HUDK 4050: CORE METHODS IN EDM

Events

Title	Date - Time	Location
Our Tech Futures	09/21 - 10:00am	Columbia
AI in the Classroom	09/20 - 3:00pm	TC
Careers in Data Science in Ed	09/25 - 3:30pm	TC
AWS: A Simplified Approach to Data Driven Research	09/30 - 9:00am	Webinar
Formal & Informal LA (lunch)	10/4 - 11:00am	GDH 449
Cornell Tech: Day of Data	10/15	Cornell Tech

In the news

Women-Led EdTech Company Poised to Take on Language and Academic Tutoring Giants

seattlepi

Sensitive Personal and Financial Data of What's Likely an Entire Country Leaked Online

SLATE

POLITICO

Online education giant 2U to disclose more data

Forbes

Bridging The Gap: Connecting Education To Recruiting

Stack Overflow's annual report shows that nearly 70% of all software engineers are at least partially self-taught, and almost 40% receive on-the-job training.
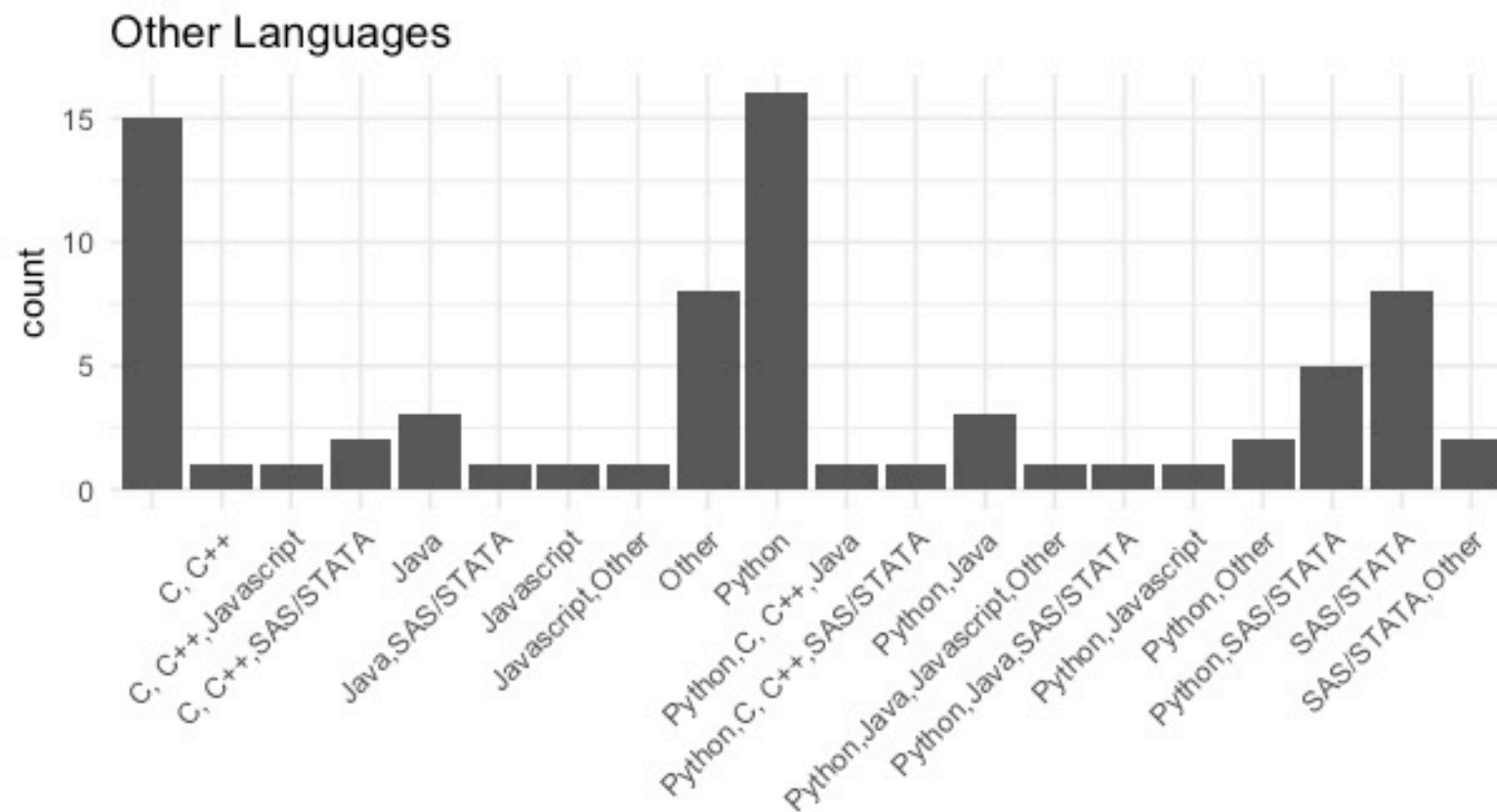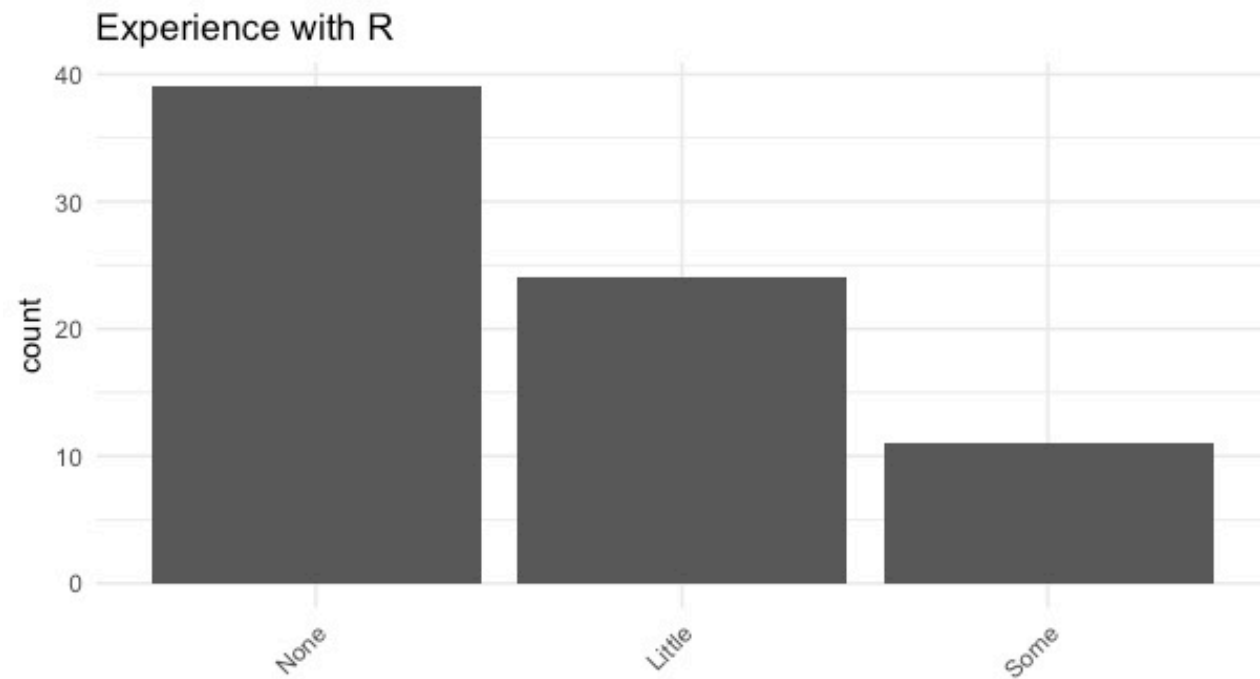
Where to find help

- Help window in R studio or help("command")

- Google/DuckDuckGo

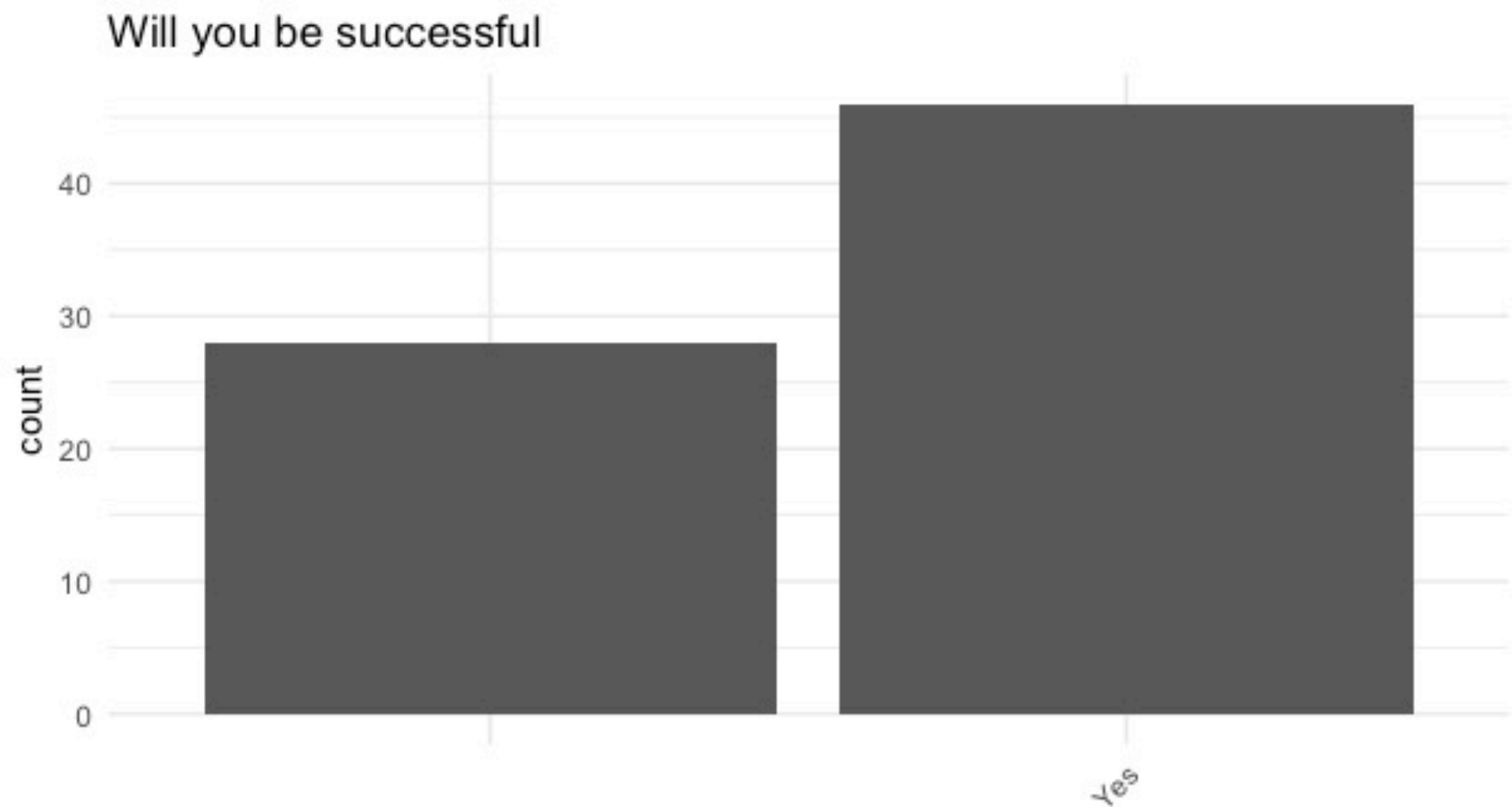- Stack Overflow/Stack Exchange

