

Practica 5: Preparando datos con ayuda de Emacs

Barajas Ibarria Juan Pedro

Hermosillo, Sonora a 06 de Marzo del 2018

76256 Empalme, Son. Observations at 2017

1. Introducción a la Actividad 5

En esta actividad haremos uso intensivo del Editor Emacs para preparar los datos de sondeos de la Atmósfera descargados en la actividad anterior los cuales fueron descargados de la página de sondeos climáticos de la Universidad de Wyoming [3], para su análisis posterior con Panda.

Emacs se ha vuelto una herramienta muy útil en esta actividad ya que nos permitió limpiar los archivos de datos y dejarlos listos con la información que necesitábamos.

2. Descripción de los conceptos físicos de CAPE y PW, que se obtienen de los datos de sondeos de la parte alta de la atmósfera

Las siglas CAPE [2] (Convective Available Potential Energy) hacen referencia a la energía potencial disponible para la convección en un momento dado. Se trata de uno de los parámetros convectivos más interesantes de todos aquellos que se derivan de los modelos meteorológicos.

La precipitación es cualquier producto de la condensación del vapor de agua atmosférico que se deposita en la superficie de la Tierra. Ocurre cuando la atmósfera (que es una gran solución gaseosa) se satura con el vapor de agua, y el agua se condensa y cae de la solución (es decir, precipita). El aire se satura a través de dos procesos: por enfriamiento y añadiendo humedad. La precipitación que alcanza la superficie de la tierra puede producirse en muchas formas diferentes, como lluvia, lluvia congelada, llovizna, nieve, aguanieve y granizo.

La lluvia convectiva proviene de nubes convectivas, como los cumulonimbos o cúmulos congestus. Cae como chaparrones con una intensidad que varía rápidamente. La precipitación convectiva cae en un tiempo relativamente corto sobre un área determinada. La mayor parte de la precipitación en zonas tropicales parece ser convectiva; sin embargo, se ha sugerido que también se da la precipitación estratiforme. Los copos de nieve y el

granizo siempre indican convección. A latitudes medias, la precipitación convectiva tiene relación con los frentes fríos (a menudo detrás del frente), las líneas de chubascos y los frentes cálidos con una significativa humedad disponible.

3. Descripción del proceso de limpieza y preparación de los datos para posteriormente analizar con Pandas

Para la limpieza de datos:

1. Primeramente realizamos la actividad anterior donde descargamos 12 archivos de datos los cuales son los datos estadísticos que nos darán la información que necesitamos para pandas.
2. Después de la descarga los preparamos con un script donde recopilamos toda la información, esto también fue parte de la actividad 4 el cual este paso fue el resultado final, dejándonos un archivo con los datos para pasar a emacs.
3. Continuando pasamos a limpiar columna a columna en emacs donde utilizamos los comandos *Esc+(shift+ %)*, *ctrl+w*, *ctrl+y*, *ctrl+space*, con los cuales ubicamos las columnas de datos sin texto y simplemente acomodarlas.
4. Después de limpiar los datos con ayuda de emacs quedan por columnas como veremos en pandas en la siguiente sección.

4. Análisis de datos utilizando Pandas

Con ayuda de pandas para analizar los datos subimos los archivos a pandas, así como se muestra a continuación.

Aquí le pedimos a pandas las bibliotecas que usaremos para analizar los datos.

```
In [01]: import pandas as pd
import numpy as np
from datetime import datetime
```

Después subimos los dos archivos de datos a pandas agregando los títulos de columnas y convirtiendo las columnas como se menciona en los comentarios.

```
In [02]: # Leer archivo de datos
# Convertir la columna CAPE de objeto a número
df00 = pd.read_csv("df2017CAPE_PW_00Z.csv", header=None, names=['Date', 'CAPE',
df00.CAPE=pd.to_numeric(df00.CAPE, errors='coerce')
df00.head()
```

```
Out [02]:
```

	Date	CAPE	PW
0	01 01 2017	10.79	24.62
1	03 01 2017	0.00	15.01
2	05 01 2017	0.00	17.29,
3	06 01 2017	0.00	18.35
4	07 01 2017	0.00	14.58

```
In [03]: # Leer archivo de datos
# Convertir la columna CAPE de objeto a número
df12 = pd.read_csv("df2017CAPE_PW_12Z.csv", header=None, names=['Date', 'CAPE',
df12.CAPE=pd.to_numeric(df12.CAPE, errors='coerce')
df12.head()
```

```
Out [03]:
```

	Date	CAPE	PW
0	01 01 2017	5.88	23.58
1	02 01 2017	0.00	15.66
2	03 01 2017	0.00	20.08
3	04 01 2017	0.00	24.33
4	06 01 2017	0.00	15.25

Aquí pasamos a convertir las cadenas de caracteres. Como se muestra en los comentarios.

```
In [04]: # Convertir la cadena de caracteres 'Date' en variable temporal 'Ndate'
df00['Ndate'] = pd.to_datetime(df00['Date'], format='%d %m %Y')
df00['month'] = df00['Ndate'].dt.month
df00.head()
```

```
Out [04]:
```

	Date	CAPE	PW	Ndate	month
0	01 01 2017	10.79	24.62	2017-01-01	1
1	03 01 2017	0.00	15.01	2017-01-03	1
2	05 01 2017	0.00	17.29	2017-01-05	1
3	06 01 2017	0.00	18.35	2017-01-06	1
4	07 01 2017	0.00	14.58	2017-01-07	1

```
In [05]: # Convertir la cadena de caracteres 'Date' en variable temporal 'Ndate'
df12['Ndate'] = pd.to_datetime(df12['Date'], format='%d %m %Y')
df12['month'] = df12['Ndate'].dt.month
df12.head()
```

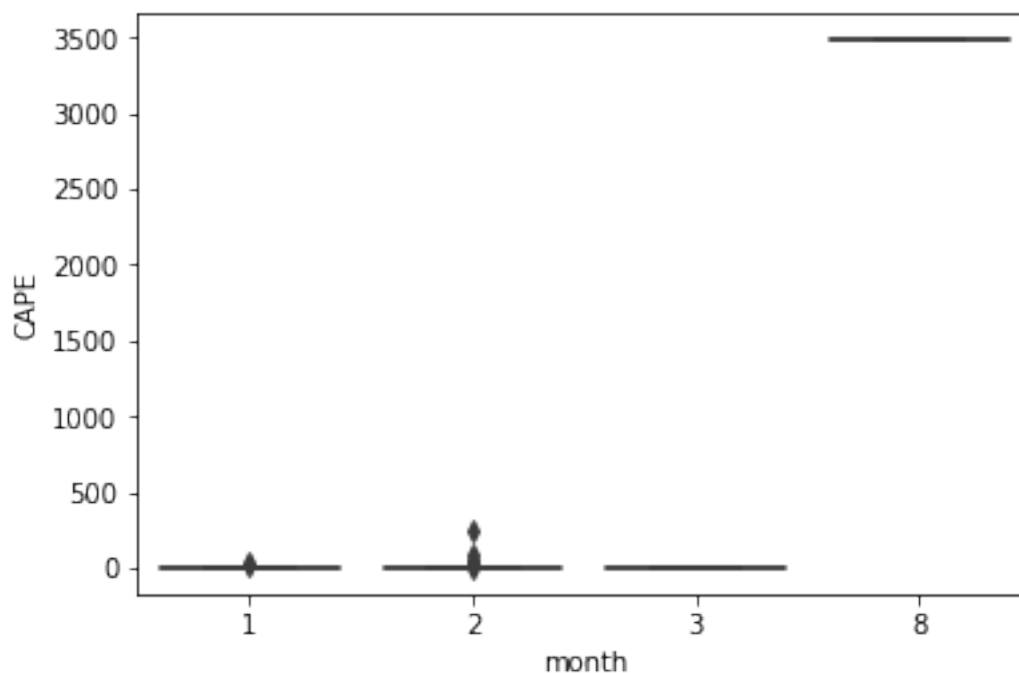
```
Out [05]:
```

	Date	CAPE	PW	Ndate	month
0	01 01 2017	5.88	23.58	2017-01-01	1
1	02 01 2017	0.00	15.66	2017-01-02	1
2	03 01 2017	0.00	20.08	2017-01-03	1
3	04 01 2017	0.00	24.33	2017-01-04	1
4	06 01 2017	0.00	15.25	2017-01-06	1

5. Resultados del análisis

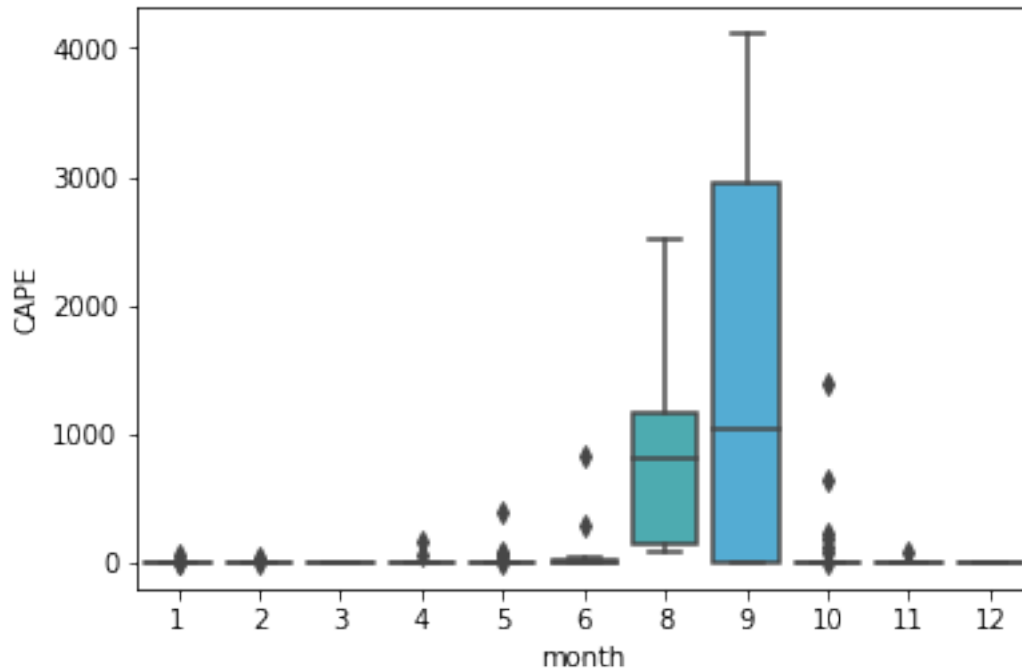
En los resultados del análisis primero graficamos CAPE por meses en donde vemos que los datos fueron un poco pequeños en las gráficas de caja de 00Z entonces no encuentro interpretación.

```
In [06]: # graficar Boxplots por mes
# Biblioteca Seaborn
import seaborn as sns
import matplotlib.pyplot as plt
ax = sns.boxplot(x="month", y="CAPE", data=df00)
plt.show()
```



