

TwoRavens
A Tool for Statistical Analysis in the Classroom

Vito D'Orazio and James Honaker

Assistant Professor of Political Science
University of Texas at Dallas

July 12, 2016



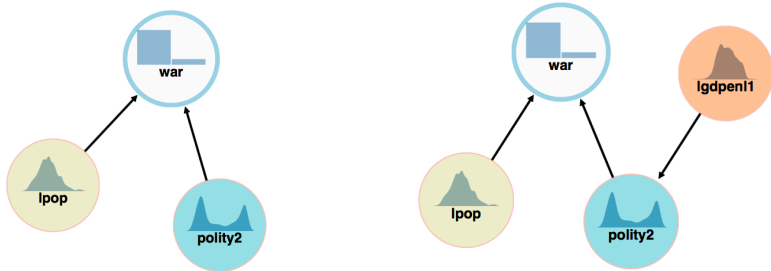


What is TwoRavens?

- Tool for estimating statistical models
- Users specify relationships among variables using the language of directed graphs

What is TwoRavens?

- Tool for estimating statistical models
- Users specify relationships among variables using the language of directed graphs

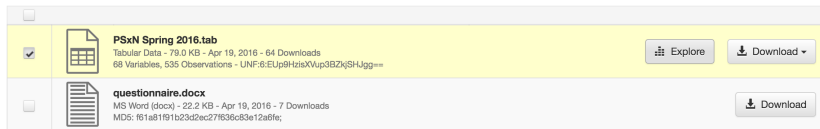


What is TwoRavens?

- Web-based and runs entirely in the browser
- Metadata sent client-side, data are never local
 - ▶ Useful for working with sensitive data and large datasets
- Backend processing using Zelig and R

What is TwoRavens?

- Web-based and runs entirely in the browser
- Metadata sent client-side, data are never local
 - ▶ Useful for working with sensitive data and large datasets
- Backend processing using Zelig and R
- Integrates with Dataverse

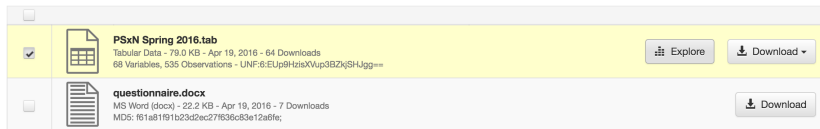


The screenshot shows a file management interface with two entries:

- PSxN Spring 2016.tab**: Tabular Data - 79.0 KB - Apr 19, 2016 - 64 Downloads
68 Variables, 535 Observations - UNF:6:EU9HziXVup3BZkqSHJgg==
Buttons: Explore, Download ▾
- questionnaire.docx**: MS Word (docx) - 22.2 KB - Apr 19, 2016 - 7 Downloads
MD5: f61a81f91b23d2ec27f636c83e12a8fe;
Button: Download

What is TwoRavens?

- Web-based and runs entirely in the browser
- Metadata sent client-side, data are never local
 - ▶ Useful for working with sensitive data and large datasets
- Backend processing using Zelig and R
- Integrates with Dataverse



The screenshot shows a file management interface with two entries:

- PSxN Spring 2016.tab**: Tabular Data - 79.0 KB - Apr 19, 2016 - 64 Downloads. 68 Variables, 535 Observations - UNF:6:EU9pHzisXVup3BZkqSHJgg==. Includes 'Explore' and 'Download' buttons.
- questionnaire.docx**: MS Word (docx) - 22.2 KB - Apr 19, 2016 - 7 Downloads. MD5: f61a81f91b23d2ec27f636c83e12a8fe;. Includes a 'Download' button.

Demo

Teaching with TwoRavens and Dataverse

Teaching with TwoRavens and Dataverse

- As instructors, we often teach with data
- Less often do we ask students to work with data

Teaching with TwoRavens and Dataverse

- As instructors, we often teach with data
- Less often do we ask students to work with data
- Why?
 - ▶ Requires a computer lab with installed statistical software
 - ▶ Requires knowledge of statistical software
 - ▶ Requires individual instruction within limited classroom time

Teaching with TwoRavens and Dataverse

- As instructors, we often teach with data
- Less often do we ask students to work with data
- Why?
 - ▶ Requires a computer lab with installed statistical software
 - ▶ Requires knowledge of statistical software
 - ▶ Requires individual instruction within limited classroom time

TwoRavens and Dataverse make teaching with data simple.

