

Friday Meet Up - Never trust only in stats

Student: Jorge Candia

Group: GroupID

Date: 29/NOV/2019

A text file "DatasaurusDozen.tsv" has been provided with the notebook. TSV files are text files in which each line contains several fields separated by tabs. You must treat it as a CSV file using the tab character ('\t') as a separator.

The content of the file is similar to this:

dataset	x	y
dino	55.3846	97.1795
dino	51.5385	96.0256
dino	46.1538	94.4872
dino	42.8205	91.4103
dino	40.7692	88.3333
dino	38.7179	84.8718

The file contains 13 datasets, each identified by a word in the first column. Columns X and Y are the coordinates of each dataset point.

The objectives of the exercise are:

- Separate the 13 datasets and plot them as 13 scatter plots inside the same figure (use subplots) (6 points)
- Calculate the basic statistics for each dataset (mean(x), mean(y), standard_deviation(x), standard_deviation(y) and correlation(x,y)) and check that despite being 13 very different datasets, their statistics are practically the same. (2 points)
- Make a well-formatted printout in the form of a table (2 points)

ENTREGA: You have to submit into MOODLE this document in PDF and in IPYNB (IMPORTANT: in that order)

TIP. Remember that the correlation coefficient is calculated as: `np.corrcoef(x, y)[0][1]`

In []:

1

In [68]:

```
1 goodf=[]
2 with open('DatasaurusDozen.tsv','r') as file:
3     f=file.readlines()
4     for i in f:
5         goodf.append(i[0:-1])
```

In [91]:

```
1 dataset=[]
2 x=[]
3 y=[]
4 for i in goodf:
5     a=i.split('\t')
6     dataset.append(a[0])
7     x.append(a[1])
8     y.append(a[2])
9
10 count=0
11 for i in x:
12     if count!=0:
13         x[count]=float(i)
14         count+=1
15
16 count=0
17 for i in y:
18     if count!=0:
19         x[count]=float(i)
20     count+=1
```

In [114]:

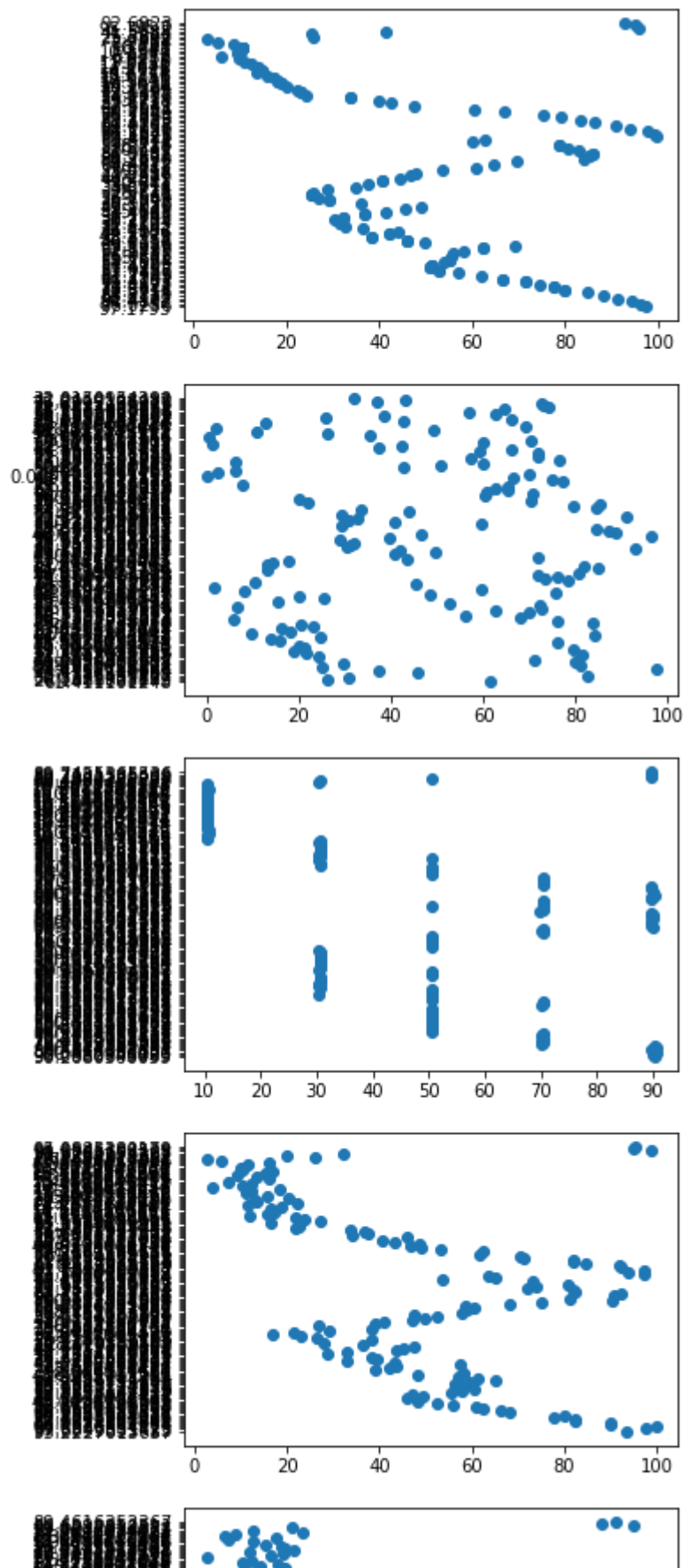
```
1 roots=[]
2 terms=[]
3 count=0
4 for i in dataset:
5     if count!=0:
6         if dataset[count]!=dataset[count-1]:
7             roots.append(count)
8             terms.append(i)
9
10     count+=1
11 roots.append(len(x))
12
```

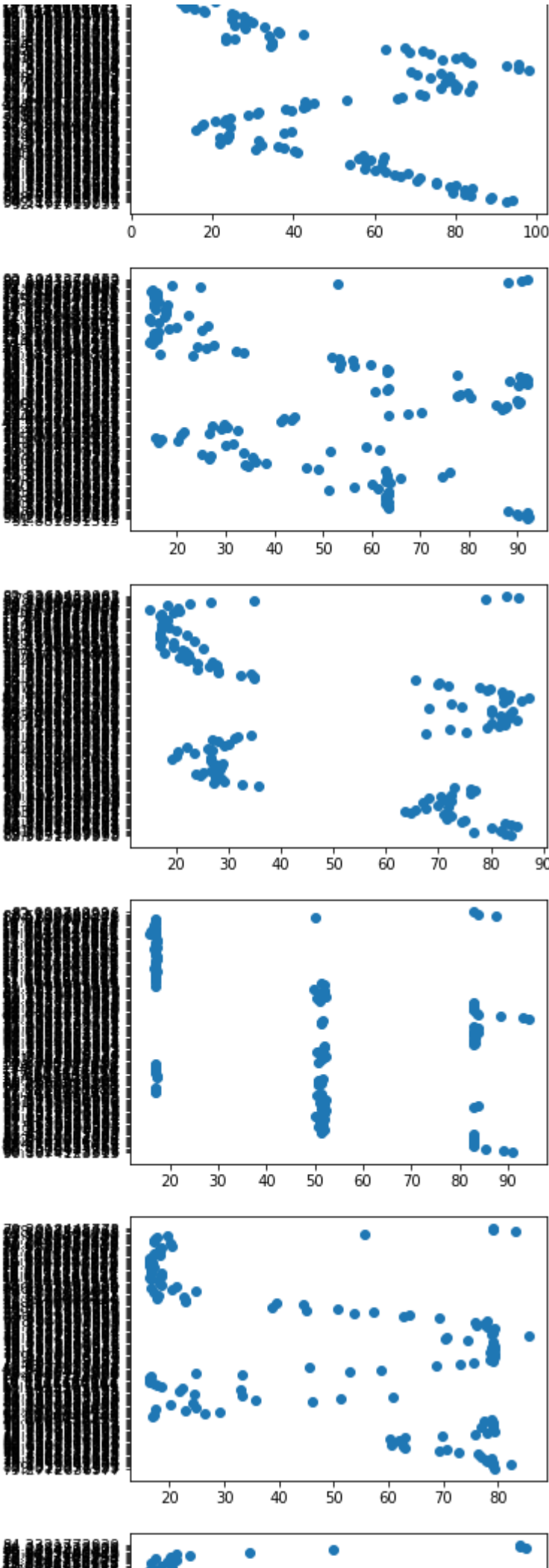
In [116]:

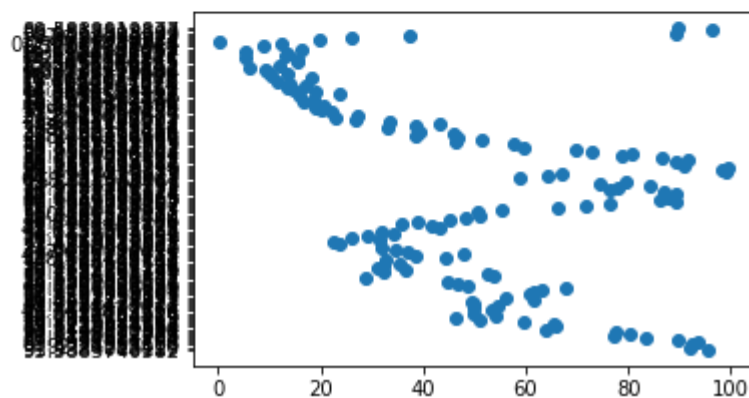
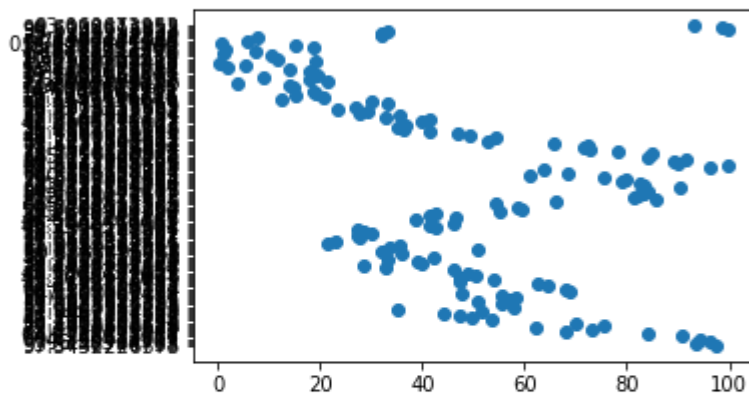
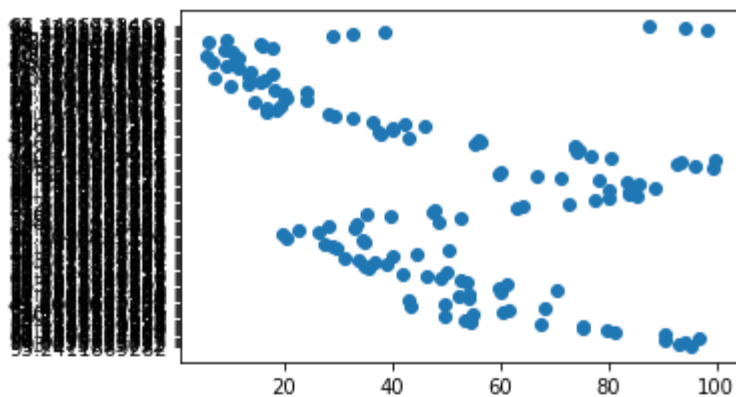
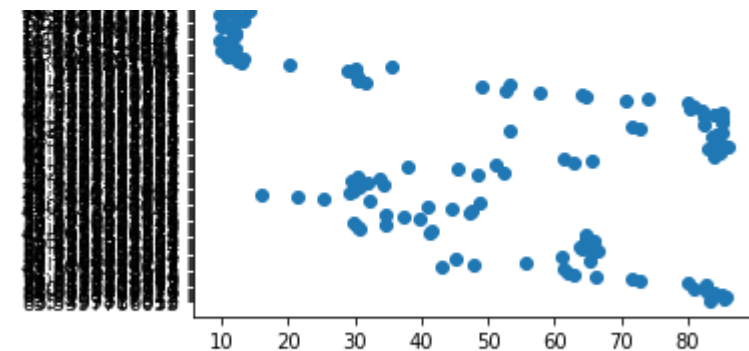
```

1 import matplotlib.pyplot as plt
2 import numpy as np
3
4 fig, axes = plt.subplots(13,1,figsize=(5,50))
5 for i in range(13):
6     axes.flat[i].scatter(x[roots[i]:roots[i+1]],y[roots[i]:roots[i+1]])

```







In [111]:

1

In []:

1

In []:

1

In []:

1	
---	--