

# 1) Plotting the Classics

Vitor Kamada

December 2019

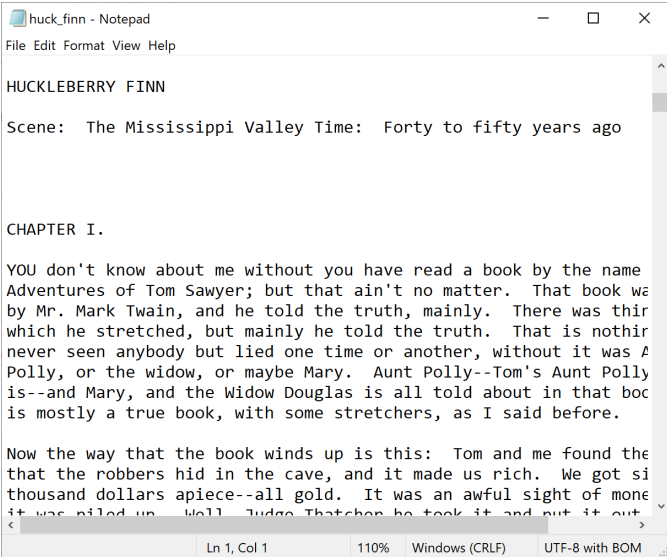
Tables, Graphics, and Figures from

**Computational and Inferential Thinking:  
The Foundations of Data Science**

Adhikari & DeNero (2019): Ch 1. Data Science

<https://www.inferentialthinking.com/chapters/01/what-is-data-science.html>

# The Adventures of Huckleberry Finn by Mark Twain



huck\_finn - Notepad

File Edit Format View Help

HUCKLEBERRY FINN

Scene: The Mississippi Valley Time: Forty to fifty years ago

CHAPTER I.

YOU don't know about me without you have read a book by the name Adventures of Tom Sawyer; but that ain't no matter. That book was by Mr. Mark Twain, and he told the truth, mainly. There was things which he stretched, but mainly he told the truth. That is nothin' never seen anybody but lied one time or another, without it was Aunt Polly, or the widow, or maybe Mary. Aunt Polly--Tom's Aunt Polly is--and Mary, and the Widow Douglas is all told about in that book; but it is mostly a true book, with some stretchers, as I said before.

Now the way that the book winds up is this: Tom and me found the place that the robbers hid in the cave, and it made us rich. We got sixteen thousand dollars apiece--all gold. It was an awful sight of money when it was piled up. Well, Judge Thatcher he took it and put it out,

Ln 1, Col 1      110%      Windows (CRLF)      UTF-8 with BOM

# Load the Book

```
from urllib.request import urlopen
import re
def read_url(url):
    return re.sub('\s+', ' ', urlopen(url).read().decode())
```

```
huck_finn_url = 'https://www.inferentialthinking.com/data/huck\_finn.txt'
huck_finn_text = read_url(huck_finn_url)
huck_finn_chapters = huck_finn_text.split('CHAPTER ')[44:]
```

**Jim:** central character

**Huck:** narrator

**Tom:** joins Huck and Jim after Chapter 30

# Display the Chapters of Huckleberry Finn in a Table

```
from datascience import *
```

```
Table().with_column('Chapters', huck_finn_chapters)
```

## Chapters

I. YOU don't know about me without you have read a book ...

II. WE went tiptoeing along a path amongst the trees bac ...

III. WELL, I got a good going-over in the morning from o ...

IV. WELL, three or four months run along, and it was wel ...

V. I had shut the door to. Then I turned around and ther ...

# Counts in each Chapter

```
import numpy as np

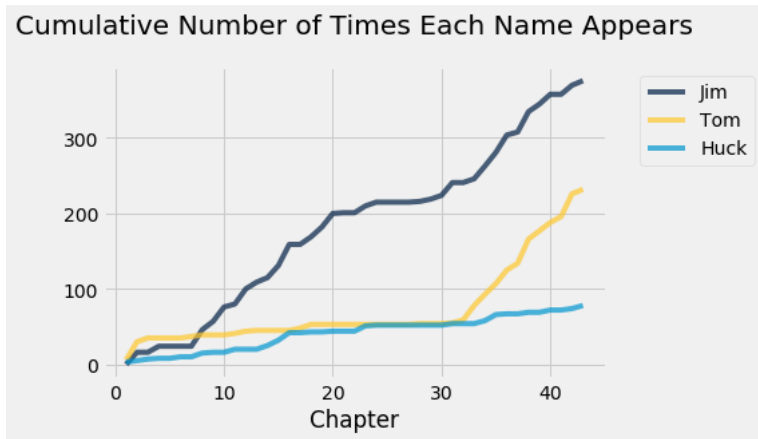
import matplotlib
matplotlib.use('Agg', warn=False)
%matplotlib inline

import matplotlib.pyplot as plots
plots.style.use('fivethirtyeight')
import warnings
warnings.simplefilter(action="ignore",
                      category=FutureWarning)

counts = Table().with_columns([
    'Jim', np.char.count(huck_finn_chapters, 'Jim'),
    'Tom', np.char.count(huck_finn_chapters, 'Tom'),
    'Huck', np.char.count(huck_finn_chapters, 'Huck')
])
```

# Cumulative Counts over Chapters

```
cum_counts = counts.cumsum().with_column('Chapter', np.arange(1, 44, 1))  
cum_counts.plot(column_for_xticks=3)  
plots.title('Cumulative Number of Times Each Name Appears', y=1.08);
```



# Little Women by Louisa May Alcott

```
path_data = 'https://www.inferentialthinking.com/data/'  
little_women_url = path_data + 'little_women.txt'  
little_women_text = read_url(little_women_url)  
little_women_chapters = little_women_text.split('CHAPTER ')[1:]  
Table().with_column('Chapters', little_women_chapters)
```

## Chapters

ONE PLAYING PILGRIMS "Christmas won't be Christmas witho ...

TWO A MERRY CHRISTMAS Jo was the first to wake in the gr ...

THREE THE LAURENCE BOY "Jo! Jo! Where are you?" cried Me ...

FOUR BURDENS "Oh, dear, how hard it does seem to take up ...

FIVE BEING NEIGHBORLY "What in the world are you going t ...

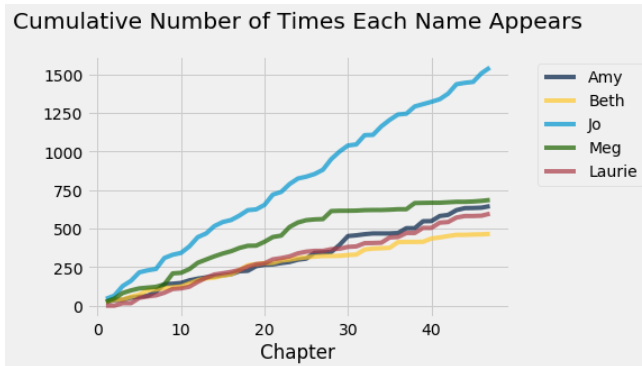


# Counts of Names and Cumulative Counts

```
counts = Table().with_columns([
    'Amy', np.char.count(little_women_chapters, 'Amy'),
    'Beth', np.char.count(little_women_chapters, 'Beth'),
    'Jo', np.char.count(little_women_chapters, 'Jo'),
    'Meg', np.char.count(little_women_chapters, 'Meg'),
    'Laurie', np.char.count(little_women_chapters, 'Laurie'),
])
```

```
cum_counts = counts.cumsum().with_column('Chapter', np.arange(1, 48,
cum_counts.plot(column_for_xticks=5)
plots.title('Cumulative Number of Times Each Name Appears', y=1.08);
```

# Jo: protagonist



**Meg, Beth, and Amy:** Jo's sisters

**Chapter 27:** Jo moves to New York alone

**Laurie:** young man who marries one of the girls in the end

# Count the Characters and Periods

```
chars_periods_little_women = Table().with_columns([
    'Little Women Chapter Length', [len(s) for s in little_women_chap
    'Number of Periods', np.char.count(little_women_chapters, '.')
])
chars_periods_little_women
```

Little Women Chapter Length	Number of Periods
21759	189
22148	188
20558	231
25526	195
23395	255

# Count the Characters and Periods

```
chars_periods_huck_finn = Table().with_columns([
    'Huck Finn Chapter Length', [len(s) for s in huck_finn_chapters],
    'Number of Periods', np.char.count(huck_finn_chapters, '.')
])
chars_periods_huck_finn
```

Huck Finn Chapter Length	Number of Periods
7026	66
11982	117
8529	72
6799	84
8166	91

# 100 to 150: Characters per Period

```
plots.figure(figsize=(6, 6))
plots.scatter(chars_periods_huck_finn.column(1),
             chars_periods_huck_finn.column(0),
             color='darkblue')
plots.scatter(chars_periods_little_women.column(1),
             chars_periods_little_women.column(0),
             color='gold')
plots.xlabel('Number of periods in chapter')
plots.ylabel('Number of characters in chapter');
```

**Twitter (140-character limit)**

