Workshop: Introduction to Stata
Instructors: Rebecca Gleit & Amy Johnson

PART 1: DATA ORGANIZATION

"Friends.dta"

	major	year_school	regions	siblings	height	temp	F_C	cheese
1	Spanish	Junior	Northeast,West	1	66	76	F	Brie
2	Math	Sophomore	Midwest,West	1	64	63	F	Parmesean
3	Sociology	Grad student	Northeast,Midwest,West	3	69	60	F	Gouda
4	Sociology	Grad student	Northeast,Midwest,West	1	65	75	F	blue
5	Sociology	Grad student	Northeast,West	4	65	75	F	Sharp cheddar
6	Sociology	Grad student	Northeast,West	2	83	78	F	Cheddar!!
7		Co-term		0	77	0	C	Gouda
8		Sophomore	South	1	88	.	.	
9	Sociology of Education	Grad student	Northeast,West	1	63	80	F	goat
10	Undeclared	Freshman	Midwest	.	38	72	F	Sharp cheddar
11	Sociology!	Grad student	West	1	68	65	F	a nice sharp gouda
12	Sociology	Grad student	Midwest,West	1	70	24	C	feta
13	Sociology of Education	Grad student	Northeast,Midwest,West	3	66	75	F	daiya

Vars: 8 Order: Dataset Obs: 13

Variable name	Variable label	Type of variable	Values & values labels
major	Major		
year_school	Year in school		
regions	Regions of US lived in		
siblings	Number of siblings		
height	Height (in)		
temp	Temperature (number)		
F_C	Temperature (unit)		
cheese	Favorite cheese		

PART 2: DATA MANIPULATION

Stata logic syntax:

==	"is equal to"	!=	"is not equal to"
>	"greater than"	<	"less than"
>=	"greater than or equal to"	<=	"less than or equal to"

Practice subsetting observations

1. **Translate this logical statement from Stata syntax into words.**

```
if temp>=70 & F_C==1
```

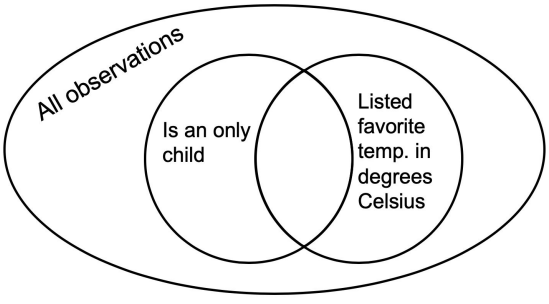
For #2-3: **Translate these logical statements from words into Stata syntax.**

2. if the student is less than 70 inches tall and has 2 siblings.

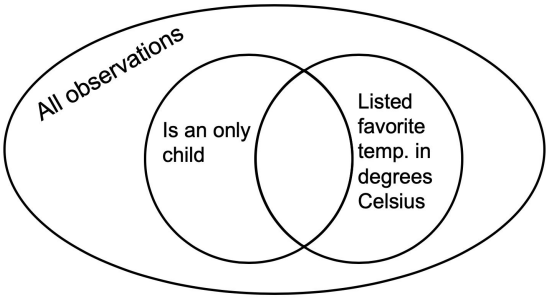
3. if the student is majoring in math or they are a freshman.

For #4-5: **Shade the area corresponding to the logical statement in Stata syntax.**

4. `if F_C==2 & siblings==0`

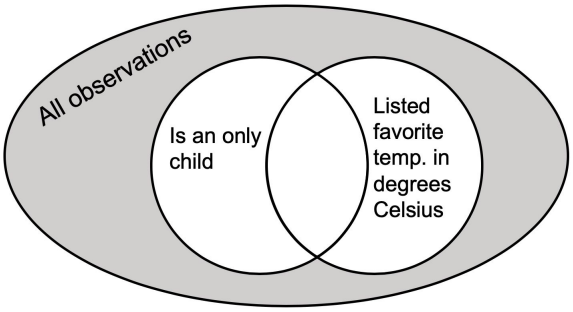


5. `if siblings>0 | F_C!=1`

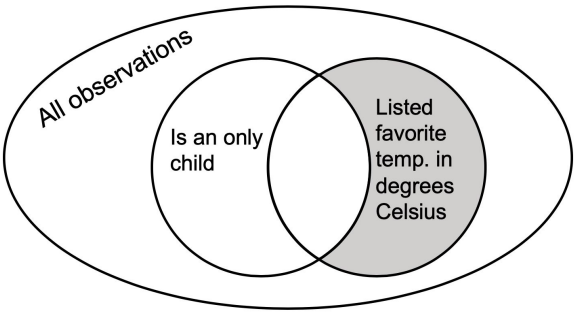


For #6-7: **Write the logical statement (in Stata syntax) that corresponds to the shaded area.**

6.



7.



For #8-11: **Look at the screenshot of "Friends.dta" on page 1 of this handout. Identify the row numbers of the observations that would be included in each subset.**

8. `if height >= 65 & height < 77`

9. `if cheese=="Gouda" | cheese== "Brie"`

10. `if temp <= 75 & siblings!=1`

11. `if major!="Sociology" | F_C==2`

CHALLENGE: Write a logical statement in Stata syntax that will capture the following subset of observations from "Friends.dta" on p. 1.

12. {10}

13. {1, 2, 7, 8, 9, 10, 11, 13}
(hint: look at the variable *major*)

Generate and replace

For #14-16: **Fill in what the new variable would look like given the Stata code. Put each new variable in a column in the table below.**

14.
```
generate softfav = 0
replace softfav = 1 if cheese==“Brie” | cheese==“goat” | cheese==“feta”
replace softfav = . if cheese==“”
```

15.
```
generate tallwsibs = 1 if height>70 & siblings!=0
replace tallwsibs = 0 if tallwsibs!=1
```

16.
```
generate likes_hot = .
replace likes_hot = 0 if temp!=.
replace likes_hot = 1 if (temp>68 & F_C==1) | (temp>20 & F_C==2)
```

For #17-18: **Write code that would create these variables included in the table below:**

17. *bigfamily*

18. enthus_tall_grad

year_school	siblings	height	temp	F_C	cheese				bigfamily	enthus_tall_grad
Junior	1	66	76	F	Brie				0	0
Sophomore	1	64	63	F	Parmesean				0	0
Grad student	3	69	60	F	Gouda				1	0
Grad student	1	65	75	F	blue				0	0
Grad student	4	65	75	F	Sharp cheddar				1	0
Grad student	2	83	78	F	Cheddar!!				0	1
Co-term	0	77	0	C	Gouda				0	0
Sophomore	1	88	.	.					0	0
Grad student	1	63	80	F	goat				0	0
Freshman	.	38	72	F	Sharp cheddar				.	0
Grad student	1	68	65	F	a nice sharp gouda				0	0
Grad student	1	70	24	C	feta				0	0
Grad student	3	66	75	F	daiya				1	0

CHALLENGE: Write a different set of code that will also create the variables in #17-18.