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

---

UCLA / Los Angeles Area Land Aknowledgement

Gabrielino-Tongva Tribe: https://gabrielinotribe.org/
L.A. Metro Gold line Basket Bridge over 210 frwy: https://www.sgvtribune.com/2013/12/05/gold-line-basket-bridge-over-210-freeway-gets-record-fifth-award/
The woven-basket look of the bridge's support columns emulate the famed woven baskets of the native Gabrielino/Tongva of the San Gabriel Valley while the underbelly of the bridge is supposed to evoke a Western diamondback rattlesnake.

Topic: Tidy Data

- *"Happy families are all alike; every unhappy family is unhappy in its own way." — Leo Tolstoy*
- *"Tidy datasets are all alike, but every messy dataset is messy in its own way." — Hadley Wickham*

Zoom link for workshop: [check registration email]
Zoom registration link if still need Zoom Room -
https://ucsd.zoom.us/meeting/register/tJMsfuypqDIoGNRMKsJeUxxDIT1ru7uV94af

Workshop website:
https://librarycarpentry.org/lc-spreadsheets/
https://hackmd.io/SGh5v7YKS1SMcU1HFwHYzQ

This Etherpad (collaborative notetaking & link sharing document):
https://pad.carpentries.org/uc-carpentries-tidy-data-2023

Links to all notes for the worshop series (with links to other etherpads):
https://hackmd.io/vfJr8dEaQn-bKG2XoWZexw

Workshop Survey:

https://ucsd.co1.qualtrics.com/jfe/form/SV_cNn33BLa3CmbCzI

Spreadsheet to download:
https://librarycarpentry.org/lc-spreadsheets/data/training_attendance.xlsx

Sign In

Name / pronouns / Dept / email /
Jamie Jamison / she/her / Library Data Science Center /UCLA / jamison@library.ucla.edu / Instructor ☆
Elizabeth "Lisa" McAulay / she/her / Digital Library Program / UCLA/emcaulay@library.ucla.edu / Instructor ☆
Kristian Allen / he/him / Library / UCLA / kallen2@library.ucla.edu / helper ☆
Shimelis Abebe Tegegn / / / shte0001@stud.slu.se
Diana Corona-Mata / she, her, hers / UCSD / dcoronamata@health.ucsd.edu
Ahmad Danesh/He/Him/ CAPS/ UCSF / ahmad.danesh@ucsf.edu
Carly Pope /she/her / Archaeology / UCLA/ c3m2pope@ucla.edu
Jonathan Morrow/ he/him / / UCLA/ jonmorrow@ucla.edu
Ramazan Yol/he/him/ Math/ IU / rmznyol@iu.edu
Alice Abate / she, her, hers / / UCSF / Alice.Abate@ucsf.edu
Nikola Marinos Raitsevits/ he,him,his/ /UCLA/ nmarinosraitsevits@ucsd.edu
Benjamin Lam / he/him / /UCSD/b6lam@ucsd.edu
Brooklyn Asai / she/her / MAE/UCSD / basai@ucsd.edu
Caitlin Nordheim / she/her / /UCSB /caitlinnordheim@ucsb.edu/ Ecology
Sofia Urgoiti Crespo / she/her/UCSB - Marine Science/surgoiticrespo@ucsb.edu
Levi Berge-Wells/he/him/ UCSD/ lbergewells@ucsd.edu/ Sociology
Emily Dovan / She/her / UCLA / edovan@library.ucla.edu
Jason Budge / he /him / UCSB / Sociology / jbudge@ucsb.edu
Elizabeth Lhost / she/her / Library /UCLA / lhost@library.ucla.edu
Melissa Hua/she/her/Environmental Science/ UCLA/ mhua27@g.ucla.edu
Emily Lam / she, her, hers / UCSD / eml048@ucsd.edu
Phoebe Valdes / she/her / UCSD / prvaldes@ucsd.edu
Joshua Love / He/him / UCSB Earth Science / jlove@ucsb.edu
Marie Kennedy / she/her / library / marie.kennedy@lmu.edu
Caitlin Hunter/ she/her / UCLA Law Library/ hunter@law.ucla.edu
Kimia Kamal / she,her / UCLA psychology / kkamal@ucla.edu
Julia Dominesey/ she her/ UCSD/jdominesey@ucsd.edu
Heather Wagner/she her/UCM/ hwagner3@ucmerced.edu
Ethel Tackie-Yarbo/UCSF/ethel.tackieyarboi@ucsf.edu
Elise Lelou / She Her / UCSF / elise.lelou@ucsf.edu
sara matsumura/she her/ UCSB/saramatsumura@ucsb.edu
Nicholas Lam / he/him / UC San Diego / n2lam@ucsd.edu
Jasper Romero / he/him / UCSC / jwromero@ucsc.edu
Anthony Swanner/ he/him / UCSF/ john.swanner@ucsf.edu
Anudari Letian/she her/ UCSF / anudari.letianu@ucsf.edu
Corinna DesCombaz/she her/ UCSF / Corinna.DesCombaz@ucsf.edu
Alyssa Ortega/ she her / UCM CIS staff / aortega59@ucmerced.edu
Gokh Amin Alshaif/ she her/ UCSB History/ galshaif@ucsb.edu

Hope Reuschel / she, they / UCM / hreuschel@ucmerced.edu
alexis roberto / she/they / UCLA / aroberto@g.ucla.edu
Nicole Avalon / she, her / UCSD / navalon@ucsd.edu
Chen Chen/she, her/UCSD/chc048@ucsd.edu
Andreina Soto /she her / UCSB / andreina@ucsb.edu
Wesley Cheng/ / / UCSD/wesleycheng@health.ucsd.edu
Leah Klement / she, her / UCSD/ lklement@ucsd.edu
Marissa Todesco / she her/ UCSD / mtodesco@ucsd.edu
Kyla Murphy / she her/ UCSB/ kylamurphy@ucsb.edu
Bianca Badajos / she her / UCLA / biancadajo@ucla.edu
Sarah Gautam / she her /UCSF / Fresno IT Data Analyst / sarah.gautam@ucsf.edu
Lucy Young / she/her / UCSF / lucy.young@ucsf.edu
Naoko Ohgama / she, her / UCB / ohgama@berkeley.edu
Iveth Estrada Reyes / she, her, they / UCLA / ivether@g.ucla.edu
Nam Do / he, him / UCSD / nad002@ucsd.edu
Jennifer Bailey / she/her / UCSD / jlbailey@ucsd.edu
Tamara Pilko/ she/her / UCSC Library / tpilko@ucsc.edu
Scott Peterson /he/him/ UC Berkeley, speterso@berkeley.edu
Zack Moreau / UCSD, NanoEngineering / zmoreau@ucsd.edu
ViviAnne Steer/ she her/ UCB / viviannesteer@berkeley.edu
Yuting Wang / she/her / UCSB / yuting_wang@ucsb.edu
Laura Cheradame /she her / UCSF / laura.cheradame@ucsf.edu
Alejandro SanchezNunez/UCLA/sanchez@humnet.ucla.edu
Charles Enyaah Amankwa/he him/UCSD/camankwa@health.ucsd.edu
Gina Nam/she her/UCLA/ginaenam@ucla.edu
Carly Garzon Vargas/she they/UCSD/cgarzonvargas@ucsd.edu
Cecilia Saavedra/ she, her, ella/ UCB / cecilia.saavedra@berkeley.edu
Martha Michel/she her/VA-UCSF/martha.michel@va.gov

Schedule

WELCOME, Sign In, Introductions

8:30 - 8:40

1. Introductions to Instructors and Helpers
2. Land Acknowledgement
3. Code of Conduct Reminder
4. Question and Answer Process

8:40am

Introduction to Tidy Data

Intro: https://librarycarpentry.org/lc-spreadsheets/instructor/00-intro.html
1. Using spreadsheet programs for data organization. What are good data practices for using spreadsheets for organizing data?

https://librarycarpentry.org/lc-spreadsheets/instructor/00-intro.html#problems-with-spreadsheets

8:50 am

Formatting and Best Practices

Best Practices
-Leave data raw
-Record steps of cleaning process to revert back if needed.

- -OpenRefine program can save steps automatically
- * We may demo OpenRefine near end (https://openrefine.org/), the carpentries also has a lesson (https://librarycarpentry.org/lc-open-refine/) on this tool worth reviewing, essentially can be thought of as Excel on steroids

-All variables to measure in columns (ex: length, width, height), attributes/observations in rows (ex: specific measurements)
-One attribute/observation per cell. Don't combine values
2. Formatting data tables in Spreadsheets How should data be formatted in spreadsheets?
https://librarycarpentry.org/lc-spreadsheets/instructor/01-format-data.html

09h 25am 3. Formatting problems. What common mistakes are made when formatting spreadsheets?
https://librarycarpentry.org/lc-spreadsheets/instructor/01-format-data.html

9:30
~~~ **Micro Break (5 min) ~~~**

10h 00m 4. Dates as data. How are dates handled by computers?
-Make a copy, make a note in a seperate note sheet
-Start cleanup by observing needs, prioritize

- -columns

https://librarycarpentry.org/lc-spreadsheets/instructor/03-dates-as-data.html

Tip - if using textpad on OSX, under top menu selection 'Format' select 'Change to Plain Text' to get a text file rather than rich text format

Keep in mind who end user of your data will be, some may be a 'human readable' version of your data, essentially a way to read your raw data and generate a human readable version, so this operation may be repeatable in a reliable manner

10h 10m 5. Basic quality assurance and controlHow can you keep data entry clean?
https://librarycarpentry.org/lc-spreadsheets/instructor/04-quality-control.html

10:50 AM
~~~ Longer Break : 20 Mins ~~~
Instructor Switch

11:10am

6. Exporting data from spreadsheets. What problems are there with Excel files?
How can we share data from spreadsheets that is useful for a variety of applications?
https://librarycarpentry.org/lc-spreadsheets/instructor/05-exporting-data.html

Dates are stored internally in different formats by different software, so it is a best practice to be aware of how dates are stored, especially when exporting

Adding a number of months to a date, the formula EDATE can be used,
=EDATE([Cell of date you want to manipulate], [integer of number of months you want to add])
=EDATE(A2, 5)
Will add 5 months to the date in cell A2

12:00pm 7. Caveats of popular data and file formatsWhat do you need to be aware of when exporting data?

12:30 pm Finish

Workshop Survey:

https://ucsd.co1.qualtrics.com/jfe/form/SV_cNn33BLa3CmbCzI

Tidy Data Notes:

Journal of Statistical Software article on TidyData:
https://www.jstatsoft.org/article/view/v059i10

Tools for organizing, storing and manipulating data:

https://git-scm.com/ - git - Tool to track changes in files
https://github.com/ - github - Cloud tool on top of git that allows group collaboration with git
https://openrefine.org/ - OpenRefine - Kind of a 'super-excel', a good midpoint if you need to do some manipulation of data but maybe not quite to the point where you need a full blown programming environment
https://www.libreoffice.org/ - LibreOffice - Free, open source equivalent of Microsoft Office. It can read and write data in MS Office format.

Advanced date manipulation:
If you are already familiar with R/Python or interested in tools to handle oddball cases, these are popular libraries we typically use:
R - lubridate - https://lubridate.tidyverse.org/
Python - arrow - https://arrow.readthedocs.io/en/latest/

Day zero in Google Sheets is 12/30/1899 0:00:00 and Day 1 is 12/31/1899 0:00:00
Google Sheets you can use negative numbers to get dates backwards from 12/30/1899.