A Novel Gateway for Visual Exploratory Analysis and Sharing of Survey Data

Ilya Zaslavsky, Side Li, Stephan Lamont, David Valentine, Thomas Whitenack

SDSC

http://suave.sdsc.edu

SuAVE

Survey Analysis via Visual Exploration

Infrastructure for Survey Data?

- **Common characteristics:**
  - Systematic data collection
  - Typically with elements of human appraisal
  - Mixture of data types: measurements, human annotations, multi-media, narratives, links to other resources, etc.
  - Typical goal is to explore and organize diverse information, develop a “lay of the land”, quickly formulate and test hypotheses, prepare for further in-depth analysis
  - Transcending granularities and spatio-temporal scales
  - Often done by teams of researchers: collaboration is typically expected
- **But: infrastructure is fragmented**

What you can do with SuAVE:

- **Publish and share** survey datasets:
  - engaging and easy to use; no need to download software
- Examine **general patterns** – and zoom in to **individual cases**, interpret deviant behaviors and outliers
- Build **explanations** of patterns, over mixed data
- Integrated visual, statistical, cartographic analysis
- **Annotate** findings and share annotations
- Help users **follow your analysis** (and reproduce it)
- A platform for **other types of analysis**

Examples

- http://suave.sdsc.edu , under the **Gallery** and **News** sections
  - General Social Survey (NORC) [lots of opinion surveys]
  - British Geological Survey (Macrofossils collection)
  - Fish gallery, Florida Museum of Natural History
  - Coral Reef Survey, Moorea LTER
  - Paintings: Picasso, Van Gogh, Dali,…
  - Sustainable Development Goals Indicators (UN)
  - Library Collections
  - Also:
    - Archaeology
    - Urban Planning
    - Portfolio analysis
    - Health informatics
    - Membership of [virtual] organizations
    - TEACHING research methods!
    - Also: conference posters; stakeholder maps; journals;

Background

- Building on HTML5 projects that attempted to rewrite MS Live Labs Pivot
- Our enhancements:
  - Rendering of large image collections (eg BGS with 8k+ images)
  - Survey-specific variable types
  - Additional analytical views
    - Key challenge is to develop a combination of analyses and visualizations that would reinforce each other
  - Connection with R
  - Survey management and publishing system
  - Capturing and annotating interesting patterns, sharing annotations and data views

Published surveys

- **Questionnaire surveys:** 2016 American National Election Pilot; GSS 2010; EarthCube Members; Health and Sanitation in Tijuana; Belmont Forum Open Data Survey; Science Gateways Institute Survey, GEO Data Management Principles survey, NHANES surveys
- **Art history:** Picasso; Bosch; Rivera; Van Gogh; Dali; Dr. Seuss cartoons and artwork; Balthus; MoMA paintings
- **Humanities** : UCSD student newspaper collection;
- **Geosciences**: British Geological Survey; AMNH rock collection; NOAA EOS catalog; domain catalog management
- **Biodiversity**: Camera trap surveys; Acacia census in Australia; Global Forest Link; Florida MNH
- **Health informatics**: dermatology, ophthalmology collections
- **Organizations**: EC Landscape Mapping; UC-Mexico faculty directory; Conference posters (EC, SI2); San Diego vacant lots survey; journals; Contamination sites in the border area; UCSD doctors; SDG Indicators
- **Archaeology** (Shiqmim survey, Ceramic photomicrographs in Negev)
- **Portfolio analysis** (many: NSF, NIH, European Commission; others)

Surveys in IT

- Belmont Forum open data survey
- GEO data management principles implementation survey
- EarthCube member survey
- Science Gateways Survey

Conclusions

- SuAVE proved to be a useful entry point for researchers from multiple domains
- **Key priorities:**
  - More flexible authoring (with mashups, functions, automatic annotations), with provenance tracking
  - Dynamically generating Suave visualizations from collections
  - Exploration of very large digital collections
  - Integrating image analytics
  - Additional data views
  - Reproducible data analysis pathways
  - Cross-survey connections and meta-analysis
- **Create accounts and experiment with it!**
  - Currently, over 40 external accounts
  - Use it to publish your surveys, to make them accessible by anybody
  - Integrate with other gateway efforts

Acknowledgement

NSF ACI-1443082, “EAGER: Development of a Novel Online Visual Survey Data Analysis Tool and Assessment of its Capabilities to Enhance Learning of Quantitative Research Methods”