R for data sciences

COMBEE R Study Group
Brian S. Yandell
Fall 2017

https://combee-uw-madison.github.io/studyGroup/

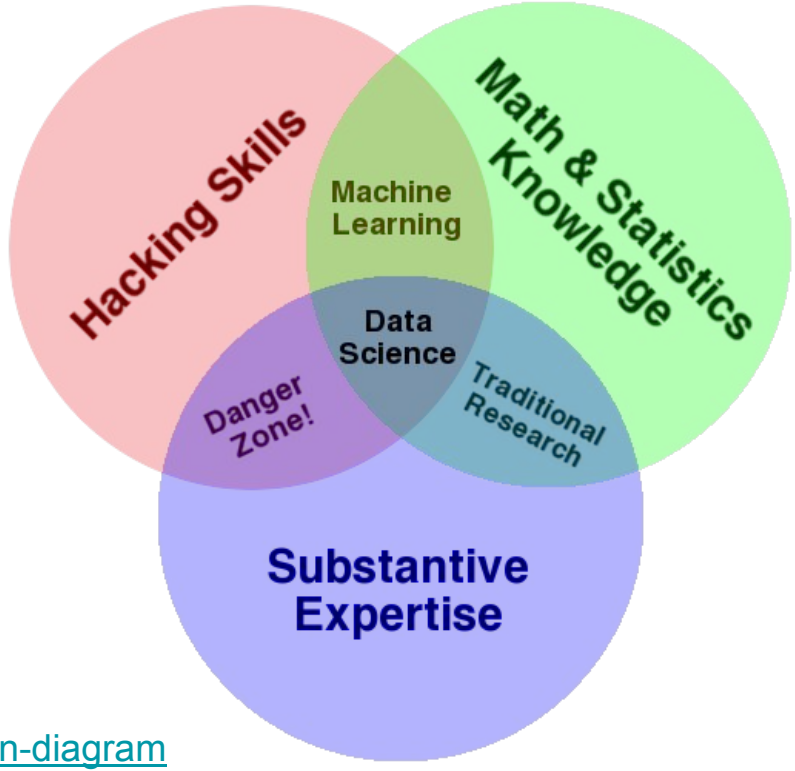
http://www.stat.wisc.edu/~yandell/R_for_data_sciences/

https://github.com/datascience-uwmadison/R_for_data_sciences

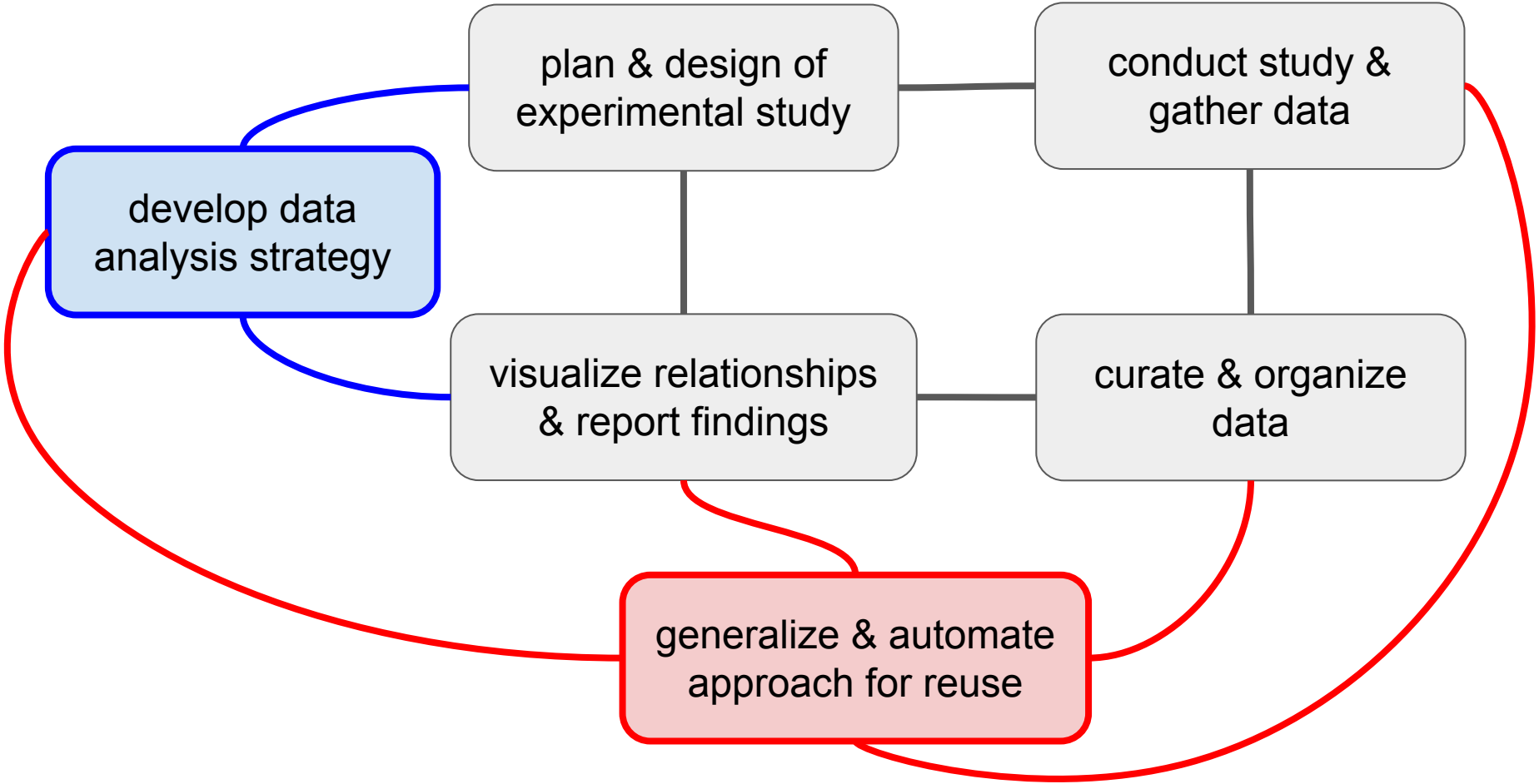
Drew Conway's Venn Diagram

- tries to capture what we do
- operational schema
- misses the bigger picture

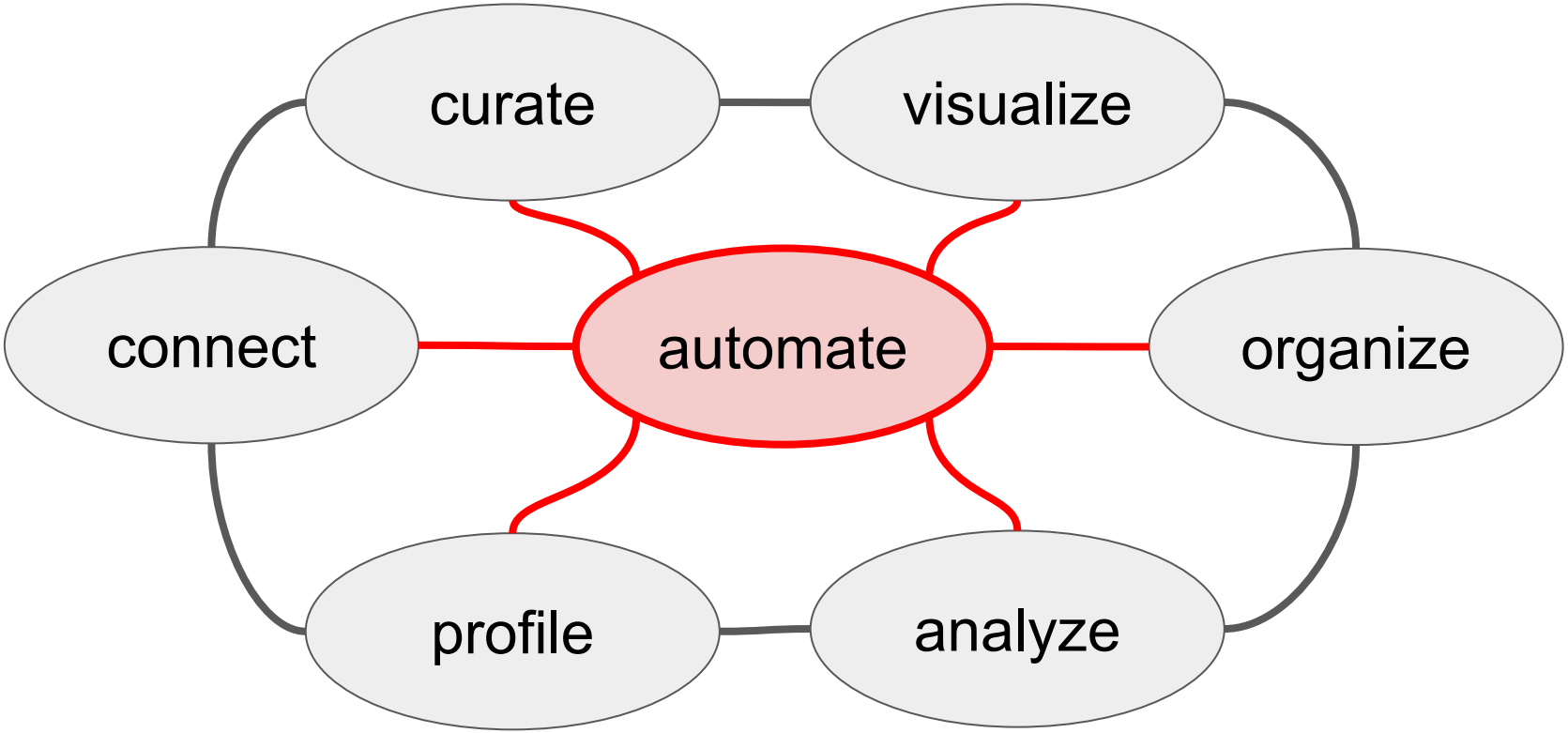
but just talking about statistics & computer science (& …) misses something as well



http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram


plan & design of experimental study
conduct study & gather data
develop data analysis strategy
visualize relationships & report findings
curate & organize data
generalize & automate approach for reuse

data science high-level verbs

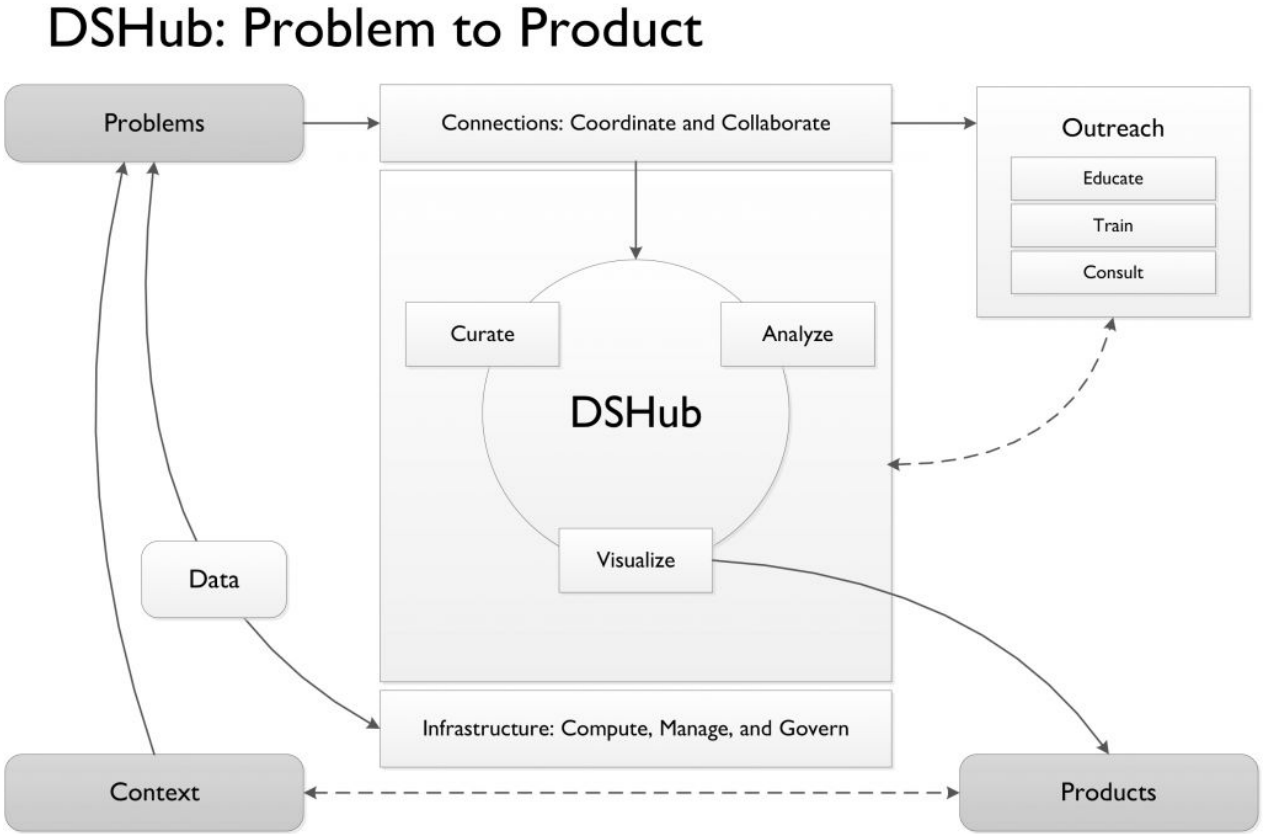


verbs of a data science ecosystem

data science hub goal:

- coordinate and execute data science strategy
- via campus-wide research network
- fill critical gaps
- support data science growth
- encourage cross-fertilization

http://datascience.wisc.edu



object-oriented approach to data environments

- S (percursor to R) and C (nee C++) were created at Bell Labs in 1970s
  - built of course on earlier ideas
  - but radically different approach than Fortran
- R and Python were developed in 1980s, Julia in late 2000s
  - R created by Ross Ihaka & Robert Gentleman using S interface but redesigned engine
- basic idea: operate on data objects as a whole
  - object contains useful information about itself
  - designing strategies (code=algorithms) for objects is smarter and cleaner
  - encourages thinking about goals rather than mechanics