

**Jorge Cimentada - 697 382 009 - Escudellers de Blancs #1, Barcelona - [cimentadaj@gmail.com](mailto:cimentadaj@gmail.com)**

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

---

## PROFILE

Data analyst with background in statistics, predictive analytics and machine learning. Experience working with causal inference, big data 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  
MSc Sociology and Demography  
**Best master's thesis:** First of class of 25. Featured in master's website.

PUCMM University Sep 2009 - Sep 2013  
B.A. in Business

---

## WORK EXPERIENCE

**Data Scientist:** Research and Expertise Centre for Survey Methodology (RECSM) Jan 2018 - Present  
[R, MySQL, Web scraping, API access]  
This project involved designing and implementing two web applications (based on Shiny apps). I applied **text mining** techniques to extract and translate text using the Google Translate API. I created a library that **reimplemented four algorithms** for adjusting **model predictions** for measurement error.

**Data Scientist:** Pompeu Fabra University Feb 2016 - Jan 2018  
[R, Rmarkdown, LaTeX]  
I exercised the complete research pipeline of accessing **noisy data**, cleaning, analyzing and reporting. In one case, I used several **classification algorithms** to estimate the propensity of mobility between social classes. I compared **Linear Discriminant Analysis, Support Vector Machines** and **Naive Bayes** algorithms and landed on LDA because the predictive classes were stable and had the lowest out-of-sample error.

**International Business Analyst:** Trovit Jul 2015 - Feb 2016  
My work was focused on analyzing click-through rates and price-per-clicks to create a better strategy for different country managers. My analysis helped launch the internationalisation of 4 new products (+2 million potential each), all decision taken from quantitative (data) and **time-series analysis**.

**Statistical consultant:** Population Services International Jul - Aug - 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 2013-2014  
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, Git, Bash, Docker, Linux, Microsoft Excel, LaTeX

---

## MY SOFTWARE

---

### R libraries:

**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](https://ropensci.github.io/essurvey), > 1000 downloads]

**perccalc:** Estimate continuous distributions from categorical variables [CRAN, [github.com/cimentadaj/perccalc](https://github.com/cimentadaj/perccalc)]

**cimentadaj:** Personal package [[github.com/cimentadaj/cimentadaj](https://github.com/cimentadaj/cimentadaj)]

**corr:** Exploring correlation in R in a 'tidy' framework. **\*\*Contributor\*\*** [[github.com/drsimonj/corr](https://github.com/drsimonj/corr)]

---

## LANGUAGES

---

English: fluent

Spanish: native tongue

---

## PRESENTATIONS, TEACHINGS AND PROJECTS

---

**"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/](https://cimentadaj.github.io/projects/)

- \* Scraping and analyzing the scripts of How I Met Your Mother

- \* Analyzing bicycle waiting time in Barcelona

- \* How I programmed a Twitter bot to clean and visualize data

- \* Web shiny application 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/](https://cimentadaj.github.io/blog/). Some example posts:

- \* Brief Analysis of Independent/Unionist Vote in Catalonia

- \* Hold and Cross-validation series of posts