

# exploRase

Exploratory Data Analysis  
Software for Systems Biology

**Michael Lawrence**

# Outline

- Background
  - Exploratory Data Analysis
  - Software Foundation
  - Systems Biology
- Introduction to **explorase**
- Demonstration
- Your turn

**expl  
o  
R  
ase**

# Exploratory Data Analysis

- Exploratory Data Analysis (EDA) is the art of being a data detective.
- We look for trends and patterns, as well as deviations from those expected.
- Interactive graphics facilitate fluid data exploration and link different views of the data by visual cues.

# Software Foundation

- **R** – Platform for statistical computing.
- **RGtk2** – Binding between **R** and **GTK** (open-source GUI toolkit).
- **GGobi** – Open-source interactive graphics software. Based on the GTK+ toolkit.
- **rggobi** – Link between **R** and **GGobi**.

# Systems Biology

- The highly computational study of biology as a complex (biochemical) system.
- Based on the analysis and modeling of metabolic, regulatory, and protein interaction networks.
- Networks are often inferred from or analyzed in conjunction with high-throughput data (transcriptomics, proteomics, metabolomics) describing system states.

# exploRase is...

- An **RGtk2**-based graphical user interface that makes **R** and **GGobi** more accessible to a systems biologist.
- Facilitates loading of experimental data into **R** and **GGobi**.
- Serves as a front-end to **BioConductor** (**R** software for Bioinformatics).

# Let's go exploring...

- **Dataset** : *Arabidopsis* transcriptomic data from 8k Affymetrix chips.
  - 2 Genotypes: WT and bio1 (deficient in biotin synth)
  - Half of samples given biotin, total of 4 treatments.
  - 2 replicates
- **Data Source**: Basil Nikolau's lab

# Acknowledgments

- **Software:** Duncan Temple Lang, Deborah Swayne, Di Cook, Hadley Wickham
- **Feedback and Testing:** Heather Babka, Eve Wurtele, Suh-yeon Choi, rest of MetNet group

---

**Money:** MGET Fellowship, NSF Arabidopsis 2010