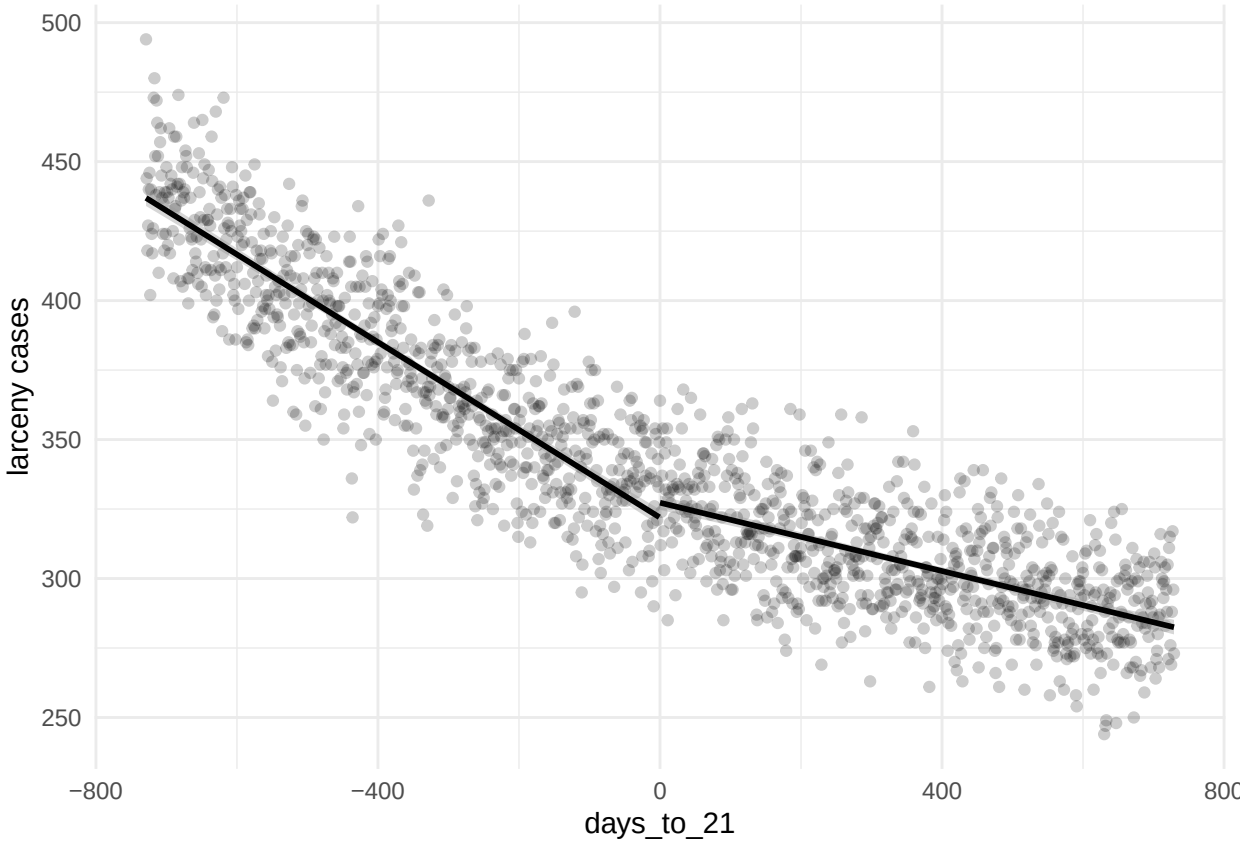


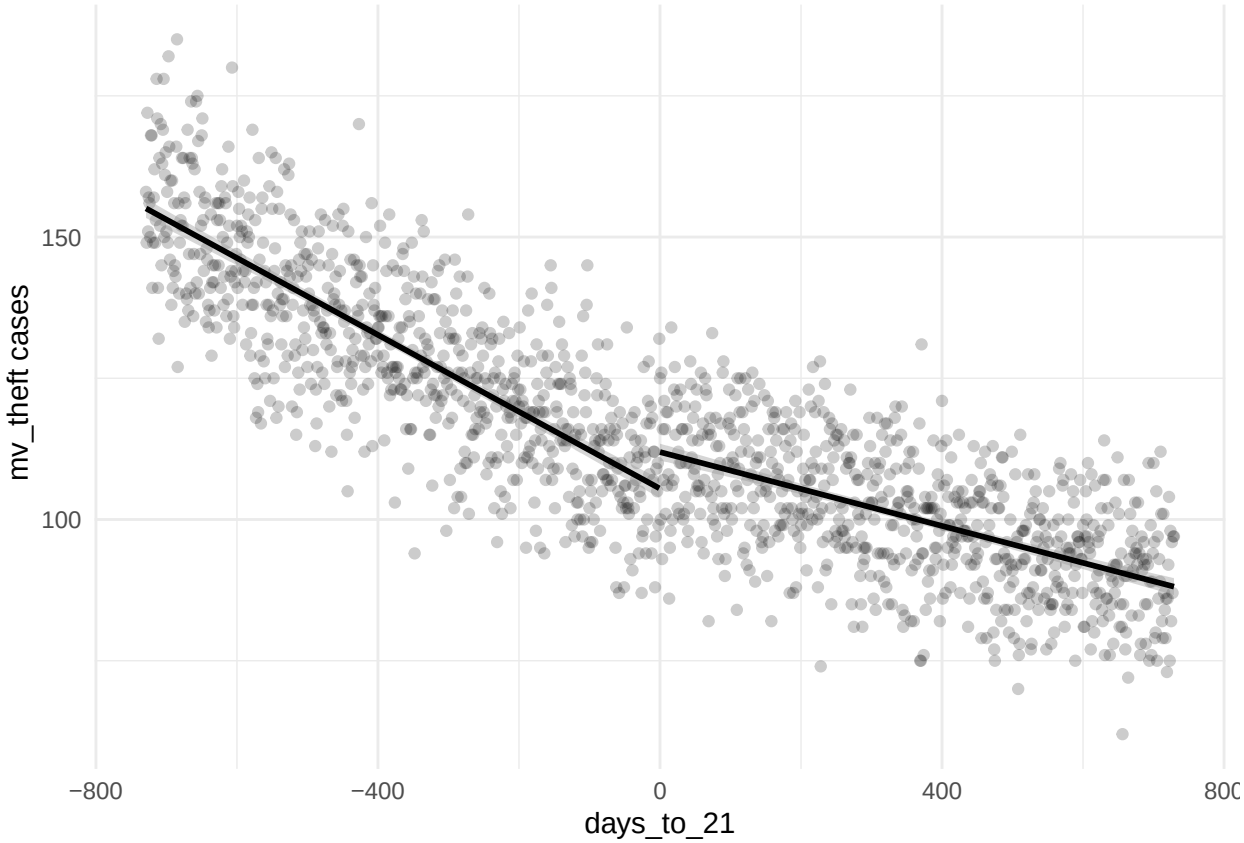

```
## # A tibble: 3 x 5
##   term        estimate std.error statistic  p.value
##   <chr>          <dbl>     <dbl>     <dbl>    <dbl>
## 1 (Intercept) 16.4      0.243        67.3  0.
## 2 days_to_21  -0.00345  0.000516     -6.69 3.13e-11
## 3 under_21     1.08     0.435         2.48 1.34e- 2
```

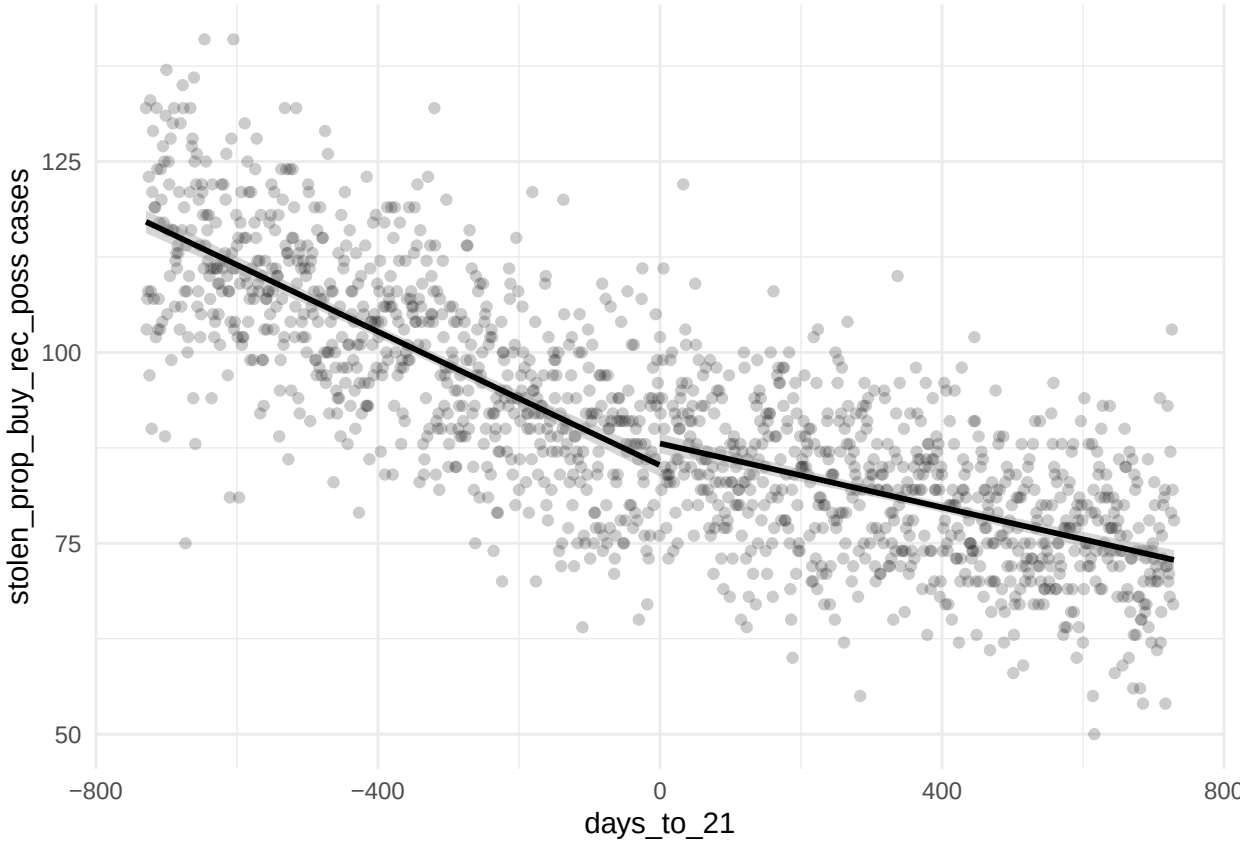
## Property cases

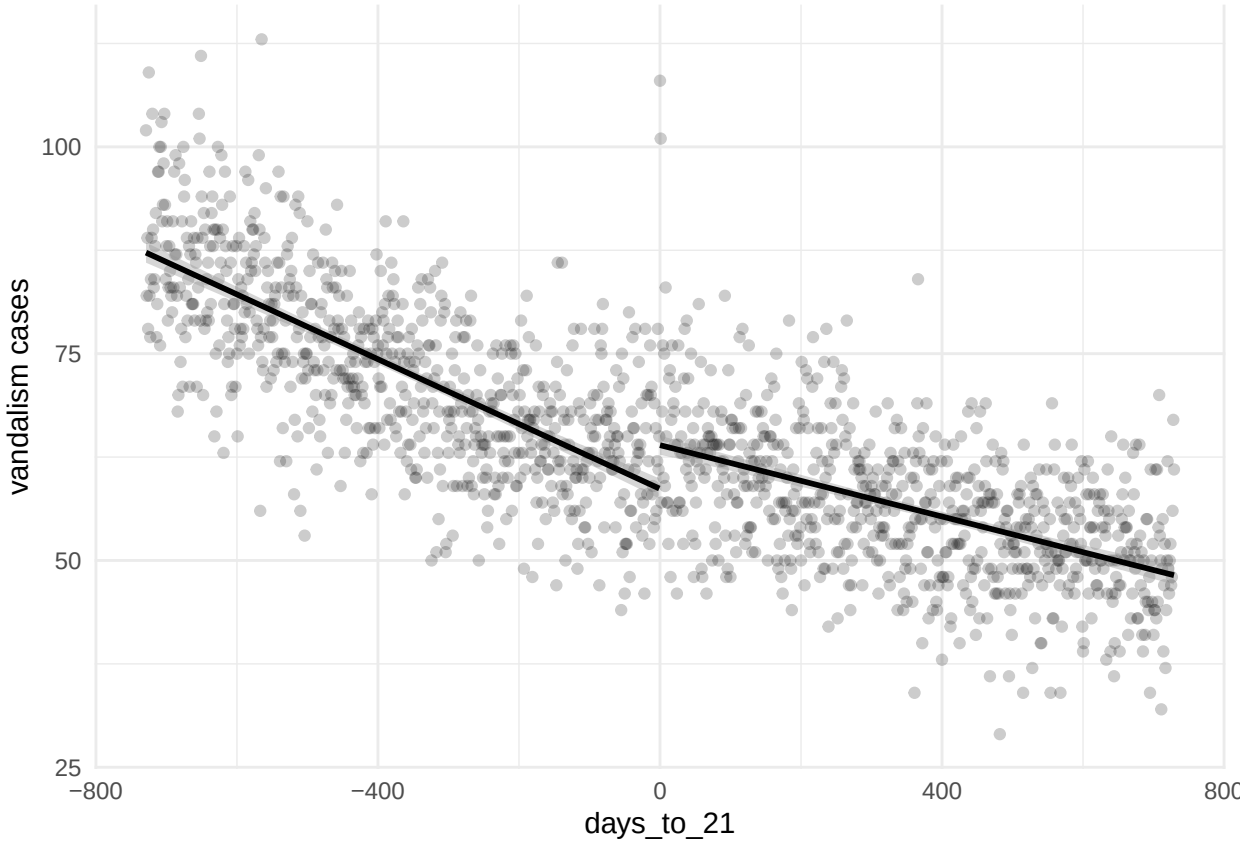
```
##   days_to_21 burglary larceny mv_theft stolen_prop_buy_rec_poss vandalism
## 1       -729      296     494      158                       132       102
## 2       -728      297     444      149                       103        82
## 3       -727      305     418      172                       107        89
## 4       -726      278     427      151                       108        78
## 5       -725      309     440      157                       123       109
## 6       -724      303     446      156                        97        82
##   under_21
## 1        0
## 2        0
## 3        0
## 4        0
## 5        0
## 6        0
```

