



The BRLESC, the Ballistic Research Laboratory Electronic Scientific Computer, of the U.S. Army, circa 1961. Source: U.S. Army, public domain image, via Wikimedia Commons. The computer contained more than 1,700 vacuum tubes, 850 transistors, and 46,500 diodes.

## **A Very Short Introduction to STATA**

Geography 450, Urban Analysis

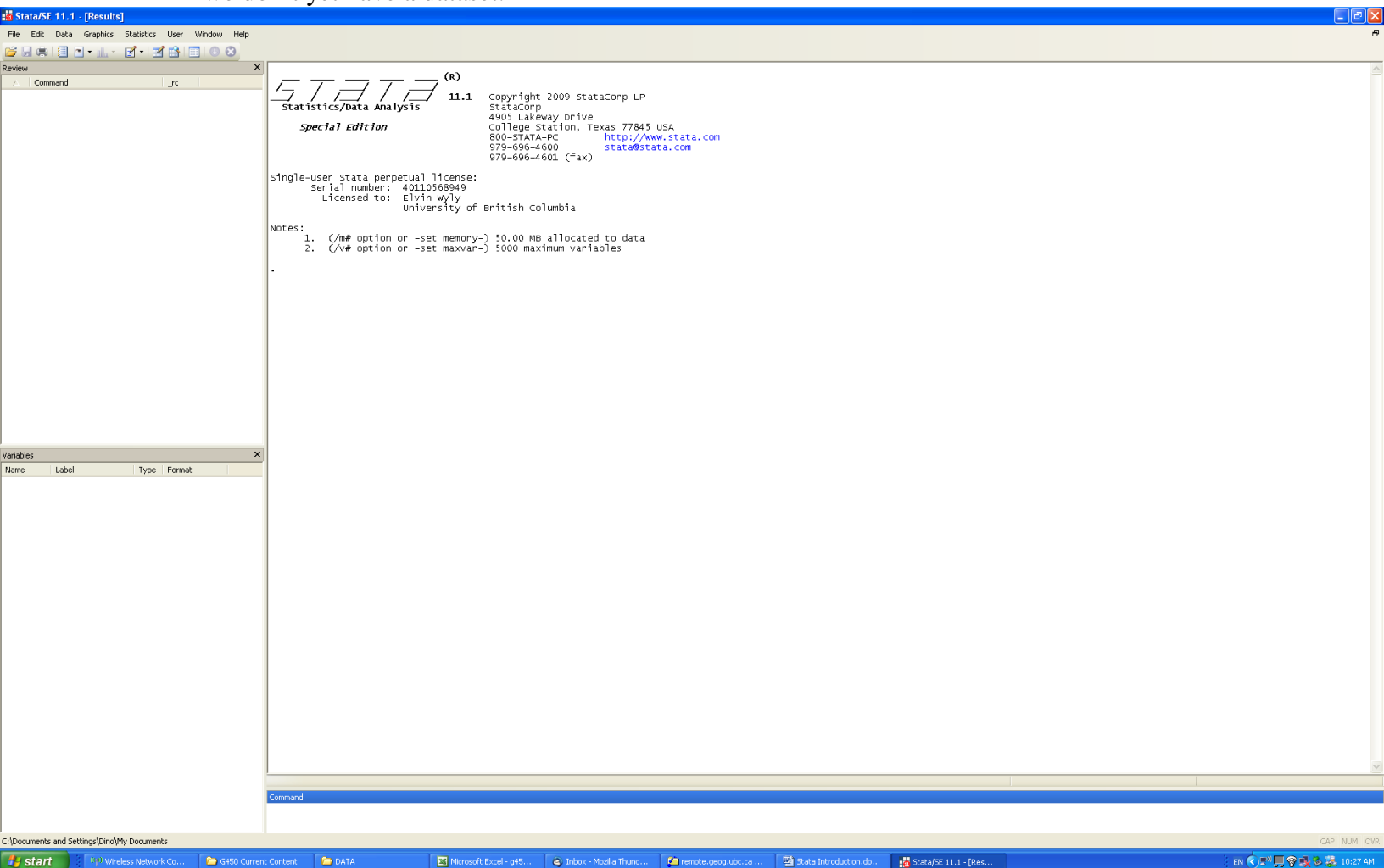
Elvin Wyly

There are dozens of software packages that are useful for organizing and analyzing large pools of information. One of the most powerful and versatile is called STATA. We will be using STATA for some of the projects in this course. You're not required to become an expert in STATA (or any other software package, for that matter) to do well in this course. But for those of you interested in working collaboratively, at least one colleague in each of the working groups should learn enough about STATA to help organize information and explore data for the group's project.

This document presents a very simple, quick introduction to just a few things that STATA can do. There are lots detailed and in-depth tutorials available for STATA. We'll just start with a few basics.

First, use your username and password to log on to one of the computers in the lab in Geography Room 115. Then find STATA on the list of programs from the Start Menu, or on the desktop if there's a program icon, and launch it.

You'll see a screen divided into four main panels. On the top left is a "review" panel, which keeps track of things we've done. We haven't done anything yet, so this is empty. On the right hand side is the "results" panel, which shows various messages and responses. It begins with a banner describing the version of STATA we're running, and a few other details. On the bottom right is the "command" panel, where we'll be telling STATA what to do. Finally, at the bottom left is a "variables" panel, which provides a quick reference for the names of variables in our dataset. This is empty, though, because we don't yet have a dataset.



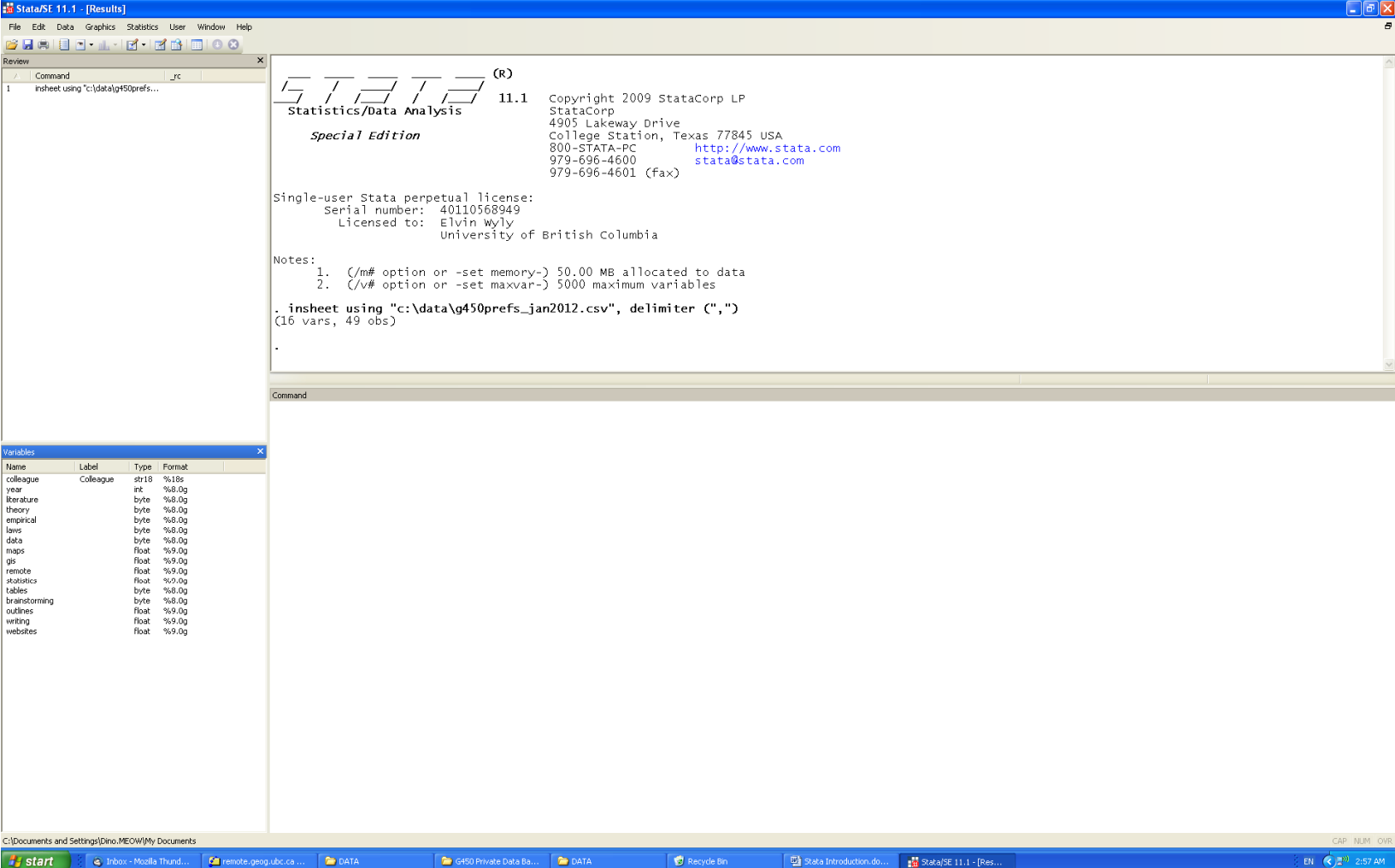
Let's import the summary of preferences our colleagues expressed for various research activities. I created a simple worksheet in Excel that looks like this:

Colleague	year	literature	theory	empirical	laws	data	maps	gis	remote	statistics	tables	brainstorming	outlines	writing	websites
Jessica Baas	2009	8	13	11	14	4	5	3	10	12	2	1	7	9	6
James Barnewall	2009	8	7	6	9	3	2	5	4	14	1	11	12	13	10
Kevin Cooper	2009	10	4	3	13	14	1	5	6	12	8	2	11	9	7
Rain de Guzman	2009	4	10	12	14	6	1	7	13	11	3	8	2	5	9
Lauren English	2009	13	14	6	5	7	1	4	8	12	2	3	9	11	10
Michael Lee	2009	7	1	3	2	6	4	12	14	13	8	5	9	10	11
Abie Luu	2009	14	13	4	12	8	2	7	5	11	3	9	10	6	1
Edda Mata	2009	1	6	7	12	8	13	14	11	10	4	3	2	5	9
Ramona Monticello	2009	1	2	3	7	8	4	14	13	12	9	5	10	11	6
Tara Mowat	2009	1	7	2	5	4	10	7.5	7.5	9	4	1	3	4	12
Tim Nugent	2009	3	6	2	1	10	5	13	9	14	11	4	8	7	12
Marnie Smith	2009	3	6	7	5	8	12	14	11	13	9	4	1	2	10
Mehdi Alaei Tafti	2009	5	10	2	1	7	13	12	11	3	8	4	6	9	14
Ren Thomas	2009	2	4	5	9	10	11	13	6	14	3	7	8	1	12
Tommy Thomson	2009	5	3	2	4	12	1	11	7.5	7.5	9	8	7	6	10
Savannah Kuchera	2011	2	7	2	1	5	6	6	11	10	7	6	6	6	12
Elliot Siemiatycki	2011	3	2	6	9	11	12	12	12	12	11	2	2	2	12
Joseph Tohill	2011	3	3	2	3	8	9	12	7	9	10	6	3	1	9
Matthew Baker	2011	2	1	4	6	6	1	6	6	10	6	4	3	2	4
Zoe Siegel	2011	1	7	8	4	3	9	12	12	12	5	3	6	2	6
Liam McGuire	2011	9	7	7	7	5	1	2	3	2	3	4	7	7	6
Brent Fairbairn	2011	3	3	3	9	9	6	9	6	12	6	3	1	1	9
Mark Stevens	2011	4	5	6	1	7	11	7.5	7.5	10	8	2	3	9	12
Joe O'Brien	2011	4	2	1	5	5	8	10	6	6	5	5	5	3	8
Gordon Clark	2011	10	10	10	10	10	13.5	13.5	2	2	2	5	6.5	6.5	4
Dylan Jones	2011	3	6	4	5	9	9	9	9	7	7	1	3	3	3
Alexandra Tyrrell	2011	7	6	5	8	9	11	12	12	12	10	1	2	3	4
Amelia Daly	2011	1	8	9	3	11	2	10	7	12	13	4	5	6	14
Janine Pham	2011	3	5	4	2	8	12	13	14	11	9	1	6	7	10
Emily Hein	2011	2	6	9	10	3	5	10	10	10	4	1	7	8	10
Jeffrey Fong	2011	12	8	5	5	3	2	2	4	7	5	1	6	8	8
Evan Goldstrom	2011	5	2	5	4	4	6	10	10	10	5	1	6	7	1
Daniela Aiello	2011	2	1	5	8	4	2	12	12	12	6	1	1	3	2
George Rahi	2011	5	1	3	6	8	7	10	11	12	9	2	4	3	4
Sam Walker	2012	9	5	4	12	13	1	2	3	11	8	6	7	10	14
Andrew Longhurst	2012	4	1	2	5	9	7.5	7.5	7.5	7.5	6	7	8	3	7.5
Henry Lebard	2012	1	1	1	2	4	1	14	1	3	5	6	5	3	4
Steffen Randahl	2012	5	1	6	8	11	12	13	14	10	9	2	3	4	7
Alex MacLeod	2012	12	13	10	8	6	2	4	9	3	1	11	7	5	14
Sophie Ellen Fung	2012	10	8	1	3	10	1	10	8	14	1	1	5	8	8
Claudia Hehn	2012	1	11	6	13	12	5	9	10	14	8	3	2	7	4
Bosco Ho	2012	14	13	3	8	7	5	10	4	11	6	1	2	9	12
Flora Lin	2012	4	10	6	9	5	13	14	12	11	8	1	3	7	2
Johanna Vortel	2012	6	10	9	7	8	3	14	13	12	4	2	5	1	11
Emily Rosenman	2012	7	5	1	2	11	6	10	13	12	9	8	4	3	14
Jacopo Miro	2012	6	7	1	5	9	11	14	12	10	8	2	3	4	13
Connor Donegan	2012	10	2	1	9	8	7	14	13	6	6	4	5	3	12
Kevin Chan	2012	1	6	5	3	9	10	13	11	12	8	2	7	4	14
Mark Bjelland	2012	7	5	3	9	10	2	1	13	11	4	6	8	12	14

Recall that I asked you to rank various kinds of methods and approaches -- reading theoretical work, reading empirical stuff, investigating laws, creating statistical models, and so on. The ranks go from 1 (you just love it) to 14 (you simply cannot stand it). I've arranged things into rows, one for each of our colleagues, and columns for the various methods or activities. The cells include the ranks, from 1 to 14, for each activity. It's standard procedure to refer to the columns as *variables*, and the rows as *observations*.

STATA prefers to have data in a slightly different form, so I exported the Excel worksheet it to a comma-delimited file. You should download this file from

[http://www.geog.ubc.ca/~ewyly/Private/g450/g450prefs\\_jan2012.csv](http://www.geog.ubc.ca/~ewyly/Private/g450/g450prefs_jan2012.csv)



and then store it in C:\DATA on the local computer.

Now, in STATA, place your cursor on the command panel, and then type in

```
insheet using "c:\data\g450prefs_jan2011.csv", delimiter (",")
```

and then press enter. This simply tells STATA to input data that was created in a spreadsheet program, to use a certain file located in the c:\data directory, and that the values are separated by commas.

STATA reads the data, and new items appear in the screens on the left.

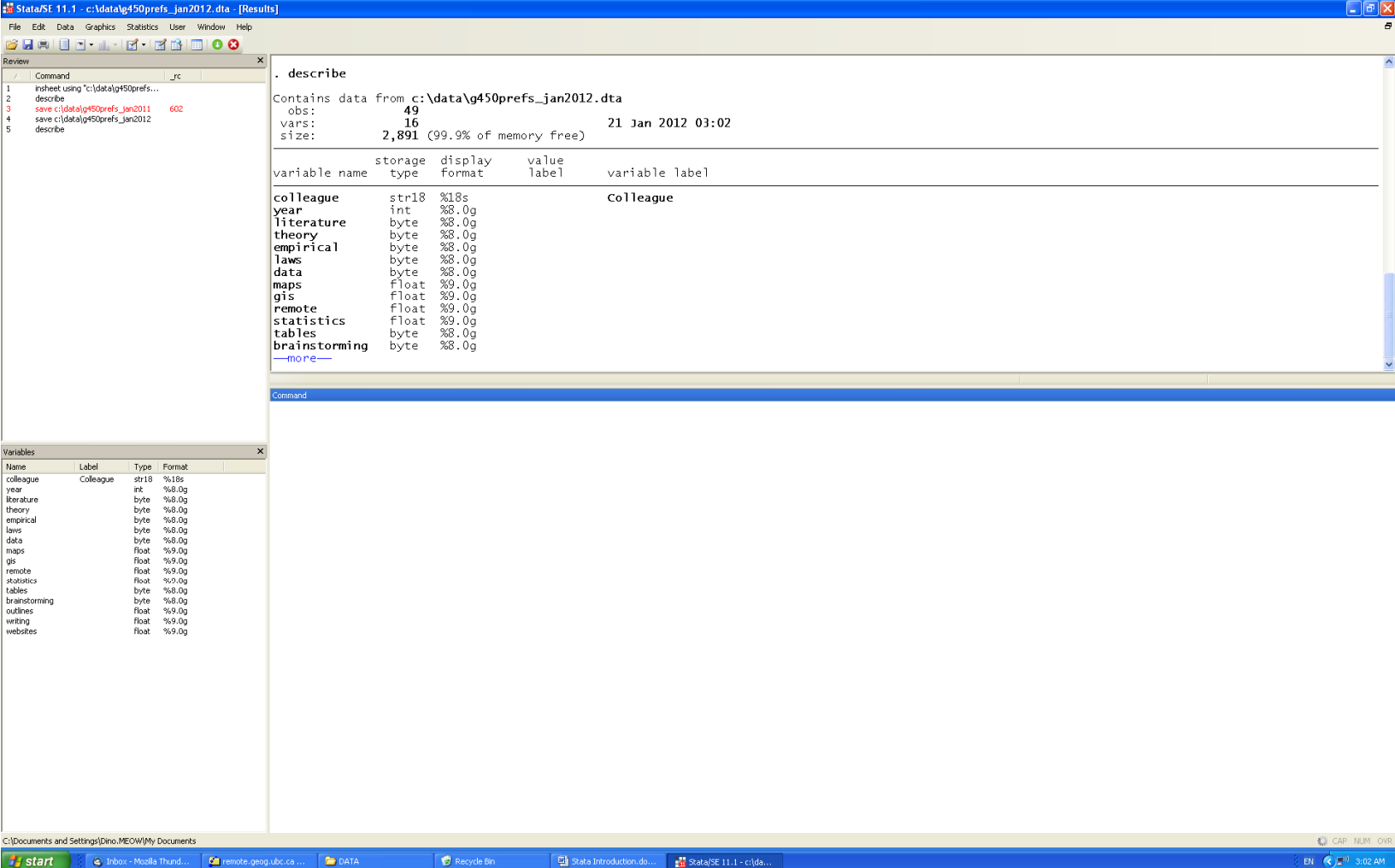
You see the variable list in the panel on the lower left, and you see the command I issued in the review panel in the top left.

Before we do anything else, let's save the data, which were imported from the comma-delimited file, in the format that STATA prefers. In the command panel, type

```
save c:\data\g450prefs_jan2011
```

and then press enter. STATA will save the file in the same directory you used to import the data, and will append ".dta" as the file extension.





There are several commands that are useful to explore what we've got. One is the describe command. Type

**describe**

and then press enter. STATA gives you a nice overview of various features of your data: the name of the dataset, the number of observations, the number of variables, and then a bunch of other details that we don't need to worry about right now.

Another way we can explore things is to use the browse command. Type

**browse**

and then press enter. STATA creates a new panel on top of all the others, and now you see something like this (next page):

Data Editor (Browse) - [g450prefs\_jan2012.dta]

File Edit Data Tools

colleague[] Jessica Baas

	colleague	year	literature	theory	empirical	laws	data	maps	gis	remote	statistics	tables	brainstorm-g	outlines	writing	websites
1	Jessica Baas	2009	8	13	11	14	4	5	3	10	12	2	1	7	9	6
2	James Barnewall	2009	8	7	6	9	3	2	5	4	14	1	11	12	13	10
3	Kevin Cooper	2009	10	4	3	13	14	1	5	6	12	8	2	11	9	7
4	Rain de Guzman	2009	4	10	12	14	6	1	7	13	11	3	8	2	5	9
5	Lauren English	2009	13	14	6	5	7	1	4	8	12	2	3	9	11	10
6	Michael Lee	2009	7	1	3	2	6	4	12	14	13	8	5	9	10	11
7	Able Luu	2009	14	13	4	12	8	2	7	5	11	3	9	10	6	1
8	Edda Mata	2009	1	6	7	12	8	13	14	11	10	4	3	2	5	9
9	Ramona Monticello	2009	1	2	3	7	8	4	14	13	12	9	5	10	11	6
10	Tara Mowat	2009	1	7	2	5	4	10	7.5	7.5	9	4	1	3	4	12
11	Tim Nugent	2009	3	6	2	1	10	5	13	9	14	11	4	8	7	12
12	Marnie Smith	2009	3	6	7	5	8	12	14	11	13	9	4	1	2	10
13	Mehdi Alael Tafli	2009	5	10	2	1	7	13	12	11	3	8	4	6	9	14
14	Ren Thomas	2009	2	4	5	9	10	11	13	6	14	3	7	8	1	12
15	Tommy Thomson	2009	5	3	2	4	12	1	11	7.5	7.5	9	8	7	6	10
16	Savannah Kuchera	2011	2	7	2	1	5	6	6	11	10	7	6	6	6	12
17	Elliot Siemiatycki	2011	3	2	6	9	11	12	12	12	12	11	2	2	2	12
18	Joseph Tohill	2011	3	3	2	3	8	9	12	7	9	10	6	3	1	9
19	Matthew Baker	2011	2	1	4	6	6	1	6	6	10	6	4	3	2	4
20	Zoe Siegel	2011	1	7	8	4	3	9	12	12	12	5	3	6	2	6
21	Liam McGuire	2011	9	7	7	7	5	1	2	3	2	3	4	7	7	6
22	Brent Fairbairn	2011	3	3	3	9	9	6	9	6	12	6	3	1	1	9
23	Mark Stevens	2011	4	5	6	1	7	11	7.5	7.5	10	8	2	3	9	12
24	Joe O'Brien	2011	4	2	1	5	5	8	10	6	6	5	5	5	3	8
25	Gordon Clark	2011	10	10	10	10	10	13.5	13.5	2	2	2	5	6.5	6.5	4
26	Dylan Jones	2011	3	6	4	5	9	9	9	9	7	7	1	3	3	3
27	Alexandra Tyrrell	2011	7	6	5	8	9	11	12	12	12	10	1	2	3	4
28	Amelia Daly	2011	1	8	9	3	11	2	10	7	12	13	4	5	6	14
29	Janine Pham	2011	3	5	4	2	8	12	13	14	11	9	1	6	7	10
30	Emily Hein	2011	2	6	9	10	3	5	10	10	10	4	1	7	8	10
31	Jeffrey Fong	2011	12	8	5	5	3	2	2	4	7	5	1	6	8	8
32	Evan Goldstrom	2011	5	2	5	4	4	6	10	10	10	5	1	6	7	1
33	Daniela Aiello	2011	2	1	5	8	4	2	12	12	12	6	1	1	3	2
34	George Rahi	2011	5	1	3	6	8	7	10	11	12	9	2	4	3	4
35	Sam Walker	2012	9	5	4	12	13	1	2	3	11	8	6	7	10	14
36	Andrew Longhurst	2012	4	1	2	5	9	7.5	7.5	7.5	7.5	6	7	8	3	7.5
37	Henry Lebard	2012	1	1	1	2	4	1	14	1	3	5	6	5	3	4
38	Steffen Landahl	2012	5	1	6	8	11	12	13	14	10	9	2	3	4	7
39	Alex MacLeod	2012	12	13	10	8	6	2	4	9	3	1	11	7	5	14
40	Sophie Ellen Fung	2012	10	8	1	3	10	1	10	8	14	1	1	5	8	8
41	Claudia Hehn	2012	1	11	6	13	12	5	9	10	14	8	3	2	7	4
42	Bosco Ho	2012	14	13	3	8	7	5	10	4	11	6	1	2	9	12
43	Flora Lin	2012	4	10	6	9	5	13	14	12	11	8	1	3	7	2
44	Johanna Vortel	2012	6	10	9	7	8	3	14	13	12	4	2	5	1	11
45	Emily Roseman	2012	7	5	1	2	11	6	10	13	12	9	8	4	3	14
46	Jacopo Miro	2012	6	7	1	5	9	11	14	12	10	8	2	3	4	13
47	Connor Donegan	2012	10	2	1	9	8	7	14	13	6	6	4	5	3	12
48	Kevin Chan	2012	1	6	5	3	9	10	13	11	12	8	2	7	4	14
49	Mark Bjelland	2012	7	5	3	9	10	2	1	13	11	4	6	8	12	14

Ready Vars: 16 Obs: 49 Filter: Off Mode: Browse CAP: NUM

start | Inbox - Mozilla Thund... | remote.geog.ubc.ca ... | DATA | Recycle Bin | Stata Introduction do... | Stata/SE 11.1 - c:\ds... | Data Editor (Browse) ... | Microsoft Excel - g45... | EN | 3:05 AM

This just looks like any spreadsheet program. If you try to select one of the cells and type in a new number, however, STATA will not respond. Browse allows you to look at the contents of the dataset, but it does not allow you to make changes. This feature comes in handy: it prevents you from accidentally making changes. If you want to change the dataset, you need to use a different feature, the **data editor**. We'll worry about that another time.

Take a close look at the values in the cells. Almost all of the numbers are *integers* -- that is, from 1 to 14, with no decimals. But there are a few exceptions. When I looked at the survey from Tara Mowat -- a student who took this course a few years ago -- I noticed that she left the ranks for GIS and remote sensing blank. She wrote, "I'm afraid I've never done geocoding, any GIS operations, or interpreted remote sensing images. Therefore, I can't assign it a score. Although I'd be interested in giving it a go." Tommy Thomson, another colleague from a previous year, also left two items blank (remote sensing and statistical models). Mark Stevens also left two items blank (remote sensing and geocoding, buffers, and other GIS operations). Similarly, Andrew Longhurst was neutral on four items -- designing websites, GIS analysis, statistics, remote sensing, and cartography.

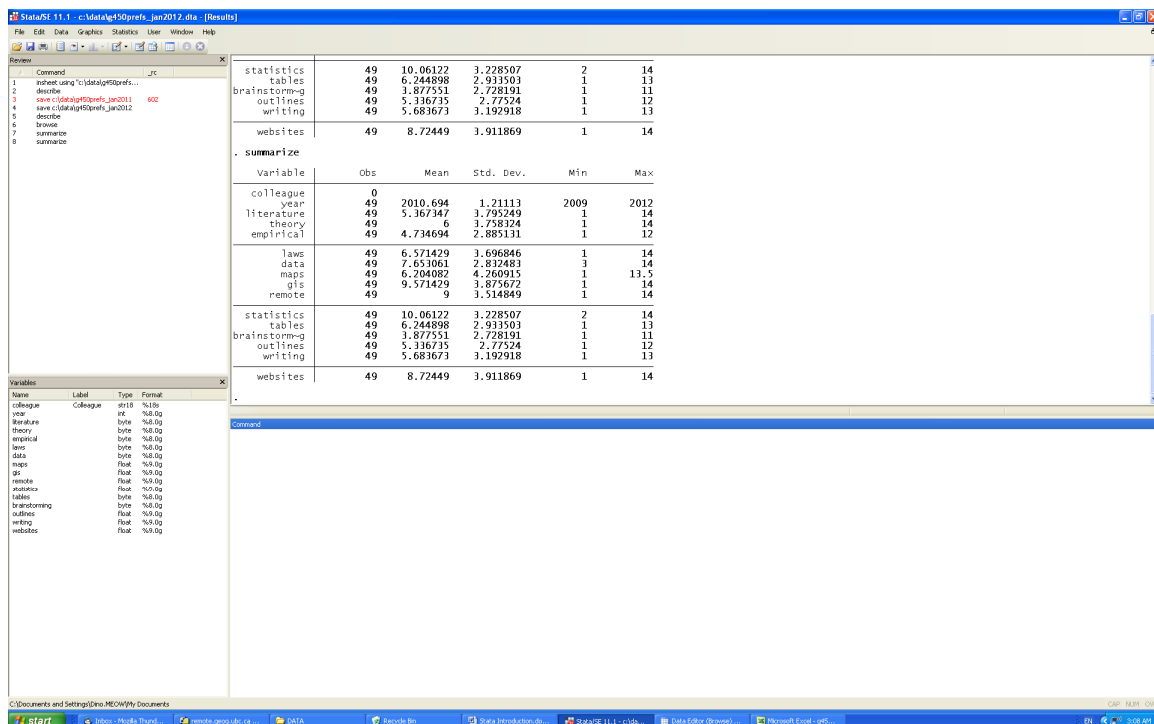
These colleagues have highlighted a crucial implicit assumption of my primitive little survey -- the questions presume sufficient familiarity to permit respondents to assign ranks to every choice. Our colleagues have brilliantly exposed the problems with this assumption. Still, unless I leave them out of the dataset -- ignoring them, and all their

responses on the other choices that they did feel comfortable ranking -- I need some consistent way of dealing with missing values. I can't assign zeros -- if I do that, then, given the way we've defined the ranks from 1 to 14, it would imply that Tara and her comrades really love these activities, and they don't. So I decided that it made sense to regard the blank responses as complete neutrality. With a scale between 1 and 14, the absolute midpoint is 7.5, so that's what I decided to enter in place of the blanks.

Now, close down the browser window -- press the red X in the top-right corner -- and you're back to the command panel. Now type

**summarize**

and then press enter. You'll see something like this:



Summarize gives us a bit more detail than the describe command. It provides the number of observations with non-missing values for each variable, along with the mean, the standard deviation, the minimum and the maximum. The mean is the same thing as the average, and the standard deviation is the average distance of the values from the average -- a measure of the “spread” of values in a particular distribution. Don't worry about the standard deviations for now; these are difficult or impossible to interpret with the measurement scale we used in this simple survey -- ranked and ordered data, also known as **ordinal** data. We'll worry about the details of ordinal data later.

We can still learn some interesting things, however, if we consider the mean values for our colleagues. The research activities with the lowest mean scores are “Brainstorming and exploring alternative ways to organize a project,” and “Reading and discussing

*empirical* or *applied* policy documents, newspaper articles, etc.” Brainstorming has a mean score of 3.88, while empirics has a mean of 4.73. By contrast, the mean value for “Designing and estimating statistical models” exceeds 10. Clearly, statistics are not your favorite activity. There is also, on average, an aversion to GIS, remote sensing, and designing and creating websites. Only one of our colleagues in recent years has identified remote sensing as their number-one preference.

Given that we’re in the Faculty of Arts, this makes sense -- most of our colleagues have strong preferences and expertise in conceptual inquiry, reading scholarship, and exploring the multidimensional aspects of society, culture, and geography. Do keep in mind, however, that many career paths do require some exposure to certain analytical and/or technical skills. Nearly every graduate program in urban planning includes a required course on quantitative methods, and many urban-oriented jobs -- in the public, private, and non-profit sectors -- require some familiarity with GIS. Some observers think this *de facto* requirement is lamentable, others applaud it. But everyone agrees that it’s something you’ll have to deal with sooner or later. I would also note that there are a wide array of companies that profit through various forms of subtle deception specifically crafted for consumers who are uncomfortable with mathematical calculations. Peter Gould was fond of saying that “learn” is not a transitive verb; I hope he forgives me for pointing out that if you don’t learn quantitative methods, they will be learned upon you.

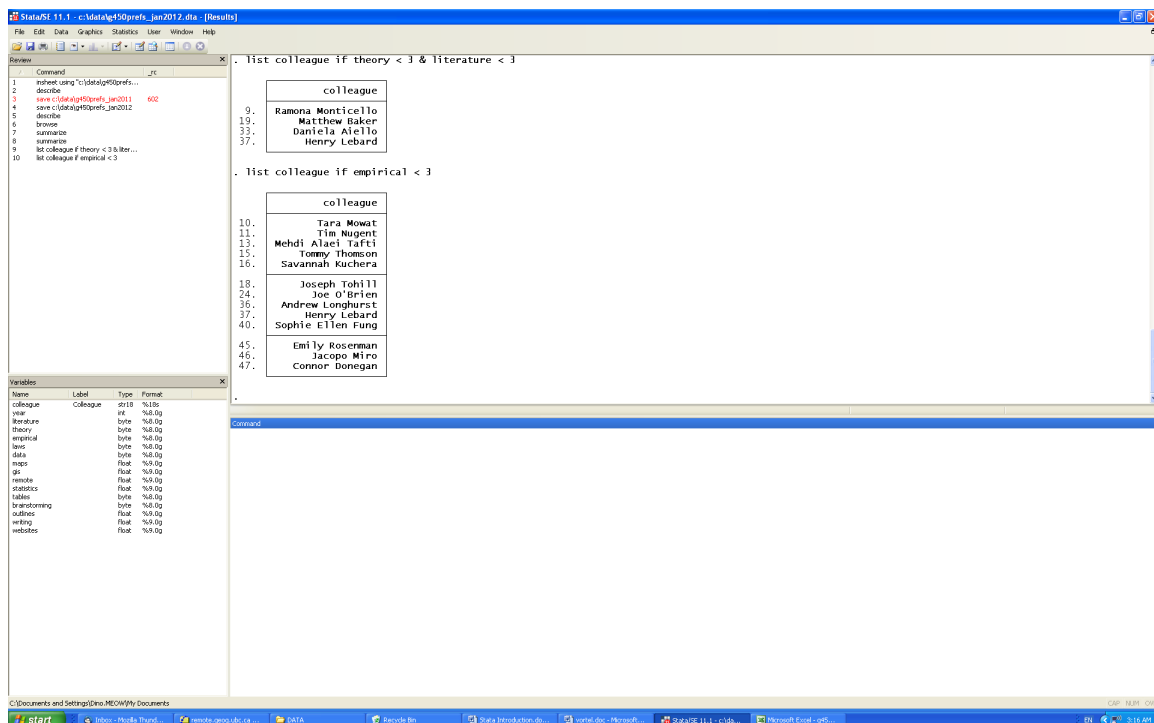
But let’s come back to that rich, multidimensional perspective on society, culture, and geography. If we wanted to visualize the multidimensional differences and similarities amongst our colleagues, how might we do it? We’ve already tried the simple univariate approach -- comparing the mean values of the ranks for the different activities, one variable at a time. But are there other ways? There are many alternatives, but let’s just consider three. First, we can divide up the list of names according to various criteria. So if we wanted to identify our colleagues who are the most conceptual and/or theoretically-inclined, we might type this into the command panel:

```
list colleague if theory < 3 & literature < 3
```

Four colleagues cited *both* reading theory and searching out literature in the top 2 (i.e., rank < 3). By contrast, typing

```
list colleague if empirical < 3
```

gives us a total of thirteen colleagues who ranked this kind of work in the top 2.



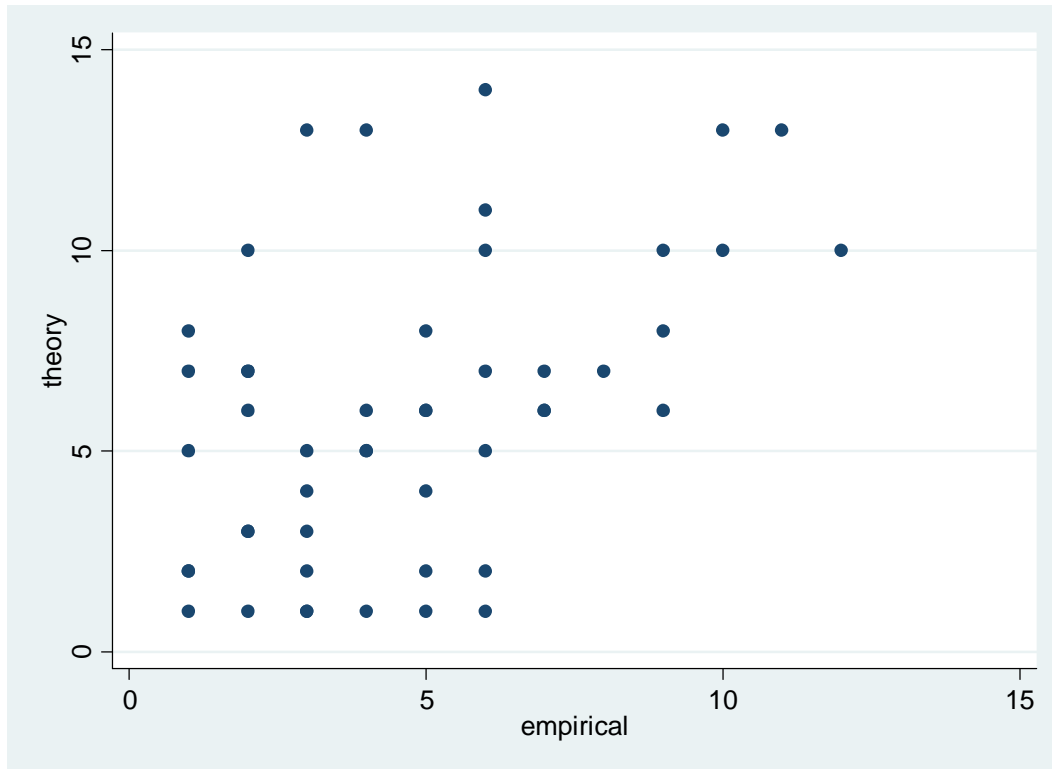
We could play around with all sorts of alternative combinations to explore similarities and differences. But is there a more systematic way of visualizing our complementary preferences? To be sure, with a small number of colleagues, we could simply have a conversation in class to draw a mental map of our diverse preferences and skills. But suppose we wanted to do this for a larger number of our friends and colleagues; in that case, we would need a systematic method that could be applied to many, many people (observations).

So this brings us to a second approach. Let's draw a simple scatterplot, comparing the ranks that our colleagues have assigned for "Reading *theoretical* scholarship and discussing it," versus "Reading and discussing *empirical* or *applied* policy documents, newspaper articles, etc." In many fields, there is a pronounced division between theoretical and applied inquiry. In the command panel, type

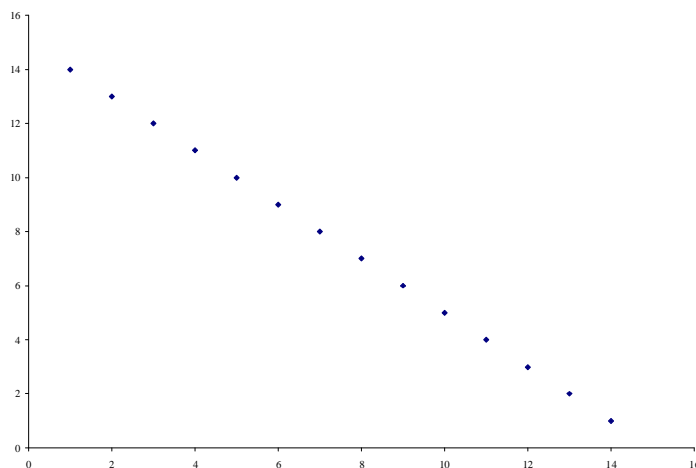
**graph twoway scatter theory empirical**

and then press enter. Don't worry, I don't expect you to memorize any of these commands, and indeed I am just learning STATA myself. You can do any of the things I've shown thus far by following the menus at the top of STATA. You can also get help by typing the wonderfully simple command **help**. If you know the topic you need help on, you can type something like **help graph**. If you don't know the precise command or procedure, then you can just do a keyword search, as in **search graph**, and STATA will provide a list of all commands that include mentions of the keyword.

I prefer to use the command panel rather than the pull-down menus. The graph command above gives us this (next page):



We see no strong evidence of a sharp divide between the empirical and theoretical. If there were a sharp divide, we would see a downward-sloping line: people who assigned high ranks to reading theory would have assigned low ranks to reading empirical stuff, and vice versa. In the extreme case, if people who loved theory hated empirics and the people who hated theory loved empirics, the graph would have looked like this:



The graph of our preferences does not look like this. Colleagues are scattered all over the place, in no apparent pattern. What we can safely say is that we do not have a clear dichotomy between theory and empirics. But we do have to be careful about drawing further conclusions, because the way I worded the questions had the effect of mixing a few distinct activities. Recall the two questions:

“Reading *theoretical* scholarship and discussing it.”

“Reading and discussing *empirical* or *applied* policy documents, newspaper articles, etc.”

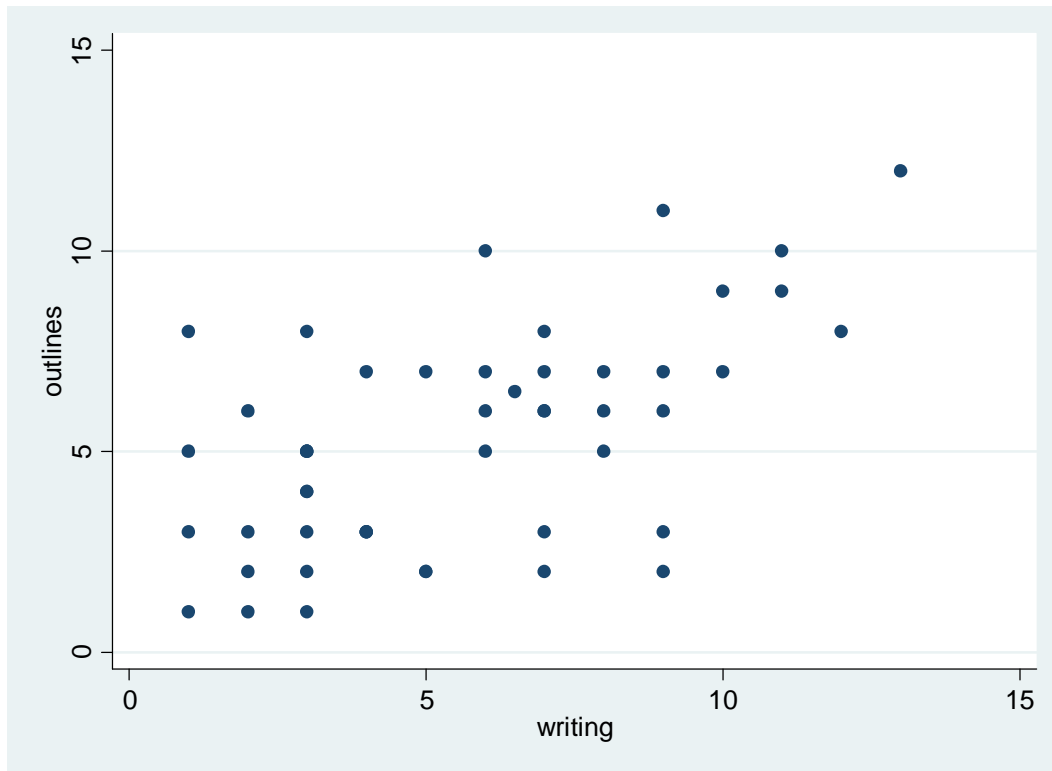
The two questions present a contrast not only between theoretical and empirical reading and discussion, but between scholarship and “policy documents, newspaper articles, etc.” If the question had been worded differently, to distinguish between reading theoretical and empirical scholarship, the responses could have been different. The present wording for the second question makes it impossible for us to know, without going back and asking everyone again, precisely what it is about this activity that is preferred or disliked; is it the empirical or applied nature? The emphasis on policy documents? The mention of newspaper articles? Or the combination of policy documents, newspaper articles, and that familiar grab-bag of everything, the “etc.”?

This is why there are entire courses, and indeed entire graduate degree programs, focused on the principles of survey research. I haven’t taken those courses, and you see the trouble that got us into. Fortunately, the public institutions that provide the results of large-scale social surveys hire people who *have* taken these courses and earned specialized degrees. They invest considerable time, money, and expertise in designing surveys that help to minimize the risks of misinterpretation.

Let’s try testing another hunch. We were also asked to rank two activities that are often viewed as quite complementary -- preparing outlines and writing. The specific wording was “Creating outlines for papers,” and “Writing and editing papers.” It seems reasonable to hypothesize that colleagues would find both of these activities enjoyable, or that they would dislike both; conversely, we would expect few people to have very strong preferences for one and intense dislike for the other. Test this hunch by typing

**graph twoway scatter outlines writing**

And you obtain the chart below. This provides fairly strong support for our hypothesis. The scatter of points slopes up to the right, and most people either like both activities (putting them in the lower left corner of the graph) or they dislike both (putting them to the top right). There are a few notable exceptions -- someone has expressed a strong preference for writing (ranked 1), for instance, along with a more neutral stance towards preparing outlines (ranked 8). If we were to exclude that exception and a few others, the positive relation -- the correlation -- between writing and editing would be quite strong. In many areas of research, the investigator will search for these exceptions -- often technically referred to as “outliers” or “residuals” -- and try to decide whether they should be excluded from a particular analysis. Sometimes there are justifiable reasons for excluding outliers. But I’m very cautious about doing this. I exclude outliers if I identify problems with missing data, or responses that are clearly contradictory or incorrect (say, if someone had left most of the choices blank on our survey of preferences). Otherwise, however, excluding outliers risks obscuring the diversity of unusual combinations. And the diversity of unusual combinations, for people and places, is what geography is all about, right?



These two graphs illustrate the many different relations we could explore between each of these research methods and activities, compared to another. With fourteen different activities, there are many different two-variable combinations -- bivariate relationships -- we could study. But can we consider more relationships at once? We could add a third variable, and create a three-dimensional scatterplot, with one axis measuring colleagues' preferences for reading theory, another axis measuring preferences for writing and editing, and another axis measuring preferences for reading empirical materials. Reproducing it here would require that we use the art and geometry of perspective to reduce a three-dimensional plot down to two dimensions, for the paper or the screen on which you are reading these words. But it turns out that this reduction -- transforming three-dimensional representations into two -- can be easily done with mathematics. That's why software programs can easily draw graphs with a three-dimensional appearance, projected onto a flat screen. Even more remarkably, however, there is nothing that restricts us from going to still more dimensions. Envision that three dimensional graph of preferences for theory, writing/editing, and empirical inquiry -- *and now try to imagine, in your mind's eye, a fourth dimension for designing and creating maps, a fifth dimension for brainstorming and organizing, and so on, until we have all of the fourteen dimensions along which we have expressed our preferences.*

The basic principles of geometry -- going all the way back to Euclid -- make it possible to measure the distance between observations in multidimensional spaces like our fourteen-dimensional space of preferences. Then, in the 1930s, psychologists and other scientists began working on ways of translating these complex, multidimensional, and impossible-



to-visualize mathematical spaces into something much more familiar -- a two-dimensional map. The challenge was similar to that of map projections. Projecting the (imperfect) sphere of the Earth's surface onto a flat, two-dimensional map requires distortion, and there is no such thing as a perfect solution: preserving directional relationships distorts area, and preserving areal relationships distorts directions (and hence shapes). But imperfect as they are, maps are extremely valuable two-dimensional representations of a three-dimensional world.

Eventually, a group of scientists developed a technique known as **multidimensional scaling**. Given any set of measures of the similarity between different observations according to scores on multiple variables, multidimensional scaling calculates the distances between the observations, and then transforms this multidimensional distance into a two-dimensional map. There will be some distortion, and the amount of distortion depends on the particular observations being measured; but it's possible to calculate numerical estimates of the overall distortion of the two-dimensional map.

Peter Gould taught multidimensional scaling in one of his Geography classes for many years. I first learned about it in the mid-1980s. The techniques had been developed by psychologists many years earlier -- an initial attempt in the late 1930s, and then a breakthrough by Warren Torgerson, who used his time on a fellowship from the Educational Testing Service (spent at Princeton University in the early 1950s) to devise a reliable, accurate, and routine method for finding the minimum-error solution.<sup>1</sup> The approach was refined considerably in the 1960s, by Torgerson and several others. We don't need to dive into the mathematical details, which Torgerson meticulously worked through in a series of equation-filled pages, deriving several alternative solutions before noting, in words that seem to reach out from the page in a deep sigh, that "All of these methods require a great deal of labor."<sup>2</sup> Torgerson's many achievements included profound understatement. Another scientist provided a more vivid description of the challenges of iterative procedures designed to move observations around in small increments, re-measuring the distances between all possible pairs at each step:

"The computational labour of examining all  $2^{N-1} - 1$  splits is enormous and the authors give  $(N-1)^2 2^{N-11}$  seconds as the time required on a computer with a  $5\mu$ -second access time. Thus with  $N$  as low as 21, over 100 hours are required and with  $N=41$  the time is over 54,000 years. Even with the fastest of projected computers these times could only be decreased by a factor of about 100, so the method is impracticable even for small values of  $N$ , and an alternative, possibly approximate algorithm must be devised.  
...<sup>3</sup>

---

<sup>1</sup> Warren S. Torgerson (1952). "Multidimensional Scaling: I. Theory and Method." *Psychometrika* 17(4), 401-419.

<sup>2</sup> Torgerson, "Multidimensional Scaling," p. 414.

<sup>3</sup> J.C. Gower (1967). "A Comparison of Some Methods of Cluster Analysis." *Biometrics* 23(4), 623-637, quote from p. 628.

Devising these alternative algorithms presented serious challenges to an entire generation of scientists. But the fundamental essence of all the complicated and sophisticated procedures is not too difficult to understand, *if* you can keep in your mind's eye a fourteen-dimensional world. Now, plot the observations -- place each of your colleagues -- into this fourteen-dimensional space. Now measure how far apart they are from one another. Now try to take those measurements and plot out everyone on a piece of paper. You'll have to do a lot of erasing and moving back and forth, as you try to keep the distances right between every individual and every other individual, when you only have two dimensions. There will be some distortion in the result, but if you follow certain rules, you can minimize this distortion. This would require a lot of erasing, a lot of re-drawing to try to get the distances as accurate as possible. But all of that repetitive, annoying measuring-drawing-and-erasing work, trying all sorts of combinations to try to minimize the distortions -- that all seems like the kind of repetitive tasks for which we have computers. Indeed, when Torgerson was asked to look back on thirty-five years of progress in multidimensional scaling, he wrote an essay for the fiftieth-anniversary issue of *Psychometrika* tracing the major themes and breakthroughs, and then offered this epilogue:

“The major developments discussed in this paper would have been of little interest, and would probably not even been publishable without the parallel development of the computer. In the early 1950's, desk calculators were used to carry out metric multidimensional analyses for similarity matrices of around ten stimuli. The computational labor was tedious and time consuming for this basic procedure even when applied to data sets of minimal size. The entire analysis would be no more than one small part of an iterative step in the present day computational routines.”<sup>4</sup>

Fortunately, this technique developed more than fifty years ago has been incorporated into STATA. So go back to the command panel and type in

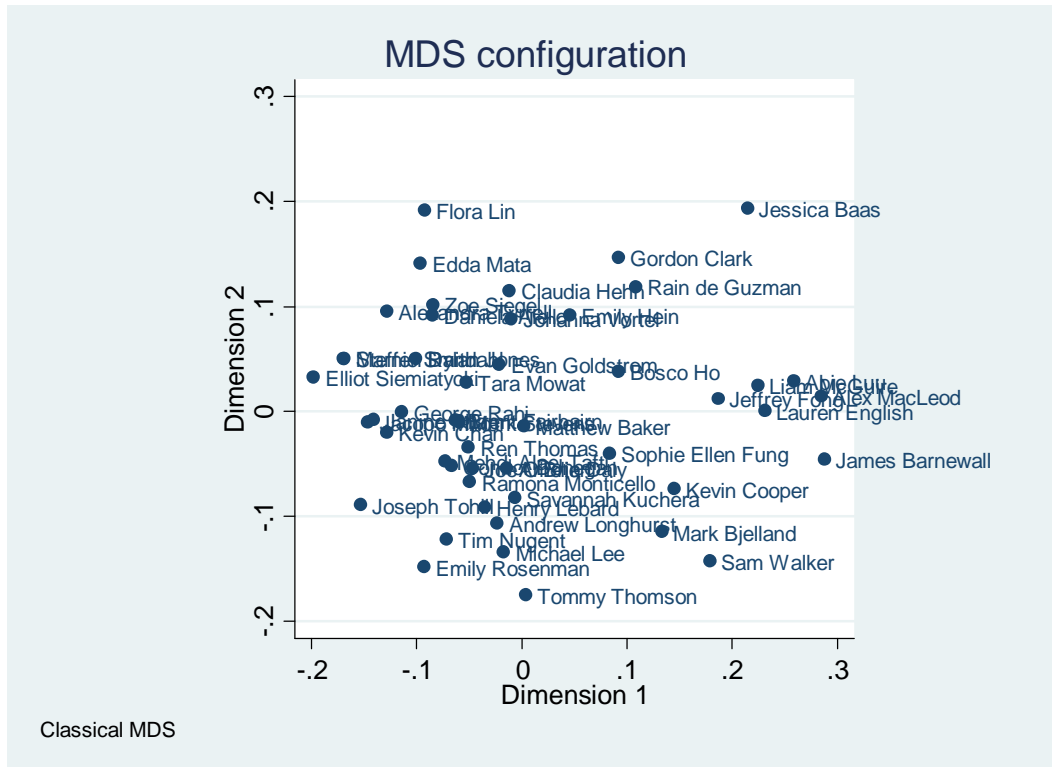
```
mds literature-websites, measure(gower) id(colleague)
```

This asks for a multidimensional scaling (mds) using all of the variables from 'literature' through 'websites.' Then the **measure** option requests that the distances between the observations be calculated using a special measure devised many years ago by J.C. Gower.<sup>5</sup> Gower was concerned with the need to find a general measure of distance between observations when some of the variables were measured on an ordinal scale -- as in the case of our rank-ordered preferences. The final option in the command above asks STATA to identify the observations using the value of the variable named 'colleague.' The result looks like this:

---

<sup>4</sup> Warren S. Torgerson (1986). "Scaling and Psychometrika: Spatial and Alternative Representations of Similarity Data." *Psychometrika* 51(1), 57-63, quote from p. 61.

<sup>5</sup> J.C. Gower (1971). "A General Coefficient of Similarity and Some of its Properties." *Biometrics* 27(4), 857-871.

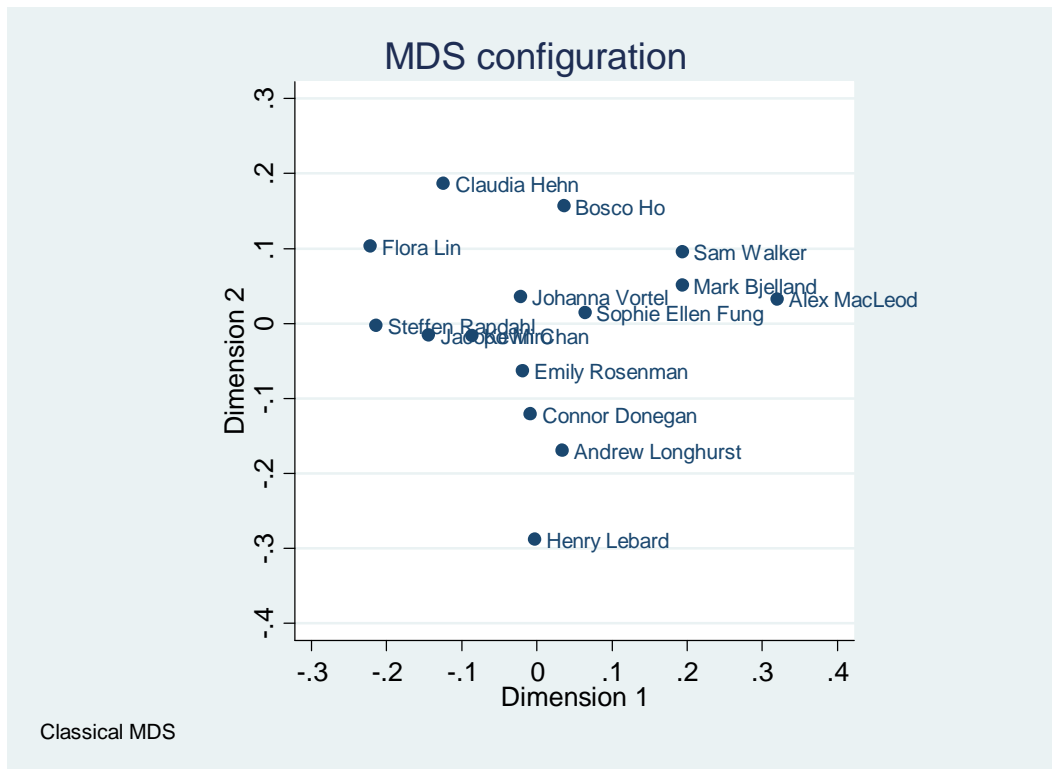


Obviously, you won't recognize all of the names on this map, because many of these colleagues who took the course in previous years have moved on to bigger and better things. So to focus on our current colleagues, issue this command:

```
mds literature-websites if year==2012, measure(gower) id(colleague)
```

Notice that STATA requires a double-equals sign for what are called 'equality tests.' STATA reserves the single equals sign to be used when you want to generate new variables.

When we restrict our analysis to 2012, this is the map we get:



This is **not** a scatterplot like the ones we did earlier. The two dimensions are not variables plotted on a chart: they are more like North-South and East-West directions on a map -- a map of the world that we have collectively created, by me asking you about your preferences, and you ranking various activities. Those decisions created a fourteen-dimensional world that can then be projected onto this two-dimensional map. There is some distortion, but not very much: one measure indicates that the two dimensions shown here are able to measure the fourteen-dimensional distances with about 74 percent accuracy. That's pretty good.

Now how to interpret this? Stata adds thoroughly unhelpful labels, like "Dimension 1" and "Dimension 2," because there is -- fortunately! -- no convenient way to ask a software program to do the fundamentally human activities of *thinking* and *interpretation*. So this is the point where I find it best to print out a few of the results produced by the software, to close things down, and then to get as far away as possible from all computers and other attention-disruption devices -- to find a nice quiet place without any distractions, to just *think*. So let's save things and exit. If you type

```
save c:\data\g450prefs.dta
```

and press enter, STATA will refuse and tell you, in that warning-red text color, that the data set already exists. You could save the dataset under a new name, if you'd like. If you're absolutely sure that you want to overwrite the old one, type

```
save g450prefs.dta, replace
```

You could also exit the program without saving your dataset, by typing

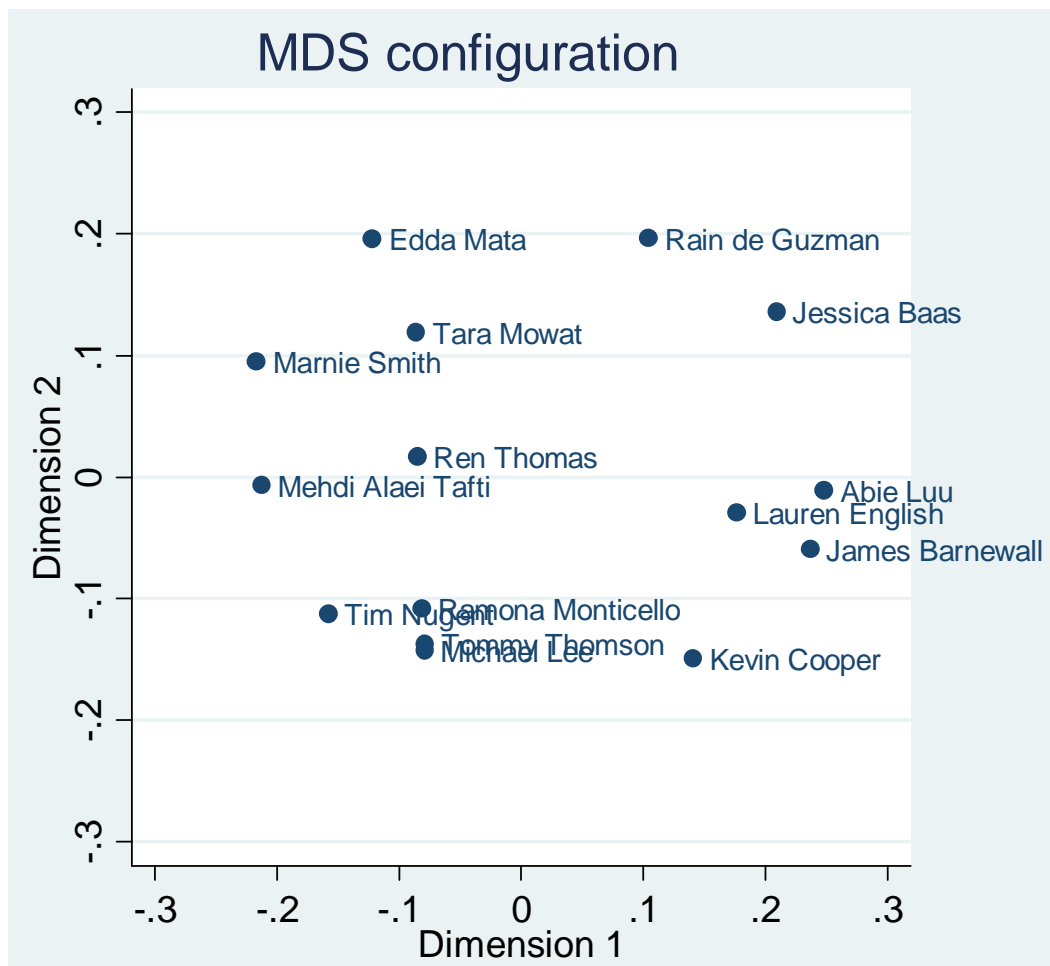
```
clear
```

which clears the current dataset from memory. *Be very, very careful with the clear command. With only a few keystrokes it's very easy to lose a lot of work if you've made changes to your data.* After you've saved the dataset, now you can type

```
exit
```

and press enter. STATA shuts down all the windows.

Now you can free yourself from the computer, find that quiet corner in the library, and sit down and think about what we've got. The multidimensional scaling procedure is measuring distances between all of the possible combinations of likes and dislikes amongst our colleagues. It's hard to think clearly about fourteen different aspects of individual preferences all at once. So when I first did this MDS of student preferences in January of 2009, I just sat down with a printed copy of the map for the students taking the course at that time. It has such an awkward, uninformative title -- "MDS Configuration." Here's what it looks like for the colleagues taking the course at that time:





## Geography 450, URBAN ANALYSIS, January 2009

So I sat there, looked at the map, and jotted down the top three preferred activities next to each name on the map. Then I tried to think through the connections, the areas of difference and similarity. What if each of us could be understood as a city, situated in a particular regional context, each with distinctive features yet still bound together by travel, trade, and other relations? What if our shared regional settings -- our common preferences for certain types of thinking -- could be viewed in topographical terms? What might the landscape look like?

This is what I came up with, in about forty minutes of sketching and thinking. Cartography has always been a blend of art and science, and this is especially true in the case of multidimensional scaling applied to survey data. Yet key features of the map of our colleagues in 'preference space' stand out. First, there is a broad division that we can imagine at the regional (or even subcontinental) scale. Nearly everyone in the east and southeast mentions designing and creating maps among their top three preferences, while

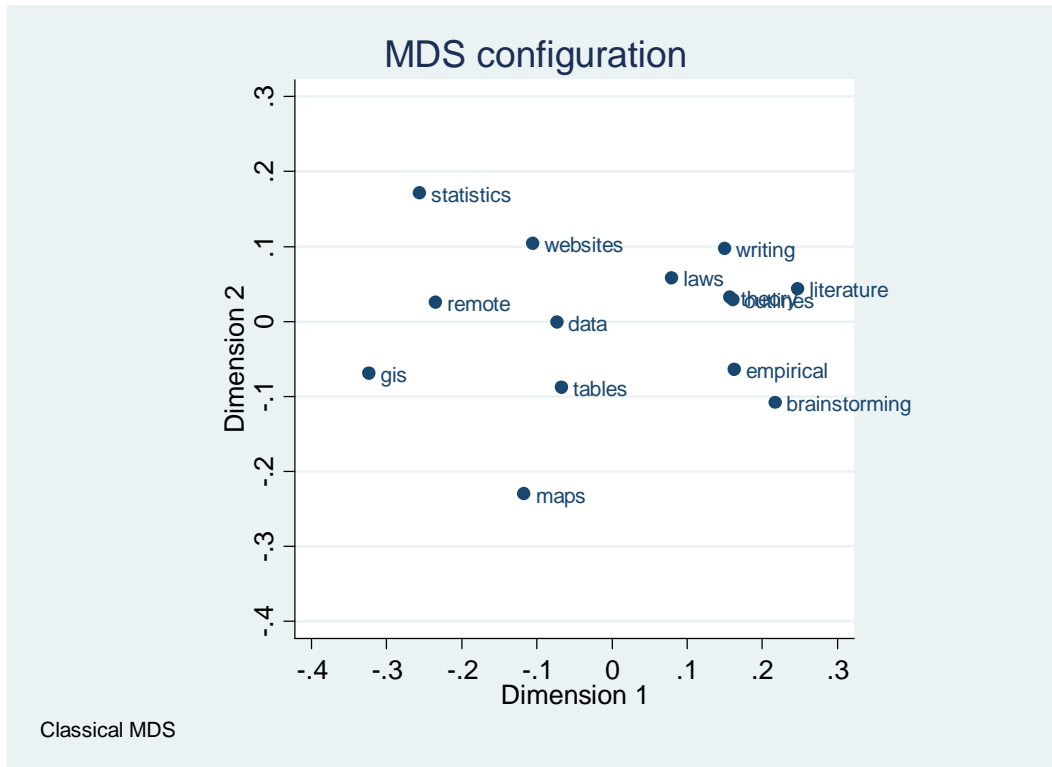
most colleagues in the west include searching for literature, and tracking down citations among their top preferences. The eastern mapping plateau includes a vast territory, and so does the land of literature to the west and northwest. Second, there are two small but high ridges, in the far east and the far west, with stunning views of the methodological possibilities of designing websites and creating statistical models. Third, prominent highlands in the southwest mark out a territory of expertise with empirical and applied policy documents, newspaper articles, and other sources. The northeastern edge of the empirical highlands is marked by a sharp escarpment of cutting-edge investigation of laws, regulations, and parliamentary and legislative records. The topography in this region is striking: recall our earlier finding that there is no sharp divide between theoretical and empirical inquiry? The empirical highlands give way at their eastern terminus to a prominent theoretical peak, which overlooks a peaceful, contemplative Thoreau-inspired landscape, a Lake Walden of writing surrounded by the lush forests and grasslands of the land of literature.

I haven't updated this cartographic interpretation for our colleagues this year. Maybe you'd like to sit down with the printout and begin drawing contour lines to help us make sense of our different interests, approaches, and expertise.

Other creative interventions are also possible. Rather than using individual's state preferences to map them in "activity space," we can transpose the variables and observations -- the rows and columns. This has the effect of mapping different activities according to the shared preferences expressed by students over the years. It takes just a few additional lines of code to change things around. Save a dataset if you've got one open, then "clear," and start over again with this:

```
clear
use c:\data\g450prefs_jan2012.dta
drop year
xpose, clear varname
mds v1-v49, measure(gower) id(_varname)
```

And now we have an alternative, complementary map of different kinds of geographical imagination and method.



Imaginative cartographies like this are inescapably subjective -- and yet also rooted in systematic ways of organizing information and measuring relationships. In turn, these measurements hinge on the decisions we make when we ask our colleagues about preferences, experience, expertise, and plans. These decisions help to shape certain aspects of the ‘reality’ we seek to observe, describe, and understand. Early decisions can have significant implications for subsequent choices and possibilities. Consider, for instance, my reluctance above to work further with the standard deviations for our colleagues’ stated preferences. I was reluctant to offer any interpretations of the standard deviations because of the measurement scale of the preferences cited by our colleagues -- each activity ranked from 1 to 14. Ranked data like these are known as “ordinal,” and they are distinctive: when you’re asked to rank things, there’s no way to know how to interpret each rank -- the difference between 1 and 2, for example, may not have the same meaning as the difference between 9 and 10. If someone really likes reading theory, investigating laws, and writing -- but passionately hates all the other activities on the list -- there will be no way to interpret the responses that my survey has forced into that ranked, 1-to-14 list. Likewise, with ordinal data, it is meaningless to say that a rank of 10 is twice as much as a rank of 5.<sup>6</sup>

<sup>6</sup> There are other ways of doing surveys like this -- using things called ‘Likert scales’ -- that solve this problem. A Likert scale would make it easy for a colleague to rank three activities as very strong preferences, and to relegate the remaining choices to a wilderness of disgust!



## Measurement and Scaling

“The phrase ‘more or less’ is a fault much in evidence in kings and geographers.”

Strabo

“The problem of measurement and scaling is the most fundamental one faced by geography and other factual sciences.”

Ron Abler, John Adams, and Peter Gould<sup>7</sup>

There are four distinct scales for observation and measurement.

**Nominal** data refer to distinct categories, with sharp differences and no clear hierarchy. An example is the set of different names for each of our colleagues.

**Ordinal** data refer to ranked information, as in the case of the research methods preferences. Analyzing ordinal data requires using a particular kind of approach called “nonparametric” statistics. The specialized measure of similarity devised by J.C. Gower is one example of a nonparametric statistic.

**Interval** data refer to numerical scales where each increment has the same meaning, and yet the values do not usually cover the full possible range. The best example is temperature, whether measured in degrees Celsius or Fahrenheit. Each degree has the same meaning -- the increment from 0°C to 5 °C has the same meaning in physical terms as the increment from 20 °C to 25 °C. But since neither the Fahrenheit nor the Celsius scales measure temperature from absolute zero, we cannot say that 25 °C is five times as warm as 5 °C. With interval data, you can add and subtract (25 °C is indeed 20 °C warmer than 5 °C) but you cannot use ratios.

Finally, **ratio** data refer to numerical scales where each increment has the same meaning, and yet the values do include the full range of possibilities. Much of the data produced by social surveys is designed to provide ratio data. One example of a variable measured on a ratio scale would be age; each unit of time has the same meaning in pure chronological terms (even if personally and metaphysically the years are felt differently at various stages of life!), and it *is* meaningful to say that someone who’s aged 20 is only half the age of someone who’s 40. *Ratio* data, in other words, allow you to safely calculate meaningful *ratios*. Most of the procedures we think of as associated with ‘statistics’ are designed to work with ratio data. This field of inquiry is known as parametric statistics; one of the most common parameters used to understand a dataset is the standard deviation.

---

<sup>7</sup> See Ron Abler, John S. Adams, and Peter Gould (1971). *Spatial Organization: The Geographer’s View of the World*. Englewood Cliffs, NJ: Prentice-Hall, quotes from p. 93.

## Another Playful Manipulation of Space

Not long ago, I wondered what multidimensional scaling might look like if we considered the strange digital footprints we're all leaving out there these days. Up to this point I had *never* Googled myself, not even once; using a corporate name as a verb has always seemed just a little bit uncomfortable.

But to "Google" another is, I guess, some sort of tribute these days, so I started with a small sample of those who have inspired me in various ways. Peter Gould was my undergraduate honors advisor at Penn State. John Adams was my Master's and Doctoral advisor at the University of Minnesota. Susan Hanson's fusion of feminist geography and the heritage of the quantitative revolution inspired my doctoral thesis -- but I was also inspired to put the work into dialogue with David Harvey's theorization of capital and residential differentiation inside the metropolis. Then I got lucky and got a job at the Center for Urban Policy Research, and one of my major collaborators was David Listokin. That led to a joint appointment with the Department of Geography, and the office assigned to me was once Neil Smith's. Neil was and is another inspiration, and at one point I had the chance to collaborate on a project led by one of his students, and we all coauthored an article on the "Camden Syndrome" diagnosing the decline of aging, inner-ring suburbs around American cities. J.K. Gibson-Graham -- the creative authorial fusion of Julie Graham and Katherine Gibson -- was also an inspiration. I was really impressed by their "How to Smash Capitalism While Working at Home in Your Spare Time" article, and I had the opportunity to bring Julie Graham for a visit and lecture to Livingston College at Rutgers University in the late 1990s. Julie passed away not long ago, and when I heard, vivid memories cascaded back of her brilliant lecture, and the conversations we had about the deeply (de)industrialized landscapes along the New Jersey Turnpike on the way to and from the airport. Then, several years later, I was fortunate to join UBC. Derek Gregory is a wonderful inspiration; he's certainly not the only colleague here who fires my imagination -- oh, we have so many brilliant and passionate geographers here, don't we? -- but at this stage my little idea was making the list a bit long and tedious. So I just added one more name -- Michael Dear -- who was then at the University of Southern California. We had great conversations when I was invited to deliver a talk at USC. Michael Dear and Jennifer Wolch even took me out on their sailboat.



**Michael Dear**, off Venice Beach, March 2008 (ElvinWyly).

This is quite a diverse cast of characters -- and this is just a tiny sample of the paradigms and personalities who have shaped my urban and geographical imagination. But what do we all look like when seen through Michael Curry's theoretical lens of the "digital individual"? Specifically, if you do a Google search of +"name" +"name" +"geography," how many shared hits do we see for each pair of names? It's a bit tedious, but with a bit of patience you can compile a nice little matrix:

Total Shared Google Hits

	Gregory	Hanson	Gould	Adams	Listokin	Smith	Harvey	Gibson-Graham	Dear	Wyly
Gregory	0	1,460	1,530	7,030	8	6,930	27,000	5,810	1,930	444
Hanson	1,460	0	611	731	41	1,060	3,480	723	149	141
Gould	1,530	611	0	5,140	5	1,340	3,030	8,920	220	86
Adams	7,030	731	5,140	0	687	16,500	2,010	289,000	184	217
Listokin	8	41	5	687	0	460	105	117	284	1,780
Smith	6,930	1,060	1,340	16,500	460	0	26,300	44,200	4,650	1,810
Harvey	27,000	3,480	3,030	2,010	105	26,300	0	262,000	5,890	5,410
Gibson-Graham	5,810	723	8,920	289,000	117	44,200	262,000	0	1,460	445
Dear	1,930	149	220	184	284	4,650	5,890	1,460	0	135
Wyly	444	141	86	217	1,780	1,810	5,410	445	135	0

*Note:* total hits from Google search of +"name" +"name" +geography, conducted January 18, 2012.

It is worth reflecting, at this point, on what these numbers might mean. The highest figures connect Adams to Gibson-Graham -- 289,000 shared hits -- and Gibson-Graham to Harvey (262,000). The latter reflects the prominence of contemporary conversations between structuralist theorists working in the tradition of political economy, and those inspired by various streams of poststructuralist thought. The large number of shared hits thus makes a lot of sense. But the linkage between Adams and Gibson-Graham is more of a puzzle. Adams was President of the Association of American Geographers some years ago, and today serves with Hanson as Editor-in-Chief of the journal *Urban Geography*. The strong connection between Adams and Gibson-Graham is surprising. Both figures (or all three) are very prominent, but in separate domains: they do very different kinds of work, and they speak to very different audiences.

This is where we need to balance our just-the-facts positivism with a sense of constructivist poststructuralism. Our inner positivist asks: *is that dominant 289,000 really number correct? Is there some sort of quirk in the search? Perhaps there is a genuinely poststructuralist John S. Adams out there somewhere, who regularly cites Gibson-Graham?* I searched through a few pages, but it was well past 2 A.M. at this point, and despite “typing like a madman” like Tom Wolfe, there seemed to be no chance that continued page scrolls through Google searches would yield anything like Wolfe’s *Kandy Kololed* brilliance.<sup>8</sup> So I stopped searching and thought about multidimensional scaling, and how these “shared hits” should be understood as *performative* phenomena. While it’s important to ask how “accurate” or “real” this 289,000 figure is, we cannot ignore the significance of people making decisions on the basis of this kind of information. If students writing term papers search for “John S. Adams,” “J.K. Gibson-Graham,” and “geography,” they’ll see this large number of hits, and it will certainly be one of many influences on their reading, thinking, and writing. Even if the 289,000 figure is just a fluke, it will acquire a certain performative reality if enough people take actions on the basis of this kind of information.

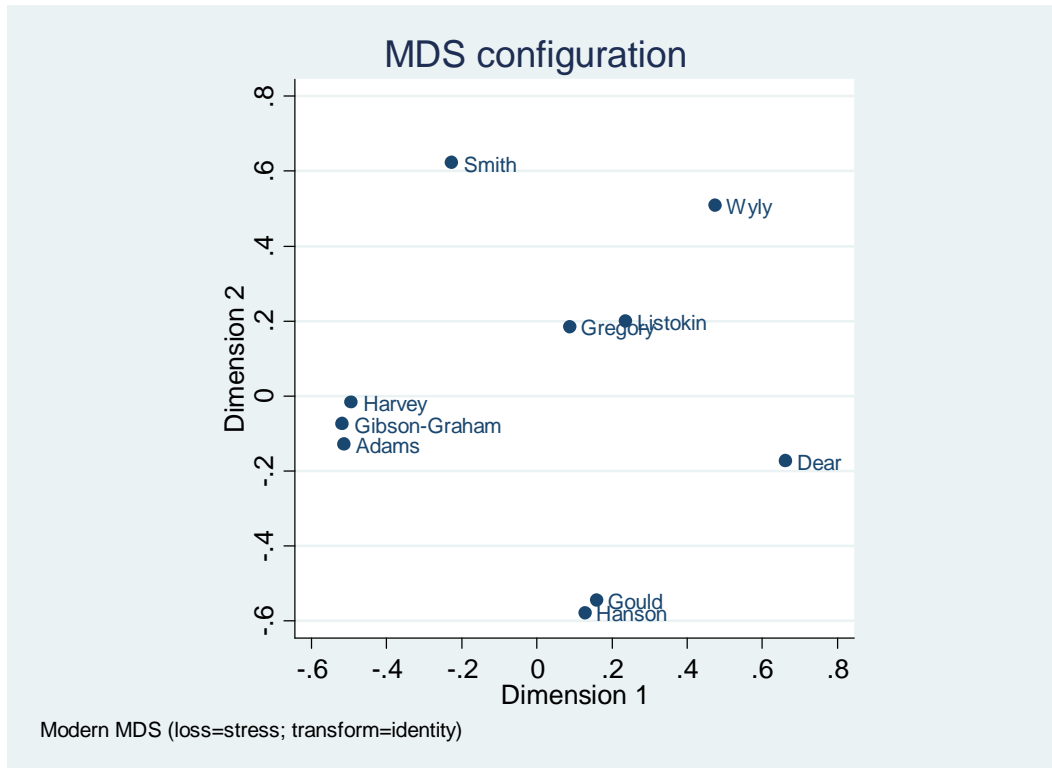
So we put down *The Kandy-Kololed Tangerine-Flake Streamline Baby*, and instead curl up with the STATA documentation and user groups. Can we do an MDS map of this matrix? As it turns out, it’s a bit more complicated than the simple example of our colleagues’ preferences described above. It was 4:00 by the time I managed to get past my mistakes and all the red-coded errors in STATA, and produced this:

```
matrix input
wyspace=(0,1460,1530,7030,8,6930,27000,5810,1930,444\1460,0,611,731,41,1
060,3480,723,149,141\1530,611,0,5140,5,1340,3030,8920,220,86\7030,731,51
40,0,687,16500,2010,289000,184,217\8,41,5,687,0,460,105,117,284,1780\693
0,1060,1340,16500,460,0,26300,44200,4650,1810\27000,3480,3030,2010,105,2
6300,0,262000,5890,5410\5810,723,8920,289000,117,44200,262000,0,1460,445
\1930,149,220,184,284,4650,5890,1460,0,135\444,141,86,217,1780,1810,5410
,445,135,0)
matrix colnames wyspace = Gregory Hanson Gould Adams Listokin Smith
Harvey Gibson-Graham Dear Wyly
matrix rownames wyspace = Gregory Hanson Gould Adams Listokin Smith
Harvey Gibson-Graham Dear Wyly
```

<sup>8</sup> See Tom Wolfe (1965). *The Kandy-Kololed Tangerine-Flake Streamline Baby*. New York: Farrar, Straus & Giroux, pp. xiii-xiv.

```
matrix wyspace=wyspace/289000
mdsmat wyspace, method(modern) s2d(oneminus) force dimension(3)
predict d1 d2 d3, config saving(digitdata,replace)
```

This is the two-dimensional map we get:



It's a curious pattern. Gibson-Graham, Adams, and Harvey are all bunched together in one part of the map, with everyone else scattered across a vast landscape to the North, East, and South. Strangely, Listokin and Gregory are placed quite close to one another, despite a total of only 8 shared hits. But recall that multidimensional scaling is similar to a map projection, with distortion as we go down from a many-dimensional hyperspace down to the simple two-dimensional summary of the page or screen. Perhaps a third dimension would help visualize a less-distorted representation of space?

For this, we need one of the add-on packages -- there are thousands to choose from! -- written by other STATA users.

We can use the net command to look for a simple three-dimensional graphics routine I read about on the user groups' bulletin boards:

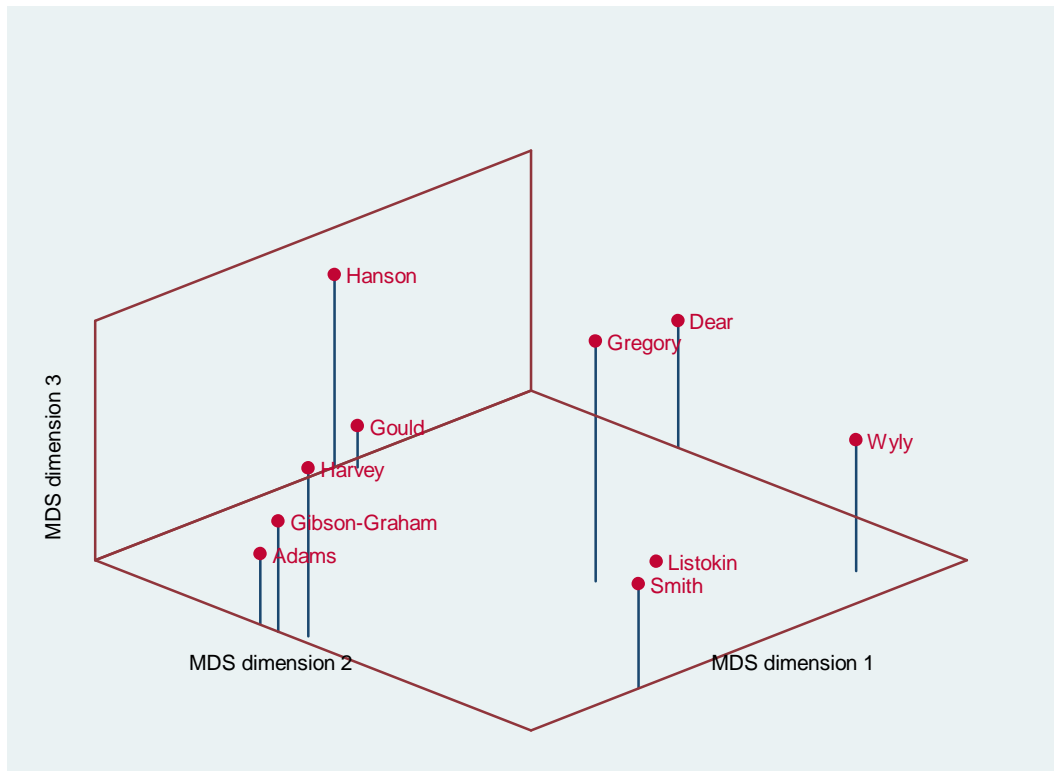
```
net search scat3
```

It gives you two options, choose the top one and it will install. Then if you issue these commands,

```
use digitdata
scat3 d1 d2 d3, mlabel(Category)
```



then you get this:



This makes a little bit more sense -- Listokin and Gregory are now clearly much farther apart when we can see that third dimension. But even so, looking across the matrix of original numbers, it's hard to make a perfect translation between intuition and the MDS results. It's not easy to think in mathematical hyperspace, especially when just a few of the digital individuals -- Adams, Gibson-Graham, and Harvey -- are such dominant avatars.

Louis Wirth would have a lot to say about those fragmented identities, which look very different depending on the circumstances in which we encounter one another in the metropolis. Urbanism is now becoming a *digital* way of life. Before most of her work was co-authored with Gibson, Julie Graham published widely on economic geography under her sole-author name. So let's go back to that original matrix of Google searches, and make that one small change:

# Total Shared Google Hits, with One Small Change.

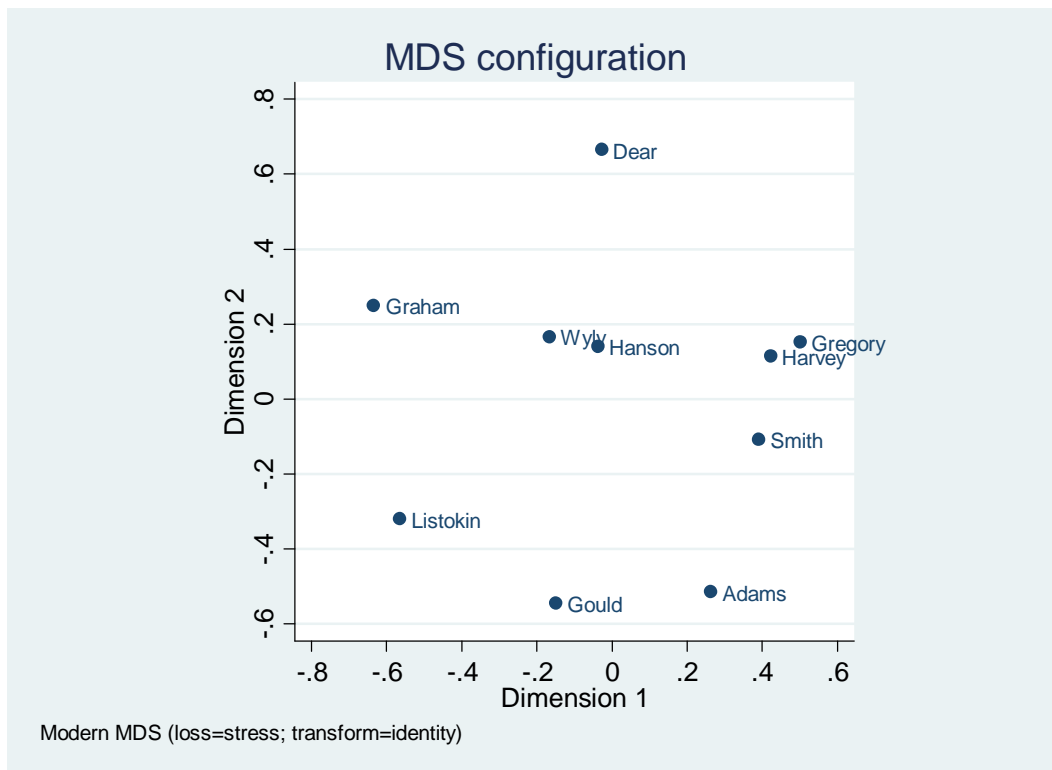
	Gregory	Hanson	Gould	Adams	Listokin	Smith	Harvey	Graham	Dear	Wyly	Hill
Gregory	0	1,460	1,530	7,030	8	6,930	27,000	366	1,930	444	3
Hanson	1,460	0	611	731	41	1,060	3,480	149	149	141	3
Gould	1,530	611	0	5,140	5	1,340	3,030	50	220	86	6
Adams	7,030	731	5,140	0	687	16,500	2,010	4	184	217	4
Listokin	8	41	5	687	0	460	105	4	284	1,780	3
Smith	6,930	1,060	1,340	16,500	460	0	26,300	522	4,650	1,810	307
Harvey	27,000	3,480	3,030	2,010	105	26,300	0	931	5,890	5,410	359
Graham	366	149	50	4	4	522	931	0	344	24	5
Dear	1,930	149	220	184	284	4,650	5,890	344	0	135	100
Wyly	444	141	86	217	1,780	1,810	5,410	24	135	0	2
Hill	3	3	6	4	3	307	359	5	100	2	-

*Note:* total hits from Google search of +"name"+"name"+geography, conducted January 18, 2012.

This small change has major consequences for what the table looks like. And since the table constitutes the constructed reality that we use when the multidimensional scaling routine kicks in, our change will result in a different map. Run the code below to see what it looks like:

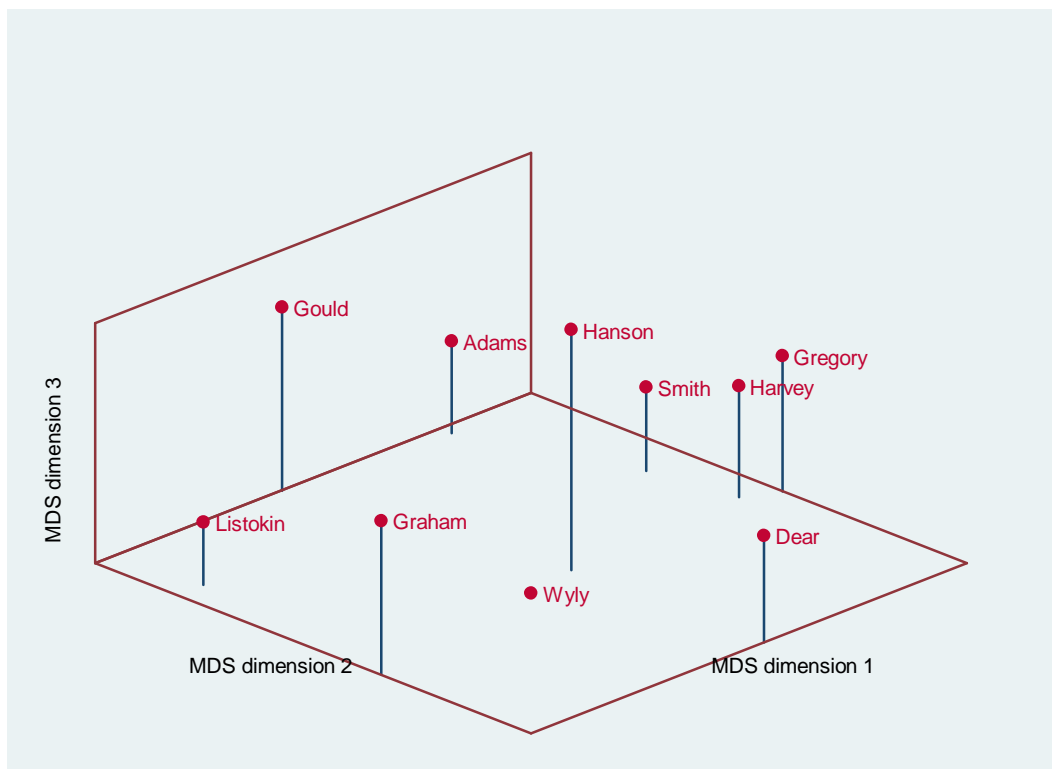
```
matrix input
wyspace2=(0,1460,1530,7030,8,6930,27000,366,1930,444\1460,0,611,731,41,1
060,3480,149,149,141\1530,611,0,5140,5,1340,3030,50,220,86\7030,731,5140
,0,687,16500,2010,4,184,217\8,41,5,687,0,460,105,4,284,1780\6930,1060,13
40,16500,460,0,26300,522,4650,1810\27000,3480,3030,2010,105,26300,0,931,
5890,5410\366,149,50,4,4,522,931,0,344,24\1930,149,220,184,284,4650,5890
,344,0,135\444,141,86,217,1780,1810,5410,24,135,0)
matrix colnames wyspace2 = Gregory Hanson Gould Adams Listokin Smith
Harvey Graham Dear Wyly
matrix rownames wyspace2 = Gregory Hanson Gould Adams Listokin Smith
Harvey Graham Dear Wyly
matrix wyspace2=wyspace2/27000
mdsmat wyspace2, method(modern) s2d(oneminus) force dimension(3)
predict d1 d2 d3, config saving(digitdata2,replace)
```

The two-dimensional map looks very different indeed:



Two more lines of code get us a three dimensional representation.

```
use digitdata2  
scat3 d1 d2 d3, mlabel(Category)
```



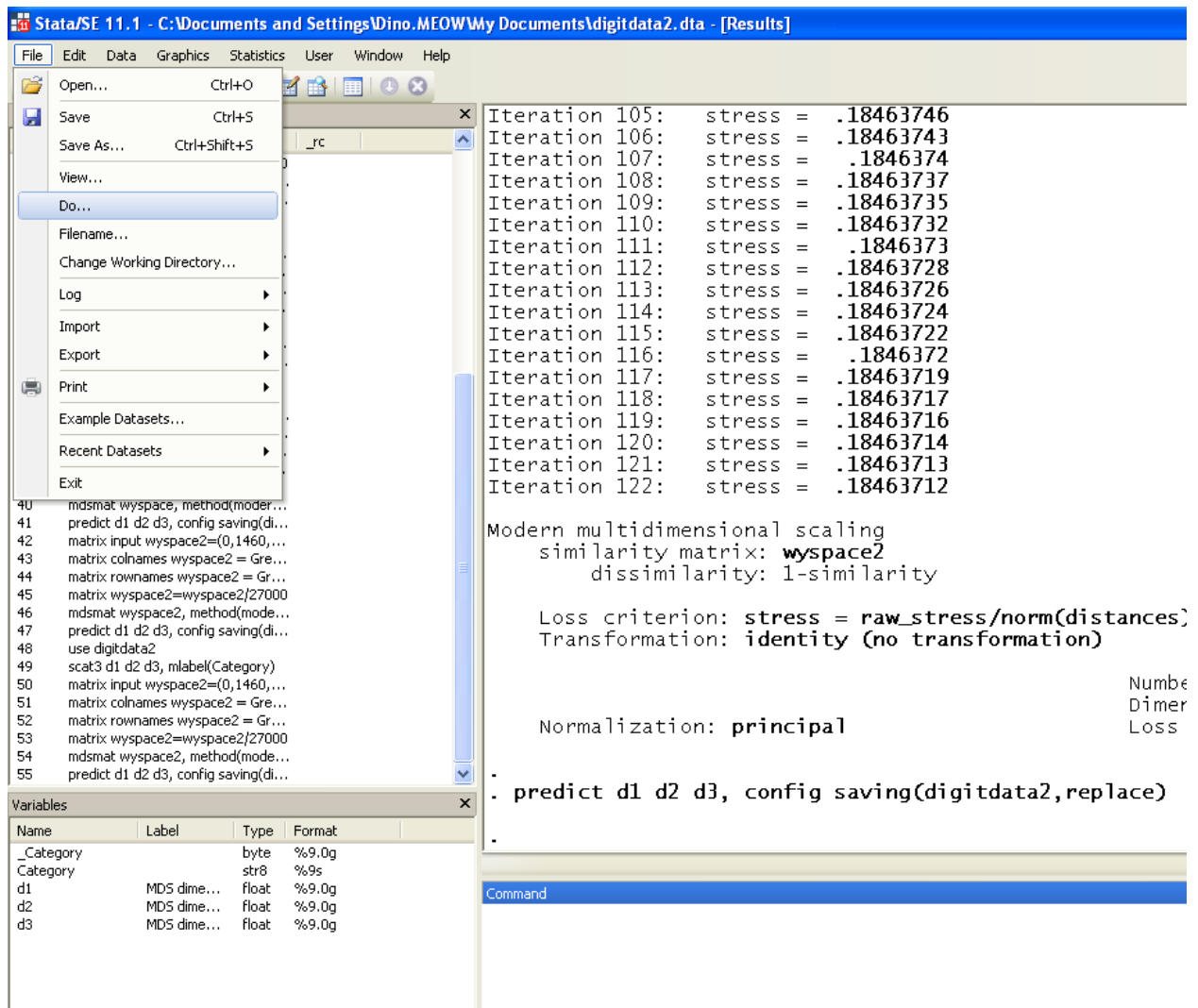


## The Do File

If you want to trace my steps in the analysis above, you could type in each of the commands to STATA, or you could cut and paste from this document into the command window. But there's another way to save time with multiple commands. This involves a "do" file, which is simply a collection of things we'd like to ask the software to do for us. You can download my do file for the first "Wyly space" example above from

<http://www.geog.ubc.ca/~ewlyly/g450/wyspace.do>

Then in STATA, choose "File/Do..." and find the file you wish to run.



The screenshot shows the Stata/SE 11.1 interface. The 'File' menu is open, and the 'Do...' option is selected. The command window displays the following results:

```
Iteration 105: stress = .18463746
Iteration 106: stress = .18463743
Iteration 107: stress = .1846374
Iteration 108: stress = .18463737
Iteration 109: stress = .18463735
Iteration 110: stress = .18463732
Iteration 111: stress = .1846373
Iteration 112: stress = .18463728
Iteration 113: stress = .18463726
Iteration 114: stress = .18463724
Iteration 115: stress = .18463722
Iteration 116: stress = .1846372
Iteration 117: stress = .18463719
Iteration 118: stress = .18463717
Iteration 119: stress = .18463716
Iteration 120: stress = .18463714
Iteration 121: stress = .18463713
Iteration 122: stress = .18463712
```

Modern multidimensional scaling  
similarity matrix: **wyspace2**  
dissimilarity: 1-similarity

Loss criterion: **stress = raw\_stress/norm(distances)**  
Transformation: **identity (no transformation)**

Normalization: **principal**

Number of dimensions: 2  
Dimer Loss

Command window output:

```
. predict d1 d2 d3, config saving(digitdata2,replace)
```

This file will take you only as far as the two-dimensional MDS map; you'll need to make sure the "scat3" package is installed if you want to proceed with the three-dimensional plot.

The second example, with Graham instead of Gibson-Graham, is at

<http://www.geog.ubc.ca/~ewyly/g450/wyspace2.do>

### **Get Creative**

This is just one brief illustration of one kind of thing we can do with STATA, and with a monstrous collection of cheap colored pens snagged at Costco.

Think creatively about the kinds of urban questions we might be able to explore through the decisions we make in measurement, scaling, and visualization. While some spatial relations are ‘out there’ waiting to be discovered and mapped through conventional approaches, other spaces are actively made by our decisions, subjectivities, and imaginations. Let’s get to work mapping and imagining the urban worlds we wish to see!

For more detailed information on STATA, consider looking at some of the resources at

<http://www.stata.com/links/resources1.html>