Ecological Health
Pitch Deck

Harisai Karthikeyan, Jackson Carey, Junkai Mo, Andrew Ranucci

Table of contents

Research Question

Do human-induced changes in ecological parameters change the migration patterns of animals?

Focus

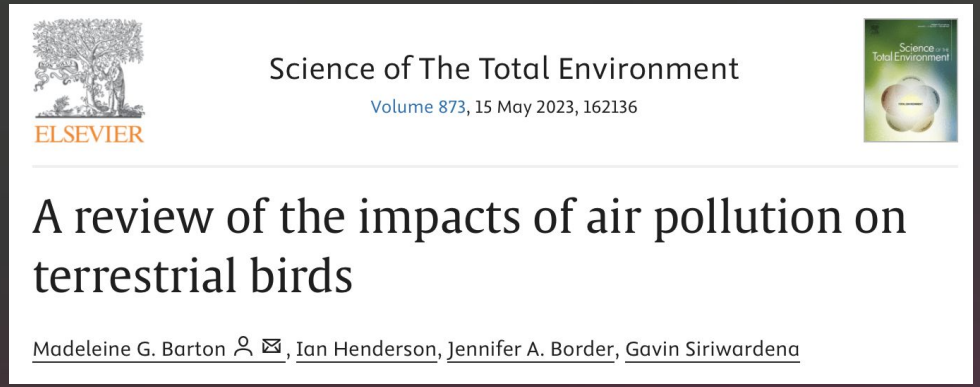
Do changes in air-pollution levels affect the migration pattern of the American robin?

Scientific Literature

Study from science of total environment shows the impacts of air pollution on terrestrial birds

Specifically, NO_2, O_3 and SO_2.

We wanted to see if it affects migration.



ML models we tried

XGB CLASSIFIER

```
grid_search = GridSearchCV(
    XGBClassifier(objective="multi:softmax", num_class=4, random_state=42),
    param_grid,
    scoring='accuracy',
    cv=5
)
grid_search.fit(x_train, y_train)
best_model = grid_search.best_estimator_

y_pred_encoded = best_model.predict(x_test)

inverse_route_mapping = {v: k for k, v in route_mapping.items()}
y_pred = pd.Series(y_pred_encoded).map(inverse_route_mapping)
y_test_original = y_test.map(inverse_route_mapping)

print(classification_report(y_test_original, y_pred))
print(confusion_matrix(y_test_original, y_pred))
```

```
              precision    recall  f1-score   support

          14       0.34      0.22      0.27       258
          22       0.35      0.61      0.45       354
          29       0.38      0.19      0.25       230
          31       0.34      0.27      0.31       273

    accuracy                           0.35      1115
   macro avg       0.35      0.32      0.32      1115
weighted avg       0.35      0.35      0.33      1115
```

RandomForestClassifier ML Model

```
y = df['Ecological_Health_Label']
x = df.drop(['Ecological_Health_Label'], axis=1)
from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(x,y,test_size=0.7, random_state=100)
from sklearn.ensemble import RandomForestClassifier
classifier = RandomForestClassifier(
    n_estimators=300,
    max_depth=10,
    min_samples_split=10,
    min_samples_leaf=10,
    max_features="sqrt",
    bootstrap=True,
    random_state=42,
    n_jobs=-1)
classifier.fit(x_train, y_train)
Eco_Label_Predictions = classifier.predict(x_test)

from sklearn.metrics import accuracy_score
train_accuracy = accuracy_score(y_train, classifier.predict(x_train))
test_accuracy = accuracy_score(y_test, Eco_Label_Predictions)
print("Training Accuracy:", train_accuracy)
print("Test Accuracy:", test_accuracy)
print("Residual", train_accuracy - test_accuracy)
```

```
Training Accuracy: 0.5025875020866952
Test Accuracy: 0.5013473874991057
Residual 0.001240114587589547
```

