**[Assignment06: Python dataset exploration](https://mymasonportal.gmu.edu/webapps/assignment/uploadAssignment?content_id=_8335369_1&course_id=_347779_1&group_id=&mode=view)-Output**

1. **Summary Statistics and Visualizations:**

**Month:**

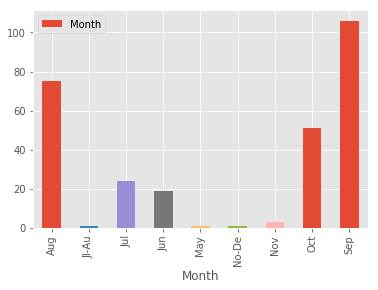
count 281

unique 9

top Sep

freq 106

Name: Month, dtype: object



**Highest Category:**

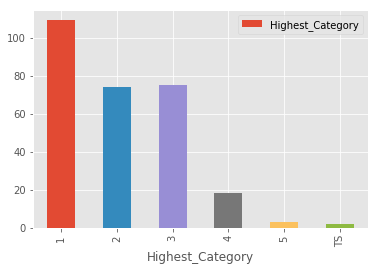
count 281

unique 6

top 1

freq 109

Name: Highest\_Category, dtype: object



**Central pressure mb:**

count 277.000000

mean 963.913357

std 54.979440

min 100.000000

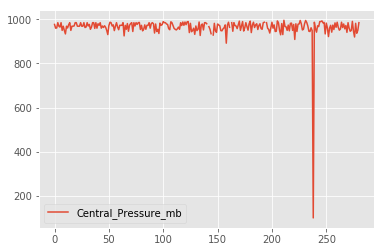
25% 955.000000

50% 970.000000

75% 982.000000

max 996.000000

Name: Central\_Pressure\_mb, dtype: float64



**Max Winds KT**:

count 182.000000

mean 88.351648

std 21.317839

min 55.000000

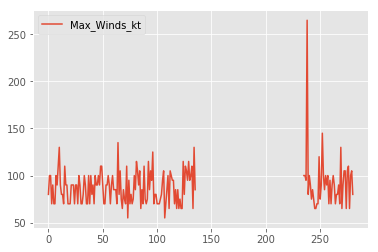
25% 70.000000

50% 90.000000

75% 100.000000

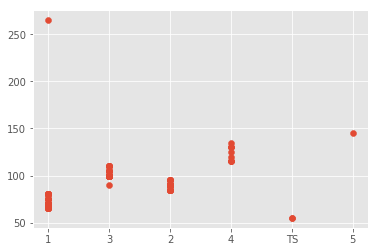
max 265.000000

Name: Max\_Winds\_kt, dtype: float64

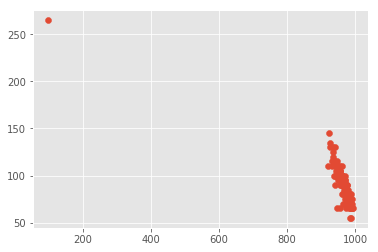


1. Relationship:

(1) Highest\_Category and Max\_Winds\_kt



(2) Central\_Pressure\_mb and Max\_Winds\_kt



**Part 2: Web/HTML Screen Scraping**

**Summary statistics output for Cores, RMax, RPeak, and Power**

1. **Cores:**

count 5.000000e+02

mean 1.161113e+05

std 5.515132e+05

min 1.195200e+04

25% 3.326400e+04

50% 4.564000e+04

75% 6.433600e+04

max 1.064960e+07

Name: Cores, dtype: float64

1. **RMax:**

count 500.00000

mean 2421.83140

std 8237.89877

min 715.60000

25% 832.50000

50% 1020.75000

75% 1649.10000

max 122300.00000

Name: Rmax (TFlop/s), dtype: float64

1. **Rpeak:**

count 500.000000

mean 3843.332000

std 12655.684899

min 778.500000

25% 1413.100000

50% 1792.000000

75% 2402.400000

max 187659.300000

Name: Rpeak (TFlop/s), dtype: float64

1. **Power:**

count 264.000000

mean 1596.928030

std 2545.841624

min 47.000000

25% 486.000000

50% 840.000000

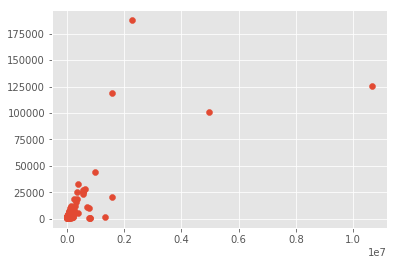
75% 1543.250000

max 19431.000000

Name: Power (kW), dtype: float64

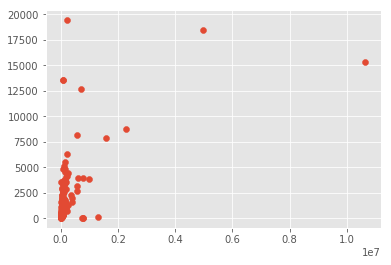
Relationship:

1. Cores and RPeak:



The scatter plots denote a fan - shaped relationship of two quantitative variables. By looking at the summary statistics, it can be said that the data is more concentrated near the mean( 1.161113e+05, 3843.332000) and more than 75% of the values are tightly coupled at the mean. Also, we can see that the data has few extreme values. We can see max values from the summary statistics.

1. Cores and Power:



The scatter plots denote a fan - shaped relationship of two quantitative variables. By looking at the summary statistics, it can be said that the data is more concentrated near the mean( 1.161113e+05, 1596.928030) and more than 75% of the values are tightly coupled at the mean. Also, we can see that the data has few extreme values. We can see max values from the summary statistics.

1. **Challenge Bonus:**

**Summary of Country:**

count 500

unique 27

top China

freq 206

Name: Country, dtype: object

