

Proposal 1

Motivation

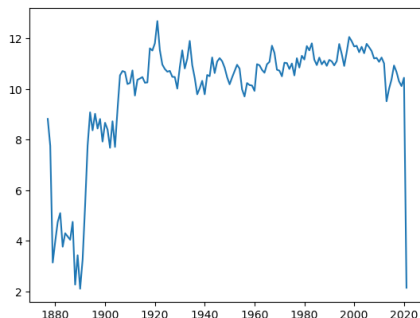
In our analysis, we try to understand to what extent the relationship between temperature and other climate indicators reflects an increase in global warming. The Global Historical Climatology Network - Daily (GHCN-Daily) dataset which we use contains a large collection of data related to climate indicators like snow, rainfall, humidity, and cloud cover. The dataset is continuously updated, giving us a large range of dates, spanning hundreds of years up to the present day. The data is collected by around 90,000 land based weather stations, with around 25,000 being continuously updated.

We want to use this data to understand long term trends in weather as it relates to changing temperature across the world, based on different weather station observations. We surmise that in places where there is a correlation between rising temperatures and reduced snow, rainfall, etc. Finding correlations between these variables would allow us to understand the extent of the impact temperature changes are actually caused by global warming. Furthermore, this would allow us to see the effects of global warming and predict future trends if we see reduction in the climate indicators mentioned above. Having the ability to detect and predict global warming trends is a powerful tool for governments deciding on climate policy, for businesses making decisions on where to invest resources and build, and communities deciding where a safe place to live long term is.

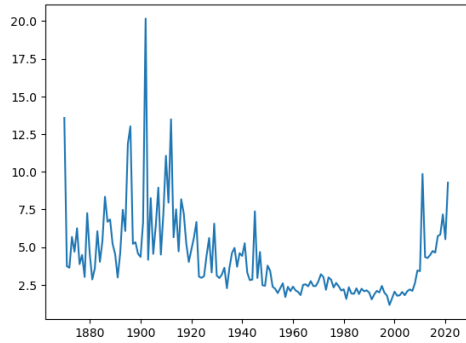
Dataset

Below are some visualizations we generated as part of our exploratory data analysis:

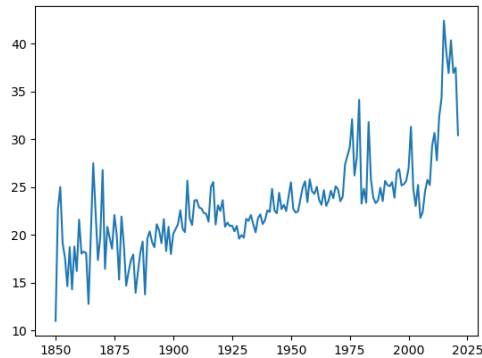
Average temperature across all stations, per year:



Average snowfall across all stations, per year:



Average precipitation across all stations, per year:



Methodology

In our project, we want to understand the relationship between temperature and certain climate indicators. We can perform correlation calculations between these variables to understand their relationships. Because of the large size of our data, we can also attempt to understand whether sampling weather station locations allows us to use the central limit theorem to see if temperatures will converge to a specific value. This is valuable in giving us a fundamental relationship between our key indicators and independent variable temperature.

We also want to understand how significant of an impact the relationship between changing temperatures and indicators is on global warming and climate change. We can use hypothesis testing and related methods to calculate the significance of the changes occurring in temperature. This would be a powerful step for us in furthering our understanding of the correlation between temperature and climate indicators.

Furthermore, we want to make predictions about how this temperature data and climate indicator data will change in the future. We can use linear regression to extrapolate on our observations and predict how snowfall or precipitation are affected by a certain value of temperature which we haven't seen yet. This would be significant because we would have estimates of how climate will be affected in the coming years, in different locations around the world, based on temperature values.