Data visualization

Problem Statement:

Visualizing data can be quite insightful for further analysis and narrowing on target problems. You are given sample data, after doing the basic analysis using pandas, visualize the outcomes to find insights and patterns in the data. Use the cars dataset for the following questions that contains the following The dataset contains information about 260 cars that include horsepower, cubic inches, time to 60, brand, make year, weight, cylinders, mpg, etc.

jupyter Untitled1 Last Checkpoint: 6 minutes ago

File Edit View Run Kernel Settings Help Trusted

Code JupyterLab Python 3 (ipykernel)

[2]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
df=pd.read_csv(r"C:\Users\Shanawaz khan\Downloads\cars (1).csv")
df
```

[2]:

	mpg	cylinders	cubicinches	hp	weightlbs	time-to-60	year	brand
0	14.0	8	350	165	4209	12	1972	US.
1	31.9	4	89	71	1925	14	1980	Europe.
2	17.0	8	302	140	3449	11	1971	US.
3	15.0	8	400	150	3761	10	1971	US.
4	30.5	4	98	63	2051	17	1978	US.
...	...	...	...	...	...	...	...	...
256	17.0	8	305	130	3840	15	1980	US.
257	36.1	4	91	60	1800	16	1979	Japan.
258	22.0	6	232	112	2835	15	1983	US.
259	18.0	6	232	100	3288	16	1972	US.
260	22.0	6	250	105	3353	15	1977	US.

261 rows × 8 columns

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

jupyter Untitled1 Last Checkpoint: 6 minutes ago

File Edit View Run Kernel Settings Help Trusted

Code JupyterLab Python 3 (ipykernel)

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

[3]:
```
(261, 8)
```

[5]:
```
df.head()
```

[5]:

	mpg	cylinders	cubicinches	hp	weightlbs	time-to-60	year	brand
0	14.0	8	350	165	4209	12	1972	US.
1	31.9	4	89	71	1925	14	1980	Europe.
2	17.0	8	302	140	3449	11	1971	US.
3	15.0	8	400	150	3761	10	1971	US.
4	30.5	4	98	63	2051	17	1978	US.

[6]:
```
df.tail()
```

[6]:

	mpg	cylinders	cubicinches	hp	weightlbs	time-to-60	year	brand
256	17.0	8	305	130	3840	15	1980	US.
257	36.1	4	91	60	1800	16	1979	Japan.
258	22.0	6	232	112	2835	15	1983	US.
259	18.0	6	232	100	3288	16	1972	US.
260	22.0	6	250	105	3353	15	1977	US.

[7]:
```
df.describe()
```

jupyter Untitled1 Last Checkpoint: 6 minutes ago

File Edit View Run Kernel Settings Help Truste

Code JupyterLab Python 3 (ipykernel)

[7]:
```
df.describe()
```

[7]:

	mpg	cylinders	hp	time-to-60	year
count	261.000000	261.000000	261.000000	261.000000	261.000000
mean	23.144828	5.590038	106.360153	15.547893	1976.819923
std	7.823570	1.733310	40.499959	2.910625	3.637696
min	10.000000	3.000000	46.000000	8.000000	1971.000000
25%	16.900000	4.000000	75.000000	14.000000	1974.000000
50%	22.000000	6.000000	95.000000	16.000000	1977.000000
75%	28.800000	8.000000	138.000000	17.000000	1980.000000
max	46.600000	8.000000	230.000000	25.000000	1983.000000

[8]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 261 entries, 0 to 260
Data columns (total 8 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   mpg          261 non-null    float64
 1   cylinders    261 non-null    int64
 2   cubicinches  261 non-null    object
 3   hp           261 non-null    int64
 4   weightlbs    261 non-null    object
 5   time-to-60   261 non-null    int64
 6   year         261 non-null    int64
```

jupyter Untitled1 Last Checkpoint: 7 minutes ago

File Edit View Run Kernel Settings Help Trusted

Code JupyterLab Python 3 (ipykernel)

```
memory usage: 16.4+ KB
```

[10]:
```
df.isnull().sum()
```

[10]:
```
mpg           0
cylinders     0
cubicinches   0
hp            0
weightlbs     0
time-to-60    0
year          0
brand         0
dtype: int64
```

[12]:
```
df["brand"].value_counts()
```

[12]:
```
brand
 US.        162
 Japan.      51
 Europe.     48
Name: count, dtype: int64
```

[13]:
```
df["cubicinches"]=pd.to_numeric(df["cubicinches"],errors="coerce")
df["weightlbs"]=pd.to_numeric(df["weightlbs"],errors="coerce")
```

[14]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 261 entries, 0 to 260
Data columns (total 8 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   mpg          261 non-null    float64
 1   cylinders    261 non-null    int64
 2   cubicinches  259 non-null    float64
 3   hp           261 non-null    int64
```

jupyter Untitled1 Last Checkpoint: 7 minutes ago

File Edit View Run Kernel Settings Help Truste

Code JupyterLab Python 3 (ipykernel)

```
 6   year          261 non-null    int64
 7   brand         261 non-null    object
dtypes: float64(3), int64(4), object(1)
memory usage: 16.4+ KB
```

[15]:
```
df.isnull().sum()
```

[15]:
```
mpg           0
cylinders     0
cubicinches   2
hp            0
weightlbs     3
time-to-60    0
year          0
brand         0
dtype: int64
```

[16]:
```
df=df.dropna()
```

[18]:
```
df.isnull().sum()
```

[18]:
```
mpg           0
cylinders     0
cubicinches   0
hp            0
weightlbs     0
time-to-60    0
year          0
brand         0
dtype: int64
```

[19]:
```
df.shape
```

[19]:
```
(256, 8)
```

Tasks To Be Performed:

1. Create scatter plots for the following data, make sure all the plots appear in the same plane with x-labels and y-labels evenly spaced.

a. Create scatter plots with between (mpg-hp, mpg-weight), and (hp-mpg, hp-time-to-60) with respect to each brand.

jupyter Untitled1 Last Checkpoint: 4 minutes ago

File Edit View Run Kernel Settings Help Trusted

Code JupyterLab Python 3 (ipykernel)

•[24]:
```
fig=plt.subplots(2,2,figsize=(20,10))
plt.subplot(2,2,1)
sns.scatterplot(x=df["mpg"],y=df["hp"],data=df)
plt.title("mpg X hp")
plt.xlabel("mpg")
plt.ylabel("hp")

plt.subplot(2,2,2)
sns.scatterplot(x=df["mpg"],y=df["weightlbs"],data=df)
plt.title("mpg X weightlbs")
plt.xlabel("mpg")
plt.ylabel("weightlbs")
plt.subplot(2,2,3)

sns.scatterplot(x=df["hp"],y=df["mpg"],data=df)
plt.title("hp X mpg")
plt.xlabel("hp")
plt.ylabel("mpg")

plt.subplot(2,2,4)
sns.scatterplot(x=df["hp"],y=df["time-to-60"],data=df)
plt.title("hp X time-to-60")
plt.xlabel("hp")
plt.ylabel("time-to-60")
```

[24]:
```
Text(0, 0.5, 'time-to-60')
```

jupyter Untitled1 Last Checkpoint: 4 minutes ago

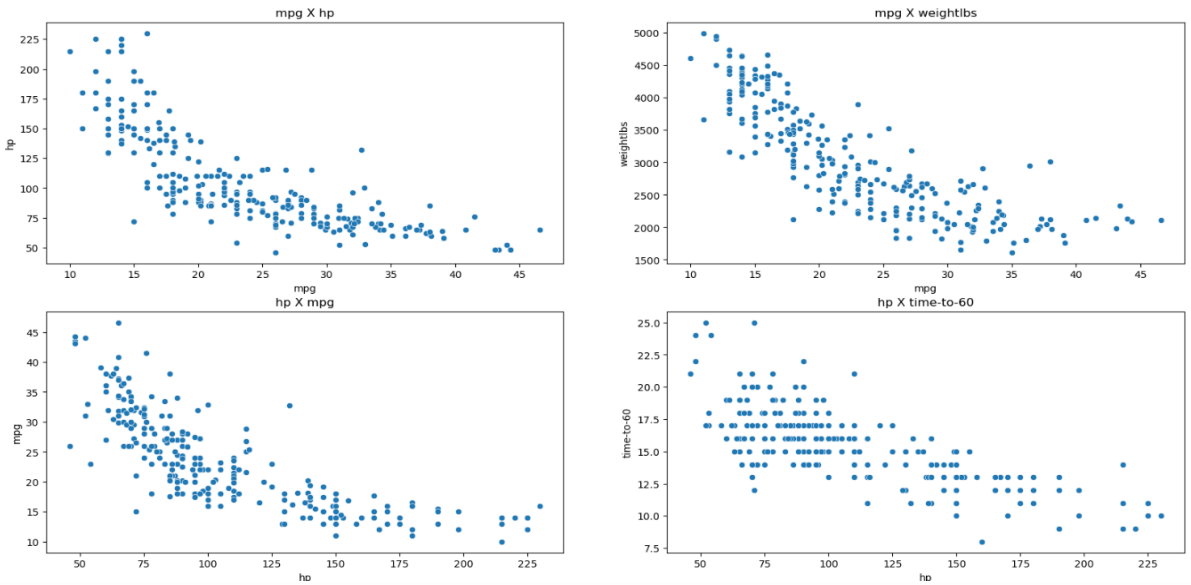
File Edit View Run Kernel Settings Help

Trusted

Code

JupyterLab Python 3 (ipykernel)

[24]: Text(0, 0.5, 'time-to-60')



jupyter Untitled1 Last Checkpoint: 5 minutes ago

File Edit View Run Kernel Settings Help

Trusted

Code

JupyterLab Python 3 (ipykernel)

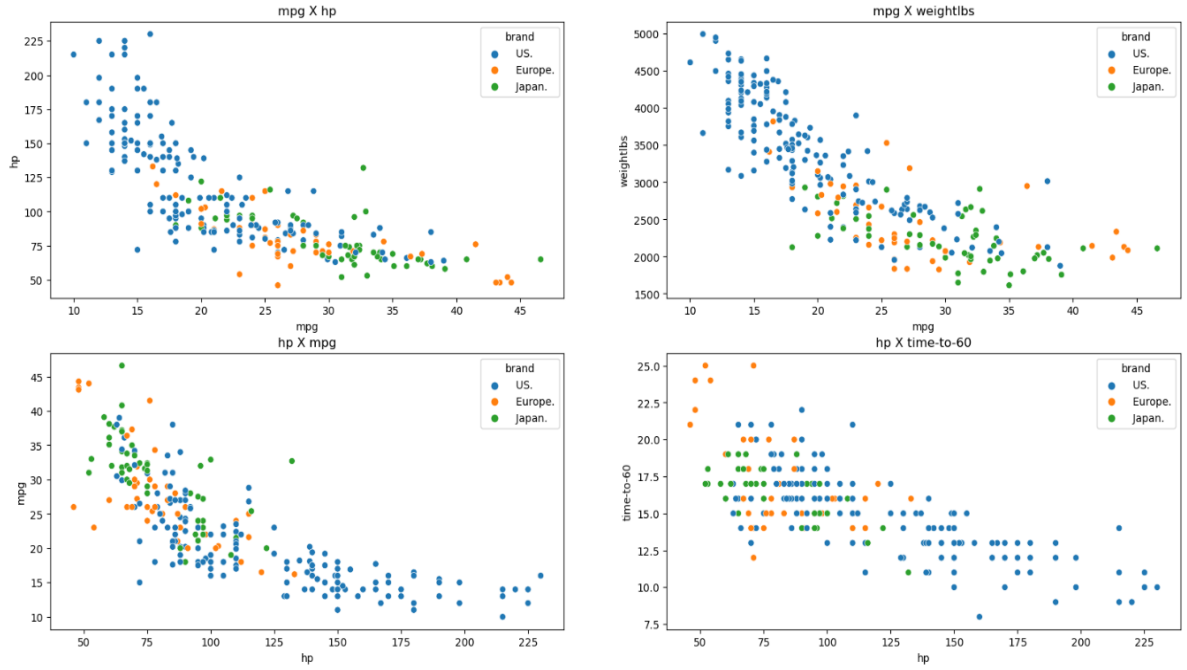
[25]:
```
fig=plt.subplots(2,2,figsize=(20,10))
plt.subplot(2,2,1)
sns.scatterplot(x=df["mpg"],y=df["hp"],data=df,hue=df["brand"])
plt.title("mpg X hp")
plt.xlabel("mpg")
plt.ylabel("hp")

