Explorable Visual Analytics (EVA)
Interactive Exploration of LEHD

Saman Amraii - Amir Yahyavi
Carnegie Mellon University
Motivation
Motivation

- 1854 Cholera outbreak in Soho, London
  - 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

START

OBSERVATIONS

MODELS

HYPOTHESIS

NULL HYPOTHESIS

EXPERIMENT

INTERPRETATION

REFUTE HYPOTHESIS AND MODEL

SUPPORT HYPOTHESIS AND MODEL

DON’T END HERE
Scientific Method

1. Hypothesis
2. Data Collection
3. Analysis: Interaction, Visualization + Contextual Knowledge

New Hypothesis
Scientific Method

Data Production >>>>>>>> Data Consumption

Big Data

Analysis:
Interaction, Visualization
+Contextual Knowledge

New Hypothesis
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)
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/](http://lehd.ces.census.gov/)

Contiguous US
- Hundreds of millions of data points
- Tens of dimensions
- 10 years (2002-2011)
Technical Information

SERVER

CLIENT
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](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
ayahyavi@andrew.cmu.edu