

```
In [2]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
In [3]: data = sns.load_dataset('car_crashes')
```

```
In [4]: data.head()
```

Out[4]:

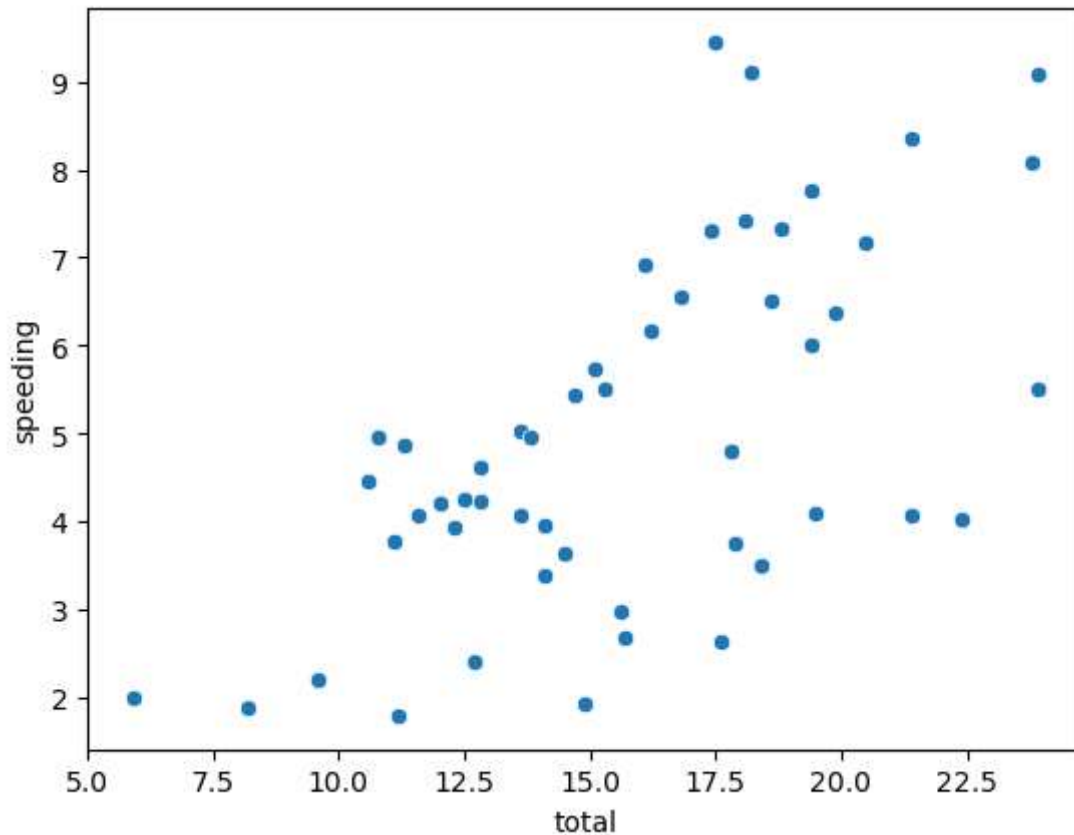
	total	speeding	alcohol	not_distracted	no_previous	ins_premium	ins_losses	abbrev
0	18.8	7.332	5.640	18.048	15.040	784.55	145.08	AL
1	18.1	7.421	4.525	16.290	17.014	1053.48	133.93	AK
2	18.6	6.510	5.208	15.624	17.856	899.47	110.35	AZ
3	22.4	4.032	5.824	21.056	21.280	827.34	142.39	AR
4	12.0	4.200	3.360	10.920	10.680	878.41	165.63	CA

```
In [5]: data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 51 entries, 0 to 50
Data columns (total 8 columns):
#   Column                Non-Null Count  Dtype
---  -
0   total                  51 non-null    float64
1   speeding               51 non-null    float64
2   alcohol                51 non-null    float64
3   not_distracted         51 non-null    float64
4   no_previous            51 non-null    float64
5   ins_premium            51 non-null    float64
6   ins_losses             51 non-null    float64
7   abbrev                 51 non-null    object
dtypes: float64(7), object(1)
memory usage: 3.3+ KB
```

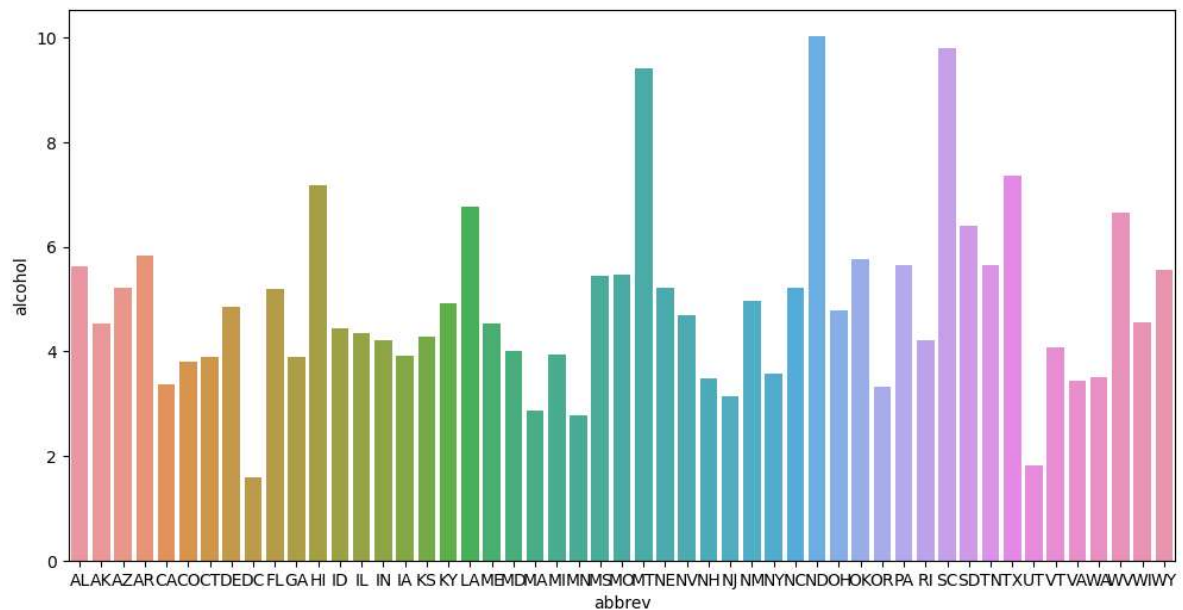
```
In [6]: '''inference: the scatter plot below suggests that there is a positive_
↳correlation between the number of drivers involved in fatal collisions and_
↳the percentage of drivers who were speeding.
This means that states with higher speeding rates tend to have more fatal_
↳collisions. '''
sns.scatterplot(x = 'total', y = 'speeding', data = data)
```

```
Out[6]: <AxesSubplot:xlabel='total', ylabel='speeding'>
```

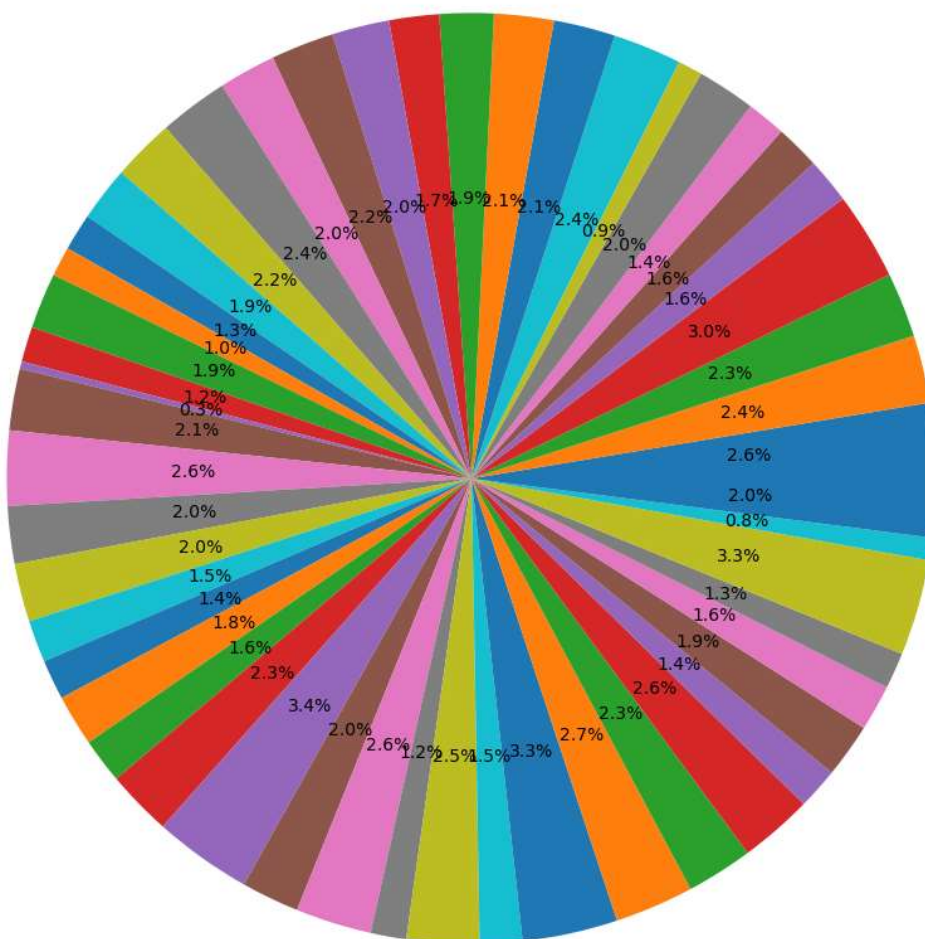


```
In [9]: ''' Inference: the percentage of drivers who were alcohol-impaired varies_
↳significantly from state to state.
This suggests that alcohol-impaired driving is a major factor in fatal_
↳collisions in some states, but not others.'''
plt.figure(figsize=(12,6))
sns.barplot(x = 'abbrev', y = 'alcohol', data = data)
```

```
Out[9]: <AxesSubplot:xlabel='abbrev', ylabel='alcohol'>
```



```
In [10]: ''' Inference: the majority of drivers involved in fatal collisions were not_
↳distracted.
This suggests that distracted driving is not the leading cause of fatal_
↳collisions.'''
plt.figure(figsize=(12, 12))
plt.pie(data['not_distracted'], autopct='%1.1f%%')
plt.show()
```

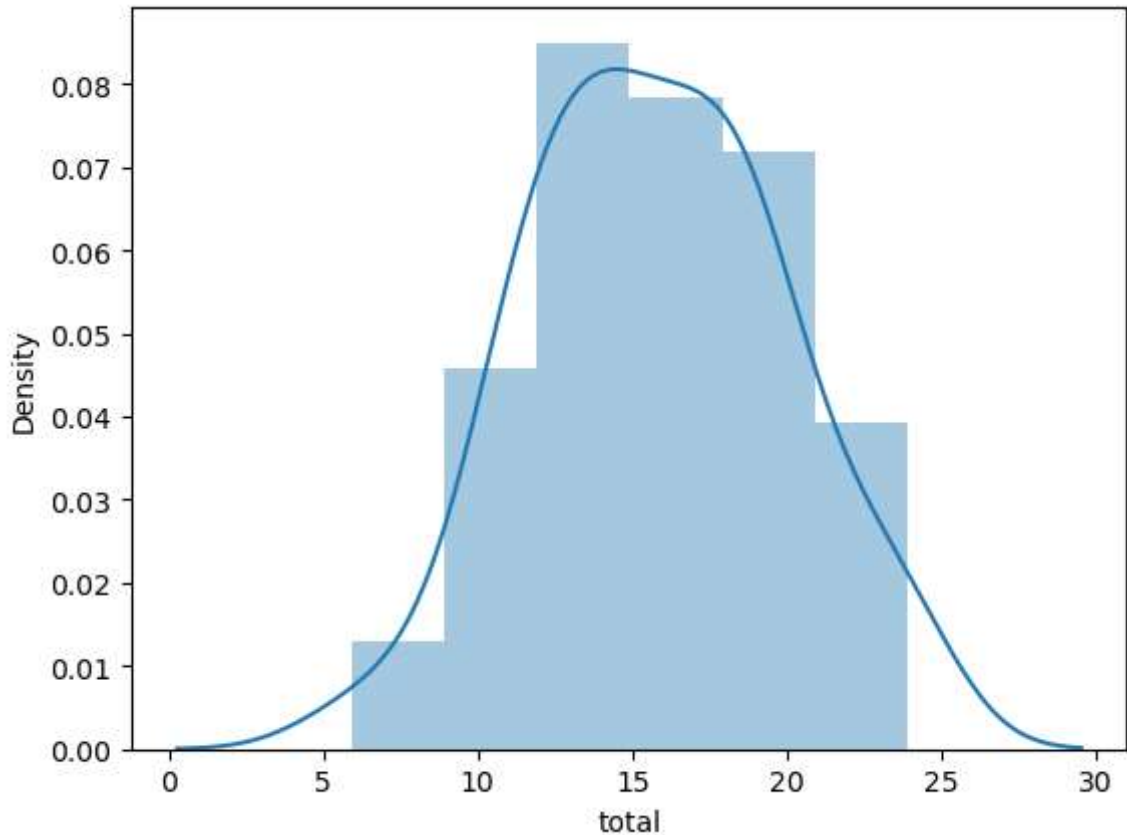


```
In [11]: sns.distplot(data['total'])
```

