README

****
About the NEWID
****

newid = province(original) + village(scrambled) + household(scrambled)

The original geographical codes were scrambled in order to conform to public data dissemination protocol.

The scrambled "NEWID" still represents a unique household in each case.
It is possible to track a given village over time using the public data without the original codes.

The "NEWID" in the Monthly Survey is composed by:

1-2 digits = Province
3-4 digits = Village
5-8 digits = Household

Example...
07610901
changwat: 07
village: 61
household: 0901

By extract the first 4 digits from the "NEWID", a user can obtain the village dummies.
In STATA, we can simply use the following commands to extract the substring:

```
gen villageid=substr(newid, 1, 4)
```

****

About the Villages

****

The Monthly Survey was initiated in August l998 in a subset of villages from the original sampling frame of the 1997 annual big survey.

In total there are *16* villages, four villages in each of the four original Changwats. Specifically, one Tambon per Changwat was chosen from the *12* Tambons of the initial 1997 annual big survey.

Below are matching scrambled codes for the 16 villages in the Monthly survey that are consistent with the NEWIDs in the public version of Townsend annual surveys.

	Newid	Changwat	Amphoe	Tambon	Village
1.	27302965	27	30	29	65
2.	27302953	27	30	29	53
3.	27302951	27	30	29	51
4.	27302956	27	30	29	56
5.	53233259	53	23	32	59
6.	53233260	53	23	32	60
7.	53233254	53	23	32	54
8.	53233253	53	23	32	53
9.	07284065	07	28	40	65
10.	07284061	07	28	40	61
11.	07284063	07	28	40	63
12.	07284062	07	28	40	62
13.	49294159	49	29	41	59
14.	49294157	49	29	41	57
15.	49294161	49	29	41	61
16.	49294160	49	29	41	60

****

About Missing Data Codes

****

For Numeric Variable	For String Variable		Missing Value Type
.a	***RF***	=	refuse to answer
.b	***DK***	=	don't know the answer
.c	***NA***	=	not applicable
.x	***IV***	=	invalid value
.y	***MI***	=	missing value

.a or ***RF***: The subject explicitly refused to answer the question or did not answer the question when he or she should have.
.b or *DK*: The subject was unable to answer the question, either because he or she had no opinion or because the required information was not available.
.c or ***NA***: The subject was never asked the question for one reason or another. Usually this results from "skip patterns" that occur.
.x or ***IV***: A response is inconsistent with related responses or is incompatible with response categories. The circumstances could be: interviewers/data-entry worker incorrectly record a response when recording or keying in the data.
.y or ***MI***: Items should be filled out but have no data entry found. This is enumerator's own mistake and nothing to do with the respondent. The circumstances can be interviewers failing to ask a question or forgetting to record a response.

****

About Dataset/Module Naming Conventions

****

Below is a summary of the module abbreviations

n	letter	name
0	N	
1	P	Household Composition
2	K	Assets
3	A	Agricultural Assets
4	R	Land
5	C	Cultivation
7	M	Crop Inventory
8	I	Livestock Inventory
9	V	Livestock Activities
10	F	Fish and Shrimp
11	J	Occupation
12	T	Business
13	G	Gifts and Remittance
14	S	Savings
15	B	Borrowing
16	L	Lending
17	H	Health
21	D	Insurance
22	E	ROSCA
23	X	Expenditures
24	Y	Income

****

****