minor project

May 18, 2020

[29]:
```
"KARTIKEYA SINGH"
"BATCH 2"
"PYTHON MINOR PROJECT"
```

[29]:
```
'PYTHON MINOR PROJECT'
```

[30]:
```
"MENTORS- GUHAN SIR AND HARSHVARDHAN SIR"
```

[30]:
```
'MENTORS- GUHAN SIR AND HARSHVARDHAN SIR'
```

[1]:
```
"minor project"
```

[1]:
```
'minor project'
```

[4]:
```
import pandas as pd
table=pd.read_csv(r"C:\Users\lenovo\Desktop\student.xlsx.csv")
table.head(10)
```

[4]:

	gender	race	education	lunch	test preparation course	maths	reading	writing
0	female	group B	bachelor's degree	standard	none	72	72	74
1	female	group C	some college	standard	completed	69	90	88
2	female	group B	master's degree	standard	none	90	95	93
3	male	group A	associate's degree	free/reduced	none	47	57	44
4	male	group C	some college	standard	none	76	78	75
5	female	group B	associate's degree	standard	none	71	83	78
6	female	group B	some college	standard	completed	88	95	92
7	male	group B	some college	free/reduced	none			
8	male	group D	high school	free/reduced	completed			
9	female	group B	high school	free/reduced	none			

```
7     40       43       39
8     64       64       67
9     38       60       50
```

[17]:
```
table.gender.unique()
```

[17]:
```
array(['female', 'male'], dtype=object)
```

[19]:
```
table.race.unique()
```

[19]:
```
array(['group B', 'group C', 'group A', 'group D', 'group E'],
      dtype=object)
```

[22]:
```
table.lunch.unique()
```

[22]:
```
array(['standard', 'free/reduced'], dtype=object)
```

[23]:
```
table.race.nunique()
```

[23]:
```
5
```

[5]:
```
table.dtypes
```

[5]:
```
gender                       object
race                         object
education                    object
lunch                        object
test preparation course      object
maths                         int64
reading                       int64
writing                       int64
dtype: object
```

[81]:
```
table.tail(6)
```

[81]:
```
     gender     race        education         lunch test preparation course  \
994    male  group A      high school      standard                    none
995  female  group E  master's degree      standard                completed
996    male  group C      high school  free/reduced                    none
997  female  group C      high school  free/reduced                completed
998  female  group D     some college      standard                completed
999  female  group D     some college  free/reduced                    none

     maths   reading  writing
994     63       63        62
995     88       99        95
996     62       55        55
```

```
997      59       71        65
998      68       78        77
999      77       86        86
```

[80]:
```
table.describe()
```

[80]:
```
            maths       reading     writing
count  1000.00000  1000.000000  1000.000000
mean     66.08900    69.169000    68.054000
std      15.16308    14.600192    15.195657
min       0.00000    17.000000    10.000000
25%      57.00000    59.000000    57.750000
50%      66.00000    70.000000    69.000000
75%      77.00000    79.000000    79.000000
max     100.00000   100.000000   100.000000
```

[30]:
```
table.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 8 columns):
 #   Column                       Non-Null Count  Dtype
---  ------                       --------------  -----
 0   gender                       1000 non-null   object
 1   race                         1000 non-null   object
 2   education                    1000 non-null   object
 3   lunch                        1000 non-null   object
 4   test preparation course      1000 non-null   object
 5   math                         1000 non-null   int64
 6   reading                      1000 non-null   int64
 7   writing                      1000 non-null   int64
dtypes: int64(3), object(5)
memory usage: 62.6+ KB
```

[31]:
```
table.shape
```

[31]:
```
(1000, 8)
```

[77]:
```
table.describe(include="all")
```

[77]:
```
        gender     race     education     lunch test preparation course  \
count     1000     1000          1000      1000                     1000
unique       2        5             6         2                        2
top     female  group C  some college  standard                     none
freq       518      319           226       645                      642
mean       NaN      NaN           NaN       NaN                      NaN
std        NaN      NaN           NaN       NaN                      NaN
```

```
min        NaN      NaN           NaN       NaN                     NaN
25%        NaN      NaN           NaN       NaN                     NaN
50%        NaN      NaN           NaN       NaN                     NaN
75%        NaN      NaN           NaN       NaN                     NaN
max        NaN      NaN           NaN       NaN                     NaN

             maths       reading     writing
count   1000.00000  1000.000000  1000.000000
unique         NaN          NaN          NaN
top            NaN          NaN          NaN
freq           NaN          NaN          NaN
mean       66.08900    69.169000    68.054000
std        15.16308    14.600192    15.195657
min         0.00000    17.000000    10.000000
25%        57.00000    59.000000    57.750000
50%        66.00000    70.000000    69.000000
75%        77.00000    79.000000    79.000000
max       100.00000   100.000000   100.000000
```

