Extracting datasets from Compustat to analyze in Stata or Excel

The Compustat database is accessed via client software called "Research Insight." This software should be installed on all Windows-based classroom and lab computers on campus.

To use Research Insight:

1. Find it under Start→All Programs→Research Insight→Research Insight. (If it tells you about updated support files and asks if you want to copy them, say yes and accept the default location.)
2. The first time you open Research Insight on a computer, it will ask you for the location of the North American database. Type in \\STORAGE\PROJECTS\ECN\Compustat\NA.
    a. If you want to use the GlobalVantage database, which has data for non-North American companies, go to Databases→Default Database and switch to that database. When it asks you for the location of that database, type in \\STORAGE\PROJECTS\ECN\Compustat\GL. Skip step 3(b) below.
3. Go to Tools→Options to set the time period and companies available for searching.
    a. Under "Period Reference," set "Annual" to 34. This will give you data for the past 34 years (back to 1970), which is all of the data available.
    b. Under "Screening," set the Default Base Set Formula to $C+$R. This gives you currently active U.S. companies (C) plus companies that are now out of business (R). Depending on your needs, you may also want to add $T (Canadian companies).
4. Select "Report Assistant" from the options that appear on the screen, or click the second button in the left sidebar, to open the Report Assistant.
5. Select "Items vs. Time for a Company" under "Historical Analysis," and click "Next."
6. Use the "Look Up" button in the upper right to find the variables that you want to include in your dataset. You can start typing in the box under "Select Items" to see variables that include the words you typed, or you can browse by category on the left. As you find variables, click "Paste" to add them to your selection. When you have all of the variables you want, close the Look Up window and click "Next."
    a. Always include at least one company identifier—name, ticker symbol, Global Vantage Company Key, etc.—in your selections.
7. Select the time periods you want to include in your report, either relative years/quarters/months (number of years before now, e.g., -14Y to 0Y to get data from 14 years ago to this year) or absolute years/quarters/months (e.g., 2000-2014).
8. Click "Finish."
9. A window will pop up asking you to select companies. Click "Cancel" to close the window. You need to modify the report before you can select companies.
10. Click once on the dark green box near the top containing your preferred identifier (ticker symbol, GV Key, etc.), then right-click on the box and select "Copy." Make sure you select the dark green box with the mnemonic for your identifier (e.g., GVKEY, TIC) rather than the light blue box with the label (e.g., "Global Value Key:" "TICKER:").
11. Right-click on an empty, white part of the screen and select "Paste." Your cursor will now show a little glue jar. Click on the screen to paste the identifier at the end of each row of data. Don't

worry if they don't line up exactly; as long as they're all the last field in the row, they'll all be in the same column when you open the file in Excel or Stata. If you accidentally paste two identifiers in the same row, or paste an identifier somewhere you don't want it, click on it to select it, press the Delete key on the keyboard, and click "OK" to get rid of it.

12. Click on each of the boxes above your data columns and press "Delete" to delete them. Click "OK" when asked if you want to delete this report object. Make sure you also delete the labels at the top of each column! (You'll put the column labels back in a later step.)
13. You may want to save your report layout at this time. Use File→Save.
14. Click on the running man icon at the top to run your report. A window will pop up asking you to select the companies you want to appear in your dataset.
    a. If you're looking for a pre-designated group of companies (e.g., the S&P 500), you can select them by clicking "Change Set" in the upper right, then clicking the "Look Up" button, selecting the group you want, and clicking "Paste." You can do this multiple times to include multiple groups. Close the Look Up window and click "OK" when you are done.
    b. To select companies by other criteria (e.g., market capitalization over a certain size), click on the "Customize" button in the lower right. Find and select the variables you want to use for screening, then click "OK" to return to the original screen. Check the boxes by those variables and enter values and screening criteria (greater than, equals, etc.) If you think you might use these screening criteria again, you may want to use the buttons in the lower left to save this screen.
15. Click OK once you have selected all of your companies. You will only see the data for the first company on your list, but data for all companies will be included when you export it.
16. Go to File→Save to save your data in a CSV file. Click on the "Text Options" button, select "Comma," and check the boxes by "Use Quotes" and "Use Labels." Delete the "txt" extension after the file name and type in "csv" to save as a CSV file. Make sure "All Companies" is selected.
17. Open your CSV file in Excel, insert a row at the top, and type in your column headers. You can select "Formula" in Research Insight to return to the formula view to see which variables are in which columns. If you've forgotten what a particular mnemonic stands for, you can use the "Look Up" button at the top of the Research Insight window to search the mnemonics and get a definition for that variable.