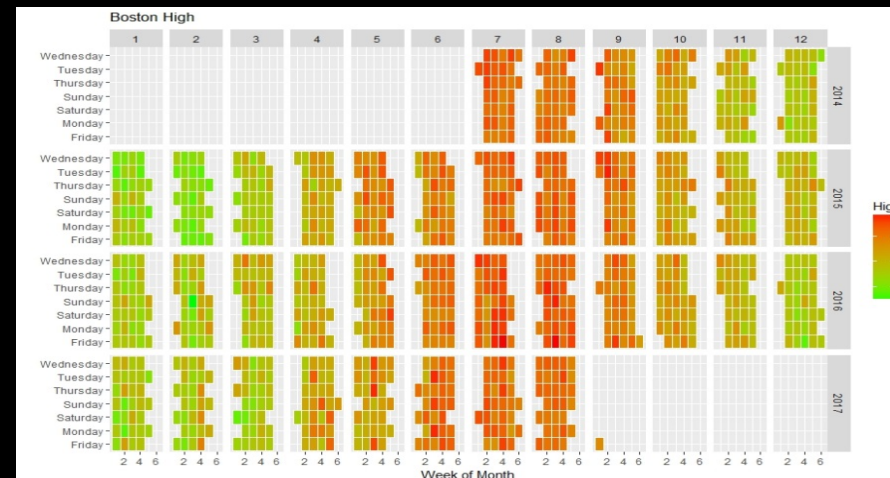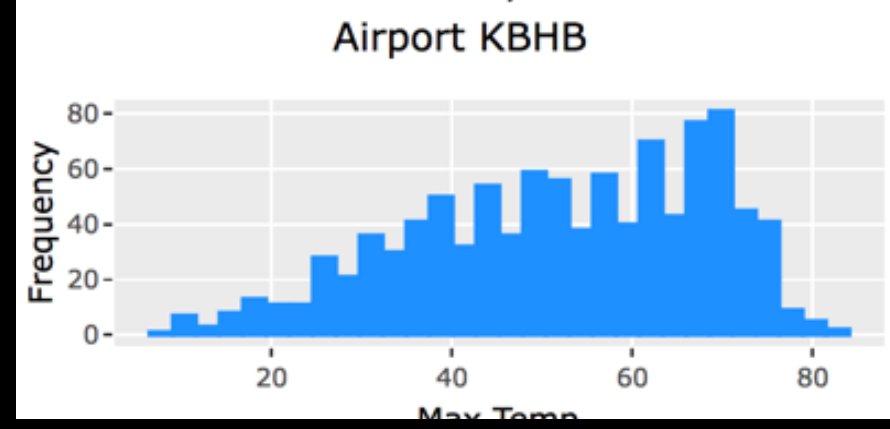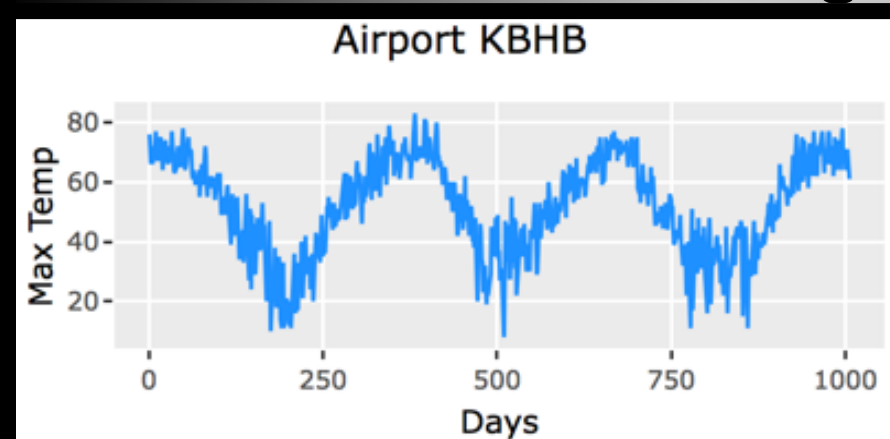# Should you pay attention to the weather forecast?

