Earth 105 Climate Report

Region - Ethiopian Highlands

Team - Ram Bharadwaj Gandikota, Lee Solbin, Gustavo Alejandro, Spencer Thomas

Index:

Introduction:

Our team was assigned the task of exploring the rainfall data of the Ethiopian Highlands and analyze to make educated inferences about the area's rainfall patterns. Raw historic rainfall data (in mm) for Addis Abada, Chisimiao, Dire Dawa, Gambela, Gondar, Hosaina, Mogadishu and Wonji were provided for the analysis. We focused our efforts on studying the yearly and monthly averages of the provided raw data after cleaning up to account for missing data to figure out if there is any clear indicator to these rainfall data such as altitude or distance from a large water body, mainly the Red Sea and Arabian Sea. In the end, we also made use of Python and it's libraries to create a program that is able to take in historic data of any region and predict the next 12 months of rainfall data.

Altitude [meters]:
- Chisimiao:
  - 6 m
- Mogadishu:
  - 61 m
- Gambela:
  - 526 m
- Dire Dawa:
  - 1,180 m
- Wonji:
  - 1,588 m
- Gondar:
  - 2,111 m
- Hosaina:
  - 2,177 m
- Addis Ababa:
  - 2,294 m

Distance from Sea (Red Sea and Arabian) [kilometers]:
- Addis Ababa:
  - Red Sea: 1252 km
  - Arabian Sea: 2816 km
- Chisimiao:
  - Red Sea: 2335 km
  - Arabian Sea: 2782 km
- Dire Dawa:
  - Red Sea: 1239 km
  - Arabian Sea: 2469 km
- Gambela:
  - Red Sea: 1402 km
  - Arabian Sea: 3281 km
- Gondar:
  - Red Sea: 861 km

    - Arabian Sea: 2917 km
- Hosaina:
    - Red Sea: 1416 km
    - Arabian Sea: 2944 km
- Mogadishu:
    - Red Sea: 2157 km
    - Arabian Sea: 2381 km
- Wonji:
    - Red Sea: 1317 km
    - Arabian Sea: 2770 km

Classification:

Based on the collected data, we can classify the regions using both altitude and distance from sea. To do this, we correlate the minimum distance from a sea (Red Sea or Arabian Sea) and the altitude in ascending order. Then these can be classified as lowest altitude and highest distance from sea to highest altitude and lowest distance from sea as these two factors can have a drastic effect on the climate.

First we have to decide which is sea between the Arabian and Red Sea has a higher importance in the context of rainfall, to do this, we can focus on the sea that is closest for each region and take the majority vote of both seas to make a decision. For example Addis Abada is 1252km from the Red Sea and 2816km from the Arabian Sea. Min(1252km, 2816) = 1252km, so +1 for Red Sea. By this process, we end up with this ascending order list:

- Gondar:
    - Red Sea: 861 km
- Dire Dawa:
    - Red Sea: 1239 km
- Addis Ababa:
    - Red Sea: 1252 km
- Wonji:
    - Red Sea: 1317 km
- Gambela:
    - Red Sea: 1402 km
- Hosaina:
    - Red Sea: 1416 km
- Mogadishu:
    - Red Sea: 2157 km
- Chisimiao:
    - Red Sea: 2335 km

We can see that Red Sea is the clear winner here and thus we can use the distance from Red Sea and altitude, like so: *Ratio = Min(Altitude)/Max(Distance)* to create an ascending order list:

- Chisimiao:
    - Altitude: 6m
    - Distance from Sea: 2335 km
    - **Ratio: 0.00257 m/km**
- Mogadishu:
    - Altitude: 61m
    - Distance from Sea: 2157 km
    - **Ratio: 0.02828 m/km**
- Gambela:

  - Altitude: 526m
  - Distance from Sea: 1402 km
  - **Ratio: 0.37517 m/km**
- Dire Dawa:
  - Altitude: 1180m
  - Distance from Sea: 1239 km
  - **Ratio: 0.95238 m/km**
- Wonji:
  - Altitude: 1588m
  - Distance from Sea: 1317 km
  - **Ratio: 1.20591 m/km**
- Hosaina:
  - Altitude: 2177m
  - Distance from Sea: 1416 km
  - **Ratio: 1.53743 m/km**
- Addis Ababa:
  - Altitude: 2294m
  - Distance from Sea: 1252 km
  - **Ratio: 1.83226 m/km**
- Gondar:
  - Altitude: 2111m
  - Distance from Sea: 861 km
  - **Ratio: 2.45181 m/km**

Using this calculated ratio, we can see a positive relationship between higher rainfall and higher ratio, this can be observed in the pronounced peaks for regions with higher ratio and lower peaks or very little rain fall averages for lower ratios when observing the monthly rainfall averages graphs.

- Higher Rainfall Higher Ratio:
  - Addis Abada: 1.83226 m/km
  - Gondar: 2.45181 m/km
  - Hosaina: 1.53743 m/km
  - *Gambela: 0.37517 m/km
- Moderate Rainfall Moderate Ratio:
  - Wonji: 1.20591 m/km
  - Dire Dawa: 0.95238 m/km
- Lower Rainfall Lower Ratio:
  - Mogadishu: 0.02828 m/km
  - Chisimiao: 0.00257 m/km

**Gambela seems to be the only region that seems to be an exception, even with peaks similar to high rain fall area, the ratio and rainfall data don't form a positive relationship, this points to the idea that there is other avenues that haven't been explored to understand this region, however we will stick to the ratio given on a majority it has proved true and shown a positive relationship with rainfall data.*

The reason the classification heavily relies on the calculated ratio is because this relationship of altitude and distance to a large water body can give us a quantitative measure of the possible orographic lifting of each region. When a highly elevated region is close to sea, the moist air formed due to these elevated regions which is cooled down creates the right conditions for higher rainfall amounts.

After this classification, we can dive into the yearly rainfall data for each classified section and it's regions and get a better understanding of the type of rainfall, meaning wether it is cyclical, monotonic or something else. Based on graphed data, we can see the following:

- Higher Rainfall Higher Ratio:
  - Addis Abada: With many peaks and valleys this can be considered cyclical.The trend is upwards, hence it can be considered positive.
  - Gondar: With many peaks and valleys this can be considered cyclical.The trend is downwards, hence it can be considered negative.
- Moderate Rainfall Moderate Ratio:
  - Hosaina: With not many many peaks and valleys this can be considered monotonic. The trend is neutral without much change.
  - Wonji: With many peaks and valleys this can be considered cyclical.The trend is neutral without much change.
  - *Dire Dawa: With many peaks and valleys this can be considered cyclical. The trend is downwards, hence it can be considered negative.
- Lower Rainfall Lower Ratio:
  - Mogadishu: With many peaks and valleys this can be considered cyclical. The trend is neutral without much change.
  - Chisimiao: With many peaks and valleys this can be considered cyclical. The trend is upwards, hence it can be considered positive.
  - *Gambela: With many peaks and valleys this can be considered cyclical. The trend is neutral without much change.

Missing Data:

- (1939-1945): Italy, which had occupied Somalia since the late 19th century, was defeated by British and Commonwealth forces in East Africa during World War II. British forces occupied Italian Somaliland in 1941, and the region came under British military administration.
- Ethiopian Civil War (1974-1991): The Ethiopian Civil War was a complex conflict that involved various factions, including the Marxist-Leninist Derg regime, which came to power in 1974. The conflict included internal strife, human rights abuses, and famine. It culminated in the overthrow of the Derg regime in 1991.
- Famine (1983-1985): The mid-1980s saw a severe famine in Ethiopia, exacerbated by a combination of factors, including drought, political instability, and civil war. The famine gained international attention, particularly due to images of suffering and malnutrition. The Live Aid concerts in 1985 were organized to raise funds for famine relief in Ethiopia.

Any missing data was handled via removing all values -1000 and leaving the cell empty.

