

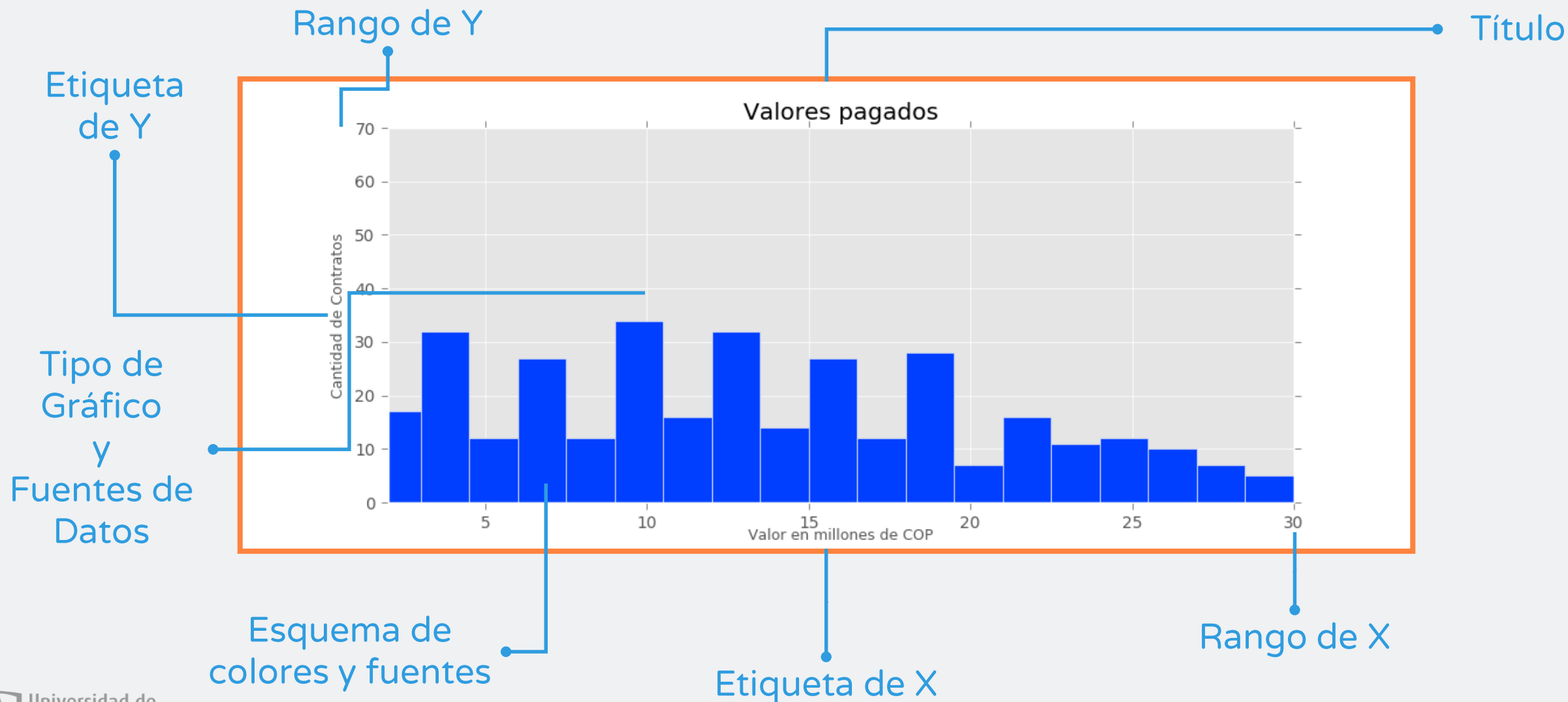
# NIVEL 4

---

## VISUALIZACIONES CON PANDAS

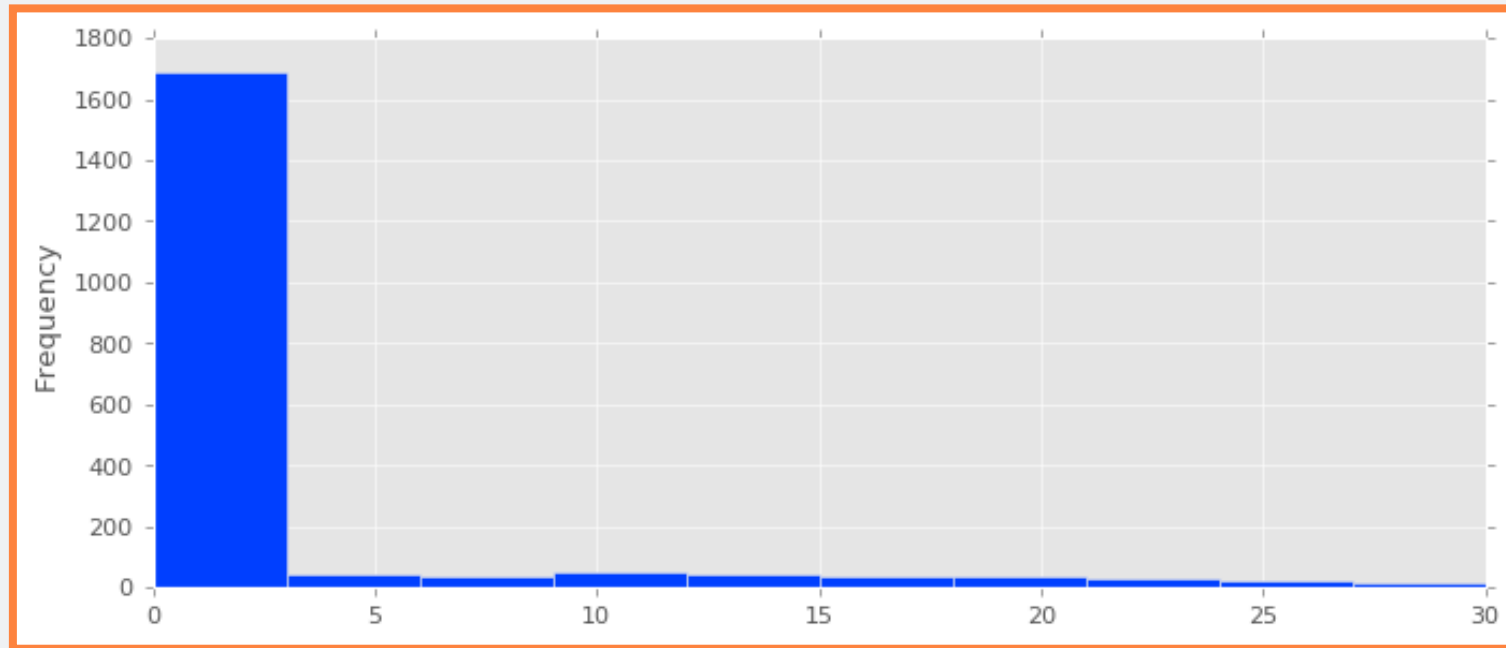


# CONCEPTOS BÁSICOS



# HISTOGRAMAHISTOGRAMA

¿Cuántas veces aparece cada valor en el conjunto de datos?



`pagado` es una Serie con los valores pagados en millones de COP

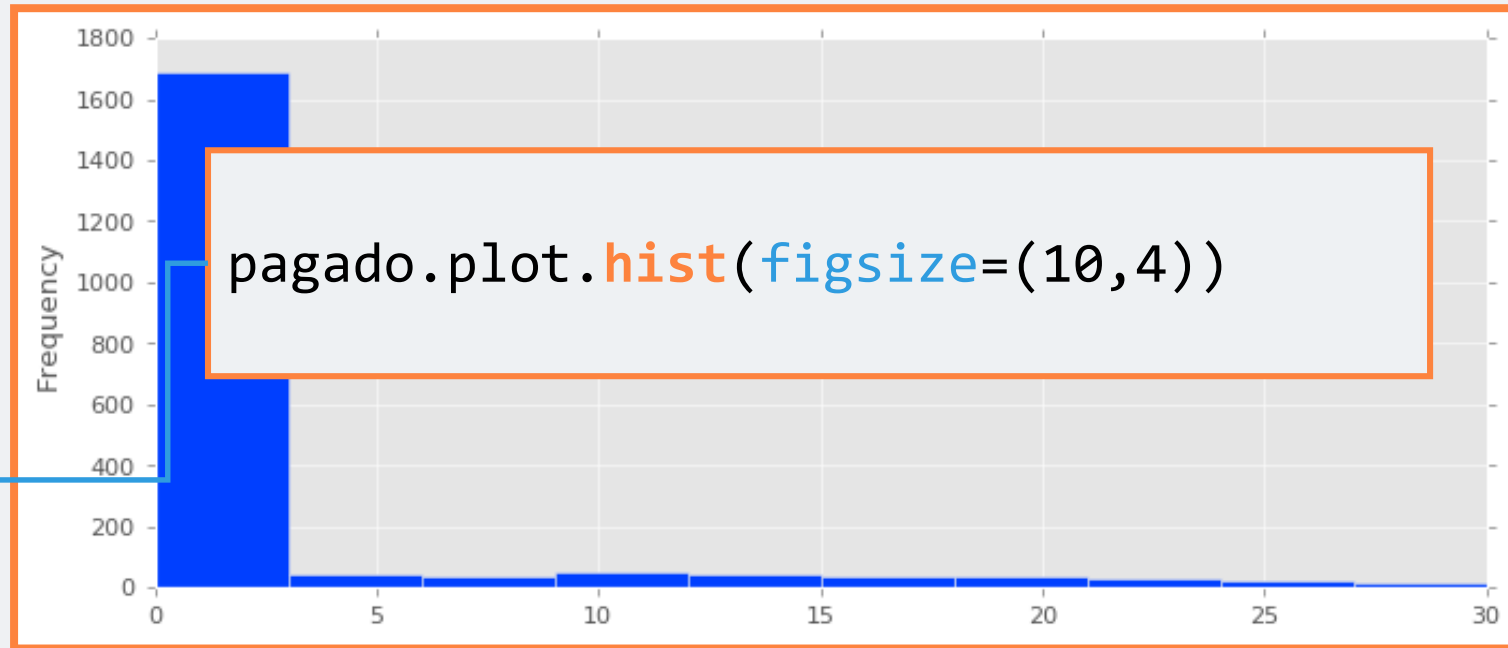
```
pagado=numericos['Valor Pagado']  
pagado.plot(kind="hist", figsize=(10,4))
```

