In [1]:

```
%load_ext watermark
%watermark
```

```
2019-05-17T17:14:34+02:00

CPython 3.6.5
IPython 6.4.0

compiler   : GCC 7.2.0
system     : Linux
release    : 5.0.13-arch1-1-ARCH
machine    : x86_64
processor  :
CPU cores  : 4
interpreter: 64bit
```

VISUALIZACIÓN BÁSICA DE DATOS

Carga de Datos y Preparacion de DataSet

Vamos a usar un dataset clasico para empezar a aprender técnicas de visualización. Se trata del Boston Housing Dataset. Recopilado en 1976 y publicado en Berkeley

Consiste en mediciones de distintas zonas del área de Boston, teniendo como variables independientes un conjunto de mediciones de la habitabilidad de dichas zonas, y como variable independiente el valor medio de las casas en dicha zona.

En concreto vamos a usar un dataset actualizado que incluye la geolocalización estimada de las mediciones.

In [2]:

```
import pandas as pd
df = pd.read_csv("boston_dataset.csv")
df.head()
```

Out[2]:

	TOWN	LON	LAT	MEDV	CRIM	ZN	INDUS	CHAS	NOX	RM	AGE	DIS	RAD	TAX	PTRATIO	B
0	Nahant	-70.955	42.2550	24.0	0.00632	18.0	2.31	0	0.538	6.575	65.2	4.0900	1	296	15.3	396.90
1	Swampscott	-70.950	42.2875	21.6	0.02731	0.0	7.07	0	0.469	6.421	78.9	4.9671	2	242	17.8	396.90
2	Swampscott	-70.936	42.2830	34.7	0.02729	0.0	7.07	0	0.469	7.185	61.1	4.9671	2	242	17.8	392.83
3	Marblehead	-70.928	42.2930	33.4	0.03237	0.0	2.18	0	0.458	6.998	45.8	6.0622	3	222	18.7	394.63
4	Marblehead	-70.922	42.2980	36.2	0.06905	0.0	2.18	0	0.458	7.147	54.2	6.0622	3	222	18.7	396.90

Renombramos las columnas para facilitar el manejo.

In [3]:

```
df = df.rename(columns={
    "TOWN":"CIUDAD",
    "CRIM":"INDICE_CRIMEN",
    "ZN":"PCT_ZONA_RESIDENCIAL",
    "INDUS":"PCT_ZONA_INDUSTRIAL",
    "CHAS":"RIO_CHARLES",
    "NOX":"OXIDO_NITROSO_PPM",
    "RM":"N_HABITACIONES_MEDIO",
    "AGE":"PCT_CASAS_40S",
    "DIS_EMPLEO":"DISTANCIA_CENTRO_EMPLEO",
```

```
    "RAD":"DIS_AUTOPISTAS",
    "TAX":"CARGA_FISCAL",
    "PTRATIO":"RATIO_PROFESORES",
    "B":"PCT_NEGRA",
    "MEDV":"VALOR_MEDIANO",
    "LSTAT":"PCT_CLASE_BAJA"
})

df.head()
```

Out[3]:

	CIUDAD	LON	LAT	VALOR_MEDIANO	INDICE_CRIMEN	PCT_ZONA_RESIDENCIAL	PCT_ZONA_INDUSTRIAL
0	Nahant	-70.955	42.2550	24.0	0.00632	18.0	2.31
1	Swampscott	-70.950	42.2875	21.6	0.02731	0.0	7.07
2	Swampscott	-70.936	42.2830	34.7	0.02729	0.0	7.07
3	Marblehead	-70.928	42.2930	33.4	0.03237	0.0	2.18
4	Marblehead	-70.922	42.2980	36.2	0.06905	0.0	2.18

In [4]:

```
df.dtypes
```

Out[4]:

```
CIUDAD                   object
LON                     float64
LAT                     float64
VALOR_MEDIANO           float64
INDICE_CRIMEN           float64
PCT_ZONA_RESIDENCIAL    float64
PCT_ZONA_INDUSTRIAL     float64
RIO_CHARLES               int64
OXIDO_NITROSO_PPM       float64
N_HABITACIONES_MEDIO    float64
PCT_CASAS_40S           float64
DIS                     float64
DIS_AUTOPISTAS            int64
CARGA_FISCAL              int64
RATIO_PROFESORES        float64
PCT_NEGRA               float64
PCT_CLASE_BAJA          float64
dtype: object
```

Cómo elegir el gráfico

Como podemos ver, la guía se divide en cuatro categorías principales y luego se clasifican los distintos métodos de visualización que mejor representan cada una de esas categorías. Veamos un poco más en detalle cada una de ellas:

