A Toolbox for Studying People and Info. Systems

MICHAEL EKSTRAND

FEBRUARY 17, 2016

http://coen.boisestate.edu/piret/

http://bit.ly/rd17pitools

BOISE STATE UNIVERSITY

PIReT People and Information Research Team

What We Study

- Recommender systems
- Information retrieval
- How people connect with them

http://bit.ly/rd17pitools

Workloads

- Data-intensive recommender evaluations
  - Much like machine learning experiments
  - Split data, train a bunch of algorithms, measure success
- Statistical analysis
  - Outcomes of the above
  - User study results
  - Exploring other data sets

Management Software

What holds it together?

- Linux compute servers
- tmux – start, disconnect, re-attach
- direnv – easy to manage per-project environment
- SSH client (PuTTY, MobaXterm, etc.)
- ngrok is useful for Jupyter notebooks
- Anaconda for installing software (Python, R)
  - Conda ‘environments’ are super-awesome

http://bit.ly/rd17pitools

Analysis Software

Mostly work in R

- Good support for statistical inference
  - As opposed to predictive model fitting
  - Has advanced stats packages like lavaan
- Very good graphics with ggplot2
- The Tidyverse is amazing
  - dplyr
  - readr

Data Processing

- Java for recommender implementations
  - Very good performance
  - Widely known
- Mix of tools for data munging
  - Python scripts
  - JavaScript (node.js is very good, if weird)
- Automation
  - Gradle for Java-based & general automation
  - Gulp for JavaScript automation

http://bit.ly/rd17pitools

Data Storage

- A lot of CSV files
- PostgreSQL for queryable storage
- Exploring Feather

Thanks!

http://bit.ly/rd17pitools