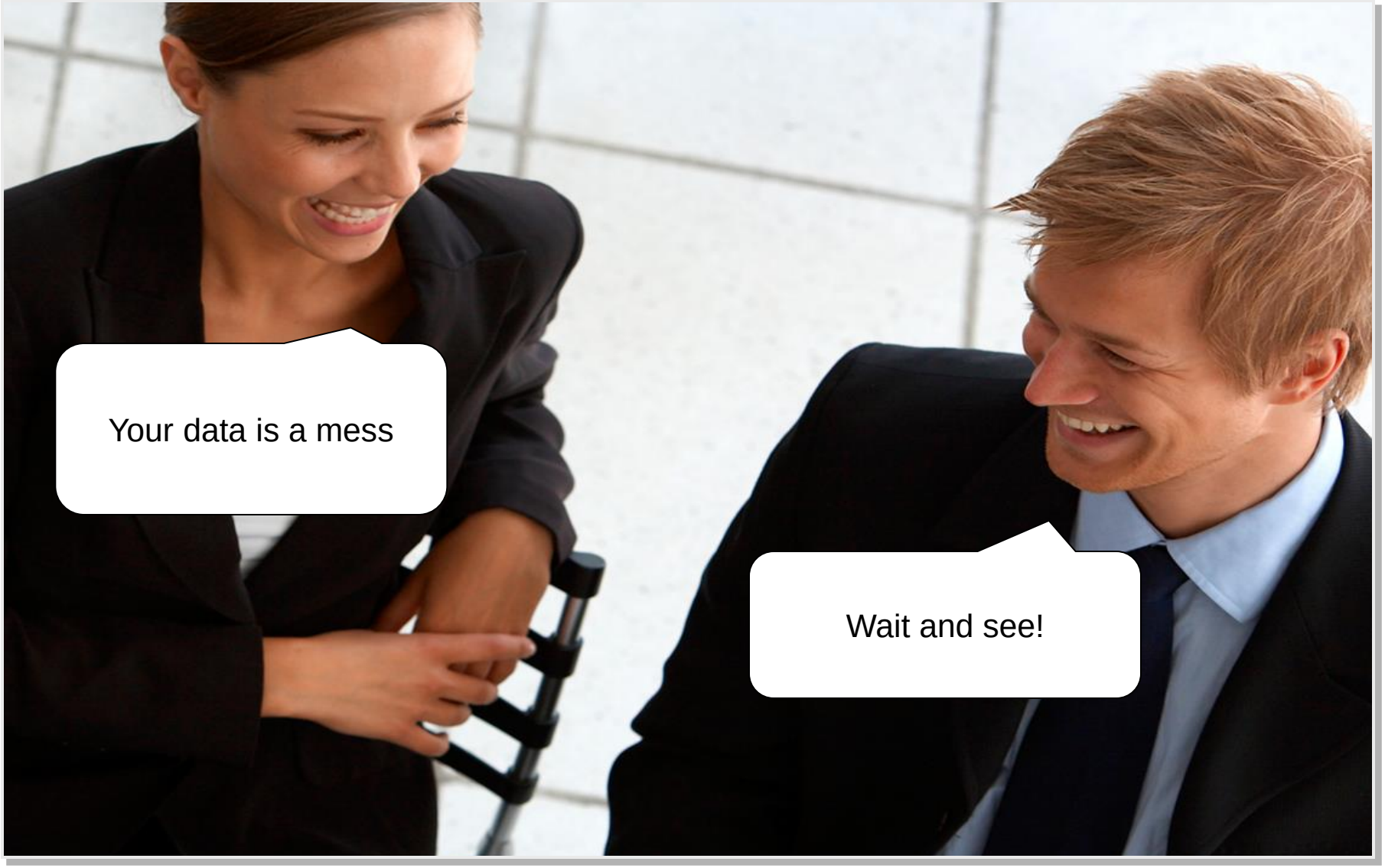
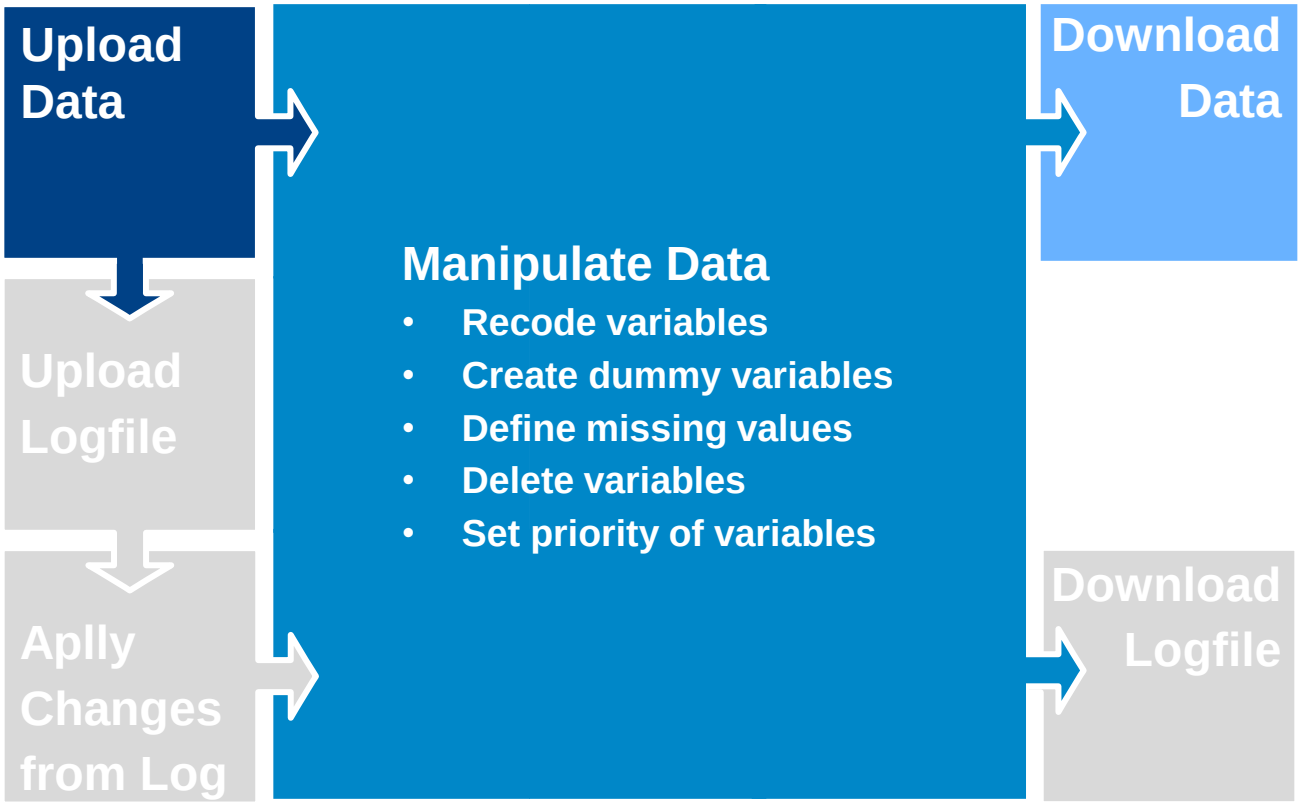
# Data Input Preperation Tool – Manual

