Analyzing Weather Prediction Abilities

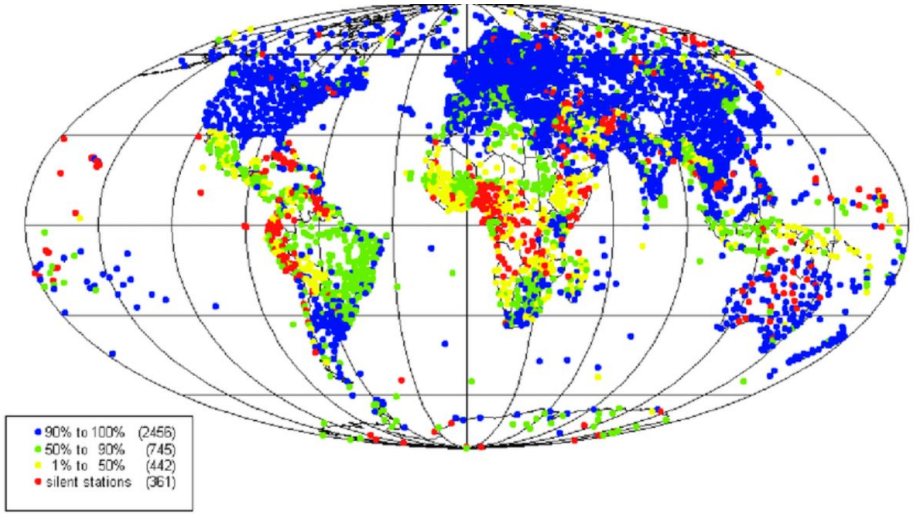
Group K

Introduction

- The National Oceanic and Atmospheric Administration runs the National Center for Environmental Information, which records weather information
- Integrated Surface dataset: over 35,000 stations, hourly measurement on data like temperature, winds, clouds, and waves
- Data goes back to 1901 (for some stations)
- Stations not equally distributed

Background

- We want to understand patterns in this dataset, and evaluate bias in this system
- We will use GIS to map these findings

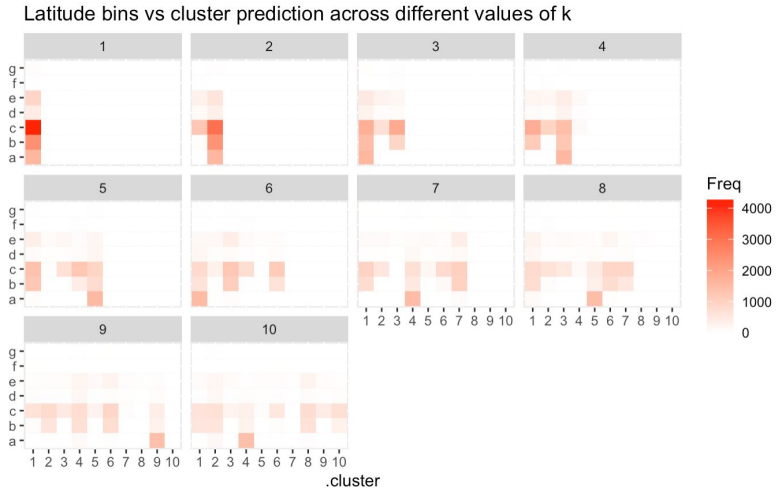


Process

- Data is available through AWS
- Data is very large, so can only take a subset
- Local versus online datasets
- Numerical vs nominal variables, and cleaning the data by removing N/As
- Exploratory clustering findings
- Mapping with ArcGIS

Clustering Cleaning and Outcomes

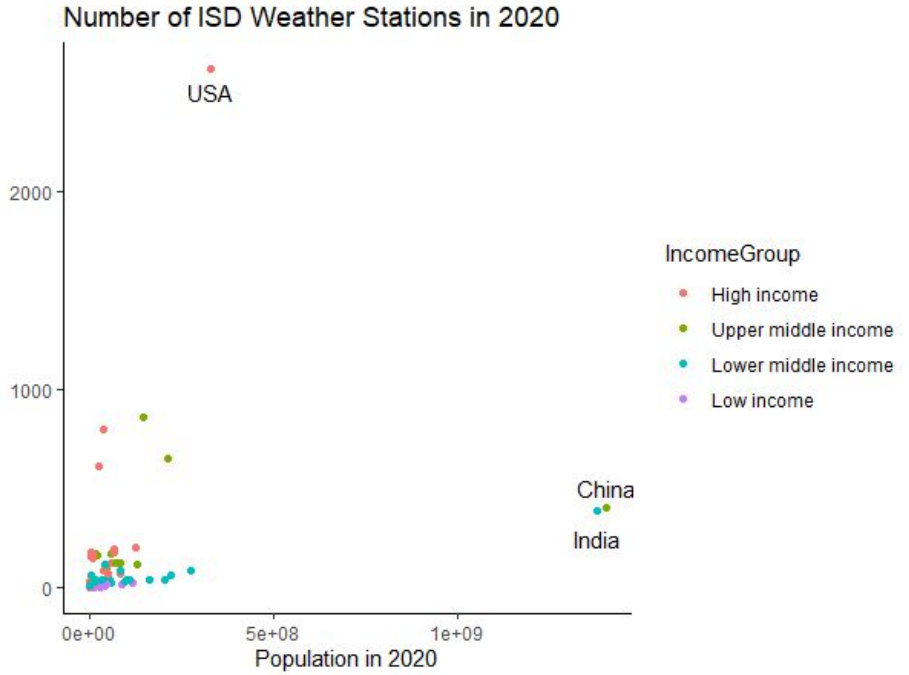
For this data set we wanted to find out whether or not the variance of the temperature throughout the year for each station was a good predictor for the distance away from the equator. We chose to use clustering to approach this problem.

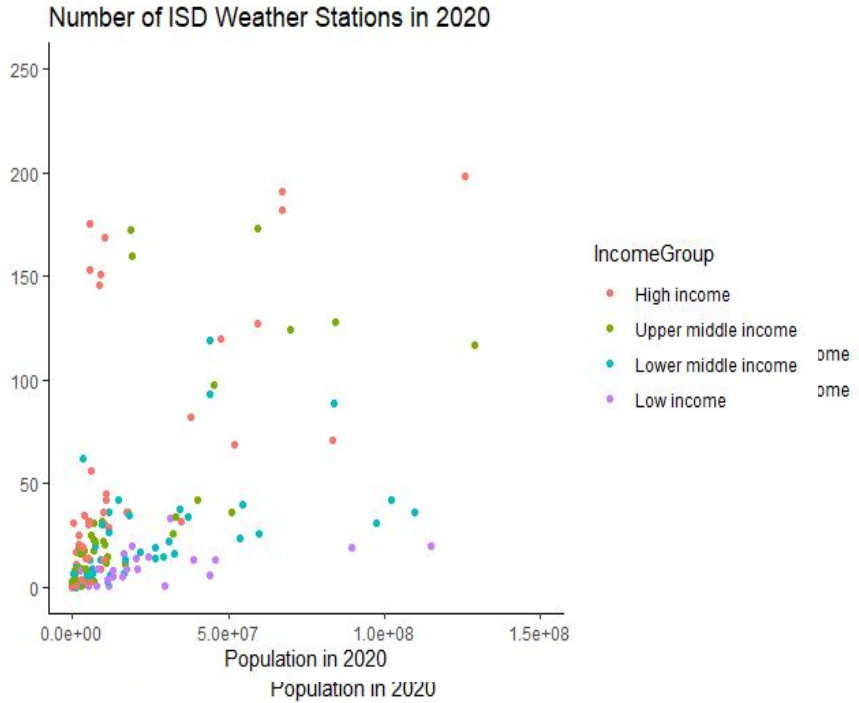


GIS Findings

- GIS = Geographical Information Systems
- Reverse geo-coding: how do we move from coordinates to countries?
- How does looking at country boundaries flatten our data?

Addressing Cartographical Bias: Population and Wealth





Mean # Weather Stations by Income Group

Income level	Mean # of weather stations (+/- standard deviation)
High	112.207 +/- 360.969
Upper middle	67.135 +/- 156.571
Lower middle	31.453 +/- 56.073
Low	9.440 +/- 7.990

Findings, Reflections, and Discussion Questions

- How does bias in different stations influence or relate to disparities in climate modeling?
- How much of this bias is explained by factors like inaccessible terrain?
- What other options exist? (inexpensive, fine-grain weather prediction systems)
- What are areas for further research?