`figsize` indica el ancho y alto de la figura

El parámetro `kind` indica el tipo de gráfica

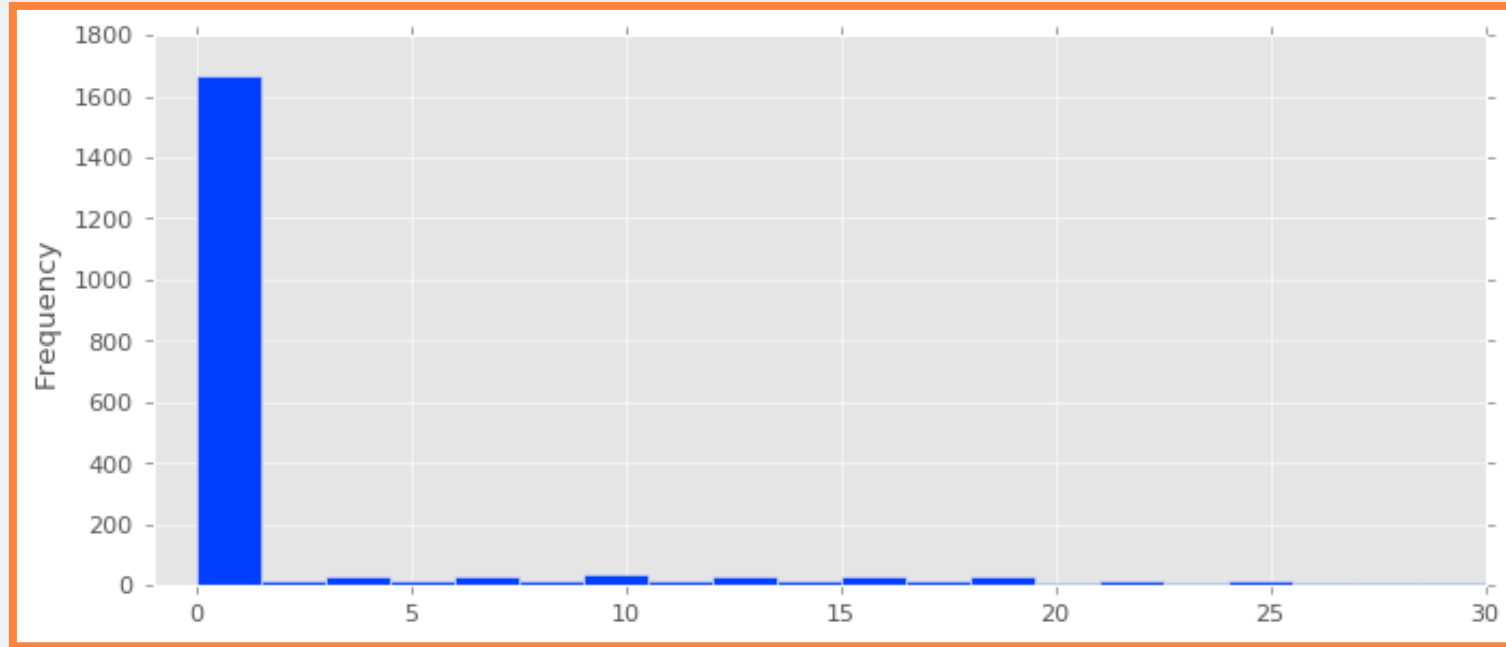
# HISTOGRAMA

Usar una función con tipo del diagrama en lugar de usar `plot` con el parámetro `kind`



```
pagado=numericos['Valor Pagado']  
pagado.plot(kind="hist", figsize=(10,4))
```

# HISTOGRAMA

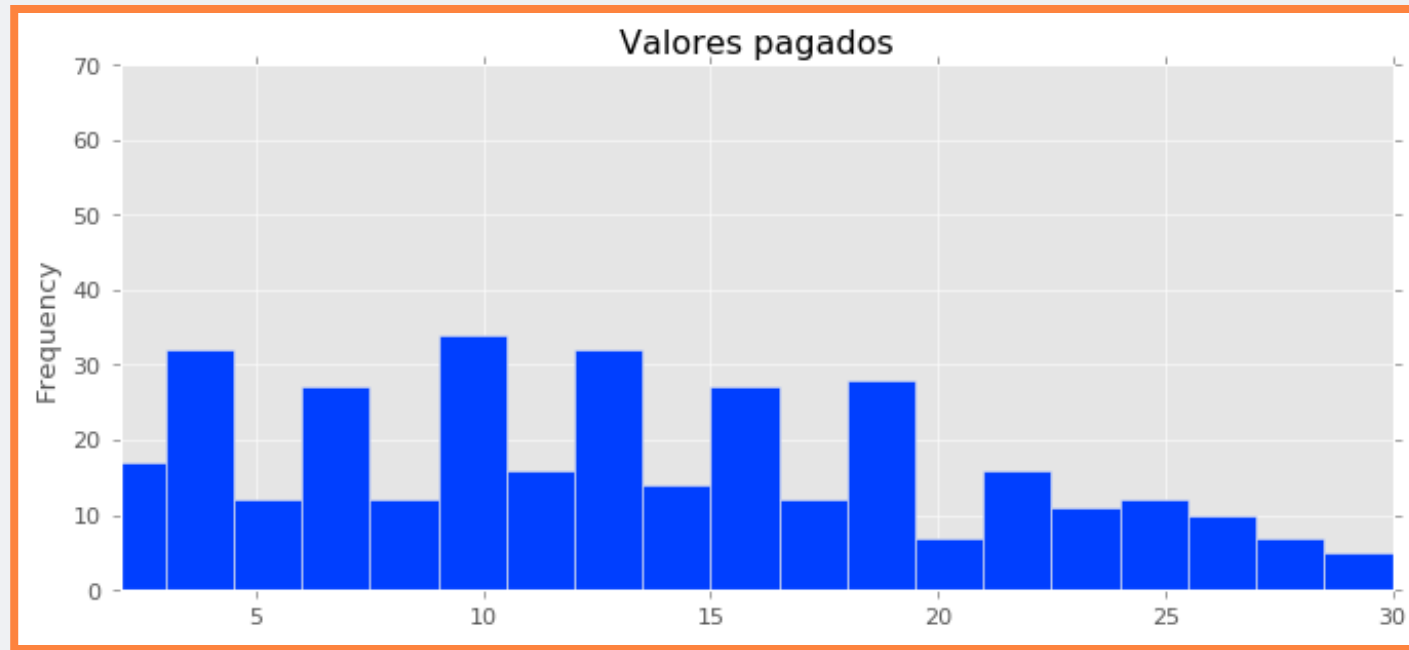


```
pagado.plot(kind="hist", figsize=(10,4),  
            xlim=(-1,30), bins=20)
```

El parámetro `xlim`  
indica el rango para x

El parámetro `bins` indica  
la cantidad de grupos

# HISTOGRAMA

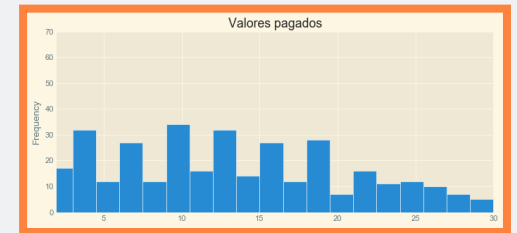
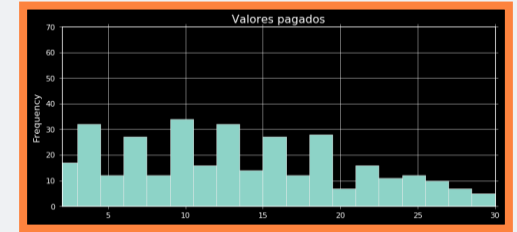
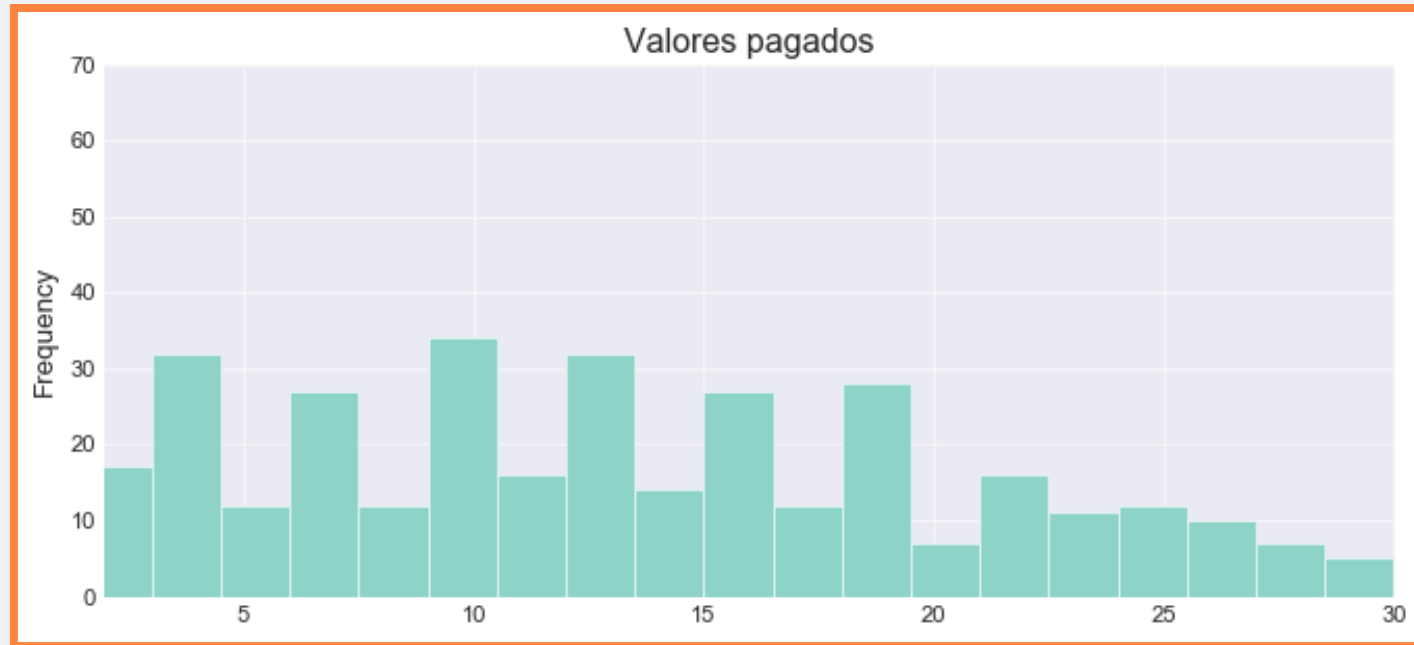


