16jkonjrg

August 3, 2023

1 20104169 - SUMESH R

2 Importing Libraries

[1]:
```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

[2]:
```
from google.colab import drive
drive.mount('/content/drive')
df=pd.read_csv("/content/drive/MyDrive/mydatasets/stations.csv")
df
```

```
Mounted at /content/drive
```

[2]:

	id	name
0	28079004	Pza. de España
1	28079008	Escuelas Aguirre
2	28079011	Avda. Ramón y Cajal
3	28079016	Arturo Soria
4	28079017	Villaverde
5	28079018	Farolillo
6	28079024	Casa de Campo
7	28079027	Barajas Pueblo
8	28079035	Pza. del Carmen
9	28079036	Moratalaz
10	28079038	Cuatro Caminos
11	28079039	Barrio del Pilar
12	28079040	Vallecas
13	28079047	Mendez Alvaro
14	28079048	Castellana
15	28079049	Parque del Retiro
16	28079050	Plaza Castilla
17	28079054	Ensanche de Vallecas
18	28079055	Urb. Embajada
19	28079056	Pza. Fernández Ladreda

```
20  28079057             Sanchinarro
21  28079058                El Pardo
22  28079059           Juan Carlos I
23  28079060             Tres Olivos

                                                 address       lon        lat  \
0                                      Plaza de España -3.712247  40.423853
1                        Entre C/ Alcalá y C/ O' Donell  -3.682319  40.421564
2    Avda. Ramón y Cajal  esq. C/ Príncipe de Vergara -3.677356  40.451475
3   C/ Arturo Soria  esq. C/  Vizconde de los Asilos  -3.639233  40.440047
4                                    C/. Juan Peñalver -3.713322  40.347139
5                          Calle Farolillo - C/Ervigio -3.731853  40.394781
6             Casa de Campo  (Terminal del Teleférico) -3.747347  40.419356
7                           C/. Júpiter, 21 (Barajas)  -3.580031  40.476928
8                  Plaza del Carmen esq. Tres Cruces.  -3.703172  40.419208
9         Avd. Moratalaz  esq. Camino de los Vinateros -3.645306  40.407947
10        Avda. Pablo Iglesias esq. C/ Marqués de Lema -3.707128  40.445544
11           Avd. Betanzos esq. C/  Monforte de Lemos  -3.711542  40.478228
12         C/ Arroyo del Olivar  esq. C/  Río Grande.  -3.651522  40.388153
13   C/ Juan de Mariana / Pza. Amanecer Mendez Alvaro -3.686825  40.398114
14                           C/ Jose Gutierrez Abascal -3.690367  40.439897
15                     Paseo Venezuela- Casa de Vacas -3.682583  40.414444
16                              Plaza Castilla (Canal) -3.688769  40.465572
17                  Avda La Gavia / Avda. Las Suertes -3.612117  40.372933
18                               C/ Riaño (Barajas)  -3.580747  40.462531
19              Pza. Fernández Ladreda - Avda. Oporto -3.718728  40.384964
20            C/ Princesa de Eboli esq C/ Maria Tudor -3.660503  40.494208
21                                   Avda. La Guardia -3.774611  40.518058
22  Parque Juan Carlos I (frente oficinas mantenim… -3.609072  40.465250
23                                Plaza Tres Olivos  -3.689761  40.500589

    elevation
0         635
1         670
2         708
3         693
4         604
5         630
6         642
7         621
8         659
9         685
10        698
11        674
12        677
13        599
14        676
```

```
15        662
16        728
17        627
18        618
19        604
20        700
21        615
22        660
23        715
```

3 Data Cleaning and Data Preprocessing

[3]:
```
df=df.dropna()
```

[4]:
```
df.columns
```

[4]:
```
Index(['id', 'name', 'address', 'lon', 'lat', 'elevation'], dtype='object')
```

[5]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 24 entries, 0 to 23
Data columns (total 6 columns):
 #   Column     Non-Null Count  Dtype
---  ------     --------------  -----
 0   id         24 non-null     int64
 1   name       24 non-null     object
 2   address    24 non-null     object
 3   lon        24 non-null     float64
 4   lat        24 non-null     float64
 5   elevation  24 non-null     int64
dtypes: float64(2), int64(2), object(2)
memory usage: 1.2+ KB
```

