Chapter 1

The Artist's Apprentice: Learning to Research

Learning Goals

By the end of this chapter, you will be able to:

1. Explain what is difficult about becoming an independent researcher #taste

2. Outline the approach that this book will take #approach

3. Recognize the core competencies you will need to do quantitative research in Peace and Conflict studies #artistry

You probably didn't get a Ph.D. to learn statistics. Statistics is an art. At least, applied research statistics is. You did not get your PhD in mathematics. You did not get your PhD in statistics. If you did, then for you statistics is a science -- you have to advance the knowledge of statistics further. But you, dear peace and conflict scholar, you have or are working on your PhD so that you can study something important to you: war, peace, peacebuilding, development, diplomacy, nuclear weapons, small arms, economic dependence, postcolonial states, negotiation, mediation, disarmament, demobilization, reintegration, peacekeeping, nonviolence, cultural violence, gender, sexuality, religion, ethnicity, memorialization, identity, cultural heritage, terrorism, technology...there are an endless number of things that are caught up in conflict, violence, and the everlasting quest for peace. This is your muse -- that thing you are pursuing; the itch you are trying to scratch.

Beyond wanting to know more about a topic, there are a lot of other good reasons to do a PhD:

- For the challenge
- For the love of academics
- Because you have a goal of applying conflict resolution and peacebuilding in your life or work
- To get the letters by your name
- To open new doors
- For a chance to study and live abroad

You probably have a mix of these motives. These are all valuable. But whatever your muse is, it probably is not statistics. Still, as a PhD, you need to at least be willing to learn it. You should, at a minimum, have a general literacy in statistics -- and more generally, you should be numerate: able to work with numbers. If you remain in academia, this knowledge will help you understand other research in your field. Even if you focus on qualitative methods, you will need

to position your contribution against developments in the entire field, not merely the qualitative ones. If you transition into government, NGOs, or industry, your PhD will make you the person people ask to understand complex developments and interpret data of all types, both qualitative and quantitative. There is no avoiding this responsibility.

At a deeper level of engagement with statistics, you may realize that learning statistics will help you achieve what you really want. If you can do statistics, then you can learn more about your topic, advancing research, pursuing your muse. Either way, statistics for you is not a science, it is an art that you need to master in order to do science.

Statistics is an art. You have to practice it to become better. But you can also learn from mentors, not just your advisor, the people on your committee, or your professors, but also books like this one. Lectures online. Blog posts from applied statisticians. Great artists steal. That is to say that they borrow ideas from the greats that came before them, and they make these ideas their own. To become a statistical artist, you must embrace the role of the apprentice, learning from the great artists in the studio (i.e., your program) as well as the wider art world (i.e., your discipline).

If you start to think this way, you should always be looking at other work to see the way they applied statistics, how they develop and present their methodology, the way they report their results. But be warned, not all examples are worth following. You have to develop your own taste –- your sense of what is good and bad. That takes time, practice, and learning. This book is here to help you develop good taste as a statistical artist.

This book is about helping you through the methodology of peace and conflict research with an emphasis on quantitative methods. This may excite you. More likely, this intimidates you. Stats is a difficult subject, and the way most people teach it doesn't help. They focus on mathematical formulas; they use the arcane but commonplace jargon of statistics; they assume you will instantly see ways to apply stats; they do all this while teaching you a computer program or software that they barely understand themselves.

I will not follow that path. I have developed my own taste as a statistical artist and mentor. What I promise you is a book focused on you, the peace and conflict researcher, not on the discipline of statistics.

The Apprenticeship

Academia is an apprentice system, make no doubt about that. Like the pupil and an artist studio, you may take classes, depending on which country you are in for your PhD. No matter where you are, you will certainly work with an advisor, your committee chair. You are the apprentice, and your chair is the artist. Your goal is to join them and become their equal in title, one day maybe surpassing them in skill.

The apprenticeship is a way to learn and grow....

However, it is also incredibly frustrating. The biggest frustration is feeling as though you will never be as good as them. Even once you overcome this self-doubt, it is not clear at what point you have become good enough for them to grant your degree. How long does the apprenticeship last? There are two ways of answering this.

First, the formal answer is: your apprenticeship lasts until you finish your coursework, complete your required exams, and defend your dissertation. Each academic program is different. Each has different expectations about how long you stay in the program. Some want to rush you through in say three or four years. That is good for you because it keeps your costs down. This includes tuition costs, if you are unfunded, but also opportunity costs, even if you are funded. You would likely make much more money than your assistantship pays you.

Other programs insisted you stay five or six years. That is not a bad thing. These programs may be more time-intensive, but they usually offer excellent professional and academic experience over that time. You will likely publish multiple articles and develop as an academic. And indeed it does take time to learn to be a researcher and to develop expertise. For programs that rush you through, you may feel as though you have finished a bit early, not entirely sure of yourself, and maybe not without many publications. But of course, then you can still work on getting those publications after you've graduated. For the programs they keep you in, it may feel like you've been in an eternity. You may feel ready to leave. That is a good sign. The only thing in your way is a successful dissertation.

However, here comes the second part of the answer: your dissertation has to be good enough. Unfortunately, no one can tell you what "good enough" means in advance. This is something you have to see for yourself.

Yes, the dissertation is a written document, so yes, you do have to write something and present something. It has to be of sufficient length, although again, it is hard to say what counts as sufficient. For some, sufficient length means a compilation of three of more articles you have published or plan to publish, all tied together by introduction and conclusion chapters that frame the work as part of a larger research agenda. For others, the dissertation takes the form of a "big book." That is, one long document, roughly the length of a book, that tells a coherent story of a large research project.

But the length is one thing, the quality is another. And the quality is difficult to judge. You start this project as an amateur, but you have to end as a professional. The determination that your chair and committee make is when you have crossed that threshold. It sounds vague because it is. Ultimately, this is a matter of quality, yes, but more importantly a it is a matter of social relations. You must convince them that you are a professional. Part of that comes down to quality, part of it comes down to presentation, but another part of it comes down to your confidence.

During my dissertation, I presented a draft to my advisor. He called me in his office after reading through it. On the table in his office was my dissertation printed out. It was open to a table presenting regression statistics. The statistics assessed how well miltiary veterans reintegrated into civilian life given a host of demographic factors, including what branch of the military they had served in. He began to ask me questions about the table, including one question: why had I left out "Army" from the branches in the table. In my mind, I knew the correct answer: Army was the "reference group," and you never include the reference group. And I told him that, but I said it in a way that was not sure of myself. When he asked me the question, I immediately doubted what I thought I knew. Instead of thinking, "here is someone who does not understand this simple thing about regression," I thought, "My advisor may know something that I do not."

These kinds of interactions occur all the time with your advisor and senior colleagues. They are not all intended to be assessments of your readiness. Still, they will make assessments, partly to know how to guide you further, but also partly to understand if you are ready to defend your dissertation, graduate, and join their ranks a fellow PhD.

After that meeting, I consulted by textbooks. Yes, it seems I was right. But I also asked another member of my committee. I told him what happened and why I was starting to doubt myself. He laughed and said, "Oh yeah. I remember when I realized that I knew more about statistics than my advisor." It was at that moment that I realized not everyone knows everything. When you start the PhD, your advisor seems all-knowing. It seems like you will never reach their level. And yes, your advisor is likely quite smart and quite experienced. But they don't know everything. They have their strengths and weaknesses just like you do. They are still learning. It turns out my advisor's strength was not statistics. My strength was. At that moment, I recognized that I had surpassed him, at least on this narrow aspect of my project. There were still other aspects that I had yet to prove myself. Now that I have the PhD and I still work with my advisor, I realize we both have our strengths and weaknesses. There are areas where he depends on me for my expertise and I depend on him for his expertise. You don't have to know everything, but you do have to know the things in your dissertation well enough to be confident when you talk about them. This includes recognizing when you need more help and when you have room to improve. This means that you have developed the skill, expertise, and awareness to judge your own work, not to wait on others to judge it. If you can judge your own work, then you can improve the parts you determine are weak. If you can do that, then you have become an independent researcher, and that's what it takes to get a PhD: you become your own apprentice and your own master at the same time.

This is not easy, but this book is here to help you. It focuses on statistics, but my goal is more general: to help you develop confidence in yourself as a quantitative researcher. That requires learning by doing, or in other words, statistical artistry.

Developing Statistical Artistry

Again, quantitative research is an art. By that, I mean you learn by doing as well as through instruction. I also mean that taste matters; there are enumerable ways to analyze the same data. Some are better than others, but it is hard to pinpoint why. You need to develop an internal sense of what is good and what is bad. That's what I call taste.

Artistry requires taste. But it takes more than that. Artistry also requires applied skills and knowledge. For quantitative methods, you have three major skillsets to develop: subject matter expertise, statistical concepts, and computing. I want you to develop all three, not just teach you statistics. Doing so, I risk repeating one of the biggest issues with how statistics taught: you have to learn equations, statistical concepts, and computing all at the same time. There are benefits to this traditional approach. Computing can help you do statistics, and doing is the pathway to artistry. And mathematics helps you learn what the computation is doing behind the scenes. It can also help your intuition for statistical concepts. The downside is that computing is difficult and so are mathematical equations. Statistical concepts themselves are not difficult. At least, we will not let them be. We also will not be afraid of math, but we will subjugate it to our understanding of statistical concepts. Concepts are supreme.

Nevertheless, we will have to learn some math. There's different levels of the mathematics that we can teach someone. You can learn the deepest math stats possible, with how to do every single thing by hand and really focus on the math of everything. We can also blackbox away the math, into some extent we almost always do this. Even if we teach you, the math at first, eventually, we just switched to calculating things with the computer and no longer need to ever calculate things by hand. The point in this instance of showing people, the math is really about giving them some sort of confidence in their understanding of what's going on underneath. The black box is nice, but it does help to know what the black box is up to. Sometimes that's just to make it yourself more confident in your interpretations. Other times is to help you diagnose issues with your computer code or to understand when the application that you're using is not a good one. But for the most part, if we add math to the equation, we're making things a lot harder. This book is not about teaching you the deepest depths of this math. We do need to know some of the math and understand what's happening at a conceptual level. We also need to be able to communicate what we're doing in a mathematical notation. That's important for two reasons. For one reason, it gives you greater confidence and specificity in the models that you are fitting. If you can express them in mathematical notation, then you can communicate to people precisely what you're doing. On the other hand, you can also learn better. You'll be able to read more mathematical notation if you can write a little bit of it. And that will allow you to dig deeper into the mathematical stats if and when you need to. The point of this course is to teach you how to write mathematical notation for your own problems. That means I'm going to provide you a lot of that to match different computer code. Again this is so that you can communicate what you're up to, as well as try to have a consistent way of understanding

what these different models do. The focus here is on the practical application of being able to communicate your mathematical model, not to be able to hand-solve mathematical equations.

Also, we do have to learn computing. Yes, this will make it difficult. But the way to learning is through doing. Computing is how you do statistics. That's actually how I learned statistics, by coding first. Through coding, I had to learn statistics, and through statistics, I was able to use coding. They reinforce each other, and it made me a very fast learner. I started from essentially no statistical understanding and within the matter of about four years was publishing research which used statistical analysis. Now those initial analyses weren't very good, at least that's what I think now. But that's a good thing. If you're actually continue to learn, then you should look back on things you did in the past and think, "wow, I was dumb back then." Learning to compute will unlock this growth trajectory for you. With guidance and dedication, you can achieve it.

We will apply all this to a focus of the subject matter, peace and conflict. In particular, we will,,,

No, that's it, I understand. It's overwhelming to learn all these three things at once. If you're a brand new beginner, it sucks to have to struggle with how to figure out how to code things. In a lot of coding doesn't have much to do with statistics at all. It has to do with computer science, with things like memory and ram. Things that aren't particularly interesting to you. I imagine as a scholar of peace and complex studies.

What will it take to learn this stuff?

What to expect

. Approach to Teaching

    1. Streamline
    2. Tripe threat
    3.

2. Approach to Statistics
    1. Bayesian
    2. GLMs
    3. Transparency

My big promise is not to make a difficult subject easier. Stats is difficult. My promise is that I will teach this differently than it is normally taught. That will make things easier, but I

- Do we really need another Bayesian textbook? After all, there are already some really good ones:

    - Statistical Rethinking, Richard McElreath ([[McElreath, 2015]]) - My personal favorite

    - Bayesian Data Analysis, Andrew Gelman and colleagues ([[Gelman et al., 2013]])- The quintessential textbook. It's heavy on math and thus not great for beginners.

    - Bayes Rules!

    - A student's guide to Bayesian statistics, Ben Lambert - Accessible

    - Doing Bayesian Data Analysis

- But none of these are focused on you, the emerging Peace and Conflict researcher. My target audience is peace and conflict studies doctoral students. You need representation. You need to see examples that are relevant to your study. You need to be inspired. You need to be "disciplined" – that is, introduced to the theories, ideas, problems, and methods that apply to your discipline. This book does that in several ways.:

    - Examples rely on real publicly available datasets that you can use. These include popular ones such as the Correlates of War and ACLED data, as well as newer ones that cover specific areas, such as the Disarmament, Demobilization, and Reintegration (DDR) dataset. For a list of these data sets, check out the [Appendix. Datasets](obsidian://open?vault=Bayesian%20Peace%20%26%20Conflict%20Research&file=Appendix%2FAppendix.%20Datasets). You can directly lift these from this book into your own projects. Likely you will need to combine these datasets with other data sets to fit your particular needs. You may even need to collect your own data. But this book should help introduce you to the wide world of peace and conflict datasets that exist publicly waiting for researchers to work with them.

    - Inter-disciplinary literacy. Examples from peace and conflict, widely conceived, to include different disciplines and theoretical approaches.

- No one textbook can do everything. This textbook has a particular approach. It's somewhere between a statistics book, a methodology book, and a PhD survival guide. My goal is not only to teach you statistics from the ground up, it is to teach you methodology, how it connects to theory, and how you can achieve your goals as a Peace and Conflict researcher.

Real Peace and Conflict Examples

- Another key feature of this book is that we're diving into real world, peace and conflict data. There are many publicly available data sets that exist. Peace and conflict scholars use these all the time. They'll take multiple data sets and combined them together to produce a data set that has all of the variables that they need to answer their specific question, to advance their specific theory. Occasionally, that will mean also collecting or creating new data yourself. By the end of this book, you'll know how to work with these publicly available data sets. How to find them, how to download them, how to use them, how to understand their contents, and how to report everything that you're doing in a way that makes it clear and transparent for your audience.

- In this book, I'm going to introduce you to peace and conflict data sets. Some of these data sets are widely used, while others are more niche. Along the way, I will introduce you to the theoretical problems and practical problems that these data sets might help us answer. I can't pretend to be an expert in all of these areas, so rely heavily on the articles and scholars that help create these data sets. My goal is to help you see the problems that they were trying to address, how compiling data and analyzing it helped them to address those problems. All of this is to inspire you. To give you examples of how people looked at the world, at a certain part of peace and conflict studies, and their particular subject, found issues with it, and then solve them using quantitative data analysis. My hope is that you'll use that as an example in your own research. Together with the statistical concepts and tools that I'm going to teach you to use, when you're done with this book, you should have a better idea of how to tackle real life research problems using basing analysis.

Practicality over Mathematics

- The emphasis of this book is on practicality. When we talk about math, it's to help you get a grasp of the concepts, but then to give you practical steps, you can take to improve the quality of your research. When we talk about coding, I'm here to give you practical ways to implement what we're talking about. We talk about research methods in theory, again. I'm hoping to help give you practical guidance that helps you avoid some of the worst parts of science and improve the quality of your research. Of course, this trip approach, while it's necessary to learn how to use this stuff, practically, also means that we probably are going as deep as some other texts can go on each of these areas. There are books devoted entirely for how to reduce statistical computing. There are books that go much deeper on the mathematics of the statistics we're talking about. There are books to discuss research methods in detail. I my goal here is to give you enough to where you can learn how all these things fit together and then to provide your resources where you can go dig for additional information. I'm also trying to point out the places where I think it's absolutely necessary for you to dig for additional information. Places where it's probably OK to cite this book as a reference, and places where it's not OK to cite this book and where you should really go dig deeper into the research. Again, I'm hoping this will be a

practical experience for you, so that you'll know hey I need to go deeper on this area because the book didn't cover it all, but that you'll at least know how that area fits into the rest of everything else. Workflows is a big part of this secure. Knowing the workflow, knowing what needs to be done at each stage things to look for pitfalls to avoid, that's my big goal for this book.

Organization of the Book

The first section is about giving you a found fundamental foundation that you can use to interpret the later chapters. One chapter focuses on the idea of modeling and the role that it plays in research. We going to depth on exploratory versus explanatory verse confirmatory modeling. We talk a lot about the role of quantitative methods in the sciences, and especially a lot about the problems with quantitative methods. There's just a couple ways to address those problems, and I think Beijing methods is one of them. But base by itself is not enough. We also need much more transparency. Transparency about the methods that are used, how they're implemented, and whether or not and analysis was exploratory or confer Matory. Finally, you need good theory. Analysis without theory is just p hacking.

The following chapter then provides the foundation for math that you're going to need for this book. It lays out precisely what math is pre-required. It's not much. And it's not difficult. It introduces the idea of mathematical notation and why that's important. And it gives you a lot of practical advice for how to learn the math that you need to learn for statistics and why to learn it. This is not a book that's going to encourage you to go super deep on math. We just need to go deep enough to understand what we're doing when we run our statistical programs. But it's really vital because this is how we've maintain transparency. We are transparent about the methods that we're using that requires specifying them mathematically. It's also really important for your understanding of concepts. You don't have to know how to calculate it every time. You just have to understand what the calculation means. It'll give you a better intuition for what statistics are actually doing. It's part of this section. Also talk a little bit about probability. Probability is the foundation of statistics, at least patient statistics. It's used in other types of statistics as well, but often in a very inconsistent and confusing way. Patient statistics is super easy. Once you understand the basics of probability. Imp probability is fun, it can help you learn all sorts of things, like how to win at Yahtzee, or poker, and how to think about the world that you care about. The world of peace and conflict studies, in a way that isn't black and white. In a way that helps you start to think scientifically about things. About their being greater and lesser chances of things occurring. Or about how certain factors are variables increased the likelihood of other factors and variables. At the end of the day, this type of thinking is scientific more than it is mathematical. But we can express it in math in a way that helps us be rigorous, clear, and consistent in our logic. That'll help you be better at theory. That'll help you be better at research. That'll help you be better at science. That'll help you have an impact through your studies on the real world.

The third and final foundational chapter is on statistical computing. Also known as coding. This chapter is gonna lay out for you how you can learn coding, and it's gonna do so in a way that makes it as easy as possible. It'll give you those dozen operations that you'll need to perform and it'll serve as a reference as you move forward in this book. We go into Y coding is a good idea, and again it's because it increases the transparency and accuracy of your work. All of this ties back in to being more transparent, more accurate, and having a greater understanding of the data that you're working with also that you can be a better researcher. Not so that you can be some computer scientist whiz.

With these foundational chapters, having established a base line for what you need for the rest of the book, the next chapters will interweave these three concepts. They'll interweave these three skill sets. Each chapter will introduce concepts to you. Each chapter will provide you the mathematical notation or ways to build a mathematical notation to support those concepts. In each chapter will teach you the fundamental coding skills that you need to perform in order to enact those concepts on a real world data set. The earlier chapters keep this thing simple and the later chapters. It starts to get a little more complicated. Well, it doesn't get more complicated, but instead of going over the fundamentals so much, does the assumption that you've kind of understood the fundamentals from previous chapters and that allows us to focus on the deeper level stuff. This means that the later chapters, although they might use some advanced techniques that you don't use will also cover some codes that you're going to need to use or some math that you're going to need to use no matter what technique you're doing. So I each chapter has a conceptual focus, I encourage you to think about ways to mix and match the concepts in the techniques introduced in different chapters. You may for instance need the coding that's introduced in chapter 10, which entails ABC. But you might want to apply that in a situation that use the concepts from chapter 5 on continuous predictors. And you'll certainly need to discuss things like in sample and out of sample fit, which is discussed in most detail in chapter on machine learning. That's even if you don't use machine learning, but instead use other types of models. The final chapter then tries to bring all these things together and give you away to express everything that we've talked about, to use all the different coding tools, and the mathematical tools, and the concepts, and help you to kind of pick and choose and build your own models in a way that allow you to answer your own questions.

Additional Resources

Practice Questions

1. What barriers to students bring with them to a statistics course that hurt their chances of learning stats?

2. What about traditional statistics courses are difficult for students?

3. What are three main competencies that you need to be a quantitative researcher?