

```
from google.colab import drive

drive.mount('/content/gdrive')

Mounted at /content/gdrive
```

▼ Atlantic Hurricane Data 1851-2015

Goals: Have hurricanes been getting stronger with time? In this dataset we are going to investigate whether or not hurricanes are getting stronger, weaker, or have remained on the same cycle for the amount of time presented in this dataset.

Dataset: <https://www.kaggle.com/datasets/maravedi/atlantic-hurricane-database>

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import math
from scipy.stats.stats import ttest_ind
from scipy import stats
sns.set()
from google.colab import drive
import warnings

warnings.filterwarnings('ignore')

#import pandas and the dataset
import pandas as pd
atl_hurr_df=pd.read_csv('/content/gdrive/My Drive/Atlantic_Hurricane_Data.csv')

atl_hurr_df.head()
```

```

      ID      Name      Date  Time  Event  Status  Latitude  Longitude  Maximum
      Wind

atl_hurr_df.shape

(49105, 22)

null_df= atl_hurr_df.isna().sum()
print(null_df)

ID      0
Name    0
Date    0
Time    0
Event   0
Status  0
Latitude 0
Longitude 0
Maximum Wind 0
Minimum Pressure 0
Low Wind NE 0
Low Wind SE 0
Low Wind SW 0
Low Wind NW 0
Moderate Wind NE 0
Moderate Wind SE 0
Moderate Wind SW 0
Moderate Wind NW 0
High Wind NE 0
High Wind SE 0
High Wind SW 0
High Wind NW 0
dtype: int64

```

Check for null values: There are no nulls in this dataset. However, there are values of -999 for years where the data was not collected. We want to collect all the data we have, but we don't want -999 to impact our analysis.

We need to use if/then statements to isolate the "-999" factors and keep them out of our analysis.

```

names_df= atl_hurr_df['Name'].value_counts()
names_df

```

```

UNNAMED    26563
FRANCES     317
ARLENE     283
BERTHA     268
DENNIS     255

```

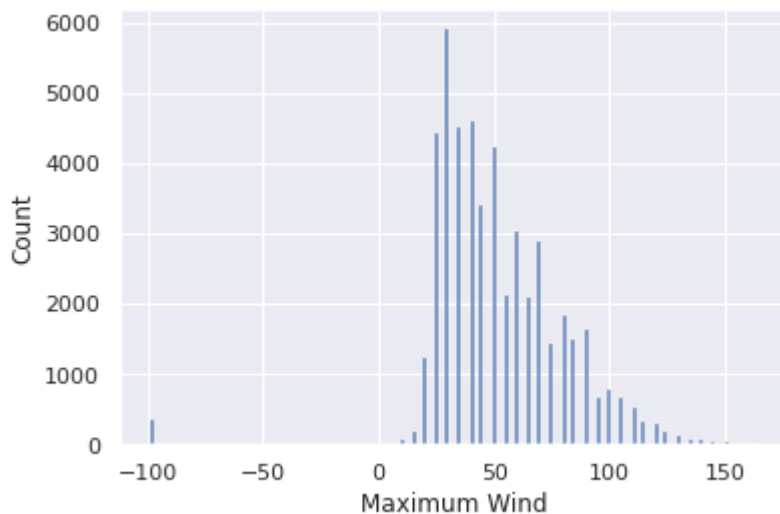
```

...
SHARY          9
TAMMY          9
SIXTEEN        8
FERNAND         7
AMELIA          6
Name: Name, Length: 288, dtype: int64

```

There are 26,563 hurricanes that are unnamed. There are many hurricanes that have the same names. This means we should categorize our data by ID number, not by name.

```
ax=sns.histplot(atl_hurr_df['Maximum Wind'])
```



Winds greater than 100mph

Let's narrow down our search for the strongest hurricanes and attempt to create a visual that will help track strenght of hurricanes over time.

```
winds_over_100 = atl_hurr_df['Maximum Wind'] >100
```

```
winds_over_100
```

```

0      False
1      False
2      False
3      False
4      False
...
49100  False
49101  False
49102  False
49103  False
49104  False
Name: Maximum Wind, Length: 49105, dtype: bool

```

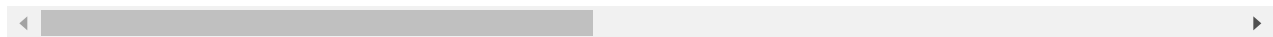
Select Data with Winds over 100mph:

Here we can see that the results for "Winds over 100mph" are demonstrated in a Boolean. Let's select ONLY the True statements, that is, every hurricane with winds over 100mph.