Future Prediction:

To ensure that the predictive model was accurate enough for forecasting rainfall patterns, it went through some changes. At first, a simple linear regression model was employed, providing a straight line relation; however this model could not adequately depict the complicated non-linearity related to such things as rainfall with regards to weather data.:%.* Therefore, to overcome this drawback, a Random Forest regressor, a sophisticated machine learning algorithm that can capture intricate, non-linear patterns found in data, was added to the model. Rolling averages and time series analyses were included in this process to show smoothly represented rainfall patterns with fluctuation short-terms and long-terms trends. The randomly formed forest of decision trees performed much better relatively than did the first linear regressions.

Issues faced in predictions: For one region in particular, there was an issue reading the csv file and it wasn't possible to make the predictions for this region, but we can use the monthly averages as a predicted rainfall amount for this region.

Code Sample:

- Naive Linear Regression:

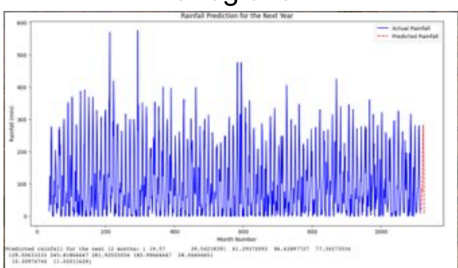
```
import pandas as pd
from sklearn.linear_model import LinearRegression
import matplotlib.pyplot as plt
import numpy as np

csv_file = "file1 - Sheet1.csv"
df = pd.read_csv(csv_file)

monthly_averages = df.mean(axis=1)

X = np.arange(len(monthly_averages)).reshape(-1, 1)
y = monthly_averages.values

model = LinearRegression()
model.fit(X, y)

future_months = np.arange(len(monthly_averages), len(monthly_averages) + 12).reshape(-1, 1)
predictions = model.predict(future_months)

plt.figure(figsize=(10, 6))
plt.scatter(X, y, color='blue', label='Actual Rainfall')
plt.plot(future_months, predictions, color='red', label='Predicted Rainfall')
plt.xlabel('Month Number')
plt.ylabel('Rainfall (mm)')
plt.title('Rainfall Prediction for the Next Year')
plt.legend()
plt.show()

print(predictions)
```

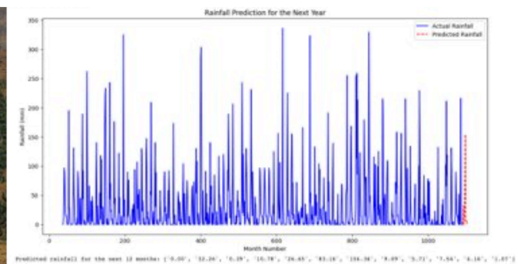
- Sophisticated Random Forest Regression with TimeSeries and Rolling Average: (Sample for Wonji) *code was written similar to this for every single region.*

