

Weather Project Report

Seattle is often mentioned as a rainy city in the United States, and this project was carried out to clearly observe the amount of rainfall in Seattle compared to New York. I want to compare the rainfall between these two cities from the beginning of 2018 to the end of 2022. This project will help me answer the question: “Does Seattle really rain more than New York?”

Before starting the data analysis, I will need to clarify how to compare which city has more rain based on two criteria:

1. **Frequency:** how many rainy days there are in each city.
2. **Amount:** the average amount of rain that falls per day in each city.

Data Sources:

This analysis used the daily precipitation records from NOAA’s Climate Data Online Daily Summaries dataset. The data were collected for:

- Seattle, WA
 - o Link: <https://www.ncei.noaa.gov/cdo-web/datasets/GHCND/stations/GHCND:US1WAKG0225/detail>
- New York, NY
 - o Link: <https://www.ncei.noaa.gov/cdo-web/datasets/GHCND/stations/GHCND:USC00308721/detail>

The cleaned dataset produced for this project to do analysis is available in the repository as `data/clean_seattle_newyork_weather.csv`. The cleaned dataset was produced by the notebook located in `Weather_Data_Preparation_SEA_NYC.ipynb`. The main process is in `Weather_Data_Processing_SEA_NYC.ipynb` to perform statistical analysis and visualization.

Method:

1. Preparation
 - Loaded data from a CSV file into a DataFrame using the pandas library.
 - Converted the date column to the correct data type.
 - Combined the two datasets into a single DataFrame based on the date column, selecting only the date and precipitation columns.
 - Reformat the dataset into tidy format with columns: date, city, and precipitation.
 - Added an additional variable, `day_of_year`, to each observation, then calculated the mean precipitation for each `day_of_year` in Seattle.
 - Imputed missing values using the corresponding `day_of_year` mean.
2. Exploratory Analysis

- Calculated the mean of precipitation for each city.
 - Extract month to allow grouping and visualization easier.
 - Plot and visualized the average rainfall amounts and proportion of rainy days.
3. Statistical Tests
- Used t-tests to get the significant differences average daily rainfall between Seattle and New York each month.
 - Used z-test to get the significant differences of the proportion of rainy days between Seattle and New York each month.

Conclusion:

1. Rainy-Day Frequency

Figure 1 show the monthly rainfall, calculated only based on the days with rain in Seattle and New York, indicates that Seattle has higher rainfall than New York in most seasons, except for summer. This suggests that Seattle generally receives more rain.

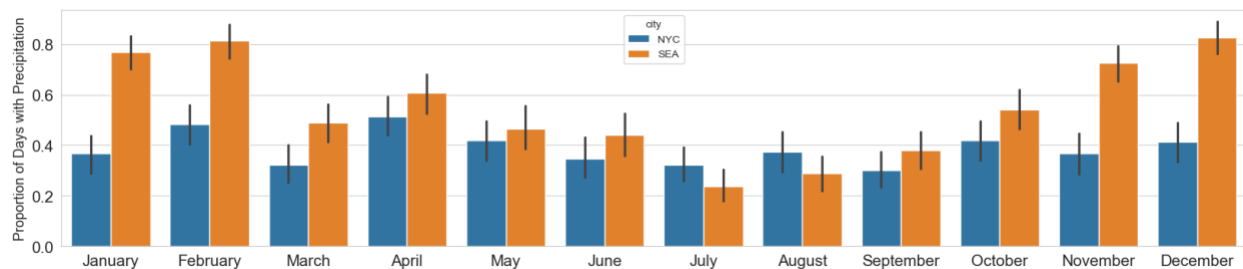


Figure 1. Comparison of monthly proportion of rainy days, Seattle vs. New York

2. Rainfall Amount per Day

Although Seattle has more rainy days, New York is often recorded as having heavier rainstorms. There are several days in New York with rainfall amounts more than twice that of Seattle. As a result, the figure 2 shows that New York has relatively consistent rainfall across the months, and particularly more than double Seattle's rainfall in the summer. This indicates that New York has fewer rainy days but tends to receive heavier rainfall.

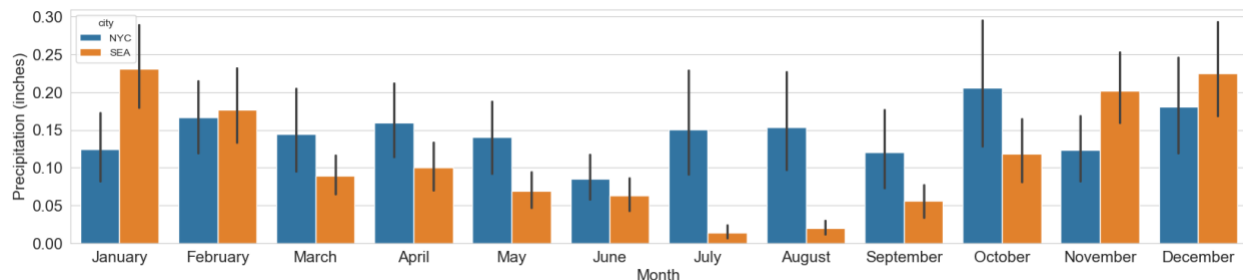


Figure 2. Comparison of daily precipitation by month for Seattle and New York