

Data Table Creation

```
In [1]: import pandas
```

```
In [2]: mydataset = {
    'Student': ["Mary", "Shane", "David" , "Rachel"],
    'Physics': [40, 20, 25, 32],
    'Chemistry': [32, 25, 20, 40],
    'History': [30, 50, 40, 20]
}

myvar = pandas.DataFrame(mydataset)

print(myvar)
```

	Student	Physics	Chemistry	History
0	Mary	40	32	30
1	Shane	20	25	50
2	David	25	20	40
3	Rachel	32	40	20

```
In [3]: print(myvar.head(3))
#head () returns the headers and a specified number of rows, starting from
```

	Student	Physics	Chemistry	History
0	Mary	40	32	30
1	Shane	20	25	50
2	David	25	20	40

```
In [4]: print(myvar.tail(1))
#tail() returns the headers and a specified number of rows, starting from t
```

	Student	Physics	Chemistry	History
3	Rachel	32	40	20

Load Data

```
In [ ]: mydata = pandas.read_csv('data.csv')

print(mydata)
```

Data Analysis

```
In [6]: print(myvar.info())
#info() gives you more information about the data set
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4 entries, 0 to 3
Data columns (total 4 columns):
#   Column      Non-Null Count  Dtype
---  ---
0   Student      4 non-null      object
1   Physics      4 non-null      int64
2   Chemistry    4 non-null      int64
3   History      4 non-null      int64
dtypes: int64(3), object(1)
memory usage: 256.0+ bytes
None
```

```
In [7]: myvar["Physics"].mean()
```

```
Out[7]: 29.25
```

```
In [8]: myvar["Physics"].median()
```

```
Out[8]: 28.5
```

```
In [9]: myvar["Physics"].max()
```

```
Out[9]: 40
```

```
In [10]: myvar["Physics"].idxmax()
```

```
Out[10]: 0
```

```
In [11]: myvar.iloc[myvar["Physics"].idxmax()]["Student"]
```

```
Out[11]: 'Mary'
```

```
In [12]: myvar["Physics"].min()
```

```
Out[12]: 20
```

```
In [13]: myvar.describe()
```

```
Out[13]:
```

	Physics	Chemistry	History
count	4.000000	4.000000	4.000000
mean	29.250000	29.250000	35.000000
std	8.693868	8.693868	12.909944
min	20.000000	20.000000	20.000000
25%	23.750000	23.750000	27.500000
50%	28.500000	28.500000	35.000000
75%	34.000000	34.000000	42.500000
max	40.000000	40.000000	50.000000

```
In [14]: myvar["Physics"].count()
```

```
Out[14]: 4
```

```
In [15]: myvar["Physics"].sum()
```

Out[15]: 117

Plots

```
In [16]: import matplotlib.pyplot as plt
```

```
In [17]: plot_data = myvar.copy()

print("Before:", '\n', plot_data.head())

plot_data = plot_data.set_index("Student")

print("After:", '\n', plot_data.head())
```

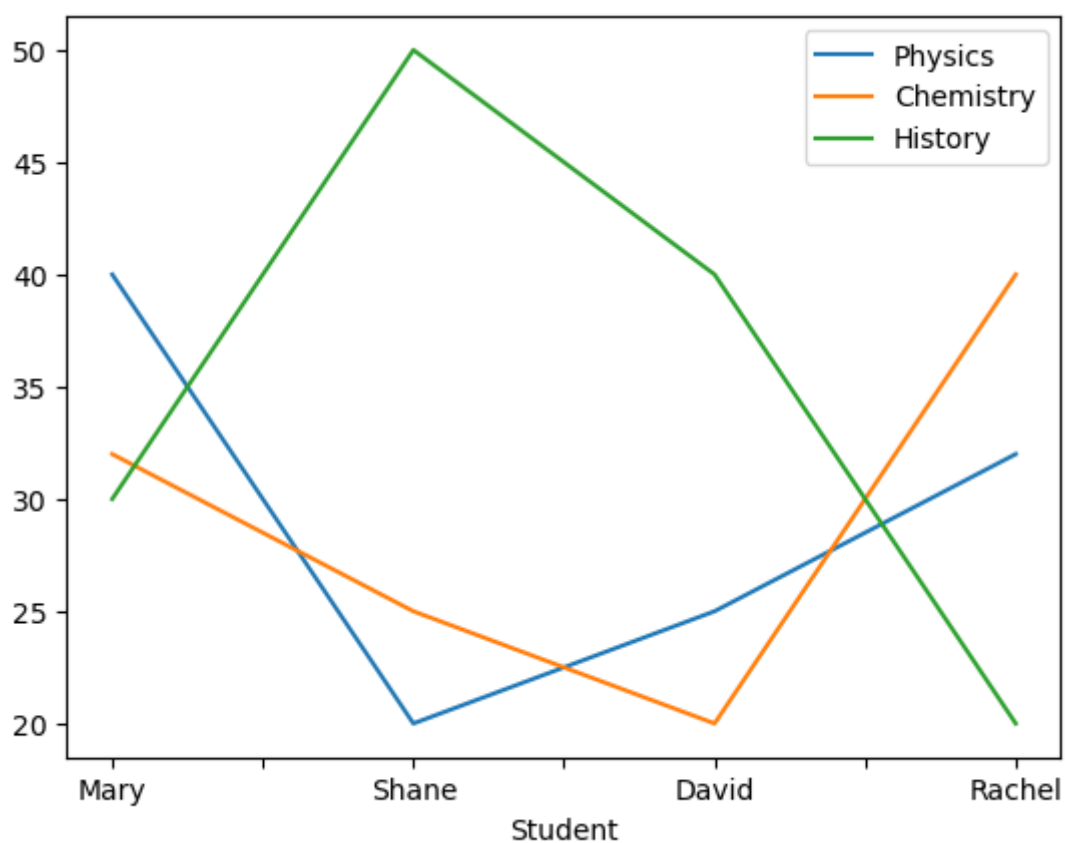
Before:

	Student	Physics	Chemistry	History
0	Mary	40	32	30
1	Shane	20	25	50
2	David	25	20	40
3	Rachel	32	40	20

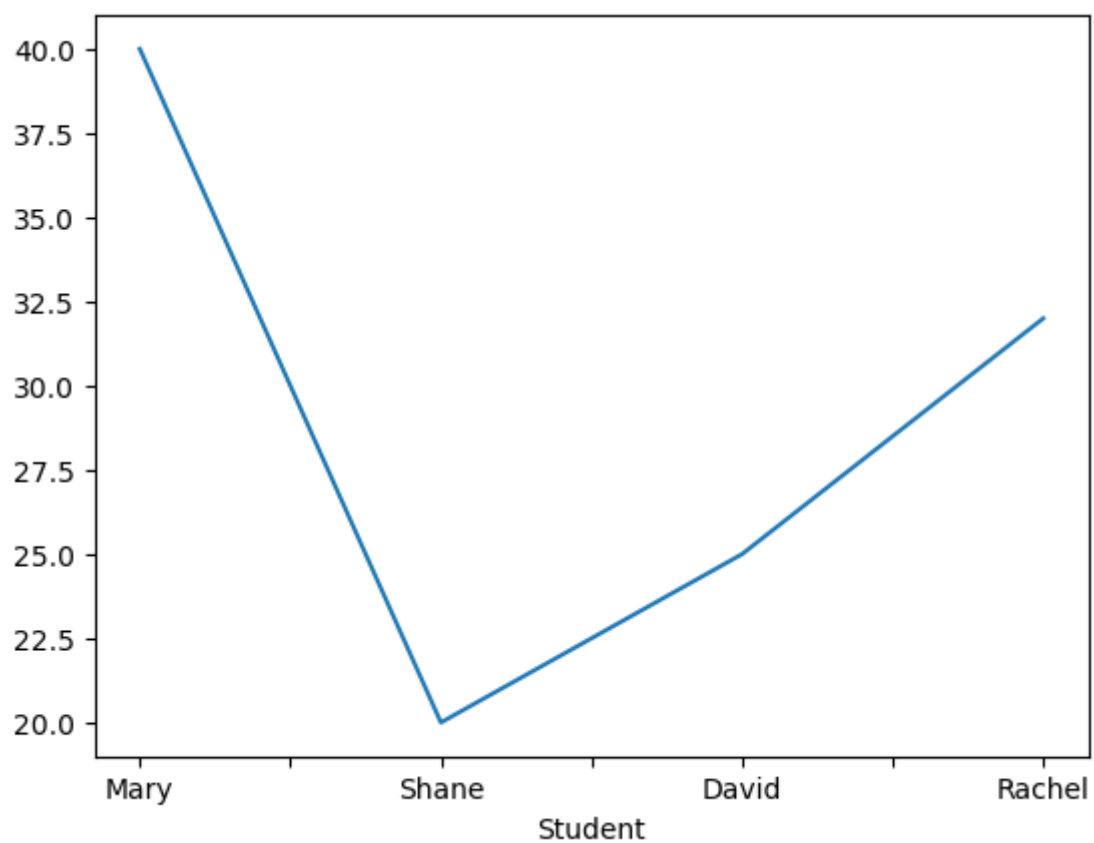
After:

	Student	Physics	Chemistry	History
	Mary	40	32	30
	Shane	20	25	50
	David	25	20	40
	Rachel	32	40	20

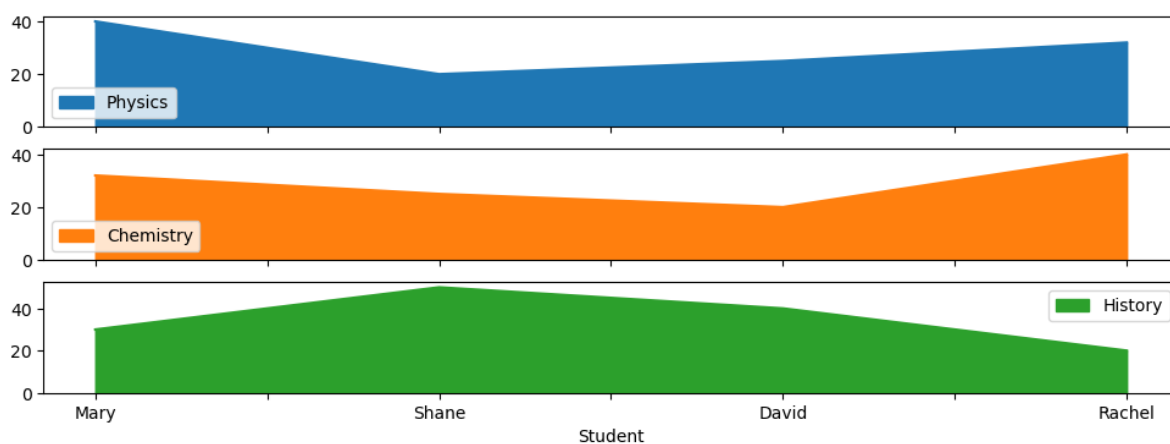
```
In [18]: plot_data.plot()
plt.show()
```



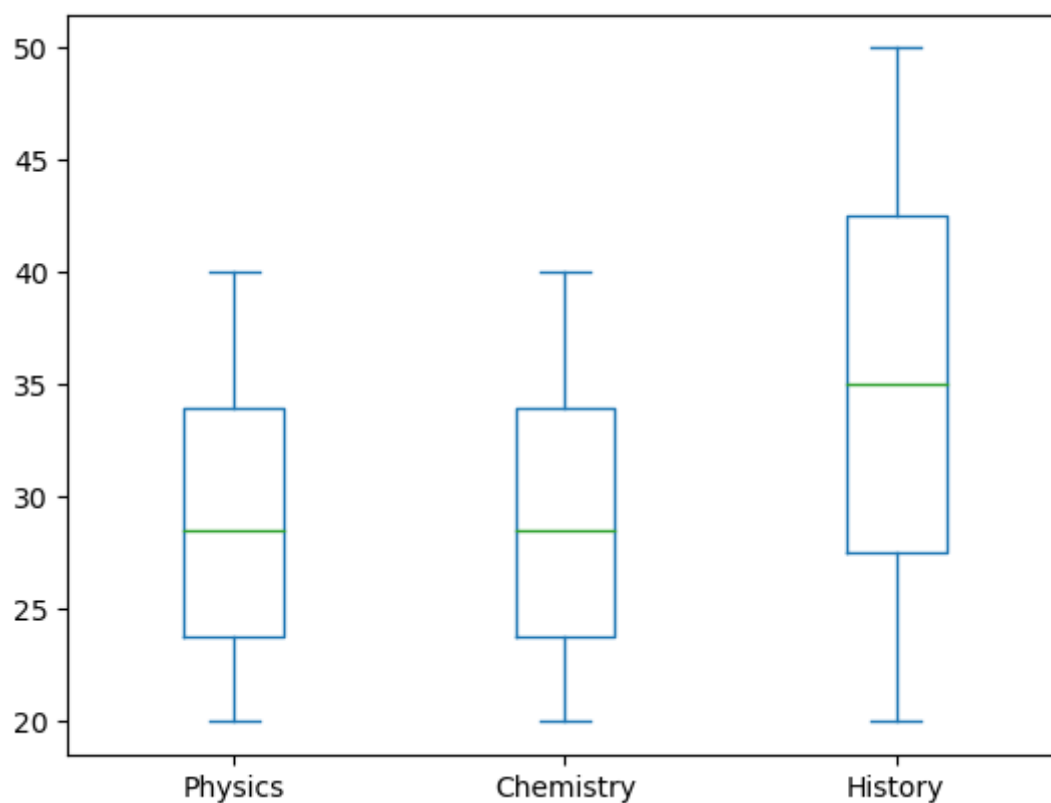
```
In [19]: plot_data["Physics"].plot()
plt.show()
```



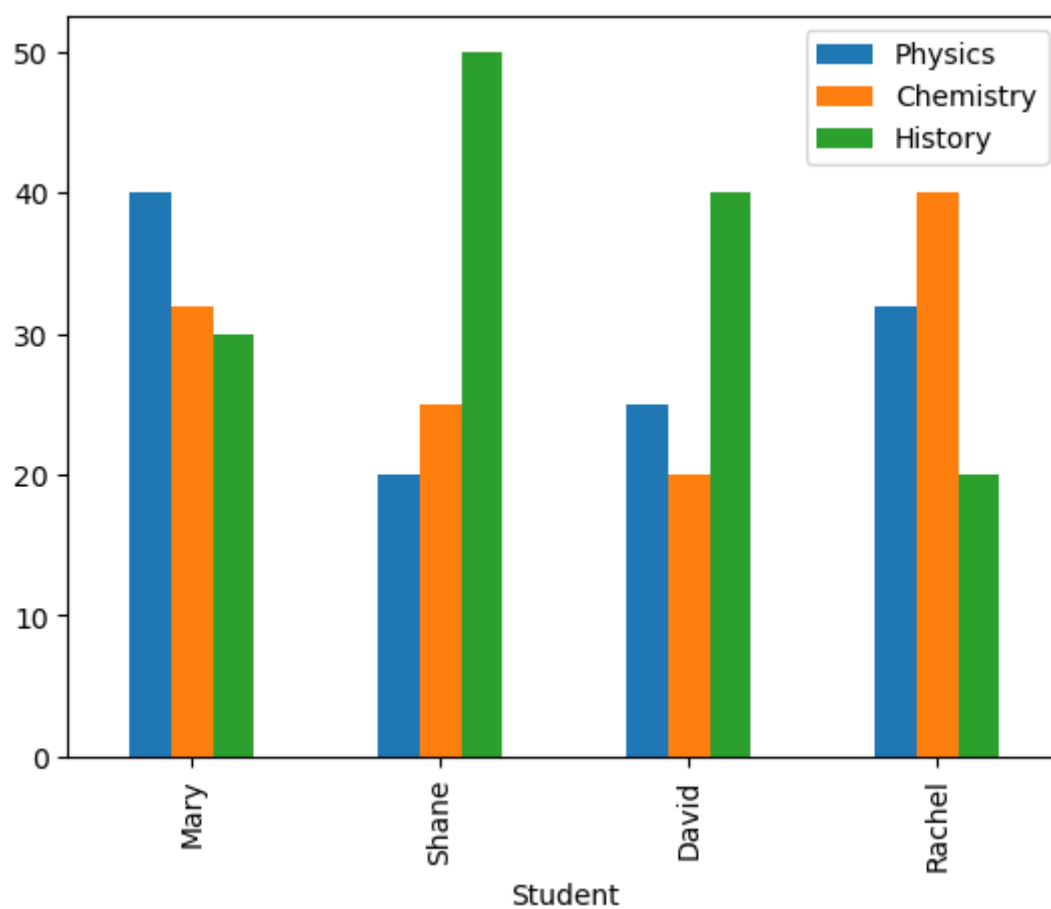
```
In [21]: plot_data.plot.area(figsize=(12, 4), subplots=True)
plt.show()
```



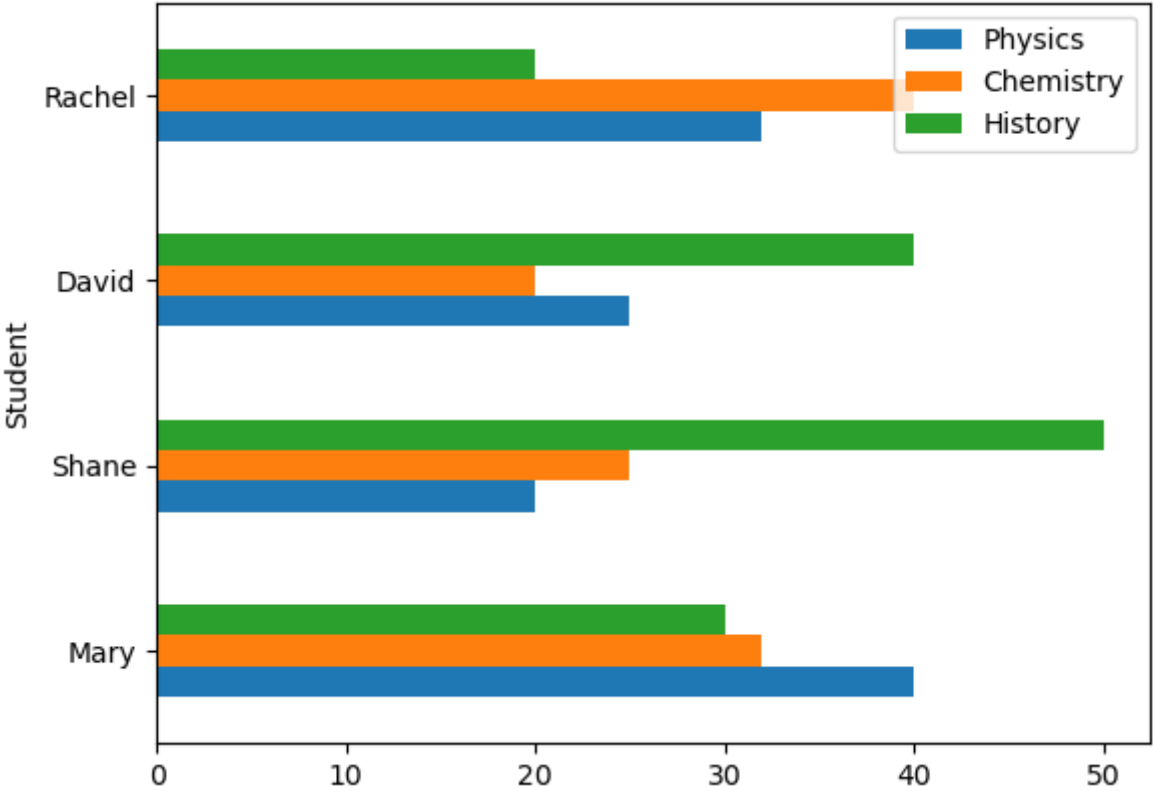
```
In [22]: plot_data.plot.box()
plt.show()
```



```
In [23]: plot_data.plot.bar()  
plt.show()
```



```
In [24]: plot_data.plot.barh()  
plt.show()
```



```
In [ ]:
```