Visualizing Learner Trajectories and Engagement to Improve Teaching & Learning

Michael Ginda[1], Michael C. Richey[2], Mark Cousino[2], Katy Börner[1]

[1]Cyberinfrastructure for Network Science Center
School of Informatics, Computing and Engineering
Indiana University
Bloomington, IN, USA
http://cns.iu.edu

[2]The Boeing Company
Everett, Washington, USA

Learning analytics in support of convergence

Learning Analytics & Knowledge Conference
Tempe, Arizona
March 5, 2019

Presentation Outline

CNS Cyberinfrastructure for Network Science Center

- Project Background and Motivation
- Understanding the Data
- Data Processing, Analysis and Visualization Pipeline
- Visualizing Course Structures
- Visualizing Student Engagement
- Visualizing Learner Trajectories
  - Related Work and Design Process
  - Demo the Learner Trajectory Network Visualization
- Future Development Efforts

PDF of Slides

- https://cns.iu.edu/presentations.html

Paper

Michael Ginda, Michael C. Richey, Mark Cousino, Katy Börner, *Visualizing Learner Engagement, Performance, and Trajectories to Evaluate and Optimize Online Course Design*, submitted to Plos One.

GitHub Repos

- edX Student and Course Analytics and Visualization Pipeline (R scripts) https://github.com/cns-iu/edx-learnertrajectorynetpipeline
- GitHub Learner Trajectory Network Project Repository (Visualization) https://github.com/cns-iu/learning-trajectories

- **Project Background and Motivation**
- Understanding the Data
- Data Processing, Analysis and Visualization Pipeline
- Visualizing Course Structures
- Visualizing Student Engagement
- Visualizing Learner Trajectories
  - Related Work and Design Process
  - Demo the Learner Trajectory Network Visualization
- Future Development Efforts

"Improving Return on Investment in Education: Measuring, Visualizing, and Optimizing Learner Trajectories"

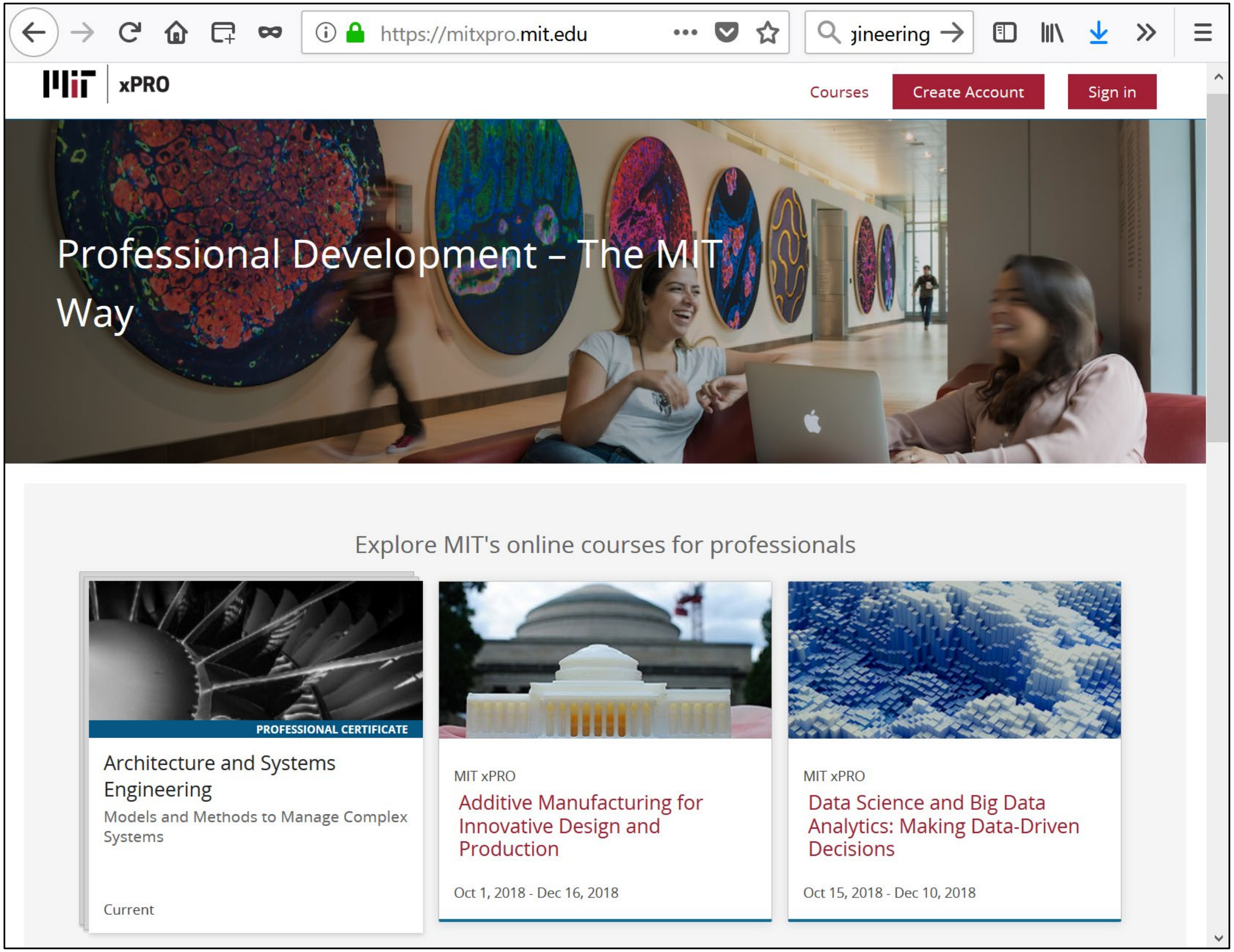
Michael C. Richey, Michael Ginda, Mark Cousino, Katy Börner

Fall/Winter 2017 our team began working with The Boeing Corporation to leverage our expertise in visual analytics to study data produced by students in online courses to understand

- the relationship between students interactions of courses resources and
- Student trajectories over the course, and
- The impact of both on student performance.

For this project, we were provided with data generated by Boeing professionals taking online courses that were designed and run on the MITx Pro platform in collaboration with Boeing.

MITxPro: Boeing Continuing Education



MITxPro: Boeing Continuing Education (cont.)

CNS Cyberinfrastructure for Network Science Center

MITxPro: Architecture of Complex Systems

> “Understanding and managing system complexity is a critical challenge today as systems continue to grow in scale and complexity. This course is designed to help engineers address changes which induce, propagate, and amplify risk in the increasingly complex products and services they are required to develop. Students will get a solid grounding in complex systems, analysis of complex systems, and complexity management.”

Analyzing course design, and performance and engagement of 3 cohorts of students between Fall 2016 and Spring 2017.

MIT, Boeing, NASA, and edX to launch online architecture and systems engineering program

Four-course program will train professionals in latest practices on models and methods to manage complex systems

For this presentation, when the Architecture of Complex System course is referenced in a visualization, we are using data from the Fall 2016 instance of the course.

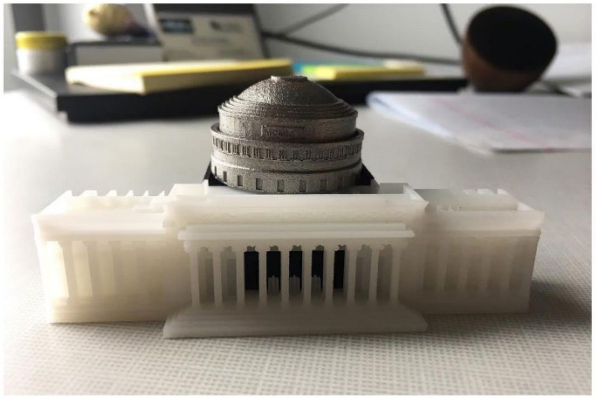
1,611 Boeing engineers registered; 1,565 were active and generated nearly **31 million click event records** while accessing videos, projects, and assessments. Some students generated over 100,000 separate events.

All but 255 engineers passed the course, resulting in a completion rate of 84.1%.

MITxPro: Additive Manufacturing

> "MIT faculty and industry experts in addressing the full spectrum of AM technologies, and connect the fundamentals of AM to its applications and business potential. Walk away with the knowledge and confidence to architect and implement innovative uses of AM across the product life cycle.
>
> You will also learn how to design parts for AM, leveraging advanced CAD, generative design, and process planning software. The course concludes with an in-depth case study, where you will solve a real-world design or business strategy problem using your new knowledge of AM. The course also describes a wide range of value-driven applications of AM, which are described according to value proposition and demonstrated using proven industry examples."

Analyzing course design, and performance and engagement from the first run of the course, in collaboration with partners at IU School of Education and University of Pennsylvania.



- Project Background and Motivation
- **Understanding the Data**
- Data Processing, Analysis and Visualization Pipeline
- Visualizing Course Structures
- Visualizing Student Engagement
- Visualizing Learner Trajectories
  - Related Work and Design Process
  - Demo the Learner Trajectory Network Visualization
- Future Development Efforts

MITxPro Courses

MITx courses are run on a *customized* open edX infrastructure.

- See *edX Research Guide* for data documentation. Note that database architecture, and data formats changed over the study periods.
- EdX was extended using LTI compliant systems:
  - Qualtrics to administer survey instruments.
  - Discussion forum systems, like Piazza, Yellow Dig, and Slack.

edX Course Data

The course database describes the course structure and user profiles.

- Basic demographics – Age, Gender, Education Level
- Enrollments, certificate status, grades, roles*, module interaction history;

Courses have a 5-level hierarchical structure

- Course, chapters, sequences, vertical pages, and content modules
- Content may include: videos, webpages, problem questions, open assessments, interactive visualizations, discussions

Course have a linear design that follow weekly schedule of materials and activities for students to complete.

edX Tracking Data

Student tracking data is captured as daily log files

- includes data for the enrollment and post course periods
- Interactions are tracked at multiple levels of the course hierarchy

Student logs include three types of interactions

- Student generated browser and mobile events
- edX LMS system server events

Logs may will include interactions with non-content resources such as OneShape.

- Project Background and Motivation
- Understanding the Data
- **Data Processing, Analysis and Visualization Pipeline**
- Visualizing Course Structures
- Visualizing Student Engagement
- Visualizing Learner Trajectories
  - Related Work
  - Demo the Learner Trajectory Network Visualization
- Future Development Efforts

The current data processing and analytics pipeline is scripted using the R statistical programming language.

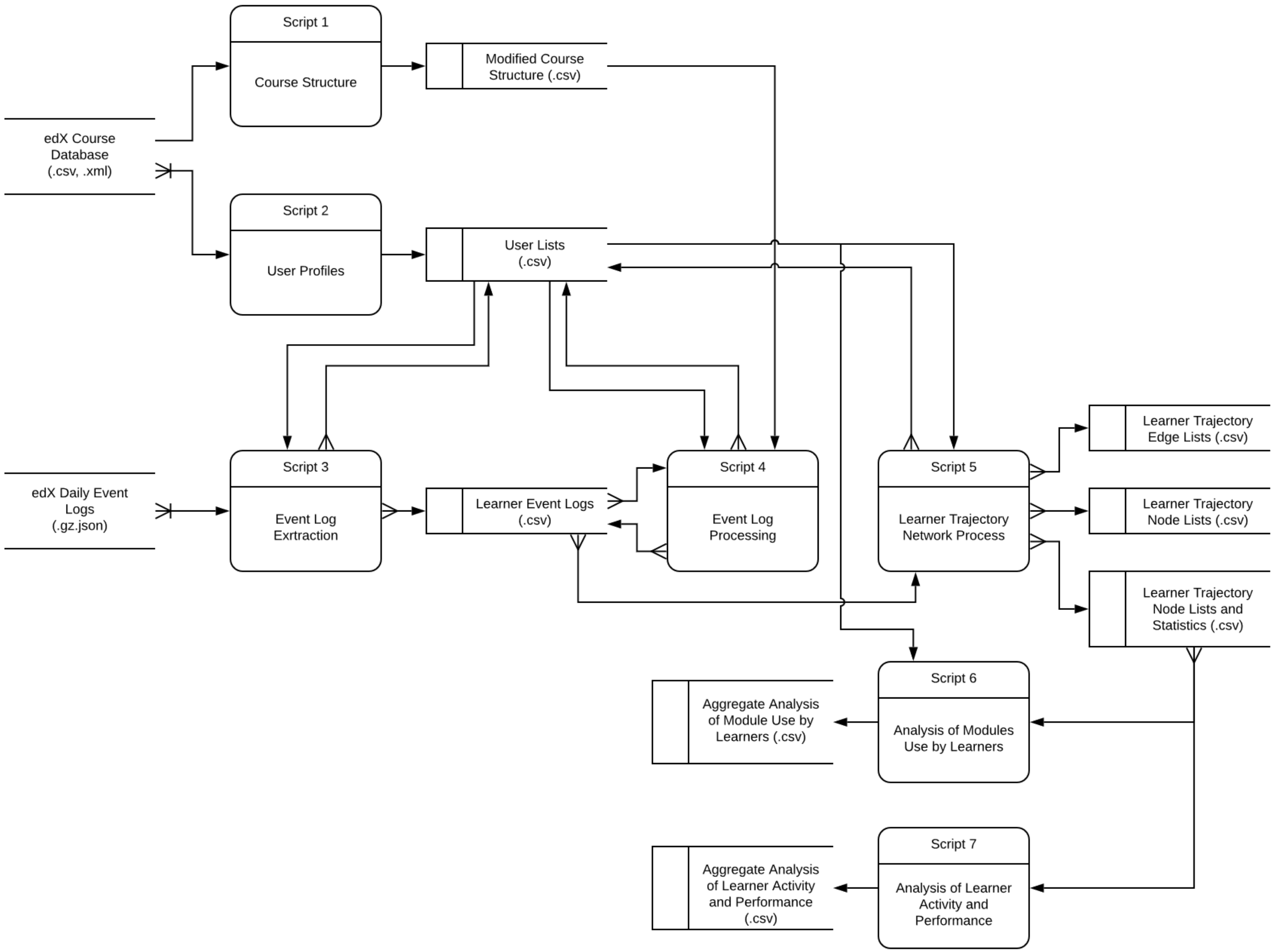
The scripts leverage a variety of packages, particularly from Hadley Wickham's Tidyverse, including:

- plyr/dplyr – data aggregations;
- stringr – string manipulation and regular expressions; &
- ggplot2 – statistical visualizations

Network visualization were computed using Gephi v. 0.8.2.

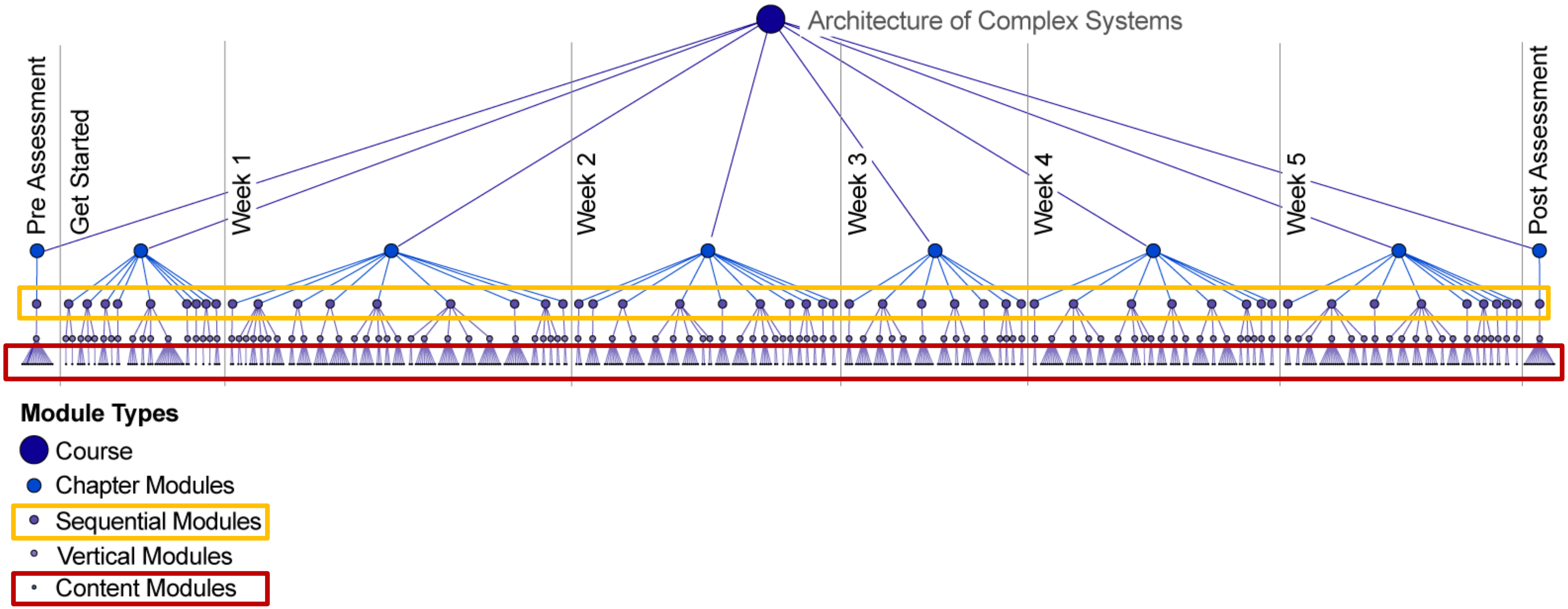
GitHub - edX Student and Course Analytics and Visualization Pipeline scripts https://github.com/cns-iu/edx-learnertrajectorynetpipeline

R Processing Pipeline – Dataflow Diagram



- Project Background and Motivation
- Understanding the Data
- Data Processing, Analysis and Visualization Pipeline
- **Visualizing Course Structures**
- Visualizing Student Engagement
- Visualizing Learner Trajectories
  - Related Work and Design Process
  - Demo the Learner Trajectory Network Visualization
- Future Development Efforts

Visualizing the Course Structure



Course Structure Tree Diagram shows 5-level hierarchical structure of the *Architecture of Complex Systems* course.

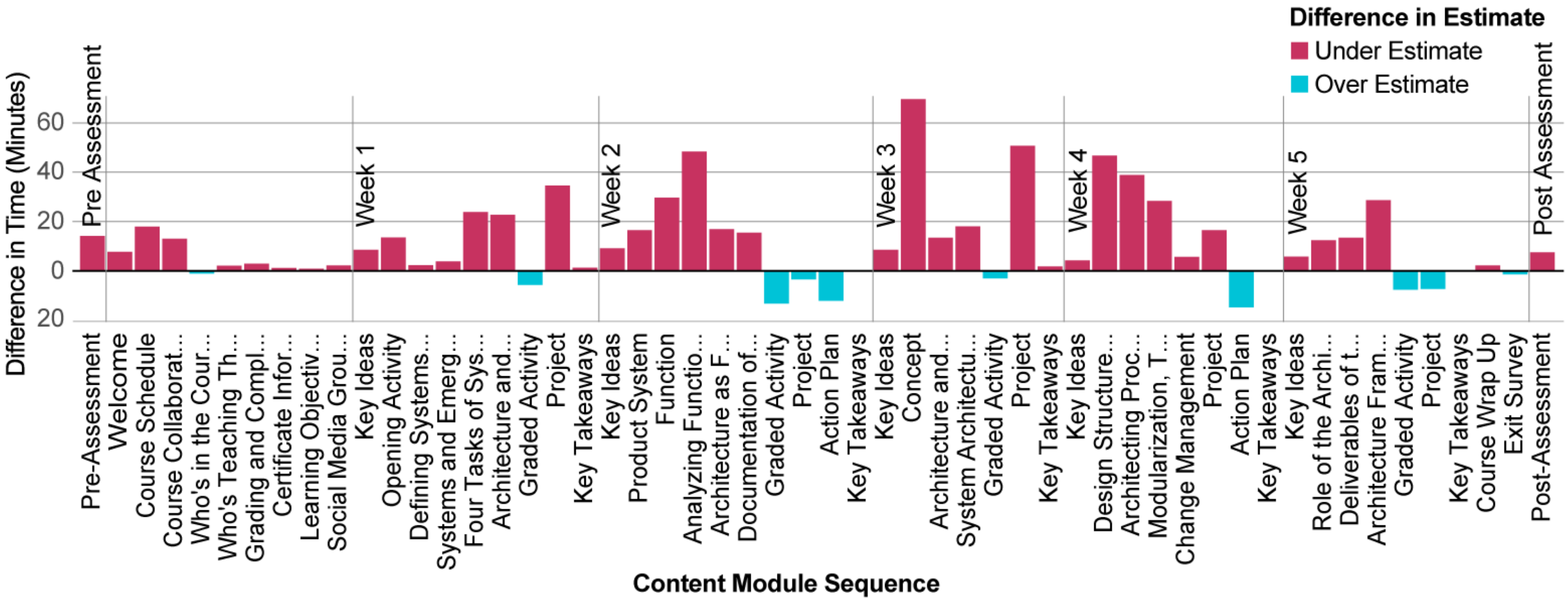
Nodes are ordered based on the sequence of learning modules presented to learners in the course.

Insights: Course structure allows for analysis and visualizations at multiple levels of granularity, temporality. We can also see that courses share similar lengths in modules presented to students.

- Project Background and Motivation
- Understanding the Data
- Data Processing, Analysis and Visualization Pipeline
- Visualizing Course Structures
- **Visualizing Student Engagement**
- Visualizing Learner Trajectories
  - Related Work and Design Process
  - Demo the Learner Trajectory Network Visualization
- Future Development Efforts
- Opportunities for Research and Collaborations

Difference in Instructor Estimates vs Avg. Time Spent

CNS Cyberinfrastructure for Network Science Center



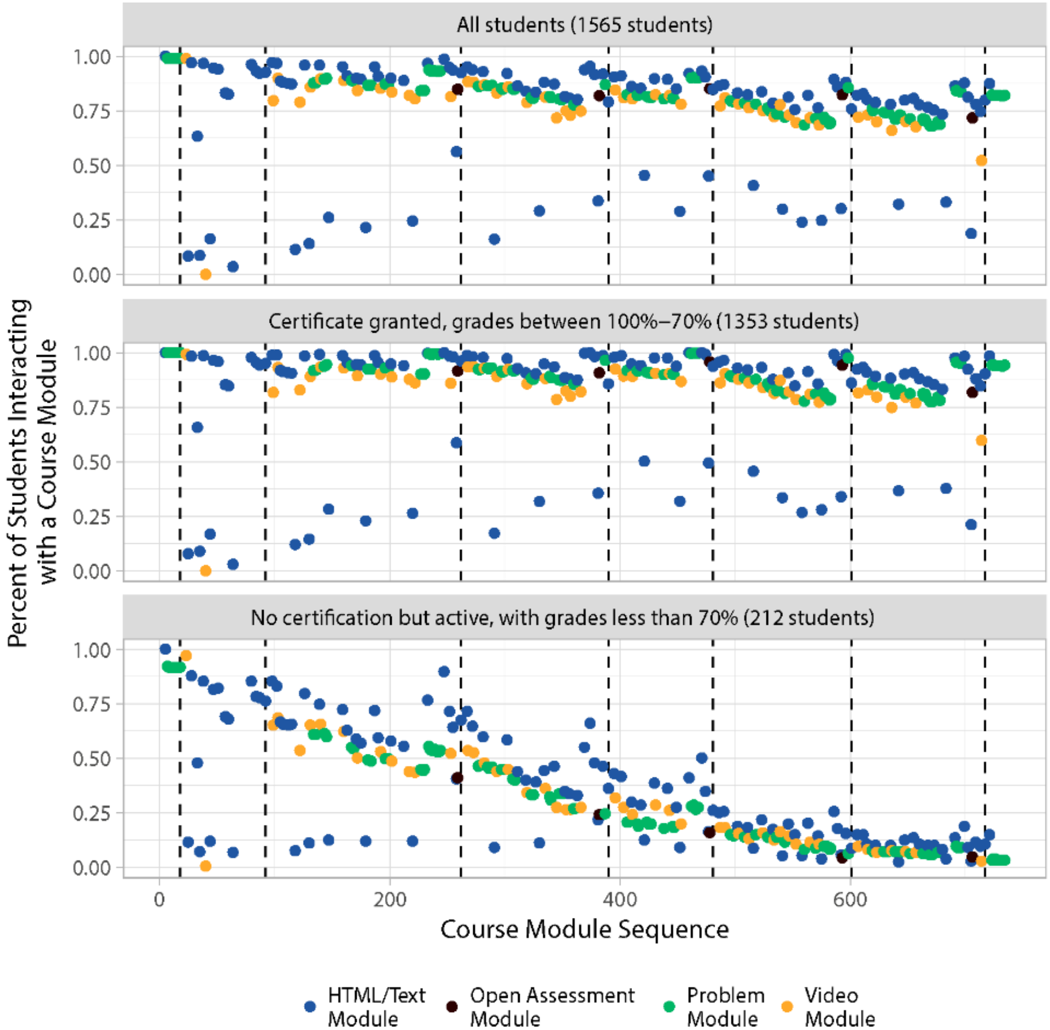
Instructors Temporal Predictions are represented in a temporal bar graph that compares course instructors *estimated* time learners would need to complete course materials, and the average time taken by learners in the course computed from data.

Insights: Instructor's temporal estimates were only made for chapter level at the sequential level (yellow box in slide 16). Instructors provided accurate estimates of time for course assessments, but did not account for studying activity of students in their estimates.

Percentage of Students Interacting with a Course Module.

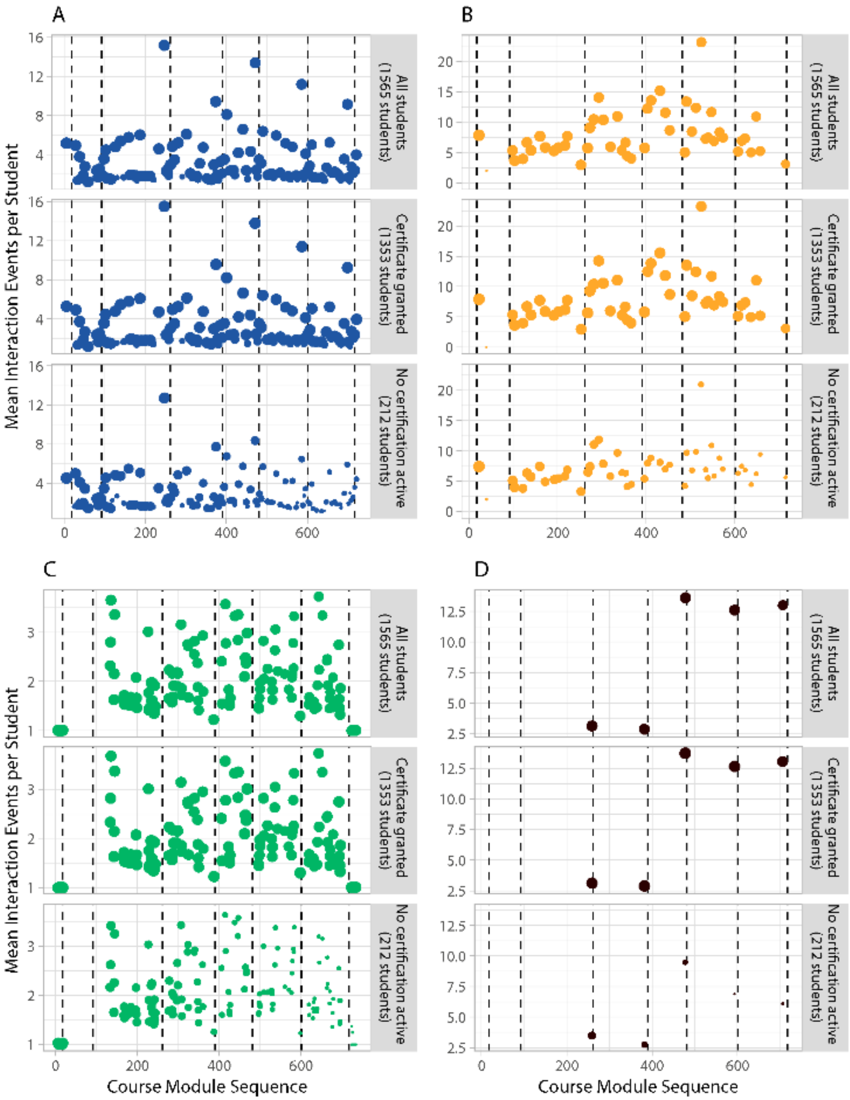
A scatter graph looks at the percentage of the learners in the *Architecture of Complex System* course accessing modules by certificate group and module type.

Insights: There is a clear difference in access patterns by students across the course by certificate and non certificate earners, as well as subtle differences between module types. Most notably, few of the students that do not earn a certificate do access the Open Assessment Modules.



Data from Architecture of Complex Systems is used in these visualizations.

Comparing Interactions by Module Type



Mean Interactions by Certificate Group and Content Type.

The visualization uses a scatter graph to visualize the mean number of interaction events for content modules along the y-axis, and by sequence order along the x-axis.

Points are sized based on the number of students that interacted with a module. The figures are split by certificate group and module type.

A. HTML Modules
B. Video Modules
C. Problem Modules
D. Open Assessment Modules

Insights: The patterns of access by both type of module in the course and by the different learner cohorts. Each type of module has a distinct range of values along the Y axis, and the number of points in each set of visualization vary.