El parámetro `title` especifica el título del gráfico

```
pagado.plot(kind="hist", figsize=(10,4),  
            xlim=(2,30), ylim=(0, 70), bins=20,  
            title="Valores pagados")
```

El parámetro `ylim` indica el rango para Y

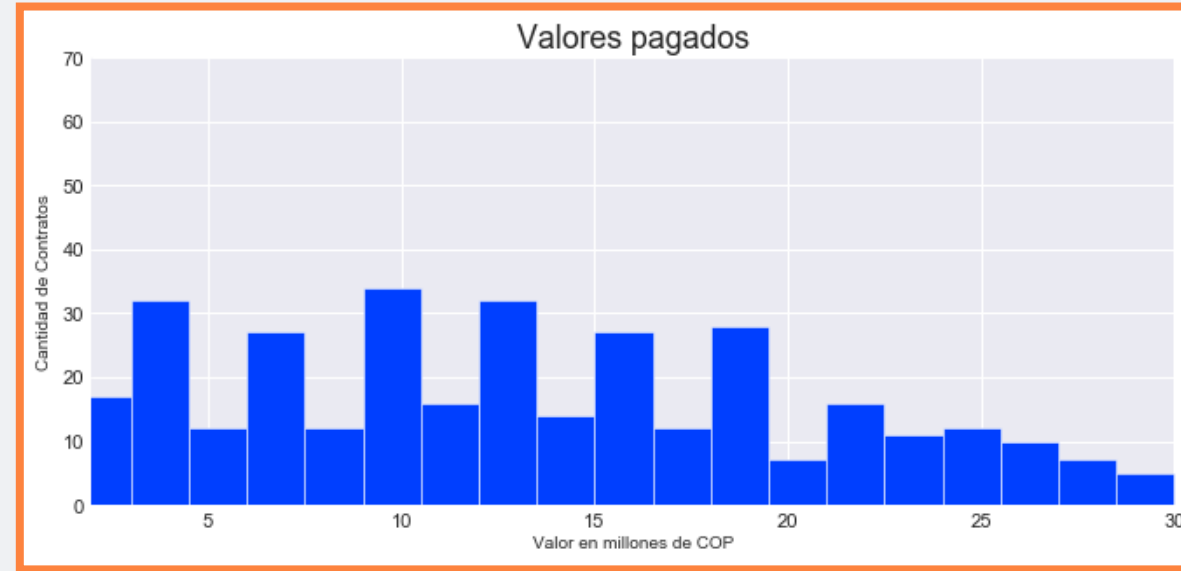
# HISTOGRAMA



Usar estilos  
predefinidos de  
Matplotlib.  
También pueden  
definirse estilos  
propios

```
matplotlib.style.use('seaborn-darkgrid')
pagado.plot(kind="hist", figsize=(10,4),
            xlim=(2,30), ylim=(0, 70), bins=20,
            title="Valores pagados")
```

# HISTOGRAMA



`ax` es un `AxesSubplot` y es el resultado de todos los llamados a `plot`

```
ax = pagado.plot(kind="hist", figsize=(10,4), xlim=(2,30), ylim=(0, 70),
                  bins=20, title="Valores pagados")
```

Configuración de etiquetas

```
ax.set_xlabel("Valor en millones de COP", fontsize = 9)
ax.xaxis.set_label_coords(0.5, -0.07)
```

Guardar la figura en un archivo

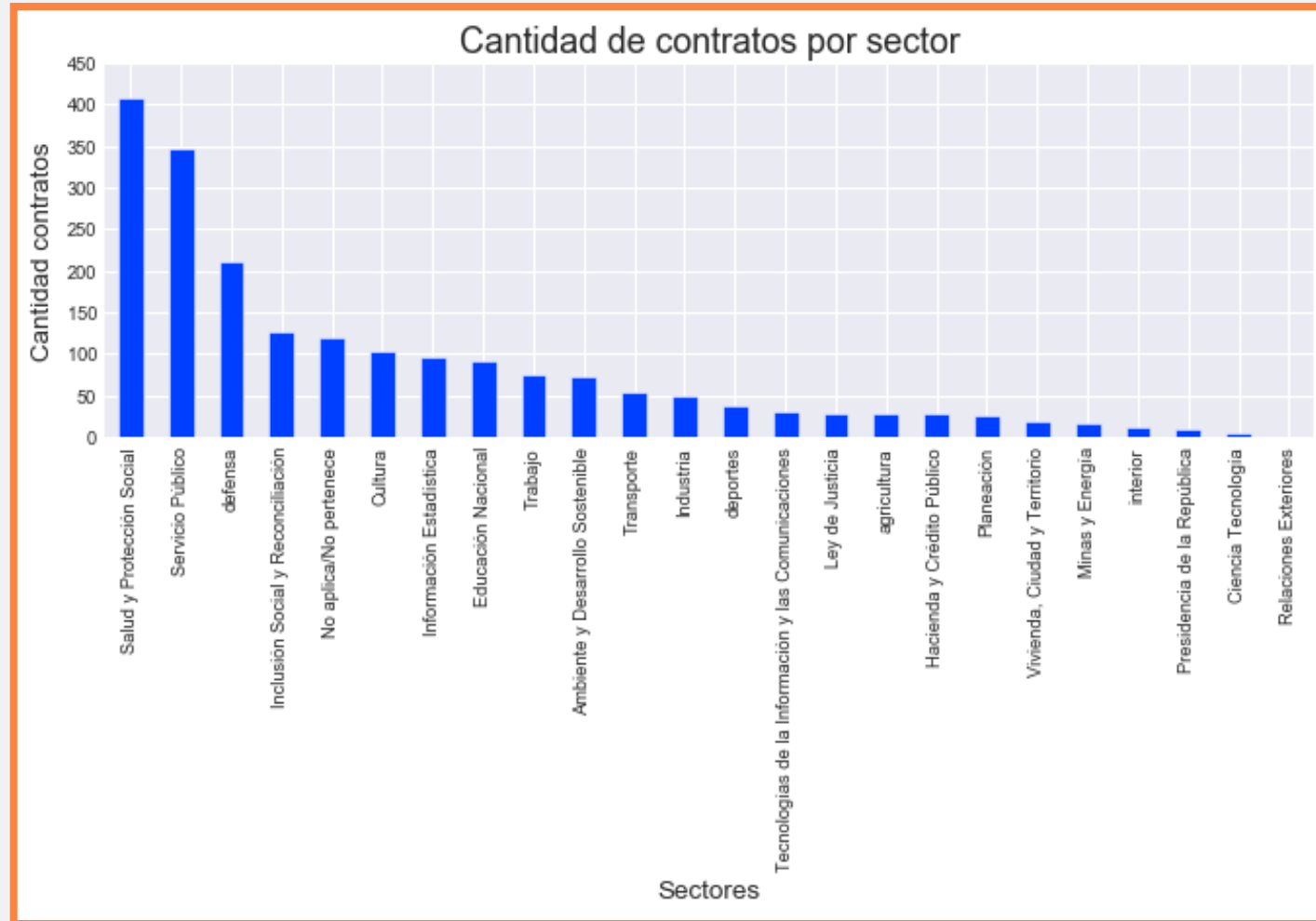
```
ax.set_ylabel("Cantidad de Contratos", fontsize = 9)
fig = ax.get_figure()
fig.savefig("histograma.svg")
```

El formato depende de la extensión que usemos



# DIAGRAMA DE BARRAS

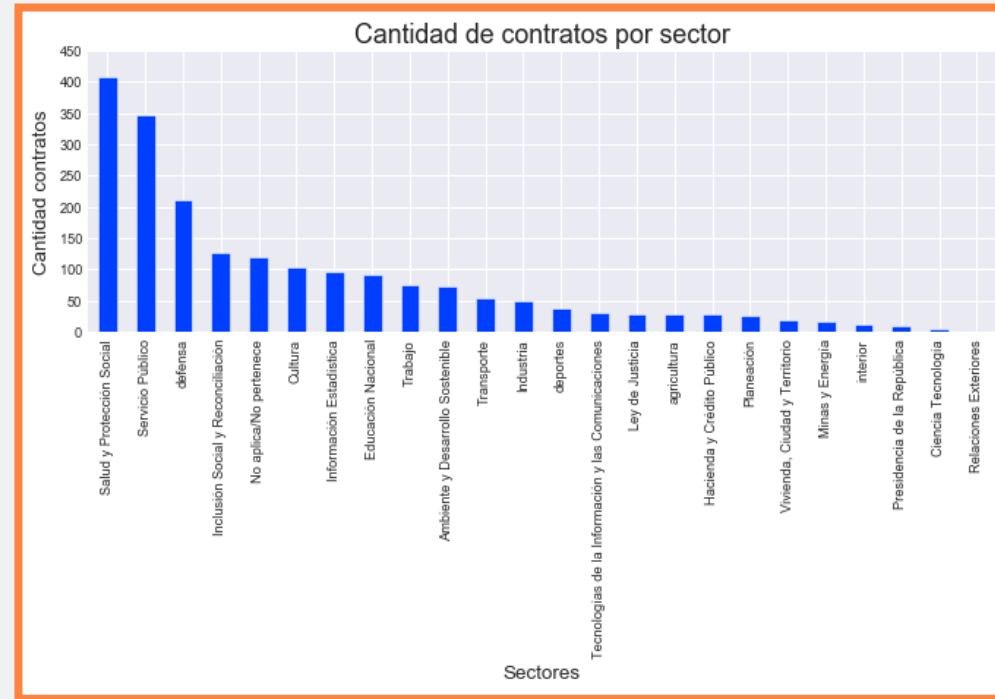
Valores numéricos asociados a valores categóricos