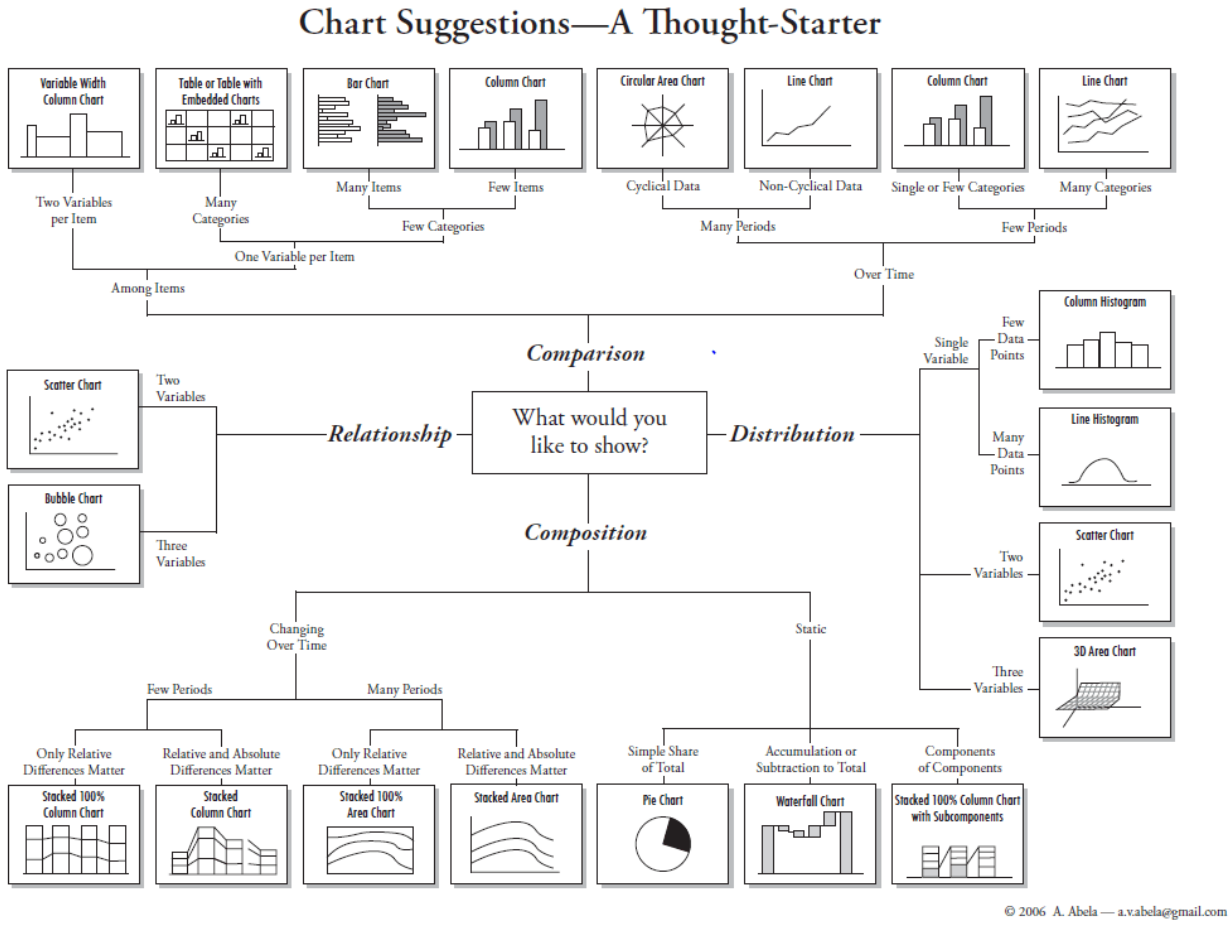
- *Distribuciones*: En esta categoría intentamos comprender como los datos se distribuyen. Se suelen utilizar en el comienzo de la etapa de exploración de datos, cuando queremos comprender las variables. Aquí también nos vamos a encontrar con variables de dos tipos cuantitativas y categóricas. Dependiendo del tipo y cantidad de variables, el método de visualización que vamos a utilizar.
- *Comparaciones*: En esta categoría el objetivo es comparar valores a través de diferentes categorías y con el tiempo (tendencia). Los tipos de gráficos más comunes en esta categoría son los diagramas de barras para cuando estamos comparando elementos o categorías y los diagramas de puntos y líneas cuando comparamos variables cuantitativas.
- *Relaciones*: Aquí el objetivo es comprender la relación entre dos o más variables. La visualización más utilizada en esta categoría es el gráfico de dispersión.
- *Composiciones*: En esta categoría el objetivo es comprender como esta compuesta o distribuida una variable; ya sea a través del tiempo o en forma estática. Las visualizaciones más comunes aquí son los diagramas de barras y los gráficos de tortas.

Fuente: https://relopezbriega.github.io/blog/2016/09/18/visualizaciones-de-datos-con-python/

In [17]:

```
from IPython.display import Image
Image("../../RESOURCES/chart-chooser-data-visualization.png")
```

Out[17]:



Aquí hay una herramienta online para ayudar a decidir el tipo de gráfico a usar

Matplotlib

Matplotlib es una librería para generar gráficos a partir de conjuntos de datos, bien sea datos contenidos en arrays de numpy o en dataframes de Pandas. En los siguientes casos utilizaremos esta librería para mostrar información del dataframe que hemos cargado con anterioridad on Pandas.

In [18]:

```
?matplotlib
```

In [19]:

```
import matplotlib.pyplot as plt
```

Scatter Plot / Gráfico de dispersión

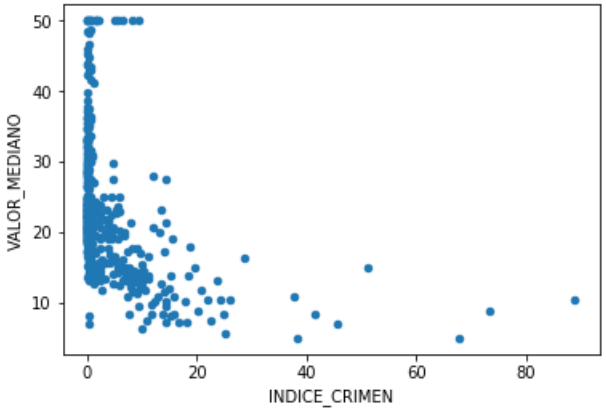
Los gráficos de dispersión son una de las mejores formas de representar la relación entre dos variables. Según Wikipedia Un diagrama de dispersión se emplea cuando una o varias variables está bajo el control del experimentador. Si existe un parámetro que se incrementa o disminuye de forma sistemática por el experimentador, se le denomina parámetro de control o variable independiente y habitualmente se representa a lo largo del eje horizontal (eje de las abscisas). La variable medida o dependiente usualmente se representa a lo largo del eje vertical (eje de las ordenadas). Si no existe una variable dependiente, cualquier variable se puede representar en cada eje y el diagrama de dispersión mostrará el grado de correlación (no causalidad) entre las dos variables.

In [20]:

```
#Usamos la funcion scatter especificando los ejex `x` e `y`
df.plot.scatter(x="INDICE_CRIMEN", y="VALOR_MEDIANO")
```

Out[20]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f8d2641cdd8>
```



In [23]:

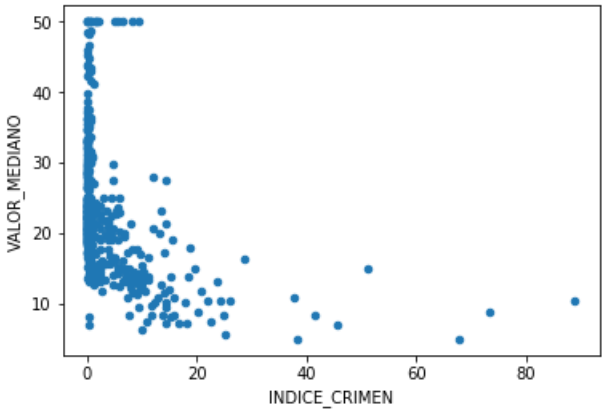
```
#Este comando adapta la grafica al documento
%matplotlib notebook
%matplotlib inline
```

In [24]:

```
df.plot.scatter(x="INDICE_CRIMEN", y="VALOR_MEDIANO")
```

Out[24]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f8d26348208>
```



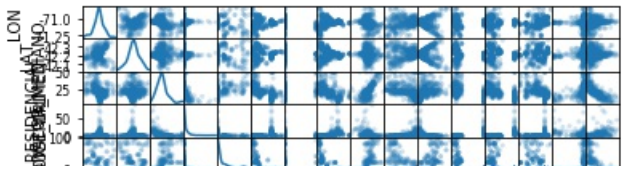
Matriz de dispersión

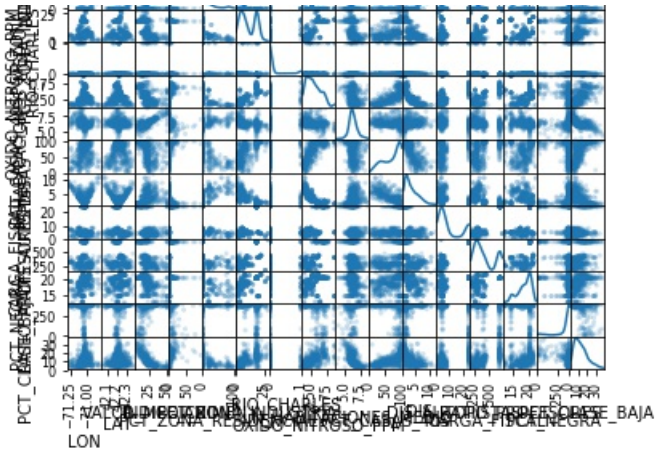
Un gráfico de matriz de dispersión es una herramienta de exploración de datos que permite comparar varios datasets para buscar patrones y relaciones.

El gráfico de matriz de dispersión (SPM) acepta una capa o una tabla como entrada. Seleccione los campos que desee utilizar en el gráfico. El gráfico tiene dos componentes principales: una matriz de gráficos de dispersión pequeños para cada uno de los campos y una ventana Vista previa mayor que muestra el gráfico de dispersión para un par de campos seleccionados con mayor detalle. También puede habilitar el trazado de histogramas, mostrando la distribución de valores para cada uno de los campos.

