

Miguel Morin

Data science researcher and computer programmer Homepage: columbia.edu/~mm3509
Ph.D. in Economics (Columbia University, 2014) GitHub: [miguelmorin](https://github.com/miguelmorin)
M.A. in Statistics (ENSAE, 2008) StackOverflow: [mmorin](https://stackoverflow.com/users/1122222/mmorin)
B.Sc. in Computer Science (Polytechnique, 2007) Last Updated: August 29, 2018

Summary

I am a data science researcher with experience in data science, machine learning, Python, Julia, Git, GitHub, XML, web (HTML, JavaScript, and CSS), SQL, and Emacs-LISP.

My current tasks consist of data exploration in distributed datasets with Hadoop using SQL and Python and research software engineering using Julia.

I follow the paradigm of test-driven development, where tests are executable documentation. Doc-tests have saved me a lot of time and effort over the years and I like that the documentation is near the code.

I also write documentation using markup languages like Org Mode (Emacs), Markdown, and LaTeX. Plain text provides easy version control and is future-proof. I can generate PDF versions with tools like Pandoc and pdflatex (as in this resume).

I have lived, studied, and worked in different countries (Portugal, France, United States, United Kingdom); I am fluent in several languages (native in English, French, and Portuguese; intermediate in Spanish and German), and my professional experience spans academia, the charity sector, start-ups, and industry.

Programming Languages

Julia: I chose this open-source and fast language for the migration of previous research projects in Matlab, Stata, and R; I am actively involved in discussing issues on GitHub and StackOverflow.

Python: I am a proficient user of versions 2 and 3 and I use libraries such as `re` (regular expressions), `BeautifulSoup`, `Pillow`, and `cookiecutter`.

Command-line, Unix, and bash: I am a proficient user of the command-line in a Unix environment. I have some exposure to bash scripting.

Emacs-LISP: I am in the initial stages of learning this language to customize Emacs, my text editor of choice.

Java: I was an advanced user over 2005-2007 during my university degree and for the ACM competition.

C#: I was an advanced user over 2007-2009 during my work for Amarile and Columbia University.

Experience

Postdoctoral research Associate, The Alan Turing Institute (London, 2017-present): I am currently a Postdoctoral Research Associate in Economics and Data Science. I work closely with HSBC and discuss the results with senior science and business stakeholders at HSBC and at The Alan Turing Institute. I converted research software from closed-source Stata to open-source Julia for an academic revision (public [GitHub repo](#)). I also coded a research project in R for the inference of network structures from panel data (private repo on GitHub). I use Git and GitHub extensively for these projects.

Inventor and designer of a children's literacy toy (2016-present) I invented a children's literacy toy (design files on [Thingiverse](#)) for two 2-year old children. The initial toy consists of a wooden board with mortises for each letter in the child's name and plastic letters that fit into those mortises. I fabricated these

at the [Cambridge MakeSpace](#) with a CNC router and a laser cutter. One year later, I was surprised that the children had taught themselves to write their own names. I decided to take the experiment further and add lower case at the back of the wooden pieces, expand to cursive type, and include numbers, math, and punctuation. I scoped available technologies and chose Python, Inkscape, and SVG. I also used the GNU Linear Programming Kit to optimize the choice of letters to maximize the total frequency of words that can be formed with those letters from the [Essex Children's Printed Word Database](#). The code is on [GitHub](#). The children are currently using it and they could teach themselves to write all letters, a result that I will assess over the next year and that could give them a two-year head start in school.

Software development intern, Simprints (2017) I did an internship as a Cloud Computing Engineer with Simprints, a technology start-up manufacturing fingerprint scanners and providing identification services in developing countries. I consolidated the data from multiple versions of the scanner into a single schema and designed solutions with different cloud tools such as Google BigQuery and Tableau.

Postdoctoral research associate, University of Cambridge (UK, 2014-2017): I was a Postdoctoral researcher in Economics over 2014-2017. I wrote research software in Stata for revisions of my doctoral dissertation. I also lectured for 1st-year Macroeconomics (2014-2017), 3rd-year Macroeconomics (2015), and PhD-level course on data analysis (2015).

Lecturer, University of Cambridge, Faculty of Economics (2014-2016): I taught 1st-year Macroeconomics (three times, 150 students), 3rd-year Macroeconomics (once, 40 students), and PhD-level course on data analysis (twice, 10 students) over 2014-2016. Student feedback available upon request.

Supervisor, Trinity Hall College, University of Cambridge (2014): I supervised a group of 3 students for 3rd-year Macroeconomics.

Teaching Assistant, Department of Economics, Columbia University (2009, 2012): I taught the breakout sessions of Intermediate Macroeconomics and Principles of Economics (2009 and 2012).

Research assistant, Columbia University (2007): I did an internship with a team of two academic researchers working on a [paper on partial identification](#). They wanted a database of prices to test the econometric theories of multiple equilibria in price competition. I designed and coded a tool in C# and .NET that scraped prices and details of every flight to and from every US airport and compiled them into a CSV file.

Software development intern, Amarile (2007): I did a 6-month part-time internship for [Amarile](#) writing commercial software in C#. The clients were oil companies that used an old simulation tool that interpreted a script with data and after several days produced estimates of the quantity of oil in a well. The interpreter did not verify syntax and could throw an error after running for several days, after which the engineers had to correct the syntax and wait several days for another result. My task was to write a C# parser to verify the syntax of the scripts. I read the documentation of the simulation tool, broke down the parts of the script, wrote objects and functions for each, and tested them with several scripts. The result was a success, the client subscribed the software, and saved hundreds of hours of work.

ACM competition (2006): I participated in the [South Western Europe Regional ACM Programming Contest 2006](#) in Lisbon in 2006 with a team of 3 from Ecole Polytechnique, coding in Java. One problem involved prime numbers and our first algorithm failed to process the inputs in the required time. We divided the effort and my task was to speed up the code while they worked on the next problem. I refactored the code to compute a list of all prime numbers up to a certain limit using the Erathostenes sieve. This change brought the algorithm below the threshold, the algorithm was correct, and we got a balloon.

Education and honors

Columbia University (New York): Economics, Ph.D.; 2014. Thesis: "General Purpose Technologies: engines of change?"

Wueller prize, Columbia University (\$2,334): winner of the best dissertation proposal (2011).

Vickrey prize, Columbia University (\$750): runner-up for the best paper of a third-year PhD student, 2011.

Harris prize, Columbia University (\$1,375): winner of the best by a second-year PhD student, 2010.

Ecole Nationale de la Statistique et de l'Administration Economique (Paris): Statistics, M.A.; 2009. This school is a top tier French engineering schools and is the branch school of Ecole Polytechnique (see below) for statistics, data science and machine learning.

Ecole Polytechnique (Palaiseau): Computer Science, B.Sc.; 2008. Ecole Polytechnique is the [#1 engineering school in France](#) with a highly selective entry process.

Bento de Jesus Caraça Prize (3,000 EUR): winner of the prize for the best paper by a high school student on the theme “complex logarithms” ([link to Portuguese press release, in Portuguese](#)).

Funding and awards

Hayek fund for scholars grant (\$400): conference travel, 2015.

Russell Sage Presidential Authority Award (\$23,000): (with Rowena Gray, UC Merced), funding for project “Technological Revolutions and Occupational Change” (2015).

Arthur H. Cole grant (\$5,000): grant to support post-doctoral research, 2015.

Keynes Fund, University of Cambridge (£20,000): grant to digitize the Census of Manufactures during the Great Depression (2014-2015).

Corpus Christi College, University of Cambridge College Research Associate, 2014-2017.

Doctoral fellowship, Columbia University (\$60,000): PhD fellowship, 2008-2009 and 2014.

Program for Economic Research, Columbia University (\$8,000): grant from Columbia University to digitize the Census of Manufactures during the Great Depression, 2013.

Fundação para a Ciência e Tecnologia, Portugal (132,000 EUR): Doctoral fellowship, 2010-2013.

Academic publications and working papers

“**How should the graduate economics core be changed?**” (2010). The Journal of Economic Education. (with Abito, J. M., K. Borovickova, H. Golden, J. Goldin, M. A. Masten, A. Poirier, V. Pons, I. Romem, T. Williams, and C. Yoon)

“**Computer Adoption and the Changing Labor Market**” (2014). Reject and Resubmit at Review of Economic Dynamics. *Media coverage: [Marginal Revolution](#). Invited presentations in 2013-2014: Midwest Macroeconomic Meetings (University of Illinois at Urbana Champaign), Instituto Superior de Economia e Gestão (Lisbon).*

“**The labor market consequences of technology adoption: concrete evidence from the Great Depression**” (2015). Revision Requested at American Economic Journal: Macroeconomics. *Media coverage: [Financial Times \(behind paywall\)](#), [Marginal Revolution](#), [European Economic Association](#), [Vox](#). Invited presentations in 2014: Economic History Association conference, Instituto Superior de Economia e Gestão (Lisbon), Center for Macroeconomics (London and Cambridge), Faculty of History (University of Cambridge); and in 2015 at Economic History Society Meetings (Wolverhampton), Royal Economic Society Meetings (Manchester), NBER Summer Institute (Boston), Econometric Society World Congress (Montréal), European Economic Association Meetings (Mannheim).*

“**Technological Revolutions and Occupational Change: Electrifying News from the Old Days**” (2016, with Paul Gaggl and Rowena Gray). *Invited presentations in 2015: Southern Economic Association meetings, UNC Charlotte, UC Merced, and UC Davis; and in 2016: NBER Summer Institute (Boston, by co-authors).*

“The Effect of Infrastructure Investment in a Low-Growth Environment: Evidence from the Great Depression” (2016, with Scott Swisher). *Invited presentations: in 2015 and 2016 at the Keynes Fund Research day.*

“Tablet language stimulation and sight-reading applications to narrow toddlers’ literacy gap.” (2016, with Sriya Iyer, ethics approval granted, on pause for lack of funding).

Hobbies

Improvised comedy (“Improv”) I attend, lead, and perform in several improvised comedy events in London and Cambridge. The guiding principle is saying “yes, and” to other people’s ideas, and as such it improves teamwork.

Bricolage, arts, and crafts I enjoy making, repairing, and inventing new things with a laser cutter to cut or engrave acrylic, a Computer-Numerical-Control (CNC) router to machine wood and plywood, and 3D printers. The kit is from the Cambridge [MakeSpace](#).

I also have an interest in film-making (see one [short video](#) of a puppet teaching emotions to children as an example) and piloting gliding aircraft (no motor).