plt.subplot(2,2,2)
sns.scatterplot(x=df["mpg"],y=df["weightlbs"],data=df,hue=df["brand"])
plt.title("mpg X weightlbs")
plt.xlabel("mpg")
plt.ylabel("weightlbs")
plt.subplot(2,2,3)

sns.scatterplot(x=df["hp"],y=df["mpg"],data=df,hue=df["brand"])
plt.title("hp X mpg")
plt.xlabel("hp")
plt.ylabel("mpg")

plt.subplot(2,2,4)
sns.scatterplot(x=df["hp"],y=df["time-to-60"],data=df,hue=df["brand"])
plt.title("hp X time-to-60")
plt.xlabel("hp")
plt.ylabel("time-to-60")
```

jupyter Untitled1 Last Checkpoint: 5 minutes ago

File Edit View Run Kernel Settings Help Trusted

Code JupyterLab Python 3 (ipykernel)

```
[25]: text(0, 0.5, 'time-to-60')
```



1. Create bar plots for the following, make sure all the plots appear in the same plane with x-labels and y-labels evenly spaced.

a. Create a bar plot that shows the visual representation of hp, mpg, weight, time-to-60 with respect to the number of cylinders for each of the brands.

b. Create a bar plot that shows the visual representation of hp, mpg, weight and time-to-60 with respect to the years, for each brand.

jupyter Untitled1 Last Checkpoint: 4 minutes ago

File Edit View Run Kernel Settings Help Trusted

Code JupyterLab Python 3 (ipykernel)

[28]:
```
fig=plt.subplots(2,2,figsize=(20,10))
plt.subplot(2,2,1)
sns.barplot(x=df["cylinders"],y=df["hp"],data=df,hue=df["brand"])
plt.title("cylinders X hp")
plt.xlabel("cylinders")
plt.ylabel("hp")

plt.subplot(2,2,2)
sns.barplot(x=df["cylinders"],y=df["mpg"],data=df,hue=df["brand"])
plt.title("cylinders X mpg")
plt.xlabel("cylinders")
plt.ylabel("mpg")

plt.subplot(2,2,3)
sns.barplot(x=df["cylinders"],y=df["weightlbs"],data=df,hue=df["brand"])
plt.title("cylinders X weightlbs")
plt.xlabel("cylinders")
plt.ylabel("weightlbs")

plt.subplot(2,2,4)
sns.barplot(x=df["cylinders"],y=df["time-to-60"],data=df,hue=df["brand"])
plt.title("cylinders X time-to-60")
plt.xlabel("cylinders")
plt.ylabel("time-to-60")
```

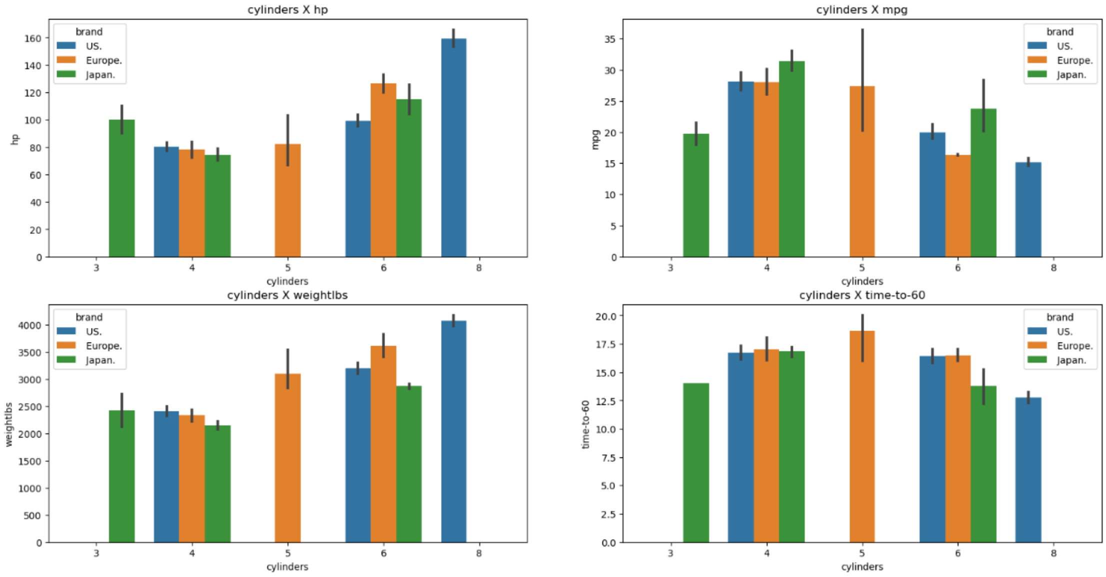
jupyter Untitled1 Last Checkpoint: 5 minutes ago

File Edit View Run Kernel Settings Help Trusted

Code JupyterLab Python 3 (ipykernel)

```
plt.xlabel("cylinders")
plt.ylabel("time-to-60")
```

[28]: Text(0, 0.5, 'time-to-60')



1. Create pair plots for the entire data to study various patterns in the data.

a. Create pair plots with respect to brand, number of cylinders, year, etc

jupyter Untitled1 Last Checkpoint: 7 minutes ago

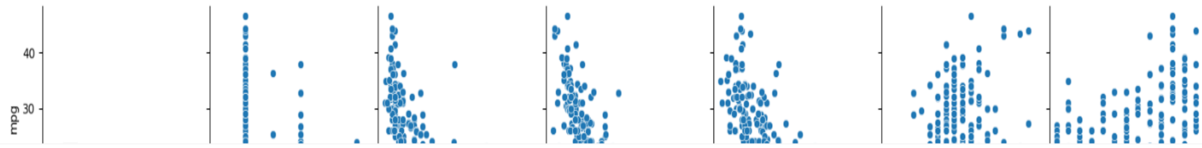
File Edit View Run Kernel Settings Help Trusted

Code JupyterLab Python 3 (ipykernel)

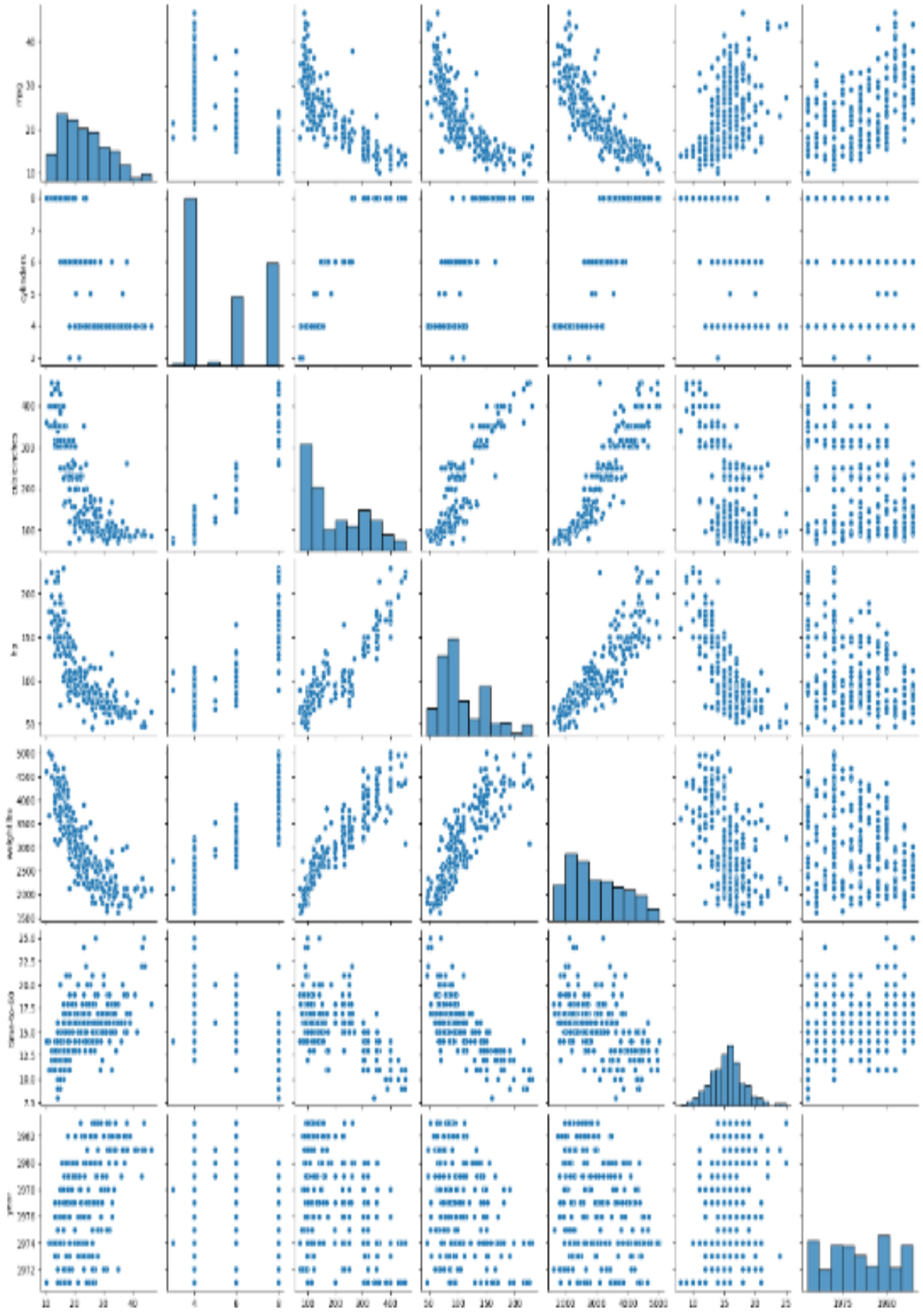
[29]: `sns.pairplot(df)`

