

How Do the Factors of Number of Police, Average Income, Unemployment and School Rating Impact the Safety of a City?

The June-ypers

June, Fang, Tigist, Christian P., Kyle

Not to be mistaken with the June-pyters

Question to Answer

- Our original question was: Can we use crime statistics to better a community?
 - The thought process behind this was can we could use a crime statistics data API to show where the most prevalent amount of crime happens in a city, and then distribute the officers in that area, according to the type/count of crime(s)?
 - We ran into the issue of insufficient data. The crime data API we found, did not have zip codes for the crimes themselves.
 - We have crime data, but it was based on the arresting police department.
- A new question formed: Which cities are safest, and why?
- Sample Size: State capitals and their respective police departments.
- Crime Variables: Rape, Simple Assaults, and Stolen Property
- How do these variables affect crime?.
 - Number of police officers
 - School Ratings
 - Avg. Unemployment Rate
 - Avg Income

Issues with data, and made assumptions

- We could not find an API to get police department codes
 - The API source we used is an FBI data source. In order to get our specific crime numbers for capital cities, we needed police department codes, the ORI. Unfortunately there was no quick way to find them all, as there were 50+ in some locations, and they are broken out by county.
 - This will cause the numbers to be skewed, mainly a rather low amount of police officers.
 - Source: <https://crime-data-explorer.fr.cloud.gov/api>
- School Ratings
 - The school ratings are based on an anonymous user rating system. Basically anyone can go in and rate the schools.
 - Statistically, you are far more likely to rate/take time to comment only if you had a bad experience.
 - <https://www.fastcompany.com/90425093/why-your-brains-so-bad-at-letting-go-of-negative-comments>
 - Schools can also be rated by someone who has no affiliation towards the school.
 - Rankings out of 10
 - Includes Public and Private schools
- Unemployment Rating
 - Unemployment ratings were taken from the census, based on the state. This one was added late in our data collection, and we didn't have enough time to lookup the rate for each city.
 - Script used from in-class instructor demo, and modified to suit just the unemployment and income ratings
- Avg. Income
 - Same as above
 - Using Household Income

Getting Data-1

Example: Alabama Capital City, Montgomery Police Department

```
response = requests.get("https://api.usa.gov/crime/fbi/sapi/api/police-employment/agencies/AL0030100/2018/2018?API_
pprint(response)

{'pagination': {'count': 1, 'page': 0, 'pages': 1, 'per_page': 0},
 'results': [{}'agency_name_edit': 'MONTGOMERY PD',
            'agency_type_name': 'City',
            'civilian_ct': 92,
            'csv_header': 'data_year,ori,state_abbr,agency_name_edit,agency_type_name,ncic_agency_name,civilian_c
t,female_civilian_ct,female_officer_ct,female_total_ct,male_civilian_ct,male_officer_ct,male_total_ct,population,po
lice_emp_per_1000,total_pe_ct',
            'data_year': 2018,
            'female_civilian_ct': 70,
            'female_officer_ct': 94,
            'female_total_ct': 164,
            'male_civilian_ct': 22,
            'male_officer_ct': 380,
            'male_total_ct': 402,
            'ncic_agency_name': 'MONTGOMERY PD',
            'ori': 'AL0030100',
            'pe_ct_per_1000': 2.85,
            'population': 198662,
            'state_abbr': 'AL',
            'total_pe_ct': 566}]}]
```

Getting Data-2

```
# Getting data from 2018
ori_list =["AL0030100",
           "AK0010300",
           "AZ0072300",
           "AR0600200",
           "CA0340400",
           "COPD0000",
           "CT0006400",
           "DE0010100",
           "FL0370300",
           "GAAPD0000"
         ]
female =[]
male =[]

begin_year = 2018
end_year = 2018

print("Beginning Data Retrieval")
print("-----")

for ori in range(len(ori_list)):
    response = requests.get(f"https://api.usa.gov/crime/fbi/sapi/api/police-employment/agencies/{ori_list[ori]}/{beg
    #print(response)
    try:
        print(f"Processing record {ori + 1}")
        female.append(response["results"][0]["female_officer_ct"])
        male.append(response["results"][0]["male_officer_ct"])
    except:
        print("Not found")
        pass
print("-----")
print("Data Retrieval Complete")
print("-----")
```

Beginning Data Retrieval

Processing record 1
Processing record 2
Processing record 3
Processing record 4

Getting Data-3

Example: Dataframe of 10 States

```
officer_2018_df = pd.DataFrame({"Male Officers": male, "Female Officers" : female})  
officer_2018_df
```

	Male Officers	Female Officers
0	380	94
1	42	5
2	2524	395
3	475	100
4	549	102
5	1309	208
6	346	48
7	95	6
8	309	68
9	1283	252

```
officer_2018_df["Total Number"] = officer_2018_df["Male Officers"]+officer_2018_df["Female Officers"]  
officer_2018_df
```

	Male Officers	Female Officers	Total Number
0	380	94	474
1	42	5	47
2	2524	395	2919
3	475	100	575
4	549	102	651
5	1309	208	1517
6	346	48	394
7	95	6	101
8	309	68	377
9	1283	252	1535

Getting Data-4

Example: Alabama Capital City, Montgomery Police Department

```
# Look up dataset
response = requests.get(f"https://api.usa.gov/crime/fbi/sapi/api/data/arrest/agencies/offense/AL0030100/all/2018/201
pprint(response)

{'pagination': {'count': 1, 'page': 0, 'pages': 1, 'per_page': 0},
 'results': [{}'aggravated_assault': 38,
            'all_other_offenses': 28,
            'arson': 0,
            'burglary': 58,
            'csv_header': None,
            'curfew': 0,
            'data_year': 2018,
            'disorderly': 0,
            'driving': 2,
            'drug_abuse_gt': 32,
            'drug_poss_m': 1,
            'drug_poss_opium': 6,
            'drug_poss_other': 23,
            'drug_poss_subtotal': 32,
            'drug_poss_synthetic': 2,
            'drug_sales_m': 0,
            'drug_sales_opium': 0,
            'drug_sales_other': 0,
            'drug_sales_subtotal': 0,
            'drug_sales_synthetic': 0,
            'drunkenness': 0,
            'embezzlement': 2,
            'forgery': 6,
            'fraud': 8,
            'g_all': 0,
            'g_b': 0,
            'g_n': 0,
            'g_t': 0,
            'ht_c_s_a': 0,
            'ht_i_s': 0,
            'larceny': 116,
            'liquor': 0,
            'manslaughter': 0,
            'murder': 5,
            'mvt': 42,
            'offense_family': 0,
            'prostitution': 0,
            'prostitution_a_p_p': 0,
            'prostitution_p': 0,
            'prostitution_p_p': 0,
            'rape': 3,
            'robbery': 43,
            'sex offense': 3,
            'simple_assault': 102,
            'stolen_property': 13,
            'suspicion': 0,
            'vagrancy': 0,
            'vandalism': 3,
            'weapons': 17}]}}
```

Getting Data-5

```
# Getting 2018 data
simple_assault_cases = []
stolen_property_cases = []
rape_cases = []

begin_year = 2018
end_year = 2018

ori_list = ori

print("Beginning Data Retrieval")
print("-----")

for ori in range(len(ori_list)):
    response = requests.get(f"https://api.usa.gov/crime/fbi/sapi/api/data/arrest/agencies/offense/{ori_list[ori]}/all")
    try:
        print(f"Processing record {ori + 1}")
        simple_assault_cases.append(response["results"][0]["simple_assault"])
        stolen_property_cases.append(response["results"][0]["stolen_property"])
        rape_cases.append(response["results"][0]["rape"])
    except:
        print("Not found")
        pass
    print("-----")
print("Data Retrieval Complete")
print("-----")
```

Beginning Data Retrieval

Processing record 1
Processing record 2
Processing record 3
Processing record 4
Processing record 5
Processing record 6
Processing record 7

