

# Introduction to Data Science



# Chapter 7

# Challenging Issues in Data Science

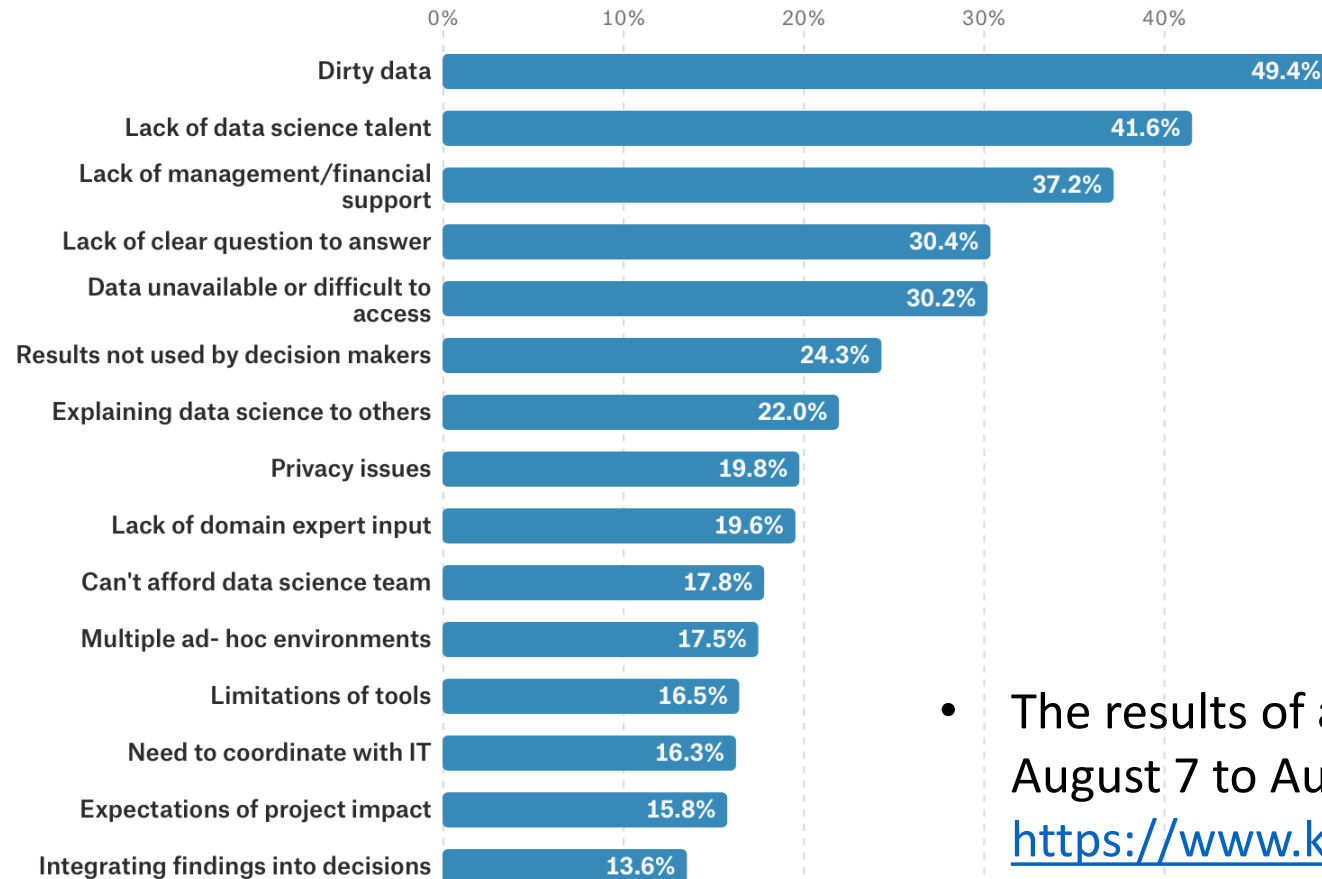
Papangkorn Inkeaw, PhD

Department of Computer Science, Faculty of Science  
Chiang Mai University



# Challenging Issues in Data Science

## What barriers are faced at work?



- The results of a survey conducted by Kaggle, from August 7 to August 25, 2017.


<https://www.kaggle.com/surveys/2017>

7,376 responses

Only displaying the top 15 answers. There are 7 answers not shown.