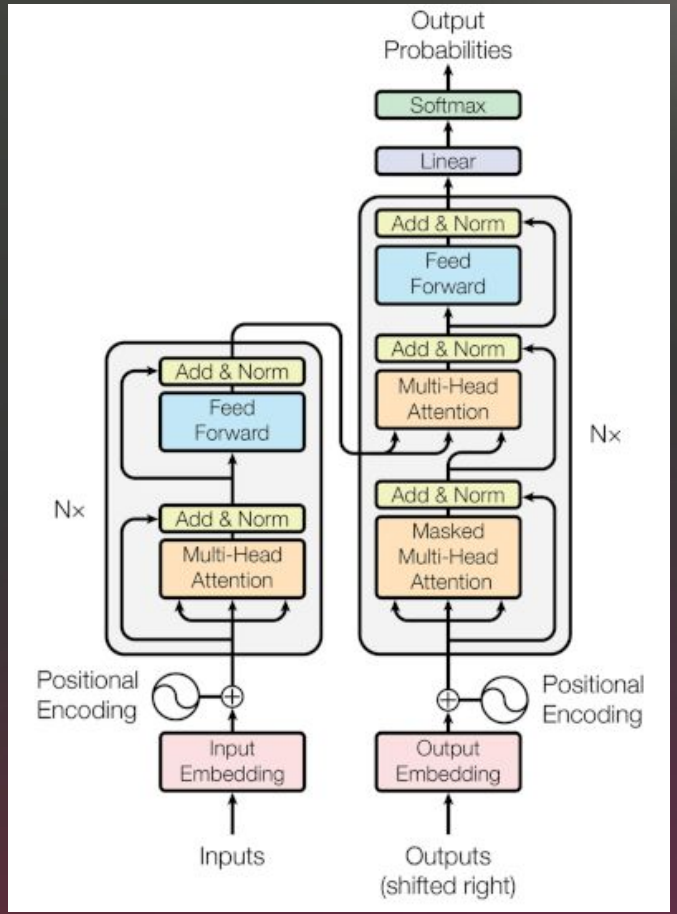
BERT

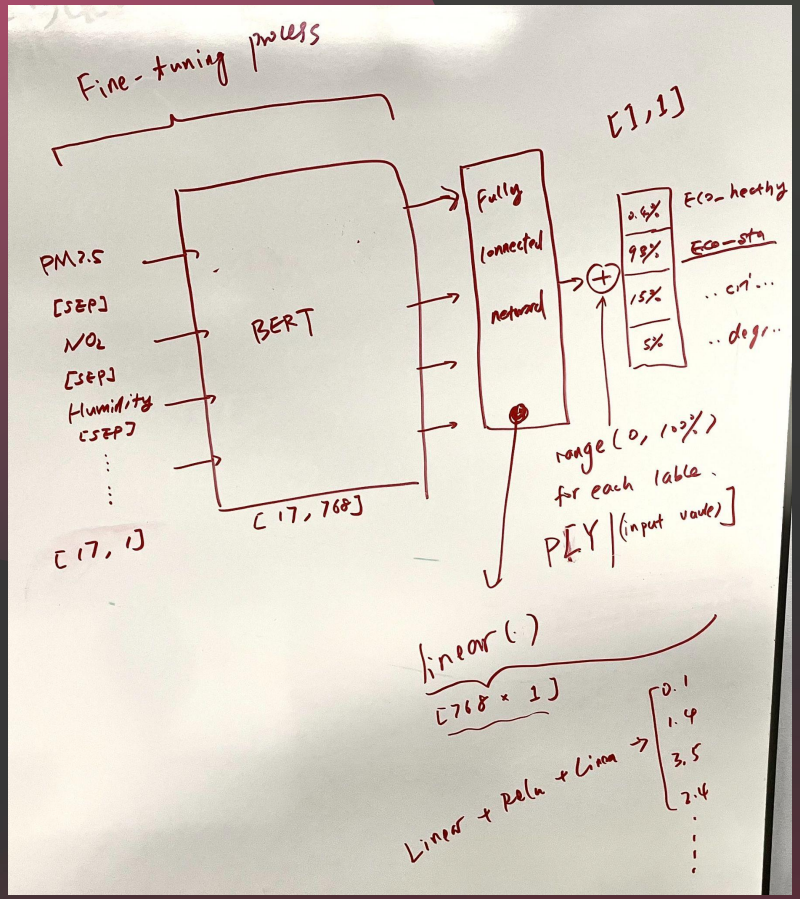
Bidirectional encoder representations from transformers

Why BERT?

