In [ ]:

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

In [ ]:

```
print(sns.get_dataset_names())
```

```
['anagrams', 'anscombe', 'attention', 'brain_networks', 'car_crashes', 'dia
monds', 'dots', 'dowjones', 'exercise', 'flights', 'fmri', 'geyser', 'glue'
, 'healthexp', 'iris', 'mpg', 'penguins', 'planets', 'seaice', 'taxis', 'ti
ps', 'titanic']
```

In [ ]:

```
dataset=sns.load_dataset('car_crashes')
dataset
```

Out[ ]:

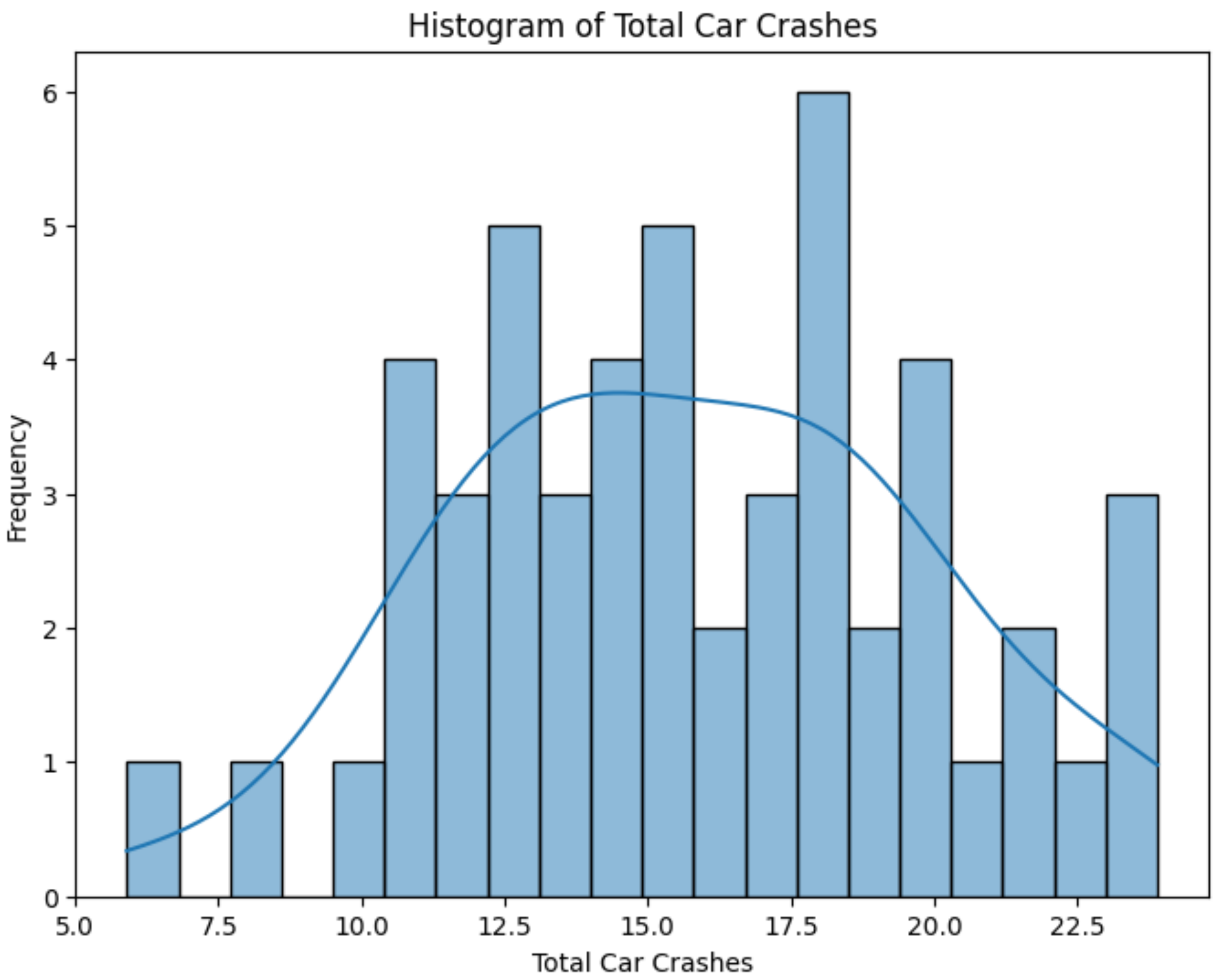
	total	speeding	alcohol	not_distracted	no_previous	ins_premium	ins_losses	abbrev
0	18.8	7.332	5.640	18.048	15.040	784.55	145.08	AL
1	18.1	7.421	4.525	16.290	17.014	1053.48	133.93	AK
2	18.6	6.510	5.208	15.624	17.856	899.47	110.35	AZ
3	22.4	4.032	5.824	21.056	21.280	827.34	142.39	AR
4	12.0	4.200	3.360	10.920	10.680	878.41	165.63	CA
5	13.6	5.032	3.808	10.744	12.920	835.50	139.91	CO
6	10.8	4.968	3.888	9.396	8.856	1068.73	167.02	CT
7	16.2	6.156	4.860	14.094	16.038	1137.87	151.48	DE
8	5.9	2.006	1.593	5.900	5.900	1273.89	136.05	DC
9	17.9	3.759	5.191	16.468	16.826	1160.13	144.18	FL
10	15.6	2.964	3.900	14.820	14.508	913.15	142.80	GA
11	17.5	9.450	7.175	14.350	15.225	861.18	120.92	HI
12	15.3	5.508	4.437	13.005	14.994	641.96	82.75	ID
13	12.8	4.608	4.352	12.032	12.288	803.11	139.15	IL
14	14.5	3.625	4.205	13.775	13.775	710.46	108.92	IN
15	15.7	2.669	3.925	15.229	13.659	649.06	114.47	IA
16	17.8	4.806	4.272	13.706	15.130	780.45	133.80	KS
17	21.4	4.066	4.922	16.692	16.264	872.51	137.13	KY
18	20.5	7.175	6.765	14.965	20.090	1281.55	194.78	LA
19	15.1	5.738	4.530	13.137	12.684	661.88	96.57	ME
20	12.5	4.250	4.000	8.875	12.375	1048.78	192.70	MD
21	8.2	1.886	2.870	7.134	6.560	1011.14	135.63	MA

22	14.1	3.384	3.948	13.395	10.857	1110.61	152.26	MI
23	9.6	2.208	2.784	8.448	8.448	777.18	133.35	MN
24	17.6	2.640	5.456	1.760	17.600	896.07	155.77	MS
25	16.1	6.923	5.474	14.812	13.524	790.32	144.45	MO
26	21.4	8.346	9.416	17.976	18.190	816.21	85.15	MT
27	14.9	1.937	5.215	13.857	13.410	732.28	114.82	NE
28	14.7	5.439	4.704	13.965	14.553	1029.87	138.71	NV
29	11.6	4.060	3.480	10.092	9.628	746.54	120.21	NH
30	11.2	1.792	3.136	9.632	8.736	1301.52	159.85	NJ
31	18.4	3.496	4.968	12.328	18.032	869.85	120.75	NM
32	12.3	3.936	3.567	10.824	9.840	1234.31	150.01	NY
33	16.8	6.552	5.208	15.792	13.608	708.24	127.82	NC
34	23.9	5.497	10.038	23.661	20.554	688.75	109.72	ND
35	14.1	3.948	4.794	13.959	11.562	697.73	133.52	OH
36	19.9	6.368	5.771	18.308	18.706	881.51	178.86	OK
37	12.8	4.224	3.328	8.576	11.520	804.71	104.61	OR
38	18.2	9.100	5.642	17.472	16.016	905.99	153.86	PA
39	11.1	3.774	4.218	10.212	8.769	1148.99	148.58	RI
40	23.9	9.082	9.799	22.944	19.359	858.97	116.29	SC
41	19.4	6.014	6.402	19.012	16.684	669.31	96.87	SD
42	19.5	4.095	5.655	15.990	15.795	767.91	155.57	TN
43	19.4	7.760	7.372	17.654	16.878	1004.75	156.83	TX
44	11.3	4.859	1.808	9.944	10.848	809.38	109.48	UT
45	13.6	4.080	4.080	13.056	12.920	716.20	109.61	VT
46	12.7	2.413	3.429	11.049	11.176	768.95	153.72	VA
47	10.6	4.452	3.498	8.692	9.116	890.03	111.62	WA
48	23.8	8.092	6.664	23.086	20.706	992.61	152.56	WV
49	13.8	4.968	4.554	5.382	11.592	670.31	106.62	WI
50	17.4	7.308	5.568	14.094	15.660	791.14	122.04	WY

Histogram of Total Car Crashes:

In this histogram, we can see that the most common range for total car crashes is approximately between 10,000 and 15,000 crashes, with a peak around 12,000 crashes. The distribution is right-skewed, indicating that a majority of states have lower total car crash counts, but a few states have significantly higher crash counts.

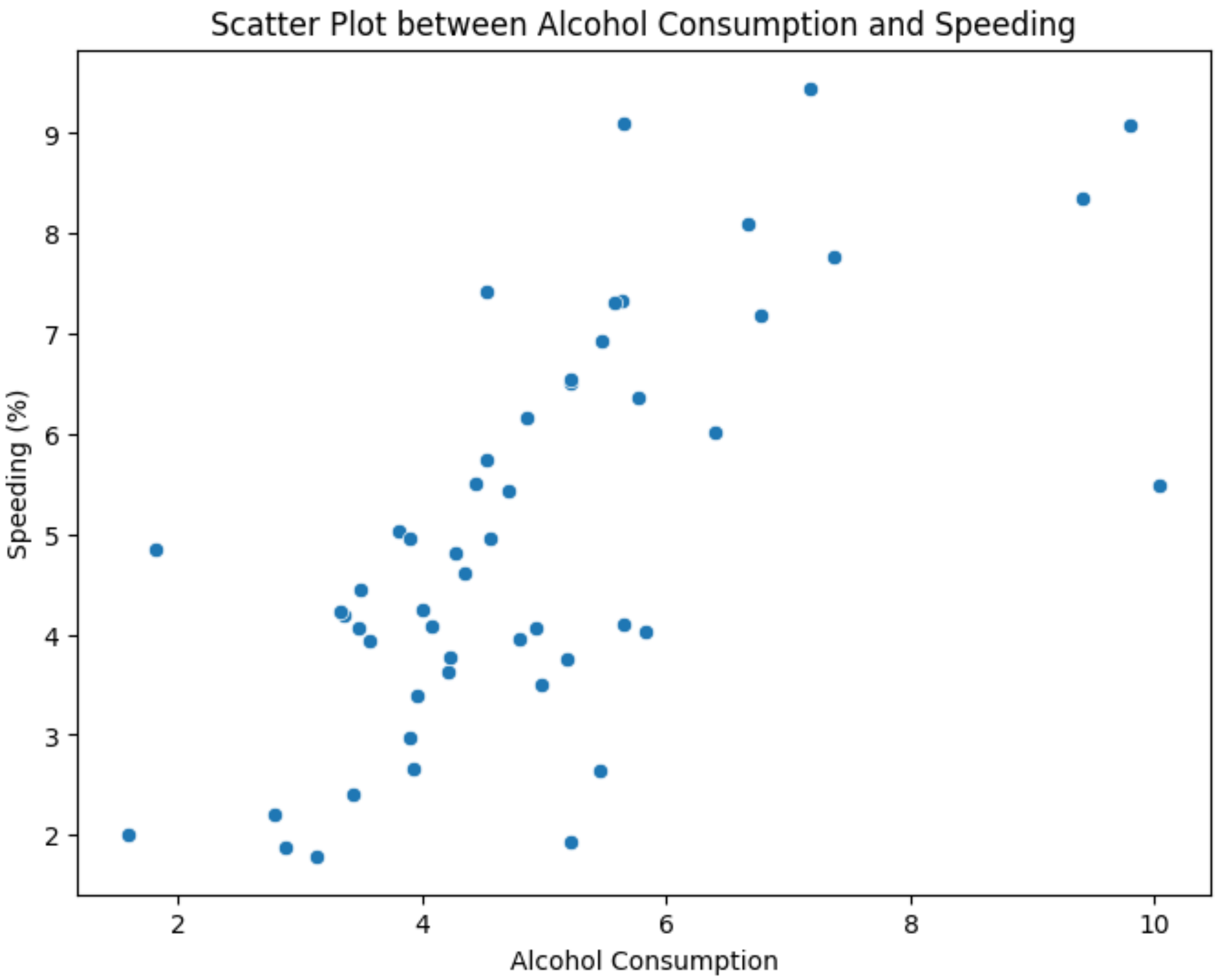
In [ ]:
```
# Example 1: Histogram of the number of car crashes
plt.figure(figsize=(8, 6))
sns.histplot(dataset['total'], bins=20, kde=True)
plt.xlabel('Total Car Crashes')
plt.ylabel('Frequency')
plt.title('Histogram of Total Car Crashes')
plt.show()
```



Scatter Plot between Alcohol Consumption and Speeding:

The scatter plot shows a weak positive correlation between alcohol consumption and the percentage of speeding-related accidents. This suggests that in states with higher alcohol consumption rates, there tends to be a slightly higher percentage of accidents caused by speeding.

In [ ]:
```
# Example 2: Scatter plot between alcohol consumption and speeding
plt.figure(figsize=(8, 6))
sns.scatterplot(x='alcohol', y='speeding', data=dataset)
plt.xlabel('Alcohol Consumption')
plt.ylabel('Speeding (%)')
plt.title('Scatter Plot between Alcohol Consumption and Speeding')
plt.show()
```

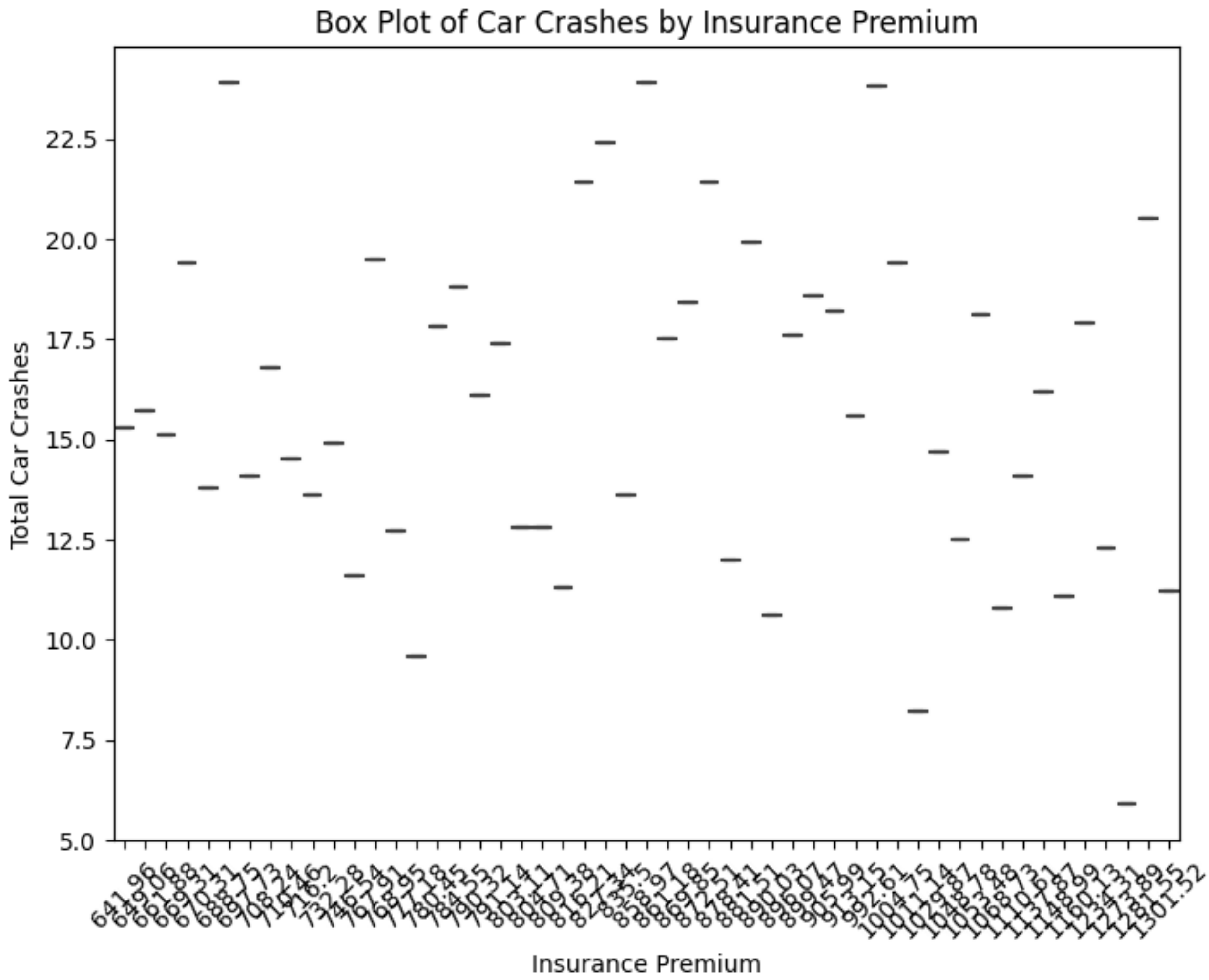


Box Plot of Car Crashes by Insurance Premium:

The box plot illustrates the distribution of total car crashes across different insurance premium levels. States with lower insurance premiums generally have a wider range of total crashes, including both low and high counts. States with higher insurance premiums have a more consistent range of total crashes, with fewer outliers.

In [ ]:

```
# Example 3: Box plot of car crashes by insurance premium
plt.figure(figsize=(8, 6))
sns.boxplot(x='ins_premium', y='total', data=dataset)
plt.xlabel('Insurance Premium')
plt.ylabel('Total Car Crashes')
plt.title('Box Plot of Car Crashes by Insurance Premium')
plt.xticks(rotation=45)
plt.show()
```

