

car

September 14, 2023

```
[ ]: import warnings
warnings.filterwarnings('ignore')
```

```
[ ]: %pip install matplotlib
      %pip install seaborn
      %pip install numpy
      import seaborn as sns
      import matplotlib.pyplot as plt
      import numpy as np
```

```
Requirement already satisfied: matplotlib in
/home/kyojuro/Desktop/Programs/REPOS/ML/SI-
GuidedProject-581974-1694665929/.venv/lib/python3.10/site-packages (3.7.3)
Requirement already satisfied: contourpy>=1.0.1 in
/home/kyojuro/Desktop/Programs/REPOS/ML/SI-
GuidedProject-581974-1694665929/.venv/lib/python3.10/site-packages (from
matplotlib) (1.1.0)
Requirement already satisfied: cyclor>=0.10 in
/home/kyojuro/Desktop/Programs/REPOS/ML/SI-
GuidedProject-581974-1694665929/.venv/lib/python3.10/site-packages (from
matplotlib) (0.11.0)
Requirement already satisfied: fonttools>=4.22.0 in
/home/kyojuro/Desktop/Programs/REPOS/ML/SI-
GuidedProject-581974-1694665929/.venv/lib/python3.10/site-packages (from
matplotlib) (4.42.1)
Requirement already satisfied: kiwisolver>=1.0.1 in
/home/kyojuro/Desktop/Programs/REPOS/ML/SI-
GuidedProject-581974-1694665929/.venv/lib/python3.10/site-packages (from
matplotlib) (1.4.5)
Requirement already satisfied: numpy<2,>=1.20 in
/home/kyojuro/Desktop/Programs/REPOS/ML/SI-
GuidedProject-581974-1694665929/.venv/lib/python3.10/site-packages (from
matplotlib) (1.25.2)
Requirement already satisfied: packaging>=20.0 in
/home/kyojuro/Desktop/Programs/REPOS/ML/SI-
GuidedProject-581974-1694665929/.venv/lib/python3.10/site-packages (from
matplotlib) (23.1)
Requirement already satisfied: pillow>=6.2.0 in
```

```

/home/kyojuro/Desktop/Programs/REPOS/ML/SI-
GuidedProject-581974-1694665929/.venv/lib/python3.10/site-packages (from
matplotlib) (10.0.0)
Requirement already satisfied: pyparsing>=2.3.1 in
/home/kyojuro/Desktop/Programs/REPOS/ML/SI-
GuidedProject-581974-1694665929/.venv/lib/python3.10/site-packages (from
matplotlib) (3.1.1)
Requirement already satisfied: python-dateutil>=2.7 in
/home/kyojuro/Desktop/Programs/REPOS/ML/SI-
GuidedProject-581974-1694665929/.venv/lib/python3.10/site-packages (from
matplotlib) (2.8.2)
Requirement already satisfied: six>=1.5 in
/home/kyojuro/Desktop/Programs/REPOS/ML/SI-
GuidedProject-581974-1694665929/.venv/lib/python3.10/site-packages (from python-
dateutil>=2.7->matplotlib) (1.16.0)
Note: you may need to restart the kernel to use updated packages.
Requirement already satisfied: seaborn in
/home/kyojuro/Desktop/Programs/REPOS/ML/SI-
GuidedProject-581974-1694665929/.venv/lib/python3.10/site-packages (0.12.2)
Requirement already satisfied: numpy!=1.24.0,>=1.17 in
/home/kyojuro/Desktop/Programs/REPOS/ML/SI-
GuidedProject-581974-1694665929/.venv/lib/python3.10/site-packages (from
seaborn) (1.25.2)
Requirement already satisfied: pandas>=0.25 in
/home/kyojuro/Desktop/Programs/REPOS/ML/SI-
GuidedProject-581974-1694665929/.venv/lib/python3.10/site-packages (from
seaborn) (2.1.0)
Requirement already satisfied: matplotlib!=3.6.1,>=3.1 in
/home/kyojuro/Desktop/Programs/REPOS/ML/SI-
GuidedProject-581974-1694665929/.venv/lib/python3.10/site-packages (from
seaborn) (3.7.3)
Requirement already satisfied: contourpy>=1.0.1 in
/home/kyojuro/Desktop/Programs/REPOS/ML/SI-
GuidedProject-581974-1694665929/.venv/lib/python3.10/site-packages (from
matplotlib!=3.6.1,>=3.1->seaborn) (1.1.0)
Requirement already satisfied: cycler>=0.10 in
/home/kyojuro/Desktop/Programs/REPOS/ML/SI-
GuidedProject-581974-1694665929/.venv/lib/python3.10/site-packages (from
matplotlib!=3.6.1,>=3.1->seaborn) (0.11.0)
Requirement already satisfied: fonttools>=4.22.0 in
/home/kyojuro/Desktop/Programs/REPOS/ML/SI-
GuidedProject-581974-1694665929/.venv/lib/python3.10/site-packages (from
matplotlib!=3.6.1,>=3.1->seaborn) (4.42.1)
Requirement already satisfied: kiwisolver>=1.0.1 in
/home/kyojuro/Desktop/Programs/REPOS/ML/SI-
GuidedProject-581974-1694665929/.venv/lib/python3.10/site-packages (from
matplotlib!=3.6.1,>=3.1->seaborn) (1.4.5)
Requirement already satisfied: packaging>=20.0 in

```

```

/home/kyojuro/Desktop/Programs/REPOS/ML/SI-
GuidedProject-581974-1694665929/.venv/lib/python3.10/site-packages (from
matplotlib!=3.6.1,>=3.1->seaborn) (23.1)
Requirement already satisfied: pillow>=6.2.0 in
/home/kyojuro/Desktop/Programs/REPOS/ML/SI-
GuidedProject-581974-1694665929/.venv/lib/python3.10/site-packages (from
matplotlib!=3.6.1,>=3.1->seaborn) (10.0.0)
Requirement already satisfied: pyparsing>=2.3.1 in
/home/kyojuro/Desktop/Programs/REPOS/ML/SI-
GuidedProject-581974-1694665929/.venv/lib/python3.10/site-packages (from
matplotlib!=3.6.1,>=3.1->seaborn) (3.1.1)
Requirement already satisfied: python-dateutil>=2.7 in
/home/kyojuro/Desktop/Programs/REPOS/ML/SI-
GuidedProject-581974-1694665929/.venv/lib/python3.10/site-packages (from
matplotlib!=3.6.1,>=3.1->seaborn) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in
/home/kyojuro/Desktop/Programs/REPOS/ML/SI-
GuidedProject-581974-1694665929/.venv/lib/python3.10/site-packages (from
pandas>=0.25->seaborn) (2023.3.post1)
Requirement already satisfied: tzdata>=2022.1 in
/home/kyojuro/Desktop/Programs/REPOS/ML/SI-
GuidedProject-581974-1694665929/.venv/lib/python3.10/site-packages (from
pandas>=0.25->seaborn) (2023.3)
Requirement already satisfied: six>=1.5 in
/home/kyojuro/Desktop/Programs/REPOS/ML/SI-
GuidedProject-581974-1694665929/.venv/lib/python3.10/site-packages (from python-
dateutil>=2.7->matplotlib!=3.6.1,>=3.1->seaborn) (1.16.0)
Note: you may need to restart the kernel to use updated packages.
Requirement already satisfied: numpy in
/home/kyojuro/Desktop/Programs/REPOS/ML/SI-
GuidedProject-581974-1694665929/.venv/lib/python3.10/site-packages (1.25.2)
Note: you may need to restart the kernel to use updated packages.

```

```

[ ]: df = sns.load_dataset('car_crashes')
df.head(1000)
# display(df.info())

```

```

[ ]:
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
5 13.6 5.032 3.808 10.744 12.920 835.50
6 10.8 4.968 3.888 9.396 8.856 1068.73
7 16.2 6.156 4.860 14.094 16.038 1137.87
8 5.9 2.006 1.593 5.900 5.900 1273.89

```

9	17.9	3.759	5.191	16.468	16.826	1160.13
10	15.6	2.964	3.900	14.820	14.508	913.15
11	17.5	9.450	7.175	14.350	15.225	861.18
12	15.3	5.508	4.437	13.005	14.994	641.96
13	12.8	4.608	4.352	12.032	12.288	803.11
14	14.5	3.625	4.205	13.775	13.775	710.46
15	15.7	2.669	3.925	15.229	13.659	649.06
16	17.8	4.806	4.272	13.706	15.130	780.45
17	21.4	4.066	4.922	16.692	16.264	872.51
18	20.5	7.175	6.765	14.965	20.090	1281.55
19	15.1	5.738	4.530	13.137	12.684	661.88
20	12.5	4.250	4.000	8.875	12.375	1048.78
21	8.2	1.886	2.870	7.134	6.560	1011.14
22	14.1	3.384	3.948	13.395	10.857	1110.61
23	9.6	2.208	2.784	8.448	8.448	777.18
24	17.6	2.640	5.456	1.760	17.600	896.07
25	16.1	6.923	5.474	14.812	13.524	790.32
26	21.4	8.346	9.416	17.976	18.190	816.21
27	14.9	1.937	5.215	13.857	13.410	732.28
28	14.7	5.439	4.704	13.965	14.553	1029.87
29	11.6	4.060	3.480	10.092	9.628	746.54
30	11.2	1.792	3.136	9.632	8.736	1301.52
31	18.4	3.496	4.968	12.328	18.032	869.85
32	12.3	3.936	3.567	10.824	9.840	1234.31
33	16.8	6.552	5.208	15.792	13.608	708.24
34	23.9	5.497	10.038	23.661	20.554	688.75
35	14.1	3.948	4.794	13.959	11.562	697.73
36	19.9	6.368	5.771	18.308	18.706	881.51
37	12.8	4.224	3.328	8.576	11.520	804.71
38	18.2	9.100	5.642	17.472	16.016	905.99
39	11.1	3.774	4.218	10.212	8.769	1148.99
40	23.9	9.082	9.799	22.944	19.359	858.97
41	19.4	6.014	6.402	19.012	16.684	669.31
42	19.5	4.095	5.655	15.990	15.795	767.91
43	19.4	7.760	7.372	17.654	16.878	1004.75
44	11.3	4.859	1.808	9.944	10.848	809.38
45	13.6	4.080	4.080	13.056	12.920	716.20
46	12.7	2.413	3.429	11.049	11.176	768.95
47	10.6	4.452	3.498	8.692	9.116	890.03
48	23.8	8.092	6.664	23.086	20.706	992.61
49	13.8	4.968	4.554	5.382	11.592	670.31
50	17.4	7.308	5.568	14.094	15.660	791.14

ins_losses abbrev

0	145.08	AL
1	133.93	AK
2	110.35	AZ

3	142.39	AR
4	165.63	CA
5	139.91	CO
6	167.02	CT
7	151.48	DE
8	136.05	DC
9	144.18	FL
10	142.80	GA
11	120.92	HI
12	82.75	ID
13	139.15	IL
14	108.92	IN
15	114.47	IA
16	133.80	KS
17	137.13	KY
18	194.78	LA
19	96.57	ME
20	192.70	MD
21	135.63	MA
22	152.26	MI
23	133.35	MN
24	155.77	MS
25	144.45	MO
26	85.15	MT
27	114.82	NE
28	138.71	NV
29	120.21	NH
30	159.85	NJ
31	120.75	NM
32	150.01	NY
33	127.82	NC
34	109.72	ND
35	133.52	OH
36	178.86	OK
37	104.61	OR
38	153.86	PA
39	148.58	RI
40	116.29	SC
41	96.87	SD
42	155.57	TN
43	156.83	TX
44	109.48	UT
45	109.61	VT
46	153.72	VA
47	111.62	WA
48	152.56	WV
49	106.62	WI

```
[ ]: # Set the style to whitegrid
sns.set(style="whitegrid")

# Load the car_crashes dataset and sort it by the 'total' column in descending
↳ order
crash_data = sns.load_dataset("car_crashes").sort_values("total",
↳ ascending=False)

# Create a subplot with a specified figure size
fig, ax = plt.subplots(figsize=(6, 15))

# Set color codes for the first barplot
sns.set_color_codes("pastel")
sns.barplot(x="total", y="abbrev", data=crash_data,
            label="Total", color="b")

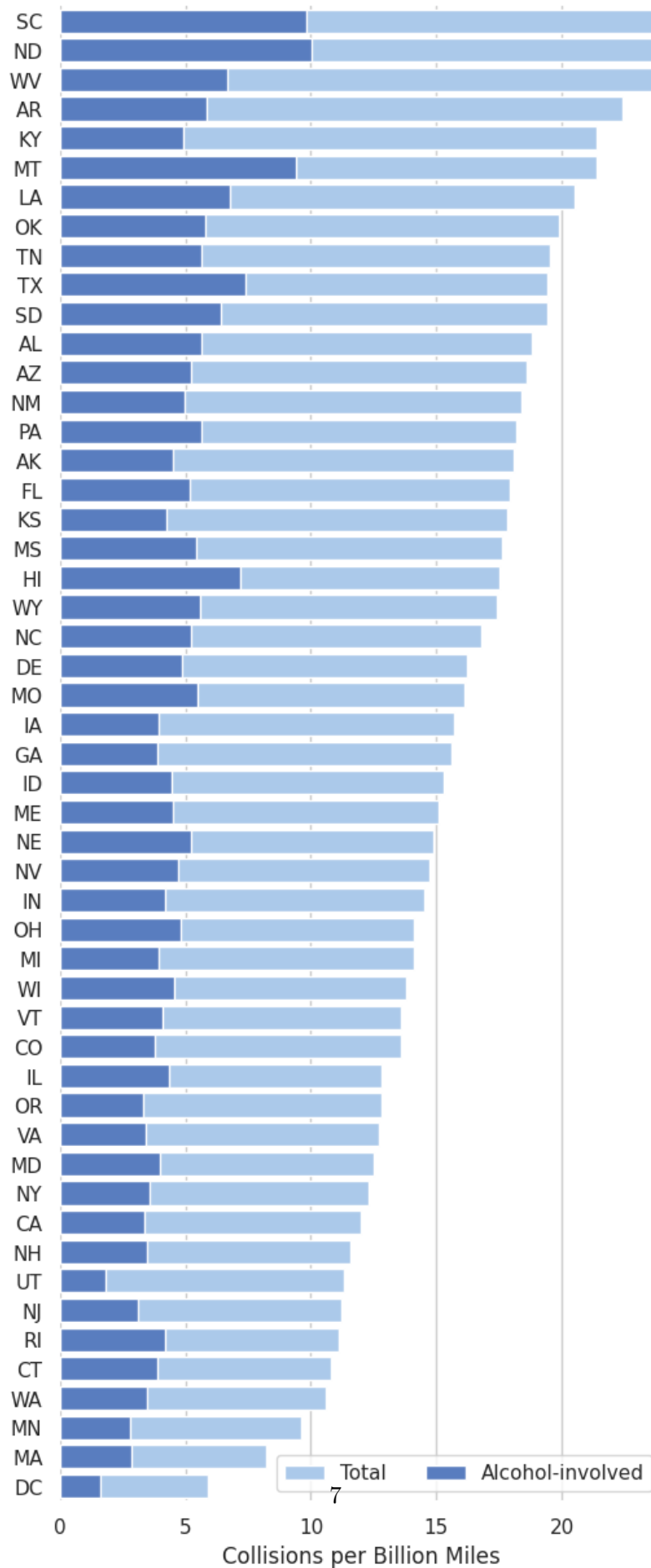
# Set color codes for the second barplot
sns.set_color_codes("muted")
sns.barplot(x="alcohol", y="abbrev", data=crash_data,
            label="Alcohol-involved", color="b")

# Add a legend to the plot with 2 columns and place it in the lower right corner
ax.legend(ncol=2, loc="lower right", frameon=True)

# Set the x-axis limit, labels, and title
ax.set(xlim=(0, 24), ylabel="",
       xlabel="Collisions per Billion Miles")

# Remove the left and bottom spines for better visualization
sns.despine(left=True, bottom=True)

# Show the plot
plt.show()
```



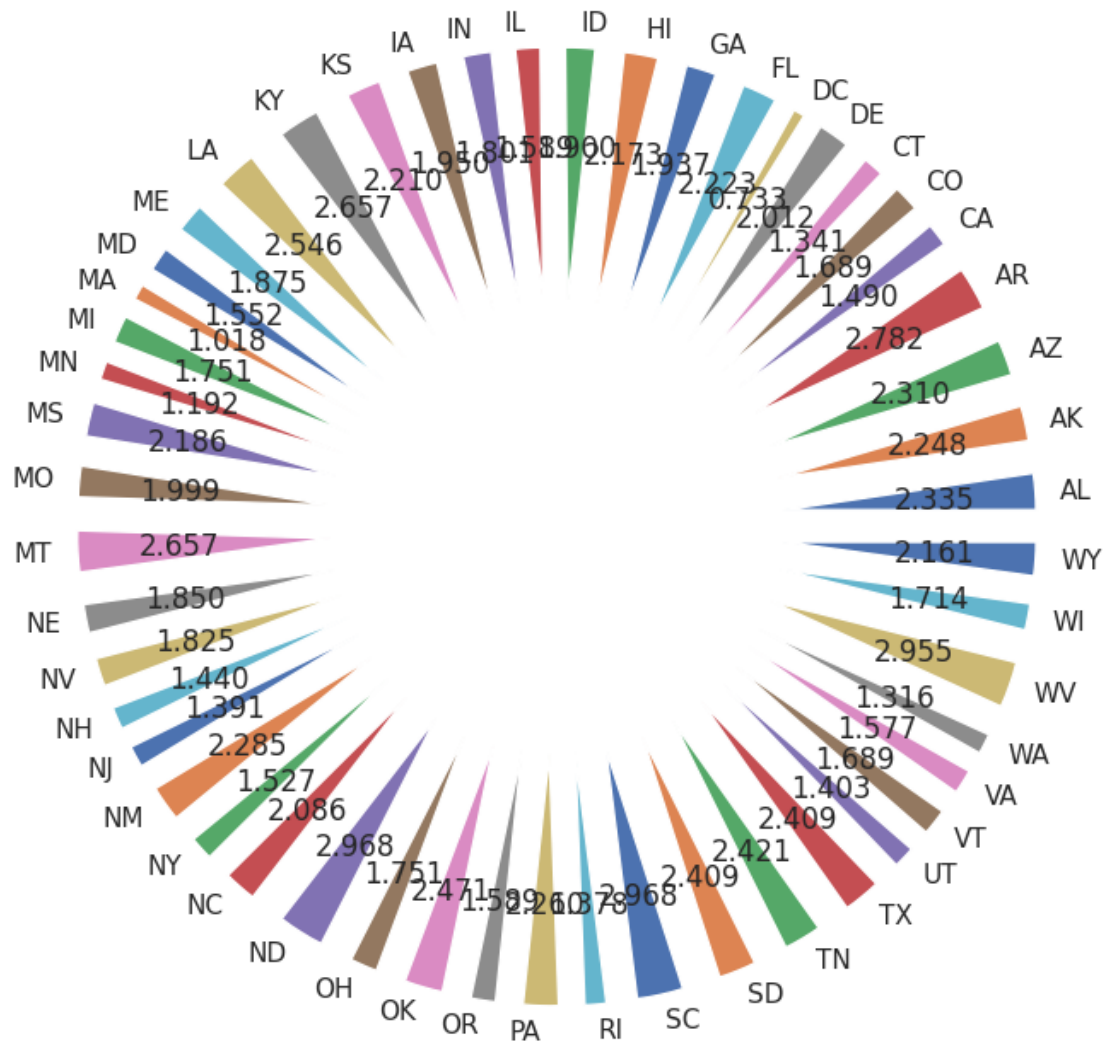
So we can see that, SC (South Carolina) has the most crashes per billion miles

```
[ ]: crash_data.head()
```

```
[ ]:      total  speeding  alcohol  not_distracted  no_previous  ins_premium  \
40    23.9      9.082    9.799          22.944      19.359      858.97
34    23.9      5.497   10.038          23.661      20.554      688.75
48    23.8      8.092    6.664          23.086      20.706      992.61
3     22.4      4.032    5.824          21.056      21.280      827.34
17    21.4      4.066    4.922          16.692      16.264      872.51

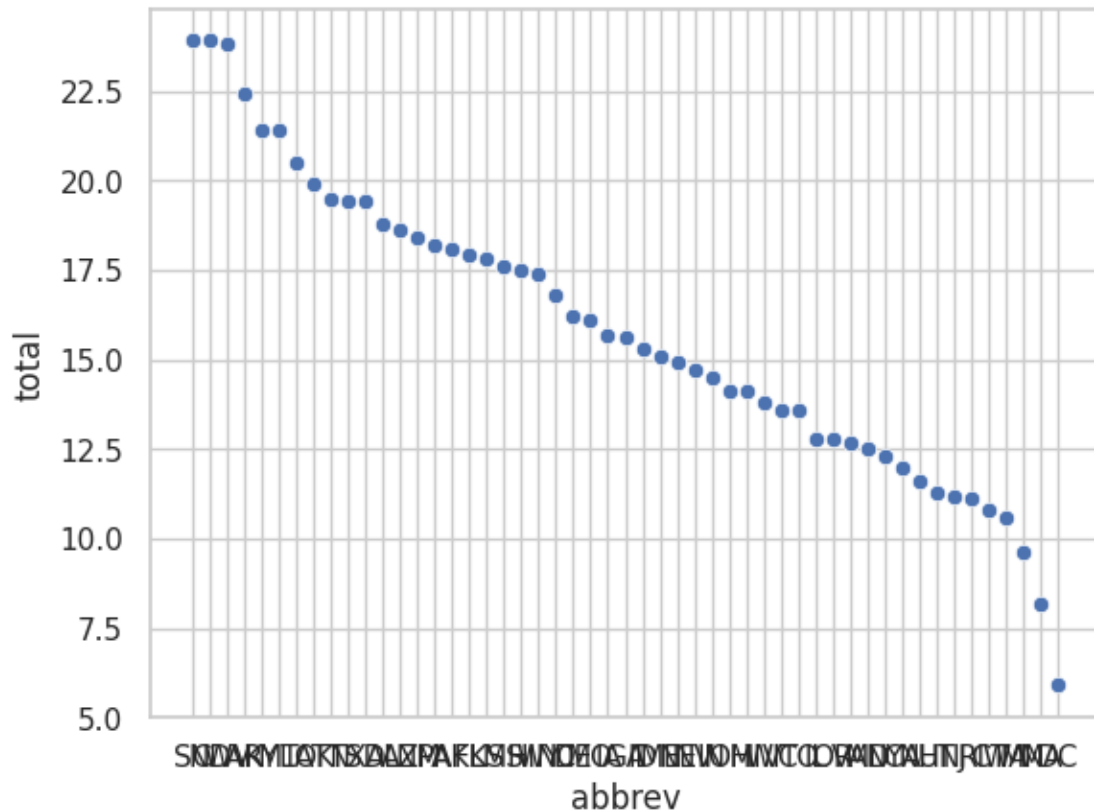
      ins_losses abbrev
40      116.29     SC
34      109.72     ND
48      152.56     WV
3       142.39     AR
17      137.13     KY
```

```
[ ]: myexplode = [0.9]*51
plt.pie(df.loc[:, "total"], labels = df.loc[:, "abbrev"], explode = myexplode,
        autopct='%0.3f')
plt.show()
```

```
[ ]: # display(crashes)
sns.scatterplot(data = crash_data, x = crash_data.loc[:, "abbrev"], y =
↳crash_data.loc[:, "total"], )
```

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

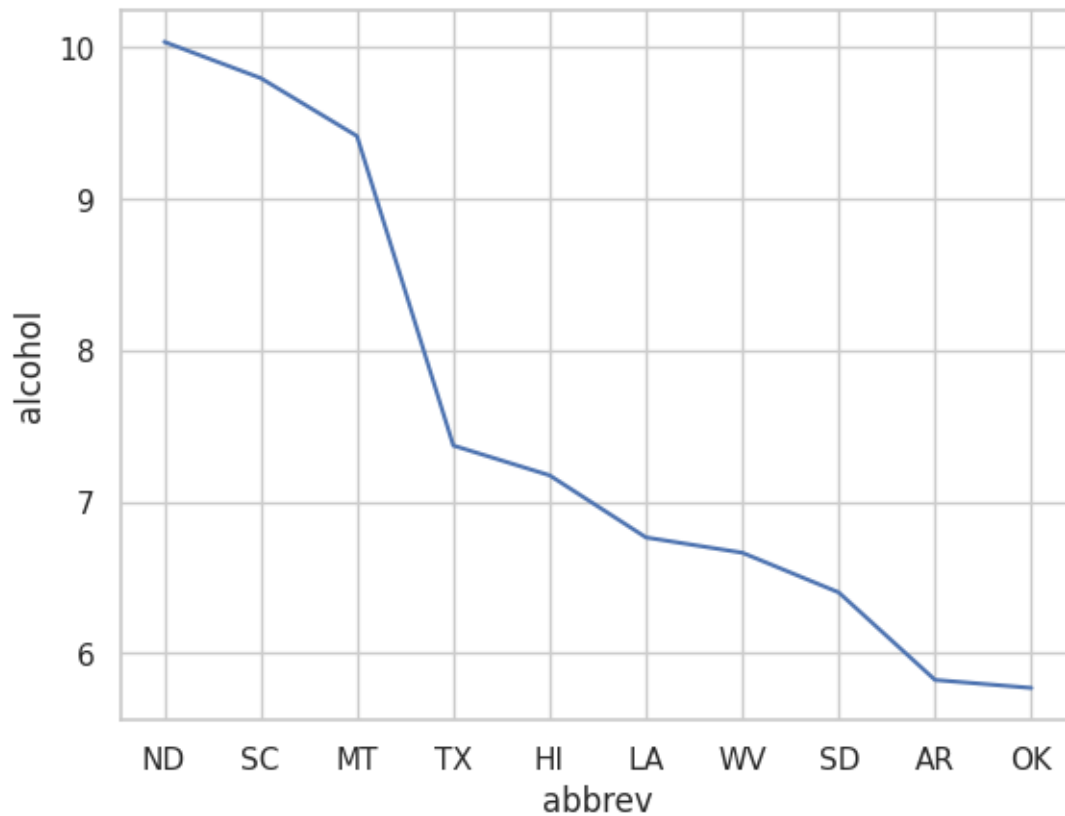


INFERENCE: Therefore we can see that SC (South Carolina) and ND (north Dakota) have the highest rate of car crashes

```
[ ]: alcohol_sorted = sns.load_dataset("car_crashes").sort_values("alcohol",
    ↪ascending=False)
alcohol_sorted = alcohol_sorted.iloc[0:10,:]
```

```
# alcohol_sorted.head()
# sns.histplot(data = alcohol_sorted, x = df.loc[:, "abbrev"], y = df.loc[:
    ↪, "alcohol"])
# plt.show()
sns.lineplot(data = alcohol_sorted, x = alcohol_sorted.loc[:, "abbrev"], y =
    ↪alcohol_sorted.loc[:, "alcohol"])
```

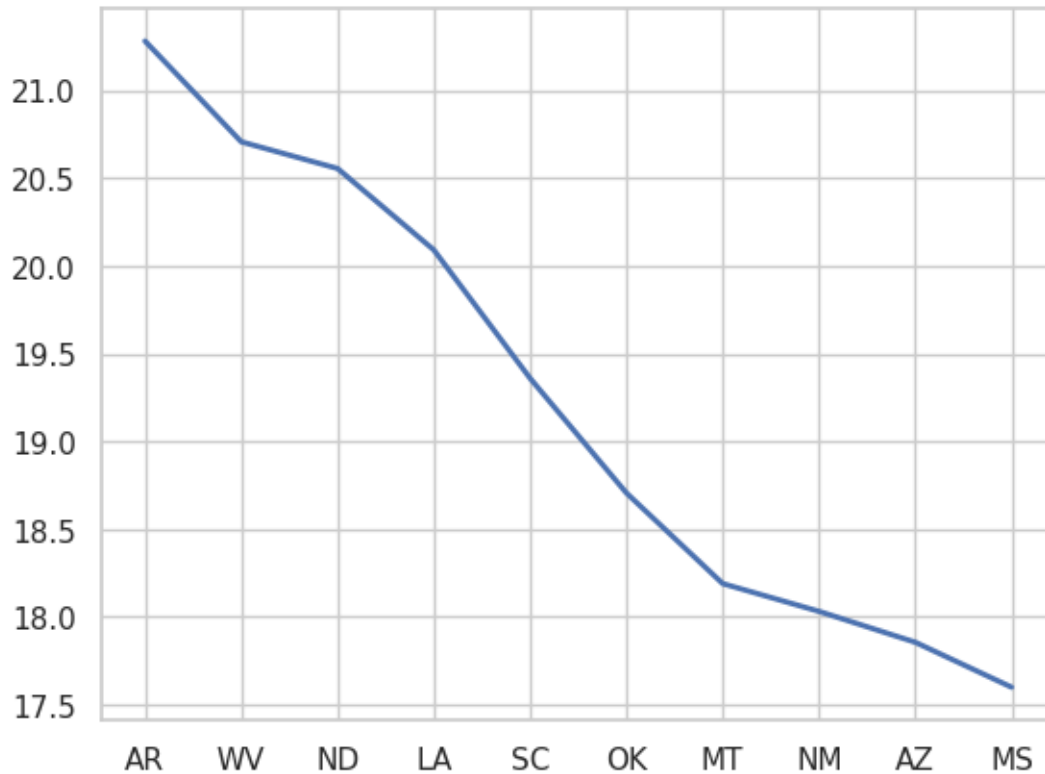
```
[ ]: <Axes: xlabel='abbrev', ylabel='alcohol'>
```



Therefore from the Lineplot we can see that ND (north Dakota) have the highest rates of car crashes which involves alcohol

```
[ ]: no_previous = sns.load_dataset("car_crashes").sort_values("no_previous",
    ↪ascending=False)
no_previous = no_previous.iloc[0:10,:]
plt.plot( no_previous["abbrev"],no_previous["no_previous"], linewidth = 2)
# display(no_previous)
```

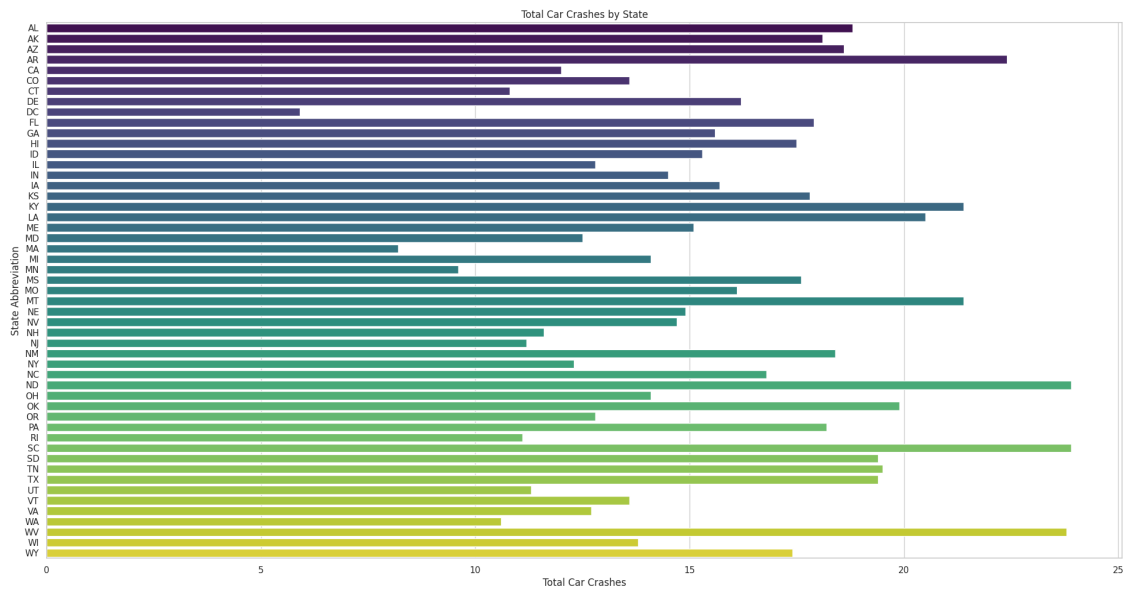
```
[ ]: [ <matplotlib.lines.Line2D at 0x7f1c14f59720>]
```



From this plot, we can see that AR (Arkansas) is the state that is most likely to have new people who meet with accidents.

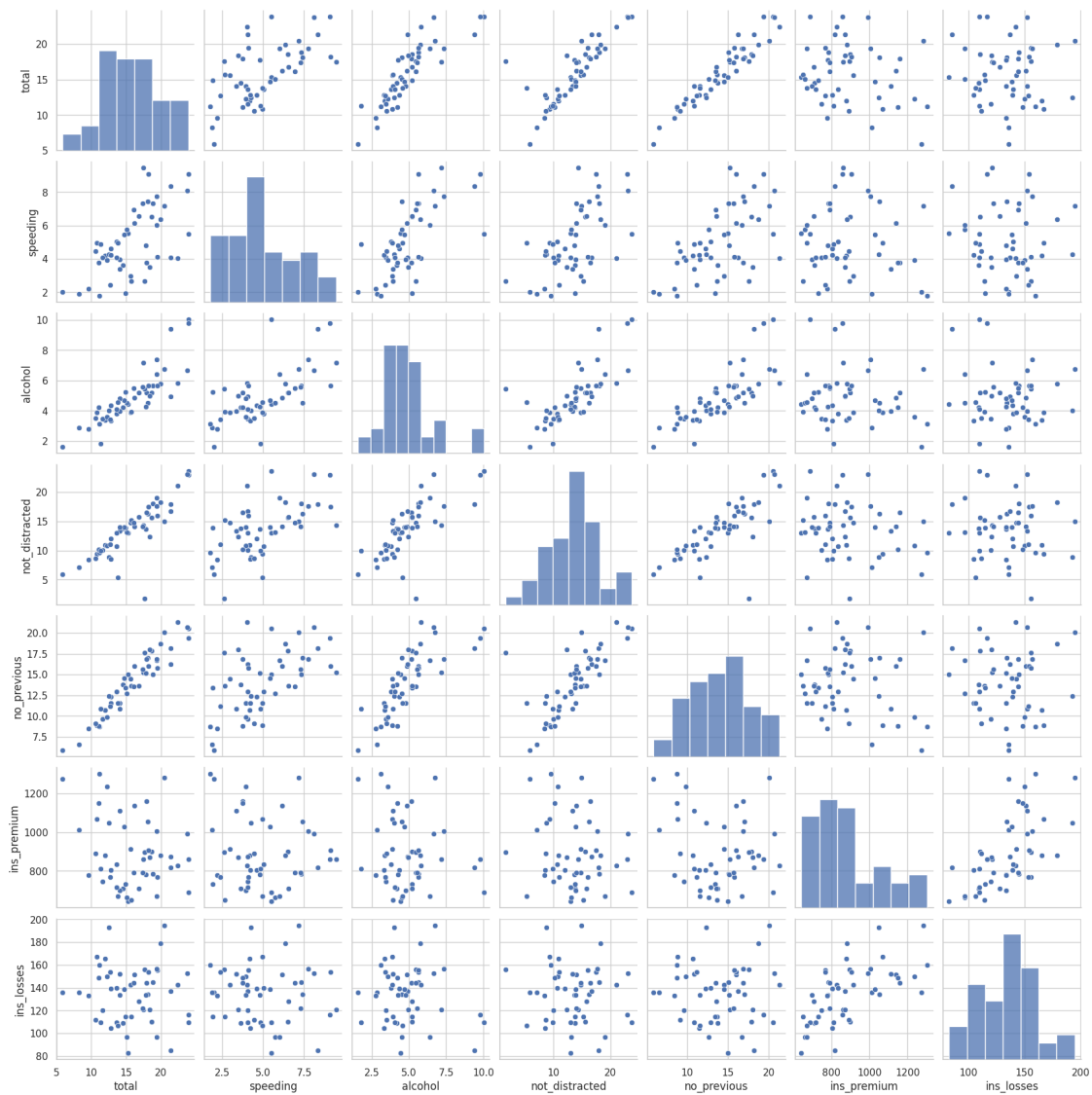
```
[ ]: # Load the car_crashes dataset
crashes = sns.load_dataset("car_crashes")

# Create a bar plot of total car crashes by state
plt.figure(figsize=(24, 12))
sns.barplot(x="total", y="abbrev", data=crashes, palette="viridis")
plt.xlabel("Total Car Crashes")
plt.ylabel("State Abbreviation")
plt.title("Total Car Crashes by State")
plt.show()
```



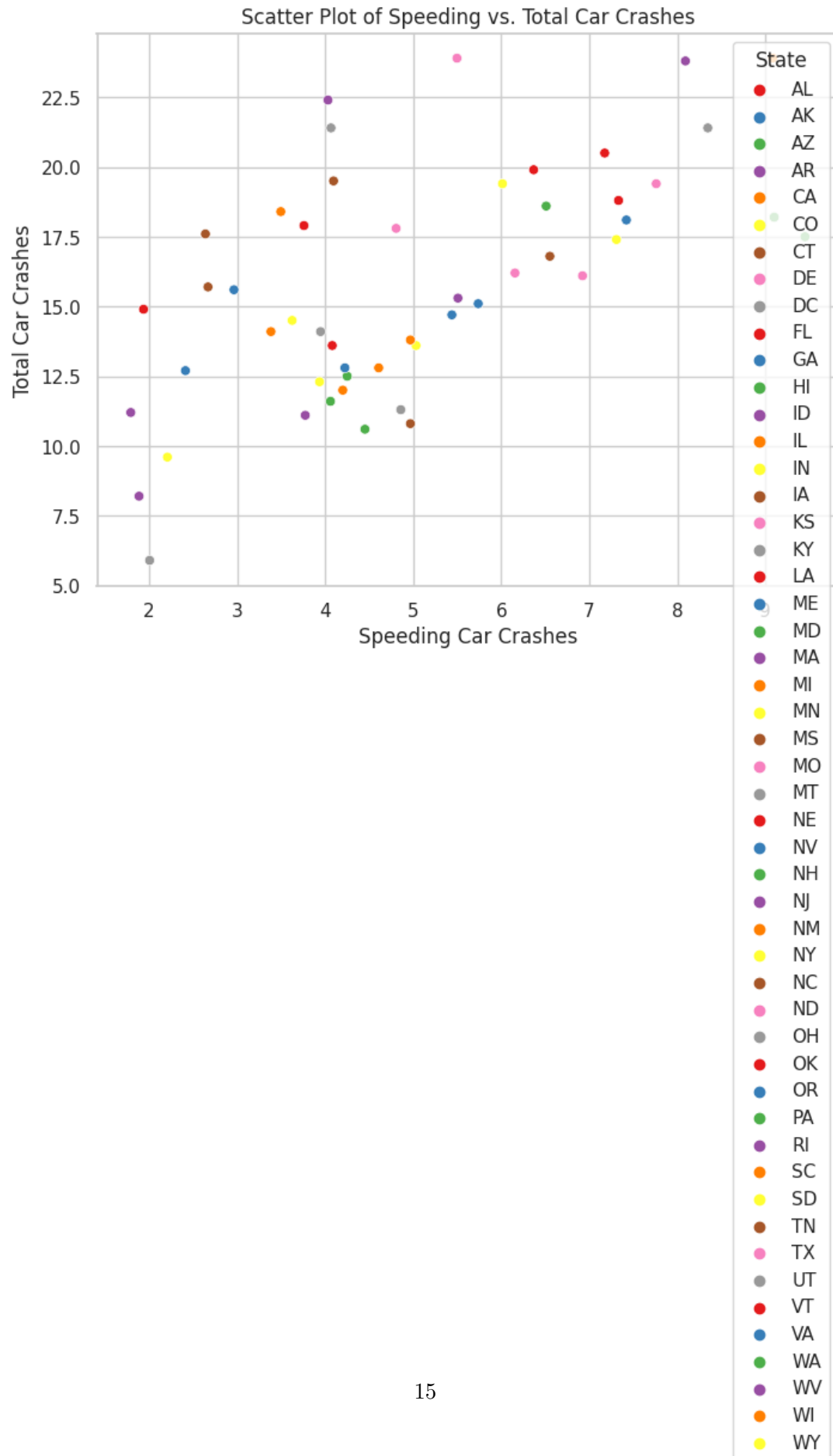
bar plot representation of car crashes of all 50 states and DC has the lowest total crashes among all the other states.

```
[ ]: crashes = sns.load_dataset("car_crashes")
sns.pairplot(crashes)
plt.show()
```



```
[ ]: crashes = sns.load_dataset("car_crashes")

# Create a scatter plot of speeding vs. total car crashes
plt.figure(figsize=(8, 6))
sns.scatterplot(data=crashes, x="speeding", y="total", hue="abbrev",
                palette="Set1")
plt.xlabel("Speeding Car Crashes")
plt.ylabel("Total Car Crashes")
plt.title("Scatter Plot of Speeding vs. Total Car Crashes")
plt.legend(title="State")
plt.show()
```



Therefore, we can see that AR (Arkansas) has the most amount of accidents related to speeding of cars.