Automobile Analysis

We will be analyzing automobile data to find different aspects of different types of vehicles. We will be focusing our analysis on following parameters:

1. Curb weight and Engine size and horsepowers
2. Type of engine and number of doors
3. Fuel type
4. Body type
5. Price range
6. Safety

Description of Data:

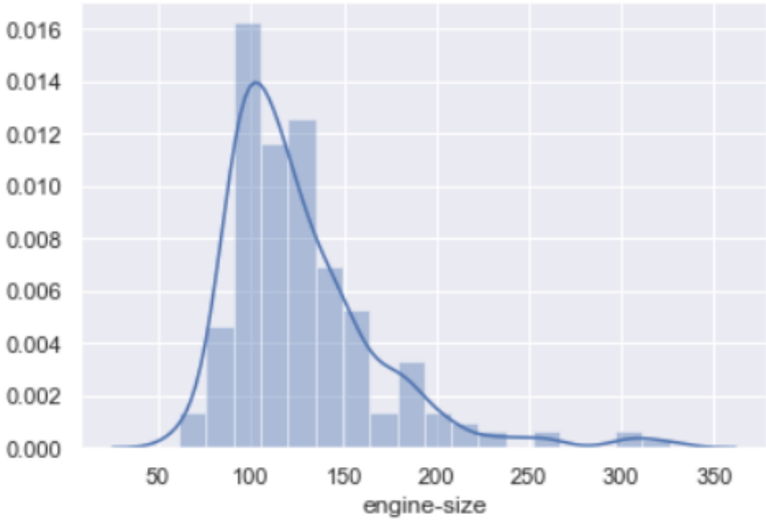
```
In [4]: auto.describe()
```

Out[4]:

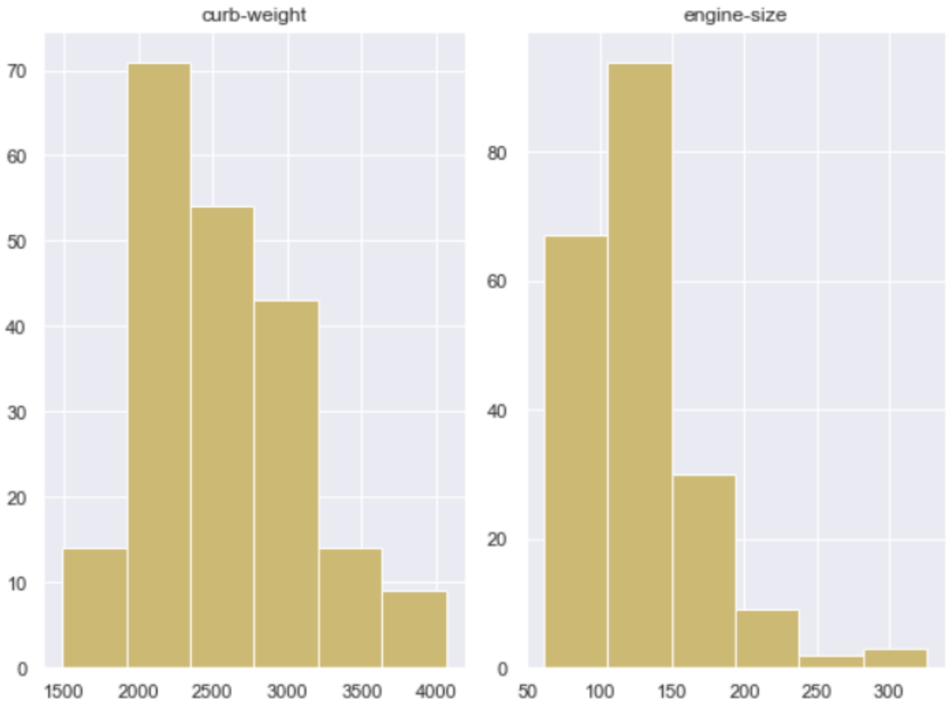
	symboling	wheel-base	length	width	height	curb-weight	engine-size	compression-ratio	city-mpg	highway-mpg
count	205.000000	205.000000	205.000000	205.000000	205.000000	205.000000	205.000000	205.000000	205.000000	205.000000
mean	0.834146	98.756585	174.049268	65.907805	53.724878	2555.565854	126.907317	10.142537	25.219512	30.751220
std	1.245307	6.021776	12.337289	2.145204	2.443522	520.680204	41.642693	3.972040	6.542142	6.886443
min	-2.000000	86.600000	141.100000	60.300000	47.800000	1488.000000	61.000000	7.000000	13.000000	16.000000
25%	0.000000	94.500000	166.300000	64.100000	52.000000	2145.000000	97.000000	8.600000	19.000000	25.000000
50%	1.000000	97.000000	173.200000	65.500000	54.100000	2414.000000	120.000000	9.000000	24.000000	30.000000
75%	2.000000	102.400000	183.100000	66.900000	55.500000	2935.000000	141.000000	9.400000	30.000000	34.000000
max	3.000000	120.900000	208.100000	72.300000	59.800000	4066.000000	326.000000	23.000000	49.000000	54.000000

