#Hurricane_Dataset

The unit of observation is a year. Each row of the dataset matches with one year, from 1971 to 2021. Each subsection below pertains to a variable describing the observations, with the heading being the variable name. As this dataset was constructed by us, it does not contain any null or missing values.

```
! git clone https://github.com/jillianhaig/Project2_DS4002 # so we can access data loaded from shared github
```

```
Cloning into 'Project2_DS4002'...
remote: Enumerating objects: 222, done.ote: Counting objects: 100% (56/56), done.ote: Compressing objects: 100% (41/41), done.ote: Total 222 (delta 31), reused 14 (delta 14), pack-reused 166 (from 1)
```

```
import pandas as pd
import matplotlib.pyplot as plt
hurricane_data = pd.read_csv("/content/Project2_DS4002/DATA/Final Dataset/Hurricane_Dataset.csv")
```

###Atlantic_Hurricane_Days

The variable takes on float values which could be anywhere from 0 to 365 and includes quarter days (.25), half days (.5), and 3/4 of day (.75). This represents the amount of days in one year that there was a hurricane in the Atlantic Ocean in the particular year indicated by the "Year" column.

```
print(hurricane_data["Atlantic_Hurricane_Days"].count())
```

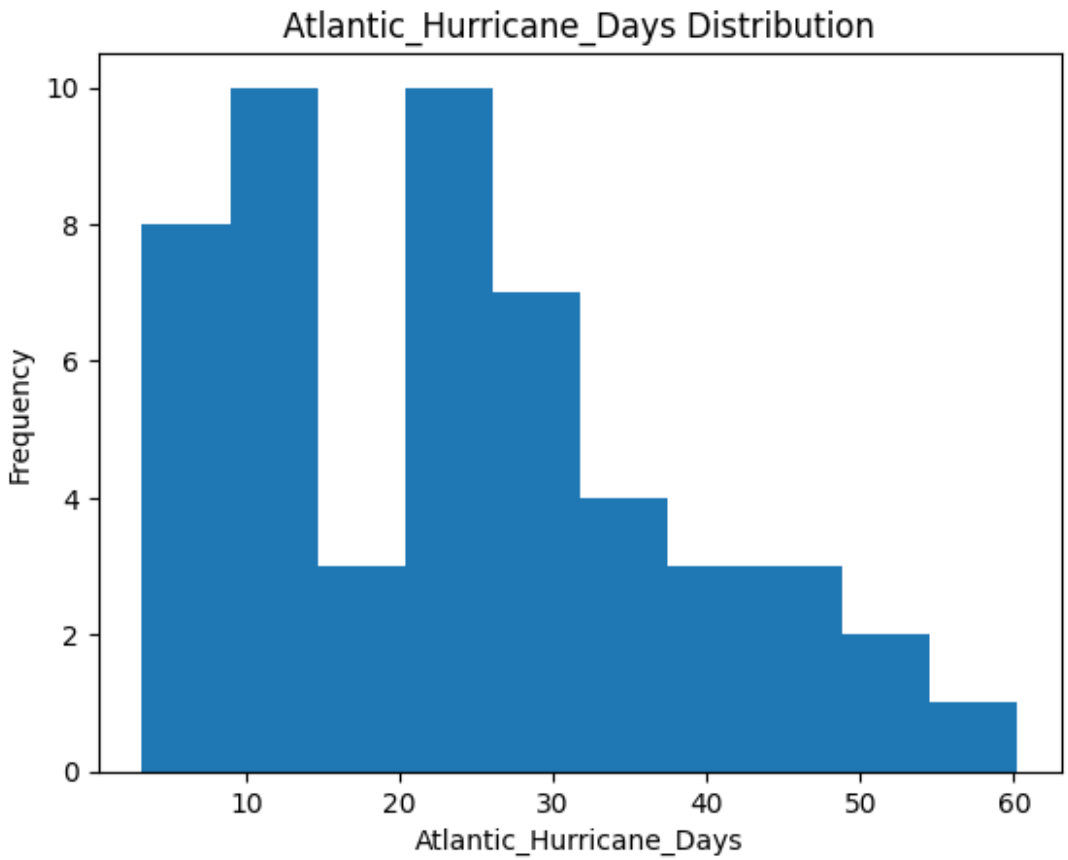
```
51
```

```
hurricane_data["Atlantic_Hurricane_Days"].describe()
```

```
count    51.000000
mean     23.328431
std      14.058226
min       3.250000
25%      11.375000
50%      21.750000
75%      31.125000
max      60.250000
Name: Atlantic_Hurricane_Days, dtype: float64
```

```
plt.hist(hurricane_data["Atlantic_Hurricane_Days"])
plt.xlabel('Atlantic_Hurricane_Days')
plt.ylabel('Frequency')
plt.title('Atlantic_Hurricane_Days Distribution')
```

```
Text(0.5, 1.0, 'Atlantic_Hurricane_Days Distribution')
```



