Exploratory data for violation

Sheng Pan

City <fctr>	Year <fctr>	Violation count <int>	total value <dbl>
Allston	2020	1837	86250
Ashmont	2020	250	14910
Back Bay East	2020	1058	48450
Back Bay West	2020	1011	36445
Beacon Hill	2020	1523	45650
Bellevue Hill	2020	60	4085
Bowdoin North/MtBowdoin	2020	358	16495
Brighton - Oak Square	2020	1195	57355
Brighton - St Elizabeth's	2020	1382	66400
Brook Farm/VFW Parkway	2020	139	5840

Violation Count versus Total Value by Neighborhood

Count versus Total Value for Violations by Neighborhood
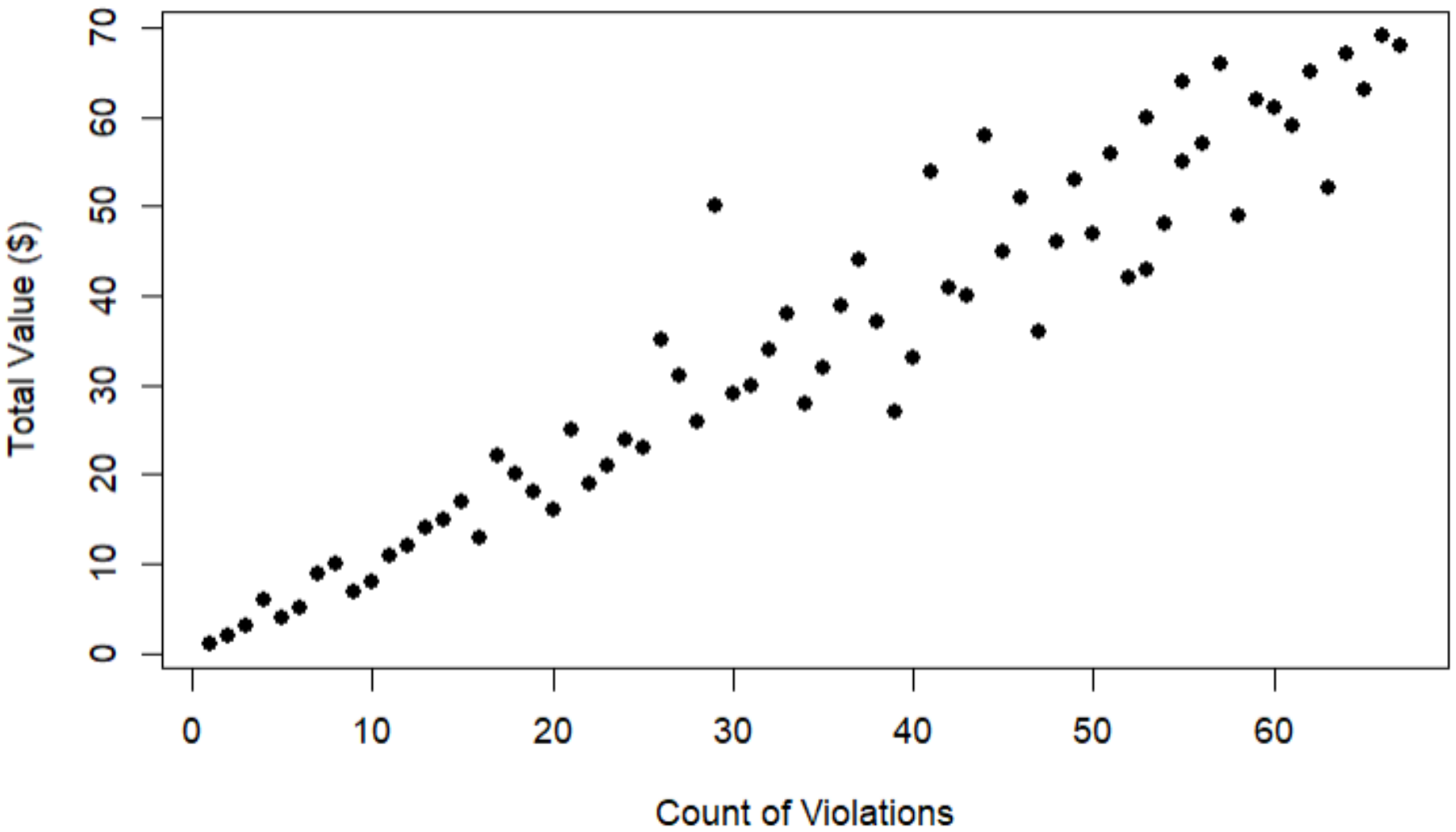
```
        Pearson's product-moment correlation

data:  NSA_violations$total_value and NSA_violations$viol_count
t = 27.417, df = 67, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.9331568 0.9739964
sample estimates:
      cor
0.9582076


Call:
lm(formula = total_value ~ viol_count, data = NSA_violations)

Residuals:
     Min      1Q   Median      3Q      Max