BERT is originally a language model developed by google for understanding context in language. Most of its applications lie in Natural Language Processing.

So why did we use it for classification based on numbers?





Why BERT?

We tweaked the CLS layer of BERT to output values based on probability. During the fine tuning process we take in the 14 inputs of the original model, and connect it with a fully connected neural network. After that the output will be generated by the last layer (P[Y| inputs]).

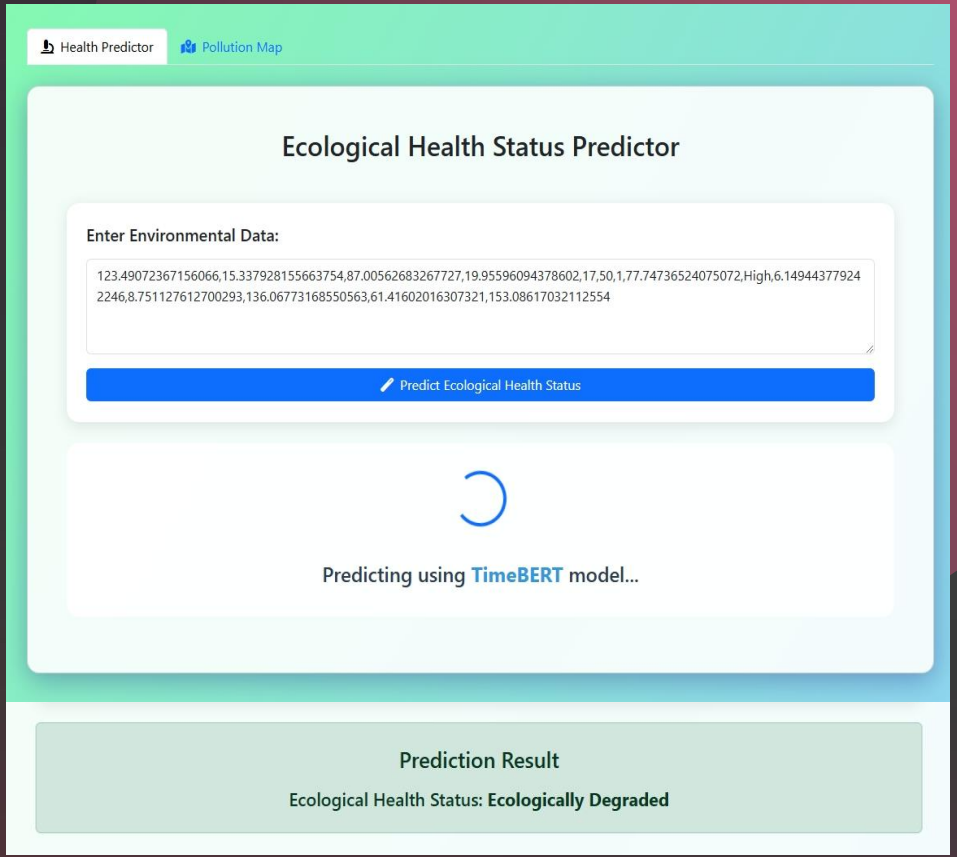
- P[Y=0 | inputs] : Eco_health_healthy
- P[Y=1 | inputs] : Eco_health_stable
- P[Y=2 | inputs] : Eco_health_critical
- P[Y=3 | inputs] : Eco_health_degrading

Baseline Prediction Model

The model must first learn what factors influence environmental quality.

To gain this understanding, we made a model to predict the Ecological Health Label from the given data.

