Import necessary libraries

In [16]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

Load dataset

In [30]:
```
df=pd.read_csv("student-mat.csv")
```

In [32]:
```
df.head()
```

Out[32]:

	school	sex	age	address	famsize	Pstatus	Medu	Fedu	Mjob	Fjob	...	famrel	freetime	goout	Dalc	Walc	health	absences	G1	G2	G3
0	GP	F	18	U	GT3	A	4	4	at_home	teacher	...	4	3	4	1	1	3	6	5	6	6
1	GP	F	17	U	GT3	T	1	1	at_home	other	...	5	3	3	1	1	3	4	5	5	6
2	GP	F	15	U	LE3	T	1	1	at_home	other	...	4	3	2	2	3	3	10	7	8	10
3	GP	F	15	U	GT3	T	4	2	health	services	...	3	2	2	1	1	5	2	15	14	15
4	GP	F	16	U	GT3	T	3	3	other	other	...	4	3	2	1	2	5	4	6	10	10

5 rows × 33 columns

Explore & Clean Data

In [34]:
```
df.shape
```

Out[34]:
```
(395, 33)
```

In [44]:
```
df.dtypes
```

Out[44]:
```
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 [48]:
```
df.isnull().sum()
```

Out[48]:
```
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 [61]:
```
df.drop_duplicates(inplace=True)
```

In [63]:
```
df.shape
```

Out[63]:
```
(395, 33)
```

Analysis questions

1. Average Final Grade (G3)

In [67]:
```
average_grade = df['G3'].mean()
print(f"Average Final Grade (G3): {average_grade:.2f}")
```

```
Average Final Grade (G3): 10.42
```

2. How many students scored above 15?

In [72]:
```
num_above_15 = (df['G3'] > 15).sum()
print(f"Number of students scoring above 15: {num_above_15}")
```

```
Number of students scoring above 15: 40
```

3. Is study time correlated with performance?

In [75]:
```
correlation = df[['studytime','G3']].corr()
print("\nCorrelation matrix:\n", correlation)
```

```
Correlation matrix:
              studytime       G3
studytime     1.00000  0.09782
G3            0.09782  1.00000
```

4. Which gender performs better on average?

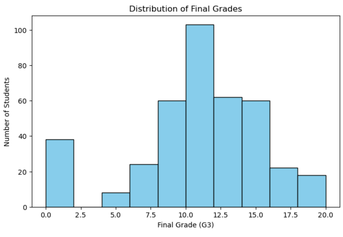
In [78]:
```
gender_performance = df.groupby('sex')['G3'].mean()
print("\nAverage Final Grade by Gender:\n", gender_performance)
```

```
Average Final Grade by Gender:
 sex
F    9.966346
M   10.914439
Name: G3, dtype: float64
```

Visualizations

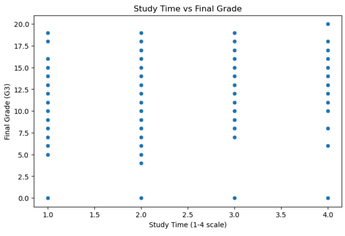
Histogram of grades

In [82]:
```
plt.figure(figsize=(8,5))
plt.hist(df['G3'], bins=10, color='skyblue', edgecolor='black')
plt.xlabel('Final Grade (G3)')
plt.ylabel('Number of Students')
plt.title('Distribution of Final Grades')
plt.show()
```



Scatterplot: Study time vs Final Grade

In [94]:
```
plt.figure(figsize=(8,5))
sns.scatterplot(x='studytime', y='G3', data=df)
plt.title('Study Time vs Final Grade')
plt.xlabel("Study Time (1-4 scale)")
plt.ylabel('Final Grade (G3)')
plt.show()
```



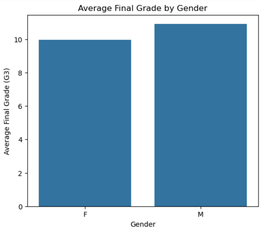
Bar chart: Male vs Female Average Score

In [97]:
```
plt.figure(figsize=(6,5))
sns.barplot(x='sex', y='G3', data=df, ci=None )
plt.title('Average Final Grade by Gender')
plt.xlabel('Gender')
plt.ylabel('Average Final Grade (G3)')
plt.show()
```

```
C:\Users\Akshay Besekar\AppData\Local\Temp\ipykernel_3612\4020269772.py:2: FutureWarning:

The `ci` parameter is deprecated. Use `errorbar=None` for the same effect.

  sns.barplot(x='sex', y='G3', data=df, ci=None )
```



Conclusion

In [ ]:
```
- The average final grade is 10.42.
- The number of students scoring above 15 is 40.
- Study time has a correlation with final grade is :
Correlation matrix:
              studytime       G3
studytime     1.00000  0.09782
G3            0.09782  1.00000
- Males and females' average scores are compared in the bar chart:
```

```
Average Final Grade by Gender:
 sex
F     9.966346
M    10.914439
```