

UNIVERSIDAD DE SONORA

---

## Actividad No.3

### Manejo de datos con en Python

---

*Alumno:*  
Paul Cañez Miranda

*Profesor:*  
Carlos Lizzárraga Celaya

16 de febrero de 2017



## Resumen

En esta práctica aprenderemos formas de manejar datos, utilizando bibliotecas de Python.

## Introducción

El uso de Emacs y los filtros de Bach, obtendremos y "limpiaremos" los datos sobre la atmósfera previamente descargados de la página de la Universidad de Wyoming.. Los datos que manejaremos serán CAPE (Convective Available Potencial Energy) y PW (Precipitable Water), obtenidos durante un año. Después con el uso de Jupyter, haremos tablas y gráficas de los datos para analizarlos de forma más práctica.

## Desarrollo

Primero, con el uso de los filtros egrep y sed, clasificaremos los datos capturados durante un año en dos grupos, uno de cada horario (00Z y 12Z) con los siguiente comandos:

```
cat sondeos.txt | egrep -i "Observations|CAPE|precipitable"| sed -e "/00Z/,+2d" > 12Zanual.txt
cat sondeos.txt | egrep -i "Observations|CAPE|precipitable"| sed -e "/12Z/,+2d" > 00Zanual.txt
```

Donde cat muestra los datos del archivo "sondeos.txt". El filtro egrep sirve para indicar las variables que se desean extraer. Sed es para indicar qué grupo (horario), se quiere quitar (en el primer comando se quita las 00Z); y por último el nombre del archivo.

Después realizaremos tablas y gráficas para los datos CAPE Y PW.

## Jupyter

Jupyter es un programa que proporciona una gran cantidad de herramientas para la informática científica mediante paquetes interactivos que combinan la ejecución de código con la creación de documentos activos. El lenguaje utilizado por Jupyter es Python.

Una vez que se tengan los datos ordenados, con las bibliotecas que ofrece Python tabularemos y graficaremos las variables CAPE y PW.

Las bibliotecas utilizadas serán Matplotlib y Pandas.

- Matplotlib  
Esta biblioteca sirve para hacer muchos tipos de gráficas. Desde plots, gráficas de barras, histogramas, etc.
- Pandas  
Las bibliotecas Pandas sirven para hacer análisis de datos.

## Tablas

Con la ayuda de Pandas, tabulamos (por días) los datos de CAPE Y PW.  
Para esto es necesario primero leer el documento en Jupyter, lo hacemos con el comando:

```
df = pd.read_csv("/home/usuario/Carpeta/archivo.csv", names=['Day', 'CAPE', 'PW'])
```

donde los nombres pertenecen a las columnas. Una vez que corra, escribimos el comando

```
df.head(10)
```

y nos aparecerá una tabla con 10 datos como la siguiente

Date	Day	CAPE	PW
0	02 Oct 2016	3039.83	60.86
1	03 Oct 2016	5422.12	63.33
2	04 Oct 2016	3308.22	59.03
3	05 Oct 2016	3433.50	59.91
4	06 Oct 2016	4492.75	60.57
5	07 Oct 2016	2169.27	57.27
6	08 Oct 2016	3477.26	60.64
7	09 Oct 2016	3941.28	55.10
8	10 Oct 2016	3511.02	60.29
9	11 Oct 2016	2984.33	56.00

Esta tabla no muestra todos los datos del año, sólo muestra los primeros 10 sondeos.

En la siguiente tabla, sin embargo, se muestra datos relevantes; mínimo, máximo, media, cuartiles, etc. de todo un año de sondeos. Esta tabla la generaremos con el comando que aparece arriba de la tabla:

```
df.describe()
```

	CAPE	PW
count	71.000000	71.000000
mean	1299.801268	42.679014
std	1354.650404	10.222792
min	0.000000	23.750000
25%	160.170000	36.645000
50%	898.960000	40.710000
75%	1987.460000	50.230000
max	5422.120000	63.330000

Con esta tabla es más sencillo y útil analizar los datos.

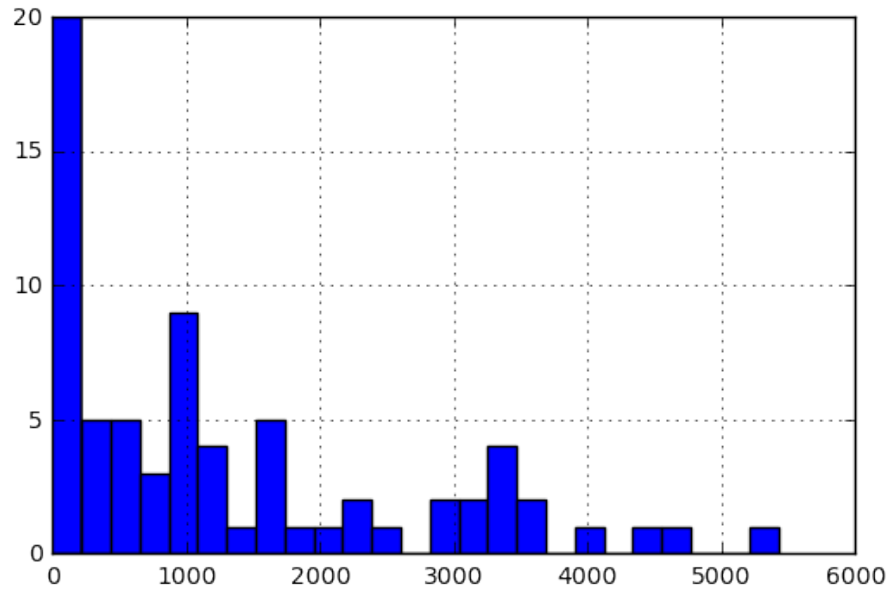
## Gráficas

Para realizar gráficas en Jupyter es necesario introducir el comando `matplotlib inline`

Que es una de las bibliotecas de Python.

En la primer gráfica vemos la cantidad de energía potencial disponible para la convektividad (CAPE) en función de la altura. El comando para obtener esta grafica es:

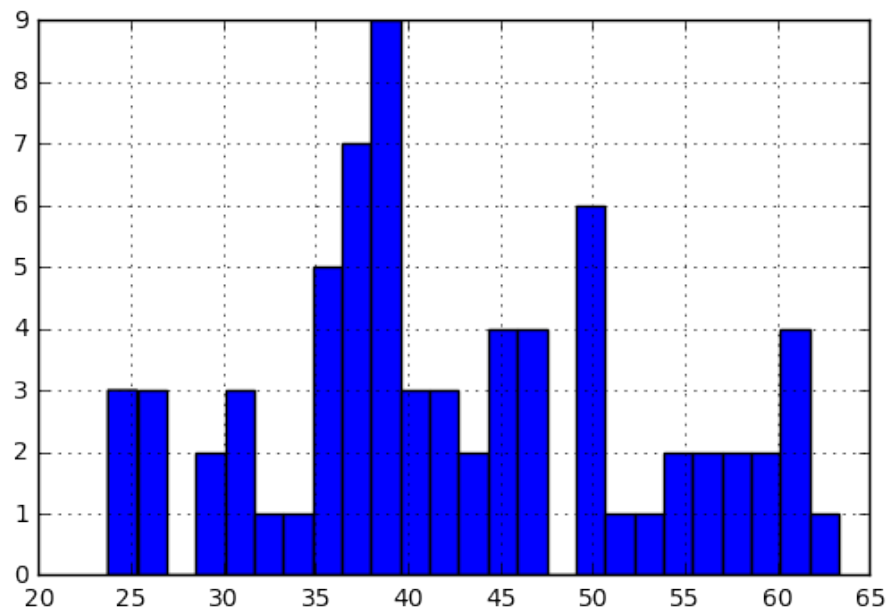
```
df_clean['CAPE'].hist(bins=25)
```



se puede observar cómo el CAPE va disminuyendo con la altura.

En la siguiente gráfica se muestra cómo varia la cantidad de agua precipitable con la altura. Introducimos el siguiente comando para obtenerla:

```
df_clean['PW'].hist(bins=25)
```



podemos observar que el PW es más disperso (varía más) en comparación del CAPE.

Otra forma de analizar de forma sencilla y eficiente un grupo de datos, es con diagramas de cajas. En ellos se puede observar el valor de los cuartiles, media, dispersión, y algunas otras cosas interesantes.

