🔒 tradestatistics - 2020 Shiny Contest Submission

shiny-contest-2020

---

pachamaltese Early Adopter Group for Evaluating RStudio   **2020-02-21**

tradestatistics

Authors: Mauricio 'Pachá Vargas Sepúlveda
Working with Shiny for 1+ years

Abstract: Open Trade Statistics is a project that includes a public API, a dashboard, and an R package for data retrieval. In particular, the dashboard was conceived as a graphical tool for people from economics and humanities that, most of the times, are used to Excel and not to using APIs. The dashboard allows users to explore the data visually and then export it to xlsx and other formats.

Full Description: *Adapted from*
https://ropensci.org/blog/2019/05/09/tradestatistics/

Open Trade Statistics (OTS) was created with the intention to lower the barrier to working with international economic trade data. It includes a public API, a dashboard, and an R package for data retrieval.

The project started when I was affected by the fact that many Latin American Universities have limited or no access to the United Nations Commodity Trade Statistics Database (UN COMTRADE).

There are alternatives to COMTRADE, for example the Base Pour L'Analyse du Commerce International (BACI) constitutes an improvement over COMTRADE as it is

Skip to main content he raw data and a method that reconciles the declarations of the

exporter and the importer. The main problem with BACI is that you need UN COMTRADE institutional access to download their datasets.

After contacting UN COMTRADE, and suggesting to them my idea of doing something similar to BACI but available for anyone and keeping commercial purposes out of the scope of the project, I got an authorization to share curated versions of their datasets.

Different projects such as The Atlas of Economic complexity and The Obervatory of Economic complexity use UN COMTRADE data and focus on data visualization to answer questions like:

- `What` did Germany export in `2016`?

- `Who` imported Electronics in `1980`?

- `Who` exported Refined Copper in `1990`?

- `Where` did Chile export Wine `to in` `2016`?

Unlike existing visualization projects, I wanted to focus on data retrieval and reproducibility, and the starting point was to study the existing trade data APIs to create something more flexible and easier to use than those tools.

I started organizing code I wrote during the last four years at **https://github.com/tradestatistics/** . There was code there that I haven't touched in more than two years, and I wrote almost no comments indicating what the parts of the code actually do, so it was not understandable for others.

My data cleaning process was not reproducible, and it was tragic to discover! I decided to start using RStudio Server to test the code line by line, in a fresh environment, and then dividing the code into smaller pieces and commenting what the different sections actually do.

Once I had reproducible results I took a snapshot of my packages by using packrat. To ensure reproducibility over time, I decided to build R from source, isolated from the system package manager and therefore avoiding accidental updates that might break the code.

Is it worth mentioning that I'm using DigitalOcean virtual machines to store the datasets and run all the services required to run an API. Under their Open Source

Skip to main content erver cost is subsidized. The base for the project is Ubuntu, the database of choice is PostgreSQL, and R constitutes 95% of the project.

Thanks to Maëlle Salmon, Amanda Dobbyn, Jorge Cimentada, Emily Riederer, Mark Padgham the overall result can be said to be top quality!

After a long reviewing process (more than six month considering initial submission!), what started as an individual process mutated into something that I consider a collective result. Thanks to the amazing team behind rOpenSci, to their constructive feedback, exhaustive software reviewing and the confidence to propose ideas that that I had never gotten, what you have now is not just a solid R package.

The hours spent as a part of the reviewing process translated into changes to the database and the API. According to the reviewers comments, there are limited opportunities to implement server-side changes and then updating the R code. With the inclusion of different API parameters that I initially didn't consider, the current API/package state provides an efficient solution way better than post-filtering. You'll always extract exactly the data you require and no more than that.

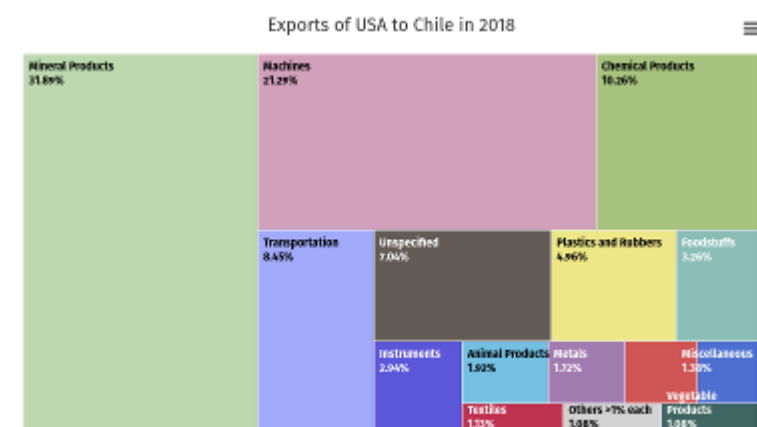You can check a long form presentation I gave at Latin R 2019 at **https://pacha.dev/latinr/tradestatistics/** .

Category: Education
Keywords: open data, international trade, highcharter, shinydashboard, sql, api
Shiny app: **https://pachamaltese.shinyapps.io/tradestatistics**
Repo: **https://github.com/tradestatistics/visualization-with-shiny/tree/master/dashboard**
RStudio Cloud: **https://rstudio.cloud/project/968512**

Thumbnail:

Full image:

Skip to main content

CLOSED ON 2020-03-02

THIS TOPIC HAS BEEN CLOSED.

IF YOU HAVE A QUERY RELATED TO IT OR ONE OF THE REPLIES, START A NEW TOPIC AND REFER BACK WITH A LINK.

IF YOU BELIEVE A REPLY REALLY BELONGS IN THIS THREAD, PLEASE SEND IT TO @ECONOMICURTIS AS A DIRECT MESSAGE.

OPENED ON 2020-03-10

CLOSED ON 2020-05-02

This topic was automatically closed after 52 days. New replies are no longer allowed.