C:\Users\hp\OneDrive\Documents\Python Scripts\lib\site-packages\seaborn\distributions.py:2619: FutureWarning: `distplot` is a deprecated function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).

warnings.warn(msg, FutureWarning)

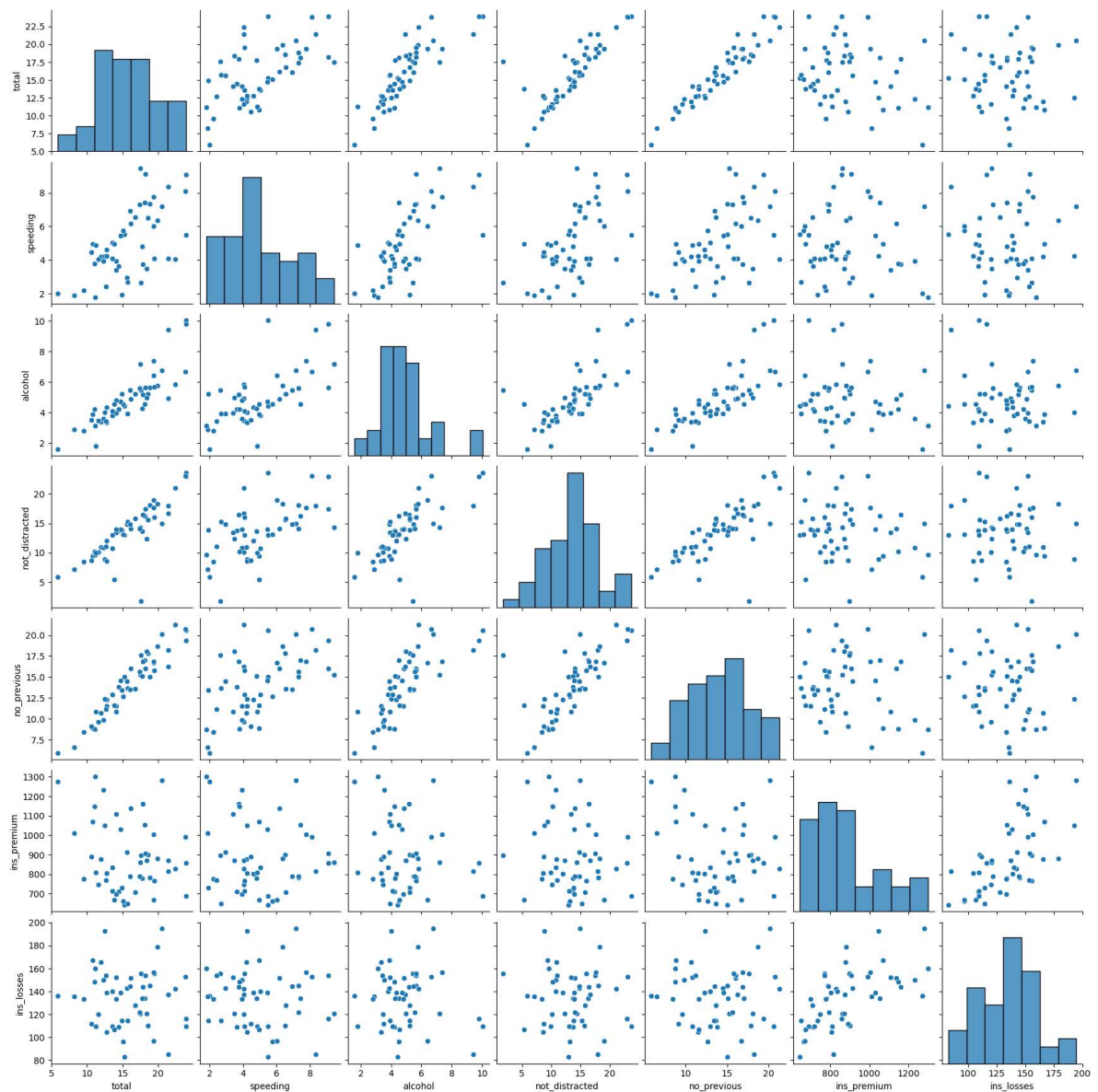
```
Out[11]: <AxesSubplot:xlabel='total', ylabel='Density'>
```



```
In [12]: #BELOW PLOTS ARE RELATION BETWEEN EVERY 2 VARIABLES IN THE DATASET
```

```
In [14]: sns.pairplot(data)
```

```
Out[14]: <seaborn.axisgrid.PairGrid at 0x1e3409edb50>
```



```
In [15]: # DONE BY CHARAN ADIMALLA(21BCE9482)
```

```
In [ ]:
```