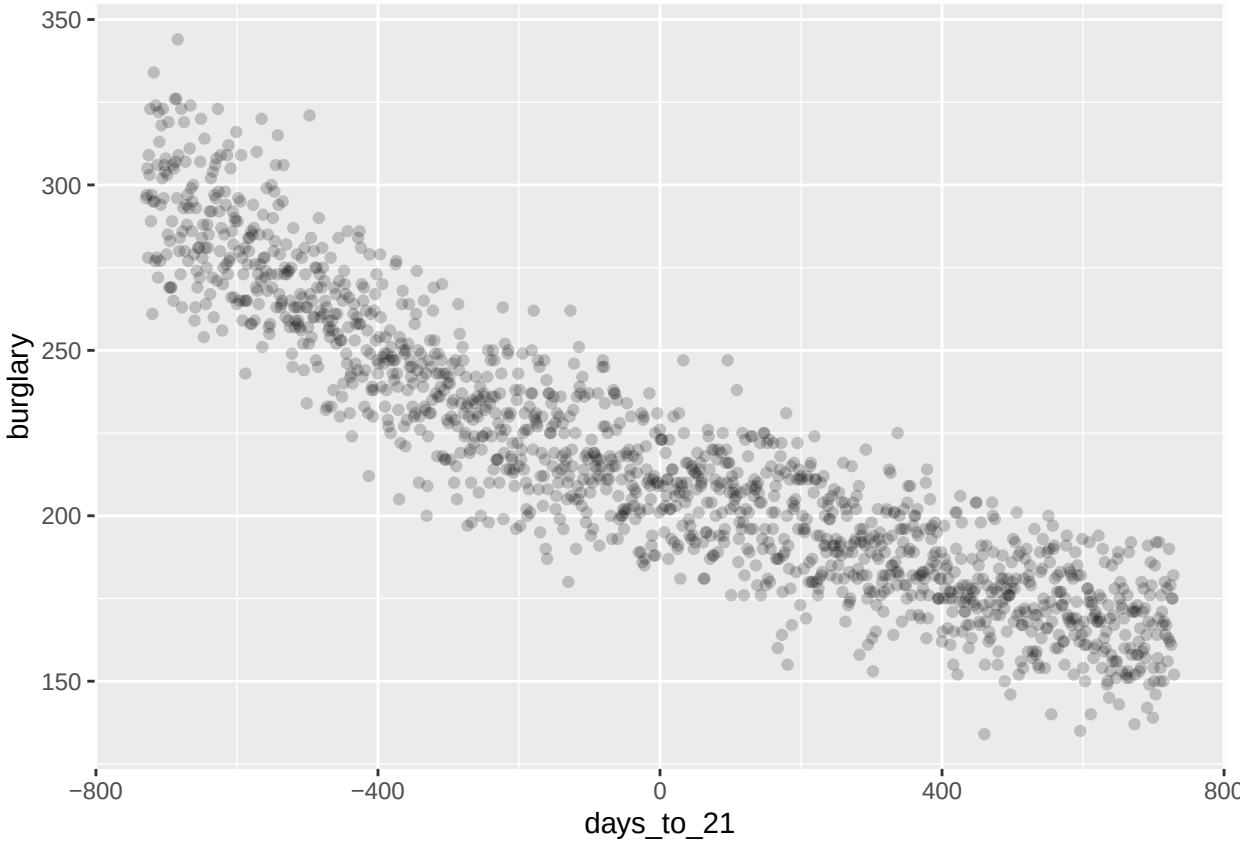

350
300
250
200
150
burglary cases
−800
−400
0
400
800
days_to_21


500
450
400
350
300
250
larceny cases
−800
−400
0
400
800
days_to_21


mv_theft cases
150
100
−800
−400
0
400
800
days_to_21


stolen_prop_buy_rec_poss cases
125
100
75
50
−800
−400
0
400
800
days_to_21


vandalism cases
100
75
50
25
−800
−400
0
400
800
days_to_21

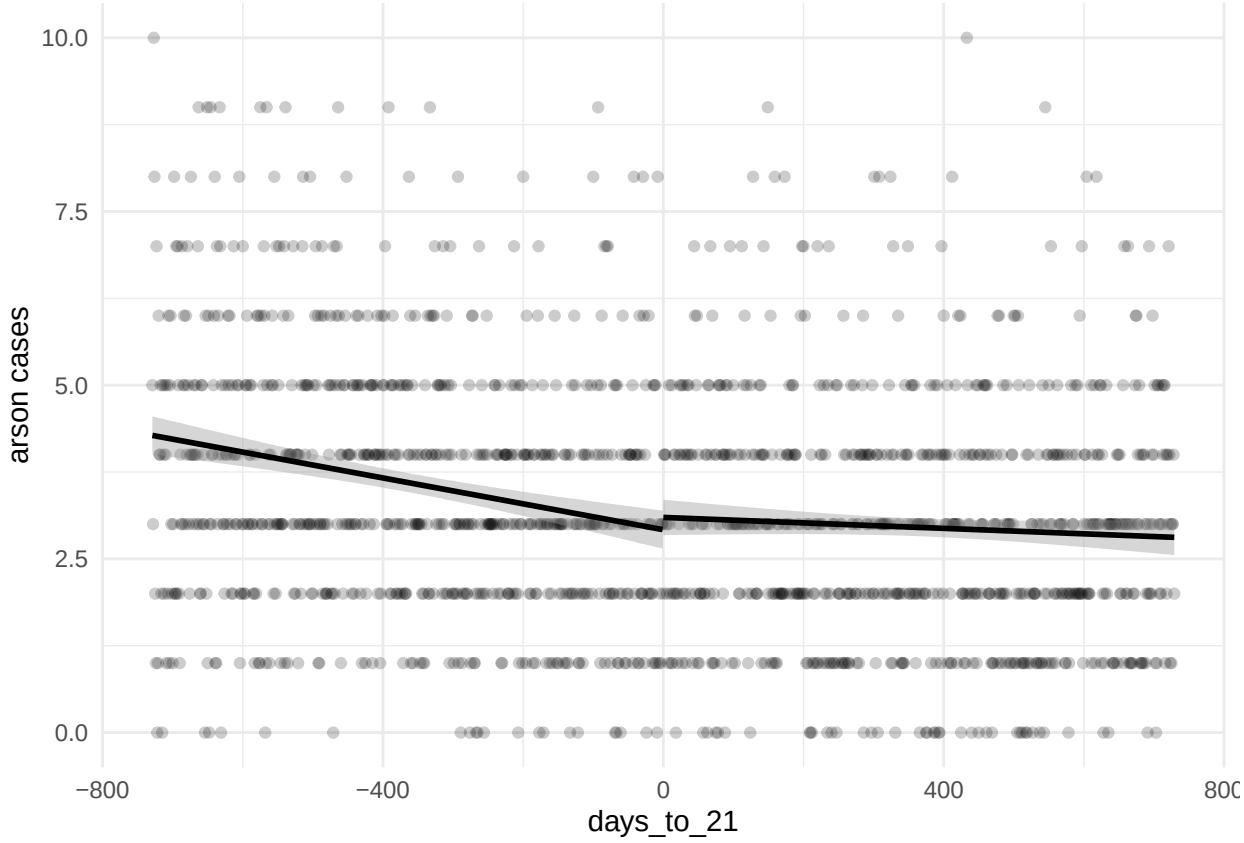


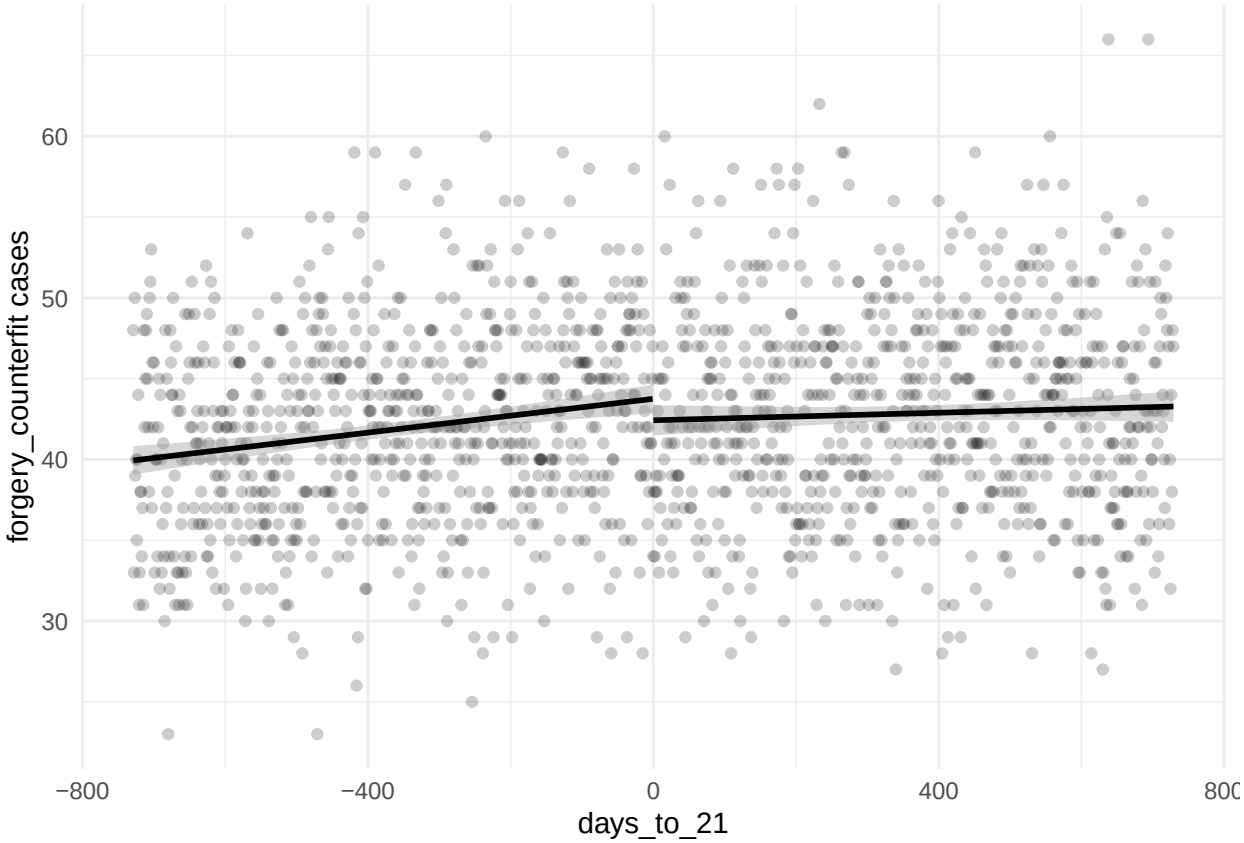

```
## # A tibble: 3 x 5
##   term        estimate std.error statistic   p.value
##   <chr>          <dbl>     <dbl>     <dbl>     <dbl>
## 1 (Intercept) 212.       0.968      219.   0.
## 2 days_to_21   -0.0995   0.00206    -48.4  1.70e-305
## 3 under_21      9.50     1.73         5.49 4.72e-  8
```

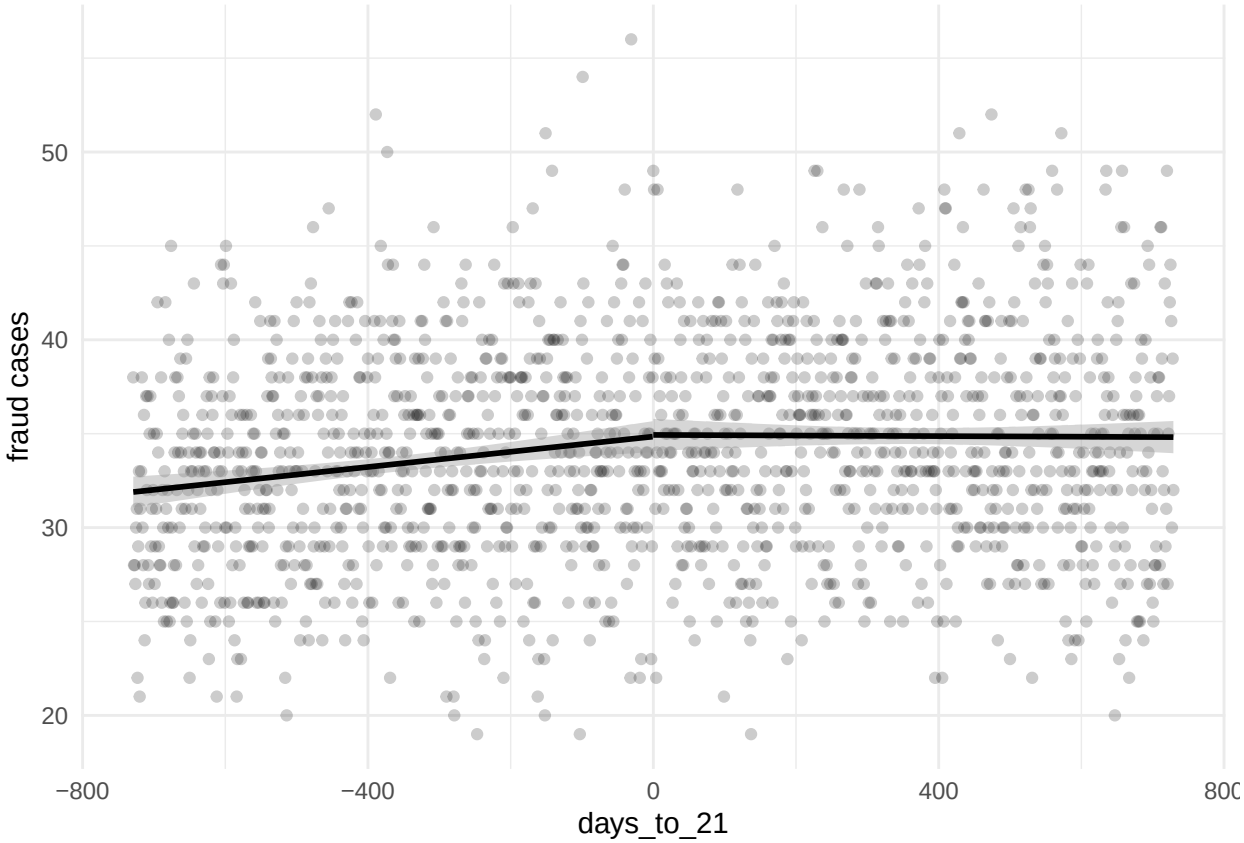
## FBI other cases

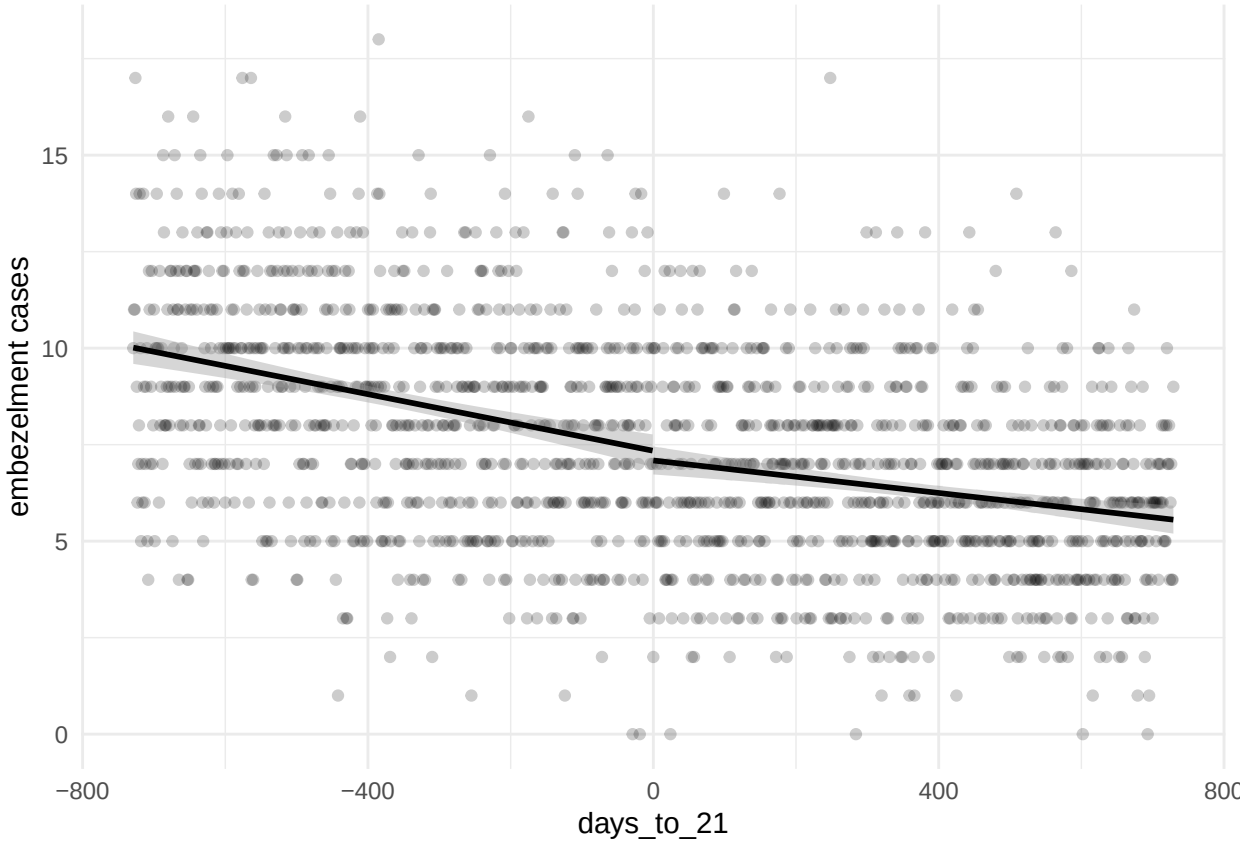
```
##   days_to_21 arson forgery_counterfit fraud embezelment weapons
## 1       -729     5                 48    38          10     129
## 2       -728     3                 33    28          11     145
## 3       -727    10                 50    28          11     157
## 4       -726     8                 39    27          17     133
## 5       -725     2                 40    30          14     132
## 6       -724     1                 35    31           9     151
##   prostitution sex_offenses bookmaking all_other_gambling aga_family_child
## 1           46           38          0                   3                1
## 2           57           38          0                   4                1
## 3           33           44          0                   0                1
## 4           42           33          1                   2                0
## 5           45           43          0                   6                0
## 6           42           40          0                   4                1
##   curfew_loitering runaways under_21
```

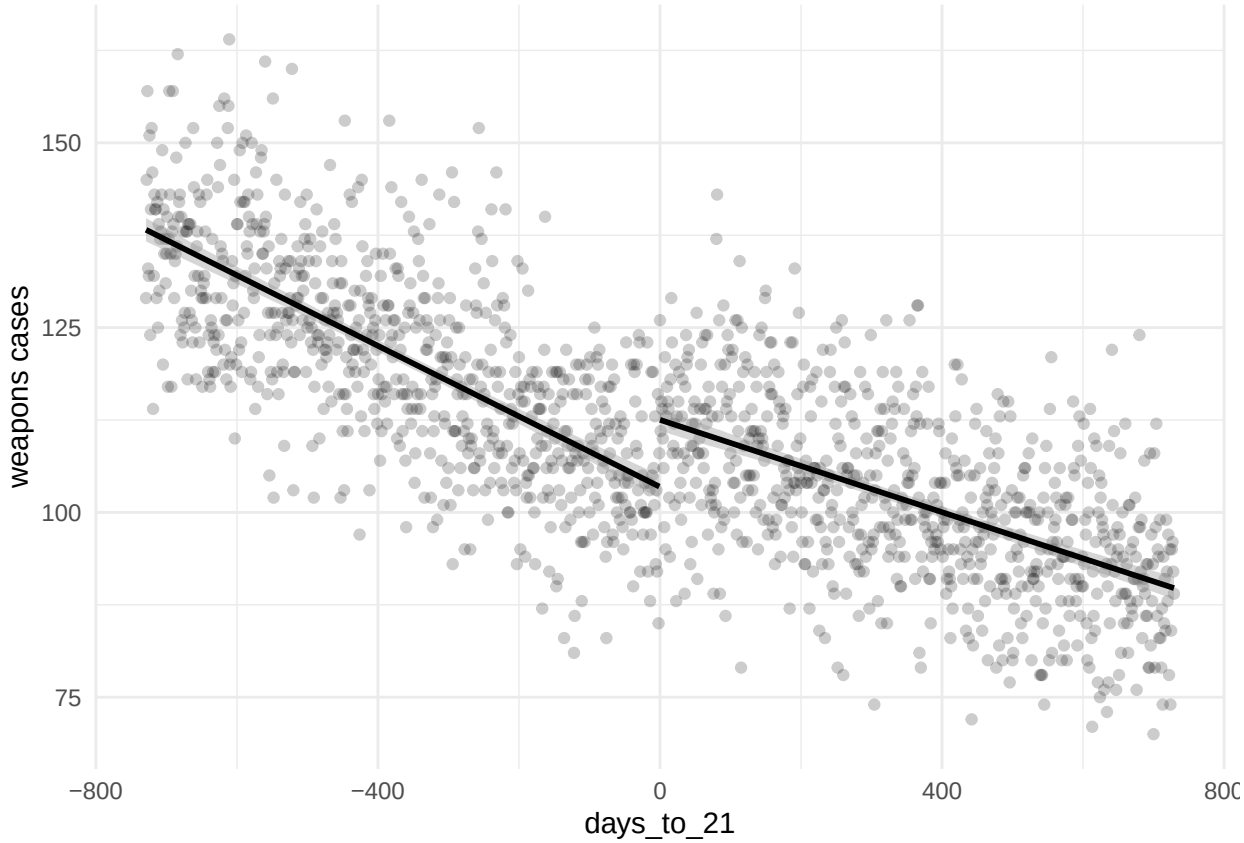
```
## 1                 0         0         0
## 2                 0         0         0
## 3                 0         0         0
## 4                 0         0         0
## 5                 0         0         0
## 6                 0         0         0
```