Data from Architecture of Complex Systems is used in these visualizations.

Learner Engagement and Performance

CNS Cyberinfrastructure for Network Science Center

We use scatter graphs to visualize statistical features calculated for each student based on analysis of their course event logs.

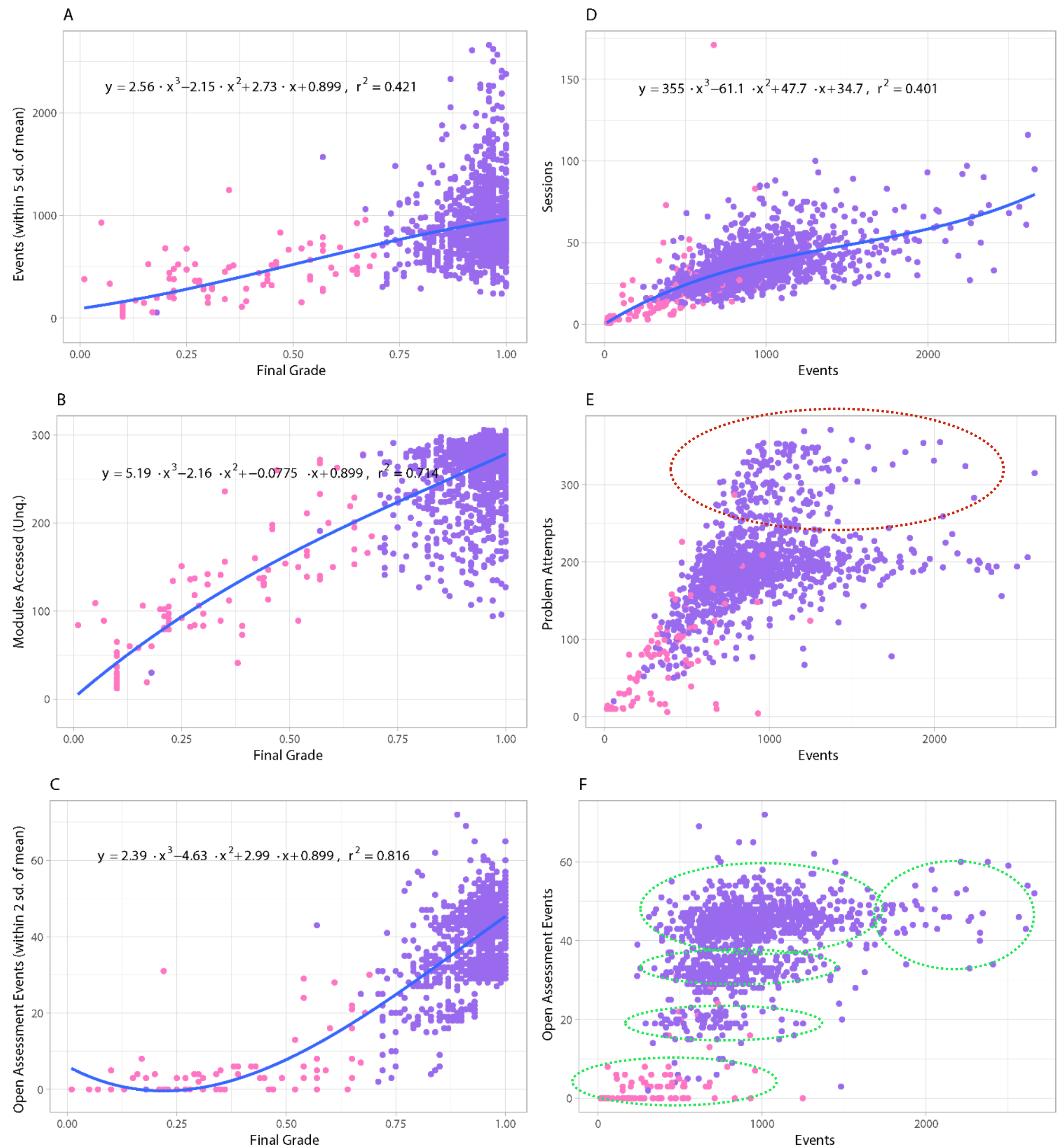
Regression analysis was performed where relevant; however, in plots E and F, clustering within the data make regressions a poor fit analysis.

Figures:

A. Events x Final Grade
B. Unique Modules Accessed x Final Grade
C. Assessment Events x Final Grade
D. Sessions x Total Number of Events
E. Problem Attempts x Total Number of Events
F. Assessment Events X Total Number of Events

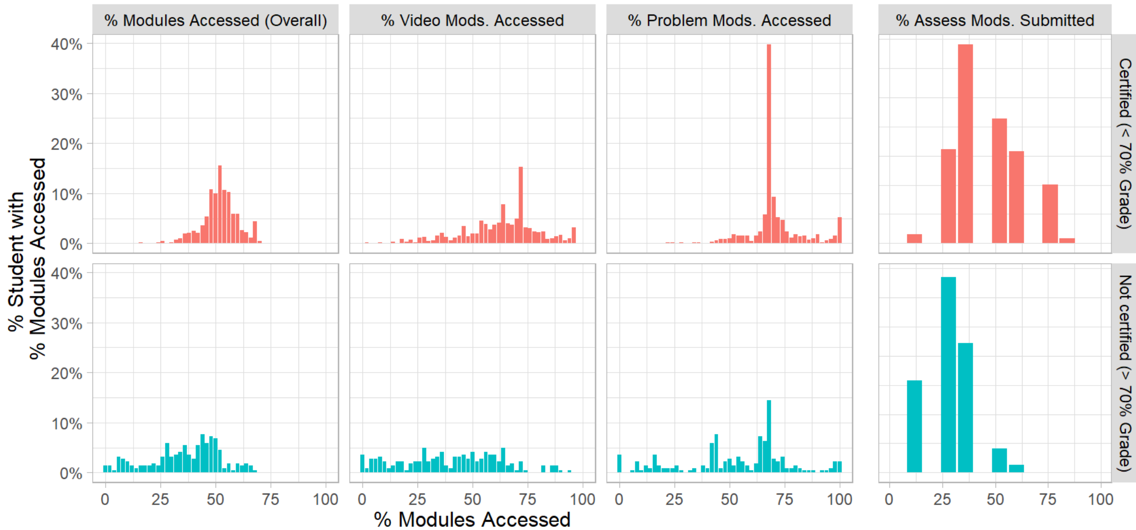
Insights: Each of these scatter plots reveals a relationship between engagement and performance in the course.

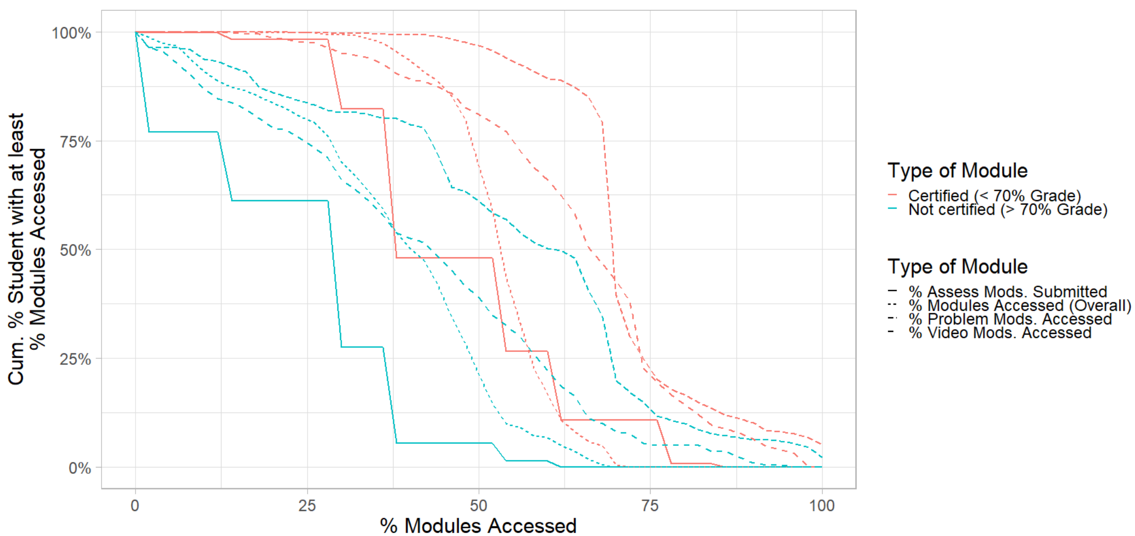
In the cases of problem modules and assessments, distinct behaviors emerge due to properties in the course structure (i.e. limited use of assessments) and student engagement (i.e. a subset engages in more attempts on average then other students.).



Data from Architecture of Complex Systems is used in these visualizations.

Distribution by % of Course Module Access





Data from Architecture of Complex Systems is used in these visualizations.

Visualizing Student Resource Access to compare student access patterns based on certificate status and the type of module accessed by student.

Bar graphs are used to show the distribution of students based on the percentage of resources that they accessed/ submitted.

Line graphs show the cumulative distribution of percentage of module access by students, by certificate group and module type.

Insights: Students that do not earn a certificate have very different access pattern distributions.

- Project Background and Motivation
- Understanding the Data
- Data Processing, Analysis and Visualization Pipeline
- Visualizing Course Structures
- Visualizing Student Engagement
- **Visualizing Learner Trajectories**
  - **Related Work and Design Process**
  - **Demo the Learner Trajectory Network Visualization**
- Future Development Efforts

Research Questions and Insight Needs

- How do learners use course materials and activities in online courses … *over time*?
  - Visualize patterns of access
  - Visualize movement across resources
    - Sequence of interactions
    - Number of interactions
    - Amount of time spent with a resources
- What patterns of activity are found across a cohort, and are they linked to learning strategies or prior knowledge?
  - Personality, job titles and skills, prior knowledge
- Do learners identified in different groups access and engage with resources differently?
  - Compare differences in access and trajectory patterns for different students and cohorts.
  - Based on aggregate access statistics or sequential/network measures.

Use Cases for Learner Trajectory Networks

Early work to visualize learner transition between types of resources used as state transitions networks that were based on learners' education software audit trails. An overview of recent work in this space follows.

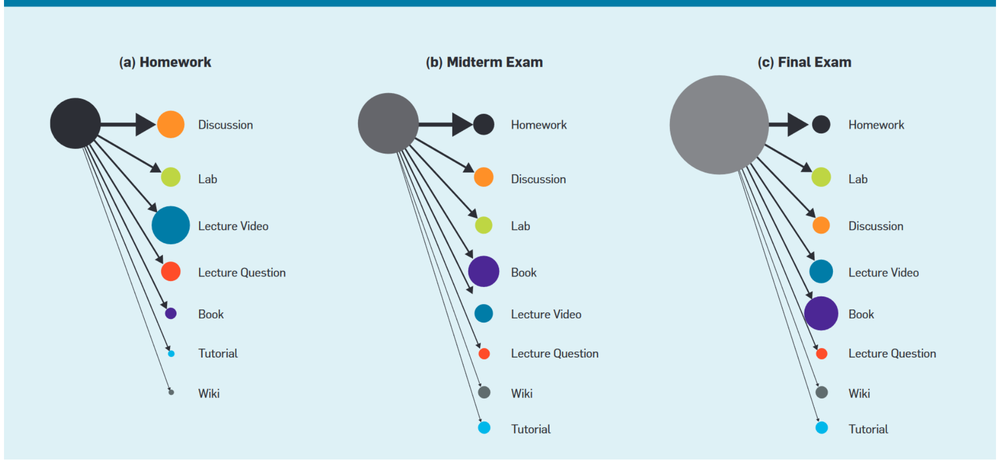
Learning trajectory analytics results can reveal patterns of engagements to data mining and visualization experts:

- learner cohorts that are are 'in flow' or and disengaged;
- spent similar dwell time and interaction patterns with learning modules; 'confused' by a set of modules/exams, e.g., jump aimlessly through course content after encountering them;
- 'never take exams' but are active otherwise;
- 'on path' or 'off-path', i.e. whether a of a student flows the linear path set by instructors and designers or deviates from it..
- 'successful' (or 'unsuccessful'), i.e., they follow trajectories that many high (or low) performing learners took.

CNS Cyberinfrastructure for Network Science Center

Seaton et. al. (2014) implemented state-transition dyad network to show transitions from problem solving activities (e.g. homework or exams) to course content, revealing the resources learners rely on when working on these activities.

Figure 6. Transitions to other components during problem solving on (a) homework, (b) midterm, and (c) final. Arrows are thicker in proportion to overall number of transitions, sorting components from top to bottom; node size represents total time spent on that component.



Kizilcec and Piech (2013) used a state transition networks to visualize the flow of learners' across engagement types that were assigned for each assignment period, based on learners' submission times. The visualization supported a K-means cluster methods were applied to learners' coded sequence of engagement to identify broad learner profiles.