###Pacific_Hurricane_Days

The variable takes on float values which could be anywhere from 0 to 365 and includes quarter days (.25), half days (.5), and 3/4 of day (.75). This represents the amount of days in one year that there was a hurricane in the Pacific Ocean in the particular year indicated by the "Year" column.

```
print(hurricane_data["Pacific_Hurricane_Days"].count())
```

```
51
```

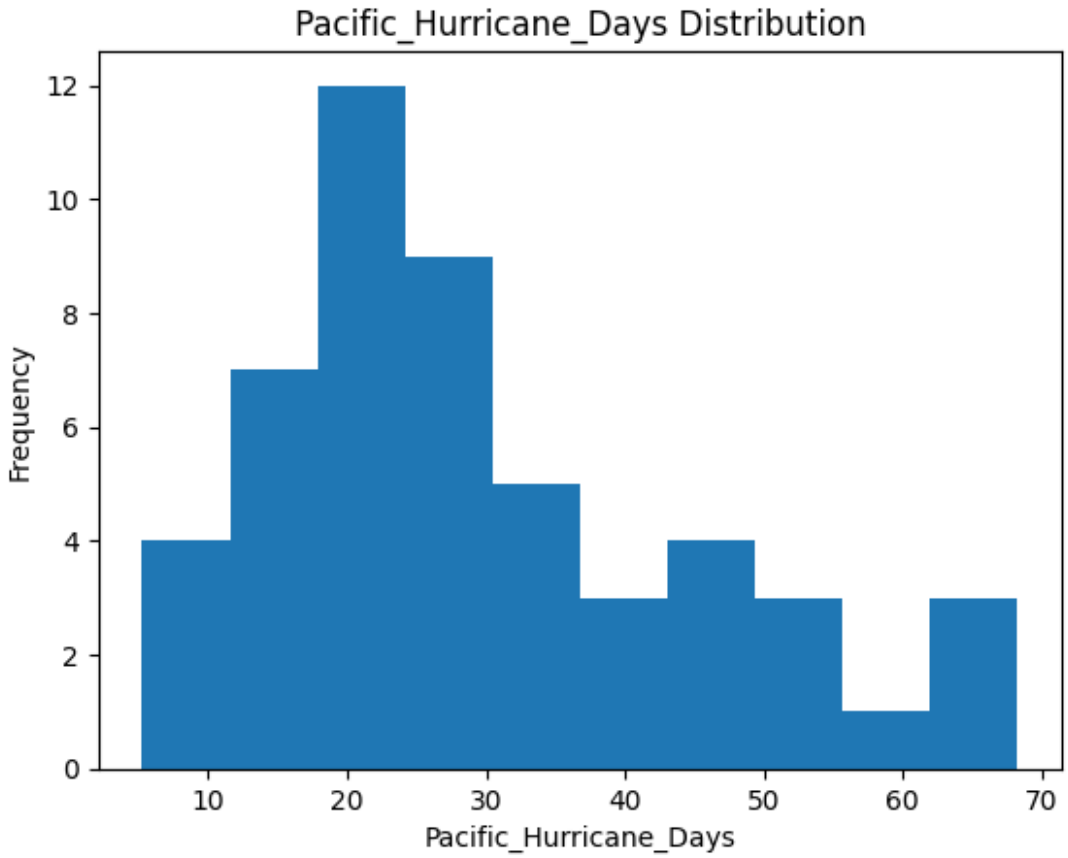
```
hurricane_data["Pacific_Hurricane_Days"].describe()
```

```
count    51.000000
mean     29.774510
std      15.181836
min       5.250000
25%      18.750000
50%      26.750000
75%      38.500000
max      68.250000
Name: Pacific_Hurricane_Days, dtype: float64
```

```
plt.hist(hurricane_data["Pacific_Hurricane_Days"])
plt.xlabel('Pacific_Hurricane_Days')
```

```
plt.ylabel('Frequency')
plt.title('Pacific_Hurricane_Days Distribution')
```

```
Text(0.5, 1.0, 'Pacific_Hurricane_Days Distribution')
```