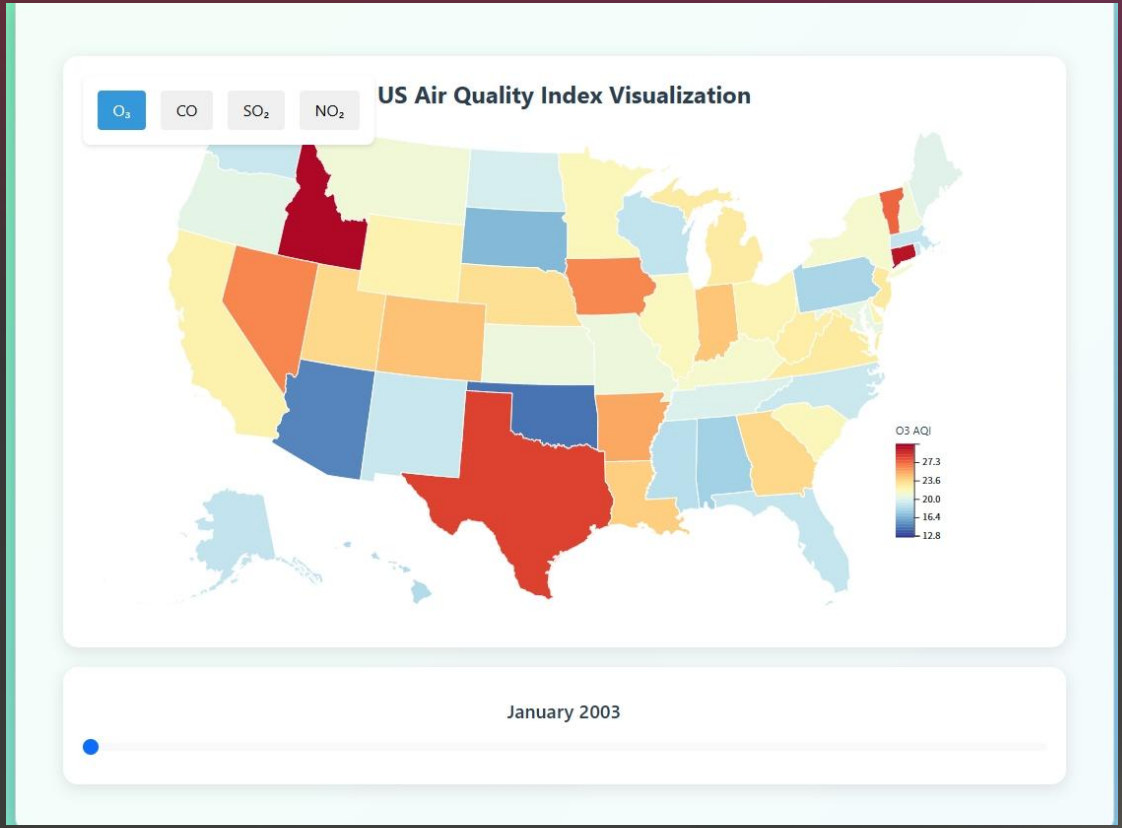
This understanding will be useful for the later models and tasks.

