

Working with Your Data in SPSS



Data (.sav)

Before doing any analysis, make sure to save a completely “clean” version of the data. If you get confused about coding or mess up relationships, you should always have a clean dataset to return to.

Data View

Includes actual numerical data (DO NOT MANIPULATE)

Variable View

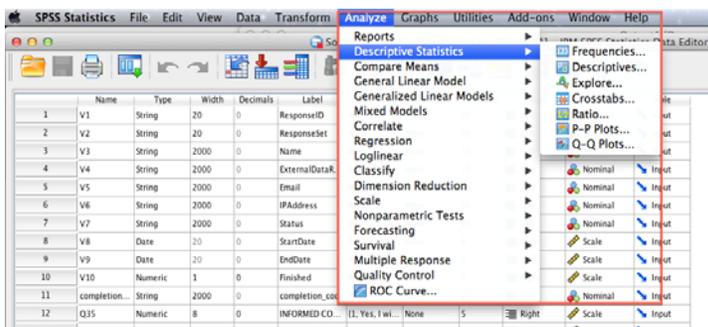
About each variable, check these in particular for correctness:

- **Type:** should be...
 - numeric for all quantitative measures
 - string for open-ended questions
- **Label:** Question wording, derived from Qualtrics
- **Values:** Attribute labels, derived from Qualtrics

1. paraid	paraid	respd	weight	int_date	lang	area	st
1	1	1	3.02	101705	2	727	
2	2	2	2.79	100605	1	396	
3	3	2	2.56	101705	2	770	
4	4	3	2.35	101705	2	718	
5	5	4	2.69	100505	1	215	
6	6	4	2.13	101705	2	786	
7	7	5	1.29	100505	1	336	
8	8	5	3.15	101705	2	716	
9	9	6	2.33	100905	1	704	
10	10	6	2.23	101705	2	973	
11	11	7	1.02	101705	2	816	
12	12	7	1.13	100505	1	850	
13	13	8	3.27	102705	1	813	
14	14	9	2.92	100505	1	276	
15	15	9	3.71	101705	2	408	
16	16	10	3.46	102205	2	626	
17	17	11	1.88	100505	1	518	
18	18	11	3.71	101805	2	908	
19	19	12	3.71	101805	2	305	
20	20	12	1.52	100505	1	845	
21	21	13	1.71	102105	2	201	
22	22	13	1.79	102605	1	919	
23	23	14	2.56	101805	2	757	
24	24	15	2.92	102905	2	956	

Drop-down menus

- **Data** is for making modifications to the dataset as a whole (like merging)
- **Transform** is for modifications to variables and their values
- **Analyze** has most all the things you need to look at the data and do statistical analysis

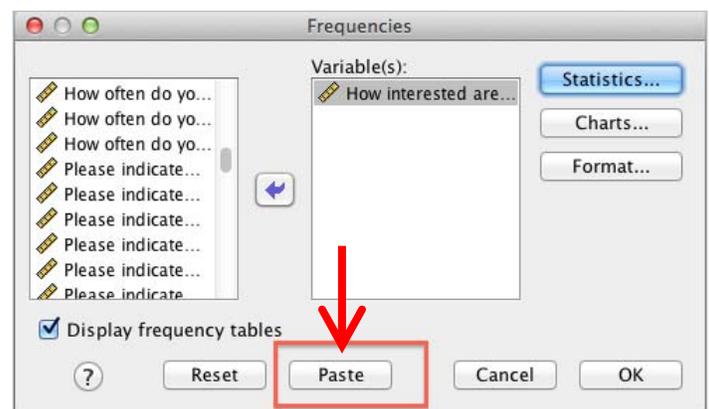


Paste

When doing anything using the drop-down menus, use **Paste** rather than **OK** to complete the task.

It is crucial to have a record of changes you made to the data so you can redo them if you made a mistake.

Paste your syntax for analyses to be able to re-run the same thing much quicker, redo your output, or quickly show someone else what you are trying to do.





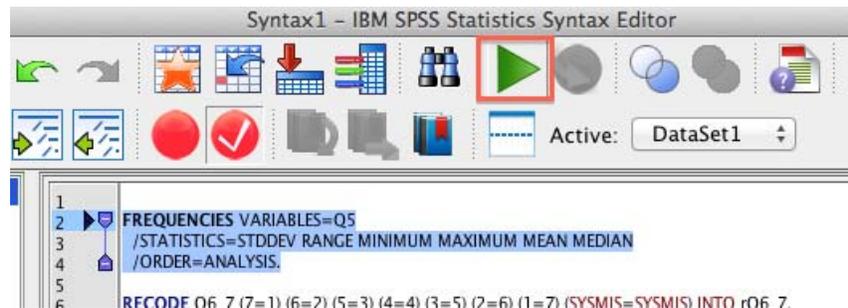
Syntax (.sps)

Where data analysis is performed.

- Always paste the syntax when using the menus. It will be put at the end of this file.
- Don't forget to save it and back it up. This is often your most important file.

When pasted into the syntax file, you then have to run the analyses to see your results:

1. Highlight the syntax you want to run
2. Click the green arrow
3. Look at the Output file for the results



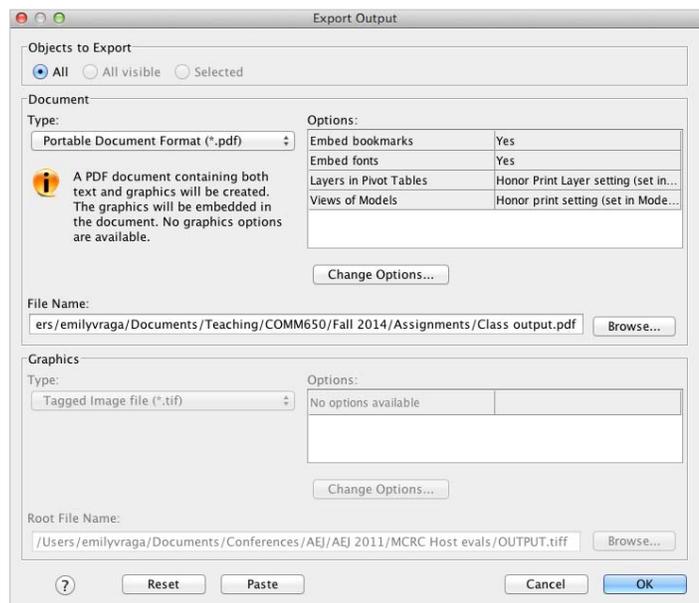
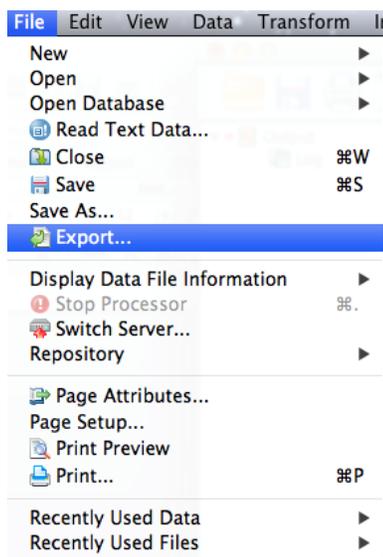
Output (.spv)

The results of the analyses you run.

This is a separate file from your data and will have a different file type extension (.spv). The output is the log of the steps you completed and the results generated, including tables or graphs you created.

You can export the output to Word or PDF. If you save your syntax, you may not need to save output.

File | Export



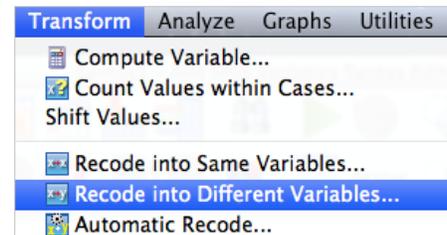
Create a New Variable w/ Recode

The first way to create a new variable is re-code, which means change the values assigned to groups. Use this when you have one variable and you want to create one new variable with different values:

- **Reduce** the number of categories (e.g., a five-point Likert scale approve-disapprove into 3 categories or divide responses of a 4-level frequency question into 2 groups).
- **Classify** knowledge questions as right or wrong, (e.g., 1 is for correct and 0 is for incorrect)
- **Reverse** code to make all values in the same direction (e.g., higher = more) before combining
- **Fill-in** missing values when they should not be missing

Transform | Recode as Different Variables

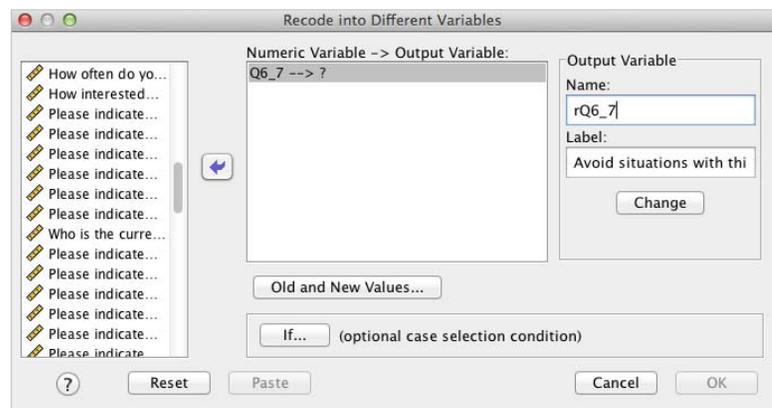
DO NOT choose Recode into Same Variable – it is too easy to forget which way the original is coded



1. Put the variable of interest into the box. You can put in other variables if they will be transformed exactly the same way.

2. Specify a new **Output Variable** name and label on the right side.

Use something easy to remember for the name that signifies it's a recoded variable, such as XX_correct for knowledge questions or rqXX for reverse coding.



3. Click **Change**

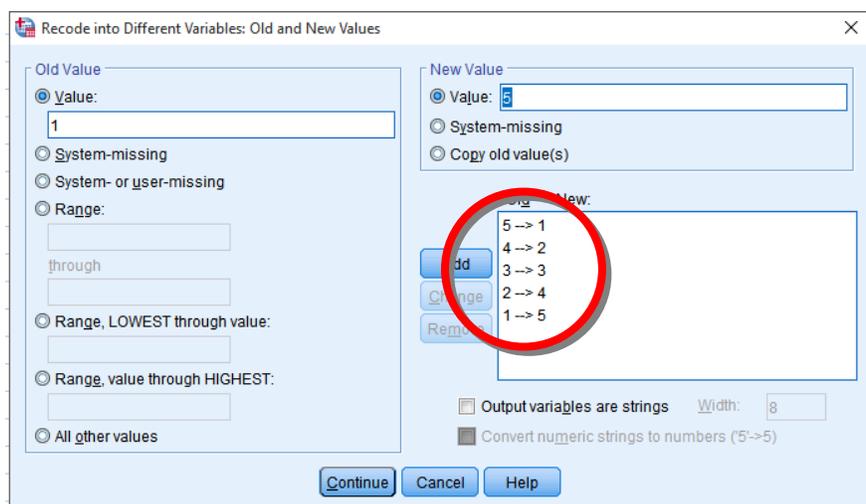
4. Click **“Old and new values”** to specify how you want the existing values to be recoded

5. Fill the Old->New box with your rules. The next page has instructions for various circumstances.

6. If your data is not “clean” or you are unsure, add a final rule at the end: Choose **“All other values”** on the left, and **“System-missing”** on the right and click **Add**.

7. Click **Continue** to return to the main window

8. Click **Paste**, then Run.



Recode: Reverse Code

Often used so that a higher value indicates more of a concept, prior to combining. For example, if you are reverse coding a variable measured from 1-5:

- Under **Old Value**, type the highest value (“5”)
- Under **New Value**, type the lowest value (“1”)
- Click **Add**
- Repeat with the next highest and next lowest until all values are accounted for. Include a value even if it doesn’t change (e.g., the midpoint of an odd-numbered scale) so that it will be copied.

Old	-->	New:
5	-->	1
4	-->	2
3	-->	3
2	-->	4
1	-->	5

Recode: Right vs Wrong

To group answers for knowledge questions as binary - either right or wrong. Must be done before calculating a score. Always [re-]code the values so that 1 means correct and 0 means incorrect. It is useful to create new variables even if the values will not be changed so it is easier to select all of them.

- Under **Old Value**, type the value that represents a correct response. For a multiple response or binary-choice question in which the correct answer was to *not* respond (e.g., check the box), choose “System-missing”.
- Under **New Value**, type “1”
- Click **Add**
- For each value that represents a wrong answer:
 - Under **Old Value**, type the value of the wrong answer
 - Under **New Value**, type “0”
 - Click **Add**

Old	-->	New:
1	-->	0
2	-->	0
3	-->	1
4	-->	0

Recode: Group Values into Ranges

To group values of an interval-level (numeric/continuous) variable into ranges.

- Under **Old Value**, click **Range**, then type the low and high values for the group in the two boxes, respectively.
- Under **New Value**, type a new ordinal value to represent that group, such as 1, 2, or 3 for low, medium, and high
- Click **Add**
- Repeat until all values are accounted for. Include values even if they do not change.

Old	-->	New:
1 thru 3	-->	1
4 thru 6	-->	2
7 thru 9	-->	3

Recode: Fill in Missing Data

For “select all that apply” style questions, sometimes the data will show a missing value when a particular box was not checked and a box number for those that were. Because analyses cannot be done with “System-missing” values, we should replace them with “0” to indicate “no response”.

- Under **Old Value**, choose “System-missing”
- Under **New Value**, type “0” (e.g., not selected)
- Click **Add**
- Under **Old Value**, choose “All other values”
- Under **New Value**, type “1” (e.g., selected)
- Click **Add**

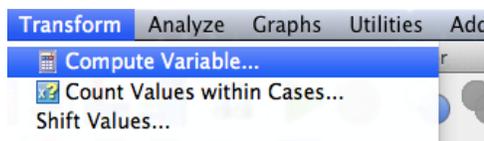
Old	-->	New:
SYSMIS	-->	0
ELSE	-->	1

Create a New Variable w/ Compute

The second way to create a new variable is to use math or logic to modify or combine values:

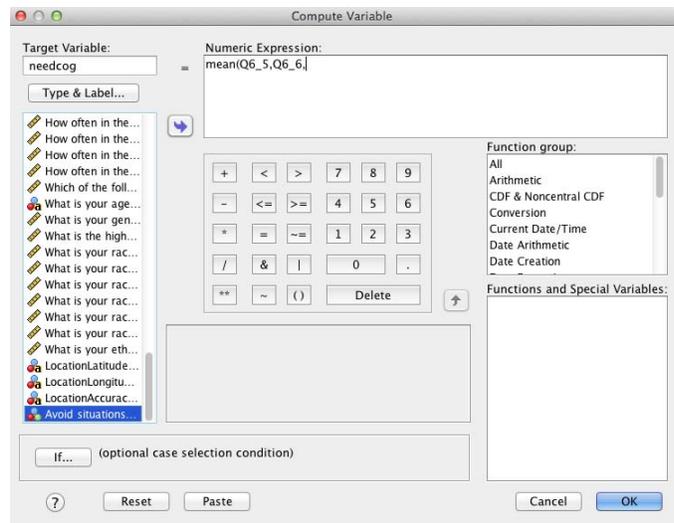
- **Convert** values from one scale to another (e.g., from kilograms to pounds)
- **Transform** values for analysis (e.g., take the natural log, or square it)
- Use other **functions** for dates or missing values
- Perform **calculations** involving several variables (e.g., calculate BMI from height and weight)
- Create an index or **composite** variable with summary statistics for a set of variables (see below)

Transform | Compute new variable



In “Compute Variable” dialogue box:

- Name the new “target variable” something short & easy to remember
- Type in or double-click on a variable on the left to include in your calculation
- Enter a **numeric expression** with variable names math symbols (you can click or type). Or, select a function from the box and provide the necessary values.



Compute: Composite Variables

It is common to have two or more variables that measure a single construct or “thing” (e.g., political knowledge, attitudes toward politics). For analysis, you often want to represent this concept with a single variable / value. This is also called a **score**, **index**, or **aggregate** variable. Follow these steps:

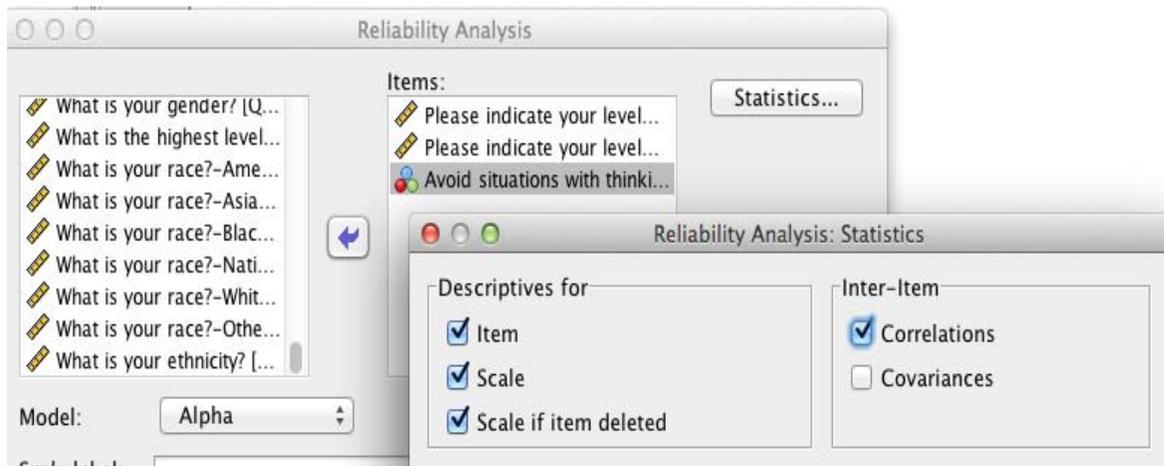
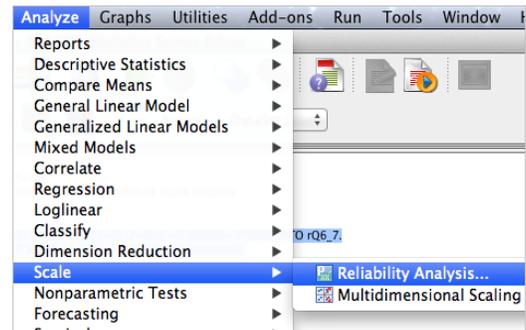
1. Make sure all variables are **coded the same way**
 - The responses must be in the same direction (e.g., higher number = more of the concept)
 - For any variables where values have the opposite meaning (e.g., higher number = less of the concept), create a new variable with Recode using Reverse Coding (see pages 3 & 4)
2. Confirm **reliability**
 - For **2 variables**: the correlation (r) is significant and at least moderately strong ($r > .30$)
 - For **3+ variables**: See “Check Reliability” on the next page.
3. Choose the **summary statistic**
 - Use **Average** if all items had the same answer choices. The composite score will be on the same scale and easier to interpret. *Numeric Expression*: **mean(var1, var2, var3)**
 - Calculate the **Total** in all other cases, or when whole numbers are useful (ex. to compare with other studies or target values). *Numeric Expression*: **sum(var1, var2, var3)**

Check Reliability

Do this to determine whether 3 or more variables “go together” and measure one concept.

Analyze | Scale | Reliability Analysis

- Enter those variables you are interested in combining
- Click **Statistics**
- Check “Descriptives for” **Item, Scale, Scale if item deleted** (so you can identify items to remove if low)
- Check “Inter-item” **Correlations**
- Click **Ok**



In the output, look in the “Reliability Statistics” box for Cronbach’s Alpha. For sufficient reliability, the alpha should be **greater than .7**.

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.761	.763	3

If you fail to meet reliability, you can also examine the **Item-Total Statistics** to see how deleting a variable might help reliability.

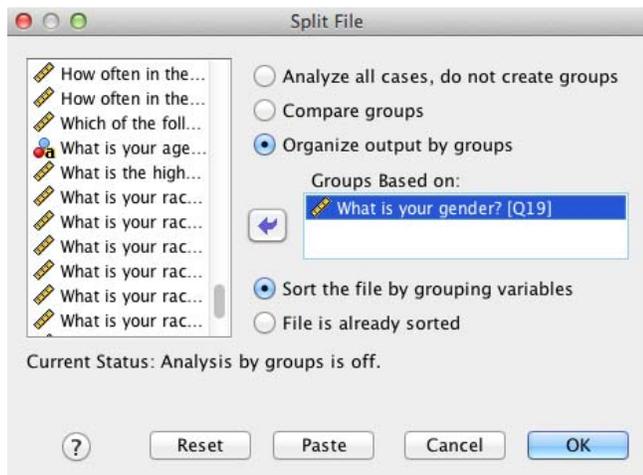
	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
Please indicate your level of agreement with each of the following statements: -I prefer complex to simple problems	9.9203	6.658	.617	.428	.653
Please indicate your level of agreement with each of the following statements: -I like to do something that challenges my thinking abilities over something that requires little thought	9.5000	7.259	.670	.465	.598
Avoid situations with thinking (reversed)	9.5652	7.985	.502	.259	.777

Compare Groups (Split File)

To compare how descriptive statistics differ across different groups (i.e., a nominal level variable)

Data | Split file

Choose **Organize output by groups** and move the variable you are interested in comparing groups within under **Groups Based on** (e.g., gender, party identification). **Compare groups** will do the same thing, but the output can sometimes be more difficult to read. Click **Paste** and run.



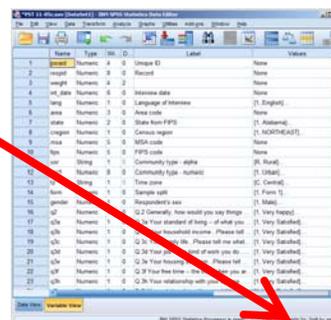
After you turn Split File on, run descriptive analyses using the other variables you are interested in (not the variable you used for Split File). You can copy or rerun the syntax from prior descriptives.

Output will produce separately for each group the data is “split” by until you turn off the split file by going back and choosing “Analyze all cases, do not create groups”.

Descriptive Statistics

What is your gender?		N	Minimum	Maximum	Mean	Std. Deviation
Male	needcog	61	2.33	7.00	5.1639	1.08802
	Valid N (listwise)	61				
Female	needcog	77	1.00	7.00	4.4892	1.38874
	Valid N (listwise)	77				

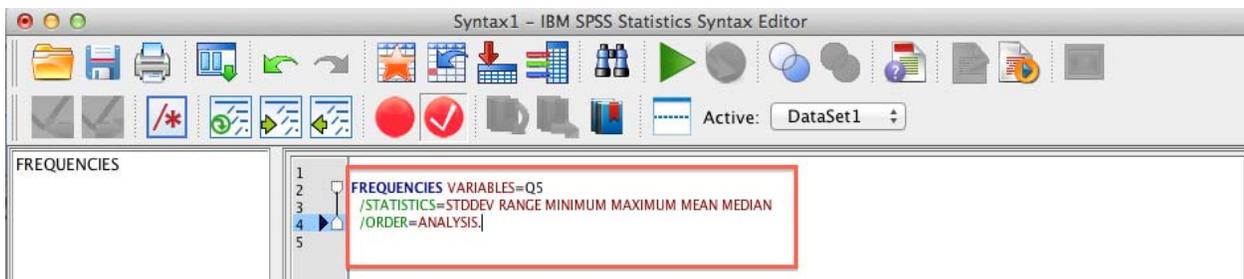
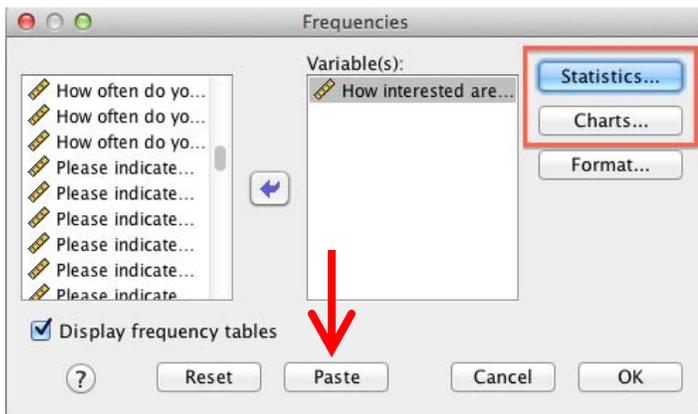
Make sure you know when your split file is on!



Descriptive Statistics

Analyze | Descriptive Statistics | Frequencies

- Select one or more variables (use Ctrl or Shift to select multiple)
- Click the “Statistics” button
- Check the appropriate central tendency and distribution statistics for that type of variable:
 - **Nominal** – mode (absolute and relative frequencies are produced automatically)
 - **Ordinal** – quintile range
 - **Interval/Ratio** – mean and standard deviation
- Click the “Charts” button as desired; choose an appropriate chart for the variable type
- Click **Paste**



When you **run** your descriptive analyses, you will be able to look at the statistics table for the measures of central tendency and dispersion as in the table on the left. The full frequency tables, on the right, has the absolute frequency (the count) and the relative frequency (“Valid Percent”).

How interested are you in politics?		
N	Valid	136
	Missing	4
Mean		4.65
Median		5.00
Std. Deviation		1.736
Range		6
Minimum		1
Maximum		7

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid				
Not at all interested	9	6.4	6.6	6.6
Fairly uninterested	16	11.4	11.8	18.4
Somewhat uninterested	7	5.0	5.1	23.5
Neutral	11	7.9	8.1	31.6
Somewhat interested	48	34.3	35.3	66.9
Fairly interested	28	20.0	20.6	87.5
Very interested	17	12.1	12.5	100.0
Total	136	97.1	100.0	
Missing	4	2.9		
System				
Total	140	100.0		