Dooti Roy*, Gregory Vaughan#, Junxian Geng*, Jianan Hui* and Xiaoyan Liu*

*Boehringer Ingelheim Inc. # Bentley University

TEMPERATURE

Region direct comparison

Data used: Historical
Plot: Time Series
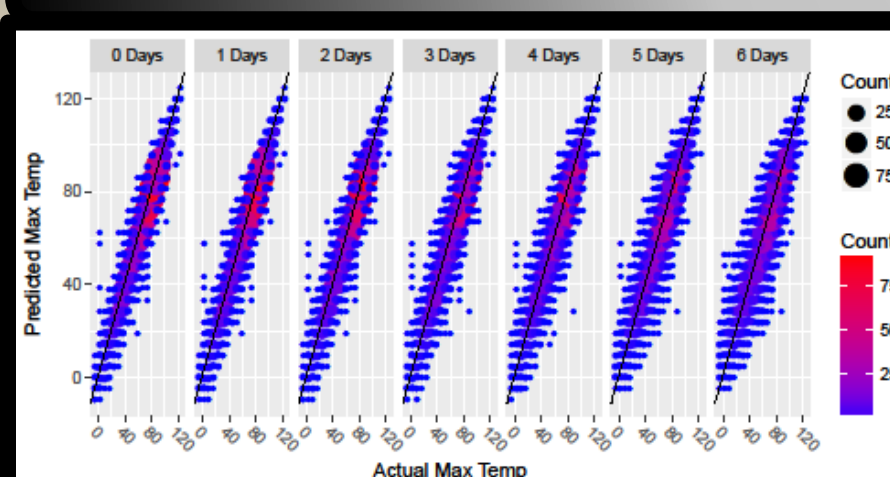**Purpose**: Show seasonal trend

Plot: Histogram
Purpose: Distributional Overview

Plot: Calendar Heat map
Purpose: Visual exploration of all the yearly consistency of seasonal and annual trends

Actual vs. Predicted
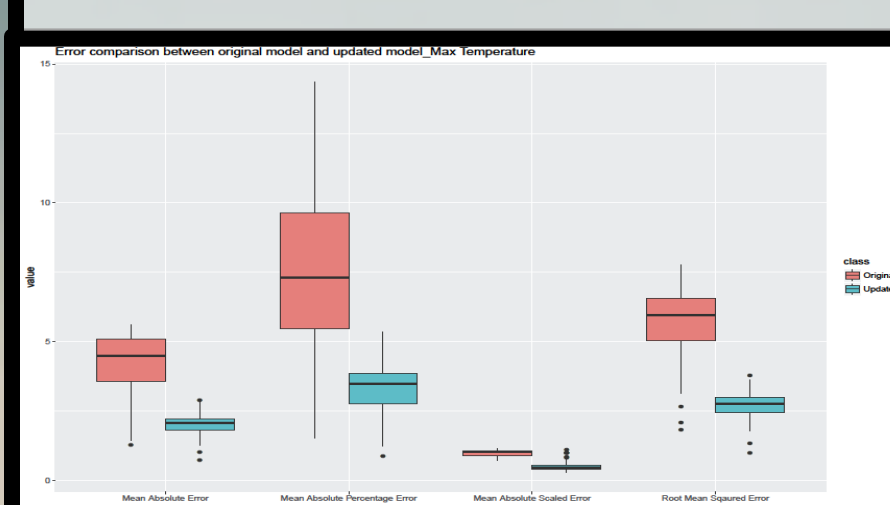

Data used: Historical, Forecast
Plot: Bubble Plot
Purpose: Compare actual with forecast, visualize the bias.

Model Comparison

ARIMA (p,d,q)

Fit separate ARIMA models within each location with and without covariates.

Method:
Merge forecast and historical data
Covariates: dew point, humidity, sea level, visibility miles, wind speed and cloud cover

Error measures (**RMSE, MAE, MAPE and MASE**) of the two models for each location.

The App was developed using R-Shiny and is meant to provide an interactive visual interface of viewing data and the in-built features help the user to discover inconsistencies, seasonal patterns, and geographical distribution of the data.

JSM WEATHER APP
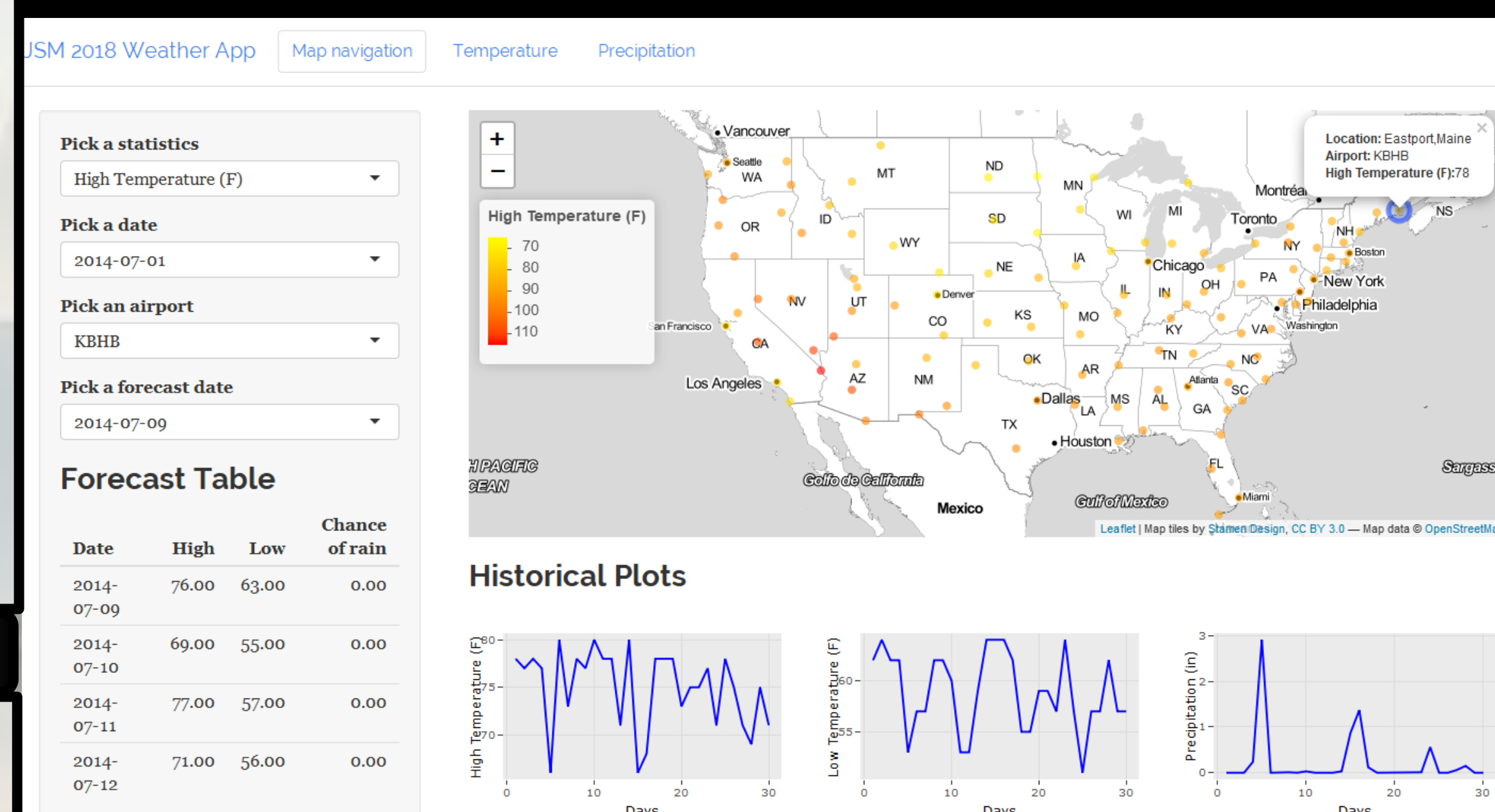
DISCUSSION

- Different locations have slightly different distributions of temperature or rain, but overall patterns are **similar** to each other.
- As the gap between the prediction date and the forecast date **increases**, the forecast grows **worse**.
- **No clear changes** in the distributions of temperature/rain over the 3 years that the data was collected.
- The original forecast tends to **underestimate** the chance of rain.
- The prediction models of temperature and rain can be **improved** by considering additional information.
- Contact: dootiroy@gmail.com, gregory.vaughan@uconn.edu

PRECIPITATION

Region direct comparison
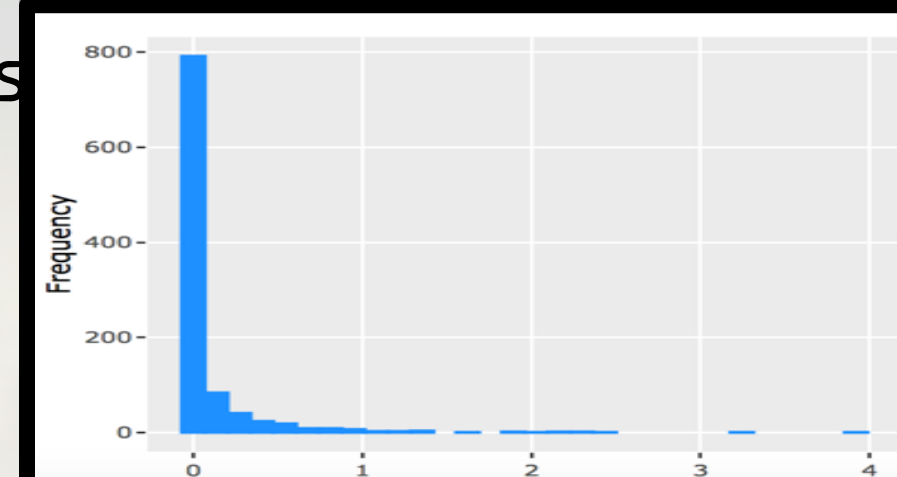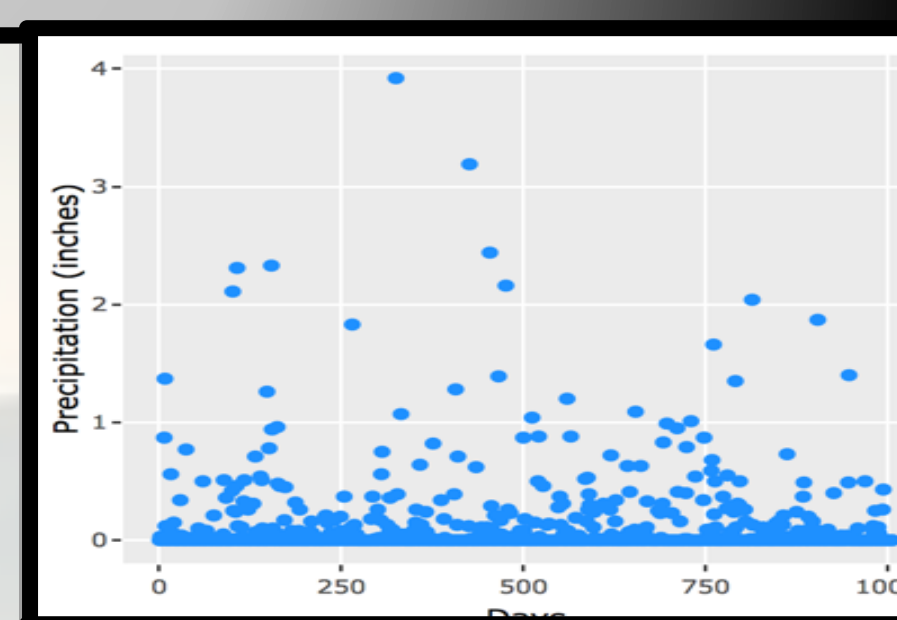
Data used: Historical
Plot: Scatter Plot
Purpose: Overview of the data, outlier detection in a particular location.

Plot: Histogram
Purpose: Distributional Overview, skewness

Historical data shows rainfall in inches.
Forecast data predicts chance of rain.
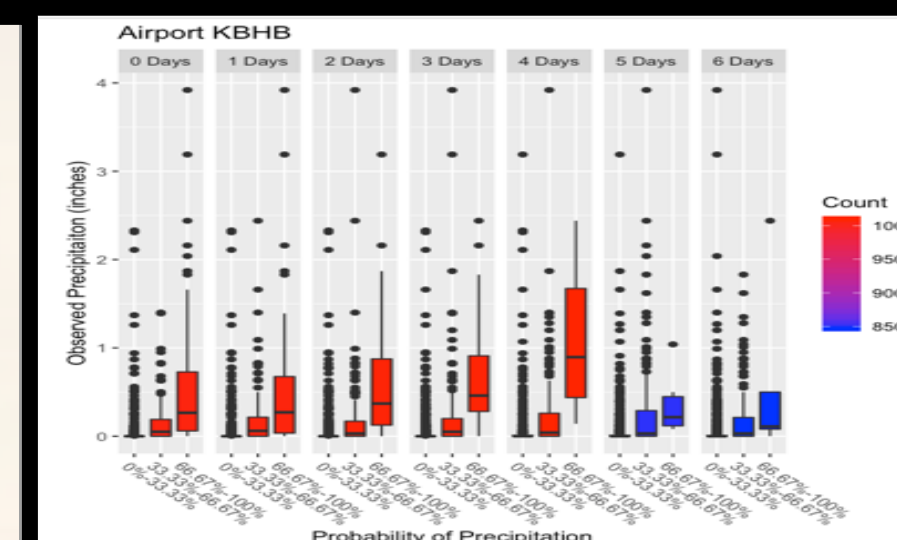Disconnect between variables.

Actual vs. Predicted

Data used: Historical, Forecast
Method: Dichotomized into rain/no rain
Plot: Box Plot
Purpose: Compare actual with forecast, visualize the bias.

Model Comparison

Method:
Logistic regression model based on the original prediction of rain

$$\text{logit}(p) = \alpha + \beta * p' + b1 * \text{mean}_{humidity} + b2 * \text{mean}_{windspeed} + b3 * \text{mean}_{cloudcover} + b4 * \text{mean}_{dewpoint} + b5 * \text{mean}_{sealevel} + b6 * \text{mean}_{visiblemiles}$$

Where p denotes the new prediction for chance of raining and p' denotes the original predicted chance of raining.