# DIAGRAMA DE BARRAS

`bar` indica que es un diagrama de barras verticales

`xlabel` y `ylabel` especifican las etiquetas de los ejes

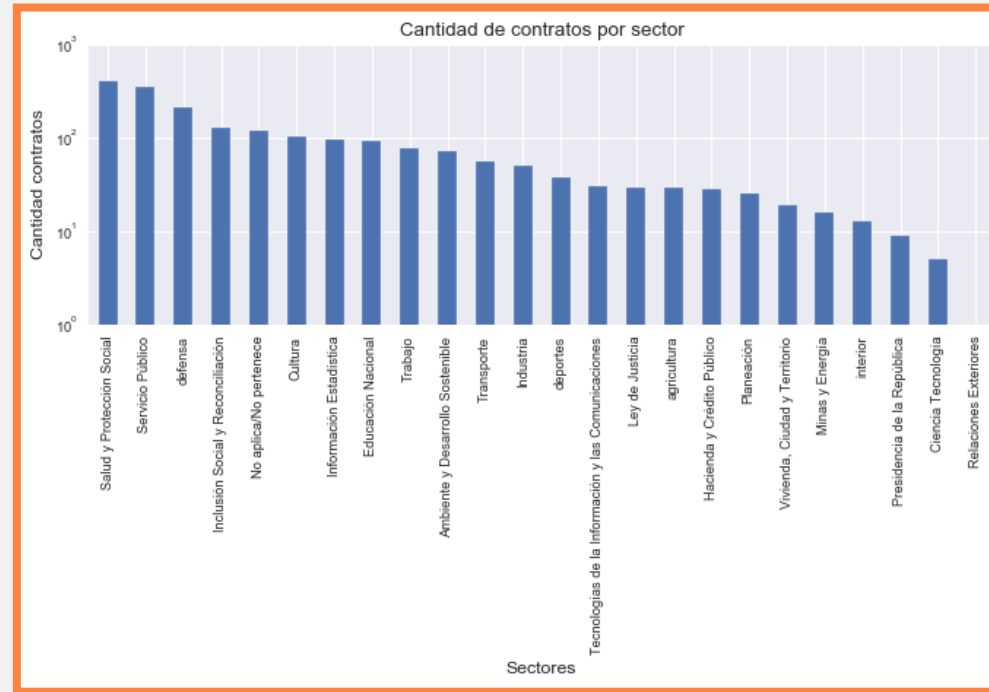


`fontsize` indica tamaño de los textos

```
cantidades = muestra['Sector'].value_counts()
cantidades.plot(kind="bar", figsize=(10, 3), fontsize="small",
                 xlabel="Sectores", ylabel="Cantidad contratos",
                 title="Cantidad de contratos por sector")
```

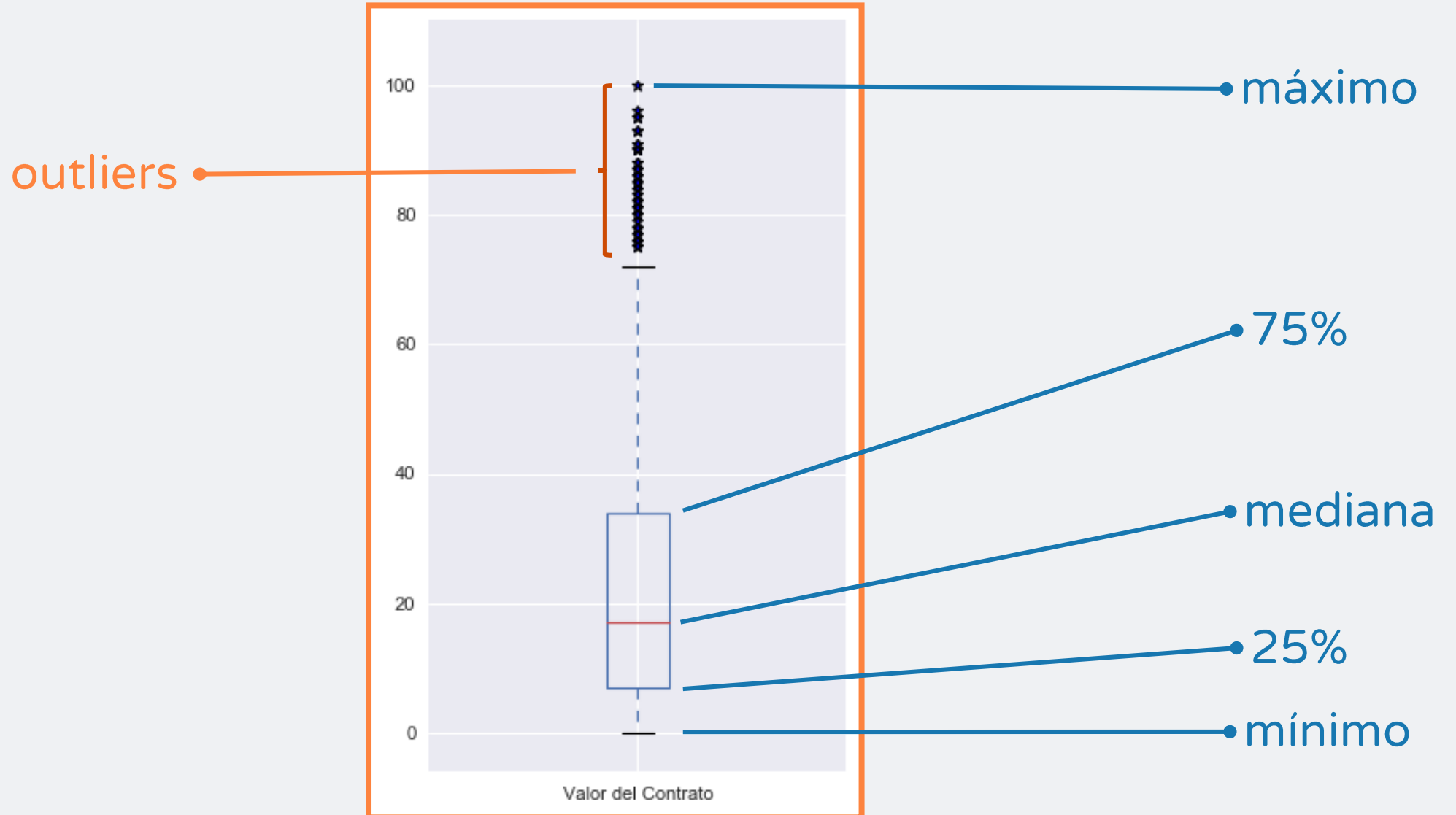
# DIAGRAMA DE BARRAS

Escala logarítmica  
en el eje Y



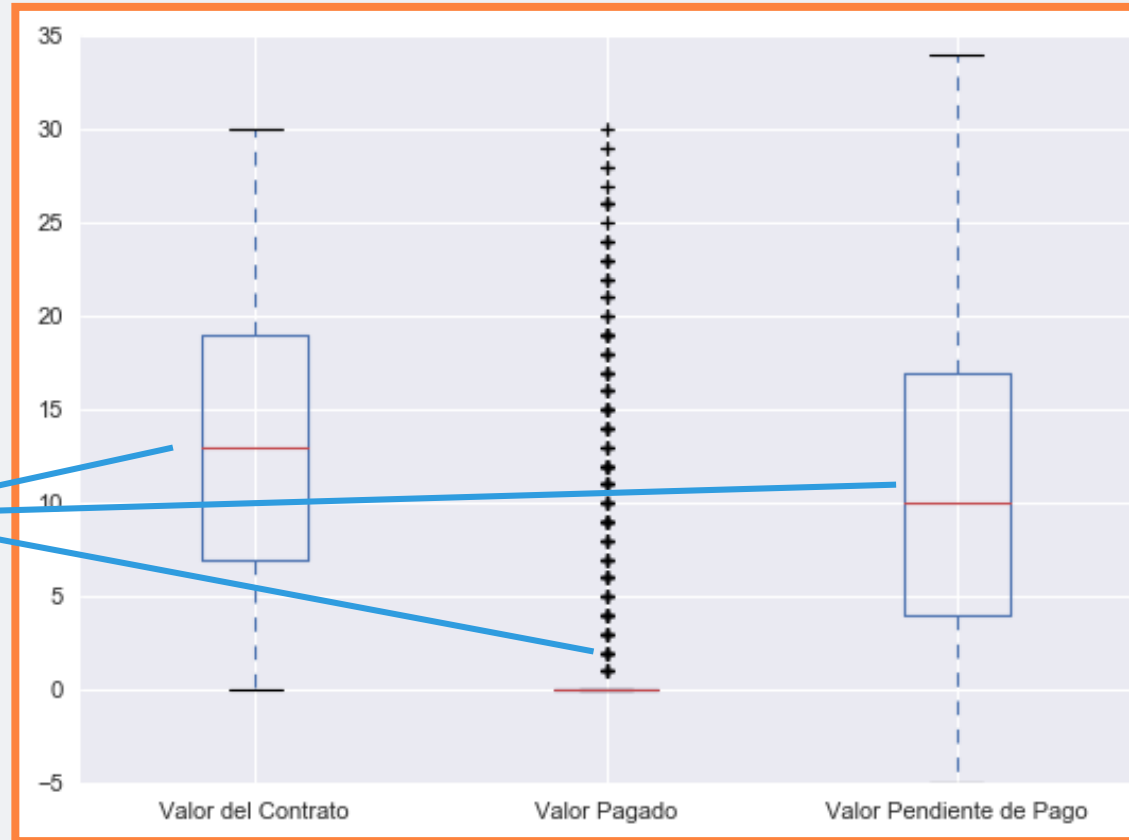
```
cantidades = muestra['Sector'].value_counts()
cantidades.plot(kind="bar", figsize=(10, 3), fontsize="small",
                xlabel="Sectores", ylabel="Cantidad contratos",
                logy=True, title="Cantidad de contratos por sector")
```

