



UNIVERSIDAD DE SONORA

DIVISIÓN DE CIENCIAS EXACTAS Y NATURALES

DEPARTAMENTO DE FÍSICA

LICENCIATURA EN FÍSICA

FÍSICA COMPUTACIONAL I

Iniciándose con Python

Martinez López Lizbeth Vanessa

Profesor del curso
Dr. Carlos Lizárraga

16 de Febrero del 2017

Resumen

En este reporte se muestran los resultados obtenidos al utilizar por primera vez Python. Nuevamente se hace uso de los datos recopilados de la Universidad de Wyoming, para hacer gráficas como histogramas y diagramas de caja, de esta forma podemos hacer un análisis y sacar conclusiones sobre el comportamiento de los datos.

1. Introducción

Python es un lenguaje de programación interpretado, su sintaxis es legible, es decir, con código fácil de leer. Python soporta distintos tipos de programación como orientada objetos, imperativa y funcional.

En Python se puede hacer un gran número de cosas, sin embargo en esta actividad nos enfocamos en la lectura de datos, creación de tablas y graficación de ciertas variables como: CAPE (Energía Potencial Disponible Convectiva o Convective Available Potential Energy por sus siglas en inglés) y Agua Precipitable.

Los datos utilizados fueron los obtenidos por la Universidad de Wyoming, siendo los datos recopilados por la estación de Albuquerque los que he tratado para la realización de esta actividad.

2. Desarrollo

2.1. Obtención y limpieza de datos

En la actividad anterior se recopilaron los datos de sondeo para el año 2016, a las 00Z y 12Z. Para esta práctica se hizo uso del archivo que contiene dichos datos de

sondeo, con el comando `egrep` se obtuvo únicamente la información de fecha, CAPE y agua precipitable. El comando empleado fue el siguiente:

```
cat sondeos.txt | egrep -i "Observations|CAPE|water"| sed -e "/12Z/,+2d"
>> Datos00Z.txt
```

Con este comando se pudo seleccionar los renglones que tenían las palabras "Observations", "CAPE" y "Water", es decir, aquellos que contienen la información requerida. La parte de:

```
sed -e "/12Z/,+2d" >> Datos00Z.txt
```

Permite excluir los datos que son para la hora que no se quiere, en este caso sólo se requieren los de 00Z, por lo que se excluyen aquellos que son para 12Z. Lo mismo se hace para obtener los datos de la hora 12Z.

Ya teniendo los archivos con los datos que se quieren, se hace uso de Emacs para la limpieza de datos, es decir, dejar sólo la fecha y los números que hay en CAPE y agua precipitable. Se hace uso de `Esc + x` y el comando `query-replace`, para reemplazar todos aquellos espacios y palabras con nada antes de la fecha, y con comas entre la fecha y CAPE, y entre CAPE y agua precipitable.

2.2. Tablas y Graficación en Python

Con los archivos limpios, se crean documentos separados por comas (csv) y con estos archivos se trabaja en Python. Por medio de la terminal en la carpeta donde se ubican los archivos, se abre Jupyter y se comienza la programación con el siguiente código:

```
import pandas as pd
import numpy as np
import matplotlib as plt
```

```
df = pd.read_csv("/home/vanemtzl/computacional/Datos de todos los meses/
Actividad 3/Datos00Z.csv")
```

Con esto, importamos todo lo que necesitamos como pandas, numpy y matplotlib, lo siguiente es leer el archivo csv.

```
df.head(20)
```

Con esto, obtenemos una tabla con los datos de los archivos csv, al ser demasiados, sólo pedimos que se muestren 20.

