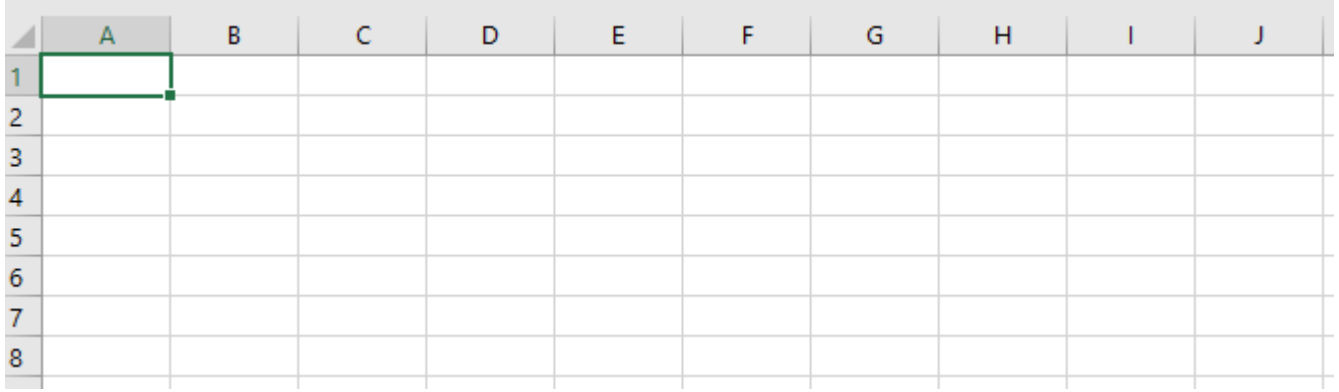
# Best Practices: Data Management for R

# Files

Assuming most of you will keep your data in spreadsheets

- .csv is great!
  - Easily read into R, non-proprietary
- .xls is ok
  - Need to convert to .txt, or use an R package to read in (not preferred)

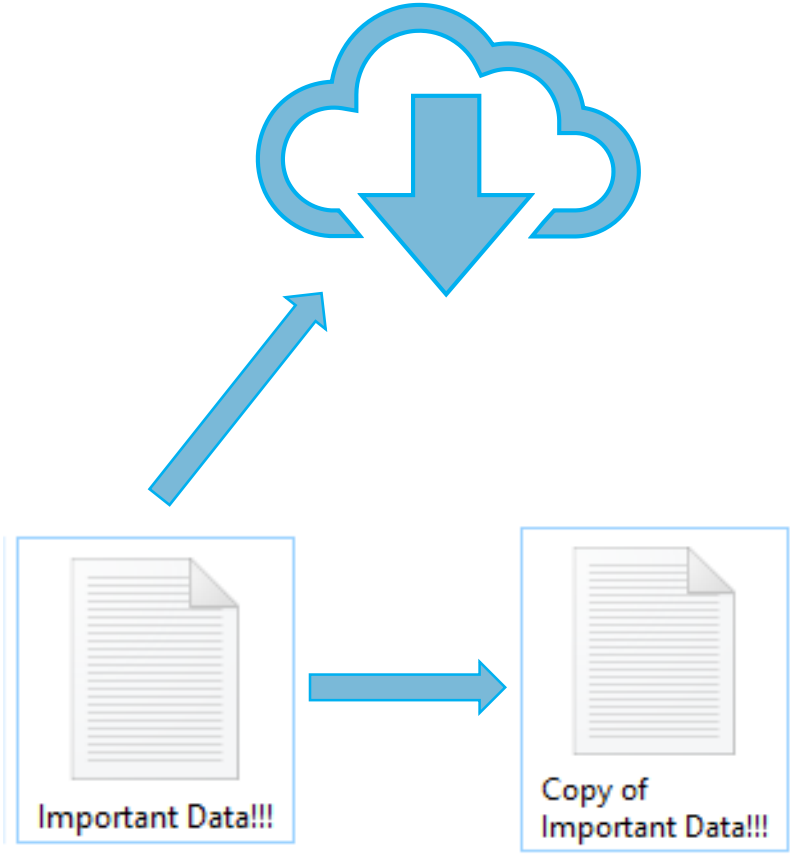


# Files

Backup!

- Maintain at least one copy of the *original* file
  - Raw data, unedited

- Keep copies of files in cloud-based storage

# Data Format

Always keep a header row

| | A | B | C | D |
|---|---|---|---|---|
| 1 | site | census | species | population |
| 2 | BCI | 1 | trichilia_tuberculata | 6523 |
| 3 | BCI | 1 | vochysia_ferruginea | 321 |
| 4 | BCI | 1 | astronium_graveolens | 3258 |

# Data Format

Always keep a header row

| | A | B | C | D |
|---|---|---|---|---|
| 1 | site | census | species | population |
| 2 | BCI | 1 | trichilia_tuberculata | 6523 |
| 3 | BCI | 1 | vochysia_ferruginea | 321 |
| 4 | BCI | 1 | astronium_graveolens | 3258 |

Plain-text for variable names, data

- Only letters, numbers, dashes - , underscores _ , periods .

# Data Format

Always keep a header row

| | A | B | C | D |
|---|---|---|---|---|
| 1 | site | census | species | population |
| 2 | BCI | 1 | trichilia_tuberculata | 6523 |
| 3 | BCI | 1 | vochysia_ferruginea | 321 |
| 4 | BCI | 1 | astronium_graveolens | 3258 |

Plain-text for variable names, data

- Only letters, numbers, dashes - , underscores _ , periods .

No spaces!!!

| species | p |
|---|---|
| trichilia tuberculata | |
| vochysia ferruginea | |
| astronium graveolens | |

| species | |
|---|---|
| trichilia_tuberculata | |
| vochysia_ferruginea | |
| astronium_graveolens | |

# Data Format

Be careful with formats for variables

- Its easier to work with complex variables when they're broken up
- Easier to combine variables than break them

# Data Format

Be careful with formats for variables

- Its easier to work with complex variables when they're broken up
- Easier to combine variables than break them

e.g. Time

| |
|---|
| 10:23 |
| 10:27 |
| 10:30 |
| 10:34 |
| 10:38 |

| hour | minute |
|---|---|
| 10 | 23 |
| 10 | 27 |
| 10 | 30 |
| 10 | 34 |
| 10 | 38 |

# Data Format

Same class of variable in each column

- Exception is NA

| species | population |
|---|---|
| trichilia_tuberculata | 6523 |
| vochysia_ferruginea | 321xx |
| astronium_graveolens | 3258 |
| trichilia_tuberculata | 6833 |
| vochysia_ferruginea | not sure |
| astronium_graveolens | 3120 |

| species | population |
|---|---|
| trichilia_tuberculata | 6523 |
| vochysia_ferruginea | 321 |
| astronium_graveolens | 3258 |
| trichilia_tuberculata | 6833 |
| vochysia_ferruginea | 402 |
| astronium_graveolens | 3120 |

# Long vs Wide

Enter your data in a long format, not a wide format

- Much easier to work across rows than columns in data frames

| census | species | population |
|---|---|---|
| 1 | trichilia_tuberculata | 6523 |
| 1 | vochysia_ferruginea | 321 |
| 1 | astronium_graveolens | 3258 |
| 2 | trichilia_tuberculata | 6833 |
| 2 | vochysia_ferruginea | 402 |
| 2 | astronium_graveolens | 3120 |

VS

| site | census | trichilia_tuberculata | vochysia_ferruginea | astronium_graveolens |
|---|---|---|---|---|
| BCI | 1 | 6523 | 321 | 3258 |
| BCI | 2 | 6833 | 402 | 3120 |

# Long vs Wide

In other words, one observation per row

1 observation

| census | species | population |
|---|---|---|
| 1 | trichilia_tuberculata | 6523 |
| 1 | vochysia_ferruginea | 321 |
| 1 | astronium_graveolens | 3258 |
| 2 | trichilia_tuberculata | 6833 |
| 2 | vochysia_ferruginea | 402 |
| 2 | astronium_graveolens | 3120 |

3 observations

| site | census | trichilia_tuberculata | vochysia_ferruginea | astronium_graveolens |
|---|---|---|---|---|
| BCI | 1 | 6523 | 321 | 3258 |
| BCI | 2 | 6833 | 402 | 3120 |

# Data Format

Include meta data if you want, but keep it within a few lines *above* your data

- Easy to tell R which lines to ignore if they come before data

| | | | |
|---|---|---|---|
| author: Andrew Muehleisen | | | |
| date entered: 9/26/2017 | | | |
| notes: enter your data smart: save time! | | | |
| | | | |
| **site** | **census** | **species** | **population** |
| BCI | 1 | trichilia_tuberculata | 6523 |
| BCI | 1 | vochysia_ferruginea | 321 |
| BCI | 1 | astronium_graveolens | 3258 |
| BCI | 2 | trichilia_tuberculata | 6833 |
| BCI | 2 | vochysia_ferruginea | 402 |
| BCI | 2 | astronium_graveolens | 3120 |

Most Importantly: Be good to your future self!