# Based off the weather, can we predict a latitude?

By Charles Phil, Bryan Mangrum, Coye Roseberry, Hayden Hinton, and Shawn Varghese

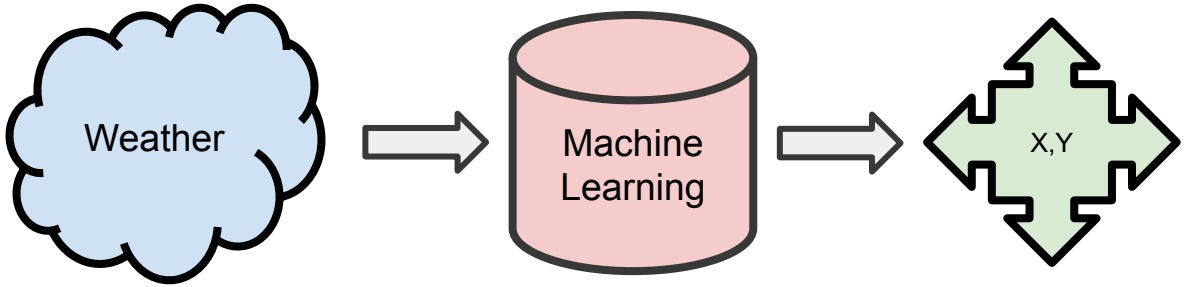
# Overview

## Goal:

- Predict a city's latitude from weather data
- Visualize predicted city locations on a map

## Methods:

- Train a model on city lat/long and weather statistics
- Render city locations on an interactive web map



# Data Wrangling

**Sourced data from OpenWeather Historical weather API:**

- 40 years of aggregated data (by month and day) for 500 Cities
  - ~183k rows of data
  - 47 Columns (potential features)

**Scoped Data:**

- Continental USA
  - 500 ⟶ 387 cities
- Average of each column grouped by city

Example API:


```
{
"cod":200,
"city_id":5400075,
"calctime":3.105838634,
"result":[
   {
      "month": 1,
      "day": 1,
      "temp":{
         "record_min":274.44,
         "record_max":290.45,
         "average_min":276.79,
         "average_max":286.5,
         "median":281.26,
         "mean":281.29,
         "p25":278.57,
         "p75":283.83,
         "st_dev":3.69,
         "num":136
      },
      "pressure":{
```

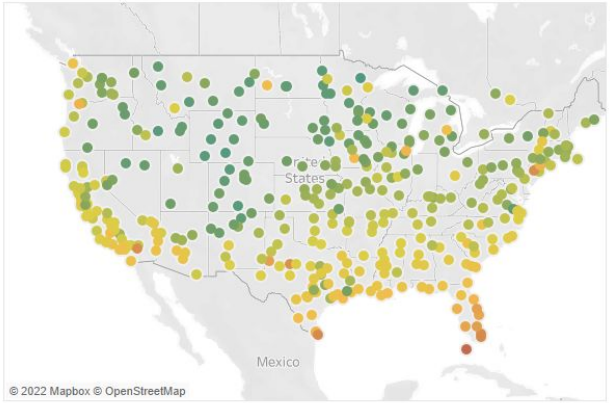
# Tableau

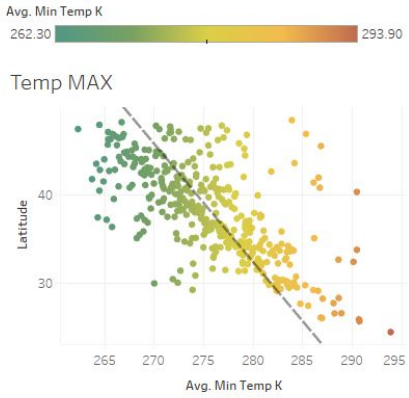
- Tableau helped us visualize the data and adjust scope
- Reduced to just the United States to prevent issues working with both the Northern and Southern hemispheres

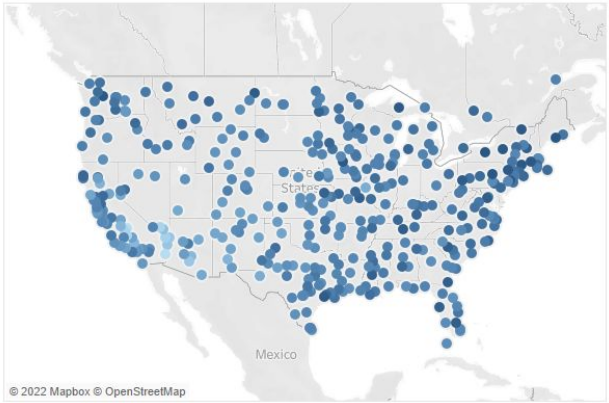
# Feature Selection

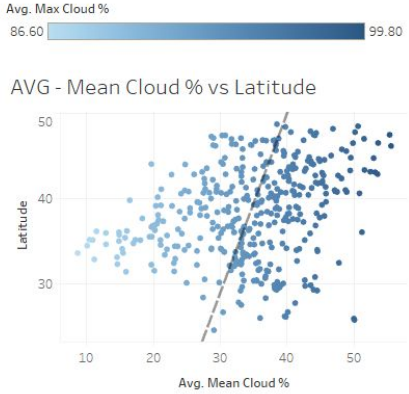
**Started with 47 Weather Stats as potential features, SelectFromModel limited it to 11:**

1. 'Longitude'
2. 'Min_temp_k'
3. 'Max_temp_k'
4. 'avgmax_temp_k'
5. 'Median_temp_k'
6. 'Mean_temp_k'
7. 'P25_temp_k'
8. 'P75_temp_k'
9. 'Stdev_wind_mps'
10. 'mean_cloud_%'
11. 'stdev_cloud_%'



# Machine Learning Process

**Tested two main models to find best fit:**

- Multivariate Linear Regression
  - Training score = 0.68
  - Testing score = 0.30
- Random Forest Regressor
  - Training score = 0.93
  - Testing score = 0.48

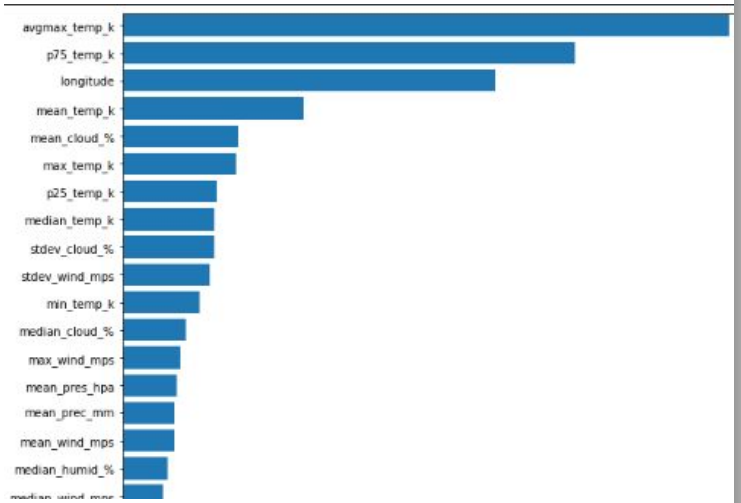
**Best fit:**

- Random Forest Regressor as the estimator for our feature selection
- Multivariate Linear Regression for the predictive model

**Final model scores:**

- Training set: 0.589
- Testing set: 0.478

Random Forest Regressor used to weight feature importance



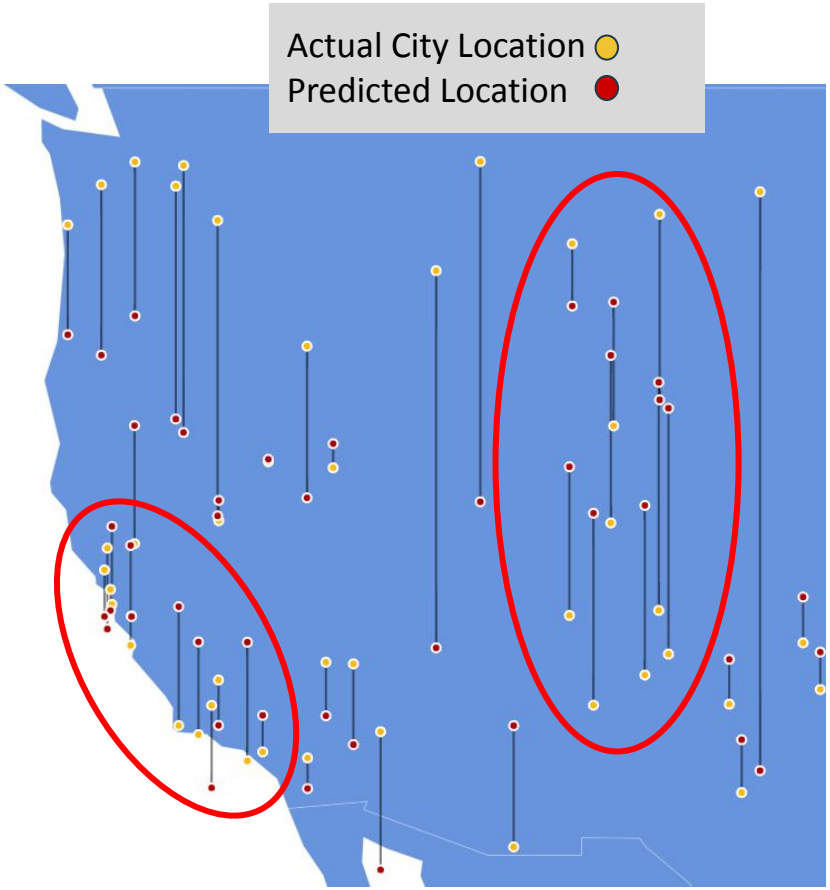
# Conclusions

**Regions with stable weather patterns were predicted more accurately:**

- Southern California, with its unnaturally consistently nice weather, had closer predictions
- The Midwest had our most far off predictions where weather fluctuates between extremes across a year.

**Climate alone is not a sufficient predictor for latitude:**

- Additional factors could include altitude, local flora, terrain, population density, etc.



# Bonus!