

# A regression analysis of the Strategic Subject Algorithm

E. Sargent, esargent at hmc dot edu

November, 2019

## 1 Introduction

- History of the SSA
- Previous reporting on the SSA

## 2 Exploratory Data Analysis

- The SSL
- Predictors and their distributions
- The predictor TREND IN CRIMINAL ACTIVITY
- Location, race, and gender

## 3 Models

- Modeling SSL score
- Modeling TREND

## 1 Introduction

- History of the SSA
- Previous reporting on the SSA

## 2 Exploratory Data Analysis

- The SSL
- Predictors and their distributions
- The predictor TREND IN CRIMINAL ACTIVITY
- Location, race, and gender

## 3 Models

- Modeling SSL score
- Modeling TREND

# History of the SSA

- Developed at the Illinois Institute of Technology in 2013, in cooperation with the Chicago Police Department.
- 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].
- Discontinued in 2019 in favor of the Crime Reduction Victimization Model (CVRM).

## 1 Introduction

- History of the SSA
- Previous reporting on the SSA

## 2 Exploratory Data Analysis

- The SSL
- Predictors and their distributions
- The predictor TREND IN CRIMINAL ACTIVITY
- Location, race, and gender

## 3 Models

- Modeling SSL score
- Modeling TREND

# Previous reporting on the SSA

- Chicago Sun Times, 2017
  - FOIA request, legal battle, release of 2017 version of Strategic Subject List (SSL)
- New York Times, 2017
  - linear regression on SSL scores
  - score mostly ( $R^2 = 89\%$ ) reflects a person's age
- Upturn, 2017
  - linear regression on SSL scores
  - discussion of the customs notification program

## 1 Introduction

- History of the SSA
- Previous reporting on the SSA

## 2 Exploratory Data Analysis

- The SSL
- Predictors and their distributions
- The predictor TREND IN CRIMINAL ACTIVITY
- Location, race, and gender

## 3 Models

- Modeling SSL score
- Modeling TREND

# The SSL

- Most recent version: Dec. 7, 2017
- 398,684 subjects
- 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 SSL Score itself



## 1 Introduction

- History of the SSA
- Previous reporting on the SSA

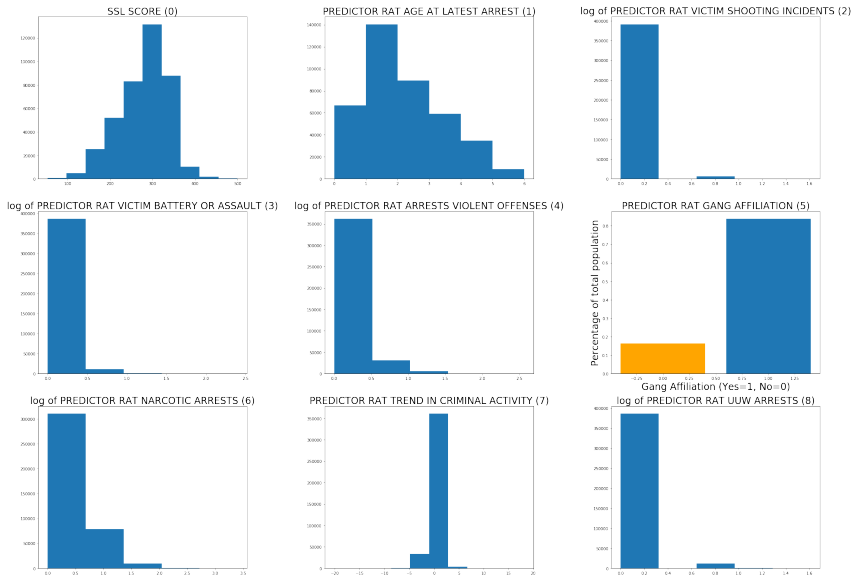
## 2 Exploratory Data Analysis

- The SSL
- Predictors and their distributions
- The predictor TREND IN CRIMINAL ACTIVITY
- Location, race, and gender

