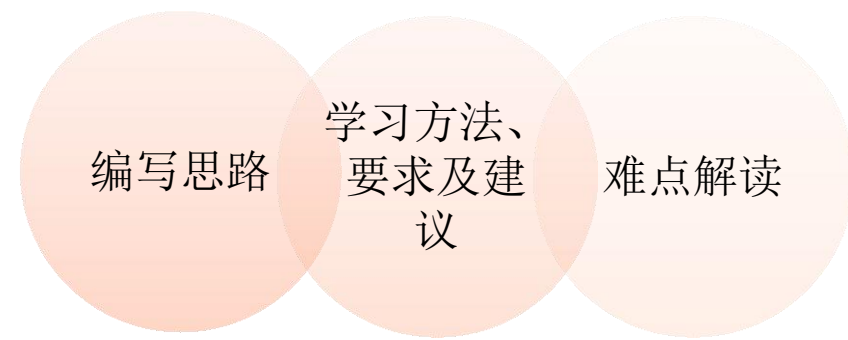


# 第一章 数据科学的基础理论



# 1.本章定位与内容简介



1.1为什么要学习数据科学

1.2数据科学的定义

1.3数据科学的知识体系

1.4 数据科学的基本流程

1.5数据科学与其他学科的区别

1.6 数据科学的人才类型

1.7数据科学的常用工具

1.8 数据科学的相关应用

1.9 继续学习本章知识

习题

## 2.本章学习提示及要求

### 了解

- 大数据时代的本质及学习数据科学的意义
- 数据科学的常用工具
- 数据科学的相关应用

