DATA SCIENCE IN PRACTICE

JESSE SPENCER-SMITH, CHIEF DATA SCIENTIST
DATA SCIENCE INSTITUTE, VANDERBILT UNIVERSITY

HTTPS://WWW.VANDERBILT.EDU/DATASCIENCE/
TWITTER: @VUDATASCIENCE

Mission

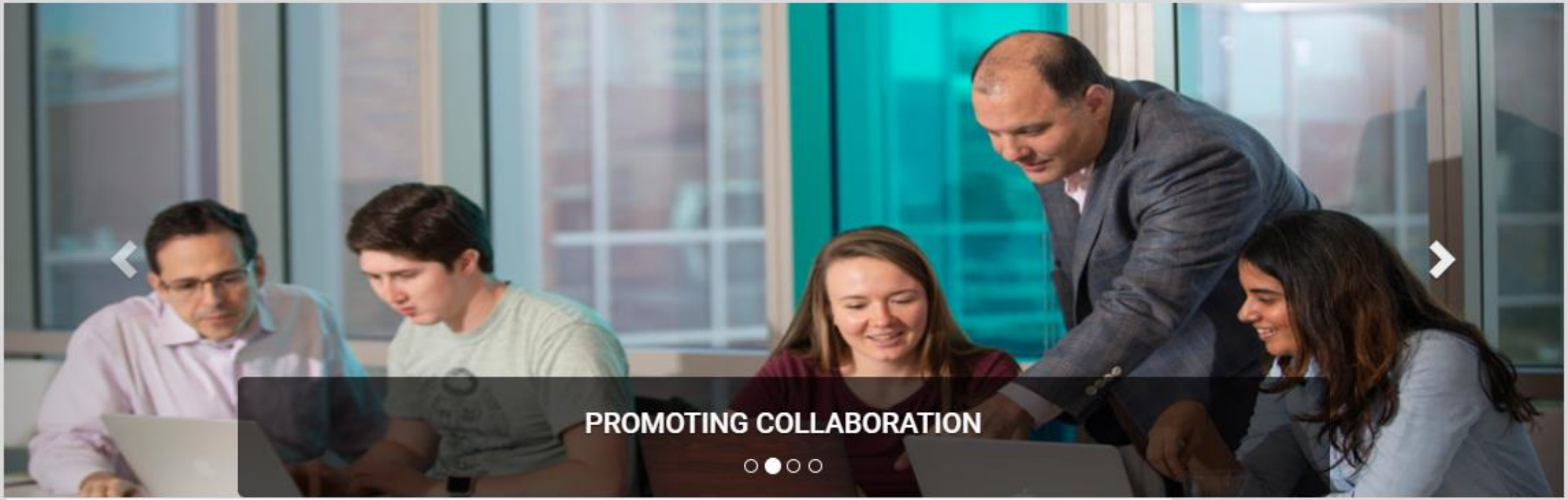
Accelerate data-driven research.

Promote collaboration.

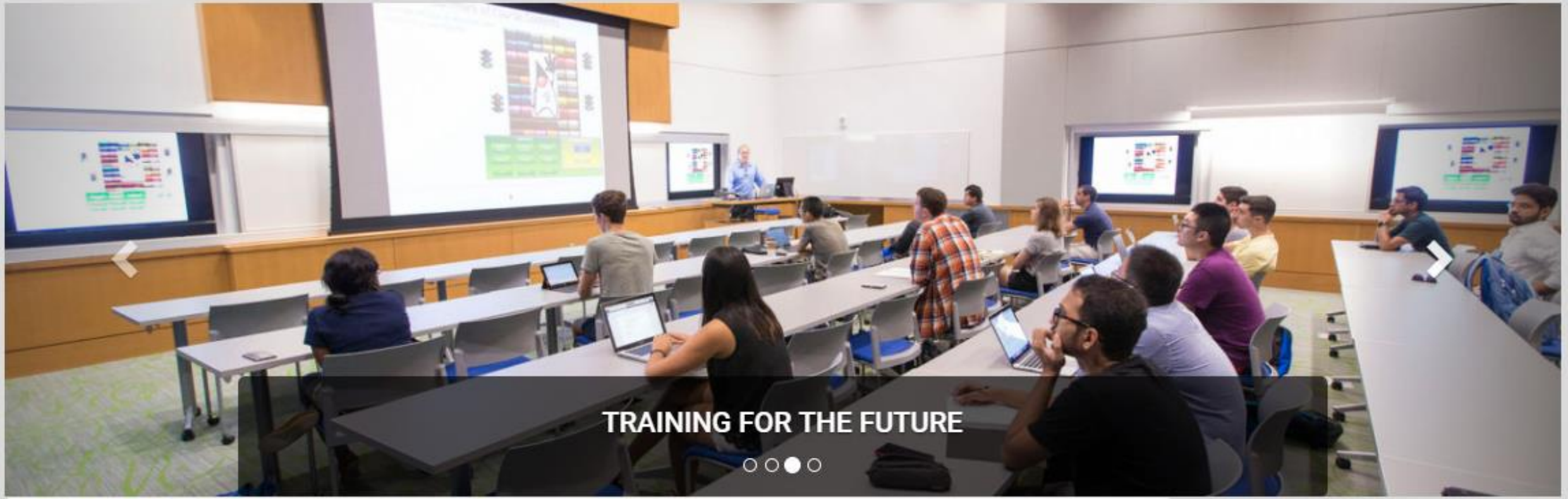
Train future leaders.

Data Science Institute

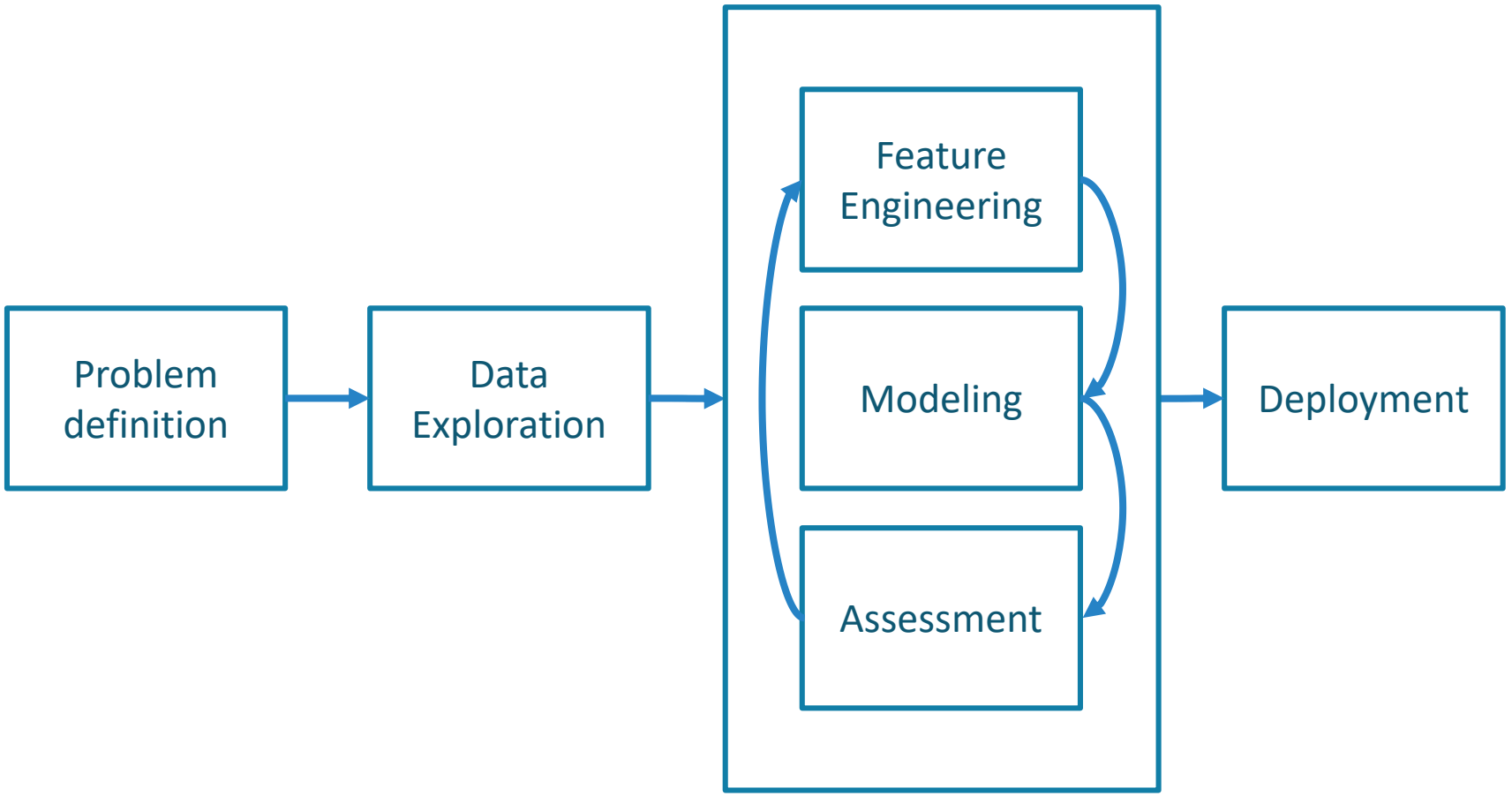
Data Science Institute



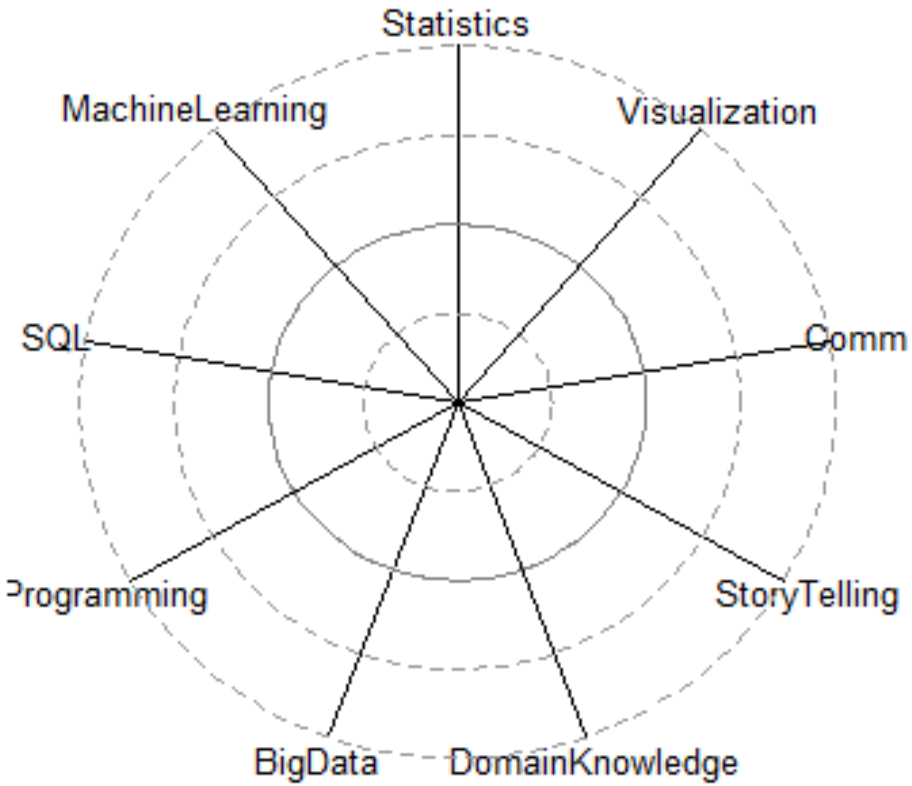
Data Science Institute

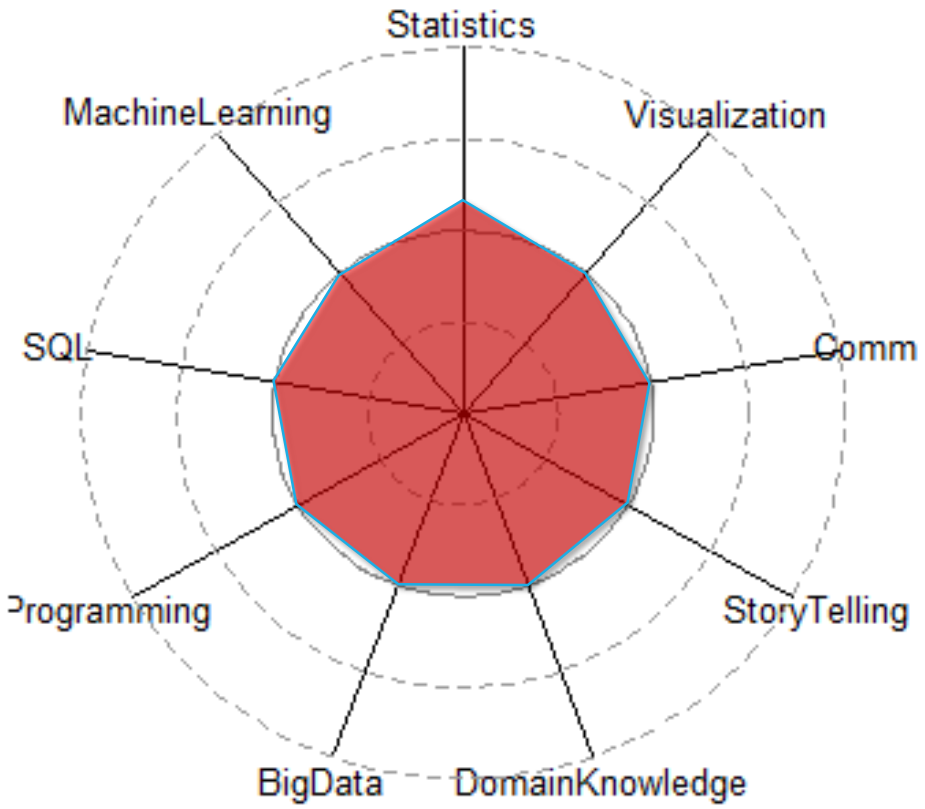


Data Science Lifecycle



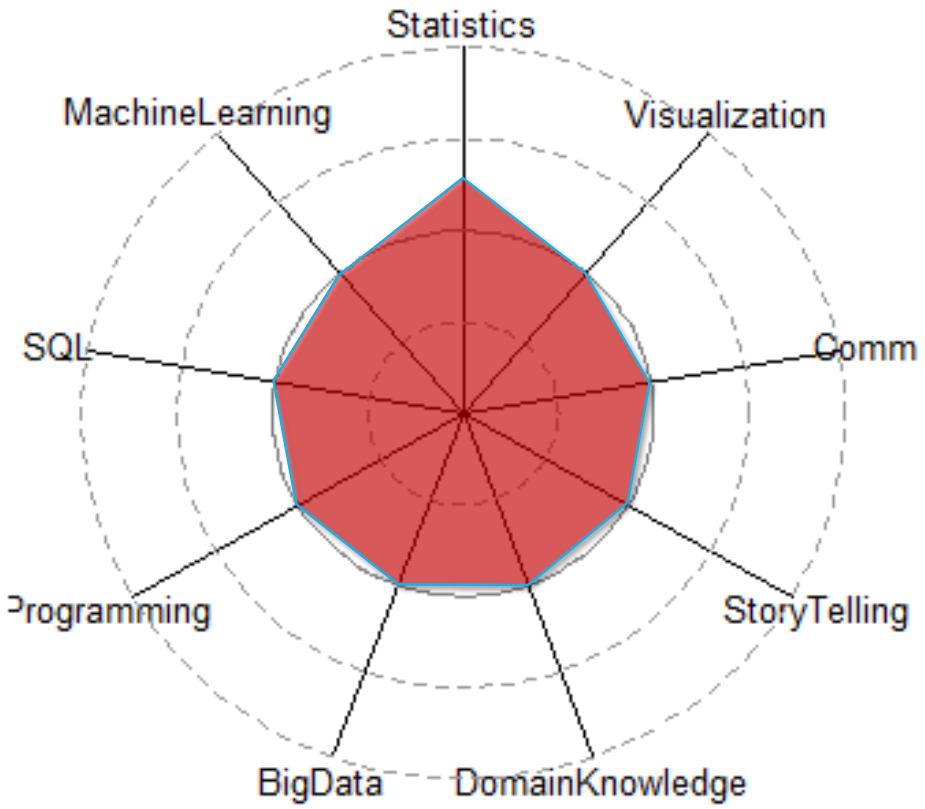
Skills Inventory





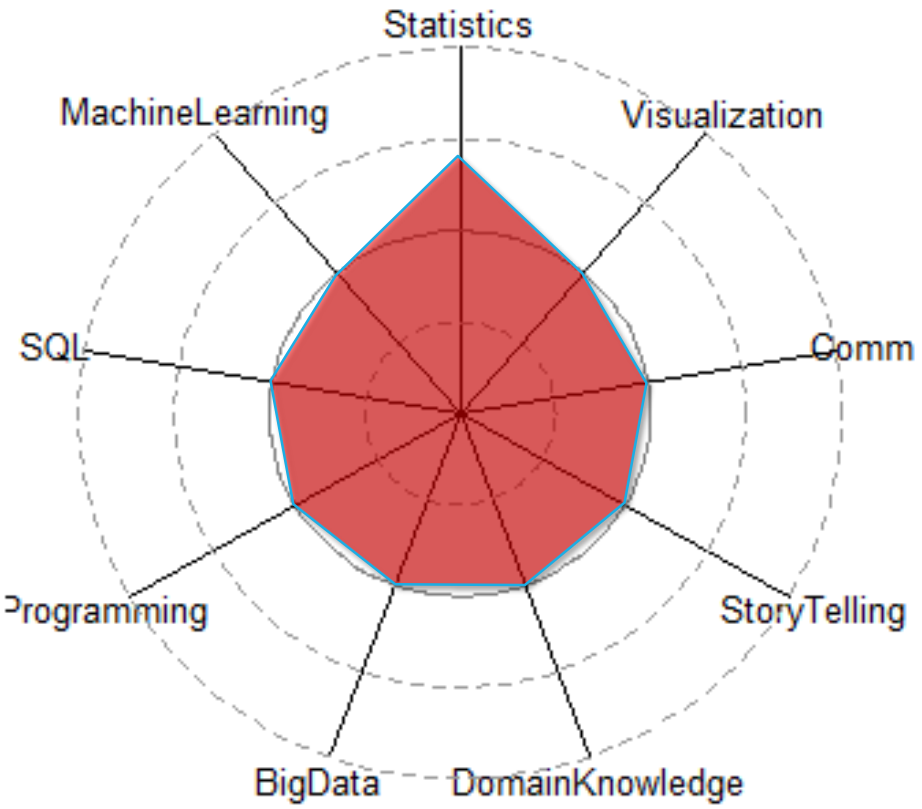
Statistics and Mathematics

Statistical modeling



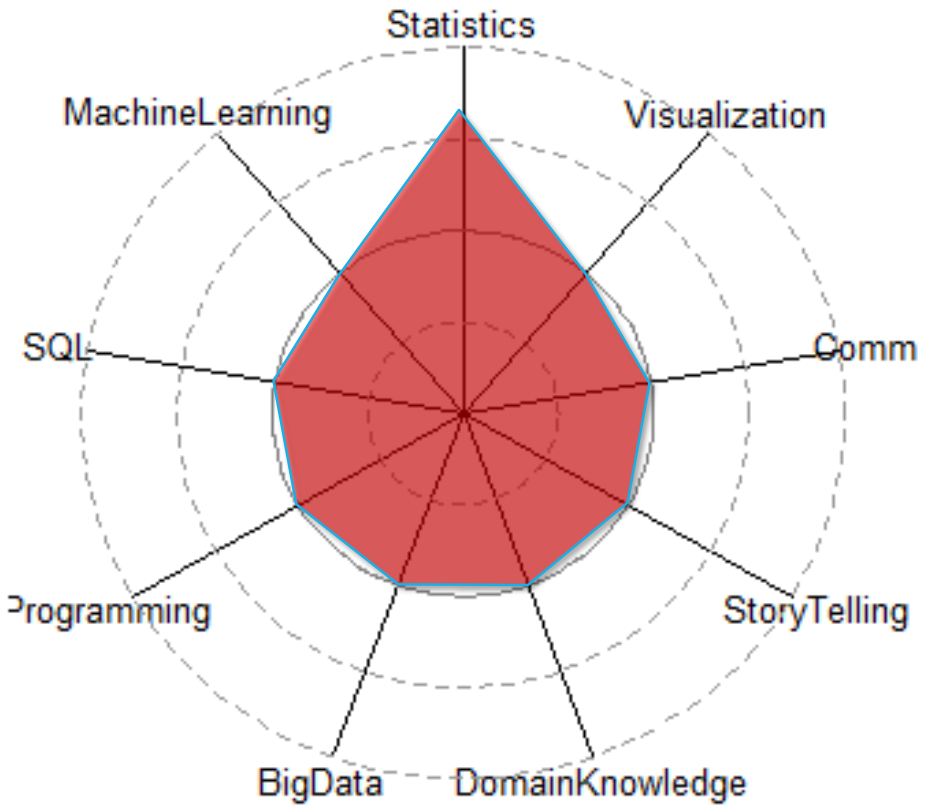
Statistics and Mathematics

Statistical modeling
Causal modeling



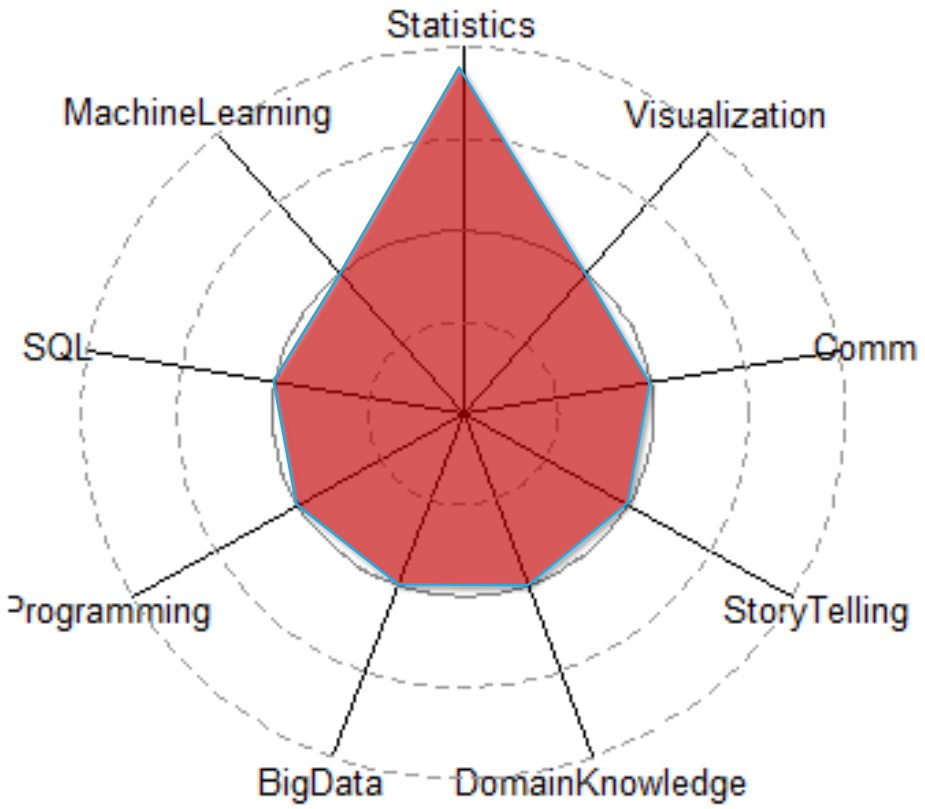
Statistics and Mathematics

Statistical modeling
Causal modeling
Linear algebra



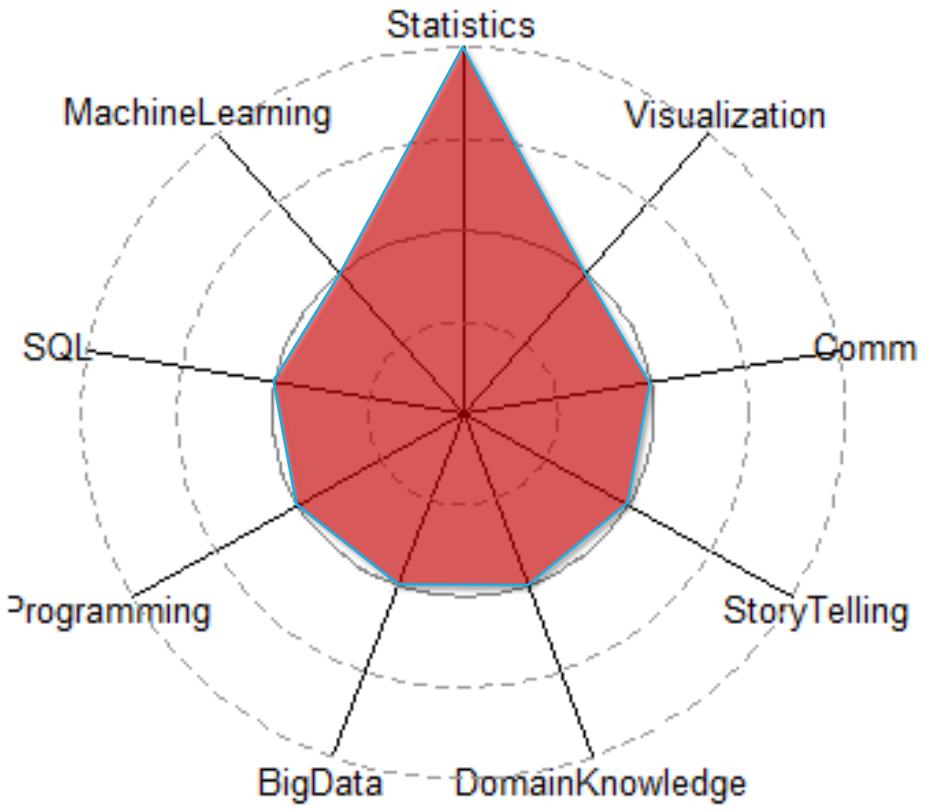
Statistics and Mathematics

Statistical modeling
Causal modeling
Linear algebra
Multi-variable calculus



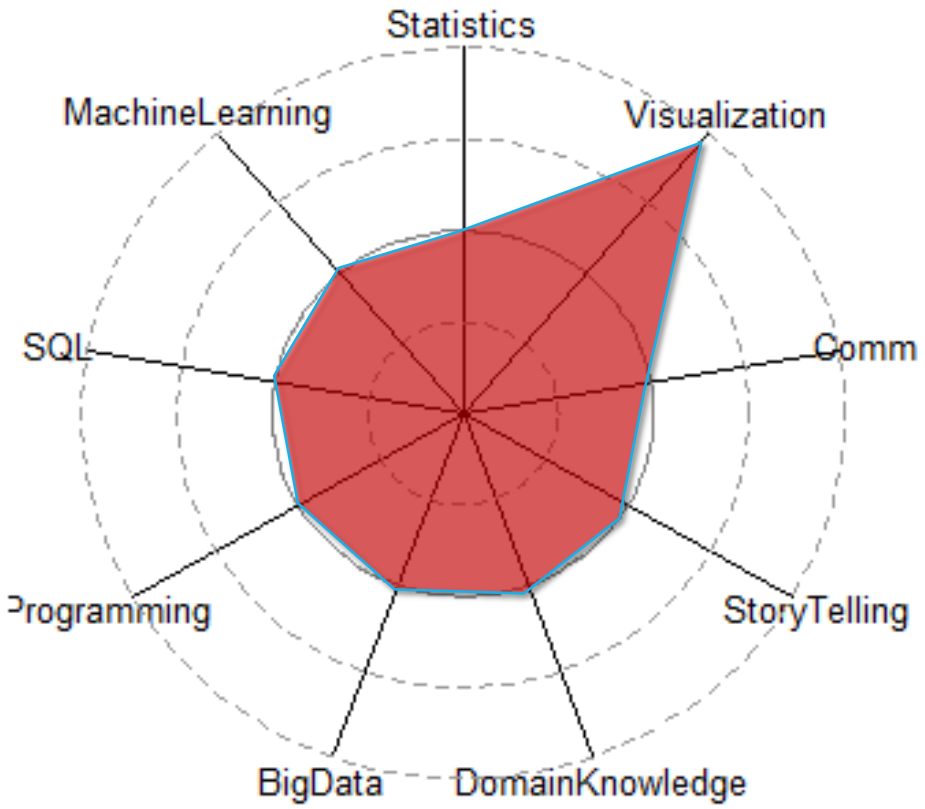
Statistics and Mathematics

Statistical modeling
Causal modeling
Linear algebra
Multi-variable calculus
Optimization



Statistics and Mathematics

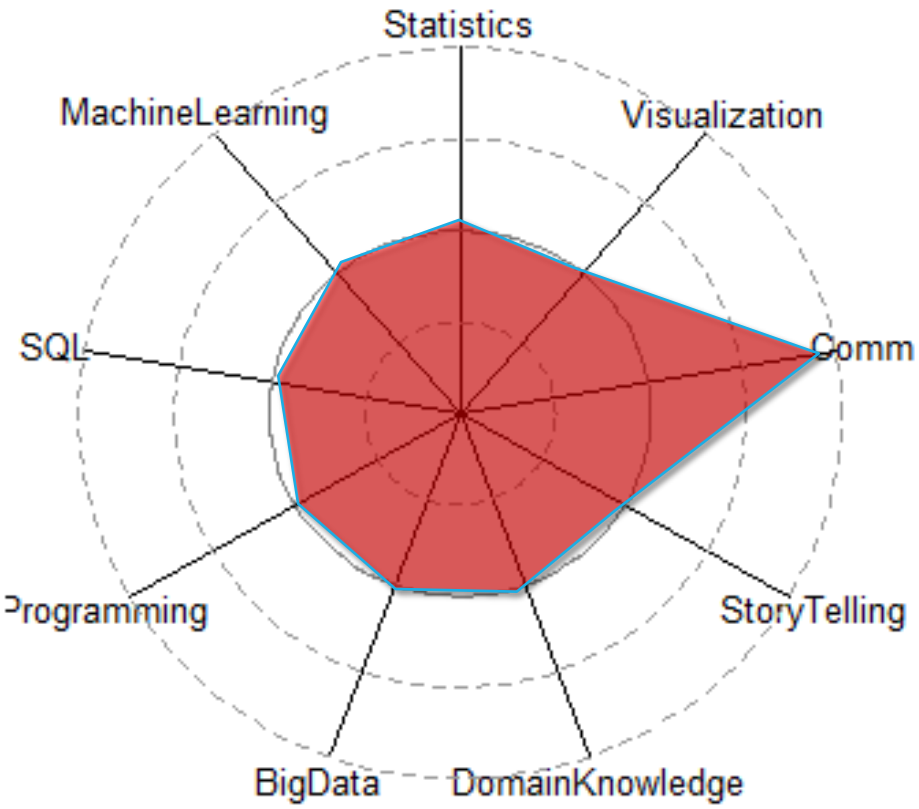
Statistical modeling
Causal modeling
Linear algebra
Multi-variable calculus
Optimization
Bayesian statistics



Visualization

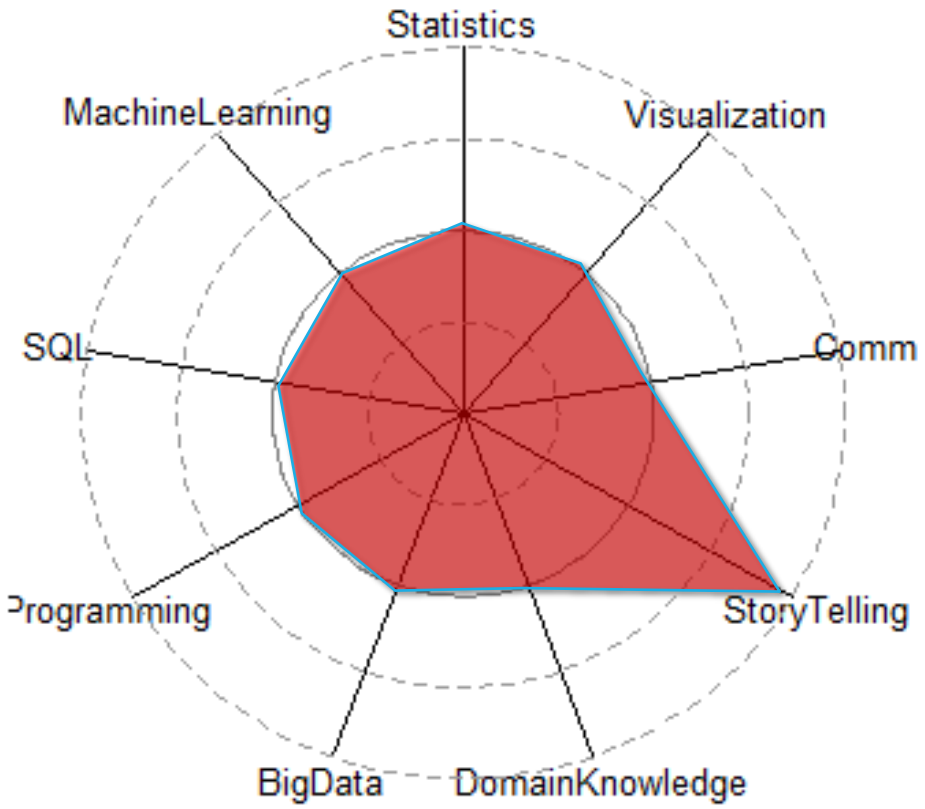
Best practices
Exploratory Data Analysis
Programming-based methods
Cognitive/Perceptual foundations

3 Skills Inventory



Communication

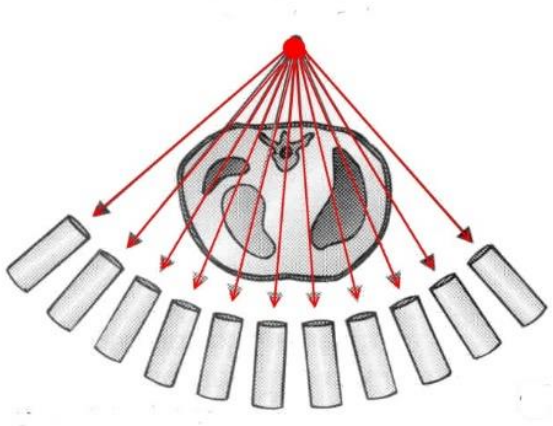
Presentations
Written reports
Accessible style
Adaptability
Executive presentations

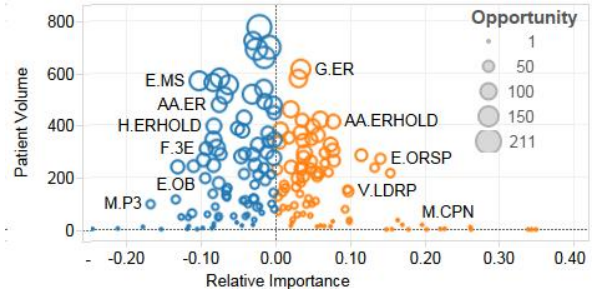


Storytelling

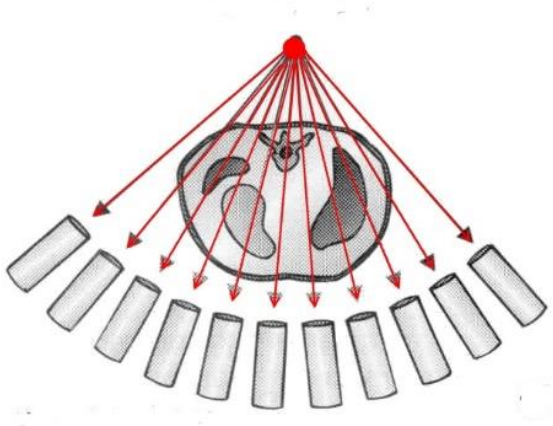
Translating
Framing

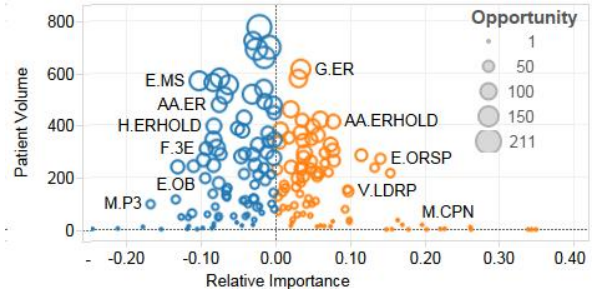
Patient Satisfaction Analysis

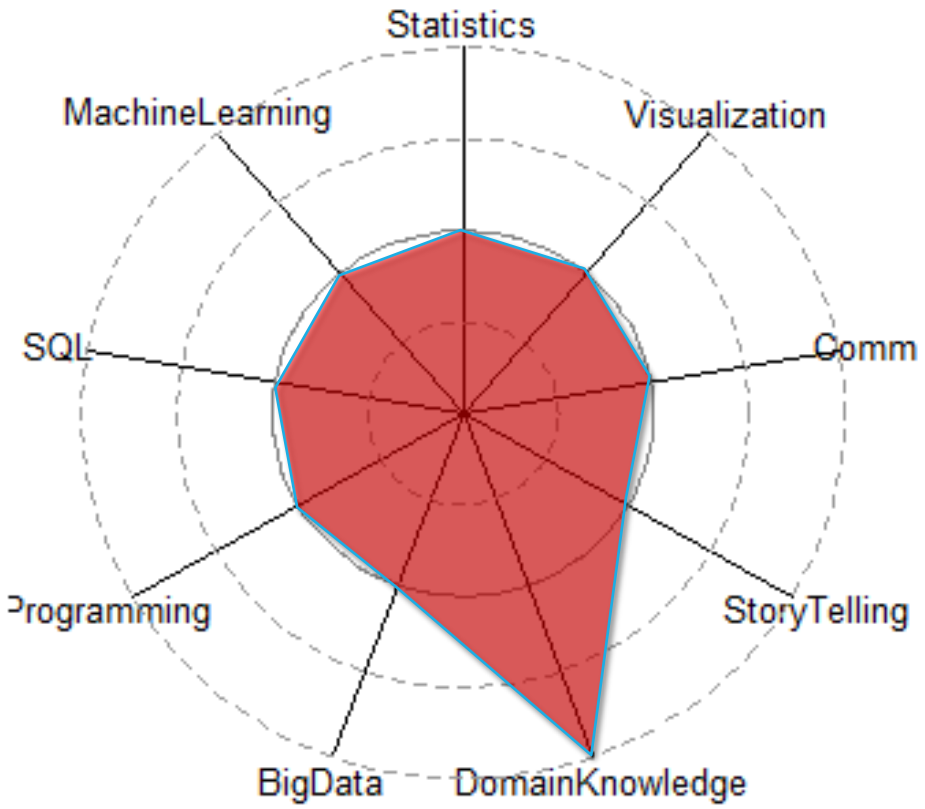




Patient Satisfaction **CT Scan**

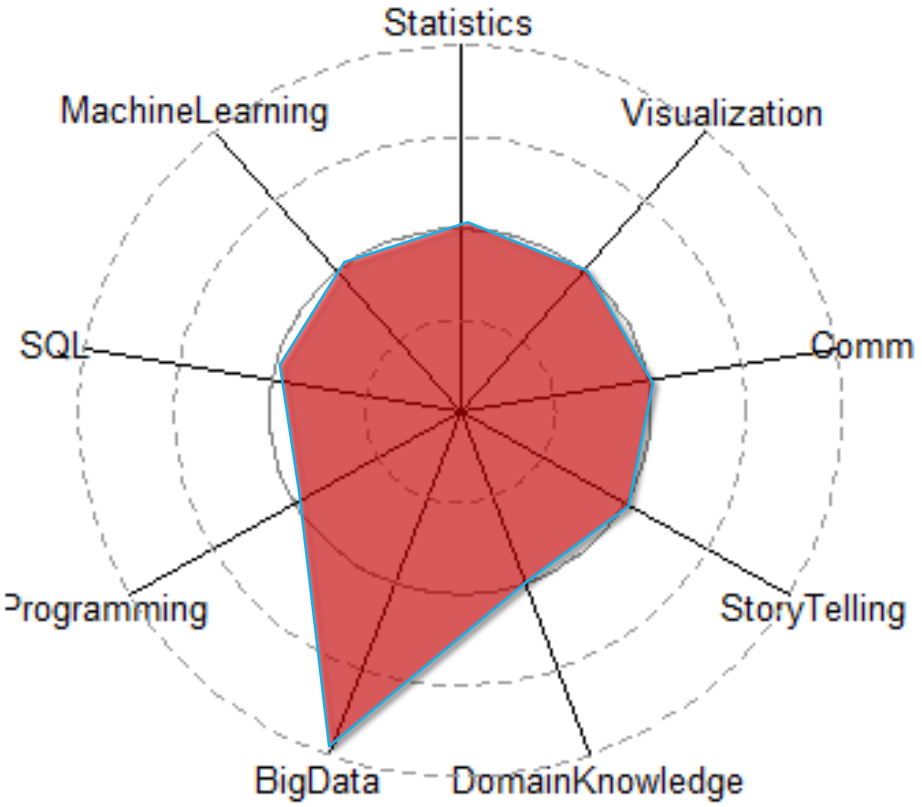






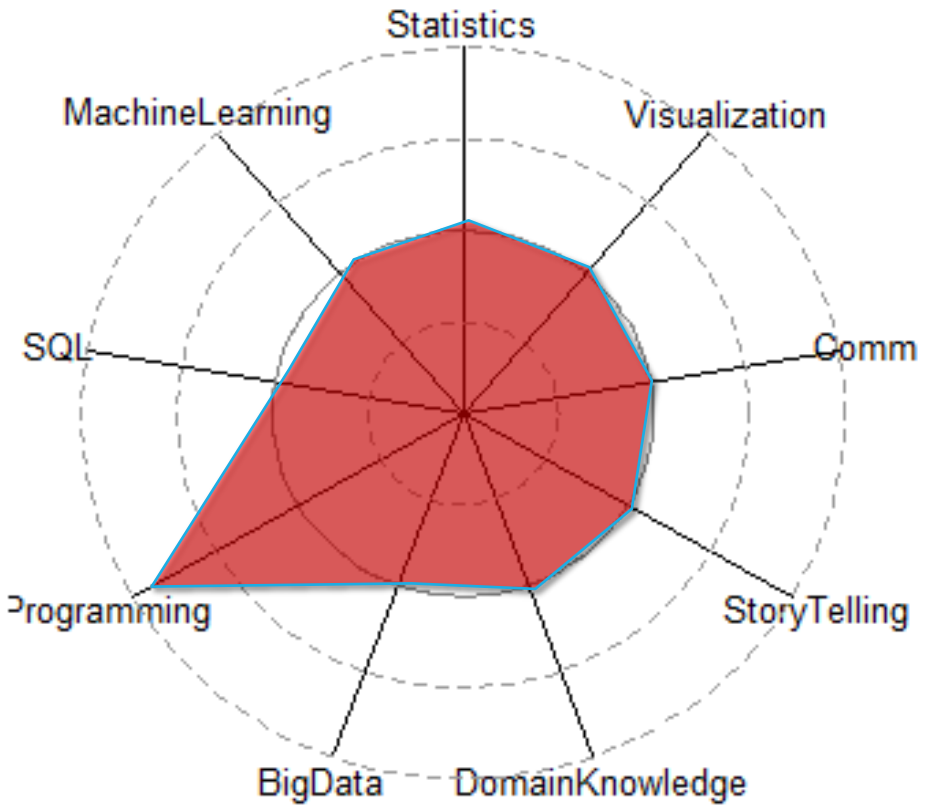
Domain Knowledge

On-the-job
Specialized knowledge



Big Data Technologies

Specialized
Most do not require

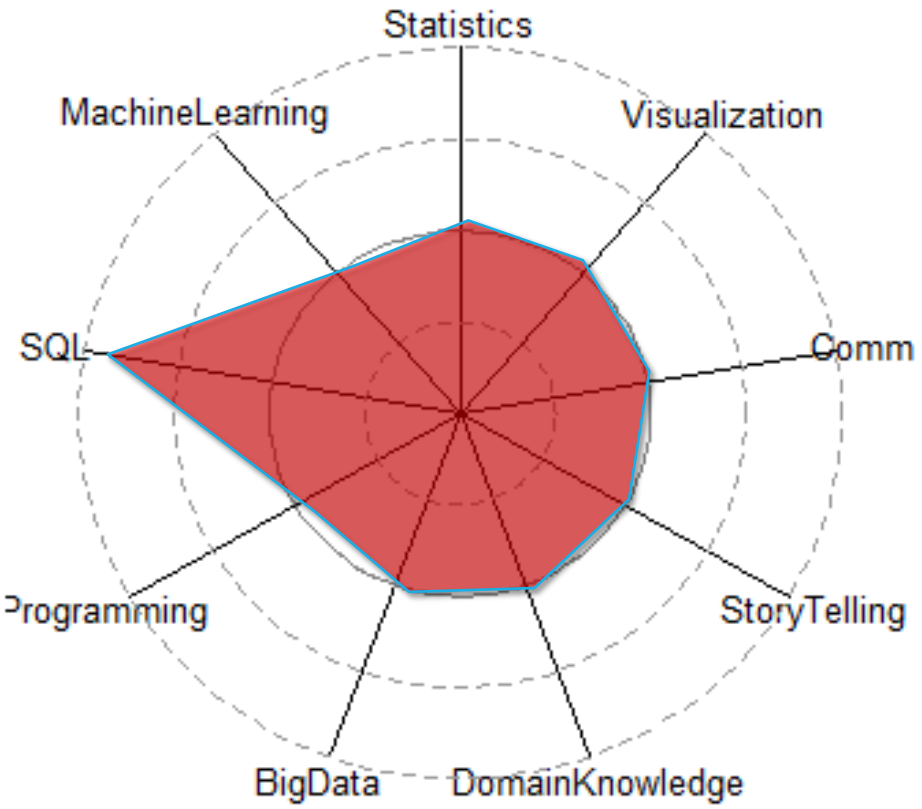


Programming

Python
R

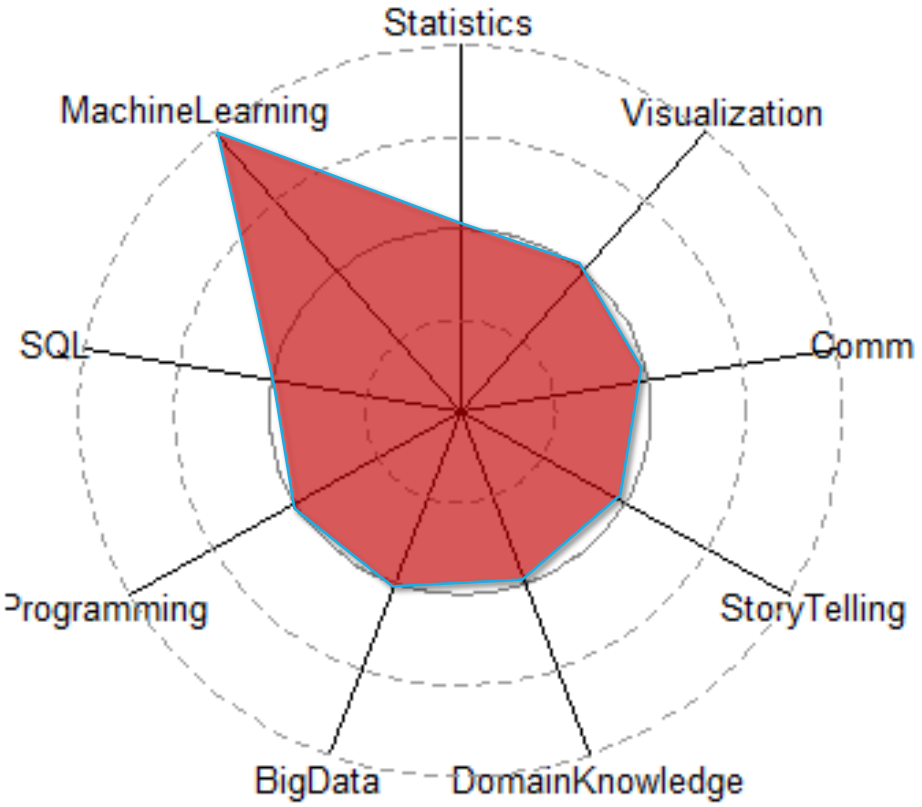
Java
Scala
(Clojure)

GPU frameworks



SQL

Some experience essential
Self-study
Working knowledge
Complex queries
Subqueries
Optimization
Expertise in multiple versions



Machine Learning/Deep Learning

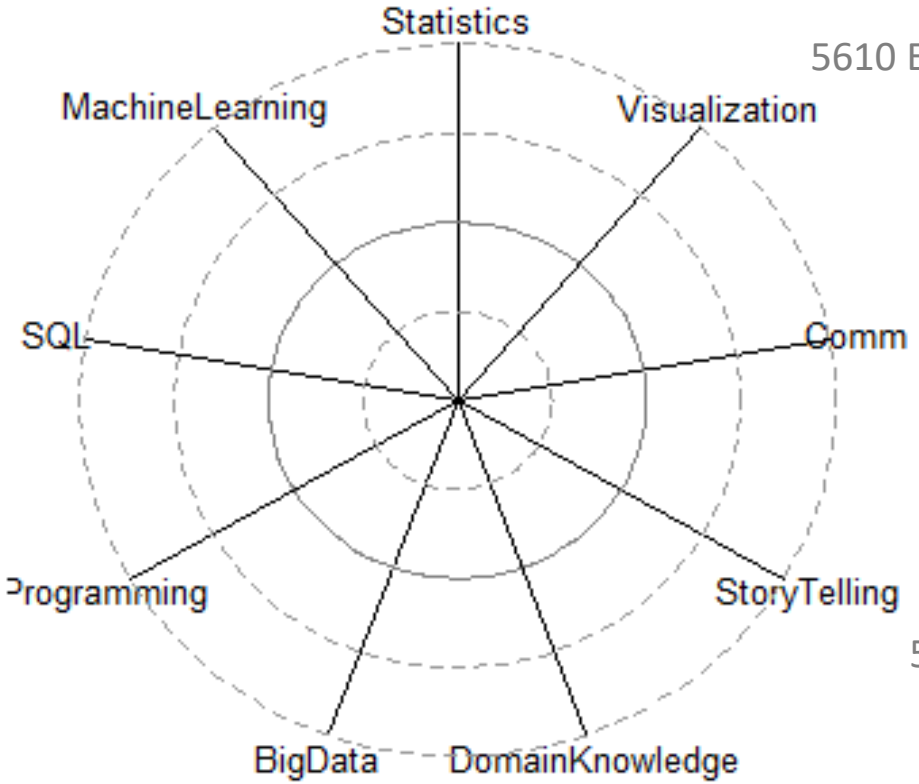
Experience with non-class projects
Understanding of major ML methods
Experience with major DL frameworks
Experience with popular pre-trained models
Well-versed in theory
Design experience with CNNs/RNNs
Expertise creating novel deep learning solutions

Skills Inventory

5620 Probability & Statistical Inference

5640 Modeling & Machine Learning I
5660 Modeling & Machine Learning II

5610 Exploratory Data Analysis



5999 Capstone Development

5420 Data Management Systems

5220 Principles of Programming and Simulation

5440 Data Science Algorithms

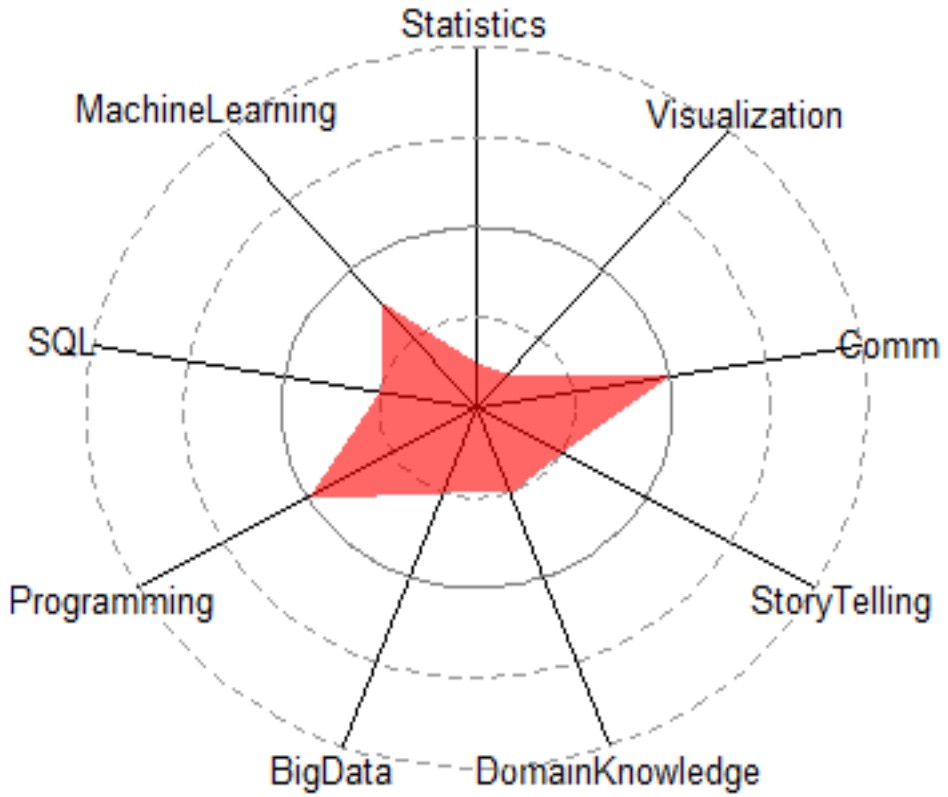
5320 Survey of Data Science Applications

5360 Case Studies in Data Science

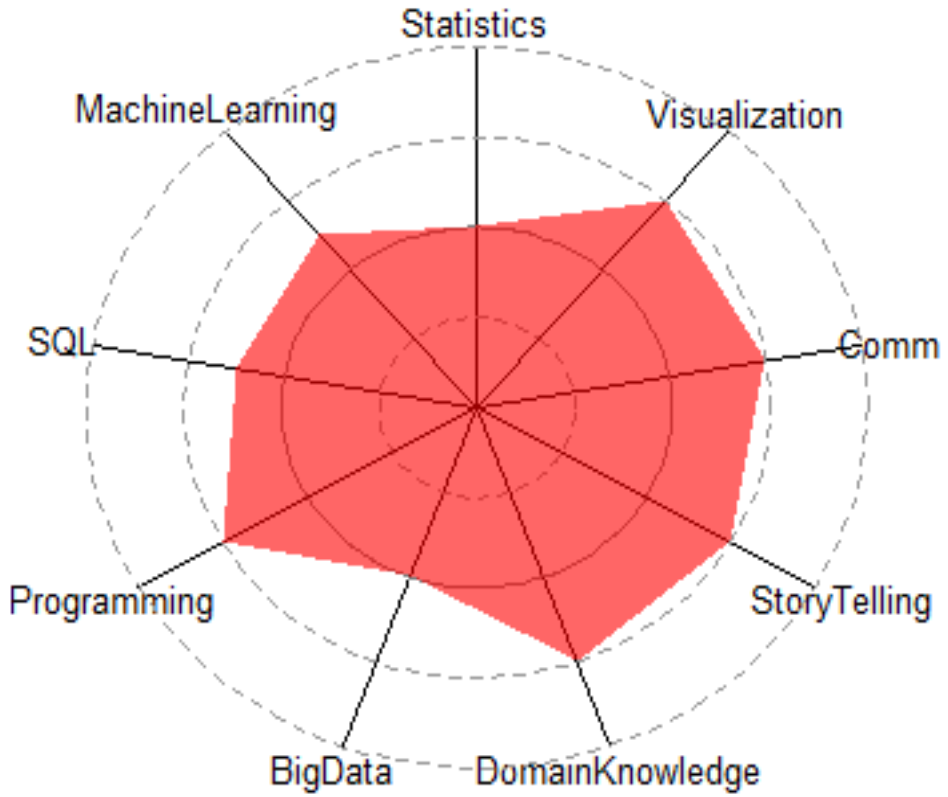
5460 Big Data & Scaling

5380 Data Science in Teams I
5390 Data Science in Teams II
5340 Data Science Rights & Responsibilities

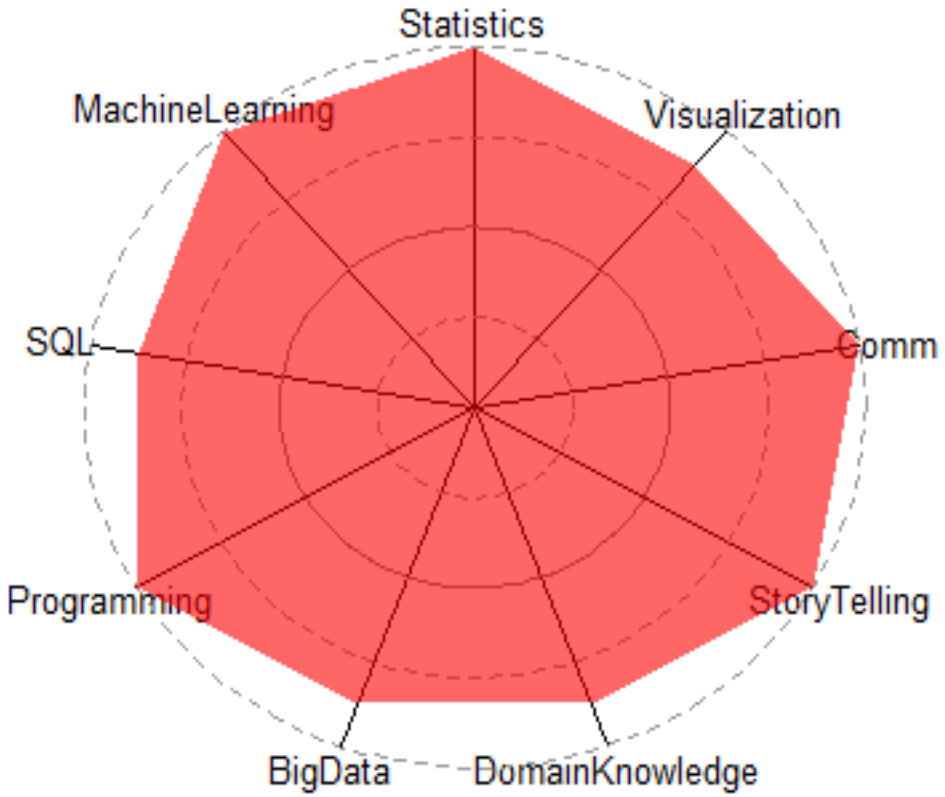
Data Scientist I



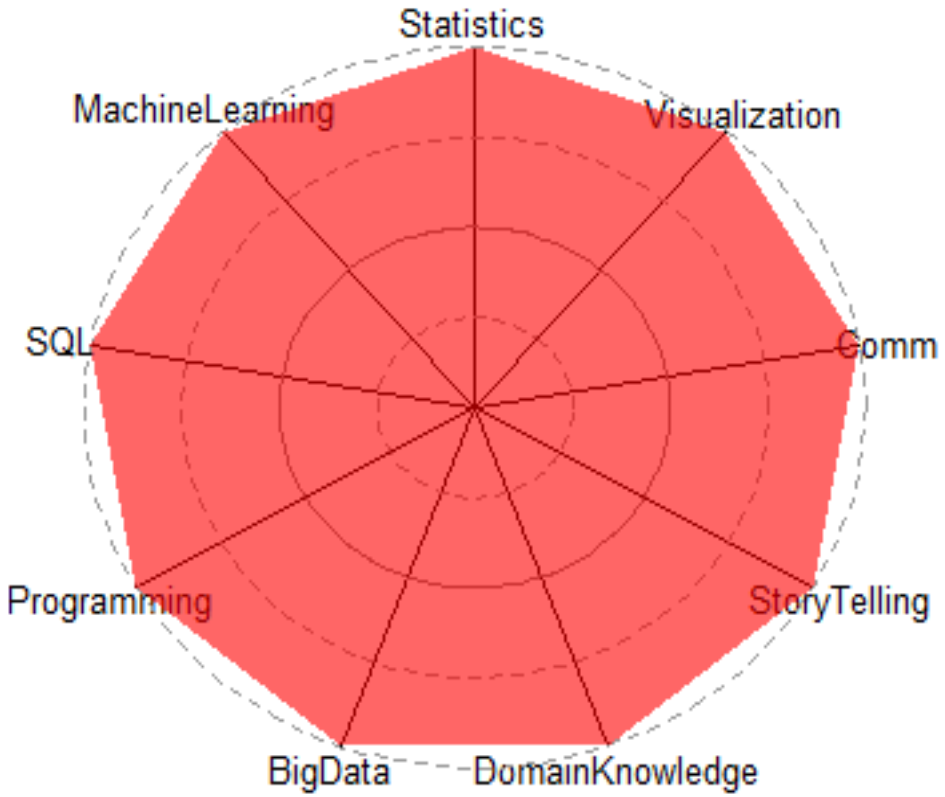
Data Scientist II

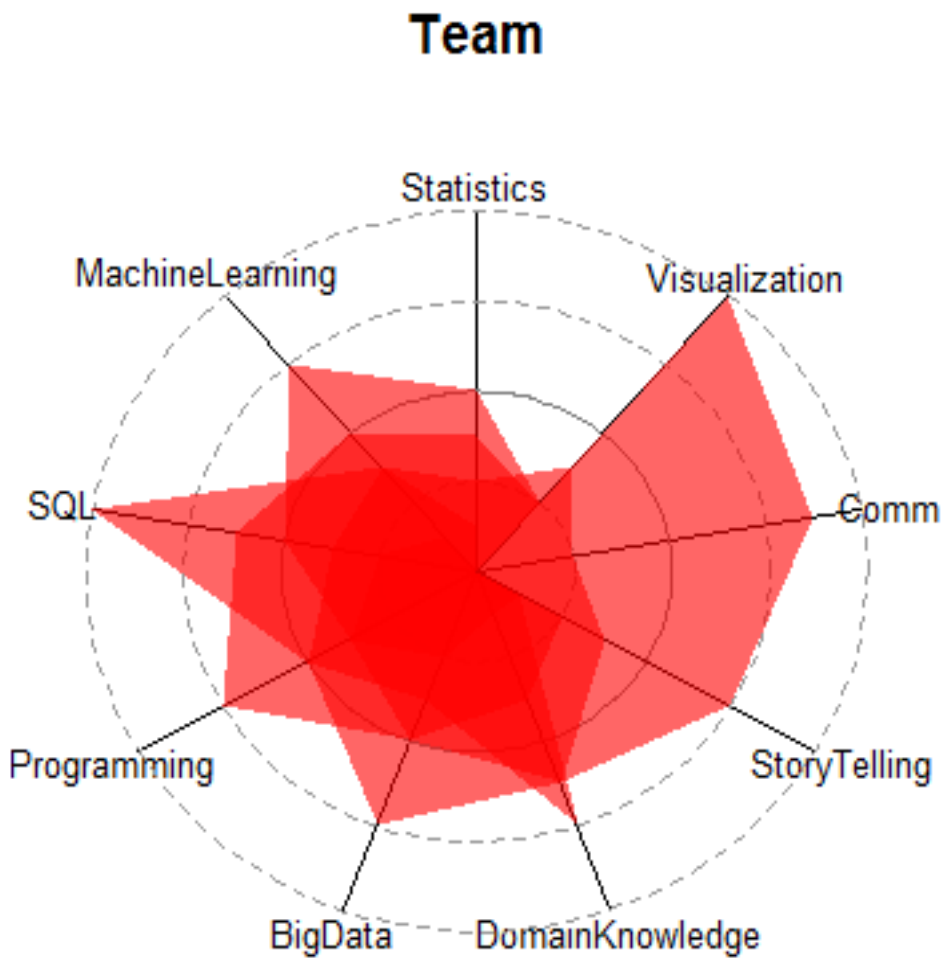


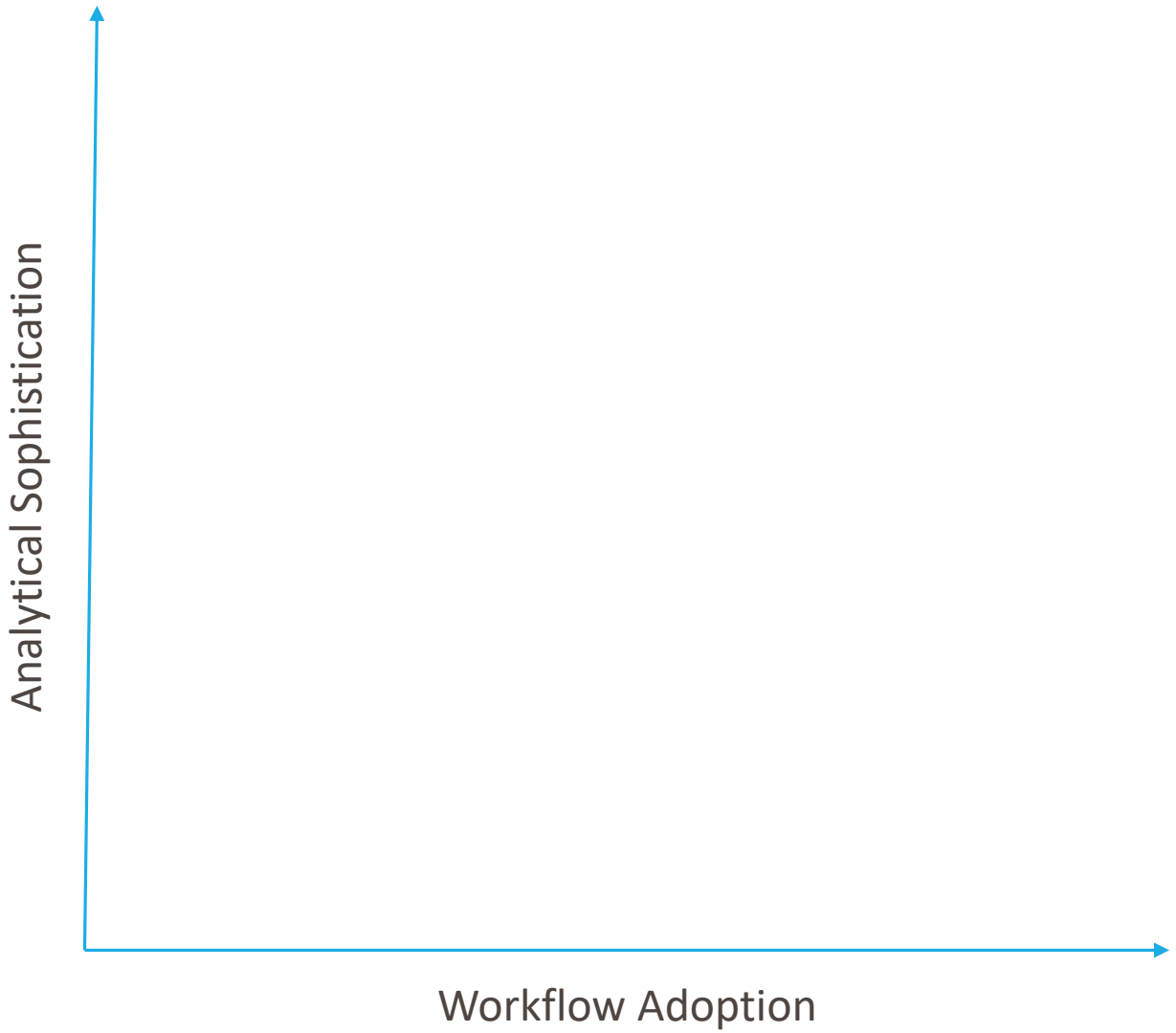
Sr. Data Scientist



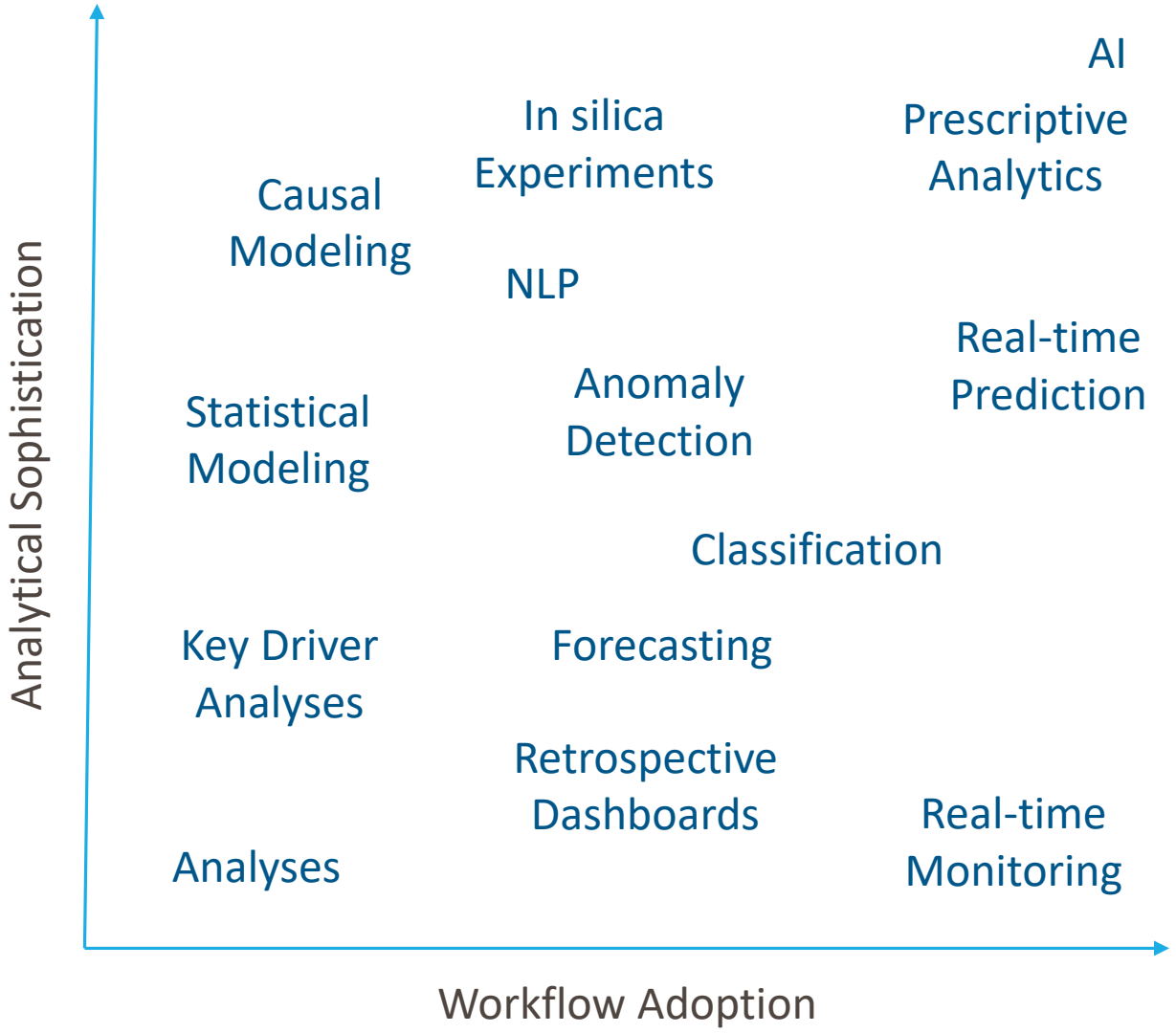
Consulting Data Scientist



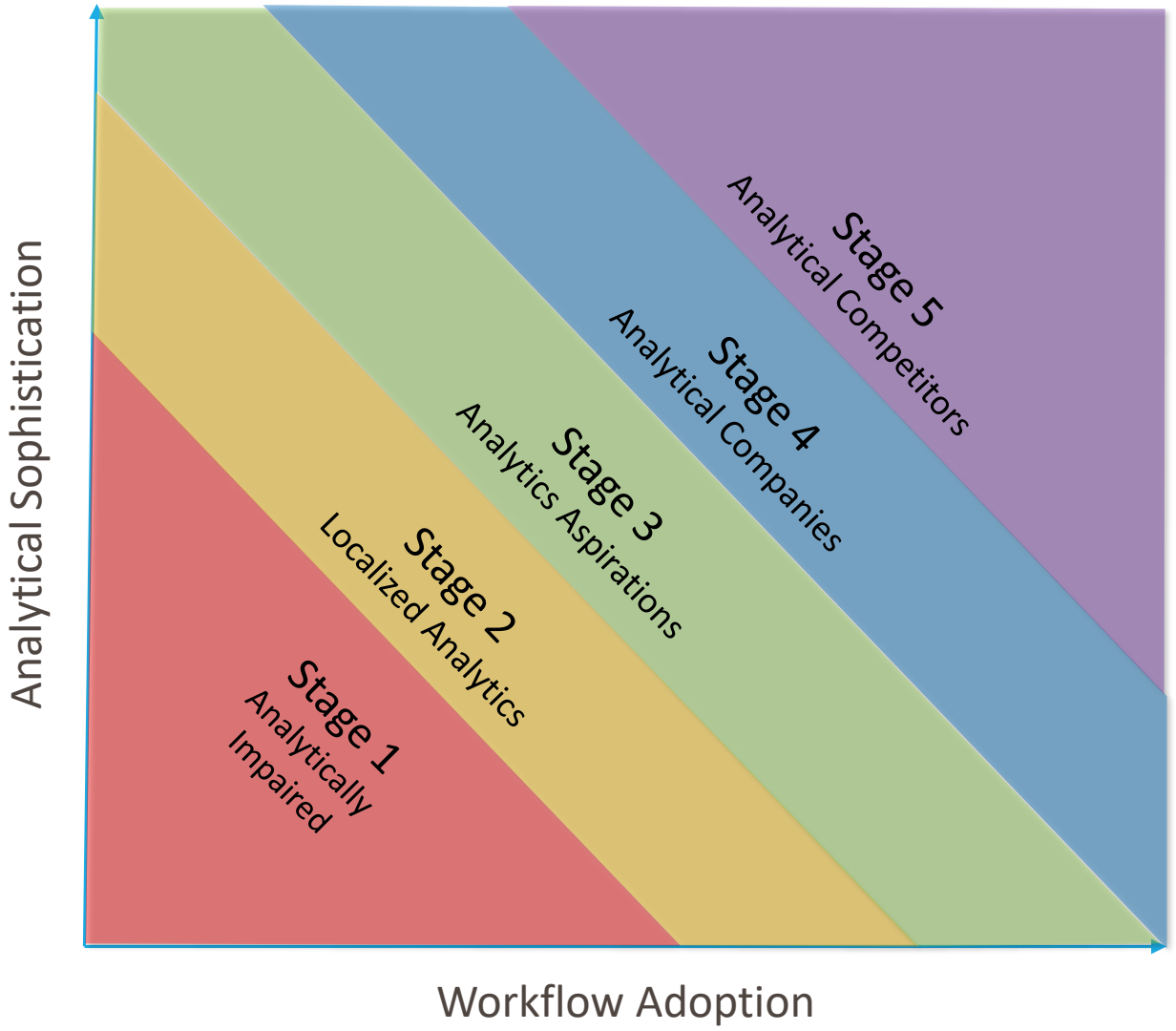

Team
Statistics
Visualization
Comm
StoryTelling
DomainKnowledge
BigData
Programming
SQL
MachineLearning


Analytical Sophistication
Workflow Adoption

Data Science Maturity Models



Data Science Maturity Models



D accessible, high-quality data

E enterprise orientation

L analytical leadership

T strategic targets

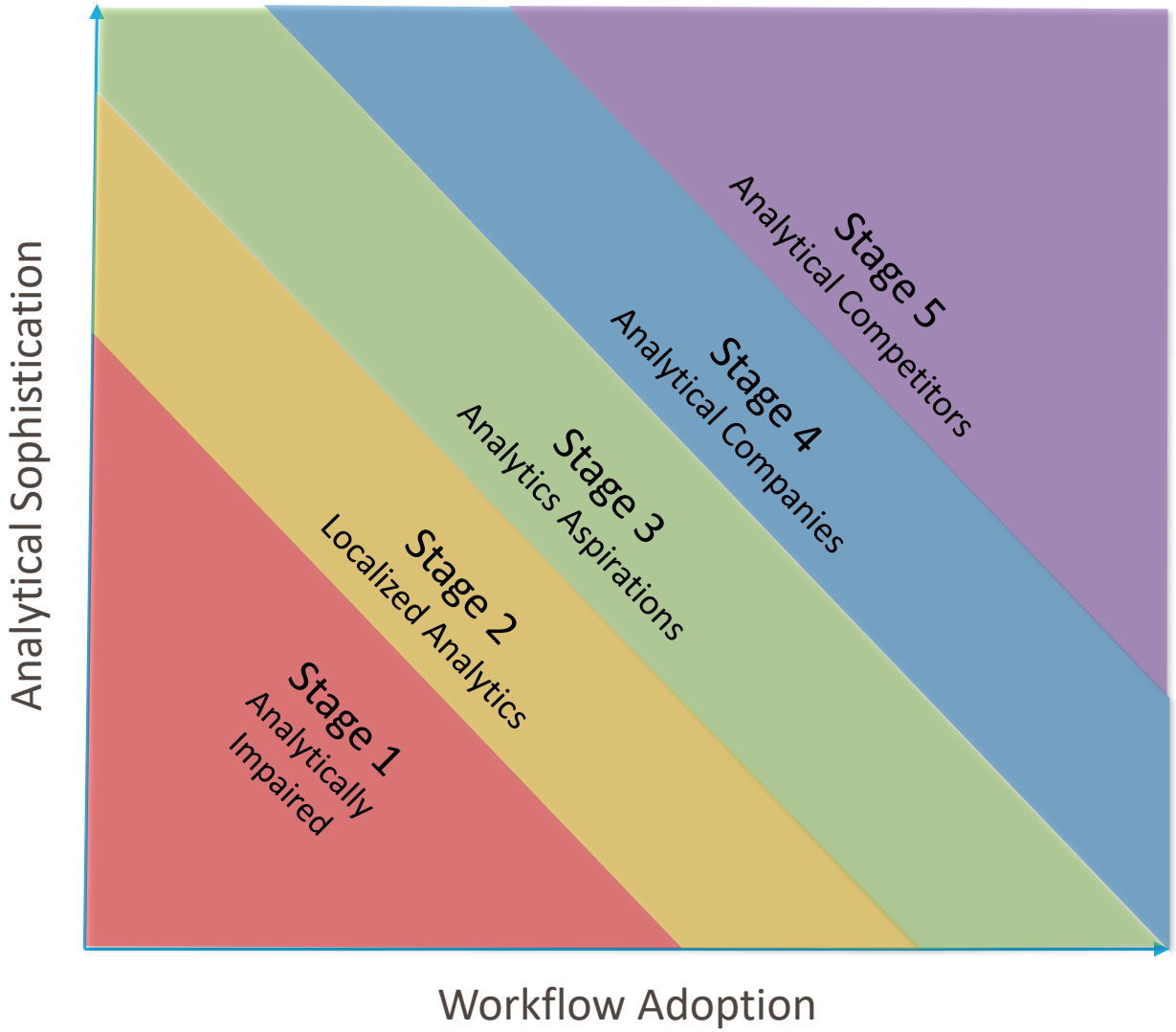
A analysts

A adoption readiness

I infrastructure to run on

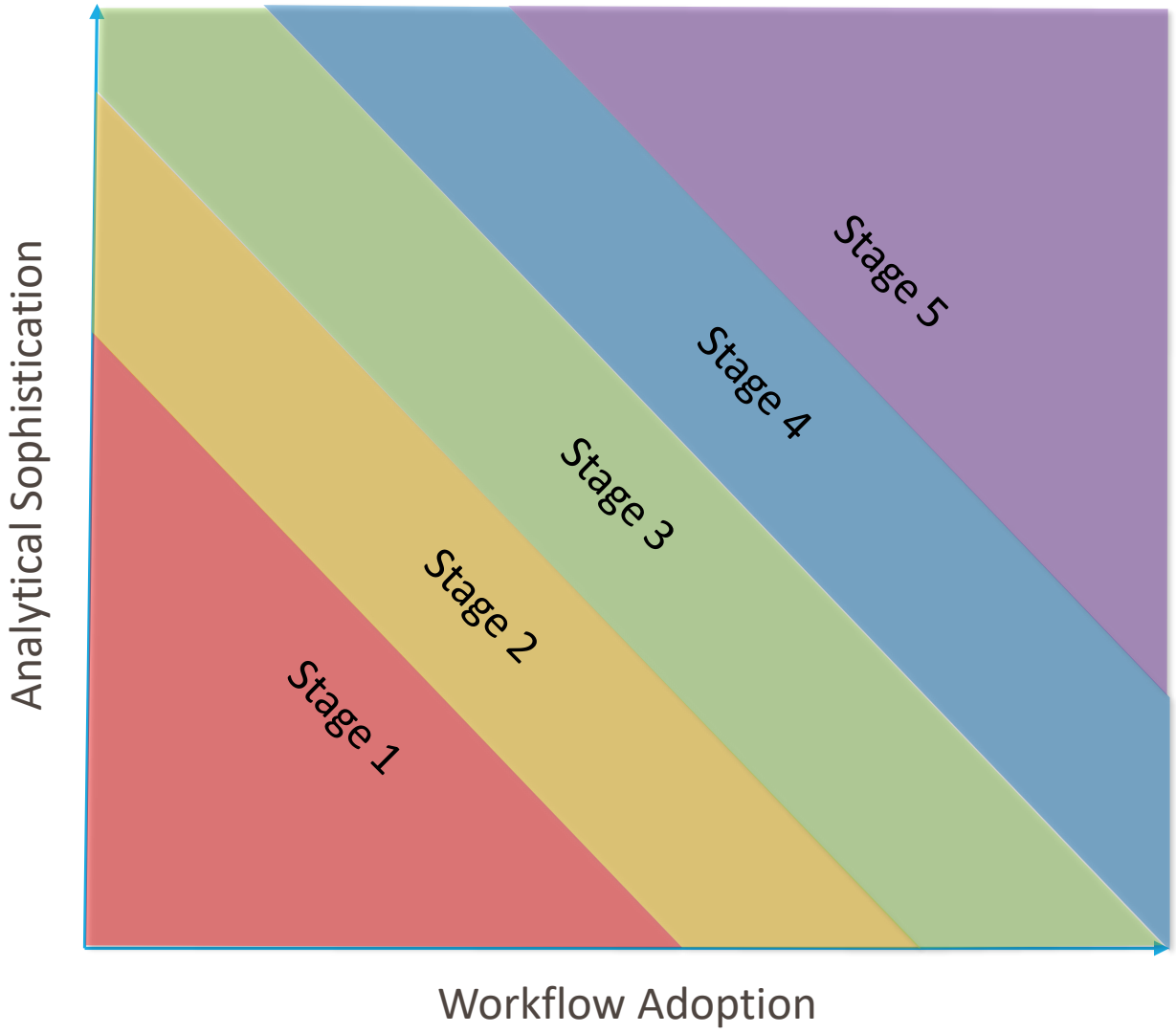
Data Science Maturity Models

2 Maturity Graph



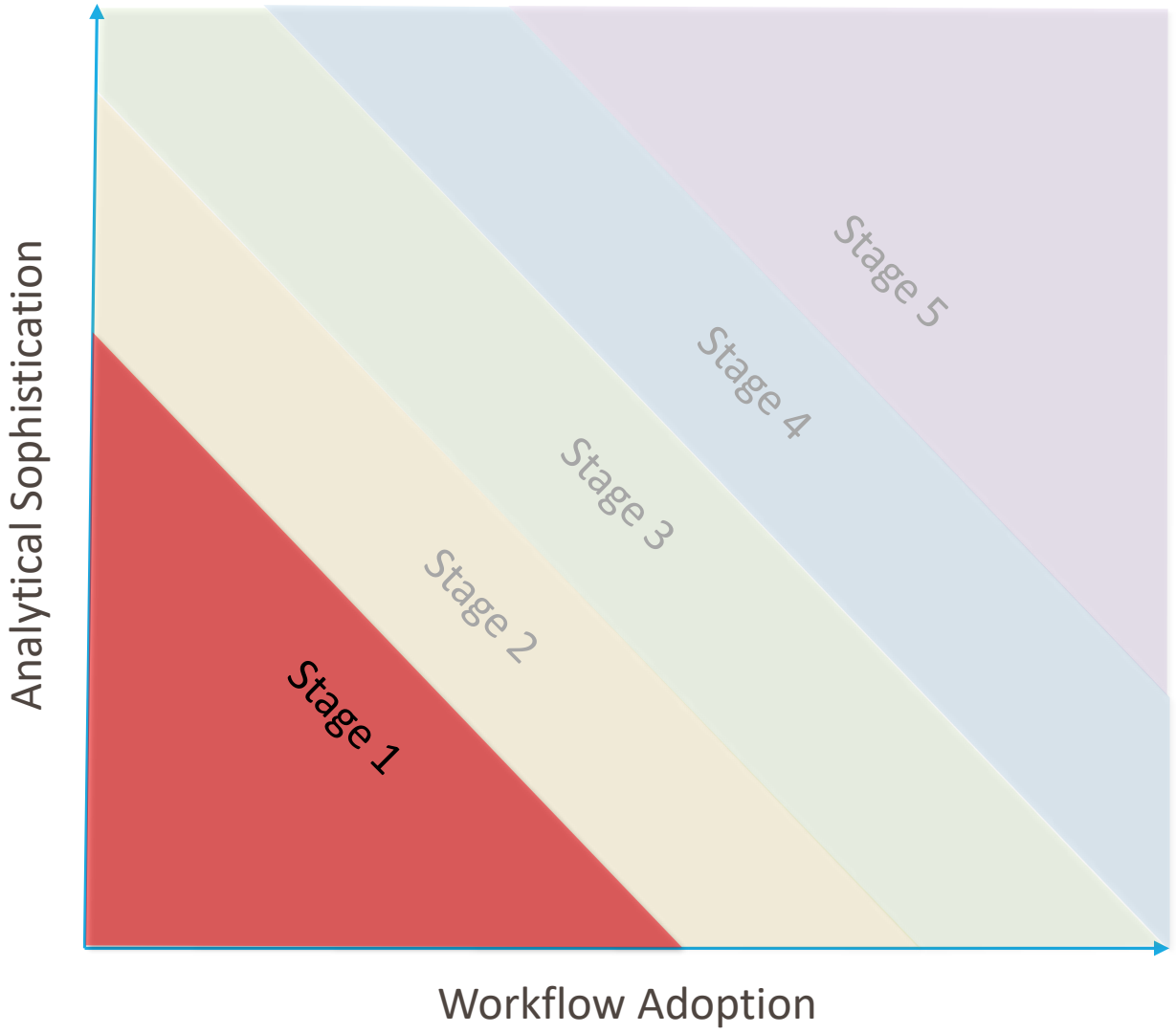
Data Science Maturity Models

2 Maturity Graph



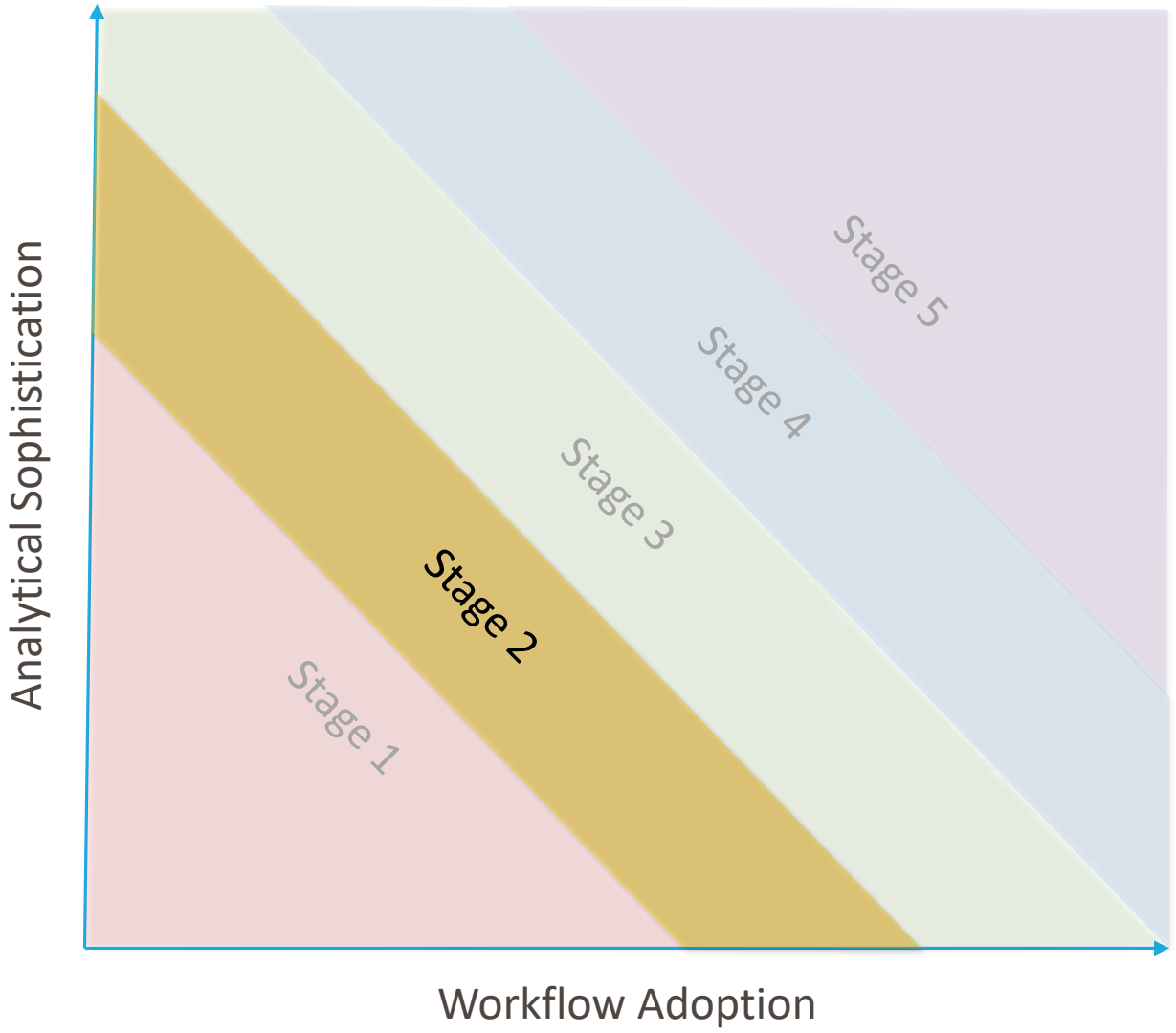
Data Science Maturity Models

2 Maturity Graph



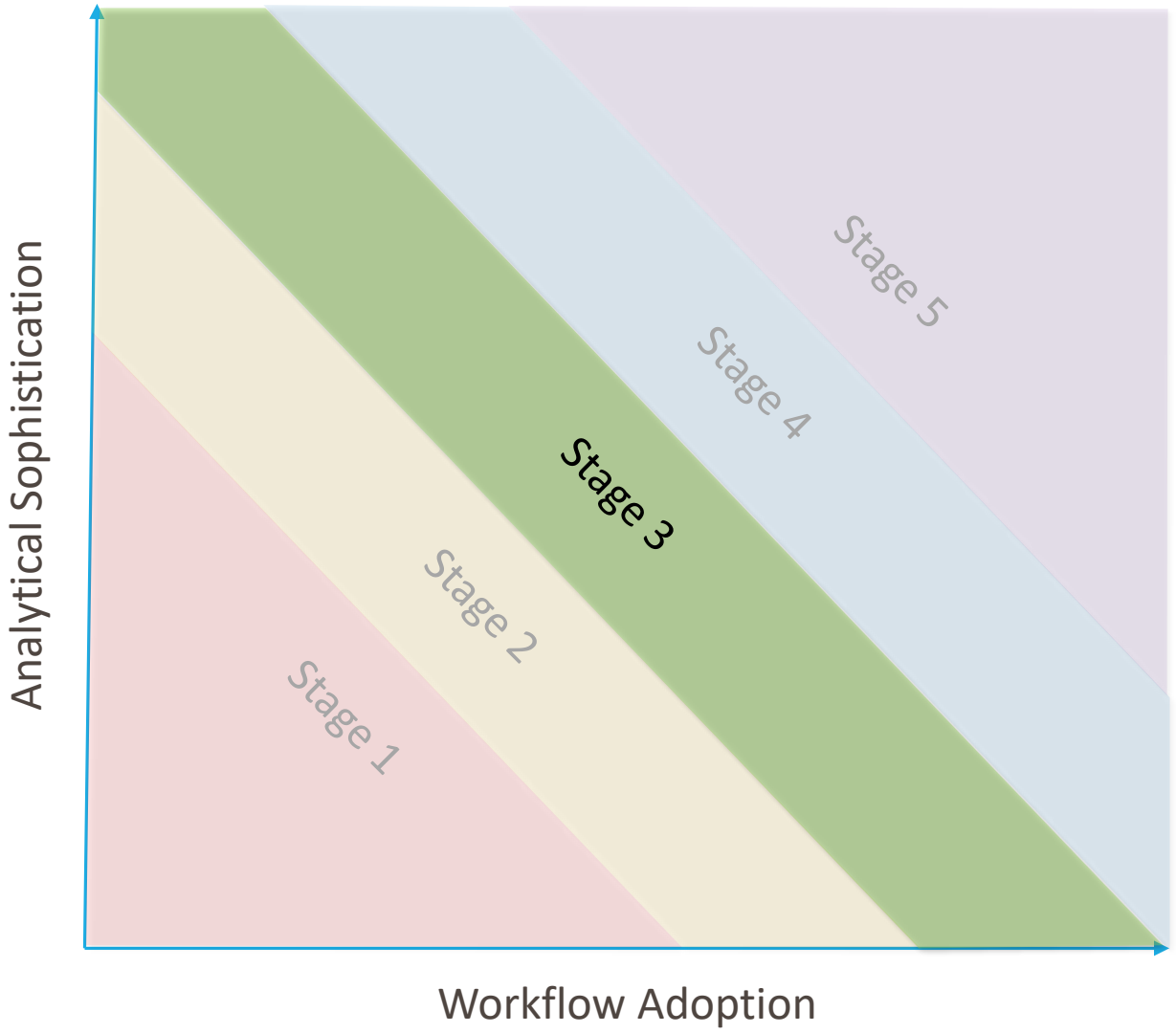
Data Science Maturity Models

2 Maturity Graph

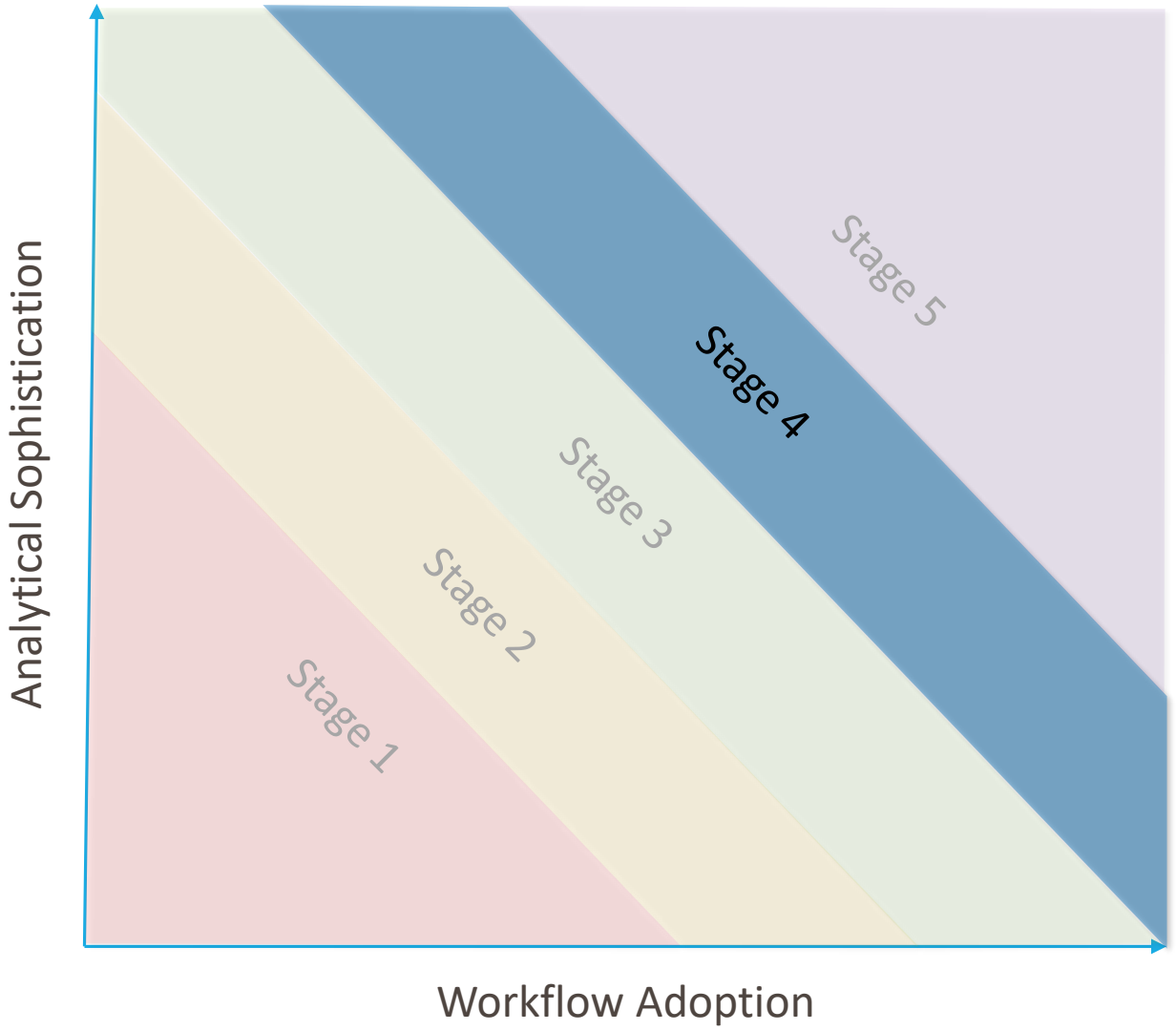


Data Science Maturity Models

2 Maturity Graph



Data Science Maturity Models

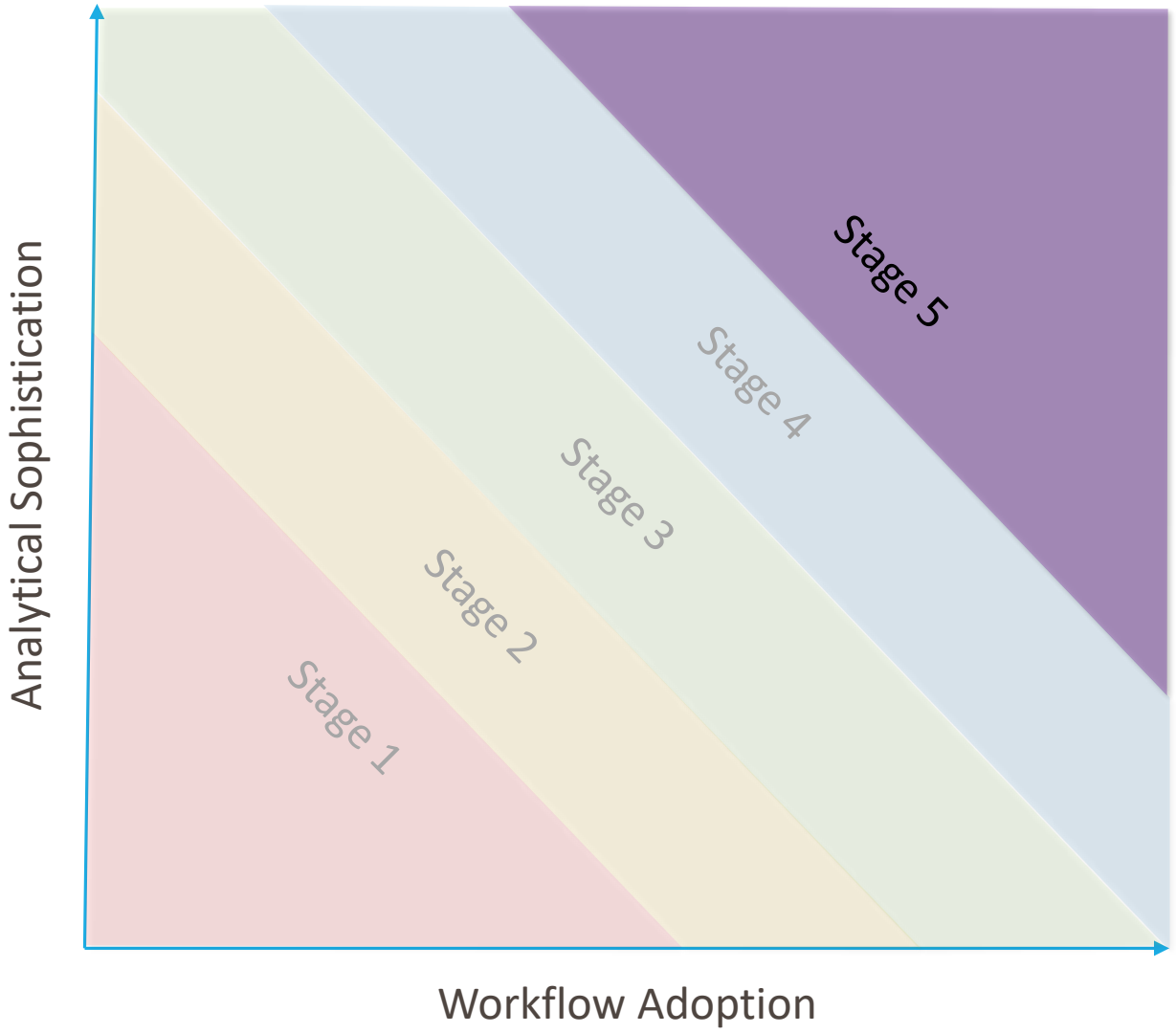


In a data-driven business, the data helps the business; in a model-driven business, the models *are* the business.

- Steven A. Cohen and Matthew W. Granade, Wall Street Journal, 8/19/2018

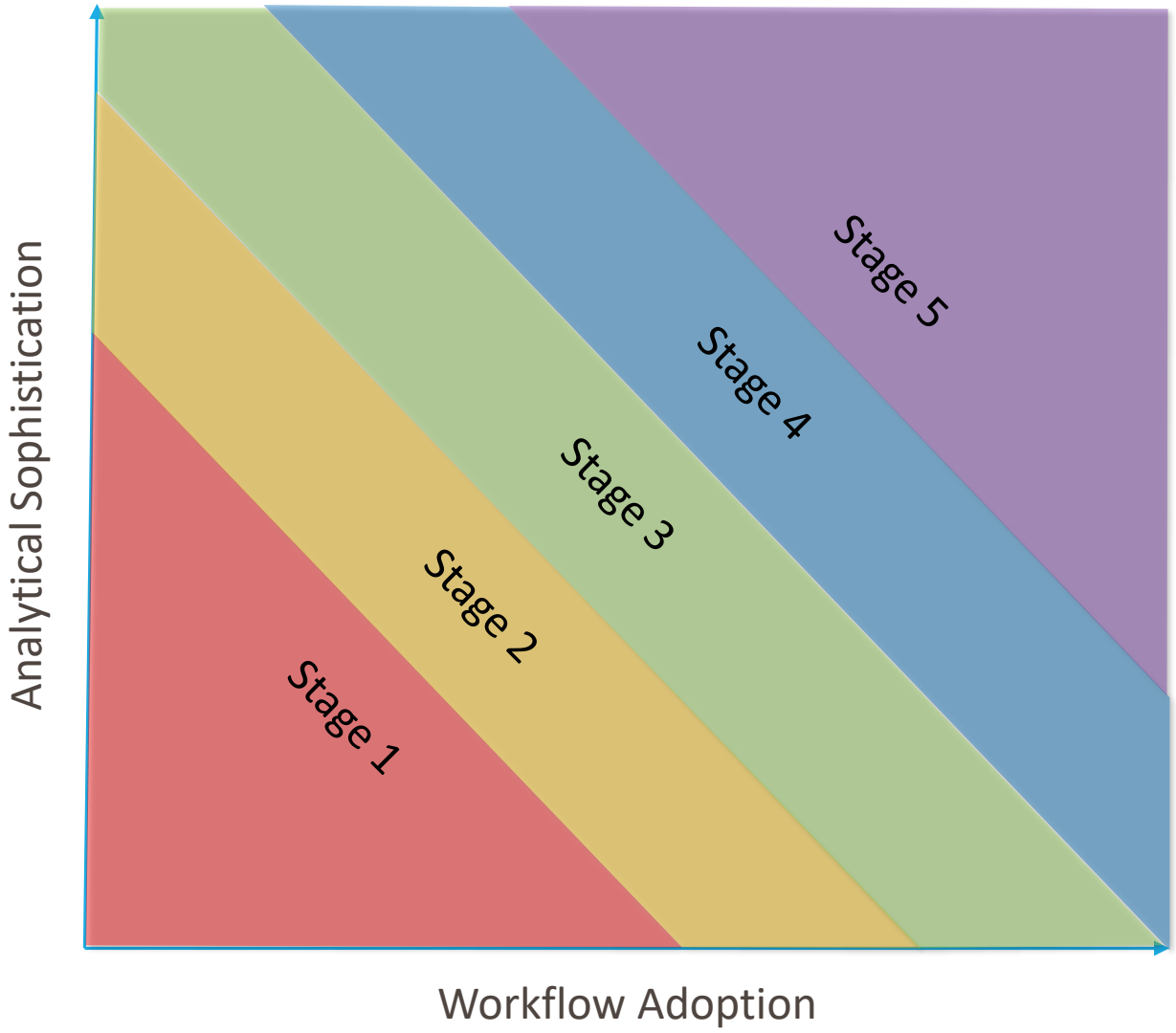
Data Science Maturity Models

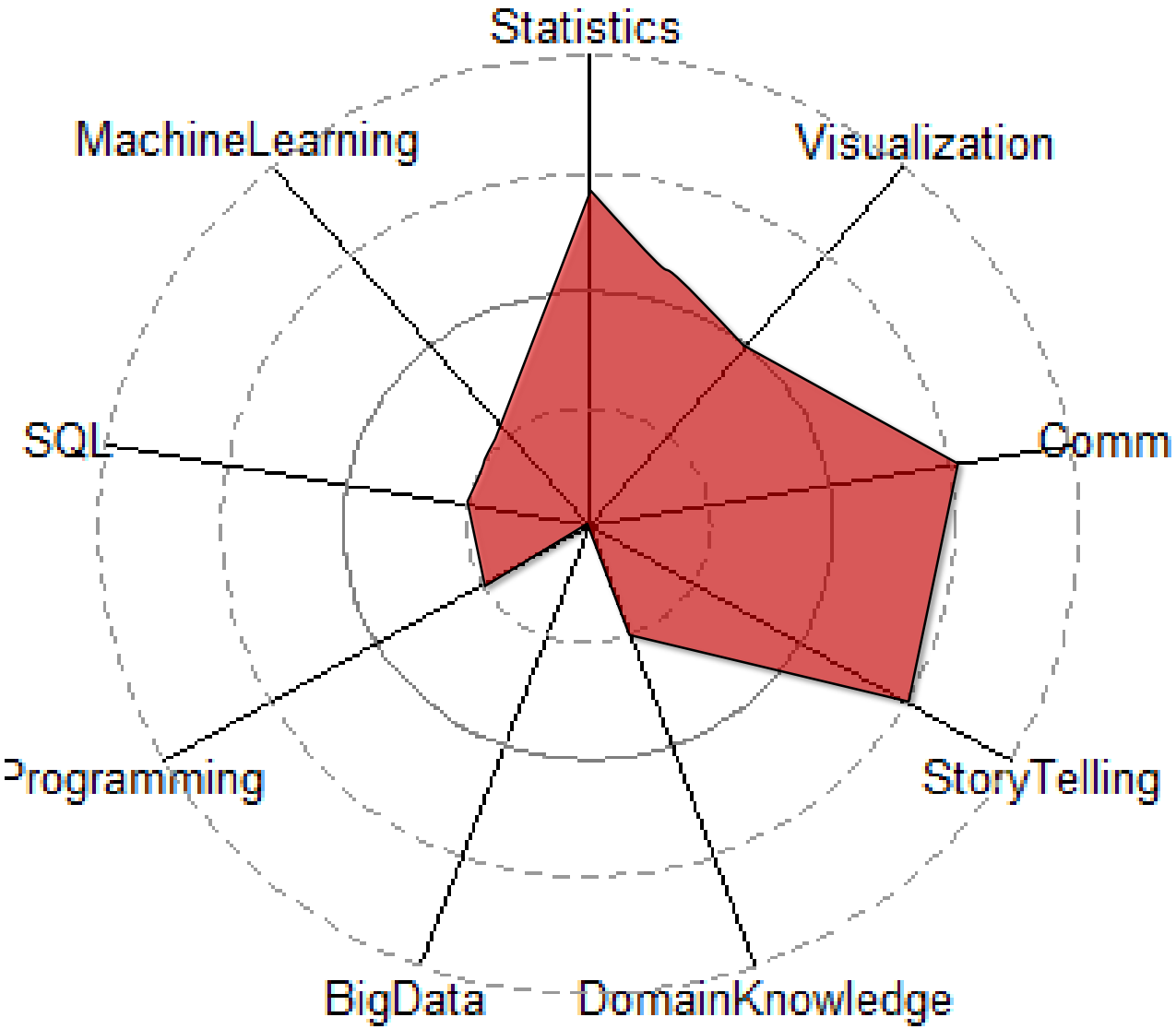
2 Maturity Graph

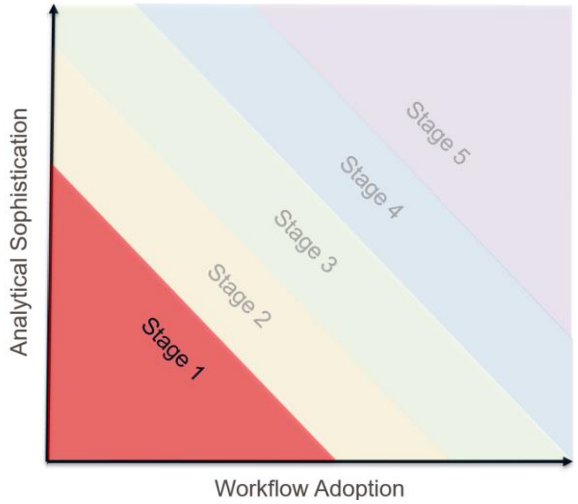


Data Science Maturity Models

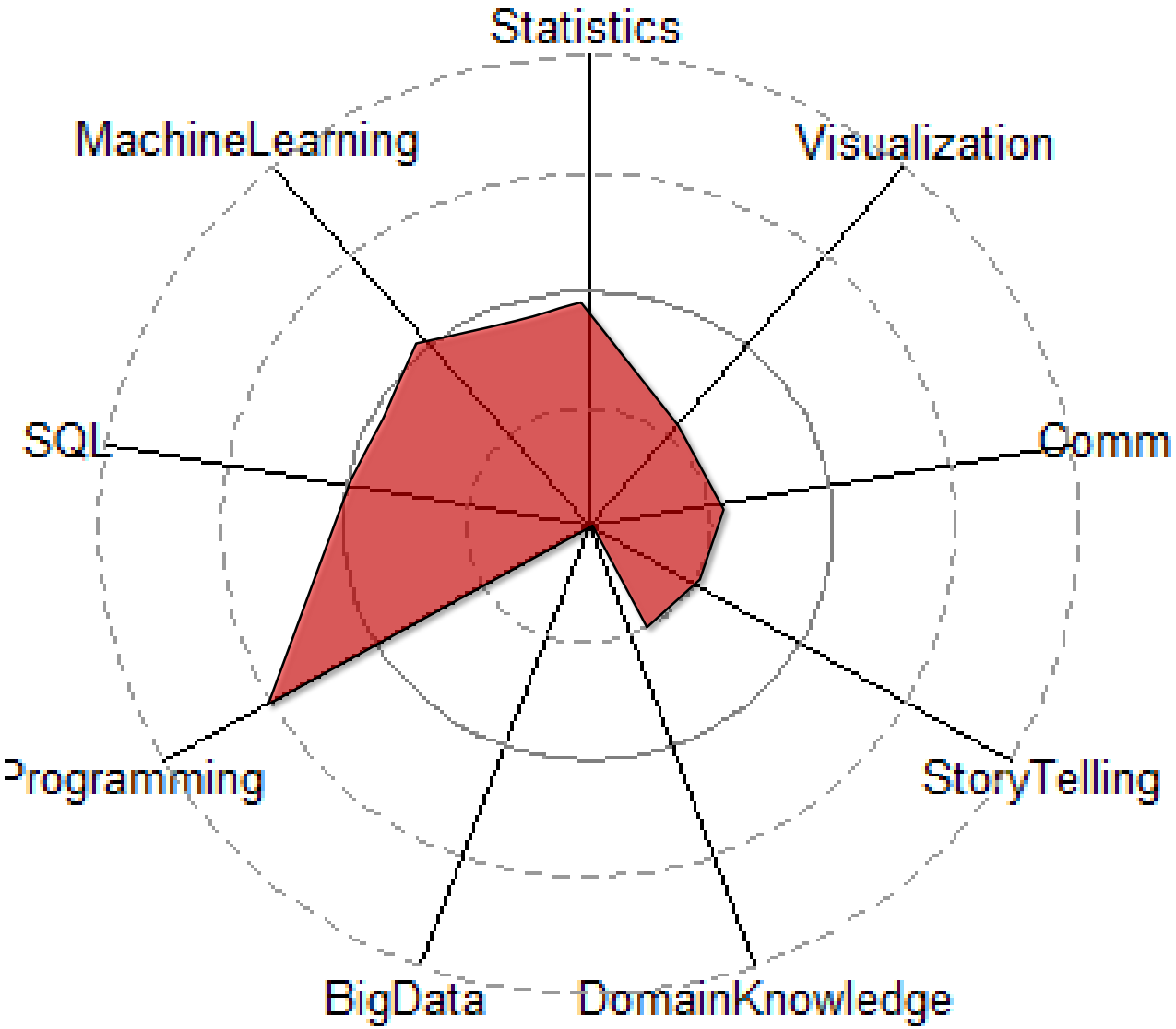
2 Maturity Graph

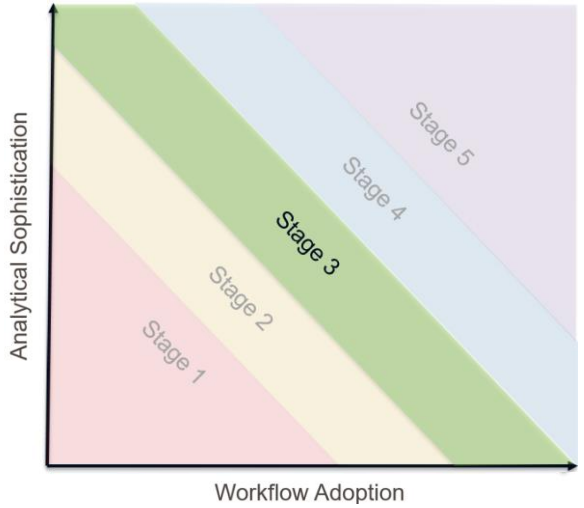



Statistics
MachineLearning
Visualization
SQL
Comm
Programming
StoryTelling
BigData
DomainKnowledge

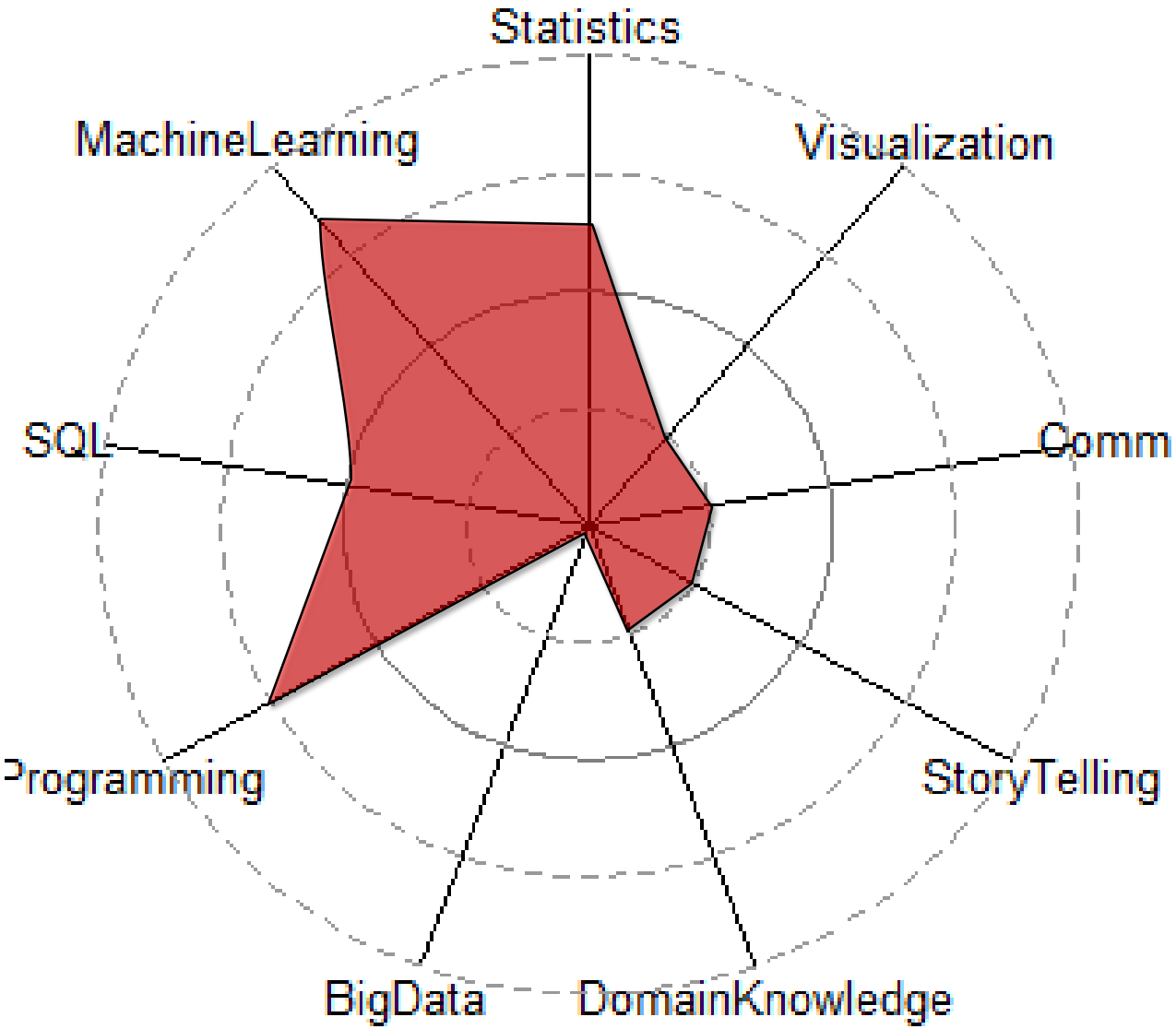

Analytical Sophistication
Workflow Adoption
Stage 1
Stage 2
Stage 3
Stage 4
Stage 5

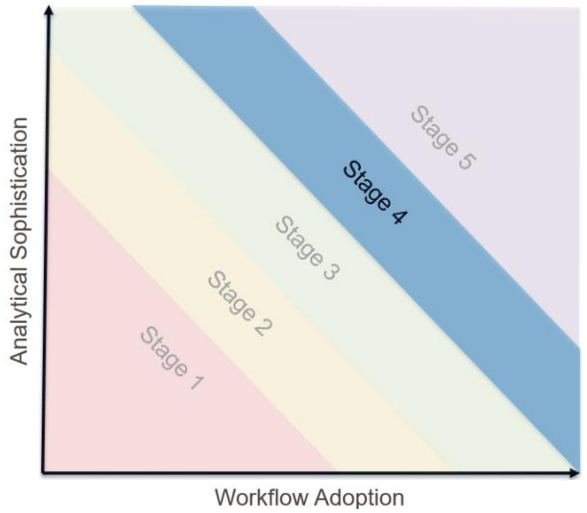
The Guide


Statistics
MachineLearning
Visualization
SQL
Comm
Programming
StoryTelling
BigData
DomainKnowledge

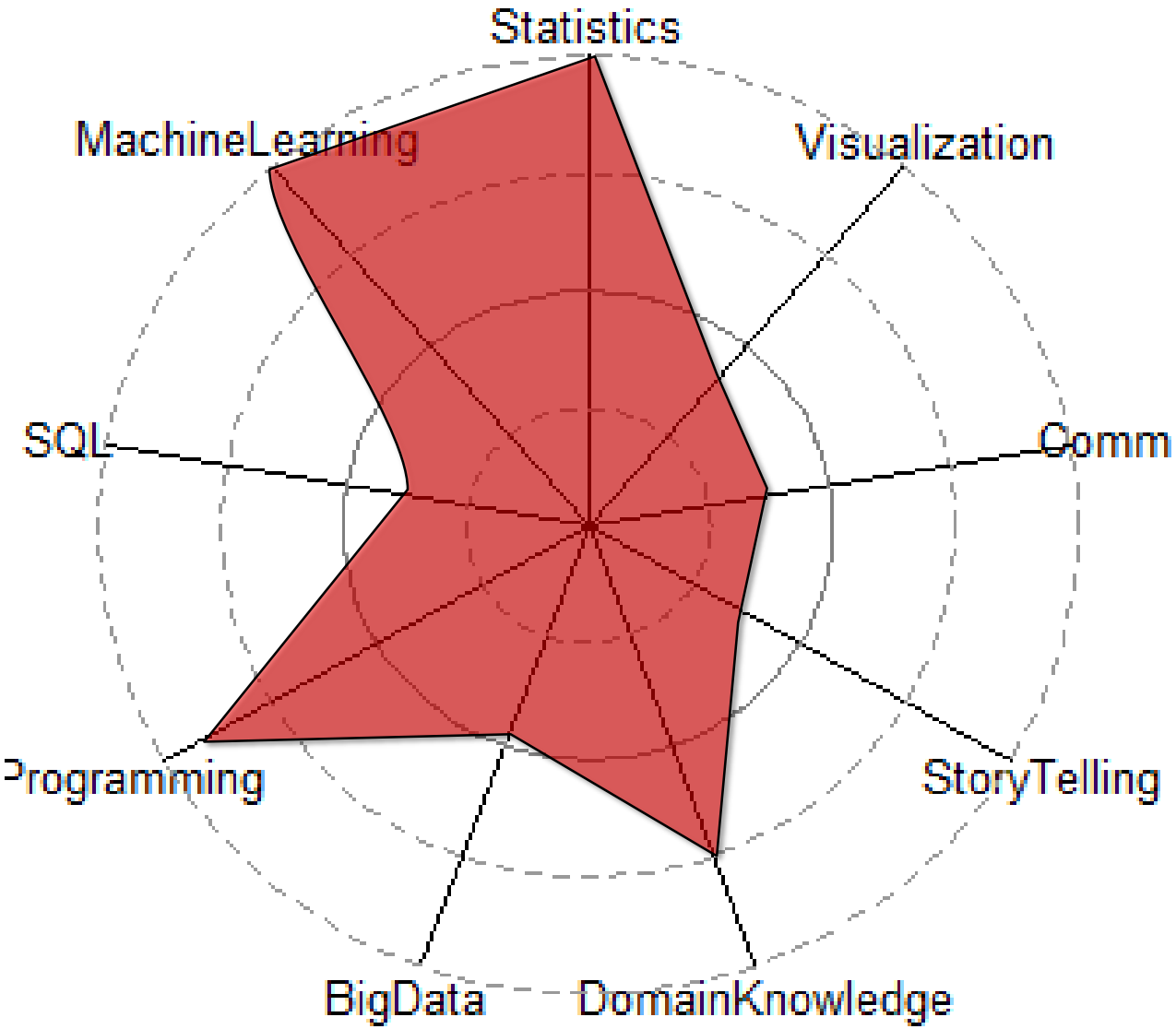

Analytical Sophistication
Workflow Adoption
Stage 1
Stage 2
Stage 3
Stage 4
Stage 5

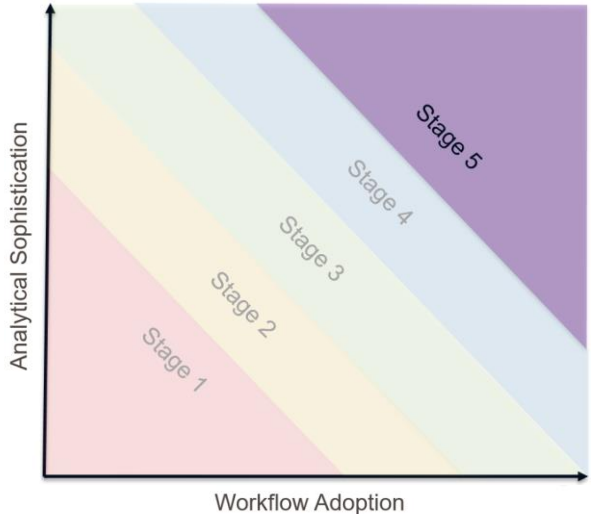
The Builder





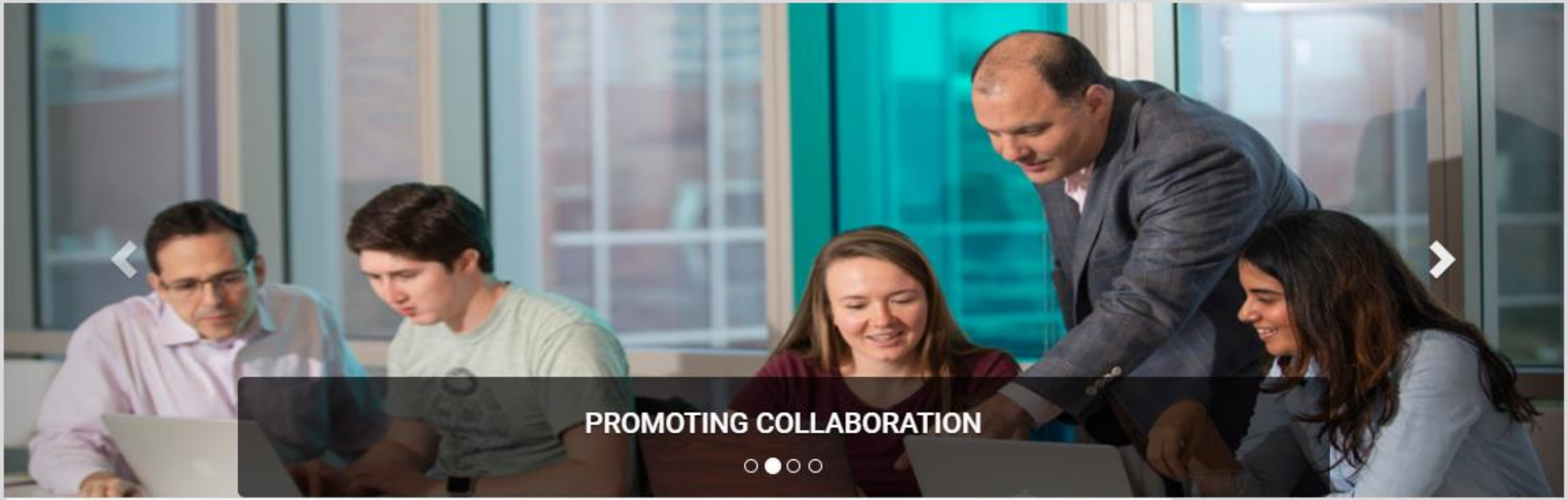
The Architect





The Visionary

Data Science Institute



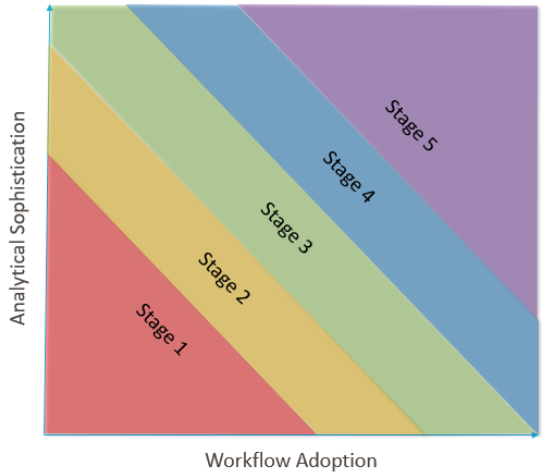
Thank you!

Jesse Spencer-Smith, Chief Data Scientist

jesse.spencer-smith@vanderbilt.edu

https://www.vanderbilt.edu/datascience/

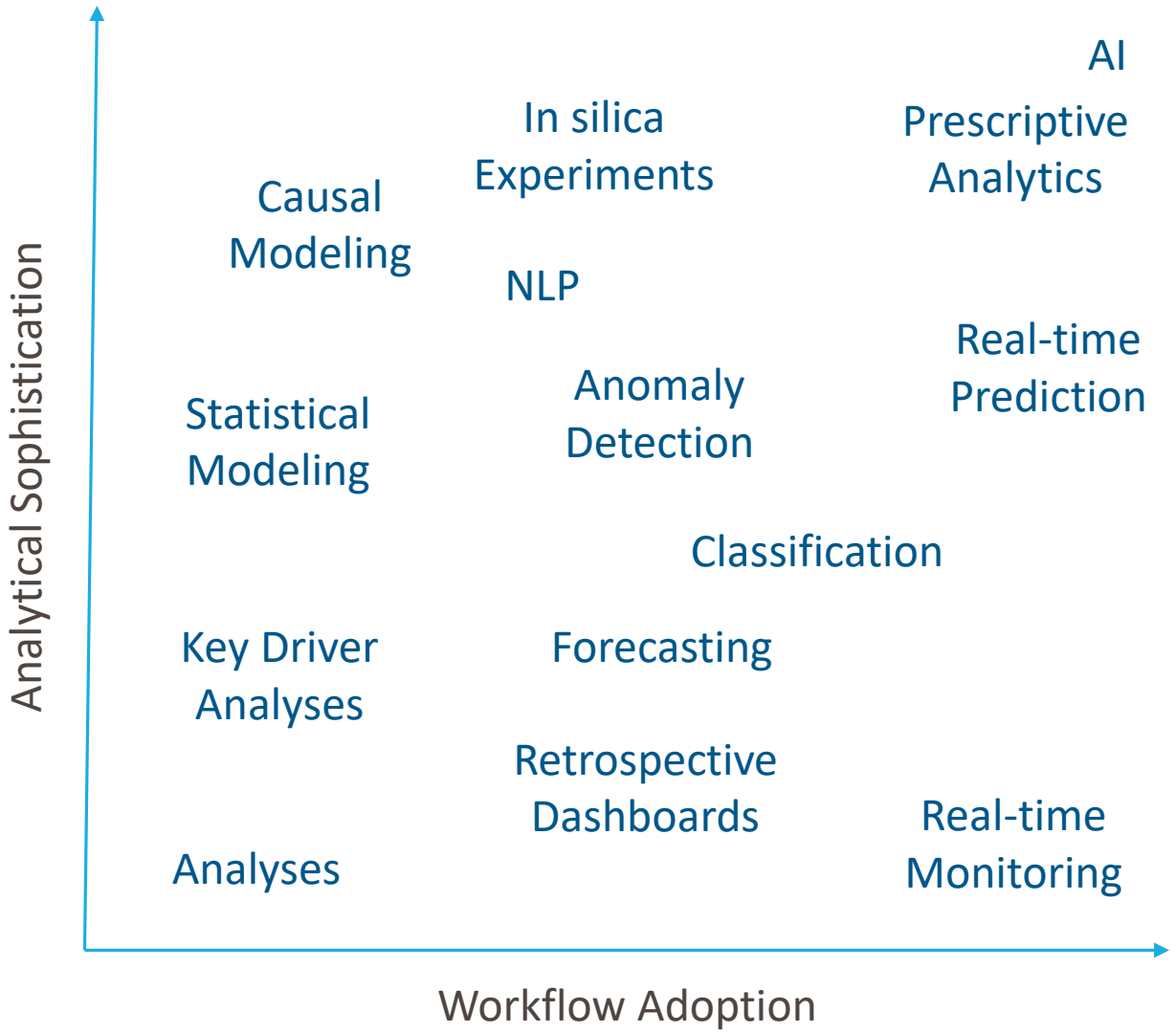
Twitter: @VUDataScience

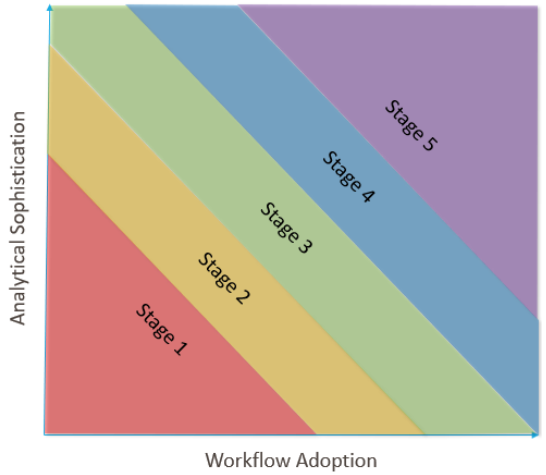


Special Case

For businesses differentiated by analysts and decision makers

Need for customized tools

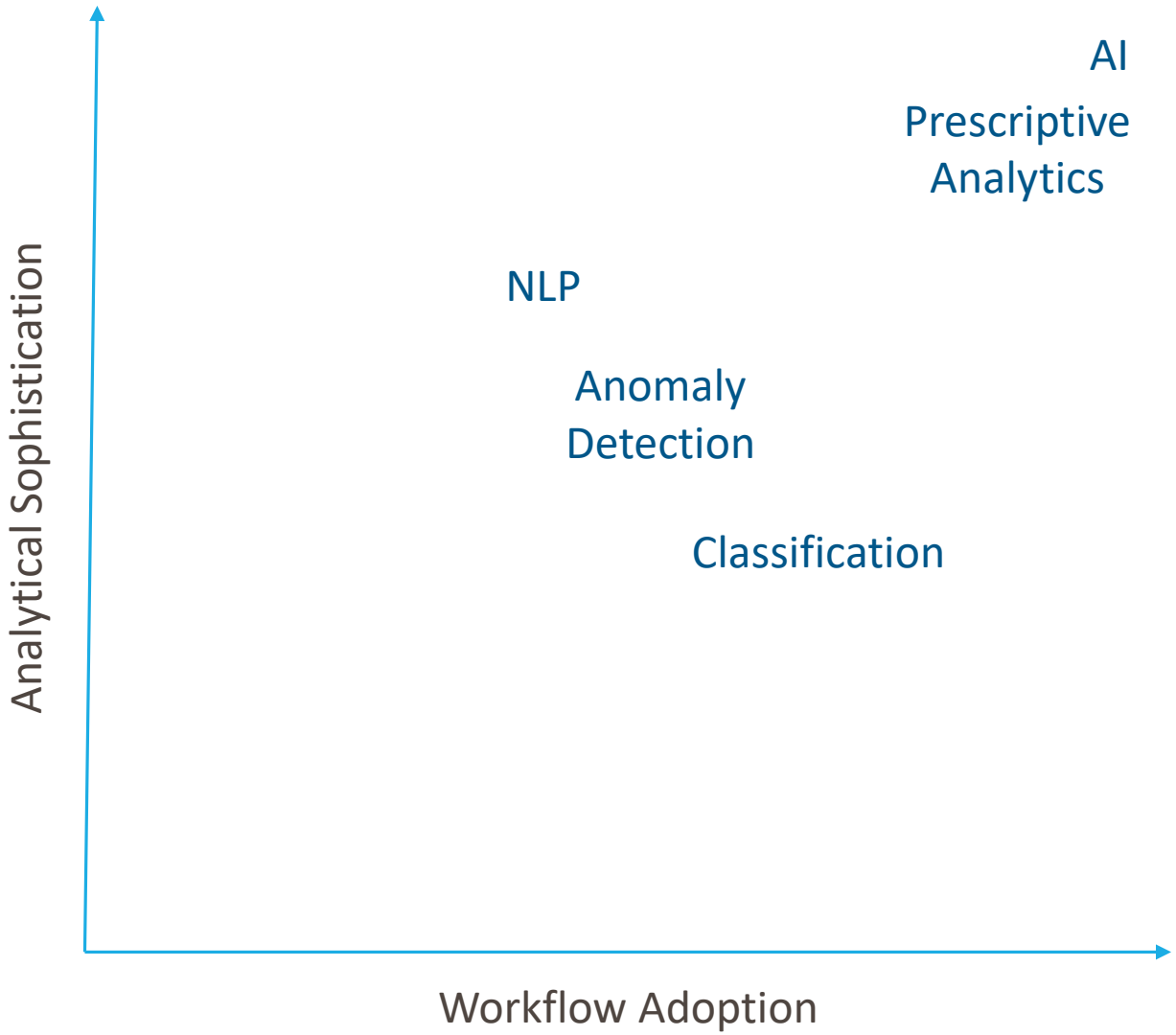


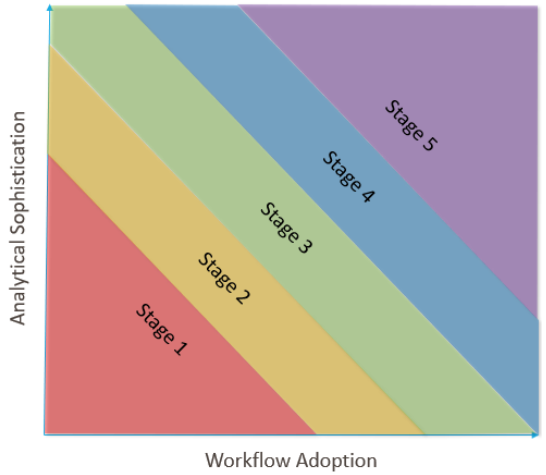


Special Case

For businesses differentiated by analysts and decision makers

Need for customized tools

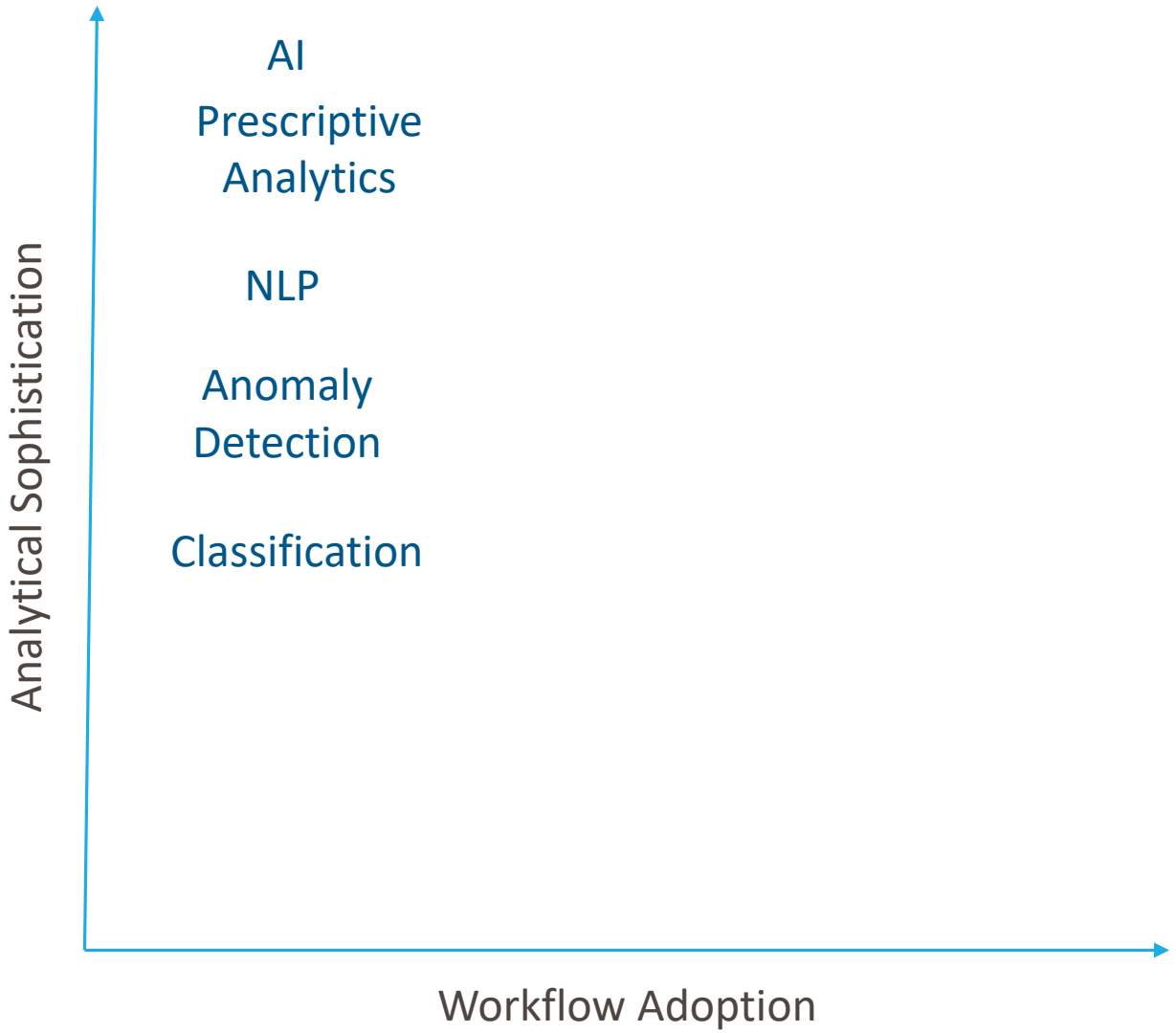


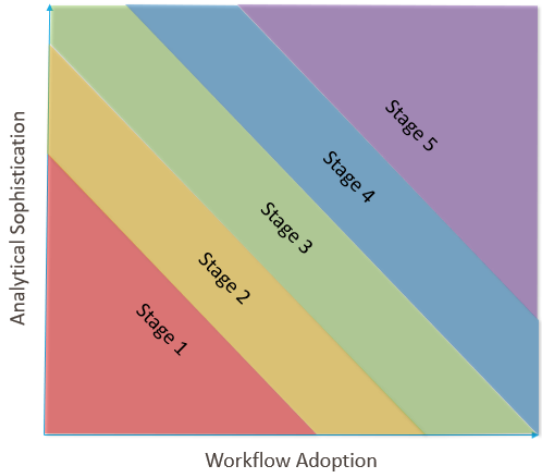


Special Case

For businesses differentiated by analysts and decision makers

Need for customized tools



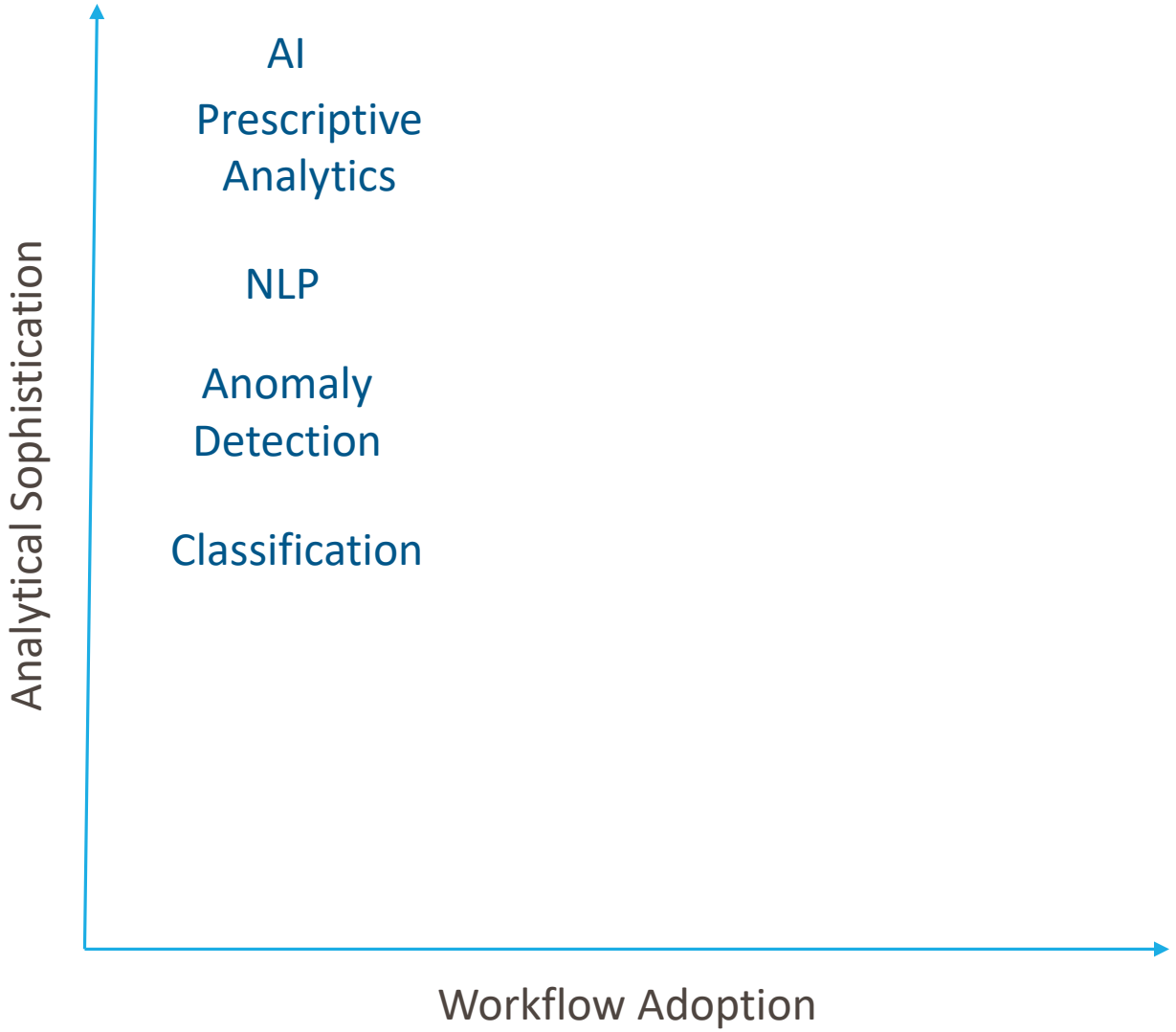


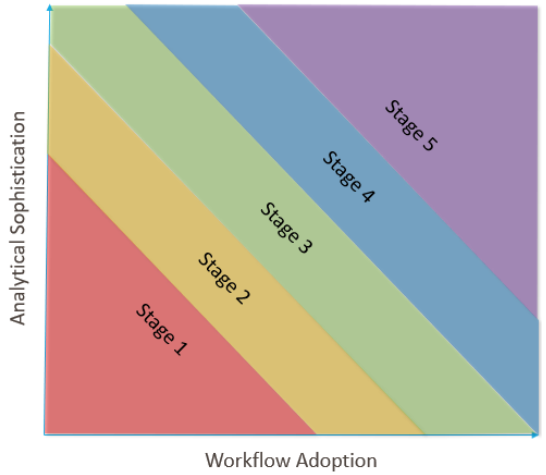
Special Case

For businesses differentiated by analysts and decision makers

Need for customized tools

Goal: Create 10x Analysts



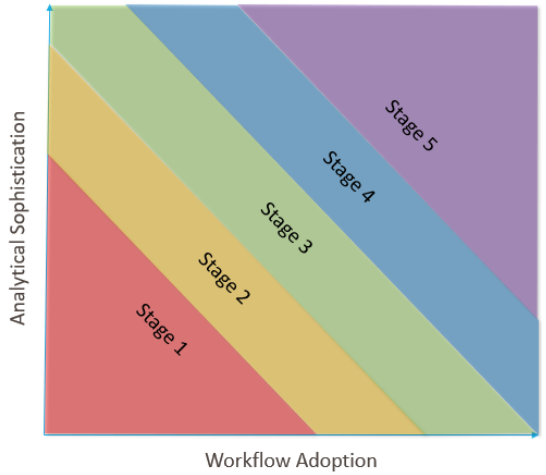


Special Case

For businesses differentiated by analysts and decision makers

Need for customized tools

Goal:
Create 10x Analysts

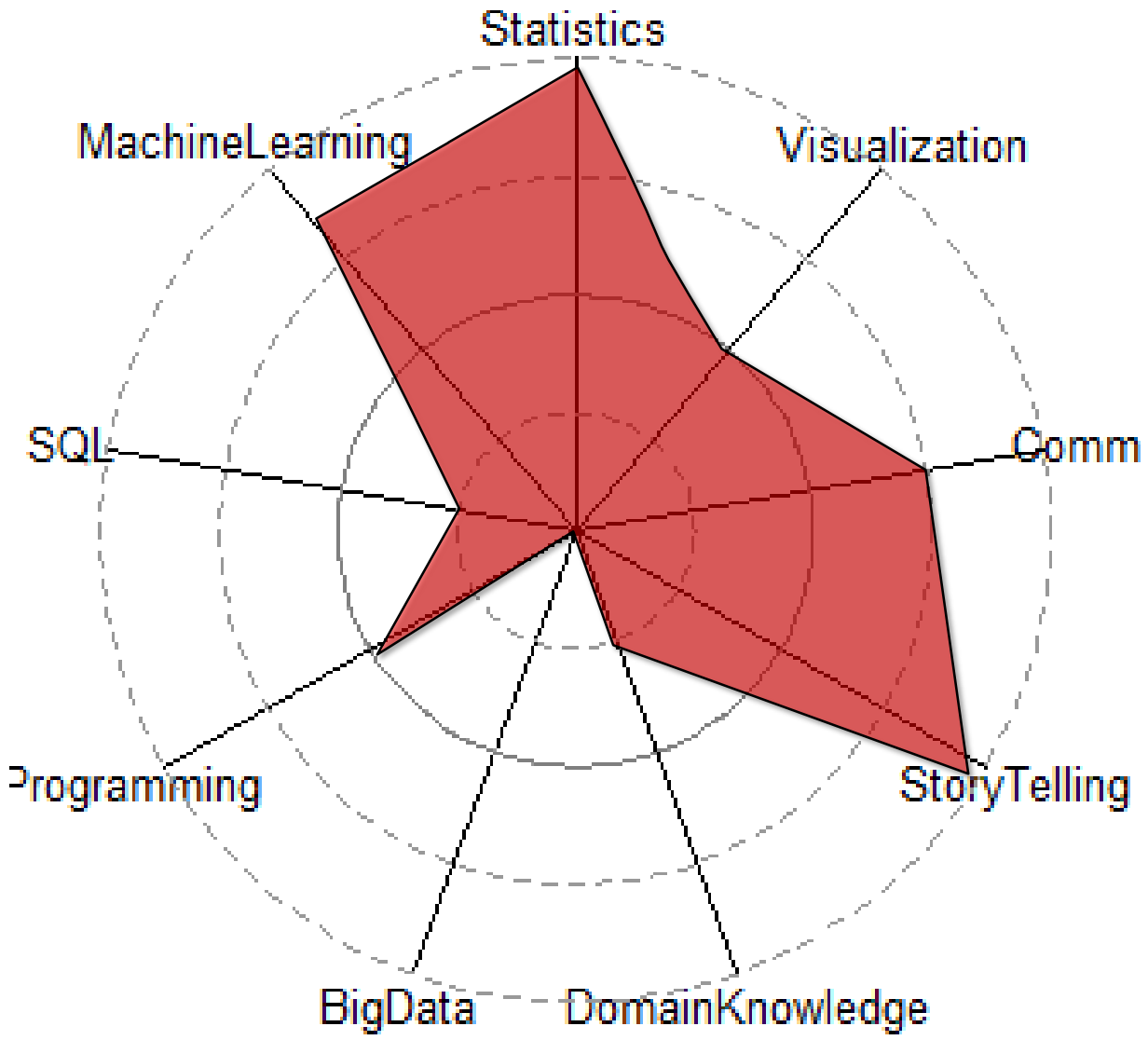


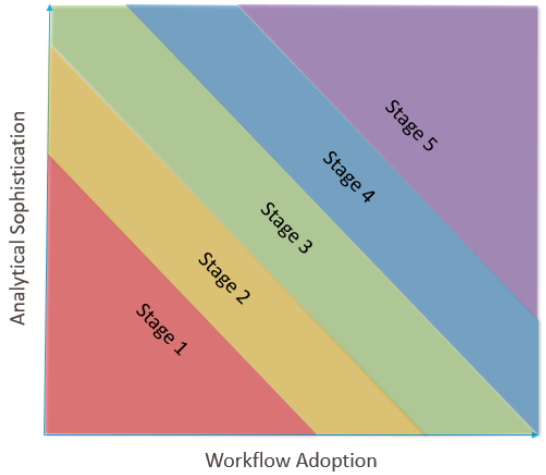
Special Case

For businesses differentiated by analysts and decision makers

Need for customized tools

Goal: Create 10x Analysts



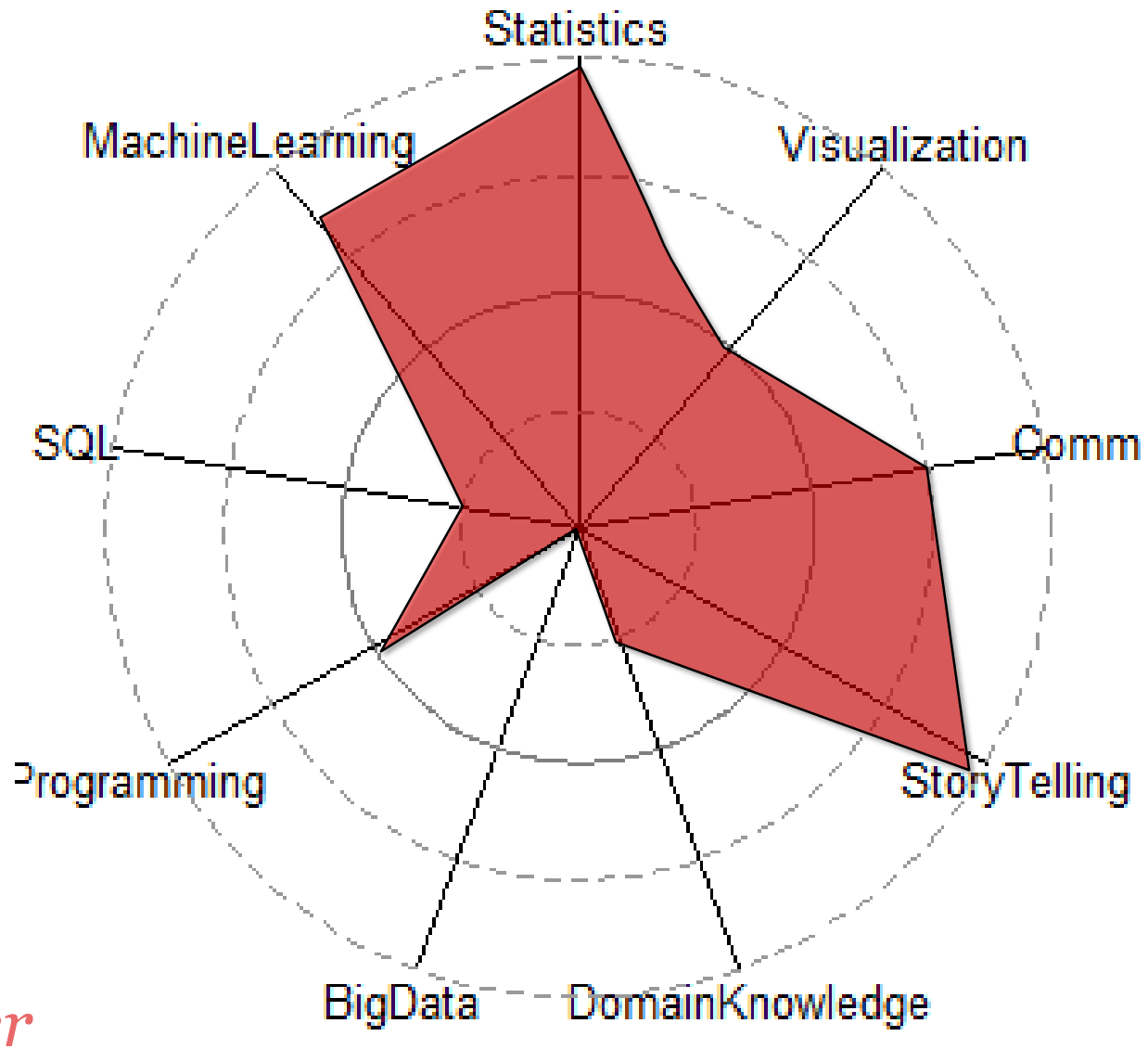


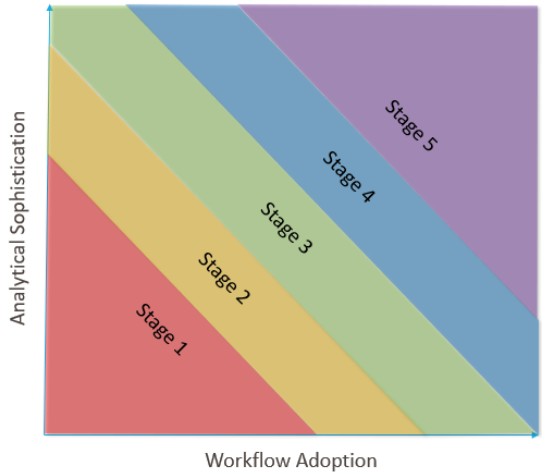
Special Case

For businesses differentiated by analysts and decision makers

Need for customized tools

Goal: Create 10x Analysts





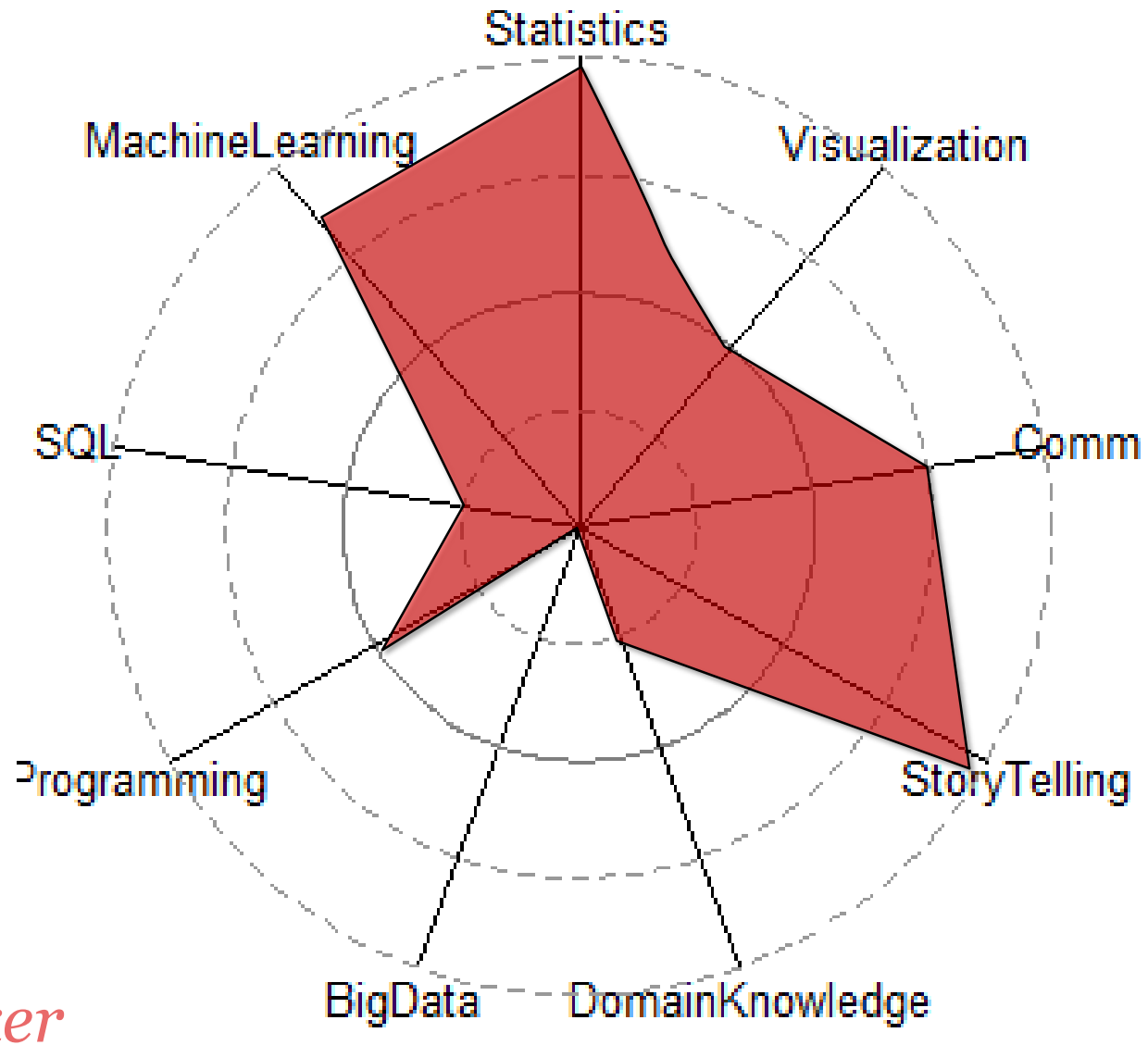
Special Case

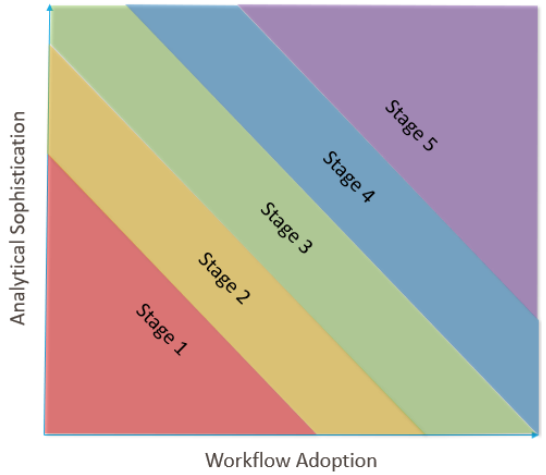
For businesses differentiated by analysts and decision makers

Need for customized tools

The Toolmaker

Goal: Create 10x Analysts





Special Case

For businesses differentiated by analysts and decision makers

Need for customized tools

The Wandmaker

Visionary, Builder, Guide:

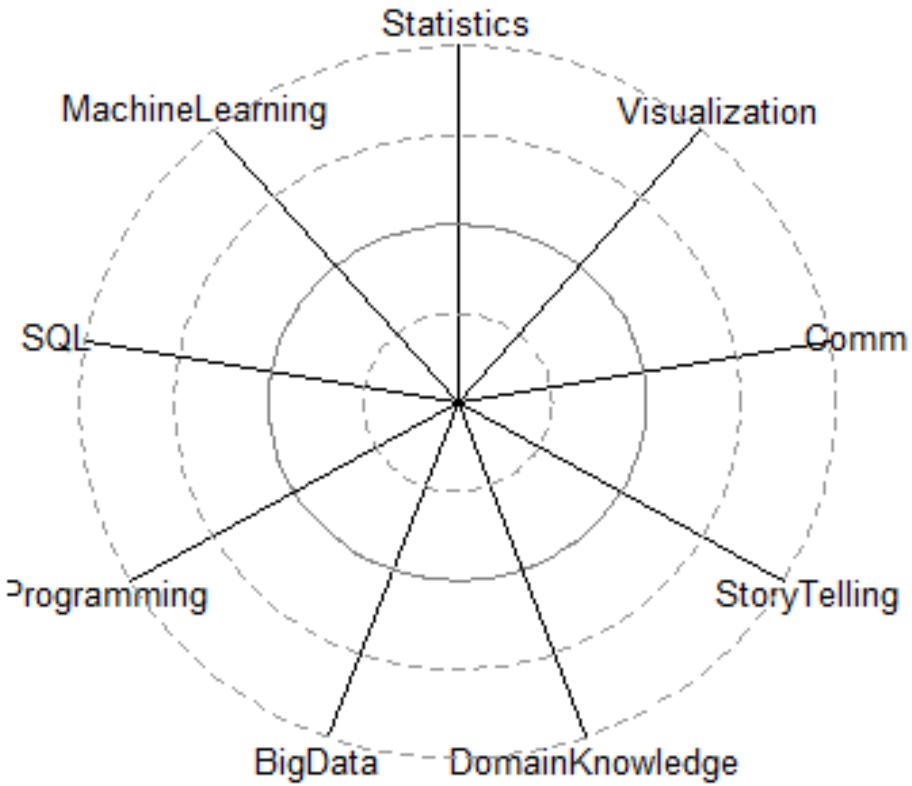
Finding the Right Data Scientist for Your Business

JESSE SPENCER-SMITH, PHD
CHIEF DATA SCIENTIST
MARCH 22, 2018

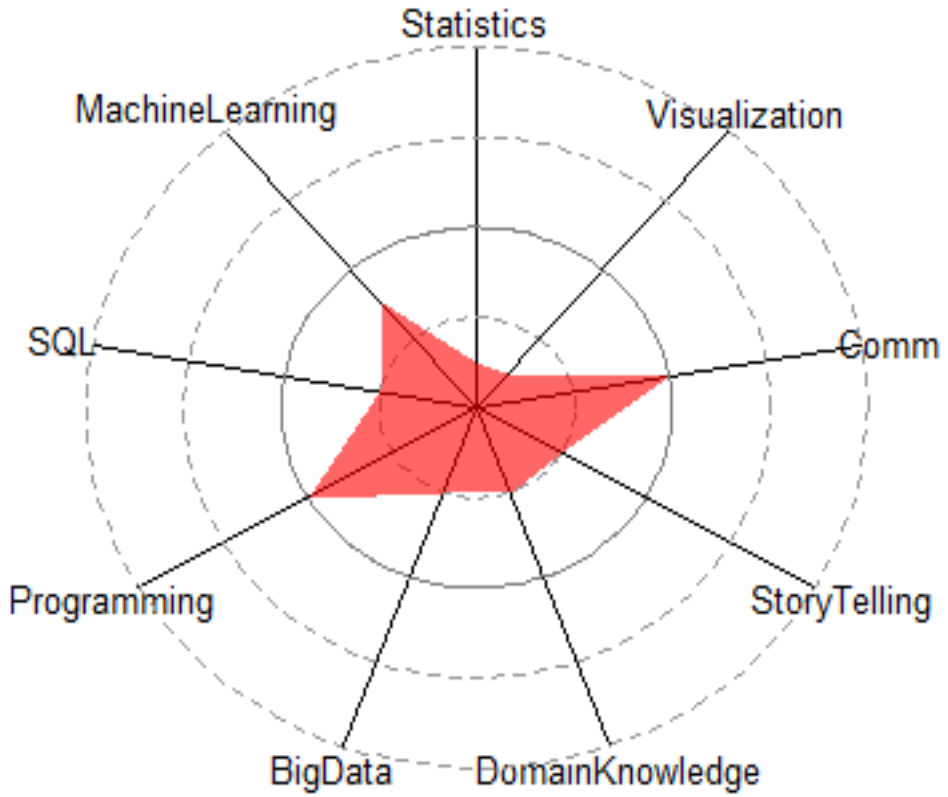
Data Science in Practice

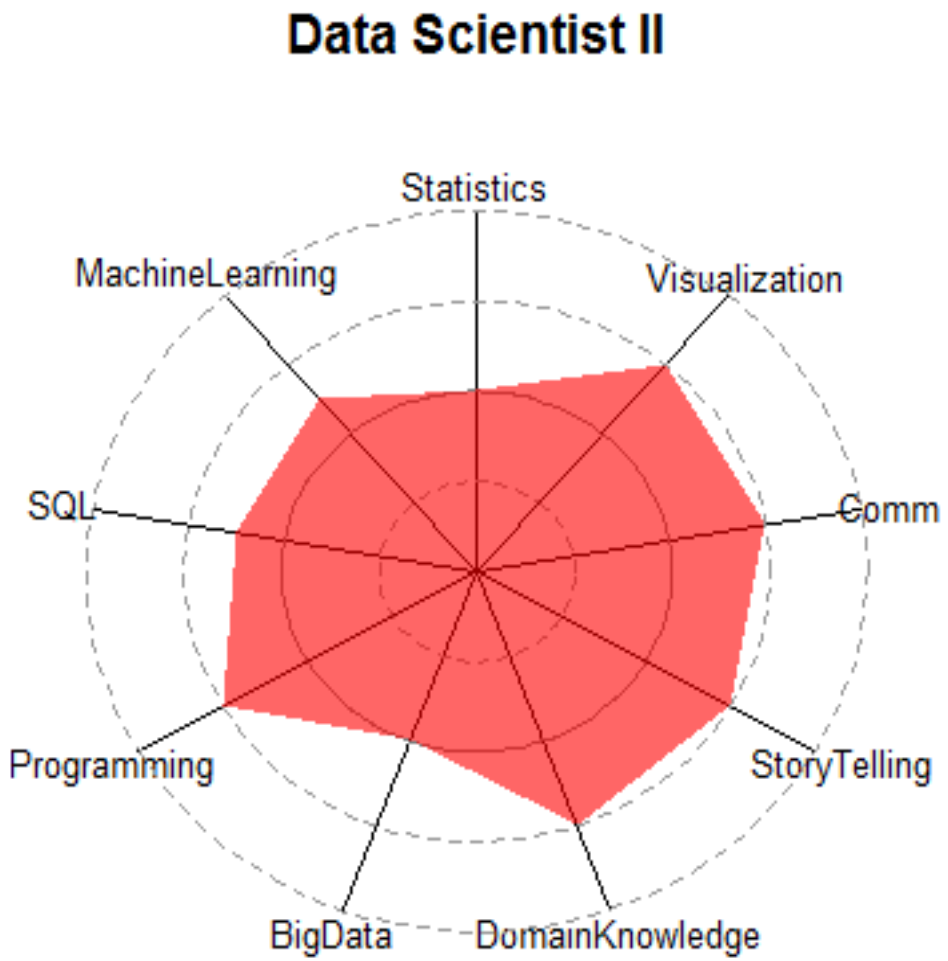
1. Organizational Data Science Maturity
2. Data Scientist Skills Inventory
3. Matching Your Skills to an Organization
4. Deep Solutions
5. Questions and Discussion

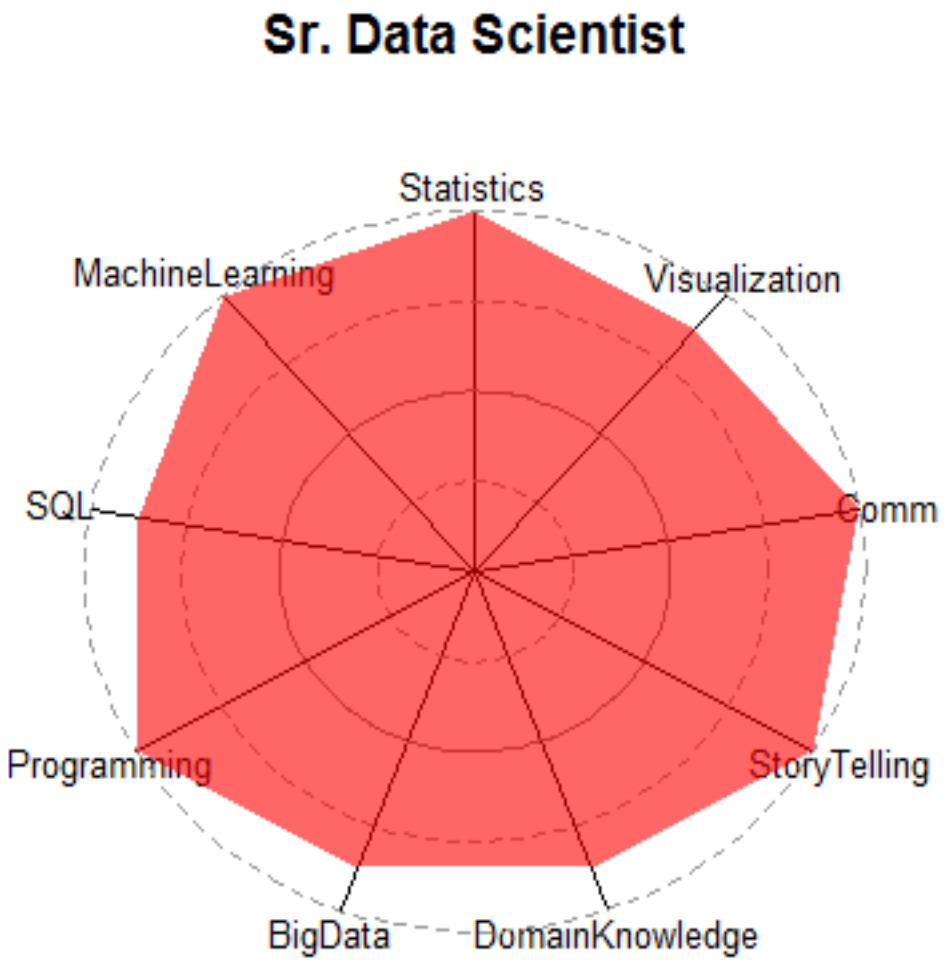
3 Skills Inventory



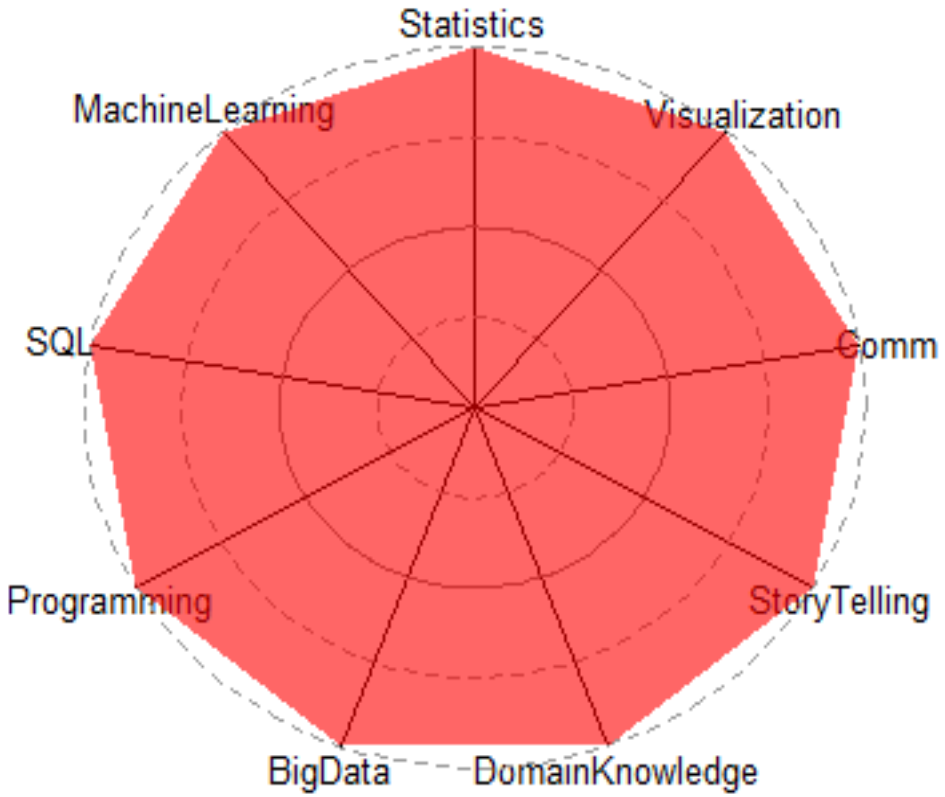
Data Scientist I




Data Scientist II
Statistics
Visualization
Comm
StoryTelling
DomainKnowledge
BigData
Programming
SQL
MachineLearning


Sr. Data Scientist
Statistics
Visualization
Comm
StoryTelling
DomainKnowledge
BigData
Programming
SQL
MachineLearning

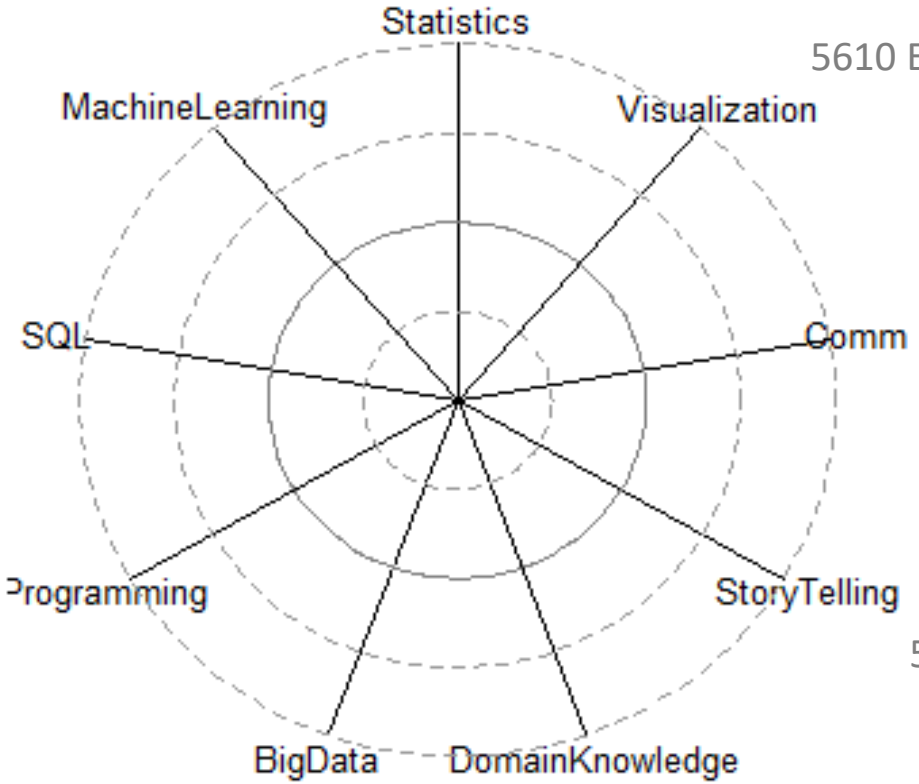
Consulting Data Scientist



5620 Probability & Statistical Inference

5640 Modeling & Machine Learning I
5660 Modeling & Machine Learning II

5610 Exploratory Data Analysis



5999 Capstone Development

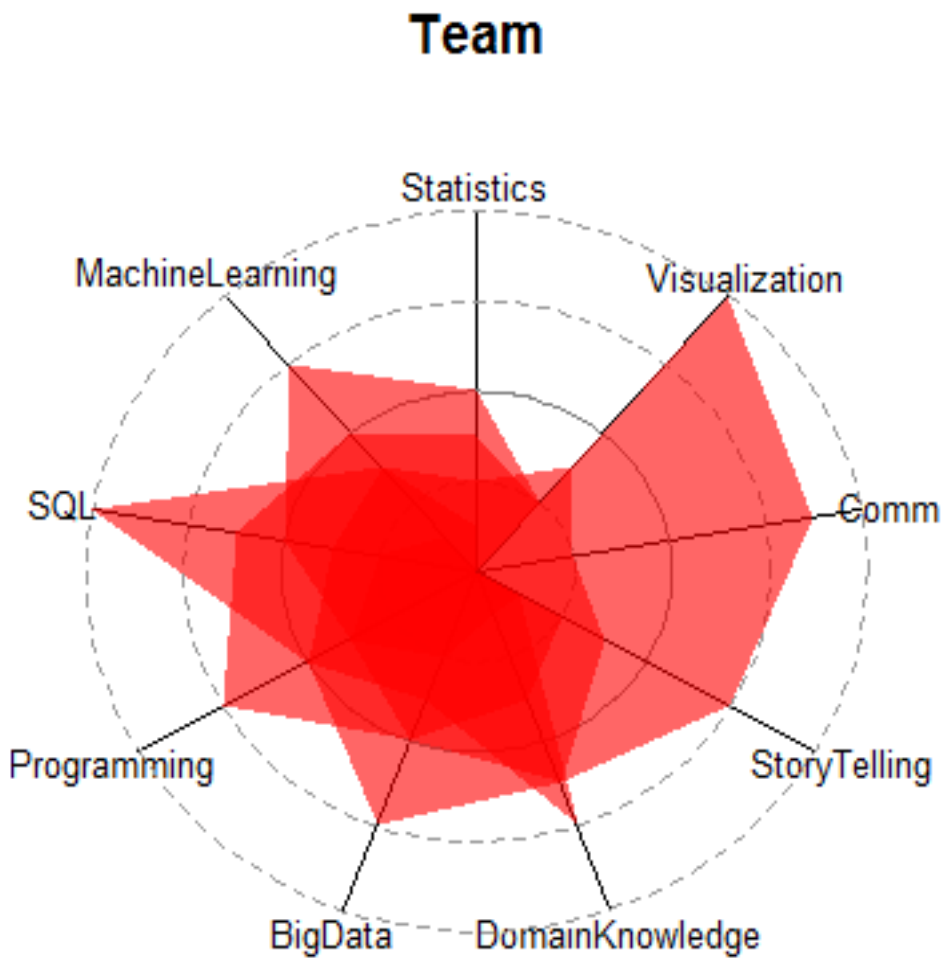
5420 Data Management Systems

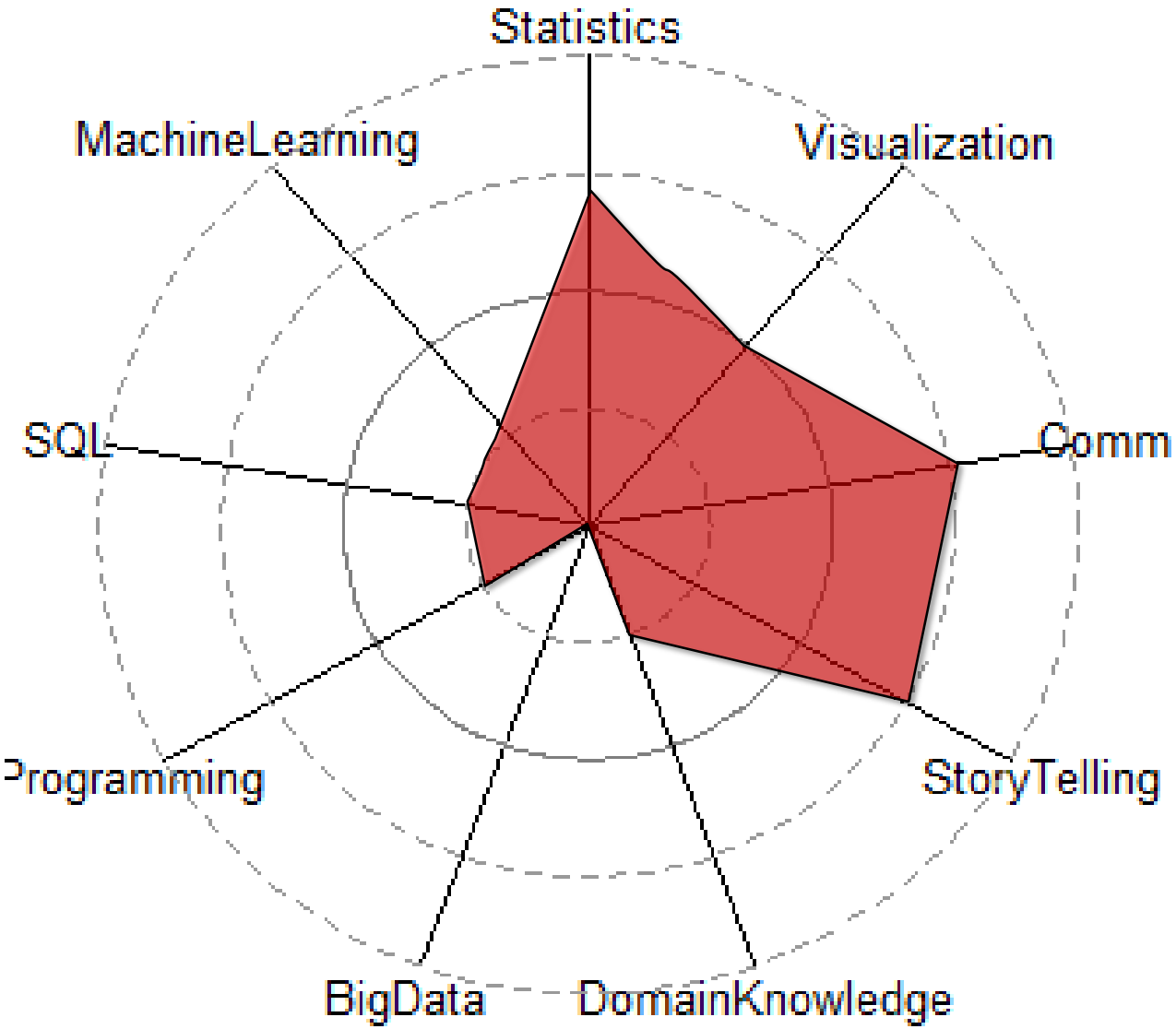
5220 Principles of Programming and Simulation
5440 Data Science Algorithms
5320 Survey of Data Science Applications

5360 Case Studies in Data Science

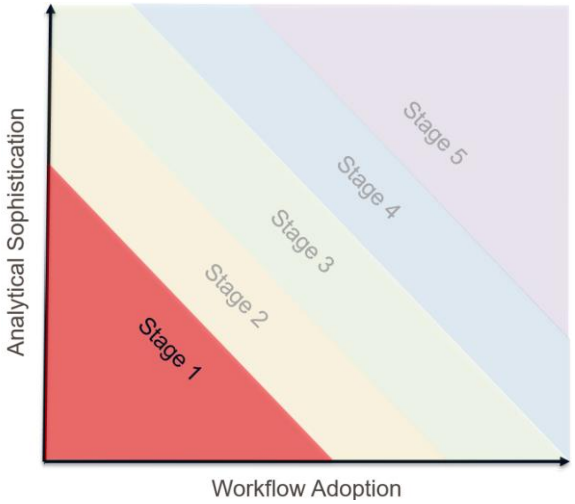
5460 Big Data & Scaling

5380 Data Science in Teams I
5384 Data Science in Teams II
5340 Data Science Rights & Responsibilities

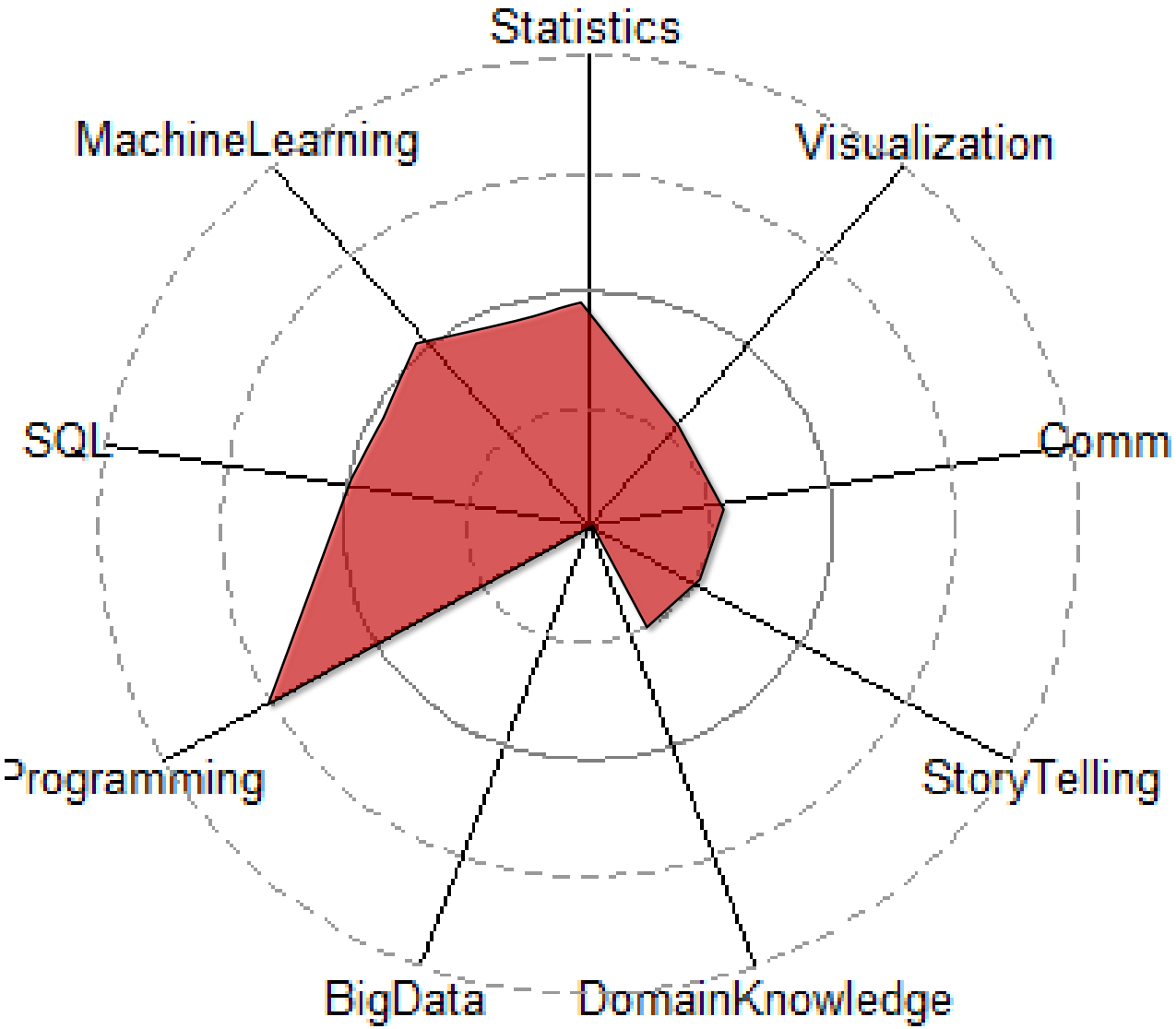

Team
Statistics
Visualization
Comm
StoryTelling
DomainKnowledge
BigData
Programming
SQL
MachineLearning



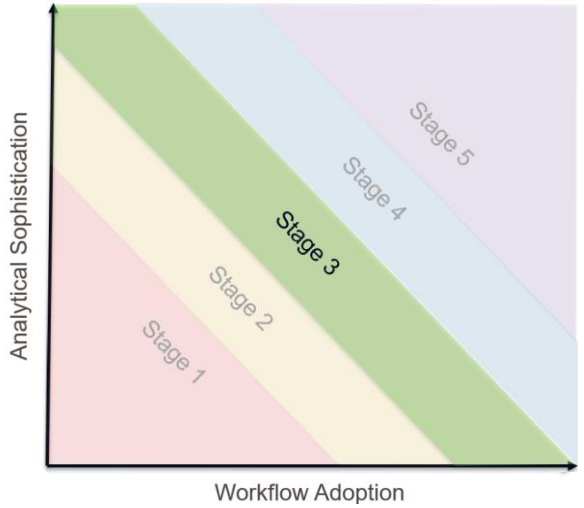
4 Needs based on Maturity



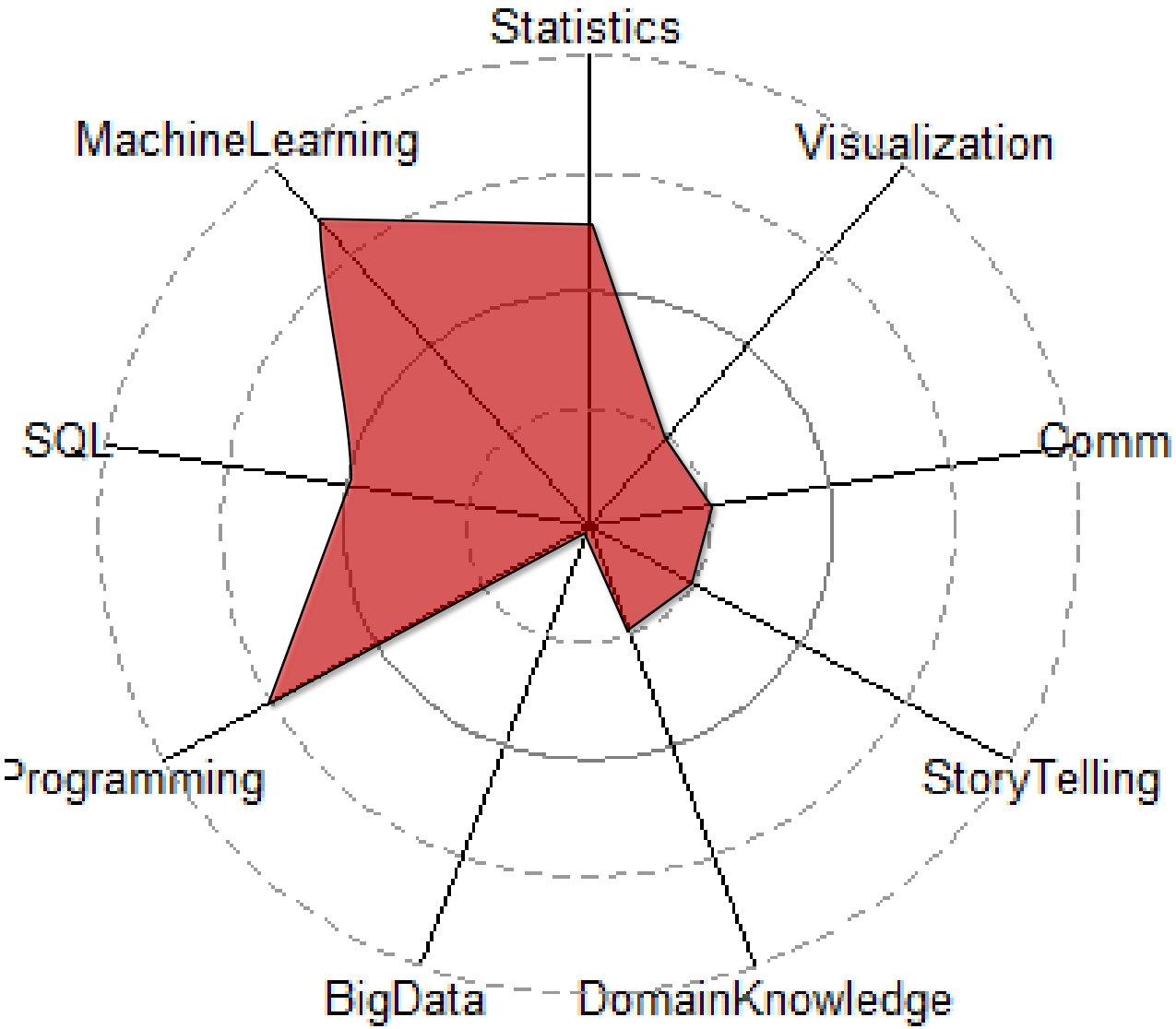
The Guide



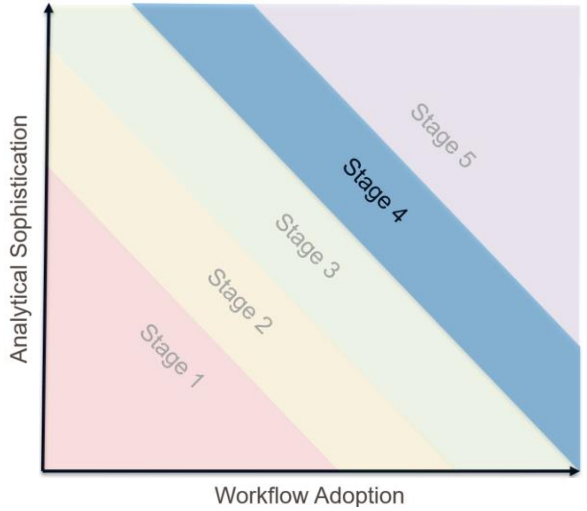
4 Needs based on Maturity



The Builder

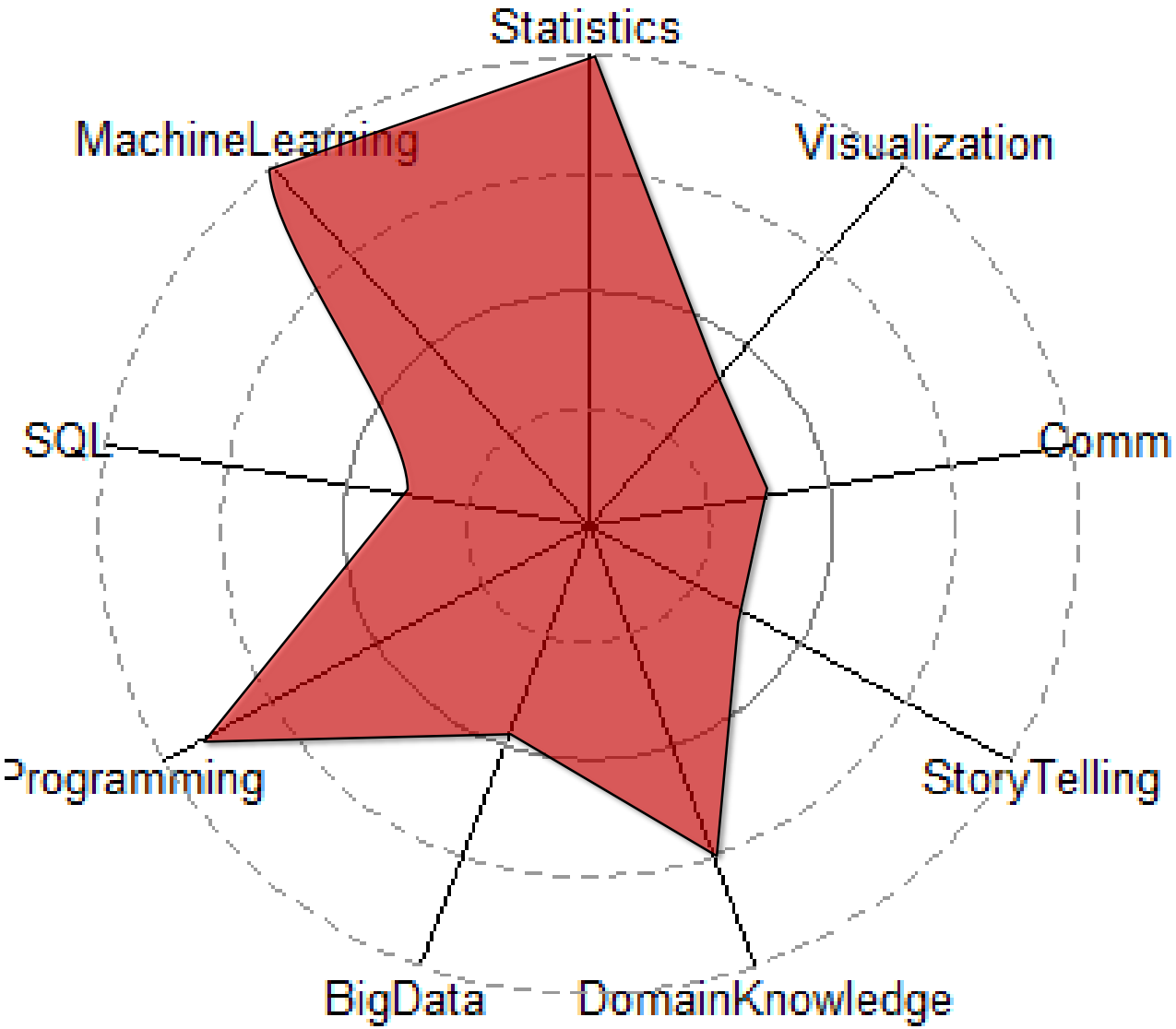


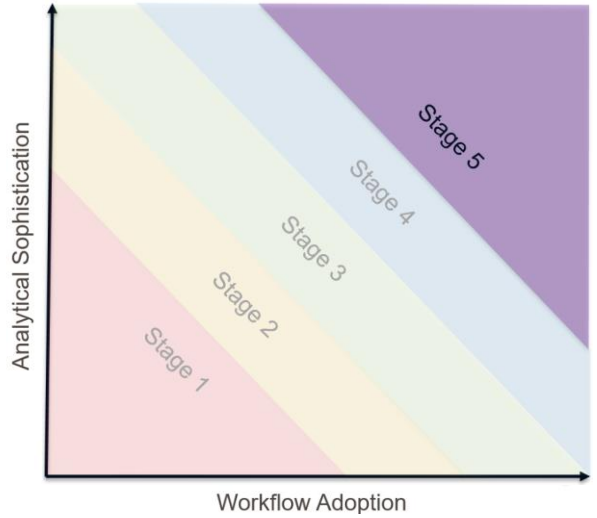
4 Needs based on Maturity



The Architect

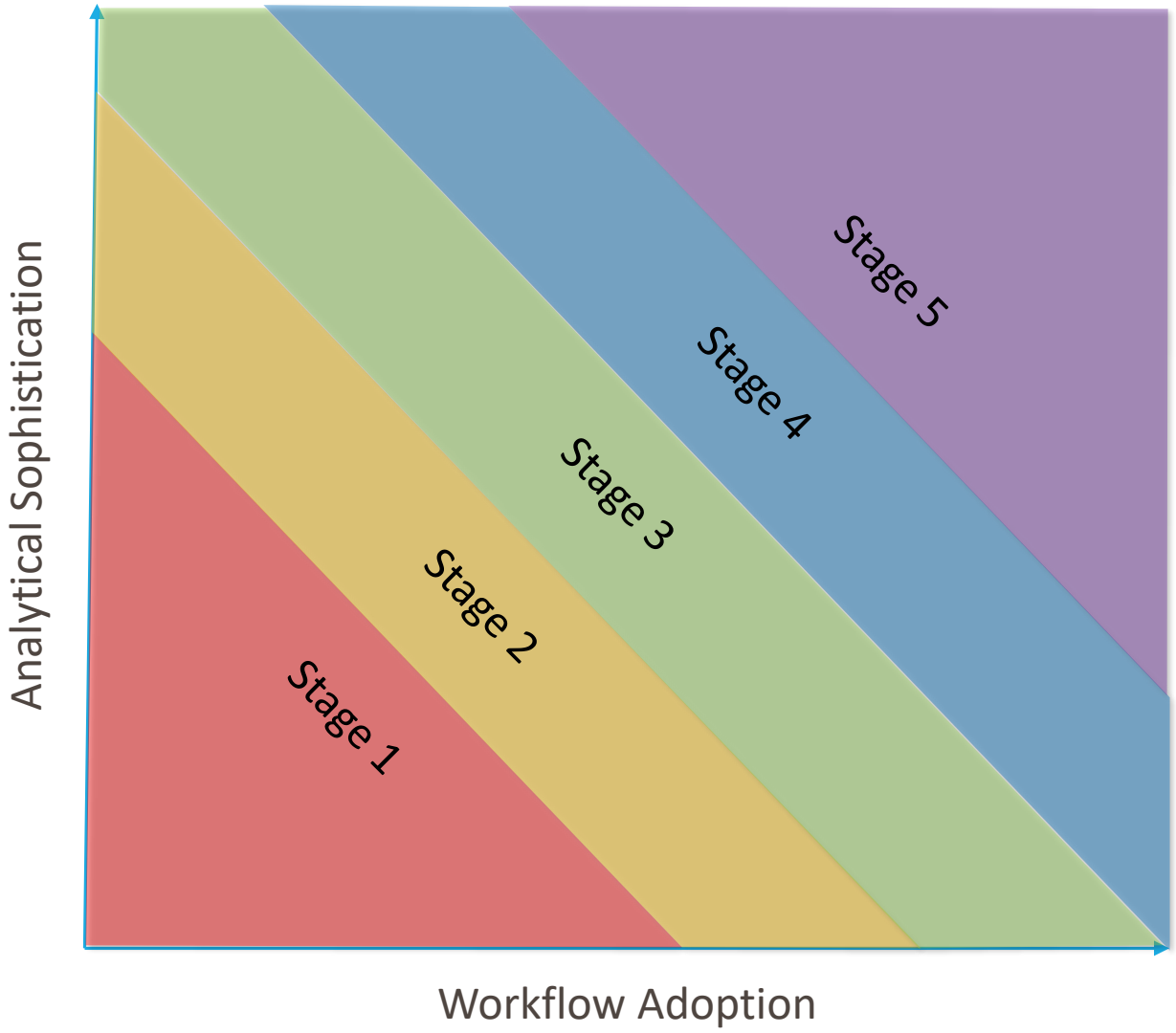
4 Needs based on Maturity



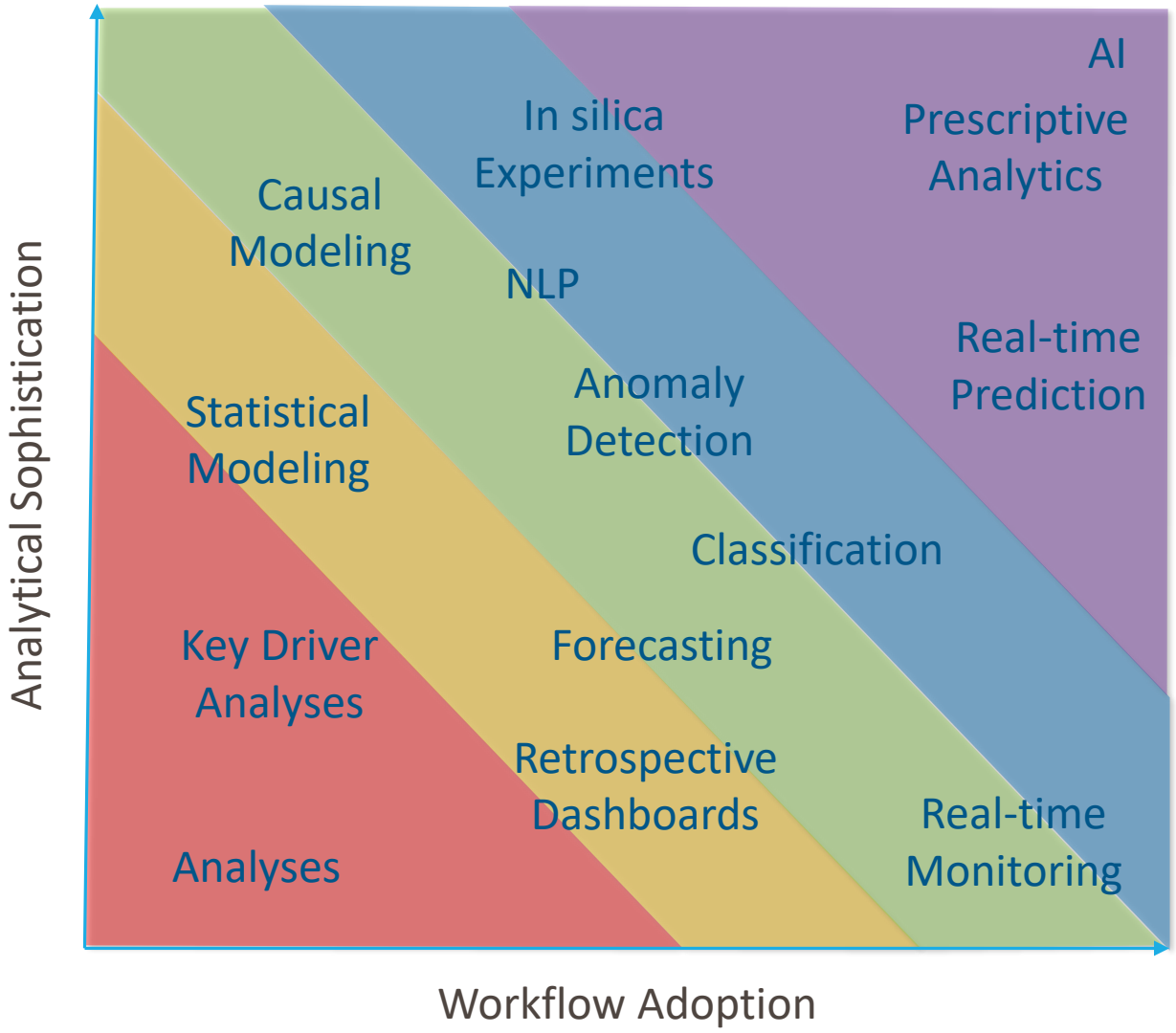


The Visionary

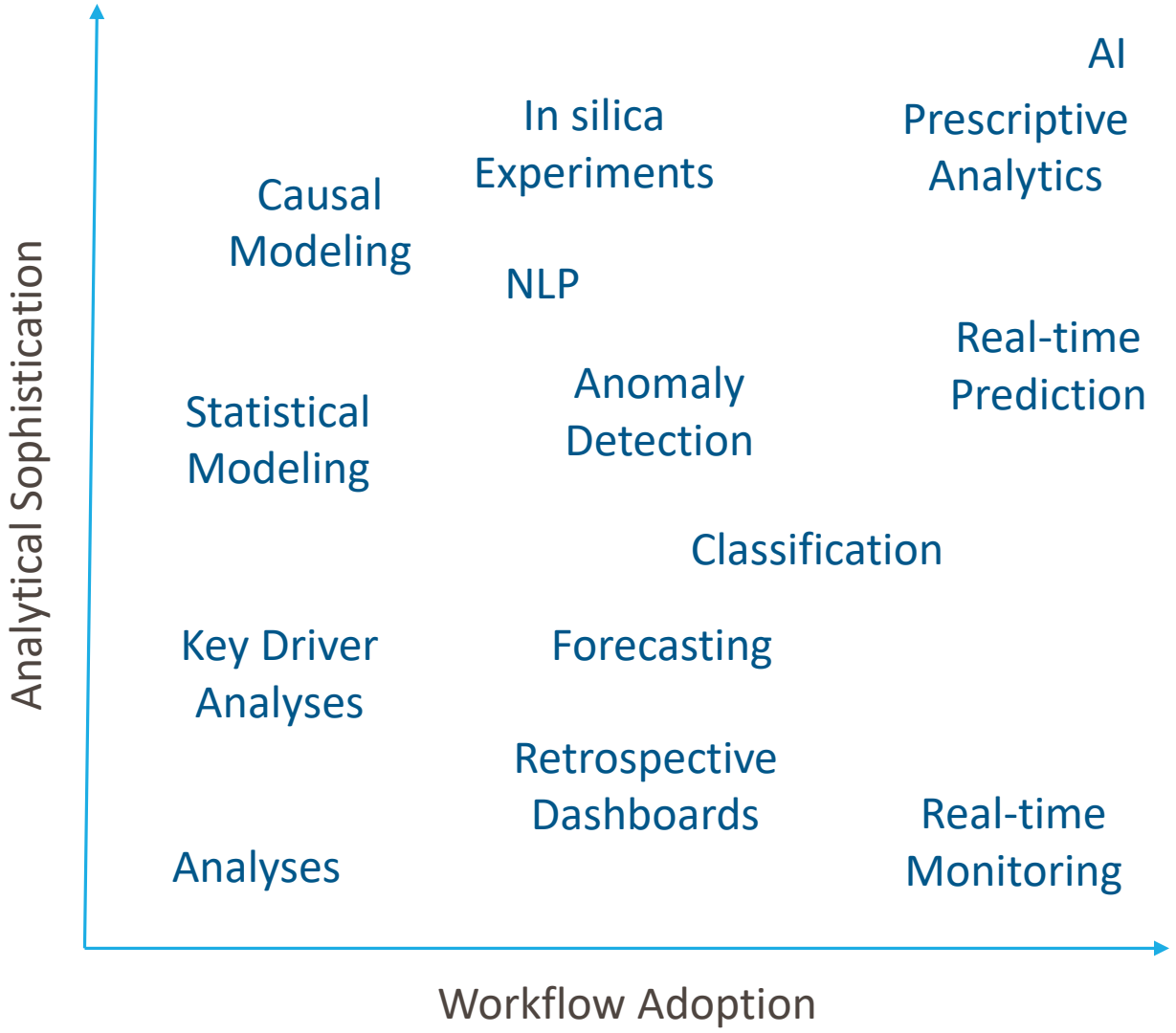
Accelerating Time to Value



Accelerating Time to Value



Accelerating Time to Value

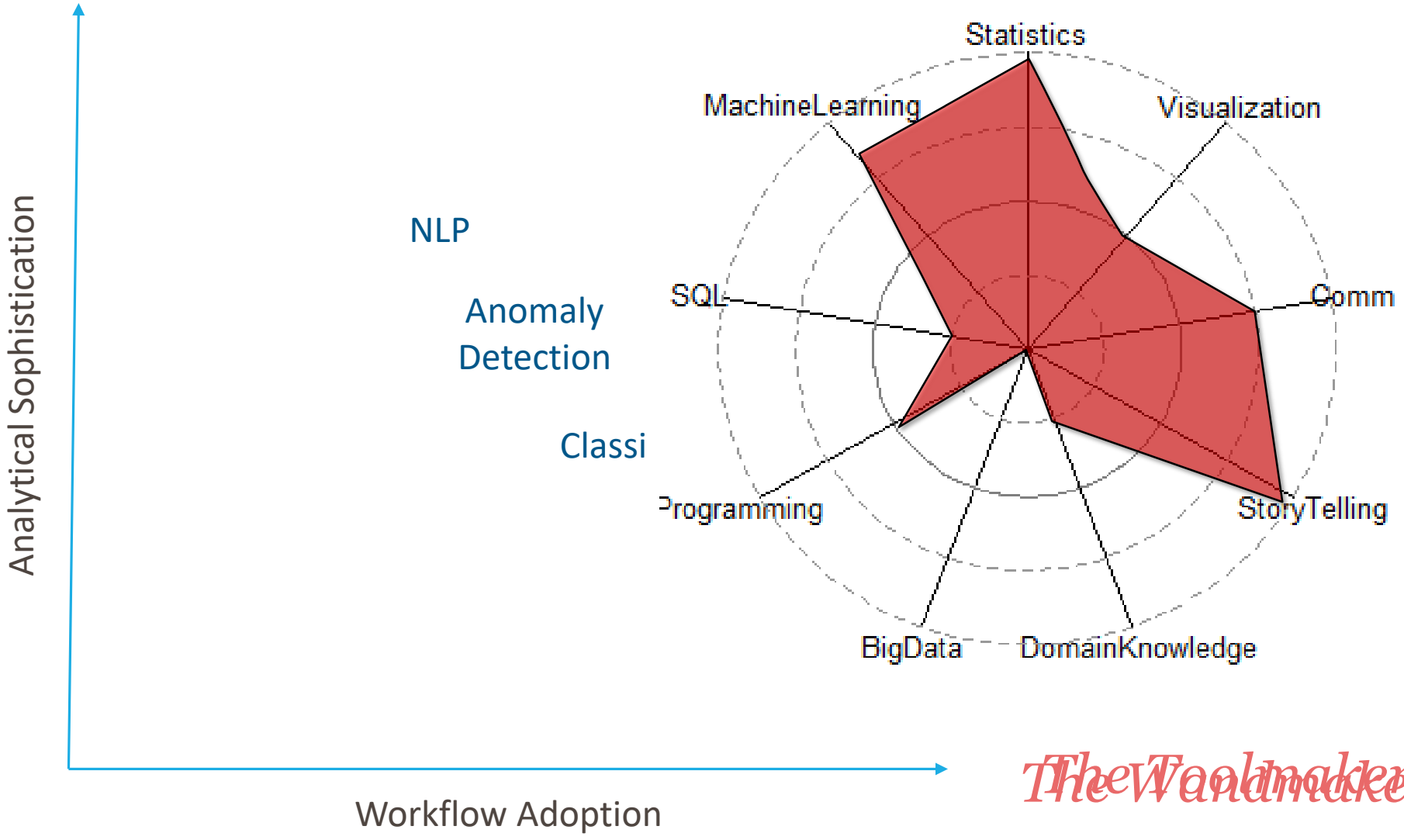


Accelerating Time to Value

Deep Solutions

Tools for Analysts and Decision makers

Goal:
Create 10x Analysts



TheToolmaker

Thank you!

Abstract:

Guide. Builder. Architect. Visionary.

Data scientists come in all types. An organization's data science needs change dramatically as they move through stages of data science maturity–their ability to consume, adopt, and deploy advanced analytics solutions. Understanding the maturity stage of your organization will help you choose projects that can bring value from your data sooner, grow your ability to derive greater value in the future, and help you make good decisions when building your analytics team. A data scientist might be a journeyman model builder, or a data scientist consultant, or a software engineer, or a developer of new deep learning algorithms. The data scientist that would be successful in a mature organization may well fail in an organization new to data science. Hiring and growing data scientists based on skill sets in line with your data science maturity stage maximizes your probability of success, and can prepare you for the next step in your evolution. We'll discuss a framework to determine your level of data science readiness, explore a tool to assess the skill sets of data scientists, and find which skills can maximize your organization's probability of success at each stage.