

Food Price Data Analysis Using STATA

Fundamentals of STATA

May 25, 2022

Goals

- Understand the Stata program and interface
- Import survey data into Stata
- Able to use basic Stata commands for summarizing data and adding variables.

Content

- Introduction to Stata
- Import survey data from Kobo ToolBox
- Data cleaning and analysis
- Data visualization

The status of food price data collection

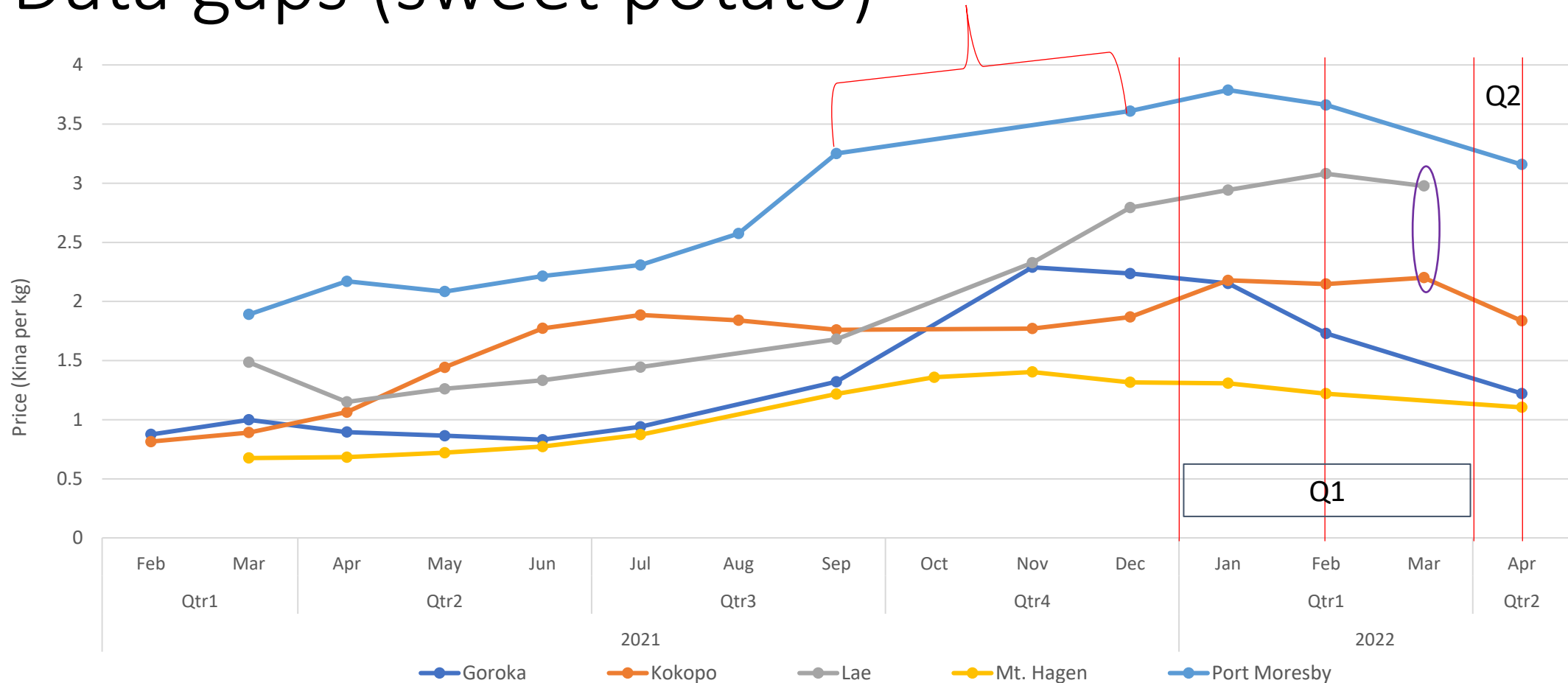
- It is designed to regularly monitor food price trends in major markets.
- Despite the fact that thorough price monitoring necessitates more frequent price information per month, it was designed to collect fortnightly owing to frequent logistical issues.
- However, the recent logistical difficulty has further disrupted fortnightly data, thereby making it monthly data.
- It lacks to represent realistic price trends, which are required for policymakers to devise informative measures.

Quarter	Start	End	Goroka	Banz	Kokopo	Lae	Mt_Hagen	Kundiawa	Port_Moresby
Quarter II	26-Mar-22	8-Apr-22							
	9-Apr-22	22-Apr-22	1	1	1	1	1	1	1
	23-Apr-22	6-May-22		1		1	1	1	1
	7-May-22	20-May-22							
	21-May-22	3-Jun-22							
	4-Jun-22	17-Jun-22							
	18-Jun-22	1-Jul-22							
Total number of rounds in quarter II			1	2	1	2	2	2	2

Fortnight data collection gaps

Quarter	Start	End	Goroka	Banz	Kokopo	Lae	Mt_Hagen	Kundiawa	Port_Moresby
Quarter I	1-Jan-22	14-Jan-22	1			1	1	1	1
	15-Jan-22	28-Jan-22	1	2	1	1	1	1	1
	29-Jan-22	11-Feb-22							
	12-Feb-22	25-Feb-22	1	1		2	1	1	1
	26-Feb-22	11-Mar-22	1	1		1	1	1	1
	12-Mar-22	25-Mar-22							
	26-Mar-22	8-Apr-22							
Total number of rounds in quarter I			4	4	5	4	4	4	4
Quarter	Start	End	Goroka	Banz	Kokopo	Lae	Mt_Hagen	Kundiawa	Port_Moresby
Quarter II	26-Mar-22	8-Apr-22							
	9-Apr-22	22-Apr-22	1	1	1	1	1	1	1
	23-Apr-22	6-May-22		1		1	1	1	1
	7-May-22	20-May-22							
	21-May-22	3-Jun-22							
	4-Jun-22	17-Jun-22							
	18-Jun-22	1-Jul-22							
Total number of rounds in quarter II			1	2	1	2	2	2	2

Data gaps (sweet potato)