Getting Data-6

Example: Dataframe of 10 States

```
df_2018 = pd.DataFrame(list(zip(ori_list, simple_assault_cases, stolen_property_cases, rape_cases)),  
                      columns = ['Police Department', "Simple Assault Cases", "Stolen Property Cases", "Rape Cases"])  
df_2018
```

	Police Department	Simple Assault Cases	Stolen Property Cases	Rape Cases
0	AL0030100	102	13	3
1	AK0010300	291	4	5
2	AZ0072300	5726	364	103
3	AR0600200	947	291	28
4	CA0340400	1309	281	30
5	CODPD0000	3435	132	143
6	CT0006400	1780	9	17
7	DE0010100	512	26	5

Getting Data-7

Example: Dataframe of 10 States

Coding reference: Python-APIs 09-Stu_Census_API_Gmap

```
# Final DataFrame
census_pd = census_pd[["State", "Name", "Household Income", "Unemployment Rate"]]
census_pd.head(10)
```

	State	Name	Household Income	Unemployment Rate
0	27	Minnesota	68411.0	2.156871
1	28	Mississippi	43567.0	3.681658
2	29	Missouri	53560.0	2.560861
3	30	Montana	52559.0	2.153433
4	31	Nebraska	59116.0	1.918614
5	32	Nevada	57598.0	3.481124
6	33	New Hampshire	74057.0	2.261946
7	34	New Jersey	79363.0	3.209739
8	35	New Mexico	48059.0	3.267487
9	36	New York	65323.0	3.071781

Getting Data-8

Example for School Rating: Alabama Capital City, Montgomery

<https://www.greatschools.org>

GreatSchools.org Schools City, zip, address or school PARENTING DONATE EN ESPAÑOL SIGN IN ListMap view Table view

Public & private schools PreK Elementary Middle High

Alabama > Montgomery Showing 1 to 25 of 34 schools found in Montgomery, AL

Overview Ratings Snapshot Equity: Test scores Sort by: GreatSchools Rating

School	Type	Grades	Total students enrolled	Students per teacher	Reviews	District
10 Booker T Washington Magnet High School Top rated 612 South Union Street, Montgomery, AL, 36104  Homes for sale	Public district	9-12	406	14:1	15 reviews 	Montgomery County School District
10 Brewbaker Tech Magnet High School Top rated 4005 Brewbaker Drive, Montgomery, AL, 36116  Homes for sale	Public district	9-12	530	15:1	18 reviews 	Montgomery County School District
10 Lavelles Academic Magnet Prog High school Top rated 215 Hall Street, Montgomery, AL, 36104  Homes for sale	Public district	9-12	507	14:1	24 reviews 	Montgomery County School District
3 Park Crossing High School Below average 8100 Park Crossing, Montgomery, AL, 36117  Homes for sale	Public district	9-12	1249	19:1	16 reviews 	Montgomery County School District
1 Carver Senior High School Below average 201 West Fairview Avenue, Montgomery, AL, 36108  Homes for sale	Public district	9-12	1098	16:1	14 reviews 	Montgomery County School District
1 Jefferson Davis High School Below average 3120 Carter Hill Road, Montgomery, AL, 36111  Homes for sale	Public district	9-12	1760	17:1	14 reviews 	Montgomery County School District
1 Lee High School Below average 215 Ann Street, Montgomery, AL, 36107  Homes for sale	Public district	9-12	1528	17:1	5 reviews 	Montgomery County School District
1 Lanier Senior High School Below average 1156 South Court Street, Montgomery, AL, 36104  Homes for sale	Public district	9-12	1008	19:1	4 reviews 	Montgomery County School District
 Fews Secondary Alternative School Current: 321 Early Street, unrated Montgomery, AL, 36104  Homes for sale	Public district	6-10	54	7:1	8 reviews 	Montgomery County School District
 Childrens Center Current: 310 North Madison Terrace, Montgomery, unrated AL, 36107  Homes for sale	Public district	PK-12	92	5:1	3 reviews 	Montgomery County School District

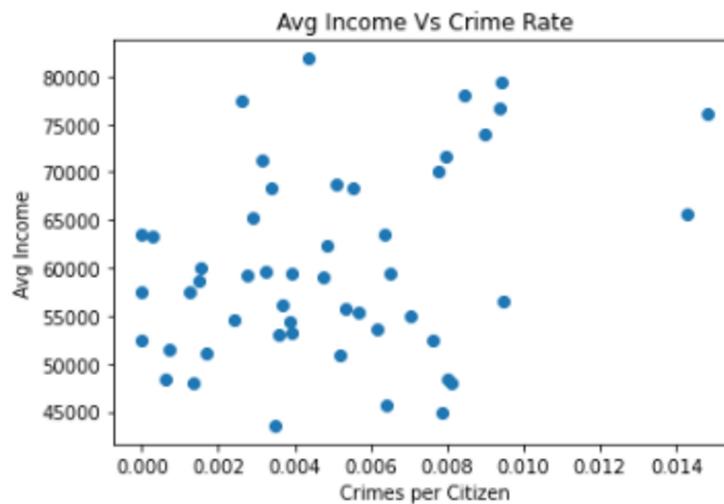
Table of Stats

State	State	State	Population	ORI Code	No. of	Year of	No. of	No. of Simple	No. of Stolen	Total	School	Unemployment	Avg.	Crimes(R,SA,SP)
	Ab.	Capital			Officers	Data	Rapes	Assaults	Property Cases	Crimes	Ratings	Rate	Income	Per Citizen
ALABAMA	AL	Montgome	198,525	AL0030100	474	2018	3	102	13	118	4.35	3.04	\$ 48,486.00	0.000594
ALASKA	AK	Juneau	32,113	AK0010300	47	2018	5	291	4	300	6.50	3.8	\$ 76,715.00	0.009342
ARIZONA	AZ	Phoenix	1,680,992	AZ0072300	2919	2018	103	5726	364	6193	4.72	3.03	\$ 56,213.00	0.003684
ARKANSAS	AR	Little Rock	197,312	AR0600200	575	2018	28	947	291	1266	3.64	2.53	\$ 45,726.00	0.006416
CALIFORNIA	CA	Sacramento	513,624	CA0340400	651	2018	30	1309	281	1620	4.63	3.38	\$ 71,228.00	0.003154
COLORADO	CO	Denver	727,211	COPDP0000	1517	2018	143	3435	132	3710	3.75	2.52	\$ 68,811.00	0.005102
CONNECTICUT	CT	Hartford	122,105	CT0006400	394	2018	17	1780	9	1806	2.80	3.52	\$ 76,106.00	0.014791
DELAWARE	DE	Dover	38,079	DE0010100	101	2018	5	512	26	543	n/a	3	\$ 65,627.00	0.014260
FLORIDA	FL	Tallahassee	194,500	FL0370300	377	2017	45	713	8	766	5.00	3.02	\$ 53,267.00	0.003938
GEORGIA	GA	Atlanta	506,811	GAAPD0000	1535	2017	45	2144	511	2700	3.85	3.14	\$ 55,679.00	0.005327
HAWAII	HI	Honolulu	345,064	HI0020000	1942	2018	51	2407	449	2907	6.33	2.25	\$ 78,084.00	0.008425
IDAHO	ID	Boise	228,959	ID0010100	292	2018	54	741	21	816	4.11	2.23	\$ 53,089.00	0.003564
ILLINOIS	IL	Springfield	114,230	IL0840200	239	2018	0	0	0	0	3.84	3.41	\$ 63,575.00	0.000000
INDIANA	IN	Indianapolis	876,384	INIPD0000	2080	2018	81	3206	124	3411	2.78	2.73	\$ 54,325.00	0.003892
IOWA	IA	Des Moines	214,237	IA0770300	0	2018	12	299	7	318	1.14	2.08	\$ 58,580.00	0.001484
KANSAS	KS	Topeka	125,310	KS0890100	267	2018	0	0	0	0	3.57	2.24	\$ 57,422.00	0.000000
KENTUCKY	KY	Frankfort	27,679	KY0370100	57	2018	0	190	32	222	6.00	2.89	\$ 48,392.00	0.008021
LOUISIANA	LA	Baton Rouge	220,236	LA0170200	631	2018	34	1467	281	1782	3.38	3.21	\$ 47,942.00	0.008091