	Fecha	CAPE	Agua Precipitable		Fecha	CAPE	Agua Precipitable
0	01 Jan 2016	0.00	3.18	0	01 Jan 2016	0.00	7.46
1	02 Jan 2016	0.00	6.78	1	02 Jan 2016	0.00	5.85
2	03 Jan 2016	0.00	5.02	2	03 Jan 2016	0.00	5.16
3	04 Jan 2016	0.00	5.37	3	04 Jan 2016	0.00	4.89
4	05 Jan 2016	0.00	9.55	4	05 Jan 2016	0.00	10.49
5	06 Jan 2016	0.00	9.31	5	06 Jan 2016	0.00	11.57
6	07 Jan 2016	84.20	7.32	6	07 Jan 2016	0.00	6.75
7	08 Jan 2016	7.81	8.46	7	08 Jan 2016	0.84	6.45
8	09 Jan 2016	0.00	7.09	8	09 Jan 2016	0.00	6.28
9	10 Jan 2016	0.00	4.94	9	10 Jan 2016	0.00	4.39
10	11 Jan 2016	0.00	5.81	10	11 Jan 2016	0.00	3.00
11	12 Jan 2016	0.00	3.96	11	12 Jan 2016	0.00	2.73
12	13 Jan 2016	0.00	2.76	12	13 Jan 2016	0.00	2.52
13	14 Jan 2016	0.00	2.69	13	14 Jan 2016	0.00	3.14
14	15 Jan 2016	0.00	4.00	14	15 Jan 2016	0.00	5.04
15	16 Jan 2016	0.00	8.11	15	16 Jan 2016	0.00	5.11
16	17 Jan 2016	0.00	6.17	16	17 Jan 2016	0.00	5.80
17	18 Jan 2016	0.00	5.07	17	18 Jan 2016	0.00	5.61
18	19 Jan 2016	0.00	6.22	18	19 Jan 2016	0.00	6.96
19	20 Jan 2016	0.00	5.43	19	20 Jan 2016	0.00	8.98

A la izquierda se encuentra la tabla para 00Z y a la derecha la tabla para 12Z.

```
df.describe()
```

Nos muestra una tabla con información de los archivos, así como la cantidad de datos, los cuartiles, el promedio, dato mínimo y máximo y desviación estándar.

```
df.hist(u'CAPE',bins=20)
```

Esto nos permite hacer un histograma de la variable que queramos, en este caso CAPE y definimos la cantidad de cajas que se quieren, en este caso 20.

	CAPE	Agua Precipitable		CAPE	Agua Precipitable
count	363.000000	363.000000	count	364.000000	364.000000
mean	90.845510	11.589311	mean	23.346346	12.339231
std	185.025789	7.146926	std	86.905569	7.687457
min	0.000000	1.800000	min	0.000000	1.530000
25%	0.000000	5.965000	25%	0.000000	5.997500
50%	0.000000	9.210000	50%	0.000000	10.225000
75%	76.535000	16.490000	75%	0.000000	17.877500
max	1092.790000	32.120000	max	801.010000	34.110000

A la izquierda se encuentra la tabla para 00Z y a la derecha la tabla para 12Z.

```
df.boxplot(column=u'CAPE',return_type='axes')
```

Con este código, podemos hacer un diagrama de cajas para todo el archivo.

```
df['month']=pd.DatetimeIndex(df[u' Fecha']).month
df.boxplot(column=u'CAPE', by=u'month')
```

Con esto, también se hacen diagramas de caja, pero en este caso comparativo. Para este trabajo se hizo un diagrama comparativo de todos los meses.

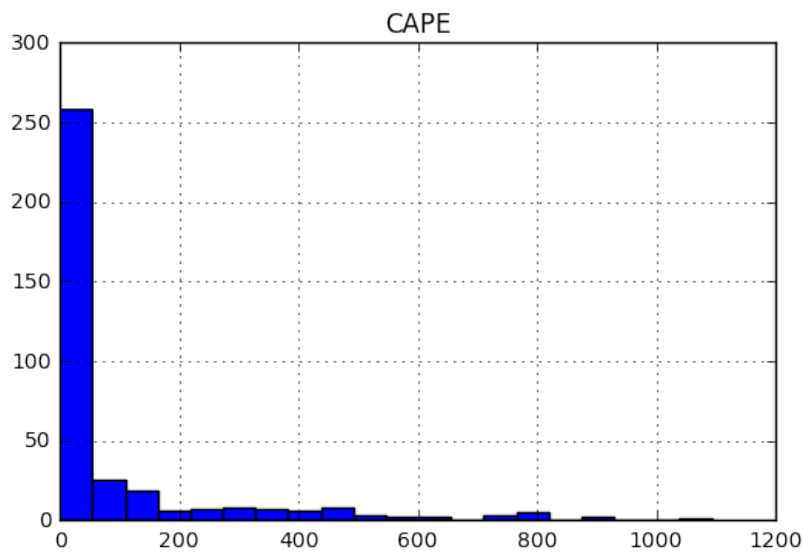
3. Resultados

3.1. CAPE

Las siglas CAPE (Convective Available Potential Energy) hacen referencia a la energía potencial disponible para la convección en un momento dado. Se trata de un parámetro que nos indica cuánta energía está disponible para la convección en caso de que ésta se inicie.

Para el caso de los datos de la estación de Albuquerque, se muestran las siguientes gráficas:

Para 00Z



Histograma

El histograma nos muestra una gran dispersión en los datos para 00Z, pues la mayoría de los datos se concentran en la primer clase (caja).

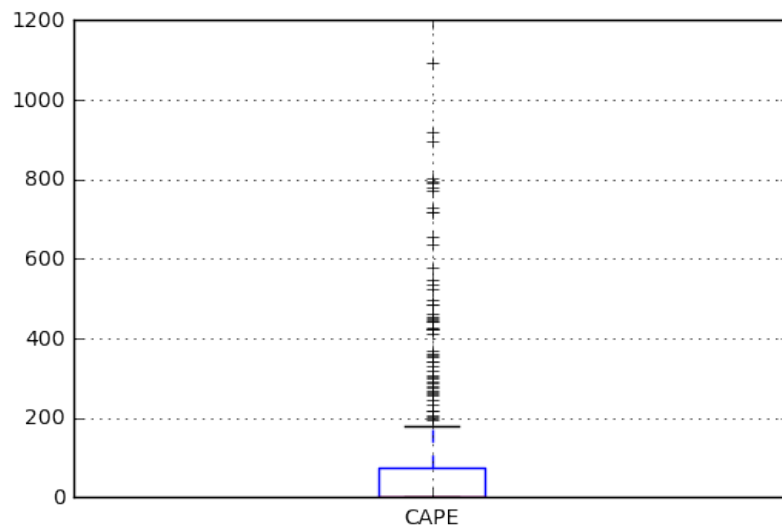


Diagrama de cajas

Este diagrama de cajas está elaborado para todos los datos del archivo. En él se puede observar que hay mucha dispersión en los datos, hay presencia de muchos ceros y es por ello que se detectan bastantes datos atípicos.

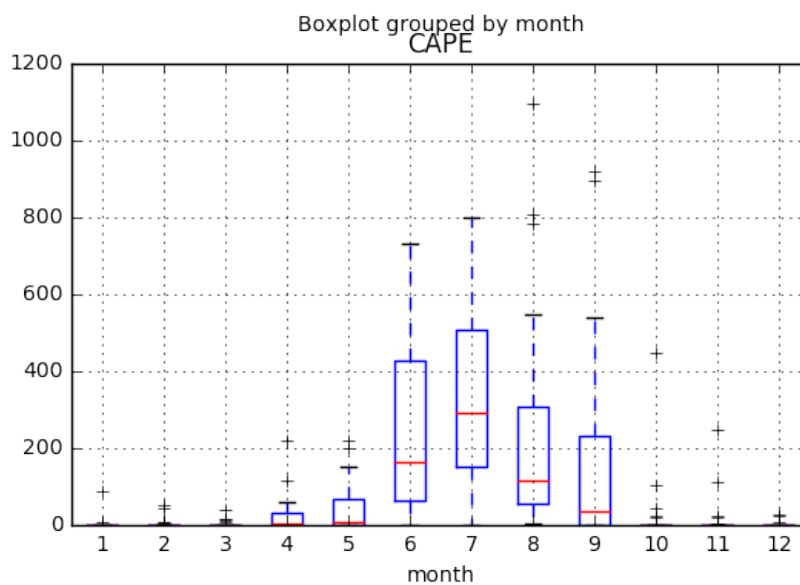
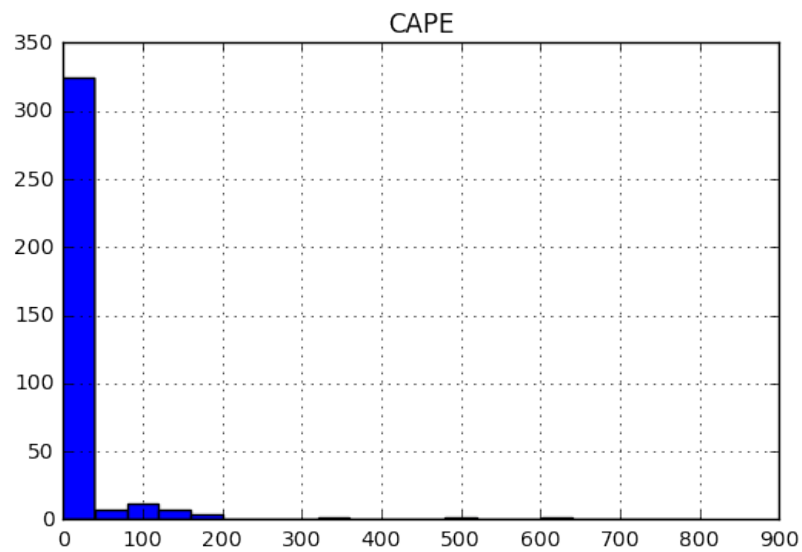