# BOX-PLOTS



# BOX-PLOTS

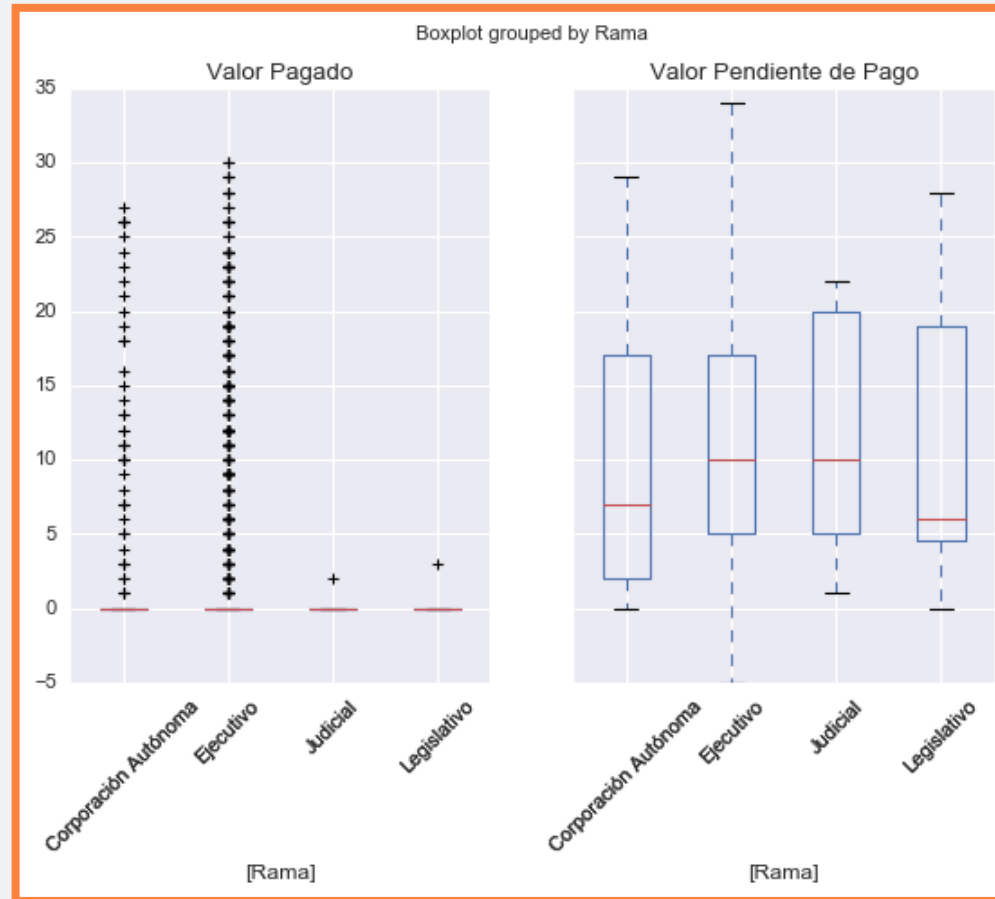
Un conjunto de valores por columna



Usamos box para indicar que es un box-plot

```
muestra[['Valor del Contrato', 'Valor Pagado',  
         'Valor Pendiente de Pago']].plot(kind="box")
```

# BOX-PLOTS AGRUPADOS



Pero... es mejor usar de la función **boxplot**. La función plot con **kind='box'** tiene errores en la implementación

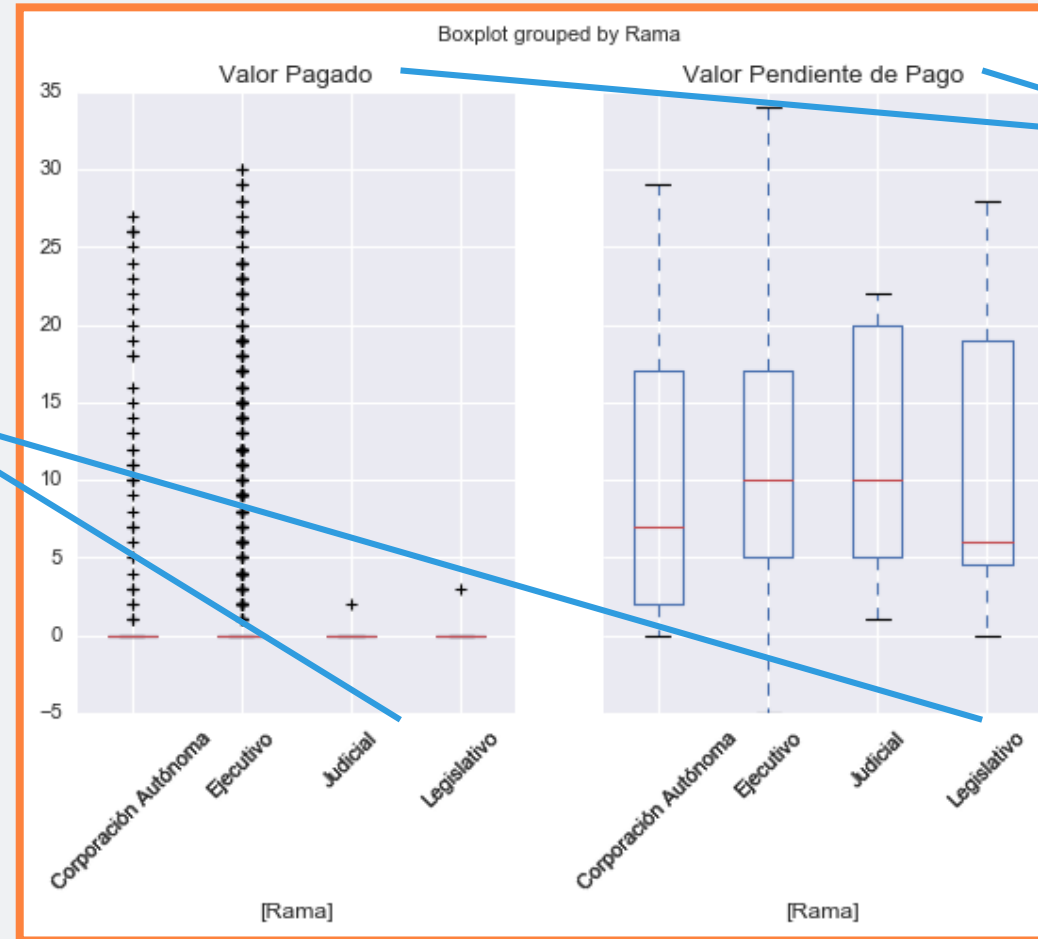
```
muestra[['Rama', 'Valor Pagado',  
         'Valor Pendiente de Pago']].boxplot(by="Rama", rot="45")
```

# BOX-PLOTS AGRUPADOS

Un conjunto de valores por categoría en 'Rama'

'Rama' tiene que hacer parte de las columnas para que se pueda usar en el parámetro `by`

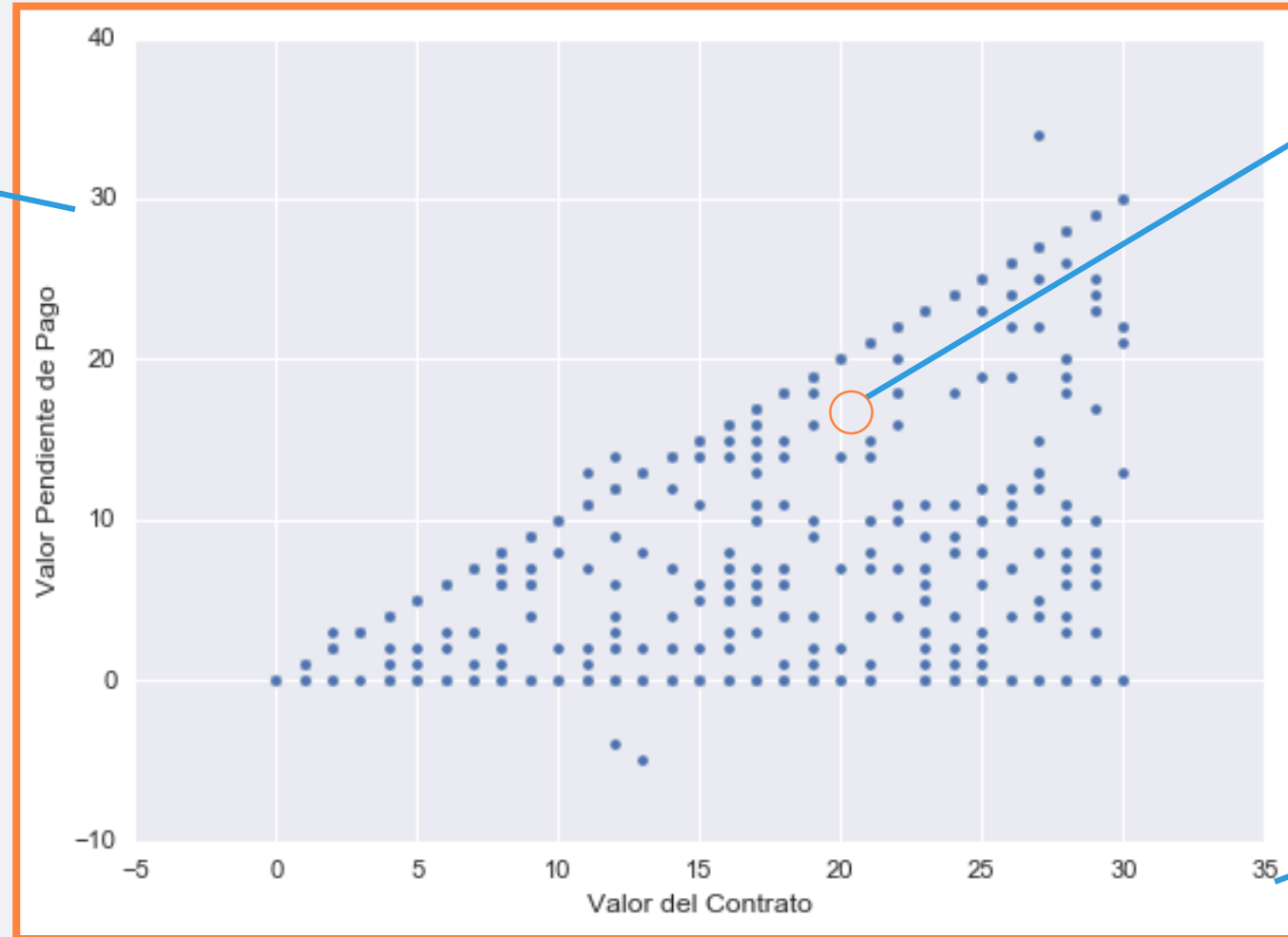
```
muestra[['Rama', 'Valor Pagado',  
        'Valor Pendiente de Pago']].boxplot(by="Rama", rot="45")
```



