

Relazione

Emanuele Artioli

May 15, 2020

ANALISI GIOCO DI CARTE COLLEZIONABILI: MAGIC THE GATHERING

ABSTRACT

Questo progetto analizza componenti estetiche e funzionali del database di carte collezionabili Magic the Gathering dal 1993, anno della sua presentazione.

GIOCO

In breve, MtG rappresenta un duello tra maghi, che a turni lanciano magie rappresentate dalle proprie carte. Queste magie richiedono mana, energia elementale di 5 colori attinta da altrettanti tipi di carte terra: Pianura per il bianco, Isola per il blu, Palude per il nero, Montagna per il rosso, e Foresta per il verde. Tra i tipi di magie troviamo creature, incantesimi, stregonerie, artefatti, e non solo. Ogni tipo di magia ha caratteristiche peculiari.

DATI

Il dataset è stato scaricato da <https://mtgjson.com/> e importato in forma di database SQLite tramite libreria RSQlite. Per la presente analisi, ci concentriamo sulle tabelle:

1. cards, contenente 70 variabili per più di 50000 carte.
2. prices, una lista dei prezzi aggiornata quotidianamente
3. legalities, la lista dei tornei nei quali ogni carta può essere giocata.
4. sets, la lista delle edizioni delle carte.

Il dataset iniziale conteneva più di 200000 righe, ma queste comprendevano ogni ristampa, nonché carte ironiche mai pensate per essere giocate e dunque non conformi alle regole del gioco, e altre tipologie di carte utilizzabili solo in contesti speciali. Vi era poi, come è facile prevedere, un gran numero di ridondanze, errori o ambiguità nel dataset. Inoltre, MtG ha negli anni introdotto molte regole aggiuntive che hanno modificato o sostituito quelle precedenti, rendendo obsolete alcune carte che sono pertanto una complicazione, peraltro difficile da individuare, non secondaria. Per tutti questi motivi, il dataset è stato decisamente rimaneggiato, mantenendo però assolutamente intatta l'informazione. Dopo le modifiche preliminari, le variabili di interesse sono le seguenti:

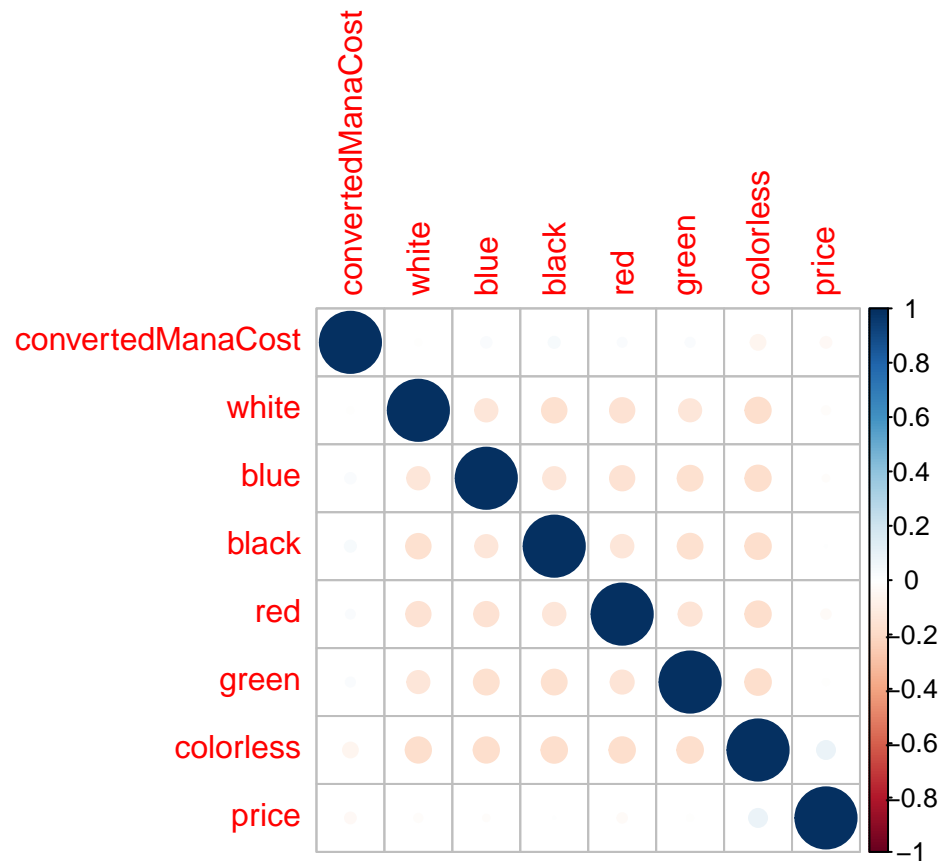
DF (unione di cards, prices e legalities, meno gli attributi relativi solo a particolari tipi di carta)

1. name (character)
2. convertedManaCost (numeric)
3. white (numeric)

4. blue (numeric)
5. black (numeric)
6. red (numeric)
7. green (numeric)
8. colorless (numeric)
9. types (character)
10. rarity (factor)
11. setCode (character)
12. price (numeric)