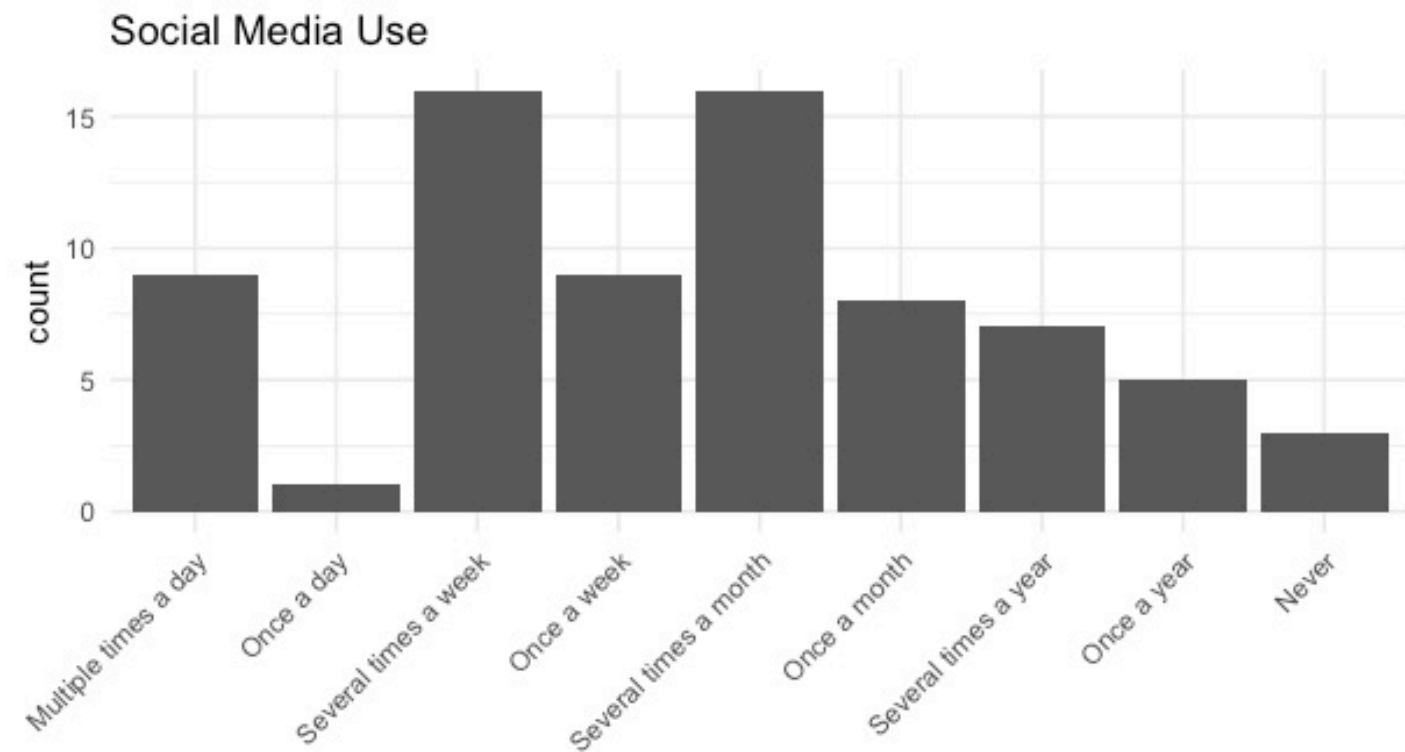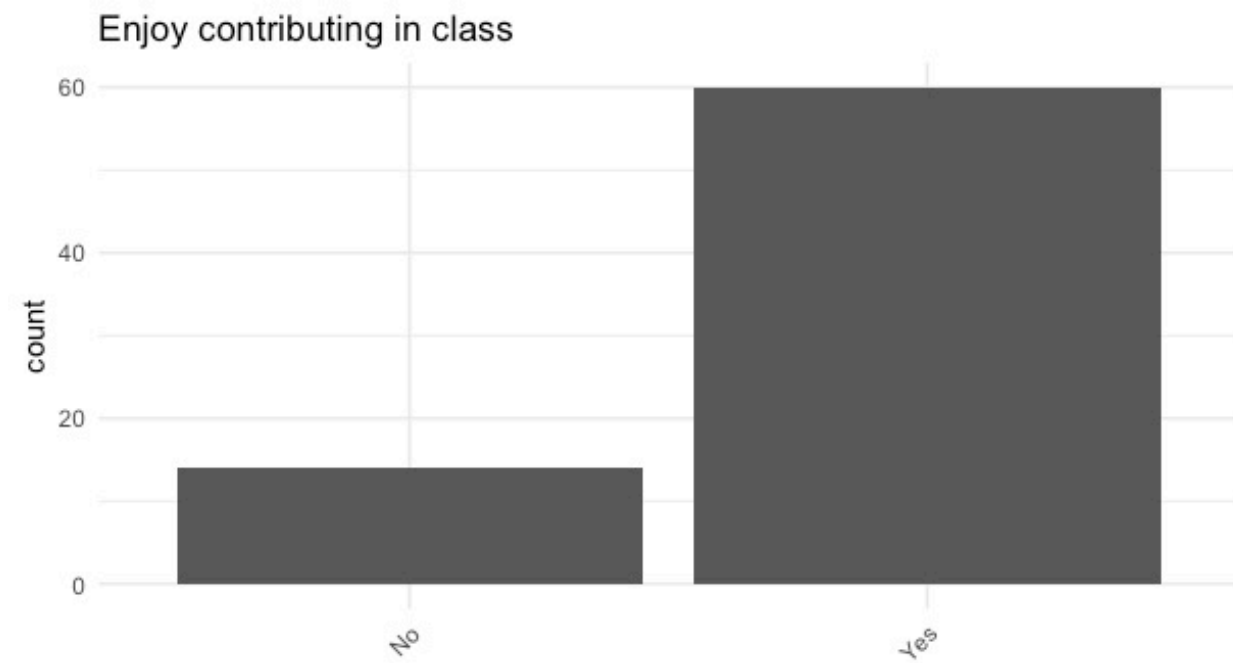- CAs & me

Class Stats
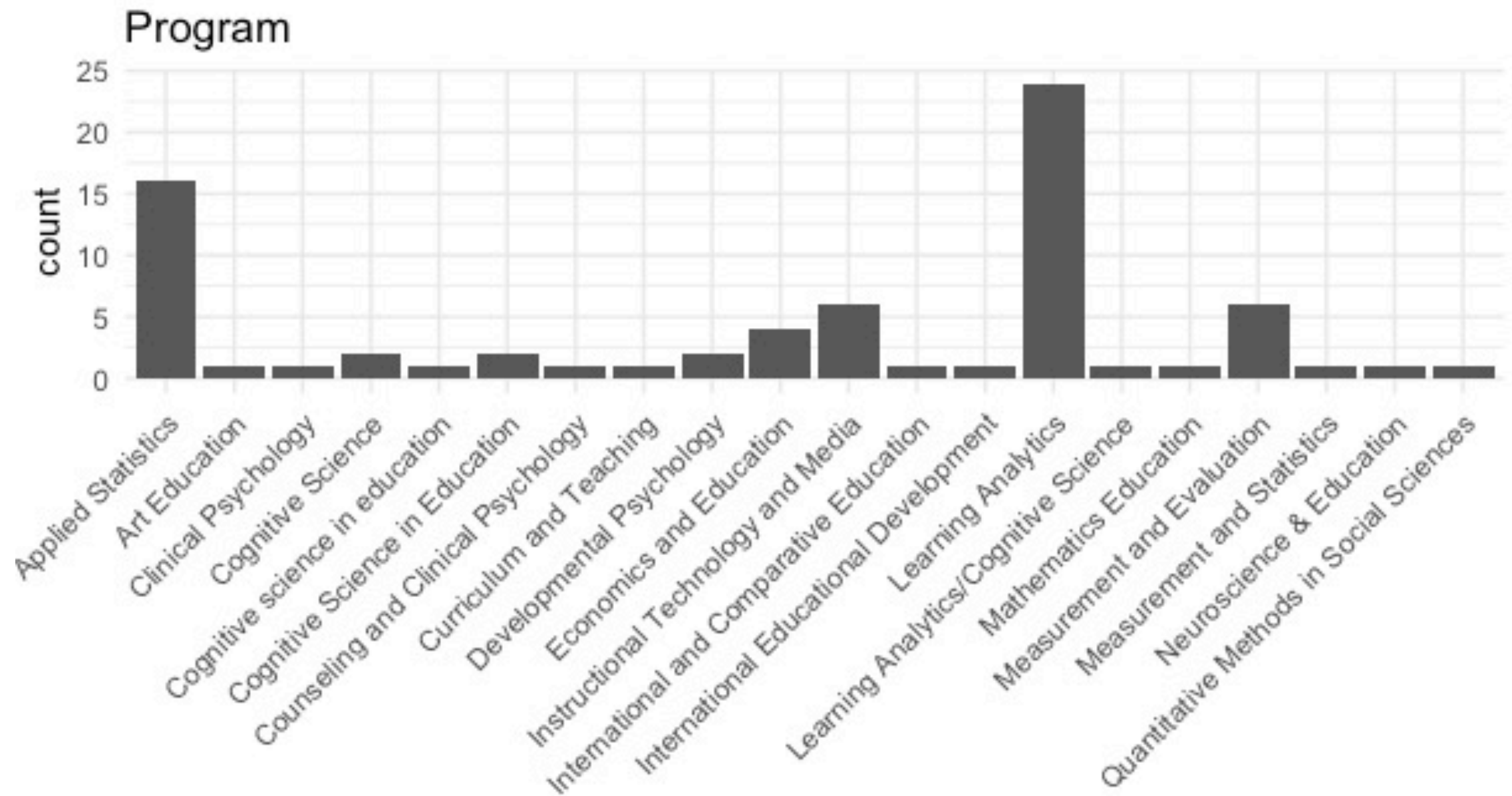
Class Stats

Class Stats

Class Stats

Zotero

- Install Firefox & Zotero (Assignment 1)

- Download the HUDK2017.rmd file from the syllabus GitHub repo

- Upload into Zotero folder

- Write a note within one of the articles you have read

- Right (cmd) click on the folder

- Choose "Export Collection…"

- Choose CSV from the drop down menu and choose a location you know the file path for

- Open R and type the code:

DF <- read.csv("FILE PATH WHERE YOU PUT CSV FILE", header = TRUE)

e.g.. DF <- read.csv("~/Desktop/HUDK2017.csv", header = TRUE)

- Behold! Your bibliography including your notes (under the "notes" column)

Randomization

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?

- <u>True-random</u>: some physical event

  - https://www.random.org/

- <u>Pseudo-random</u>: 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

```
> 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 from the first class (or invent one if you weren't there)

- 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)
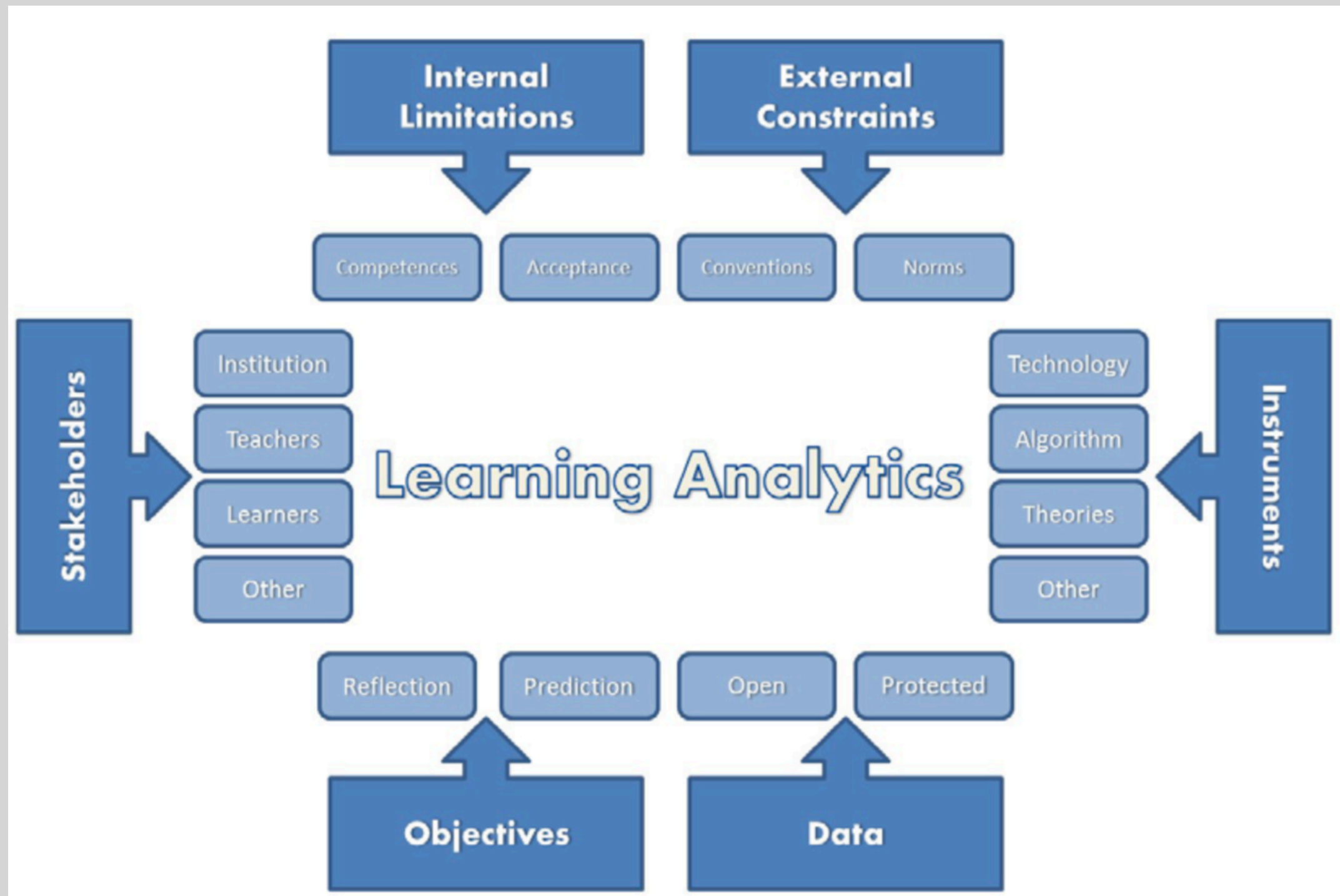
Exercise 2

- 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?

- Write a note summarizing the article in Zotero, include thoughts based on your answers to the above

Ed Data Science Cycle
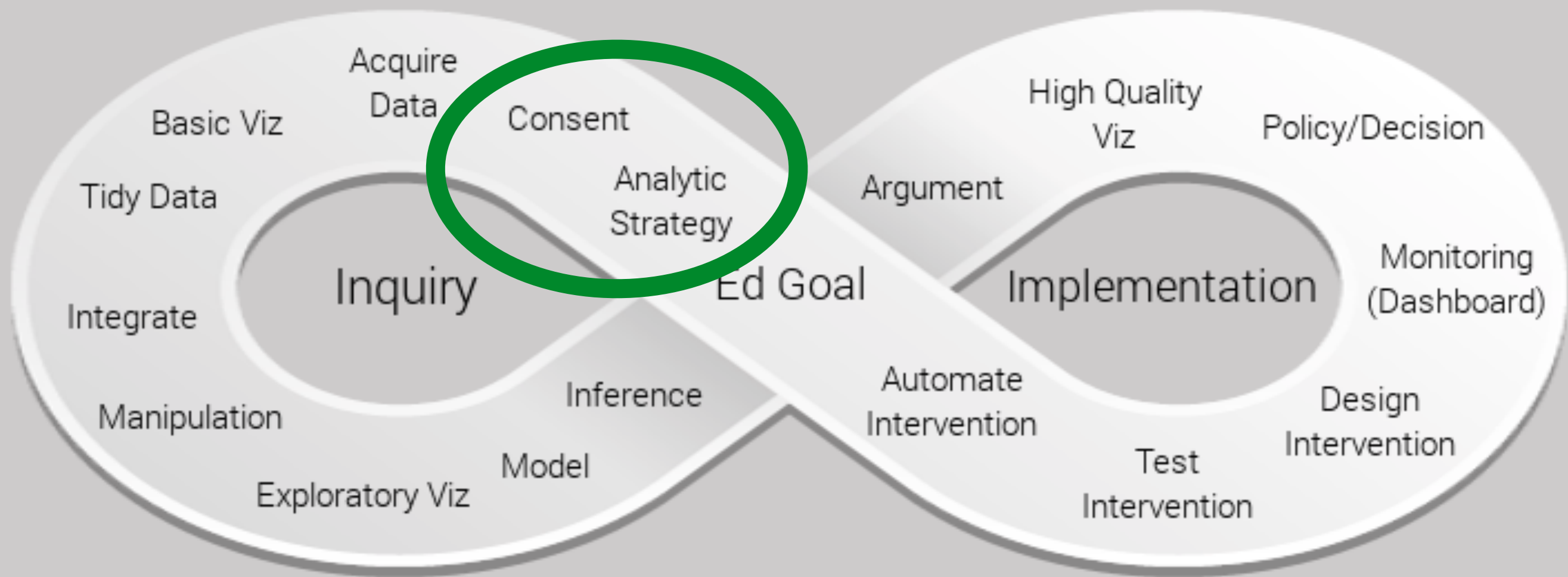
Code of Ethics

- 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