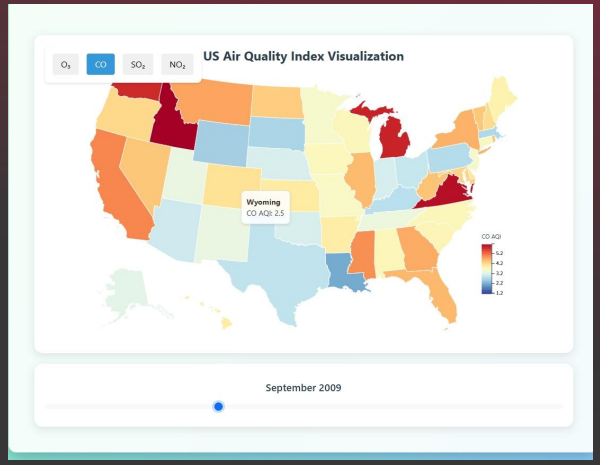


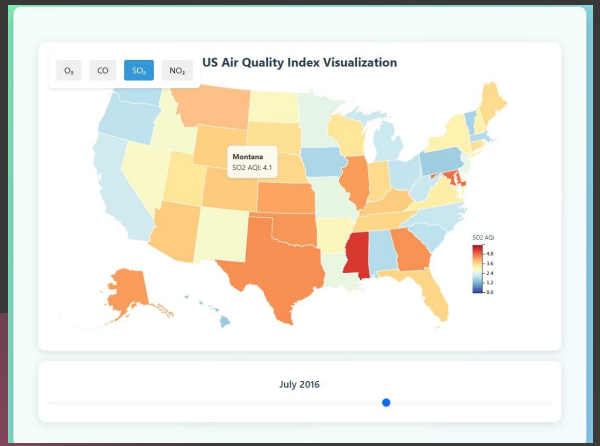
AQI Prediction Model

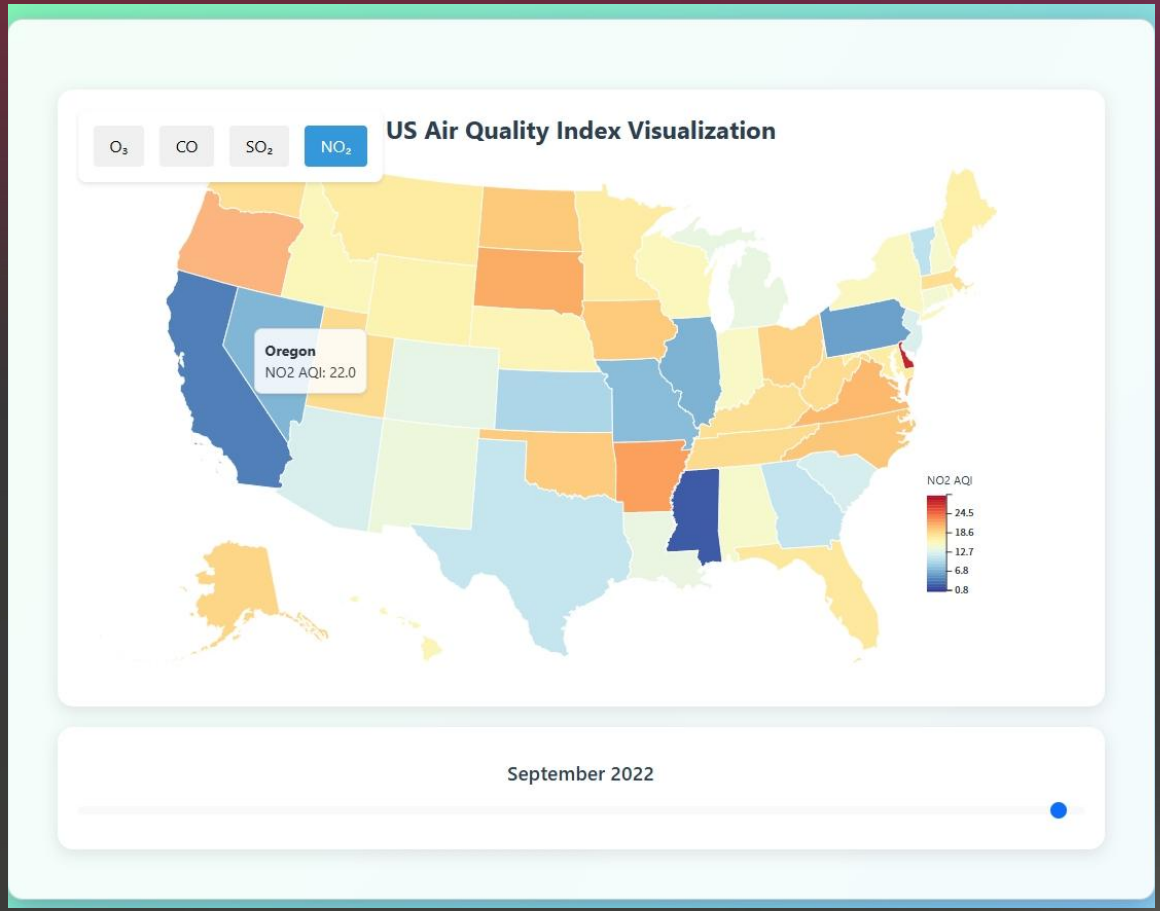
We download the daily pollution data by state and convert it to a monthly average.

