In [1]:

```
import numpy as np
import linalg as la
```

Create 5 matrices with five different dimensions (1-D,2-D,...5-D)

In [2]:

```
a=np.arange(5)
print(a)
```

```
[0 1 2 3 4]
```

In [3]:

```
b=np.arange(4).reshape(2,2)
b
```

Out[3]:

```
array([[0, 1],
       [2, 3]])
```

In [4]:

```
x=np.arange(1,28)
c=x.reshape(3,3,3)
c
```

Out[4]:

```
array([[[ 1,  2,  3],
        [ 4,  5,  6],
        [ 7,  8,  9]],

       [[10, 11, 12],
        [13, 14, 15],
        [16, 17, 18]],

       [[19, 20, 21],
        [22, 23, 24],
        [25, 26, 27]]])
```

In [5]:

```
y=np.arange(5,69)
d=y.reshape(4,4,4)
d
```

Out[5]:

```
array([[[ 5,  6,  7,  8],
        [ 9, 10, 11, 12],
        [13, 14, 15, 16],
        [17, 18, 19, 20]],

       [[21, 22, 23, 24],
        [25, 26, 27, 28],
        [29, 30, 31, 32],
        [33, 34, 35, 36]],

       [[37, 38, 39, 40],
        [41, 42, 43, 44],
        [45, 46, 47, 48],
        [49, 50, 51, 52]],

       [[53, 54, 55, 56],
        [57, 58, 59, 60],
        [61, 62, 63, 64],
        [65, 66, 67, 68]]])
```

In [6]:

```
z=np.arange(5,130)
e=z.reshape(5,5,5)
e
```

Out[6]:

```
array([[[  5,   6,   7,   8,   9],
        [ 10,  11,  12,  13,  14],
        [ 15,  16,  17,  18,  19],
        [ 20,  21,  22,  23,  24],
        [ 25,  26,  27,  28,  29]],

       [[ 30,  31,  32,  33,  34],
        [ 35,  36,  37,  38,  39],
        [ 40,  41,  42,  43,  44],
        [ 45,  46,  47,  48,  49],
        [ 50,  51,  52,  53,  54]],

       [[ 55,  56,  57,  58,  59],
        [ 60,  61,  62,  63,  64],
        [ 65,  66,  67,  68,  69],
        [ 70,  71,  72,  73,  74],
        [ 75,  76,  77,  78,  79]],

       [[ 80,  81,  82,  83,  84],
        [ 85,  86,  87,  88,  89],
        [ 90,  91,  92,  93,  94],
        [ 95,  96,  97,  98,  99],
        [100, 101, 102, 103, 104]],

       [[105, 106, 107, 108, 109],
        [110, 111, 112, 113, 114],
        [115, 116, 117, 118, 119],
        [120, 121, 122, 123, 124],
        [125, 126, 127, 128, 129]]])
```

Find determinants of 5 matrices and display your output

In [7]:

```
print(np.linalg.det(b))
print(np.linalg.det(c))
print(np.linalg.det(d))
print(np.linalg.det(e))
```

```
-2.0
[ 0.00000000e+00  0.00000000e+00 -2.13162821e-14]
[ 0.00000000e+00  0.00000000e+00 -6.05845175e-28  0.00000000e+00]
[ 0.00000000e+00 -7.17464814e-42  0.00000000e+00  0.00000000e+00
  5.73971851e-41]
```

Find inverse of the above 5 matrices and display your

In [8]:

```
print(np.linalg.inv(b))
```

```
[[-1.5  0.5]
 [ 1.   0. ]]
```

Find the rank, diagonal and trace of the 5 matrices

Rank

In [9]:

```
print(np.linalg.matrix_rank(a))
print(np.linalg.matrix_rank(b))
print(np.linalg.matrix_rank(c))
print(np.linalg.matrix_rank(d))
print(np.linalg.matrix_rank(e))
```

```
1
2
[2 2 2]
[2 2 2 2]
[2 2 2 2 2]
```

Diagonal

In [10]:

```
print(np.diag(a))
print(np.diag(b))
```

```
[[0 0 0 0 0]
 [0 1 0 0 0]
 [0 0 2 0 0]
 [0 0 0 3 0]
 [0 0 0 0 4]]
[0 3]
```

In [11]:

```
dia=[np.diag(i) for i in c]
for j in dia:
  print(j)
```

```
[1 5 9]
[10 14 18]
[19 23 27]
```

In [12]:

```
dia=[np.diag(i) for i in d]
for j in dia:
  print(j)
```

```
[ 5 10 15 20]
[21 26 31 36]
[37 42 47 52]
[53 58 63 68]
```

In [13]:

```
dia=[np.diag(i) for i in e]
for j in dia:
  print(j)
```

```
[ 5 11 17 23 29]
[30 36 42 48 54]
[55 61 67 73 79]
[ 80  86  92  98 104]
[105 111 117 123 129]
```

Find Eigen value and eigen vector for 5 matrices

Eigen Values

In [14]:

```
print(np.linalg.eigvals(b))
print(np.linalg.eigvals(c))
print(np.linalg.eigvals(d))
print(np.linalg.eigvals(e))
```

```
[-0.56155281  3.56155281]
[[ 1.61168440e+01 -1.11684397e+00 -1.30367773e-15]
 [ 4.24242853e+01 -4.24285286e-01 -8.76087811e-16]
 [ 6.92598907e+01 -2.59890679e-01  3.45459719e-15]]
[[ 5.15518361e+01+0.00000000e+00j -1.55183609e+00+0.00000000e+00j
   1.14139605e-15+7.74891647e-16j  1.14139605e-15-7.74891647e-16j]
 [ 1.14697487e+02+0.00000000e+00j -6.97486947e-01+0.00000000e+00j
   4.04539332e-15+0.00000000e+00j  1.70279653e-15+0.00000000e+00j]
 [ 1.78448309e+02+0.00000000e+00j -4.48309095e-01+0.00000000e+00j
  -1.78612149e-14+0.00000000e+00j -1.48008554e-15+0.00000000e+00j]
 [ 2.42330128e+02+0.00000000e+00j -3.30128163e-01+0.00000000e+00j
   7.54701580e-15+0.00000000e+00j -4.43308381e-15+0.00000000e+00j]]
[[ 8.78458929e+01+0.00000000e+00j -2.84589287e+00+0.00000000e+00j
   2.73009873e-15+0.00000000e+00j -2.94436099e-15+0.00000000e+00j
  -1.40262056e-15+0.00000000e+00j]
 [ 2.11183803e+02+0.00000000e+00j -1.18380291e+00+0.00000000e+00j
   2.79017982e-15+0.00000000e+00j -5.26498327e-15+0.00000000e+00j
   5.70697559e-15+0.00000000e+00j]
 [ 3.35744614e+02+0.00000000e+00j -7.44613584e-01+0.00000000e+00j
  -8.45894330e-15+0.00000000e+00j -5.01414612e-15+1.52241436e-15j
  -5.01414612e-15-1.52241436e-15j]
 [ 4.60542838e+02+0.00000000e+00j -5.42837668e-01+0.00000000e+00j
  -1.38672957e-14+0.00000000e+00j  2.68436920e-15+8.79224795e-15j
   2.68436920e-15-8.79224795e-15j]
 [ 5.85427039e+02+0.00000000e+00j -4.27038697e-01+0.00000000e+00j
   1.21503441e-14+0.00000000e+00j -5.56406501e-15+0.00000000e+00j
   1.00681571e-14+0.00000000e+00j]]
```

Eigen Vector

In [15]:

```
print(np.linalg.eig(e))
print(np.linalg.eig(b))
print(np.linalg.eig(c))
print(np.linalg.eig(d))
```

```
EigResult(eigenvalues=array([[ 8.78458929e+01+0.00000000e+00j, -2.84589
287e+00+0.00000000e+00j,
         2.73009873e-15+0.00000000e+00j, -2.94436099e-15+0.00000000e+00
j,
        -1.40262056e-15+0.00000000e+00j],
       [ 2.11183803e+02+0.00000000e+00j, -1.18380291e+00+0.00000000e+00
j,
         2.79017982e-15+0.00000000e+00j, -5.26498327e-15+0.00000000e+00
j,
         5.70697559e-15+0.00000000e+00j],
       [ 3.35744614e+02+0.00000000e+00j, -7.44613584e-01+0.00000000e+00
j,
        -8.45894330e-15+0.00000000e+00j, -5.01414612e-15+1.52241436e-15
j,
        -5.01414612e-15-1.52241436e-15j],
       [ 4.60542838e+02+0.00000000e+00j, -5.42837668e-01+0.00000000e+00
j,
        -1.38672957e-14+0.00000000e+00j,  2.68436920e-15+8.79224795e-15
j,
```

EDA:

Frame a problem statement, clean, preprocess and visulaize the data and interpret your conclusion

Write a python program to visualize the breast cancer prediction dataset with the help of pandas and matplotlib library and understand the relationship between the parameters to define the tumor is malignant or benign

In [16]:

```
import pandas as pd
from matplotlib import pyplot as plt
```

In [17]:

```
df=pd.read_csv("8_BreastCancerPrediction.csv")
df
```

Out[17]:

	id	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothn
0	842302	M	17.99	10.38	122.80	1001.0	
1	842517	M	20.57	17.77	132.90	1326.0	
2	84300903	M	19.69	21.25	130.00	1203.0	
3	84348301	M	11.42	20.38	77.58	386.1	
4	84358402	M	20.29	14.34	135.10	1297.0	
...	...	...	...	...	...	...	
564	926424	M	21.56	22.39	142.00	1479.0	
565	926682	M	20.13	28.25	131.20	1261.0	
566	926954	M	16.60	28.08	108.30	858.1	
567	927241	M	20.60	29.33	140.10	1265.0	
568	92751	B	7.76	24.54	47.92	181.0	

569 rows × 33 columns

In [18]:

```
df.columns
```

Out[18]:

```
Index(['id', 'diagnosis', 'radius_mean', 'texture_mean', 'perimeter_mean',
       'area_mean', 'smoothness_mean', 'compactness_mean', 'concavity_mea
n',
       'concave points_mean', 'symmetry_mean', 'fractal_dimension_mean',
       'radius_se', 'texture_se', 'perimeter_se', 'area_se', 'smoothness_s
e',
       'compactness_se', 'concavity_se', 'concave points_se', 'symmetry_s
e',
       'fractal_dimension_se', 'radius_worst', 'texture_worst',
       'perimeter_worst', 'area_worst', 'smoothness_worst',
       'compactness_worst', 'concavity_worst', 'concave points_worst',
       'symmetry_worst', 'fractal_dimension_worst', 'Unnamed: 32'],
      dtype='object')
```

In [19]:

```
df.isna().sum()
```

Out[19]:

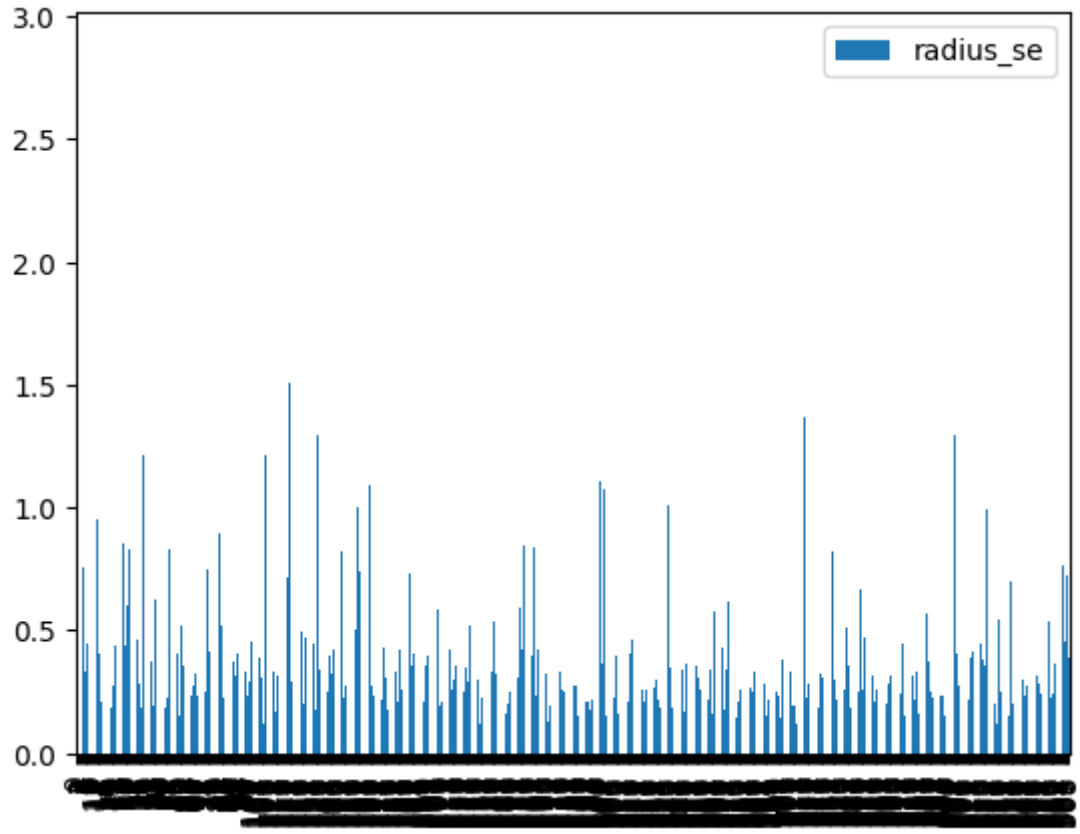
```
id                           0
diagnosis                    0
radius_mean                  0
texture_mean                 0
perimeter_mean               0
area_mean                    0
smoothness_mean              0
compactness_mean             0
concavity_mean               0
concave points_mean          0
symmetry_mean                0
fractal_dimension_mean       0
radius_se                    0
texture_se                   0
perimeter_se                 0
area_se                      0
smoothness_se                0
compactness_se               0
concavity_se                 0
concave points_se            0
symmetry_se                  0
fractal_dimension_se         0
radius_worst                 0
texture_worst                0
perimeter_worst              0
area_worst                   0
smoothness_worst             0
compactness_worst            0
concavity_worst              0
concave points_worst         0
symmetry_worst               0
fractal_dimension_worst      0
Unnamed: 32                569
dtype: int64
```

In [20]:

```
dat=df[["diagnosis","radius_se"]]
dat.plot.bar()
```

Out[20]:

<Axes: >

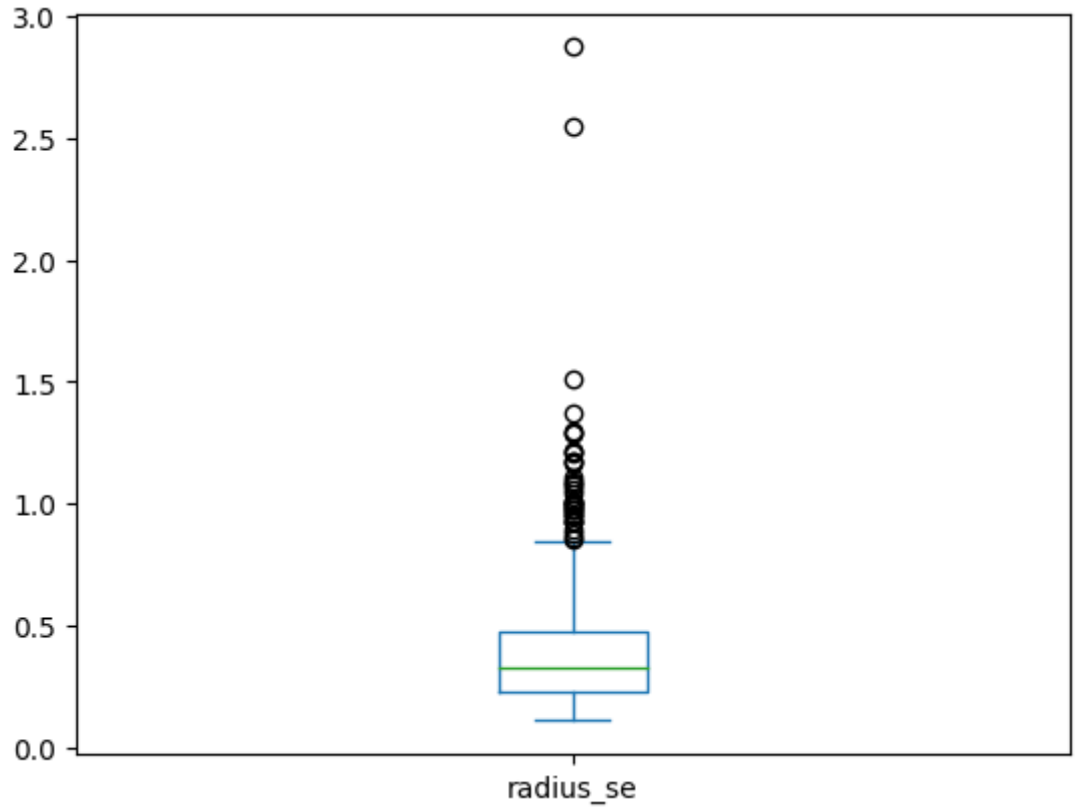


In [21]:

```
dat.plot.box()
```

Out[21]:

<Axes: >

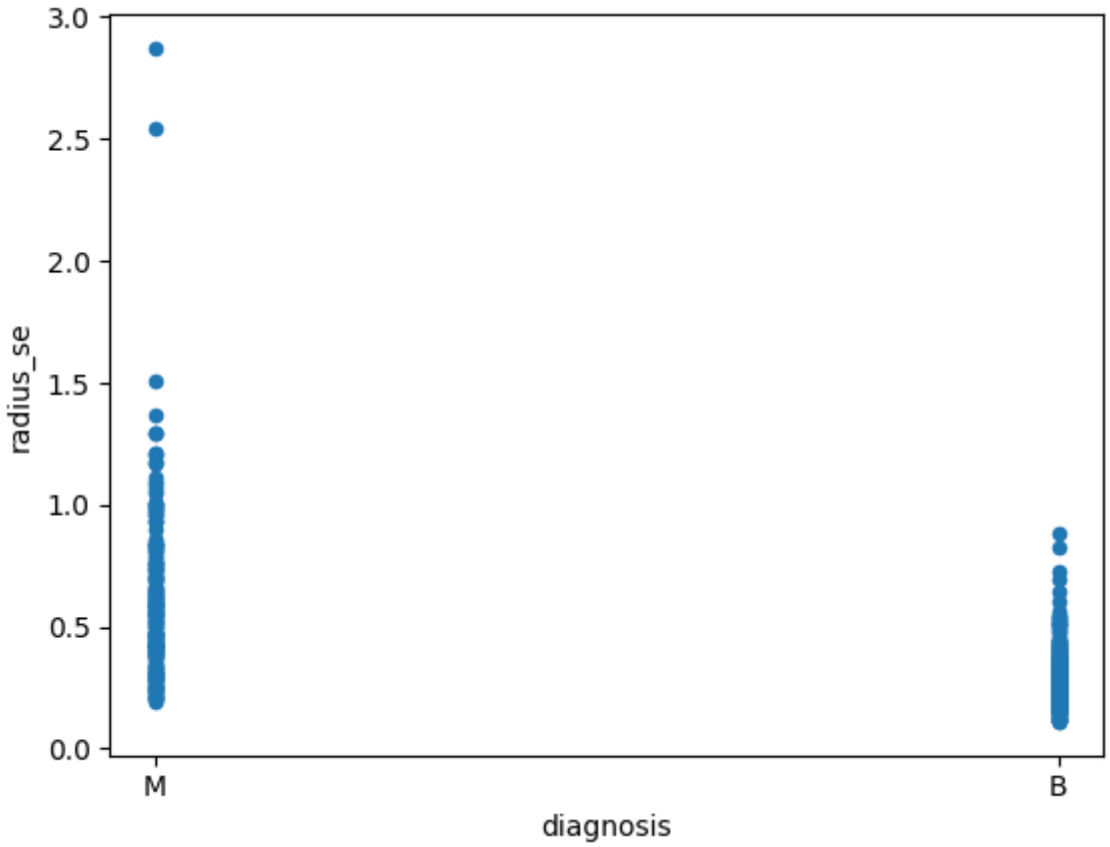


In [22]:

```
dat.plot.scatter("diagnosis","radius_se")
```

Out[22]:

```
<Axes: xlabel='diagnosis', ylabel='radius_se'>
```



** The malignant tumor has the highest radius_se values than benign

Bottle Dataset

In [23]:

```
df2=pd.read_csv("C:/Users/Gokul Jana/Downloads/9_bottle.csv")
df2
```

```
C:\Users\Gokul Jana\AppData\Local\Temp\ipykernel_8\3343413708.py:1: DtypeW
arning: Columns (47,73) have mixed types. Specify dtype option on import o
r set low_memory=False.
  df2=pd.read_csv("C:/Users/Gokul Jana/Downloads/9_bottle.csv")
```

Out[23]:

	Cst_Cnt	Btl_Cnt	Sta_ID	Depth_ID	Depthm	T_degC	Salnty	O2ml_L	STheta
0	1	1	054.0 056.0	19-4903CR-HY-060-0930-05400560-0000A-3	0	10.500	33.4400	NaN	25.64900
1	1	2	054.0 056.0	19-4903CR-HY-060-0930-05400560-0008A-3	8	10.460	33.4400	NaN	25.65600
2	1	3	054.0 056.0	19-4903CR-HY-060-0930-05400560-0010A-7	10	10.460	33.4370	NaN	25.65400
3	1	4	054.0 056.0	19-4903CR-HY-060-0930-05400560-0019A-3	19	10.450	33.4200	NaN	25.64300
4	1	5	054.0 056.0	19-4903CR-HY-060-0930-05400560-0020A-7	20	10.450	33.4210	NaN	25.64300
...	...	...	...	...	...	...	...	...	...
864858	34404	864859	093.4 026.4	20-1611SR-MX-310-2239-09340264-0000A-7	0	18.744	33.4083	5.805	23.87055
864859	34404	864860	093.4 026.4	20-1611SR-MX-310-2239-09340264-0002A-3	2	18.744	33.4083	5.805	23.87072
864860	34404	864861	093.4 026.4	20-1611SR-MX-310-2239-09340264-0005A-3	5	18.692	33.4150	5.796	23.88911
864861	34404	864862	093.4 026.4	20-1611SR-MX-310-2239-09340264-0010A-3	10	18.461	33.4062	5.816	24.01426

In [24]:

```
df2.isna().sum()
```

Out[24]:

```
Cst_Cnt          0
Btl_Cnt          0
Sta_ID           0
Depth_ID         0
Depthm           0
                ...
TA1         862779
TA2         864629
pH2         864853
pH1         864779
DIC Quality Comment    864808
Length: 74, dtype: int64
```

In [25]:

```
df2.columns
```

Out[25]:

```
Index(['Cst_Cnt', 'Btl_Cnt', 'Sta_ID', 'Depth_ID', 'Depthm', 'T_degC',
       'Salnty', 'O2ml_L', 'STheta', 'O2Sat', 'Oxy_µmol/Kg', 'BtlNum',
       'RecInd', 'T_prec', 'T_qual', 'S_prec', 'S_qual', 'P_qual', 'O_qua
l',
       'SThtaq', 'O2Satq', 'ChlorA', 'Chlqua', 'Phaeop', 'Phaqua', 'PO4u
M',
       'PO4q', 'SiO3uM', 'SiO3qu', 'NO2uM', 'NO2q', 'NO3uM', 'NO3q', 'NH3u
M',
       'NH3q', 'C14As1', 'C14A1p', 'C14A1q', 'C14As2', 'C14A2p', 'C14A2q',
       'DarkAs', 'DarkAp', 'DarkAq', 'MeanAs', 'MeanAp', 'MeanAq', 'IncTi
m',
       'LightP', 'R_Depth', 'R_TEMP', 'R_POTEMP', 'R_SALINITY', 'R_SIGMA',
       'R_SVA', 'R_DYNHT', 'R_O2', 'R_O2Sat', 'R_SIO3', 'R_PO4', 'R_NO3',
       'R_NO2', 'R_NH4', 'R_CHLA', 'R_PHAEO', 'R_PRES', 'R_SAMP', 'DIC1',
       'DIC2', 'TA1', 'TA2', 'pH2', 'pH1', 'DIC Quality Comment'],
      dtype='object')
```

In [26]:

```
df2=df2.drop(df2.iloc[:,60:].columns,axis=1)
```

	Cst_Cnt	Btl_Cnt	Sta_ID	Depth_ID	Depthm	T_degC	Salnty	O2ml_L	STheta
864862	34404	864863	093.4 026.4	1611SR-MX-310-2239-09340264-0015A-3	15	17.533	33.3880	5.774	24.15297

In [27]:

```
per=(df2.isna().sum()/len(df2)*100)
pr=pd.DataFrame(per,columns=['values'])
per1=pr[pr['values']>=80].index
per1
```

Out[27]:

```
Index(['BtlNum', 'T_qual', 'S_qual', 'SThtaq', 'NH3uM', 'C14As1', 'C14A1
p',
       'C14As2', 'C14A2p', 'DarkAs', 'DarkAp', 'MeanAs', 'MeanAp', 'IncTi
m',
       'LightP'],
      dtype='object')
```

In [28]:

```
df2=df2.drop(['BtlNum', 'T_qual', 'S_qual', 'SThtaq', 'NH3uM', 'C14As1', 'C14A1p',
       'C14As2', 'C14A2p', 'DarkAs', 'DarkAp', 'MeanAs', 'MeanAp', 'IncTim',
       'LightP'],axis=1)
df2.isna().sum()/len(df2)*100
```

Out[28]:

```
Cst_Cnt         0.000000
Btl_Cnt         0.000000
Sta_ID          0.000000
Depth_ID        0.000000
Depthm          0.000000
T_degC          1.267600
Salnty          5.475318
O2ml_L         19.501586
STheta          6.092179
O2Sat          23.540029
Oxy_µmol/Kg    23.540723
RecInd          0.000000
T_prec          1.267600
S_prec          5.475318
P_qual         22.096910
O_qual         78.646791
O2Satq         74.817168
ChlorA         73.952869
Chlqua         26.096272
Phaeop         73.952984
Phaqua         26.095809
PO4uM          52.210119
PO4q           47.762131
SiO3uM         59.058140
SiO3qu         40.930991
NO2uM          60.967691
NO2q           38.779437
NO3uM          60.987694
NO3q           38.726365
NH3q            6.540227
C14A1q          1.879835
C14A2q          1.877754
DarkAq          2.823915
MeanAq          2.824031
R_Depth         0.000000
R_TEMP          1.267600
R_POTEMP        5.324196
R_SALINITY      5.475318
R_SIGMA         6.111488
R_SVA           6.101660
R_DYNHT         5.394727
R_O2           19.501586
R_O2Sat        22.941784
R_SIO3         59.057215
R_PO4          52.209194
dtype: float64
```

In [29]:

```
df3=df2.iloc[:,:11]
df3
```

Out[29]:

	Cst_Cnt	Btl_Cnt	Sta_ID	Depth_ID	Depthm	T_degC	Salnty	O2ml_L	STheta
0	1	1	054.0 056.0	19-4903CR-HY-060-0930-05400560-0000A-3	0	10.500	33.4400	NaN	25.64900
1	1	2	054.0 056.0	19-4903CR-HY-060-0930-05400560-0008A-3	8	10.460	33.4400	NaN	25.65600
2	1	3	054.0 056.0	19-4903CR-HY-060-0930-05400560-0010A-7	10	10.460	33.4370	NaN	25.65400
3	1	4	054.0 056.0	19-4903CR-HY-060-0930-05400560-0019A-3	19	10.450	33.4200	NaN	25.64300
4	1	5	054.0 056.0	19-4903CR-HY-060-0930-05400560-0020A-7	20	10.450	33.4210	NaN	25.64300
...	...	...	...	...	...	...	...	...	...
864858	34404	864859	093.4 026.4	20-1611SR-MX-310-2239-09340264-0000A-7	0	18.744	33.4083	5.805	23.87055
864859	34404	864860	093.4 026.4	20-1611SR-MX-310-2239-09340264-0002A-3	2	18.744	33.4083	5.805	23.87072
864860	34404	864861	093.4 026.4	20-1611SR-MX-310-2239-09340264-0005A-3	5	18.692	33.4150	5.796	23.88911
864861	34404	864862	093.4 026.4	20-1611SR-MX-310-2239-09340264-0010A-3	10	18.161	33.4062	5.816	24.01426

	Cst_Cnt	Btl_Cnt	Sta_ID	Depth_ID	Depthm	T_degC	Salnty	O2ml_L	STheta
864862	34404	864863	093.4 026.4	20-1611SR-MX-310-2239-09340264-0015A-3	15	17.533	33.3880	5.774	24.15297

In [30]:

```
print(df3.columns)
df3.isna().sum()/len(df3)*100
```

864863 rows × 11 columns

```
Index(['Cst_Cnt', 'Btl_Cnt', 'Sta_ID', 'Depth_ID', 'Depthm', 'T_degC',
       'Salnty', 'O2ml_L', 'STheta', 'O2Sat', 'Oxy_µmol/Kg'],
      dtype='object')
```

Out[30]:

```
Cst_Cnt          0.000000
Btl_Cnt          0.000000
Sta_ID           0.000000
Depth_ID         0.000000
Depthm           0.000000
T_degC           1.267600
Salnty           5.475318
O2ml_L          19.501586
STheta           6.092179
O2Sat           23.540029
Oxy_µmol/Kg     23.540723
dtype: float64
```

In [31]:

```
df3=df3.fillna(df3['T_degC'].mean())
df3=df3.fillna(df3['Salnty'].mean())
df3=df3.fillna(df3['O2ml_L'].mean())
df3=df3.fillna(df3['STheta'].mean())
df3=df3.fillna(df3['Oxy_µmol/Kg'].mean())
df3=df3.fillna(df3['O2Sat'].median())

df3.isna().sum()
```

Out[31]:

```
Cst_Cnt         0
Btl_Cnt         0
Sta_ID          0
Depth_ID        0
Depthm          0
T_degC          0
Salnty          0
O2ml_L          0
STheta          0
O2Sat           0
Oxy_µmol/Kg     0
dtype: int64
```

In [32]:

```
df3
```

Out[32]:

	Cst_Cnt	Btl_Cnt	Sta_ID	Depth_ID	Depthm	T_degC	Salnty	O2ml_L	STheta
0	1	1	054.0 056.0	19-4903CR-HY-060-0930-05400560-0000A-3	0	10.500	33.4400	10.799677	25.64900
1	1	2	054.0 056.0	19-4903CR-HY-060-0930-05400560-0008A-3	8	10.460	33.4400	10.799677	25.65600
2	1	3	054.0 056.0	19-4903CR-HY-060-0930-05400560-0010A-7	10	10.460	33.4370	10.799677	25.65400
3	1	4	054.0 056.0	19-4903CR-HY-060-0930-05400560-0019A-3	19	10.450	33.4200	10.799677	25.64300
4	1	5	054.0 056.0	19-4903CR-HY-060-0930-05400560-0020A-7	20	10.450	33.4210	10.799677	25.64300
...	...	...	...	...	...	...	...	...	...
864858	34404	864859	093.4 026.4	20-1611SR-MX-310-2239-09340264-0000A-7	0	18.744	33.4083	5.805000	23.87055
864859	34404	864860	093.4 026.4	20-1611SR-MX-310-2239-09340264-0002A-3	2	18.744	33.4083	5.805000	23.87072
864860	34404	864861	093.4 026.4	20-1611SR-MX-310-2239-09340264-0005A-3	5	18.692	33.4150	5.796000	23.88911
864861	34404	864862	093.4 026.4	20-1611SR-MX-310-2239-09340264-0010A-3	10	18.161	33.4062	5.816000	24.01426

	Cst_Cnt	Btl_Cnt	Sta_ID	Depth_ID	Depthm	T_degC	Salnty	O2ml_L	STheta
864862	34404	864863	093.4 026.4	20-1611SR-MX-310-2239-09340264-0015A-3	15	17.533	33.3880	5.774000	24.15297

In [33]:

```
ds=df3[["O2Sat","Oxy_µmol/Kg"]]
ds=ds.iloc[:4000]
```

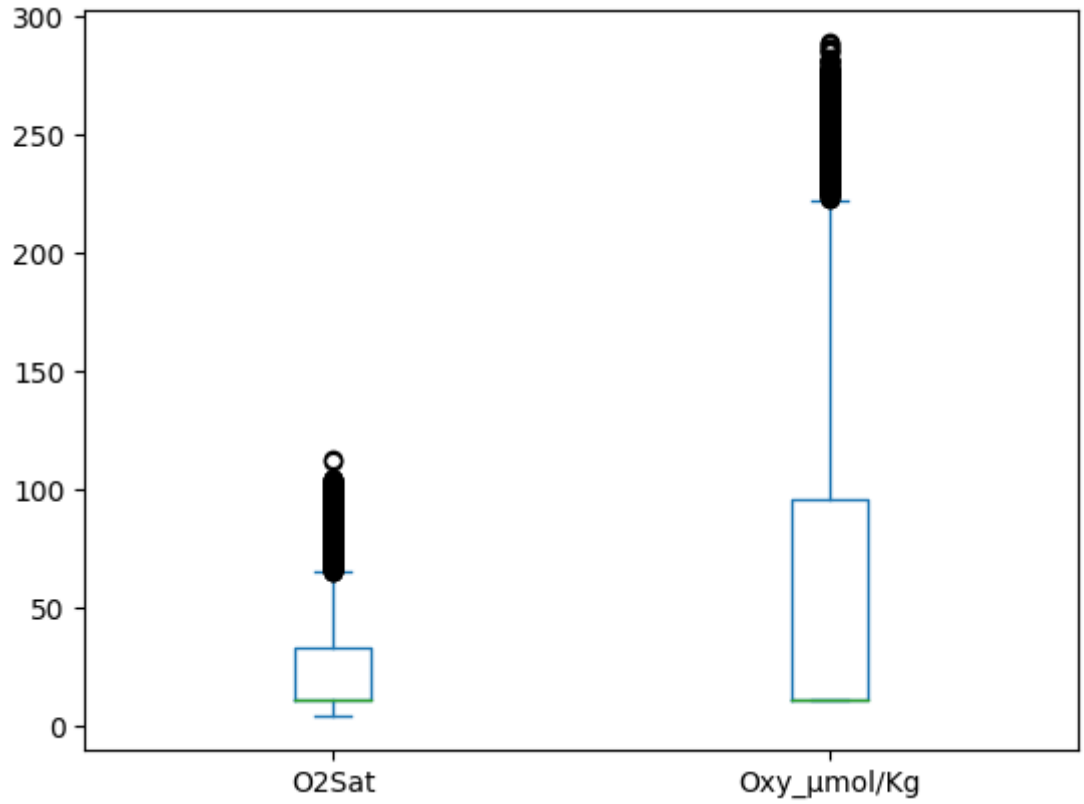
864863 rows × 11 columns

In [34]:

```
ds.plot.box()
```

Out[34]:

<Axes: >

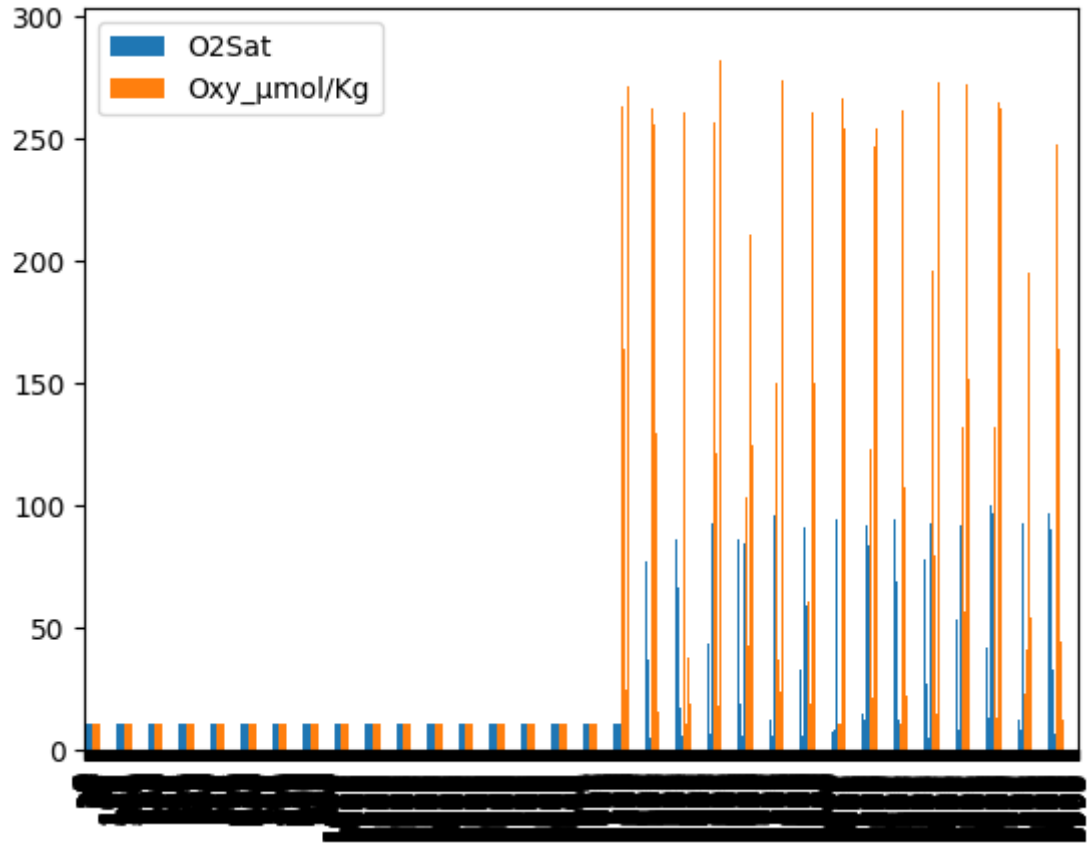


In [35]:

```
ds.plot.bar()
```

Out[35]:

<Axes: >



In [36]:

```
ds.plot.line()
```

Out[36]:

```
<Axes: >
```