###Total_Hurricane_Days

The variable takes on float values which could be anywhere from 0 to 730 and includes quarter days (.25), half days (.5), and 3/4 of day (.75). This represents the amount of days that there was a hurricane in either the Atlantic Ocean or Pacific Ocean in the particular year indicated by the "Year" column. If there was a hurricane in both oceans on the same day for the whole year, this would equal 730, but this is extremely unlikely.

```
print(hurricane_data["Total_Hurricane_Days"].count())
```

```
51
```

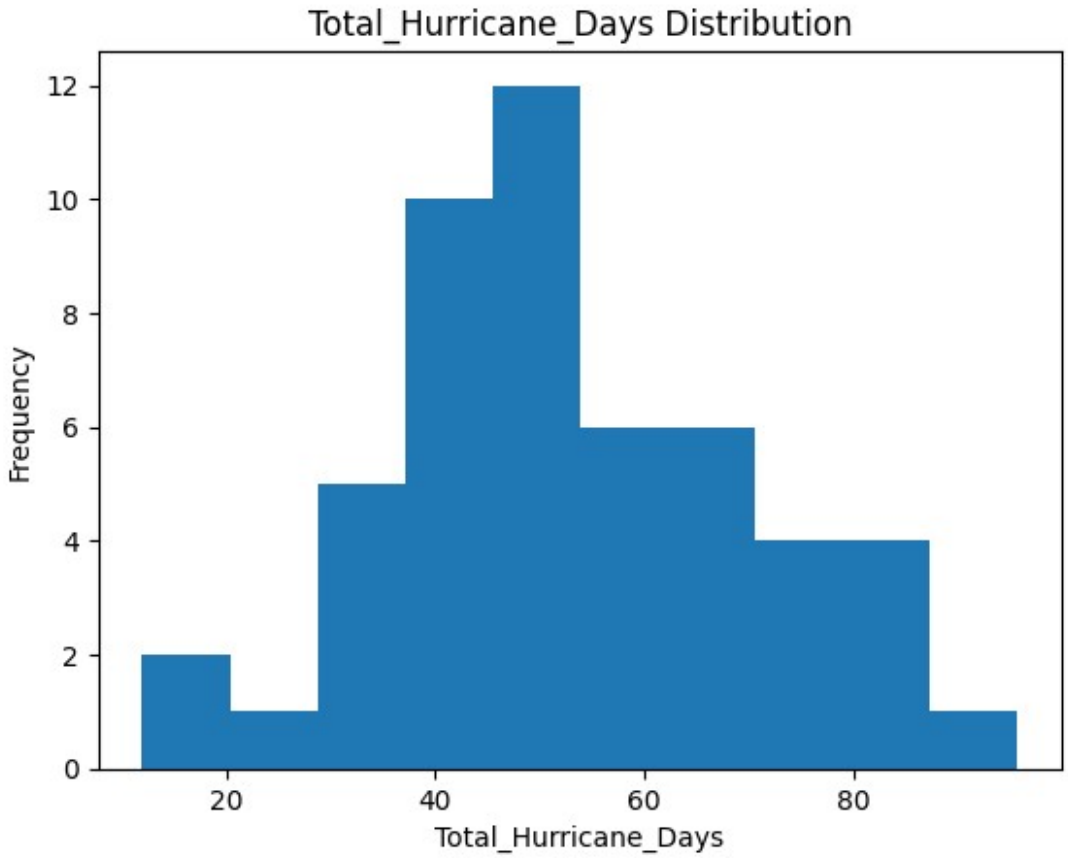
```
hurricane_data["Total_Hurricane_Days"].describe()
```

```
count    51.000000
mean     53.102941
std      17.265803
min      12.000000
```

```
25%      40.500000
50%      50.750000
75%      64.500000
max      95.750000
Name: Total_Hurricane_Days, dtype: float64
```

```
plt.hist(hurricane_data["Total_Hurricane_Days"])
plt.xlabel('Total_Hurricane_Days')
plt.ylabel('Frequency')
plt.title('Total_Hurricane_Days Distribution')
```

```
Text(0.5, 1.0, 'Total_Hurricane_Days Distribution')
```



