

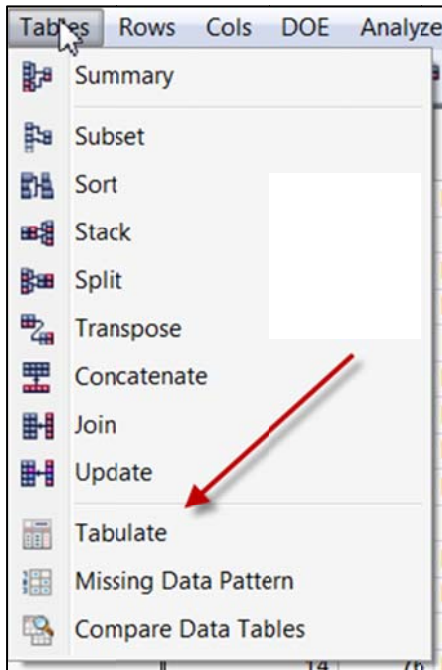
Converting individual-level data to summary data

Chong Ho Yu, Ph.D. (2013)

Azusa Pacific University

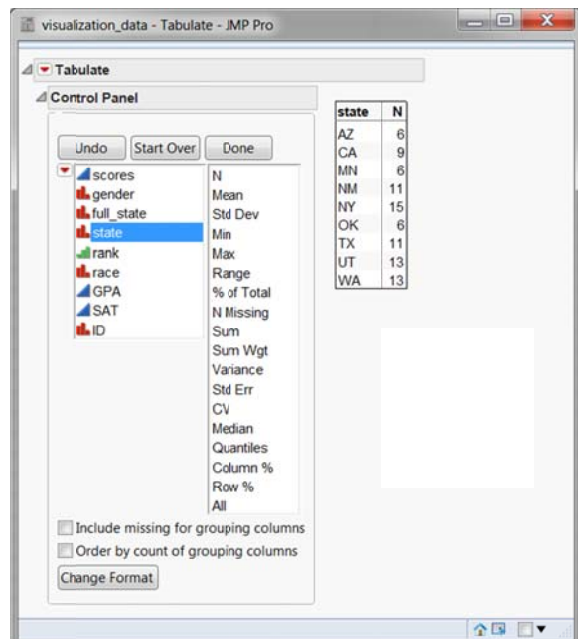
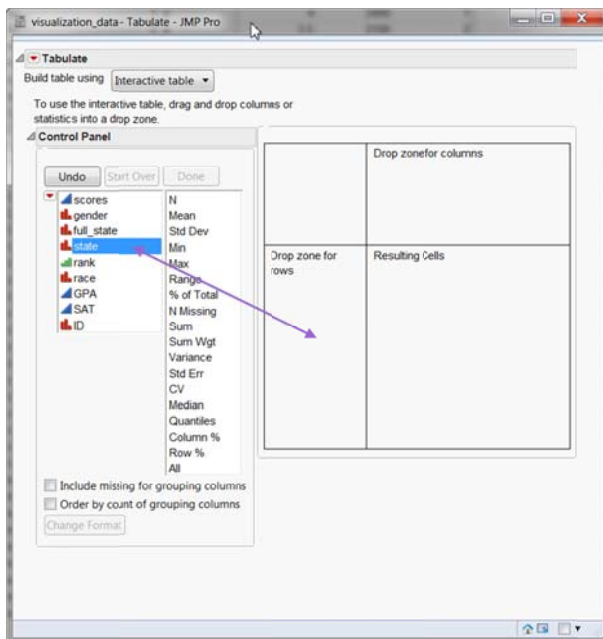
chonghoyu@gmail.com

