

Workflow of Data Preparation

Indiana University, Workshop in Methods

3 February 2023

Bianca Manago

Vanderbilt University

Table of Contents

- Part 1. Introduction
- Part 2. Best Practices
- Part 3. Steps and Order of Data Preparation

Part 1. Introduction

What is data preparation?

What is data preparation?

- Addressing missing data
- Labeling variables and values
- Re-coding variables
- Importing data
- Dropping variables
- Descriptive statistics
- Validating data
- Graphing data
- Outliers

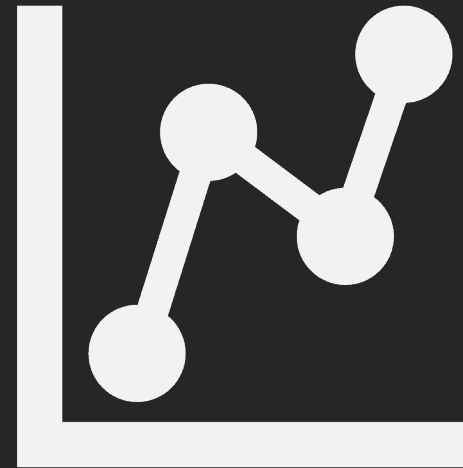
What is data preparation?

Time consuming, behind-the-scenes work



What is data analysis?

- Estimating models to address research questions
- Creating tables
- Creating figures



Myths about data preparation

Myths about data preparation

Data preparation is unimportant & boring

- Data preparation is as important as analysis
- Data preparation requires knowledge
- Data preparation requires skill
- Data preparation requires *thinking*

Myths about data preparation

There are “pre-cleaned” datasets

- Data needs to be cleaned in context
- Clean data for Project A \neq clean data for Project B

Myths about data preparation

Data preparation is separate from data analysis

- Theory affects data preparation & analysis
- Decisions about data preparation affect analysis
- Decisions about data analysis affect preparation
- Dual workflow (more later) (Long 2009)

Myths about data preparation

Data preparation is straightforward

- Data preparation requires dozens of decisions
- No standard method to data preparation
- Researchers make different decisions in same situations (Leahey, et al. 2003)

Why is data preparation
important?

Case Studies

NB: The examples here are not to shame anyone

The Impact of Chief Executive Officer Personality on Top Management Team Dynamics: One Mechanism by Which Leadership Affects Organizational Performance

CEOs' personality traits (e.g., extraversion) affected the group dynamics of the top management team

The Impact of Chief Executive Officer Personality on Top Management
Team Dynamics: One Mechanism by Which Leadership Affects
Organizational Performance

Mishandling of Outliers

CEOs' personality traits (e.g., extraversion) affected the
group dynamics of the top management team



PEDIATRICS

Poverty and Trends in 3 Common Chronic Disorders

Reported prevalence of autism up 400%

PEDIATRICS

Poverty and Trends in 3

Common Chronic Disorders

Miscoding of Variables
Error in Transposing

Reported prevalence of autism up 400%

Police Violence and Citizen Crime Reporting in the Black Community

American Sociological Review
2016, Vol. 81(5) 857–876
© American Sociological
Association 2016
DOI: 10.1177/0003122416663494
<http://asr.sagepub.com>



Police violence reduces calls to police

Police Violence and Citizen Crime Reporting in the Black Community

American Sociological Review
2016, Vol. 81(5) 857–876
© American Sociological
Association 2016
DOI: 10.1177/032216331663494
<http://as.sagepub.com>



**Unclear Inclusion Criteria
Mishandling of Outliers**

Police violence, escape, calls to police

Police Violence and Citizen Crime Reporting in the Black Community

American Sociological Review
2016, Vol. 81(5) 857–876
© American Sociological
Association 2016
DOI: 10.1177/00031224166663494
<http://asr.sagepub.com>



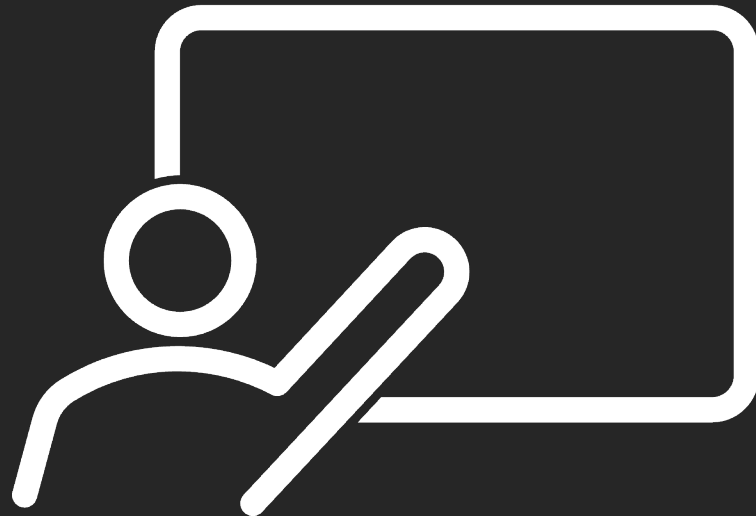
**Finding eventually
checks out!**

Police violence requires calls to police

How/why do data preparation errors occur?

Training

- Apprenticeship model to data preparation
- Standard/simple practices \neq best practices
- No overall guide



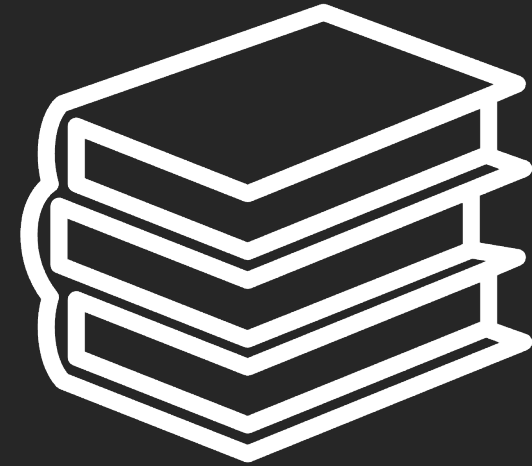
Literature

No formal reporting mechanism

No examples

No peer review

No way to catch errors



Human error

- Humans make errors
- Many opportunities for error
- No way to catch error



So what?

Humans make mistakes

Why errors matter

Incorrect information

- Confuse record
- Decisions/decision-makers
- Future research



Lose trust of public (Hendriks et al. 2020, Wingen et al. 2020)

How does a standardized approach help?

What is a standardized approach?

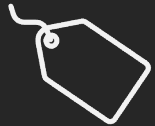
- Accounts for all data preparation steps
- Considers order of steps
- Follows best practices



CLEANR Method



Compile



Label/Name



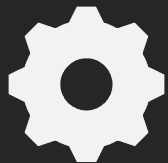
Examine



Alter



New variables



Re-configure & re-examine

CLEANR Method



Compile



Label/Name



Examine



Alter



New variables



Re-configure & re-examine

More to Come...

Back to the question....

How does a standardized approach help?

Enhancing reproducibility



What is reproducibility?

- Analysis yields same results
- Same or different samples (Plesser 2018)
- Relies on strength of evidence

Enhancing reproducibility



How does a standardized approach help?

- Data preparation decisions affect findings
- Easy to miss or forget steps
- Easy to overlook things
- Negative effect on reproducibility/accuracy

Enhancing replication

What is replication?

- Analysis yields identical results (Plesser 2018)
- Same Data → Same Analyses → Same Results



Enhancing replication

How does a standardized approach help?

- Know what decisions we made
- Why we made them



Why do reproducibility and replication matter?

Why do reproducibility/replication matter?

It is the only thing we can guarantee

Not interesting results

Not significant results

Not even correct results*

Why do reproducibility/replication matter?

Selfish motivations

Save time

Repetition

Lost work

Review process

Protect Reputation

Errors are human

People are judgey

Trust in your research

(Anvari, et al. 2018)

Why do reproducibility/replication matter?

External motivations

Publishing

Journal requirements

Funding

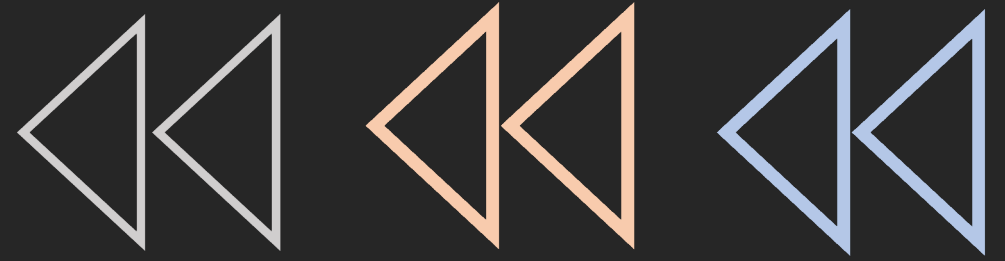
Agency requirements

Why do reproducibility/replication matter?

Internal motivations

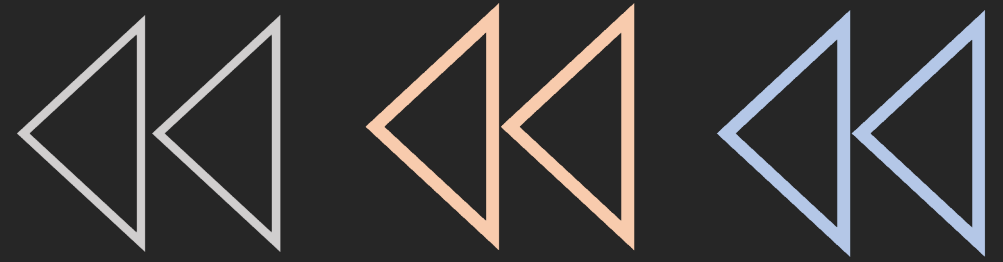
It is the right thing to do.

Without replication, knowledge does not accumulate, and veracity is questioned



Summary

Summary



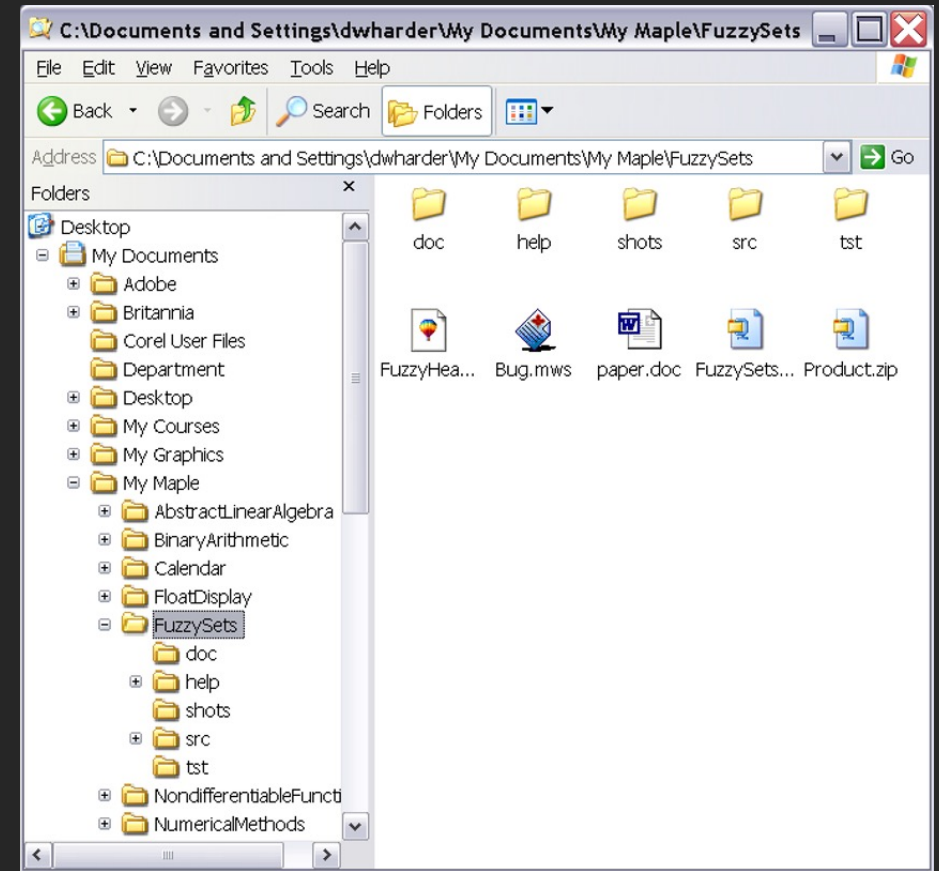
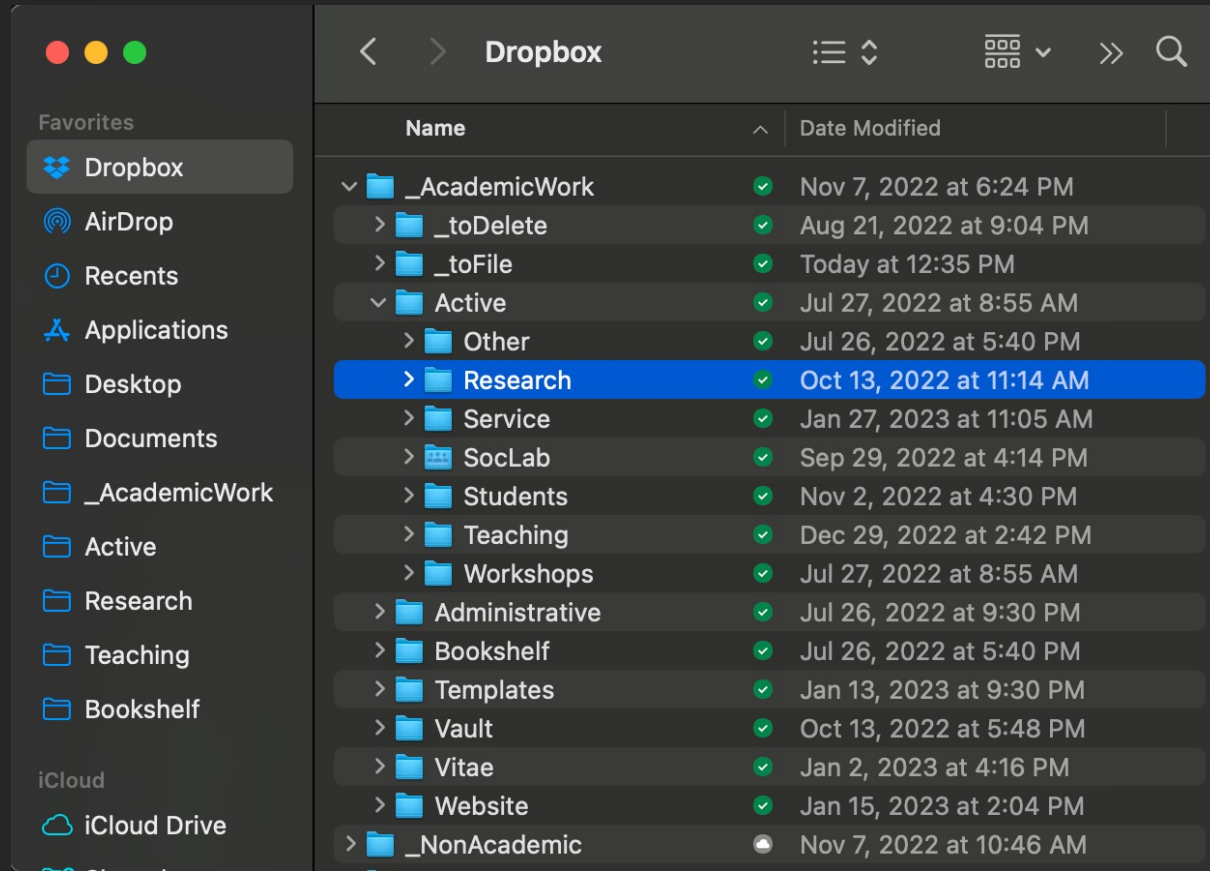
- What is data preparation?
- Myths about data preparation
- Why is data preparation important?
- Why do errors occur/matter?
- How standardized approach helps.
- Importance of reproducibility and replication.

Part 2. Best Practices

enhancing replication through workflow

2.1 File Organization/ Directory Structure

What is a directory structure?



Evaluating your directory structure

“Someone unfamiliar with your project should be able to look at your computer files and understand in detail what you did and why. This ‘someone’ could be any of a variety of people:.... Most commonly, however, that ‘someone’ is you.”

William Stafford Noble
in PLOS Computational Biology (2009)