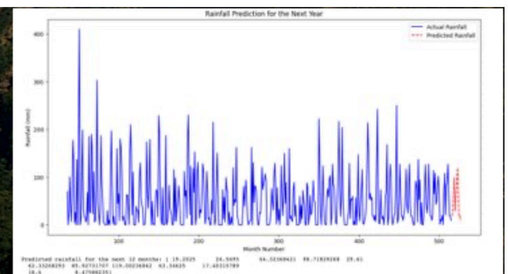
```
#Region Wonji
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.ensemble import RandomForestRegressor
from sklearn.model_selection import TimeSeriesSplit
from sklearn.metrics import mean_squared_error
from sklearn.impute import SimpleImputer

file_path = 'file1 - Sheet1.csv'
df = pd.read_csv(file_path)
imputer = SimpleImputer(strategy='mean')
df_filled = pd.DataFrame(imputer.fit_transform(df), columns=df.columns)

window_size = 3
rolling_averages = df_filled.rolling(window=window_size).mean().shift(1)
rolling_averages.dropna(inplace=True)
X = rolling_averages.values
rows_dropped_in_X = df_filled.shape[0] - rolling_averages.shape[0]
num_years = df_filled.shape[1]
y_dropped_values = rows_dropped_in_X * num_years
y = df_filled.iloc[rows_dropped_in_X:].values.flatten()
y_reshaped = y.reshape(-1, num_years)
model = RandomForestRegressor(n_estimators=100, random_state=42)
tscv = TimeSeriesSplit(n_splits=5)
cross_val_scores = []

for train_index, test_index in tscv.split(X):
    X_train, X_test = X[train_index], X[test_index]
    y_train, y_test = y_reshaped[train_index], y_reshaped[test_index]
    model.fit(X_train, y_train)
    predictions = model.predict(X_test)
    mse = mean_squared_error(y_test, predictions)
    rmse = np.sqrt(mse)
    cross_val_scores.append(rmse)
    print(f"RMSE: {rmse}")

model.fit(X, y_reshaped)

last_rolling_average = rolling_averages.iloc[-1].values.reshape(1, -1)
future_values = model.predict(np.tile(last_rolling_average, (12, 1)))

plt.figure(figsize=(14, 7))
plt.plot(range(rows_dropped_in_X * 12, rows_dropped_in_X * 12 + len(y)), y, label='Actual Rainfall', color='blue')
plt.plot(range(rows_dropped_in_X * 12 + len(y), rows_dropped_in_X * 12 + len(y) + 12), future_values.flatten()[:12], label='Predicted Rainfall', color='red', linestyle='--')

plt.xlabel('Month Number')
plt.ylabel('Rainfall (mm)')
plt.title('Rainfall Prediction for the Next Year')
plt.legend()
plt.show()
print("Predicted rainfall for the next 12 months:", future_values.flatten()[:12])
```

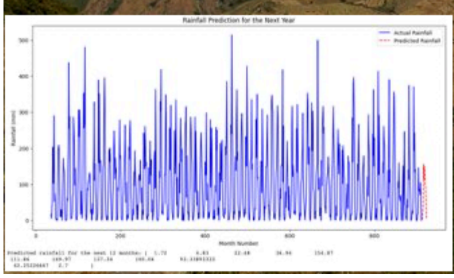
- Predictions made by Sophisticated Python coded model:
  - Addis Abada - 19.57, 39.54, 61.29, 86.64, 77.35, 129.51, 265.82, 281.93, 183.00, 28.05, 10.31, 11.02 (mm)
  - Chisimiao - 0.00, 32.26, 0.39, 10.78, 26.65, 83.16, 154.36, 9.09, 5.71, 7.54, 4.16, 1.07 (mm)
  - Dire Dawa - 19.20, 26.57, 64.32, 98.72, 29.61, 62.33, 85.93, 119.00, 63.35, 17.40, 18.60, 8.48 (mm)
  - Gambela - 1.72, 4.83, 22.48, 36.94, 154.87, 111.86, 149.97, 137.34, 100.06, 93.34, 42.25, 2.70 (mm)
  - Gonda - 0.75, 3.46, 17.32, 45.00, 87.62, 143.83, 263.04, 242.93, 31.80, 154.89, 35.58, 11.91 (mm)
  - Hosaina - 6.85, 53.91, 50.18, 115.88, 107.57, 121.95, 168.31, 187.33, 152.25, 43.42, 20.07, 24.09 (mm)
  - Mogadishu - 1.72, 4.83, 22.48, 36.94, 154.87, 111.86, 149.97, 137.34, 100.06, 93.34, 42.25, 2.70 (mm)
  - Wonji - 36.10, 71.60, 242.70, 222.40, 1246.00, 717.85, 2033.90, 1607.40, 861.30, 93.27, 76.20, 62.17 (mm)

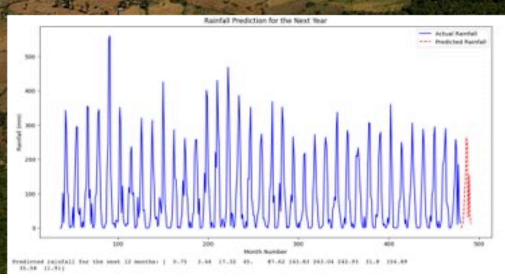
You will notice the prediction values for Mogadishu and Gambela are the same, this is because the code written for Mogadishu was tested using the Gambela csv due to issues loading Mogadishu csv values into the code, and unfortunately no amount of fixing the code resolved the issues to predict the proper values for this region. Below are the graphs and outputs for all the regions.