<http://www.creative-wisdom.com/computer/sas/sas.html>

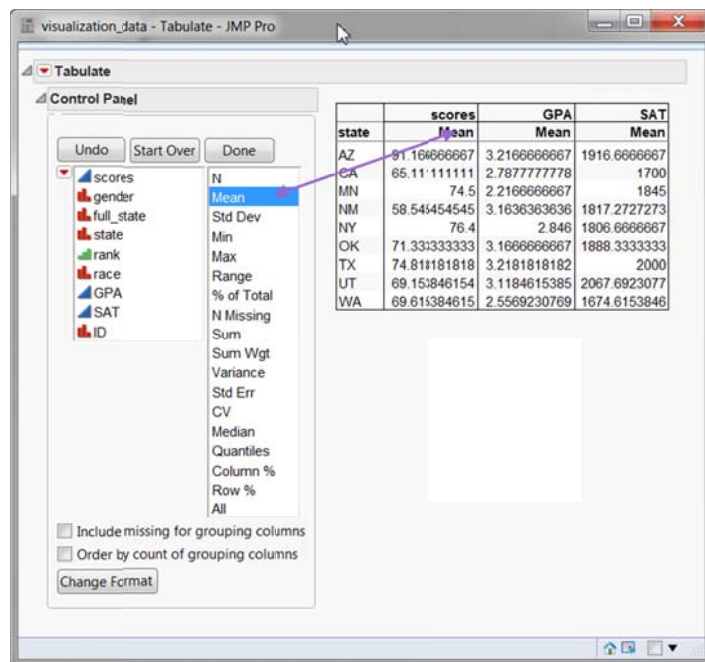
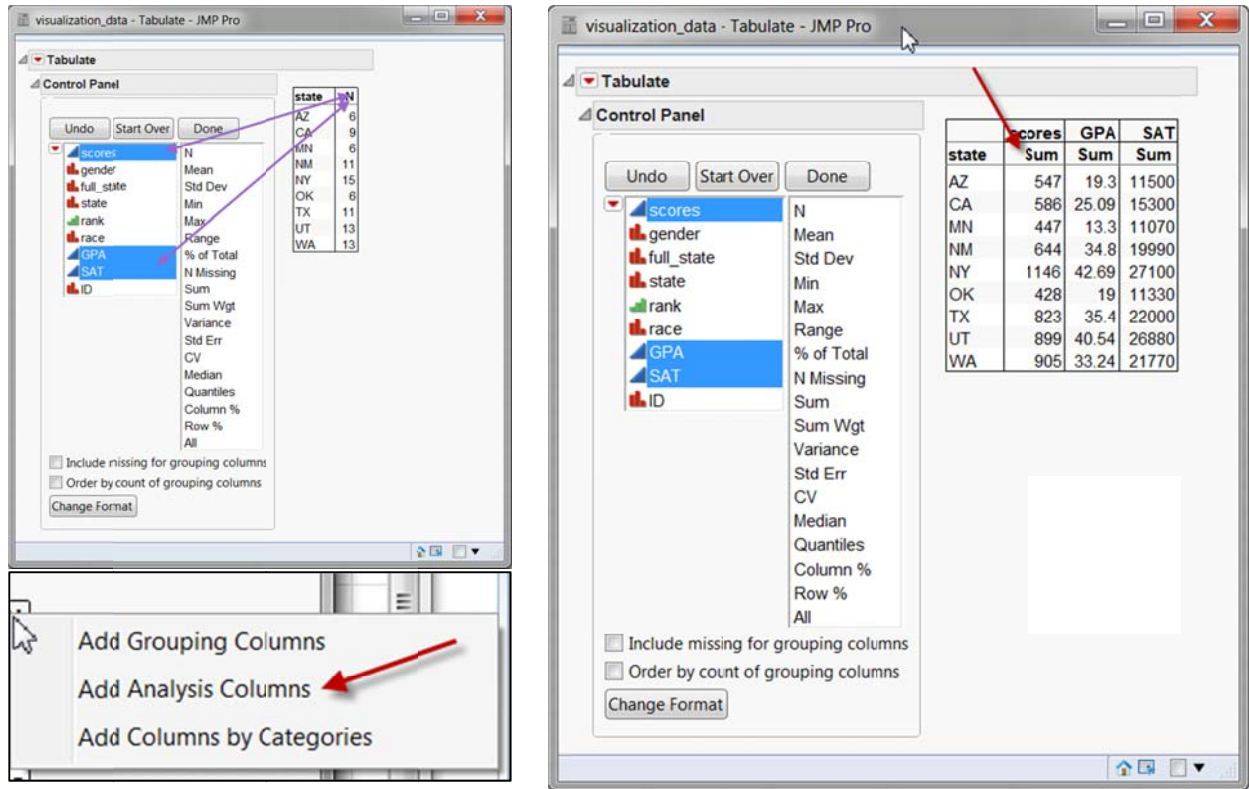


When a summary data set (e.g. by country, state...etc.) is used, the interpretation of the result is subject to the ecological fallacy. However, sometimes we have to use summary-level data. For example, the researcher may want to investigate the inter-relationships between GPA, SAT, assessment test scores, and National Assessment of Educational Progress (NAEP) scores. Even though the researcher has already collected individual-level data for each student, he has access to the NAEP data only (summary data by state). In this case, he needs to collapse student-level data into state-level data in order to match the NAEP data set.

He can make the data conversion in JMP easily. First, choose **Tabulate** from **Tables**. In the pop-up window, drag the variable "state" into **Drop zone for rows**. By default, the number of students (N) shows up in a column.

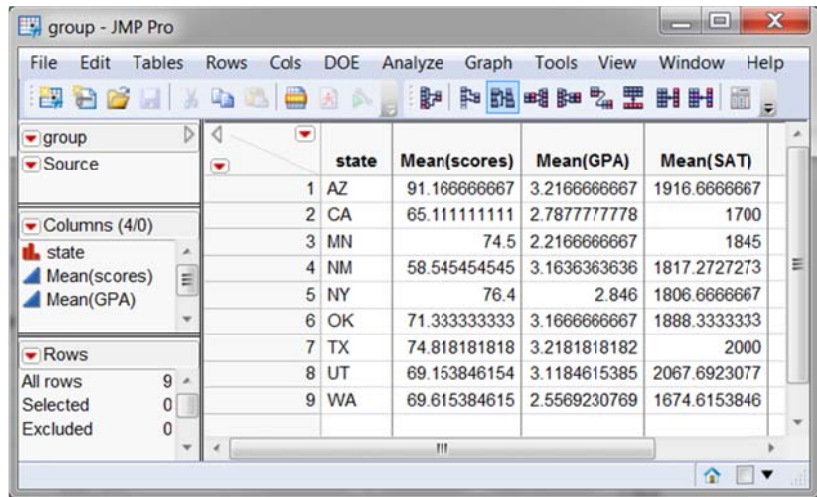
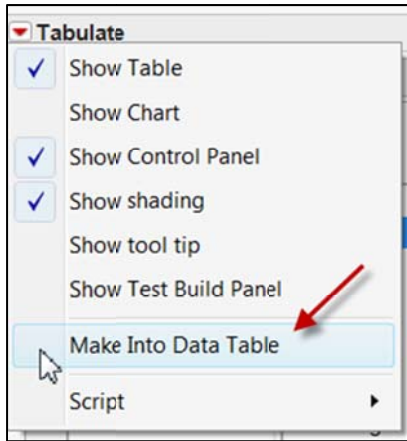


Next, drag “scores”, “GPA”, and “SAT” into the **column zone**, which is occupied by N in this example. In the drop down menu, choose **Add Analysis Columns**. By default, the sum scores by state show up in three columns. But this is not what the researcher wants.

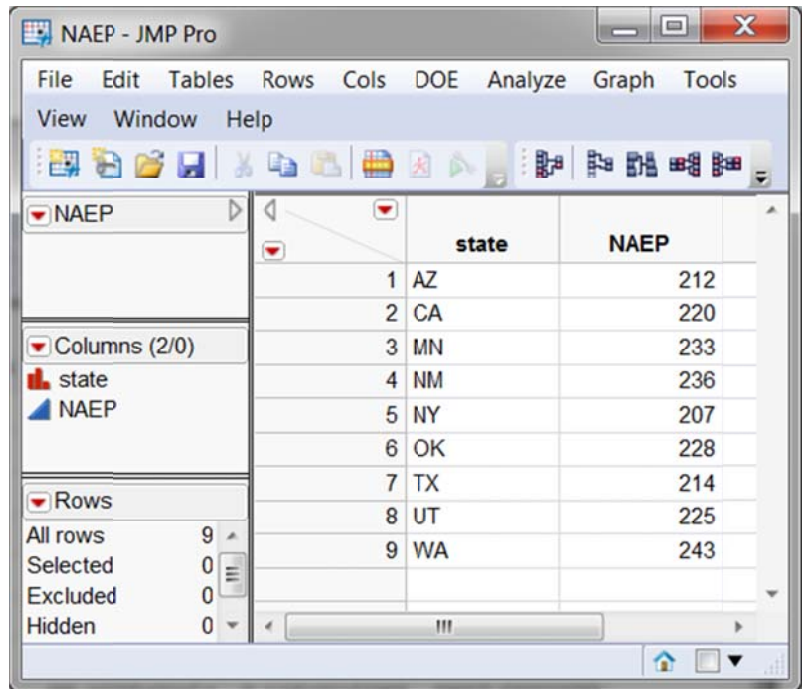
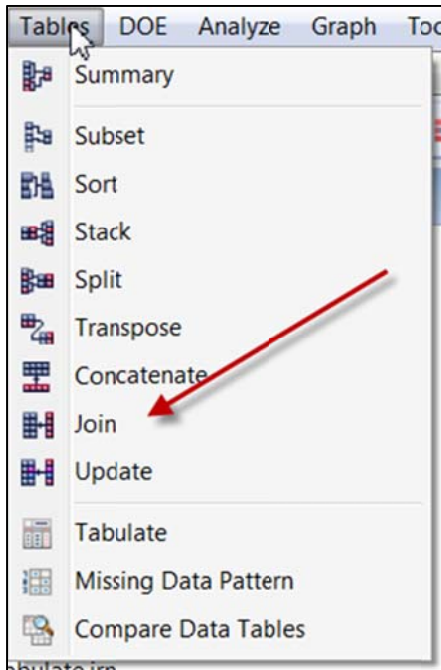


To obtain the mean score of all the column variables, drag the **Mean** from the left panel into the **Mean** in the **column zone**. Now the sum scores are replaced by the mean scores.

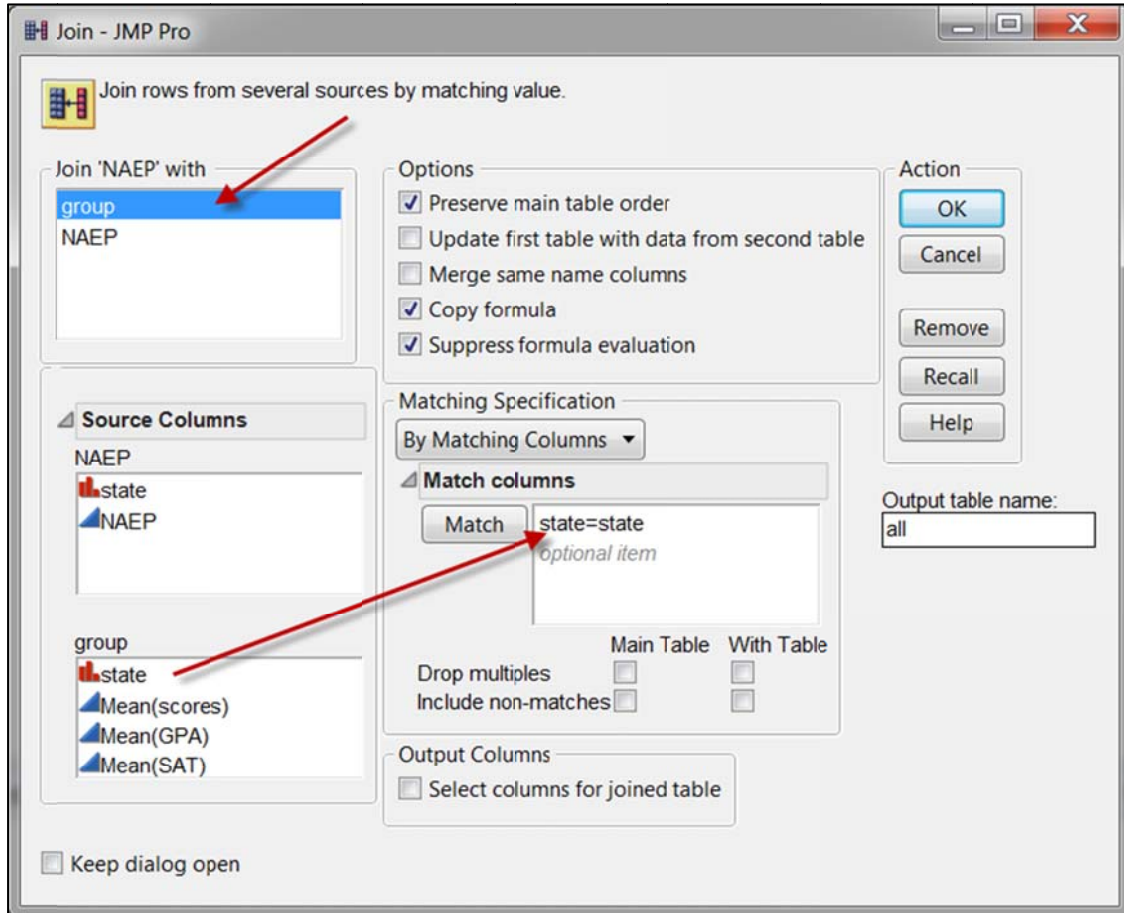
Afterwards, choose **Make Into Data Table** from the red triangle, and the summary data set is ready to go.



The right panel below is the NAEP data set. To merge the NAEP data set with the newly created data set, choose **Join** from **Tables**.



Drag “state” into **Match** because both data sets have the state ID in common. You can merge the data into one of the original tables. But it is a good idea to keep the existing data intact, just in case any error happens. To join the two tables into a new data set, type a new name in **Output table name**. In this example the new data set is called “all”. Now the research has what he needs for state-level analysis.



	state of NAEP	NAEP	state of group	Mean(scores)	Mean(GPA)	Mean(SAT)
1	AZ	212	AZ	91.166666667	3.216666667	1916.6666667
2	CA	220	CA	65.111111111	2.787777778	1700
3	MN	233	MN	74.5	2.216666667	1845
4	NM	236	NM	58.545454545	3.163636363	1817.2727273
5	NY	207	NY	76.4	2.846	1806.6666667
6	OK	228	OK	71.333333333	3.166666667	1888.3333333
7	TX	214	TX	74.818181818	3.218181818	2000
8	UT	225	UT	69.153846154	3.118461538	2067.6923077
9	WA	243	WA	69.615384615	2.556923077	1674.6153846