

SI 618: Project Proposal
Exploratory Data Analysis on Weather of U.S.

Jun Wang wanjun@umich.edu 52383890

Introduction

Michigan's winter is coming. Getting accustomed to the relative mild climate in China, I still cannot live well together with the severe cold. Hence, I want to perform the exploratory data analysis on the weather of U.S. in some specific locations in some time.

The Data

The datasets are pulled from the API provided by NOAA (National Centers for Environmental Information). The weather data from weather stations is available at the specific location of the weather station across all the country, and it is often useful to have estimates of daily or hourly weather aggregated on a larger spatial level, so that it can be particularly useful to be able to pull time series of weather by county. The base URL of API is <http://www.ncdc.noaa.gov/cdo-web/api/v2/{endpoint}>, and the endpoint can be the specific category of the weather data we request, such as datasets, locations and stations. Here, the R packages, *noaa* and *countyweather* provide useful and integrated weather data access methods, the data we pull from the API is in JSON format. The main variables the datasets include are *time* in date format, *positions* in character format, *precipitation*, *wind speed*, *snow fall*, *snow depth*, *maximum temperature*, *minimum temperature* in numeric format and etc..

Exploratory Questions and Methods

1. How does the weather like in Michigan's winter, especially nearby Ann Arbor?

To look at this, I will collect the weather data including snow fall, maximum temperature and minimum temperature of the county near Ann Arbor in a winter. To visualize this, I will use the line plot and the spectrum plot.

2. When the hurricane stuck, how did the weather like in Florida?

To analyze this, I will collect the weather data including precipitation, wind speed for a county station in Florida when the hurricane stuck. To visualize and analyze this, I will use the line plot, the histogram and maybe the geographical map with the weather data.

3. The annual weather analysis on the Silicon Valley in California.

For this, I want to analysis the precipitation and the temperature in California in a year. I may choose a county station near the Silicon Valley then collect the annual weather data for it. The pie chart, the line plot, the histogram may be used in the visualization.

4. The difference between the weather of the urban area and the rural area in New York.

For this, I will choose a county station in the urban area and a county station in the rural area, then pull the weather data of these two stations. The precipitation and the other weather variables will be analyzed. To perform this, the comparison between the line plots, the histograms, the pie charts will be visualized.