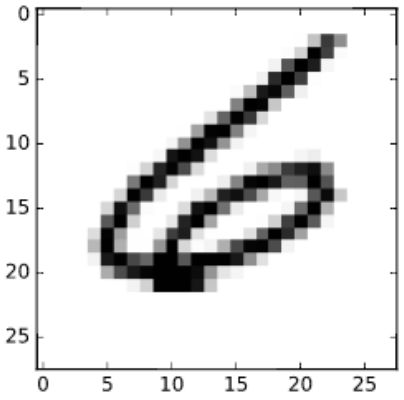
Importing Data in Python (Part 1)

November-15-17 10:14 AM

```
#Ch 1 Introduction and flat files
# Open a file: file
file = open('moby_dick.txt', mode = 'r')

# Print it
print(file.read ())

# Check whether file is closed
print(file.closed)

# Close file
file.close ()

# Check whether file is closed
print (file.closed)
```

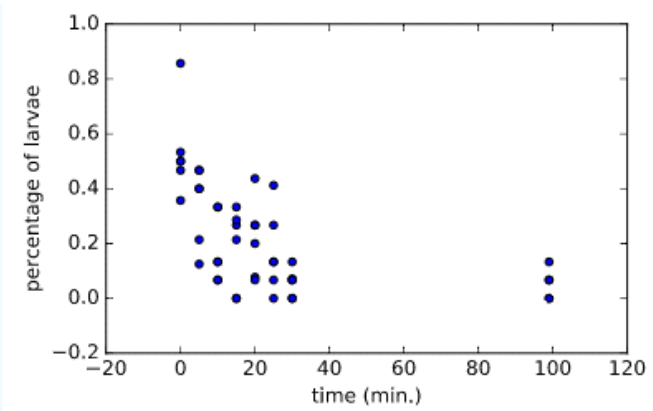
```
# Read & print the first 3 lines
with open('moby_dick.txt') as file:
    print(file.readline())
    print(file.readline())
    print(file.readline())
```

```
#Using NumPy to import flat files
# Import package
import numpy as np

# Assign filename to variable: file
file = 'digits.csv'

# Load file as array: digits
digits = np.loadtxt(file, delimiter=',')

# Print datatype of digits
print(type(digits))

# Select and reshape a row
im = digits[21, 1:]
im_sq = np.reshape(im, (28, 28))

# Plot reshaped data (matplotlib.pyplot already loaded as plt)
plt.imshow(im_sq, cmap='Greys', interpolation='nearest')
plt.show()
```




```
#Customizing your NumPy import
# Import numpy
import numpy as np

# Assign the filename: file
file = 'digits_header.txt'

# Load the data: data
data = np.loadtxt(file, delimiter="\t", skiprows=1, usecols=[0, 2])

# Print data
print(data)
```

```
#Importing different datatypes
# Assign filename: file
file = 'seaslug.txt'

# Import file: data
data = np.loadtxt(file, delimiter='\t', dtype=str)

# Print the first element of data
print(data[0])

# Import data as floats and skip the first row: data_float
data_float = np.loadtxt(file, delimiter='\t', dtype=float, skiprows=1)

# Print the 10th element of data_float
print(data_float [9])

# Plot a scatterplot of the data
plt.scatter(data_float[:, 0], data_float[:, 1])
plt.xlabel('time (min.)')
plt.ylabel('percentage of larvae')
plt.show()
```




```
#Working with mixed datatypes (2)
# Assign the filename: file
file = 'titanic.csv'

# Import file using np.recfromcsv: d
d = np.recfromcsv (file, delimiter = ',', names = True, dtype = None)

# Print out first three entries of d
print(d[:3])
```

```
#Using pandas to import flat files as DataFrames (1)
# Import pandas as pd
import pandas as pd

# Assign the filename: file
file = 'titanic.csv'

# Read the file into a DataFrame: df
df = pd.read_csv(file)

# View the head of the DataFrame
print (df.head ())
```

```
#Using pandas to import flat files as DataFrames (2)
# Assign the filename: file
file = 'digits.csv'

# Read the first 5 rows of the file into a DataFrame: data
data = pd.read_csv (file, nrows = 5, header = None)

# Build a numpy array from the DataFrame: data_array
data_array = np.array (data.values)

# Print the datatype of data_array to the shell
print(type(data_array))
```

```
#Customizing your pandas import
# Import matplotlib.pyplot as plt
import matplotlib.pyplot as plt

# Assign filename: file
file = 'titanic_corrupt.txt'

# Import file: data
data = pd.read_csv(file, sep='\t', comment='#', na_values='Nothing')

# Print the head of the DataFrame
print(data.head())

# Plot 'Age' variable in a histogram
pd.DataFrame.hist(data[['Age']])
plt.xlabel('Age (years)')
plt.ylabel('count')
plt.show()
```