[78]:
```
table.columns
```

[78]:
```
Index(['gender', 'race', 'education', 'lunch', 'test preparation course',
       'maths ', 'reading', 'writing '],
      dtype='object')
```

[6]:
```
import pandas as pd
table=pd.read_csv(r"C:\Users\lenovo\Desktop\student.xlsx.csv")
table.head(5)
```

[6]:
```
   gender     race           education         lunch test preparation course  \
0  female  group B   bachelor's degree      standard                    none
1  female  group C        some college      standard               completed
2  female  group B     master's degree      standard                    none
3    male  group A  associate's degree  free/reduced                    none
4    male  group C        some college      standard                    none

   maths  reading  writing
0     72       72       74
1     69       90       88
2     90       95       93
3     47       57       44
4     76       78       75
```

[100]:
```
table.maths
```

[100]:
```
0      72
1      69
```

```
2      90
3      47
4      76
       ..
995    88
996    62
997    59
998    68
999    77
Name: maths, Length: 1000, dtype: int64
```

[102]:
```
table.reading
```

[102]:
```
0      72
1      90
2      95
3      57
4      78
       ..
995    99
996    55
997    71
998    78
999    86
Name: reading, Length: 1000, dtype: int64
```

[134]:
```
table.writing
```

[134]:
```
0      74
1      88
2      93
3      44
4      75
       ..
995    95
996    55
997    65
998    77
999    86
Name: writing, Length: 1000, dtype: int64
```

[125]:
```
table.maths.sort_values(ascending=False).head(100)
```

[125]:
```
149    100
962    100
451    100
623    100
```

```
625    100
       …
426     86
990     86
447     86
638     86
509     86
Name: maths, Length: 100, dtype: int64
```

[128]:
```
table.reading.sort_values(ascending=False).head(25)
```

[128]:
```
458    100
179    100
962    100
594    100
106    100
970    100
381    100
712    100
114    100
566    100
916    100
903    100
957    100
886    100
165    100
546    100
149    100
403     99
685     99
995     99
580     97
855     97
625     97
514     97
373     97
Name: reading, dtype: int64
```

[135]:
```
table.writing.sort_values(ascending=False).head(25)
```

[135]:
```
114    100
916    100
957    100
106    100
970    100
179    100
903    100
```

```
458    100
685    100
403    100
165    100
377    100
962    100
566    100
717     99
712     99
594     99
625     99
652     98
110     98
451     97
546     97
580     96
373     96
855     96
Name: writing, dtype: int64
```

[7]:
```
# male student with above 90 percent marks in all 3 subject
table[(table.maths >90) & (table.reading >90) & (table.writing >90) & (table.
↪gender =='male')]
```

[7]:
```
     gender     race           education         lunch test preparation course  \
149    male  group E  associate's degree  free/reduced                completed
571    male  group A   bachelor's degree      standard                     none
625    male  group D        some college      standard                completed
864    male  group C  associate's degree      standard                     none
916    male  group E   bachelor's degree      standard                completed
919    male  group B        some college      standard                completed

     maths  reading  writing
149    100      100       93
571     91       96       92
625    100       97       99
864     97       93       91
916    100      100      100
919     91       96       91
```

[8]:
```
# male student with above 90 percent marks in all 3 subject
table[(table.maths >90) & (table.reading >90) & (table.writing >90) & (table.
↪gender =='female')]
```

