ls -lhclear Welcome to The Carpentries Etherpad!

This pad is synchronized as you type, so that everyone viewing this page sees the same text. This allows you to collaborate seamlessly on documents.

Use of this service is restricted to members of The Carpentries community; this is not for general purpose use (for that, try https://etherpad.wikimedia.org).

Users are expected to follow our code of conduct: https://docs.carpentries.org/topic_folders/policies/code-of-conduct.html

---

Good afternoon, or good morning!!

Find the Workshop website and schedule here: https://smithsonianworkshops.github.io/2021-06-21-smithsonian-online/

Welcome to Day 1: The Unix Shell
Please add your name to the top-right of the etherpad window.

Before the workshop begins, please see the Setup page (https://swcarpentry.github.io/shell-novice/setup.html) an ensure you have: (1) the data file downloaded and moved to your Desktop and (2) either the Terminal app or some Terminal emulator (such as "Git for Windows")

The Unix Shell: Lesson page can be found here: https://swcarpentry.github.io/shell-novice/

ATTENDANCE DAY 1
Add your name, unit, and your favorite type of cookie (or other sweet treat)
Sammy Arnold, REU-UMCES, Chocolate Chip
Rebecca Dikow, OCIO Data Science Lab, peanut butter cookie
Maddy Bursell, OCIO Data Science Lab, Snickerdoodle
Jennifer Spillane, Data Science Lab, oatmeal chocolate chip
Arsh Suri, SCBI, Chocolate Chip Cookie + Milk
Alex R, OCIO DSL/SCBI, Professional (Oatmeal Chocolate Chip) Cookie Monster
avery j, Archives of American Art, chocolate chip cookie
Fiona Skerrett, STRI, hard sugar cookies (with the royal icing)
Dalila Lara , Data Science Lab, Choclate Chip
Chandini Palem, DPO, Chocolate Chip
Rachel Miller, NMNH, Chocolate chip cookies
Christina Balentine, NMNH, Snickerdoodle
Elizabeth Gutierrez, Smithsonian Astrophysical Observatory, Oreo
Milena Nunes, NHRE, chocolate chip cookies
Mike Trizna, OCIO Data Science Lab, mint oreo
Jenna Ekwealor, OCIO Data Science Lab, peanut butter cookie
Carlos Arias, OCIO Data Sciece Lab, Cocossete.

Paula Pappalardo, IZ-NMNH, peanut butter cookie
Marina Ellis, NMNH, blackberry sorbet (I'm not a sweets person)
Bella Schrader, NMNH Data Science Lab, Chocolate Chip cookies
Taaj Clark, Smithsonian Conservation of Genomics, Key Lime Pie
Karen Holm SCBI/NZP chocolate chip
Christelle Inema, Intern at Data Science lab, Oreo
Felipe Mello, STRI, Butter cookies!
Melissa Anderson, Pinwheel

- Episode 1: Introducing the Shell

Shared notes:
ls [listing]
pwd [print working directory]
ls -F [listing + flag]
clear
man ls
ls --help
ls -l
ls -h
ls -lh
ls -F _ [list directory]
cd _ [change directory]
cd .. [move up one level in the file system]
ls -F -a [show all directories including hidden ones / a dot before the file name means it is hidden]
cd . [current directory]
mkdir _ [make directoru]
ls -FR [recursive flagging]
nano _.txt [creates + opens text file]
mv _ _ [move to new location and/or name]
cp _ _ [copy allows you to copy a file to a new location without the risk of losing the original file]

- Episode 2: Navigating Files and Directories

Shared notes:

Exercise: Exploring More ls Flags

- You can also use two options at the same time. What does the command ls do when used with the -l option? What about if you use both the -l and the -h option?
- Some of its output is about properties that we do not cover in this lesson (such as file permissions and ownership), but the rest should be useful nevertheless.

Exercise: Absolute vs. Relative Paths

- Starting from /Users/amanda/data, which of the following commands could Amanda use to navigate to her home directory, which is /Users/amanda?
- Put an x next to command(s) you think would work:
- cd .x

- cd /
- cd /home/amandaxcskn
- cd ../..x
- cd ~xxxxxxxxx
- cd home
- cd ~/data/..
- cd xxxxxxxxxx
- cd .. xxxxxxxxx

- Episode 3: Working With Files and Directories

Shared notes:

Exercise: Moving Files to a New Folder

- After running the following commands, Jamie realizes that she put the files sucrose.dat and maltose.dat into the wrong folder. The files should have been placed in the raw folder.
- $ ls -F
- Output:
- analyzed/ raw/
- $ ls -F analyzed
- Output:
- fructose.dat glucose.dat maltose.dat sucrose.dat
- $ cd analyzed
- Fill in the blanks to move these files to the raw/ folder (i.e. the one she forgot to put them in)
- $ mv sucrose.dat maltose.dat ____/____

- $ mv sucrose.dat maltose.dat ../raw

DAY 1 FEEDBACK: https://forms.gle/yU2F5Re6AXyqmZ6Y8

----------WELCOME BACK! DAY 2----------

Good afternoon, or good morning! Today we'll be going over Data Organization in Spreadsheets and an Introduction to using OpenRefine!

Our Workshop Schedule/webpage: https://smithsonianworkshops.github.io/2021-06-21-smithsonian-online/

Your instructor today (he/him/his): https://datascience.si.edu/people/alex-robillard

If you havent already done so, please download the Spreadsheet needed for part 1:
Download this data file to your computer: https://ndownloader.figshare.com/files/2252083

If you havent already done so please download OpenRefine: http://openrefine.org/download.html.

If you're having difficulty installing it, please follow this link:
https://github.com/SmithsonianWorkshops/binders/tree/open_refine34

ATTENDANCE DAY 2
Add your name, unit, and your favorite national or state park (Green Space)?
Maddy Bursell, OCIO Data Science Lab, Zion National Park
Taaj Clark, Smithsonian Conservation Of Genomics , Yellow Stone
Sammy Arnold, REU UMCES, Acadia
Christina Balentine, NMNH, Rocky Mountain National Park
Felipe Mello, STRI, Sequoia National Park.
Fiona Skerrett, STRI, Mt. Rainier
Chandini Palem, Smokey Mountains
Jennifer Spillane, OCIO Data Science Lab, North Cascades National Park
Rachel Miller, NMNH, Gorges National Park NC
avery j., Archives of American Art, Kristenbosch National Garden
Dalila Lara , OCIO Data Science Lab, Yellow Stone National Park
Karen HOlm, My own 45 acres, Rocky Mountain National Park
Vanessa González, NMNH Global Genome Initiative, Joshua Tree National Park
Elizabeth Gutiérrez, SAO, Grand Canyon
Milena Nunes, NHRE, Great Falls National park
Melissa Anderson, Grandfather Mtn. State Park in NC
Christelle Inema, Intern Data Science Lab, Zion National Park and Nyungwe National Park
Isabella Schrader, NMNH NHRE Data Science Intern, New River Gorge National Park
Marina Ellis, NMNH, Great Smoky Mountains National Park

Challenge- Why would the TSV selection change the preview?

Separates values by tab as opposed to comma
Isn't recognizing the commas that are defining the breaks in the spreadsheet

http://openrefine.org/download.html
Aqui!!

END OF DAY TWO FEEDBACK FORM: https://forms.gle/B2pFctFAzNQGmN4C6

ATTENDANCE DAY 3
Add your name, unit, and your favorite kind of ice cream
Rebecca Dikow, OCIO Data Science Lab, strawberry
Mike Trizna, OCIO Data Science Lab, pistachio
Carlos Arias, OCIO Data Science Lab, Chocolate
Jennifer Spillane, OCIO Data Science Lab, basil
Maddy Bursell, OCIO DataScience Lab, Mint Chocolate Chip
Taaj Clark, Smithsonian Conservation Of Genomics, Cookies and Cream
Felipe Mello, STRI, Guanabana
Rachel Miller, NMNH, Ben and Jerry's Phish Food

Christina Balentine, NMNH, Ben & Jerry's Half Baked
Chandini Palem, DPO, Cookies & Creame
Meissa Anderson, pistachio
Sammy Arnold, REU-UMCES, Cherry Vanilla
Dalila Lara, OCIO Dara Science Lab, Cookies and Cream
Milena Nunes, NHRE, Chocolate
Elizabeth Gutiérrez, SAO, cookie dough
Jenna Ekwealot, OCIO Data Science Lab, coffee
avery, AAA, vanilla
Paula Pappalardo, IZ-NMNH, dulce de leche (caramel)
Christelle Inema, shortcake
Karen Holm SCBI chocolate chip cookie dough

Python setup instructions:
https://datacarpentry.org/python-ecology-lesson/setup.html

Create a list of unique site ID's ("plot_id") found in the surveys data. Call it site_names.

- How many unique sites are there in the data? How many unique species are in the data?
- What is the difference between len(site_names) and surveys_df['plot_id'].nunique()?

1. How many recorded individuals are female F and how many male M?

2. What happens when you group by two columns using the following syntax and then calculate mean values?

    - grouped_data2 = surveys_df.groupby(['plot_id', 'sex'])
    - grouped_data2.mean()

3. Summarize weight values for each site in your data. HINT: you can use the following syntax to only create summary statistics for one column in your data.by_site['weight'].describe()

1. Create a plot of average weight across all species per site.
2. Create a plot of total males versus total females for the entire dataset.

END OF DAY 3 FEEDBACK FORM: https://forms.gle/RSuDMgAT5vvdm5tb7

----------WELCOME BACK! DAY 4----------

ATTENDANCE DAY 4
Add your name, unit, and one skill you've learned so far that you plan on using during your internship
avery, AAA, Unix/Bash
Sammy Arnold, REU-UMCES, python, notebooks
Jenna Ekwealor, OCIO Data Science Lab, Python making copies vs
Melissa Anderson, python
Christelle Inema, Data Science lab, notebooks
Chandini Palem, DPO, Jupyter Notebooks
Felipe Mello, STRI, OpenRefine will be useful!
Fiona Skerrett, STRI, Utilizing terminal
Jennifer Spillane, OCIO Data Science Lab, manipulating pandas dataframes

Dalila Lara, OCIO Data Science Lab, Terminal management
Taaj Clark, Smithsonian Conservation Of Genomics, plotting data
Karen Holm SCBI/NZP Jupyter notebook and python
Maddy Bursell, OCIO Data Science Lab, jupyter notebooks
Christina Balentine, NMNH, jupyter notebooks
Milena Nunes, NHRE, Terminal
Elizabeth Gutiérrez, SAO, python

https://ndownloader.figshare.com/files/10717177

1. What happens when you execute:
    - surveys_df[0:1]
    - surveys_df[:4]
    - surveys_df[:-1]
2. What happens when you call:
    - surveys_df.iloc[0:4, 1:4]
    - surveys_df.loc[0:4, 1:4]
- How are the two commands different?

1. Select a subset of rows in the surveys_df DataFrame that contain data from the year 1999 and that contain weight values less than or equal to 8. How many rows did you end up with? What did your neighbor get?

5 rows++++

https://ndownloader.figshare.com/files/3299483

conda install -c conda-forge plotnine

In case this is useful for someone, while trying to install plotnine I got a message to update conda:
$ conda update -n base -c defaults conda
It does ask you directly if you want to update after you run the previous code to install plotnine (if your conda is not very updated I imagine)

FEEDBACK FORM DAY 4: https://forms.gle/qeBZXeFkdopP11Uc7

Resources for after the workshop:
For announcements of upcoming SI Carpentries workshops and events:
https://si-listserv.si.edu/cgi-bin/wa?SUBED1=CARPENTRIES-ANNOUNCE&A=1

Sign up for the Smithsonian Slack. If you already are a part of the Smithsonian Slack workspace, skip to Step #3. Go to https://smithsonian.slack.com/signup, and enter your SI email account (must be @si.edu).

Monthly lunchtime gatherings to discuss topics relating to Smithsonian Data or Software Carpentry

workshop content.
Fourth Thursday of every month from 1pm-2pm ET.
Zoom links provided via Carpentries-Announce mailing list and Slack/Teams channels.

Schedule: https://github.com/SmithsonianWorkshops/carpentries-brown-bag/wiki/Schedule
Archive of past talks: https://github.com/SmithsonianWorkshops/carpentries-brown-bag

Email addresses of Instructors:

- Mike: triznam@si.edu
- Rebecca: dikowr@si.edu
- Alex: Robillarda@si.edu
- Jenna: EkwealorJ@si.edu