[6]:
```
data=df[['lat', 'elevation']]
data
```

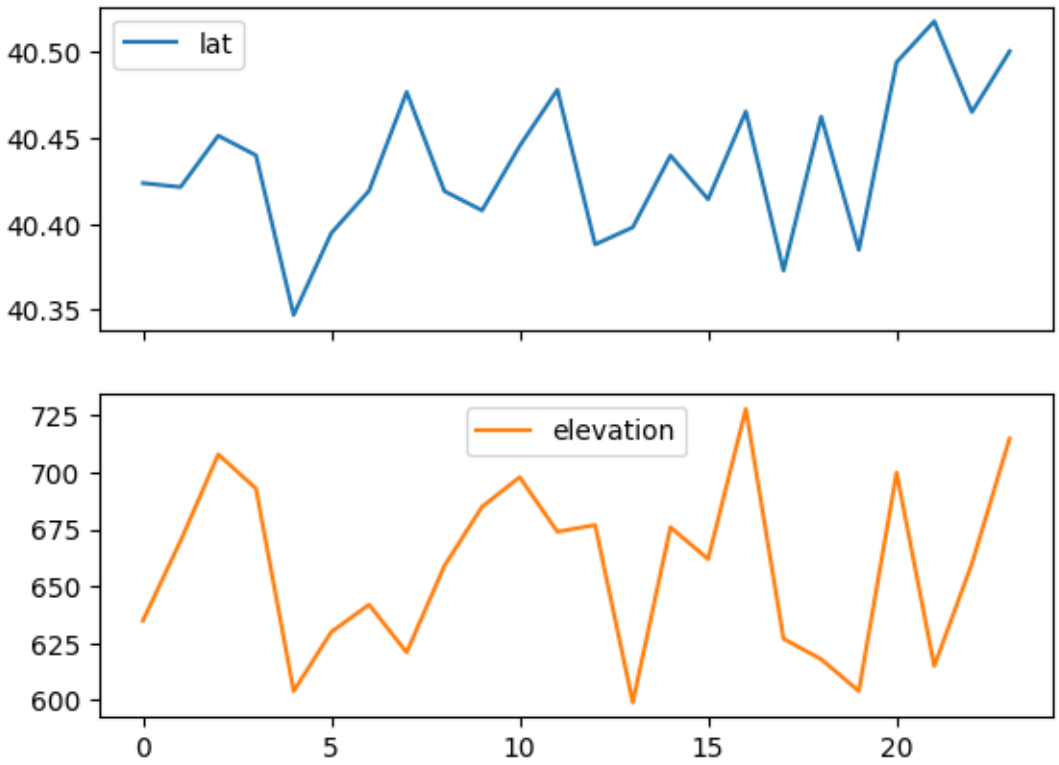
[6]:
```
          lat  elevation
0   40.423853        635
1   40.421564        670
2   40.451475        708
3   40.440047        693
4   40.347139        604
5   40.394781        630
6   40.419356        642
7   40.476928        621
```

8	40.419208	659
9	40.407947	685
10	40.445544	698
11	40.478228	674
12	40.388153	677
13	40.398114	599
14	40.439897	676
15	40.414444	662
16	40.465572	728
17	40.372933	627
18	40.462531	618
19	40.384964	604
20	40.494208	700
21	40.518058	615
22	40.465250	660
23	40.500589	715

4 Line chart

[7]:
```
data.plot.line(subplots=True)
```

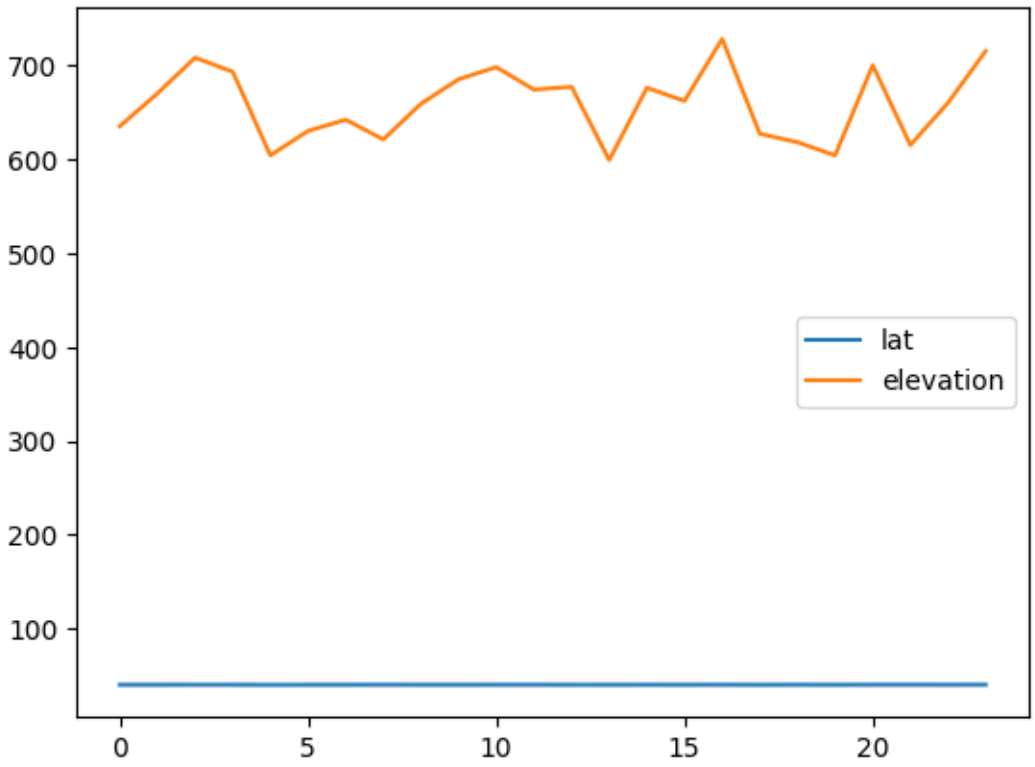
[7]:
```
array([<Axes: >, <Axes: >], dtype=object)
```



5 Line chart

[8]:
```
data.plot.line()
```

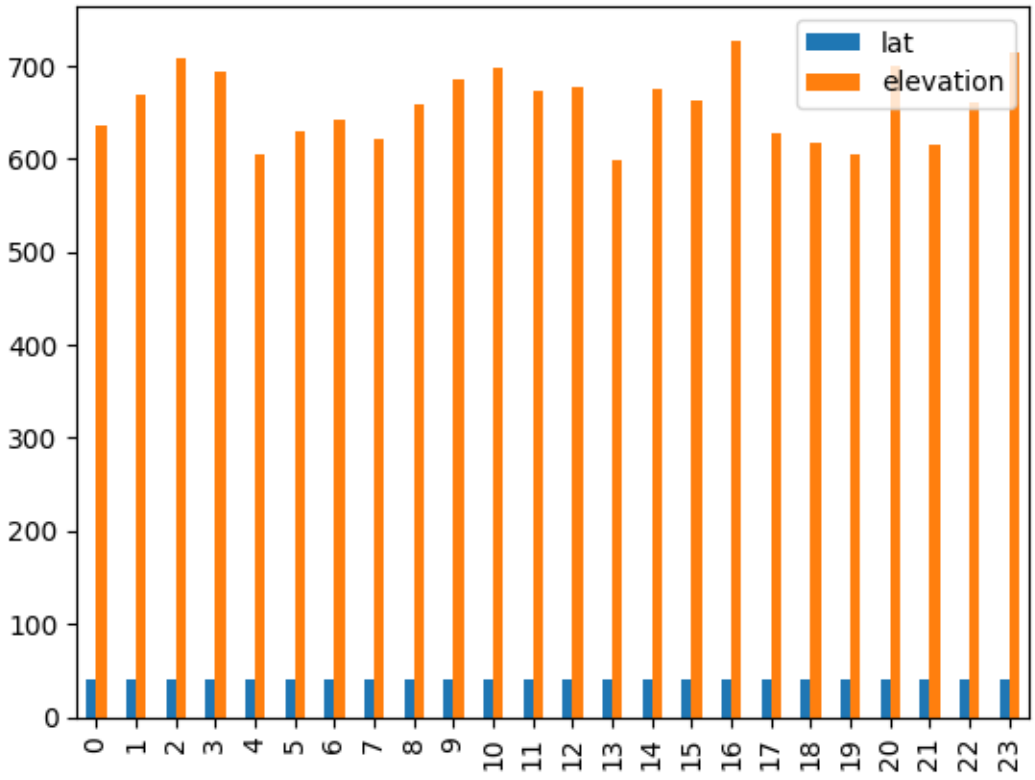
[8]:
```
<Axes: >
```



6 Bar chart

[9]:
```
data.plot.bar()
```

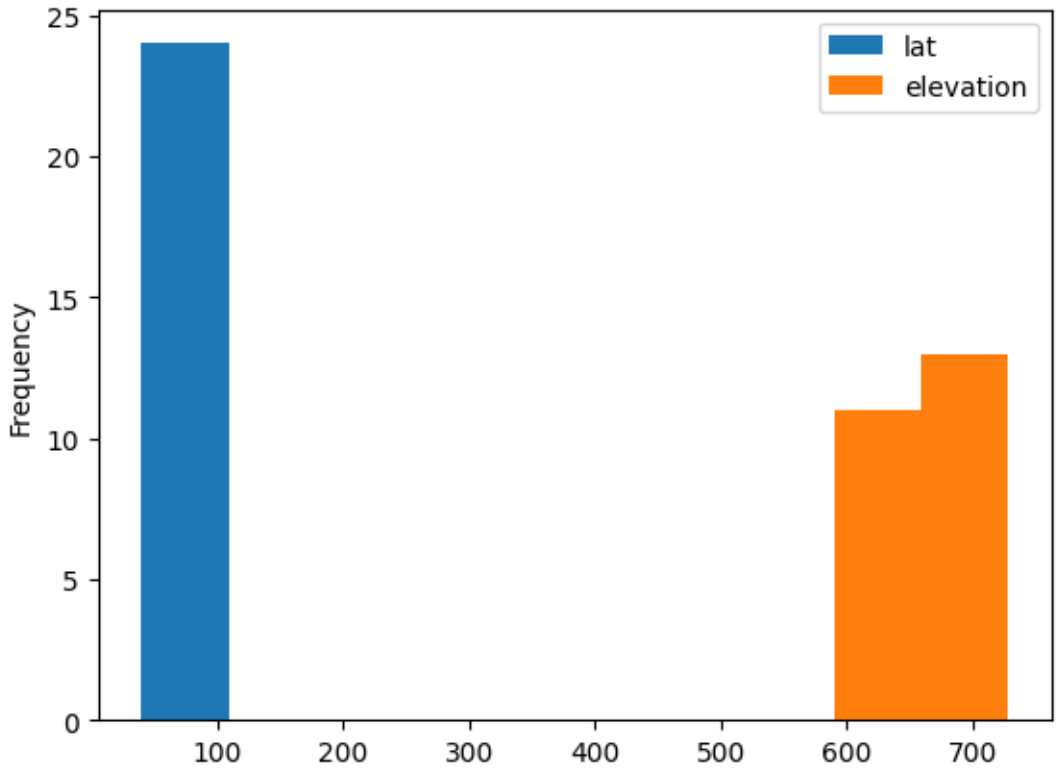
[9]:
```
<Axes: >
```



7 Histogram

