purushothamreddy-21bce5289

September 14, 2023

[7]:
```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

[8]:
```
data=pd.read_csv("/content/car_crashes.csv")
```

car_crashes Accidents in the states of the USA are examined. This is the data set of the cause of the accidents and the cost to the accident insurance companies.

· total -> Number of drivers involved in fatal collisions per billion miles (5.900–23.900)

· speeding -> Percentage Of Drivers Involved In Fatal Collisions Who Were Speeding (1.792–9.450)

· alcohol -> Percentage Of Drivers Involved In Fatal Collisions Who Were Alcohol-Impaired (1.593–10.038)

· not_distracted -> Percentage Of Drivers Involved In Fatal Collisions Who Were Not Distracted (1.760–23.661)

· no_previous -> Percentage Of Drivers Involved In Fatal Collisions Who Had Not Been Involved In Any Previous Accidents (5.900–21.280)

· ins_premium -> Car Insurance Premiums (641.960–1301.520)

· ins_losses -> Losses incurred by insurance companies for collisions per insured driver (82.75–194.780)

· abbrev -> USA states **bold text**

[9]:
```
data.head()
```

[9]:

	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		
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		

```
2      110.35     AZ
3      142.39     AR
4      165.63     CA
```

[10]:
```
data.tail()
```

[10]:
```
    total  speeding  alcohol  not_distracted  no_previous  ins_premium  \
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
46      153.72     VA
47      111.62     WA
48      152.56     WV
49      106.62     WI
50      122.04     WY
```

[11]:
```
x = data["total"].values
y= data["alcohol"].values
z = data["speeding"].values
```

[12]:
```
x
```

[12]:
```
array([18.8, 18.1, 18.6, 22.4, 12. , 13.6, 10.8, 16.2,  5.9, 17.9, 15.6,
       17.5, 15.3, 12.8, 14.5, 15.7, 17.8, 21.4, 20.5, 15.1, 12.5,  8.2,
       14.1,  9.6, 17.6, 16.1, 21.4, 14.9, 14.7, 11.6, 11.2, 18.4, 12.3,
       16.8, 23.9, 14.1, 19.9, 12.8, 18.2, 11.1, 23.9, 19.4, 19.5, 19.4,
