GENDERED LANGUAGE IN ANIMATED MOVIES

Cassie Maz

University of Pittsburgh

18 April 2019

PROJECT ORIGINS

- I'm borderline obsessed with Disney
- Language Gender and Society Final Project
  - Do princesses become more authoritative over time?
  - Analyzed commands in 3 Disney Princess films
  - Inconclusive results

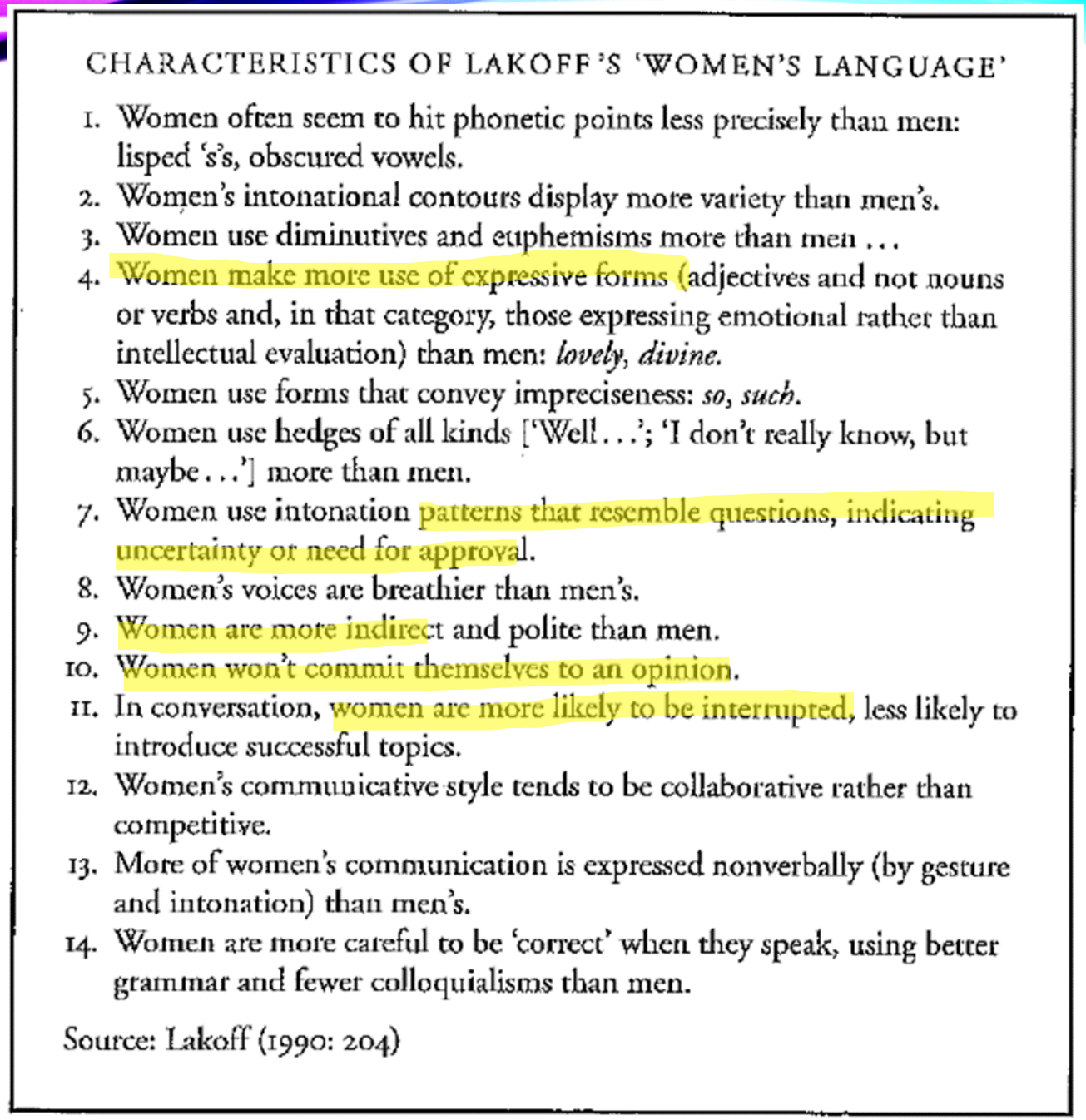
PROJECT ORIGINS (CONT.)

			Speakers										Total
Addressees			Male				Female						
			Stefan	Hubert	Phillip	O	Flora	Fauna	Merryweather	Aurora	Queen	Maleficent	
	Male	Stefan	0	3	0	0	1	0	0	0	0	0	4
		Hubert	5	0	0	0	0	0	0	0	0	0	5
		Phillip	0	10	0	0	5	2	0	0	0	1	18
		O	1	0	3	0	0	0	0	0	0	14	18
		Total Male	**6**	**13**	**3**	**0**	**6**	**2**	**0**	**0**	**0**	**15**	**45**
	NG*	**Animals/Objects**	**0**	**0**	**1**		**3**	**1**	**4**	**0**	**0**	**0**	**9**
	Female	Flora	0	0	0	0	0	2	0	0	0	0	2
		Fauna	0	0	0	0	12	0	0	0	0	0	12
		Merryweather	0	0	0	0	17	0	0	0	0	0	17
		Aurora	0	0	0	0	4	0	2	0	0	2	8
		Queen	0	0	0	0	1	0	0	0	0	0	1
		Maleficent	0	0	0	0	0	0	0	0	0	0	0
		Total Female	**0**	**0**	**0**	**0**	**34**	**2**	**2**	**0**	**0**	**2**	**40**
	Total		**6**	**13**	**4**	**0**	**43**	**5**	**6**	**0**	**0**	**17**	**94**
						23						**71**	**94**

Commands by Gender in Sleeping Beauty

PROJECT PROPOSAL

- Do female characters in animated movies adhere to what was traditionally known as "feminine" language?
- Does this change over time and across production companies?



CHARACTERISTICS OF LAKOFF'S 'WOMEN'S LANGUAGE'

1. Women often seem to hit phonetic points less precisely than men: lisped 's's, obscured vowels.
2. Women's intonational contours display more variety than men's.
3. Women use diminutives and euphemisms more than men ...
4. Women make more use of expressive forms (adjectives and not nouns or verbs and, in that category, those expressing emotional rather than intellectual evaluation) than men: *lovely, divine.*
5. Women use forms that convey impreciseness: *so, such.*
6. Women use hedges of all kinds ['Well ...'; 'I don't really know, but maybe ...'] more than men.
7. Women use intonation patterns that resemble questions, indicating uncertainty or need for approval.
8. Women's voices are breathier than men's.
9. Women are more indirect and polite than men.
10. Women won't commit themselves to an opinion.
11. In conversation, women are more likely to be interrupted, less likely to introduce successful topics.
12. Women's communicative style tends to be collaborative rather than competitive.
13. More of women's communication is expressed nonverbally (by gesture and intonation) than men's.
14. Women are more careful to be 'correct' when they speak, using better grammar and fewer colloquialisms than men.

Source: Lakoff (1990: 204)

Cameron and Kulick, 2003

HYPOTHESIS: FEMALE CHARACTERS WILL ADHERE LESS TO "WEAK" LANGUAGE OVER TIME. FEMALE ANTAGONISTS WILL ADHERE LESS TO "WEAK" LANGUAGE THAN FEMALE PROTAGONISTS DO.

BIG QUESTIONS

- Do women tend to talk more than men?
- Do women use a wider vocabulary than men?
- Do women tend to be less certain than men?
- Do women use adjectives more than men?

- Some trickier topics….
  - Are women interrupted more than men?
  - Do women use more commands than men?

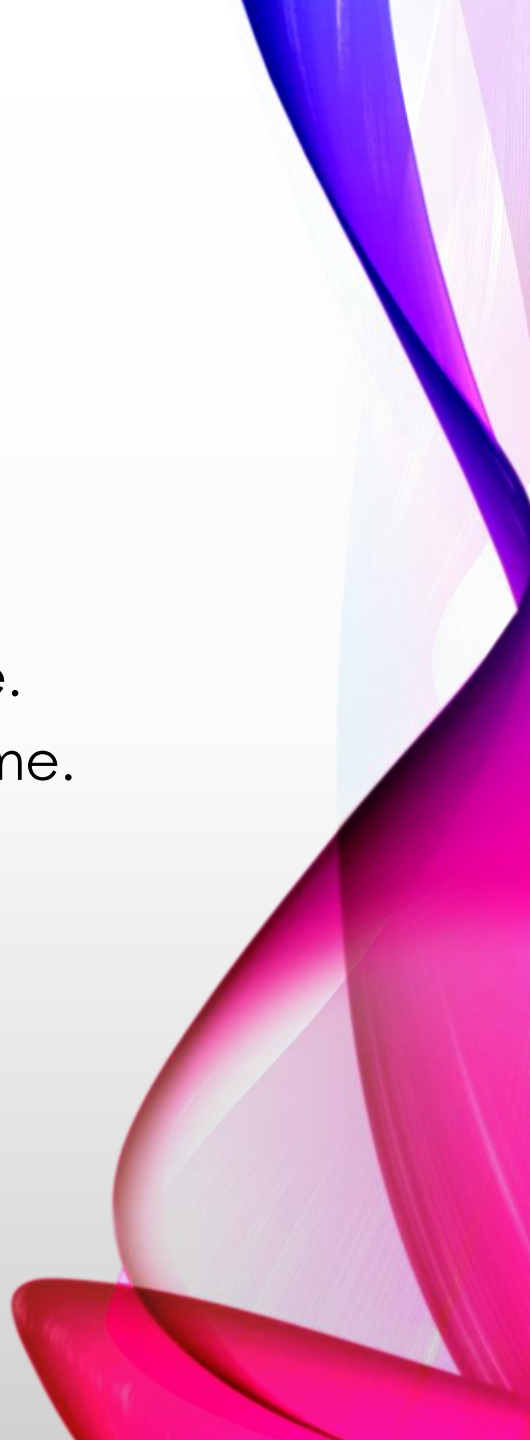
MY DATA

- CSV File of 12 Disney Princess Films
  - Created by Stanford linguist Lelia Glass
  - Line by line
  - Moana manually added
- Natural Language and Dialogue Systems' Film Corpus 2.0
  - Scripts of 1068 films from imsdb.com (text files)
  - 35 animated film scripts
  - 9 DreamWorks scripts
- Overall: 22 Movies

SOME**BODY** ONCE TOLD ME DATA CLEAN-UP WAS GONNA ROLL ME…

It was Na-Rae.

Na-Rae told me.



CLEANING THE DREAMWORKS DATA

- Formatted for humans to read
- Solution: Analyze white space
- Ran into a few “hiccups” in How to Train Your Dragon

```
shrek_script[:2000]
```

```
"                                        SHREK\n\n                                        Written by\n\n
William Steig & Ted Elliott\n\n                                        SHREK\n\n                                        Once upon a time there
was a lovely \n\n                                        princess. But she had an enchantment \n\n                                        upon her of a fe
arful sort which could \n\n                                        only be broken by love's first kiss. \n\n                                        She wa
s locked away in a castle guarded \n\n                                        by a terrible fire-breathing dragon. \n\n
Many brave knights had attempted to \n\n                                        free her from this dreadful prison, \n\n
but non prevailed. She waited in the \n\n                                        dragon's keep in the highest room of \n\n
the tallest tower for her true love \n\n                                        and true love's first kiss. (laughs) \n\n
Like that's ever gonna happen. What \n\n                                        a load of - (toilet flush)\n\n                                        Allstar - by S
mashmouth begins to play. Shrek goes about his \n\n                                        day. While in a nearby town, the villagers get together to
go \n\n                                        after the ogre.\n\n                                        NIGHT - NEAR SHREK'S HOME\n\n                                        MA
N1\n\n                                        Think it's in there?\n\n                                        MAN2\n\n
All right. Let's get it!\n\n                                        MAN1\n\n                                        Whoa. Hold on. Do you know
what that \n\n                                        thing can do to you?\n\n                                        MAN3\n\n
Yeah, it'll grind your bones for it's \n\n                                        bread.\n\n                                        Shrek sneaks up behind them and
laughs.\n\n                                        SHREK\n\n                                        Yes, well, actually, that would be a \n\n
giant. Now,"
```

```
SHREK

Written by

William Steig & Ted Elliott

SHREK

Once upon a time there was a lovely

princess. But she had an enchantment

upon her of a fearful sort which could

only be broken by love's first kiss.

She was locked away in a castle guarded

by a terrible fire-breathing dragon.

Many brave knights had attempted to

free her from this dreadful prison,

but non prevailed. She waited in the
```

EDITING THE DISNEY DATA

- Marked Songs vs Dialogue
- Marked Gender (male, female, neutral), and Role (pro, ant, or helper)
- Mostly minor typos

```
In [33]: tangled_lines['Text'].tail()

Out[33]: 188    hmmmm. let me just get this straight. i take y...
         189    i promise. and when i promise something, i nev...
         190    all right, listen, i didn't want to have to do...
         191                                          really? oops.
         192                                you broke my 'smoulder'.
         Name: Text, dtype: object
```

CLEANING THE DISNEY DATA

Tangled

In [15]:
```
for line in frozen_lines.head().itertuples():
    print(line[0])
    print(line[1])
    print(line[2])
    print('\n')
```

```
0
ice harvesters
born of cold and winter air and mountain rain combining, this icy force both foul and fair has a frozen heart worth mining.
the men drag giant ice blocks through channels of water. ice harvesters cut through the heart, cold and clear. strike for lo
ve and strike for fear. see the beauty sharp and sheer. split the ice apart! and break the frozen heart. hup! ho! watch your
step! let it go!


1
ice harvesters
, and his reindeer calf, sven, share a carrot as they try to keep up with the men.


2
ice harvesters
ice harvesters hup! ho! watch your step! let it go! kristoff struggles to get a block of ice out of the water. he fails, end
s up soaked. sven licks his wet cheek. ice harvesters beautiful! powerful! dangerous! cold! ice has a magic can't be control
led.
```

CLEANING THE DISNEY DATA

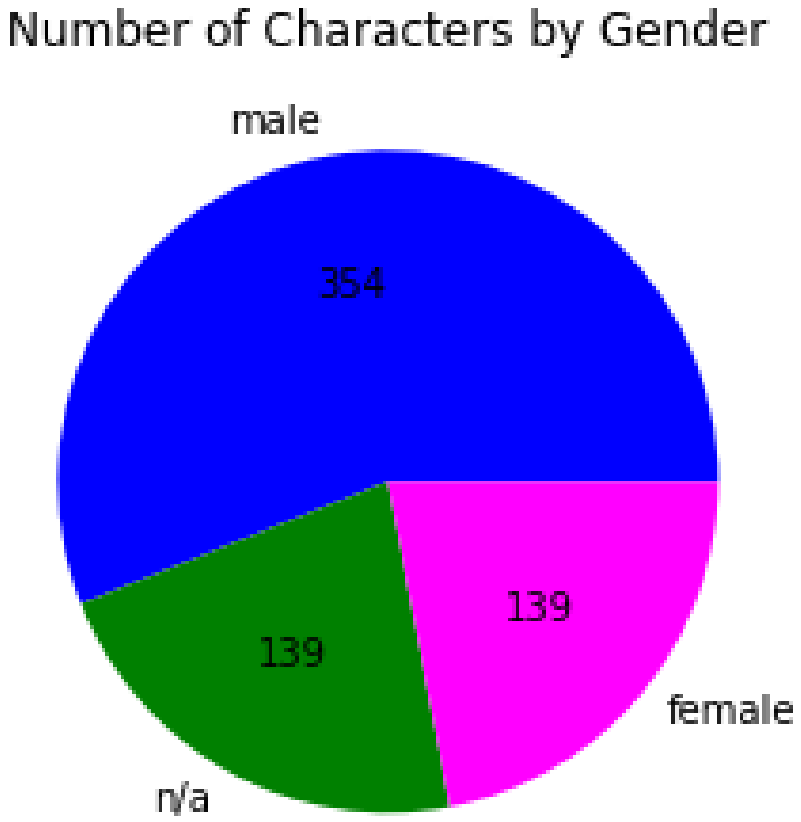
Frozen

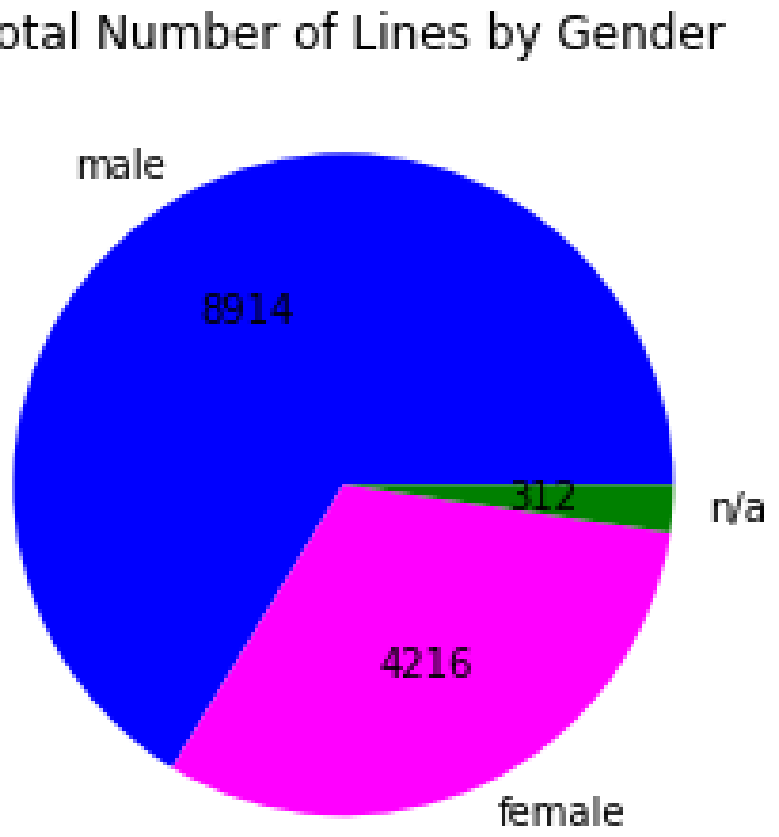
FINAL DATA FRAME

- Filtered out songs
- Dropped empty lines
  - 13442 lines
  - 632 characters
- With marked gender:
  - 13310 lines
  - 493 characters
- With marked gender & role:
  - 11320 lines
  - 216 characters

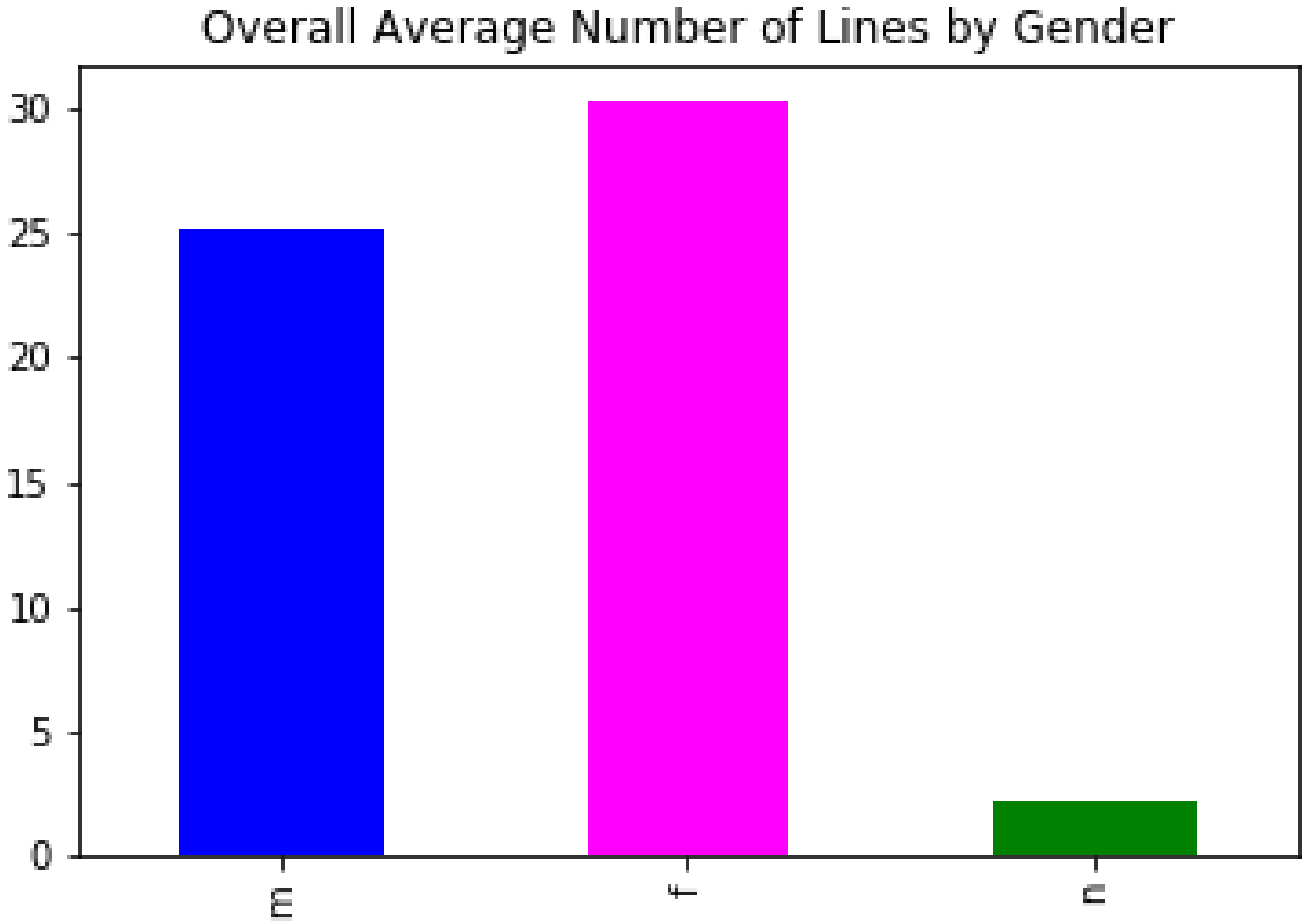
```
In [21]: movie_df_dialogue.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 13442 entries, 0 to 14095
Data columns (total 14 columns):
Disney_Period       13442 non-null object
Gender              13442 non-null object
Movie               13442 non-null object
Role                13442 non-null object
Song                13442 non-null object
Speaker             13442 non-null object
Speaker_Status      13442 non-null object
Text                13442 non-null object
UTTERANCE_NUMBER    13442 non-null int64
Year                13442 non-null int64
Tokens              13442 non-null object
Types               13442 non-null object
Token_Count         13442 non-null int64
Type_Count          13442 non-null int64
dtypes: int64(4), object(10)
memory usage: 1.0+ MB
```

DATA DISTRIBUTION: GENDER

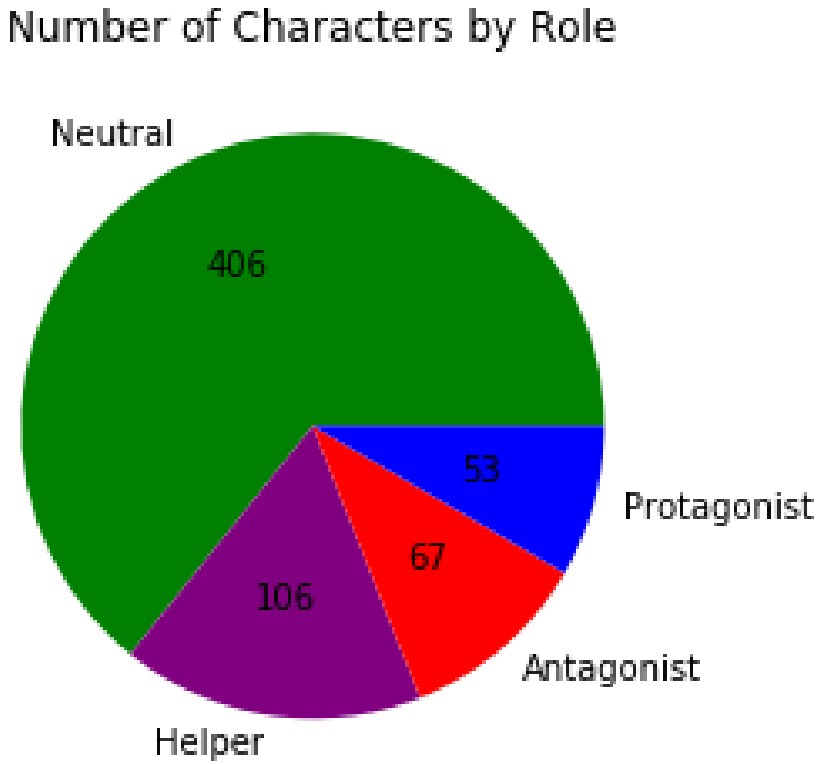


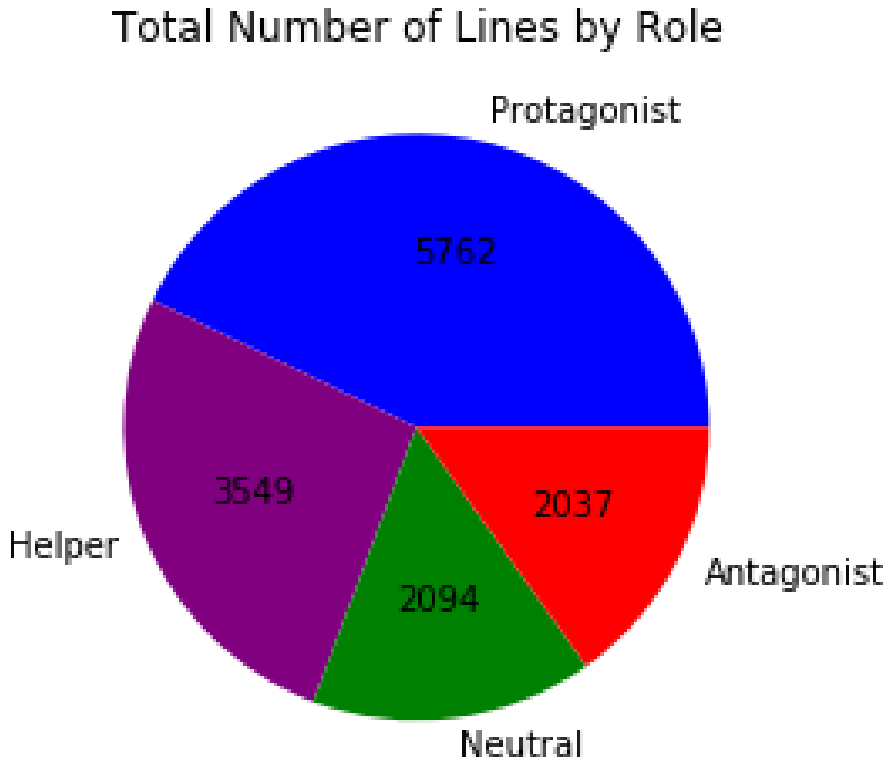


DATA DISTRIBUTION: GENDER

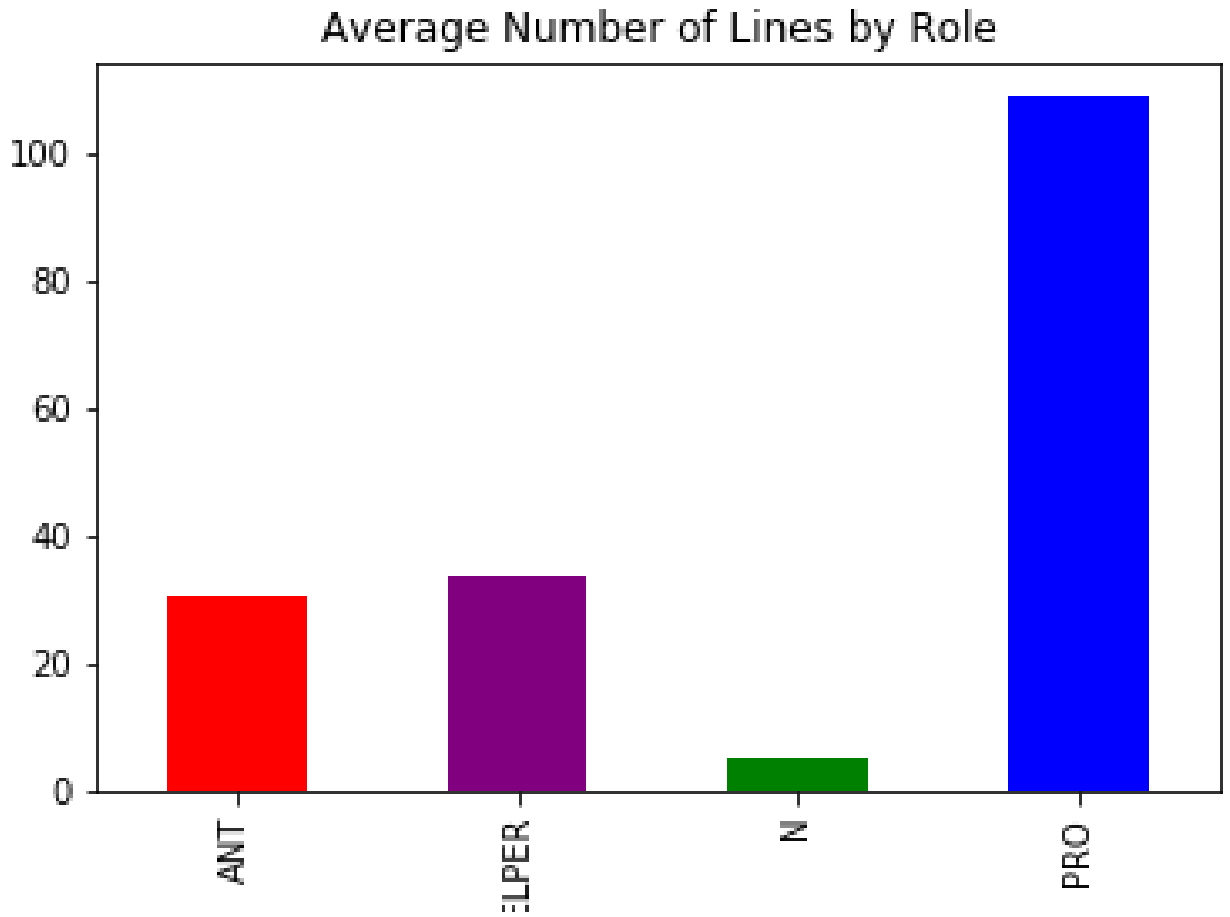


DATA DISTRIBUTION: ROLE





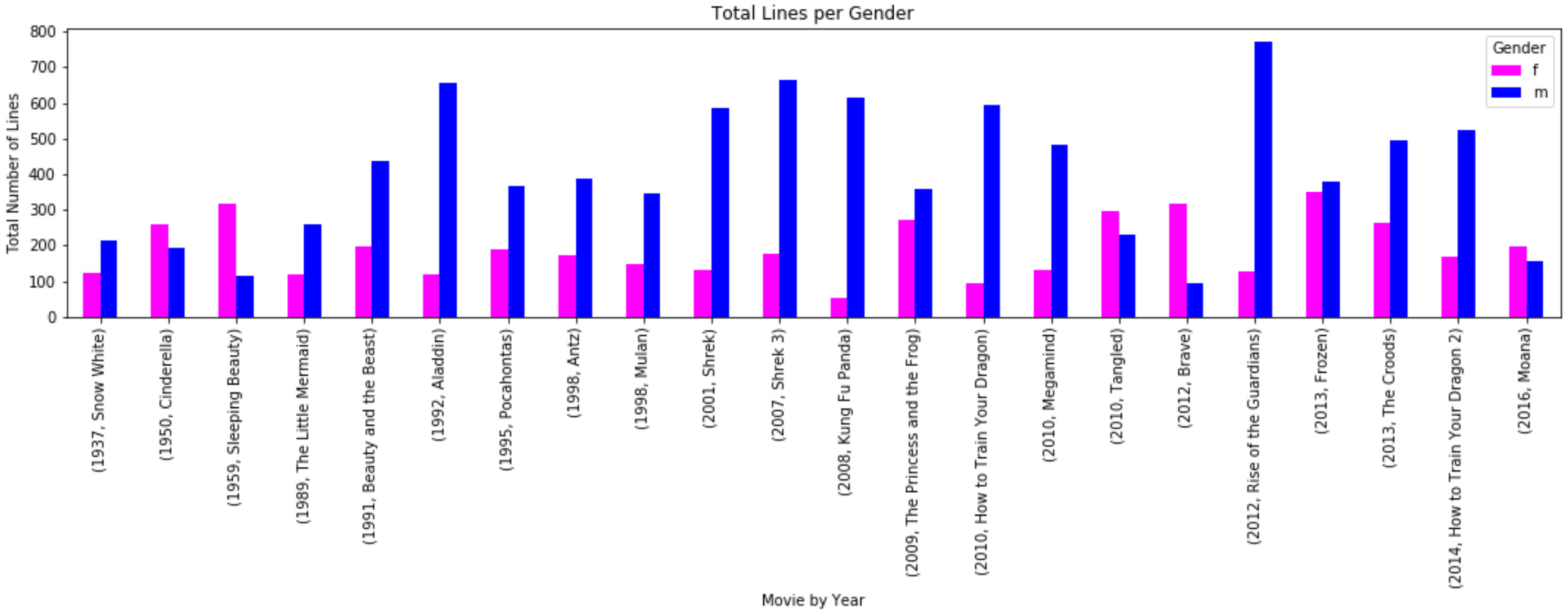
DATA DISTRIBUTION: ROLE



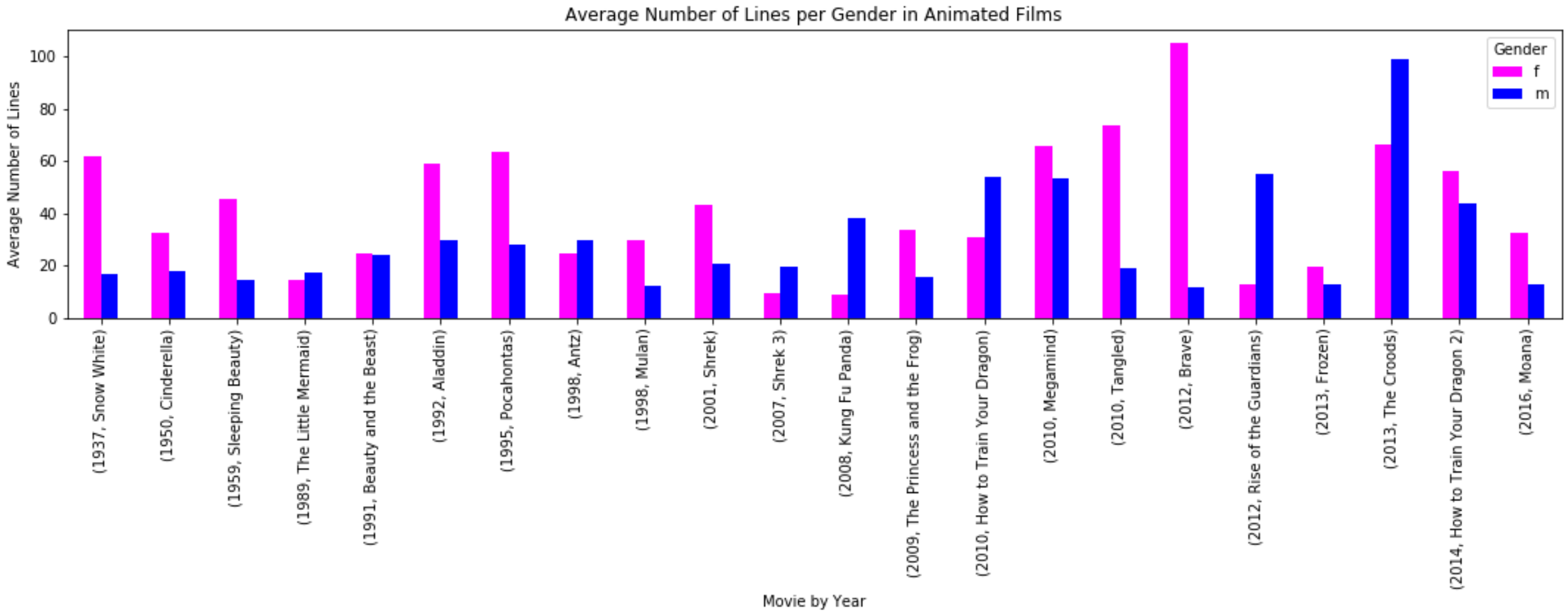
ANALYSIS:

Total Lines, Line Length, and Vocabulary

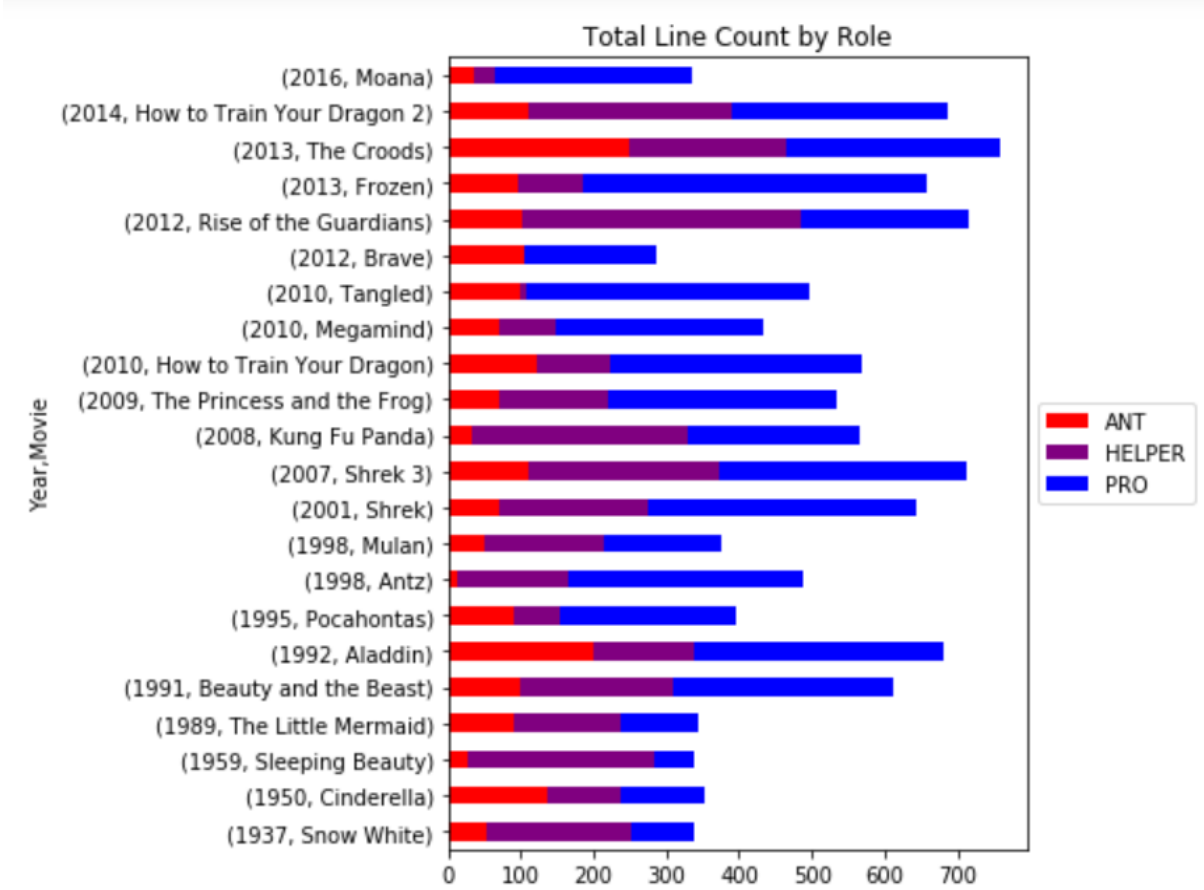
TOTAL NUMBER OF LINES: GENDER

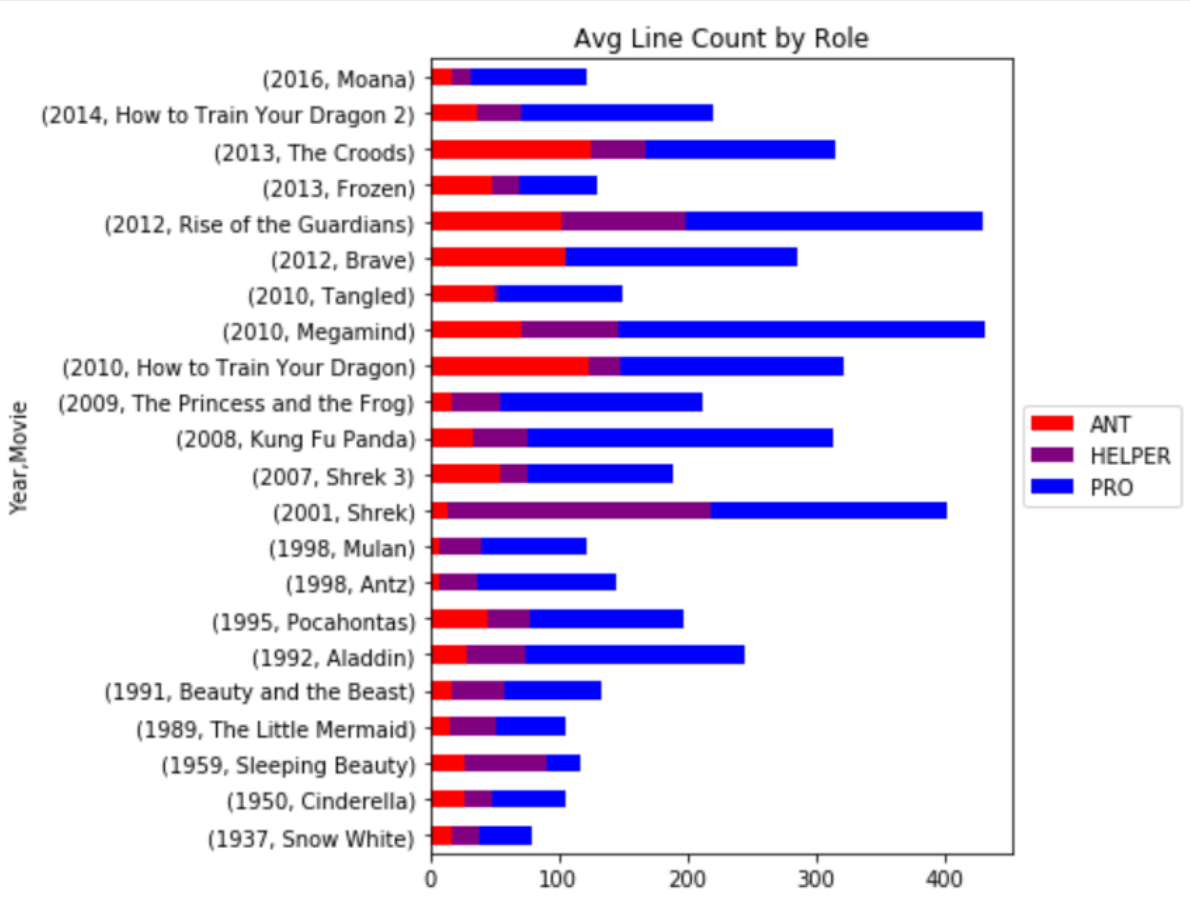


AVG NUMBER OF LINES: GENDER

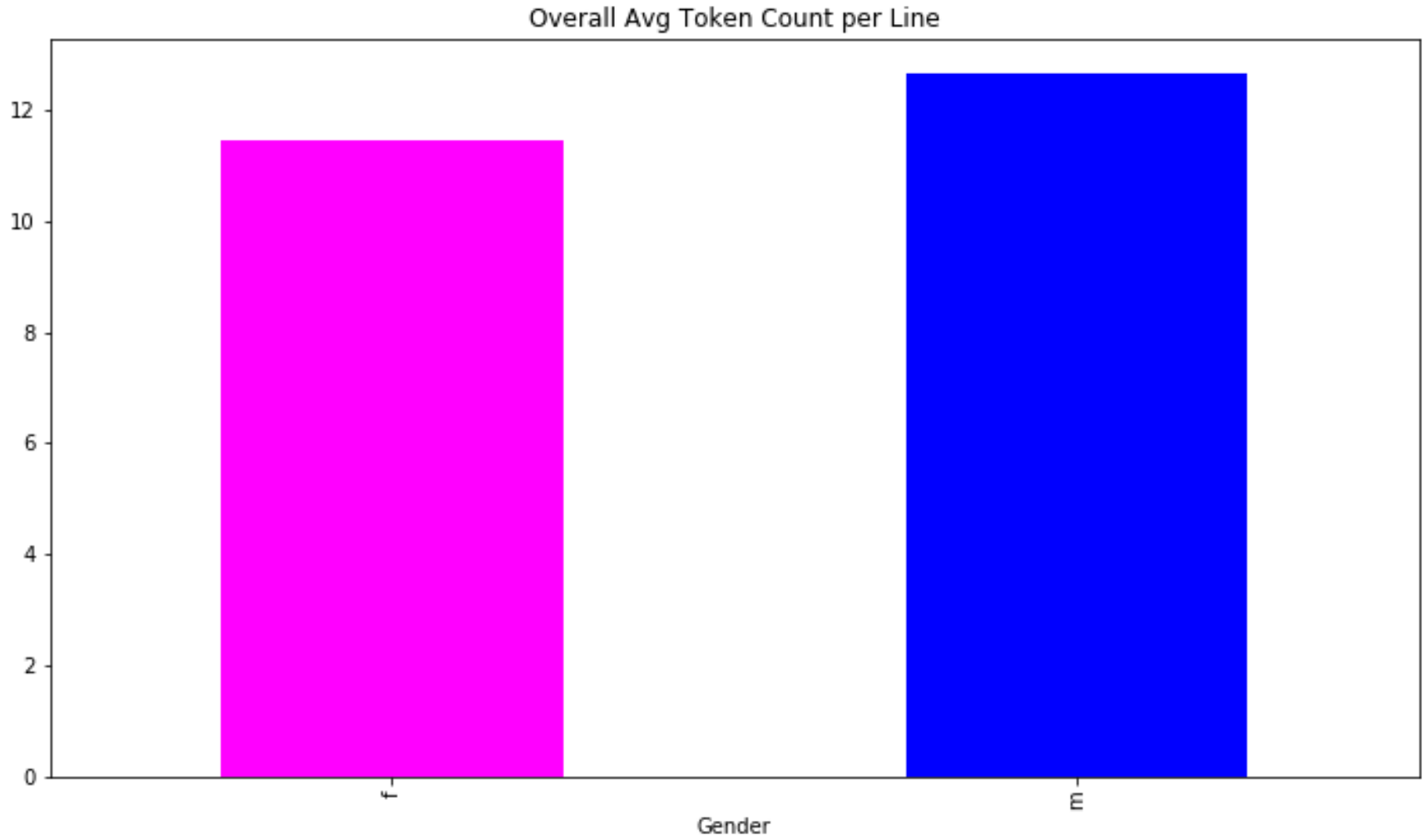


LINE COUNT: ROLE

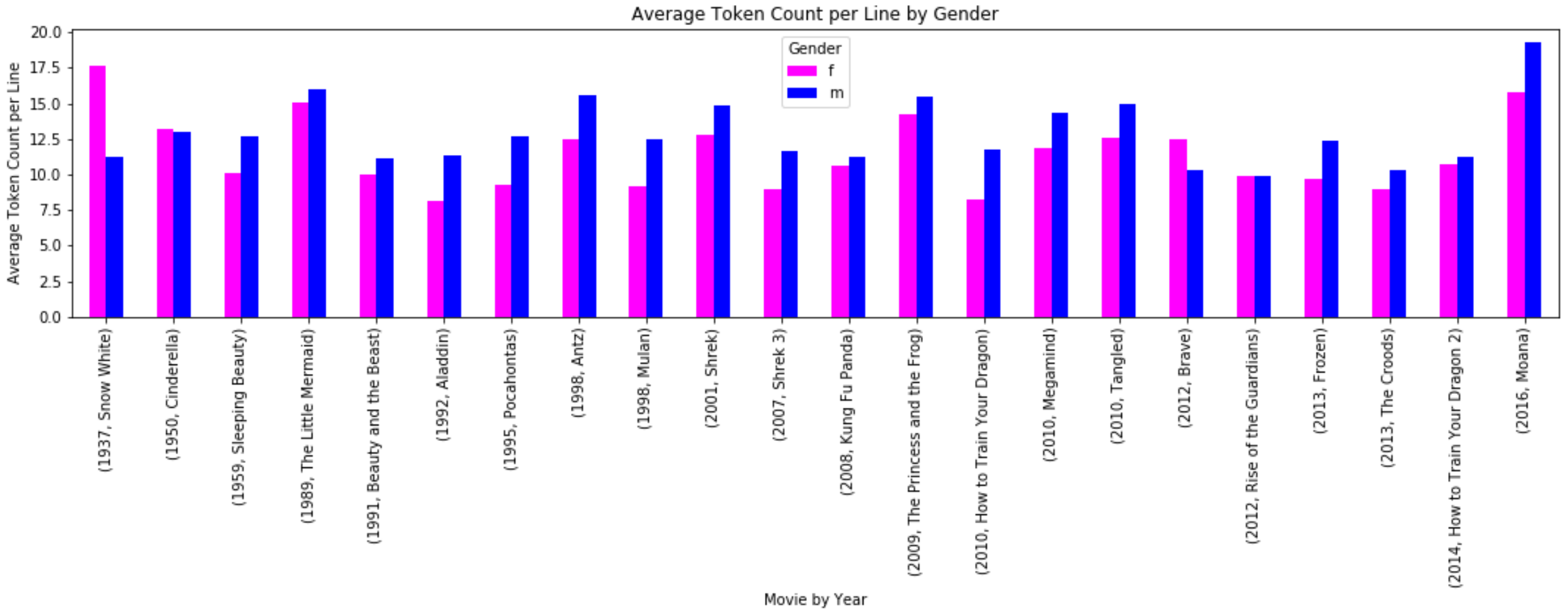




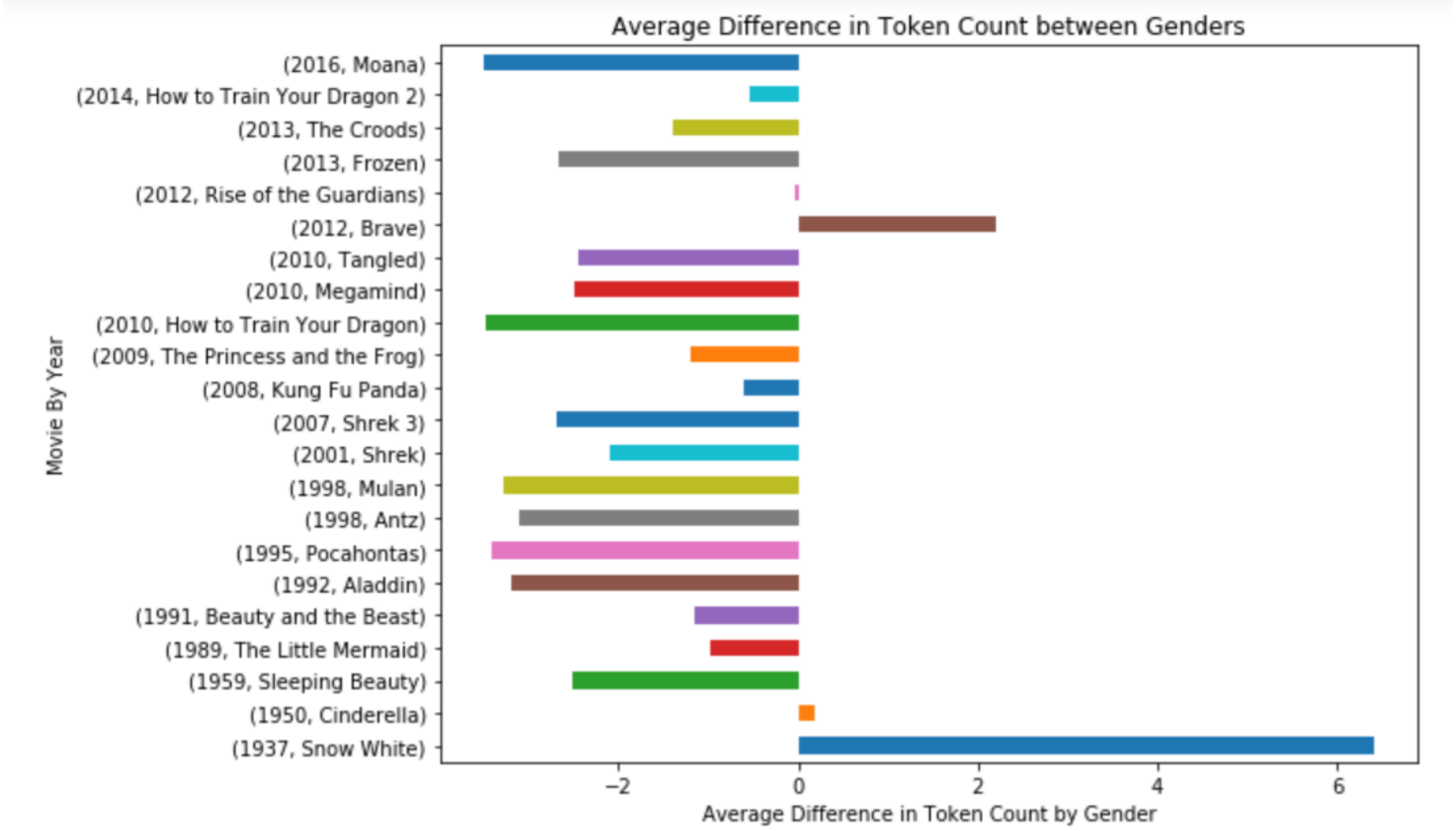
LINE LENGTH



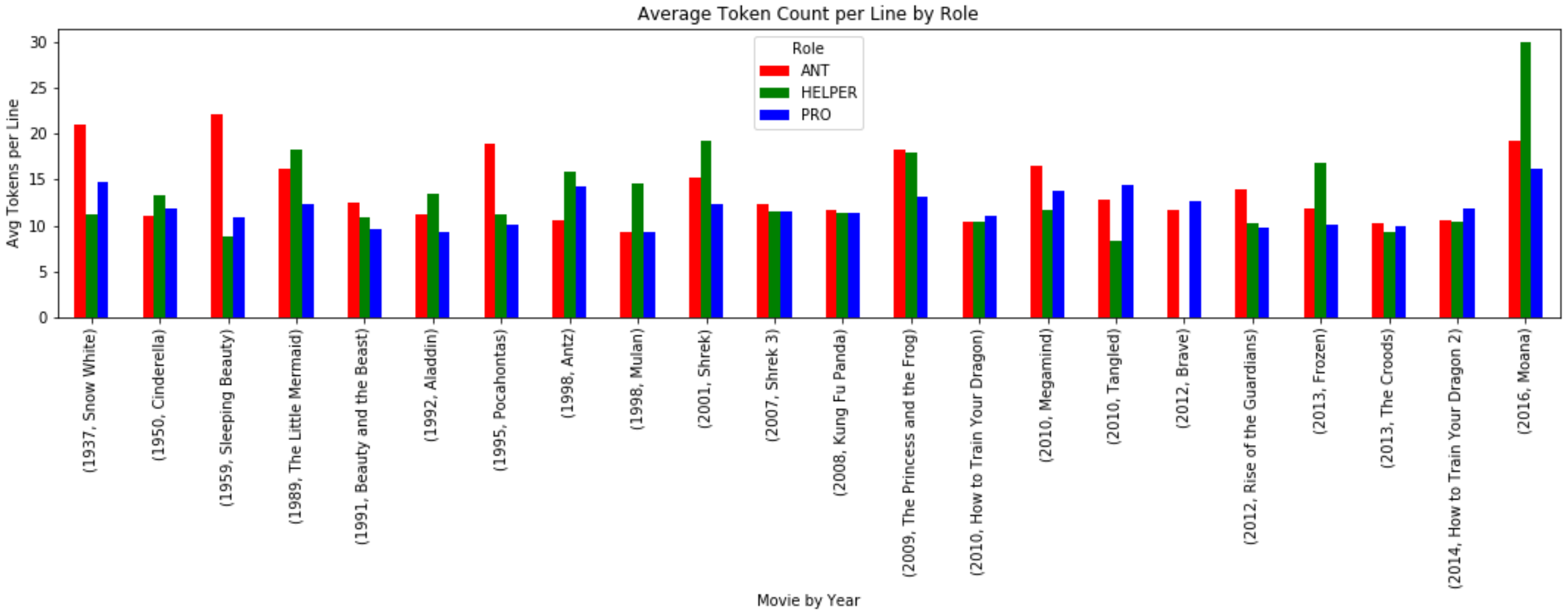
LINE LENGTH


Average Token Count per Line by Gender
Gender
f
m
Average Token Count per Line
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
(1937, Snow White)
(1950, Cinderella)
(1959, Sleeping Beauty)
(1989, The Little Mermaid)
(1991, Beauty and the Beast)
(1992, Aladdin)
(1995, Pocahontas)
(1998, Antz)
(1998, Mulan)
(2001, Shrek)
(2007, Shrek 3)
(2008, Kung Fu Panda)
(2009, The Princess and the Frog)
(2010, How to Train Your Dragon)
(2010, Megamind)
(2010, Tangled)
(2012, Brave)
(2012, Rise of the Guardians)
(2013, Frozen)
(2013, The Croods)
(2014, How to Train Your Dragon 2)
(2016, Moana)
Movie by Year

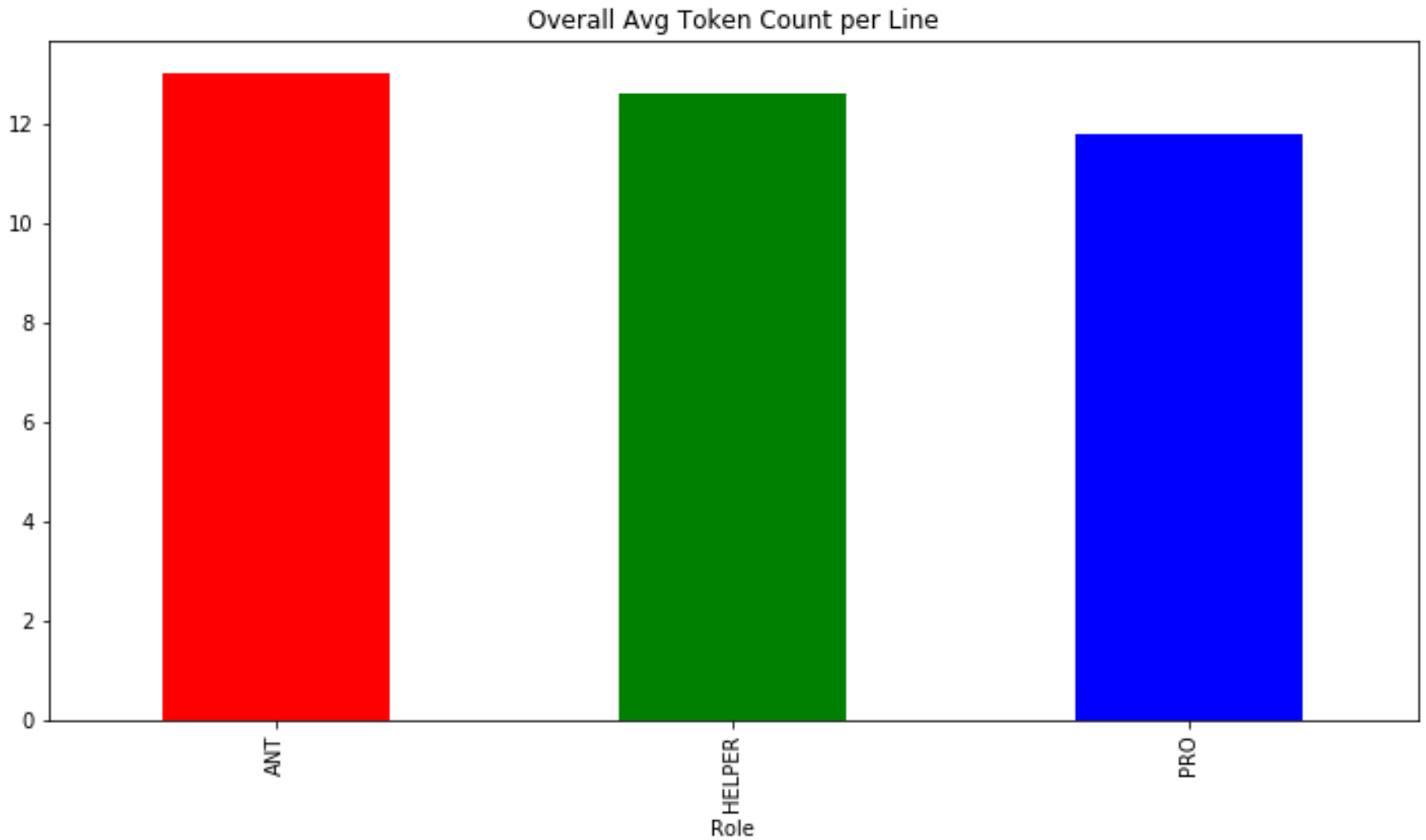
LINE LENGTH


Average Difference in Token Count between Genders
(2016, Moana)
(2014, How to Train Your Dragon 2)
(2013, The Croods)
(2013, Frozen)
(2012, Rise of the Guardians)
(2012, Brave)
(2010, Tangled)
(2010, Megamind)
(2010, How to Train Your Dragon)
(2009, The Princess and the Frog)
(2008, Kung Fu Panda)
(2007, Shrek 3)
(2001, Shrek)
(1998, Mulan)
(1998, Antz)
(1995, Pocahontas)
(1992, Aladdin)
(1991, Beauty and the Beast)
(1989, The Little Mermaid)
(1959, Sleeping Beauty)
(1950, Cinderella)
(1937, Snow White)
Movie By Year
−2
0
2
4
6
Average Difference in Token Count by Gender

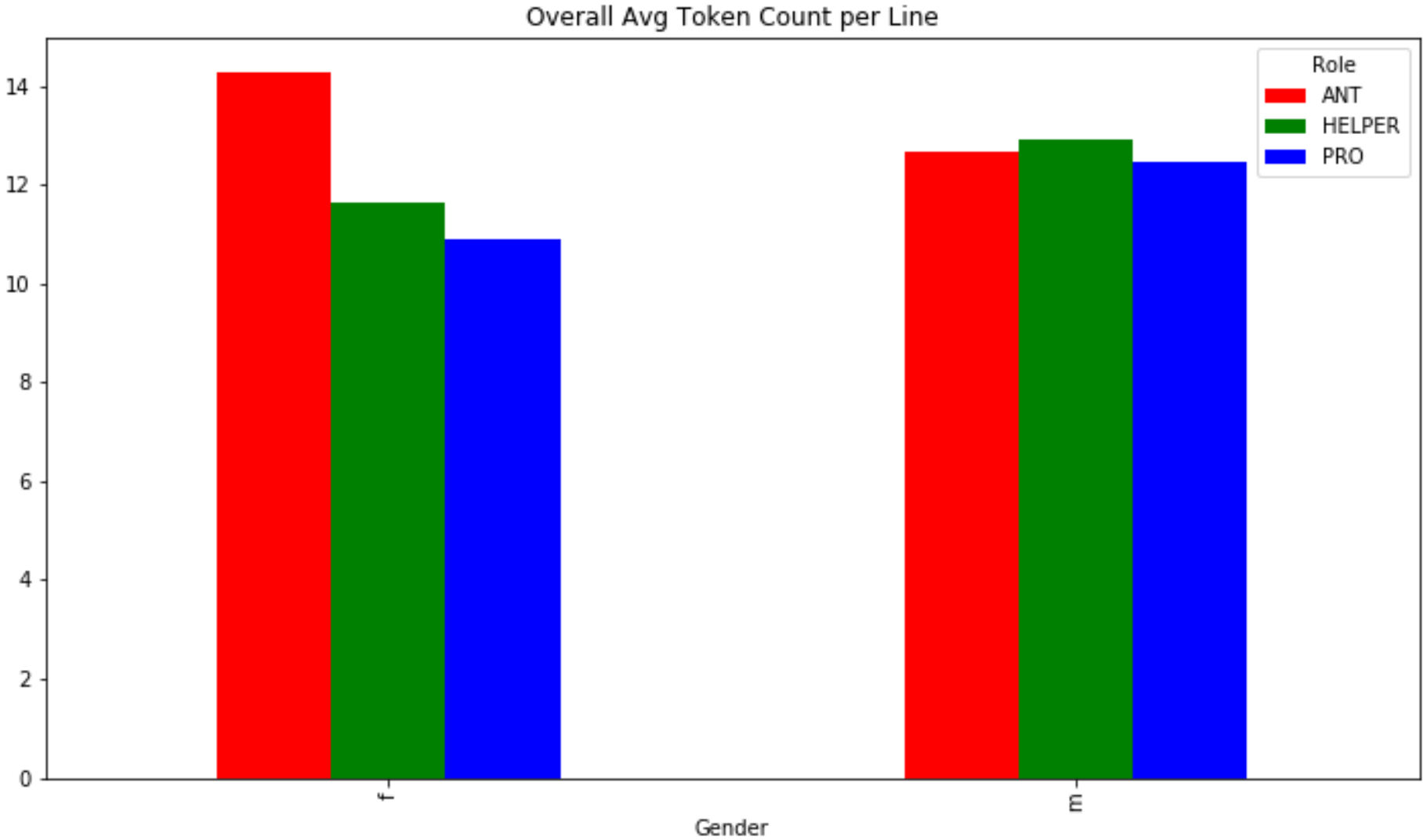
LINE LENGTH


Average Token Count per Line by Role
Role
ANT
HELPER
PRO
Avg Tokens per Line
0
5
10
15
20
25
30
(1937, Snow White)
(1950, Cinderella)
(1959, Sleeping Beauty)
(1989, The Little Mermaid)
(1991, Beauty and the Beast)
(1992, Aladdin)
(1995, Pocahontas)
(1998, Antz)
(1998, Mulan)
(2001, Shrek)
(2007, Shrek 3)
(2008, Kung Fu Panda)
(2009, The Princess and the Frog)
(2010, How to Train Your Dragon)
(2010, Megamind)
(2010, Tangled)
(2012, Brave)
(2012, Rise of the Guardians)
(2013, Frozen)
(2013, The Croods)
(2014, How to Train Your Dragon 2)
(2016, Moana)
Movie by Year

LINE LENGTH

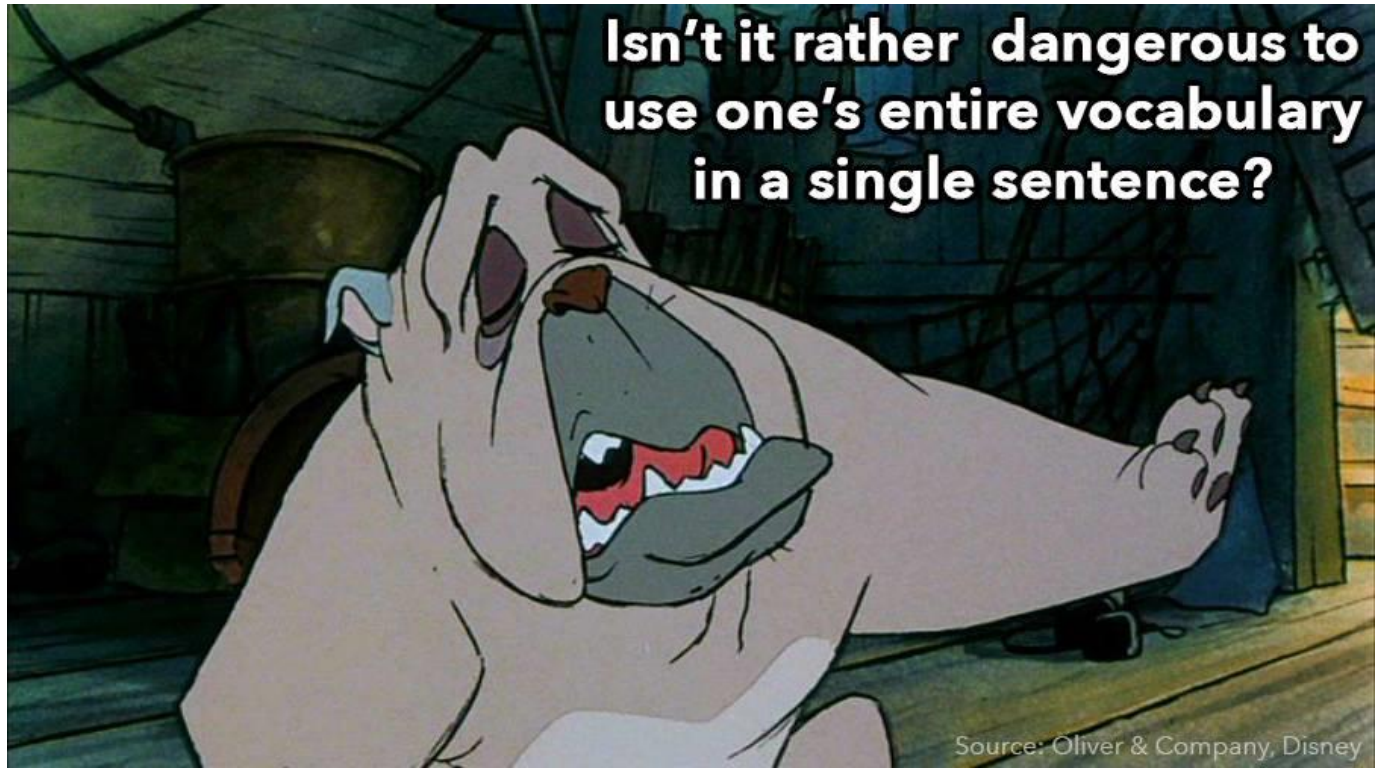

Overall Avg Token Count per Line
12
10
8
6
4
2
0
ANT
HELPER
PRO
Role

LINE LENGTH

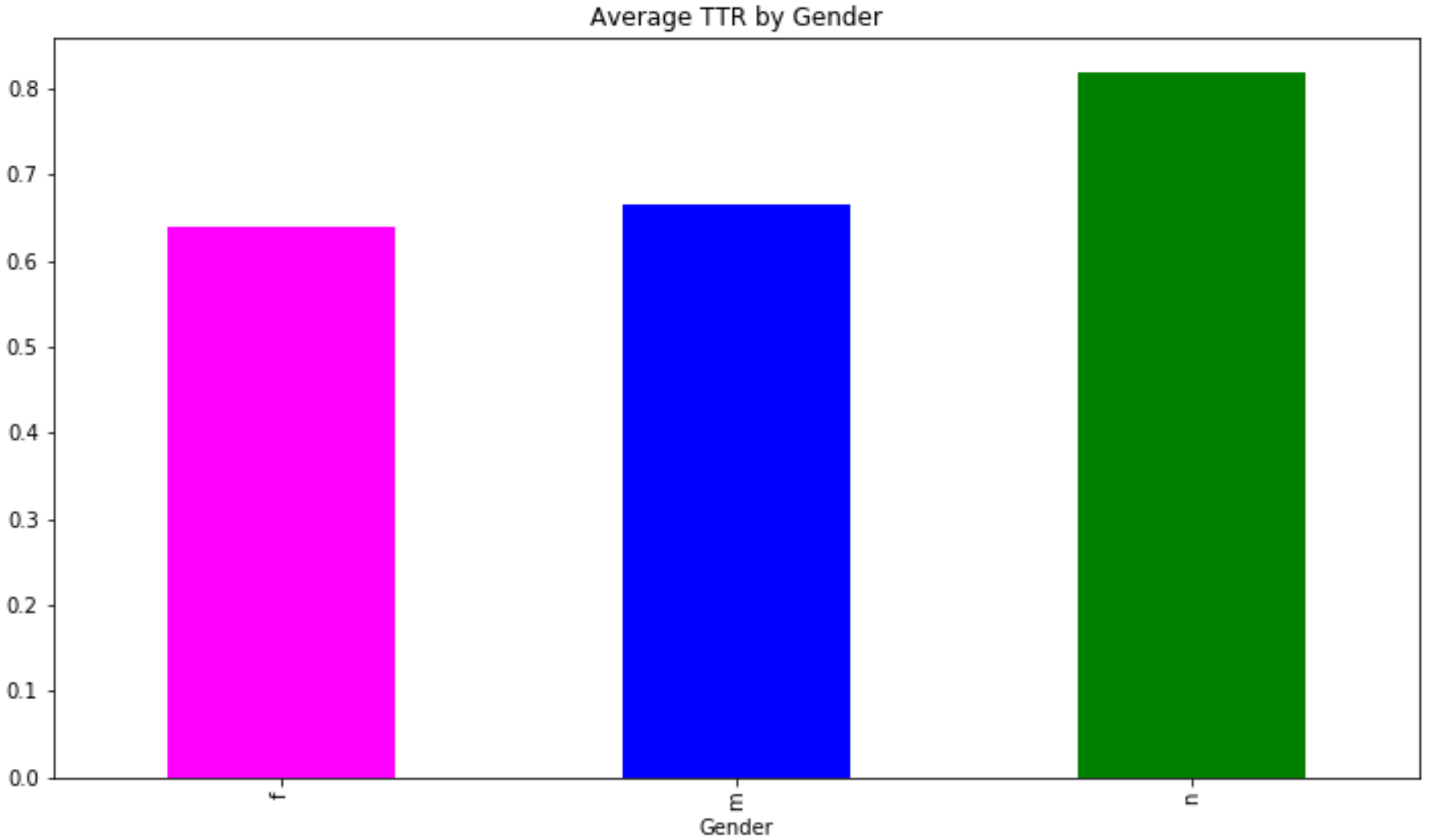

Overall Avg Token Count per Line
Role
ANT
HELPER
PRO
14
12
10
8
6
4
2
0
f
m
Gender

VOCABULARY LEVEL

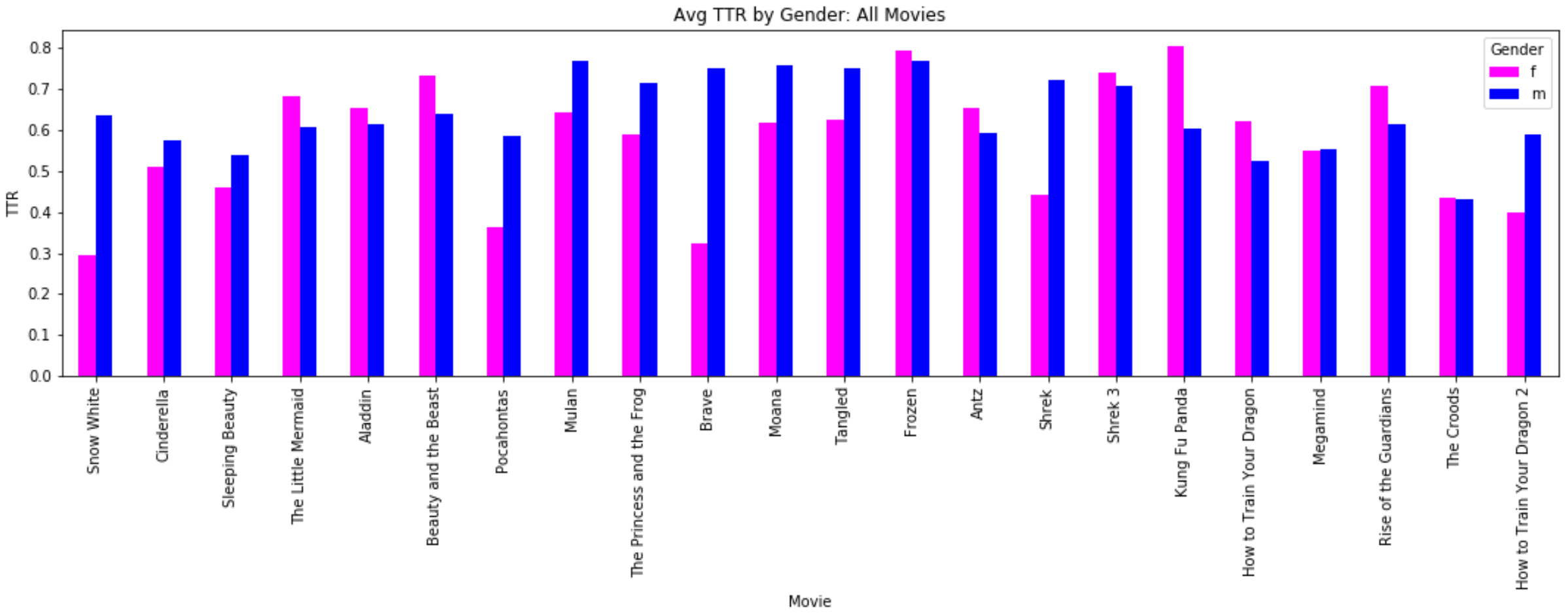
- Hypothesis: female speakers will have a more complex vocabulary
- Total Token/Type Counts used
- Analyzed with
  - TTR
  - K-bands



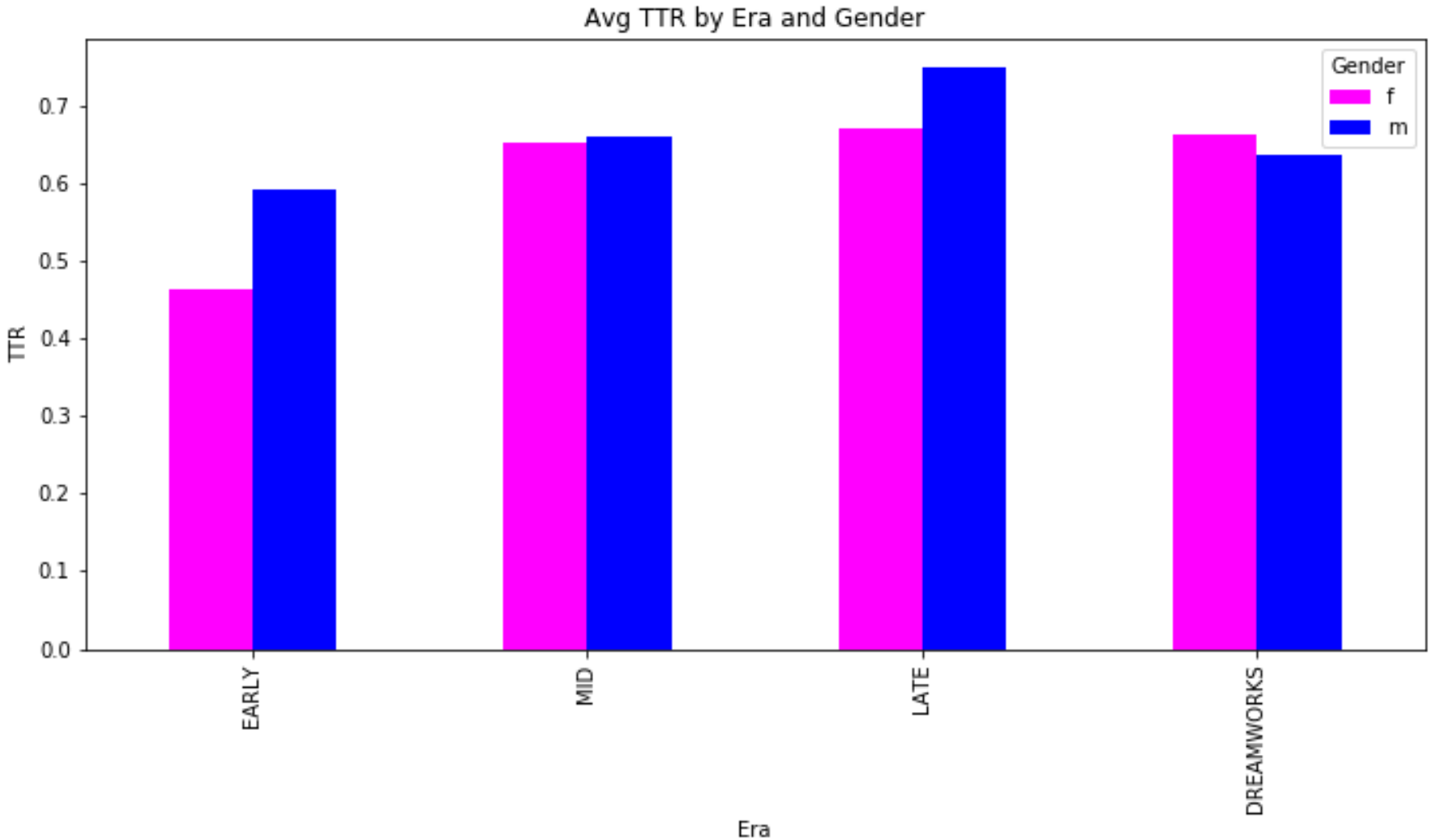
ATTEMPT 1: TTR



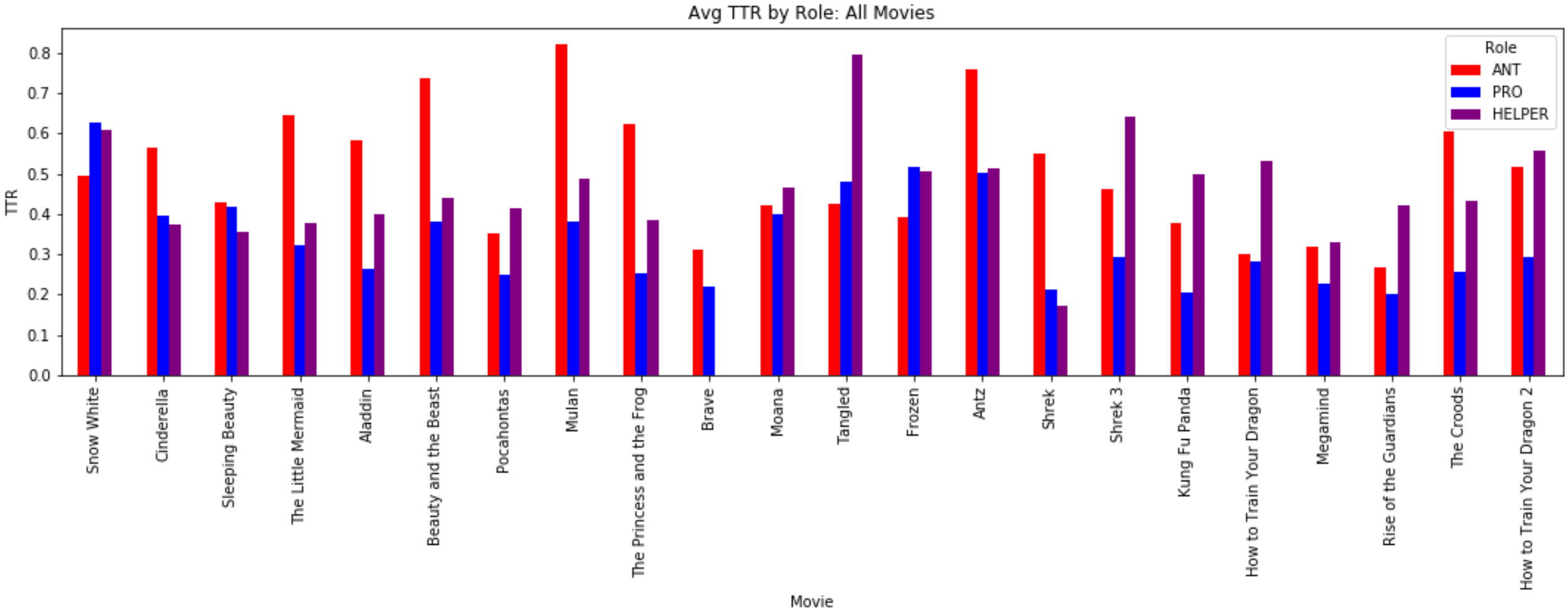
ATTEMPT 1: TTR



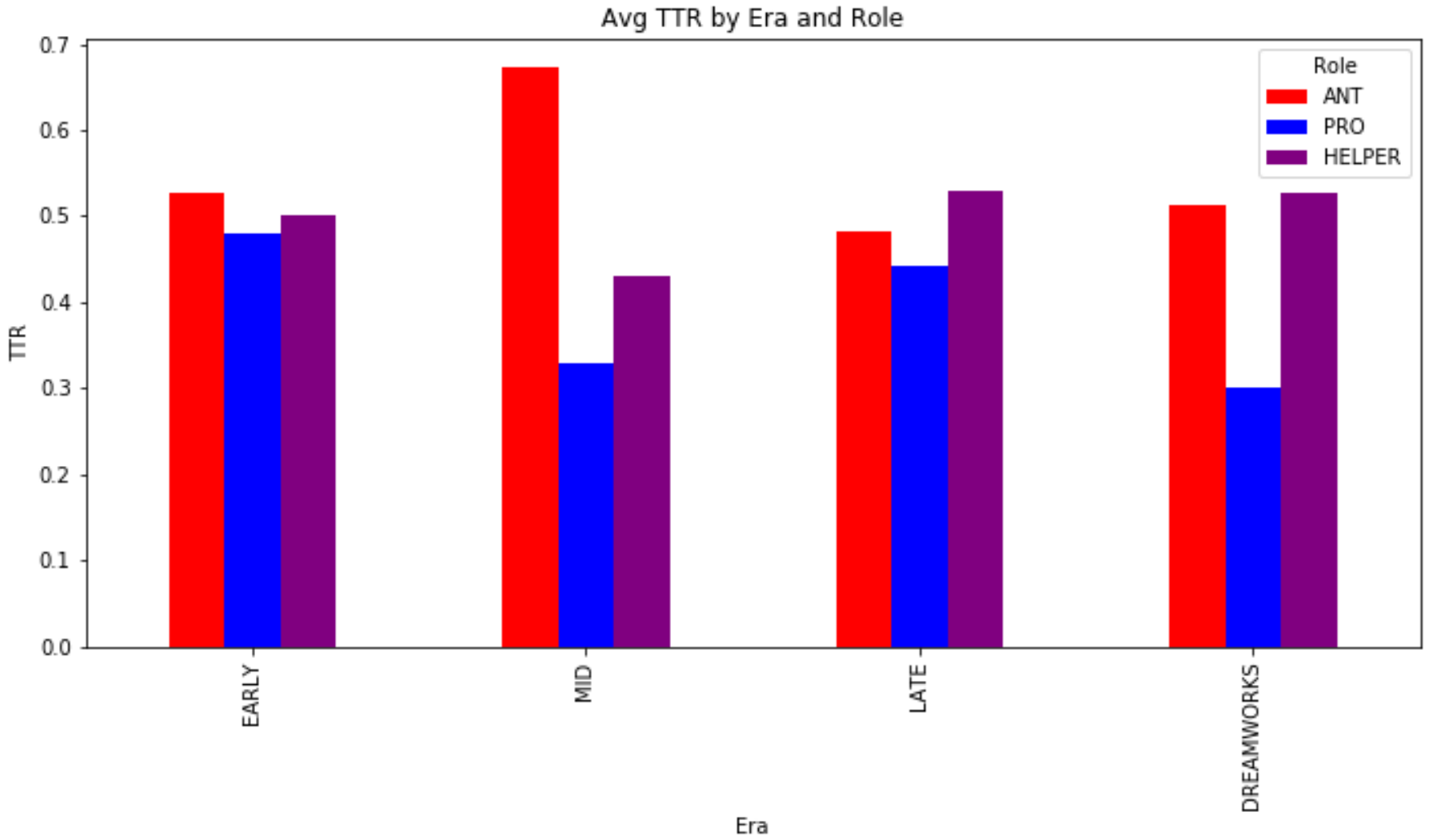
ATTEMPT 1: TTR



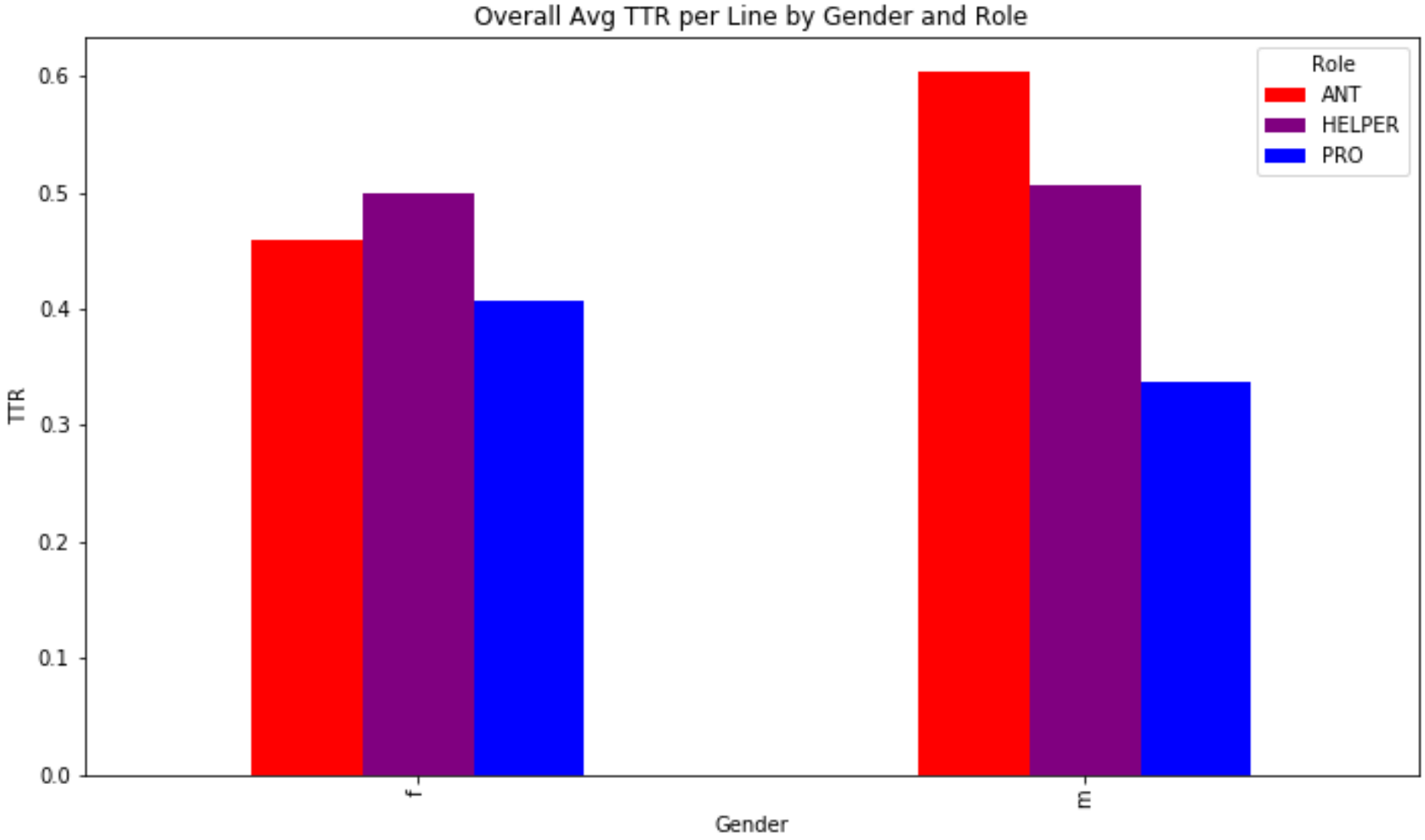
ATTEMPT 1: TTR



ATTEMPT 1: TTR



ATTEMPT 1: TTR



ANALYSIS:

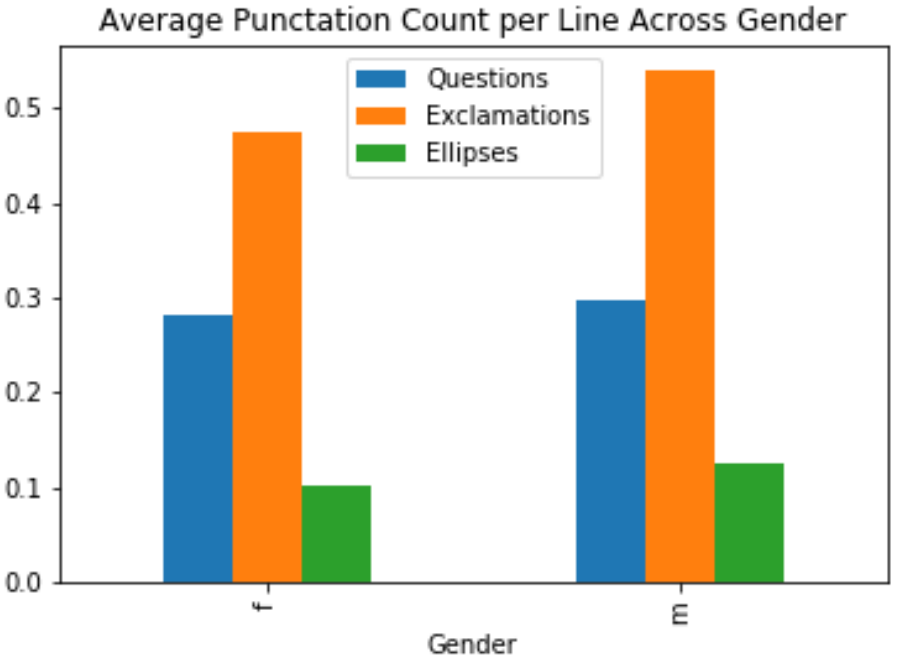
Punctuation: Exclamations, Questions, and Ellipses

THE IDEA

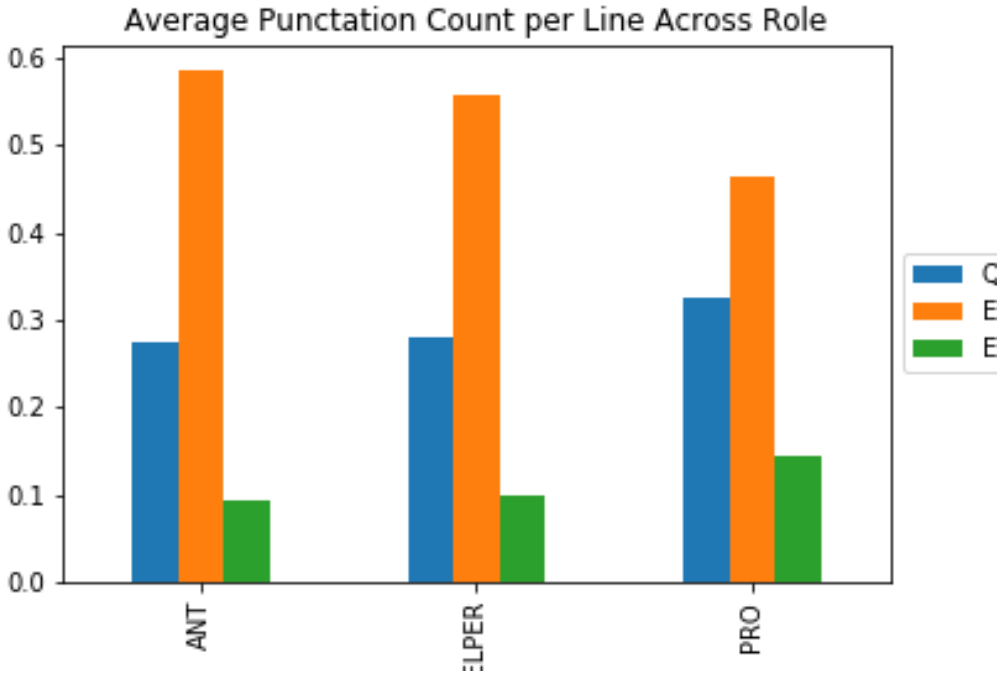
- A character with more authority/confidence will use more exclamations
- A character who is uncertain will use more questions and hesitation
- Some obvious issues
  - “I can do it!” vs “Aaah!”
  - “Can I help you?” (intonation)
  - Discretion of the transcriber

PUNCTUATION IN GENERAL

By Gender



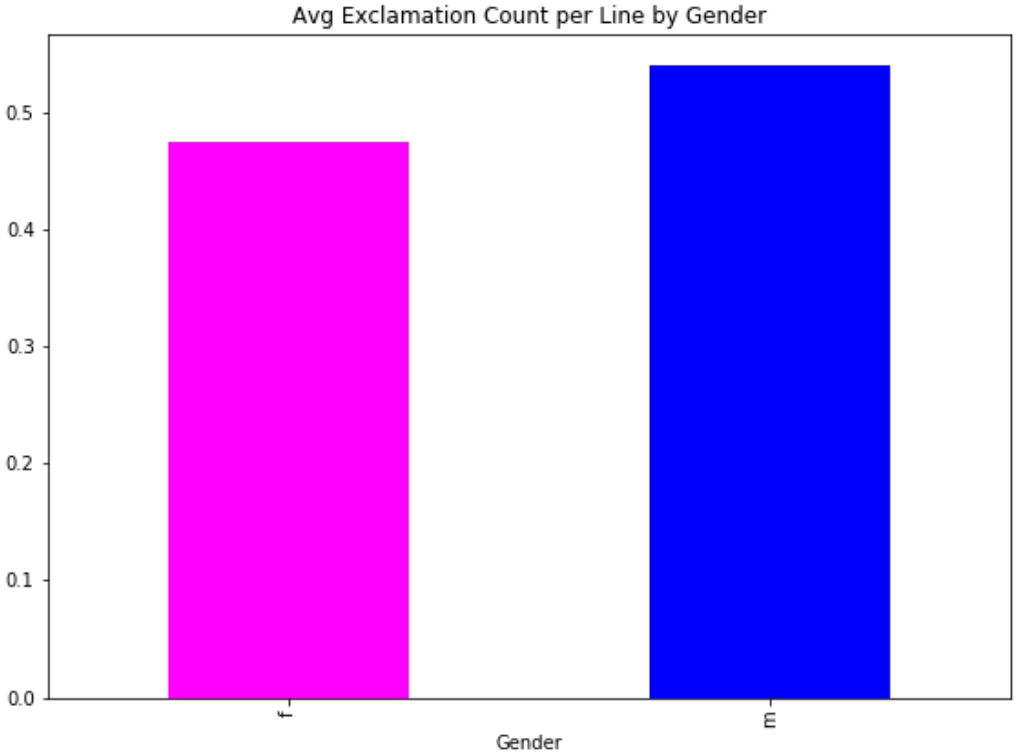
By Role

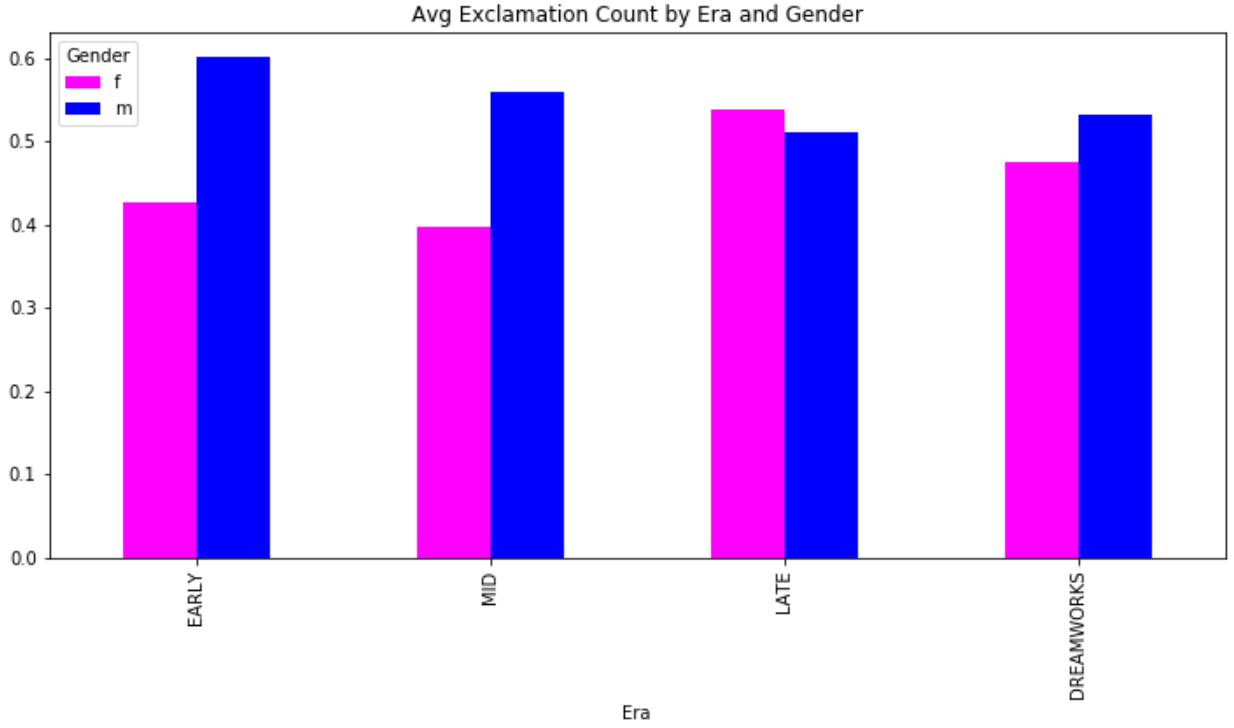


EXCLAMATIONS

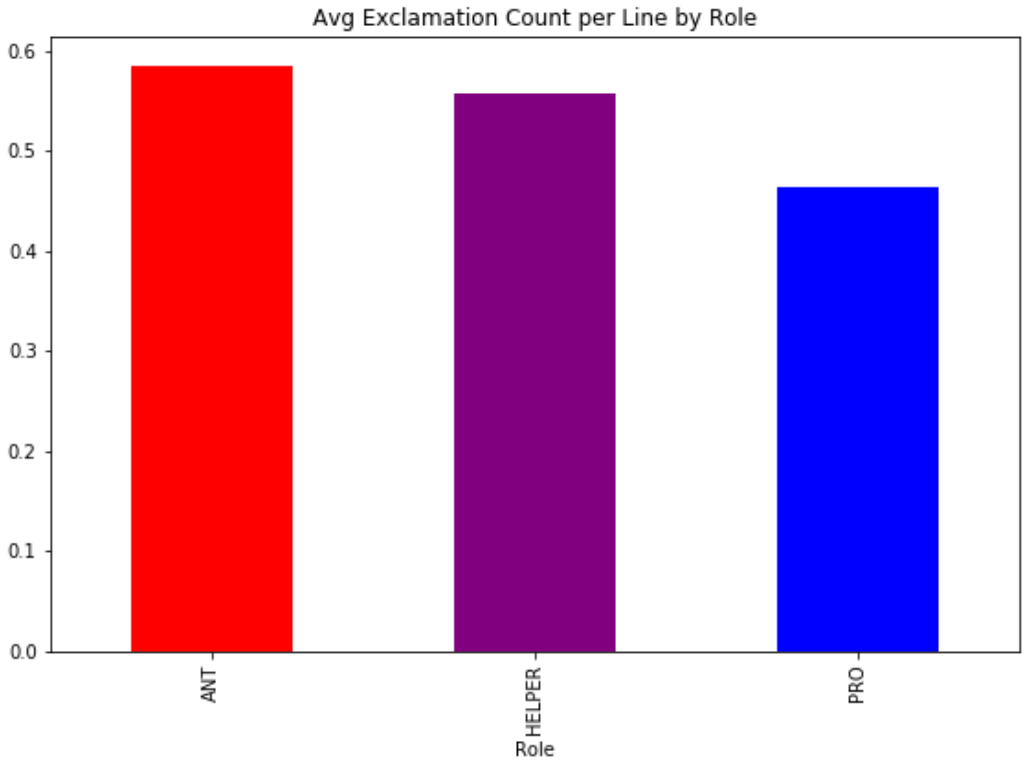
- Hypothesis: Men will use more exclamations than women
- The Idea:
- More shouting -> More Authority
- Found average number of exclamation points in each utterance

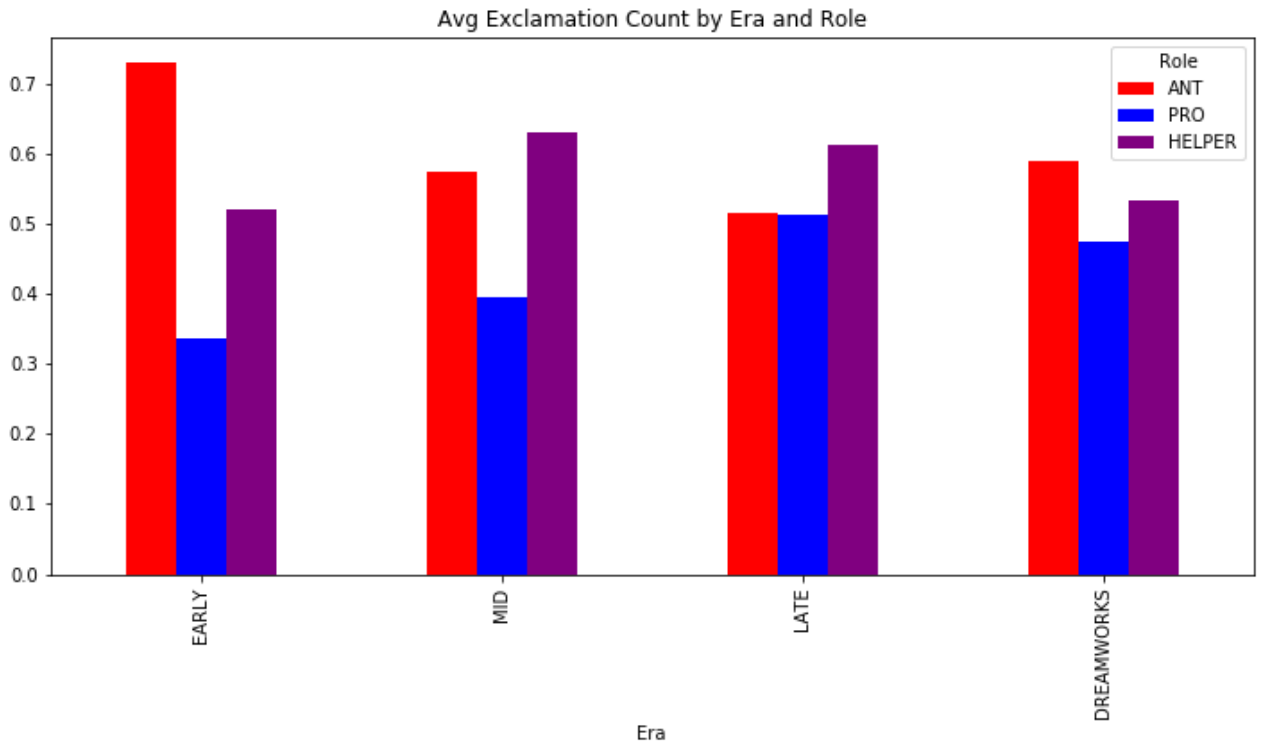
EXCLAMATIONS: GENDER



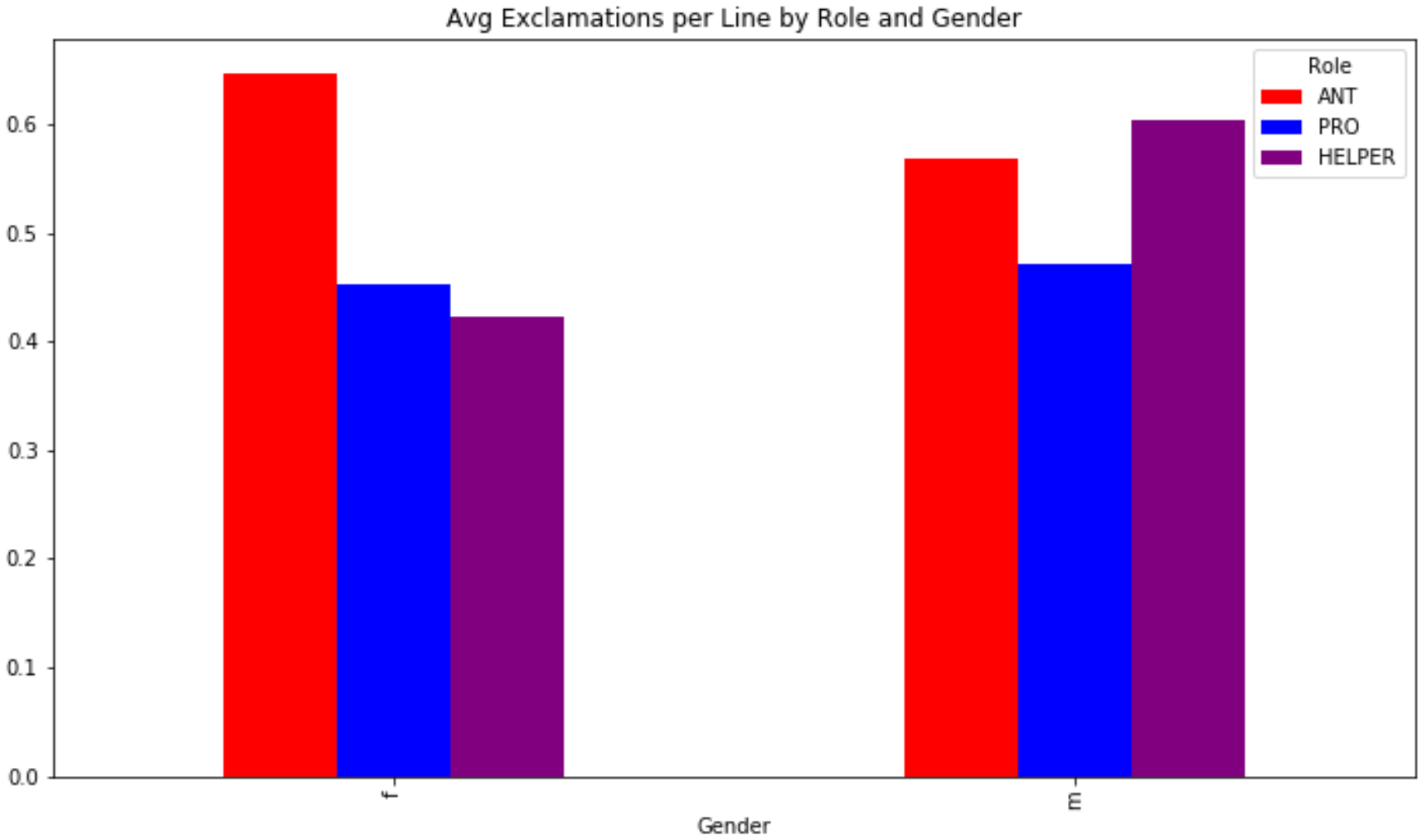


EXCLAMATIONS: ROLE





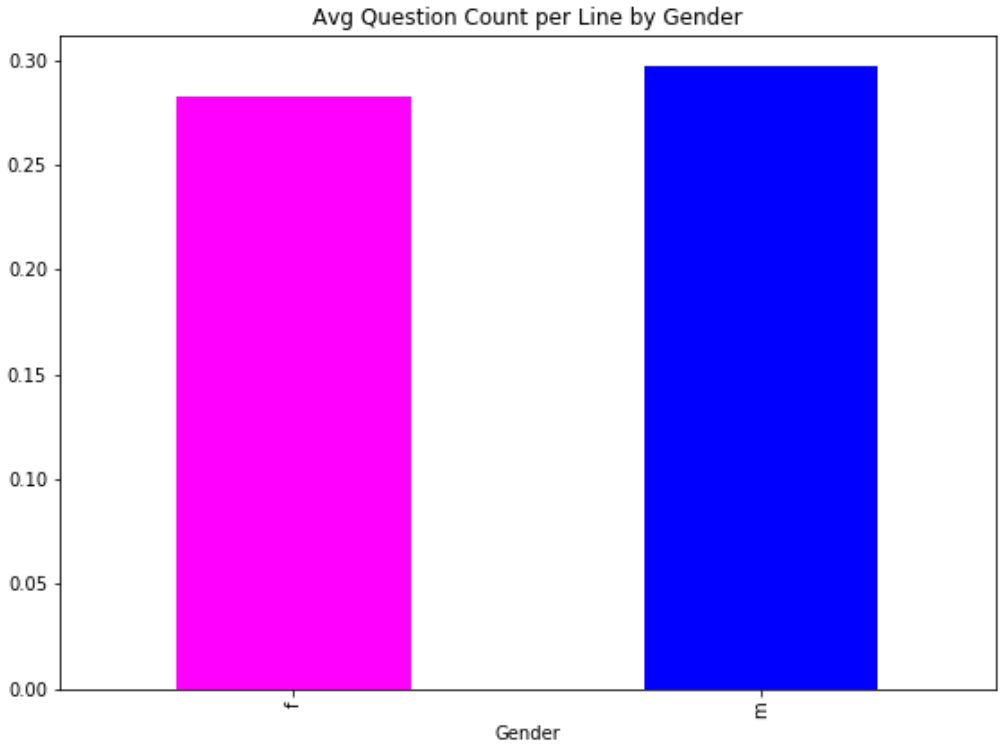
EXCLAMATIONS: ROLE & GENDER

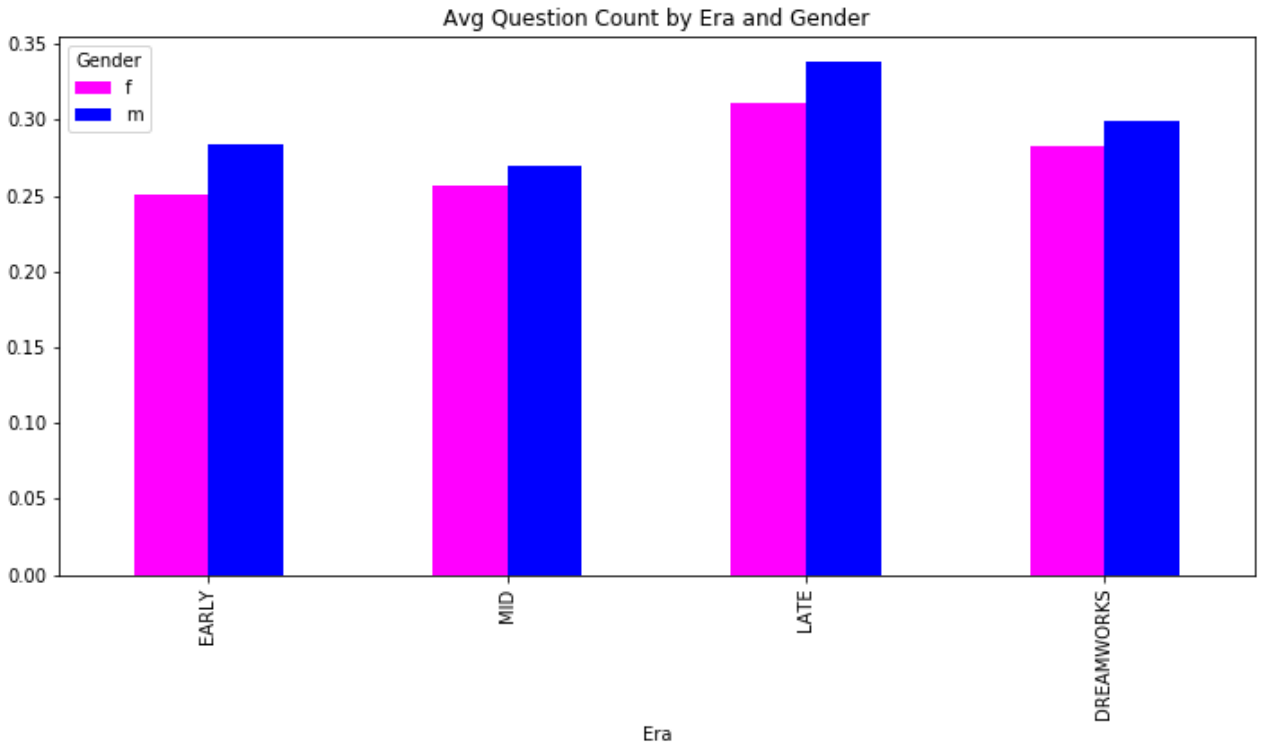


QUESTIONS

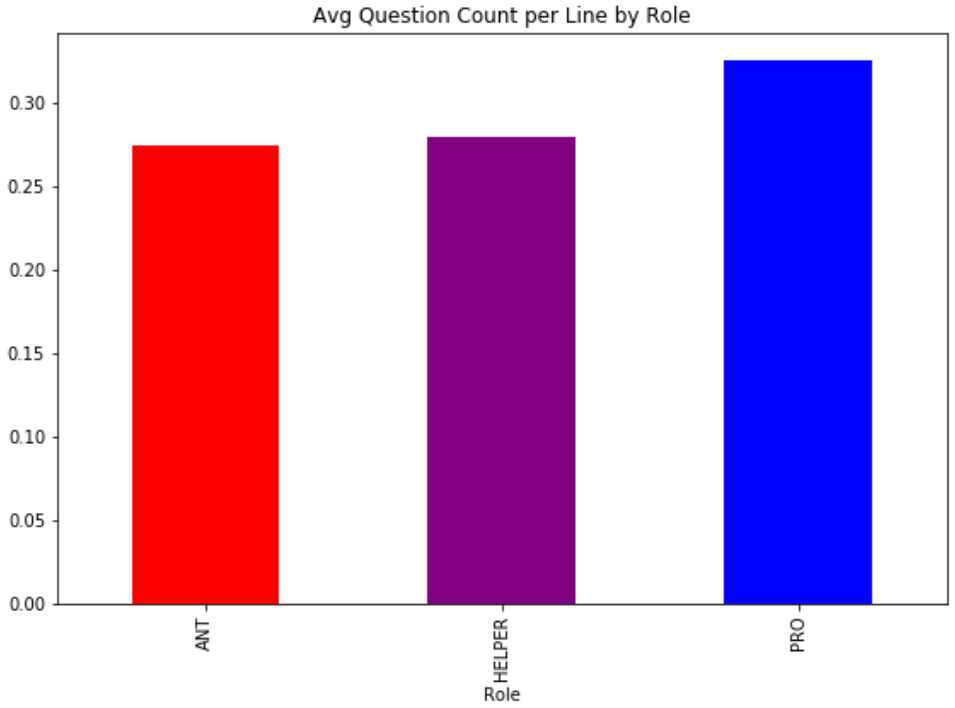
- Hypothesis: Women will use more questions than men. Female antagonists won't use as many questions as female protagonists
- The Idea:
  - Fewer Questions -> More Authority / Less Uncertainty
- Found average number of question marks in each utterance

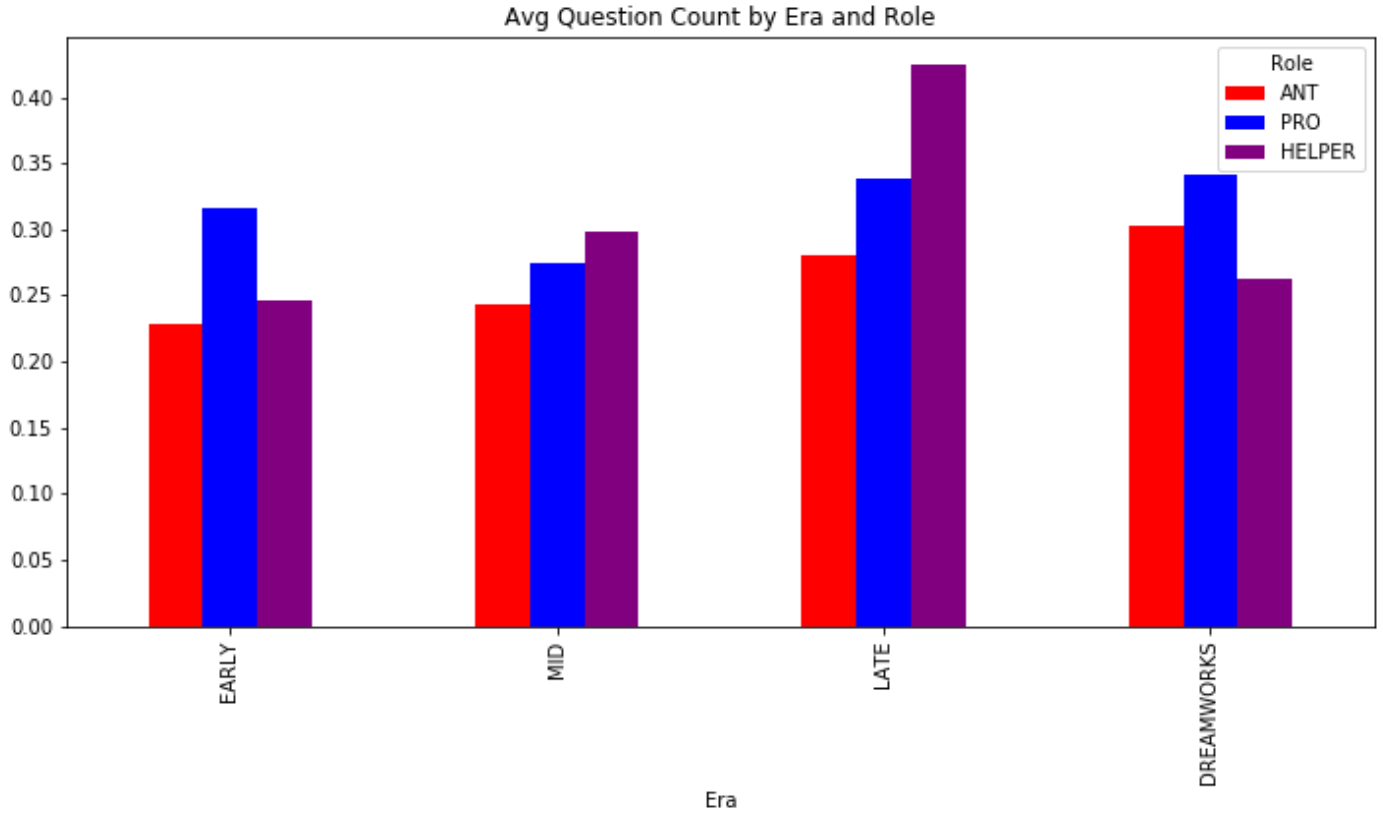
QUESTIONS: GENDER



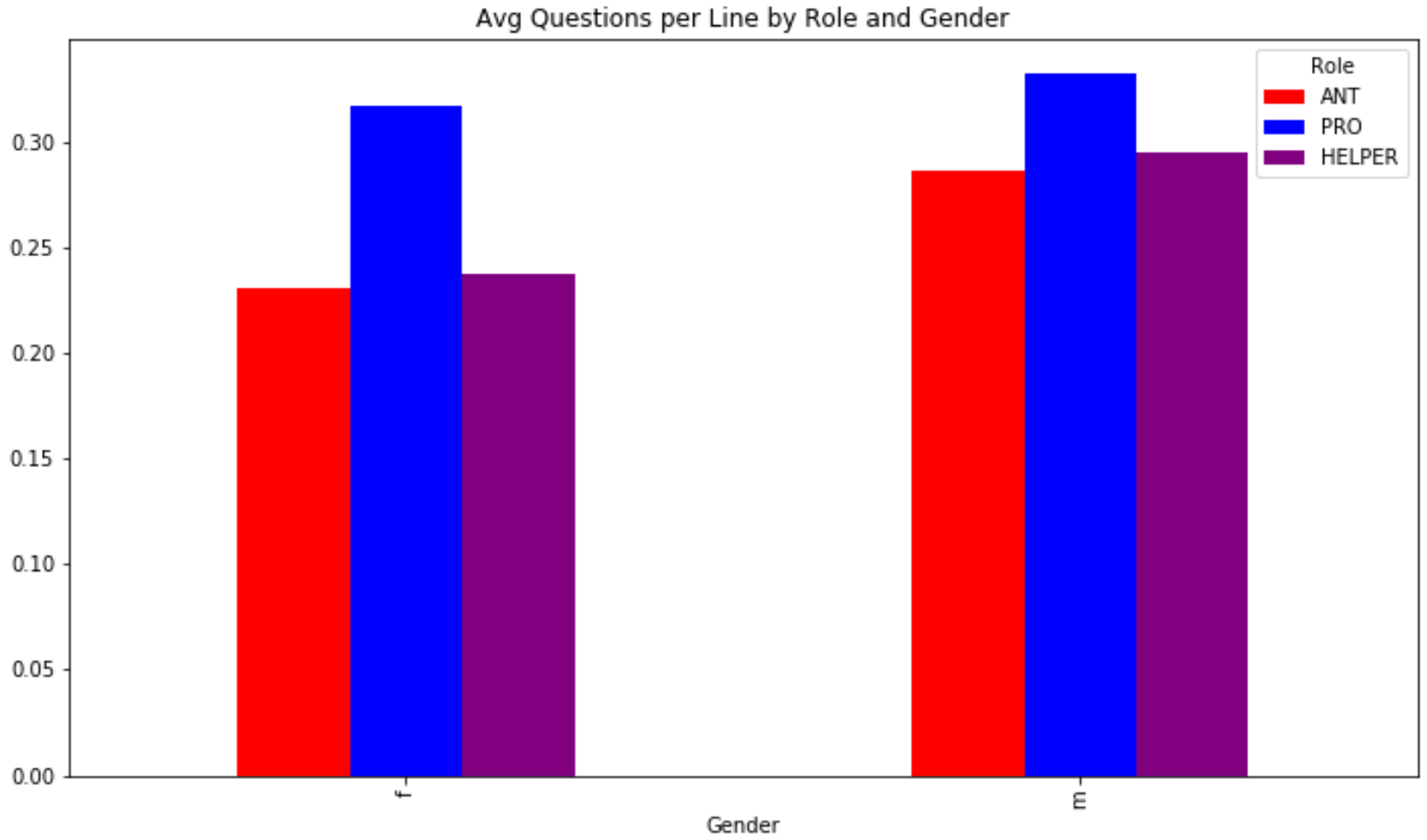


QUESTIONS: ROLE





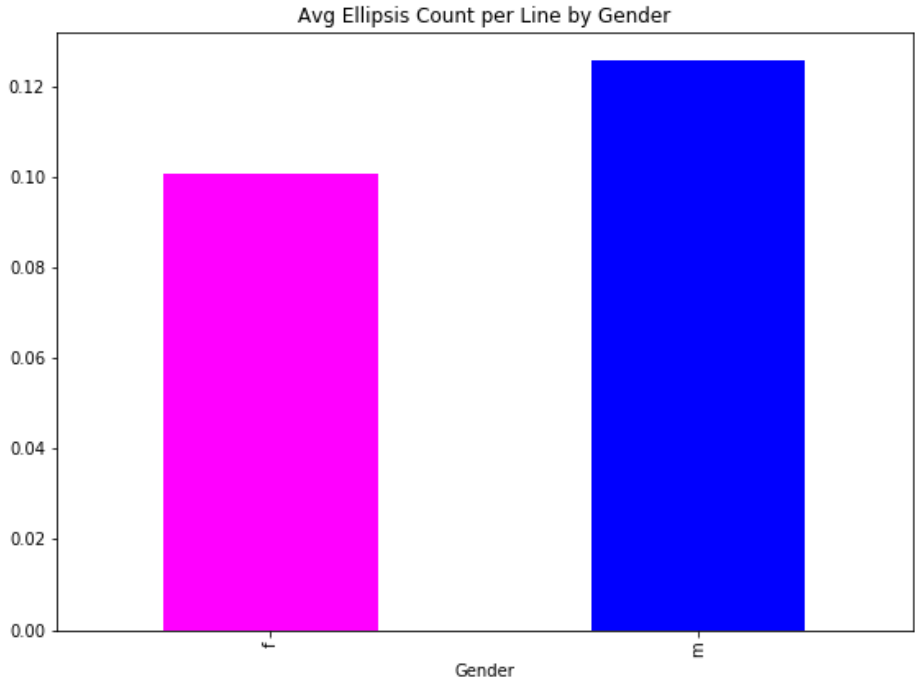
QUESTIONS: ROLE & GENDER

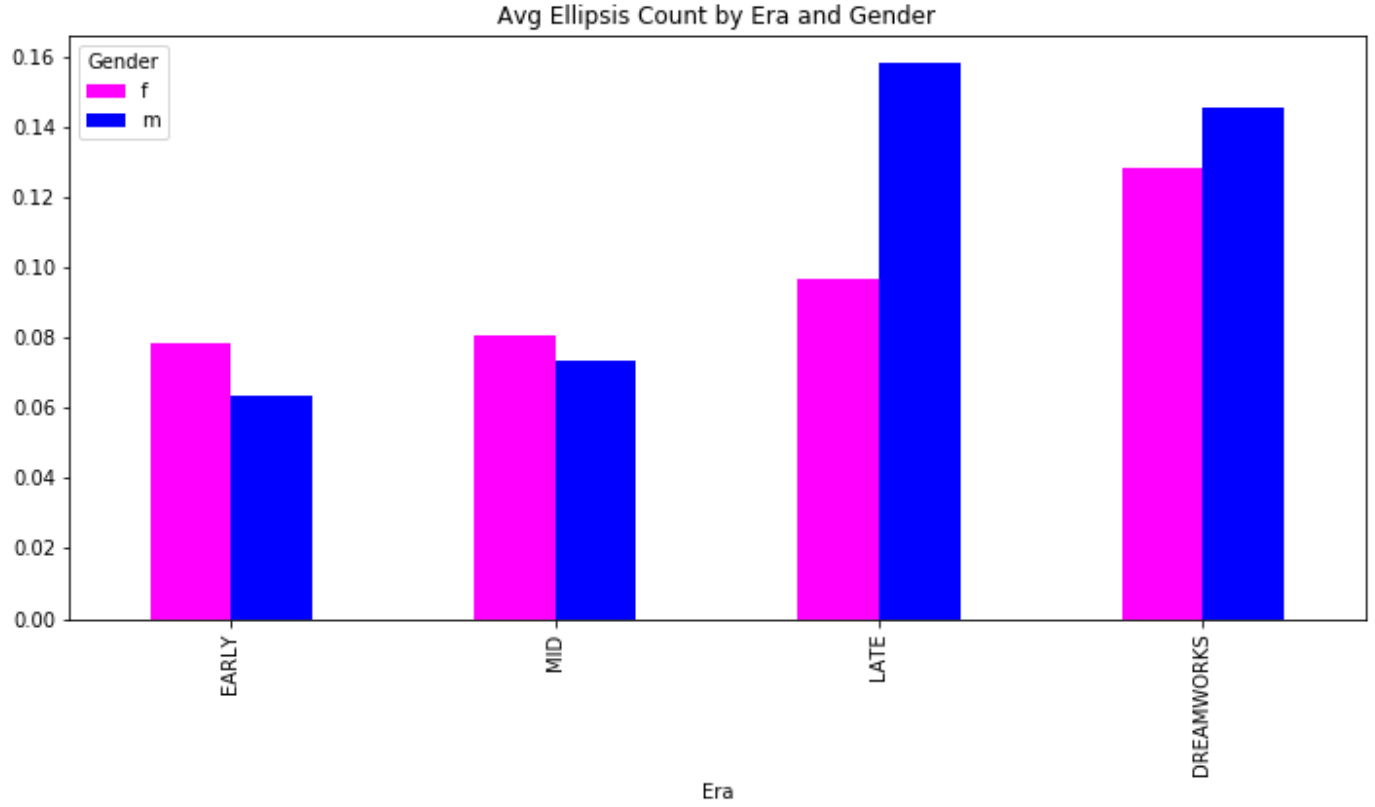


ELLIPSIS

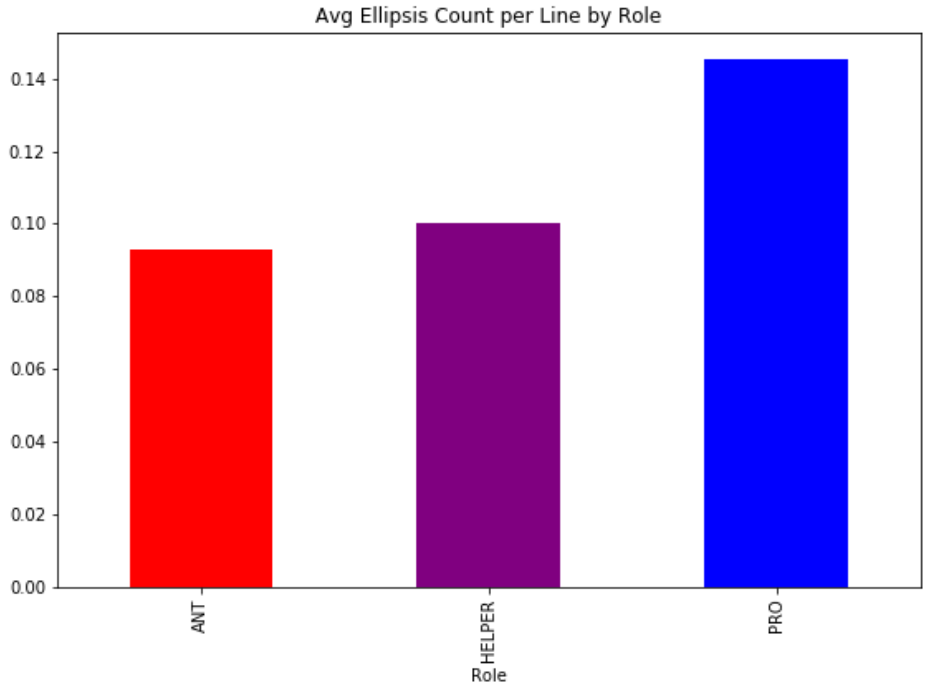
- Hypothesis: Women will use more than men. Female antagonists won't use as many as female protagonists
- The Idea:
  - Less Hesitation -> More Authority / Less Uncertainty
- Found average number of ellipses in each utterance

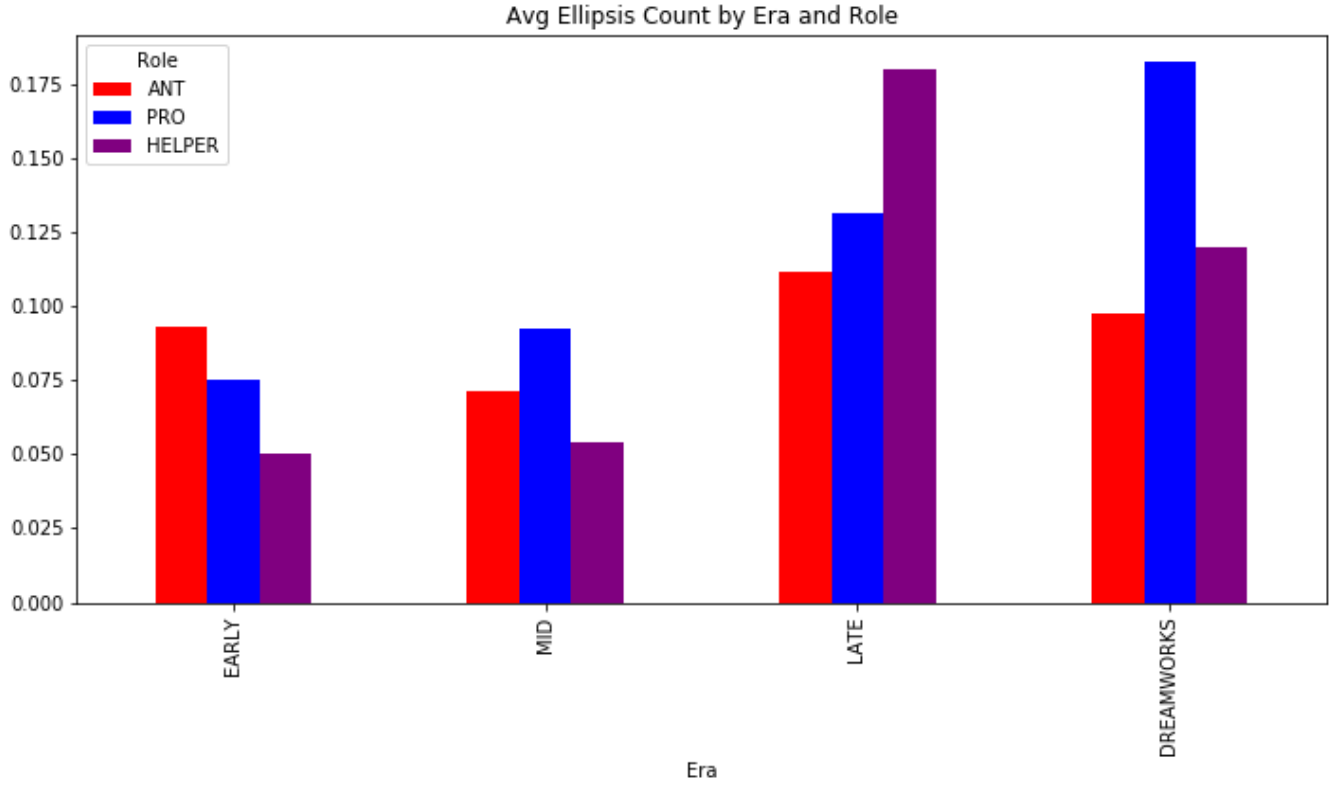
ELLIPSIS: GENDER



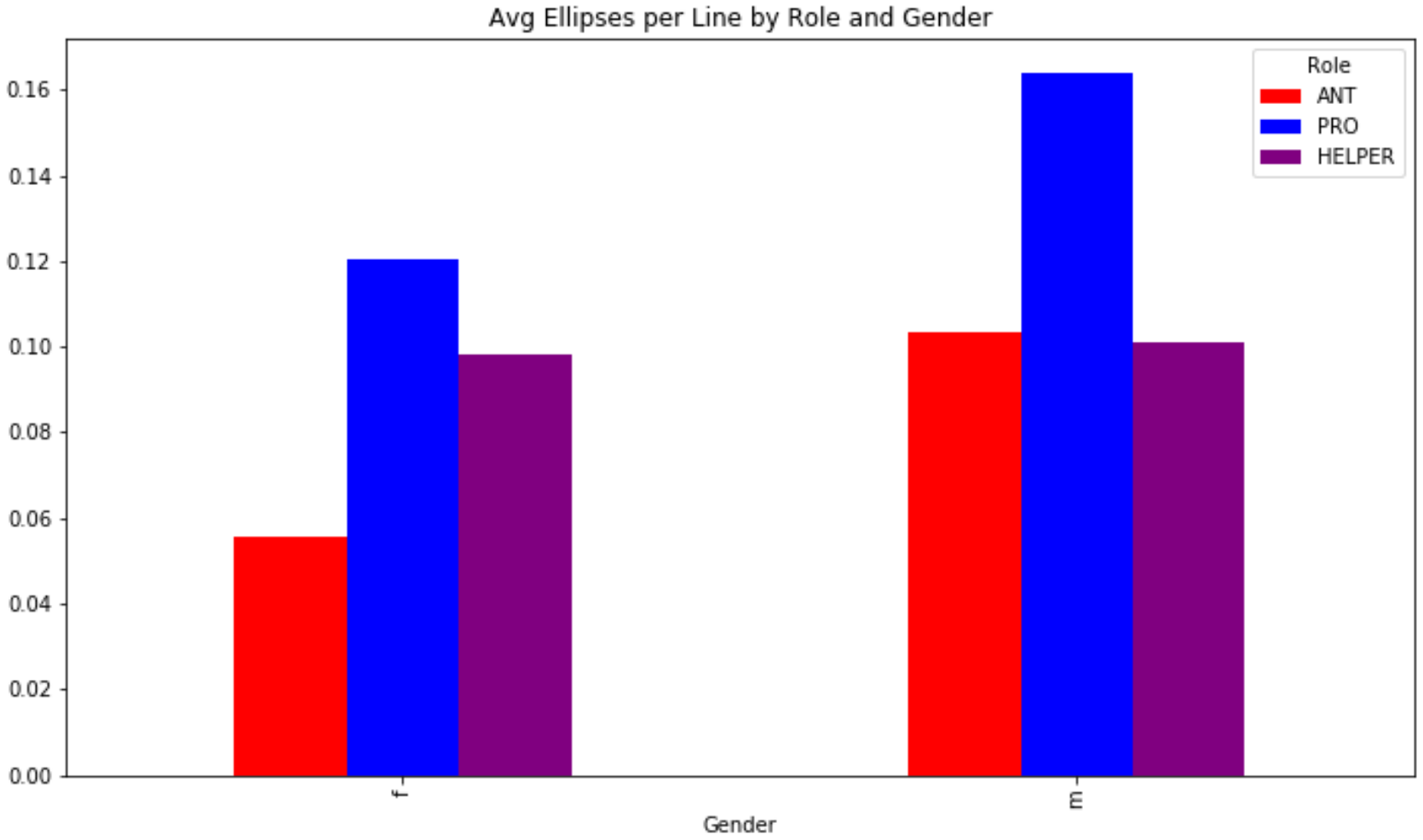


ELLIPSIS: ROLE





ELLIPSIS: ROLE & GENDER



INTERRUPTIONS—AN ATTEMPT

- Only 152 interruptions in total
- Many movies don't mark interruption

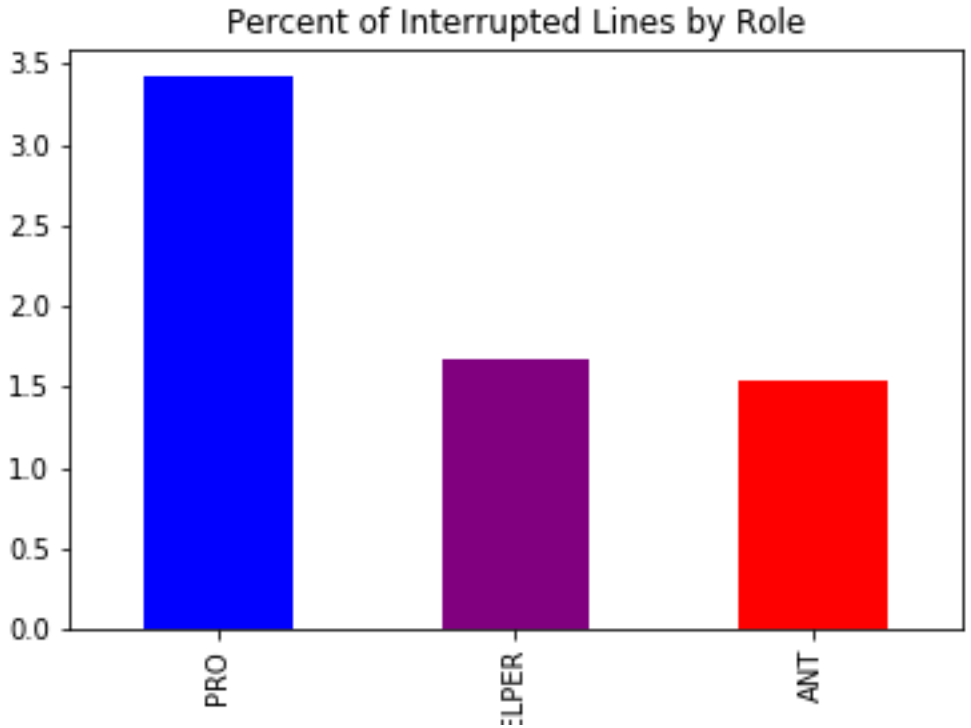
```
def is_interrupted(line):
    if line.endswith('-'):
        return 'Yes'
    else:
        return 'No'
```

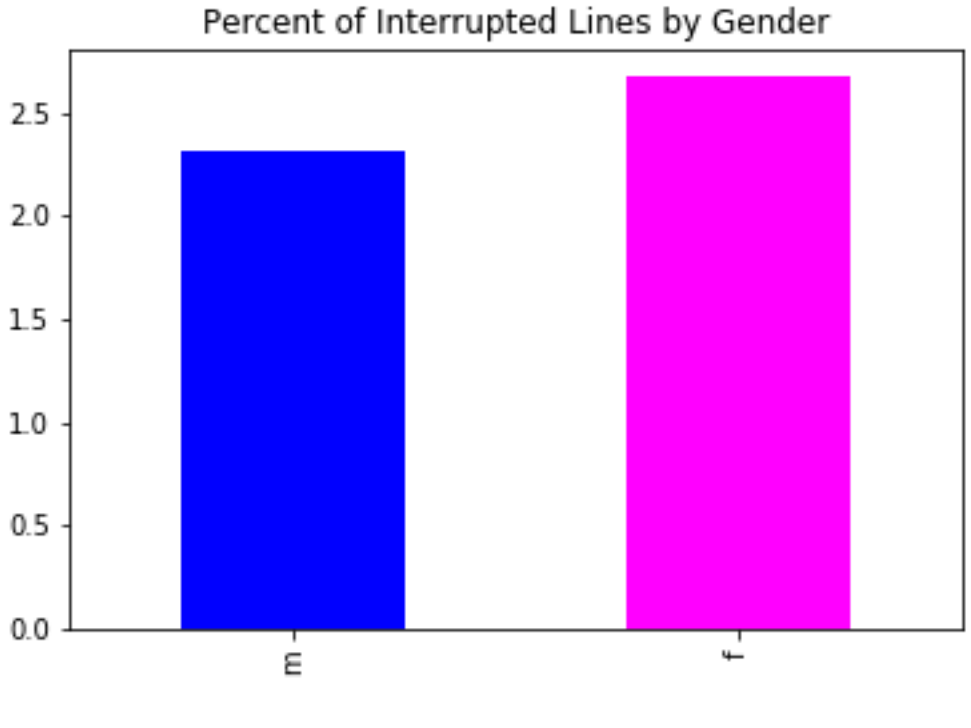
```
movies_df['Interrupted'] = movies_df.Text.map(is_interrupted)
```

```
movies_df[movies_df['Interrupted'] == 'Yes'].Movie.unique()
```

```
array(['Moana', 'Antz', 'Shrek', 'Shrek 3', 'Kung Fu Panda',
       'How to Train Your Dragon', 'Rise of the Guardians', 'The Croods',
       'How to Train Your Dragon 2'], dtype=object)
```

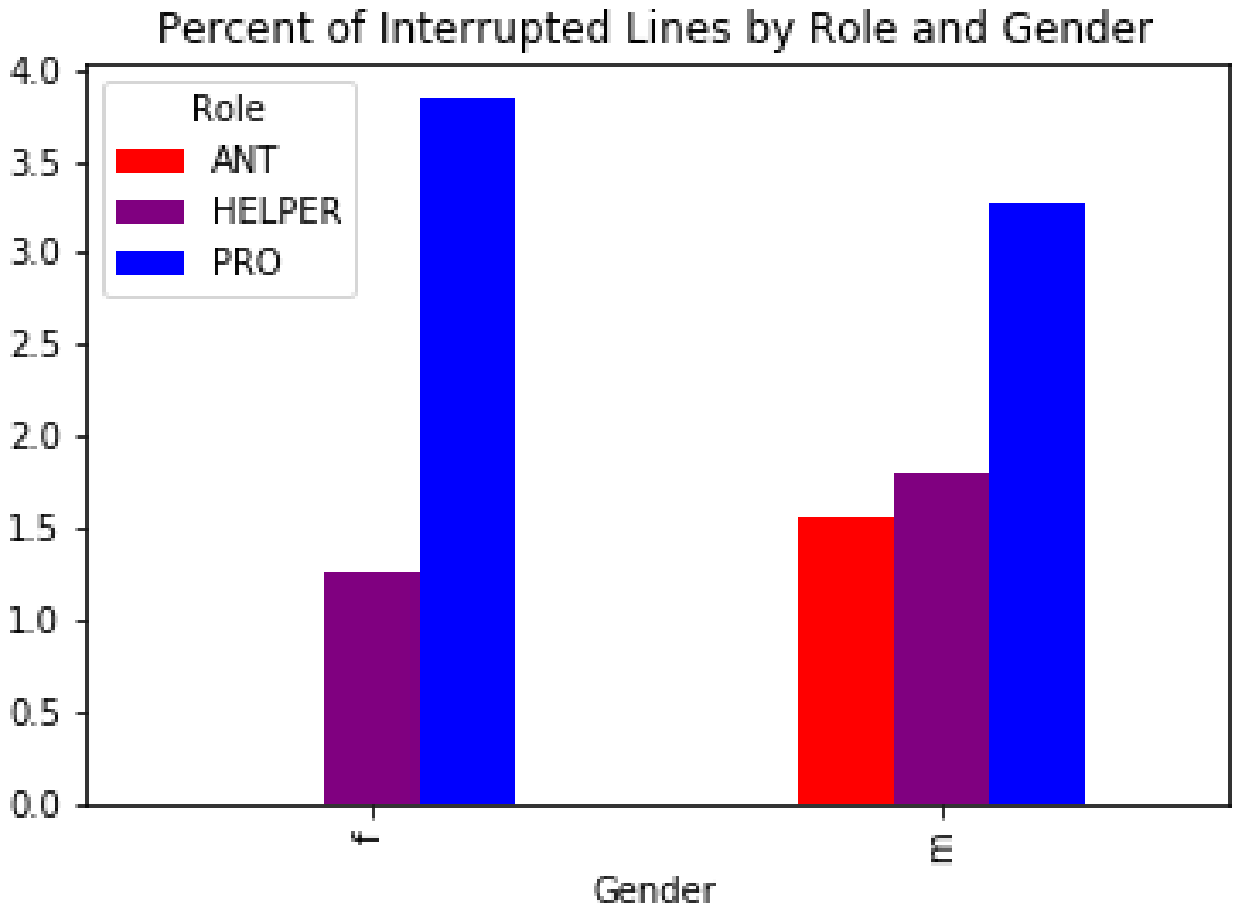
INTERRUPTIONS—AN ATTEMPT



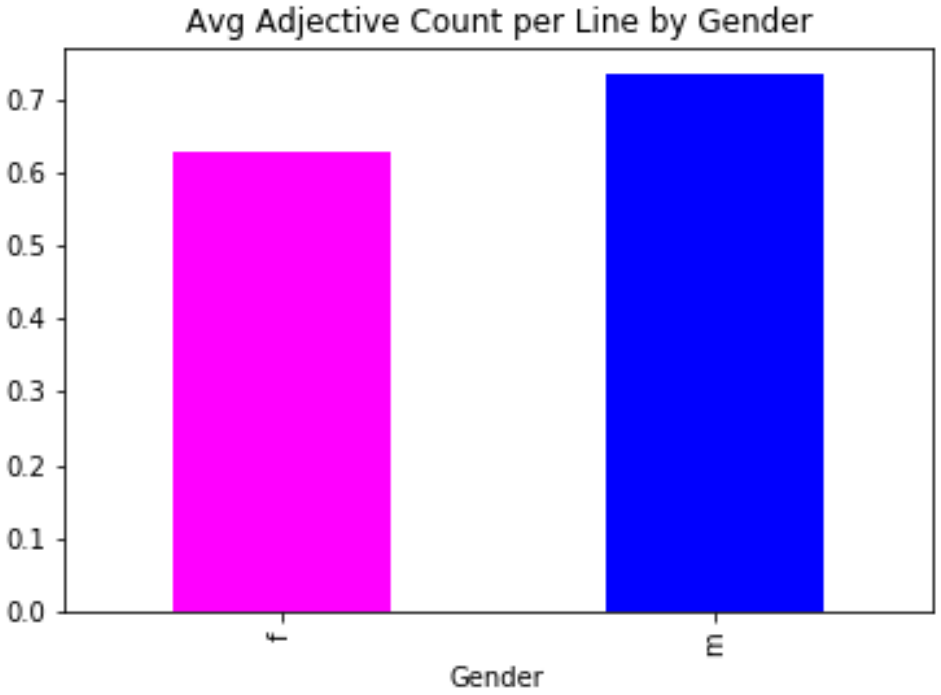


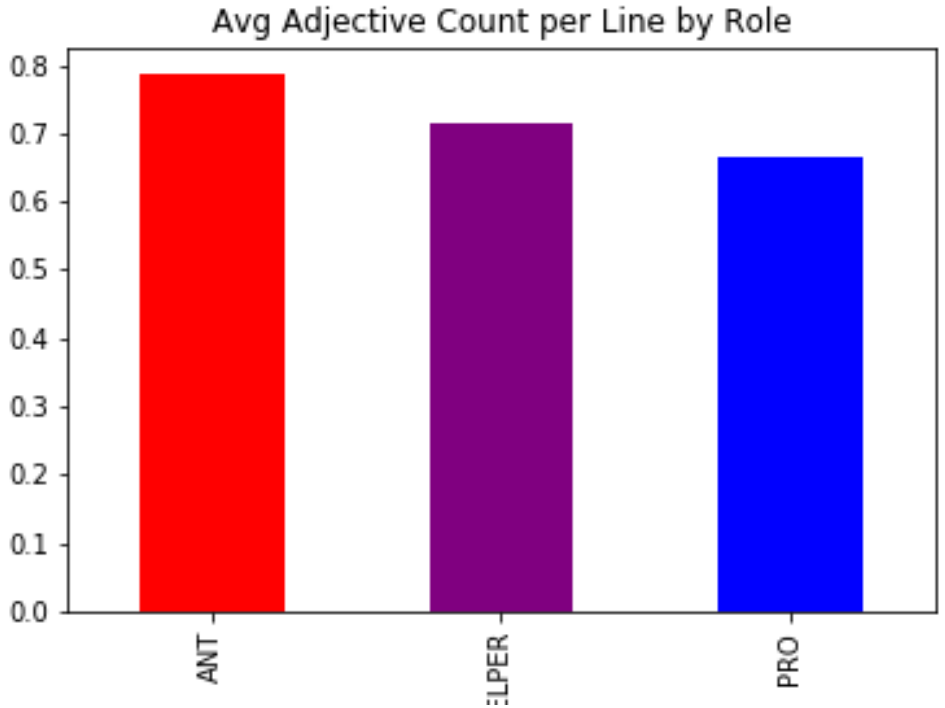
INTERRUPTIONS—AN ATTEMPT

- No female antagonists are interrupted?!?
  - Careful—not significant!
  - Overall, not enough data for anything conclusive…



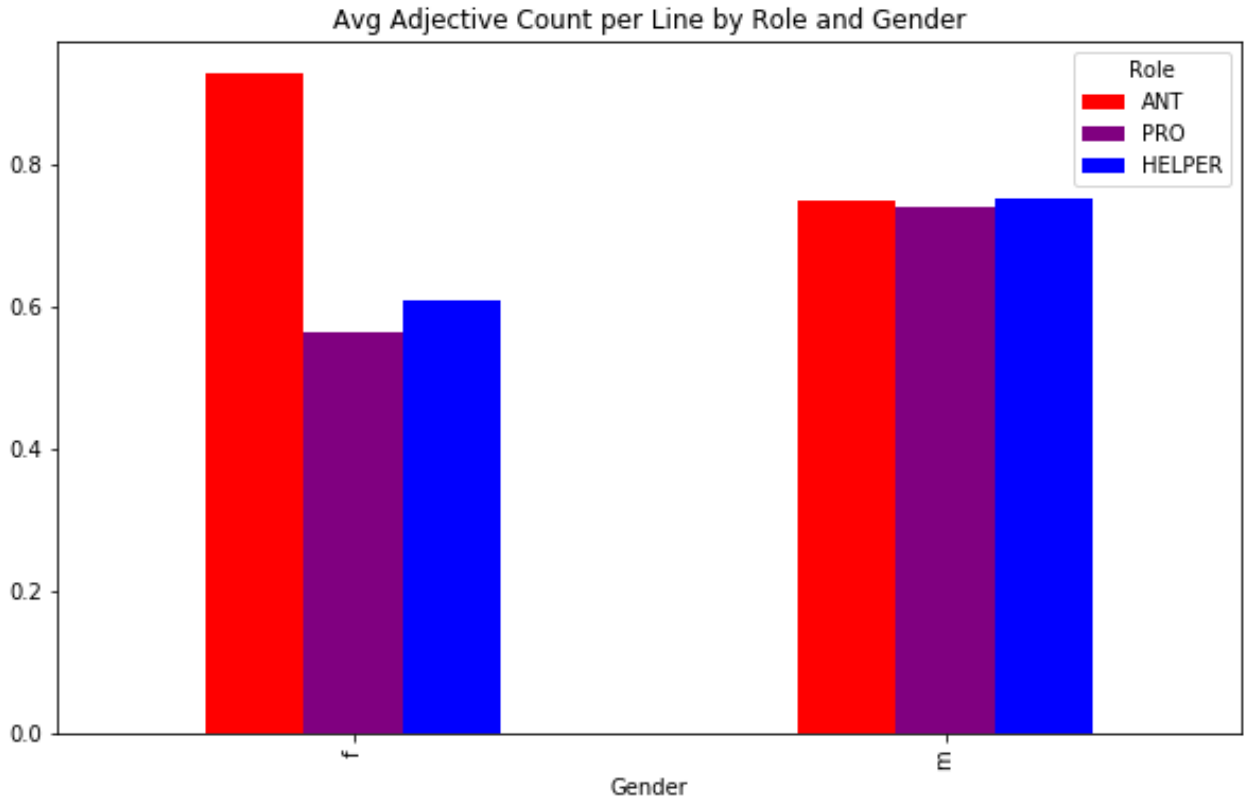
ADJECTIVES


Avg Adjective Count per Line by Gender
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
f
m
Gender


Avg Adjective Count per Line by Role
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
ANT
ELPER
PRO

ADJECTIVES

- This may have to do with line lengths….
  - Longer line -> more variety in words



COMMANDS?

- Tricky to find…
- Imperative Forms
  - Stop!
  - Go do your homework!
- Modals
  - You must stop
  - You should do your homework
  - You have to help
- False positives
  - You may be right

- One approach
  - Mark POS, find all VB's (base-forms)
    - Issue: infinitives
- Alternate approach
  - Regular expressions
    - Issue: differentiate between infinitives and commands

- I'm open to suggestions!

```
commands_4 = []
for line in movie_df.Text:
    if re.search(r'you (must|could)(?! be )', line) or re.search(r'you (should|shall|will|have to)', line):
        commands_4.append(line)
```

WHAT NEXT?

- Statistical Test – are these results significant?
- Politeness
- More work on commands
- Comparisons Across Production Companies

FURTHER QUESTIONS…

- Dialogue Interaction?
- How does intonation affect dialogue?
- How does royal status affect dialogue
- Machine Learning
  - Gender?
  - Role?
- More movies!


ANY
QUESTIONS?