# Challenging Issues in Data Science

## What barriers are faced at work?

- 
- Dirty Data
  - Data unavailable or difficult to access
  - Privacy issues
  - Lack of clear question to answer
  - Lack of domain expert input
  - Results not used by decision makers
  - Explaining data science to others
  - Need to coordinate with IT

- 
- Lack of data science talent
  - Integration finding into decisions
  - Multiple ad-hoc environments
  - Limitations of tools
  - Lack of management/financial support
  - Can't afford data science team
  - Expectation of project impact

# Challenging Issues in Data Science

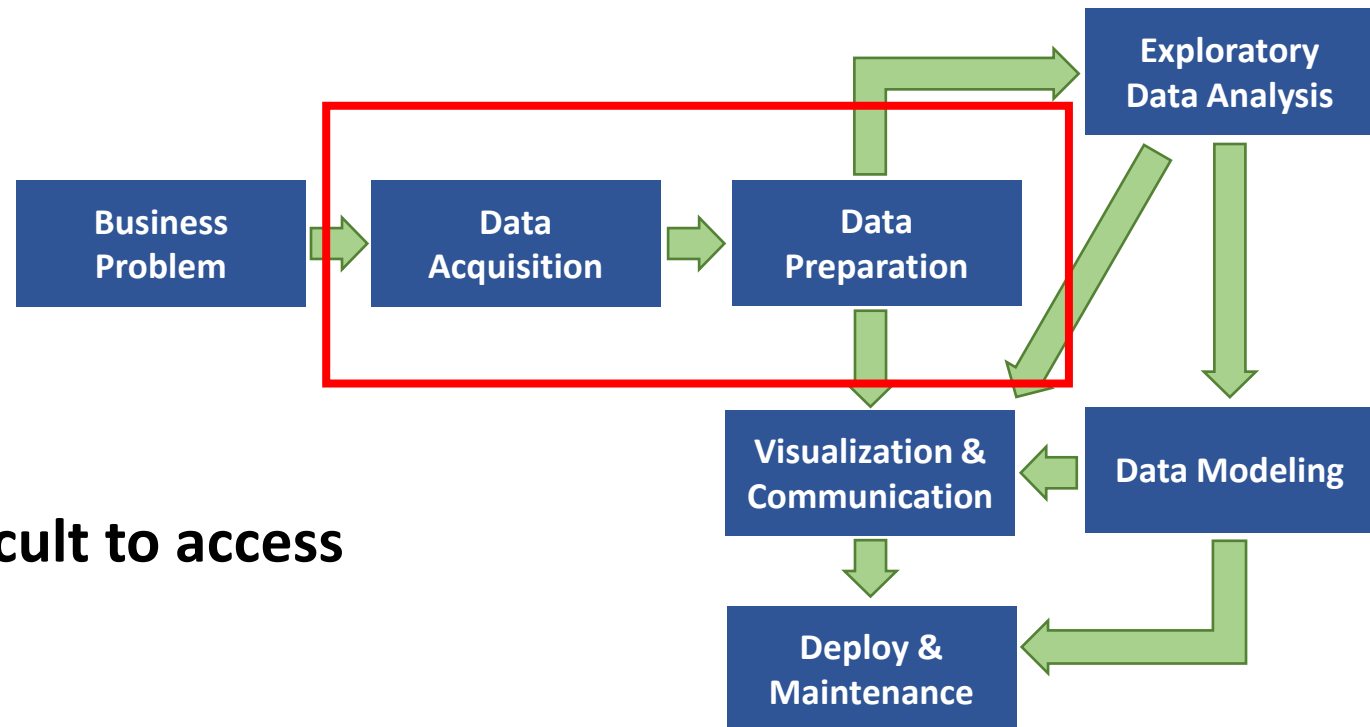
**Data is the biggest challenge in data science**