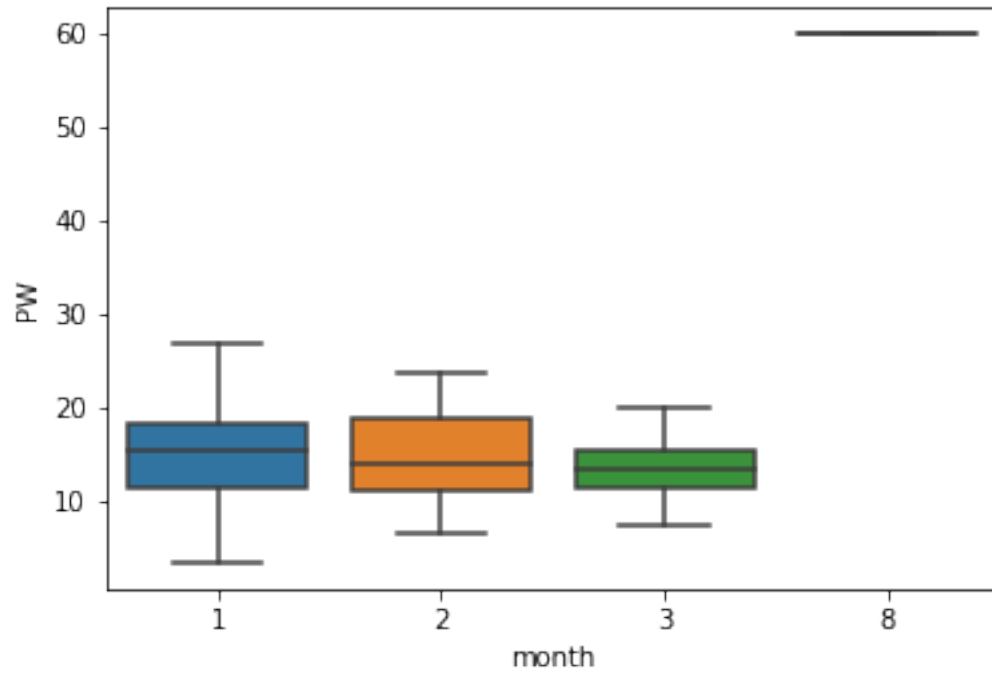
En las graficas de caja 12Z vemos que la distribución no es tan sesgada.

```
In [07]: ax = sns.boxplot(x="month", y="CAPE", data=df12)
plt.show()
```

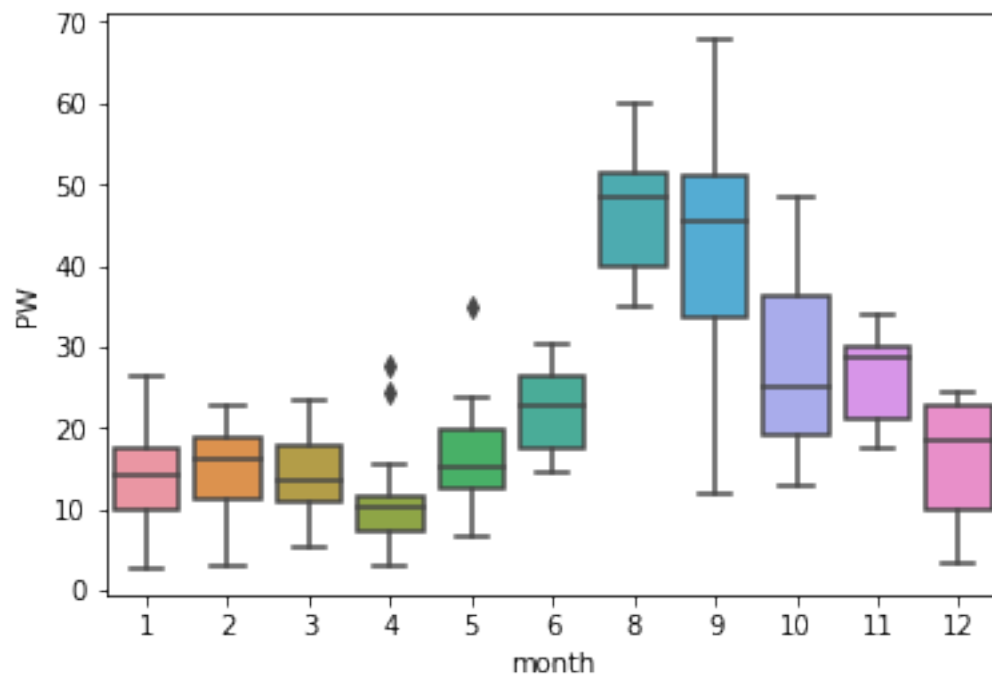


Ahora graficamos las distribuciones en diagramas de caja de la PW(Precipitable water [1]) donde vemos que la primera gráfica de 00Z la distribución no esta muy sesgada, en cambio en 12Z hay mas variaciones y mas en agosto, septiembre, octubre y en noviembre aun mas esta sesgada la distribución.

```
In [08]: # graficar Boxplots por mes
# Biblioteca Seaborn
import seaborn as sns
import matplotlib.pyplot as plt
ax = sns.boxplot(x="month", y="PW", data=df00)
plt.show()
```



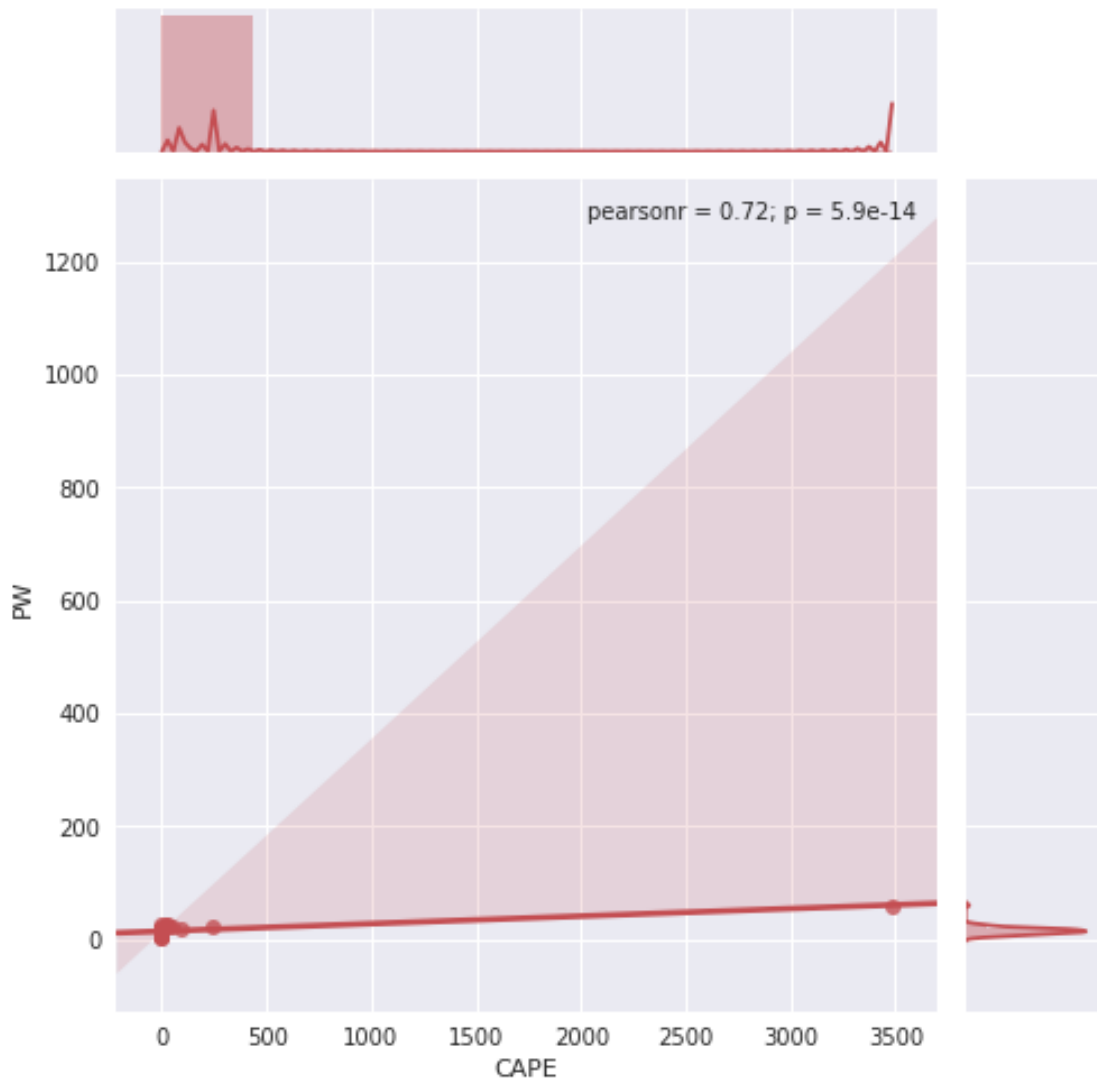
```
In [09]: ax = sns.boxplot(x="month", y="PW", data=df12)
plt.show()
```



Aquí vemos las gráficas de relación entre PW y CAPE donde las distribuciones son menores en 00 pero en 12Z es mayor la relación y distribución por la relación entre pearsonr y p.