Un sub-plot por columna

# GRÁFICAS DE DISPERSIÓN

Valores del  
Eje Y



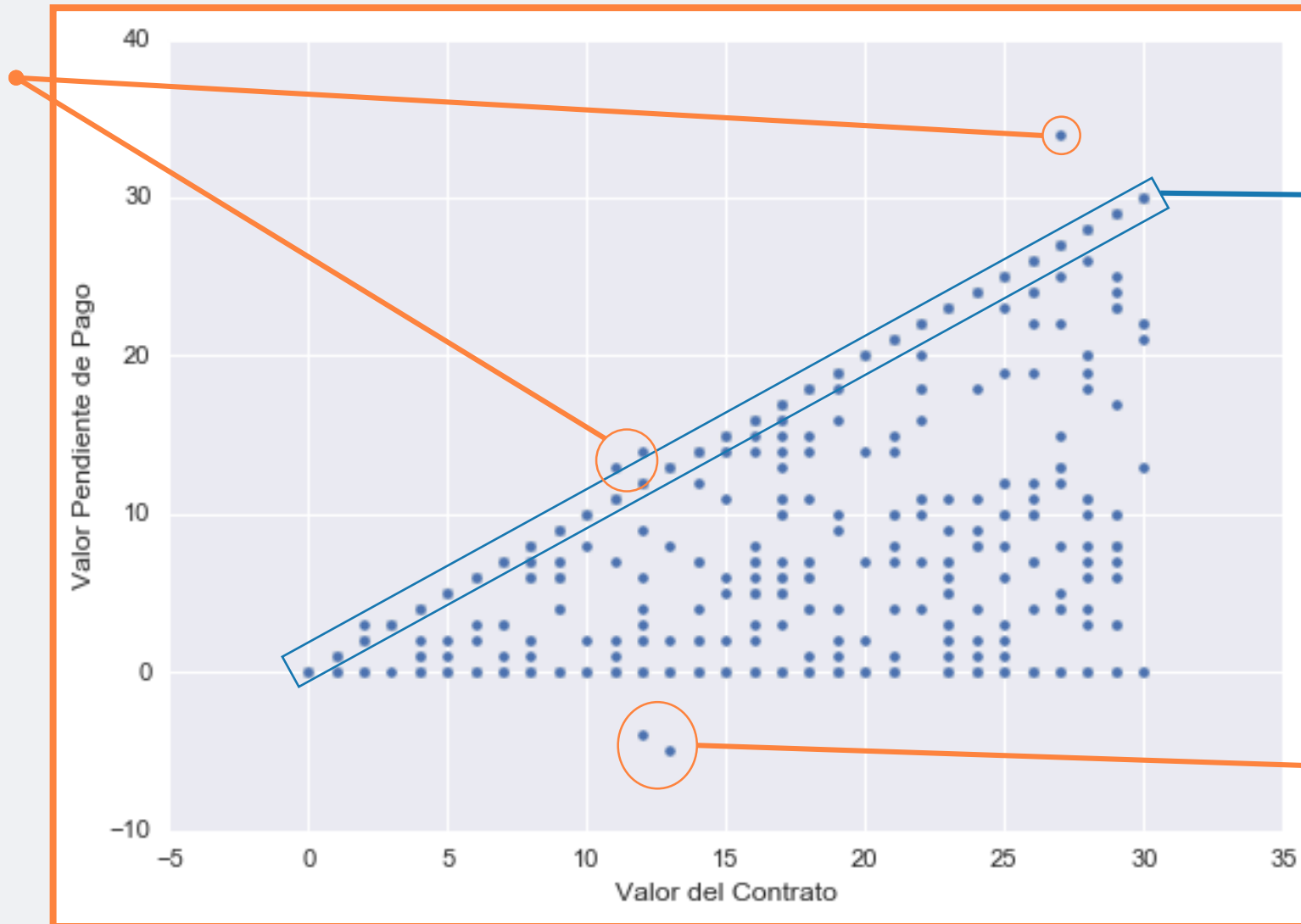
Valor en la  
posición ( X, Y)

Valores del  
Eje X



# GRÁFICAS DE DISPERSIÓN

Se les debe  
más de lo  
que vale el  
contrato

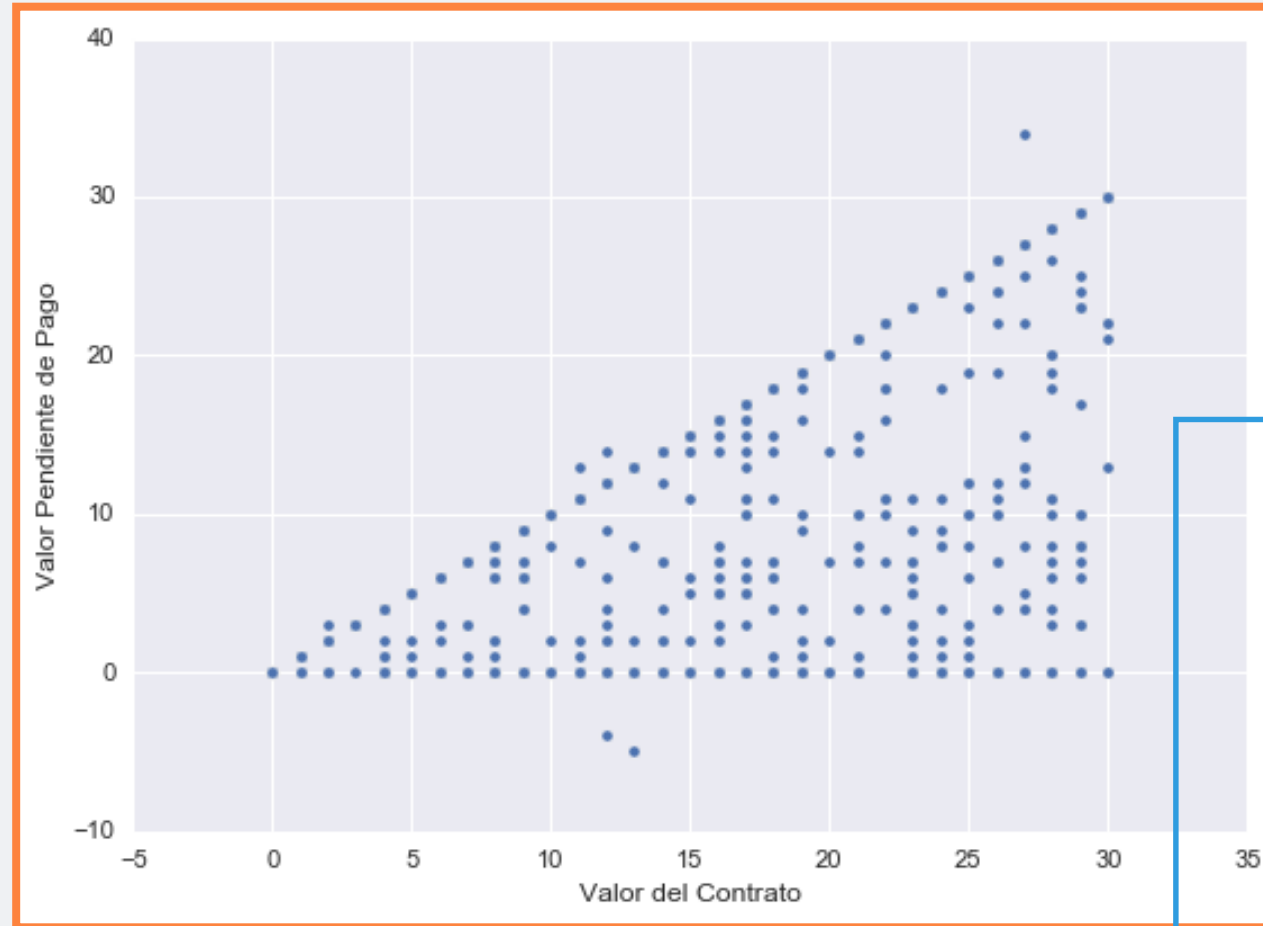


Se les debe el  
total del  
contrato

Se les pagó "de  
más" y ahora  
"deben"

# GRÁFICAS DE DISPERSIÓN

Usamos **scatter** para indicar que es un diagrama de dispersión

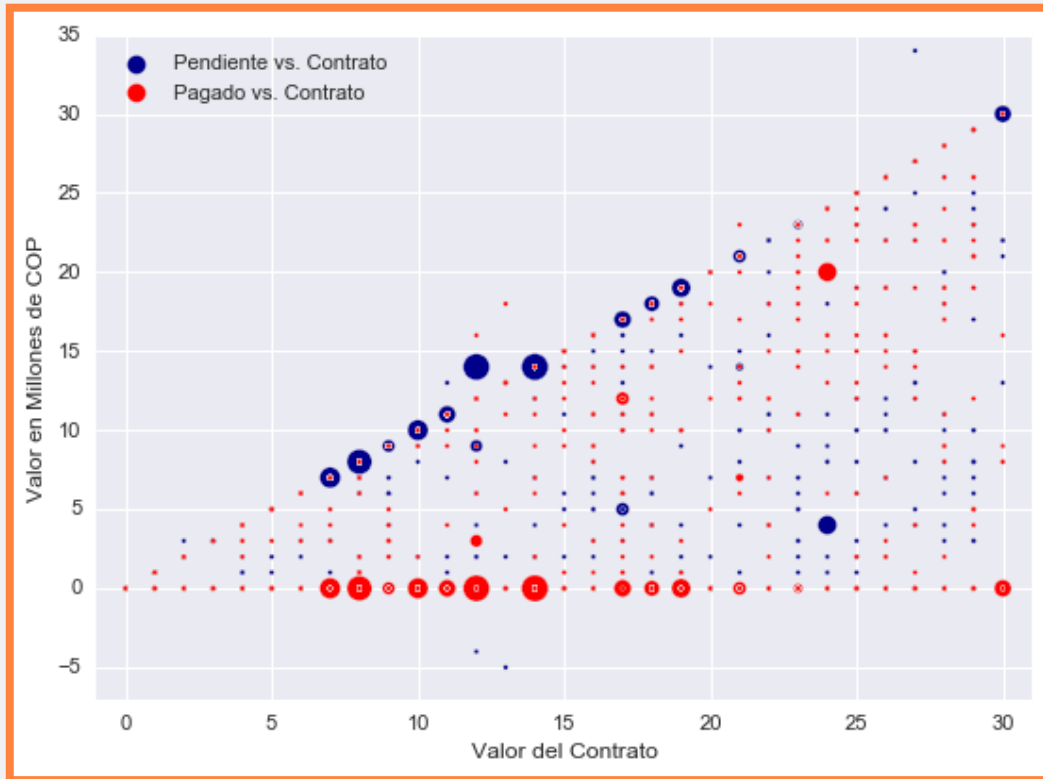


Columna que usaremos para el eje X

Columna que usaremos para el eje Y

```
muestra.plot(kind="scatter", x='Valor del Contrato',  
y='Valor Pendiente de Pago')
```