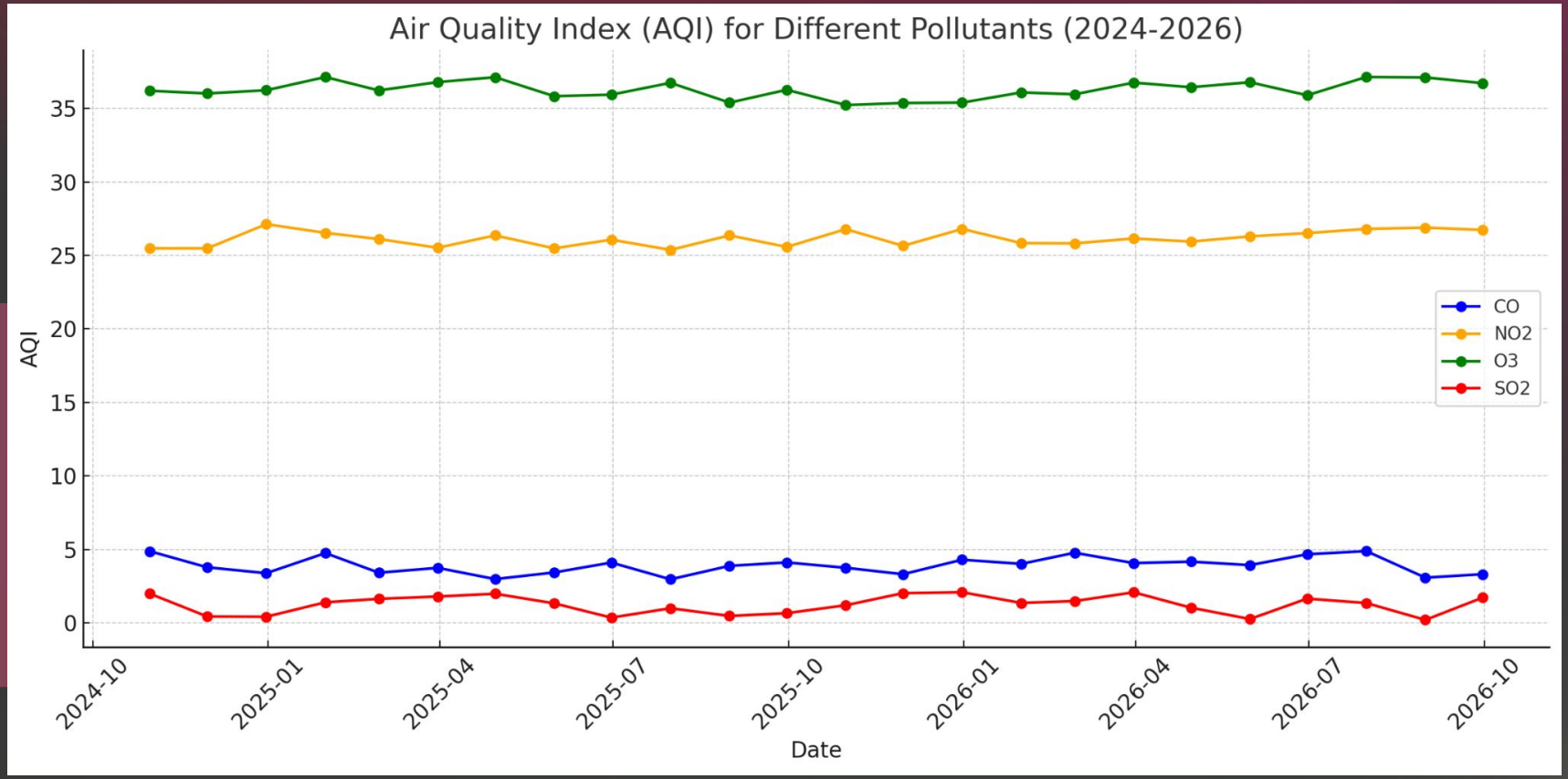
We then find correlations between our existing ecological data and pollution level and predict future pollution values.

