Climate Trends

An exploration of weather data

Earth’s Changing Climate

Can we use weather data to classify its time period by decade?

Weather data:

- Noisy, high variance
- Notoriously hard to predict
- Changes continuously
- Follows reliable patterns (e.g. daily, seasonal)
- Baseline changes occur gradually

Our assumption: with enough well-recorded data over long time stretches, we can build models to find patterns of changing signal and figure out “when” we are

We collected weather data from multiple sources, building classification models to see if there are *distinguishable* changes from decade to decade

Earth’s Changing Climate

The *climate change hypothesis* in its basic form supposes Earth’s climate changes over time. Thus, climate outputs (like **weather**) may differ in some time periods versus others

The statement is obviously true:

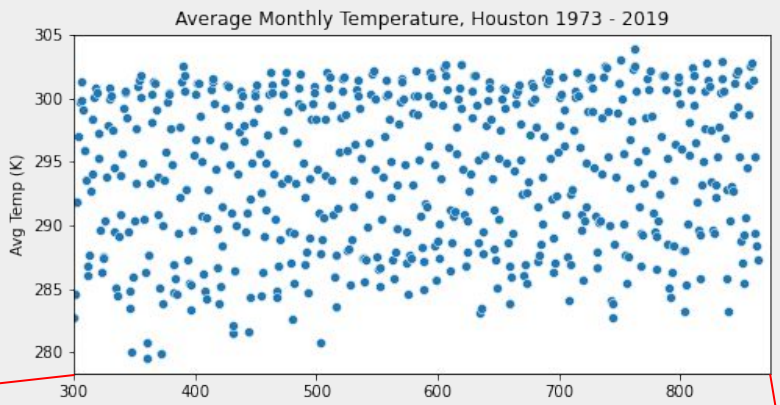
- Daily weather changes are forecast a week ahead
- Summer months warmer than Winter; Spring and Autumn rather mild
- ‘Little ice ages’ can last 300 - 500 years and longer ice ages many thousands

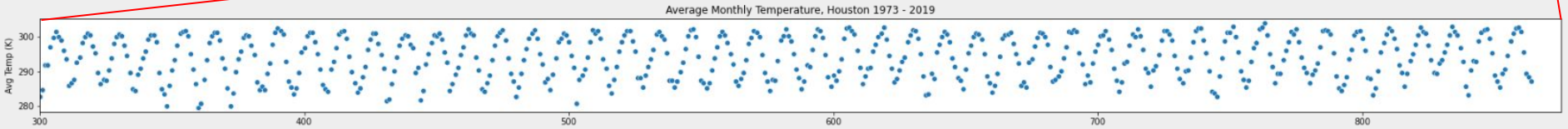
The typical context of a *climate change discussion* involves long timescales, on the order of decades or centuries

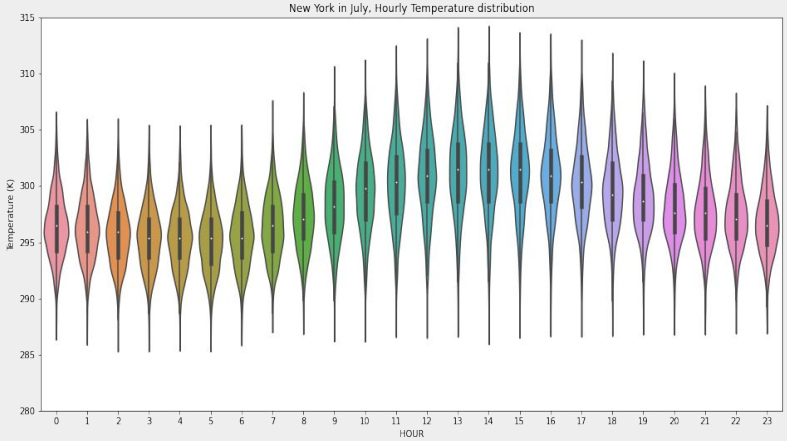
We asked the inverse question, *if we know the weather patterns over a given time period, can we reliably determine which time period?*

