

JORGE CIMENTADA - 697 382 009

- **Languages:** Native Spanish, High-level English

- **Age:** 27

cimentadaj@gmail.com | @linkedin: [Jorge Cimentada](#) | cimentadaj.github.io | github.com/cimentadaj | Twitter: [@cimentadaj](#)

PROFILE

Data analyst with background in statistics, predictive analytics and machine learning. Experience working with causal inference, big data, business intelligence and R programming. Passionate about teaching R, data science and explaining my research to non-technical audiences

EDUCATION

Pompeu Fabra University Jan 2016 - Present

PhD Candidate Quantitative Sociology - Expected autumn 2018

My research involved collapsing data on over **3 million students** to study educational policies and its impact on inequality. The key takeaway is that inequality depends heavily on the level of vocational education of the country.

Pompeu Fabra University Sep 2014 - Sep 2015

M.A Sociology and Demography

Best master's thesis (2015): Only 1 in a class of 25. Featured in master's website.

PUCMM University Sep 2009 - Sep 2013

B.A. in Business

WORK EXPERIENCE

R developer & Data Analyst: Research and Expertise Centre for Survey Methodology (RECSM) Jan 2018 - Present

[R, MySQL, Web scraping, API access]

This project involved designing and implementing two Shiny apps. I applied **text mining** techniques to extract and translate text using the Google Translate API and created a library that **reimplemented several algorithms** for adjusting **model predictions** for measurement error.

Data Scientist: Pompeu Fabra University Jan 2016 - Jan 2018

[R, Rmarkdown, LaTeX]

As the **data analyst** for a research project, I exercised the complete research pipeline of accessing **dirty data**, cleaning, analyzing and reporting. In one case, I used several **classification algorithms** to estimate the propensity of mobility between social classes, landing on LDA because the predictive classes were stable and had the lowest out-of-sample error.

Statistical consultant: Population Services International Aug - Sep - 2014

[Causal Inference, Reporting]

I designed, implemented and analyzed the **randomized controlled trial** related to HIV prevention. The project increased HIV awareness by 7% in the population of armed forces resulting in a **renovation of funding** from the Department of Defense.

Junior analyst: Newlink Research Jan - Aug - 2013

My main job was **statistical consulting** within the company. I worked on projects related to **customer churn**, banking, retail and **public opinion/political forecasts**. The latter involved live-analysis of opinion polls in combination with the Miami Headquarters.

TECHNICAL SKILLS

Machine Learning: classification, regression, clustering, feature engineering

Statistical Methods: regression models, hypothesis testing and confidence intervals, principal component analysis and dimensionality reduction, text mining, multilevel models, propensity score matching.

Software and Programming Languages: R (+ 3 years, dplyr, ggplot2, Rmarkdown, caret, tidyverse), Python (< 6 months), SQL (6 months), Regular Expressions, Bash, Git, Linux, Microsoft Excel, LaTeX

SOFTWARE

sqpr: Access over 40,000 question from the SQP API and adjust your estimations for measurement error
essurvey: Download data from the European Social Survey [CRAN, ropensci.github.io/essurvey/, > 1000 downloads]
perccalc: Estimate continuous distributions from categorical variables [CRAN, github.com/cimentadaj/perccalc]
cimentadaj: Personal package [github.com/cimentadaj/cimentadaj]
corr: Exploring correlation in R in a 'tidy' framework. ****Contributor**** [github.com/drsimonj/corr]

SELECTED PRESENTATIONS AND TEACHING

"Introduction to R programming", RECSM Data Science Summer School 2016 [Jul 2018 booked]
"Introduction to the tidyverse and functional programming", UPF PhD training, March 2017
"Introduction to the tidyverse" - PyData Conference, May-2017 [invited talk]
"The essurvey package" - ESS conference at the University of Southampton, January 2018 [invited talk]

SELECTED INDEPENDENT PROJECTS

See at cimentadaj.github.io/projects/

- * Scraping and analyzing the scripts of How I Met Your Mother
- * How I programmed a Twitter bot to clean and visualize data
- * Shiny App to visualize educational inequality for over 70 countries in 15 years
- * All solutions to the R for Data Science book

Or check out my blog for shorter posts on machine learning and data analysis at cimentadaj.github.io/blog/. Some example posts:

- * Brief Analysis of Independent/Unionist Vote in Catalonia
- * Hold and Cross-validation series of posts