



Programación como  
herramienta

# AYUDANTÍA 2

Pandas y limpieza de datos





Programación como  
herramienta

# ¿Para qué usar pandas?

	nombre	apellido	edad	aprobado	nota	examen
0	Alonso	Zamorano	23	True		NaN
1	Francisco	Concha	21	False		2.1
2	Nicolás	Buzeta	18	False		3.7
3	Fernanda	Lillo	20	True		NaN

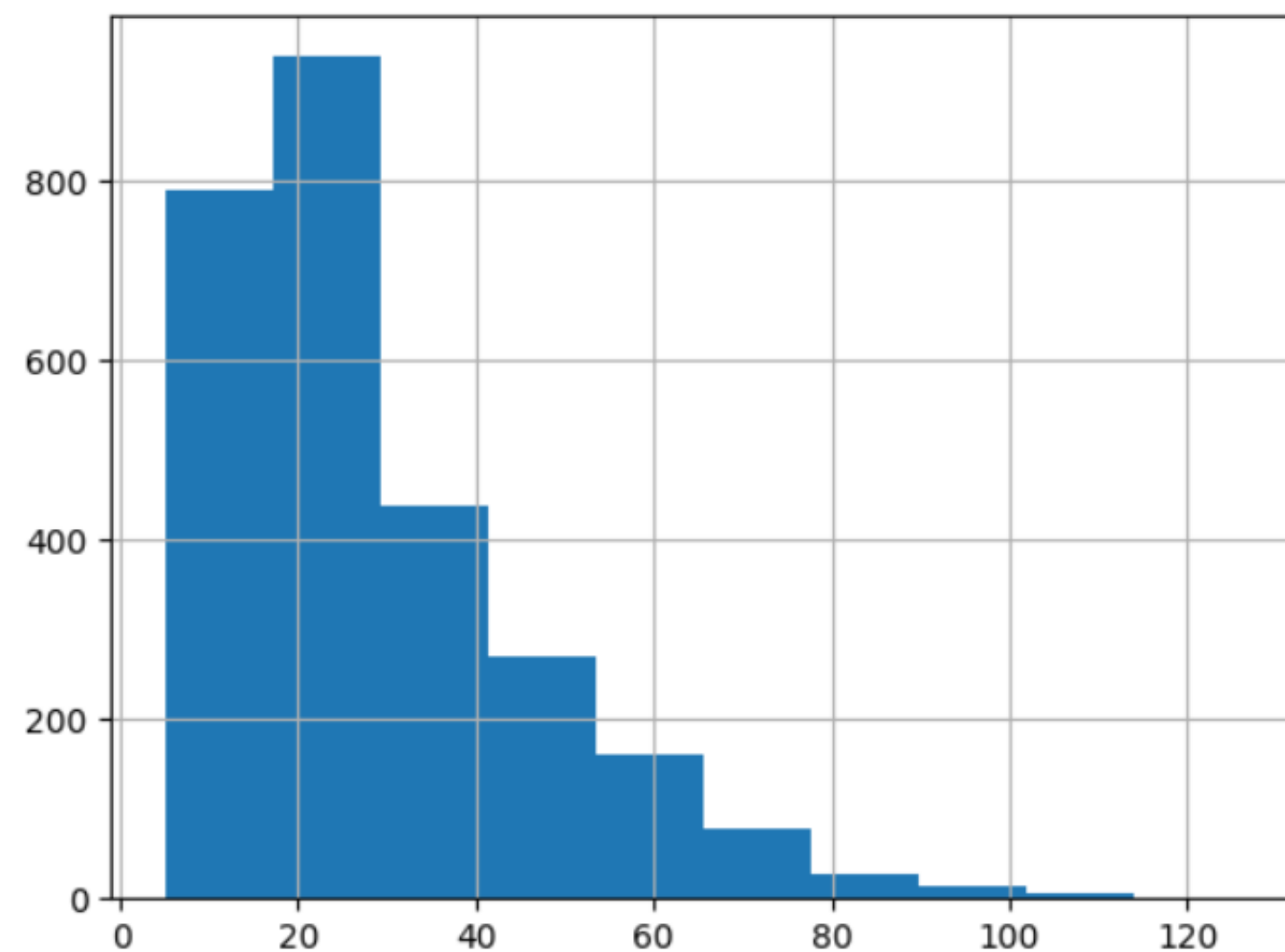
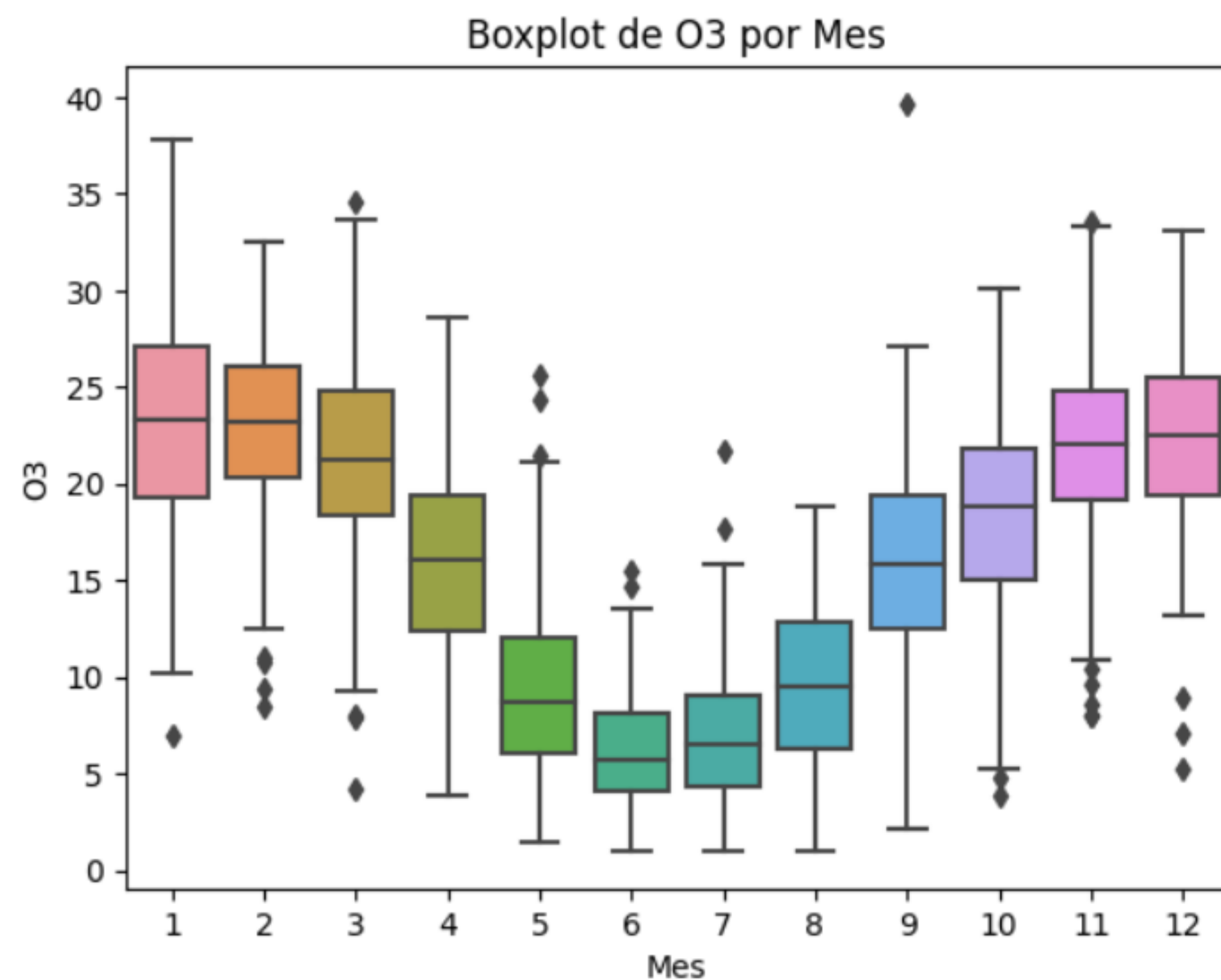
	edad	nota	examen
count	4.000000	2.000000	
mean	20.500000	2.900000	
std	2.081666	1.131371	
min	18.000000	2.100000	
25%	19.500000	2.500000	
50%	20.500000	2.900000	
75%	21.500000	3.300000	
max	23.000000	3.700000	



