EVANS SCHOOL OF PUBLIC POLICY & GOVERNANCE

UNIVERSITY of WASHINGTON

Evans School Policy Analysis and Research (EPAR)

Making Exploratory Data Analysis More Accessible Through Interactive Visualization

THE CHALLENGE

Several factors constrain broader use of increasingly available agricultural and household survey data:

- Complex data structures require a substantial time investment to understand
- Important decisions surrounding data cleaning, trimming, and model selection are persistent challenges
- Software can be expensive, and requires both technical and statistical competence
- Current tools for data exploration are limited in scope

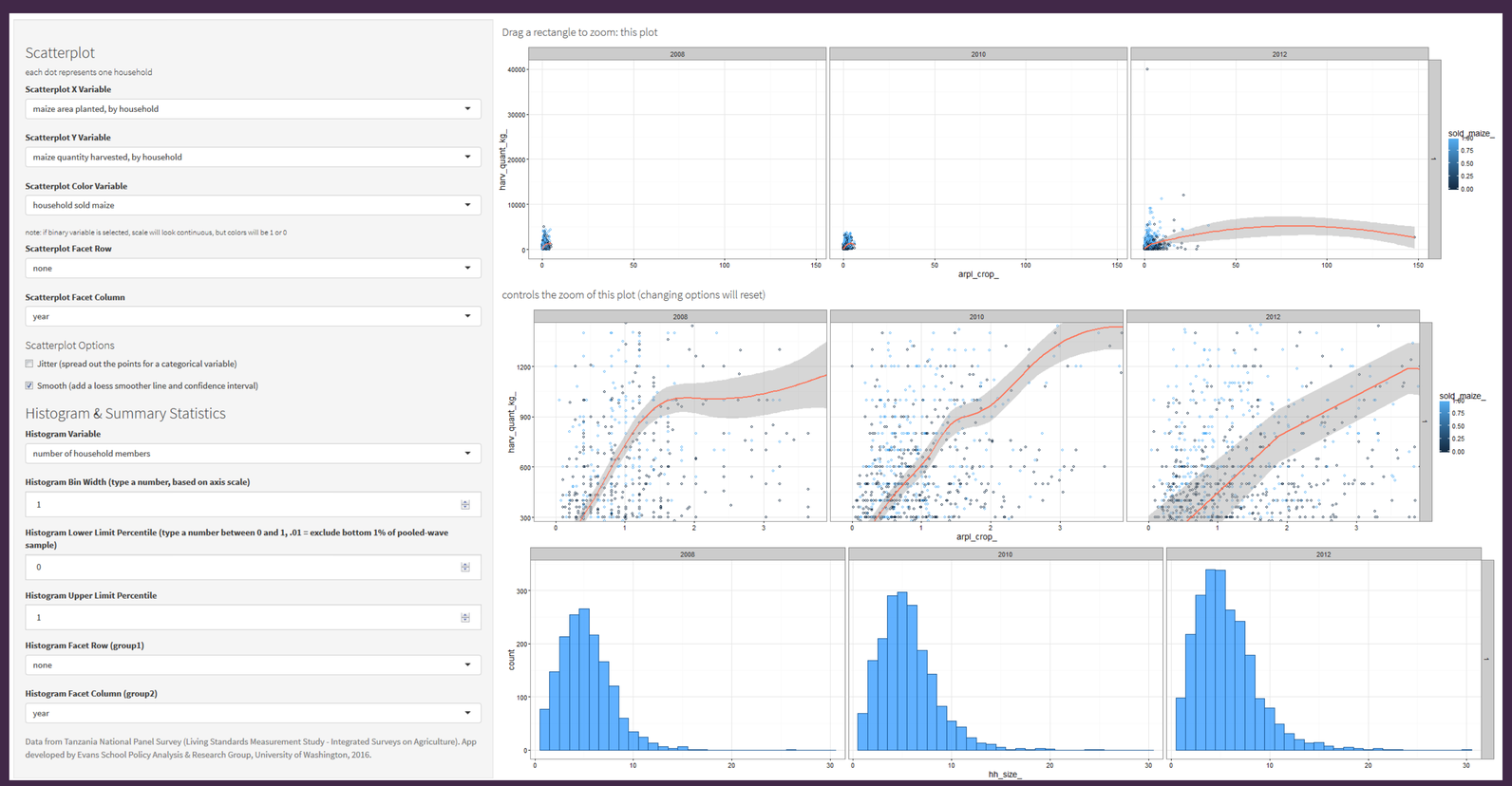
These constraints affect even researchers and practitioners with specialized statistical and data analysis skills, the primary users of these data sources.

OUR SOLUTION

A point-and-click web interface allows for accessible, interactive data exploration. Users can:

- Examine relationships between variables using scatterplots
- "Facet" charts by categorical variables to compare across groups
- Zoom in on any section of a plot
- Add smoother curves and confidence intervals
- Study distributions through histograms with user-selected bin sizes and percentile-based trimming
- Display summary statistics by grouping variable

The app is built using the Shiny interactivity package for R, an open-source software environment for statistical computing and graphics.



OUR DATA

The World Bank's Living Standards Measurement Study – Integrated Surveys on Agriculture (LSMS-ISA) is a publicly available household panel survey dataset for seven countries in Sub-Saharan Africa. The survey includes linked plot, household, and community level modules that provide information on crops, livestock, farming practices, and socio-demographics.

Our app displays cleaned data from three panels of the Tanzania National Panel Survey, one of the LSMS-ISA datasets.

With simple modifications and basic data preparation, the app can visualize any dataset.

ADVANTAGES OF INTERACTIVE VISUALIZATION

Our tool goes beyond summary tables and pre-programmed chart extraction to allow any user to interactively explore not just measures of central tendency, but complete variable distributions and relationships within a complex dataset.

Advantages of this app include:

- App runs in web browser – no software, spreadsheets, or coding knowledge required
- Users can interact with raw survey data to understand outliers and skew, and make choices about cleaning and trimming
- App generates customized charts and summary tables for any combination of variables in the dataset: not limited to pre-programmed choices

Further, our code is openly available for further development.

- Easy to adapt to any dataset with a similar basic structure
- Intermediate programmers can modify scripts to do any statistical operation possible in R software
- The app can be run locally on a secure computer for data that are not yet public

WHO WE ARE

The University of Washington's Evans School Policy Analysis and Research Group (EPAR) uses an innovative student-faculty team model to provide rigorous, applied research and analysis for international development stakeholders, including both researchers and practitioners. Our work helps to inform investment decisions and development strategies aimed at alleviating some of the world's most complex, collective, and pressing economic, agricultural, public health, and environmental challenges. Established in 2008, EPAR has prepared more than 250 technical reports covering topics in agriculture, aid and development policy, financial services, adoption, gender, and measurement.

Learn more at *http://bit.ly/EPAREvans*