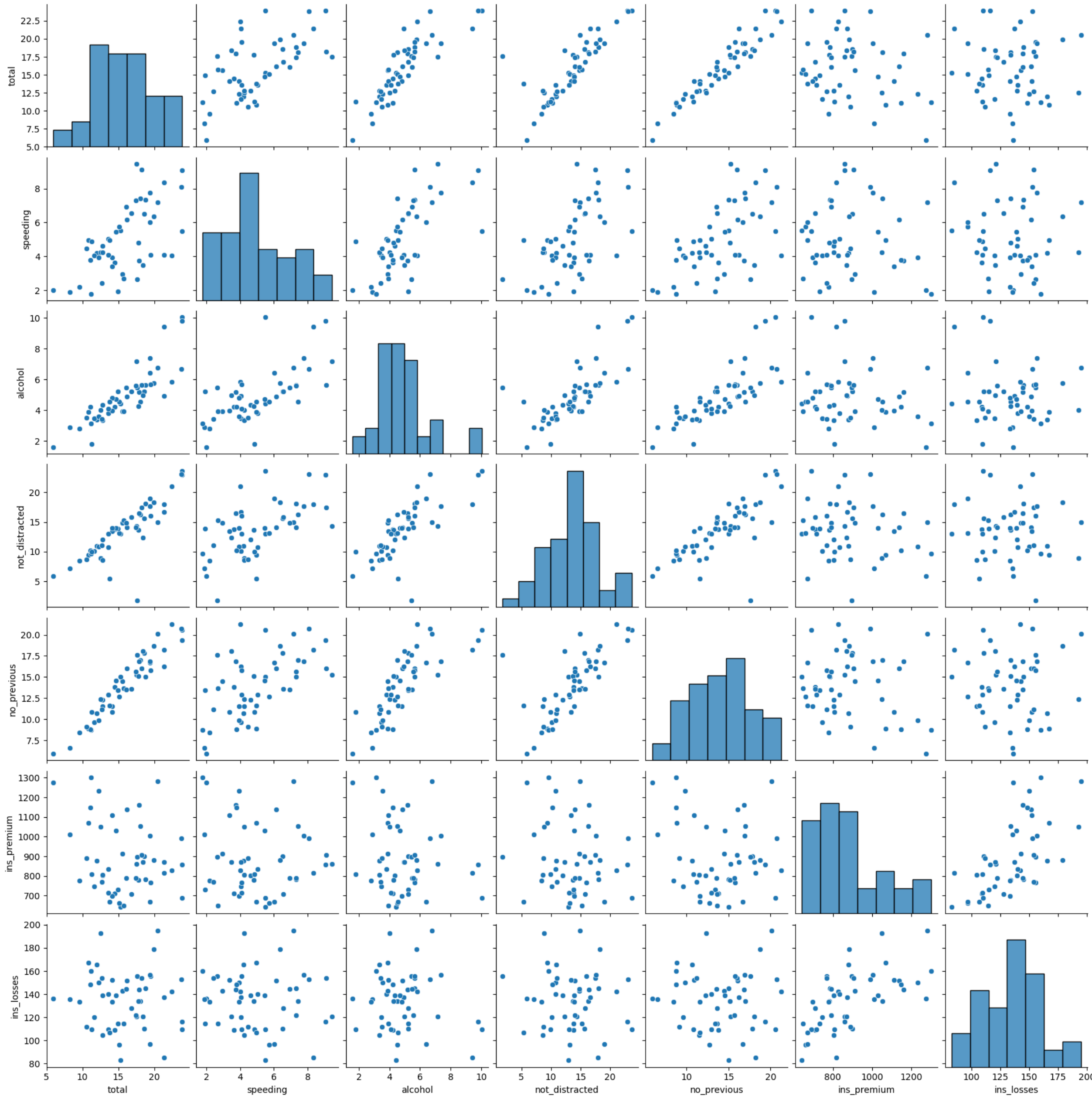


Pairplot (Multiple Variable Relationships):

The pairplot provides a comprehensive view of relationships between multiple variables in the dataset. Notable patterns include a positive correlation between 'total' car crashes and 'alcohol' consumption, as well as a negative correlation between 'total' car crashes and 'ins_premium.'

In [ ]:

```
sns.pairplot(dataset)
plt.show()
```

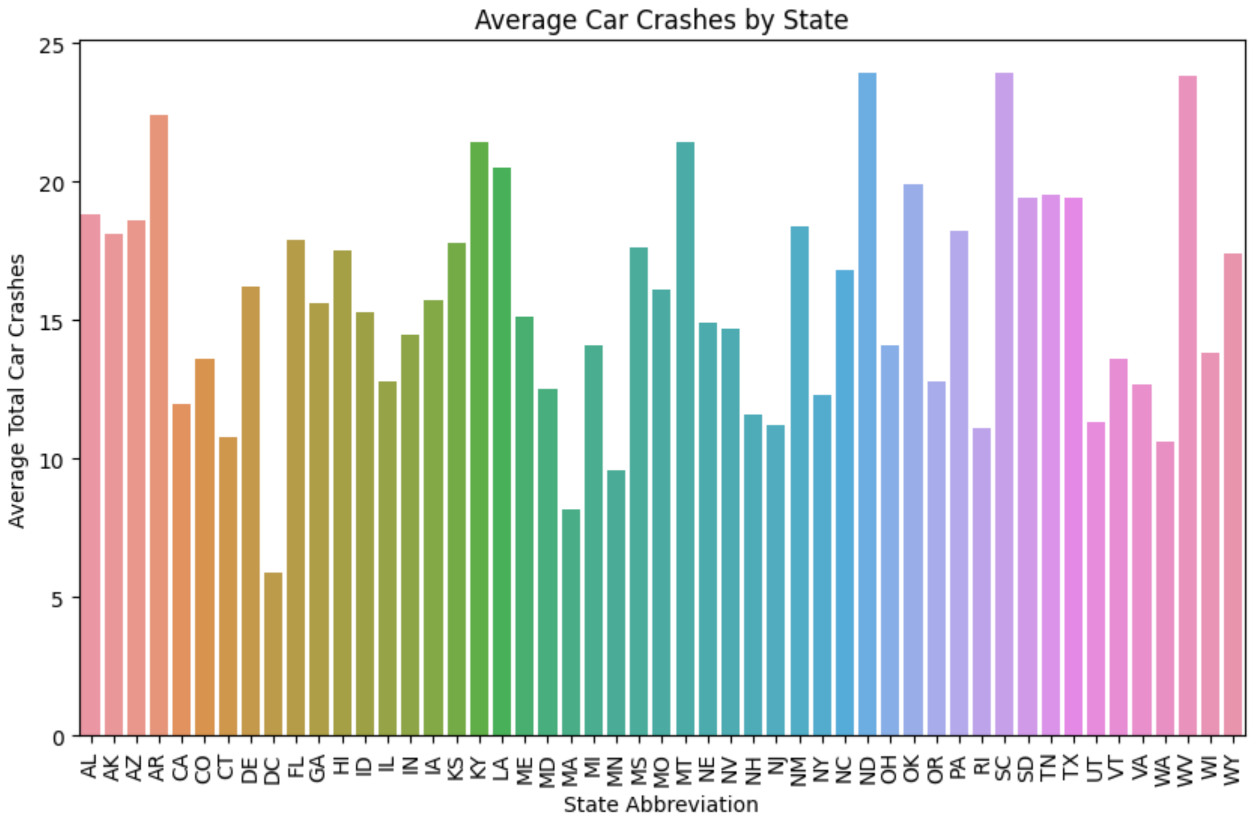


Bar Plot of Average Car Crashes by State (Abbreviation):

This bar plot displays the average total car crashes for each state, represented by their abbreviations. States like 'DC' (District of Columbia) and 'RI' (Rhode Island) have notably higher average crash counts compared to other states. Conversely, 'ND' (North Dakota) and 'VT' (Vermont) have lower average crash counts.