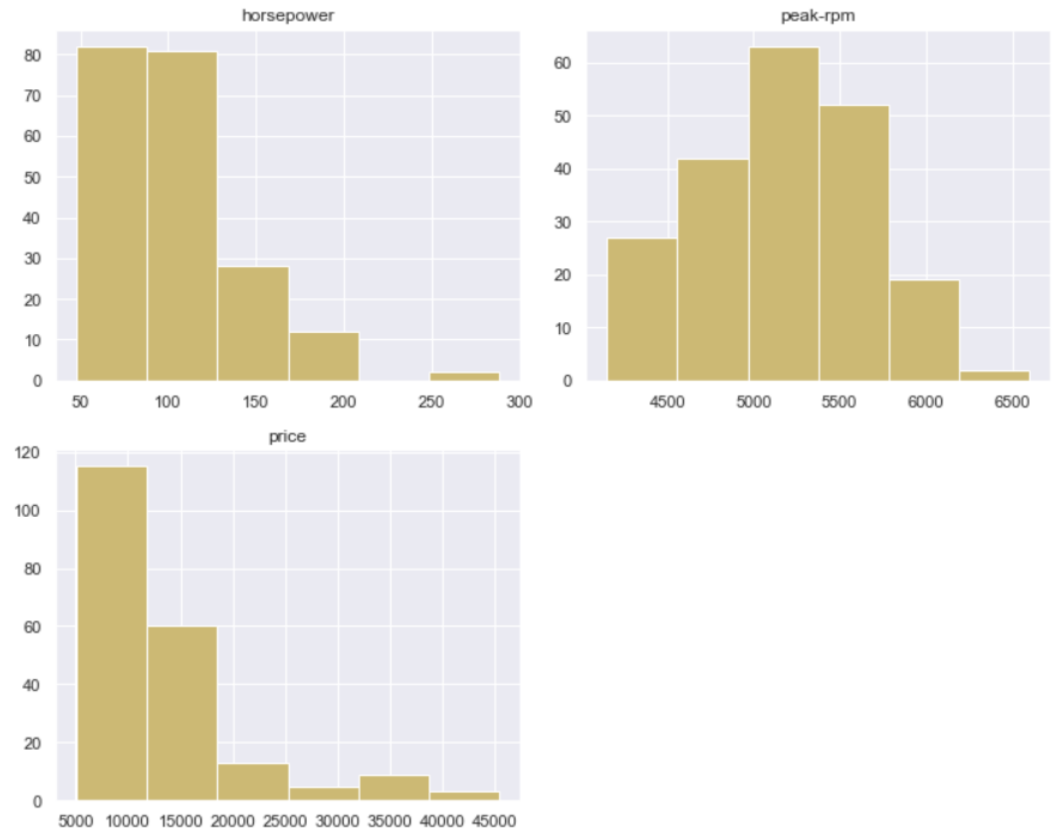
<u>Engine Size, Power parameters and wheel base:</u>

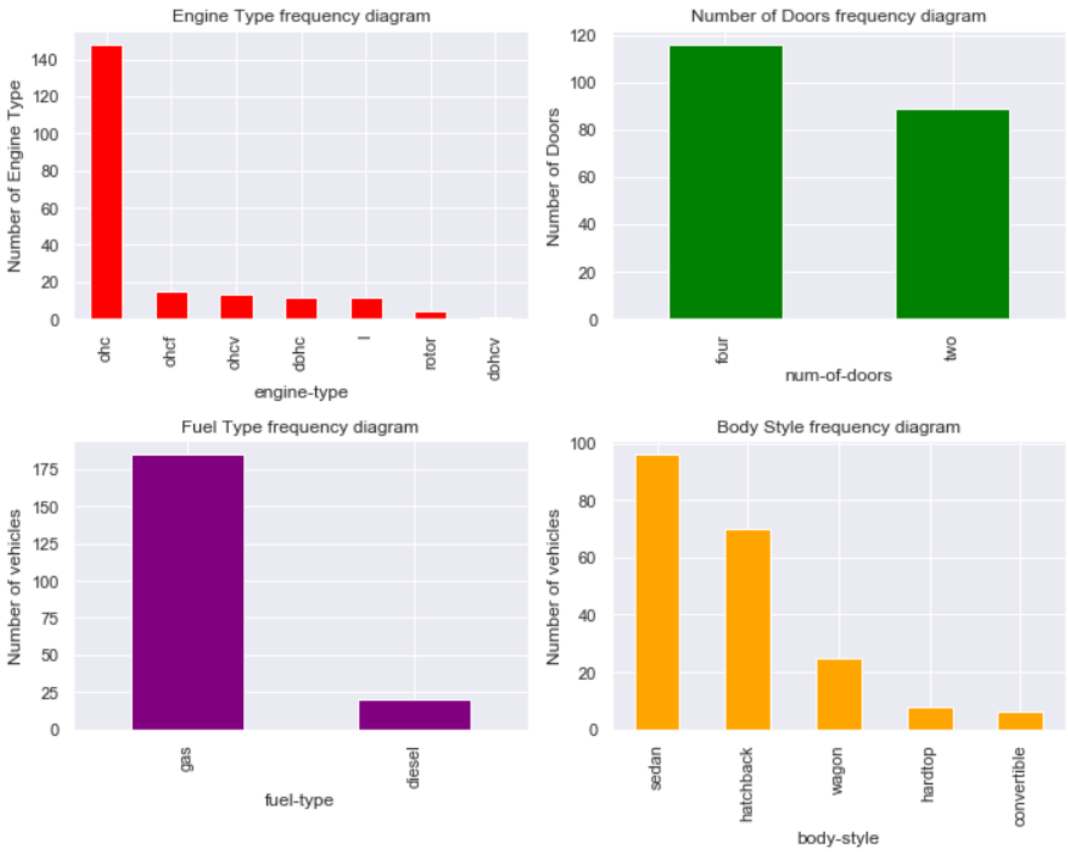
In [9]:
```
sns.distplot(df['engine-size']);
```



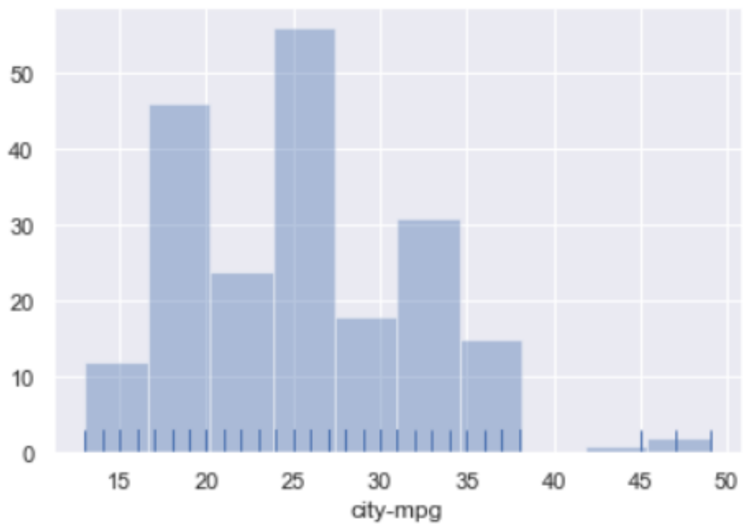
In [61]:
```
df2[['engine-size','curb-weight']].hist(figsize=(8,6),bins=6,color='Y')
plt.tight_layout()
plt.show()
```





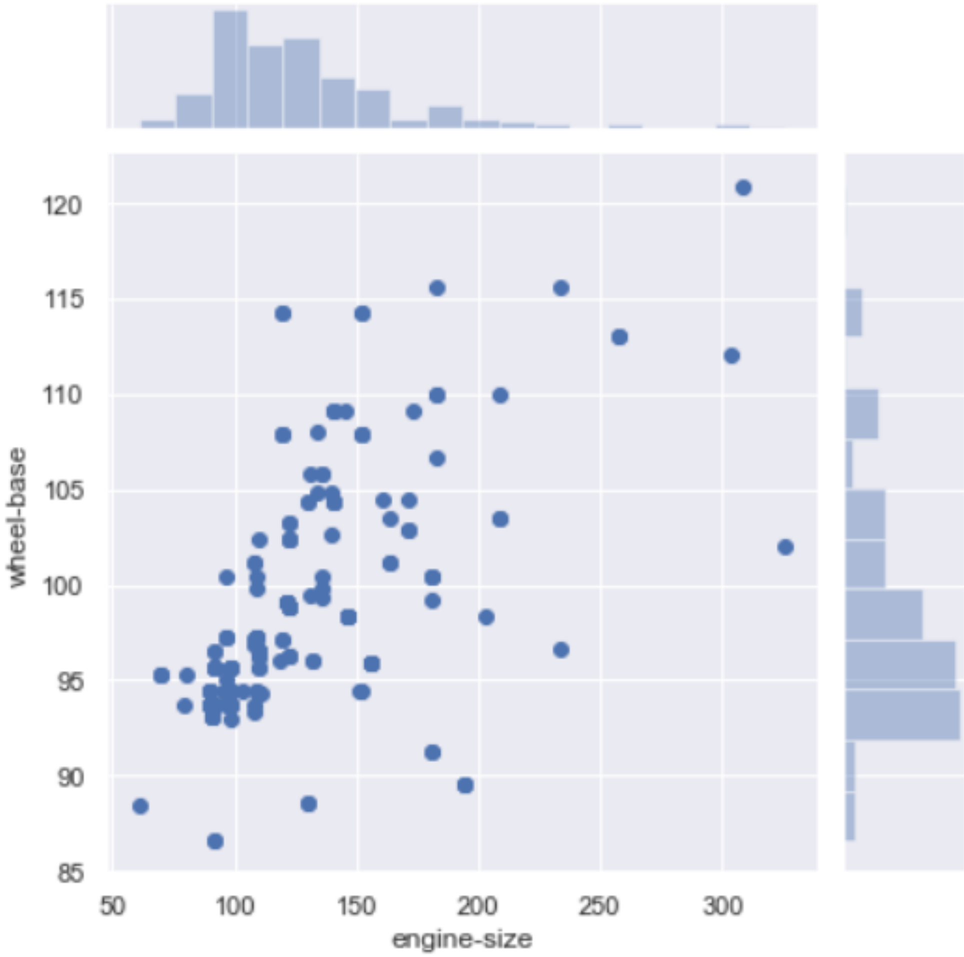


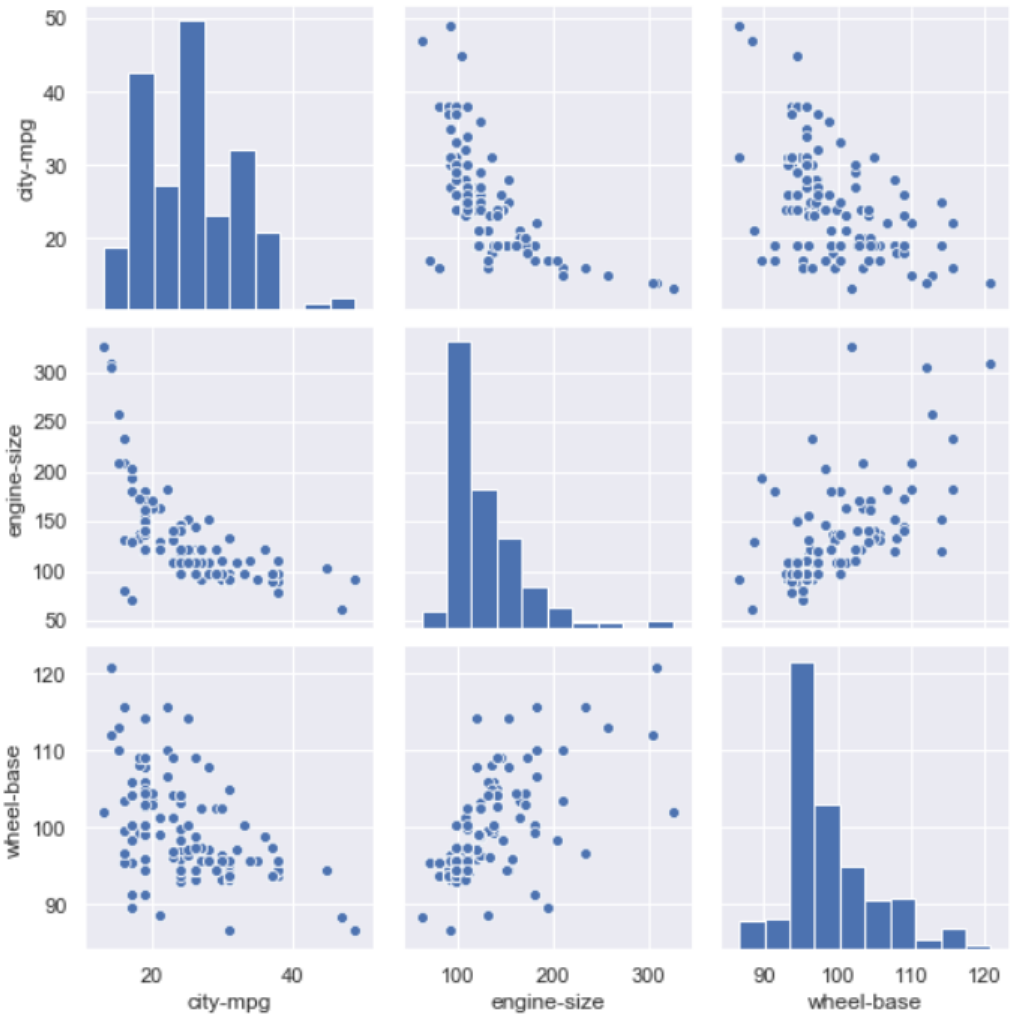
In [10]:
```
sns.distplot(auto['city-mpg'], kde=False, rug=True);
```



In [6]:
```
sns.jointplot(auto['engine-size'],auto['wheel-base'])
```

Out[6]: <seaborn.axisgrid.JointGrid at 0x13332788508>





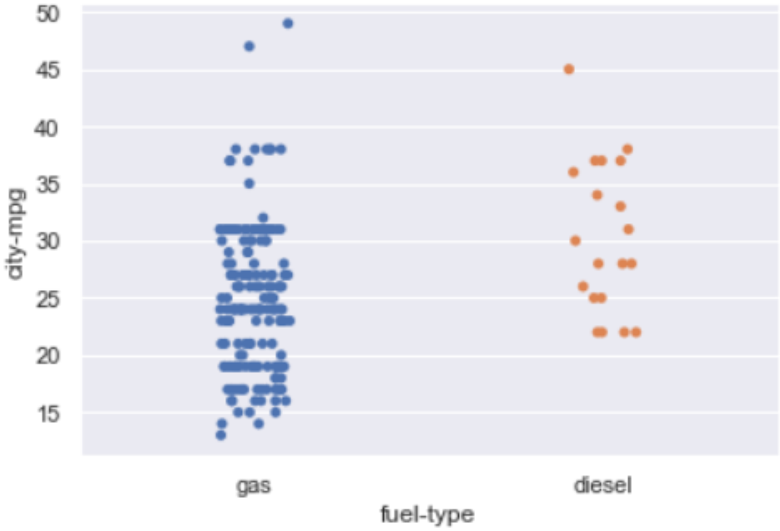
Conclusion:

- Most of the car has a Curb Weight is in range 1900 to 3100 kg.
- The Engine Size is in the range 60 to 190 cc.
- Most vehicle has horsepower 50 to 125 bhp.
- Most Vehicle are in price range 5000 to 18000 $.
- peak rpm is mostly distributed between 4600 to 5700.
- Most cars have engine size around 100 cc and wheel base around 95 cm.
- City mileage is concentrated around 20-30 miles per gallon.
- Most vehicles are running on gas.

Engine type, fuel parameters and body type:

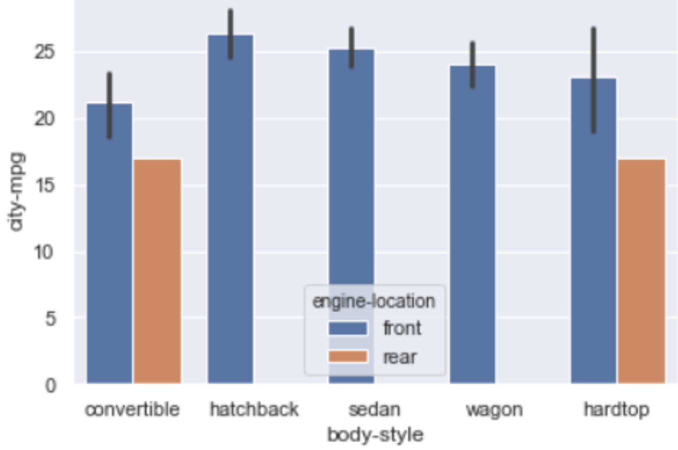
In [12]:
```
sns.stripplot(auto['fuel-type'], auto['city-mpg'],jitter=True)
```

Out[12]: <matplotlib.axes._subplots.AxesSubplot at 0x1333320e808>



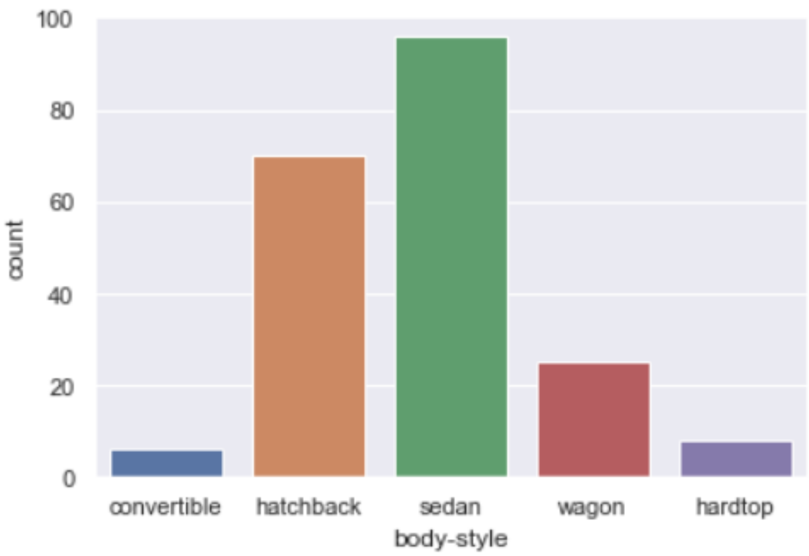
In [53]:
```
sns.barplot(df['body-style'], df['city-mpg'], hue=df['engine-location'])
```

Out[53]: <matplotlib.axes._subplots.AxesSubplot at 0x176d7f226c8>



```
In [18]: sns.countplot(auto['body-style'])
```

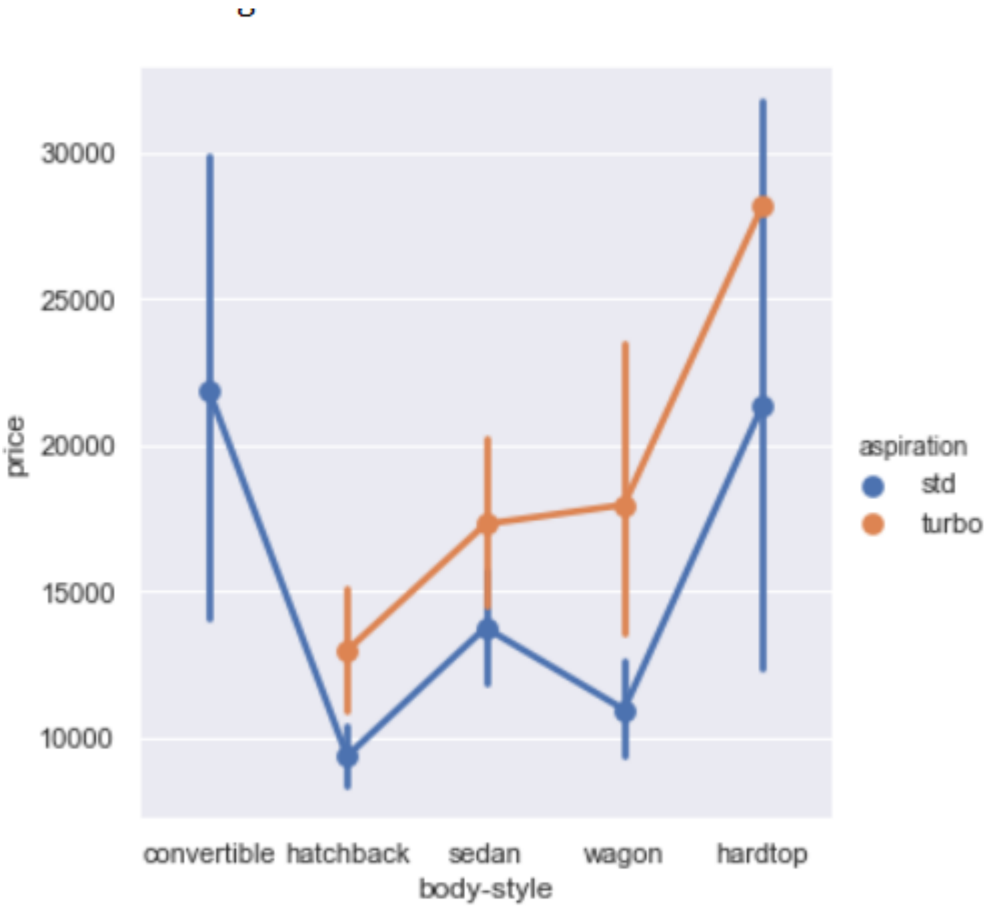
```
Out[18]: <matplotlib.axes._subplots.AxesSubplot at 0x13334469d88>
```