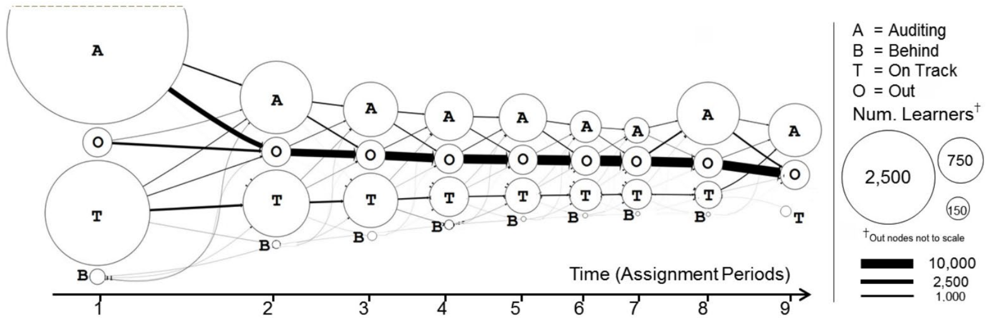
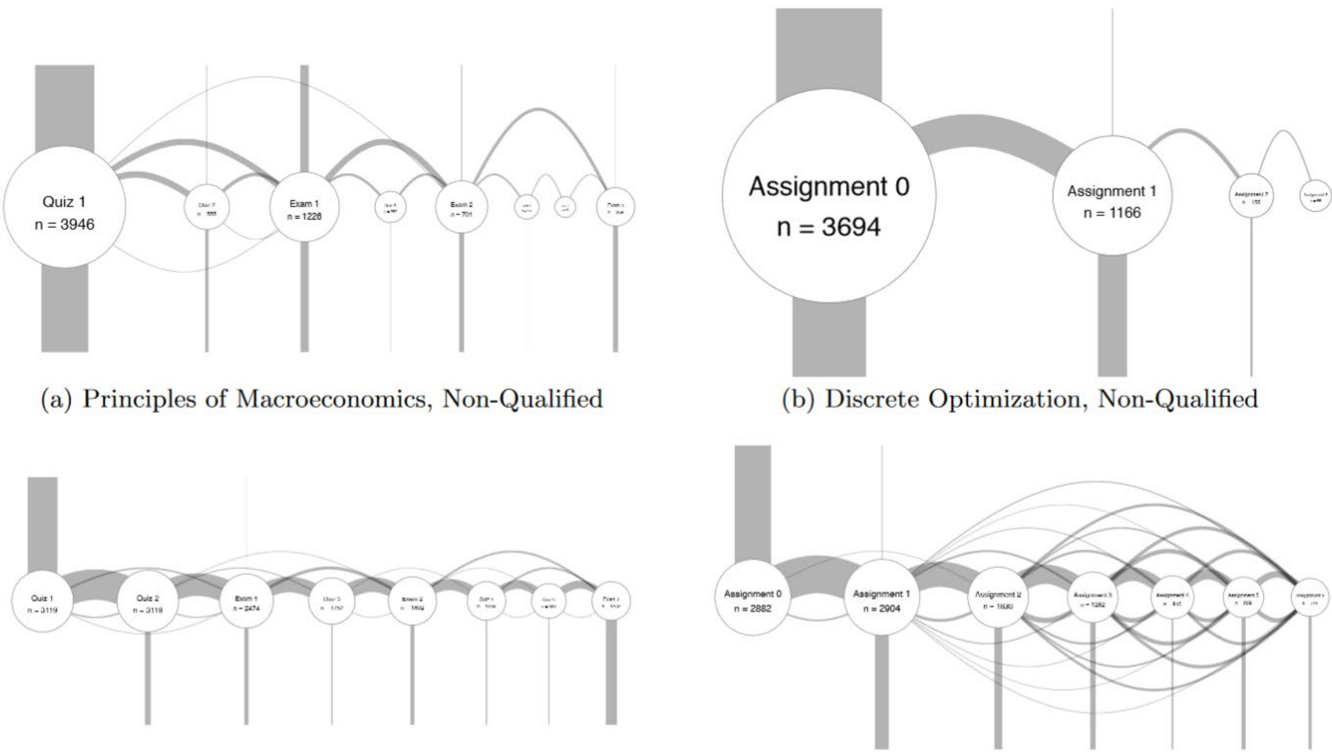


Figure 1: Labels for learners (GS-level)–arcs show movement of students from one assignment period to next.

Coffrin and Corrin (2014) use state transition networks to understand the how learner cohorts (e.g. qualified and non-qualified) use resources and transition between the same type of resources (videos & assignments). Activity is aggregated to the units and number of learners accessing materials; edges show the proportion of learners forward or backwards movement between units in the course.



(a) Principles of Macroeconomics, Non-Qualified

(b) Discrete Optimization, Non-Qualified

(c) Principles of Macroeconomics, Qualified

(d) Discrete Optimization, Qualified

Figure 7: Student Assignment Submissions Transitions Broken Down by Subgroups, Non-Qualified (top) and Qualified (bottom) for Principles of Macroeconomics (left) and Discrete Optimization (right).

Blot, Saurel, & Rousseaux (2014a, 2014b) used time-graph network analysis model learners' use & average transition times between course materials based on based on all clickstream from two small online courses. The work reveals patterns of resource uses and transition times for students that could be leveraged for predictive modeling

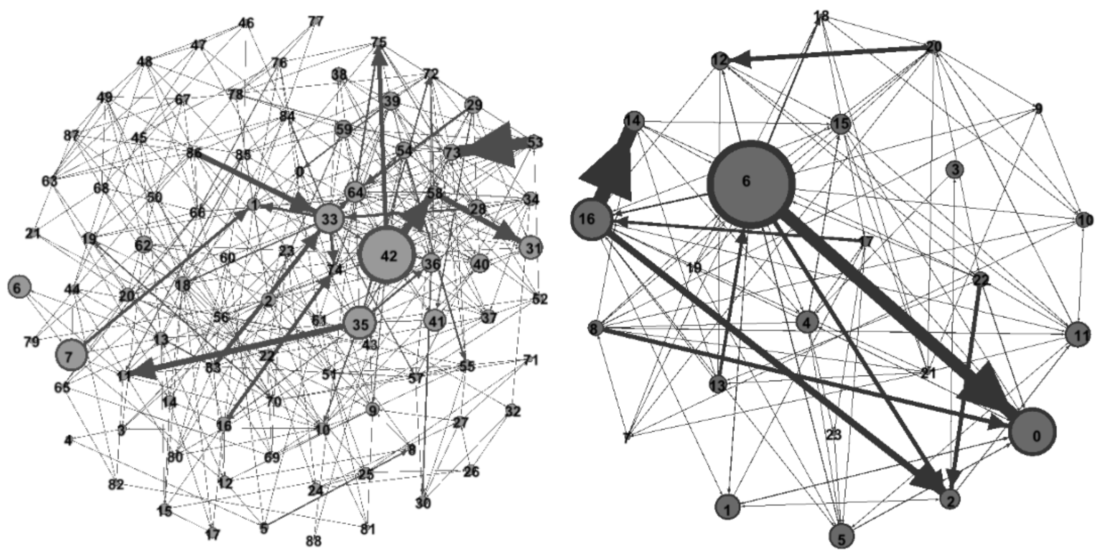


Figure 2. The time-graph produced from course A (left) and course B (right). Big nodes are resources that collected the more visits and thick edges are longest connections. As an example, for course A: $x_{35\text{-}11}$ = 31 days and $x_{33\text{-}35}$ = 60sec and for course B: $x_{16\text{-}14}$ = 250 days and $x_{18\text{-}19}$ = 60sec.

Node represent resources in the course, and are sized by the number of times learners visited the material; edge indicates transitions between materials that use size to indicate the average length of time taken by learners between transition.

Davis and Chen (2016) use state transition diagram edges (arcs) to map designed course sequence; which allows for standardized set of activities and layout for cross course comparison.

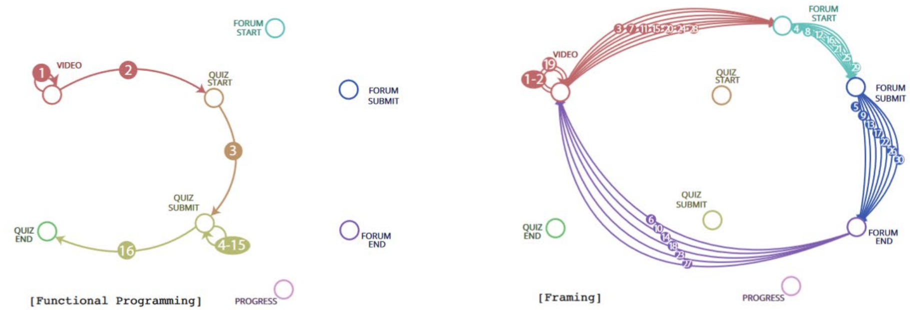


Figure 2: The designed learning path for a standard week (Week 4) of each MOOC. The circled numbers indicate the step number of each transition in that week's sequence. Notice the diversity in course designs that characterize these four MOOCs.

The same layout is used to visualize and compare results of a discrete time Markov Models for passing and non-passing, which used learners' clickstreams data, within and across courses. Edges represent the Markov model transition probabilities between event states, which are filtered to show only high weighted transitions.

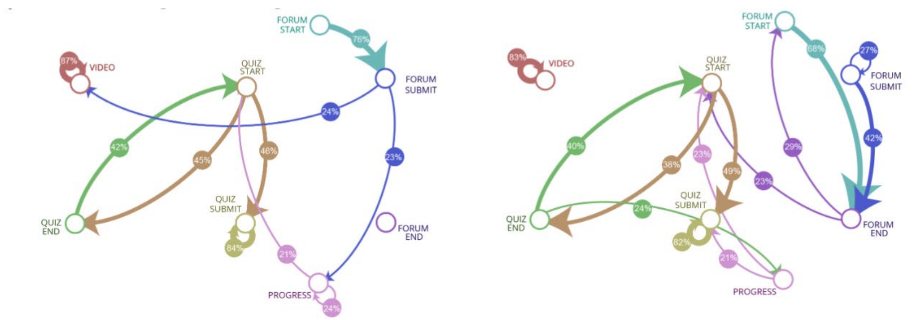


Figure 8: Markov Model state visualization of non-passing (left) and passing (right) learners in the Functional Programming MOOC. Edges with weights below 20% are hidden from view.

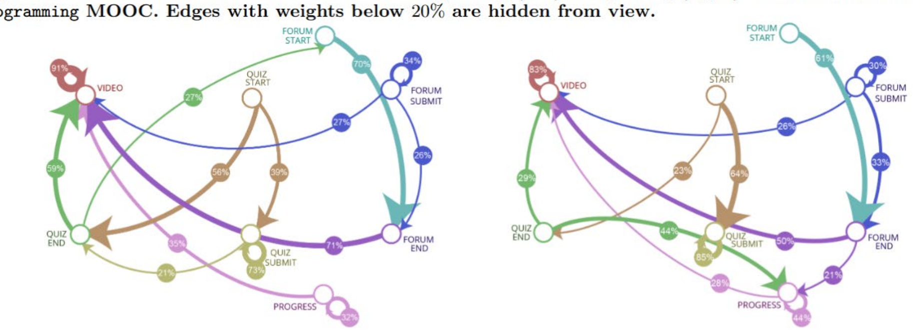
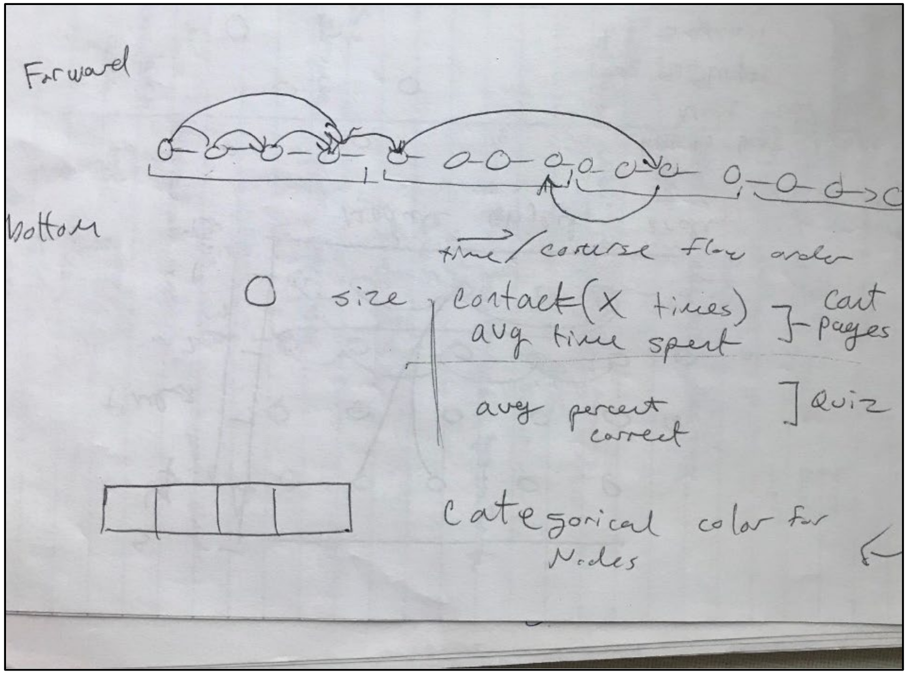
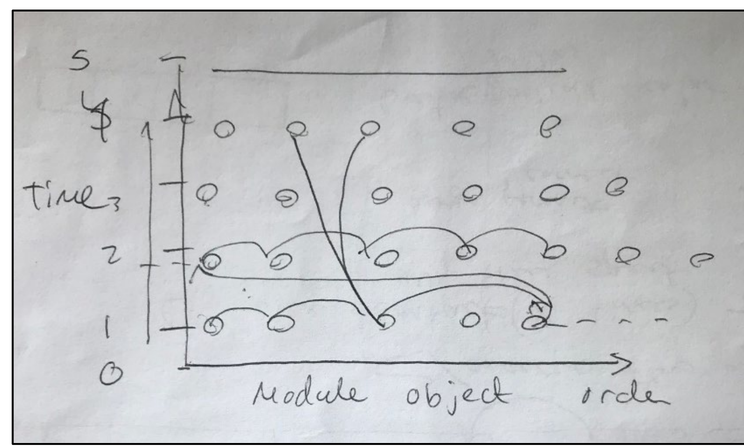


Figure 9: Markov Model state visualization of non-passing (left) and passing (right) learners in the Framing MOOC. Edges with weights below 20% are hidden from view.

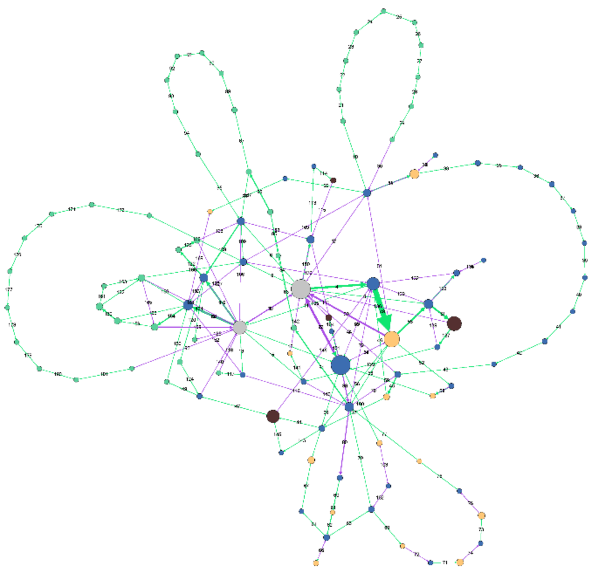


Early sketches for learner trajectory network visualizations provide different layouts considered during the initial design process that took place in Fall 2017.

