

TwoRavens and Software Development on Academic Projects

Vito D'Orazio

Assistant Professor
University of Texas at Dallas

December 11, 2017

TwoRavens Project

- Privacy Tools, Dataverse, Zelig
 - ▶ 2013–2015 saw initial TwoRavens development
 - ▶ I was a postdoc at IQSS
 - ▶ Honaker was a senior research scientist at IQSS

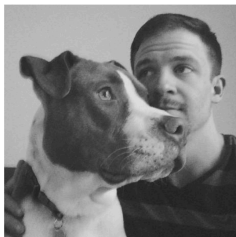
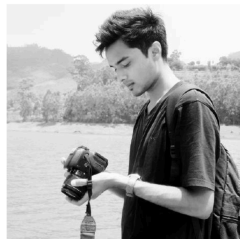
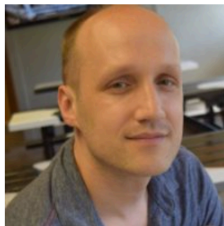
TwoRavens Project

- Privacy Tools, Dataverse, Zelig
 - ▶ 2013–2015 saw initial TwoRavens development
 - ▶ I was a postdoc at IQSS
 - ▶ Honaker was a senior research scientist at IQSS
- Minor UTD funding (2015-2016)
- Modernizing Political Event Data
 - ▶ NSF RIDIR grant, 2016–2018

TwoRavens Project

- Privacy Tools, Dataverse, Zelig
 - ▶ 2013–2015 saw initial TwoRavens development
 - ▶ I was a postdoc at IQSS
 - ▶ Honaker was a senior research scientist at IQSS
- Minor UTD funding (2015-2016)
- Modernizing Political Event Data
 - ▶ NSF RIDIR grant, 2016–2018
- Data Driven Discovery of Models (D3M)
 - ▶ DARPA grant, 2017–2021

The Team





A tool for statistical analysis:

- Explore and annotate data



A tool for statistical analysis:

- Explore and annotate data
- Construct and estimate statistical models



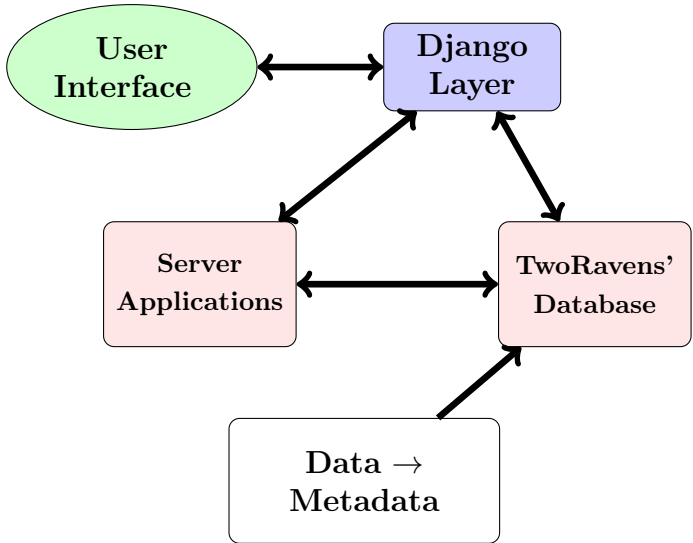
A tool for statistical analysis:

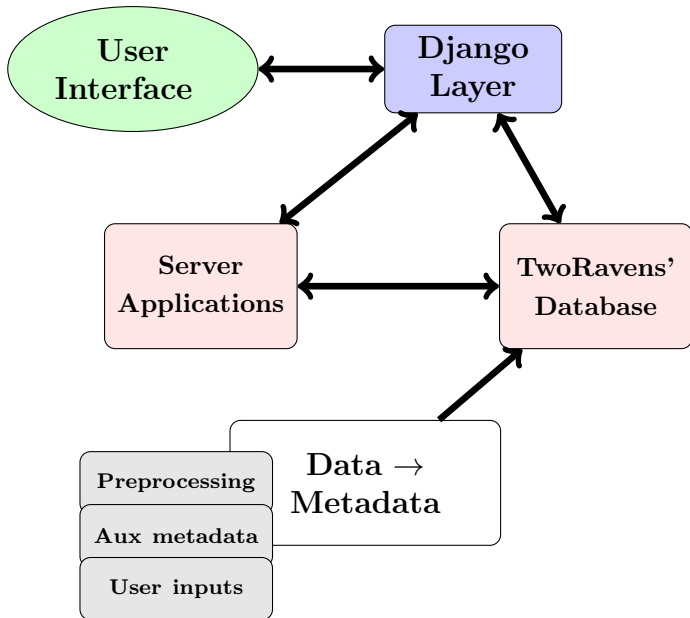
- Explore and annotate data
- Construct and estimate statistical models
- Compare and curate empirical results

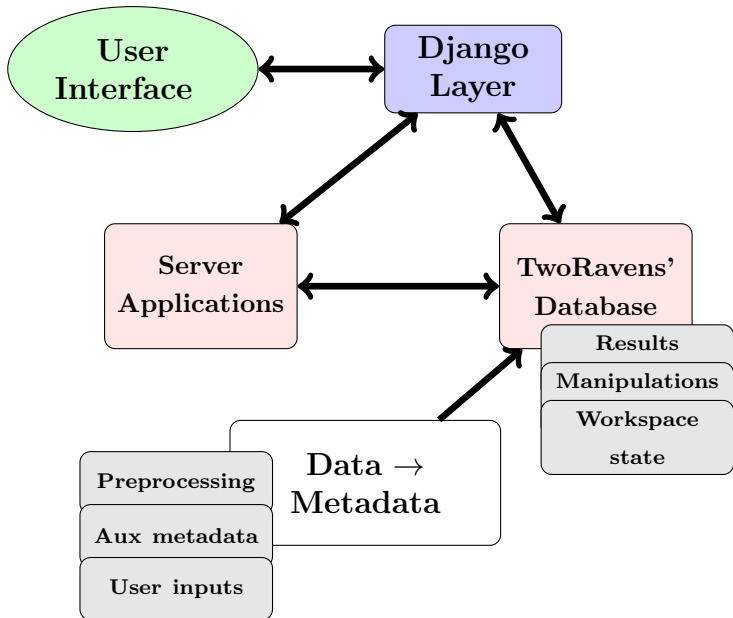


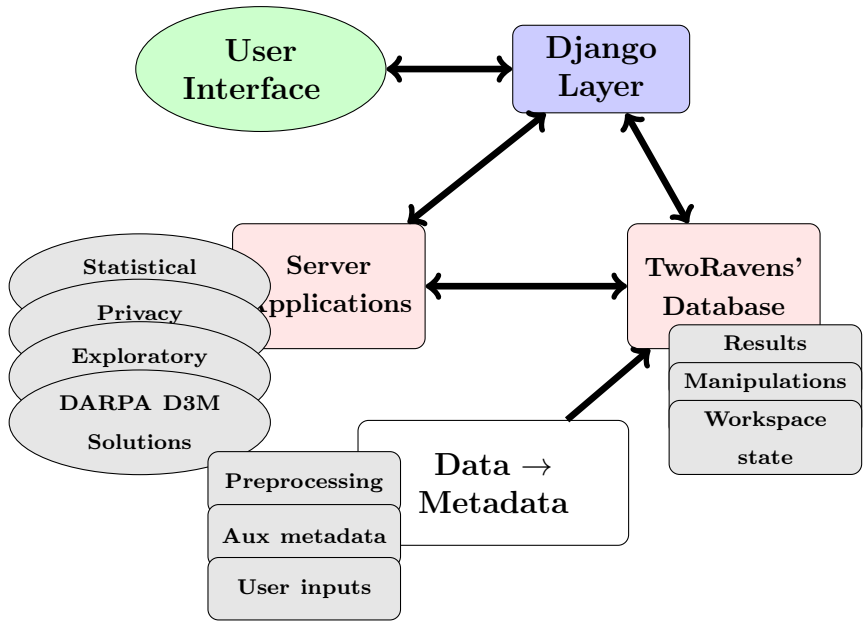
Our goals:

- Connect people with data—a tool for education as much as for research
- Accumulate knowledge and learn patterns from users and the broader research community
- Improve the quality of results in the literature and disseminate that knowledge

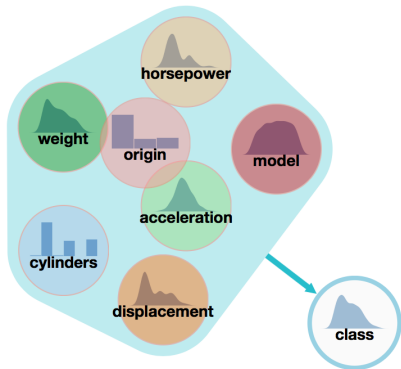
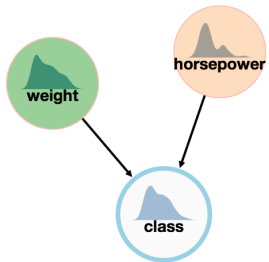








- Users specify relationships among variables using the language of directed graphs



TwoRavens Demo in Dataverse domain:

[https://rserve.dataverse.harvard.edu/
dataexplore/gui.html?dfId=2723294&](https://rserve.dataverse.harvard.edu/dataexplore/gui.html?dfId=2723294&)

TwoRavens Demo in D3M domain:

<http://2ra.vn/demos/d3mintegrationdemo.mp4>



Our goals:

- Connect people with data—a tool for education as much as for research
- Accumulate knowledge and learn from users and the broader research community
- Improve the quality of results in the literature and disseminate that knowledge

Recent Progress

- Django Web framework
 - ▶ Routing requests, configurations, user logins
- Mithril frontend framework
 - ▶ Manages the rendering of frontend components
- TwoRavens database
 - ▶ Save, restore, share workspace
 - ▶ Store results
- Explore mode
 - ▶ A mode focused on descriptive data analysis
- D3M domain
 - ▶ New grammar for less-restricted problem solving
 - ▶ Integration with other D3M teams
- API documentation

Future Goals

Education—Teaching with TwoRavens

- Workspaces are shareable
 - ▶ Single data file for instructor and student
 - ▶ Link directly on syllabus
- Only requires an Internet connection
 - ▶ Data analysis into any classroom
 - ▶ Community colleges, high schools
- Software is both free and easy-to-use
 - ▶ R and Python are not easy to use
 - ▶ Matlab, Stata, SAS, SPSS are not free

Future Goals

Education—Teaching with TwoRavens

- Workspaces are shareable
 - ▶ Single data file for instructor and student
 - ▶ Link directly on syllabus
- Only requires an Internet connection
 - ▶ Data analysis into any classroom
 - ▶ Community colleges, high schools
- Software is both free and easy-to-use
 - ▶ R and Python are not easy to use
 - ▶ Matlab, Stata, SAS, SPSS are not free
- **Pebbles are fun**

Future Goals

Replication

- Simple idea: TwoRavens-readable replication files
- Harvest and launch workspace containing published results
- Trajectory of a research problem over time
 - ▶ Google Ngram Viewer for data
- Automated recommendations for future researchers working on the same problem

Future Goals

Replication

- Simple idea: TwoRavens-readable replication files
- Harvest and launch workspace containing published results
- Trajectory of a research problem over time
 - ▶ Google Ngram Viewer for data
- Automated recommendations for future researchers working on the same problem
- Sensitivity analysis on a scale that would satisfy Edward Leamer

Lessons Learned

Interdisciplinary projects and software:

- Language is tough, but incentives are tougher
 - ▶ Graduate students are not aspiring developers
 - ▶ Large numbers of students *are* aspiring developers

Lessons Learned

Interdisciplinary projects and software:

- Language is tough, but incentives are tougher
 - ▶ Graduate students are not aspiring developers
 - ▶ Large numbers of students *are* aspiring developers
- Design documents at the proposal phase
 - ▶ Neither faculty nor students know how to do this
 - ▶ Where I see the greatest need for professional developers or institutional assistance



Website: 2ra.vn

Github: www.github.com/tworavens

Vito D'Orazio
dorazio@utdallas.edu

James Honaker
james@hona.kr