Data Privacy in the Age of Big Data

Jeanette Baez
Executive Director of Institutional Research, Planning & Analytics @ Cal Poly Pomona

Today's Agenda

- What is data analytics?
- What is the role of big data?
- What are the implications for data privacy?
- What other issues are related to data analytics?
- How can we better implement the use of data analytics, particularly with big data?

DATA ANALYTICS

Systematic examination of data to identify patterns and gain insight on a specific phenomena, especially to inform decision-making

About me

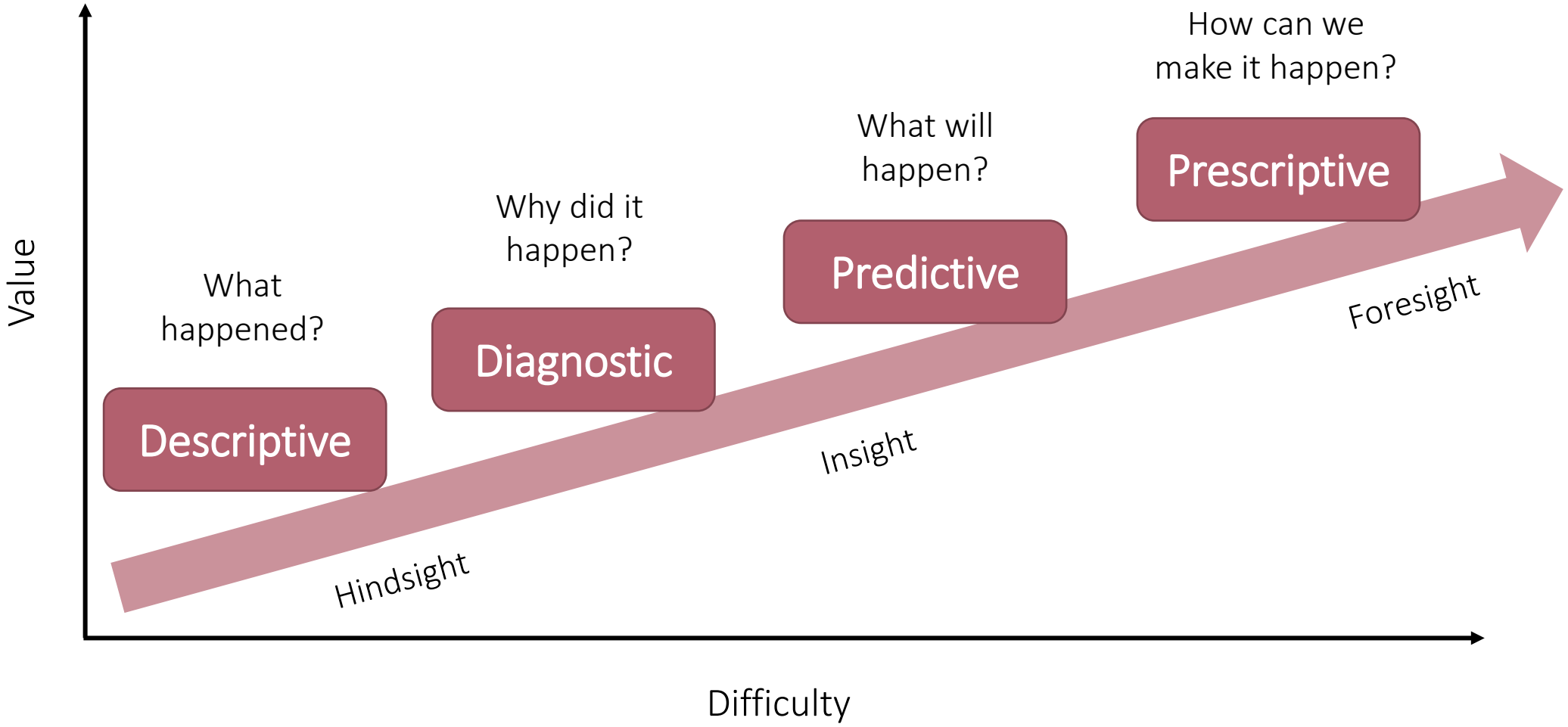
Institutional Research, Planning, and Analytics (IRPA) supports data-informed decision making to help Cal Poly Pomona fulfill its commitment to student access and achievement. We support continuous improvement by seeking innovative ways to make data accessible, intelligible, and actionable. We also help demonstrate accountability to accreditors, external entities, and the public.

About you

What industry do you currently employed in, or seeking employment in?

- Financial Services/Insurance
- Defense
- Entertainment
- Education
- Government
- Manufacturing
- Professional Services
- Technology
- Utilities
- Other

Levels of analytics



Gartner Analytic Ascendancy Model (2012)

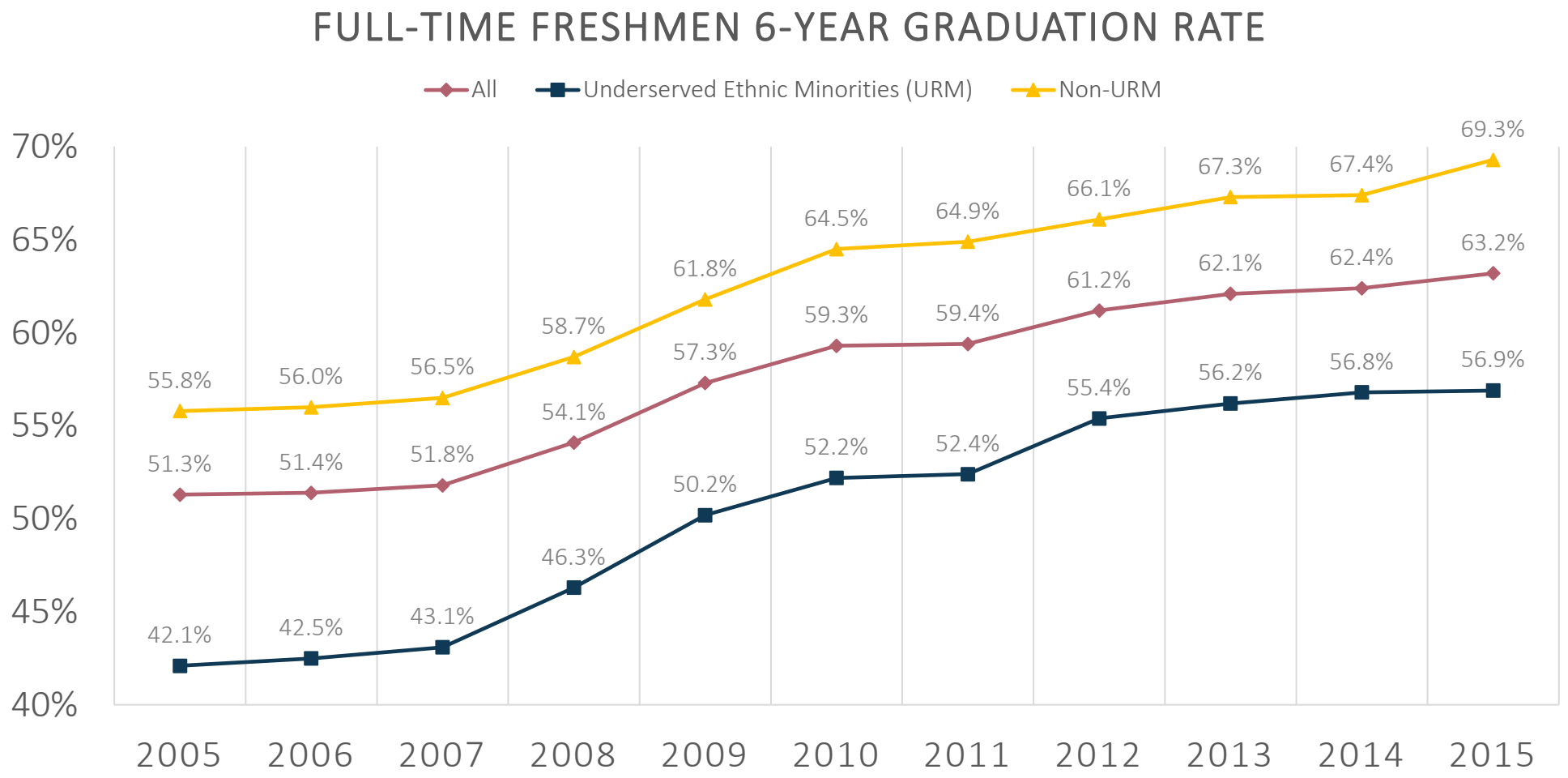
Descriptive Analytics

- Helps to categorize and summarize data in a meaningful manner to better understand a phenomena
- Descriptive analytics typically is represented by counts and percentages and can be displayed in a table or graph
- The development of descriptive analysis can involve small to large amounts of data and accomplished through unsophisticated to sophisticated means

Example of Descriptive Analytics

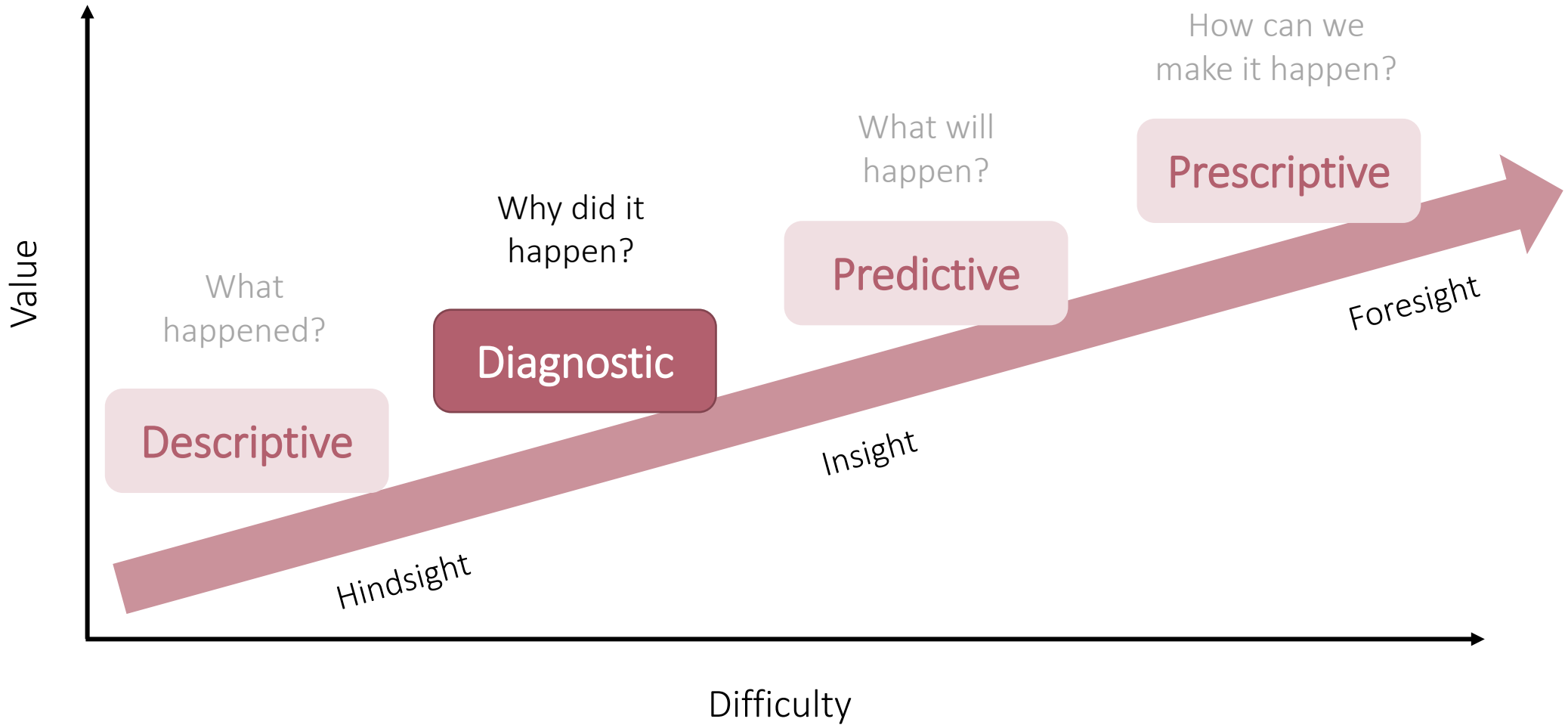
Source: NBC News as of February 7, 2022

Using data to support student success



California State University. What are our Historical Graduation Rates? Retrieved from https://public.dashboards.calstate.edu/public/csu-by-the-numbers/graduation-rates

Levels of analytics



Gartner Analytic Ascendancy Model (2012)

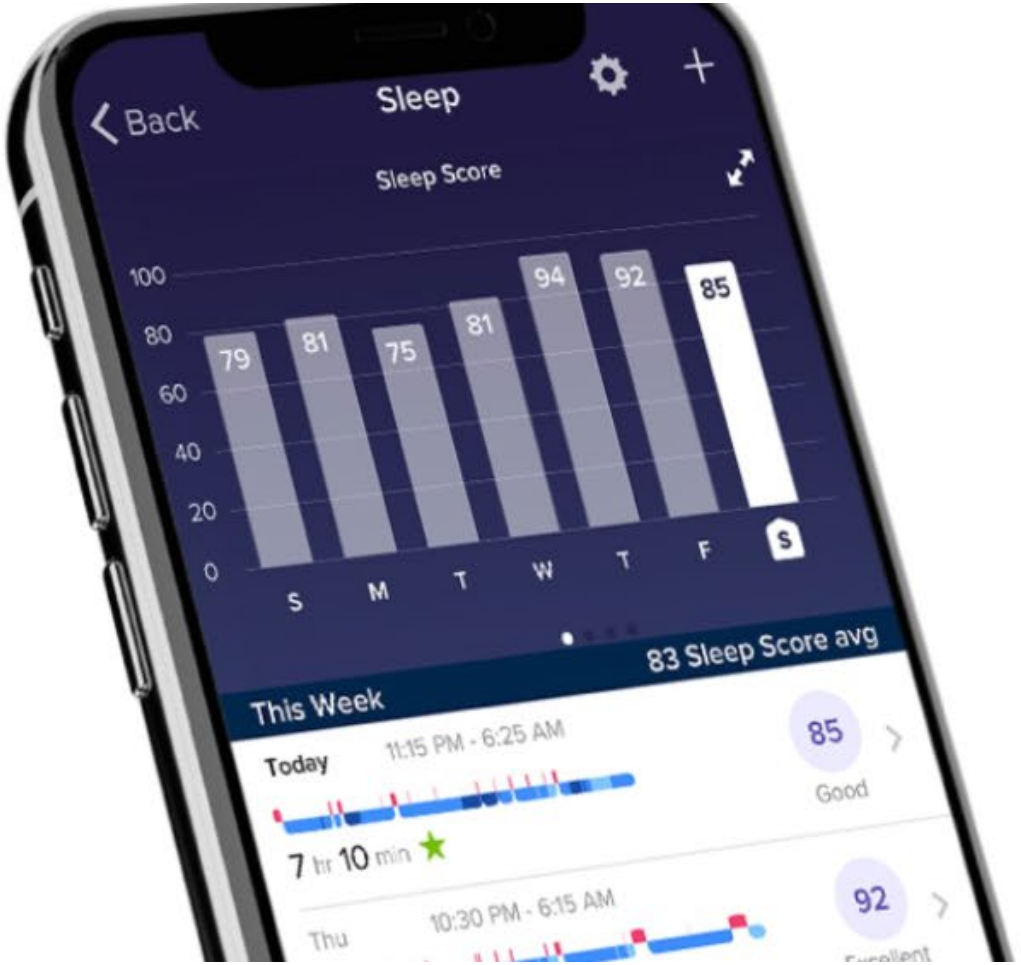
Diagnostic Analytics

- Diagnostic analytics uses data to explain trends and relationships witnessed in descriptive analysis, aka root cause analysis
- Statistical analysis is more complex than descriptive analysis and is typically conducted utilizing software

Why don’t students graduate

- What is the relationship between high school performance and college outcomes?
- How does enrollment from one term to the next relates to graduation?
- What is the relationship between course taking patterns and college outcomes?

Example of Diagnostics Analytics

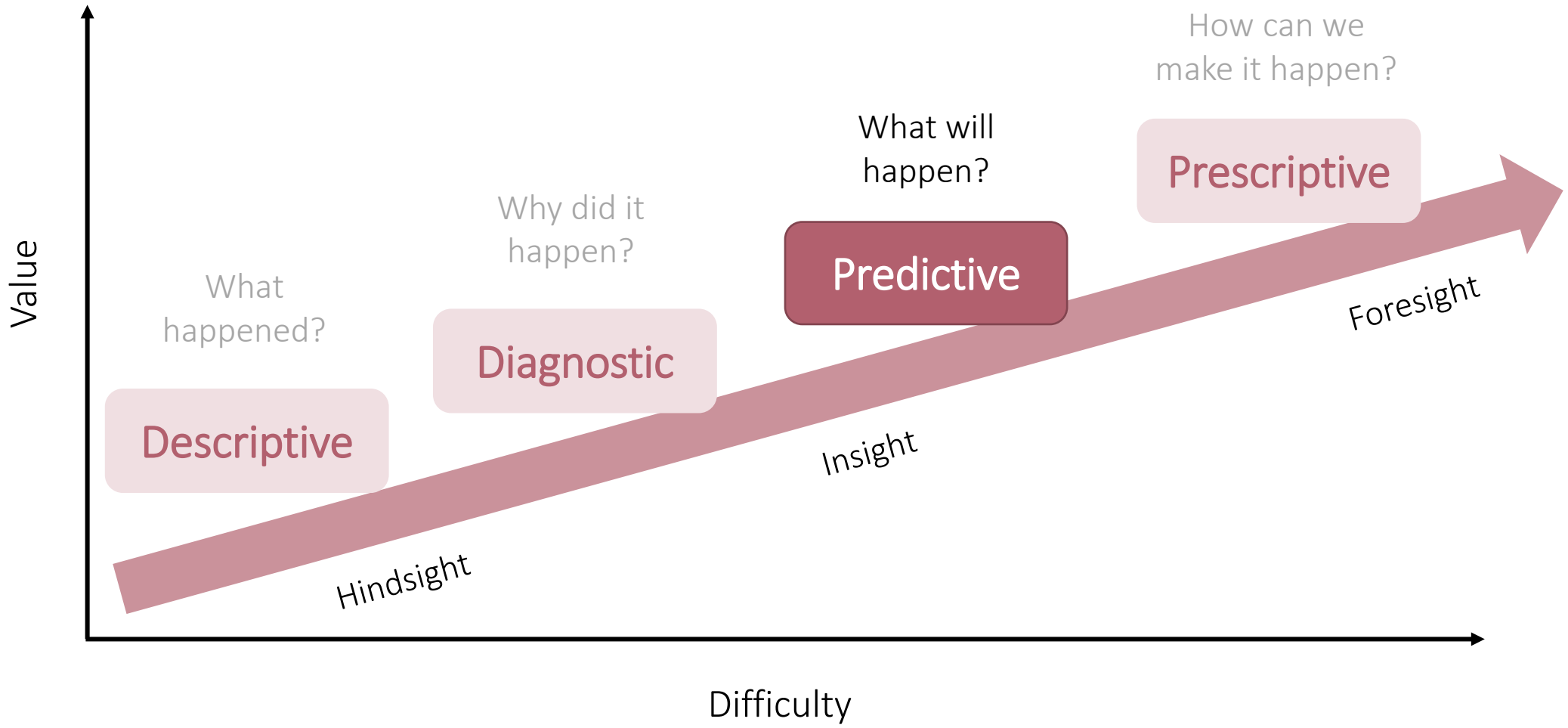


Use your heart to reveal the quality of your ZZZs

Fitbit trackers and watches use your sleeping heart rate, movement and more to measure your time spent in each sleep stage and give you a personalized Sleep Score that shows how well you slept. Plus, view your trends over time in the app, and see how your stats compare to others.*

Source: www.fitbit.com

Levels of analytics



Gartner Analytic Ascendancy Model (2012)

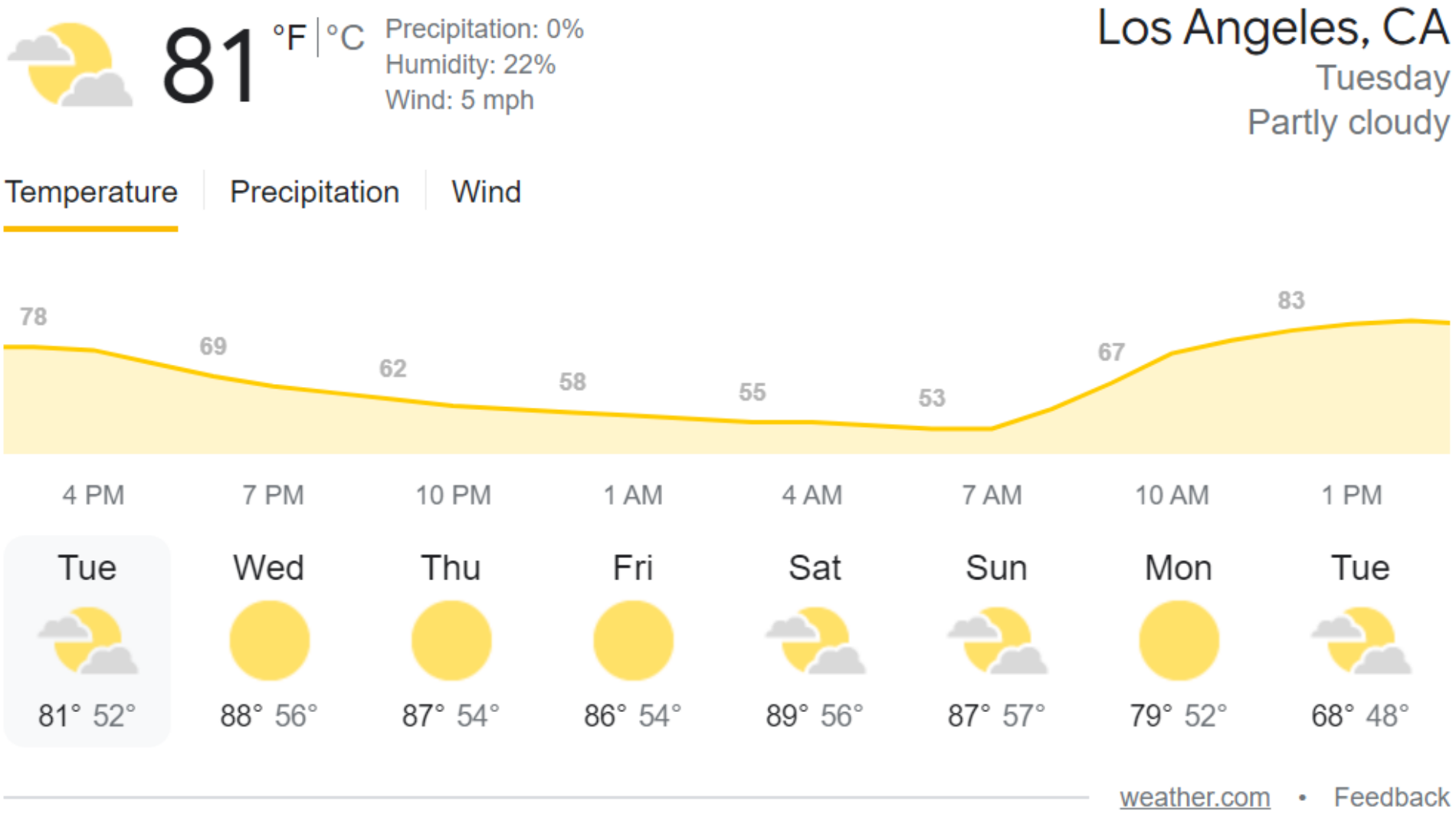
Predictive Analytics

- Proactive approach that allows an organization to anticipate the possibility of an outcome before it happens
- Predictive analytics can be utilized to view outcomes in the aggregate, or even down to an individual record
- The scale of predictive analytics can vary from a handful of variables used to in a statistical model to the utilization of machine learning and artificial intelligence that can handle dozens or hundreds of variables