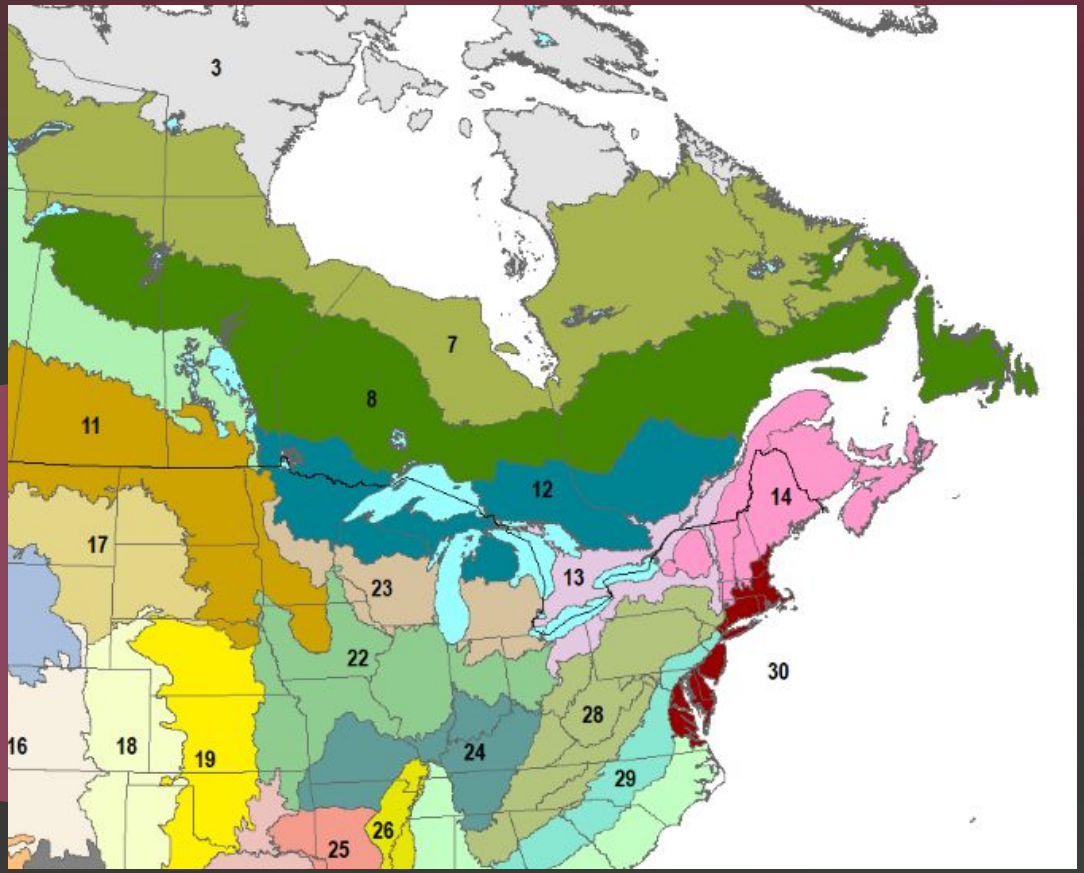



O_3
CO
SO_2
NO_2
US Air Quality Index Visualization
Wyoming
CO AQI: 2.5
CO AQI
September 2009


O_3
CO
SO_2
NO_2
US Air Quality Index Visualization
Montana
SO2 AQI: 4.1
SO2 AQI
July 2016


O_3
CO
SO_2
NO_2
US Air Quality Index Visualization
Oregon
NO2 AQI: 22.0
NO2 AQI
24.5
18.6
12.7
6.8
0.8
September 2022


Air Quality Index (AQI) for Different Pollutants (2024-2026)
AQI
Date
CO
NO2
O3
SO2
0
5
10
15
20
25
30
35
2024-10
2025-01
2025-04
2025-07
2025-10
2026-01
2026-04
2026-07
2026-10