### 理解

- 数据科学的人才类型及其主要职责
- 数据科学与其他学科的区别与联系

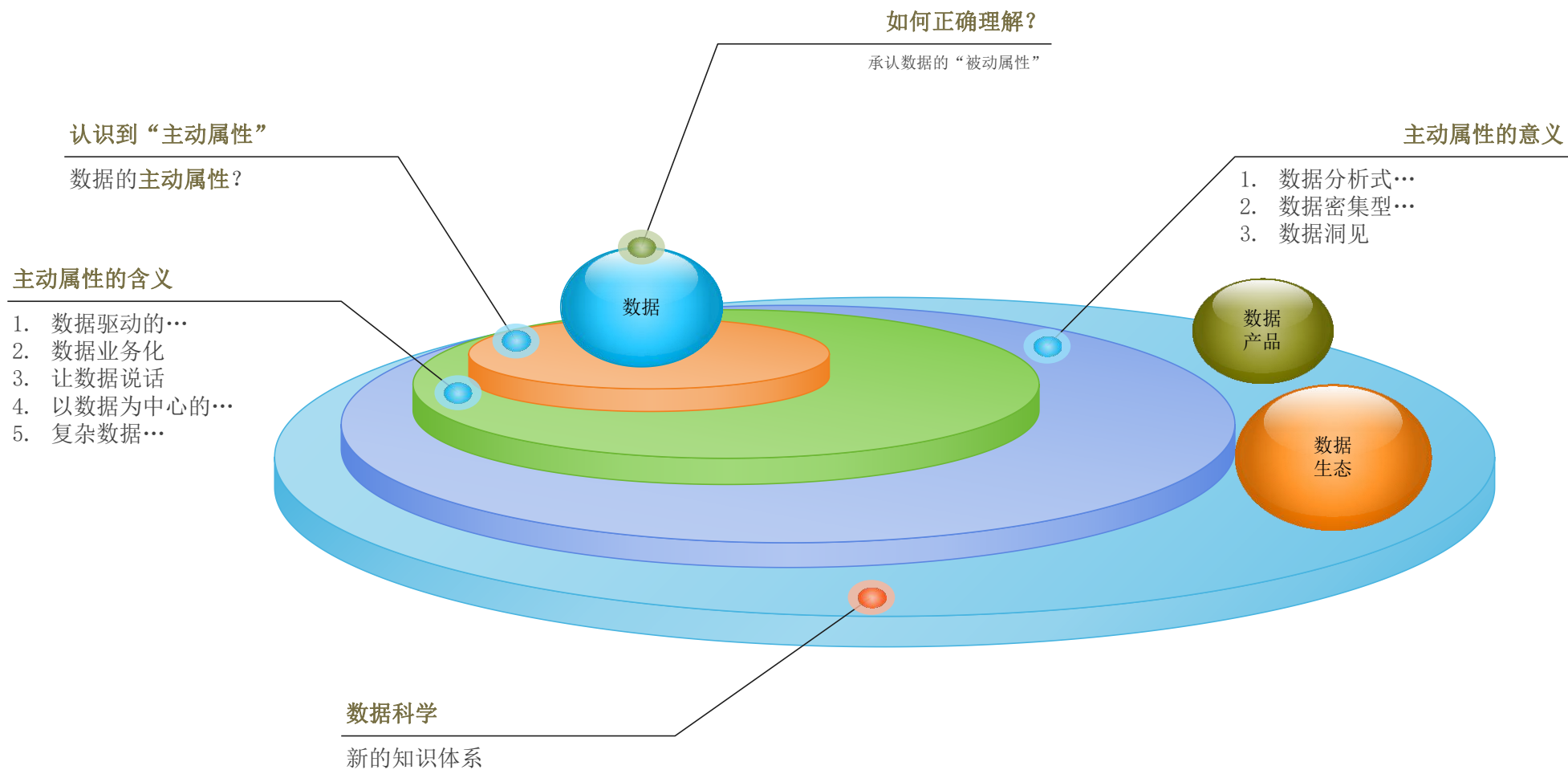
### 掌握

- 大数据和数据科学的定义与特征
- 数据科学的知识体系

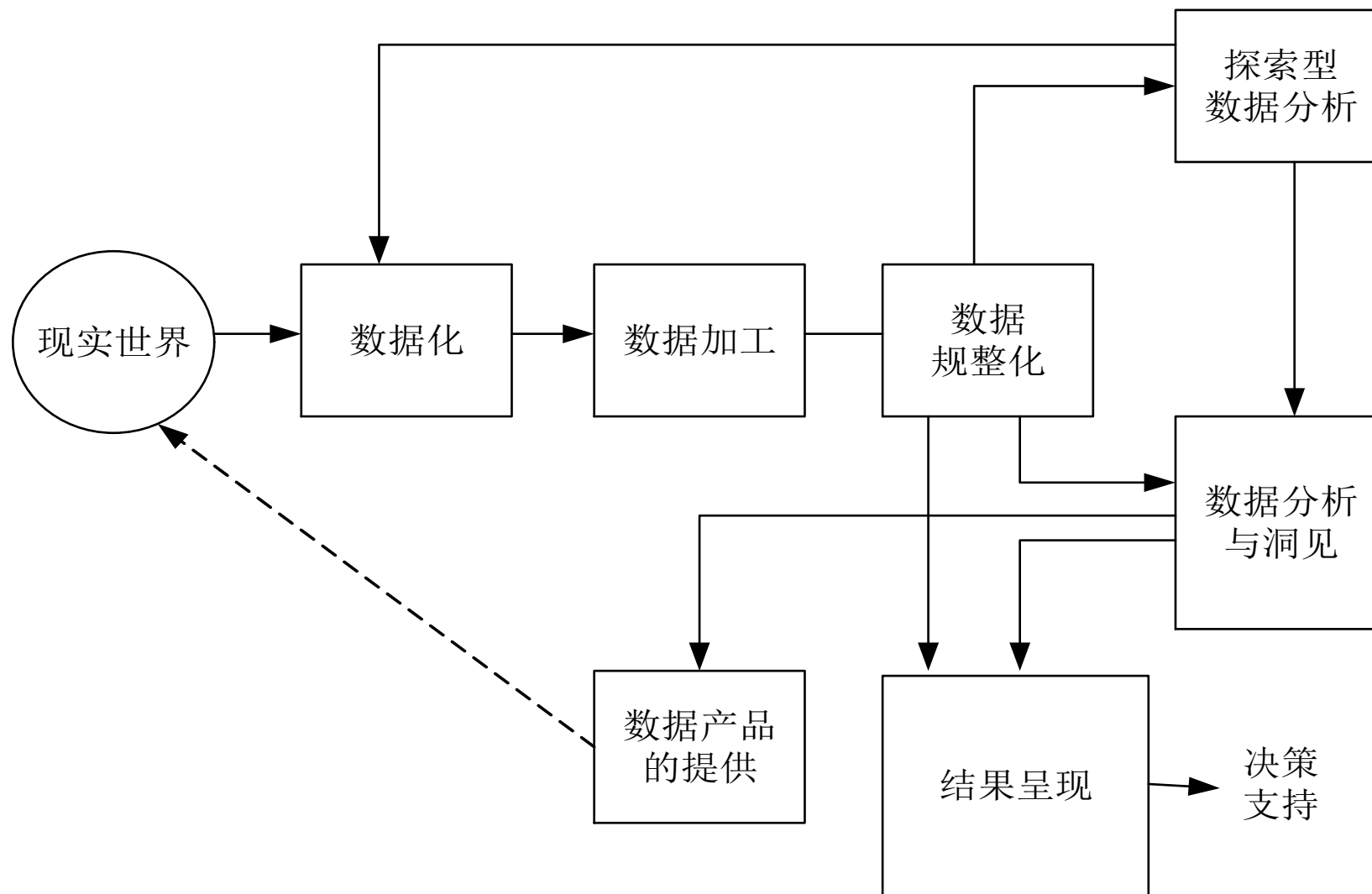
### 熟练掌握

- 数据科学的基本流程（图1-5）

### 3.数据科学的主要特征



## 4. 数据科学的基本流程



## 5.数据科学的三个要素

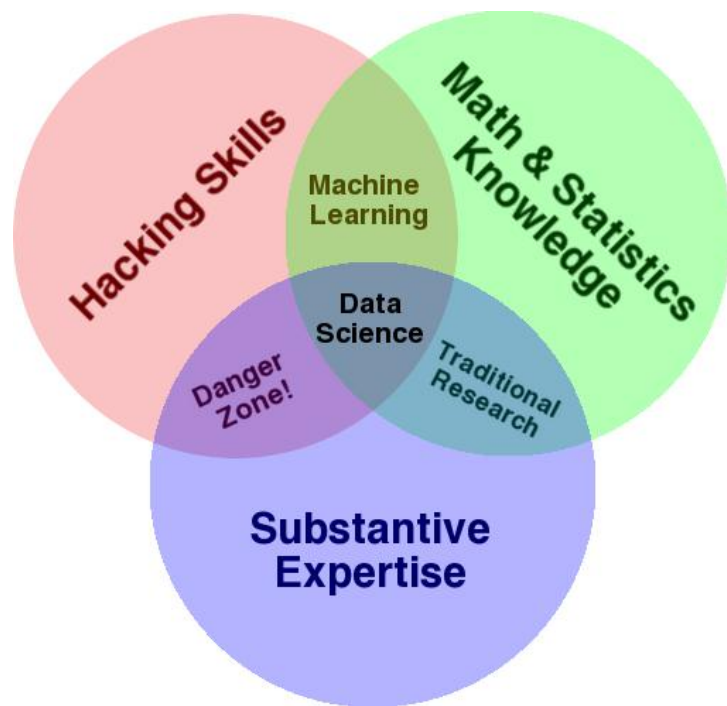


