Discussion:

1. I create a word-character matrix and rank characters one to another. There are 935 characters provided by the csv file and I write a rank\_character function to rank which character is most like the target character.

For example,

**The top most similar character to "KING CLAUDIUS" using compute\_cosine\_similarity are:**

**1: DUKE VINCENTIO**

**The top most similar character to "KING CLAUDIUS" using compute\_jaccard\_similarity are:**

**1: ANTONIO**

**The top most similar character to "KING CLAUDIUS" using compute\_dice\_similarity are:**

**1: ANTONIO**

From the results, they are all men and have a high social status. So what they said in the play are similar to each other.

**The top most unsimilar character to "Groom" using compute\_cosine\_similarity are:**

**935: Thieves**

**The top most unsimilar character to "Groom" using compute\_jaccard\_similarity are:**

**935: ESCANES**

**The top most unsimilar character to "Groom" using compute\_dice\_similarity are:**

**935: ESCANES**

From this result we can see that a groom has lower social status compare to the Escanes. (Escanes is a lord of Tyre, whom Helicanus trusts a great deal. When both Pericles and Helicanus are absent from Tyre, he governs the state in their place.(<http://www.playshakespeare.com/license)>) and also very different from thieves.

1. I use k-mean to cluster the play’s name. I get three categories listed below. I mark each play different which represent their original category(orange means Histories, green means tragedies and blue means comedies). And I find something interesting from this table:
2. Historical subjects and tragedies are easily confused，because historical subjects usually “sad”…
3. Tragedy subjects performed not so good in this cluster experiment, maybe because most of them have been taken away by historical dramas.
4. Comedy is sometimes confused with tragedies. I guess this has something to do with the author's writing habits or something like “The core of comedy is tragedy”……

|  |  |  |
| --- | --- | --- |
| **Cluster One**  (Histories) | **Cluster Two**  (Tragedies) | **Cluster Three**  (Comedies) |
| Henry IV  Antony and Cleopatra  Coriolanus  Hamlet  Richard II  Henry VI Part 2  Henry VIII  Richard III  Henry V  Troilus and Cressida  Henry VI Part 3  Cymbeline  King Lear | A Midsummer nights dream  King John  macbeth  Timon of Athens  The Tempest  Julius Caesar  A Comedy of Errors  Two Gentlemen of Verona  Henry VI Part 1  Pericles  Titus Andronicus | Alls well that ends well  Loves Labours Lost  Taming of the Shrew  Merry Wives of Windsor  Romeo and Juliet  A Winters Tale  As you like it  Measure for measure  Much Ado about nothing  Twelfth Night  Merchant of Venice  Othello |