

Exploring Data For Fun And Profit: Case Study of Jeopardy!

Joshua Appleman, Anubhav Gupta, Anand Rajagopal, Juan Shishido, Marti A. Hearst*

School of Information, UC Berkeley

ABSTRACT

This case study describes a new, emerging way to shed light on a complex dataset and communicate with a broad audience: first perform exploratory data analysis, then create interactive visualizations and embed these within a narrative using infographics. We illustrate the idea using data from 30 seasons of the *Jeopardy!* TV game show, showing interesting patterns both qualitative and quantitative.

Index Terms: H.5.0 [Information Interfaces and Presentation]: General

1 INTRODUCTION

Information visualization has escaped the research lab and is now widely used by practitioners across a wide spectrum of fields. New design paradigms are rapidly gaining acceptance and evolving. Today, a powerful new way to communicate insights about data is to interweave interactive visualizations within a narrative using well-motivated infographics [1].

In this poster, we demonstrate an example of this emerging design paradigm in which data is first explored using techniques from exploratory data analysis, then trends are confirmed using interactive graphics, and a story is tied together using narrative infographics, all while using sound information visualization and HCI principles. We demonstrate this idea using as a case study a large multivariate dataset about the TV game show *Jeopardy!*

Multivariate data, which often includes both qualitative and quantitative features, can be difficult to view from multiple perspectives. Our goal was to present data about the show in a manner that presents the essence of the game statistics while also highlighting less obvious trends. We also wanted to present an informative tour through the game’s history that is also interactive so that users could explore their particular interests.

Using a combination of Tableau and Python to explore the dataset, D3 and Highcharts to develop interactive visualizations, and Illustrator to create additional graphics, we produce a visualization with a coherent narrative.¹

2 DATA

Jeopardy! is an American TV game show in which three contestants are presented general knowledge clues in the form of answers and, in response, must phrase questions to the given answers. The game consists of three rounds: Jeopardy and Double Jeopardy, each of which consist of six categories with five questions each, and Final Jeopardy. In Final Jeopardy, contestants answer a single clue and are able to wagers part or all of their earnings. Details about betting rules and strategy can be found in [3].

We obtained qualitative data from the fan-driven website J! Archive² consisting of more than 200,000 Jeopardy! questions and

* {japple,anubhav.rg,anand_rajagopal,juanshishido,hearst}@berkeley.edu

¹Details about software tools used can be found at: d3js.org, tableau-software.com, highcharts.com, and www.adobe.com/illustrator

²j-archive.com/; Earnings data from jeopardy.com

related information such as the air date for each episode, the prize money for each question, and contestants’ names across 30 seasons. *Jeopardy!* is not just about the questions and answers, so we augmented this with quantitative data about question-by-question scores as well as the total earnings for about 4,700 episodes.

We used several data sources and strategies to determine contestants’ genders, including probability based on the US Government Baby Names corpus, online APIs, entering ambiguous names into social media sites to see which gender was more probable and manually checking contestant’s photos on J! Archive.³

Given the diverse nature of the entire data set, our initial challenge was to identify its most interesting aspects so that we could focus our efforts towards that front.

3 EXPLORATORY DATA ANALYSIS WITH TABLEAU

We began with a set of hypotheses about the data and explored them using Tableau. One hypothesis was that some topics or domains are preferred by the show’s content creators (and if this were true, this would prove beneficial to anyone preparing to compete in the show). Traditionally, fans have suspected “Potpourri” to be a popular category, but our analysis showed “Before and After” to be at the top followed by “Science,” “Literature,” and “American History” [4]. We then flipped the analysis and looked at the most frequent *answers*. Surprisingly, the 20 most frequent answers were all geographic locations.

While the dashboards helped highlight interesting trends, they focused on specific areas of the dataset and did not present a holistic view. To explore other facets of the data, such as a potential gender divide or wagering dynamics, we decided to implement a richer interactive view, described below, to better characterize this data.

4 VISUALIZATIONS WITH D3 AND HIGHCHARTS

In order to ensure that the visualization was not too narrowly focused, we developed D3 visualizations that presented multiple variables in a simple yet organized manner. We made a variation on the parallel coordinates visualization [2] that connects quantitative variables to qualitative ones. We used this to highlight the achievements of the top 50 contestants and how they fared across various parameters (see Figure 1).

Working with 30 seasons of Jeopardy! challenged us to think of effective ways of displaying the data. As our primary goal for visualizing the earnings data was to provide multiple levels of detail, we provide a slider to allow users to filter the data by season. A Javascript-based heatmap provides an overview of the earnings data where each square represents a single episode and the layout shows patterns by week and month (see Figure 2). Users can click on a square to update a Highcharts-based line chart that gives insight into the game dynamics and how each contestant performed for all questions for that episode. A trend of higher earnings over time can be seen by moving the slider from left to right.

The Final Jeopardy round is where the biggest potential for earnings gains and position changes are. To explore how contestants make wagers, we created a scatter plot of wagers against earnings going into the Final Jeopardy round. Wagers are shown in both absolute and percentage (of earnings) terms and the dots are shaded to represent the position of a contestant at the time they placed the

³www.ssa.gov/oact/babynames; genderchecker.com; gender-api.com

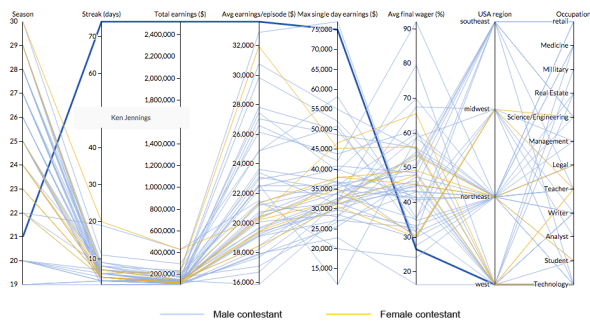


Figure 1: Exploring the characteristics of the top 50 contestants with a parallel coordinates view. It is easy to see how well-distributed their occupations are, the distribution of female contestants, and how much of an outlier Ken Jennings is.

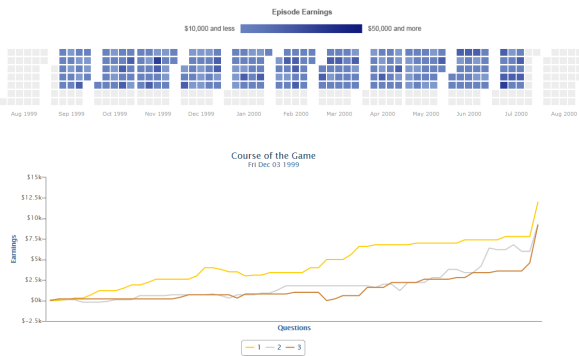


Figure 2: Heatmap of earnings; a slider changes the view over time.

wager. The scatter plot is linked to the slider, showing wager tendencies by season and position. As a way to summarize the wager information, we use a bipartite graph to show how player positions changed before and after the Final Jeopardy round.

5 NARRATIVE INFOGRAPHIC WITH ILLUSTRATOR

Having created dashboards with Tableau that provide depth within the dataset and D3 visualizations that cover the breadth of the data, we came to the most essential part of the process—developing a narrative. It is crucial that the user have enough context to understand the purpose of the visualization and is sufficiently engaged to want to interact with the data. Creating infographics with Illustrator and charts using Highcharts, we developed a compelling story which stitched the different pieces together.

Part of that story involved explaining the audition process. The goal was to convey the difficulty of being selected to be on the show. Within this context, we also explored how the proportion of women on the show has changed over the years.

6 UNIFYING DESIGN

In order to display the content, we designed a web page that contains the entire narrative with the different dashboards and visualization components.⁴ We organized the page into two sections: the players, which provides an overview and features more qualitative data, and the earnings, where users can explore the individual episodes. At this stage, it was important to make good design choices such as a common color theme and uniform spacing and

⁴<http://people.ischool.berkeley.edu/~japple/jeopardy/>

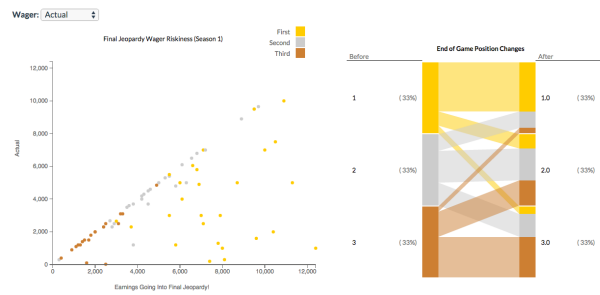


Figure 3: The bipartite visualization shows the movement of players between different positions before and after the Final Jeopardy round. While contestants typically maintain their positions, third place contestants are more likely than those in second place to move up.

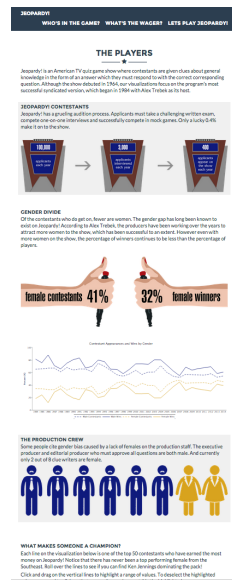


Figure 4: The infographics help draw attention to the key points in the narrative.

grouping. By using Gestalt principles of proximity and similarity, we distinguished between the different threads in the visualization.

7 CONCLUSION

This case study shows a general approach to provide insight about a large data set: use findings from exploratory data analysis to create dashboards that provide focused analysis of the data, then create interactive information visualizations which provide a broader view overall. The final step is to knit the pieces together with a narrative and the context it provides to make the visualization coherent and engaging.

REFERENCES

- [1] A. Cairo. *The Functional Art: An introduction to information graphics and visualization*. New Riders, 2012.
- [2] A. Inselberg. Multidimensional detective. In *Information Visualization, 1997. Proceedings., IEEE Symposium on*, pages 100–107. IEEE, 1997.
- [3] A. Metrick. A natural experiment in Jeopardy! *The American Economic Review*, pages 240–253, 1995.
- [4] J. Singer-Vine. I'll take jeopardy! trivia for \$200, alex. In *Slate*. Feb 2011.