```
C:\Users\Shanawaz khan\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119: FutureWarning: use_inf_as_na option is deprecated and will be removed in a future version. Convert inf values to NaN before operating instead.
  with pd.option_context('mode.use_inf_as_na', True):
C:\Users\Shanawaz khan\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119: FutureWarning: use_inf_as_na option is deprecated and will be removed in a future version. Convert inf values to NaN before operating instead.
  with pd.option_context('mode.use_inf_as_na', True):
C:\Users\Shanawaz khan\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119: FutureWarning: use_inf_as_na option is deprecated and will be removed in a future version. Convert inf values to NaN before operating instead.
  with pd.option_context('mode.use_inf_as_na', True):
C:\Users\Shanawaz khan\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119: FutureWarning: use_inf_as_na option is deprecated and will be removed in a future version. Convert inf values to NaN before operating instead.
  with pd.option_context('mode.use_inf_as_na', True):
C:\Users\Shanawaz khan\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119: FutureWarning: use_inf_as_na option is deprecated and will be removed in a future version. Convert inf values to NaN before operating instead.
  with pd.option_context('mode.use_inf_as_na', True):
C:\Users\Shanawaz khan\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119: FutureWarning: use_inf_as_na option is deprecated and will be removed in a future version. Convert inf values to NaN before operating instead.
  with pd.option_context('mode.use_inf_as_na', True):
C:\Users\Shanawaz khan\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119: FutureWarning: use_inf_as_na option is deprecated and will be removed in a future version. Convert inf values to NaN before operating instead.
  with pd.option_context('mode.use_inf_as_na', True):
```

[29]: <seaborn.axisgrid.PairGrid at 0x2b8e6e84150>



[29]: <seaborn.axisgrid.PairGrid at 0x2b8e6e84158>



jupyter Untitled1 Last Checkpoint: 9 minutes ago

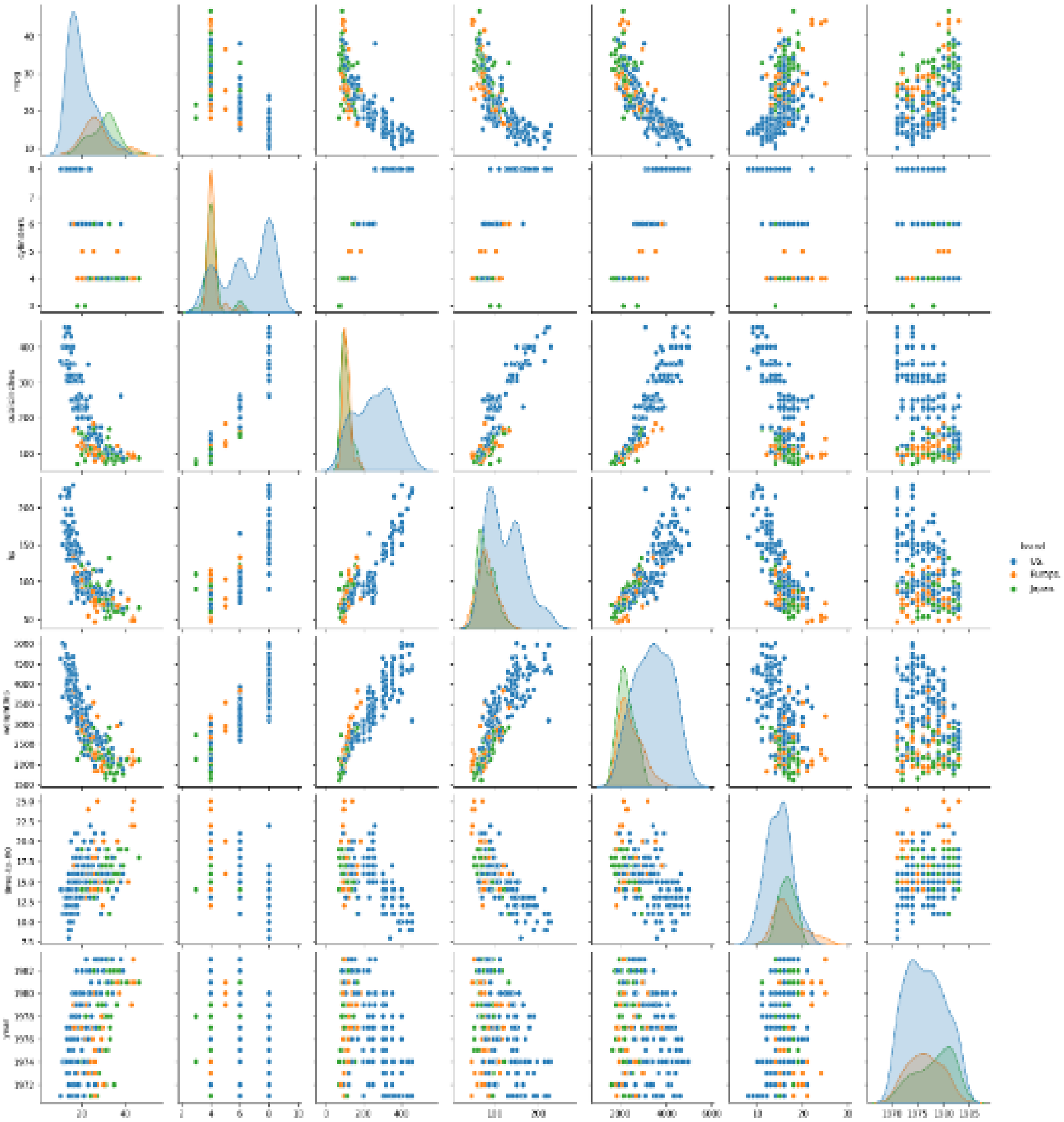
File Edit View Run Kernel Settings Help

Code | JupyterLab | Python 3 (ipykernel)

[30]:
```
sns.pairplot(df,hue="brand")
```

