Data set 9

In [1]:

```
import numpy as np
import pandas as pd
```

In [2]:

```
a=pd.read_csv(r"C:\Users\user\Downloads\9_bottle.csv")
```

```
C:\ProgramData\Anaconda3\lib\site-packages\IPython\core\interactiveshell.p
y:3165: DtypeWarning: Columns (47,73) have mixed types.Specify dtype optio
n on import or set low_memory=False.
  has_raised = await self.run_ast_nodes(code_ast.body, cell_name,
```

To print top rows:

In [23]:

```
b=a.head(1000)
b
```

Out[23]:

	Cst_Cnt	Btl_Cnt	Sta_ID	Depth_ID	Depthm	T_degC	Salnty	O2ml_L	STheta	O2Sat
0	1	1	054.0 056.0	19-4903CR-HY-060-0930-05400560-0000A-3	0	10.50	33.440	NaN	25.649	NaN
1	1	2	054.0 056.0	19-4903CR-HY-060-0930-05400560-0008A-3	8	10.46	33.440	NaN	25.656	NaN
2	1	3	054.0 056.0	19-4903CR-HY-060-0930-05400560-0010A-7	10	10.46	33.437	NaN	25.654	NaN
3	1	4	054.0 056.0	19-4903CR-HY-060-0930-05400560-0019A-3	19	10.45	33.420	NaN	25.643	NaN
4	1	5	054.0 056.0	19-4903CR-HY-060-0930-05400560-0020A-7	20	10.45	33.421	NaN	25.643	NaN
...	...	...	...	...	...	...	...	...	...	...
995	33	996	092.0 088.0	19-4903NS-HY-061-0906-09200880-0300A-7	300	7.22	34.040	NaN	26.636	NaN
996	33	997	092.0 088.0	19-4903NS-HY-061-0906-09200880-0379A-3	379	6.58	34.040	NaN	26.724	NaN
997	33	998	092.0 088.0	19-4903NS-HY-061-0906-09200880-0400A-7	400	6.44	34.049	NaN	26.750	NaN
998	33	999	092.0 088.0	19-4903NS-HY-061-0906-09200880-0500A-7	500	5.85	34.113	NaN	26.876	NaN

	Cst_Cnt	Btl_Cnt	Sta_ID	Depth_ID	Depthm	T_degC	Salnty	O2ml_L	STheta	O2Sat
999	33	1000	092.0 088.0	19-4903NS-HY-061-0906-09200880-0552A-3	552	5.60	34.160	NaN	26.944	NaN

To print Last rows:

In [4]:
1000 rows × 74 columns

```
a.tail()
```

Out[4]:

	Cst_Cnt	Btl_Cnt	Sta_ID	Depth_ID	Depthm	T_degC	Salnty	O2ml_L	STheta	(
864858	34404	864859	093.4 026.4	20-1611SR-MX-310-2239-09340264-0000A-7	0	18.744	33.4083	5.805	23.87055	1
864859	34404	864860	093.4 026.4	20-1611SR-MX-310-2239-09340264-0002A-3	2	18.744	33.4083	5.805	23.87072	1
864860	34404	864861	093.4 026.4	20-1611SR-MX-310-2239-09340264-0005A-3	5	18.692	33.4150	5.796	23.88911	1
864861	34404	864862	093.4 026.4	20-1611SR-MX-310-2239-09340264-0010A-3	10	18.161	33.4062	5.816	24.01426	1
864862	34404	864863	093.4 026.4	20-1611SR-MX-310-2239-09340264-0015A-3	15	17.533	33.3880	5.774	24.15297	1

5 rows × 74 columns

Statistical Summary:

In [6]:

```
a.describe()
```

Out[6]:

	Cst_Cnt	Btl_Cnt	Depthm	T_degC	Salnty	O2
count	864863.000000	864863.000000	864863.000000	853900.000000	817509.000000	696201.0(
mean	17138.790958	432432.000000	226.831951	10.799677	33.840350	3.3(
std	10240.949817	249664.587267	316.050259	4.243825	0.461843	2.0(
min	1.000000	1.000000	0.000000	1.440000	28.431000	-0.0(
25%	8269.000000	216216.500000	46.000000	7.680000	33.488000	1.3(
50%	16848.000000	432432.000000	125.000000	10.060000	33.863000	3.4(
75%	26557.000000	648647.500000	300.000000	13.880000	34.196900	5.5(
max	34404.000000	864863.000000	5351.000000	31.140000	37.034000	11.1(

8 rows × 70 columns

To print no of rows and columns

In [7]:

```
a.shape
```

Out[7]:

```
(864863, 74)
```

To print no of elements

In [8]:

```
a.size
```

Out[8]:

```
63999862
```

Missing no of values

In [10]:

```
a.isna()
```

Out[10]:

	Cst_Cnt	Btl_Cnt	Sta_ID	Depth_ID	Depthm	T_degC	Salnty	O2ml_L	STheta	O2Sa
0	False	False	False	False	False	False	False	True	False	Tru
1	False	False	False	False	False	False	False	True	False	Tru
2	False	False	False	False	False	False	False	True	False	Tru
3	False	False	False	False	False	False	False	True	False	Tru
4	False	False	False	False	False	False	False	True	False	Tru
...	...	...	...	...	...	...	...	...	...	.
864858	False	False	False	False	False	False	False	False	False	Fals
864859	False	False	False	False	False	False	False	False	False	Fals
864860	False	False	False	False	False	False	False	False	False	Fals
864861	False	False	False	False	False	False	False	False	False	Fals
864862	False	False	False	False	False	False	False	False	False	Fals

864863 rows × 74 columns

In [11]:

```
a.dropna(axis=1)
```

Out[11]:

	Cst_Cnt	Btl_Cnt	Sta_ID	Depth_ID	Depthm	RecInd	R_Depth	R_PRES
0	1	1	054.0 056.0	19-4903CR-HY-060-0930-05400560-0000A-3	0	3	0.0	0
1	1	2	054.0 056.0	19-4903CR-HY-060-0930-05400560-0008A-3	8	3	8.0	8
2	1	3	054.0 056.0	19-4903CR-HY-060-0930-05400560-0010A-7	10	7	10.0	10
3	1	4	054.0 056.0	19-4903CR-HY-060-0930-05400560-0019A-3	19	3	19.0	19
4	1	5	054.0 056.0	19-4903CR-HY-060-0930-05400560-0020A-7	20	7	20.0	20
...	...	...	...	...	...	...	...	...
864858	34404	864859	093.4 026.4	20-1611SR-MX-310-2239-09340264-0000A-7	0	7	0.0	0
864859	34404	864860	093.4 026.4	20-1611SR-MX-310-2239-09340264-0002A-3	2	3	2.0	2
864860	34404	864861	093.4 026.4	20-1611SR-MX-310-2239-09340264-0005A-3	5	3	5.0	5
864861	34404	864862	093.4 026.4	20-1611SR-MX-310-2239-09340264-0010A-3	10	3	10.0	10
864862	34404	864863	093.4 026.4	20-1611SR-MX-310-2239-09340264-0015A-3	15	3	15.0	15

864863 rows × 8 columns

In [12]:

```
import matplotlib.pyplot as pp
```

In [15]:

```
c=a[['Cst_Cnt','Btl_Cnt']]
c
```

Out[15]:

	Cst_Cnt	Btl_Cnt
0	1	1
1	1	2
2	1	3
3	1	4
4	1	5
...	...	...
864858	34404	864859
864859	34404	864860
864860	34404	864861
864861	34404	864862
864862	34404	864863

864863 rows × 2 columns

In [16]:

```
b.plot.line()
```

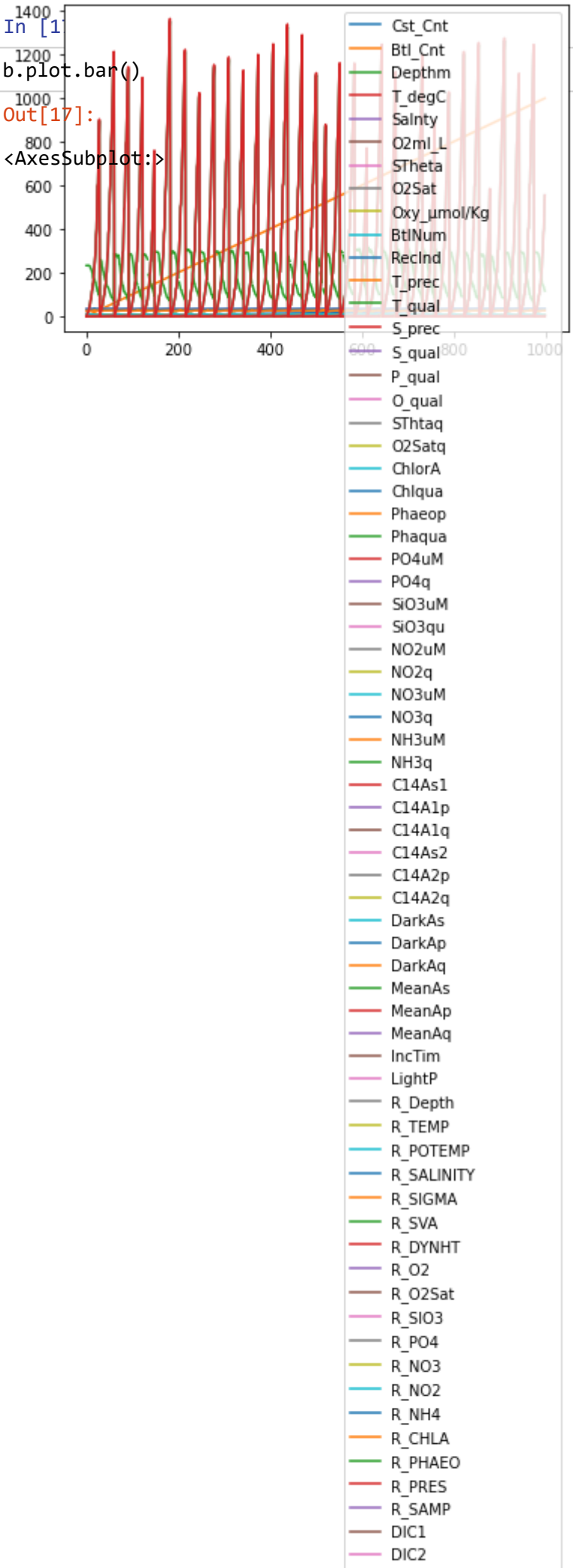
Out[16]:

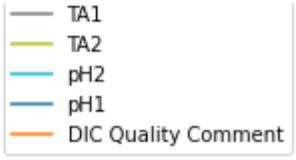
```
<AxesSubplot:>
```

In [17]:
```
b.plot.bar()
```

Out[17]:

<AxesSubplot:>



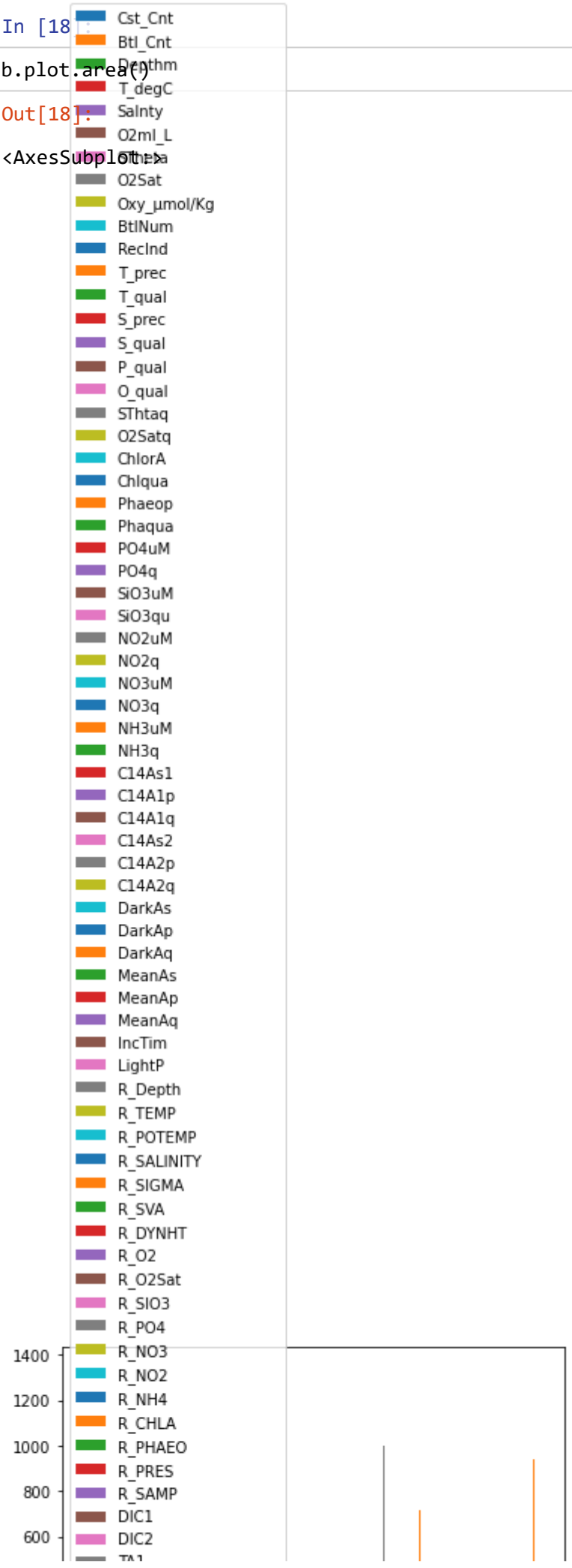

TA1
TA2
pH2
pH1
DIC Quality Comment

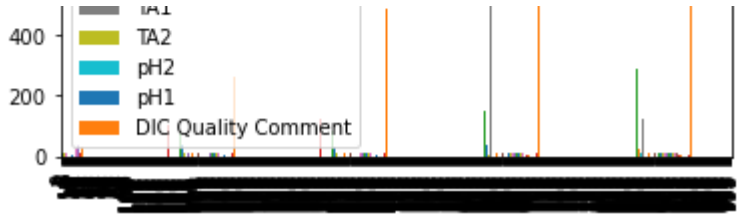
In [18]:

```
b.plot.area()
```

Out[18]:

```
<AxesSubplot:>
```



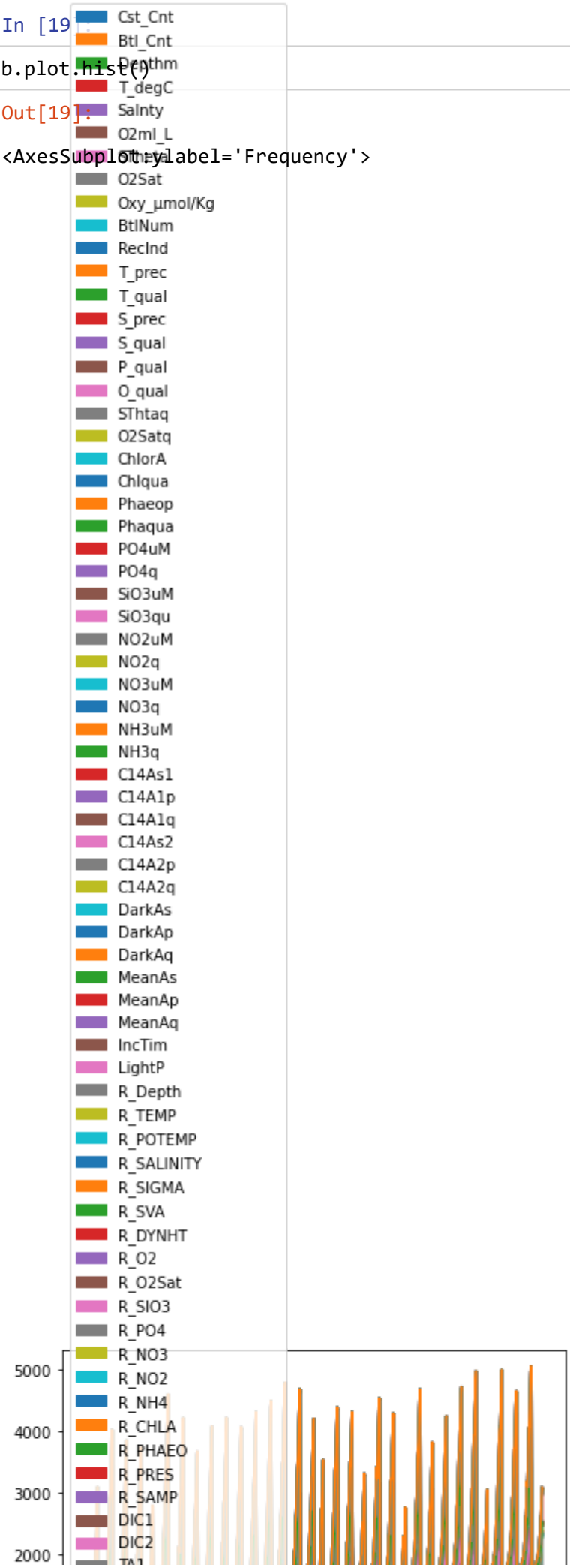

400
200
0
TA2
pH2
pH1
DIC Quality Comment

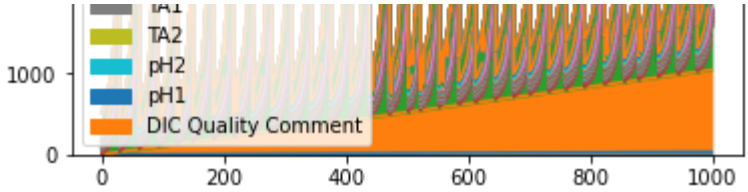
In [19]:

```
b.plot.hist()
```

Out[19]:

```
<AxesSubplot:ylabel='Frequency'>
```



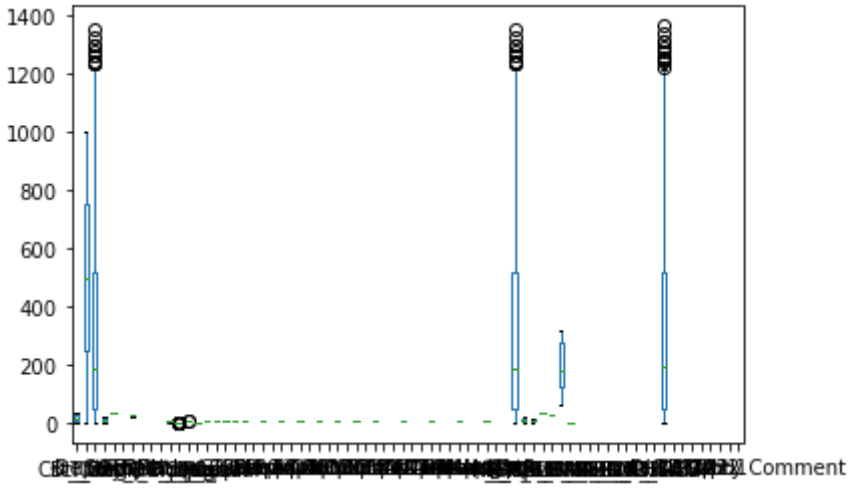


In [20]:

```
b.plot.box()
```

Out[20]:

```
<AxesSubplot:>
```

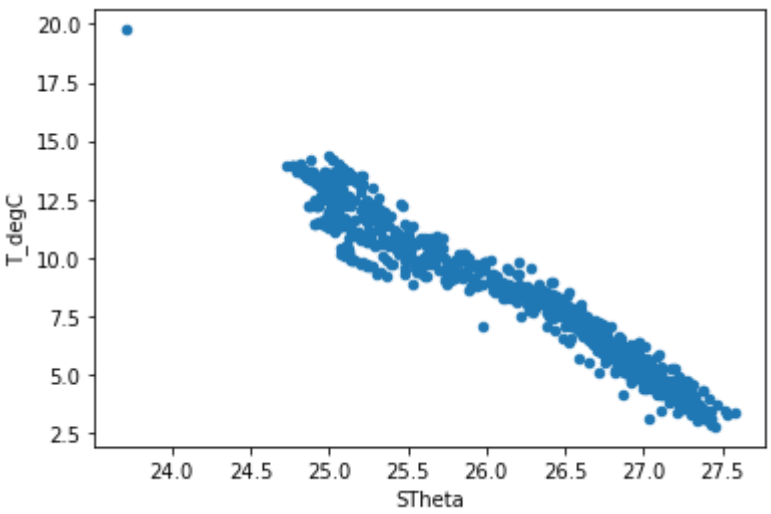


In [21]:

```
b.plot.scatter(x='STheta',y='T_degC')
```

Out[21]:

```
<AxesSubplot:xlabel='STheta', ylabel='T_degC'>
```

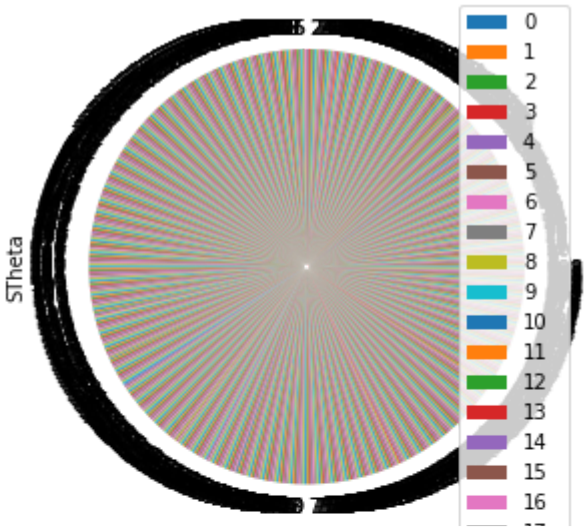


In [22]:

```
b.plot.pie(y='STheta',figsize=(5,5))
```

Out[22]:

```
<AxesSubplot:ylabel='STheta'>
```



In [ ]: