What Flu Forecasting Taught Us About Visualizing Uncertainty

Andrea Brennen
IQT Labs / In-Q-Tel
Waltham, MA USA
abrennen@iqt.org

George Sieniawski
IQT Labs / In-Q-Tel
Waltham, MA USA
gsieniawski@iqt.org

Dylan George
B.Next / In-Q-Tel
Arlington, VA USA
dgeorge@iqt.org

ABSTRACT

Data visualizations are an increasingly popular means of communicating information to decision-makers and the public. However, today's tools for creating visualizations offer limited capabilities for conveying uncertainty. We believe that when visualizations do not convey known uncertainties they can inadvertently misrepresent the output of analytical efforts by implying more certainty than is (or can be) known about underlying data. In our work we use design prototyping and empirical evaluation to develop and validate visualization tools to convey uncertainties in data and data analysis. In our **"Viziflu"** project, we collaborated with the Centers for Disease Control and Prevention (CDC) Influenza Division to: (1) design and build a web-based tool for visualizing uncertainty in crowdsourced seasonal influenza forecast data; and (2) evaluate how the use of that tool influenced the way viewers interpreted uncertainty in the forecasts. The intented users of Viziflu were analysts working in CDC's Influenza Division. However, we believe our findings may be of interest to Computation & Journalism attendees because the visualization tool was designed to support CDC's efforts to communicate with journalists and the public about the spread and timing of seasonal influenza outbreaks and to underscore the need for precautionary measures such as vaccinations.

Viziflu displays publicly-available forecast data generated through FluSight, a flu forecasting contest. Academic teams submit weekly forecasts for targets set by CDC, such as onset, peak and intensity. Submitted forecasts are then compared against observed cases of influenza-like illness (ILI) within the ILINet database, CDC's main influenza surveillance system. During the 2018-2019 season, 23 teams produced 38 distinct forecasting models. Experimentation with different modeling approaches will ultimately help improve forecasting results. However, in recent years, variations across models have created significant discrepancies in resulting forecasts, exacerbating the difficulty of synthesizing actionable information. Based on user research we conducted with the CDC Influenza Division, seasonal influenza forecasts can help inform operational decision-making by providing timely information about *when flu is likely to peak*. In response, Viziflu emphasizes temporal predictions and helps viewers understand two different types of uncertainty: (1) the predicted probability that flu will peak in a particular week of the flu season, (a quantitative output provided by each forecast model) and (2) uncertainty created when multiple forecast models output conflicting predictions. The Viziflu tool and source code are available under the Apache 2.0 Open Source License (see https://github.com/BNext-IQT/Viziflu).

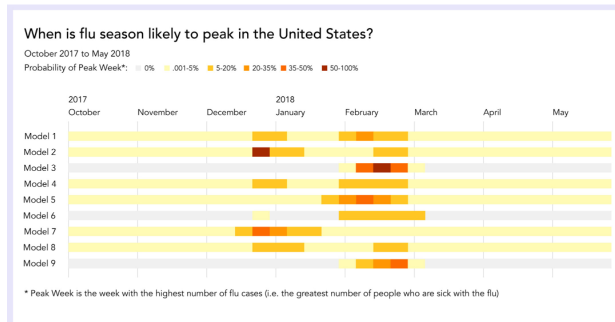

Figure 1: Viziflu displays crowdsourced forecasts for when seasonal influenza is likely to peak nationwide.

We conducted an exploratory user study via Amazon's Mechanical Turk to investigate how the number of forecasts displayed in Viziflu might influence participants' interpretations of the information they saw. Participants were randomly assigned to one of three groups and each group was shown a version of Viziflu containing a different number of forecasts (1, 3 or 9). We asked participants both to answer four information retrieval questions (with definitive answers) and to make a series of judgments about when flu was likely to peak, given the forecasts they saw. We then asked them to report their *conviction* in these judgments and the *cognitive load* required to make the judgments.

For the information retrieval questions, we saw an overall response accuracy rate of 80%, which we interpreted as an indication that participants were able to interpret the Viziflu UI effectively. Our results also suggested three hypotheses about how the number of models displayed might influence participants' interpretations of forecast data: (1) that viewing more forecasts increases participants' reported cognitive load; (2) that viewing more forecasts may be correlated with a decrease in response accuracy, even for questions that refer only to a single forecast; and (3) that the number of forecasts displayed may influence participants' conviction in significant ways. We observed that when forecasts were aligned, participants who saw multiple forecasts expressed 10% greater conviction than those who saw only one, but when forecasts predicted different outcomes, participants who saw multiple forecasts expressed 17% less conviction, even when they came to the same general conclusion about the data.