MAINE	ME	Augusta	18,681	ME0060100	45	2018	0	104	2	106	3.83	2.42	\$ 55,425.00	0.005674
MARYLAND	MD	Annapolis	39,174	MD0020100	119	2018	5	162	3	170	3.67	3.02	\$ 81,868.00	0.004340
MASSACHUSETTS	MA	Boston	692,600	MA0130100	2122	2018	43	1612	169	1824	2.97	2.98	\$ 77,378.00	0.002634
MICHIGAN	MI	Lansing	118,210	MI3351900	188	2018	17	791	22	830	3.34	3.22	\$ 54,938.00	0.007021
MINNESOTA	MN	Saint Paul	308,096	MN0620900	631	2018	16	984	43	1043	3.69	2.16	\$ 68,411.00	0.003385
MISSISSIPPI	MS	Jackson	160,628	MS0250100	0	2018	8	445	106	559	2.00	3.68	\$ 43,567.00	0.003480
MISSOURI	MO	Jefferson C	42,838	MO0260100	88	2018	3	260	0	263	5.50	2.56	\$ 53,560.00	0.006139
MONTANA	MT	Helena	32,315	MT0250100	52	2018	5	239	2	246	5.50	2.15	\$ 52,559.00	0.007613
NEBRASKA	NE	Lincoln	289,102	NB0550100	344	2018	19	1301	55	1375	2.50	1.92	\$ 59,116.00	0.004756
NEVADA	NV	Carson City	55,916	NV0130000	95	2018	6	46	17	69	3.00	3.48	\$ 57,598.00	0.001234
NEW HAMPSHIRE	NH	Concord	43,627	NH0071600	82	2018	1	340	51	392	3.00	2.26	\$ 74,057.00	0.008985
NEW JERSEY	NJ	Trenton	83,203	NJ0111100	287	2018	6	652	125	783	2.97	3.2	\$ 79,363.00	0.009411
NEW MEXICO	NM	Santa Fe	84,683	NM0260100	150	2018	3	66	43	112	5.35	3.27	\$ 48,059.00	0.001323
NEW YORK	NY	Albany	96,460	NY0010100	316	2018	9	234	35	278	4.74	3.07	\$ 65,323.00	0.002882
NORTH CAROLINA	NC	Raleigh	474,069	NC0920100	732	2018	0	0	0	0	6.13	3.08	\$ 52,413.00	0.000000
NORTH DAKOTA	ND	Bismarck	73,529	ND0080100	127	2018	7	406	53	466	6.50	1.55	\$ 63,473.00	0.006338
OHIO	OH	Columbus	898,553	OHCOP0000	1870	2018	53	1966	139	2158	3.33	2.95	\$ 54,533.00	0.002402
OKLAHOMA	OK	Oklahoma	655,057	OK0550400	96	2018	3	417	34	454	4.48	2.53	\$ 51,424.00	0.000693

Scatter Plots

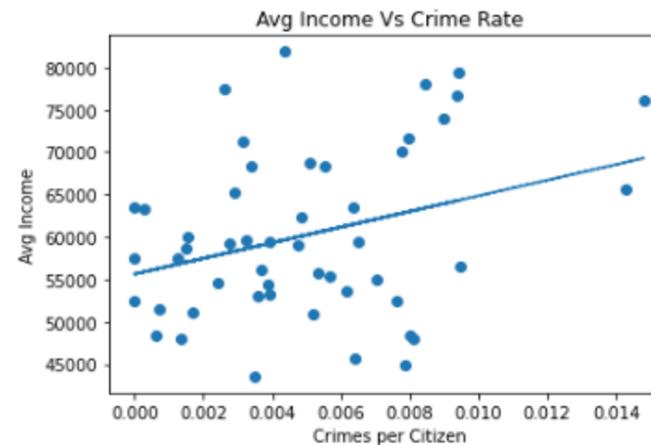
We started with our original Plots

```
1 avg_income_ratings = file_df.iloc[:,13]
2 crime_rate = file_df.iloc[:,14]
3 plt.scatter(crime_rate, avg_income_ratings)
4 plt.ylabel('Avg Income')
5 plt.xlabel('Crimes per Citizen')
6 plt.title('Avg Income Vs Crime Rate')
7 plt.show()
```



Then we added our linear regression script

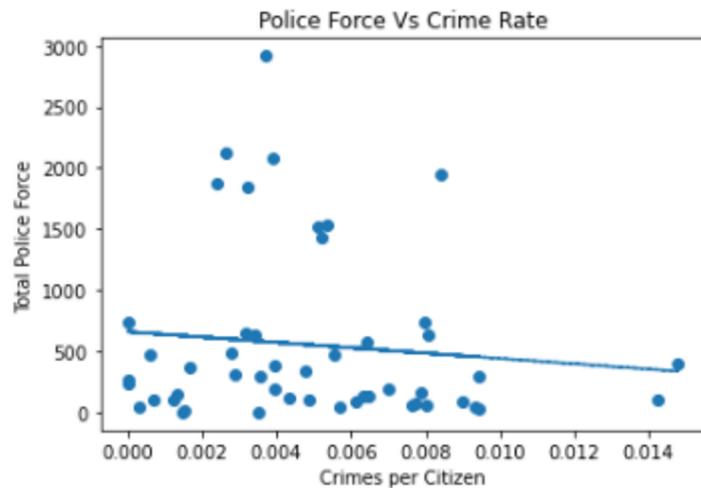
```
1 avg_income_ratings = file_df.iloc[:,13]
2 crime_rate = file_df.iloc[:,14]
3 ai_slope, ai_int, ai_r, ai_p, ai_std_err = stats.linregress(crime_rate, avg_income_ratings)
4 ai_fit = ai_slope * crime_rate + ai_int
5 plt.scatter(crime_rate,avg_income_ratings)
6 plt.plot(crime_rate,cr_fit,"--")
7 plt.ylabel('Avg Income')
8 plt.xlabel('Crimes per Citizen')
9 plt.title('Avg Income Vs Crime Rate')
10 plt.show()
```



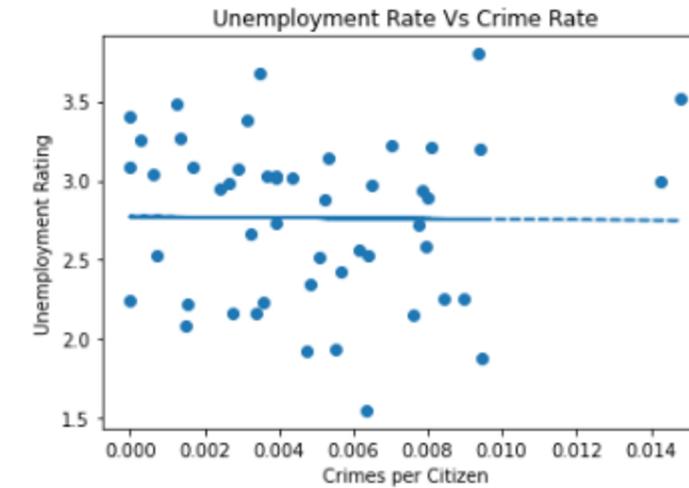
The line shows us the relationship between the two factors.

Linear Regression Plots Cont.

