

Daisy

Data Made Easy

Grace, Shelby & Jen

Daisy Is A Complete Workflow Interface

Cloud Based
Data



Analysis &
Visualization



Collaboration



Complete Analysis Toolkit
for Domain Experts

Easily Obtain Datasets

OSDC

- Query various databases (Twitter, PDB, Astrodata)
 - ex. Twitter data only extends back 1 month
- Retain data on OSDC
- Find past data and save time



Protein Data Bank

The PDB is a major repository for experimentally determined structures of biological macromolecules such as proteins and nucleic acids. It is updated weekly and the [PDB Holdings Report](#) contains a high level overview of the current contents.

Source: <http://www.rcsb.org/pdb/home/home.do>

Keywords: [biology](#)

Size: 243.4GB

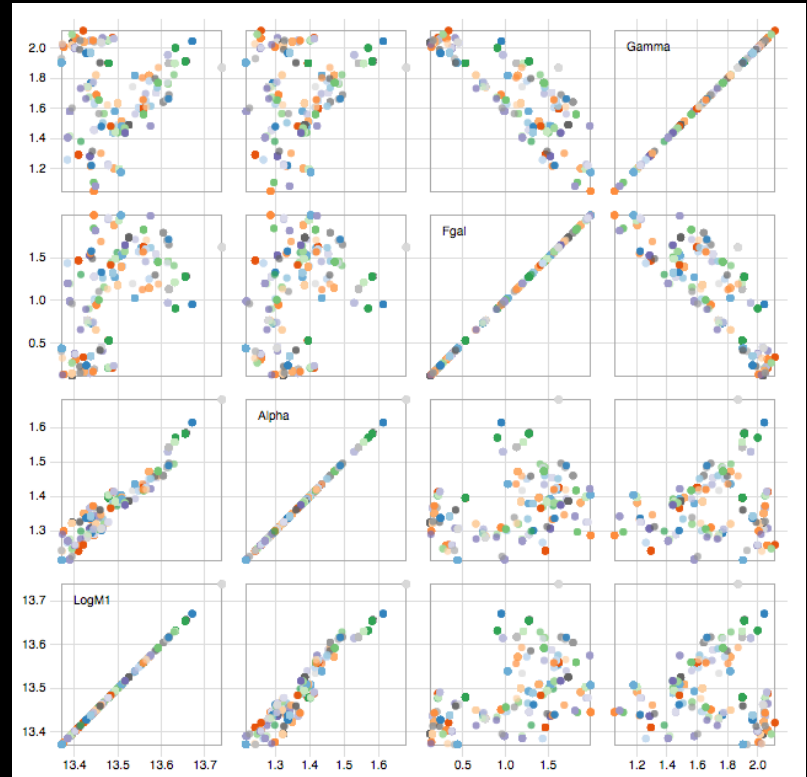
Identifiers:

- <ark:/31807/osdc-bf242fd3>

Analyze and Visualize

Find programs to analyze data
(ex. text analysis, statistical
insights, visual analytics)

Interactive Scatterplot using d3.js: Four
Different Parameters Plotted in a matrix to
quickly see correlations in the data



Problem Identification and Collaboration

- See projects other people have worked on
- Ask for help or seek opportunities to work together
- Cross-discipline, international collaboration

BROWSE AND UPLOAD DATASETS

Title	Description	User	Date	See Code	Data Format	See Data
Twitter Keyword Data	Nouns and verbs from tweets found using keyword 'data'	gracelu	6/9/15	tstream_time4.py	txt	data_sample.txt
Twitter Public Data	Nouns and verbs from tweets streamed from public timeline using NLTK	gracelu	6/9/15	public_tweet_format3.py	txt	public_sample2.txt
PDB Protein Data	Information about proteins	shelby	6/10/15	ids_to_nr.py	txt	ids_to_nr.py
Astroata	Plotting correlations between LogM1, Alpha, Fgal, and Gamma in astrodata	jennifer	6/11/15	scatter_script.js	txt	chains.csv