graph_stats

January 10, 2023

[7]:
```
%%capture
%pip install seaborn
```

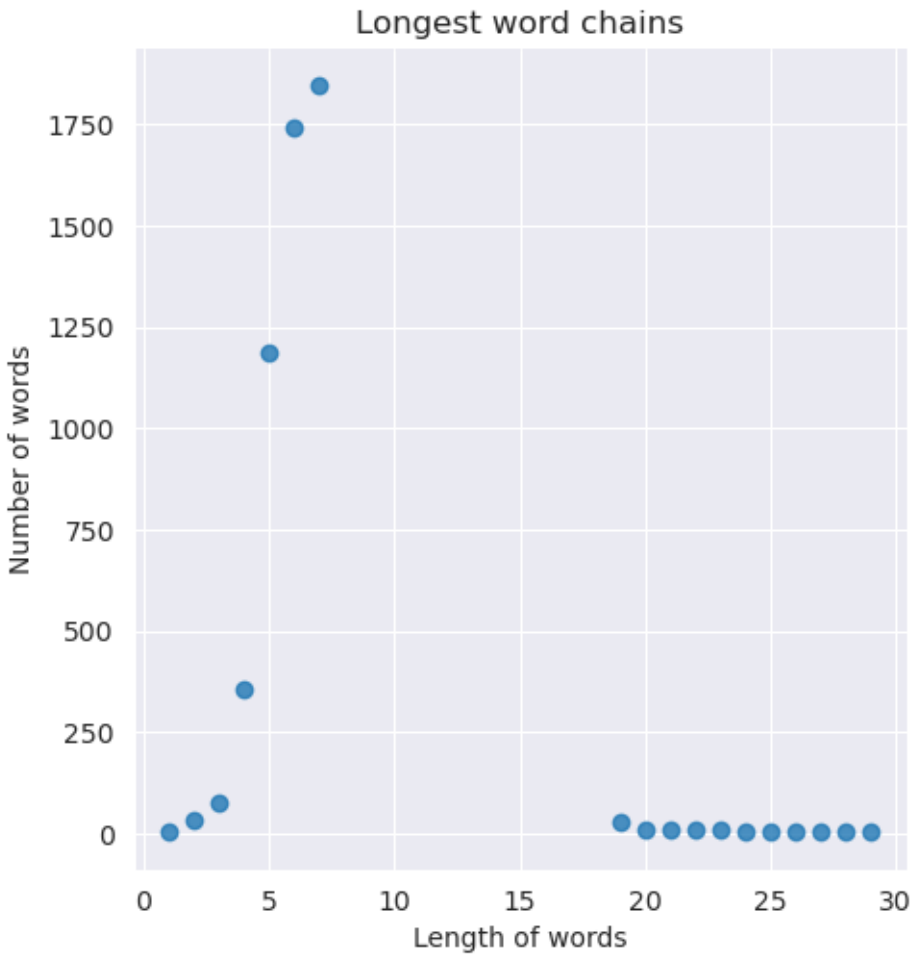
[8]:
```
import re
import sys
import seaborn as sns
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

sns.set_style("darkgrid")

class Chain():
    def __init__(self, length, words) -> None:
        self.length = length
        self.words = words
        self.size = len(words)
```

[9]:
```
chains = []
with open("../Server/longest.txt") as f:
    lines = f.readlines()
    for idx, line in enumerate(lines):
        if line.startswith("Longest"):
            length = int(re.search(r"\d+", line)[0])
            chain = Chain(length, lines[idx+2].strip().split(" -> "))
            chains.append(chain)
```

[10]:
```
df = pd.DataFrame([{"length": chain.length, "words": chain.words, "size": chain.
 ↪size} for chain in chains])
```

[11]:
```
sns.lmplot(x="length", y="size", data=df, fit_reg=False)
plt.xlabel("Length of words")
plt.ylabel("Number of words")
plt.title("Longest word chains")
```

[11]:
```
Text(0.5, 1.0, 'Longest word chains')
```



[11]: Text(0.5, 1.0, 'Longest word chains')