

Gov329: Empirical memo handout

What is this? You will write an empirical memo of no more than 5 pages in which you use data to examine an international security hypothesis.

Why do it? This assignment gives you the opportunity to work with data on an international security issue of your choice. After completing the assignment, you should have a better sense of where international security data comes from, the limitations political scientists face in drawing firm conclusions based on these data, and how to present data in a graphical way.

What is it worth? This assignment is worth 20% of your course grade.

When is it due? The empirical memo is due before class on October 4.

For the empirical memo, you will:

1. **Formulate at least one hypothesis about international security, and explain why it is interesting.** You can cite some of the existing international security literature if you want, but this is not necessary.

You should express an association between two variables in the world, but you do not need a full-fledged causal theory. Your hypothesis might posit changes in international security phenomena over time, or a difference in the international security behavior of different subgroups. For example, you might hypothesize that international conflict has decreased over time, or that international conflict is more prevalent among poorer states.

2. **Find a data source that will help you investigate your hypothesis.** Good options are listed in the course schedule for 9-10-21, and include:

The Correlates of War Project: www.correlatesofwar.org

International Crisis Behavior (ICB) Project: sites.duke.edu/icbdata/

Armed Conflict Data (UCDP): ucdp.uu.se/downloads

World Development Indicators (World Bank): databank.worldbank.org

You are not limited to these data sources. Think about what kind of data you are interested in, and then Google it—it probably exists somewhere. Contact me if you want suggestions for particular kinds of data.

3. **Analyze the data to see whether it supports your hypothesis.** This can be as simple as examining the trend over time in a spreadsheet. You might also try looking at your data by subgroup. Does this trend in international security apply to all states, or just to democracies? Do all wars generate many casualties, or just those involving territorial disputes?

Gov329: Empirical memo handout

This is the hardest part of the assignment! Give yourself plenty of time to find and analyze the data. If you are having trouble understanding what the data means for your hypothesis, try graphing the data over time or by other factors. Sometimes a picture can help you see patterns that you might have missed otherwise.

No statistics are required, but you are welcome to conduct more advanced analysis of the data if you wish. All of the analysis I'm looking for can be conducted using Excel or Google Sheets. If you have some experience with statistical software and want to give more advanced tools a try, please feel free. Because of the short time frame for completing this assignment, however, this probably isn't the best time to learn Python or R if you've never used them before.

4. **Describe your findings.** Include a chart or graph that shows how the data bears on your hypothesis and explain this chart or graph in your memo. Discuss whether the data supports your hypothesis.

Please label the axes of your chart or graph, give it a title, and provide enough of a description in the text of the memo that I can understand what the chart or graph represents. When designing the chart or graph, think about whether you've chosen the best approach to highlighting the important patterns in your data, and try to avoid common data visualization errors (refer to the reading and slides for 9-10-21 for more on this). You may use more than one chart or graph, if that helps you illustrate your findings, but only one is required.

You must create your own chart or graph. Don't use one that someone else made that you found on the internet. The purpose of the assignment is to give you experience working with data. Your memo, including your chart or graph, should show me you did that.

Your empirical memo should be no more than 5 pages of double-spaced text, using a standard 12-point font, not including the chart or graph. Please number your pages. You can use whatever citation style you prefer, but don't forget to cite the source for the data you employed. Submit the memo via Blackboard before class on October 4.

I am flexible about how you format your memo. But in general, I'd recommend spending about 1–2 pages on your hypothesis and why it is interesting, and 3–4 pages describing your findings and your chart or graph. You don't need an introduction or a conclusion.

A few tips:

- *Don't drop data (without good reason).* Good analysis requires simplifying complex data, but don't oversimplify. If you're going to drop data from your analysis, you should have a good reason for doing so (for example, if your hypothesis is about the behavior of democracies, it might make sense to drop autocratic states from your

Gov329: Empirical memo handout

data). Don't limit yourself to a small number of countries or a small number of years. If dropping data makes your analysis much easier, then you're probably not approaching the analysis in the right way. One goal of the assignment is to give you experience working with a large dataset—don't try to avoid doing this.

- *Combining datasets can be tricky (but you might need to do it).* If possible, try to get all your data from one source—this avoids the pain of having to merge two different datasets. But if your hypothesis involves the relationship of two different substantive variables, you might have trouble finding a single data source that includes both.

To combine two datasets in Excel:

1. Filter them so they cover the same time range
2. Sort them by country (or conflict or dyad or whatever) and year
3. Paste them next to each other in a spreadsheet
4. Go through and fix any mismatches

I'll talk more about this approach in class.

Another option is to use a dataset builder or data source that can combine datasets for you. Some options include:

- World Bank data: databank.worldbank.org
- NewGene: newgenesoftware.org
- UCDP event data: www.ucdp.uu.se

Finally, if you can't find data for your hypothesis or if it's too difficult to combine the datasets that would make it work, just change the hypothesis to something else you're interested in with data that's easier to work with. I'll never know!

- *Don't wait until the last minute (really!).* Yes, you can write five pages the night before the assignment is due, but data analysis doesn't work like that. Something will go wrong. You might not be able to find the data you want easily. You'll have some false starts and have to backtrack and try a different approach. Excel will do something annoying. If you're trying to do the data analysis part of this assignment on a short timeframe, you will get frustrated and angry. Please, give yourself plenty of time for at least the data analysis part of the assignment.

Need help?

- Check your class notes from 9-10-21 and review the data walkthrough script posted on Blackboard.
- Email me or contact me to set up a time to meet.