In [ ]:
```
# QUESTION - 1

import pandas as pd

us_cereal_df = pd.read_csv("UScereal.csv")

print(us_cereal_df)
```

```
                         Name mfr  calories  protein   fat  sodium  fibre  \
0                   100% Bran   N    212.12    12.12  3.03  393.94  30.30
1                    All-Bran   K    212.12    12.12  3.03  787.88  27.27
2   All-Bran with Extra Fiber   K    100.00     8.00  0.00  280.00  28.00
3     Apple Cinnamon Cheerios   G    146.67     2.67  2.67  240.00   2.00
4                  Apple Jacks   K    110.00     2.00  0.00  125.00   1.00
..                         ...  ..       ...      ...   ...     ...    ...
60                     Triples   G    146.67     2.67  1.33  333.33   0.00
61                        Trix   G    110.00     1.00  1.00  140.00   0.00
62                  Wheat Chex   R    149.25     4.48  1.49  343.28   4.48
63                    Wheaties   G    100.00     3.00  1.00  200.00   3.00
64         Wheaties Honey Gold   G    146.67     2.67  1.33  266.67   1.33

    carbo  sugars  shelf  potassium  vitamins
0   15.15   18.18      3     848.48  enriched
1   21.21   15.15      3     969.70  enriched
2   16.00    0.00      3     660.00  enriched
3   14.00   13.33      1      93.33  enriched
4   11.00   14.00      2      30.00  enriched
..    ...     ...    ...        ...       ...
60  28.00    4.00      3      80.00  enriched
61  13.00   12.00      2      25.00  enriched
62  25.37    4.48      1     171.64  enriched
63  17.00    3.00      1     110.00  enriched
64  21.33   10.67      1      80.00  enriched

[65 rows x 12 columns]
```

In [ ]:
```
import pandas as pd

us_cereal_df = pd.read_csv("UScereal.csv")

summary_statistics = us_cereal_df.describe()

missing_values = us_cereal_df.isnull().sum()

print("Summary Statistics:")
print(summary_statistics)

print("\nMissing Values:")
print(missing_values)
```

```
Summary Statistics:
         calories    protein        fat      sodium      fibre      carbo  \
count   65.000000  65.000000  65.000000   65.000000  65.000000  65.000000
mean   149.408615   3.684000   1.422462  237.838308   3.870923  19.967538
std     62.411936   2.642821   1.647561  130.629537   6.133094   8.468190
min     50.000000   0.750000   0.000000    0.000000   0.000000  10.530000
25%    110.000000   2.000000   0.000000  180.000000   0.000000  15.000000
50%    134.330000   3.000000   1.000000  232.000000   2.000000  18.670000
75%    179.100000   4.480000   2.000000  290.000000   4.480000  22.390000
max    440.000000  12.120000   9.090000  787.880000  30.300000  68.000000

          sugars      shelf   potassium
count  65.000000  65.000000   65.000000
mean   10.051077   2.169231  159.119692
std     5.835252   0.839815  180.288575
min     0.000000   1.000000   15.000000
25%     4.000000   1.000000   45.000000
50%    12.000000   2.000000   96.590000
75%    14.000000   3.000000  220.000000
max    20.900000   3.000000  969.700000

Missing Values:
Name          0
mfr           0
calories      0
protein       0
fat           0
sodium        0
fibre         0
carbo         0
sugars        0
shelf         0
potassium     0
vitamins      0
dtype: int64
```

In [ ]:
```python
import pandas as pd

us_cereal_df = pd.read_csv("UScereal.csv")

average_protein_by_manufacturer = us_cereal_df.groupby('mfr')['protein'].mean()

print(average_protein_by_manufacturer)
```

```
mfr
G    2.885000
K    3.919048
N    7.026667
P    4.698889
Q    3.460000
R    2.604000
Name: protein, dtype: float64
```

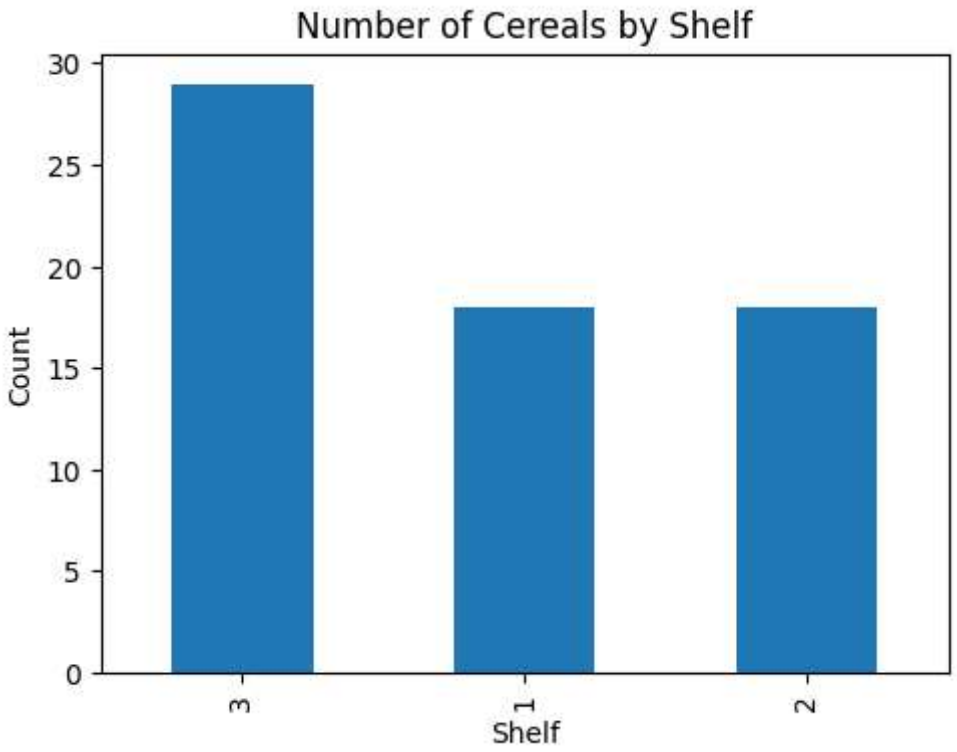
In [ ]:
```python
import pandas as pd

us_cereal_df = pd.read_csv("UScereal.csv")
```

```
cereals_from_manufacturer_G = us_cereal_df[us_cereal_df['mfr'] == 'G']

highest_sugar_cereal = cereals_from_manufacturer_G[cereals_from_manufacturer_G['sug

name_of_highest_sugar_cereal = highest_sugar_cereal['Name'].values[0]

print(f"The cereal with the highest sugar content from Manufacturer G is: {name_of_
```

```
The cereal with the highest sugar content from Manufacturer G is: Oatmeal Raisin Cri
sp
```

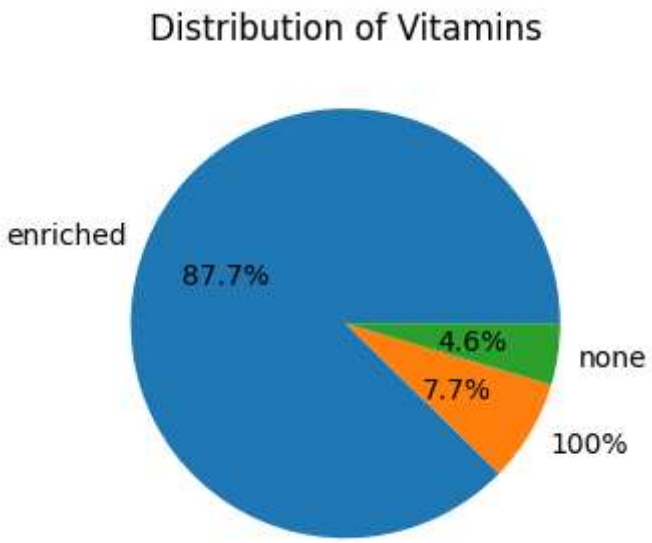
In [ ]:
```
import pandas as pd
import matplotlib.pyplot as plt

us_cereal_df = pd.read_csv("UScereal.csv")
plt.figure(figsize=(12, 4))

# Subplot 1: Bar plot for "shelf"
plt.subplot(1, 2, 1)
us_cereal_df['shelf'].value_counts().plot(kind='bar')
plt.title('Number of Cereals by Shelf')
plt.xlabel('Shelf')
plt.ylabel('Count')
```

