

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

```
In [2]: %matplotlib inline
```

```
In [18]: df = pd.read_csv("C:\\Users\\Shay\\Documents\\Yari\\Data analysis Python\\Data\\Original 1
```

```
In [4]: #XYP=pd.read_csv("C:\\Users\\Shay\\Documents\\Yari\\Python\\XYPlot.csv")
```

```
In [5]: #XYP
```

Out[5]:

	BAND	MCW_nm
0	0	58.733654
1	1	55.678649
2	2	57.542610

```
In [5]: df
```

Out[5]:

	HDDSN	ENDDATE	PROCID	DRIVEMODEL	PFCODE	TESTCODE	TESTC	
0	2FA056NA20210117124423	2FA056NA	2.021010e+13	6400	2F	0	PDQC3E8M	PDQ
1	2FA056NA20210117124423	2FA056NA	2.021010e+13	6400	2F	0	PDQC3E8M	PDQ
2	2FA056NA20210117124423	2FA056NA	2.021010e+13	6400	2F	0	PDQC3E8M	PDQ
3	2FA056NA20210117124423	2FA056NA	2.021010e+13	6400	2F	0	PDQC3E8M	PDQ
4	2FA056NA20210117124423	2FA056NA	2.021010e+13	6400	2F	0	PDQC3E8M	PDQ
...
3115	2FA059LA20210117124210	2FA059LA	2.021010e+13	6400	2F	0	PDQC3E8M	PDQ
3116	2FA059LA20210117124210	2FA059LA	2.021010e+13	6400	2F	0	PDQC3E8M	PDQ
3117	2FA059LA20210117124210	2FA059LA	2.021010e+13	6400	2F	0	PDQC3E8M	PDQ
3118	2FA059LA20210117124210	2FA059LA	2.021010e+13	6400	2F	0	PDQC3E8M	PDQ
3119	2FA059LA20210117124210	2FA059LA	2.021010e+13	6400	2F	0	PDQC3E8M	PDQ

3120 rows × 78 columns

```
In [19]: df = df[["HDDSN", "PROCID", "TESTCODEC", "PFCODE", "MFGID", "HDDTRIAL", "QUALIFIER", "LHD", "PHD",
```

```
In [8]: df
```

Out[8]:

	HDDSN	PROCID	TESTCODEC	PFCODE	MFGID	HDDTRIAL	QUALIFIER	LHD	PHD	BAND	DataTP
0	2FA0491A	6400	PDQC3CX	2525	KQBT03	AQ0N	10N0	0	0	0	1188 1
1	2FA0491A	6400	PDQC3CX	2525	KQBT03	AQ0N	10N0	0	0	1	1188 1
2	2FA0491A	6400	PDQC3CX	2525	KQBT03	AQ0N	10N0	0	0	2	1188 1
3	2FA0491A	6400	PDQC3CX	2525	KQBT03	AQ0N	10N0	1	1	0	1188 1
4	2FA0491A	6400	PDQC3CX	2525	KQBT03	AQ0N	10N0	1	1	1	1188 1
...
		6400	PDQC3CX	2521	KQBT03	AQ0N	2300	7	17	2	986

	HDDSN	PROCID	TESTCODEC	PFCODE	MFGID	HDDTRIAL	QUALIFIER	LHD	PHD	BAND	DataTP
956	2FA049GA	6400	PDQC3CX	2521	KQBT03	AQ0N	2300	7	17	3	987
957	2FA049GA	6400	PDQC3CX	2521	KQBT03	AQ0N	2300	7	17	4	986
958	2FA049GA	6400	PDQC3CX	2521	KQBT03	AQ0N	2300	7	17	5	986
959	2FA049GA	6400	PDQC3CX	2521	KQBT03	AQ0N	2300	7	17	6	987

960 rows × 12 columns

```
In [20]: df['MCW_nm'] = (df['MCW']/100)*(1000/580)*(25.4)*(df['DataTP']/1024.0)
```

```
In [21]: df
```

	HDDSN	PROCID	TESTCODEC	PFCODE	MFGID	HDDTRIAL	QUALIFIER	LHD	PHD	BAND	DataTP
0	2FA056NA	6400	PDQC3E8	0	KKBT01	MW11	10N0	0	2	0	1188
1	2FA056NA	6400	PDQC3E8	0	KKBT01	MW11	10N0	0	2	1	1188
2	2FA056NA	6400	PDQC3E8	0	KKBT01	MW11	10N0	0	2	2	1188
3	2FA056NA	6400	PDQC3E8	0	KKBT01	MW11	10N0	1	3	0	1188
4	2FA056NA	6400	PDQC3E8	0	KKBT01	MW11	10N0	1	3	1	1188
...
3115	2FA059LA	6400	PDQC3E8	0	KKBT01	MW13	X010	6	14	1	1255
3116	2FA059LA	6400	PDQC3E8	0	KKBT01	MW13	X010	6	14	2	1310
3117	2FA059LA	6400	PDQC3E8	0	KKBT01	MW13	X010	7	15	0	1139
3118	2FA059LA	6400	PDQC3E8	0	KKBT01	MW13	X010	7	15	1	1144
3119	2FA059LA	6400	PDQC3E8	0	KKBT01	MW13	X010	7	15	2	1149

3120 rows × 13 columns

```
In [22]: df.drop(['MCW', 'DataTP'], axis = 1, inplace=True)
```

```
In [23]: df
```

	HDDSN	PROCID	TESTCODEC	PFCODE	MFGID	HDDTRIAL	QUALIFIER	LHD	PHD	BAND	MCW_nm
0	2FA056NA	6400	PDQC3E8	0	KKBT01	MW11	10N0	0	2	0	62.863545
1	2FA056NA	6400	PDQC3E8	0	KKBT01	MW11	10N0	0	2	1	59.005896
2	2FA056NA	6400	PDQC3E8	0	KKBT01	MW11	10N0	0	2	2	64.476066
3	2FA056NA	6400	PDQC3E8	0	KKBT01	MW11	10N0	1	3	0	61.685163
4	2FA056NA	6400	PDQC3E8	0	KKBT01	MW11	10N0	1	3	1	58.819836
...
3115	2FA059LA	6400	PDQC3E8	0	KKBT01	MW13	X010	6	14	1	57.695007
3116	2FA059LA	6400	PDQC3E8	0	KKBT01	MW13	X010	6	14	2	59.676360
3117	2FA059LA	6400	PDQC3E8	0	KKBT01	MW13	X010	7	15	0	53.135247
3118	2FA059LA	6400	PDQC3E8	0	KKBT01	MW13	X010	7	15	1	53.105720
3119	2FA059LA	6400	PDQC3E8	0	KKBT01	MW13	X010	7	15	2	54.573497

```
In [13]: df.to_csv("C:\\Users\\Shay\\Documents\\Yari\\Data analysis Python\\Data\\CCB_CI_MCW (1-28",
```

```
In [24]: def RADIUS (row):  
    if row['BAND'] == 0 :  
        return 'OD'  
    if row['BAND'] == 1:  
        return 'MD'  
    if row['BAND'] == 2 :  
        return 'ID'
```

```
In [25]: df['RADIUS'] = df.apply (lambda row: RADIUS(row), axis=1)
```

```
In [26]: df
```

Out[26]:

	HDDSN	PROCID	TESTCODEC	PFCODE	MFGID	HDDTRIAL	QUALIFIER	LHD	PHD	BAND	MCW_nm
0	2FA056NA	6400	PDQC3E8	0	KKBT01	MW11	10N0	0	2	0	62.863545
1	2FA056NA	6400	PDQC3E8	0	KKBT01	MW11	10N0	0	2	1	59.005896
2	2FA056NA	6400	PDQC3E8	0	KKBT01	MW11	10N0	0	2	2	64.476066
3	2FA056NA	6400	PDQC3E8	0	KKBT01	MW11	10N0	1	3	0	61.685163
4	2FA056NA	6400	PDQC3E8	0	KKBT01	MW11	10N0	1	3	1	58.819836
...
3115	2FA059LA	6400	PDQC3E8	0	KKBT01	MW13	X010	6	14	1	57.695007
3116	2FA059LA	6400	PDQC3E8	0	KKBT01	MW13	X010	6	14	2	59.676360
3117	2FA059LA	6400	PDQC3E8	0	KKBT01	MW13	X010	7	15	0	53.135247
3118	2FA059LA	6400	PDQC3E8	0	KKBT01	MW13	X010	7	15	1	53.105720
3119	2FA059LA	6400	PDQC3E8	0	KKBT01	MW13	X010	7	15	2	54.573497

3120 rows × 12 columns

```
In [27]: df.to_csv("C:\\Users\\Shay\\Documents\\Yari\\Data analysis Python\\Data\\CCB_CI_MCW (1-28",
```

```
In [28]: pd.value_counts(df['QUALIFIER'])
```

Out[28]:

X000	456
10N0	456
C0N0	456
A0N0	456
X080	432
X070	432
X010	432

Name: QUALIFIER, dtype: int64

```
In [11]: new_table = df[df['MCW'] > 0]
```

```
In [12]: new_table
```

Out[12]:

	HDDSN	PROCID	TESTCODEC	PFCODE	MFGID	HDDTRIAL	QUALIFIER	LHD	PHD	BAND	DataTP
0	2FA0491A	6400	PDQC3CX	2525	KQBT03	AQ0N	10N0	0	0	0	1188 1
1	2FA0491A	6400	PDQC3CX	2525	KQBT03	AQ0N	10N0	0	0	1	1188 1
2	2FA0491A	6400	PDQC3CX	2525	KQBT03	AQ0N	10N0	0	0	2	1188 1
		6400	PDQC3CX	2525	KQBT03	AQ0N	10N0	1	1	0	1188 1

	HDDSN	PROCID	TESTCODEC	PFCODE	MFGID	HDDTRIAL	QUALIFIER	LHD	PHD	BAND	DataTP	
4	2FA0491A	6400	PDQC3CX	2525	KQBT03	AQ0N	10N0	1	1	1	1188	1
...
899	2FA049GA	6400	PDQC3CX	2521	KQBT03	AQ0N	10N0	6	16	1	1188	1
900	2FA049GA	6400	PDQC3CX	2521	KQBT03	AQ0N	10N0	6	16	2	1188	1
901	2FA049GA	6400	PDQC3CX	2521	KQBT03	AQ0N	10N0	7	17	0	1188	1
902	2FA049GA	6400	PDQC3CX	2521	KQBT03	AQ0N	10N0	7	17	1	1188	1
903	2FA049GA	6400	PDQC3CX	2521	KQBT03	AQ0N	10N0	7	17	2	1188	1

288 rows × 13 columns

```
In [13]: new_table.to_csv('C:\\Users\\Shay\\Documents\\Yari\\Python\\new.csv')
```

```
In [14]: BAND0 = new_table[new_table['BAND'] == 0]
```

```
In [15]: BAND0
```

```
Out[15]:
```

	HDDSN	PROCID	TESTCODEC	PFCODE	MFGID	HDDTRIAL	QUALIFIER	LHD	PHD	BAND	DataTP	
0	2FA0491A	6400	PDQC3CX	2525	KQBT03	AQ0N	10N0	0	0	0	1188	1
3	2FA0491A	6400	PDQC3CX	2525	KQBT03	AQ0N	10N0	1	1	0	1188	1
6	2FA0491A	6400	PDQC3CX	2525	KQBT03	AQ0N	10N0	2	8	0	1188	1
9	2FA0491A	6400	PDQC3CX	2525	KQBT03	AQ0N	10N0	3	9	0	1188	1
12	2FA0491A	6400	PDQC3CX	2525	KQBT03	AQ0N	10N0	4	10	0	1188	1
...
889	2FA049GA	6400	PDQC3CX	2521	KQBT03	AQ0N	10N0	3	9	0	1188	1
892	2FA049GA	6400	PDQC3CX	2521	KQBT03	AQ0N	10N0	4	10	0	1188	1
895	2FA049GA	6400	PDQC3CX	2521	KQBT03	AQ0N	10N0	5	11	0	1188	1
898	2FA049GA	6400	PDQC3CX	2521	KQBT03	AQ0N	10N0	6	16	0	1188	1
901	2FA049GA	6400	PDQC3CX	2521	KQBT03	AQ0N	10N0	7	17	0	1188	1

96 rows × 13 columns

```
In [16]: BAND1 = new_table[new_table['BAND'] == 1]
```

```
In [17]: BAND1
```

```
Out[17]:
```

	HDDSN	PROCID	TESTCODEC	PFCODE	MFGID	HDDTRIAL	QUALIFIER	LHD	PHD	BAND	DataTP	
1	2FA0491A	6400	PDQC3CX	2525	KQBT03	AQ0N	10N0	0	0	1	1188	1
4	2FA0491A	6400	PDQC3CX	2525	KQBT03	AQ0N	10N0	1	1	1	1188	1
7	2FA0491A	6400	PDQC3CX	2525	KQBT03	AQ0N	10N0	2	8	1	1188	1
10	2FA0491A	6400	PDQC3CX	2525	KQBT03	AQ0N	10N0	3	9	1	1188	1
13	2FA0491A	6400	PDQC3CX	2525	KQBT03	AQ0N	10N0	4	10	1	1188	1
...
890	2FA049GA	6400	PDQC3CX	2521	KQBT03	AQ0N	10N0	3	9	1	1188	1

	HDDSN	PROCID	TESTCODEC	PFCODE	MFGID	HDDTRIAL	QUALIFIER	LHD	PHD	BAND	DataTP
893	2FA049GA	6400	PDQC3CX	2521	KQBT03	AQ0N	10N0	4	10	1	1188 1
896	2FA049GA	6400	PDQC3CX	2521	KQBT03	AQ0N	10N0	5	11	1	1188 1
899	2FA049GA	6400	PDQC3CX	2521	KQBT03	AQ0N	10N0	6	16	1	1188 1
902	2FA049GA	6400	PDQC3CX	2521	KQBT03	AQ0N	10N0	7	17	1	1188 1

96 rows × 13 columns

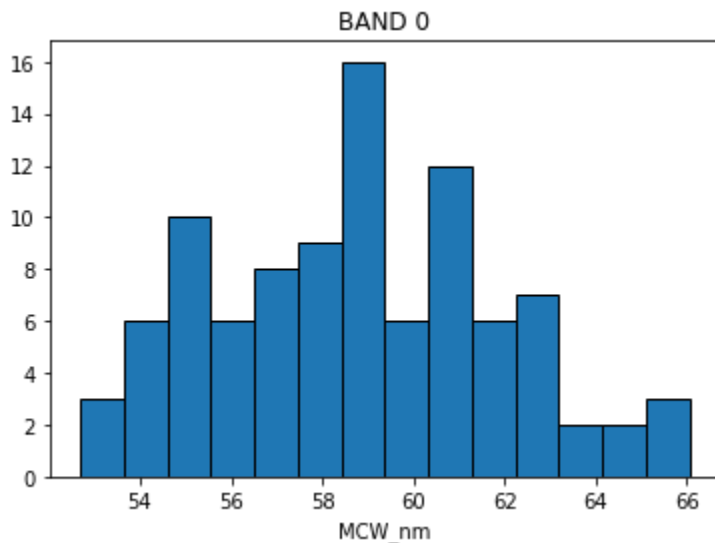
```
In [18]: BAND2 = new_table[new_table['BAND'] == 2]
```

```
In [19]: BAND2
```

	HDDSN	PROCID	TESTCODEC	PFCODE	MFGID	HDDTRIAL	QUALIFIER	LHD	PHD	BAND	DataTP
2	2FA0491A	6400	PDQC3CX	2525	KQBT03	AQ0N	10N0	0	0	2	1188 1
5	2FA0491A	6400	PDQC3CX	2525	KQBT03	AQ0N	10N0	1	1	2	1188 1
8	2FA0491A	6400	PDQC3CX	2525	KQBT03	AQ0N	10N0	2	8	2	1188 1
11	2FA0491A	6400	PDQC3CX	2525	KQBT03	AQ0N	10N0	3	9	2	1188 1
14	2FA0491A	6400	PDQC3CX	2525	KQBT03	AQ0N	10N0	4	10	2	1188 1
...
891	2FA049GA	6400	PDQC3CX	2521	KQBT03	AQ0N	10N0	3	9	2	1188 1
894	2FA049GA	6400	PDQC3CX	2521	KQBT03	AQ0N	10N0	4	10	2	1188 1
897	2FA049GA	6400	PDQC3CX	2521	KQBT03	AQ0N	10N0	5	11	2	1188 1
900	2FA049GA	6400	PDQC3CX	2521	KQBT03	AQ0N	10N0	6	16	2	1188 1
903	2FA049GA	6400	PDQC3CX	2521	KQBT03	AQ0N	10N0	7	17	2	1188 1

96 rows × 13 columns

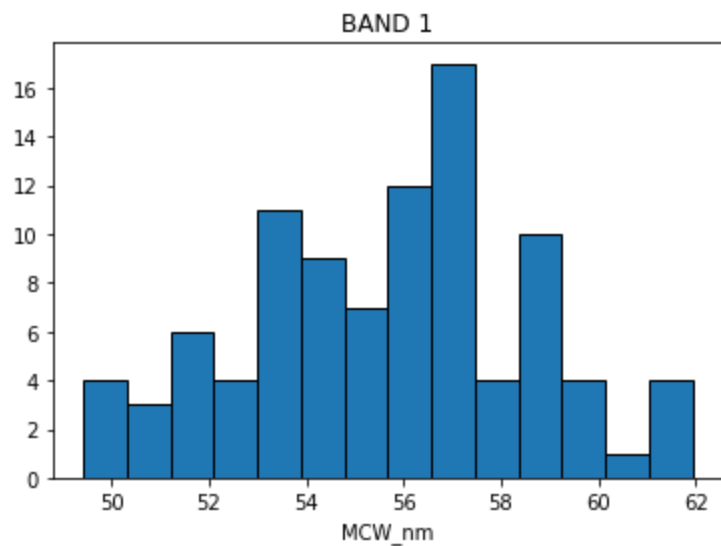
```
In [20]: plt.hist(BAND0['MCW_nm'], bins=14, edgecolor='black')
plt.xlabel('MCW_nm')
plt.title('BAND 0')
plt.show()
```



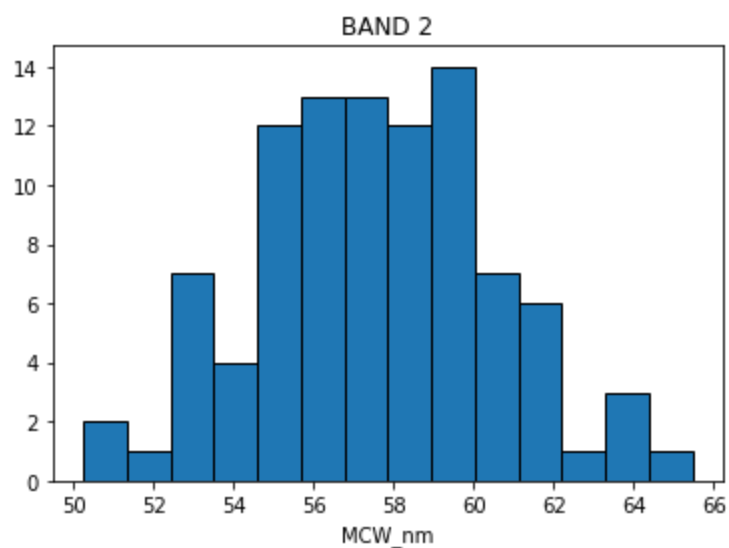
```
In [21]: plt.hist(BAND1['MCW_nm'], bins=14, edgecolor='black')
```

```
Loading [MathJax]/extensions/Safe.js
```

```
plt.title('BAND 1')
plt.show()
```



```
In [22]: plt.hist(BAND2['MCW_nm'], bins=14, edgecolor='black')
plt.xlabel('MCW_nm')
plt.title('BAND 2')
plt.show()
```



```
In [23]: new_table.describe()
```

```
Out[23]:
```

	PROCID	LHD	PHD	BAND	DataTP	MCW	MCW_nm
count	288.0	288.000000	288.000000	288.000000	288.0	288.000000	288.000000
mean	6400.0	3.500000	9.000000	1.000000	1188.0	112.816111	57.318304
std	0.0	2.295276	5.754562	0.817918	0.0	6.353221	3.227871
min	6400.0	0.000000	0.000000	0.000000	1188.0	97.265625	49.417593
25%	6400.0	1.750000	6.250000	0.000000	1188.0	108.734131	55.244379
50%	6400.0	3.500000	9.500000	1.000000	1188.0	112.390137	57.101880
75%	6400.0	5.250000	12.250000	2.000000	1188.0	116.857910	59.371815
max	6400.0	7.000000	17.000000	2.000000	1188.0	130.004883	66.051376

```
In [24]: new_table['MCW_nm'].mean()
```

Out[24]: 57.3183041537222

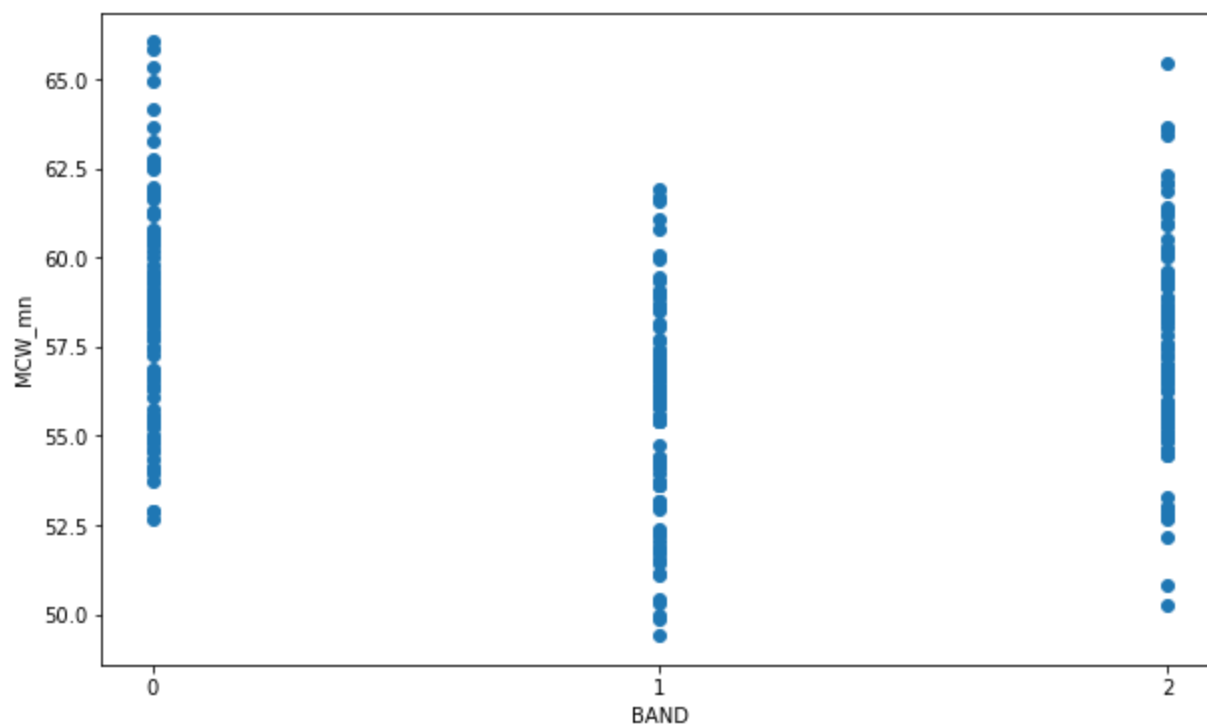
```
In [25]: new_table['BAND'].mean()
```

Out[25]: 1.0

```
In [50]: x = new_table['BAND']  
y = new_table['MCW_nm']
```

```
In [51]: plt.figure(figsize=(10,6))  
plt.scatter(x,y)  
plt.xlabel("BAND")  
plt.ylabel("MCW_mn")  
plt.xticks([0,1,2])
```

Out[51]: ([<matplotlib.axis.XTick at 0x1e256103160>, <matplotlib.axis.XTick at 0x1e2561036a0>, <matplotlib.axis.XTick at 0x1e2563b48e0>], [Text(0, 0, ''), Text(0, 0, ''), Text(0, 0, '')])



```
In [25]: BAND0['MCW_nm'].mean()
```

Out[25]: 58.73365387064688

```
In [29]: BAND1['MCW_nm'].mean()
```

Out[29]: 55.678648509776366

```
In [30]: BAND2['MCW_nm'].mean()
```

Out[30]: 57.54261008074344

```
In [70]: x = XYP['BAND']  
y = XYP['MCW_nm']  
x = new_table['BAND']  
y = new_table['MCW_nm']
```

```
In [71]: plt.figure(figsize=(10,6))
```

Loading [MathJax]/extensions/Safe.js

```

plt.plot(x,y,color='r',alpha=0.25)
plt.xlabel("BAND")
plt.ylabel("MCW_mn")

plt.xticks([0,1,2])

plt.legend(["MCW_nm", "BAND"])

plt.grid(b=True)
plt.show()

```

