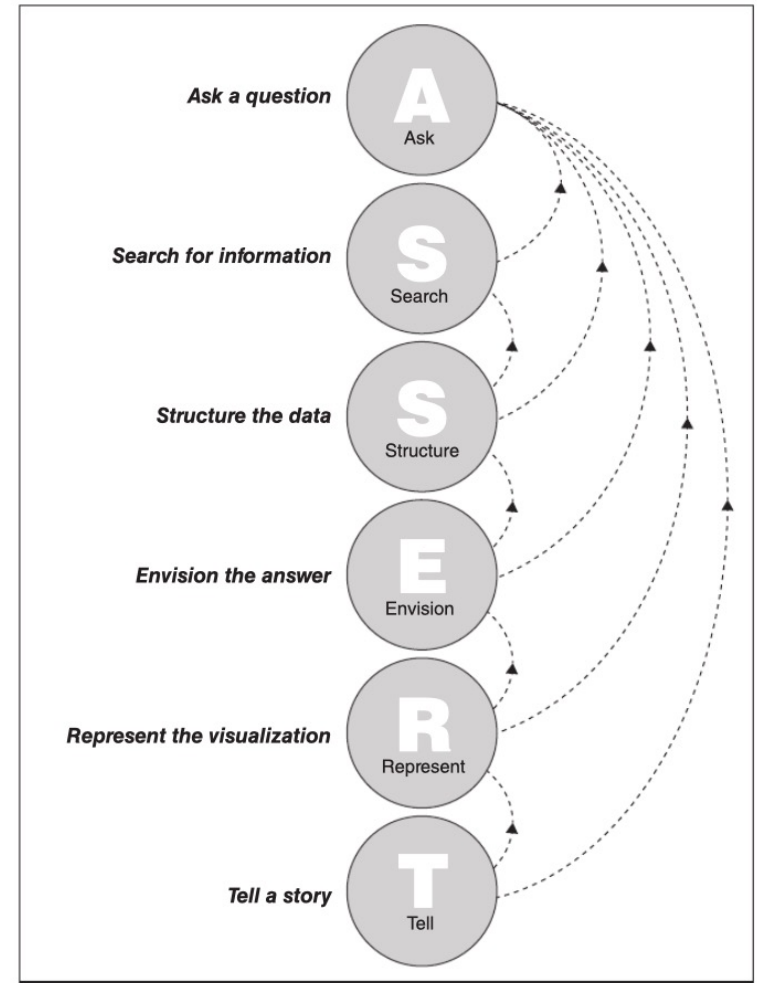


# CP321 Data Visualization

Jiashu (Jessie) Zhao

# The ASSERT model

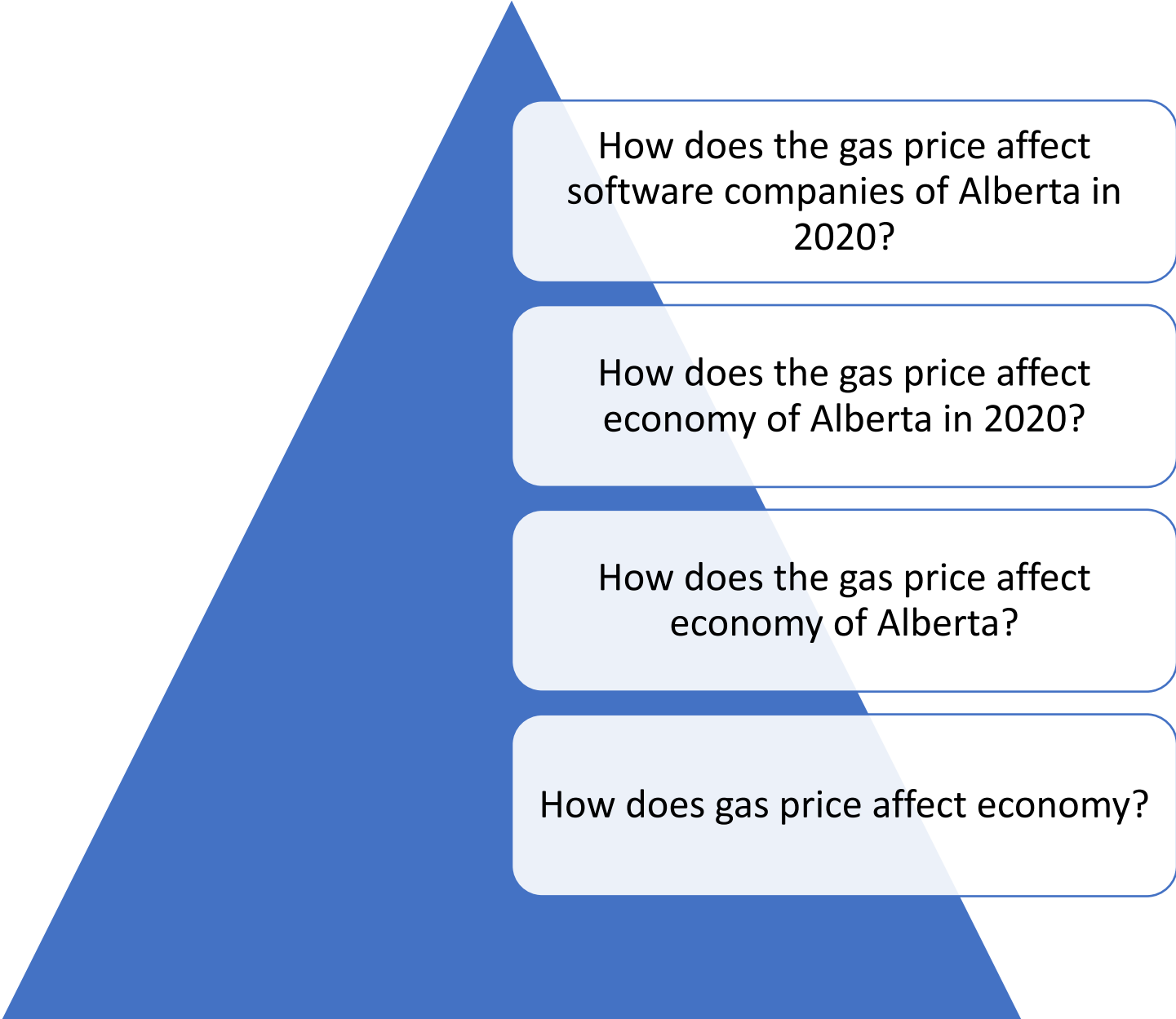
## -Ask a Question



- Good visualizations are driven by good questions.
  - The goal is to ask specific questions about the topic that can be answered by accessible and reliable **data**.
  - Questions should be able to address why you or anyone would care.
  - The questions should guide the subsequent phases and be used as a test for inclusion or rejection of a source or method.

- Defining the Audience: it is important to have a good understanding of who its audience will be.
  - Audience Needs and Capabilities
  - Their Background Knowledge
  - The Information They Want to Know
  - Their Content Sensitivity
  - Their Availability of Technology

- Narrowing the Question ' s Scope: It is critical to narrow the scope of the topic so that a meaningful (and presumably answerable) question can be found
  - Time
  - Location
  - People
  - other attributes that might constrict the topic
- On the other hand, a too focused topic may not have enough information to support



How does the gas price affect software companies of Alberta in 2020?

How does the gas price affect economy of Alberta in 2020?

How does the gas price affect economy of Alberta?

How does gas price affect economy?

# Turning Topics into Questions

- Extract specific questions that are suitable for visualization
- In general, questions that offer more complex answers than a simple “ yes ” or “ no ” tend to be more interesting pursuits.

“ Was slavery the cause  
of the Civil War? ”



“ What degree of  
economic advantage  
did slavery offer the  
South to maintain it? ”

- Strategies for methodically generating and assessing the quality of questions
  - The three-part query
  - K-W-L



- Strategies for methodically generating and assessing the quality of questions

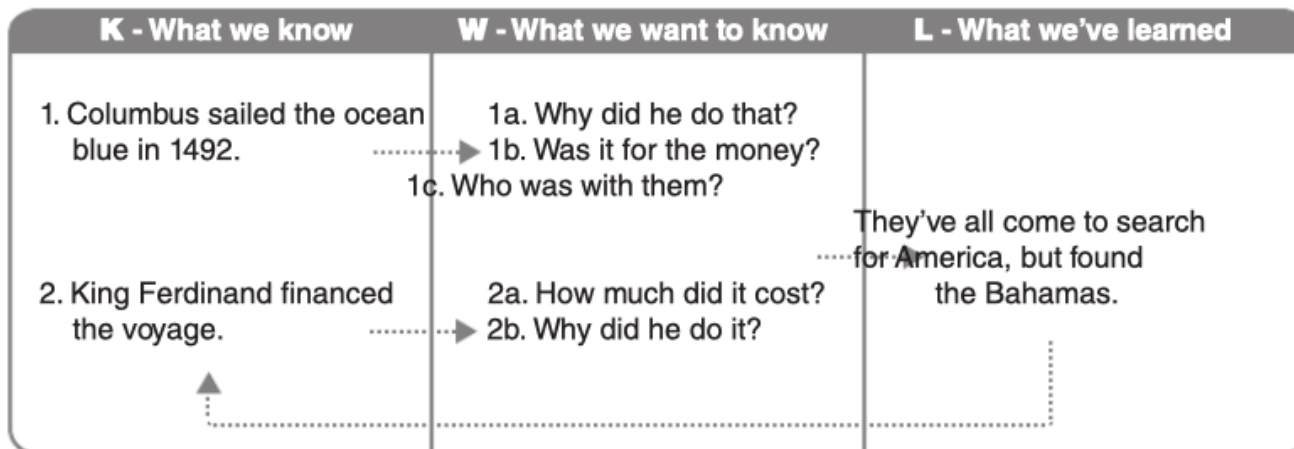
- The three-part query

- identifying a focused **topic** , the **questions** to ask about information needed to explore, and the **reason** to care about exploring it
- We are looking at [ *topic* ] because we want to find out [ *who/where/when/why/what/how* ] in order for my audience to understand [ *significance* ].

Example: We are looking at [the University of Virginia 1828 library](#) to find out [where the books were published](#) in order for my audience to see [the geographic spread of eighteenth century scholarship](#). (Booth et al., 2003)

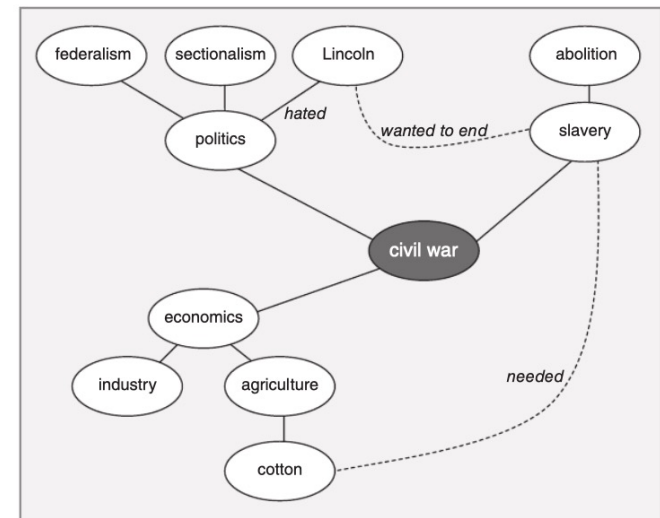
- K-W-L

- Strategies for methodically generating and assessing the quality of questions
  - The three-part query
  - K-W-L
    - what we already **K**now about the topic
    - what we **W**ant to learn about the topic
    - what we have **L**earned



# Creative Problem-Solving Techniques

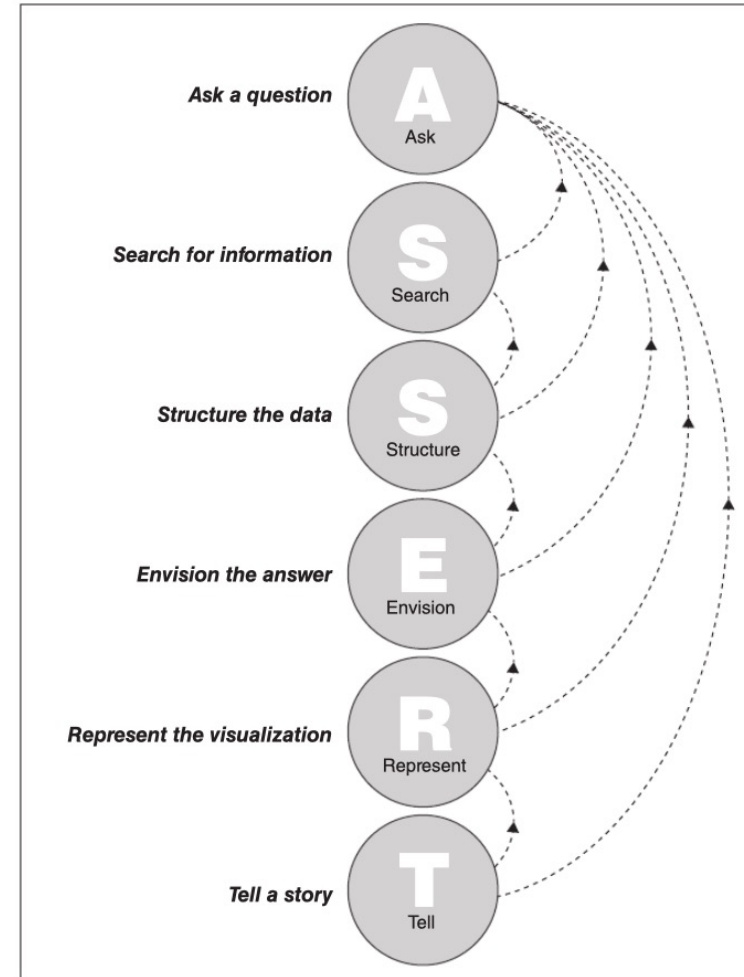
- Creative Problem-Solving (CPS) is to generate ideas (options) and then systematically evaluates the options.
- The CPS framework begins by identifying the broad goal in order to understand the problem to be solved – through data exploration and problem framing
- CPS tools: Brainstorming, Group discussion, Concept Map



# Generating Questions from Questions

- The best source of new questions about the topic may come from questions already asked.
  - drilling down deeper into the question
  - examining the question from multiple perspectives
    - i.e., northerners, southerners, slaves, and slave-holders
  - asking the question from an opposite point of view

# Search for Information

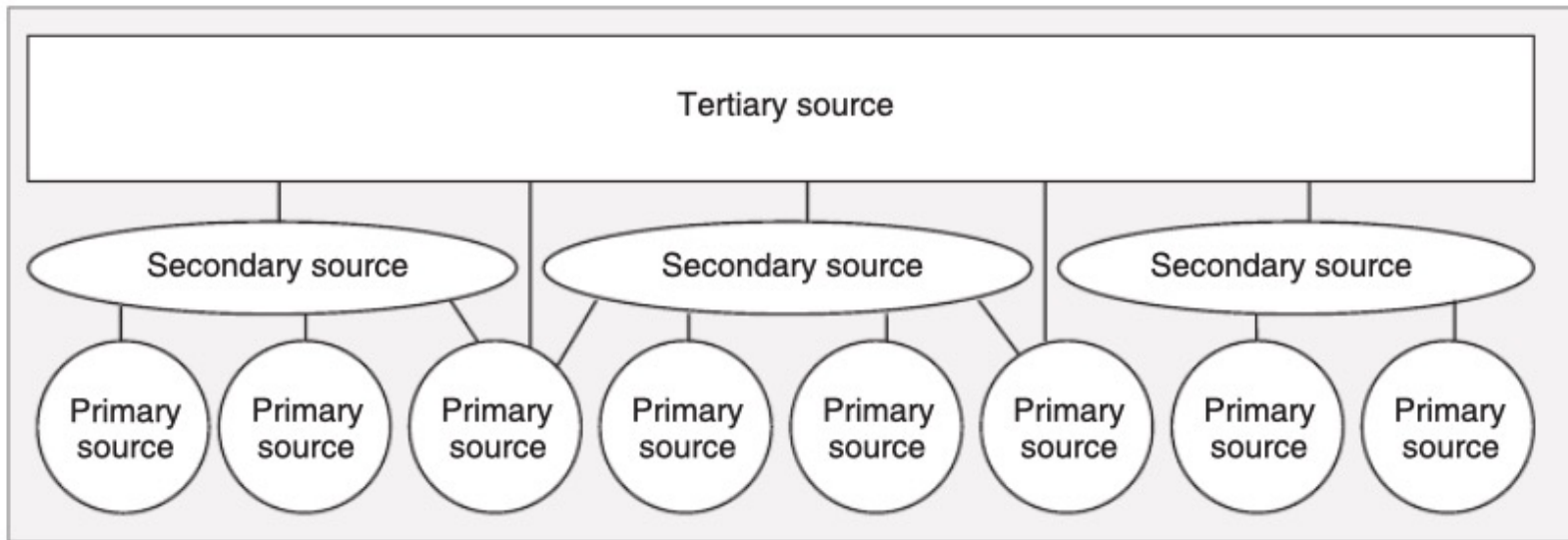


# Types of Raw Data

- *Numbers*: such as populations and rainfall amounts
- *Documents*: such as court records, birth and death records, letters, diaries and other personal writings,
- *Multimedia*: photographs, drawings, and audio and video recordings used in oral histories, news coverage, and documentaries.

# Types of Data Sources

- Primary sources are data taken directly from the event, such as a first-hand witness account or temperature readings.
- Secondary sources are books and other resources that aggregate the information from primary sources, and finally,
- Tertiary resources such as encyclopedias pull from both primary and secondary sources to provide broad but shallow overviews of a topic.





- Successful visualizations depend on the availability of effective and sufficient sources of data to draw on as evidence
- Secondary and even tertiary sources are a good starting point for finding relevant primary sources
- Many places to find data: libraries, museums, historical societies, courthouses, public records, etc.

# Effective Quantitative Data

- Not all data are useful enough to be included in an information visualization.
- Effective Data: more likely to yield insights when visualized
  - High volume
  - Historical
  - Consistent
  - Multivariate
  - Atomic
  - Clean
  - Structured
  - Sourced
  - Accessible
  - Representative

# Project Data Sets

- <https://open.canada.ca/en/open-data>
- <https://data.ontario.ca/>