HUDK 4050:

CORE METHODS

IN EDM

9/18/18 11:52 AM

Events

Title	Date	Link
California Safe Drinking Water Challenge	Due Oct 1	http://waterchallenge.data.ca.gov/
Gigged: The End of the Job and the Future of Work	4:00pm September 19	https://datasociety.net/events/databite-no-113-sarah-kessler/
AI: Building Products that Understand	6:00pm September 19	https://meet.meetup.com/wf/click?upn=pEEcc35imY7Cq0tG1vyTt6kYBBpMSWKNFnrHIgtlsQcV2vLEHAeU4PTsxc6f5OrFX7zI7QSmevdc3nb1zTPXNiXlxpp2sZSSdoTKpJBysvsuJyytM8jetPacPzbp51ZXp8Qju6g3ugF-2FESMea-2FltjZeLn4L-2B-2BdOROXtm0sn-2FXJS-2FGv7XdowrZ3qYTyLGtrojim0D0gL9cHri6w8M4saX2L9KlZ1YkNKroDi72blxLD9Lz6Rv-2BMOtKff4gIkDWwPazJ0nChogE4bPmVxBrIzP0LLSdElyXeas6oekT7u36wUUNK-2FyzDg64ji-2BZoIVAFSyZ-2B9gRuayrXsVvLYDU73BSw-3D-3D_I3fs1g-2F466j3y5fD5Q61KTvREVidSz8eOmcqorpVEXbt5H1apoGNSAiFy94udlsHIY1PSkmM0wPZ-2FQuxZTuFZkEjGhVpzXS1aDcO8D7IERoAkNk0W5nGsGSpBxyCvr11zY-2FT49LaaSYfiOfXriPdDdkiroJ8RFeOPYN9GssUBjtQOHV-2Bn34Id0mm0RaBiT8bUCDW9HyM5OLOTK0wxIZ6SyOfB2We8IMQCqn2cdjAPlQ-3D	
Classifying brain waves using AI - Exploratory data analysis of EEG data	5:30pm September 19	https://meet.meetup.com/wf/click?upn=pEEcc35imY7Cq0tG1vyTt6zEs68RbcMfjPcajNHTKtkAchTKu1RCXrveUReTXBSBqntUCEEUQKuJ9AwcnMji1FmFWzvsTCg0tPUuSlVr1XLlPDNL15aFnBvVRzhIJJdjeeUZ7f0EfiRDSKpIIF1-2Bnn69qaauozUcASwLYO3t-2FoY76CTK8KLFr2NbEL8QqQ4xbpyt4IFYbNxPtMQSvoqZAxpt7hrBEbCYCpmZAHbBcGPgm4qtvVAtCG1i5CcNHHV2Q1gli25SSULXq5JmA0M3e-2FmIpHLsCBZxG7PaeICHmRkOTAjnyayMauPoov6FVJ5Y_I3fs1g-2F466j3y5fD5Q61KTvREVidSz8eOmcqorpVEXbt5H1apoGNSAiFy94udlsHIY1PSkmM0wPZ-2FQuxZTuFZgu91KBIfpRlrcG0GsOJ4QJw6d2kjZoEDqySkPD9JfoQR4x8dftwdKolQ4ZZhnCLU00hzm9Te-2BLUVI9N-2BFKFxE9xGs-2Fi6dbHV5zx4rYyljOYfm9TsfLjphnNNFfGSCwwnfXtl6uh2fnOlpiAkAvJpAQ-3D	
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=
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=

In the news

U.S.News & WORLD REPORT

When the Data's No Good

Lauren Camera · Sept. 14, 2018, at 6:00 a.m.

AP

Bill Gates calls for more global education assessments data

By: SALLY HO, Associated Press

Updated: Sep 18, 2018 - 10:24 AM

When 12-Year-Olds Can Breach School IT Systems, Who's Responsible?

By **Doug Levin** Sep 15, 2018

EdSurge

THE TECH EDVOCATE

WHAT ARE THE LONG-TERM EFFECTS OF EDTECH USE?

BY MATTHEW LYNCH / SEPTEMBER 18, 2018 / 0

Vectr

- Log into Vectr
- Change your profile from your UNI to your real name

  My Profile (top right) -> Preferences -> Display Name

Vectr Experiment

- Randomly assign weeks to be anonymous or non-anonymous

- Look at how student behavior changes on the platform

Why randomize?

- To guard against selection or unintended bias
- To create comparable groups (if the sample is large)
- Allows the calculation of probabilities

What is random?

- Must be unconnected to the event under study
  - Is this possible?
- True-random: some physical event
  - https://www.random.org/
- Pseudo-random: algorithm with a seed value

Random samples in R

- `set.seed()`
- `runif(n, min = 0, max = 1)`
- `sample(x, size, replace = FALSE, prob = NULL)`
- `rbinom(n, size, prob)`, `rnorm(n, mean = 0, sd = 1)`

Random samples for Vectr

```
> set.seed(123)

> coin <- c(“heads”,”tails”)

> sample(coin, 1, replace = FALSE, prob =
NULL)
```

Sequences in R

- `seq(from, to, by)`

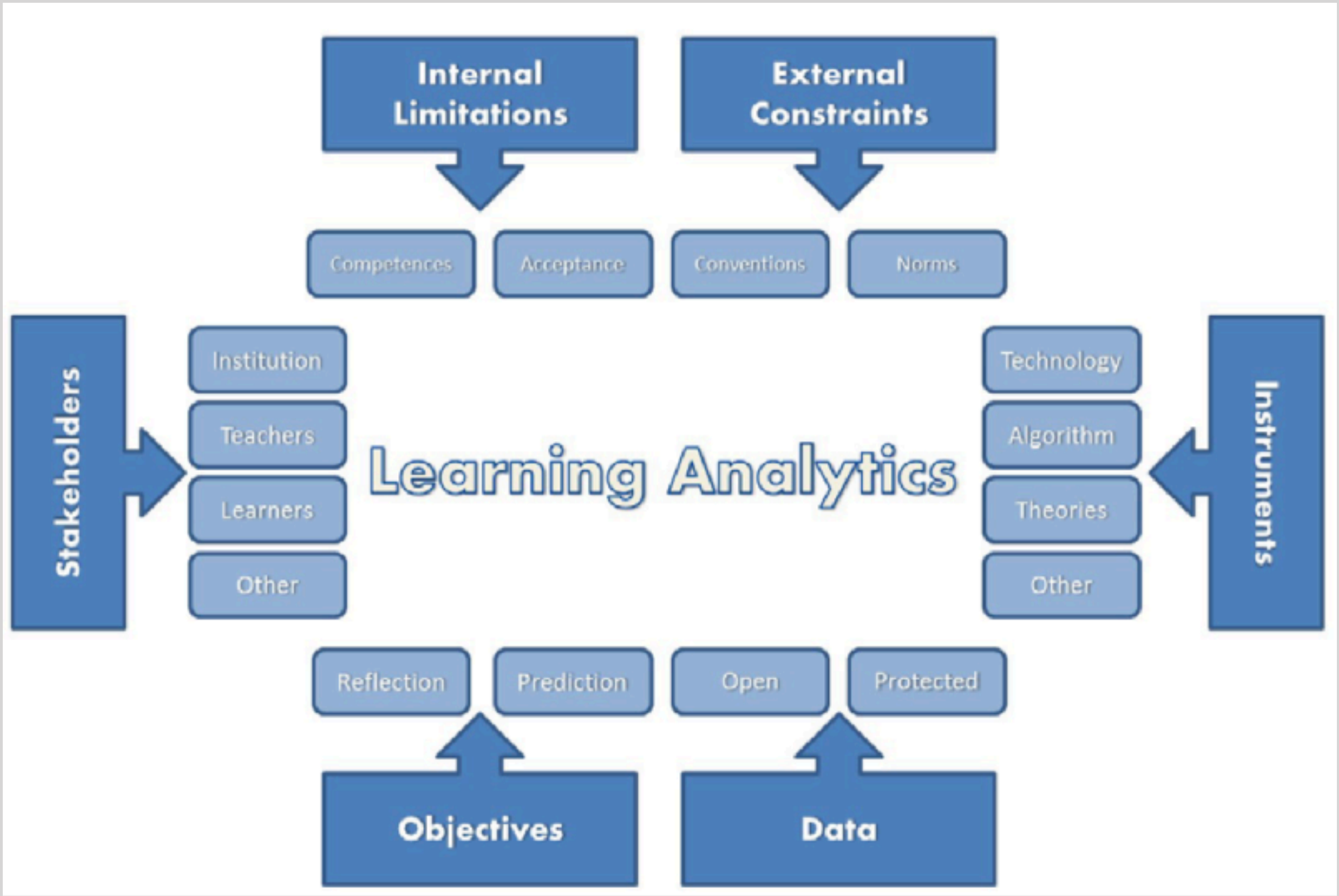
- Generates a sequence of numbers

Exercise 1

- Return to your educational goal: http://bit.ly/2flChcj
- What are you counting?
- Generate a sequence that represents those numbers in R
- Add some random noise to each number using one of the random number generators

Translating Learning Into Numbers

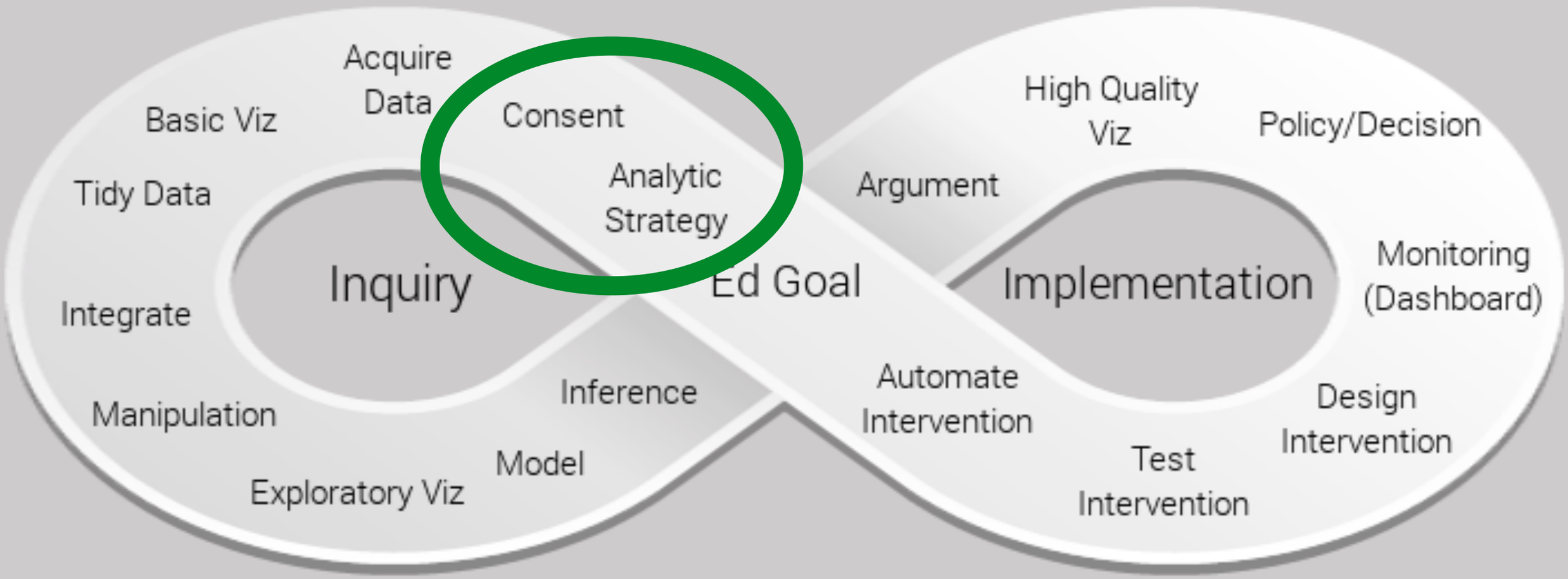
Greller & Draschler (2012)



Exercise 2

- Consider the fake data you generated and the diagram on page 44 of Greller & Draschler
- Work your way through each of the boxes
- Which would pose problems for you to actually acquire the data you want?
- Include these problems under “complications" in the spreadsheet

Ed Data Science Cycle



Code of Ethics

- There have been several Learning Analytics Codes of Ethics drawn up for institutions:
  - Open University
  - JISC
  - American Library Association
  - Data for Good

D — **DETERMINATION** – Why you want to apply Learning Analytics?
- What is the added value (Organisational and data subjects)?
- What are the rights of the data subjects (e.g., EU Directive 95/46/EC)

E — **EXPLAIN** – Be open about your intentions and objectives
- What data will be collected for which purpose?
- How long will this data be stored?
- Who has access to the data?

L — **LEGITIMATE** – Why you are allowed to have the data?
- Which data sources you have already (aren't they enough)?
- Why are you allowed to collect additional data?

I — **INVOLVE** – Involve all stakeholders and the data subjects
- Be open about privacy concerns (of data subjects)
- Provide access to the personal data collected (about the data subjects)
- Training and qualification of staff

C — **CONSENT** – Make a contract with the data subjects
- Ask for a consent from the data subjects before the data collection
- Define clear and understandable consent questions (Yes / No options)
- Offer the possibility to opt-out of the data collection without consequences

A — **ANONYMISE** – Make the individual not retrievable
- Anonymise the data as far as possible
- Aggregate data to generate abstract metadata models (Those do not fall under EU Directive 95/46/EC)

T — **TECHNICAL** – Procedures to guarantee privacy
- Monitor regularly who has access to the data
- If the analytics change, update the privacy regulations (new consent needed)
- Make sure the data storage fulfills international security standards

E — **EXTERNAL** – If you work with external providers
- Make sure they also fulfil the national and organisational rules
- Sign a contract that clearly states responsibilities for data security
- Data should only be used for the intended services and no other purposes

Draschler & Greller, 2016

Code of Ethics

bit.ly/HUDK4050COE

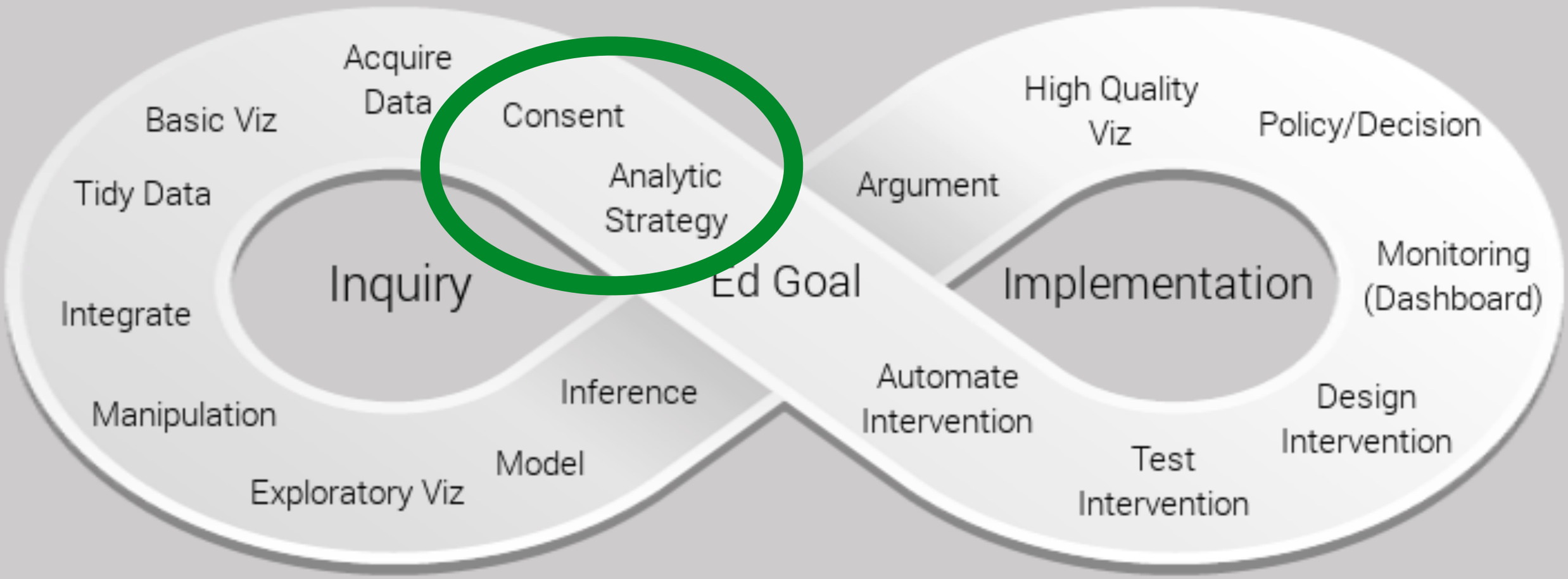
Exercise 3

- Read over the code
- Does it seem reasonable?
- Is there anything missing?
- Do you believe it is useful?

Anonymous Code of Ethics Survey

http://bit.ly/2w3GR51

Ed Data Science Cycle



Knowledge Check

http://bit.ly/2f7Ohxm