Handling Large-Scale Unit Level Data Using STATA Professor Pratap C. Mohanty Department of Humanities and Social Sciences, Indian Institute of Technology, Roorkee Lecture 13 Managing Data in Stata - I

Welcome once again dear learners to the NPTEL module on Handling Large-Scale Data Using Stata, and we are explaining how to manage data with the Stata. So far in last two lectures we have been trying our best to introduce you with Stata and there are many pointer options or command options we discussed, but now onwards it will be more practical oriented. Though earlier classes were also practical oriented, but the appetite of understanding Stata was through data. So we introduced data earlier. Now onwards we are using data with the help of Stata software. So, deliberately the title is defined as Managing Data in Stata.

So, once again, for your understanding, I am going to use the Stata 15 software and I am again suggesting you to operate yourself through the demo version or if you have a purchased version or licensed version, please open that and start working and go through our instructions and also interpretations.

(Refer Slide Time: 01:53)



When you start your work with the dataset, first of all, it is always good to understand the dataset and try to clean up as much as possible for smooth functioning and smooth interpretation of the results. So each data file contains some documentation, because without any structured presentation of the data, which is provided by the surveyor or the organization which make data, if they do not make it in a structure, it is very difficult to understand. So to make it, put it in a structure, they sometimes compress, they go by systematic commands and so that the data is in structured work for us that just cannot be extracted and used. You have to mine the data or simplify the data a bit and start explaining it. So here we are to help you how to simplify the data and make it in a structure.

So the tip here is given that always go through the documentation file, every data. We have already shown those documentations with NSS, National Sample Survey, IHDS and National Family Health Survey data. Please check our previous lectures, you will be guided with the documentation very clearly. And before starting to work, it is once again suggested you to please check the documentation. We will also guide you side by side. Unfortunately, not all datasets have good documentation or any documentation at all, so understanding the nature of data itself is a vital skill for all of us.

(Refer Slide Time: 03:48)



As highlighted it is always advisable to examine your data when you first read it into Stata to check all the variables and observations are present and in the correct format that is very essential.

(Refer Slide Time: 04:05)



The practice dataset for today's lecture we are going to use is National Sample Survey 73rd round that was published in 2015-16 and we are focusing on block number two. There is a purpose behind using this. Our title of the module is handling large-scale unit level data with Stata. Since the title is large-scale, we are trying to equip you all with the exact and core dataset which are available for social science discipline as well as for management discipline largely. Others can also get the advantage out of it.

What I am suggesting you, we will also provide whatever we are using, we also provide these database to you. These are going to be your added materials. NSS 73rd round block number two. This data is also available as I already guided you in earlier lectures, this is already available but all the blocks are available. If you go by all the blocks, the data will be very large enough and it consumes huge byte space, the memory is consumed. So when you operate, it takes time. So, we will guide you side by side how to filter and simplify.

So, for your simplicity, we have deliberately drawn a sample of 25% and the command is given. You can experiment on your own. So from the original dataset we have reduced to 25% of the total sample for easy operation and understanding of the data. (Refer Slide Time: 06:00)





Let me proceed to start working with Stata. Let me open the dataset first. We will also provide you the data, the 25% of the sample data to you along with this PPT and since you have already downloaded the Stata version. I assume that you have downloaded so far. The demo version you have got so far or you have purchased.

So, please keep three things together. First is one sample data. Otherwise, if you do not have sample data, it is becoming difficult for you. I guided you in the last class about Stata directory and we used life expectancy data. You can also operate from there as well. This is a very simple

small dataset and very quickly you can understand applying the command we are going to suggest.

Let me tell you what you are supposed to collect at this moment. A Stata version must be there. Maybe 13 onwards version, but at this moment even older version is going to be useful except few commands, like I told you already help command, search command, find it command was earlier used. But still most of the operation you can do with the earlier version. If those are freely available by any reason, you can also procure that. You can write it to StataCorp and they will guide you.

So here is the window for you to start working with the data I have opened. Once again let me tell you, I will just guide you here since I have already opened here. So you go to the file and file menu, then you open. Once you open, I have opened here the practice_dataset_NSS 73_block2.dta. I have already guided you .dta file must be there. Otherwise, it might be problematic. Some data can be also imported with different versions, like excel or maybe other whichever are supported, you can also open that. But at this moment I am working with .dta. So I have this.

I am also going to use some other aspect to it. Later on, I will be discussing about IHDS data, practice dataset IHDS. I will tell you later, but at this moment I am using NSS 73rd. So these, after opening that all the variables list are visible at the right hand side under variable heading and their respective properties are also getting visible to all of you.

Here is the data review file as I told you, if you click on tab, it keeps on rotating with the availability of review, then variables, then properties, then command, then result. Result is not visible through tab, but others are visible. So let me proceed. Start with understanding browse or edit, as I already told you. Browse starts a pop-up window in which you can examine the raw data. Edit allows you to manually change the dataset.

So, in the last lecture, I specifically mentioned that if you just want to check the data and their nature, try to go by the browse not by the edit. So if you want to change the data, then you click on edit or you type edit then the variable or you wanted to change the particular value within the variable, then you go by edit that is separate, but always it is suggested to open the browse data unless it is necessary.

So most of the time, we only want to view a few variables at a time. All the variables like let me start with. I will operate now. There are two options, simply you can click. At the tool bar menu, there are two. First one is I can highlight here data edit, then data browse, that is one option. If I just click on browse it will show you the data. The data fitted in .dta format and it is now visible to us.

Instead of open that easiest way you can simply type br and enter, it will also direct you to the same window. Else if I just click on edit, it will also give you the same data. But there is a difference, here you can change the data, but in br command, browse command, you are not going to change the data. You can only see the data. So, as I told you already that seeing the data is different than that of editing the data. It is safe to see the data. When you are doubly sure that you are going to change the data then open the edit command.

So, let me move on to our next operation. As I told you, usually when you open the data there are so many variables. You look at this. I have already told you that just by br command I have opened the dataset before you. There are so many variables listed. You can just scroll to right you will find so many variables.

But so many variables you may not be interested to look at a given time. You may simply type br, any two variable name or three variable name or even one variable name. Whatever you are interested to carefully check, just click on or type that particular variable. I am just putting a double click here. I am there to switch the enter button. So it has given me only two variables information in the browse. Similarly, in the edit command as well you can go through.

(Refer Slide Time: 12:52)



Let me proceed further to other details. So here is the detail. next important aspect for our understanding is list, listing unlike browse and edit command, list command helps you to examine the data within the result window. Like help window and chelp window we told you earlier one is in another new window will open up and in the c command, chelp, it gives you the command within that same window.

So similarly list command helps you to examine the data within the result window. Like when I browse or edit another window popped up. But if I am interested to look at some variables and a specific aspect of it, I can go by list, because it will be visible on the same window, result window. Result window we mean here only on your screen. So, I will operate and let all this.

So this command will produce a listing of every variable, value for every subject in the dataset. This command is not visible for large datasets. Why? Because so many information will be visible on the screen and all your screen space is blocked and it will be very difficult to scroll down till the end and it will be very uneasy. That does not mean it will not be visible. It will be very uneasy to check. I am not going to do that all together. If I do it all will be opening on the same screen it will be suggested not to open that if you want to take entire, better to open in the new window. So, list is important.

(Refer Slide Time: 14:47)

Listing is more manageable when used for some variables or options. E.g. This command will produce a listing of list var1 var2... the values of the two variables (variable1 and variable 2) for every subject in the dataset. □ list var1 var2 in 1/5 This command will produce a listing of the values of the two variables (variable1 and variable 2) for the first 5 subjects in the dataset. 33 212 244 181 194 362 321 274 031 131 272 102 293 272 093 091 061 095 272 022 101 022 285 221 235 195 184 085 🛞 🖉 🖸 🛹 🛄 🌍 🛄 🖪 🟭 • 2 📀 4 🖬 h 😽



We can do the listing more manageable when used for some variables or options. As I just told you a couple of minutes back, options are important. List variable one and two, you will get the listing of the values of two variables for every subject in the dataset like this. Suppose I do here for you. So I just typed list variable one here, then variable two let it be here, I just enter. Now look at this, these are visible on the same screen. But I am interested in so many variables I wanted to check I will suggest you to go by the browse.

List variable, this is going to be more interesting. , if there are some lakhs of observations, usually the dataset we are handling has more than 2 lakhs or even 6 lakhs observations are there. So, 6 lakhs observations and you are going to show by the variable on the screen is also going to be cumbersome. So what do we do, if some clarity are required particularly from five observation or maybe last five observation or maybe some portion of the observation, we have some separate command.

Here this command will enable you to check a listing of the values of two variables for the first five subjects. Two variables were put here as variable one and variable two, so two variables are there for you. For the first five subjects in the dataset with the command in 1/5, list variable in 1/5. Let me operate it for you. So what we have been dealt so far is understanding some portion of our data, let it be 1/5. So I am deliberately trying to open the Stata once again to show you how 1/5 is being displayed in the command. How it is important, let me clarify before you. Yes, it is here.

So, how to do it? So I will discuss list then two variables, then in 1/5. So let me click here by list, then as I told you two variables, just double click on two variables, this and this let it be two variables, then I am using in 1/5. So, let us enter. So it has given us the details of 1/5 of the observations.

(Refer Slide Time: 18:14)



Let me understand all the details from our explanation. Similarly, if you want to list the values of the two variables for the last five subjects in the datasets, last five subjects instead of 1/5 here last means l, let me help you, l is important. So it is even highlighted letter l in lowercase, not in

capital, because Stata is very sensitive with the exact command. If you are deviating from it, there may be some errors.

So if I just show you once again with the command like what I try to do, I will go by the previous command. If you use page up, you will get that command but here also I can show you page up. So page up command is now visible to me. For you just click on the page up key in your keyboard you will get that. Now I am going to delete this 5, I will just show you minus 5 slash small l, now enter.

It has given you the last five information. So it is very important, though it is visible as not so important, but you have to be very handy in your operation. You will certainly use it. Those who wish to have a mastery on Stata, these commands are very very important for better explanation.

Similarly, list if you want observation 60 till the last, then 60 slash l is going to be important. small l, l in lowercase we are suggesting you to enter. Similarly, list variable one if variable one double equal to certain value. If you want to get certain values and how variable one looks like with certain values is very important. This command will produce result for variable one with limited observation that specifies certain values, any value as a center and you will certainly get that.

But if command is essential, because you are conditioning to a particular value. So I am not going to explain you. If you are stuck, we will clarify in the doubt class.

(Refer Slide Time: 21:00)

□ <u>d</u> escrib)e-			
□This data	command p aset.	provides ove	rall information	about the
dos			nan	ne
. des				
Contains > a obs: vars: size:	data from F:\Dropbo 72,528 42 19,437,504	X\NPTEL on data and	1 analysis\MOOC\practice	Creation date/time
variable	storage disp name type form	lay value at label y	variable label	
				9
File Edit Data Graphics Statistics User	Window Help			
Review T 4 x	Sample State_Region in -5/1			Variables T 4 > K Fiter variables here
Command Jr. Lore "Gipractice, dataert,N. Lore "Gipractice, dataert,N. Lore Same, Region in. List Sample Same, Region in. T2524. T2526. T2526.	Sample State_*n 1 221 1 225 1 195 1 184 1 085			Name Lubel RNID Ørinari key - un. Centró colek. Centre cole Round VII. Bia Sino - FSU Senai No. Round Schedule Stangle Sample Sample
. d Contain obs: vars:	s data from G:\practice_dataset_NSS 72,528 42	73_block2.dta 6 Aug 2020 14:26		State, Region State, Region State, State State, Danict State, Danict State, Danict Statum Statum Substatum SubStatu
5120:	storage display value			Properties 4 >
BNTID CentreC Vill_Bi Sound Schedul Sample	<pre>w name vype format interi str30 %30s deated str5 %5s % str5 %15s str6 %6s * str6 %9s str3 %3s</pre>	Primary key - unique identifi Centre code, Round FSU Serial No. Round Schedule Sample Sample	er for an enterprise)	Variables Name State Report Label Quarter State Report Label Quarter State Report Type strip format State Varies Notes
Sactor State_N State State_D Stratum	strs x3s sqion str9 k9s str6 k6s istrict str12 k12s str6 k6s	==COF State-Region State State-District Stratum		Data Sename practice_dataset_N Label Notes Vanables 42
Commany	d			Coservisions 12.528 Size 18.51M Memory 54M Sorted by
(Inner)	9 📋 🗖 🔡			- ♥ ♥ ↓ ₩ № 4119M

s 🖬 📾 🗐 🖻 • ⊥ • 💥 • 🖻	iii ⊡ O · O					
eview T a x	SubStratunNo	stró	163	Sub-Stratum-no	Variables	•
Citer community here	SubRound	str3	\$35	Sub-Round	A Discounted	
Inter contraints nere	SubSample	str3	\$35	Sub-Sample	The manage	
Command	FODSubRegion	str12	\$125	FOD-Sub-Region	Name	Label
use "G\practice_dataset_N	NG_SubBlkNo	str3	\$35	Hamlet group/Sub-block no.	ENTID	Primary key - u
list Sample State_Region in	second_stratum	stro	603	Second-Stage-Stratum-no	CentreCodeR.	. Centre code,Rou
list Sample State Region in	ent_no	Stro	603	Sample enterprise No.	Vill Blk Sino	FSU Serial No.
4	Level	SULO	603	Lievel 2-block 2.	Round	Round
•	b2_q201	3123	433	whether pursuing mixed activity/(yes-1,no-2)	f shadeda	future de la
	b2_q202	strl	6103	Description of Major activity during the reference month code (5 digit as per N1	screose	schedule
	b2 g207	atr3	23	if item 204 is 11: to 15: social group of the owner/major martner	Sample	Sample
	b2 c208	atr3	\$3+	Shathar the enterprise maintain any hant account/most office saving hant account	Sector	Sector
	b2 a210	strl	131	location of the enterprise	State_Region	♦ State-Region
	h2 a211	strl	834	Enterprise type during the last 365 days	State	State
	b2 g212	str12	\$125	Year of initial operation under the current owner	State_District	State-District
	b2 q213	str3	135	Nature of operation	Stratum	Stratum
	b2 q216	str3	135	Whether accounts maintained? (yes-1, no-2)	California	n Sub-Orabaman
	b2 q218	str3	135	Did the enterprise use computers during the last 365 days? (yes-1, no-2)	5005030000	1 340-30 800000
	b2_q219	str3	835	Did the enterprise use internet during the last 365 days ? (yes-1, no-2)	Subikound	Sub-Round
	b2_q220	str3	835	Did the enterptise face any problem in its operation during last 365 days? (yes-	Properties	
	b2_q223	str3	133	Did the enterprise receive any assistance from the government during last three		
	b2_q226	str3	\$3s	Status of the enterprise over the last 3 years	A Variables	
	b2_q227	str3	\$35	whether registered under any act/authority? (yes-1, no-2)	- variaures	
	b2_q237	str3	\$35	Does the enterprise undertake any work on contract basis (yes-1, no-2)	name	
	b2_q244	Strb	\$35	does the enterprise have access to toilet facility?	Label	⊕State-Regi
	N\$5	double	%10.0g	number of first stage units surveyed within a substratum for the sub-sample	Type	
	NSC	double	%10.0g	number of first stage units surveyed within a substratum for sub-sample combined	Format	%95
	MLT	double	810.0g	weight or multiplier within a substratum for the sub-sample	 Value label 	
	Weight_SS	double	\$10.0g	Final weight for sub-sample wise estimates	Notes	
	Weight_SC	double	810.0g	Final weight for sub-sample combined estimates	4 Data	
	and here				- Filename	practice data
	sorted by:				Label	
					Noter	
					North New York	
						42
	Command				Uoservations	12,528
					2/24	18.54M
					Memory	64M
					Sorted by	
141-15						7.10 MI
18181.7						U.V. NUI

Other details are there, like describe. You might have heard in many articles, many journal papers, description is very important. Though variable description is a different one, but some information we can derive from the describe table, describe table is always very important. The command is highlighted in blue color. For you, I am going to show you how describe is going to clarify you.

So, let me type it here, simply d. D is also going to give you the result. underscore of d was there. So you just click on, you will get the data or ds or describe whatever you type you will get the information for sure. Look at how clearly you get the information of your data which we are trying to work. Look at this, the data is visible before you generated on August 6, 2020 at 2:26 or 14:26 hours. So, it is also showing that practice dataset NSS 73rd which I have already told you.

It also gives information like storage types, whether a string or in double entry or in numeric all those details are also going to be there. So, the second column is explaining that. So variable name on the first column, but variable label is very very important for a researcher to understand the data, labeling of the variables, which label, which name is there and you can also rename those labeling. I will guide you in between.

Also look at other entries here, like all the variables information are given, double, as I told you, string 3, I told you one string str1 till 2,045 entries are there for string values, string 1. Sometimes the long string, large string is coded as strL. So, other details we will guide you in between, let

me proceed. Even the dataset name as I already told you with their location, its directory or the file name is clearly highlighted here.

(Refer Slide Time: 23:49)



riew ₹₽×	Level	stró	\$6s		Clevel 2-block 20	• Variables	T I
Fulter commands here	b2_q201	str3	\$35		Whether pursuing mixed activity?(yes-1,no-2)	A Litter sanable	es here
	b2_q202	stris	\$155		Description of Major activity during the reference month code (5 digit as per NI		
Command	b2_q204	SUL3	833		Type of ownership code	Name	Label
use 'G\practice_dataset_N	b2_q208	4742	\$2.		Whathar the enterprise maintain any hank account/most office saving hank account	ENTID	Primary key - un
list Sample State_Region in	h2 m210	str3	135		Location of the enterprise	CentreCodeR.	. Centre code,Roun
list Sample State_Region in	b2 a211	str3	131		Enterprise type during the last 365 days	Vil_Bik_Sino	FSU Serial No.
d	b2 g212	str12	\$12.5		Year of initial operation under the current owner	Round	Round
d State	b2 g213	str3	135		Nature of operation	Schedule	Schedule
	b2 g216	str3	135		Whether accounts maintained? (ves-1, no-2)	Samola	Samola
	b2 q218	str3	\$35		Did the enterprise use computers during the last 365 days? (yes-1, no-2)	Sartas	Sample
	b2 q219	str3	135		Did the enterprise use internet during the last 365 days ? (yes-1, no-2)	Sector	sector
	b2_q220	str3	135		Did the enterptise face any problem in its operation during last 365 days? (yes-	State_Kegion	♦ State-Region♦
	b2_q223	str3	135		Did the enterprise receive any assistance from the government during last three	State	State
	b2_q226	str3	\$35		Status of the enterprise over the last 3 years	State_District	State-District
	b2_q227	str3	\$35		whether registered under any act/authority? (yes-1, no-2)	Stratum	Stratum
	b2_q237	str3	135		Does the enterprise undertake any work on contract basis (yes-1, no-2)	SubStratumNo	o Sub-Stratum-no
	b2_q244	str3	135		does the enterprise have access to toilet facility?	SubRound	Sub-Round
	NSS	double	\$10.0g		number of first stage units surveyed within a substratum for the sub-sample		
	NSC	double	\$10.0g		number of first stage units surveyed within a substratum for sub-sample combined	Properties	
	MLT	double	\$10.0g		weight or multiplier within a substratum for the sub-sample		
	Weight_SS	double	\$10.0g		Final weight for sub-sample wise estimates	4 Variables	
	Weight_SC	double	\$10.0g		Final weight for sub-sample combined estimates	Name	
						Label	
	Sorted by:					t and	
	4 0000					1994	
	. u state					Pormat	
		torne	dimlau	and the		Value label	
	wariable name	tume	format	label	variable label	Notes	
						4 Data	
	State	stré	165		State	Filename	practice_datase
						Label	
						 Notes 	
						Variables	42
							22.528
	Command					4 Site	LR SEM
							0+M

The primary goal of running describe before you is to find what variables you have and what they are called, variable labels as I already guided you. If you want to know about only one variable or more to describe variable list, like in the screen which we have just shown you describes entire variable.

You might forget one thing, last class I started guiding you, when you operate any kind of trial and error, even any kind of operation or management with Stata, I always suggest that you start with a log file. I will guide you in between do we mean by log file. Last class I started, but I did not guide you in detail.

Here is the option I am just pointing out it will show you log begin, close. Otherwise, file then log then you begin. If you do that, and there are two options, I will guide you all the details are going to be dealt in this lecture only, like log files comes with smcl or in text format. Text format is suggested because we generally take those result to word format or excel format for further operations, for further manipulations. So I will guide you all detail, but I will suggest you just click on this and save with your filename. Rest of the details I will guide you in due course of time.

Variable list, as I told you, variable list or the describe variable, as I already told you, start with a d, enter with a d command or des command or even describe command it will give you the result. If you are interested to go for only one variable, you simply click on that variable, like click here, yes I will again type d or you just go by your page up command, you will get that

same command, but now I am only using one variable. One variable, let me use another variable called state, now enter. It gives the information of one variable.

When you are quite sure you wanted to operate with a particular variable, so why to check so many. So many may not be going to be very important, one is important, you just click on that and find out. Otherwise, if less number of variables are there you can check at a go. So, variable list is a list of one or more variables on which the command is to operate, the subject of the verb.

Stata works on the concept of a single set of variable currently defined and contained in memory, each of which has a name. We have already mentioned that basically this suggests that you click on the variable, you will get the list. If you want more variables, you click more variables then you will get the information. The varies specifies which of the defined variable are to be used in the command.

(Refer Slide Time: 27:33)



view T # x							Variables	τņ
Filter commands here	HG_SubBlkNo	•					A Filter variable	
Command In	ent no	0					Name	Label
use 'G\oractice dataset N.	Level						ENTID	Primary key - un
list Sample State Region in.	b2_q201	0					CentreCodeR.	Centre code.Round
list Sample State Region in	b2 a202						VIL8Ik,Sino	FSU Serial No.
đ	b2_q202 b2_q204	ŏ					Round	Round
d State	b2_q207	0					Schedule	Schedule
94	b2_q208	0					Sample	Sample
	b2_q210	٥					Sector	Sector
	b2 a211	٥					State_Region	♦State-Region♦
	b2 g212						State	State
	b2_q213	0					State_District	State-District
	b2_q216	0					Stratum	Stratum
	b2_q218	0					SubStratumNo	Sub-Stratum-no
	b2 a219	0					SubRound	Sub-Round
	b2 q220	0					Properties	
	b2_q223	٥						
	p5_d556	0					 Variables 	
	P5_d521	0					Name	State
	b2 q237	0					Label	State
	b2_q244	0					Type	str6
	N.5.5	72,528	2.47845	3.76205	1	31	Format	5.63
	NSC	72,528	4.95206	7.516038	1	62	Value label	
	RL1	12,520	00013.00	313336.7		2.320107	Notes	
	Weight SS	72,528	868.1966	3153.567	.84	231973.5	4 Data	
	Weight_SC	72,528	434.4895	1576.872	. 42	115986.8	C Filename	
							Label	
							Notes	
							Churanation	12.528
	Command						- Sune	
							Memory	64M
itata15			100					CAP NUM
🕑 🖉 🔘 🧇 I	🖳 🗿 📗						- 1 🖸 📢 🕇	8/10/2020
								6/16/2020

	ENTID	CentreCode-d	Vill_81k_51no	Round	Schedule	Sample	Sector	State_Region	State	State_Dist+t	Stretum	SubStratundo	SubRound	SubSample 1	. V	ariables	
1	6950911502	738	69509	73	234	1	1	091	0.9	0902	02	02	4	1	Ш,	+ Fiter variable	
2	8383911101	730	83839	73	234	1	2	031	03	0902	02	01	1	2	s	7 Name	Label
3	5116511403	738	51165	73	234	1	2	212	21	2129	29	01	3	2	l.	ENTID	Primary key - un.
1	6778391101	738	67703	73	234	1	1	245	24	2412	12	01	3	2	5	CentreCodeR.	Centre code Round
5	6768911403	734	67689	73	234	1	1	241	24	2421	21	01	1	1		VII Bik Sino	FSU Serial No.
6	8777411403	738	87774	73	234	1	2	272	27	2725	39	02	1	1		Round	Round
7	5915310702	73£	59153	73	234	1	2	091	09	0909	72	13	3	2		Schedule	Schedule
	6861511404	73 F	68615	73	234	1	1	151	15	1506	06	02	1	1	-	Sample	Sample
9	5355411603	730	53554	23	234	1	2	061	06	0602	02	02	1	2		Sector	Sector
10	5912111502	738	59121	23	234	1		093	09	0957	\$7	03	3	2	5	State Region	éState-Regioné
11	5963811404	738	59638	23	234	1		092	09	0933	75	10	3	1		State	State
12	9603711601	730	96037	73	234	1	1	321	92	3203	03	02	1	1		State District	éState-Districté
13	8193510201	137	81935	73	234	1	2	193	19	1916	20	01	2	1		Stratum	Stratum
14	3876711602	738	38767	73	234	1	1	142	14	1408	08	03	2	2		SubStratumNo	Sub-Stratum-no
15	9530711502	738	95307	73	234	1	1	333	33	3311	11	04	1	1		SubRound	Sub-Round
16	9519310201	738	95193	73	234	1	1	341	34	3404	04	03	2	2		Variables Snaps	hots
17	5355811605	730	53558	73	234	1	2	061	06	0601	01	03	1	2			
18	5823610801	738	58236	73	234	1	2	194	19	1909	09	02	3	1		roperties	
19	6966711501	738	62667	73	234	1	1	091	0.9	0909	09	03	1	1	1	Variables	
20	5733211401	73A	57332	23	234	1	2	242	24	2407	07	01	1	2		Name	
21	5736210603	738	57342	73	234	1	2	241	24	2425	29	11	4	1		Label	Primary key -
22	6893921501	730	68939	23	235		1	111	11	1101	01	03	3	2		lype	
23	6198920101	738	61989	23	234	1	1	192	19	1907	07	01	3	2			
24	6379811502	730	63798	23	234	1	1	032	0.9	0322	22	03	4	2			
25	8969411302	738	83691	23	234	1	2	095	0.9	0915	73	01	2	2		Notes	
26	9194321601	738	91943	73	234	1	1	195	19	1914	14	0.5	2	2		Data	
27	9858821402	735	38588	73	231	1	1	161	16	1602	02	03	1	1		hiename	
28	6687321401	733	66873	23	234	1	1	285	28	2811	11	01	4	2		Label	
29	8601711302	738	86017	23	234	1	2	331	33	3302	33	01	2	1			
50	6701211301	733	67012	23	234	1	1	275	27	2709	09	01	3	2			12.528
31	6952811401	738	69528	23	235	4	1	092	0.9	0924	24	02	4	2			18.510

nie Edit Data Graphics Statisti	cs User Window	нер				
3 🗟 😂 🗏 🖻 - 🖬 - 🕺 - 📓	🗎 🔲 O · O					
eview T # x	b2_q237	str3 %3s	D	oes the enterprise undertake any work on contract basis (yes-1, no-2)	 Variables 	T 0
	b2_q244	str3 %3s	6	ces the enterprise have access to toilet facility?	A Line small	
The commands nere	NSS	double %10.0g	n/	umber of first stage units surveyed within a substratum for the sub-sample	The Parage	
Command _rc	NSC	double \$10.0g	2/	under of first stage units surveyed within a substratum for sub-sample combined	Name	Label
use 'G\practice_dataset_N_	Weight SS	double \$10.0g		inal weight for sub-sample wise estimates	State	State
list Sample State_Region in	Weight_SC	double \$10.0g	r	inal weight for sub-sample combined estimates	State_District	State-District
list Sample State_Region in					Stratum	Stratum
d	Sorted by:				SubStratumNo	Sub-Stratum-no
d State	4 19 19				SubRound	Sub-Round
92	. d state				SubSample	Sub-Sample
	5	torage display	value		FODSubRegi	FOD-Sub-Region
	variable name	type format	label v	ariable label	HG_Sub8ikNo	Hamlet group/Su
					. Second_Strat.	Second-Stage-Str
	State	str6 %65	\$	tate	ent_no	Sample enterprise
					Level	Level 2-block 20
	. 50				b2_q201	Whether pursuing
	Variable	Obs	Mean Std. D	ev. Min Max	b2_q202	Description of Maj
					Properties	ů.
	ENTID	0				
	CentreCode-d	0			Variables	
	Vill_Bik_S-0	ŝ			Name	State
	Schedule	ě			Label	State
		•			Type	str6
	Sample	0			Format	5.63
	Sector	٥			Value label	
	State_Region	0			Notes	
	State	•			4 Data	
	state_pist+t	•			: Filename	practice_dataset,
	Stratum	0			Label	
	SubStratumNo	0			 Notes 	
					Variables	42
	c				Observations	72.528
	Command				Size	18.54M
					Memory	64M
					Sorted by	
Stata15						CAP NUM 0

So, this is the window we have already shown. Similar to this kind of window, we have shown you, but this is a little different, because we are going to operate with the summarize. Earlier one was describe. This kind of window we have shown you. I am going to talk about summarize.

So, summarize gives you descriptive statistics and this is going to be very very useful for all the researchers while writing papers and communicating to journals, even for books or even for assignment. If you do not give a descriptive statistic of all your variables you are using for modeling, I think your modeling is completely redundant. How another author is going to understand your variable descriptions?

This carries with different name, variable descriptions, descriptive statistics. And usually this gives information like how many observations are there, the mean, standard deviation, minimum and maximum. So far as standard command that is summarize you are going to enter. This is called a standard command for summarize your descriptive statistics. There may be a detailing of the summary descriptions as well.

Sometimes some researcher in the paper write a summary. So in a table, in the heading they simply write summary. So summary also give you some kind of description as highlighted here summarize, you can write summarize or only su. Underscore su is mentioned or underlined with su, so with the variable list we will give you the detailed information about it.

Now, let me guide you first before going to the detailing of it. If I just type su or summarize, sum also, usually in my understanding many researchers usually enter sum. So let me exercise with su. If I simply start with su, it gives all the variables and their summary. How to read those, just have a look, how to read these? Look at this. As I told you, it gives information like observation, mean, standard deviation, mean maximum.

Another important thing is to be noted here, why observations boils down to zero because of string information. As I told you, wherever you have numeric values or labeled numeric values, you will get the result, because there are, then only we can able to enter some statistics on those data, some mathematical operations are made on the numeric data or even on the labeled numeric data, but if it is a string variable, it only represents characteristics, or some coding with a picturization of it or the time, roman characters are there, it is very difficult to do any mathematical operations.

So here the summary does not give information about the string variable, like observation is now showing zero. I will tell you when we go by the numeric value, it gives you observe, like numeric value here for NSS, NSC. As I told you already, here number of first stage, you need units surveyed within a stratum for sub-sample then number of first stage unit surveyed within a stratum for sub-sample then a stratum for the multiplier within a stratum, then weight of those indicators these are a numeric entry.

Even we can, you can check with our data browse. So, as I told you, if it is in red color, these are clearly the string data. If you browse through other variables, I am going by the right scroll, not the bottom or vertical, I am going by the horizontal scroll. Look at the black fonts, as I already guided you in the last class that this indicates numeric values. Had it been blue color, these are called labeled numeric, still operations are possible with this command. So we will clarify further details.

Let me now go to the window once again. What I wanted to say here once again to you that if you are not interested to summarize all the variables, you are interested into the particular variable, then you can simply write down the variable name. But usually when you have already filter and you have used the particular variable for your model and these variables are important, so better to include all the variables, which variables are important for you, like in your dataset so many variables are available.

If you simply go by sum or su, it gives the variable descriptive statistics, description, not variable description, descriptive statistics of entire variable available in the data. But in your model, like in a regression model you have used only selected few. Why to include the descriptive statistics of entire data you are interested for only those lists, so enter only those selected variables. So that is suggested at this moment.

If you want further details, like some data are categorical in nature, I told you from here, is numerical, I told you, some data are categorical by string, string categorical. But if string categorical, then you cannot operate it. There are certain other approaches to do it. I will show you in a short while.

But usually the standard variable or the summary statistics is not going to give you the right result, like categorical data, our result only gives mean value. The mean value of categorical data is redundant. It is not right. Mean value of, for example, gender of a person, mean value of a gender of a person is going to give you absurd result. Like gender of a person are generally coded as 1, 2, 3, so average might be boils down to 1.39. So 1.39 indicates nothing. So you require some detailed command, detailed summarize, summary statistics.

