Visual (dis)Confirmation: Validating Models and Hypotheses with Visualizations

In Kwon Choi, Nirmal Kumar Raveendranath, Jared Westerfield, and Khairi Reda
School of Informatics and Computing
Indiana University-Purdue University Indianapolis, Indiana, USA
Email: {inkwchoi, niraveen, jartwest, redak}@iu.edu

***Abstract*—Data visualization provides a powerful way for analysts to explore and make data-driven discoveries. However, current visual analytic tools provide only limited support for hypothesis-driven inquiry, as their built-in interactions and workflows are primarily intended for exploratory analysis. Visualization tools notably lack capabilities that would allow users to visually and incrementally test the fit of their conceptual models and provisional hypotheses against the data. This imbalance could bias users to overly rely on exploratory analysis as the principal mode of inquiry, which can be detrimental to discovery. In this paper, we introduce *Visual (dis)Confirmation*, a tool for conducting confirmatory, hypothesis-driven analyses with visualizations. Users interact by framing hypotheses and data expectations in natural language. The system then selects conceptually relevant data features and automatically generates visualizations to validate the underlying expectations. Distinctively, the resulting visualizations also highlight places where one's mental model disagrees with the data, so as to stimulate reflection. The proposed tool represents a new class of interactive data systems capable of supporting confirmatory visual analysis, and responding more intelligently by spotlighting gaps between one's knowledge and the data. We describe the algorithmic techniques behind this workflow. We also demonstrate the utility of the tool through a case study.**

***Index Terms*—Visual analytics, hypothesis-driven reasoning, sensemaking**

I. Introduction

Visualization plays an increasingly important role in supporting data-driven science and decision making [1], [2]. Visualization tools enable people to interactively explore large amounts of information and look for patterns that might suggest new findings. These tools effectively facilitate a *bottom-up* discovery process, where apparent signals in the data are interpreted in a new way yielding unexpected insights.

Yet, in addition to engaging in exploratory analysis (bottom-up), analysts also conduct confirmatory analyses (top-down), where they explicitly test their beliefs and predictions against the data [3]. In fact, analysts often switch between these two modes of discovery during sensemaking [4]. However, while current visualization tools have built-in workflows to scaffold exploratory analysis, they provide no affordances for users to explicitly test hypotheses and models they have in mind. Visualization designers almost exclusively focus on supporting data-driven tasks (e.g., browsing clusters or finding anomalies [5]). The most common visualization design patterns (such as Shneiderman's "overview first" paradigm [6]) emphasize the data as a starting point, but neglect the role that existing user hypotheses and mental models might play into the analysis. Accordingly, the prevailing theory of visual analytics is that of a "model building" activity [7], where mental or formal models are seen as products that are constructed from data in an almost purely bottom-up fashion.

The overemphasis on data-driven tasks in visual analytics serves to privilege exploratory analysis as the principal mode of discovery, while thoroughly understating the need for confirmatory analysis. This imbalance can be dangerous to discovery [8], as it prevents people from explicitly testing their beliefs, leaving them with potentially faulty models [3]. Crucially, cognitive science shows that it can be difficult for one to recognize 'holes' and deficiencies in their conceptual models just by looking at data in a bottom-up fashion [9]. Instead, to repair one's model and ascertain a better understanding, **one needs to actively test their belief against the data**, and explicitly look for places where their models and the data disagree [3]. Successful scientists naturally employ confirmatory workflows [10]. However, techniques for highlighting model-data discrepancies—so as to encourage users to attend to them—are absent from visualizations.

To support a richer and more diverse analytic discourse, visualization tools must allow analysts to proactively share and externalize their models. This can be done using natural language (e.g., by speaking or typing hypotheses and data expectations directly into the interface). The system can then interpret these formulations, select relevant data attributes and features, and visualize the fit of these models to the data. Such workflow would allow users to interactively test their predictions, identify deficiencies in their models, and actively revise those models to reconcile mismatches.

In this paper, we describe an example of such a visualization tool, which we dub *Visual (dis)Confirmation*. The tool employs natural language processing (NLP) to parse hypotheses and extract the implied data relations, trends, and values. It then interactively generates pertinent visualizations in the form of familiar data charts. Encoded in these charts is a visual representation of the user's expectation alongside the actual data. This enables the user to validate or disconfirm hypotheses they might have, by visually relating the two encodes. We describe the design and implementation of the system. We then illustrate through a case study how the system might be used in a realistic analysis. We conclude with a discussion of challenges and future research directions.

978-1-7281-2850-4/19/$31.00 ©2019 IEEE
DOI 10.1109/IV-2.2019.00032

II. Related Work

Visual analytics tools are primarily intended to support sensemaking. Accordingly, these tools are often designed to mimic the natural workflow of analysts. Several models exist to explicate how people analyze and make sense of information. One of the most popular is Pirolli and Card's, which posits that people generally start by filtering the source data for relevant information, highlighting nuggets of evidence, and re-expressing that evidence to a way that aids reasoning [4]. In this model, the analysts iteratively funnels the data into increasingly sparser and more structured representations (referred to as the 'Schema'). From the Schema, the analyst can more easily generate hypotheses or makes decisions. While the model allows for feedback, it is generally regarded as bottom-up and data-driven sensemaking.

Pirolli and Card's model serves as the design basis for many visualization tools (e.g., Jigsaw [11]). Yet, an equally important (but less known) model is Klein et al's Data-Frame theory, which posits that, when analysts make sense of data, "they often begin with a perspective, viewpoint, or framework—however minimal" [3]. This initial "frame", which can take the form of a narrative, timeline, or hypothesis, defines the main relationships one expects to see in the data. Here, sensemaking is primarily a confirmatory activity: the analyst iteratively questions his/her frame by testing its fit against the data. Poor fit can lead one to revise the frame or, alternatively, adopt an entirely new frame.

It is believed that analysts often mix the two types of workflows, switching between data-driven (bottom-up) and confirmatory (top-down) analyses [4]. However, most visualization tools have been designed to solely support the former. Such designs can be problematic, as they could discourage analysts from deliberately testing their expectations, which is essential to refining one's model [12]. For instance, in cognitive experiments, subjects who did not explicitly test their working hypothesis failed to correct their mental model and missed the chance of discovery [9]. By contrast, those who frequently tested their predictions against the data were able to attend to discrepancies and, accordingly, revise their model to reach the correct conclusion. Unfortunately, no such confirmatory workflows exist in current visualization tools—a missed opportunity. Our work aims to fill this gap.

Another area that we build upon is natural language interfaces, which represents an emerging method for interacting with visualizations [13]. There are tools now that allow users to speak or type their queries and, accordingly, receive pertinent data plots [14]–[16]. In addition to communicating queries, natural language interfaces may also allow us to tap into users' mental models. Understanding what users are thinking enables us to create more effective (and potentially confirmatory) visualizations. We contribute techniques to parse hypotheses and data expectations, and convert them to concept graphs so that they can be processed algorithmically. These techniques are based on an empirical study we conducted to inform the design of the system [17].

III. System Design

In this section, we introduce Visual (dis)Confirmation, an interactive tool supporting confirmatory visual analysis. The tool allows users to specify hypotheses and expectations about data in natural language. It then translates these hypotheses to concept maps, and generates visualizations to validate (or disconfirm) the implied data relations. We describe the user interface and discuss the implementation of the system.

A. User Interface

Fig. 1 illustrates the user interface. The tool allows the user to "upload" a dataset in CSV format. Alternatively, the user can choose from two default datasets about world development and health risk factors in the US **(1)**. The user initiates the analysis by typing an expectation or hypothesis in natural language **(2)**. For instance, when analyzing economic development data, the user can enter "I expect New Business Density in Ireland to be lower than France, Australia, Netherlands Sweden, and Spain." To aid the user in specifying data attributes, the interface auto-completes words corresponding to attributes and provides suggestions. Additionally, the interface auto-completes geographic locations (e.g., country and city names) and other named entities within the dataset. A map in the form of a rotating globe is also included to facilitate geo-referencing **(3)**, enabling the user to easily specify expectations about localities. The interface also remembers past expectations and their resulting visualizations **(4)**, giving the user quick access to his/her analysis history.