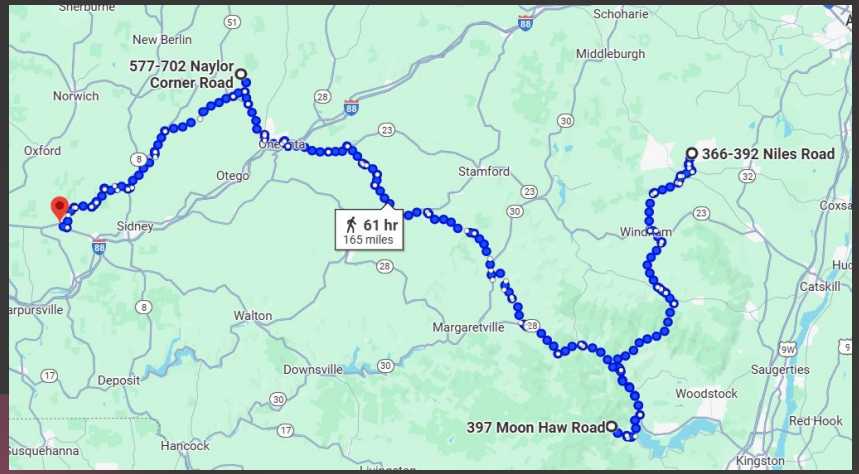


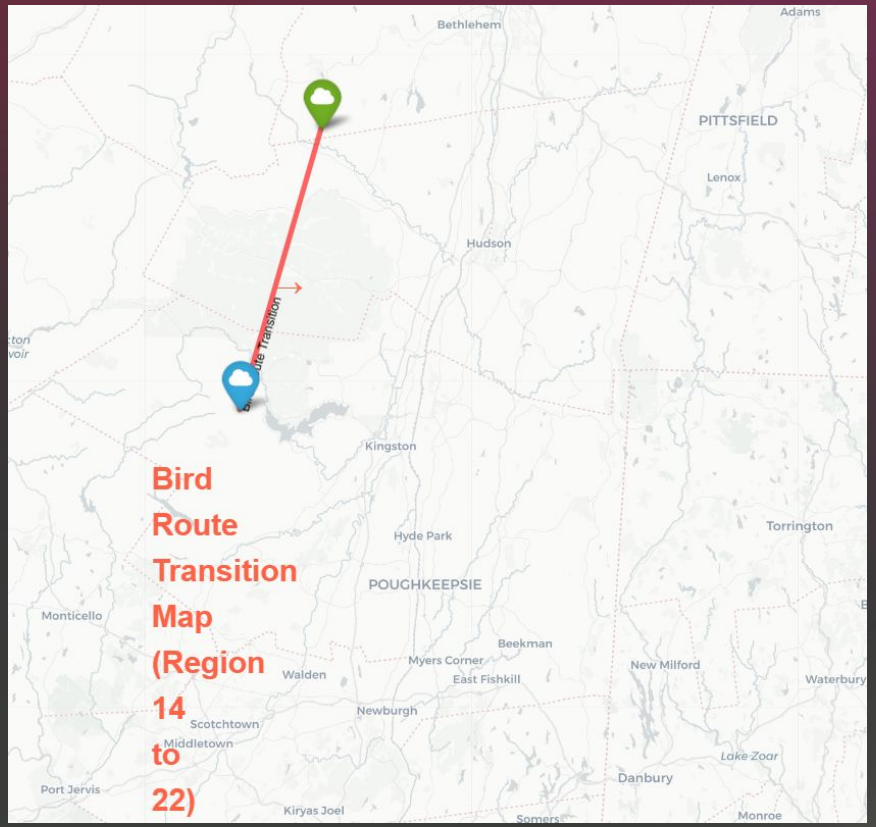
American Robin Migration Model

We use the AQI prediction from the previous model as input to our migration model.

We then predict the route the birds will take and then associate the region they will travel in.

Change in Migration Route





80.43%

accuracy

Citation

Bird Migration - https://www.sciencebase.gov/catalog/item/66d9ed16d34eef5af66d534b

Air Quality Index - https://ephtracking.cdc.gov/

Air Quality by State - https://www.epa.gov/outdoor-air-quality-data/download-daily-data