Matthew Dulchinos
12-15-2018
Dr. Dan Venning

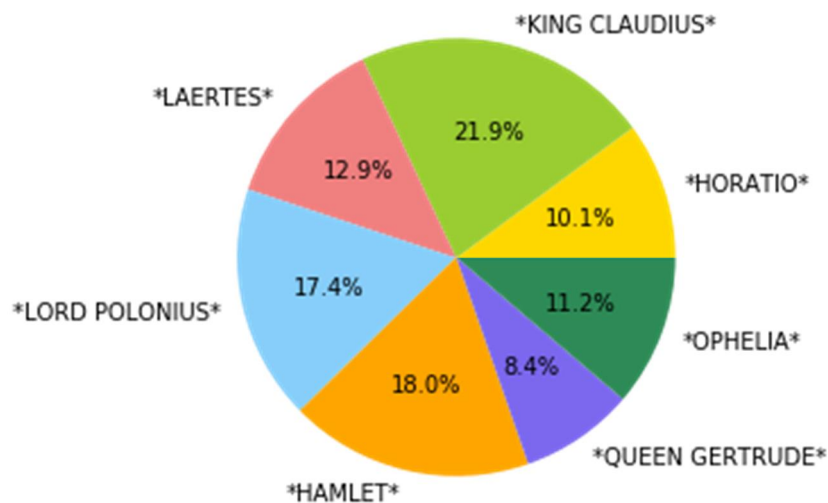<div align="center">Shakespeare's Brevity</div>

Introduction:

"Brevity is the soul of wit," uttered Lord Polonius "and tediousness the limbs that outwards flourishes" in his famous monologue pertaining to Hamlet's apparent insanity. It's one of my favorite Shakespeare monologues of all time. I knew about the saying "brevity is the soul of wit" prior to reading Hamlet, but I never realized how ironically funny the saying was until getting to read it with my own eyes; which got me thinking about whether or not this saying was actually true in Shakespeare's writing. Does the amount of words a character speaks per line of dialogue have anything to do with a character's intelligence? My hypothesis was that characters such as Polonius, or Rosencrantz and Guildenstern would all have a very high average of words per line, where a line is defined as all of the words spoken until someone else starts speaking . Judging from my results, this sadly does not appear to be the case.

I started my project by writing a program that would take a full Shakespeare play, with credits to shakespeare.mit.edu for the plays, and parse the text file to figure out exactly how many words were spoken by each character and how many times each character spoke. This was not a straightforward task unfortunately. Stage directions and Scene/Act changes both had to be coded around, as to not include them as words spoken by characters. In addition, I had to worry about times when two characters would speak at once, together. Finally, I also wanted to make my code robust enough where I could give it any script, so I wasn't able to tell the computer what characters would be in the show beforehand, and instead it had to learn as it went. Eventually though, I finished the basics of my program and instantly found some results which interested me greatly. I continued adding onto the program so that it would be able to create some graphs to aid the data, and then I decided to write this report.

Hamlet:

First off, my theory surrounding Lord Polonius turned out to be a complete flop. Despite his long monologues, his average was only 31 words per line. Compare this to Hamlet, who averaged just one word higher, at 32 words per line. Of the main characters, (who I arbitrarily decided were people with 50+ lines), King Claudius actually had the highest average at 39 words per line. What was quite intriguing to me, was that the Player King actually had the highest average of any character in the show, at a whopping 84 words per line. Of course, this character had only 4 lines in the entire show, which is why I decided to make an arbitrary cut of 50 lines required to be a main character. Still, it was interesting to see the King with the largest average line count of the main cast, and his impersonator have the largest of the whole cast. Below we can see a pie chart of each character, and we can see just how much talking the average character says in a line to one another.

And while this chart is interesting, it fails to really prove anything towards my hypothesis. If I had known what the results would have looked like coming into the project, I would have tried to record how long each of a character's lines were, and then plotted them on a bell curve to see where the mean,
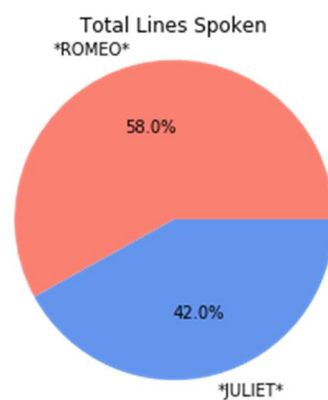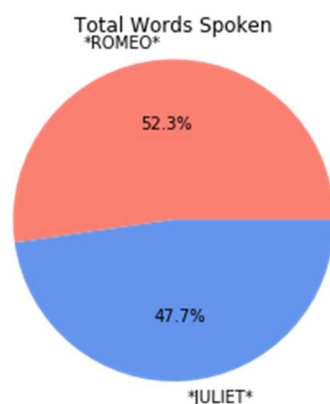
median, and mode lay. Unfortunately, due to the way I wrote my program, implementing this feature would have required an almost complete rewrite of the program. Perhaps next time I am allowed to do a project like this, I will go back and revisit the code!

I decided to look at other plays as well, and see if I could find any other interesting data, but before I move on, a couple of other fun facts:
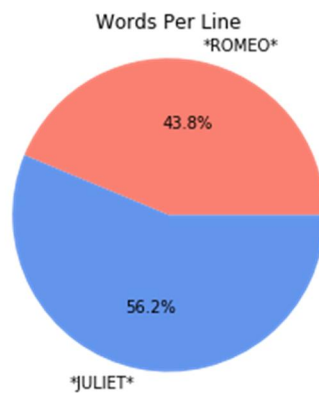
- Of any show I looked at, Hamlet has the highest line count of any character, with a total of 11,678 words spoken

- Despite how similar Rosencrantz and Guildenstern are portrayed, Rosencrantz has speaks more than double the amount of words (704 – 335) and has an average of 4 more words per line (14 – 10)

- Danes had the lowest average, with only 4 words spoken per line.

Romeo and Juliet:

The other play that I was really interested in was Romeo and Juliet, and more specifically, how the two main characters compared. After all, it's a love story between the two, so they should be about equal, right? Surprising that was not the case. Romeo beats out Juliet in both words spoken and lines spoken.

I found it especially surprising with how many more lines Romeo has than Juliet, from my recollection of the play, he doesn't speak all that much more than Juliet, but according to my data, he actually holds an eight point lead over Juliet in lines spoken. However, when it comes to lines spoken, Juliet actually wins out by quite a fair margin.

Words Per Line

ROMEO

43.8%

56.2%

JULIET

When thinking about it, I figure that this probably has to do with Romeo's fight scenes he gets himself into, while Juliet often just chats up the nurse for long periods of time. Fight scenes often involve quick quips back and forth, while Juliet instead discusses love for long monologues. This would explain the line difference as well, although the fact that Romeo speaks more than 400 more words than Juliet is not something I have a good explanation for (4669 – 4263).

Conclusion:

Overall though, I was a little disappointed with my findings. I will admit, given how oddly specific Shakespeare was with his writing, I almost expected to find some major discoveries through running these tests, but unfortunately, nothing that was too exciting. I feel like this data could, however, be potentially useful for a future paper, perhaps written by someone else. Even though I didn't discover any hidden messages in the information and my hypothesis was proven false, I could see the info being useful if I were to ever write a paper regarding sexism in Shakespeare's plays or something similar.

Appendix:

Figure 1: Copy of the script, minus MatPlotLib library. Should work in both Python 2.7 and 3.

```
#Shakespeare analyzer
#By Matthew Dulchinos 12-15-2018
#Credits to Shakespeare.mit.edu for providing the scripts used in the program

#Global Variables
playFile = 'HamletEntire.txt'
play = open(playFile,'r')
lineByLine = play.readlines()
playLists = []
name = ""
recordingName=False
speaking=False
statsIndex=[]
peopleTalking = []

def StartOfSceneOrAct(line):
    if len(line) >= 6:
        for letter in line[0:5]:
            if letter != ' ':
                return False
        #print("New Act:",line)
        return True

def StageDirection(line):
    if len(line) >= 5:
        if line[4] == '/':
            #print("Stage Direction:",line)
            return True
        else:
            return False
"""
returns true if the name already exists in the statsIndex
"""
def hasActor(name):
    if len(statsIndex) != 0:
        for actor in statsIndex:
            if actor[0] == name:
                return True
    return False

def createActor(name):
    return [name,0,0]
    #[name,words,lines]

def getPos(name):
    pos = 0
    for i in statsIndex:
        if i[0] == name:
            return pos
        pos+=1
    return -1
```

```
#Adds a line to the character, using number rep of character
def addLinePos(pos):
    statsIndex[pos][2]+=1

def addWordPos(pos):
    statsIndex[pos][1]+=1

#print (lineByLine)
for line in lineByLine:
    if line[0]=='*':
        #print("Starting a new person")

        peopleTalking=[]
        for letter in line:
            if recordingName:
                name += letter
                if letter == '*':
                    recordingName=False
                    if not hasActor(name):
                        statsIndex.append(createActor(name))
                    nPos = getPos(name)
                    peopleTalking.append(nPos)
                    addLinePos(nPos)
                    name = ""
                    #print(name)
                    #name = ""
                    #activate recording software
            else:
                if letter == '*':
                    name += letter
                    recordingName=True
    elif (not StartOfSceneOrAct(line)) and (not StageDirection(line)):
        words=line.split()
        for word in words:
            for x in peopleTalking:
                addWordPos(x)
            #print(word)
        #print("!!!!!!!!!!!!!!!!!!!!!")
        #print(line)


print (statsIndex)
for actor in statsIndex:
    print(str(actor[0])+" Words:"+str(actor[1])+" Lines:"+str(actor[2])+" Avg:"+str(actor[1]/actor[2]))
```

Figure 2: Hamlet Stats

```
*BERNARDO* Words:220 Lines:23 Avg:9
*FRANCISCO* Words:54 Lines:8 Avg:6
*HORATIO* Words:2049 Lines:112 Avg:18
*MARCELLUS* Words:440 Lines:36 Avg:12
*KING CLAUDIUS* Words:4053 Lines:102 Avg:39
```

```
*CORNELIUS* Words:10 Lines:1 Avg:10
*VOLTIMAND* Words:159 Lines:2 Avg:79
*LAERTES* Words:1442 Lines:62 Avg:23
*LORD POLONIUS* Words:2668 Lines:86 Avg:31
*HAMLET* Words:11678 Lines:359 Avg:32
*QUEEN GERTRUDE* Words:1050 Lines:69 Avg:15
*All* Words:11 Lines:4 Avg:2
*OPHELIA* Words:1178 Lines:58 Avg:20
*Ghost* Words:677 Lines:14 Avg:48
*REYNALDO* Words:65 Lines:13 Avg:5
*ROSENCRANTZ* Words:704 Lines:48 Avg:14
*GUILDENSTERN* Words:335 Lines:32 Avg:10
*First Player* Words:365 Lines:8 Avg:45
*Prologue* Words:16 Lines:1 Avg:16
*Player King* Words:337 Lines:4 Avg:84
*Player Queen* Words:238 Lines:5 Avg:47
*LUCIANUS* Words:40 Lines:1 Avg:40
*ROSENCRANTZ:* Words:4 Lines:1 Avg:4
*GUILDENSTERN:* Words:4 Lines:1 Avg:4
*PRINCE FORTINBRAS* Words:193 Lines:6 Avg:32
*Captain* Words:84 Lines:7 Avg:12
*Gentleman* Words:176 Lines:3 Avg:58
*Danes* Words:13 Lines:3 Avg:4
*Servant* Words:9 Lines:1 Avg:9
*First Sailor* Words:39 Lines:2 Avg:19
*Messenger* Words:36 Lines:2 Avg:18
*First Clown* Words:734 Lines:33 Avg:22
*Second Clown* Words:97 Lines:12 Avg:8
*First Priest* Words:90 Lines:2 Avg:45
*OSRIC* Words:356 Lines:25 Avg:14
*Lord* Words:67 Lines:3 Avg:22
*First Ambassador* Words:40 Lines:1 Avg:40
```
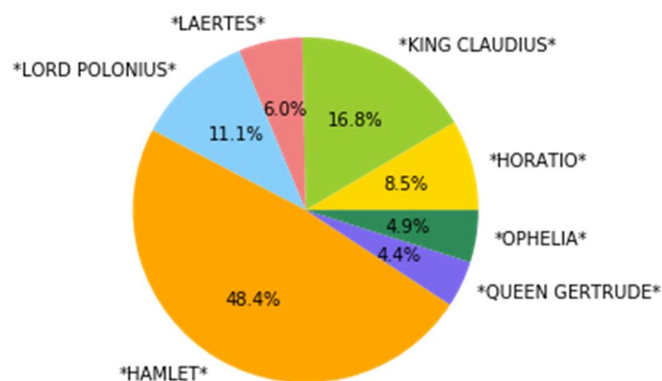
Figure 3: Percentage of words spoken by Hamlet main characters

Figure 4: R and J Stats

```
*SAMPSON* Words:259 Lines:20 Avg:12
*GREGORY* Words:149 Lines:15 Avg:9
*ABRAHAM* Words:24 Lines:5 Avg:4
*BENVOLIO* Words:1153 Lines:64 Avg:18
*TYBALT* Words:265 Lines:17 Avg:15
*First Citizen* Words:43 Lines:3 Avg:14
*CAPULET* Words:2123 Lines:50 Avg:42
*LADY CAPULET* Words:872 Lines:45 Avg:19
*MONTAGUE* Words:316 Lines:10 Avg:31
*LADY MONTAGUE* Words:28 Lines:2 Avg:14
*PRINCE* Words:585 Lines:16 Avg:36
*ROMEO* Words:4669 Lines:163 Avg:28
*PARIS* Words:539 Lines:23 Avg:23
*Servant* Words:183 Lines:10 Avg:18
*Nurse* Words:2121 Lines:89 Avg:23
*JULIET* Words:4263 Lines:118 Avg:36
*MERCUTIO* Words:2094 Lines:62 Avg:33
*First Servant* Words:79 Lines:4 Avg:19
*Second Servant* Words:99 Lines:6 Avg:16
*Second Capulet* Words:17 Lines:2 Avg:8
*Chorus* Words:113 Lines:1 Avg:113
*FRIAR LAURENCE* Words:2727 Lines:55 Avg:49
*PETER* Words:245 Lines:13 Avg:18
*NURSE* Words:87 Lines:1 Avg:87
*First Musician* Words:62 Lines:9 Avg:6
*Second Musician* Words:34 Lines:3 Avg:11
*Musician* Words:8 Lines:1 Avg:8
*Third Musician* Words:7 Lines:1 Avg:7
*BALTHASAR* Words:232 Lines:12 Avg:19
*Apothecary* Words:53 Lines:4 Avg:13
*FRIAR JOHN* Words:96 Lines:4 Avg:24
*PAGE* Words:81 Lines:4 Avg:20
*First Watchman* Words:134 Lines:6 Avg:22
*Second Watchman* Words:9 Lines:1 Avg:9
*Third Watchman* Words:26 Lines:1 Avg:26
```

Figure 5: Macbeth Stats:

```
*First Witch* Words:349 Lines:23 Avg:15
*Second Witch* Words:125 Lines:15 Avg:8
*Third Witch* Words:129 Lines:13 Avg:9
*ALL* Words:131 Lines:13 Avg:10
*DUNCAN* Words:483 Lines:18 Avg:26
*MALCOLM* Words:1505 Lines:40 Avg:37
*Sergeant* Words:234 Lines:3 Avg:78
*LENNOX* Words:502 Lines:22 Avg:22
*ROSS* Words:905 Lines:39 Avg:23
*MACBETH* Words:5300 Lines:146 Avg:36
*BANQUO* Words:780 Lines:33 Avg:23
*ANGUS* Words:137 Lines:4 Avg:34
*LADY MACBETH* Words:1897 Lines:59 Avg:32
*Messenger* Words:176 Lines:6 Avg:29
*FLEANCE* Words:15 Lines:2 Avg:7
*Porter* Words:305 Lines:4 Avg:76
*MACDUFF* Words:1156 Lines:59 Avg:19
*DONALBAIN* Words:59 Lines:3 Avg:19
*Old Man* Words:81 Lines:4 Avg:20
*ATTENDANT* Words:8 Lines:1 Avg:8
*First Murderer* Words:169 Lines:19 Avg:8
*Second Murderer* Words:86 Lines:6 Avg:14
*Both Murderers* Words:8 Lines:2 Avg:4
*Servant* Words:21 Lines:5 Avg:4
*Third Murderer* Words:42 Lines:6 Avg:7
*Lords* Words:13 Lines:3 Avg:4
*HECATE* Words:265 Lines:2 Avg:132
*Lord* Words:156 Lines:3 Avg:52
*First Apparition* Words:13 Lines:1 Avg:13
*Second Apparition* Words:23 Lines:2 Avg:11
*Third Apparition* Words:32 Lines:1 Avg:32
*LADY MACDUFF* Words:291 Lines:19 Avg:15
*Son* Words:143 Lines:14 Avg:10
*Doctor* Words:335 Lines:20 Avg:16
*Gentlewoman* Words:182 Lines:11 Avg:16
*MENTEITH* Words:71 Lines:5 Avg:14
*CAITHNESS* Words:73 Lines:3 Avg:24
*SEYTON* Words:32 Lines:5 Avg:6
*SIWARD* Words:203 Lines:11 Avg:18
*Soldiers* Words:4 Lines:1 Avg:4
*YOUNG SIWARD* Words:43 Lines:4 Avg:10
```