

# Predicting Vessel Delays at Army Corps of Engineers Locks

A Machine Learning Classification  
Study



Cargo vessel traffic on US waterways and through US Army Corps of Engineers-managed locks plays an important role in the transport of foreign and domestic goods into and out of the United States.

Lock and Dam 1, Minneapolis. Photo by Patrick Loch.  
<http://www.mvp.usace.army.mil/Media/Images/igphoto/2001935893/>



Goal: to determine what factors predict extended vessel delays at locks.

Locations of 196 locks owned or managed by the US Army Corps of Engineers and monitored with the Lock Performance Monitoring System (LPMS).

# US Army Corps of Engineers

IWR: Institute of Water Resources

NDC: Navigation and Civil Works  
Decision Support Center

LPMS: Lock Performance  
Management System

Maintain Data for:

- Waterborne commerce
  - Vessel Characteristics
  - Navigation locks data
  - Commodity and Transport data
-

# US Army Corps of Engineers

IWR: Institute of Water Resources

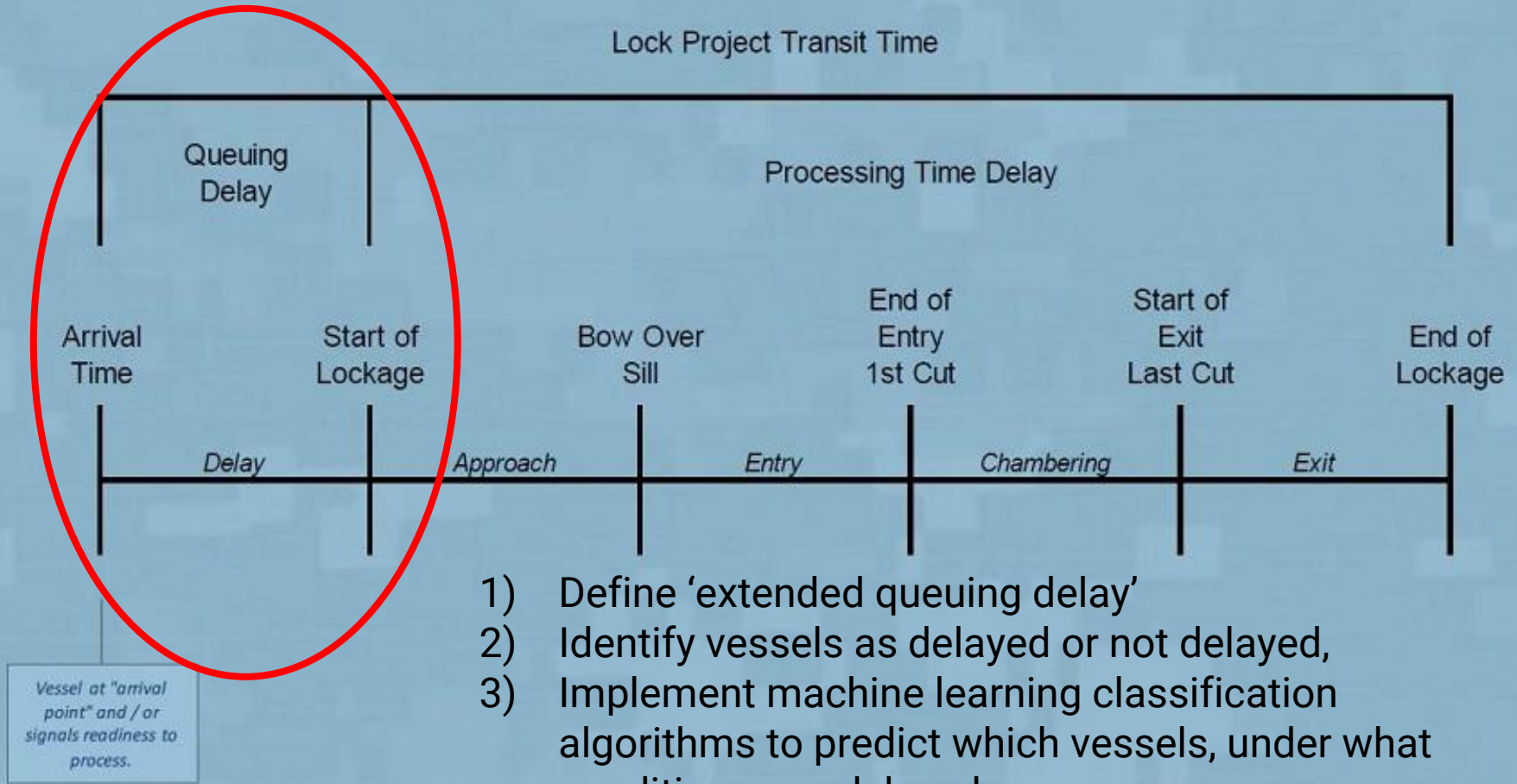
NDC: Navigation and Civil Works  
Decision Support Center