## 1. Dirty Data

- Inaccurate
- Incomplete
- Inconsistency
- Invalid
- Nonuniform
- Duplicate

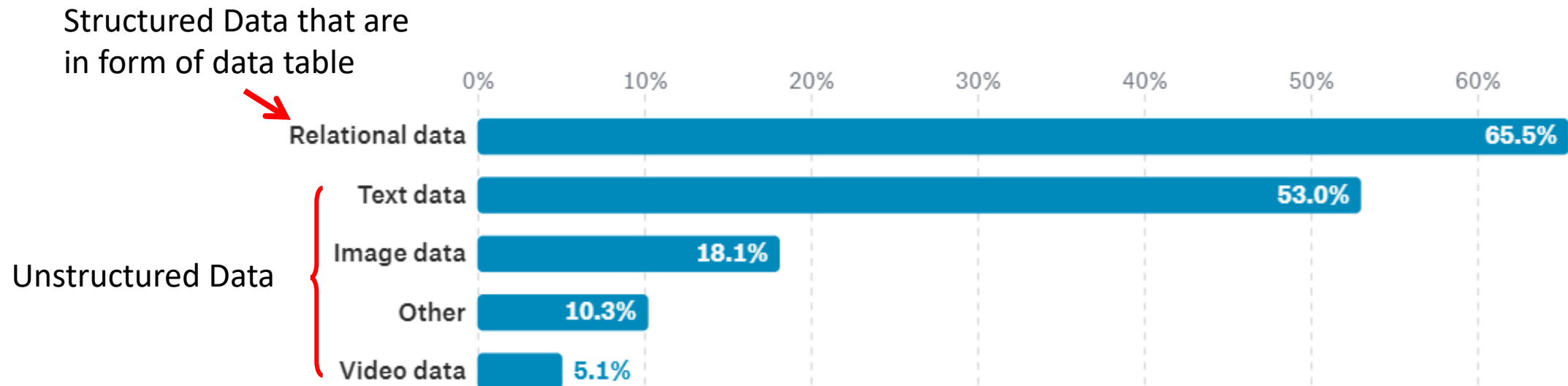
**2. Data unavailable or difficult to access**

**3. Privacy issues**



# Challenging Issues in Data Science

## Types of data used at work



A survey conducted by Kaggle, from August 7 to August 25, 2017.

<https://www.kaggle.com/surveys/2017>





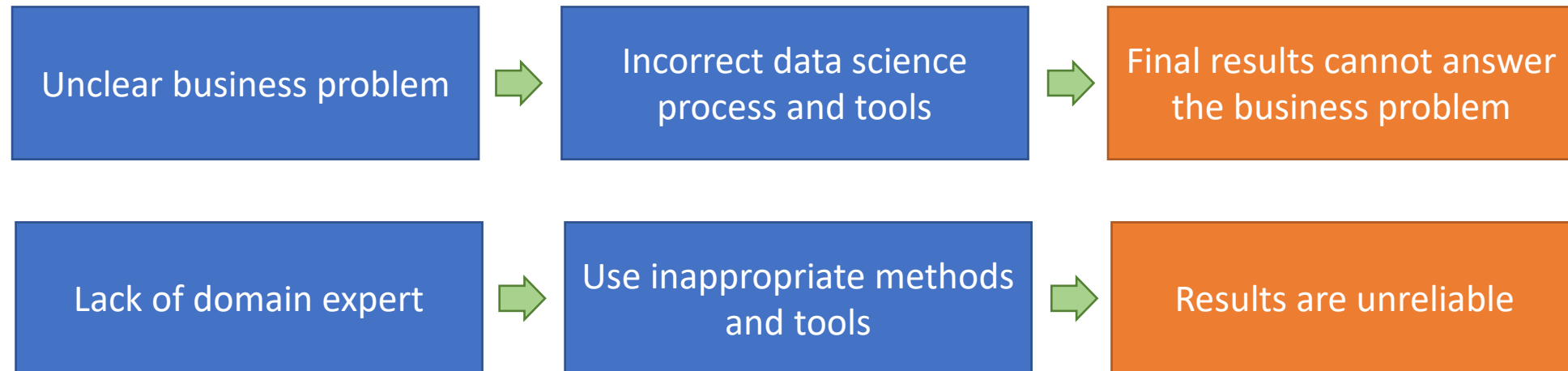


# Challenging Issues in Data Science

## Results not used by decision makers

This issue may be caused by

- Lack of clear question to answer
- Lack of domain expert input



# Challenging Issues in Data Science

## Collaboration and organizational issues

1. Explaining data science to others
2. Need to coordinate with IT

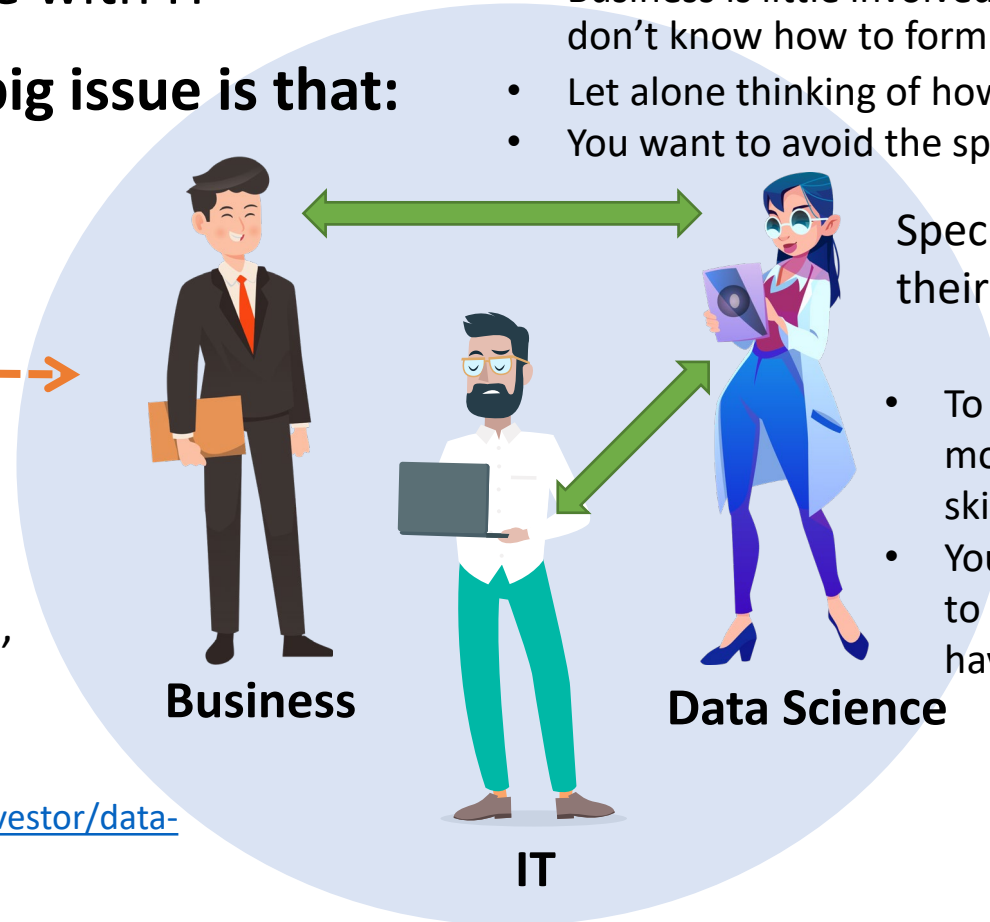
**The reason it is a big issue is that:**



Need to add to the mix generalists, who understand business, data science and IT.

Source: <https://medium.com/datadriveninvestor/data-science-challenges-b7622b85b807>

- Business is little involved because they don't know how to formulate the problem.
- Let alone thinking of how to solve it.
- You want to avoid the specialist bias.



Specialists tend to have their own language

- To minimize the time required to put a model in production, good software design skills are required.
- You do not want your development team to have to rewrite the code data scientists have produced.

# Challenging Issues in Data Science

## Talent

1. Lack of data science talent
  - The generalist talent is required to do the initial problem solving
2. Integration finding into decisions
  - Need more experiences.



# Challenging Issues in Data Science

## Tools

### 1. Multiple ad-hoc environments

Ad-hoc code needs to be written to optimize the computing time required to run the models.