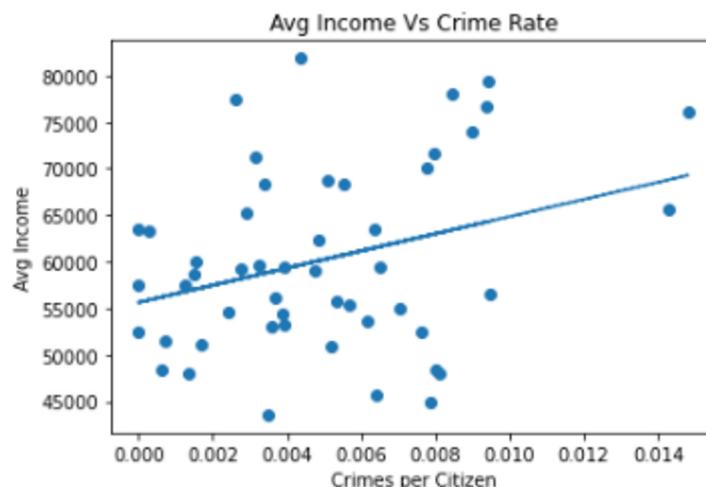
Number of Police Officers v Crime Rate



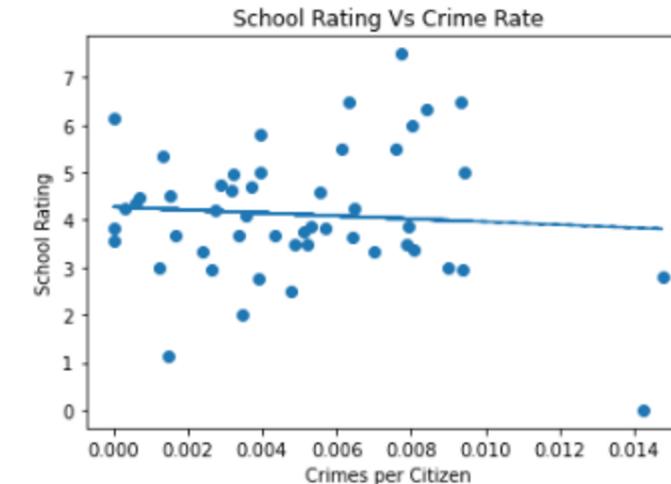
Unemployment Rate v Crime Rate



Avg Income Rate v Crime Rate



School Ratings v Crime Rate



R values & Correlations

R Value table for reference

Absolute Value of r	Strength of Correlation
$r < 0.3$	None or very weak
$0.3 \leq r < 0.5$	Weak
$0.5 \leq r < 0.7$	Moderate
$r \geq 0.7$	Strong

R Values for our factors

```
1 school_ratings = file_df.iloc[:,11]
2 unemployment_ratings = file_df.iloc[:,12]
3 avg_income_ratings = file_df.iloc[:,13]
4 total_police = file_df.iloc[:,5]
5 crime_rate = file_df.iloc[:,14]
6 correlation_sr = stats.pearsonr(school_ratings,crime_rate)
7 correlation_ur = stats.pearsonr(unemployment_ratings,crime_rate)
8 correlation_pf = stats.pearsonr(total_police,crime_rate)
9 correlation_ai = stats.pearsonr(avg_income_ratings,crime_rate)
10 print(f" The correlation between School Ratings & Crime Rate is {round(correlation_sr[0],2)}")
11 print(f" The correlation between Unemployment Rate & Crime Rate is {round(correlation_ur[0],2)}")
12 print(f" The correlation between Police Force & Crime Rate is {round(correlation_pf[0],2)}")
13 print(f" The correlation between Avg. Incomes & Crime Rate is {round(correlation_ai[0],2)}")
```

The correlation between School Ratings & Crime Rate is -0.08
The correlation between Unemployment Rate & Crime Rate is -0.01
The correlation between Police Force & Crime Rate is -0.11
The correlation between Avg. Incomes & Crime Rate is 0.32

Conclusion

- **What makes a city safe?**
 - How does number of police, average income, unemployment, and school rating affect crime rates?
- **Minimal correlation between test variables and crime rate.**
 - Fail to reject the null hypothesis
- **Improvements?**
 - Larger sample size with more crime variables
 - Find a full collection of police data from the city rather than one station
 - Using more relevant cities and ORIs

Adding laterCrime Rate (R,SA,SP) per Citizen