```
C:\Users\Shanawaz khan\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119: FutureWarning: use_inf_as_na option is deprecated and will be removed in a future version. Convert inf values to NaN before operating instead.
  with pd.option_context('mode.use_inf_as_na', True):
C:\Users\Shanawaz khan\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119: FutureWarning: use_inf_as_na option is deprecated and will be removed in a future version. Convert inf values to NaN before operating instead.
  with pd.option_context('mode.use_inf_as_na', True):
C:\Users\Shanawaz khan\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119: FutureWarning: use_inf_as_na option is deprecated and will be removed in a future version. Convert inf values to NaN before operating instead.
  with pd.option_context('mode.use_inf_as_na', True):
C:\Users\Shanawaz khan\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119: FutureWarning: use_inf_as_na option is deprecated and will be removed in a future version. Convert inf values to NaN before operating instead.
  with pd.option_context('mode.use_inf_as_na', True):
C:\Users\Shanawaz khan\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119: FutureWarning: use_inf_as_na option is deprecated and will be removed in a future version. Convert inf values to NaN before operating instead.
  with pd.option_context('mode.use_inf_as_na', True):
C:\Users\Shanawaz khan\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119: FutureWarning: use_inf_as_na option is deprecated and will be removed in a future version. Convert inf values to NaN before operating instead.
  with pd.option_context('mode.use_inf_as_na', True):
C:\Users\Shanawaz khan\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119: FutureWarning: use_inf_as_na option is deprecated and will be removed in a future version. Convert inf values to NaN before operating instead.
  with pd.option_context('mode.use_inf_as_na', True):
```

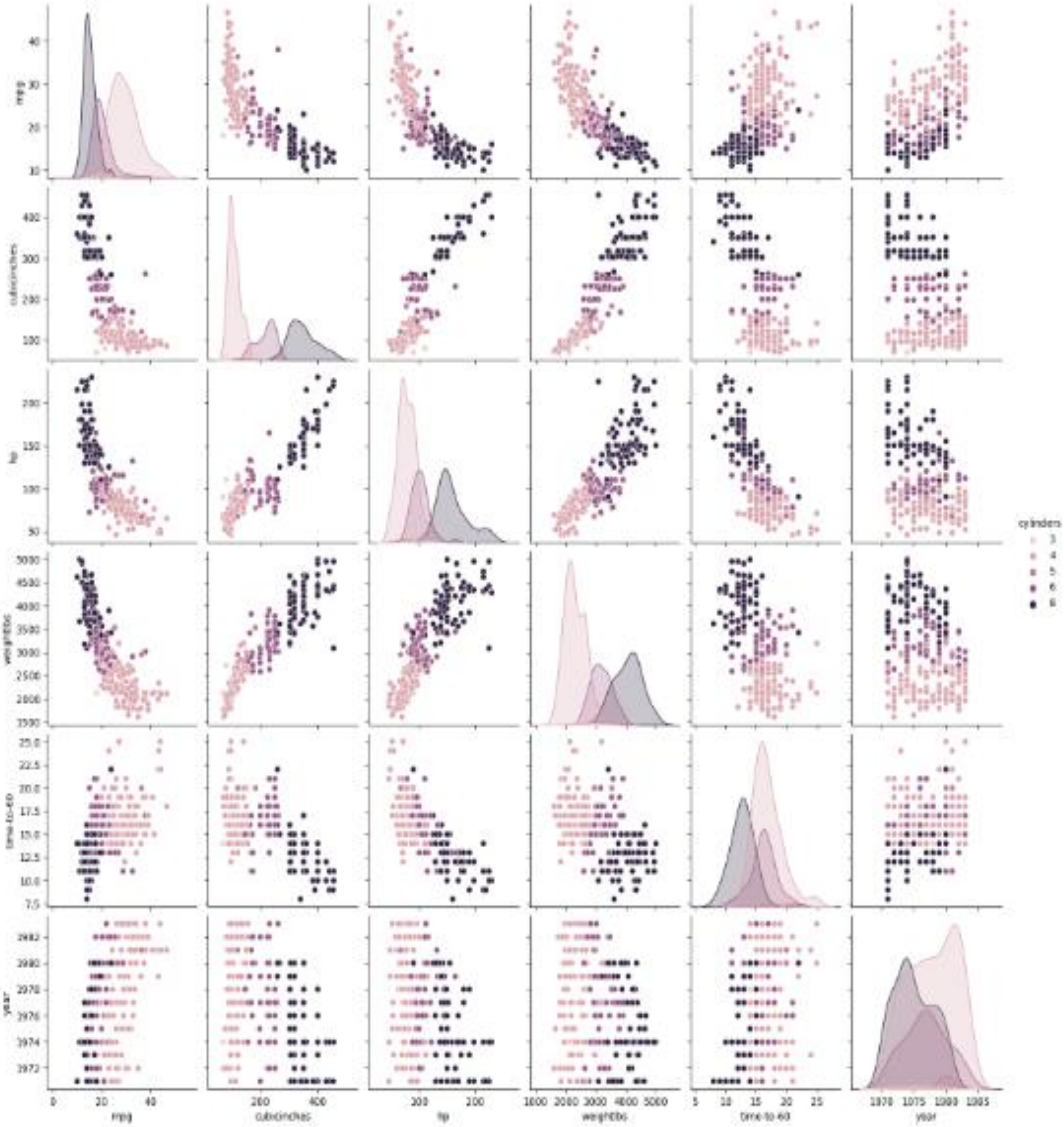
[30]: <seaborn.axisgrid.PairGrid at 0x2b8e9d81910>



[31]:
```
sns.pairplot(df,hue="cylinders")
```

```
C:\Users\Shanawaz khan\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119: FutureWarning: use_inf_as_na option is deprecated and will be removed in a future version. Convert inf values to NaN before operating instead.
  with pd.option_context('mode.use_inf_as_na', True):
C:\Users\Shanawaz khan\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119: FutureWarning: use_inf_as_na option is deprecated and will be removed in a future version. Convert inf values to NaN before operating instead.
  with pd.option_context('mode.use_inf_as_na', True):
C:\Users\Shanawaz khan\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119: FutureWarning: use_inf_as_na option is deprecated and will be removed in a future version. Convert inf values to NaN before operating instead.
  with pd.option_context('mode.use_inf_as_na', True):
C:\Users\Shanawaz khan\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119: FutureWarning: use_inf_as_na option is deprecated and will be removed in a future version. Convert inf values to NaN before operating instead.
  with pd.option_context('mode.use_inf_as_na', True):
C:\Users\Shanawaz khan\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119: FutureWarning: use_inf_as_na option is deprecated and will be removed in a future version. Convert inf values to NaN before operating instead.
  with pd.option_context('mode.use_inf_as_na', True):
C:\Users\Shanawaz khan\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119: FutureWarning: use_inf_as_na option is deprecated and will be removed in a future version. Convert inf values to NaN before operating instead.
  with pd.option_context('mode.use_inf_as_na', True):
```

[31]: <seaborn.axisgrid.PairGrid at 0x2b8f1ec2f10>



Jupyter Untitled1 Last Checkpoint: 10 minutes ago

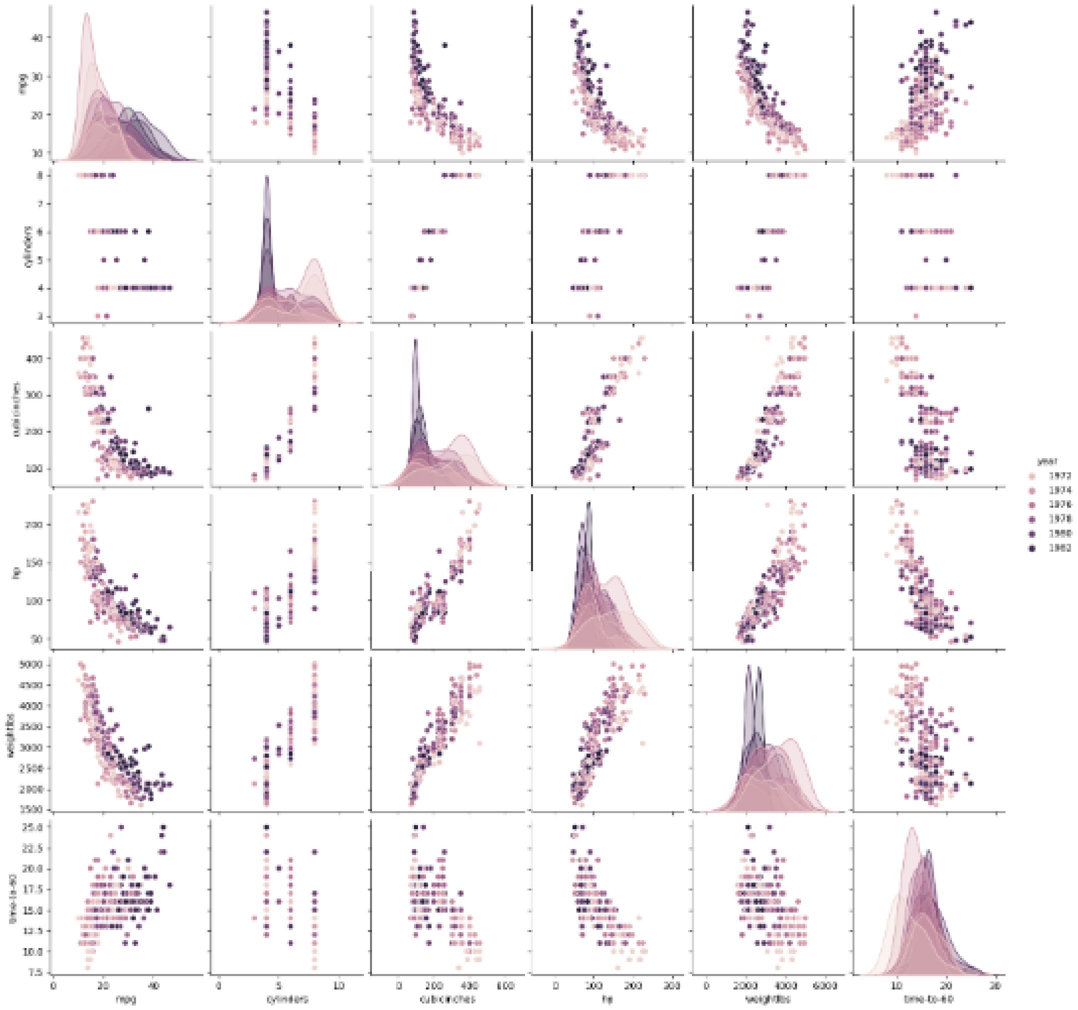
File Edit View Run Kernel Settings Help

Code JupyterLab Python 3 (ipykernel)

[32]:
```
sns.pairplot(df,hue="year")
```

```
C:\Users\Shanawaz khan\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119: FutureWarning: use_inf_as_na option is deprecated and will be removed in a future version. Convert inf values to NaN before operating instead.
  with pd.option_context('mode.use_inf_as_na', True):
C:\Users\Shanawaz khan\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119: FutureWarning: use_inf_as_na option is deprecated and will be removed in a future version. Convert inf values to NaN before operating instead.
  with pd.option_context('mode.use_inf_as_na', True):
C:\Users\Shanawaz khan\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119: FutureWarning: use_inf_as_na option is deprecated and will be removed in a future version. Convert inf values to NaN before operating instead.
  with pd.option_context('mode.use_inf_as_na', True):
C:\Users\Shanawaz khan\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119: FutureWarning: use_inf_as_na option is deprecated and will be removed in a future version. Convert inf values to NaN before operating instead.
  with pd.option_context('mode.use_inf_as_na', True):
C:\Users\Shanawaz khan\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119: FutureWarning: use_inf_as_na option is deprecated and will be removed in a future version. Convert inf values to NaN before operating instead.
  with pd.option_context('mode.use_inf_as_na', True):
C:\Users\Shanawaz khan\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119: FutureWarning: use_inf_as_na option is deprecated and will be removed in a future version. Convert inf values to NaN before operating instead.
  with pd.option_context('mode.use_inf_as_na', True):
```

[32]: <seaborn.axisgrid.PairGrid at 0x2b8f47d63d0>



1. Create a heatmap for the entire data to study correlation between each of the columns

[33]:
```
correlation=df.corr(numeric_only=True)
```

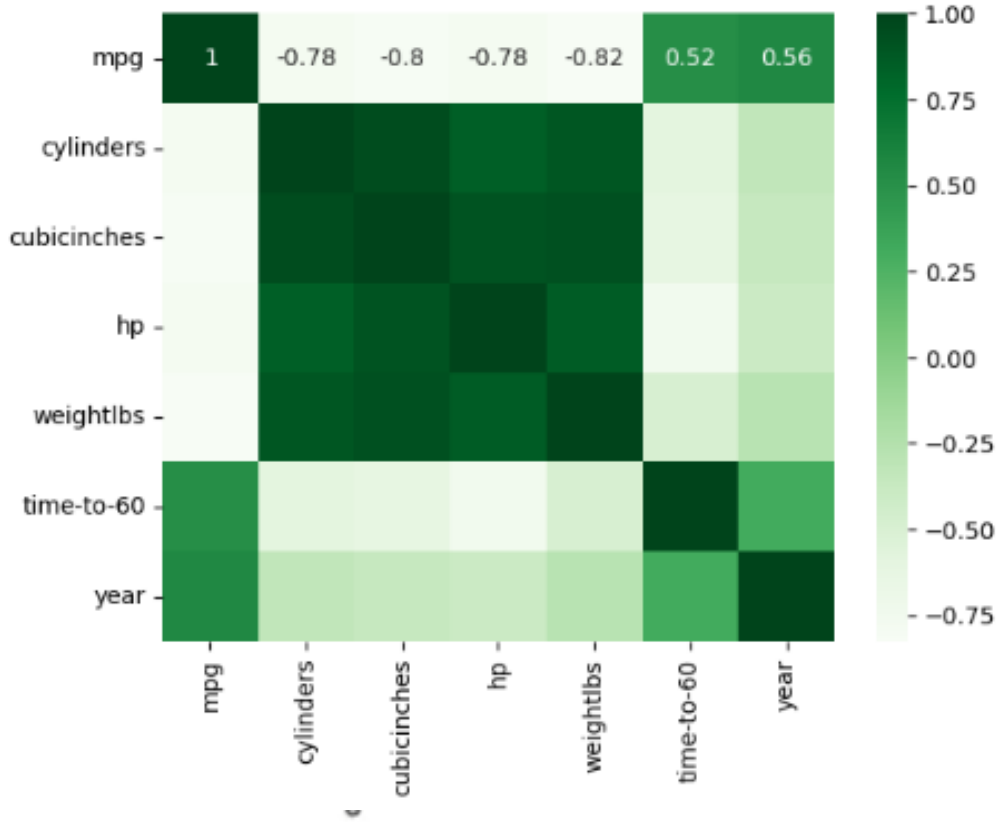
[34]:
```
correlation
```

[34]:

	mpg	cylinders	cubicinches	hp	weightlbs	time-to-60	year
mpg	1.000000	-0.776599	-0.803830	-0.779954	-0.824945	0.520401	0.561719
cylinders	-0.776599	1.000000	0.951529	0.847450	0.897247	-0.583449	-0.329193
cubicinches	-0.803830	0.951529	1.000000	0.907341	0.930027	-0.613344	-0.359215
hp	-0.779954	0.847450	0.907341	1.000000	0.863467	-0.745310	-0.393079
weightlbs	-0.824945	0.897247	0.930027	0.863467	1.000000	-0.488671	-0.281156
time-to-60	0.520401	-0.583449	-0.613344	-0.745310	-0.488671	1.000000	0.315549
year	0.561719	-0.329193	-0.359215	-0.393079	-0.281156	0.315549	1.000000

[39]:
```
sns.heatmap(correlation,annot=True,cmap="Greens")
```

[39]:
```
<Axes: >
```



1. Create a pie chart for the following columns and their distribution in the data:

[50]:
```
df["brand"].value_counts()
```

[50]:
```
brand
 US.        158
 Japan.      51
 Europe.     47
Name: count, dtype: int64
```

[51]:
```
df["brand"].value_counts().plot.pie(autopct="%1.1f%%")
```

[51]:
```
<Axes: ylabel='count'>
```