Explorable Visual Analytics (EVA) Interactive Exploration of LEHD

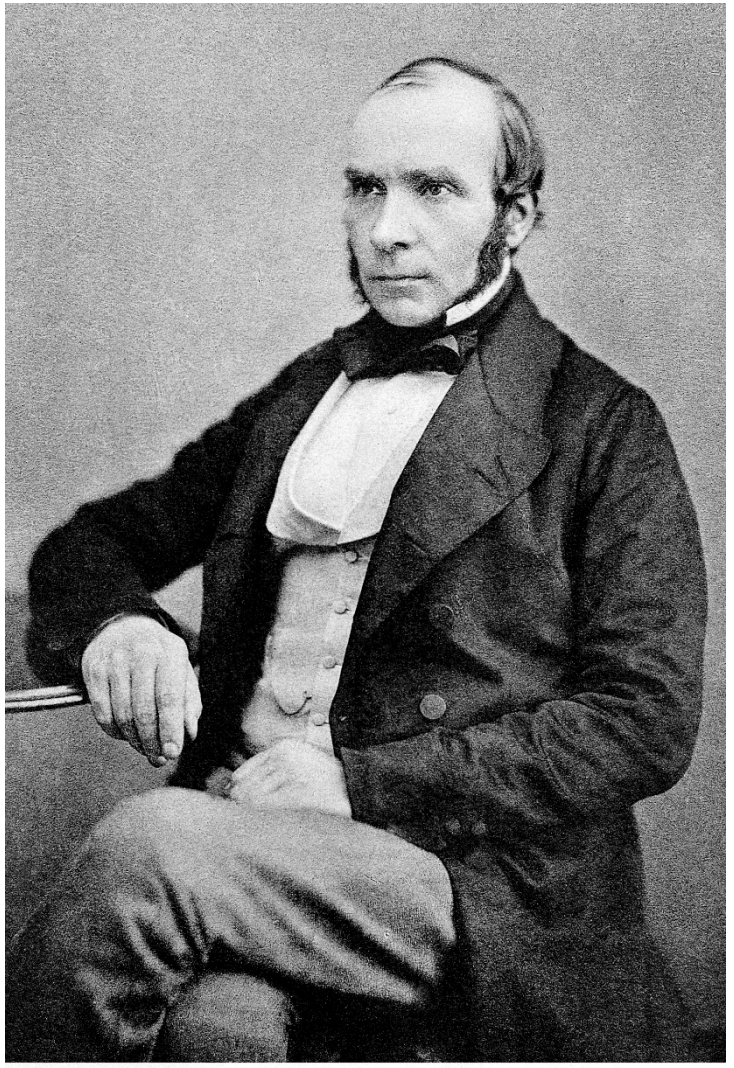
Saman Amraii - Amir Yahyavi

Carnegie Mellon University

Motivation

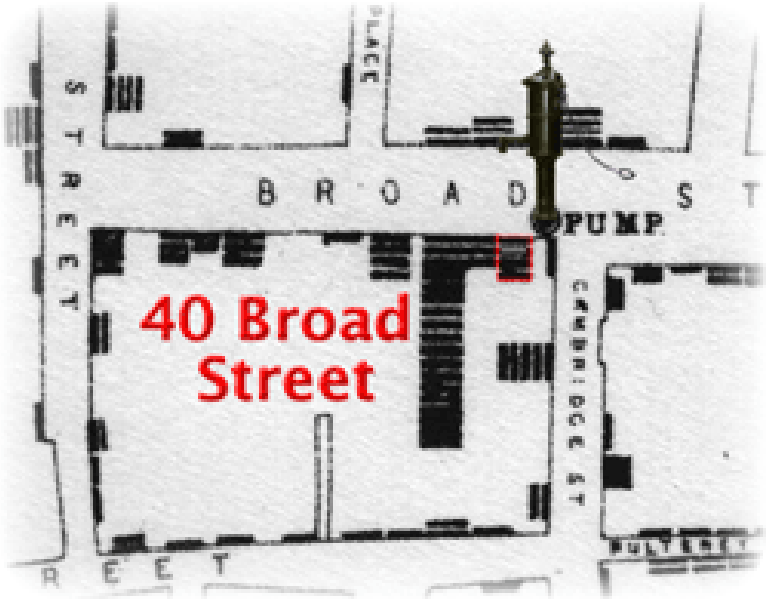
Motivation

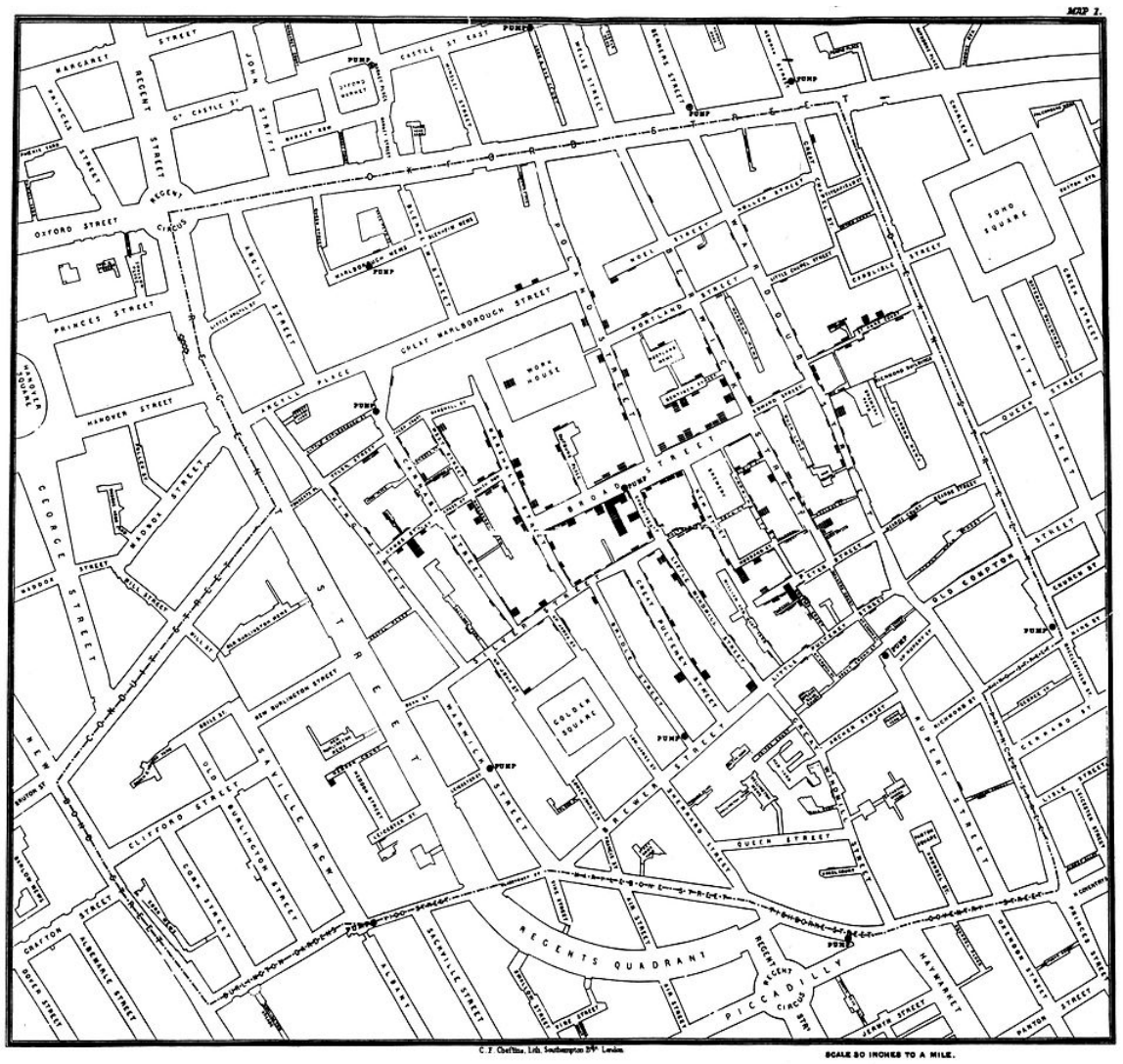
- 1854 Cholera outbreak in Soho,
  - 616 Dead
- No Germ Theory
  - Miasmatic Theory (bad air)
- John Snow
  - Spread through water supply
  - Study of water didn't help



Motivation

- Drew a map
- Broad Street Pump



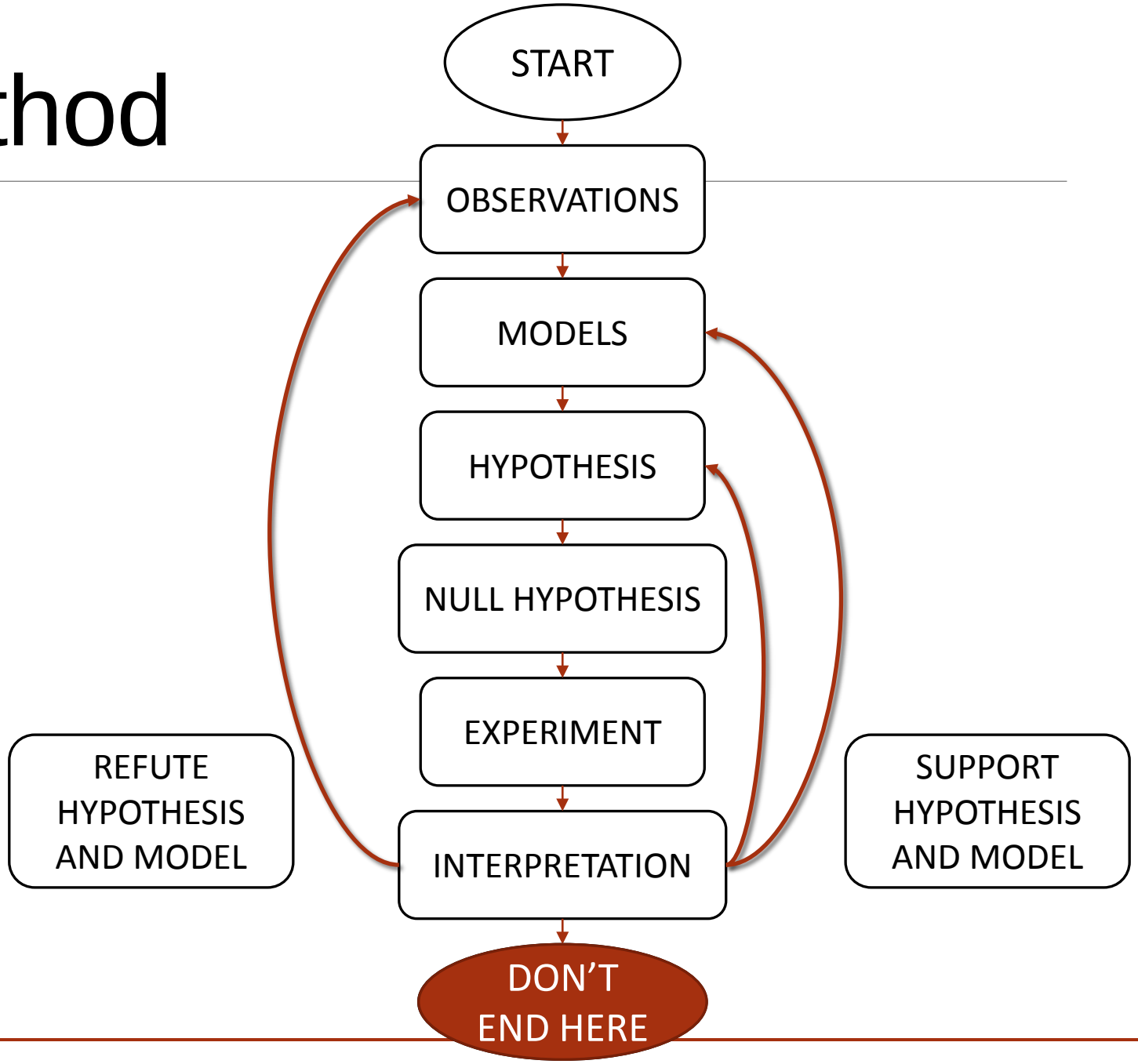


Motivation

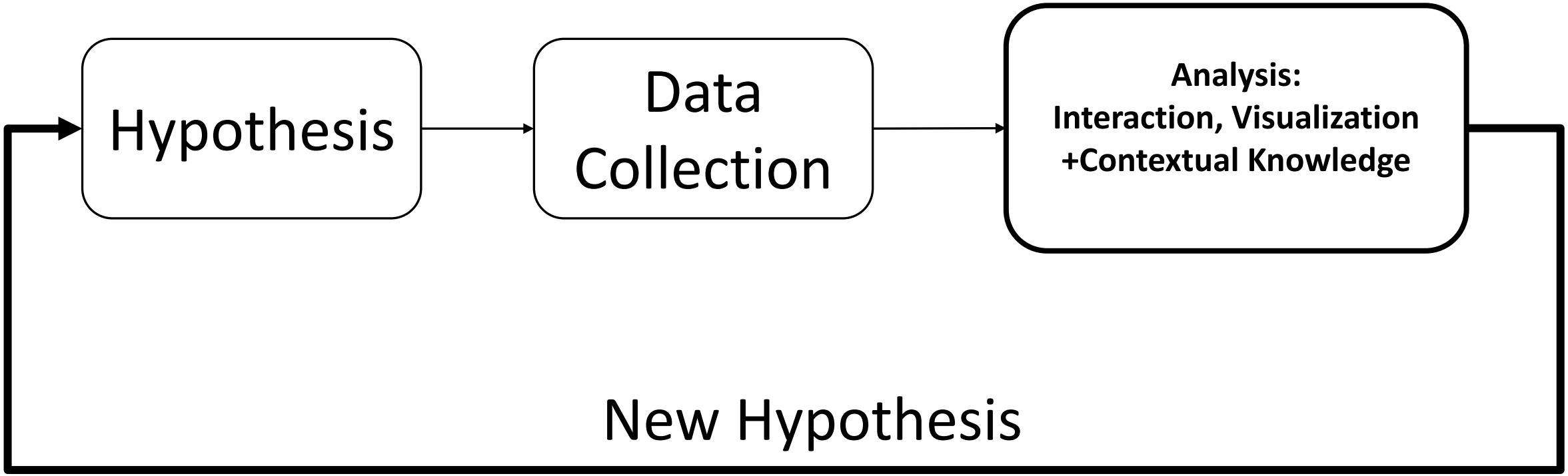
- Solved by removing a handle
- Political controversy
  - Officials rejected the theory
  - Accepted 12 years later in another outbreak

John Snow memorial and public house on Broadwick Street, Soho

Scientific Method

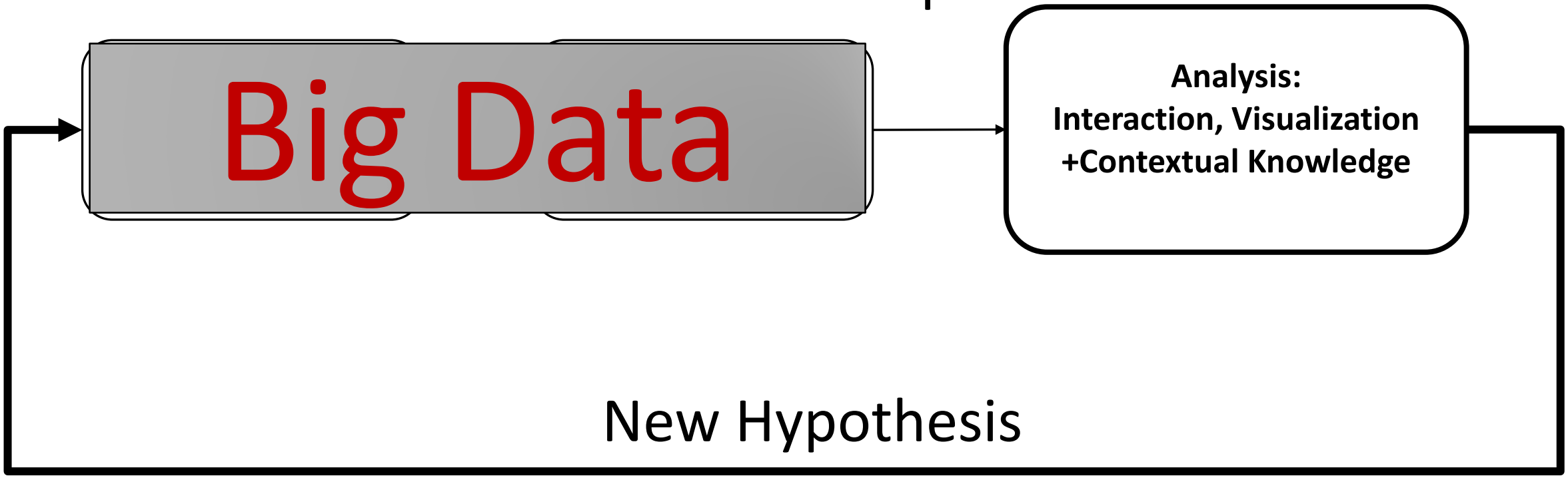


Scientific Method

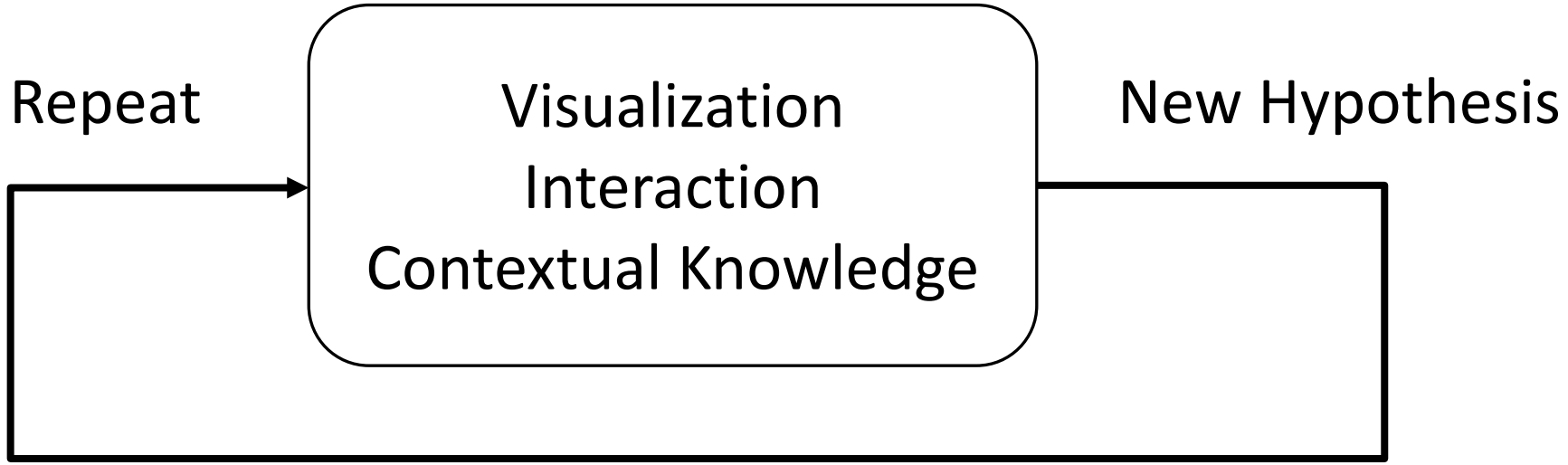


Scientific Method

Data Production >>>>>> Data Consumption



Sensemaking Loop

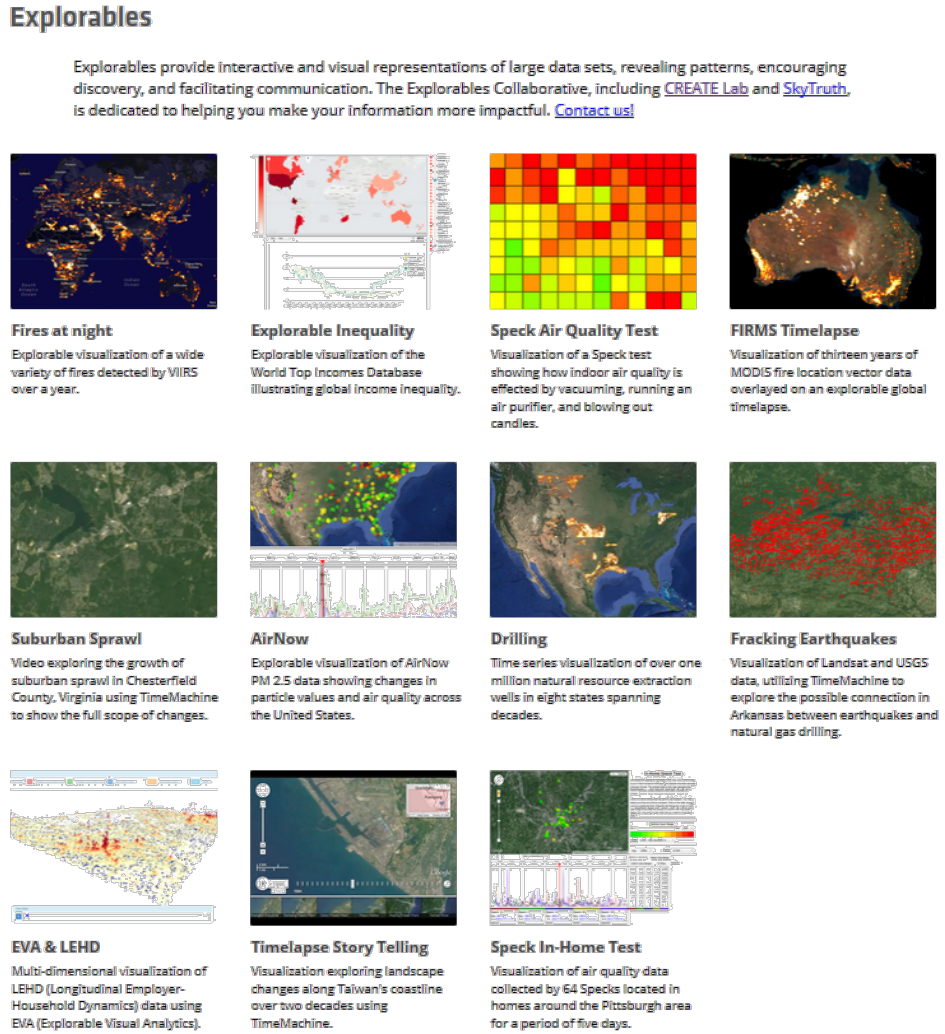


This loop should happen fast, otherwise we hesitate to explore or lose our train of thought

The Explorables Collaborative

http://explorables.cmucreatelab.org/

An effort to understand the challenges in visualizing, exploring, and analyzing **large** and **complex** data.



