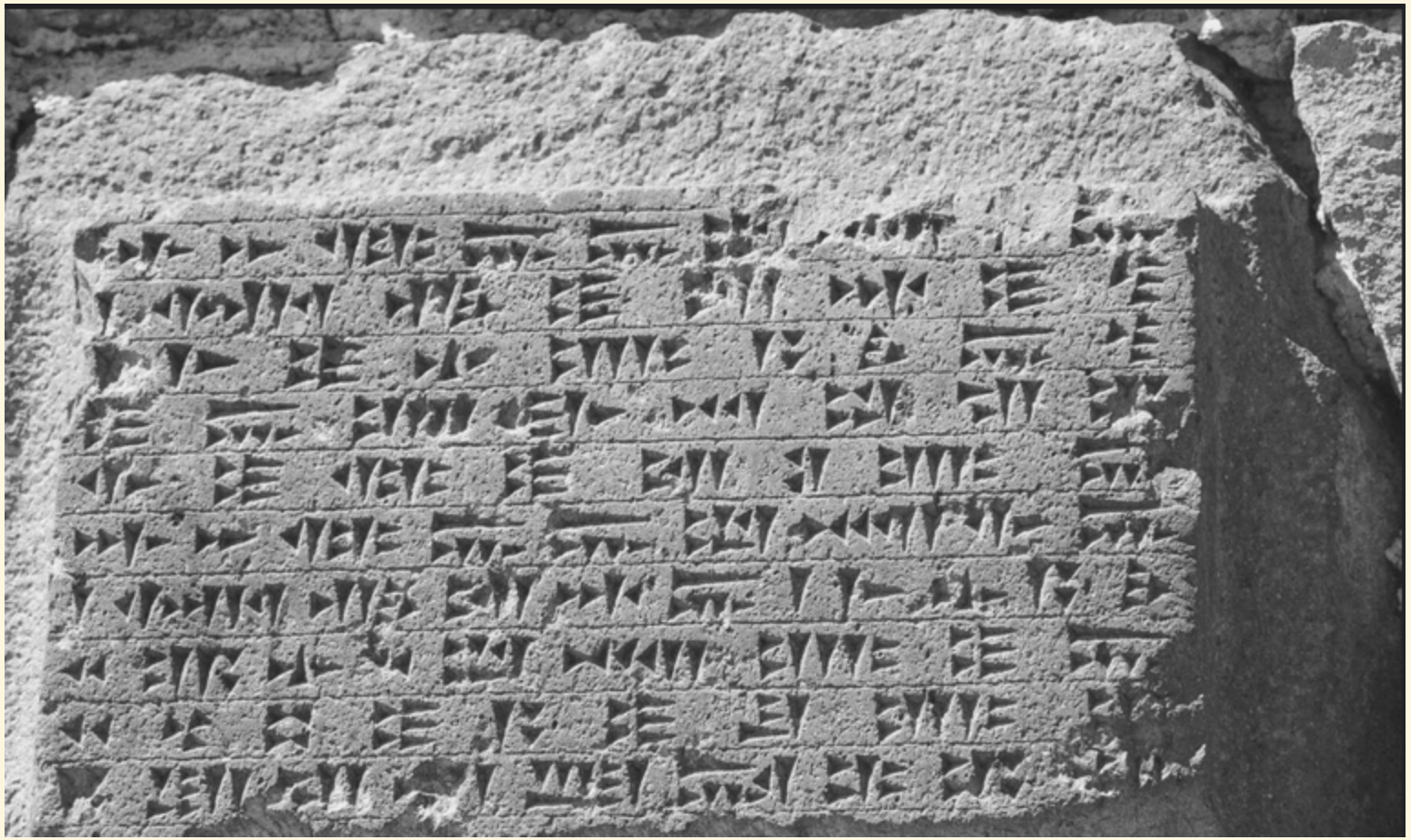
Data Analysis and Visualization

1. Introduction to -- *REAL* -- Data Analysis

1.1 A Brief History of the Data Revolution

1.1.1 From Writing to the Internet



1. **Writing Invention - Mesopotamia**: The earliest form of data recording—cuneiform tablets used to track trade, taxes, and inventories.

2. **Gutenberg's Printing Press**: Made books accessible, empowering the spread of knowledge and enhancing religious and political influence.

3. **Telegraph / Phone**: Instant communication across distances revolutionized commerce and personal connectivity.