Programación como  
herramienta

# ¿Qué puedo hacer con pandas?

## Visualización





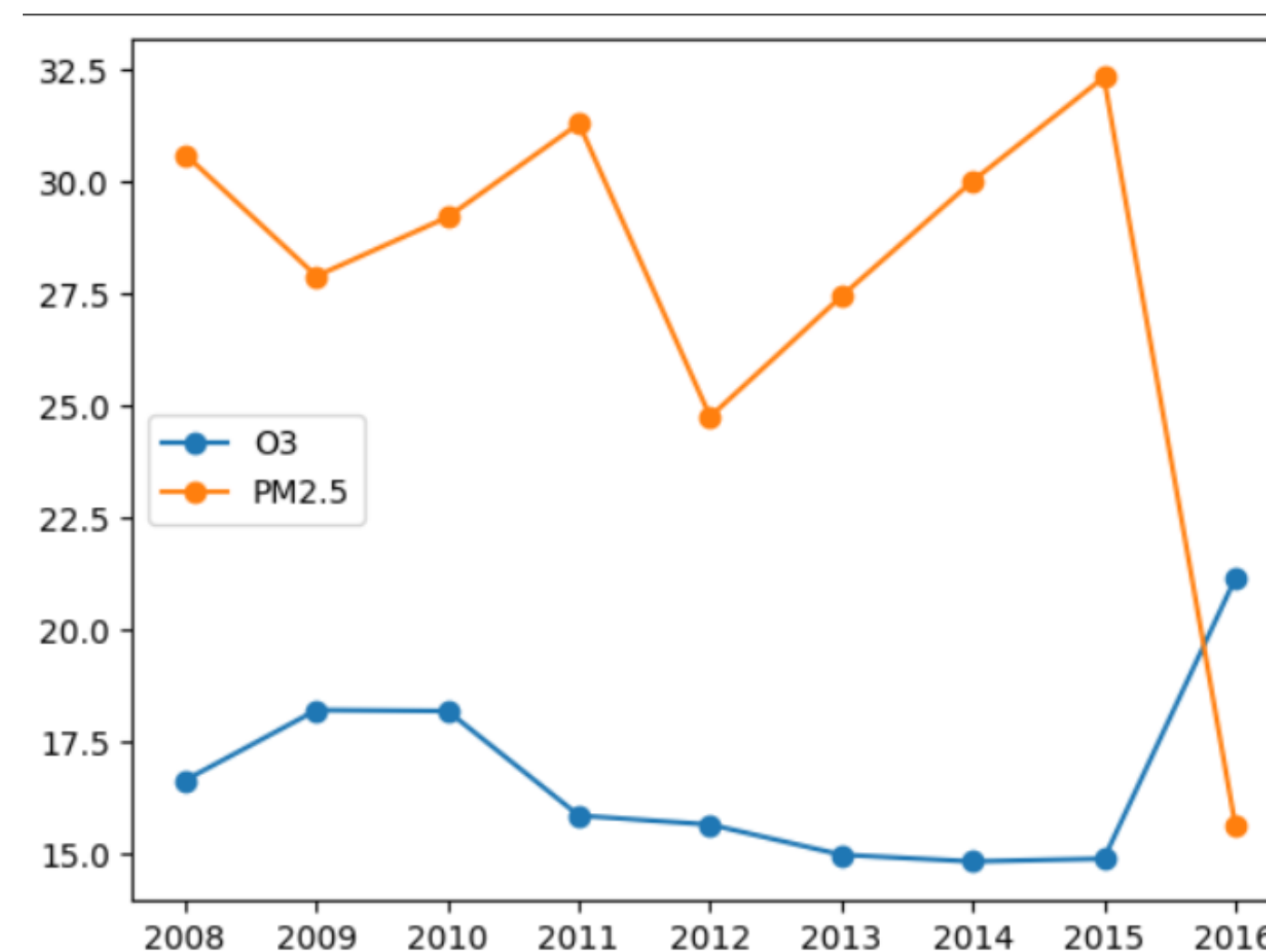
Programación como  
herramienta

# ¿Qué puedo hacer con pandas?

## Exploración

Media O3	Desviación Estándar O3	Máximo O3	Mínimo O3	Kurtosis O3	Media PM2.5	Desviación Estándar PM2.5	Máximo PM2.5	Mínimo PM2.5	Kurtosis PM2.5	Correlación
16.250821	7.857782	39.65	1.0	-0.944356	28.950613	16.868082	126.2	5.0	2.354072	-0.478362

	Year	Month	Day	O3	PM2.5
count	2984.000000	2984.000000	2984.000000	2878.000000	2725.000000
mean	2011.592493	6.419236	15.715818	16.313767	28.814594
std	2.357040	3.487660	8.804066	7.995680	17.479388
min	2008.000000	1.000000	1.000000	1.000000	5.000000
25%	2010.000000	3.000000	8.000000	9.392500	16.170000
50%	2012.000000	6.000000	16.000000	16.960000	23.210000
75%	2014.000000	9.000000	23.000000	22.357500	37.210000
max	2016.000000	12.000000	31.000000	56.300000	126.200000





Programación como  
herramienta

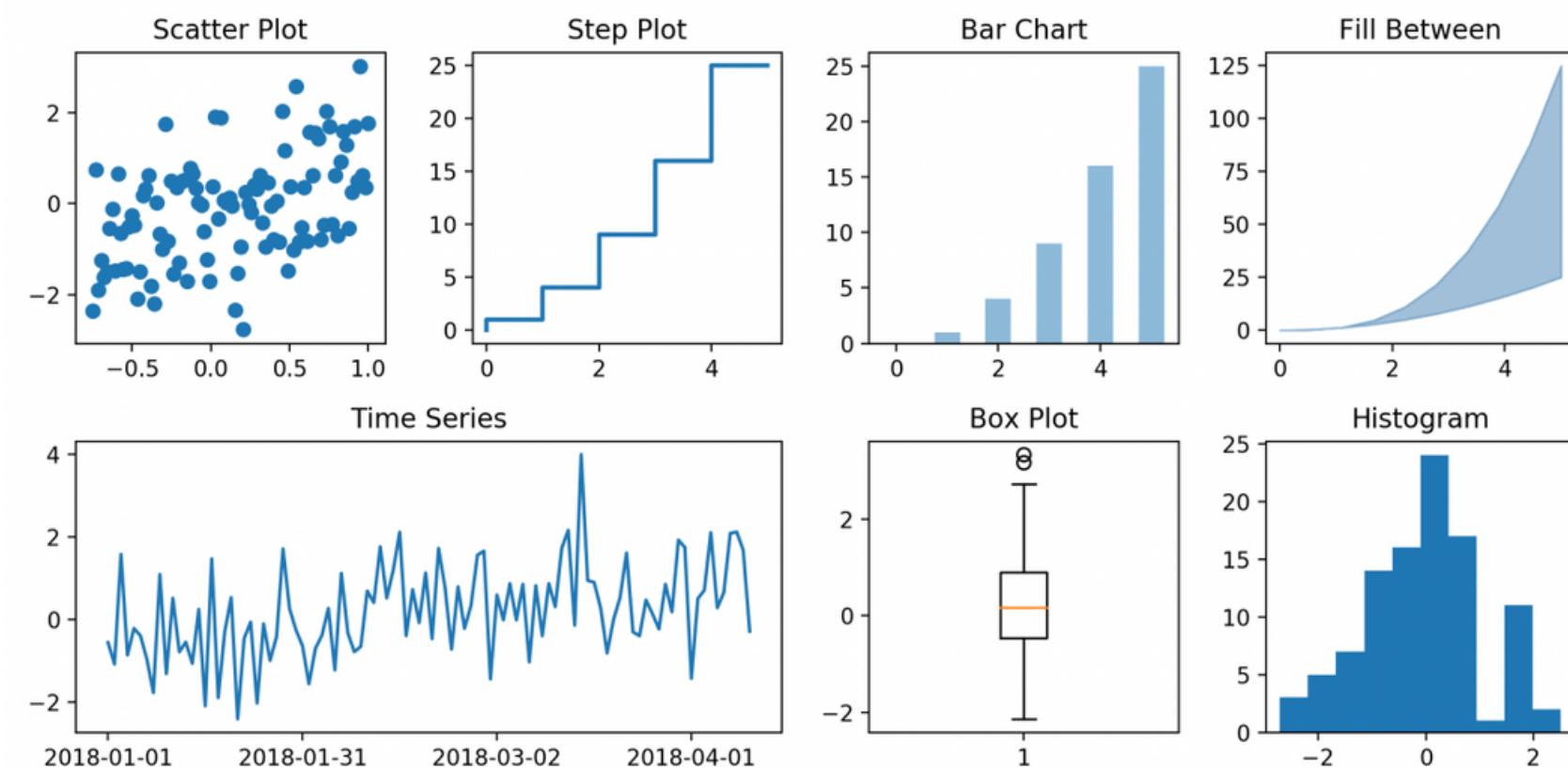
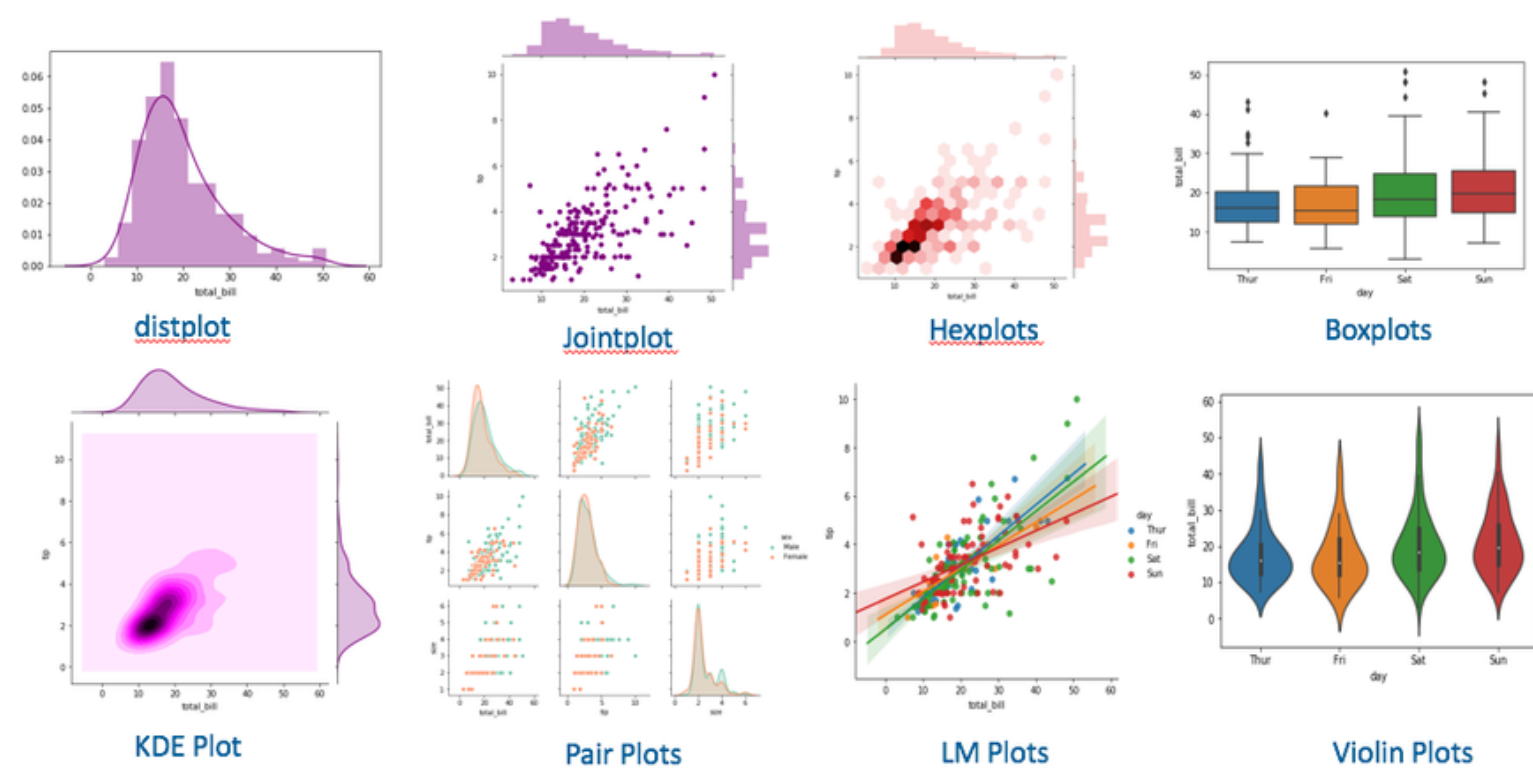
# ¿Qué puedo hacer con pandas?

Limpieza

Seaborn

Matplotlib

## Seaborn Plots





# ¿Qué puedo hacer con pandas?

## Funciones útiles

### **dropna()**

Función para eliminar valores vacíos de un dataframe

### **fillna()**

Rellenar valores vacíos de un dataframe en base a las condiciones que le especifiquemos

### **apply()**

Podemos escribir sobre el dataframe de manera eficiente y en base a la función que necesitemos

### **hist()**

Función para graficar un histograma del dataframe

### **boxplot()**

Función para graficar un boxplot del dataframe

### **plot()**

Función para graficar acorde a los datos del dataframe

### **read\_csv()**

### **read\_excel()**





Programación como  
herramienta

# MUCHAS GRACIAS!