-12.3594  -2.5798  -0.7312   2.9573  20.5029

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.89650    1.42540   0.629    0.532
viol_count   0.98623    0.03597  27.417   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.782 on 67 degrees of freedom
Multiple R-squared:  0.9182,    Adjusted R-squared:  0.9169
F-statistic: 751.7 on 1 and 67 DF,  p-value: < 2.2e-16
```
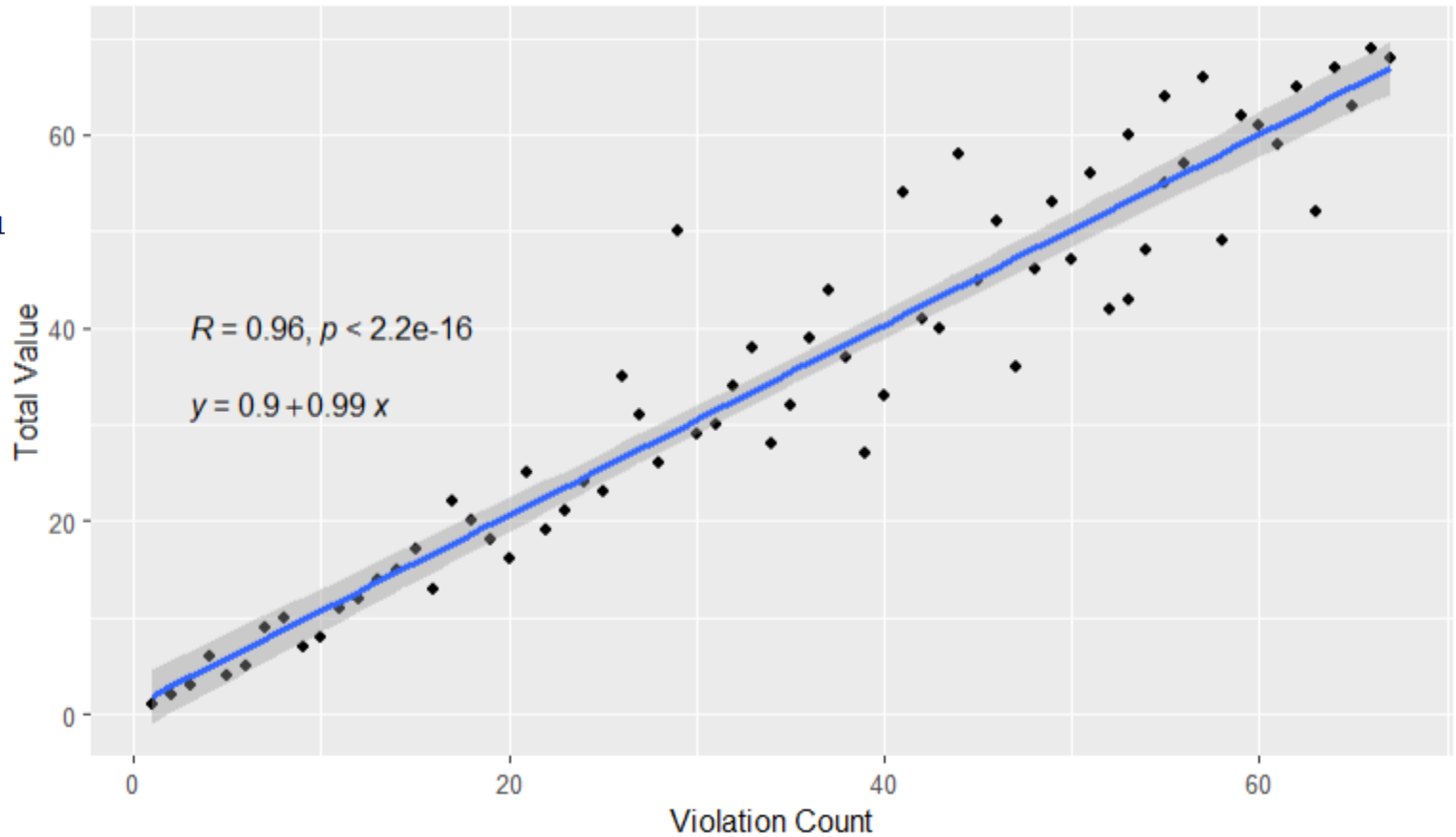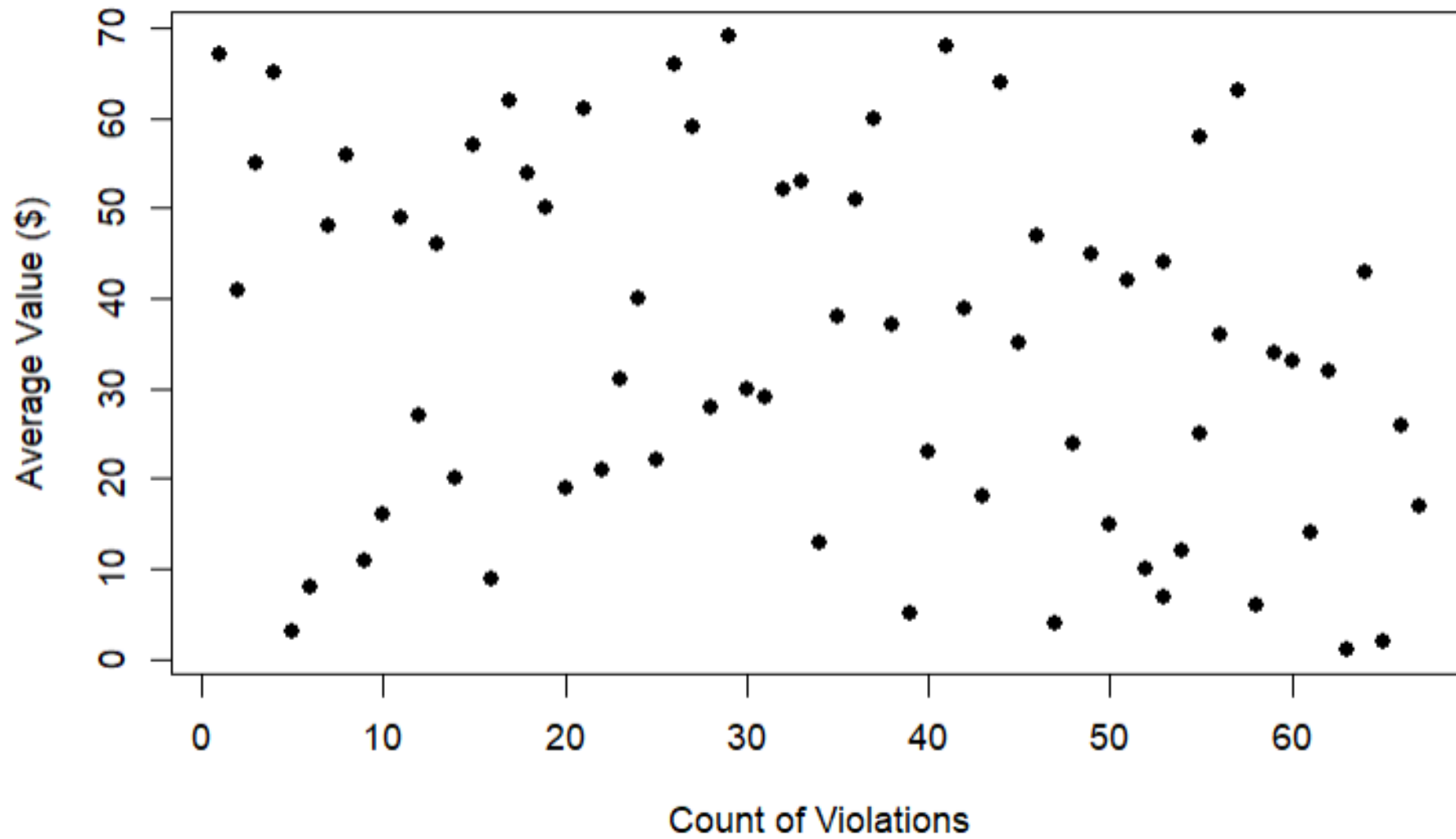

