```
MAHESH SRINIVAS UPPADA
21BCE9465
```

In [2]:

```
import seaborn as ms
```

In [4]:

```
ms.get_dataset_names()
```

Out[4]:

```
['anagrams',
 'anscombe',
 'attention',
 'brain_networks',
 'car_crashes',
 'diamonds',
 'dots',
 'dowjones',
 'exercise',
 'flights',
 'fmri',
 'geyser',
 'glue',
 'healthexp',
 'iris',
 'mpg',
 'penguins',
 'planets',
```

In [6]:

```
d=ms.load_dataset("car_crashes")
```

In [8]:

```
d.head()
```

Out[8]:

	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

In [10]:

```
d['not_distracted'].value_counts()
```

Out[10]:

```
14.094    2
18.048    1
17.472    1
13.965    1
10.092    1
9.632     1
12.328    1
10.824    1
15.792    1
23.661    1
13.959    1
18.308    1
8.576     1
10.212    1
17.976    1
22.944    1
19.012    1
15.990    1
17.654    1
9.944     1
13.056    1
11.049    1
8.692     1
23.086    1
13.857    1
14.812    1
16.290    1
1.760     1
15.624    1
21.056    1
10.920    1
10.744    1
9.396     1
5.900     1
16.468    1
14.820    1
14.350    1
13.005    1
12.032    1
13.775    1
15.229    1
13.706    1
16.692    1
14.965    1
13.137    1
8.875     1
7.134     1
13.395    1
8.448     1
5.382     1
Name: not_distracted, dtype: int64
```

DATA VISUALISATION

In [12]:

```
import numpy as np
print(np.array(d).shape)
print(len(d.columns))
print(len(d.index))
```

```
(51, 8)
8
51
```

In [14]:

```
# checking missing values
np.sum(d.isnull())
```

Out[14]:

```
total             0
speeding          0
alcohol           0
not_distracted    0
no_previous       0
ins_premium       0
ins_losses        0
abbrev            0
dtype: int64
```

In [16]:

```
d.corr()
```

```
C:\Users\thanu\AppData\Local\Temp\ipykernel_9692\2319908662.py:1: FutureWa
rning: The default value of numeric_only in DataFrame.corr is deprecated.
In a future version, it will default to False. Select only valid columns o
r specify the value of numeric_only to silence this warning.
  d.corr()
```

Out[16]:

	total	speeding	alcohol	not_distracted	no_previous	ins_premium	ins
total	1.000000	0.611548	0.852613	0.827560	0.956179	-0.199702	-(
speeding	0.611548	1.000000	0.669719	0.588010	0.571976	-0.077675	-(
alcohol	0.852613	0.669719	1.000000	0.732816	0.783520	-0.170612	-(
not_distracted	0.827560	0.588010	0.732816	1.000000	0.747307	-0.174856	-(
no_previous	0.956179	0.571976	0.783520	0.747307	1.000000	-0.156895	-(
ins_premium	-0.199702	-0.077675	-0.170612	-0.174856	-0.156895	1.000000	
ins_losses	-0.036011	-0.065928	-0.112547	-0.075970	-0.006359	0.623116	

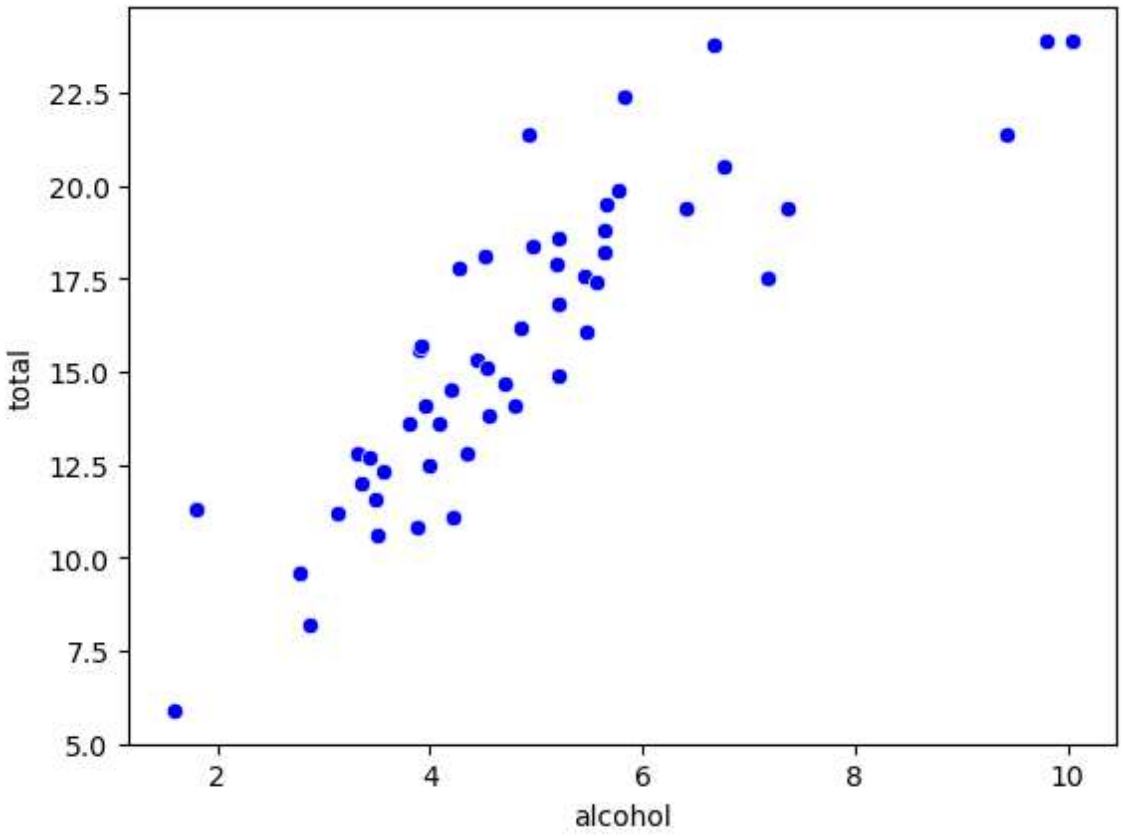
SCATTER PLOT

In [23]:

```
ms.scatterplot(x="alcohol",y="total",data=d,color='blue')
```

Out[23]:

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



LINE PLOT

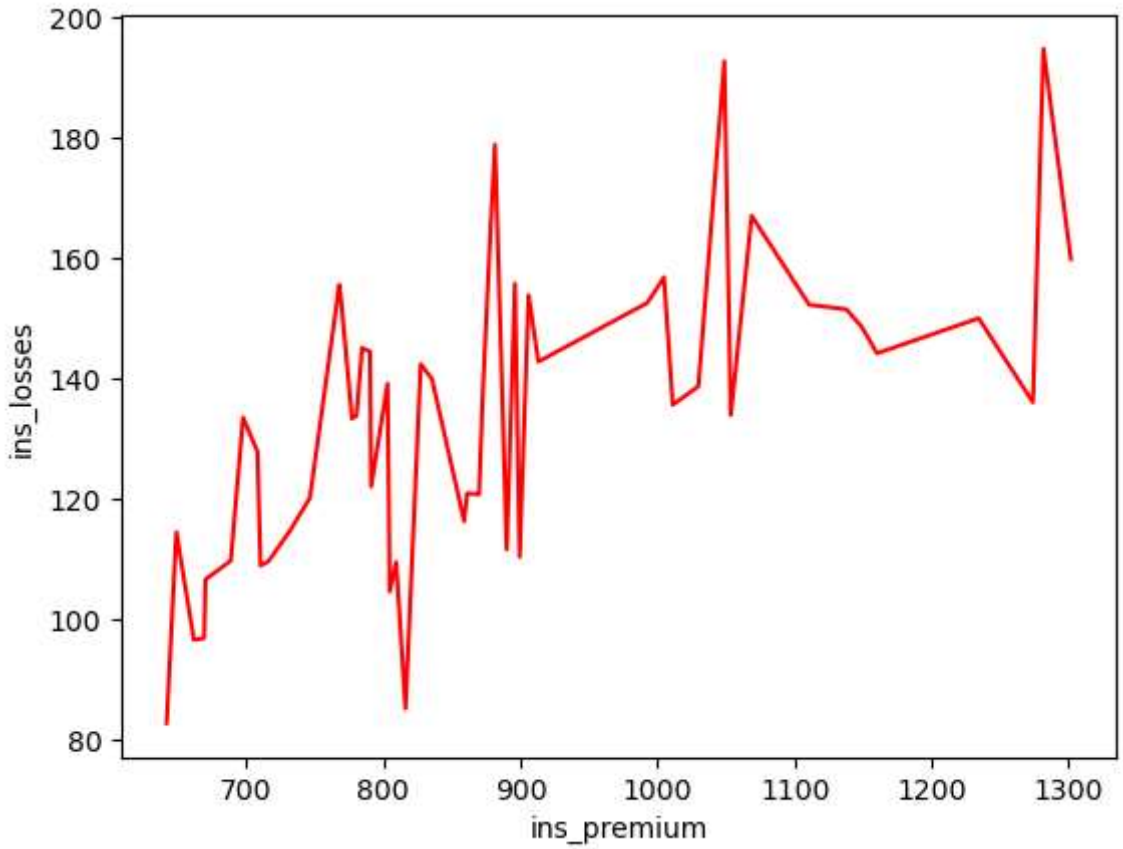
INFERENCE: As the rate of alcohol consumption increasing, the total accedents are also increasing.

In [25]:

```
ms.lineplot(x="ins_premium",y="ins_losses",data=d,color='red')
```

Out[25]:

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



INFERENCE: As the rate of insurence_premium increasing, the insurrence_losses are also increasing.

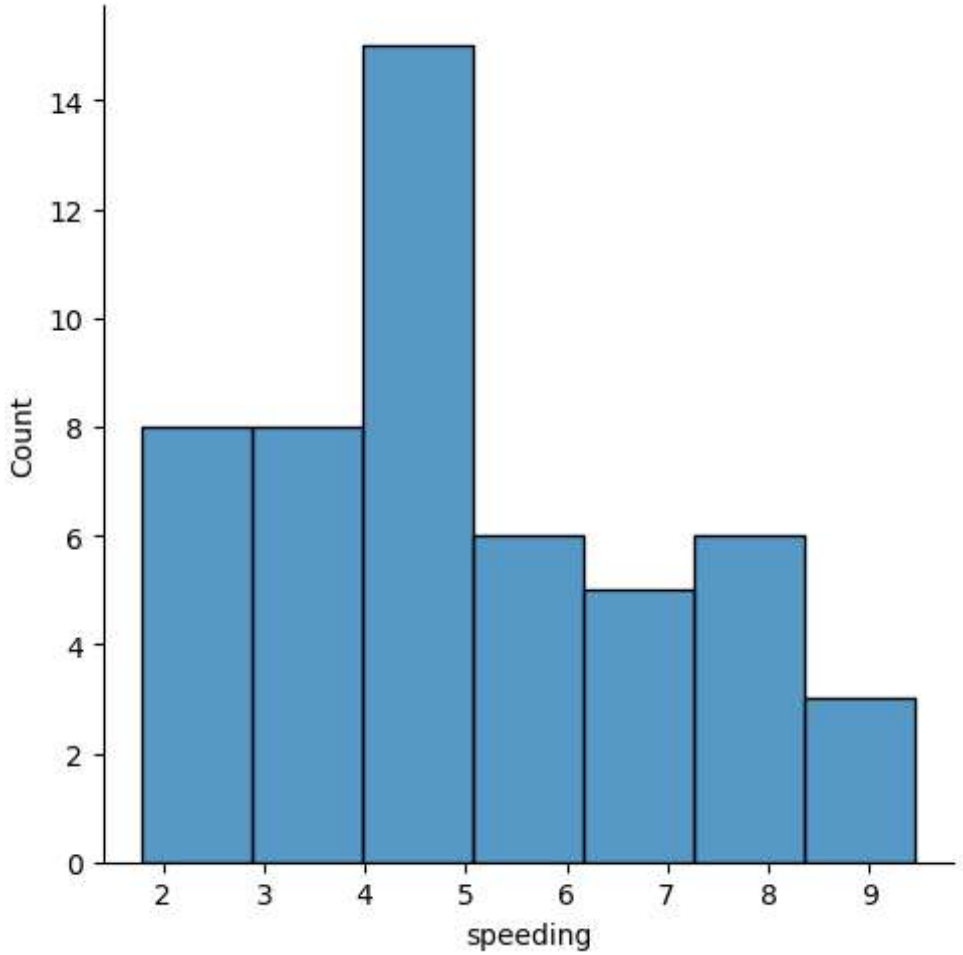
DISPLOT

In [27]:

```
ms.displot(d['speeding'])
```

Out[27]:

```
<seaborn.axisgrid.FacetGrid at 0x2b3ca252ad0>
```



INFERENCE: Maximum car crashes are happened in between 4-5 speed

DISTRIBUTION PLOT

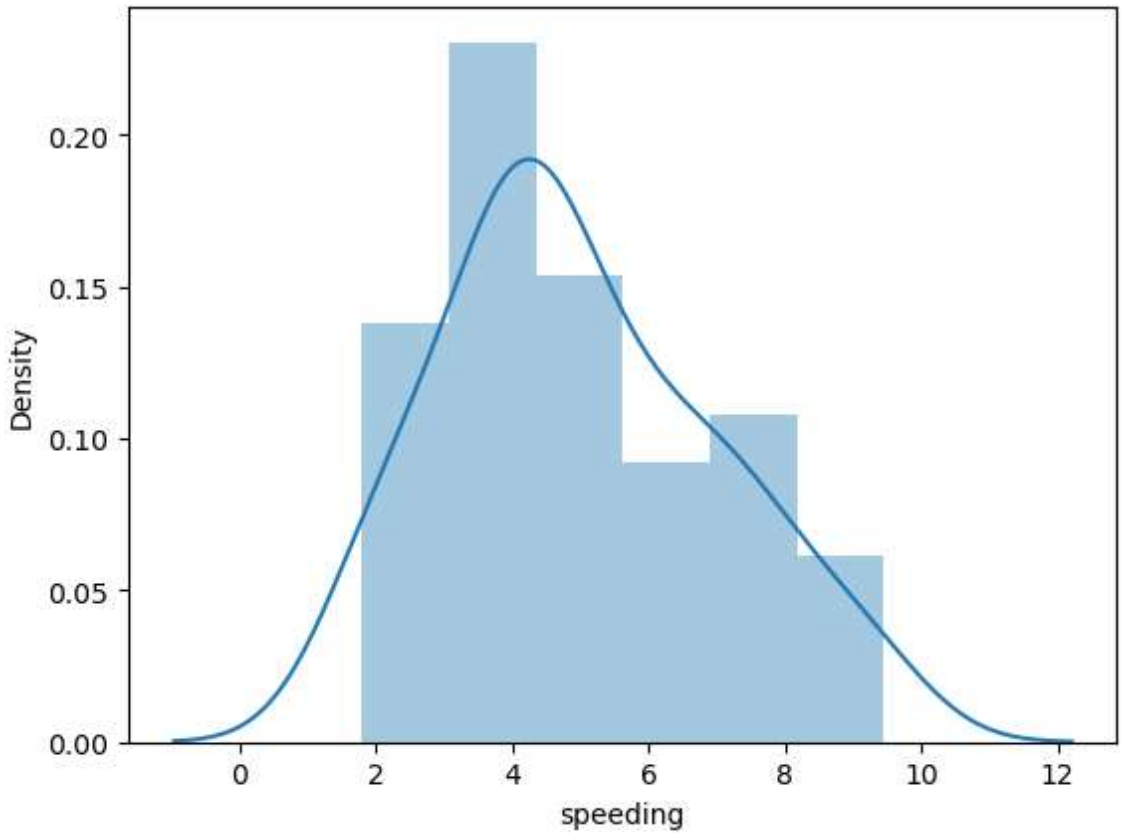
In [30]:

```
ms.distplot(d['speeding'])
```

```
C:\Users\thanu\AppData\Local\Temp\ipykernel_9692\2289468554.py:1: UserWarn
ing:

`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 wi
th
similar flexibility) or `histplot` (an axes-level function for histogram
s).

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

  ms.distplot(d['speeding'])
```

Out[30]:

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



In [32]:

```
d['speeding'].mean()
```

Out[32]:

```
4.998196078431373
```

In [34]:

```
d['speeding'].median()
```

Out[34]:

```
4.6080000000000005
```

In [36]:

```
d['speeding'].mode()
```

Out[36]:

```
0    4.968
Name: speeding, dtype: float64
```

INFERENCE:The distribution is almost symmetrical

RELPLOT

In [39]:

```
ms.relplot(x="alcohol",y="total",data=d,hue="abbrev")
```

Out[39]:

```
<seaborn.axisgrid.FacetGrid at 0x2b3cb378250>
```

INFERENCE: The plotting is done based on states

BARPLOT

In [41]:

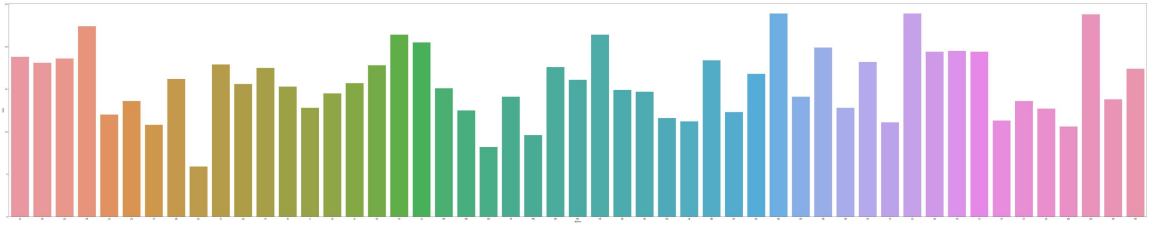
```
import matplotlib.pyplot as p
```

In [44]:

```
p.subplots(figsize=(106,20))
ms.barplot(x="abbrev",y="total",data=d)
```

Out[44]:

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

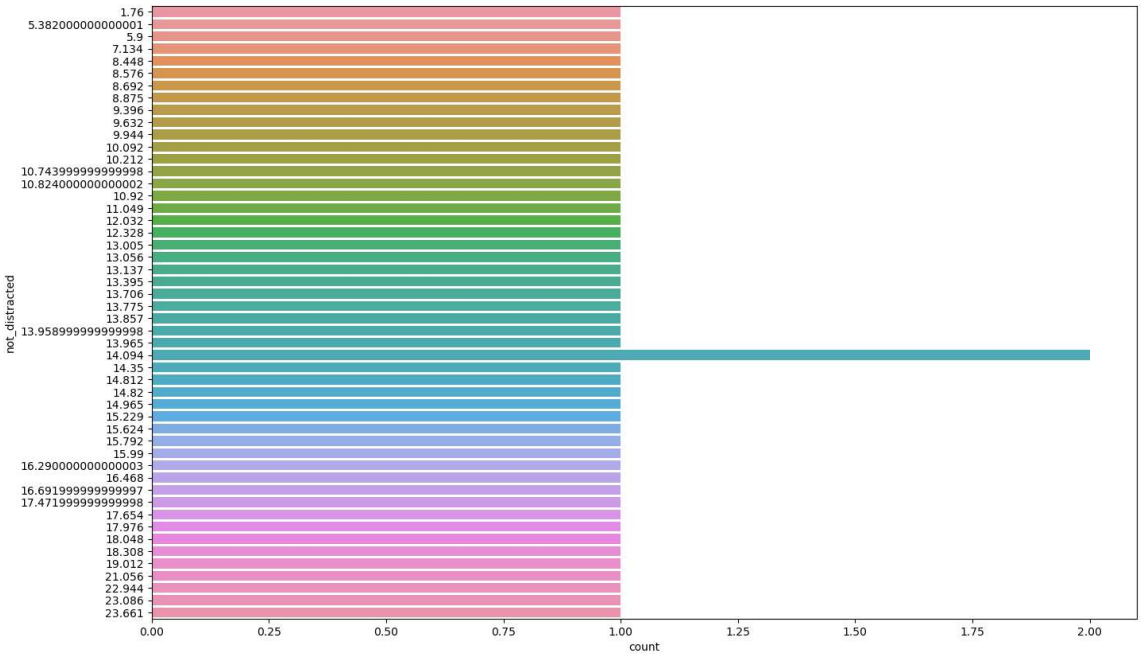


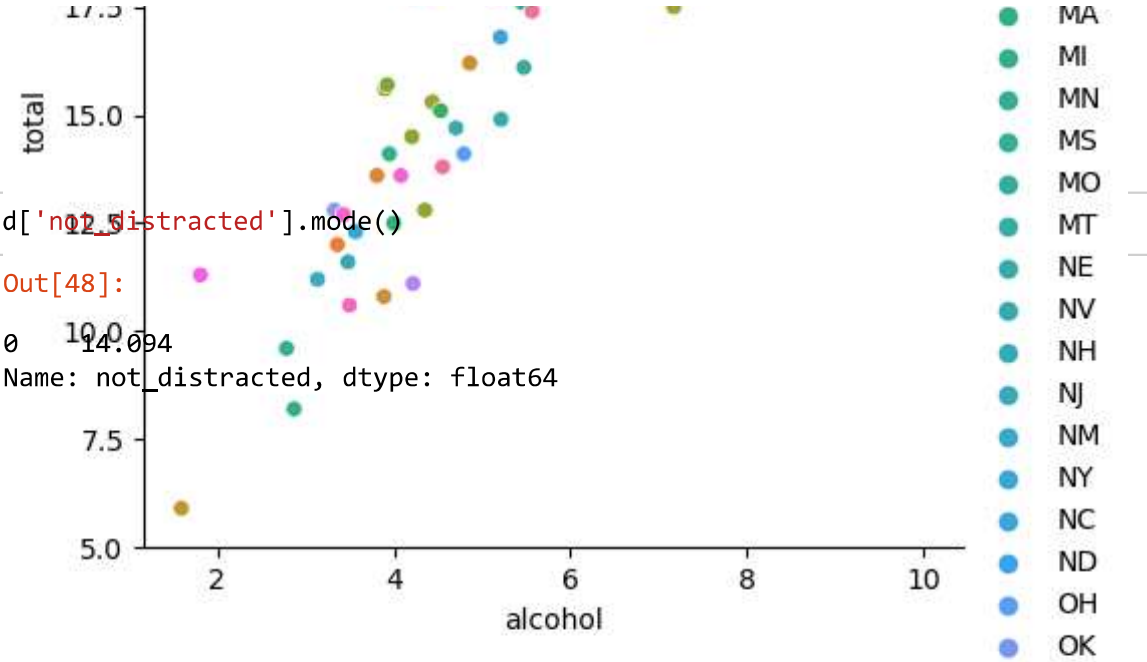
COUNTPLOT

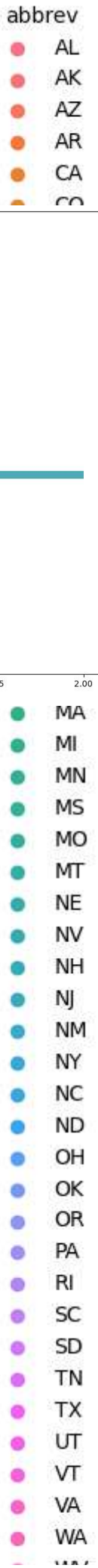
```
p.subplots(figsize=(16,10))
ms.countplot(y="not_distracted",data=d,orient='h')
```

Out[46]:

```
<Axes: xlabel='count', ylabel='not_distracted'>
```








```
d['not_distracted'].mode()
```

Out[48]:

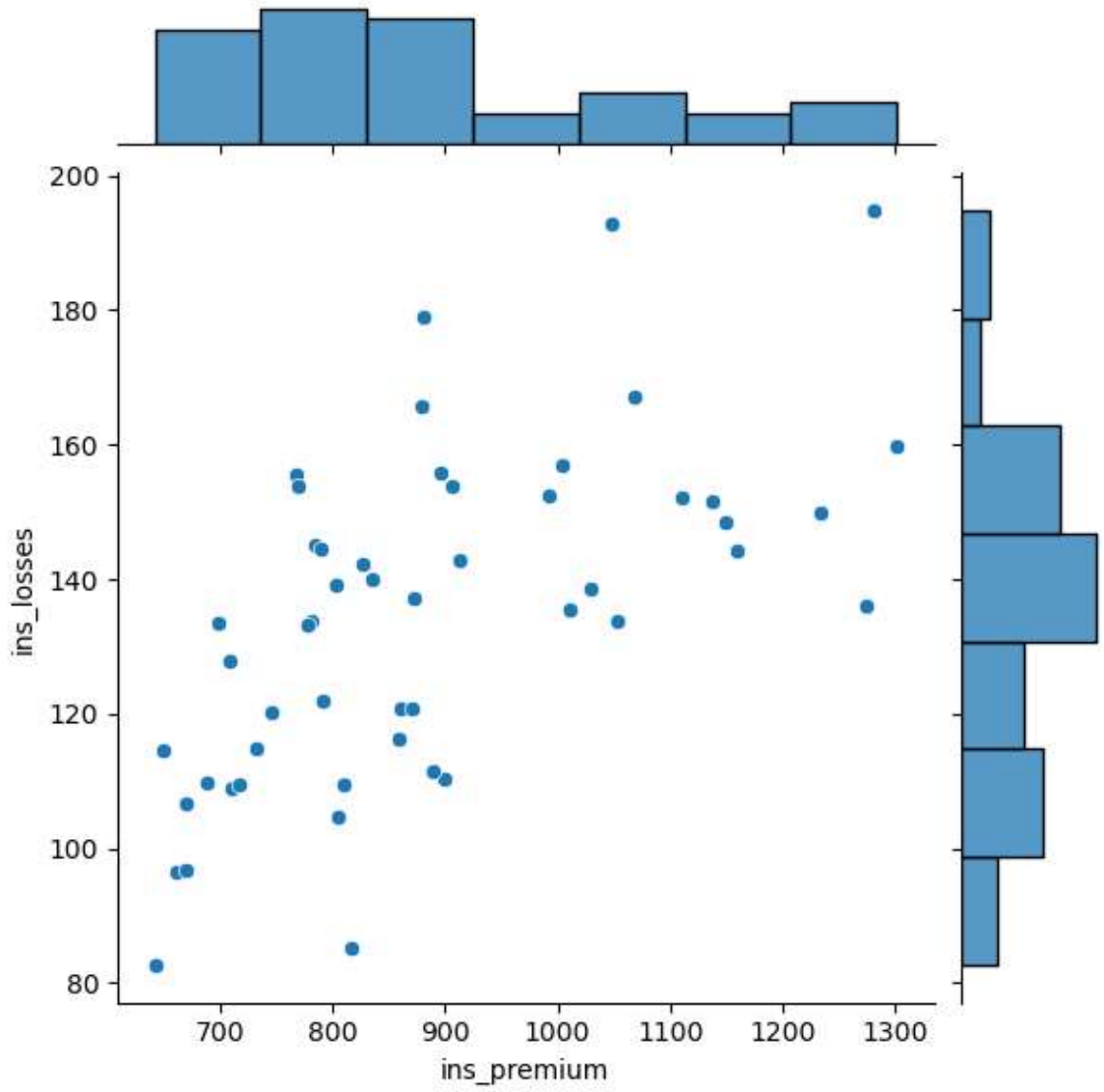
```
0    14.094
Name: not_distracted, dtype: float64
```

WV
WI
WY

```
ms.jointplot(x="ins_premium",y="ins_losses",data=d)
```

Out[50]:

```
<seaborn.axisgrid.JointGrid at 0x2b3d07cb5d0>
```



INFERENCE: As the Insurance Premium is taken but the Insurance losses are also high in state wise

HEATMAP FOR CORRELATION

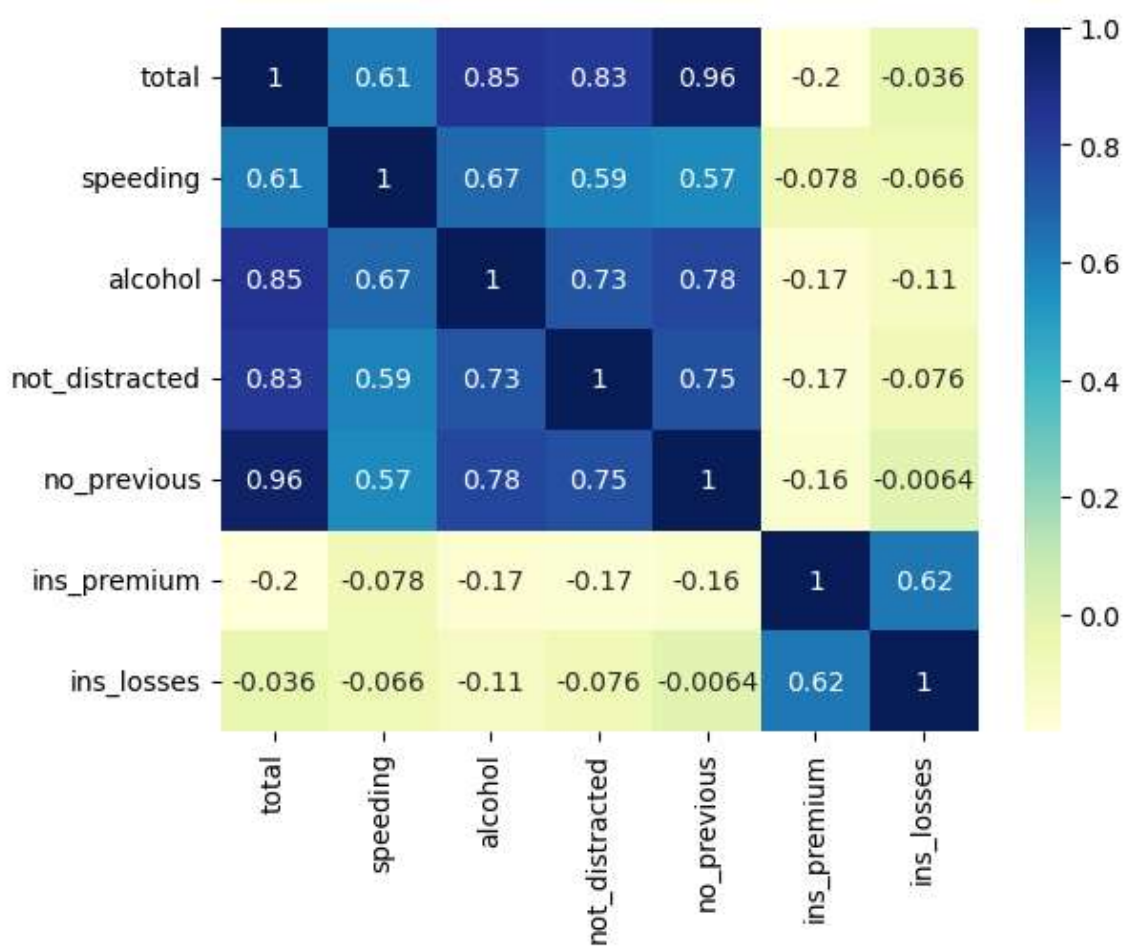
In [53]:

```
ms.heatmap(d.corr(),annot=True,cmap="YlGnBu")
```

```
C:\Users\thanu\AppData\Local\Temp\ipykernel_9692\1006457501.py:1: FutureWa
rning: The default value of numeric_only in DataFrame.corr is deprecated.
In a future version, it will default to False. Select only valid columns o
r specify the value of numeric_only to silence this warning.
  ms.heatmap(d.corr(),annot=True,cmap="YlGnBu")
```

Out[53]:

```
<Axes: >
```



INFERENCE: Total is highly correlated with Alcohol to test car crash

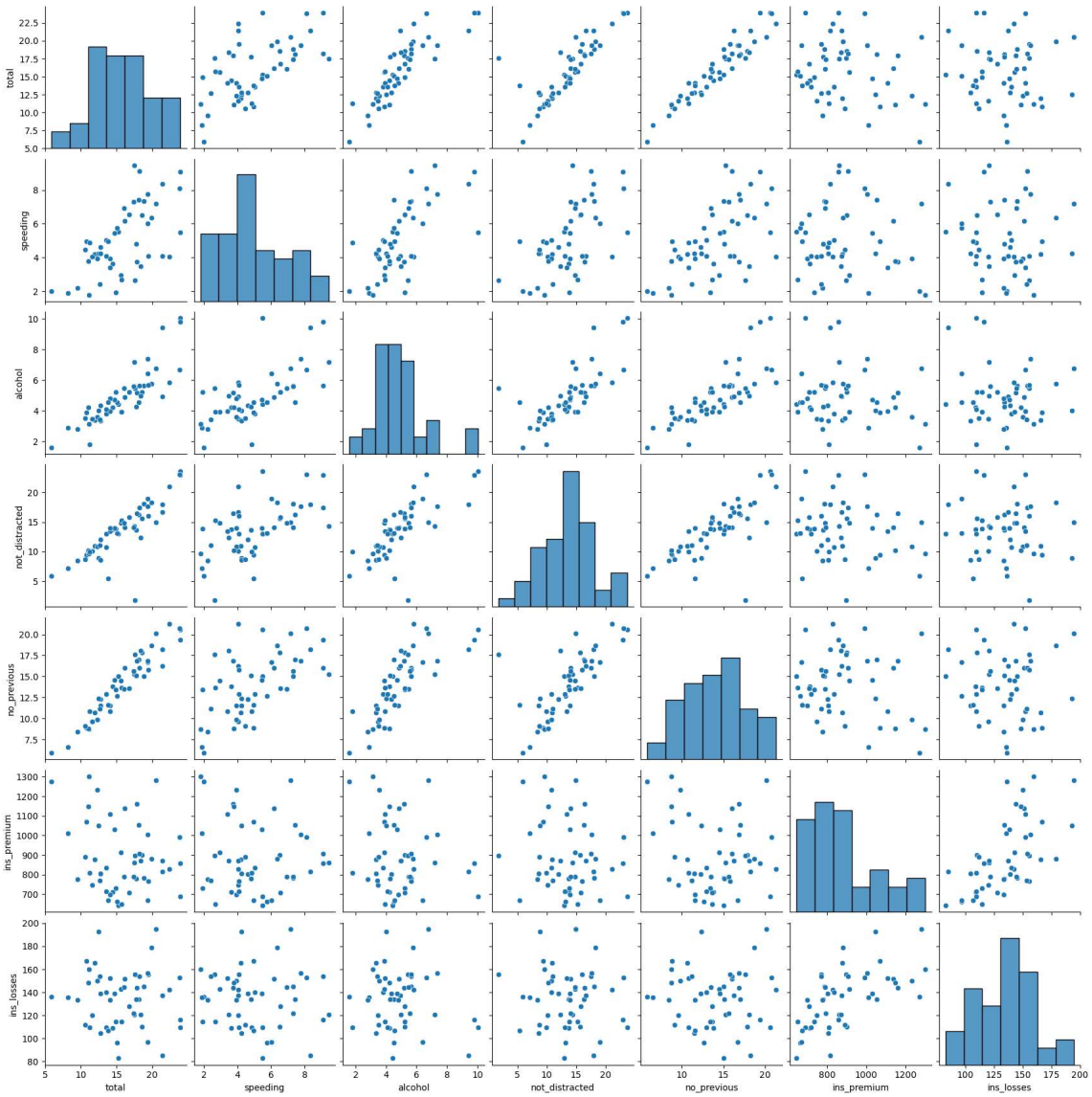
PAIRPLOT

In [58]:

```
ms.pairplot(d)
```

Out[58]:

```
<seaborn.axisgrid.PairGrid at 0x2b3deffbad0>
```




```
INFERENCE: Above methods tells us the Trend between all Numerical Analasys
```

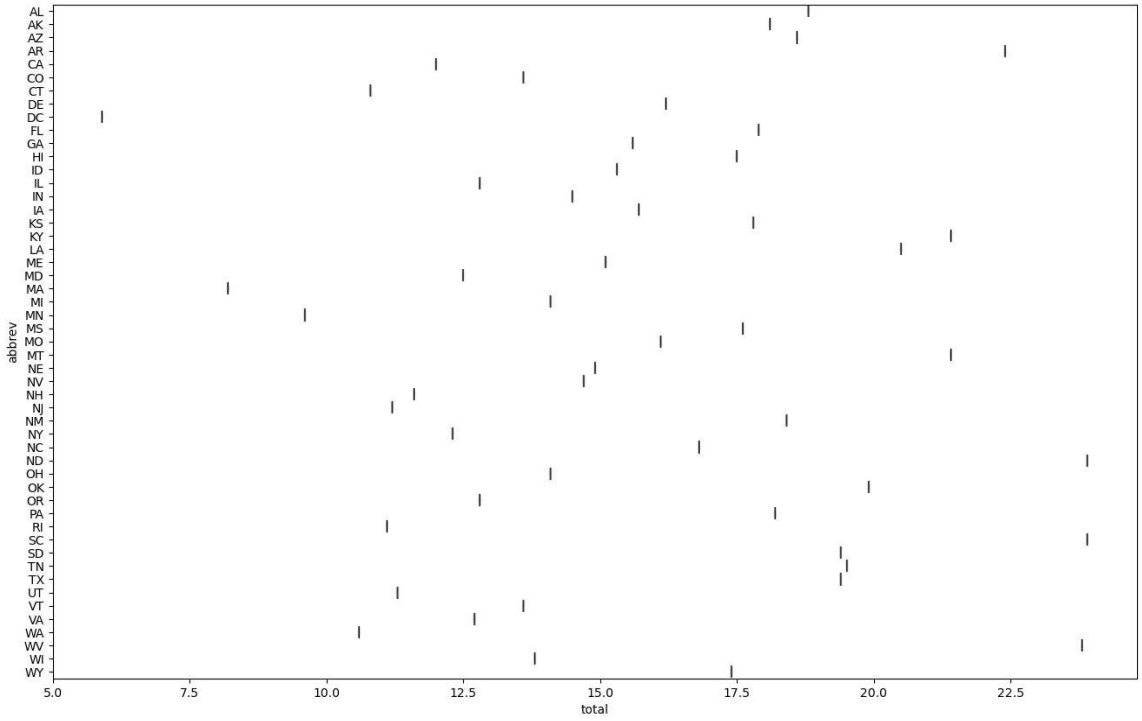
BOXPLOT

In [60]:

```
p.subplots(figsize=(16,10))
ms.boxplot(x="total",y="abbrev",data=d)
```

Out[60]:

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



INFERENCE: As the states of USA are unique so as mean, median and mode are equal. It indicates only one value