# GRÁFICA MÚLTIPLE



```
sub_plot = muestra.plot(kind="scatter",
                        x='Valor del Contrato',
                        y='Valor Pendiente de Pago',
                        color="DarkBlue",
                        label="Pendiente vs. Contrato",
                        xlim=(-1,31), ylim=(-7,35),
                        s=5+prestacion['Dias Adicionados'])

sub_plot2 = muestra.plot(kind="scatter",
                        x='Valor del Contrato', y='Valor Pagado',
                        color="Red",
                        label="Pagado vs. Contrato",
                        s=5+prestacion['Dias Adicionados'],
                        ax=sub_plot)

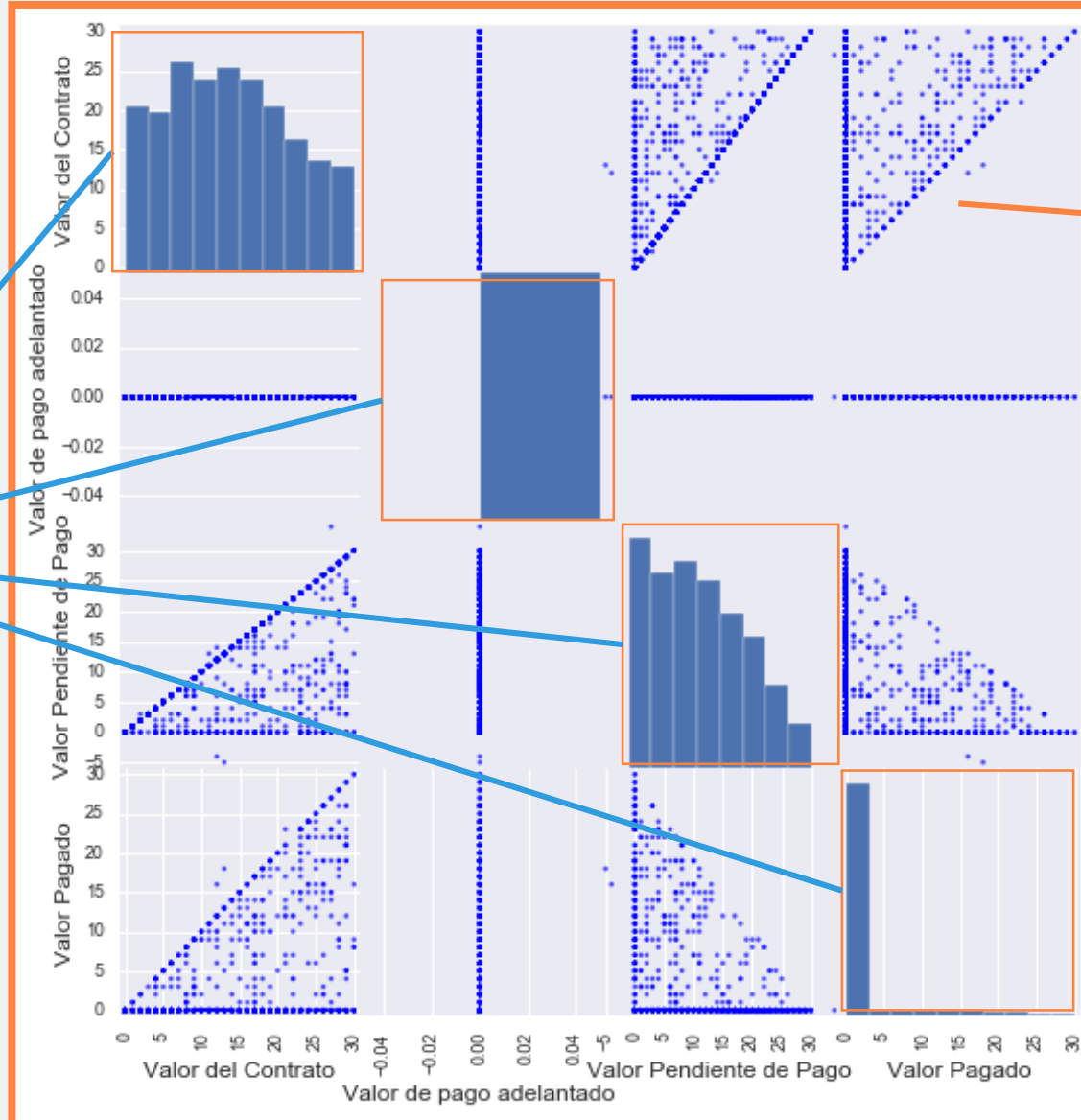
sub_plot2.set_ylabel("Valor en Millones de COP")
```

