Lab-5

Correlation Analysis

- Name: Daniel Mehta
- Student ID: n01753264

Import Libraries

In [14]:
```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

In [2]:
```

```

Import Titanic dataset

In [2]:
```
fp = 'Titanic.csv'
df = pd.read_csv(fp)
```

Read head of the dataset

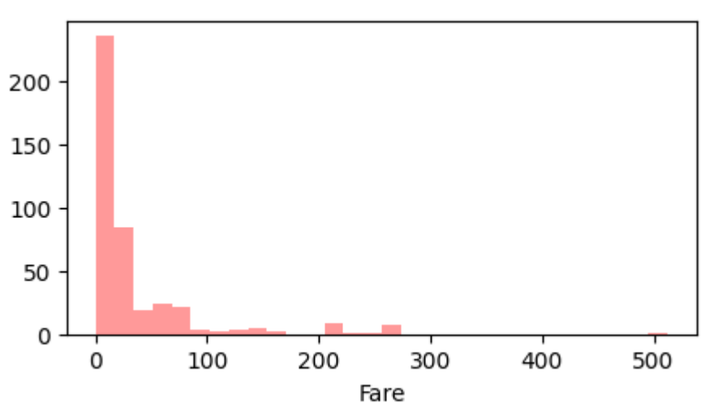
In [3]:
```
df.head()
```

Out[3]:

	PassengerId	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	892	3	Kelly, Mr. James	male	34.5	0	0	330911	7.8292	NaN	Q
1	893	3	Wilkes, Mrs. James (Ellen Needs)	female	47.0	1	0	363272	7.0000	NaN	S
2	894	2	Myles, Mr. Thomas Francis	male	62.0	0	0	240276	9.6875	NaN	Q
3	895	3	Wirz, Mr. Albert	male	27.0	0	0	315154	8.6625	NaN	S
4	896	3	Hirvonen, Mrs. Alexander (Helga E Lindqvist)	female	22.0	1	1	3101298	12.2875	NaN	S

Exercise 1

In [8]:
```
# CODE HERE
plt.figure(figsize=(5, 2.5))
plt.hist(df["Fare"].dropna(), bins=30, color="red", alpha=0.4)
plt.xlabel("Fare")
plt.show()
```



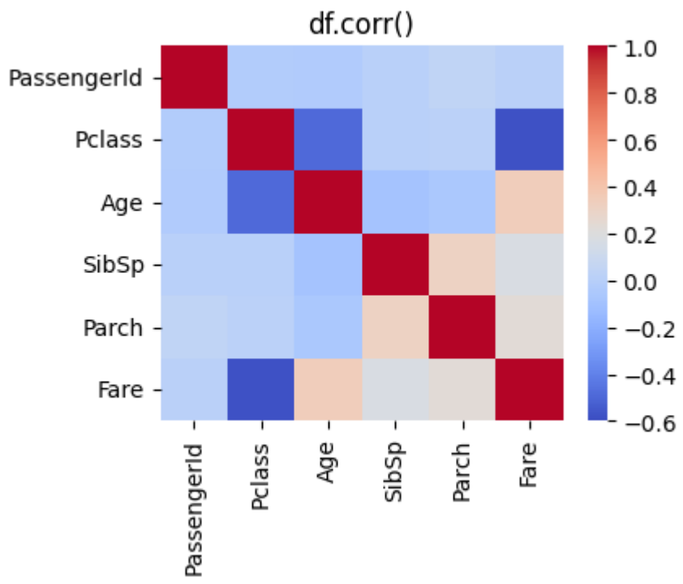
Exercise 2

In [70]:
```
# CODE HERE
numeric_df = df.select_dtypes(include=["number"])
corr_matrix = numeric_df.corr()
plt.figure(figsize=(4, 3))
sns.heatmap(corr_matrix, annot=False, cmap="coolwarm",vmin=-0.6,vmax=1, center=0.2)

plt.xticks(rotation=90)
plt.yticks(rotation=0)

plt.title("df.corr()")

plt.show()
```



Exercise 3

Find "Pearson correlation" and "Spearman correlation" between "Age" and "Parch" column?

In [73]:
```
# CODE HERE
pearson_corr = df[['Age', 'Parch']].corr(method='pearson').iloc[0, 1]
spearman_corr = df[['Age', 'Parch']].corr(method='spearman').iloc[0, 1]
pearson_corr, spearman_corr

print(f"Pearson Correlation: {pearson_corr}")
print(f"Spearman correlation: {spearman_corr}")
```

```
Pearson Correlation: -0.06124863292998425
Spearman correlation: -0.1304079630779184
```

Exercise 4

Calculate the standard deviation, variance and mean of column "Fare" and "Age"

In [78]:
```
# CODE HERE
stats = df[['Fare', 'Age']].agg(['std', 'var', 'mean'])
stats
print(stats)
```

```
             Fare         Age
std     55.907576   14.181209
var   3125.657074  201.106695
mean    35.627188   30.272590
```

In [ ]: