Pandas Assignment

import pandas as pd #Importing the pandas library

country\_d,=pd.read\_html("<https://github.com/jennybc/gapminder/blob/master/dataraw/01_pop.tsv>") # Getting data from the repository

 print(country\_d.shape) # To get number of rows and columns

print(country\_d.head(100)) # To print first 100 rows

print(country\_d.tail(100)) # To print bottom 100 rows

[country\_d.info](http://c_d.info/)()

country\_d.to\_csv("con\_pop.csv", index=False) # To save the data into the file

country\_p = pd.read\_csv("con\_pop.csv", usecols=['country', 'year', 'pop'])

[country\_p.info](http://c_p.info/)() # To read a file and take only the required

p = country\_p.groupby(by=['year'])['pop'].sum()

print(p) #To calculate the total population by year of all countries.

UNIX Commands

1. ls -al | wc –l #To find files in current directory
2. grep -l \_\_init\_\_ \*.py # to find all the files with init in the file
3. file1.html -> <head>

file2.txt -> <head> -- you shouldnt print this

find . -name "\*.html" | xargs grep -in "<head>" \*.html # to find files that have head in file

1. delete\_all.sh /tmp/newDir

vi delete\_all.sh

if test -d $1

then

cd $1

rm -R \*

echo"deleted"

else

echo"directory not found"

fi

. delete\_all.sh /tmp/newdir # a script to delete all the content in the directory