# Provisional Battleplan

for

## Analyzing Politics

**MEETINGS:** Knapp 105/106, WF 1:00-2:50pm
**INSTRUCTOR**: Dr. Paul Djupe
**PHONE**: 740-587-6310
**OFFICE AND HOURS:** Knapp Hall 305-B, WF 12:00-1:00, 2:50-3:30 (and by appointment)
**E-MAIL**: djupe@denison.edu
**LOCATION OF COURSE FILES**: S:\Djupe\POSC201-01

## CLASS PURPOSE

We are faced with three general, interrelated tasks in this course. The first is to familiarize you with the conduct of social scientific inquiry. Upon completion of this course, you will have taken a big step toward becoming a political scientist, one who can conduct and understand research on things political (and otherwise) and think clearly about the nature of causation. The second purpose, naturally, is to prepare you for upper level political science courses in which your professors (and you!) can take advantage of your new found research and critical thinking skills. We are learning how to become political scientists, which suggests becoming familiar with the ways political science is done and thus being able to practice those skills from here on out. Third, I argue that learning how to be political scientists also enables you to be an independent citizen, not dependent on elites, who can contribute in significant ways to holding government and other institutions accountable. Toward these ends we hope to accomplish the following:

- Come to some understanding of political science as a discipline, as something more than your coursework, and where it should be headed;
- wrestle with the social scientific method and what that means for the study of an often emotional, value-filled subject;
- gain some mastery over the tools of the political science trade, which include quantification and statistics, logic of inquiry, among other things;
- be able to apply such knowledge to published research and problems of your own design;
- attain proficiency of language appropriate for professional communication.

## READINGS

- Johnson, Janet B., H.T. Reynolds, and Jason D. Mycoff. 2012. *Political Science Research Methods.* 7th ed. Washington, D.C.: CQ Press.
- We are using the statistical program Stata 14ic(!) for this class. It is available on the lab machines in Knapp 105 and Higley basement. If you wish to have it on your personal machine, you can purchase a 6 month license of Stata IC for ~$75 or a permanent copy for ~$198.
- Additional readings will be available from journals available online (i.e., International Political Science Abstracts, J-Stor, and/or Psycinfo; you might also find the Electronic Journal Center, Consort, and Lexis-Nexis helpful in class).

## ASSIGNMENTS AND WORTH

| Assignment | Worth | Grade Distribution |
|---|---|---|
| Methodology Assignments | 75% | (see p. 5) |
| Exams | 20% | (10% each for midterm and final) |
| Class attendance and participation | 5% | |

## EXPLANATION OF ASSIGNMENTS

***Attendance and Participation*** (5%): This course demands your active participation. I will lecture infrequently. I place a significant amount of attention on participation based on the theory of learning through **apprenticeship**, which provides a great deal of directed guidance with opportunities for independent exploration. While only 5 percent of your grade is directly at stake here, your grade will suffer considerably without perfect attendance. To earn the five percent, you must attend perfectly and ask questions, suggest answers, and continue the dialogue of this course. Also see **Attendance Policy** below.

***Methodology Assignments*** (75%): Statistics has become the common language of political science. These assignments consist of actually doing political science. The content includes how to conduct research, field implementation of research methodology, interpreting statistical tables, and creating and interpreting your own statistics and graphical techniques. The length of each assignment will range from short to long. <u>Do not underestimate the time they will take you</u>. The rigor of thought and language required is most likely on a higher level than you are used to for other courses. Due dates are listed in the schedule below (pp. 4-5). The last assignment is done in competition for the *V.O. Key Award in Political Science*. Specifics of each assignment are listed on pages 5 through 24 of this battleplan.

***Exams*** (20% total: 10% each for a midterm and final): The midterm covers material from the JR book and any articles we read, while the final (a take home) is functionally comprehensive and consists of a critique of a journal article, which I will distribute during the last class.

## IMPORTANT COURSE POLICIES

***Due Dates and Excuses*** The due dates are prescribed to fit into a larger schedule. The semester is fairly evenly spaced so that the work of the course (and there's a lot) is paced and not crammed into one week. If you see a conflict coming (a campus event, 10 papers due the same day, etc.), PLAN FOR IT IN ADVANCE! Extensions are granted on a case by case basis, and not for, "I'm really busy." If you are ill, I expect a short call/email alerting me of your absence (emergencies obviously excepted, but get someone to call and make arrangements with me soon). I will not reschedule the course for you. Your late, unexcused work will not be able to earn top grades – see **Late Assignment Policy** below.

***Computer-Based Excuses*** It is <u>not acceptable</u> anymore to claim a cd, flash drive, or hard drive error for late work since the network is reliable and accessible to everyone. Therefore, use your network account (U:).

***Late Assignment Policy*** Each day (including weekends and holidays) your work is late and unexcused, you lose 5 percent from your final grade for that assignment. So: the top possible grade for an assignment 3 days late is an 85 (or B). You must alert me to your valid excuse <u>before</u> the assignment is due, not when you are handing it in late. Failure to complete all assignments results in a failing grade for the class. These administrative penalties aside, I cannot emphasize strongly enough how important it is to stay on top of assignments for this class.

***Electronic Submission*** Submit your assignments by email. Use the following naming structure for your assignments: "lastname_A#.docx". I accept doc, docx, rtf, and pdf, though I strongly prefer docx. Include all figures and tables within the same document—thus, <u>I will only accept one document for each assignment</u>. The usual late submission penalties apply to electronic submissions. Failing to attach a file is your responsibility.

***Attendance Policy*** My rule is that you forfeit your attendance points (5%) if you miss more than 2 unexcused classes (a week's worth of class). However, mere attendance is not enough; if you attend religiously and speak nary a word you will earn about two-thirds of your attendance points (that's a D). Make an effort to be an active participant in class and you will benefit, not just by earning your attendance points, but in all your coursework. You will be at a severe disadvantage should you miss the classes on statistical procedures starting midway through the course.

***Accomodations*** Any student wishing an accomodation based on the impact of a disability should contact me privately as soon as possible to discuss your specific needs. I rely on the Office of Academic Support to verify the need for accomodation based on their office files.

***Academic Honesty*** Proposed and developed by Denison students, passed unanimously by DCGA and Denison's faculty, the Code of Academic Integrity requires that instructors notify the Associate Provost of cases of academic dishonesty. Cases are typically heard by the Academic Integrity Board which determines whether a violation has occurred, and, if so, its severity and the sanctions. In some circumstances the case may be handled through an Administrative Resolution Procedure. Further, the code makes students responsible for promoting a culture of integrity on campus and acting in instances in which integrity is violated. Any incidence of academic dishonesty in this course will result in your failure of this course and referral to the Denison judicial process. There is never a problem if you keep your eyes to yourself and properly cite and reference sources of information and ideas.

***Grading Policy*** Final grades are based roughly on a 90 (lowest A), 80 (lowest B), 70 (lowest C), etc., scale. I prefer not to specify exactly what the grading scale will be so I have the flexibility to curve the scale a bit to fit the class distribution.

***Grading Expectations***

☺ A: An 'A' paper presents a clear question and justifies it, articulates clear and appropriate hypotheses, presents clear reasons for gathering appropriate evidence to test the hypotheses, analyzes the evidence in a clear and organized fashion using appropriate methods, and thinks broadly about the implications of the findings. If appropriate, it uses relevant and sufficient sources, appropriately cited and referenced.

☺ B: A 'B' paper misses at least one of A paper categories or presents a muddied version of A paper characteristics. It might be a solid piece except for doing some of the following: presents a hypothesis but may not be completely logical or organized in analyzing evidence regarding that hypothesis; has writing problems; is sloppy in presentation; skips pieces of mandatory information/analyses; has the bare minimum of sources; or fails to properly cite information. Or it may merely have a less clear question, weaker justification, etc. than an A paper.

☹ C: A 'C' paper misses several of the A paper categories. It might have no clear question, unclear hypotheses, may not investigate a question in much depth, might have logical or organizational flaws, might have significant writing problems, may have skipped multiple steps or requested information/analyses, might have the bare minimum of sources, or no consistent citation of sources.

☽ D: A 'D' paper combines serious logical flaws, superficiality, and writing problems.

☠ F: An 'F' paper fails to display much of any thought or effort on your part.

**A Note on Citation**

Any material or ideas you use from an outside source should be cited appropriately in the text and included in a works cited list at the end of the paper (you can begin this on the last page of your paper). ALL written assignments completed out of class should include citations whenever you

reference the work of another person. Citations must adhere to the Turabian/Chicago Manual of Style **parenthetical** approach with a works cited section. Note that in this approach footnotes/endnotes are not used for citations, but clarification of an idea. More information and examples of this approach can be found at:
http://www.press.uchicago.edu/books/turabian/turabian_citationguide.html.

## D A I L Y  S C H E D U L E

| Week | Topic | | Reading | Notes/Due Dates |
|---|---|---|---|---|
| 1.20 | Why study society scientifically and what does that mean? | W | Skim JR 1, read JR 2 https://goo.gl/qHnKxL http://huff.to/UrmYMg http://nyti.ms/V0pJGs http://nyti.ms/Zf7y2O | |
| | | F | JR 3 (just the first part on research questions) & JR 4 (CVH) Zaller, John R. 1998. "Monica Lewinsky's Contribution to Political Science." *PS: Political Science and Politics* 31(2): 182-9. | |
| 1.27 | Measurement, Interviewing and the Art of Survey Construction | W | JR 5 (measurement) Bring Zaller again. Mendez, Jeanette and Tracy Osborn. 2010. "Gender and the Perception of Knowledge in Political Discussion." *Political Research Quarterly* 63(2): 269-279. | |
| | | F | JR 10 (interviewing) and Qualtrics introduction (preparation steps tbd) Pasek, Josh and Jon A. Krosnick. 2010. "Optimizing survey questionnaire design in political science: Insights from psychology." In Jan Leighley, ed. *Oxford Handbook of American Elections and Political Behavior.* New York: Oxford University Press. https://pprg.stanford.edu/wp-content/uploads/2009-03-Leighly-Chapter03.pdf | |
| 2.3 | Research Design | W | **A1 due in class (but email me a copy from one team member)** Qualtrics introduction, continued | |
| | | F | JR 6 (design) | |
| 2.10 | Finish (not Finnish) Design; Sampling; Literature Reviews | W | Design, continued Maltzman, Forrest, Lee Sigelman, and Sarah Binder. 1996. "Leaving Office Feet First: Death in Congress." *PS: Political Science & Politics* 29(4): 665-71. Hetherington, Marc J., and Jason A. Husser. "How Trust Matters: The Changing Political Relevance of Political Trust." *American Journal of Political Science* 56(2): 312–325. Bloom, Pazit Ben-Nun and Gizem Arikan. 2012. "Priming Religious Belief and Religious Social Behavior Affects Support for Democracy." *International Journal of Public Opinion Research.* http://bit.ly/17fqeRf | |
| | | F | JR 7 (sampling), JR 3 (the lit review back half) [for sampling] Birch, Sarah. 2010. "Perceptions of Electoral Fairness and Turnout." *Comparative Electoral Studies* 43(12): 1601-1622. Re-read Bloom and Arikan (2012) for sampling as well. | |
| 2.17 | Observational issues, and Qualitative techniques; Data Presentation | W | JR 8 (observation), JR 9 (content analysis) [for content] Sigelman, Lee and Emmett H. Buell. 2003. "You Take the High Road and I'll Take the Low Road? The Interplay of Attack Strategies and Tactics in Presidential Campaigns." *Journal of Politics* 65(2): 518-531. | |
| | | F | Data presentation, reading tbd | **F-2.19 A2 Due at Midnight** |
| 2.24 | Oral Lit Review of | W | First half | **W-2.24 & F 2.26 A3 will be** |

| | | | | |
|---|---|---|---|---|
| | Controversies | F | Second half | **performed** |
| 3.2 | MIDTERM! | W | MIDTERM! | **W-3.2 MIDTERM** |
| | | F | Intro to Statistics, reading tbd | |
| 3.9 | Enter the World of Stata | W | Intro to Stata | |
| | | F | Begin A4 | |
| 3.16 | SPRING BREAK | | NO CLASS | |
| 3.23 | Re-intro to Stata and A4 | W | Lab work | |
| | | F | Lab work | |
| 3.30 | A5 | W | Lab work – intro to regression | **Tue 3-29, midnight, A4 Due** |
| | | F | Lab work – more regression | |
| 4.6 | ¡MPSA! | | I don't know if we have class on Wednesday or not yet. I'll be in Chicago for the back end of the week, presenting a project with Ben Hsiung's ('16) that jumps off his summer research: Djupe, Paul A. and Benjamin O. Hsiung. 2016. "How Religious Styles Promote Policy Threat: The Case of Religious Liberty." Paper presented at the annual meeting of the MPSA, Chicago. | |
| 4.13 | Begin A6 | W | Lab work | **Tue 4.12, midnight, A5 due** |
| | | F | Lab work | |
| 4.20 | A6 in earnest | W | Lab work | |
| | | F | Lab work | |
| 4.27 | Wrap A6 Final Exam Review | M | Lab work | **Wed 4.27, midnight, A6 due** |
| | | W | Lab work | **Fri 4.29, Hand out final article** |
| | | F! | Discuss sample journal article (tbd); discuss VO Key findings and hand out Award; hand out final article to review. | |
| | **Final Due: Saturday, May 7, 4pm (by email only)** | | | |

**\* Note:** The days on which I will introduce new procedures are not set in stone, but will occur under days now marked as "lab work." Their timing entirely depends on how quickly the class is moving. Don't miss these days unless you are actually dying.

# ASSIGNMENT PACKET

### METHODOLOGY ASSIGNMENT SUMMARY (100 PTS TOTAL; 75% OF GRADE)

| No. | Assignment | Points | Due Date |
|---|---|---|---|
| 1 | Team Measurement and Sampling | 5 | **Wednesday 2.3 (bring results to class + email a copy)** |
| 2 | The research design section | 15 | **Friday 2.19 midnight** |
| 3 | Presentation on Data Controversy | 15 | **Wednesday 2.24 & Friday, 2.26** |
| 4 | Telling a story with basic descriptive and inferential statistics | 15 | **Tuesday 3.29 midnight** |
| 5 | Multivariate statistical analysis | 15 | **Tuesday 4.12 midnight** |
| 6 | Research paper | 35 | **Wednesday 4.29 midnight** |
| | *Total=* | 100 | |

## Important Notes

o Submit all assignments by email, but bring your results for A1 to class; one file per assignment only.

**ASSIGNMENT 1**
# MEASUREMENT AND SAMPLING EXERCISE
### (TO BE COMPLETED IN TEAMS OF 2)

*When you can measure what you are speaking about and express it in numbers,*
*you know something about it; but when you cannot measure it, when you cannot*
*express it in numbers, your knowledge is of a meager and unsatisfactory kind.*
(Lord Kelvin, 19th century English physicist who discovered the 2nd law of thermodynamics)

## ROADMAP

Answering questions, conducting research, depends heavily on proper measurement. Proper measurement incorporates both accuracy and precision—measuring what you want to and doing it carefully, respectively. Clearly, accurate and precise measurement depends heavily on definitions. What constitutes the concept, problem, or target within your area of interest? For instance, what is political participation? Is it directly targeted at government or would any activity that involves politics like political discussion count? Is giving money to a candidate or campaign the same as direct expression of views like protest? Is participation in official channels like voting the same as protest? Start with the concept in its ideal form and then compose a measure that is a reasonable approximation given your (limited) resources.

Successful measurement also depends heavily on the instrument used. For instance, to measure distance, we are fortunate enough to have rulers with decent quality control governing length. In the study of human behavior, we are not so lucky. Though we have some direct measures, we mostly rely on indirect measures based on perception, self-reports, reports of others, etc. Without the quality controls imposed on instruments capturing distance, how can you be certain you are measuring the right amount or even the right thing?

Intertwined with proper measurement is proper sampling technique. You may want to measure how all voters feel about welfare reform, but such a task is beyond even the most wealthy researchers (the federal government can't even count everyone!). Good sampling allows you to take accurate measurements of very large, cumbersome objects, such as American voters, interest groups, markets, federal bureaucracies, or countries. We will come to sampling in more detail later, so for now concentrate on why you choose certain items to measure—of what are they representative? By that choice, what are you choosing not to measure and is that selection important?

## SPECIFICS

You are to answer the following research questions, paying close attention to your measurement and scheme for inclusion of items. Provide operational definitions for each concept in question and justify that definition: what are you going to measure and why did you choose that definition? Do you introduce or exclude any biases by using that definition? Then answer the question using the data you collect based on your definitions. Show your data in a way that illustrates your conclusions. Divide up the tasks as you see fit, or do them all together. Group members receive the same grade for the assignment (unless a member fails to do her share). Turn in one typed paper (3-4 pages total, including tables) incorporating all of the 4 elements. It's essential to provide tables (or figures) describing the data you used to reach your conclusions.

- How many doors are on the Academic Quad?
- Which pizza place gives a better deal, Papa John's (hurl!) or Domino's (hurl$^2$!)?
- Which grocery store is cheapest, Kroger or IGA?
- Which school provides a better education, Denison or Kenyon?

NOTE: If choices above, e.g. Papa John's or Kroger, don't work for you, choose ones that do.

Paper Due **WEDNESDAY, FEBRUARY 3 (email it to me and bring a copy to class)**

ASSIGNMENT 2
# The Research Design Section

The purpose of A2 is to 1) gain a working knowledge in the use of Qualtrics and 2) gain a working knowledge of how to compile the research design section of a research paper. To maximize the use of Qualtrics and to write a more detailed research design, you will choose one of these experiments. I'm choosing experiments because they generally require much more detailed design sections than other kinds of research (why?).

Option 1 – Americans have classically been noted to evince strong support for civil liberties like free speech and are not shy about their prejudice for controversial groups. These two do not mix well and Americans' support for the civil liberties of controversial groups is often meager. But perhaps Americans simply need to be reminded of Constitutional imperatives and their supposed commitments. Does framing a controversial activity by a controversial group in terms of civil liberties increase the expression of political tolerance?

Option 2 – Evaluating the work of a politician is a complex task. Presidents, for instance, have extensive portfolios of issues on which they advocate and their acheivements and failures may have little to do with their own actions, such as when it concerns the economy. The complexity of the job makes evaluation all the more difficult and, therefore, subject to variation. Does presidential approval vary upon the information individuals are primed with? Or does approval reflect ideological consistency?

Option 3 – The new Pope Francis has made quite a stir with his apparent more inclusive and open approach to certain issues and groups than his predecessors. How far does goodwill toward Pope Francis extend? Does Papal priming$^©$ affect evaluations of the Catholic Church in the US? Does Papal priming affect the support for the standard issue agenda of the American Catholic Church?

By the time of our introduction to Qualtrics, you will have your own Qualtrics account (in fact, you already do, you just don't know it yet). The first order of business is to develop your own survey that explores the question you choose from the above list of three. Your survey will have randomization, naturally, since it has an embedded experiment. As with most experiments, it does not have to be long, but it does need to be well developed, using industry standard measures. The survey that you supply to me will be annotated (use the comment feature in Word or something like it) to record the source for each measure and why the measure is asked.

You will write a research design section of a research paper. Make sure it opens with the research question. Justify the choice of an experimental design – what advantages and disadvantages do experiments offer? Then segue to talk about your particular design and the nature of the treatment(s). Show how the treatment(s) test your hypothesis or hypotheses. Then discuss what is needed to assess whether your design was successfully implemented (what does that mean?) on the sample (and sample size) you specify. You'll finally spend some time with measurement, discussing the nature and source of your DVs (and their validity) and any other (very few) variables that come into play in the study. As part of the design or wherever it comes up, walk through survey construction issues.

Experiments always take a bit longer to describe than other designs since more of the analytic power comes with the design rather than in the analysis as is true for non-experimental work. I suspect that

this will take 5 pages or a bit less to write up. The annotated survey will probably be 2 pages. Make sure to include the preview link in the exported survey (a word doc).

The write up is due **Friday 2.19** at midnight.

---

**ASSIGNMENT 3**
ORAL LITERATURE REVIEW OF CONTROVERSIES

The purpose of this assignment is to orally present a literature review. Since reviews are designed to lay out a controversy in a literature, you will discuss the controversy, its importance, and articulate the implications of different theories, research designs, measurement schemes, data sources, etc. – there are multiple ways in which a literature may vary and thereby introduce controversy. These presentations will take place on Wednesday 2-24 and Friday 2-26, giving you ample time to research them and create an informative presentation. You should plan to have either handouts or some sort of visual display to help tell your story. Plan to speak for about 10-15 minutes, and not a second longer about your controversy. We'll sign up for these a short way into the semester.  You will present these in teams of two, working together and sharing the burden as equally as possible. I encourage you to report shirking. Moreover, if presentation performance is markedly different, I reserve the right to assign individualized grades. I'll have more to say about this in class.

**¡¡YOU SHOULD TALK TO ME ABOUT THIS SO WE'RE ON THE SAME PAGE!!**

| Date | Controversies |
|------|---------------|
| 2.24 | 1. Is Islam compatible with democracy? |
| 2.24 | 2. Does exposure to disagreement demobilize people? |
| 2.24 | 3. What's a democracy? How democratic is the U.S. compared to others? |
| 2.24 | 4. Does racism shape public opinion? |
| 2.24 | 5. Would liberalizing voter registration laws boost turnout? |
| 2.26 | 6. What is the effect of abstinence-only sex education? |
| 2.26 | 7. Do political advertisements matter? |
| 2.26 | 8. Are judges rational political actors? |
| 2.26 | 9. Is there a crisis in Social Security funding? |
| 2.26 | 10. Do democracies fight? |

**GIVEN WEDNESDAY 2.24 AND FRIDAY 2.26**

**ASSIGNMENTS 4**

# What States Regulate Religion (And Why)?

In addition to freedom to associate, speak, and publish, liberties for religion and conscience stand at the core of free societies. Without the ability to believe in and practice religious dictates, people are put in a near impossible position with regard to the state. That is, it is difficult to see the state as legitimate and participate within it when the state does not reciprocally recognize the legitimacy of a group's belief system. This is not just a problem that minorities face, as some states regulate majority religions in an attempt to maintain purity as well as restrain the power of an organized religious group. These concerns make understanding the degree and spread of religious regulation essential in the project of extending freedom and democracy around the world.

We will draw on two datasets to explore these concerns while building our descriptive and inferential skills. The International Religious Freedom Dataset, collected by Brian Grim and Roger Finke, come from the American Religion Data Archive[1]. They coded US State Department country reports, produced in response to Congressional mandate about a decade ago. The website describes the project in some more detail, while the codebook (IRFD_codebook.docx) has some further variable description (see also Grim and Finke 2006 – in our S: A4 folder).

Given that we're political scientists, we might want some governmental measures, such as what type of government describes the state. I went to one industry standard measure – Polity IV.[2]  Data are available back to the early 1800s in some cases, but we're going to focus on 2008 to match the IRFD data availability. There's a well developed "manual" about the Polity IV (aka p4) in our S: A4 folder. The resulting dataset is in our S: folder called "irfd_08_merged.dta".

| **Stata Notes** | You should open Stata first, then open your desired data and do files. |
|---|---|
| | I will put menu directions in all caps. Mostly we won't be using menus, so there should be little of this. |
| | I will bold face variable names. |
| | Stata is capitalization sensitive. V4 is different than v4. The default among experienced users is to use lowercase. |
| | All Stata commands will be preceded by → |
| | You will want to use and maintain well annotated "do files." Click   to start a new one. Keep it all in the do file. Keep your code. Stick results in there. Take notes about the code and results. |

There's a composite religious regulation measure in the dataset called "GRI_08". I really don't like capitalizing variable names, so I'm going to change it:
→rename **GRI_08 gri_08**

By the way, if you had a dataset that came in with all capitalized variable names, you can make them all lowercase this way (the * is a wildcard that includes everything in this case):
→rename *, lower

---

[1] http://www.thearda.com/Archive/Files/Descriptions/IRF2008.asp
[2] http://www.systemicpeace.org/inscrdata.html

**Descriptive Results**

First, we need to get a sense of this variable (**gri_08**). We can do that in a variety of ways. Try all of them and report the ones that help tell the story the best (what's best?).

| | |
|---|---|
| → tab1 **gri_08** | [frequencies of responses] |
| → sum **gri_08** | [mean, standard deviation, etc.] |
| → hist **gri_08** | [histogram – you can vary the number of bins which defaults here to 25. To do that, use this and put whatever number you want in the bin parentheses: hist **gri_08**, bin(10) ] |
| → graph box **gri_08** | [the white line is the median, the box represents 25$^{th}$ and 75$^{th}$ percentiles, the whiskers extend to the next values, and the dots beyond are simply observations outside of that range] |
| → graph hbox **gri_08** | [hbox shifts the direction to horizontal] |
| → gen place1=1 | [It's useful to have placeholder variables now and again] |
| → scatter **gri_08** place1, jitter(10) | [place1 is simply a constant=1; the jitter can be adjusted up or down] |

I want us to try one more graphical presentation, but we need to make some new variables first.

| | |
|---|---|
| →egen grim=mean(gri_08) | [makes a variable containing the mean of v28_1 – it's a constant – tab1 it if you don't believe me] |
| →egen grisd=sd(gri_08) | [makes a variable containing the standard deviation of v28_1 – it's similarly a constant] |
| →gen gri_low=grim-grisd | [makes the mean minus the standard deviation] |
| →gen gri_high=grim+grisd | [makes the mean plus the standard deviation] |

Now we can graph this! The command is all four of the following lines and can only be done through a do file. Read the notes after it for a full explanation.

→twoway scatter gri_08 place1, mcolor(gs14) jitter(20) ///
|| scatter grim place1, mc(black) ///
|| rcap gri_low gri_high place1, lc(black) lw(medthick) ///
||, yscale(range(0 100)) ytitle("Religious Regulation") xtitle("")

That requires some explanation!

Overview: "Twoway" is a graph type in Stata that, somewhat obviously, shows some graphical representation at the junction of two values in a graph space. The "///" at the end of each line tells Stata to keep going to the next line. These are not necessary, but they often help me keep track of all the components in a complex graph. You can use them for any command, actually. The "||" tells Stata that this is the next graph to pile on.

Graph 1 (line 1) – "Scatter" is the mother of the twoway group and this one scatters all values of **gri_08** over the constant **place1**. It's weird to scatter over a constant, but this means you can scatter across groups (think government type) which will work well for you in the next part of the assignment. I tweaked the marker color (mcolor or mc) so it would be a lighter color (type "help colorstyle" for a list of color options). And then I included a "jitter" so the individual dots could be a bit better distinguished from one another. The code also runs in order, so this scatter will occupy the background (hence why I wanted a lighter color).

Graph 2 (line 2) – This simply places a marker at the value of **grim** (the mean of gri_08). I want to see the mean value in the midst of that scatter and changed the color of the marker (mc) to black so it stands out.

Graph 3 (line 3) – This one uses the variables (**gri_low/high**) we just created to draw a capped line (rcap) one standard deviation out from the mean. I changed the color of the line (lc) to black and the thickness of the line (lw) to medium thick so it stands out. You could also pick an "rspike", "rcapsym", or "rbar" style for this graph. I'm partial to the rspike, fwiw.

Global options (line 4) – I decided to mandate that Stata display the full possible variable range (hence the yscale command). If you ran the rcap graph by itself (twoway rcap gri_low gri_high place1) it might cut off at less (it's up to you how to display this). So, it's useful to keep this in your back pocket in case you want to show a different portion of the range.  I changed the title for the y axis and eliminated the x axis title by including blank quotes. I could even eliminate all labeling for the x axis by adding the command "xlabels(none)" in that last line – I actually like that better in this case.

Other options: I can't stand the Stata default light blue-gray background. I change it to white always: graphregion(color(white)). You'll notice that the legend is less than informative without some further labeling. We can either do that in the graph command with the "order" subcommand:
        legend(order(1 "Something" 2 "Something else" 3"Something in addition"))
In this case, I like the first label, so I want to keep that and will tweak the rest. In that case, it would look like this (note, you always need a space after the number in this command):
        legend(order(1 2 "GRI mean" 3 "+/- 1 SD"))

A composite last line to this graph might look like (make sure to put "///" on the line above):
    legend(order(1 2 "Mean" 3 "+/- 1 SD") col(1)) graphregion(color(white)) xlabels(none)

This graph is probably something you could include in a paper at this point. It's clean. However, if you wanted to tweak small things you can use the graph editor described in the note below.

**Stata Notes**   When the graph is open, you can edit it from within the pop-up window. I'll walk through this in class.
Make sure you save it. Save a Stata version (.gph file) so you can go back and edit it quickly as well as one suitable to stick in a Word file (.emf or .png) – you'll have to do a "FILE-SAVE AS" for that. While you can add text for a title or notes while in Stata, I'd skip it and just do that in the Word file.

**What to Write**
You may have lost the thread here a bit. You were trying a bunch of different descriptive statistics and presentations to showcase the variation religious regulation around the world. What works best to tell that story? Now to what you might want to write. You need to start this paper with an introduction that pitches a research question. This should address the classic concerns of research question statements as we've discussed in class. It should be followed by a brief statement of a paragraph or two laying out your hypotheses of what shapes religious regulation. In longer papers, this is where you would develop a theory in a lit review setting that builds toward hypotheses. In this short paper, we'll cut to the chase. That is followed by a research design statement – how are you going to answer your research question? With what design, sampling method, data gathering

instrument, and measurements? What limitations are there to the data collected? This need not be long – perhaps a good long paragraph. That's followed by the results. Each of these chunks should have a heading in your paper. Bold face that text and give it its own line.

Results sections almost always begin with a depiction of the DV with the intention of 1) showing that there is actual variation to explain and 2) describing the extent of the variation. How you demonstrate that is your choice here. I gave you almost a dozen different ways of doing it.

The descriptive section is followed by a section that tries to explain the DV by testing your hypotheses. For this paper, we're keeping the analysis very simple to work on our data presentation skills. Let's turn to that now.

**Inferential Results**
Now we need to test your hypotheses. We're going to focus on a few ways of displaying relationships through tables and figures and you'll decide which ones you want to use. I want you to see how religious regulation varies by polity type (**polity**). You are going to pick at least 2 other sensible measures to test for effects on religious regulation. Let's start with graphs and it will become clear that we need to do something different to make tables.

You can choose whatever other IVs you want. I would recommend **msri_08**. It might also be interesting to see what relationship there is to being a victim of terrorism (**teror108**). I might also be tempted to engage the first answer in political science using **lg1pct08** (the percent of the population belonging to the largest religious group – this is thus a measure of diversity).

First, let's see what **polity** looks like. What is polity anyway? Check out the codebook.
→ tab polity
Most of the scores range from -10 to +10. What are those -66 and beyond numbers? You'll find out that we want to get rid of them for our purposes. Let's recode them to missing:
→ recode polity (-99/-66=.a), gen(p4)                    [".a" is a missing value in Stata]

It would be nice to have some labels attached to **p4** scores:
→label define p4 -8"Autocracy" 0"Anocracy" 8"Democracy"    [this creates a labeling scheme called p4, attaching some text in parentheses to the number out front]
→label values p4 p4                                       [applies to a variable the label scheme]

So, for instance, we can now check out how the polity score relates to religious regulation.
→ scatter gri_08 p4, xlabels(-10(2)10, values) graphregion(color(white))

We could just summarize the amount of regulation by government type. First, let's collapse down the government type:
recode p4 (-10/-6=0)(-5/5=1)(6/10=2), gen(p4_3)          [p4 is often broken down into 3 categories matching our label scheme and using the break points I just put in parentheses here]

Can you tweak the "label define" code above to apply to our new **p4_3** variable?

Maybe we could do the same scatter routine above (with means and standard deviations) but across the three polity categories. The code is only slightly different and that one change gives this a LOT of power (adding "by").

→ egen p4_3_grim=mean(gri_08), by(p4_3)
→ egen p4_3_grisd=sd(gri_08), by(p4_3)

→ gen p4_3_grilo=p4_3_grim-p4_3_grisd
→ gen p4_3_grihi=p4_3_grim+p4_3_grisd

→ twoway scatter gri_08 p4_3, mc(gs10) jitter(10) xscale(range(-.25 2.25)) xlabels(#3, val) ///
        || scatter p4_3_grim p4_3, mc(black) ///
        || rspike p4_3_grilo p4_3_grihi p4_3, lc(black) ytitle("Religious Regulation") ///
        xtitle("Polity Type") legend(off) graphregion(color(white))

I don't like it, but you could do a bar graph.
→ twoway bar p4_3_grim p4_3, barw(.5)

Want to just show a line connecting the means? Here are two options
→ sort p4_3 // for some reason, you need to sort the x axis variable first
→ twoway line p4_3_grim p4_3
→ twoway connected p4_3_grim p4_3 // minor tweak to use the "connected" option

Now what if we wanted to show a simple old cross-tabulation.
The general form of cross-tabulations will be:
→tab2 *varname1 varname2*              [the first var appears in rows, the second in columns]

It is conventional to put the IV in the rows and the DV in the columns (if you can identify an IV and DV, which is not always the case). There are a number of options that are available that you'll find useful. The first is to give percentages beyond raw frequencies and let's translate the general form to the specific form required to test how class prep varies by gpa:
→ tab2 **p4_3 gri_08**, row chi2 nofreq ["row" gives percentages going across in the rows, "col" (not shown) gives percentages going from top to bottom. "chi2" gives you a chi-square ($\chi^2$) test. "nofreq" shuts off the raw numbers.]

Oh man. What a mess. We need to collapse down the **gri_08** measure into something manageable for a table. This shows you how inflexible tables can be – you sometimes have to throw away variation in order to make it presentable. Here are two strategies for simplifying **gri_08**.
→ recode gri_08 (0/3.3=1)(3.31/6.6=2)(6.61/10=3), gen(gri_08_3)
→ xtile gri_083x=gri_08, n(3)

In the first, I have very tight control over what goes into what category. In the second, Stata does the work for me, trying to create 3 equally sized categories (hence "n(3)"). How do they stack up? There is a tension here between theoretically meaningful categories and statistically meaningful categories. You can tweak this code as much as you want. But try this then:
→ tab2 p4_3 gri_08_3, row chi2

→ tab2 p4_3 gri_08_3x, row chi2    [some options for tabs: "row" gives percentages going across in the rows, "col" gives percentages going from top to bottom. "chi2" gives you a chi-square ($\chi^2$) test. "nofreq" shuts off the raw numbers.]

Here's an alternate table format to see if values of one variable move with another. Put differently, this gives you the mean values of one variable for each category of another. It's called a tab-sum and it's particularly handy if one variable has a lot of categories and thus is more amenable to computing averages:

→tab p4_3, sum(gri_08)

Now that you've tried all of these options for presenting relationships, you get to choose which one is the most effective presentation style (or styles). I'd like to see that you've tried them all in class, but you need not hand them all in, which also means that they don't all need to be super polished unless you're handing them in. You also need not replicate this search process for the other two independent variables you want to test.

**What to Write**
Make sure to describe the test, point to the table/figure where the results can be found, tell us the lesson of the table/figure, and hedge your lesson as much as needed. Then move on to the next test. For your conclusion, you will naturally reiterate a bit of what you found, but be sure to push beyond a summary to think big. What are the big picture implications of your findings?

All told, I suspect that this paper will be about 6-7 pages of text, a page appendix that describes how variables used are coded, and numerous pages of tables and figures. You may not have any references, but if you do they should be placed after the appendix and before the tables and figures.

Let's aim for Sunday, March 27 at midnight as a due date. Submit 1 document per person by email only.

## A5
## What explains religious regulation?

Yes, this is a continuation from A4. We are staying more closely connected to the A4 data in the hopes that this new technique comes more easily to you. The introductory material to this assignment is therefore the same and we are iterating the question with a new technique in the toolkit: OLS regression.

The goal is the same, too – to explain why states regulate religion at different levels. You recognized in A4 that while you could see relationships in the data, they were messy until you took into account multiple variables. This technique (OLS regression) allows us to do that in an extremely powerful way.

We'll skip right to the heart of things, recognizing that the activities you are doing in this assignment come directly after what you just did in A4 (well, in between the results and the conclusion of A4).

To estimate your model of religious regulation, you would do the following (where *varnames* is a list of all of your independent variables):
→reg gri_08 *varnames*

For all variables you include in a regression model, they must:
1) Have a linear trend (no nominal variables, unless they're dichotomous);
2) Have missing values set to ".a";
3) Have to make sense in this model (i.e., you have a good, stated reason for them being there – as in there is a hypothesis);
4) Are coded in a way that makes sense to readers at first glance (e.g., is democracy coded so that a higher number is more democratic?).

Make sure you interpret the quality of the model (model statistics) as well as the significance, direction, and strength of each variable of interest. Recall, you don't discuss the direction and strength of the variable unless it's significant.

Now, here are a few tips for showing the effect of your variables. Of course, one way is to discuss what is significant, but that doesn't tell you the strength. Strength is the size of the effect – how much the DV is changed by movement across the IV. The slope coefficient tells you this – it shows how much the DV moves for a one point change in the IV. Thus if your IV is coded from 0-1 you can see the full effect right in the slope coefficient estimate. If there are more categories to your IV, then you need to do some multiplication (or recode the range to be 0-1 as described below). Strength can be thought of as the effect across the full range of the IV:

$$\text{Strength} = |\text{slope}*highvalue - \text{slope}*lowvalue|$$

"|" stands for the absolute value. Another simple way is to have Stata do this for you using the margins command. Margins can be run after a regression model to provide estimates from the model using combinations of values of the IVs you specify. Let's say you wanted to examine the strength of the polity variable. It would look like this (assuming it's still coded -10 to +10):
→margins, at(polity=(-10 10))
This will pop out two values – the estimate for the DV when efficacy is -10 and the estimate for the DV when polity is +10. You can see (and calculate) the difference in predictions and that's the "strength" of polity. You could also see the prediction for each value of polity like this (the "(1)" I added below tells Stata to go from -10 to +10 using 1 step increments):
→margins, at(polity=(-10(1)10))

Yes, this can be graphed! And now that we have Stata 14 (and not 11) we can use one simple, magical command:
→marginsplot

What's neat about the margins setup is that you can use combinations of variables in there. So, let's say that I wanted to see what the effect of religious majority size looked like for democracies and autocracies (assuming **p4_3** is coded 0=auto, 1=ano, 2=demo). First, I have to run a different regression model that includes **p4_3**:
→reg gri_08 p4_3 lg1pct08

Then to see the effect of the two variables together, this is the code:
→margins, at(lg1pct08=(0(10)100) p4_3=(0 2))

It should pop out 20 estimates – 10 values of religious majority size for autocracies and then democracies. Yes, this can be graphed!
→marginsplot

Of course we can do all the same stuff we would with twoway graphs in terms of labeling etc.:
→marginsplot, graphregion(color(white)) ytitle("Predicted Religious Regulation") ///
        xtitle("Religious Majority Size") legend(order(1 "Autocracies" 2 "Democracies"))
(Quick note: if you copy this text into your do file, you may have to replace the quotes with ones that Stata recognizes – for some reason these quotation marks don't work).

The limitation is that it's simply additive and the slopes differ in a constant way given the average difference between autocracies and democracies. What we would rather have is a way to assess whether the slope for the religious majority size differs given whether the state is a democracy vs autocracy (which is what many of you found in your A4). Remember that the relationship appeared flat (null) for autocracies and was positive for democracies. Let's simplify this a bit. First, let's make a dummy variable for whether the state is a democracy.

→recode p4_3 (2=1)(0 1=0), gen(dem1)

Now, let's re-run the last analysis, but allowing the regression model to calculate different slopes for each group in **dem1** – that is, using an interaction term:
→reg gri_08 c.lg1pct08##i.dem1

Let me unpack that. This now uses Stata's "factor" notation. It's a simple way of doing multiplication. I'll explain this more in class. The "c." and the "i." stand for the type of data in the variable. "c" stands for continuous and "i" stands for interval but we will generally just use it for dummy variables. The "##" tells Stata to do the "full factorial" treatment, which gives estimates for three variables. That is, when you run that code, it will produce an estimate for **lg1pct08** and **dem1** and then the product (multiplication) of the two variables (so, any values of the two variables would get run through both). It would produce the exact same estimates if you had done this:
             gen lg1pct08xdem1=lg1pct08*dem1 // gens the product of the two variables
           reg gri_08 lg1pct08 dem1 lg1pct08xdem1 // the two individually and then their product

The meaning of the **dem1** variable is perhaps easiest to understand here: the coefficient is the effect of a democracy when **lg1pct08** is zero. You can see how that is limited information (we want the effect of all combinations!), so we need to think of this now as a complete prediction system and visualize it as such. It should show the difference between democracies and other governments across the full range of religious majority sizes. So let's re-run the margins and marginsplot to see how these estimates differ:
→margins, at(lg1pct08=(0(10)100) dem1=(0 1))
→marginsplot[3]

This shows "the moderating effect of polity on religious majority effects on religious regulation." Note that the difference between the lines grows as the religious majority size grows! The difference is not constant as before but changes. The two groups have different slopes! Autocracies are, in general, not different from democracies when the society is religiously pluralistic (note the indistinguishable confidence intervals), but they differ tremendously in their religious regulation rates when characterized as religious monopolies. This opens up an entirely new world of nuanced empirical story telling.

**Deliverables**
By the end of this, you should have at least 4 graphs (one descriptive graph of the DV[4] and at least two showing the effect of an independent variable from the regression estimates), one table or coefplot (showing the regression results), and 6-8ish pages of write up (with a research question, a bit of theorizing with specific hypotheses, design section, results, and conclusion. Your regression model should have no more than 10 independent variables. Try one interaction term. Make sure to write a conclusion where you bring the discussion back to the theory and then expand beyond that to write about bigger questions like the ability of democracy to solve social problems). In addition to the tables and figures, you should include an appendix that tells us how your variables were coded. You might find this table handy – you can

---

[3] Here's perhaps a prettier way to show the CIs; at least it shows a few different commands:
→marginsplot, recast(line) recastci(rline) ciopt(lp(dash) lc(gs10)) graphregion(color(white))
[4] Yes, you can pilfer one from A4.

just fill in the details. It also serves as a useful check for you to make sure your variables are ready for the regression model.

---

Table A.1: Descriptive Statistics for Model Variables

| Variable Name | Mean | Std Dev | Min (#) | Min Label | Max (#) | Max Label |
| --- | --- | --- | --- | --- | --- | --- |

There is an incredibly easy way to make reasonably polished regression tables using Stata using a command called outreg2. So, let's say I have a regression model below that I would like to display the results of in a table.
reg gri_08 dem1 ano1 lg1pct08 teror108

First make sure that the variables are labeled for presentation purposes:
la var dem1 "Democracy"
la var ano1 "Anocracy"
la var lg1pct08 "Religious majority size"
la var teror108 "Experienced terrorism"

outreg2 using table1.doc, dec(2) label replace

So, you have to specify a document name – it will save the file in the default directory, which is probably where the data are, but to be safe, I would use the "cd" command first to specify where it should be saved. Then the "dec(2)" option says to report out 2 decimal places for each stat. I told it to use variable "label"'s and then "replace" writes over the document IF it is not open in Word (so close it when you're done inspecting). If you take out replace, then Stata will continue to add models in extra columns to the right. That's pretty cool if you want to compare "specifications" of your model (with certain extra variables perhaps).

Here's another way to display regression results – with a graph that shows the coefficients. It's called…wait for it…a coefplot! After running the reg model, you can run:
→ coefplot, drop(_cons) xline(0)

I almost always drop the constant, add an xline at 0 (which shows what we're comparing our regression coefficients to), and a few other things. You'll notice that you want to clean up the variable labels first (la var…). Also, I tend to standardize my variables to a 0-1 scale (the full range goes from 0-1) so that the coefficients are showing the same thing – a one unit shift in the x variable shows the full effect of the IV on the DV. It makes sense to do this given, say, the difficulty in comparing a "democracy" effect (0-1) compared to a religious majority size effect (0-100). The coefficients are telling you radically different things unless you standardize them. For instance, if you wanted to do this to the **lg1pct08** variable, it would look like this:
→gen lg1pct08_1=(lg1pct08)/100      // then use lg1pct08_1 in your models.

Otherwise, let's say you had a variable (var2) that ran from 1-5 (like strongly agree to strongly disagree). This is the way to standardize it to a 0-1 scale. You first make sure the low value is 0

(which is why you subtract 1), and then divide by the top number (which after you subtract 1 is now 4). I generally use the "_1" suffix to indicate that I collapsed the scale to 1:

→gen var2_1=(var2-1)/4

**Assignment 6**
**Who is Willing to Deliberate?**

In the fall of last year, my senior seminar was interested in the question of deliberation – debate over controversial issues governed by norms of reciprocity. The course's research design was a complex experiment that first engaged a survey of the student body in order to gather pre-test information as well as gauge who is willing to deliberate over public affairs. Within the survey, we took the 2ⁿᵈ answer in political science to heart and randomized treatments related to the key dependent variable for our A6: who is willing to deliberate. The willingness to debate could be explained by a wide range of other variables, though, so we will be sure to look for observational effects as well as treatment effects. Our goal, then, is to ask what explains Denison students' willingness to deliberate?

**The Dependent Variable**
Let's make the DV. Since it was part of an experiment, there are 8 versions of it (as described below). We just need to combine all versions of the question, which all have the same response categories, of course. By the way, "<." just means that the variable is non-missing (i.e., it has values). For short, I'm calling this "del."[5]

gen del=q23 if q23<.
replace del=q24 if q24<.
replace del=q25 if q25<.
replace del=q26 if q26<.
replace del=q28 if q28<.
replace del=q29 if q29<.
replace del=q30 if q30<.
replace del=q31 if q31<.

Woo! We have a DV! What does it look like? Are Denison students willing to come together and debate? Is there any variance to explain? Choose a suitable display method to show your audience what the analysis is about.

**The Experiment**
We randomized the conditions under which respondents were invited to a prospective forum. There were 8 versions of the question in a 2x2x2 between-groups design. The versions begin with Q23:

> As part of a joint effort between DCGA, The Bullsheet, and The Denisonian, we are going to sponsor a forum devoted to one of the issues listed on the previous page, in which participants will deliberate with a small group to decide what Denison's policy on this should be. The results will be covered by campus media and DCGA members will be informed about the outcomes you decide as a group. Invitations will be issued in the coming weeks. Given this, how likely do you think you would accept an invitation and attend?

Another version mentioned President Weinberg; one indicated that "You will be asked to give your opinion and justify it to others who may not agree with you;" and one had both present. Then another set of these 4 question versions appeared directly after a short battery of questions (Q27) that primed risk attitudes.

---

[5] As in "the funky homosapien."

Now, we need to get a handle on this experiment, so we need to combine these 8 treatment cells into variables. I'm going to make some dummies and one indicator variable (the latter first):

gen treat=1 if q23<.
replace treat=2 if q24<.
replace treat=3 if q25<.
replace treat=4 if q26<.
replace treat=5 if q28<.
replace treat=6 if q29<.
replace treat=7 if q30<.
replace treat=8 if q31<.

Now let's make some dummies for each of the different treatments, using "t_" to indicate to us that they are treatments. As always, use whatever nomenclature you want. Run one set of these codes – they'll do the same thing, which is to make "t_adam":

recode treat (1 3 5 7=0)(2 4 6 8=1), gen(t_adam)

or (btw, "|" stands for "or" below)

gen t_adam=0 if q23<. | q25<. | q28<. | q30<.
replace t_adam=1 if q24<. | q26<. | q29<. | q31<.

Can you code the "You will be asked to give your opinion and justify it to others" treatment yourself? I'd call it "t_justify".

Here's how to make the treatment for getting the questions after the "risk" battery.

recode treat (1 2 3 4=0)(5 6 7 8=1), gen(t_risk)

Now we have 3 treatments coded: t_adam, t_justify, and t_risk. We'd like to see if any of them made a difference, but first we need to check to see if the experimental technology was effective. Remember, treatments need to be randomized. We told Qualtrics to do that and now we need to assess if it worked. Effectively, this means that their compositions (types of people in each cell) need to be statistically equal (non-distinguishable). Let's check it out in a few ways. First, what sort of variables would be useful? How about gender? Why or why not?[6]

tab treat, sum(q61)

That's great, but is a 20 point difference significant? There's a test that helps us assess whether all of the means in a grouping (e.g., the treatment cells) are statistically indistinguishable: ANOVA (analysis of variance). If at least 2 groups are different from each other, then the ANOVA test will turn up significant ($p < .05$). If not (which is what we want), then there will be an insignificant ANOVA test ($p > .05$).

---

[6] You need to decide for yourself what to do with our gender variable (q61), which is coded 1=male, 2=female, and 3=other. There are 7 people who identified as "other," which is not really enough to do any statistical analysis with. I would (sadly) remove them from the dataset and just go with a m/f variable: recode q61(1=0)(2=1)(3=.a), gen(female)

anova q61 treat

It turns out that p=.12, which we say is not significant, but that's getting awfully close to p=.05. While we can technically say that the treatments were randomized along gender lines, there is clearly variation by cell so I would make sure to include gender as a control in the model. This also begs the question whether the treatments were randomized with respect to other variables – check 2 more and report the results of all of this in a footnote in the paper or in the design section.

Let's check out some treatment effects! There are a lot of ways to do this. If the treatments are randomized, then we can just use t-tests (assuming dichotomous treatments). If the cells are unbalanced in some way (even insignificantly), then we should include some controls and a regression context is more appropriate. Let's see how the estimates change with and without controls.

reg del t_adam t_justify t_risk
reg del t_adam t_justify t_risk female <plus other controls>

Do the results change? How stable are the treatment effects?[7] How do you want to show the treatment effects? I think a coefplot or a marginsplot would be nice (depending on how you display it, they're the same thing), but a table would work too (remember outreg2 will put them side by side). Showing the results side by side (with and without controls) would shore up confidence in the findings. Here's something I think is killer – construct a coefplot that puts the estimates from the two models side by side:
reg del t_adam t_justify t_risk
estimates store del1
reg del t_adam t_justify t_risk female <plus other controls>
estimates store del2
coefplot del1 del2, drop(_cons) xline(0)

**Observational Effects**
The survey also included a wide range of dispositional elements that may bear on whether people want to deliberate. Is it risky to engage in debate with others? Do Republicans want to do it more than others? How about those with strong opinions or many arguments? Is the willingness to debate a function of smarts? Civic engagement and civic skills? Majority status? Personality attributes? School pride? Democratic norms? The kind of issue they want to debate? There is so much cool stuff to check out in this dataset – look through the codebook and see what looks interesting. I would like to see two indexes in your final model.[8] An index is a combination of other variables that taken together capture a concept better than one variable alone. Have some fun with this.

---

[7] This evidence gives us a sense of whether cell composition imbalance has to do with the differences in means in the DV. The other thing that might be going on is that the treatments either interact with the covariates (hence have "conditional effects" rather than "sample-level effects") or the treatments interact with each other. I'm not mandating that you try these things, which is why this discussion is in a footnote, but if you do there are some interesting conditionalities.

[8] If you want to make one with risk attitudes, it's modestly tricky because half the sample got the questions before "del" and half got it afterward. We have to combine those measures. I made a little loop to put them together:
foreach x in 1 2 3 {
gen risk_`x'=q27_`x' if q27_`x'<.
replace risk_`x'=q36_`x' if q36_`x'<.

For instance, let's say you wanted to create an index of what my species calls civic skills (the q41s). Here are two ways to make it, but first let's recode them so that 0=no, 1=yes:
recode q41_1 q41_2 q41_3 q41_4 q41_5 (2=0)
egen q41_tot1=anycount(q41_1 q41_2 q41_3 q41_4 q41_5), values(1)

gen q41_tot2=q41_1+q41_2+q41_3+q41_4+q41_5

Tab them individually. Note the difference? We would need to adjust the egen version (which just adds up all the 1's and counts everything else as zero. That may not be appropriate since some people dropped out of the survey before that point and didn't answer the question). I would use this adjustment since Stata drops the case if there's a missing value in the addition formula:
replace q41_tot1=.a if q41_1==. | q41_2==. | q41_3==. | q41_4==. | q41_5==.
Now, when you tab them, they have the same values. You have to be sensitive to the missing values.

Your final model will have the three treatments, some demographic controls, and then about 6 other variables that would best explain willingness to deliberate. The goal is not just the best explanation, but the best theorizing about deliberation. I'm looking for a good, coherent package.

**Deliverables**
The final paper will be in the range of 10-12 pages of text (that, as always, is a rough estimate). It should follow the traditional layout of a research paper that we've developed in class (I will send you some description of the design you can use). The techniques should include the following used as many times as appropriate (and at least once): one variable figure, a two variable figure, regression model, and regression effects figure. Ideally it would include an interaction term and effects plot, though that is not required. You will include a variable coding appendix as in MA5. The completed assignment, with page numbers, flawless writing, and all the good stuff, competes for the VO Key Award.

**It's due Wednesday, 4.27 at midnight.**

---

}
This will make risk1 risk2 and risk3 (risk1 combines q27_1 and q36_1, for instance). Now you can combine those as you see fit. Do they all three belong together? Are they running in parallel or is one in the opposite direction? (Yeah, for the first one, a high score is less risk acceptance, so it would need to be reversed before combining them). How to combine them? Add them up. While you're doing that, do it in such a way that the scale runs from 0-1.