```
winds_100_df = atl_hurr_df.iloc[winds_over_100.values]
winds_100_df
```

	ID	Name	Date	Time	Event	Status	Latitude	Longitude	M
249	AL031853	UNNAMED	18530902	1800		HU	17.4N	51.0W	
250	AL031853	UNNAMED	18530903	0		HU	18.2N	52.8W	
251	AL031853	UNNAMED	18530903	600		HU	19.1N	54.6W	
252	AL031853	UNNAMED	18530903	1200		HU	19.7N	56.2W	
253	AL031853	UNNAMED	18530903	1800		HU	20.7N	58.0W	
...
49037	AL112015	JOAQUIN	20151003	600		HU	24.8N	73.6W	
49038	AL112015	JOAQUIN	20151003	1200		HU	25.4N	72.6W	
49039	AL112015	JOAQUIN	20151003	1800		HU	26.3N	71.0W	
49040	AL112015	JOAQUIN	20151004	0		HU	27.4N	69.5W	
49041	AL112015	JOAQUIN	20151004	600		HU	28.9N	68.3W	

2254 rows × 22 columns



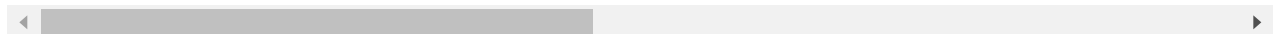
Now let's sort the data. We need to find the top wind speeds by sorting "Maximum Wind" from highest wind speed to smallest.

```
#DataFrame.sort_values(by, axis=0, ascending=True, inplace=False, kind='quicksort',
#na_position='last', ignore_index=False, key=None)
```

```
max_wind_df = winds_100_df.sort_values('Maximum Wind', axis=0, ascending=False)
max_wind_df
```

	ID	Name	Date	Time	Event	Status	Latitude	Longitude	M
33731	AL041980	ALLEN	19800807	1800		HU	21.8N	86.4W	
44435	AL252005	WILMA	20051019	1200		HU	17.3N	82.8W	
17943	AL031935	UNNAMED	19350903	600		HU	25.1N	81.1W	
17942	AL031935	UNNAMED	19350903	200	L	HU	24.8N	80.8W	
17941	AL031935	UNNAMED	19350903	0		HU	24.6N	80.5W	
...
19641	AL051941	UNNAMED	19411005	1800		HU	24.2N	75.4W	
20037	AL031943	UNNAMED	19430823	0		HU	24.5N	65.0W	
20038	AL031943	UNNAMED	19430823	600		HU	25.5N	65.5W	
20068	AL041943	UNNAMED	19430904	600		HU	32.5N	61.2W	
49041	AL112015	JOAQUIN	20151004	600		HU	28.9N	68.3W	

2254 rows × 22 columns



```
#>>> df = DataFrame({'a': [1, 10, 8, 11, -1],
#...                  'b': list('abdce'),
#...                  'c': [1.0, 2.0, np.nan, 3.0, 4.0]})
#>>> df.nlargest(3, 'a')
#   a  b  c
#3  11  c  3
#1  10  b  2
#2   8  d NaN
```

Top Ten Toughest Hurricanes: Maximum Wind

Let's look at the top ten hurricanes with maximum wind speed.

```
max_wind_df.nlargest(10, 'Maximum Wind')
```

	ID	Name	Date	Time	Event	Status	Latitude	Longitude	Maximum Wind
33731	AL041980	ALLEN	19800807	1800		HU	21.8N	86.4W	160
44435	AL252005	WILMA	20051019	1200		HU	17.3N	82.8W	160
17943	AL031935	UNNAMED	19350903	600		HU	25.1N	81.1W	160
17942	AL031935	UNNAMED	19350903	200	L	HU	24.8N	80.8W	160
17941	AL031935	UNNAMED	19350903	0		HU	24.6N	80.5W	160
36445	AL081988	GILBERT	19880914	0		HU	19.7N	83.8W	160
33722	AL041980	ALLEN	19800805	1200		HU	15.9N	70.5W	150
44434	AL252005	WILMA	20051018	600	L	HU	17.3N	82.8W	160

Hurricane Allen in 1980 clocks in with the highest wind speeds.

10 rows × 22 columns

Conclusion

More comprehensive data needs to be explored to see if hurricanes have trended towards higher winds in the latter half of the 20th century and the beginnings of the 21st. Since we only have data until 2015, it would be useful to join the most recent data tables and investigate this further.

[Colab paid products](#) - [Cancel contracts here](#)

