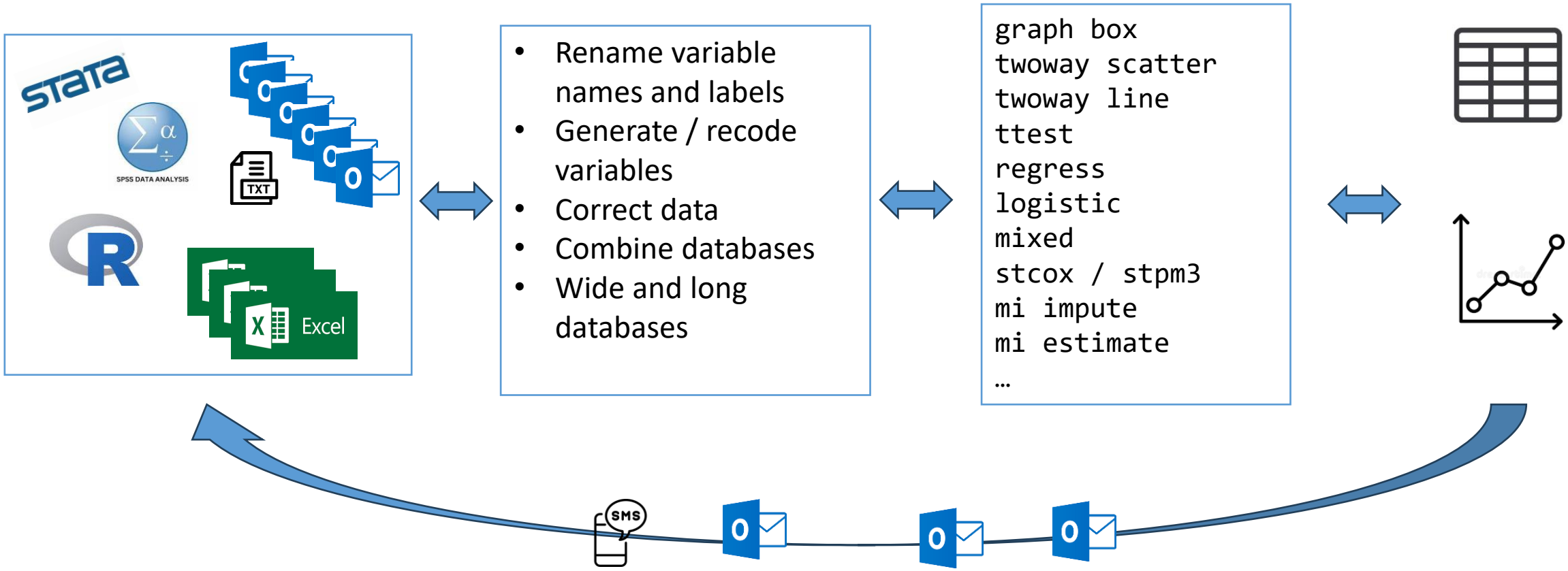
Are Hugo Pripp

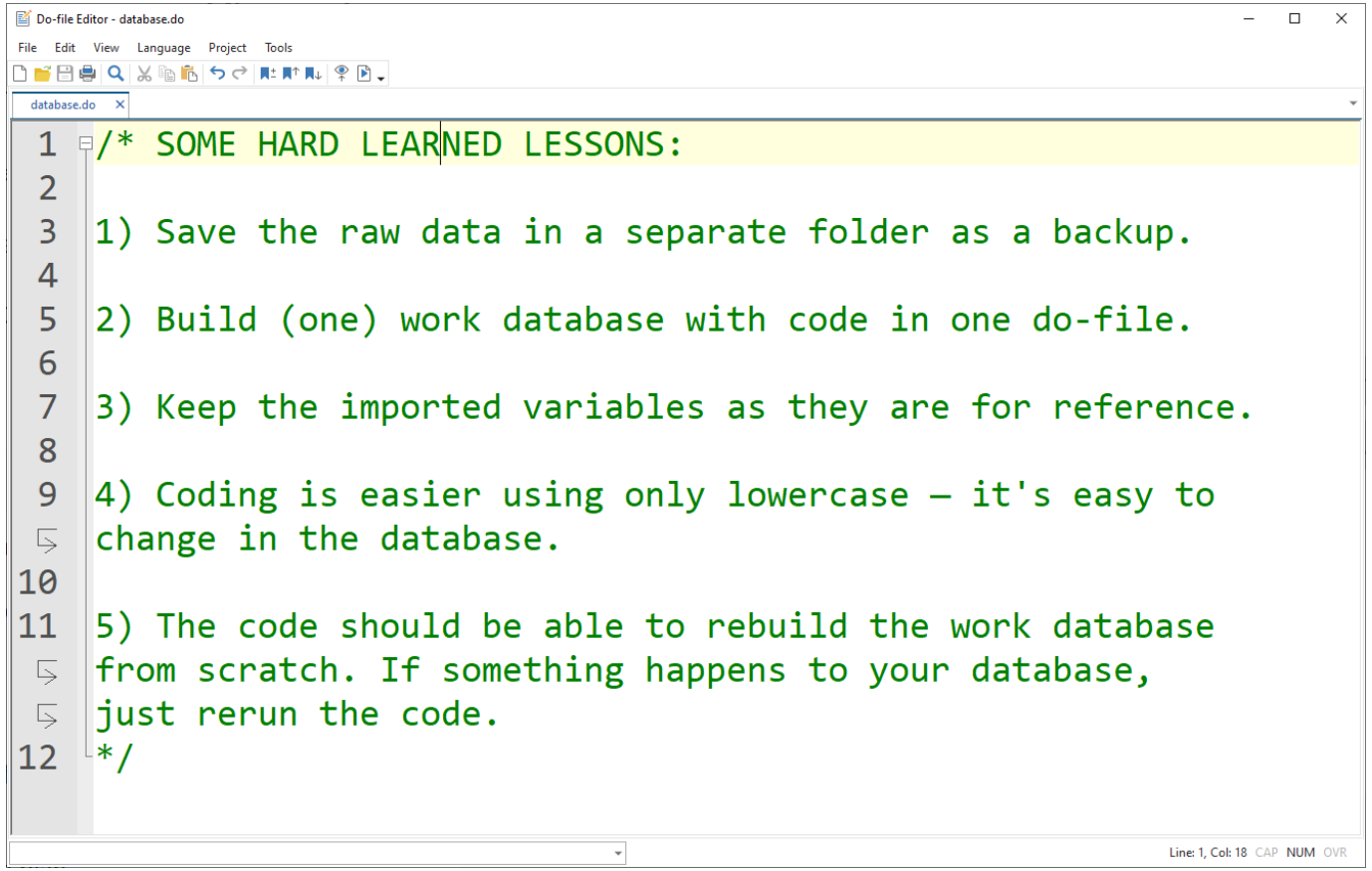
Using Stata with many datasets, methods, and variables

ChatGPT image generator

Workflow

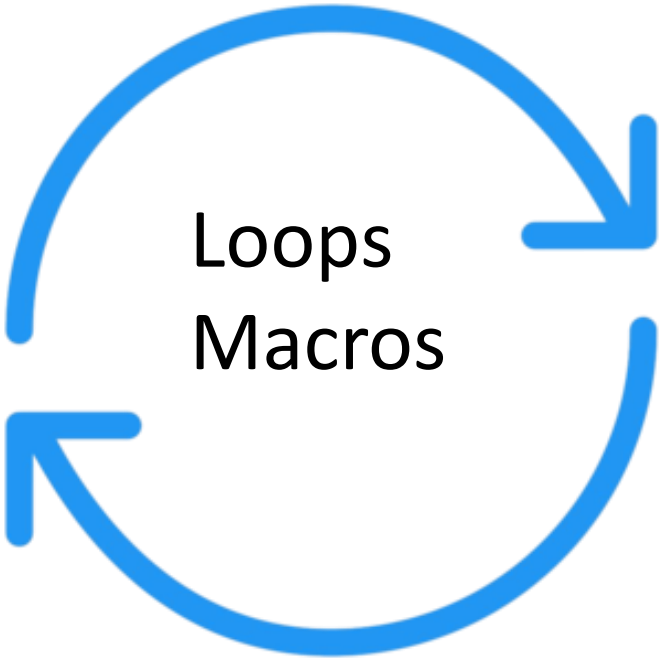

STATA
SPSS DATA ANALYSIS
TXT
R
Excel
• Rename variable names and labels
• Generate / recode variables
• Correct data
• Combine databases
• Wide and long databases
graph box
twoway scatter
twoway line
ttest
regress
logistic
mixed
stcox / stpm3
mi impute
mi estimate
…
SMS

Building the database(s)



- use
- import excel
- import spss
- duplicates
- merge
- append
- generate
- replace
- rename
- label variable
- label define
- label values
- drop
- keep

DRY or WET



- DRY (Don’t Repeat Yourself)
- WET (Write Everything Twice)

- A DRY STATA
  - `foreach – Loop of items`
  - `forvalues – Loop over consecutive values`
  - `local (define macro)`

- `STATA Project Manager`

Export results to an Excel file

```
. regress mpg weight foreign
      Source |       SS       df       MS      Number of obs   =        74
-------------+------------------------------   F(2, 71)        =     69.75
       Model |  1619.2877     2  809.643849   Prob > F        =    0.0000
    Residual |  824.171761    71  11.608053   R-squared       =    0.6627
-------------+------------------------------   Adj R-squared   =    0.6532
       Total |  2443.45946    73  33.4720474   Root MSE        =    3.4071

------------------------------------------------------------------------------
         mpg |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      weight |  -.0065879   .0006371   -10.34   0.000    -.0078583   -.0053175
     foreign |  -1.650029   1.075994    -1.53   0.130    -3.795505    .4954422
       _cons |   41.6797   2.165547    19.25   0.000     37.36172    45.99768
------------------------------------------------------------------------------
```

putexcel

dtable – Create a table of descriptive statistics

A case study

- Many outcome variables
- Several follow-up time points
- “It would be nice to have”:
  - Basic descriptive graphs and statistics for all outcome and time points
  - T-tests and ANCOVA (adjusted for covariates) at each time point, in addition to mixed models for repeated measurements.
- Manuscript in progress

Do-file Editor - example.do*

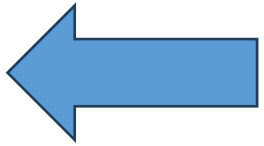
File Edit View Language Project Tools

example.do* database.do* Untitled

```
putexcel set "results.xlsx", replace

putexcel E1 = "Descriptive statistics", bold
putexcel I1 = "Linear mixed model of repeated measurement", bold
putexcel Y1 = "ANCOVA", bold
putexcel AO1 = "Regression t-test", bold

local row = 2
putexcel A`row' = "Measurement"
putexcel B`row' = "Time of observation"
putexcel C`row' = "Group"
*D
putexcel E`row' = "Number of observation"
putexcel F`row' = "Mean"
putexcel G`row' = "SD"
*H
*I
putexcel J`row' = "Mean"
putexcel K`row' = "Standard error"
putexcel L`row' = "z"
putexcel M`row' = "P-value"
putexcel N`row' = "ll 95ci"
putexcel O`row' = "ul 95ci"
*P
*Q
putexcel R`row' = "Mean difference"
putexcel S`row' = "Standard error"
putexcel T`row' = "z"
putexcel U`row' = "P-value"
putexcel V`row' = "ll 95ci"
```



Structure the Excel sheet

Do-file Editor - example.do*

File Edit View Language Project Tools

example.do* database.do* Untitled

```
67  local timeobs timeobs2
68  local rand ran_trt
69  local outcome pain strength bloodpressure bmi /// (and many more...)
70
71  foreach y of varlist `outcome'   {
72
73  * Boxplot
74  graph box `y' , over(`rand', label(angle(vertical) labsize(vsmall)) gap(40))
    over(`timeobs', label(labsize(small)))  ytitle("`y'- `:variable label `y''",
    size(small)) ylabel(, angle(horizontal))
75  graph export "`y' boxplot.png", as(png) width(2000) replace
76
77  mixed `y' `y'_bl i.`timeobs'##i.`rand'  if `timeobs' > 1  || subjectid: , base
78  margins i.`timeobs'#i.`rand'
79  marginsplot,  title("") ytitle("`y' - `:variable label `y''", size(small))
    ylabel(, angle(horizontal)) xtitle("")  legend( rows(1) position(6))
80  graph export "`y' marginsplot.png", as(png) width(2000) replace
81
82  * Baseline
83  local t 1
84  local row1 = `row' + 1
85  local row2 = `row' + 2
86  local row = `row2'
87
88  capture: putexcel A`row1' = "`y' - `:variable label `y''"
89  capture: putexcel A`row2' = "`y' - `:variable label `y''"
90
91  capture: putexcel B`row1' = "`:label (`timeobs') `t''"
92  capture: putexcel B`row2' = "`:label (`timeobs') `t''"
93
```

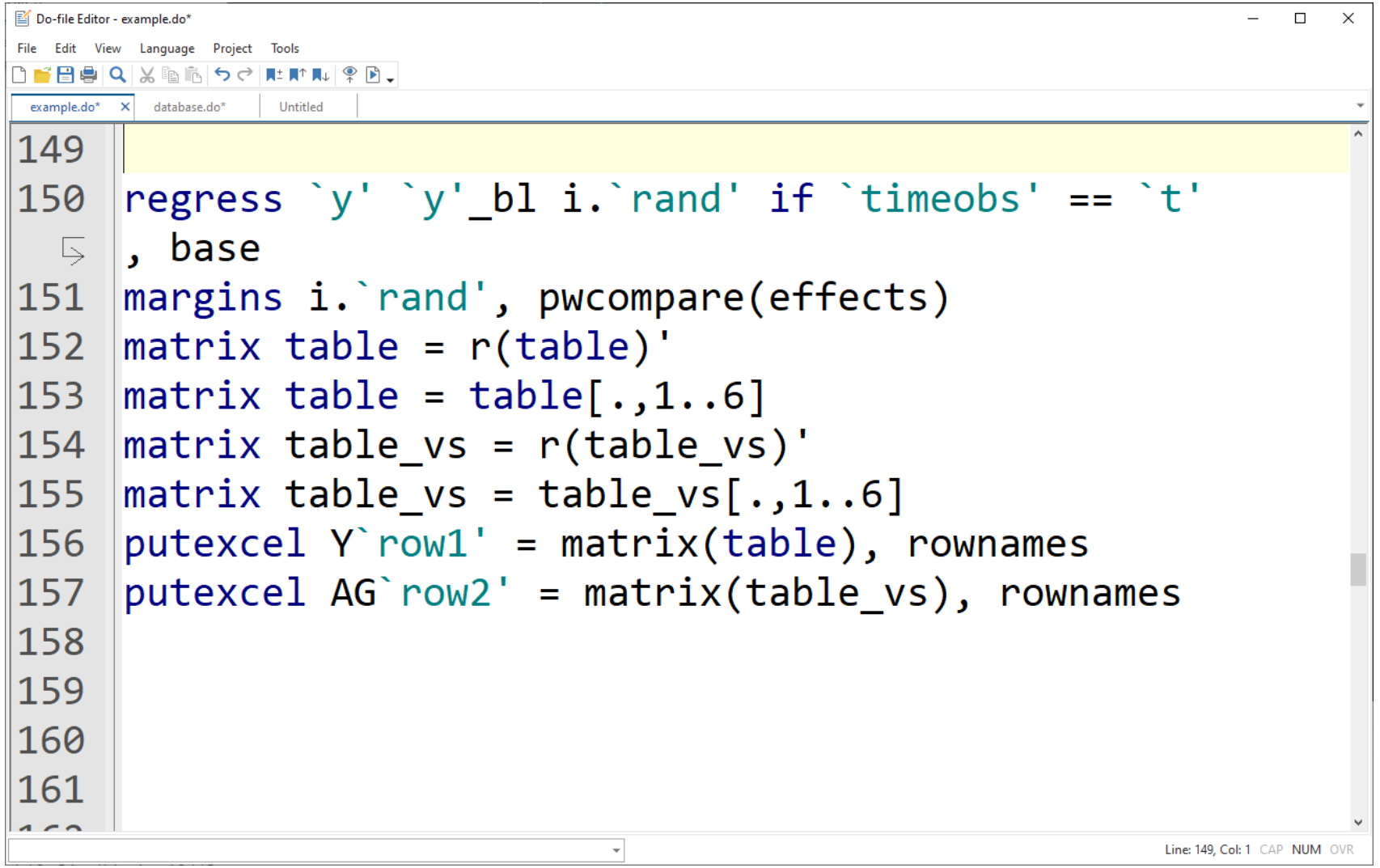
Line: 70, Col: 1 CAP NUM OVR

Local marcos for flexibility and DRY coding

Loop through all outcome variables

- Make and save graphs with variable label / name
- Run statistical model

Export to Excel sheet
Shift rows for each new analysis



Analyses looped within outcome variable and time points

`return – Return stored results`

Export selected results to Excel sheet

Reporting the results

Excel document with selected results					
Y	Time	Group	N	mean	sd
pain	1	Trmt	50	8.5	2.1
pain	1	Ctrl	49	8.2	2.0
pain	2	Trmt	48	6.3	2.0
pain	2	Ctrl	49	7.3	2.3
pain	3	Trmt	47	5.1	1.9
pain	3	Ctrl	42	7.0	2.2
pain	4	Trmt	45	4.2	2.0
pain	4	Ctrl	40	7.1	2.1

List of graphical files
pain boxplot.png
pain marginsplot.png
strength boxplot.png
strength marginsplot.png
bmi boxplot.png
bmi marginsplot.png
…

A case study

Hi,

I’m sorry, but I found a mistake in the journal for patient 152.

Could you please correct the data and check the results?

- Update the do-file for the database.
- Rerun the database do-file.
- Run the analysis do-file to "automatically" export results to Excel or other formats.
- No need for "copy and paste.“
- Ensure traceability and reproducibility.

Concluding remarks

ChatGPT image generator

Data management with traceable and reproduceable results

– The Unsong Hero of Statistics