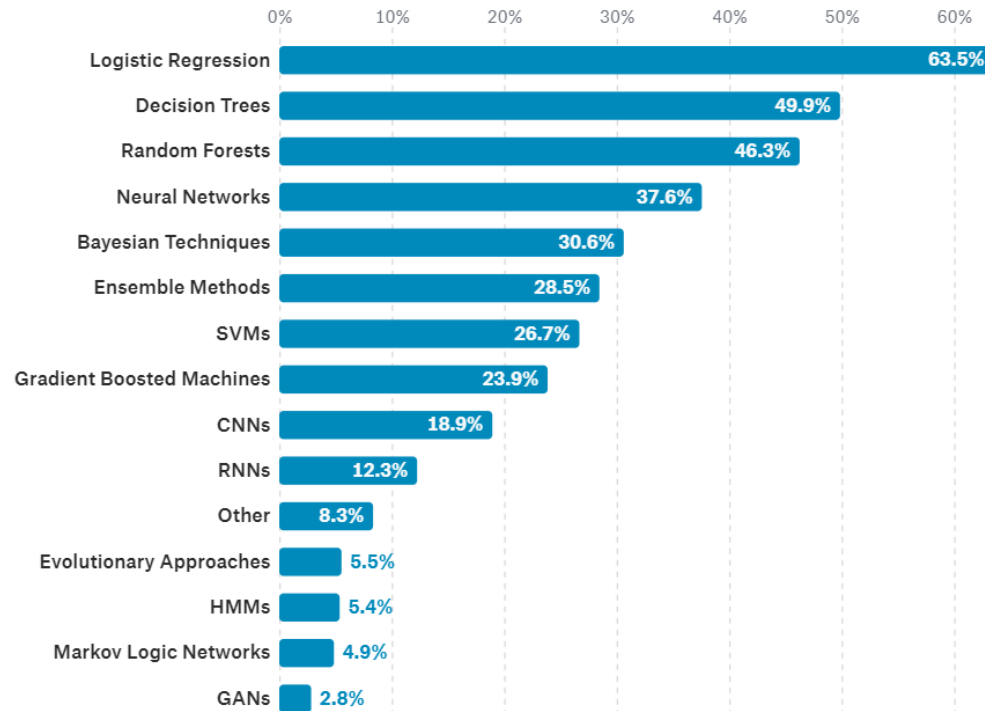
### 2. Limitations of tools

- Limitation of software
  - A software has both advantages and drawbacks.
  - We need to use more than one software to perform a data science project.
- Limitation of hardware
  - High computing power
  - Expensive technology stack
- Limitation of algorithms/methods
  - Simpler algorithms might limit your findings.
  - Most advanced algorithms do not scale well with the size of the datasets.

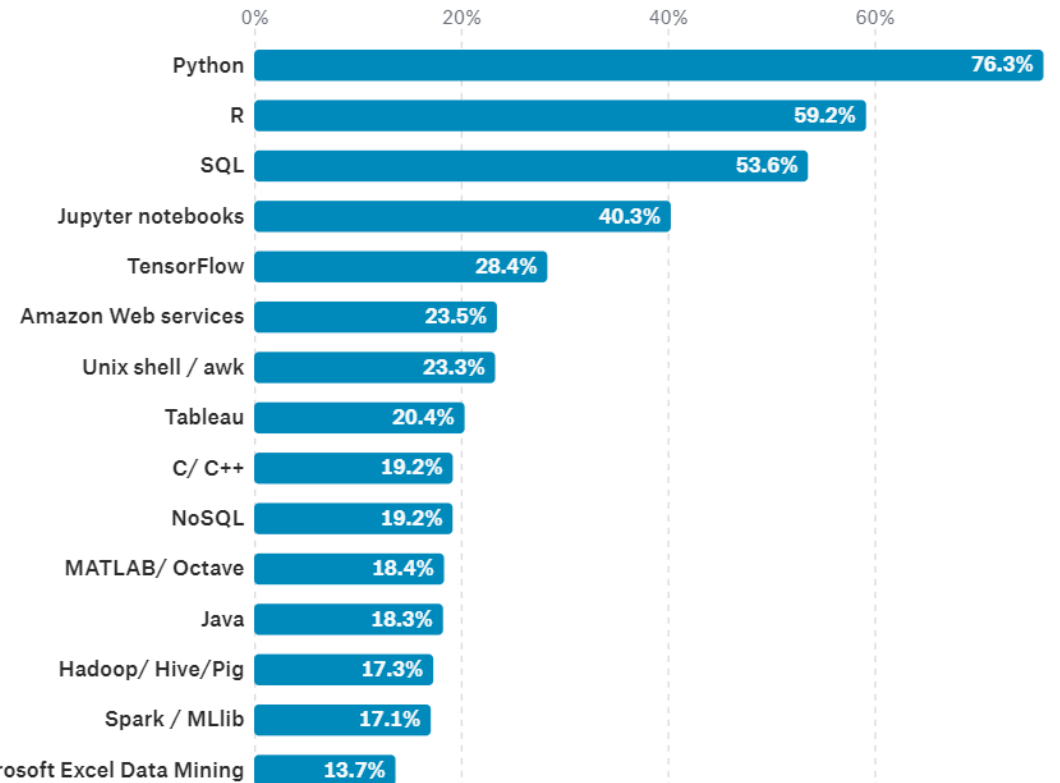
# Challenging Issues in Data Science

## Tools

Most commonly data science method used at work



Most commonly used data analysis tool



A survey conducted by Kaggle, from August 7 to August 25, 2017.

<https://www.kaggle.com/surveys/2017>

# Challenging Issues in Data Science

## Budget

1. Lack of management/financial support
2. Can't afford data science team
3. Expectation of project impact
  - If the costs are higher than the expected outcome of the project, then there is no point doing this project.



# We need your comments

ขอความร่วมมือนักศึกษาทุกคนเข้าไปประเมินและให้ข้อเสนอแนะ/ความคิดเห็นต่อ  
กระบวนวิชาและอาจารย์ผู้สอน ผ่านทาง CMU SIS

<https://sis.cmu.ac.th/cmuisis>

# Further Study

## Website:

- <https://www.kaggle.com/surveys/2017>
- <https://medium.com/datadriveninvestor/data-science-challenges-b7622b85b807>