       11.3, 13.6, 12.7, 10.6, 23.8, 13.8, 17.4])
```

[13]:
```
y
```

[13]:
```
array([ 5.64 ,  4.525,  5.208,  5.824,  3.36 ,  3.808,  3.888,  4.86 ,
        1.593,  5.191,  3.9  ,  7.175,  4.437,  4.352,  4.205,  3.925,
        4.272,  4.922,  6.765,  4.53 ,  4.   ,  2.87 ,  3.948,  2.784,
        5.456,  5.474,  9.416,  5.215,  4.704,  3.48 ,  3.136,  4.968,
        3.567,  5.208, 10.038,  4.794,  5.771,  3.328,  5.642,  4.218,
        9.799,  6.402,  5.655,  7.372,  1.808,  4.08 ,  3.429,  3.498,
        6.664,  4.554,  5.568])
```

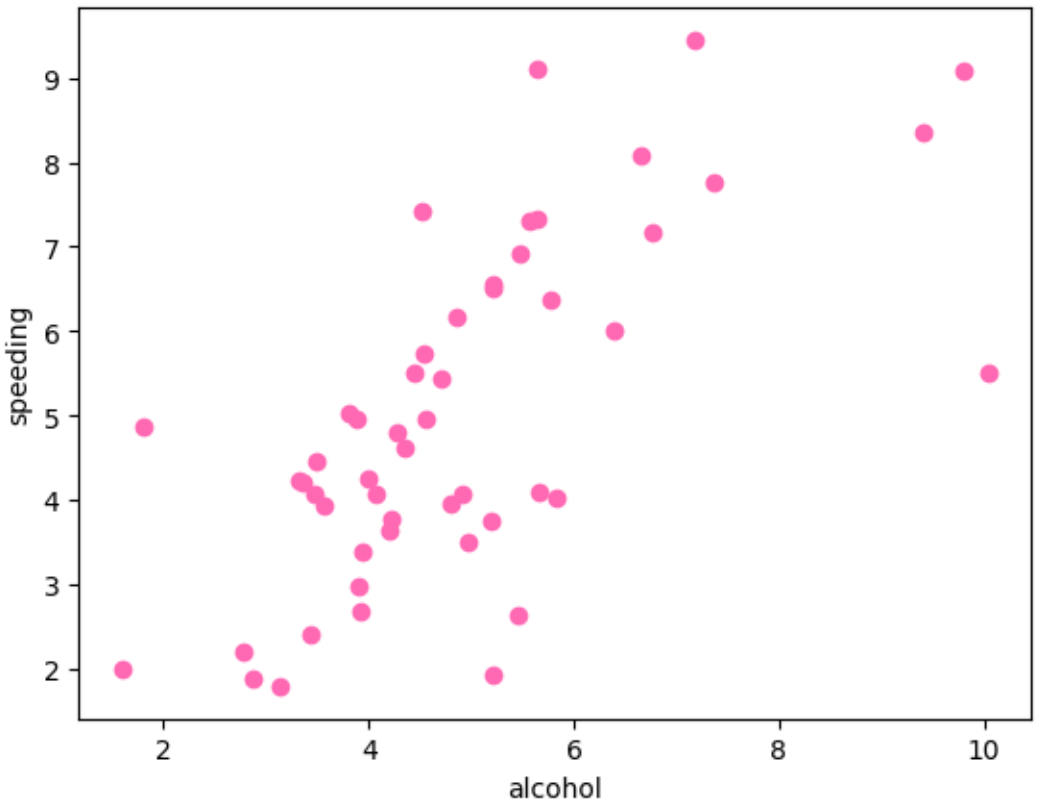
[14]:
```
z
```

[14]:
```
array([7.332, 7.421, 6.51 , 4.032, 4.2  , 5.032, 4.968, 6.156, 2.006,
       3.759, 2.964, 9.45 , 5.508, 4.608, 3.625, 2.669, 4.806, 4.066,
       7.175, 5.738, 4.25 , 1.886, 3.384, 2.208, 2.64 , 6.923, 8.346,
```

```
1.937, 5.439, 4.06 , 1.792, 3.496, 3.936, 6.552, 5.497, 3.948,
6.368, 4.224, 9.1  , 3.774, 9.082, 6.014, 4.095, 7.76 , 4.859,
4.08 , 2.413, 4.452, 8.092, 4.968, 7.308])
```

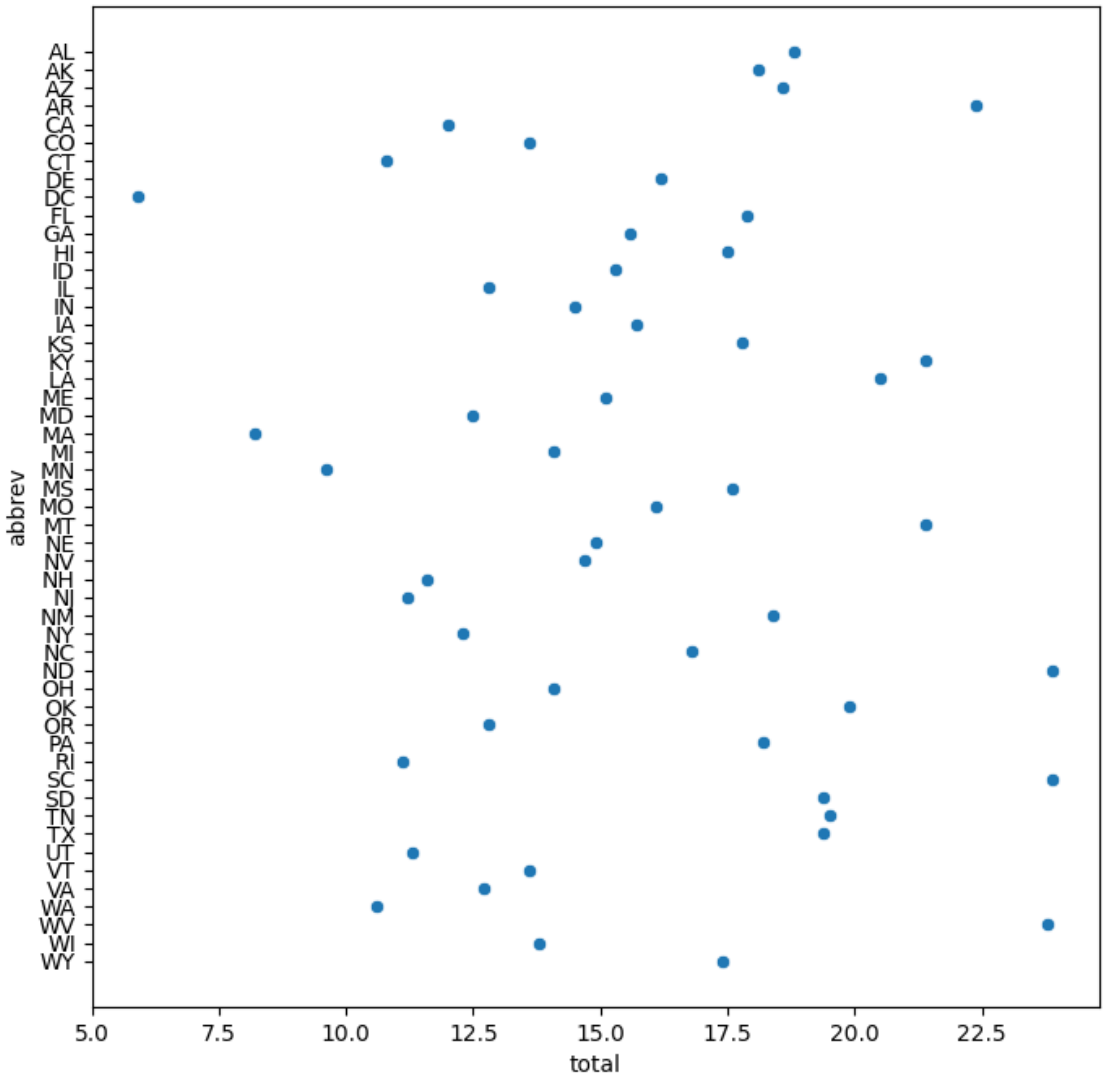
[15]:
```
plt.scatter("alcohol","speeding",data=data,color ='hotpink')
plt.xlabel("alcohol")
plt.ylabel("speeding")
```

[15]:
```
Text(0, 0.5, 'speeding')
```



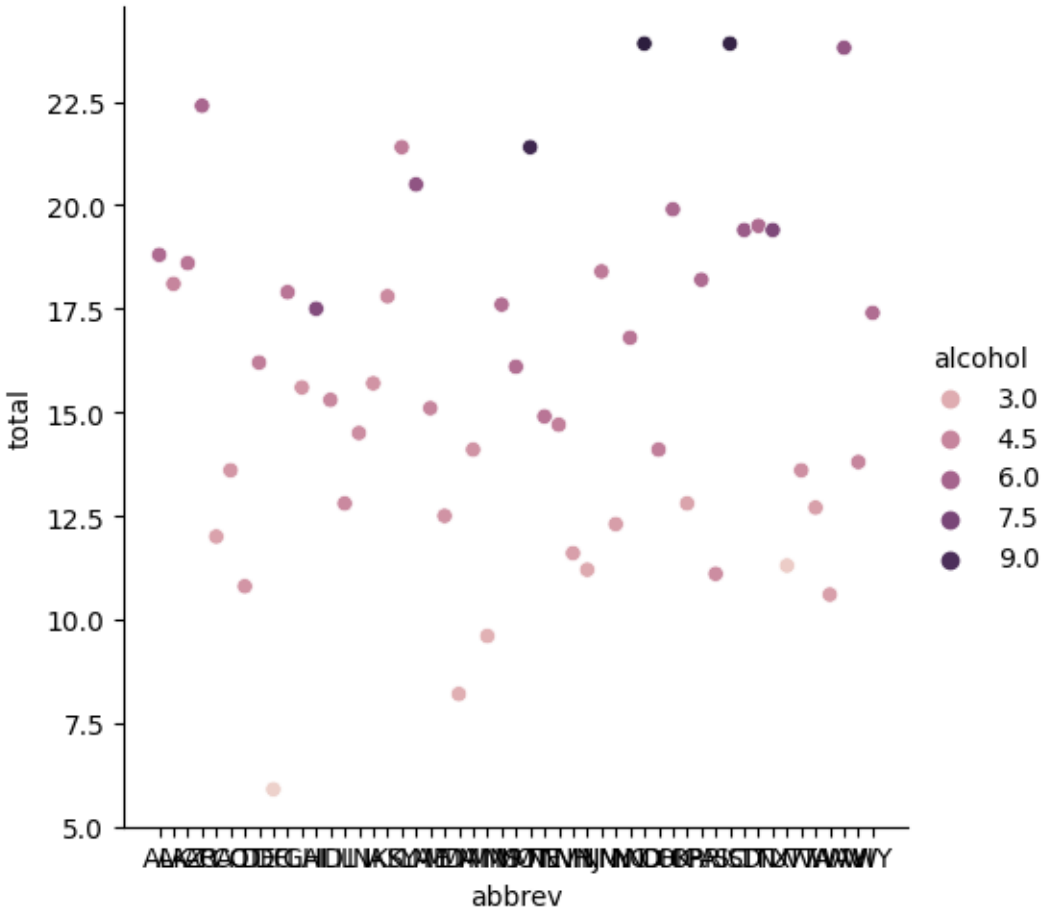
The scatter plot indicates a negative correlation between alcohol and speeding. As the alcohol level increases, the instances of speeding decrease.

[16]:
```
plt.figure(figsize=(8,8))
sns.scatterplot(x="total",y="abbrev",data=data)
plt.show()
```



The scatter plot shows a positive correlation between the variables 'total' and 'state'. As 'total' increases, the corresponding value for 'state' also increases.

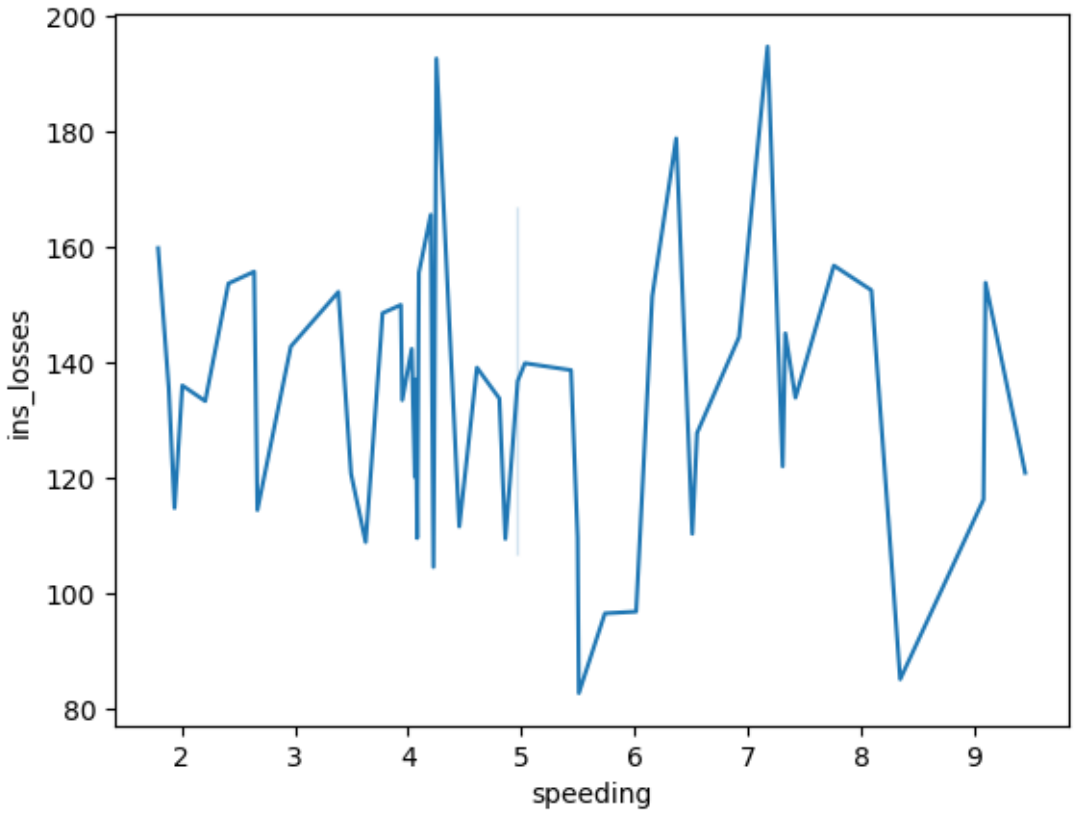
[44]:
```
sns.relplot(y='total', x='abbrev', data=data ,hue ='alcohol');
```



The scatter plot indicates a correlation between the variables 'abbrev' and 'total'. The color and size of the points represent the alcohol content and the total number of drinks consumed, respectively.

