**Background**

Hearthstone is an online collectible card game produced by Blizzard Entertainment. Since its release in 2014, it has attracted over 100 million players worldwide, garnering enough interest for Blizzard to sponsor international championship tournaments and the development of a large online community of players dedicated to developing new competitive strategies and skills.

The game is a turn-based card game between two opponents, with each player constructing a deck of 30 cards and a selecting a hero character with a unique power. Players are paired via an online queue and then use limited resources each turn to play their cards, with the ultimate goal of destroying the opponent’s hero. With over 1000 cards to choose from, and with developers regularly adding more, the space of possible decks is enormous. However, conventional wisdom among players is that at any one time there will be a small number of deck “archetypes” which perform outstandingly against most of the other decks commonly used, and that this set of overperforming decks will change as new cards are released and as players adapt to the current state of the game (referred to as the “meta-game”). Multiple websites have arisen for players to collaboratively identify and track the success of decks over time, and to discuss deck lists and strategies.

**The Data**

Data on all 1800+ collectible cards in the game was obtained from hearthstonejson.com, a site maintained by developers interested in hearthstone, and 5000 fan-submitted decklists were scraped from hearthpwn.com. Relevant fields in the card data included the cost, class, and game mechanics of each card. The cards data were flattened into a wider format (with one column per game mechanic instead of one column containing lists of game mechanics for each card). Decks scraped from hearthpwn.com had to be checked for validity under the rules of the game (as a joke, users will sometimes intentionally submit invalid decks) and then also saved in a sparse matrix with one row per deck and one column per card, where entry (i,j) in the table was equal to the number of copies of card j in deck i. The decklist data was further processed to create a square matrix of size equal to the number of cards; entry (i,j) in this matrix is the number of decks containing at least one copy of both cards i and j (diagonal elements are the number of decks containing a specific card). The information obtained on the decks was simply the set of cards the author chose to include.

**The Application**

The shiny application contains three tabs for exploring one of each of the three datasets. The first tab allows the user to understand the population of minion cards available in the game. Controls allow filtering for particular game mechanics and card costs. Plot axes can be set to show bivariate relationships between any pair of minion health, cost, and attack values, or a univariate histogram of one of those. Points are sized by the number of minions at the corresponding value and can be colored by the value of the aforementioned three numerical variables or by their class. Finally, the plot can be faceted by the parity of the card cost (game mechanics exist that provide bonuses for decks that are all even-cost or odd-cost cards) or the class. To the right, a menu exists to view any of the cards that match the current filtering criteria (game mechanics are typically in boldface font on a card’s text).

Some possible uses of this tool include helping the user:

* Develop an intuitive grasp the variation that exists among minions’ stats and attributes at a given cost;
* Find minions similar to a given minion in terms of attributes, stats, etc. If a player does not own a particular card but wants to build a deck that includes it, they can find a substitute in this manner;
* Compare distributions of minion stats or mechanics across classes, which has strategic implications for playing the game or constructing decks.

The second tab gives the user the choice to view pairwise associations between the most popularly included cards. This was designed to be helpful to a player identify cards that synergize well or complement each other’s roles in decks, as well as identify which cards are more popular among submitted decks in general. The heatmap is colored by the frequency with which each pair of cards occurs in the data, and the tree diagrams on each side show higher-level clustering of cards. Since a deck’s class necessarily restricts it to a certain subset of cards, the lower-level clustering of cards is less interesting than the higher-level clusters – it’s not surprising that mage cards often appear with other mage cards, for example, but it is interesting if neutral cards (which can be played with any class) that are associated more with one class are also associated with another class. This would perhaps indicate that classes with cards closer together on the tree diagram have more similar strategies or playstyles in the current state of the meta-game (since their deck makeups are similar to a higher degree). As strategies evolve, new cards are released and players adapt to these changes, the clustering could also change.

The final tab borrows techniques from text analysis. It treats each deck as a document in a corpus, and each card included in the deck as a word in that document. Given a user-inputted number of deck archetypes to find and an option to restrict the analysis to decks that contain specific cards, the application then uses latent Dirichlet allocation (LDA) for topic modeling of the “documents.” LDA identifies topics in a corpus by finding sets of words that are common within particular topics and rare outside of those topics. In this context, topic modeling can be used to find sets of cards that form the “core” of many decks on hearthpwn.com; these cards are critical to the strategy behind the deck and are more stable across different instances of similar decks than other cards that are more optional. For a player using the app, this “archetype finder” tool, beyond simply identifying the cores of the most common deck archetypes, also identifies and displays the ten most important cards for that topic. If a player is considering building a deck that follows a particular strategic archetype but is working with an incomplete card collection, this tool can help them identify which cards are truly critical to the working of the deck and which they can more easily get away with substituting.