

## assignment-02

September 13, 2023

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

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

```
[32]: data.head()
```

```
[32]:
```

	total	speeding	alcohol	not_distracted	no_previous	ins_premium	\
0	18.8	7.332	5.640	18.048	15.040	784.55	
1	18.1	7.421	4.525	16.290	17.014	1053.48	
2	18.6	6.510	5.208	15.624	17.856	899.47	
3	22.4	4.032	5.824	21.056	21.280	827.34	
4	12.0	4.200	3.360	10.920	10.680	878.41	

	ins_losses	abbrev
0	145.08	AL
1	133.93	AK
2	110.35	AZ
3	142.39	AR
4	165.63	CA

```
[33]: 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
```

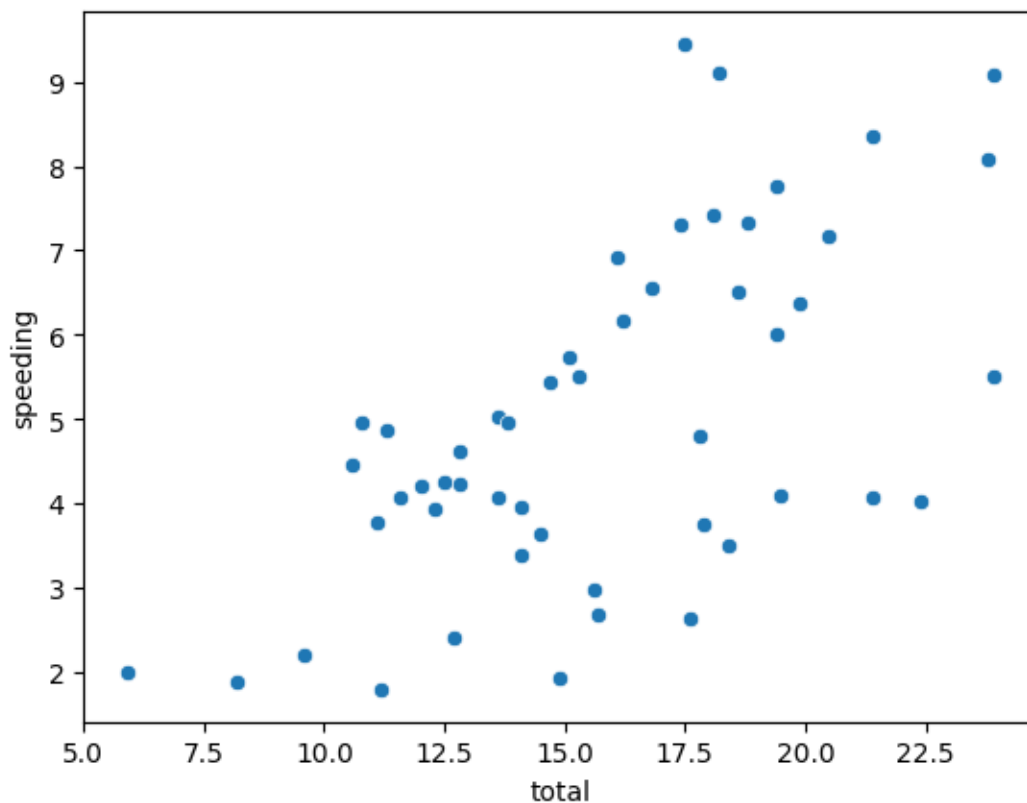
```
dtypes: float64(7), object(1)
memory usage: 3.3+ KB
```

we can plot the number of drivers involved in fatal collision(total) against the percentage of drivers who were speeding(speeding).

```
[34]: '''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)
```

```
[34]: <Axes: xlabel='total', ylabel='speeding'>
```

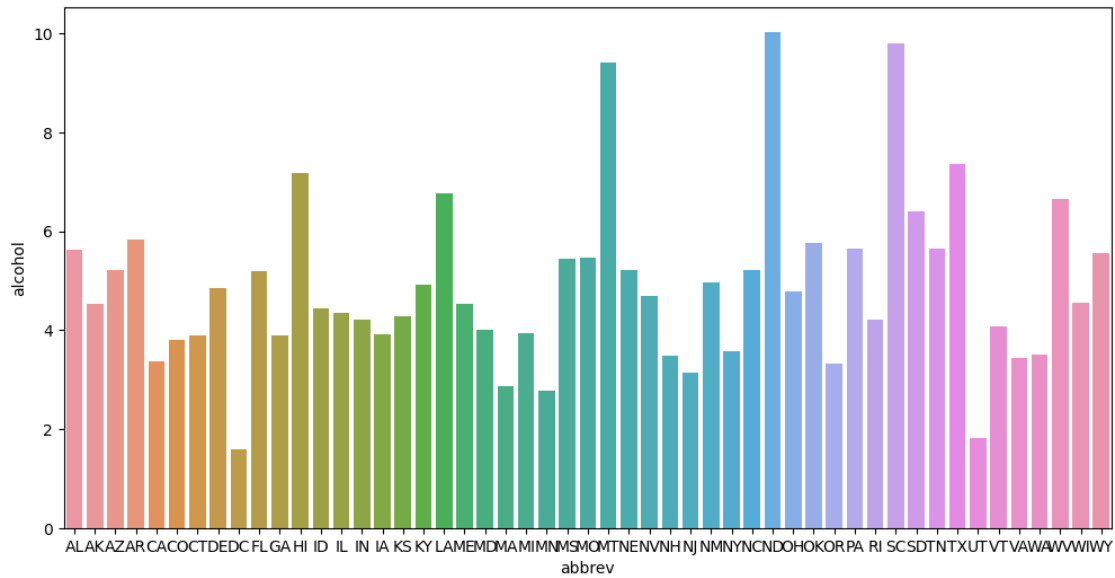


```
[35]: ''' Inference: the percentage of drivers who were alcohol-impaired varies
      ↳significantly from state to state.
      This suggests that alcohol-impaired driving is a major factor in fatal
      ↳collisions in some states, but not others. '''
```

```
plt.figure(figsize=(12,6))
```

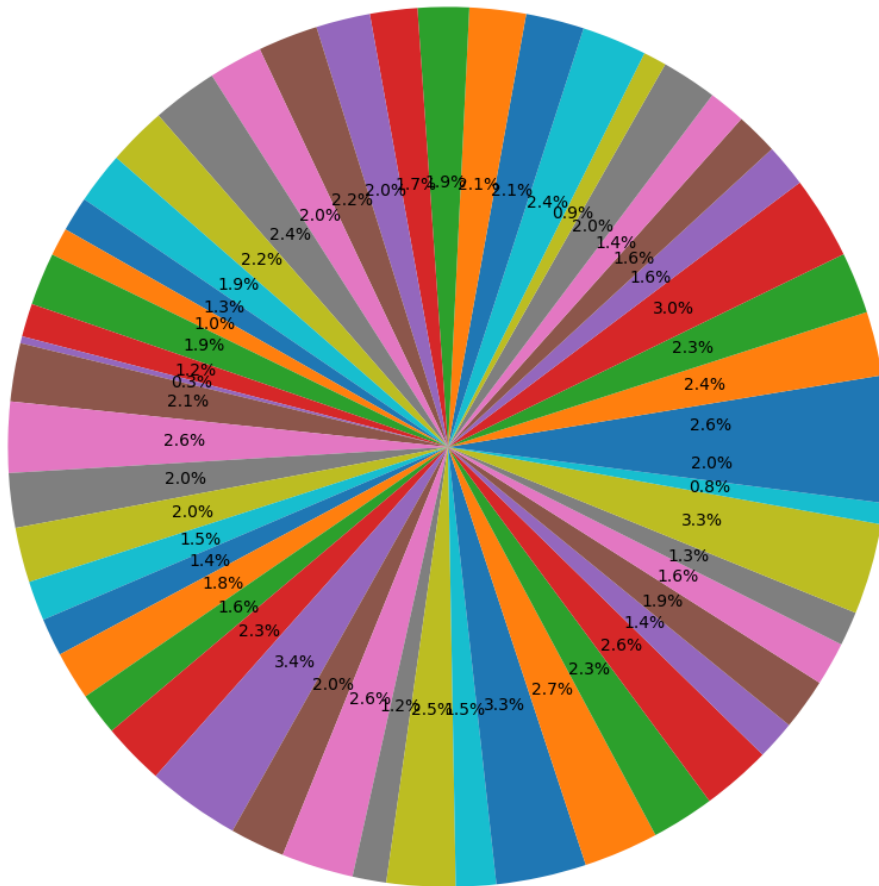
```
sns.barplot(x = 'abbrev', y = 'alcohol', data = data)
```

[35]: <Axes: xlabel='abbrev', ylabel='alcohol'>



[36]: *''' 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()
```



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

C:\Users\sivar\AppData\Local\Temp\ipykernel\_4884\3477427589.py:1: UserWarning:

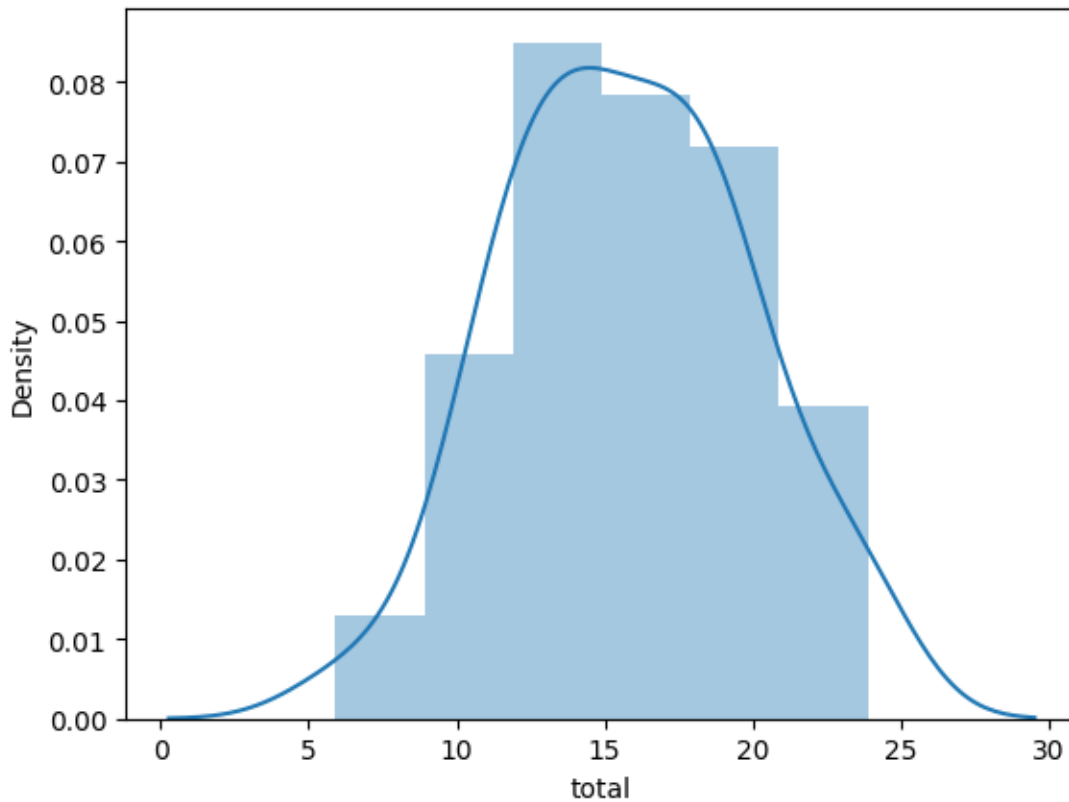
`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see <https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751>

```
sns.distplot(data['total'])
```

```
[37]: <Axes: xlabel='total', ylabel='Density'>
```



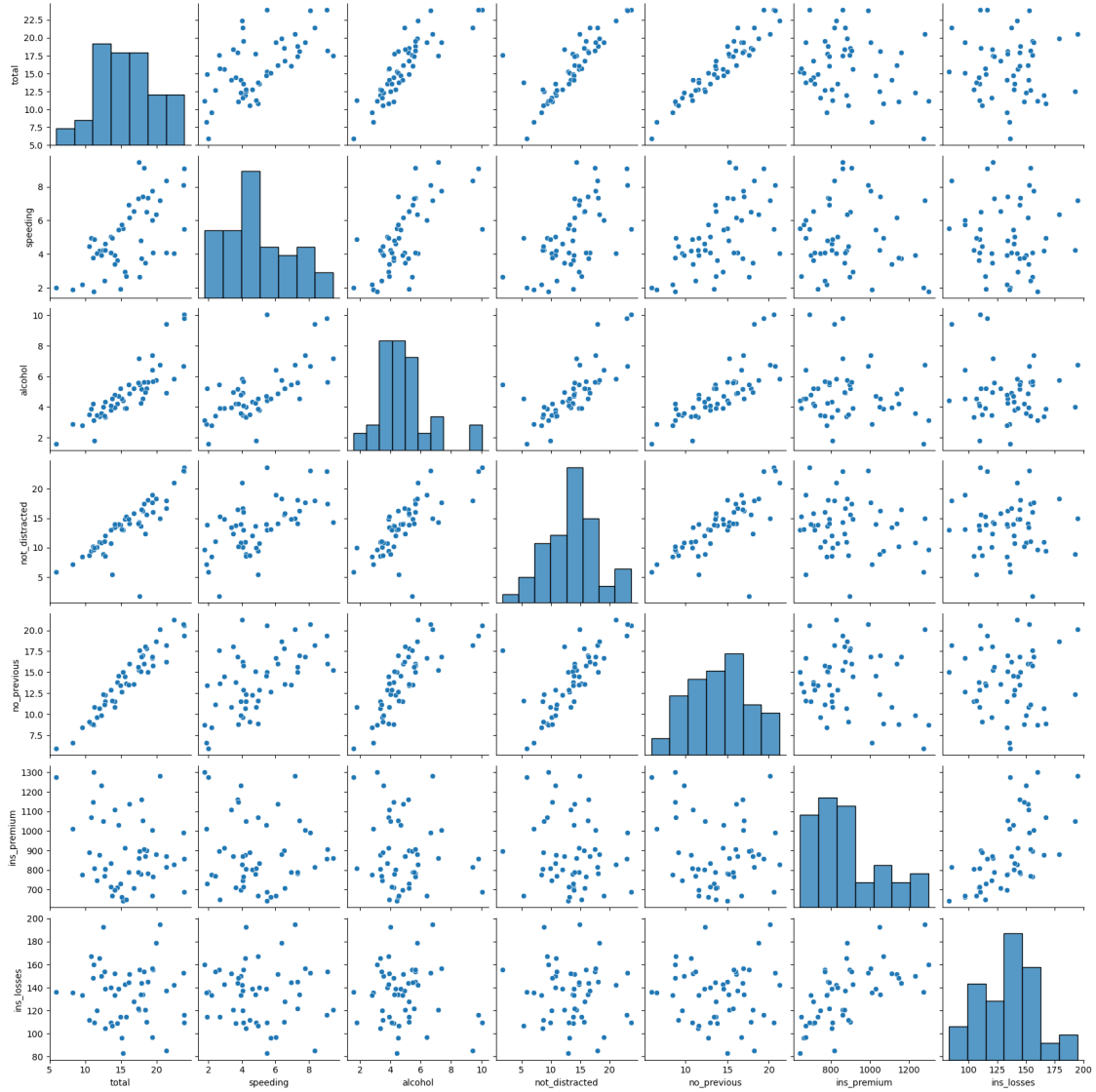
Below plots are relation between every two variables in the dataset

```
[38]: sns.pairplot(data)
```

```
c:\Users\sivar\AppData\Local\Programs\Python\Python311\Lib\site-packages\seaborn\axisgrid.py:118: UserWarning: The figure layout has changed to tight
```

```
self._figure.tight_layout(*args, **kwargs)
```

```
[38]: <seaborn.axisgrid.PairGrid at 0x17f74068610>
```



Completed by : Maguluri Venkata Siva Rama Krishna (21BCE9322)