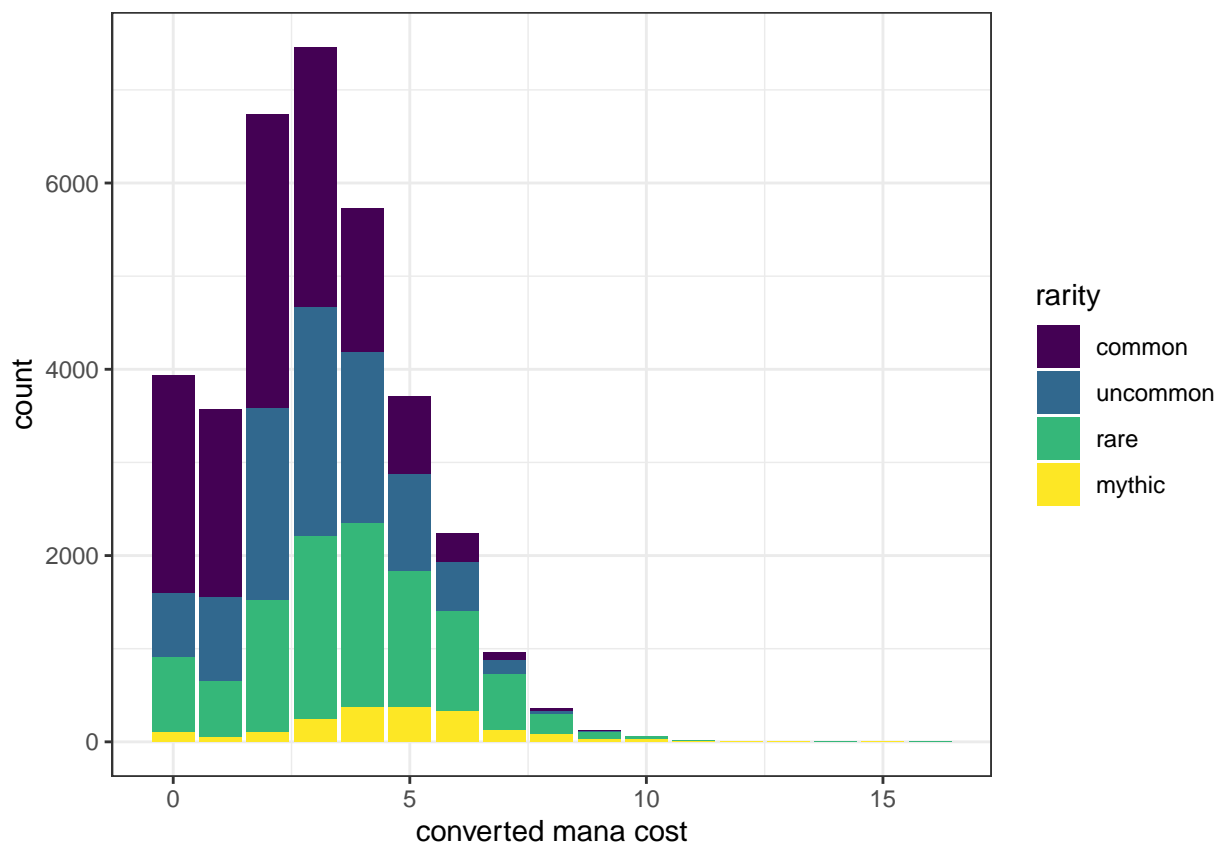
PANORAMICA

Iniziamo la nostra analisi con una concisa rappresentazione delle principali caratteristiche comuni ad ogni carta, per avere una visione ad alto livello del dataset.



Da una preliminare analisi di correlazione si vede come gli attributi numerici non diano molte informazioni, tranne quella data dalla leggera correlazione negativa tra i colori, che ci indica una tendenza delle carte ad avere pochi colori. L'analisi si dovrà allora concentrare sugli attributi non numerici, che analizzeremo uno per volta. Dati così poco correlati potrebbero scoraggiare dall'indagare ulteriormente il dataset, poniamo però l'attenzione su un limite della correlazione tramite un esempio lampante: se consideriamo la correlazione, come detto leggermente negativa, tra gli attributi white e colorless, potremmo essere portati a pensare che questi possano coesistere relativamente spesso, e invece è ovvio che essi siano mutuamente esclusivi. La

negatività della correlazione viene attenuata soltanto dall'ingente numero di casi in cui entrambi gli attributi sono assenti (ovvero la carta ha colore, ma non bianco).

