How to Prioritize Locations for Railroad Safety Inspection Programs

Ruby Li, Federal Railroad Administration, USDOT
Jianqiang (Tony) Ye, Federal Railroad Administration, USDOT
Young-Jun Kweon, Bureau of Transportation Statistics, USDOT

U.S. Department of Transportation
Federal Railroad Administration

October 15, 2024

Disclaimer

The opinions, findings, and conclusions expressed in this presentation are those of the authors and not necessarily those of the United States Department of Transportation, the Federal Railroad Administration, or the Bureau of Transportation Statistics. The United States Government assumes no liability for its contents or use thereof.

Our Safety Mission

The mission of the Federal Railroad Administration (FRA) is to enable the safe, reliable, and efficient movement of people and goods for a strong America, now and in the future.

Reducing and eliminating risk is paramount to enhancing safety across the rail industry.

Nearly 400 Federal safety inspectors focus on compliance and enforcement in:

- Grade Crossings
- Hazardous Materials
- Motive Power and Equipment
- Operating Practices
- Signal and Train Control
- Track

Inspection Locations: AIRS Database

Asset Inventory of Railroads and Shippers (AIRS):

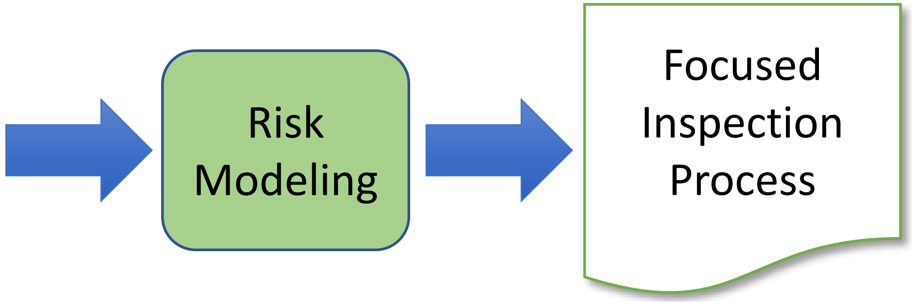
- ☑ Developed by discipline subject matter experts to identify the highest priority characteristics of inspection points
- ☑ Is the basis for assigning inspectors territories and planning inspection activities
- ☑ As an inventory system, collects appropriate data regarding each inspection point
- ☑ AIRS ensures inner-discipline consistency in collection methods, data structures and data maintenance
- ☑ Provides GPS coordinates for each inspection point which can be mapped and joined with other relevant geospatial data

- ☑ **There are 16,000 AIRS locations and over 126,000 open, at-grade crossings for inspection in the US - with only 400 inspectors the FRA must deploy resources effectively and efficiently**

Objective of Risk Models

- FRA developed a risk model to support inspection planning within each discipline
  - The models focus on the data relevant to each discipline's specific needs
  - Different datasets may be used for different disciplines
- The models support prioritization of inspection locations so resources can be focused on high-risk areas
- Inspection locations are defined in FRA's AIRS database or the Grade Crossing Inventory System (GCIS)

FRA Data
AIRS by Discipline
GCIS (FRA Form 6180.71)
Inspections (FRA Form 6180.96)
Incidents (FRA Form 6180.54/57)
Casualties (FRA Form 6180.55a)
NARN (North American Railroad Network)
Other Datasets
Population (Census Bureau)
Non-Accident Release (NARS) (PHMSA)
Waybill Sample (Surface Transportation Bureau)
Additional data sources may be used based on the model. Data sources are continually evaluated for usage in models, and thus not all current sources may be represented



Risk Score Definition

- Risk scoring
  - Predicts which Inspection Locations are most at risk of having a safety incident based on data analytics and modeling
  - Estimates the potential negative impact of an incident

To quantify risk the following calculation assigns a risk score:

$$\text{Risk Score}(r) = \text{Probability}(r) * \text{Severity}(r)$$

Where:

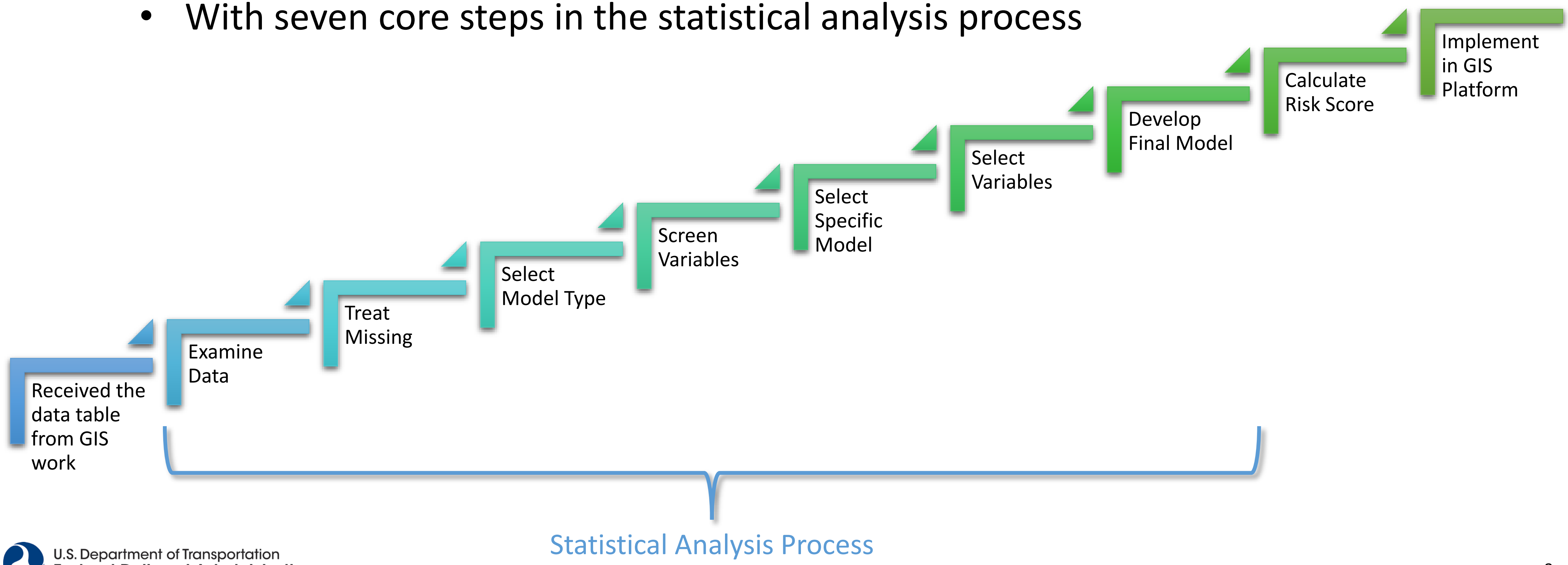
- $\text{Probability}(r) = \sum \text{Probability Factors}\ (r)$
- $\text{Severity}(r) = \sum \text{Severity Factors}\ (r)$
- r: AIRS point, i.e., inspection location

FRA Risk Model Process: First Generation

1. Research Methodologies by discipline including literature review
2. Develop and determine the methodologies
3. Explore data source(s) and build data table in SAS including data cleaning and mining
4. Explore and determine the decision variables
5. Develop a mathematical model with a good fit
6. Validate the model
7. Produce the risk score by AIRS point
8. Implement visualization in Tableau

FRA Risk Model Process: Second Generation

- FRA implemented a 10-step process for the next generation of risk models
  - With seven core steps in the statistical analysis process



Deciding On the Final Predictive Model

- Many AIRS locations do NOT have data on reported incidents (blank fields at the time of analysis), while many locations also have zero reported incidents – *creating a dataset with excessive zeros*
- We have considered and tested many candidate regression models for V1 and V2 such as Poisson, Negative Binomial, etc.
- The Zero-inflated Negative Binomial (ZINB) has turned out to be the best fit model for most of our disciplines.

The mathematical model for the ZINB:

$$\Pr(y_i = j) = \begin{cases} \pi_i + (1-\pi_i) g(y_i = 0) & if\ j = 0 \\ (1-\pi_i) g(y_i) & if\ j > 0 \end{cases} \qquad (1)$$

$$g(y_i) = \Pr(Y = y_i \mid \mu_i, \alpha) = \frac{\Gamma(y_i + \alpha^{-1})}{\Gamma(\alpha^{-1})\Gamma(y_i + 1)} \left(\frac{1}{1+\alpha\mu_i}\right)^{\alpha^{-1}} \left(\frac{\alpha\mu_i}{1+\alpha\mu_i}\right)^{y_i} \qquad (2)$$

$$\mu_i = \exp(x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki}) \qquad (3)$$

Risk Score Metrics Example: MP&E Risk Model Version 2

Risk scores are calculated for each AIRS data point based on multiple factors that influence the probability or severity of an event.

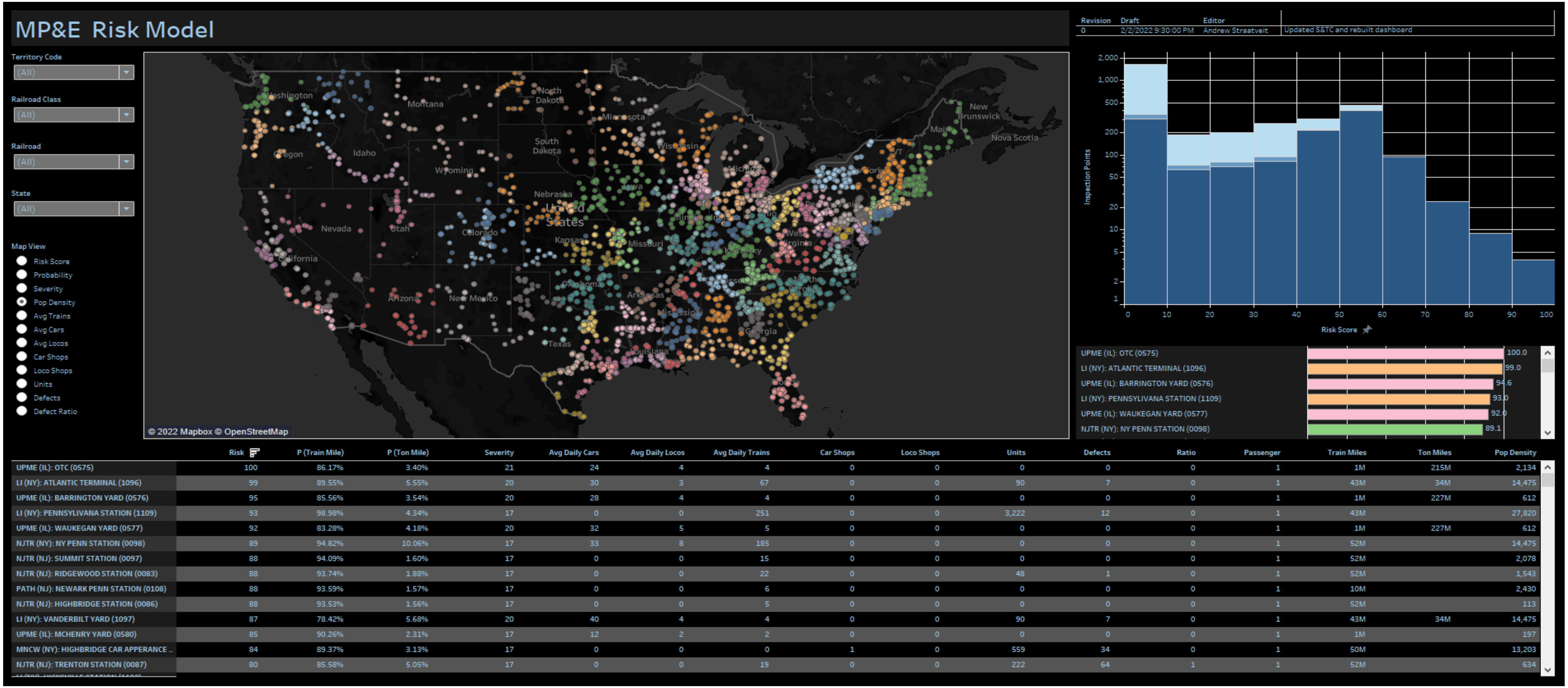
Probability: a calculation of the coefficients in a statistical model using ZINB influenced by the factors listed below:

- Daily Freight Traffic
- Hazmat Cars
- Accountable Incidents
- Average Daily Trains
- Locomotive Shops
- District
- Signal Type
- Track Class
- Average Daily Cars
- Number of Car Shops

Severity: a summation of the severity factors listed below. Each factor is analyzed and assigned an appropriate bin scale.

- Days Since Last Inspection
- Passenger Service
- Accident Total Cost
- Toxic Inhalation Hazard (TIH) Ton Miles
- Class 3 Flammable Ton Miles
- Fatalities/Injuries
- Stracnet Miles
- Population Density

Tableau Dashboard for Users: First Generation

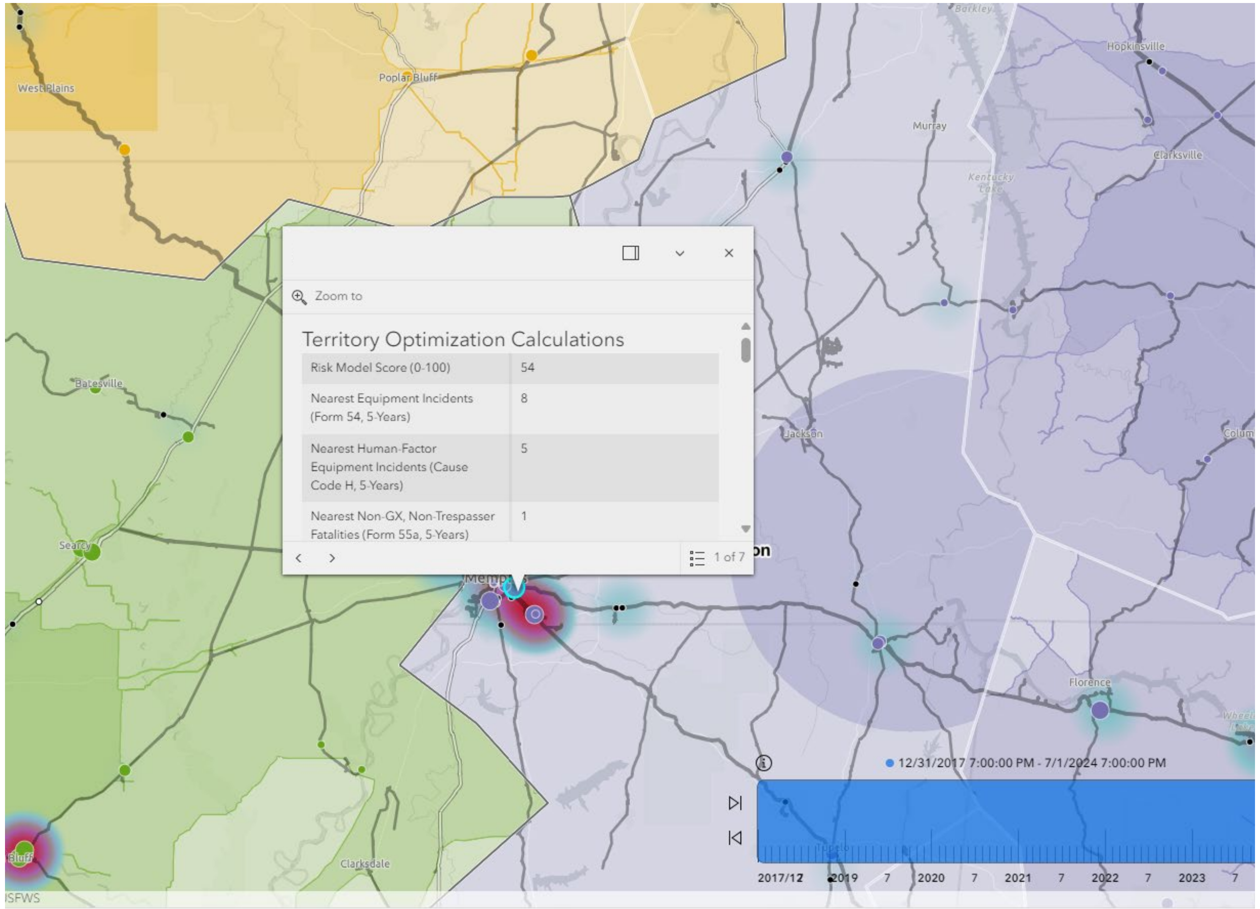


	Risk	P (Train Mile)	P (Ton Mile)	Severity	Avg Daily Cars	Avg Daily Locos	Avg Daily Trains	Car Shops	Loco Shops	Units	Defects	Ratio	Passenger	Train Miles	Ton Miles	Pop Density
UPME (IL): OTC (0575)	100	86.17%	3.40%	21	24	4	4	0	0	0	0	0	1	1M	215M	2,134
LI (NY): ATLANTIC TERMINAL (1096)	99	89.55%	5.55%	20	30	3	67	0	0	90	7	0	1	43M	34M	14,475
UPME (IL): BARRINGTON YARD (0576)	95	85.56%	3.54%	20	28	4	4	0	0	0	0	0	1	1M	227M	612
LI (NY): PENNSYLIVANA STATION (1109)	93	98.98%	4.34%	17	0	0	251	0	0	3,222	12	0	1	43M		27,820
UPME (IL): WAUKEGAN YARD (0577)	92	83.28%	4.18%	20	32	5	5	0	0	0	0	0	1	1M	227M	612
NJTR (NY): NY PENN STATION (0098)	89	94.82%	10.06%	17	33	8	185	0	0	0	0	0	1	52M		14,475
NJTR (NJ): SUMMIT STATION (0097)	88	94.09%	1.60%	17	0	0	15	0	0	0	0	0	1	52M		2,078