[10]:
```
data.plot.hist()
```

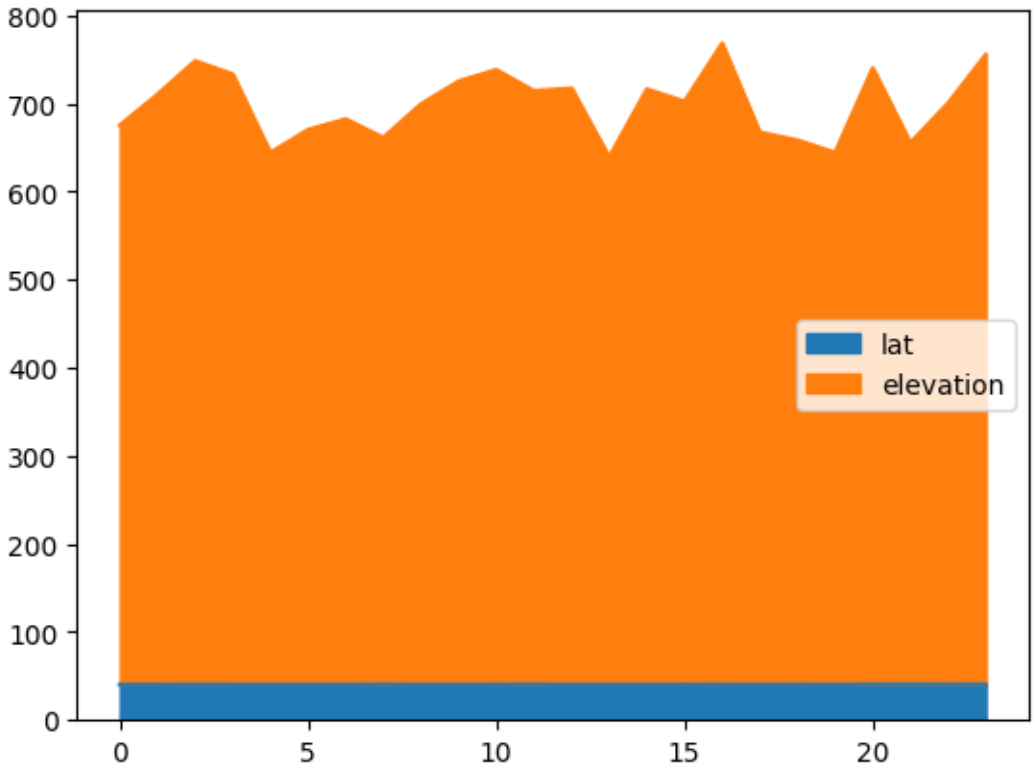
[10]: <Axes: ylabel='Frequency'>



8 Area chart

[11]:
```
data.plot.area()
```

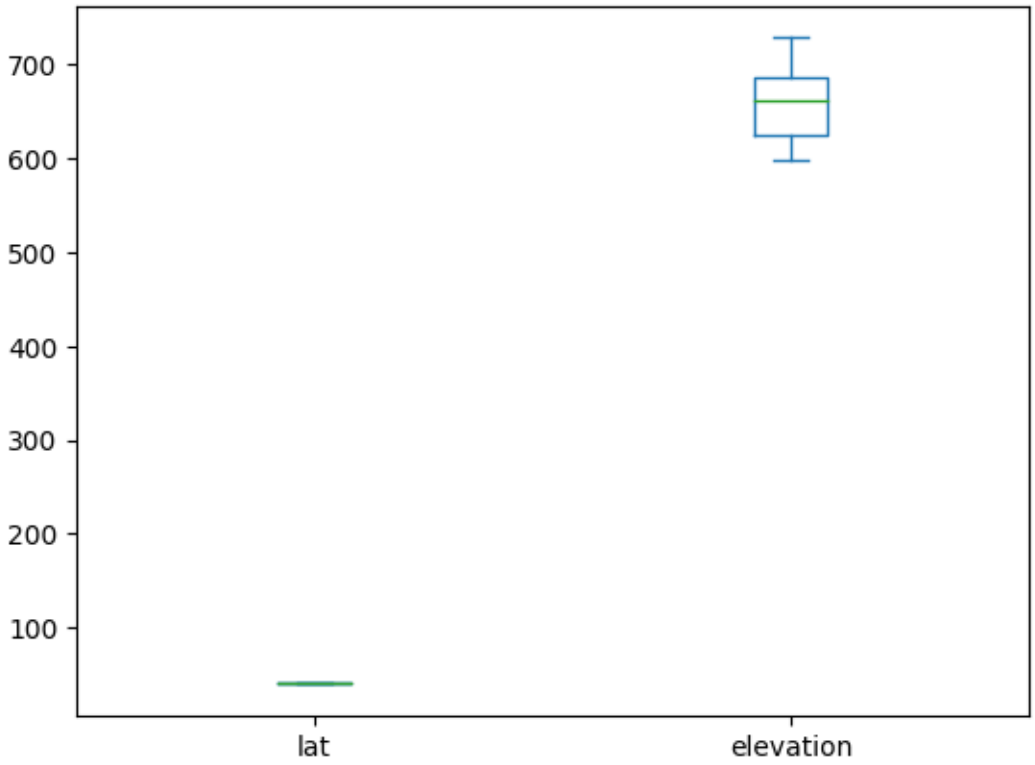
[11]: <Axes: >



9 Box chart

[12]:
```
data.plot.box()
```

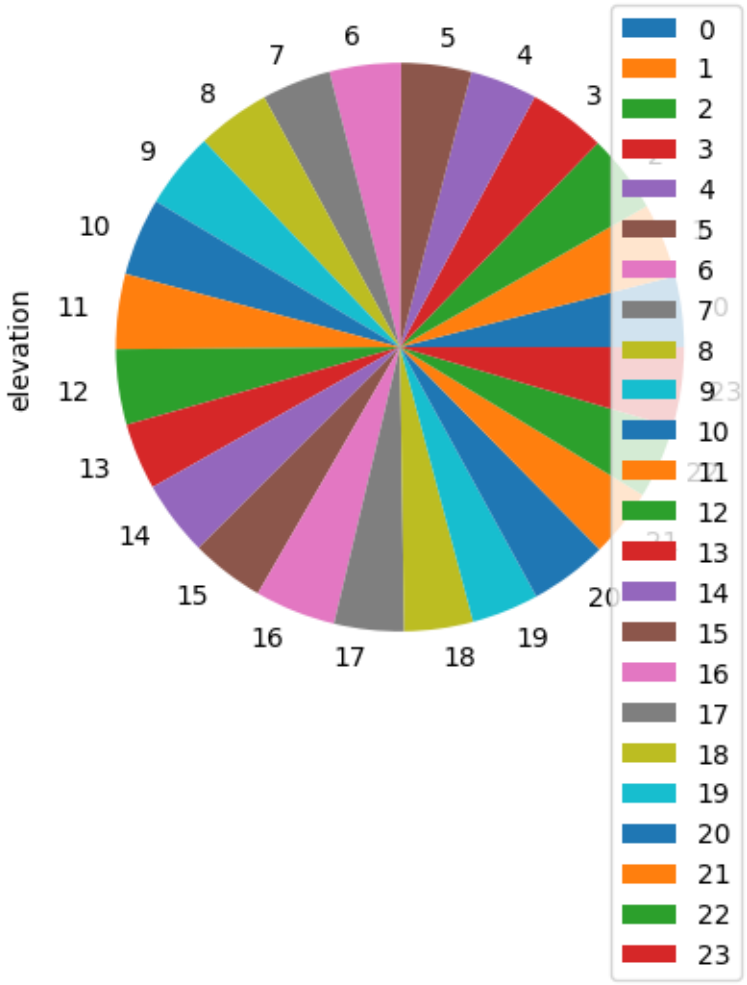
[12]: <Axes: >



10 Pie chart

[13]:
```
data.plot.pie(y='elevation' )
```

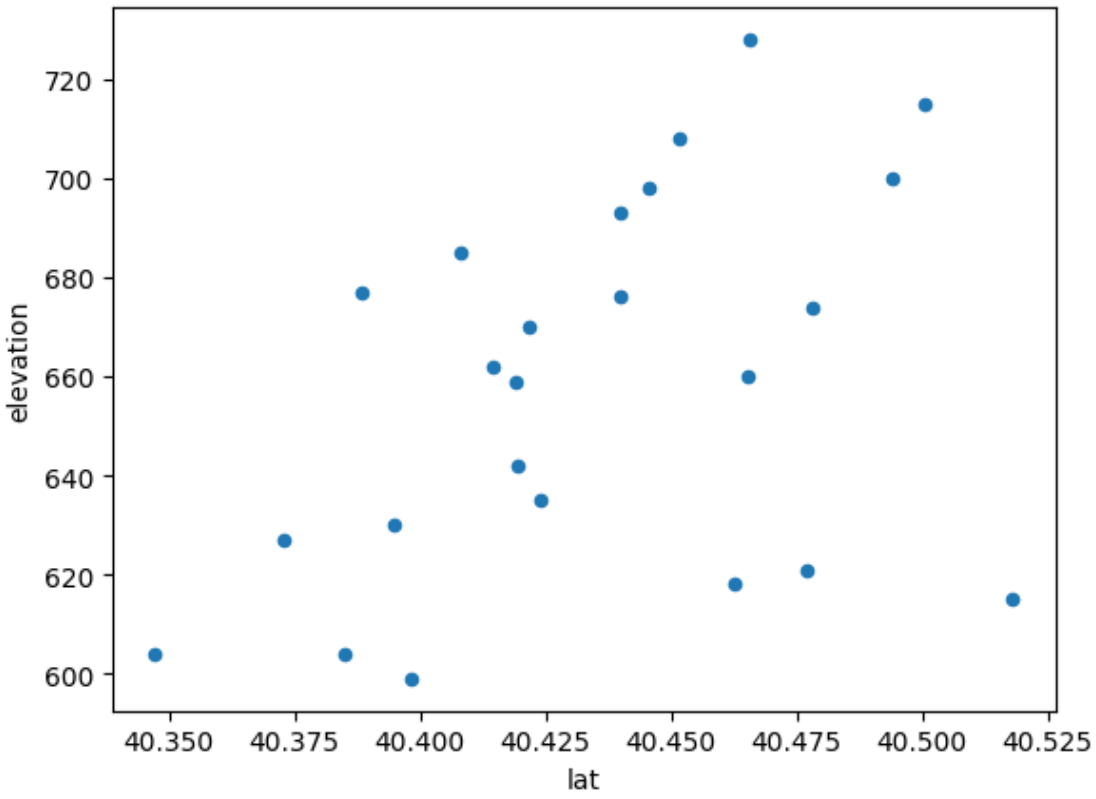
[13]:
```
<Axes: ylabel='elevation'>
```



11 Scatter chart

[14]:
```
data.plot.scatter(x='lat' ,y='elevation')
```

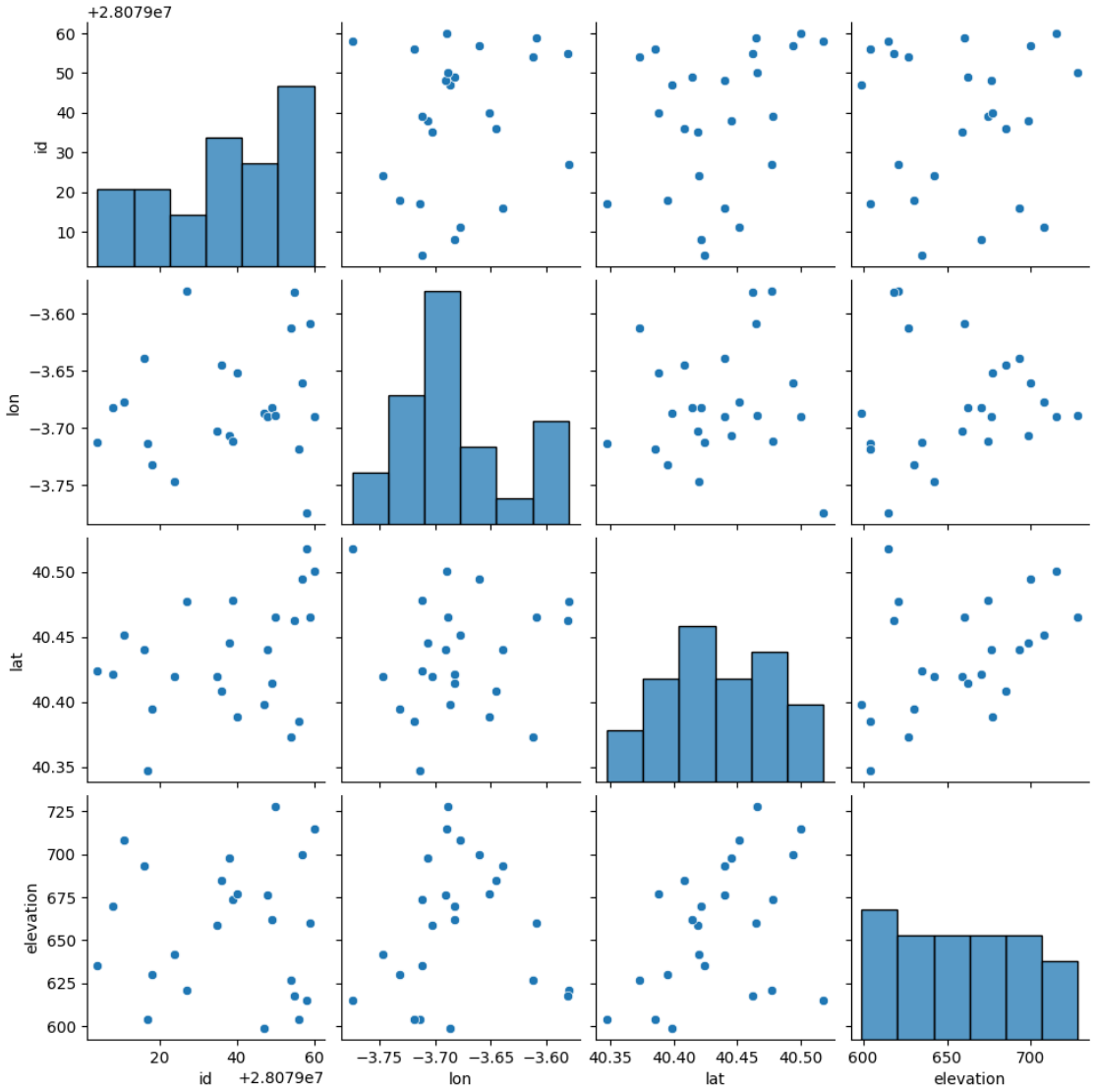
[14]: <Axes: xlabel='lat', ylabel='elevation'>



12 Seaborn

[15]:
```
sns.pairplot(df[0:50])
```

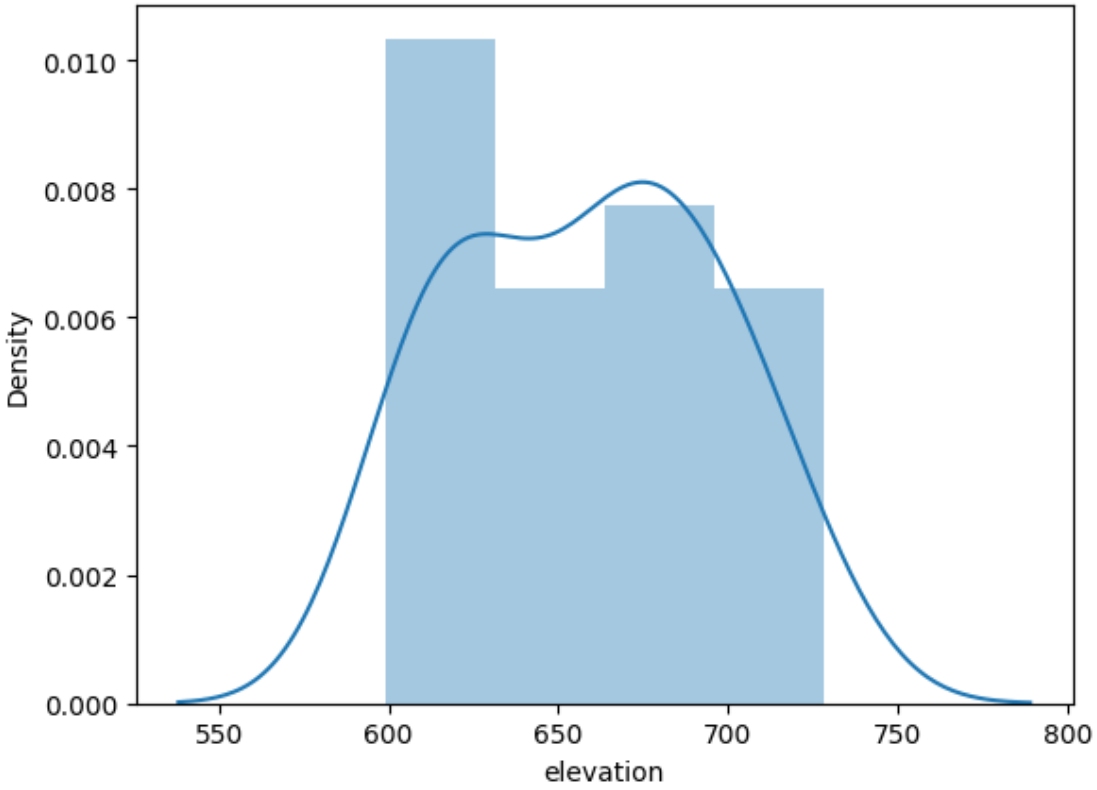
[15]:
```
<seaborn.axisgrid.PairGrid at 0x7f55a08fbcd0>
```



[16]:
```
sns.distplot(df['elevation'])
```

```
<ipython-input-16-67528c7a127b>:1: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(df['elevation'])
```

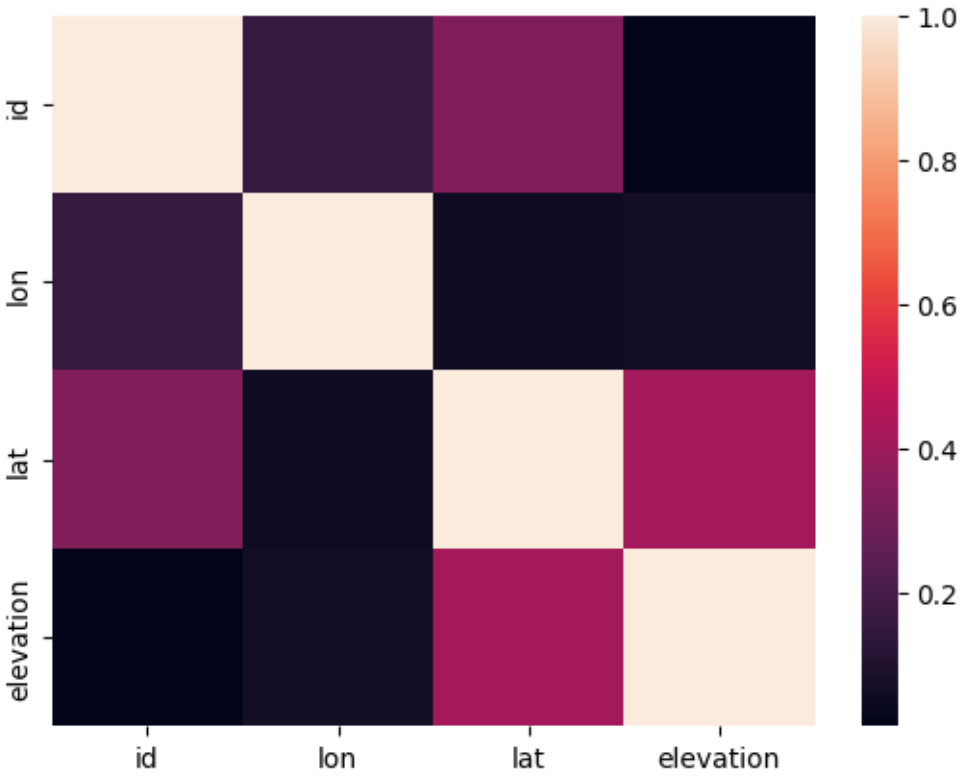
[16]: <Axes: xlabel='elevation', ylabel='Density'>



[17]:
```
sns.heatmap(df.corr())
```

```
<ipython-input-17-aa4f4450a243>:1: FutureWarning: The default value of
numeric_only in DataFrame.corr is deprecated. In a future version, it will
default to False. Select only valid columns or specify the value of numeric_only
to silence this warning.
  sns.heatmap(df.corr())
```

[17]: <Axes: >


id
lon
lat
elevation
id
lon
lat
elevation
1.0
0.8
0.6
0.4
0.2