

- 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

All content is publicly available under the Creative Commons Attribution License:
<https://creativecommons.org/licenses/by/4.0/>

SI Winter Carpentries

Introduction slides: <https://smithsonianworkshops.github.io/workshop-slides/#/>

Daily feedback form: <https://forms.gle/LjFaH9oZpU8WfB4C6>

SI Carpentries: <https://datascience.si.edu/carpentries>.

Library Carpentry Lessons: <https://librarycarpentry.org/lessons/>

Workshop Website: <https://smithsonianworkshops.github.io/2025-01-smithsonian-online/>

DAY 1: Icebreaker - put your name, unit, and a favorite snow day activity

- Hernán D. Capador-Barreto, STRI, not much snow in Panamá :)
- Adena B. Collens, Natural History (NMNH-IZ), crafting and staying cozy under blankets
- Sue Graves, SLA, long ago-skating, now making soup!
- Alex Lawrence, NMNH Paleo, winter wonderland walk
- Corey Schmidt, ODT, eating fresh Mac n'Cheese after shoveling the driveway
- Keri Thompson, ODT, staring out the window at the squirrels
- Richard Naples, SLA, making stew
- Jennifer Lynn Bartlett, ADS at SAO, sledding, baking
- Julio Carrión, National Zoo and Conservation Biology Institute (NZCBI). Not much snow in Ecuador.
- Silky Sullivan, OA, stay warm while staring out the window
- Timothy Nguyen, SLA, cuddle with dog and play video games
- Gabe Johnson, NMNH, making spiced cider
- Meg Phillips, NMNH, looking at all the snow on trees
- Jennifer Wodzianski, NPG, hot chocolate
- Yelena Pacheco, entomology, drinking hot chocolate
- Sonia J Rowley - NMNH-IZ - at Univeristy of Hawai'i at Manoa - sandy snow here :)
- Abi Pocasangre, OA, bird watching out the window
- Kat Cook, Product Manager - Learning Lab, OET
- Kristina Heinricy, SLA, watching movies and playing in the snow with my kid

- Camille Leal, NMNH-IZ, read a good book
- Janice Hussain, CHSDM, inside
- Corey DiPietroi, DPO, cross country skiing (aspirational, b/c i havent actually done this in years)
- Crystal Sanchez, OCIO DAMS, reading sci-fi
- Becca Stout, NMAI-NY, reading a good book with a cup of tea
- Jean Beard, OSHM, playing with my pups in the snow
- Erin Bordeaux, NMAI, sitting by the fireplace with a good book or article
- Adam Hager, HMSG, sledding with my 2 and 4 year old
- Sky diving !!!! Marc Tartaro cram oratrat
- Shruti Dube, NMNH, Snowball fight

Day 1: Tidy Data

Tidy Data Excel sheet data file download:

https://librarycarpentry.github.io/lc-spreadsheets/data/training_attendance.xlsx

SLIDES: <https://github.com/SmithsonianWorkshops/2025-01-smithsonian-online/blob/gh-pages/Carpentries%20Tidy%20Data%202025.pdf>

QUESTIONS

- How many people have used spreadsheets in their work?
- What kind of operations do you do in spreadsheets?
- Which ones do you think spreadsheets are good for?
- What are some things that you've accidentally done in a spreadsheet, or have been frustrated that you can't do easily?

Yelena: Use them often to keep track of collection data and character data (measurements, morphotypes, etc.) Frustrated with sorting data by different characters and respective rows don't match.

Erin: Use all the time. Mostly for budgets but sometimes for project work. Calculations, organizing data to assign tasks. Tried to count rows with text data (count "Yes") - don't always know the formulas (have to google for help). I haven't tried AI with it yet....

Janice: Occasionally, mass digitization projects mostly; sort without messing other data

Bartlett: Use all the time: sorting, calculating, & graphing for budgets, scientific data, historic records, & surveys. Poor performance with large datasets, working with abstracts, and storing flat file data that probably should be a database.

Richard Naples: Use them all the time! I create reports with Excel, clean data, split and join data, run calculations, etc. I think they are great for most of these activities. +I have definitely sorted only some of my data, thus putting it out of sync and ruining everything! Also, those damn auto-date creation. If i have pages 1-12, it's not January 12th!!! Grr. So sometimes WYSIWYG but other times it thinks for you and I hate that!

Tartaro programming data and analysis, contracts. Formula and cell problems with bad results that are not obvious

Crystal: Spreadsheets every day! good for tabular data, pivot tables!

Camille: I use spreadsheets all the time, I do calculations, make graphs, organize data.

