

Statistics 341dataset_SPSS.htm 341dataset_SPSS.pdf	Michael Hallstone, Ph.D. hallston@hawaii.edu
----------------------------------------------------------	-------------------------------------------------

How to Create a Data Set Using PASW/SPSS

You should understand the material in the following lectures before doing this exercise: [lecture 1: populations and samples](#) and ["lecture 2: types of variables."](#)

The purpose of this exercise is to show you how to create a data set on PASW/SPSS. You will need a data set for the take home exams.

By the way if you want to know what to do with missing values for income or any other variable see the very end of this document.

Please oh please oh please [did I say please?] read this whole document BEFORE you start to create your spss data set!!! Believe it or not it will save you much time in the long run. And believe it or not, despite me saying this here, some will ignore this advice, get a lower hw 2 grade, and still have to go back and fix their spss data set anyway, thus spending more time and effort in the long run.

1. Submitted a proposed questionnaire that is ethical
2. Had that questionnaire approved by me and instructed to “go collect your data”
3. Have completed questionnaires from your non-random sample. [See HW #1 for the exact number of questionnaires you need filled out. It is changed every semester to prevent cheating.](#)
4. Access to a computer with PASW/SPSS
5. How many variables will be in your data set? Count the exact number of questions you have on your survey. You have your questions on top and the mandatory demographic variables on the bottom of the survey that I made you collect. Count the total number of questions on your survey. That is how many variables will be in your data set. Each question on the survey will be a single variable. If you had 12 questions in your survey you will have 12 variables in your data set! If you had 11 questions you would have 11 variables, and so on.

If you have not completed those steps you can practice using the data from my fictitious project below.

Fake Project Topic

The administration, in its infinite wisdom, has decided that the lack of local cuisine on campus is a crying shame. So they invite *Arno's* restaurant in Waianae to open a concession on campus. Only problem is Arno has enough to do running his restaurant out in Waianae, so he doesn't have the time to cook his reasonably priced and *onolicious* food here on campus. So Arno unwisely hires me to run his concession. But I don't have the time either (as I must still conduct research) so I have to figure out a way to do both. I can only offer four different lunch plates: *Uncle Butch's LauLau*, *Waianae Backyard Aquaculture Sunfish*, *Auntie Pua's Paddling Pasta Salad*, and *Hallstone's Haole Hamburger Plate*. Now I know that poi fills you up, is one of the world's best complex carbohydrates, and is just down right good tasting so I'm going to serve it with every lunch plate I make.

Here is my hypothesis: People who order the Hawaiian food (lau lau and sunfish) will be smarter than those who order burgers and pasta salad. Furthermore, those who eat poi will be smarter than those who don't eat it. There's this guy in sensible shoes, Ray-Ban sunglasses, and gaudy tourist Aloha shirts with perpetual sunburn who is always hanging around anything "Hawaiian" on campus; he has this little machine that can read your finger prints from a bizzillion miles away. He has some way of using fingerprints to access information about people that is stored in all of these government computers (as well as UH's main computer). Well, even this guy can't resist Hawaiian food and pretty soon I've got him giving me everyone's Grade Point Average (GPA) who buys food in exchange for *Arno's* ono plate lunches. I'll use everyone's GPA to measure how smart they are.

Below are the 3 variables for my study, which we will code into the columns on the top of the spreadsheet. Since the statistics deals with numbers we will have to code the variables.

Variable 1 “kinefood” is nominal

- *kinefood* = Kind of plate lunch ordered: 1 = Uncle Butch’s Chicken and Butterfish Lau Lau, 2= Waianae Backyard Aquaculture Brand Sunfish, 3=Auntie Pua’s Paddling Pasta Salad, and 4= Hallstone’s Haole Hamburger Plate

Variable 2 “eatdpoi” is nominal

- *eatdpoi*= does the customer eat their poi: 1= yes , 0= no

Variable 3 “GPA” is ratio

- *GPA*= the customer’s GPA: it can be any number from 0.0 to 4.0

People who filled out survey are called “subjects”

In statistics the people who provide data for your study are called “subjects.” So, below the terms “subjects” and “people who filled out your survey” mean the same thing!

For brevity, pretend I handed out 4 surveys with 3 variables each:

I have surveya (and thus data) from 4 people (n=4) for the three variables above. Below are each of the four completed surveys. **For the sake of brevity, notice that my surveys do not have the mandatory demographic variables such as age, gender, income, etc; your surveys will have these demographic variables**

Survey from person #1

Person/subject 1 likes Lau Lau, eats the poi, and their GPA is 4.0. Notice how they filled in their survey below

Hawaiian Food Survey

My name is Michael Hallstone and I am collecting some data about Hawaiian Food for a statistics class at the University of Hawaii West Oahu.. This is not a real scientific survey and the data will only be used so that I can complete the course. Please do not write your name on this paper as all of your answers will be anonymous and confidential.

If you have any concerns about this survey, please do not fill it out. You are also free to withdraw your participation at any time while you are filling it out. Thank you very much.

What is your favorite plate lunch? (check only one)

Uncle Butch’s Chicken and Butterfish Lau Lau
Backyard Aquaculture Brand Sunfish

Waianae

Auntie Pua's Paddling Pasta Salad
 Hallstone's Haole Hamburger Plate

Do you eat poi yes no

GPA 4.0 (enter exact GPA)

All *pau*! Thank you!!!!

How to input data for Person/subject 1

The way person/subject 1 filled out their survey indicates that s/he likes Laulau, eats the poi, and their GPA is 4.0:

Using the coding above their answers to the questions are converted to numbers and are entered into the computer in the following format:

	kinefood	eatdpoi	GPA
Person/subject 1	1	1	4.0

Survey from person #2

Person/subject 2 likes Sunfish, eats the poi, and their GPA is 3.8. Notice how they filled in their survey below

Hawaiian Food Survey

My name is Michael Hallstone and I am collecting some data about Hawaiian Food for a statistics class at the University of Hawaii West Oahu.. This is not a real scientific survey and the data will only be used so that I can complete the course. Please do not write your name on this paper as all of your answers will be anonymous and confidential.

If you have any concerns about this survey, please do not fill it out. You are also free to withdraw your participation at any time while you are filling it out. Thank you very much.

What is your favorite plate lunch? (check only one)

Uncle Butch's Chicken and Butterfish Lau Lau
 Waianae Backyard Aquaculture Brand Sunfish
 Auntie Pua's Paddling Pasta Salad
 Hallstone's Haole Hamburger Plate

Do you eat poi yes no

GPA 3.8 (enter exact GPA)

All *pau*! Thank you!!!!

