

HUDK 5053: FEATURE ENGINEERING STUDIO

9/13/16 8:50 AM

In the news

SEPTEMBER 7, 2016

Stanford explores case for code of ethics to tackle big data's deluge in higher ed

Stanford partnership launches effort to help universities deal responsibly with student data.

Students use data analytics, algorithms to crack top colleges' admissions codes

Akiko Fujita

11 Hours Ago

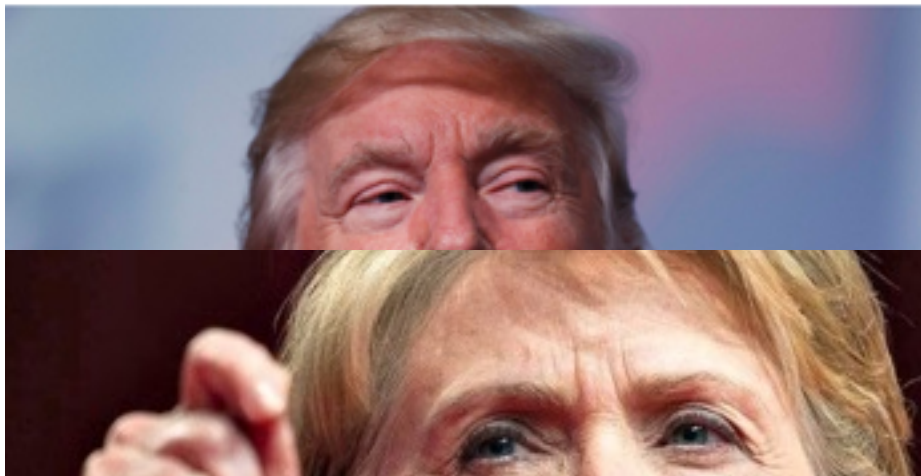


Clinton's and Trump's Plans to Help Education Differ Sharply

By THE ASSOCIATED PRESS
WASHINGTON — Sep 12, 2016, 5:33 AM ET

Share with Facebook

Share with Twitter



News > Edtech Business > Mergers and Acquisitions

Frontline Education Acquires Excent Corporation, Expands Special Ed Data Management

News > Edtech Business > Financing

Check I'm Here Scores \$1M Check to Help Universities Boost Engagement



Education ministry to promote use of big data



Native Hawaiians, Pacific Islanders Push for Recognition In Education Data

Predictive Analytics World for Government's 6th Annual Conference:
October 17-20, 2016 in Washington, DC



How to Maximize Your Salary as a Data Scientist



TOOLS USED

Learn Spark, Scala, D3, AWS. Stay away from proprietary tools and go towards open source. Learn 15+ tools.



INDUSTRY

Look towards social media, search, hardware, finance, and consulting companies.



LOCATION

Work in the West Coast of the United States. Choose between San Jose, Seattle and San Francisco, in that order.



ROLES

Make sure you're applying to data scientist roles, and not data analyst or engineering roles.



EXP. + DEGREES

Get an advanced technical degree to get your foot in the industry. Aim for 20+ years of experience.



NEGOTIATION

Practice your negotiation skills, and negotiate all the time.

Huang, R (2016)

Today

- Workflow: **Github** - RStudio - **Zotero**

The GitHub Octocat logo is centered in the background. It is a dark blue silhouette of an octopus with a light blue cat face in the center. The text "Git & Github" is overlaid on the logo.

Git & Github

THIS IS GIT. IT TRACKS COLLABORATIVE WORK
ON PROJECTS THROUGH A BEAUTIFUL
DISTRIBUTED GRAPH THEORY TREE MODEL.

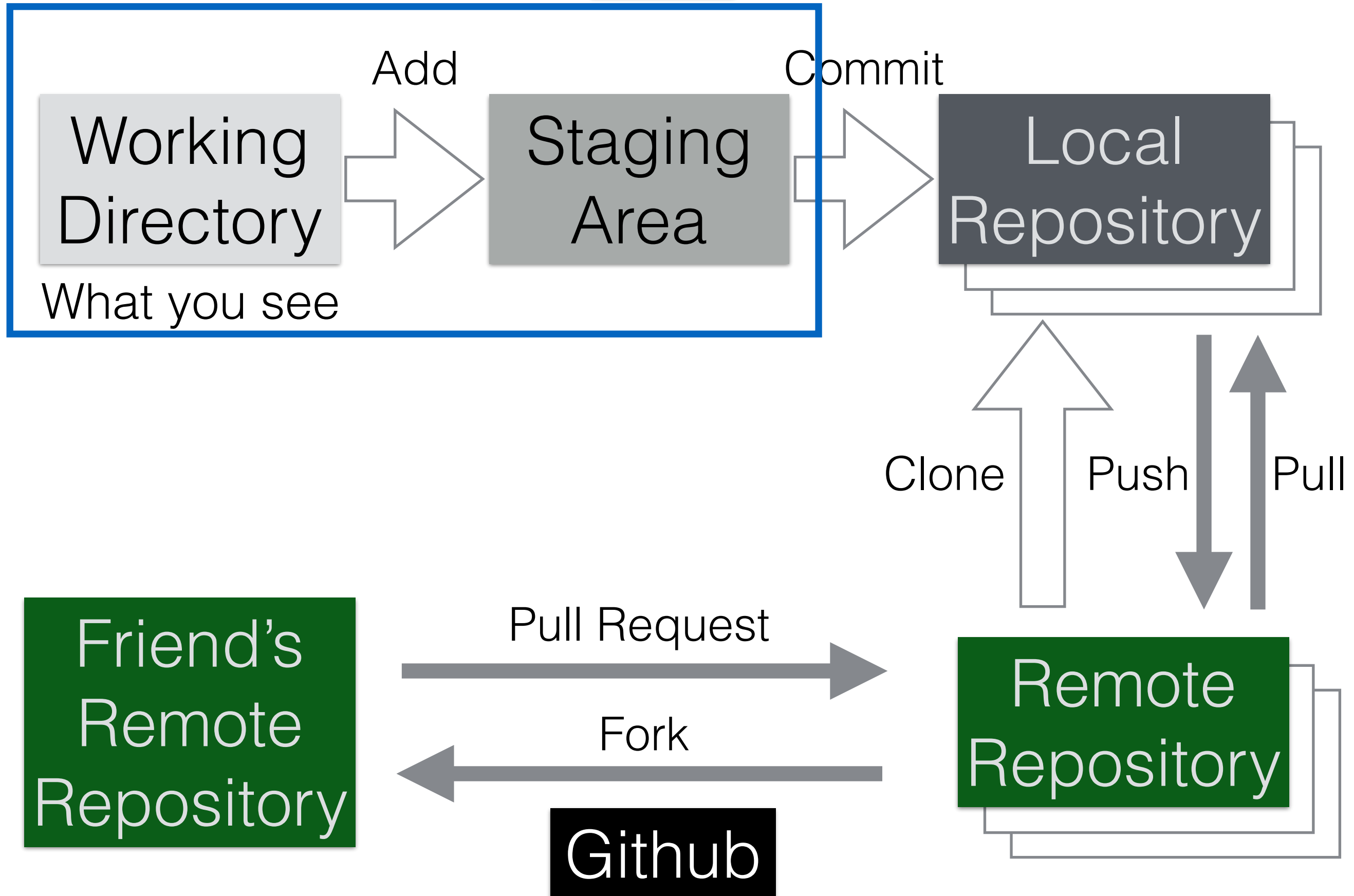
COOL. HOW DO WE USE IT?

NO IDEA. JUST MEMORIZE THESE SHELL
COMMANDS AND TYPE THEM TO SYNC UP.
IF YOU GET ERRORS, SAVE YOUR WORK
ELSEWHERE, DELETE THE PROJECT,
AND DOWNLOAD A FRESH COPY.



RStudio

Git




Functions

- Open syllabus
- RStudio integration

[https://github.com/
feature-engineering-
studio/workflow](https://github.com/feature-engineering-studio/workflow)

[https://
csgillespie.github.io/
efficientR/](https://csgillespie.github.io/efficientR/)

Github Forking

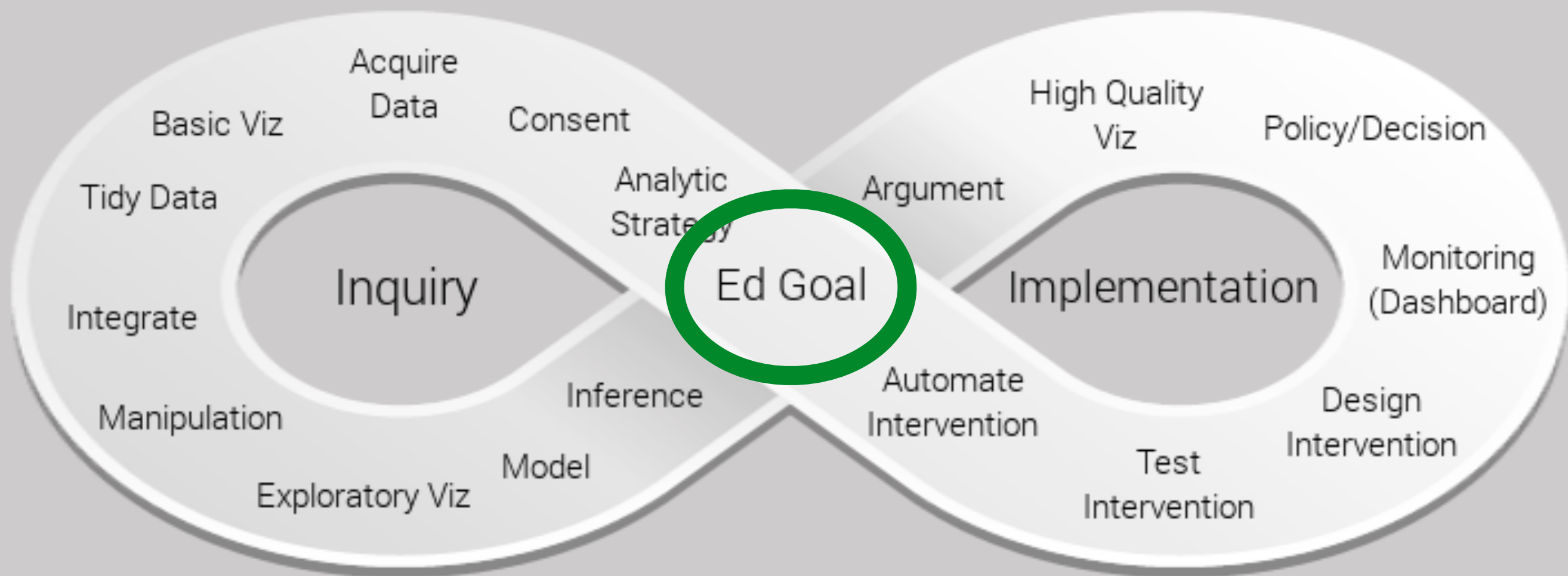
- Log into Github/Create account if needed
- “Fork” feature-engineering-studio/syllabus page to your personal account
- Click on the “README.md” file in your clone’s version
- Find something you would like to change
- Click  and change it!
- Scroll down and click “Commit Change”
- Click on the “Pull Requests” tab and then “New pull request”
- Describe the change you made in the message and click “Create Pull Request”

Zotero

- Install
- Download the .rmd file
- Upload into your account
- Write a note on educational goals

Educational Goals

Ed Data Science Cycle



“It’s better to be at the bottom of the ladder you want to climb than at the top of the one you don’t.”

–Stephen Kellogg, 2013

Google Books Ngram Viewer



Why?

- Education is controversial
- Concepts cross disciplinary lines
- Vocabulary is slippery
- Systems are complex

Characteristics

- Specific/tightly defined
- No jargon
- Measurable
- Attainable
- Relevant
- Time-limited
- Defined level of inference
- Contextual framework
- They can change (BUT keep a record of the change)
- Can be to answer a question (To determine the time of day students write the most words.)
- Needs to be workshopped with others

Activity

1. Find a partner
2. Determine one educational goal that you would like to define for yourself, for this semester
3. What information would you need to allow you to determine if you had met that goal