Becca: Use them sometimes, but I think there are areas of my job that I could use them more frequently

for; mostly for expenses and general tables

Adam Hager: I use them daily for almost all of my work and home tasks - general formulas, financial docs, expense tracking, inventory

mistakes: erased a cell or formula, used an incorrect formula, inconsistent data entry, translating live data to graphs and charts has been difficult to achieve the results that I'm looking for

Sue Graves: I use them to look at data generated from sql queries

Pivot tables in Excel are so useful, but they often make my hair hurt! I also struggle with using filters incorrectly

Corey Schmidt: Mostly use spreadsheets for data storage/retrieval and use python for data manipulation. Frustrated that I cannot cross compare columns/rows from different sheets - one sheet has data I want to compare to another and highlight differences, but conditional formatting seems too limited to do this.

Julio: all the time. / metadata associated with lab assays. // coordinates formatting, dates formatting, ES and US formatting.

Sonia: Use spreadsheets all the time; data, collections, budgets, organising metadata, preparing data and data files for uploading to other programs

Frustrations = Strange things Excel does with dates (looking forward to this section today)

Shruti-use all the time for data collection. Formatting is difficult.

Meg: All the time- mostly to look at collections data. Mostly use simple formulas like X/ Vlookup:

Frustrated with: repetitive series of steps for a particular formatted outcome

Use them for metadata capture and updating, standardized vocabularies, filtering, basic math functions, project management, reporting. Human error is usually when things go awry, inconsistent spellings, etc. Dates and number formatting are for sure issues, duration of timebased media number formatting

Jenn: use them regularly to organize, track and report data, formatting is problematic

Hernán: Use them regularly for data entry,
dates are frustrating !!!

Alex: Use spreadsheets daily; sorting, using formulas, reorganizing data; works well for the most part minus dates and other formatting issues

(2) Have accidentally lost sheets by saving as csv instead of xls

Abi: yes, mailing data, content research and metadata

Jean: I use them all the time in work and for orgs I volunteer for, mostly budgeting and to track project progress question 2 - erased formula's by mistake, also issues with some add-ons to make bar codes

Gabe: 3 days a week. I record data about DNA samples. Spreadsheets are good for auto-complete lists, and do basic arithmetic; it is difficult to format cells so that "E" is not automatically assumed to be an exponential function, copy-and-pasting dates can often be corrupted if they come from different sources.

Tim: use spreadsheets frequently, primarily as a format to deliver data to someone else, generally I fetch the data in a spreadsheet format and use them to sort and filter data based on column values.

Tim - question 2: formatting data can sometimes be frustrating and how best to handle NULL values, also filtering based on conditionals or based on relations to other data is always tricky, but I guess that's what SQL is for, also things that are specific to Excel spreadsheets and not CSVs generally

Silky: Utilize spreadsheets for budget management, image management, mailing fulfillment. Seems helpful for both.

(2) Not as nimble as I think they could be for formula input. Formatting.

Date format questions:

- What do you notice about the display of the date information above? what information changes between the columns?
- What aspects of the display lack specificity and may introduce ambiguity?

Day 1: Basics of the Unix Shell

<https://librarycarpentry.github.io/lc-shell/>

<https://librarycarpentry.org/lc-shell/data/shell-lesson.zip>

Bash is the default shell on most Linux distributions and older versions of macOS. Windows users will need to install Git Bash to provide a Unix-like environment.

- **Linux:** The default shell is usually Bash, but if your machine is set up differently you can run it by opening a terminal and typing bash followed by the enter key. There is no need to install anything. Look for Terminal in your applications to start the Bash shell.
- **macOS:** Open Terminal from /Applications/Utilities or Spotlight Search. In versions before Catalina, Bash is the default shell, so you do not need to do anything further. In Catalina and onwards, the default shell is zsh, which is similar but may behave differently from Bash in some cases. To switch to Bash, enter the command bash in your terminal window followed by the enter key.
- **Windows:** On Windows, CMD or PowerShell are normally available as the default shell environments. These use a syntax and set of applications unique to Windows systems and are incompatible with the more widely used Unix utilities. However, a Bash shell can be installed on Windows to provide a Unix-like environment. For this lesson we suggest using Git Bash, part of the >Git for Windows package:
 - Download the latest Git for Windows installer.
 - Double click the .exe file to run the installer (for example, Git-2.42.0.2-64-bit.exe) using the default settings.
 - Once installed, open the shell by selecting Git Bash from the start menu (in the Git folder).

COMMANDS:cd

- Pwd. - where you are in your file system,
- Cd – change directory
- o cd Downloads
- o cd .. (go up a level)
- ls - a list of files and folders in your current directory
- o ls -l
- o ls -a
- o ls -lh
- o

man ls. – get to the manual to see the flag options; Q to get out of Manual
on a PC: ls --help.

<https://tldr.sh/>

now list the files in a directory ordered by their filesize. LONG listing by filesize

ls -lS

- `mkdir` – make directory
- § (ctrl-c to quit "cat" view)
- `open`
 - `clear`
 - `cat` – concatenate- (show contents)
-
- `head` – show first ten lines
 - `head -n 20` will show first 20 lines
 - `tail` – show last 10 lines
 - `less.` – view contents one screen at a time - hit the Q to exit
 - (“q” to quit “less” view)
 - `mv` – change file name OR move file into a different directory
 - `cp` – make a copy of a file
 - eg: `cp gulliver.txt gulliver-backup.txt`

The shell supports wildcards!

? (matches exactly one character)

* (matches zero or more characters)

- `cp` – make a copy of a file
 - `mv` - move files

`wc` is the “word count” command: it counts the number of lines, words, and bytes.
`sort`

`-n` flag for numerical sorting

Pipes

Combine `wc -l *.tsv | sort -n`

Use the `>` to output the results to a file

`wc -l *.tsv > lengths.txt`

Search for all case sensitive instances of a whole word you choose in all four derived .tsv files in this directory. Print your results to the shell.

DAY 2: OpenRefine

Lesson: <https://librarycarpentry.github.io/lc-open-refine/>

Slides:

https://docs.google.com/presentation/d/1OfM34MIHHNr8j80vDE32_kfOquyCJl6bo8Yx0RiwaVs/edit?usp=sharing

Dataset: <https://librarycarpentry.github.io/lc-open-refine/data/doaj-article-sample.csv>

OpenRefine: <http://127.0.0.1:3333/>

Cloud Service: <https://github.com/SmithsonianWorkshops/binders>

SLA Shared OpenRefine: <http://sil-viweb-wikibase.si.edu:3333/> (internal only, shared, and not guaranteed to stick around long term!)

Feedback form: <https://forms.gle/LjFaH9oZpU8WfB4C6>

Slides for wrap up and how to stay involved with SI Carpentries

<https://github.com/SmithsonianWorkshops/workshop-slides/blob/main/outro.pdf>

Instructor: naplesr@si.edu

Day 2 Icebreaker - put your name, unit and the last thing you ordered online

- Richard Naples, SLA, oil pastels
- Bea Bock, Northern Arizona University, sunglasses
- Keri Thompson, ODT CAAS, pencil eraser replacements
- Crystal Sanchez, OCIO DAMS, replaced broken w/d filter
- Adam Mansur, NMNH Mineral Sciences, dishwasher tablets
- Alicia Hodson, NMNH IDSC-ITIS, hand soap refills (exciting I know)
- Adam Hager, HMSG, diaper pail refills :)
- Alex Lawrence, NMNH Paleo, mp3 (fake iPod for my kid)
- Becca Stout, NMAI, saddle soap
- Jennifer Lynn Bartlett, ADS @ SAO, water bottle
- Carrie Sims, Naos Panama, book
- Kayla Henry-Griffin, AVMPI @ SLA, Polaroid 600 film and 120 film
- Jennifer Wodzianski, NPG, books
- Walter Forsberg, AVMPI @ SLA, LPs
- Paulina Segarra, OPA, dog food!
- Silky Sullivan, OA, Candle
- Abi Pocasangre, OA, groceries
- Natalia Ovalle, STRI Panama, cat food
- Kira Sobers, SLA - kids games
- Timothy Nguyen, SLA, a raspberry pi ☺
- Marc Tartaro OPDC tea
- Alvin Hutchinson, Smithsonian Libraries and Archives. Dog stuff
- Sonia J Rowley - NMNH-IZ - rebreather diving equipment
- Felicia Boretzky - AVMPI, SLA - 5 lbs hand weights and Hello Panda snacks
- Siobhan Hagan, SLA, organic cheeseits
- Kristina Heinricy, SLA, clothes
- Janice Hussain, CHSDM, cat litter
- Sue Graves, SLA, space heater
- Jean Beard, OSHM, last thing I ordered online was gardening supplies
- Shruti Dube, NMNH, dresses
- Isabela Buxbaum, naos panama, credit card
- Anna Davis, SERC, Valentines for my kids' classes!
- Piper Mullins, PSCI (all science units?), pastry piping tips and bags
- Jenny Koch, ADS/SciX at Smithsonian Astrophysical Observatory, cordless vacuum

DAY 3: Python for Libraries

Intro slides: <https://smithsonianworkshops.github.io/workshop-slides/#/>

Workshop website: <https://smithsonianworkshops.github.io/2025-01-smithsonian-online/>

Lesson: <https://librarycarpentry.github.io/lc-python-intro/aio.html>

Feedback form: <https://forms.gle/LjFaH9oZpU8WfB4C6>

Slides for wrap up and how to stay involved with SI Carpentries

<https://github.com/SmithsonianWorkshops/workshop-slides/blob/main/outro.pdf>

Day 3 Icebreaker - put your name, unit and your favorite soup

- Crystal Sanchez, OCIO DAMS, ramen
- Sue Graves, SLA, corn chowder
- Kristina Heinricy, SLA Web-IT, spicy ramen
- Alicia Hodson, NMNH-IDSC-ITIS, Green Chile Chicken Soup
- Kira Sobers, SLA, broccoli cheese
- Julio Carri
- Sergio dos Santos, STRI, Pumpkin
- Andres Diaz, STRI-CTPA, carrot cream
- Adam Mansur, NMNH, lentil
- Corey Schmidt, ODT, Clam chowder
- Alex Lawrence, NMNH Paleo, red pea soup
- Brianna Toth, AVMPI SLA, Pozole
- Amanda Lawrence, NMAH, Potato
- Kelly Revak, ODT, french onion
- marc tartaro cram oratrat, OPDC
- Jess Shue, SERC, kale potato
- Amanda Reynolds SERC Turkey Kale
- Sonia J Rowley - NMNH-IZ - chunky vegetable, pumpkin, oh so many decisions :)
- Lauren Robinson, CHSDM, matzoh ball soup
- Becca Stout, NMAI, chicken noodle
- Timothy Nguyen, SLA, pho
- Stephanie Fares, NMAH, French onion
- Erin Bordeaux, NMAI, Split Pea
- Carrie Sims stri Minestrone
- Mike Trizna, chicken noodle
- Meg Phillips, NMNH, Minestrone
- Anna Davis - Lentil soup
- Tim Nielsen, SAAM, Tomato Soup with a grilled cheese sandwich dunked in
- Janice Hussain, CHSDM, gumbo
- Julia Blum, SERC, chicken noodle
- Isabela Buxbaum, naos panama, i dont really like soup...
- Shruti Dube, NMNH, tomato
- Richard Naples, SLA, minestrone
- Natalia Ovalle, STRI Panamá, Pozole
- Jennifer Koch, SAO, ramen

-
-

If you had trouble installing Anaconda or JupyterLab, here is a back-up solution:

<https://mybinder.org/v2/gh/SmithsonianWorkshops/binders/python>

This will launch a temporary Jupyter interface, which will time out after 10 or 15 minutes of non-use.

Prompt: Can you think of ways to use Python in libraries?

I use it to get data from APIs, to work with files in bulk, and to work with our Wikibase. I'd love to we make file checksums and validate in bulk

Challenge: Length and Indexing

1. Create a list named colors containing the strings 'red', 'blue', and 'green'.
2. Print the length of the list.
3. Print the first color using indexing.

Solution:

```
colors = ['red', 'blue', 'green']
print(len(colors))
print(colors[0])
```

Challenge: Working With the End

Run the following code and answer the following questions.

```
resources = ['books', 'DVDs', 'maps', 'databases']
print(resources[-1])
```

1. How does Python interpret a negative index value?
2. If resources is a list, what does del resources[-1] do?

1. Predict what each of the print statements in the program below will print.

2. Does max(len(cataloger), assistant_librarian) run or produce an error message? If it runs, does its result make any sense?

```
cataloger = "metadata_curation"
assistant_librarian = "archives"
print(max(cataloger, assistant_librarian))
print(max(len(cataloger), assistant_librarian))
```

```
df = pd.read_csv('https://raw.githubusercontent.com/SmithsonianWorkshops/2025-01-smithsonian-online/refs/heads/gh-pages/data/2011_circ.csv')
```

If you want to learn more about the **describe** method for pandas data frames, you can view its help by

running:
help(pd.DataFrame.describe)
(assuming you aliased pandas to pd when you import-ed it)

Challenge

1. Fill in the blanks so that the program below prints 0123456789.
2. Rewrite the program so that it uses import *without* as.
3. Which form do you find easier to read?

```
import string as s  
numbers = _____.digits  
print(_____)
```

Feedback Form:

<https://docs.google.com/forms/d/e/1FAIpQLScP7ndSbTWwcp5jKKwYHJpx0YboFkSk76r7jj6AsoJ2GVXIpQ/viewform>