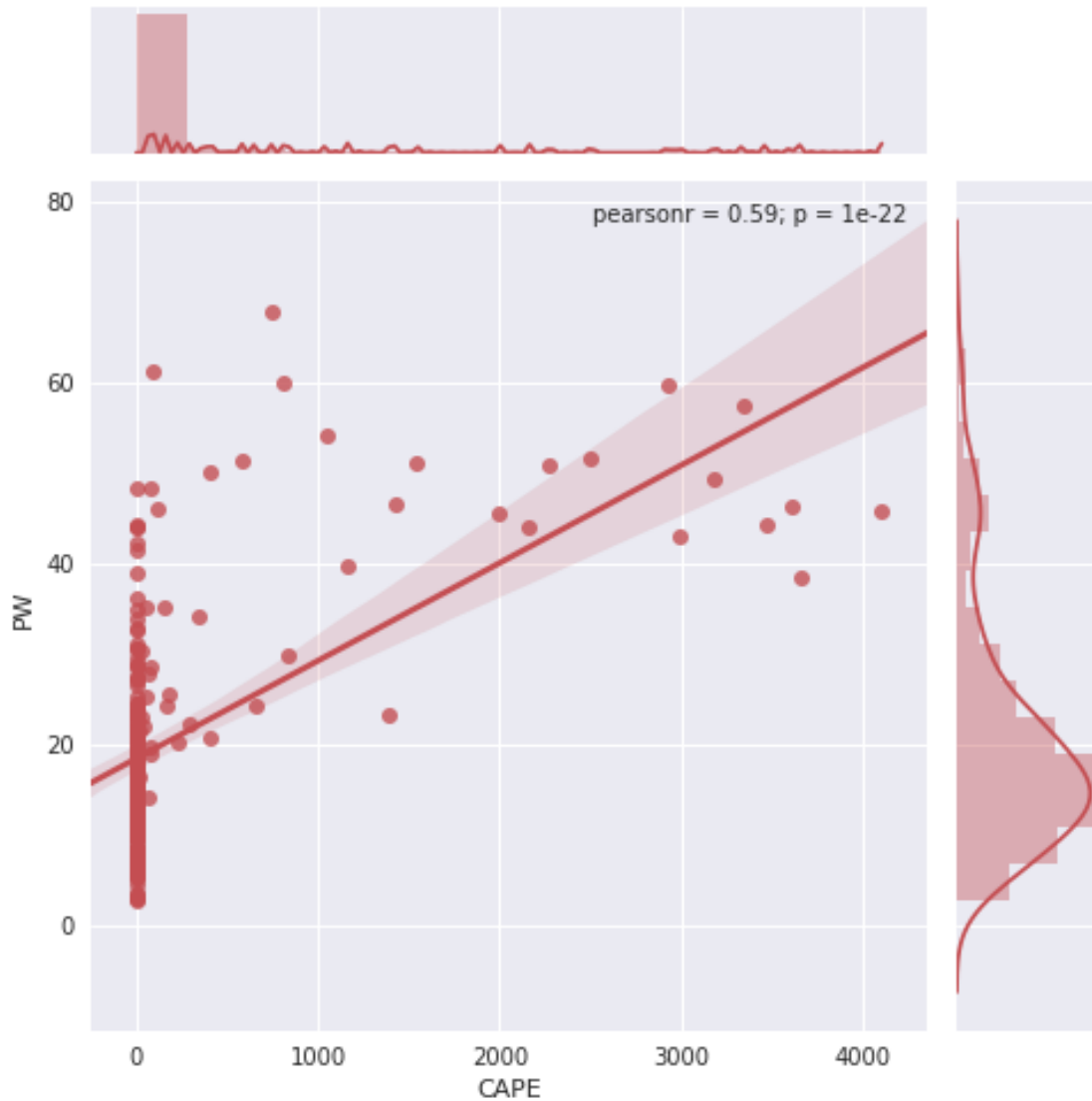
```
In [10]: import seaborn as sns
sns.set(style="darkgrid", color_codes=True)

g = sns.jointplot("CAPE", "PW", data=df00, kind="reg",
                  color="r", size=7)
plt.show(g)
```