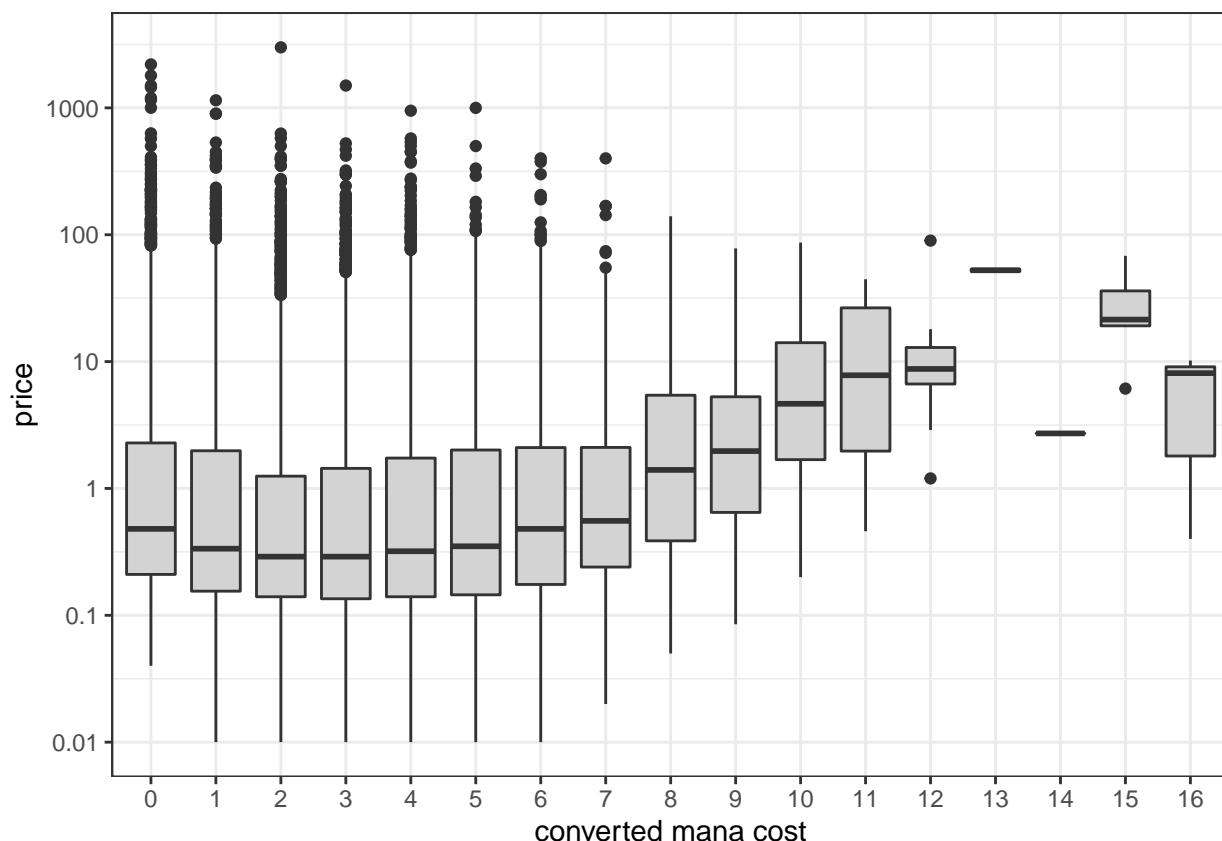


E' subito evidente come la distribuzione del costo di mana sia positivamente asimmetrica, per quanto media e mediana siano quasi coincidenti (mediana = 3, media = 3.047). La variabilità è contenuta, con scarto quadratico medio di 1.95 e coefficiente di variazione 0.64.

Un'altra osservazione degna di nota è relativa alla distribuzione delle rarità: tralasciando le carte mitiche, che sono state introdotte dopo molti anni dalla nascita di MtG, vediamo come il numero di carte sia pressapoco lo stesso per ogni rarità, semplicemente distribuito in maniera differente: al crescere del costo di mana, cresce la rarità media. In ogni busta però si trovano più di 10 carte comuni, 3 non-comuni e 1 sola rara. Se uniamo questo alla grande varietà di carte rare, ricaviamo che trovare una particolare carta rara è particolarmente difficile, e questo senza dubbio contribuirà ad alzare il prezzo di alcune di queste, le più ricercate.

Obiettivo dell'analisi: Prezzo

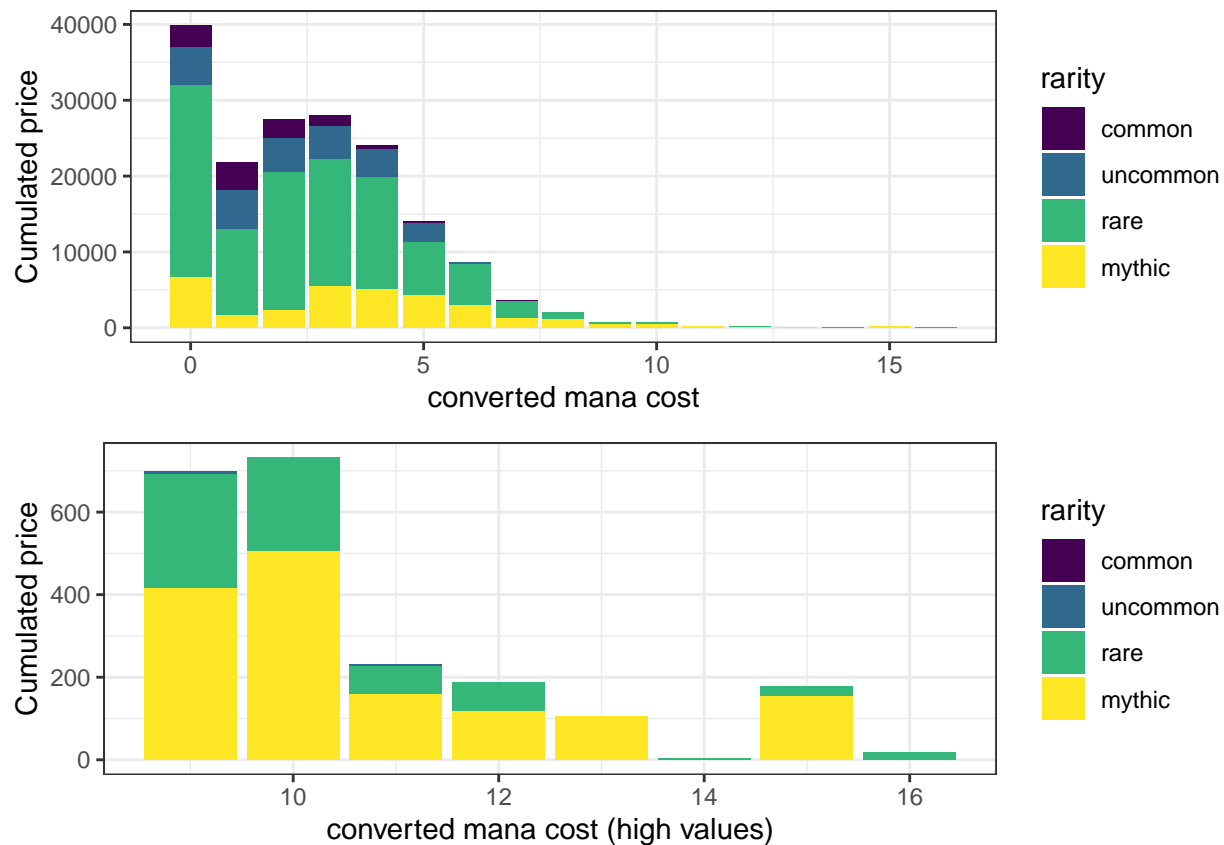
Il punto di partenza per un'analisi più approfondita sarà il prezzo delle carte. Infatti, l'analisi è inquadrata nella ricerca di caratteristiche comuni alle carte che possano rispondere alla seguente domanda: è possibile prevedere il prezzo di una carta dalle sue caratteristiche? Abbiamo già visto come il prezzo non sia correlato al costo di mana nè al colore, ma ciò non toglie che possano sussistere legami che sfuggono a questa prima analisi. Un'altra possibilità è che il legame sia da ricercare negli attributi non numerici.



Il grafico mostra una dispersione di prezzo elevata, con molte carte di nessun valore e picchi che invece superano le migliaia di euro per singolo pezzo. Questa dispersione è però fortemente influenzata da questi ultimi, che possiamo considerare outlier una volta calcolato il range interquartile e verificato che questo è inferiore ad 1€.

Potrebbe stupire che, sebbene le carte ad alto costo di mana siano quelle che dovrebbero portare un maggiore impatto nella partita, i prezzi maggiori si riscontrino tra le carte a costo inferiore.

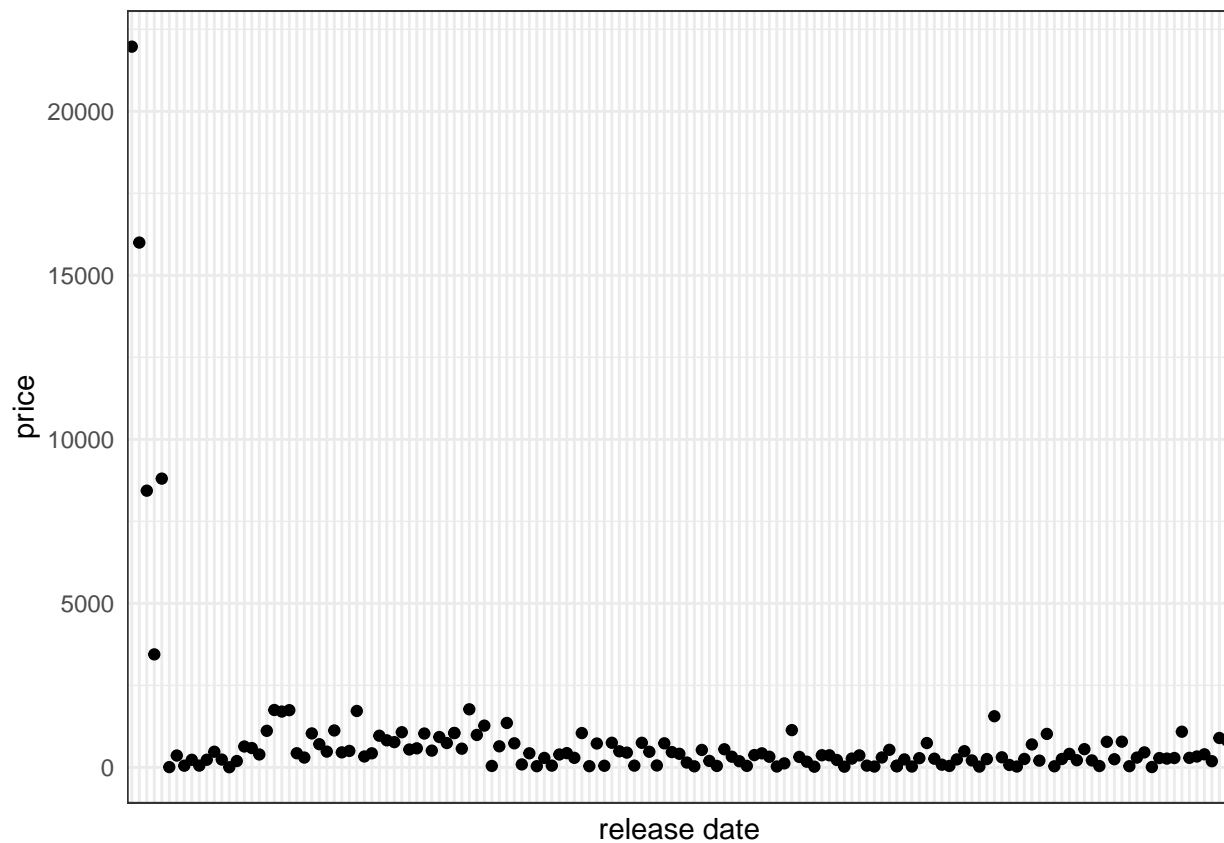
Possiamo spiegare questo fenomeno soltanto in termini di mole di dati? Quindi avere valori di prezzo così alti per carte a basso costo di mana nonostante una media molto bassa, soltanto in virtù del fatto che, come visto dalla precedente visualizzazione, ci sono molte più carte a basso costo? Questa risposta non spiega ad esempio i prezzi medi relativamente alti delle carte a bassissimo costo, 0 e 1 (si tenga presente che essendo il grafico logaritmico, la differenza tra le medie è più accentuata di ciò che appare, ma il trade-off era necessario). La conoscenza di dominio ci permette di fare luce su questa apparente contraddizione, che è in realtà una caratteristica del gioco nota alla maggior parte dei giocatori esperti: la lunga vita di MtG ha portato a carte di ogni genere, che assolvono qualsiasi funzione i giocatori richiedano dal loro mazzo. Pertanto, il gioco è divenuto progressivamente più rapido, in quanto il risultato ottenuto con una carta ad alto costo può tipicamente essere raggiunto dalla sinergia di carte a costo minore, e il limite massimo, imposto dal mana disponibile, alle carte giocabili in un turno, rende la seconda opzione molto più efficiente e preferibile. Da questo una domanda maggiore per carte più “veloci”, che ne aumenta i prezzi. Addirittura, una media di prezzo così elevata per carte ad alto costo, raramente incluse nei mazzi di alto livello, pare in contrasto con l’esperienza comune, e merita di essere ulteriormente analizzata.