## 3 Models

- Modeling SSL score
- Modeling TREND

# Predictors and their distributions



## 1 Introduction

- History of the SSA
- Previous reporting on the SSA

## 2 Exploratory Data Analysis

- The SSL
- Predictors and their distributions
- **The predictor TREND IN CRIMINAL ACTIVITY**
- Location, race, and gender

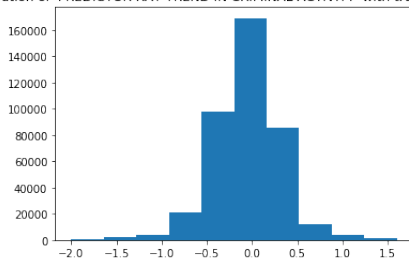
## 3 Models

- Modeling SSL score
- Modeling TREND

# Predictors and their distributions

- Unknown distribution (many outliers)
- Possibly computed from missing data

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



## 1 Introduction

- History of the SSA
- Previous reporting on the SSA

## 2 Exploratory Data Analysis

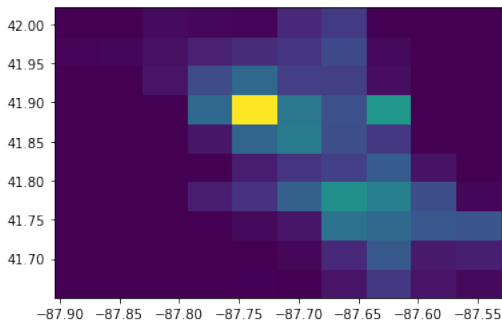
- The SSL
- Predictors and their distributions
- The predictor TREND IN CRIMINAL ACTIVITY
- Location, race, and gender

## 3 Models

- Modeling SSL score
- Modeling TREND

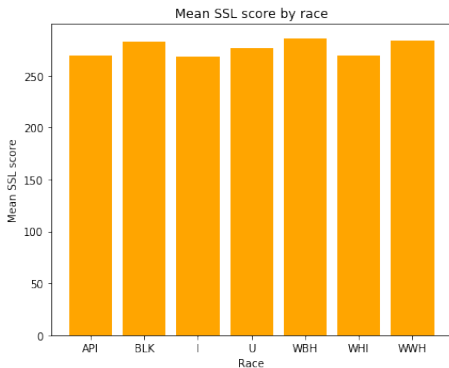
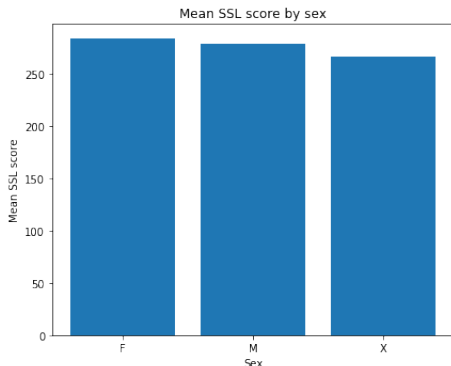
# Location

- lat/long of most recent arrest for only 224,235 subjects, or 56.24% of the total.



# Sex and race

- Between-group differences are slight, but statistically significant



## 1 Introduction

- History of the SSA
- Previous reporting on the SSA

## 2 Exploratory Data Analysis

- The SSL
- Predictors and their distributions
- The predictor TREND IN CRIMINAL ACTIVITY
- Location, race, and gender

## 3 Models

- Modeling SSL score
- Modeling TREND



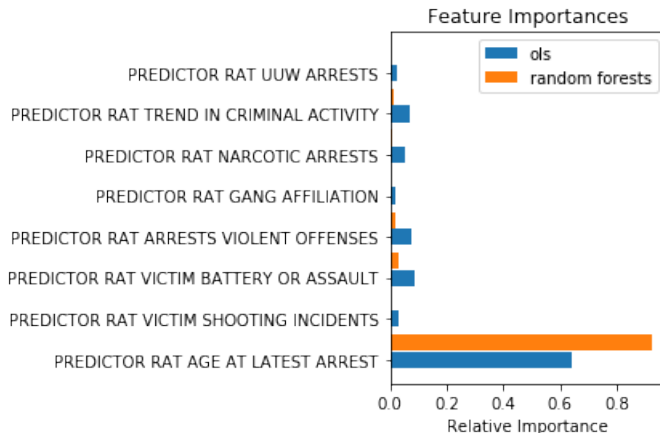
# Modeling SSL score

- best-performing model: xgboost

Model	RMSE	Cross-validated
OLS (age only)	19.55	No
OLS	12.97	No
OLS w/ polynomial features ( $\text{deg} \leq 3$ )	12.48	No
Random Forests	12.47	Yes
XGBoost	12.38	No

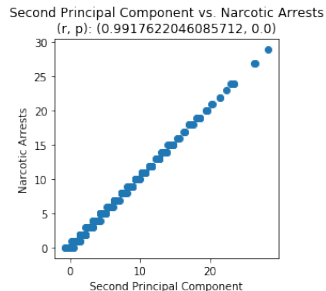
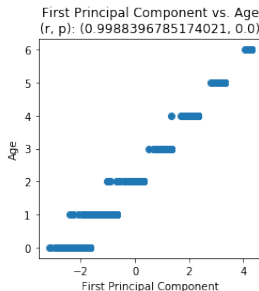
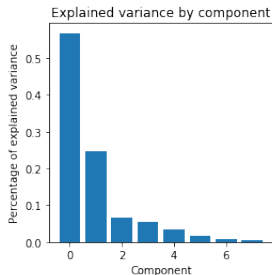
# Modeling SSL score (cont.)

- Age is by far the most important feature



# Modeling SSL score (cont.)

- Narcotic arrests, the second principal component of a principal component decomposition of the predictors, is not an input to the CVRM.



## 1 Introduction

- History of the SSA
- Previous reporting on the SSA

## 2 Exploratory Data Analysis

- The SSL
- Predictors and their distributions
- The predictor TREND IN CRIMINAL ACTIVITY
- Location, race, and gender

## 3 Models

- Modeling SSL score
- Modeling TREND

# Modeling TREND

- Per [4], "the 'trend' variable is the slope of a line obtained by a least-squares fit to the individuals numbers of arrests each year for the past four years."
- Data to compute this is missing
- Two approximations with our incomplete data:
  - 0 substitution
  - slope substitution

# Modeling TREND (cont.)

Model	Substitution	RMSE	Cross-validated
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

# For Further Reading I



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>



B. Posadas, "How strategic is Chicagos 'Strategic Subjects List' ? Upturn investigates." Medium, 22 June 2017;  
<https://medium.com/equal-future/how-strategic-is-chicagos-strategic-subjects-list-upturn-investigates-9e5b4b235a7c>



Chicago Data Portal, 7 December 2017;  
<https://data.cityofchicago.org/Public-Safety/Strategic-Subject-List/4aki-r3np>

# For Further Reading II



"Crime and Victimization Risk Model (CVRM)." Illinois Institute of Technology, n.d.;  
<https://home.chicagopolice.org/wp-content/uploads/2019/01/FACT-SHEET-Crime-and-Victimization-Risk-Model-1.pdf>



"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>



"Crime in Chicago: Explore your community." Chicago Tribune, 01 April 2019. <https://www.chicagotribune.com/news/ct-crime-in-chicago-20171114-storygallery.html>



"Gun Charge in Illinois FAQs." Robert J. Callahan: Chicago Criminal Defense Attorney, n.d.;  
<https://www.defenselawyersite.com/gun-charge-in-illinois-faqs/>



# For Further Reading III



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>