[18]:
```
sns.lineplot(x="speeding",y="ins_losses",data=data)
```

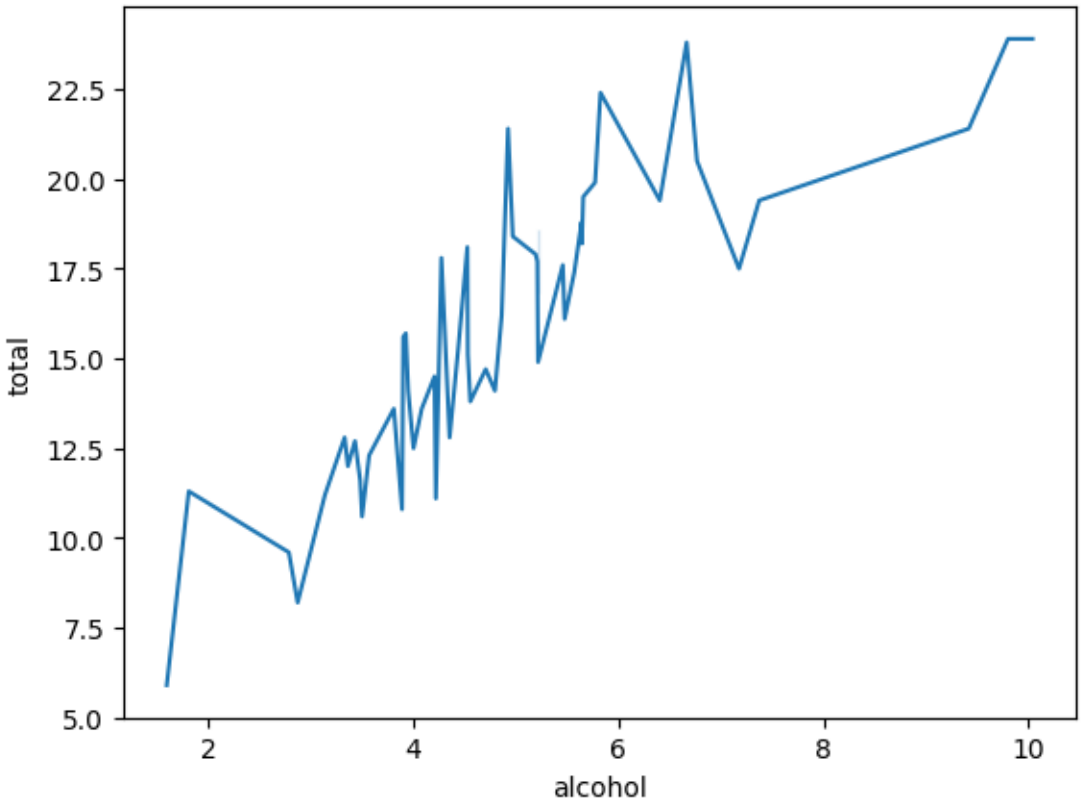
[18]:
```
<Axes: xlabel='speeding', ylabel='ins_losses'>
```



The line graph shows a fluctuating relationship between 'speeding' and 'insurance losses'. There are peaks and valleys, indicating variations in insurance losses with changes in speeding.

[19]:
```
sns.lineplot(x="alcohol",y="total",data=data)
```

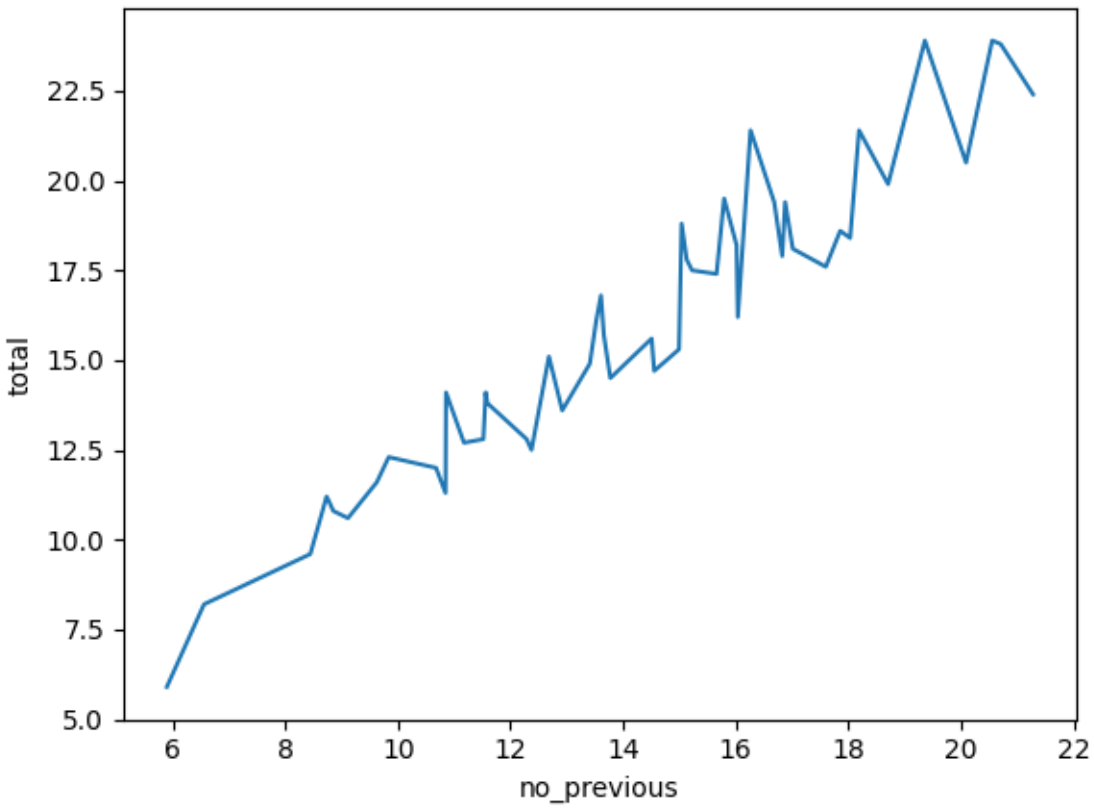
[19]:
```
<Axes: xlabel='alcohol', ylabel='total'>
```



The line graph shows a positive correlation between 'alcohol' and 'total'. As the alcohol level increases, the total also increases.

[20]:
```
sns.lineplot(x="no_previous",y="total",data=data)
```

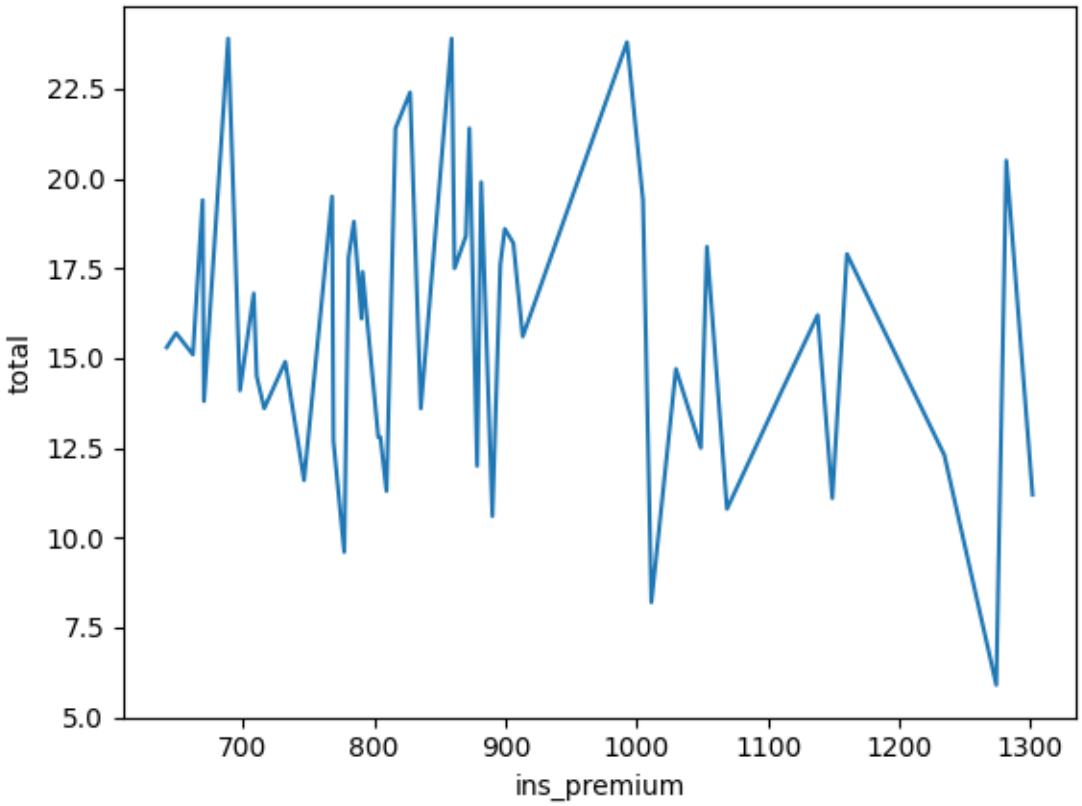
[20]:
```
<Axes: xlabel='no_previous', ylabel='total'>
```



The line graph shows a positive correlation between 'alcohol' and 'total'. As the alcohol level increases, the total also increases.

[21]:
```
sns.lineplot(x="ins_premium",y="total",data=data)
```

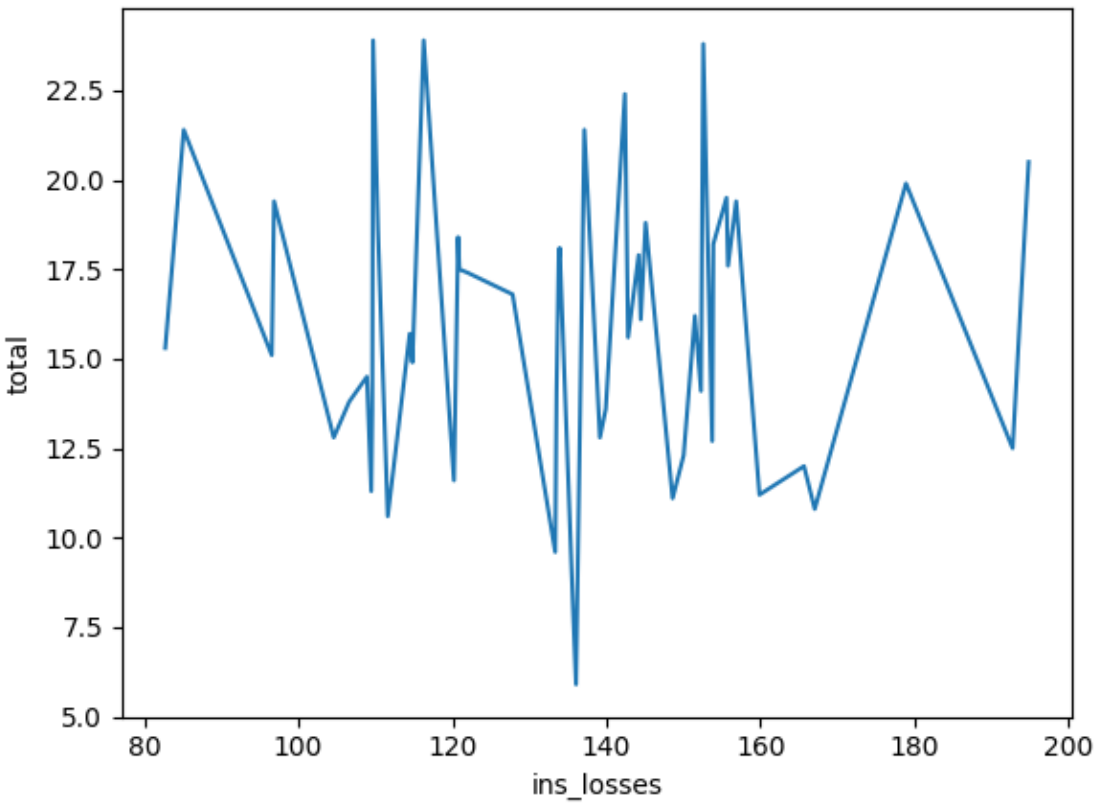
[21]:
```
<Axes: xlabel='ins_premium', ylabel='total'>
```



The line graph shows a decreasing trend between 'ins_premium' and 'total'. As the insurance premium increases, the total decreases.

[22]:
```
sns.lineplot(x="ins_losses",y="total",data=data)
```

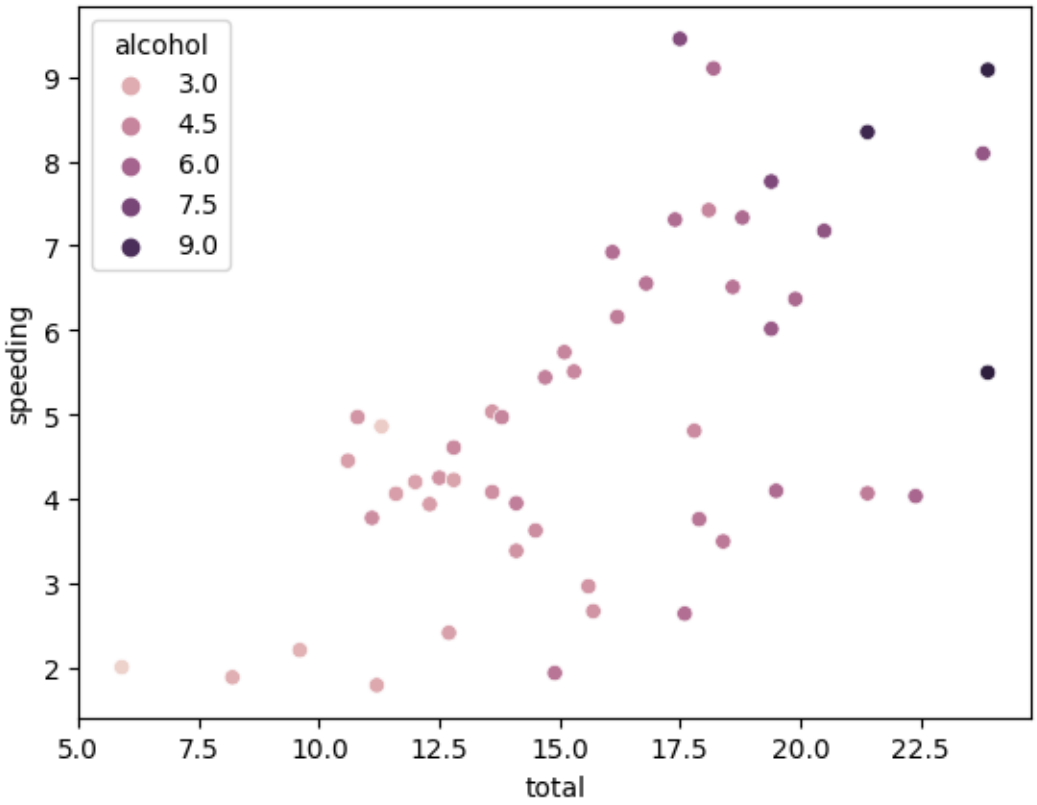
[22]:
```
<Axes: xlabel='ins_losses', ylabel='total'>
```



The line graph shows a slight overall downward trend in 'total' with respect to 'ins_losses'. This indicates that as insurance losses increase, the total decreases.

[23]:
```
sns.scatterplot(x="total",y="speeding",data=data,hue="alcohol")
```

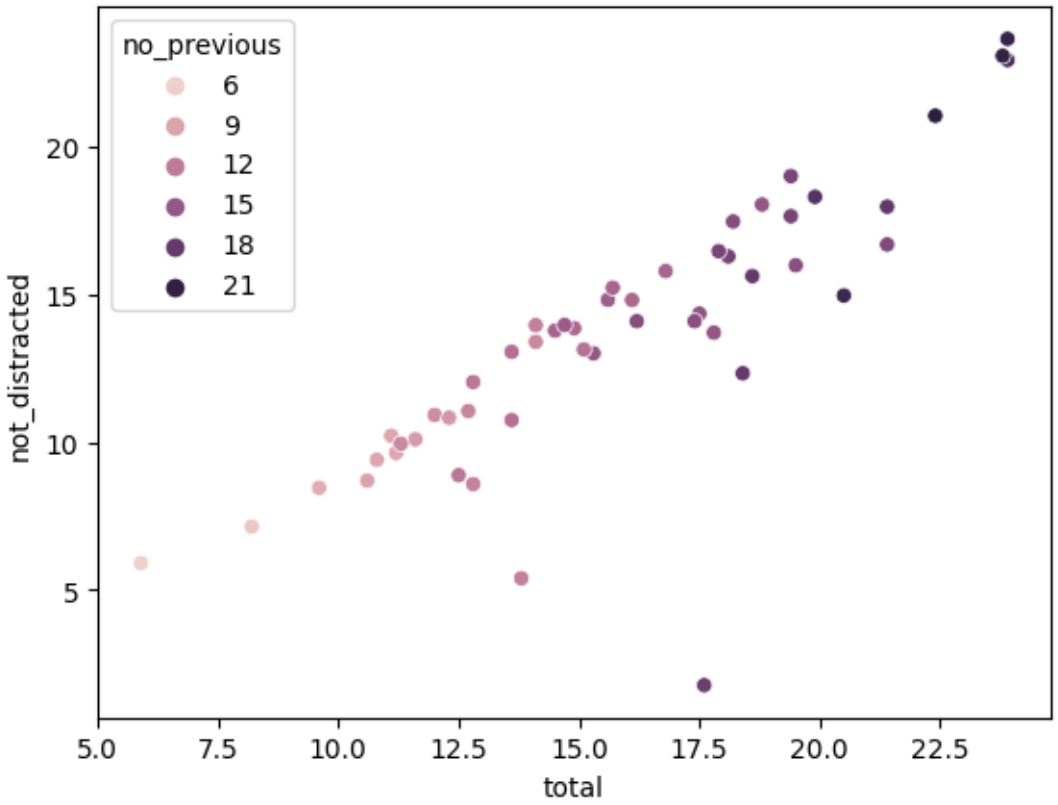
[23]:
```
<Axes: xlabel='total', ylabel='speeding'>
```



The scatter plot indicates a correlation between 'total' and 'spending'. The data points are scattered across the plot, with a higher concentration of points in the lower left corner. The larger points labeled with numbers: 3.0, 4.5, and 9.0 are located in the upper right corner of the plot.

[24]:
```
sns.scatterplot(x="total",y="not_distracted",data=data,hue="no_previous")
```

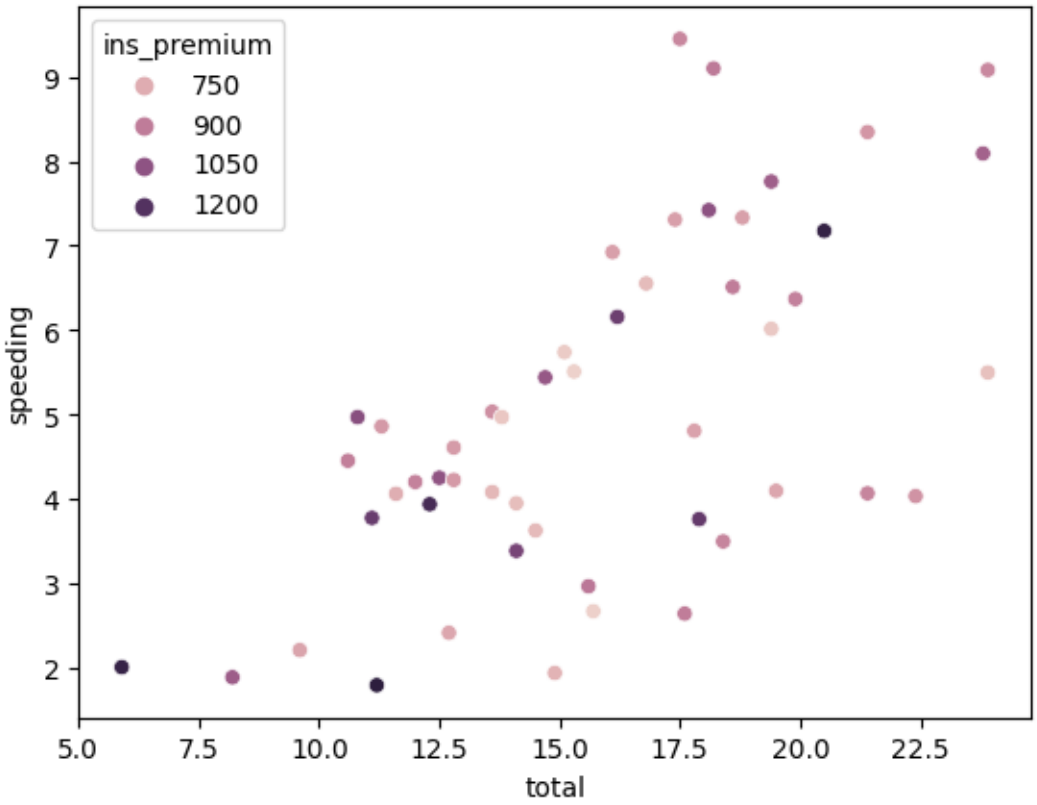
[24]:
```
<Axes: xlabel='total', ylabel='not_distracted'>
```



The scatter plot shows a positive correlation between 'total' and 'no_previous'. As the 'total' increases, the 'no_previous' also increases.

[25]:
```
sns.scatterplot(x="total",y="speeding",data=data,hue="ins_premium")
```

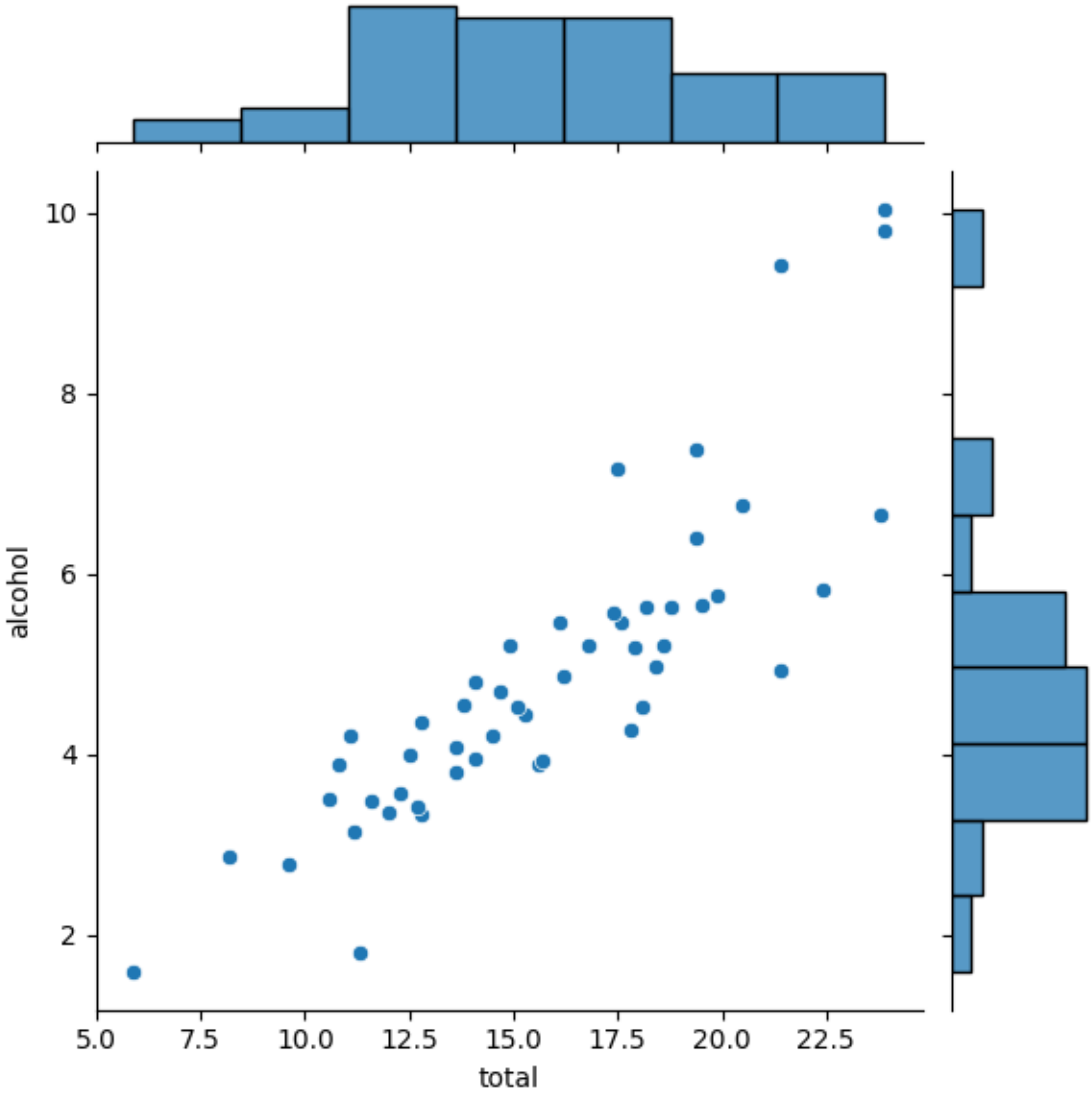
[25]:
```
<Axes: xlabel='total', ylabel='speeding'>
```



The scatter plot shows the relationship between 'total' and 'spending' variables, with data points colored based on the 'ins_premium' variable. There is no clear pattern or correlation observed between the variables in the plot. The distribution of data points is quite scattered, indicating a high degree of variability in the data.

[26]:
```
sns.jointplot(x="total",y="alcohol",data=data)
```

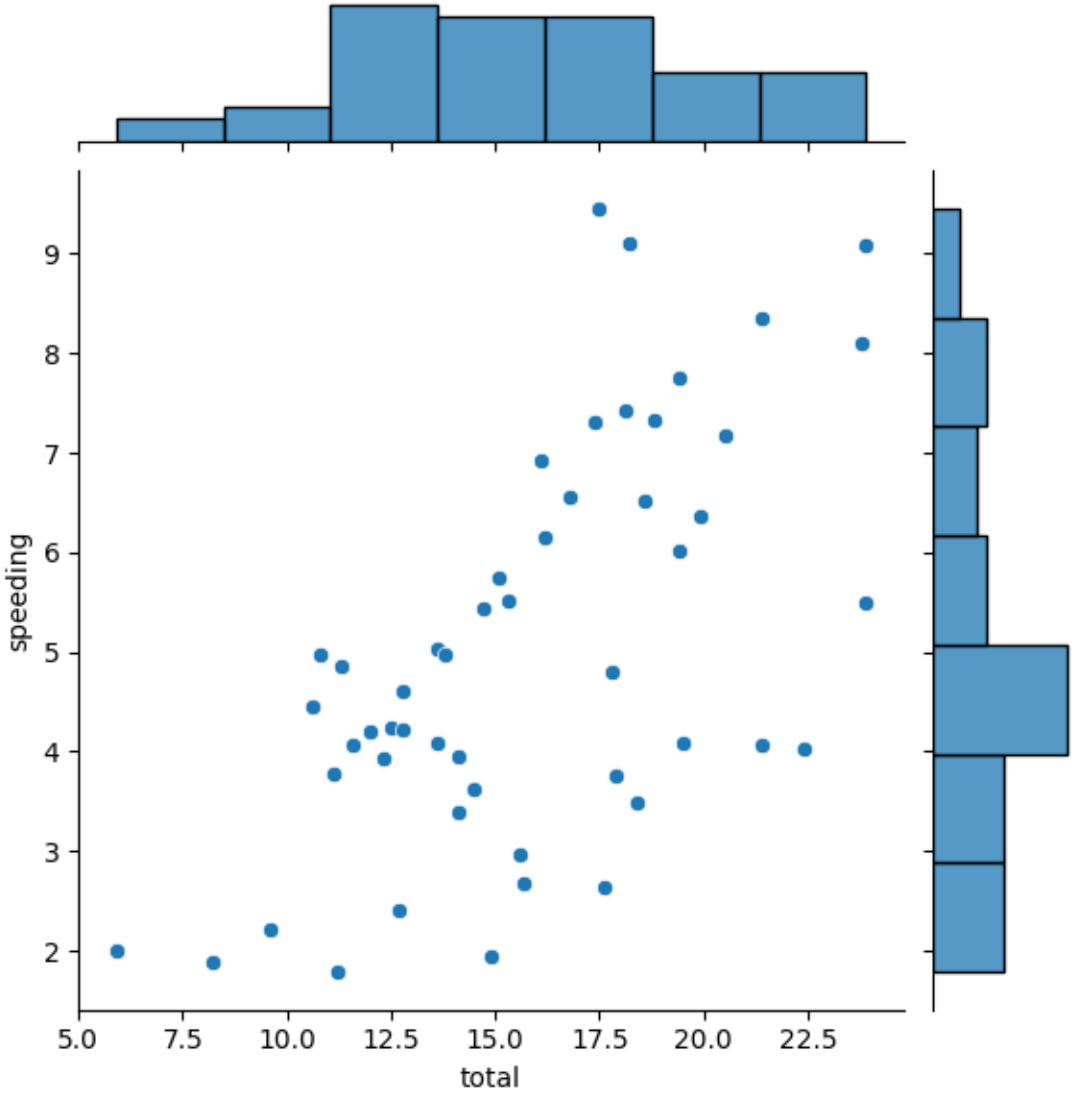
[26]:
```
<seaborn.axisgrid.JointGrid at 0x7a037f42bf40>
```



The scatter plot shows a positive correlation between 'total' and 'alcohol'. The histograms indicate that the distribution of 'total' is fairly uniform, while 'alcohol' decreases as we move from high to low.

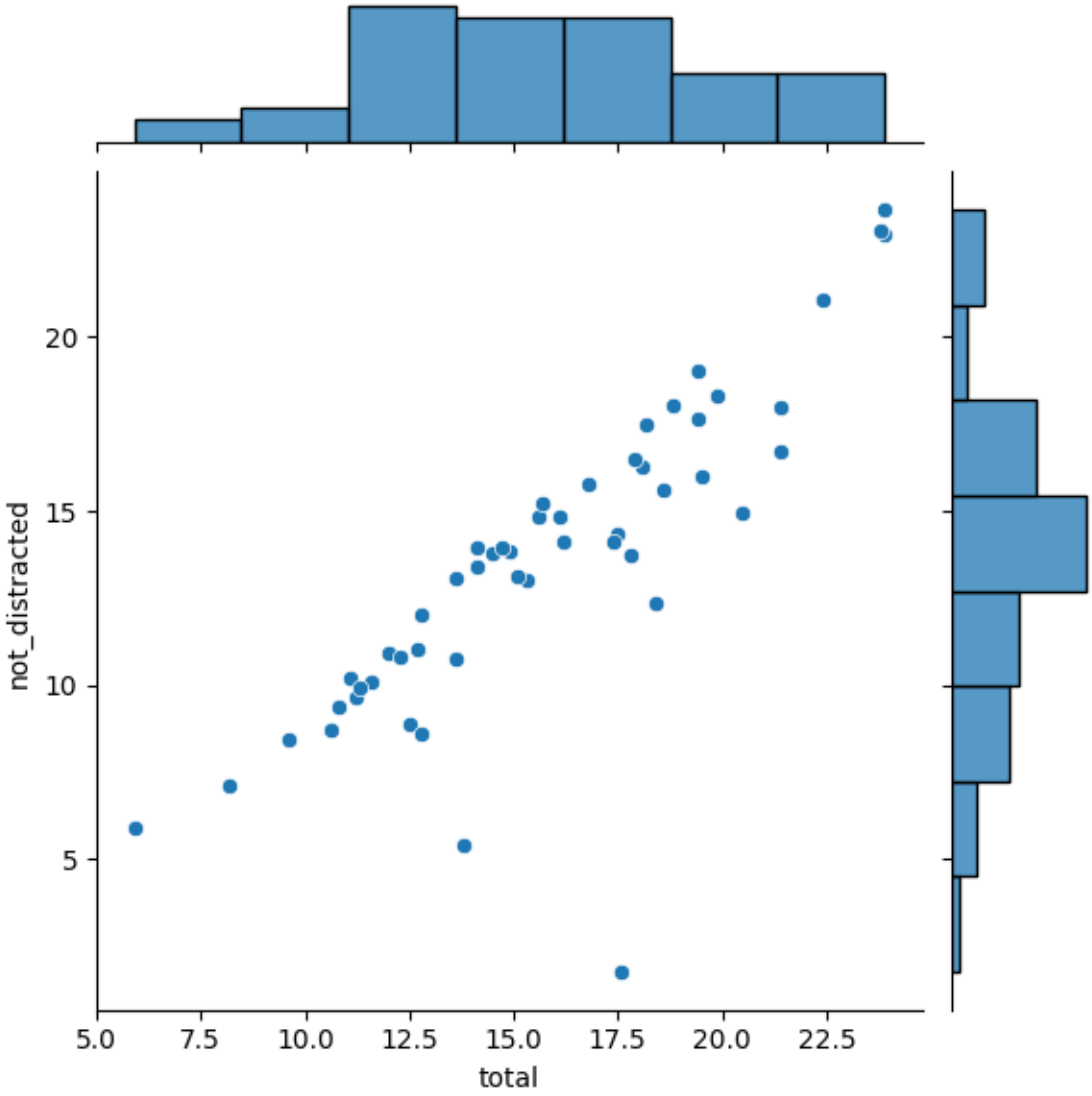
[27]:
```
sns.jointplot(x="total",y="speeding",data=data)
```

[27]:
```
<seaborn.axisgrid.JointGrid at 0x7a037f48a6b0>
```



[28]:
```
sns.jointplot(x="total",y="not_distracted",data=data)
```

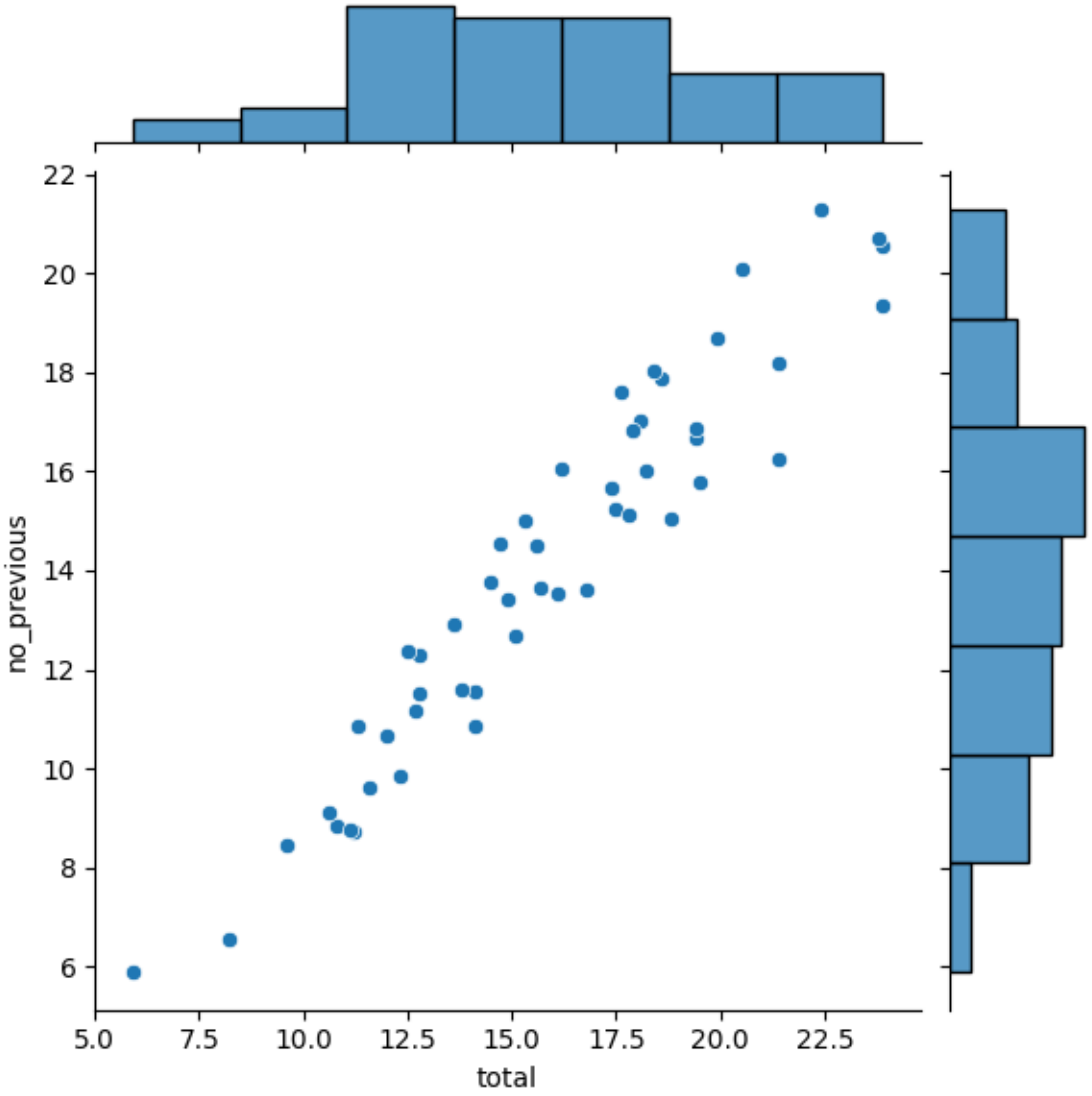
[28]: <seaborn.axisgrid.JointGrid at 0x7a037f236b00>



The scatter plot indicates a positive linear relationship between 'total' and 'not satisfied'. The histograms show the distributions of these two variables.

[29]:
```
sns.jointplot(x="total",y="no_previous",data=data)
```

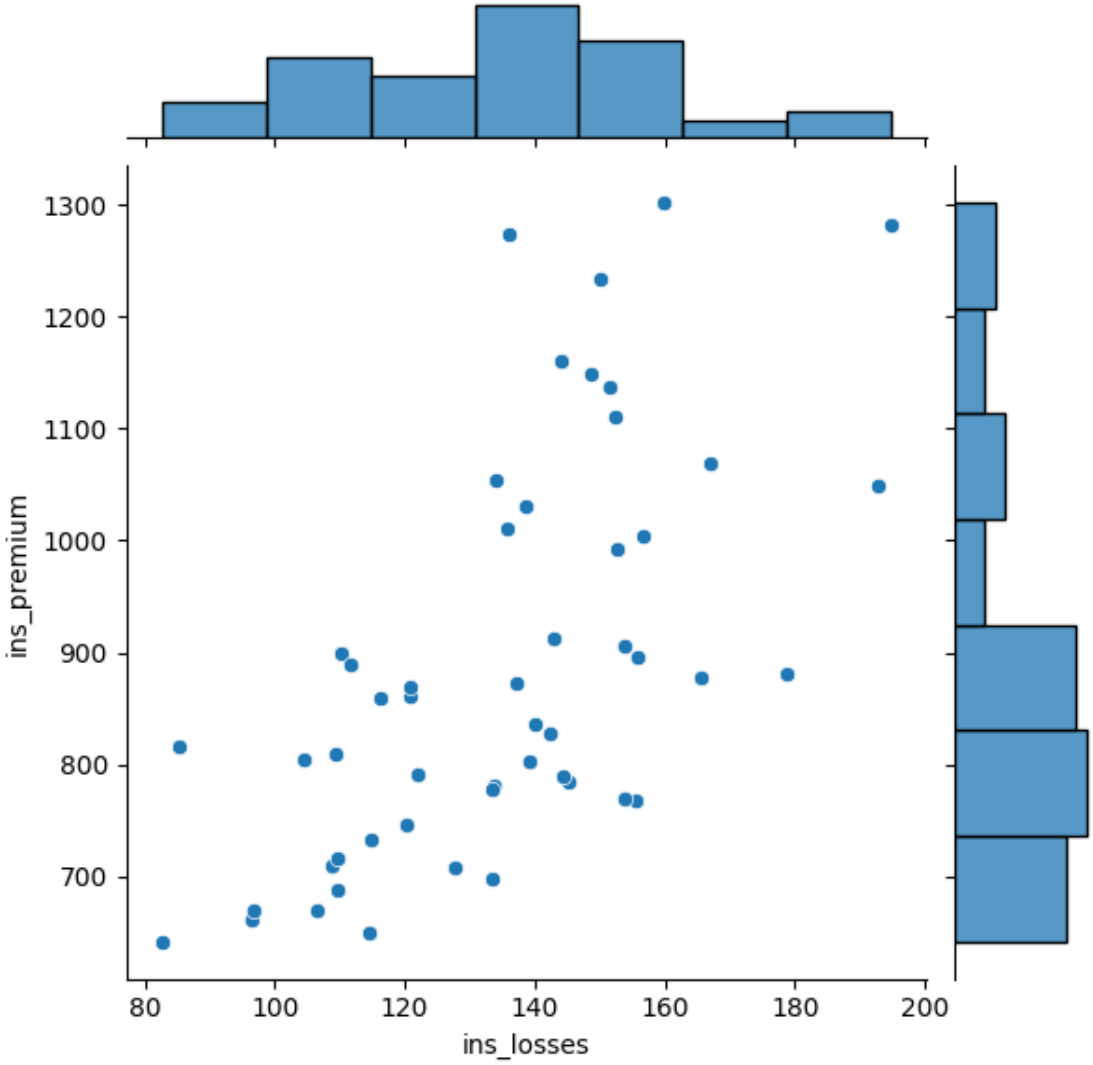
[29]:
```
<seaborn.axisgrid.JointGrid at 0x7a037eed1ae0>
```



The scatter plot shows a positive linear relationship between 'total' and 'no_previous'. The histograms represent the distributions of these two variables.

[30]:
```
sns.jointplot(x="ins_losses",y="ins_premium",data=data)
```

[30]:
```
<seaborn.axisgrid.JointGrid at 0x7a037f45d750>
```



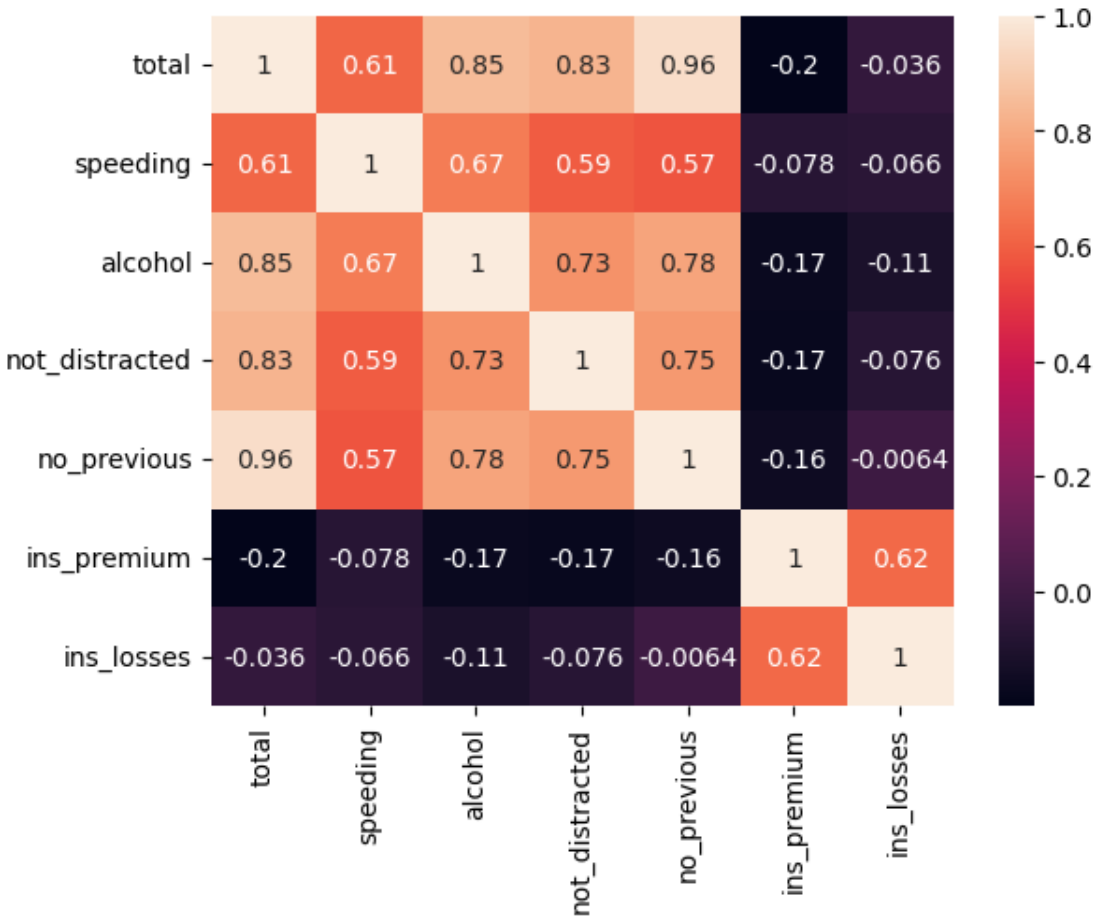
The scatter plot shows a scattered distribution of 'ins_losses' and 'ins_premium', indicating no clear correlation. The histograms represent the frequency distribution of these variables.

[31]:
```
corr = data.corr()
```

```
<ipython-input-31-17182710d970>:1: 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.
  corr = data.corr()
```

[32]:
```
sns.heatmap(corr,annot=True)
```

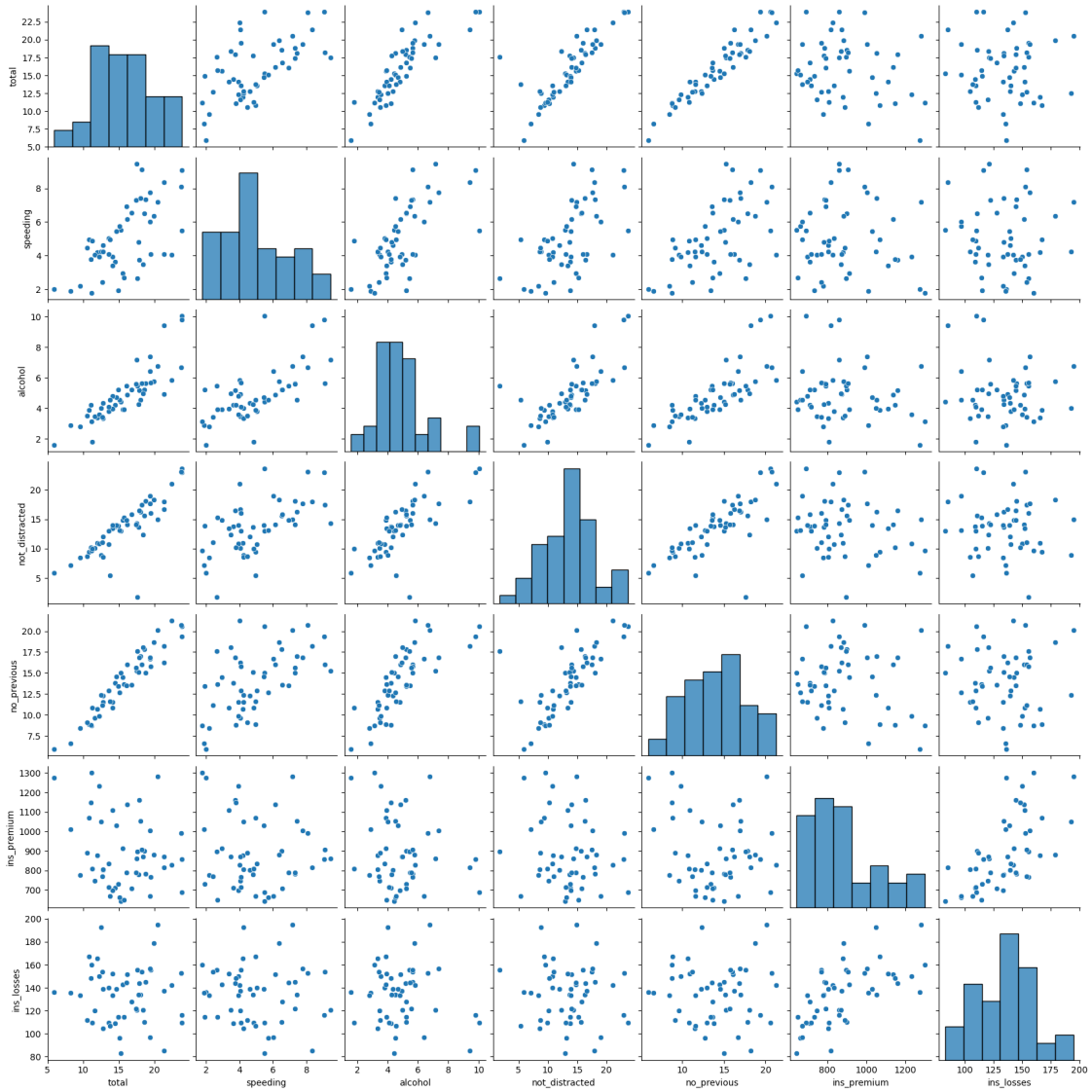
[32]: <Axes: >



The heatmap shows the correlation matrix of seven variables: total, speeding, alcohol, not_distracted, no_previous, ins_premium, and ins_losses. Dark blue indicates a strong positive correlation, while dark red indicates a strong negative correlation. The diagonal line represents a perfect positive correlation of a variable with itself

[33]:
```
sns.pairplot(data)
```

[33]:
```
<seaborn.axisgrid.PairGrid at 0x7a037eaf5150>
```



The grid of scatter plots and histograms shows the relationships between different variables. The scatter plots indicate linear relationships, while the histograms show the distributions of individual variables.