In [ ]:

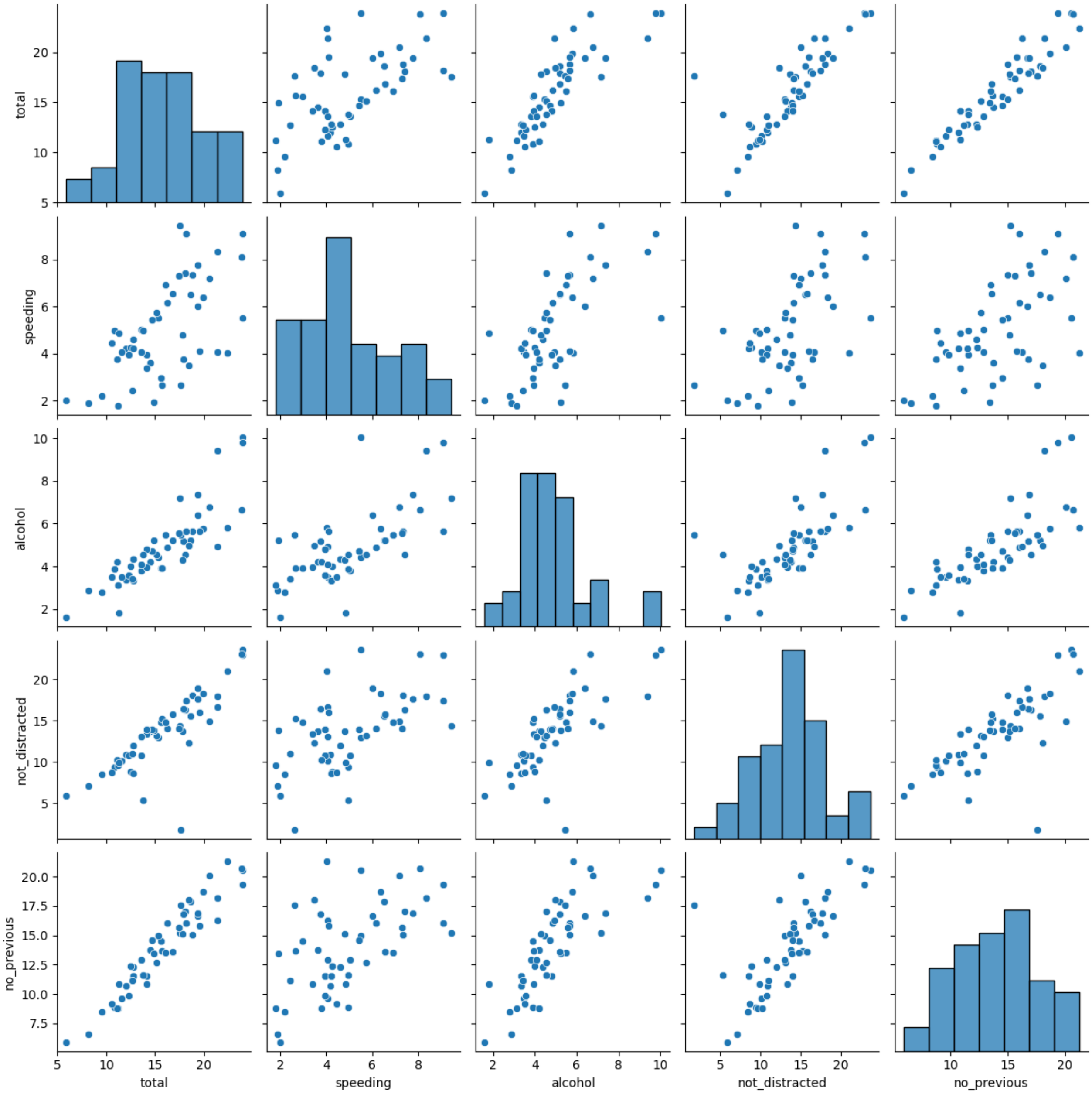
```
plt.figure(figsize=(10, 6))
sns.barplot(x='abbrev', y='total', data=dataset)
plt.xlabel('State Abbreviation')
plt.ylabel('Average Total Car Crashes')
plt.title('Average Car Crashes by State')
plt.xticks(rotation=90)
plt.show()
```



The pairplot shows scatterplots and histograms for the numeric variables. It appears that there is a positive correlation between 'alcohol' and 'speeding,' while 'not_distracted' and 'no_previous' show some patterns of negative correlation with 'total' car crashes.

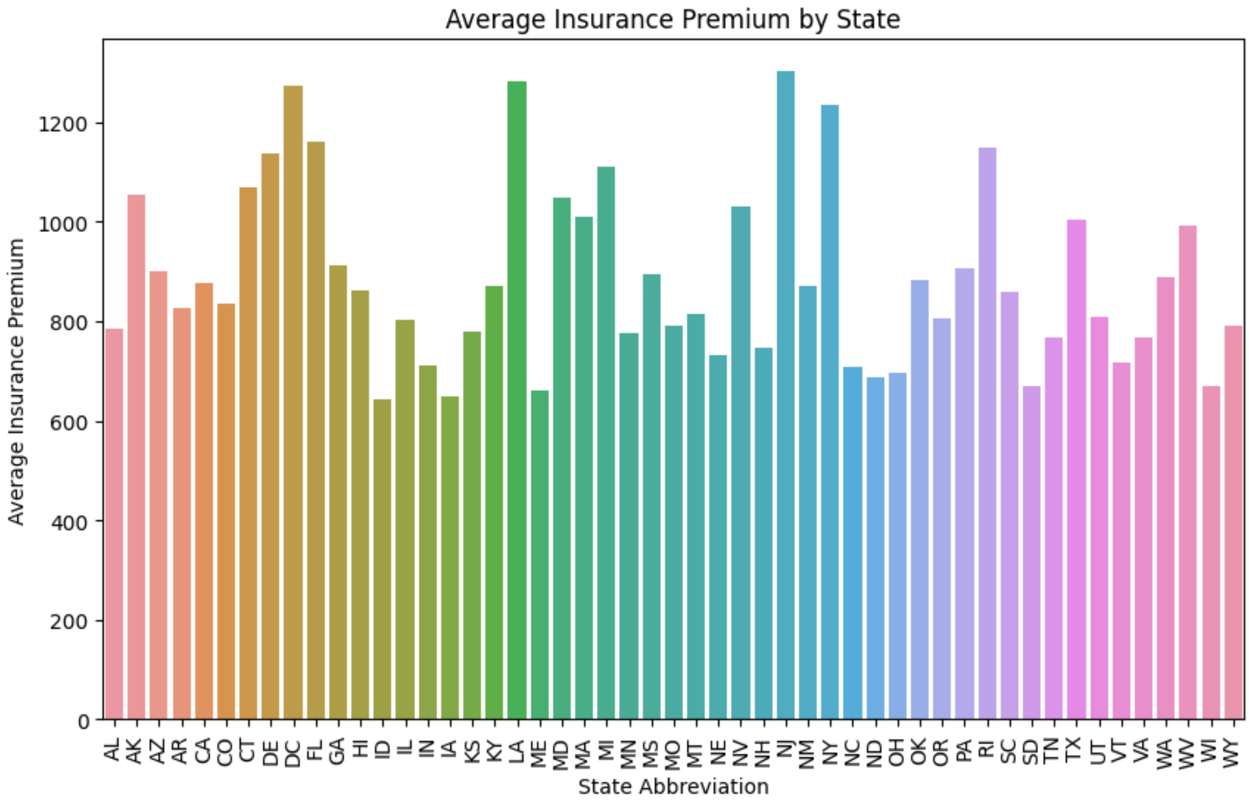
In [18]:

```
sns.pairplot(dataset[['total', 'speeding', 'alcohol', 'not_distracted', 'no
plt.show()
```