[8]:
```
     gender     race           education         lunch  \
114  female  group E   bachelor's degree      standard
165  female  group C   bachelor's degree      standard
```

```
     gender     race           education         lunch  \
179  female  group D    some high school      standard
451  female  group E        some college      standard
458  female  group E   bachelor's degree      standard
546  female  group A    some high school      standard
566  female  group E   bachelor's degree  free/reduced
594  female  group C   bachelor's degree      standard
685  female  group E     master's degree      standard
712  female  group D        some college      standard
717  female  group C  associate's degree      standard
855  female  group B   bachelor's degree      standard
886  female  group E  associate's degree      standard
903  female  group D   bachelor's degree  free/reduced
957  female  group D     master's degree      standard
962  female  group E  associate's degree      standard
979  female  group C  associate's degree      standard

    test preparation course  maths  reading  writing
114               completed     99      100      100
165               completed     96      100      100
179               completed     97      100      100
451                    none    100       92       97
458                    none    100      100      100
546               completed     92      100       97
566               completed     92      100      100
594               completed     92      100       99
685               completed     94       99      100
712                    none     98      100       99
717               completed     96       96       99
855                    none     97       97       96
886               completed     93      100       95
903               completed     93      100      100
957                    none     92      100      100
962                    none    100      100      100
979                    none     91       95       94
```

[9]:
```
# female are more studious than male
```

[41]:
```
x=table[(table.maths >90) & (table.reading >90) & (table.writing >90)]
x.head(10)
```

[41]:
```
     gender     race           education         lunch  \
114  female  group E   bachelor's degree      standard
149    male  group E  associate's degree  free/reduced
165  female  group C   bachelor's degree      standard
179  female  group D    some high school      standard
451  female  group E        some college      standard
458  female  group E   bachelor's degree      standard
```

```
546  female  group A    some high school      standard
566  female  group E   bachelor's degree  free/reduced
571    male  group A   bachelor's degree      standard
594  female  group C   bachelor's degree      standard

    test preparation course  maths  reading  writing
114               completed     99      100      100
149               completed    100      100       93
165               completed     96      100      100
179               completed     97      100      100
451                    none    100       92       97
458                    none    100      100      100
546               completed     92      100       97
566               completed     92      100      100
571                    none     91       96       92
594               completed     92      100       99
```

[35]:
```
# student of group A with above 90 percent in all 3 subject
x[x.race =='group A']
```

[35]:
```
     gender     race          education     lunch test preparation course  \
546  female  group A   some high school  standard               completed
571    male  group A  bachelor's degree  standard                    none

     maths  reading  writing
546     92      100       97
571     91       96       92
```

[36]:
```
# student of group B with above 90 percent in all 3 subject
x[x.race =='group B']
```

[36]:
```
     gender     race          education     lunch test preparation course  \
855  female  group B  bachelor's degree  standard                    none
919    male  group B       some college  standard               completed

     maths  reading  writing
855     97       97       96
919     91       96       91
```

[37]:
```
# student of group C with above 90 percent in all 3 subject
x[x.race =='group C']
```

[37]:
```
     gender     race           education     lunch test preparation course  \
165  female  group C   bachelor's degree  standard               completed
594  female  group C   bachelor's degree  standard               completed
717  female  group C  associate's degree  standard               completed
864    male  group C  associate's degree  standard                    none
```

```
979  female  group C  associate's degree  standard                    none

     maths  reading  writing
165     96      100      100
594     92      100       99
717     96       96       99
864     97       93       91
979     91       95       94
```

[38]:
```
# student of group D with above 90 percent in all 3 subject
x[x.race =='group D']
```

[38]:
```
     gender     race           education         lunch test preparation course  \
179  female  group D   some high school      standard                completed
625    male  group D       some college      standard                completed
712  female  group D       some college      standard                     none
903  female  group D  bachelor's degree  free/reduced                completed
957  female  group D    master's degree      standard                     none

     maths  reading  writing
179     97      100      100
625    100       97       99
712     98      100       99
903     93      100      100
957     92      100      100
```

[39]:
```
# student of group E with above 90 percent in all 3 subject
x[x.race =='group E']
```

[39]:
```
     gender     race           education         lunch  \
114  female  group E   bachelor's degree      standard
149    male  group E  associate's degree  free/reduced
451  female  group E        some college      standard
458  female  group E   bachelor's degree      standard
566  female  group E   bachelor's degree  free/reduced
685  female  group E     master's degree      standard
886  female  group E  associate's degree      standard
916    male  group E   bachelor's degree      standard
962  female  group E  associate's degree      standard

    test preparation course  maths  reading  writing
114               completed     99      100      100
149               completed    100      100       93
451                    none    100       92       97
458                    none    100      100      100
566               completed     92      100      100
685               completed     94       99      100
```

```
886              completed     93      100       95
916              completed    100      100      100
962                   none    100      100      100
```

