

Doing data science with F#

Tomas Petricek

tomas@tomasp.net | @tomaspetricek

PhD Student at Cambridge & Coordinator of <http://fsharp.org>

software stacks

trainings teaching F# user groups snippets

mac and linux community books and tutorials

F# Software Foundation

consulting open-source MonoDevelop

<http://www.fsharp.org>

contributions research support

cross-platform mailing lists

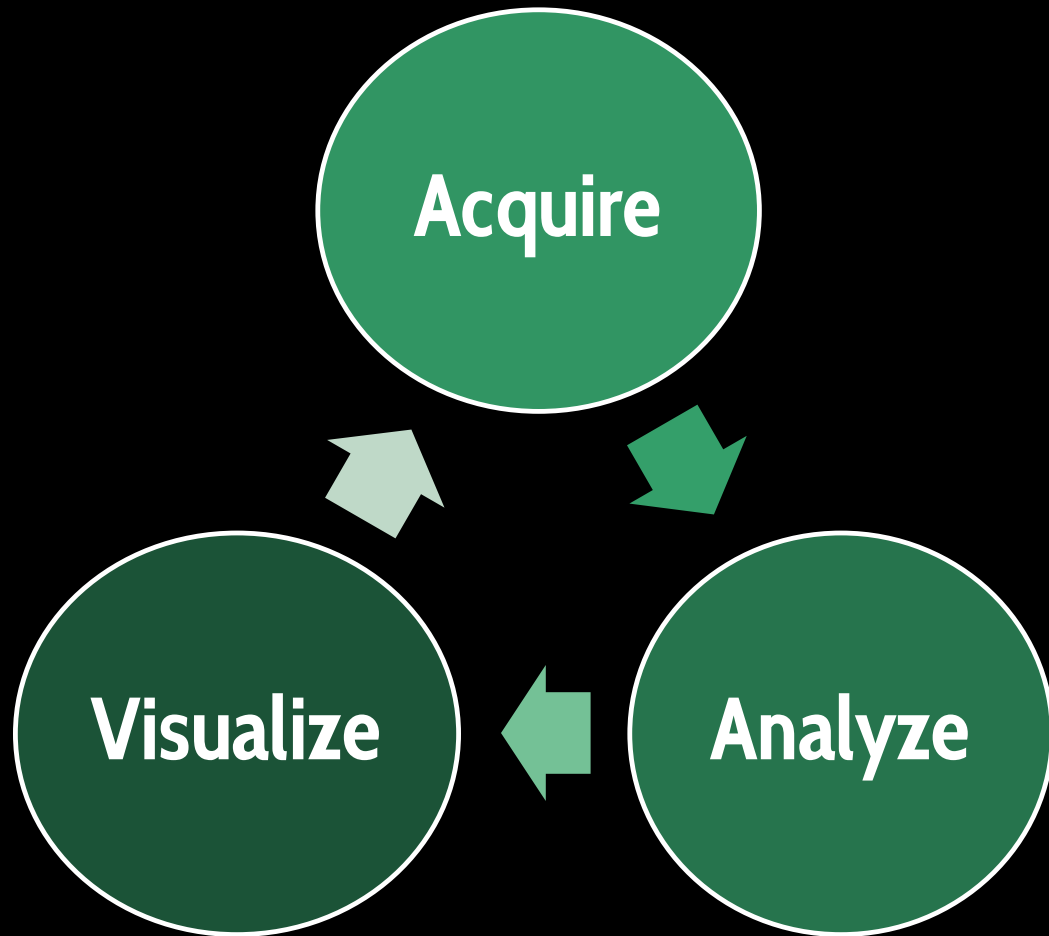
Community matters!



A satellite image of Earth, centered on North and Central America. The image shows the continents of North America, Central America, and parts of South America, surrounded by the Atlantic and Pacific Oceans. Swirling white clouds are visible over the oceans. The text "All the Data of the World" is overlaid in white, bold, sans-serif font across the center of the image.

All the Data of the World

data acquisition
statistics data cleaning machine learning
data transformation visualization type providers
F# Data Science Working Group
kaggle vega grammar data sources presentation
R provider **www.fslab.org** visualization
time-series data aggregation





Demo: Analyzing Titanic survivors



Deedle data frame

Data exploration

Indexing and aggregation



F# Charting library

Simple & composable

Interactive style

www.fslab.org



Demo: Understanding the world



F# Data type providers

First-class data

CSV, REST, WorldBank...

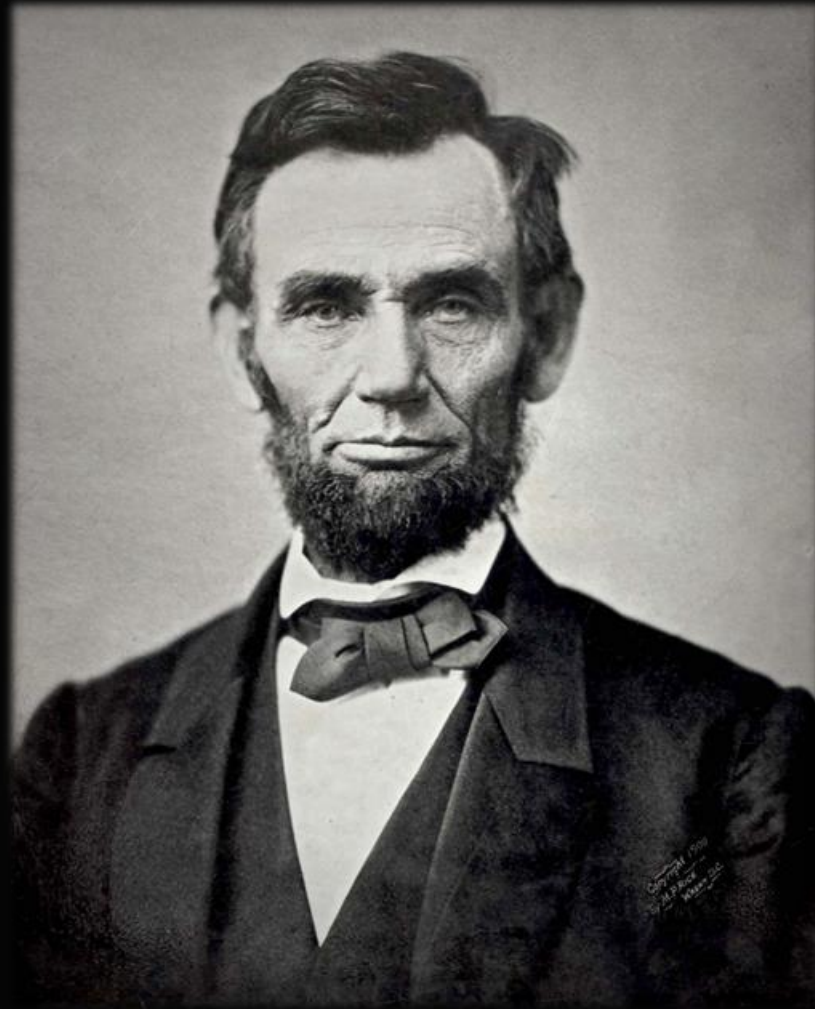


R Type provider

Statistics & visualization

5000 tested packages

www.fslab.org



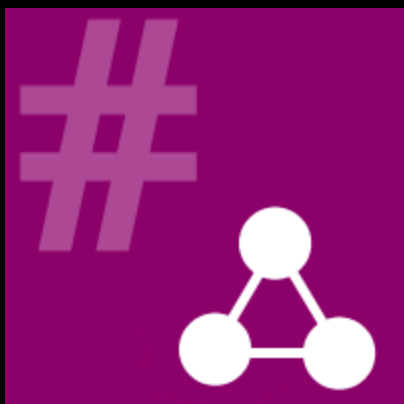
Demo: US debt over the last century



Deedle data frame

Time-series alignment

Data transformations



Vega visualization

F# wrapper for Vega

Pre-alpha version

www.fslab.org

F# for Data Science

acquire, analyze, visualize

interactive experience

safety and efficiency of .net

ready for production

@tomaspetricek

Going forward

Use #fsharp for fun & profit

Join local user groups

Help us build data science tools

fsharp.org | fslab.org | tomasp.net

@tomaspetricek