(Refer Slide Time: 34:49)



Here we are suggesting some detailing of it, detailing like su then detail. Summarize a detail. I think let me again go through that, like I have done here. I told you that su then you, key here, so here is a key I have entered then detail, so enter. What it gives? it gives other information as well. Like let me show you, since there are so many variables, I have gone through entire one, so many pages of ours unnecessarily occupied.

So what I do, I will suggest you stick to the particular variable and operate this, like su the variable you want, like 1, 2, 3, 4, whatever you just simply click on your variable tab, then you

comma detail, it gives you more information like percentiles, then it gives like skewness, kurtosis it is visible here in the window. I am just showing you variance, mean standard deviation we have already shown you in the standard command. But in addition to that It gives percentile, variance, kurtosis, skewness along with the above information as I already guided.

So, summarize without argument, please take a note I am going to read this carefully. Summarize without arguments provide summary statistics for all numeric variables, like for non-numeric value variables zero was the result. For string variable, it shows no observation with their variable labels.

(Refer Slide Time: 37:49)

A userur too	I for identifying variable	e types is the codebook com	mand
🗅 This comma	nd will produce basic i	nformation about every vari	able in
your data se	et, it will also produce a	a limited set of descriptive st	atistics
that are help	oful – n, range, mean, so	d.	
b2_	q204	Type of ownership code	
	type: numeric (byte)		
	range: [1,9] unique values: 8	units: missing . 012,528	
	tabulation: Freq. Value 60.350 1		
	tabulation: Freq. Value 60,350 1 8,804 2 9 3		
	tabulation: Freq. Value 60,350 1 8,804 2 9 3 1,346 4		
	tabulation: Freq. Value 60,350 1 8,804 2 9 3 1,346 4 938 5 533 6		

