

Thiago Veiga Marzagão

tmarzagao@gmail.com, <https://thiagomarzagao.com>, @tmarzagao

PROFILE

- Data scientist with strong statistics background and experience with machine learning and data products. Ph.D. from the Ohio State University. Professor of statistics and data mining at Brasília universities. Researcher w/ peer-reviewed publications in AI conferences and econometrics journals.

RECENT WORK EXPERIENCE

- Observatory of Public Spending (May '16 - present). Position: data scientist. The OPS is the data intelligence unit of Brazil's Comptroller-General. I write algorithms and apps that mine government data looking for evidence of corruption and waste.
- Administrative Council for Economic Defense (Sep '14 - May '16). Position: data scientist. The ACED is the Brazilian antitrust agency. At the ACED I developed and maintained CÉREBRO, an app that mines government procurement data looking for evidence of bid rigging.
- Institute of Higher Education of Brasília (Feb '18 -). Position: professor of machine learning.
- University of Brasília (Mar '16 - Dec '16). Position: professor of data mining.
- Brasília Institute of Public Law (Mar '16 - Dec '16). Position: professor of statistics.

SKILLS

- Data science: statistical inference, clustering, linear regression, logistic regression, random forest, SVM, neural networks, dimensional analysis, spatial econometrics, time series econometrics, survival analysis, Monte Carlo, topic modeling, feature engineering, feature selection.
- Programming: Python, R, Matlab, Stata, Cypher (Neo4j), SQL, HTML/CSS, JavaScript, basic notions of C/C++.
- Systems: Linux, cloud computing (AWS, GCP), web app platforms (Google App Engine, Heroku).
- Languages: English (fluent), Portuguese (native), Spanish (read and listen), French (read and listen).

CODE (<https://github.com/thiagomarzagao/>, <http://thiagomarzagao.com/code/>)

- ev3py. Python module that lets users interact with LEGO Mindstorms EV3 bricks using intuitive, easy-to-understand methods.
- CATMATfinder. Web app that suggests likely CATMAT classes based on product descriptions (CATMAT is the taxonomy the Brazilian government uses to classify the merchandise it buys).
- Wordscores in Python. Python implementation of the popular Wordscores algorithm, used to analyze political texts.
- Fightin' Words in Python. Python implementation of the popular Fightin' Words algorithm, used to analyze political texts.
- CÉREBRO. App I developed in-house at the ACED. It mines public procurement data looking for evidence of bid rigging. It is a web app, mostly Python-based, with some of the screening algorithms implemented in R. It only runs on the ACED's intranet and its source code cannot be publicized.

PUBLICATIONS (<http://thiagomarzagao.com/publications>)

- Automated Democracy Scores. *Brazilian Review of Econometrics*, 37(1), 31-43, 2017. Here I text-mined 42 million news articles and used machine learning algorithms to assign democracy scores to every independent country from 1993-2012.
- Deep Learning Anomaly Detection as Support Fraud Investigation in Brazilian Exports and Anti-Money Laundering. (With Ebberth Paula, Marcelo Ladeira, and Rommel Carvalho.) ICMLA 2016 paper. Here we mined export data to find fake exports (which is a possible indicator of money laundering).
- A dimensão geográfica das eleições brasileiras. (The spatial dimension of Brazilian elections.) *Opinião Pública*, 19(2), 270-290, 2013. Here I used spatial econometrics to understand voting patterns in Brazil.
- Lobby e protecionismo no Brasil contemporâneo. (Lobbying and protectionism in contemporary Brazil.) *Revista Brasileira de Economia*, 62(3), 263-178, 2008. Here I used regression to understand tariff policy in Brazil.
- Several tutorials (on webscraping, Django, robotics, other topics): <http://thiagomarzagao.com/index/>

DEGREES

- Ph.D. in Political Science, Ohio State University (Aug '14), with a Fulbright scholarship. The first two years I took lots of stats classes. The last two years I mined news articles to automate the measurement of democracy.
- M.A. in Political Science, Ohio State University (Dec '12).
- M.A. in International Relations, University of Brasília (Jun '07).
- B.A. in International Relations, University of Brasília (Jul '03).