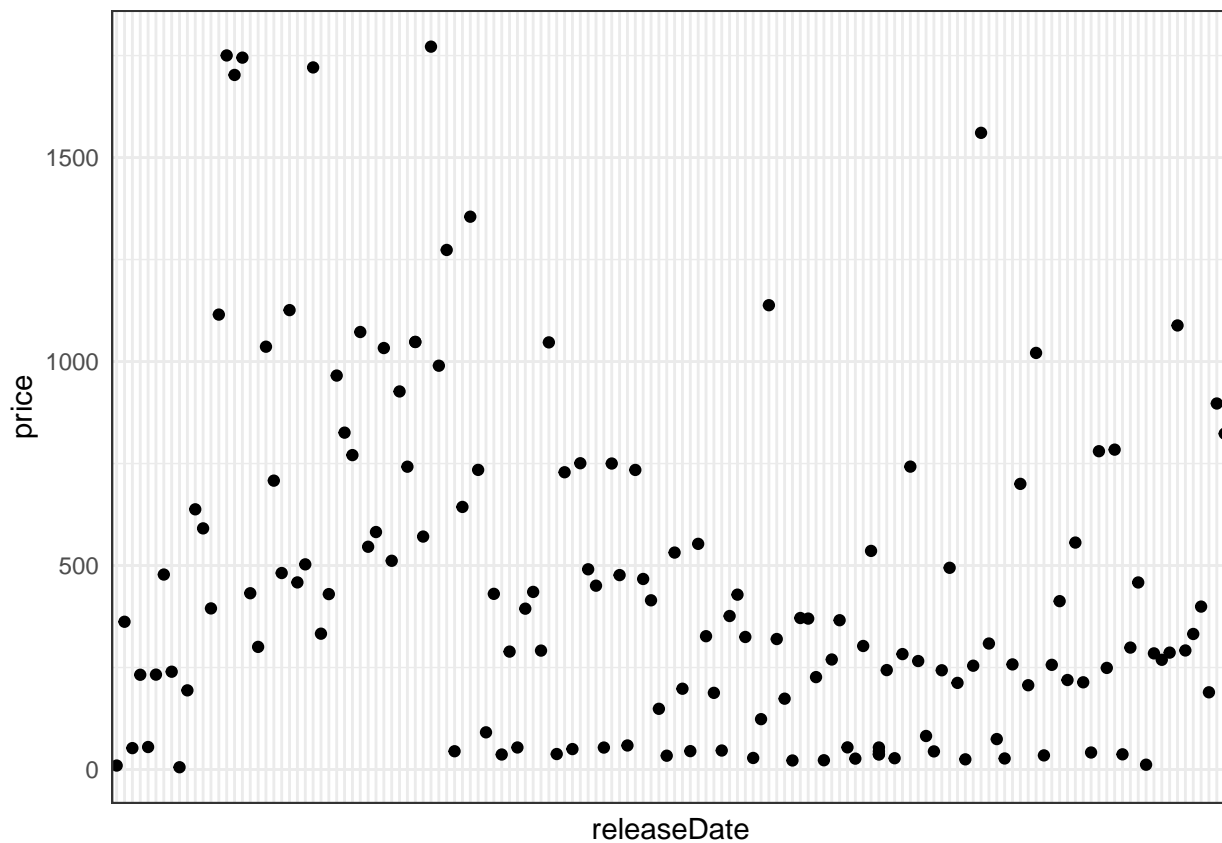
Ecco spiegato anche l'ultimo dubbio: sebbene le carte ad alto costo siano meno richieste nella formazione di mazzi, esse sono tutte carte molto rare, che come abbiamo visto sono caratterizzate da una disponibilità molto limitata, da cui il prezzo mediamente alto. La prima risposta che possiamo offrire alla domanda in esame è quindi: le carte ad alto costo di mana hanno un valore medio molto alto, a causa della propria rarità, nonostante non vengano giocate.