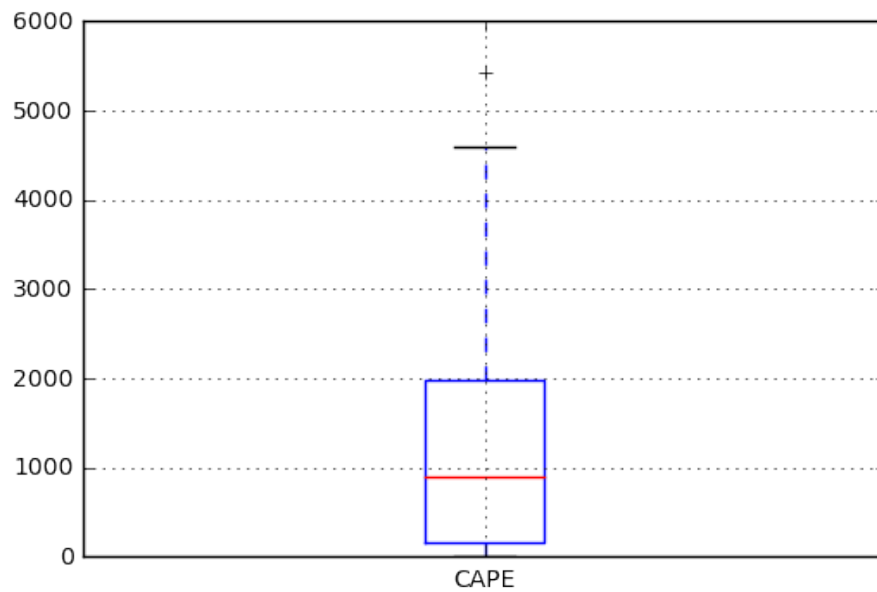
Para graficar éstas cajas de CAPE y PW, utilizaremos los comandos:

```
df_clean.boxplot(column='CAPE')
```

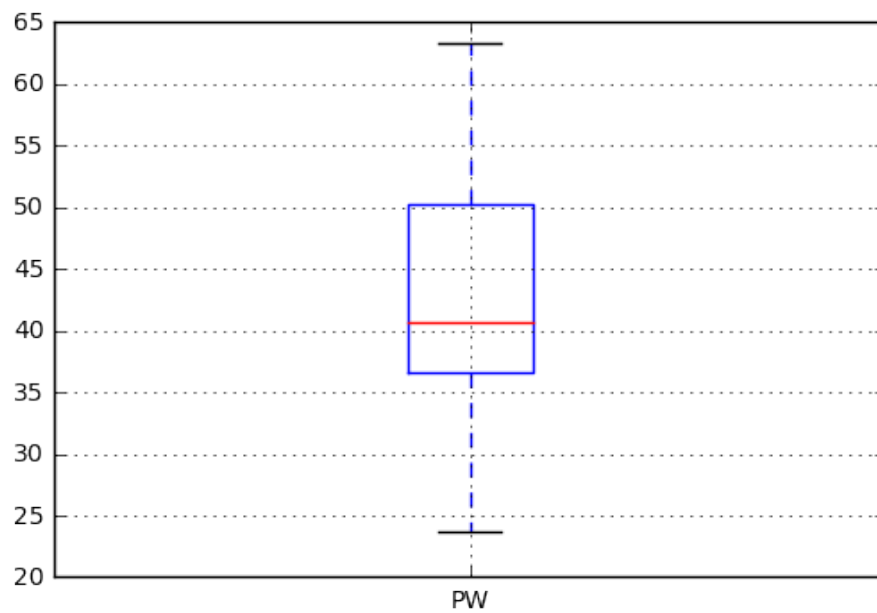
```
df_clean.boxplot(column='PW')
```

respectivamente.

Caja de CAPE



Caja de PW



## Sondeos durante el año 2016

Mes	00Z	12Z
Enero	0	22
Febrero	0	18
Marzo	0	26
Abril	0	29
Mayo	0	15
Junio	0	26
Julio	0	29
Agosto	0	28
Septiembre	0	29
Octubre	21	30
Noviembre	27	30
Diciembre	0	0

Cuadro 1: observaciones realizadas a las 00Z y 12Z horas.

## Conclusión

Las bibliotecas de Python son de gran utilidad para manejar datos de forma sencilla. Además Jupyter me pareció práctico, porque puede correr todos los comandos a la vez.

## Referencias

- [1] *Wikipedia.org. Convective Available Potential Energy*
- [2] *[http : //matplotlib.org/](http://matplotlib.org/)*