

# Statistical Analysis of Temperature Trends in Madison Metro Area

Dave Huwe & Nick Exley

## Motivation & Goal:

Our goal for this project is to analyze temperature changes to see if there is any statistical relationship between temperature and a number of different independent variables. We hope to find out which statistical model can produce the best results for predicting a relationship between temperature and the other independent variables pulled from our data source.

## Required Data:

Over the span of 36 years from 1979 – 2014, The National Centers for Environmental Prediction (NCEP) has collected data on air temperature, precipitation, wind, relative humidity and solar radiation as part of their Climate Forecast System Reanalysis (CFSR).

We will be using data collected by NCEP to use in a couple different statistical models. Our main variables obtained from these data sources are maximum & minimum air temperature, relative humidity, measurement date (standardized by Julian date), wind speed, precipitation and solar radiation level.

## Methods:

We have created four different models for the data:

- Simple/Multiple Linear Regression (using regular date & date adjusted using highest avg max temp)
- Random Forest Regression
- ARIMA
- TBATS

## Results:

Model Type	MSE (Mean Square Error)	R <sup>2</sup>
Simple Linear Regression (Date)		0.0006555
Simple Linear Regression (Adj. date)		0.7731
Multiple Linear Regression (Adj. date & solar)	34.87576	0.7971
Random Forest Regression	20.92942	
ARIMA (1,0,1; daily)	144.9181	
ARIMA (1,0,1; monthly avg)	6.82499	
TBATS (daily)	31.04402	
TBATS (monthly avg)	9.121892	