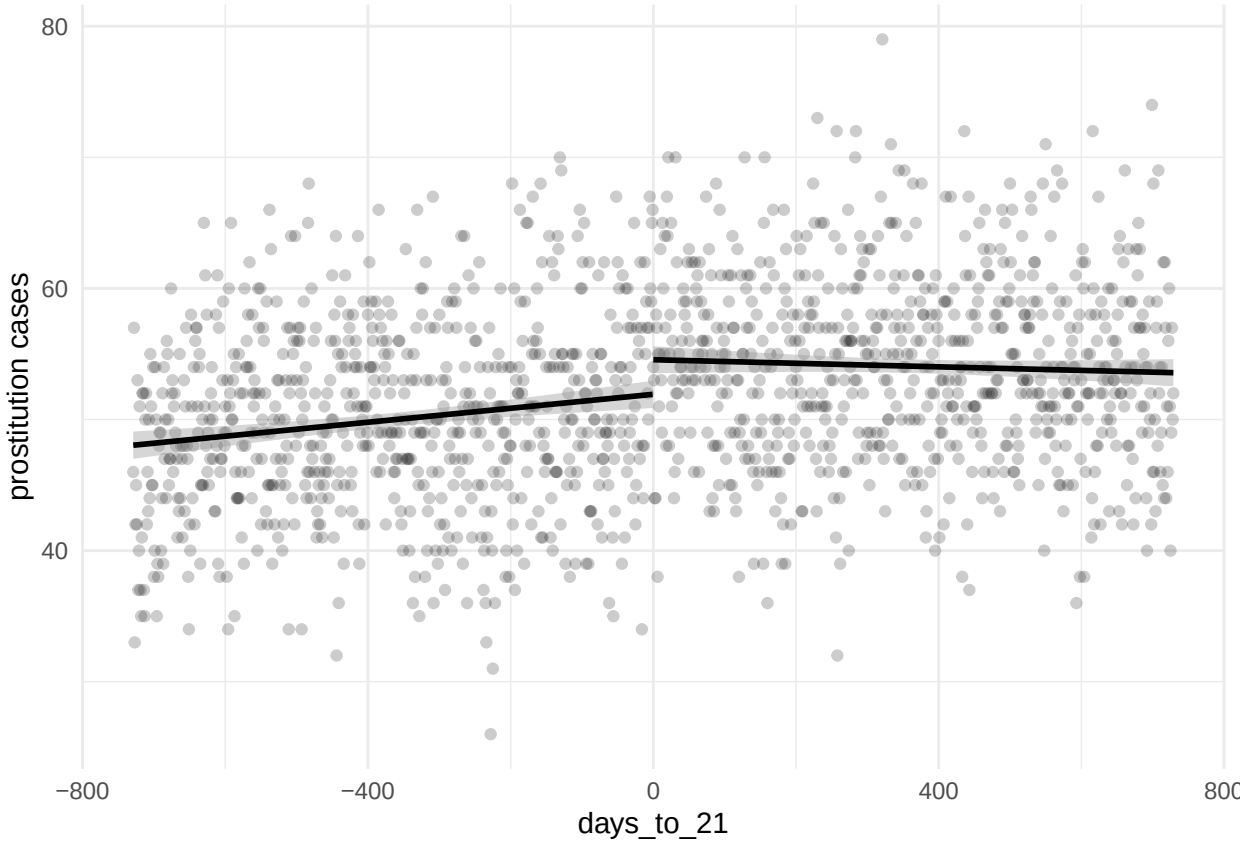


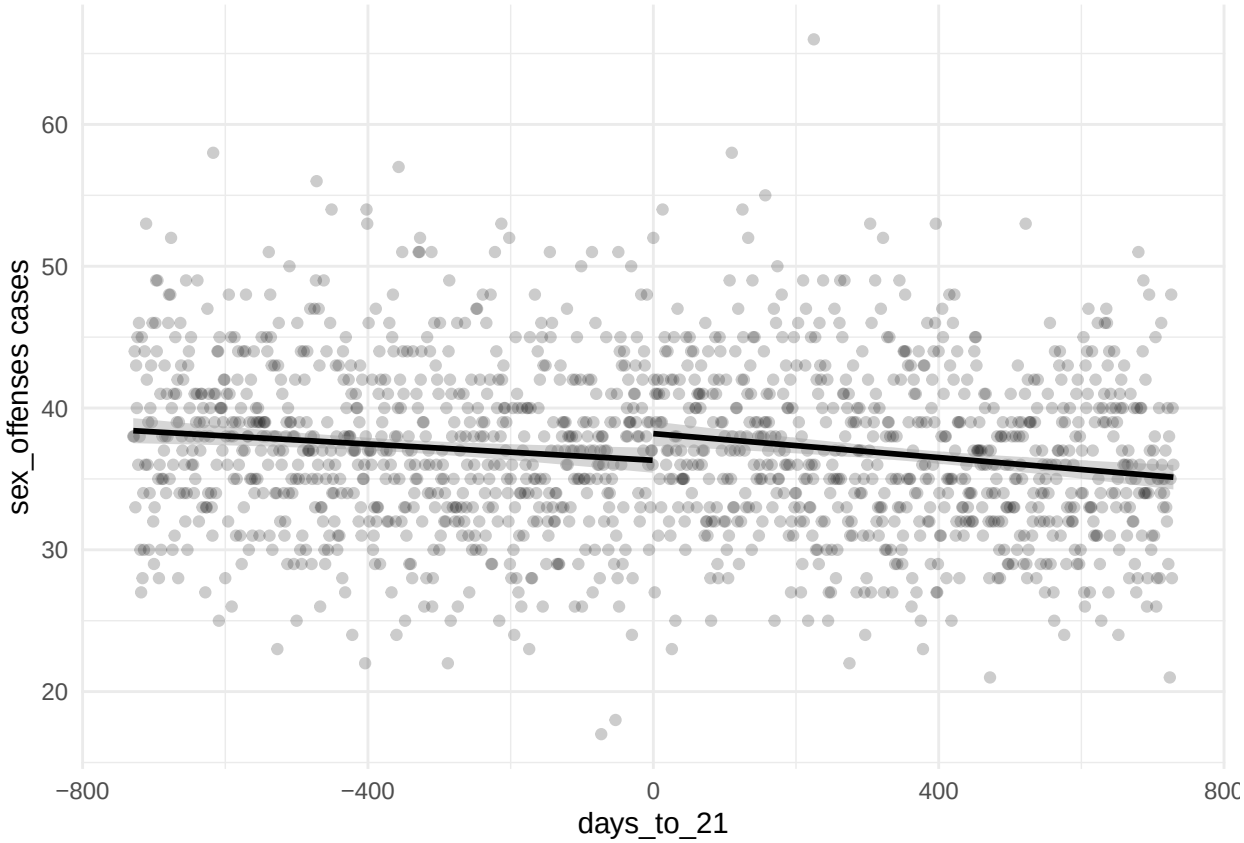

60
50
40
30
forgery_counterfit cases
−800
−400
0
400
800
days_to_21

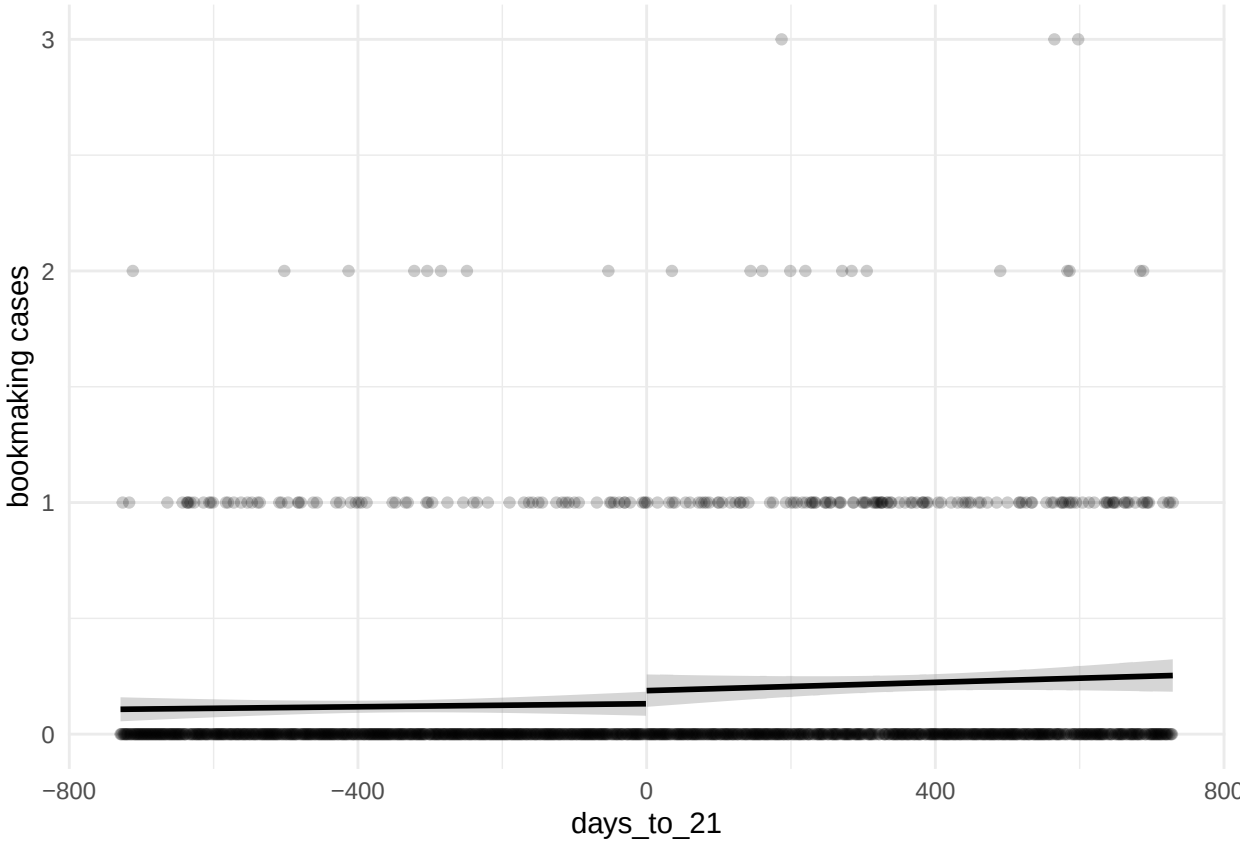

fraud cases
50
40
30
20
−800
−400
0
400
800
days_to_21

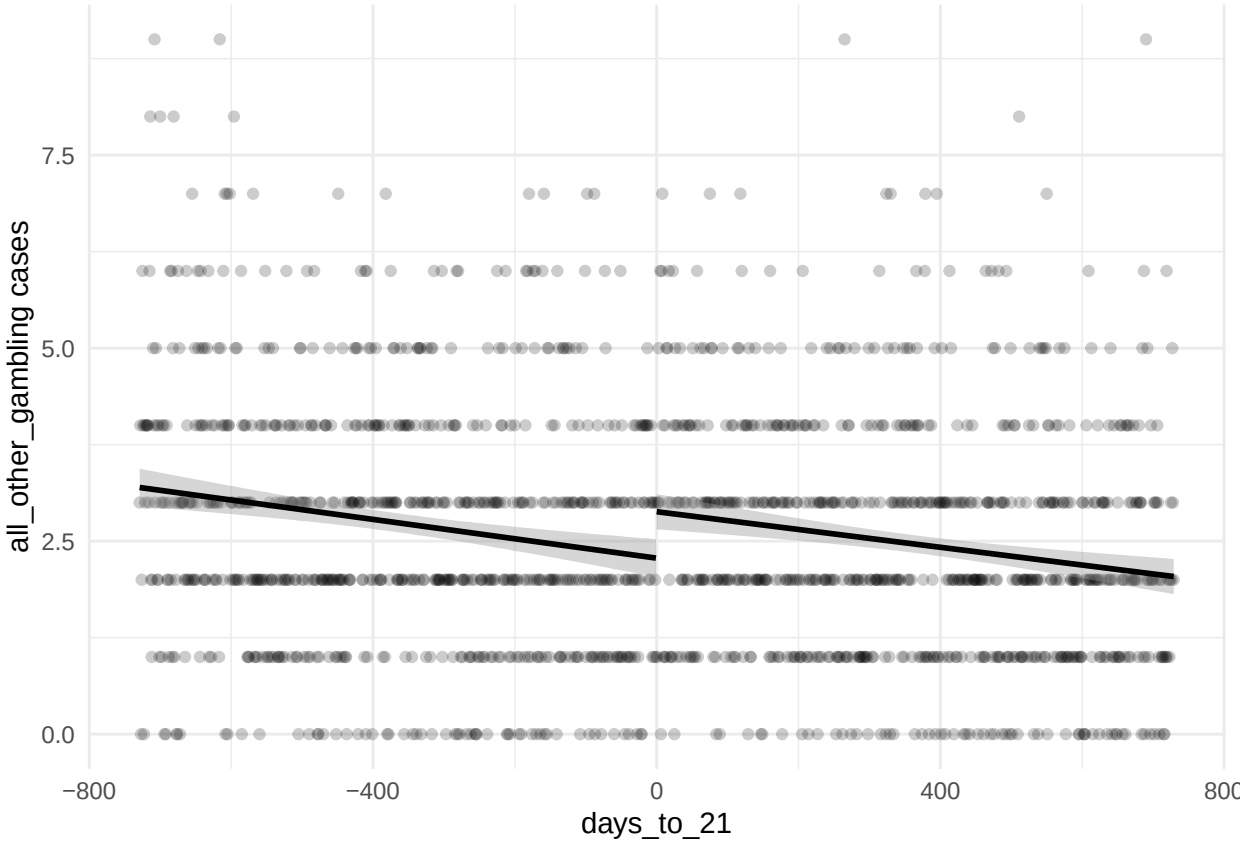

embezelment cases
15
10
5
0
−800
−400
0
400
800
days_to_21

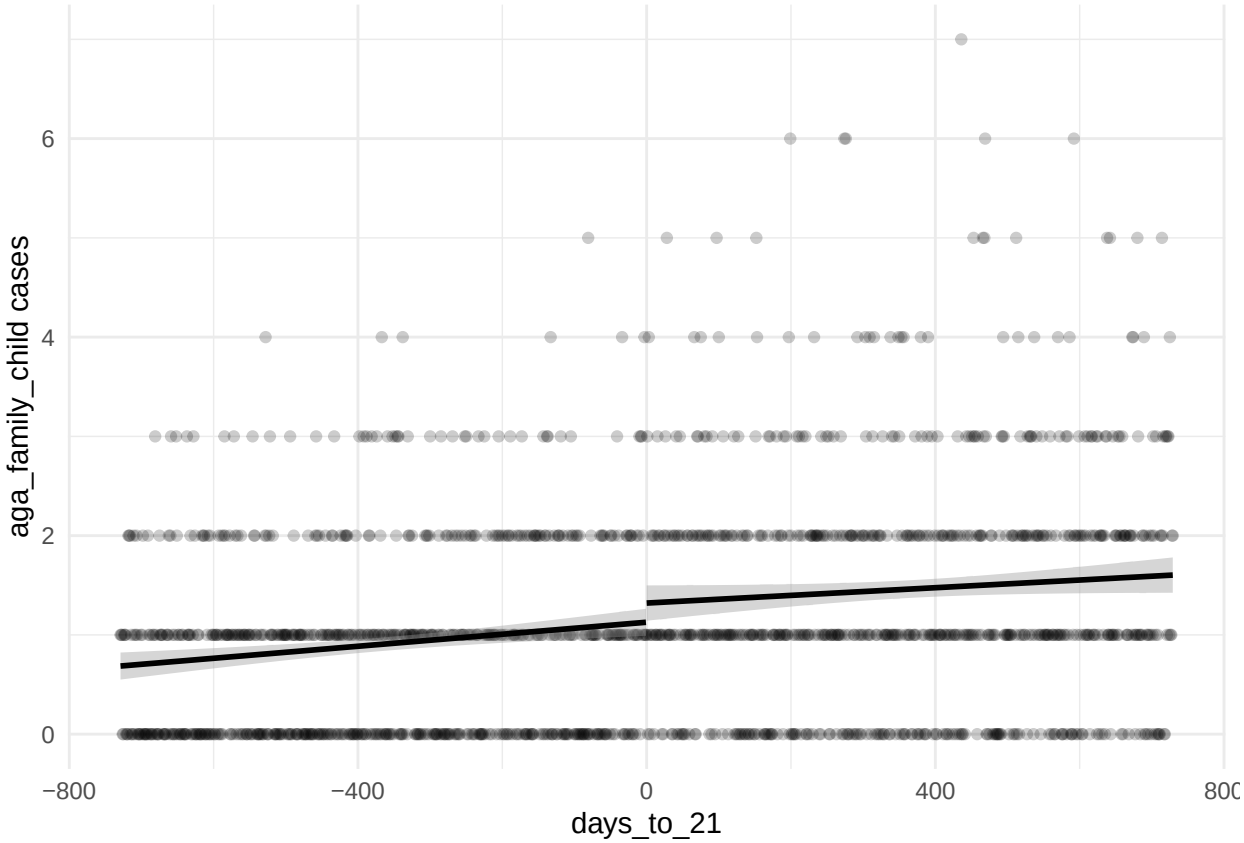

150
125
100
75
weapons cases
−800
−400
0
400
800
days_to_21

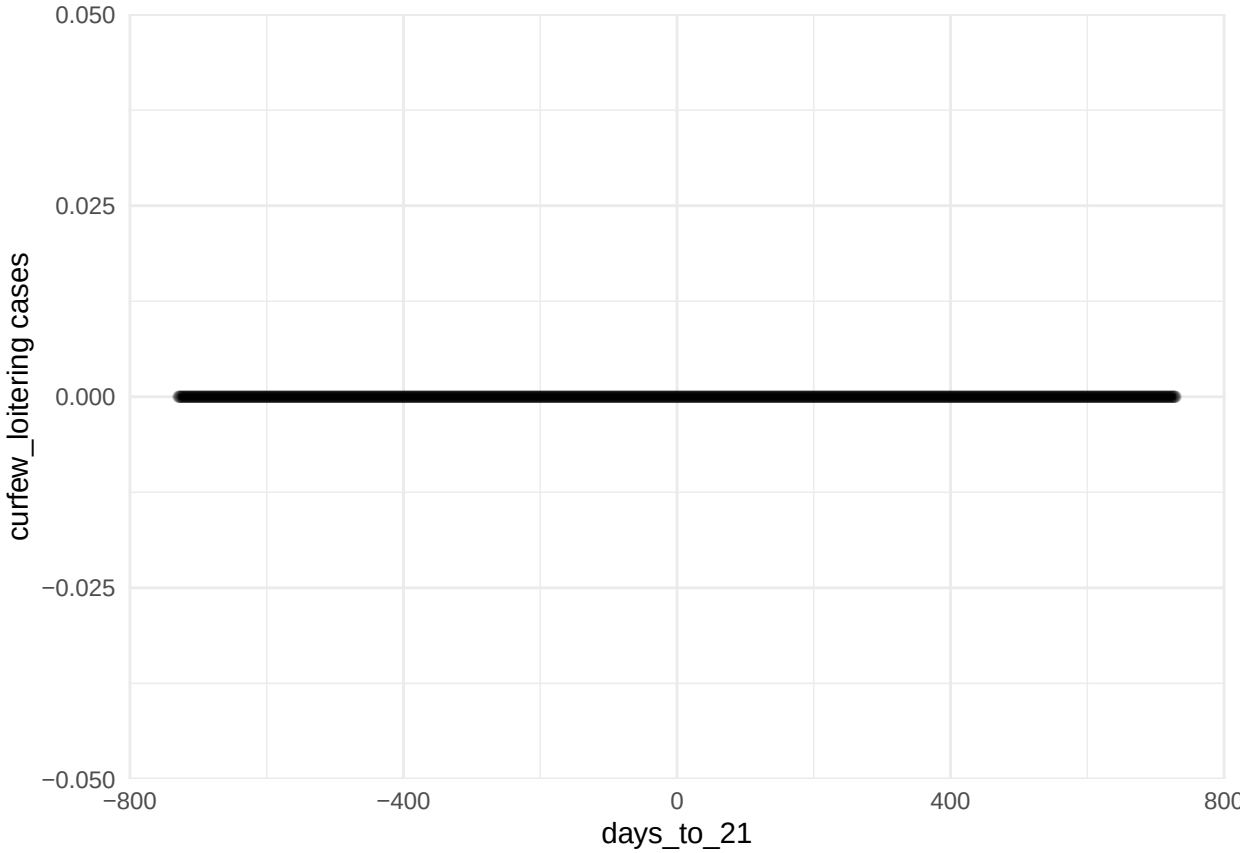

80
60
40
prostitution cases
−800
−400
0
400
800
days_to_21

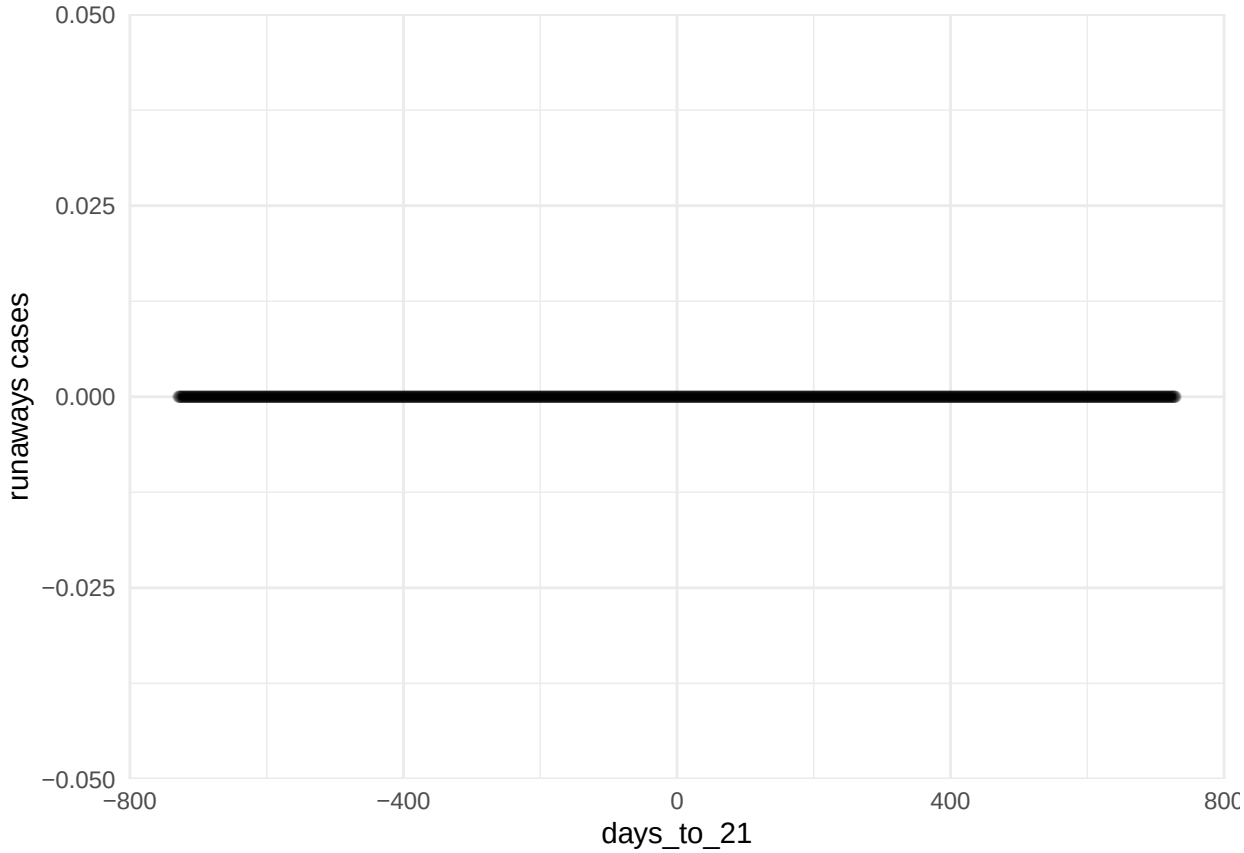

60
50
40
30
20
sex_offenses cases
−800
−400
0
400
800
days_to_21


3
2
1
0
bookmaking cases
−800
−400
0
400
800
days_to_21


all_other_gambling cases
7.5
5.0
2.5
0.0
−800
−400
0
400
800
days_to_21


aga_family_child cases
6
4
2
0
−800
−400
0
400
800
days_to_21


0.050
0.025
0.000
−0.025
−0.050
curfew_loitering cases
−800
−400
0
400
800
days_to_21


0.050
0.025
0.000
−0.025
−0.050
runaways cases
−800
−400
0
400
800
days_to_21

# Appendix A

```
knitr::opts_chunk$set(echo = FALSE, include = TRUE)
# import libraries
library(broom)
library(tidyverse)
library(gridExtra)
library(pastecs)


setwd("~/Experimental Design for Data Science/ProblemSet5")

drinking <- read.csv("P01 Age Profile of Arrest Rates 1979-2006.csv")
drinking <- janitor::clean_names(drinking)

head(drinking)
drinking_19_to_23 <- filter(drinking, days_to_21 > -730 & days_to_21 <
    730)

drinking_19_to_23 %>% ggplot(aes(x = days_to_21, y = vandalism)) +
    geom_point(alpha = 0.2) + geom_smooth(data = drinking_19_to_23 %>%
    filter(days_to_21 < 0), method = "lm", color = "black") +
    geom_smooth(data = drinking_19_to_23 %>% filter(days_to_21 >=
        0), method = "lm", color = "black") + theme_minimal() +
    labs(x = "days_to_21", y = "felony cases")
drinking_19_to_23 <- filter(drinking, days_to_21 > -730 & days_to_21 <
    730)

drinking_19_to_23 %>% ggplot(aes(x = days_to_21, y = violent_r)) +
    geom_point(alpha = 0.2) + geom_smooth(data = drinking_19_to_23 %>%
    filter(days_to_21 < 0), method = "lm", color = "black") +
    geom_smooth(data = drinking_19_to_23 %>% filter(days_to_21 >=
        0), method = "lm", color = "black") + theme_minimal() +
    labs(x = "days_to_21", y = "felony cases")
drinking <- drinking %>% mutate(under_21 = if_else(days_to_21 <
    0, 0, 1))

lm(felony ~ days_to_21 + under_21, data = drinking) %>% tidy()
violent_cases <- select(drinking_19_to_23, days_to_21, murder,
    manslaughter, rape, robbery, aggravated_assault, ot_assault)

violent_cases <- violent_cases %>% mutate(under_21 = if_else(days_to_21 <
    0, 0, 1))

head(violent_cases)

# Plot murder cases
violent_cases %>% ggplot(aes(x = days_to_21, y = murder)) + geom_point(alpha = 0.2) +
    geom_smooth(data = violent_cases %>% filter(days_to_21 <
        0), method = "lm", color = "black") + geom_smooth(data = violent_cases %>%
    filter(days_to_21 >= 0), method = "lm", color = "black") +
    theme_minimal() + labs(x = "days_to_21", y = "murder cases")
```

```r
# Plot manslaughter cases
violent_cases %>% ggplot(aes(x = days_to_21, y = manslaughter)) +
    geom_point(alpha = 0.2) + geom_smooth(data = violent_cases %>%
    filter(days_to_21 < 0), method = "lm", color = "black") +
    geom_smooth(data = violent_cases %>% filter(days_to_21 >=
        0), method = "lm", color = "black") + theme_minimal() +
    labs(x = "days_to_21", y = "manslaughter cases")

# Plot rape cases
violent_cases %>% ggplot(aes(x = days_to_21, y = rape)) + geom_point(alpha = 0.2) +
    geom_smooth(data = violent_cases %>% filter(days_to_21 <
        0), method = "lm", color = "black") + geom_smooth(data = violent_cases %>%
    filter(days_to_21 >= 0), method = "lm", color = "black") +
    theme_minimal() + labs(x = "days_to_21", y = "rape cases")

# Plot robbery cases
violent_cases %>% ggplot(aes(x = days_to_21, y = robbery)) +
    geom_point(alpha = 0.2) + geom_smooth(data = violent_cases %>%
    filter(days_to_21 < 0), method = "lm", color = "black") +
    geom_smooth(data = violent_cases %>% filter(days_to_21 >=
        0), method = "lm", color = "black") + theme_minimal() +
    labs(x = "days_to_21", y = "robbery cases")

# Plot aggravated_assault cases
violent_cases %>% ggplot(aes(x = days_to_21, y = aggravated_assault)) +
    geom_point(alpha = 0.2) + geom_smooth(data = violent_cases %>%
    filter(days_to_21 < 0), method = "lm", color = "black") +
    geom_smooth(data = violent_cases %>% filter(days_to_21 >=
        0), method = "lm", color = "black") + theme_minimal() +
    labs(x = "days_to_21", y = "aggravated_assault cases")

# Plot ot_assault cases
violent_cases %>% ggplot(aes(x = days_to_21, y = ot_assault)) +
    geom_point(alpha = 0.2) + geom_smooth(data = violent_cases %>%
    filter(days_to_21 < 0), method = "lm", color = "black") +
    geom_smooth(data = violent_cases %>% filter(days_to_21 >=
        0), method = "lm", color = "black") + theme_minimal() +
    labs(x = "days_to_21", y = "ot_assault cases")


lm(manslaughter ~ days_to_21 + under_21, data = violent_cases) %>%
    tidy()

lm(rape ~ days_to_21 + under_21, data = violent_cases) %>% tidy()

# Compute descriptive statistics - boxplots
p1 <- ggplot(violent_cases) + aes(y = murder) + geom_boxplot() +
    labs(x = "", y = "murder cases (%)")
p2 <- ggplot(violent_cases) + aes(y = rape) + geom_boxplot() +
    labs(x = "", y = "Low Income (%)")
p3 <- ggplot(violent_cases) + aes(y = manslaughter) + geom_boxplot() +
    labs(x = "", y = "Visible minority (%)")
grid1 <- grid.arrange(p1, p2, p3, ncol = 1, nrow = 3)
```

```r
grid1

# Compute descriptive statistics - table
tt1 <- ttheme_default()
stats_table <- stat.desc(violent_cases)
stats_table <- round(stats_table, 2)
stats_table <- mutate(stats_table, stats = row.names(stats_table))
# stats_table <- select(stats_table, stats,
# no_certificate_diploma_or_degree_2,
# x18_to_64_years_percent, total_visible_minority_population)
# colnames(stats_table) <- c('stats','No certificate','Low
# Income','Visible minority')
grid2 <- grid.arrange(tableGrob(stats_table, theme = tt1, rows = NULL),
    ncol = 1, nrow = 1)

# Join the boxplots and the table
grid.arrange(arrangeGrob(grid1, ncol = 2, nrow = 1), arrangeGrob(grid2,
    ncol = 1, nrow = 1), heights = c(30, 1), widths = c(1, 3),
    bottom = "Figure 1: Descriptive statistics of the data")
drugs_cases <- select(drinking_19_to_23, days_to_21, cocaine_opio_sale_manuf,
    mj_sale_manuf, dang_non_narc_sale_manuf, cocaine_opio_posses,
    mj_posses, dang_non_narc_posses)

drugs_cases <- drugs_cases %>% mutate(under_21 = if_else(days_to_21 <
    0, 0, 1))

head(drugs_cases)


# Plot cocaine_opio_sale_manuf cases
drugs_cases %>% ggplot(aes(x = days_to_21, y = cocaine_opio_sale_manuf)) +
    geom_point(alpha = 0.2) + geom_smooth(data = drugs_cases %>%
    filter(days_to_21 < 0), method = "lm", color = "black") +
    geom_smooth(data = drugs_cases %>% filter(days_to_21 >= 0),
        method = "lm", color = "black") + theme_minimal() + labs(x = "days_to_21",
    y = "cocaine_opio_sale_manuf cases")

# Plot mj_sale_manuf cases
drugs_cases %>% ggplot(aes(x = days_to_21, y = mj_sale_manuf)) +
    geom_point(alpha = 0.2) + geom_smooth(data = drugs_cases %>%
    filter(days_to_21 < 0), method = "lm", color = "black") +
    geom_smooth(data = drugs_cases %>% filter(days_to_21 >= 0),
        method = "lm", color = "black") + theme_minimal() + labs(x = "days_to_21",
    y = "mj_sale_manuf cases")

# Plot dang_non_narc_sale_manuf cases
drugs_cases %>% ggplot(aes(x = days_to_21, y = dang_non_narc_sale_manuf)) +
    geom_point(alpha = 0.2) + geom_smooth(data = drugs_cases %>%
    filter(days_to_21 < 0), method = "lm", color = "black") +
    geom_smooth(data = drugs_cases %>% filter(days_to_21 >= 0),
        method = "lm", color = "black") + theme_minimal() + labs(x = "days_to_21",
    y = "dang_non_narc_sale_manuf cases")
```

```r
# Plot cocaine_opio_posses cases
drugs_cases %>% ggplot(aes(x = days_to_21, y = cocaine_opio_posses)) +
    geom_point(alpha = 0.2) + geom_smooth(data = drugs_cases %>%
    filter(days_to_21 < 0), method = "lm", color = "black") +
    geom_smooth(data = drugs_cases %>% filter(days_to_21 >= 0),
        method = "lm", color = "black") + theme_minimal() + labs(x = "days_to_21",
    y = "cocaine_opio_posses cases")

# Plot mj_posses cases
drugs_cases %>% ggplot(aes(x = days_to_21, y = mj_posses)) +
    geom_point(alpha = 0.2) + geom_smooth(data = drugs_cases %>%
    filter(days_to_21 < 0), method = "lm", color = "black") +
    geom_smooth(data = drugs_cases %>% filter(days_to_21 >= 0),
        method = "lm", color = "black") + theme_minimal() + labs(x = "days_to_21",
    y = "mj_posses cases")


# Plot dang_non_narc_posses cases
drugs_cases %>% ggplot(aes(x = days_to_21, y = dang_non_narc_posses)) +
    geom_point(alpha = 0.2) + geom_smooth(data = drugs_cases %>%
    filter(days_to_21 < 0), method = "lm", color = "black") +
    geom_smooth(data = drugs_cases %>% filter(days_to_21 >= 0),
        method = "lm", color = "black") + theme_minimal() + labs(x = "days_to_21",
    y = "dang_non_narc_posses cases")



alcohol_cases <- select(drinking_19_to_23, days_to_21, dui, liquor_laws,
    drunkeness_pc, disorderly_cond, vagrancy)

alcohol_cases <- alcohol_cases %>% mutate(under_21 = if_else(days_to_21 <
    0, 0, 1))

head(alcohol_cases)

# Plot dui cases
alcohol_cases %>% ggplot(aes(x = days_to_21, y = dui)) + geom_point(alpha = 0.2) +
    geom_smooth(data = alcohol_cases %>% filter(days_to_21 <
        0), method = "lm", color = "black") + geom_smooth(data = alcohol_cases %>%
    filter(days_to_21 >= 0), method = "lm", color = "black") +
    theme_minimal() + labs(x = "days_to_21", y = "dui cases")

# Plot liquor_laws cases
alcohol_cases %>% ggplot(aes(x = days_to_21, y = liquor_laws)) +
    geom_point(alpha = 0.2) + geom_smooth(data = alcohol_cases %>%
    filter(days_to_21 < 0), method = "lm", color = "black") +
    geom_smooth(data = alcohol_cases %>% filter(days_to_21 >=
        0), method = "lm", color = "black") + theme_minimal() +
    labs(x = "days_to_21", y = "liquor_laws cases")

# Plot drunkeness_pc cases
alcohol_cases %>% ggplot(aes(x = days_to_21, y = drunkeness_pc)) +
    geom_point(alpha = 0.2) + geom_smooth(data = alcohol_cases %>%
```

```
    filter(days_to_21 < 0), method = "lm", color = "black") +
    geom_smooth(data = alcohol_cases %>% filter(days_to_21 >=
        0), method = "lm", color = "black") + theme_minimal() +
    labs(x = "days_to_21", y = "drunkeness_pc cases")

# Plot disorderly_cond cases
alcohol_cases %>% ggplot(aes(x = days_to_21, y = disorderly_cond)) +
    geom_point(alpha = 0.2) + geom_smooth(data = alcohol_cases %>%
    filter(days_to_21 < 0), method = "lm", color = "black") +
    geom_smooth(data = alcohol_cases %>% filter(days_to_21 >=
        0), method = "lm", color = "black") + theme_minimal() +
    labs(x = "days_to_21", y = "disorderly_cond cases")

# Plot vagrancy cases
alcohol_cases %>% ggplot(aes(x = days_to_21, y = vagrancy)) +
    geom_point(alpha = 0.2) + geom_smooth(data = alcohol_cases %>%
    filter(days_to_21 < 0), method = "lm", color = "black") +
    geom_smooth(data = alcohol_cases %>% filter(days_to_21 >=
        0), method = "lm", color = "black") + theme_minimal() +
    labs(x = "days_to_21", y = "vagrancy cases")


lm(vagrancy ~ days_to_21 + under_21, data = alcohol_cases) %>%
    tidy()
property_cases <- select(drinking_19_to_23, days_to_21, burglary,
    larceny, mv_theft, stolen_prop_buy_rec_poss, vandalism)

property_cases <- property_cases %>% mutate(under_21 = if_else(days_to_21 <
    0, 0, 1))

head(property_cases)

# Plot burglary cases
property_cases %>% ggplot(aes(x = days_to_21, y = burglary)) +
    geom_point(alpha = 0.2) + geom_smooth(data = property_cases %>%
    filter(days_to_21 < 0), method = "lm", color = "black") +
    geom_smooth(data = property_cases %>% filter(days_to_21 >=
        0), method = "lm", color = "black") + theme_minimal() +
    labs(x = "days_to_21", y = "burglary cases")

# Plot larceny cases
property_cases %>% ggplot(aes(x = days_to_21, y = larceny)) +
    geom_point(alpha = 0.2) + geom_smooth(data = property_cases %>%
    filter(days_to_21 < 0), method = "lm", color = "black") +
    geom_smooth(data = property_cases %>% filter(days_to_21 >=
        0), method = "lm", color = "black") + theme_minimal() +
    labs(x = "days_to_21", y = "larceny cases")

# Plot mv_theft cases
property_cases %>% ggplot(aes(x = days_to_21, y = mv_theft)) +
    geom_point(alpha = 0.2) + geom_smooth(data = property_cases %>%
    filter(days_to_21 < 0), method = "lm", color = "black") +
    geom_smooth(data = property_cases %>% filter(days_to_21 >=
```

```
        0), method = "lm", color = "black") + theme_minimal() +
    labs(x = "days_to_21", y = " mv_theft cases")

# Plot stolen_prop_buy_rec_poss cases
property_cases %>% ggplot(aes(x = days_to_21, y = stolen_prop_buy_rec_poss)) +
    geom_point(alpha = 0.2) + geom_smooth(data = property_cases %>%
    filter(days_to_21 < 0), method = "lm", color = "black") +
    geom_smooth(data = property_cases %>% filter(days_to_21 >=
        0), method = "lm", color = "black") + theme_minimal() +
    labs(x = "days_to_21", y = "stolen_prop_buy_rec_poss cases")

# Plot vandalism cases
property_cases %>% ggplot(aes(x = days_to_21, y = vandalism)) +
    geom_point(alpha = 0.2) + geom_smooth(data = property_cases %>%
    filter(days_to_21 < 0), method = "lm", color = "black") +
    geom_smooth(data = property_cases %>% filter(days_to_21 >=
        0), method = "lm", color = "black") + theme_minimal() +
    labs(x = "days_to_21", y = "vandalism cases")

property_cases %>% ggplot(aes(x = days_to_21, y = burglary)) +
    geom_point(alpha = 0.2)

lm(burglary ~ days_to_21 + under_21, data = property_cases) %>%
    tidy()
fbi_other_cases <- select(drinking_19_to_23, days_to_21, arson,
    forgery_counterfit, fraud, embezelment, weapons, prostitution,
    sex_offenses, bookmaking, all_other_gambling, aga_family_child,
    curfew_loitering, runaways)

fbi_other_cases <- fbi_other_cases %>% mutate(under_21 = if_else(days_to_21 <
    0, 0, 1))

head(fbi_other_cases)

# Plot arson cases
fbi_other_cases %>% ggplot(aes(x = days_to_21, y = arson)) +
    geom_point(alpha = 0.2) + geom_smooth(data = fbi_other_cases %>%
    filter(days_to_21 < 0), method = "lm", color = "black") +
    geom_smooth(data = fbi_other_cases %>% filter(days_to_21 >=
        0), method = "lm", color = "black") + theme_minimal() +
    labs(x = "days_to_21", y = "arson cases")

# Plot forgery_counterfit cases
fbi_other_cases %>% ggplot(aes(x = days_to_21, y = forgery_counterfit)) +
    geom_point(alpha = 0.2) + geom_smooth(data = fbi_other_cases %>%
    filter(days_to_21 < 0), method = "lm", color = "black") +
    geom_smooth(data = fbi_other_cases %>% filter(days_to_21 >=
        0), method = "lm", color = "black") + theme_minimal() +
    labs(x = "days_to_21", y = "forgery_counterfit cases")

# Plot fraud cases
fbi_other_cases %>% ggplot(aes(x = days_to_21, y = fraud)) +
    geom_point(alpha = 0.2) + geom_smooth(data = fbi_other_cases %>%
```

```
    filter(days_to_21 < 0), method = "lm", color = "black") +
    geom_smooth(data = fbi_other_cases %>% filter(days_to_21 >=
        0), method = "lm", color = "black") + theme_minimal() +
    labs(x = "days_to_21", y = "fraud cases")

# Plot embezelment cases
fbi_other_cases %>% ggplot(aes(x = days_to_21, y = embezelment)) +
    geom_point(alpha = 0.2) + geom_smooth(data = fbi_other_cases %>%
    filter(days_to_21 < 0), method = "lm", color = "black") +
    geom_smooth(data = fbi_other_cases %>% filter(days_to_21 >=
        0), method = "lm", color = "black") + theme_minimal() +
    labs(x = "days_to_21", y = "embezelment cases")

# Plot weapons cases
fbi_other_cases %>% ggplot(aes(x = days_to_21, y = weapons)) +
    geom_point(alpha = 0.2) + geom_smooth(data = fbi_other_cases %>%
    filter(days_to_21 < 0), method = "lm", color = "black") +
    geom_smooth(data = fbi_other_cases %>% filter(days_to_21 >=
        0), method = "lm", color = "black") + theme_minimal() +
    labs(x = "days_to_21", y = "weapons cases")

# Plot prostitution cases
fbi_other_cases %>% ggplot(aes(x = days_to_21, y = prostitution)) +
    geom_point(alpha = 0.2) + geom_smooth(data = fbi_other_cases %>%
    filter(days_to_21 < 0), method = "lm", color = "black") +
    geom_smooth(data = fbi_other_cases %>% filter(days_to_21 >=
        0), method = "lm", color = "black") + theme_minimal() +
    labs(x = "days_to_21", y = "prostitution cases")


# Plot sex_offenses cases
fbi_other_cases %>% ggplot(aes(x = days_to_21, y = sex_offenses)) +
    geom_point(alpha = 0.2) + geom_smooth(data = fbi_other_cases %>%
    filter(days_to_21 < 0), method = "lm", color = "black") +
    geom_smooth(data = fbi_other_cases %>% filter(days_to_21 >=
        0), method = "lm", color = "black") + theme_minimal() +
    labs(x = "days_to_21", y = "sex_offenses cases")

# Plot bookmaking cases
fbi_other_cases %>% ggplot(aes(x = days_to_21, y = bookmaking)) +
    geom_point(alpha = 0.2) + geom_smooth(data = fbi_other_cases %>%
    filter(days_to_21 < 0), method = "lm", color = "black") +
    geom_smooth(data = fbi_other_cases %>% filter(days_to_21 >=
        0), method = "lm", color = "black") + theme_minimal() +
    labs(x = "days_to_21", y = "bookmaking cases")

# Plot all_other_gambling cases
fbi_other_cases %>% ggplot(aes(x = days_to_21, y = all_other_gambling)) +
    geom_point(alpha = 0.2) + geom_smooth(data = fbi_other_cases %>%
    filter(days_to_21 < 0), method = "lm", color = "black") +
    geom_smooth(data = fbi_other_cases %>% filter(days_to_21 >=
        0), method = "lm", color = "black") + theme_minimal() +
    labs(x = "days_to_21", y = "all_other_gambling cases")
```

```r
# Plot aga_family_child cases
fbi_other_cases %>% ggplot(aes(x = days_to_21, y = aga_family_child)) +
    geom_point(alpha = 0.2) + geom_smooth(data = fbi_other_cases %>%
    filter(days_to_21 < 0), method = "lm", color = "black") +
    geom_smooth(data = fbi_other_cases %>% filter(days_to_21 >=
        0), method = "lm", color = "black") + theme_minimal() +
    labs(x = "days_to_21", y = "aga_family_child cases")

# Plot curfew_loitering cases
fbi_other_cases %>% ggplot(aes(x = days_to_21, y = curfew_loitering)) +
    geom_point(alpha = 0.2) + geom_smooth(data = fbi_other_cases %>%
    filter(days_to_21 < 0), method = "lm", color = "black") +
    geom_smooth(data = fbi_other_cases %>% filter(days_to_21 >=
        0), method = "lm", color = "black") + theme_minimal() +
    labs(x = "days_to_21", y = "curfew_loitering cases")


# Plot runaways cases
fbi_other_cases %>% ggplot(aes(x = days_to_21, y = runaways)) +
    geom_point(alpha = 0.2) + geom_smooth(data = fbi_other_cases %>%
    filter(days_to_21 < 0), method = "lm", color = "black") +
    geom_smooth(data = fbi_other_cases %>% filter(days_to_21 >=
        0), method = "lm", color = "black") + theme_minimal() +
    labs(x = "days_to_21", y = "runaways cases")
```

# References