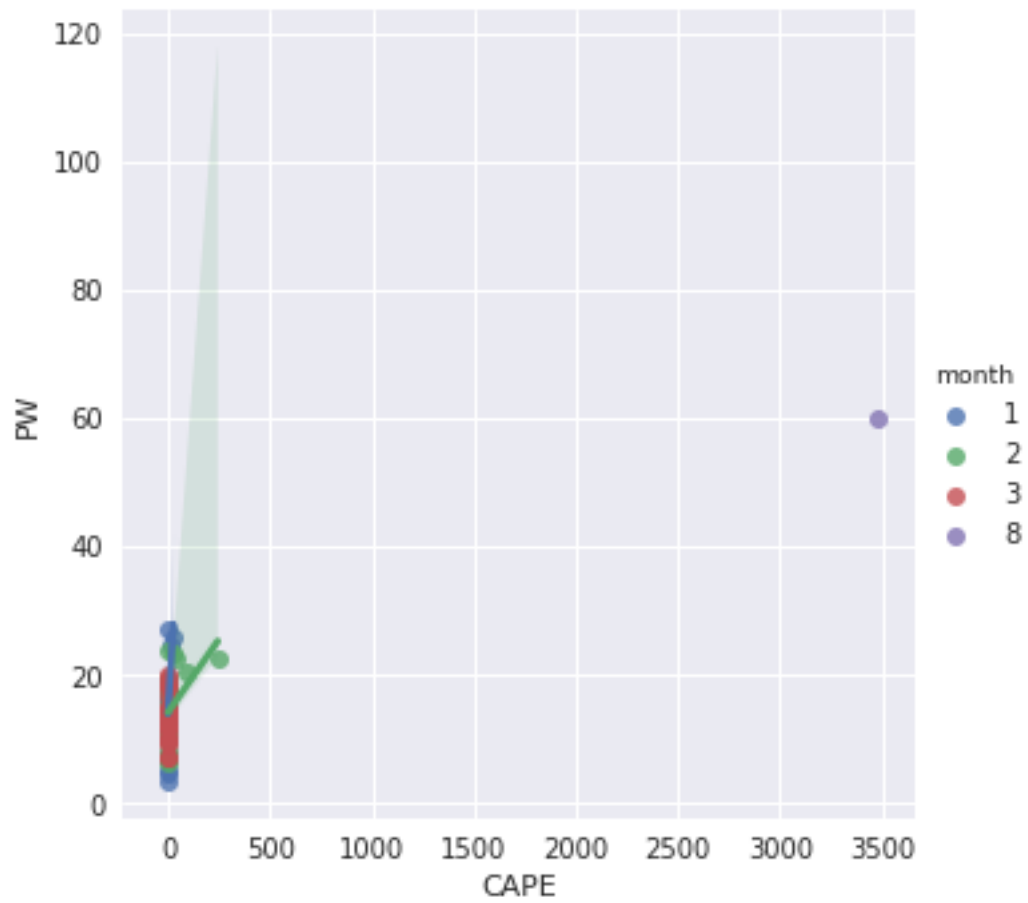


```
In [11]: import seaborn as sns
sns.set(style="darkgrid", color_codes=True)
```

