Determinants of Spatio-Temporal Patterns of Cybercrimes in the USA: Implications for Cybersecurity Personnel Resource Allocation

Stefano Chiaradonna

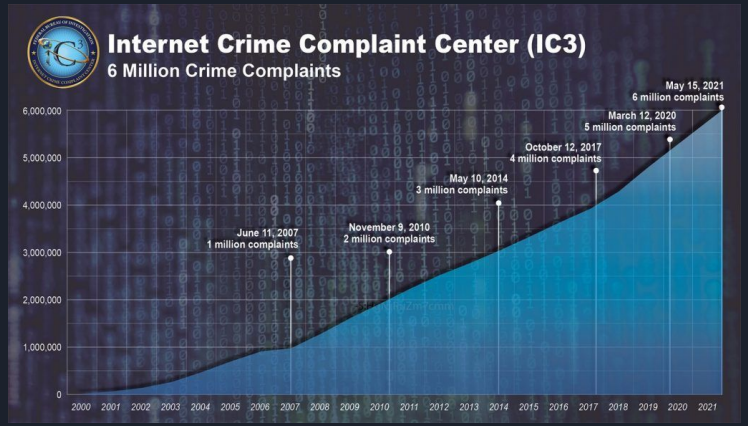
10 March 2025

“The U.S. Is Less Prepared to Fight Cybercrime Than It Could Be” – U.S. Government Accountability Office 2023.

Problem – Cybercrime agents are continuously overwhelmed with limited resources.

- Increasing cybersecurity workforce shortage: 4 million needed positions [1].
- FBI personnel is outnumbered 50 to 1 [2].
- Static Assignment: only 12 personnel per field office [3].

Goal – Determine where to effectuate cybersecurity personnel by state and field office.



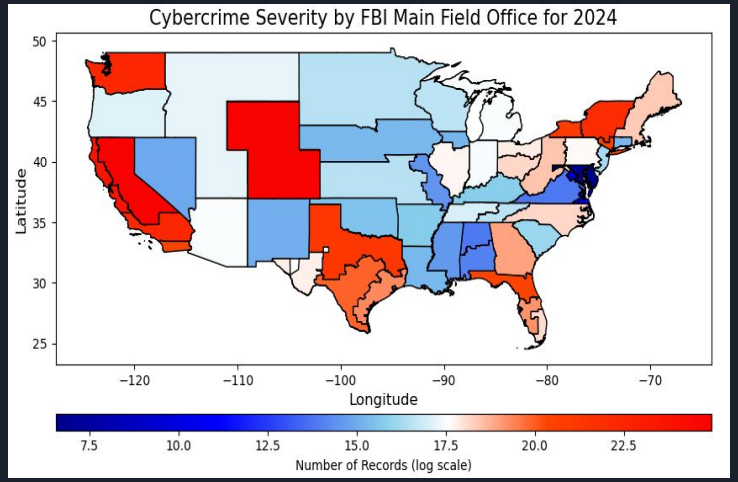
Benefit to DHS – *Improved cybersecurity law enforcement resource allocation and response.*

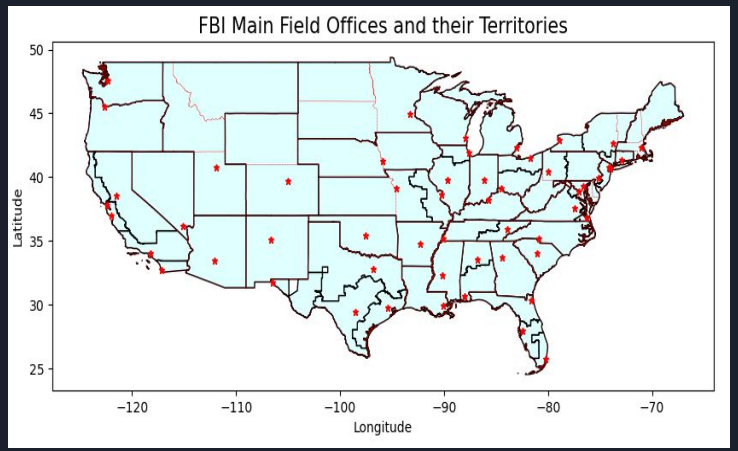
- *Help reduce burnout and increase workforce retention efforts.*

