



Next Steps for Big Data in Education

Chris Dede

Harvard University

Chris_Dede@harvard.edu

www.gse.harvard.edu/faculty/christopher-dede

My Current Role in Data-Intensive Research in Education

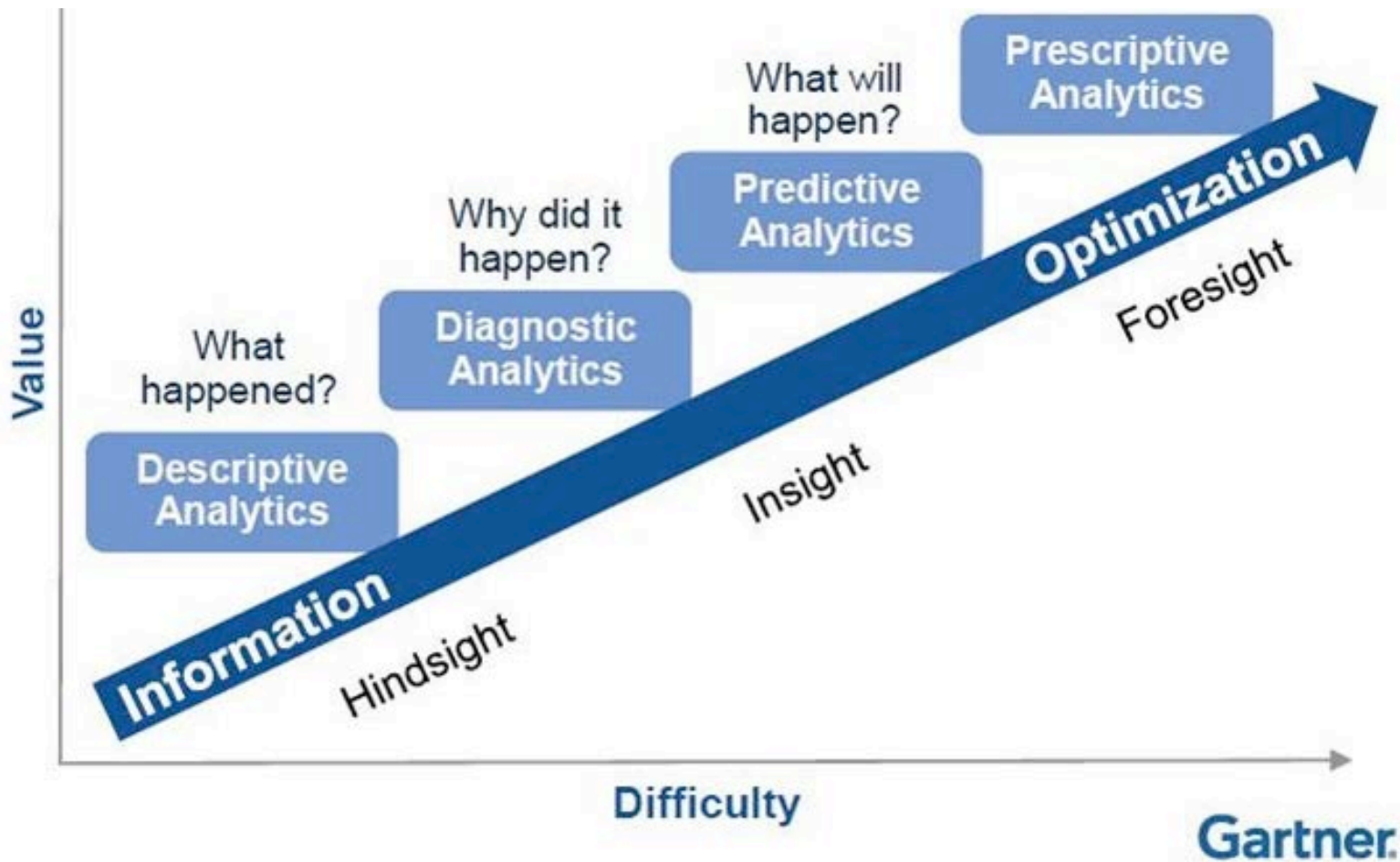
- Organized a two workshop sequence on data-intensive research for NSF and the field: insights from relatively mature data-intensive research initiatives in the sciences and engineering were applied to nascent data-intensive research efforts in education
- Confront “big data” issues in my design-based research in ecosystems science education

Definitions

- Big Data is characterized by the ways in which it allows researchers to do things not possible before (i.e., Big data enables the discovery of new information, facts, relationships, indicators, and pointers that could not have been realized previously).
- Data-intensive research involves data resources that are beyond the storage requirements, computational intensiveness, or complexity that is currently typical of the research field.
- Data science is the large-scale capture of data and the transformation of those data into insights and recommendations in support of decisions.



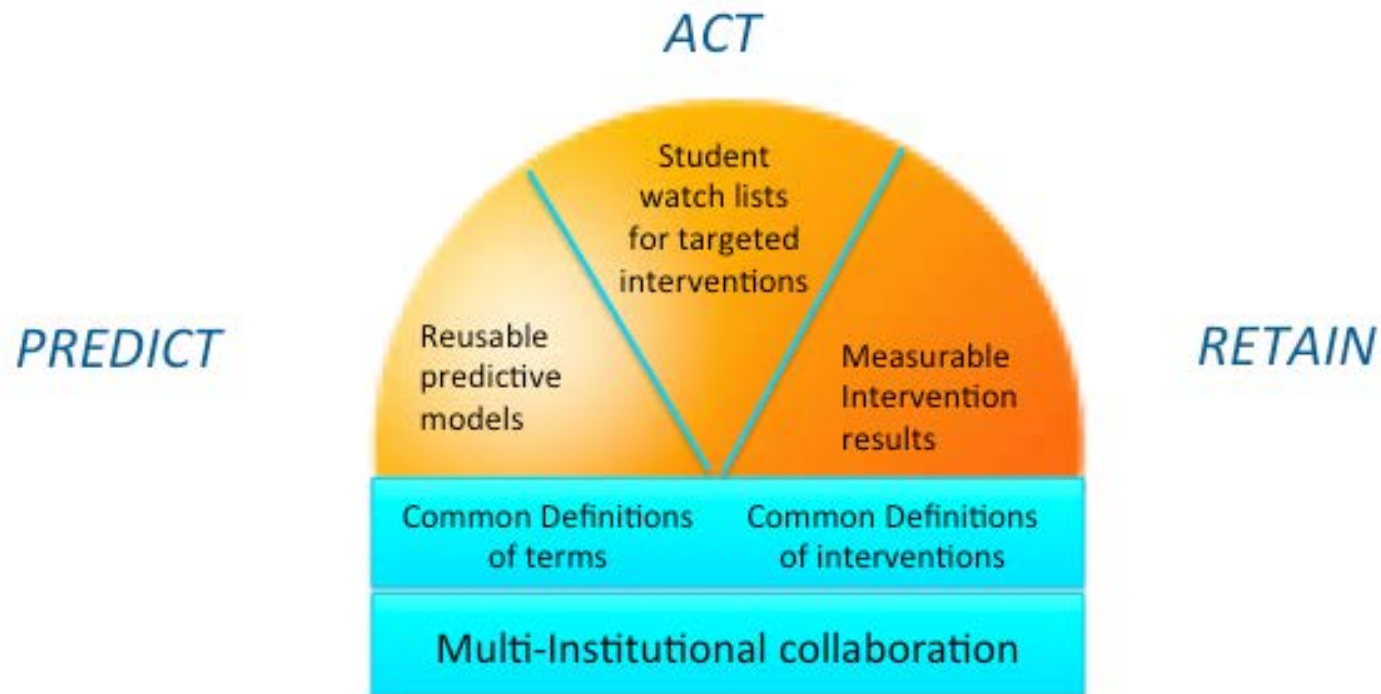
From Hindsight to Foresight



From Description to Prescription

- Determine students' probabilities of failure (*predictions*)
- Determine which students respond to which interventions (*uplift modeling*)
- Determine which interventions are most effective (*explanatory modeling*)
- Allocate resources accordingly (*cost benefit analysis*)

PAR Framework



Scalable, cross-institutional improvements that support each individual student's success

PAR Data Inputs

Student Demographics & Descriptive

Gender
Race
Prior Credits
Perm Res Zip Code
HS Information
Transfer GPA
Student Type

Student Course Information

Course Location
Subject
Course Number
Section
Start/End Dates
Initial/Final Grade
Delivery Mode
Instructor Status
Course Credit

Student Financial Information

FAFSA on File – Date
Pell Received/Awarded – Date

Student Academic Progress

Curent Major/CIP
Earned Credential/CIP

Course Catalog

Subject
Course Number
Subject Long
Course Title
Course Description
Credit Range

Lookup Tables

Credential Types Offered
Course Enrollment Periods
Student Types
Instructor Status
Delivery Modes
Grade Codes
Institution Characteristics

Possible Additional **

Placement Tests
NSC Information
SES Information
Satisfaction Surveys
College Readiness Surveys
Intervention Measures

** Future

Descriptive and Predictive Insight

PAR Benchmarks Descriptive Analytics

Cross Institutional

Student/degree/major level insight into:

1. What **did** the retention look like for students entering in the same cohort
2. How does your institution compare to peer institutions / institutions in other sectors
3. How **did** performance vary by student attributes

PAR Models Predictive Analytics

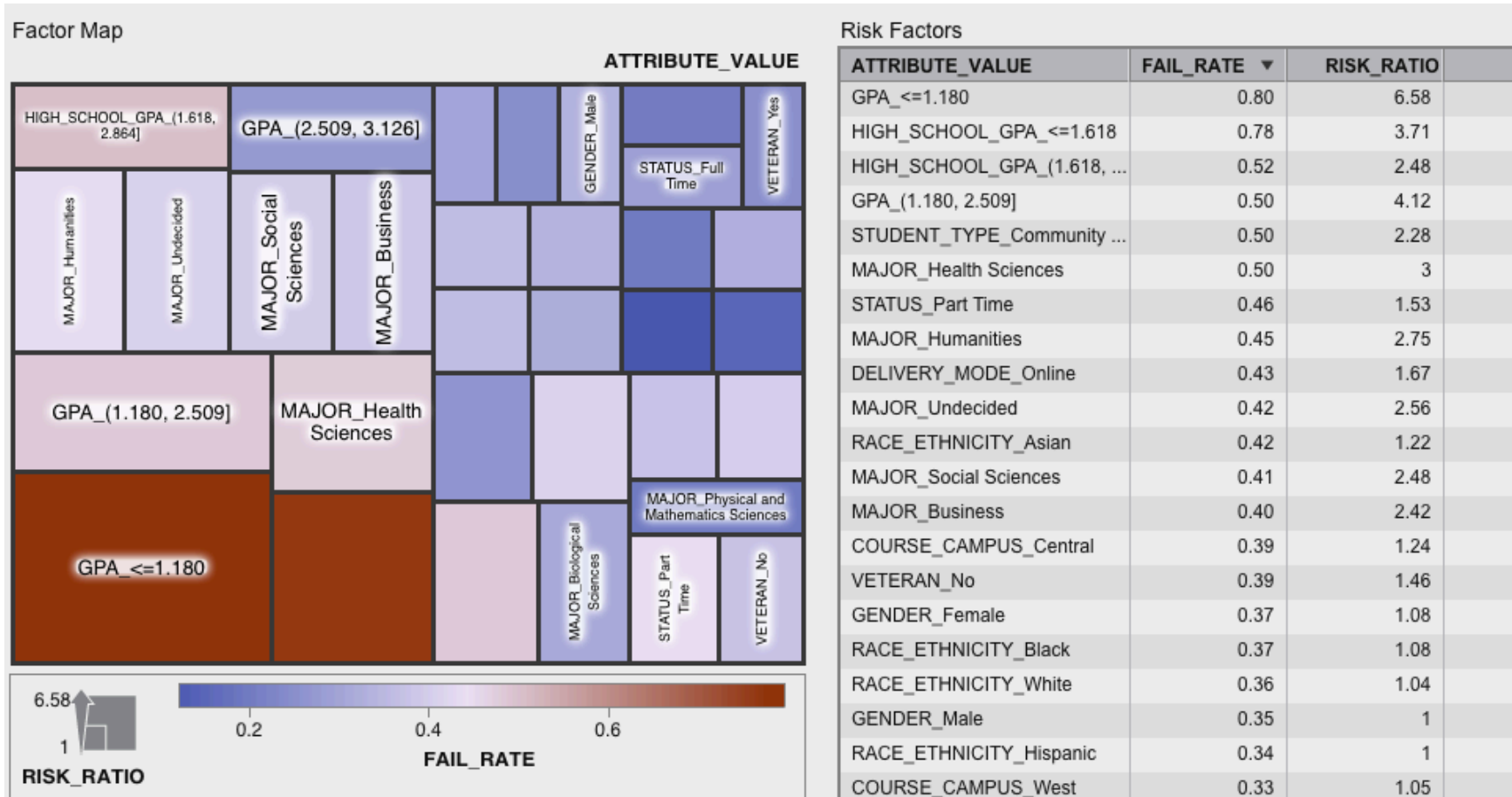
Institutional Specific insight into:

1. What students are being retained over time?
2. Which students are currently at risk for completing and why?
3. Which factors are directly correlated to student success?
4. What is the predicted course completion rate for a particular program?

Outcome Measures

- Credits Earned
- Credits Earned- No Dev Ed
- Credit Ratio
- Credit Ratio- No Dev Ed
- Retention
- Credentials Earned (Any, Bachelors, Associates, Certificate)

Predicting retention aimed at taking action - finding the most important factors



Key Research Questions

- Can we detect problems that students are having *as they are happening*, through automated analysis?
- Can we provide real-time feedback to students and educators in response to the problem detection?
- Is the feedback effective in helping students attain more sophisticated behaviors? Does it *make sense* to the students and educators? Is it *actionable* in that they are able to do something useful with it?

Key Next Steps

- Mobilize Communities around Opportunities based on New Forms of Evidence
- Develop New Forms of Educational Assessment
- Develop New Types of Analytic Methods
- Build Human Capacity to Do and to Understand Data Science
- Develop Advances in Privacy, Security, and Ethics
- Infuse Evidence-based Decision-Making throughout Organizations and Systems

3 E's of Immersive Learning

- **Engagement**
Students are motivated to do well, see the relevance of their learning, and increase in self-efficacy
- **Evocation**
Immersive interfaces can evoke a wide spectrum of authentic performances with embedded support
- **Evidence**
Log files, chat logs, shared notebooks, and similar artifacts provide a rich evidentiary trail



EcoMUVE – Multi-User Virtual Environment



Drag

Connect

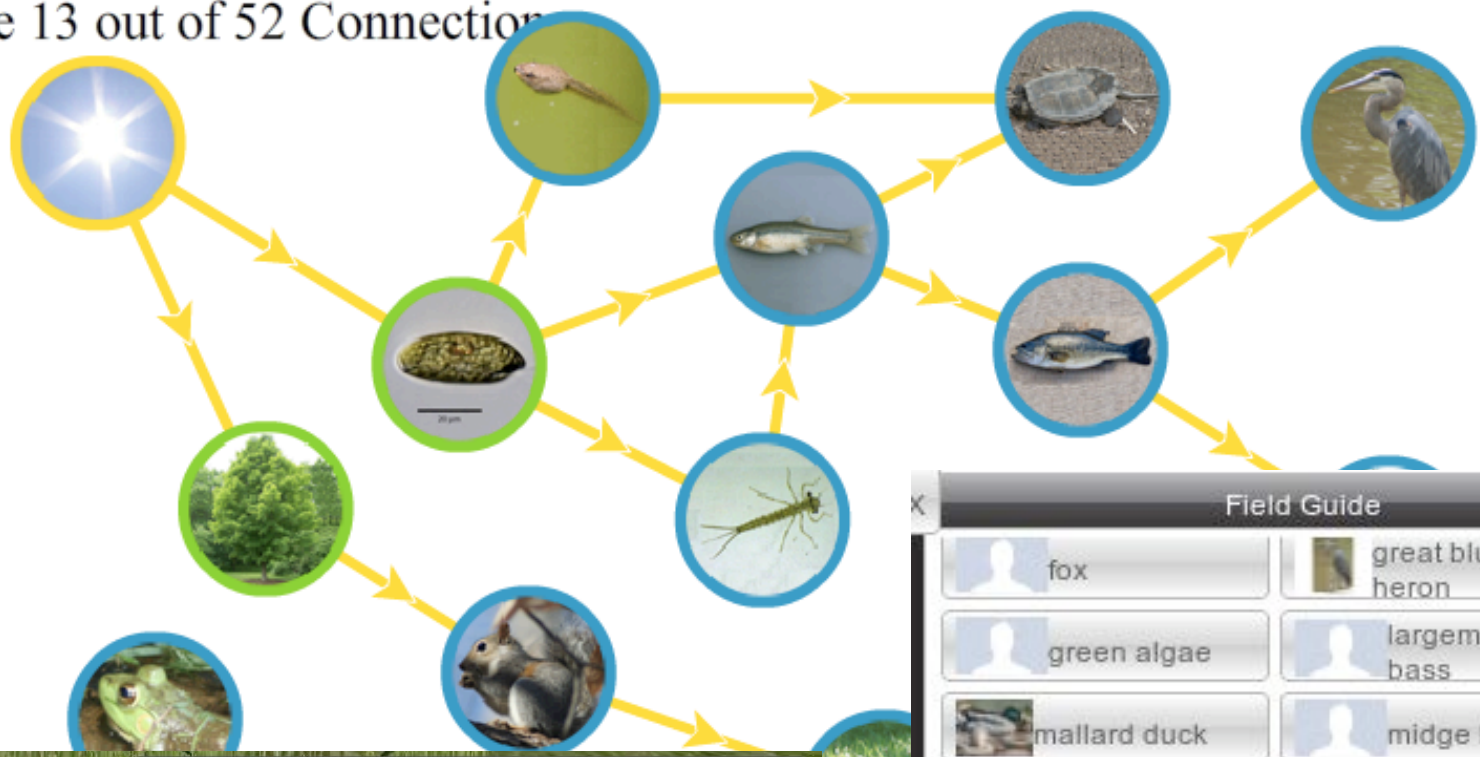
Disconnect

Check

Print

Field Guide

You have 13 out of 52 Connection



snapping turtle

Save

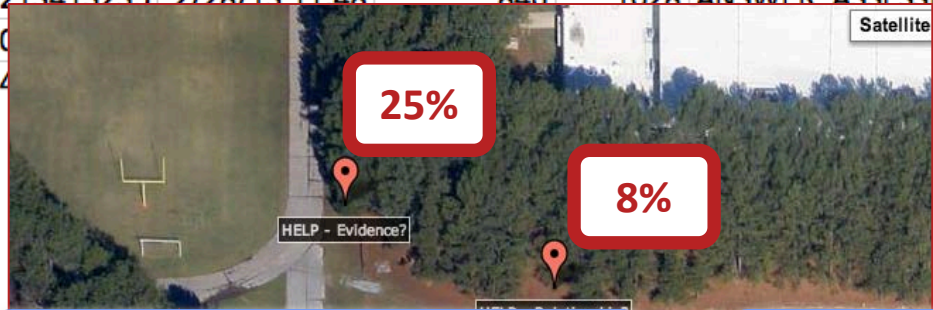
Cancel

Field Guide

fox	great blue heron
green algae	largemouth bass
mallard duck	midge larvae
mosquito larvae	predaceous diving beetle
protist	red-tailed hawk
rotifer	snail
snapping turtle	squirrel
sugar maple	tadpole
water chestnut	white pine

Log File Data

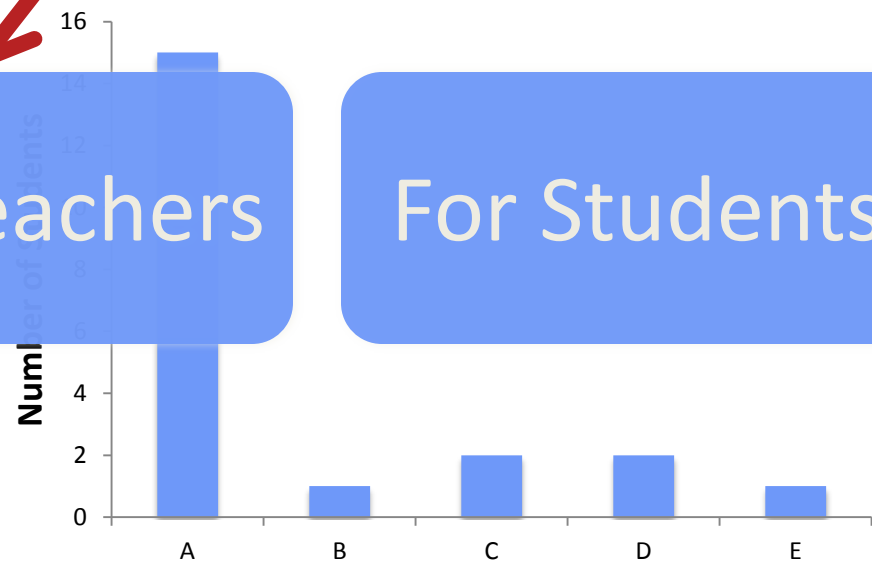
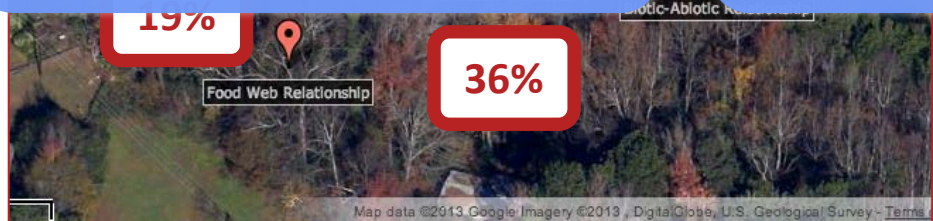
timestamp	latitude	longitude	event	details
196451875 2/28/13 11:47	33.890183	-84.272432	Before you	are there any bugs
196451875 2/28/13 11:47	33.8901883	-84.272389	Before you	are there any animals
196451875 2/28/13 11:47	33.8902795	-84.272475	Before you	do animals live there?
196451875 2/28/13 11:47	33.8902795	-84.272363	Before you	do bugs live there ?
196451875 2/28/13 11:47	33.8902259	-84.272304	Before you	do they die fast
196451875 2/28/13 11:47	33.8902313	-84.272389	Before you	does an animal live th
196451875 2/28/13 11:47	33.8900381	-84.272309	Before you	does it rain alot.
196451875 2/28/13 11:47	33.8902205	-84.272443	Before you	how could animals sur
196451875 2/28/13 11:47	33.8899791	-84.271837	Before you	how do animals live in
196451875 2/28/13 11:47	33.8902634	-84.272266	Before you	how do you know wha
196451875 2/28/13 11:47	33.8902795	-84.272427	Before you	how does it survive?
196451875 2/28/13 11:47	33.8902473	-84.272534	Before you	how many
196451875 2/28/13 11:47	33.890225	-84.272438	Before you	how many plants does
196451875 2/28/13 11:47	33.8902713	-84.272373	Before you	how old is the baby pi
196451875 2/28/13 11:47	33.8902205	-84.272481	Before you	see a lot of dead plan
196451875 2/28/13 11:47	33.8903868	-84.272711	Before you	i want to know if their



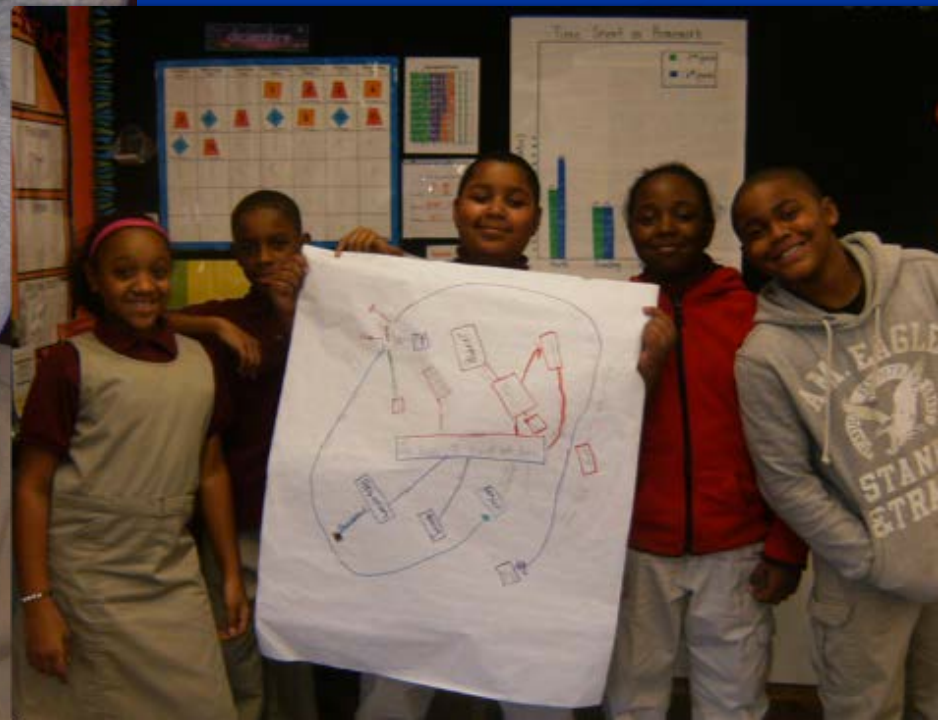
For Researchers

For Teachers

For Students



Collaborative construction of concept maps





(Conner Flynn)

Augmenting Real World Ecosystems

<http://ecomobile.gse.harvard.edu>

GoPro Cameras Capture EcoMOBILE Experience



P2: Now we need to write two things we see that could affect dissolved oxygen.

P1: Plants.

S: Plants.

P2: Plants. And...you guys would rather say...rain, or the dead matter?

S: Mm, dead matter, maybe. 'Cause—

P1: (Why the) dead matter is bacteria?

S: —Yeah. The bacteria. And we don't know, you know, how long...this has been...

P2: You got plants already? Plants, 'cause they release dissolved oxygen into the water.

P1: This could

P2: ... Provide food for bacteria, increasing their population and increasing their need for dissolved oxygen.

P1: The bacteria and —

P2: And um —

P1: And causing an increase in population.

P2: Yeah, increasing their population and their need for dissolved oxygen.

S: [Student talking to other student] Quinn.

P1: Um, provide food for bacteria, increasing population?

P2: Mhm. [Partner 1 continues typing in Evernote]

Evernote: Plants could release dissolved oxygen into the water and dead matter could provide food for bacteria, increasing the bacteria population and their need for dissolved oxygen.



EcoMUVE

- MUVES promote self-efficacy in science
- Simulate experiences otherwise impossible in school settings.
- Explore time and scale
- Opportunities to take on roles, work in teams
- Shared immersive experience that contextualizes learning and supports inquiry

(Ketelhut et al. 2010, Metcalf et al. 2011)

EcoMOBILE

- Greater fidelity and sensory richness, physical interactions with organisms and environments.
- Self-directed collection of real-world data and artifacts.
- Facilitated use of cameras, recording devices, probes, GPS, mapping, graphing, augmented reality.



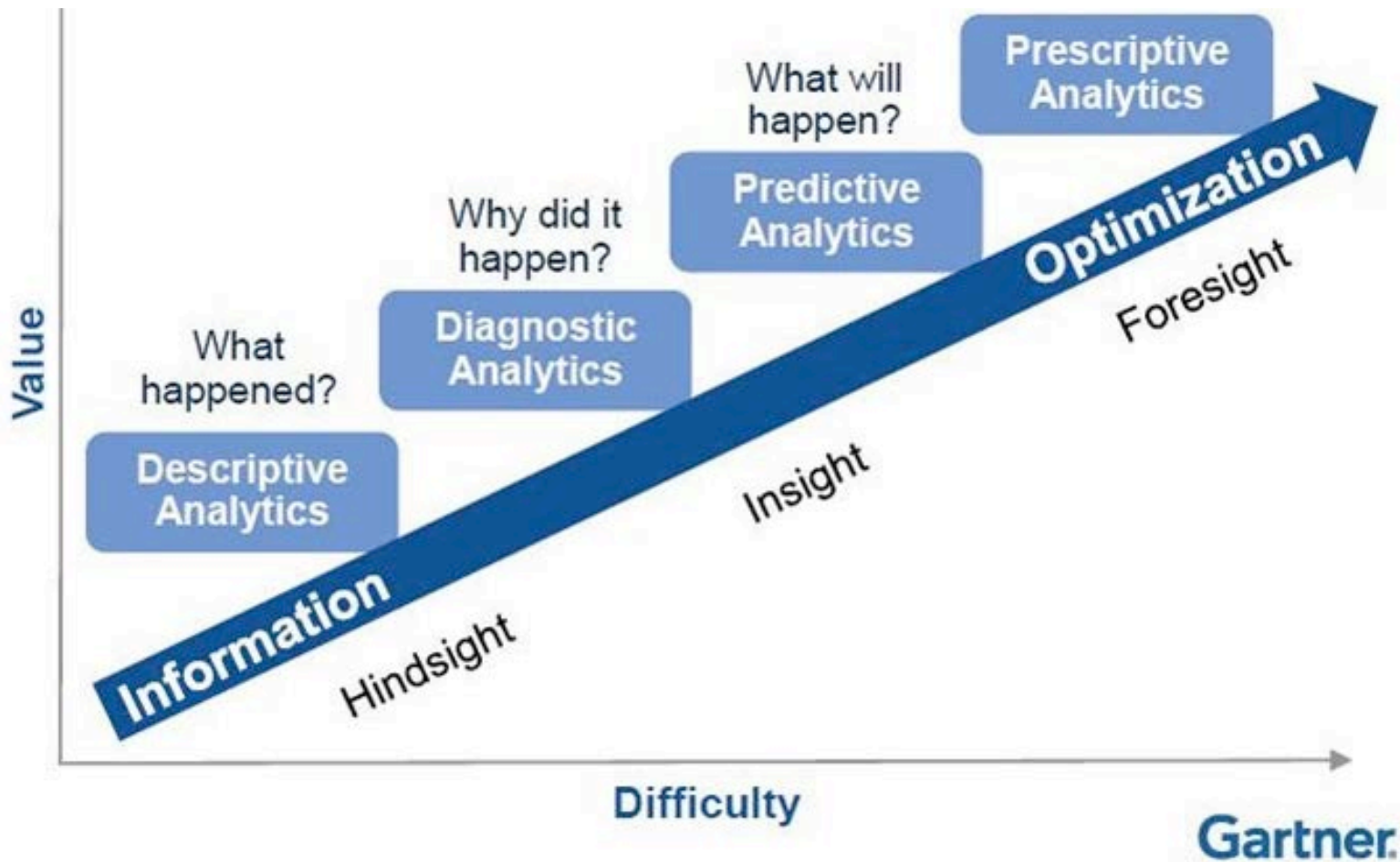
What Can We Inculcate and Assess?

- Inquiry skills?
- Collaboration?
- Leadership?
- Self-efficacy?
- Metacognition?

Key Research Questions

- Can we detect problems that students are having *as they are happening*, through automated analysis?
- Can we provide real-time feedback to students and educators in response to the problem detection?
- Is the feedback effective in helping students attain more sophisticated behaviors? Does it *make sense* to the students and educators? Is it *actionable* in that they are able to do something useful with it?

From Hindsight to Foresight



Imagining the Possible
















Edith S. Gummer
Ewing Marion Kauffman
Foundation












NSF Advancing Data Intensive Research in Education

June 1-2 2015

Macro-Level Data

District and School Information		
Type	Name	Description
	District Accreditation	Raw Data: Current accreditation status for districts.
	District Calendar Days and Hours	Raw Data: Comparison of district's calendar information.
	School District by Representative District	List of school districts by Missouri Representative District.
	School District by Senatorial District	List of school districts by Missouri Senatorial District.
	School District by Congressional District	List of school districts by United States Congressional District.
	School Districts with Congressional District	List of School Districts with their United States Congressional District.
	School Districts Map - County Boundaries	Map of Missouri school districts with county boundaries.
	School Districts Map - House District Boundaries	Map of Missouri school districts with House district boundaries.
	School Districts Map - Senate District Boundaries	Map of Missouri school districts with Senate district boundaries.
	School Districts Map - Congressional District Boundaries	Map of Missouri school districts with congressional district boundaries.
	District Enrollment by Grade	Raw Data: District enrollment by grade level for 1991-2014 (data as of 10/15/2014).
	District Enrollment 2007-2013	Raw Data: Multi-year enrollment by district.
	District Enrollment 2014	Raw Data: District Enrollment by district - ranked by percent of total state enrollment.
	District Dropout and Graduation Rates 2007-2011	Raw Data: 2007-2011 Dropout and Graduation Rates by District.
	Excess of Technology 2011	Report: Summary Report of annual Census of Technology measures.

District and Building Graduation and Dropout Indicators

Type	Name	Description
	District Annual Dropout Rate	Raw Data: District dropout information. Includes total dropouts and race breakdowns.
	District Graduate Analysis	Raw Data: District graduate analysis. Includes placement status on previous year's graduates.
	District Graduation Rate	Raw Data: District graduation rates. Includes total graduates and race breakdowns.
	Four-Year District Graduation Rate	Raw Data: Four-Year State and District Graduation Rates.
	Four-Year State Graduation Rate	PDF: Four-Year State and District Graduation Rates.
	Building Annual Dropout Rate	Raw Data: Building dropout information. Includes total dropouts and race breakdowns.
	Building Graduate Analysis	Raw Data: Building graduate analysis. Includes placement status on previous year's graduates.
	Building Graduation Rate	Raw Data: Building graduation rates. Includes total graduates and race breakdowns.
	Graduates 10 Month Postsecondary Follow-up (SPSP-511)	Report: Number of graduates who have enrolled in Postsecondary education within 10 months of graduation. This report is being published as an

District and Building Student Indicators

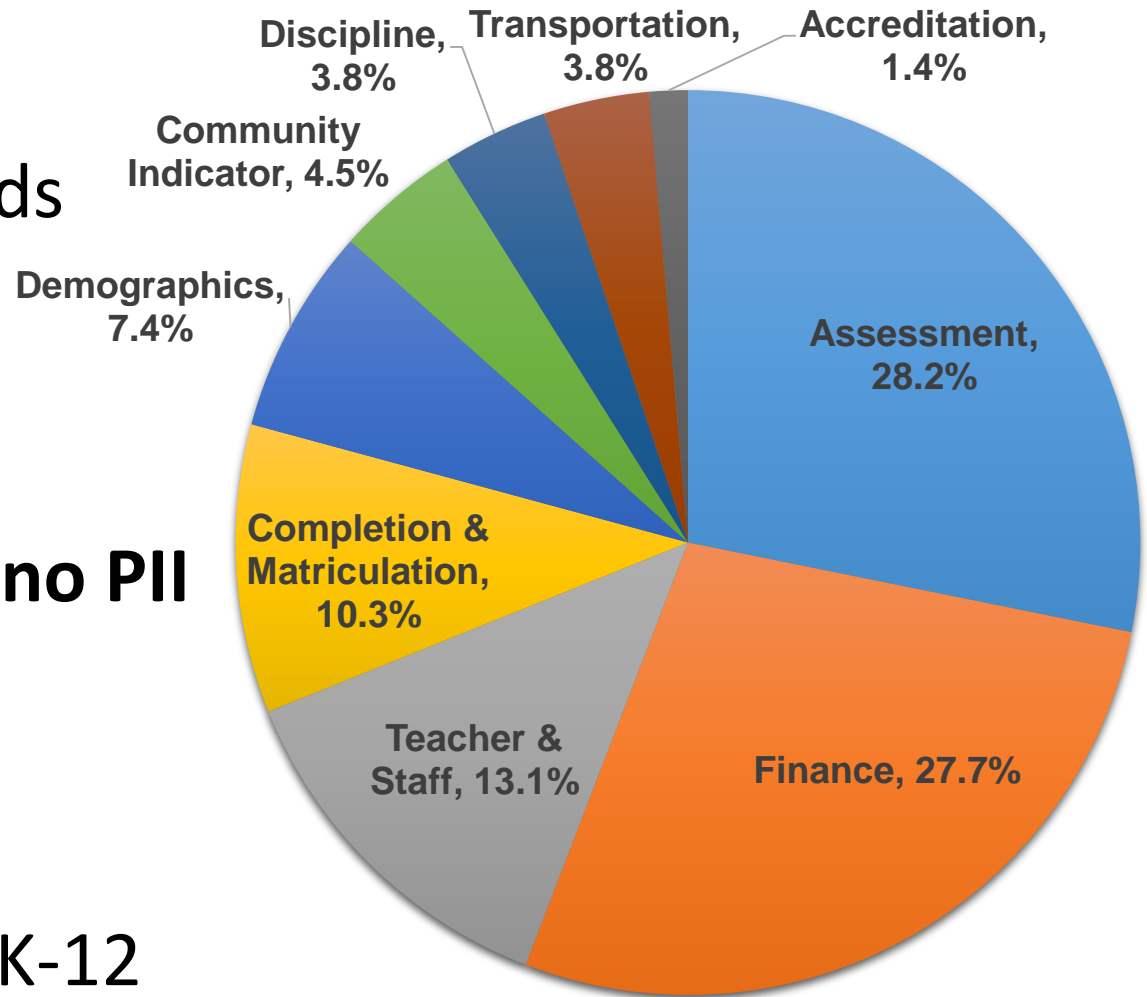
Type	Name	Description
	District Attendance Rate	Raw Data: District attendance rates. Includes K-12 and K-8 rates.
	District Demographic Data	Raw Data: District demographic data. Includes enrollment information by race. Data as of 10/09/2014.
	District Discipline Incidents	Raw Data: District discipline information. Includes enrollment, number of incidents, and rates by offense and removal type.
	District Proportional Attendance Rate	Raw Data: District proportional attendance rate (students attending 90% of the time). Years 2009-2013. Includes K-8 and 9-12 rates.
	Building Attendance Rate	Raw Data: Building attendance rates.
	Building Demographic Data	Raw Data: Building demographic data. Includes enrollment information by race. Data as of 10/09/2014.
	Building Discipline Incidents	Raw Data: Building discipline information. Includes enrollment, number of incidents, and rates by offense and removal type.
	Building Proportional Attendance Rate	Raw Data: Building proportional attendance rate (students attending at least 90% of the time). Years 2009-2013. Includes K-8 and 9-12 rates.

District and Building Education Staff Indicators

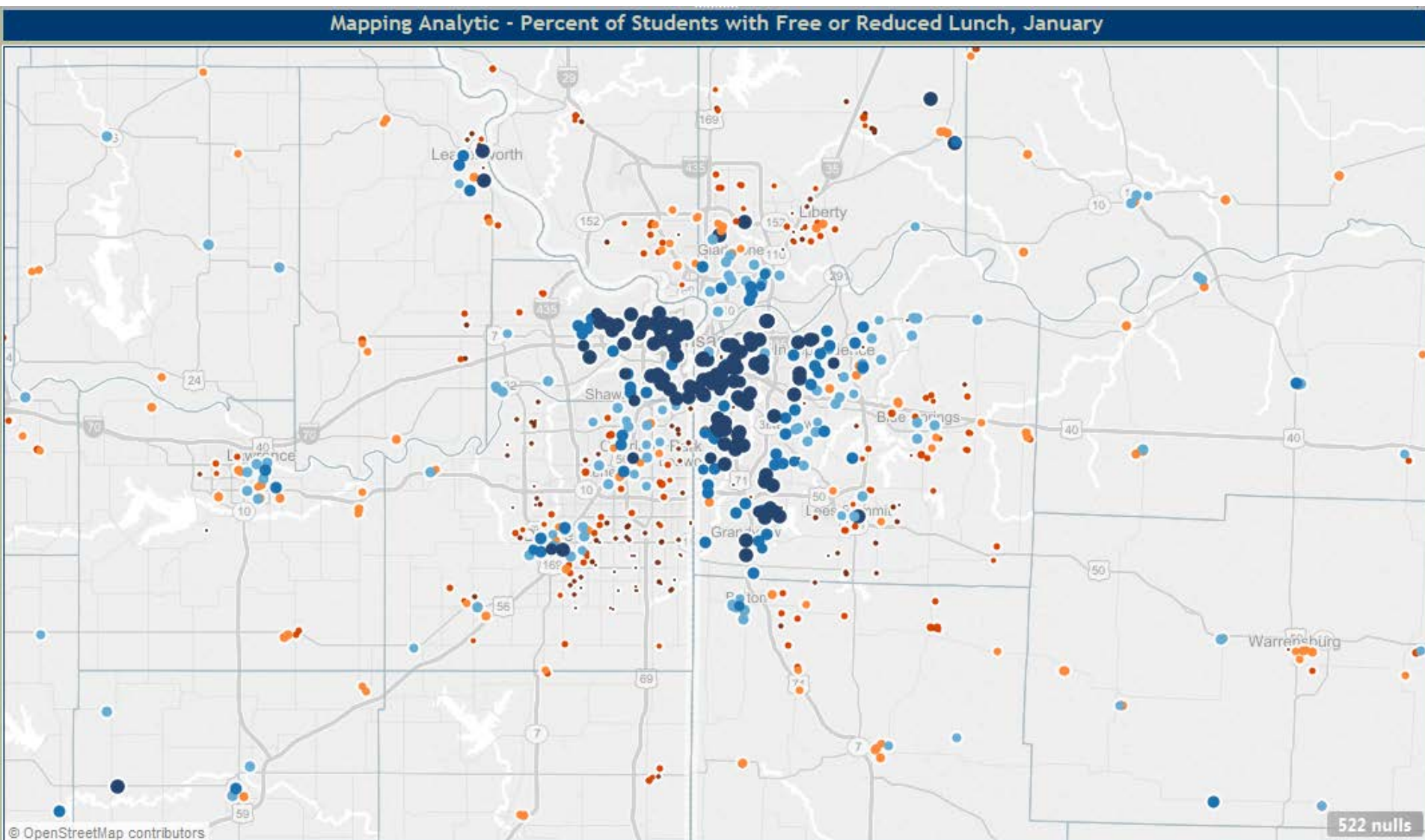
Type	Name	Description
	District Faculty Information	Raw Data: District faculty information. Includes salary, degree, and years of experience. Data as of 10/09/2014.
	District Student Staff Ratios	Raw Data: District staff ratios. Include student to teachers and student to administrator ratios. Data as of 10/09/2014.
	Building Faculty Information	Raw Data: Building faculty information. Includes salary, degree, and years of experience. Data as of 10/09/2014.
	Building Student Staff Ratios	Raw Data: Building staff ratios. Include student to teachers and student to administrator ratios. Data as of 10/09/2014.
	Building Certification	Raw Data: Building teacher certification information. Includes number of teachers with a valid certificate and highly qualified teachers.
	District Certification	Raw Data: Building teacher certification information. Includes number of teachers with a valid certificate and highly qualified teachers.

The Data Behind EdWise

- 14 million records
- 600+ metrics
- Aggregate only, **no PII**
- Nine themes
- 100% of data is K-12



Tool 1: Mapping

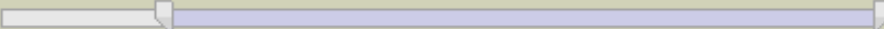


Tool 2: Find a District/School

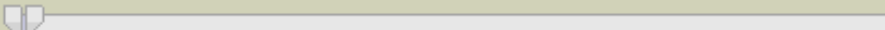
Metric Measurement Level	Beginning of School Year	State	Grade/Assessment Group	Districts/Schools Matching Selected Metrics	
Building	(All)	(All)	Null	State	District/School
				Kansas	Abilene/Abilene High School

Metric Measurement Level	Beginning of School Year	State	Grade/Assessment Group	Districts/Schools Matching Selected Metrics	
District	2013	Missouri	Null	State	District/School
				Missouri	Ferguson-Florissant R-II
					Hazelwood
					Kansas City 33
					St. Louis City

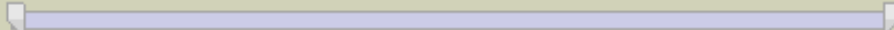
Metric Category [A]
Fall Enrollment

Metric Value [A]
10000  50749


Metric Category [B]
Students Per Teacher

Metric Value [B]
0  15

Metric Category [C]
Percent of Students with Free or Reduced Lunch, January

Metric Value [C]
0.5  31.167

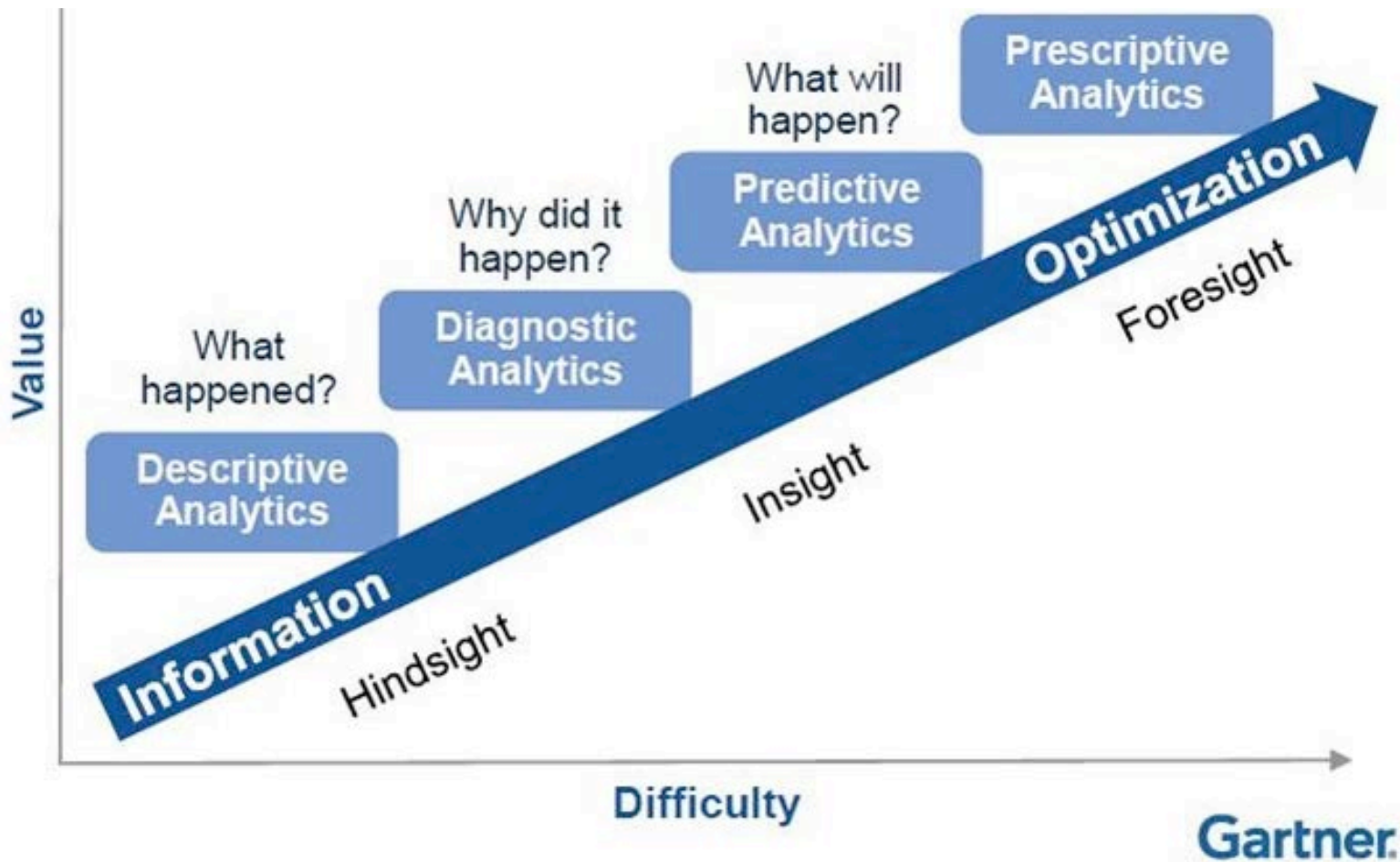
Metric Category [D]
Percent of Non-White Students within the Fall Enrollment

Metric Value [D]
0.5  1

Key Research Questions

- Can we detect problems that students are having *as they are happening*, through automated analysis?
- Can we provide real-time feedback to students and educators in response to the problem detection?
- Is the feedback effective in helping students attain more sophisticated behaviors? Does it *make sense* to the students and educators? Is it *actionable* in that they are able to do something useful with it?

From Hindsight to Foresight



Key Next Steps

- Mobilize Communities around Opportunities based on New Forms of Evidence
- Develop New Forms of Educational Assessment
- Develop New Types of Analytic Methods
- Build Human Capacity to Do and to Understand Data Science
- Develop Advances in Privacy, Security, and Ethics
- Infuse Evidence-based Decision-Making throughout Organizations and Systems

Questions to Ponder

- To what types of behavioral data could we now apply these methods?
 - *Micro-level* data (e.g., each student's second-by-second behaviors as they learn)
 - *Neso-level* data (e.g., teachers' patterns in instruction; students' patterns in retention)
 - *Macro-level* data (e.g., aggregated student outcomes for accountability purposes) *Gummer's work with EdWise*
- What are the barriers to collecting, storing, sharing and analyzing these data?
- How can we build human and organizational capacity to use evidence-based findings effectively?