数据分析基础

主讲人：刘宏志
liuhz@ss.pku.edu.cn

Data Analysis
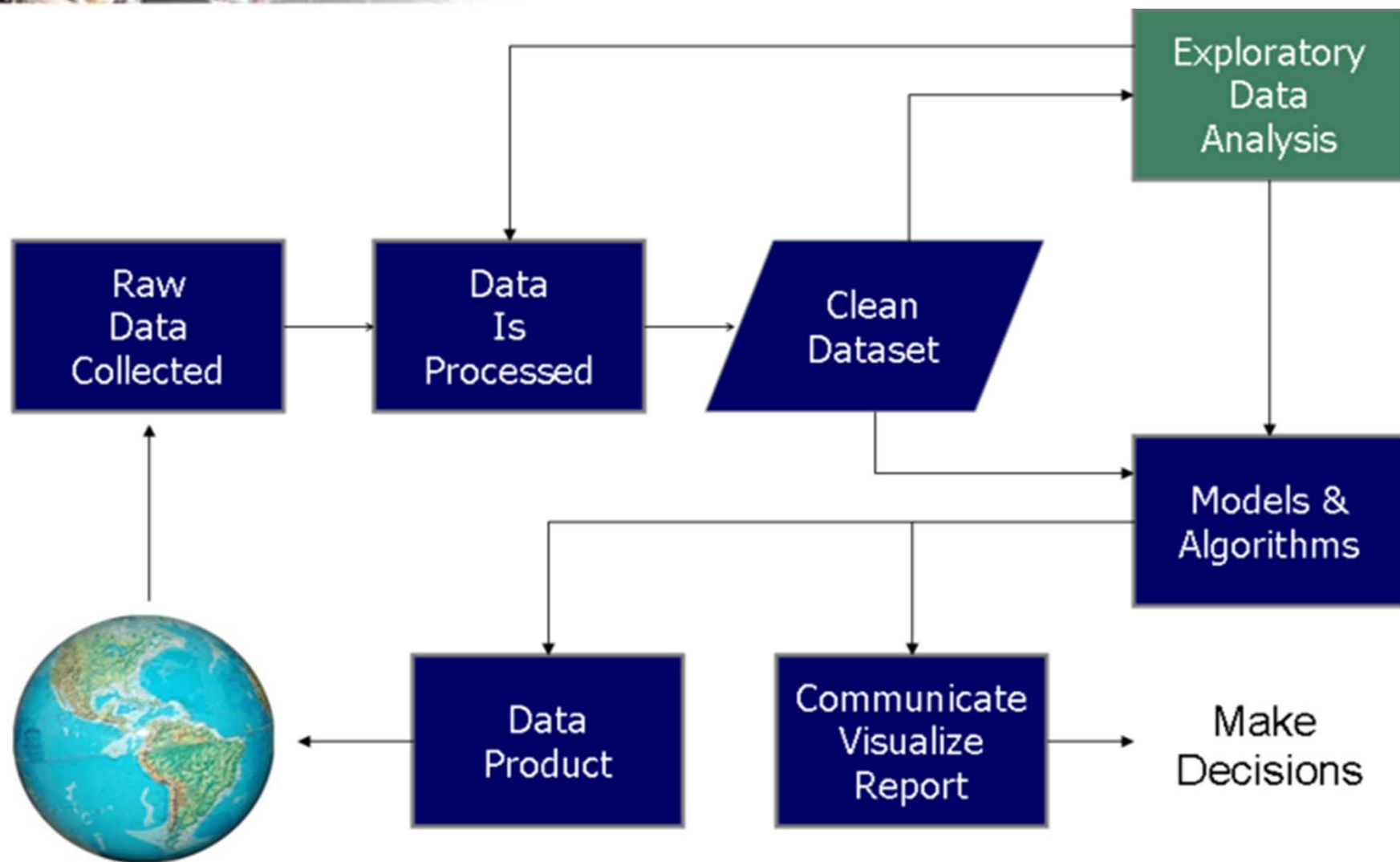
- A process of inspecting, cleaning, transforming, and modeling data with the goal of discovering useful information, suggesting conclusions, and supporting decision-making. –Wiki

- Varieties of data analysis:

  ➢ **Data mining** focuses on modeling and knowledge discovery for predictive rather than purely descriptive purposes.

  ➢ **Business intelligence** focuses on business information and relies heavily on aggregation.

  ➢ **Text analytics** applies statistical, linguistic, and structural techniques to extract and classify information from textual sources, a species of unstructured data.

  ……

Phases of Data Analysis
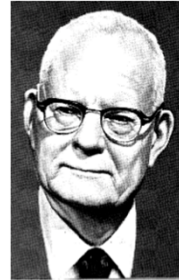
A Short History of Data Analysis

1935: "The Design of Experiments"

R.A. Fisher
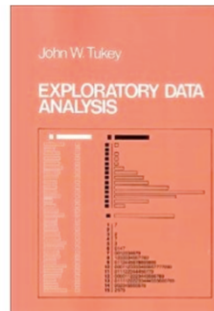
1939: "Quality Control"

W.E. Demming

1958: "A Business Intelligence System"

Peter Luhn

1977: "Exploratory Data Analysis"

1989: "Business Intelligence"

Howard Dresner

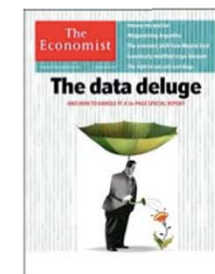1997: "Machine Learning"

2010: "The Data Deluge"

2007: "The Fourth Paradigm"

2009: "The Unreasonable Effectiveness of Data"

1996: Google
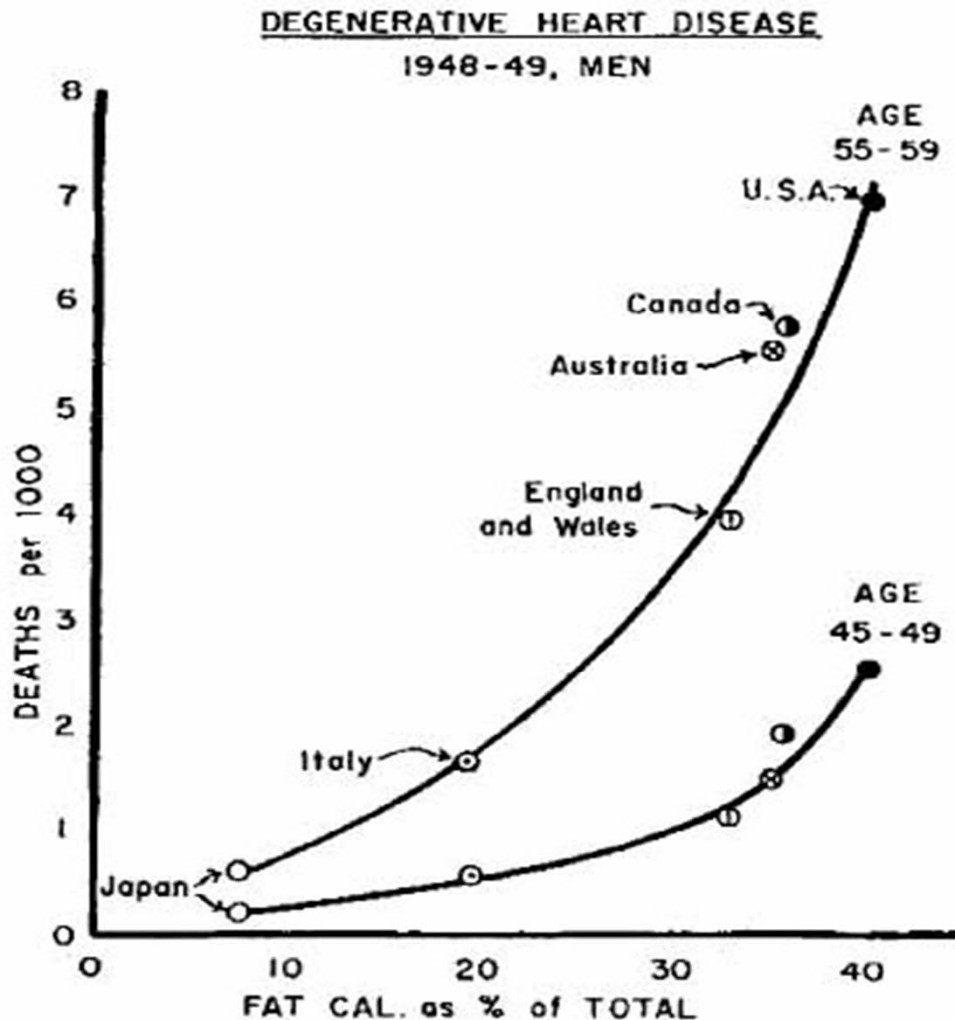
北京大学

Data Makes Everything Clearer

- ## Seven Countries Study (Ancel Keys)
- 13,000 subjects total, 5-40 years follow-up.

DEGENERATIVE HEART DISEASE
1948-49, MEN

课程介绍

- 教材
  - ➢ 数理统计与数据分析，John A. Rice （第3版）

- 参考书
  - ➢ 深入浅出数据分析
  - ➢ 统计模型：理论与实践（第2版）
  - ➢ Data Analysis: What Can Be Learned from th
  - ➢ Data Analysis: A Bayesian Tutorial

- 评分体系：
  - ➢ 平时成绩 40%： 作业、课堂练习、表现
  - ➢ 项目成绩 20%： 小组完成
  - ➢ 期末笔试 40%： 闭卷考试