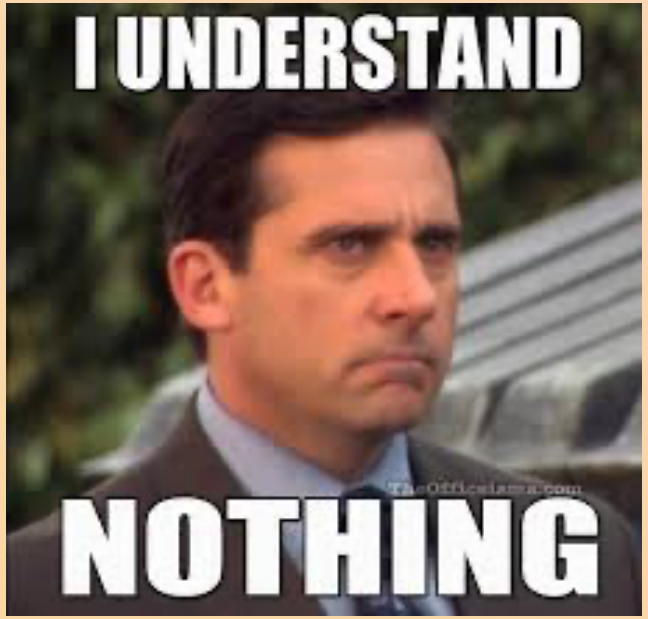
Datapalooza 2020

Finding Data Balance

Tyson S. Barrett, PhD

Photo by Bekir Dönmez on Unsplash



Microsoft Excel or Google Sheets

SPSS, Jamovi, or JASP

R or SAS

Data in Spreadsheets

A million ways to make your life harder

File	Data Modified
dataFINAL_v5_comments.xlsx	02/09/2019
dataFINAL_FINAL_v2.xlsx	02/28/2019
dataFINAL_FINAL_FINAL.xlsx	02/20/2019
dataForAnalysis.xlsx	12/01/2018
data_Jim_FINAL.xlsx	12/15/2018

Data in Spreadsheets

A million ways to make your life harder

File	Data Modified
dataFINAL_KS_Comments.xlsx	02/09/2019
dataFINAL_FINAL_v2.xlsx	02/28/2019
dataFINAL_FINAL_FINAL.xlsx	02/28/2019
dataForAnalysis.xlsx	12/01/2018
data_Jim_FINAL.xlsx	12/15/2018

Copy-and-Pasting data

Color coding without indicators

Dates and things that look like dates

Hiding columns

Data in Spreadsheets

A million ways to make your life harder

Cop

Color coding w
indicato

things that look like dates

Data in Spreadsheets

Broman et al., 2018

Some Basic Principles

Be consistent

Write dates like YYYY-MM-DD

Don't leave any cells empty

Put just one thing in a cell

Organize data as a single rectangle

Create a data dictionary

Don't use color as data

Don't include calculations in raw data

Data in Spreadsheets

Broman et al., 2018

Some Basic Principles

Create a data

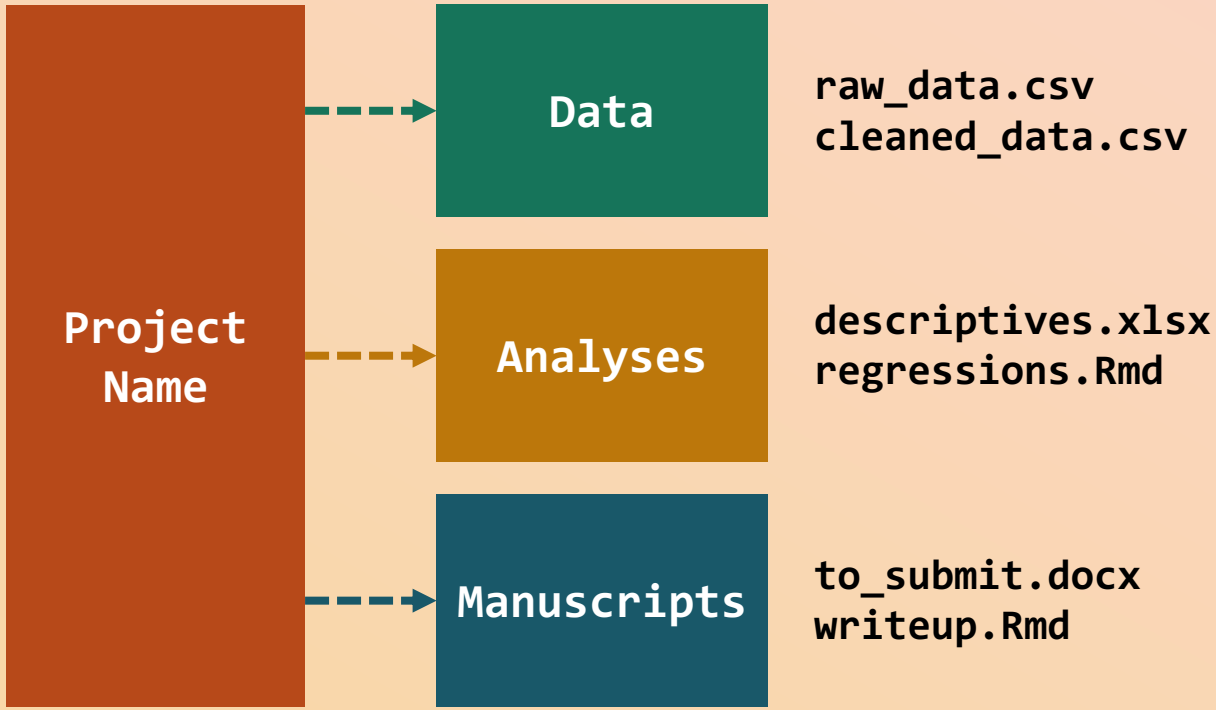
These are ways to be kind to your future self and obtain data balance!

Organize data as a single rectangle

calculations in raw data

Avoid Headaches!

Have a consistent folder structure for where Spreadsheet analyses are vs Spreadsheet data are



A Grammar of Data

`Wickham, SQL, dplyr`

Term	Definition	Synonyms
data wrangling	obtaining and cleaning data to get ready for analyses	data cleaning, data manipulating
selecting	grabbing columns (”variables”) and leaving others out	
filtering	grabbing rows (“observations”) and leaving others out	
mutating	creating a new variable and adjusting an existing one	transform, recoding
summarizing	reducing the amount of data using summary information (e.g. means, medians, etc.)	aggregating
pivoting	Reformatting the data to make it longer or wider (not changing any values in the data)	reshaping, elongating, widening
tidying	Creating a data set that has: 1. All columns are variables 2. All rows are observations	rectangulating, cleaning, pivoting

A Grammar of Data

Wickham, SQL, dplyr

Term	Definition	Synonyms
data wrangling	obtaining and cleaning data to get ready for analyses	data cleaning, data manipulating
selecting	grabbing columns (”variables”) and leaving others out	
filtering	grabbing rows (“observations”) and leaving others out	
mutating	creating a new variable and adjusting an existing one	transform, recoding
summarizing	reducing the amount of data using summary information (e.g. means, medians, etc.)	aggregating
pivoting	Reformatting the data to make it longer or wider (not changing any values in the data)	reshaping, elongating, widening
tidying	Creating a data set that has: 1. All columns are variables 2. All rows are observations	rectangulating, cleaning, pivoting

Tyson S. Barrett

tyson.barrett@usu.edu

tysonbarrett.com

@healthandstats

github.com/tysonstanley

Slides at tysonbarrett.com/teaching