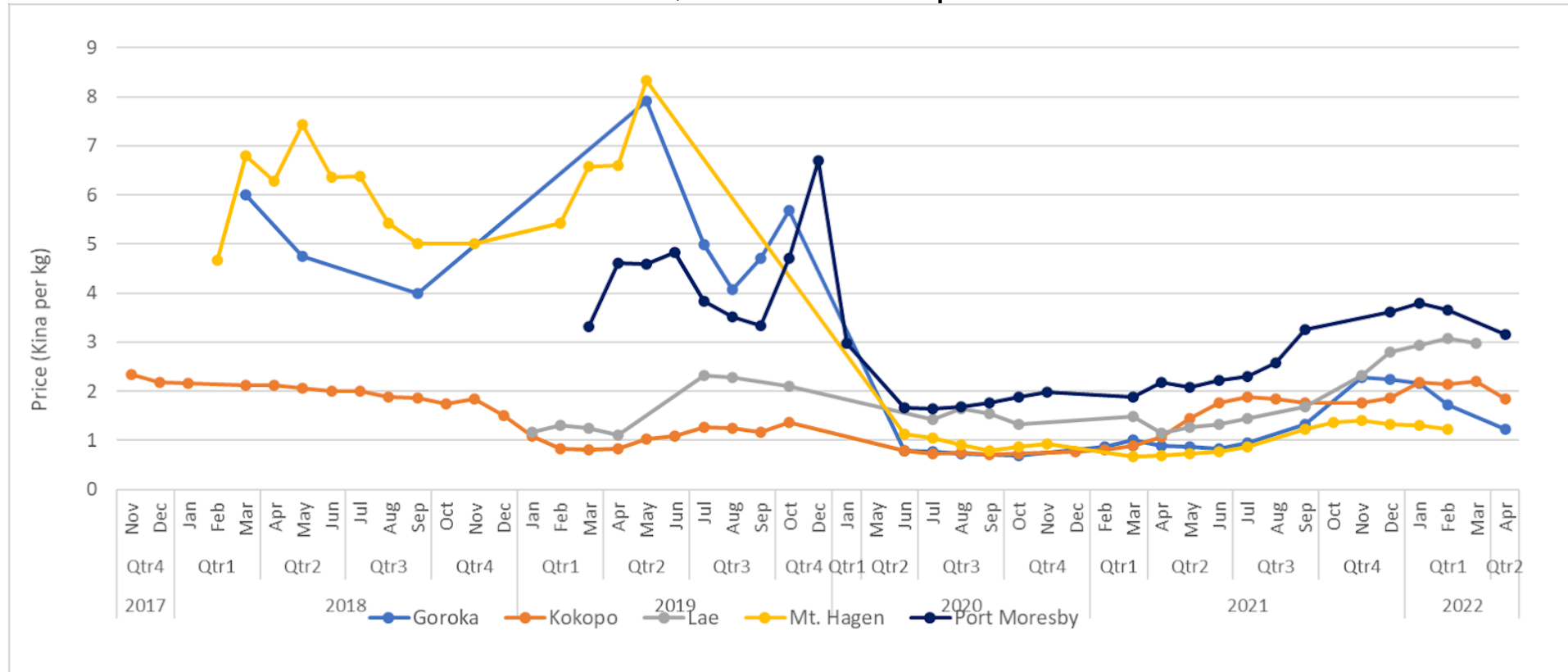
- A significant price data gap reduces the fortnightly quality to monthly data.
- Any conceivable option to maintain the relevance and quality of such an established data collection scheme?

Data quality issues

- It is essential that the data be high quality for research, planning and policy making.
- Even if the quality of price data collection is improving, data entry concerns must still be addressed.
- The key concern noted in the price survey database, among others,
 - Entering unit price instead of total price
 - Incorrectly entering the number of pieces as the weight and vice versa

Example

Observation: The price of sweet potato appears to be more expensive in the producing regional markets of Mt Hagen and Goroka than in the consuming markets of POM, Lae, and Kokopo in 2017-2019, would this be possible?



Example 2 (potato price in Mt Hagen market)

	crops	croptype	unitprice	pieces	weight	availability	quality	year	month	day	market
5482	Potato	Root Crop	10	14	2.2	High	Good	2019	6	19	Mt. Hagen
5483	Potato	Root Crop	10	14	1.88	High	Good	2019	6	19	Mt. Hagen
5484	Potato	Root Crop	5	13	1.22	High	Good	2019	6	19	Mt. Hagen
5494	Potato	Root Crop	4	6	1.64	Low	Excellent	2018	2	20	Mt. Hagen
5495	Potato	Root Crop	4	6	1.4	Low	Excellent	2018	2	20	Mt. Hagen
5496	Potato	Root Crop	3	6	1.5	Low	Excellent	2018	2	20	Mt. Hagen
5497	Potato	Root Crop	4	6	1.5	Low	Excellent	2018	2	20	Mt. Hagen
5498	Potato	Root Crop	4	6	1.52	Low	Excellent	2018	2	20	Mt. Hagen
5499	Potato	Root Crop	4	6	1.5	Low	Excellent	2018	2	20	Mt. Hagen
5500	Potato	Root Crop	3	9	1.26	High	Good	2018	2	20	Mt. Hagen
5501	Potato	Root Crop	3	11	1.3	High	Good	2018	2	20	Mt. Hagen
5502	Potato	Root Crop	3	9	1.2	High	Good	2018	2	20	Mt. Hagen
5503	Potato	Root Crop	2	16	1.4	Medium	Good	2018	2	23	Mt. Hagen
5504	Potato	Root Crop	2	15	1.3	Medium	Good	2018	2	23	Mt. Hagen
5505	Potato	Root Crop	2	16	1.4	Medium	Good	2018	2	23	Mt. Hagen
5506	Potato	Root Crop	5	1	1.8	Medium	Good	2018	2	28	Mt. Hagen
5507	Potato	Root Crop	5	1	1.71	Medium	Good	2018	2	28	Mt. Hagen
5508	Potato	Root Crop	5	1	1.71	Medium	Good	2018	2	28	Mt. Hagen
5518	Potato	Root Crop	10	1	2.3	Medium	Good	2018	3	29	Mt. Hagen
5519	Potato	Root Crop	10	1	3.1	Medium	Good	2018	3	29	Mt. Hagen
5520	Potato	Root Crop	10	1	2.6	Medium	Good	2018	3	29	Mt. Hagen
5521	Potato	Root Crop	5	1	1.7	Medium	Good	2018	4	13	Mt. Hagen
5522	Potato	Root Crop	5	1	1.7	Medium	Good	2018	4	13	Mt. Hagen
5523	Potato	Root Crop	5	1	1.71	Medium	Good	2018	4	13	Mt. Hagen
5524	Potato	Root Crop	3	9	1.5	High	Excellent	2018	4	23	Mt. Hagen
5525	Potato	Root Crop	3	9	1.5	High	Excellent	2018	4	23	Mt. Hagen
5526	Potato	Root Crop	2	9	.9	High	Excellent	2018	4	23	Mt. Hagen
5527	Potato	Root Crop	3	9	1.2	Medium	Good	2018	4	23	Mt. Hagen
5528	Potato	Root Crop	3	9	1.26	Medium	Good	2018	4	23	Mt. Hagen
5529	Potato	Root Crop	3	9	1.32	Medium	Good	2018	4	23	Mt. Hagen
5533	Potato	Root Crop	5	1	1.8	Medium	Good	2018	5	14	Mt. Hagen
5534	Potato	Root Crop	5	1	1.72	Medium	Good	2018	5	14	Mt. Hagen
5535	Potato	Root Crop	5	1	1.78	Medium	Good	2018	5	14	Mt. Hagen
5536	Potato	Root Crop	5	1	1.8	Medium	Good	2018	5	24	Mt. Hagen
5537	Potato	Root Crop	5	1	1.6	Medium	Good	2018	5	24	Mt. Hagen
5538	Potato	Root Crop	5	1	2.1	Medium	Good	2018	6	7	Mt. Hagen
5539	Potato	Root Crop	5	1	1.87	Medium	Good	2018	6	7	Mt. Hagen
5540	Potato	Root Crop	5	1	1.93	Medium	Good	2018	6	7	Mt. Hagen

Current survey implementation

▼ » » Sample 1: price for Taro True

S1. Price of one (bunch / heap / piece / etc.)

S1. No of Pieces per (bunch / heap / piece / etc.)

S1. Weight (Kg) of one (bunch / heap / piece / etc.)

S1. Supply Source (Province)



- Western
- Gulf
- Central
- National Capital District
- Milne Bay
- Northern (Oro)
- Southern Highlands
- Enga
- Western Highlands
- Chimbu (Simbu)

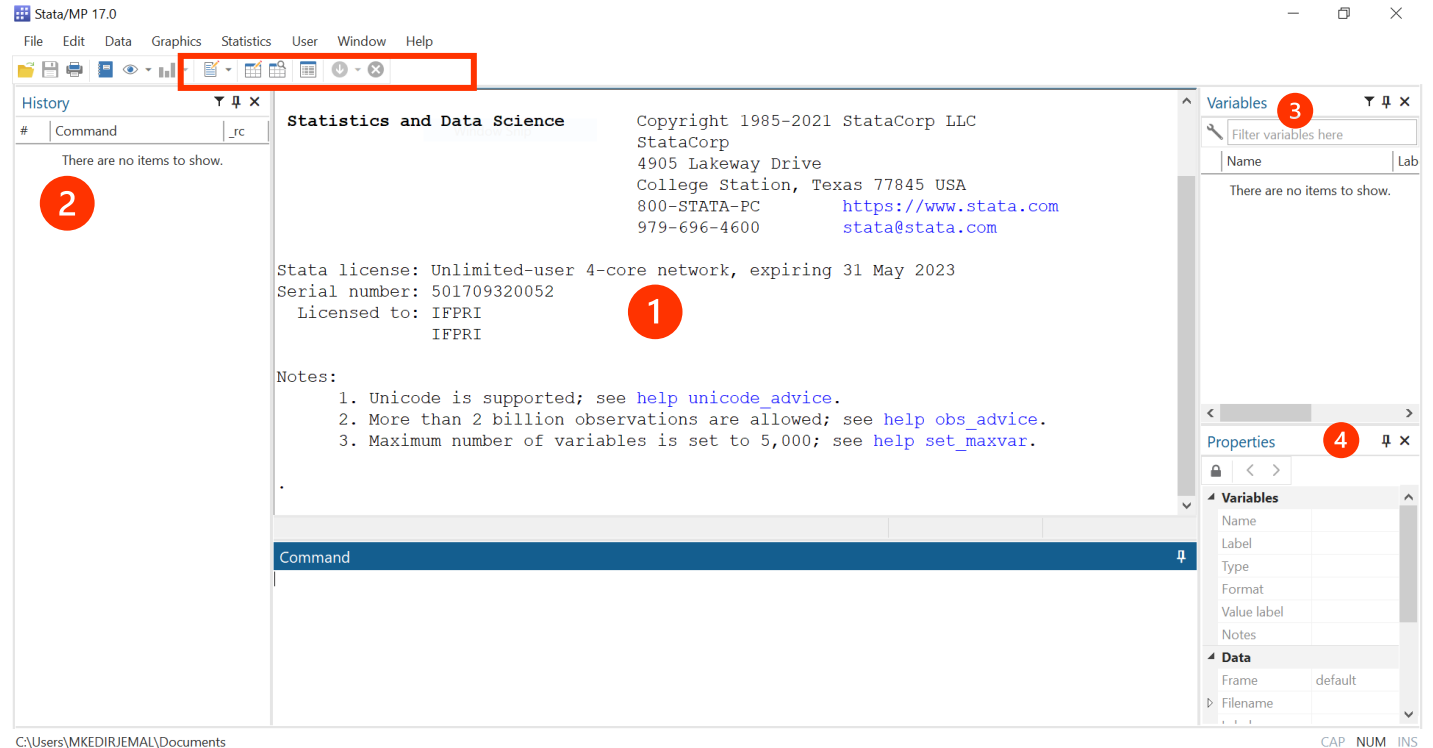
Stata

- STATA is a powerful command driven statistical package with data management, statistical analysis and graphics capabilities.
- It is a complete, integrated software package that provides all your data science needs—data manipulation, visualization, statistics, and automated reporting.
- It is a fast, accurate, and easy to use software.
- STATA's commands for performing tasks are intuitive and easy to learn.

The Stata Interface

- 1 Results window:** shows the results in the larger window
- 2 History:** keep track of command operations used
- 3 Variables:** located on the top right that lists the variables in the dataset being used
- 4 Properties:** displays properties of the variables and datasets

Icons in menu options: Data Editor (edit and browse) , Do file Editor 



Three ways to enter commands

1. Stata has a graphical user interface (GUI) for command entry via menus and dialogs

File >> open

Statistics >> summaries, tables, tests

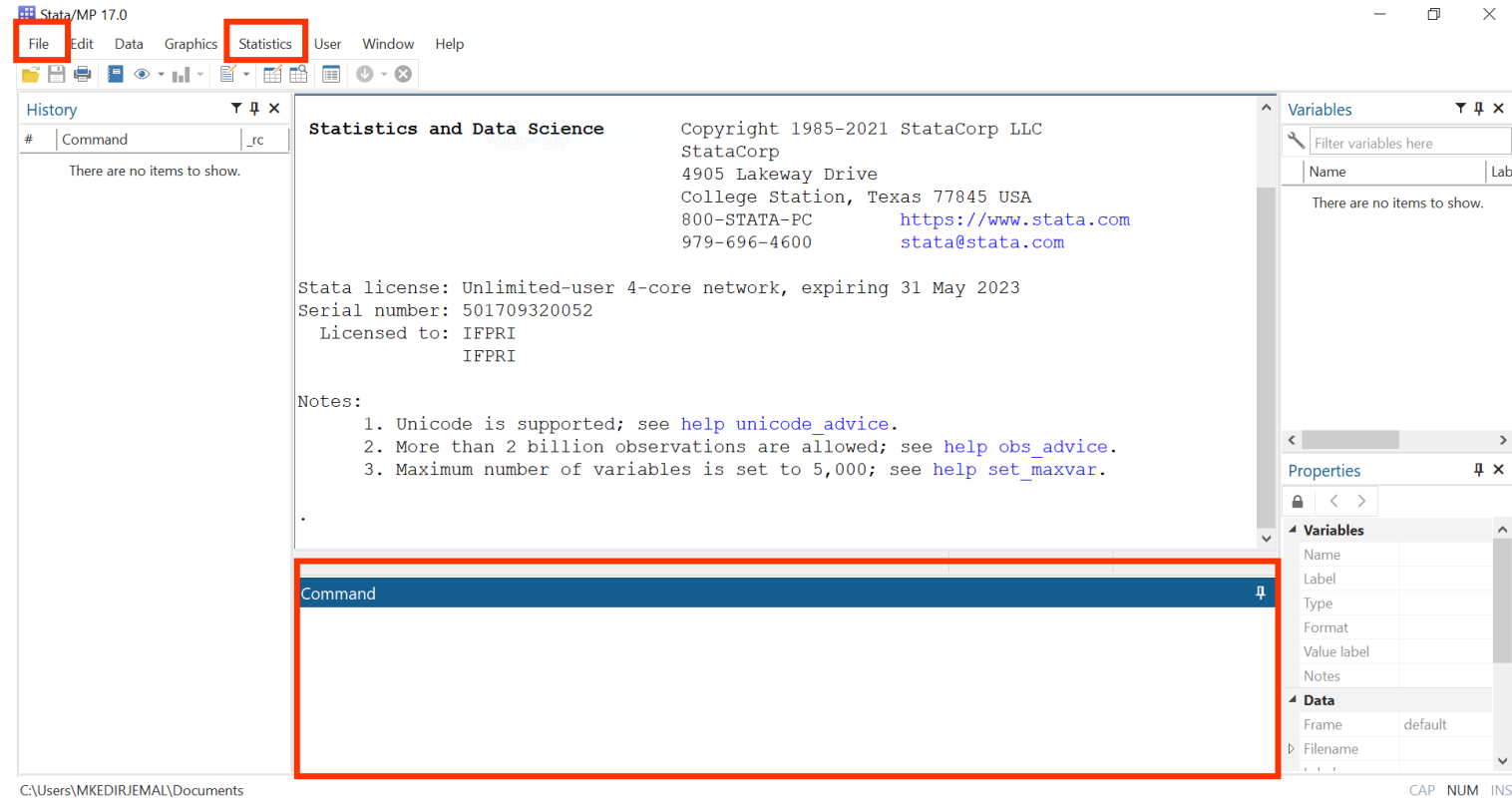
File >> Import >> Excel

2. Command window: type the commands in the boxed listings below into the small window labeled Command

browse

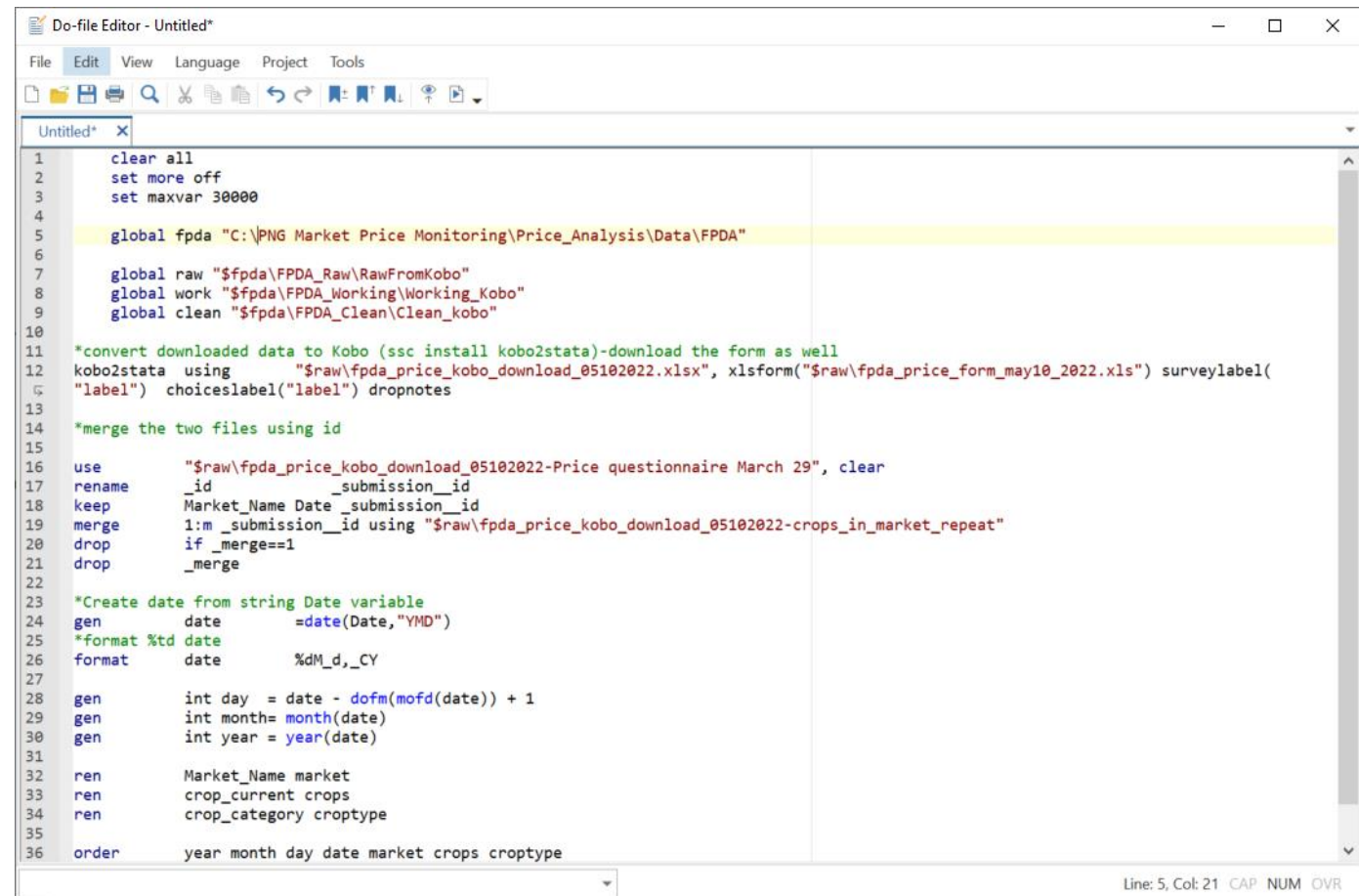
describe

tabulate



3. Do-file: write commands/script in a “do-file” and execute the do-file

- It is a good practice to use do files when performing long or repetitive tasks
- Since do files are stored as permanent records, they are editable in the future.
- Manipulation driven by menus or commands is useful for testing commands/menus on the fly, and then using the command in a do file later.



```
Do-file Editor - Untitled*
File Edit View Language Project Tools
Untitled* x
1 clear all
2 set more off
3 set maxvar 30000
4
5 global fpda "C:\PNG Market Price Monitoring\Price_Analysis\Data\FPDA"
6
7 global raw "$fpda\FPDA_Raw\RawFromKobo"
8 global work "$fpda\FPDA_Working\Working_Kobo"
9 global clean "$fpda\FPDA_Clean\Clean_kobo"
10
11 *convert downloaded data to Kobo (ssc install kobo2stata)-download the form as well
12 kobo2stata using "$raw\fpda_price_kobo_download_05102022.xlsx", xlsform("$raw\fpda_price_form_may10_2022.xls") surveylabel(
13 "label") choiceslabel("label") droptypes
14
15 *merge the two files using id
16 use "$raw\fpda_price_kobo_download_05102022-Price questionnaire March 29", clear
17 rename _id _submission_id
18 keep Market_Name Date _submission_id
19 merge 1:m _submission_id using "$raw\fpda_price_kobo_download_05102022-crops_in_market_repeat"
20 drop if _merge==1
21 drop _merge
22
23 *Create date from string Date variable
24 gen date =date(Date,"YMD")
25 *format %td date
26 format date %dM_d,_CY
27
28 gen int day = date - dofm(mofd(date)) + 1
29 gen int month= month(date)
30 gen int year = year(date)
31
32 ren Market_Name market
33 ren crop_current crops
34 ren crop_category croptype
35
36 order year month day date market crops croptype
Line: 5, Col: 21 CAP NUM OVR
```

Exercise:

Start by loading the 'fpda_price_sample_2021.dta' dataset, located in the training folder.

Note that Stata datasets always have '.dta' extension.

To access existing Stata data file using GUI

1. Select File > your working directory
2. Click on 'fpda_price_sample_2021' in the directory folder.

Explore the data

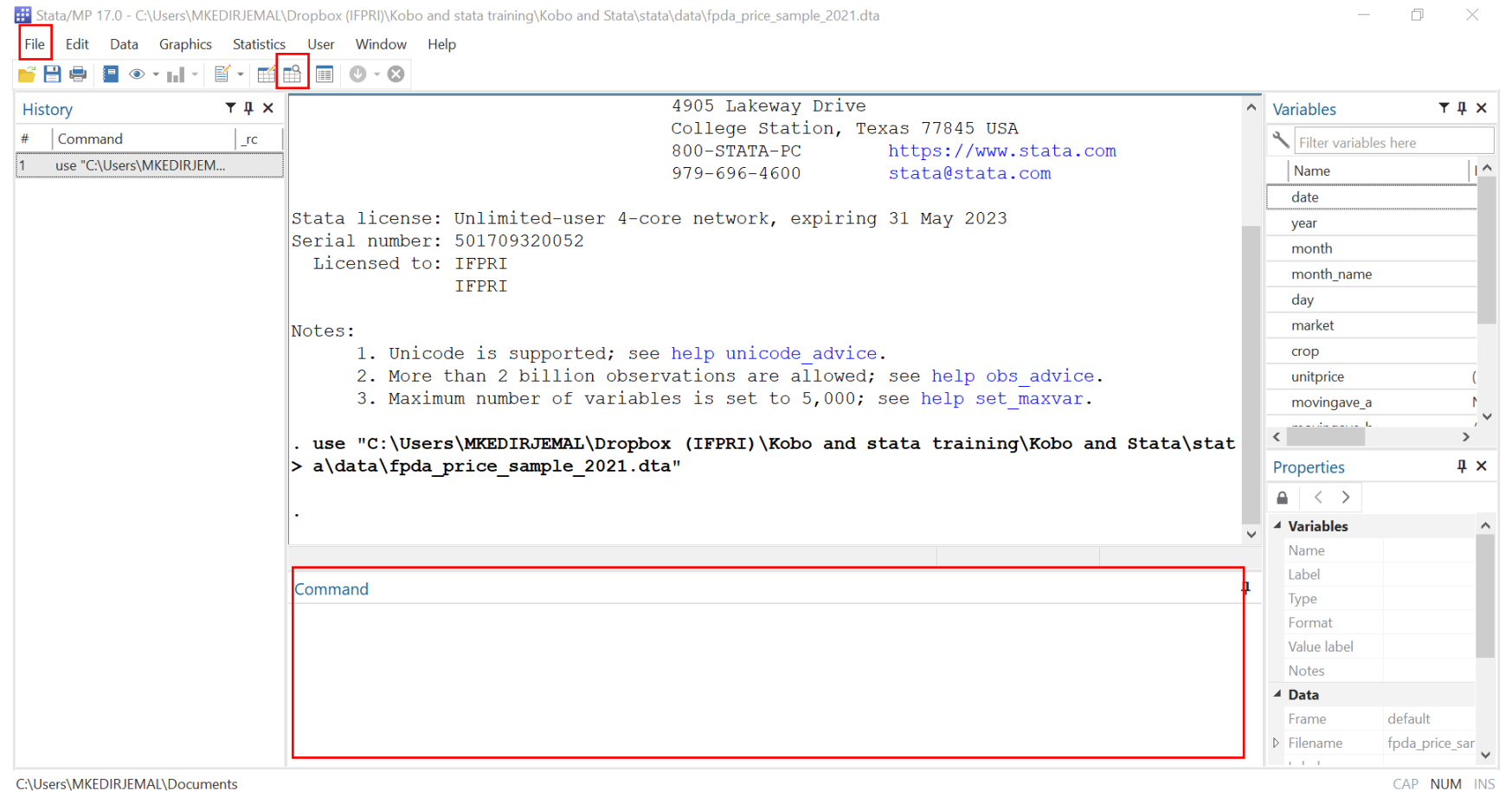
Click Data Editor (Browse)



Or

From the main menu click
Data > Data Editor > Data
Editor (Browse)

Type 'browse' in the
command window



The screenshot shows the Stata software interface. The main window displays the following information:

4905 Lakeway Drive
College Station, Texas 77845 USA
800-STATA-PC <https://www.stata.com>
979-696-4600 stata@stata.com

Stata license: Unlimited-user 4-core network, expiring 31 May 2023
Serial number: 501709320052
Licensed to: IFPRI
IFPRI

Notes:

1. Unicode is supported; see [help unicode_advice](#).
2. More than 2 billion observations are allowed; see [help obs_advice](#).
3. Maximum number of variables is set to 5,000; see [help set_maxvar](#).

```
. use "C:\Users\MKEDIRJEMAL\Dropbox (IFPRI)\Kobo and stata training\Kobo and Stata\stat  
> a\data\fpda_price_sample_2021.dta"
```

The command window at the bottom is empty and highlighted with a red box.

The right-hand side of the interface shows the Variables and Properties panels. The Variables panel lists variables such as date, year, month, month_name, day, market, crop, unitprice, movingave_a, and movingave_b. The Properties panel shows the current variable selected is 'Data' with a frame of 'default' and a filename of 'fpda_price_sar'.



	date	year	month	month_name	day	market	crop	unitprice	movingave_a	movingave_b	movingave_c	movingave_d	movingave_e
1	22apr2021	2021	4	April	22	Banz	Aibika	2.173913	2.173913
2	26nov2021	2021	11	November	26	Banz	Aibika	2.279719	2.279719
3	01dec2021	2021	12	December	1	Banz	Aibika	1.136364	1.955319	1.955319	1.955319	1.955319	1.955319
4	29dec2021	2021	12	December	29	Banz	Aibika	2.347222	1.955319	1.955319	1.955319	1.955319	1.955319
5	17dec2021	2021	12	December	17	Banz	Aibika	2.057971	1.955319	1.955319	1.955319	1.955319	1.955319
6	17jan2022	2022	1	January	17	Banz	Aibika	2.932423
7	24feb2021	2021	2	February	24	Goroka	Aibika	3.270202	3.231061	3.231061	.	.	.
8	11feb2021	2021	2	February	11	Goroka	Aibika	3.191919	3.231061	3.231061	.	.	.
9	24mar2021	2021	3	March	24	Goroka	Aibika	3.611111	3.203914	3.203914	3.203914	3.203914	3.203914
10	10mar2021	2021	3	March	10	Goroka	Aibika	2.742424	3.203914	3.203914	3.203914	3.203914	3.203914
11	28apr2021	2021	4	April	28	Goroka	Aibika	3.020202	3.305976	3.305976	3.305976	3.305976	3.305976
12	21apr2021	2021	4	April	21	Goroka	Aibika	4	3.305976	3.305976	3.305976	3.305976	3.305976
13	19may2021	2021	5	May	19	Goroka	Aibika	3.130159	3.339875	3.339875	3.339875	3.339875	3.339875
14	26may2021	2021	5	May	26	Goroka	Aibika	3.535354	3.339875	3.339875	3.339875	3.339875	3.339875
15	30jun2021	2021	6	June	30	Goroka	Aibika	2.555556	3.051133	3.051133	3.051133	3.051133	3.051133
16	23jun2021	2021	6	June	23	Goroka	Aibika	2.065527	3.051133	3.051133	3.051133	3.051133	3.051133
17	16jul2021	2021	7	July	16	Goroka	Aibika	2.705255	2.786033	2.786033	2.786033	2.786033	2.786033
18	28jul2021	2021	7	July	28	Goroka	Aibika	2.724349	2.786033	2.786033	2.786033	2.786033	2.786033
19	22sep2021	2021	9	September	22	Goroka	Aibika	3.247863	2.892489	2.892489	2.892489	2.892489	2.892489
20	23nov2021	2021	11	November	23	Goroka	Aibika	5.351515	4.390028	4.390028	4.390028	4.390028	4.390028
21	29nov2021	2021	11	November	29	Goroka	Aibika	4.570707	4.390028	4.390028	4.390028	4.390028	4.390028
22	17dec2021	2021	12	December	17	Goroka	Aibika	4.004273	4.123158	4.123158	4.123158	4.123158	4.123158
23	29dec2021	2021	12	December	29	Goroka	Aibika	2.566138	4.123158	4.123158	4.123158	4.123158	4.123158
24	18feb2021	2021	2	February	18	Kokopo	Aibika	.6666667	1.039239	1.039239	1.039239	1.039239	1.039239
25	26feb2021	2021	2	February	26	Kokopo	Aibika	1.581197	1.039239	1.039239	1.039239	1.039239	1.039239
26	31mar2021	2021	3	March	31	Kokopo	Aibika	1.217949	1.203304	1.203304	1.203304	1.203304	1.203304
27	11mar2021	2021	3	March	11	Kokopo	Aibika	1.347403	1.203304	1.203304	1.203304	1.203304	1.203304
28	21apr2021	2021	4	April	21	Kokopo	Aibika	.969697	1.235769	1.235769	1.235769	1.235769	1.235769
29	28apr2021	2021	4	April	28	Kokopo	Aibika	1.631702	1.235769	1.235769	1.235769	1.235769	1.235769
30	26may2021	2021	5	May	26	Kokopo	Aibika	1.242424	1.219711	1.219711	1.219711	1.219711	1.219711
31	20may2021	2021	5	May	20	Kokopo	Aibika	.9090909	1.219711	1.219711	1.219711	1.219711	1.219711
32	23jun2021	2021	6	June	23	Kokopo	Aibika	2.166667	1.432336	1.432336	1.432336	1.432336	1.432336
33	30jun2021	2021	6	June	30	Kokopo	Aibika	1.674437	1.432336	1.432336	1.432336	1.432336	1.432336
34	19jul2021	2021	7	July	19	Kokopo	Aibika	1	1.322002	1.322002	1.322002	1.322002	1.322002
35	31jul2021	2021	7	July	31	Kokopo	Aibika	.9393939	1.322002	1.322002	1.322002	1.322002	1.322002
36	27aug2021	2021	8	August	27	Kokopo	Aibika	1	1.356099	1.356099	1.356099	1.356099	1.356099
37	30sep2021	2021	9	September	30	Kokopo	Aibika	2	1.50303	1.50303	1.50303	1.50303	1.50303
38	22sep2021	2021	9	September	22	Kokopo	Aibika	2.575758	1.50303	1.50303	1.50303	1.50303	1.50303

Variables

Filter variables here

- Name
- date
- year
- month
- month_name
- day
- market
- crop
- unitprice (mean) unitprice
- movingave_a No restriction on minimum # of values averaged during a 3-month period
- movingave_b At least 2 values averaged during a 3-month period
- movingave_c At least 2 values from at least 2 months averaged
- movingave_d At least 3 values averaged during a 3-month period
- movingave_e At least 3 values from at least 2 months averaged
- movingave_f At least 3 values from at least 3 months averaged

Properties

Variables

Name	date
Label	
Type	float
Format	%td
Value label	
Notes	

Data

Frame	default
Filename	fpda_price_sample_2021.dta
Label	
Notes	
Variables	14
Observations	2,310
Size	230.10K
Memory	64M
Sorted by	

Data Manipulation

- The most important arithmetic, logical and relational operators in Stata.
 - Among the most common are **&** (and), **|** (or), and **!** (not).

Logical Operators in Stata	
And	&
Or	
Not	! or ~
Multiplication	*
Division \	\
Addition	+
Subtraction	-
Less Than	<
Greater Than	>
Less Than or Equal	<=
More Than or Equal	>=
To The Power Of	^
Wildcard	*

Describing Data

- **Browse**: there are several ways in Stata to investigate and describe data. You have begun browsing the data in the previous section.

//type browse in the command window

- **browse**
- **browse** market crop year

```
Command  
browse|
```

```
Command  
browse market crop year
```

List

If the dataset is not too large, you can also examine it in the results window by using the *list* command

- For example the observation 1 to 50:

- *//type list market in the command window*

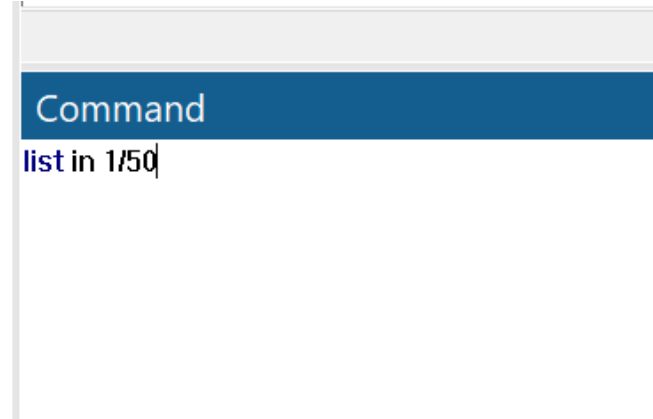
```
list market
```

```
list in 1/50      //lists observation from 1 to 50
```

```
list in 10/29    //lists observation from 10 to 29
```

```
list in 100/1    //lists observation from 100 to last  
observation (lower case 'l')
```

```
list in -10/1    //lists observation from last 10 observation  
(lower case 'l')
```



Assert

One useful command is *assert*, which verifies whether a certain statement is true or false. Check that all unitprice values are positive, for example:

```
assert unitprice > 0
```

```
assert unitprice < 0
```

Describe

This command provides a brief overview of the dataset and its variables (size, number of variables, observations, storage types of variables, etc.).

describe

Observations:	2,310			
Variables:	14			May 2022 17:52
Variable name	Storage type	Display format	Value label	Variable label
date	float	%td		
year	int	%8.0g		
month	int	%8.0g		
month_name	float	%tmMonth		
day	int	%8.0g		
market	str36	%36s		
crop	str24	%24s		
unitprice	float	%9.0g		(mean) unitprice
movingave_a	float	%9.0g		No restriction on minimum # of values averaged during 3-month period
movingave_b	float	%9.0g		At least 2 values averaged during a 3-month period
movingave_c	float	%9.0g		At least 2 values from at least 2 months averaged during a 3-month period
movingave_d	float	%9.0g		At least 3 values averaged during a 3-month period
movingave_e	float	%9.0g		At least 3 values from at least 2 months averaged during a 3-month period
movingave_f	float	%9.0g		At least 3 values from at least 3 months averaged during a 3-month period

Codebook

It provides additional information about the variables, such as summary statistics for numeric, examples of data points for strings, etc. Without a list of variables, codebook will give information on all variables in the dataset.

//the command displays the frequency of each market

```
codebook market
```


Summarize

This command gives summary statistics, such as means, standard deviations, and so on.

You can also specify a variable (e.g. *sum* unitprice) after the command to summarize only specific variables.

If you would like more precise information (e.g. percentiles) then you can add the detail option to the end of that command, i.e. *sum* unitprice, detail.

//in the command window type sum

sum

sum unitprice

sum unitprice, detail

Tabulate

Use this command to create a frequency table or a cross-tabulation of two variables.

Typing 'tab market' will show how many markets are there in the data.

The 'tab market crop', for example, will display the crop and number of observations for each market.

```
tab market
```

```
tab crop
```

```
tab market crop
```

With the tabulate command and the sum(varname) option, it is possible to get a quick idea of the descriptive statistics of some subgroups.

For example, the average sweet potato unitprice per market:

```
tab market if crop=="Sweet Potato", sum(unitprice)
```

```
tab market if crop=="Sweet Potato", sum(unitprice)
```

	Summary of (mean) unitprice		
	Mean	Std. dev.	Freq.
Banz	1.2962642	.19958018	7
Goroka	1.2873904	.63626853	17
Kokopo	1.6164387	.48731619	19
Kundiawa	1.2256389	.15180282	8
Lae	1.8077708	.79371596	16
Mt. Hagen	1.0134973	.34365974	16
Port Moresby	2.7486693	.86287757	17
Total	1.633446	.81793677	100

- Another tabulate command is “tabstat”. This is used for continuous variables, and it is mainly used to determine mean values. For example `tabstat unitprice` will give the average unit price in the dataset. The 'if' command can be used to subset the data to get the average price of each crop by a market.

```
tabstat unitprice, by(crop)
```


```
tabstat unitprice if market=="Lae", by(crop)
```

tabstat unitprice, by(crop)

crop	Mean
Aibika	2.680144
Amaranthus-Aupa	2.510823
Asparagus	21.59091
Banana-Cooking	2.727463
Banana-Ripe	2.192472
Broccoli	7.538422
Cabbage-English	2.699988
Capsicum	7.270709
Carrot	6.220998
Cassava	1.179519
Choko-Tips	2.612188
Cucumber	2.427502
Fern	2.421382
French/Dwarf Bea	5.624007
Garlic	29.28704
Ginger	6.740609

crop	Mean
Karakap	2.577978
Lemon	4.917943
Lettuce	11.64506
Mandarine	7.818021
Onion Bulb	7.531552
Orange	4.842498
Pakchoi	4.208492
Pawpaw	2.32537
Peanut	7.747872
Pineapple	3.982939
Potato	3.21211
Pumpkin-Tips	2.018589
Rice	3.979762
Sweet Potato	1.633446
Taro True	3.721406
Tomato	5.892871
Watermelon	3.935437
Wongbok	5.702879
Total	4.417942

Working with do file

- Rather than typing commands one by one interactively, you can enter them all at one time in a do-file, then run the do-file once.
- Upon completion of the do-file, the results of each command can be reviewed in a log file.
- Stata has a built-in editor - simply click the pad-and-pencil icon  on the from the main menu.
- Text editors such as Word and Notepad can be used to create Do-files.

Major commands

- *Clear* - clears any data currently in Stata's memory, closes open files, windows, and dialog boxes while clearing data, labels, and stored results. Then you can run a new do-file and clear all previous work.
- *cd "[path]/[working directory]"* sets your working directory, the location from which you
- *capture log close* - closes an already-opened log. you will retrieve and save your data/files.
- *log using sample1.log, replace text* - creates a log file of all the results.
- *set more off* - Stata will not pause and display the ---more--- message in the results window for you to review each page on-screen and press a key to get more. Stata will instead run the entire do-file without pausing.
- *set memory 100m* - specify 100m as Stata's default memory might not be enough for large datafiles.
- Use *"[pathway]/[filename]"* - to open or retrieve data from the current directory
- Save *"[pathway]/[filename]"* - saves data to working directory

Comments/Notes

- `//` - Stata ignores two consecutive slashes (or a slash followed by an asterisk `*`)
- `/* */` - a note can be inserted inside these pseudo-parentheses after a command.
- `***` - describe a task or line of code.
- `*` - notes or explanations of what a line of code does.


```
1
2 *****
3 * FPDA food price survey data: simple analysis training
4
5 *Stata programs are referred to as do-files. The files have the suffix .do and contains the Stata commands that you want to execute. off of your work.
6
7 * Do file provides basic information on how to get started with Stata:
8
9 * Commands
10 * 1. Review a) changing directory: cd "...directory..." and b) changing paths
11 * 2. commenting commands and commenting "good habit"
12 * 3. description of files: do, log, scml, dta
13 * 4. basic statistics: means, standard deviations, min, max
14 * 5.a-variable information: # of observations, describe, codebook
15 * 5.b-types of variables: binary, categorical, ordinal, continuous
16 * 6. logical & relational operators: =, ==, >, >=, <, <=, !=, &, |
17 * link to set theory: union and intersection
18 * 7. Arithmetic/String operators: +, -, /, *, ^
19 * 8. if (link to set theory of sub-sets) and by commands
20 * 9. generatiing variables: gen (across variables); egen (across observations)
21 * g byte (binary variables)
22 * ordinal variables
23 * 10. test of means of the variables: ttests (continuous) and tab var1 var2, row col chi2 (binary)
24 * 11. saving data
25 * 12. collapsing data
26
27 *Begin typing your commands. This clears your workspace, frees Memory for faster calculations, opens your data file, and displays a Stata log.
28 clear all
29
30 set more off
31 set memory 100m
32
33 capture log close
34
35
36
37 **1: DIRECTORY
38 *****
39
40
41 * A. The following command, cd, changes your working directory. This enables us to tell Stata where to work.
42 * from where to get the files, save the files, and send the output
43 *cd "C:\Users\MKEDIRJEMAL\Dropbox (IFPRI)\Kobo and stata training\"
44
45 * B. setting a global path is another way to more efficiently refer to files
46 global dir "C:\Users\MKEDIRJEMAL\Dropbox (IFPRI)\Kobo and stata training\"
47 global data "$dir\data"
48 global analysis "$dir\analysis"
49 global graphs "$dir\graphs"
50
51 **2: COMMENTING
52 *****
53 * An asterisk (*) at the start of a line allows you to remark on the line but does not execute the command.
54 * as shown below. Text in GREEN means that it is a comment. Text in BLUE means that it is a Stata key word to be run
55 * clear
56
```