Bivariate analysis

Pearson correlation and scatter plot

In [26]:
```
import pandas as pd
import numpy as np
```

In [27]:
```
data = pd.read_csv('star_dataset.csv')
data.head()
```

Out[27]:

	Temperature (K)	Luminosity(L/Lo)	Radius(R/Ro)	Absolute magnitude(Mv)	Star type	Star color	Spectral Class
0	3068	0.002400	0.1700	16.12	Red Dwarf	Red	M
1	3042	0.000500	0.1542	16.60	Red Dwarf	Red	M
2	2600	0.000300	0.1020	18.70	Red Dwarf	Red	M
3	2800	0.000200	0.1600	16.65	Red Dwarf	Red	M
4	1939	0.000138	0.1030	20.06	Red Dwarf	Red	M

In [28]:
```
dataFrame = pd.DataFrame({'Temperature (K)':data['Temperature (K)'], 'Luminosity(L/Lo)':
correlations = dataFrame.corr()
correlations.style.background_gradient(cmap='coolwarm', axis=None).format(precision=2)
```

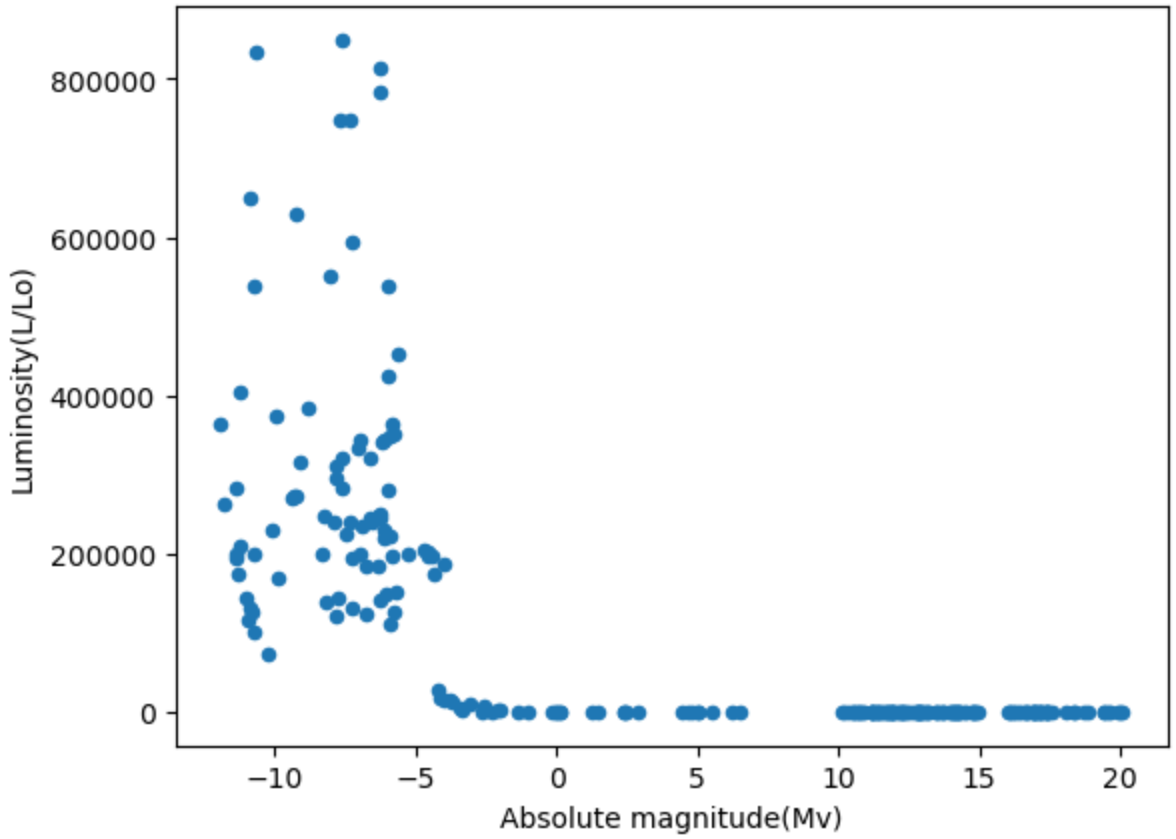
Out[28]:

	Temperature (K)	Luminosity(L/Lo)	Radius(R/Ro)	Absolute magnitude(Mv)
Temperature (K)	1.00	0.39	0.06	-0.42
Luminosity(L/Lo)	0.39	1.00	0.53	-0.69
Radius(R/Ro)	0.06	0.53	1.00	-0.61
Absolute magnitude(Mv)	-0.42	-0.69	-0.61	1.00

De hoogste correlatie die we hier zien is de absolute magnitude met de luminosity.

In [29]:
```
data.plot(kind='scatter', x='Absolute magnitude(Mv)', y='Luminosity(L/Lo)')
```

Out[29]:
```
<AxesSubplot:xlabel='Absolute magnitude(Mv)', ylabel='Luminosity(L/Lo)'>
```

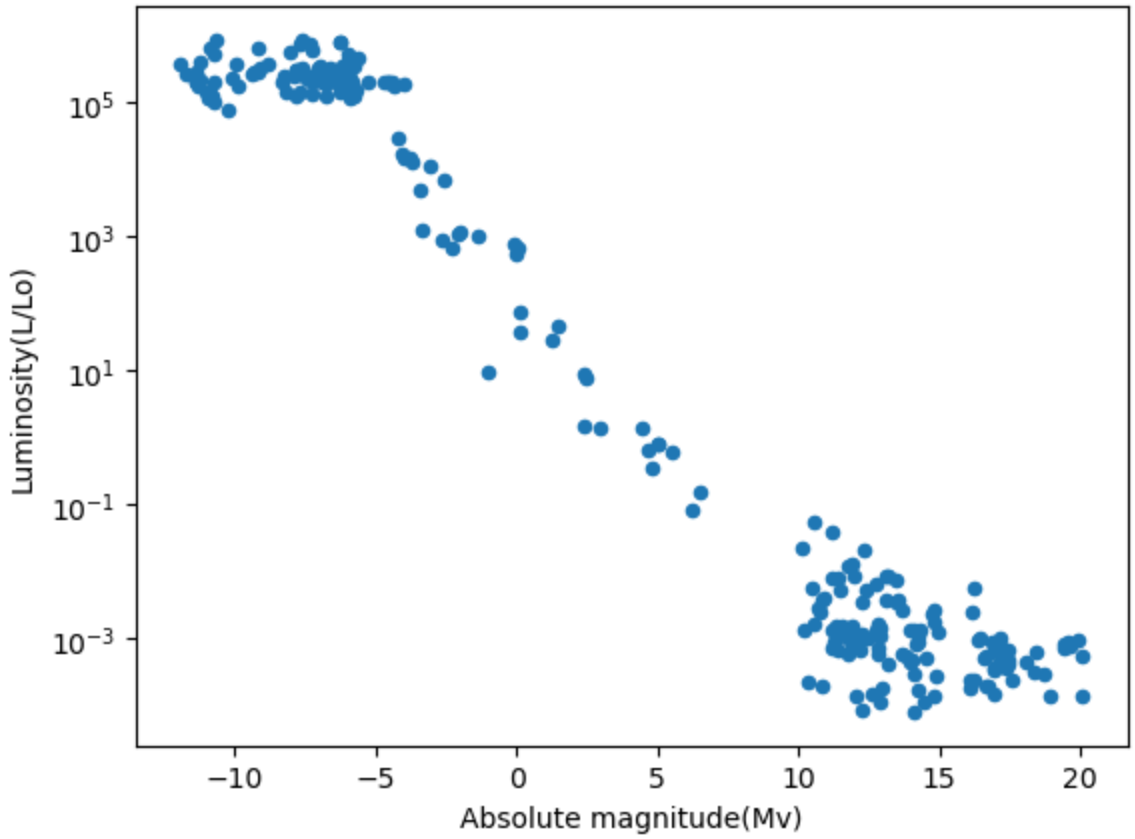


Deze correlatie lijkt op een logaritmische schaal, en dat is logisch, want de formule voor Absolute Magnitude is:

$$M_v = 4.74 - 2.5 * \log_{10} \frac{L}{Lo}$$

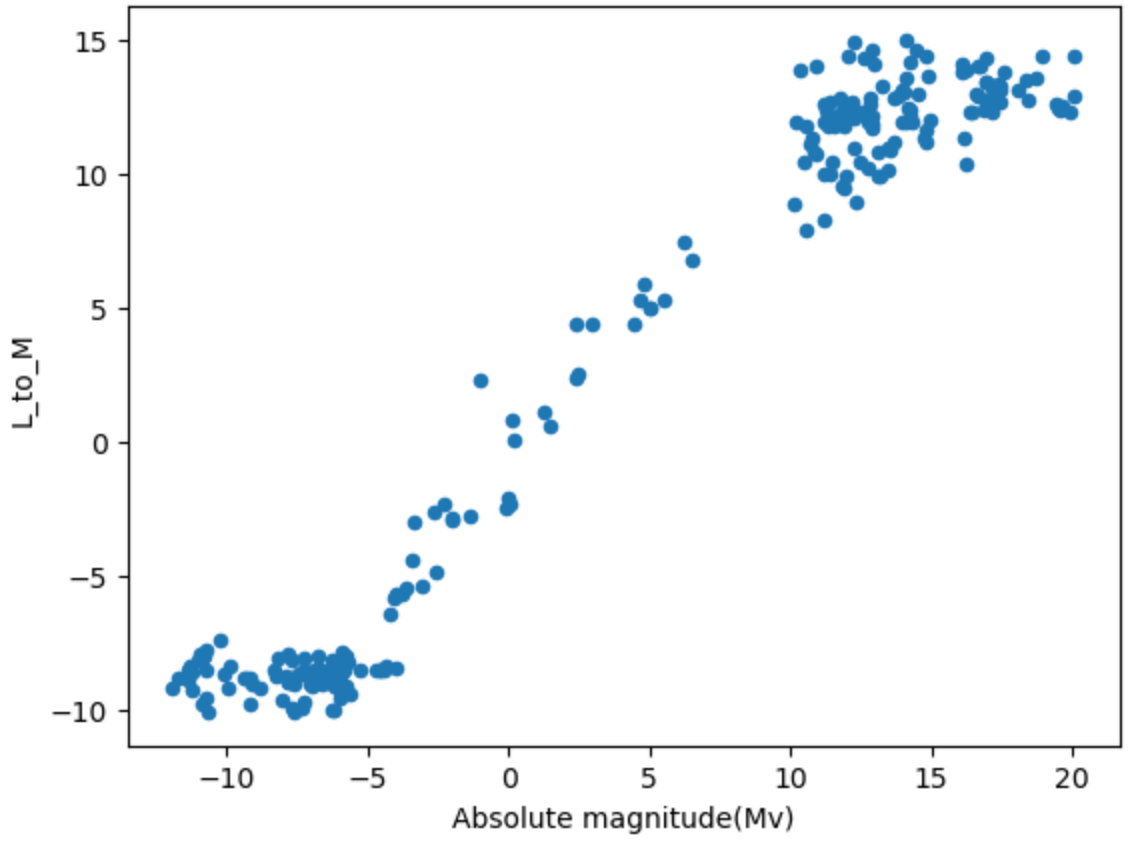
In [30]:
```
data.plot(kind='scatter', x='Absolute magnitude(Mv)', y='Luminosity(L/Lo)', logy=True)
```

Out[30]:
```
<AxesSubplot:xlabel='Absolute magnitude(Mv)', ylabel='Luminosity(L/Lo)'>
```



In [31]:
```
data['L_to_M'] = 4.74 - 2.5*np.log10(data['Luminosity(L/Lo)'])
data.plot(kind='scatter', x='Absolute magnitude(Mv)', y='L_to_M')
```

Out[31]:
```
<AxesSubplot:xlabel='Absolute magnitude(Mv)', ylabel='L_to_M'>
```



In [32]:
```
dataFrame_L_to_M = pd.DataFrame({'Temperature (K)':data['Temperature (K)'], 'Luminosity(
correlations_L_to_M = dataFrame_L_to_M.corr()
correlations_L_to_M.style.background_gradient(cmap='coolwarm', axis=None).format(precisi
```

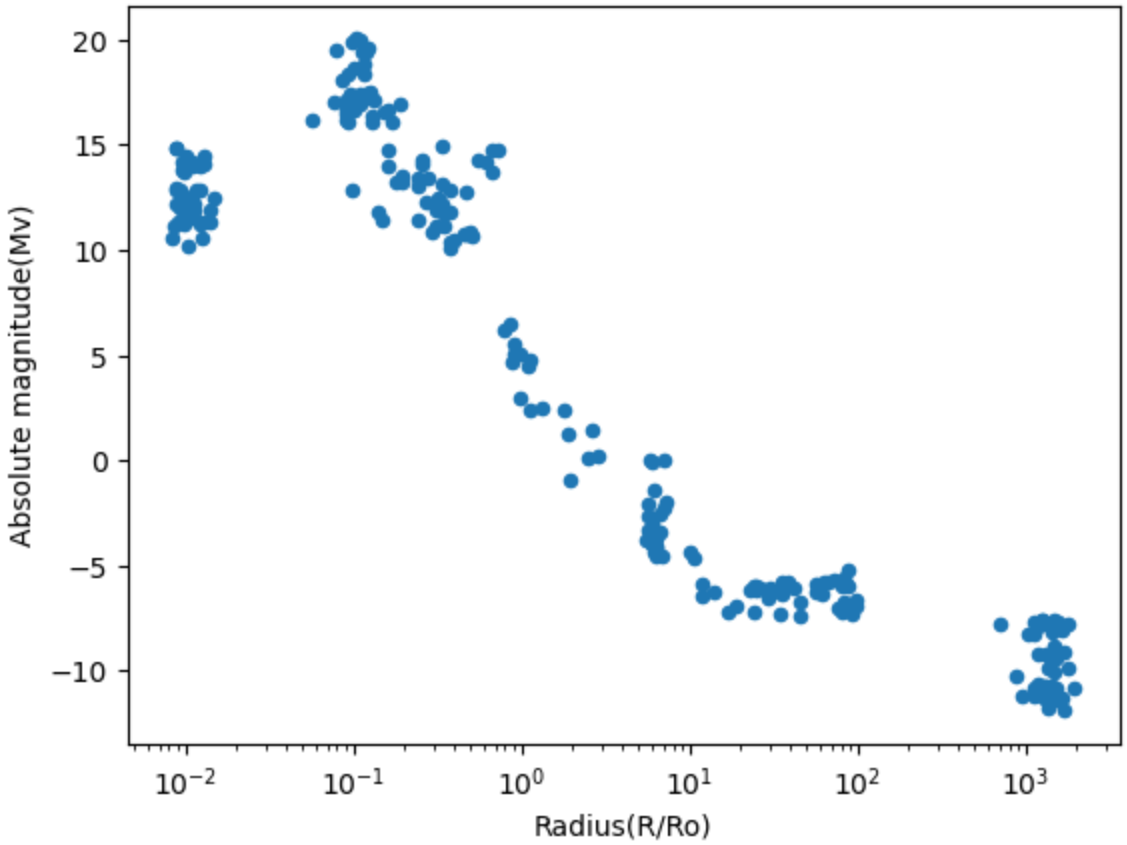
Out[32]:

	Temperature (K)	Luminosity(L/Lo)	Radius(R/Ro)	Absolute magnitude(Mv)
Temperature (K)	1.00	-0.44	0.06	-0.42
Luminosity(L/Lo)	-0.44	1.00	-0.55	0.98
Radius(R/Ro)	0.06	-0.55	1.00	-0.61
Absolute magnitude(Mv)	-0.42	0.98	-0.61	1.00

De radius heeft ook een correlatie met beide absolute magnitude en de luminosity, wat ook logisch is, want bij grotere sterren verwacht je dat de totale hoeveelheid energie die per seconde wordt uitgestraald ook groter is.

In [33]:
```
data.plot(kind='scatter', x='Radius(R/Ro)', y='Absolute magnitude(Mv)', logx=True)
```

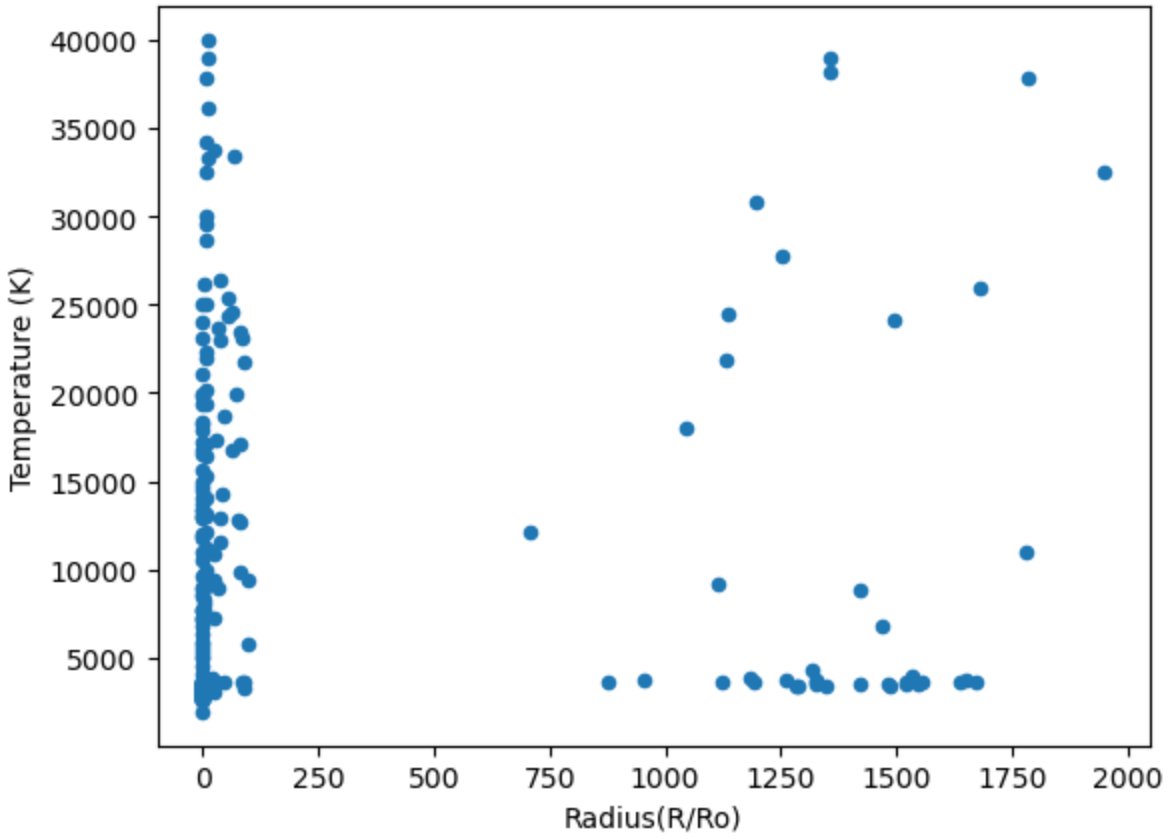
Out[33]:
```
<AxesSubplot:xlabel='Radius(R/Ro)', ylabel='Absolute magnitude(Mv)'>
```



De kleinste correlatie (bijna 0) is tussen Radius en Temperature. We verwachten hier ook geen correlatie omdat sterren van verschillende types (Main Sequence, Giants) dezelfde temperatuur kunnen hebben.

In [34]:
```
data.plot(kind='scatter', x='Radius(R/Ro)', y='Temperature (K)')
```

Out[34]:
```
<AxesSubplot:xlabel='Radius(R/Ro)', ylabel='Temperature (K)'>
```


40000
35000
30000
25000
20000
15000
10000
5000
Temperature (K)
0
250
500
750
1000
1250
1500
1750
2000
Radius(R/Ro)