La segunda gráfica debe construirse sobre la primera

**s** = Representa el tamaño de cada punto

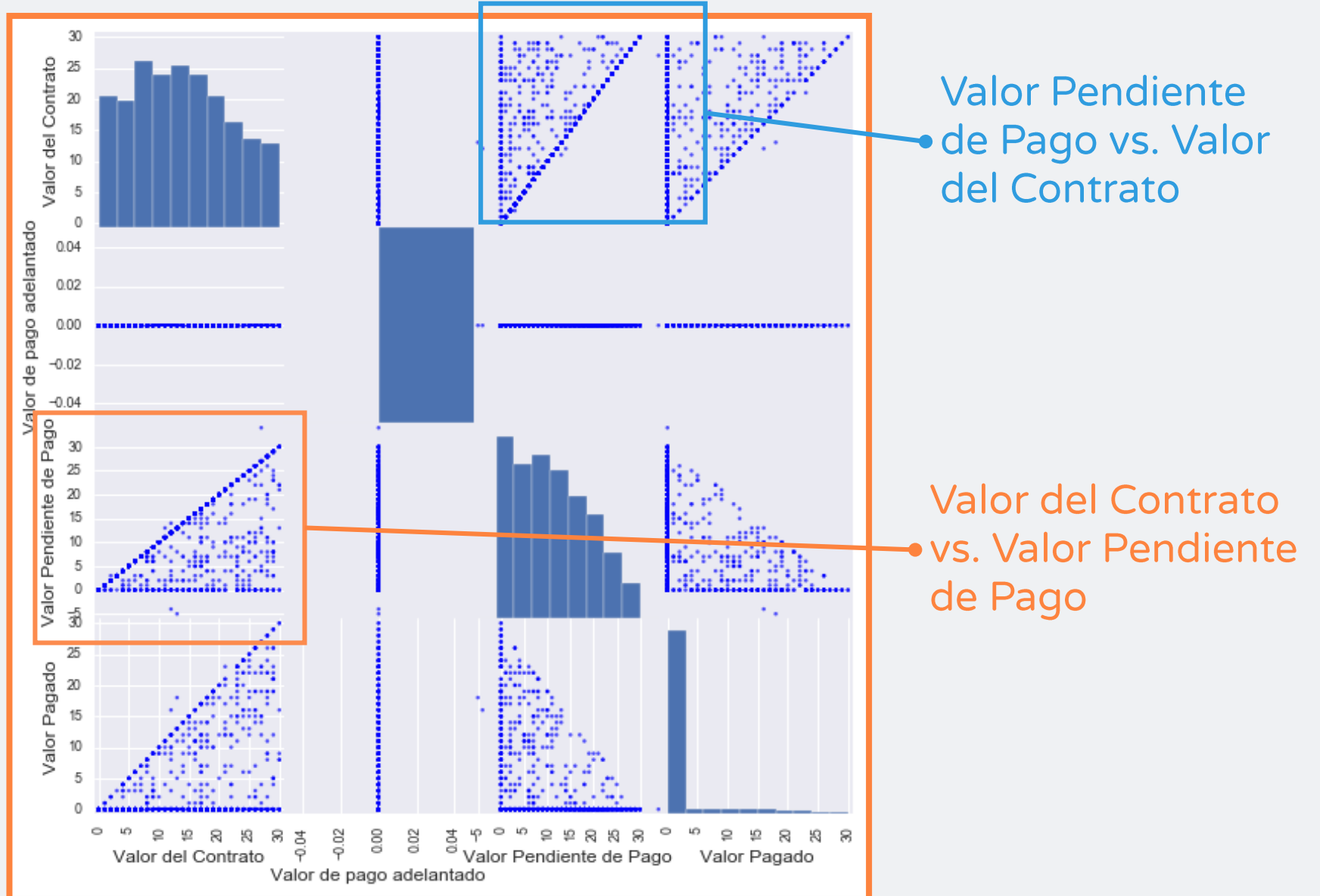
# MATRICES DE DISPERSIÓN

Histogramas  
o  
KDE

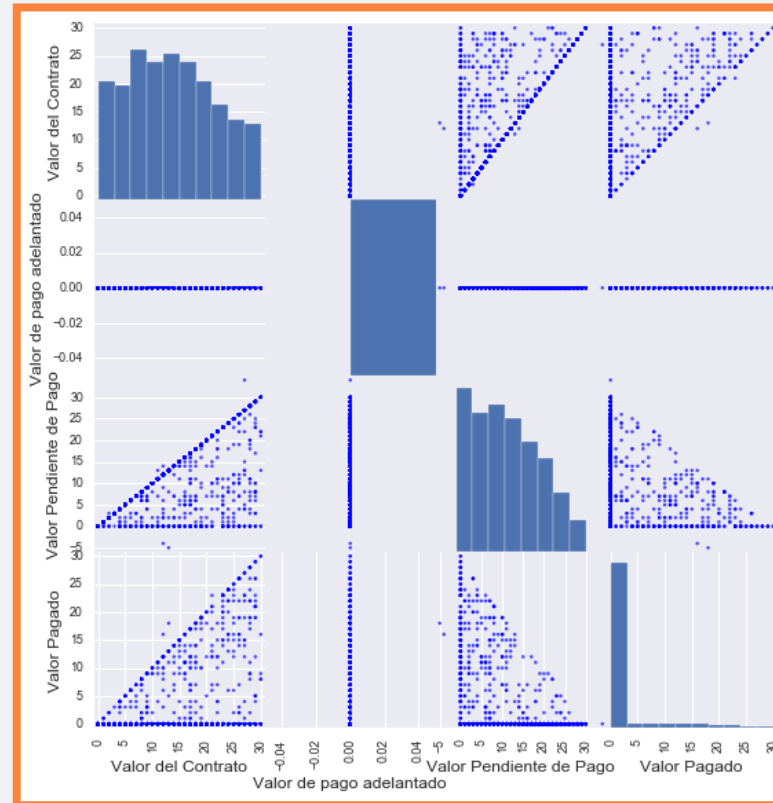


Dispersión:  
Variable vs.  
Variable

# MATRICES DE DISPERSIÓN

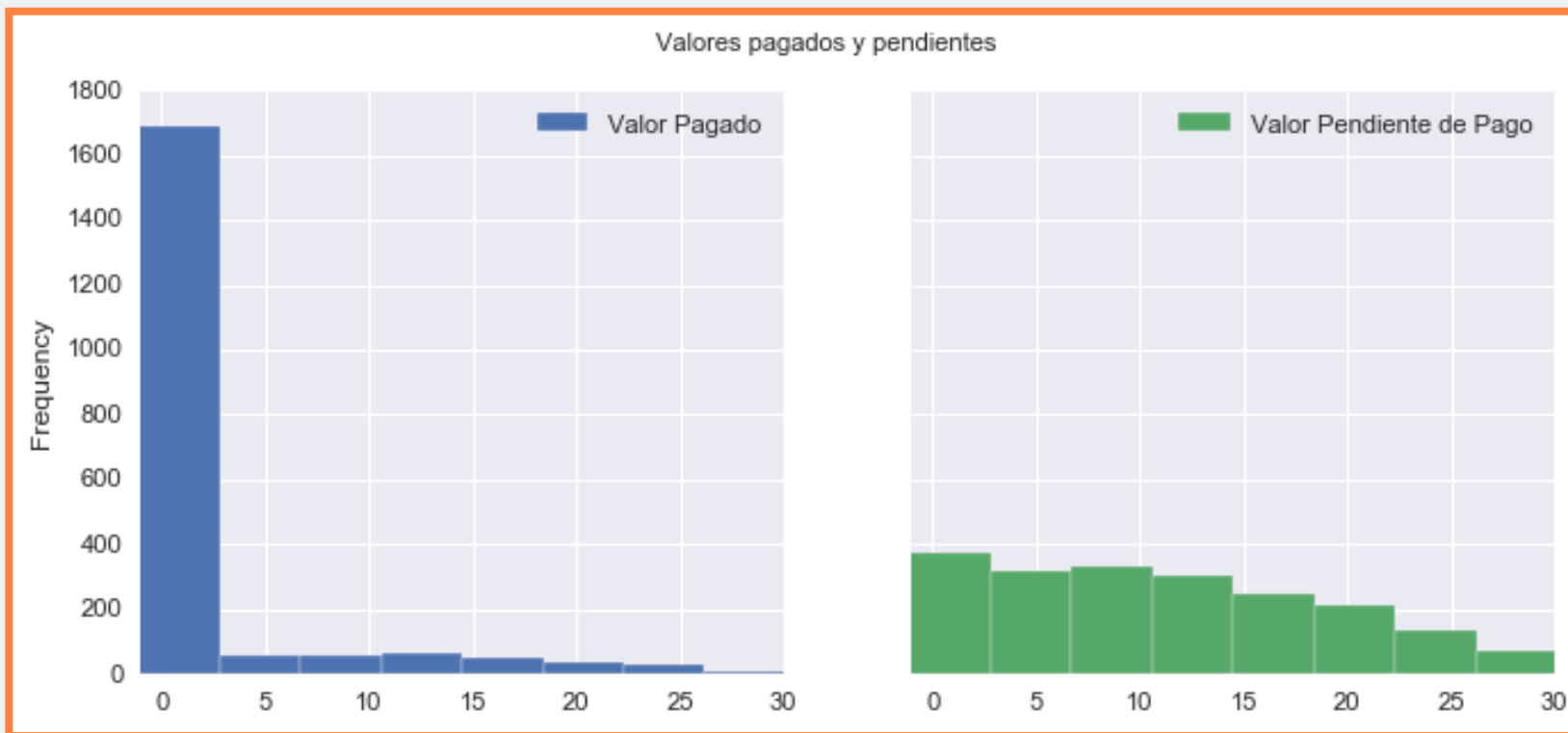


# MATRICES DE DISPERSIÓN



```
from pandas.plotting import scatter_matrix  
scatter_matrix(numericos, alpha=0.7, figsize=(8, 8), diagonal='hist')
```

# MATRICES DE DISPERSIÓN



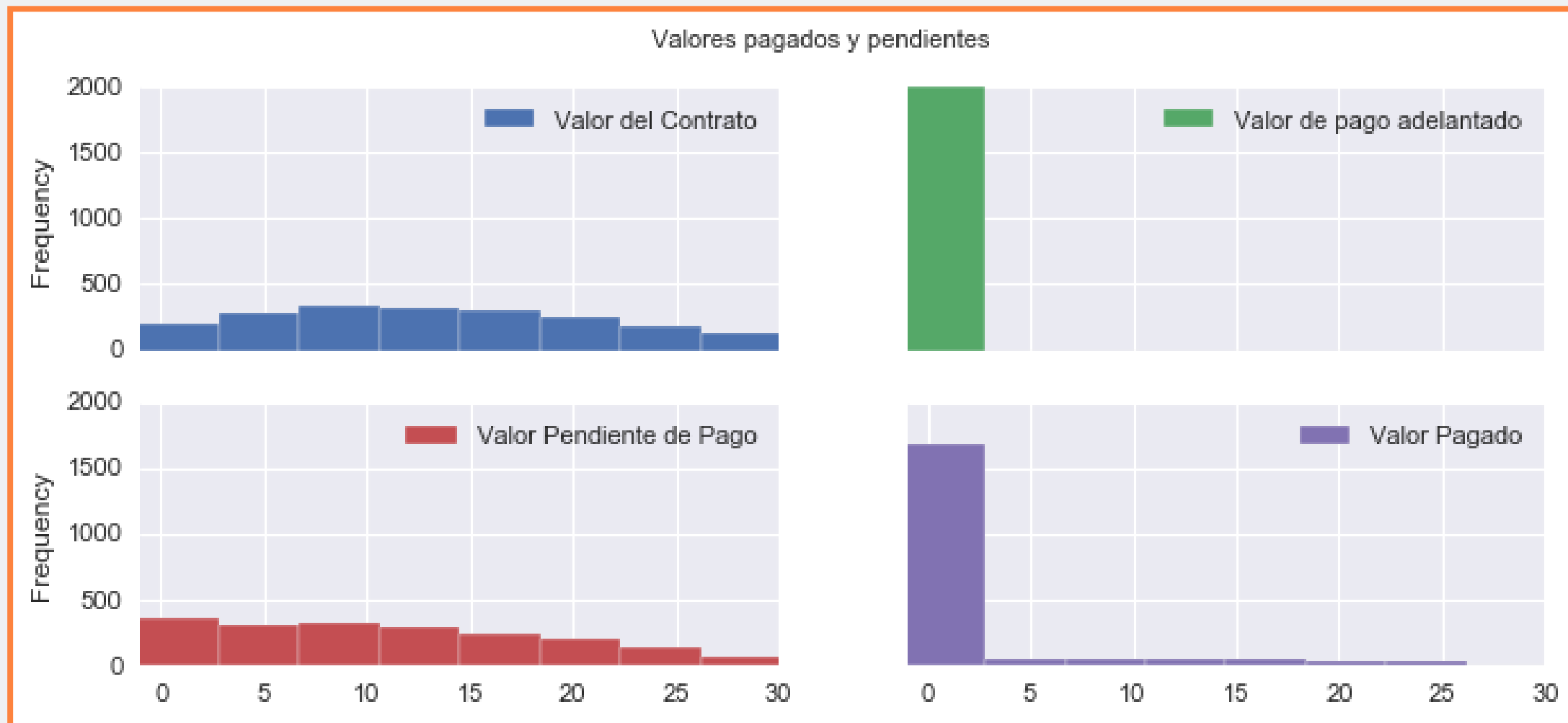
Si **subplots** es True, **layout** tiene una tupla con la cantidad de filas y columnas

Si **subplots** es True, la figura puede tener varias gráficas

```
columnas = numericos[['Valor Pagado', 'Valor Pendiente de Pago']]
columnas.plot(kind="hist", figsize=(10,4), xlim=(-1,30), bins=10,
               subplots=True, layout=(1,2), sharey=True,
               title="Valores pagados y pendientes")
```

Indica que se debe compartir el eje Y

# MATRIZ DE HISTOGRAMAS





# MÁS INFORMACIÓN

[https://pandas.pydata.org/docs/user\\_guide/visualization.html](https://pandas.pydata.org/docs/user_guide/visualization.html)

[https://pandas.pydata.org/docs/user\\_guide/visualization.html#box-plots](https://pandas.pydata.org/docs/user_guide/visualization.html#box-plots)

[https://matplotlib.org/api/pyplot\\_api.html#matplotlib.pyplot.boxplot](https://matplotlib.org/api/pyplot_api.html#matplotlib.pyplot.boxplot)

[https://pandas.pydata.org/docs/user\\_guide/visualization.html#scatter-plot](https://pandas.pydata.org/docs/user_guide/visualization.html#scatter-plot)

[https://pandas.pydata.org/docs/user\\_guide/visualization.html#scatter-matrix-plot](https://pandas.pydata.org/docs/user_guide/visualization.html#scatter-matrix-plot)

[https://pandas.pydata.org/docs/user\\_guide/visualization.html#subplots](https://pandas.pydata.org/docs/user_guide/visualization.html#subplots)