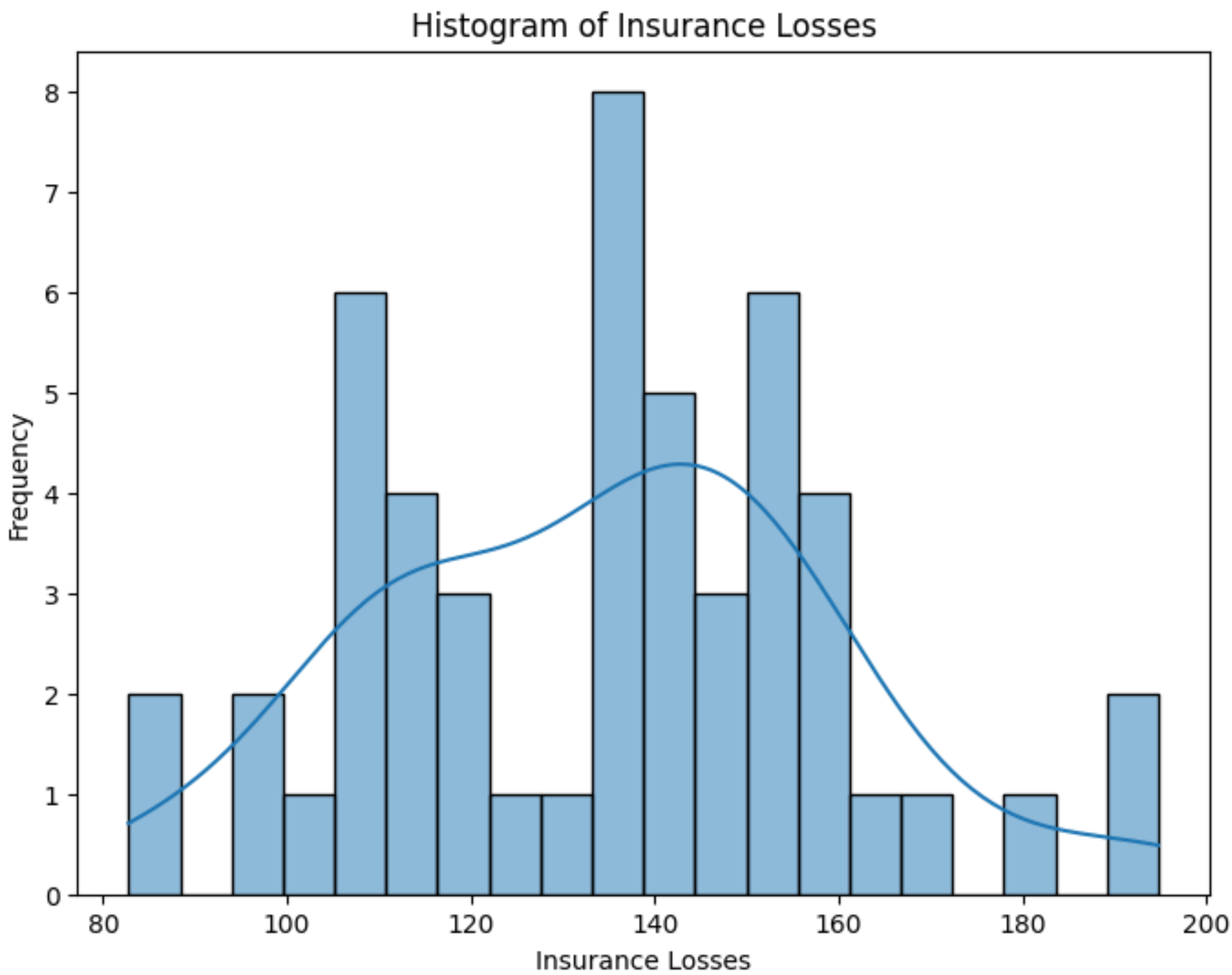
This bar plot displays the average insurance premium for each state. It can help identify variations in insurance premiums across different states.

In [19]:
```
plt.figure(figsize=(10, 6))
sns.barplot(x='abbrev', y='ins_premium', data=dataset)
plt.xlabel('State Abbreviation')
plt.ylabel('Average Insurance Premium')
plt.title('Average Insurance Premium by State')
plt.xticks(rotation=90)
plt.show()
```



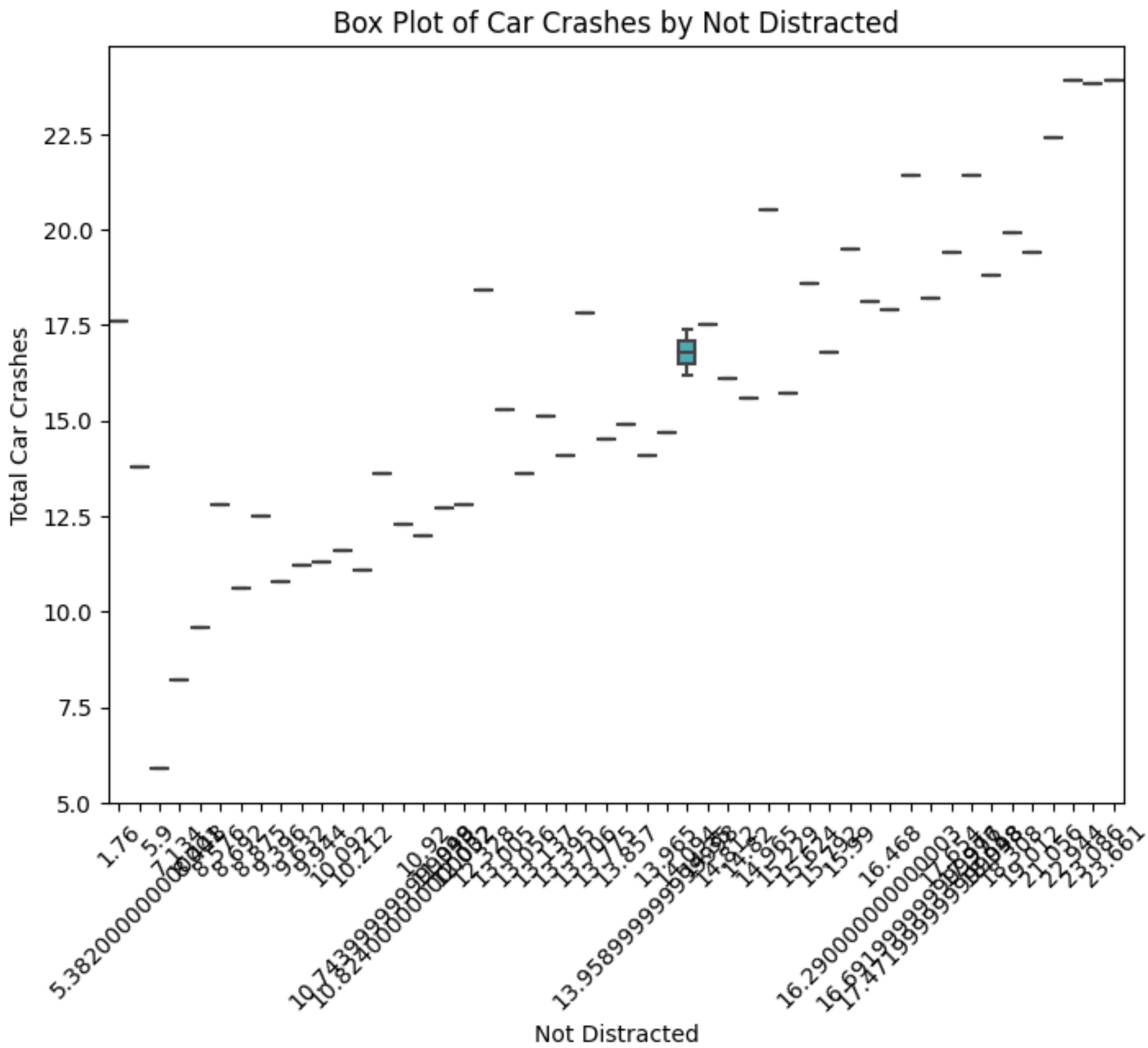
The histogram of insurance losses shows the distribution of insurance losses across the dataset. It may reveal any predominant ranges of insurance losses.

In [23]:
```
plt.figure(figsize=(8, 6))
sns.histplot(dataset['ins_losses'], bins=20, kde=True)
plt.xlabel('Insurance Losses')
plt.ylabel('Frequency')
plt.title('Histogram of Insurance Losses')
plt.show()
```



This box plot helps visualize the distribution of total car crashes based on the 'not_distracted' variable. It can highlight any differences in crash counts between distracted and not distracted states.

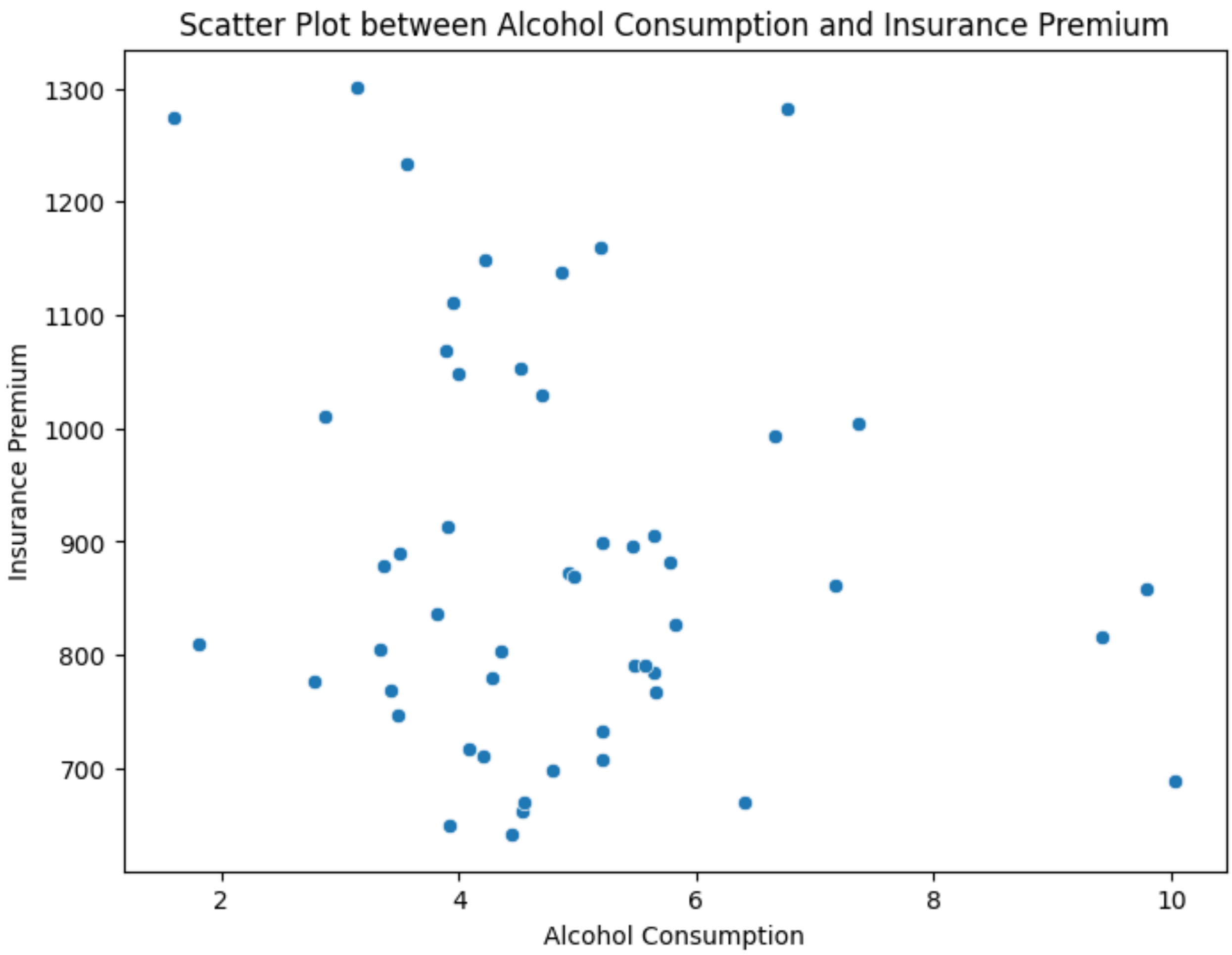
In [22]:
```
plt.figure(figsize=(8, 6))
sns.boxplot(x='not_distracted', y='total', data=dataset)
plt.xlabel('Not Distracted')
plt.ylabel('Total Car Crashes')
plt.title('Box Plot of Car Crashes by Not Distracted')
plt.xticks(rotation=45)
plt.show()
```



This scatter plot explores the relationship between alcohol consumption and insurance premiums. It may indicate whether there is any connection between these two variables.

In [24]:

```
plt.figure(figsize=(8, 6))
sns.scatterplot(x='alcohol', y='ins_premium', data=dataset)
plt.xlabel('Alcohol Consumption')
plt.ylabel('Insurance Premium')
plt.title('Scatter Plot between Alcohol Consumption and Insurance Premium'
plt.show()
```

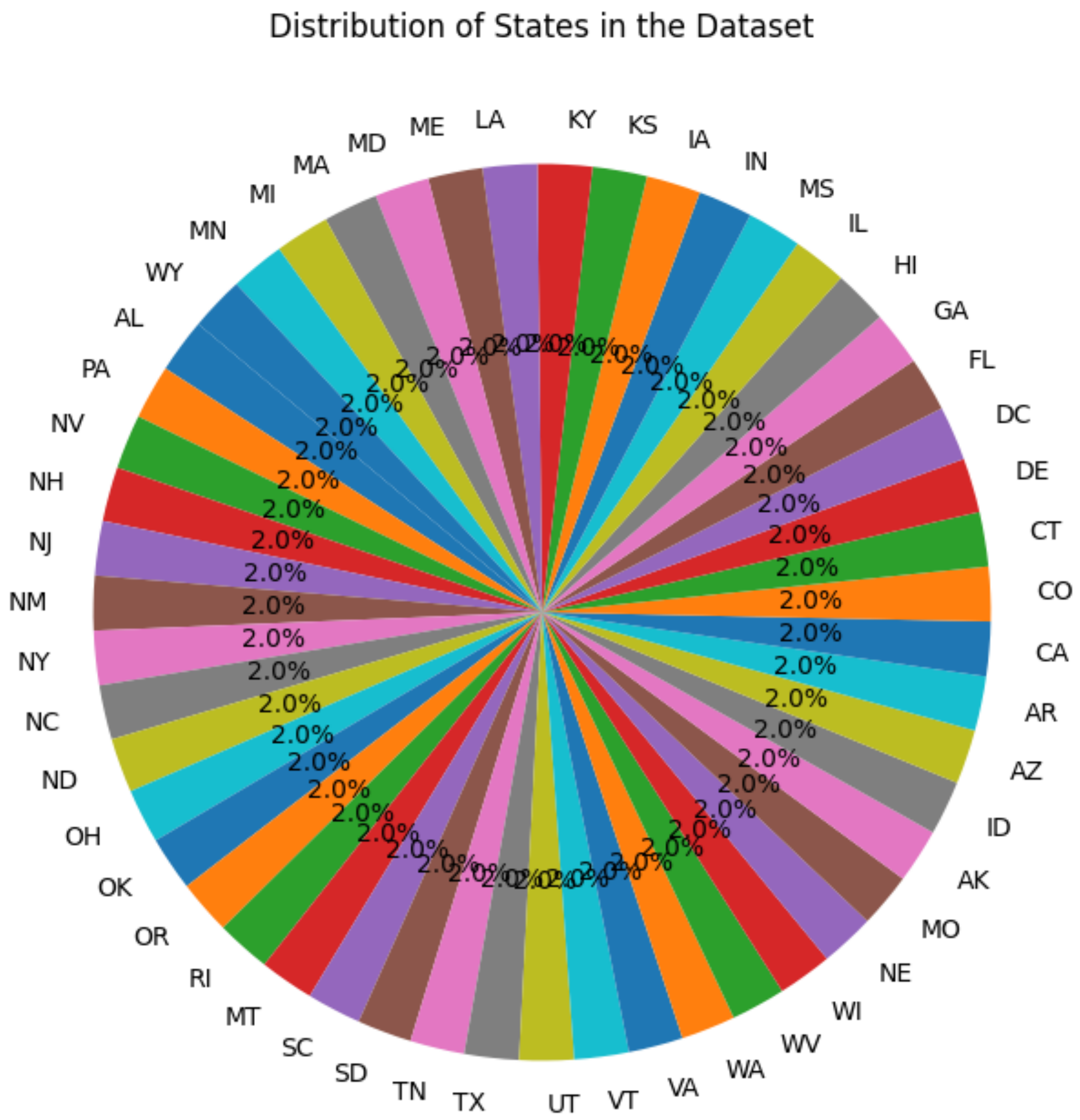


we've created a pie chart to visualize the distribution of states in your dataset based on their abbreviations. The chart displays the proportion of each state relative to the total number of states in the dataset.

In [25]:

```
state_counts = dataset['abbrev'].value_counts()

# Create a pie chart
plt.figure(figsize=(8, 8))
plt.pie(state_counts, labels=state_counts.index, autopct='%1.1f%%', startan
plt.title('Distribution of States in the Dataset')
plt.show()
```


Distribution of States in the Dataset
KY
KS
IA
IN
MS
IL
HI
GA
FL
DC
DE
CT
CO
CA
AR
AZ
ID
AK
MO
NE
WI
WV
WA
VA
VT
UT
TX
TN
SD
SC
MT
RI
OR
OK
OH
ND
NC
NY
NM
NJ
NH
NV
PA
AL
WY
MN
MI
MA
MD
ME
LA
2.0%