LPMS: Lock Performance  
Management System

## LPMS Analysis Request:

- Deeper analysis of the national network of waterway locks
  - Explore relationships relevant to budgeting, maintenance, scheduling.
  - Identify factors causing delays at locks.
-

## Lock Project Transit Time



- 1) Define 'extended queuing delay'
- 2) Identify vessels as delayed or not delayed,
- 3) Implement machine learning classification algorithms to predict which vessels, under what conditions, are delayed.

# LPMS Data Details

---

## Vessel Traffic Table:

- Limit to 5 years
- Random subsample
- Vessel, River, Lock IDs
- Vessel Function Types
- Lockage Types
- Time marks during lockage
- Calculated: Delay Time

## Stall Stoppage Table:

- Limit to 5 years
- Vessel, River, Lock IDs
- Begin/End Stop Dates
- Scheduled Stop or not?
- Reason Code for Stops
- Calculated: Duration of Stop



# LPMS Data & NOAA Weather Data



Daily Summaries at 7 nearby weather stations:

- Min/Max Temp
- Precipitation
- Wind speeds
- Presence/Absence of 19 extreme weather types.



# Define Extreme Delay

15% of vessels are delayed  
awaiting lockage

**Define 'delay' on a lock-by-lock basis:**

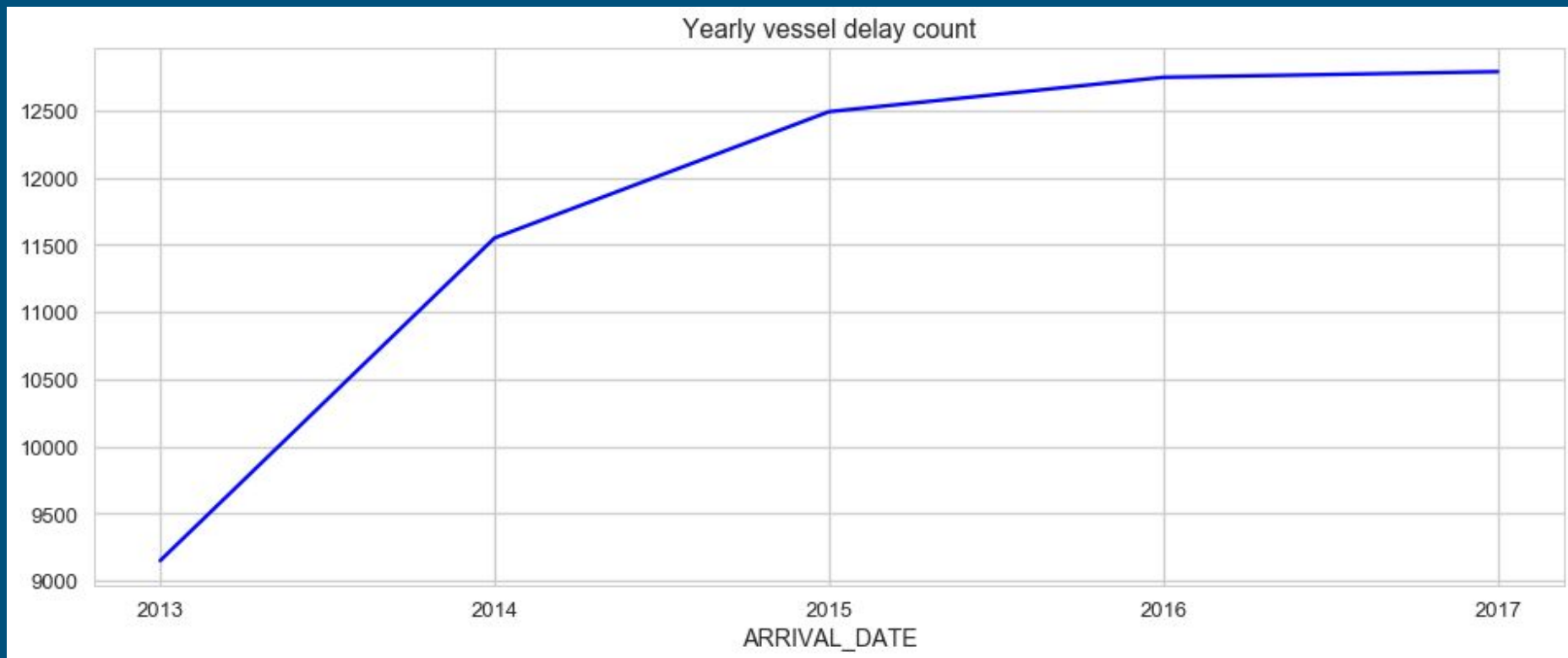
- True delay: duration greater than 75th percentile/lock.
  - False delay: delay-time less than 75th percentile/lock.
  - Extreme delay ranges from minutes to 29 hours.
-

