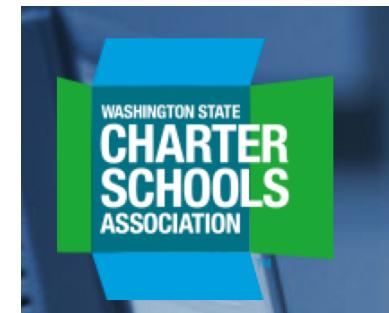


EDCT GE2550: DATA SCIENCE IN EDUCATION

Big Data, Learning Analytics & The Information Age

2/25/16 12:42 PM

In the news



**Symposium Examines
Technology, Privacy, and
the Future of Education -
March 4**

[Blog](#) | [News](#) | [Press Releases](#) | [Updates](#) | February 8, 2016

Washington's Public Charter Schools on Path to
Successful First Year Despite Threat of Closure
Following State Supreme Court Ruling
Early Data Indicates Rapid Gains in Student Achievement

Today

In the news 6:45 - 6:50

Quiz 6:50 - 7:00

Self guided learning 7:00 - 7:10

Markdown & Basic Viz 7:10 - 7:20

Basic Viz with Class Data 7:20 - 7:30

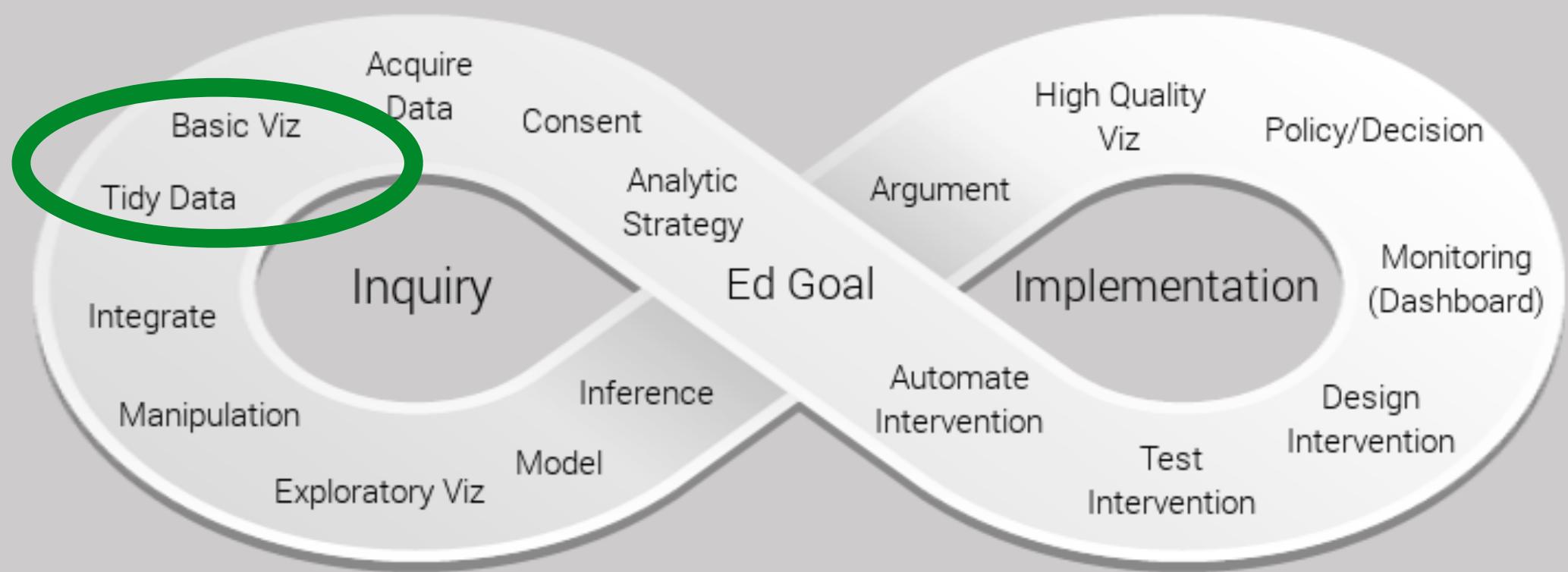
Tidy Data 7:30 - 7:40

Github Viz 7:40 - 7:50

Cloning & Plotting Twitter 7:50 - 8:00

What to do with Twitter 8:00 - 8:25

Ed Data Science Cycle

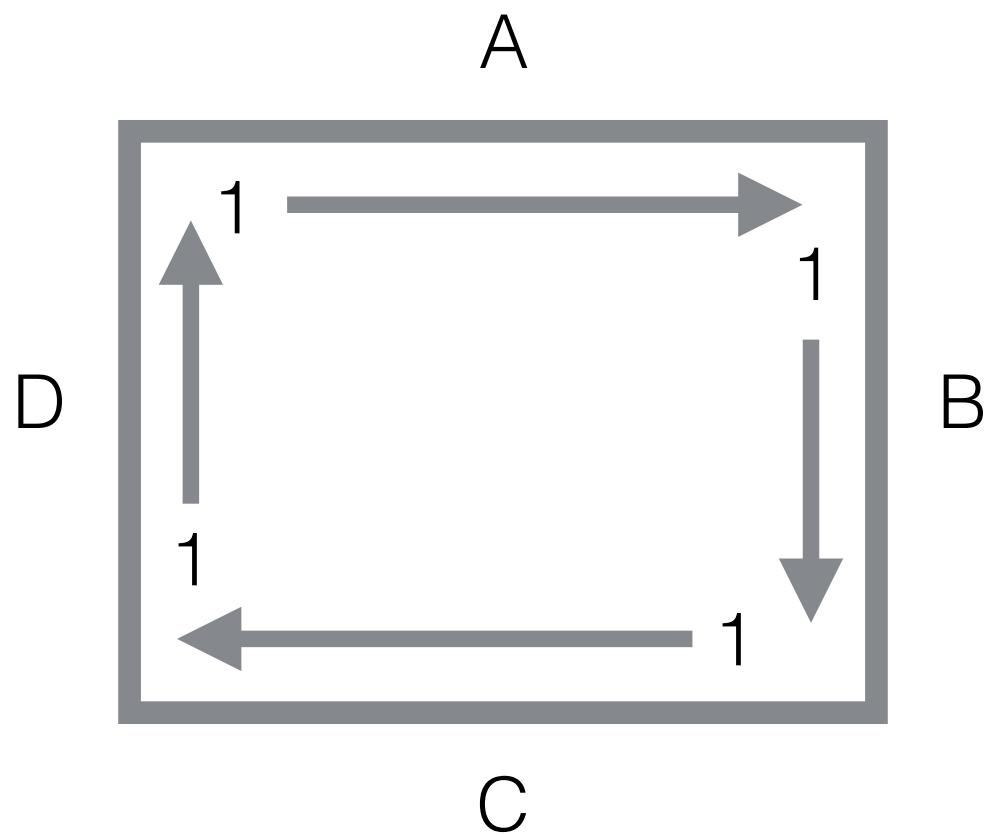


Quiz

bit.ly/1QsINaw

Table coordinates

— Projection



Available Variables (Take II)

All Anonymous

Demographics

- Gender, other classes, current degree, work, educational background

Environmental

- Seat & table location, time of day you do homework, comfort with group work

Learning Data

- Assignment, quizzes, interactions with Github, Tweets, reading notes, concerns about course

Troubleshooting

Course Goals

Short Term		Long Term	
<i>Content</i>	<i>Skills</i>	<i>Abstractions</i>	<i>Habits</i>
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 (trouble- shooting)

Troubleshooting

- Find an expert (phone a friend)
- Web search
- Crowd source
 - Within community (EG - Twitter)
 - StackOverflow



How to ask a good question:

- Do an exhaustive search
- Format correctly (EG - indent code)
- Include a “reproducible example”
- Don’t be intimidated

Tidy Data

(#notidynodessert)

Data Frames & Vectors

Why is tidy data?

- Difference between “clean” and “tidy”
- Data comes in a lot different structures, some which are difficult to analyze
- We want to make them manageable
- We want them to be “intuitive” to R (vectorized)
- BUT we want to keep a very precise record of how we did that

What is tidy data?

1. Observations are in rows
2. Variables are in columns
3. In a single data set

But...?

- What is a variable?
- What is an observation?
- What goes where in a data matrix?

Wide Format

- Repeated measures are in a single row

Student	Quiz 2-1-16	Quiz 2-10-16	Quiz 2-20-16
Francis	10	10	11
Alex	14	15	18
Kaji	11	17	14
Miriam	8	10	8

Long (Narrow) Format

- Each row is one time point per subject

Student	Quiz	Date
Francis	10	2-1-16
Francis	10	2-10-16
Francis	11	2-20-16
Alex	14	2-1-16

Generalize

Male	Female
4	10
7	10

How many variables are in the above matrix?

1. Male
2. Female
3. Count

Types of Messiness

- Column headers are values, not variable names
- Multiple variables are stored in one column
- Variables are stored in both rows and columns
- Multiple types of experimental unit stored in the same table
- One type of experimental unit stored in multiple tables

Tidy System

- There are many commands and several packages for doing this in R
- We are going to try to stick to two: `tidyr` & `dplyr` (we may end up using more)
- Reshape, Subset, Variable generation, Combine, Summarize

Reshape

- Similar to generating pivot tables
- Long format \leftrightarrow Wide format

Student	Quiz 1	Quiz 2	Quiz 3
Francis	10	10	11
Alex	14	15	18
Kaji	11	17	14
Miriam	8	10	8

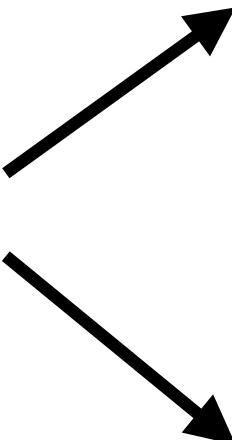


Student	Quiz	Date
Francis	10	2-1-16
Francis	10	2-10-16
Francis	11	2-20-16
Alex	14	2-1-16

Subset

- Splitting data frames

Student	Quiz	Date
Francis	10	2-1-16
Francis	10	2-10-16
Francis	11	2-20-16
Alex	14	2-1-16



Student	Quiz	Date
Francis	10	2-1-16
Francis	10	2-10-16
Francis	11	2-20-16

Student	Quiz	Date
Alex	14	2-1-16

Variable Generation

- Create new variable from current variables

Student	Quiz 2-1-16	Quiz 2-10-16	Quiz 2-20-16		mean
Francis	10	10	11		10.3
Alex	14	15	18		15.7
Kaji	11	17	14		14
Miriam	8	10	8		8.7

Combine

- Merge and bind dataframes
- Mutate or Filter

Student	Quiz 2-1-16	Quiz 2-10-16
Francis	10	10
Alex	14	15
Kaji	11	17

+

Student	Quiz 2-1-16	Quiz 2-20-16
Francis	10	9
Suchi	14	5
Kaji	11	10

Summarize

- Collapse data into a limited number of values according to a function

Student	Quiz 2-1-16	Quiz 2-10-16
Francis	10	10
Alex	14	15
Kaji	11	17

→

Av(Score/ Quiz/ Student)
12.8