Corrlation plot between total value and violation count

$R = 0.96, p < 2.2e\text{-}16$

$y = 0.9 + 0.99\,x$

Count versus Average Value for Violations by Neighborhood

```
        Pearson's product-moment correlation

data:  NSA_violations2$avg_value and NSA_violations2$viol_count
t = -2.2019, df = 67, p-value = 0.03112
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.46769883 -0.02460604
sample estimates:
       cor
-0.2597739


Call:
lm(formula = avg_value ~ viol_count, data = NSA_violations2)

Residuals:
    Min      1Q  Median      3Q     Max
-39.909 -15.877   2.914  15.053  34.717

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  44.2456     4.8116    9.196 1.71e-13 ***
viol_count   -0.2674     0.1214   -2.202   0.0311 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '

Residual standard error: 19.52 on 67 degrees of freedom
Multiple R-squared:  0.06748,   Adjusted R-squared:  0.05356
F-statistic: 4.849 on 1 and 67 DF,  p-value: 0.03112
```
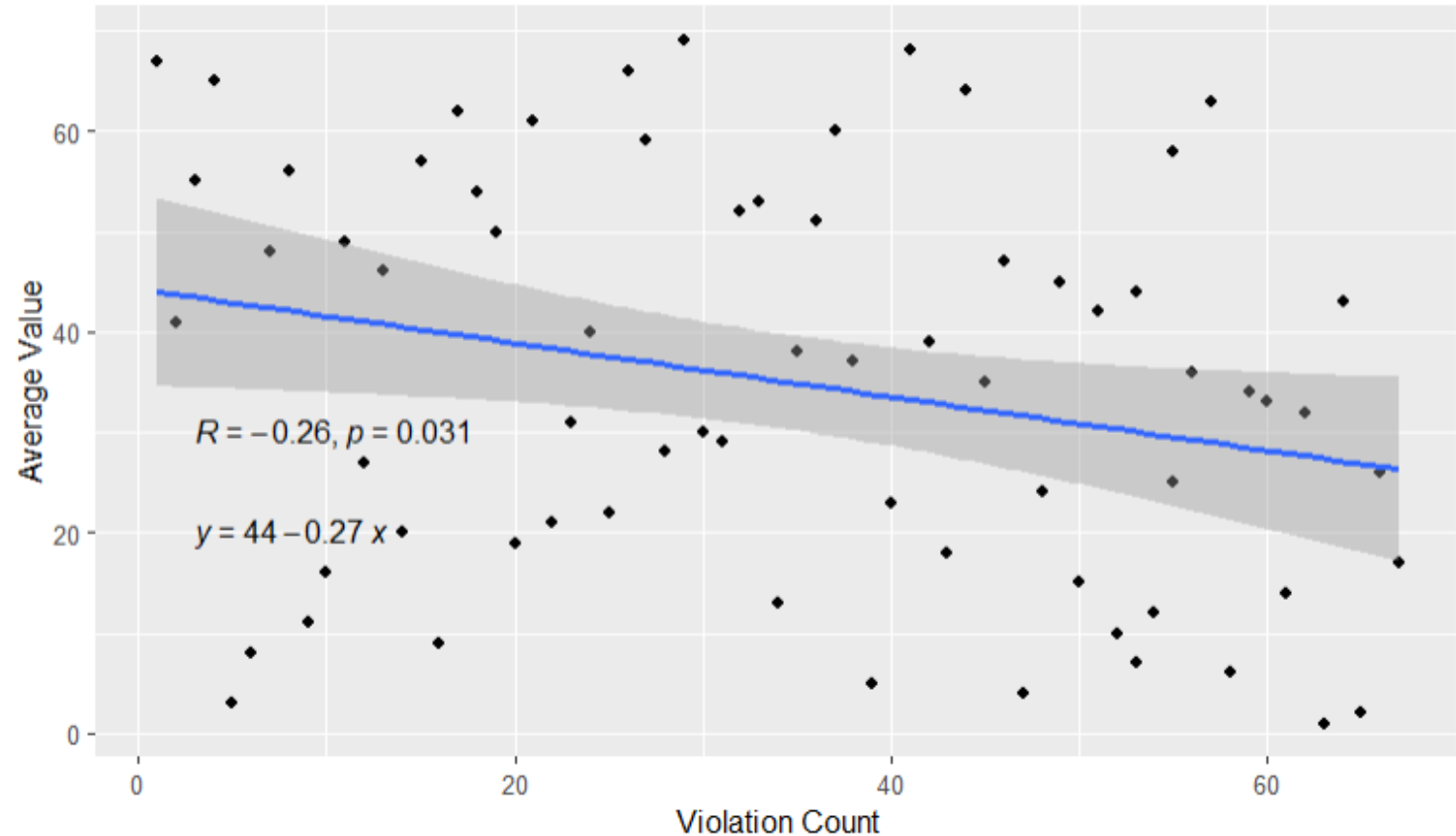

Corrlation plot between average value and violation count

$R = -0.26, p = 0.031$

$y = 44 - 0.27\,x$

```
        Pearson's product-moment correlation

data:  countn$violation_count and countn$month
t = 0.8862, df = 22, p-value = 0.3851
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.2353702  0.5480090
sample estimates:
      cor
0.1856539
```

```
Call:
lm(formula = violation_count ~ month + status, data = countn)

Residuals:
    Min      1Q  Median      3Q     Max
-38.538 -10.168  -5.001   7.627  65.710

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   28.545     13.292   2.148   0.0436 *
month          1.416      1.635   0.866   0.3963
statusOpen     1.167     11.289   0.103   0.9187
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 27.65 on 21 degrees of freedom
Multiple R-squared:  0.03496,   Adjusted R-squared:  -0.05695
F-statistic: 0.3804 on 2 and 21 DF,  p-value: 0.6882
```
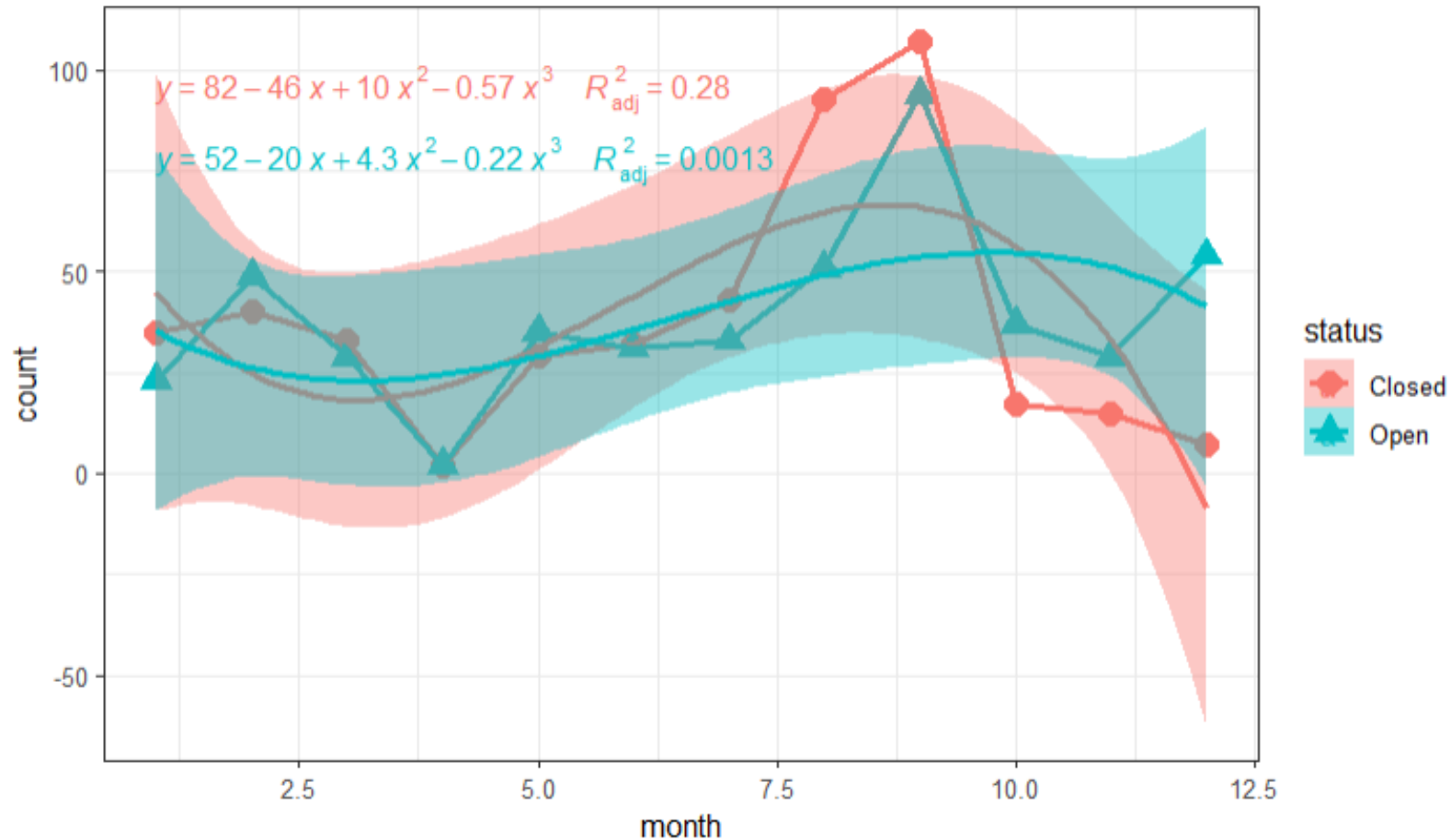
Line chart between month and violation count

$y = 82 - 46\,x + 10\,x^2 - 0.57\,x^3 \quad R^2_{adj} = 0.28$

$y = 52 - 20\,x + 4.3\,x^2 - 0.22\,x^3 \quad R^2_{adj} = 0.0013$

status
Closed
Open

Violation Count versus Total Value by description in Allston,2020

description <chr>	month <fctr>	viol_count <int>	total_value <dbl>
overfilling of barrel/dumpster	9	55	5500
improper storage trash: res	9	132	3300
improper storage trash: res	8	103	2575
overfilling of barrel/dumpster	2	25	2500
overfilling of barrel/dumpster	7	25	2500
overfilling of barrel/dumpster	8	25	2500
overfilling of barrel/dumpster	6	21	2100
overfilling of barrel/dumpster	5	18	1800
overfilling of barrel/dumpster	3	16	1600
failure clear sidewalk - snow	12	10	1250

Count versus Total Value for Violations by description in Allston,2020
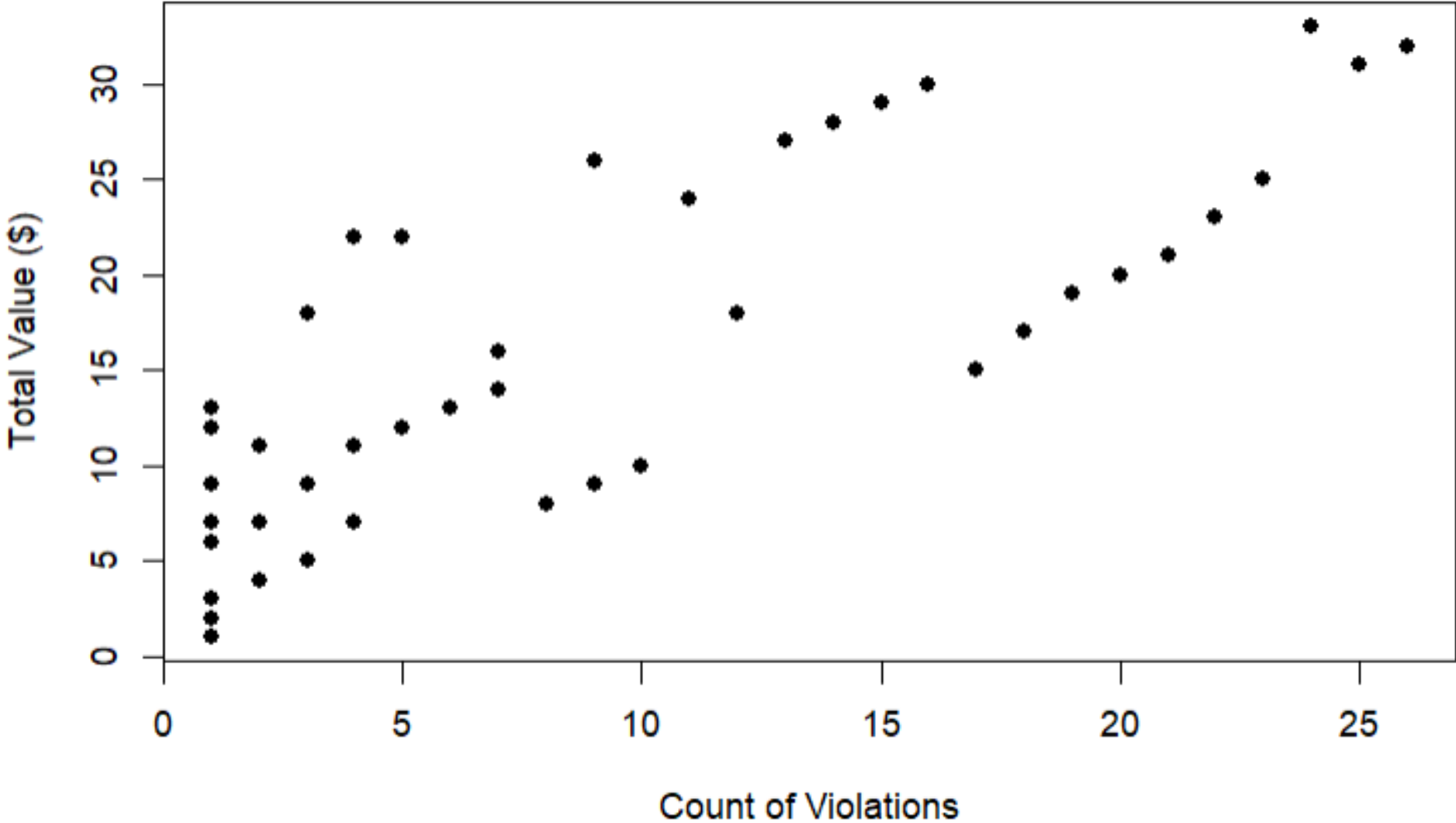
```
      Pearson's product-moment correlation

data:  viool_description$total_value and viool_description$viol_count
t = 9.3748, df = 54, p-value = 6.388e-13
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.6608677 0.8699311
sample estimates:
      cor
0.7870281


Call:
lm(formula = total_value ~ viol_count, data = viool_description)

Residuals:
    Min      1Q  Median      3Q     Max
-8.0003 -5.5480 -0.4704  4.7231 10.5296