In [25]:

```
from pandas.plotting import scatter_matrix

sm = scatter_matrix(df, alpha=0.2, figsize=(6, 6), diagonal='kde')
```





In [26]:

```
sm
```

Out[26]:

```
array([[<matplotlib.axes._subplots.AxesSubplot object at 0x7f8d262aa8d0>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d2640db38>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d2eeb2f98>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d2ebdf0b8>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d2627f748>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d2627f780>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d262584a8>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d26201b38>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d261b3208>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d261da898>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d26185f28>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d261355f8>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d2615bc88>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d2610f358>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d260b79e8>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d260680b8>],
       [<matplotlib.axes._subplots.AxesSubplot object at 0x7f8d2608f748>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d26039dd8>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d25fe94a8>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d2600fb38>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d25fc4208>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d25f6d898>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d25f96f28>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d25f475f8>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d25eeec88>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d25ea0358>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d25ec89e8>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d25e7a0b8>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d25e22748>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d25e4ddd8>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d25dfb4a8>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d25da3b38>],
       [<matplotlib.axes._subplots.AxesSubplot object at 0x7f8d25dd8208>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d25d7e898>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d25d27f28>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d25d595f8>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d25d02c88>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d25cb3358>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d25cdc9e8>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d25c8e0b8>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d25c36748>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d25c5ddd8>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d25c0e4a8>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d25bb5b38>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d25b69208>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d25b91898>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d25b3bf28>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d25aeb5f8>],
       [<matplotlib.axes._subplots.AxesSubplot object at 0x7f8d25b13c88>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d25ac4358>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d25a6c9e8>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d25a9f0b8>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d25a48748>,
        <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d259f2dd8>,
```

```
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d259a24a8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d259c9b38>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d2597b208>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d25923898>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d2594cf28>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d258fd5f8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d258a6c88>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d258d7358>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d2587f9e8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d258320b8>],
[<matplotlib.axes._subplots.AxesSubplot object at 0x7f8d25859748>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d25801dd8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d257b34a8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d257dbb38>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d2578d208>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d25736898>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d2575ef28>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d257105f8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d256b6c88>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d25669358>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d256919e8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d256420b8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d255ec748>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d25615dd8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d255c54a8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d2556fb38>],
[<matplotlib.axes._subplots.AxesSubplot object at 0x7f8d25521208>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d25545898>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d254f0f28>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d254a15f8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d254c8c88>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d2547a358>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d254249e8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d254560b8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d253fc748>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d253a7dd8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d253d84a8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d2537fb38>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d25333208>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d2535a898>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d25301f28>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d252b25f8>],
[<matplotlib.axes._subplots.AxesSubplot object at 0x7f8d252dcc88>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d2528e358>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d252339e8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d251e70b8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d25210748>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d251b8dd8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d2516b4a8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d25192b38>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d25144208>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d250eb898>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d25113f28>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d250c55f8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d2506dc88>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d25020358>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d250469e8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d24ffa0b8>],
[<matplotlib.axes._subplots.AxesSubplot object at 0x7f8d24fa2748>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d24fcbdd8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d24f7c4a8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d24f25b38>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d24f56208>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d24efd898>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d24ea6f28>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d24ed85f8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d24e82c88>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d24e31358>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d24e5a9e8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d24e0d0b8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d24db4748>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d24ddedd8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d24d904a8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d24d37b38>],
[<matplotlib.axes._subplots.AxesSubplot object at 0x7f8d24ce9208>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d24d12898>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d24cb9f28>,
```

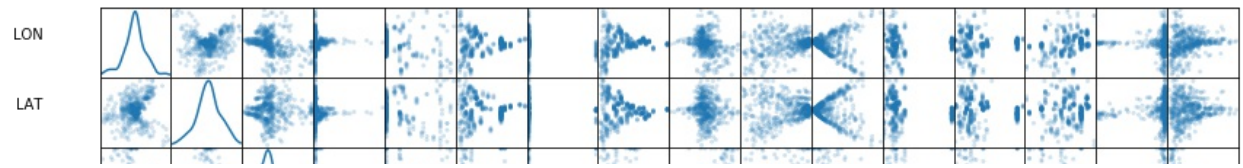
```
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d24c6b5f8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d24c92c88>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d24c43358>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d24bec9e8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d24c1f0b8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d24bc5748>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d24b6fdd8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d24b214a8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d24b48b38>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d24afc208>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d24aa1898>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d24acdf28>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d24a7e5f8>],
[<matplotlib.axes._subplots.AxesSubplot object at 0x7f8d24a26c88>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d24a57358>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d249fe9e8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d249b10b8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d249d9748>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d2497fdd8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d249334a8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d2495cb38>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d2490d208>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d248b5898>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d248def28>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d248905f8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d24837c88>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d247e7358>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d248129e8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d247c30b8>],
[<matplotlib.axes._subplots.AxesSubplot object at 0x7f8d2476a748>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d24792dd8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d247464a8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d246ebb38>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d2471d208>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d246c7898>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d24672f28>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d246205f8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d2464cc88>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d245fd358>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d245a29e8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d245d60b8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d2457e748>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d24524dd8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d245574a8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d24500b38>],
[<matplotlib.axes._subplots.AxesSubplot object at 0x7f8d244b1208>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d244d9898>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d24482f28>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d244335f8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d2445bc88>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d2440d358>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d243b59e8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d243680b8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d2438f748>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d24336dd8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d242e94a8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d2430fb38>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d242c2208>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d2426b898>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d24296f28>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d242475f8>],
[<matplotlib.axes._subplots.AxesSubplot object at 0x7f8d241efc88>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d2421f358>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d241c79e8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d2417a0b8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d24121748>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d2414bdd8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d240fb4a8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d240a3b38>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d240d6208>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d2407d898>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d24026f28>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d240575f8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d24000c88>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d23fb1358>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d23fdc9e8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d23f8d0b8>],
```

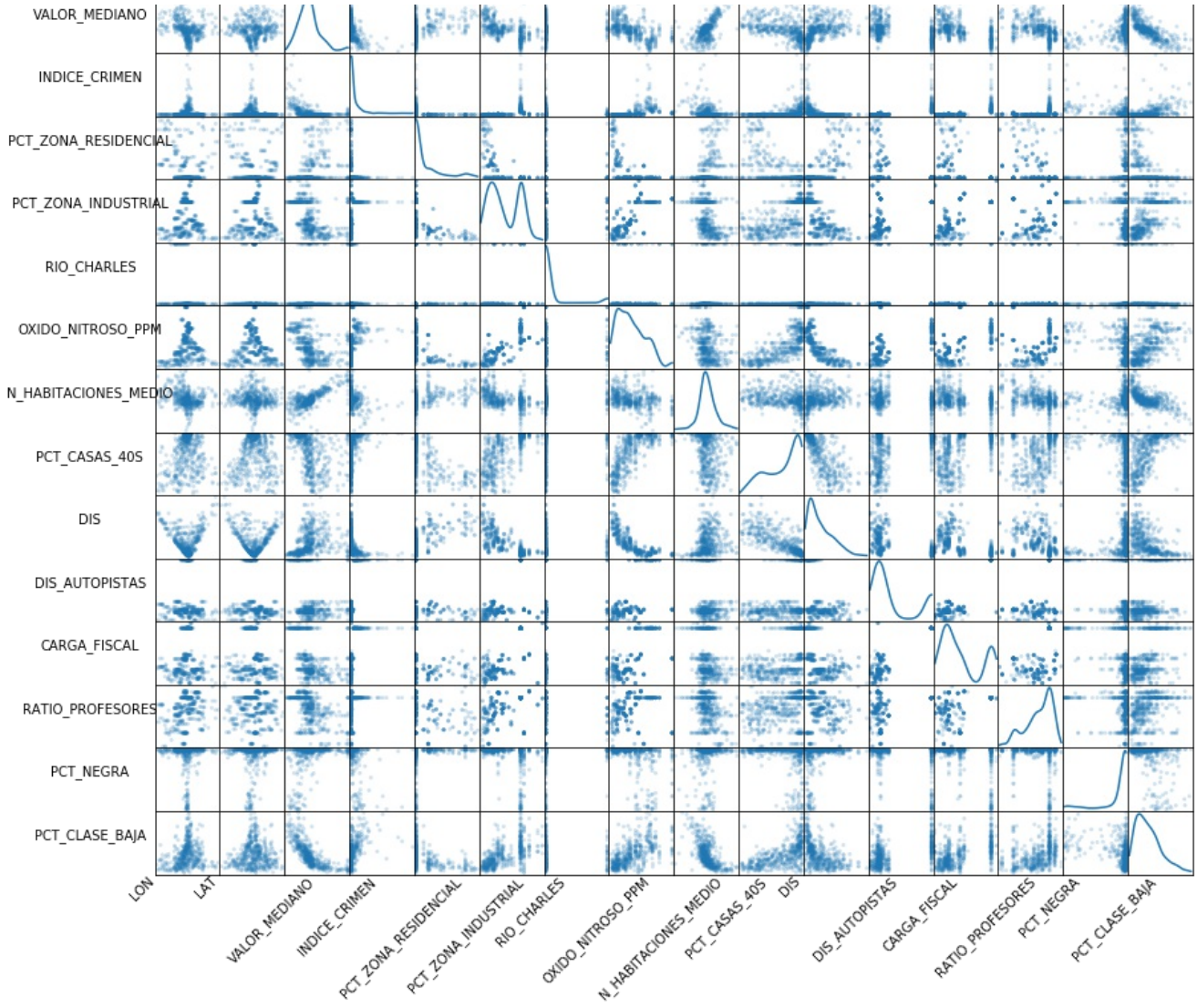
```
[<matplotlib.axes._subplots.AxesSubplot object at 0x7f8d23f35748>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d23f5ddd8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d23f0d4a8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d23eb4b38>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d23e6a208>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d23e90898>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d23e3bf28>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d23dec5f8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d23e13c88>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d23dc4358>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d23d6d9e8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d23d9e0b8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d23d46748>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d23cefdd8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d23ca14a8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d23cc8b38>],
[<matplotlib.axes._subplots.AxesSubplot object at 0x7f8d23c7b208>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d23c21898>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d23c4bf28>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d23bfc5f8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d23ba4c88>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d23bd5358>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d23b7f9e8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d23b310b8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d23b59748>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d23b00dd8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d23ab34a8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d23addb38>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d23a8c208>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d23a35898>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d23a5ef28>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d23a0d5f8>],
[<matplotlib.axes._subplots.AxesSubplot object at 0x7f8d239b6c88>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d2396b358>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d239909e8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d239410b8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d238ed748>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d23913dd8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d238c44a8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d2386eb38>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d2389f208>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d23846898>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d237eef28>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d2381f5f8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d237c6c88>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d2377a358>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d237239e8>,
 <matplotlib.axes._subplots.AxesSubplot object at 0x7f8d237550b8>]],
dtype=object)
```

Este es uno de los problemas de matplotlib, que su api es bastante complicada si quieres hacer algo que se salga de lo corriente. Por ejemplo, en este caso necesitamos un montón de código para rotar las etiquetas de los ejes

In [27]:

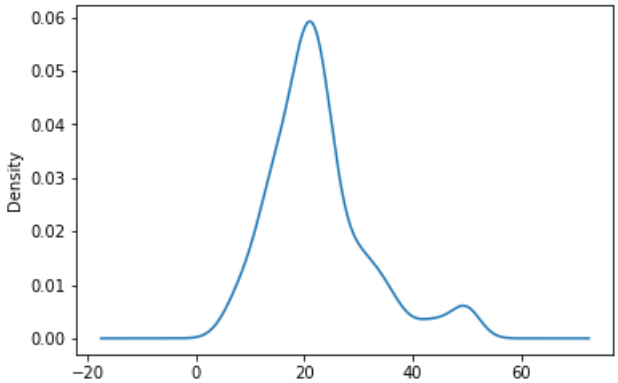
```
sm = scatter_matrix(df, alpha=0.2, figsize=(14, 14), diagonal='kde')

#https://stackoverflow.com/questions/32560932/how-to-customize-a-scatter-matrix-to-see-all-titles
#Change label rotation
[s.xaxis.label.set_rotation(45) for s in sm.reshape(-1)]
[s.yaxis.label.set_rotation(0) for s in sm.reshape(-1)]

#May need to offset label when rotating to prevent overlap of figure
[s.get_yaxis().set_label_coords(-1,0.5) for s in sm.reshape(-1)]
[s.get_xaxis().set_label_coords(-0.2,0) for s in sm.reshape(-1)]

#Hide all ticks
[s.set_xticks(()) for s in sm.reshape(-1)]
[s.set_yticks(()) for s in sm.reshape(-1)];
```





In [28]:

```
df.VALOR_MEDIANO.plot.kde();
```



Histograma

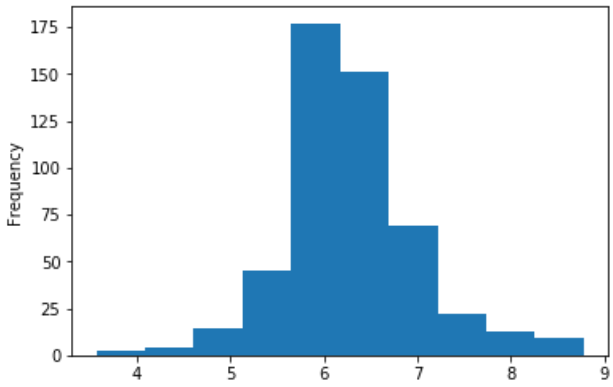
Los histogramas se usan para representar la distribucion de una variable, esto es, que rango de valores tiene, cuales son los valores más comunes Segun Wikipedia un histograma es una representación gráfica de una variable en forma de barras, donde la superficie de cada barra es proporcional a la frecuencia de los valores representados. Sirven para obtener una "primera vista" general, o panorama, de la distribución de la población, o de la muestra, respecto a una característica, cuantitativa y continua (como la longitud o el peso). De esta manera ofrece una visión de grupo permitiendo observar una preferencia, o tendencia, por parte de la muestra o población por ubicarse hacia una determinada región de valores dentro del espectro de valores posibles (sean infinitos o no) que pueda adquirir la característica. Así pues, podemos evidenciar comportamientos, observar el grado de homogeneidad, acuerdo o concisión entre los valores de todas las partes que componen la población o la muestra, o, en contraposición, poder observar el grado de variabilidad, y por ende, la dispersión de todos los valores que toman las partes, también es posible no evidenciar ninguna tendencia y obtener que cada miembro de la población toma por su lado y adquiere un valor de la característica aleatoriamente sin mostrar ninguna preferencia o tendencia, entre otras cosas.

In [29]:

```
df.N_HABITACIONES_MEDIO.plot.hist()
```

Out[29]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f8d19526780>
```



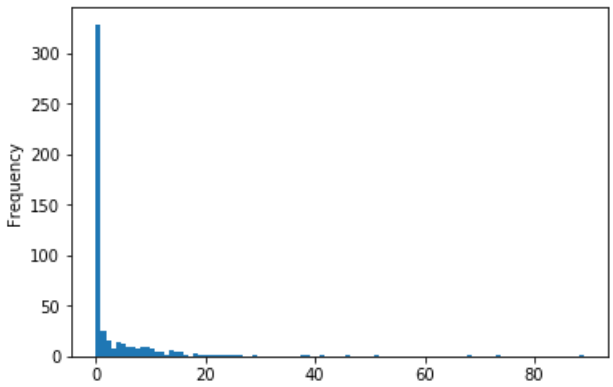
podemos especificar cuantos grupos queremos en el histograma

In [30]:

```
df.INDICE_CRIMEN.plot.hist(bins=100)
```

Out[30]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f8d194be7f0>
```



Tambien podemos filtrar el gráfico poniendo límites a los ejes

In [31]:

```
df.INDICE_CRIMEN.plot.hist(bins=100, xlim=(0,20))
```

Out[31]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f8d199e8be0>
```

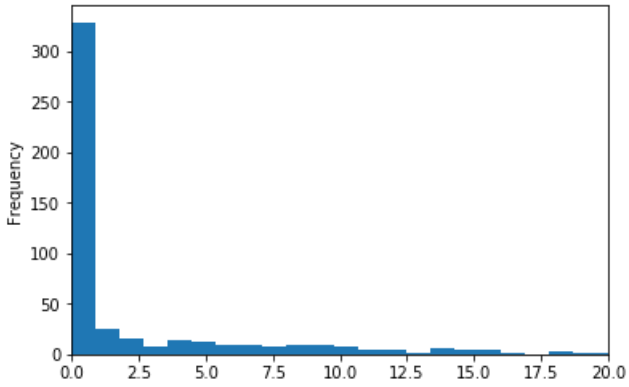


Gráfico de barras/columnas

Los gráficos de barras se utilizan comúnmente para representar y comparar una variable entre distintos grupos

In [32]:

```
valor_por_ciudad = df.groupby("CIUDAD")["VALOR_MEDIANO"].mean()
valor_por_ciudad.head()
```

Out[32]:

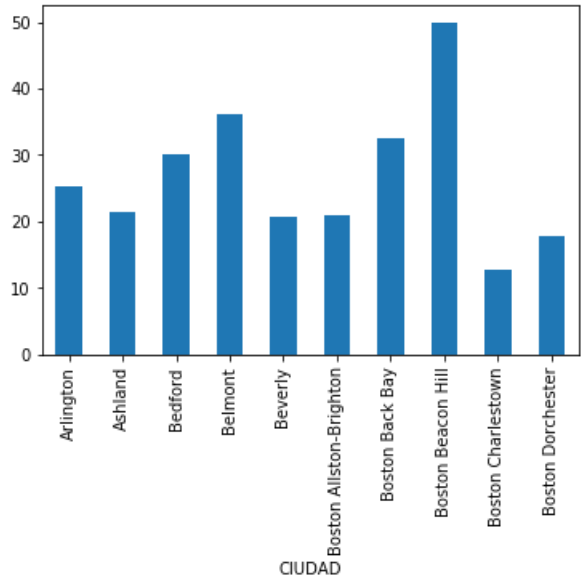
```
CIUDAD
Arlington     25.2
Ashland       21.4
Bedford       30.1
Belmont       36.2
Beverly       20.8
Name: VALOR_MEDIANO, dtype: float64
```

In [33]:

```
valor_por_ciudad.head(10).plot.bar()
```

Out[33]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f8d1980c208>
```



In [34]:

```
valor_por_ciudad.head(10).plot.barh()
```

Out[34]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f8d19860080>
```

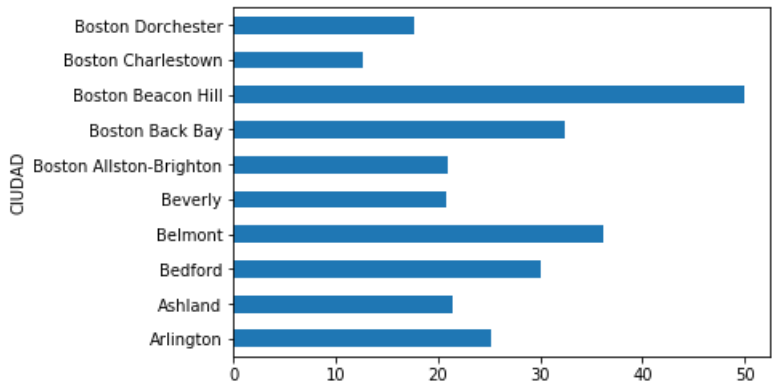


Gráfico de linea

Los gráficos de linea se usan principalmente para representar tendencias, esto es, se usan para variables que varian con el tiempo

In [35]:

```
df.groupby("RATIO_PROFESORES").VALOR_MEDIANO.mean().plot.line();
```

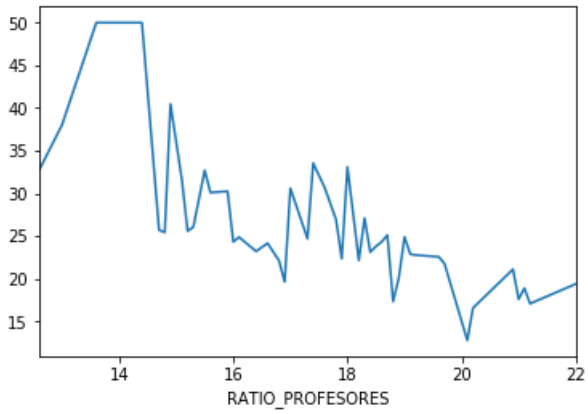


Diagrama de caja (Box Plot)

Los diagramas de caja son útiles a la hora de representar grupos de datos y comparar entre ellos. Otra ventaja de los boxplots es que identifican de forma sencilla si una variable tiene muchos outliers, esto es, elementos que se alejan de los valores frecuentes de dicha variable.

In [36]:

```
df["VALOR_CUANTILES"] = pd.qcut(df.VALOR_MEDIANO, 5)
```

In [37]:

```
df.boxplot(column="INDICE_CRIMEN", by="VALOR_CUANTILES", figsize=(10,10))
```

Out[37]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f8d19738b38>
```

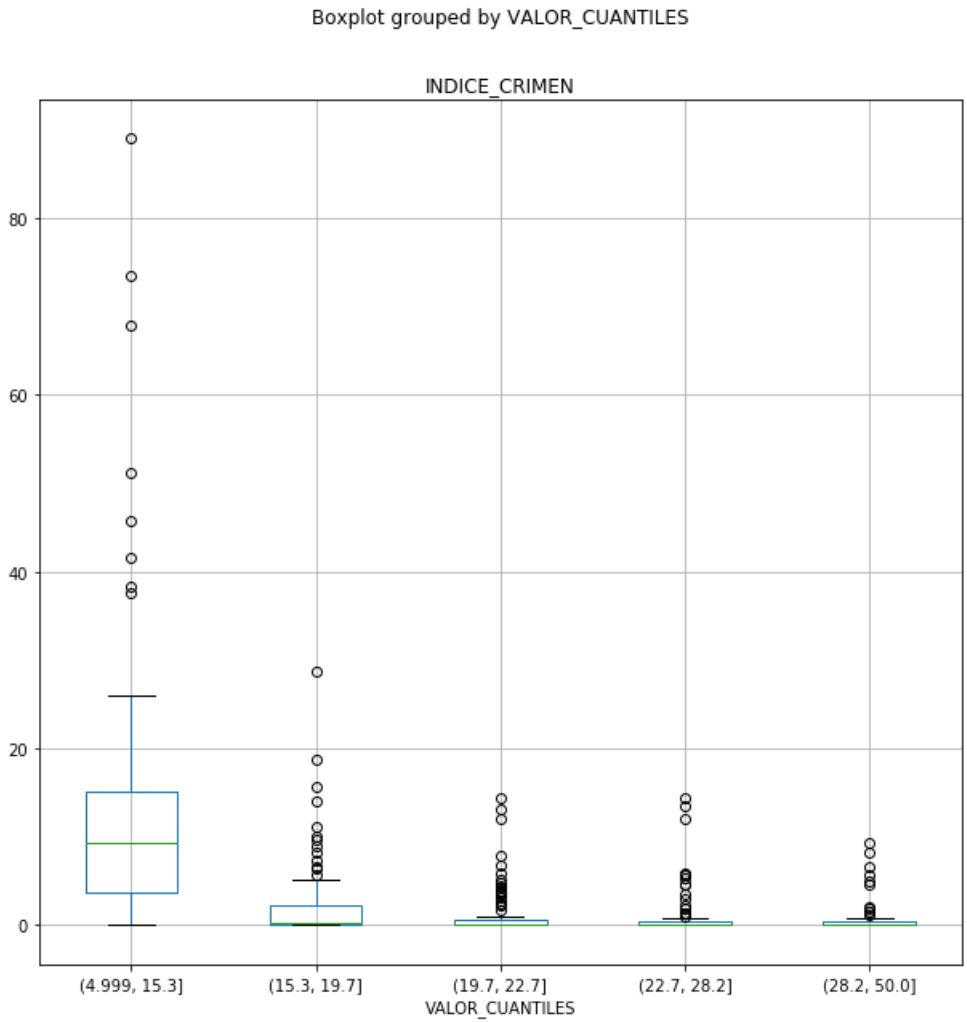


Gráfico circular

Un gráfico circular es una representación gráfica de una serie de cantidades y consiste en un círculo dividido en varios sectores, cuyo tamaño se corresponde con las proporciones de las cantidades. Básicamente, este tipo de gráfico muestra la relación porcentual entre las partes con relación a su conjunto.

In [38]:

```
df.RIO_CHARLES.value_counts().plot.pie()
```

Out[38]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f8d1912b588>
```

https://conda-forge.github.io/

```
https://anaconda.org/conda-forge/repo
    https://seaborn.pydata.org/

        http://bokeh.pydata.org/en/latest/

            https://altair-viz.github.io/

                https://plot.ly/python/getting-started/
```

https://ipywidgets.readthedocs.io/en/latest/examples/Using%20Interact.html `conda install -c conda-forge ipywidgets`

https://github.com/bokeh/bokeh/blob/0.12.5/examples/howto/notebook_comms/Jupyter%20Interactors.ipynb