

Actividad 5

Jesús Adrián Zatarain Alvarado

6 de marzo del 2018

1 Introducción

En esta práctica se tiene como objetivo el conocimiento del uso de ciertos comandos de emacs. Se continúa desde donde se dejó la actividad anterior, con los datos recolectados. Se hará uso intensivo de comandos de recortes para filtrar ciertas palabras. Al final se analizará el archivo en Pandas y se realizarán algunas gráficas relacionadas a los datos filtrados. En total se tendrán dos archivos; uno con los datos de cierta hora del día y otros con otra.

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 (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 y por este motivo lo intentaremos explicar a lo largo de las siguientes líneas.

Es importante empezar remarcando que se trata de un parámetro que nos indica cuanta energía está disponible para la convección en caso de que esta se inicie. Por lo tanto, la consulta de este parámetro se tiene que complementar siempre con la lectura de otros campos del modelo que nos permitan determinar la probabilidad de que la convección se inicie. De hecho, es muy habitual que en los distintos portales meteorológicos se pueda ver ya combinado con otro tipo de información.

El agua precipitable es la cantidad de agua, expresada como altura o masa, que se obtendría si todo el vapor de agua contenido en una columna específica de la atmósfera, de sección transversal horizontal unitaria, se condensase y precipitase.

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

Del archivo obtenido de la actividad anterior, se prosiguió a seleccionar sólo las columnas que coincidieran con las palabras clave "CAPE" y "Precipital Water (PW)", donde creó un archivo con esas variables más la fecha en que fueron recolectados los datos, y también muchos datos que venían por defecto en él.

Tenían los siguientes datos antes de que se filtrará CAPE y PW:

```
<LINK REL="StyleSheet" HREF="/resources/select.css" TYPE="text/css">
<H2>87576 SAEZ Ezeiza Aero Observations at 00Z 01 Jan 2017</H2>
  Station number: 87576
  Showalter index: -6.25
  LIFT computed using virtual temperature: -9.05
  SWEAT index: 351.60
  K index: 19.90
  Totals totals index: 54.00
  CAPE using virtual temperature: 3466.80
  CINS using virtual temperature: -6.27
  LFCT using virtual temperature: 848.36
  Bulk Richardson Number using CAPV: 95.09
  Temp [K] of the Lifted Condensation Level: 293.09
  Precipitable water [mm] for entire sounding: 48.42
<H2>87576 SAEZ Ezeiza Aero Observations at 12Z 01 Jan 2017</H2>
  Station number: 87576
  Showalter index: 1.03
  LIFT computed using virtual temperature: -3.73
  SWEAT index: 207.62
  K index: 30.90
  Totals totals index: 46.20
  CAPE using virtual temperature: 622.65
  CINS using virtual temperature: -101.63
  LFCT using virtual temperature: 796.14
  Bulk Richardson Number using CAPV: 299.60
  Temp [K] of the Lifted Condensation Level: 291.17
  Precipitable water [mm] for entire sounding: 40.67
<H2>87576 SAEZ Ezeiza Aero Observations at 00Z 02 Jan 2017</H2>
  Station number: 87576
  Showalter index: 1.01
  LIFT computed using virtual temperature: -3.51
  SWEAT index: 362.54
  K index: 30.30
  Totals totals index: 45.80
  CAPE using virtual temperature: 1090.55
  CINS using virtual temperature: -32.50
  LFCT using virtual temperature: 809.53
  Bulk Richardson Number using CAPV: 83.62
  Temp [K] of the Lifted Condensation Level: 288.04
  Precipitable water [mm] for entire sounding: 35.46
<H2>87576 SAEZ Ezeiza Aero Observations at 12Z 02 Jan 2017</H2>
  Station number: 87576
  Showalter index: 6.08
  LIFT computed using virtual temperature: 2.71
  SWEAT index: 95.39
  K index: -3.90
  Totals totals index: 39.80
  CAPE using virtual temperature: 0.00
  CINS using virtual temperature: 0.00
  Bulk Richardson Number using CAPV: 0.00
  Temp [K] of the Lifted Condensation Level: 286.31
```

Figure 1:

Lo siguiente consiste en borrar los datos no necesarios para la siguiente parte de la actividad. Sólo se necesitan el CAPE, PW, y la fecha, al igual que su correspondiente hora del día a la que fueron tomados los datos.

Para seleccionar sólo CAPE y PW, se usó el comando:

```
~$ egrep -v 'PRES|hPa' df2017_2.csv | egrep '76225|CAPE|Precip' > df2017_PWs.csv
```

Que sólo filtrará las palabras claves entre comillas y crea un archivo a partir de otro archivo, el nuevo archivo tendrá sólo los caracteres seleccionados.

Posteriormente, se iniciará el proceso de eliminación de texto innecesario en el archivo nuevo, por medio del editor Emacs. Se usarán los comandos intrínsecos en él, que sirven para guardar caracteres en la memoria volátil, la reemplazación de letras por otras y la utilización de un comando que sirve para aplicar los mismo para todos los demás caracteres de la misma condición. Los comandos utilizados están descritos a continuación:

- Control-W: Sirve para guardar el texto seleccionado en la memoria para su posterior utilización
- Control-Y: Se usa para desplegar el texto guardado en la memoria en una parte en específico.
- Esc-shift-5: Sirve para entrar en un modo, donde lo que se encuentra después de éste, puede ser reemplazado por los caracteres que el usuario introduzca por otros a su placer.
- shift-1: La acción a realizar se hace íntegramente en todos los demás de su condición.

Al final el archivo lucirá así:

```

<H2>87576 SAEZ Ezeiza Aero Observations at 00Z 01 Jan 2017</H2>
      Station number: 87576
      CAPE using virtual temperature: 3466.80
Precipitable water [mm] for entire sounding: 48.42
<H2>87576 SAEZ Ezeiza Aero Observations at 12Z 01 Jan 2017</H2>
      Station number: 87576
      CAPE using virtual temperature: 622.65
Precipitable water [mm] for entire sounding: 40.67
<H2>87576 SAEZ Ezeiza Aero Observations at 00Z 02 Jan 2017</H2>
      Station number: 87576
      CAPE using virtual temperature: 1090.55
Precipitable water [mm] for entire sounding: 35.46
<H2>87576 SAEZ Ezeiza Aero Observations at 12Z 02 Jan 2017</H2>
      Station number: 87576
      CAPE using virtual temperature: 0.00
Precipitable water [mm] for entire sounding: 17.83
<H2>87576 SAEZ Ezeiza Aero Observations at 00Z 03 Jan 2017</H2>
      Station number: 87576
      CAPE using virtual temperature: 0.00
Precipitable water [mm] for entire sounding: 35.72
<H2>87576 SAEZ Ezeiza Aero Observations at 12Z 03 Jan 2017</H2>
      Station number: 87576
      CAPE using virtual temperature: 1675.90
Precipitable water [mm] for entire sounding: 53.48
<H2>87576 SAEZ Ezeiza Aero Observations at 00Z 04 Jan 2017</H2>
      Station number: 87576
      CAPE using virtual temperature: 1195.56
Precipitable water [mm] for entire sounding: 38.44
<H2>87576 SAEZ Ezeiza Aero Observations at 12Z 04 Jan 2017</H2>
      Station number: 87576
      CAPE using virtual temperature: 770.89
Precipitable water [mm] for entire sounding: 28.03
<H2>87576 SAEZ Ezeiza Aero Observations at 12Z 05 Jan 2017</H2>
      Station number: 87576
      CAPE using virtual temperature: 0.00
Precipitable water [mm] for entire sounding: 12.85
<H2>87576 SAEZ Ezeiza Aero Observations at 00Z 06 Jan 2017</H2>
      Station number: 87576
      CAPE using virtual temperature: 0.00
Precipitable water [mm] for entire sounding: 15.24
<H2>87576 SAEZ Ezeiza Aero Observations at 00Z 07 Jan 2017</H2>
      Station number: 87576
      CAPE using virtual temperature: 2.10
Precipitable water [mm] for entire sounding: 17.55
<H2>87576 SAEZ Ezeiza Aero Observations at 12Z 07 Jan 2017</H2>
      Station number: 87576
      CAPE using virtual temperature: 497.99
Precipitable water [mm] for entire sounding: 38.03
<H2>87576 SAEZ Ezeiza Aero Observations at 00Z 08 Jan 2017</H2>
      Station number: 87576
      CAPE using virtual temperature: 1771.81

```

Después, se separará el archivo en dos, cada uno con una hora diferente del día. Para ello se utiliza el comando grep como sigue a continuación:

```
~$ egrep -v '00Z' df2017_2.csv > df2017_PW_12Z.csv
```

Que recortará todas las líneas que coincidan con "00Z" en el archivo. Lo que quede en el archivo se colocará en un nuevo archivo que se usará más adelante. Para los correspondientes a "12Z", se prosigue a realizar casi lo mismo; sólo que se cambia las palabras clave y el nombre del archivo para su diferenciación y no se reemplaze por el anterior creado.

```
~$ egrep -v '12Z' df2017_2.csv > df2017_PW_00Z.csv
```

Ya con los dos archivos creados, filtrados y listos para usarse; se prosigue a crear una sesión en Jupyter notebook donde se analizará el archivo por medio de bibliotecas del lenguaje Python.

```
01 01 2017 3466.80 48.42
02 01 2017 1099.55 35.46
03 01 2017 0.00 35.72
04 01 2017 1195.56 38.44
05 01 2017 0.00 19.24
06 01 2017 2.19 17.55
08 01 2017 1771.81 48.98
09 01 2017 1535.83 37.71
10 01 2017 0.00 32.33
12 01 2017 0.00 23.43
14 01 2017 0.00 22.97
15 01 2017 0.00 18.74
17 01 2017 347.28 42.25
18 01 2017 1504.68 32.48
19 01 2017 0.00 19.54
20 01 2017 2671.23 41.10
21 01 2017 30.76 29.01
22 01 2017 334.54 30.09
23 01 2017 475.73 33.55
25 01 2017 291.09 31.35
26 01 2017 0.00 16.38
27 01 2017 0.00 18.29
28 01 2017 0.00 20.15
29 01 2017 0.00 18.17
30 01 2017 171.46 18.78
31 01 2017 1059.45 35.90
01 02 2017 306.61 39.44
02 02 2017 199.70 36.20
03 02 2017 0.00 20.83
04 02 2017 62.71 30.39
05 02 2017 7.01 50.24
06 02 2017 0.00 20.99
07 02 2017 0.00 15.63
08 02 2017 0.00 20.27
09 02 2017 0.00 23.98
10 02 2017 62.16 34.65
11 02 2017 1252.15 32.11
12 02 2017 0.69 29.56
13 02 2017 57.21 35.06
14 02 2017 155.16 47.47
15 02 2017 51.57 48.65
16 02 2017 398.65 28.88
17 02 2017 742.74 39.24
18 02 2017 2111.05 34.86
19 02 2017 650.55 24.39
20 02 2017 1045.42 25.13
21 02 2017 196.47 31.73
22 02 2017 141.90 43.21
24 02 2017 1070.23 28.73
25 02 2017 1711.39 29.78
26 02 2017 1198.84 37.18
```

4 Análisis de datos utilizando Pandas

Para el análisis y creación de gráficas, se utilizó la biblioteca Pandas, que permite analizar archivos de texto. Se usó el código proporcionado por el docente para poder realizar lo faltante de la actividad.

```
import pandas as pd
import numpy as np
from datetime import datetime
```

Que contiene las bibliotecas que se usarán en la actividad. Está la biblioteca de pandas, que es para el análisis del archivo, numpy para el manejo de datos, y por último, datetime, que permite manejar las fechas contenidas en el archivo.

```
# Leer archivo de datos
# Convertir la columna CAPE de objeto a número
df = pd.read_csv("df2017_PV_002.csv", header=None, names=['Date', 'CAPE', 'PV'])
df.CAPE = pd.to_numeric(df.CAPE, errors='coerce')
df.head()
```

Esta parte lee el archivo en cuestión. Se tiene que introducir el nombre del archivo con su correspondiente extensión. Despliega cinco líneas del archivo para verificar si lo ha leído correctamente.