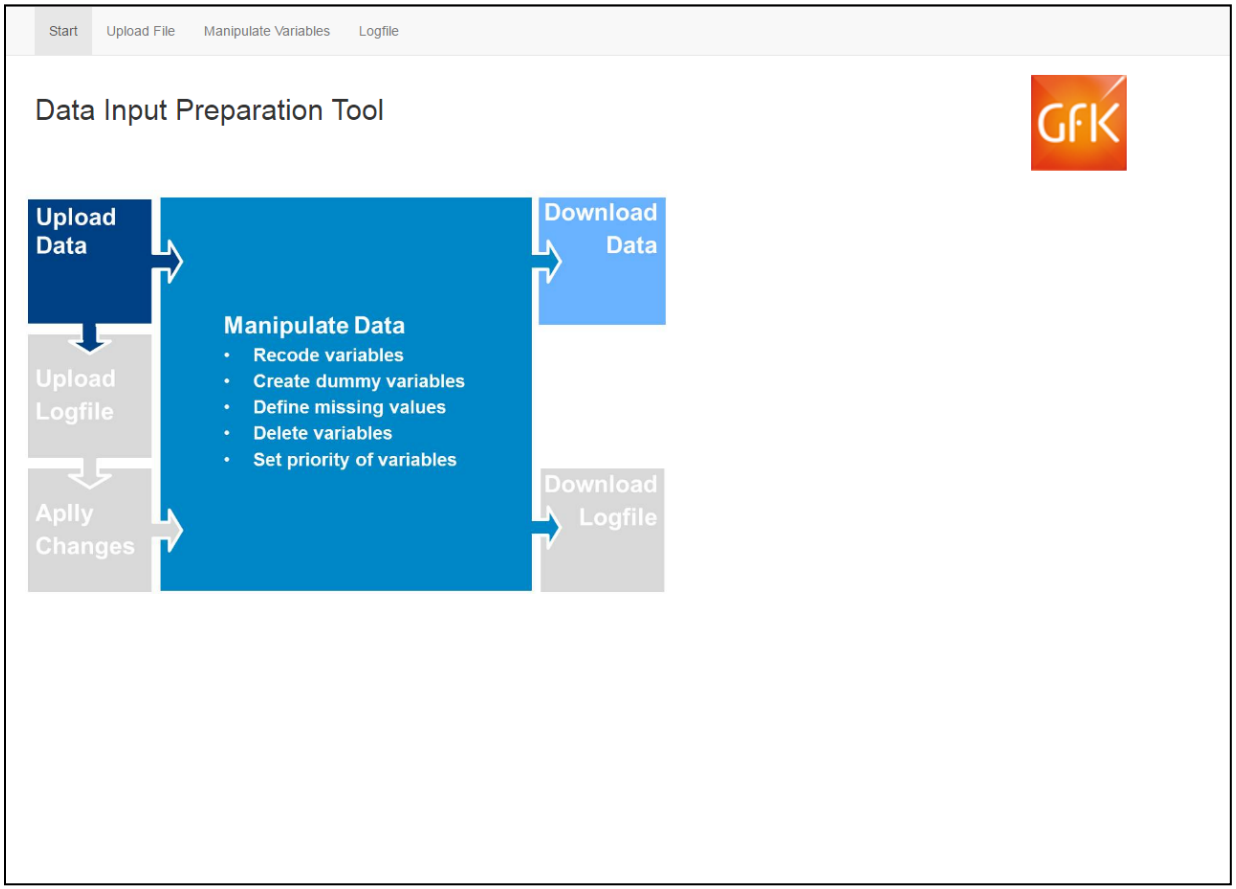


# Outline

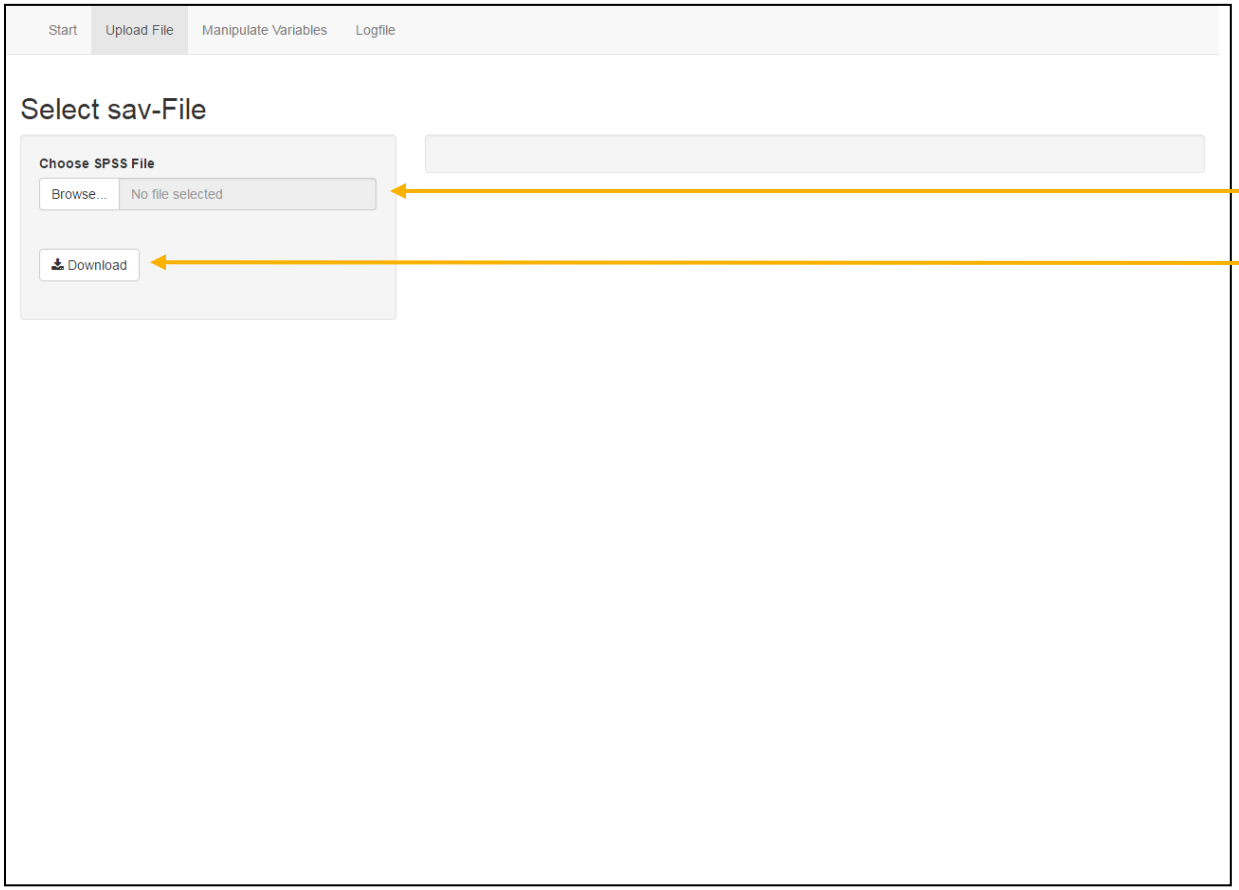
# Workflow


Upload Data
Upload Logfile
Aplly Changes from Log
Manipulate Data
• Recode variables
• Create dummy variables
• Define missing values
• Delete variables
• Set priority of variables
Download Data
Download Logfile

# Data Input


Start
Upload File
Manipulate Variables
Logfile
Data Input Preparation Tool
Upload Data
Upload Logfile
Aplly Changes
Manipulate Data
Recode variables
Create dummy variables
Define missing values
Delete variables
Set priority of variables
Download Data
Download Logfile

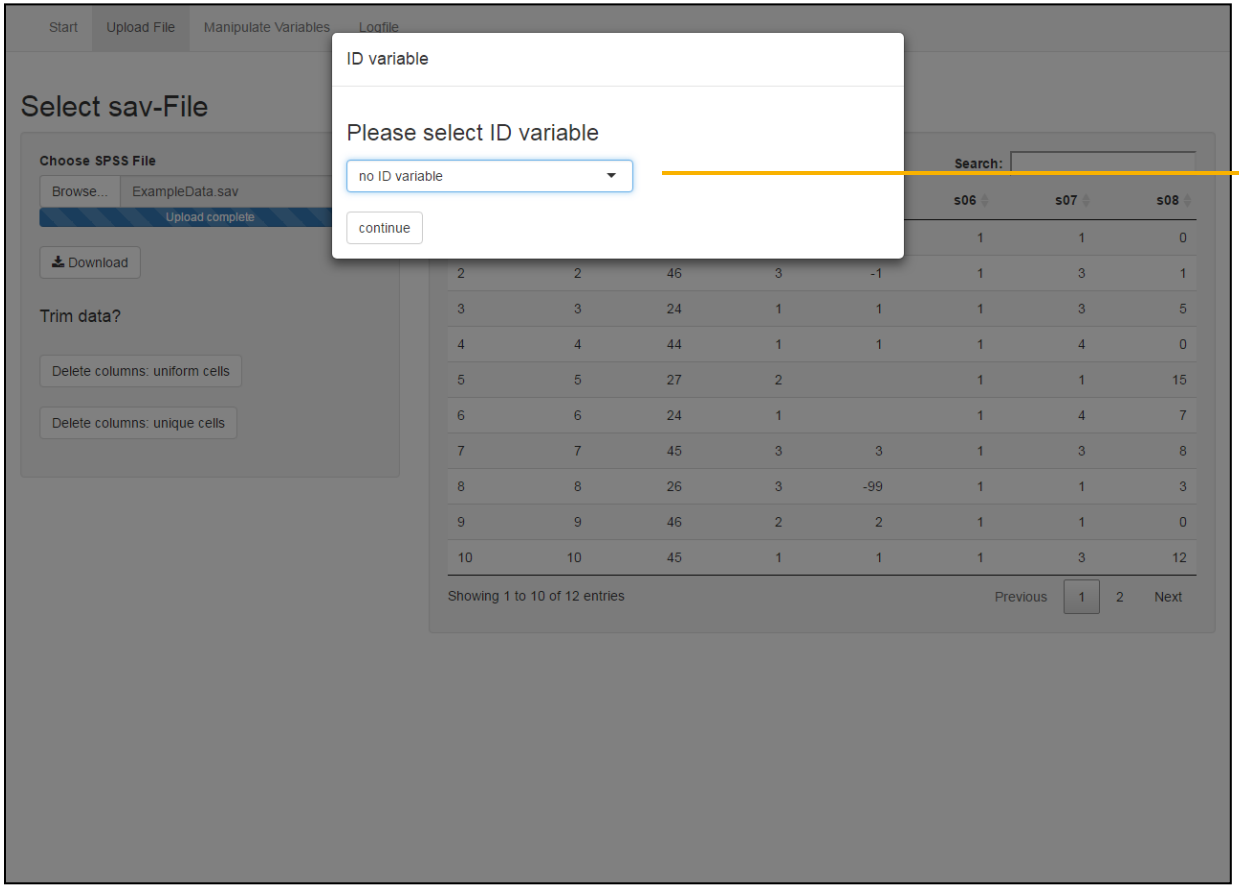
# Data Input

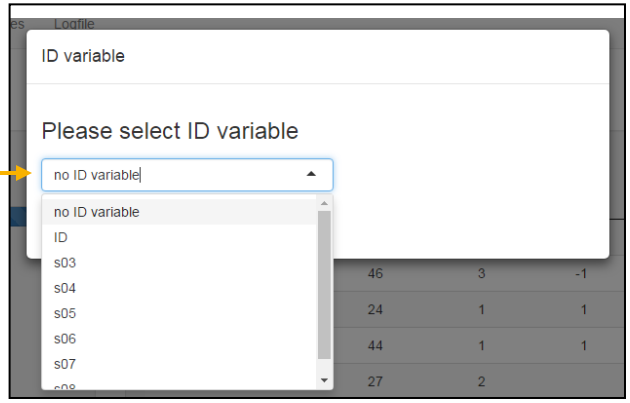



Upload your sav-file here

Later, you can download the revised file here.

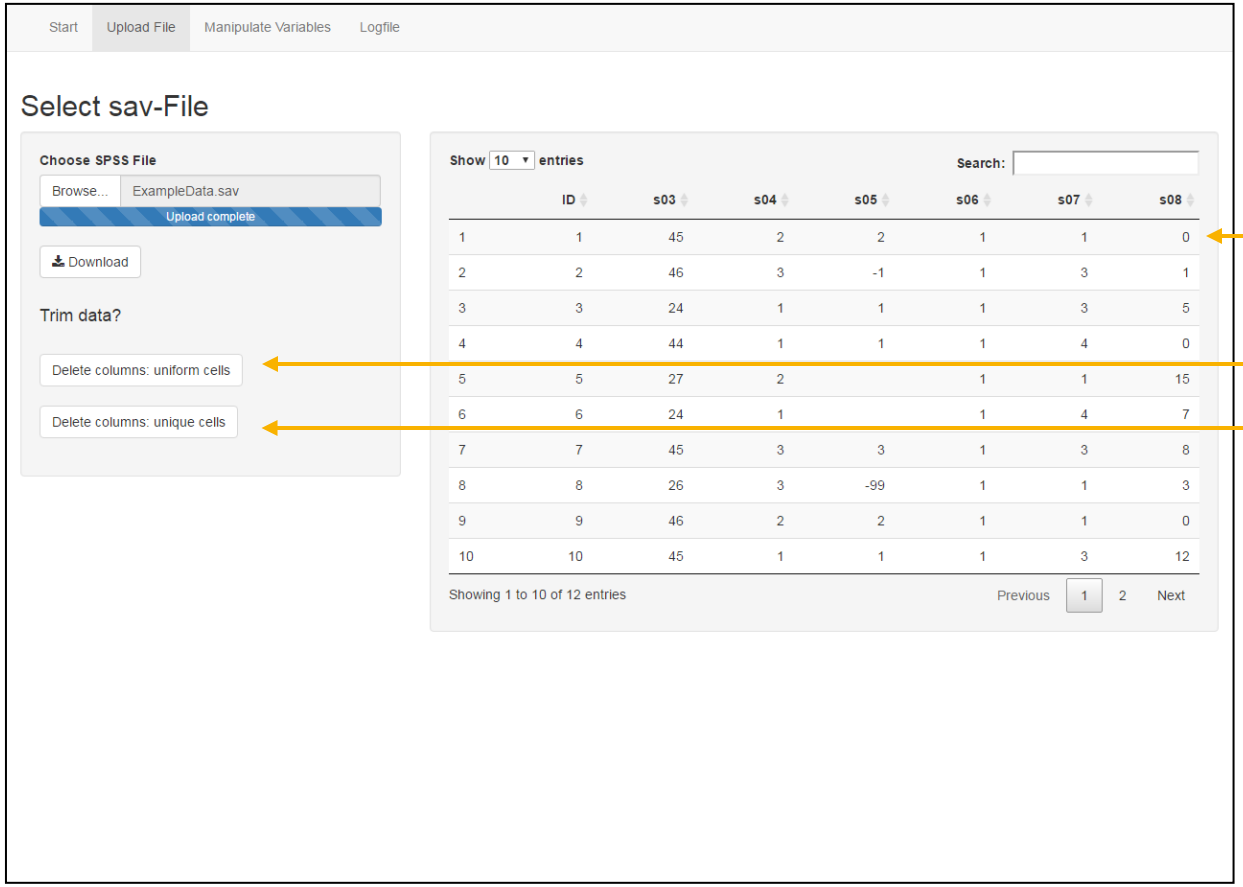
# Data Input




When you have uploaded a sav-file, select the ID variable of your dataset. This variable will kept untouched by all manipulations. If there is no ID variable or you do not want to define one, simply choose ‚no ID variable‘.

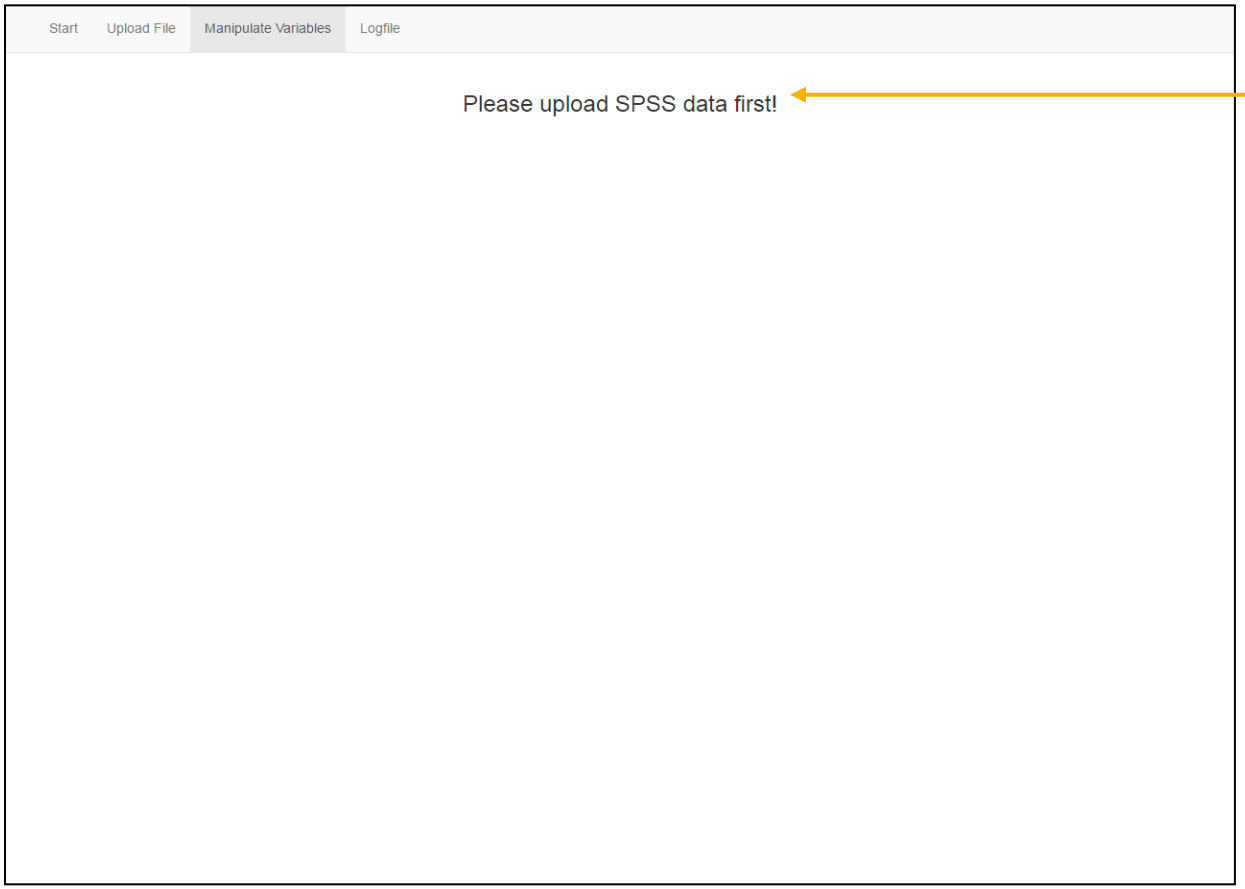
# Data Input




This gives you an overview of your dataset. Note that changes carried out by the app will not appear here until you upload your data again.

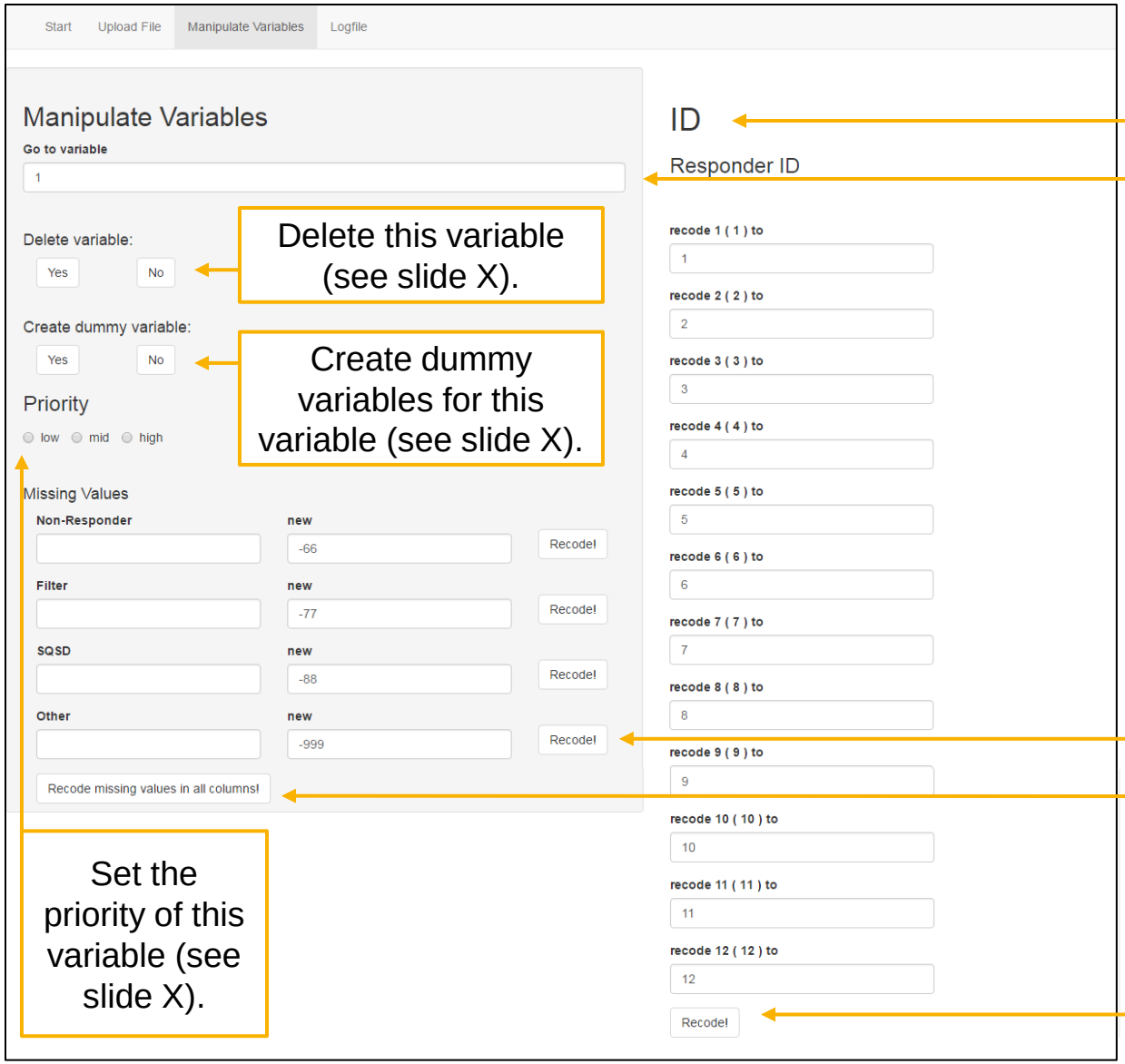
Here, you can delete all variables that have the same entry, like variable ‚s06' ( → **Delete columns: uniform cells**), or you can delete all variables that differ in all cells, like ‚ID' ( → **Delete columns: unique cells**; Of course, it makes no sense to delete the ID variable. Therefore, you can make variables untouchable. See the previous slide for details.)

# Data Input




Oops! Seems like you have to upload your data first.

# Data Input




The header of the current variable.

Navigate through your variables.

Delete this variable (see slide X).

Create dummy variables for this variable (see slide X).

Set the priority of this variable (see slide X).

Recode different missing values in this variable or in all variables (see slide X).

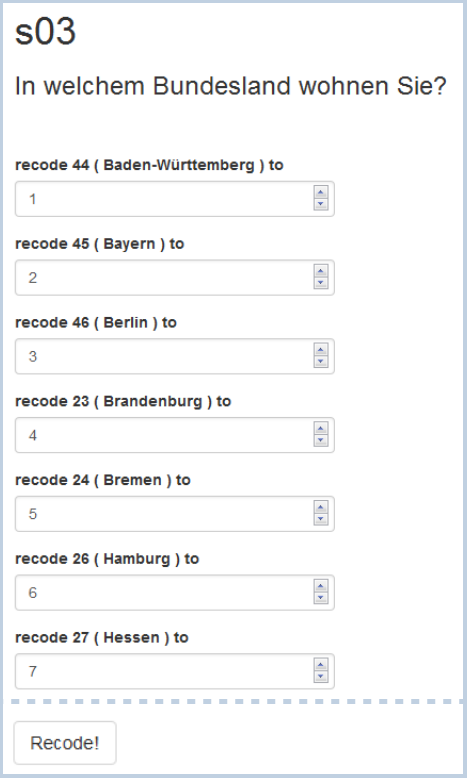
Recode single values in this variable (see slide X).

# Manipulate Variables

## Recode Variables

- Recode values in single variables

### App



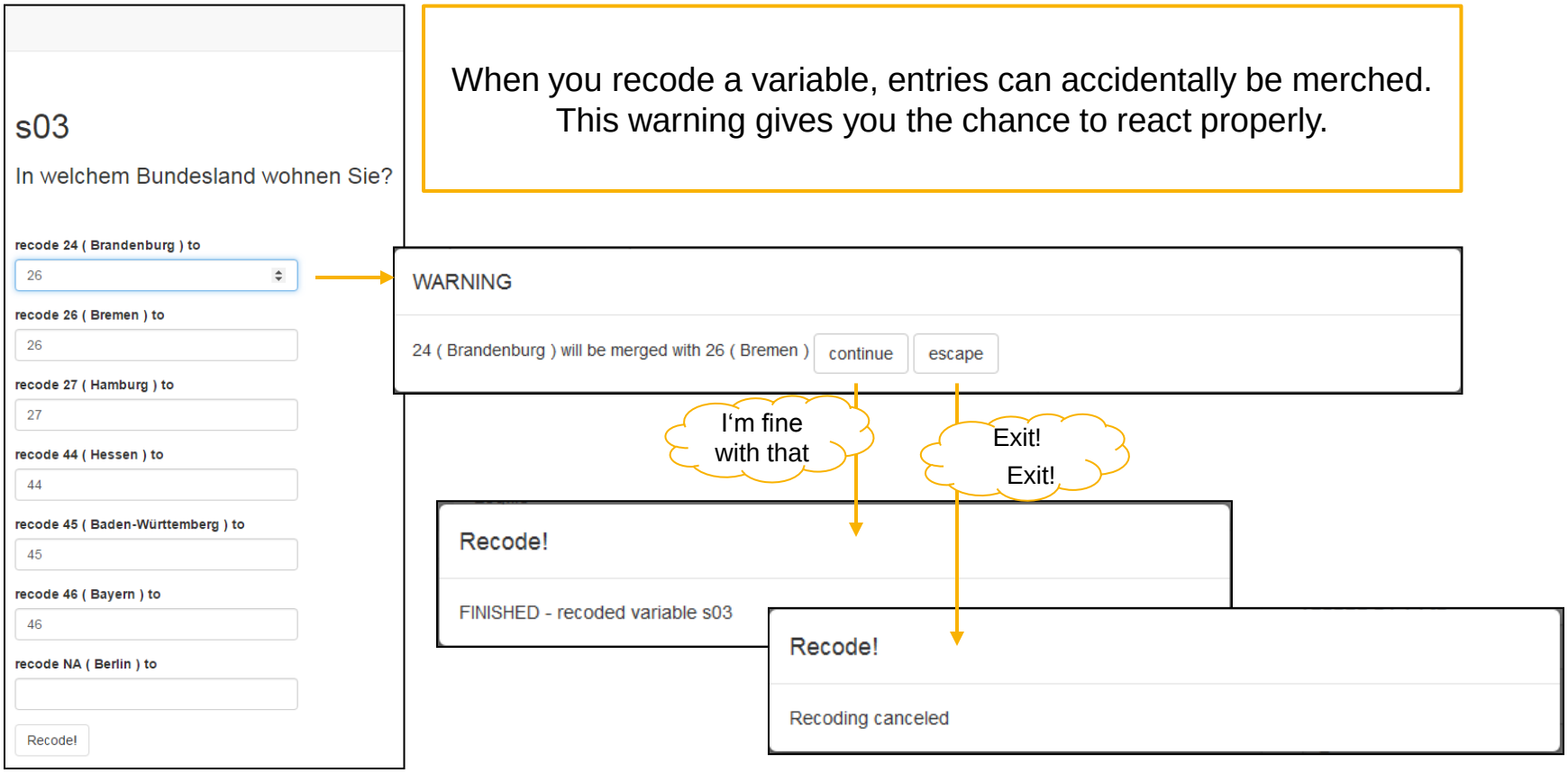

### Data

| s03 | s03 |
|---|---|
| 45 | 2 |
| 46 | 3 |
| 24 | 5 |
| 44 | 1 |
| 27 | 7 |
| 24 | 5 |
| 45 | 2 |
| 26 | 6 |
| 46 | 3 |
| 45 | 2 |
| 27 | 7 |
| ... | ... |

### Logfile

| Variable | Level | oldvalue | newvalue |
|---|---|---|---|
| s03 | Baden-Württemberg | 44 | 1 |
| s03 | Bayern | 45 | 2 |
| s03 | Berlin | 46 | 3 |
| s03 | Brandenburg | 23 | 4 |
| s03 | Bremen | 24 | 5 |
| s03 | Hamburg | 26 | 6 |
| s03 | Hessen | 27 | 7 |
| ... | ... | ... | ... |

# Data Input


s03
In welchem Bundesland wohnen Sie?
recode 24 ( Brandenburg ) to
26
recode 26 ( Bremen ) to
26
recode 27 ( Hamburg ) to
27
recode 44 ( Hessen ) to
44
recode 45 ( Baden-Württemberg ) to
45
recode 46 ( Bayern ) to
46
recode NA ( Berlin ) to
Recode!
When you recode a variable, entries can accidentally be merched.
This warning gives you the chance to react properly.
WARNING
24 ( Brandenburg ) will be merged with 26 ( Bremen )
continue
escape
I'm fine with that
Exit!
Exit!
Recode!
FINISHED - recoded variable s03
Recode!
Recoding canceled

# Manipulate Variables

## Create Dummy Variable

- Create a dummy variables for a specific variable
- Saved in Variable „Create Dummy“ in logfile (1 = yes; 0 = no)

**App**


**Data**

| s04 | s04__1 | s04__2 | s04__3 |
|---|---|---|---|
| 2 | 0 | 1 | 0 |
| 3 | 0 | 0 | 1 |
| 1 | 1 | 0 | 0 |
| 1 | 1 | 0 | 0 |
| 2 | 0 | 1 | 0 |
| 1 | 1 | 0 | 0 |
| 3 | 0 | 0 | 1 |
| 3 | 0 | 0 | 1 |
| 2 | 0 | 1 | 0 |
| 1 | 1 | 0 | 0 |
| 3 | 0 | 0 | 1 |
| ... | ... | | |

**Logfile**

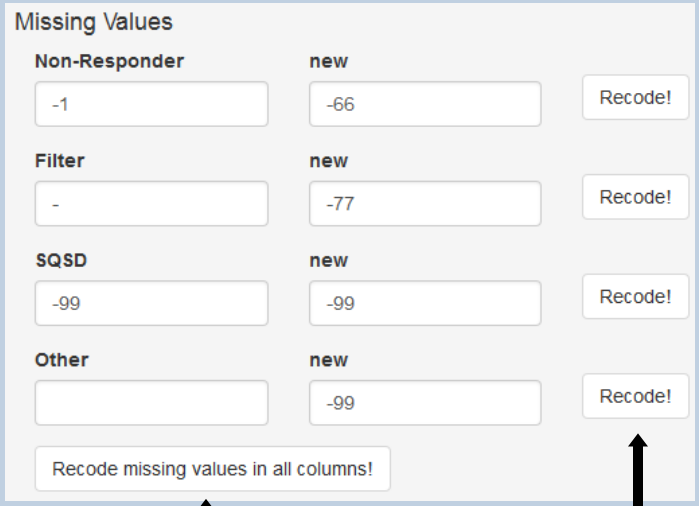
| Variable | ... | Create Dummy |
|---|---|---|
| s03 | ... | 0 |
| s04 | ... | 1 |
| s04 | ... | 1 |
| s04 | ... | 1 |
| s05 | ... | 0 |
| s05 | ... | 0 |
| s05 | ... | 0 |
| ... | ... | ... |

# Manipulate Variables

## Define Missing Values

- Recode values in the complete data set or in single variables
- Recode different missing values with different meanings

### App




Recode missing values in the complete data set.

Recode missing values in current variable.

### Data

| s05 | s05 |
|---|---|
| 2 | 2 |
| -1 | -66 |
| 1 | 1 |
| 1 | 1 |
| - | -77 |
| - | -77 |
| 3 | 3 |
| -99 | -99 |
| 2 | 2 |
| 1 | 1 |
| 3 | 3 |
| ... | ... |

### Logfile

| Variable | ... | MV1_old | MV1_new | MV2_old |
|---|---|---|---|---|
| s05 | ... | -1 | -66 | - |
| s05 | ... | na | na | na |
| s05 | ... | na | na | na |
| s05 | ... | na | na | na |
| s05 | ... | na | na | na |
| s05 | ... | na | na | na |
| s05 | ... | na | na | na |
| ... | ... | ... | ... | ... |

# Manipulate Variables

## Delete Variables

- Delete unnecessary variables
- Saved in Variable „Remain“ in logfile (1 = yes; 0 = no)

**App**


**Data**

| s06 | s07 | s08 |
|---|---|---|
| 1 | 1 | 0 |
| 1 | 3 | 1 |
| 1 | 3 | 5 |
| 1 | 4 | 0 |
| 1 | 1 | 15 |
| 1 | 4 | 7 |
| 1 | 3 | 8 |
| 1 | 1 | 3 |
| 1 | 1 | 0 |
| 1 | 3 | 12 |
| 1 | 4 | 1 |
| ... | | |

| s06 | s08 |
|---|---|
| 1 | 0 |
| 1 | 1 |
| 1 | 5 |
| 1 | 0 |
| 1 | 15 |
| 1 | 7 |
| 1 | 8 |
| 1 | 3 |
| 1 | 0 |
| 1 | 12 |
| 1 | 1 |
| ... | |

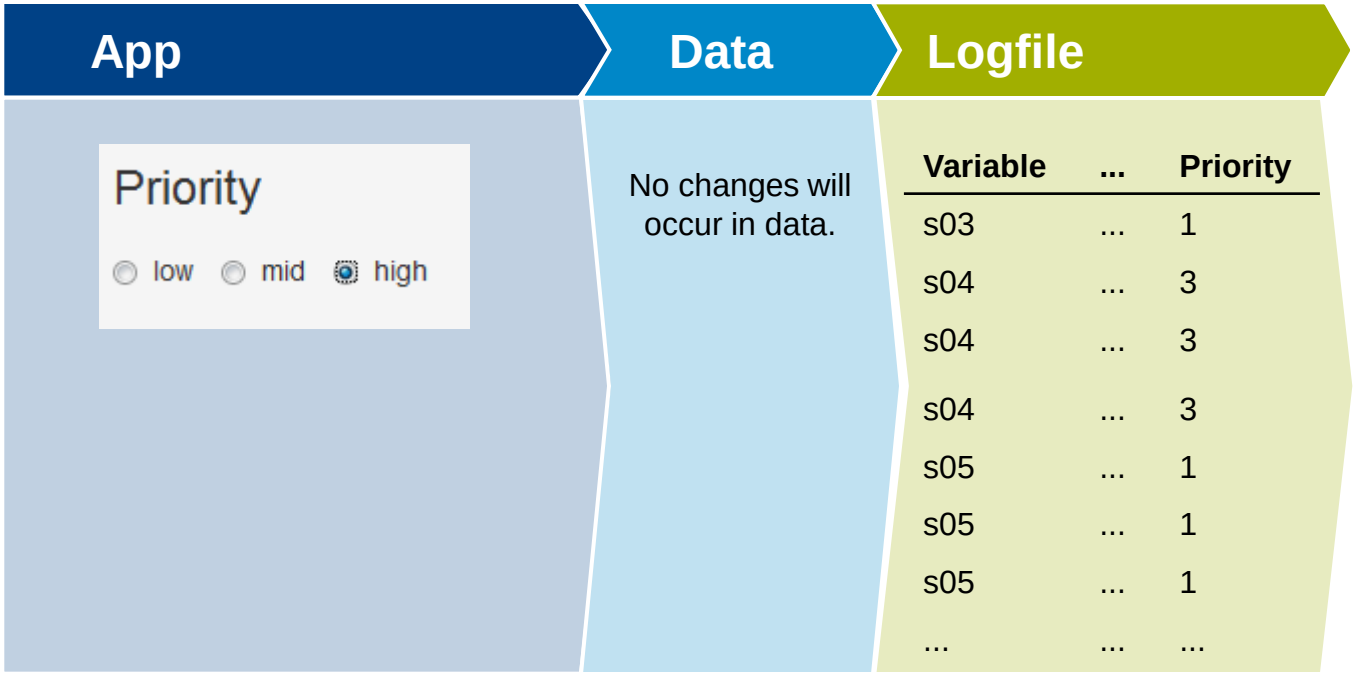
**Logfile**

| Variable | ... | Remain |
|---|---|---|
| s06 | ... | 1 |
| s07 | ... | 0 |
| s07 | ... | 0 |
| s07 | ... | 0 |
| s08 | ... | 1 |
| s08 | ... | 1 |
| s08 | ... | 1 |
| ... | ... | ... |

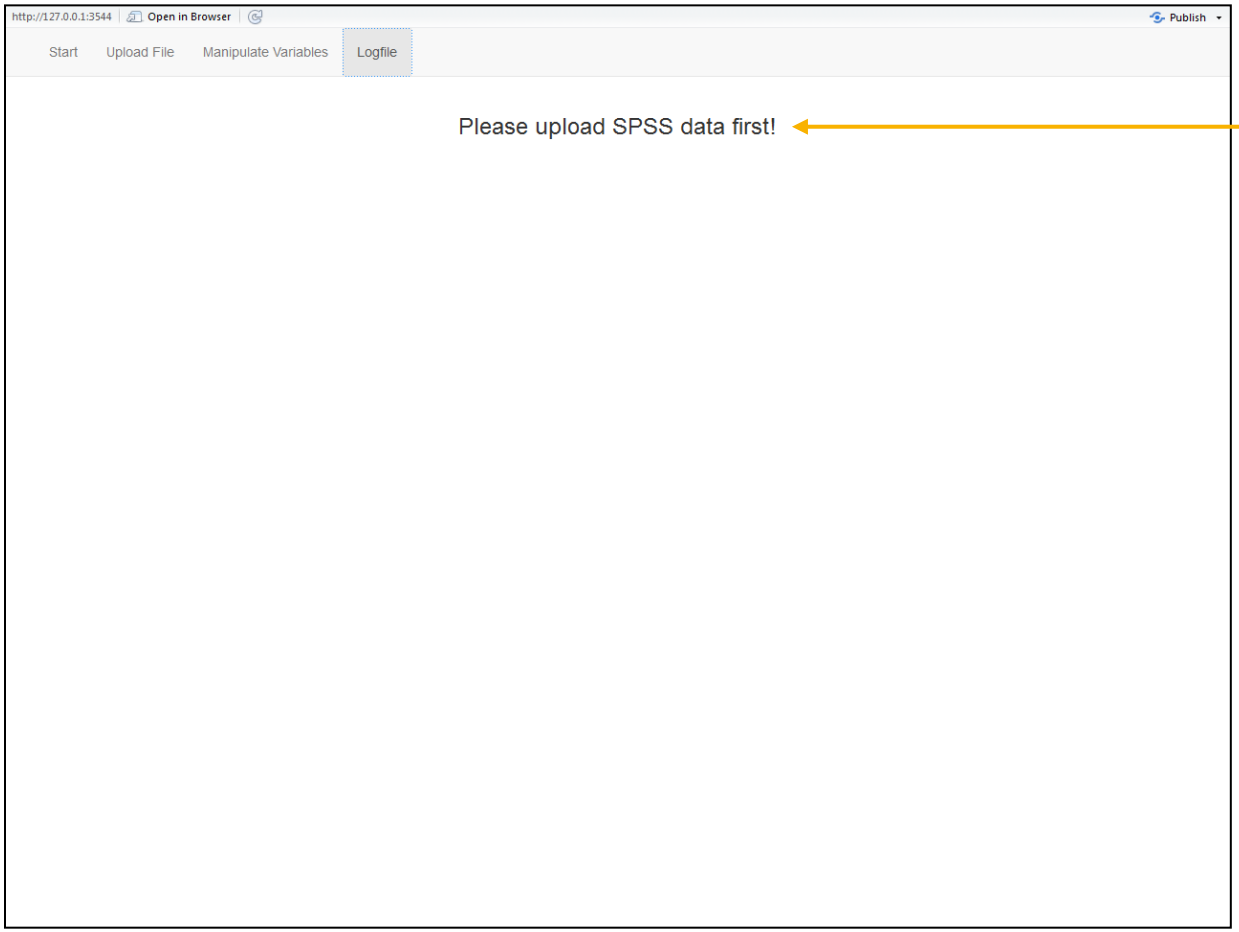
# Manipulate Variables

## Set Priority of Variable

- For further purposes, the priority of single variables can be defined
- Saved in Variable „Priority“ in logfile (1 = low; 2 = mid; 3 = high)
- Default is „low“.

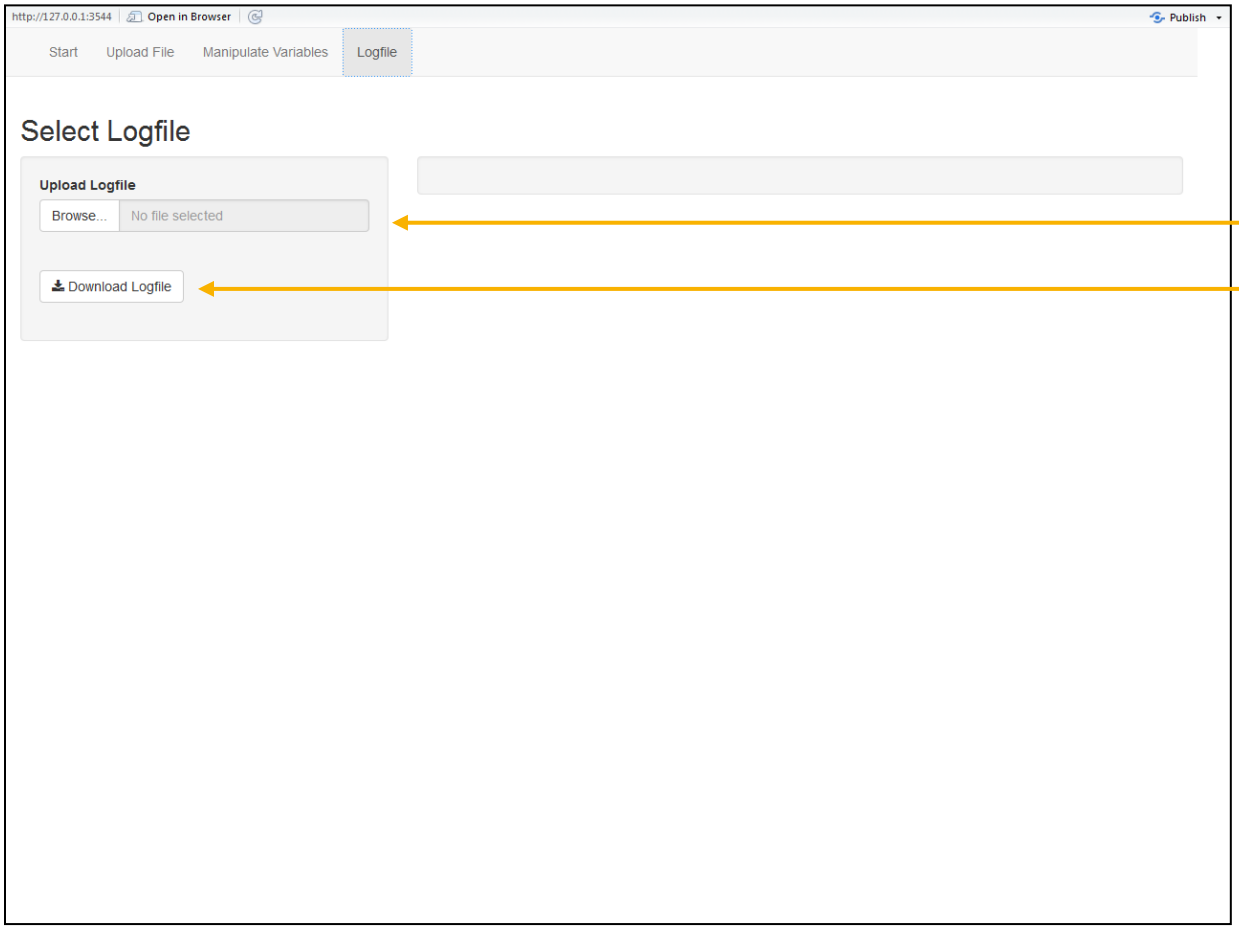


# Data Input




Oops! Seems like you have to upload your data first.

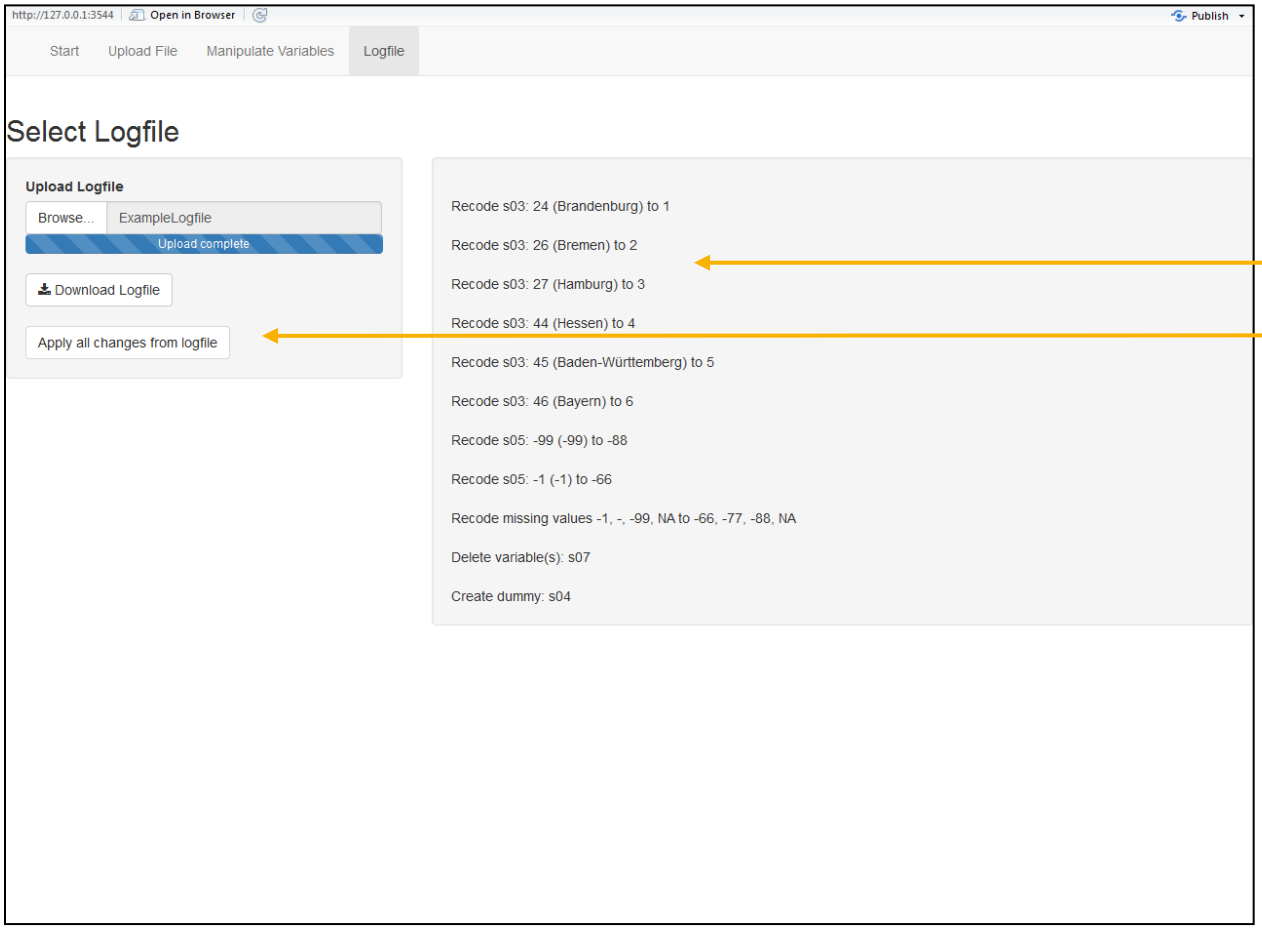
# Data Input




Upload an existing logfile here.

Or download the logfile of your current manipulations here.

# Data Input




After uploading a logfile, this gives you a summary of the changes that will be executed on the current dataset.

You can apply all these changes on your current dataset.

While all recodings will be done immediatly, you can change the selection of variables that will be deleted or dummy coded under ‚manipulate variables‘ later on.

# Logfile

- The logfile is a simple csv-table.
- Each variable takes n rows, where n is the number of different levels.

| variable | level | oldvalue | newvalue | MV1_old | MV1_new | MV2_old | MV2_new | MV3_old | MV3_new | MV4_old | MV4_new | Remain | Create Dummy | Priority |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ID | 1 | 1 | 1 | NA | NA | NA | NA | NA | NA | NA | NA | 1 | 0 | 1 |
| ID | 2 | 2 | 2 | NA | NA | NA | NA | NA | | | | 1 | 0 | 1 |
| ID | 3 | 3 | 3 | NA | NA | NA | NA | NA | | | | 1 | 0 | 1 |
| ID | 4 | 4 | 4 | NA | NA | NA | NA | NA | | | | 1 | 0 | 1 |
| ID | 5 | 5 | 5 | NA | NA | NA | NA | NA | | | | 1 | 0 | 1 |
| ID | 6 | 6 | 6 | NA | NA | NA | NA | NA | | | | 1 | 0 | 1 |
| ID | 7 | 7 | 7 | NA | NA | NA | NA | NA | | | | 1 | 0 | 1 |
| ID | 8 | 8 | 8 | NA | NA | NA | NA | NA | | | | 1 | 0 | 1 |
| ID | 9 | 9 | 9 | NA | NA | NA | NA | NA | | | | 1 | 0 | 1 |
| ID | 10 | 10 | 10 | NA | NA | NA | NA | NA | | | | 1 | 0 | 1 |
| ID | 11 | 11 | 11 | NA | NA | NA | NA | NA | | | | 1 | 0 | 1 |
| ID | 12 | 12 | 12 | NA | NA | NA | NA | NA | | | | 1 | 0 | 1 |
| s03 | Brandenburg | 24 | 1 | NA | NA | NA | NA | NA | | | | 1 | 0 | 1 |
| s03 | Bremen | 26 | 2 | NA | NA | NA | NA | NA | | | | 1 | 0 | 1 |
| s03 | Hamburg | 27 | 3 | NA | NA | NA | NA | NA | | | | 1 | 0 | 1 |
| s03 | Hessen | 44 | 4 | NA | | | | NA | | | | 1 | 0 | 1 |
| s03 | Baden-Württemberg | 45 | 5 | NA | | | | NA | NA | NA | NA | 1 | 0 | 1 |
| s03 | Bayern | 46 | 6 | NA | | | | NA | NA | NA | NA | 1 | 0 | 1 |
| s03 | Berlin | NA | NA | NA | | | | NA | NA | NA | NA | 1 | 0 | 1 |
| s04 | 1 | 1 | 1 | NA | | | | NA | NA | NA | NA | 1 | 1 | 3 |
| s04 | 2 | 2 | 2 | NA | | | | NA | NA | NA | NA | 1 | 1 | 3 |
| s04 | 3 | 3 | 3 | NA | NA | NA | NA | NA | NA | NA | NA | 1 | 1 | 3 |
| s05 | -99 | -99 | -88 | -1 | -66 | - | -77 | -99 | -88 | NA | NA | 1 | 0 | 1 |
| s05 | -1 | -1 | -66 | NA | NA | NA | NA | NA | NA | NA | NA | 1 | 0 | 1 |

It seems like you can delete, dummify or set the priority of single levels of a variable, but you can only define it for the whole variable. However, here, only the first row of a variable will be considered by the app.

Missing values are saved in the first row of a variable. The following rows are meaningless.