###CO2_emissions

The variable takes on float values that are any positive number. It measures the CO2 emissions for the entire world in tonnes on a yearly basis.

```
print(hurricane_data["CO2_emissions"].count())
```

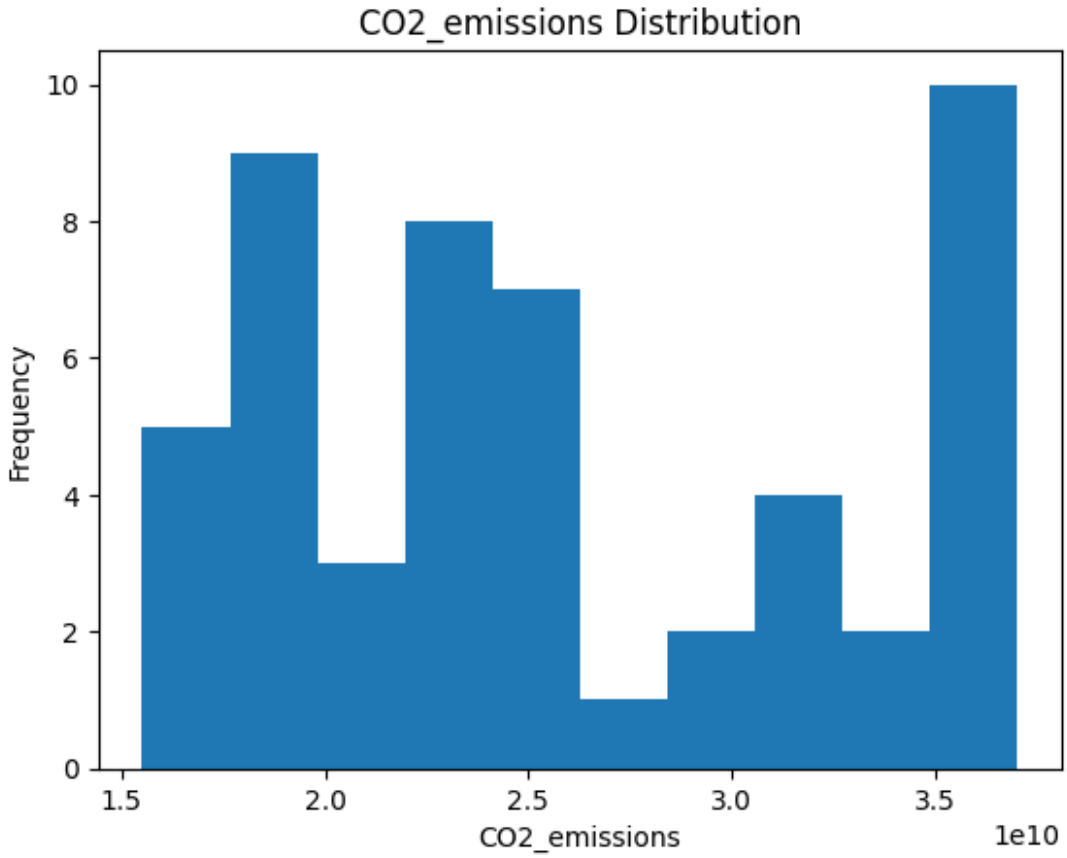
```
51
```

```
hurricane_data["CO2_emissions"].describe()
```

```
count    5.100000e+01
mean     2.571027e+10
std      6.777399e+09
min      1.550279e+10
25%      1.962237e+10
50%      2.425016e+10
75%      3.177082e+10
max      3.704010e+10
Name: CO2_emissions, dtype: float64
```

```
plt.hist(hurricane_data["CO2_emissions"])
plt.xlabel('CO2_emissions')
plt.ylabel('Frequency')
plt.title('CO2_emissions Distribution')
```

```
Text(0.5, 1.0, 'CO2_emissions Distribution')
```



