

# Playing nice with others: Initializing your work with external configurations

Sven O. Spieß (*sven[dot]spiess[at]dpc-software[dot]de*)

DPC Software GmbH



# Motivation

- Encourage Stata users to think in terms of re-usable (software) applications
- Facilitate quality of collaboration  
(with others/across programs)
- ...by demonstrating three ways of implementing easily accessible setups/configurations in Stata, namely (a) macros within dedicated do-files, (b) INI files, and (c) Excel Sheets
- Promote discussion of good (and maybe less good) practices  
...and offer some humble suggestions

# Agenda

1. An Innocuous Example («*Use Case*»)
2. Four Plus X Approaches
3. Summary
4. Discussion

# An Innocuous Example

- Grading single-choice exams
- Three crucial parts:
  1. Student answers
  2. Model solution correct choices
  3. Grading, i.e. analysis
- Outputs, e.g.
  - List of grades
  - Item statistics/figures (distributions, selectivities, ...)
- Fictitious data of 15 undergrad students of dramatics taking a 10 question sc-exam

→ *data*

→ «***settings***»

→ *do-file*

# An Innocuous Example – Exam Data

Data Editor (Browse) - [omr\_exam shksp 2021-01]

File Edit View Data Tools

name[1] Margaret

	name	ID	question_1	question_2	question_3	question_4	question_5	question_6	question_7	question_8	question_9	question_10	
1	Margaret	34463	c	d	c	b	d	e	a	b	c	c	
2	Conrad	84768	c	c	c	c	b	c	a	e	d	c	
3	Don John	17098	e	c	c	c	e	c	b	b	b	a	
4	Ursula	80821	c	c	c	b	e	c	b	b	e	a	
5	Balthasar	96078	c	c	c	c	b	c	b	d	e	a	
6	Hero	53733	c	c	c	b	b	c	b	b	b	a	
7	Leonato	59459	c	c	c	b	e	c	b	d	b	d	
8	Dogberry	61549	c	c	c	c	e	c	c	b	e	a	
9	Beatrice	43237	c	c	c	c	b	c	c	b	e	a	
10	Verges	56603	c	c	c	c	b	c	c	b	d	a	
11	Benedick	74808	c	c	c	c	b	c	c	b	e	a	
12	Antonio	33809	c	c	c	c	e	c	c	b	e	a	
13	Don Pedro	31585	c	c	c	c	e	c	d	b	e	a	
14	Claudio	97777	a	c	c	b	b	c	d	a	b	d	
15	Borachio	02110	c	c	c	c	e	c	d	b	d	a	

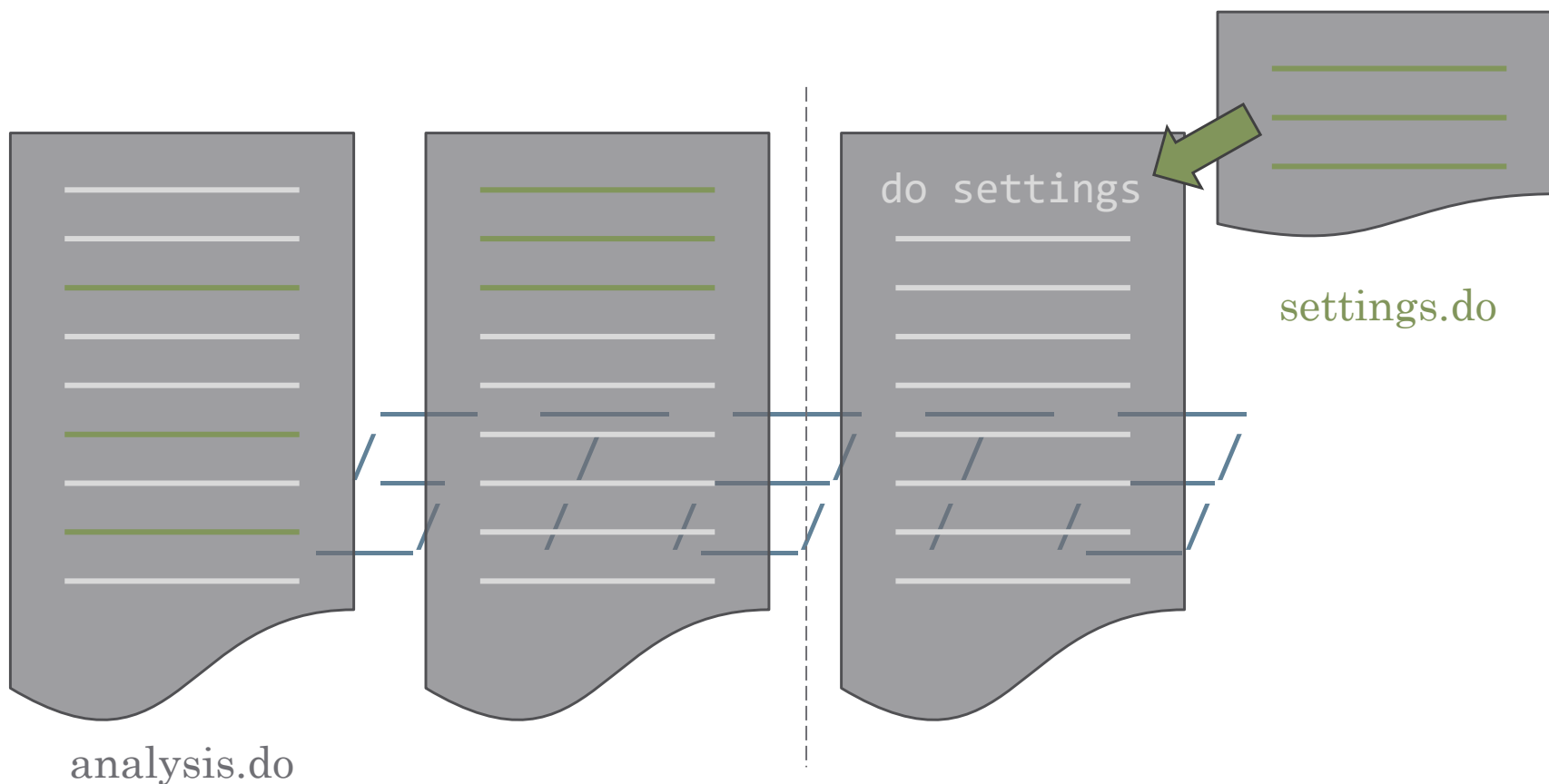
Ready Length: 9 Vars: 12 Order: Dataset Obs: 15 Filter: Off Mode: Browse CAP NUM

# Four Plus X Approaches – Stata Do-Files

- Macro variables go-to feature to define things we want to reuse again (and again)
  - `global question_1 = "c"`
- Compatible to general programming principle of key-value pairs
- In simple cases issue than becomes a matter of organizing our do-files

# Four Plus X Approaches – Stata Do-Files

```
. global question_1 = "c"
```

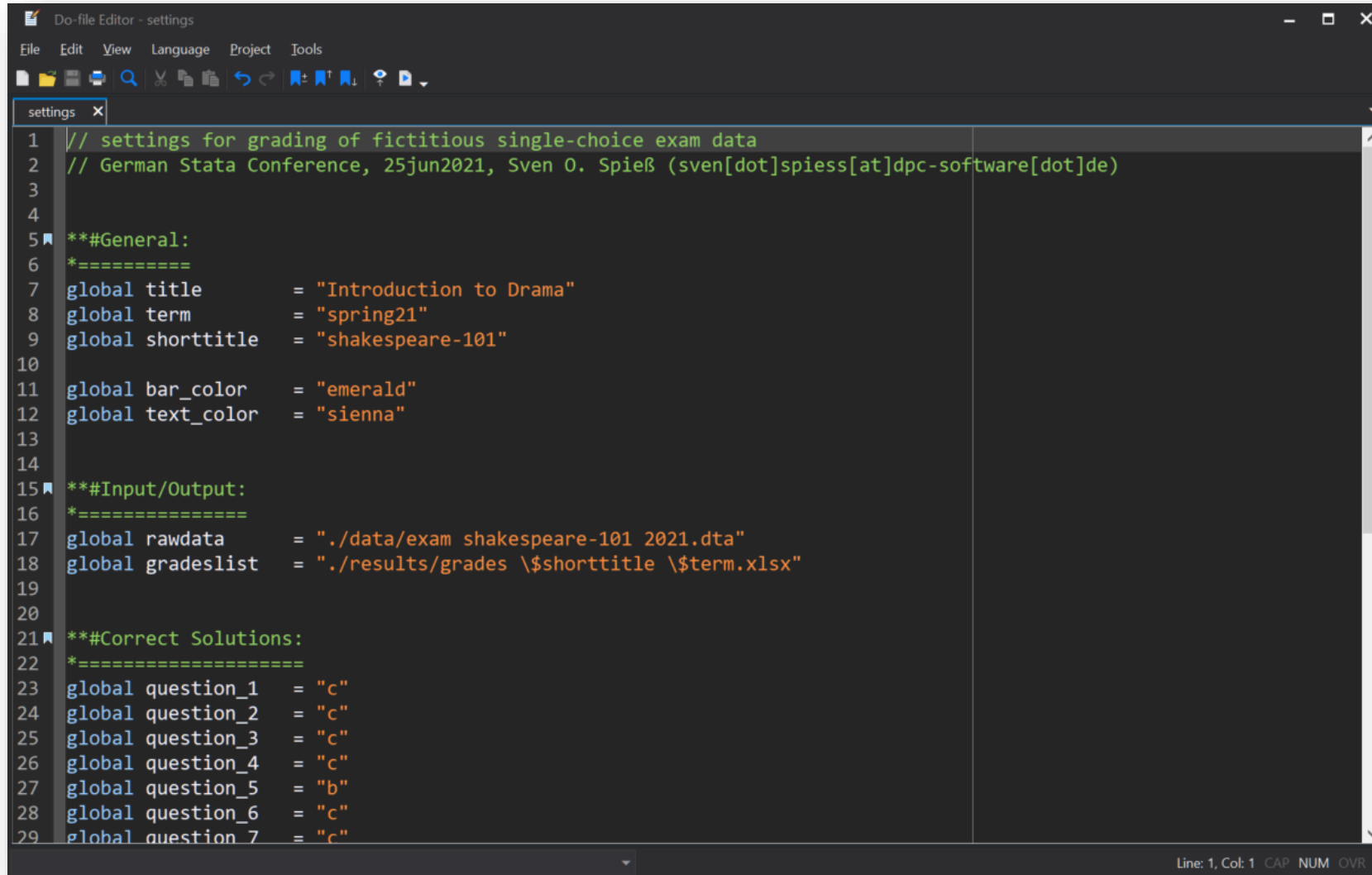


(a) «*just in time*»

(b) «*top-loading*»

(c) «*outsourcing*»

# Four Plus X Approaches – Stata Do-Files



```
Do-file Editor - settings
File Edit View Language Project Tools

settings x
1 // settings for grading of fictitious single-choice exam data
2 // German Stata Conference, 25jun2021, Sven O. Spieß (sven[dot]spiess[at]dpc-software[dot]de)
3
4
5 **#General:
6 *=====
7 global title = "Introduction to Drama"
8 global term = "spring21"
9 global shorttitle = "shakespeare-101"
10
11 global bar_color = "emerald"
12 global text_color = "sienna"
13
14
15 **#Input/Output:
16 *=====
17 global rawdata = "./data/exam shakespeare-101 2021.dta"
18 global gradeslist = "./results/grades \${shorttitle} \${term}.xlsx"
19
20
21 **#Correct Solutions:
22 *=====
23 global question_1 = "c"
24 global question_2 = "c"
25 global question_3 = "c"
26 global question_4 = "c"
27 global question_5 = "b"
28 global question_6 = "c"
29 global question_7 = "c"

Line: 1, Col: 1 CAP NUM OVR
```

*Outsourcing example «settings.do»*

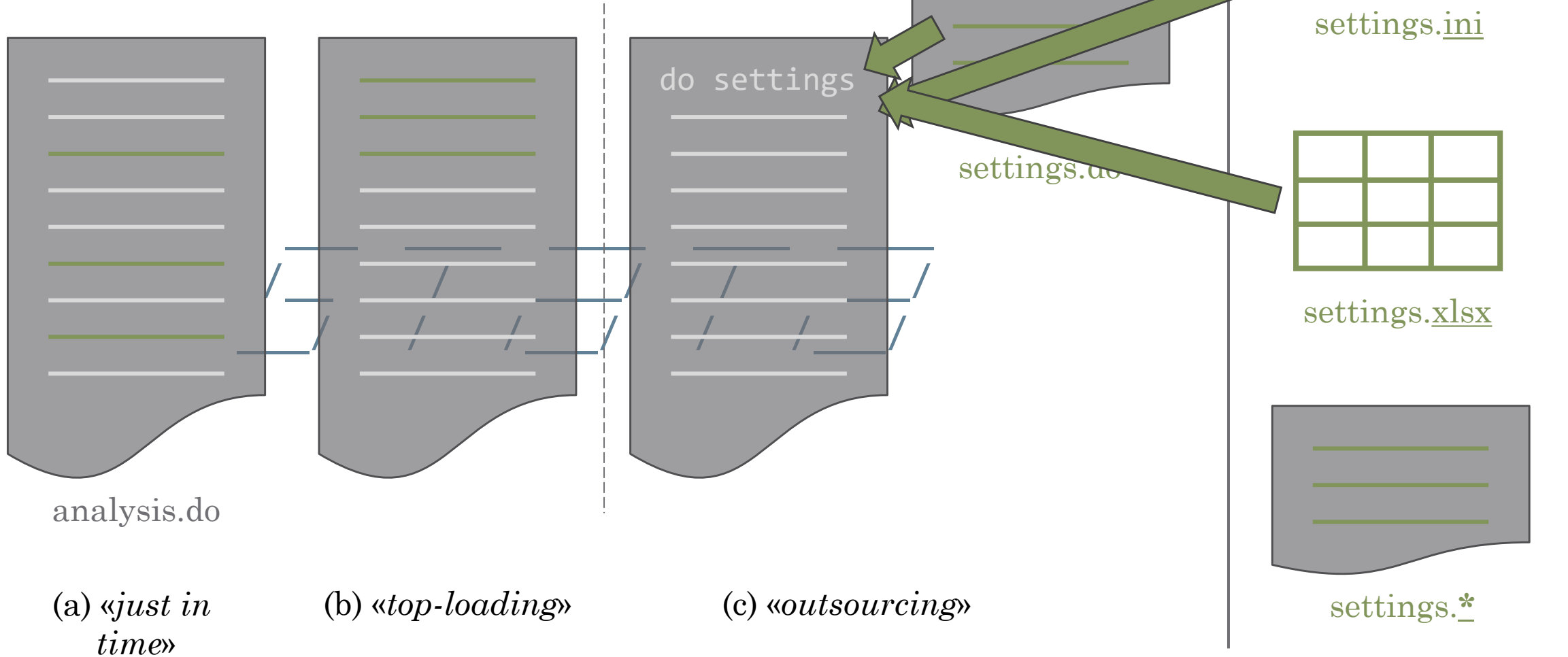


# Four Plus X Approaches – Other Formats

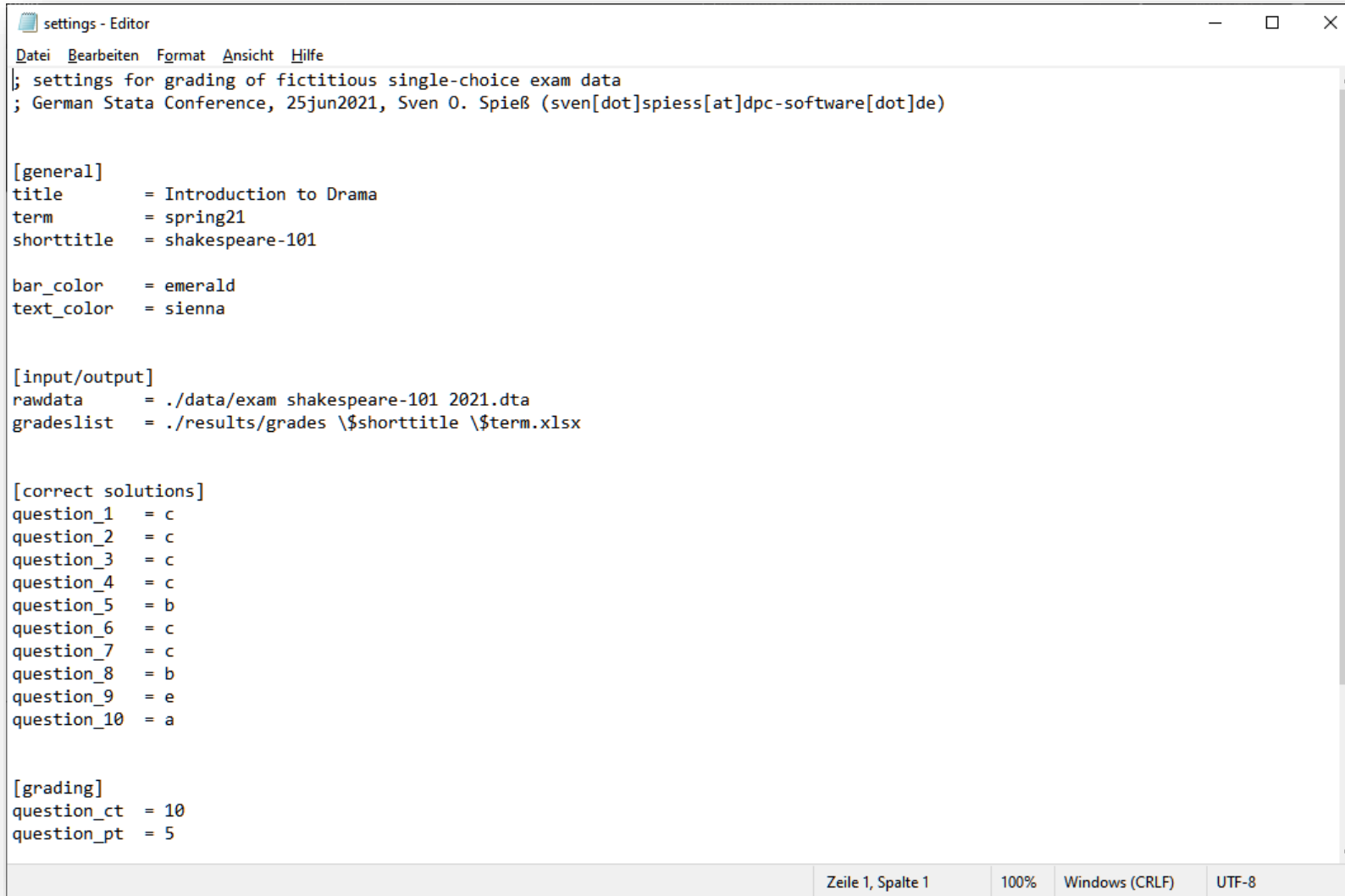
- Macro variables go-to feature define things we want to reuse again (and again)
  - `global question_1 = "c"`
- Compatible to general programming principle of key-value pairs
- In simple cases issue than becomes a matter of organizing our do-files
- However, if our goal/demand is (a) collaborating with people unfamiliar with Stata or (b) interacting with other programs, we might have to consider using other file formats such as Excel sheets or INI files, respectively

# Four Plus X Approaches

```
. global question_1 = "c"
```



# Four Plus X Approaches – INI Files



```
settings - Editor
Datei Bearbeiten Format Ansicht Hilfe
; settings for grading of fictitious single-choice exam data
; German Stata Conference, 25jun2021, Sven O. Spieß (sven[dot]spiess[at]dpc-software[dot]de)

[general]
title      = Introduction to Drama
term       = spring21
shorttitle = shakespeare-101

bar_color  = emerald
text_color = sienna

[input/output]
rawdata     = ./data/exam shakespeare-101 2021.dta
gradeslist  = ./results/grades \${shorttitle} \${term}.xlsx

[correct solutions]
question_1  = c
question_2  = c
question_3  = c
question_4  = c
question_5  = b
question_6  = c
question_7  = c
question_8  = b
question_9  = e
question_10 = a

[grading]
question_ct = 10
question_pt = 5

Zeile 1, Spalte 1    100%    Windows (CRLF)    UTF-8
```

*Example «settings.ini»*

# Four Plus X Approaches – INI Files

- As INI files are plain-text based we can use Stata's tools to read settings *relatively* conveniently and split lines with key-value pairs at position of equal sign:

```
file open ini using "settings.ini", read
```

```
file read ini line
```

```
while r(eof)==0 {  
    local pos = strpos("`line'", "=")  
    local name = substr("`line'", 1, `='pos'-1')  
    local content = substr("`line'", `='pos'+1', .)  
    global `name' = "`content'"
```

```
    file read ini line  
}
```

```
file close ini
```

```
macro list
```

# Four Plus X Approaches – INI Files

- As INI files are plain-text based we can use Stata's tools to read settings *relatively* conveniently and split lines with key-value pairs at position of equal sign:
- However:
  - INI standard only loosely defined (e.g., equal signs (=) vs. colons (:)) as delimiters
  - Need to adhere to Stata's quoting rules – or (much) more advanced parsing
  - Stata naming limitations (e.g., length, characters)
  - No advanced features (e.g., sections, defaults, anchors, etc.)

# Four Plus X Approaches – Excel Sheets

Automatisches Speichern | Suchen (Alt+M) | Spieß, Sven Oliver

Datei Start Einfügen Seitenlayout Formeln Daten Überprüfen Ansicht Hilfe

C21 | x | ✓ | f

Section	Name	Value
General	title	Introduction to Drama
	term	spring21
	shorttitle	shakespeare-101
	bar_color	emerald
	text_color	sienna

**invalid value**

only lowercase letters "a" to "e" allowed

Wiederholen Abbrechen Hilfe

*NB: to refer to other settings prefix respective key-name with "\$"  
only lowercase letters a-e*

question	answer
question_1	f
question_2	c
question_3	c
question_4	c
question_5	b
question_6	c

*Example «settings.xlsx» with data validation error*

# Four Plus X Approaches – Excel Sheets

- With Excel, we first need to take a little detour by importing our settings as an “artificial” data set:

```
frame create settings  
frame change settings
```

```
import excel using settings.xlsx
```

- Subsequently we can again loop over the rows (i.e., observations) to assign macros:

```
forvalues line = 1/`c(N)' {  
    global `=name[`line']' = value[`line']  
}
```

```
macro list
```

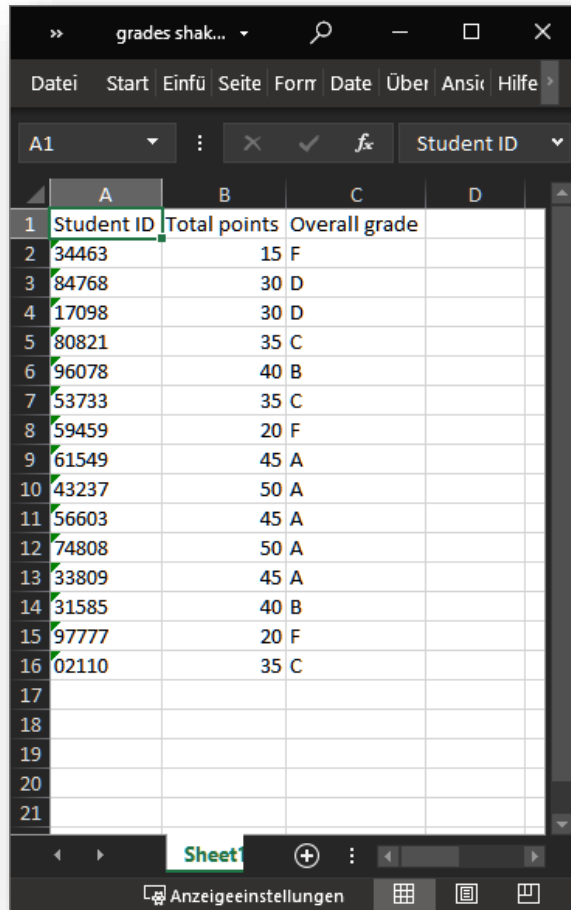
# Four Plus X Approaches – Excel Sheets

- With Excel, we first need to take a little detour by importing our settings as an “artificial” data set
- However, using Excel sheets, we can set many defaults and data validation rules to reduce entry errors/misspecifications
- Same limitations of Stata macro variables apply



# Four Plus X Approaches – Sample Outputs

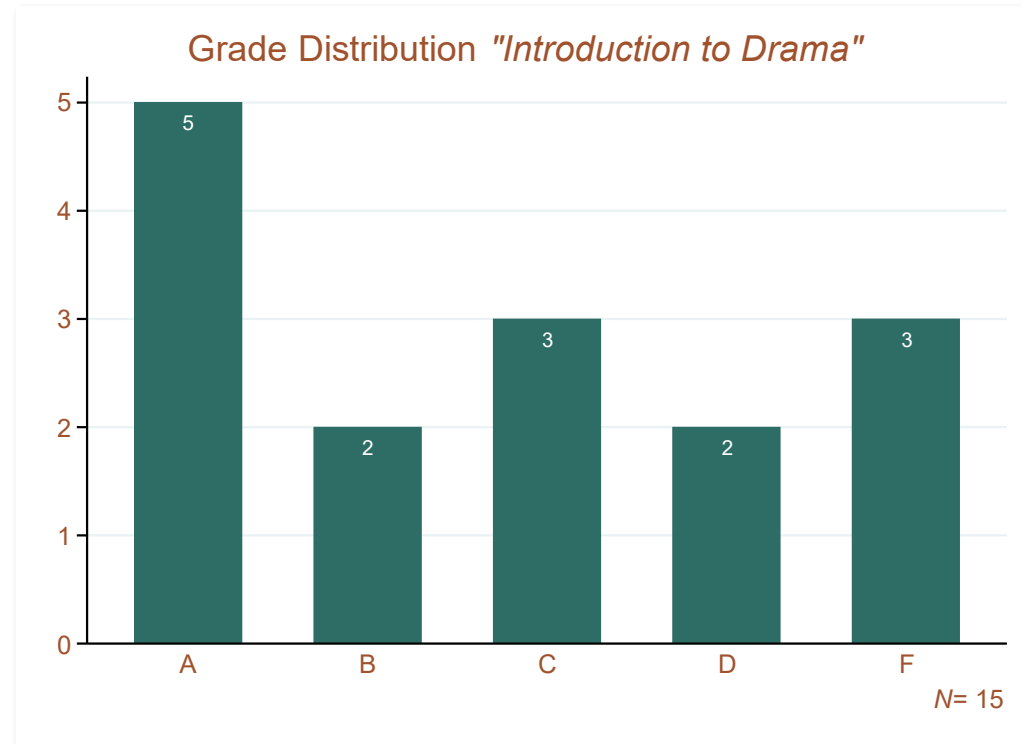
- By only getting the respective settings for each exam, we can from now on produce consistent outputs without “messing” with our analysis syntax



The screenshot shows a spreadsheet window titled 'grades shak...'. The menu bar includes 'Datei', 'Start', 'Einfü', 'Seite', 'Form', 'Date', 'Über', 'Ansik', and 'Hilfe'. The toolbar shows 'A1', a dropdown arrow, a formula icon, and a dropdown menu set to 'Student ID'. The spreadsheet has columns A, B, and C. Column A is labeled 'Student ID', column B is labeled 'Total points', and column C is labeled 'Overall grade'. The data is as follows:

Student ID	Total points	Overall grade
34463	15	F
84768	30	D
17098	30	D
80821	35	C
96078	40	B
53733	35	C
59459	20	F
61549	45	A
43237	50	A
56603	45	A
74808	50	A
33809	45	A
31585	40	B
97777	20	F
02110	35	C

*Example list of grades*



*Example plot*

# Summary – Decision Aid

	Top-loading	Outsourcing	INI files	Excel Sheets
Stata compatibility	✓✓✓	✓✓✓	✓✓	✓
Maintainability	✓✓	✓✓✓	✓✓✓	✓
Ease of setup	✓✓✓	✓✓	✓	✗
Ease of sharing	(✓)	✓✓	✓✓✓	✓✓
Ease of use (collaborators)	✓	✓✓	✓✓	✓✓✓
Validation/control	(✓)	✗	✗	✓✓✓
Compatibility/ interoperability	✗	✓	✓✓✓	✓✓

# Discussion

- With a changing data science ecosystem, maybe we should revise some conventions in using and teaching Stata
  - e.g., more liberal use of global macros, hierarchical do-files, & `macro drop _all`
- Future releases of Stata should drop all user-defined macro variables with `clear all`
- Do we need more (native) support for other data formats, such as INI, and YAML files?
- Limitation of macros regarding more complex structures (e.g., lists of dictionaries, etc.)