# Unplanned Stall Stoppage Metric

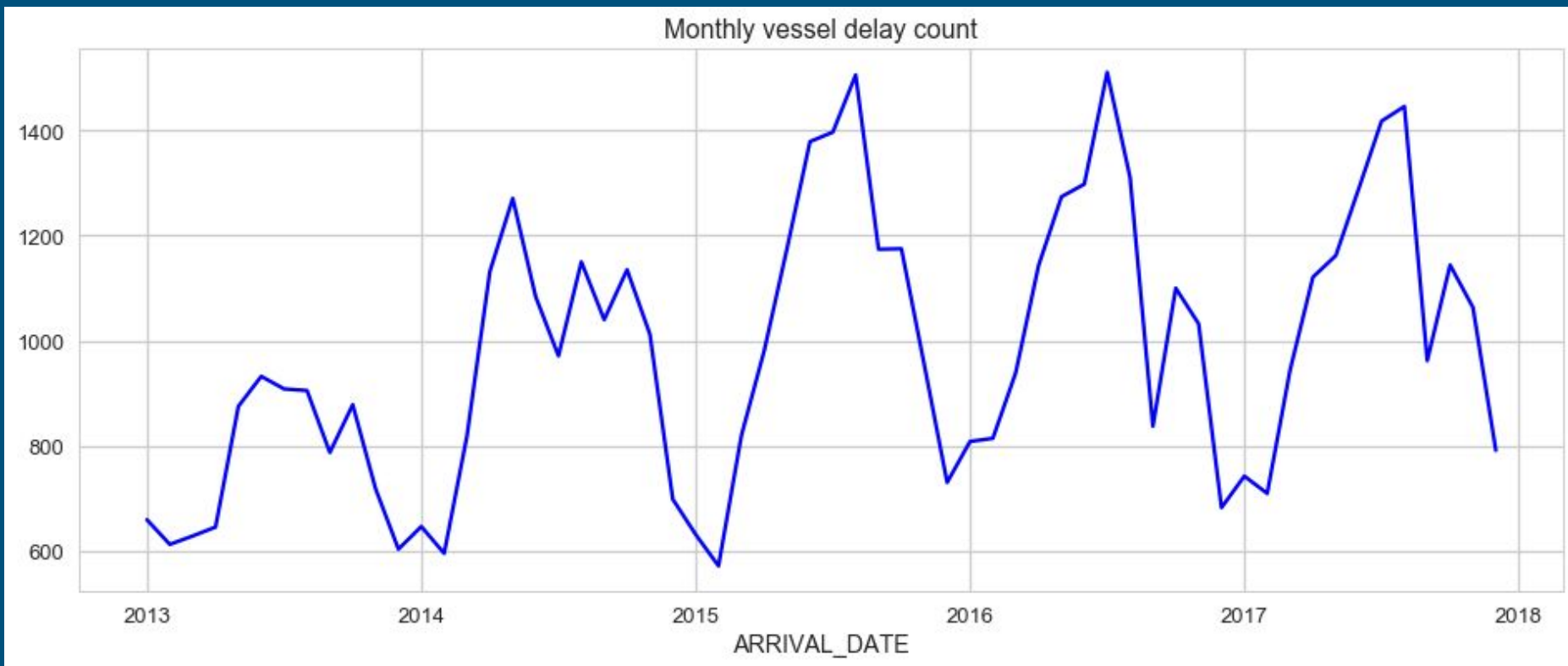
Average duration of unplanned  
stops: 24 hours

## Stall Stoppage Data:

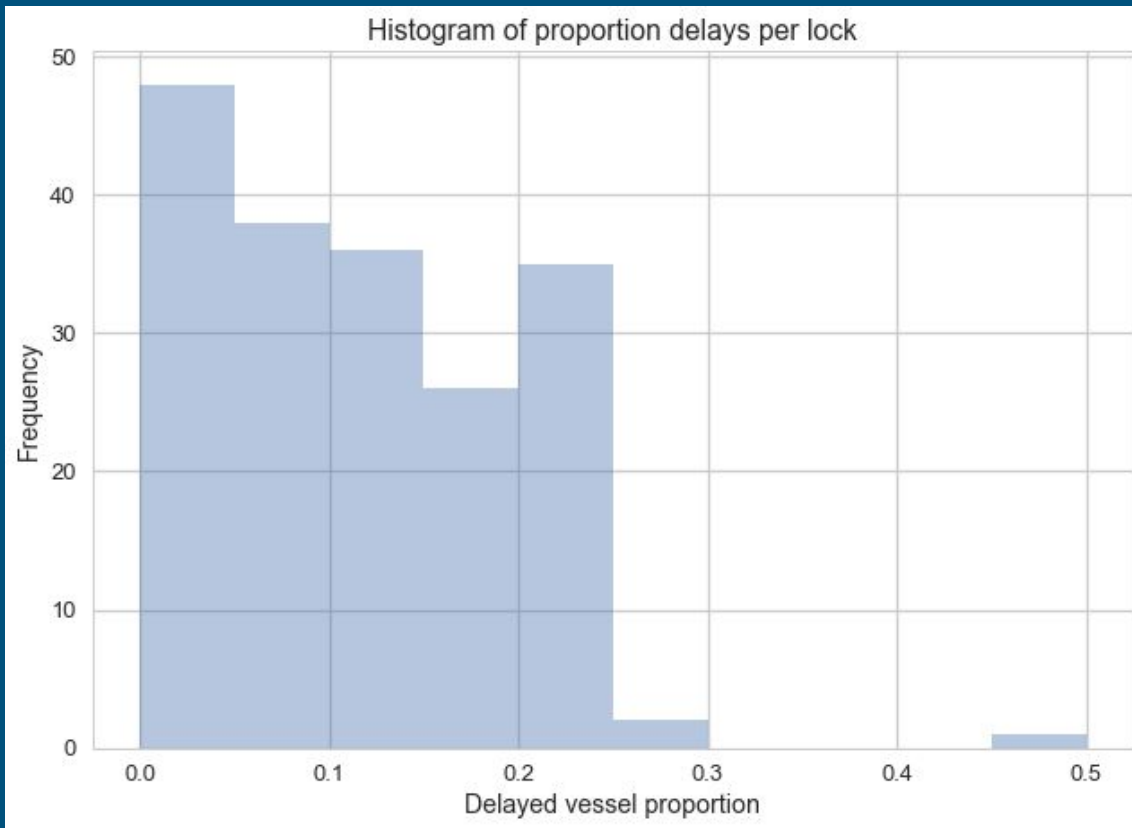
- Removed vessel records where there was a 'Scheduled' stop on a day.
  - Included maximum duration of unplanned stops at a lock/day.
  - Longest unplanned stoppage in 5-year study was 3 months long.
-



Overall extreme-delay counts are increasing



Extreme-delay counts have a seasonal pattern with fewer delays, and perhaps less vessel traffic during the colder seasons.



- Determined the artificial cut-off for 'low' and 'high' proportions of delays
- Dropped all locks with delay proportions less than 0.15 from the pool of locks for machine learning.
- Selected 'focal locks' from locks with high vessel traffic & proportion delayed.

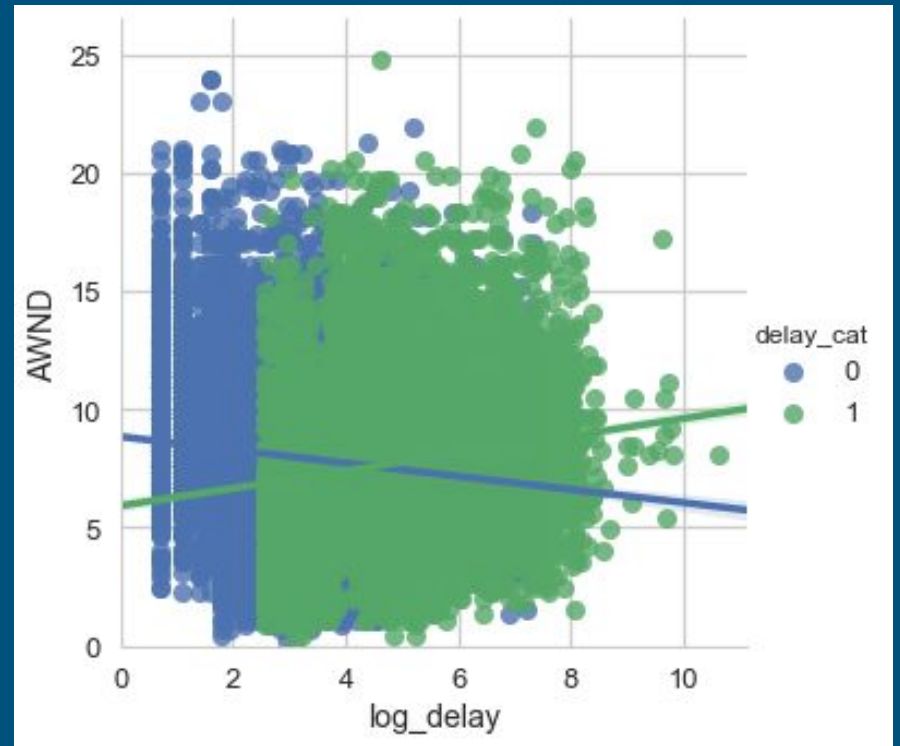
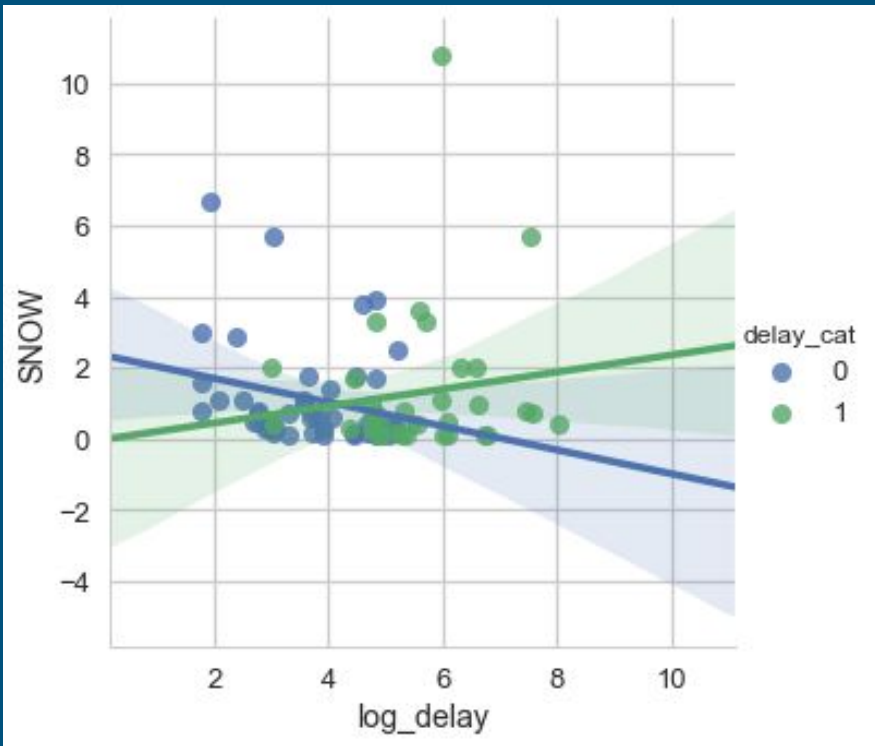
The distribution of delay proportions for locks.



