CHRISTOPHER PRENER, PH.D.
FALL, 2016

WEEK 03
LECTURE 04

QUANTITATIVE ANALYSIS

STRUCTURING DO-FILES FOR REPLICATION

AGENDA

1. Replication's Importance
2. Do-file "Stacks"
3. Structuring Master Files
4. Structuring Sub-files
5. The Zen of Stata

1 REPLICATION'S IMPORTANCE

FUNDAMENTAL QUESTION:

IF YOU HAD TO RE-CREATE A LARGE-SCALE STATISTICS PROJECT IN FIVE YEARS FROM MEMORY, COULD YOU DO IT?

PLANNING REPLICATION

1. Allows you to confident that you can return to projects after an (extended) absence and pickup where you left off.

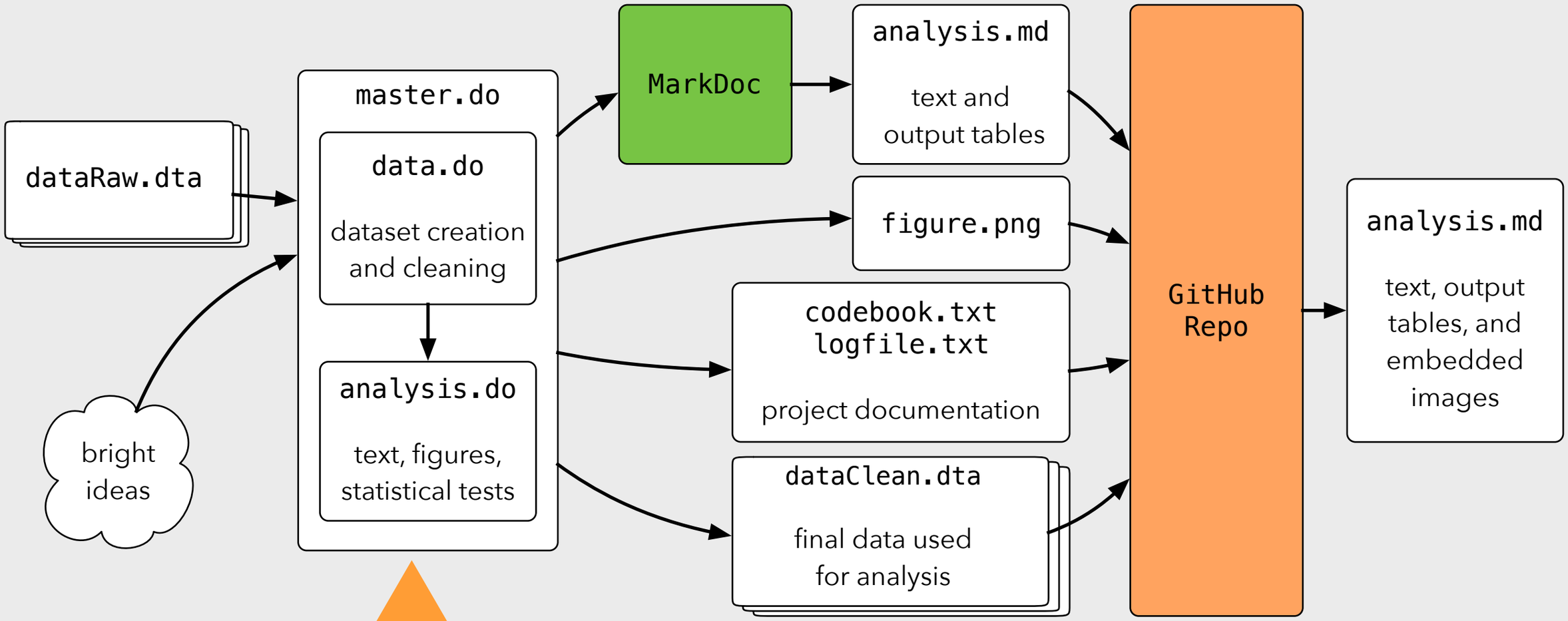
2. Makes the process of collaboration easier.

3. Allows you to easily make corrections or alterations to your work due to feedback / article reviews / new publications.

4. Allows you to easily share your process with others who seek to recreate / replicate / adapt your research.

2 DO-FILE “STACKS”

STATISTICAL WORKFLOW



A COMPLEX PROJECT MAY NEED MULTIPLE DATA AND ANALYSIS DO-FILES. I REFER TO THESE AS A "STACK" OF DO-FILES.

3 STRUCTURING MASTER FILES

STANDARD OPENING OPTIONS

```
. version 14

. log close _all

. graph drop _all

. clear all

. set more off

. set linesize 80
```

CONFIRM NECESSARY PACKAGES

```
. which markdoc
```

Tips:

- If you use other custom packages, add them on separate lines below the `Markdoc` package

CHANGE WORKING DIRECTORY

```
. if "`c(os)'" == "MacOSX" {
.  cd "/Users/`c(username)'/Documents/Working"
. }
. else if "`c(os)'" == "Windows" {
.  cd "E:\Users\`c(username)'\Documents\Working"
. }
```

Tips:

- Edit the file path for your primary OS so that it points to your Stats course “Working” folder where you stage files while they are in progress (see Long 2009)

CHECK DATA SOURCE

```
. global sourceData “[filePath/]sourceData.dta“

. capture confirm file $sourceData
{ ERROR CODE OMITTED }
```

Tips:

- Keep the full file paths for course data handy use in lieu of entering just the file name.
- Use previously cleaned data as a starting place!

CREATE PROJECT IN WORKING DIRECTORY

```
. global projName "projectName"

. capture mkdir $projName

. capture mkdir "$projName/CodeArchive"

. capture mkdir "$projName/Plots"
```

Tips:

- Keep directory names short (follow filename guidelines)

DOCUMENT, DOCUMENT, DOCUMENT

```
. log using “$rootDir/$projName.txt”, text replace
```

Tips:

- Log files are records of your output
- Save these as text files rather than in Stata’s default `.smcl` file format
- Plain text can be opened by virtually all text editing applications across all major operating systems - important for reproducibility.

DOCUMENT, DOCUMENT, DOCUMENT

```
/*

file name - master.do


project name - SOC5050: Quantitative Analysis, Fall 2016

purpose - /*what does this do-file do?*/

created - /*what is today's date?*/

updated - /*what is today's date?*/

author - /*what is your name?*/

*/
```

DOCUMENT, DOCUMENT, DOCUMENT

```
/*

superordinates  -

- /*enter the source data filename here*/.dta

*/
```

```
/*

subordinates -

- data.do

- analysis.do

*/
```

DOCUMENT, DOCUMENT, DOCUMENT

```
/*

full description -

/*enter a description of what this do-file does*/

*/


/*

updates -


*/
```

MAKE A COPY OF YOUR SOURCE DATA

```
copy sourceFile newFile
```

```
. global $newData “[filepath/]newDataName.dta"
. copy $sourceData "$rootDir/$newData", replace
```

Tips:

- Multiple copies of your data, when clearly differentiated, have limited downsides given the relatively cheap price of hard drive storage.
- Large projects may require external hard drives for storage.

EXECUTE OTHER DO-FILES

```
do filename.do
```

```
. do data.do
. do analysis.do
```

Tips:

- Make sure **all** of these files are saved in the working directory.

ARCHIVE CODE

```
copy "sourceFile" "newFile"
```

```
. copy "master.do" "$rootDir/Code/master.do"
. copy "data.do" "$rootDir/Code/data.do"
. copy "analysis.do" "$rootDir/Code/analysis.do"
```

Tips:

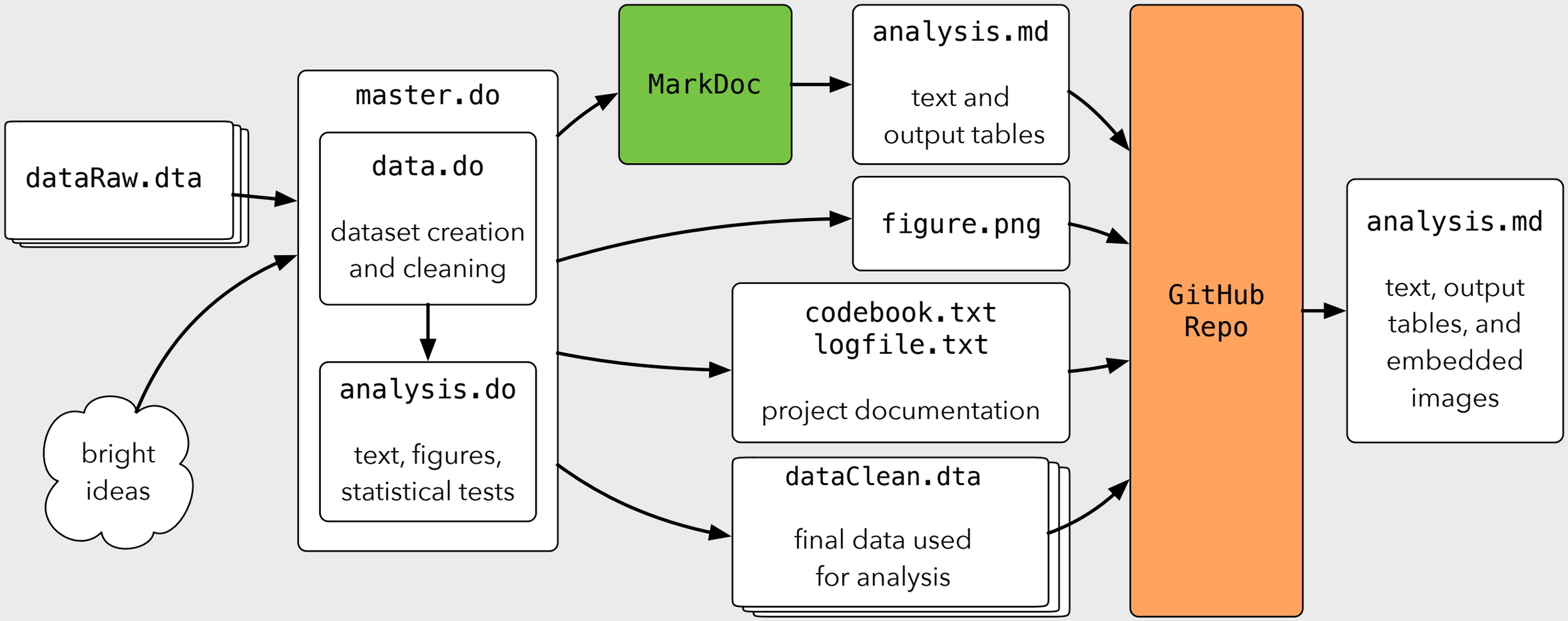
- After you execute this do-file stack, copies of the three do-files will be archived in your project's folder in the working directory.
- Once you post your files, you can delete the original do-files.

STANDARD CLOSING OPTIONS

```
. log close _all

. graph drop _all

. set more on

// ================================================================================

. exit
```

4 STRUCTURING SUB-FILES

STATISTICAL WORKFLOW



Tips:

- Each sub-filed needs the `set more off` and `set linesize` opening options as these preferences are not passed from file to file.
- The file `analysis.do` file also needs the MarkDoc commands!

USING ATOM EFFECTIVELY

- The `Week-03` repository contains a file named `stata-snippets.cson` that can be loaded into Atom (open this file in Atom and read the top for instructions!)

- After you change the language of your file to `Stata`, type "`head`" and select from the auto-complete dropdown:

    - `headMaster` for a `master.do` file template
    - `headData` for a `data.do` file template
    - `headAnalysis` for an `analysis.do` file template

5 THE ZEN OF STATA

A COUPLE SUGGESTIONS

- Explicit is better than implicit - spell out commands fully and use comments broadly.
- Readability counts - wrap lines to 80 characters and use /// to accommodate very long commands.

SEE THE COURSE USER'S GUIDE FOR DESCRIPTION OF 'THE ZEN OF STATA'

COMMENTS

```
// this is a comment


/*
This is a multi-line comment.

*/
```

DIVIDER LINES

```
//  ==============================================================
//  ==============================================================
//  ==============================================================
```

Level 1

```
//  ==============================================================
//  ==============================================================
```

Level 2

```
//  ==============================================================
```

Level 3

```
//  ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
```

Level 4

TEXT WRAPPING

```
summarize var1 var2 var3 var4 var5 var6 var7 var8 var9 var10 var11 ///
var12 var13 var14 var15
```

DOCUMENT DETAILS

Document produced by Christopher Prener, Ph.D for the Saint Louis University course SOC 5050: QUANTITATIVE ANALYSIS - APPLIED INFERENTIAL STATISTICS. See the course wiki and the repository README.md file for additional details.