

Eric Layne
CSE40637 Data Science
Prof. Jiang
16 September 2018

Homework #1: Programming Assignment

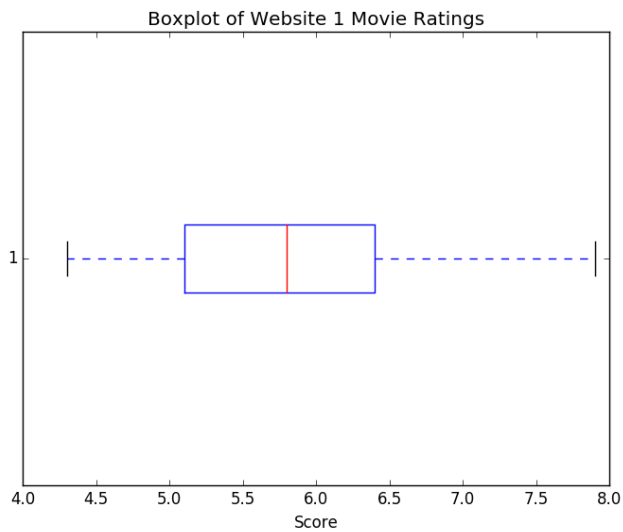
Data Description

1:

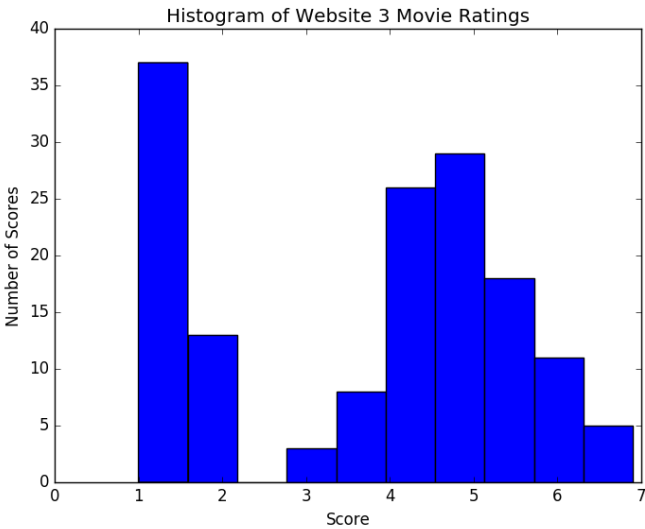
zscore1		max: 2.49201920212	min: -1.87002413385
zscore2		max: 3.0907752483	min: -2.43394714191
zscore3		max: 1.78583195363	min: -1.56757623428
zscore4		max: 1.71309868604	min: -1.44954504204

Data Visualization

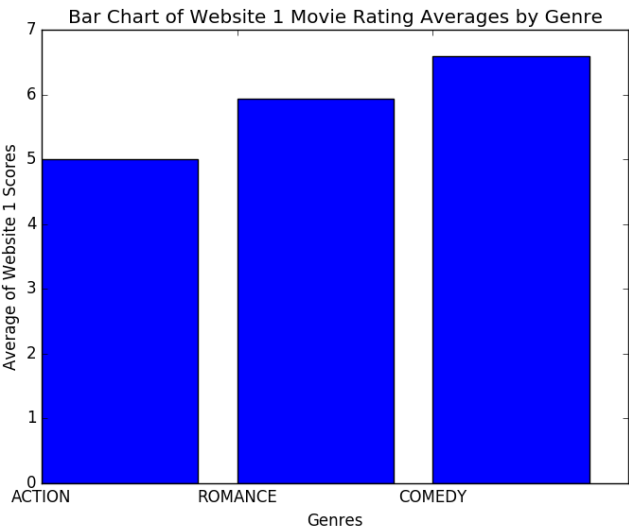
2-1:



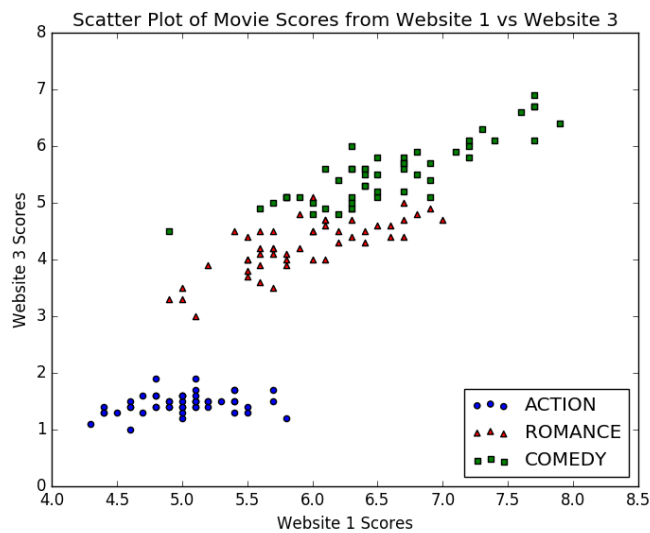
2-2:



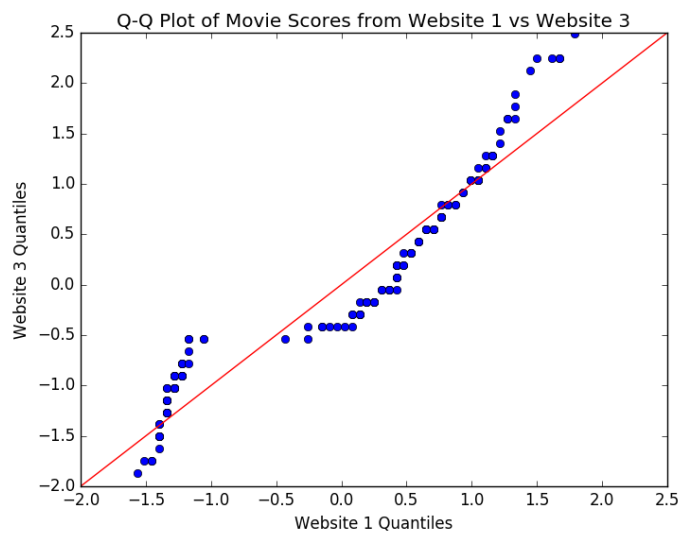
2-3:



2-4:



2-5:



2-6:

KL Divergence of 1 compared to 3: 0.09013557003510739
KL Divergence of 3 compared to 1: 0.0747357693699291

3-1:

	score1	score2	score3	score4
score1	0.685694	-0.042434	1.274315	0.515705
score2	-0.042434	0.189979	-0.329656	-0.121275
score3	1.274315	-0.329656	3.116278	1.294027
score4	0.515705	-0.121275	1.294027	0.579732

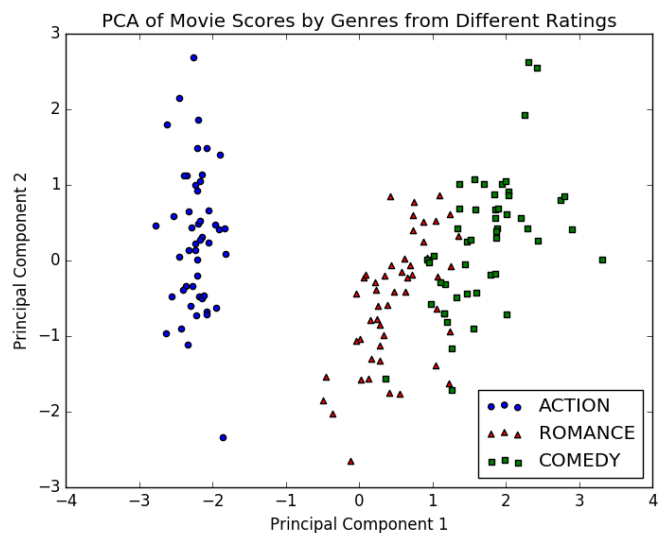
3-2:

	score1	score2	score3	score4
score1	1.006711	-0.118359	0.877604	0.823432
score2	-0.118359	1.006711	-0.431316	-0.367883
score3	0.877604	-0.431316	1.006711	0.969207
score4	0.823432	-0.367883	0.969207	1.006711

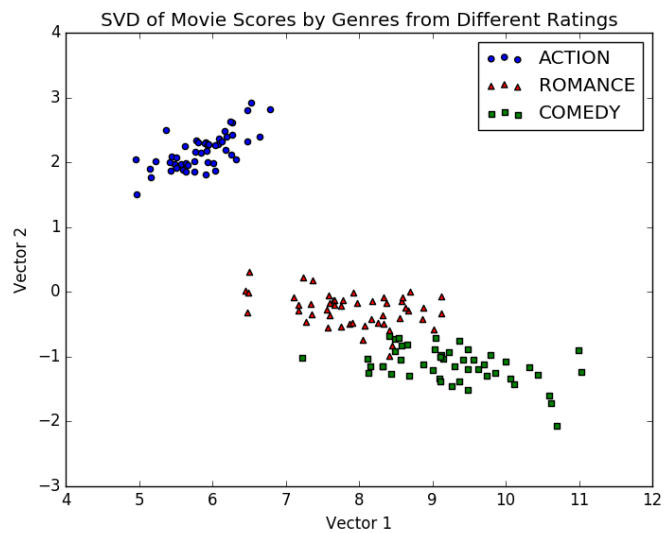
3-3:

The results are not the same due to the z-score normalization. Converting the values to z-scores removes the variations in scoring between the different metrics. Analyzing the differences between normalized z-scores tells much more than the difference between arbitrary scoring where values might range depending on the metric used.

4-1:



4-2:



4-3:

PCA Eigenvalues:

```
[[ 0.52113894 -0.26919025  0.58044885  0.56482736]
 [ 0.37703631  0.92342771  0.02416808  0.06738537]]
```

Top 3:

```
0.92342771
0.58044885
0.56482736
```

SVD Singular Values:

```
[95.96094362 17.75656516]
```

4-4:

My code did not work. My estimate is that it is neither.