UC San Diego

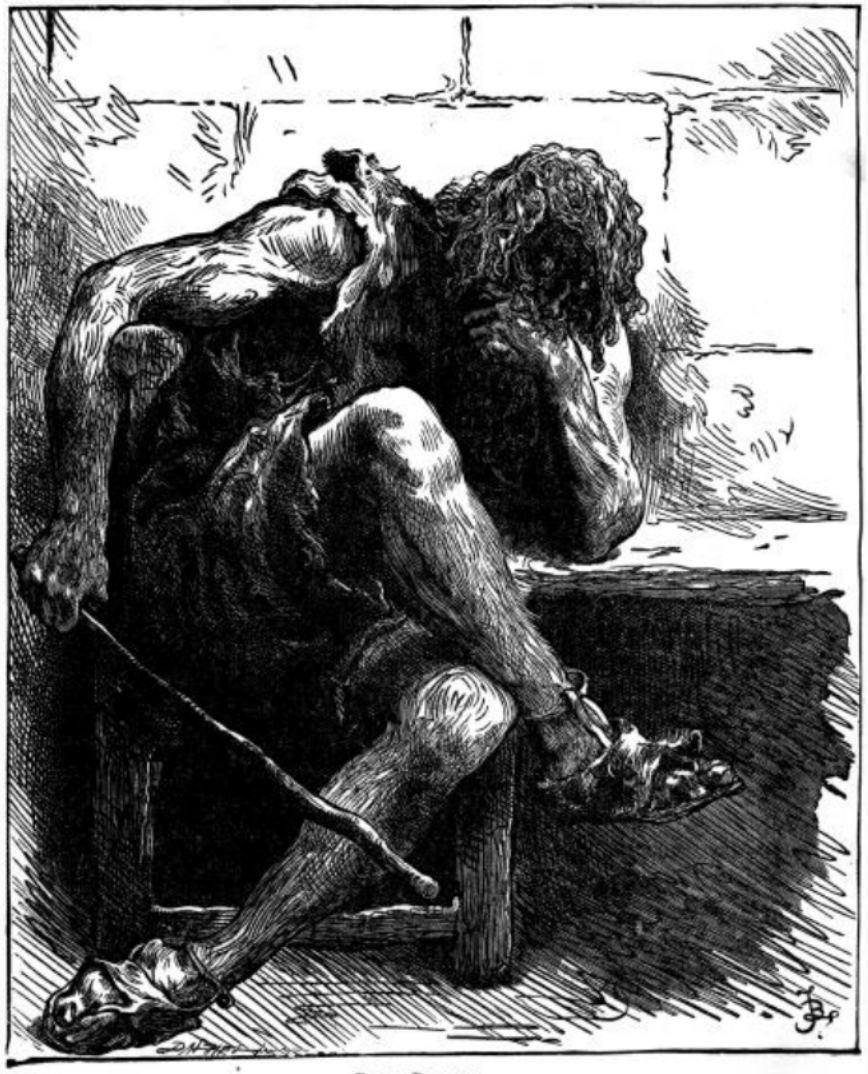
Canvas Data and Student Retention Analysis: A Match Made In Heaven!

UC San Diego Student Activity Hub
March, 2023

Vince Kellen, Ph.D., Chief Information Officer
Amin Qazi, Enterprise Architect
Kevin Chou, J.D., Deputy CIO

AGENDA

1. Problems
2. Approach
3. SIS data analysis
4. Canvas LMS visuals
5. Canvas LMS & SIS data visuals
6. Wrap up



GIANT DESPAIR.

"Over this stile is the way to Doubting Castle, which is kept by Giant Despair."

COMPLEX AND DISTRIBUTED NATURE OF INSTITUTIONS

At UC San Diego, we are not unique. Like other places, we have data that lies in nested technical and organizational silos. The reasons for this tend to fall into two categories:

1. Significant technical differences between different IT systems makes joining different data difficult or impossible

2. Organizational silos prevent bringing data from different units together

WHAT WE DID: A THREE-PRONGED APPROACH TO THIS PUZZLE

1.
Establish a collaborative evidence-based culture that supports safe data sharing and democratization of insights

2.
Separate data analytic use from data warehouse construction through governance, enabling bringing data together

3.
Implement fast and affordable technology, making large-scale data warehousing possible, leading to insights quickly

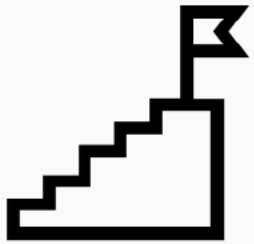
UC SAN DIEGO CAMPUS DATA GOVERNANCE

- Governance and collaboration are very location-specific. Institutional context is everything
- There is nothing like data to make the heart beat faster!
- Good governance is foundational for bringing different units and hence data together
- It is difficult to provide "cookie cutters" for governance
- We grew ours over time. It isn't perfect and we are always seeking improvement
- Key governance duties:
  - Establish data access rules and procedures
  - Help define data
  - Establish use case priorities

UC San Diego Data Community of Practice Principles

- Be safe and secure
- Be collegial
- Help improve data quality
- Be open-minded and inquisitive
- Value individual uniqueness
- Share

How can we improve student data management collaboratively?



OUR DESIRE

To advance the state of student data management and student analytics in order to achieve educational goals, as diverse as they may be, while protecting district and school uniqueness, control over data and privacy

PROBLEMS

Bringing disparate data from different systems and unlocking broad educational and learning analytics is difficult. Managing the governance and politics surrounding data takes hard work and time

APPROACH

Using next generation analytic technologies, we were able to clear the technical hurdles. We wanted to support diverse analysis use cases: including enrollment, retention, graduation, majors, grades, course outcomes and digital curriculum and LMS data

FUTURE

Make it easier to address organization-specific outcomes without excessive lock in with as much control over all aspects of the infrastructure as needed. Our goal is not to make a profit with this initiative, but to help organizations make a difference in their educational mission the way they want to

SAH IS A NEW RULES ARCHITECTURE

Today's designers of data analytics systems are using thirty-year-old mental models around scarcity of compute and are thus crippling their designs, not fully realizing how radically different 21st-century analytics has become (New Rules)

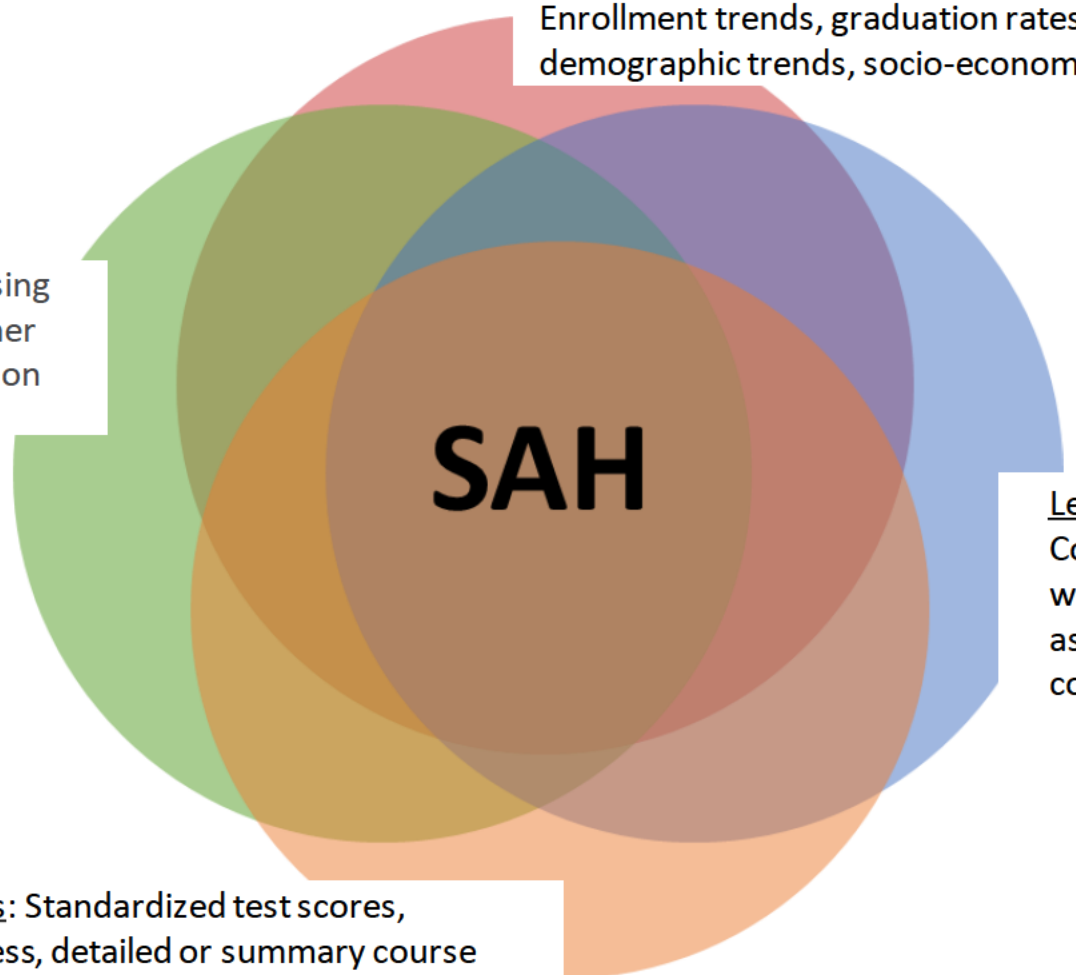
1. Everything is a verb. Events, streams, replayable log, idempotency, eventual consistency
2. Express maximum semantic complexity. All attributes, lowest level of granularity
3. Build provisionally. Highly reusable, overlapping, curated views
4. Design for the speed of thought. Sub-second clicks, real-time ingestion as needed
5. Waste is good. Data explosion. Overlapping curated views. No user joins. Feral de-normalization
6. Democratize the data. Equal access for all. Ease of use. Information sharing is power

https://er.educause.edu/articles/2019/5/21st-century-analytics-new-technologies-and-new-rules

SIS & INSTRUCTURE CANVAS: THE STUDENT ACTIVITY HUB (SAH)

Education institution analytics:
Enrollment trends, graduation rates, retention rates, demographic trends, socio-economic analysis, etc.

Engagement analytics: tutoring or advising interactions, co-curricular activities, other service center engagement (depending on source systems)



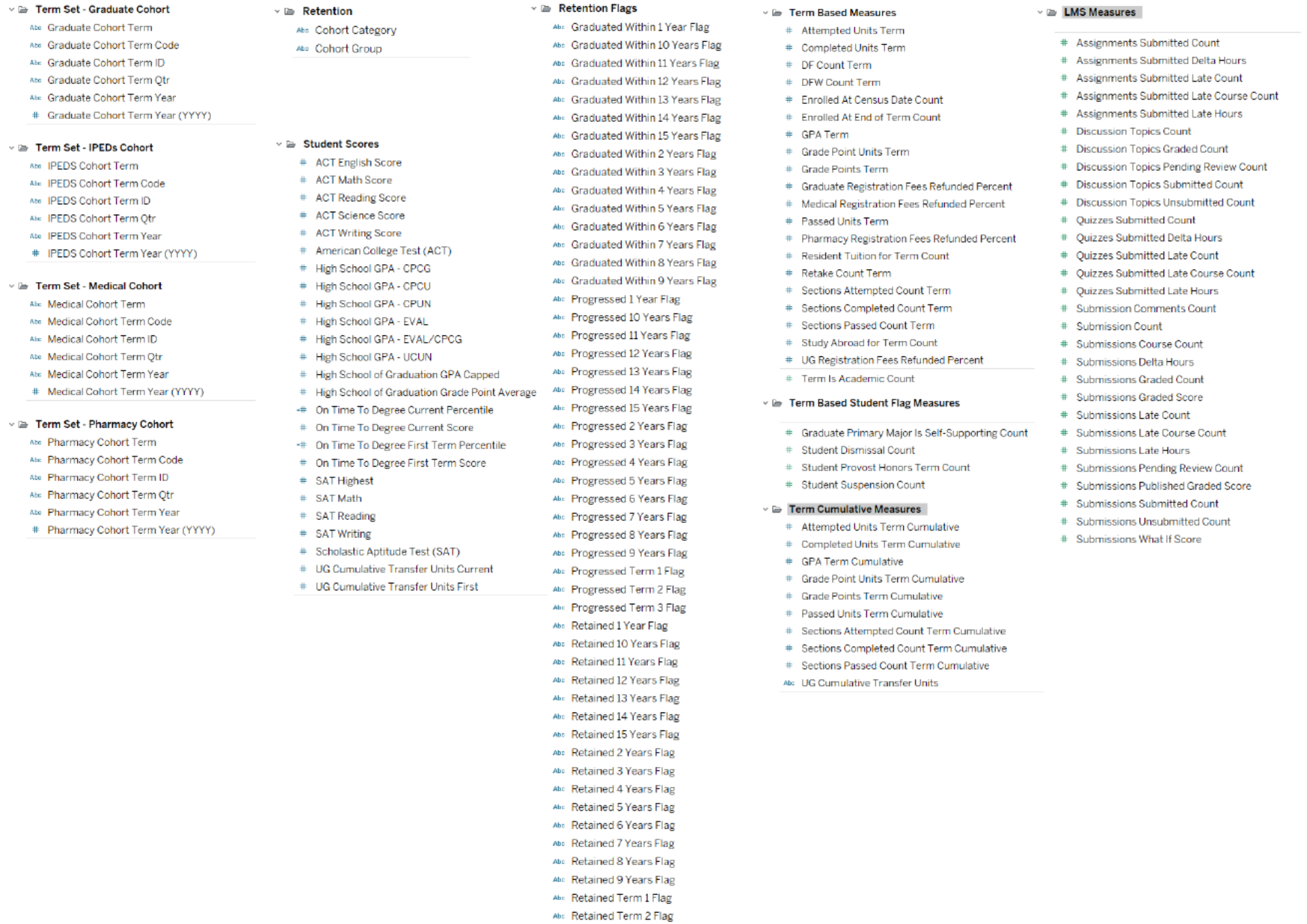
Learning analytics:
Course engagement, assignments, submissions, within-course grades, learning outcomes assessment, digital tool engagement analysis, course content page views, video views

Academic analytics: Standardized test scores, satisfactory progress, detailed or summary course grades, courses needing pedagogical improvement, post-secondary readiness

A SAMPLING OF SAH MEASURES

A sampling of some SAH measures:

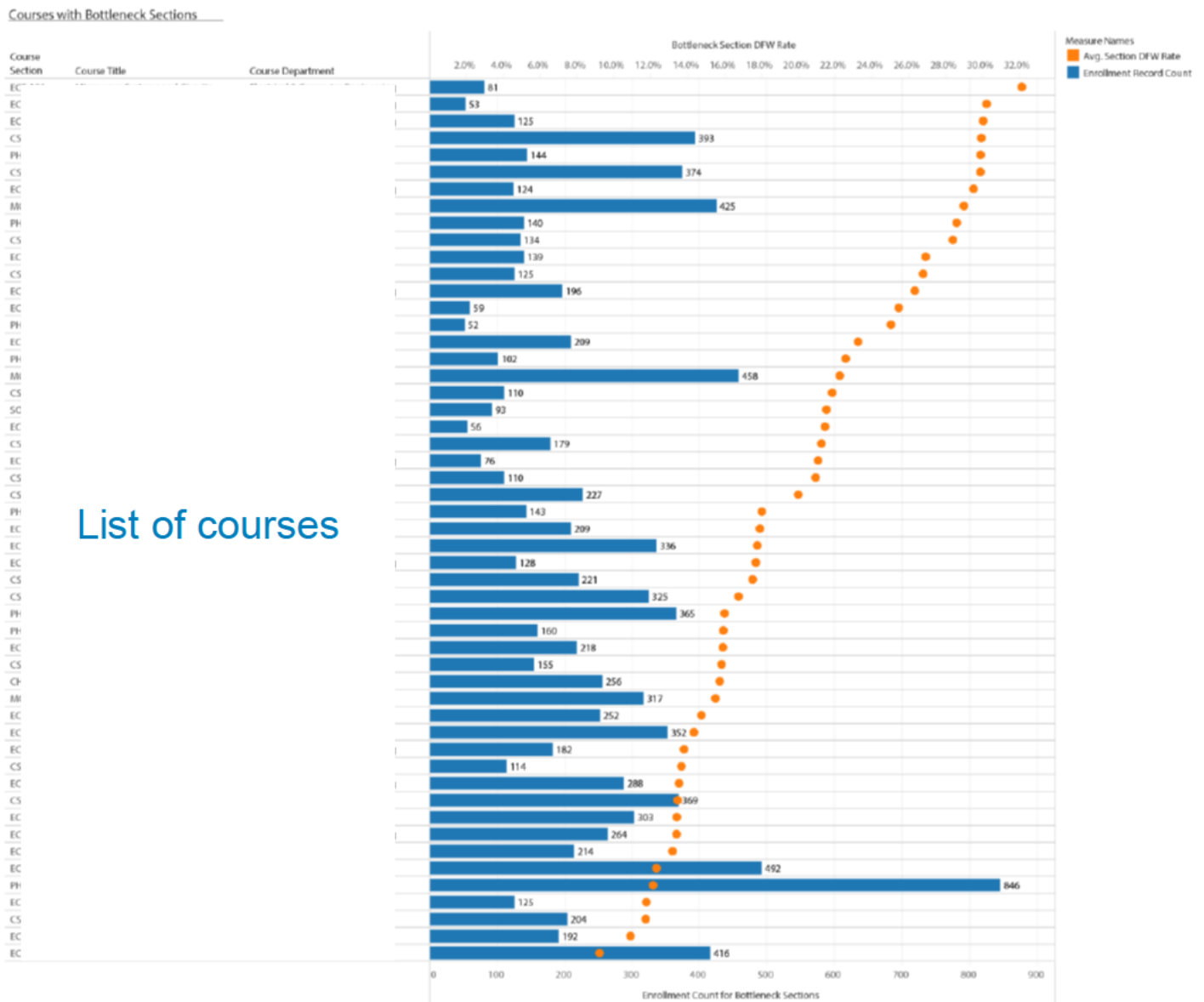
1. How student cohorts are categorized (unlimited, but common categories here)
2. Entrance measures
3. Retention flags (which each have measures)
4. Measures for each term
5. Cumulative measures up to the given term
6. Canvas assignment LMS measures