图1 Drew Conway的数据科学韦恩图（2010）

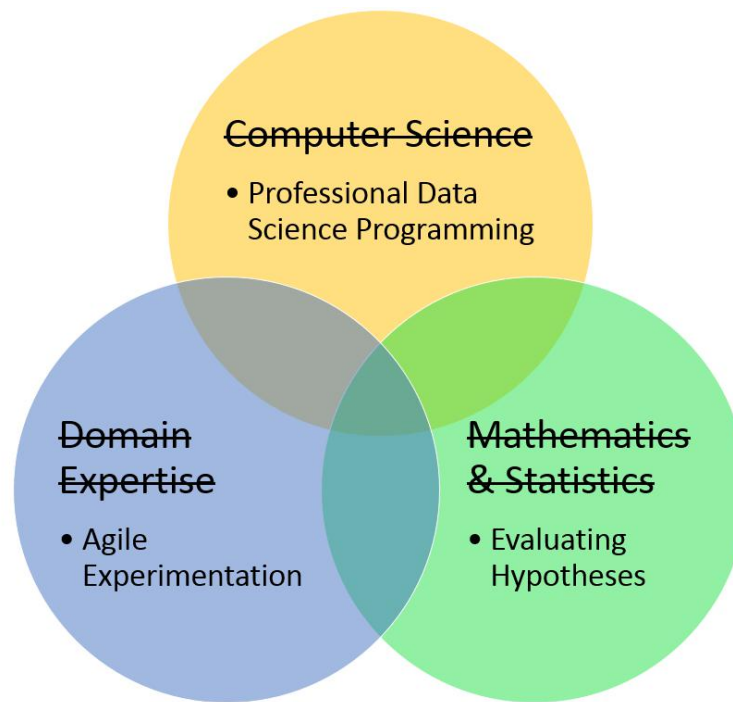


图2 Jerry Overton的数据科学韦恩图（2016）

来源:

图1: Schutt R, O'Neil C. Doing data science: Straight talk from the frontline[M]. O'Reilly Media, Inc., 2013:7.

图2: Jerry Overton.Going Pro in Data Science [M].O'Reilly Media, Inc,2016:12.