Edizione

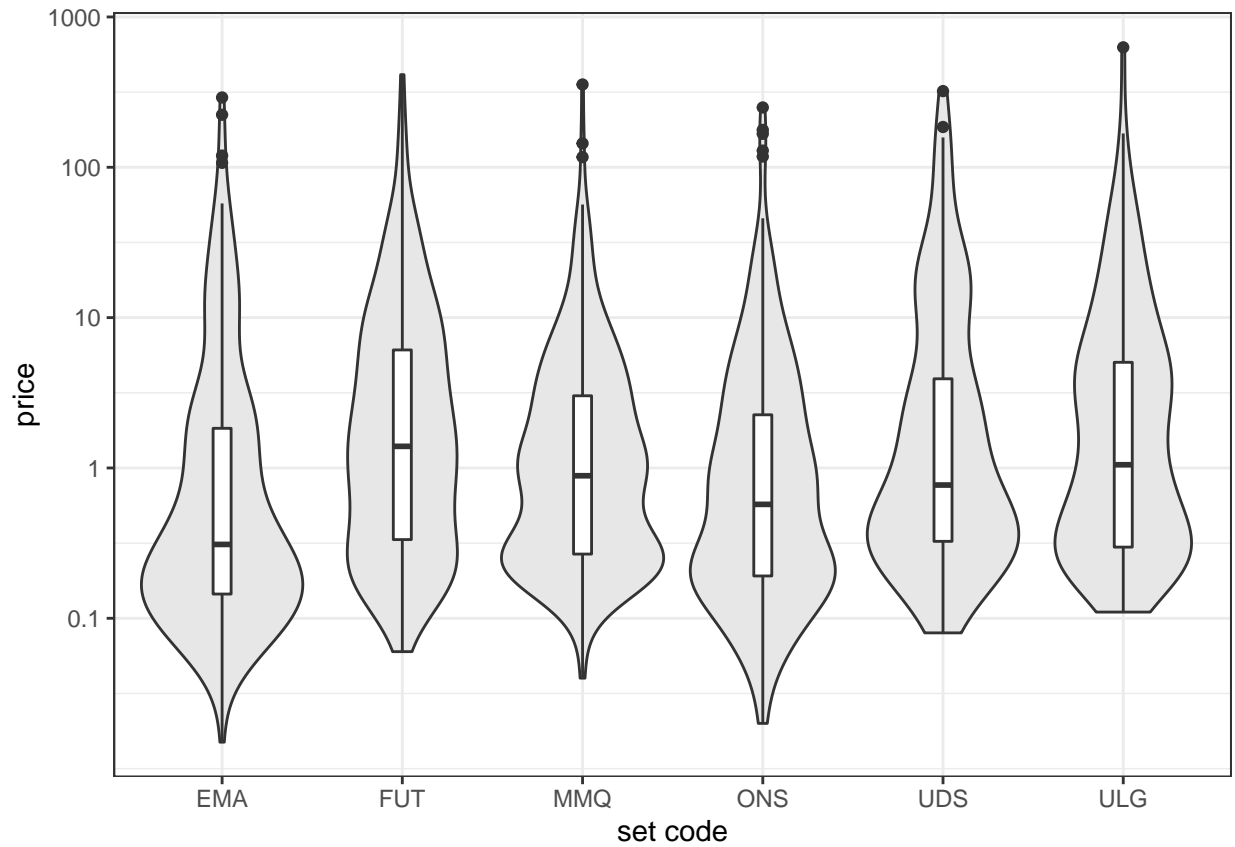
La prossima variabile non numerica che consideriamo è l'edizione: le varie edizioni di carte hanno un valore costante?



Ordinandole per data di rilascio, ci appare chiaro che le prime edizioni abbiano un valore spropositato rispetto alle altre. Questo è noto: quando le tirature erano limitatissime e il gioco non aveva ancora i contorni odierni, alcune carte erano totalmente fuori scala in quanto a rapporto costo-beneficio e non sono mai state ristampate. La somma di impatto enorme e offerta ridottissima hanno generato i prezzi che vediamo. Rimuoviamo questi outlier per poter osservare meglio l'andamento delle altre edizioni.



