



University of Central Florida  
**STARS**

---

Scholarly Communication Brown Bag Series

Scholarly Communication

---

6-23-2021

## Stay Savvy with Scholarly Communication: Data Organization in Spreadsheets

Sandy Avila

University of Central Florida, [savila@ucf.edu](mailto:savila@ucf.edu)

Lee Dotson

University of Central Florida, [lee.dotson@ucf.edu](mailto:lee.dotson@ucf.edu)

Find similar works at: <https://stars.library.ucf.edu/scholcom-brownbag>

University of Central Florida Libraries <http://library.ucf.edu>

This Presentation is brought to you for free and open access by the Scholarly Communication at STARS. It has been accepted for inclusion in Scholarly Communication Brown Bag Series by an authorized administrator of STARS. For more information, please contact [STARS@ucf.edu](mailto:STARS@ucf.edu).

---

### Recommended Citation

Avila, Sandy and Dotson, Lee, "Stay Savvy with Scholarly Communication: Data Organization in Spreadsheets" (2021). *Scholarly Communication Brown Bag Series*. 15.

<https://stars.library.ucf.edu/scholcom-brownbag/15>



# Data Organization in Spreadsheets for Social Scientists

• • •

**Lee Dotson**

Interim Dept. Head

Technology Solutions & Digital Initiatives

**Sandy Avila**

Science Librarian

UCF Libraries

Stay Savvy with Scholarly Communication

June 23, 2021



# Getting Started

Post its =



Have Excel open & ready to go

Have files ready to work with

Open Etherpad and be ready to respond as instructed:

[https://etherpad.wikimedia.org/p/Stay Savvy Data Organization in Spreadsheets](https://etherpad.wikimedia.org/p/Stay_Savvy_Data_Organization_in_Spreadsheets)

# Introduction

- Good data organization is the foundation of any research project
- This workshop will show us how to use features in Excel to check your data quality
- The lesson will cover
  - Formatting Data Tables in Spreadsheets
  - Dates as Data
  - Quality Assurance
  - Exporting Data

# Formatting Data Tables in Spreadsheets

## Structuring Data

- Put all variables in columns
- Put each observation in its own row
- Don't combine info in one cell
- Leave raw data alone- don't change!
- Export the cleaned data consider

Tidy Data Principles:

<https://cran.r-project.org/web/packages/tidyr/vignettes/tidy-data.html>

## Metadata

- Recording data about your data is important
- Store metadata in a separate file in the same directory
- Follow a standard like Data Documentation Initiative (DDI)-

[www.ddialliance.org](http://www.ddialliance.org)

# Dates as Data



<https://i.kym-cdn.com/photos/images/original/001/247/306/867.png>

Your goal = Reduce ambiguity!

- Single column
- Regional variations
- Different versions/suites of software
- “Useful features”

# Date Formatting



<https://memeguy.com/photo/262993/perfect-date>

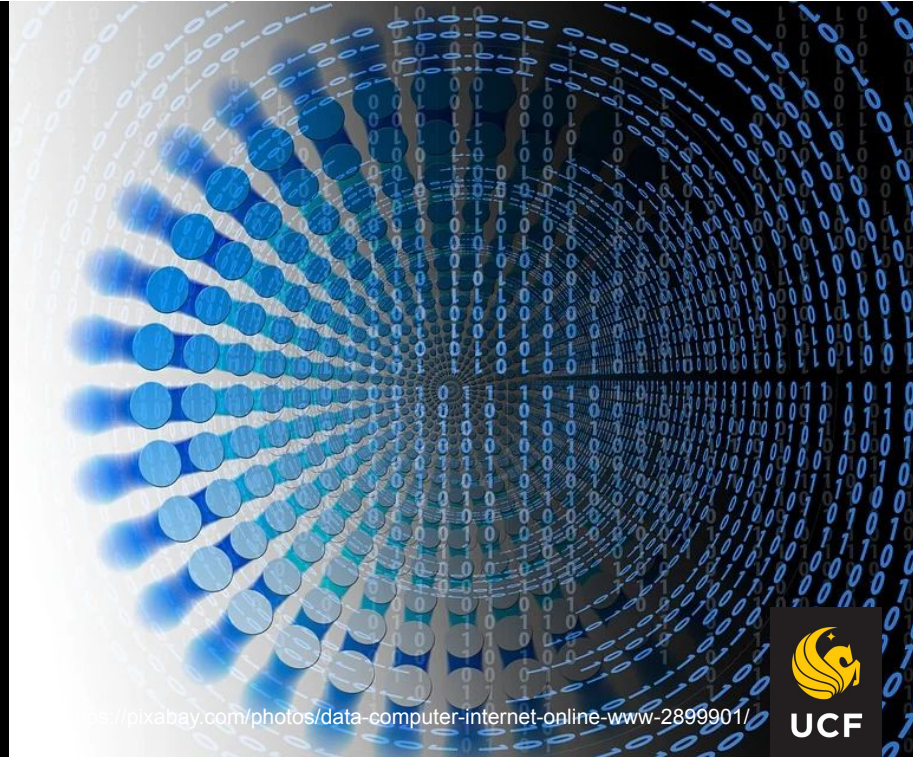
- Dates stored as integers
- Excel's two date systems
- Regional date formatting
- Historical data
- Exercises using the SAFI\_dates.xlsx file



# Quality Assurance

## Using Data Validation Restrictions

- Restricting data to a numeric range
- Restricting data to entries from a list





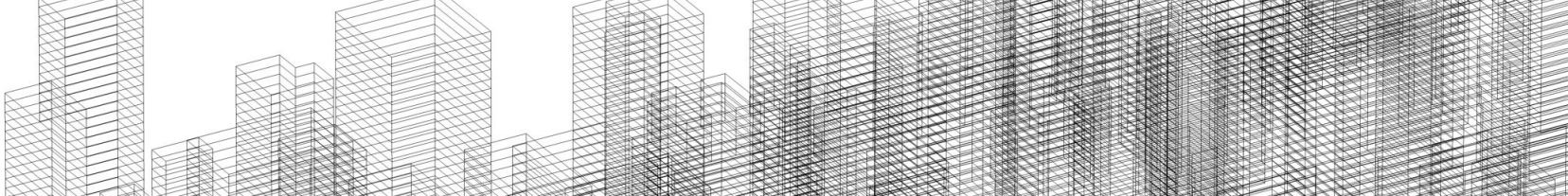
# Exporting Data

---

## Using Universal File Formats for Portability and Reproducibility

### Why not save as .xls or .xlsx?

- Proprietary format
- Variations in software versions
- Publication/data repository requirements
- Inconsistencies in data storage



# Final Thoughts

- Good data entry practices are important
- Watch out for common formatting mistakes
- Utilize quality control measures & data manipulation when possible
- Follow metadata standards
- Export data properly

# Thanks for your time & patience! Any questions?

**Sandy Avila**

Science Librarian

[savila@ucf.edu](mailto:savila@ucf.edu)

**Lee Dotson**

Interim Dept. Head

Technology Solutions & Digital Initiatives

[lee.dotson@ucf.edu](mailto:lee.dotson@ucf.edu)