Why Teach with TwoRavens?

Valuable features of TwoRavens for instruction:

1. Gesture-driven and graphical
2. Web application, runs in browser
3. Uses data in remote repositories
4. Customizable metadata

TwoRavens' Features for Instruction

Gesture-driven and graphical

TwoRavens' Features for Instruction

Gesture-driven and graphical

- An intuitive way to construct a model

TwoRavens' Features for Instruction

Gesture-driven and graphical

- An intuitive way to construct a model
- With minimal training:
 - ▶ Explore data
 - ▶ Construct statistical models
 - ▶ Interpret results

TwoRavens' Features for Instruction

Gesture-driven and graphical

- An intuitive way to construct a model
- With minimal training:
 - ▶ Explore data
 - ▶ Construct statistical models
 - ▶ Interpret results
- Instructors focus:
 - ▶ On the theory, the model, and the analysis
 - ▶ *Not* on statistical software

TwoRavens' Features for Instruction

Gesture-driven and graphical

- An intuitive way to construct a model
- With minimal training:
 - ▶ Explore data
 - ▶ Construct statistical models
 - ▶ Interpret results
- Instructors focus:
 - ▶ On the theory, the model, and the analysis
 - ▶ *Not* on statistical software
- Easily used on tablets and phones

TwoRavens' Features for Instruction

Web application, runs in browser

TwoRavens' Features for Instruction

Web application, runs in browser

- Instructor in front of the room may launch TwoRavens while her students follow along

TwoRavens' Features for Instruction

Web application, runs in browser

- Instructor in front of the room may launch TwoRavens while her students follow along
- Requirements:
 - ▶ An internet connection
 - ▶ A device with a Web browser

TwoRavens' Features for Instruction

Web application, runs in browser

- Instructor in front of the room may launch TwoRavens while her students follow along
- Requirements:
 - ▶ An internet connection
 - ▶ A device with a Web browser
- No need to purchase proprietary software
 - ▶ Free to use and our software is open-source
- No need to install anything
- No need to be in a computer lab, or even a classroom

TwoRavens' Features for Instruction

Remote data repositories

TwoRavens' Features for Instruction

Remote data repositories

- Everybody uses the same data file
 - ▶ Not corruptible, never local

TwoRavens' Features for Instruction

Remote data repositories

- Everybody uses the same data file
 - ▶ Not corruptible, never local
- Currently integrated with Harvard's Dataverse
 - ▶ May use TwoRavens on 30,996 datasets
 - ▶ Up from 28,944 in February!
 - ▶ Instructor may create a "Teaching Course" dataverse
 - ▶ Datasets are versioned, permanently archived, provided with a DOI, file conversion capabilities (Stata to SPSS to R to Excel)

TwoRavens' Features for Instruction

Remote data repositories

- Everybody uses the same data file
 - ▶ Not corruptible, never local
- Currently integrated with Harvard's Dataverse
 - ▶ May use TwoRavens on 30,996 datasets
 - ▶ Up from 28,944 in February!
 - ▶ Instructor may create a "Teaching Course" dataverse
 - ▶ Datasets are versioned, permanently archived, provided with a DOI, file conversion capabilities (Stata to SPSS to R to Excel)
- Link to your dataverse in your syllabus
 - ▶ One time, one place, and you're done for the course
 - ▶ No migrating files from course to course on Blackboard
 - ▶ No fighting with file versions or compatibility

TwoRavens' Features for Instruction

Customizable metadata (in progress)

- Data owners may refine metadata through our interface
 - ▶ Variable descriptions
 - ▶ Variable types (e.g., *time*, *nominal*)
 - ▶ Pre-drawn statistical models
- Students see this customized metadata

Teaching with TwoRavens

2ra.vn

- Replicating “Ethnicity, Insurgency, and Civil War” by James Fearon and David Laitin
- Exploring and instructing with the Quality of Government dataset

“Teaching Course” Dataverse

- *Political Science by Numbers* by Yamil Velez at Wesleyan University

Future Directions

- We want to:
 - ▶ Make this a way to *show students data*
 - ▶ Allow them to *discover* data and *interact* with data
 - ▶ Not just read about data or listen to it talked about

Future Directions

- We want to:
 - ▶ Make this a way to *show students data*
 - ▶ Allow them to *discover* data and *interact* with data
 - ▶ Not just read about data or listen to it talked about

2ra.vn

Vito D'Orazio
dorazio@utdallas.edu

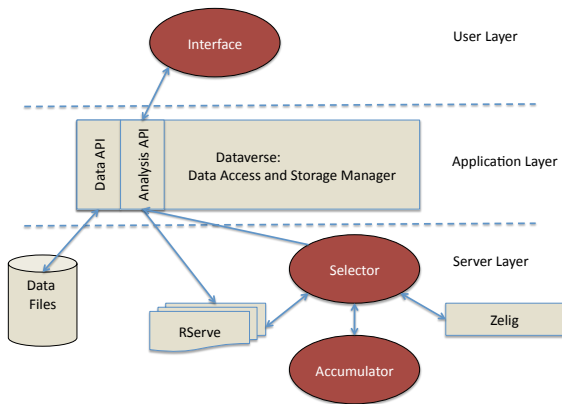
James Honaker
james@hona.kr

The TwoRavens Interface

The screenshot displays the TwoRavens software interface for the 'fearonLaitinData' dataset. The interface is divided into several sections:

- Data Selection:** A list of variables is shown, with 'war' selected. The list includes: ccode, country, cname, cmark, year, wars, war, war1, onset, ethonset, darest, aim, casename, ended, ethwar, and waryrs.
- Model Selection:** A list of models is shown, including: ls, logit, probit, poisson, normal, gamma, negbinom, exp, lognorm, tobit, quantile, logitgee, probitgee, zgammagee, znormalgee, and poissongee.
- Causal Diagram:** A directed acyclic graph (DAG) showing relationships between variables: 'lqdpwin1' (orange circle) points to 'polity2' (blue circle); 'lpop' (green circle) points to 'war' (blue circle); 'polity2' points to 'war'; and 'mtnest' (pink circle) points to 'war'.
- Legend:** A legend indicates that a blue circle represents a 'Dep Var'.
- Buttons:** At the top right, there are buttons for 'Variable transformation', 'C' (refresh), and 'Estimate'.

The TwoRavens Architecture



<https://github.com/IQSS/tworavens>

TwoRavens Community

- The Dataverse Project
 - ▶ dataverse.org
- Privacy Tools for Sharing Research Data
 - ▶ privacytools.seas.harvard.edu
- Modernizing Political Event Data for Big Data Social Science Research
 - ▶ <http://goo.gl/12NLwm>
- Zelig
 - ▶ zeligproject.org
 - ▶ Zelig drives our back-end statistical analyses

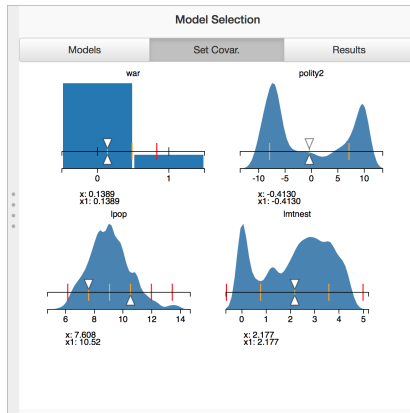
Integration with Zelig

- R library that facilitates *estimation* and *interpretation* of regression models
- Estimation: common call structure for all included models
- Interpretation: quantities of interest are simulated and visualized
- Also facilitates a quantitative workflow that includes *multiple imputation* for missing data, *matching* for causal reasoning, and *counterfactual analysis* bc who doesn't love to ask what if?

Integration with Zelig

Model Selection

Models	Set Covar.	Results
ls		
logit		
probit		
Model Description		
Negative Binomial Regression for Event Count Dependent Variables		
negbinom		
exp		
lognorm		
tobit		
quantile		
logitgee		
probitgee		
zgammagee		
znormalgee		
poissongee		



The TwoRavens Interface

