**Introduction**

The purpose of this project is to compare the precipitation levels of Seattle, Washington, with that of Detroit, Michigan, to determine whether it rains more in Seattle than in Detroit. The choice of Detroit as a comparison city stemmed from personal curiosity. Having lived in Windsor, Ontario, Canada—a border city neighboring Detroit—for eight years, I observed that the area seemed to receive a moderate amount of rainfall. This prompted me to compare it with Seattle, a city well-known for its frequent rainfall, especially during the winter months. The data used for this analysis were obtained from the National Centers for Environmental Information (NCEI) archive.

**Data Description**

The data used in this analysis were obtained from the National Oceanic and Atmospheric Administration (NOAA), a U.S. government agency responsible for collecting and maintaining environmental data. Specifically, daily precipitation data for both cities were sourced from the National Centers for Environmental Information (https://www.ncei.noaa.gov/cdo-web/search). Using the search tool provided on the website, we selected the Daily Summaries dataset, which includes daily precipitation information for each city. The selected period spans from January 1, 2018, to December 31, 2022, providing a sufficient sample to draw conclusions about general rainfall patterns.

During the data collection process, we selected weather stations within each city that offered the best data coverage over the chosen time period. Both selected stations provided more than 95% data coverage, ensuring that our dataset was comprehensive and reliable.

We also selected the relevant attributes for our analysis. Initially, most of the default attributes were included to ensure completeness. During the data cleaning stage, we filtered the dataset to retain only the necessary attributes—specifically, date and precipitation. It was important to download the data in .csv (Comma-Separated Values) format, as this is one of the simplest and most common formats for storing tabular data (rows and columns).

The CSV files for both cities were then uploaded to my GitHub repository for storage. From there, the raw data files were imported into a Jupyter Notebook to begin the data cleaning and analysis process.

Here is the link to where our data is saved:

Seattle: https://raw.githubusercontent.com/kembaoak/weather/refs/heads/main/data/seattle_rain.csv

Detroit: https://raw.githubusercontent.com/kembaoak/weather/refs/heads/main/data/detroit_rain.csv

**Analysis**

After the data was successfully imported as a data frame, it was cleaned to ensure there were no missing values. Missing values were then replaced with the mean precipitation for that particular date across the five years, as leaving them blank would have resulted in incomplete or invalid analysis. It was also important to ensure that the date and precipitation columns were in the correct formats (date in datetime format and precipitation in inches). The data frames for both cities were merged to allow a more direct comparison of daily precipitation levels. The cleaned data was then exported for further analysis.

In the exploratory data analysis, we used descriptive statistics, a pie chart, and a bar chart to draw inferences from our data. Descriptive statistics from Figure 1.1 were used to answer questions such as how many days of rainfall data were recorded, which city receives more rain on a daily average, how much daily rainfall varies in each city, and how the average rainfall in Seattle compares to that in Detroit. The pie chart in Figure 1.2 provides a visual comparison of the composition of total precipitation for Detroit and Seattle over the five-year period, helping us better understand the overall distribution of rainfall. A bar chart was also used to visualize the monthly proportion of days with precipitation, which allows us to identify seasonal trends and compare rainfall patterns across months.

**Results**

| city | precipitation count | mean | std | min | 25% | 50% | 75% | max |
|------|------|------|------|------|------|------|------|------|
| DTW | 1826.0 | 0.085308 | 0.246427 | 0.0 | 0.0 | 0.00 | 0.04 | 2.6 |
| SEA | 1826.0 | 0.113270 | 0.240516 | 0.0 | 0.0 | 0.01 | 0.12 | 2.6 |

Figure 1.1 - Descriptive statistics table

The data from Figure 1.2 shows that there are an equal number of records for both cities. Seattle has a higher daily mean precipitation than Detroit, with values of 0.113 inches for SEA and 0.085 inches for DTW. The minimum values indicate that both cities experience days without rain, as both have 0.0 inches recorded. About 25% of days in both cities have no rain(1st quartile). Approximately 25% of days in both cities have no rain (1st quartile). However, around 50% of days in Detroit are dry, while in Seattle, 50% of days receive about 0.01 inches of precipitation. The 75th percentile indicates that Detroit receives about 0.04 inches of rainfall on wetter days, compared to 0.12 inches in Seattle. The maximum daily precipitation is the same for both cities, at 2.6 inches.

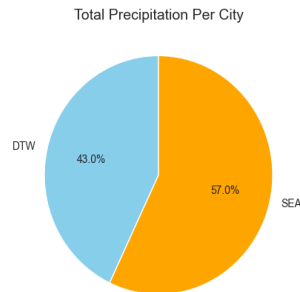Total Precipitation Per City



Figure 1.2 - Pie chart showing total precipitation per city

Another important visual is the pie chart, which shows that Seattle accounts for 57% of the total precipitation compared to Detroit's 43%.
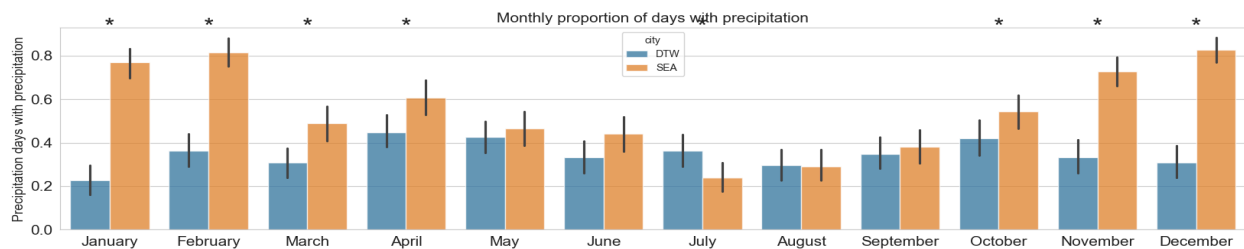


Figure 1.3 - Bar chart presenting the monthly proportion of days with precipitation

Figure 1.3 illustrates the monthly proportion of days with precipitation in Seattle and Detroit, with asterisks indicating months where statistical tests revealed significant differences between the two cities.

Some key observations include:
The data shows that there are distinct seasonal patterns between the two cities. From December through March, Seattle consistently demonstrates a higher proportion of rain days compared to Detroit. The different climates experienced play a major role, as Seattle's climate brings more frequent winter rainfall, while Detroit's continental climate results in more winter precipitation in the form of snow, which is not accounted for in the precipitation measurement.

The marked months (indicated by asterisks) highlight where the difference in precipitation frequency is statistically significant. During the summer months, Seattle experiences significantly less rainfall than in winter. In Detroit, the proportion of rainy days appears relatively consistent, with values ranging between 0.2 and 0.4.

**Conclusion**
The analysis of our precipitation data for Seattle and Detroit over a five-year period reveals clear differences in rainfall patterns between the two cities. Seattle receives more total and average daily rainfall than Detroit, as reflected in the data provided in both descriptive statistics and the pie chart, which shows Seattle accounts for 57% of total precipitation. The bar chart also indicates a seasonal trend: Seattle experiences more frequent rainfall during the winter months, while Detroit has lower precipitation, which could be due to the different climates that results in snow during the winter months. The monthly analysis further highlights statistically significant differences in precipitation frequency for certain months, particularly in winter. Overall, Seattle's climate results in more consistent and higher rainfall throughout the year, whereas Detroit experiences more variable precipitation, with relatively equal proportions of rainy days in the summer. These insights provide a comprehensive understanding of the rainfall dynamics in both cities. Finally, the data cleaning and imputation of missing values with mean precipitation could slightly bias the results. Despite these limitations, the analysis provides meaningful insights into the comparative rainfall patterns and seasonal trends in Seattle and Detroit.