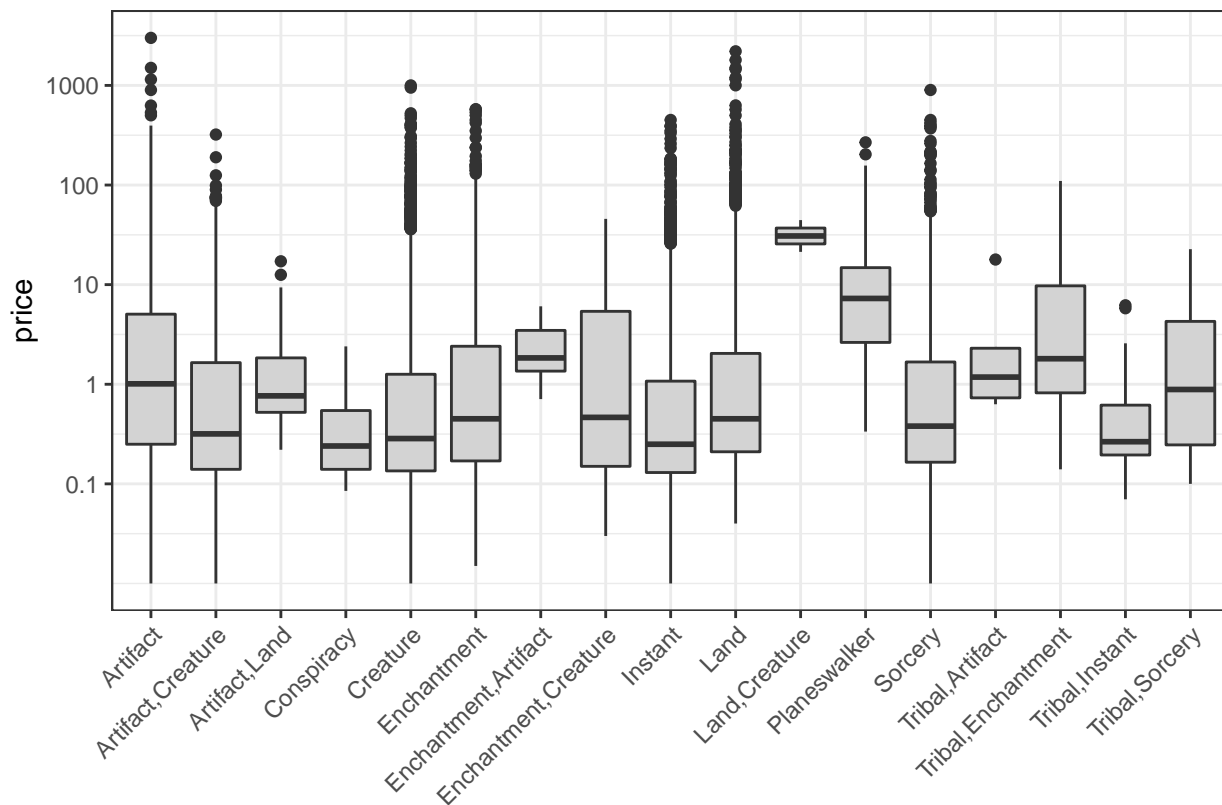
Quello che notiamo immediatamente è che resta una grande variabilità di prezzo tra le edizioni, e questo ci porta alla seguente domanda: le edizioni di prezzo elevato lo sono per la presenza di un numero ridotto di carte, o per un valore medio? In entrambi i casi, cercheremo di estrapolare se possibile la caratteristica di questi sotto insiemi.



Il violin plot ci restituisce una situazione ibrida: le edizioni continuano ad avere mode molto basse, ma mentre alcune hanno solo delle sottili code ad alti valori, altre hanno effettivamente numeri più consistenti. Questa condizione rende difficile stabilire la presenza e l'origine di eventuali pattern nei risultati basandosi sulle edizioni.

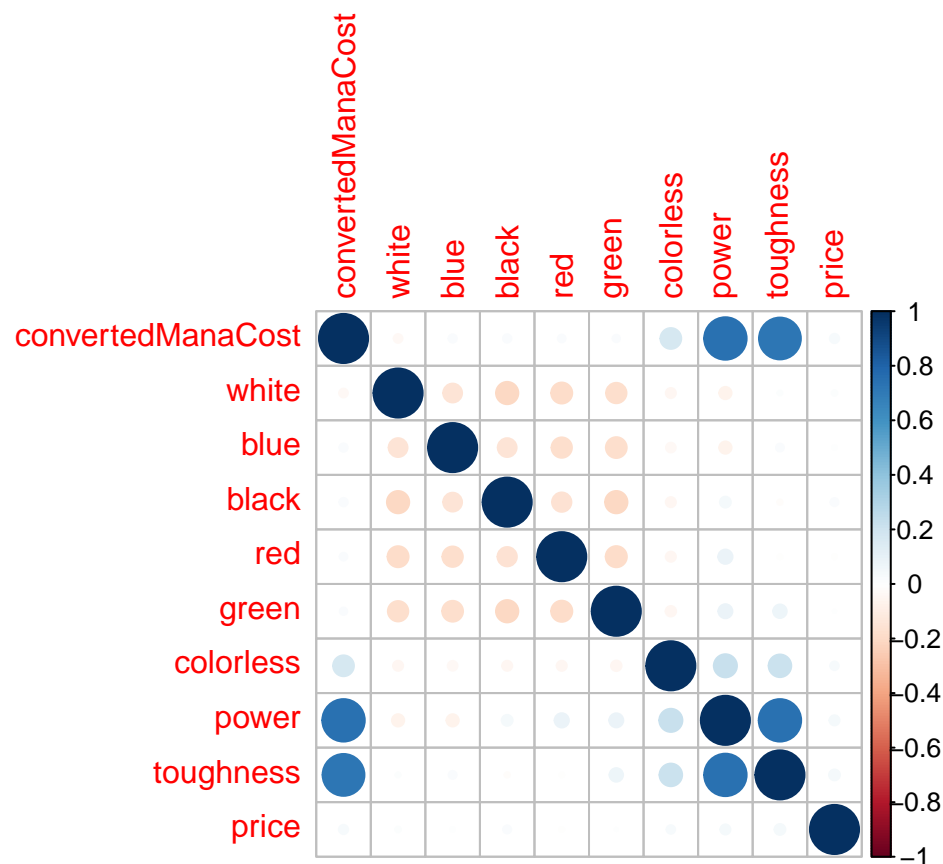
Tipo

L'ultima analisi rimasta è quella relativa al tipo delle carte.



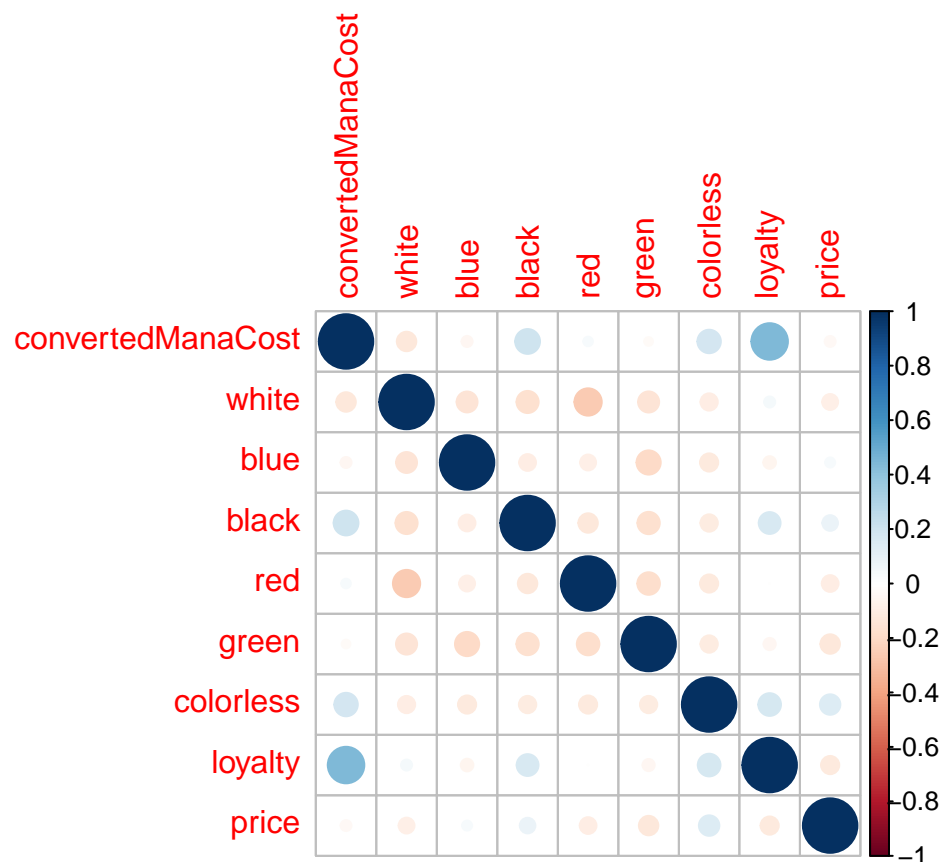
Questo grafico ci dà effettivamente dei risultati interessanti: Notiamo infatti come, sebbene senza dubbio influenzato dalla ridotta quantità di elementi, la combinazione di tipi Terra-Creatura comporti un prezzo medio elevatissimo, e in misura minore anche le combinazioni Terra-Artefatto, Incantesimo-Artefatto e Tribale-Artefatto si discostano nettamente dal resto del panorama. Già ad un primo impatto abbiamo una risposta alquanto soddisfacente: Il tipo è un buon indice del prezzo delle carte e in particolare lo sono le seguenti combinazioni: Terra-Creatura, Terra-Artefatto, Incantesimo-Artefatto e Tribale-Artefatto.

Scendiamo nello specifico dei tipi che comportano nuove caratteristiche per le carte, ovvero Creatura, che introduce gli attributi Forza e Costituzione, e Planeswalker, che introduce Lealtà.



Le forti correlazioni sono alquanto triviali: ci dicono che carte dall'alto costo di mana tendono ad avere alta forza e costituzione, e che queste vanno di pari passo tra loro. Impariamo anche essere presente una correlazione tra la mancanza di colore e il costo di mana, ovvero che le carte senza colore tendono ad essere leggermente più costose (in partita) delle altre. Questo dato è interessante ed è probabilmente una scelta precisa degli ideatori del gioco per non rendere carte senza colore troppo semplici da lanciare per i maghi, dato che esse non richiedono specificità sul tipo di carte terra richieste per poter essere giocate. Purtroppo nulla emerge per quanto riguarda il prezzo.

Passiamo allora ai planeswalker.



Come per le creature, il legame costo di mana - lealtà è alquanto triviale, si affacciano però leggere correlazioni di diversi attributi con il prezzo, attributi che tra loro non dimostrano avere forti legami. Queste sono condizioni promettenti per un modello di regressione multipla.

```
##
## Call:
## lm(formula = price ~ white + blue + black + red + green + colorless +
##     loyalty, data = planeswalkers.numeric)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -39.870 -12.445  -6.783   2.196 242.425
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  37.9536     6.0786   6.244 1.05e-09 ***
## white        -7.3369     3.6049  -2.035  0.04246 *
## blue         -0.7658     3.2403  -0.236  0.81328
## black         3.6386     3.4715   1.048  0.29519
## red          -7.9421     3.3809  -2.349  0.01928 *
## green        -9.8774     3.5301  -2.798  0.00538 **
## colorless    22.0355     8.8665   2.485  0.01334 *
## loyalty      -3.7911     1.2661  -2.994  0.00291 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 27.32 on 418 degrees of freedom
```

```
## (3 observations deleted due to missingness)
## Multiple R-squared:  0.07986,    Adjusted R-squared:  0.06445
## F-statistic: 5.182 on 7 and 418 DF,  p-value: 1.121e-05
```

Il test F è significativo, ma un R-quadro di appena 0.08 non rende questo un modello adatto a spiegare la variabile dipendente prezzo. Purtroppo il modello multiplo non aiuta nella previsione del prezzo delle carte. I passi successivi potrebbero essere una regressione con modelli non lineari, o l'analisi delle serie storiche dei prezzi, caratteristica pressochè fondamentale in ogni tecnica di previsione del prezzo di un bene. Entrambe queste tecniche sono però al di fuori degli scopi della presente trattazione, e vengono perciò riservate a futuri approfondimenti.