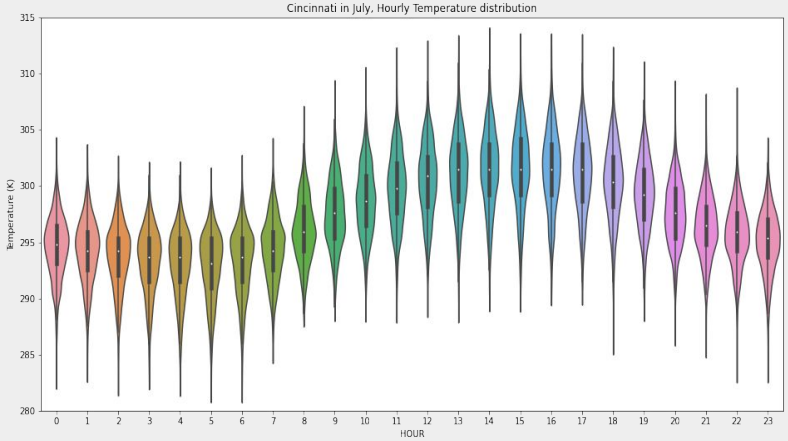
Noise or Signal?

Weather data has unpredictable elements embedded in a reliable pattern



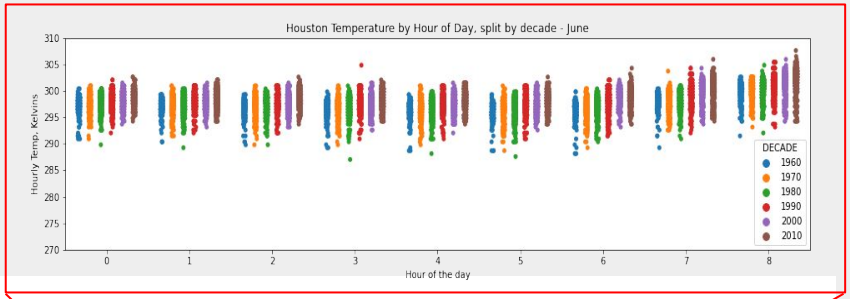


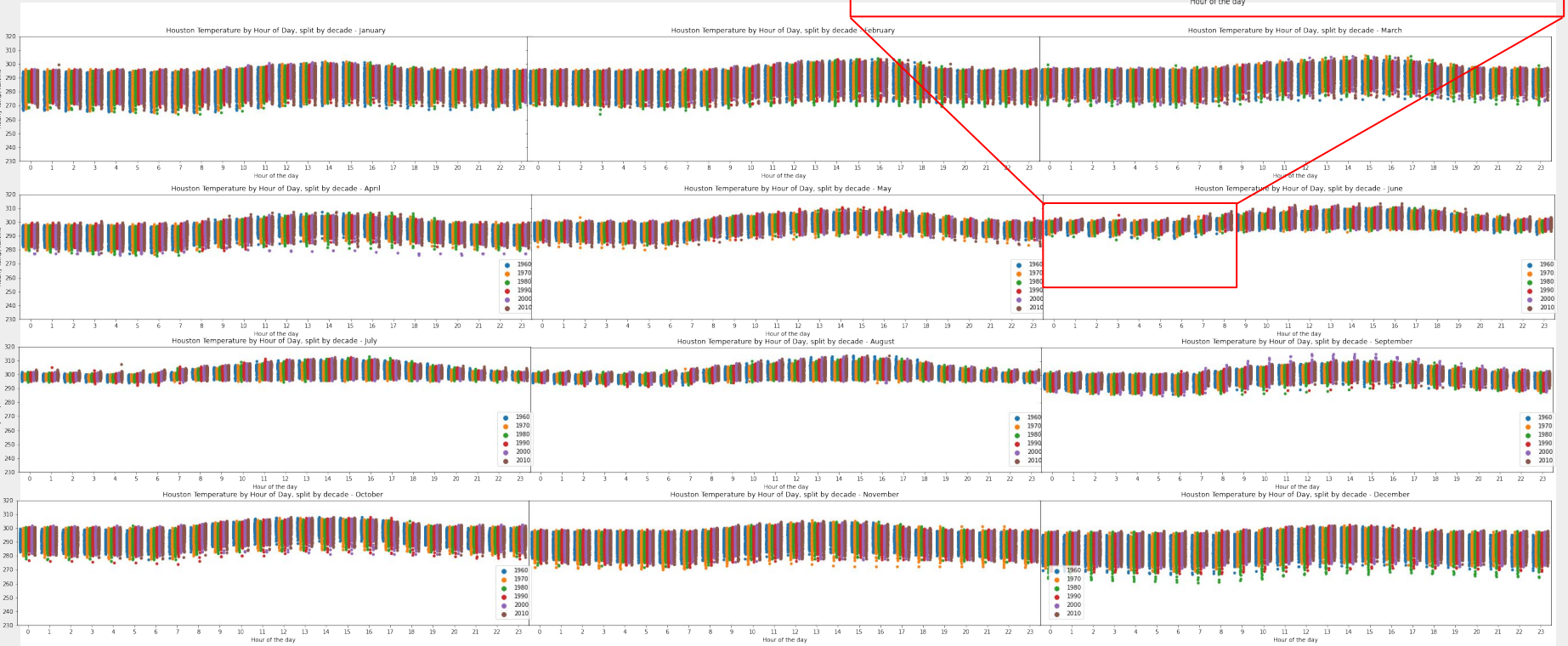




Weather Changes by the Hour

And by month, by season, by decade





Data Sources

Nic:

- DataWorld(originally from Canada Government)
- Monthly weather data of all provinces and available stations **across Canada**, from 1917 - 2017

Stephen:

- NOAA (manual)
- Hourly weather data sets of 4 US **cities** from 1948 - 2020

Trevor:

- NOAA (API)
- Daily weather data from stations located **across the world**

EDA - common themes

Missing Values: Tons of nulls, no way to infer the missing values.

Choice of variable: Average/Minimum/Maximum temps, precipitation, snowfall/snow depth.

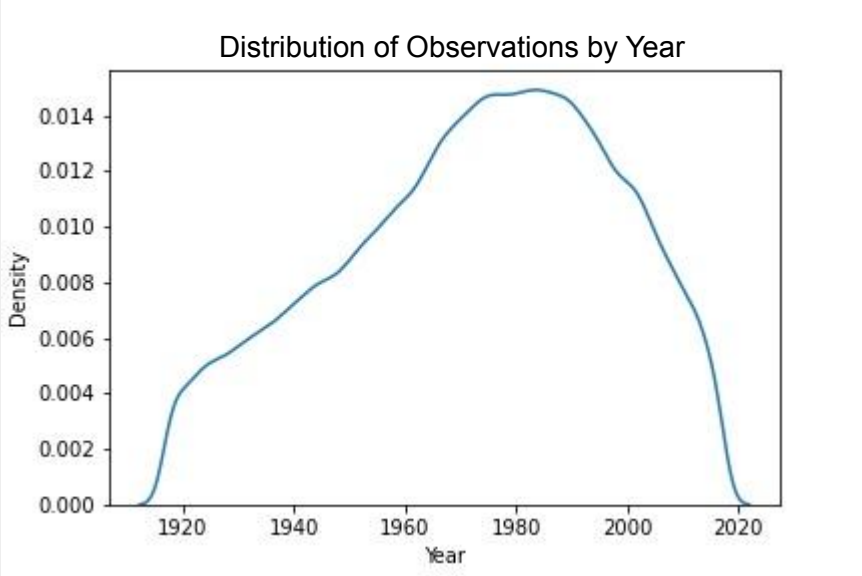
EDA also uncovered summary findings such as rising temperatures over decades, which is a common theme across datasets as well

Evidence of a warming trend across decades

Houston, TX	Cincinnati, KY	Queens, NY	Atlanta, GA	Canada
Avg Max Temp (K) 1950 - 282.05 1960 - 281.79 1970 - 281.50 1980 - 281.80 1990 - 283.06 2000 - 283.33 2010 - 283.55	Avg Max Temp (K) 1950 - 271.33 1960 - 269.75 1970 - 270.13 1980 - 270.26 1990 - 271.10 2000 - 271.93 2010 - 272.25	Avg Max Temp (K) 1950 - 274.93 1960 - 274.98 1970 - 274.29 1980 - 274.57 1990 - 275.82 2000 - 276.03 2010 - 276.67	Avg Max Temp (K) 1950 - 277.41 1960 - 275.60 1970 - 276.29 1980 - 276.89 1990 - 277.59 2000 - 277.17 2010 - 278.88	Avg temp (K) 1950 - 291.47 1960 - 291.51 1970 - 291.57 1980 - 291.89 1990 - 291.70 2000 - 291.41 2010 - 291.11
Avg temp (K) 1950 - 293.74 1960 - 293.40 1970 - 293.22 1980 - 293.57 1990 - 294.20 2000 - 294.41 2010 - 294.77	Avg temp (K) 1950 - 285.42 1960 - 284.72 1970 - 284.65 1980 - 285.10 1990 - 285.42 2000 - 285.60 2010 - 286.06	Avg temp (K) 1950 - 285.80 1960 - 285.24 1970 - 285.36 1980 - 285.59 1990 - 286.36 2000 - 286.27 2010 - 286.74	Avg temp (K) 1950 - 289.72 1960 - 288.39 1970 - 288.88 1980 - 289.67 1990 - 290.13 2000 - 289.85 2010 - 290.77	Avg temp (K) 1950 - 276.65 1960 - 278.69 1970 - 276.80 1980 - 277.24 1990 - 277.27 2000 - 276.89 2010 - 277.07
Avg Min Temp (K) 1950 - 282.06 1960 - 281.80 1970 - 281.50 1980 - 281.80 1990 - 283.07 2000 - 283.34 2010 - 283.55	Avg Min Temp (K) 1950 - 271.33 1960 - 269.75 1970 - 270.13 1980 - 270.26 1990 - 271.10 2000 - 271.93 2010 - 272.25	Min temp (K) avg 1950 - 274.937 1960 - 274.984 1970 - 274.294 1980 - 274.575 1990 - 275.825 2000 - 276.034 2010 - 276.677	Avg Min temp (K) 1950 - 277.41 1960 - 275.61 1970 - 276.30 1980 - 276.89 1990 - 277.60 2000 - 277.18 2010 - 278.89	Avg temp (K) 1950 - 261.70 1960 - 262.02 1970 - 262.00 1980 - 262.48 1990 - 262.88 2000 - 262.81 2010 - 263.26

Canada Dataset

- Taken from different weather stations across Canada
- Over 850,000 data points used
- Data was broken up by month
- 16 features used
- Some features included
  - Average monthly temp
  - Monthly low
  - Monthly high
  - Precipitation
  - Snowfall
- Most data came from 1960 or later



NOAA API -- GHCND Dataset

GHCN-Daily: Global Historical Climatology Network + Daily

Stations: 118,486 weather stations all around the world

Date range: 1763 - Now

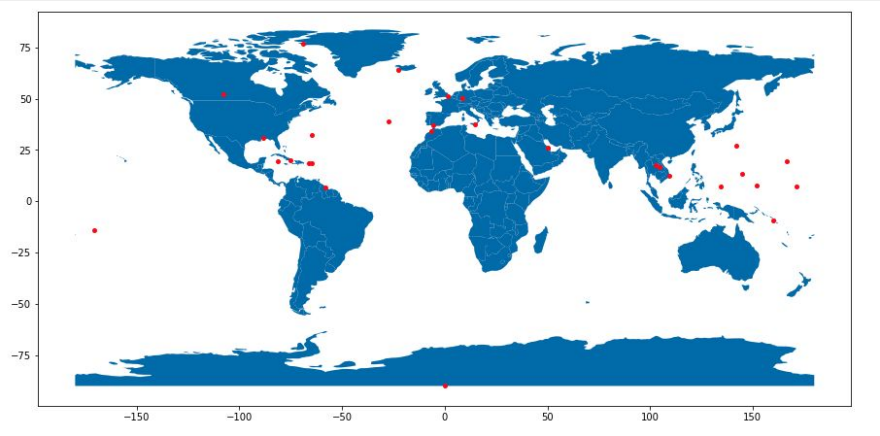
Values Provided: 137 variables

Data from NOAA API

“Only” 1676 stations provide all variables we want.

1676 stations are from 29 countries, and mostly are US and Canada.

Randomly selected one station from each country, then using API to get their data from 1979 - 1999



Modeling for Decade

We categorized decade (1940, 1950, etc) as our target variable for every model, using multiple models to reflect the strengths/weaknesses of different algorithms

Dataset	Log.Regression	RandomForest
Atlanta, GA	train score: 0.404 test score: 0.405	train score: 1.000 test score: 0.813
Cincinnati, KY	train score: 0.313 test score: 0.312	train score: 1.000 test score: 0.794
Queens, NY	train score: 0.326 test score: 0.326	train score: 1.000 test score: 0.796
Houston, TX	train score: 0.230 test score: 0.230	train score: 0.999 test score: 0.837
Cities combined	train score: 0.424 test score: 0.423	train score: 0.999 test score: 0.809

These datasets have a baseline accuracy score of 15% - 20%

600K hourly observations per city

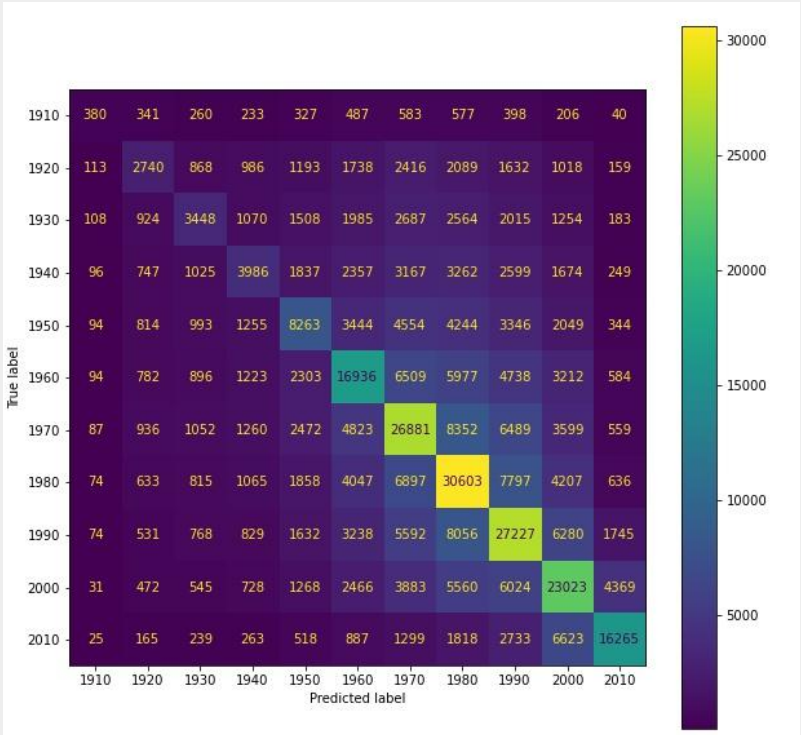
16 features, including:

- Month and Hour of the day
- Temperature
- Humidity

Data from 1948 - 2019

Model Using Monthly Canadian Data

- Baseline accuracy of 15%
- 16 features, including:
  - Average monthly temp
  - Monthly low / high
  - Precipitation
  - Snowfall
- 3 hidden layers, 512 nodes each and 1 output layer
- Neural net was able to get a train accuracy of 48% and a test accuracy of 42%



The ‘wrong’ model predictions were still close to the true decade. And the model was more likely to predict its true decade compared to a different one

Predicting Decade Around The Globe

- Neural net model, we identified Decade using global data from the NOAA
- Using a Decade prediction model that was built on Canadian weather we were still able to achieve accuracy of 38% on global data
    - This accuracy on the data from around the world is extremely close to the original testing data from our model accuracy of 42%.
- This could imply the same country-wide weather changes identified in Canada are also happening globally on the timescale of many decades

Summary of our Model Findings

We created different models from different datasets because we wanted to look for changing climate from many perspectives

Our models:

- Produced results better than baseline for predicting DECADE as a target
  - We currently lack interpretability to determine the causes for these changes in weather
- The projected focused on whether there is a detectable difference in weather over the course of decades, NOT what is causing the change
- Shows the effects happening on a country wide scale are also happening on a global scale

Closing Thoughts

Further explorations could include analyzing what are factors are causing the change, and the downstream effects of these changes. E.g. How does this affect growing food, our food supply, other economic impacts.

Exploring the data and the predictions the model was able to make strongly suggest there is change on a decade to decade basis. Not only on a city level but also on a country and global level also.

Thank you

- Chang, Trevor
- Reagin, Stephen
- Steele, Nicholas