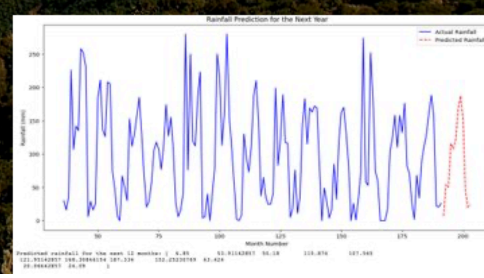


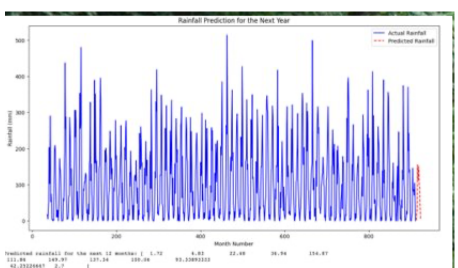


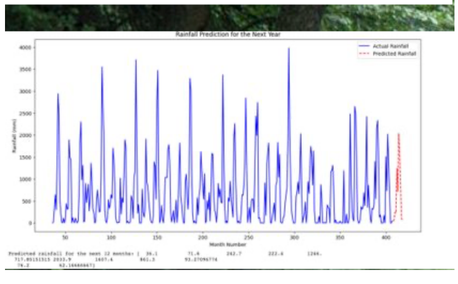






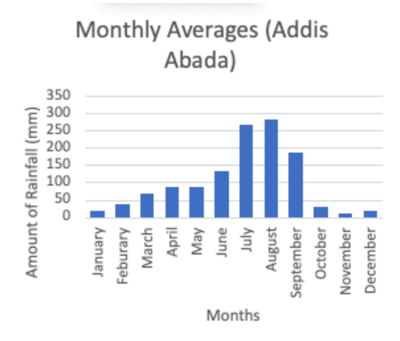


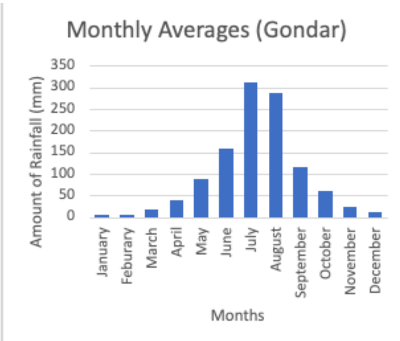


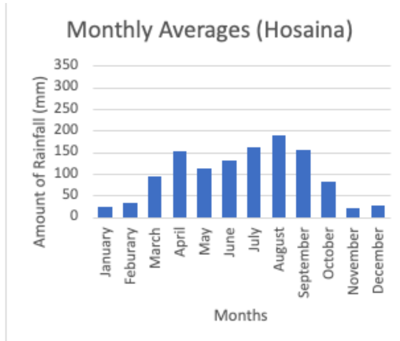


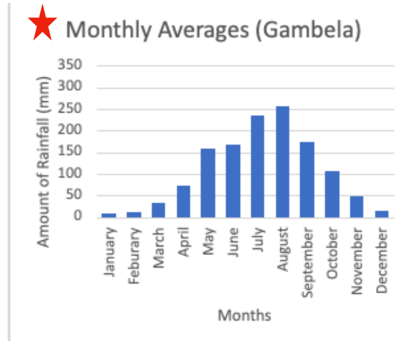
Graphs: (Post-Classification using ratio)

Higher Ratio-Higher RainFall: [Gondar, Addis Abada, Hosaina, Gambela]

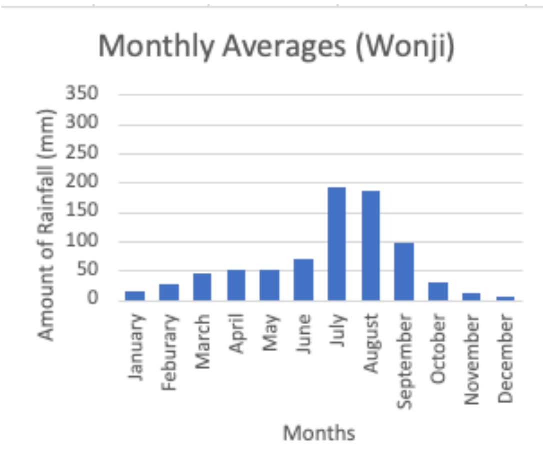


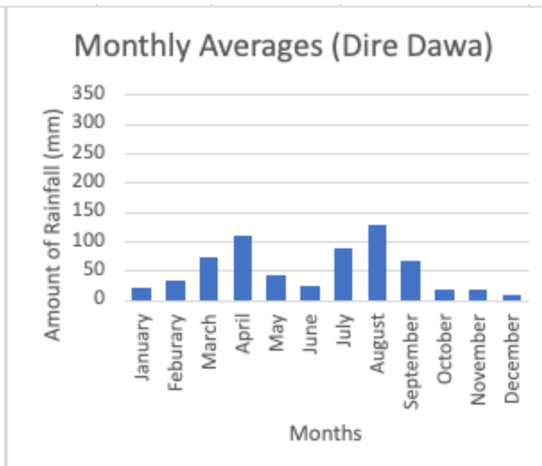




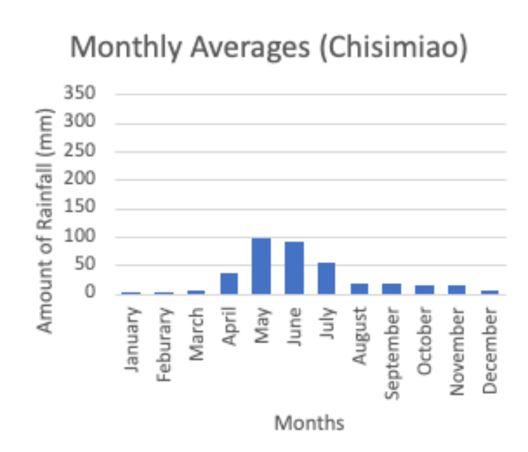


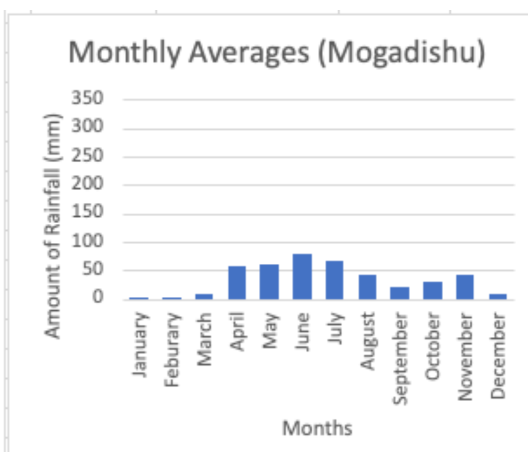
Moderate Ratio-Moderate Rainfall: [Wonji, Dire Dawa]





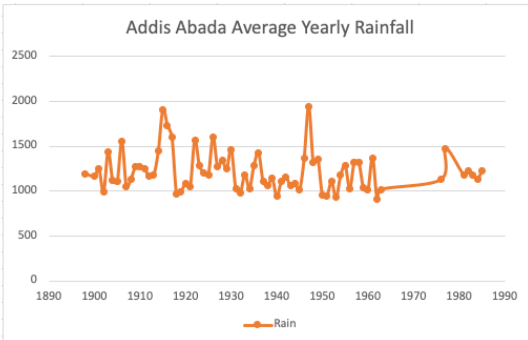
Lower Ratio-Lower Rainfall: [Mogadishu, Chisimiao]

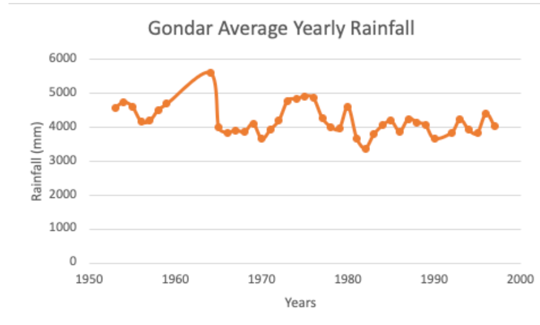


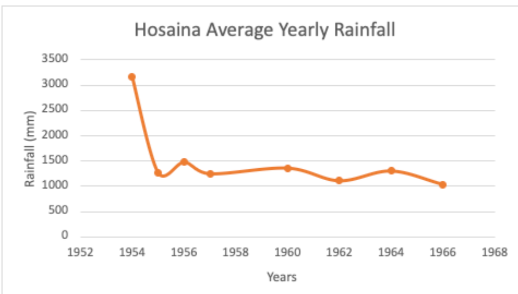


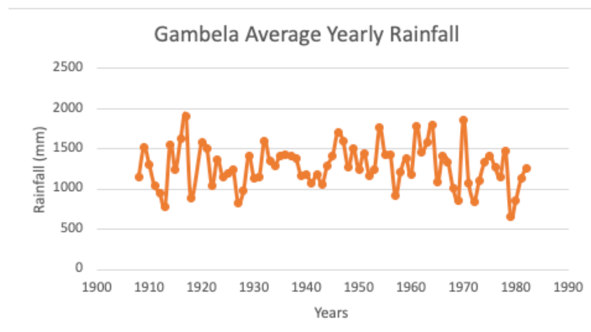
Yearly Scatter Plots:

Higher Ratio-Higher RainFall: [Gondar, Addis Abada, Hosaina, Gambela]

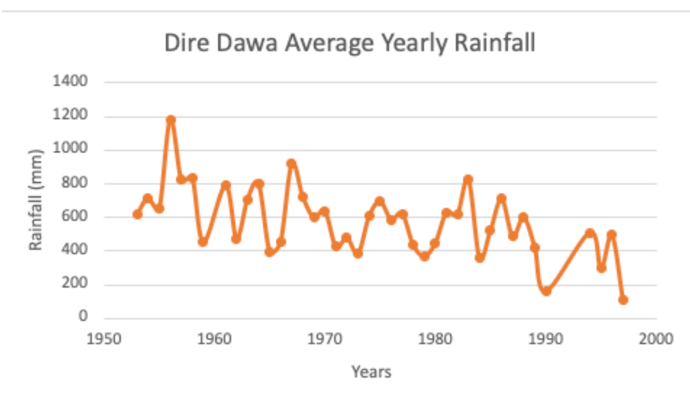


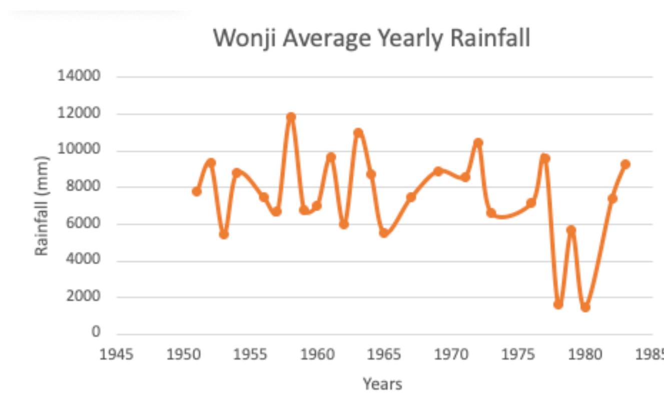






Moderate Ratio-Moderate Rainfall: [Wonji, Dire Dawa]





Lower Ratio-Lower Rainfall: [Mogadishu, Chisimiao]