- 21 - H - 1	T # x TSb					1	0		Del				12	Variables	
		9 9	v e r	1	y u		ο ρ	1111	1 Der					A Liter variable	
ornmand	Caps	a	s d	f	a h	i k	TIT							Name	Label
e 'G\practice_dataset_N.	1977 - 19				9			2 1						State	State
Sample State_Region in	Shift	Z	x c	v	b n	m	< >	? 1	Shift					State_District	State-Distric
Sample State_Region in	· 📰													Stratum	Stratum
1110	Ctrl	🛿 Alt				Alt 🗟	Ctrl	* F	* fn					SubStratumNo	Sub-Stratum-I
														SubSample	Sub-Sample
Setail	100				Marca									FOOSubRegi	FOD-Sub-Reg
debook State	204	125.165	Larges	t	Mean Std. Dev.	43	76.872							HG_Sub81kNo	Hamlet group
	758	353.025	94383.2	6										Second_Strat	Second-Stage
	908	913.26	105757	7	Skewness	2	486525							ent_no Level	Al evel 2-bio
	998	5027.15	115986	8	Kurtosis	14	97.748							b2_q201	Whether purs
	code	book State												b2_q202	Description of
														Properties	
	State													 Variables 	
														Name	State
			type: strin	g (str	6), but lon	gest is	str2								
		unique va	lues: 36			missi	ng **: 0/7	2,528						Format	5.63
														Value label	
		6.42	*18*											Notes	
			-23-											Data	
			-29-										8	Label	
													•	Notes	
														Variables	42
	Comma	nd											4		72,528 18 Satu
														Memory	64M
														Sorted by	
Itor (Browse) - (practice Srt View Data 1	e dataset, NSS73, bior Tocos	🧿 []												- 2041	43 8/10
tor (Browse) - (praction dr. View Data 1 @ D 2 2 2 ENTID(1)	e dataset, NSS73, bios Tools Tools	()												- 1041	431 8/10
tor (Browst) - (practice N View Data 1 ENTID(1) ENTID	e dataset NSS73 bloc Tools T. Basson Cean	9 1007 1507 1911	1.51X,5120 B	iound '	Schedule Sr	ample 5	ector State	Region State	State_Diat-t :	Stretum - Subb	itretundio Sul	blound SubJam	ple P(*	 Variables 	431 8/10
tor (browst) - (practice N View Data 1 ENTID ENTID	dataset NSS73 bio dots fors for		1_818_5100 R	iound '	Schedule S4 234	Ample 5	ector State	Region State 091 01	State_Diat-t :	Steatum Subb	itratumlis Sull 02	DRound SubJam 4	ple R*	 N ⊕ ↓ Variables N Fiter variable 	431 A/Loy - 0
tor (Browse) - (practice bit View Data (NTID)] ENTID	dataset NSS/2 bio c dataset NSS/2 bio Tools 695091 cas cas		1_B1¥_51200 PA (5559 23399	lound 1 73 73	Schedule St 234 234	ample 5	ector State	Region State 091 04 033 01	State_Diat-t 0 0902 0 0322	Stretum Sub O2 O2	itretumlio Sul OZ Ol	Blound JubJan 4 1	ple Pr#	 Name 	shee Label
tor (Browse) - (practice it View Data ENTID(1) ENTID			1,912,3260 F 6555 33335 2116 6773	iound 1 73 73 73 73	Schedule Sr 234 234 234 234	Ample 5 1 1	ector State 1 2 2	Region State 091 04 031 0 212 2 245 2	State_DLat-t 0 0902 0 0322 1 2129 1 2432	Stratum Suit 02 02 29 12	1 1781-0016 Jul 02 01 01 01	BRound JubJan 4 3 3	Die N.4	 Name Name Name 	es here Label Primary key
tor Recoved - practice it View Data 1 ENTID(1) ENTID	dnuet NS73, poo toos		1,838,5320 B 69509 5335 53165 67753 6785	iound ' 73 73 73 73 73 73	Schedule 54 234 234 234 234 234 234	ample 5 1 1 1	ector State 1 2 1 1	Region State 091 01 001 0: 212 2: 245 2: 241 2:	State_Dist-t 9 0992 9 0352 1 2129 1 2422 9 2423	5tratum - Sub 02 02 12 21	tratumbo Sum 02 01 01 01 01 01	Blound JubJan 4 3 3 4	Die P(* 1 2 2 2 1	Yariables Variables	es here Label Querter codes
tor (Brown) - Ipactor It Vew Data 1 O 2 2 3 ENTO(1) ENTO	dataset NSS7) Boo dataset NSS7) Boo dots dataset NSS7) dots	82,000 ************************************	1.514,3300 B 60509 53165 67753 67754	iound 73 73 73 73 73 73	Schedule 31 234 234 234 234 234 234 234	ample 5 1 1 1 1	ector State 1 2 1 1 2 2 1 2	Argion State 001 001 021 01 242 21 241 21 272 27	3teste Diset-t 0402 0302 2129 2412 2411 2225	Stratum Sub 02 02 12 21 29	tratumble Sum 02 01 01 01 01 02	DRound DubJam 4 3 4 1 1	Die P(*	Variables Variables Virtual Variables	es here Label FSU Senal No
ror (Brower) - granting A Yew Data Exhto(1) Exhto(1) Exhto(1)	defauer (NST) door toos to	Image: Non-State Vill 1507 1507 1507 730 730 737 733 738 734 738 735 738 736 738 737 733	1.514,3300 B 60505 53165 67753 67754 53155	iound 5 73 73 73 73 73 73 73 73 73 73	Schedule 34 234 234 234 234 234 234 234 234	ample 5 1 1 1 1 1 1	ector State 1 2 1 1 2 2 2 2 2 2 2 2 2	Argion State 001 001 021 01 242 24 242 24 272 272 091 01	State_Dist-1 0002 2103 2212 2212 2212 2212 2212 221	Stratum Subl 02 02 29 12 21 29 21 29 72	TERCUNNIG Jul 02 01 01 01 02 02 13	bhund Dublam 4 1 3 4 1 3 4 1 3	Die D	Variables Variables Inter-smath Variables Vari	es here Label SDu Senia No. SDu Senia No. Round
tor (Browd) - Spanfac Int Vew Data Vew Data Evilip(1) Evilip(1)	dataset 145573 pilos dataset 145573 pilos dos dos v		1.832, 5320 B 65559 13339 67659 67659 67659 67659 67659 67659 67659 67659 67659 67659 67659 67659 67659	lound 73 73 73 73 73 73 73 73 73 73 73 73 73 7	Sabebile 3 234 234 234 234 234 234 234 234 234 23	ample 5 1 1 1 1 1 1 1 1 1	ector State 1 2 1 1 2 2 1 1 2 2 1	Argion State 061 01 021 01 212 212 245 21 245 21 272 21 265 21 275 21 265 21 275 21 275 21 261 11 151 11	36446_DIA4-4 0402 20032 2012 2012 2012 2012 2012 2012 2	Stretum Subl 02 02 29 12 21 29 72 06	Tretunió Jul 02 01 01 01 02 13 02	bhund Dublan 4 1 3 4 1 3 4 4	Die 1 1 1 2 2 2 1 1 1 2 1 1 2 1 1 1 2 1 1 1 2 1	Variables Variables If the variable If the variable If the variable Variables	es here Label Gestre code, Schoud Schould Schoule Sample
tor (Rowell - Ipractice in Vew Data 1 e Nillo(1) e Millo(1)			L. B1K, 52200 F. 64555 64555 67653 67653 67653 67653 67653 67653 67653 67653 67653	iound 53	Sobedule 34 234 234 234 234 234 234 234 234 234 2	NADIE 5	ector State 1 2 1 1 2 2 1 2 2 1 2 2 1 2 2 2	Jegion State 081 01 031 01 212 21 245 21 245 21 261 21 151 11 061 01	State Diart-1 0322 0332 2138 2128 2431 2235 0339 1344 1344 1345	Stratum Juli 02 29 12 21 33 72 06 02 23	TERCUMUNE Sul 02 01 01 01 02 13 02 02 02 03	bhund Dublan 4 3 3 4 1 3 4 4 4 4	DL+ P(#)	Variables Variables (Tetr sprach Variables) (Tetr spra	s here Label Schelle Schelle Schelle Sector
tor (Bowed - pactor it Vew Data C 2 2 NITO(1) ENTO		(()	L DI 2300 P. 0 69509 53166 67785 53165 67785 53155 53155 53155 53155 53155 53155	event 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	50be011e 50 234 234 234 234 234 234 234 234 234 234	smple 5 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	eccor State 1 2 1 1 2 2 1 1 2 2 2 2 2 2 2 2 2 2 2	Region Diskie 091 01 031 01 032 21 245 21 241 21 241 01 041 01 043 01 044 01	Jease Director 0 0402 2 2323 2 2411 2 2431 2 2431 0 0495 0 0495 0 0495 0 0495 0 0495 0 0495 0 0495	Stratum Subl 02 29 12 21 39 72 06 02 57 75	tretunife 3ul 00 01 01 01 02 13 02 02 02 03 02 03 00	BRound DoubJam 4 3 3 4 1 3 4 4 4 3 3 4 3 3	DLe IV. 1 1 2 2 2 2 1 1 2 1 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 2 1 2	Variables Variables Inter-sanato Mane Mane Mane Mane Mane Second Seco	s here Label Subel SU Serial Nc Round Schedule Samje Sector
or Rowel - partice R Vew Data 1 Entities Entities Entities	distance NASS72 door tools	(100) (100) (100)	1,012,0200 P 6000 50165 6015 50165 6015 50165 6015 5015 5	event 1 73 73 73 73 73 73 73 73 73 73 73 73 73	54566048 54 234 234 234 234 234 234 234 234 234 23	καρία 5 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	ector State 1 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1	Argin State 091 01 031 01 241 21 245 21 245 21 091 01 151 11 041 01 043 01 043 01 043 01 043 02 043 03	State_Dist-14 0 0802 0 0802 0 0802 0 2121 0 2412 0 0895 0 0895 0 0895 0 0895 0 0895 0 0857 0 0857 0 0857 0 0857	Stratum Sub3 02 02 29 12 21 29 72 06 02 57 73 03	TETECUMING Sul 02 01 01 01 02 13 02 02 02 02 03 02 02 02	Blound JubJan 4 1 3 4 1 3 4 4 4 5 3 3 1	Die R.+ 1 2 2 2 2 1 1 2 2 1 2 2 1 1 2 2 1 1 2 2 1 1 2 2 1 1 2 2 1 1 2 2 2 1 1 1 2 2 2 2 1 1 1 1 2 2 2 2 2 1	Variables Variables Toter variable Vill Dirac Sino Provid Scherolde Scherolde State Repon State State State	s here Label Autor Aber Label Abimary key Centre codel FSU Senial No Round Schedule Samje Sector & State-Rep State-Rep
or Bowel - packar a Vew Data e NID(1) entities entities	datase (053) (050) Tools	(1007 (L 112 1240 P	iound 73 73 73 73 73 73 73 73 73 73 73 73 73	500e6034 50 234 234 234 234 234 234 234 234 234 234	mmp1+ 5 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	ector State 1 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2	Argion Disate 041 01 031 01 212 21 245 21 245 21 246 21 241 21 245 21 041 11 043 04 044 04 045 21 046 11 041 04 042 04 043 04 044 13 145 14	State Dist-te 0022 0322 2124 2412 2412 2412 1014 1346 0032 0333 10333 1346	Stratum Subb 02 02 12 21 29 72 20 6 02 57 73 20 20	TERECUMING Sul 02 01 01 01 02 13 02 02 02 03 10 02 02 03 10 02 01	Blound Publish 4 1 3 4 1 3 4 4 4 5 3 3 1 2	ble P.*.	Variables Tate unado Marine	 431 441 441
evillouit Evillouit Evillouit Evillouit	dnuer (NSS7) Size dnuer (NSS7) Size dnuer (NSS7) Size dnuer (NSS7) Size dnuer (NSS7)		1,832,2356 F 6555 6555 6555 6555 6555 6555 6555	lound 73 73 73 73 73 73 73 73 73 73 73 73 73	Saturdiale 5 234 234 234 234 234 234 234 234 234 234	mple 5 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	ecter State 1 2 1 1 2 2 1 2 2 1 2 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 1 2 2 1 1 2 2 1 1 2 2 1 1 2 2 1 1 2 2 1 1 2 2 1 1 2 2 1 1 2 2 1 1 2 2 1 1 1 2 2 1 1 1 2 2 1 1 1 2 2 1 1 1 1 2 2 1 1 1 1 2 2 1	Angion State 001 01 013 01 212 21 245 21 241 21 241 21 241 21 041 01 043 01 044 21 245 21 246 21 241 21 041 01 042 01 043 01 044 10 124 11 142 12	Jase Dist- 0102 0102 0202 0202 2412 2412 2423 2423 0495 0455 0452 0453 0453 0453 0453 0453 0453 1455	5teatum Subb 02 29 12 39 72 06 57 75 57 75 03 20 08	tratuatio gui 02 01 01 03 03 03 03 03 03 03 03 03 03 03 03 03	Blound Publism 4 1 3 3 4 1 3 4 4 3 3 3 3 3 2 2 2	ble R*	Variables Variables Inter-construction Inter-constr	 431 431
ex (Browd practice 1 Vew Data 1 E NTD(1) E PTTD	Adducer (KGS7) Data Adducer (KGS7) Data Tools		L 2014 2000 F	event 73 73 73 73 73 73 73 73 73 73 73 73 73	50000010 3 234 234 234 234 234 234 234 234 234 23	mple 5 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	ector State 1 2 2 1 1 2 2 2 2 1 2 2 1 2 2 1 2 1 1 2 1 1 1 1 1	Region Diskle 081 01 032 01 245 22 245 22 051 01 041 01 043 01 044 01 045 01 046 01 047 01 048 01 049 01 041 01 042 01 134 11 142 12 343 24	Base Director 0302 0302 2022 2123 2022 2021 2022 2021 0405 0405 0405 0405 0405 0405 0405 0405 0405 0405 1566 1466 1566 1466 1566 1466	Stantum Sub0 02 02 29 21 21 21 21 21 21 20 6 6 02 57 72 06 02 57 75 03 20 08 11	tretundio 94 02 01 01 01 03 03 03 03 03 03 03 03 03 03 03 03 03	Rhund Dublam 4 1 3 4 1 3 4 4 5 3 3 1 2 2 2 2 1 2	ble R*	Variables I fate variable Name WithD Via (historecode), With (historecode), With (historecode), State (4313 4315 43 4315 43
	Additional Possibility	2000 2000	L 112 1244 P 4000 P 4000 P 114 124 P 4000 P 114 124 P 114 P 11	event 73 73 73 73 73 73 73 73 73 73 73 73 73	545409344 234 234 234 234 234 234 234 234 234 234	Magle 5	ector State 1 2 1 1 2 2 2 2 2 2 2 1 2 2 1 1 2 1 1 2 1 1 2 1 1 2 2 1 2 1 2 2 1 2 2 1 2 2 1 2 2 2 1 2	Argin State 091 01 091 01 201 21 241 21 243 21 241 11 041 01 042 01 153 12 143 12 143 12 143 12 143 12 143 12 144 12 145 12 144 12 145 12 1461 12	Jean 0402 0122 0122 0123 0122 2412 2412 2412 2412 1566 0402 0402 0403 1566 0402 0415 1566 1566 0402 1566 0402 1566 0402 1566 0402 1566 0402 1566 0402 1566 0402 1566 0402 1566 0402 1566 0402 1566 0402 1566 0402 1566 0402 1566 0402 1566 0402 1566 0402 1567 0402 1568 0402 1568 0402	Proteina data 03 03 12 13 13 13 13 13 13 13 13 13 13 13 13 13	tretado 3 10 2 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 0 1 0 0 1 0 0 1 0	08/und 2-ub344 4 3 3 4 1 3 4 4 4 4 3 3 1 2 2 2 2 2 1 2 2 2 4		Variables Variables Variables Variables Variables Variables Variables Variables Sample Sample Samples	A 433 A 470 A 470
erto Brower: packato Vere Data O C C C C Entro(1) Entro	Construct (NSU) Joo Tools Construct (NSU) Joo Construct	2000 2000 2000 2000 200	La 18 12, 2020 K 4000 K 1000 K 100	event 73 73 73 73 73 73 73 73 73 73 73 73 73	Satestile 3 234 234 234 234 234 234 234 234 234 23	Magle	ecter State 1 2 1 2 1 2 2 2 2 2 2 1 2 2 2 1 1 1 1	Jegion State 051 051 053 07 245 27 245 27 045 07 045 07 045 07 045 07 045 07 047 07 142 07 143 07 144 07 041 07 042 07 043 07 044 07 045 07	Jease Diarc-t 0192 0322 022 0322 2412 2412 2421 2412 0395 0452 0442 0451 0332 1233 1445 1446 1544 0451 1544 0451	Deretam Pak 02 02 23 23 23 23 23 23 23 23 23 23 23 23 23	tretuelle 32 31 31 31 32 32 33 33 33 33 33 33 33 33 33 33 33	Blound Jublan 4 1 3 4 1 3 4 4 3 4 4 5 3 3 3 2 2 2 4 4 3 3 3 3 2 2 4 3 3 3 3 3		Variables Variables Inter-sensor Variables Variable	2 No. 4310 Artiki Ar
tor diversit - portion it Verw Data (NID)11 DUTD	Additect NXX73 along Aditect NXX73 along Aditect NXX73 along Aditect NXX73 alon	Image: Second	1, 212, 33,000 F. 60503 51505 51505 64515 51515 64515 51515 64515 51515 64515 51515 51515 64515 51555 51555 51555 51555 51555 51555 51555 51555 51555 51555 51555 51555 51555 51555 51555 51555 51555 515555 515555 515555 5155555 5155555 5155555 5155555 5155555555	event 73 73 73 73 73 73 73 73 73 73 73 73 73	3000034 30 134 234 234 234 234 234 234 234 234 234 2	mmple 5 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	ector Society 1 2 1 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2	Region Factor 051 01 033 01 210 21 245 22 041 01 153 01 043 01 044 01 045 01 046 01 047 01 149 11 142 12 043 01 154 01 154 01 154 12 155 14 041 01 041 11	Rest Dist-14 0432 0432 2123 2123 2123 2121 2121 2121 2122 2121 2123 2121 2124 2125 2125 2125 2126 2125 2127 2121 2128 2121 2129 2121 2121	Stratum Sold 62 62 63 62 73 64 64 62 63 62 64 62 64 64 64 64 69 64	tretails 54 61 61 61 61 61 61 62 62 63 62 63 62 64 63 63 63 63 63 63 63 63 63 63 63 63 63	about Soltan 4 3 3 4 4 3 3 4 4 3 3 3 3 3 3 3 3 3 3	Die D	Variables Variables Variables Variables Virtual Virtual Virtual Virtual Virtual Virtual Virtual Virtual State Post State District State District State State State State State State State State State State Virtual State Stat	2 In All All All All All All All All All Al
In the second se	Comment (Loss) (Loss) Comment (Loss) (Loss) Comment (Loss) (Loss) Comment (Loss) Commen	0 100 100 101	L 101 21/0 L 10 4 010 21 101	event 23 23 23 23 23 23 23 23 23 23 23 23 23	5000000 10 2134 2134 2134 2134 2134 2134 2134 2134	mmp1+ 5 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	eccer 20000 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 1 2 2 2 1 2 2 2 1 2 2 2 2 1 2	Pequation Picate 001 01 001 01 001 01 001 01 011 01 021 01 021 01 021 01 021 01 021 02 023 02 033 02 034 01 042 02 043 02 044 02 045 04 046 04 041 04 044 04 044 04 044 04 045 04	Jease Bart-1 0002 0002 2129 2129 2129 2124 0002 2129 2129 2124 0002 2129 2129 2124 0003 2126 0013 0015 0013 1016 1016	Steria Sol 02 02 02 02 02 02 03 02 04 03 05 04 07 03 08 04 09 04 01 04 02 04 03 04 04 04 05 04 06 04 07 04 08 04 09 04 09 04	trateadile 3 al 02 3 01 3 02 3 03 3 03 3 03 3 03 3 03 3 03 3 03 3 03 3 03 3 03 3 03 3 03 3 03 3 04 3 05 3 06 3 07 3	alices factors 4 3 3 4 3 4 4 3 4 4 3 3 4 4 3 3 1 2 2 2 1 2 4 3 4 4 3 4 4 3 4 4 5 3 4 4 5 7 4 5 7 7 7 7 7 7 7 7 7 7 7 7 7 7	Die P.*.	Variables Totor unado Totor unado Variables Variables Variables Variables Variables Sarap S	All And And And And And And And And And And
Conference - Specific In View Data - 1 NOTO: NOTO: DIFFE		Image: 2016	2,012,3140 P	event 23 23 23 23 23 23 23 23 23 23 23 23 23	5000000 50 134 5 234 5 2	Angle 5 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	HOF DOG 1 2 2 1 1 2 2 2 2 1 1 2 2 2 1 1 2 2 2 1 1 1 2 2 2 1 1 2 2 2 2 1 1 2	Region Pieste 081 01 081 01 212 21 245 22 241 22 253 01 153 11 084 01 083 01 154 23 084 01 154 12 154 12 154 12 154 12 154 12 154 12 155 12 154 12 155 12 154 12 154 12 154 12 154 12 154 12 154 12 154 12 154 12 154 12 154 12 154 12 155 12 156 12 <td>State Dist-1 0000 0000 2010 2010 2011 2011 2012 2011 2013 2011 2014 2011 2015 2011 2016 2011 2017 2000 2018 2011 2019 2011 2010 2011 2011 2011 2012 2011 2014 2011 2015 2011 2016 2011 2017 2011 2018 2011 2019 2011 2010 2011 2011 2011 2012 2011 2014 2014 2015 2011 2014 2014 2015 2014 2014 2014 2015 2014</td> <td>Stratum Stratum Stratum 0.2 0.2 0.2 2.3 0.2 0.2 2.1 0.2 0.2 2.3 0.4 0.2 2.4 0.2 0.2 2.5 0.1 0.2 2.6 0.1 0.1 2.6 0.1 0.1 2.6 0.1 0.1 2.7 0.1 0.1 2.8 0.1 0.1 2.9 0.1 0.1 2.9 0.1 0.1 2.9 0.1 0.1 2.9 0.1 0.1 2.9 0.1 0.1 2.9 0.1 0.1 2.9 0.1 0.1 2.9 0.1 0.1 2.9 0.1 0.1 2.9 0.1 0.1 2.9 0.1 0.1 2.9 0.1 0.1 2.9 0.1 1</td> <td>triticalis 5 4 4 4 5 4 4 5 4 4 4 4 4 4 4 4 4 4 4 4</td> <td>blood Polla 4 3 3 4 4 4 3 4 4 3 3 4 4 3 3 4 4 4 3 3 4 4 4 3 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 5 4 5 4 5 4 5</td> <td>Die Die 1 III 2 2 1 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 2 2 3 2 2 2 3 2 2 2 3 2 3 2 4 2 5 2</td> <td>Variables Variables Inter-sanato References Variables Variables Variables Sector Society State, Reveloate Society Soci</td> <td>All And And And And And And And And And And</td>	State Dist-1 0000 0000 2010 2010 2011 2011 2012 2011 2013 2011 2014 2011 2015 2011 2016 2011 2017 2000 2018 2011 2019 2011 2010 2011 2011 2011 2012 2011 2014 2011 2015 2011 2016 2011 2017 2011 2018 2011 2019 2011 2010 2011 2011 2011 2012 2011 2014 2014 2015 2011 2014 2014 2015 2014 2014 2014 2015 2014	Stratum Stratum Stratum 0.2 0.2 0.2 2.3 0.2 0.2 2.1 0.2 0.2 2.3 0.4 0.2 2.4 0.2 0.2 2.5 0.1 0.2 2.6 0.1 0.1 2.6 0.1 0.1 2.6 0.1 0.1 2.7 0.1 0.1 2.8 0.1 0.1 2.9 0.1 0.1 2.9 0.1 0.1 2.9 0.1 0.1 2.9 0.1 0.1 2.9 0.1 0.1 2.9 0.1 0.1 2.9 0.1 0.1 2.9 0.1 0.1 2.9 0.1 0.1 2.9 0.1 0.1 2.9 0.1 0.1 2.9 0.1 0.1 2.9 0.1 1	triticalis 5 4 4 4 5 4 4 5 4 4 4 4 4 4 4 4 4 4 4 4	blood Polla 4 3 3 4 4 4 3 4 4 3 3 4 4 3 3 4 4 4 3 3 4 4 4 3 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 5 4 5 4 5 4 5	Die Die 1 III 2 2 1 2 1 2 2 2 1 2 2 2 1 2 2 2 1 2 2 2 2 2 3 2 2 2 3 2 2 2 3 2 3 2 4 2 5 2	Variables Variables Inter-sanato References Variables Variables Variables Sector Society State, Reveloate Society Soci	All And
to theme is particular in Vew Data 1 in Vew Data 1 in Vew Data 1 in Unit EPTID		0 100 100 100		abound 1 1 1 1 1 1 1 1 1 1 1 1 1 1	20000310 10 10 10 10 10 10 10 10 10 10 10 10 10 1	mmple 5 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	Netter Dester 1 2 2 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2	Matter Last 401 00 401 0 401 0 401 0 402 0 403 0 404 0 405 0 404	2444 0144-4 042 042 242 242 242 242 242 242 242 24	Otom Data 02 02 23 02 24 02 25 02 26 02 27 03 04 02 05 03 04 04 05 04 04 04 05 04 07 03 08 04 09 04 01 04 02 04 03 04 04 05 05 04 07 05	201 201 201 201 201 201 201 201 201 201	blood Indian 4 1 3 4 4 3 3 4 4 3 3 3 1 2 2 4 4 4 4 4 4 4 5 3 3 1 1 2 2 4 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3		Variables Variables Variables Vitil Bite vandatio Vitil Bite vandati	In All All All All All All All All All Al
Conformation in State	Control (Control (Contro) (Control (Contro) (Contro) (Contro) (Contro) (Contro) (Contro)	Image: Section 2016		ivent 1 3 3 3 3 3 3 3 3 3 3 3 3 3	Jonebie J 134 - 134 - 134 - 134 - 134 - 134 - 134 - 134 - 134 - 134 - 134 - 134 - 134 - 134 - 134 - 134 - 134 - 135 - 136 - 137 - 138 - 139 - 131 - 132 - 134 -	ample 5 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	HEHE - 2006 1 2 2 1 1 2 2 2 2 2 2 1 1 1 2 2 2 2 2	Mattern Justo Bit 0 Bit 0 Total 0	Rest Dist 0.002 0.002	Statum 2nd G2 22 23 23 23 23 23 23 23 23 23 23 23 23	tretuils 5 4 2 3 4 3 4 3 4 3 4 4 3 4 4 3 4 4 4 4 4	about 2016 4 3 3 4 4 4 3 3 4 4 3 3 3 3 3 3 3 3 3		Vandales Vandal	All All Andrew All Andrew Andre
Tor Howes: - Spatial If Vew Data 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0		Image: Control of the second		Apund 1 1 1 1 1 1 1 1 1 1 1 1 1	30000000 9 30 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 1	Migle 5 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	eccar - Source - 	Jugan Jugan 66 00 712 21 712 21 714 21 714 21 714 21 714 21 714 21 714 21 714 21 714 21 714 21 714 21 714 21 714 21 714 21 714 21 715 21 716 21 717 21 718 21 719 21 710 21 711 21 712 21 713 21 714 21 715 21 714 21 715 21 714 21 715 21 714 21	J444 J147 0122 J128 J124 J127 J127 J127 J128 J128 J129 J128 J129 J128 J129 J128 J126 J128 J127 J128 J126 J128 J127 J128 J127 J128 J127 J128 J128 J128 J128 <td>State Dadi 02 02 03 02 12 12 13 02 04 02 05 02 04 02 05 02 04 02 05 02 06 02 07 03 08 02 09 02 01 02 02 03 03 03 04 04 05 04 06 04 07 04 08 05 09 04 01 04 02 04 03 04 04 05 05 04 05 04 05 04 05 04 05 04 05 04 05</td> <td>Freedom Fall 42 42 43 43 43 43 43 43 43 43 43 43 43 43 43 43 43 43 44 43 43 43 44 43 43 43 44 43 45 43 46 43 47 43 48 43 49 44 41 43 42 44 43 44 44 45 45 45 46 45 47 45 48 45 49 45 41 45 42 45</td> <td>Blood Judse 4 3 3 4 3 4 4 3 3 4 4 3 3 3 4 4 3 3 4 4 3 3 4 4 4 3 3 4 4 3 3 4 4 3 3 4 4 3 3 4 4 3 3 4 4 4 3 3 4 4 4 4 4 4 5 3 3 3 4 4 4 4</td> <td>D.4 D.4 1 III 2 2 1 III 2 2 1 III 2 III 2 III 2 IIII 2 IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII</td> <td>Vaniales Van</td> <td> A 333 A 100 A 100</td>	State Dadi 02 02 03 02 12 12 13 02 04 02 05 02 04 02 05 02 04 02 05 02 06 02 07 03 08 02 09 02 01 02 02 03 03 03 04 04 05 04 06 04 07 04 08 05 09 04 01 04 02 04 03 04 04 05 05 04 05 04 05 04 05 04 05 04 05 04 05	Freedom Fall 42 42 43 43 43 43 43 43 43 43 43 43 43 43 43 43 43 43 44 43 43 43 44 43 43 43 44 43 45 43 46 43 47 43 48 43 49 44 41 43 42 44 43 44 44 45 45 45 46 45 47 45 48 45 49 45 41 45 42 45	Blood Judse 4 3 3 4 3 4 4 3 3 4 4 3 3 3 4 4 3 3 4 4 3 3 4 4 4 3 3 4 4 3 3 4 4 3 3 4 4 3 3 4 4 3 3 4 4 4 3 3 4 4 4 4 4 4 5 3 3 3 4 4 4 4	D.4 D.4 1 III 2 2 1 III 2 2 1 III 2 III 2 III 2 IIII 2 IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	Vaniales Van	 A 333 A 100 A 100
Kor Noveri S pratori Kor Noveri S pratori Kor Die Verw Data 1 Kor		Image: Second		event 73 73 73 73 73 73 73 73 73 73 73 73 73	500000 0 10 210 1 213 1 213 1 213 1 213 1 213 1 214 1	megle 5 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	HILF 2006 2 2 3 3 2 3 2 3 3 2 3 3 3 3 3 3 3 3 3	Jugim Jost (1) (1) (2) (2) (2) (2) (2) (2) (3) (2) (4) <td>Date, Rote, J. 002 002 012 012 014 012</td> <td>Statum Solution 02 02 03 02 04 02 05 02 04 02 05 02 04 02 05 02 04 02 05 02 06 02 07 03 08 07 07 02 08 07 07 02 08 07 09 07 01 07 02 07 03 07 04 07 05 07 07 03 08 07 09 04 01 05 02 05 03 05 04 05</td> <td>tretonio 34 42 42 43 42 43 42 43 42 43 42 43 42 43 42 43 42 43 42 43 42 43 42 43 42 43 42 43 42 43 42 43 42 43 42 43 42 43 42 44 42 45 42</td> <td>ahood Jodies 4 3 3 4 4 3 3 4 4 3 3 3 3 3 3 3 3 3 3</td> <td>Le P.A. 1 1 2 2 2 1 2 2 1 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2</td> <td>Vanades Vanades Marce Ma</td> <td> In All All All All All All All All All Al</td>	Date, Rote, J. 002 002 012 012 014 012	Statum Solution 02 02 03 02 04 02 05 02 04 02 05 02 04 02 05 02 04 02 05 02 06 02 07 03 08 07 07 02 08 07 07 02 08 07 09 07 01 07 02 07 03 07 04 07 05 07 07 03 08 07 09 04 01 05 02 05 03 05 04 05	tretonio 34 42 42 43 42 43 42 43 42 43 42 43 42 43 42 43 42 43 42 43 42 43 42 43 42 43 42 43 42 43 42 43 42 43 42 43 42 43 42 44 42 45 42	ahood Jodies 4 3 3 4 4 3 3 4 4 3 3 3 3 3 3 3 3 3 3	Le P.A. 1 1 2 2 2 1 2 2 1 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2	Vanades Vanades Marce Ma	 In All All All All All All All All All Al
tor (forward) = grantized Int Verw Data 1 PATIOL BATTO		(a) (b) (b)		event 73 73 73 73 73 73 73 73 73 73 73 73 73	20000.00 P	Magile 9 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	HENE Rose 1 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2	Applie Lott 401 0 401 0 402 2 402 2 403 2 404 2 404 2 405 2 406 0 401 0 403 0 404 0 405 0 406 0 406 0 406 0 401 0 402 0 403 0 404 0 405 0 404 0 405 0 406 0 407 0 408 0 409 0 400 0 401 0 402 0 403 0 404 0 405 0 404	Juie Juie 1 002 002 002 002 002 002 002 002 002 00	Offertum Bub 02 02 02 02 02 02 02 02 03 02 04 02 05 02 06 02 07 03 08 02 09 02 01 02 02 02 03 02 04 02 05 02 06 02 07 03 08 04 09 04 01 02 02 02 03 02 04 02 05 02 06 02 07 03 08 02 09 02 01 02 02 02 03 02	Iteration JA 41 41 42 41 41 41 42 41 42 41 42 41 43 41 43 41 43 41 43 41 43 41 43 41 43 41 43 41 44 41 43 41 44 41 45 41 46 41 47 41 48 41 49 41 41 41 42 41 43 41 44 41 43 41 44 41 45 41 46 41 47 41 48 41 49 41 41	shout future 4 3 3 4 4 3 3 4 4 4 4 4 4 4 4 4 4 4 4		Variadas Variad	en here En here En here En here En here En here En here En here En here Solet Sole
TO KNOWE - PARLER II VAW DAS II VAW DAS		Image: Control of the second		event 1 1 1 1 1 1 1 1 1 1 1 1 1	300000 10 13 13 13 13 13 13 13 13 13 13 13 13 13	mmple 5 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1000 9000 1 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2	Impair Josef 61 0 61 0 61 0 70 0 <t< td=""><td>Faire Autor 002 002 002 002 002 002 002 002 002 002 002 002 002 002 002 002 002 002 002 002 003 002 004 002 004 002 004 002 004 002 004 002 004 002 004 002 004 002 004 002 004 002 004 002 004 002 004 002 004 002 004 002</td><td>Statum 986 02 03 03 03 03 03 03 03 03 03 03 03 03 03</td><td>treation pl 22 - 43 - 41 - 42 - 43 - 43 - 43 - 43 - 43 - 43 - 43 - 43 - 43 - 43 - 43 - 43 - 44 - 45 - 46 - 47 - 48 - 49 - 41 - 42 - 43 - 43 - 44 - 45 - 46 - 47 - 48 - 49 - 41 - 42 - <tr< td=""><td>Alone 7.0156 4 3 3 4 4 3 3 4 4 3 3 3 4 4 4 4 4 4 5 3 3 4 4 4 4</td><td></td><td>Variables Variables</td><td>In Park State Set</td></tr<></td></t<>	Faire Autor 002 002 002 002 002 002 002 002 002 002 002 002 002 002 002 002 002 002 002 002 003 002 004 002 004 002 004 002 004 002 004 002 004 002 004 002 004 002 004 002 004 002 004 002 004 002 004 002 004 002 004 002	Statum 986 02 03 03 03 03 03 03 03 03 03 03 03 03 03	treation pl 22 - 43 - 41 - 42 - 43 - 43 - 43 - 43 - 43 - 43 - 43 - 43 - 43 - 43 - 43 - 43 - 44 - 45 - 46 - 47 - 48 - 49 - 41 - 42 - 43 - 43 - 44 - 45 - 46 - 47 - 48 - 49 - 41 - 42 - <tr< td=""><td>Alone 7.0156 4 3 3 4 4 3 3 4 4 3 3 3 4 4 4 4 4 4 5 3 3 4 4 4 4</td><td></td><td>Variables Variables</td><td>In Park State Set</td></tr<>	Alone 7.0156 4 3 3 4 4 3 3 4 4 3 3 3 4 4 4 4 4 4 5 3 3 4 4 4 4		Variables Variables	In Park State Set
Conformation in State				event 1 1 1 1 1 1 1 1 1 1 1 1 1	30000.00 9 303 303 304 304 304 304 304 304 304 304	mg2+ 5 1 1 1 1 1 1 1 1 1 1 1 1 1	Hear Date 1 2 2 1 1 2 2 2 2 2 1 1 2 2 2 2 1 1 2 2 2 2 1 1 2 2 2 2 2 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2	Japan Jaar 101 0 101 0 101 0 101 0 102 0 103 0 104 0 105 0 106 0 107 0 108 0 109 0 101 0 102 0 103 0 104 0 105 0 106 0 107 0 108 0 109 0 101 0 102 0 103 0 104 0 105 0 106 0 107 0 108 0 109 0	Rote Arter 002 002 010 002	Pretain 945 02 03 13 13 13 14 04 04 04 04 04 04 04 04 04 04 04 04 04	tretaule Ja 21 2 41 2 42 2 43 2 43 2 43 2 43 2 43 2 43 2 44 2 43 2 44 2 43 2 44 2 45 2 46 3 47 3 48 3 49 3 41 3 42 3 43 3 43 3 43 3 44 3 45 3 45 3 46 3 47 3 48 3 49 3 41 3 42 4	Advad Paldad 4 3 3 4 3 4 4 3 4 4 3 3 4 4 4 4 4 4 4		Vanaldes Van	c) Park 100 Control
tor (Rowell - Spatial If Vew Data 1 Control - Spatial Control - Spa	Control (Control (Contro) (Contro) (Contro) (Contro) (Contro) (Contro) (Contro) (Contro)	Image: second		event 73 73 73 73 73 73 73 73 73 73	Standardar Ja 133 - 134 - 135 - 136 - 137 - 138 - 139 - 130 - 131 - 132 - 134 - 135 - 136 - 137 - 138 - 139 - 134 - 135 - 136 - 137 - 138 - 139 - 131 - 132 - 134 - 135 - 134 - 135 - 136 - 137 - 138 - 139 - 1314 - 1315 <td>Migle 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3</td> <td>1000 1000 1 2 2 1 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2</td> <td>Perior Perior Perior 511 0 0 511 0 0 512 0 0 743 0 0 744 0 0 747 2 2 747 2 2 743 0 0 744 0 0 745 0 0 741 0 0 743 0 0 744 0 0 745 0 0 744 0 0 745 0 0 744 0 0 744 0 0 744 0 0 744 0 0 745 0 0 744 0 0 744 0 0 744 0 0 744 0 0</td> <td>Parcy Jure 4 002 002 002 002 002 002 002 00</td> <td>Stretus Data 02 02 03 02 04 02 05 02 04 02 05 02 06 02 07 03 08 02 09 02 01 03 02 03 03 03 04 03 05 04 06 03 07 04 08 04 09 04 01 04 02 04 03 04 04 04 05 04 05 04 05 04 05 04 05 04 05 04</td> <td>tretadit Juli 12 - 13 - 14 - 13 - 14 - 15 - 16 - 17 - 18 - 19 - 10 - 11 - 12 - 13 - 14 - 15 - 16 - 17 - 18 - 19 - 10 - 11 - 12 - 13 - 14 - 15 - 16 - 17 - 18 - 19 - 10 - 11 - 12 - <tr td=""> - 1</tr></td> <td>alon Abland 4 3 3 4 4 3 3 4 4 4 3 3 3 3 3 4 4 4 3 3 3 4 4 4 3 3 4 4 4 3 3 4 4 4 3 3 4 4 4 4 3 3 3 3 4 4 4 4 5 4 5</td> <td>Die P.+ 1 2 2 2 2 3 3 4 2 4 2 4 2 4 2 4 2 4 2 4 2 4 2</td> <td>Vanadas Vanadas Contro construinte Vanadas Vanadas Vanadas Vanadas Vanadas San</td> <td>Image: Provide and Provide and</td>	Migle 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3	1000 1000 1 2 2 1 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2	Perior Perior Perior 511 0 0 511 0 0 512 0 0 743 0 0 744 0 0 747 2 2 747 2 2 743 0 0 744 0 0 745 0 0 741 0 0 743 0 0 744 0 0 745 0 0 744 0 0 745 0 0 744 0 0 744 0 0 744 0 0 744 0 0 745 0 0 744 0 0 744 0 0 744 0 0 744 0 0	Parcy Jure 4 002 002 002 002 002 002 002 00	Stretus Data 02 02 03 02 04 02 05 02 04 02 05 02 06 02 07 03 08 02 09 02 01 03 02 03 03 03 04 03 05 04 06 03 07 04 08 04 09 04 01 04 02 04 03 04 04 04 05 04 05 04 05 04 05 04 05 04 05 04	tretadit Juli 12 - 13 - 14 - 13 - 14 - 15 - 16 - 17 - 18 - 19 - 10 - 11 - 12 - 13 - 14 - 15 - 16 - 17 - 18 - 19 - 10 - 11 - 12 - 13 - 14 - 15 - 16 - 17 - 18 - 19 - 10 - 11 - 12 - <tr td=""> - 1</tr>	alon Abland 4 3 3 4 4 3 3 4 4 4 3 3 3 3 3 4 4 4 3 3 3 4 4 4 3 3 4 4 4 3 3 4 4 4 3 3 4 4 4 4 3 3 3 3 4 4 4 4 5 4 5	Die P.+ 1 2 2 2 2 3 3 4 2 4 2 4 2 4 2 4 2 4 2 4 2 4 2	Vanadas Vanadas Contro construinte Vanadas Vanadas Vanadas Vanadas Vanadas San	Image: Provide and

So, let me proceed further. let us understand codebook. So far we discussed browse, edit, then listing, then we discussed summarize. We are trying to guide you, codebook. How codebook, these commands are most often used by the users. We usually use all those commands simultaneously in our everyday life for working with the data or while working with the data.

So, codebook, let me guide you, is a useful tool for identifying variable types and is the codebook for identifying variables with a codebook command. This command will produce basic information about every variable in your dataset. It also produced a limited set of descriptive statistics that are helpful, like n number of observation, range, mean, standard deviation.

Look at the screen that is visible here. This is different than that of the summary statistics we have. These keeps like starting with which kind of data it is, then its range maximum, minimum value, then the unique values, then the frequency of the particular variable and their individual value with label is clearly given. Sometimes those labeling are not clearly written, only in value they have written. So labeling should have been mentioned. Usually the researcher will label, I will tell you.

This also give you information like missing values out of total observation you just look at carefully here, total observation information is there, how many are missing? Zero is missing. And this is important, if missing is there, instead of zero you have got some information like out of 72,528 samples, you are near about 50% is missing, around 30,000 or 40,000 is missing. So it must be noted as a serious issue.

And you must cross check your data or if it is like 2% or 1% is missing and this number is given like out of this, let it be 7 or 10 out of this much is given so better to handle this missing data. Usually Stata reads missing as dot in the data. I will show you if it is available in our data.

You just have a look, since we have filter for our use, I think we have already corrected the dot. All the entries are filled up. There is no dot here. But if any places you have seen any dot, only dot is entered instead of any other entries, then please note that this is a missing data. also let me operate this before you. Likewise, codebook, variable name if you give, it gives you the information.

Let me go back to the start up window once again and operate. So I will start with codebook here, if I simply enter codebook it gives all the variables and their codebooks. So I am interested for a particular variable, let me just go by state again. So let me enter.

So it gives me the information of, let us say string data and its different examples or entries are also visible and their frequency or their labeling, since labeling is not mentioned here, so not giving, but most importantly to be noted here, the sample and there is no missing value. Similarly, you can check please do experiment with other variables, you will find some interesting part of it and you may raise this question in our discussion time.

(Refer Slide Time: 41:50)



So let me explain you another interesting part that most of the researchers generally do not know. I claim this as important because there are certain important aspects for you to do some advance analysis or advance simplification of your data. For better simplification of our data or operating our data this command is going to be useful. Codebook provides extra information on the variables such as summary statistics of numerics, example data points of strings as well.

Codebook, compact is another aspect where all the variables are opened in another window with all such information, variable, then observation, unique ID and information mean, minimum value, then maximum value and the labeling. I will tell you right now. Codebook comma compact, this has given me information in the window, in the same window that how result is, the data, the codes can be read in different approach like a number of observations is visible, variable and whether it is unique or not. Look at the first one ID. ID is expected to be unique, but not necessarily if there are other way of checking whether it is unique or not, I will tell you in a short while. But at this moment these two are same. I am going to tell you why this is going to be important.

It seems that this variable is uniquely identified, but I am emphasizing now on the compact information. This gives in a vertical window of all that information and their labeling also. At a go, you can check all those details. So let me come back to our explanation. Please take a note here, if you do not say what to describe or summarize or codebook, Stata assumes that you want to perform those commands for every variable in the memory. So, if you specify, then it will give you. Otherwise, all details will be displayed on the window.

(Refer Slide Time: 44:32)



This is going to be very interesting for all of us to note that identifying the type of each variable, like I told you that some colors are visible, colors in red, colors in black, colors in blue. But let us understand, what are those variables? Why string? Where to get the string? How to read a string? So most common variable types are continuous, categorical, string and identifier variable. Usually these four categories are available in the existing datasets we are going to work with.

(Refer Slide Time: 45:25)



So the continuous variables mean that can take infinite numbers or values. Those also referred to as quantitative variables, they contain correspond to some quantity in the real world and mathematical operations are really possible, because these are continuous variable and these are quantitative in nature. Some example we are already discussed like income of the consumer, expenditure, gross value added of an enterprise etc. are called continuous variable.

(Refer Slide Time: 46:02)



Categorical variable, the word categorical again of three types. the meaning category maybe in order or may not be in order, so unordered or ordered or indicator variable. So like these are also called, categorical variable, also called factor variables. And take on finite set of values, often called levels. So since its ordered or unordered or even indicator, some boundary limits are defined.

Based on the boundary limits the ordering or non-ordering is possible and this is separate kind of variable. So I am using these words for clarifying the concept of categorical variable. These are sometimes also called close ended variables. The end is closed it is not open ended unlike the continuous variable.

The levels are typically stored as numbers, basically labeled sometimes with one or two if it is taken gender as male and female or third is transgender. In our data, we are including all, the

National Sample Survey has already included transgender, for your clarity, so three is already there. You may operate with that or you may also operate with another one.

Please take a note that the numbers do not represent quantities. These are just though coded or labeled with a number, but they are never represents the magnitude of the quantity or the degree of the quantity or the value of the quantity. Two does not mean higher than that of one. It is not greater than one.

So please do not read in case of order or even unordered, ordered generally have certain higher values, is usually higher values ordered or in lesser value ordered, but in case of unordered one, any number, those number never represents any value. So mathematical operations just by this number is not possible. There are some specific mathematical operations suggested. We will tell you in that particular lecture.

(Refer Slide Time: 48:36)

 Categorical variables can also be stored as strings. With Unordered categorical variables, the numbers assigned are
completely arbitrary.
□ E.g., 1= female, 2= transgender and 3= male.
So, mathematical operations has no meaning.
With Ordered categorical variables, the levels have some natural order. For instance, Likert scale (1=Very Dissatisfied, 2=Dissatisfied, 3=Neither Satisfied nor Dissatisfied, 4=Satisfied, 5=Very Satisfied).
Indicator variables, also called binary variables or dummy variables, are just categorical variables with two levels.
Often they answer the question "Is some condition true for this observation?"
Swayam 🧕 18

Categorical variable can also be stored as strings. As I mentioned this word while understanding string variable, categorical variables like education in case of standard of education, first standard, second standard, third standard, those are categorical. But first, instead of one, two, three, they can carry with first roman letter, second roman letter, third roman letter. So if those entries are there, you cannot just operate all conventional statistical or stata commands will give

you the result is a wrong notion. You have to convert those string to numeric values for better operation.

With unordered categorical variables, I have already discussed about categorical, if it is unordered one, the numbers assigned are completely arbitrary. It is completely arbitrary, like one for female, for gender I have already mentioned. So, mathematical operation has no meaning. I have already clarified that.

With ordered categorical variable, like standard as I told you or maybe customer care and their ratings usually in 5 point Likert scale, 7 point Likert scale these are in order, in different order. And so far as indicator variables are concerned, like if we are asking this type of question, are you satisfied with our service. So you have to indicate one.

Though options are generally yes and no, but you have to indicate one. It is nothing to do with the ordering or with its category. It does signal an indicator, so generally called dummy variables or binary variables. So often the answering the question is like some condition true for this observation or options will be yes or no.

(Refer Slide Time: 51:00)





String variable, string variables contain, some further clarity to string variable is essential. I have already mentioned two types here out of four, continuous, categorical we explained, we are explaining string variables. Then we will explain identifier variables also. The string variables contain text, as I clarified several times.

Sometimes the text is just labels for categories like, the categories can be simply labeled and put in label, but the labels are not numeric values. So they can be treated like categorical variables only. Other times, they contain actual information. Even if actual information, some description information is there, descriptions are there, numerical operations just based on string variable is not possible. We have to operate in a different way, we will suggest in a short while.

So last one, understanding the variables which are entered in the data is called identifier variables. This identifier variable allows you to find observation rather than containing information about them. So, regarding observation, whether this is the identifier or not, is very much understood through the identifier variables. It has nothing to do with the number or nothing to do with the string number, nothing to do with the categories or not, only identifier, where this identify the particular row or identify the particular person or the household is very very important for understanding the data. I think I am little faster for it, but since I am providing everything together, friends you can understand from my slides.

So I will tell you through the data. I have already given you the background of different type of variables, different type of data. I have shown you, data in the sense the colors in the data. In the browsing window I have shown you the data. So I am clarifying the identifier variables.

(Refer Slide Time: 53:19)

FINDING IDENTIFIER VARIABLES	
 Identifier variables, also known as primary keys or index are variables whose purpose is to identify observations. If you have a dataset on IITR students then stud enrollment id would be a unique identifier, that allows you identify a single row. 	r es , ent u to
sometimes a dataset is not uniquely identified by a sir identifier variable. e.g.	gle
🚳 swajani 🔮	20

Identifier variables also known as primary key or the indexes and are the variables whose purpose is to identify the particular observation. In the row which particular observation is unique or there are so many observations of the same category. If it is not unique, then there will be some problem. We are going to tell you please hold for some time. In our lecture, we will certainly clarify.

If you have a dataset on IITR student, one example, IIT Roorkee students we are citing here, student enrollment ID would be a unique identifier, like student ID number, unique ID they define is generally the unique identifier, which allows you to identify as a single row. If all the students and their information are mentioned, but the starting one is your enrollment number against to that your age, your class name and your batch name, your subject name, your teacher name or your previous percentage whatever is written, but unique ID is already defined then that is perfectly fine.

But sometimes the dataset may not be having with a single unique ID. It might be coming with a compound variable. You may have to go by composite variable to define a unique ID. In the IITR student, for example, here let me read out once then I will clarify.

(Refer Slide Time: 55:00)



Suppose a dataset describing IITR students, but there is one row for each class the students are taking, with each class being identified by a number. In order to identify a specific row you need to specify both a student ID number and a class number. Student ID is there, but you need to identify which particular class.

Class, you wanted to add further information to the ID, like this represents also the class number, also the student ID number. If we are also attaching some information to the ID from, if I am the administrator, I wanted to track the person about his particular class, if that is attached with the unique ID, some additional information is attached with the unique ID, then from the new unique ID, you can easily track the particular observation or the student.

So basically what I wanted to say that combining this two information that is unique student ID with the class number. Combination of student ID and the class number would be a compound identifier for the dataset and the combination of the two is a unique ID or unique identifier. Similarly, in our dataset like IHDS data, there are so many information required, not just one ID,

the first column usually the ID, but that is not enough for observation to be defined as unique. I will guide you. It is there in our content.

So we have tried to clarify the previous one where how to define the unique ID. I will also test it through ISID command we have in our PPT, in our operation. I will tell you in our next class. So let me stop here and thank you so much we will continue from the next class.