

Hi everyone, I'm Santiago Ortiz. I lead Moebio Labs, where we constantly experiment with data and interaction; our aim is to create tools that connect Big Data and Cognition. Ask Me Anything!

moebio ¹ and r/Science AMAs¹

¹Affiliation not available

April 17, 2023

Abstract

Santiago Ortiz is a mathematician, data scientist, information visualization researcher and developer. He uses his background in mathematics and complexity sciences to push the boundaries of information visualization and data based storytelling. In 2005 he co-founded Bestiario (Barcelona), the first company in Europe devoted to information visualization. He currently leads Moebio Labs. Moebio Labs is a team of data scientists, data visualization developers and designers. We develop advanced interactive visualization projects that connect with huge data sets. Our methodology and projects are designed to get deep insight from data in collaboration with the client, solve real problems and answer strategic questions. We work for clients around the world. To see my work check Moebio.com — there is a navigation widget on the bottom left. These are my projects from the past 3 years. I recommend seeing Lostalgic and Twitter using Twitter. Now Moebio is a team, and we are delivering similar interactive experiences as in these experimental projects, except that: data is real (with people in companies as opposed to people in islands), we aim to align with companies' strategies and goals, and we are infusing predictive modeling into the visualizations. It's not only that we visualize prediction model results, but that the visualizations allow users to modify and tune the models. Our goal is to help companies becoming collaborative-data-driven. We're about to open the Moebio framework, a JS framework for data wrangling, exploration and visualization (working hand in hand with Bocoup on this). We're also close to start sharing Lichen (mail subscription), our modular environment, in which data projects (wrangling, modeling, analysis, visualizations) can be built in seconds, even by non-developers... and developers can add their own technology-agnostic modules. Here's proof that it's me. I'm here to talk about managing teams of data scientists, working with big data, predictive modeling, or anything else. Ask Me Anything! Today, coinciding with this AMA, we are releasing the free open source version of the Moebio Framework. Ask Me Anything about this as well! Thanks everyone for the clever questions, It was very interesting and fun! Always happy to continue conversations via twitter: <https://twitter.com/moebio>

the
WINNOWER

[REDDIT](#)

Hi everyone, I'm Santiago Ortiz. I lead Moebio Labs, where we constantly experiment with data and interaction; our aim is to create tools that connect Big Data and Cognition. Ask Me Anything!

MOEBIO [R/SCIENCE](#)

ABSTRACT

Santiago Ortiz is a mathematician, data scientist, information visualization researcher and developer. He uses his background in mathematics and complexity sciences to push the boundaries of information visualization and data based storytelling. In 2005 he co-founded Bestiario (Barcelona), the first company in Europe devoted to information visualization. He currently leads Moebio Labs.

Moebio Labs is a team of data scientists, data visualization developers and designers. We develop advanced interactive visualization projects that connect with huge data sets. Our methodology and projects are designed to get deep insight from data in collaboration with the client, solve real problems and answer strategic questions. We work for clients around the world.

To see my work check [Moebio.com](#) — there is a navigation widget on the bottom left. These are my projects from the past 3 years. I recommend seeing [Lostalgic](#) and [Twitter using Twitter](#). Now Moebio is a team, and we are delivering similar interactive experiences as in these experimental projects, except that: data is real (with people in companies as opposed to people in islands), we aim to align with companies' strategies and goals, and we are infusing predictive modeling into the visualizations. It's not only that we visualize prediction model results, but that the visualizations allow users to modify and tune the models. Our goal is to help companies becoming collaborative-data-driven.

We're about to open the Moebio framework, a JS framework for data wrangling, exploration and visualization (working hand in hand with Bocoup on this). We're also close to start sharing Lichen (mail subscription), our modular environment, in which data projects (wrangling, modeling, analysis, visualizations) can be built in seconds, even by non-developers... and developers can add their own technology-agnostic modules.

[Here's proof that it's me.](#)

I'm here to talk about managing teams of data scientists, working with big data, predictive modeling, or anything else. Ask Me Anything!

Today, coinciding with this AMA, we are releasing the free open source version of the [Moebio Framework](#). Ask Me Anything about this as well!

Thanks everyone for the clever questions, It was very interesting and fun! Always happy to continue conversations via twitter: <https://twitter.com/moebio>

[READ REVIEWS](#)

[WRITE A REVIEW](#)

CORRESPONDENCE:

DATE RECEIVED:
September 16, 2015

DOI:
10.15200/winn.144232.28402

ARCHIVED:
September 15, 2015

CITATION:
moebio , r/Science , Hi everyone, I'm Santiago Ortiz. I lead Moebio Labs, where we constantly experiment with data and interaction; our aim is to create tools that connect Big

Thanks for the AMA - I love your field of interest. John Snow's Ghost Map was one of the earliest and most elegant of the statistics + visualization = eureka! As an interest user, how could I plug into your Moebio Framework?

[dogthistle](#)

The best way to get introduced to the Moebio Framework is via this post by @bocoup <https://bocoup.com/weblog/introducing-moebio-framework/>

Then you can download it and just start using it. You have the documentation here: http://moebiolabs.github.io/moebio_framework/docs/

And if you want to have a fast glance, check the examples and play with them. For instance, you can change the number of nodes and probability of relations in this network example: http://moebiolabs.github.io/moebio_framework/examples/playground/#/spanning_tree

Ah, the Ghost Map indeed; I love it, but not as much as Alberto Cairo that has, I think, a collection with

Data and Cognition. Ask Me Anything!, *The Winnower* 2:e144232.28402 , 2015 , DOI: [10.15200/winn.144232.28402](https://doi.org/10.15200/winn.144232.28402)

© et al. This article is distributed under the terms of the [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and redistribution in any medium, provided that the original author and source are credited.



t-shirts with the map on it.

Can you remember a time where the use of statistics dramatically changed your opinion on something? A scenario where the stats disproved many of your preconceived notions about a topic?

[rhiever](#)

It's happening all the time... and to many of us. Take only the general statistics on violence, and how they dramatically contrast to the general perception. It's not so different to the time it was already known that the Earth was not plane, except that it was not know... a huge discrepancy of views. It also stroke me, when first read that datasets from different origins were telling that violence is declining worldwide, and constantly at least since the 50s, I didn't believe it. So I read Pinker's book. http://www.amazon.com/Better-Angels-Our-Nature-Violence-ebook/dp/B005HHSYMW/ref=sr_1_1?s=books&ie=UTF8&qid=1442338448&sr=1-1&keywords=better+angels+of+our+nature and got convinced. And then an entire trend started, including Max Roser's Our World in Data <http://ourworldindata.org/> and many other communicating data that basically contradict this deeply engrained idea, fed by mass media and news, that the world is now on its worst moment.

And that's just one example. All the time I'm contradicting by data, in big scale or small scale. Very often in my work, with assumptions I have about a reality I'm analyzing and visualizing.

why did you decide to start bestario?

[Nisha_the_lawbringer](#)

I think the real reason to start that company (in which I do not longer work since 2012) was that I met the best and most complementary people to do so! We wanted to have fun, we knew data would be big, we had ideas and tools. I personally was a little bit tired of doing experimentation and research, and wanted to work for clients (never stopped experimenting though). It was a great adventure, there was no visualization market, specially in Spain, so we had to create it.

Is there a big demand for the type of data visualization you do? Who are you primary clients?

[poopvotes](#)

There is. Moebio website features the more experimental projects we had created. In more private presentations we show work done for clients. Which is actually not completely different. It's also very interactive and to some extent experimental, but it's meant to really answer questions and help on the strategy. They also contain much more data science than the moebio.com featured projects show.

There's a lot of demand, organizations are becoming frustrated with two things: conventional BI, and cryptic data science that only delivers unidimensional answers.

For someone who does a lot of d3, and other JS based visualizations, what does moebio framework offer that those tools might lack?

edit: also what's up with the 'M's in your verification pic? :). Cheers

[nthitz](#)

I can't really talk for what these other tools miss. I think the best way to understand the framework strengths is by looking to the projects it has produced. For the time being these projects are Moebio's, see <http://moebio.com/>, since we have been the only users of the framework; but even so, I think there are some clear hints about what the framework allows to do.

Again, I'm not very aware of other frameworks, but I think a differential factor is that with Mo you always know the type of the data, and for each type you have a very clear list of operations you can make. So if you have an array of array of numbers, it's actually a NumberTable, and you can do things such as `numberTable.getNormalized()` and obtain a new NumberTable. Then with the NumberTable you can build a network (using the #T as adjacent matrix), then obtain clusters from the network, then rebuild a table this time form the nodes in clusters and reading some property from them, etc... I think this kinds of flows are easy with Mo because for each type there's a clear way to (combined with other types, optionally) create new types. There's a sort of combinatoric grammar of types, and then a flow

(each type can be converted into other types without losing information).

No idea why my Ms are so baroque, I think it's because I borrowed my wife's calligraphy marker.

What are you favourite books on data visualisation and analysis? In general, what is the best way to learn that type of thinking you need to create useful reports?

[sztanko](#)

Please check this answer:

https://www.reddit.com/r/dataisbeautiful/comments/3l1pmw/hi_everyone_im_santiago_ortiz_i_lead_moebio_labs/cv2mpha

My favorite is probably the aforementioned data science for business, I also like very much this one:

<https://www.goodreads.com/book/show/17397466-an-introduction-to-statistical-learning>

Why do people in the information systems industry like to use buzz words?

[60daygoal](#)

Only in information systems industry? I guess we need them in order to show other people we are not behind... but, more importantly, we need them to not longer use them, and show other people we are ahead.

What is your favorite statistical anomaly?

[rhiever](#)

Maybe our Universe, capable of generating Life. It's not even clear if that's an statistical anomaly, but the way some "parameters" are tuned, so complexity and in particular life (and even consciousness) can exist, it's so unlikely, its probability so low, that scientists come up with the most craziest ideas. Very famously the anthropic principle, which many disregard as non scientific and even esoteric. And then you have the theory of evolution through selection of nested universes!

[http://evodevouniverse.com/wiki/Cosmological_natural_selection_\(fecund_universes\)](http://evodevouniverse.com/wiki/Cosmological_natural_selection_(fecund_universes)) (it seems that black holes and life require "similar" level of complexity to exist, and that they reproduce and are selected by that complexity level)

What type of work is being done for data visualizations for the blind? Other than our sense of sight, what other sense do you think is a great way to convey data? Or is accessibility a pretty low-priority item right now in the field of dataviz research?

[samxli](#)

Data visualization for the blind is an interesting oxymoron... except it's not if we reckon visualization in a broader sense, as obviously you do. Data Somatosensoring?

Interesting indeed! I have little information about visualization done for the blind, except perhaps for this collection of material visualization. Probably not meant to be devices for blind people, but most of them would work for that purpose. <http://dataphys.org/list/>

UPDATE: it seems that some maps there are meant to be used at night, and I adventure to say that very probably were used by blind people, that were/are archetypically considered as capable of "seeing further".

UPDATE 2: one of the items in the list is actually for blind people: <http://dataphys.org/list/tactile-infographics/>

Many years ago I tried to build a sound game... actually it started being visible and faded to black. It was a sort of sound-pong, with spatial sonorization. The algorithm for spatial sound was bad and the game didn't work.

Coincidentally I'm reading now a book about a blind girl that lives in Saint-Malo and whose father build for her a complete 3D model of the city, so she would learn to memorize it, and then learn to walk it. Each Tuesday she practices trying to get to her home from a distant point in the city (her father always nearby) and guess what happens,... read the book: <https://www.goodreads.com/book/show/18143977-all-the-light-we-cannot-see>

Also, in one client's project we tried sound. Just because the time patterns were very interesting, clearly something special, and actually the first visualization results spelled "music". Maybe the ear can at least re-inforce somehow what the eye sees? The hypothesis was that by listening the rhythm you won't get any insight... but you could eventually recognize a pattern in time that's somehow special or different from others, and that would be interesting to analyze. I won't say that sonorization was a success or that the hypothesis was proven... but sonorization did something interesting for the project: it made people become more interested on it, in a way it helped 'opening the eyes' to the interesting time patterns of transactions.

Thanks for the AMA! As someone who routinely works with small and medium sized data sets, I have a couple questions:

- **How do you think the industry as a whole can make robust data analysis more accessible to small and medium-sized organizations?**
- **At Moebio, how do you approach the trade off between information delivery and simplicity for data visualization?**

[ckwilson912](#)

- Data science will become more accessible for multiple reasons. The knowledge is spreading, executive people in organizations are more aware and many are actually getting data savvy... inversely, data scientists and statisticians are becoming more connected with the context of the companies, coming closer to the business side. This encounter of usually isolated people and teams will transform data science and the tools, that are also becoming easier and faster.
- Simplicity, in many cases, means that a clear, communicable insight, has been found in the data. So a good visualization should always be, in some way, simple. Obviously complexity lives in the data, specially when there is collinearity in more than two or three dimensions. In those cases, on good strategy is to build 'multiple simplicities', different visualizations that offer different views on the data, each one being simple, but building in the overall a more comprehensive perspective on the data. Imagine a table that contains multiple categories a numerical variables. One can focus on the correlations of the numerical variables, that could produce a network... and maybe that perspective is interesting. Another approach would be a matrix of scatter plots... and maybe some of those are particularly interesting and worth be represented isolated. Then, from the categorical lists one can find interesting relations, such as one being a sub-category of other, with some numbers being weights. That could lead to an interactive treemap in which the user can actually switch from numeric variables to assign weights, or color nodes according to a numerical threshold etc... I advocate for the multiplicity of approaches, specially in early fases of a project in which one wants to discover things in the data, and each approach being simple.

Hi Santiago,

I've been following your work for some time now, which includes peeking occasionally at the code of moebio.com . I noticed that you do not use d3 but rather pure JS in an object oriented approach. I had myself some trouble using d3's selection and data binding paradigm, as I was mostly used to code in Processing or Openframeworks. Is there a reason why you favor one approach over the other?

[sergiobd](#)

When I moved from actionscript to javascript, working with DOM objects and with SVG in particular, seemed to be the natural path. But I immediately recognized lots of limits, specially in terms of the freedom I wanted to draw and build interaction. I do not operate (at least not always) under a linear 1-1, datapoint to graphical object, approach. I needed something more fluid. I also build spaces that are neither 2D or 3D but that operate under different logics, so I really needed 100% freedom in presentation. The canvas provides that for you, in principle, controls each single pixel. The very idea of an element, that can be hovered, selected, dragged. it's just a pattern, it's a representation and an interactive set of rules, that makes people perceive that there's indeed an element.

There were (and are) performance reason too, specially for the complexity of the visualizations we at Moebio build.

Hi Santiago! Thanks you are doing this.

Can you say a bit more about Lichen, and why you decided to develop this. If I understand correctly (simply by looking at the teaser images) it seems to be a combination of graphical and text based programming environment for datavis. Is that the main USP? How does it compare to other datavis constructing environments, such as Quadrigram (graphical programming, from your former Bestiario), or Tableau Public (graphical interface), or d3 (text based programming)?

[yesnewyarseve](#)

Lichen is a modular platform (any one will be able to add modules), that will be released soon, and that plays very well with the Moebio Framework (although is not tied to it). It has some similarities to the tools you mention, but the main difference is that it will be framework agnostics: anyone will be able to add modules that could be built with any framework or web technology.

How many people work in a typical team on a product? Is it one designer with one data scientist with many developers? Do people switch between roles?

[frostickle](#)

In Moebio Labs we develop project in iterations that last ~1 month each, and that normally (specially the first iteration) contain a data scientist half-time, data visualization developer full-time, advanced data visualization developer directing the project half-time.

Depending on the project we could need a database person, an UX person or more time from the data scientist.

Which is your favorite buzzword: Big Data, Cloud, or Innovate?

[nxwtypx](#)

Innovative, for sure. For as a concept it goes way beyond our idiosyncratic tech/start-up era. Innovation is the study of relations among ideas that have the power of producing changes in human life; how can that not be interesting?, regardless if it also happens to be (ab)used as a buzzword. Innovation, creativity, imagination and curiosity are resilient words, they continue being meaningful in spite of their biased uses.

What do you think about the usability of data visualizations on touchscreen devices?

[gkatlauskas](#)

short answer: it's a pain in the arse. Unfortunately for me it's more a problem than an interesting feature to research. For the time being, and except is a client's project requires it, I prefer to work for non-reactive screens.

Now you have the 3D/force pressure sensors coming. At some point one expects that these interactions will become standard.

One interesting thing about this pressure datum, is that in a way it mirrors something that often happens in datavis, in which you have elements that visually weighed (by position, size, opacity, etc...). Now people could also touch elements and weigh them (maybe conveying the idea of how much I'm interested in such element).

Suppose somebody wants to work in this field and is starting with no math except old fashioned engineering/physics stuff, and very little CS or Programming. What are some subjects to learn? Please be as specific as you like.

Also, any subjects people *think* would be important that are not?

[terrifiedbyvajajays](#)

A good way to start is by understanding what information is, and why is so important. For that purpose I recommend this book: https://www.goodreads.com/book/show/8701960-the-information?from_search=true&search_version=service

Then, a further step would get you closer to data and statistics. These books can help: <https://www.goodreads.com/book/show/13588394-the-signal-and-the-noise> and <https://www.goodreads.com/book/show/13707560-naked-statistics>

At this point learning to code is not only a tool for datavis, but also a means to learn it. It's by playing with data that you learn to build images out of datapoints (at least is easier that way). Nowadays there are so many good courses online. And yes, you'll have to learn some math; but if you do it by building stuff, you'll enjoy it. A lot.

Then, to get a better understanding of how data can provide value, this book is a really good one: <https://www.goodreads.com/book/show/17912916-data-science-for-business>.

And in parallel with all these you can read this very solid book about datavis: https://www.goodreads.com/book/show/21927046-information-visualization?from_search=true&search_version=service

if you had unlimited time what would you like to build next?

[jcukier](#)

having unlimited time is a good idea to build something accumulative. Let's call it Infinite Castle. But it would be also something modular, so each Nth element you add to the castle, BOOM, it opens new N-1 pair combination possibilities, and 2^N combinations containing the new element.

Mmm no, I'll learn to build sailing boats, then learn to sail, then travel each port in the world.

Ok no, I will read Infinite Jest.

Besides jokes, we're kind of building that modular system, it's called Lichen and it integrates very well with the Moebio Framework!

What about the performance of predictive modeling into the data visualizations? Are there any possibilities of integration between the Moebio framework and predictive models built on Python/R/etc?

[gkatlauskas](#)

We currently integrate those technologies with the Moebio framework. For that purpose we use a modular platform called Lichen and that will be also released soon.

As someone who is planning to pursue data science as a graduate degree (currently an actuarial science undergraduate), I have a question regarding what employers are generally looking for in the skills of data scientists. If you had the pick the top 3 skills that you want your employees to have, what would they be?

[Xexxar](#)

We are hiring now, so I can give you a fresh answer. It totally depends on the company, specially the moment of the company. In our case, we are looking for developers starting to do data science, or mathematicians/statisticians that are learning to code. A reason to do this is that for the time being we are hiring more in Brasil, so we don't expect to find lots of already formed and experienced data scientists. But it's also because is part of our strategy to build a team that learn. So the first skill is learning... which is more a motivation than a skill. We seek for people craving to learn, curious and eclectic. We also value the capacity to connect math, code an data with real life, being able to understand client's context and strategy.

So we would not be interested, for instance, in a very strong data scientist with poor communication skills and that only wants to work with mathematical and code problems, while displaying little interest for the reality behind the data. And that's something you can detect on an interview.

Your visualizations are stunning, but have you run accross an specific design idea, that though beautiful, obscured the relationships in the data, or even worse it encouraged certain misreadings?

What are the heuristics you use to balance these two constraints beauty vs. clarity?

[SantiagoPeirce](#)

Never! if the geometric/visual approach does not help, doesn't give any insight about the information

used to build it, I don't even reckon it as beautiful! For me beauty in datavis is a cognitive asset. No understanding, no beauty. I'm not impressed by simetric and colorful shapes. I, for instance, never liked any Pi visualization (I'm referring to the ones that visualize Pi figures) for they tell nothing about Pi! (or maybe that's the story, that there's no pattern on Pi and that's what is visualized). On the other hand, as an example, I love the Mandelbrot Set, because the astonishing image is in perfect and deep connection with the concept, the equation (the construction process) and the marvelous properties it has... all these are one, but our consciousness can only grasp these qualities separately at first, and then join them together again... beauty lies in that encounter.

I just spent about 45 minutes playing with visual crawler.. All I can say is wow. It took a bit of time to figure out what the hell I was looking at but after learning about web crawlers it started to make more sense.. That is a cool looking tool for sure, congrats to your team for making that work. Do you know any good business use for this or is it just a sweet visualization?

[insperaeshun](#)

There are professional, much more robust crawlers out there. So this experiment is just that. But notice that the structural part of the project, that deals with "nested networks", and even the representation side, can be useful in many other contexts. And indeed some of tools developed to build this experiment were incorporated into the framework, and are already used in commercial projects.

Existe en el horizonte de Moebio Labs interés por ir más allá de las compañías y empresas, y plantear posibles herramientas de visualización y análisis dirigidos a usuarios particulares o instituciones públicas?. Moebio Framework podría ser útil en otros contextos además del empresarial?. Gracias.

[Javingka](#)

(Questions is about Moebio Labs interest of opening our tool to a not corporative context)

Sí claro, el framework se puede usar para cualquier propósito, y si miras la página de moebio verás que hay mucha experimentación: <http://moebio.com>

Yes, the framework can be used in more creative contexts, the moebio website features several projects that are quite experimental: <http://moebio.com>

How do you see people making the connection between what you do and their own work? There's a huge disconnect between the existing data visualization capabilities and everyday business practices. I'm trying to bridge that gap, but most people just don't understand how powerful this technology is.

[5steelBI](#)

please see this answer

https://www.reddit.com/r/dataisbeautiful/comments/3l1pmw/hi_everyone_im_santiago_ortiz_i_lead_moebio_labs/cv2ox4i