Design Challenges in Playable Data

Nicholas Diakopoulos

School of Communication and Information

Rutgers University

diakop@rutgers.edu

Position Statement

Work published this year at CHI has introduced the notion of game-y information graphics which take raw datasets from sources such as data.gov and create playable visualizations by adding elements of goals, rules, rewards, and mechanics of play [2]. One example is Salubrious Nation, which uses geographically tagged public health data such as smoking and obesity rates, to create a guessing game. The goal of the game is to accurately guess the magnitude of the given health parameter for a randomly selected target county. A player's guess can be informed by looking at the map (See Figure 1) for visual clues as a slider is changed, or by using hover-over information on correlated variables (e.g. poverty rate or elderly population rate).

In addition to allowing players to use the map-based graphic to arrive at insights about the data and to redistributing players attention to different aspects of the data, such an approach also offers the promise of reducing the amount of effort needed to repurpose that data into new playable experiences. Interested readers can see [2] for all of the details.

In the remainder of this position paper, however, I would like to expound on and explore the design difficulty associated with creating a *challenging and balanced* game experience when drawing on raw
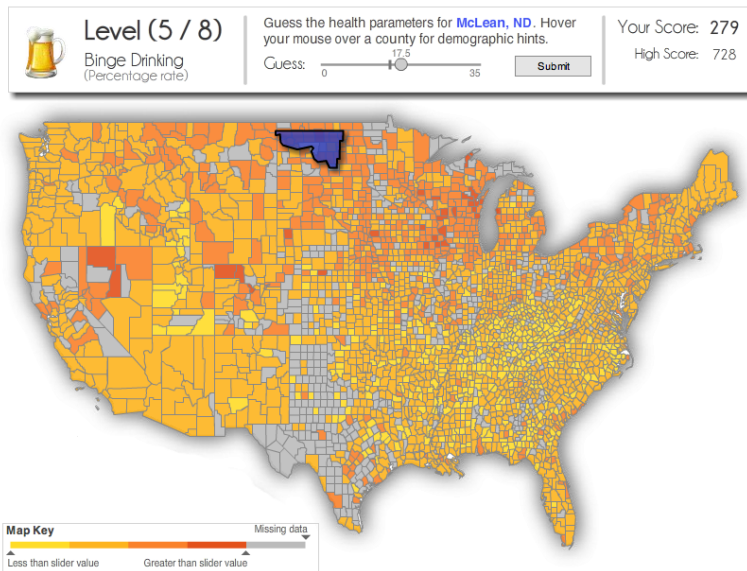
Figure 1. Salubrious Nation. http://www.salubriousnation.com

datasets as input for the construction of a game. Ordinarily when designing games, substantial effort is directed to level design. In fact, many games employ dedicated level designers who work with the game designer in order to provide the right amount of challenge, reward, and balance to the game experience [3].

In contrast to such heavily authored experiences, gamified data experiences (whether they be based on infographics as in Salubrious Nation, or not), may draw on data that is incomplete, inconsistent, or dynamic. For instance, if a dataset is missing values, such missing values must be taken into account so that this does not completely break the game, or at least does not substantially reduce the engagement of the experience. Salubrious Nation relies on correlations between health variables to demographic variable such

as poverty rate, to help users predict the public health variable (e.g. smoking rate). If the data were updated in such a way that relationships (i.e. such as a correlation) was diminished or removed, this would affect the playability of the game.

Dealing with data that is updated, refreshed, or otherwise dynamic represents a design challenge. Another example, the California Stimulus Map Game [1] was a game-y infographic created for the Sacramento Bee newspaper website. In this trivia game players had to answer a series of trivia questions about stimulus funds by interacting with a visual map of the state of California. Two weeks after the initial publication the data for the map in the game had already been updated by the government. Not only did this affect the visual representation of the map, but it also impacted the answers to some of the trivia questions, thus forcing the designers to update the game in order to accommodate the new data. One approach to dealing with this issue would be to devise better automatic authoring routines so that trivia answers could be extracted directly from the data without human intervention (e.g. "What is the county with the largest (or smallest) amount of stimulus money"). More research needs to be done to determine the best way for dealing with changes to data which can impact a play experience. Methods developed should be robust to incomplete, inconsistent, or dynamic data and should provide for a playable experience regardless of reasonable changes to such data.
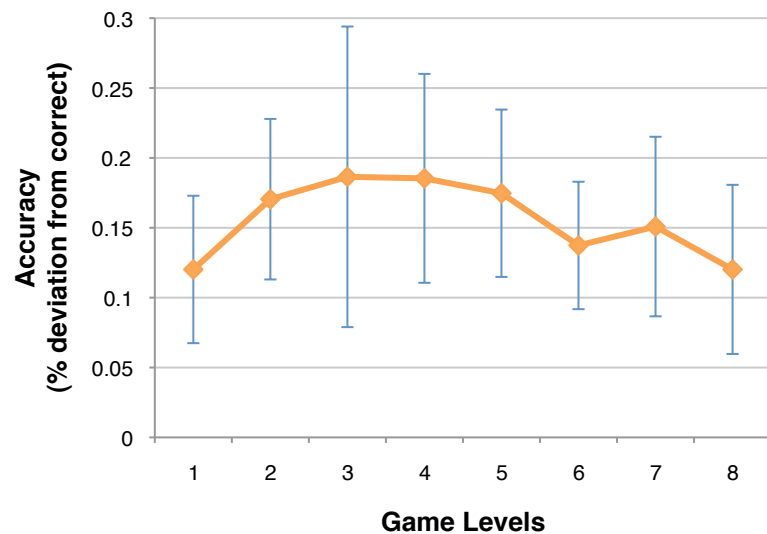
Figure 2. The average accuracy of players' guesses for each level. Error bars show the standard deviation

A more general issue with raw data is that the challenge or difficulty of the experience produced in the game is hard to control. With one set of data as an input a game may be too easy but with another it could become too hard. For instance, in Salubrious Nation there were 8 levels, each using a different public health parameter. For each of the levels we measured the average accuracy of the guesses that were produced by the 41 players in our experiment. This is shown in Figure 2. As can be seen in the graph, some levels were more difficult than others, even considering some potential learning and improvement by players in the latter levels. This is in contrast to the typical game design pattern of increasing difficulty of levels. Indeed, based on the collected data it may be advisable to re-order the levels in Salubrious Nation so that easier levels are first and more difficult ones later.

In the absence of carefully authored levels of a game, we can still collect log data from players in order to infer difficulty and challenge. While this is relatively straightforward for a puzzle where there is a correct answer and a relatively simple metric can be used to infer difficult, there remain open questions for research. How can log data be used to infer other measures of difficulty (frustration even)? How can playable data games be rapidly and perhaps automatically re-adjusted to assess difficulty so that in a short period when a game is first being played it is able to evolve and adjust itself to provide an appropriately balanced and challenging experience?

These questions apply generally to the gamification of any data-based resource. When gamifying a dynamic, perhaps arbitrarily defined data source, how can we arrive at estimates for the challenge, balance, and playability of those experiences? Properly instrumented such games could perhaps automatically adapt their levels and difficulty to compensate for differences in the input data. I believe that answering these questions will be essential to being able to more rapidly create compelling gamified data experiences in the future.

References

1. Diakopoulos, N., Game-y Information Graphics. *CHI Extended Abstracts*, (2010).

2. Diakopoulos, N., Kivran-Swaine, F. and Naaman, M., Playable Data: Characterizing the Design Space of Game-y Infographics. *Proc. CHI*, (2011).

3. Fullerton, T. *Game Design Workshop*. Morgan Kaufmann, 2008.