HUDK 4050:
CORE METHODS
IN EDM

9/25/18 2:56 PM

News

EdSurge

Tear Down That Wall? Why Data Walls May Cause More Harm Than Good.

By **Tina Nazerian** Sep 7, 2018

Return on investment in higher education: a student's view

THE DAILY SENTINEL

ELEARNING INSIDE

NEVADA TAKES STRONG APPROACH TOWARD VIRTUAL CHARTER SCHOOL DATA

Analytics Insight

THE TOP 10 DATA MINING TOOLS OF 2018

REUTERS

The artist as an algorithm: robot-made Rembrandt for sale

Vectr Q&A Format

- Headline

space

- Question

space

- Link

- Anonymity: anonymous

Events

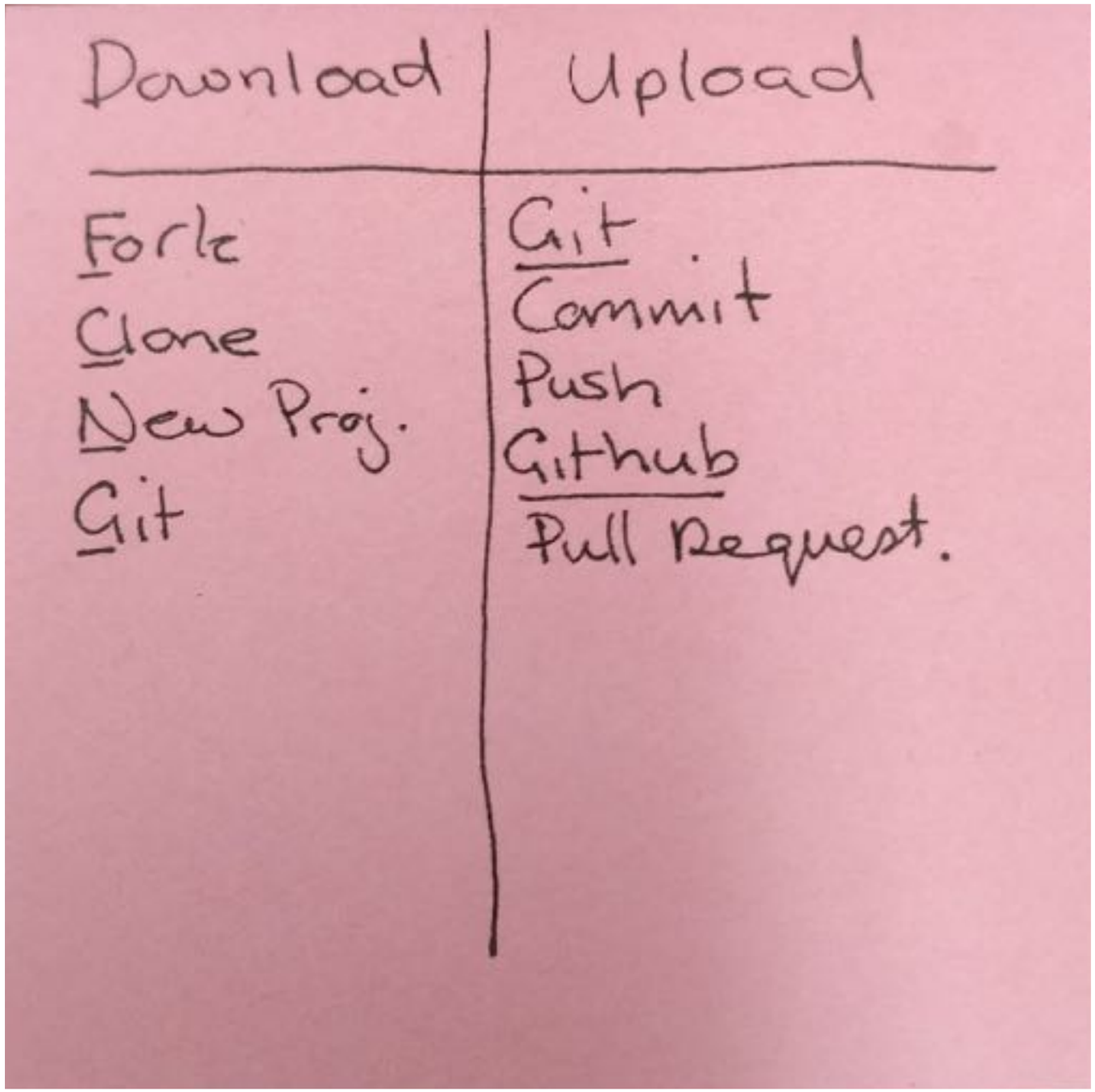
Title	Date	Link
Semi-automated exploration and extraction of data in scientific tables	5:00pm September 26	https://cvn.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=c3b4fa1d-5fac-4a48-93f8-a9530130e792
Data Science Institute Town Hall	10:00am September 28	https://events.columbia.edu/cal/event/eventView.do?b=de&calPath=%2Fpublic%2Fcals%2FMainCal&guid=CAL-00bb9e28-657a7368-0165-7c79c918-0000182cevents@columbia.edu&recurrenceId=
California Safe Drinking Water Challenge	Due Oct 1	http://waterchallenge.data.ca.gov/
Changing the Airline Industry Beyond the Aircraft	5:00pm October 4	https://events.columbia.edu/cal/event/eventView.do?b=de&calPath=%2Fpublic%2Fcals%2FMainCal&guid=CAL-00bb9e24-655b8449-0165-5e0596df-00001917events@columbia.edu&recurrenceId=
Learning Analytics in Physical Spaces	12:00pm October 2	https://events.nyu.edu/#!view/event/event_id/215979
People centric approach to optimize Data Science, Commercial impact and Leadership	10:30am November 14	https://events.columbia.edu/cal/event/eventView.do?b=de&calPath=%2Fpublic%2Fcals%2FMainCal&guid=CAL-00bb9e24-655b8449-0165-5e0ea7e9-00001957events@columbia.edu&recurrenceId=
Big Data for Intelligence Symposium	Washington DC, October 16 - 17	http://bigdatasymposium.dsigroup.org/

Announcements

- Swirl

- Assessment

Git/Github


Download
Upload
Fork
Clone
New Proj.
Git
Git
Commit
Push
Github
Pull Request.

Class Activity Answers

- Will post to Github

Data Wrangling II

- Matrices

Matrix vs. Data Frame

Matrix

- Uses less memory
- Operations are faster
- Requires same data type (character or numeric)
- Useful for matrix algebra

Data Frame

- Convenient
- Intuitive
- Can have different data types in one format
- Useful for referring to columns individually

Create Matrix

- `matrix()`
- `as.matrix()`

Other Useful Operations

Transpose Function

- `t()`
- Transposes a matrix or data frame
- rows ⟶ columns, columns ⟶ rows
- Output = matrix

Diagonal Function

- `diag()`

- Replace or extract the diagonal of a matrix

Matrix Multiplication

- %*%
- Two matrices must be the same size
- Multiplies the rows of the first matrix by the columns of the second
- Multiplies two matrices together
- Will become useful when we get to Social Network analysis

Activity

- Create a data frame called **A** of three variables, each having three values

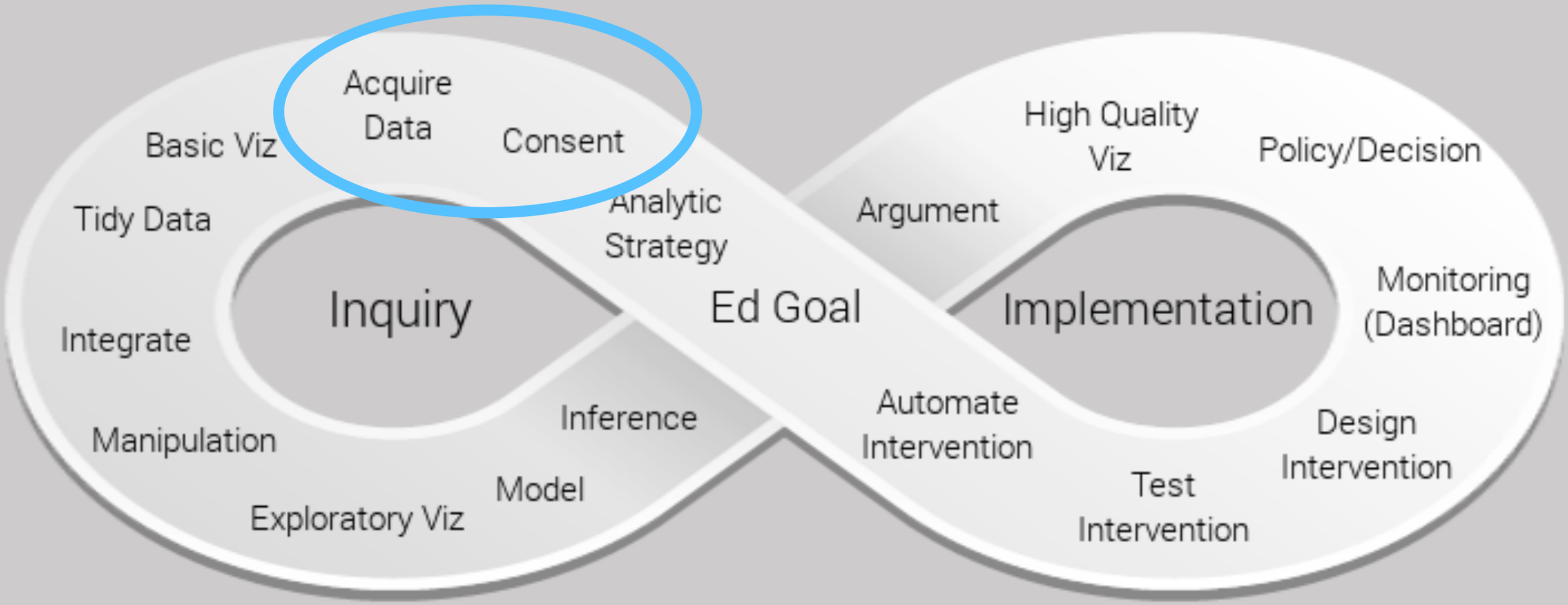
- Convert the data frame to a matrix called **B**

- Create a matrix called **C** that is the transposition of **A**

- Create a matrix called **D** that is the multiplication of **C** and **B**

- Replace the diagonal values in **D** with missing values

Ed Data Science Cycle



Activity

- Educational goal:
- Baseline: What is your baseline measure?
- What level should we be looking at?
- What variation do we need?
- Is there any other information we need
- What information should we track?
- Do we need to ask permission? Of whom?
- Do you wants to share results with students? Why/why not