Key Insights

- Cybercrime agents in all field offices will face an overwhelming surge
  - Particularly in California, Texas, and Colorado

- **Current policy** of uniformly static personnel allocation is critically **insufficient**
  - 12 agents per field office not enough





Approach

01	**Aggregate a Large, Representative Dataset**	• **Cybercrime data** from Privacy Rights Clearninghouse (PRC) • **Socio-economic data** from U.S. Census Bureau and U.S. Bureau of Economic Analysis • **Technological data** from U.S. Census Bureau
02	**Clean and Select Significant Covariates**	• Fill in missing values. • From the aggregated dataset of ~1M data points, determine significant covariates at state level.
03	**Leverage Geospatial Temporal Regression Model**	• Apply regression model. • Forecast coefficients for 2025 to predict cybercrime hotspots and workload. • Apply to field office jurisdictions.

Aggregate Dataset

Target Variable:

- **Cybercrime data** from Privacy Rights Clearninghouse (PRC)
  - Number of records compromised (severity)
  - 2009 - 2022

Covariates:

- **Socio-economic data** from U.S. Census Bureau and U.S. Bureau of Economic Analysis
- **Technological data** from U.S. Census Bureau

Considered **26** covariates in total

- ~1M data points

Covariate	Source	Years
% Households ($75,000+)	ACS S1901	2010-2022
% Bachelors or Above	ACS DP02	2010-2022
% Single Parent Households[3]	ACS DP02	2010-2022
% Uninsured Population	ACS S2701	2010-2022
% Below Poverty	ACS S1702	2010-2022
% Households No Vehicle	ACS DP04	2010-2022
% African American	ACS DP05	2010-2022
% Indigenous	ACS DP05	2010-2022
% Asian	ACS DP05	2010-2022
% Hispanic/Latino	ACS DP05	2010-2022
GSP (Millions of Dollars)	BEA	1998-2022
Unemployment Rate	ACS S2301	2010-2022
% Agriculture, Forestry, Fishing, Mining	ACS DP03	2010-2022
% Construction	ACS DP03	2010-2022
% Manufacturing	ACS DP03	2010-2022
% Wholesale Trade	ACS DP03	2010-2022
% Retail Trade	ACS DP03	2010-2022
% Transportation, Warehousing, Utilities	ACS DP03	2010-2022
% Information	ACS DP03	2010-2022
% Finance, Insurance, Real Estate, Leasing	ACS DP03	2010-2022
% Professional, Scientific, Management Services	ACS DP03	2010-2022
% Educational Services, Health Care	ACS DP03	2010-2022
% Arts, Entertainment, Recreation, Food Services	ACS DP03	2010-2022
% Other Services	ACS DP03	2010-2022
% Public Administration	ACS DP03	2010-2022
% Households With Internet	ACS DP02	2013-2022

Methodology

- **Statistical Regression–** Implement a geographically and temporally weighted regression (GTWR) approach
  - *Useful technique to calibrate regional spatio-temporal models.*
  - *Comparison with*
    - *Geographically Weighted Regression (GWR)*
    - *Ordinary Least Squares (OLS)*

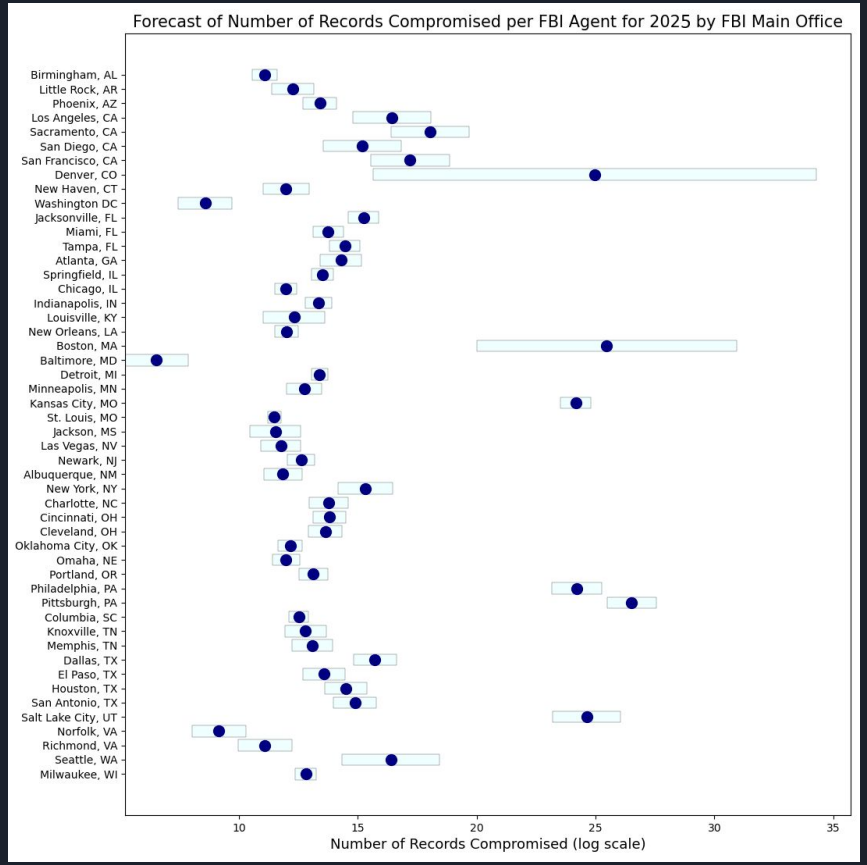
Method	R^2	AIC	RSS
OLS	0.3745	1919.73	2058.08
GWR	0.8507	1831.20	2866.18
GTWR	0.7320	1831.20	2601.12

- **Forecasting–** Implement Vector Autoregressive (VAR) model to forecast coefficients.

2025 Forecast Workload

FBI Field offices with great variability:

- Denver, CO
- Boston, MA



Actionable Recommendations & Next Steps

Immediate Actions:

- Dynamic personnel resource allocation efforts.

- Great need for an **increased number of cybercrime agents** across all field offices (and states)
  - Particularly in Colorado, California, Texas, and Massachusetts.
  - Increasing the number of cybercrime agents may help **alleviate burnout and increase workforce retention.**

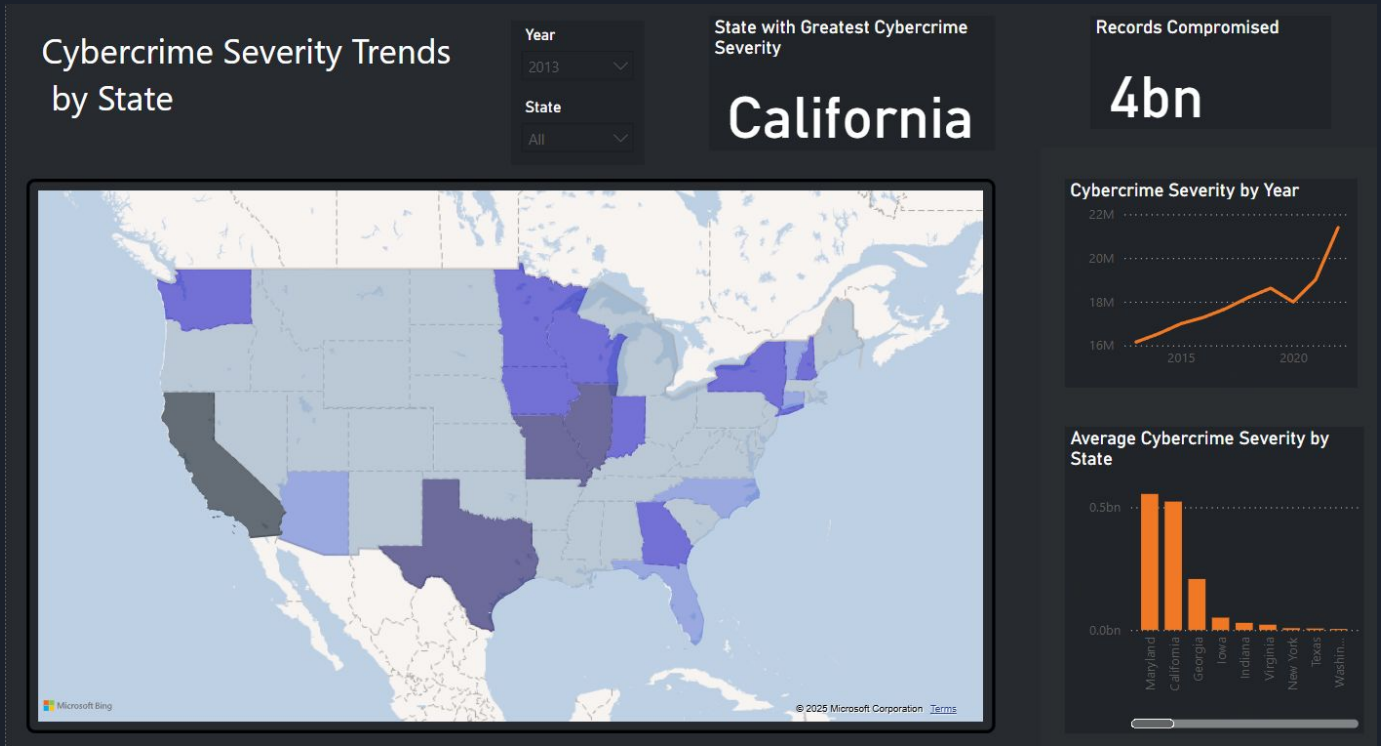
Strategic Alignment:

- **DHS Strategic Plan – Goal 3: Secure Cyberspace and Critical Infrastructure**

Future Considerations:

- Integrate **employee data** with skill sets for more strategic alignment.
- Integrate **field office data** for more accurate reporting and workload capacity.

Dashboard


Cybercrime Severity Trends by State
Year
2013
State
All
State with Greatest Cybercrime Severity
California
Records Compromised
4bn
Microsoft Bing
© 2025 Microsoft Corporation Terms
Cybercrime Severity by Year
22M
20M
18M
16M
2015
2020
Average Cybercrime Severity by State
0.5bn
0.0bn
Maryland
California
Georgia
Iowa
Indiana
Virginia
New York
Texas
Washin...

Thank you.

Questions?

Multicollinearity

	Selected Covariates				Response
	GSP	% Professional	% Educational	% Other	Records Compromised
GSP	1.00	0.57	0.02	-0.05	0.39
% Professional	0.57	1.00	-0.10	0.23	0.49
% Educational	0.02	-0.10	1.00	-0.49	-0.39
% Other	-0.05	0.23	-0.49	1.00	0.28
Records	0.39	0.49	-0.39	0.28	1.00

Table 4: The Pearson correlations of each selected covariate and response variable.

	Selected Covariates				Response
	GSP	% Professional	% Educational	% Other	Records Compromised
GSP	-	1.23 (1.38)	1.51 (1.95)	1.52 (1.93)	1.54 (1.93)
% Professional	2.01 (2.11)	-	2.28 (3.02)	2.58 (2.96)	2.61 (3.05)
% Educational	3.09 (3.48)	2.77 (3.24)	-	3.17 (3.37)	3.16 (3.33)
% Other	1.86 (2.05)	1.86 (1.97)	1.88 (1.82)	-	1.81 (2.12)
Records	1.05 (1.62)	1.05 (1.54)	1.05 (1.49)	1.03 (1.66)	-

Table 5: The Variance Inflation Factor (corresponding spatial lags) for selected covariates and response variable.

Morans' I - Spatial dependencies

Selected Covariates				Response
GSP	% Professional	% Educational	% Other	Records Compromised
0.0948^{***}	0.3455^{***}	0.6152^{***}	0.3390^{***}	0.0165^{**}

Table 6: The global Moran's I estimate for each selected covariate and reponse variable. $*$, $**$, and $***$ indicate significance level of 10%, 5%, and 1%. We use the *spdep* (R.S. Bivand et al., 2008) *R* package with queen adjacency.

- Significant spatial dependency in the severity of cybercrime, with the log1p transformation, at the 5% significance level.

Geographically and Temporally Weighted Regression (GTWR)

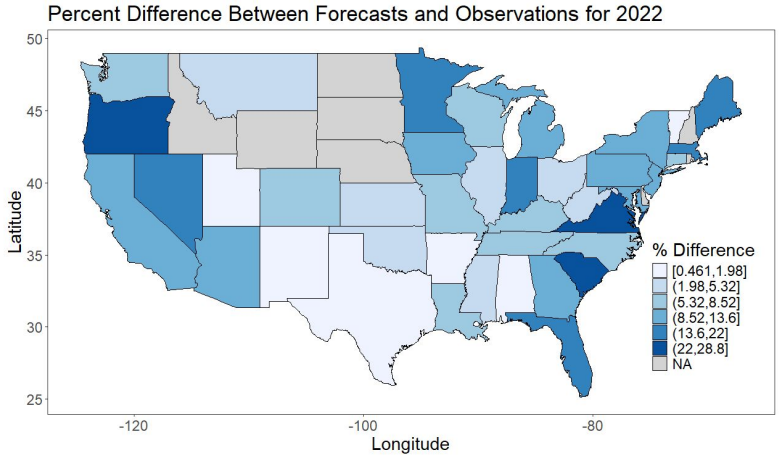
Geographically and Temporally Weighted Regression (GTWR):

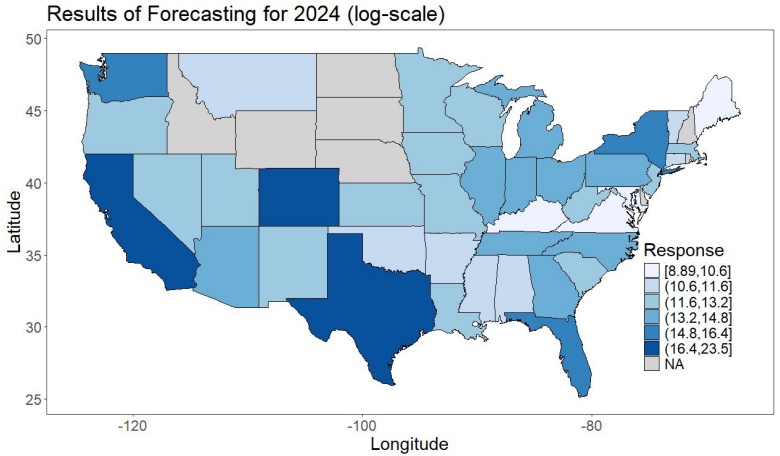
$$Y_j = \beta_0(u_j, v_j, t_j) + \sum_{k=1}^{d} \beta_k(u_j, v_j, t_j) X_{jk} + \epsilon_j, \qquad j = 1, \dots, n$$

where

- Y_j – the severity of the cybercrime (number of compromised records)
- β_0 – estimates the intercept
- (u_j, v_j, t_j) – spatio-temporal coordinates for the *j*-th cybercrime observation at state-level coordinates
- β_k – **estimates the relationship between cybercrime severity and determinants**
- X_{jk} – covariate of state-level determinant (e.g. technological, socio-economic, etc.)
- ϵ_j – error term

Backtesting

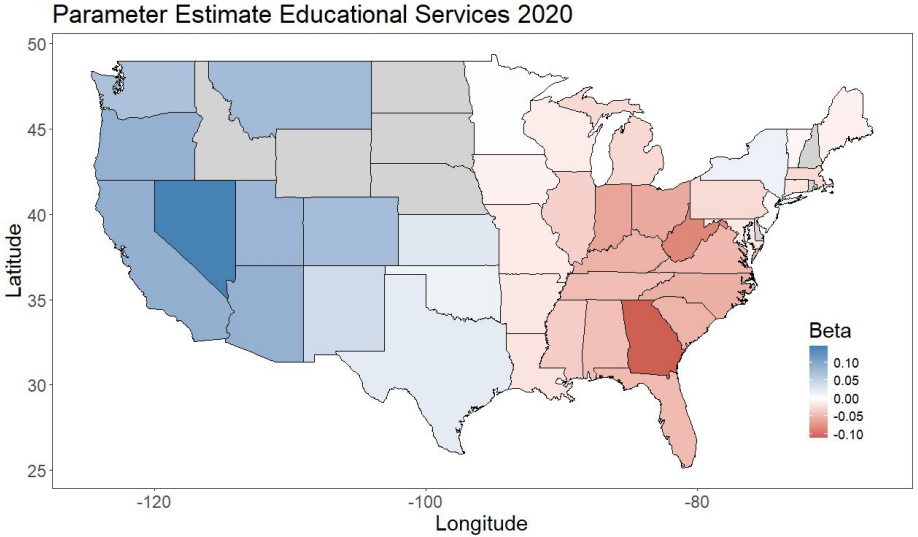




Error Difference

Minimum: 0.461%, Average: 9.5%, Maximum: 29.82%

Parameter Estimates Visualized



- Parametric estimates are given for the industry penetration of the education services by state and time.
- Shows the relationship between cyber attack severity and independent variables (sign and magnitude).