In [31]:
```
import seaborn as sns

df = sns.load_dataset('mpg')
df
```

Out[31]:

| | mpg | cylinders | displacement | horsepower | weight | acceleration | model_year | origin |
|---|---|---|---|---|---|---|---|---|
| **0** | 18.0 | 8 | 307.0 | 130.0 | 3504 | 12.0 | 70 | usa |
| **1** | 15.0 | 8 | 350.0 | 165.0 | 3693 | 11.5 | 70 | usa |
| **2** | 18.0 | 8 | 318.0 | 150.0 | 3436 | 11.0 | 70 | usa |
| **3** | 16.0 | 8 | 304.0 | 150.0 | 3433 | 12.0 | 70 | usa |
| **4** | 17.0 | 8 | 302.0 | 140.0 | 3449 | 10.5 | 70 | usa |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... |
| **393** | 27.0 | 4 | 140.0 | 86.0 | 2790 | 15.6 | 82 | usa |
| **394** | 44.0 | 4 | 97.0 | 52.0 | 2130 | 24.6 | 82 | europe |
| **395** | 32.0 | 4 | 135.0 | 84.0 | 2295 | 11.6 | 82 | usa |
| **396** | 28.0 | 4 | 120.0 | 79.0 | 2625 | 18.6 | 82 | usa |
| **397** | 31.0 | 4 | 119.0 | 82.0 | 2720 | 19.4 | 82 | usa |

398 rows × 9 columns

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

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 398 entries, 0 to 397
Data columns (total 9 columns):
 #   Column        Non-Null Count  Dtype  
---  ------        --------------  -----  
 0   mpg           398 non-null    float64
 1   cylinders     398 non-null    int64  
 2   displacement  398 non-null    float64
 3   horsepower    392 non-null    float64
 4   weight        398 non-null    int64  
 5   acceleration  398 non-null    float64
 6   model_year    398 non-null    int64  
 7   origin        398 non-null    object 
 8   name          398 non-null    object 
dtypes: float64(4), int64(3), object(2)
memory usage: 28.1+ KB
```

In [33]:
```
from scipy.stats import skew, kurtosis

skewness = df[['mpg', 'cylinders', 'displacement', 'horsepower', 'weight', 'acceler
skewness
```

Out[33]:
```
mpg             0.455342
cylinders       0.524934
displacement    0.716930
horsepower           NaN
weight          0.529059
acceleration    0.277725
model_year      0.011491
dtype: float64
```

In [34]:
```
kurt = df[['mpg', 'cylinders', 'displacement', 'horsepower', 'weight', 'acceleratio
kurt
```

Out[34]:
```
mpg            -0.519425
cylinders      -1.374466
displacement   -0.752288
horsepower           NaN
weight         -0.790733
acceleration    0.399208
model_year     -1.181482
dtype: float64
```

In [35]:
```
import pandas as pd

df['origin'] = df['origin'].astype('category').cat.codes
df['name'] = df['name'].astype('category').cat.codes
df
```

Out[35]:

| | mpg | cylinders | displacement | horsepower | weight | acceleration | model_year | origin |
|---|---|---|---|---|---|---|---|---|
| **0** | 18.0 | 8 | 307.0 | 130.0 | 3504 | 12.0 | 70 | 2 |
| **1** | 15.0 | 8 | 350.0 | 165.0 | 3693 | 11.5 | 70 | 2 |
| **2** | 18.0 | 8 | 318.0 | 150.0 | 3436 | 11.0 | 70 | 2 |
| **3** | 16.0 | 8 | 304.0 | 150.0 | 3433 | 12.0 | 70 | 2 |
| **4** | 17.0 | 8 | 302.0 | 140.0 | 3449 | 10.5 | 70 | 2 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... |
| **393** | 27.0 | 4 | 140.0 | 86.0 | 2790 | 15.6 | 82 | 2 |
| **394** | 44.0 | 4 | 97.0 | 52.0 | 2130 | 24.6 | 82 | 0 |
| **395** | 32.0 | 4 | 135.0 | 84.0 | 2295 | 11.6 | 82 | 2 |
| **396** | 28.0 | 4 | 120.0 | 79.0 | 2625 | 18.6 | 82 | 2 |
| **397** | 31.0 | 4 | 119.0 | 82.0 | 2720 | 19.4 | 82 | 2 |

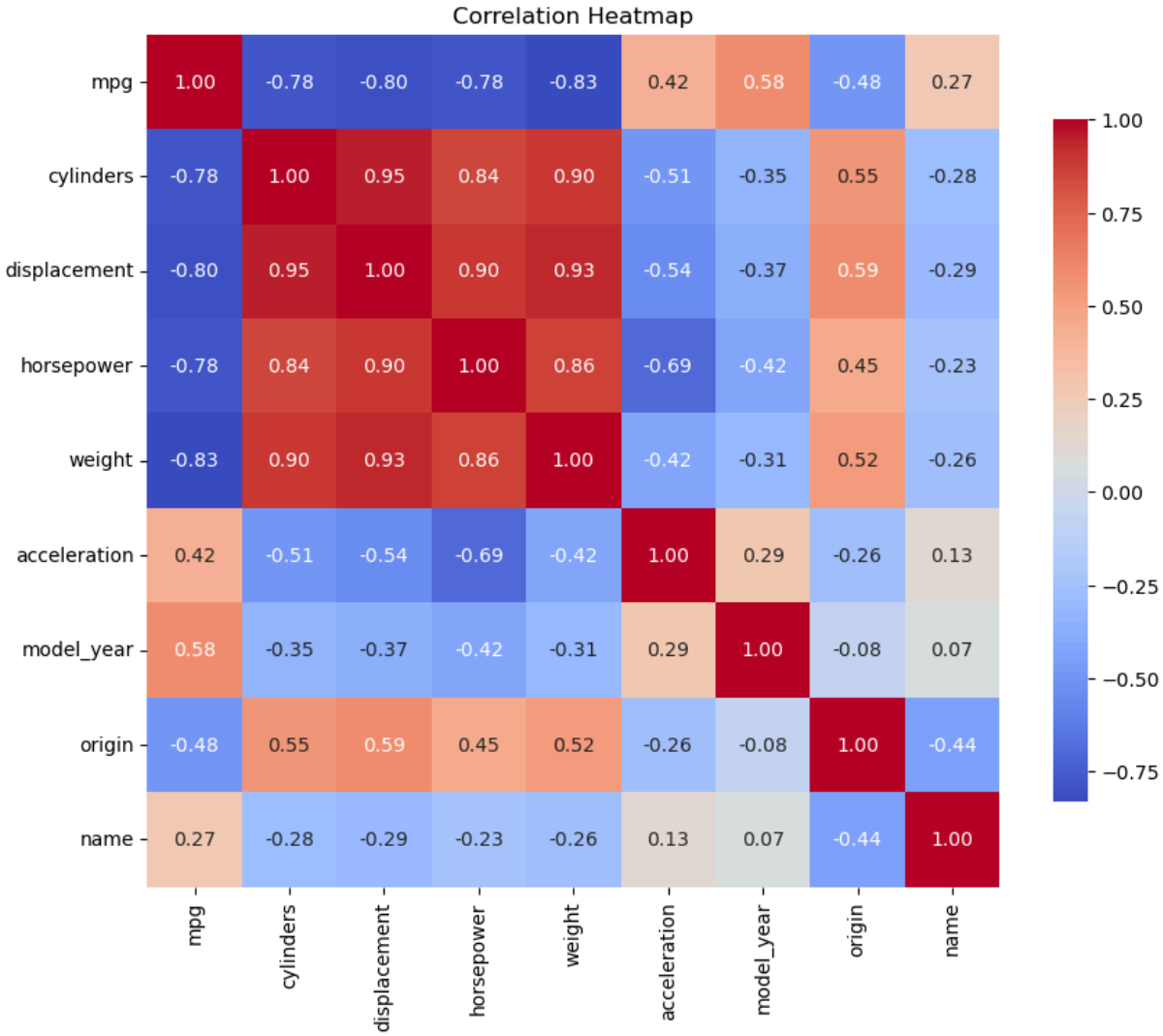
398 rows × 9 columns

In [37]:
```
correlation_matrix = df.corr()
correlation_matrix
```

Out[37]:

| | mpg | cylinders | displacement | horsepower | weight | acceleration | mo |
|---|---|---|---|---|---|---|---|
| **mpg** | 1.000000 | -0.775396 | -0.804203 | -0.778427 | -0.831741 | 0.420289 | ( |
| **cylinders** | -0.775396 | 1.000000 | 0.950721 | 0.842983 | 0.896017 | -0.505419 | -( |
| **displacement** | -0.804203 | 0.950721 | 1.000000 | 0.897257 | 0.932824 | -0.543684 | -( |
| **horsepower** | -0.778427 | 0.842983 | 0.897257 | 1.000000 | 0.864538 | -0.689196 | -( |
| **weight** | -0.831741 | 0.896017 | 0.932824 | 0.864538 | 1.000000 | -0.417457 | -( |
| **acceleration** | 0.420289 | -0.505419 | -0.543684 | -0.689196 | -0.417457 | 1.000000 | ( |
| **model_year** | 0.579267 | -0.348746 | -0.370164 | -0.416361 | -0.306564 | 0.288137 | |
| **origin** | -0.482619 | 0.551378 | 0.591137 | 0.447330 | 0.521088 | -0.257365 | -( |
| **name** | 0.273936 | -0.275754 | -0.292064 | -0.234663 | -0.255247 | 0.128285 | ( |

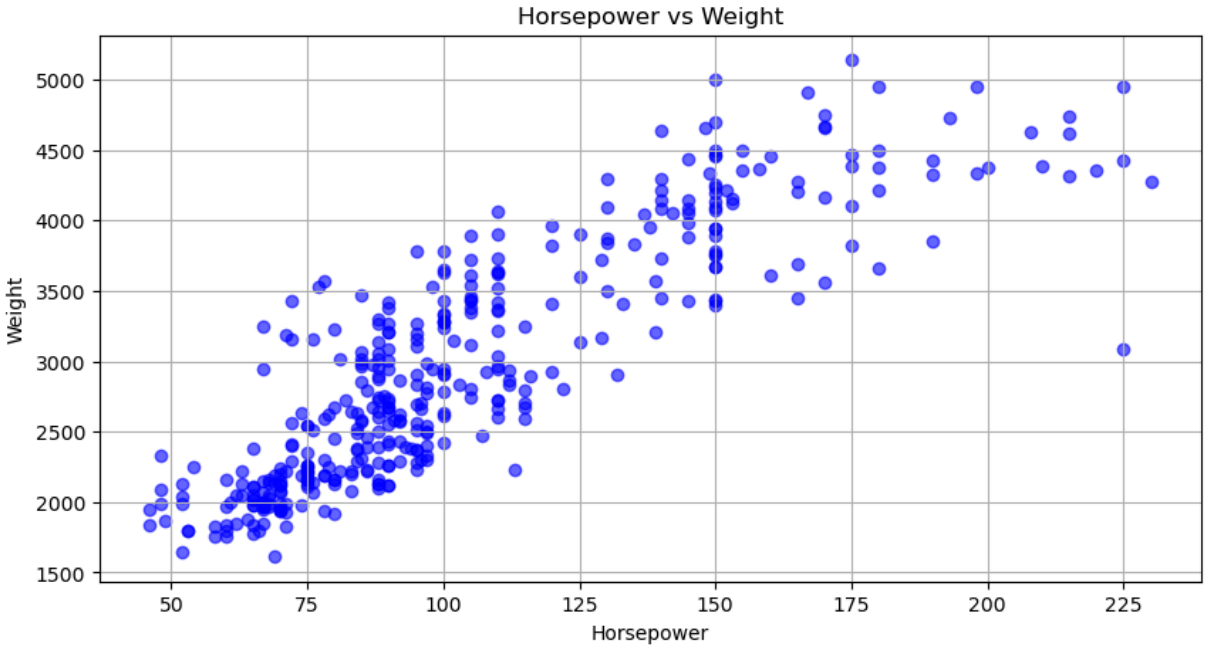
In [38]:
```
plt.figure(figsize=(10, 8))
sns.heatmap(correlation_matrix, annot=True, fmt=".2f", cmap='coolwarm', square=True
plt.title('Correlation Heatmap')
plt.show()
```



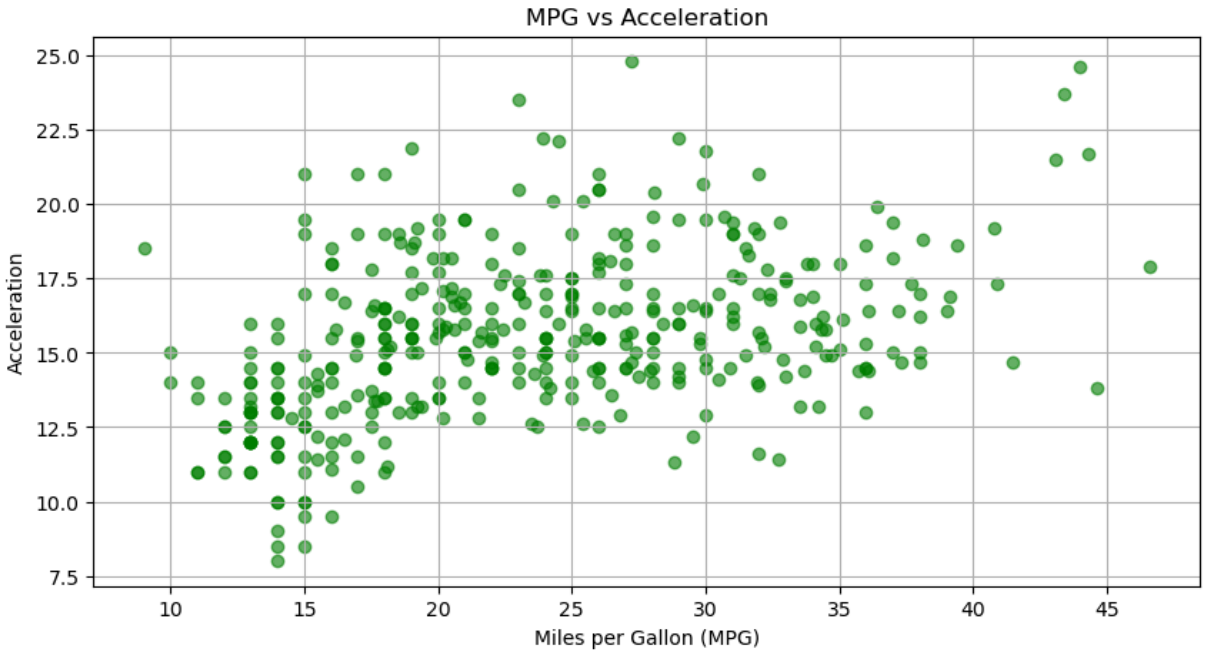

In [40]:

```python
import matplotlib.pyplot as plt

#Scatterplot: Horsepower vs Weight
plt.figure(figsize=(10, 5))
plt.scatter(df['horsepower'], df['weight'], color='blue', alpha=0.6)
plt.title('Horsepower vs Weight')
plt.xlabel('Horsepower')
plt.ylabel('Weight')
plt.grid(True)
plt.show()
```

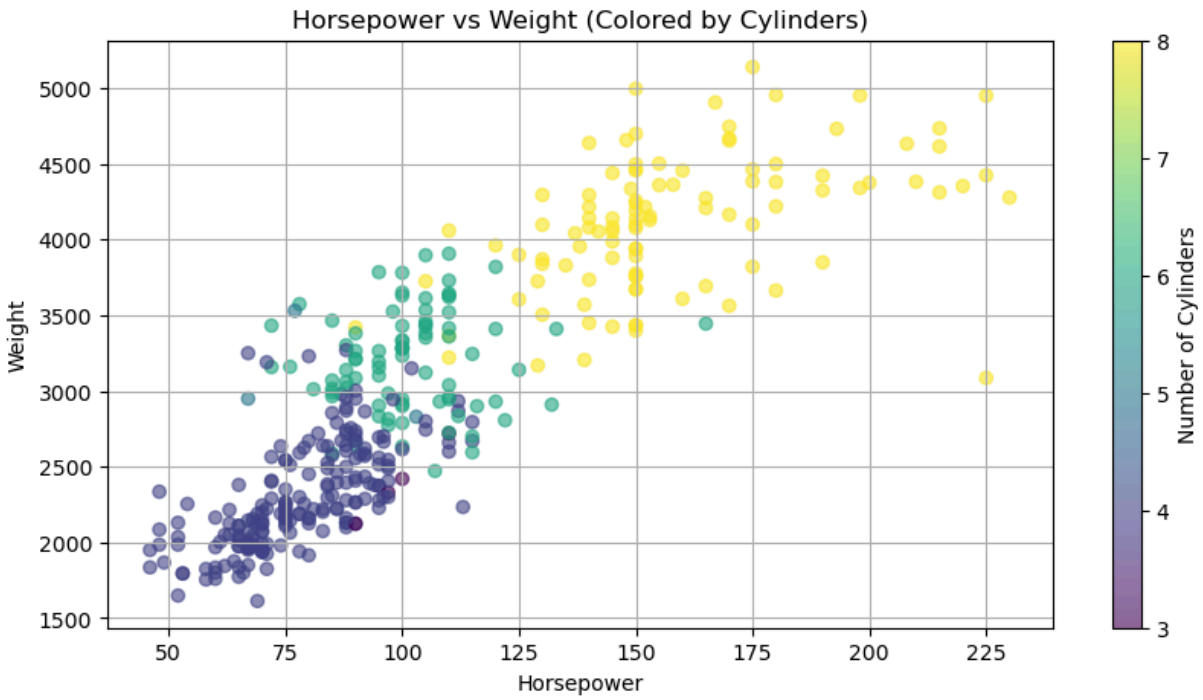



In [41]:
```
#Scatterplot: MPG vs Acceleration
plt.figure(figsize=(10, 5))
plt.scatter(df['mpg'], df['acceleration'], color='green', alpha=0.6)
plt.title('MPG vs Acceleration')
plt.xlabel('Miles per Gallon (MPG)')
plt.ylabel('Acceleration')
plt.grid(True)
plt.show()
```




In [43]:
```
#Scatterplot: Horsepower vs Weight colored by number of cylinders
plt.figure(figsize=(10, 5))
scatter = plt.scatter(df['horsepower'], df['weight'], c=df['cylinders'], cmap='viri
plt.title('Horsepower vs Weight (Colored by Cylinders)')
plt.xlabel('Horsepower')
plt.ylabel('Weight')
```

```
plt.colorbar(scatter, label='Number of Cylinders')
plt.grid(True)
plt.show()
```




In [ ]: