

```
import seaborn as sns
```

```
df=sns.load_dataset('car_crashes')
df.info()
```

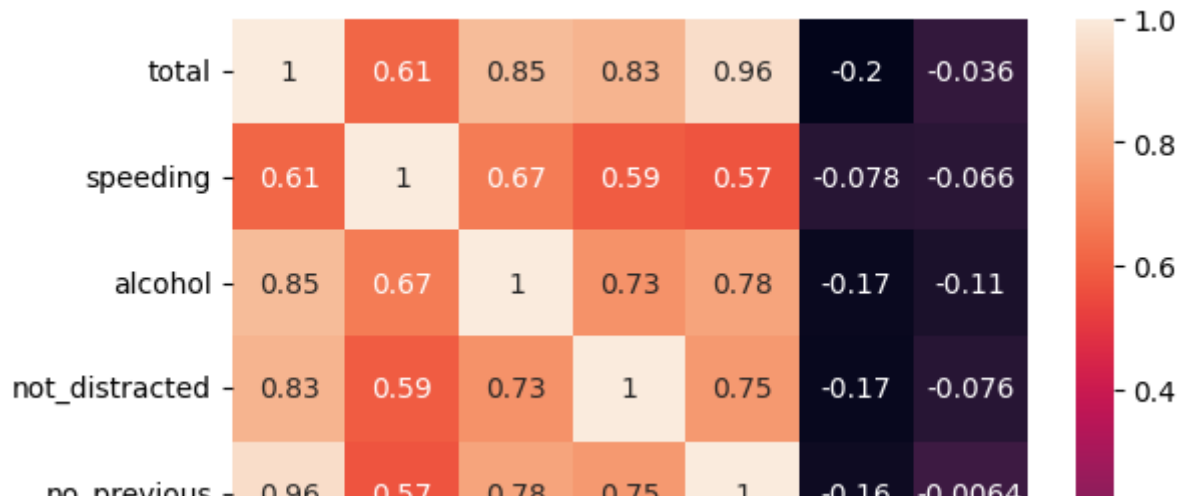
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 51 entries, 0 to 50
Data columns (total 8 columns):
#   Column                Non-Null Count  Dtype
---  -
0   total                  51 non-null    float64
1   speeding               51 non-null    float64
2   alcohol                51 non-null    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
```

```
df.head()
```

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

```
correlation = df.corr()
sns.heatmap(correlation,annot=True)
```

```
<ipython-input-89-a1367c46b0c7>:1: FutureWarning: The default value of numeric_only is deprecated
correlation = df.corr()
<Axes: >
```



Inference: The heatmap shows the correlation between all the features.

High positive correlation (features are directly proportional) is observed when the value of coefficient is greater than 0.5

High negative correlation (features are indirectly proportional) is observed when the value of coefficient is greater than -0.5

Here, a general inference is that all causes of accidents is more in the states where the total accidents are more, especially, accidents where the drivers weren't involved in any previous accident (Since, corr. between total and no_previous = 0.96)

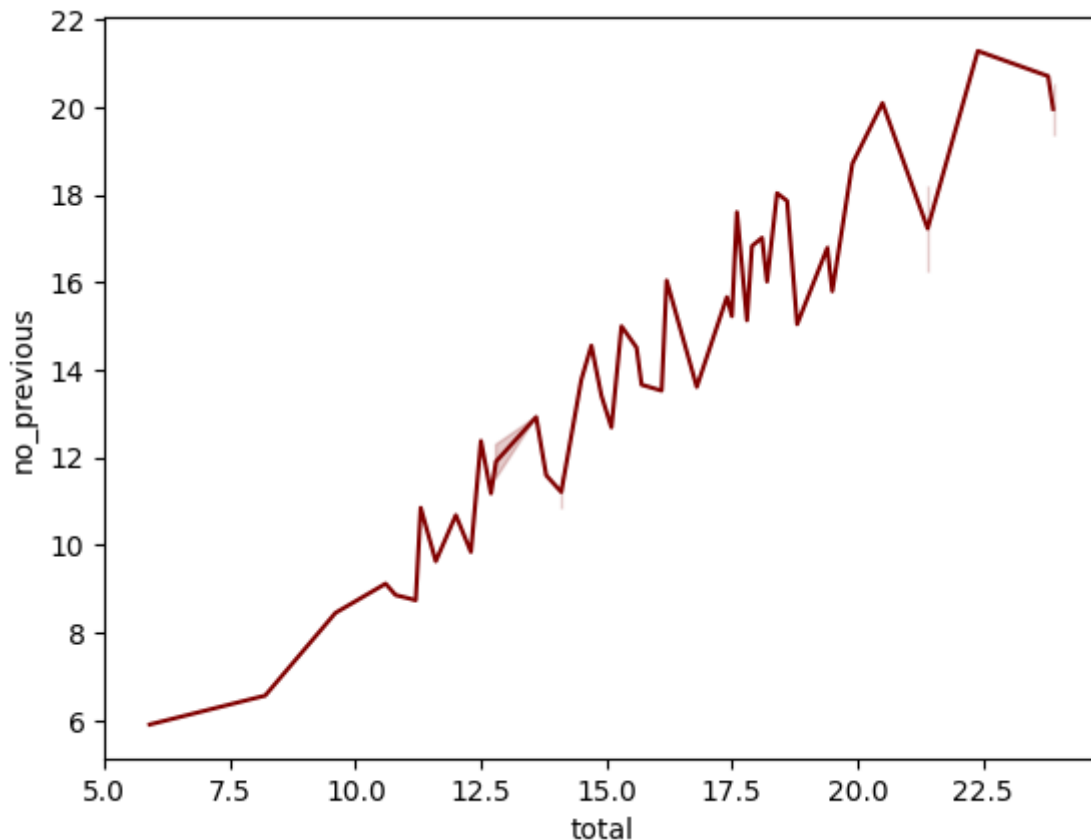
```
sns.scatterplot(x='speeding',y='alcohol',data=df,color='crimson')
```

```
<Axes: xlabel='speeding', ylabel='alcohol'>
```

Inference: The percentages of alcohol-related incidents and instances of speeding among drivers are directly proportional for most of the states, this suggests drunk drivers are more prone to exceeding speed limits causing accidents.

```
sns.lineplot(x='total',y='no_previous',data=df,color='maroon')
```

```
<Axes: xlabel='total', ylabel='no_previous'>
```



Inference: The states with more total accidents, in general, tend to have more accidents where drivers have no previous record of involvement in such accidents.

```
sns.distplot(df["total"],color='skyblue')
```

```
<ipython-input-92-eee39b9eef3b>:1: UserWarning:
```

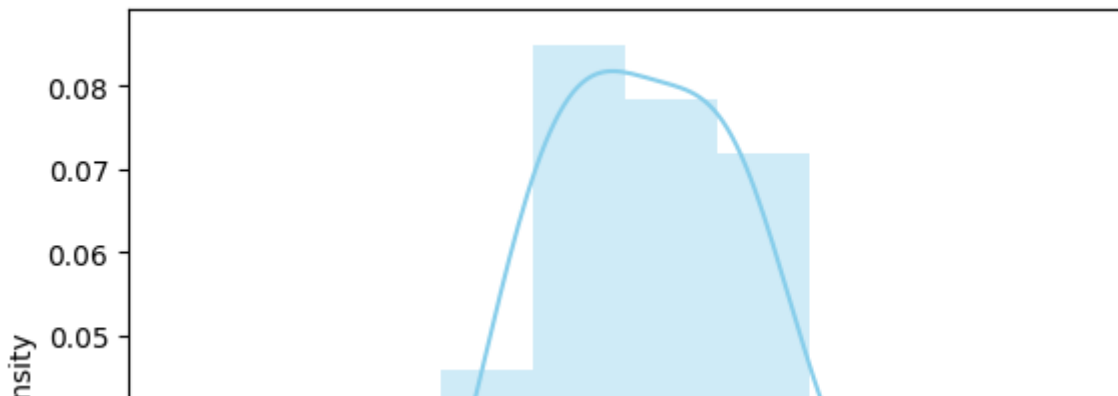
```
`distplot` is a deprecated function and will be removed in seaborn v0.14.0.
```

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(df["total"],color='skyblue')
<Axes: xlabel='total', ylabel='Density'>
```



Inference: Most of the states have a total car crashes somewhere in the range of 10-15. The peak represents the mode.

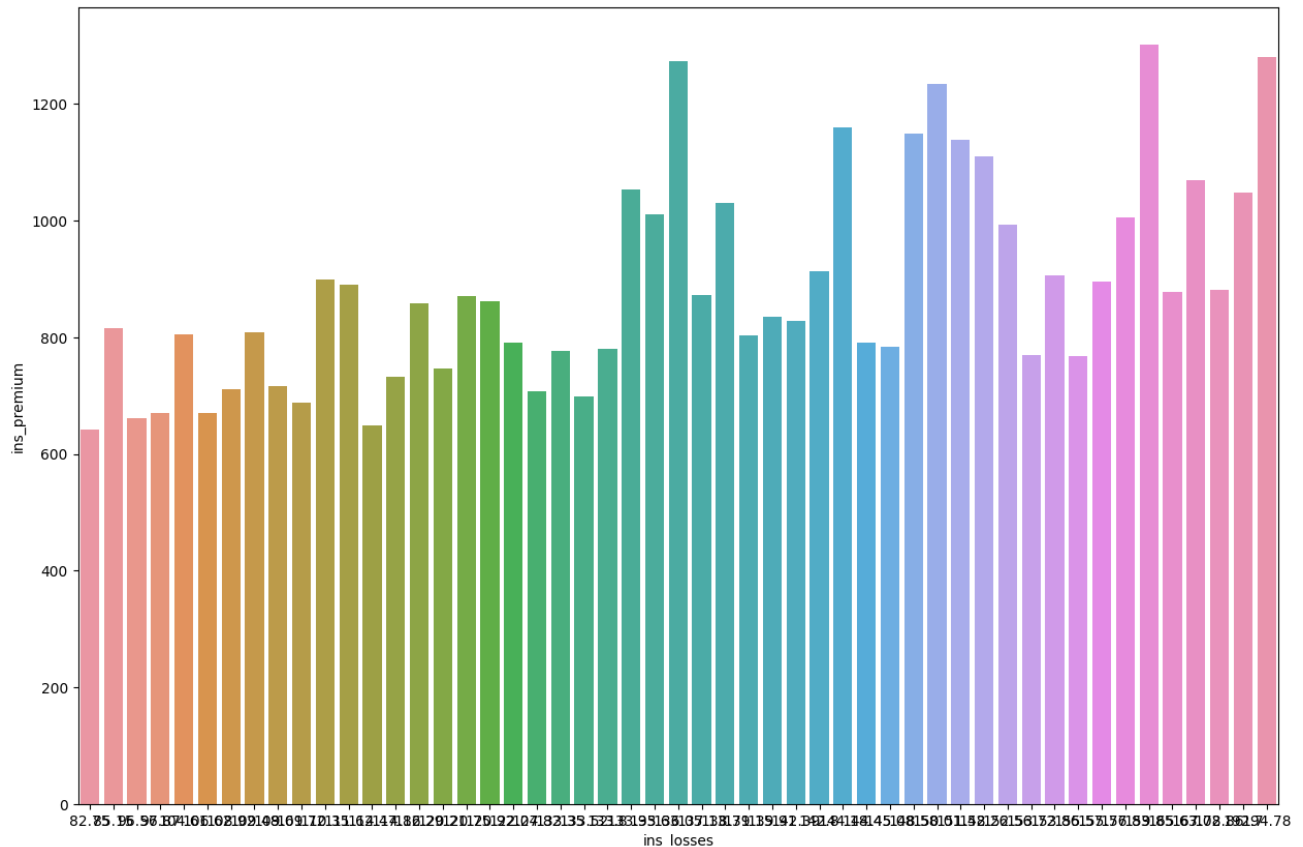
```
sns.relplot(x='speeding',y='not_distracted',data=df,hue='no_previous')
```

```
<seaborn.axisgrid.FacetGrid at 0x7d0e4b8e9720>
```

Inference: This plot suggests that among the accidents caused by speeding, the states where percentage of drivers who were not distracted also had high percentage of drivers who had no previous accidents.

```
import matplotlib.pyplot as plt
plt.subplots(figsize=(15,10))
sns.barplot(x='ins_losses',y='ins_premium',data=df)
```

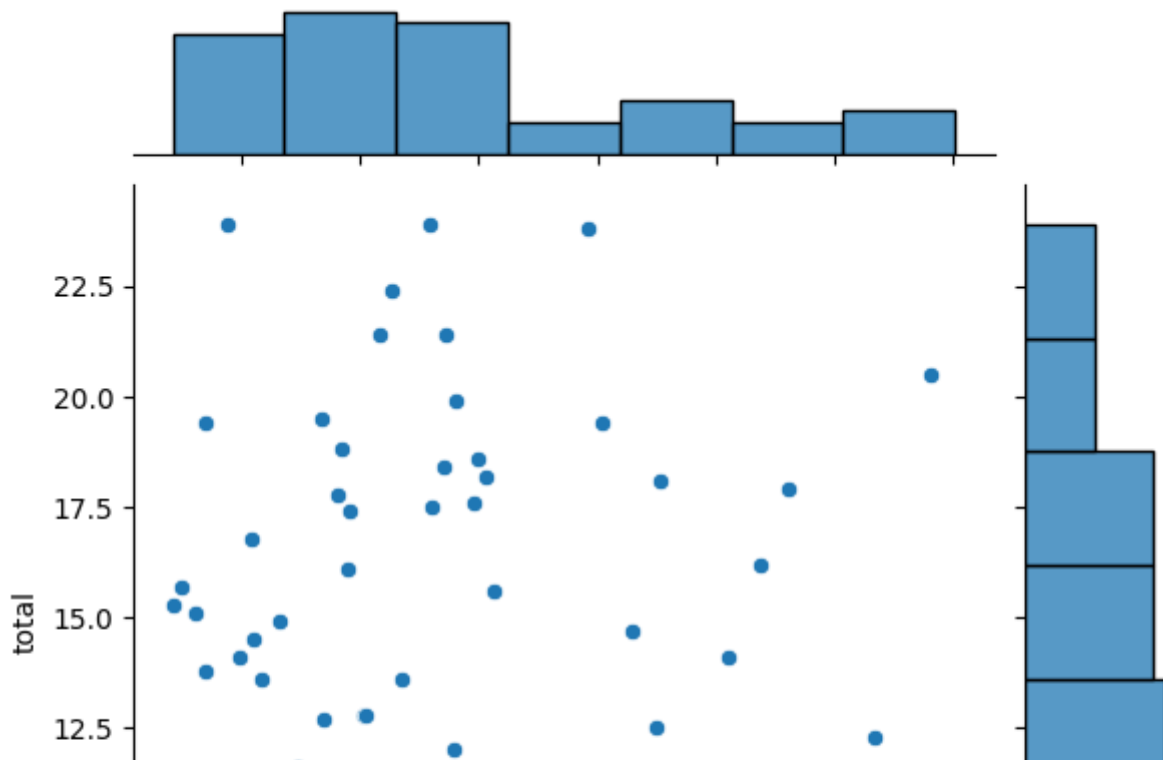
```
<Axes: xlabel='ins_losses', ylabel='ins_premium'>
```



Inference: Although there is no specific relationship between insurance premium and insurance losses, it can be observed that the likelihood of encountering higher insurance premiums within a particular state is more if the insurance loss incurred is high.

```
sns.jointplot(x='ins_premium',y='total',data=df)
```

<seaborn.axisgrid.JointGrid at 0x7d0e4c42f970>

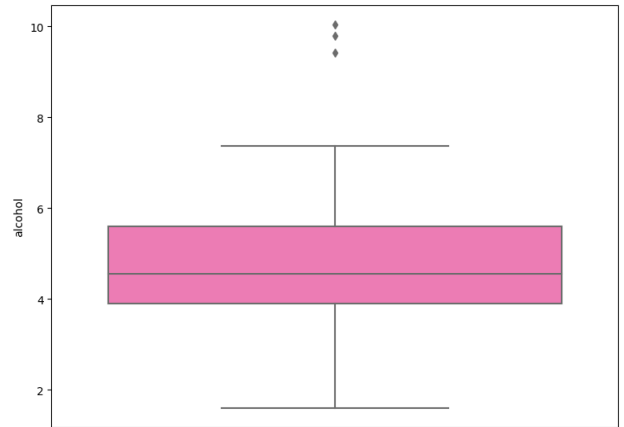
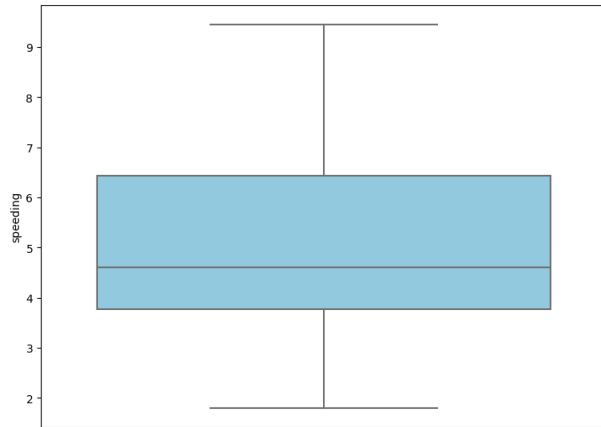


Inference: Most states have total accidents between 11.25 and 18.75, while most insurance premiums are between 600 and 920 (Univariate data analysis)

It can be observed that places with low insurance premium tend to have more accidents. (Bivariate data analysis)

```
plt.subplots(figsize=(20,15))
plt.subplot(2,2,1)
sns.boxplot(y='speeding',data=df,color='skyblue')
plt.subplot(2,2,2)
sns.boxplot(y='alcohol',data=df,color='hotpink')
plt.subplot(2,2,3)
sns.boxplot(y='not_distracted',data=df,color='yellow')
plt.subplot(2,2,4)
sns.boxplot(y='no_previous',data=df,color='lime')
```

```
<ipython-input-96-1c4deeba3ccc>:2: MatplotlibDeprecationWarning: Auto-removal of over
plt.subplot(2,2,1)
<Axes: ylabel='no_previous'>
```



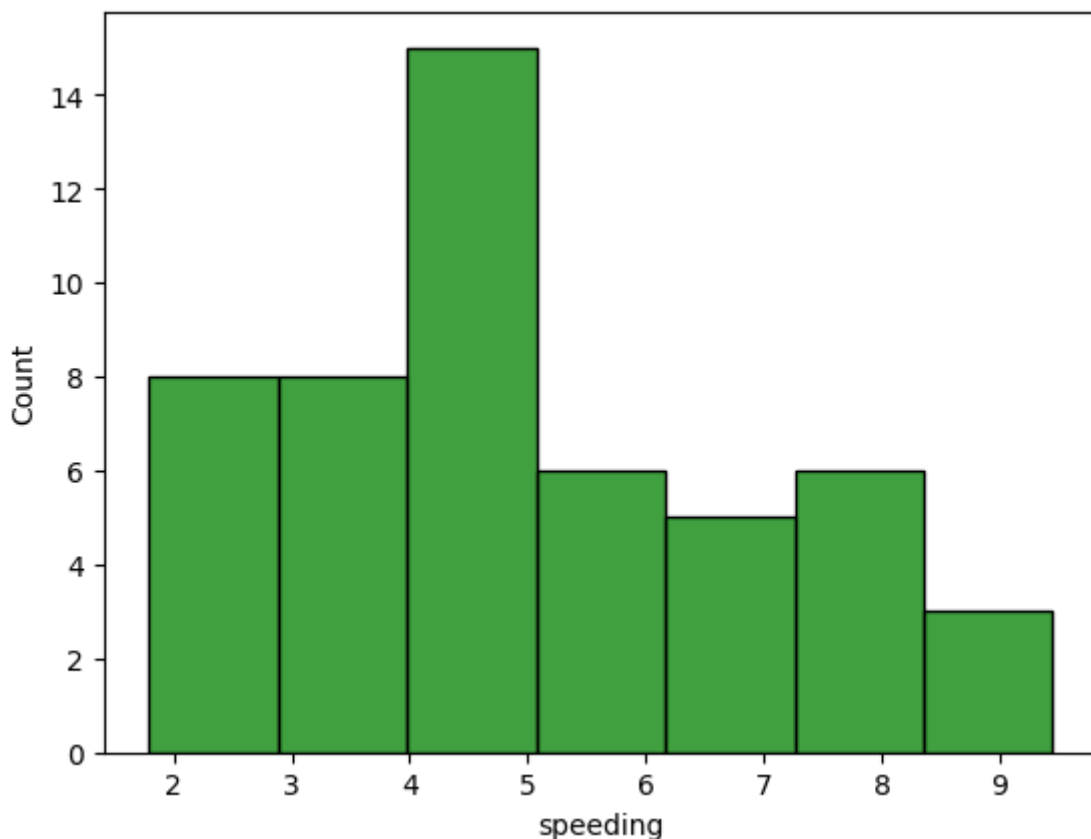
Inference: The boxplots show the the median, quartiles, and the maximum and minimum value for all the four features.

Among alcohol feature, there are 3 outliers while not_distracted feature has 1 outlier.



```
sns.histplot(df['speeding'],color='green')
```

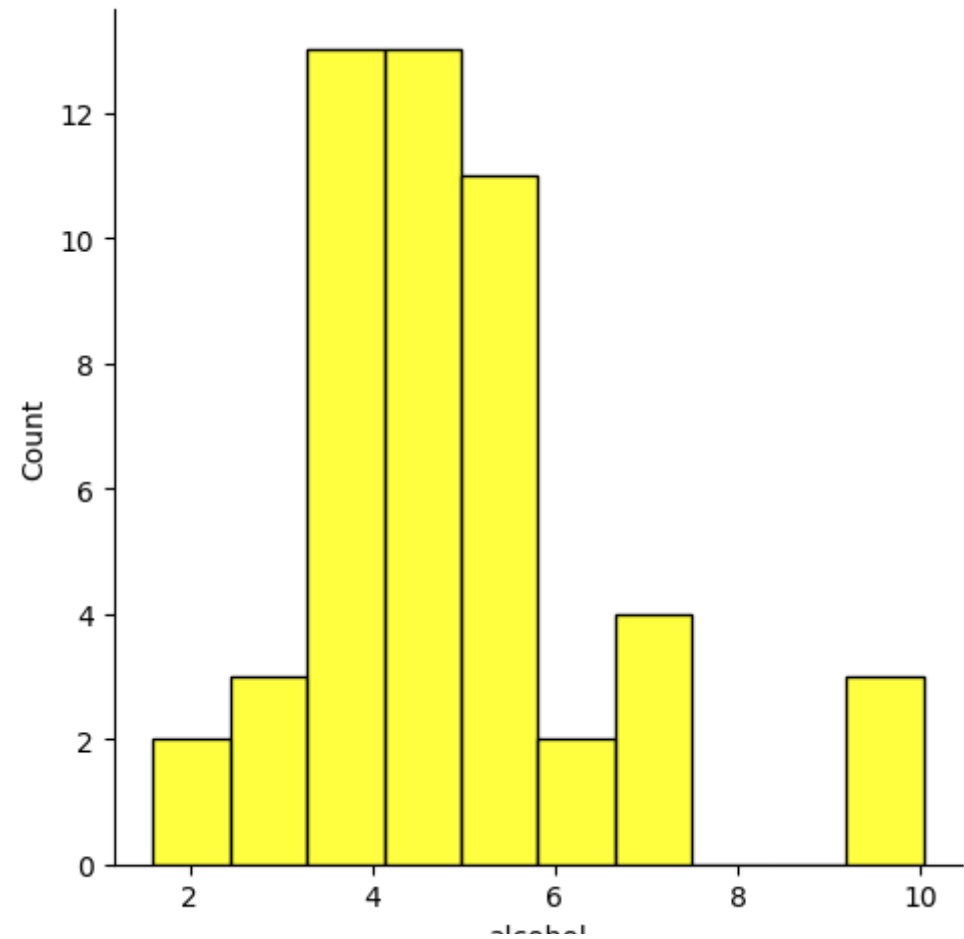
```
<Axes: xlabel='speeding', ylabel='Count'>
```



Inference: Most states have the percentage of accidents caused by speeding between 4 and 5.

```
sns.displot(df['alcohol'],color='yellow')
```

<seaborn.axisgrid.FacetGrid at 0x7d0e583bead0>



Inference: Most states have the percentage of accidents caused by drunk drivers between 3 and 5.