

Crime_Dataset_EDA

the_principal_components

2022-07-15

```
correlation_table(data=communities, target="ViolentCrimesPerPop")
```

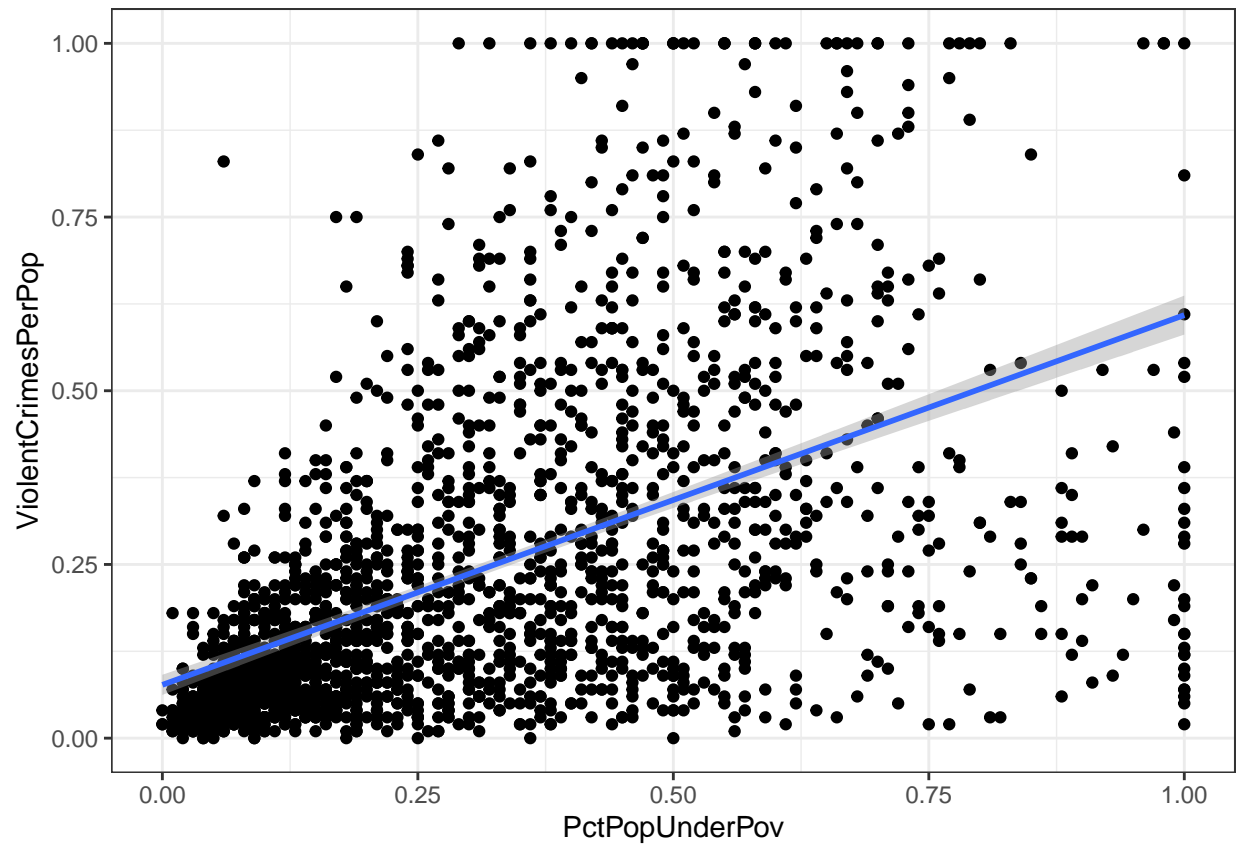
##	Variable	ViolentCrimesPerPop
## 1	ViolentCrimesPerPop	1.00
## 2	PctIlleg	0.83
## 3	pctWPubAsst	0.79
## 4	PctPopUnderPov	0.76
## 5	PctUnemployed	0.76
## 6	racepctblack	0.74
## 7	PctVacantBoarded	0.74
## 8	PctHousNoPhone	0.71
## 9	FemalePctDiv	0.70
## 10	PctLargHouseFam	0.70
## 11	TotalPctDiv	0.69
## 12	PctPolicBlack	0.69
## 13	PctPolicMinor	0.69
## 14	MalePctDivorce	0.66
## 15	NumIlleg	0.64
## 16	PctLargHouseOccup	0.60
## 17	PctWOFullPlumb	0.60
## 18	MalePctNevMarr	0.57
## 19	NumInShelters	0.57
## 20	NumUnderPov	0.56
## 21	PctNotHSGrad	0.55
## 22	PctPersDenseHous	0.55
## 23	PopDens	0.53
## 24	PersPerFam	0.52
## 25	HousVacant	0.50
## 26	NumStreet	0.49
## 27	MedRentPctHousInc	0.44
## 28	population	0.43
## 29	numbUrban	0.43
## 30	PolicPerPop	0.43
## 31	PersPerRentOccHous	0.42
## 32	LemasSwFTPerPop	0.42
## 33	PolicCars	0.42
## 34	PctHousLess3BR	0.41
## 35	LemasSwornFT	0.41
## 36	PctImmigRec10	0.40
## 37	PolicOperBudg	0.39
## 38	PctLess9thGrade	0.38
## 39	LemasSwFTFieldPerPop	0.38
## 40	LemasTotalReq	0.38

## 41	PctUsePubTrans	0.38
## 42	PctImmigRec8	0.37
## 43	LemasTotReqPerPop	0.37
## 44	PctOccupManu	0.36
## 45	NumImmig	0.36
## 46	OfficAssgnDrugUnits	0.36
## 47	agePct12t21	0.35
## 48	PctImmigRec5	0.32
## 49	agePct12t29	0.31
## 50	racePctHisp	0.28
## 51	PctImmigRecent	0.27
## 52	PctNotSpeakEnglWell	0.26
## 53	PctPolicHisp	0.26
## 54	NumKindsDrugsSeiz	0.25
## 55	PolicBudgPerPop	0.25
## 56	agePct16t24	0.24
## 57	PctVacMore6Mos	0.23
## 58	PctPolicAsian	0.20
## 59	PolicAveOTWorked	0.20
## 60	householdsize	0.19
## 61	pctUrban	0.19
## 62	PersPerOwnOccHous	0.18
## 63	PctSameCity85	0.15
## 64	PctRecImmig10	0.14
## 65	PersPerOccupHous	0.14
## 66	PctEmplProfServ	0.13
## 67	PctRecImmig8	0.13
## 68	MedOwnCostPctIncNoMtg	0.13
## 69	PolicReqPerOffic	0.12
## 70	PctRecImmig5	0.11
## 71	PctRecentImmig	0.09
## 72	LemasGangUnitDeploy	0.09
## 73	PctSameState85	0.07
## 74	PctBornSameState	0.06
## 75	county	0.02
## 76	LandArea	0.02
## 77	indianPerCap	-0.01
## 78	PctForeignBorn	-0.01
## 79	LemasPctOfficDrugUn	-0.01
## 80	state	-0.03
## 81	fold	-0.03
## 82	PctEmplManu	-0.06
## 83	MedOwnCostPctInc	-0.06
## 84	PctWorkMomYoungKids	-0.08
## 85	pctWSocSec	-0.13
## 86	racePctAsian	-0.14
## 87	PctSpeakEnglOnly	-0.15
## 88	LemasPctPolicOnPatr	-0.15
## 89	community	-0.17
## 90	MedNumBR	-0.23
## 91	PctSameHouse85	-0.24
## 92	PctWorkMom	-0.27
## 93	pctWRetire	-0.28
## 94	agePct65up	-0.30

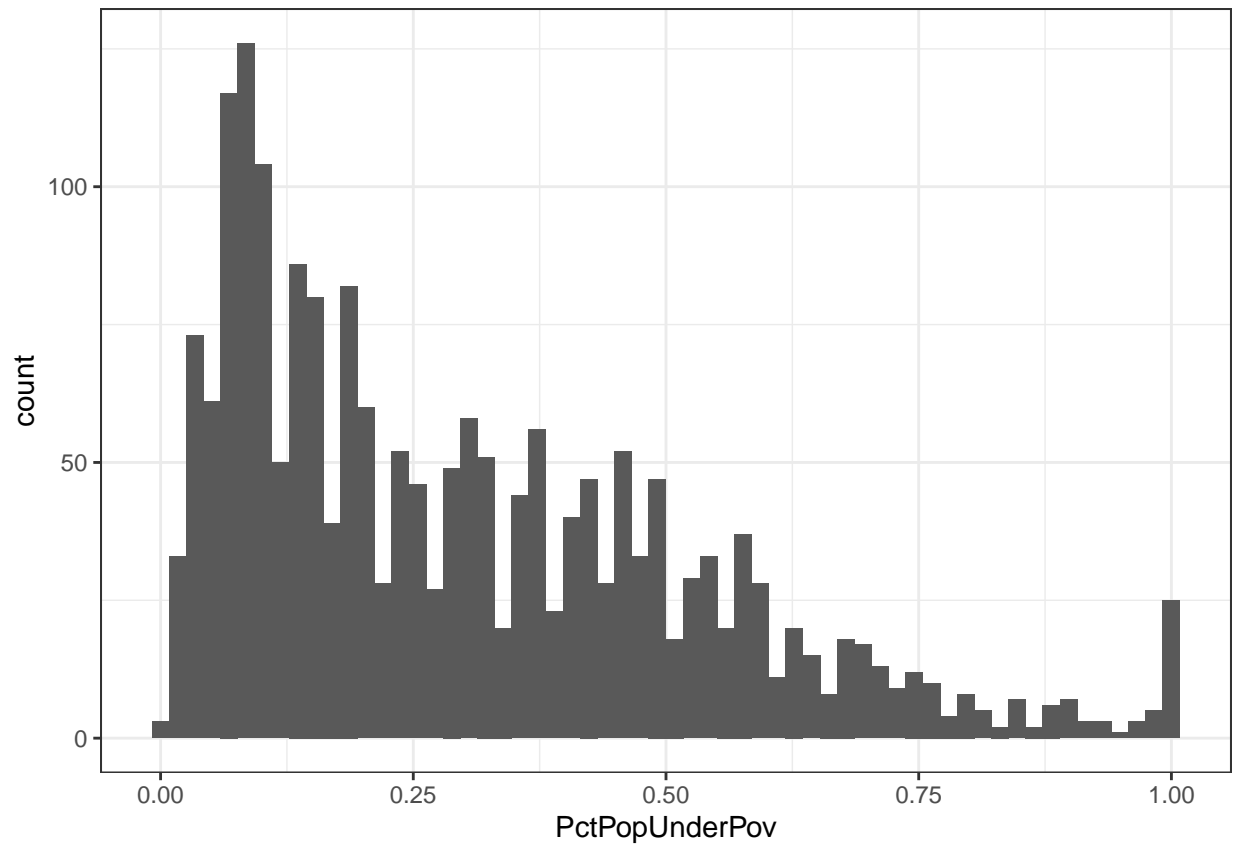
## 95	pctWFarmSelf	-0.31
## 96	whitePerCap	-0.31
## 97	AsianPerCap	-0.32
## 98	OwnOccHiQuart	-0.34
## 99	blackPerCap	-0.37
## 100	PctBSorMore	-0.38
## 101	RentHighQ	-0.38
## 102	PctOccupMgmtProf	-0.39
## 103	OwnOccMedVal	-0.39
## 104	MedRent	-0.39
## 105	LemasSwFTFieldOps	-0.39
## 106	OtherPerCap	-0.40
## 107	MedYrHousBuilt	-0.40
## 108	RentMedian	-0.40
## 109	pctWWage	-0.42
## 110	OwnOccLowQuart	-0.42
## 111	HispPerCap	-0.44
## 112	RentLowQ	-0.48
## 113	perCapInc	-0.53
## 114	PctEmploy	-0.53
## 115	RacialMatchCommPol	-0.54
## 116	PctPolicWhite	-0.54
## 117	PctHousOwnOcc	-0.55
## 118	PctHousOccup	-0.57
## 119	medIncome	-0.58
## 120	PctPersOwnOccup	-0.58
## 121	medFamInc	-0.60
## 122	pctWInvInc	-0.76
## 123	racePctWhite	-0.80
## 124	PctYoungKids2Par	-0.82
## 125	PctFam2Par	-0.85
## 126	PctTeen2Par	-0.85
## 127	PctKids2Par	-0.86

```
ggplot(communities, aes(x=PctPopUnderPov, y=ViolentCrimesPerPop)) +
  geom_point()+
  geom_smooth(method=lm)
```

```
## `geom_smooth()` using formula 'y ~ x'
```

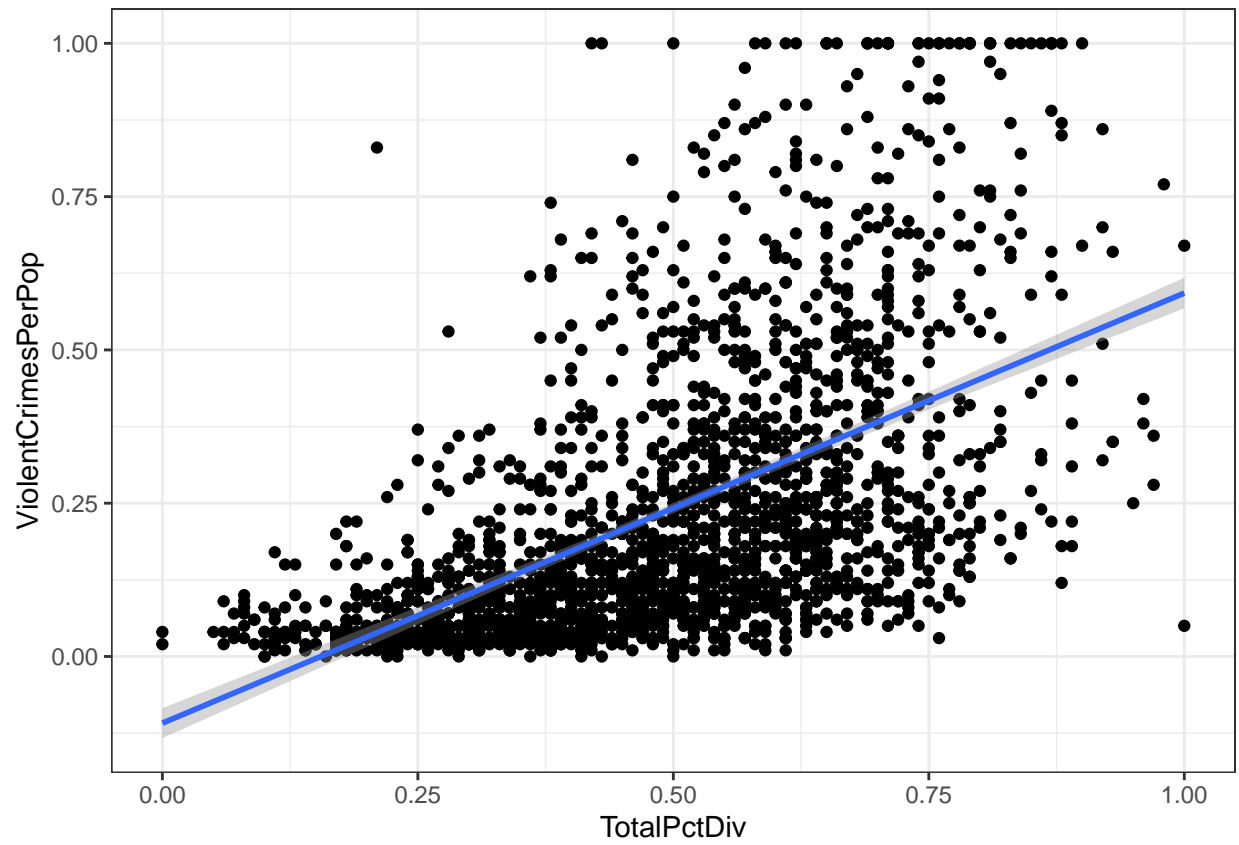


```
ggplot(communities, aes(x=PctPopUnderPov)) +  
  geom_histogram(bins=60)
```

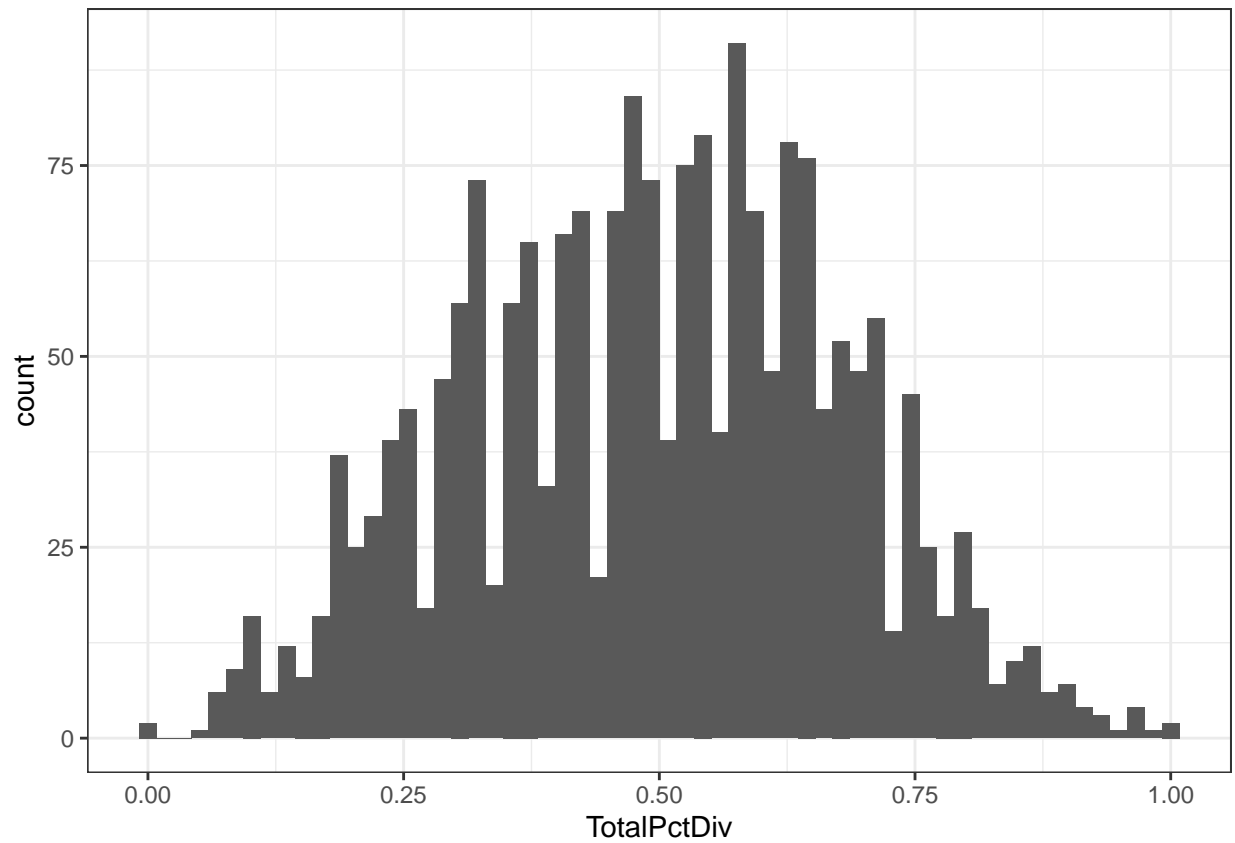


```
ggplot(communities, aes(x=TotalPctDiv, y=ViolentCrimesPerPop)) +  
  geom_point()+  
  geom_smooth(method=lm)
```

```
## `geom_smooth()` using formula 'y ~ x'
```



```
ggplot(communities, aes(x=TotalPctDiv)) +  
  geom_histogram(bins=60)
```

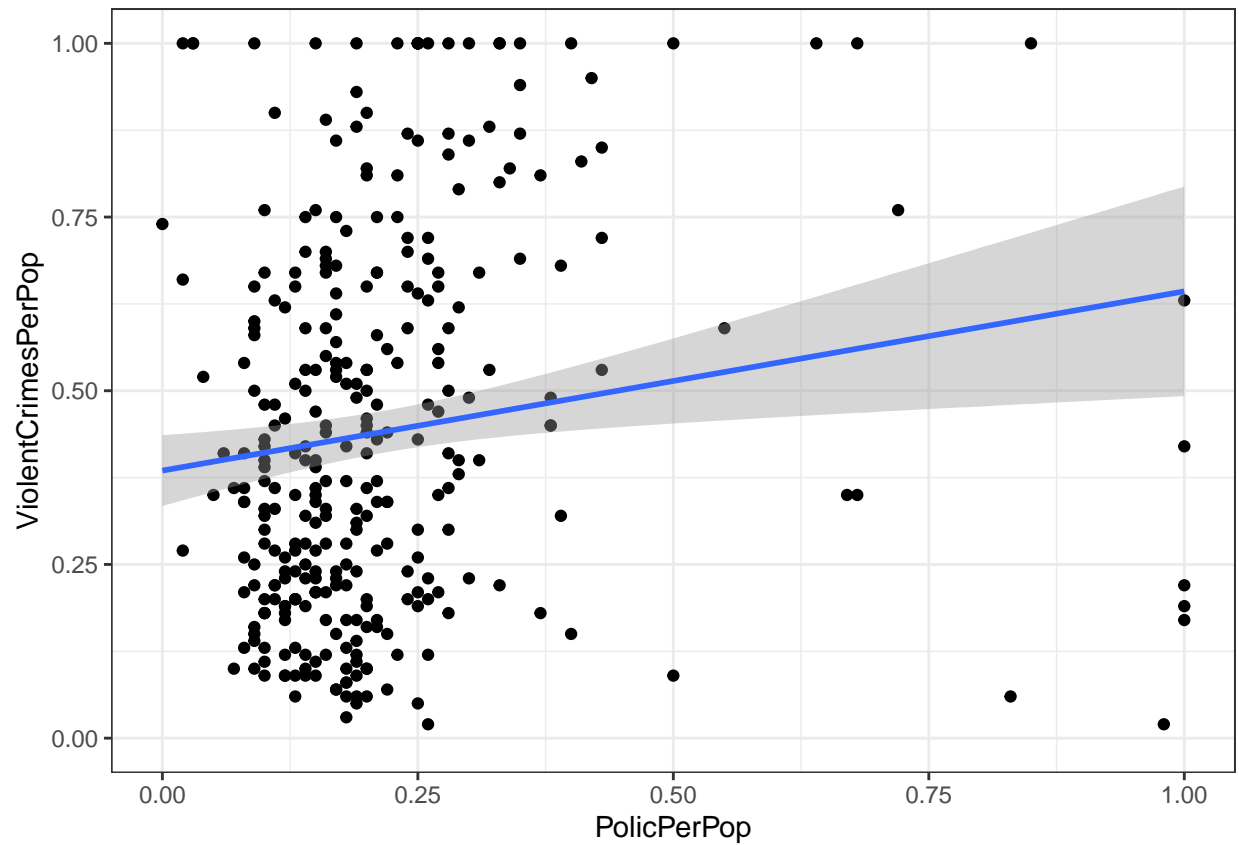


```
ggplot(communities, aes(x=PolicePerPop, y=ViolentCrimesPerPop)) +  
  geom_point()+  
  geom_smooth(method=lm)
```

```
## `geom_smooth()` using formula 'y ~ x'
```

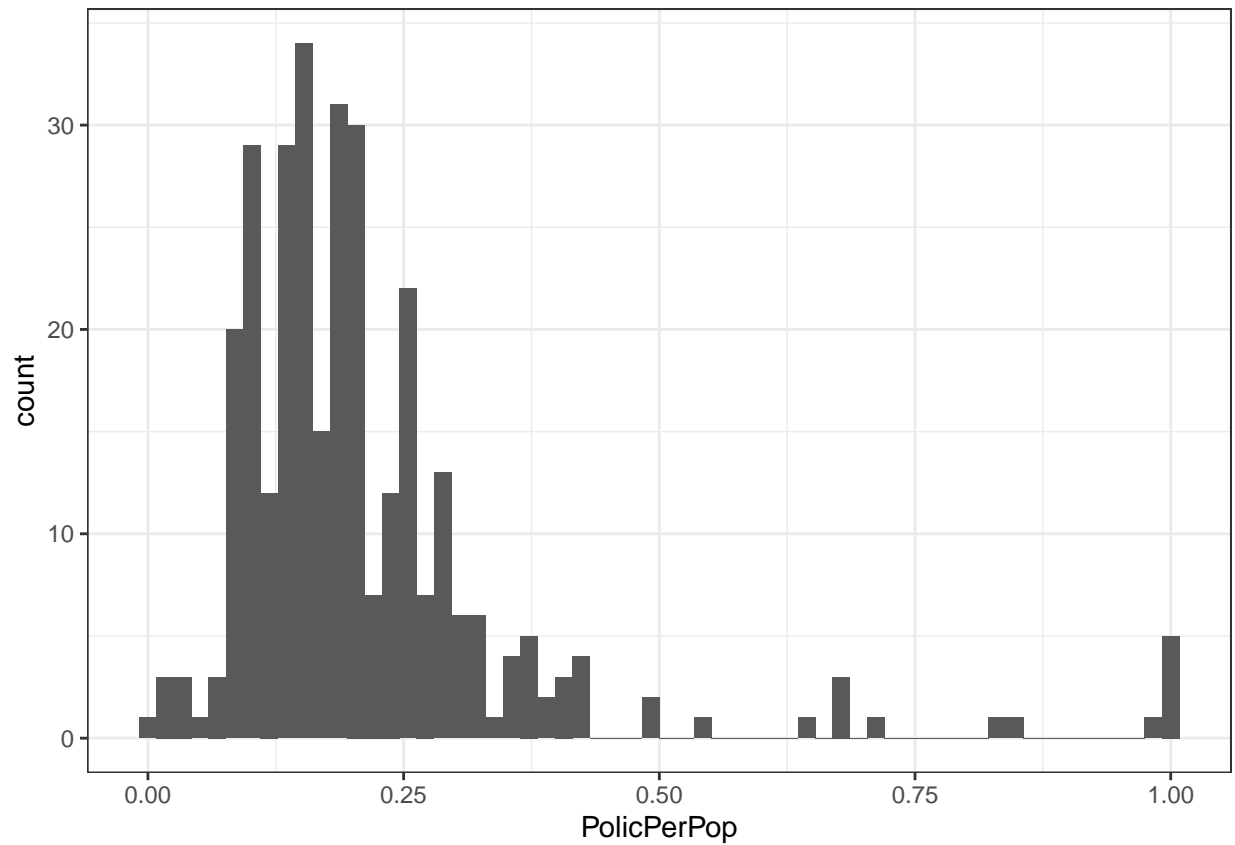
```
## Warning: Removed 1675 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 1675 rows containing missing values (geom_point).
```



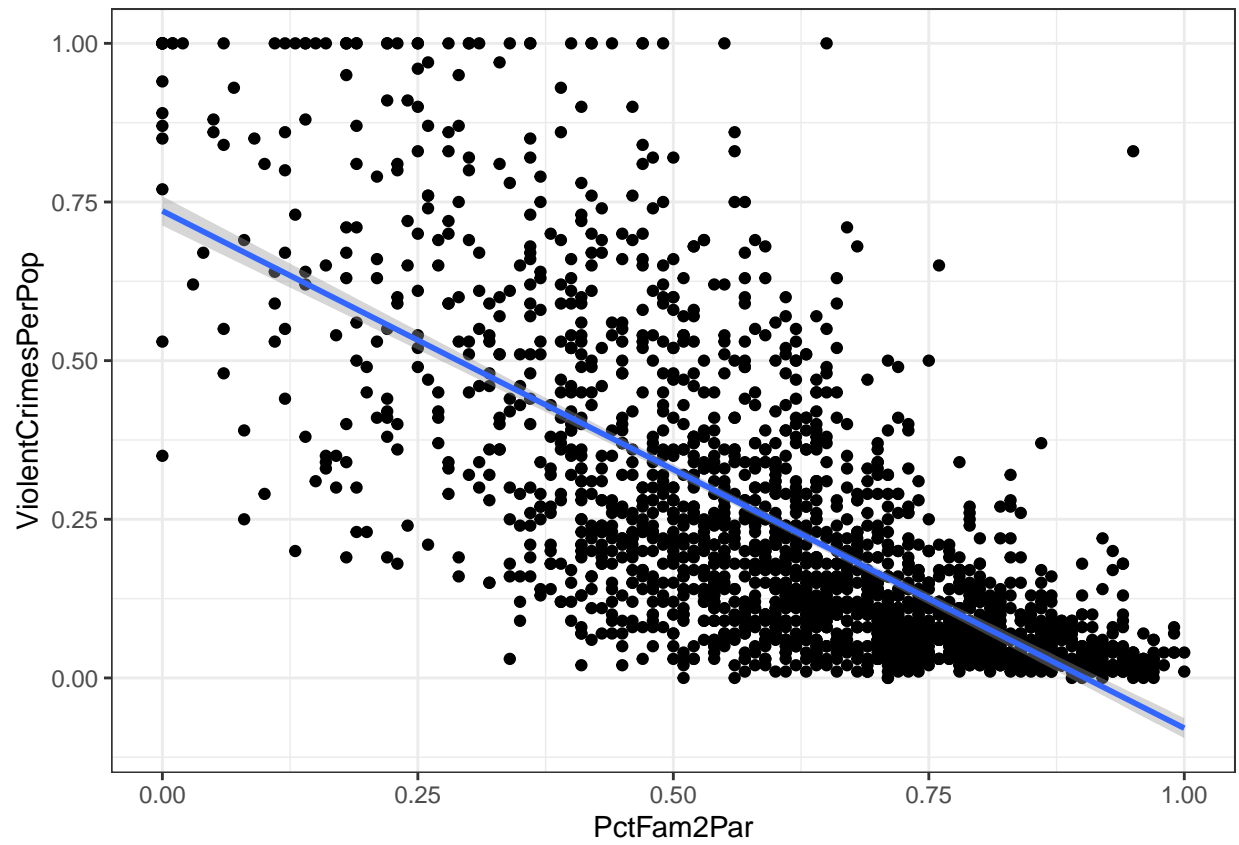
```
ggplot(communities, aes(x=PolicPerPop)) +  
  geom_histogram(bins=60)
```

```
## Warning: Removed 1675 rows containing non-finite values (stat_bin).
```

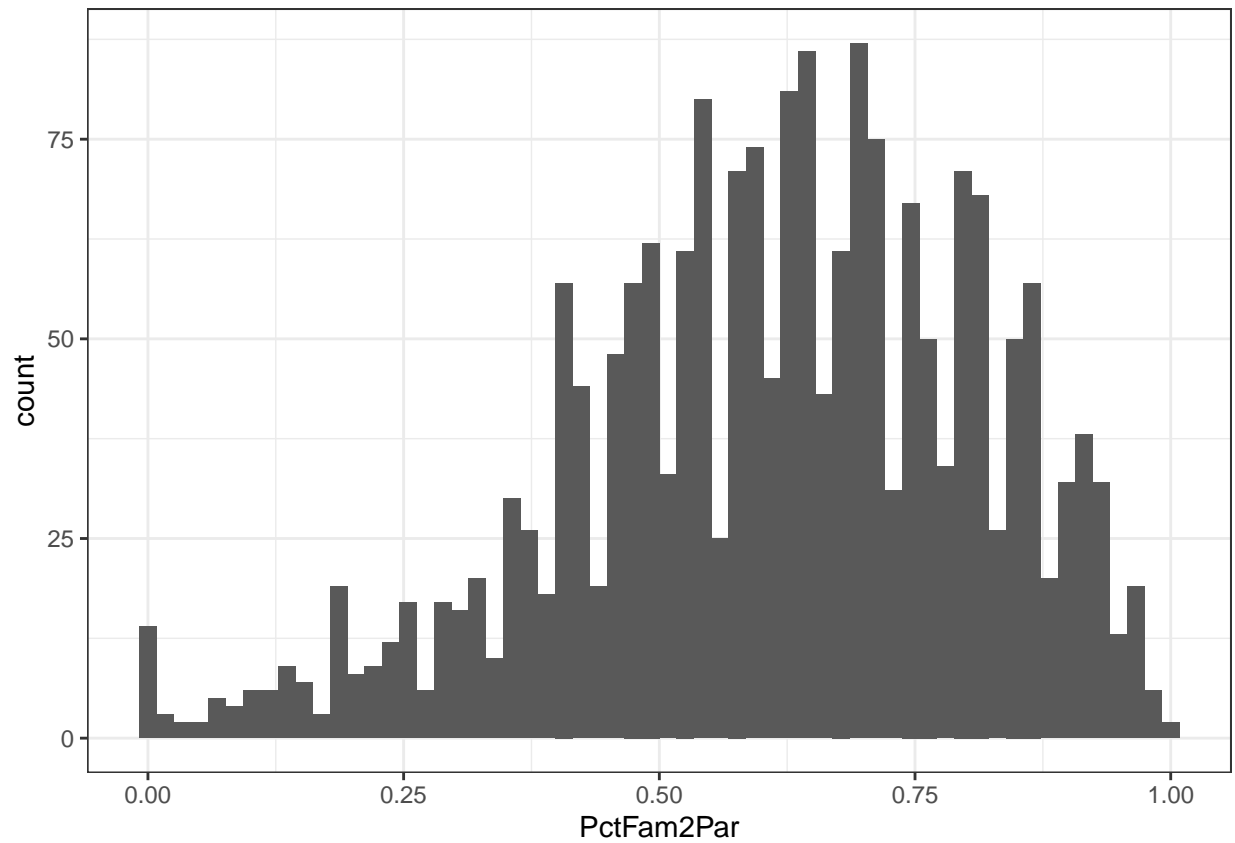



```
ggplot(communities, aes(x=PctFam2Par, y=ViolentCrimesPerPop)) +  
  geom_point()+  
  geom_smooth(method=lm)
```

```
## `geom_smooth()` using formula 'y ~ x'
```



```
ggplot(communities, aes(x=PctFam2Par)) +  
  geom_histogram(bins=60)
```

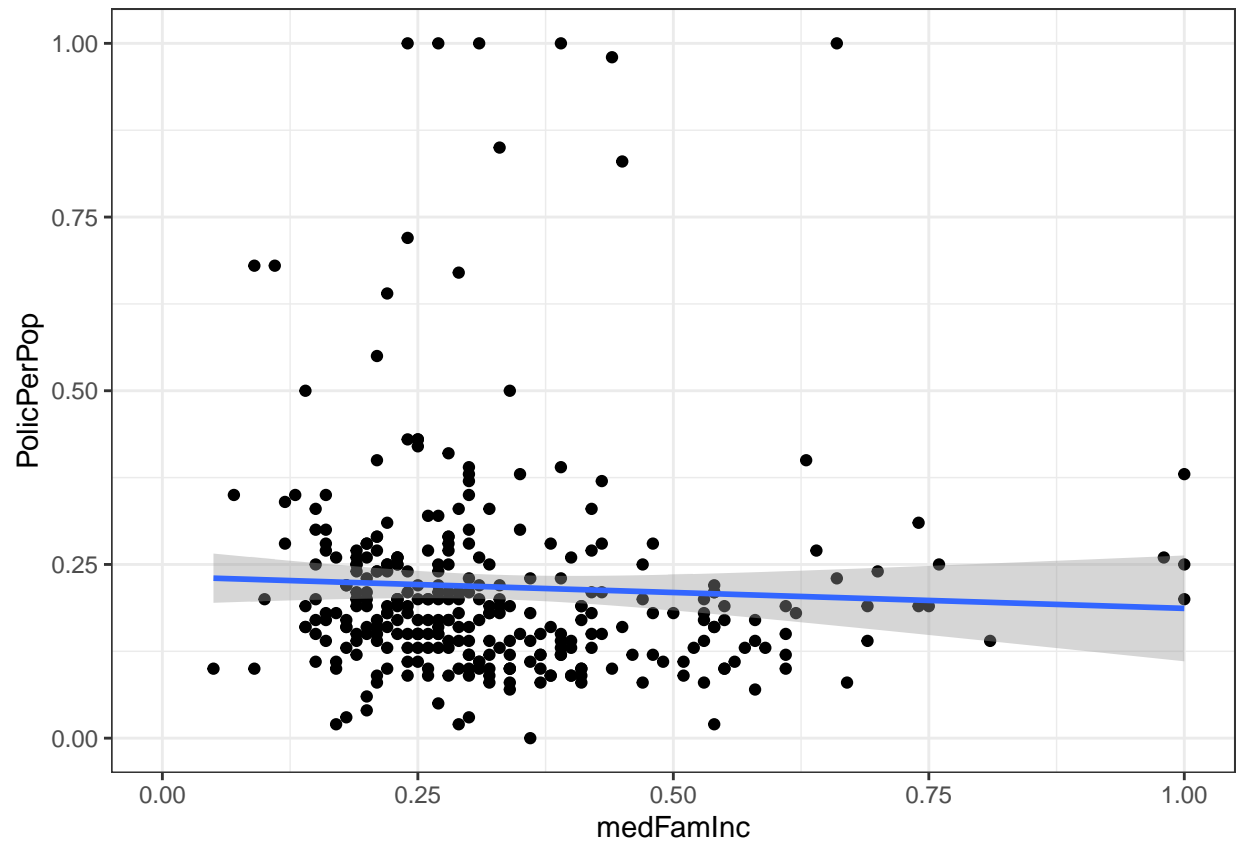


```
ggplot(communities, aes(x=medFamInc, y=PolicPerPop)) +
  geom_point()+
  geom_smooth(method=lm)
```

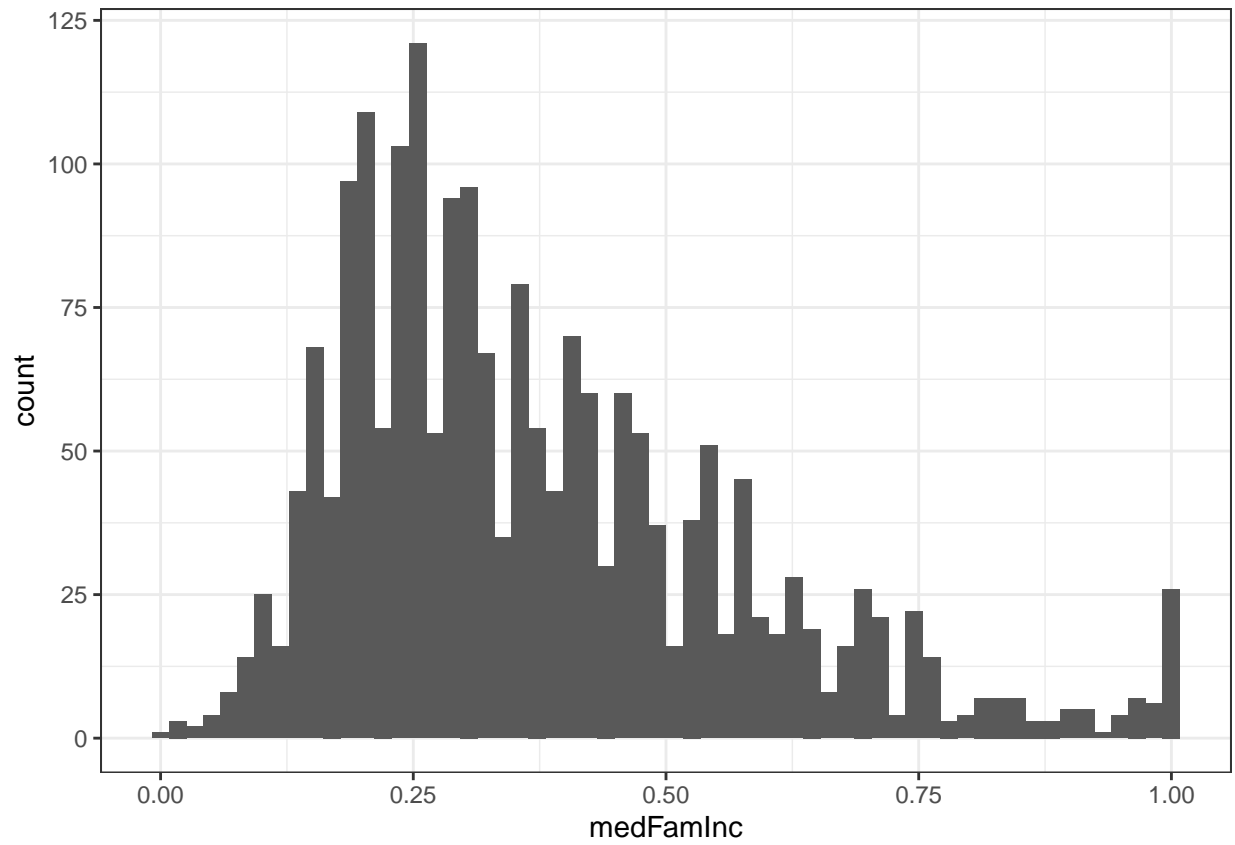
```
## `geom_smooth()` using formula 'y ~ x'
```

```
## Warning: Removed 1675 rows containing non-finite values (stat_smooth).
```

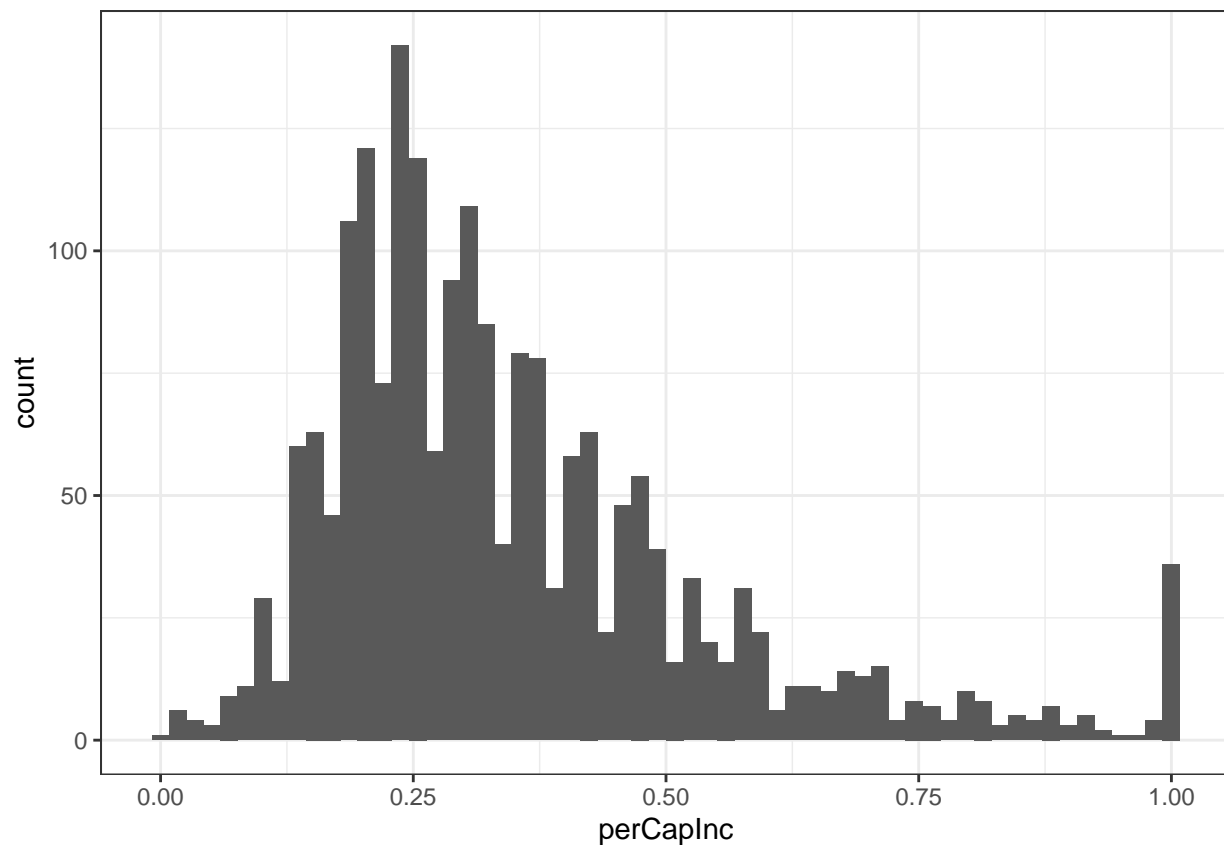
```
## Removed 1675 rows containing missing values (geom_point).
```



```
ggplot(communities, aes(x=medFamInc)) +  
  geom_histogram(bins=60)
```



```
ggplot(communities, aes(x=perCapInc)) +  
  geom_histogram(bins=60)
```



```
crime_fit_01 <- communities %>% lm(ViolentCrimesPerPop ~ PctIlleg + HousVacant + PctUnemployed + NumStr
crime_fit_01_se <- crime_fit_01 %>% vcovHC(type = "HC1") %>% diag() %>% sqrt()
stargazer(crime_fit_01, type="text", se = list(crime_fit_01_se))
```

```
##
## =====
##                               Dependent variable:
##                               -----
##                               ViolentCrimesPerPop
##                               -----
## PctIlleg                      0.535***
##                               (0.026)
##
## HousVacant                    0.166***
##                               (0.036)
##
## PctUnemployed                 0.098***
##                               (0.021)
##
## NumStreet                     0.191***
##                               (0.052)
##
## TotalPctDiv                   0.245***
##                               (0.021)
##
## Constant                      -0.069***
```

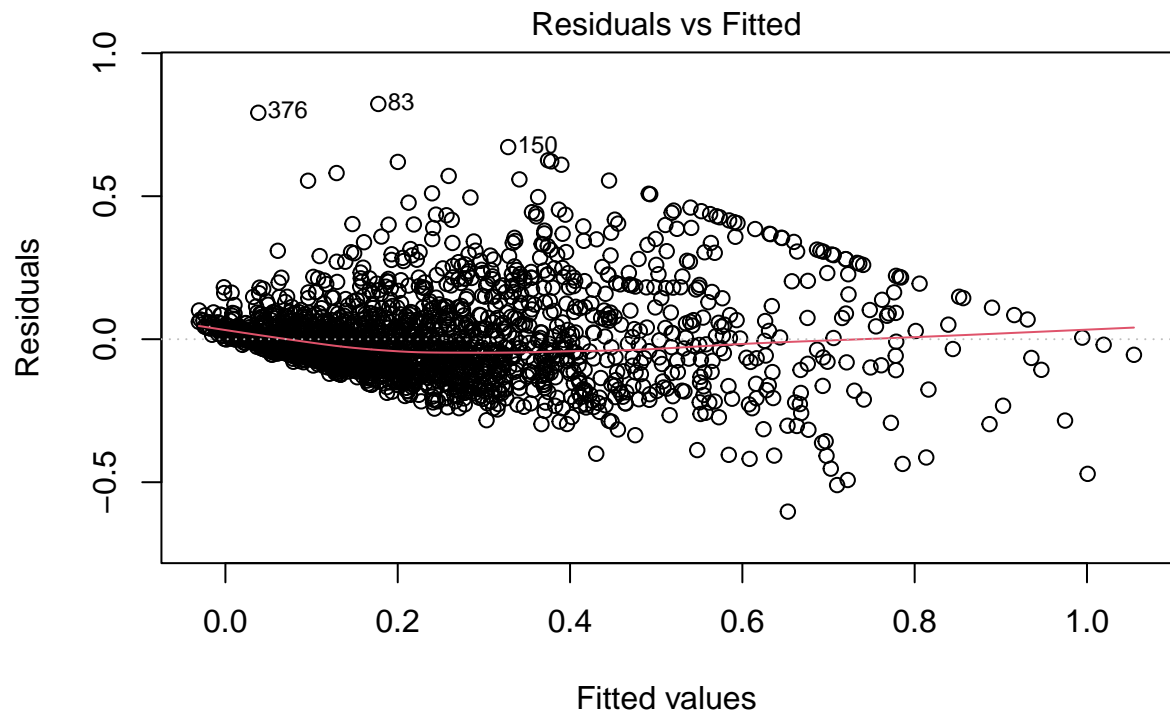
```

##                                     (0.009)
##
## -----
## Observations                1,994
## R2                          0.612
## Adjusted R2                 0.611
## Residual Std. Error        0.145 (df = 1988)
## F Statistic                 628.017*** (df = 5; 1988)
## =====
## Note:                       *p<0.1; **p<0.05; ***p<0.01
coeftest(crime_fit_01, vconv = vcovHC(type = "HC1"))

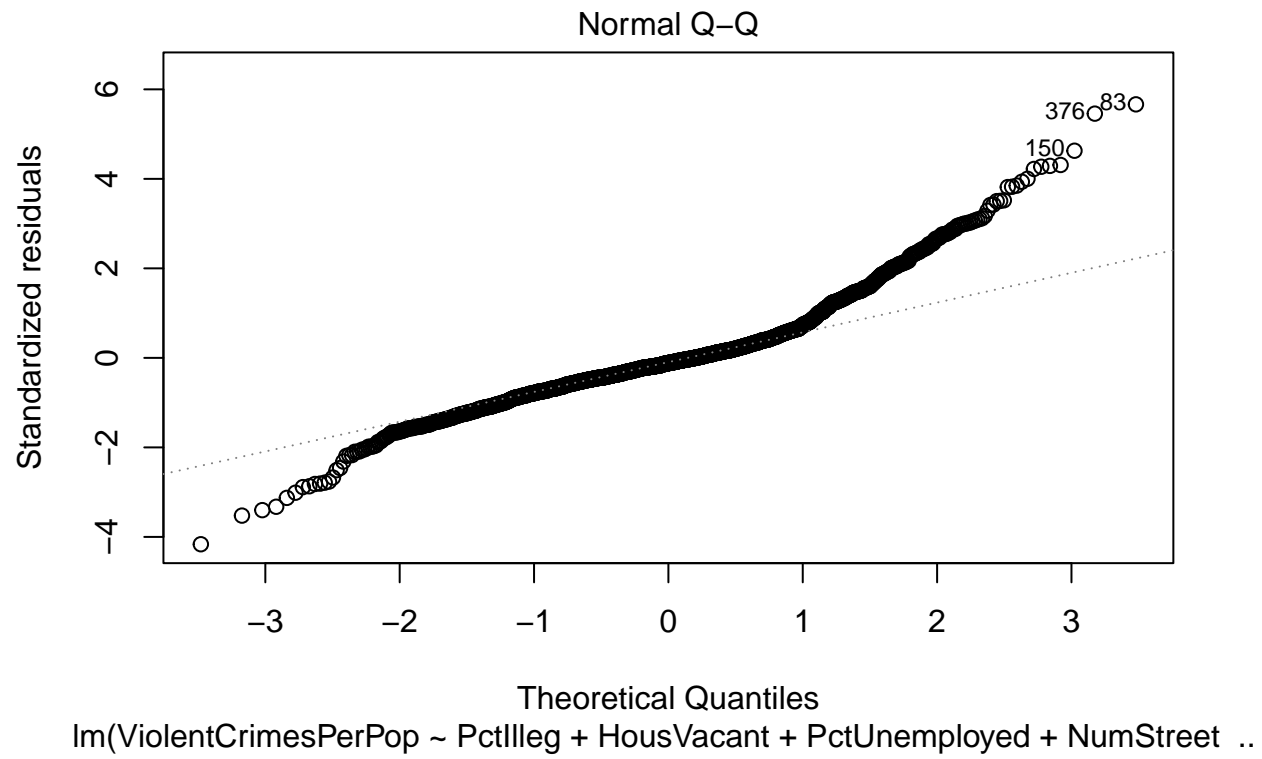
##
## t test of coefficients:
##
##      Estimate Std. Error t value  Pr(>|t|)
## (Intercept)  -0.069422   0.010138 -6.8479 9.952e-12 ***
## PctIlleg      0.535219   0.019517 27.4233 < 2.2e-16 ***
## HousVacant    0.166001   0.028133  5.9006 4.247e-09 ***
## PctUnemployed 0.097642   0.020132  4.8502 1.329e-06 ***
## NumStreet     0.190724   0.040448  4.7153 2.581e-06 ***
## TotalPctDiv   0.244818   0.021241 11.5256 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

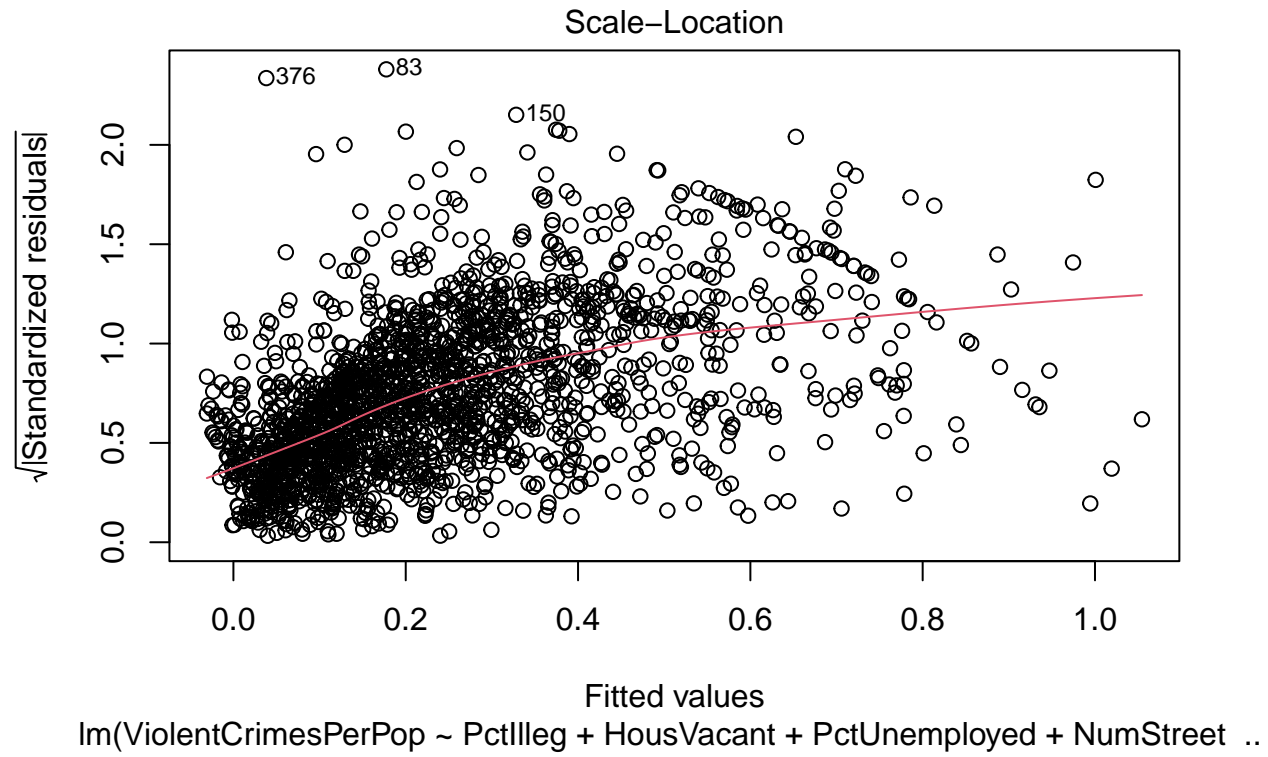
fit_o1_vif = ols_vif_tol(crime_fit_01)
plot(crime_fit_01)

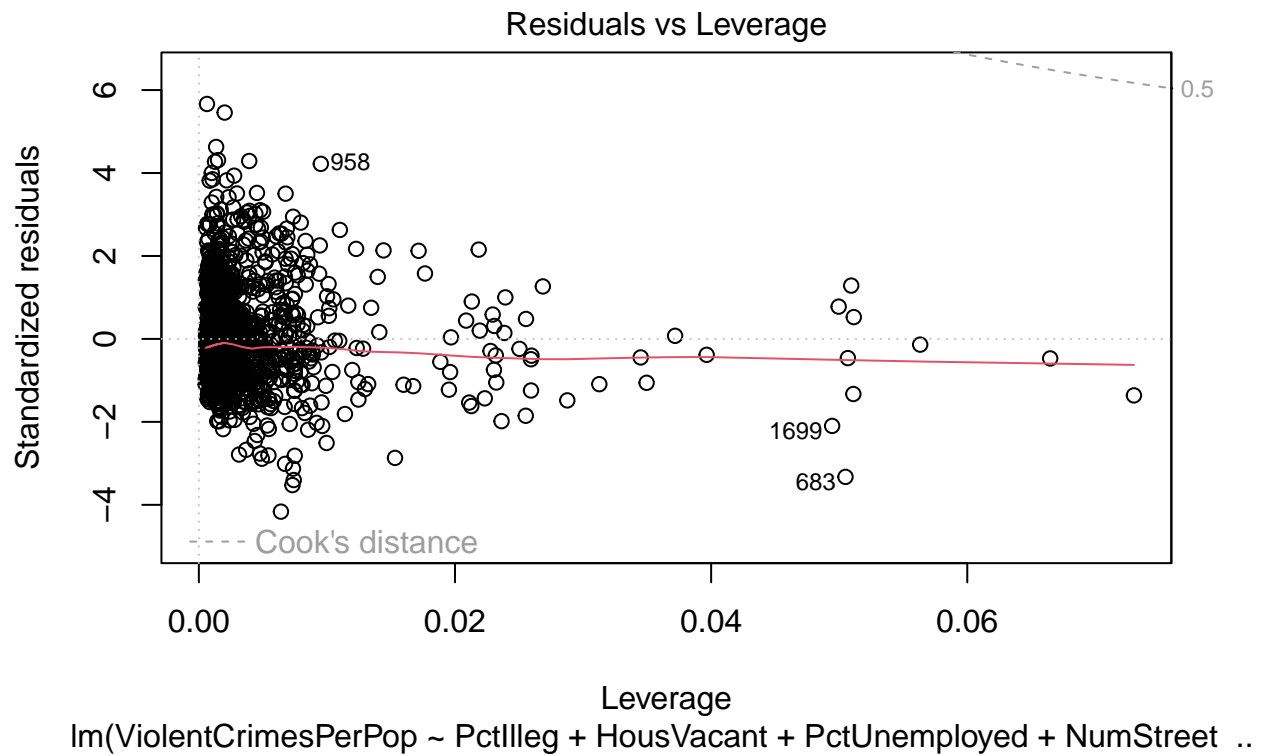
```



lm(ViolentCrimesPerPop ~ PctIlleg + HousVacant + PctUnemployed + NumStreet ..







```
lmtest::bptest(crime_fit_01)
```

```
##
## studentized Breusch-Pagan test
##
## data: crime_fit_01
## BP = 240.92, df = 5, p-value < 2.2e-16
```