###Population

The variable takes on integer values that can only be positive representing the worlds population on a yearly basis.

```
print(hurricane_data["Population"].count())
```

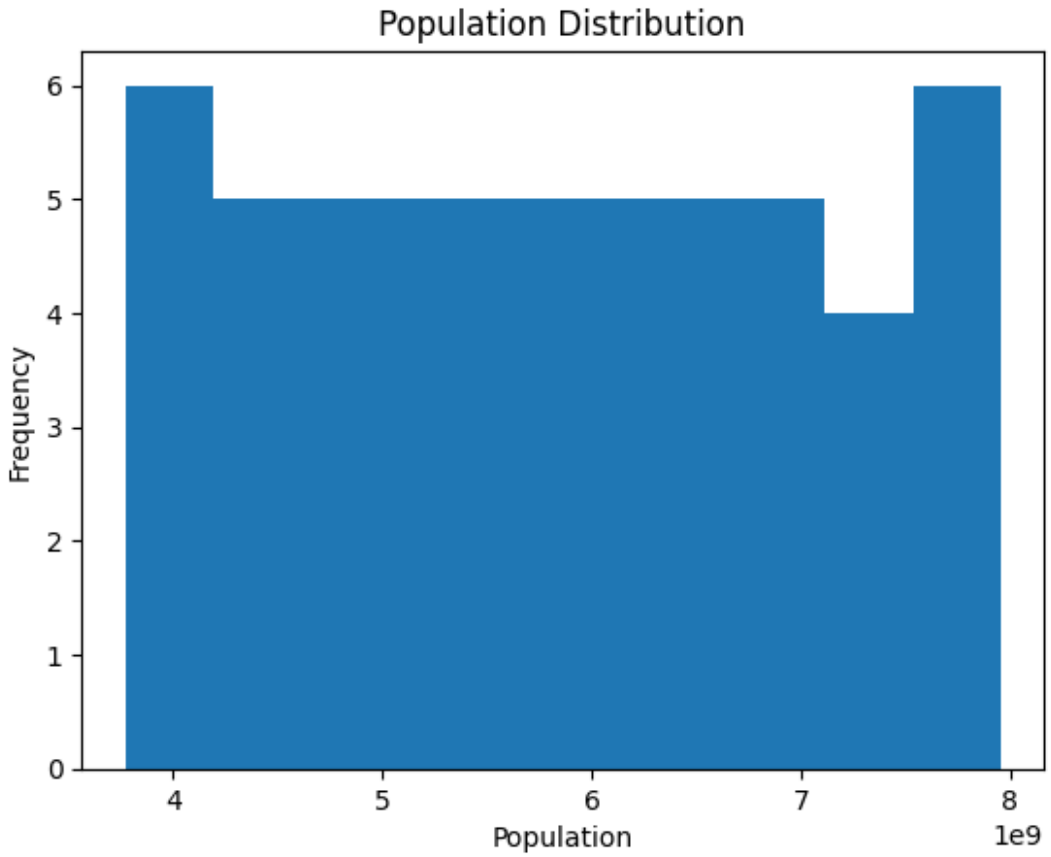
```
51
```

```
hurricane_data["Population"].describe()
```

```
count    5.100000e+01
mean     5.835966e+09
std      1.265055e+09
min      3.769848e+09
25%      4.739752e+09
50%      5.842056e+09
75%      6.888612e+09
max      7.954448e+09
Name: Population, dtype: float64
```

```
plt.hist(hurricane_data["Population"])
plt.xlabel('Population')
plt.ylabel('Frequency')
plt.title('Population Distribution')
```

```
Text(0.5, 1.0, 'Population Distribution')
```



###CSIRO_Adjusted_SeaLevel_Inch

The variable takes on float values and represents how much the sea level (in inches) has increased worldwide per year since 1880. Since 1880 is the base year, the value would be 0 for that year, and it increases from there.

```
print(hurricane_data["CSIRO_Adjusted_SeaLevel_Inch"].count())
```

