want to frame as an effect of climate change on epigenomic mechanisms/ disease

(Disparate) Data Sources on open AWS

miRNA expression data (CPACT2 and TARGET)

DNA methylation data (array, WGBS) (CPACT2 and TARGET, TCGA, ENCODE)

exposome data set (TBD)

NOAA Joint Polar Satellite System (JPSS)

has ozone measurements, pollutant info, ambient temperature/ climate data

Bakeoff of BRNNs, compare performance of a few, time allowing

Clean and impute missing data using statistical models (ex. MICE)

subsect 80% of data for training

TRAIN ML model: Bidirectional Recurrent Neural Networks (RNNs, LSTMs, GRUs)

TEST ML model: Bidirectional Recurrent Neural Networks (select one of RNNs, LSTMs, GRUs)

subsect 20% of data for testing

Experiment: What effect does (exposome variable) have on miRNA and methylation activity? Are they near any specific genes. what pathways are those involved in?

Interpret Results (Bioinformatics Analyses, GSEA, GO...)

Visualize Results (time allowing)

Package model into a standarized reusable module

Publish a Python Package!!!

Ideas and examples