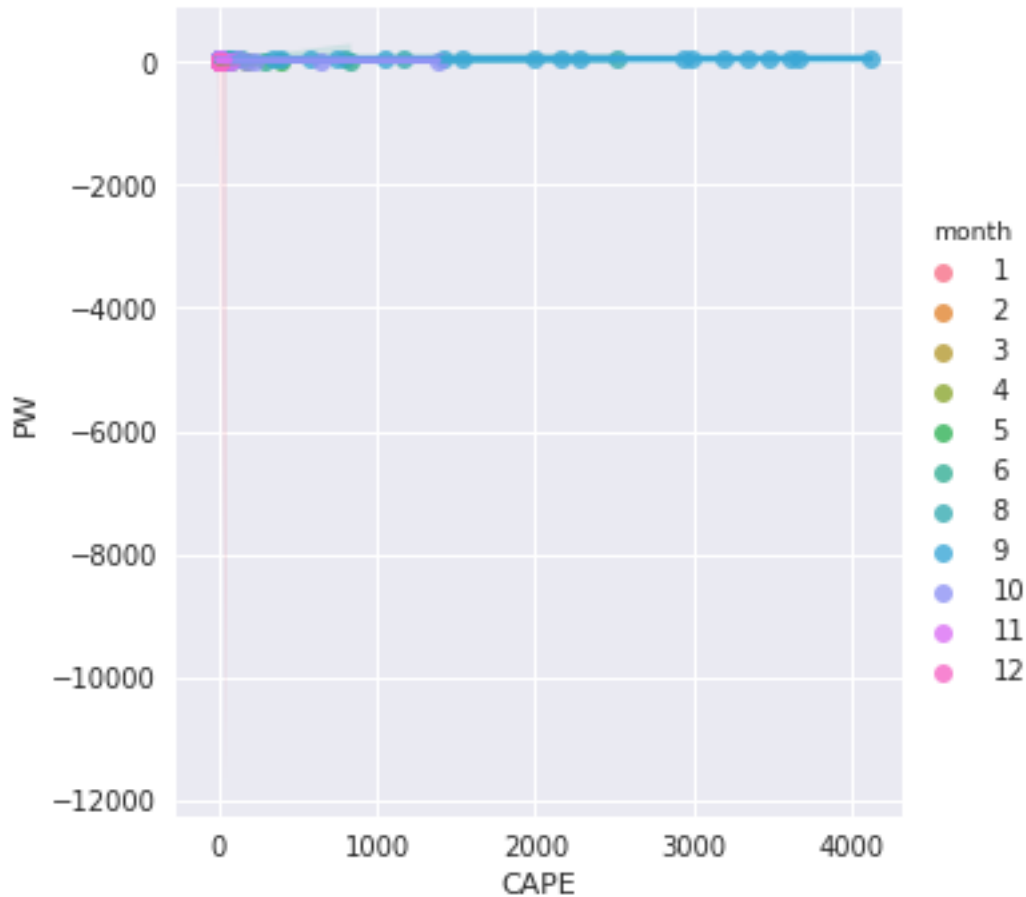
```
g = sns.jointplot("CAPE", "PW", data=df12, kind="reg",
                  color="r", size=7)
plt.show(g)
```



```
In [12]: g = sns.lmplot(x="CAPE", y="PW", hue="month",
                        truncate=True, size=5, data=df00)
plt.show(g)
```

```
In [13]: g = sns.lmplot(x="CAPE", y="PW", hue="month",  
                        truncate=True, size=5, data=df12)  
plt.show(g)
```



6. Conclusiones

Podemos concluir que la relación atmosférica en esta ciudad con los datos CAPE y PW es mayor en los lanzamientos 12Z

Ademas que el uso de pandas esta vez para métodos estadísticos ayudo bastante ya que la interpretación fue mucho mas clara.

Referencias

- [1] Precipitable water. https://en.wikipedia.org/wiki/Precipitable_water.
- [2] admin/anon. Cape y algunas nociones básicas sobre convención atmosférica. <http://www.meteoillesbalears.com/?p=623>, Mayo 2016.
- [3] Department of Atmospheric Science University of Wyoming. Weather sounding. <http://weather.uwyo.edu/upperair/sounding.html>, 2017.

Apéndice

1. ¿Cómo se te hizo esta actividad? ¿Compleja, Difícil, Sencilla?
 - Sencilla, solo que termine tarde por la falta de tiempo.
2. ¿Qué te llamó más la atención?
 - El uso de emacs y la interpretación en pandas para utilizar métodos estadísticos
3. ¿Qué parte fue la que menos te interesó hacer?
 - El reporte.
4. ¿Cómo mejorarías esta actividad? ¿Qué le faltó? ¿Qué sobró??
 - Utilizar otro tipo de datos, otro tipo de relación.
5. ¿Hasta este punto, que te parece el uso de Jupyter para programar en Python?
 - Esta muy bien, interfaz agradable y sencilla.