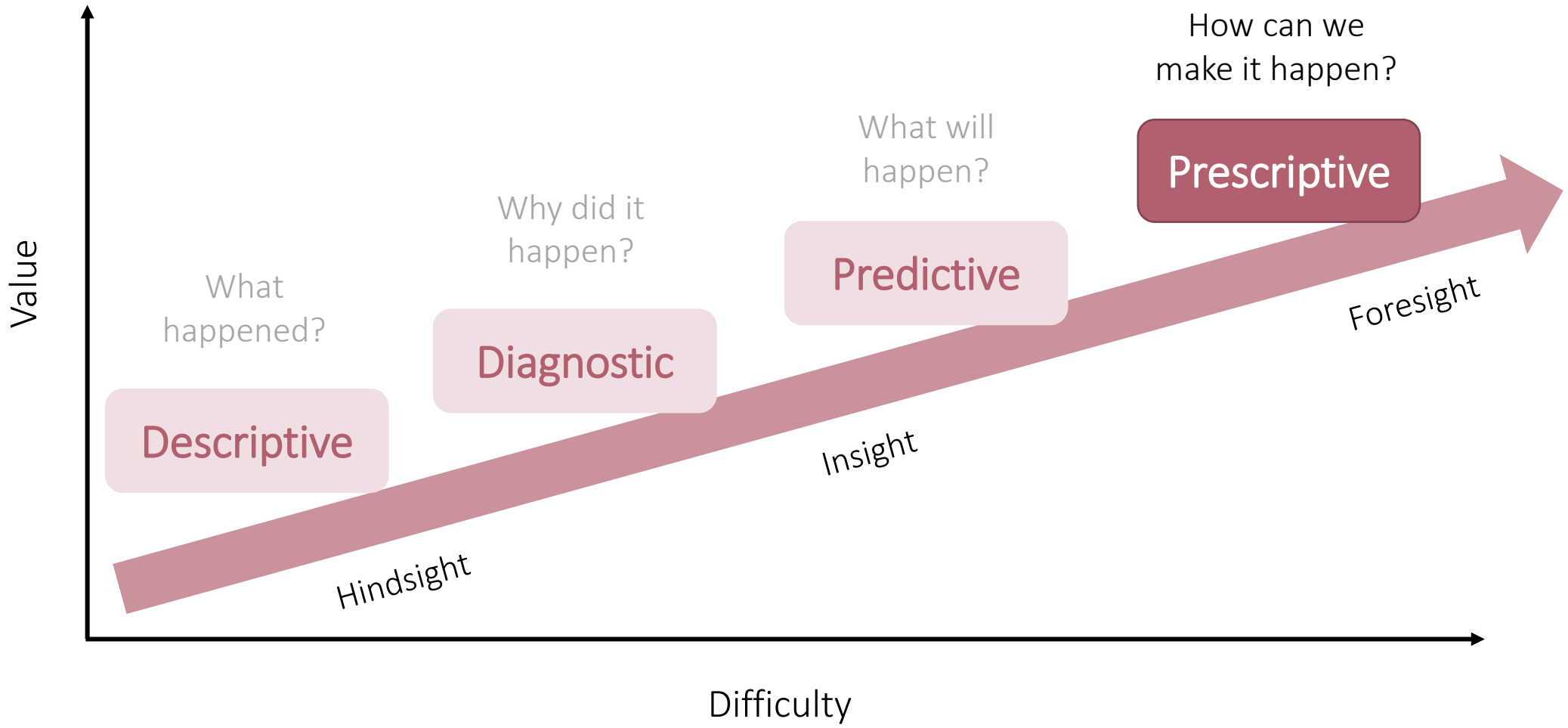
Predictive analytics applied to student success

- Which applicants would do well at our institution?
- Which math course would a particular student succeed in?
- Which students are most likely to not be enrolled after one year, two years, etc.?
- Which students are most likely to graduate in four years?

Example of Predictive Analytics



Levels of analytics



Gartner Analytic Ascendancy Model (2012)

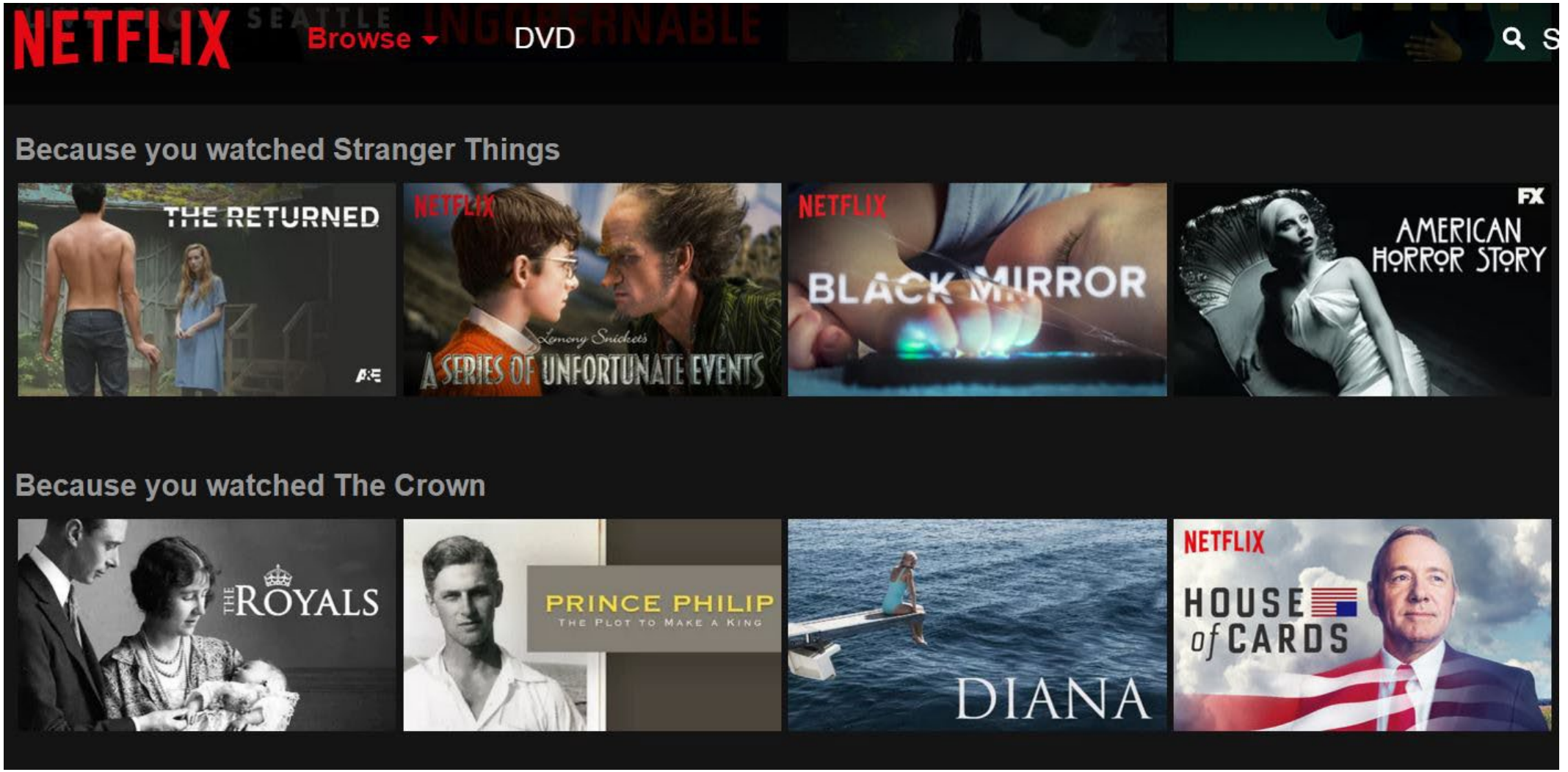
Prescriptive Analytics

- Prescriptive analytics utilizes historical data patterns and outcomes to provide recommendations for a current situation
- Prescriptive analytics are often used to provide consumer recommendations for next steps, or to help inform decisions or strategies within the organization
- Large data sets and sophisticated analytic techniques are required for prescriptive analytics

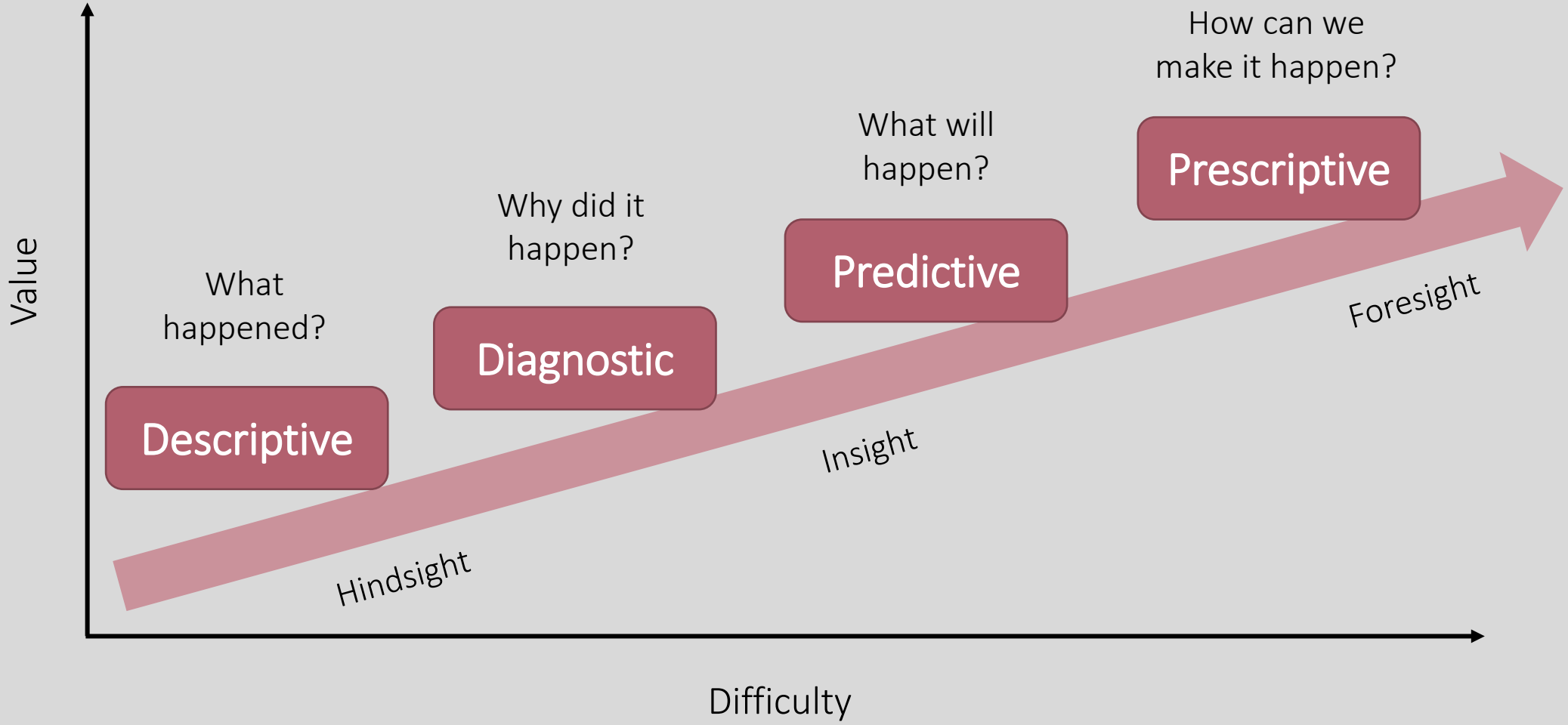
Prescriptive analytics in higher education

- Recommending certain courses to students during a student's registration period
- Automatically trigger advising or tutoring appointments based on course attendance, engagement, and grades
- Send personalized messaging to students based on patterns of behavior and transactions

Example of Prescriptive Analytics



What is the highest level of analytics employed by your organization?