BOTTLENECK CLASSES EXAMPLE

Courses with high failure rates (orange dots) showing course enrollment (blue bar)

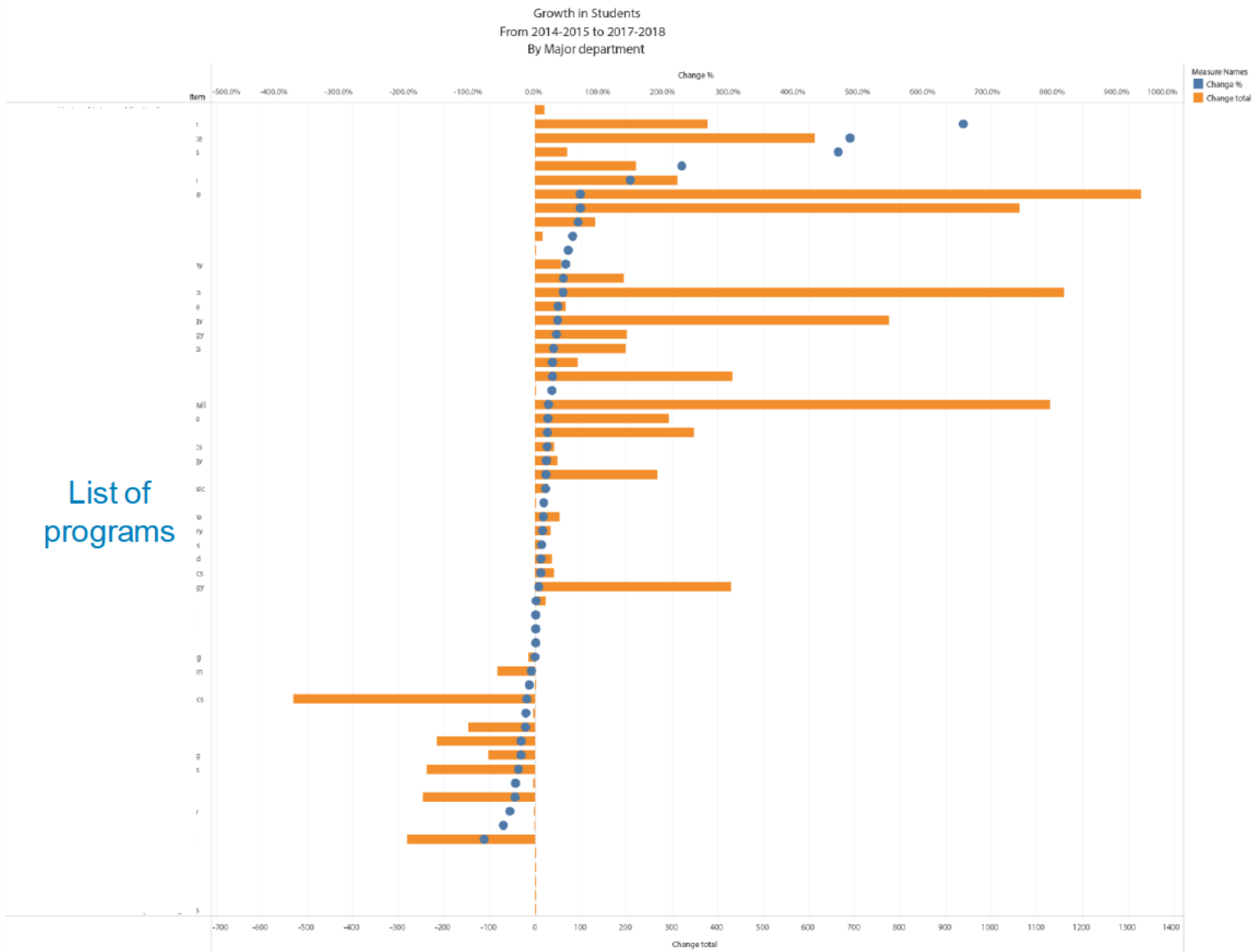
List of courses



ACADEMIC PROGRAM GROWTH EXAMPLE

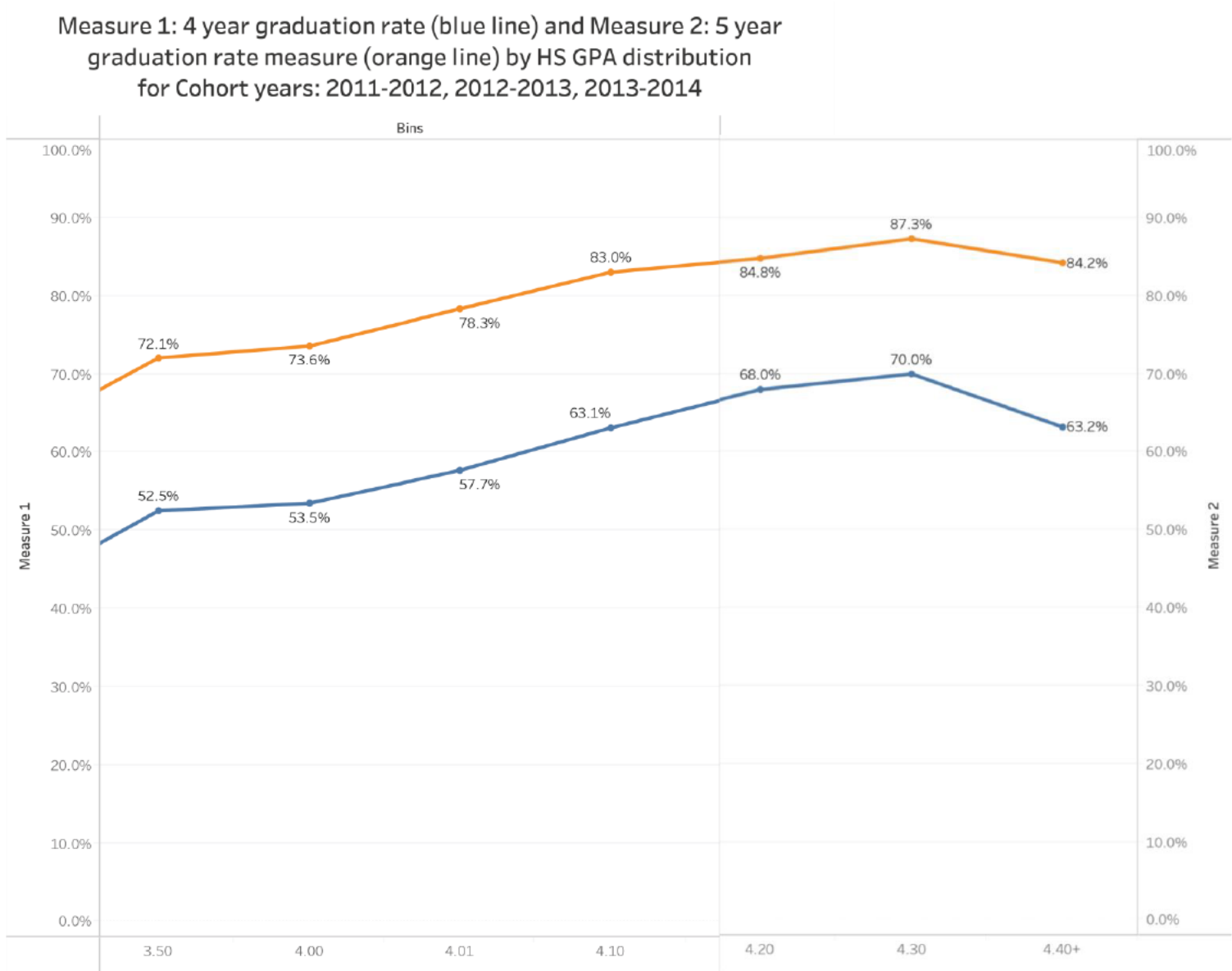
Program enrollment growth (orange bar) and growth percentage change (blue dot)

List of programs



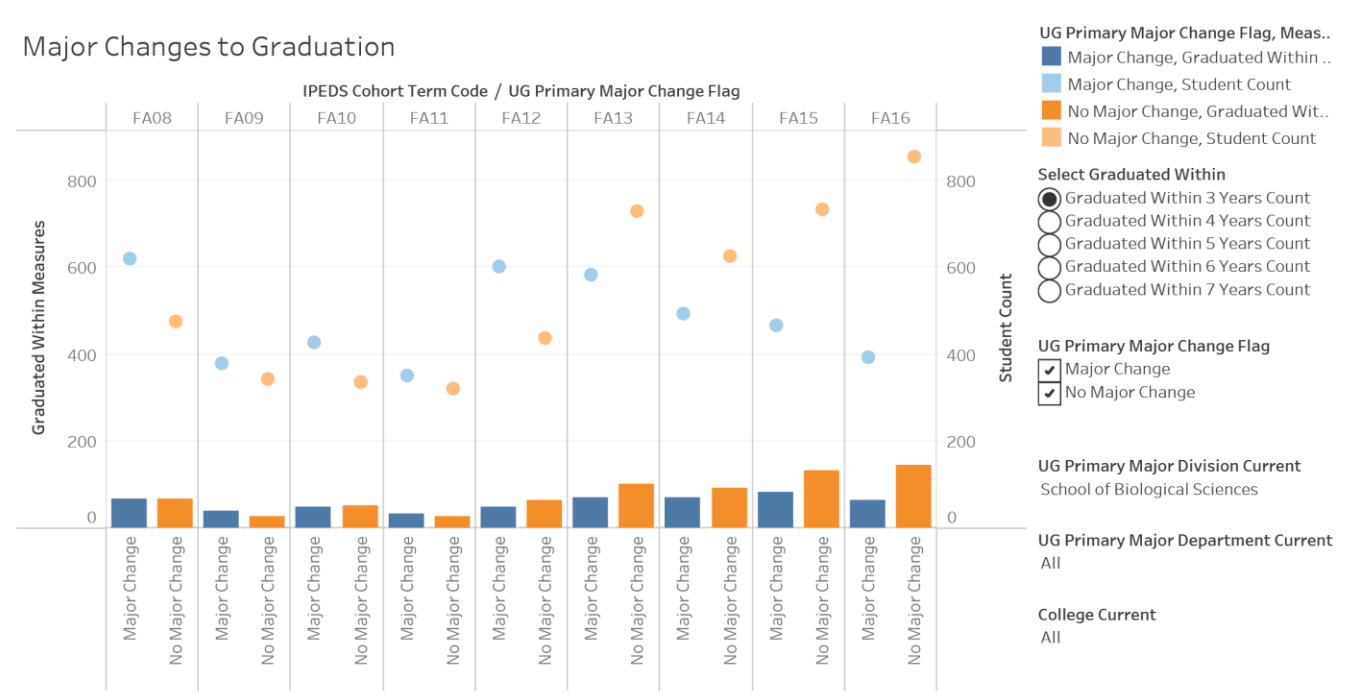
RETENTION AND ENTRANCE SCORES EXAMPLE

5-year (orange line) and 4-year graduation rates (blue lines) by high school GPA (binned)



MAJOR SWITCHING EXAMPLE

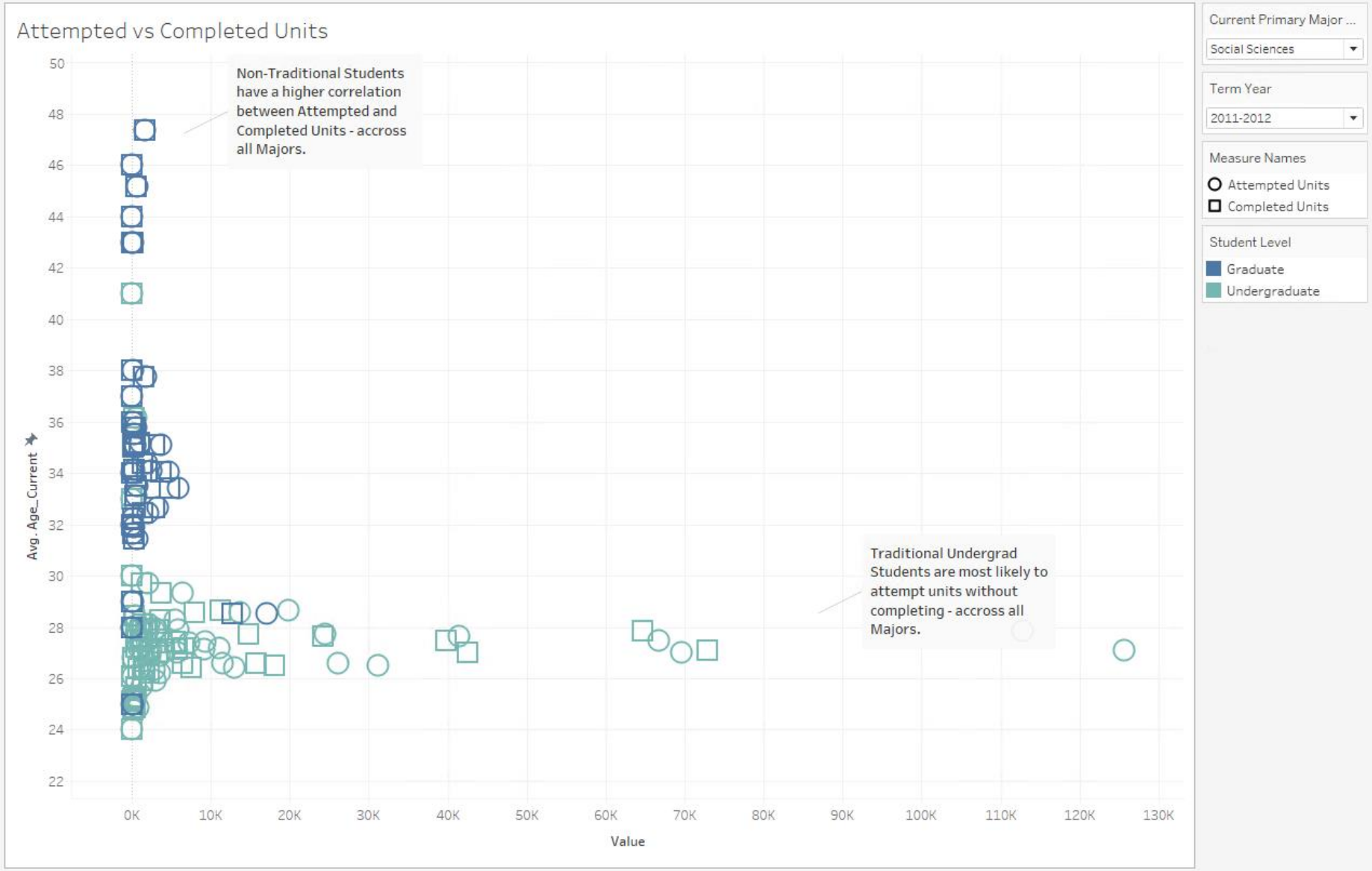
IPEDS cohort dashboard comparing students who changed their major versus those who did not across different graduation time frames



IPEDS Cohort Term Code	UG Primary Major Change Flag	Student Count	Graduated Within 3 Years Count	Graduated Within 4 Years Count	Graduated Within 5 Years Count	Graduated Within 6 Years Count	Graduated Within 7 Years Count
FA08	Major Change	620.0	69.0	378.0	547.0	568.0	575.0
	No Major Change	476.0	69.0	351.0	417.0	427.0	429.0
FA09	Major Change	380.0	40.0	267.0	352.0	359.0	361.0
	No Major Change	344.0	29.0	248.0	309.0	312.0	313.0
FA10	Major Change	428.0	51.0	293.0	396.0	415.0	416.0
	No Major Change	337.0	53.0	260.0	319.0	320.0	321.0
FA11	Major Change	352.0	34.0	219.0	301.0	312.0	319.0
	No Major Change	322.0	27.0	218.0	276.0	280.0	280.0
FA12	Major Change	602.0	51.0	432.0	532.0	546.0	550.0

ATTEMPTED VERSUS EARNED CREDITS (UNITS) EXAMPLE

Correlation analysis between Attempted Units and Completed Units, by student age, primary major, term year and student level.



