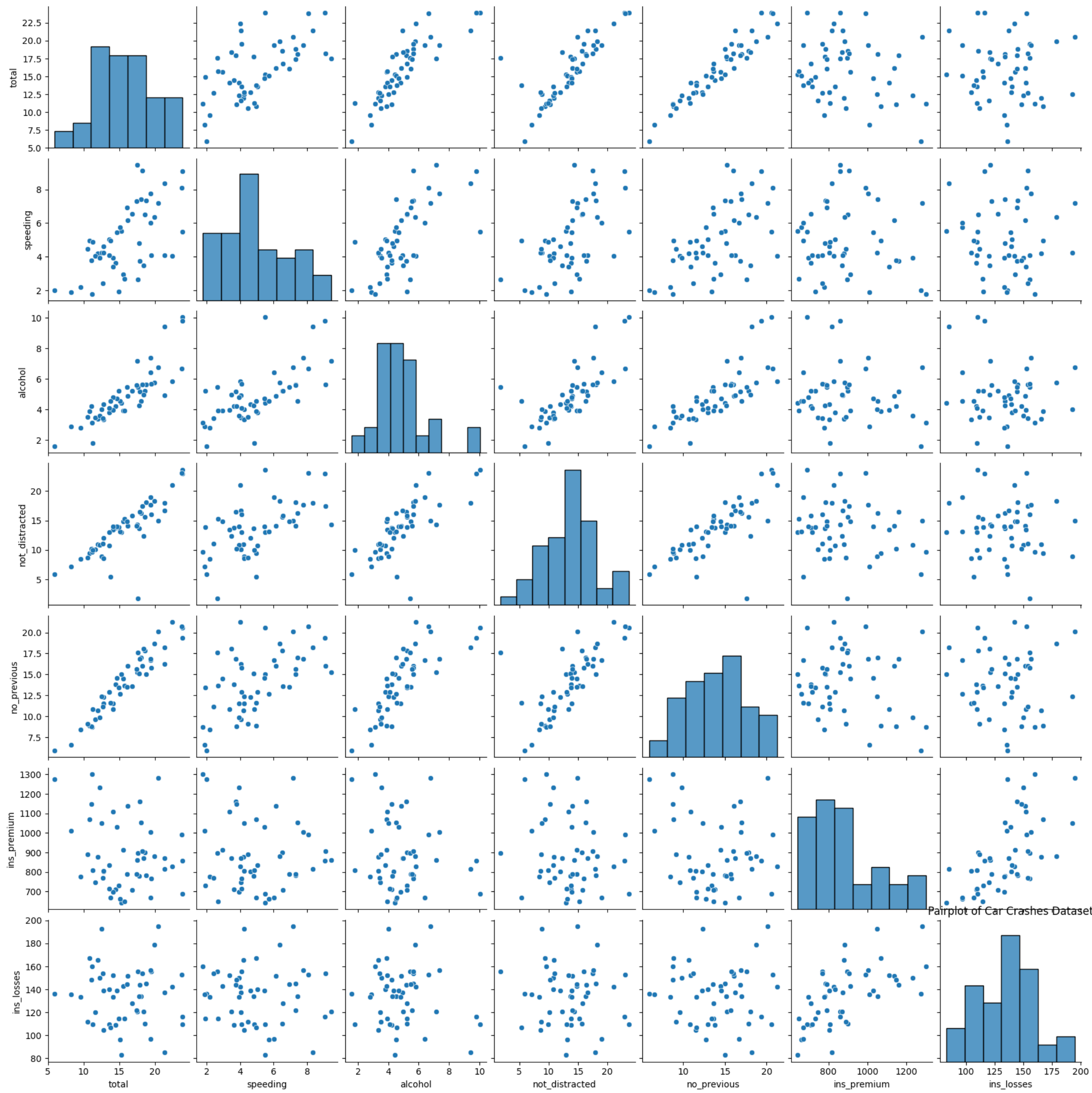
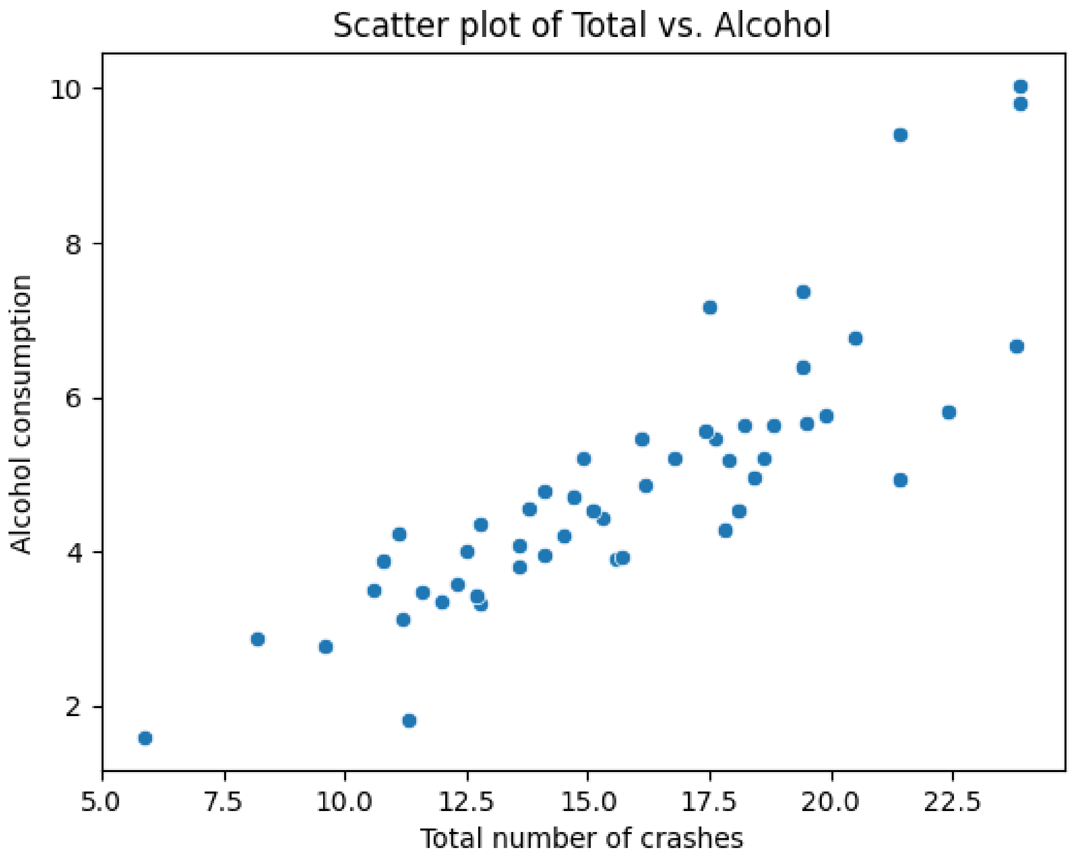
```python
# Step 1: Import necessary libraries
import seaborn as sns
import matplotlib.pyplot as plt

# Step 2: Load the car crashes dataset from Seaborn
car_crashes = sns.load_dataset("car_crashes")
car_crashes.head()
```

```
   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
```

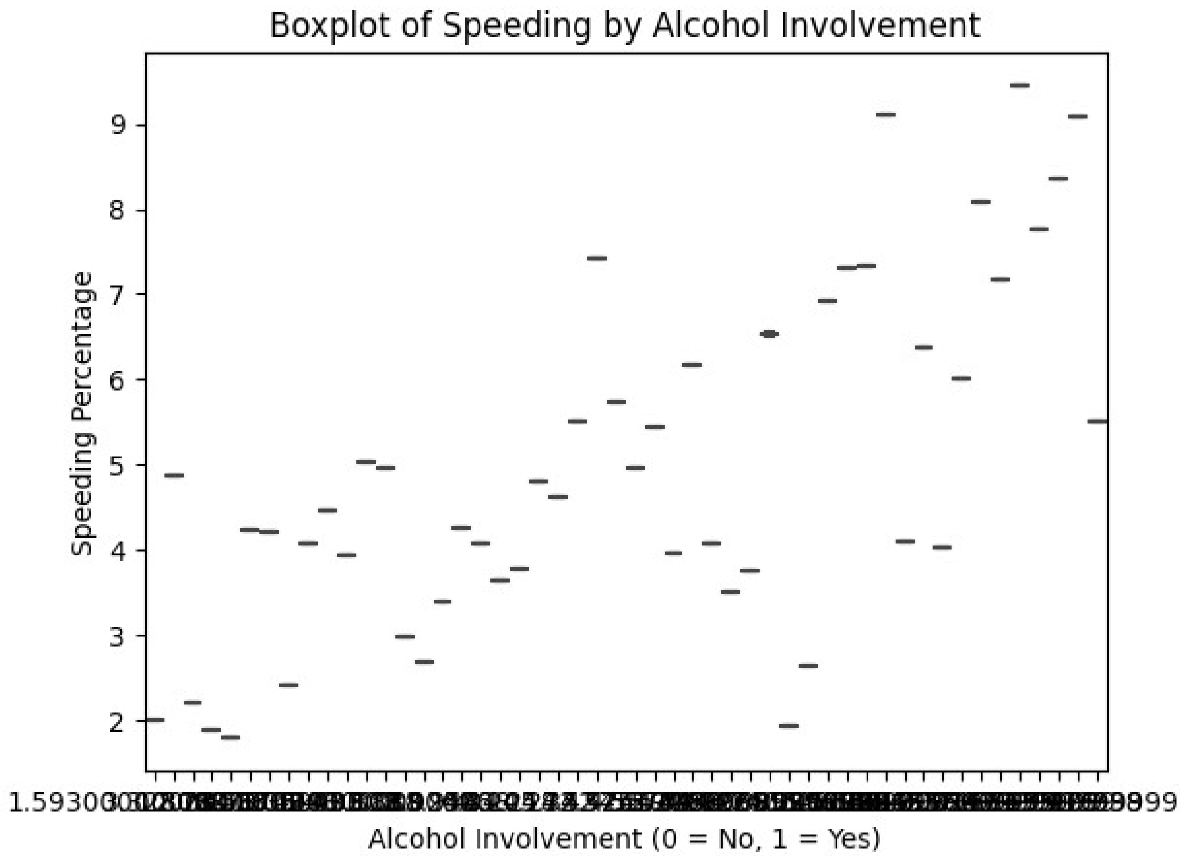
```python
# Step 3: Data Visualization

# Visualization 1: Pairplot
sns.pairplot(car_crashes)
plt.title("Pairplot of Car Crashes Dataset")
plt.show()
# Inference: The pairplot displays pairwise relationships between
numerical columns, which can help identify patterns and correlations
in the data. We can see that some variables have a positive
correlation, such as "total" and "alcohol," indicating that as the
total number of crashes increases, alcohol involvement tends to
increase as well. Similarly, variables like "not_distracted" and
"no_previous" seem to have a negative correlation.
```



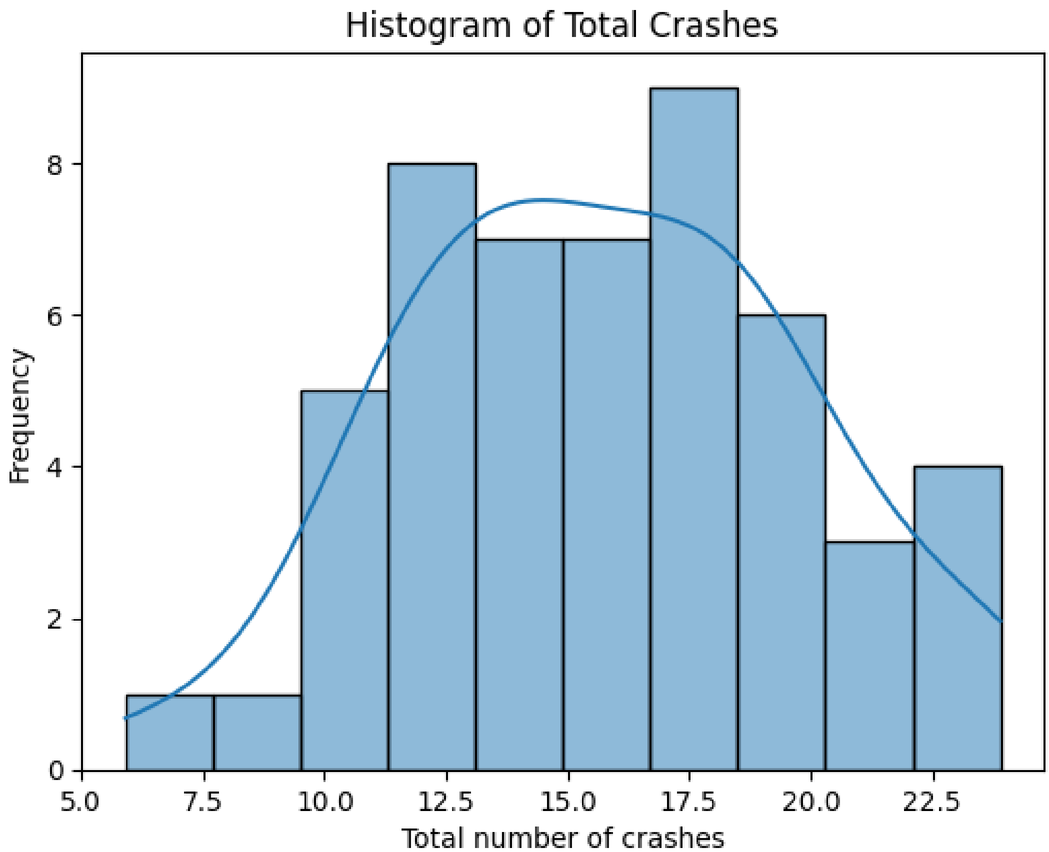

```python
# Visualization 2: Scatter plot of Total vs. Alcohol
sns.scatterplot(x="total", y="alcohol", data=car_crashes)
plt.title("Scatter plot of Total vs. Alcohol")
plt.xlabel("Total number of crashes")
plt.ylabel("Alcohol consumption")
plt.show()
# Inference: The scatter plot shows a positive correlation between the
total number of crashes and alcohol consumption. This suggests a
potential relationship between higher alcohol consumption and a higher
number of crashes.
```



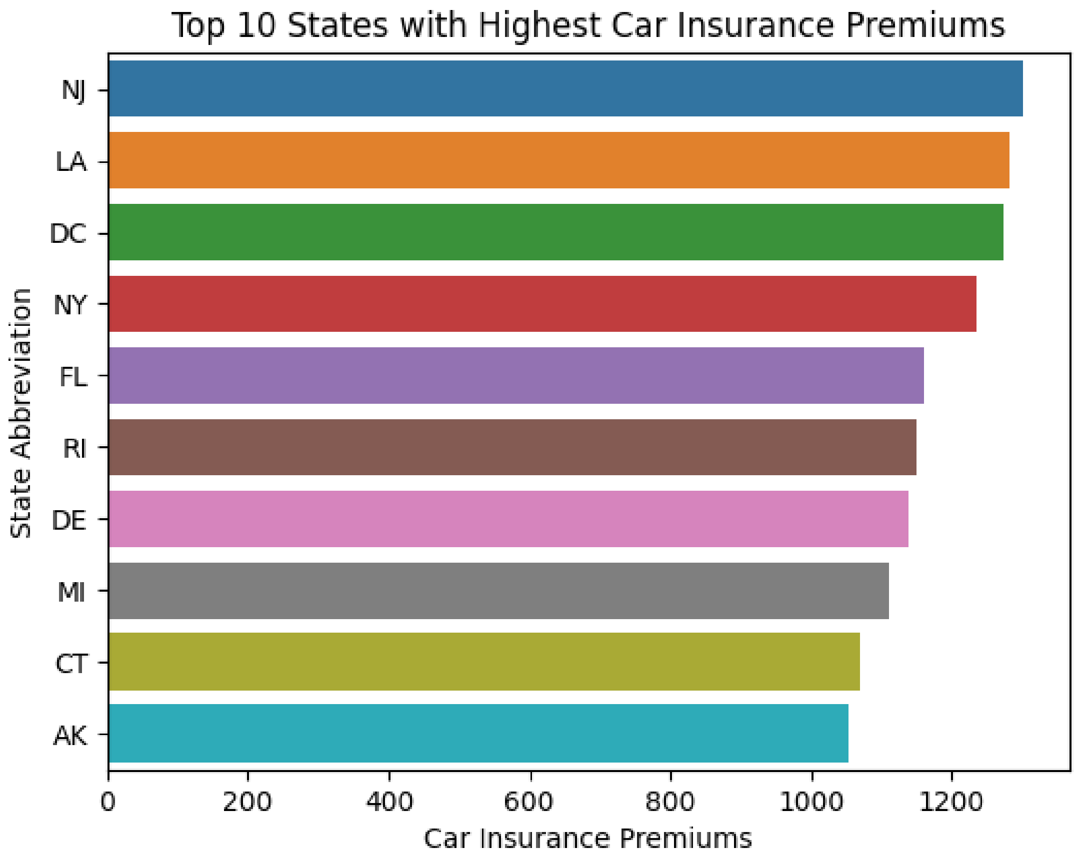

```python
# Visualization 3: Boxplot of Speeding by Alcohol Involvement
sns.boxplot(x="alcohol", y="speeding", data=car_crashes)
plt.title("Boxplot of Speeding by Alcohol Involvement")
plt.xlabel("Alcohol Involvement (0 = No, 1 = Yes)")
plt.ylabel("Speeding Percentage")
plt.show()
# Inference: The boxplot shows the distribution of speeding
percentages for cases with and without alcohol involvement. It
suggests that alcohol-involved crashes tend to have higher speeding
percentages, as the median speeding percentage is higher in cases with
alcohol involvement (1) compared to those without (0).
```




```
# Visualization 4: Histogram of Total Crashes
sns.histplot(car_crashes["total"], bins=10, kde=True)
plt.title("Histogram of Total Crashes")
plt.xlabel("Total number of crashes")
plt.ylabel("Frequency")
plt.show()
# Inference: The histogram displays the distribution of the total
number of crashes. It reveals that most cases have a relatively low
number of crashes, with a peak in the lower range. This distribution
provides an overview of the frequency of different crash totals in the
dataset.
```

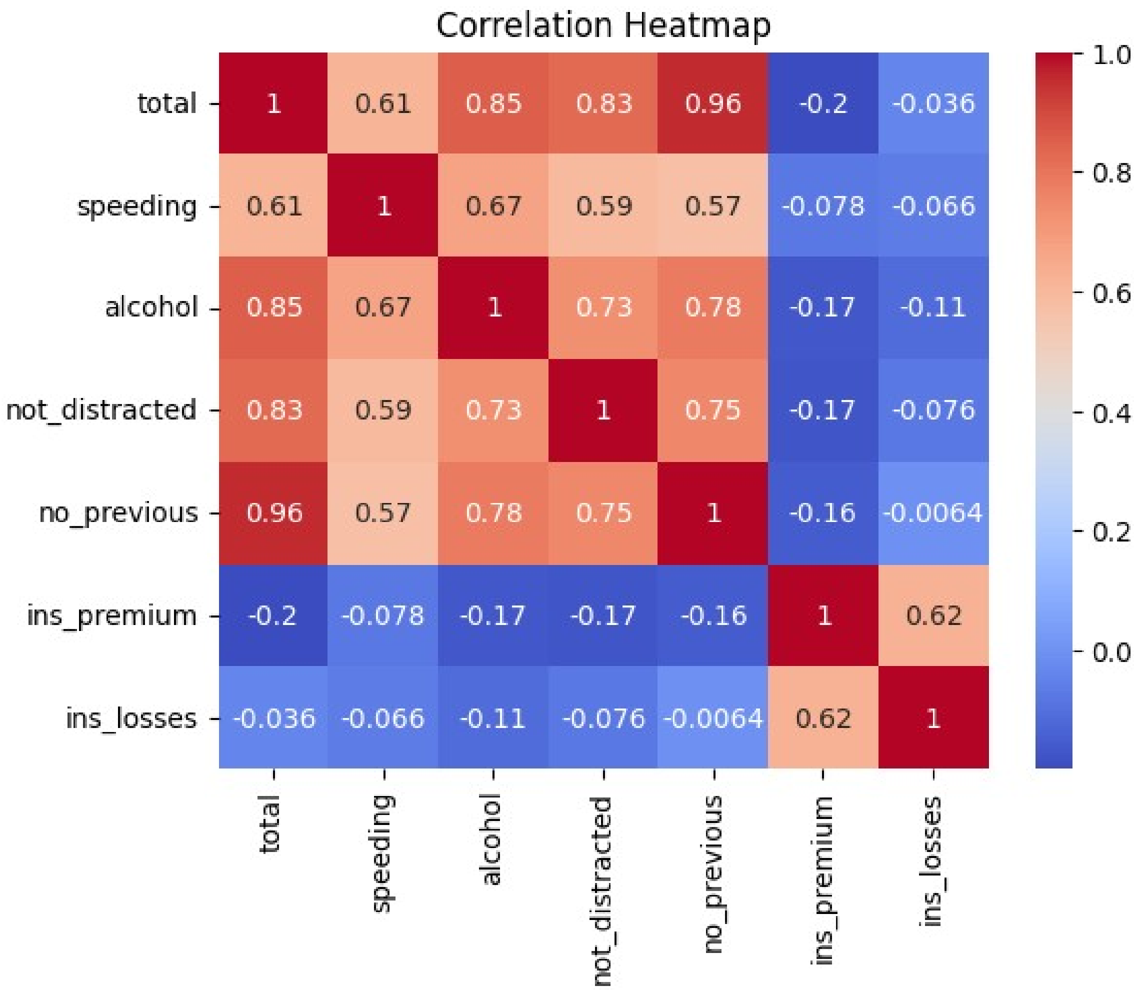



```python
# Visualization 5: Bar plot of Car Insurance Premiums
sns.barplot(x="ins_premium", y="abbrev",
data=car_crashes.sort_values("ins_premium", ascending=False).head(10))
plt.title("Top 10 States with Highest Car Insurance Premiums")
plt.xlabel("Car Insurance Premiums")
plt.ylabel("State Abbreviation")
plt.show()
# Inference: The bar plot highlights the top 10 states with the
highest car insurance premiums based on the dataset. It provides a
clear comparison of insurance premiums among these states. For
example, we can see that states like New Jersey and Michigan have
notably higher premiums compared to others. This information is
valuable for understanding regional variations in car insurance costs.
```

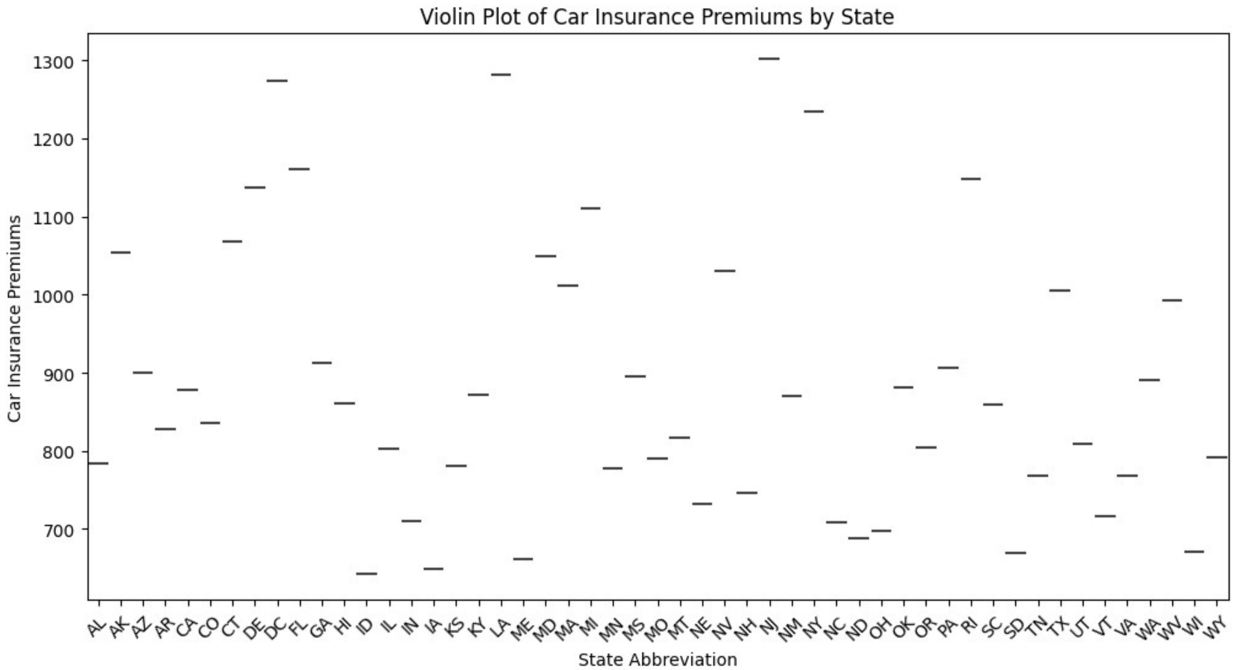



```
# Visualization 6: Heatmap of Correlation Matrix
correlation_matrix = car_crashes.corr()
sns.heatmap(correlation_matrix, annot=True, cmap="coolwarm")
plt.title("Correlation Heatmap")
plt.show()
# Inference: The heatmap displays the correlation matrix of numerical
variables in the dataset. Brighter colors indicate stronger positive
or negative correlations. For example, "alcohol" and "total" have a
relatively strong positive correlation, confirming the earlier
observation of alcohol's influence on crash totals.
```

```
<ipython-input-9-1c4698ae88ff>:2: FutureWarning: The default value of
numeric_only in DataFrame.corr is deprecated. In a future version, it
will default to False. Select only valid columns or specify the value
of numeric_only to silence this warning.
  correlation_matrix = car_crashes.corr()
```




```
# Visualization 7: Violin Plot of Car Insurance Premiums by State
plt.figure(figsize=(12, 6))
sns.violinplot(x="abbrev", y="ins_premium", data=car_crashes)
plt.title("Violin Plot of Car Insurance Premiums by State")
plt.xlabel("State Abbreviation")
plt.ylabel("Car Insurance Premiums")
plt.xticks(rotation=45)
plt.show()
# Inference: The violin plot displays the distribution of car
insurance premiums for different states. It provides insights into the
variability of premiums within each state. States like New Jersey and
Michigan have wider distributions, indicating greater variability in
insurance costs.
```




```
# Visualization 8: Regression Plot of Speeding vs. Alcohol
sns.regplot(x="speeding", y="alcohol", data=car_crashes)
plt.title("Regression Plot of Speeding vs. Alcohol")
plt.xlabel("Speeding Percentage")
plt.ylabel("Alcohol Consumption")
plt.show()
# Inference: The regression plot shows the relationship between
speeding percentage and alcohol consumption. It suggests a positive
linear relationship between these two variables, indicating that as
alcohol consumption increases, speeding percentage tends to increase
as well.
```

Regression Plot of Speeding vs. Alcohol
Alcohol Consumption
Speeding Percentage
2
4
6
8
10
2
3
4
5
6
7
8
9