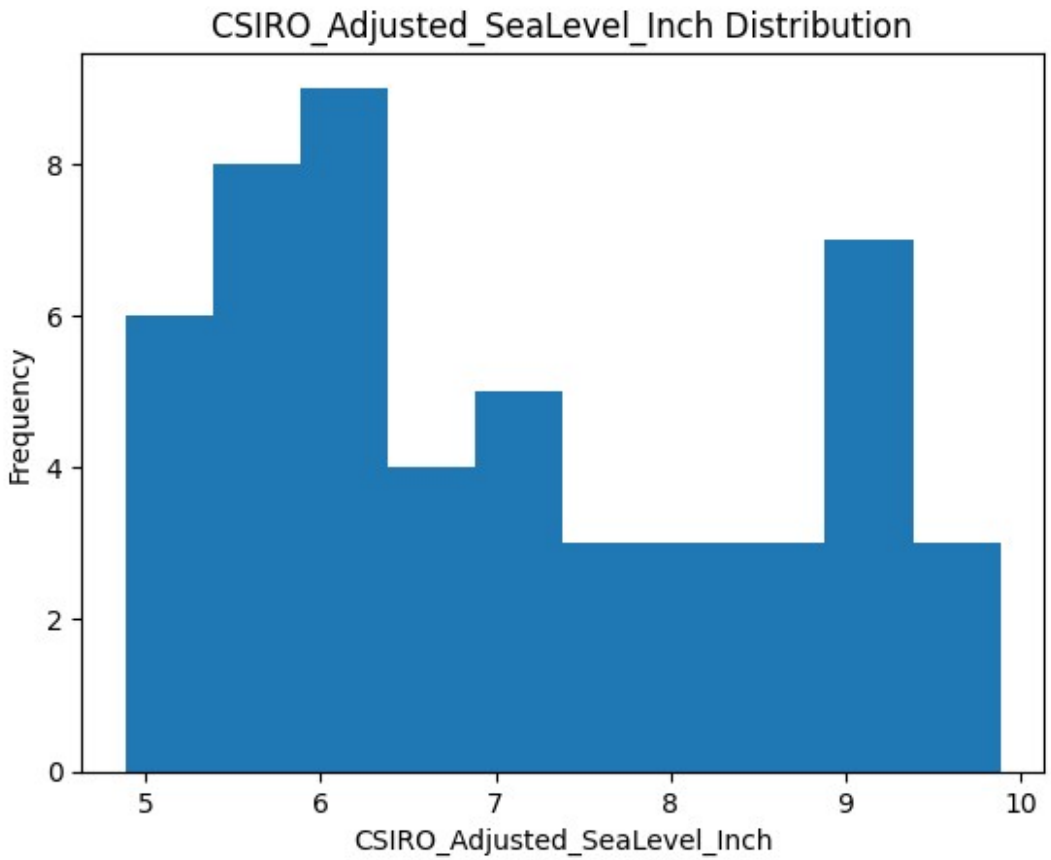
```
51
```

```
hurricane_data["CSIRO_Adjusted_SeaLevel_Inch"].describe()
```

```
count    51.000000
mean      7.078656
std       1.474016
min       4.881890
25%       5.826772
50%       6.669291
75%       8.417323
max       9.886688
Name: CSIRO_Adjusted_SeaLevel_Inch, dtype: float64
```

```
plt.hist(hurricane_data["CSIRO_Adjusted_SeaLevel_Inch"])
plt.xlabel('CSIRO_Adjusted_SeaLevel_Inch')
plt.ylabel('Frequency')
plt.title('CSIRO_Adjusted_SeaLevel_Inch Distribution')
```

```
Text(0.5, 1.0, 'CSIRO_Adjusted_SeaLevel_Inch Distribution')
```



###Surface_Temperature_Change

The variable takes on float values which could be any negative or positive number. It represents the change in the Earth's land temperature over time. 1880 is also the base year, so the value for each year represents the change in land temperature since 1880.

```
print(hurricane_data["Surface_Temperature_Change"].count())
```

```
51
```