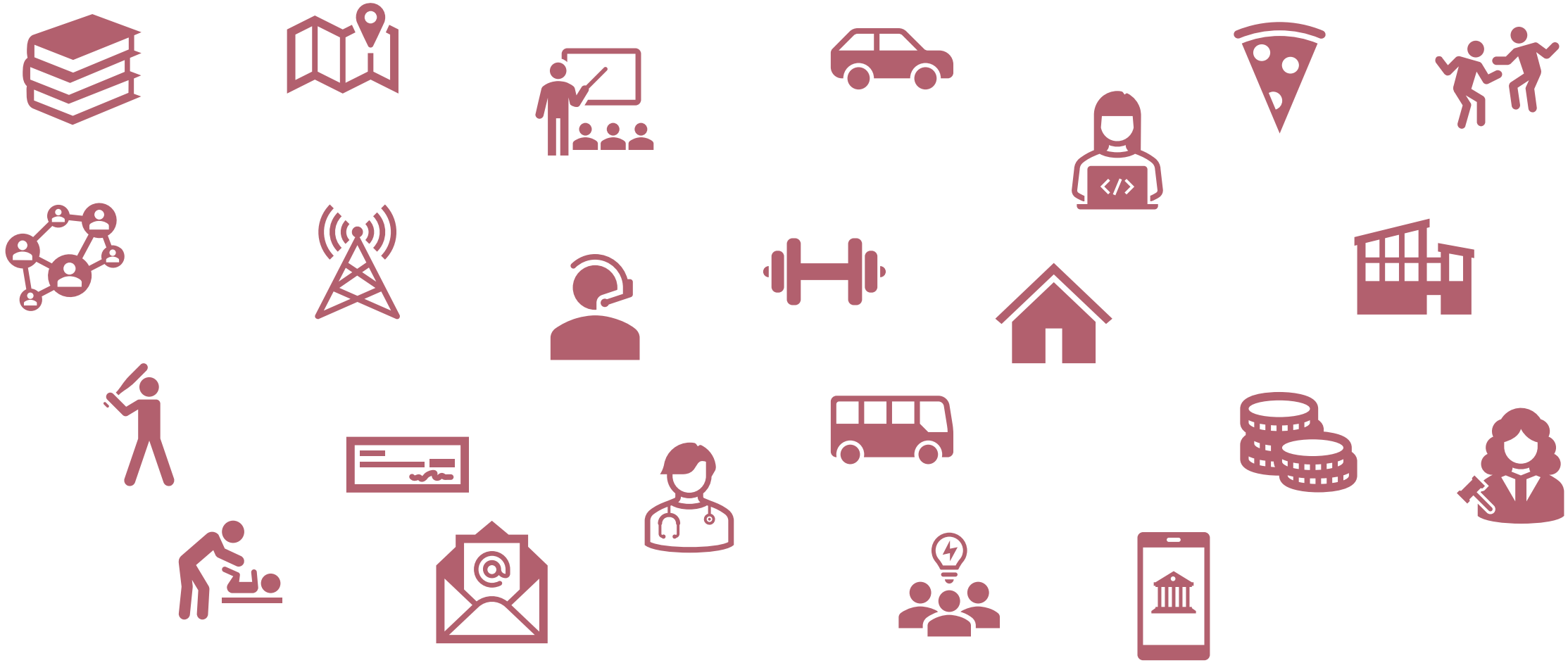


Gartner Analytic Ascendancy Model (2012)

Big Data

- Big data is the collection of extremely large sets of information from various sources that is constantly streaming in and out
- Sophisticated software and analytical methods are required to examine trends and patterns

The amount of data at a college campus is endless



Data privacy concerns

- Big data poses a high risk due to the sheer amount of personal data collected which may result in the los of anonymity
- Loss of data can lead to reputational harm, legal action, loss of revenue, disruption of operations, and regulatory sanctions
- Need for oversight, monitoring, and updating data security and privacy measures

Data privacy concerns, continued

- May grow consumer's mistrust of the organization with the collection of a significant amount of personal information, especially if it's not disclosed
- Regardless of analytic level and even if data is de-identified, anonymity is not always a guarantee
- Education and training must be routine and constantly evolving

What do you feel is the top data privacy concern for your organization?

- Data breaches
- Accidental exposure
- Exponential data growth
- Compliance with laws and regulations
- Data governance, including data oversight and policies
- Ethical use of data
- Unsure/Other

Additional concerns and considerations

- Complex analytic models are not completely accurate, can be biased, may be based on inaccurate data, and may not include all the necessary data
- Predictions do not define a consumer; for some industries, it should in no way impede access
- Signals to consumers takes away autonomy and can discourage them from taking certain paths

Concerns and Considerations, continued

- Being mindful of diversity, equity, and inclusion of minoritized populations throughout the data analysis process, but also in relation to data privacy
- Data never tells the entire story

How to move forward

- Data governance
- Data ethics standards
- Data security policies
- Education and training
- Infrastructure investment
- Consumer disclosures

Data governance

- Typically, a collection of individuals from across the organization that develop internal standards and policies pertaining to data
- Examples include data definitions, type of data that will be collected, what the data will be used for, how it will be protected, etc.

Standards in data ethics

- Ethical considerations for collecting, handling, analyzing, reporting, interpreting, and archiving data
- Data governance committees are often involved with developing data ethics standards
- Can be included in consumer disclosures

Data security policies

- Wide ranging, including policies relating to users and technology
- Typically includes auditing policies, access control management, incident reporting, system security, vulnerability scans, etc.
- Build a culture where everyone sees data security as part of their responsibility as well

Education and training

- Educating staff, top to bottom, on data handling, privacy, and security.
- Keeping abreast on new technologies available for collecting and analyzing data, data security, and data privacy
- Routine skill development on new technologies

Infrastructure Investment

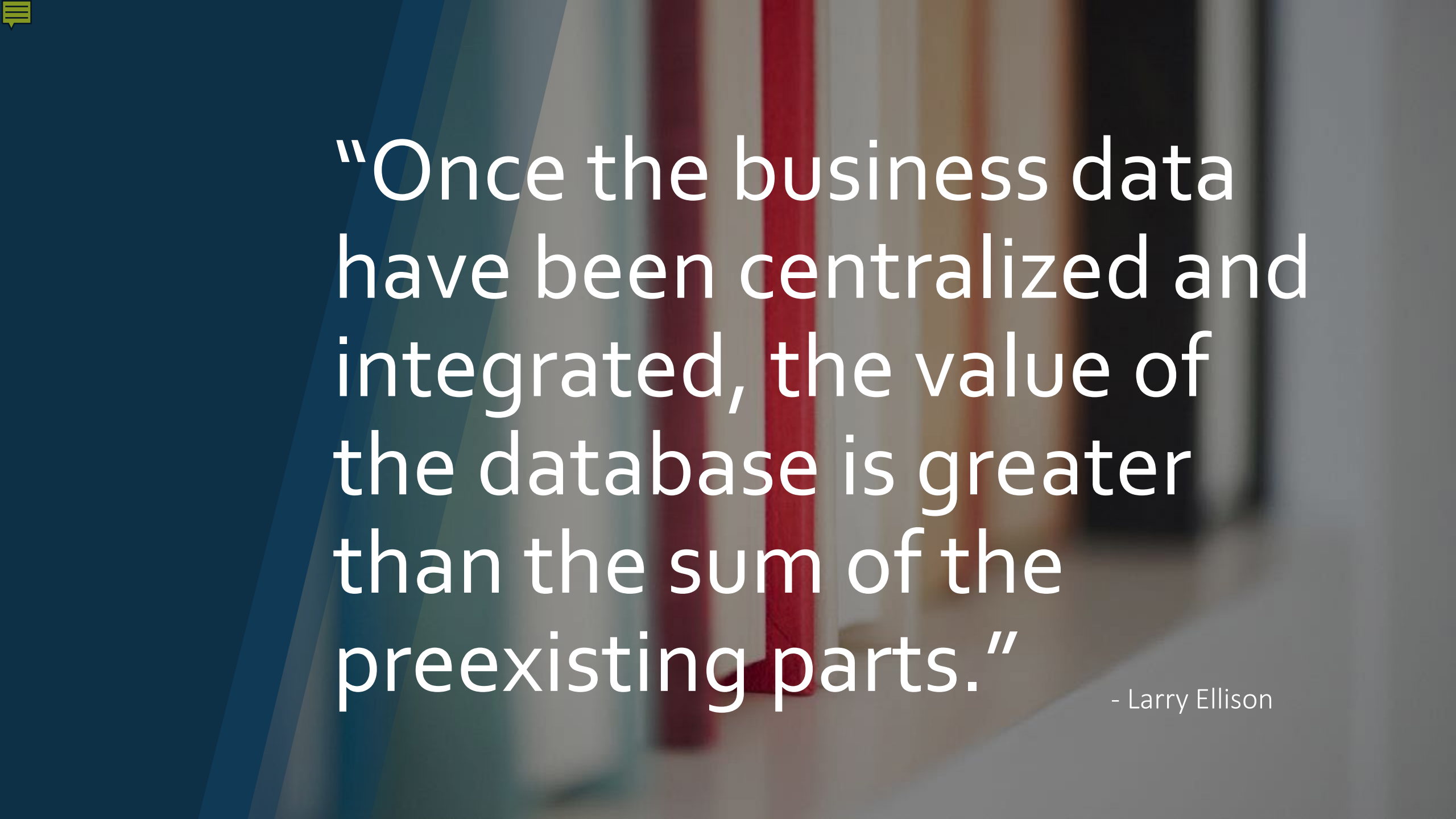
- Big data analytics and data privacy require investment in resources such as skilled personnel, professional development programs, physical and cloud technology
- Requires taking stock of where you're at and where you want to be
- With constantly evolving and new technologies, investments are continual as well

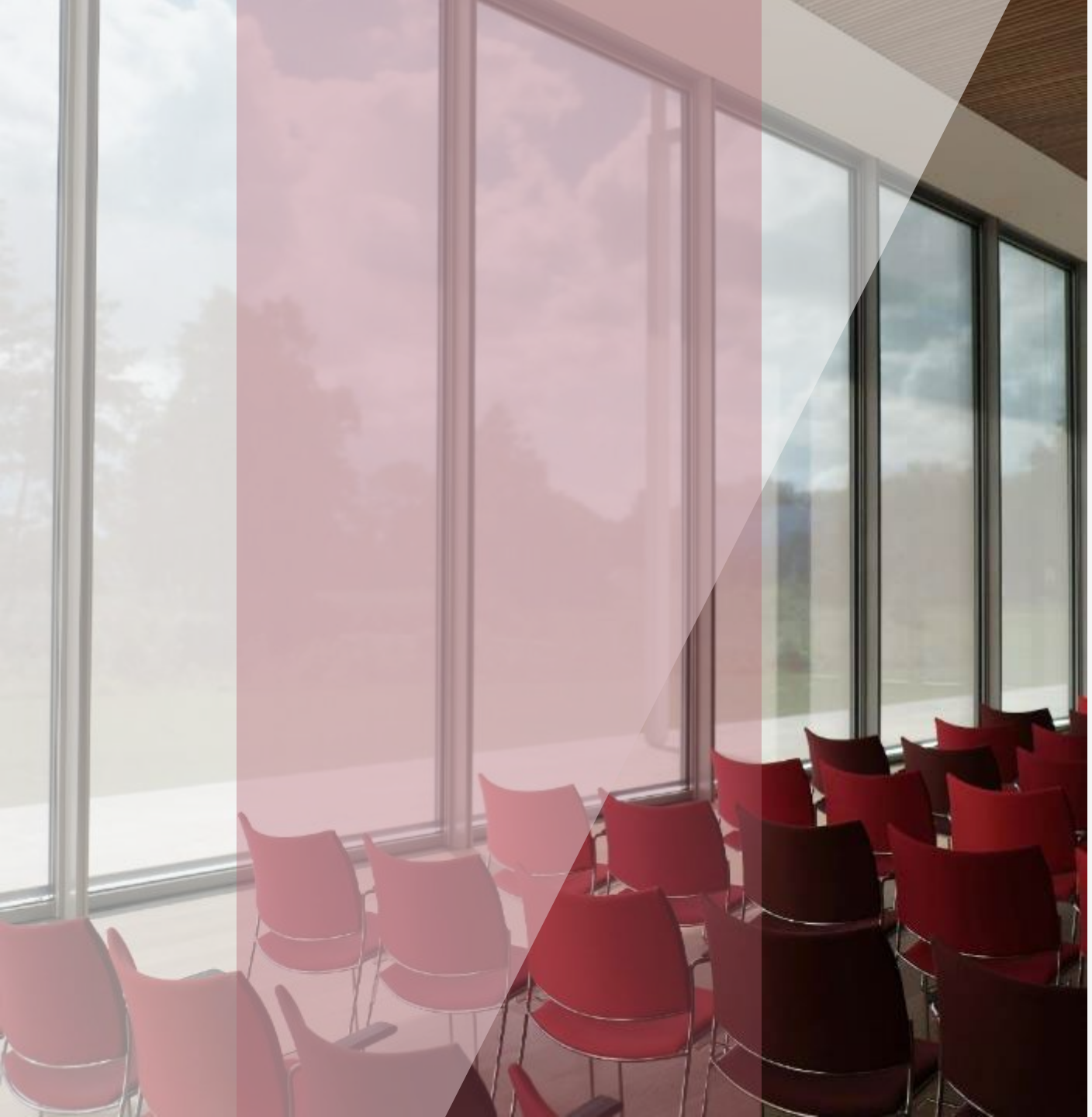
Disclosure

- Data usage, privacy, and security policies followed by the organization
- Disclose to consumers the type of data that is being collected, when it is being collected, how it's being collected, what will be done with the data, how the data will be protected, and who will have access to their data, including third parties
- Builds trust and transparency

Based on today's session, which area do you plan to explore further?

- Data governance
- Data ethic standards
- Data security policies and procedures
- Education and training
- Infrastructure investment
- Consumer disclosure
- Unsure/Other


"Once the business data have been centralized and integrated, the value of the database is greater than the sum of the preexisting parts."
- Larry Ellison



Questions

THANK YOU

Jeanette Baez

jgbaez@cpp.edu

www.cpp.edu/data