```
In [38]: import pandas as pd
         import matplotlib.pyplot as plt
         import numpy as np
         import seaborn as sns
```

```
In [2]: df  = pd.read_csv("student-mat.csv")
        df.head()
```

Out[2]:

	school	sex	age	address	famsize	Pstatus	Medu	Fedu	Mjob	Fjob	...	famrel	freetime
0	GP	F	18	U	GT3	A	4	4	at_home	teacher	...	4	3
1	GP	F	17	U	GT3	T	1	1	at_home	other	...	5	3
2	GP	F	15	U	LE3	T	1	1	at_home	other	...	4	3
3	GP	F	15	U	GT3	T	4	2	health	services	...	3	2
4	GP	F	16	U	GT3	T	3	3	other	other	...	4	3

5 rows × 33 columns

```
In [3]: df.shape
```

Out[3]: (395, 33)

```
In [8]: df.isnull().sum()
```

```
Out[8]:   school        0
          sex           0
          age           0
          address       0
          famsize       0
          Pstatus       0
          Medu          0
          Fedu          0
          Mjob          0
          Fjob          0
          reason        0
          guardian      0
          traveltime    0
          studytime     0
          failures      0
          schoolsup     0
          famsup        0
          paid          0
          activities    0
          nursery       0
          higher        0
          internet      0
          romantic      0
          famrel        0
          freetime      0
          goout         0
          Dalc          0
          Walc          0
          health        0
          absences      0
          G1            0
          G2            0
          G3            0
          dtype: int64
```

In [4]: `df.dtypes`

```
Out[4]:  school          object
         sex             object
         age              int64
         address         object
         famsize         object
         Pstatus         object
         Medu             int64
         Fedu             int64
         Mjob            object
         Fjob            object
         reason          object
         guardian        object
         traveltime       int64
         studytime        int64
         failures         int64
         schoolsup       object
         famsup          object
         paid            object
         activities      object
         nursery         object
         higher          object
         internet        object
         romantic        object
         famrel           int64
         freetime         int64
         goout            int64
         Dalc             int64
         Walc             int64
         health           int64
         absences         int64
         G1               int64
         G2               int64
         G3               int64
         dtype: object
```

In [16]:
```python
df.drop_duplicates(inplace = True)
```

In [17]:
```python
df.head()
```

Out[17]:

	school	sex	age	address	famsize	Pstatus	Medu	Fedu	Mjob	Fjob	...	famrel	freetime
0	GP	F	18	U	GT3	A	4	4	at_home	teacher	...	4	3
1	GP	F	17	U	GT3	T	1	1	at_home	other	...	5	3
2	GP	F	15	U	LE3	T	1	1	at_home	other	...	4	3
3	GP	F	15	U	GT3	T	4	2	health	services	...	3	2
4	GP	F	16	U	GT3	T	3	3	other	other	...	4	3

5 rows × 33 columns

In [15]:
```python
df.describe()
```

	age	Medu	Fedu	traveltime	studytime	failures	famrel	freetime
count	395.000000	395.000000	395.000000	395.000000	395.000000	395.000000	395.000000	395.000000
mean	16.696203	2.749367	2.521519	1.448101	2.035443	0.334177	3.944304	3.235443
std	1.276043	1.094735	1.088201	0.697505	0.839240	0.743651	0.896659	0.998862
min	15.000000	0.000000	0.000000	1.000000	1.000000	0.000000	1.000000	1.000000
25%	16.000000	2.000000	2.000000	1.000000	1.000000	0.000000	4.000000	3.000000
50%	17.000000	3.000000	2.000000	1.000000	2.000000	0.000000	4.000000	3.000000
75%	18.000000	4.000000	3.000000	2.000000	2.000000	0.000000	5.000000	4.000000
max	22.000000	4.000000	4.000000	4.000000	4.000000	3.000000	5.000000	5.000000

Q1 What is the average score in math (G3)?

```python
df['G3'].mean()
```

Out[18]:
```
10.415189873417722
```

Q2 How many students scored above 15 in their final grade (G3)?

```python
count = df[df['G3']>15].shape[0]
print(count)
```

```
40
```

Q3 Is there a correlation between study time (study time) and the final grade (G3)?

```python
corr = df['studytime'].corr(df['G3'])
print(corr)
```

```
0.0978196896531963
```

Q4 Which gender has a higher average final grade (G3)?

```python
gender_avg = df.groupby('sex')['G3'].mean()
print(gender_avg)
```

```
sex
F       9.966346
M      10.914439
Name: G3, dtype: float64
```

Q5 Plot a histogram of final grades (G3).

```
In [62]:   plt.figure(figsize = (4,5))
           plt.hist(df['G3'],bins = 10,color = 'skyblue',edgecolor = 'black')
           plt.legend('G3')
           plt.title('Histogram chart')
           plt.xlabel('x_value')
           plt.ylabel('number_of_student')
           plt.show()
```
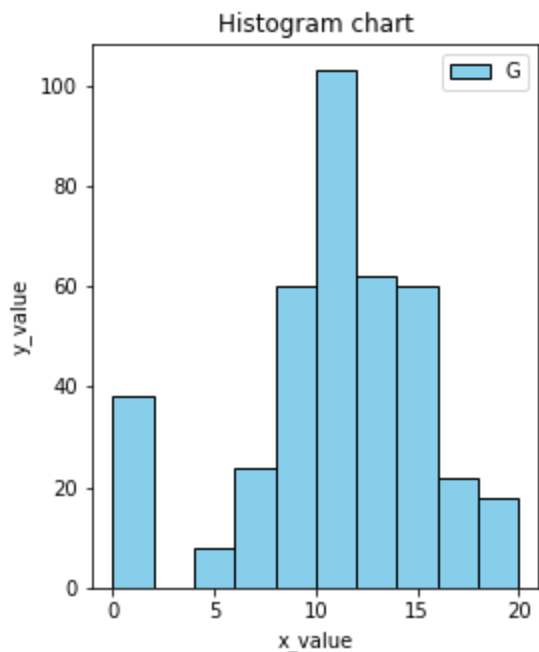

Q6 Create a scatter plot between study time (study time) and final grade (G3).

```
In [64]:   plt.figure(figsize = (4,5))
           plt.scatter(df['studytime'],df['G3'],alpha = 0.6,color = 'skyblue',edgecolor = 'black'
           plt.legend('chart')
           plt.title('study and grade chart')
           plt.xlabel('studytime hour per week')
           plt.ylabel('grade_value')
           plt.show()
```
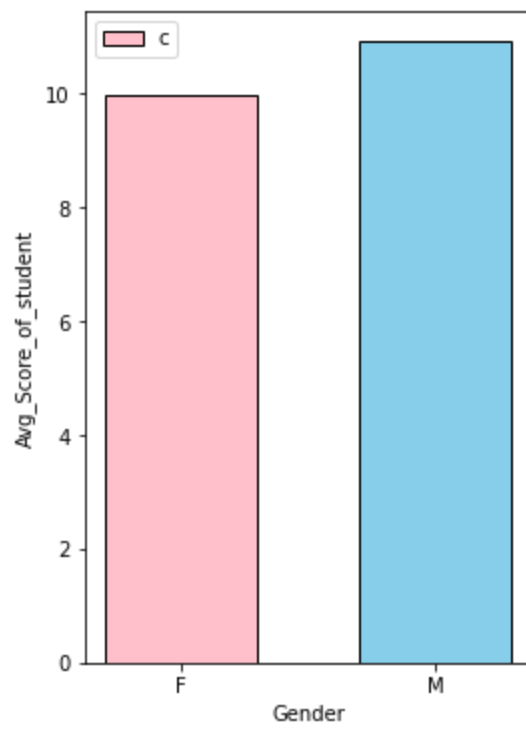
study and grade chart
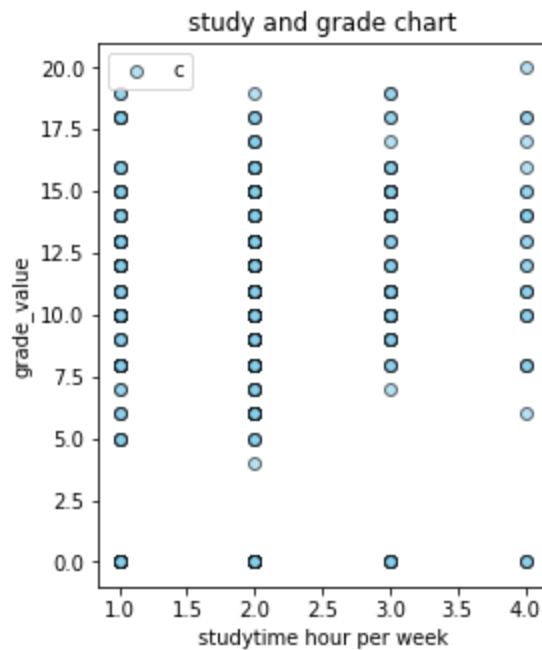
Q7 Create a bar chart comparing the average scores of male and female students.

In [65]:
```python
gender_avg = df.groupby('sex')['G3'].mean()
print(gender_avg)
```

```
sex
F     9.966346
M    10.914439
Name: G3, dtype: float64
```

In [73]:
```python
plt.figure(figsize = (4,6))
plt.bar(gender_avg.index,gender_avg.values,color = ['pink','skyblue'],width = 0.6,edge
plt.xlabel('Gender')
plt.ylabel('Avg_Score_of_student')
plt.legend('chart')
plt.show()
```

In [ ]: