

The Characters of Russian Folklore

Pranav Gonepalli, Ben O'Neill, Matthew Rosenthal



Our Research Question

How can we contextualize Russian culture by analyzing the frequency of certain types of characters appearing in folklore?

Through distant reading analysis, we will attempt to draw conclusions about the roles of characters in Russian folklore. Using these conclusions, we can make tentative assumptions about broader Russian culture



Our Methods

- The source texts Analyzed were pulled from the Project Gutenberg website. These texts were already translated into English for ease of understanding the data
- An estimated 50 different folktales were used for this project
- We will be keeping a particular eye out for specific characters and names, as well as actions (verbs)



Code



```
with open('text.txt', 'r') as file:  
    text = file.read()
```

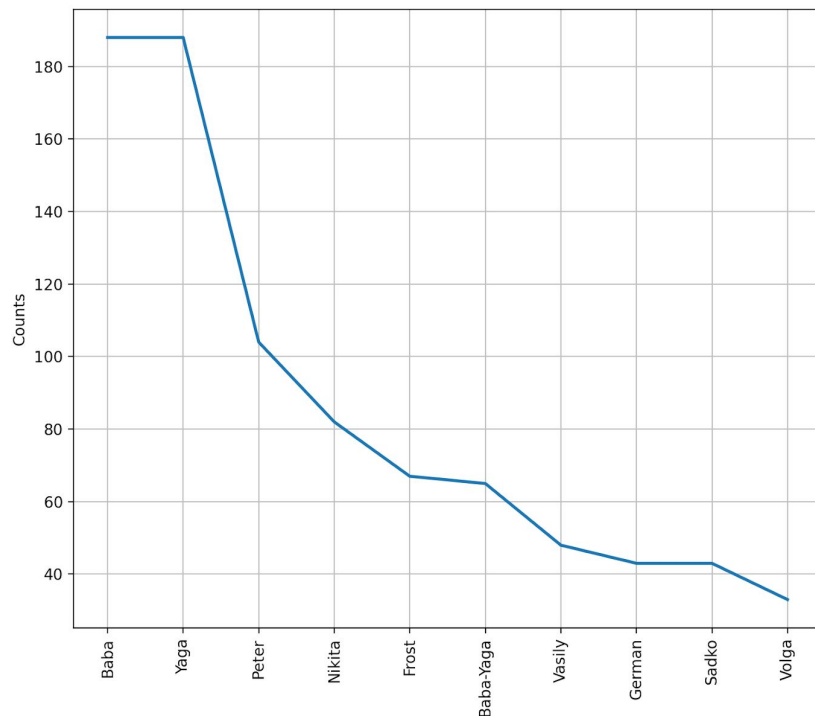
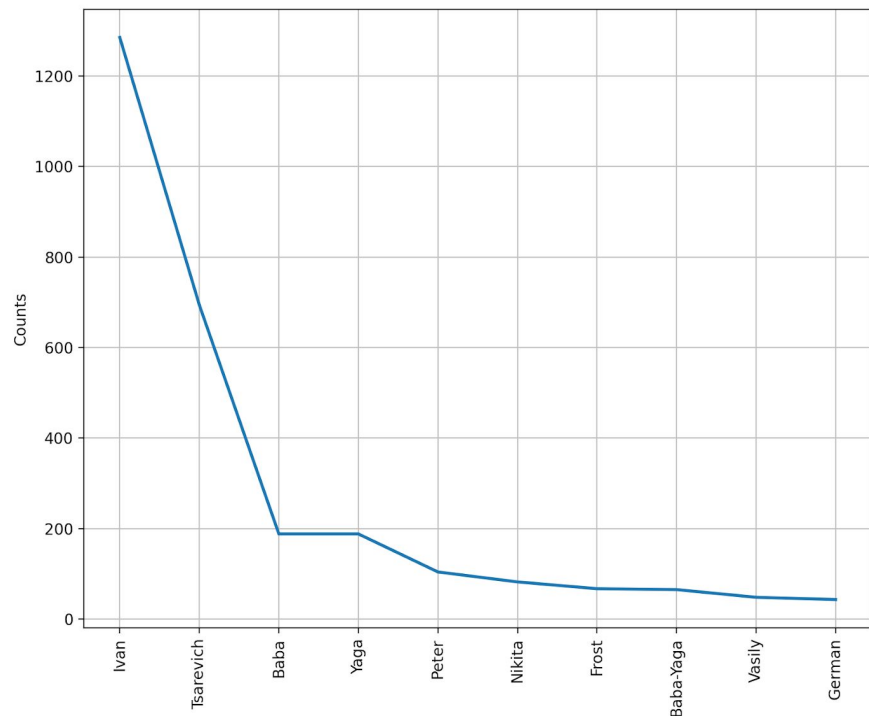
```
with open('verbs.txt', 'r') as file:  
    verbs = file.read()
```

```
with open('russiancreatures.txt', 'r') as file:  
    russiancreatures = file.read()
```

```
stops = stopwords.words('english')
stops.append('would')
stopverbs = ['is', 'was', 'said', 'says', 'say', 'has', 'have', 'had', 'or']
punctuation = string.punctuation + '""'
word_tokens = [token for token in word_tokenize(
    text) if token not in punctuation and token not in stops]
verb_tokens = {token for token in word_tokenize(verbs)}
russian_creatures_tokens = {token.lower()
    for token in word_tokenize(russiancreatures)}
```

```
character_occurrences = [word for word in word_tokens
|   |   |   |   |   |   | if word.lower() in russian_creatures_tokens]
without_ivan_and_tsarevich = [
|   word for word in character_occurrences if word != 'Ivan'
|   and word != 'Tsarevich']
fd = nltk.FreqDist(character_occurrences)
wit_fd = nltk.FreqDist(without_ivan_and_tsarevich)
fd.plot(10)
wit_fd.plot(10)
```

Character Occurrences




```
Ivan_occurrences = [word_tokens[word] for word in range(
    1, len(word_tokens)) if word_tokens[word - 1].lower()
    == 'ivan' and word_tokens[word] in verb_tokens and
    word_tokens[word] not in stopverbs]
fd = nltk.FreqDist(Ivan_occurrences)
fd.plot(10)
```

Data

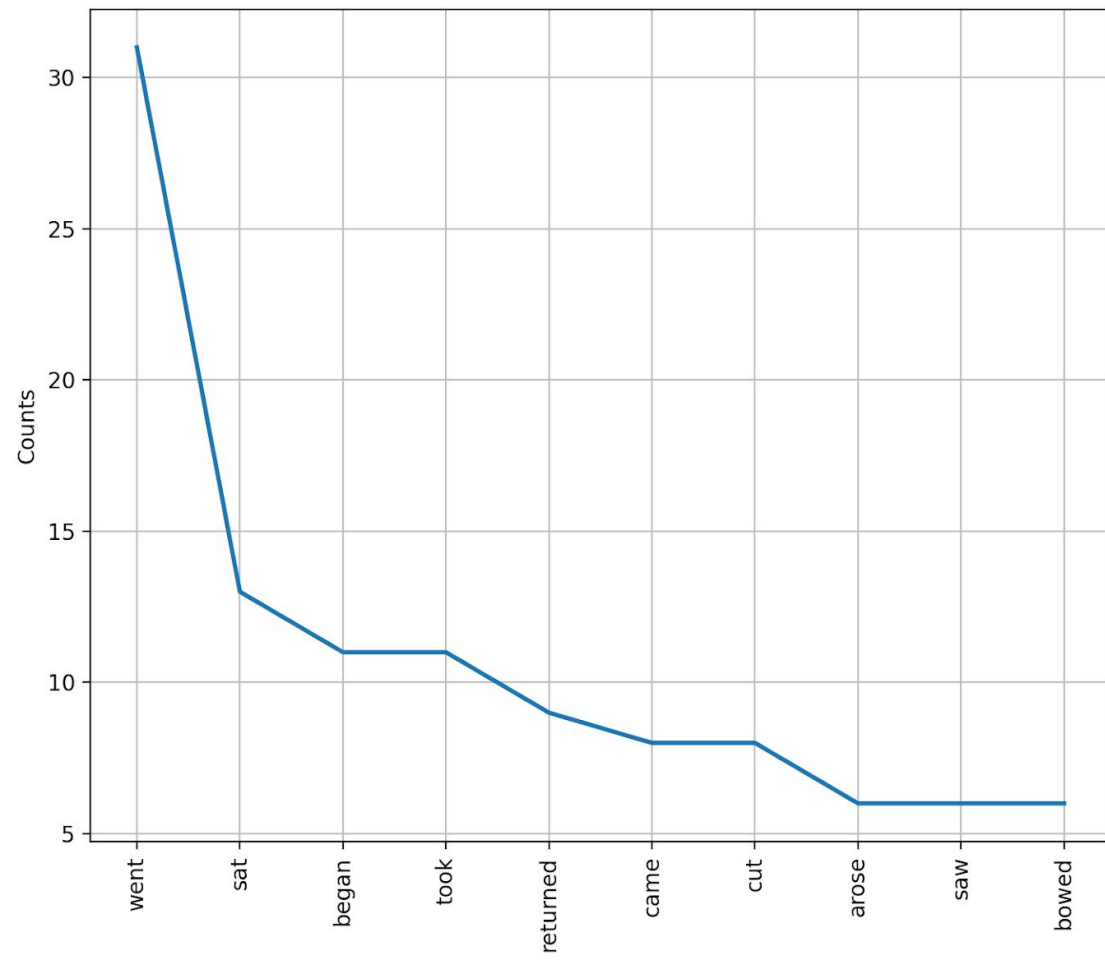


Character Overview

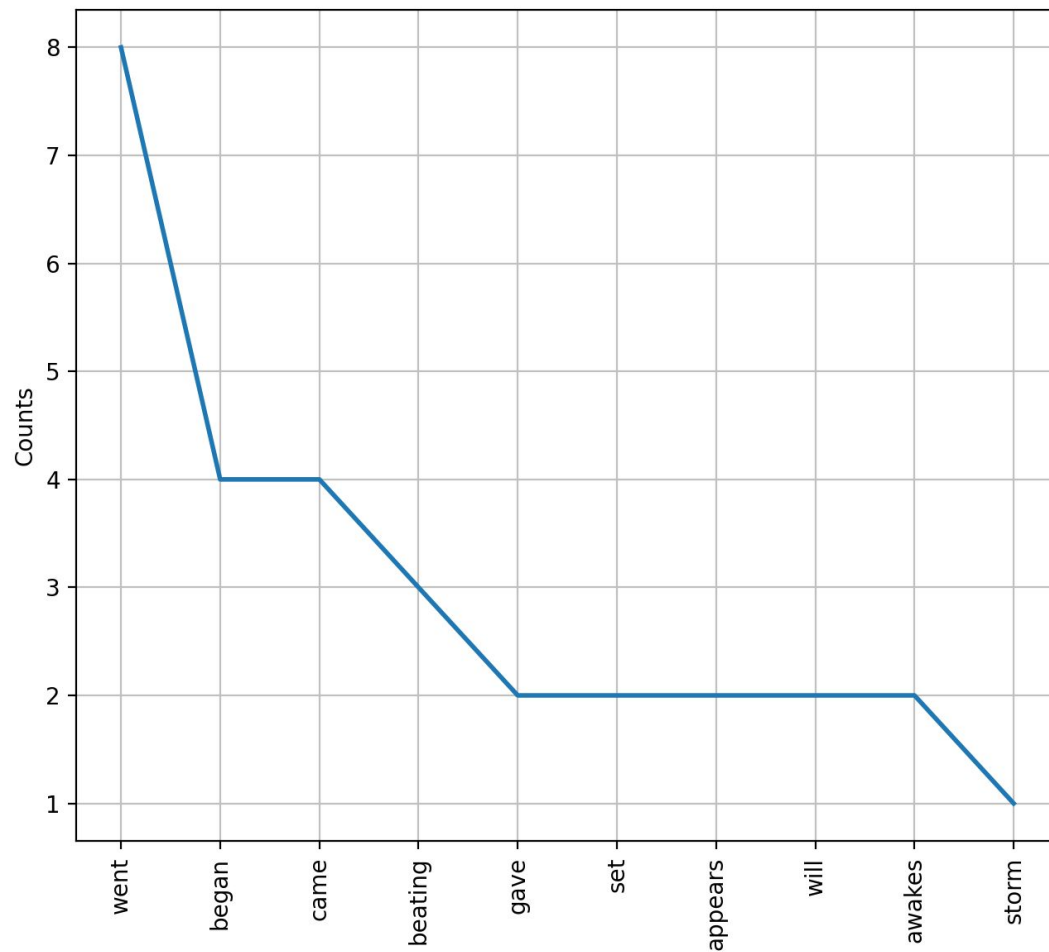
- Ivan
- Baba Yaga
- Peter (Pyotr)
- Nikita
- Frost (Ded Moroz)
- Vasily



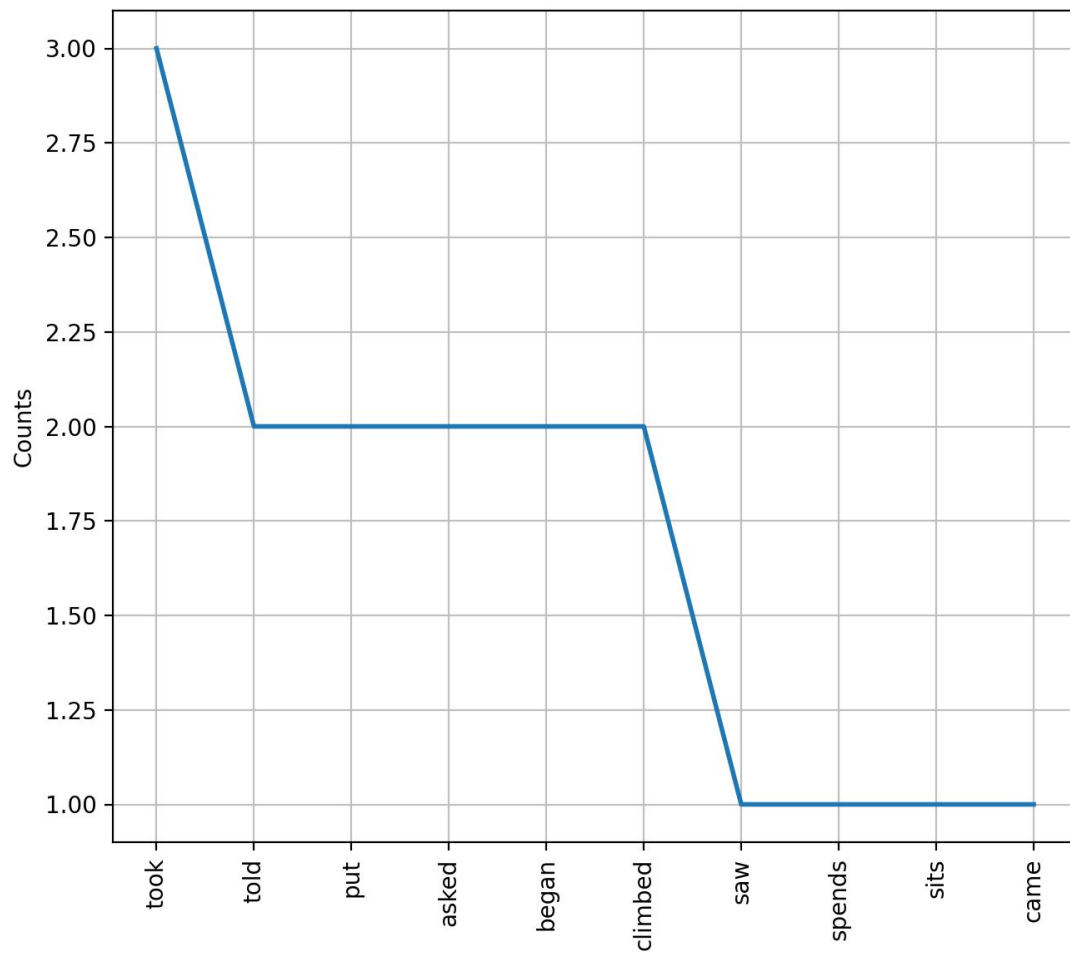
Ivan Graph



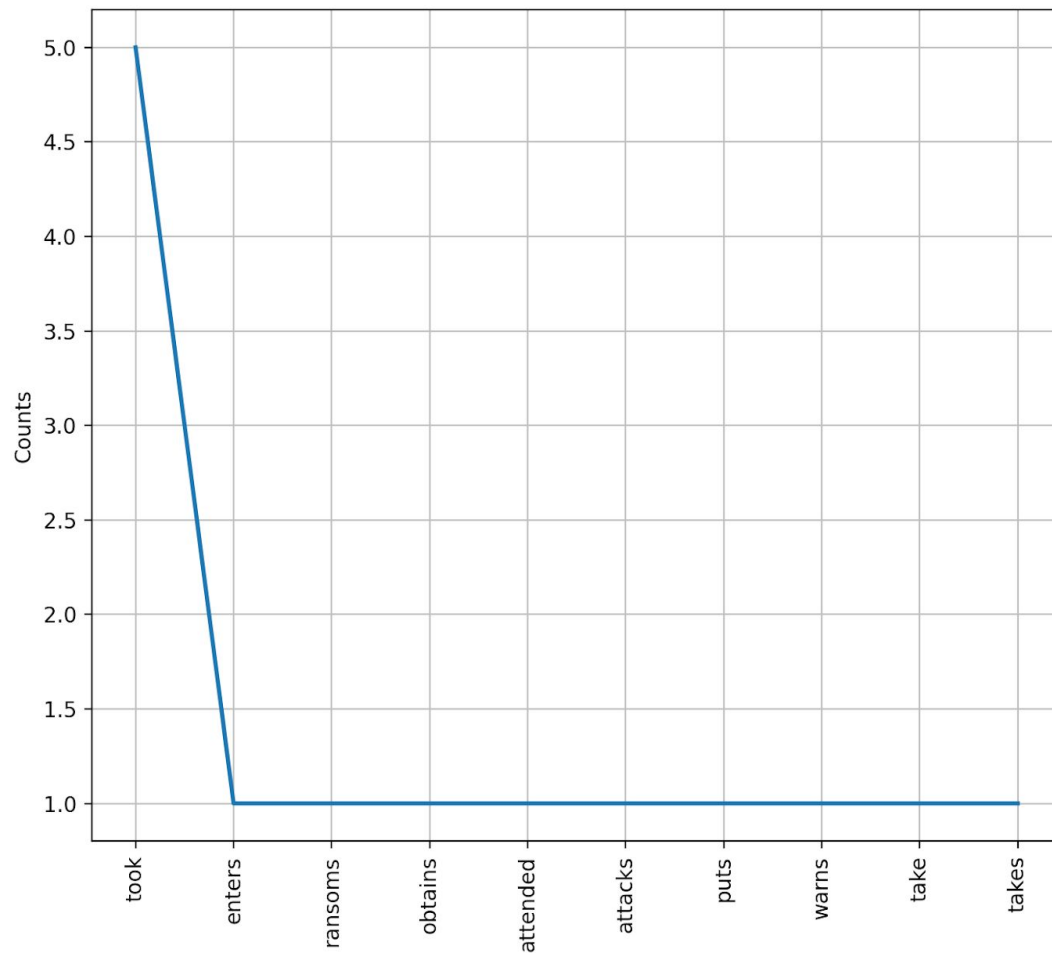
Baba Yaga Graph



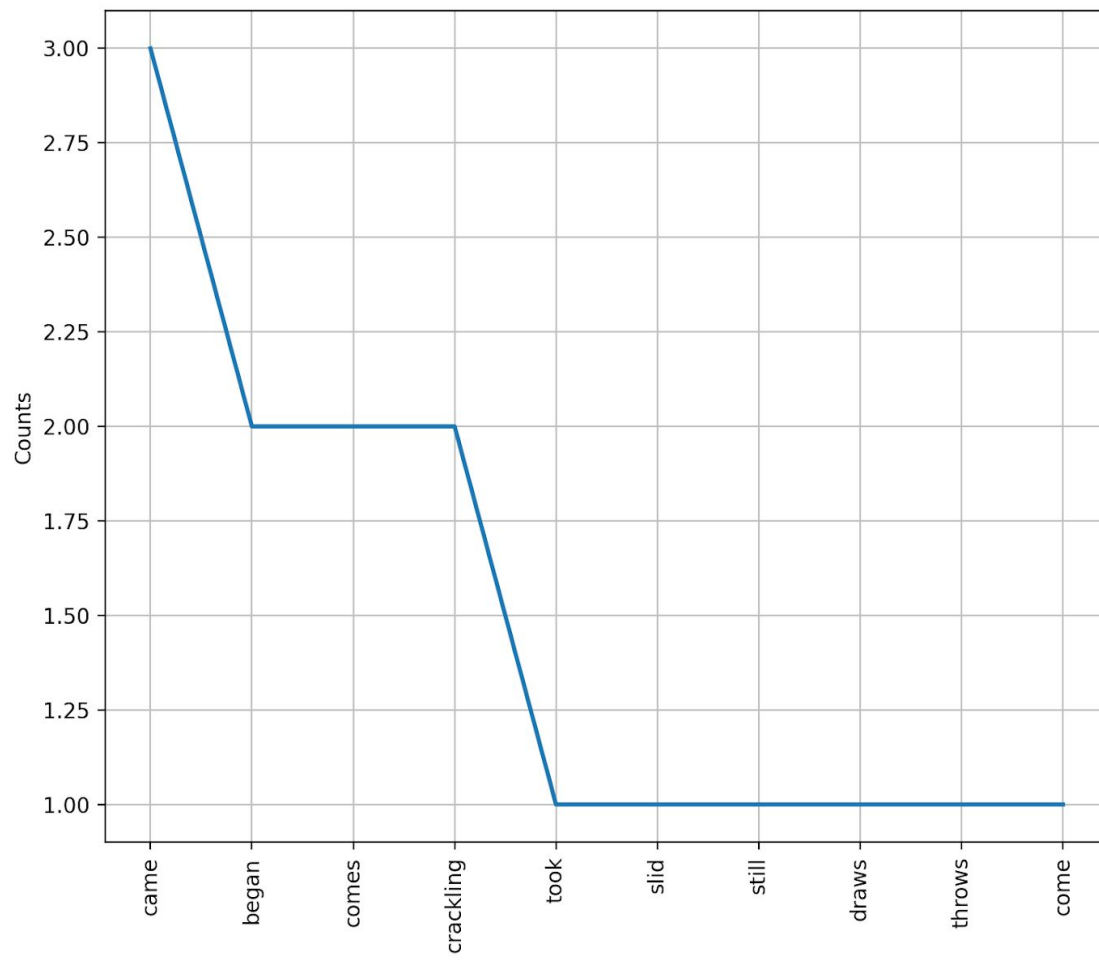
Peter Graph



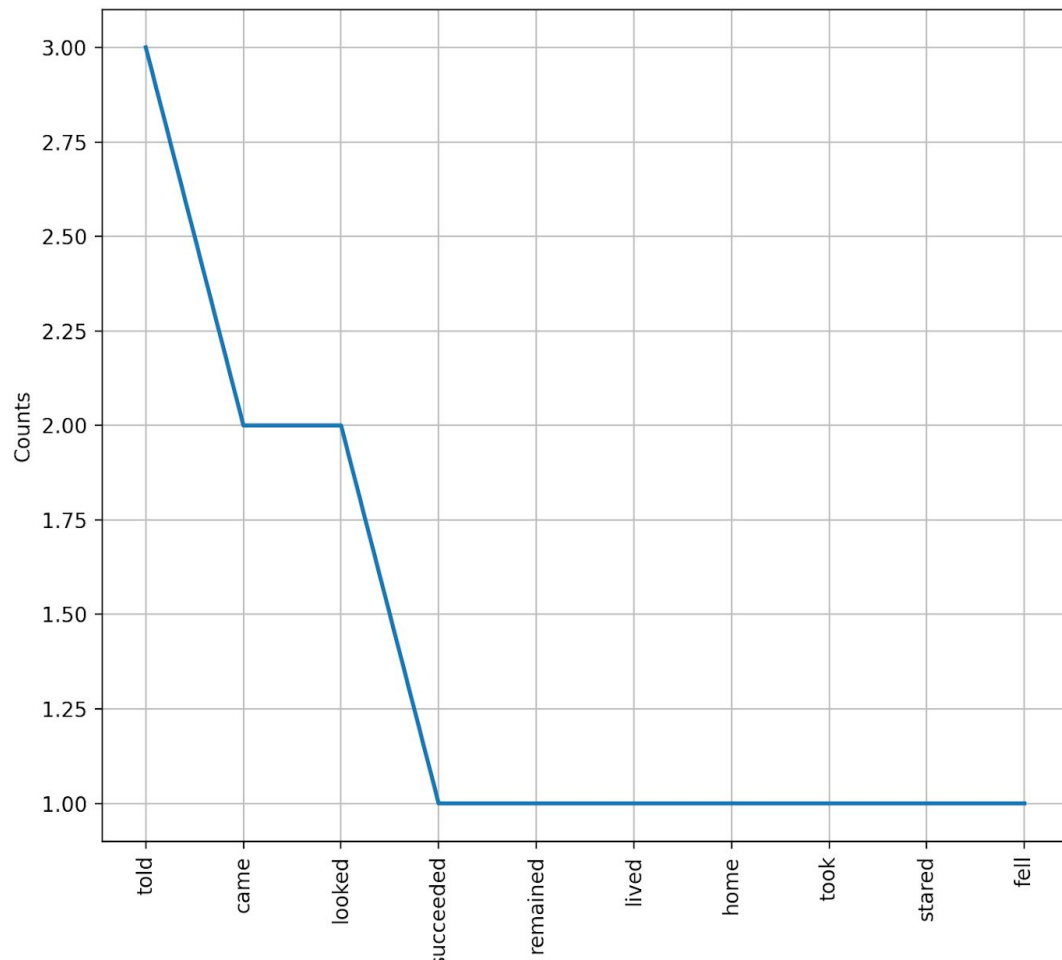
Nikita Graph



Frost Graph



Vasily Graph



Analysis



Big Ideas



- Use of generic “John Doe” names
- Prevalence of male names over female names
- Higher prevalence of Baba Yaga over Ded Moroz
- Translation of names

Name Analysis

- Ivan Tsarevich and Pyotr
 - Staple names in Russian folklore
 - Commentary on setting of folklore
 - No female equivalents
- Parallels between name meanings and roles in storyline

Male vs Female Occurrences

- Besides “Baba Yaga”, no female names in 10 most frequent
 - Not a positive character
 - Similar to Ded Moroz, deemed more “evil”
- Patriarchal society
- Epic stories required a male hero

Website



Website Link

<https://sites.google.com/email.wm.edu/thecharactersofrussianfolklore/home?authuser=1>