# ANAIYICS PROCESS &

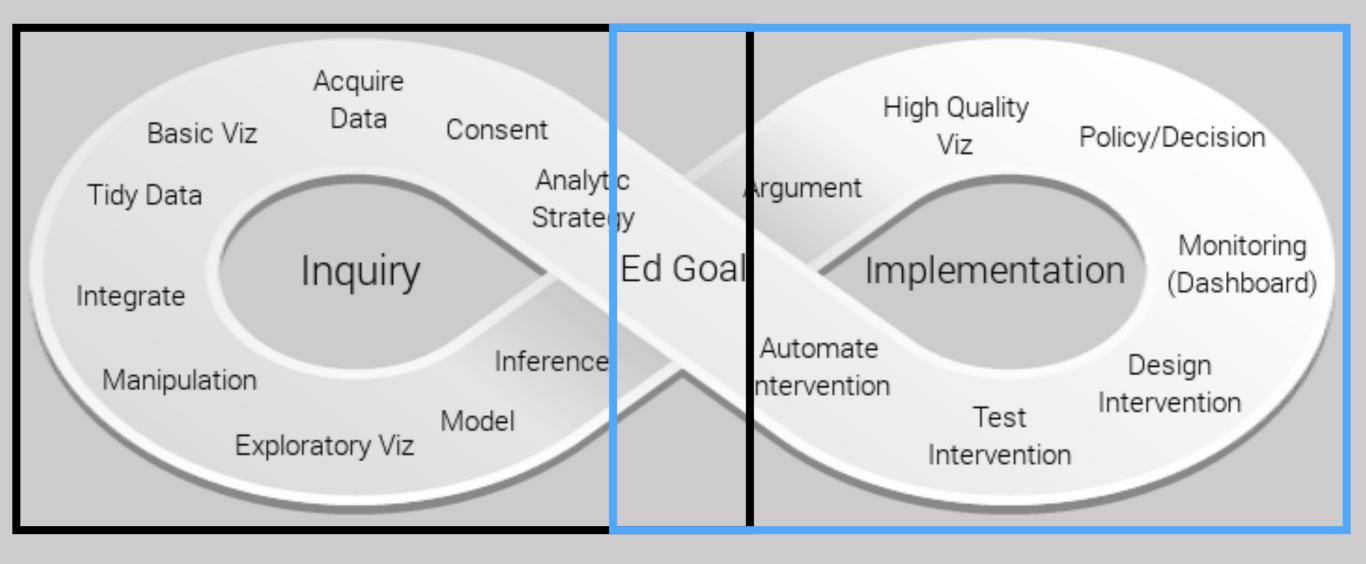
### Course Goals

Short Term		Long Term	
Content	Skills	Abstractions	Habits
remember, understand	demonstrate	synthesize, argue	organize, implement
Conceptual basis of methods, use cases	SQL, R, git, application of methods	Evaluate broader implications, have opinions, methods schema	Workflow, documentation, JIT learning

## Ed Data Science Cycle

HUDK4050

**HUDK4051** 



### Purpose

Build a portfolio of projects that you can show prospective employers

#### Slides

https://github.com/la-process-and-theory/lecture-slides

### Setup

https://github.com/la-process-and-theory/setup

# Syllabus

https://github.com/la-process-and-theory/syllabus



### Units

- · SQL
- Some old friends: Prediction & SNA
- Natural Language Processing
- Neural Networks
- Closing the Loop
- Interactive Visualization
- Recommender systems

#### What is the work?

- Attend class
- In person meeting with instructor (sign up)
- Complete 8 Units
  - Readings/Videos
  - Assignment
  - Self Test

- Real World Project
  - Set up a SQL database an analysis
  - Present to class
  - Analyzing Podcasts for the MASCLAB
  - Will kick off 1/30

#### Format

W1 W9 W15

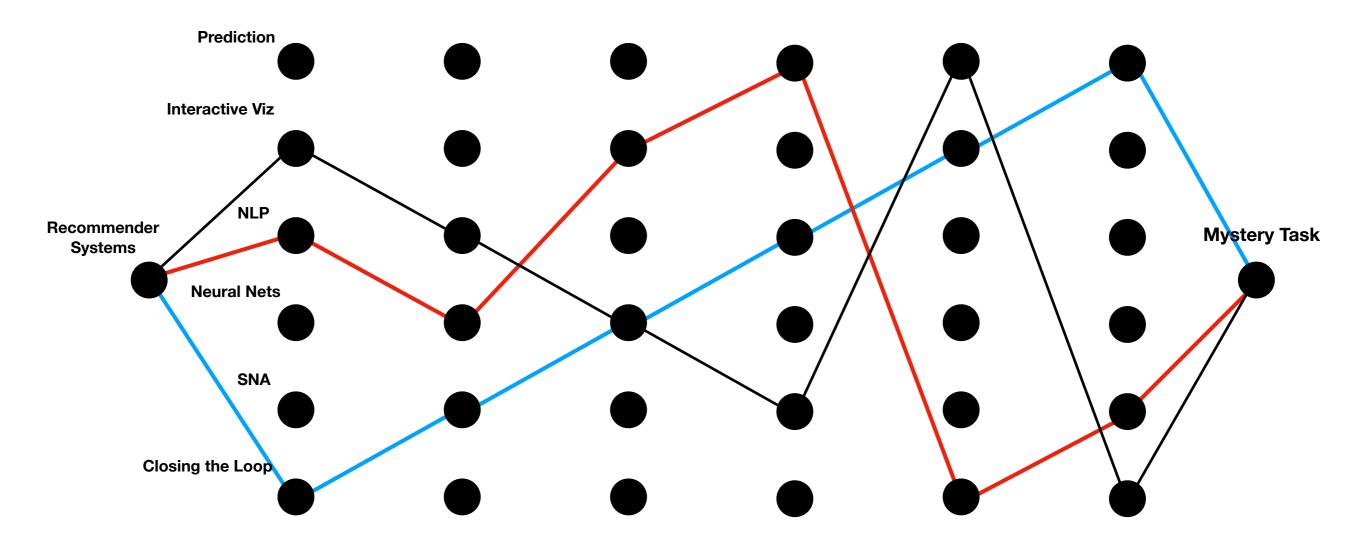
**Second Pass of Topics/ First Pass of Topics Class Choice** Group Group Group Group Group Group **Project Project Project Project Project Project Unit Deliverables Due** 

> Mystery Unit

### Units

- <u>Disaggregated</u>: You do them in any order that takes your fancy
- BUT there is no choice in the first and last units
- Complete one unit every week for six weeks
- One mystery unit

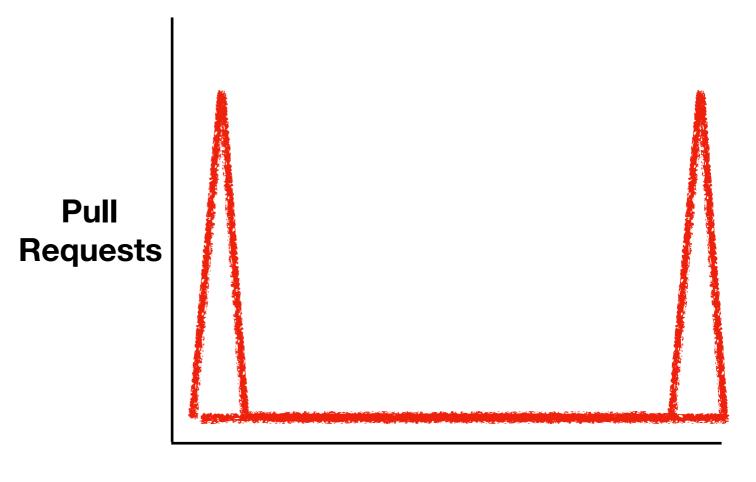
### Units



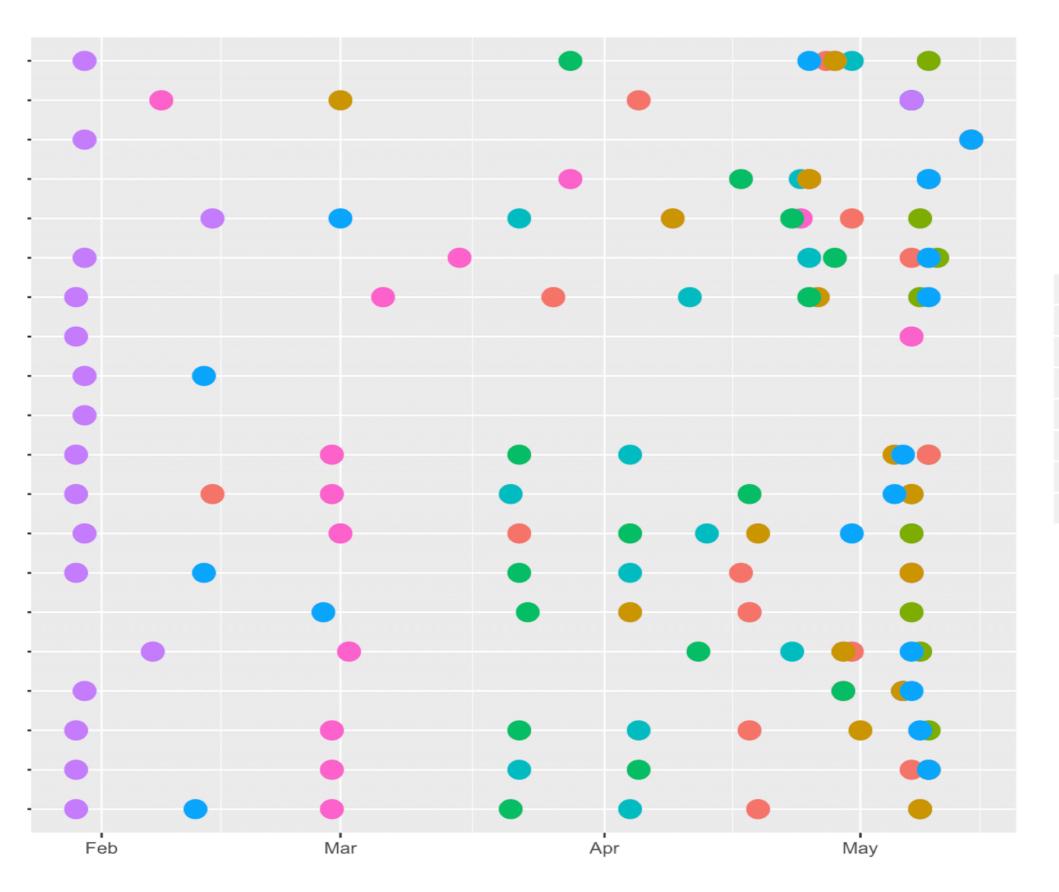
#### Assessment

**Quantity** & **Content** of Pull Requests

Pull request early, pull request often



Weeks



- interactive-visualization
- loop-closing
- mystery-unit
- natural-language-processing
- neural-networks
- prediction
- recommender-systems
- social-network-analysis

# Why?

- Content isn't necessary cumulative
- Learning is more effective if you can follow your own interest
- Students are better arbiters of what they need to know than instructor
- Promotes self-motivated learning
- More time getting instructor feedback
- Get to test the tool you build in unit 1

### If you did not take HUDK4050

Must see me at the end of class

### If you did take HUDK4050

- Assignments are more difficult in HUDK4051
- But you will get more assistance

#### Immediate to dos:

**Setup:** https://github.com/la-process-and-theory/setup

- You need accounts: Github, AWS
- You need to install: R/RStudio & Git
- Download data fro next week

#### Survey

#### Take a look at Recommender Systems

- Issues tab
- Additional material

### Questions?

#### To Dos

#### Survey

http://bit.ly/hudk4051survey1