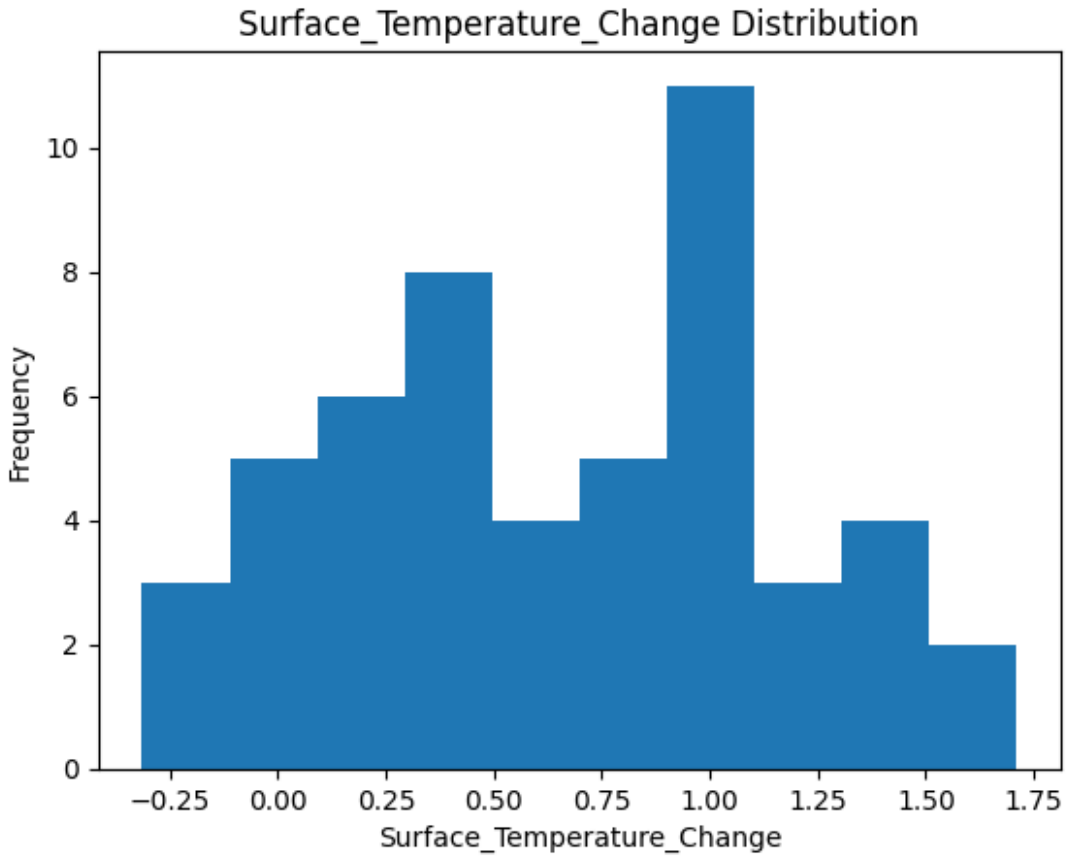
```
hurricane_data["Surface_Temperature_Change"].describe()
```

```
count    51.000000
mean      0.667255
std       0.514895
min      -0.314000
25%       0.269000
50%       0.656000
75%       1.018500
max       1.711000
Name: Surface_Temperature_Change, dtype: float64
```

```
plt.hist(hurricane_data["Surface_Temperature_Change"])
plt.xlabel('Surface_Temperature_Change')
```

```
plt.ylabel('Frequency')
plt.title('Surface_Temperature_Change Distribution')

Text(0.5, 1.0, 'Surface_Temperature_Change Distribution')
```



###SurfaceSeaTemp_Annual_Anomaly

The variable takes on float values which could be any negative or positive number. It represents the change in the Earth's sea temperature over time. 1880 is also the base year, so the value for each year represents the change in sea temperature since 1880.

```
print(hurricane_data["SurfaceSeaTemp_Annual_Anomaly"].count())

51

hurricane_data["SurfaceSeaTemp_Annual_Anomaly"].describe()

count    51.000000
mean      0.229431
std       0.352201
min      -0.519000
25%       0.019500
50%       0.183000
```

```
75%       0.441000
max       0.930000
Name: SurfaceSeaTemp_Annual_Anomaly, dtype: float64
```

```
plt.hist(hurricane_data["SurfaceSeaTemp_Annual_Anomaly"])
plt.xlabel('SurfaceSeaTemp_Annual_Anomaly')
plt.ylabel('Frequency')
plt.title('SurfaceSeaTemp_Annual_Anomaly Distribution')
```

```
Text(0.5, 1.0, 'SurfaceSeaTemp_Annual_Anomaly Distribution')
```