NJTR (NJ): RIDGEWOOD STATION (0083)	88	93.74%	1.88%	17	0	0	22	0	0	48	1	0	1	52M		1,543
PATH (NJ): NEWARK PENN STATION (0108)	88	93.59%	1.57%	17	0	0	6	0	0	0	0	0	1	10M		2,430
NJTR (NJ): HIGHBRIDGE STATION (0086)	88	93.53%	1.56%	17	0	0	5	0	0	0	0	0	1	52M		113
LI (NY): VANDERBILT YARD (1097)	87	78.42%	5.68%	20	40	4	4	0	0	90	7	0	1	43M	34M	14,475
UPME (IL): MCHENRY YARD (0580)	85	90.26%	2.31%	17	12	2	2	0	0	0	0	0	1	1M		197
MNCW (NY): HIGHBRIDGE CAR APPERANCE ..	84	89.37%	3.13%	17	0	0	0	1	0	559	34	0	1	50M		13,203
NJTR (NJ): TRENTON STATION (0087)	80	85.58%	5.05%	17	0	0	19	0	0	222	64	1	1	52M		634

Web Map Applications for Second Generation

Territory Optimization Planning System (TOPS)

- Web mapping application
- Updated monthly with new data
- Risk scores recalculated monthly as new incidents and inspections occur
- Visualizes risk scores within inspector territories
- Allows inspectors to provide additional “Priority Scores” based on field experience
- Calculates “Combined Score” from risk model and priority inputs



U.S. Department of Transportation
Federal Railroad Administration

Contact Us

Federal Railroad Administration
1200 New Jersey Avenue, SE
Washington, DC 20590

Connect with us at USDOTFRA

Ruby Li

Email: ruby.li@dot.gov

Jianqiang (Tony) Ye

Email: jianqiang.ye@dot.gov

Young-Jun Kweon

Email: young-jun.Kweon@dot.gov

U.S. Department of Transportation
Federal Railroad Administration