## 6.数据科学的基本原则

三世界原则

三要素原则

数据驱动原则

数据复杂性  
原则

数据资产原则

DIKUW原则

3C原则

协同原则

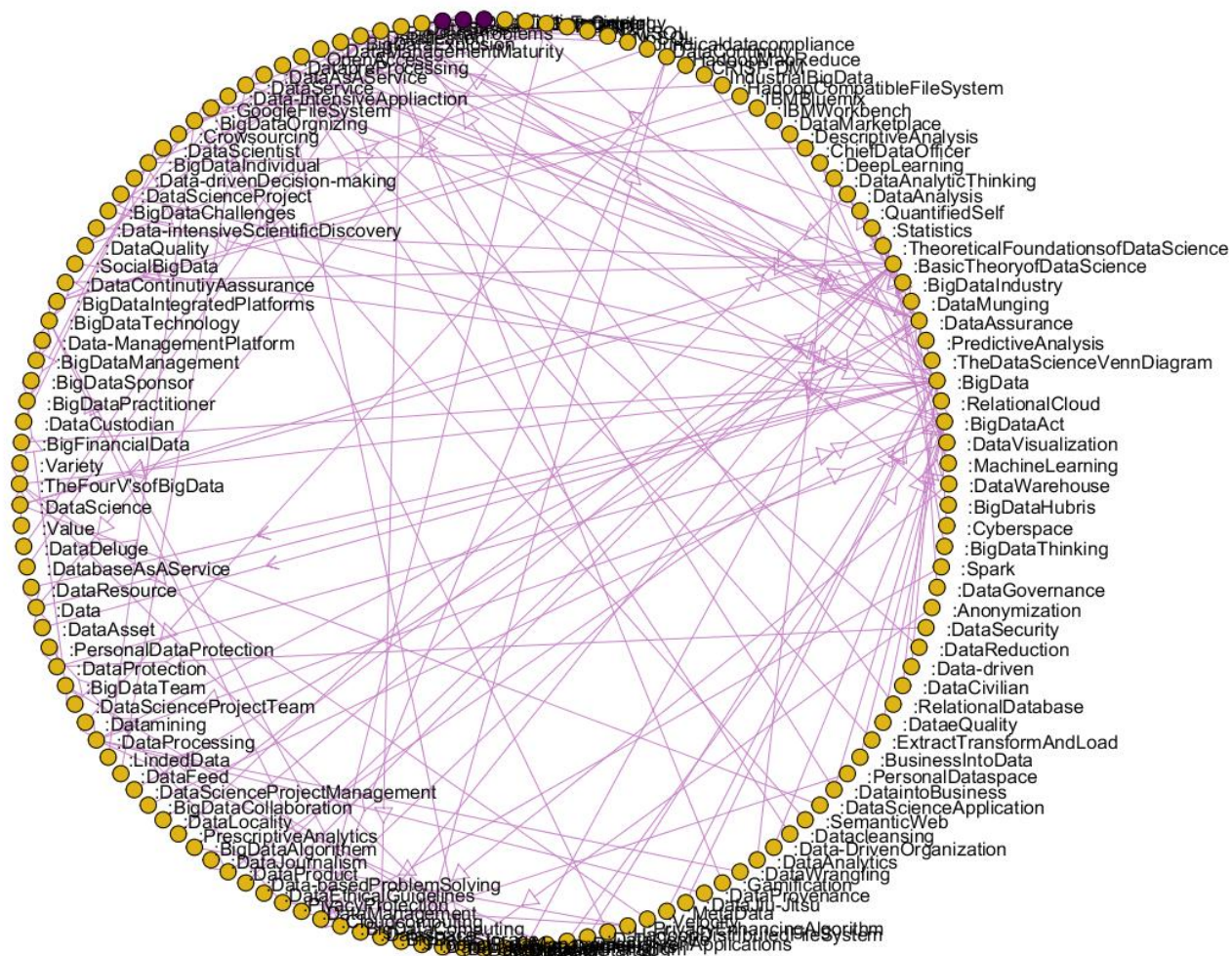
从简原则

数据范式原则

详见. 朝乐门.数据科学理论与应用[M].北京:清华大学出版社,2017.



## 7.数据科学的本体



(研发者：朝乐门等)

## 数据科学领域本体诞生——小编以5W1H方式采访研发团队（附送互动游戏）





## 8.如何继续学习本章知识

经典理论  
×  
最佳实践

理论学习  
+  
动手操作

全集知识  
—  
领域差异性知识

最深奥理论  
÷  
最基本逻辑

学习“数据科学”的四则运算基本原则  
(来源: 朝乐门.数据科学[M].清华大学出版社,2016)

一张图告诉你如何轻松学习《数据科学》



# 参与开源项目和竞赛的重要性

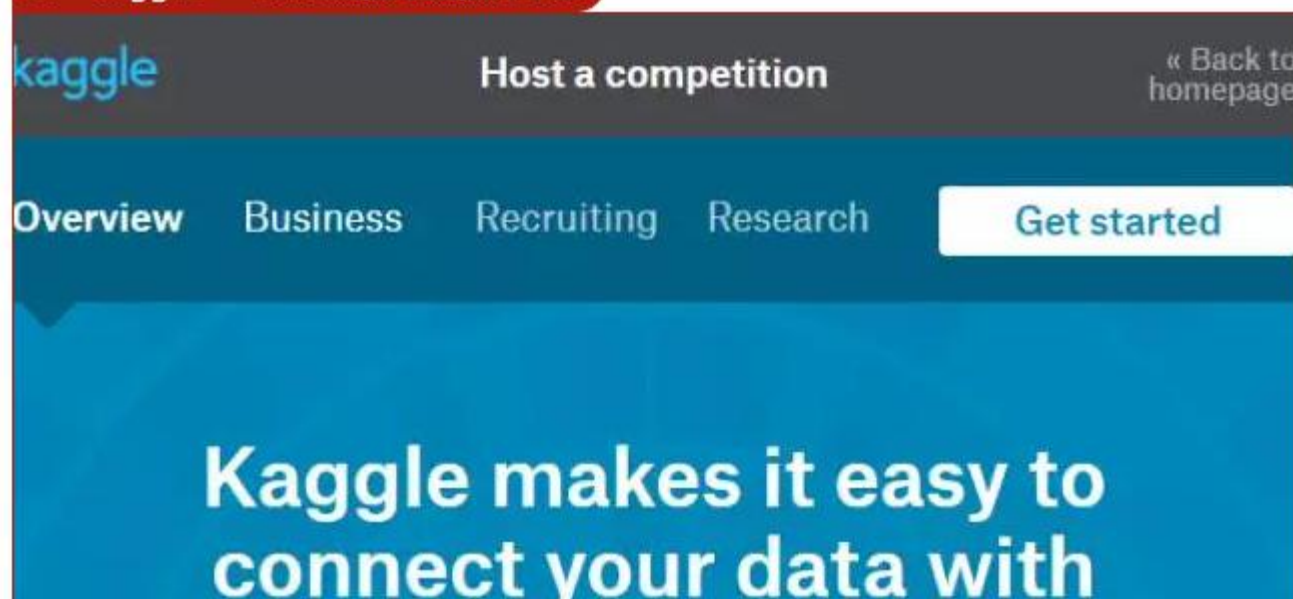
“

竞赛与开源项目是通往数据科学家的两条最佳捷径。

——摘自《数据科学理论与实践》

国内外数据科学领域的重要竞赛

01 Kaggle——连接数据与数据科学家



# 小结