[6]:
```
table.maths.mean()
```

[6]:
```
66.089
```

[7]:
```
table.reading.mean()
```

[7]:
```
69.169
```

[8]:
```
table.writing.mean()
```

[8]:
```
68.054
```

[9]:
```
# using groupby function
```

[12]:
```
table.groupby('gender').agg(['count', 'mean', 'min', 'max'])
```

[12]:
```
       maths                     reading                     writing  \
       count       mean min  max   count       mean min  max   count
gender
female   518  63.633205   0  100     518  72.608108  17  100     518
male     482  68.728216  27  100     482  65.473029  23  100     482


             mean min  max
gender
female  72.467181  10  100
male    63.311203  15  100
```

[13]:
```
table.groupby('race').agg(['count', 'mean', 'min', 'max'])
```

[13]:
```
        maths                     reading                     writing  \
        count       mean min  max   count       mean min  max   count
race
group A    89  61.629213  28  100      89  64.674157  23  100      89
group B   190  63.452632   8   97     190  67.352632  24   97     190
group C   319  64.463950   0   98     319  69.103448  17  100     319
group D   262  67.362595  26  100     262  70.030534  31  100     262
group E   140  73.821429  30  100     140  73.028571  26  100     140


              mean min  max
race
group A  62.674157  19   97
```

```
group B  65.600000  15   96
group C  67.827586  10  100
group D  70.145038  32  100
group E  71.407143  22  100
```

[14]:
```
table.groupby('education').agg(['count', 'mean', 'min', 'max'])
```

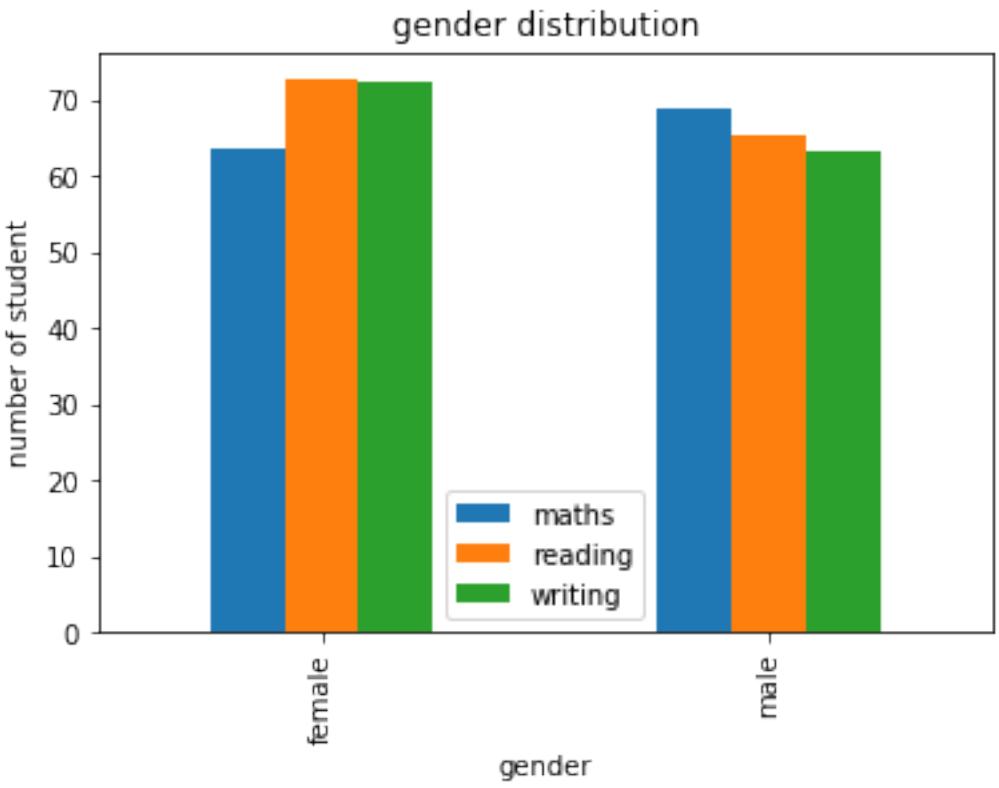
[14]:
```
                    maths                      reading                      \
                    count       mean min  max   count       mean min  max
education
associate's degree    222  67.882883  26  100     222  70.927928  31  100
bachelor's degree     118  69.389831  29  100     118  73.000000  41  100
high school           196  62.137755   8   99     196  64.704082  24   99
master's degree        59  69.745763  40   95      59  75.372881  42  100
some college          226  67.128319  19  100     226  69.460177  23  100
some high school      179  63.497207   0   97     179  66.938547  17  100

                    writing
                      count       mean min  max
education
associate's degree      222  69.896396  35  100
bachelor's degree       118  73.381356  38  100
high school             196  62.448980  15  100
master's degree          59  75.677966  46  100
some college            226  68.840708  19   99
some high school        179  64.888268  10  100
```

[15]:
```
# with the help of above tables we can analyze all types of  relation (using␣
↪groupby)
```

[1]:
```
# drawing some graphs/plots
```

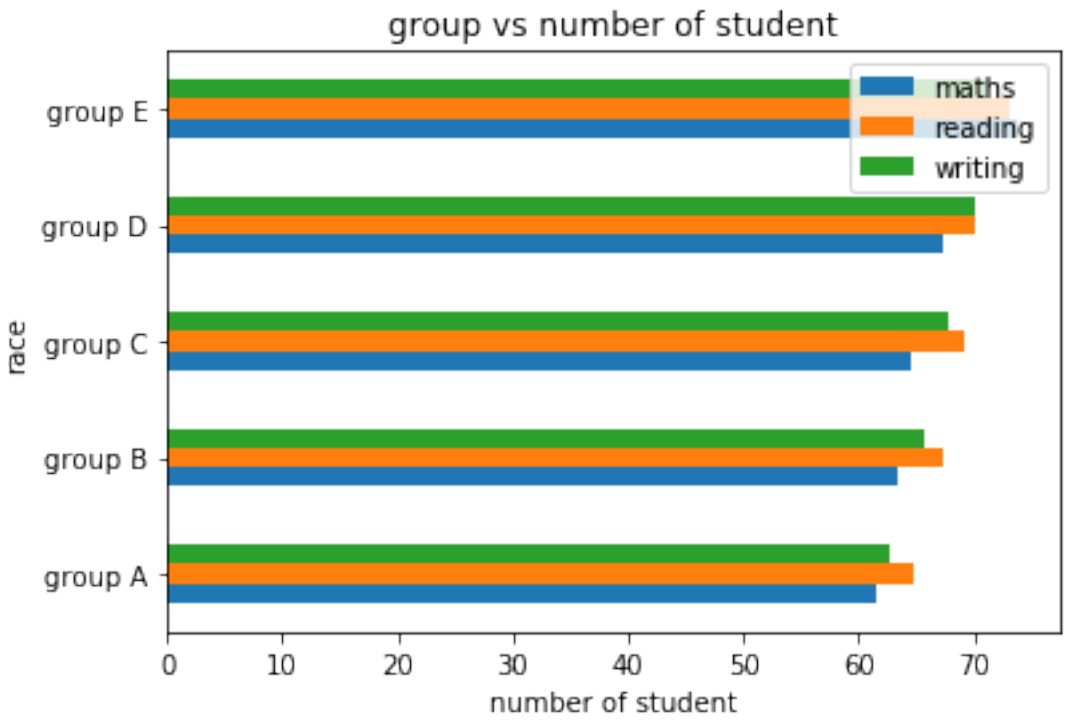
[24]:
```
import matplotlib.pyplot as plt
%matplotlib inline
table.groupby('gender').mean().plot(kind ='bar')
plt.ylabel('number of student')
plt.title('gender distribution')
```

