

A REGRESSION ANALYSIS OF THE STRATEGIC SUBJECT LIST*

Ethan Sargent

October 30, 2019

*This work was not supported by any organization. Ethan Sargent graduated from Harvey Mudd College in 2019 with a B.S. in Mathematics. Contact: `esargent at hmc dot edu`

1 Statement of Intent

The Strategic Subject Algorithm (SSA) was developed at the Illinois Institute of Technology in 2013 and ranked hundreds of thousands of Chicagoans based on their estimated likelihood of involvement in a shooting, as either as a victim or a perpetrator [1].

After a FOIA request by the Chicago Sun-Times [2], the City of Chicago released the 2017 version of the Strategic Subject List (SSL), which is available at [3]. Each row of the list represents a subject. Fields include each subject’s SSL score, the 8 fields used as inputs to the SSA in computing this score, and 44 other fields including such sensitive information as race, gender, and geographical location that the algorithm does not incorporate. The details of the algorithm itself are not public, nor are the criteria for inclusion on the list - per [2], “127,513 individuals on the list have never been arrested or shot.” These criteria and the algorithm itself are periodically updated - only 102 of the 398,684 subjects on the most recent list have no recent arrest date.

The New York Times [1] and Upturn [2] conducted highly successful regression analyses of SSL score, although they did not publicize the details of their models. What follows is an additional regression analysis which is therefore not intended as original research, but is instead offered in the spirit of exposition on the Strategic Subject Algorithm. We begin with exploratory data analysis of the SSL and its 53 fields. Next, we present several accurate models of SSL score as a function of its 8 predictor variables. Finally, we attend to the obfuscated predictor variable **TREND IN CRIMINAL ACTIVITY**. As with SSL score, the mechanism by which this variable is computed is unknown, so we construct a model which predicts this variable.

In a Chicago Police Department directive [5], the SSL was announced to have been discontinued effective January 9th, 2019. The SSL and SSA were replaced by the Subject Assessment and Information Dashboard (SAID) and the Crime and Victimization Risk Model (CVRM), which model is described in detail at [4]. The CVRM takes as inputs 6 of the 8 inputs to the SSA, was also designed by IIT, and appears to have the same function of predicting subjects’ short-term likelihood of either committing or being a victim of a shooting. Unlike the SSA, the CVRM represents the subject list as an undirected graph, reflecting IIT’s understanding that “[i]n prior research on violent crime, [correlations] have been extensively demonstrated to exist among individuals who are close to one another in a graph defined by co-arrests.” To my knowledge no other information about the SAID or the CVRM is available to the public.

2 Table of Contents

Contents

1	Statement of Intent	2
2	Table of Contents	3
3	Exploratory Data Analysis	4
3.1	A word on the dataset	4
3.1.1	Missing values	4
3.1.2	Preprocessing	4
3.2	The distributions of SSL score and its 8 predictor variables	4
3.2.1	SSL scores: a summary	5
3.3	Exploring a model of PREDICTOR RAT TREND IN CRIMINAL ACTIVITY	6
3.3.1	The distribution of TREND	6
3.3.2	Comparing TREND to the SSA's unused fields	7
3.3.3	LATEST DATE vs. TREND	8
3.3.4	Arrest counts and TREND:	9
3.3.5	Recent arrest dates and TREND:	9
3.4	Age	10
3.4.1	The relationship between AGE CURR and PREDICTOR RAT AGE AT LATEST ARREST	11
3.5	Location	12
3.6	Race and gender	12
3.7	Gang affiliation	13
4	Modeling SSL score	14
4.1	Summary	14
4.2	Predictor Multicollinearity	14
4.3	Predictor Strengths	15
5	Modeling Trend	16
5.1	Preprocessing	16
5.1.1	0 substitution	16
5.1.2	Slope substitution	17
5.2	Summary	17

3 Exploratory Data Analysis

3.1 A word on the dataset

3.1.1 Missing values

The Strategic Subject List has 398,684 rows, each of which represents an individual on the list. The list has 53 fields: 8 predictor fields, which the CPD claims are the only fields used in computing the SSL Score, 44 other fields including variables like race, sex, and geographical location, and the SSL Score itself. Among the 8 predictor columns and the SSL Score column, there are 102 rows with missing values. Since the dataset is large, we elect to drop these rows rather than impute or interpolate these missing values.

3.1.2 Preprocessing

There are two predictor variables that need to be interpreted before they are useful as inputs to a regression analysis. First, the field `PREDICTOR RAT AGE AT LATEST ARREST` is given as a string: ‘less than 20’ if the individual was younger than 20 at the time of their most recent arrest, and otherwise as a decade-long range; for instance, ‘20-30’ if the subject was between 20 and 30 at the time of their most recent arrest. These obfuscations are presumably to preserve anonymity. To convert these values to numerical inputs that preserve distance, we convert ‘less than 20’ observations to the value 0.0, ‘20-30’ observations to the value 1.0, ‘30-40’ observations to the value 2.0, and so on.

Second, the field `PREDICTOR RAT GANG AFFILIATION` is a binary categorical variable, with ‘1’ indicating an outstanding gang affiliation. We use this as a dummy variable in the regression analysis, treating it like any other continuous input to the model. The coefficient on the dummy variable in a linear regression analysis then becomes a differential intercept coefficient. That is to say, our predictions for gang affiliated individuals differ from our predictions for non-gang affiliated individuals only by a constant value fitted by the model. Per [2], gang affiliation is no longer used as an input to the model and was never a particularly effective predictor, however, the most recent publicly available SSL has scores computed using gang affiliation, so we include it in our model.

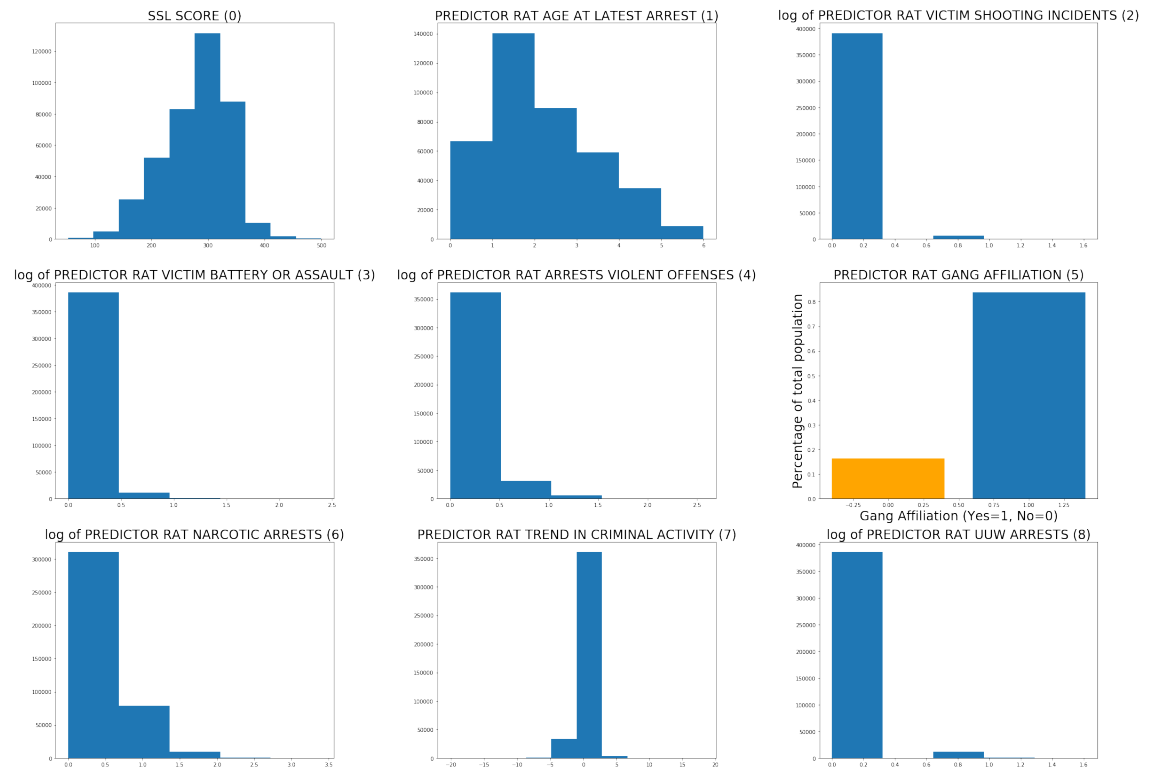
3.2 The distributions of SSL score and its 8 predictor variables

SSL scores are computed using an algorithm which takes the following eight fields as inputs

1. `PREDICTOR RAT AGE AT LATEST ARREST`
2. `PREDICTOR RAT VICTIM SHOOTING INCIDENTS`
3. `PREDICTOR RAT VICTIM BATTERY OR ASSAULT`

4. PREDICTOR RAT ARRESTS VIOLENT OFFENSES
5. PREDICTOR RAT GANG AFFILIATION
6. PREDICTOR RAT NARCOTIC ARRESTS
7. PREDICTOR RAT TREND IN CRIMINAL ACTIVITY
8. PREDICTOR RAT UUV ARRESTS

The distributions of SSL Scores and the above variables are shown below. Predictors 2-4, 6, and 8 are heavily skewed right. Visibility is improved a little by putting them on a log-scale. Normalizing predictor 7 is unhelpful due to the existence of extreme outliers.



3.2.1 SSL scores: a summary

SSL Scores have minimum 53, maximum 500, mean 279.90, and standard deviation 57.88. They fail a normality test. 153 individuals have the maximum score of 500, while only 132 have scores between 480 and 499, suggesting some degree of bunching at the right tail. 95% of individuals have an SSL score between 157 and 372.

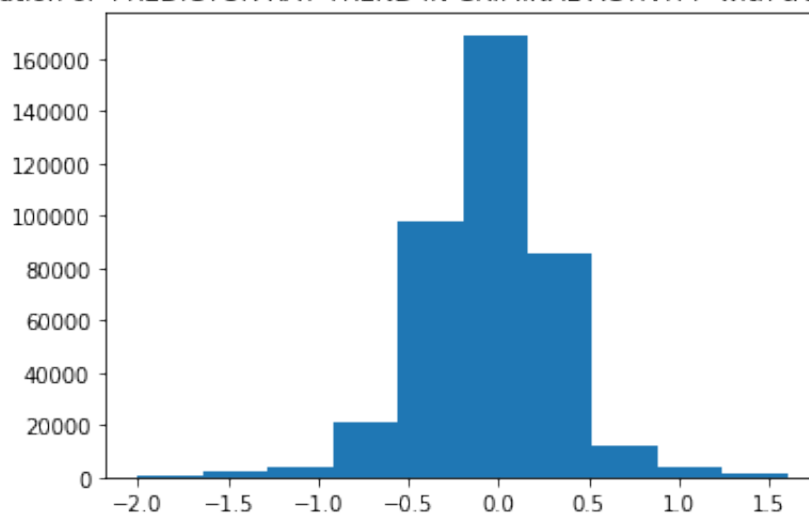
3.3 Exploring a model of PREDICTOR RAT TREND IN CRIMINAL ACTIVITY

The way PREDICTOR RAT TREND IN CRIMINAL ACTIVITY (hereafter, **TREND**) is computed is unknown. A regression analysis of **TREND** will eventually factor in to our analysis. For now, we examine its distribution and remark on its relationships to other fields in the list.

3.3.1 The distribution of **TREND**

The distribution fails a normality test. 95% of the data lies between -0.9 and 0.7 , and 99% of the data between -1.6 and 1.3 . However, the distribution has many extreme outliers. The maximum and minimum are both approximately twenty standard deviations from the mean, and there are 238 values 8 or more standard deviations from the mean. A histogram of **TREND** is a little more revealing with 1000 left and right outliers truncated.

Distribution of 'PREDICTOR RAT TREND IN CRIMINAL ACTIVITY' with truncated outliers



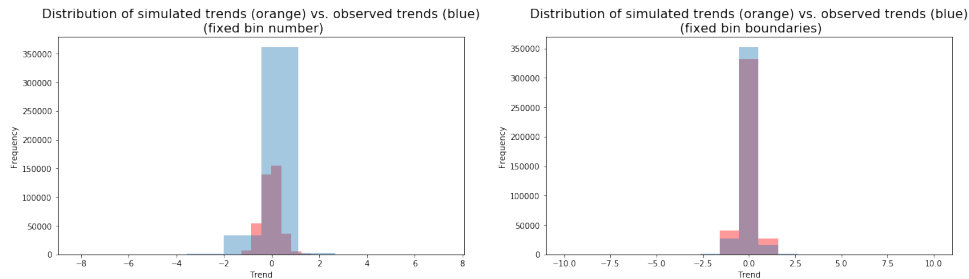
An upward spike in criminal activity (the highest 100 of observed **TREND** values) almost always is associated with a **LATEST DATE** of 2016. The low end **TREND** has a mean **LATEST DATE** of 2014.76. In contrast, the list as a whole has a mean **LATEST DATE** of 2014.20. One possible explanation: an active criminal who has not committed a crime recently ought to have a lower trend than someone who has simply not committed a crime recently.

We use the **fitter** module to investigate possible distributions. The **fitter** module compares the observed distribution to **scipy**'s 80 distributions - for each of **scipy**'s continuous distributions, the `.fit` function returns a maximum likelihood estimate for the parameters of the distribution given that the observed data is of that distribution. Then, **fitter** finds the distribution that minimizes the squared-error discrepancy between the observed and

predicted data.

The `exponnorm` distribution is perhaps the most intuitive of these fits since it has a low squared error and an interpretation as the sum of normal and exponential random variables, which variables are possibly inputs to the calculation of `TREND`.

Below we compare the distribution of simulated, fitted `exponnorm` values to that of observed `TREND` values. As evidenced by the plots, the `exponnorm` fails to replicate the heavy tails of the `TREND` values. While imperceptible in the left histogram, the outliers of `TREND` skew the bin width much higher than in the histogram of the simulated values. We correct for this in the right histogram - nevertheless, `exponnorm` is a poor choice of distribution.



3.3.2 Comparing `TREND` to the SSA's unused fields

In addition to the seven `PREDICTOR` fields, our analysis will require interpretation of the remaining 44 fields. We are specifically interested in whether there is evidence that `PREDICTOR RAT TREND IN CRIMINAL ACTIVITY` is derived from these fields, which the CPD claims are not inputs to the model. In addition to the predictor fields, the following fields warrant investigation:

1. `LATEST DATE`: (latest date of police contact)
2. `WEAPONS ARR CNT` (counts are computed from arrests in last 10 years)
3. `LATEST WEAPON ARR DATE`
4. `NARCOTICS ARR CNT`
5. `LATEST NARCOTIC ARR DATE`
6. `DOMESTIC ARR CNT`
7. `LATEST DOMESTIC ARR DATE`
8. `SSL LAST PTV DATE` (from the third footnote: "Most recent date that the subject was matched with a victim or arrest record making the subject a 'Party to Violence'")

Roughly 75.8% of the entries in these fields are `NaN`. This makes sense - most people will not have, for instance a `LATEST DOMESTIC ARR DATE`, since most people have not been arrested

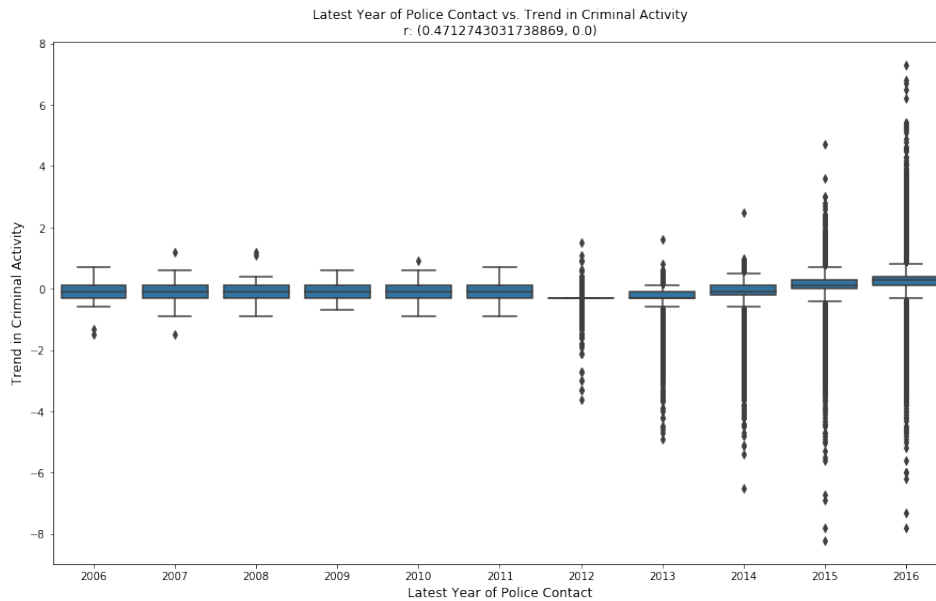
for domestic violence. These are therefore not “missing values” in the traditional sense, and so it would be inappropriate to simply drop them from the model. Dropping every row with an NaN value in any of the above fields would leave only individuals with a large, diverse criminal record - this sampling bias would corrupt the resulting analysis. Our representation of these values will ultimately depend on model choice.

3.3.3 LATEST DATE vs. TREND

The below box plot indicates that the vast majority of subjects have a **TREND** near 0, regards of the year of their **LATEST DATE** of police contact. There is a slight positive correlation between 'trend' and **LATEST DATE** of .4712, with a p value of 0.

Yet it is possible that not all police contact is bad for one's 'trend'. In the year 2016, for example, many subjects have an incredibly low 'trend' value that coincides with a recent date of police contact. It is unknown what qualifies as police contact: an arrest might warrant an increase in 'trend', whereas an interview leading the CPD to conclude a subject is no longer gang affiliated, for instance, might warrant a decrease in 'trend'. Per [2]:

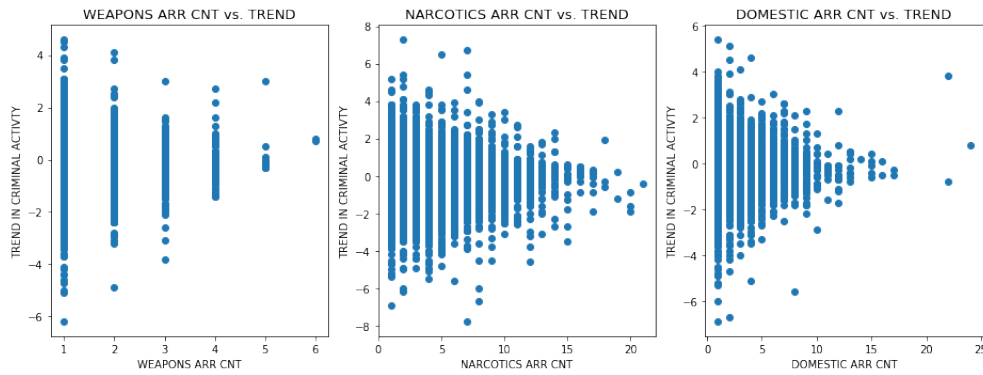
The CPD claims the SSL is used in conjunction another program, called Custom Notifications (Special Order S10-05), where police officers, social workers, and community leaders “deliver a joint message ... informing [people on the list] of their risk for prosecution based on criminal history, and explaining their opportunities for community help and support.” The CPD itself describes Custom Notifications as “a process that identifies potential criminal actors and victims associated with the continuum of violence. Once identified, the individual is notified of the consequences that will result should violent activity continue.” Between 2013 and 2016, the CPD is reported to have made over 1,400 of these visits.



The boxplot for the year 2013 is not an error - roughly 29,000 of the 30,000 subjects for that year have `TREND` value $-.3$, which encompasses the second and third quartiles. This anomaly in `TREND` computation will doubtless affect our model.

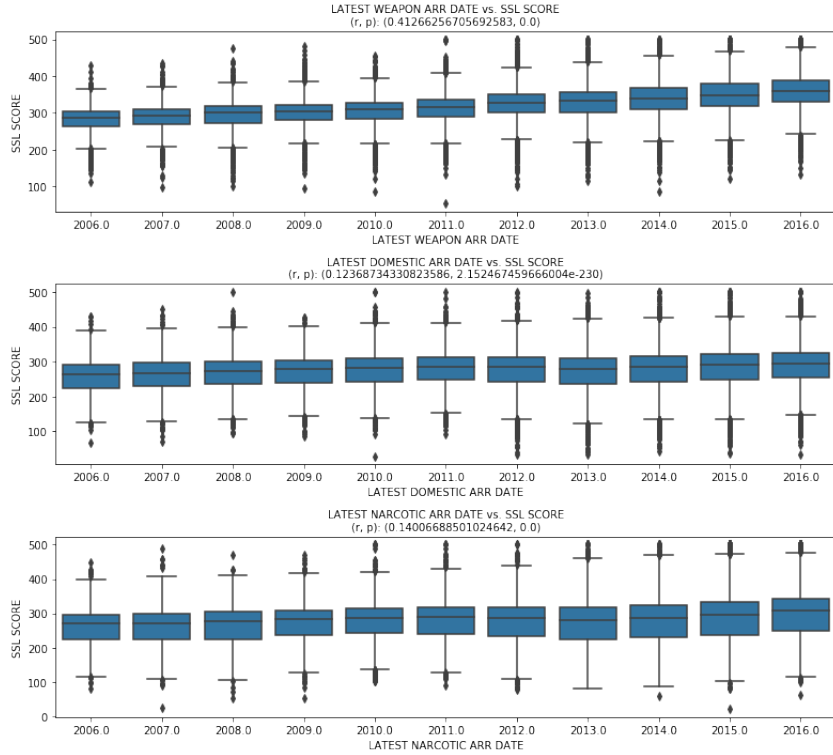
3.3.4 Arrest counts and `TREND`:

Below are three scatterplots. They depict one of three types of arrest count (weapons, narcotics, domestic) vs. `TREND`. These plots reveal they are unlikely to be helpful in any model. We note that the distribution of `WEAPONS ARR CNT` is much tighter than those of `NARCOTICS ARR CNT` and `DOMESTIC ARR CNT`. Illinois has relatively strict gun control laws; per a Chicago criminal defense lawyer, possession of a loaded gun with no FOID/CCL card carries a penalty of 1-3 years in prison with steeper penalties for repeat offenders and other aggravating factors.[7] It is likely that subjects don't accumulate many weapons charges due to the high probability of lengthy incarceration that accompanies each arrest.



3.3.5 Recent arrest dates and `TREND`:

Here we come across more related variables. `LATEST WEAPON ARR DATE` has the strongest correlation we have seen yet with `TREND`: .41. `LATEST DOMESTIC ARR DATE` and `LATEST NARCOTICS ARR DATE` have correlations of .12 and .14, respectively. In the boxplots below the correlations are visible in the slight positive trends of the means.



3.4 Age

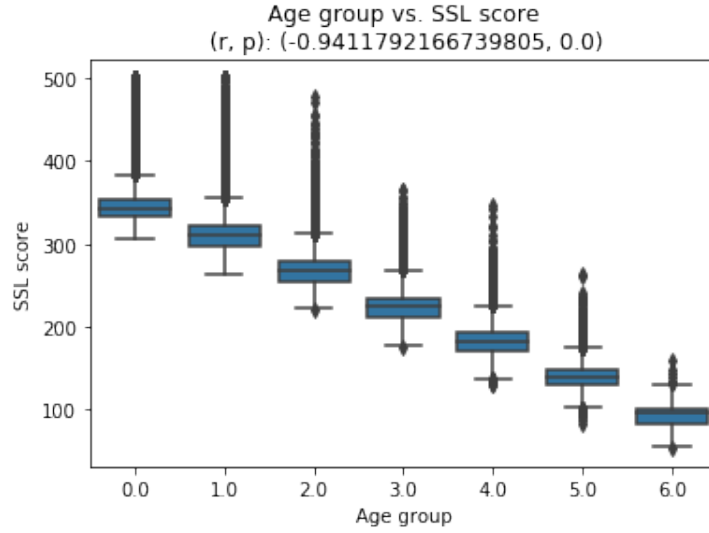
Both Upturn and the New York Times found that the variable most closely related to SSL score was age. Concretely, [1] found that

The most significant characteristic for computing an S.S.L. risk score is the age of a potential victim or offender. For every decade of age, the risk score declined by about 40 points. Practically speaking, this variable limits the list to young people: No one older than 30 falls within the highest-risk category with a score at or above 480.

And similarly, [2] found that

age accounts for roughly 89% of variance in SSL scores. In other words, the score mostly just reflects each person's age.

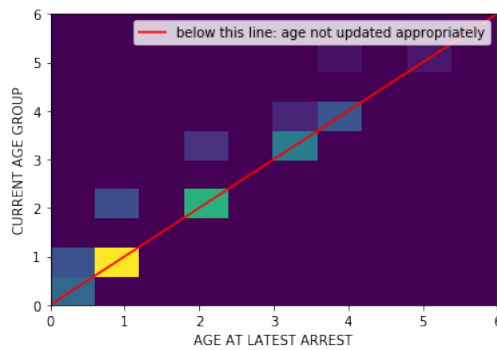
So, age is a crucial ingredient for modeling SSL score. Below we have a scatter plot of age group vs. SSL score. We replicate Upturn's finding that age accounts for $.9412^2 \sim 89\%$ of the variance in SSL scores, although we are unsure whether they used the same age-group-to-continuous-variable conversion. Recall that ours was 'less than 20' \rightarrow 0, '20-30' \rightarrow 1, '30-40' \rightarrow 2 and so on.



There are two other fields on the list pertaining to age. `AGE TO` (per [3]: “If Age is an estimated range, this is the upper end of that range”) and `AGE GROUP` (per [3]: “Subject’s Age as of their latest arrest record”). In any case, these fields contain which exactly the same values. Since `AGE CURR` values are higher or equal, we assume these fields were deprecated and will use `AGE CURR` exclusively.

3.4.1 The relationship between `AGE CURR` and `PREDICTOR RAT AGE AT LATEST ARREST`

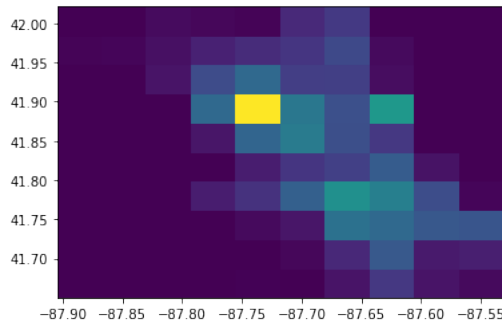
About 77% of subjects are, as of May 1, 2017 when the data was last updated, the same age as when they were last arrested. Interestingly, 924 subjects are listed to have a current age less than their age at the time of their latest arrest. Since this is impossible, it is possible when a subject is arrested, their current age is not updated. The 2D histogram below indicates that there is a high density of subjects along the line `AGE CURR = PREDICTOR RAT AGE AT LATEST ARREST`, with over a quarter of subjects ((28.6%, the yellow square) both between 20 and 30 and with a most recent arrest date while between 20 and 30.



3.5 Location

Latitude and longitude data are provided for location of the subject’s most recent arrest, but only for 224,235 subjects, or 56.24% of the total. The situation is not much improved for census tract and community area fields, which are non-null for just 227,763 and 224,311 rows, respectively. Since 398,582 subjects have a non-null entry for **PREDICTOR RAT AGE AT LATEST ARREST**, we assume at least 398,582 subjects have been arrested, and therefore these location fields are woefully incomplete.

Below is a heatmap of of arrests based on the available latitude and longitude data. The one yellow square is adjacent to West and East Garfield Park - per the Chicago Tribune, these are two of the highest-crime neighborhoods in Chicago.[6]

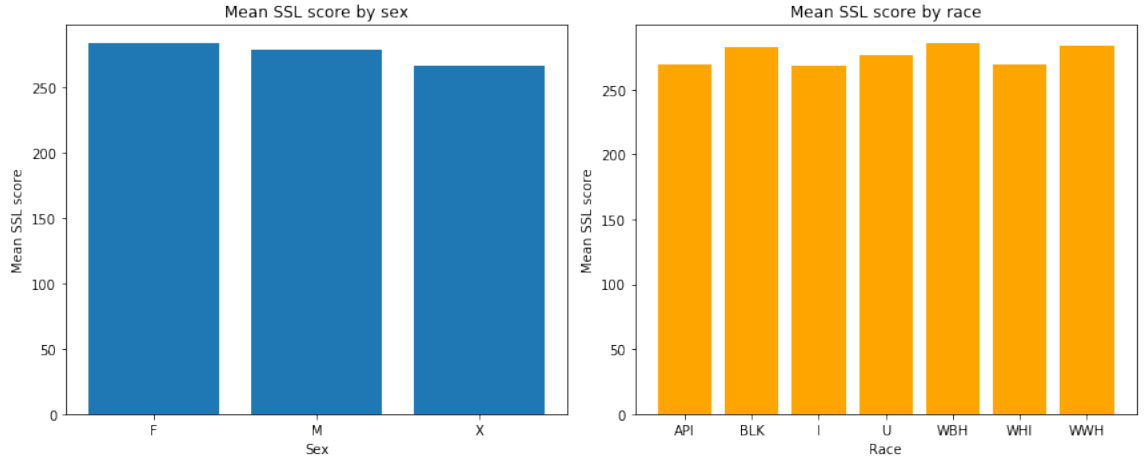


3.6 Race and gender

Per the CPD, the algorithm does not take this sensitive information as input. Some analysis is recorded here for posterity, although it does not suggest vast differences in SSL score across race and sex. Nevertheless, we conduct two F -tests below which lead us to reject the null hypotheses that SSL scores do not depend on race or sex. Since SSL scores fail a normality test, we do not observe the normality assumption required for the F -test, which luckily is robust to violations of this assumption especially when sample size is large.

The field **SEX CODE CD** has value **M** if the subject is male, **F** if the subject is female, and **X** in 57 cases. Presumably in these latter cases either the subject’s sex was not recorded or the subject chose not to identify as male or female.

Per [3], the field **RACE CODE CD** uses the encoding “BLK - Black; WHI - White; API - Asian/Pacific Islander; WBH - Black Hispanic; WWH - White Hispanic; I - American Indian/Alaskan Native; U - Unknown.” There are 1899 individuals for whom race is unknown. Bar charts of the mean SSL score by race are shown below.



The differences are slight; nevertheless, race and sex have statistically significant impacts on SSL score. One-way F-tests lead us to reject the null hypotheses that SSL score means are the same across these groups. Somewhat surprisingly, females have a higher mean SSL score (283) than males (278). Racial differences in SSL score are more pronounced. Black Hispanics have the highest mean SSL score of 285, whereas Native Americans have the lowest: 268. A table summarizing these results is shown below.

Race	Mean SSL Score
API	269.433831
BLK	282.426208
I	268.340580
U	276.681938
WBH	285.431104
WHI	269.723842
WWH	283.267756

3.7 Gang affiliation

Gang affiliation is associated with roughly a 10% increase in SSL SCORE - unaffiliated subjects ($n=333737$) have a mean SSL score of 275.170092, whereas gang affiliated subjects ($n=64947$) have a mean SSL score of 303.835235. After an F -test ($p = 0.0$) we reject the null hypothesis that the difference is statistically insignificant.

4 Modeling SSL score

Except for ordinary least squares (OLS), data was randomly divided into training (80% of data) and test (20% of data) sets, and final performance was measured as root-mean-squared error on the test set. A summary of various modeling efforts is below.

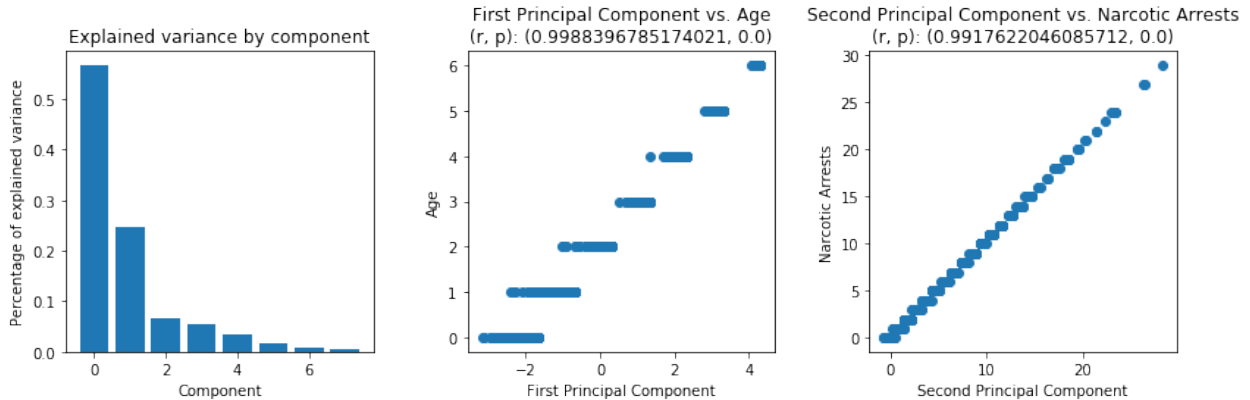
4.1 Summary

As far as we know, [1] and [2] only employed linear regression in their analyses of SSL score, and with good reason: even ordinary least squares is quite effective at predicting SSL score, and more complicated models do not significantly outperform linear regression. So we devote most of our analysis to the results of the linear model, which results are most easily interpreted in context.

Model	RMSE	Cross-validated (n folds= 5)
OLS (age only)	19.55	No
OLS	12.97	No
OLS w/ polynomial features (deg ≤ 3)	12.48	No
Random Forests	12.47	Yes
XGBoost	12.38	No

4.2 Predictor Multicollinearity

A principal component analysis reveals high multicollinearity among the predictors. Furthermore, the first principal component is essentially age, and the second principal component is essentially narcotic arrests. It seems that the true dimensionality of the data is much lower than 8.

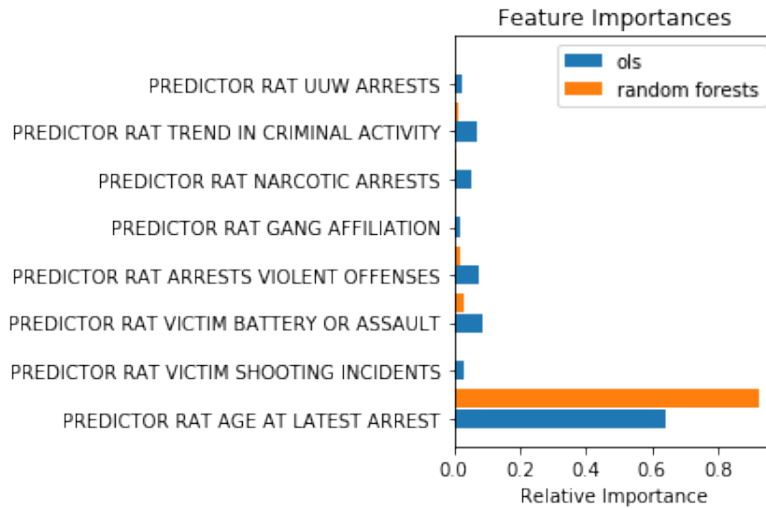


In the public description of the CVRM, we find the remark that “inclusion of narcotics arrests and gang affiliation [has] only marginal impact on the results beyond what the model discerns from the other risk factors; therefore, in the current CVRM version, these two risk

factors have been omitted.” [4] If age and narcotic arrests capture most of the variance in the predictors as indicated by the figure above, the omission of narcotic arrests in the CVRM may reflect an understanding that narcotic arrests is approximately a weighted combination of other inputs.

4.3 Predictor Strengths

Recall [2] found that age accounts for 89% of the variance in SSL score - unsurprisingly, age at latest arrest is by far the most important predictor across all our models. We can assess the relative importance of the predictors for our models concretely: a multiple linear regression on standardized predictors yields coefficients that are just the relative strengths of the predictors. Furthermore, we can assess a predictor’s importance to a random forests regression by its total variance reduction across all nodes that split on that feature. Concretely, we have the below barplots.



In the linear model, age is the strongest predictor (63.91% relative importance), while being a victim of a battery or assault is the second strongest predictor (7.07%), followed by arrests for violent offenses (6.04%) and trend in criminal activity (5.75%). In a random forests regression, age is even more dominant at 92.67% relative importance - the next highest predictor, being a victim of a battery or assault, only has 2.80% relative importance.

5 Modeling Trend

Overall, our model of **TREND** was less successful. There is good reason to believe this is due to lack of data. In [4], which describes the CVRM, an algorithm which has replaced the SSA but takes many of the same inputs including **TREND**, the authors write that “the ‘trend’ variable is the slope of a line obtained by a least-squares fit to the individual’s numbers of arrests each year for the past four years.” If the trend variable is computed in the same way for the SSL, then the necessary data to compute it is absent from the SSL; in particular, the SSL only contains the year of a subject’s most recent arrest, by arrest type, and their arrest count, by arrest type.

Our pessimism assumes that the trend variables are computed in the same way for the SSL, which may not be the case - if trend is computed strictly from arrests as with the CVRM, it is difficult to see why, for instance, according to the SSL, a man over 60 with no arrests on record has a **TREND** of 1.2, which puts him in the 99th percentile for that variable. Yet we saw in (3.4) that the SSL has non-predictor fields that are internally inconsistent for specific subjects - it is certainly possible that arrest data is not updated properly in the non-predictive fields.

In sum, the evidence of missing data and the unreliability of the non-predictive entries of the SSL mean that we do not devote as much time and analysis to a model of **TREND**.

5.1 Preprocessing

Recall that our model of **trend** used 8 fields in addition to the 8 predictor fields of SSL score, and that 75.8% of the entries in these additional fields were **NaN**.

If an arrest count field, for instance **NARCOTIC ARR CNT**, is **NaN**, we assume that the subject has does not have a narcotics arrest on record, and we substitute 0. However, interpreting a time-sensitive fields like **LATEST DATE** of police contact requires more subtlety. We employed two different strategies for representing these fields as inputs to regression models.

5.1.1 0 substitution

We want to preserve distance between values like 2015 and 2016, while also ensuring that a subject’s **TREND** is not affected if they have no **LATEST DATE** - a subject with no police contact is not becoming more or less criminally active over time. Substituting 0 for **NaN** values satisfies the second criterion but not the first - a linear regression fitted to points 0 and 2015 will not be able to adequately distinguish between 2015 and 2016. This substitution may be adequate for a nonlinear model like random forests or **xgboost**.

5.1.2 Slope substitution

In the spirit of the definition of **TREND** in [4], we re-express a subject’s most recent arrest date as the slope of the best fit line through the one-hot encoding for that year, where the encoding spans the years 2012-2016, the four years considered by the model described in [4]. For example, if a subject was arrested in 2015, we substitute the slope of the best-fit line through

$$x = \{2012, 2013, 2014, 2015, 2016\}, y = \{0, 0, 0, 1, 0\}.$$

For subjects with a **LATEST DATE** earlier than 2012 ($n=5556$), say, y , we compute the best-fit slope through

$$x = \{y, y + 1, \dots, 2011, \dots, 2016\}, y = \{1, 0, \dots, 0\}.$$

This substitution has many desirable properties - first, the contribution to **TREND** of a subject without a recent arrest year is 0, that is, the slope of the best fit line through the zero vector. Second, recent arrest years have positive slopes which are higher the more recent the year of arrest. Finally, slope tends toward 0 as arrest year decreases. A subject’s total **TREND** will be a weighted combination of the subject’s slope for each time-sensitive input - **LATEST DATE**, **LATEST NARCOTIC ARR DATE**, **LATEST DOMESTIC ARR DATE**, **LATEST WEAPON ARR DATE**, and **SSL LAST PTV DATE** - together with the other inputs.

5.2 Summary

Below is a summary of various models of **TREND**. The random model computes the root-mean-squared error of the difference between the **TREND** labels and a bootstrap replicate of the **TREND** labels. A cross-validated random forests regression without a slope substitution is the best performing model.

Model	Substitution	RMSE	Cross-validated (nfolds = 5)
Random model	n/a	0.5726	No
OLS	0	0.32786	No
OLS	slope	0.3135	No
OLS w/ polynomial features	0	0.2895	No
OLS w/ polynomial features	slope	0.2843	No
Random forests	0	0.2832	Yes
Random forests	slope	0.2848	Yes

References

- [1] J. Asher and R. Arthur, “Inside the Algorithm That Tries to Predict Gun Violence in Chicago.” New York Times, 13 June 2017; <https://www.nytimes.com/2017/06/13/upshot/what-an-algorithm-reveals-about-life-on-chicagos-high-risk-list.html>
- [2] B. Posadas, “How strategic is Chicago’s ‘Strategic Subjects List’? Upturn investigates.” Medium, 22 June 2017; <https://medium.com/equal-future/how-strategic-is-chicagos-strategic-subjects-list-upturn-investigates-9e5b4b235a7c>
- [3] Chicago Data Portal, 7 December 2017; <https://data.cityofchicago.org/Public-Safety/Strategic-Subject-List/4aki-r3np>
- [4] “Crime and Victimization Risk Model (CVRM).” Illinois Insitute of Technology, n.d.; <https://home.chicagopolice.org/wp-content/uploads/2019/01/FACT-SHEET-Crime-and-Victimization-Risk-Model-1.pdf>
- [5] “SUBJECT ASSESSMENT AND INFORMATION DASHBOARD (SAID).” Chicago Police Department, 09 January 2019. <http://directives.chicagopolice.org/directives/data/a7a57b85-155e9f4b-50c15-5e9f-7742e3ac8b0ab2d3.html?hl=true>
- [6] “Crime in Chicago: Explore your community.” Chicago Tribune, 01 April 2019. <https://www.chicagotribune.com/news/ct-crime-in-chicago-20171114-storygallery.html>
- [7] “Gun Charge in Illinois FAQs.” Robert J. Callahan: Chicago Criminal Defense Attorney, n.d.; <https://www.defenselawyersite.com/gun-charge-in-illinois-faqs/>
- [8] M. Dumke and F. Main, “A look inside the watch list Chicago police fought to keep secret.” Chicago Sun Times, 18 May 2017. <https://chicago.suntimes.com/2017/5/18/18386116/a-look-inside-the-watch-list-chicago-police-fought-to-keep-secret>