The system parses the user's natural language hypothesis (see Section III-B for a description of the NLP pipeline), and generates a corresponding concept map. The concept map encodes the expected data relations using a node-link diagram **(5)**; nodes denote attributes, geographic locations (e.g., countries), temporal features (e.g., years or full dates), or numeric quantities. Different node shapes (ellipses, rectangles, or clouds) are used to distinguish between different concept types (attributes, named entities, or entities that are external to the dataset). Labeled edges connect the concepts explicating the relationships expected among them. For instance, in the above hypothesis, the attribute of interest ("New Business Density" is depicted with an ellipsoid node, whereas the different countries (Ireland, France, etc...) are each depicted with rectangular concept nodes. The implied relationship (inequality in this example) is expanded and encoded as edges between the countries involved. Note that in this particular example the expected relation is qualified to a single year, hence the conditional edges in the concept map (labeled 'in').

The concept map constitutes a re-expression of the user's natural language hypothesis in a non-ambiguous format. While the two representations (natural language and concept map) are somewhat synonymous, our goal in including the latter is to provide feedback to the user on how his/her hypothesis was interpreted by the system. Moreover, the node-link concept diagram is editable, allowing the user to revise his/her formulation or correct errors introduced by the NLP process. It is

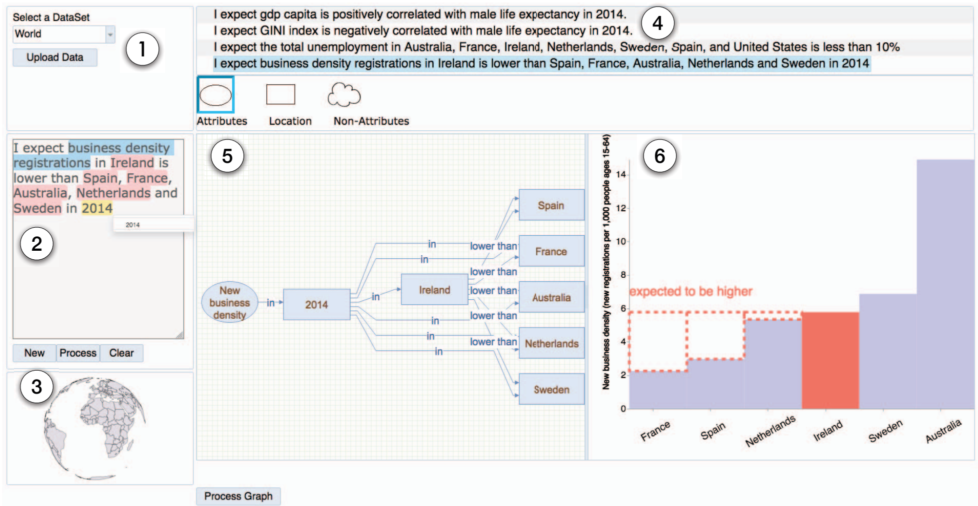

Fig. 1. User interface of the *Visual (dis)Confirmation* tool showing its various components: Dataset upload and selection (1), a text box for hypothesis/expectation entry in natural language (2), a map widget to aid the selection of geographic locations (3), history of prior queries (4), a concept map representing the implied relationship in the entered expectation (5), and a chart area to display the resulting visualization (6).

also possible for the user to specify hypotheses manually by creating the corresponding concept maps from scratch.

Once the user is finished specifying a hypothesis, he/she clicks the "Process Graph" to initiate the validation process (see Section III-C). The result of the validation is a data chart visualizing the attributes referenced in the expectation **(6)**. Distinctively, charts generated with *Visual (dis)Confirmation* explicitly incorporate what the user expects into the visualization, highlighting the gap between the expectation and the data. For instance, in the bar chart shown in Fig. 1, Ireland is highlighted. Furthermore, countries that violate the expected inequality (France, Spain, and the Netherlands, in this instance) are spotlighted (see Fig. 1).

B. *Expectation Parsing and Concept Map Generation*

The first step in validating user's hypotheses is to parse their data expectations and generate an equivalent concept map. We developed an NLP pipeline using the Stanford CoreNLP toolkit [18]. The pipeline is depicted in Fig. 2 and illustrated through an example. The first step is cleaning the provided expectation (Fig. 2-A). This includes expansion of abbreviated words ("I'd" to "I would") to ensure grammatical completeness, removal of special characters (such as quotes), and transformation of quantitative units (e.g., '3 million') to their numeric equivalent (e.g., 3,000,000). These steps are achieved using a Named Entity Recognizer [18].

Second, after basic transformations and cleanups, the text is parsed using a Stanford Dependency Parser to extract the relationships among the words. The dependency parser allows us to find compound nouns, which often correspond to data attributes or geographic locations. Words belonging to a single compound noun are concatenated to facilitate matching against name entities in the dataset (B).

Third, the sequence is tokenized into *n-grams*. The resulting tokens are checked for exact match against features in the dataset, including attributes and geographic locations (C). If no exact matches are found, cosine and semantic similarity is performed over all tokens and against all named entities in the dataset. Matching tokens are added as concept nodes (D).

Using a typology developed from an earlier empirical study [17], we identify the data model implied by the expectation. The system currently supports 4 major types and 13 subtypes (depicted in Fig. 3). Each model is represented by a template containing a number of 'slots' that can be filled with specific attributes, locations, time periods, and other qualifiers, as specified by the user's hypothesis. For a given expectation, we identify the most closely related template by inspecting the stem words along with their position in the sentence (E). For instance, the expectation in Fig. 2 can be matched against the *V1* template (in Fig. 3). We employ template-specific heuristics to fill template slots with the corresponding concept nodes. Here, the `X` slot in *V1* is mapped to 'total population', `V` is mapped to 3 million, the geographic location is mapped to Ireland, the time period to 2008—2012, and the inequality is set to 'less'. Note that the system does not require user input to match the exact wording of the template. Rather, template matching is performed through a process of constraint satisfaction, which provides verbal flexibility while allowing the selection of the closest model. Once the template is 'filled', relationship edges are added to the concept map according to the model implied by the template (F).

The pipeline relies on user hypotheses (partially) matching one or more predefined templates in our typology (Fig. 3). In practice, the templates capture over 91% of the expectations we have observed empirically [17]. Thus, despite the assumptions, we expect the pipeline to correctly process a majority of expectations users may want to test.

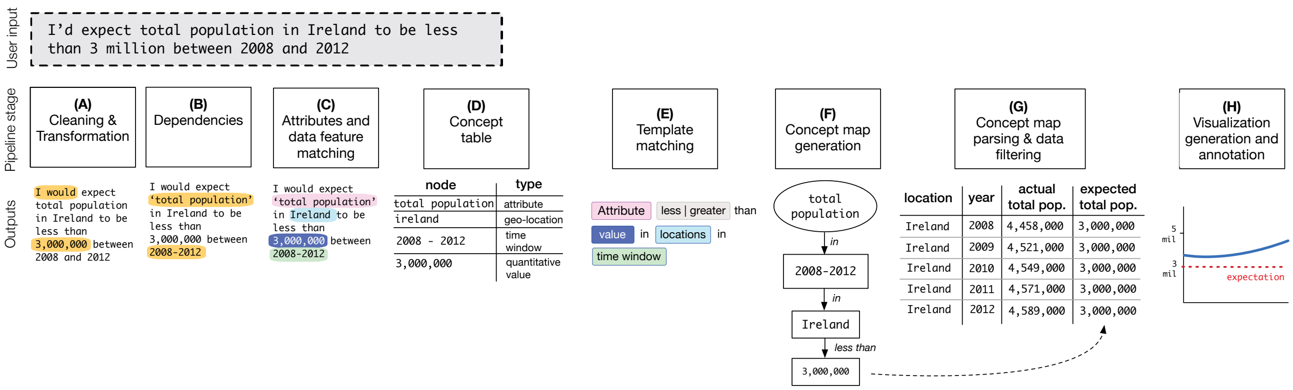

Fig. 2. A pipeline for visual hypothesis validation. Data expectations, specified in natural language, are first translated to concept maps (steps A through F). The concept maps are then used to filter the dataset for relevant records and features (G). Lastly, the actual and expected data are compared, and discrepancies are annotated onto the resulting visualization (H).

C. Model Validation and Visual Feedback

The previous part of the pipeline concludes with the generation of a concept map that encodes the user's hypothesis. To test the expectation implied and generate an appropriate validation chart, the concept map is analyzed as follows:

The concept graph is traversed depth-first starting from the root node, which by convention corresponds to a data attribute (e.g., "total population"). Traversal enables the system to determine the subset of data points that relate to the encoded expectation. This is achieved by simulating a source-sink flow model onto the concept graph; root nodes act as data sources whereas leaf nodes serve as sinks. Intermediate nodes act as filters (e.g., "Ireland", "2008-2012"), progressively restricting the 'flow' of data to, for instance, specific countries or time periods, as indicated by nodes in the graph. Ultimately this process yields a table of data points (a subset from the full dataset) comprising records that are necessary to validate the user's hypothesis. Concurrent with the traversal process, the system generates 'expected' data values at every step, comparing them against the actual data. When mismatches occur, annotations are added to the visualization to highlight the discrepancy. In the above example, because the actual population of Ireland is higher than expected, a horizontal line is added to the chart with its Y value at 3 million, along with an annotation reading "expected to be less".

To determine an appropriate output visualization, the system considers the number of attributes involved and the size of the resulting data table, employing a rule set of established practices for chart selection. The system then generates a visualization specification in the Vega grammar [19]. For example, the expectation in the example above involves one location, one attribute, and multiple years. Therefore, the system chooses a line chart as the most appropriate to represent the relevant data. The Vega specification includes both the actual data and the mismatch annotations. The latter are encoded with a salient color (e.g., bright red) so as to draw attention to model-data discrepancies.

IV. Case Study

To illustrate the utility of the system for confirmatory visual analysis, we present the following case study.

A socio-economic researcher is interested in confirming her hypothesis that life expectancy is linearly associated with economic growth indicators. She further believes this correlation will be strongly pronounced in recent years, following an economic recovery. But she suspects that this recovery has not been shared equitably among the developed economies. Based on her prior experience, she enters the following expectation into the system: "I expect GDP per capita to be positively correlated with male life expectancy in 2014". The system parses the sentence and produces a concept map along with a scatterplot showing the two attributes, and annotating a linear positive expectation for reference (see Fig. 4-A).

From the scatterplot, she finds that, while there is a relationship between the two attributes, it is not strictly linear when compared to the expectation. She thinks there might be another economic indicator, Gini, that is better correlated with life expectancy. Her hypothesis is that Gini, which accounts for the dispersion of wealth in society, will better reflect her mental model, although the correlation is expected to be negative this time (lower Gini index is more equitable distribution). She writes "I expect Gini index is negatively correlated with male life expectancy in 2014".

The system produces a concept map and a corresponding scatterplot shown in Fig. 4-B. Here, the researcher sees that the data reasonably fits her revised model. Nevertheless, she notices that the fit is less satisfying at the upper echelons of life expectancy, where a number of countries exhibit very similar expectancy levels despite having different Gini values (between 25–35). She identifies these countries by hovering over their respective points in the scatterplot, and selects six of them: France, Ireland, United States, Australia, Sweden and Netherlands. She believes the higher life expectancy in these countries (than can be explained by Gini alone) is partially driven by a relatively high labor participation. She thus expects the six countries to have an unemployment level that is well

Model	Subtype	ID	Template
Relationships	Cause and effect	R1.	X causes Y (to increase | decrease) [in locations C+] [between T1-T2]
	Mediation	R2.	X [pos | neg] affects Y, which in turn [pos | neg] affects Z [in locations C+]
	Causal sequence	R3.	A series of past events (E1, E2, E3, ...) caused X (to increase | decrease) [in locations C+] [between T1-T2]
	Correlation	R4.	X [pos | neg] correlates with Y [in locations C+] [between T1-T2]
Partial Ordering	Direct comparison	C1.	Locations C1+ are (higher | lower | similar | dissimilar) compared to locations C2+ in X [between T1-T2]
		C2.	X is (higher than | lower than | similar to | different from) Y [in locations C+] [between T1-T2]
	Ranking	K1.	Locations (C1, C2, ...) are in (ascending | descending) order with respect to X
		K2.	Location A is ranked (Nth | lowest | highest) from [a set of locations C+ | all available locations] in X
Trends & Values	Values	V1.	[Average of] X (equals | greater than [equal] | less than [equal]) V [in locations C+] [between T1-T2]
		V2.	X is in the range of V1-V2 [in locations C+]
		V3.	X is [not] (high | low) [in locations C+]
	Trends	T1.	X (increases | decreases) between T1-T2 [in locations C+]
		T2.	X exhibits a (geometric shape) between T1-T2 [in locations C+]

Fig. 3. A typology of data models representing commonly occurring expectation types. Each type of expectation is depicted with a template. Color-coded slots in the templates represent data attributes (purple), geographies (cyan), time periods (green), event sequences (yellow), quantitative data values and trends (blue), and other qualifiers (grey). 'Optional' slots are enclosed in square brackets. The typology categorizes the most common data expectations people tend to externalize in visual. analytics [17].

below 10%: "I expect the total unemployment in Australia, France, Ireland, Netherlands, Sweden, and United States to be lower than 10%". In response, the system generates a line graph (Fig. 4-C), showing indeed that her expectation is met throughout the 9 years.

Observing a downward unemployment trend that accelerates in 2014, she decides to dig further. She focuses on an attributes that could indicate investment activities. Believing that Ireland, in particular, had been struggling to attract new businesses, she chooses to compare the island nation with the six other countries, typing: "I expect business density registrations in Ireland to be lower than France, Australia, Netherlands, Sweden, and United States in 2014". A bar graph is generated (Fig. 4-D) with Ireland highlighted. Additionally, the bar chart contains annotations where the system found unexpected differences. Looking at the annotation, she finds that, counter to what she thought, there were two countries with lower new business activity (France and the Netherlands in this case). Not sure what to make of these results, she decides to repeat her analysis of unemployment, but this time she focuses on women labor participation. Additionally, she also widens the focus to include additional countries: "I expect female unemployment in 2014 in France, Australia, Netherlands, Sweden, United States, Ireland, and Spain to be less than 10%." The striking feature of the resulting chart (Fig. 4-E) is Spain (highlighted in red), as it significantly exceeded the expectation.

In a final analysis, she wonders whether rural population is decreasing in the these countries: "I expect the rural population in Ireland, France, Australia, Netherlands, Sweden, and United States to be decreasing." She finds that all countries indeed have decreasing trends, although Ireland still stands as an outlier with a higher rural population (Fig. 4-F).

This example case study illustrates how various data models can be quickly, visually, and interactively validated. Visual (dis)Confirmation currently supports a variety of data expectations, including trends, inequalities, and correlations. Its key feature is the ability to highlight counter examples, drawing more scrutiny to data instances that violate user assumptions.

V. Limitations and Future Work

Visual (dis)Confirmation represents the first attempt at creating a visual analytics tool that explicitly scaffolds model- and hypothesis-based reasoning. While the system is functional, there are a number of limitations that can be addressed with future research. First, the system relies on a predefined typology of expectations that we synthesized in a Wizard-of-Oz study. Although the typology is expansive, it is somewhat inflexible, and can limit the types of models and hypotheses that can be validated. Second, the system currently supports a limited number of charts (bar, line graphs, and scatterplots). Support for additional visualizations (e.g., node-link diagrams, parallel coordinates) can improve the utility of the system, by allowing a wider variety of models to be validated. The annotation strategy, which is used to highlight model-data discrepancies can also be improved. For instance, annotations could be placed considering the emerging visual structure of the emerging visualization to avoid occlusions, as opposed to simple superimposition as in the current system. Lastly, while we have conducted preliminary testing of the tool, there is a need for formal evaluation to ensure usability and utility.

VI. Conclusion

Current visualization tools facilitate exploratory data analysis, but fall short at fully supporting hypothesis- and model-based reasoning. We presented *Visual (dis)Confirmation*, a tool for confirmatory visual analysis. Users interact with the system by framing hypotheses and expectations in natural language, prior to seeing the data. In response, the system selects conceptually relevant data features and automatically generates visualizations to validate the underlying hypotheses. The resulting visualizations highlight places where one's

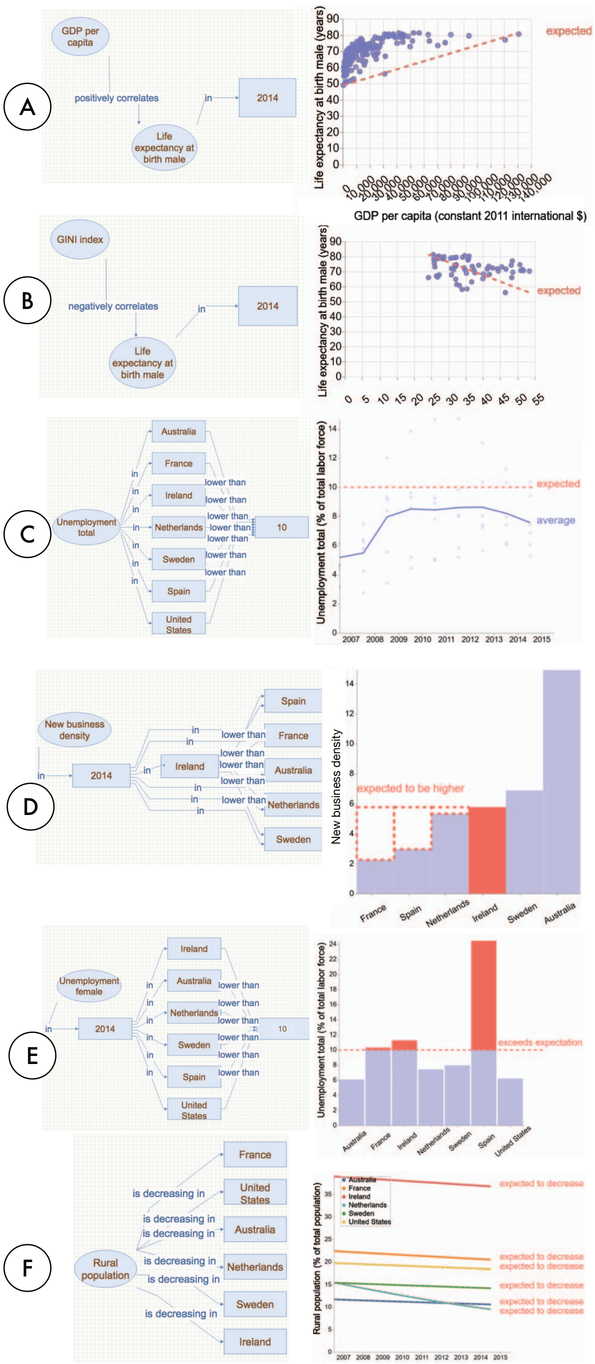

Fig. 4. A series of confirmatory visualizations generated in a hypothesis-driven case study.

model disagrees with the data, so as to stimulate conceptual reflection and model revision. We described algorithmic techniques for parsing expectations and converting them to meaningful confirmatory visualizations. We also demonstrated the utility of the system through a case study. Lastly, we discussed challenges and future research opportunities to enable a richer, bidirectional discourse between people and data through visualization.

Acknowledgement

This paper is based upon work supported by the US National Science Foundation under award 1755611.

References

[1] S. Card, J. Mackinlay, and B. Shneiderman, *Readings in information visualization: using vision to think*. Morgan Kaufmann, 1999.

