# Globalization: A Data Analysis Approach

Author: Scott Siegel

Department of International Relations

Humanities 331

California State University, San Francisco

San Francisco, CA 94132

Email: snsiegel@sfsu.edu

## Instructional Materials

The following four exercises integrate key concepts related to the study of international political economy by applying quantitative methods. These exercises help students understand the dynamics of international trade, contrasting patterns of international development, and how the winners and losers from globalization behave at the ballot box. Students will also learn how to present quantitative information graphically.

Exercises are aimed for those at the beginner level. We do not assume any existing knowledge of statistics, econometrics, or other research methods. In addition, each exercise involves the use of the *STATA* software package, for which students are not required to have any previous familiarity.

Each assignment consists of three parts. First, there is an assigned reading to accompany each exercise. It provides the theoretical background and the intellectual motivation for the exercise. In addition, there is a short book excerpt explaining the quantitative method used in the exercise and underlying explanation about the statistics used. In-class lectures will supplement this information. The exercise’s second part involves analyzing quantitative data. This part of the exercise will be done in the classroom with some additional time spent working outside of class hours. The third part consists of a brief “write-up” of the results, which briefly summarizes your findings and that brings together the conceptual material with the empirical evidence.

### Exercise 1. Understanding and Visualizing Patterns of International Trade

### (100 points)

**Note to Instructor: This first lesson is aimed at helping new users become familiar working with statistical software packages, especially their graphing tools. Students learn how to navigate data sources online, download and transform for analysis. This exercise introduces students to basic concepts such as what a variable is, the difference between a dependent and independent variable, the nature of panel and time-series data, and how to present information visually. Students can approach the exercise in different ways, but a template to follow is included in the attached *STATA* .do file.**

We will use the [World Bank’s Data Bank](https://databank.worldbank.org/data/home.aspx) to analyze one element of globalization—international trade. We look at how levels of international trade change over time, which countries trade most with the rest of the world, and their main trading partners. For this assignment you will actually construct panel data that will be used in Exercise 2 (trade and socio-economic development). The point of this exercise to help you become comfortable with locating data, how and why to choose different indicators for a concept, and how to present information in both written and visual format.

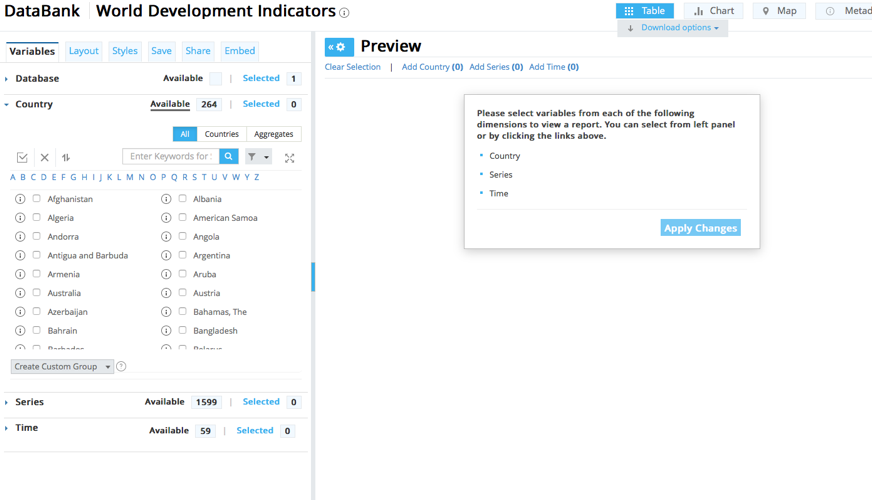
#### Reading Component:

* + Robert O’Brien and Marc Williams, *Global Political Economy*, 5th Edition, 2016, Palgrave Macmillian, Introduction, Chapters 1 & 2 [Textbook]
  + Dani Rodrik, *The Globalization Paradox*, pp. 3-23 (on iLearn).
  + Esteban Ortiz-Ospina, Diana Beltekian and Max Roser, “[Trade and Globalization](https://ourworldindata.org/trade-and-globalization#trade-around-the-world-today),” *Blogpost*, October 2018; Esteban Ortiz-Ospina, “[Does Trade Cause Growth?”](https://ourworldindata.org/trade-and-econ-growth) *Blogpost*, October 22, 2018.
  + Peter Galderisi, *Understanding Political Science Statistics*, 2015, Taylor and Francis, Chapter 2.

#### Data Analysis Component:

**Step 1**

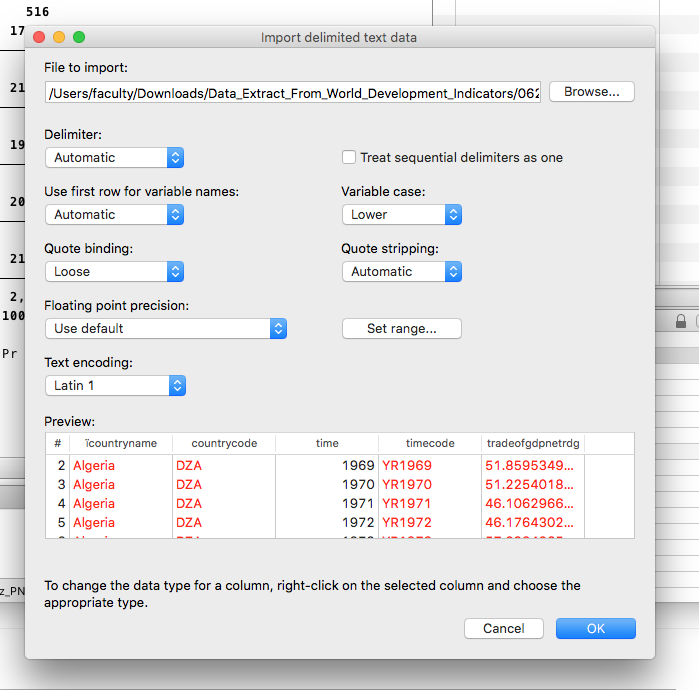
* 1. Visit the World Bank’s Data Website and click on “Data Bank” and then “World Development Indicators.” You will see a screen that looks like this:



* 1. Our first step is to create some panel data that we can analyze. We will be choosing different variables for multiple countries and regions over a specified time period. From the left-side panel:
     1. Under “Country,” click on the “All” tab if it is not already selected and select 5 countries from 7 different regions of the world (North America, South America; Europe; East Asia; Southeast Asia or Oceania; the Middle East and North Africa or South Asia) and the “World.”
     2. Under “Year,” select the last 50 years.
     3. Under “Series,” there are many different indicators that measure a country’s trade levels. We are interested in how important trade (exports + imports/GDP) is to a country’s economy, how trade levels have changed over time, and a country’s main trading partners.
     4. If necessary change the layout so that your independent variable (Trade % GDP) is in a column and country-year appears as a row. This is called panel-time series data. See illustration below.



* + 1. Download the data as comma-separated values (CSV).



* 1. Open *Stata* and ***import*** the Trade file making sure that the first row are variable names. Now, let’s clean this data up for analysis. *STATA* can’t recognize data that is both text (“string”) and a number (“integer”). We need to adjust some cells so that *STATA* recognizes them as missing data. Alternatively, download as an Excel file and clean it up in that program before importing into *STATA*.
  2. Use the ***destring*** command to convert text variables into numerical format. Use the Edit window to delete any information in a cell that is not in numerical format. Make sure you delete each cell that contains non-numerical information or the *destring* command won’t work. We can also easily take *string* formatted data for each country name and assign it a specific numerical code for ease in sorting and labeling later.

encode countrycode, generate(country) label (l\_country)

You can see the labels created by using ***label list*** and the value label assigned. You can change the value labels using the label define command.Let’s also start renaming and labeling our variables if they don’t make sense or didn’t translate well during the importing process. Use ***label variable X “XX”*** to change it to a title you prefer.

* 1. Let’s get some basic information about this trade data, such as each country’s mean, standard deviation, minimum and maximum. First, we need to tell *STATA* we are using panel data, observations of multiple countries in multiple years: use the ***tsset*** command.

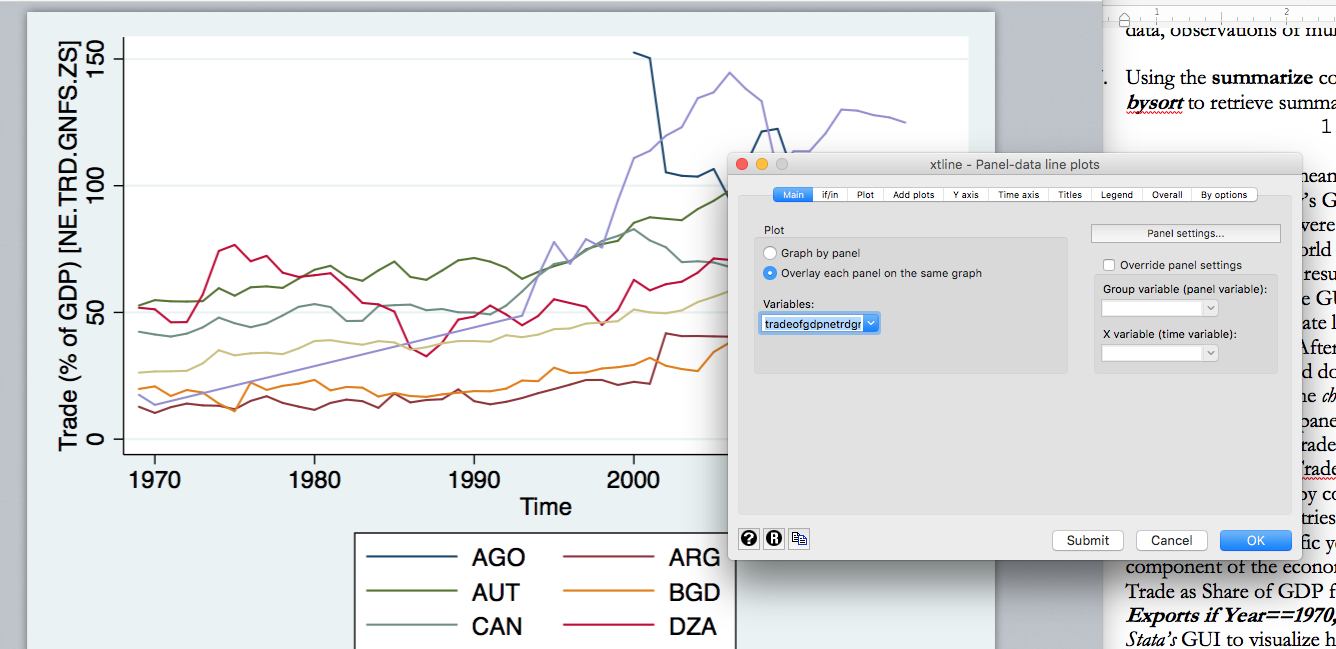
tsset country year

* 1. First, use the **summarize** command to get an overall look at the data. Then use the ***bysort*** prefixto retrieve summary statistics by each country and by year:

1. bysort country: summarize
2. bysort year: summarize

Note the differences in means between countries and years. Note where trade is a large share of a country’s GDP and where it is the least. You may also want to take note of which years trade levels were highest (as shares of country’s GDP) and how each country’s share compares to the world average.

* 1. Now, let’s graph our results by country over time. The fastest and easiest way to graph panel data is using the GUI option “Panel data-line” menu option. *STATA* will know to graph the time-series data for each country on the same graph. You may have to edit the scales for the countries you selected to see the trends more clearly.



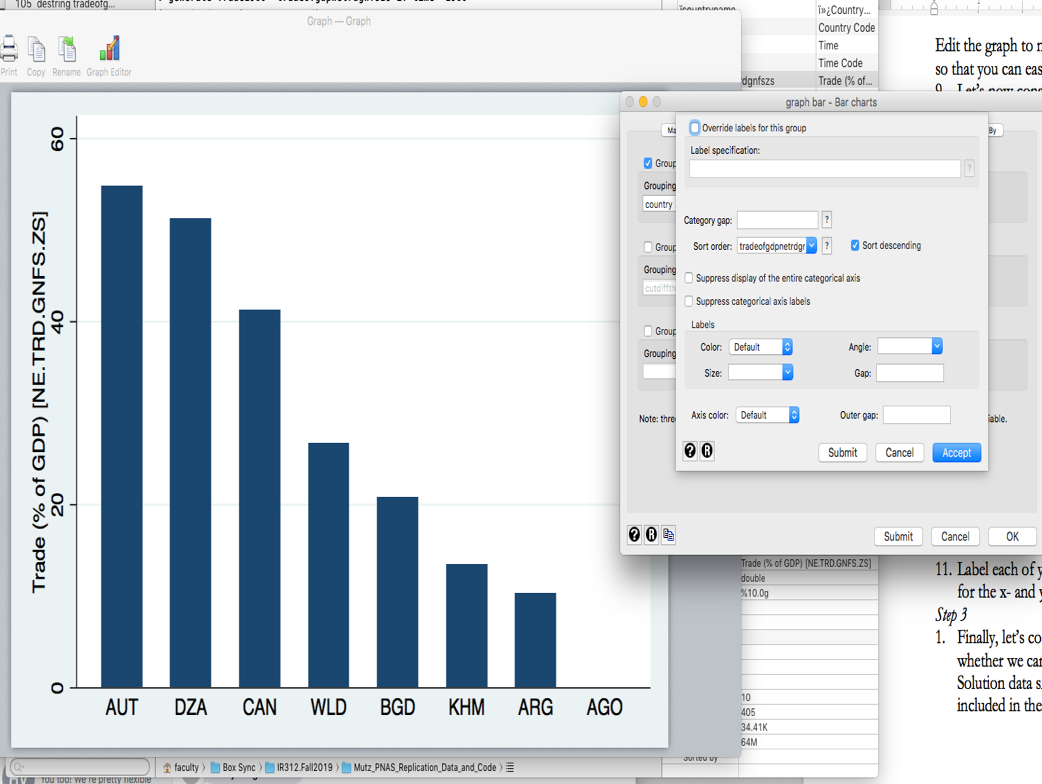
Edit the graph to make all the information legible. After labeling the graph, save as a *.jpeg* file so that you can easily insert into a MS Word document.

* 1. Let’s now consider the *change* in levels of trade over time. Generate a new variable for changes in levels of trade using the ***bysort*** command:

bysort country (time): generate ChangeTrade =(TradeGDP- [\_n-1])

Again, graph this variable by country over year. What trends do you observe?

* 1. Now, let’s rank countries according the importance of trade to their economies. We can best see this visually. We can either choose a specific year, the average over specific time span, or how trade changed as a component of the economy between the first and last year data is available. Do all three! Use the bar graph command in *Stata’s* GUI to visualize how trade as share of GDP has changed over time.



You can also calculate the *average* level, *absolute change*, and *average change* in trade as a country’s share of GDP using the ***collapse*** command and then create bar graphs. First, use the ***preserve*** command so that you can easily restore the data set to its format before transforming it.

collapse (mean) TradeGDP, by(country)

collapse (mean) TradeGDP, by(country) if year==2000

* 1. Label each of your graphs correctly with persuasive and concise titles, units of analysis for the x- and y-axes, legend and sources. You can combine them all into one file using the ***graph combine*** command, if you wish.

**Step 3**

* 1. Finally, let’s consider the major trading partners of the countries you selected and whether we can detect any patterns. Visit the [World Bank’s World Integrated Trade Solution data site:](https://wits.worldbank.org/countrystats.aspx) For each country included in the analysis above, identify their 5 main trading partners for the latest year available, total imports and exports, the trade balance, and share of all trade. Create table in a word processing program that summarize the data.

*Extra-Credit (20 points): Use the STATA software package to create a pie-graph for each country that shows each trading partner’s share of total trade. You can download the data from the website in .csv format, edit in MS Excel, and then import into STATA.*

*More Extra-Credit (10 points): Create a series of pie graphs showing change over time in which countries are the main trading partners for each country chosen for analysis.*

*This part of the assignment requires you to work more independently, but your instructor will provide some limited guidance and instruction. This will help you become more familiar with the graphics features associated with STATA*.

#### Writing Component:

Write up! It’s time to describe your findings. In a one-page essay (max. 1000 words). Describe the dynamics you observe. Some of the questions you want to answer are:

* For which countries is trade a major component of their economy?
* Which countries export more than others?
* For which countries has trade increased as a share of their economy over time? Which countries have seen stagnation or decline in trade as a share of their economies over time? When?
* Which countries trade with each other most?

In a final paragraph, construct some hypotheses (we will test these later).

* Why is trade a big component of some countries’ economies than others?
* Why do countries trade with some countries more than others? Which types of countries are most likely to trade with each other and why?
* Why have levels of trade increased, decreased or remained level for each country?
* Why has total international trade increased since 1960?

### Exercise 2. Globalization and Economic Development (100 points)

#### Reading Assignment Component

* O&W: Chapter 11
* Dani Rodrik, *The Globalization Paradox*, pp. 135-183 (iLearn).
* Nancy Birdsall and Francis Fukuyama, “The Post-Washington Consensus,” *Foreign Affairs*, March/April 2011 (online).
* Esteban Ortiz-Ospina, “[Is Globalization an Engine of Economic Development](https://ourworldindata.org/is-globalization-an-engine-of-economic-development),” *Blogpost*, October 22, 2018.
* Peter Galderisi, *Understanding Political Science Statistics*, 2015, Taylor and Francis, Chapters 4 & 11 (on iLearn).

#### Data Analysis Component

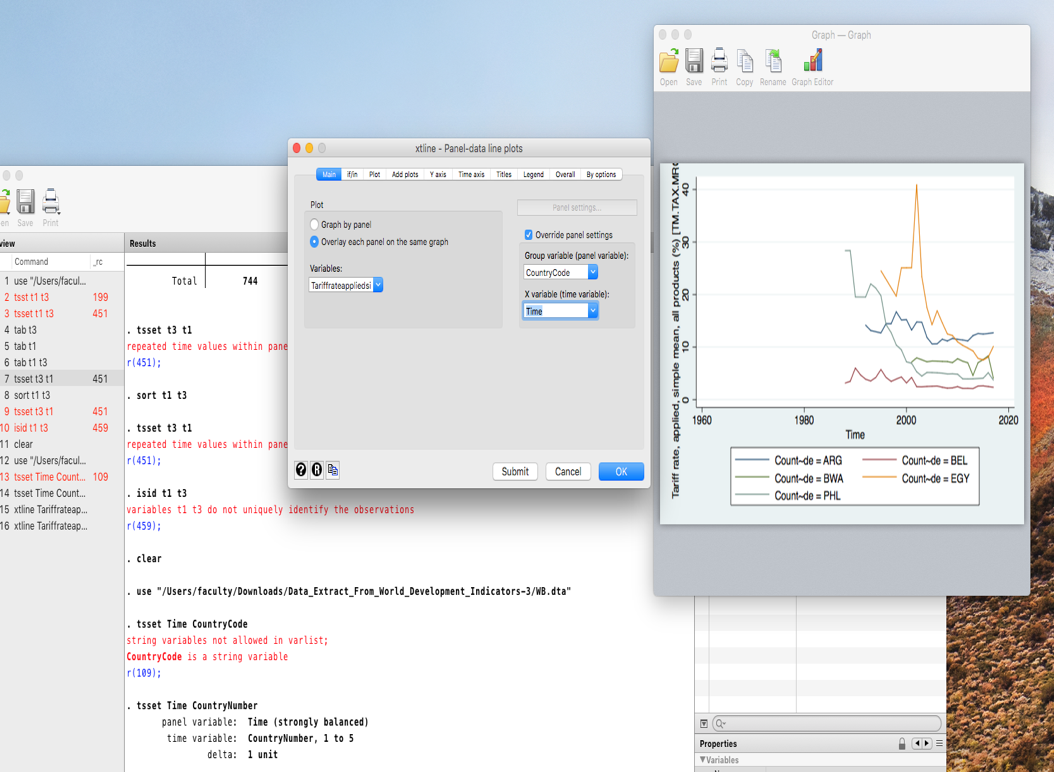
For this exercise, we will draw on the dataset you created in Exercise 1. Our goal is to identify the relationship between globalization—here, international trade—and a country’s socio-economic development. Does globalization make a country better off, according to liberal and neoliberal theories of international trade? Or does trade hinder development? Or have no impact at all? Methodologically, we will learn about dispersion, variation, goodness of fit, and the basics of linear regression.

* 1. Open the *.dta* file containing the data on the countries you selected for Exercise 1. If you have not already done so, make sure included the variables “Tariff rate, applied all products” and 2 measures of wealth, “Poverty headcount ratio at “**$1.90 a day (2011 PPP) (% of population)”** and “**GDP per capita (constant 2010 US**$**).”** For the written part of the assignment, you may want to consider alternative or additional indicators of socio-economic development, such as mortality rates, literacy rates, or rates of personal income growth.
  2. As in Exercise 1, let’s become familiar with our variables. Then use the ***bysort*** to retrieve summary statistics by each country and by year:

1. bysort country: summarize
2. bysort year: summarize

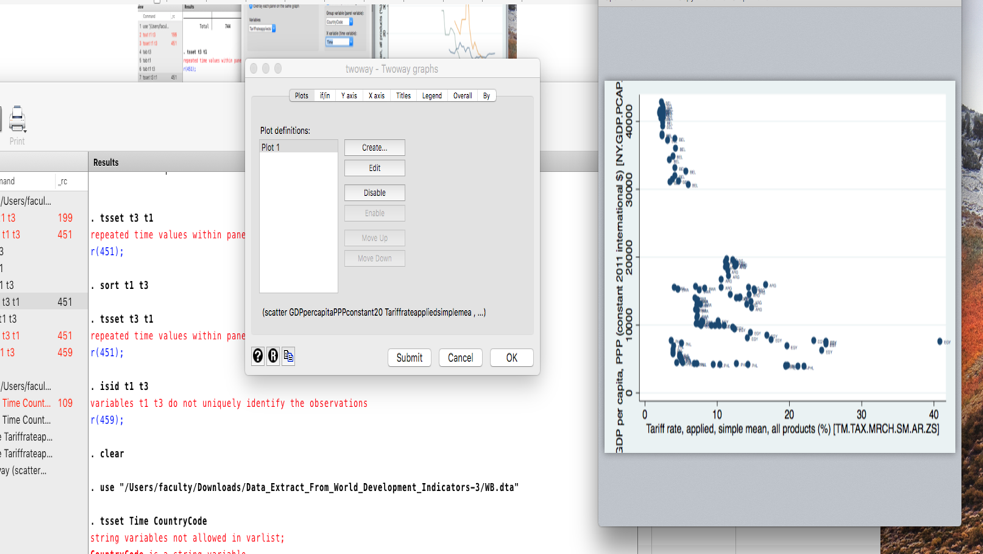
What were the mean tariff levels in the time period selected? Which country had the highest tariff levels and when?

* 1. Let us take a look at the data visually.
     1. Let’s consider the general trends in tariffs and poverty over time in each country, respectively. First, choose exports as your *dependent* variable and time (year) as your independent variable. Use *STATA*’s GUI interface to create a series of lines for each country [Don’t forget to tell *STATA* your data is panel data using the ***tsset*** command.] Click on “Graphics,” select “Panel data line-plots,” and then select “line” under the “Plot” tab. Also, make sure that each country’s data appears on the same plot (check “*Overlay each panel on the same graph*”). You will produce a figure that looks similar to this:



Create similar line plots for other variables, such as for poverty or other measures of socio-economic development you chose for your analysis.

* + 1. Now let’s create a scatterplot that shows the relationship between free trade and economic development. We will ignore, for the moment, the panel nature of the data. Instead, we will assume that each observation is independent or unrelated to another. Again, use the *STATA’s* GUI to create a scatter plot. Visually, do you see a pattern? Is there a general trend you see in the data? Create similar graphs using poverty or other indicators of socio-economic development of your choosing and note any patterns you see in the data. You will produce a figure that resembles the example below:



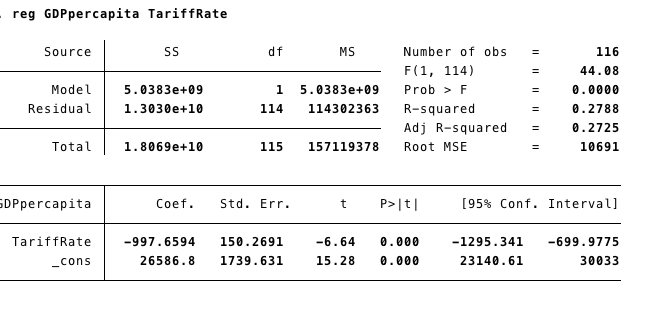
You can also you the “Panel-data plot lines” tool under the “Graphics” menu to create scatter plots for each country, which will be useful for the next part of the assignment.

* 1. What is the mathematical equation that minimizes the sum of the squared residuals? Using simple linear regression, we can identify the relationship between two variables. In the *STATA* command window, we can run a regression using the **regress** or **reg** command (but ignores the “fixed effects” related to each country and the serial correlation between each year):

reg GDPpercapita TariffRate

This produces the statistical output similar to what is shown below:

R-squared is the “goodness of fit” of the linear model (ranges from 0 to 1). Tariff rates explain 27% of the variation between in GDP per capita across multiple countries and years. This leaves 73% of the variation unexplained.



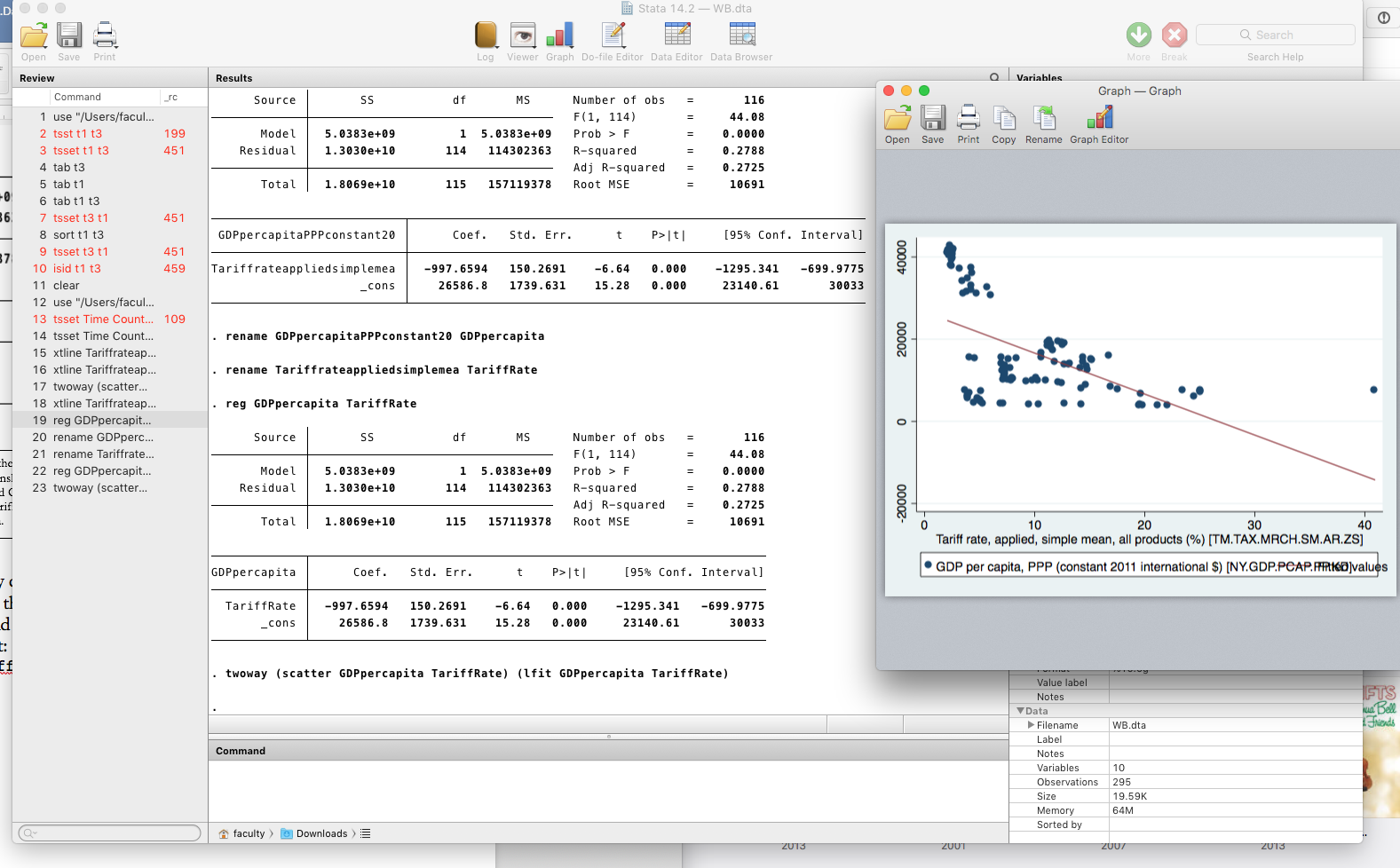
The “p” value shows that the relationship is statistically significant and the set of values the coefficient could be does not include 0.

The coefficient shows that there is a negative (decreasing) relationship between tariff rate levels and GDP per capita. That is, the higher tariffs are, the lower is GDP per capita.

The constant represents what is the (mean) predicted value if the tariff rate equals 0 among the sample of countries and years in the model.

We can also see the relationship graphically by drawing the “best-fitted” line that best predicts the relationship between the two variables. In the *STATA* command window, create a scatter plot of the relationship between tariff rates and GDP per capita, but also include a command that asks STATA to produce the linear fit plot:

twoway (scatter GDPpercapita TariffRate) (lfit GDPpercapita TariffRate)



You can also ask STATA to create linear fit plots for each country by including the country code # you assigned and the “**if**” statement. For example:

twoway (scatter GDPpercapita TariffRate if code==1) (lfit GDPpercapita TariffRate if code ==1)

twoway (scatter GDPpercapita TariffRate if code==2) (lfit GDPpercapita TariffRate if code ==2)

twoway (scatter GDPpercapita TariffRate if code==3) (lfit GDPpercapita TariffRate if code ==3)

From the graph above, and perhaps in your own data, you may observe a few outlier observations or even influential observations (see class lecture for explanation). You can address these using several techniques:

* + - 1. Sort the data from lowest to highest values and then use the *Data Browser* window to locate the outlier observations.
      2. To identify observations with significant leverage, create predicted values with:

predict h if e(sample), hat

Then graph a box plot to locate those observations that are much lower or higher than then the general pattern or average covariance pattern. Values larger than .5 should be avoided, while values between 0.2 and 0.5 are risky.

* + - 1. Using Cook’s distance is also a way of measuring the degree of influence an observation has on our model:

predict cooksd if e(sample), cooksd

Then re-run the regression by excluding any observations > 1 using the **if** command.

*Note: Consider the figure above. Do you think the relationship between the two variables is best described as linear? Or would a different type of equation better describe the relationship. Consider this issue when estimating your own equations and crating your own graphs.*

Repeat these steps for any other independent variables you chose to indicate socio-economic development.

#### Writing Assignment Component (between 500 and 1000 words)

Compose a short report describing the relationship you discovered between tariffs, trade, and socio-economic development. In your write-up, include the statistical results and copies of your figures. Some of the questions you want to consider answering:

* + 1. Is there a positive or negative relationship between free trade and economic development? Is the relationship strong or weak? How well does your regression equation fit the data? Does the relationship vary by country?
    2. Which countries best resemble the pattern you discovered? Which countries don’t appear to following the pattern very well? Are there any outliers, either regarding specific countries or years? If so, why do you think that is?
    3. Are there other factors that affect socio-economic development that you think should be included in the model such that you could better estimate the model?

### Exercise 3. The Party Politics of International Trade in the United States: Who are the Winners and Losers? (100 points)

In this exercise we will explore whether the losers from globalization account for why Donald Trump won the US Presidential election in 2016. President Trump’s victory in the 2016 election was a surprise to many. Despite Hilary Clinton winning the popular vote, Trump secured an electoral college majority. His electoral college victory came by winning several states in the US in the Midwest that Obama had won in 2008 and 2012. Why the switch?

Some political scientists argue that Trump’s appeal to white, working class voters in these states led to his victory over Hilary Clinton in key swing states in the US electoral college. White, working class voters are assumed to have borne the brunt of costs of adjusting to competitive wage pressures caused by globalization. Unemployment in manufacturing has gone up, incomes are stagnating, and the community in which there were once factors are dying, some say, because of foreign competition

In the 2016 election, there were clear policy differences between the two candidates over free trade. Hilary Clinton argued the US should maintain open trade with its partners, while Donald Trump publicly committed to renegotiating trade deals and protectionist policies. Thus, we should expect the losers from globalization, especially from free trade, to choose Trump over Clinton at the polls. This argument is highly disputed in the scholarly literature, however. Instead, many scholars argue that President Trump’s use of racist rhetoric and anti-immigrant policies motivated many whites—both college and non-college educated—to vote for him and economic, or “pocketbook” issues, played little role.

To test these competing hypotheses, we will analyze a unique set of survey data that polled voters along both of these issue dimensions to assess which school of thought is better at predicting why people voted for Trump or Clinton. The assigned reading discusses the domestic politics of international trade, rightwing populism, and how the two are possibly related. Then, we draw on an existing data set to identify any patterns in how the so-called losers—the white, working class—voted in the 2016 election. We will compare their voting behavior to other social groups and assess whether there are any statistically significant differences. You are asked to summarize your results in a short essay.

#### Assigned Reading Component

* [Mutz, Diana C. “Status Threat, Not Economic Hardship, Explains the 2016 Presidential Vote,” Proceedings of the National Academy of Sciences, 115, no. 19, (May 8, 2018): E4330-39 and supplemental material.](https://www.pnas.org/content/115/19/E4330)
* [Dani Rodrik, “Populism and the Economics of Populism,” Journal of International Business Policy, June 2018, 1, 1-2, pp. 12-33.](https://link.springer.com/article/10.1057%2Fs42214-018-0001-4)
* Justin Gest, *The New Minority: White Working Class Politics in an Age of Immigration and Inequality*, Chapter 4, pp. 74-115 (on iLearn).
* Peter Galderisi, *Understanding Political Science Statistics*, 2015, Taylor and Francis, Chapter 7 (on iLearn).

#### Data Analysis Component

Mutz’s (2018) dataset include responses to two surveys. First, Mutz (2018) collected panel data, that is, repeated observations of the same people over time and how they voted. The author relies on a second, cross-sectional survey of voters by AmericaSpeak/NORC. Both surveys asked questions about voters’ candidate preferences, demographic characteristics, and policy opinions about globalization. We will work with the panel data first, provided by your instructor. Your task is to identify how different groups of voters behaved at the ballot box and whether there are significant differences between them regarding their attitudes to the two candidates and whether their views on globalization affected how they voted. A full description of all variables and how different indicators are measured is included in the assigned reading.

* 1. Open the file “Panels.dta” in *STATA*. You will see all of the variables included in Mutz’s (2019) article the support one part of the author’s analysis. For ease of analysis and interpretation, you may want to start assigning labels to the variables.
  2. Let’s start by summarizing some of the data. Use the **summarize** command to get an idea of how each variable is measured and their range of values.
  3. Using the **tabulate** command and one variable, we gain a sense of the *frequency* of observations at different values and their share of the sample. **tabulate** makes most sense for using with ordinal or categorical variables. For example, to assess what shares of the sample population voted for Trump over Clinton, enter the command:

tabulate demrepvote

* 1. Before beginning our analysis, let’s make one of our dependent variables easier to interpret. This variable ranges from 0 (very negative) to 100 (very positive) feeling a voter has about each candidate. So, let’s combine the responses into five categories—from those who strongly favored Clinton to those who strongly favored Trump. Let’s begin by exploring one of the dependent variables—the “feeling thermometer.” Use the **replace** command to create a new categorical variable of 5 different sub-groups:

egen cutdifftherm= cut(thermdiffTC), group(10)

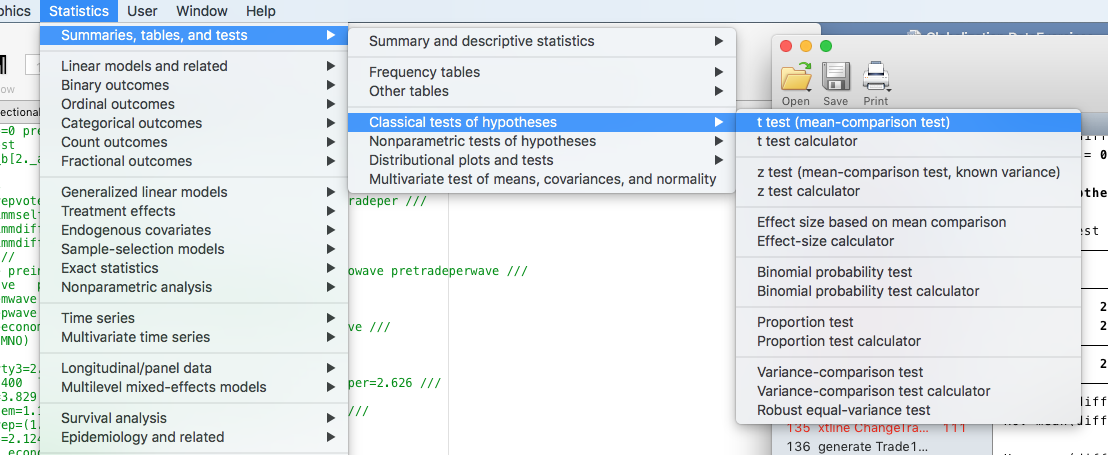
* 1. Now we can start exploring how different groups voted and whether the differences between the mean values of two different variables are statistically significantly different from each other using *STATA’s* **ttest** command (*Background: see Galderisi reading and lectures*).
     1. Was the mean feeling for Donald Trump significant different from 50% (i.e., neutral)?

ttest reptherm=50

* + 1. Was the mean feeling for Clinton significantly different from the meaning feeling for Trump?

ttest reptherm= demtherm

Use can also use *STATA*’s GUI to produce the same results:



* + 1. Was the mean difference in feelings towards either Trump or Clinton different according to the respondent’s party affiliation (Democrat, Republican or Independent)

pwmean demtherm, over(xparty3)

* 1. Next, let’s see how different groups voted and whether they are significantly different from each other. Using the **tabulate** command, we can produce contingency tables, or “crosstabs,” that display the cross-classification frequencies of two more variables. For each of your tables, include the option **chi** to test whether the relationships are statistically significant. Produce a table of frequencies and sample percentages for how different gender groups felt about Hilary Clinton versus Donald Trump.

tabulate cutdiffthermTC xparty3, column chi

The first row reports the *frequencies*, or the total number of respondents, that falls into each thermometer category sorted by party. Repeat this process for other independent variables. *Do you notice any patterns? In the sample, which groups had higher opinions of Clinton than Trump?*

* 1. Second, let’s consider how people possibly affected by globalization felt about each of the candidates. Recall, our theory is that those most harmed by free trade found candidate Trump’s message of tariffs and protectionism most appealing. These people include those who do not hold a college degree, believe that trade has been bad for the United States, and are less likely to perceive China as an economic opportunity. Once again, use the **tabulate** command to compare frequencies and percentages according to different values of each of these variables. For instance, China is rated as either a threat (1) neutral (2), or an opportunity (3). Create tables for each of these categories, again, include the results of chi-square tests.
  2. Finally, we can create some tables that look at how different policy attitudes vary by social group by using the **bysort** prefix. You can examine how attitudes towards China or free trade vary by social group. For example, you want to compare how attitudes towards free trade vary according to the person’s political party affiliation.

#### Writing Component

In a one-page essay (not more than 1000 words), summarize your results. As guidance, consider the following questions:

* Which groups were more likely to favor Trump over Hillary and vote for the Trump at the polling booth?
* Does it appear that policy attitudes are strong predictors who will vote for whom?
* What are the differences in probabilities between different groups and their attitudes towards Trump or likelihood of voting for him over Clinton in 2016?
* Specify some hypotheses or “guesses” for your observations. Why do some groups prefer to Trump to Clinton? Why are people with different social attitudes gravitating towards one candidate over another? What types of evidence would you look for test these hypotheses?

### Exercise 4. Does Free Trade or Racism Predict Who Votes for Donald Trump?

#### Reading Assignment:

* 1. [Morgan, Stephen L., and Jiwon Lee. 2019. “Economic Populism and Bandwagon Bigotry: Obama-to-trump Voters and the Cross Pressures of the 2016 Election.” SocArXiv. June 2. doi:10.31235/osf.io/nmxcr.](https://osf.io/preprints/socarxiv/nmxcr/)
  2. Emmit Riley III and Clarissa Peterson, “Economic Anxiety or Racial Predispositions? Explaining White Support for Donald Trump in the 2016 Presidential Election,” Journal of Race and Policy, forthcoming (iLearn).
  3. Mehmet Mehmetoglu and Tor Georg Jakobsen, *Applied Statistics Using STATA*, Chapter 8, pp. 161-191.

#### Data Analysis: Replication and Interpretation

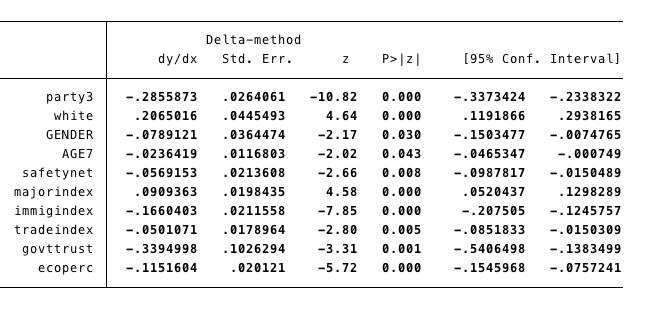
**Note to Instructor: In this exercise, students learn how to interpret statistical output and weigh the relative effectiveness of different models. The data, methodology, and commands are provided. This exercise encourages students to engage and think about the data critically as well as perform some simple analyses of how well the models explain variation in the data. There is no emphasis or focus on regression diagnostics.**

*Step 1. Replication [Students should come prepared to answer the following questions in class.]*

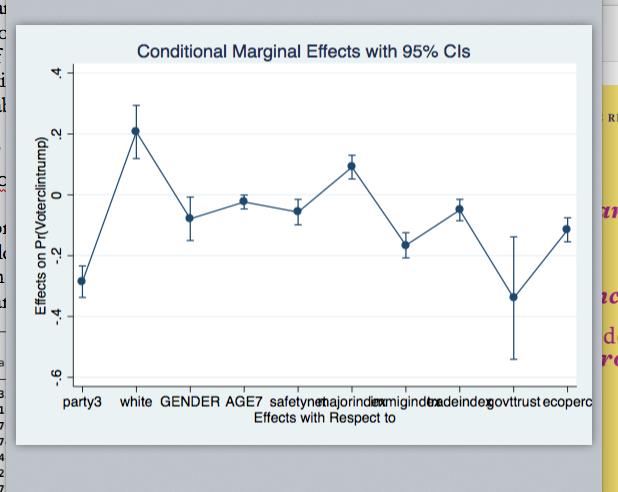
1. Open Mutz’s other data file, “AmericaSpeaks.Exercise4.dta” into *STATA*. Explore each variable in the data. Let us explore the data together in class, first.
   1. What is the dependent variable and how is it measured?
   2. What are the independent variables? What does each independent variable measure? How does each measure reflect the different schools of thought or hypotheses for voters’ choice in 2016? Do these measures accurately measure the concepts?
   3. Using the **tabulate** command from Exercise 3, *evaluate* which portions of the population felt their status most threatened. Which groups were most likely to express economic hardship? Which voter groups are more likely to identify as identifying with the Republican or Democratic Party?
2. Replication is a very important quality and principle in scientific research. When the same results can be produced, we can be more confident in the validity and reliability of the data analysis. Accompanying the data file is a *STATA* **.do** file. The **.do** file allows other researchers and social scientists see exactly how the statistical results were produced. A *STATA*’s ***.do*** file contains all of the statistical commands a researcher used to produce her statistical results. By running a **.do** file we repeat the same steps and see if we obtain the same results, while also using the same data to test alternative explanations.[[1]](#footnote-1)
3. In *STATA*, use the GUI to open the ***.do*** file: “Exercise4.do.” This will run most of the statistical models that Mutz (2018) used in her *PNAS* article. Running the file produces 10 different statistical models that model how voters chose Donald Trump over Hilary Clinton in the 2016 election.
   1. Across each of the models, which factors are statistically significant predictors of who will vote for Donald Trump? Which factors are not statistically significant? What is the directional relationship between the independent variables that are statistically significant and the choice for Donald Trump as president in 2016?
   2. The coefficients produced the results table produce the *natural logarithm* of the *odds* for Y=1 for each one-unit change in *X*. However, those numbers can be hard to interpret into meaningful numbers that tells us about the probability *Y* will happen. To assess the independent effect of *X* on *Y*, we need to transform the logit into a predicted probability. Thankfully, we do not have to do this by hand. *STATA* can do this for us with the **margins** command. Let’s look at the independent effect of “loss-of-social-status” variables in a model that also includes “pocketbook” factors on the probability that a person votes for Hilary Clinton
      1. Re-run the first statistical model only with using the commands in the **.do** file. Simply copy and paste the command from the output window (you’ll have to remove some of the added characters to make it one line of code).
      2. Type **margins** and include only the variables that are statistically significant followed by the sub-command, while holding all other variables at their mean value:

margins, dydx(party3 white GENDER AGE7 safetynet majorindex immigindex tradeindex govttrust ecoperc) atmeans.

The output window shows the *change in probability* for each one-unit change in one of the variables selected when all other variables are held at their mean (see figure below). Which factors have the largest effect on voting for Donald Trump (or reduce the probability of voting for Hilary Clinton)? You can also use the **listcoef, help** command to see the change in *odds* of a person’s choice to vote for Donald Trump or Hilary Clinton.

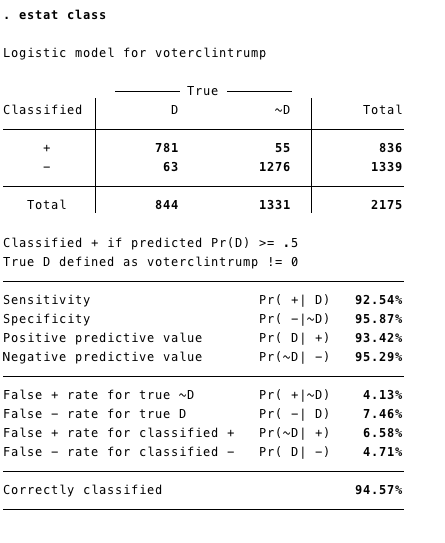


* + 1. You can also graph the results using the command **maginsplot**. This will produce a graph similar to this:



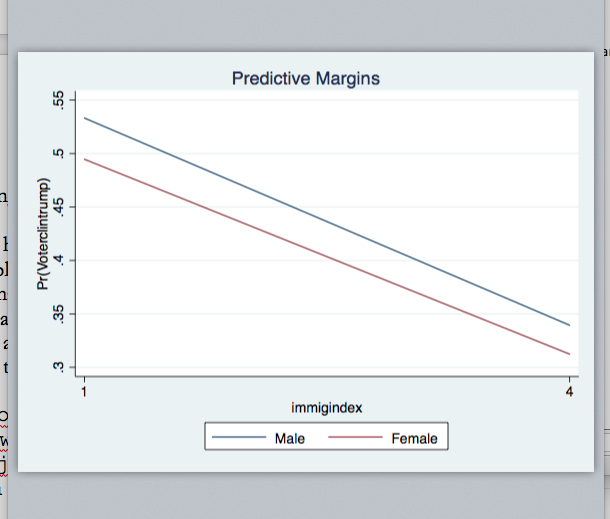
Using the graph you produced, in class, tell us which factors increase the probability of voting for Trump. Which factors have the largest effect? Which have the smallest? Which variables are we less confident *statistically* have an effect on a person’s voting choice?

* 1. All statistical models are simplifications of reality. However, we can measure how will the model fit our data by assessing its goodness of fit by using the command **estat class**. *STATA* will produce output that looks similar to this:



As you can see, our chosen model did a very good job of predicting correctly how an individual would vote. It explains almost ***95%*** of all cases.

* 1. Finally, we can create a graph showing the effect different factors have on the probability of voting for Donald Trump when independent variables are given specific values, which allows us to observe how the same factors have different levels of effect depending on the groups to which voters belong. For example, let us compare the effect of attitudes towards immigrants between men and women on the probability a voter will choose Trump in the 2016 election. Using the following commands:
     1. quietly logit voterclintrump party3 noncollegegrad white GENDER AGE7 religion INCOME lookwork ecoworry perecoperc safetynet medianincome majorindex pt4r sdoindex prejudice isoindex china immigindex tradeindex natsupindex govttrust ecoperc terrorthreat
     2. margins, at(immigindex=(1(3)5) GENDER=(1 2))
     3. marginsplot, noci (recast(line)



#### Independent Analysis and Writing Assignment

This assignment consists of two parts:

* 1. Conduct your own replication of Mutz’s data. Choose either the Preference for Trump at the voting booth or the actual vote choice as your dependent variables. Then choose a model of an individual’s electoral behavior that includes only “pocketbook issues,” one that only tests the role of “social status threat,” and one that includes both economic and social variables. Repeat the analysis we did in class for each of the models you choose.
  2. Given your results to the tests above, what do you think best explains why white working class Americans voted for Donald Trump in the 2016 election? Do pocketbook issues or threats to their social status better explain how people voted? Write up your conclusions in an essay of between 2000 and 2500 words. You may want to address the following issues in your essay:
     1. What are the best predictors or factors of whether one votes for Donald Trump or not? Which “pocketbook” factors have strongest effect on the probability of voting for Donald Trump? Which “status” factors best explain why someone voted for Donald Trump? When both types of factors included, which factors have the largest, statistically significant impact?
     2. Compare the effects of your main predictors across different social groups, such as gender, age, or levels of education. For example, do attitudes that immigrants are a threat to society have a larger effect on the probability of voting for Donald Trump over others?
     3. How well do each of the models explain a voter’s choice in 2016? Include tables and graphs like the ones we created in class to support your conclusions.

1. Because of issues of confidentiality, Mutz (2018) cannot share all of the data used in the published article. The publicly available data is missing some core variables related to economic vulnerability, such as employment in the manufacturing sector by county and ZIP code. However, we can still use different indicators to test the “pocketbook” hypothesis against the “cultural status” argument. [↑](#footnote-ref-1)