Out[ ]:
```
Text(0, 0.5, 'Count')
```



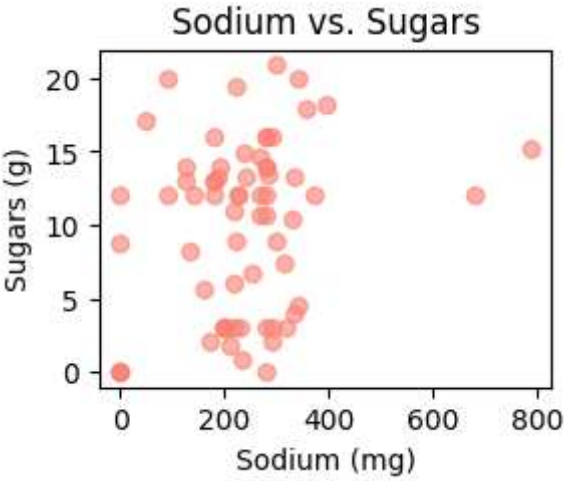
In [ ]:
```
plt.subplot(1, 2, 2)
us_cereal_df['vitamins'].value_counts().plot(kind='pie', autopct='%1.1f%%')
plt.title('Distribution of Vitamins')
plt.ylabel('')
```

```
plt.tight_layout()
plt.show()
```



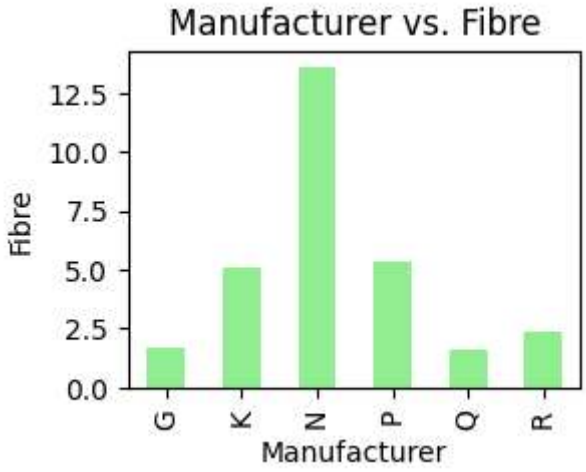
In [ ]:
```
plt.subplot(2, 2, 2)
plt.scatter(df['sodium'], df['sugars'], color='salmon', alpha=0.6)
plt.title("Sodium vs. Sugars")
plt.xlabel("Sodium (mg)")
plt.ylabel("Sugars (g)")
```

Out[ ]:
```
Text(0, 0.5, 'Sugars (g)')
```



In [ ]:
```
plt.subplot(2, 2, 2)
mfr_fiber = df.groupby(['mfr'])['fibre'].mean().plot(kind='bar', color='lightgreen'
plt.title("Manufacturer vs. Fibre")
plt.xlabel("Manufacturer")
plt.ylabel("Fibre")
```

Out[ ]:
```
Text(0, 0.5, 'Fibre')
```


Manufacturer vs. Fibre
Fibre
12.5
10.0
7.5
5.0
2.5
0.0
G
K
N
P
Q
R
Manufacturer