[24]:
```
Text(0.5, 1.0, 'gender distribution')
```



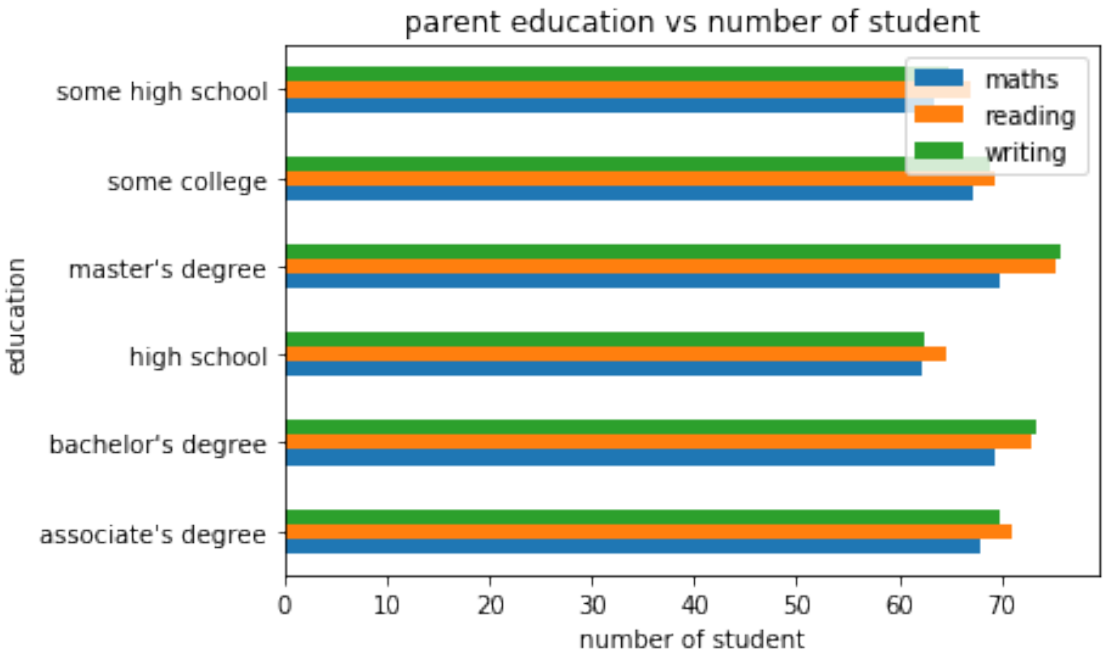
[26]:
```
table.groupby('race').mean().plot(kind ='barh')
plt.xlabel('number of student')
plt.title('group vs number of student')
```

[26]: Text(0.5, 1.0, 'group vs number of student')



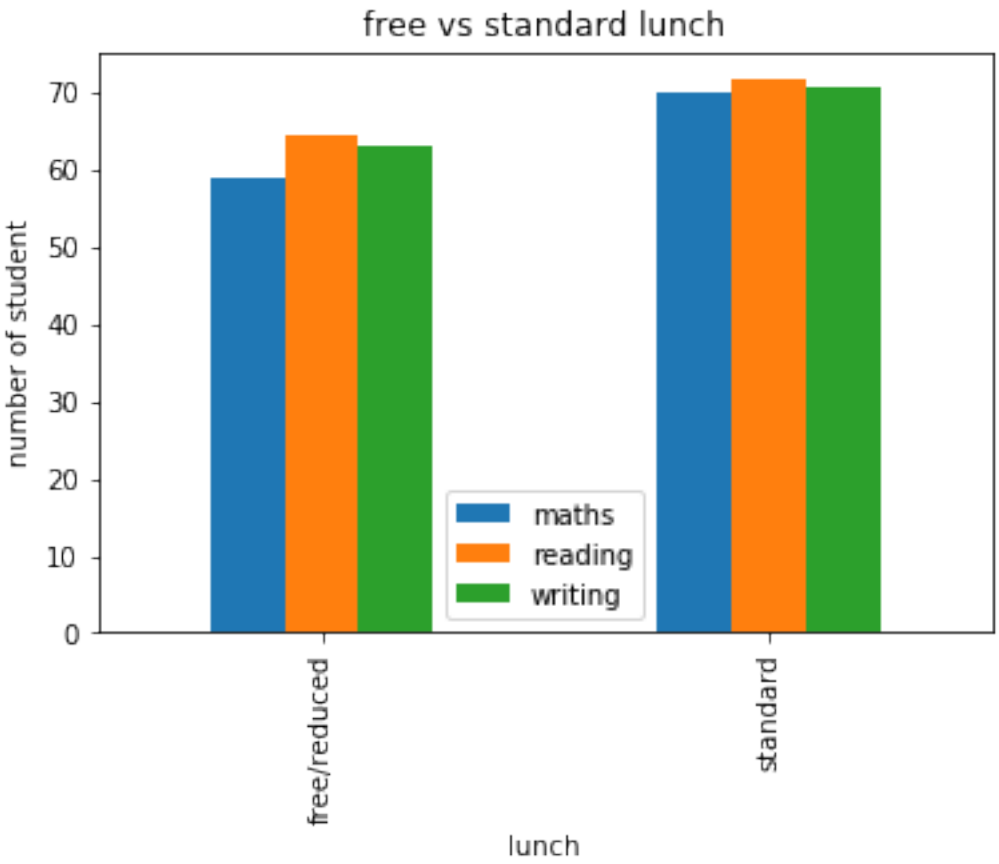
[27]:
```
table.groupby('education').mean().plot(kind ='barh')
plt.xlabel('number of student')
plt.title('parent education vs number of student')
```

[27]: Text(0.5, 1.0, 'parent education vs number of student')



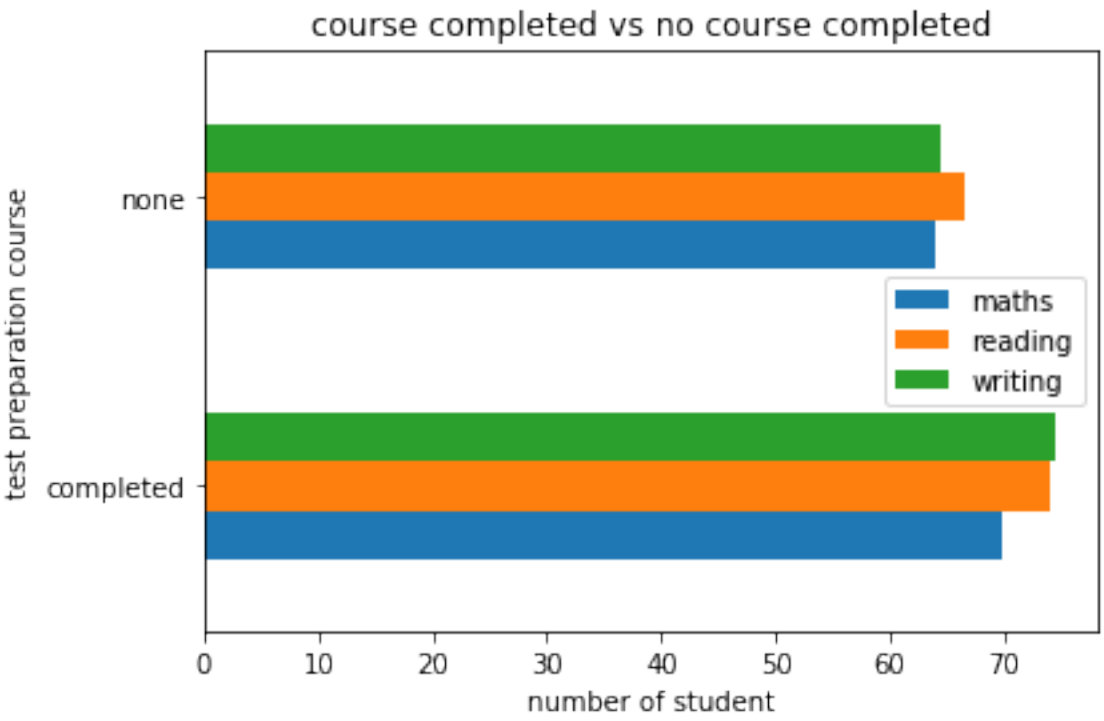
[21]:
```
table.groupby('lunch').mean().plot(kind ='bar')
plt.title('free vs standard lunch')
plt.ylabel('number of student')
```

[21]: Text(0, 0.5, 'number of student')



[28]:
```
table.groupby('test preparation course').mean().plot(kind ='barh')
plt.title('course completed vs no course completed')
plt.xlabel('number of student')
```

[28]: Text(0.5, 0, 'number of student')


course completed vs no course completed
test preparation course
none
completed
maths
reading
writing
0
10
20
30
40
50
60
70
number of student

[ ]: