In [1]:

```
# import all the library files needed

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

In [2]:

```
# read the Automobile price data _Raw_.csv file and assign it as car_insurance

car_insurance = pd.read_csv(r"Automobile price data _Raw_.csv")
```

In [3]:

```
# show the first 5 rows of car_insurance

car_insurance.head()
```

Out[3]:

	symboling	normalized-losses	make	fuel-type	aspiration	num-of-doors	body-style	drive-wheels	engine-location	wheel-base	...	engine-size	fuel-system	bore
0	3	?	alfa-romero	gas	std	two	convertible	rwd	front	88.6	...	130	mpfi	3.47
1	3	?	alfa-romero	gas	std	two	convertible	rwd	front	88.6	...	130	mpfi	3.47
2	1	?	alfa-romero	gas	std	two	hatchback	rwd	front	94.5	...	152	mpfi	2.68
3	2	164	audi	gas	std	four	sedan	fwd	front	99.8	...	109	mpfi	3.19
4	2	164	audi	gas	std	four	sedan	4wd	front	99.4	...	136	mpfi	3.19

5 rows × 26 columns

In [4]:

```
car_insurance.info()

# We are able to notice object data type for some of the columns
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 205 entries, 0 to 204
Data columns (total 26 columns):
symboling            205 non-null int64
normalized-losses    205 non-null object
make                 205 non-null object
fuel-type            205 non-null object
aspiration           205 non-null object
num-of-doors         205 non-null object
body-style           205 non-null object
drive-wheels         205 non-null object
engine-location      205 non-null object
wheel-base           205 non-null float64
length               205 non-null float64
width                205 non-null float64
height               205 non-null float64
curb-weight          205 non-null int64
engine-type          205 non-null object
num-of-cylinders     205 non-null object
engine-size          205 non-null int64
```

```
fuel-system          205 non-null object
bore                 205 non-null object
stroke               205 non-null object
compression-ratio    205 non-null float64
horsepower           205 non-null object
peak-rpm             205 non-null object
city-mpg             205 non-null int64
highway-mpg          205 non-null int64
price                205 non-null object
dtypes: float64(5), int64(5), object(16)
memory usage: 41.8+ KB
```

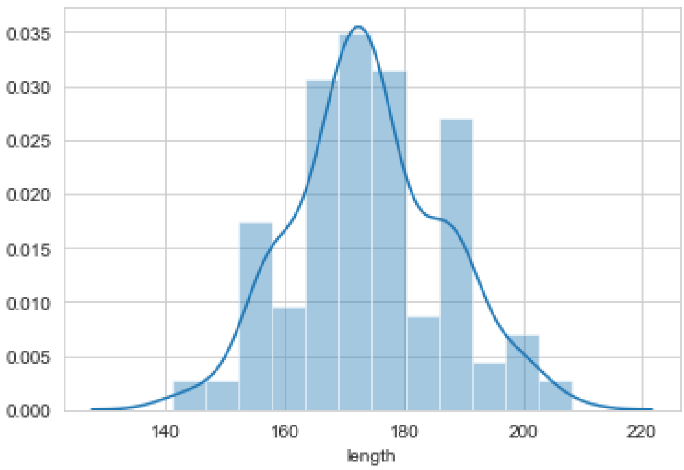
Histogram

In [5]:

```
sns.set_style("whitegrid")
sns.distplot(car_insurance['length'])

# We can observe the positive kurtosis here due to its high peakness in the plot
# Max length of making a car is 208
# Majority of length of cars made in range between 163-180
# Min length of making a car is 141
```

Out[5]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x2684ff94788>
```



In [6]:

```
from scipy.stats import skew
from scipy.stats import kurtosis
print(skew(car_insurance['length']))
print(kurtosis(car_insurance['length']))
```

```
0.15481031885453517
-0.11001300115343327
```

In [7]:

```
car_insurance['length'].describe()
```

Out[7]:

```
count    205.000000
mean     174.049268
std       12.337289
min      141.100000
25%      166.300000
50%      173.200000
75%      183.100000
max      208.100000
Name: length, dtype: float64
```

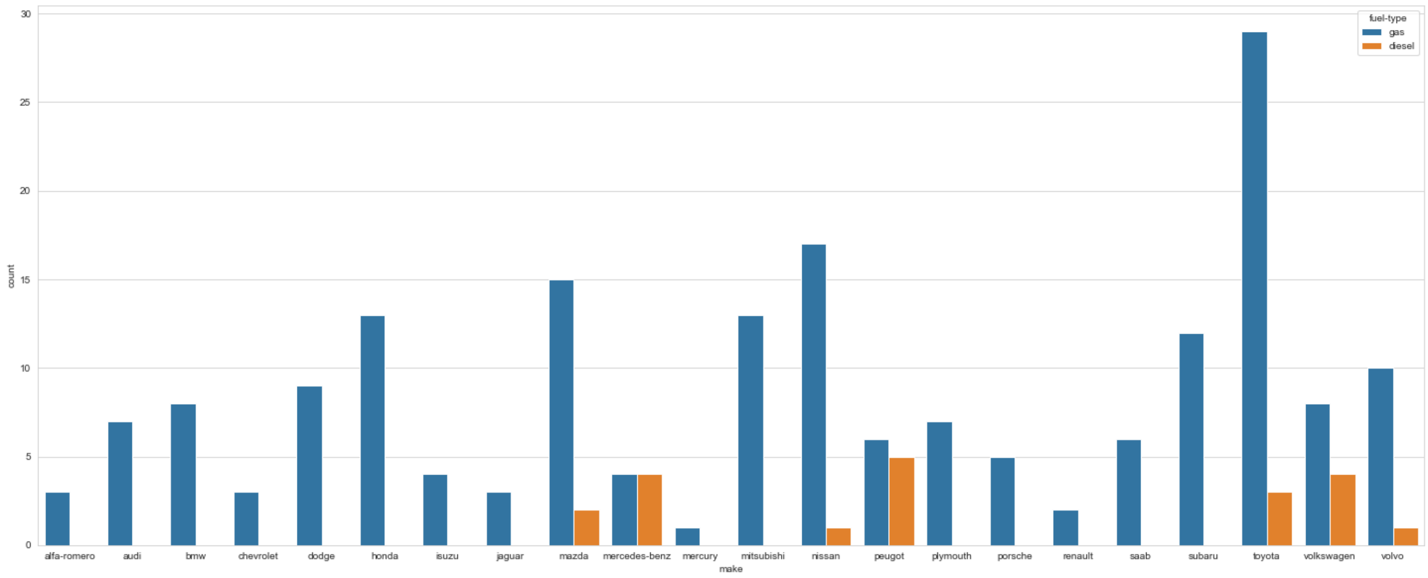
Countplot

In [8]:

```
plt.figure(figsize=(25,10))
sns.countplot(data=car_insurance,x='make',hue='fuel-type')

# We can observe here majority of cars have been insured in 'Toyota' brand with 'gas' fue
l-type
```

Out[8]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x26850308b48>
```

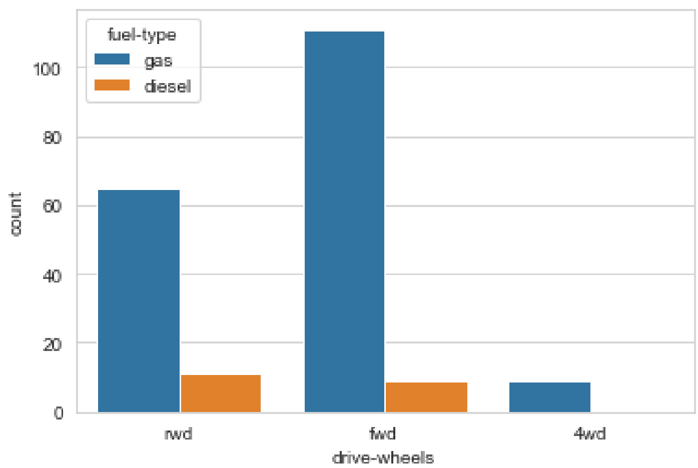


In [9]:

```
sns.countplot(data=car_insurance,x='drive-wheels',hue='fuel-type')

# Comparison based on the drive-wheels with fuel-type
# We can see many companies prefer the 'FWD' and 'RWD' drive wheels with gas engine
# Companies drive-wheels with 'FWD' and 'RWD' uses very less count of diesel engines.
# Companies drive-wheels with '4WD' has only gas engine and there is no diesel engine use
d.
```

Out[9]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x26850985e08>
```



Scatterplot

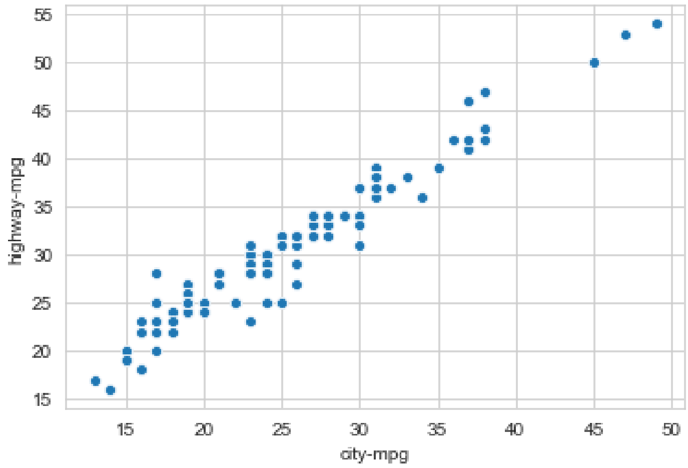
In [10]:

```
sns.scatterplot(data=car_insurance,x='city-mpg',y='highway-mpg')
```

```
# Here we can observe the moderate positive correlation between 'city-mpg' and 'highway-m
pg'
```

Out[10]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x2685049e388>
```



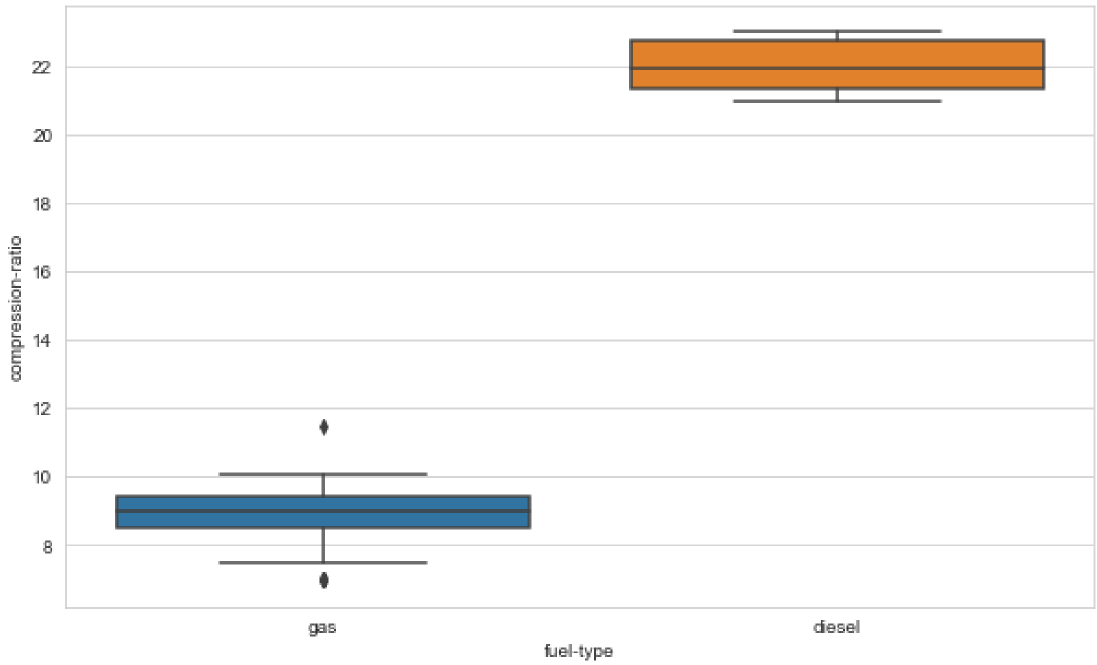
Boxplot

In [11]:

```
plt.figure(figsize=(10,6))
sns.boxplot(data=car_insurance,x='fuel-type',y='compression-ratio')

# We can notice two outliers in the x-axis (gas)
# We can observe here the compression ratio is higher in the diesel engine than the gas e
ngine
# Higher the compression-ratio lower will be the consumption of fuel
```

Out[11]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x2685051b0c8>
```



In [12]:

```
car_insurance['compression-ratio'].describe()
```

Out[12]:

```
count    205.000000
mean      10.142537
```

```
std          3.972040
min          7.000000
25%          8.600000
50%          9.000000
75%          9.400000
max         23.000000
Name: compression-ratio, dtype: float64
```

Heatmap

In [13]:

```
plt.figure(figsize=(10,6))
g = car_insurance.corr()
sns.heatmap(g,cmap='viridis',annot=True,linewidths=2)
plt.axis('scaled')

# This gives the correlation of the insurance data and the intensity of each column that is mapped.
```

Out[13]:

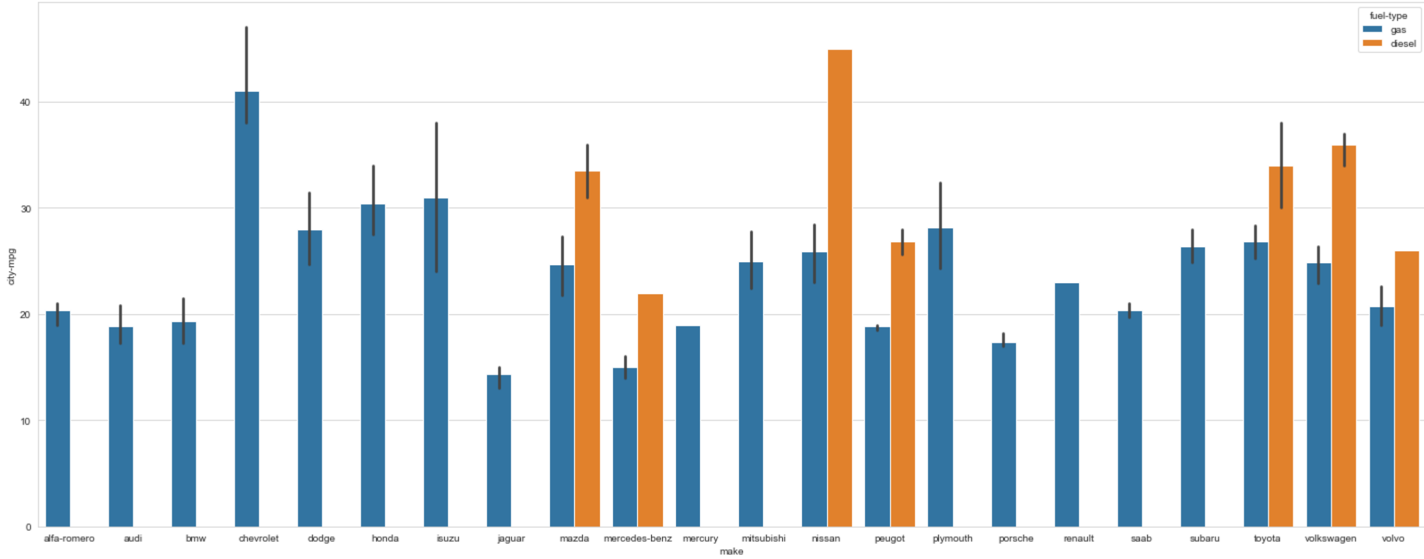
```
(0.0, 10.0, 10.0, 0.0)
```

Barplot

In [14]:

```
plt.figure(figsize=(20,8))
sns.barplot(data=car_insurance,x='make',y='city-mpg',hue='fuel-type')
plt.tight_layout()

# we can see majority of cars manufactured with 'gas' fuel-type because it uses spark-plug to ignite the engine.
# Gasoline engines are quieter and nippier compared to diesel engines.
# Pollution, Diesel cars emit roughly 13% more CO2 gas per gallon of fuel compared to gasoline cars.


# nissan company makes more no_ of_cars with diesel engine performs good in city test drive
# chevrolet also makes more no_ of_cars with gas engine but this gives better results tha
```

```
n the nissan in the city test drive
# although the diesel engines usage is low but its performance is higher than the gas eng
ines.
```

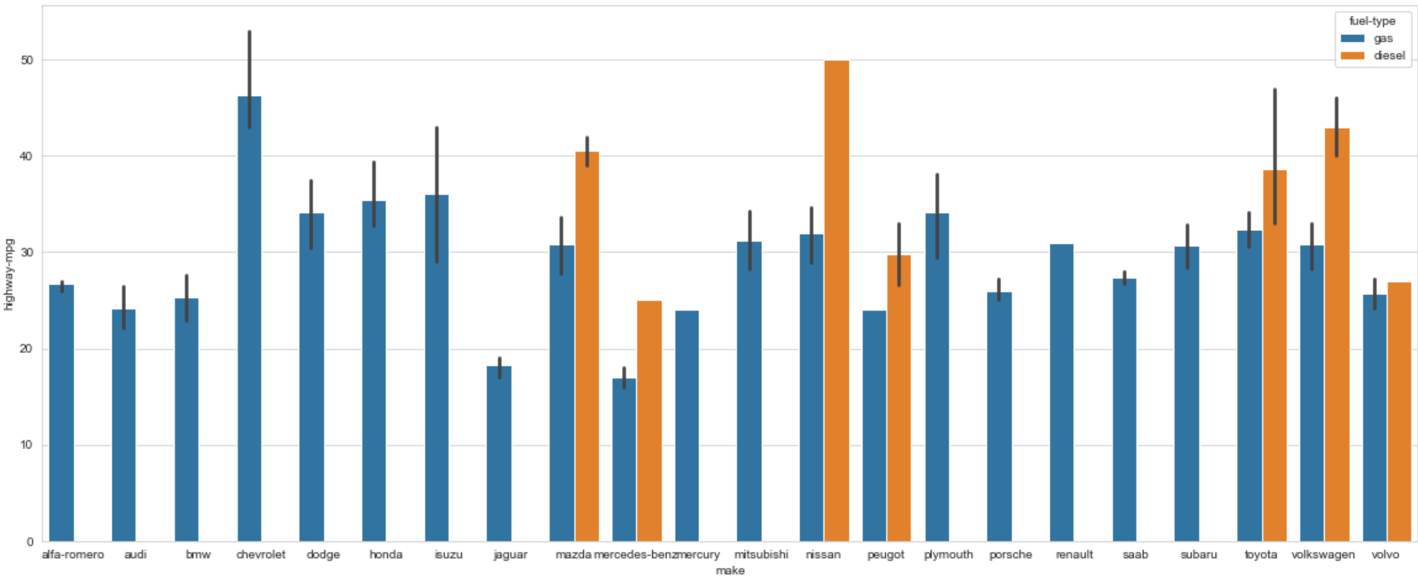


In [15]:

```
plt.figure(figsize=(20,8))
sns.barplot(data=car_insurance,x='make',y='highway-mpg',hue='fuel-type')

# nissan company makes more no_ of_cars with diesel engine performs good in highway test
drive
# chevrolet also makes more no_of_cars with gas engine but this gives better results than
the nissan in the highway test drive
# although the diesel engines usage is low but its performance is higher than the gas eng
ines.
```

Out[15]:

<matplotlib.axes._subplots.AxesSubplot at 0x268508e2e48>



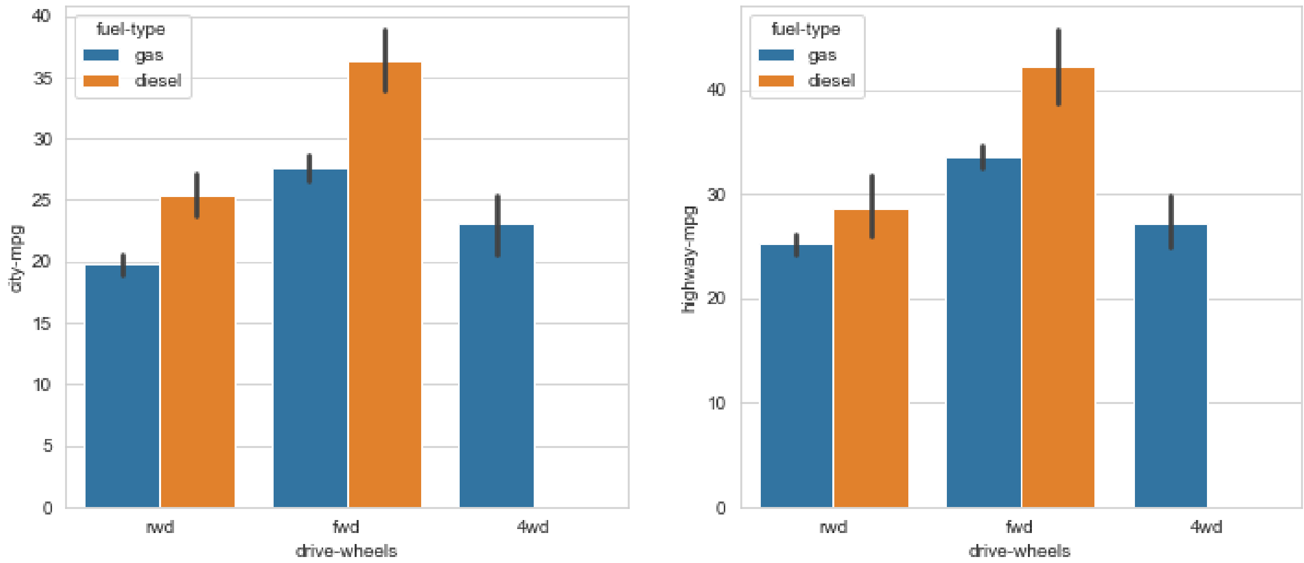
In [16]:

```
plt.figure(figsize=(12,5))

plt.subplot(1,2,1)
sns.barplot(x="drive-wheels", y="city-mpg",hue="fuel-type",data=car_insurance)

plt.subplot(1,2,2)
sns.barplot(x="drive-wheels", y="highway-mpg",hue="fuel-type",data=car_insurance)

# Comparison based on drive-wheels
# Cars with 'FWD' using diesel type engines gives high mpg in both city and highway
# Cars with 'RWD' using gas type engines gives very low mpg in both city and highway
```

```python
# Majority of the cars use the 'FWD' - forward wheel drive for long drives
```

Out[16]:

<matplotlib.axes._subplots.AxesSubplot at 0x26850e45a48>



In [ ]:

```python

```