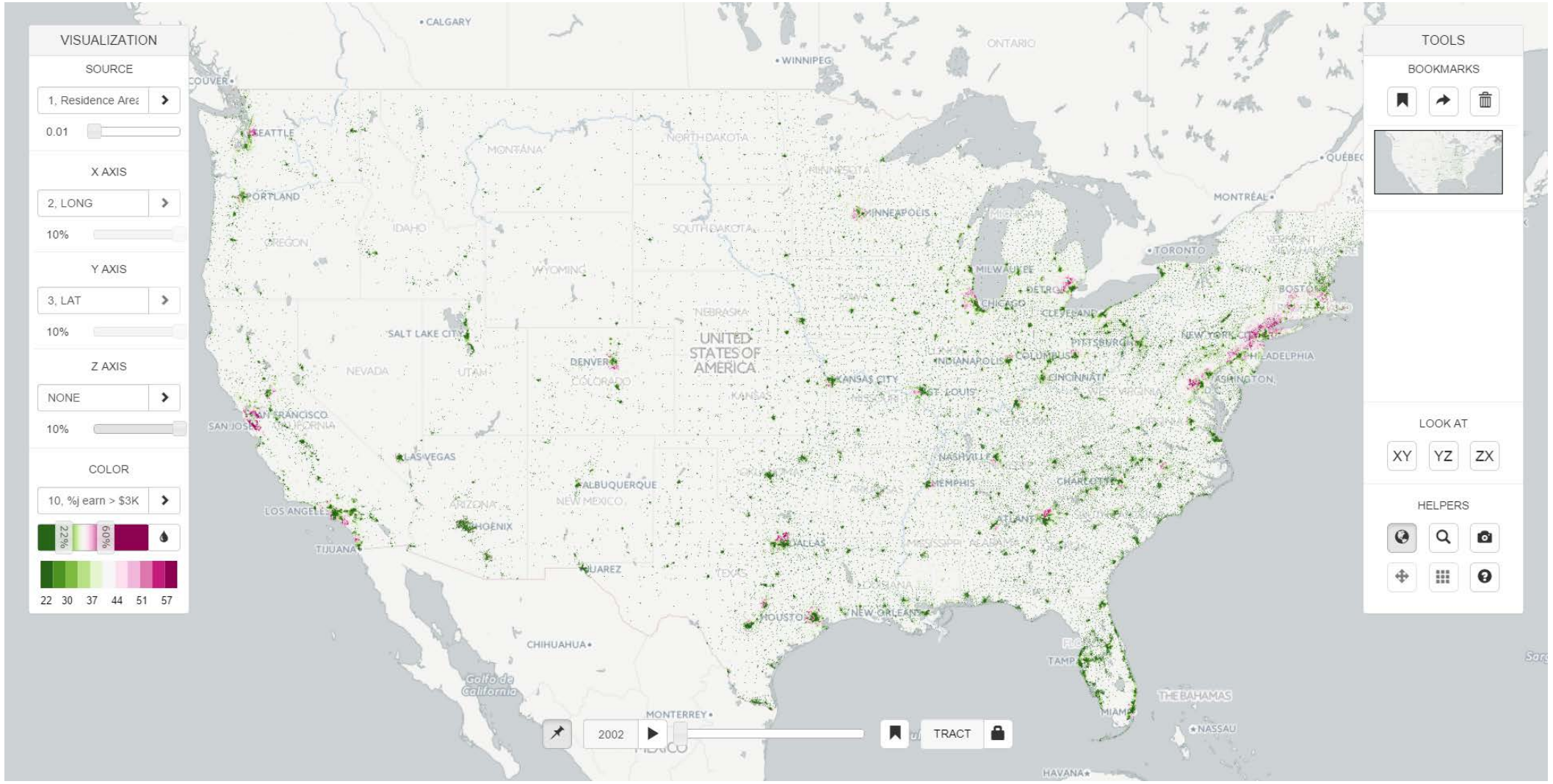
Explorable Visual Analytics (EVA)

http://eva.cmucreatelab.org

Goal: Improving **Hypothesis Generation**

- Easy Exploration
- Sharing Discoveries
- Quick Intuitions Testing

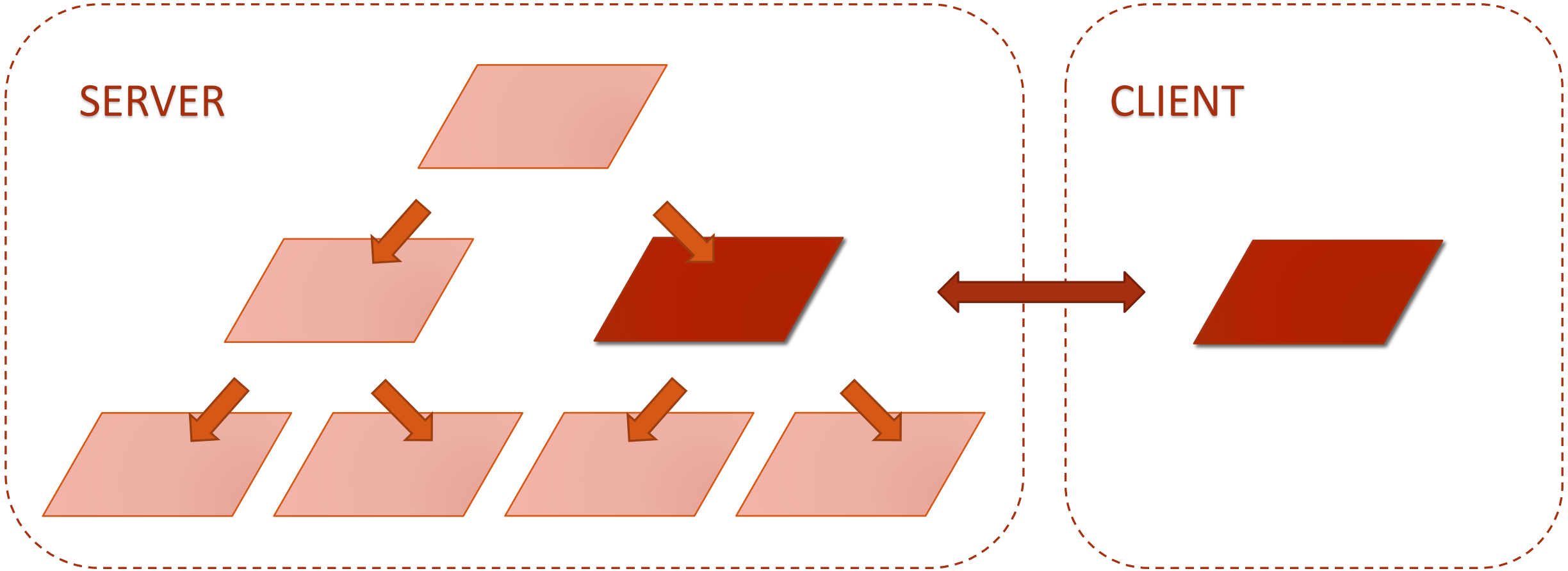
Explorable Visual Analytics (EVA)



EVA Demo↗

- **Data?** large, complex, high spatial and temporal resolution
  - Opportunities for real and meaningful discoveries
- Census Longitudinal Employer-Household Dynamics (LEHD)
  - http://lehd.ces.census.gov/
- Contiguous US
  - Hundreds of millions of data points
  - Tens of dimensions
  - 10 years (2002-2011)

Technical Information



Technical Information

- **Big Technical Challenges:**
  - **Massive size of Data:** Big Data Processing on Server + Client Side Analysis
- **Open Source**
  - https://github.com/CMU-CREATE-Lab/EVA-for-Census
- **Compatibility:** Any Modern Desktop Browser (Any OS, 1~2 GB of RAM)

Key Aspects

Improving Hypothesis Generation, How?

- **Scalable:** Interactive visualization of large, high-dimensional datasets
- **High Resolution:** Don't aggregate if you can → Less data loss
- **Intuitive:** Intuitive navigation in high dimensional space
- **Responsive:** Removing the delay between forming a hypothesis and seeing the visualization → Aiding our limited working memory
- **Accessible:** Using the web with no additional installation
- **Shareable:** Easy to share **explorable** discoveries and tell a story

Why?

DISCOVERY	DISSEMINATION
Human-Centered Data Mining	Guided Tours
Curiosity-Driven Discovery	Participatory Learning
Collaborative Exploration	Accessibility

Data-Driven Decision Making

Thank You

samirpou@andrew.cmu.edu

ayahyavi@andrew.cmu.edu