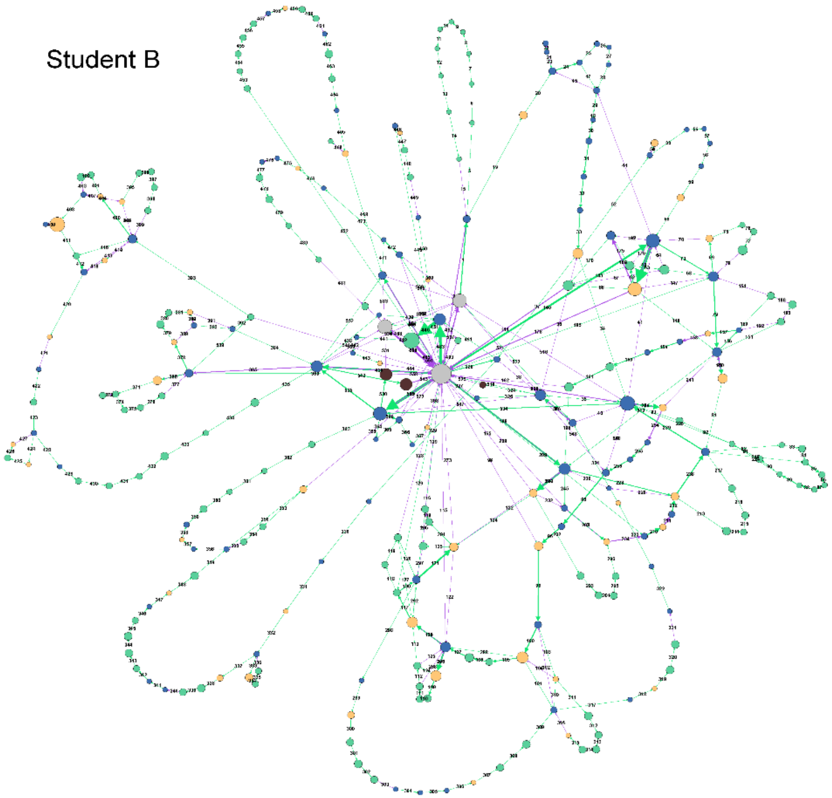


Initial prototype visualizations use a force directed network layout, using Gephi

Student A



Student B



#Event Interactions

Student A		Student B
21	○	77
1	◦	1

Edge Color

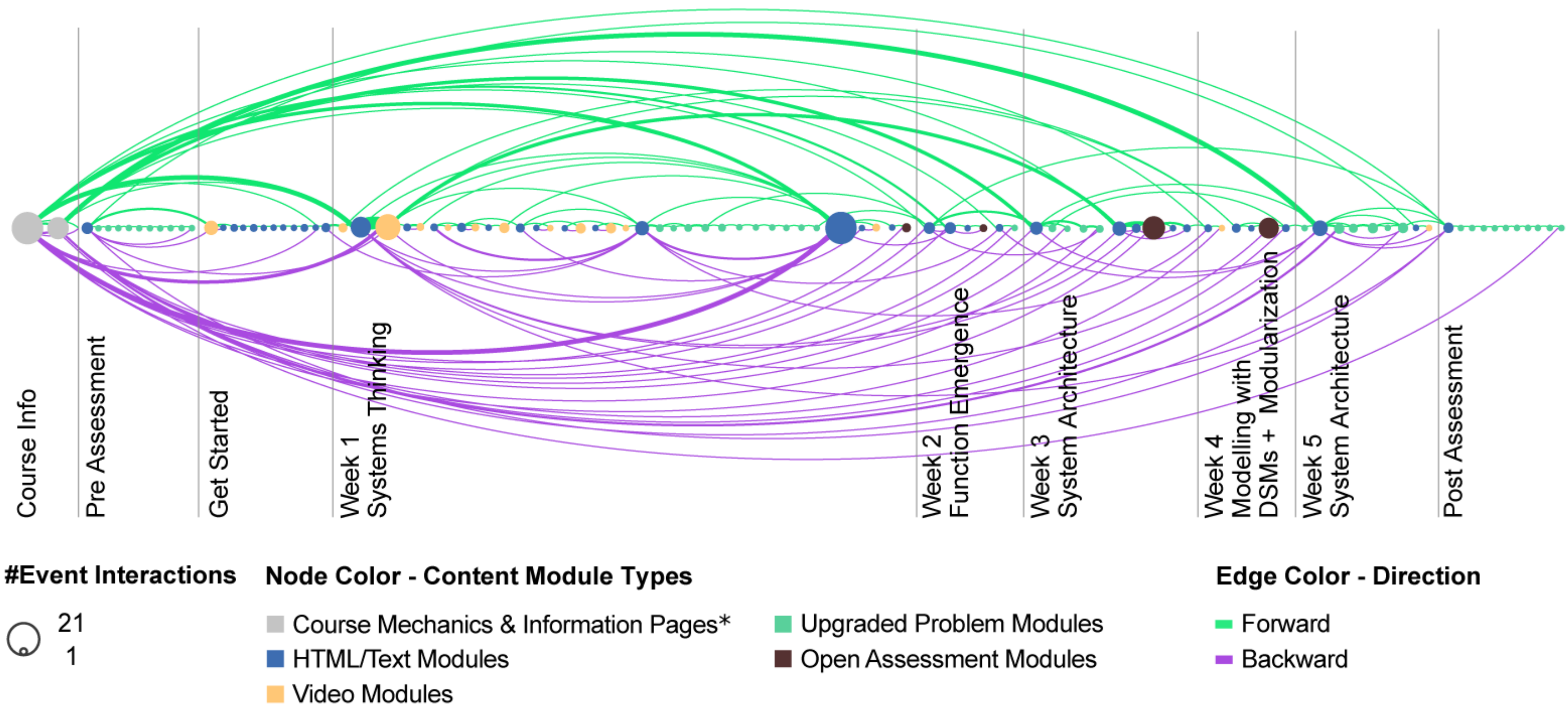
- Forward Direction
- Backward Direction

Node Color

- Course Mechanics and Information Pages^
- HTML/Text Modules
- Video Modules
- Ungraded Problem Modules
- Open Assessment Modules

^These modules do not appear in the course structure, but appear in the event logs of student interactions.

Learner Trajectory Networks – Designs and Prototypes



*These modules do not appear in the course structure, but appear in the event logs of student interactions.

This prototype uses a linear node layout, with node order based on sequence in course structure. Nodes are sized by the number of interactions for a given piece of content or activity.

The visualization represents the aggregate interactions and movement through course content of Student A from the previous slide.

Learner Trajectory Visualization – Dashboard Dev.

Implements an *animated* SVG network visualization and paired legend.

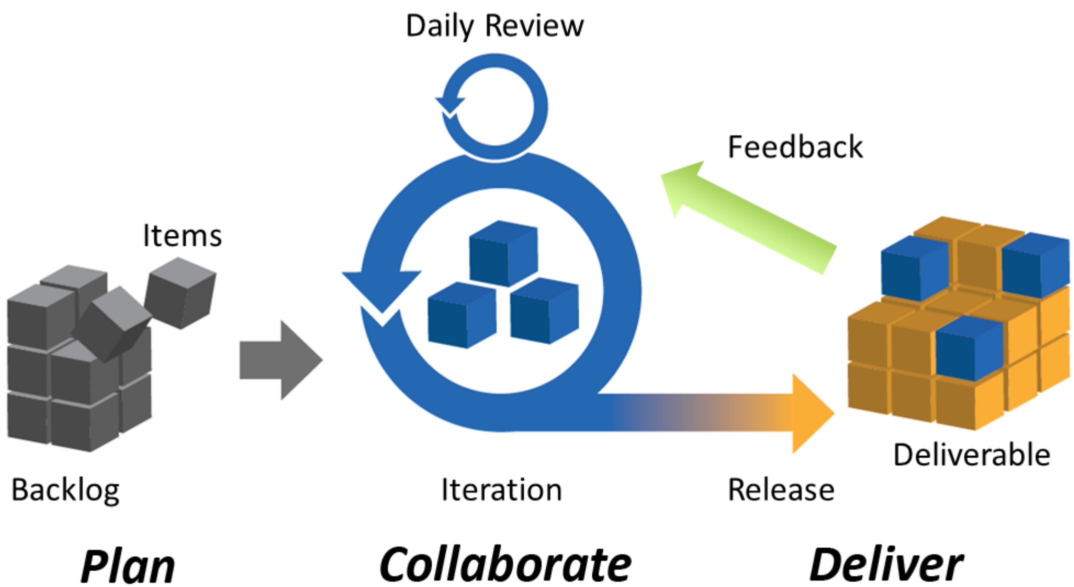
Networks are visualized at the course content level of analysis, based on an individual learner's event logs.

The visualization is deployed as an HTML5 web application that uses:

- NGX-DINO – CNS's in-house visualization framework & library
- Angular 6 Java Script
- HTML5 Web Animations

CNS Cyberinfrastructure for Network Science Center

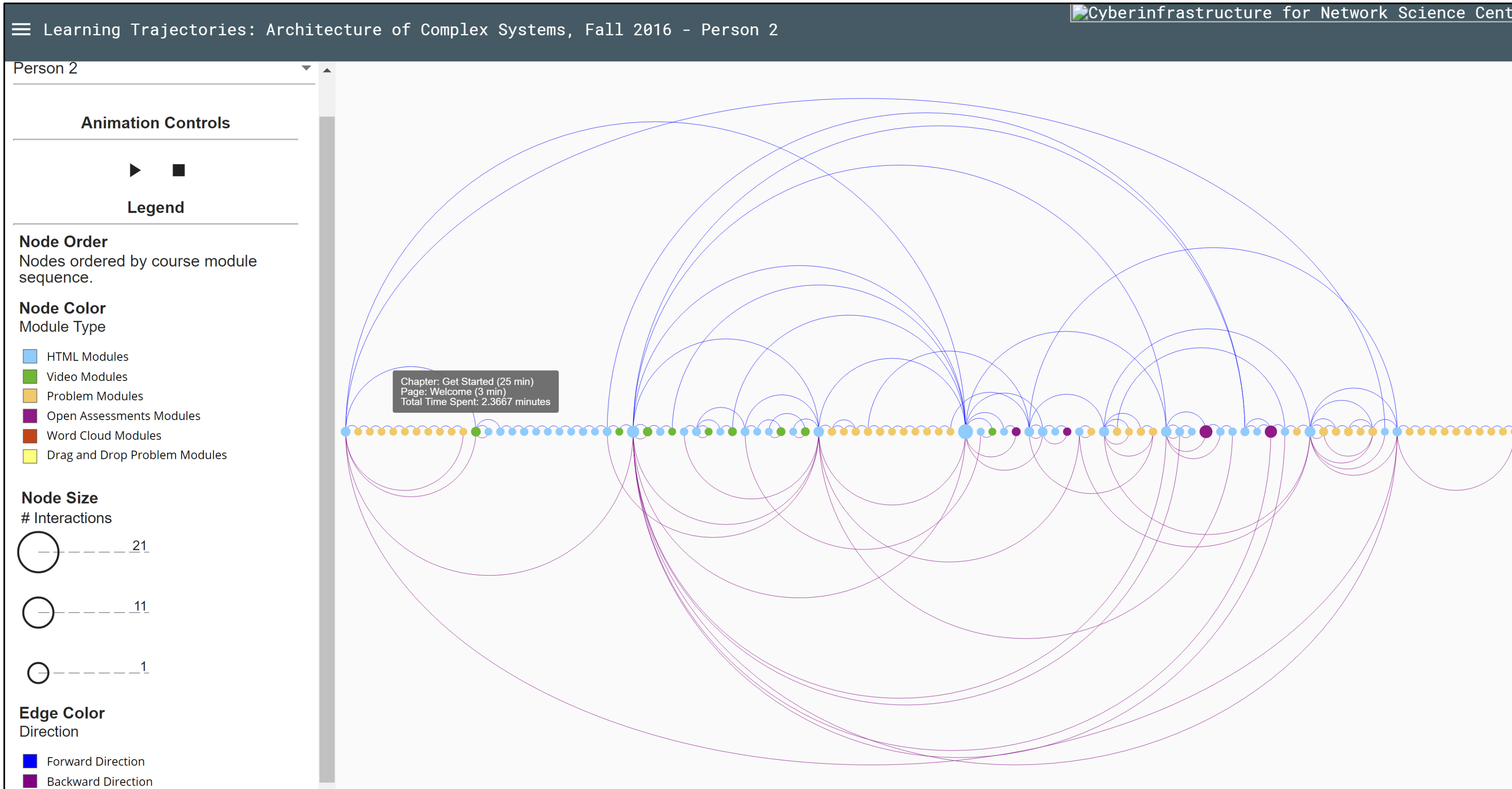
Scrum Development Cycle



Agile Project Management: Iteration

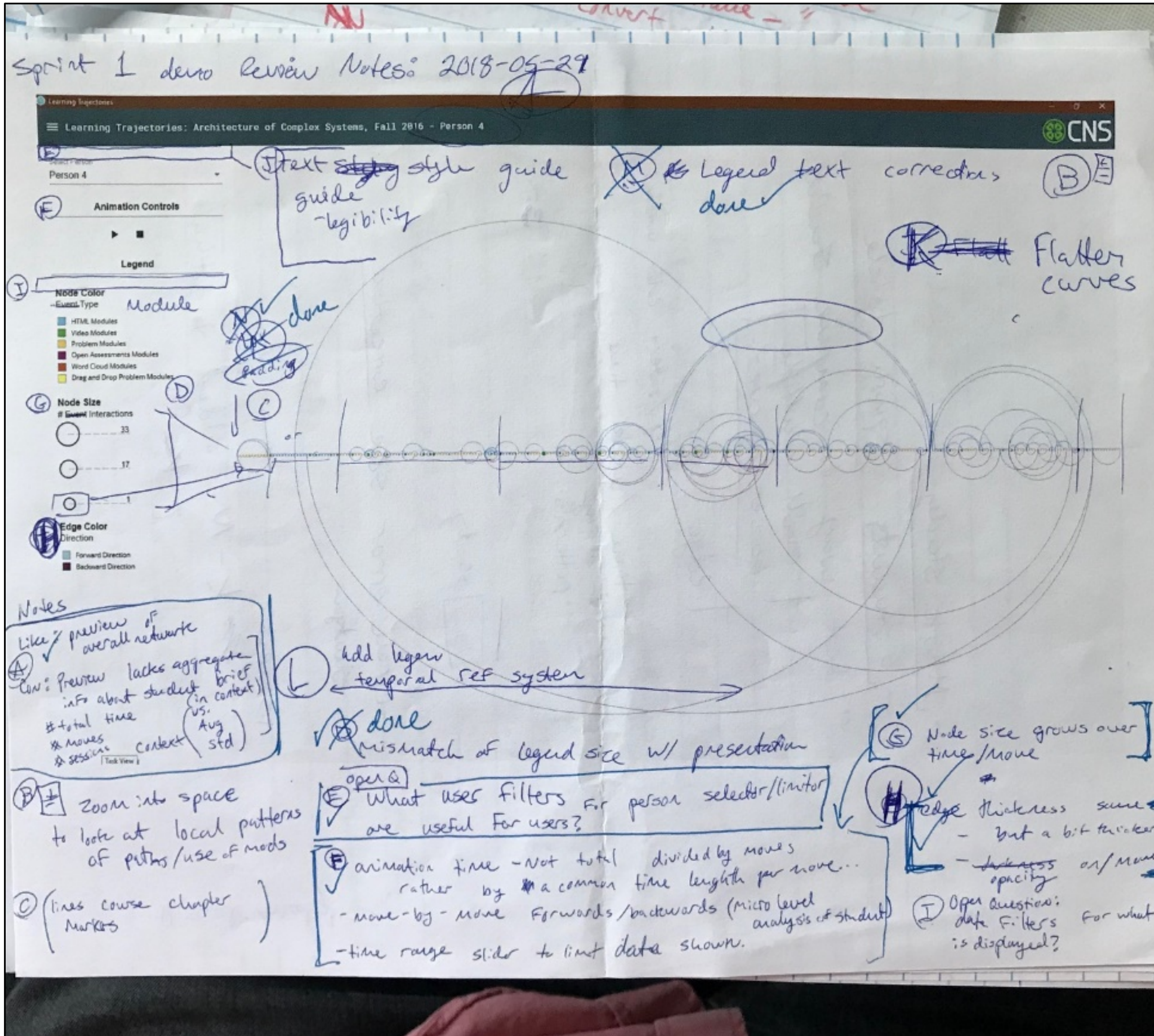
- Short development cycles (1-4 weeks).
- Iterative process with built in review processes to gain stakeholder feedback in the design and development process.
- Collecting user requirements is an essential part of planning stage.

v.0.0.1


Cyberinfrastructure for Network Science Cente
Learning Trajectories: Architecture of Complex Systems, Fall 2016 - Person 2
Person 2
Animation Controls
Legend
Node Order
Nodes ordered by course module sequence.
Node Color
Module Type
HTML Modules
Video Modules
Problem Modules
Open Assessments Modules
Word Cloud Modules
Drag and Drop Problem Modules
Node Size
Interactions
21
11
1
Edge Color
Direction
Forward Direction
Backward Direction
Chapter: Get Started (25 min)
Page: Welcome (3 min)
Total Time Spent: 2.3667 minutes

Sprint 1 Review Demo Notes

At the end of each sprint cycle, our team completes a sprint review, where we see a demo of the current project and provide feedback on the current deployment, and develop questions to take to stakeholders.



Learner Trajectory Visualization – Dashboard Demo

v.0.0.3

- Project Background and Motivation
- Understanding the Data
- Data Processing, Analysis and Visualization Pipeline
- Visualizing Course Structures
- Visualizing Student Engagement
- Visualizing Learner Trajectories
  - Related Work and Design Process
  - Demo the Learner Trajectory Network Visualization
- **Future Development Efforts**

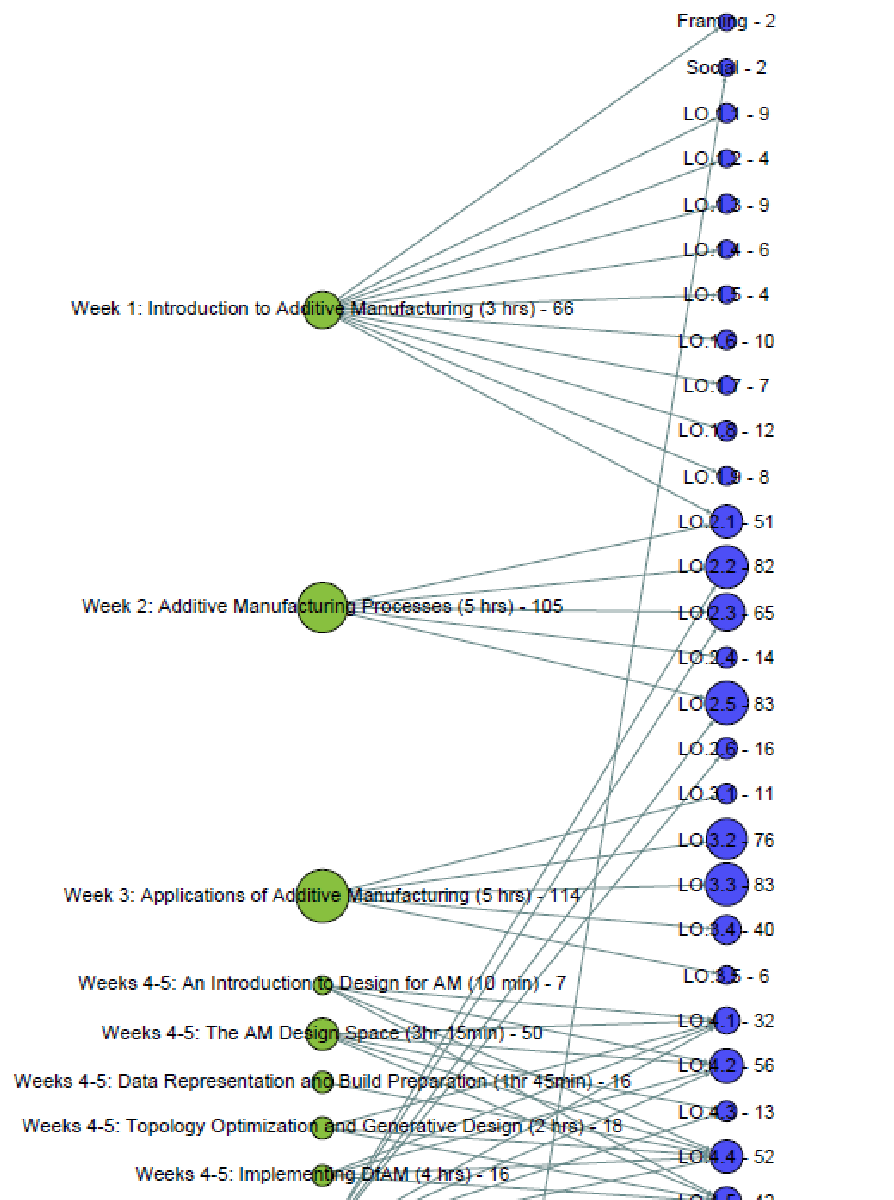
Future Development Efforts

Implement and test new analytic models and visualizations with our research partners on MITxPro's Additive Manufacturing course.

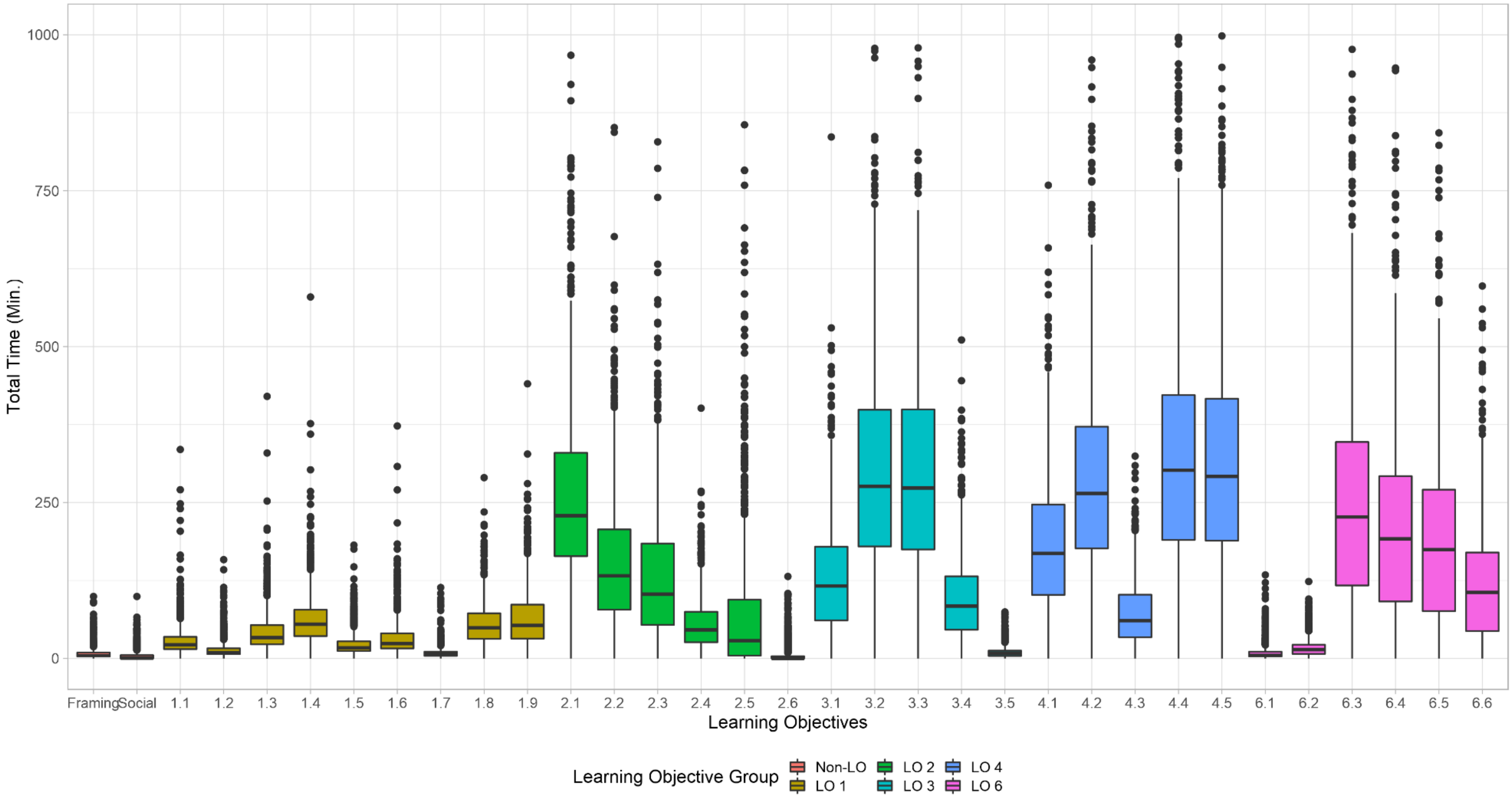
Visualizations of student activity represents 930 students enrolled and active in the course.

- Mapping student course activity and grade performance data to learning objectives in coordination with Dr. Kylie Peppler, Janice Watson, and Joey Huang.

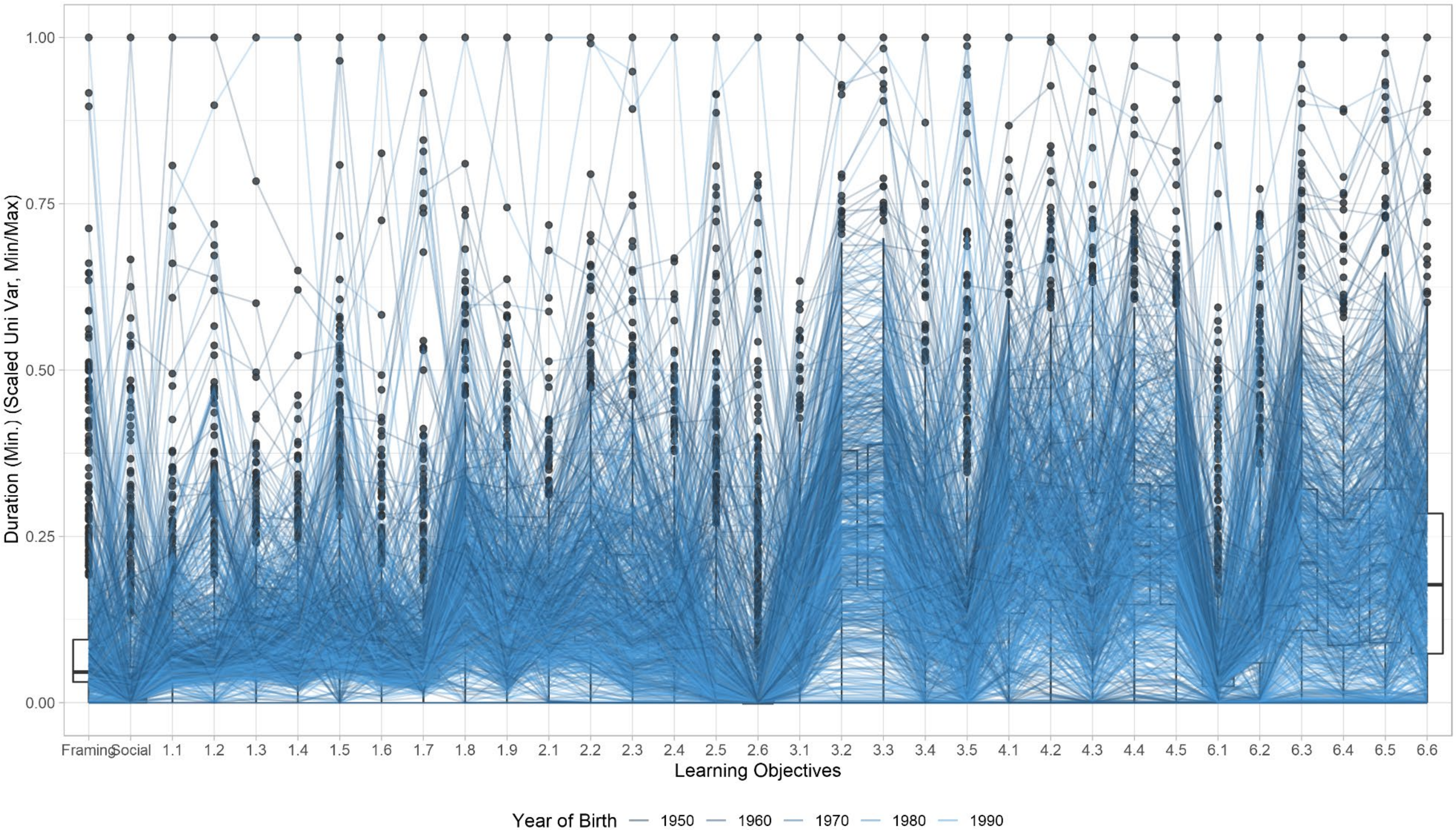
- Visualizing learner behavior models with Dr. Ryan Baker, University of Pennsylvania.



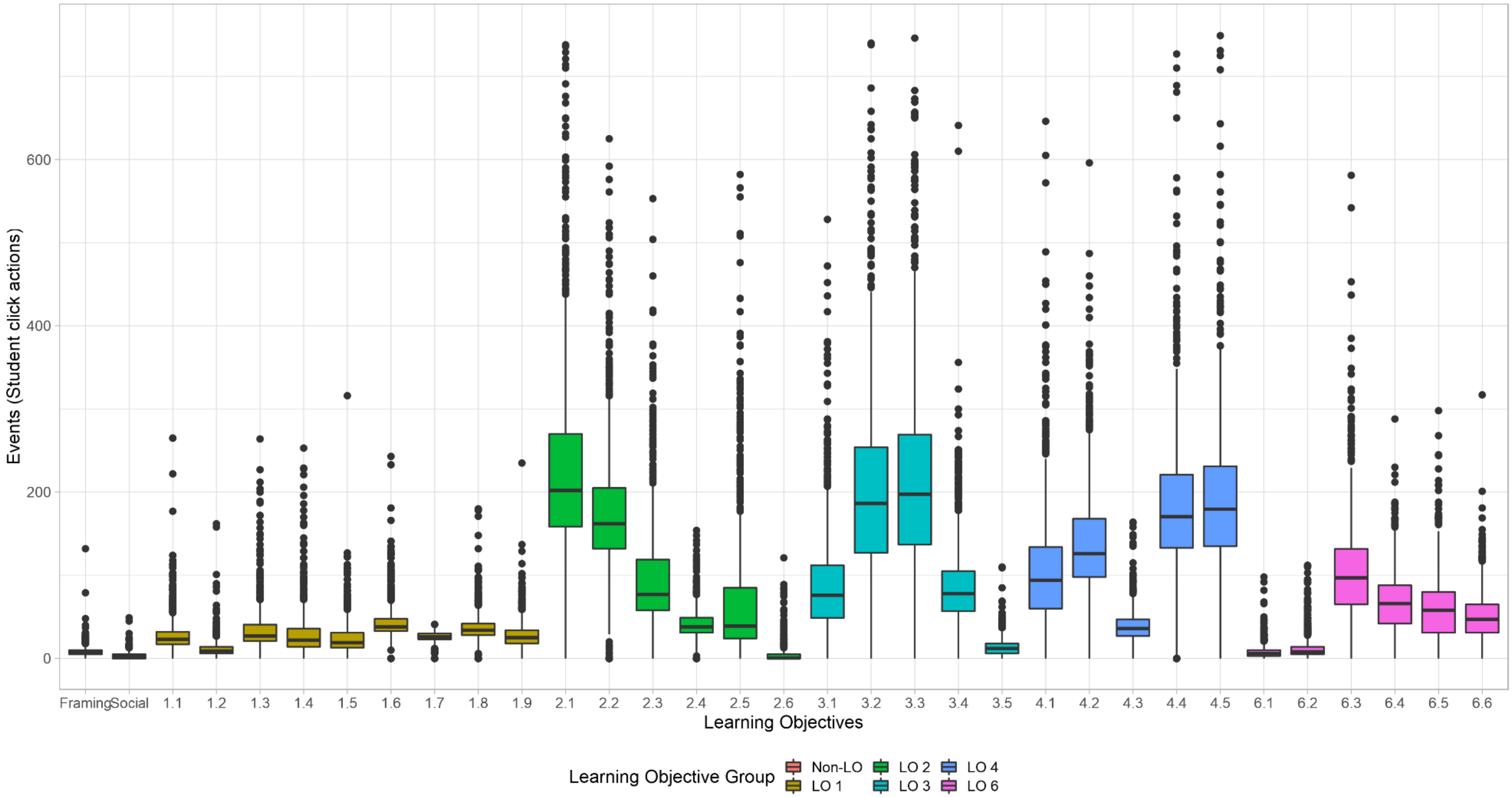
Student Learning Objectives - Duration Boxplots



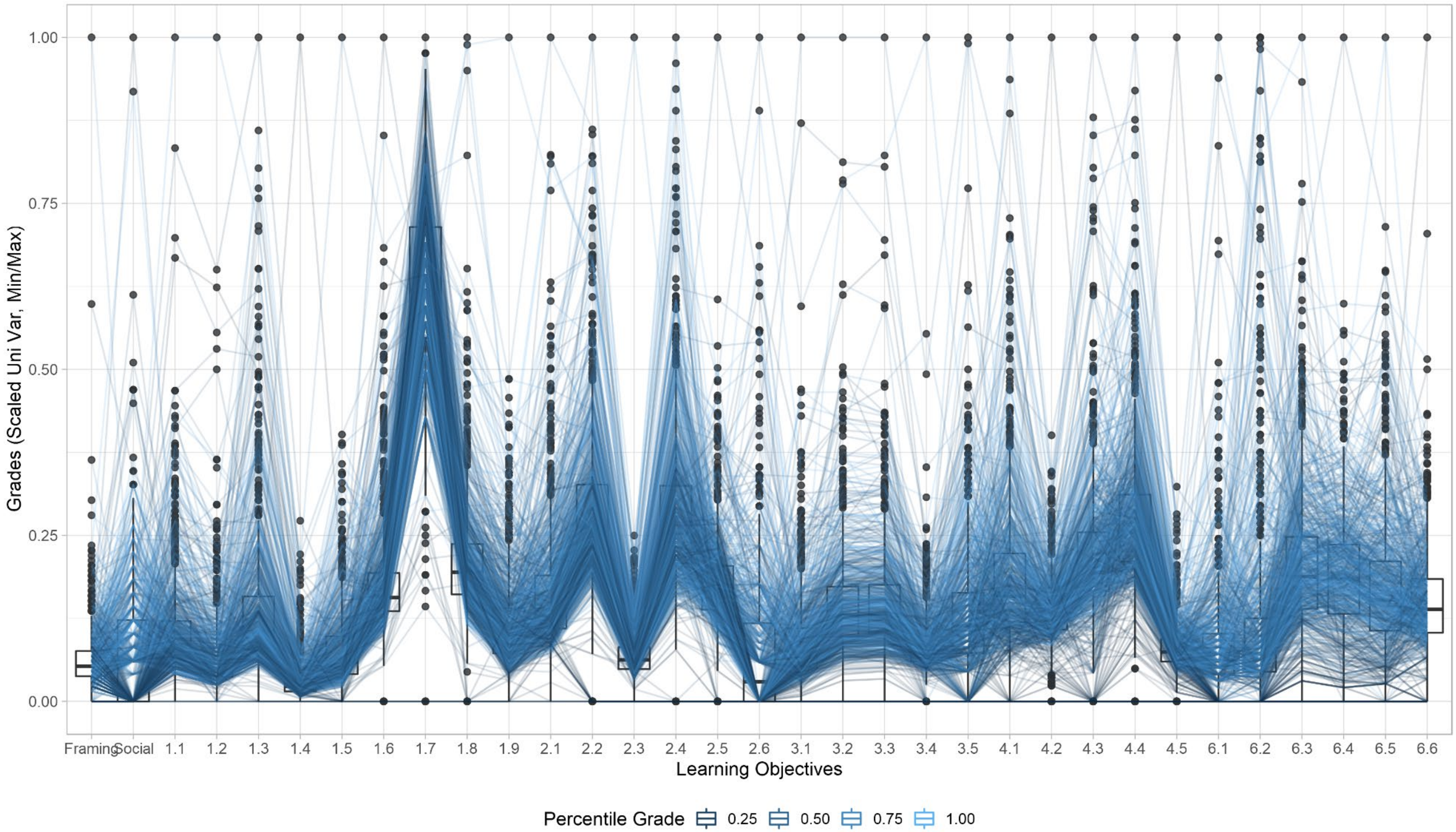
Student Learning Objectives – Duration Parallel Coordinate Chart



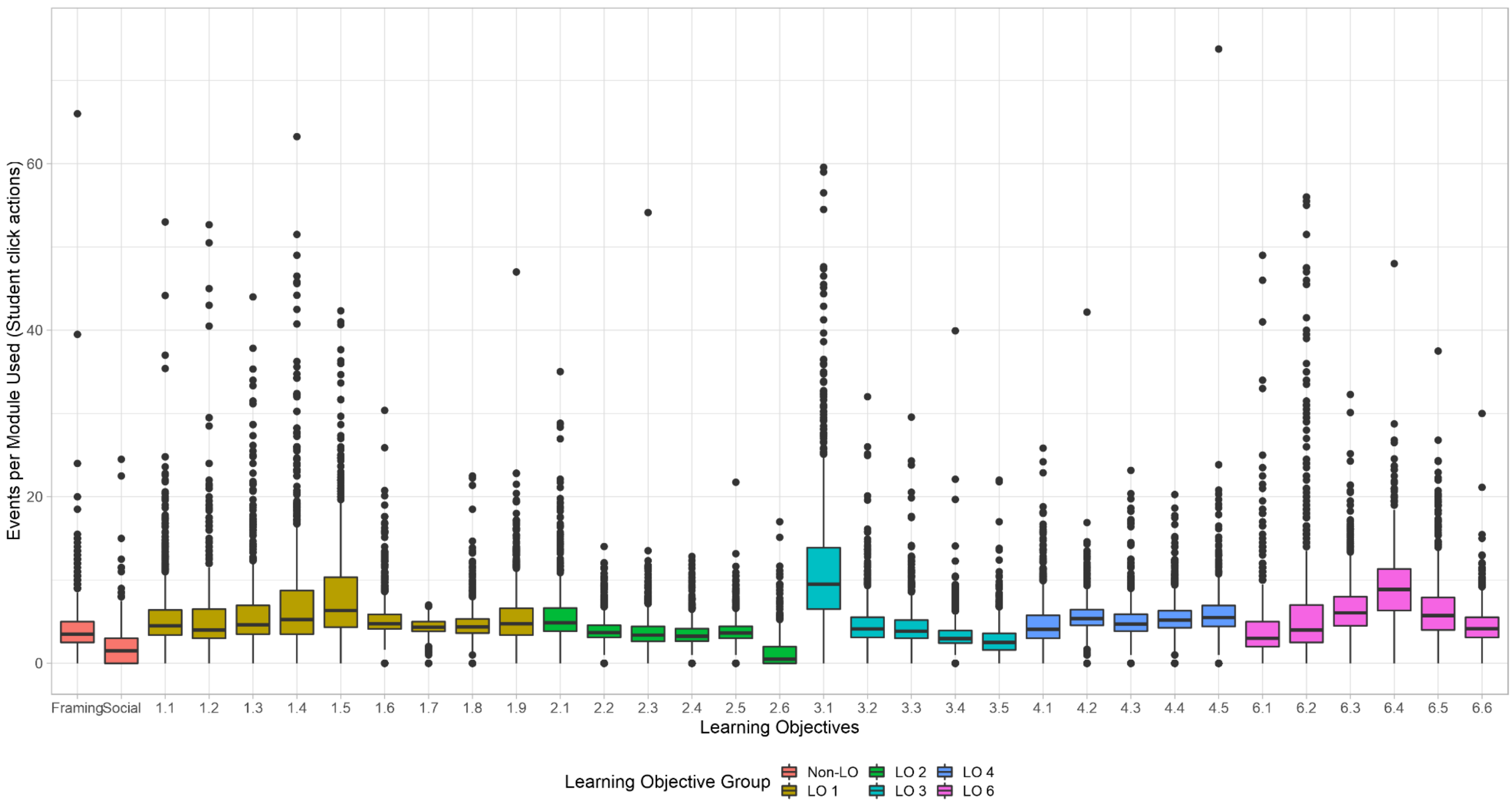
Student Learning Objectives - Events Boxplots



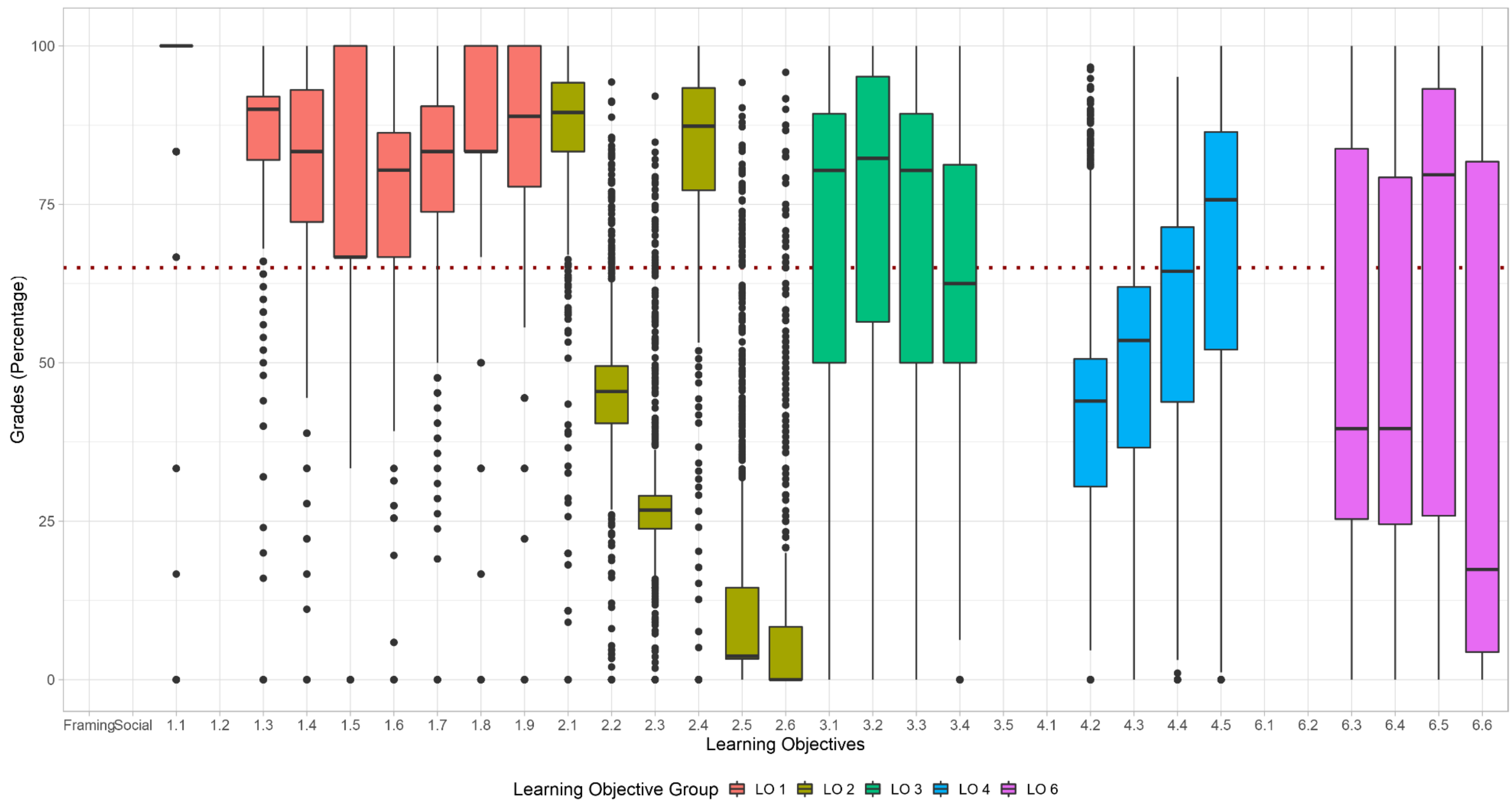
Student Learning Objectives – Events Parallel Coordinate Chart



Student Learning Objectives - Events per Module Used Boxplots

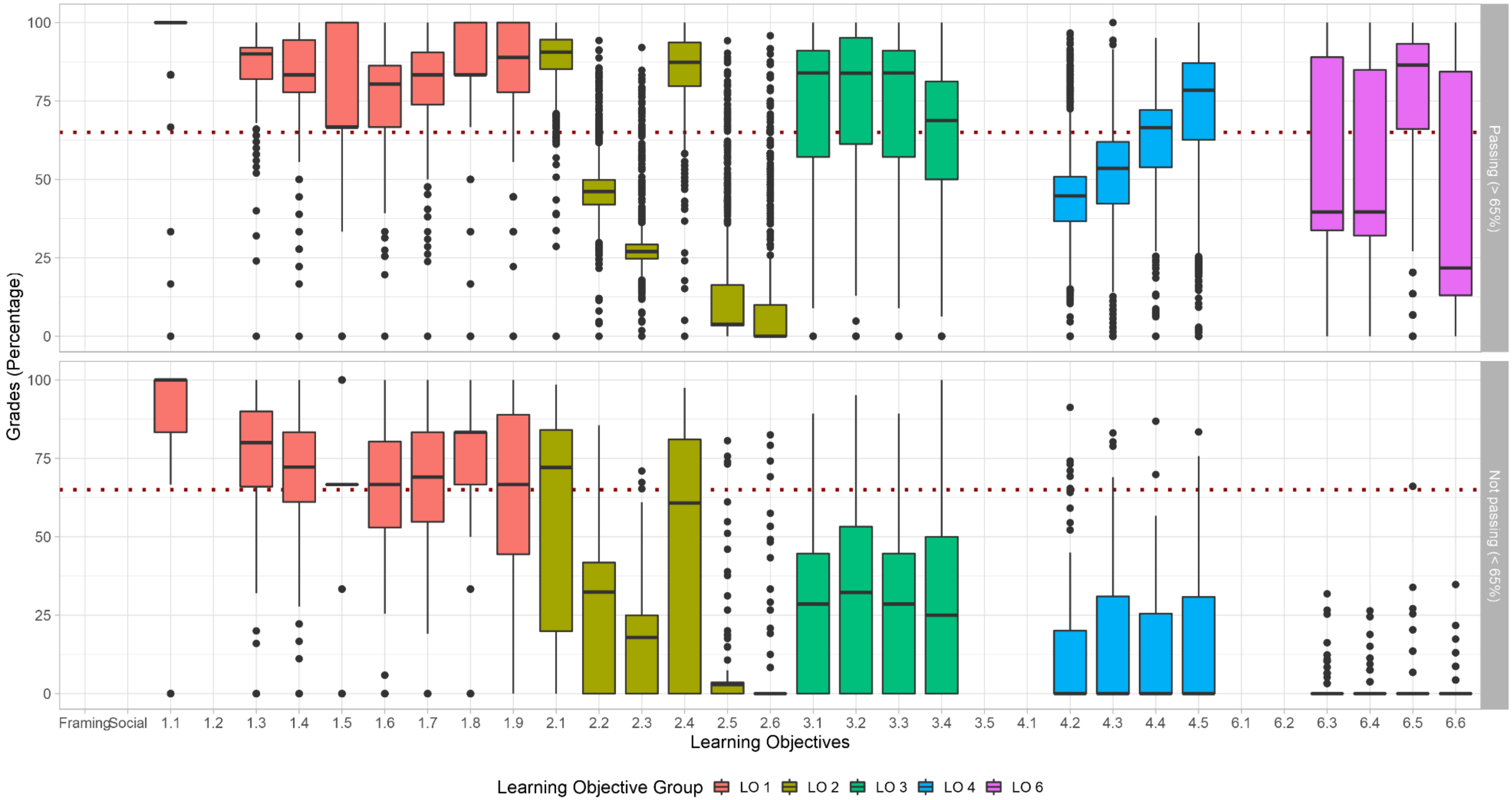


Student Learning Objectives – Grade Performance Boxplots

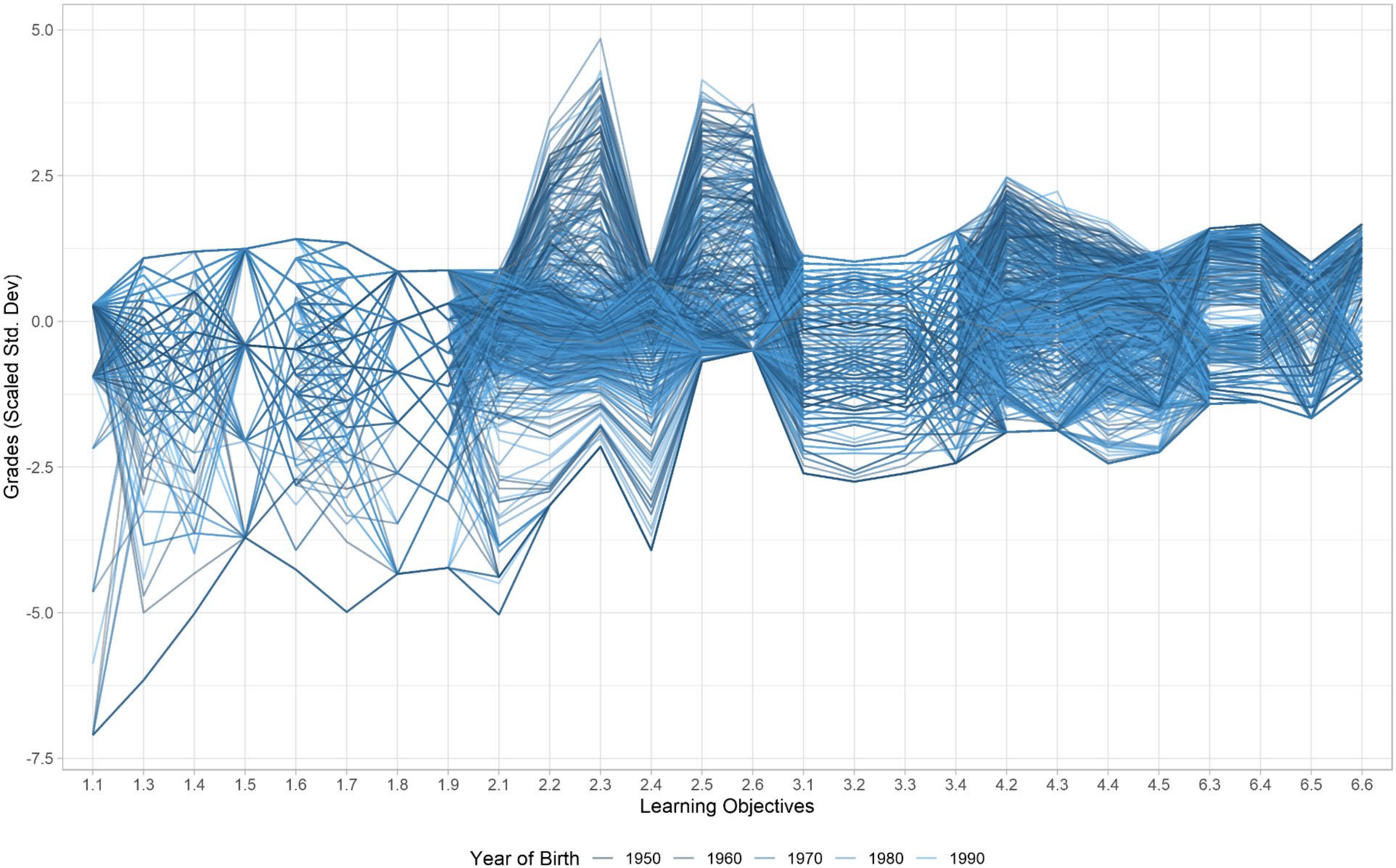


Student Learning Objectives – Grade Performance Boxplots

Cyberinfrastructure for Network Science Center



Student Learning Objectives – Grade Performance Parallel Coordinate Chart



Cloud Infrastructure

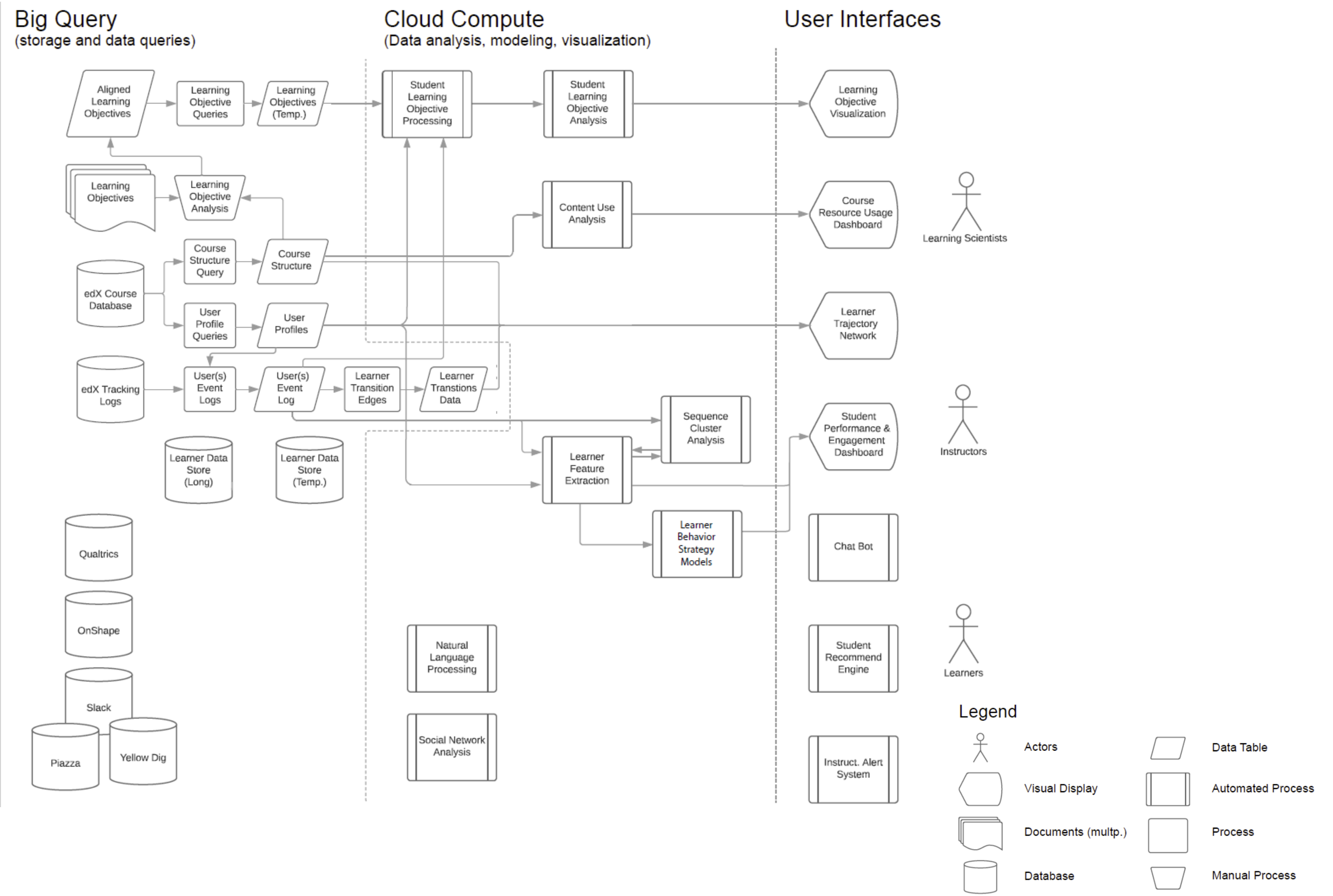
- Transitioning analytics workflow from R scripts into production strength cloud computing infrastructure
  - handle streaming data sets
- Development of analytics and visualization processing pipeline using Google Cloud Platform.
  - streamline design new and replicate older learning analytic workflows and visualizations
- Extend processing, analytics, and visualizations (where possible) to other platforms (e.g. Canvas Data Product)
  - Collaboration exist within IU UITS Learning Technologies, and eLearning Design and Services group as well as data consortia.

Google Cloud Platform

Google BigQuery

Google Cloud Platform – Dataflow Diagram

CNS Cyberinfrastructure for Network Science Center



Questions?

Data and Tool Documentations

- edX. edX Research Guide. edX Inc.; 2016. Available from: https://edx.readthedocs.io/projects/devdata/en/stable/.
- Bastian M., Heymann S., Jacomy M. Gephi: an open source software for exploring and manipulating networks. ICWSM. 2009 May 17, 8(2009):361-2.
- Team RC. R: A language and environment for statistical computing.
- Wickham H., 2009. plyr: Tools for splitting, applying and combining data. *R package version 0.1*, *9*, p.651.
- Wickham H., Francois R, Henry L, Müller K. dplyr: A Grammar of Data Manipulation. R package version 0.5. 0.
- Wickham H. stringr: Simple, consistent wrappers for common string operations. R package version. 2015., 1(0).
- Wickham H. ggplot2: elegant graphics for data analysis. Springer, 2016 Jun 8.
- Cyberinfrastructure for Network Science Center. GitHub Learner Trajectory Network Project Repository (Visualization). 2018 Sept 29. https://github.com/cns-iu/learning-trajectories

Prior Work

- Seaton D.T., Bergner Y., Chuang I., Mitros P., Pritchard D.E. Who does what in a massive open online course? Communications of the ACM. 2014;57(4):58-65. doi: 10.1145/2500876.
- Kizilcec R.F., Piech C., Schneider E. Deconstructing disengagement: analyzing learner subpopulations in massive open online courses. Proceedings of the Third International Conference on Learning Analytics and Knowledge; Leuven, Belgium: ACM; 2013. p. 170-9. doi: 10.1145/2460296.2460330.
- Coffrin C., Corrin L., de Barba P., Kennedy G., editors. Visualizing patterns of student engagement and performance in MOOCs. 2014: ACM Press. doi: 10.1145/2567574.2567586.
- Blot G, Saurel P, Rousseaux F, editors. Resource Connectivism in E-learning Courses Based on an Analytical Time-Graph. ICT, Society of Human Beings; 2014a.
- Blot G, Saurel P, Rousseaux F, editors. Pattern discovery in e-learning courses: a timebased approach. 2014 International Conference on Control, Decision and Information Technologies (CoDIT); 2014b: IEEE.
- Davis D., Chen G., Hauff C., Houben G-J. Gauging MOOC Learners' Adherence to the Designed Learning Path. EDM. 2016; 16:9th.

CNS Cyberinfrastructure for Network Science Center

MG and KB are partially supported by The Boeing Company, the National Institutes of Health under awards P01AG039347 and U01CA198934 and the National Science Foundation under AISL-1713567. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.