4. **The Internet**: The advent of websites, email, and social media made data ubiquitous, driving modern global communication and business.

1.1.2 Data as a Source of Power and Wealth

- **Data → Insight → Decision → Value**:
  - Analyze situations and options.
  - Make informed decisions.
  - Drive actions that generate power and wealth.

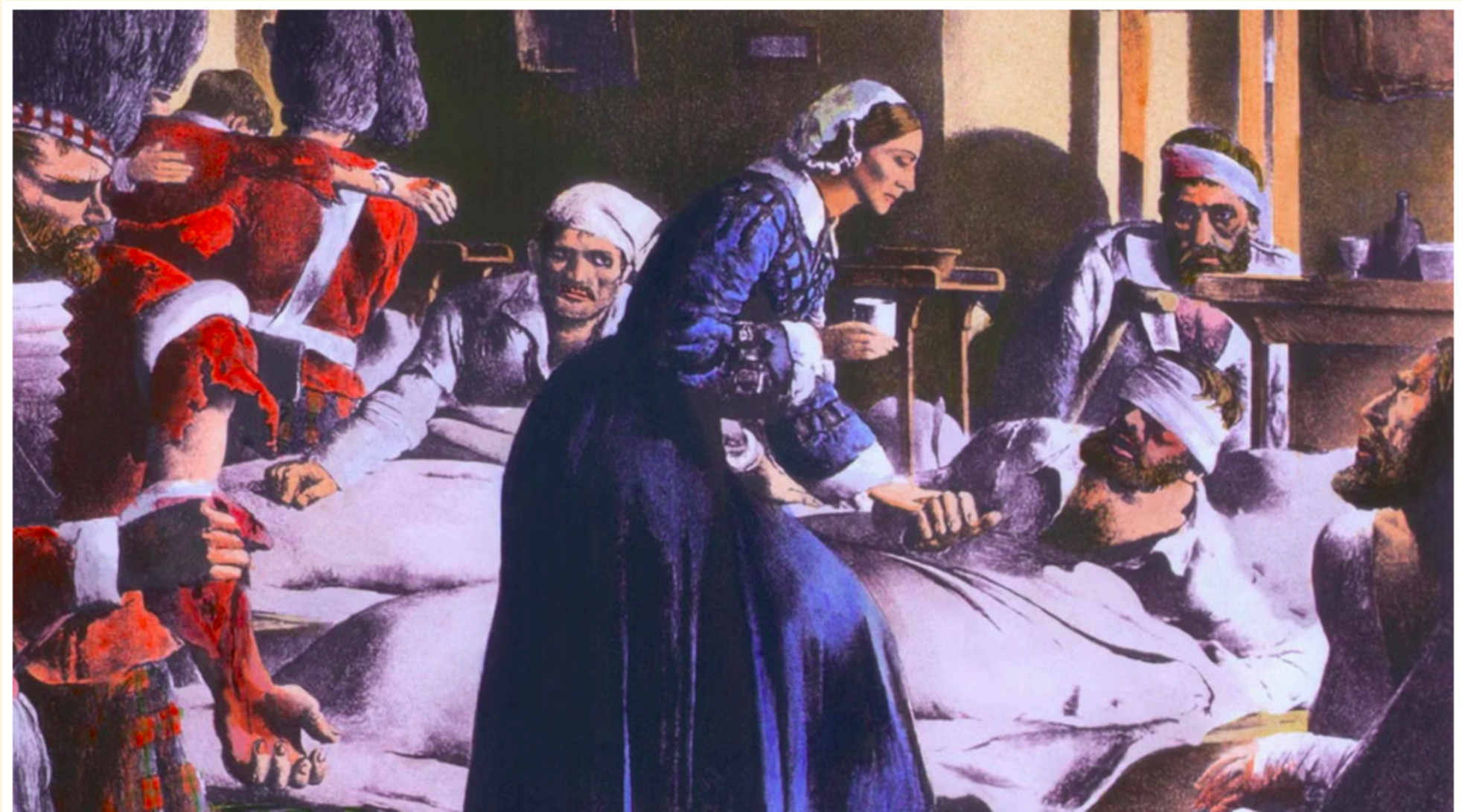
- **The Double-Edged Sword**:
  - Wrong data can lead to successful outcomes by luck.
  - Good data and analysis might fail due to unforeseen circumstances.

1.1.4 Short Stories from Data History

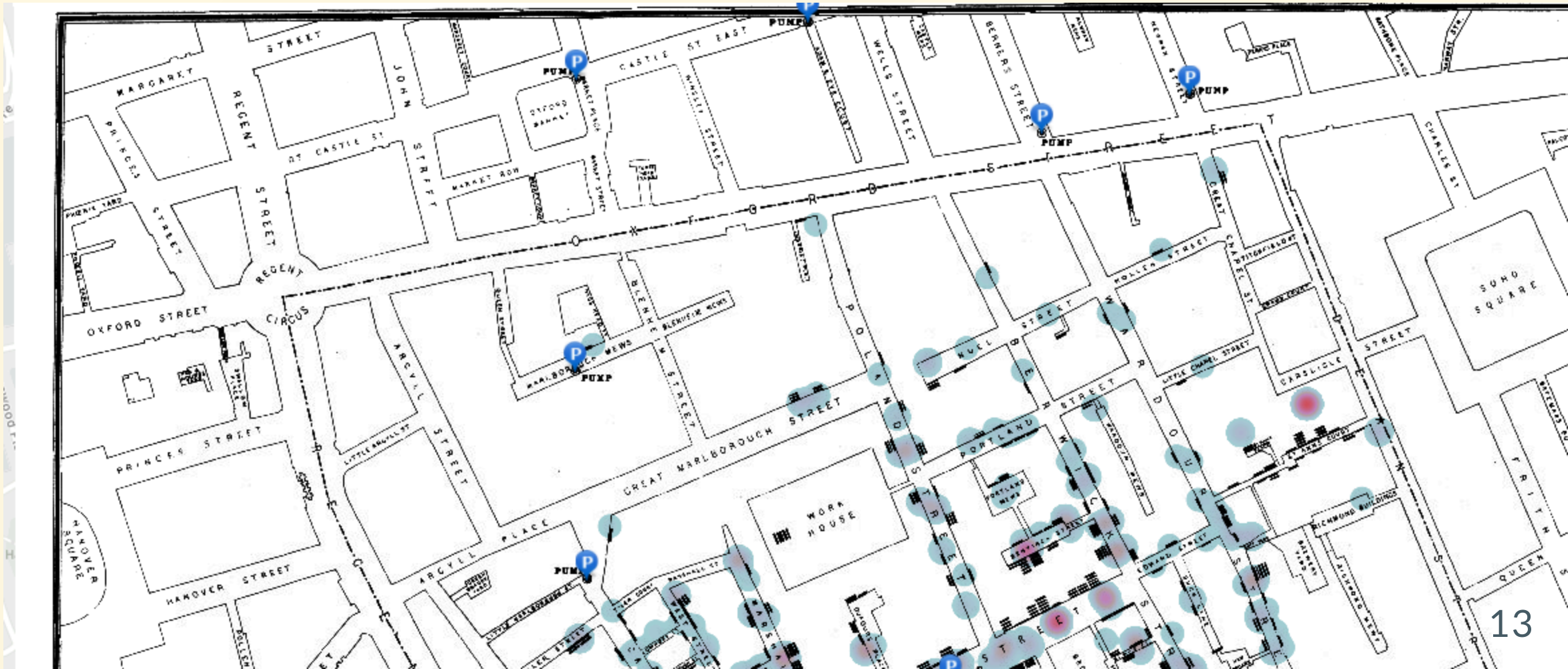
1. **Mesopotamian Trade Records**: Early accounting systems to track wealth.

2. **Domesday Book (England)**: Commissioned by William the Conqueror to assess land and taxes.

3. **Nurse in Wartime**: Florence Nightingale's use of data visualization to highlight sanitary conditions over battlefield injuries.



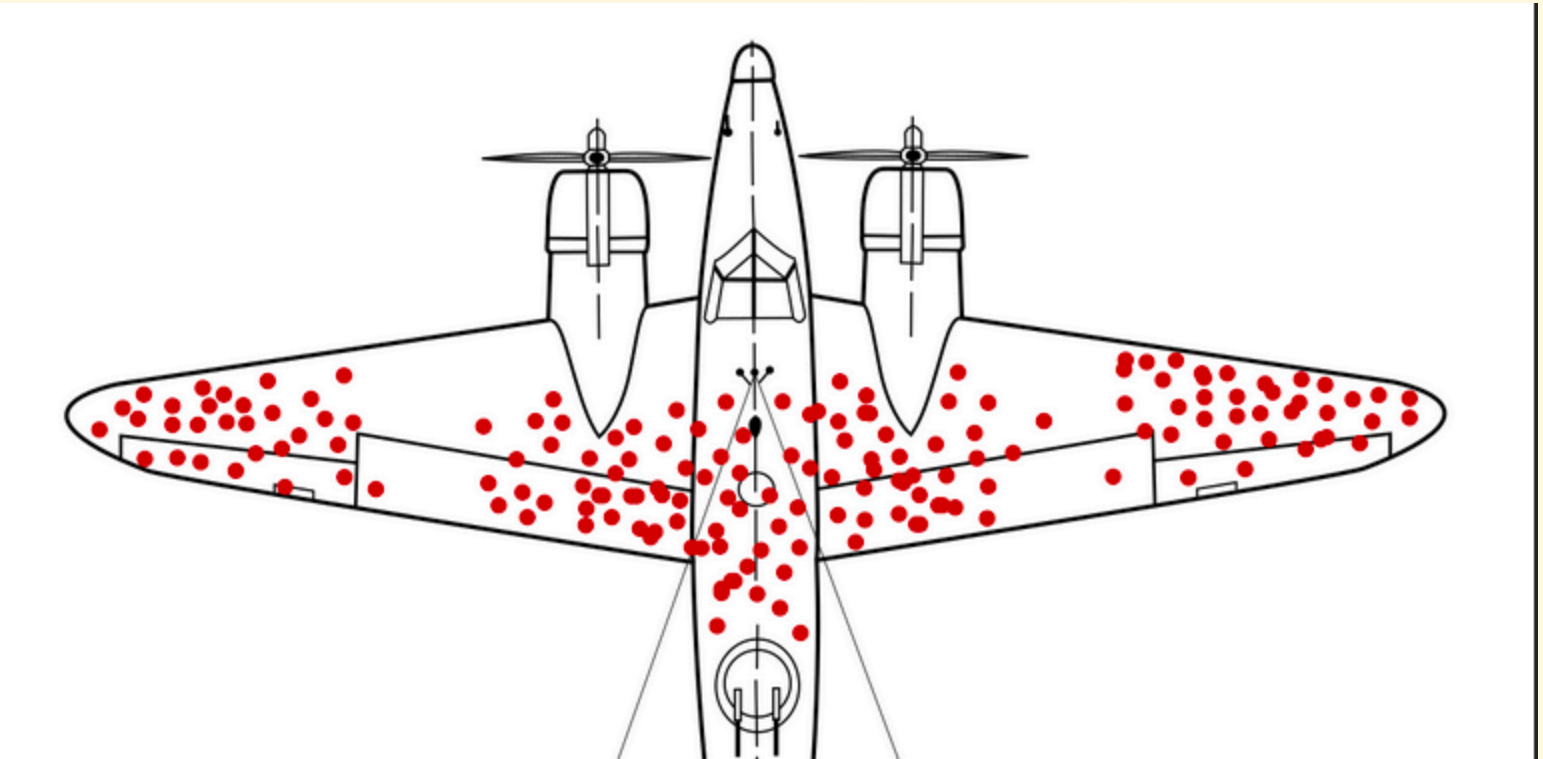
4. **London Cholera Epidemic**: John Snow's mapping of cholera clusters linked to contaminated water—a cornerstone of modern epidemiology.



1.2 Misinterpretations of Data

1.2.1 WWII Plane Armor

- Observing bullet holes on returning planes led to the assumption armor should be placed there.

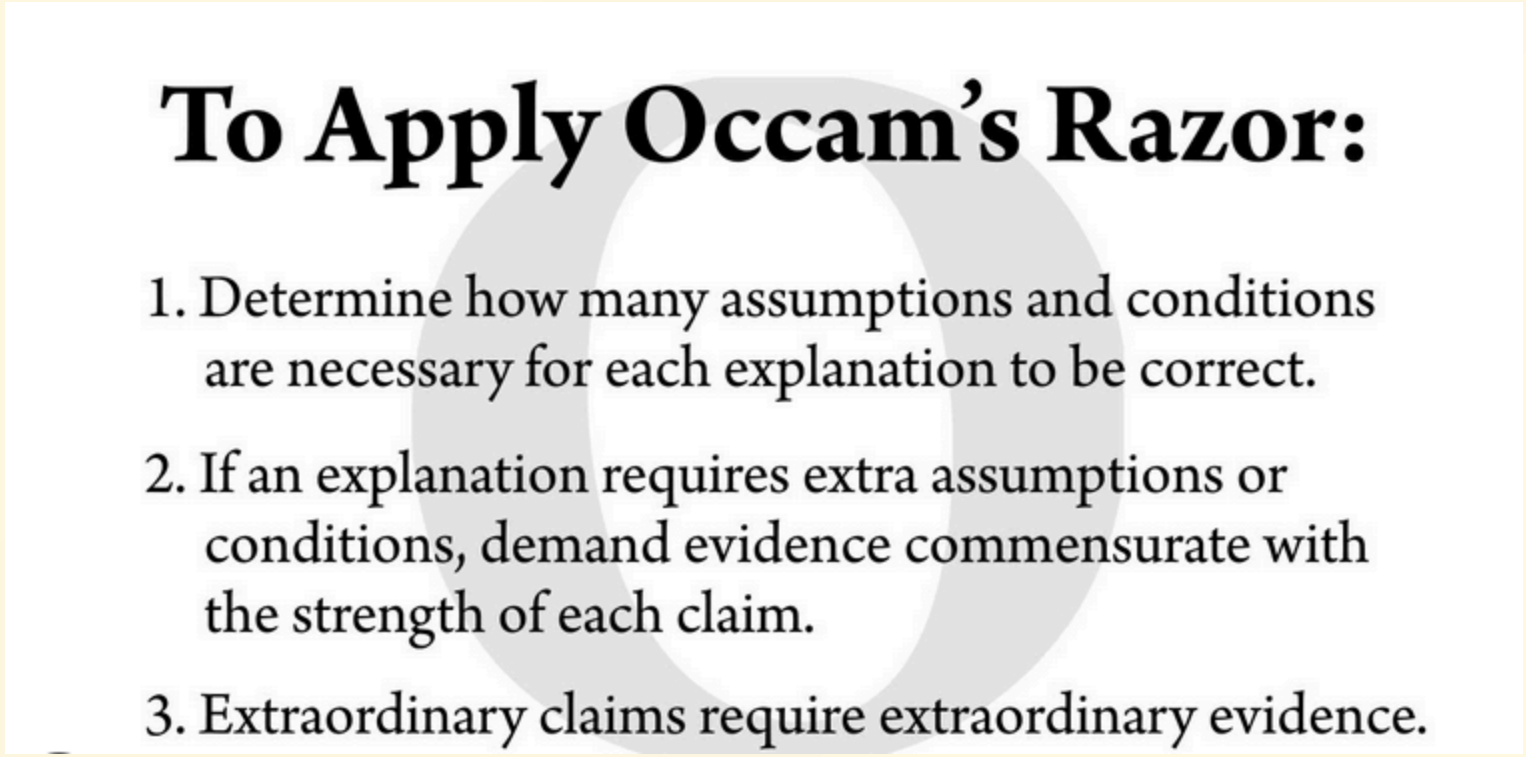


- **The Insight**: Focus on parts where planes didn’t sustain damage but were still crucial to survival—the non-returned planes provided key missing data.

1.2.2 Being Fooled by Data:

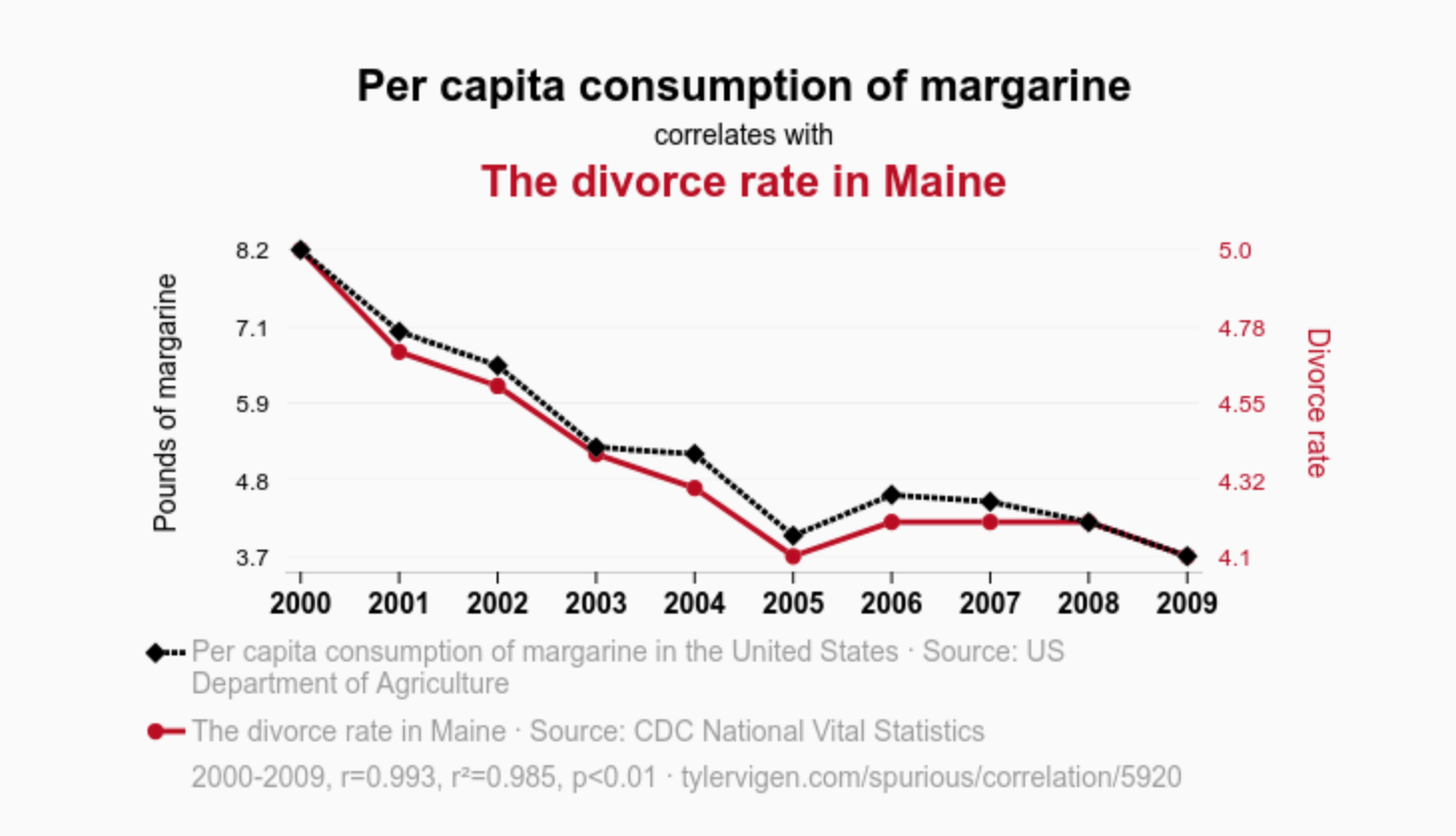
- Common cognitive biases:
  - **Confirmation Bias**: Seeking data that confirms existing beliefs.
  - **Availability Heuristic**: Overvaluing easily accessible information.
  - **Survivorship Bias**: Focusing on successful outcomes while ignoring failures.
  - **Loss Aversion**: Overvaluing potential losses over gains.
  - **Sunk Cost Fallacy**: Justifying continued investment in a failing endeavor.
  - **Anchoring**: Relying too heavily on the first piece of information.

1.2.3 Occam's Razor:



- "The simplest explanation is usually the best." --> MISSUNDERSTOOD !
- Be wary of convoluted interpretations.
- Simplest theory are always easiest to refute

1.2.4 Spurious Correlations and Hidden Variables



https://www.tylervigen.com/spurious-correlations

Example 1: Chocolate Consumption vs. Nobel Laureates

Example 2: Smoking and Speed

- Data suggested smokers are faster, but the hidden variable was **age** (younger individuals smoked more and were faster).
- Other hidden variables include gender, socioeconomic status, and education.

Famous Data Misunderstanding

1. **Modern World Data Explosion**: With data everywhere, misinterpretations have amplified in scope and impact.

1.4 Conclusion

By understanding the power of data, historical lessons, and the risks of misinterpretation, we can develop a more informed, critical approach to decision-making in a data-driven world.

2. Not so Basic Reminder about Not so Basic data Analysis

2.1 Data is not about mean, IQ ort stat test !

Data -> insignt -> decision -> value
So you ave to think reverse

- What is my uporiise ?
- ◦ Optimising n, finding
- ◦ Business, money medial
- Whatt is my problem ?
- What do i know about my probme

2.2 Getting some relevant data