CANVAS LMS EVENTS BY DAY OF TERM EXAMPLE

The ‘heart beat’ of LMS activity

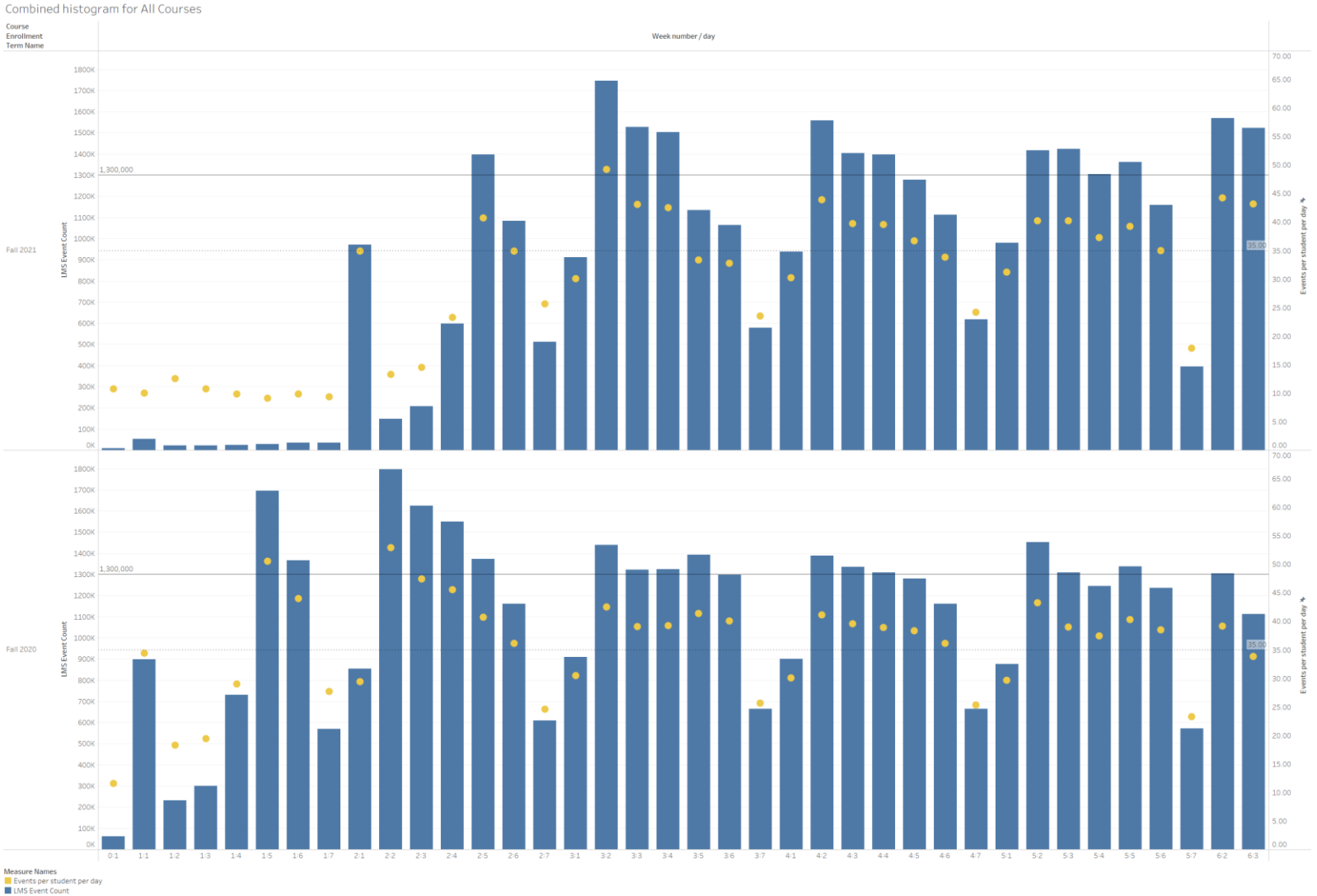
Fall 2020 – all classes remote

Fall of 2021 – back in the class, but with hybrid options

Yellow dot: average Canvas events per day per student

Blue bar, total events overall per day

Canvas and Kaltura activity activity table size: 1.6 billion rows!

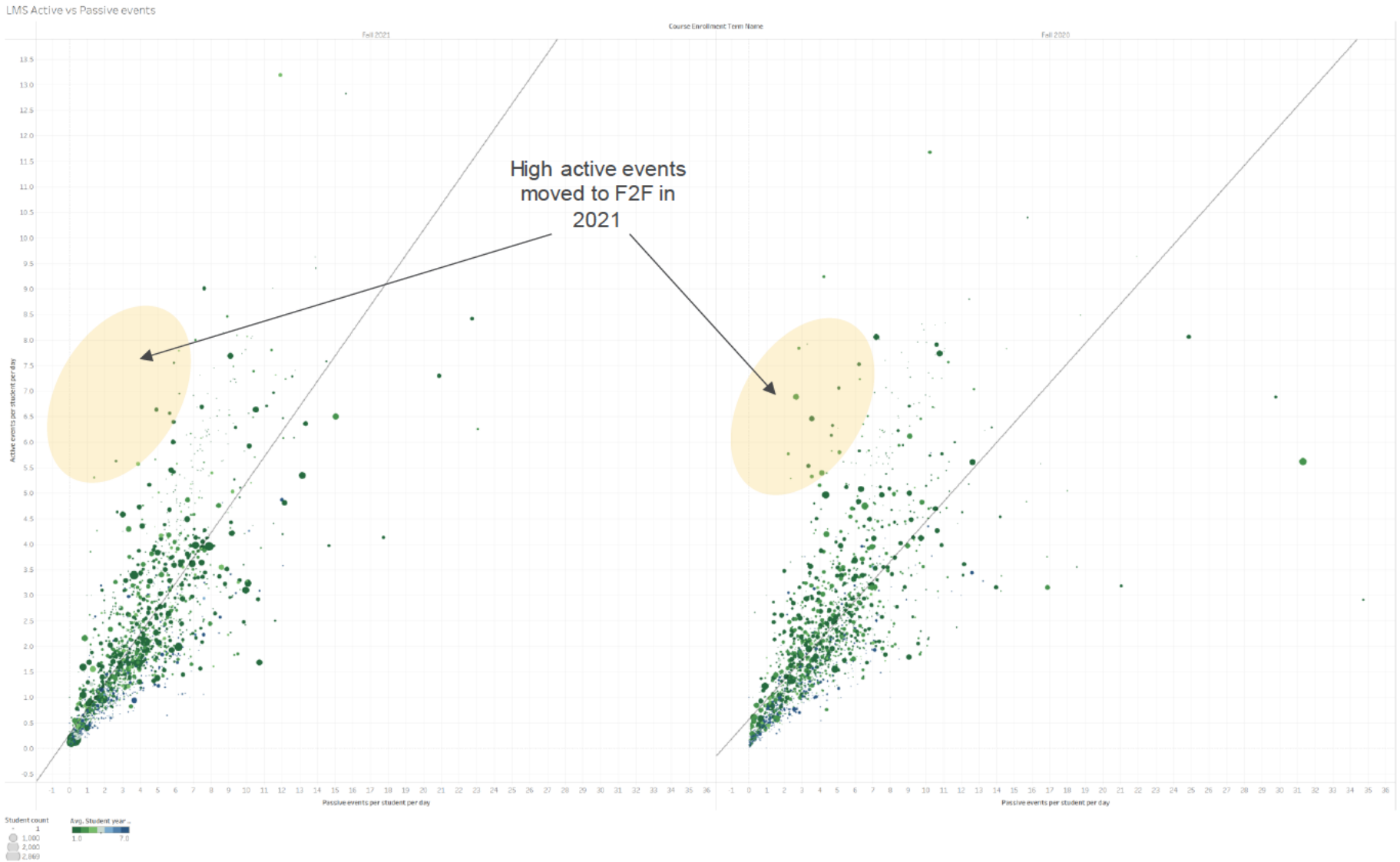


CANVAS LIVE EVENTS ANALYSIS EXAMPLE

Canvas events by day Fall 2021 vs Fall 2020

Active vs Passive events

Active events include things like assignments submitted or done online, discussion posts, searching and tagging pages. Passive events include things like viewing course announcements, syllabi, etc. Course interactivity is slightly more active this term (slope of the line is steeper for Fall 2021), although a small number of courses with high interactivity has dropped (orange oval), which is probably appropriate



CANVAS LIVE EVENTS BY DAY OF WEEK AND HOUR EXAMPLE

Count of active students and student activity in Canvas LMS by day of the week and hour of the day, for a specific course

Total enrollment	First-gen students	Transfer students	International students
487	135	478	98

Active Students by Day and Hour

for MMW121R, FA20

Hour of..	Sunday	Monday	Tuesday	Wednesd..	Thursday	Friday	Saturday
0	187	153	228	231	238	196	197
1	114	113	171	152	155	140	164
2	71	89	150	122	116	99	129
3	59	73	114	103	96	78	103
4	44	56	99	84	74	67	80
5	41	66	110	71	64	65	73
6	39	64	108	82	79	75	88
7	81	124	135	123	112	104	109
8	96	212	166	174	163	149	126
9	115	214	238	189	210	200	153
10	123	232	274	256	308	218	202
11	161	268	309	301	319	234	239
12	195	274	316	283	327	235	286
13	193	290	323	265	312	245	283
14	199	298	324	278	306	243	285
15	207	315	323	265	343	226	294
16	211	329	324	258	326	253	327
17	213	312	338	286	342	237	341
18	220	313	319	273	350	219	345
19	244	289	324	298	349	258	364
20	224	296	342	312	357	261	368
21	246	298	372	328	349	270	368
22	234	293	354	330	354	247	352
23	209	268	320	290	315	241	305

% of Total N student

8.01% 76.39%

Course Activity Count by Day and Hour

for MMW121R, FA20

Hour of..	Sunday	Monday	Tuesday	Wednesd..	Thursday	Friday	Saturday
0	1,456	1,187	1,925	1,968	2,070	1,682	1,436
1	806	888	1,595	1,254	1,302	944	1,156
2	530	647	1,130	856	738	791	925
3	344	384	894	730	615	392	610
4	272	363	610	539	501	349	552
5	359	392	818	508	634	423	491
6	277	478	930	571	645	427	768
7	471	1,326	955	706	880	580	922
8	661	2,636	1,446	1,095	1,265	1,040	950
9	748	2,131	1,761	1,517	1,766	1,294	1,299
10	850	2,507	2,829	2,157	2,886	1,687	1,473
11	1,050	3,116	3,541	2,749	3,527	1,696	1,972
12	1,168	3,024	2,989	2,436	3,206	1,781	2,288
13	1,286	2,933	3,350	2,089	3,188	1,745	2,633
14	1,221	2,792	3,677	2,326	2,780	1,747	2,795
15	1,572	3,495	3,555	2,248	3,458	1,726	3,015
16	1,714	3,427	3,175	2,062	2,966	1,877	3,094
17	1,946	2,942	3,300	2,405	3,160	1,477	3,310
18	1,780	2,875	2,986	2,249	3,124	1,570	3,856
19	1,617	2,670	3,145	2,500	2,809	1,816	4,133
20	1,813	2,965	3,731	2,486	3,071	1,909	4,291
21	1,965	2,582	4,344	2,971	3,346	2,222	4,383
22	1,954	2,595	3,938	2,968	3,527	1,927	4,210
23	1,602	2,644	3,491	2,793	3,447	1,738	3,553

Course activity count

272 4,383

Course ID T

MMW121R

Course Term Code

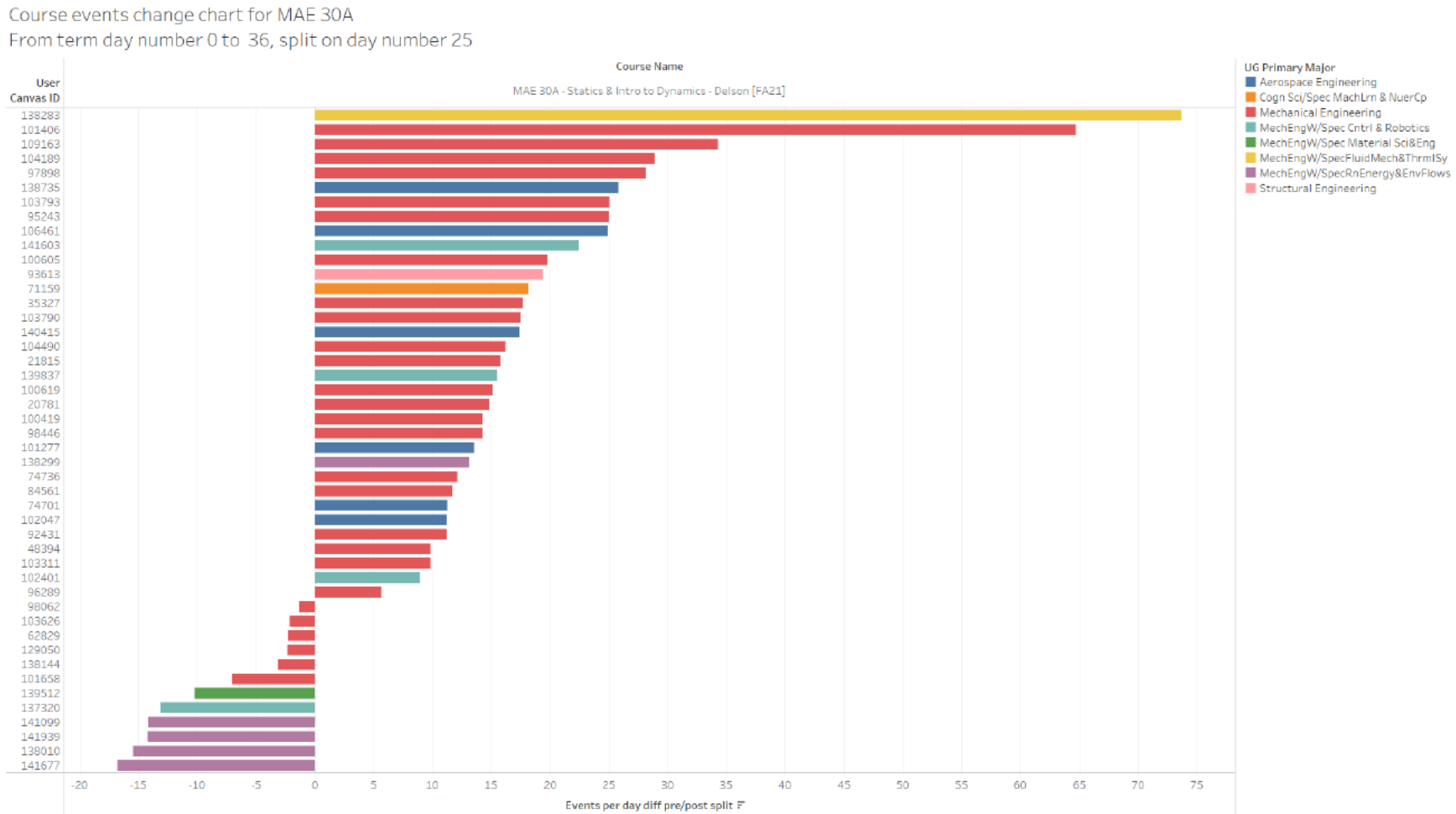
FA20

Course activity name

- (All)
- Announcement viewed
- Assignment viewed
- Course collaborations viewed
- Course grade viewed
- Course outcome viewed
- Course quizzes viewed
- Course roster viewed
- Course speed grader viewed
- Course topics viewed
- Course updated
- File viewed
- Module viewed
- Not labeled activity
- Page content search viewed
- Page viewed
- Syllabus viewed

CANVAS LIVE EVENTS AND STUDENT DISENGAGEMENT EXAMPLE

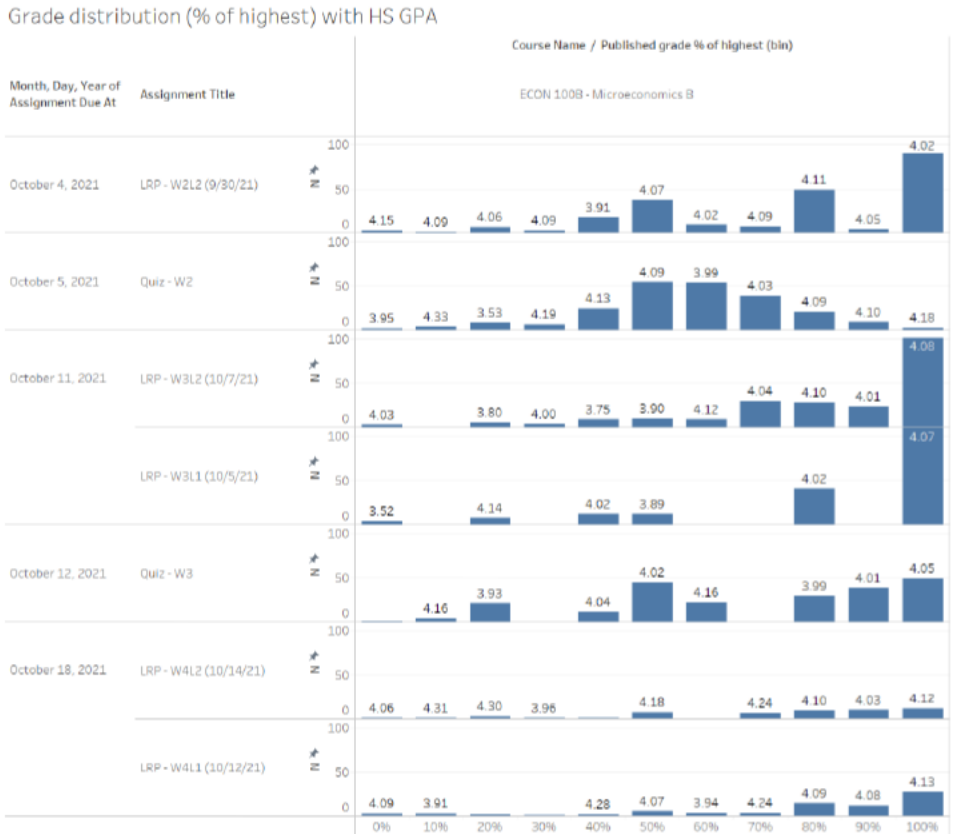
Pre-post split, which students are decreasing versus increasing their interactivity, comparing early in the term to later in the term, showing the undergraduate major (color)

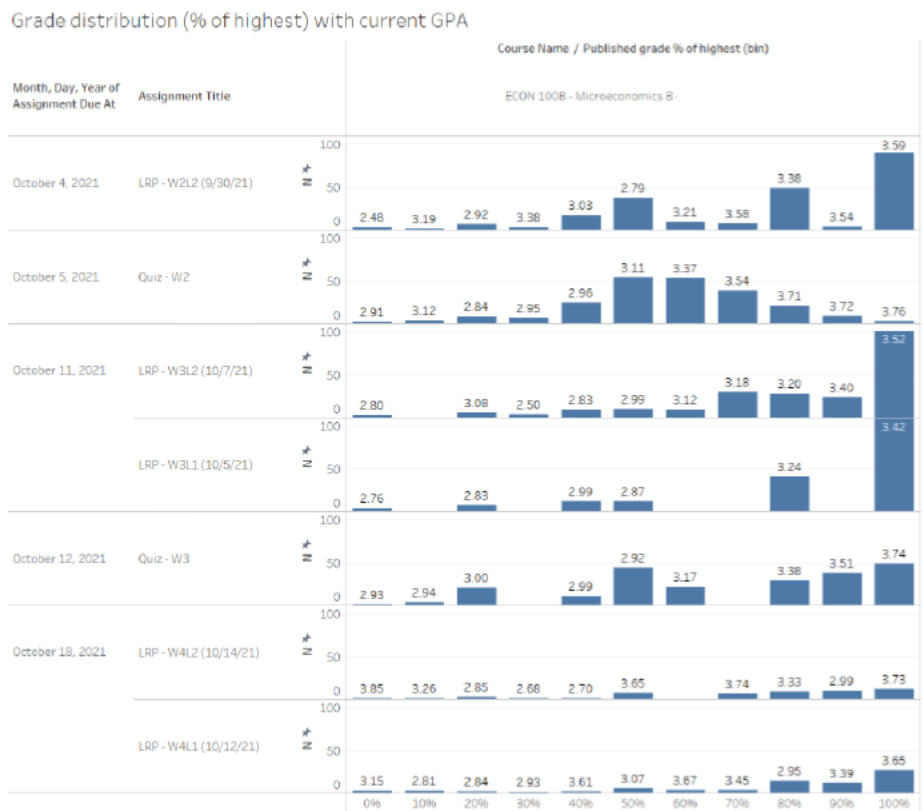


CANVAS ASSIGNMENT AND SIS DATA EXAMPLE

Assignments in canvas grades annotated with student HS GPA (left) and current university GPA (right)

Joining of SIS data and Canvas data





CANVAS DATA AND RETENTION METRICS EXAMPLE

Students higher number of Canvas late assignment submissions and retention, progression and GPA metrics

(Sample data)

Submission & Progression Summary

	Excessive late Canvas submissions		
	Yes	No	Grand Total
Student Count	212	4,487	4,699
Progressed terms per student	6.39	8.59	8.78
Passed units (credits) per student	4.4	15.1	9.8
1 year retention %	87.3%	86.4%	86.5%
2 term retention %	93.4%	97.2%	97.1%
Avg. GPA Current Cumulative	3.61	3.70	3.70

SUMMARY

- Over the past seven years, we worked hard to design the Student Activity Hub to seamless integrates data from our student information system (SIS) with Canvas data and Canvas Live events
- The SAP HANA platform makes short work of very large table sizes, billions of rows of Caliper Events. For us this includes Canvas, Kaltura, OpenEdX and our advising system (VAC). It speeds up analysis by our users and it also dramatically speeds up our own development
- With our platform, we have found that the linkages between a SIS and Canvas are fairly straightforward. The inclusion of Canvas data (nightly data and real-time Live Events) is also simple. We can handle data from a variety of systems and we can handle different integration demands
- The new rules architecture unlocks the potential in the data. It easy for analysts, who do not have to fashion core metrics. The user never has to join data. All this is handled in the SAH views
- Canvas and SIS data is a match made in heaven!

Questions?

Fedot Sychkov. *Collective Farm Market.* 1936
<https://www.facebook.com/moscowart/photos/a.471911309576816.1073742330.276754345759181/654936557940956/>

FOLLOW-UP

People you can contact

- Kevin Chou, Deputy CIO (kcchou@ucsd.edu)
- Greg Stinsa, Collaborations Manager (gstinsa@ucsd.edu)
- Joe Salwach, Collaborations Support (jsalwach@ucsd.edu)
- Vince Kellen, CIO, UC San Diego (vkellen@ucsd.edu)

Partners

- Moran Technology Consulting (https://www.morantechnology.com)
- Tata Consultancy Services (https://www.tcs.com)
- Slower Inc. (https://slower.ai)

Web site

https://studentactivityhub.com

Upcoming webinars

- https://csc.ucsd.edu/webinars/index.html

Additional slides

RECENT CANVAS EVENT COUNTS

Canvas Events

LMS Event Category L1	LMS Event Category L2	LMS Event Name	
Announcements	Announcements	Announcement viewed	1,868,112
Assessments	Assessment	Assessment grade created	3,184,147
		Assessment quiz viewed	8,774,536
		Assessment submitted	1,355,748
Assignments	Assignments	Assignment created	15,960
		Assignment edited	88,887
		Assignment grade override created	7,398
		Assignment grade override edited	15,299
		Assignment submitted	1,264,657
		Assignment viewed	22,511,415
Collaborations	Collaborations	Collaboration viewed	9,151
Document	Document	Document viewed	24,866,699
Enrollment	Course enrollment	Course enrollment created	435,465
		Course enrollment state created	12
		Course enrollment updated	70,315
		Course enrollment viewed	80,937
Forums	Forum thread	Forum thread created	30,445
		Forum thread viewed	10,385,738
Groups	Groups	Group announcements viewed	23,165
		Group calendar feed viewed	4,695,702
		Group category created	980
		Group collaborations viewed	129
		Group created	9,569
		Group files viewed	26,300
		Group membership created	34,051
		Group pages viewed	21,883
		Group roster viewed	36,961
		Group topics viewed	83,708
Media	Files	File viewed	1,296,918
Messages	Messages	Message sent	298,127
Module	Module	Module viewed	9,246,870
Other Event	Attachment	Attachment created	884,126
		Attachment deleted	13,355
		Attachment modified	5,922
	External tool	External tool accessed	7,022,495
	LTI tool	LTI tool accessed	61,424
	Submission	Submission modified	4,948,590
Pages	Section/course page	Calender event viewed	9,042
		Calender feed viewed	31,549,730
		Course collaborations viewed	196
		Course grade viewed	3,330,962
		Course outcome viewed	480
		Course quizzes viewed	737,358
		Course roster viewed	628,312
		Course speed grader viewed	246,285
		Course topics viewed	1,619,229
		Page content search viewed	37,151,629
		Page tagged	807,741
		Page viewed	65,347
	Wiki page	Wiki page created	5,044
		Wiki page deleted	1,818
		Wiki page updated	29,045
	Wikipage	Wiki page viewed	8,331,977
Session	Session	User log in	3,109,893
		User log off	56,329
Syllabi	Syllabi	Syllabus viewed	1,035,280
System	Course	Course updated	5,542
	User	User account association created	284,584
		User calendar feed viewed	6,849,972
		User files viewed	23,558
Grand Total			199,554,549

SAP HANA and Activity Hubs

UC San Diego Activity Hubs

For trained decision makers and analysts across campus

Designed for "blended" analysis and data curation

Works in tandem with app-specific reports

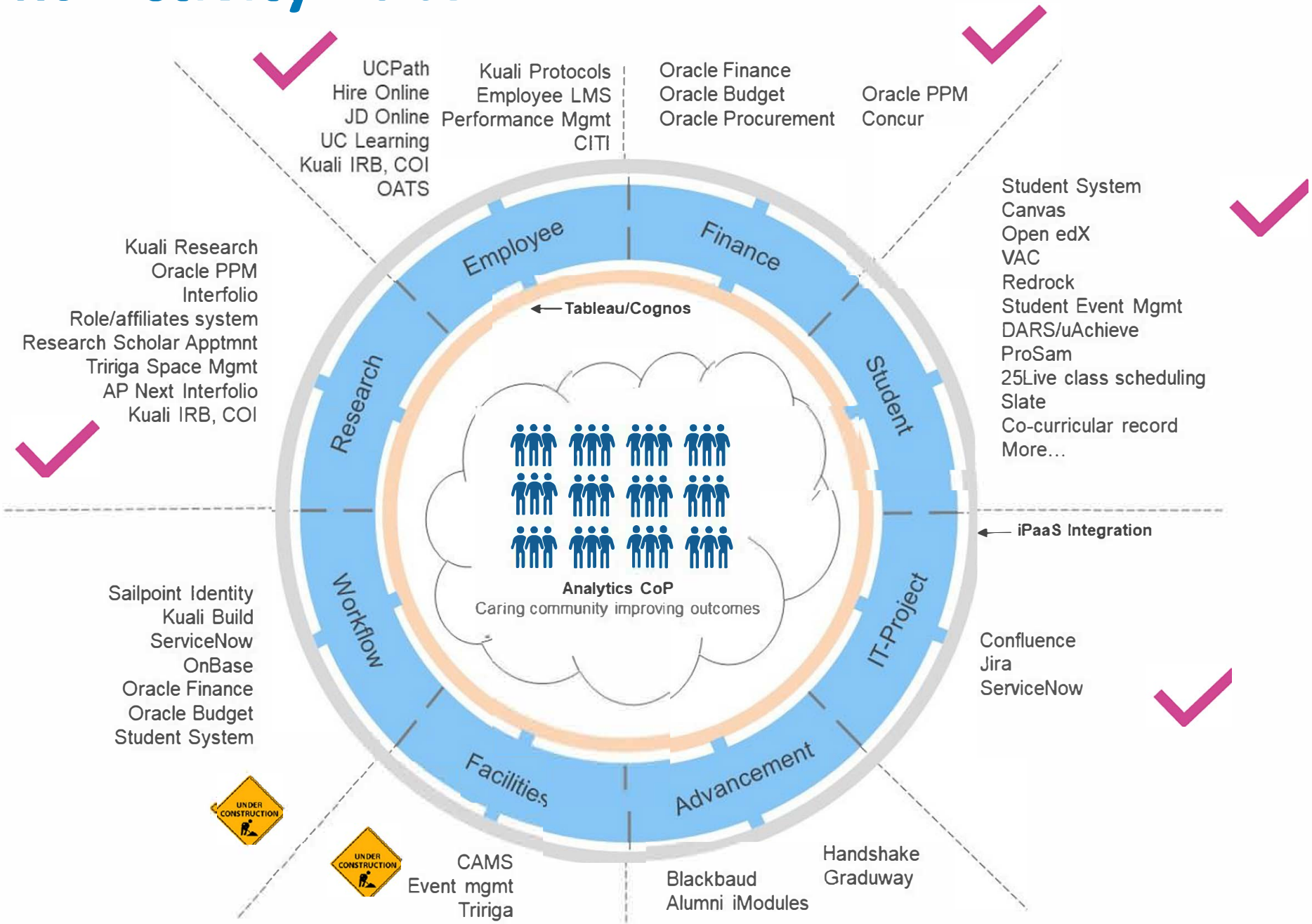
Community-driven environment

Secure. Private. Safe.

Includes data from all major systems

Local systems access via secure integration (iPaaS)

Integrated with mobile and advanced analytic & ML technologies



Platform predictive capabilities

Classification Analysis
- CART
- C4.5 Decision Tree Analysis
- CHAID Decision Tree Analysis
- K Nearest Neighbour
- Logistic Regression Elastic Net
- Back-Propagation (Neural Network)
- Naïve Bayes
- Support Vector Machine
- Random Forests
- Gradient Boosting Decision Tree
- Linear Discriminant Analysis (LDA)
- Confusion Matrix
- Area Under Curve (AUC)
- Parameter Selection/Model Evaluation

Regression
- Multiple Linear Regression Elastic Net
- Polynomial, Exponential, Bi-Variate Geometric, Bi-Variate Logarithmic Regression
- Generalized Linear Model
- Cox Proportional Hazards Model

Cluster Analysis
- ABC Classification
- DBSCAN
- K-Means/Accelerated K-Means
- K-Medoid Clustering
- K-Medians
- Kohonen Self-Organized Maps
- Agglomerate Hierarchical
- Affinity Propagation
- Latent Dirichlet Allocation (LDA)
- Gaussian Mixture Model (GMM)
- Cluster Assignment

Time Series Analysis
- Single/Double/Brown/Triple Exponential Smoothing
- Forecast Smoothing
- Auto – ARIMA/ Seasonal ARIMA
- Croston Method
- Forecast Accuracy Measure
- Linear Regression with Damped Trend and Seasonal Adjustment
- Test for White Noise, Trend, Seasonality
- Fast Fourier Transform (FFT)
- Correlation Function

Association Analysis
- Apriori
- Apriori Lite
- FP-Growth
- KORD – Top K Rule Discovery
- Sequential Pattern Mining

Probability Distribution
- Distribution Fit/Weibull analysis
- Cumulative Distribution Function
- Quantile Function
- Kaplan-Meier Survival Analysis

Outlier Detection
- Inter-Quartile Range Test (Tukey's)
- Variance Test
- Anomaly Detection
- Grubbs Outlier Test

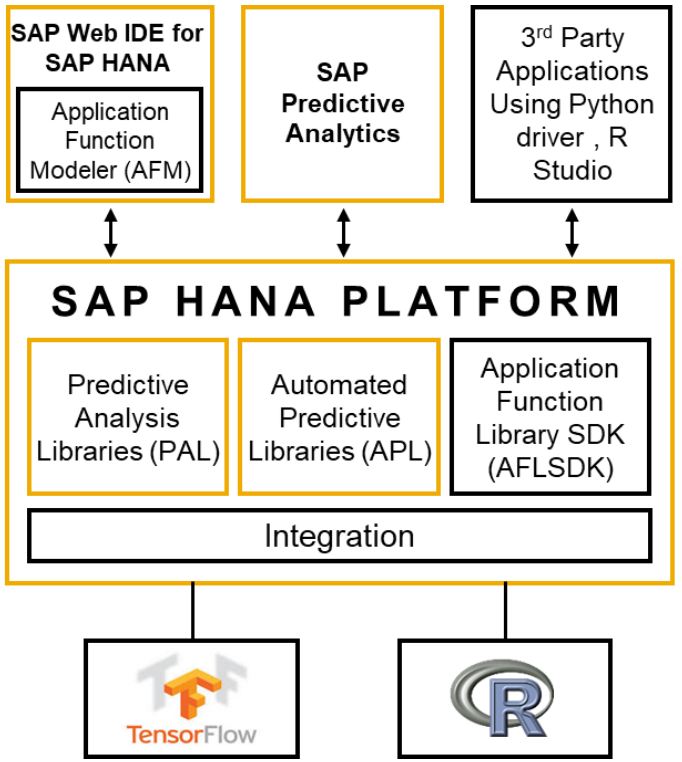
Recommender
- Factorized Polynomial Regression Models

Link Prediction
- Common Neighbors
- Jaccard's Coefficient
- Adamic/Adar
- Katzβ

Statistical Functions
- Mean, Median, Variance, Standard Deviation, Kurtosis, Skewness
- Covariance Matrix
- Pearson Correlations Matrix
- Chi-squared Tests:
  - Test of Quality of Fit
  - Test of Independence
- F-test (variance equal test)
- Data Summary
- ANOVA
- One-sample Median Test
- T Test
- Wilcox Signed Rank Test

Data Preparation
- Sampling
- Binning
- Scaling
- Partitioning
- Principal Component Analysis (PCA)/ PCA Projection

Other
- Weighted Scores Table
- Substitute Missing Values



- 90+ prepackaged machine learning/predictive algorithms
- Supports association, clustering, classification, regression, time series, ...
- Supports different types of data – structured, streaming and series data
- Real-time scoring for several algorithms
- Integrated with open source machine learning libraries – TensorFlow and R
- Integration with other tools via data virtualization or other access, e.g., Google Big Query, Google TPUs, AWS Trainium, etc.

UC San Diego