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  7.92274    1.13570   6.976 4.51e-09 ***
viol_count   0.88691    0.09461   9.375 6.39e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.537 on 54 degrees of freedom
Multiple R-squared:  0.6194,    Adjusted R-squared:  0.6124
F-statistic: 87.89 on 1 and 54 DF,  p-value: 6.388e-13
```
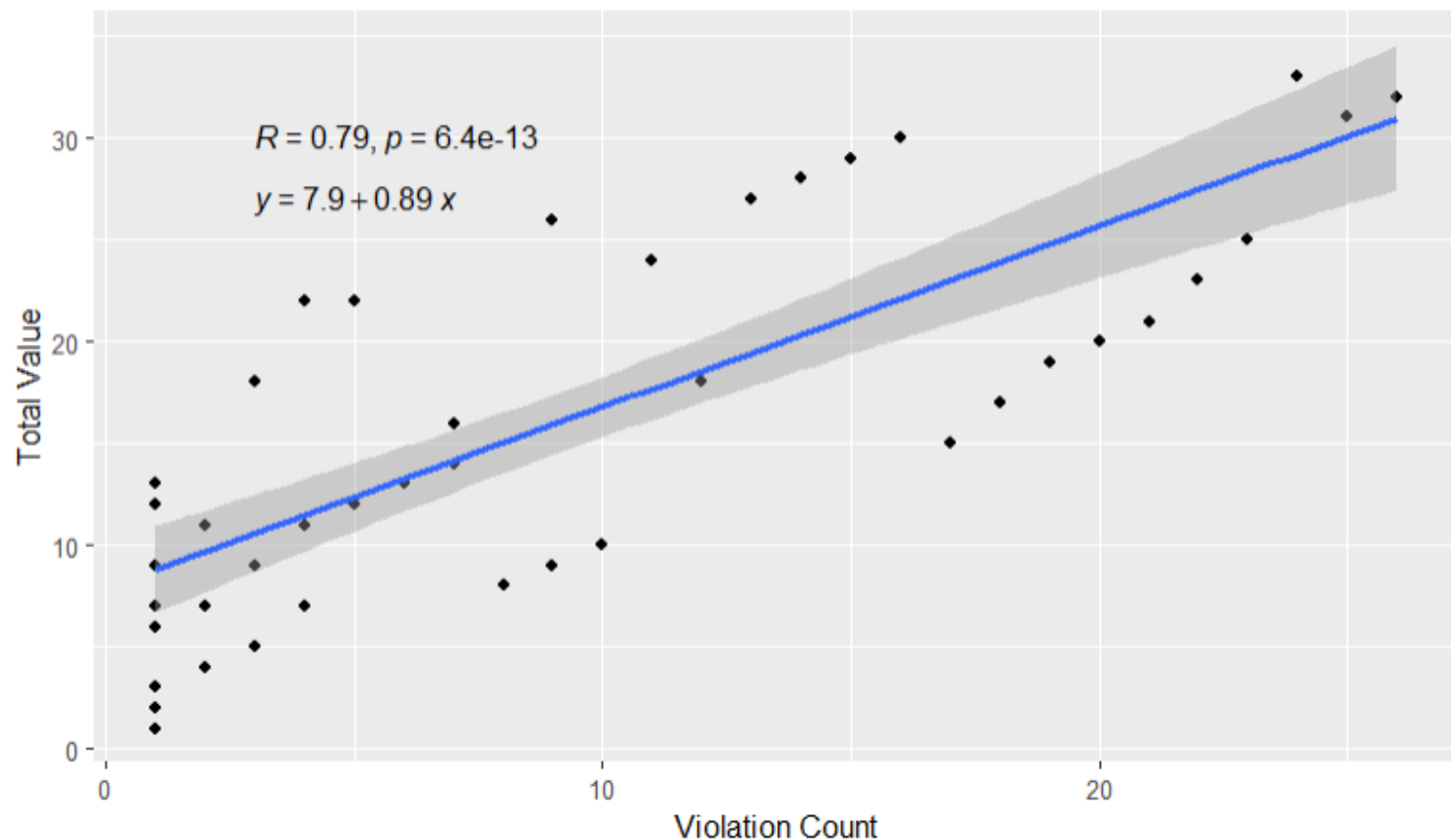
Corrlation plot between total value and violation count in Allston 2020

$R = 0.79, p = 6.4e\text{-}13$

$y = 7.9 + 0.89\,x$

Total Value

Violation Count

Conclusion