[2] K. Reda, A. Febretti, A. Knoll, J. Aurisano, J. Leigh, A. Johnson, M. E. Papka, and M. Hereld, "Visualizing large, heterogeneous data in hybrid-reality environments," *IEEE Computer Graphics and Applications*, vol. 33, no. 4, pp. 38–48, 2013.

[3] G. Klein, B. Moon, and R. R. Hoffman, "Making sense of sensemaking 2: A macrocognitive model," *IEEE Intelligent systems*, vol. 21, no. 5, pp. 88–92, 2006.

[4] P. Pirolli and S. Card, "The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis," in *Proceedings of international conference on intelligence analysis*, vol. 5. McLean, VA, USA, 2005, pp. 2–4.

[5] R. Amar, J. Eagan, and J. Stasko, "Low-level components of analytic activity in information visualization," in *Information Visualization, 2005. INFOVIS 2005. IEEE Symposium on*. IEEE, 2005, pp. 111–117.

[6] B. Shneiderman, "The eyes have it: A task by data type taxonomy for information visualizations," in *Proceedings of the IEEE Symposium on Visual Languages*. IEEE, 1996, pp. 336–343.

[7] N. Andrienko, T. Lammarsch, G. Andrienko, G. Fuchs, D. Keim, S. Miksch, and A. Rind, "Viewing visual analytics as model building," in *Computer Graphics Forum*, vol. 37, no. 6. Wiley Online Library, 2018, pp. 275–299.

[8] J. W. Tukey, "We need both exploratory and confirmatory," *The American Statistician*, vol. 34, no. 1, pp. 23–25, 1980.

[9] K. Dunbar, "Concept discovery in a scientific domain," *Cognitive Science*, vol. 17, no. 3, pp. 397–434, 1993.

[10] Dunbar, "How scientists think: On-line creativity and conceptual change in science," *Creative thought: An investigation of conceptual structures and processes*, vol. 4, 1997.

[11] J. Stasko, C. Görg, and Z. Liu, "Jigsaw: supporting investigative analysis through interactive visualization," *Information visualization*, vol. 7, no. 2, pp. 118–132, 2008.

[12] G. Klein, B. Moon, and R. R. Hoffman, "Making sense of sensemaking 1: Alternative perspectives," *IEEE intelligent systems*, no. 4, pp. 70–73, 2006.

[13] J. Aurisano, A. Kumar, A. Gonzales, K. Reda, J. Leigh, B. Di Eugenio, and A. Johnson, ""show me data": Observational study of a conversational interface in visual data exploration," in *IEEE VIS posters*, 2015.

[14] Y. Sun, J. Leigh, A. Johnson, and S. Lee, "Articulate: A semi-automated model for translating natural language queries into meaningful visualizations," in *International Symposium on Smart Graphics*. Springer, 2010, pp. 184–195.

[15] K. Dhamdhere, K. S. McCurley, R. Nahmias, M. Sundararajan, and Q. Yan, "Analyza: Exploring data with conversation," in *Proceedings of the 22nd International Conference on Intelligent User Interfaces*. ACM, 2017, pp. 493–504.

[16] V. Setlur, S. E. Battersby, M. Tory, R. Gossweiler, and A. X. Chang, "Eviza: A natural language interface for visual analysis," in *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*. ACM, 2016, pp. 365–377.

[17] I. K. Choi, T. Childers, N. K. Raveendranath, S. Misrha, K. Harris, and K. Reda, "Concept-driven visual analytics: an exploratory study of model- and hypothesis-based reasoning with visualizations," in *Proceedings of the ACM Conference on Human Factors in Computing Systems (to appear)*. ACM, 2019.

[18] C. Manning, M. Surdeanu, J. Bauer, J. Finkel, S. Bethard, and D. McClosky, "The stanford corenlp natural language processing toolkit," in *Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations*, 2014, pp. 55–60.

[19] A. Satyanarayan, R. Russell, J. Hoffswell, and J. Heer, "Reactive vega: A streaming dataflow architecture for declarative interactive visualization," *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 1, pp. 659–668, 2016.