How to input data for Person/subject 2

The way person/subject 2 filled out their survey indicates s/he likes Sunfish, eats the poi, and their GPA is 3.8

Using the coding above their answers to the questions are converted to numbers and are entered into the computer in the following format:

	kinefood	eatdapo	GPA
Person/subject 2	2	1	3.8

Survey from person #3

Person/subject Person/subject 3 likes pasta salad, does NOT eat the poi, and their GPA is 1.2. Notice how they filled in their survey below

Hawaiian Food Survey

My name is Michael Hallstone and I am collecting some data about Hawaiian Food for a statistics class at the University of Hawaii West Oahu.. This is not a real scientific survey and the data will only be used so that I can complete the course. Please do not write your name on this paper as all of your answers will be anonymous and confidential.

If you have any concerns about this survey, please do not fill it out. You are also free to withdraw your participation at any time while you are filling it out. Thank you very much.

What is your favorite plate lunch? (check only one)

- Uncle Butch's Chicken and Butterfish Lau Lau Waianae
 Backyard Aquaculture Brand Sunfish
 Auntie Pua's Paddling Pasta Salad
 Hallstone's Haole Hamburger Plate

Do you eat poi yes no

GPA 1.2 (enter exact GPA)

All *pau*! Thank you!!!!

How to input data for Person/subject 3

The way person/subject 3 filled out their survey indicates s/he Person/subject 3 likes pasta salad, does NOT eat the poi, and their GPA is 1.2

Using the coding above their answers to the questions are converted to numbers and are entered into the computer in the following format:

	kinefood	eatdapo	GPA
Person/subject 3	3	0	1.2

Survey from person #4

Person/subject 4 likes hamburger, does NOT eat the poi, and their GPA is 0.75. Notice how they filled in their survey below

Hawaiian Food Survey

My name is Michael Hallstone and I am collecting some data about Hawaiian Food for a statistics class at the University of Hawaii West Oahu.. This is not a real scientific survey and the data will only be used so that I can complete the course. Please do not write your name on this paper as all of your answers will be anonymous and confidential.

If you have any concerns about this survey, please do not fill it out. You are also free to withdraw your participation at any time while you are filling it out. Thank you very much.

What is your favorite plate lunch? (check only one)

- Uncle Butch's Chicken and Butterfish Lau Lau Waianae
 Backyard Aquaculture Brand Sunfish
 Auntie Pua's Paddling Pasta Salad
 Hallstone's Haole Hamburger Plate

Do you eat poi yes no

GPA 0.75 (enter exact GPA)

All *pau*! Thank you!!!!

How to input data for Person/subject 4

The way person/subject 4 filled out their survey indicates that s/he likes hamburger, does NOT eat the poi, and their GPA is 0.75:

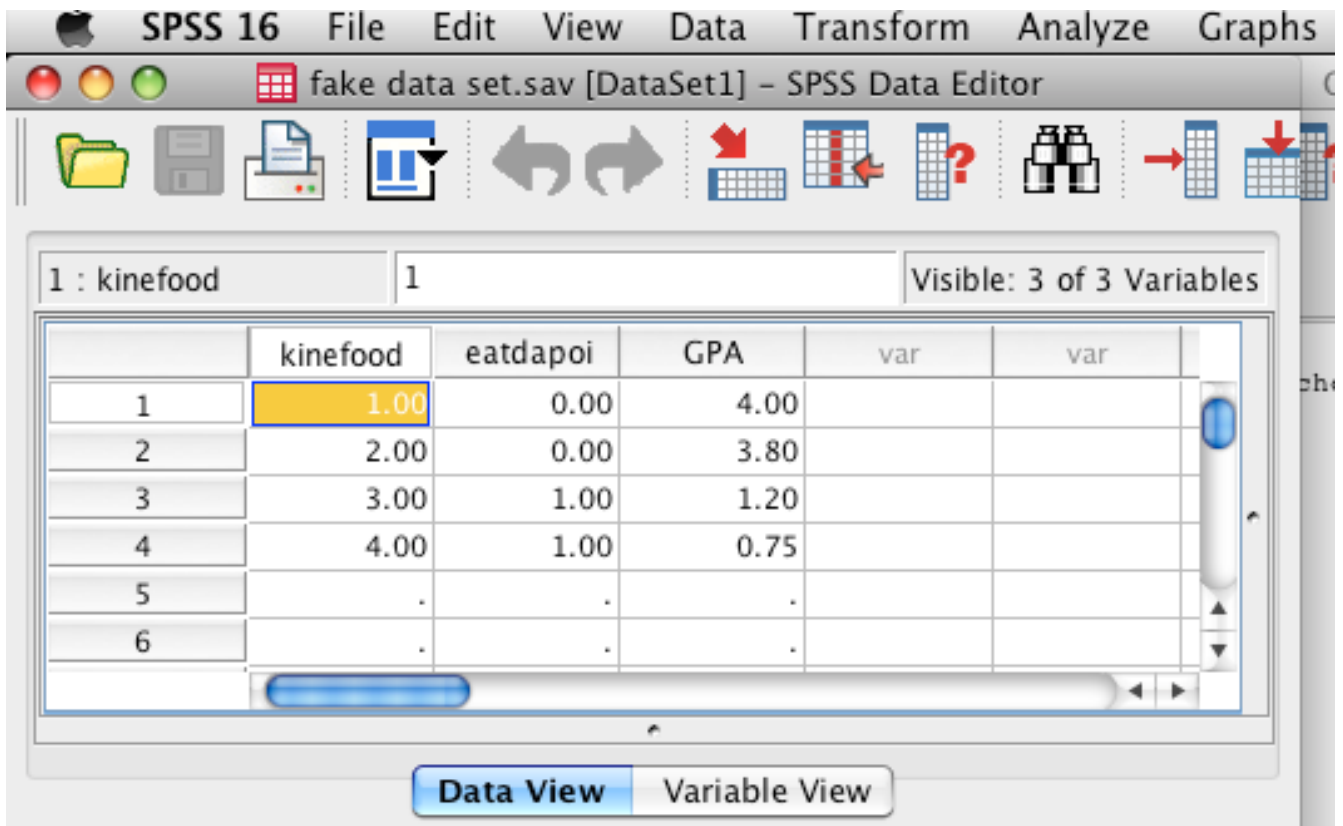
Using the coding above their answers to the questions are converted to numbers and are entered into the computer in the following format:

	kinefood	eatdpoi	GPA
Person/subject 4	4	0	0.75

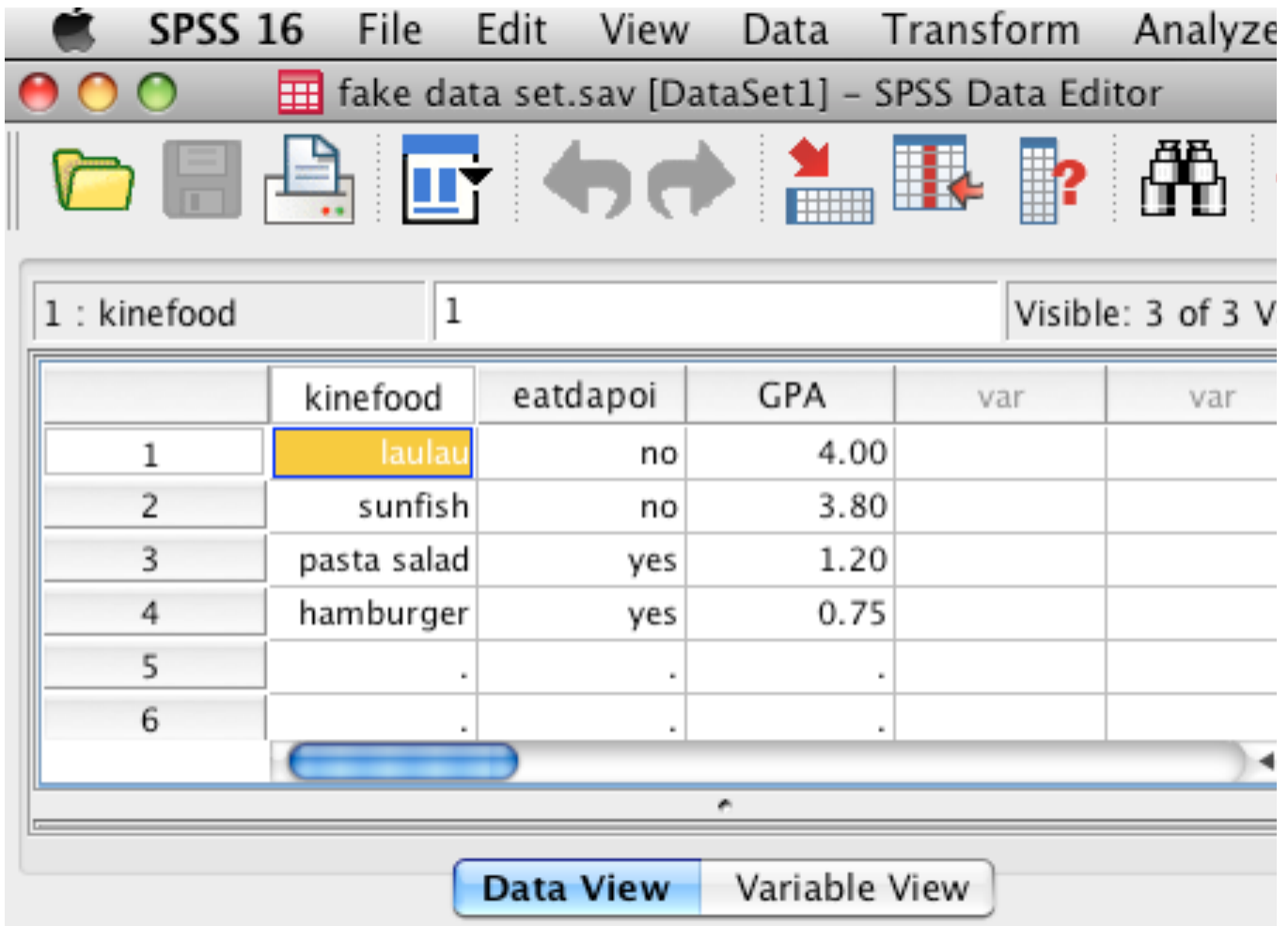
Completed fake data set will look like this

Below we will learn how to create a data set using PASW/SPSS, *but when we're done* the data set will eventually look like the picture below. Note that you have to complete all of the steps below to create the PASW/SPSS data set, but this is a preview.

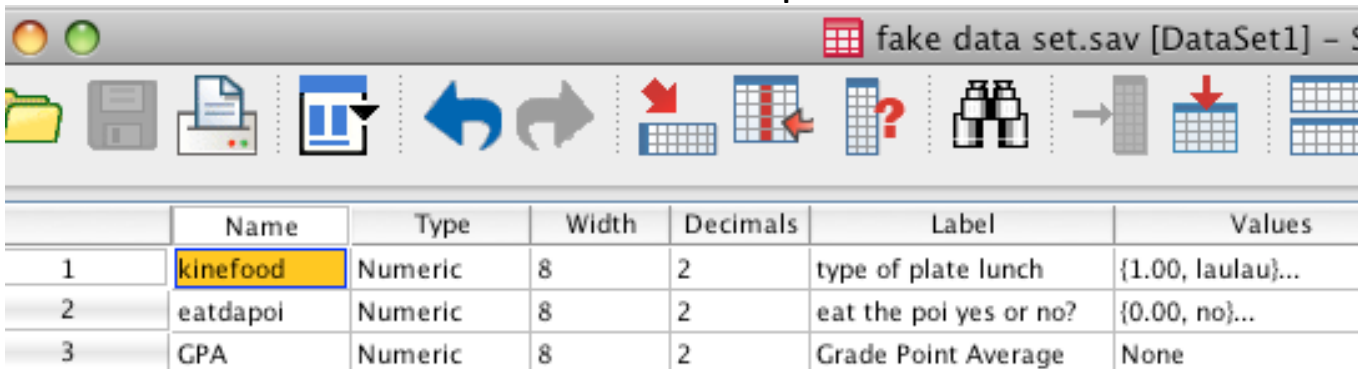
Data set EVENTUALLY looks like this in the Data View option



If you choose to view the value labels (see end of this document for how to do this) it will look like this in Data View option



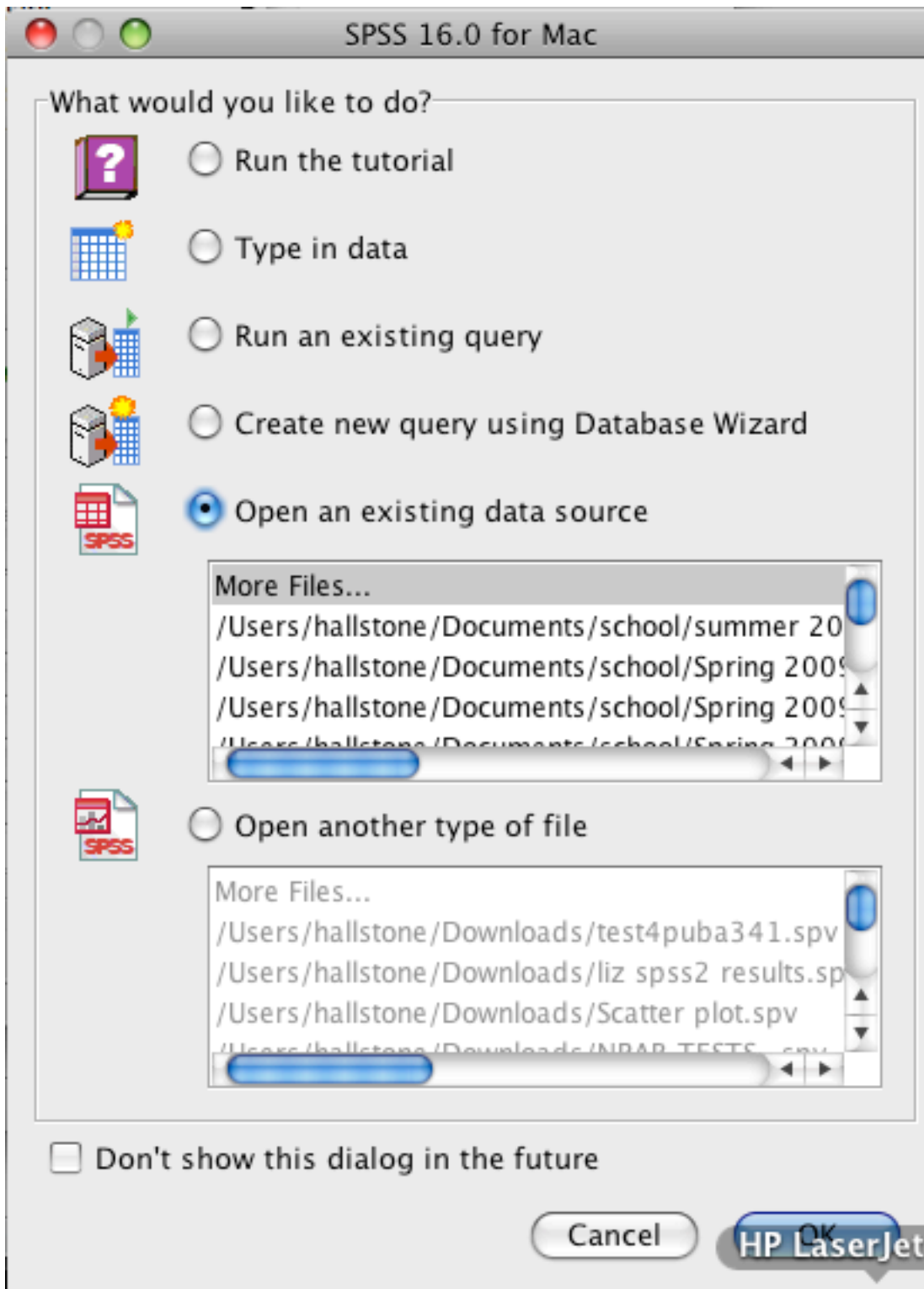
Data set EVENTUALLY looks like this in the Variable View option



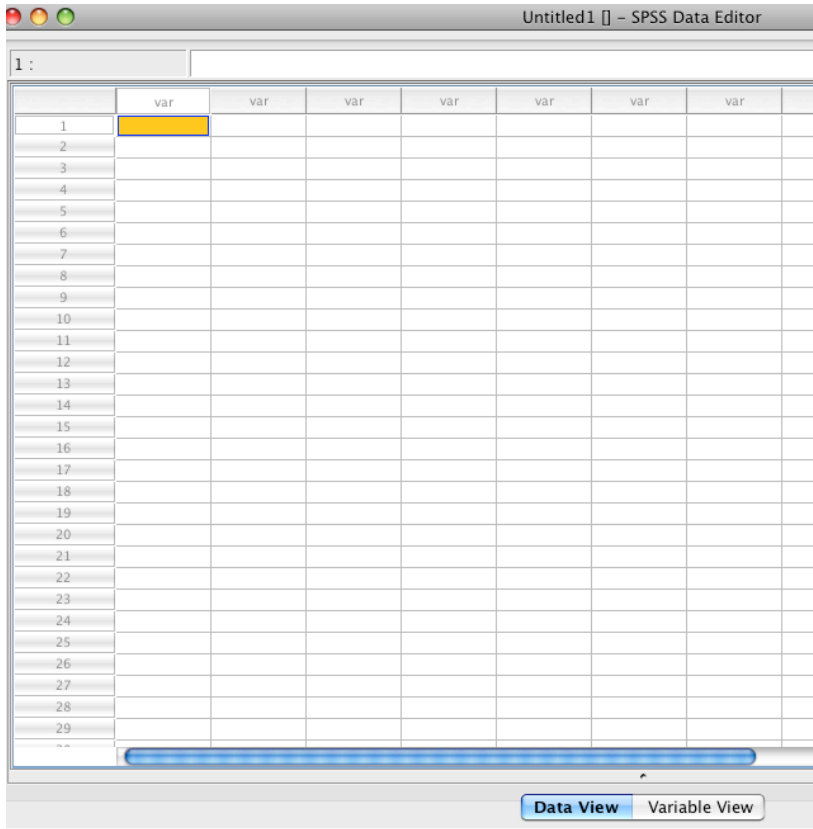
Creating a data set using PASW/SPSS

(The following instructions are based upon PASW/SPSS version 16.0 for MAC)

Open PASW/SPSS and the following screen will show up. Select "Type in Data"



Select "Variable View" at the bottom of the screen.

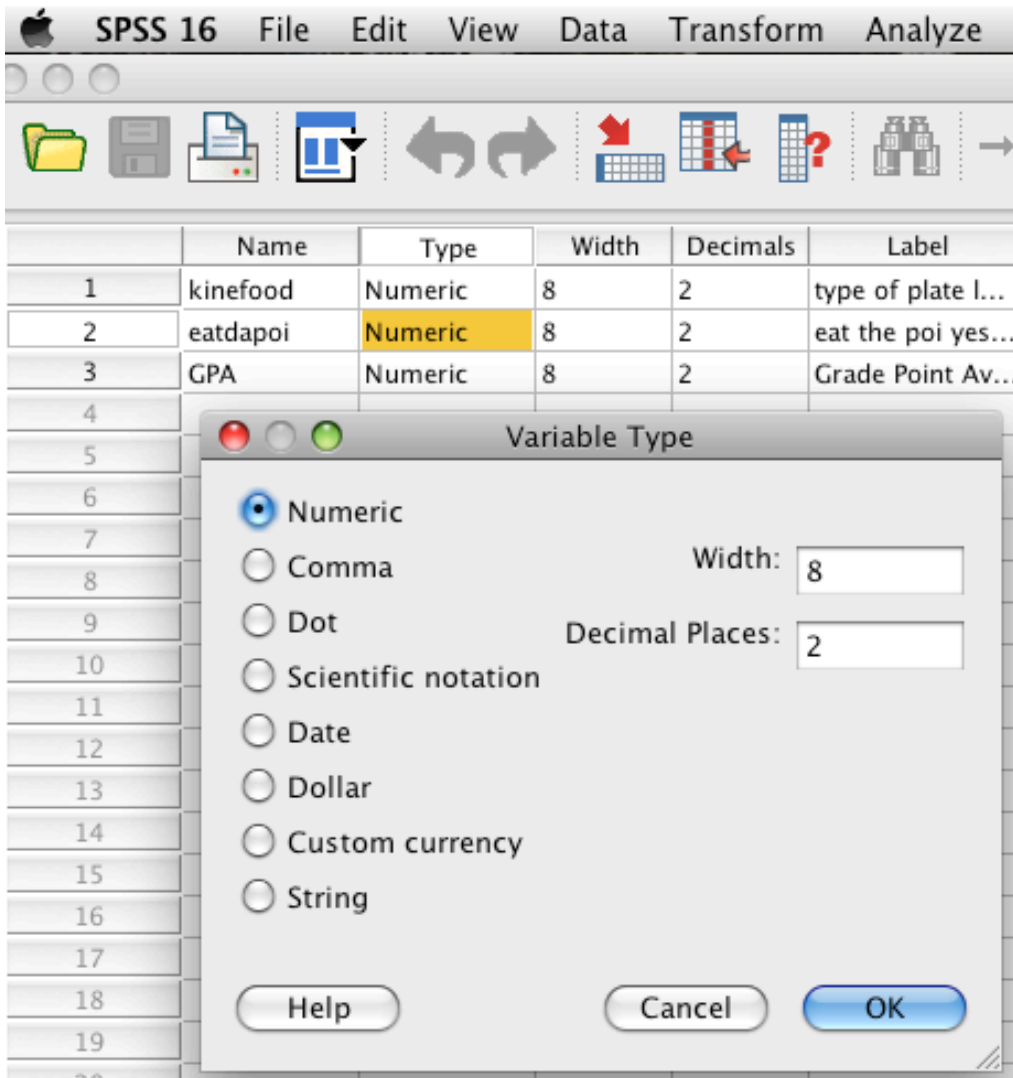


In the next screen type in the names of your variables under the column heading "Name." NOTE YOU CANNOT HAVE SPACES OR SPECIAL CHARACTERS IN YOUR VARIABLE NAMES. So you have to use short cut names

Make all variables "Numeric" in the Type column

While you are in the Variable View screen, make sure all of your variables are "Numeric" in the Type column [the second column over]. If you choose "String" or another variable type you may have problems in later PASW/SPSS exercises where I show you how to actually analyze data.

When you double click on the Type box the following screen will pop up.



Labeling variables in PASW/SPSS

Next under “Label” is a chance to type a plain English phrase to remind you what your variable name means. So for example, it is hard to know what “kinefood” means or what the variable measures. Under label you can do this in plain language. I’ve labeled all three variables in the picture below. (For our class, simply cut and paste the actual question from your survey into this box. Please just do this for all of your variables as it helps with grading your test; if I don’t understand your plain English I won’t be able to grade your test. I did not cut and paste the whole question from my survey below because I wanted the screen shot to be bigger.)

Untitled1 [DataSet0] - SPSS Data Editor

	Name	Type	Width	Decimals	Label
1	kinefood	Numeric	8	2	type of plate lunch
2	eatdpoi	Numeric	8	2	do they eat the poi?
3	GPA	Numeric	8	2	Grade Point Average

Coding variables in PASW/SPSS (assigning values)

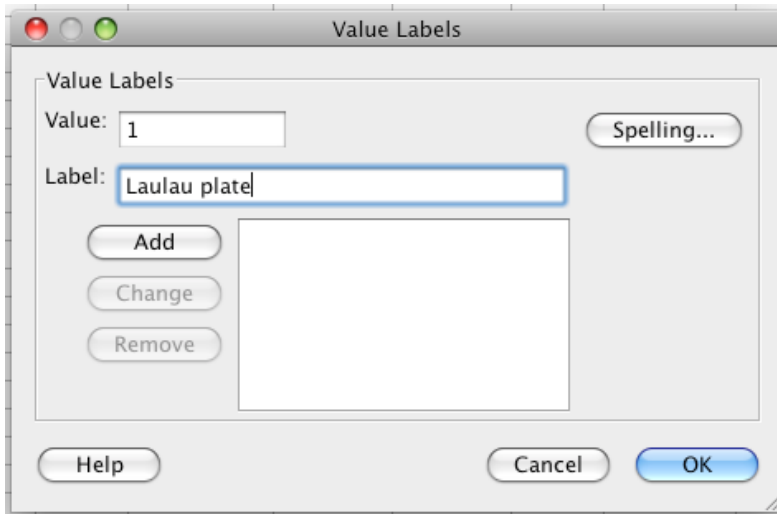
The variables *kinefood* and *eatdpoi* both require coding and we will have to tell PASW/SPSS what the numbers mean. You must tell PASW/SPSS the coding for all nominal and ordinal level variables.

There is no need to code interval or ratio level variables. Remember, the variable *GPA* is “ratio level” so the numbers “naturally code” and do not require coding. The number itself is the “coding.” For the variable *GPA* there is no need to put in value labels because there is an infinite amount of possible values between 0 and 4.0. This is true for any ratio level variable where the number is self-explanatory. For example if you have a variable “Age_____” the number means the “age in years.” There is no need to explain what 29 means!

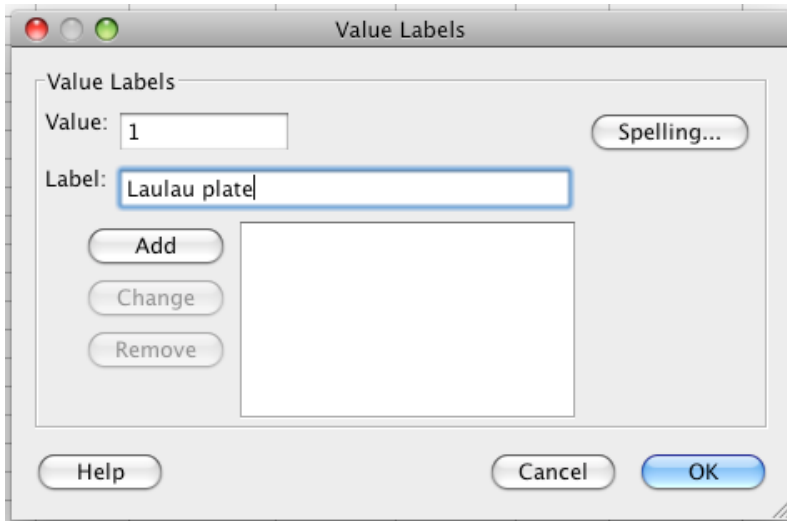
For *kinefood* click on “Values” and then on the square with the dots

	Name	Type	Width	Decimals	Label	Values
1	kinefood	Numeric	8	2		None <input type="checkbox"/>
2	eatdpoi	Numeric	8	2		None
3	GPA	Numeric	8	2		None

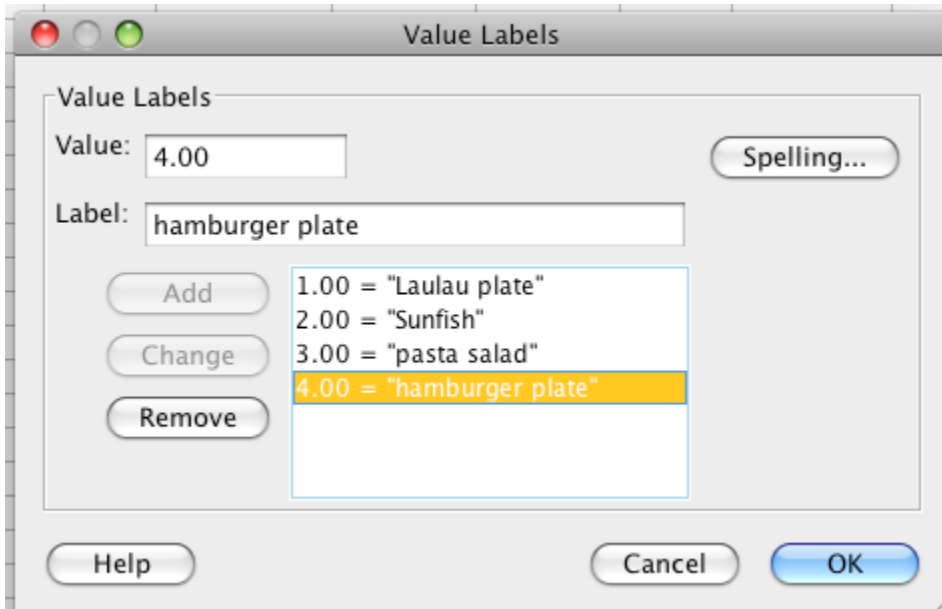
Tell PASW/SPSS that 1=lalau plate, then push add



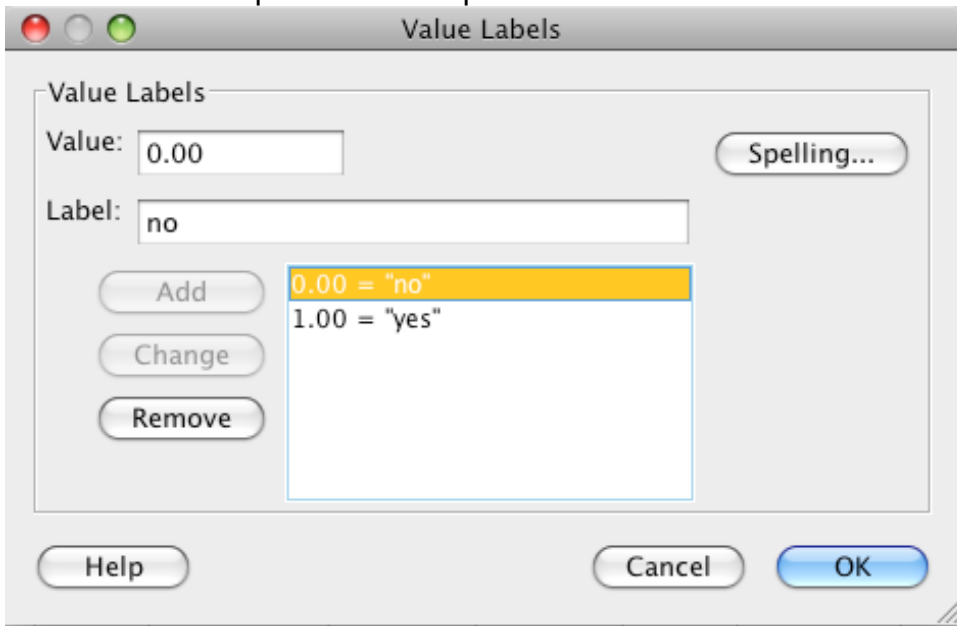
After you push add it will look like this:



Tell PASW/SPSS that 2=sunfish and push add. Repeat for 3=pasta salad and 4= hamburger and it will look like this:



Push "OK" and repeat for "eatdapo"



Push OK and then push "Data View" at the bottom of the screen, which will return you to the following screen:

1 :

	kinefood	eatdpoi	GPA	var	var	var	var
1		?	?				
2		?	?				
3		?	?				
4		?	?				
5		?	?				
6		?	?				
7		?	?				
8		?	?				
9		?	?				
10		?	?				
11		?	?				
12		?	?				
13		?	?				
14		?	?				
15		?	?				
16		?	?				
17		?	?				
18		?	?				
19		?	?				
20		?	?				
21		?	?				
22		?	?				
23		?	?				
24		?	?				
25		?	?				
26		?	?				
27		?	?				
28		?	?				
29		?	?				
--		?	?				

Data View Variable View

Typing data into PASW/SPSS

(At any point during this process save this PASW/SPSS data file to your computer.)

The numbers in the grey column on the far left hand side represent a single person/subject in your study. Take the questionnaire for the first person/subject and fill in their responses horizontally.

To keep subject in the proper row fill in data “horizontally”

You must keep data from each subject in it’s own row. ***Do not mix the rows up.***

To do this you fill in the data horizontally across rows.

Subject #1’s data

Recall that person/subject 1 likes laulau, eats poi, and has a GPA of 4.0. ***Note how I filled in their data horizontally or across the page.***

	kinefood	eatdapo	GPA
1	1.00	1.00	4.00
2			
3			

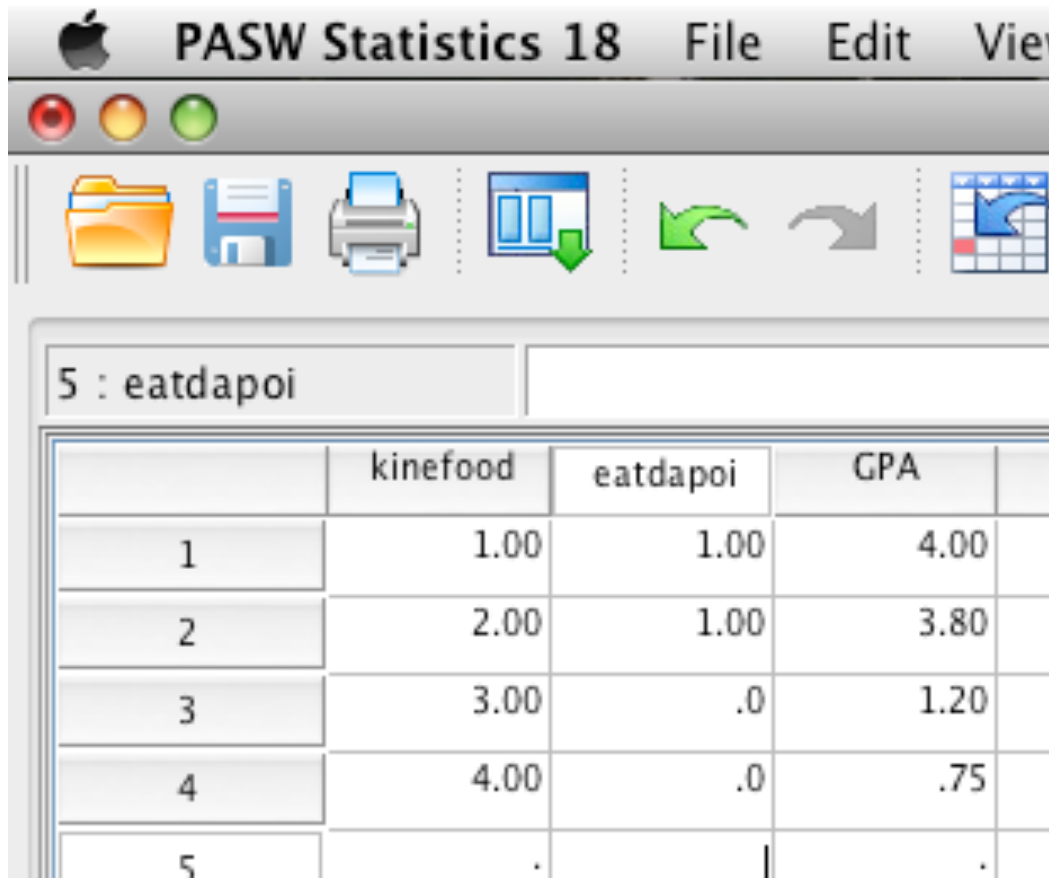
Subject #2's data

Recall that person/subject 2 likes sunfish, eats poi, and has a GPA of 3.8.

	kinefood	eatdapo	GPA
1	1.00	1.00	4.00
2	2.00	1.00	3.80
3			

Subject #3's data goes in Row #3 and so on and so on.

Person/subject 3 likes pasta salad, does not eat poi, and has a GPA of 1.2 Person 4 likes hamburger plate, does not eat poi, and has GPA of 0.75.



The screenshot shows the PASW Statistics 18 software interface. The title bar includes the Apple logo, the text "PASW Statistics 18", and menu options "File", "Edit", and "View". Below the title bar are three window control buttons (red, yellow, green) and a toolbar with icons for folder, save, print, zoom, undo, redo, and a grid. The main window displays a data table with the following structure:

	kinefood	eatdapo	GPA	
1	1.00	1.00	4.00	
2	2.00	1.00	3.80	
3	3.00	.0	1.20	
4	4.00	.0	.75	
5	.		.	

What to do with “missing values” for income [or any variable actually]

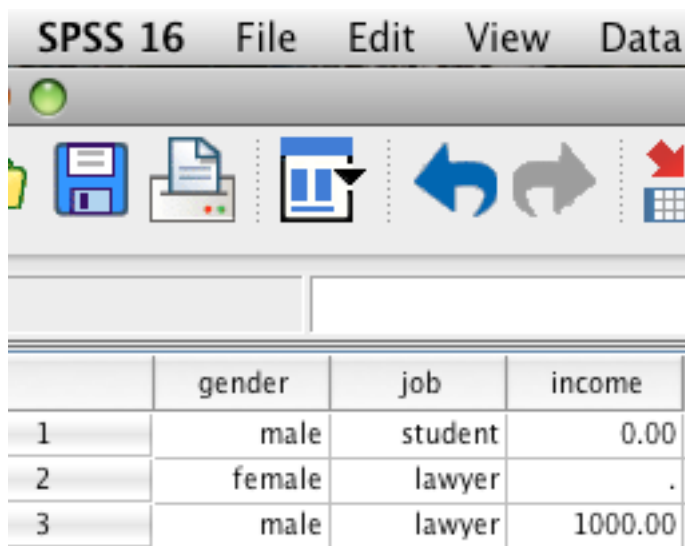
I suggest that almost all students have a variable in their survey that measures income such as:

Monthly Income _____ [enter exact dollar amount]

However this not the best way to measure income and is actually one of the worst ways to do so! I just want you to have a ratio level variable. It just so happens that income is a very sensitive topic in the US. In fact folks will tell you about things you would think are private, such as deviant sexual practices, but when you ask them their income they say, “That’s none of your business!”

Making a long story short, it is likely that you will have some people who do not answer this question. Sometimes they leave it blank and sometimes they write “n/a” or something.

This is okay! If someone refuses to answer a question leave the space blank and a little dot will show up. Do not enter 0 if they refused to answer the question!



The image shows a screenshot of the SPSS 16 software interface. The menu bar at the top includes 'SPSS 16', 'File', 'Edit', 'View', and 'Data'. Below the menu bar is a toolbar with icons for saving, opening, printing, undo, redo, and deleting. The main window displays a data table with the following content:

	gender	job	income
1	male	student	0.00
2	female	lawyer	.
3	male	lawyer	1000.00

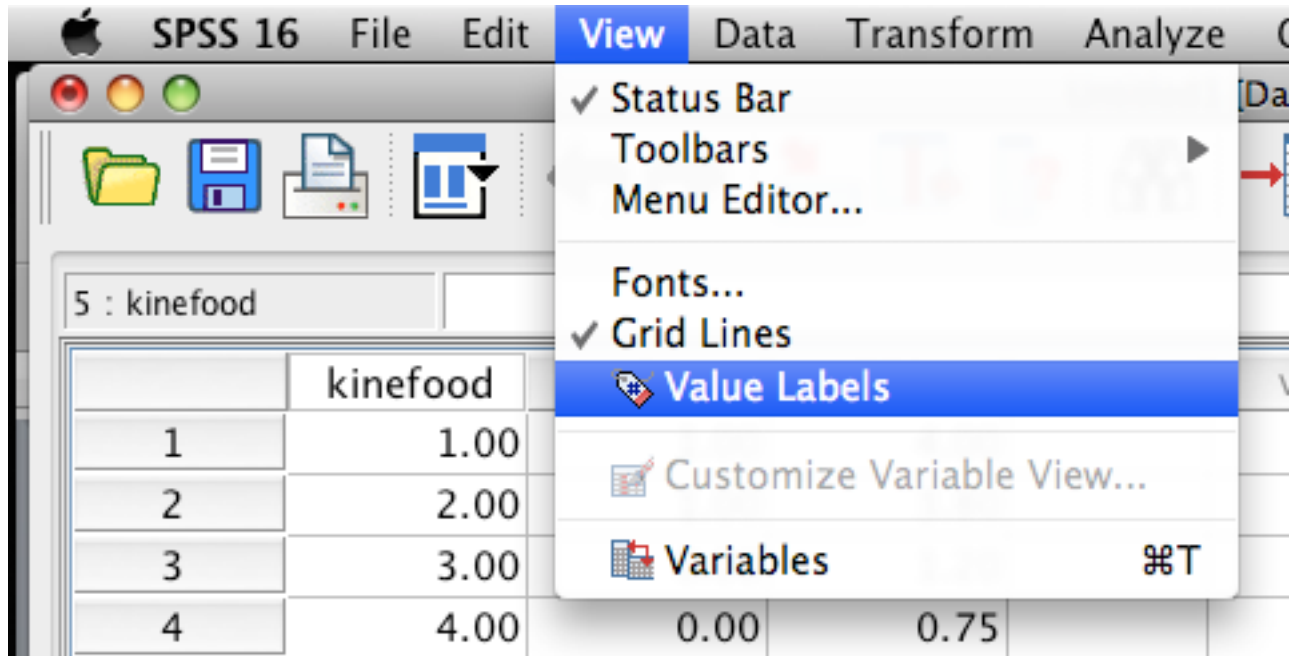
So above we have three people. Let’s pretend
Person#1: a male student does not have a job and entered 0 for income. This is probably correct as a student without a job is likely to have 0 income.

Person#2: a female lawyer either left income blank or entered 0. It is unlikely that she has zero income so it is best to leave it blank as a missing value. **The missing value shows up as a “dot” or period.**

Person #3: a male stockbroker put in 1,000 , so that is what we enter.

You can tell PASW/SPSS you want to view variable labels (coding) instead of numbers

If you prefer to see the coding of the variables you can do so in PASW/SPSS. Go to “View” and select “Value Labels”



Notice how the numbers are replaced with the value labels we typed in earlier during this exercise. Also notice that we see the coding or the labels for the variables *kinefood* and *eatdapo* but NOT for GPA (as it is a ratio level variable).

The image shows a screenshot of the SPSS 16 software interface. The title bar at the top reads "SPSS 16" and includes standard menu items: "File", "Edit", "View", "Data", "Transform", and "An". Below the title bar is a toolbar with various icons for file operations (like folders, save, print) and data manipulation (like undo, redo, delete, insert). The main window displays a data editor with a single data file named "5 : kinefood". The data is organized into a table with three columns: "kinefood", "eatdapoi", and "GPA". The table contains four rows of data, with the first row highlighted in blue.

	kinefood	eatdapoi	GPA
1	Laulau plate	yes	4.00
2	sunfish	yes	3.80
3	pasta salad	no	1.20
4	hamburger plate	no	0.75

Reminder of what to send me

You are done creating a data set on PASW/SPSS! Now create a PASW/SPSS data set using your questionnaires. When you are done, save the file and send it to me via an email attachment.

- **Remember all PASW/SPSS data files end in *.sav. Name the file like this: "Lastname, firstname.sav" so if your name was Michael Hallstone you would name the file "Hallstone, Michael.sav"**
- **Remember to indicate whether each variable is nominal, ordinal, interval, or ratio (or you can change the variables in the "variable view" in the "measure column" [the far right column] of PASW/SPSS. Interval and Ratio are called "scale." Nominal and Ordinal are called Nominal or Ordinal. Click on the box next to the variable you want to change and you'll be able to change it. See picture below.**

Values	Missing	Columns	Align	Measure
1.00, unite...	None	8	Right	Nominal
1.00, no ba...	None	8	Right	Nominal
1.00, unite...	None	8	Right	Nominal
Vone	None	8	Right	Scale
Vone	None	8	Right	Scale
Vone	None	8	Right	Scale
Vone	None	8	Right	Scale
Vone	None	8	Right	Scale
1.00, fema...	None	8	Right	Nominal
1.00, polyn...	None	8	Right	Nominal
Vone	None	8	Right	Scale
0.00, no}...	None	8	Right	Nominal
Vone	None	8	Right	Scale

Variable View