

Econometric Analysis – Dr. Sobel

Econometrics Session 1:

1. Building a data set

Which software - usually best to use Microsoft Excel (XLS format) but CSV is also okay

Variable names (first row only, 15 character max in gretl software we will be using, 1st must be letter, no spaces but _ symbol is ok)

First column is sometimes dates or labels (state names) and it is okay if this has spaces, but give it a title too

EXAMPLE DATA SET IN EXCEL:

The screenshot shows an Excel spreadsheet with the following data:

Country	Entrepreneurship	GDP_PerCapita	GovSpend_PerCap	Percent Male	Median_Age
Australia	6.68	24000	4344.548191	49.94	36
Belgium	2.99	26100	10332.63878	49.06	40
Canada	8.82	27700	5108.758149	49.54	37.8
Denmark	6.53	28000	9583.742386	49.45	39.1
Finland	4.56	25800	5989.431937	48.76	40.3
France	3.2	25400	4030.143661	48.58	38.3
Germany	5.16	26200	9936.223177	49.10	41.3
Hungary	6.64	12000	1424.893704	47.70	38.4
Iceland	11.32	30200	11874.51872	50.71	34
Ireland	9.14	27300	7029.715911	49.48	33.1
Italy	5.9	24300	8963.272687	48.53	41
Japan	1.81	27200	5663.72633	48.89	42
Mexico	12.4	9000	1374.176867	48.63	23.8
Netherlands	4.62	25800	8384.709494	49.47	38.6
New Zealand	14.01	10500	1295.922033	48.89	33.1
Norway	8.69	30800	12790.22259	49.55	37.7
Poland	4.44	8800	1353.732959	48.58	36
South Korea	14.52	19600	1997.730061	50.35	33.2
Spain	4.59	18900	2722.414047	48.90	38.7
Sweden	4	24700	12394.29218	49.63	40.1
Switzerland	7.13	31100	4119.02669	49.63	40.2
United Kingdom	5.37	24700	9053.143461	49.14	38.4
United States	10.51	36300	6124.602077	48.91	35.8

Variable names must be 15 letters or less, begin with a letter, and cannot have any spaces (but underscores " _ " are okay). They must be on the first row, with data starting on the second row. A column of names that label the data is okay and it doesn't matter much what's in it, and it is okay to have spaces in it.

The number of "observations" (data points) matters. More is better. Always report the number of observations you use in a research paper.

Missing observations – usually best to have none, but most programs can deal with them – leave cell blank

Types of datasets: cross section, time series, panel (usually must specify, and label)

– in gretl software we will be using, go to Data menu, Dataset structure to change or specify

Source information – how to document it

Descriptions: what is included, what year, how measured, etc.

Keep track of units of measurement (data in millions of dollars? Percentages: would 5% be as 0.05 or 5.0?)

EXAMPLE OF HOW TO DOCUMENT YOUR DATA IN YOUR RESEARCH PAPER:

Appendix 1: List of Variables and Sources

Table 5

Variable name	Minimum	Maximum	Average	Obs.	Source	Definition
Dependent variables						
Average business start-up rate	0.114	0.229	0.145	50	a	Average business births as a percentage of total businesses with 1–9 employees (2002–2003)
Net business creation rate	0.001	0.052	0.018	50	a	Average net business creation as a percentage of total businesses with 1–9 employees (2002–2003)
Venture capital per capita	0	379.39	46.635	48	b	Venture capital invested (by destination) per capita (2005)
Patents per capita	0.044	1.086	0.249	50	c	Number of patents granted (all types) per 1,000 people (2005)
Productive entrepreneurship	3.2	43	23.583	48	d	Index of entrepreneurial activity from Sobel (2008) based on several measures of entrepreneurship
Independent variables						
Cultural diversity	0.021	0.44	0.132	50	e	Probability that 2 randomly chosen individuals from a state were born in different countries, calculated like a standard Herfindahl–Hirshman Index (2000)
Median age	27.1	37.9	35.536	50	f	Median age of state population (2000)
Percent male	88.8	107.6	94.114	50	f	Percent of 18 years and over state population that is male (2000)
Percent college degree	14.8	33.2	23.776	50	f	Percent of state population with a bachelor's degree or higher (2000)
Population density	0.0011	1.1344	0.1819	50	f	Population density per square mile of land area in state (2000), in thousands
Median household income	29.696	55.146	41.371	50	e	Median household income (1999), in thousands
Economic freedom	5.6	7.8	6.684	50	f	Economic freedom of North America (EFNA) index (2005)

Data Sources: *a* Company Statistics Division of the United States Census Bureau (<http://www.census.gov/csd/susb/susb.htm>), *b* Venture Capital Association's (NVCA) Yearbook, *c* the US Patent and Trademark Office (USPTO), *d* Index constructed in Sobel (2008), *e* US Census 2000 (<http://www.census.gov>), *f* Fraser Institute (Karabegovic and McMahon 2005)

Transformations (getting units right, like per capita, logs, changes, percent changes)

- generally try to not have the units be to different (billions for one variable, hundreds for another)
- use per capita to adjust for differences due to like the size of states or changes thru time
 - or sometimes per something else (school spending per student)
- sometimes the effects are “percentage” based (e.g., a 10% weight loss) – use logs or percent changes
- sometimes when we use “time series” data we do the year to year changes as the variables

2. Reading the data into “econometric” software that can run “regressions”

Everyone must obtain or use software called “gretl”, downloadable at <http://gretl.sourceforge.net/>

- will also be available on computers in Beaty 120 computer lab and Beatty Center Atrium
- if you know another program (like EViews) you are welcome to use it, but I can't help much

DOWNLOAD SOFTWARE FROM GRETL WEBSITE EXAMPLE (homepage):



URL: gretl.sourceforge.net

Spanish Italiano Português Türkçe Polish site

Gnu Regression, Econometrics and Time-series Library

gretl

Download

gretl for Windows

gretl on Mac OS X

gretl manual

Do you use gretl?

Contributions to gretl

gretl community

gretl and X-12-ARIMA

gretl and TRAMO/SEATS

gretl and GNU R

More screenshots

Numerical accuracy

Project page

Files listing

Change log

Latest source via CVS

Libgretl API docs

Mailbox lists

gretl wiki

Acknowledgements

Links to download the program are here

Is a cross-platform software package for econometric analysis, written in the C programming language. It is free, open-source software. You may redistribute it and/or modify it under the terms of the GNU General Public License (GPL) as published by the [Free Software Foundation](#).

Warning: gretl on Ubuntu 11.04

The default gretl package for Ubuntu 11.04 (natty) is broken; it should be replaced with gretl version 1.9.5 (or higher), available via packages.debian.org.

gretl conference 2011

This took place at Nicolaus Copernicus University, Toruń, Poland on 16-17 June 2011; [details here](#).

Features

- Easy intuitive interface (now in French, Italian, Spanish, Polish, German, Basque, Portuguese, Russian, Turkish, Czech, Traditional Chinese, Albanian and Greek as well as English)
- A wide variety of estimators: least squares, maximum likelihood, GMM; single-equation and system methods
- Time series methods: ARMA, GARCH, VARs and VECMs, unit-root and cointegration tests, etc.
- Limited dependent variables: logit, probit, tobit, interval regression, models for count and duration data, etc.
- Output models as LaTeX files, in tabular or equation format
- Integrated scripting language: enter commands either via the gui or via script
- Command loop structure for Monte Carlo simulations and iterative estimation procedures
- GUI controller for fine-tuning Gnuplot graphs
- Links to [GNU R](#), [GNU Octave](#) and [Ox](#) for further data analysis

Data formats

Supported formats include: own XML data files; Comma Separated Values; Excel, Gnumeric and Open Document worksheets; Stata .dta files; SPSS .sav files; Eviews workfiles; Multi data files; own format binary databases (allowing mixed data frequencies and series lengths), RATS 4 databases and PC-Give databases. Includes a sample US macro database. See also the [gretl data page](#).

Download

gretl is available for both Windows and Mac, and can even be installed on computers on which you do not have administrative rights. There are even older versions available for older operating systems.

DOWNLOAD SOFTWARE FROM GRETL WEBSITE EXAMPLE (Windows download page):

System requirements

As of version 1.9.4, gretl requires Windows XP or higher and a processor that supports the SSE2 instruction set. SSE2 support is older systems, [see below](#).

Downloads

If you have the rights of a "power-user" or better on Windows, choose a self-installer from the first column below; just download

If you have no administrator rights on Windows choose a zip archive from the second column; unzip this in any location where you called "enable folders". The whole archive is in a directory called gretl. For example, if you unzip the archive into a directory name

The current "snapshot" of gretl is more up to date than the release: often it will contain bug-fixes but sometimes it will contain ne

	<i>self-installer</i>	OR	<i>zip archive (no admin rights)</i>
latest release (Dec 22, 2011)	gretl-1.9.7.exe		gretl-1.9.7-win32.zip
OR current snapshot	gretl-install.exe		gretl-win32.zip

The executables were cross-compiled under GNU/Linux using [mingw32](#) and [GTK for Windows](#) (thanks Tor Lillqvist!). The free insta

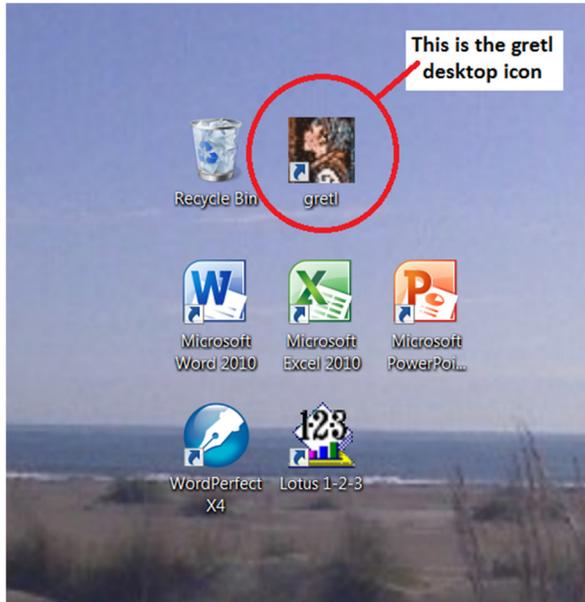
Optional extras you may wish to install

- X-12-ARIMA (seasonal adjustment, ARIMA)
- TRAMO/SEATS (seasonal adjustment, ARII)
- Datasets for Wooldridge, [Introductory Ec](#)
- Datasets for Gujarati, [Basic Econometrics](#)
- Datasets + scripts for Stock and Watson,
- Datasets for Davidson and MacKinnon, [Ec](#)
- Datasets for Marno Verbeek's [Guide to Mc](#)

choose to download “latest release” and use the one for “self-installer” if you do have administrative rights do not worry about downloading any of the “optional extras”

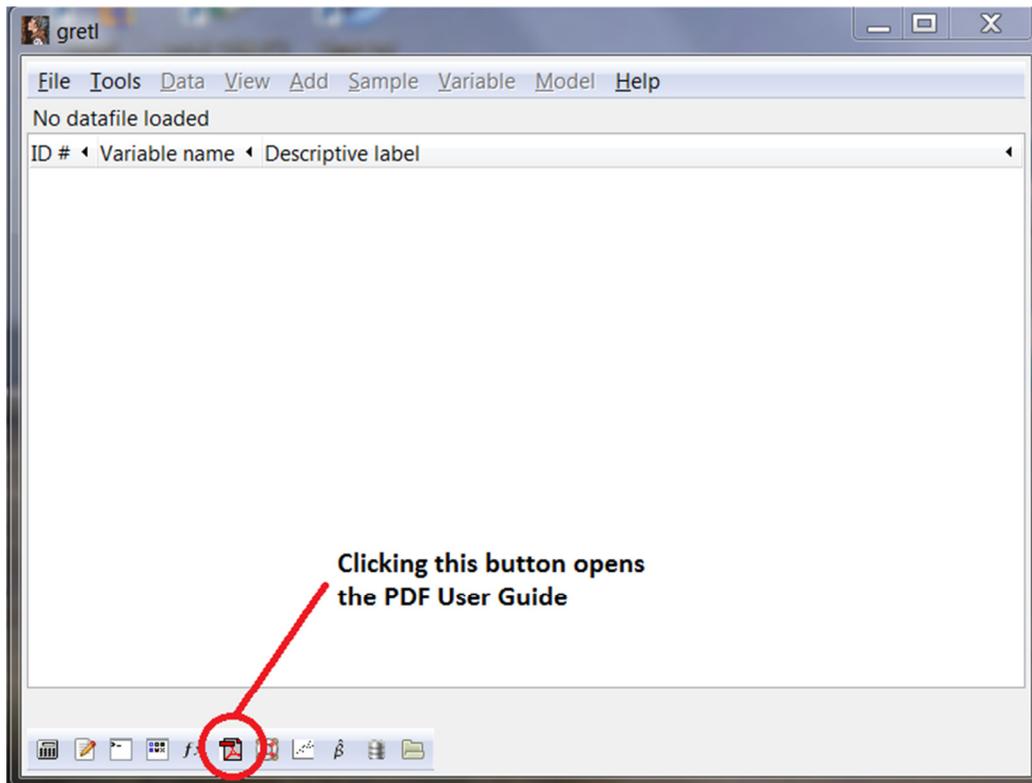
Once installed, open gretl by clicking on the gretl icon on your desktop:

GRETL DESKTOP ICON EXAMPLE:



When you open gretl you will get a main page that looks like this below:

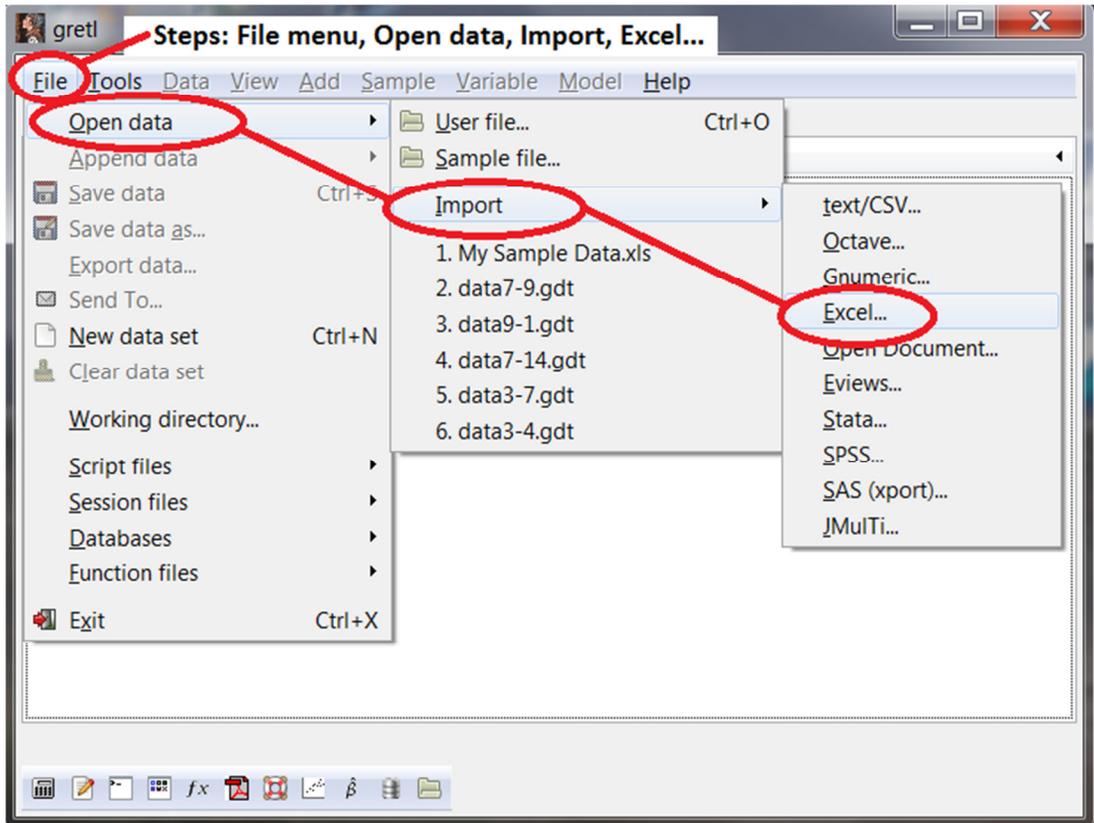
GRETL MAIN OPENING SCREEN EXAMPLE:



gretl has a built in user manual if you ever have questions or problems (PDF "users guide" button along bottom)

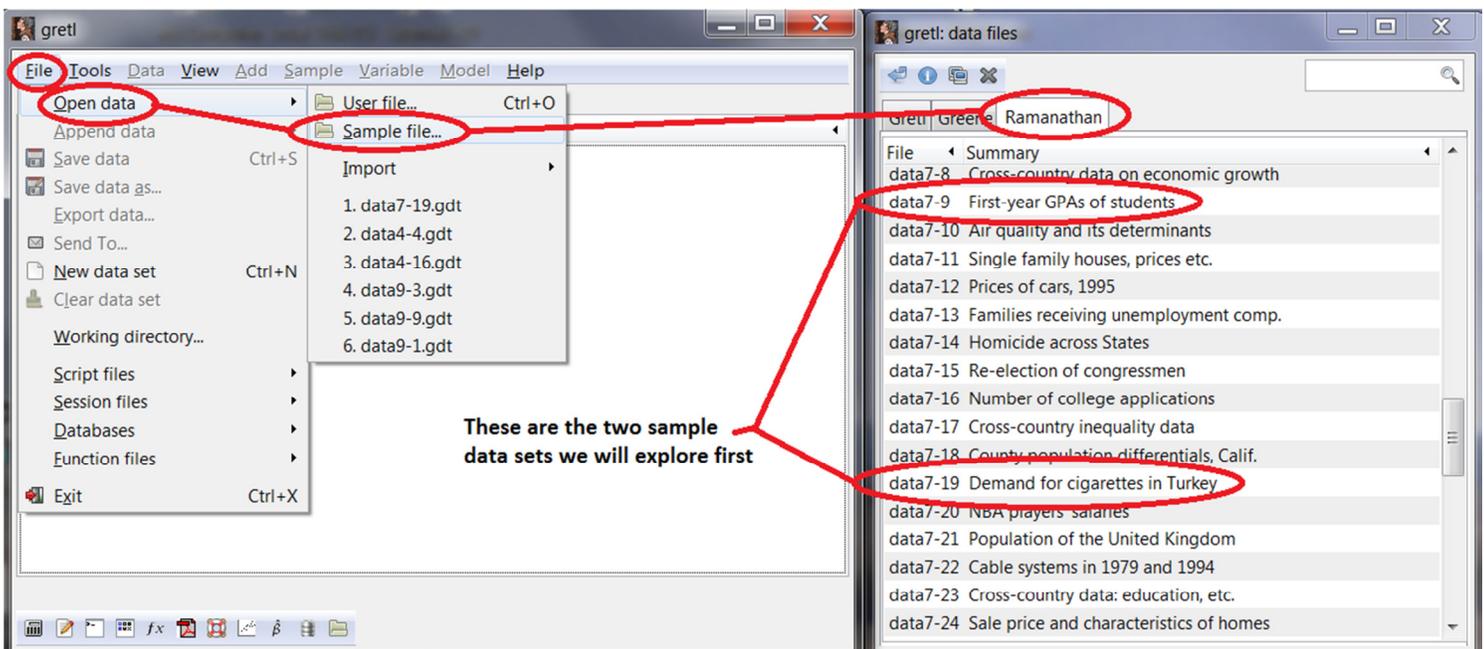
Reading data into gretl (File menu, Open Data, Import, then Excel if it's your data)

GRETl IMPORTING EXCEL DATA EXAMPLE:



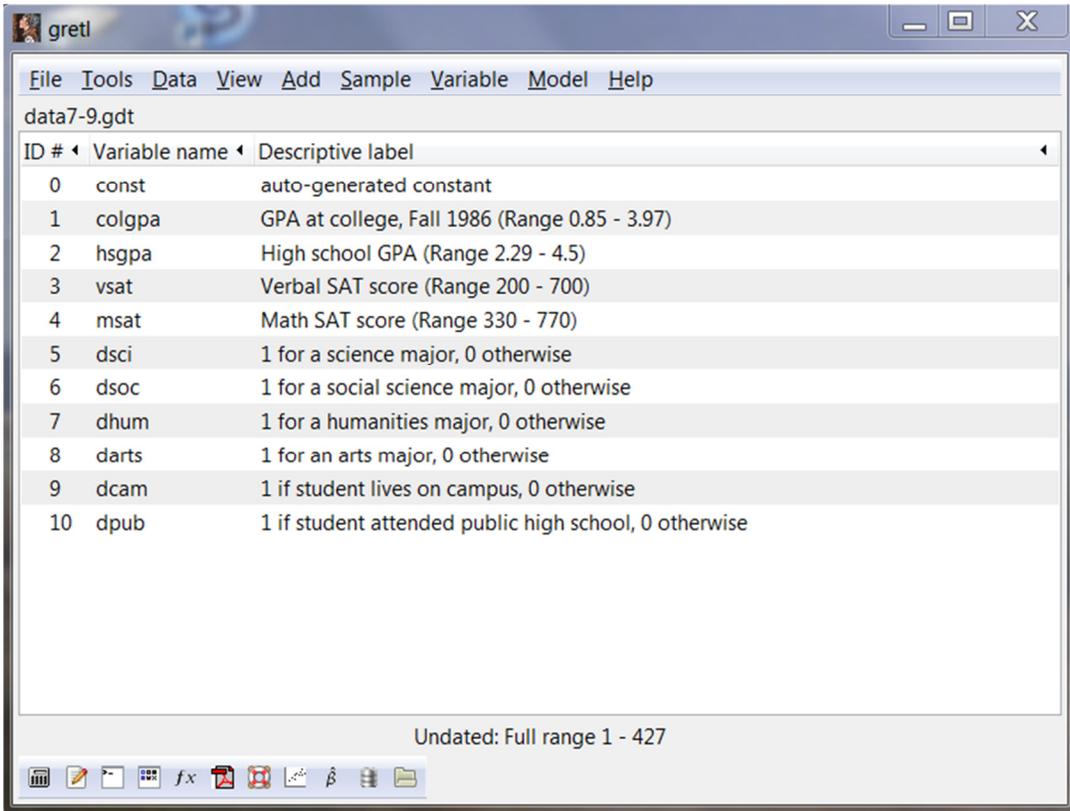
- For now, use sample dataset (File menu, Open Data, Sample file): Ramanathan data7-9 “First year GPAs of students”, we may also use data7-19 “Demand for cigarettes in Turkey” but we will start with the GPA data

GRETl OPENING RAMANATHAN GPA SAMPLE DATA EXAMPLE:



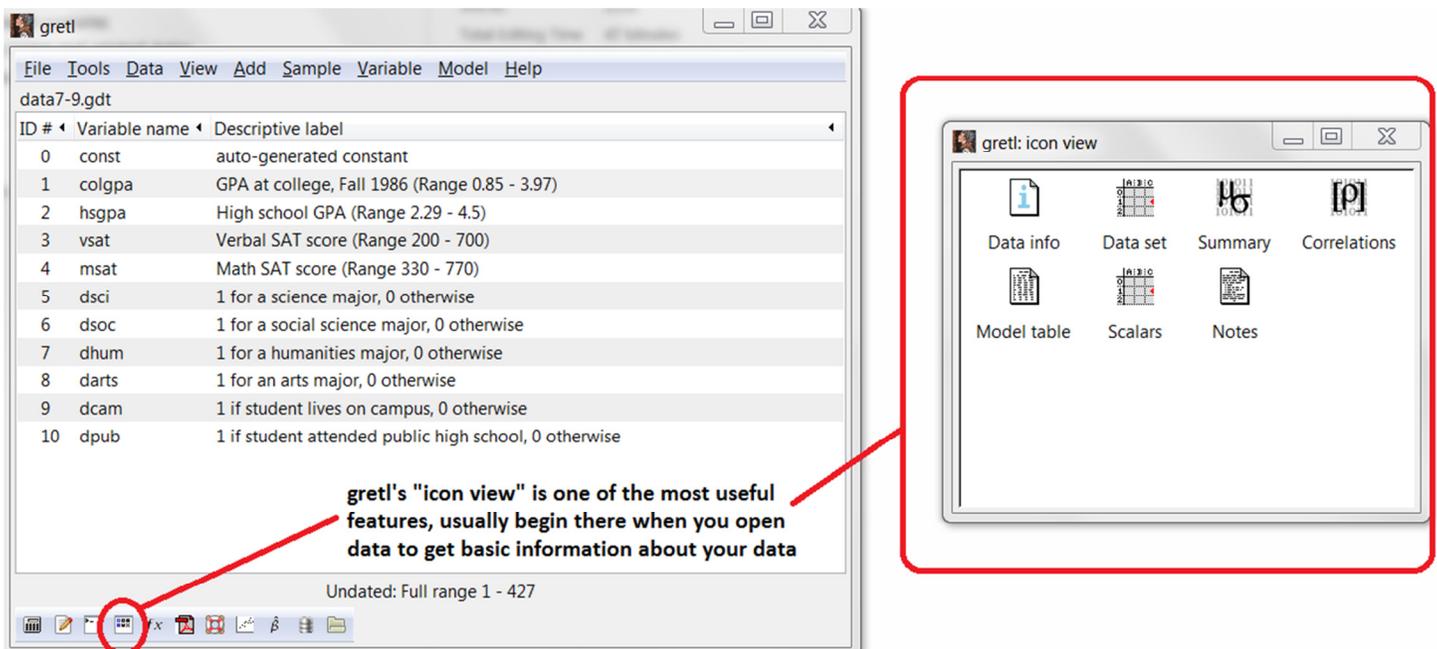
once you open the data, the main screen will look like this below:

GRET MAIN SCREEN AFTER OPENING SAMPLE GPA DATA:



one of the most used features is the "icon view", so let's open it. From this window you can look at your data and get basic statistical information about your data:

OPENING GRET'S ICON VIEW:

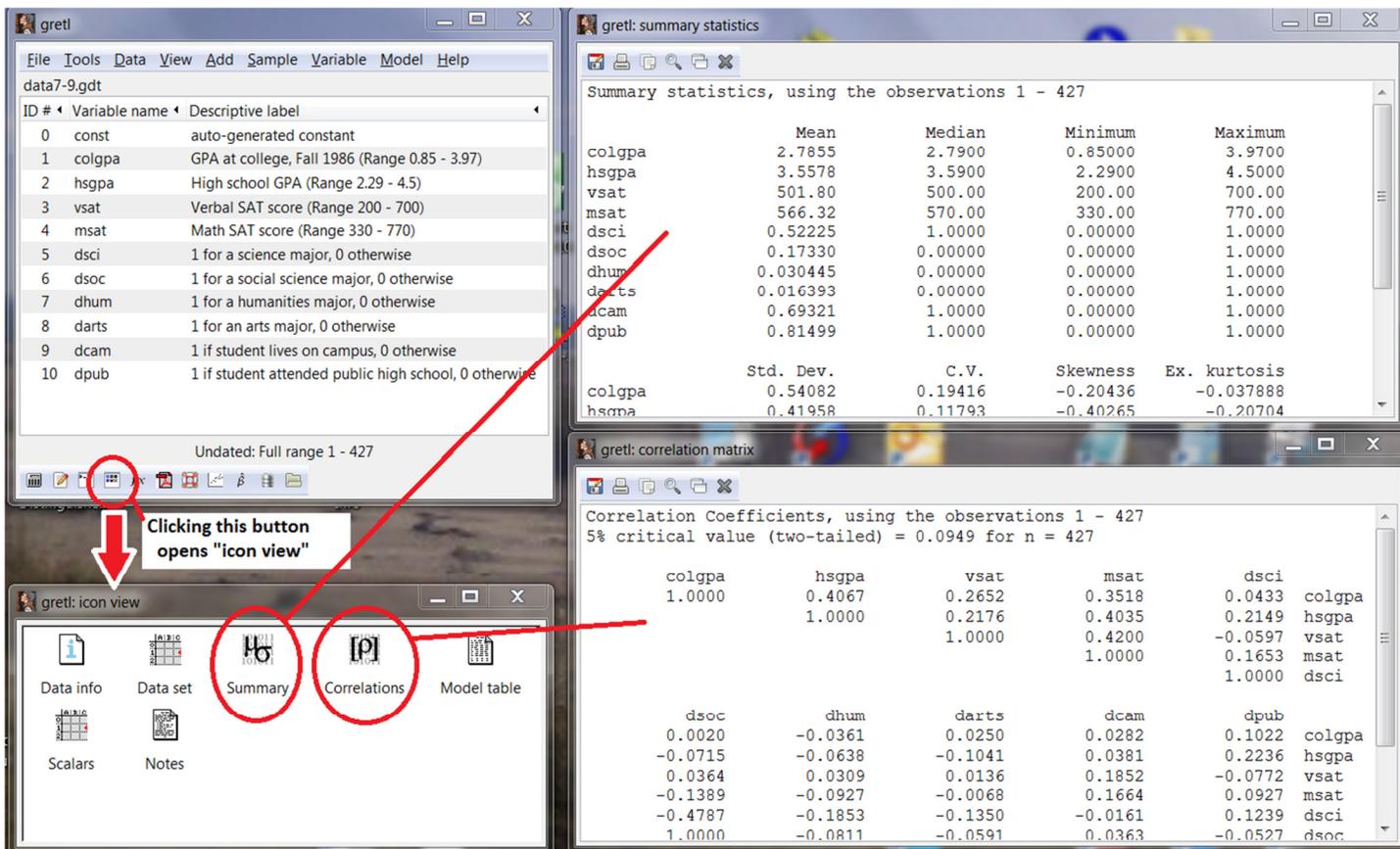


3. First steps: Examine basic descriptive statistics, and examine correlations, creating new variables
 Note: these can be done in Excel too but are easier in gretl [Excel commands shown in brackets]

Examine “descriptive statistics” such as the mean (average), maximum, minimum of each variable

- in gretl open “session icon view” and click “Summary” (or can use “View” menu then “Summary Statistics”)
- [in Excel for data in cells A1 to A20 the commands are: =AVERAGE(A1:A20); =MAX(A1:A20); =MIN (A1:A20)]

USING GRETL'S ICON VIEW TO GET SUMMARY STATISTICS AND CORRELATIONS:



Examine “correlation coefficients”

- in gretl either use “View” menu then “Correlation Matrix” or open “session icon view” and click “Correlations”
- [in Excel for data in cells A1 to A20 and B1 to B20 the command is: =CORREL(A1:A20,B1:B20)]

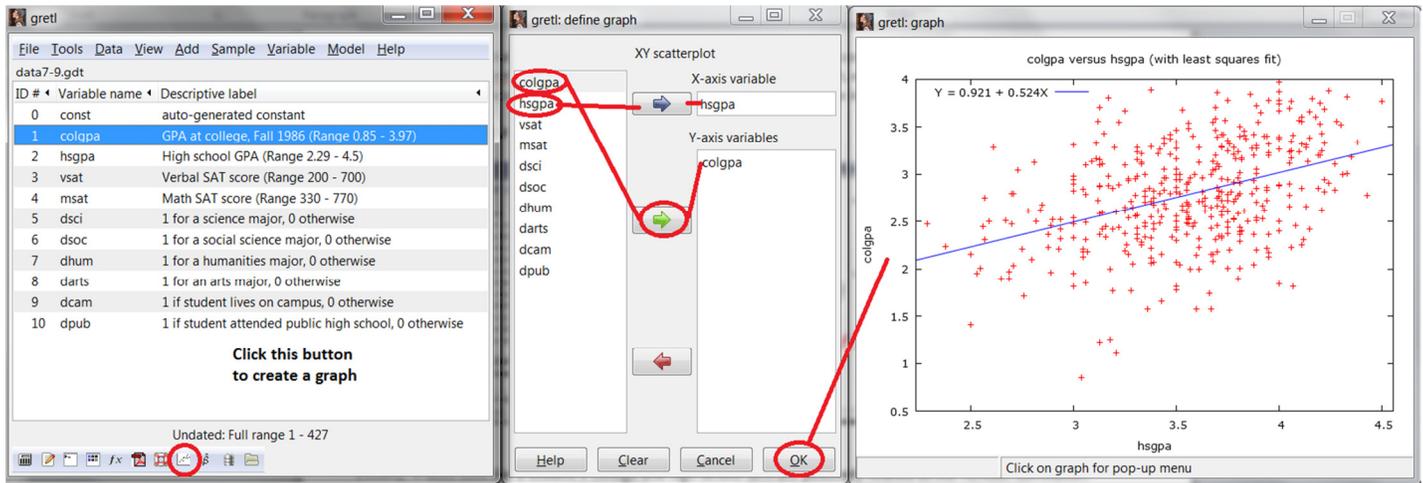
We will use these later, but for now just know the correlation coefficients measure whether two variables are positively or negatively related, and how strongly on a scale of zero to one. It shows the correlations between all of your variables, but most of the time you will only be interested in one or a few.

Example: what is the correlation between a student’s college GPA (colgpa) and highschool GPA (hsgpa)? +0.4067

Graph key variables against each other as XY plot – examine visual correlations

- in gretl click X-Y graph icon along bottom to create a chart, and choose the variables
- in Excel you have to create a “Scatter” type chart from the insert menu, after highlighting the data

CREATING A GRAPH OF TWO VARIABLES IN GRETL:

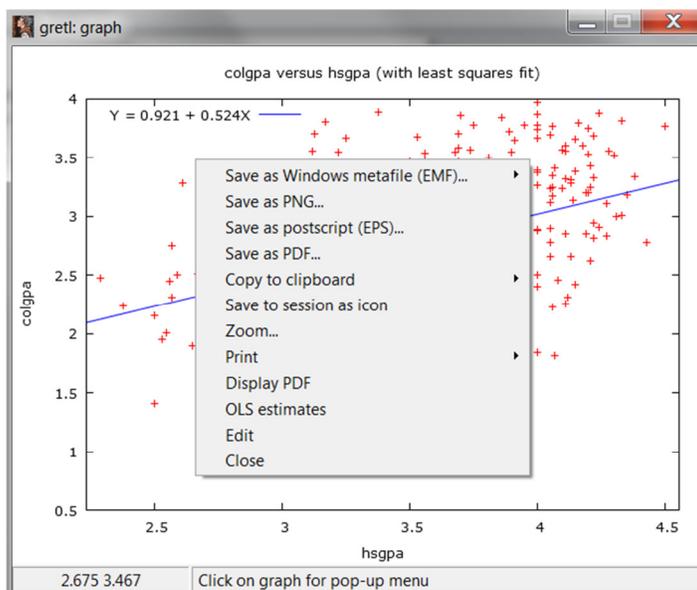


Visually, it does look like a student’s college and high school GPA are positively related as the +0.4067 correlation coefficient suggested.

Showing trend lines or “least squares fit” (which is an Ordinary Least Squares regression we are about to perform) in the graph is helpful to show you the relationship. The blue line in the graph above is a trend line. gretl usually adds a trend line automatically (in Excel you have right click any one data point in the scatter chart you create and choose “Add trendline”) In gretl it also gives you the equation of the trend line at the top of the chart.

Saving your graph: in gretl, click anywhere on the graph, can save as PDF, copy to clipboard (to say paste in your paper), and you can save it so you can reopen it later by choosing “Save to session as icon”

RIGHT CLICK IN THE GRAPH ANYWHERE TO BRING UP A MENU OF OPTIONS TO SAVE OR PRINT YOUR GRAPH:

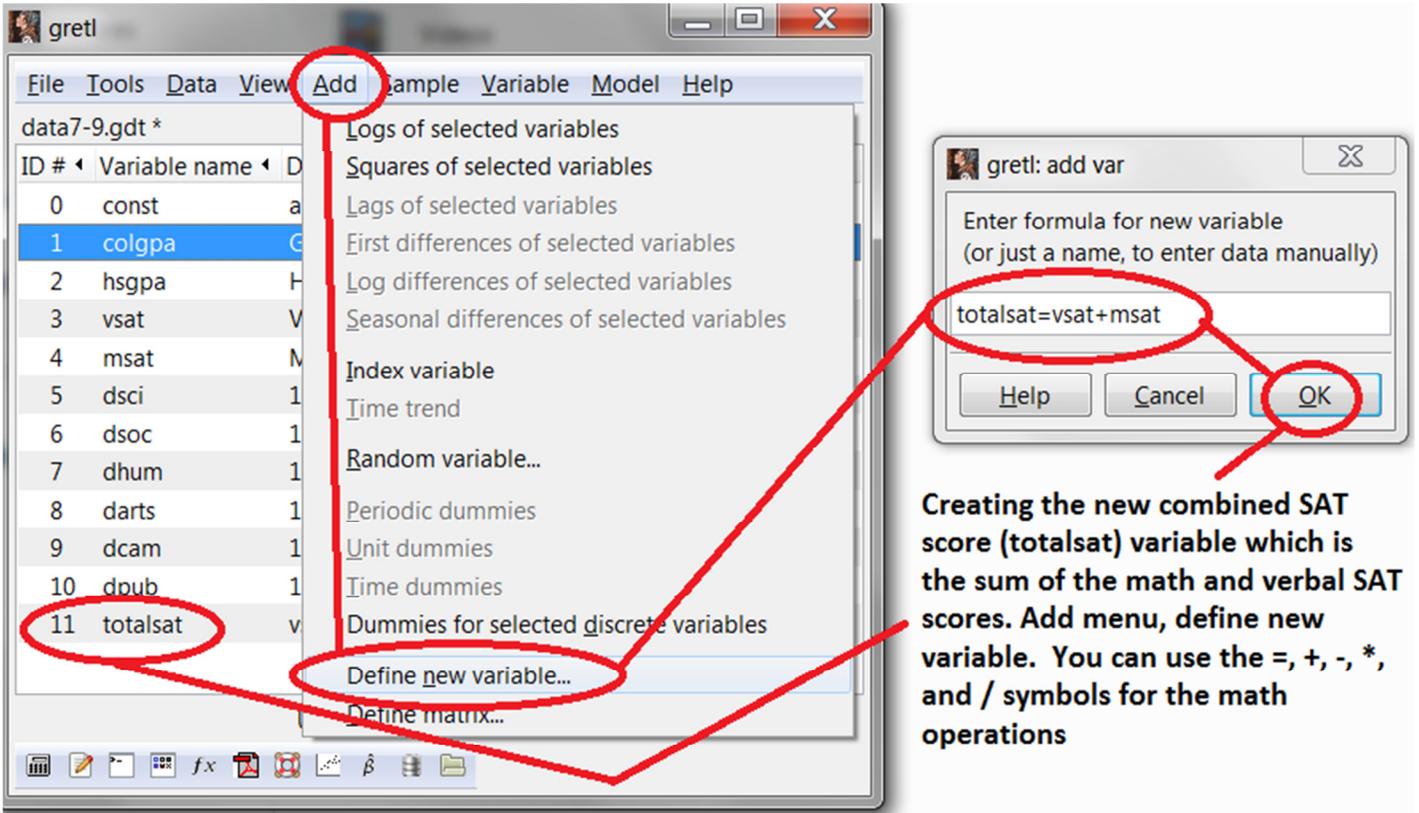


Creating a new variable in gretl from your existing variables

- Add menu, Define new variable, do as equation (e.g., $Z=X+Y$ if X and Y are in your data and Z is new)

In our sample data, there is both the student's verbal SAT (vsat) and math SAT (msat) score. But there is no variable for the student's combined math plus verbal score. So let's create a new variable named "totalsat" for that.

CREATING A NEW VARIABLE IN GRETL (COMBINED SAT SCORE EXAMPLE):



gretl can also create "logs" or "squared" versions of your variables automatically, these are the first two entries in the Add menu shown above where we picked "Define new variable".

gretl can also create random variables for you.