### Focal Lock Data:

- One vessel per day per lock.
- Kept delayed vessels in data.
- 8435 non-delays
- 7103 delays

Locations of 10 focal locks located in 5 states: Texas (3 locks), Indiana (1), Illinois (1), Louisiana (4 locks – two pairs of locks), and Washington (1).



Snowfall depth and average wind speed may have linear relationships by delay-type, but differences are not meaningfully different between the groups.



# Classifying vessels: extreme delays

---

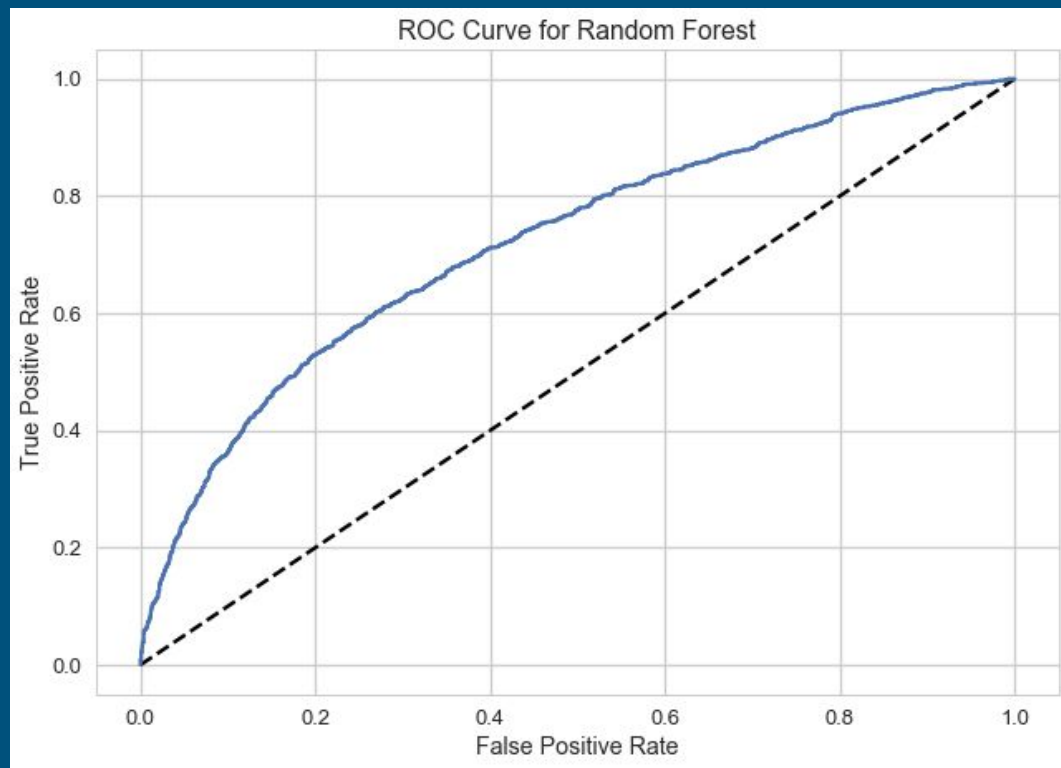
- Machine Learning Approach: supervised learning classification problem.
- Binary target variable (y): 'delay' as defined in EDA steps.
- Feature groups:
  - Lock or vessel related
  - Weather metrics
  - Date & Time derived features
- Matrix of predictive features:
  - 44 features
  - 11 continuous variables
  - 33 binary, including dummy variables for categorical variables.

# Model Selection

---

Compare: Logistic Regression and Random Forest Classifier

- Model 10 focal locks
- Scaled & standardized continuous variables
- Explore models after Recursive Feature Elimination
- Examine feature importances
- Model individual locks separately
- Compare models with ROC-AUC scores



Random Forest Classifier with Grid Search Cross Validation, Full Focal Lock Dataframe, ROC-AUC = 0.729

# Important Features for Predicting Delays

---

Feature Name	Importance
Average Wind Speed	0.1202
Temperature Minimum	0.1151
Temperature Maximum	0.1120
Fastest 5-second wind speed	0.1011
Fastest 2-minute wind speed	0.0898
Temperature Average	0.0628
Precipitation	0.0508
Night	0.0262

# Recommendations

---

# Caveats & Future Directions

---