

Assignment_2

gzahn

December 9, 2018

Assignment 2

In this assignment, you will use the command-line of your computer to:

- Use relative file paths that are portable between computers sharing a Git repository
- Demonstrate the use of “wildcards” to perform actions on groups of files simultaneously
- Demonstrate redirection of output to files (“>” and “>>”)
- Use the pipe “|” to chain output of one program to input of another
- Use the programs “wc” “ls” “sort” “uniq” “head” “tail” “find” “mkdir”
- Create a “for-loop” that combines all of the above skills to run an existing script on many files and summarize the output

This assignment will be completed using the Shell terminal in the “Data_Course” directory.

All file paths should be relative, starting from the Assignment_2 directory!! (where you found this file)

For credit...

1. Upload a .txt file to Canvas that has the code you wrote to complete each of the following tasks. Your plaintext file should be numbered (1-6) and the code for each task should be on the line(s) below the task number!
2. You will also be scored on whether new files created as part of this assignment are in their proper place in your own repository. That means you will need to “stage,” “commit,” and “push” all changes in your local repository to your GitHub repository.

Complete the following 6 tasks:

1. For each of the filenames that begin with “M” in the directory “../Data/data-shell/data/pdb” list the number of lines in each file.
2. Redirect the output from the previous task to a new file named “m_lines.txt” in “../Data_Course_YOUR_LAST_NAME/Assignments/Assignment_2/” (note that all new files in your personal repository will be checked by the instructor...also, you will need to create that directory)
3. For the file “animals.txt” in “../Data/data-shell/data/” get an alphabetical list of the unique animal types that were observed and redirect this list to a new file called “unique_animals.txt” in “../Data_Course_YOUR_LAST_NAME/Assignments/Assignment_2/”
4. In the directory “../Data/” there is a file named “BioLog_Plate_Data.csv” Write a command that gives a list of the unique carbon substrates the biolog plate measured, and which saves this list as a new file named “unique_C-substrates.txt” in “../Data_Course_YOUR_LAST_NAME/Assignments/Assignment_2/”

5. Write a command that finds the longest .txt file (by number of lines) in the directory “../Data/data-shell/data/”

Hint: wc, sort, tail, head

6. In the directory “../Data/data-shell/many_files/” there are a large number of subdirectories. Buried annoyingly inside these subdirectories are some .txt files that all contain a list of numbers (one integer per line). There is a shell script in the directory “../Data/data-shell/scripts” called “sum_lines.sh”

Your task is to run this script on **ALL** the .txt files buried within the “many_files/” directory to output the sums of all the lines in each file. This output (on all those files) should be saved in a single file named “file_sums.txt” in “../Data_Course_YOUR_LAST_NAME/Assignments/Assignment_2/”

This is a tough one. Remember that to run a shell script you need to type “bash PATH_TO_SCRIPT”

Hint: use find -exec and/or a for-loop

find -exec reminder:

```
find ../../Data/data-shell/data/animal-counts/ -type f -exec head -1 {} \;
```

```
## 2012-11-05,deer,5
```

for-loop reminder:

```
# this will cycle through the files in ../../Data/ with *.txt in the filename  
# with each new value, it will echo the filename to the terminal,  
# and then give the first line of the file
```

```
for i in ../../Data/*.txt  
do  
  echo $i  
  head -1 $i  
done
```

```
## ../../Data/MOMS-PI_16S_STIRRUPS_profiles_POP1.txt  
## SampleID Taxa Threshold-Status No_of_Reads Percentage_Abundance Avg._Score  
## ../../Data/MOMS-PI_README.txt  
## #####  
## ../../Data/MOMS-PI_subject_kit_sample_mapping_POP1.txt  
## SubjectID KitID KitType SampleID SampleType BodySite VisitNumber  
## ../../Data/Soil_Predators_README.txt  
## Data from: Testing the validity of functional response models using molecular gut content analysis f  
## ../../Data/thatch_ant_info.txt  
##  
## ../../Data/trait_data_README.txt  
## When using this data, please cite the original article:
```

On one line, it would look like this:

```
for i in ../../Data/*.txt;do echo $i;head -1 $i;done
```