

Understanding the Role of Alternatives in Data Analysis Practices

Jiali Liu, Nadia Boukhelifa, James R. Eagan

IEEE Trans. Visualization and Computer Graphics (Proc. InfoVis 2020)
doi: 10.1109/TVCG.2019.2934593

SPEAKER Francis Nguyen DATE November 12, 2019

Understanding the Role of Alternatives in Data Analysis Practices

Jiali Liu, Nadia Boukhelifa, James R. Eagan

IEEE Trans. Visualization and Computer Graphics (Proc. InfoVis 2020)
doi: 10.1109/TVCG.2019.2934593

SPEAKER Francis Nguyen DATE November 12, 2019

↑
Data sources, Tools,
Methods, Analysis,
Visual designs ...

Raise your hand if...

Raise your hand if...

You have analyzed data

Raise your hand if...

You have analyzed data
Used more than one dataset?

Raise your hand if...

You have analyzed data
Used more than one dataset?
Visualized your data in more than one way?

Raise your hand if...

You have analyzed data
Used more than one dataset?
Visualized your data in more than one way?
Wondered what others have done?

Raise your hand if...

You have analyzed data
Used more than one dataset?
Visualized your data in more than one way?
Wondered what others have done?
You consider yourself a **data scientist**

Data Workers & Alternatives

- People who come from a variety of domains and perform data analysis as part of their daily work
- But don't consider themselves data scientists

The iterative sense-making process

- Exploratory data analysis
- Open-ended, actual steps are selected segments of a branching, tree-like pattern of possible actions

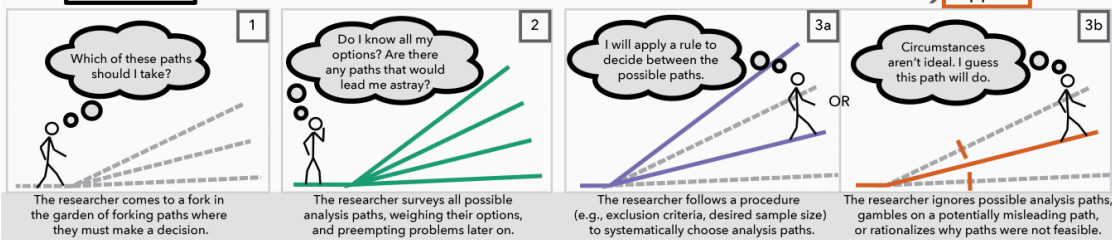


Fig 1. Kale et al. 2019, Decision-Making Under Uncertainty in Research Synthesis: Designing for the Garden of Forking Paths

Exploration of many alternatives

- What do we mean by alternatives?
- How do data workers consider and explore alternatives in their real analytic practices?


Objective

Understand **alternatives** used by data workers

4 Research Questions to Understand Alternatives

- Q1: To what extent do data workers explicitly consider alternatives in their workflow?
- Q2: When do they consider alternatives? Are there specific **triggers & barriers** for exploring alternatives?
- Q3: What **types of alternatives** are considered?
- Q4: What **strategies** are deployed to cope with alternatives?

Contributions



Semi-structured interviews and analysis to **understand and define** alternatives

Semi-structured Interviews

Semi-structured Interviews

Interviews with 12 data workers from various domains

- Participants:
- Project management, HCI research, humanitarian, marketing, education, topological research
 - Ranges in expertise – 9 ‘novices’, 3 ‘experts’
 - Diverse levels of domain and computational experience

- Other Details:
- In-situ or teleconference
 - 50 mins - 2 hours in length
 - Voice recorded + photos

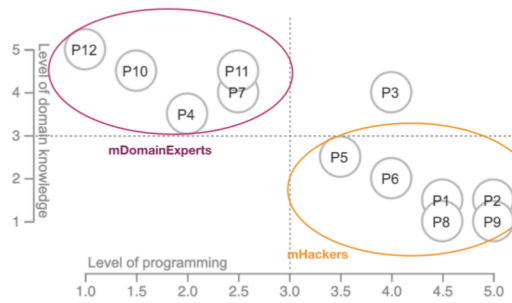
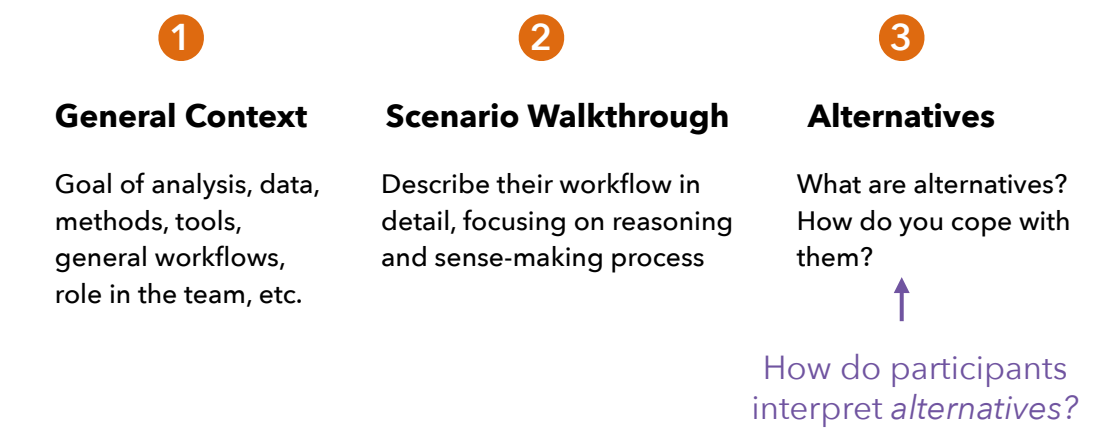


Fig 1. Liu et al. Understanding the Role of Alternatives in Data Analysis Practices.

Participants self-reported level of expertise in terms of domain knowledge and computational skill.

Semi-structured Interviews – 3 Phases



17

Data Collection & Analysis

- Recorded **827** mins of audio
- Transcribed in **73,895 words**
- Open-coding process cross-checked for **585 extracts**

Affinity Diagrams Analysis + Process Workflow Analysis

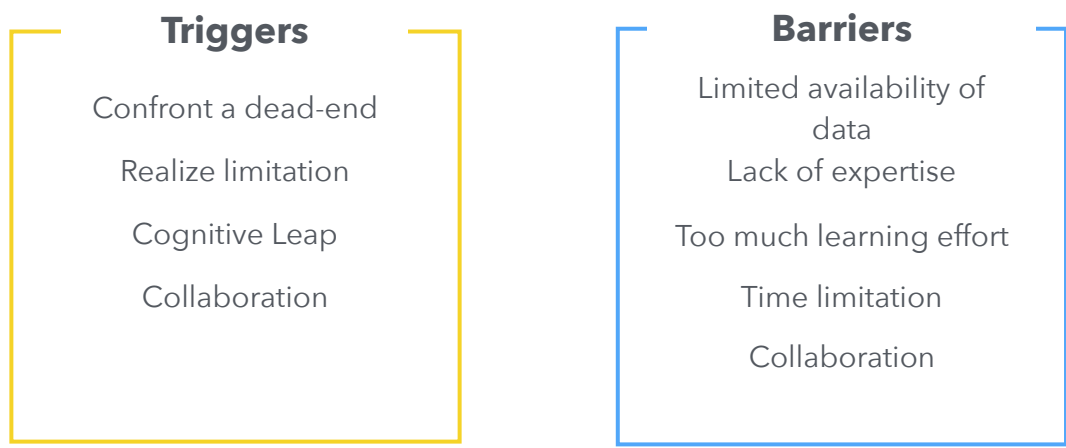


Findings – Reasons

- Clarify goals and processes
- Delay decision making
- Build confidence in a solution
- Partition the sense-making workload

20

Findings – Triggers & Barriers



21

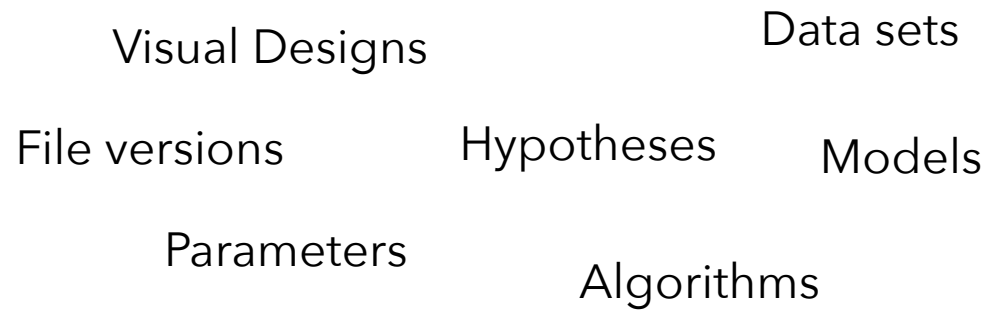
Alternatives

"Encompass multiple iterative versions of the same artifact or refined versions of a given hypothesis or altogether distinct methods to analyze data"

22

Alternatives

"Encompass multiple iterative versions of the same artifact or refined versions of a given hypothesis or altogether distinct methods to analyze data"

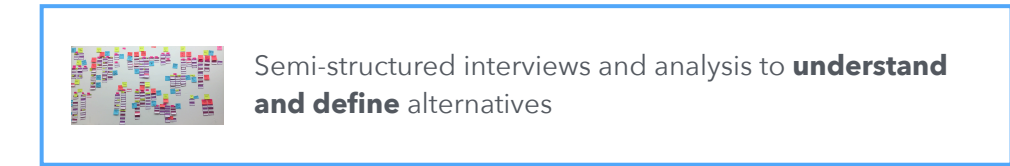


23

Findings – Strategies

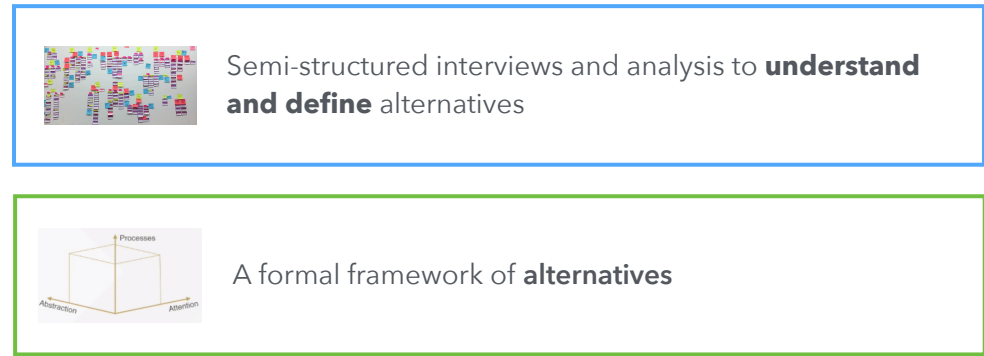
- Depth-first* (concentrate on a given alternative, consider others when needed)
- Breadth-first* (generate many multiples eventually focusing on one choice)
- Generating, update and reduce* alternatives using external resources

Contributions



25

Contributions



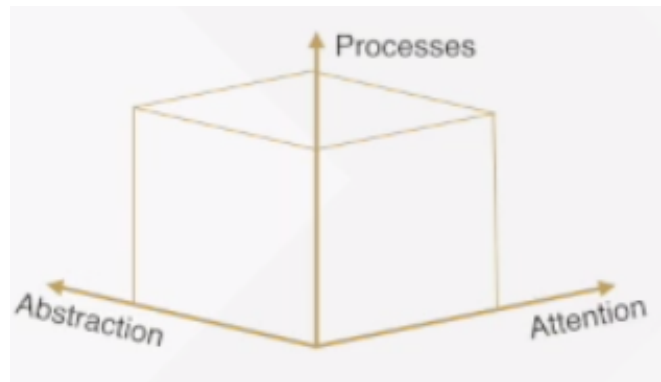
26

A Framework of Alternatives

Degree of attention

Abstraction level

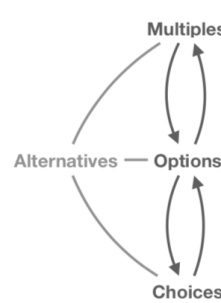
Analytic processes



Liu et al. Screenshot from VAST 2019 Presentation

27

Framework – Degree of Attention



Multiples: possibilities data workers are aware of, but no attention is put on one specific one

Options: multiples brought into attention for closer inspection

Choices: options actively pursued in analysis at a given moment in time

Fig 2: Liu et al. Understanding the Role of Alternatives in Data Analysis Practices.

28

Framework – Degree of Attention

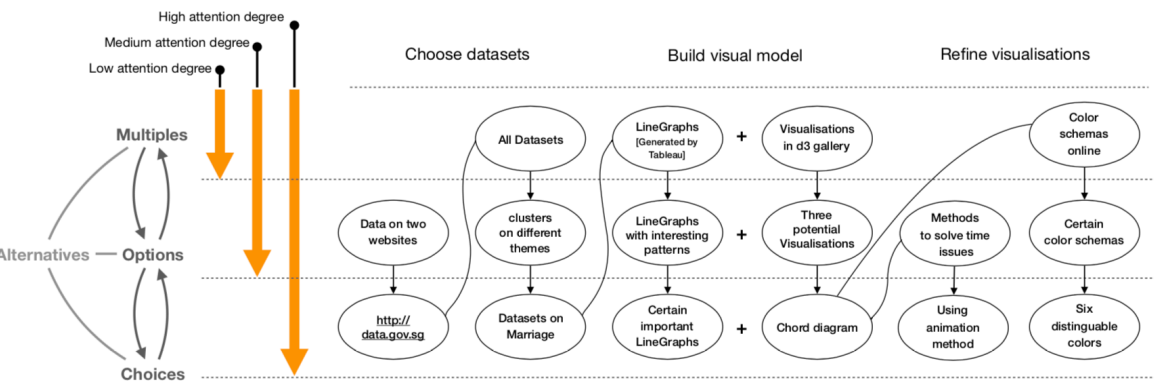


Fig 2: Liu et al. Understanding the Role of Alternatives in Data Analysis Practices.

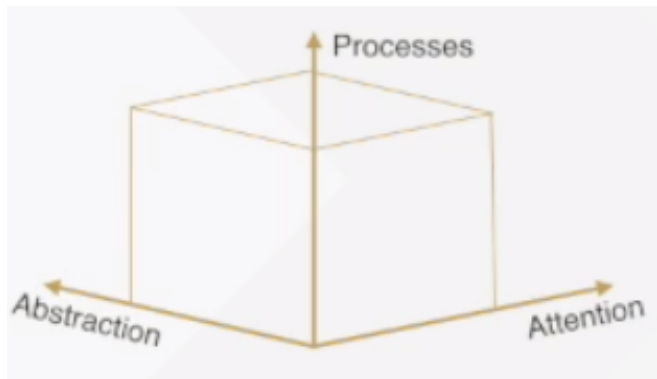
29

A Framework of Alternatives

Degree of attention

Abstraction level

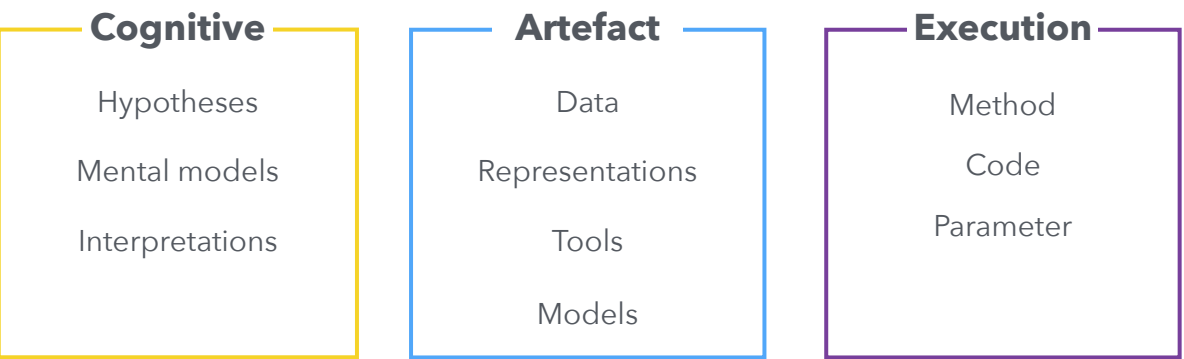
Analytic processes



Liu et al. Screenshot from VAST 2019 Presentation

30

Framework – Abstraction Level of Alternatives



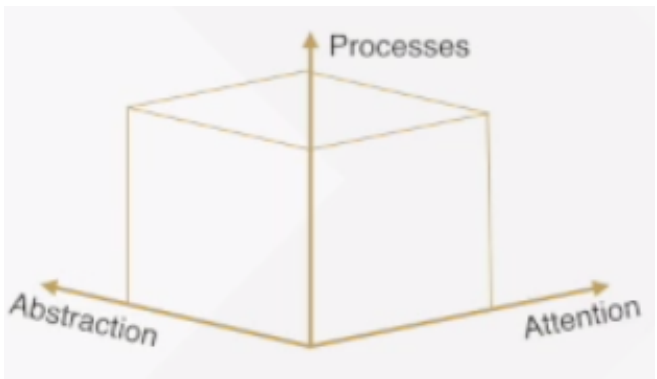
31

A Framework of Alternatives

Degree of attention

Abstraction level

Analytic processes



Liu et al. Screenshot from VAST 2019 Presentation

32

Framework – Processes

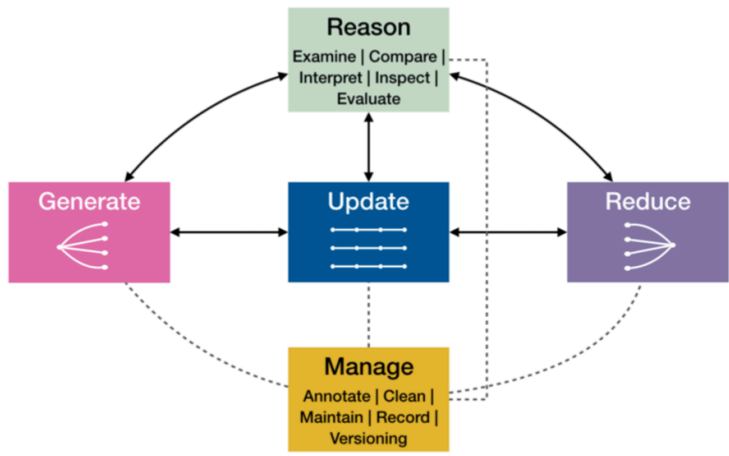


Fig 3. Liu et al. Understanding the Role of Alternatives in Data Analysis Practices.

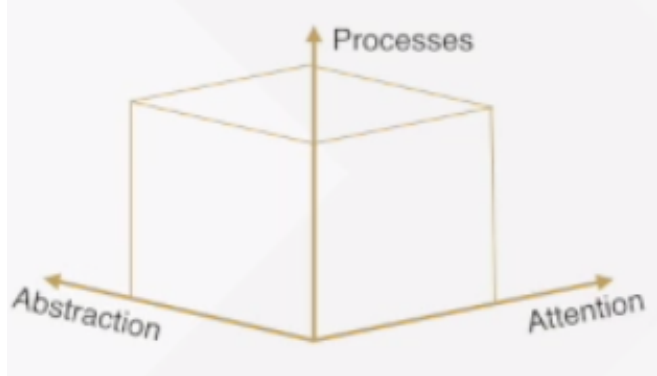
33

A Framework of Alternatives

Degree of attention

Abstraction level

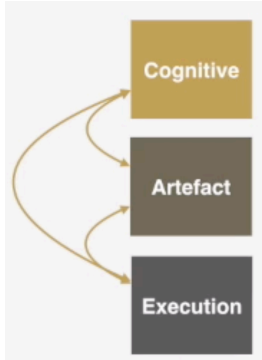
Analytic processes



Liu et al. Screenshot from VAST 2019 Presentation

34

Implications for Analysis Tools



Liu et al. Screenshot from VAST 2019 Presentation

Analysis tools lack of have limited support for alternatives at the cognitive layer

Analysis environments break the chain of alternatives across different abstraction levels

Limitations –

- **Generalizability** – Only 12 data workers analyzed
- **Limited group of individuals** – what about other data workers with other areas of expertise?
- Process seems **highly linked** to the set of tools and domains

35

Thanks! Questions?

Understanding the Role of **Alternatives** in Data Analysis Practices

Jiali Liu, Nadia Boukhelifa, James R. Eagan

IEEE Trans. Visualization and Computer Graphics (Proc. InfoVis 2020)
doi: 10.1109/TVCG.2019.2934593

Slides adapted from Liu et al. [VAST 2019 Presentation](#)

37