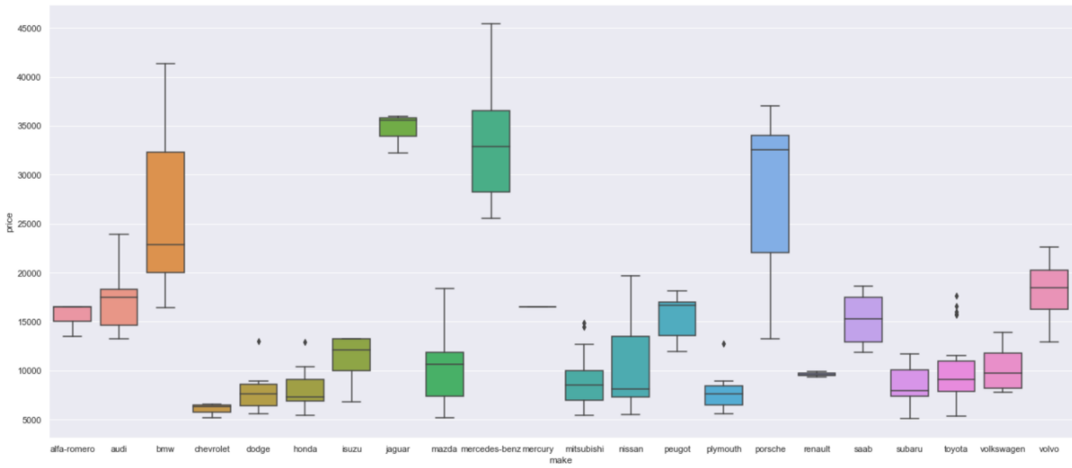
Conclusion:

- More than 70 % of the vehicle has Ohc type of Engine
- 57% of the cars has 4 doors
- Gas is preferred by 85 % of the vehicles
- Most produced vehicle are of body style sedan around 48% followed by hatchback 32%.

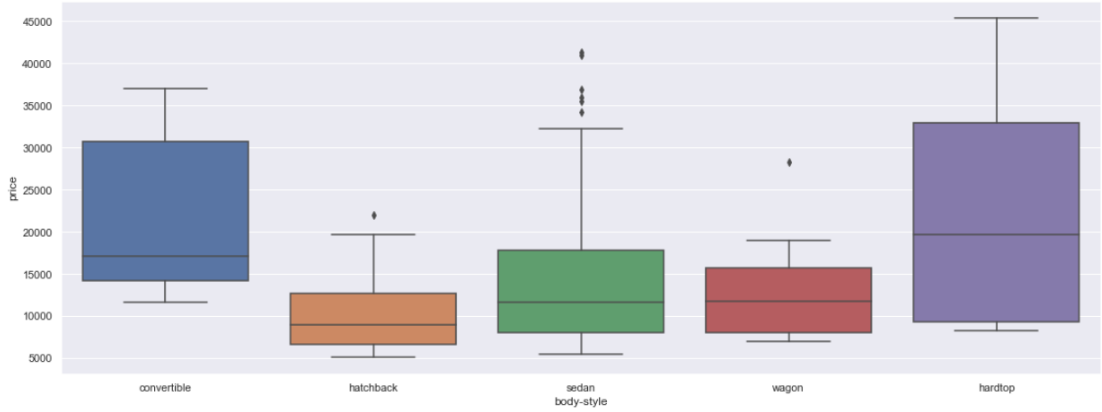
Price:



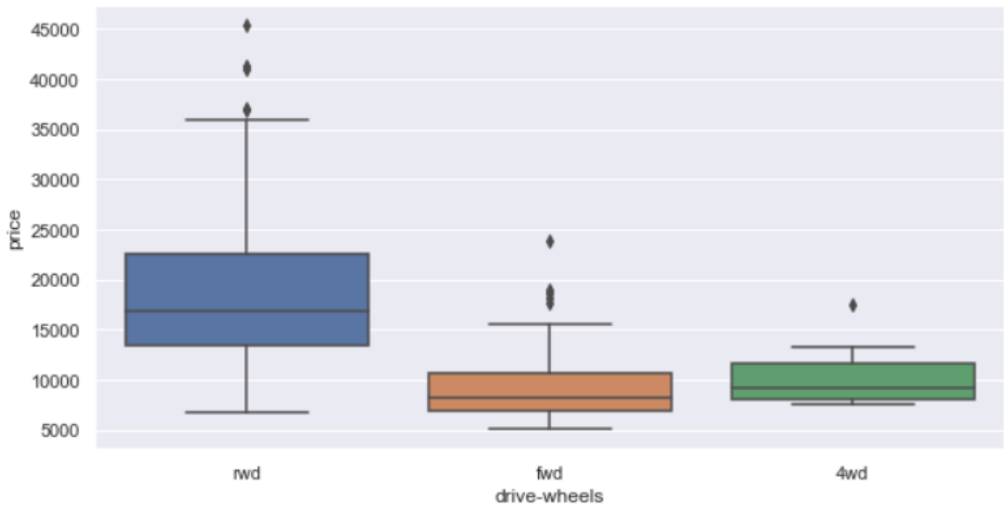
In [56]:
```
plt.rcParams['figure.figsize']=(23,10)
ax = sns.boxplot(x="make", y="price", data=df)
```



```
In [57]: plt.rcParams['figure.figsize']=(19,7)
ax = sns.boxplot(x="body-style", y="price", data=df)
```




```
In [58]: plt.rcParams['figure.figsize']=(10,5)
ax = sns.boxplot(x="drive-wheels", y="price", data=df)
```



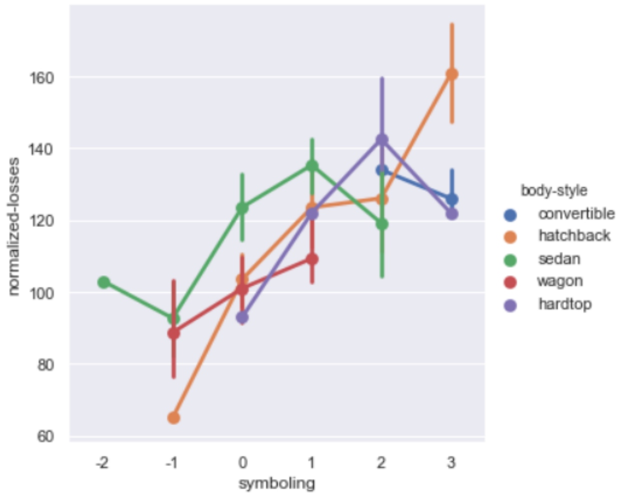
Conclusion:

- Mercedez-Benz ,BMW, Jaguar, Porshe produces expensive cars more than 25000
- Cheverolet, dodge, honda, mitbushi, nissan, plymouth subaru, toyata produces budget models with lower prices
- most of the cars comapanies produce car in range below 25000
- Hardtop model are expensive in prices followed by convertible and sedan body style
- Turbo models have higher prices than for the standard model

- Convertible has only standard edition with expensive cars
- hatchback and sedan turbo models are available below 20000
- rwd wheel drive vehicle have expensive prices

Safety:

```
In [59]: sns.catplot(data=df, y="normalized-losses", x="symboling" , hue="body-style" ,kind="point")
Out[59]: <seaborn.axisgrid.FacetGrid at 0x176d6eca808>
```



Conclusion:

Note :- here +3 means risky vehicle and -2 means safe vehicle

- Increased in risk rating linearly increases in normalised losses in vehicle
- covertible car and hardtop car has mostly losses with risk rating above 0
- hatchback cars has highest losses at risk rating 3
- sedan and Wagon car has losses even in less risk (safe)rating