

Algorithms and Data Structures for Data Science

lab_parsing

CS 277
Brad Solomon

August 27, 2021



Department of Computer Science

Learning Objectives

Review fundamentals of Python I/O

Explore common string data formats (.txt and .csv)

Identifying and accounting for errors in input datasets

Python File I/O

Readable:

Writable:

Appendable:

```
1
2
3 readableFile = open('inputFile.txt', 'r')
4
5
6
7 writableFile = open('outputFile.txt', 'w')
8
9
10
11 carefulWritableFile = open('outputFile.txt', 'x')
12
13
14
15 appendableFile = open('outputFile.txt', 'a')
16
17
18
```

Python File I/O

Which approach is better?

```
1 # Approach 1
2
3 readableFile = open('inputFile.txt', 'r')
4
5
6 fileData = readableFile.read()
7
8
9 readableFile.close()
10
11
12 # Approach 2
13
14 with open('inputFile.txt', 'r') as myFile:
15     fileData = myFile.read()
16
17
18
```

Python File I/O

```
1
2
3 with open('dataFile.txt') as myFile:
4     print(myFile.read(10))
5
6
7
8 with open('dataFile.txt') as myFile:
9     print(myFile.readline())
10
11     print(myFile.readline())
12
13     print(myFile.readline())
14
15
16
17
18
```

```
1 A B C D E F G
2 1234567890
3
4 Mint Chocolate Banana
5
```

dataFile.txt

Python File I/O

```
1
2 with open('dataFile.txt') as myFile:
3     line = myFile.readline()
4
5     while line:
6         print(line)
7         line = myFile.readline()
8
9 with open('dataFile.txt') as myFile:
10     print(myFile.readlines())
11
12
13
14
15
16
17
18
```

```
1 A B C D E F G
2 1234567890
3
4 Mint Chocolate Banana
5
```

dataFile.txt

Python File I/O

Whats wrong here:

```
1  
2 with open('dataFile2.txt','w') as myFile:  
3     data = myFile.readlines()  
4  
5 for line in data:  
6     print(line)  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18
```

Python File I/O

What will get written:

```
1 with open('dataFile2.txt','w') as myFile:  
2     myFile.write("Hello!")  
3  
4     x = 2  
5     y = 3  
6     z = 5  
7     myFile.write("{} , {} , {}".format(x,y,z))  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18
```


Python File I/O

split():

strip():

```
1 text = "1,2,3,,4,"
2
3 print(len(text.split(',')))
4
5 print(text.split(',')
6
7
8
9 text = " a b c \n "
10
11 print("'" + text.strip() + "'")
12
13
14
15
16
17
18
```

Coding the lab

- 1) Complete parts 1 and 2 — test using provided files (or your own)
- 2) Consider how to parse and store grades (.csv) and roster (.txt)
- 3) Consider how to identify valid students (by roster or grade)