Diagrama de cajas comparativo

Con este diagrama de cajas comparativo, podemos concluir que el mes en el que hay mejor distribución de los datos es en Julio, pues a pesar de que hay dispersión, es menor que en los demás meses.

Para 12Z



Histograma

El histograma nos muestra que hay una gran dispersión en los datos para 12Z, pues la mayoría de los datos se concentran en la primer clase (caja).

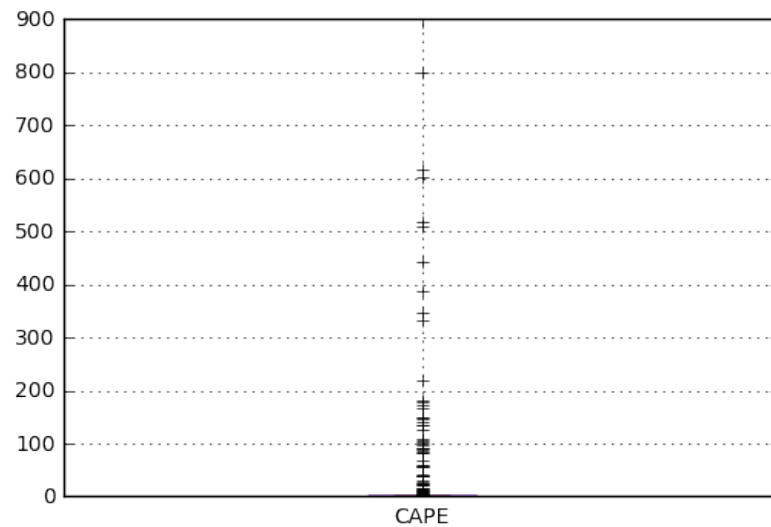


Diagrama de cajas

Este diagrama de cajas está elaborado para todos los datos del archivo. En él se puede observar que hay mucha dispersión en los datos, ni siquiera se presenta la

caja, pues sólo se muestran datos atípicos, esto indica que hay presencia de muchos ceros en los datos.

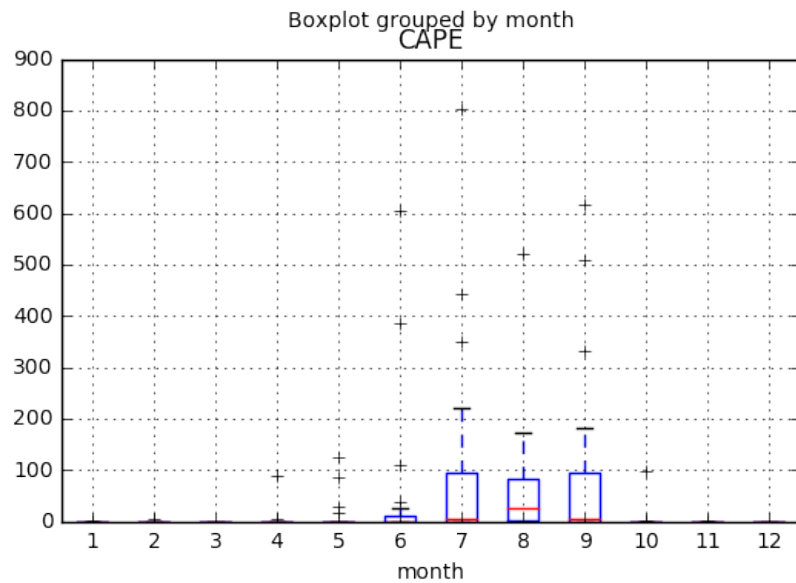


Diagrama de cajas comparativo

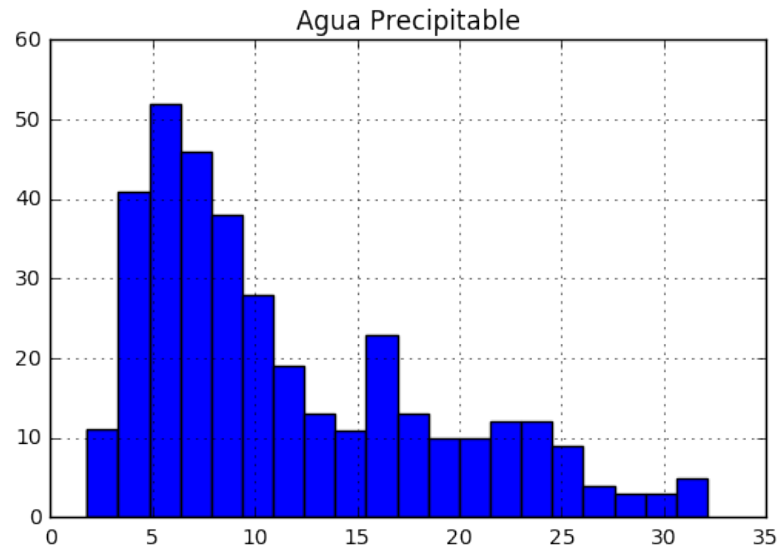
Con este diagrama de cajas comparativo, podemos concluir que el mes en el que hay mejor distribución de los datos es en Agosto, pues a pesar de que hay dispersión, es menor que en los demás meses, además sólo hay presencia de un dato atípico.

3.2. Agua precipitable

El agua precipitable es el contenido de humedad en la atmósfera; se mide como el espesor vertical que ocuparía si toda el agua cayera.

Para el caso de los datos de la estación de Albuquerque, se muestran las siguientes gráficas:

Para 00Z



Histograma

El histograma nos muestra que para el agua precipitable, los datos están mejor distribuidos, sin embargo, la clase 3 (tercer caja) es la que contiene más datos, y las clases 18 y 19 son las que contienen menos datos.

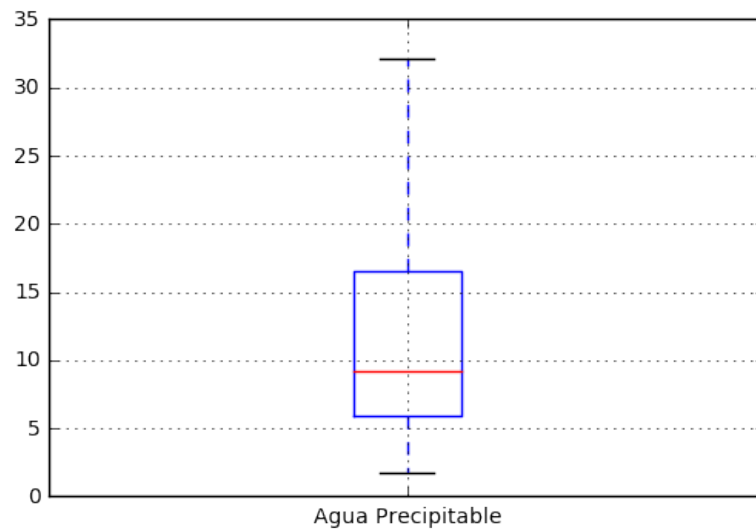


Diagrama de cajas

Este diagrama de cajas está elaborado para todos los datos del archivo. En él se puede observar que hay mayor dispersión en los datos sobre la media, pues hay

mayor concentración de datos hacia abajo de la media (línea roja), esto se puede ver, porque la media tiende más hacia los datos menores.

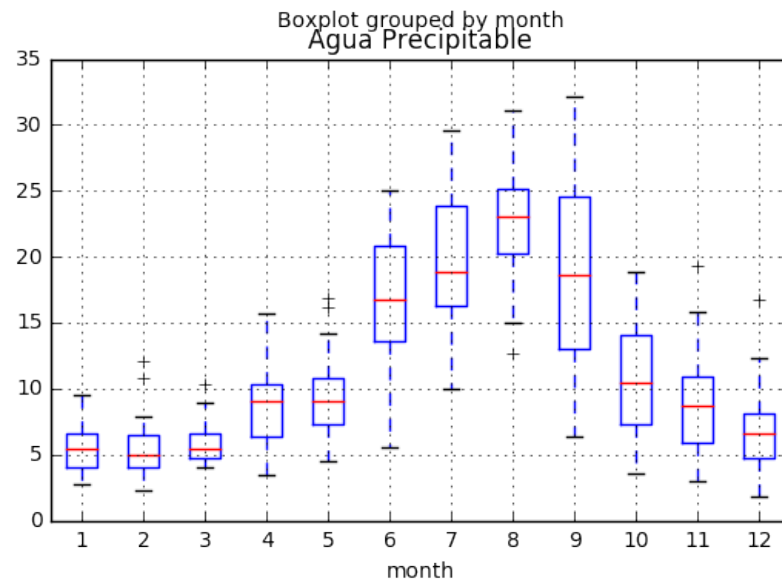
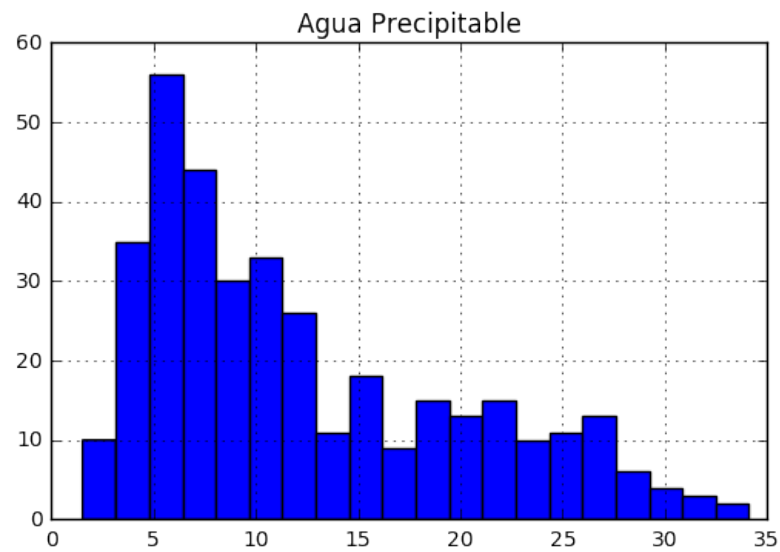


Diagrama de cajas comparativo

En este diagrama de cajas comparativo, podemos ver que para todos los meses los datos se ven distribuidos de manera casi simétrica, sin embargo hay leve dispersión en la mayoría. Se puede concluir que Septiembre y Octubre son los meses que tienen mejor distribución en los datos respecto a la media.

Para 12Z



Histograma

El histograma nos muestra que para el agua precipitable, hay más datos en la clase 3 (tercer caja) y la clase 20 es la que contiene menos datos.

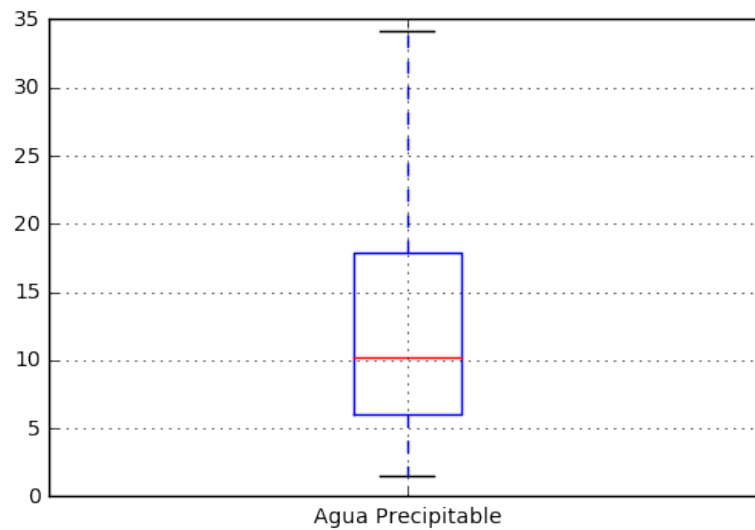


Diagrama de cajas

Este diagrama de cajas está elaborado para todos los datos del archivo. En él se puede observar que hay mayor dispersión en los datos sobre la media, pues hay

mayor concentración de datos hacia abajo de la media (línea roja), esto se puede ver, porque la media tiende más hacia los datos menores.

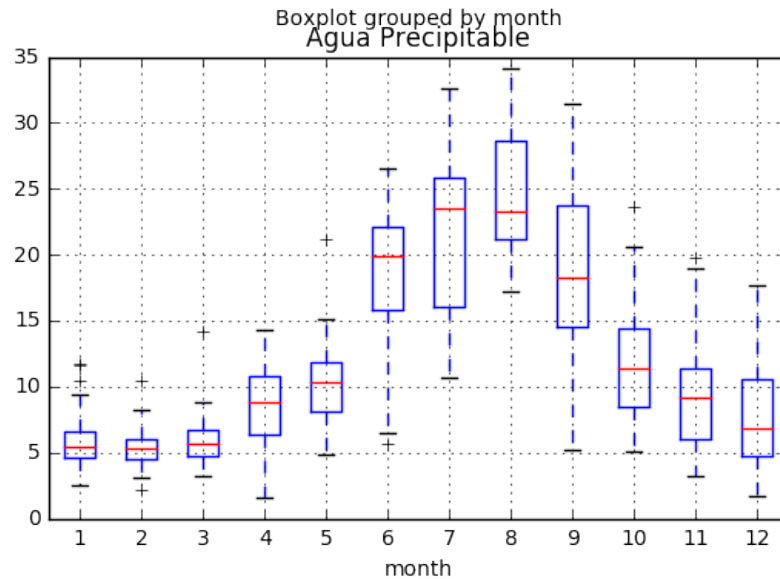


Diagrama de cajas comparativo

En este diagrama de cajas comparativo, podemos ver que es el mes de Octubre el que tiene mejor distribución de los datos respecto a la media, sin embargo hay presencia de datos atípicos. En sí para todos los meses se puede ver dispersión en los datos.

4. Conclusión

Python es un lenguaje de programación bastante sencillo y con código fácil de leer para nosotros. En lo personal, creo que es un buen lenguaje de programación y tiene muchas herramientas que nos son y serán útiles. En esta práctica, se pudieron emplear conocimientos que se han estado adquiriendo en estadística, y dichos conocimientos pudieron ser plasmados gracias a la graficación en Python.

5. Referencias

- [1] *[https : //es.wikipedia.org/wiki/Python](https://es.wikipedia.org/wiki/Python)*
- [2] *[http : //www.meteoillesbalears.com/?p = 623](http://www.meteoillesbalears.com/?p=623)*
- [3] *[http : //weather.wvgo.edu/upperair/sounding.html](http://weather.wvgo.edu/upperair/sounding.html)*
- [4] *[http : //bibliotecadigital.ilce.edu.mx/sites/ciencia/volumen3/ciencia3/127/htm/sec18.htm](http://bibliotecadigital.ilce.edu.mx/sites/ciencia/volumen3/ciencia3/127/htm/sec18.htm)*