

# **Explorable Visual Analytics (EVA)**

## **Interactive Exploration of LEHD**

**Saman Amraii - Amir Yahyavi**

**Carnegie Mellon University**

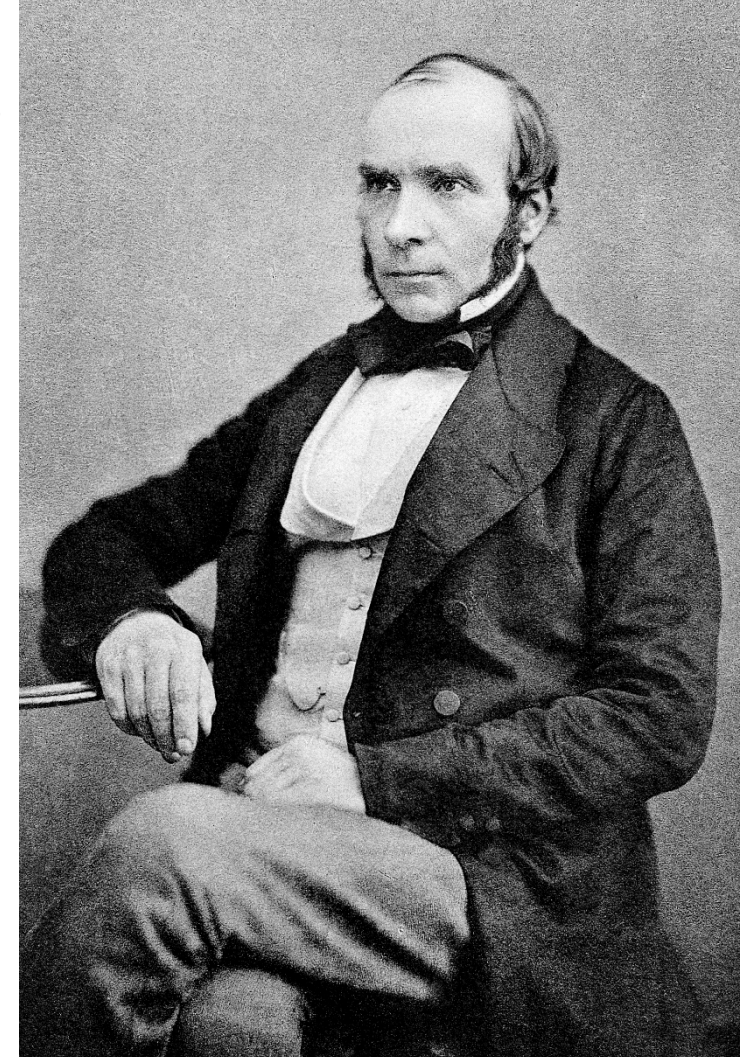
# Motivation

---



# 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

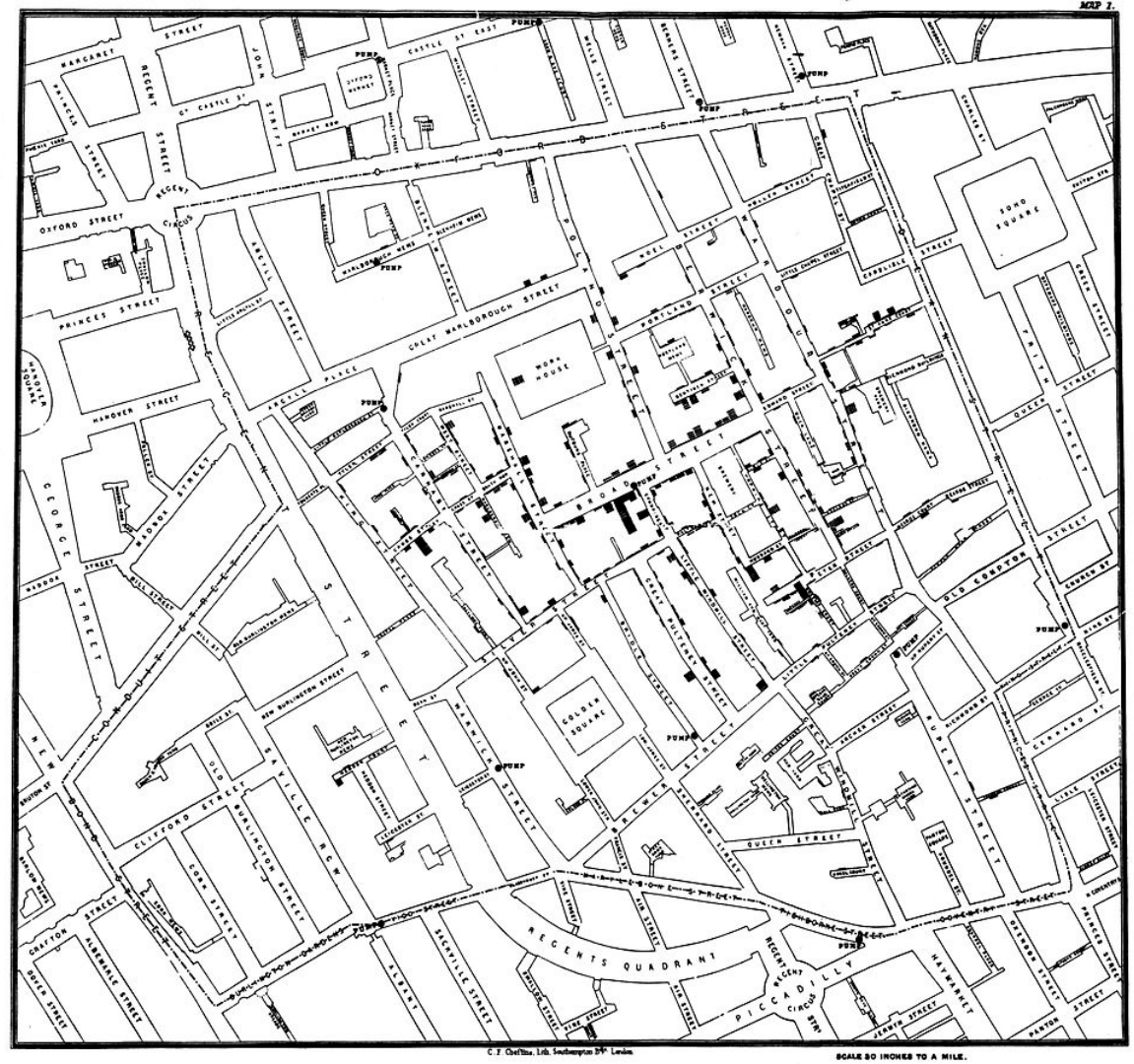


*John Snow*



# Motivation

- Drew a map
- Broad Street Pump



# Motivation

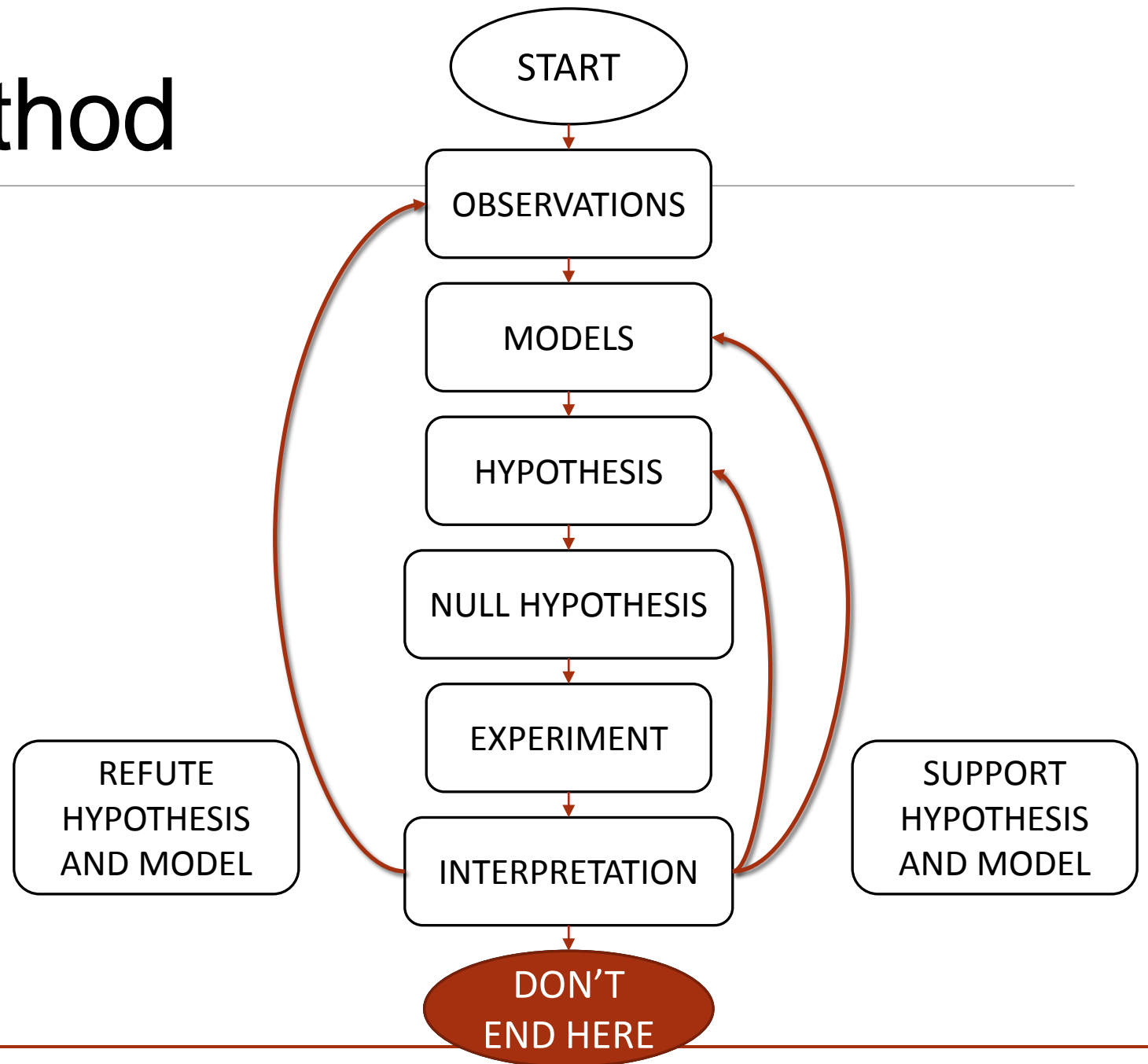
---

- 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

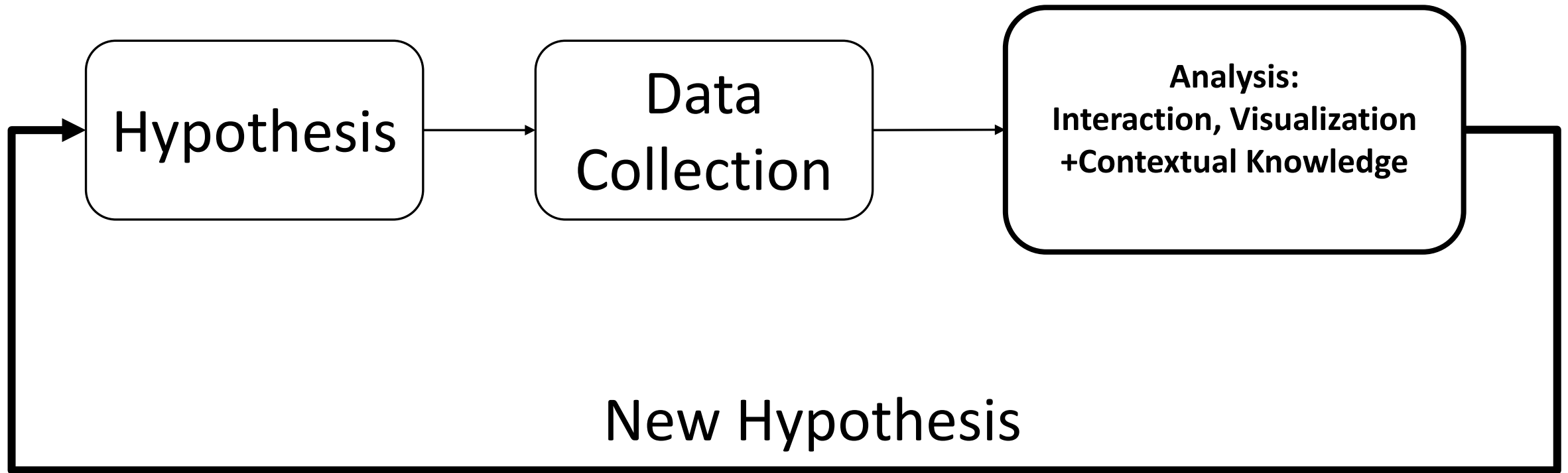


# Scientific Method



# Scientific Method

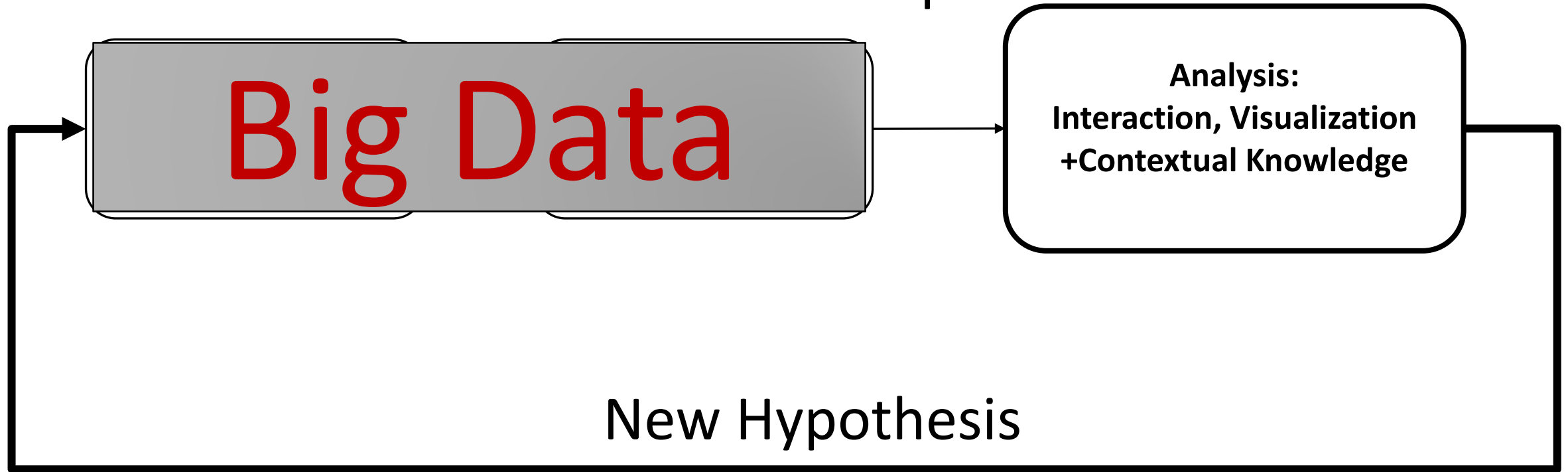
---



# Scientific Method

---

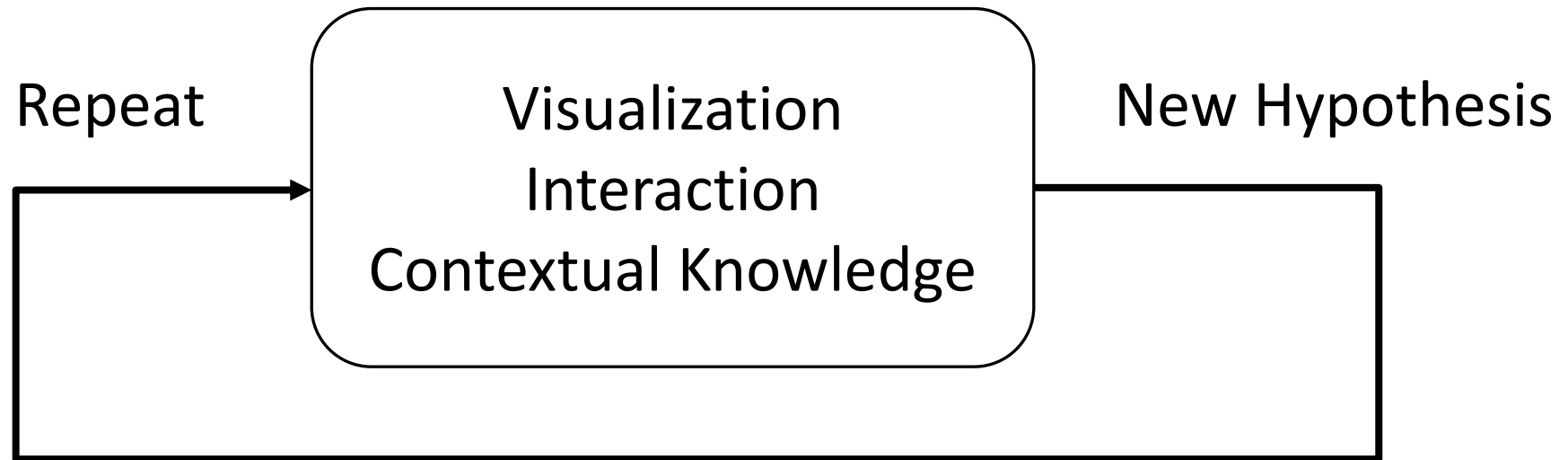
Data Production >>>>> Data Consumption





# Sensemaking Loop

---



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

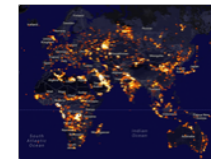
# The Explorables Collaborative

<http://explorables.cmucreatelab.org/>

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

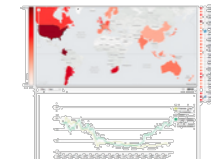
## Explorables

Explorables provide interactive and visual representations of large data sets, revealing patterns, encouraging discovery, and facilitating communication. The Explorables Collaborative, including [CREATE Lab](#) and [SkyTruth](#), is dedicated to helping you make your information more impactful. [Contact us!](#)



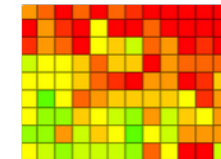
### Fires at night

Explorable visualization of a wide variety of fires detected by VIIRS over a year.



### Explorable Inequality

Explorable visualization of the World Top Incomes Database illustrating global income inequality.



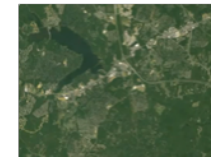
### Speck Air Quality Test

Visualization of a Speck test showing how indoor air quality is affected by vacuuming, running an air purifier, and blowing out candles.



### FIRMS Timelapse

Visualization of thirteen years of MODIS fire location vector data overlaid on an explorable global timelapse.



### Suburban Sprawl

Video exploring the growth of suburban sprawl in Chesterfield County, Virginia using TimeMachine to show the full scope of changes.



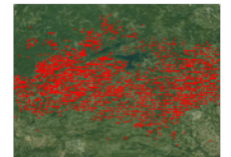
### AirNow

Explorable visualization of AirNow PM 2.5 data showing changes in particle values and air quality across the United States.



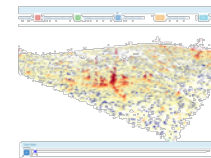
### Drilling

Time series visualization of over one million natural resource extraction wells in eight states spanning decades.



### Fracking Earthquakes

Visualization of Landsat and USGS data, utilizing TimeMachine to explore the possible connection in Arkansas between earthquakes and natural gas drilling.



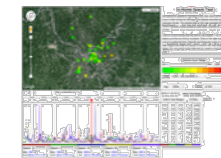
### EVA & LEHD

Multi-dimensional visualization of LEHD (Longitudinal Employer-Household Dynamics) data using EVA (Explorable Visual Analytics).



### Timelapse Story Telling

Visualization exploring landscape changes along Taiwan's coastline over two decades using TimeMachine.



### Speck In-Home Test

Visualization of air quality data collected by 64 Specks located in homes around the Pittsburgh area for a period of five days.

# Explorable Visual Analytics (EVA)

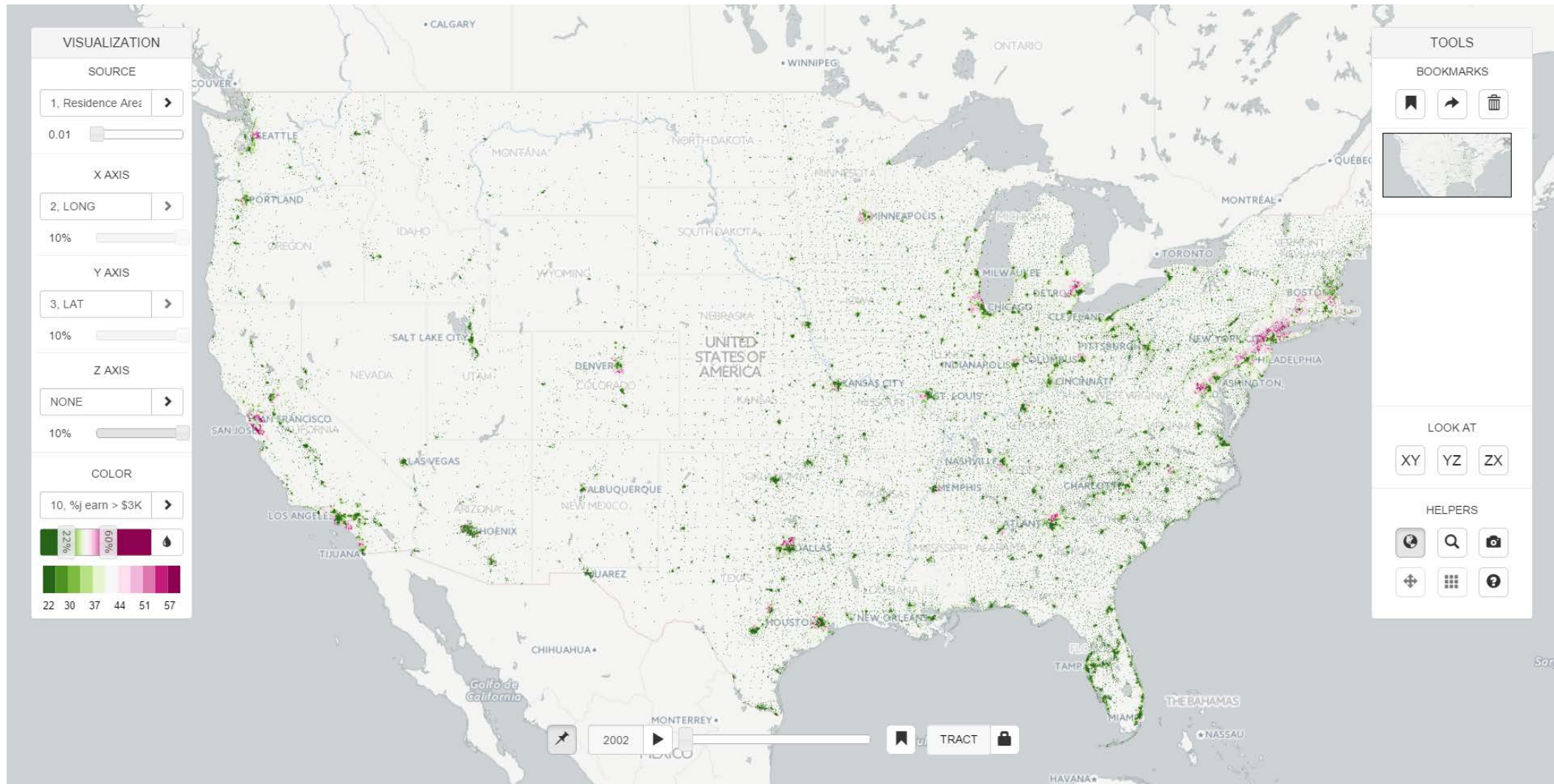
---

<http://eva.cmucreatelab.org>

## **Goal: Improving Hypothesis Generation**

- Easy Exploration
- Sharing Discoveries
- Quick Intuitions Testing

# Explorable Visual Analytics (EVA)



# EVA Demo<sup>↗</sup>

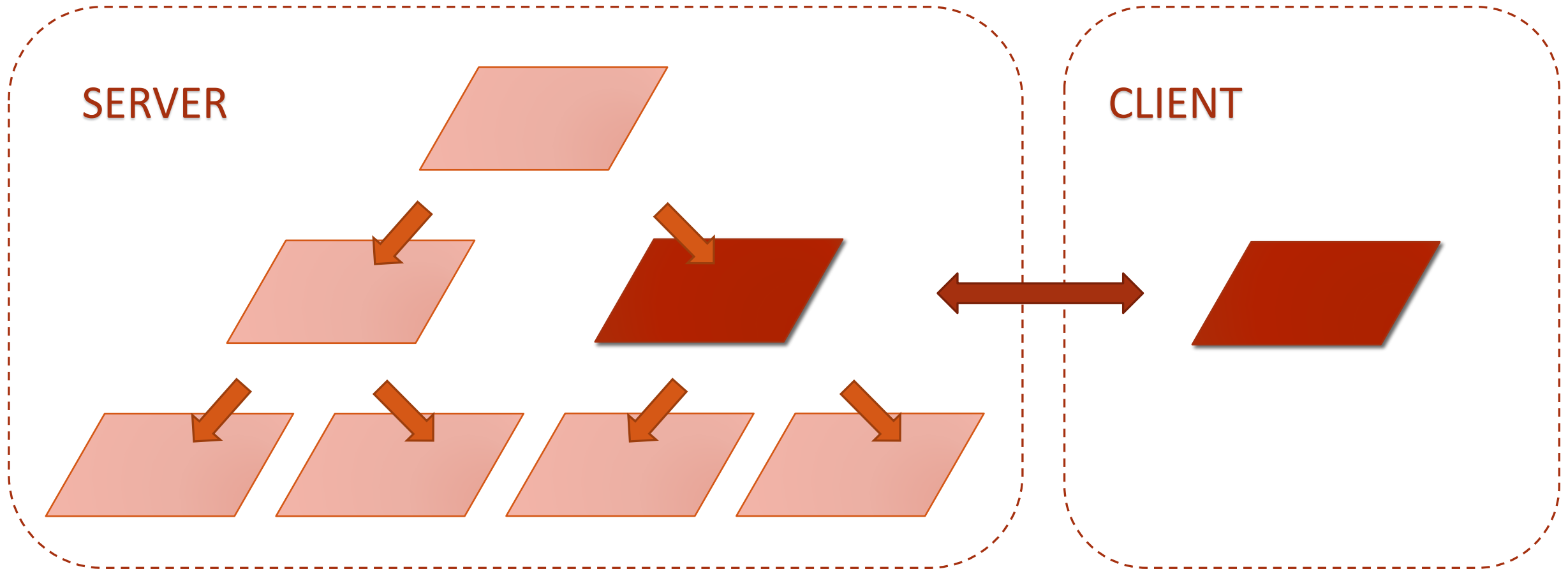
---

- **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

Human-Centered Data Mining

Curiosity-Driven Discovery

Collaborative Exploration

## DISSEMINATION

Guided Tours

Participatory Learning

Accessibility

# Data-Driven Decision Making

---

# Thank You

[samirpou@andrew.cmu.edu](mailto:samirpou@andrew.cmu.edu)

[ayahyavi@andrew.cmu.edu](mailto:ayahyavi@andrew.cmu.edu)