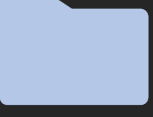



How to organize research project folders

1. Overall directory structure
2. Research project folder naming
3. Subfolders
4. Working directories



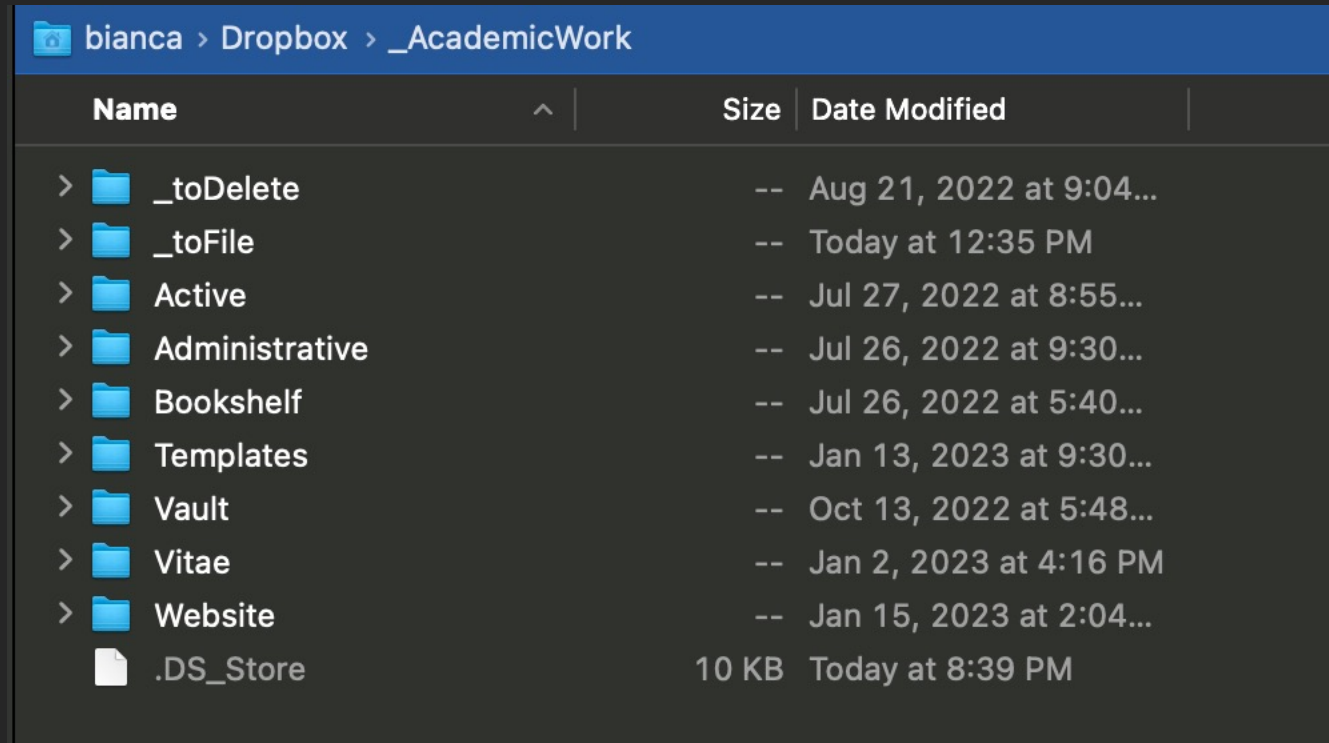
Directory Structure

Main Level

-  _toDelete
-  _toFile
-  Active
-  Admin
-  Bookshelf
-  Templates
-  Vault
-  Vitae
-  Website

Directory Structure

Main Level



The screenshot shows a file explorer window with the path "bianca > Dropbox > _AcademicWork". The window displays a list of folders and files with columns for Name, Size, and Date Modified. The folders are: _toDelete, _toFile, Active, Administrative, Bookshelf, Templates, Vault, Vitae, and Website. The file is: .DS_Store (10 KB).

Name	Size	Date Modified
> _toDelete	--	Aug 21, 2022 at 9:04...
> _toFile	--	Today at 12:35 PM
> Active	--	Jul 27, 2022 at 8:55...
> Administrative	--	Jul 26, 2022 at 9:30...
> Bookshelf	--	Jul 26, 2022 at 5:40...
> Templates	--	Jan 13, 2023 at 9:30...
> Vault	--	Oct 13, 2022 at 5:48...
> Vitae	--	Jan 2, 2023 at 4:16 PM
> Website	--	Jan 15, 2023 at 2:04...
.DS_Store	10 KB	Today at 8:39 PM

Directory Structure

Second Level



Active



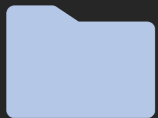
Grants



Learning



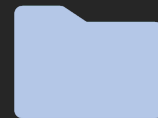
Research



Service



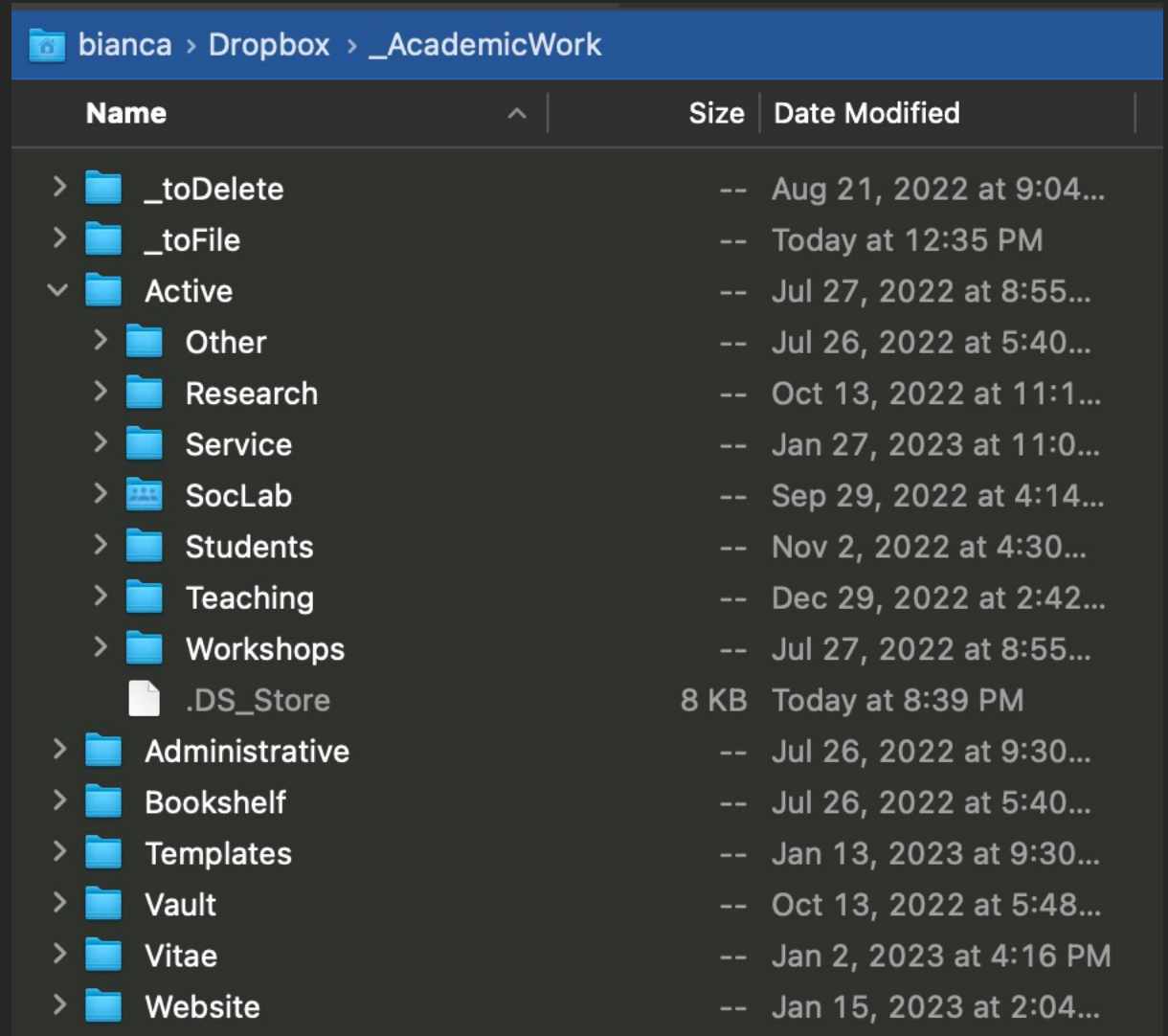
Students



Teaching

Directory Structure

Second Level




Name	Size	Date Modified
> _toDelete	--	Aug 21, 2022 at 9:04...
> _toFile	--	Today at 12:35 PM
∨ Active	--	Jul 27, 2022 at 8:55...
> Other	--	Jul 26, 2022 at 5:40...
> Research	--	Oct 13, 2022 at 11:1...
> Service	--	Jan 27, 2023 at 11:0...
> SocLab	--	Sep 29, 2022 at 4:14...
> Students	--	Nov 2, 2022 at 4:30...
> Teaching	--	Dec 29, 2022 at 2:42...
> Workshops	--	Jul 27, 2022 at 8:55...
.DS_Store	8 KB	Today at 8:39 PM
> Administrative	--	Jul 26, 2022 at 9:30...
> Bookshelf	--	Jul 26, 2022 at 5:40...
> Templates	--	Jan 13, 2023 at 9:30...
> Vault	--	Oct 13, 2022 at 5:48...
> Vitae	--	Jan 2, 2023 at 4:16 PM
> Website	--	Jan 15, 2023 at 2:04...

Directory Structure

Third Level


 Active

 Research

 -OnHold

 -UnderReview

 RP1-ResProj1

 RP2-ResProj2

 RP3-ResProj3

 RP4-ResProj4

Research Project Directory Structure

Fourth Level



Active



Research



RP1-ResProj1



_posted



_toDelete



_toFile



admin



bookshelf



design



manuscript



presentation

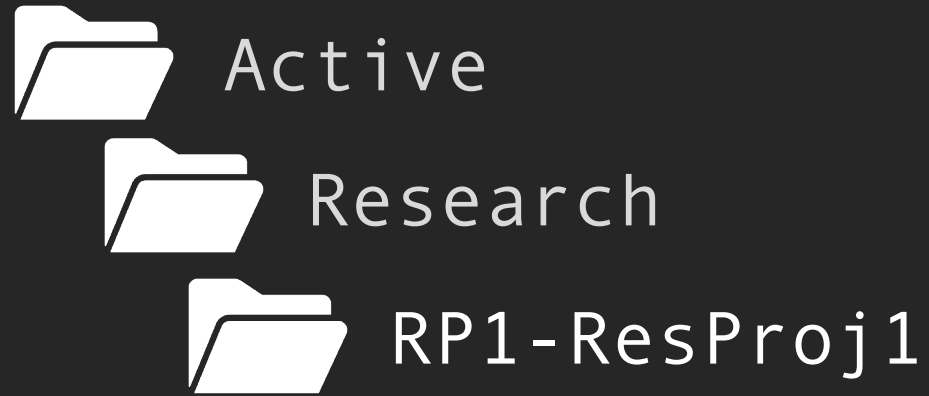












resources



work

Working Directory



-  _posted
-  _toDelete
-  _toFile
-  admin
-  bookshelf
-  design
-  manuscript
-  presentation
-  resources
-  work

Working Directory

When computing, folder that you pull files from & save files to

Where you set your file path



work

File path

Directions to your working directory

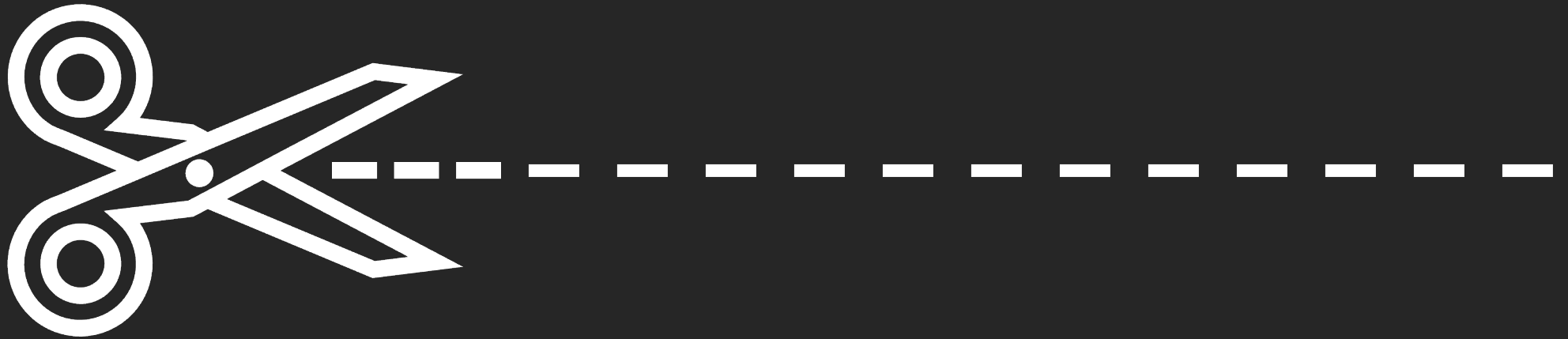
```
Users
├── admin
├── guest
└── biancamanago
    ├── project01
    └── project02
        └── work
            ├── _rawdata
            ├── -data
            ├── analysis
            └── management
```

```
cd "/Users/biancamanago/project02/work/"
setwd("/Users/biancamanago/project02/work")
```

Dual workflow

What is dual workflow?

Separate data management and data analysis

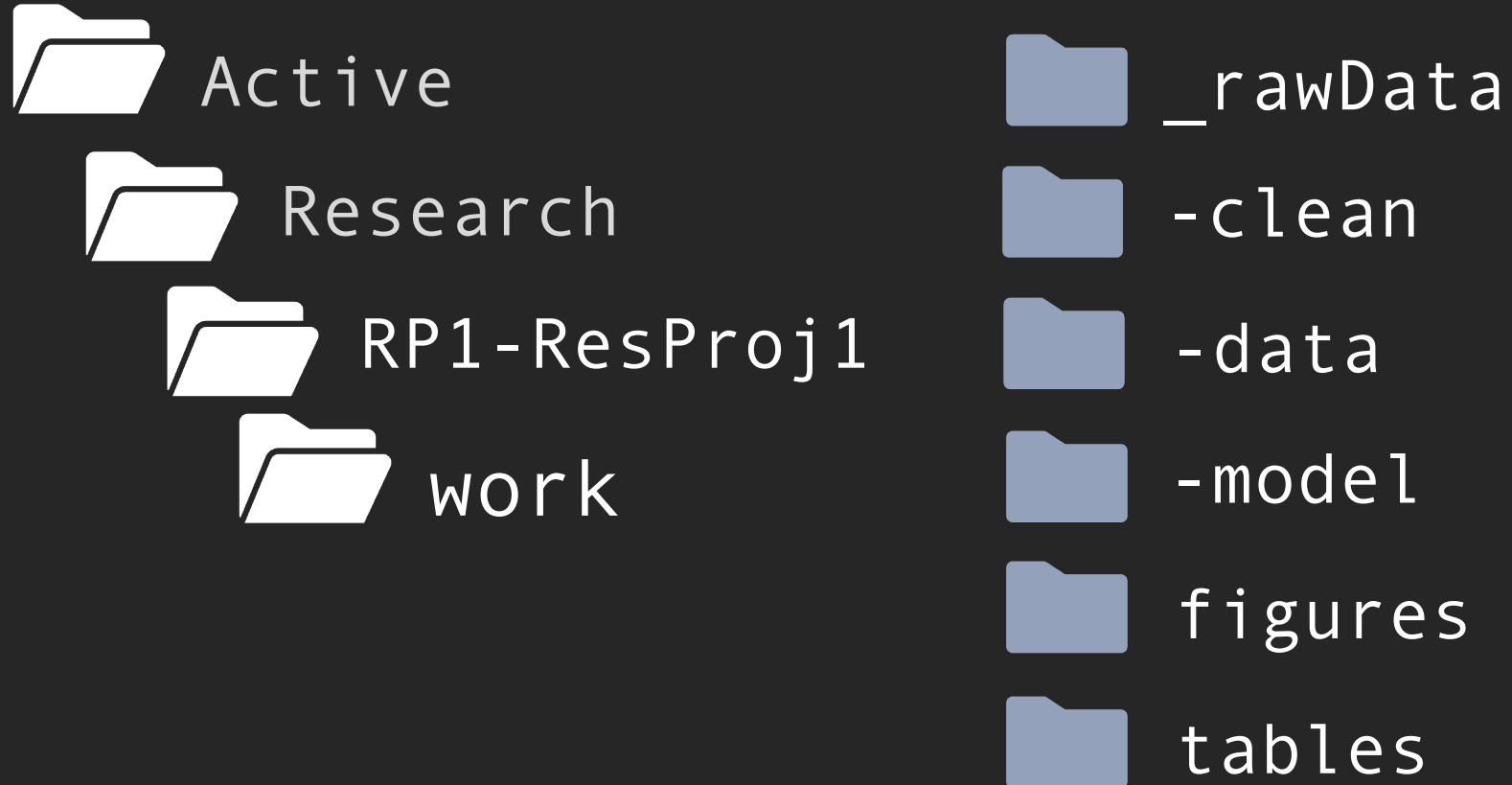


Why dual workflow?

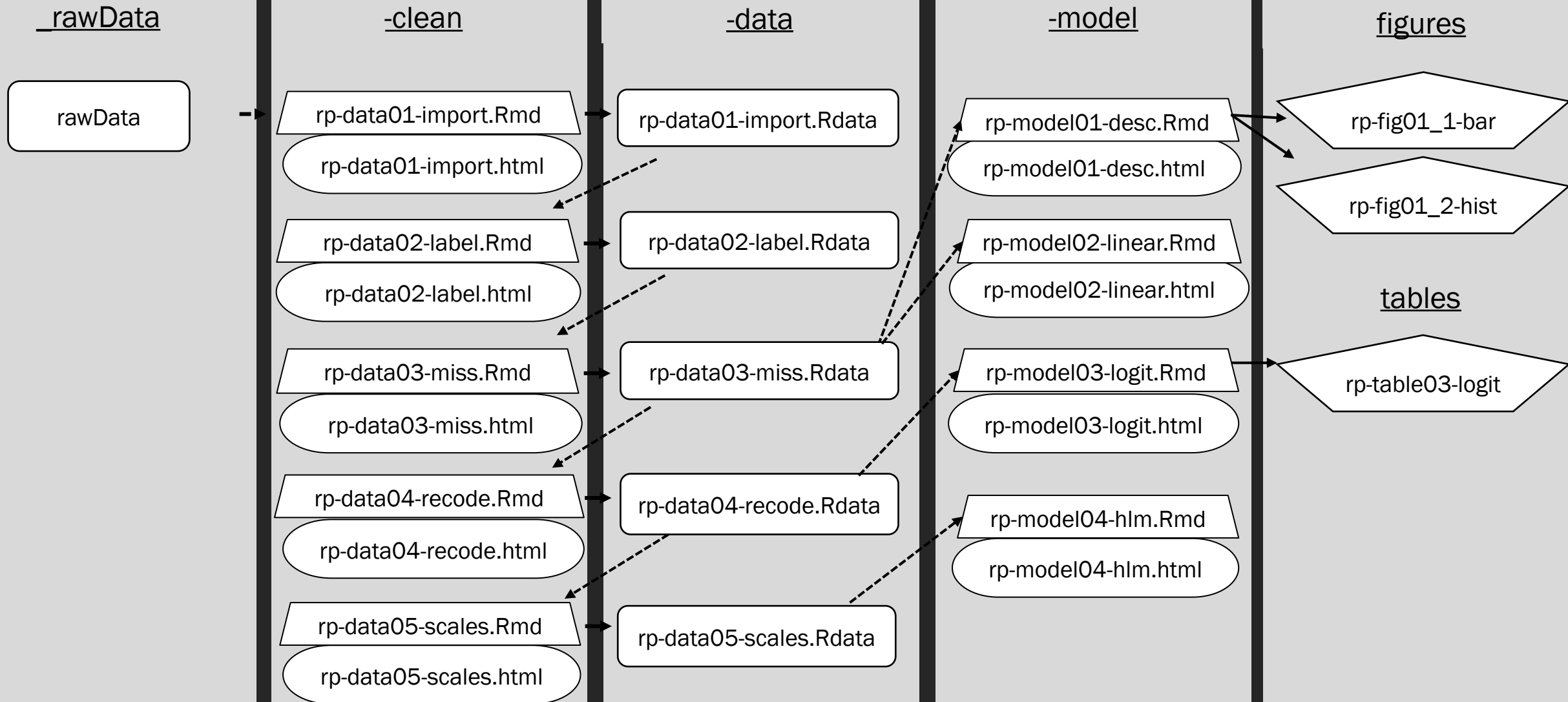
- Makes work more **efficient**
- Facilitates **replication**
- Prevents errors (**accuracy**)
- **Simplifies** organization and documentation
- Encourages **planning**

(Long 2009)

Dual workflow in practice



Dual Workflow in Work Folder



Dual workflow in practice

Data Cleaning/Preparing

rp-data01-import.do
rp-data02-drop.do
rp-data03-label.do
rp-data04-inspect.do
rp-data05-recode.do
rp-data06-outliers.do

Data Modeling/Analysis

rp-model01-descriptives.do
rp-model02-demog.do
rp-model03-ttest.do
rp-model04-anova.do
rp-model05-chi2.do
rp-model06-regression.do

Dual workflow in practice

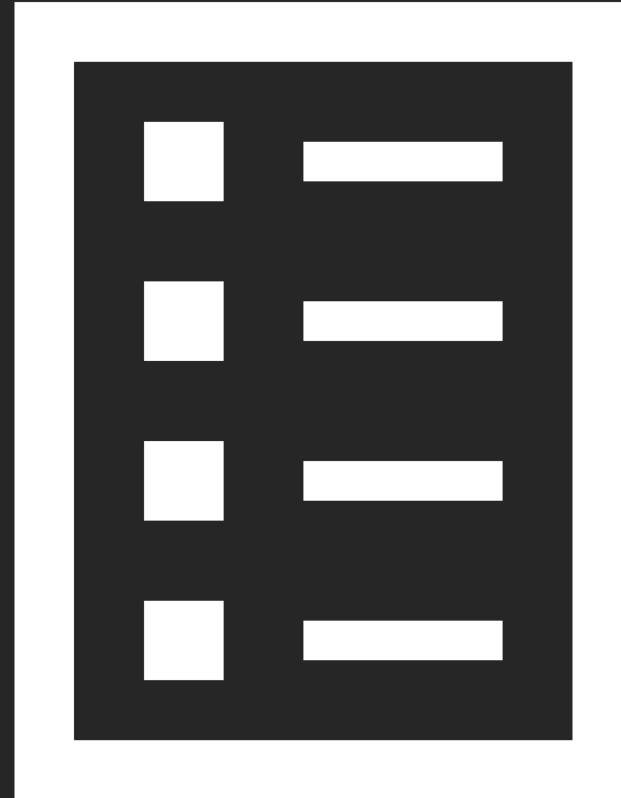
Why so many script files?

- Fewer errors
- Easier to debug
- Easier to track
 - what you did
 - where you did it

Dual workflow in practice

Keeping track of files

Use a primary script file



Dual workflow in practice

```
// Data Cleaning
```

```
do rp-data01-import.do  
do rp-data02-label.do  
do rp-data03-miss.do  
do rp-data04-recode.do  
do rp-data05-scales.do
```

```
// Data Analysis
```

```
do rp-model01-describe.do // uses rp-data02-label  
do rp-model02-linear.do // uses rp-data02-label  
do rp-model03-logit.do // uses rp-data03-miss  
do rp-model04-hlm.do // uses rp-data05-scales
```

2.2 Planning/Preregistration

Planning

Why plan?

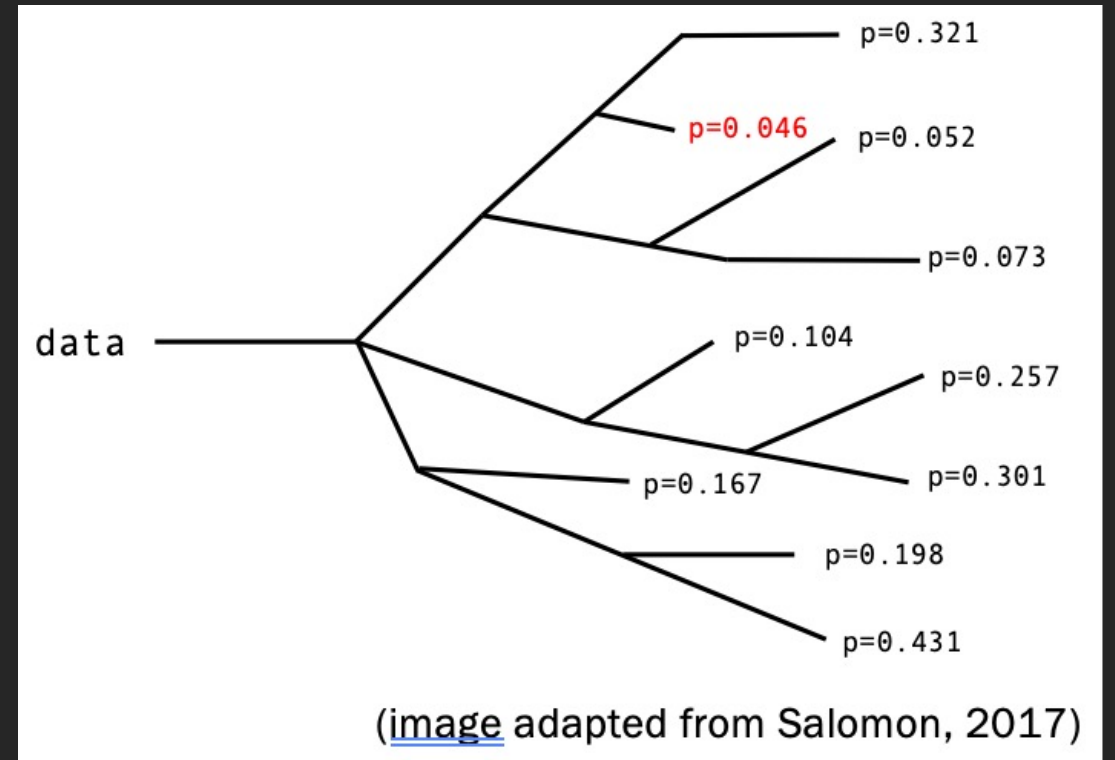
- Strategic
- Thoughtful
- Efficient
- Reduce multiple comparisons



Multiple Comparisons

Garden of Forking Paths

- Frequentist methods
- Effect was zero – still find difference
- <5% of time
- Over 100 comparisons



(Gelman and Loken 2014; Nosek, Ebersole, DeHaven, and Mellor 2018; Simmons, Nelson, and Simonsohn 2011)

What is preregistration?

Process of stating:

- hypotheses
- desired samples
- exclusion criteria
- study design
- analysis plans

on a public repository before embarking on a study

(Kavanagh and Kapitány 2019)



“Preregistration separates hypothesis-generating (exploratory) from hypothesis-testing (confirmatory) research. Both are important...

Center for Open Science 2021

... But the same data cannot be used to generate and test a hypothesis, which can happen unintentionally and reduce the credibility of your results.

Center for Open Science 2021

Addressing this problem through planning improves the quality and transparency of your research.”

Center for Open Science 2021

What isn't preregistration?

- A strict set of rules
- Preventor of sensitivity analyses
- Preventor of exploratory analyses
- A panacea for errors in scientific inquiry (Kavanagh, et al. 2019, p.12; Lakens 2019, p. 226)

Why use preregistration?

- Improves planning of studies
- Restricts researcher degrees of freedom
- Provides independent verification of a-priori predictions

(Kavanagh, et al. 2019)

How to pre-register a study?



More on preregistration

Forthcoming paper:

“Preregistration and registered reports in sociology: strengths, weaknesses, and other considerations.” The American Sociologist.

More on preregistration

Forthcoming paper:

“Preregistration and registered reports in sociology: strengths, weaknesses, and other considerations.” The American Sociologist.

shameless self plug

2.3 Transparency

Transparency

What is transparency?

- Makes research easy to reproduce
- Sharing information about study, e.g.:
 - Data
 - Data preparation methods
 - Sensitivity analyses
 - Planned and exploratory analyses

Tools for transparency

Publishing all study, data, and script files

Why?

- Others can follow all decisions
- Can reproduce & replicate work

Tools for transparency

Data preparation appendix

What it is

Place to describe data preparation decisions for reviewers and/or scholars that are interested in building on your work.

Tools for transparency

Data preparation appendix

Why it is needed

Data analysis decisions can be reviewed in-depth:

- during review process
- with publication

Tools for transparency

Data preparation appendix

What to include

- All decisions related to data preparation
- Reasoning for decisions
- Can copy plans from pre-registration and edit

Tools for transparency

Sensitivity analysis/robustness check appendix

What it is

- Appendix to report all additional tests (null or not)

Tools for transparency

Sensitivity analysis/robustness check appendix

Why it is needed

- Multiple tests
- Assess strength of findings

Tools for transparency

Sensitivity analysis/robustness check appendix

What to include

- ALL robustness checks
- In an organized fashion

2.4 Automation

What is automation?

Uses code to reduce:

- number of clicks
- lines of code
- need for copy and paste

Why automation?

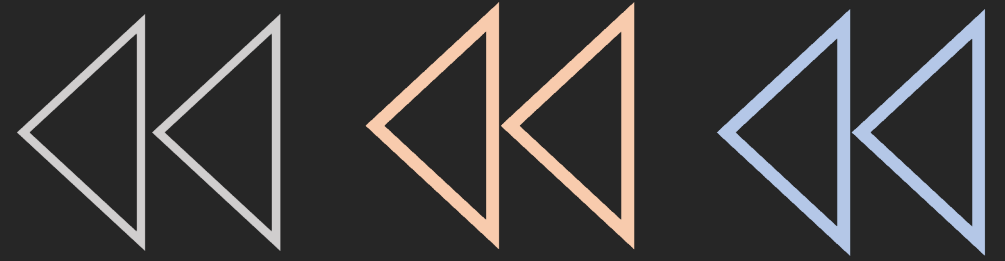
- Easier to see mistakes
- Faster to respond to changes (doing so in 1 place, rather than 100)
- Fewer bugs (but potentially bigger)
- Fewer points of origin for mistakes

Automation using software/packages

- Stata: putdocx/markstat
- R: Rmarkdown, etc.

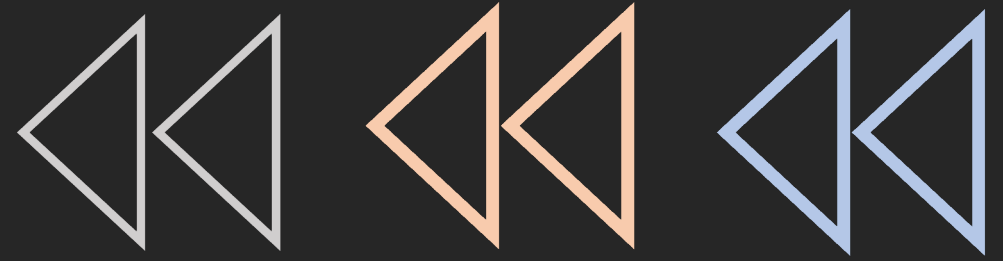
Automation using code

- for loops
- functions
- locals/vectors
- graph schemes
- table commands



Summary

Summary



- Good file organization
 - Directory structure
 - Dual workflow
- Planning/pre-registration
- Transparency
 - Publish script files
 - Data preparation appendix
 - Sensitivity analysis appendix
- Automation

Part 3. Steps and Order of Data Preparation

Overview

- Steps associated with data preparation
- Considerations for the order of steps
- Reconsider the rules/order
- Other tips

Data Cleaning Steps



Compile



Label/Name



Examine



Alter



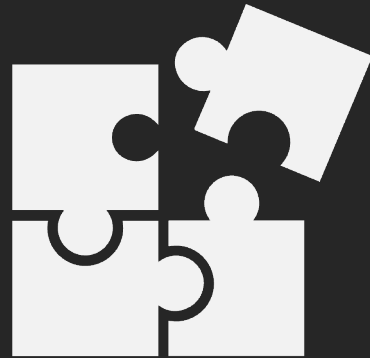
New variables



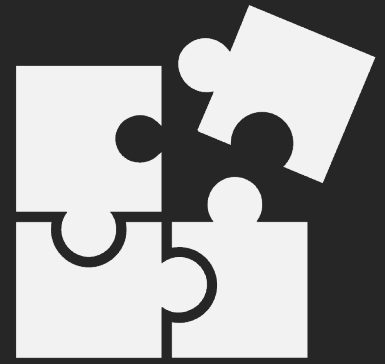
Re-configure & re-examine

Compiling

All the data you need, and none of the data you do not, are available in your data set.



Steps in compiling

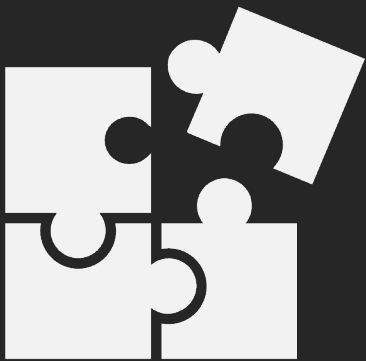


- importing data
- dropping completely irrelevant variables
- merging/ appending (sometimes later)
- reshaping data (sometimes later)

Notes about compiling

ID	var1	var2	var3
1	50	80	70
2	60	70	65
3	65	60	70
4	70	50	45
5	65	40	55

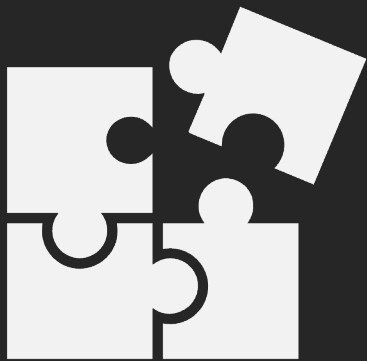
ID	num	var
1	1	50
2	1	60
3	1	65
4	1	70
5	1	65
1	2	80
2	2	70
3	2	60
4	2	50
5	2	40
1	3	70
2	3	65
3	3	70
4	3	45
5	3	55



Notes about compiling

id	var1	var2	var3
1	3	5	1
2	5	3	2
3	3	6	2
4	2	7	3
5	7	3	6

id	var4	var5
1	4	5
2	5	4
3	2	4
4	6	6
5	7	7

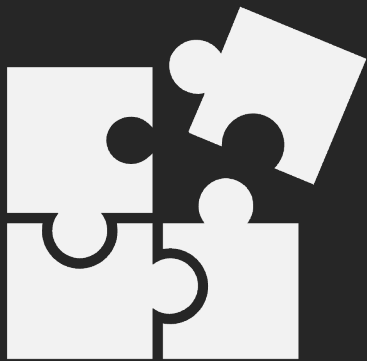


id	var1	var2	var3	var4	var5
1	3	5	1	4	5
2	5	3	2	5	4
3	3	6	2	2	4
4	2	7	3	6	6
5	7	3	6	7	7

Notes about compiling

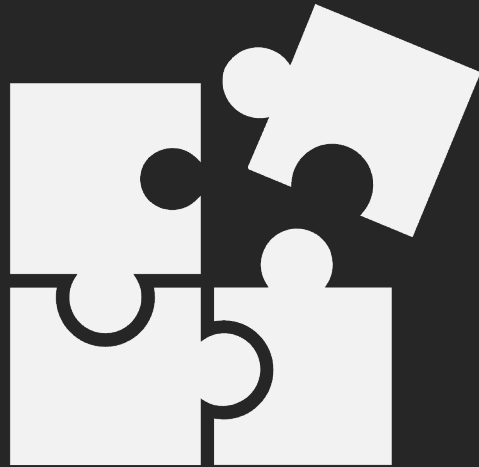
id	var1	var2	var3
1	3	5	1
2	5	3	2
3	3	6	2

id	var4	var5
1	4	5
4	6	6
5	7	7



id	var1	var2	var3	var4	var5
1	3	5	1	4	5
2	5	3	2	.	.
3	3	6	2	.	.
4	.	.	.	6	6
5	.	.	.	7	7

Notes about compiling



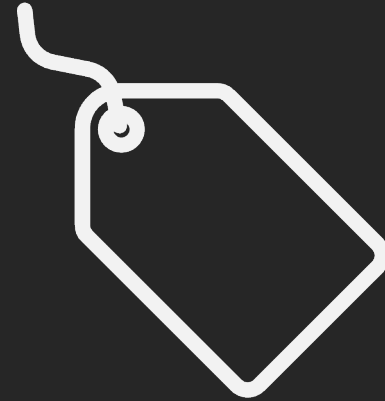
id	var 1	var 2	var 3
1	3	5	1
2	5	3	2
3	3	6	2
4	2	7	3
5	7	3	6

id	var 1	var 2	var 3
6	5	3	2
7	3	2	2
8	6	2	2
9	7	3	3
10	3	6	6

id	var 1	var 2	var 3
1	3	5	1
2	5	3	2
3	3	6	2
4	2	7	3
5	7	3	6
6	5	3	2
7	3	2	2
8	6	2	2
9	7	3	3
10	3	6	6

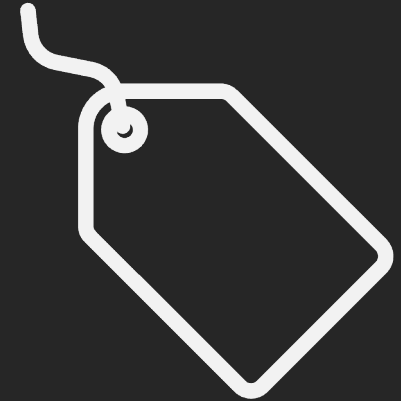
Labeling & Naming

Making it easy to see what you have (and find it again)



Steps in Labeling & Naming

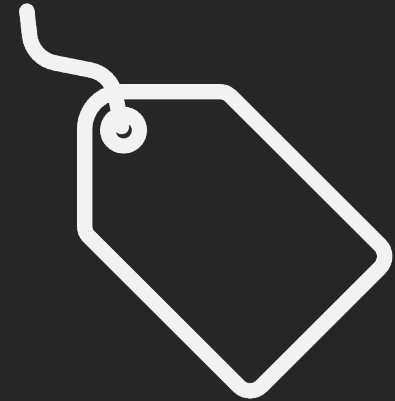
- (re)naming variables
- labeling variables
- labeling values
- encoding variables
- changing variable type (string, integer)



Do not recode in this step.

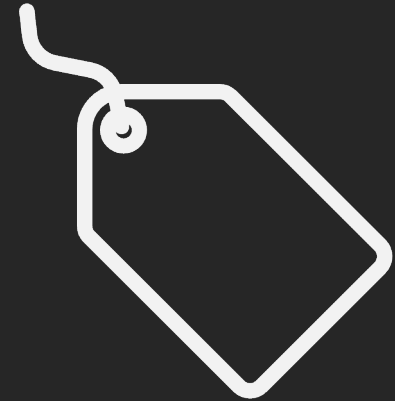
Notes about Labeling & Naming

- Use variable names that mean something
 - `age > q1`
 - `age > k2902`



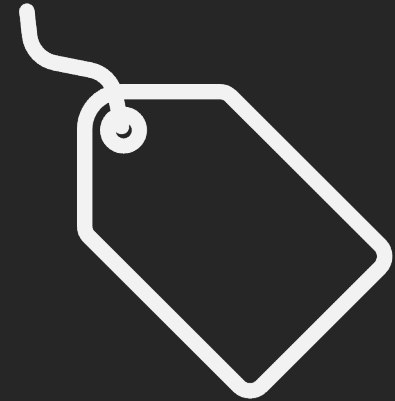
Notes about Labeling & Naming

- Difficult to choose good variable names
 - Take your time
 - Short but meaningful
 - `educ_yrs` > `educationinyears`



Notes about Labeling & Naming

- Label all variables
 - All of them, really
- Label values of categorical variables
 - Short and meaningful labels



Examining

Getting to know your data



Steps in Examining



- graphing
- range constraints
- descriptive statistics
- missing data
- outliers
- duplicates
- cross-validation

*Do not recode in this step.
Make note of all issues in Data Preparation Appendix.*

Notes on Examining



Reality checks

How would you know if your data wasn't quite right?

(Firebaugh 2008)

Examples:

- Do you have more people who graduated than were enrolled?
- Were people reporting their height in the same units?

Notes on Examining



Curious cases

Examine strange data in detail

Examples:

- If someone has an age of 5.8, think about why that might be the reported age. Did they mean 58? How can you be sure?
- Did people who reported being born male also report being on hormonal birth control? Why?



Notes on Examining

Missing data

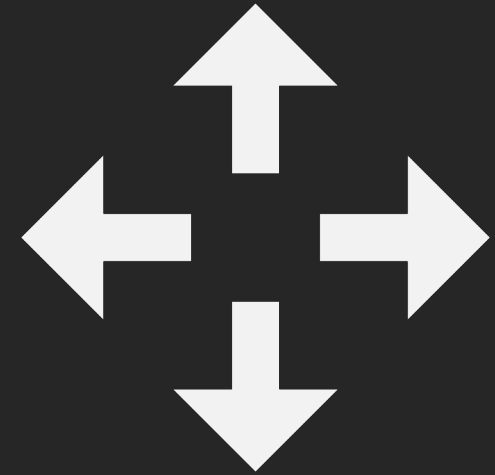
Determine why and what it means

Examples:

- Are a large portion of people are missing on a question? Was there some skip logic?
- What does it mean if people refused to answer about past mental illness?

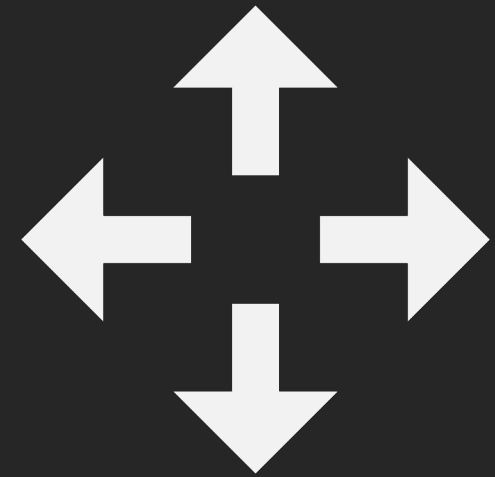
Altering

Ensure data is ready for
analysis



Steps in Altering

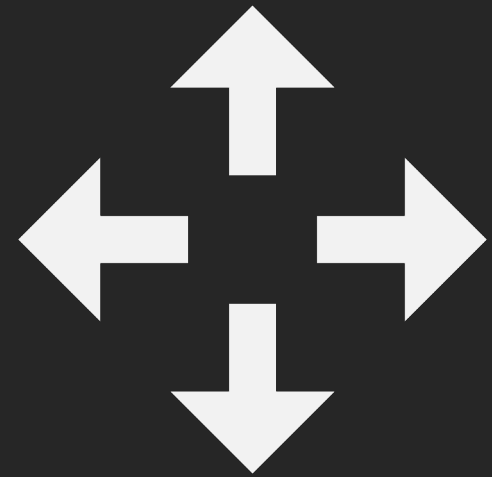
- recoding variables
- transforming variables
- removing duplicates
- addressing anomalous data
- address missing data



Plan your alterations.
Document in Data Preparation Appendix.

Notes on altering

- keep track of alterations
- do sensitivity analyses
- look up best practices



New Variables

Create any new variables of substantive interest





Steps in New Variables

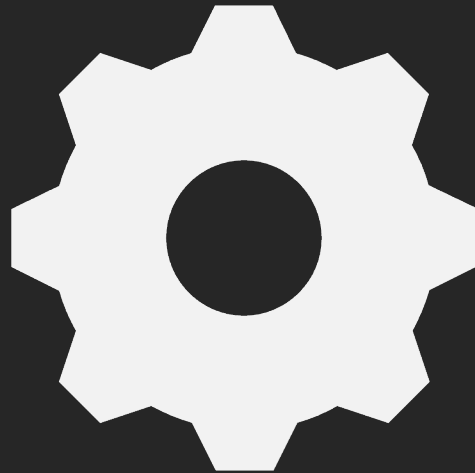
- Create scales
 - Using confirmatory and/or exploratory techniques
- Combine variables in substantive ways

Plan your variables.

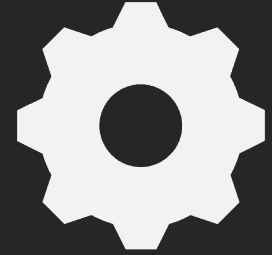
Document in Data Preparation Appendix.

Re-configure and...

Getting data into appropriate form for analysis



Steps in Re-configuring



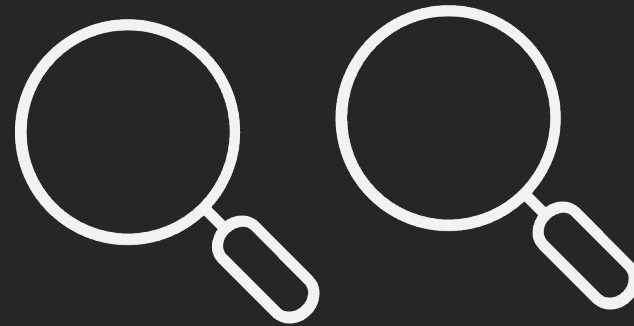
- Merging
- Reshaping

Plan your variables.

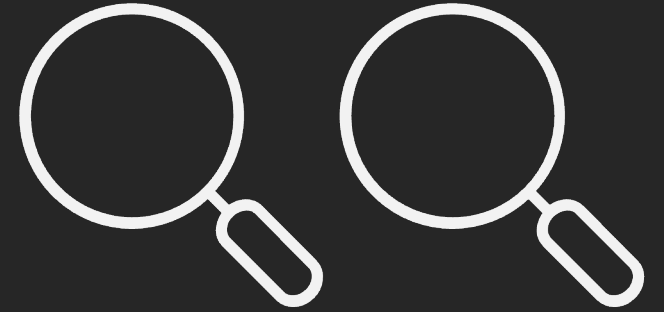
Document in Data Preparation Appendix.

Re-examine

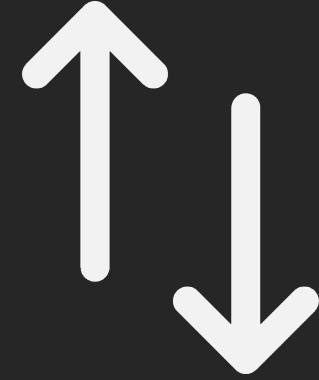
Again, ensuring data makes sense and alterations are appropriate.



Steps in Re-examining



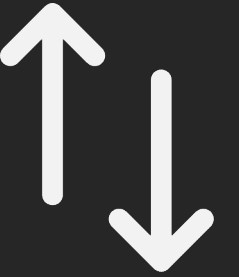
- Compare new variables against original variables.
- Make sure transformations seem correct.
- Re-examine missing data
 - How has your treatment (or lack thereof) of missingness changed your data?
 - Has your recoding changed your ideas about reasons for missingness?



Why in this order?

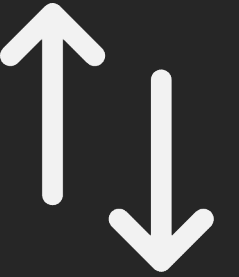
Pay attention to the reasons for the rules, rather than the rules themselves.

Why compile before label/name?



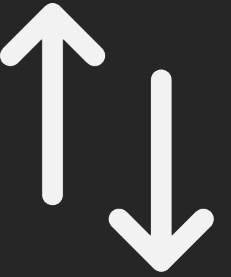
- May make naming easier
- Easier to clean data in long vs. wide form
- Easier to choose names when all data is in one place

Why label/name before examining?



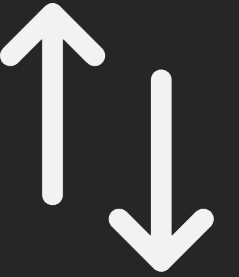
- Inaccurate labels make data examination difficult
- Value labels prevent interpretation errors
- Easier to type/remember

Why examine before altering/creating new variables?



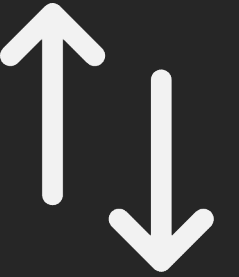
- Need to address/evaluate missing data before creating new variables
- Need to address outliers before using factor analysis-type techniques
- Need to address outliers before imputing missing data

Why re-configure towards the end?



- Need analysis-ready data in multiple formats (e.g., long and wide)
- Merging may be easier with clean data
- Duplicates and missing data cause problems for merging

Why re-examine towards the end?



- Re-check model assumptions after alternations
- Examine data after imputations
- Examine new variables

Reconsider the rules

Reconsider the rules

- The order of steps depends on your research question
- General guidelines, things to consider
- Be thoughtful, plan ahead

The rules of ordering

“The applied researcher who routinely adopts a traditional procedure without giving thought to its associated assumptions may unwittingly be filling the literature with nonreplicable results.”

- Keselman

(Keselman, Huberty, Lix, Olejnik, Cribbie, Donahue, Kowalchuk, Lowman, Petoskey, Keselman, and Levin (1998), p. 351; Osborne (2013a), p. 1)

Why not compile before label/name?

- Variable renaming may make merging datasets easier
- Perhaps do all CLEAN steps before re-configuring/merging data

Why create new variables during examination?

- Create indicator variables for missing data/outliers
- Create categorical variables for low-quality data
- Create substantive combo variables

When merging comes earlier:

- Multiple datasets needed to create a variable
- If merging addresses missing data
- If merging helps with imputation

When reshaping comes earlier:

- May help with data examination
- May assist with merging

Other tips and principles

Plan first

- CLEANR is a general framework
- Use it as a starting point
- Decide your end goal
- Move towards it

Prepare data in groups

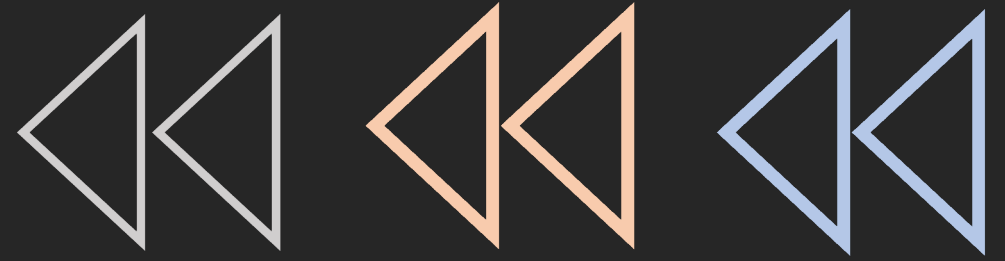
What are groups?

- demographic variables
- independent variables
- covariates
- dependent variables
- auxiliary variables

Prepare data in groups

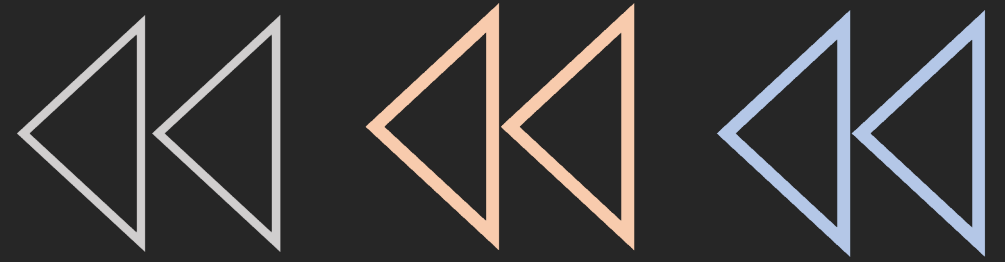
Why proceed in groups?

- Keep inventory of variables
- Look at associations between similar variables
- Thoughtful naming and sorting conventions



Summary

Summary



- Label and name variables/values clearly
- Understand your data (strengths and weaknesses)
- Make changes only after you understand what you'd be changing
- Consider workflow – make changes once (more later)

The last step will always be to re-examine your data

Sources

- Anvari, F., & Lakens, D. (2018). The replicability crisis and public trust in psychological science. *Comprehensive Results in Social Psychology*, 3(3), 266–286. <https://doi.org/10.1080/23743603.2019.1684822>
- Firebaugh, G. (2008). *Seven rules for social research*. Princeton University Press.
- Gelman, A., & Loken, E. (2014). The Statistical Crisis in Science. *American Scientist*, 102(6), 460–465.
- Hendriks, F., Kienhues, D., & Bromme, R. (2020). Replication crisis = trust crisis? The effect of successful vs failed replications on laypeople's trust in researchers and research. *Public Understanding of Science*, 29(3), 270–288. <https://doi.org/10.1177/0963662520902383>
- Kavanagh, C. M., & Kapitány, R. (2019). *Promoting the Benefits and clarifying misconceptions about Preregistration, Preprints, and Open Science for Cognitive Science of Religion*. [Preprint]. PsyArXiv. <https://doi.org/10.31234/osf.io/e9zs8>
- Keselman, H. J. et al. (1998). "Statistical Practices of Educational Researchers: An Analysis of Their ANOVA, MANOVA, and ANCOVA Analyses". En. In: Review of Educational Research 68.3, pp. 350-386.
- Lakens, D. (2019). *The Value of Preregistration for Psychological Science: A Conceptual Analysis* [Preprint]. PsyArXiv. <https://doi.org/10.31234/osf.io/jbh4w>
- Leahey, E., Entwisle, B., & Einaudi, P. (2003). Diversity in Everyday Research Practice: The Case of Data Editing. *Sociological Methods & Research*, 32(1), 64–89. <https://doi.org/10.1177/0049124103253461>
- Long, J. Scott. 2009. *The Workflow of Data Analysis Using Stata*. College Station, TX: Stata Press Books.
- Noble, W. S. (2009). A Quick Guide to Organizing Computational Biology Projects. *PLoS Computational Biology*, 5(7), e1000424. <https://doi.org/10.1371/journal.pcbi.1000424>
- Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences*, 115(11), 2600–2606. <https://doi.org/10.1073/pnas.1708274114>
- Osborne, J. W. (2013a). "Is data cleaning and the testing of assumptions relevant in the 21st century?" En. In: *Frontiers in Psychology* 4. ISSN: 1664-1078. DOI: 10.3389/fpsyg.2013.00370.
- Plesser, H. E. (2018). Reproducibility vs. Replicability: A Brief History of a Confused Terminology. *Frontiers in Neuroinformatics*, 11, 76. <https://doi.org/10.3389/fninf.2017.00076>
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. *Psychological Science*, 22(11), 1359–1366. <https://doi.org/10.1177/0956797611417632>
- Wingen, T., Berkessel, J. B., & Englich, B. (2020). No Replication, No Trust? How Low Replicability Influences Trust in Psychology. *Social Psychological and Personality Science*, 11(4), 454–463. <https://doi.org/10.1177/1948550619877412>