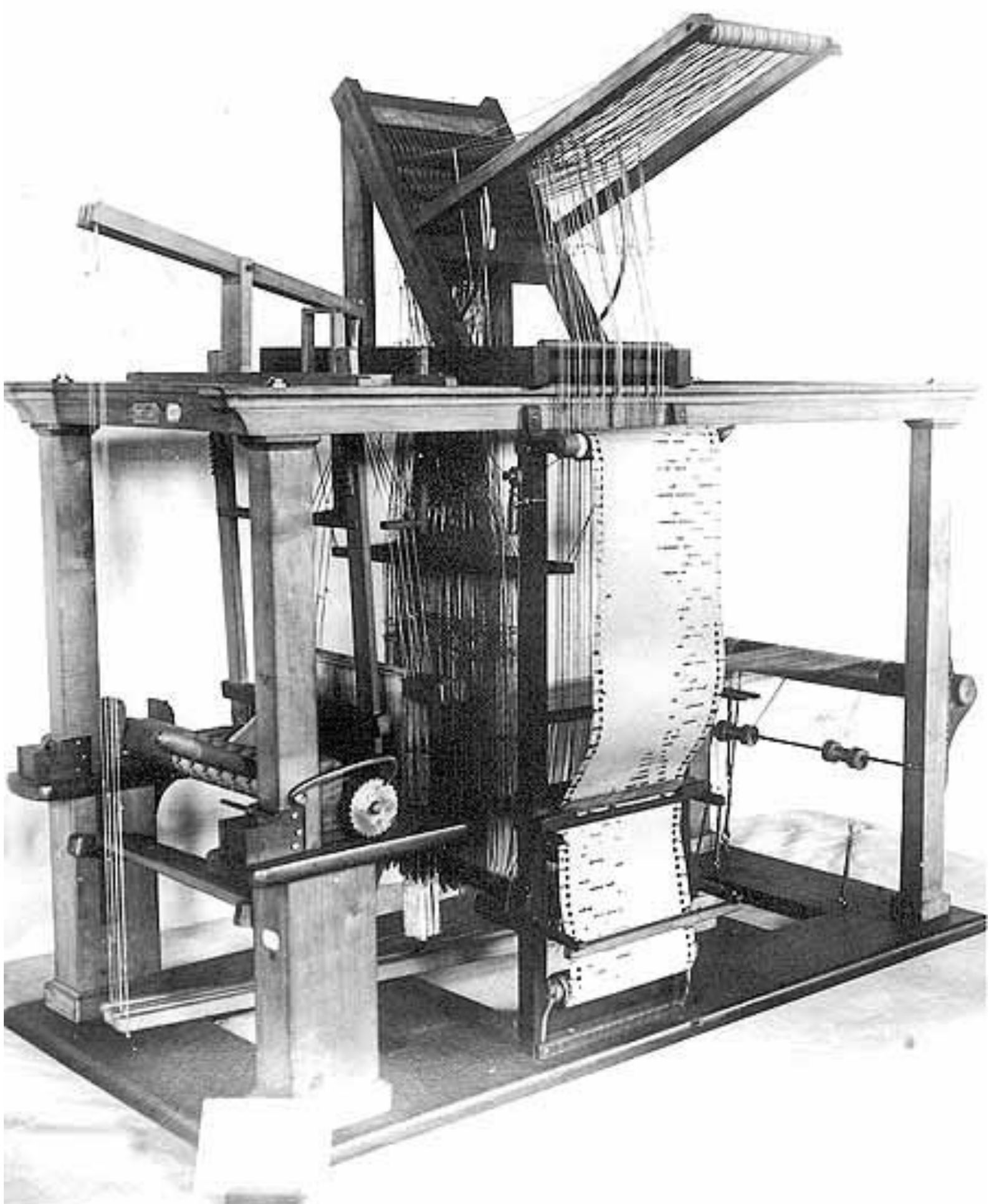
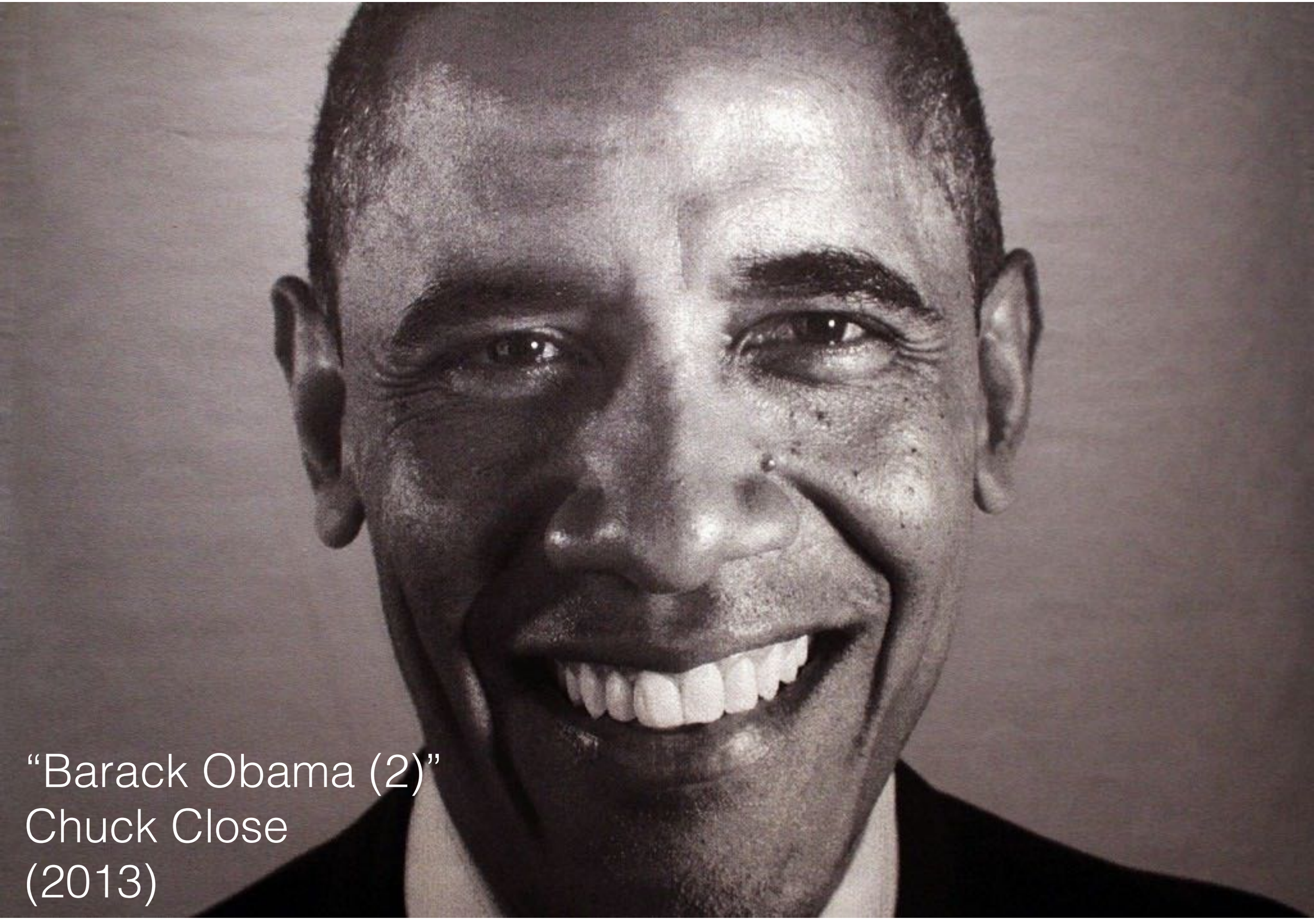
HUDK 4050: CORE METHODS IN EDM

9/5/18 11:04 PM

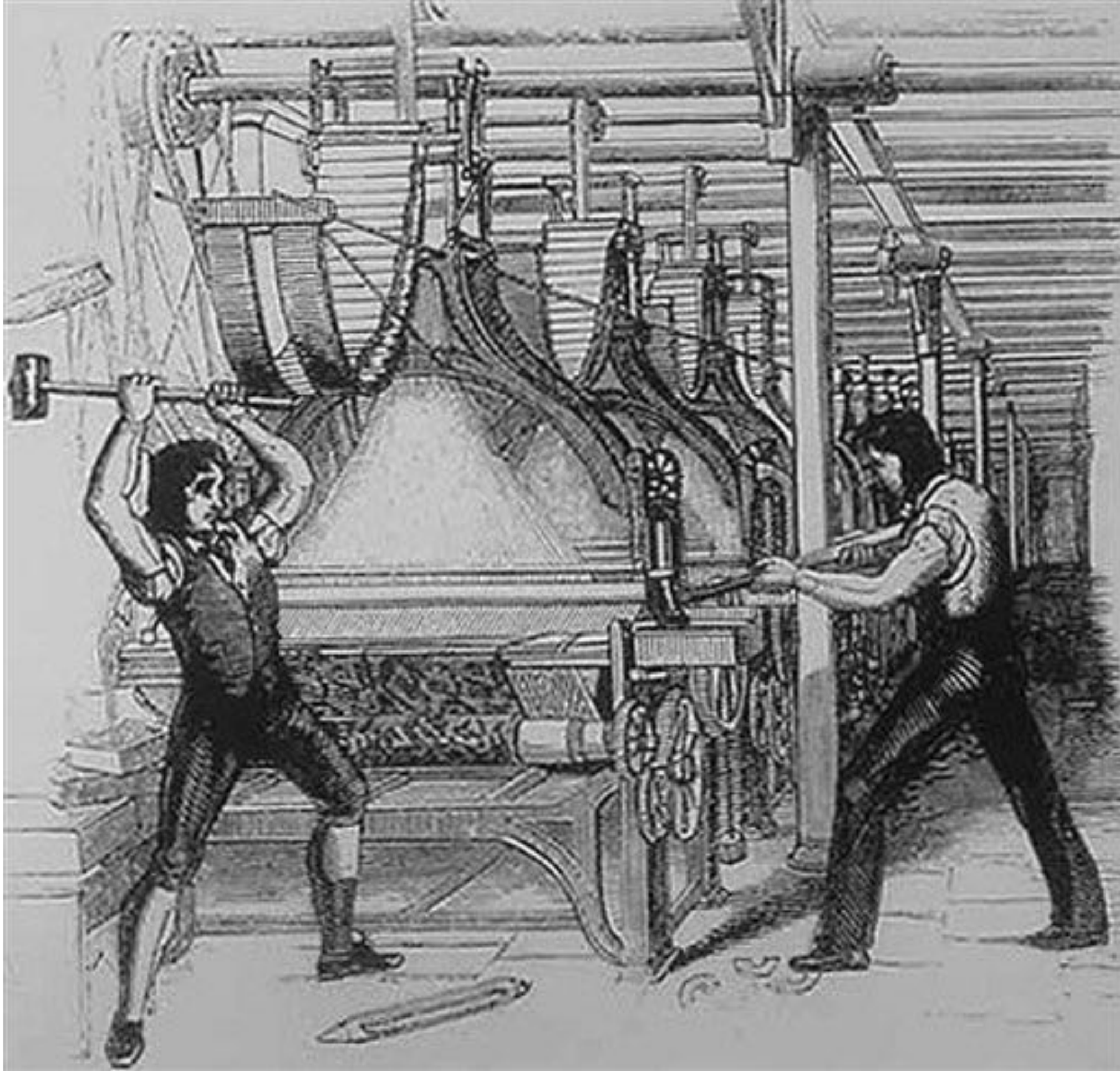
Jacquard Loom
(1801)





“Barack Obama (2)”
Chuck Close
(2013)

Luddites
(1811-17)



Tweets | Media | Likes

leonie haimson @leoniehaimson 2d
Treating kids like cattle
@Parents4Privacy

> **Lori Lalama** @TechEducator1
>
> #DataMining #PrivacyAware
> @leoniehaimson @DianeRavitch
> wired.com/2012/09/rfid-c...

2

leonie haimson Retweeted

Sheila Resseger @sheilaresseger 2d
@leoniehaimson I had no illusions 2
the contrary! We r up against a
relentless & callous assault on
human dignity, preventing self-
mpowrment

2

Mark Zuckerberg
December 3, 2015 at 5:24pm · Palo Alto, CA ·

I want to thank you all for your heartwarming congratulations on Max's birth and on starting the Chan Zuckerberg Initiative. This whole community has been so loving and supportive.

If you're interested in following the philanthropy work we're doing with the Chan Zuckerberg Initiative, I encourage you to like the page here:

https://www.facebook.com/chanzuckerberginitiative

Since we announced this a couple days ago, many people have asked about what we're planning to focus on and how we're structuring our work.

Our initial focus areas are personalized learning, curing disease, connecting people and building strong communities. We've already made many investments over the past five

1 THE RAPID GROWTH OF GLOBAL DATA

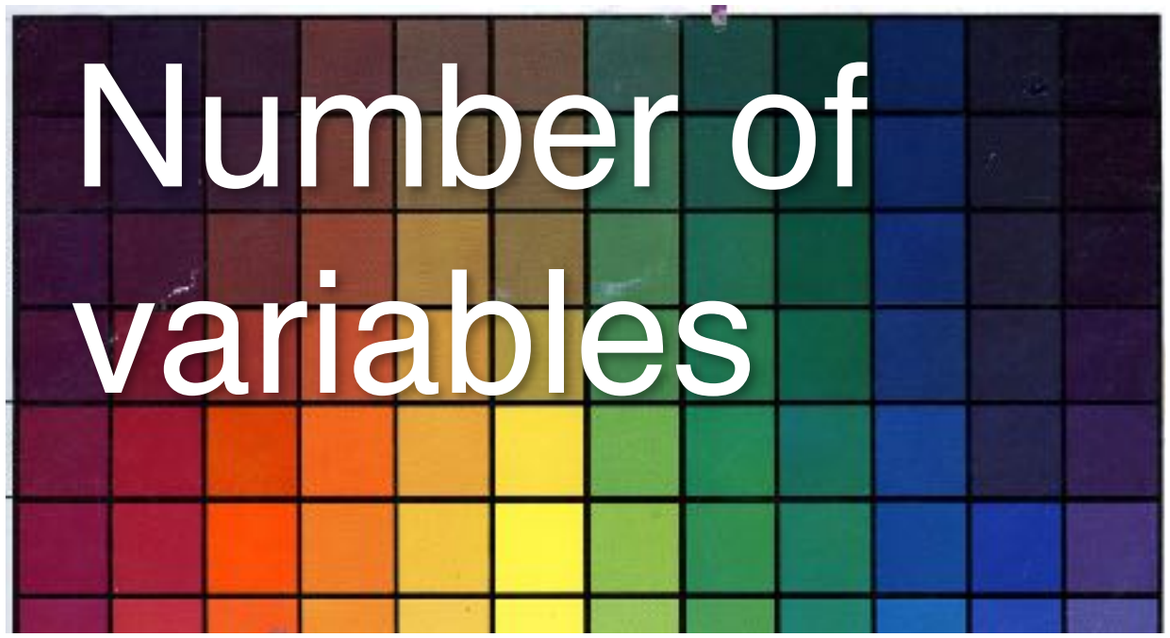
CSC

The production of data is expanding at an astonishing pace. Experts now point to a 4300% increase in annual data generation by 2020. Drivers include the switch from analog to digital technologies and the rapid increase in data generation by individuals and corporations alike.

Computer Sciences Corporation, 2012

What's new?


Number of
data points


Number of
variables


Track
Annotate
News
Zoom
Number of
time points

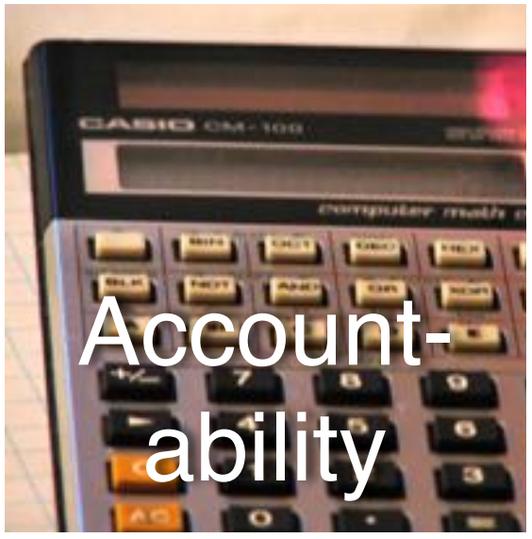

Feed-
back

MR. MESSY

by Roger Hargreaves

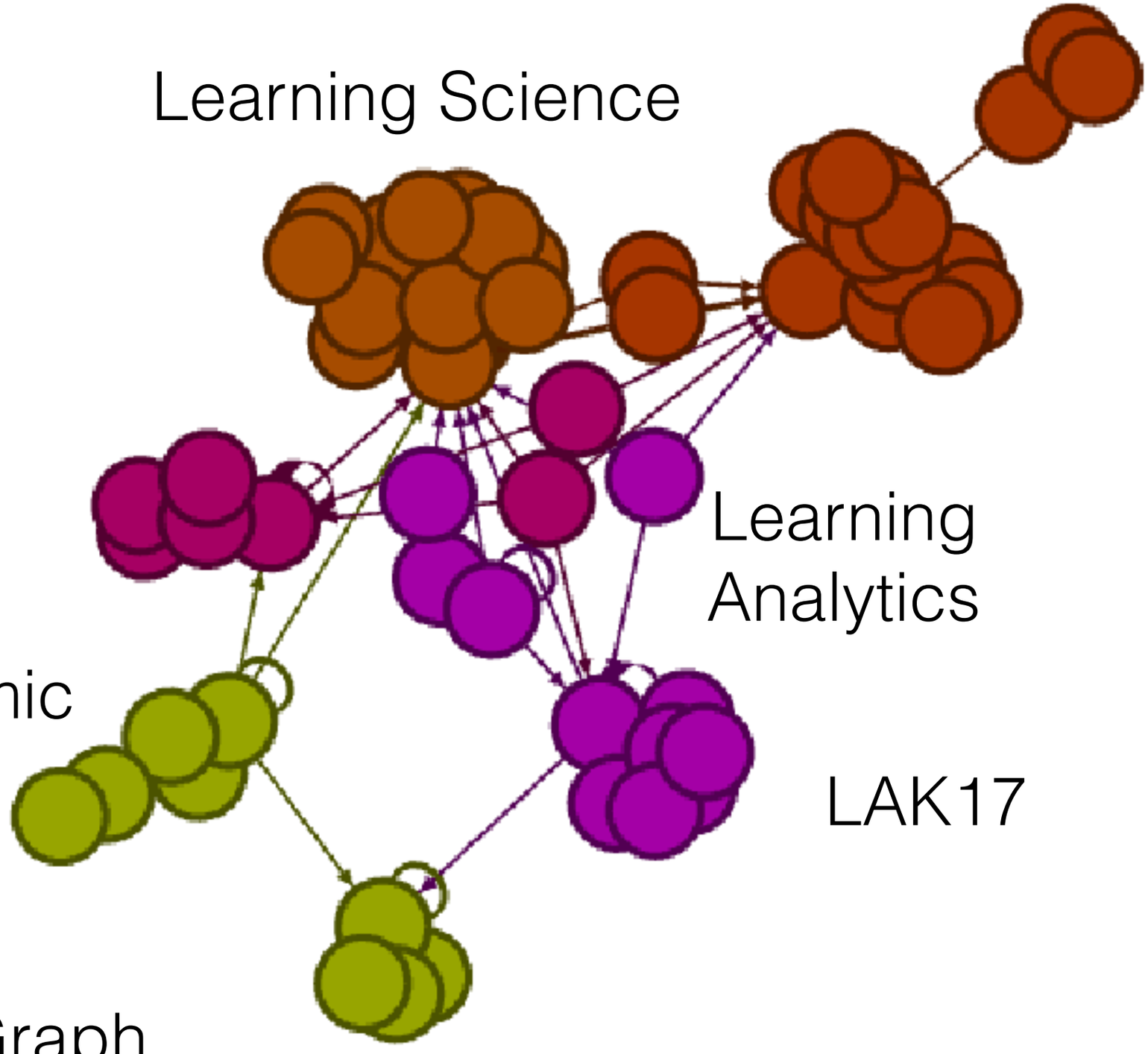
- Volume
- Velocity
- Variety
- Veracity

Possibilities



Computer Science

Learning Science



Learning
Analytics

EDM/Academic
Analytics

LAK17

Twitter Social Graph
#learninganalytics
2 weeks worth of tweets

Learning Analytics

- Grew from BI
- Early-mid 2000s in Canada
- Large domain umbrella - EDM, MOOCs, learning sciences, game design
- Systems oriented
- Methodologically broad (qual, quant, CS, stat)
- Journal: JAL
- Society: SOLAR

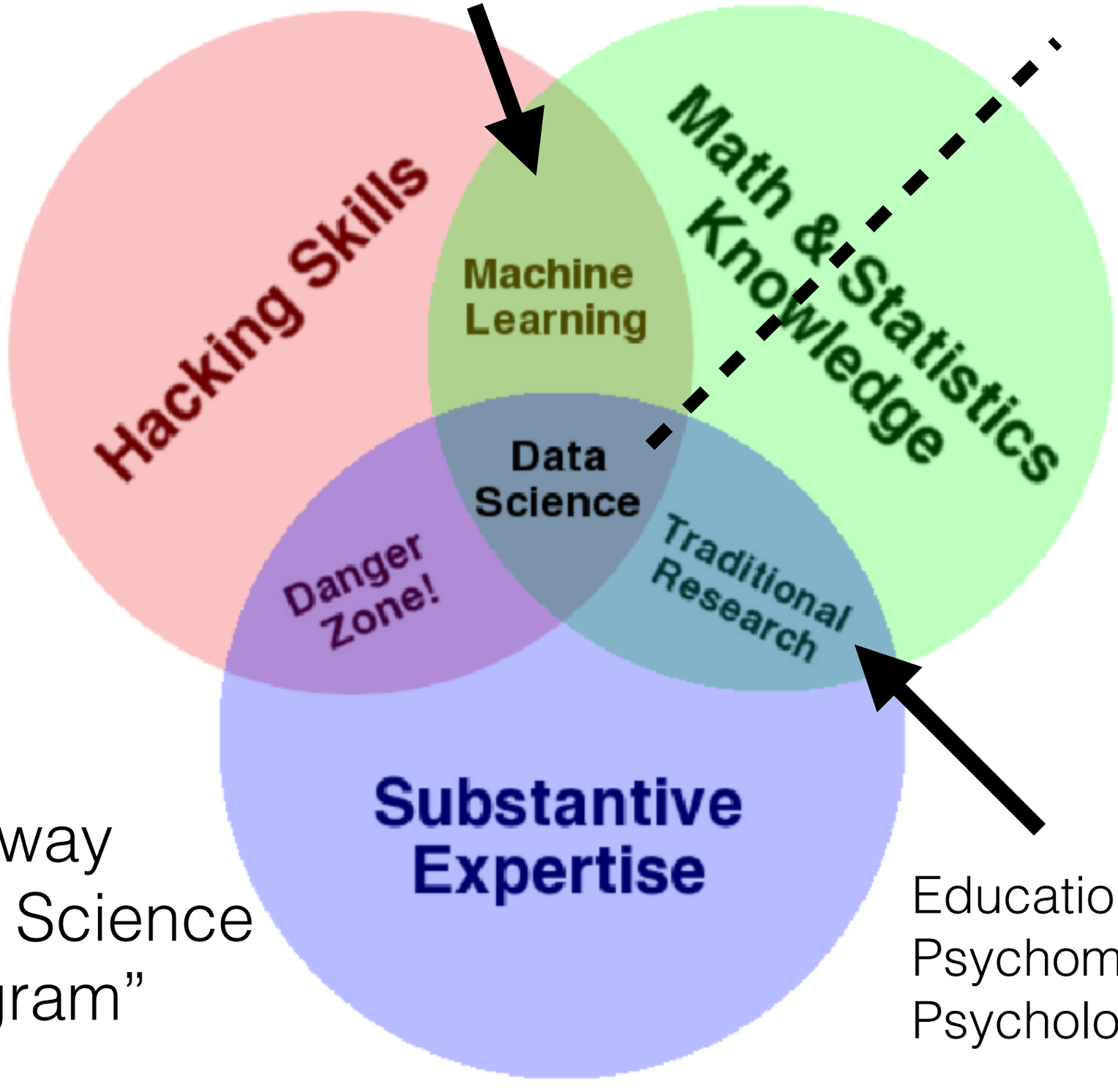
Ed. Data Mining

- Grew from Knowledge Mining or KDD
- Late 90s at CMU
- Intelligent tutoring, log-file
- Software oriented
- CS methods
- Journal: JEDM
- Society: IEDMS

A Data Scientist is a statistician who lives in San Francisco

–Sean Owen (Cloudera), 2014

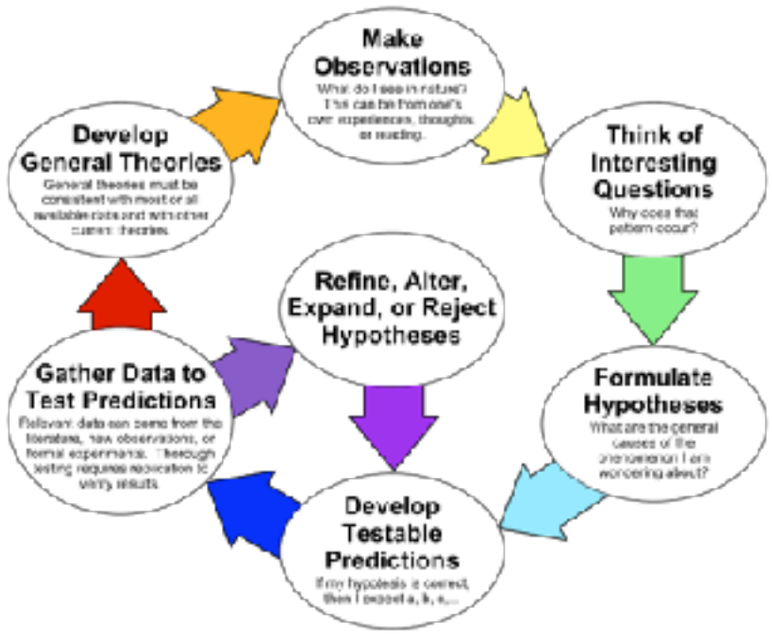
EDM
Academic Analytics



Drew Conway
"The Data Science Venn Diagram"
(2010)

Educational Statistics
Psychometrics
Psychology

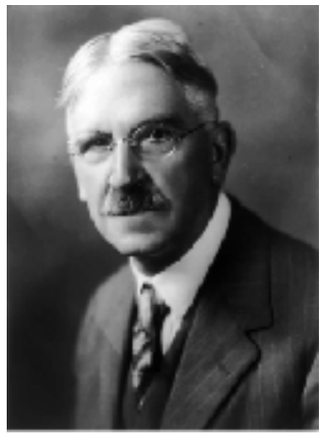
Philosophy of Science Spectrum



HUDK 4050

Mr Vargas

Delores Etter



John Dewey

William James

Paul Feyerabend

Rick

Relationships

Explanations

Prediction

Intro Survey

http://bit.ly/HUDK4050INTRO

Today

- Course logic & structure
- What a lesson will look like
- Assessment

Instructor

- email: charles.lang@tc.columbia.edu

- appointment: sign up using link on syllabus

Syllabus

https://github.com/core-methods-in-edm/syllabus

Course Goals

- Foundation in methods
- Provide a solid workflow for doing LA moving forward
- Not scare you away!

Course Goals

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

Assessment

<u>Weekly:</u>

- Attend class
- Readings
- Post article to Vectr with Questions & Answers
- Automated R course (Swirl) } Competition
- Maintain documentation of work (Github)

Assessment

<u>One time:</u>

- Ask one question on Stack Overflow
- 7 x short individual assignments
- 1 x group assignment (3-5 students)

Assessment

How to assess a course that is all about how difficult it is to measure learning?

- Assess your preparedness to do the work after the course has ended

- Two measures:

    - Contribution: assignments, comments, quizzes

    - Organization: keep a record of what you have done (Zotero, Git, Markdown)

Assessment

Purpose of grades

- Not performance based

- Entirely a tool to overcome **immunity to change**

Assessment

<u>Purpose of grades</u>

- You will be assessed based on:
- Timeliness (5pm)
- Comprehensiveness
- Organization

Tools

- Git/Github

- R/RStudio

- Vectr

- Tools that are worth learning in & of themselves
- Tools that we can easily extract data from
- Tools that you can use unrestricted in the future

Git/Github

- Git is a common version control system
- Github is an online hosting service for Git files

We will be using Git/Github to:

- Keep track of work
- Document problems
- As an LMS (assignment submission, etc)

R/RStudio 

- R is a scientific programming language

- RStudio is an Integrated Development Environment

Why use R?

<u>Pros</u>

- Free
- Platform independent
- Actively developed by a large community of users
- Functionality is VAST
- Help resources are VAST
- Best graphics (At the moment)
- All the cool kids use it
- Dominant language in Dept.

Why use R?

Cons

- Slow (for some tasks)
- Not designed to build tools
- Relies on vectorization
- Non-intuitive structures & syntax with respect to other programming languages

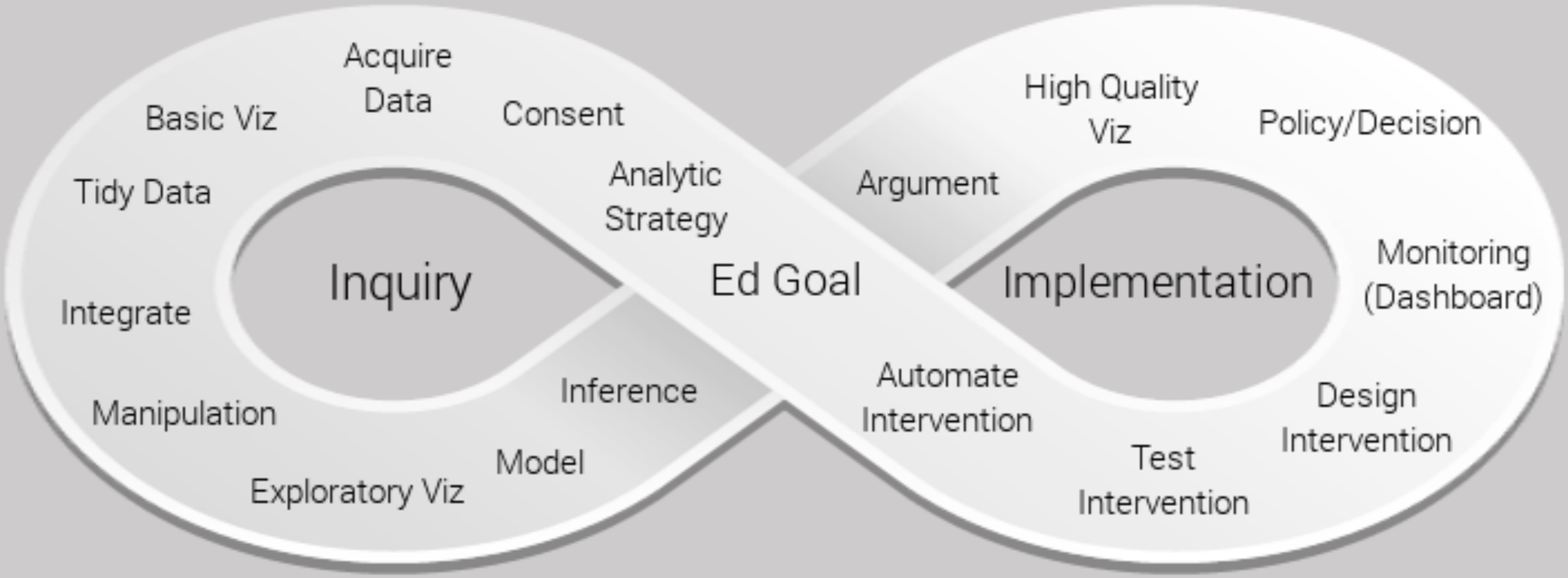
Why use R?

It is designed for people who are *learning* a method *as they apply it*.

Vectr 

- Q & A Platform
- App/Website
- Synthesizing methods
- Putting methods in context

Ed Data Science Cycle



Questions?

Assignment 1

https://github.com/core-methods-in-edm/assignment1