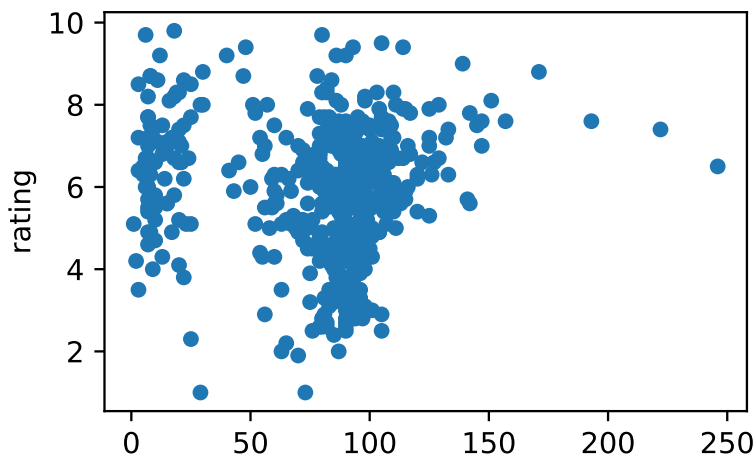*Practical 3*

Jumping Rivers

Graphics

We will continue to investigate the movies data from earlier. To begin
we will load the data and then take a random sample of 500 values to
help keep the plots a bit cleaner.

```
import jrpyintroduction.datasets as dat
movies = dat.movies.load_data()
movies = movies.sample(500)
```

Also load all of the packages that we might need for this practical

```
import numpy as np
import matplotlib.pyplot as plt
```

1. Start with a simple scatter plot of movie ratings against lengths. It
   should look something like the one below.

2. Use the `xlim` and `ylim` arguments to change the axis ranges to
   (0,200) and (0,10) respectively.

3. Change the colours of points such that we have one colour for
   Comedy films and another colour for non Comedy films.

4. Relabel your axes with the units given on the x axis

5. Finally give your graph a title.

6. Create a boxplot of movie lengths using the `.boxplot()` method.

7. You can amend the axes after the plot with `plt.ylim()` and the corresponding `.xlim()`.

```
p7 = movies.boxplot('length')
plt.ylim(
  # calculate the axes limits from the data
  movies.length.min(),
  movies.length.max()
)
plt.show(p7)
```

8. Add a new column to your movies `DataFrame` which corresponds to the decade the film was released. Hint: You can do this by taking the year value, dividing by 10, rounding down using `np.floor()` and then multiplyng by 10 again.

9. Use the `by` argument of the `.boxplot()` method to create a separate boxplot for each decade.

If you finish

If you finish, feel free to explore some of the other graphics that we have discussed.