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

Que toma los datos correspondientes a las fechas y los hace utilizables en el posterior análisis y gráficas a realizar.

Ahora se pasa a la parte de graficar los datos, para poder observar su distribución. En la primera se tiene el siguiente código:

```
# graficar Boxplots por mes
# Biblioteca Seaborn
import seaborn as sns
import matplotlib.pyplot as plt
ax = sns.boxplot(x="month", y="Pw", data=df)
plt.show()
```

EL cual va a desplegar una gráfica que contiene los datos registrados de agua precipitada por mes. Los siguiente en la actividad es la realización de una gráfica de PW contra CAPE, que tiene todos los datos en forma de puntos en la gráfica, al igual que una recta con ajuste lineal.

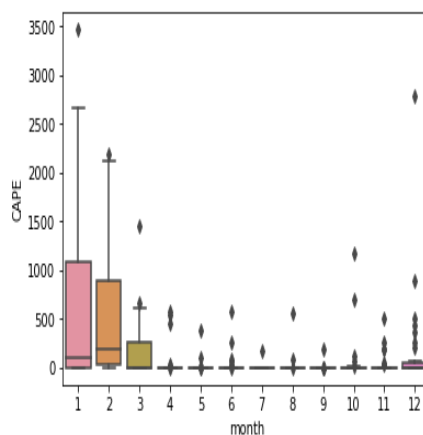
```
import seaborn as sns
sns.set(style="darkgrid", color_codes=True)
g = sns.jointplot("CAPE", "Pw", data=df, kind="reg",
                  color="r", size=7)
plt.show(g)
```

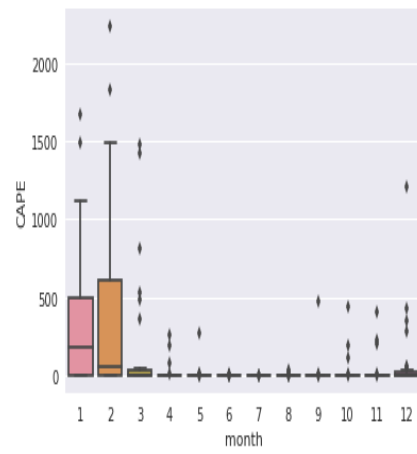
Al final se hizo una gráfica de con el comando: `sns.lmplot`, para hacer la gráfica correspondiente.

5 Resultados del análisis

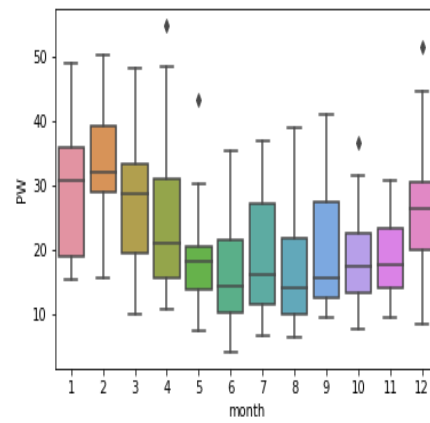
Se leyeron los dos archivos para poder realizar las gráficas correspondientes.

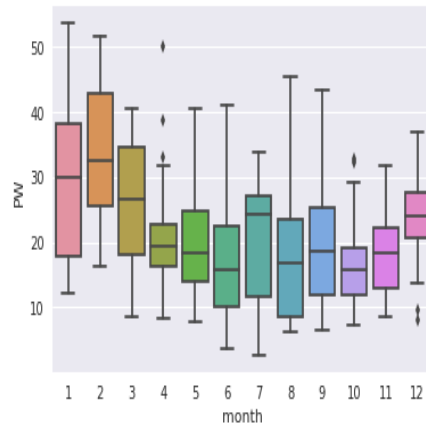
La primera fue por medio de `boxplot`. En primera, se tiene a los datos correspondientes a 00Z ,seguidos de los 12Z. Las gráficas a continuación son para el CAPE. Las dos son diferentes, y se presetan en forma de caja.



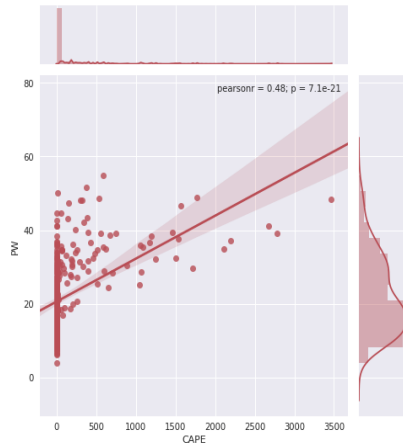


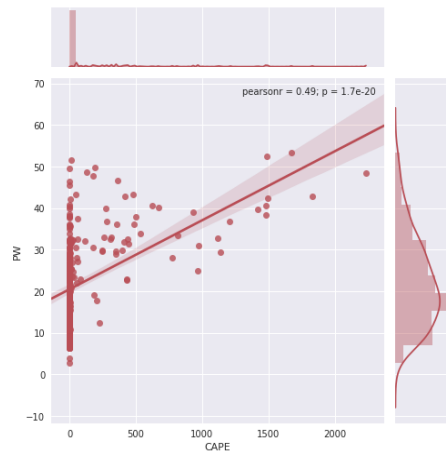
En la siguiente, se tienen los datos correspondientes a el agua precipitada por mes para cada uno.



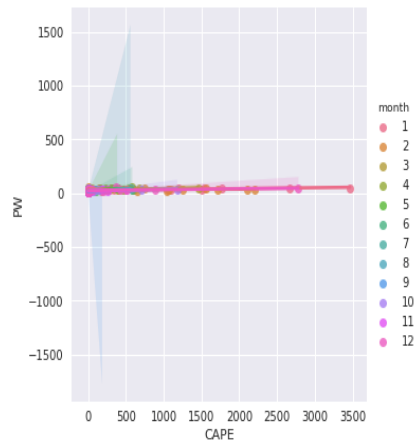


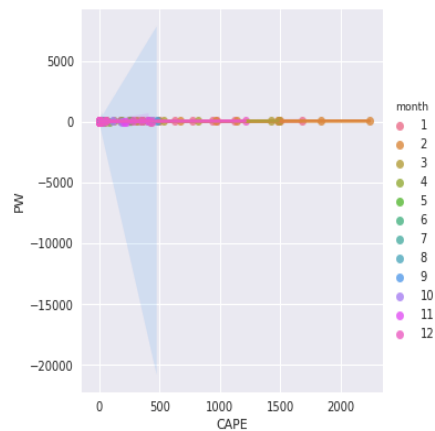
Las dos gráficas a continuación se asemejan mucho. Son gráficas de PW contra CAPE que muestran los datos en forma de puntos y hace un ajuste lineal de ellos. En la parte superior se tiene el CAPE respecto al tiempo, al lado derecho el PW respecto al tiempo, su distribución





Por último se tienen estas dos gráficas. Muestran la dispersión de los datos por medio de los triángulos tenues.





6 Conclusiones

Fue una práctica donde se aprendió mucho sobre los comandos de Emacs, que puede servir para recortar caracteres con patrones similares, y hacer los recortes necesarios para poder utilizar los datos correspondientes con más facilidad. También se vio un poco sobre graficación de datos, que despliega una nueva forma no conocida hasta entonces por el estudiante.

7 Bibliografía utilizada

<http://www.meteoillesbalears.com/?p=623>

8 Apéndice

¿Cómo se te hizo esta actividad? ¿Compleja, Difícil, Sencilla?

Fue sencilla, pero tardada; más por los nuevos comandos que se aprendieron.

¿Qué te llamó más la atención?

La fácil utilización de Emacs para la manipulación de archivos. Al igual que las distintas gráficas que se hicieron.

¿Qué parte fue la que menos te interesó hacer?

Todo estuvo bien

¿Cómo mejorarías esta actividad? ¿Qué le faltó? ¿Qué sobró?

Fue buena actividad, no le faltó nada

¿Hasta este punto, que te parece el uso de Jupyter para programar en Python?

Es una buena herramienta que siempre puedes tener disponible.