

The Data Science of Board Games

Scott Burger

Data Scientist, Microsoft



This talk

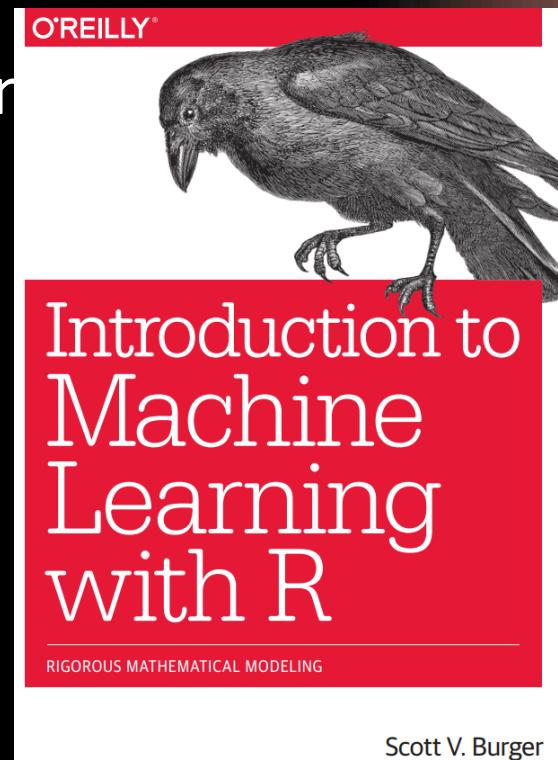
- Intro to modern board games
- The BoardGameGeek database
- Top dogs and bottom feeders
- Measuring the ‘most popular’ games
- Ratings and game ‘behaviors’
- Designing a way to measure board game balance



UNPAKER

About me

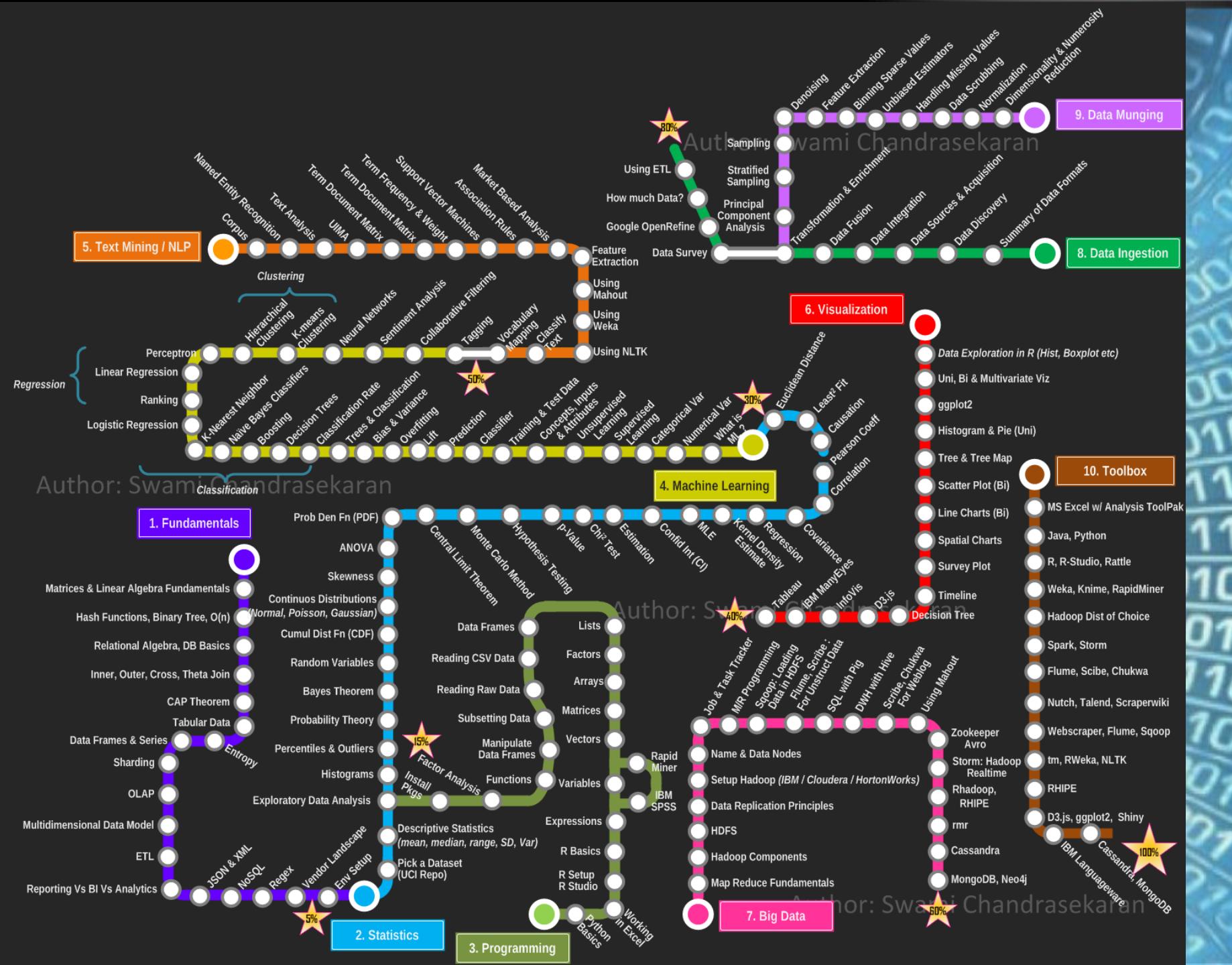
- Western Washington Univ - Physics
- University College, London - Astrophysics
- Data Scientist - Microsoft
- Race bikes on the Seattle Cancer Care Alliance
- Writing book on machine learning (2018)
- Like games of all kinds (digital, analog)
- Blog about it all on svburger.com



What is data science?

- Catch-all for scientific approaches to using data to solve questions
 - 90% just data manipulation problems
 - 4% statistics
 - 4% machine learning
 - 2% other
-
- Typical tools: SQL, Python, R, Excel
 - Tools used here: R (code online)





The Gateway

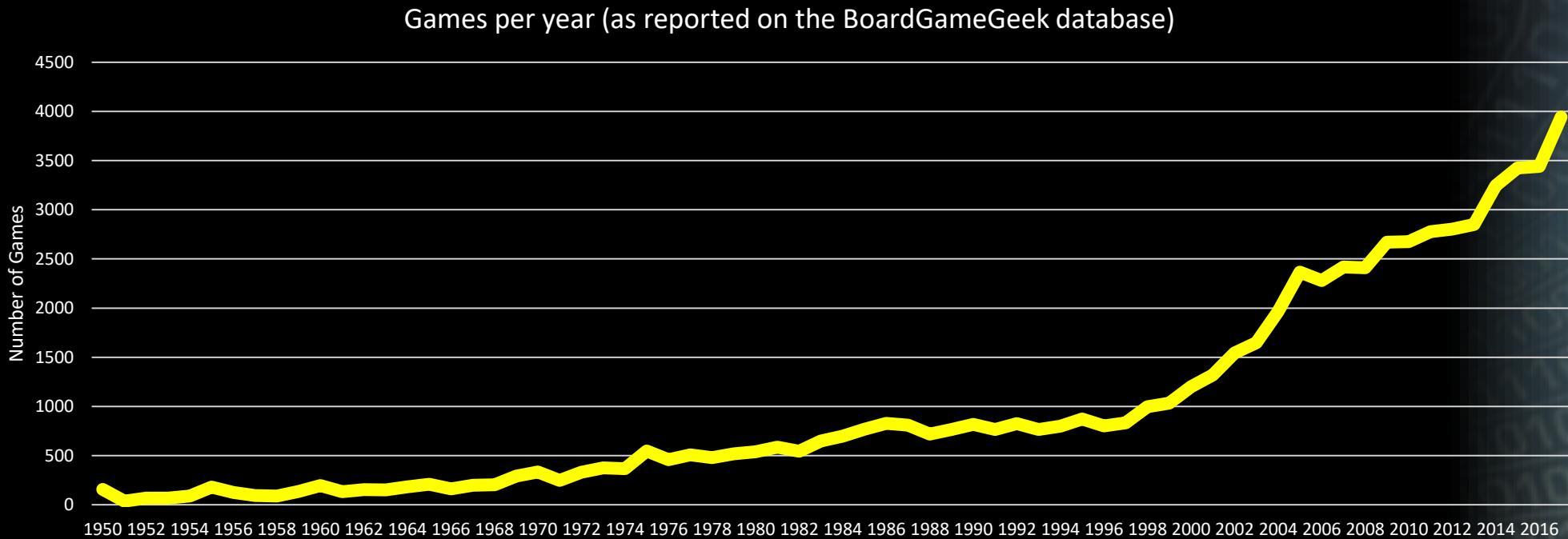
- Risk, Stratego having to compete with prime SNES and N64 games
- Had played games like Life, Sorry!, Mousetrap, Battleship, but never really ‘hooked’
- Until that is... Power Grid
- Showed that games could be difficult, different, deterministic
- So what defines a ‘gateway’ game exactly?
- Rough criteria: simple to learn, well-rated
- Examples to follow!



Why Board Games?

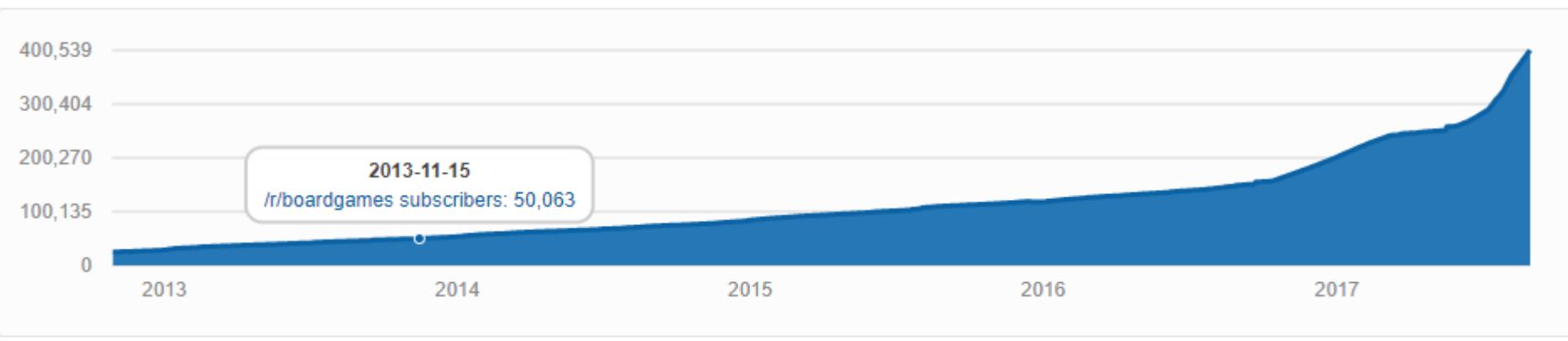
- The number of board game releases has seen an explosive growth in the past few years
- Games released since 2006 account for **52%** of all board games made since 1950
- Big advancements in the different types of games developed (kickstarter, legacy games)
- Explosion in popularity reflected in social sites like reddit





/r/boardgames metrics (boardgames)

The #1 reddit source for news, information, and discussion about modern board games. Join our community! Come discuss games like Codenames, Eldritch Horror, Terra Mystica, and **your** favorite games!



Why BoardGameGeek.com?

- Huge amount of data: over 500k items (games, expansions, other stuff)
- Lots of great info from player ratings, number of plays, game complexity, if a game is recommended for certain numbers of players
- Enthusiast community keeps the database on the cutting edge
- Cons: enthusiast community biased towards hottest new releases, your favorite game might not be their favorite game, lots of noise in the dataset that needs to be cleaned out

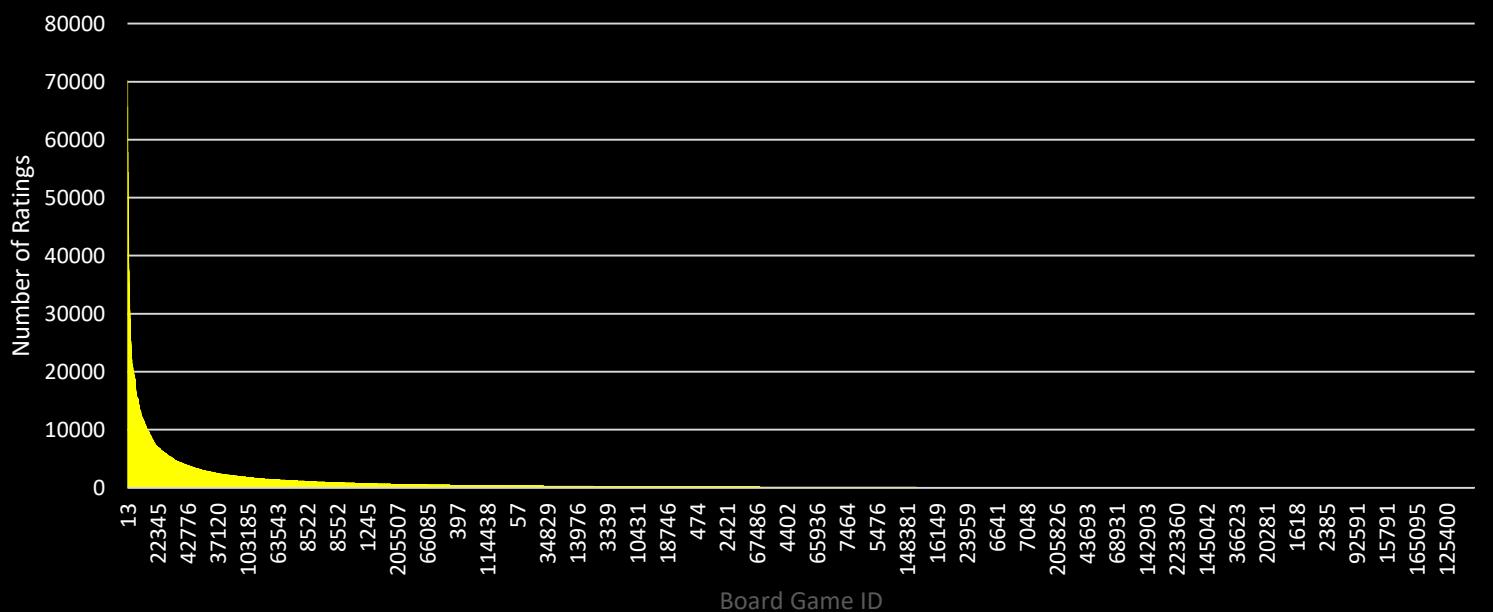


The Top Dogs and Bottom Feeders

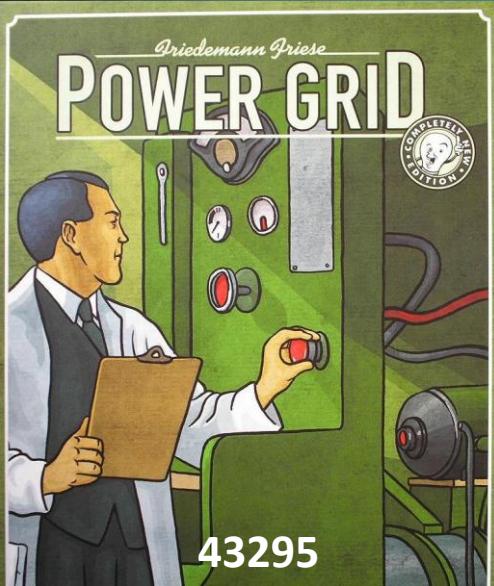
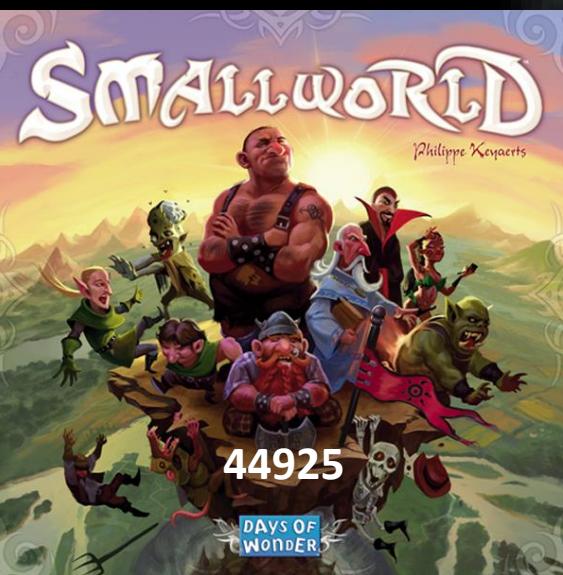
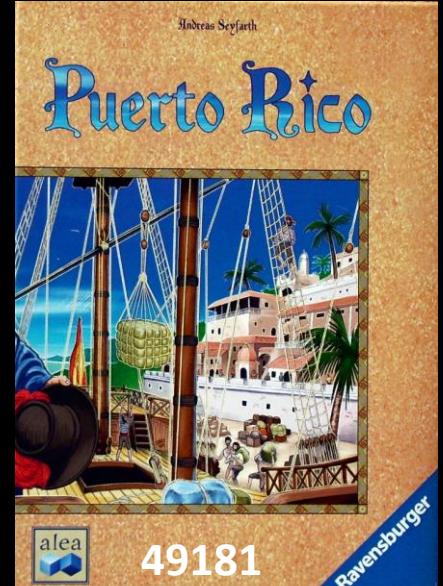
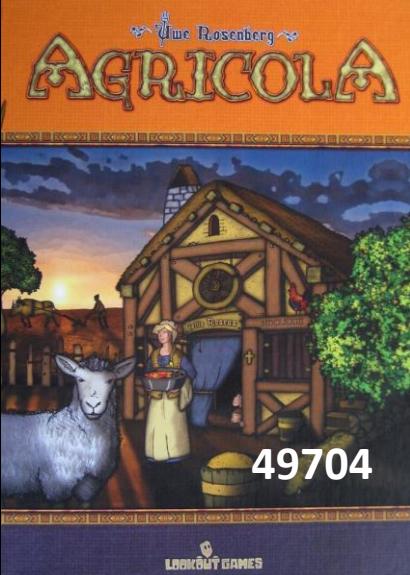
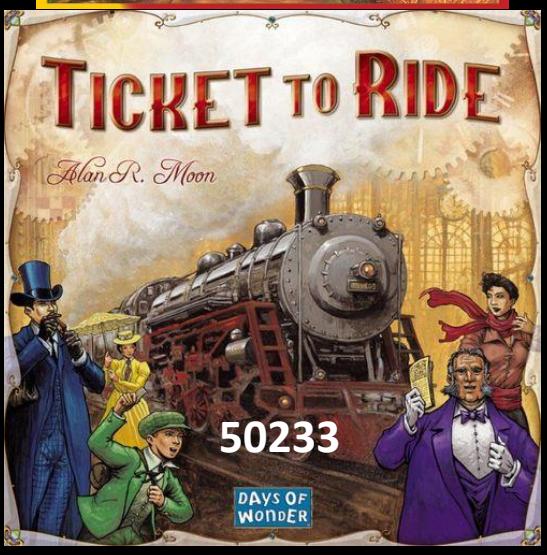
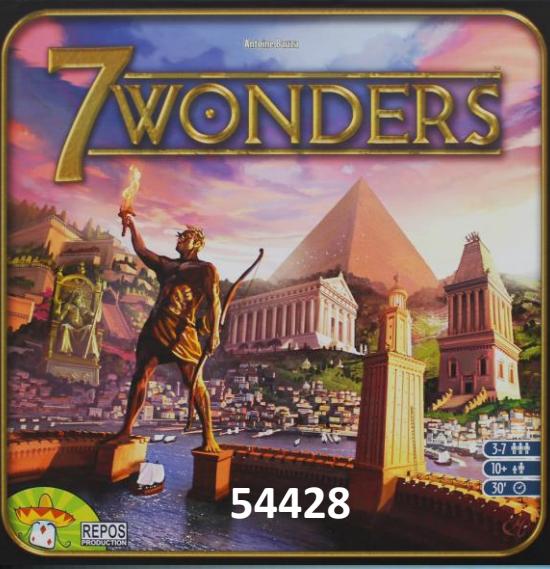
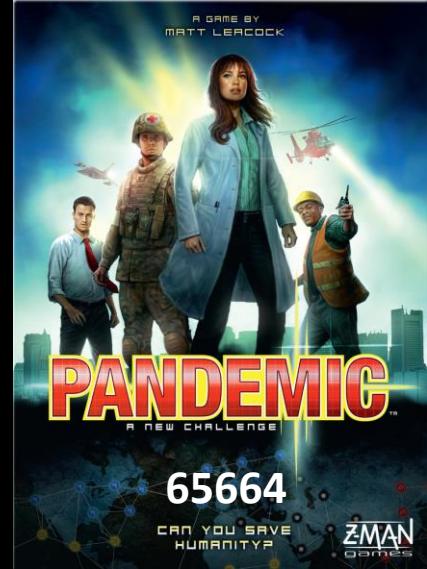
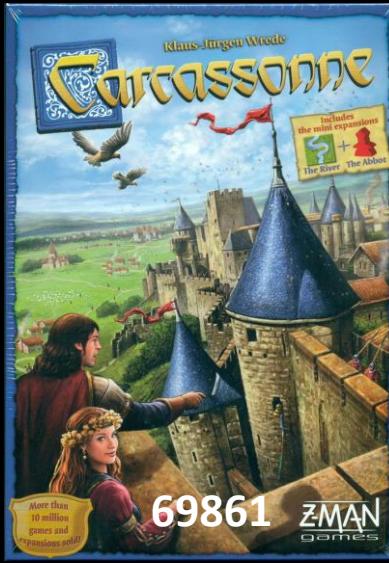
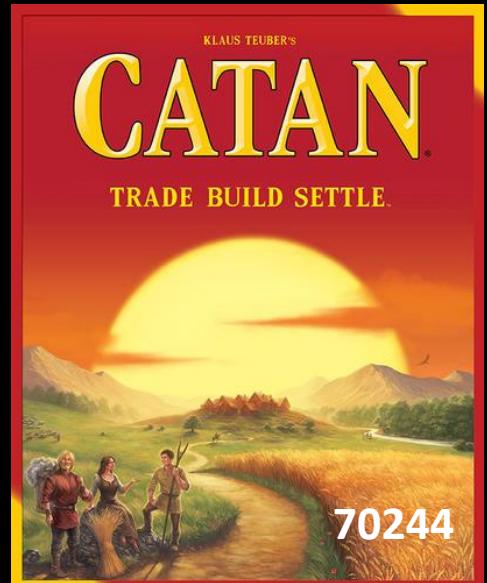


What are the most popular games?

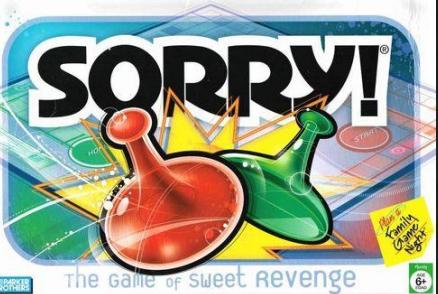
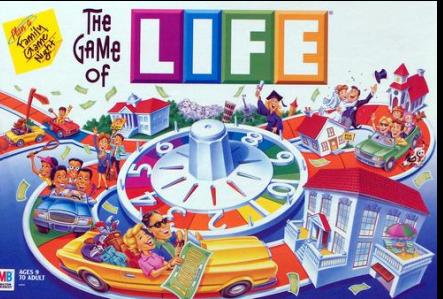
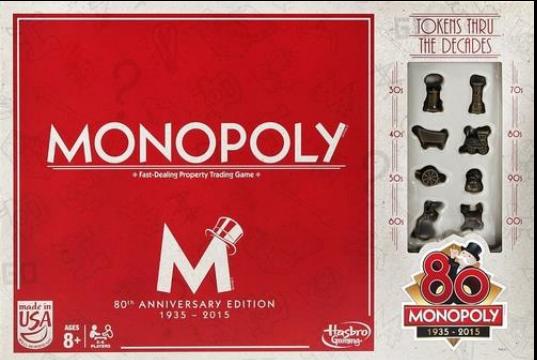
- Games with the highest number of ratings show the current landscape of what's been consistently the most popular
- Issue: very long tailed dataset



Top 10 by number of ratings



Top 10 owned games with ratings less than 5.0



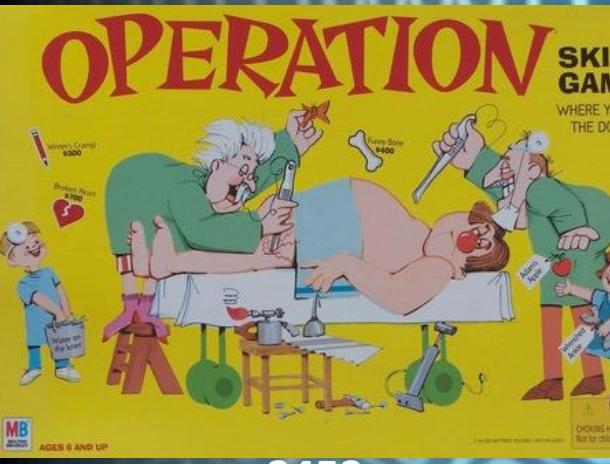
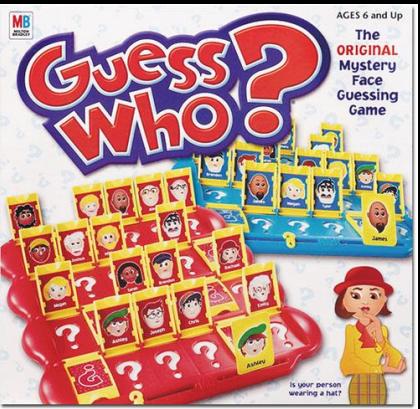
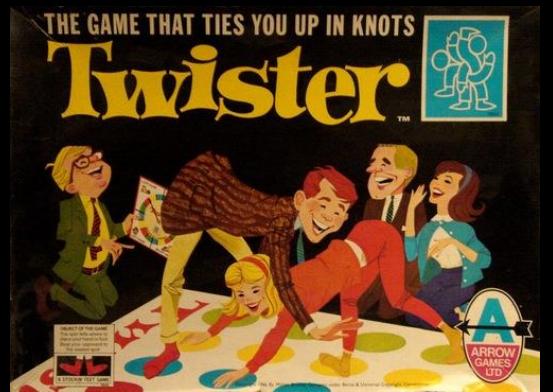
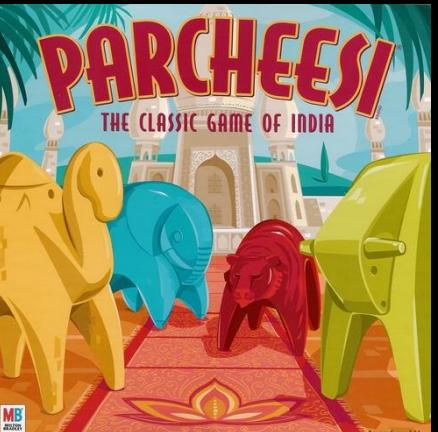
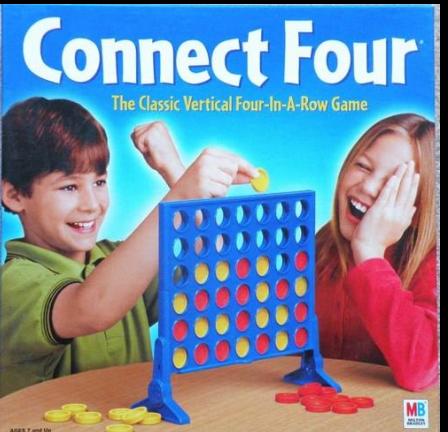
23539

9630

7543

5998

5858



5182

4752

3829

3794

3453

[–] realzequel  118 points 5 days ago

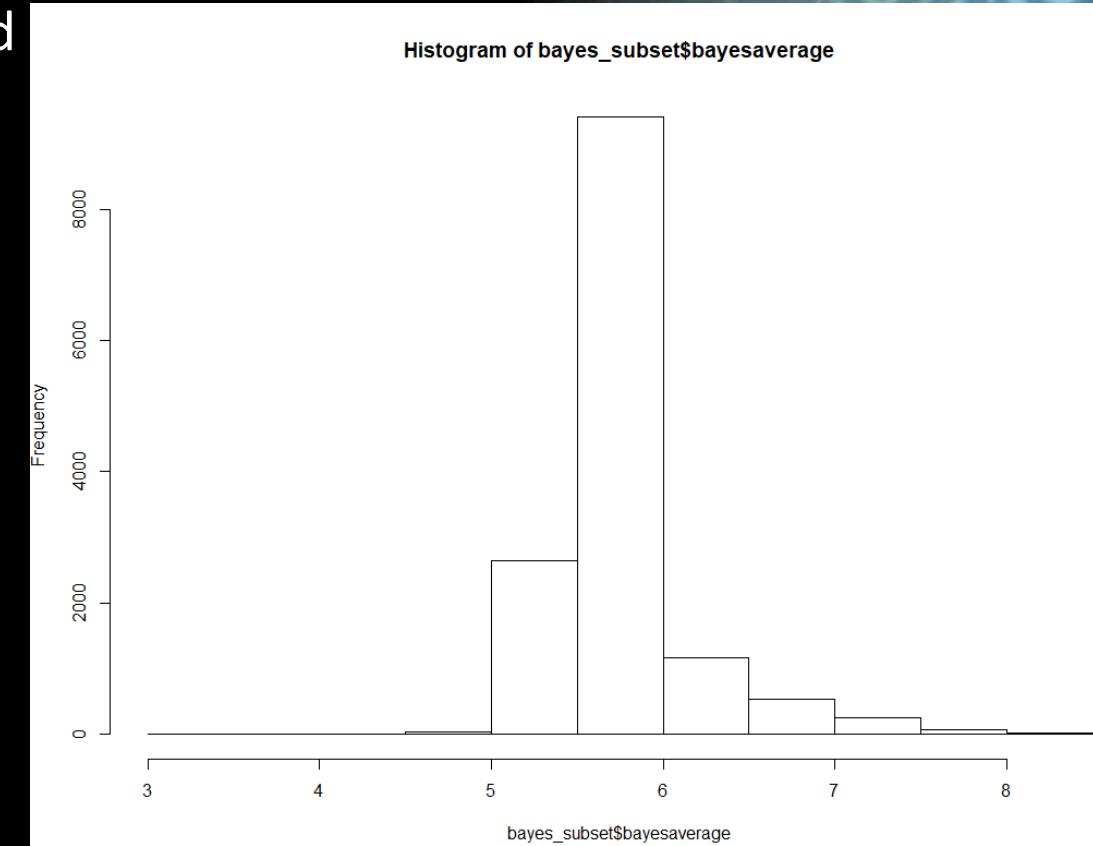
Well we don't play in a vaccuum, other non-gamers will play games they know. Even myself, as a hobbyist, will not turn down a game of:

- Sorry!
- Risk
- Scrabble
- Stratego
- Careers
- Pictionary

And I'll play Battleship or Clue with the kids. Just because games aren't my favorites or well-designed doesn't mean they're not fun.

Caveats

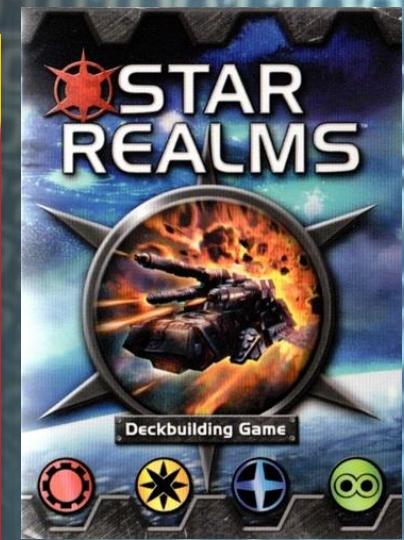
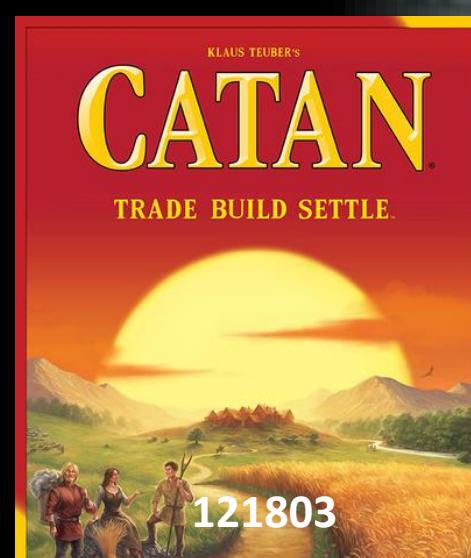
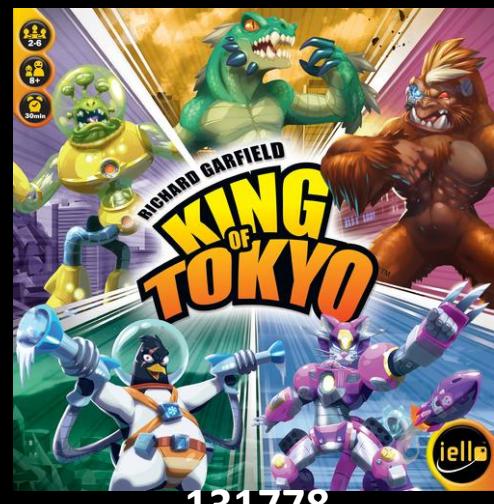
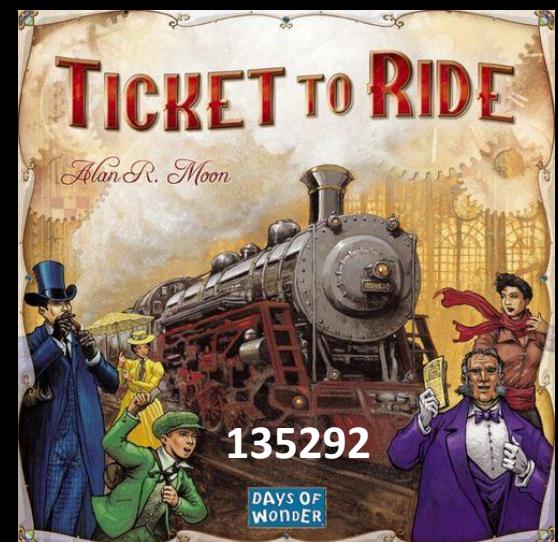
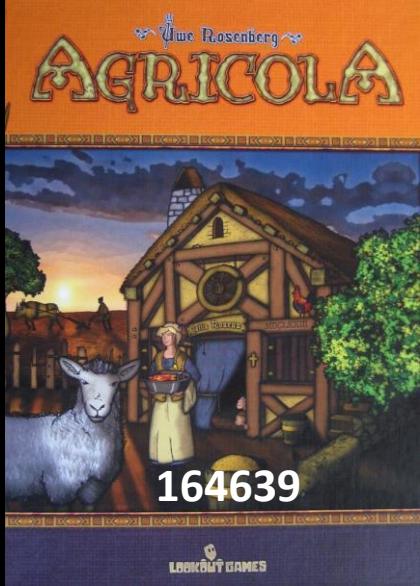
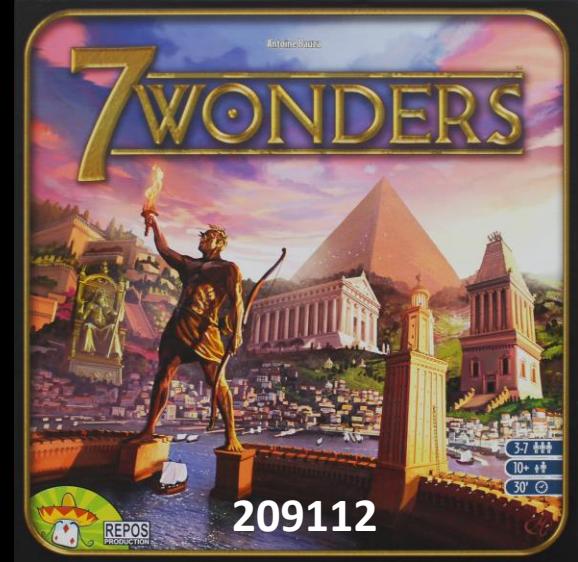
- BGG database is skewed heavily to enthusiasts and hobbyists at the top end
- Global ratings are, surprisingly, well balanced!
- “IGN effect”: a 7 rated game shouldn’t be thought of as ‘bad’
 - 1: Awful – defies game description
 - 2: Very bad – won’t ever play again
 - 3: Bad – likely won’t play this again
 - 4: Not so good – but could play again
 - 5: Mediocre – take it or leave it
 - 6: Ok – will play if in the mood
 - 7: Good – usually willing to play
 - 8: Very good – enjoy playing and would suggest it
 - 9: Excellent – very much enjoy playing
 - 10: Outstanding – will always enjoy playing



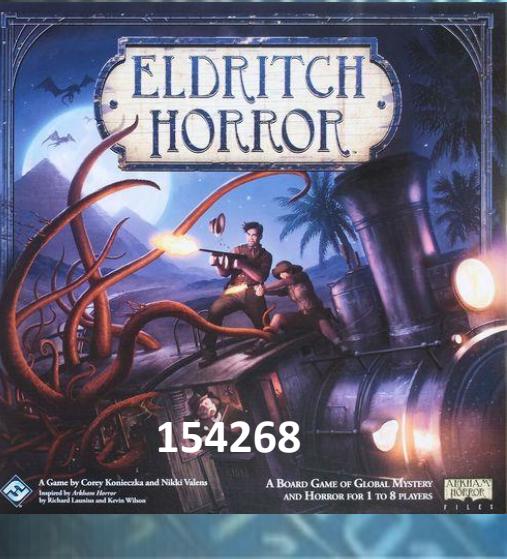
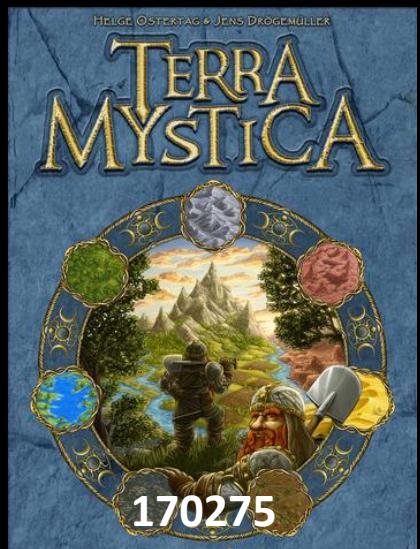
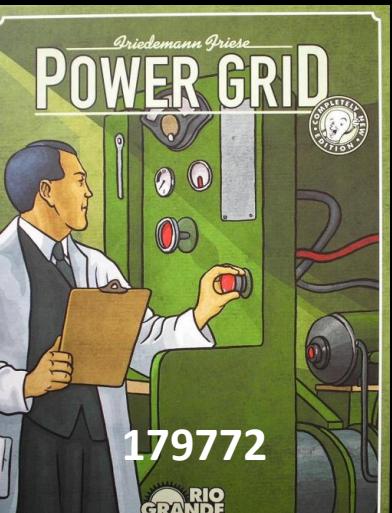
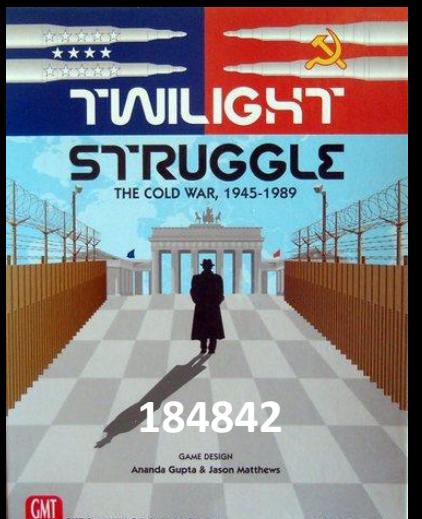
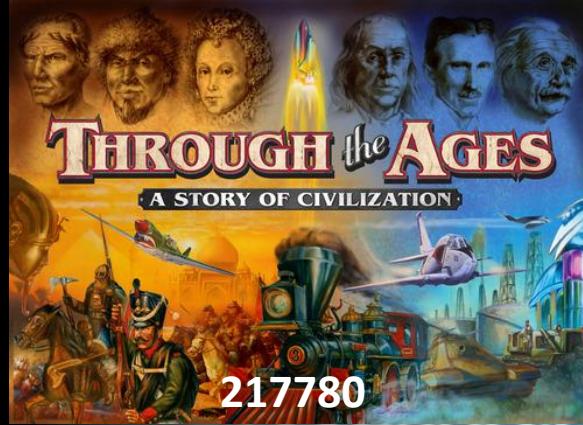
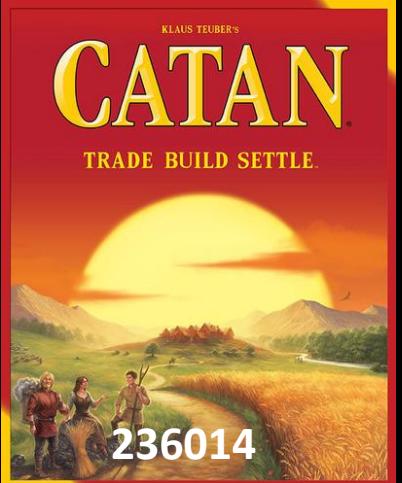
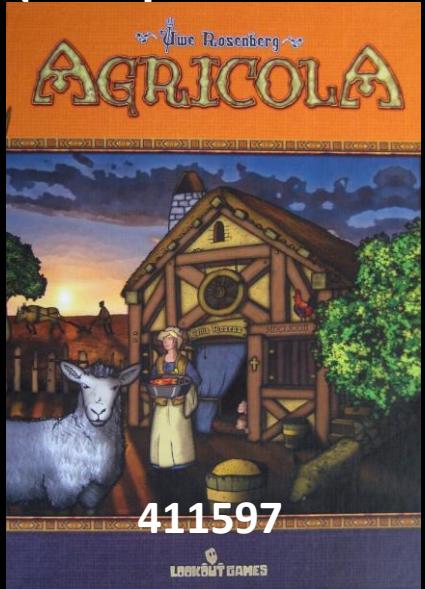
Games and Popularity



Measuring Popularity (top 10 by play count)



Measuring Popularity (top 10 by playtime hours)



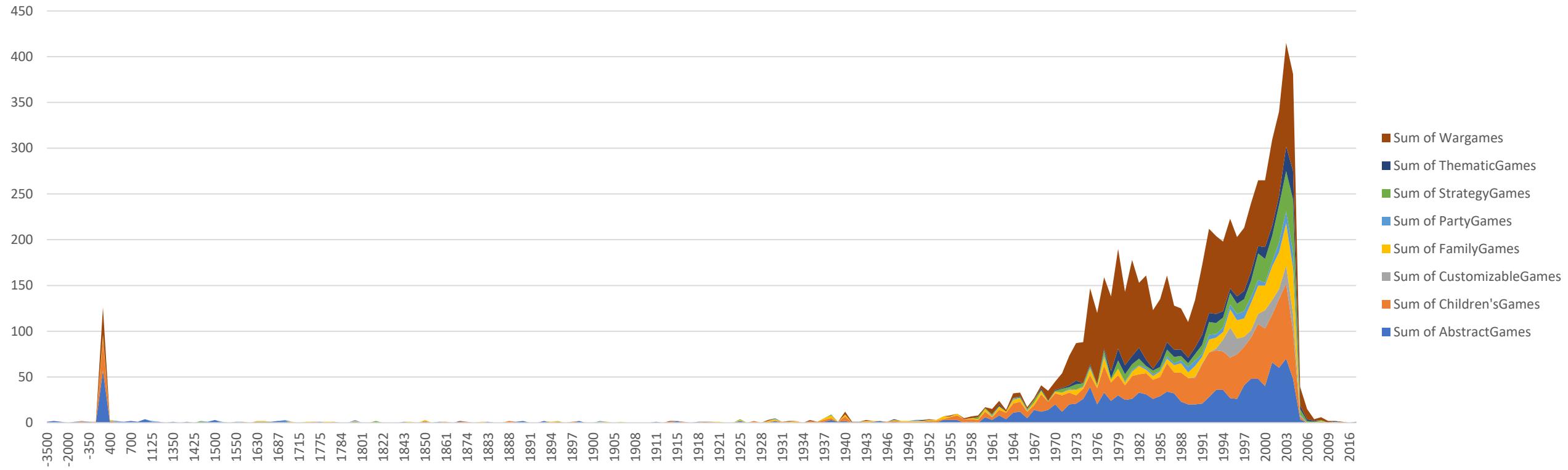
Evolving Nature of Board Games



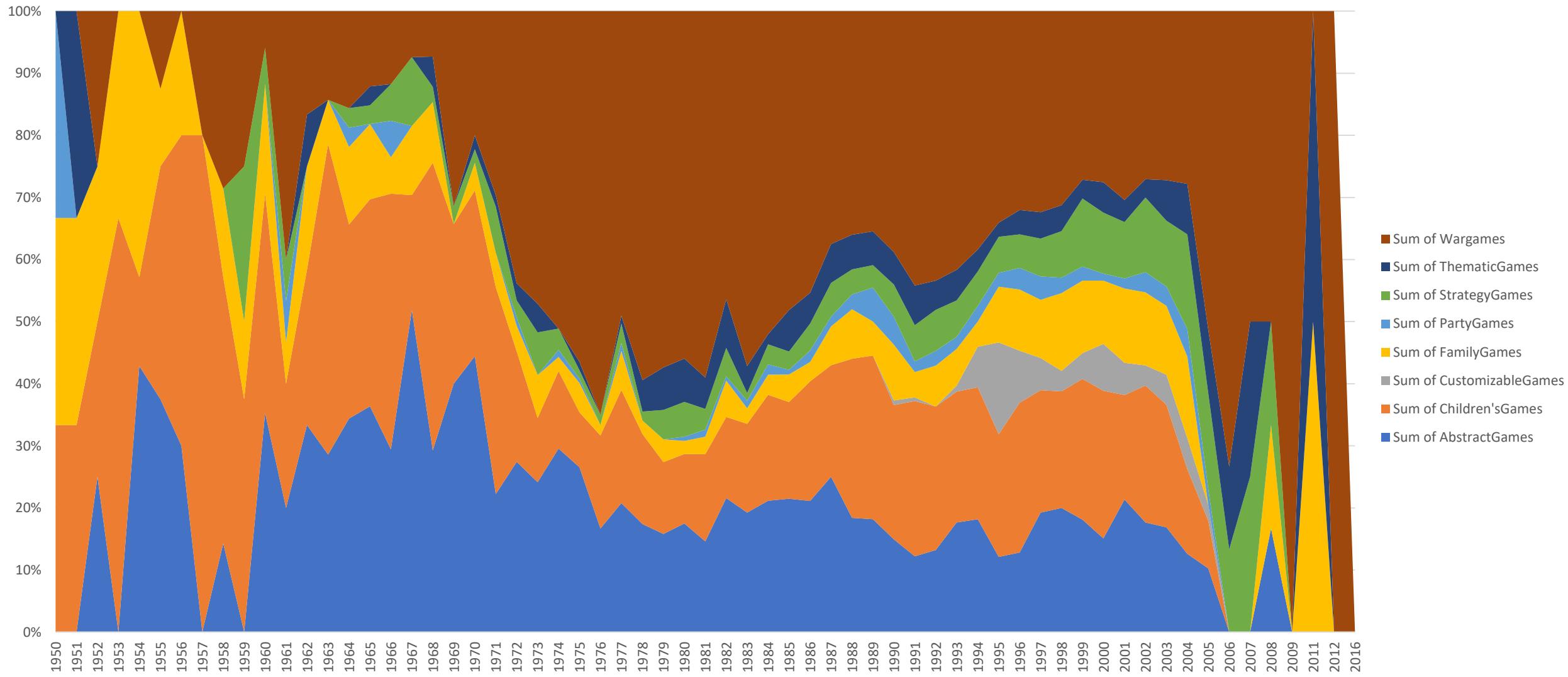
Genres over time

- Type, category, mechanics since 1950
 - Genres are spotty before then
 - And for some reason after 2006???
- The landscape has been mostly stable
- Information overload with number of categories to choose from
 - Type: 8
 - Categories: 84
 - Mechanics: 52

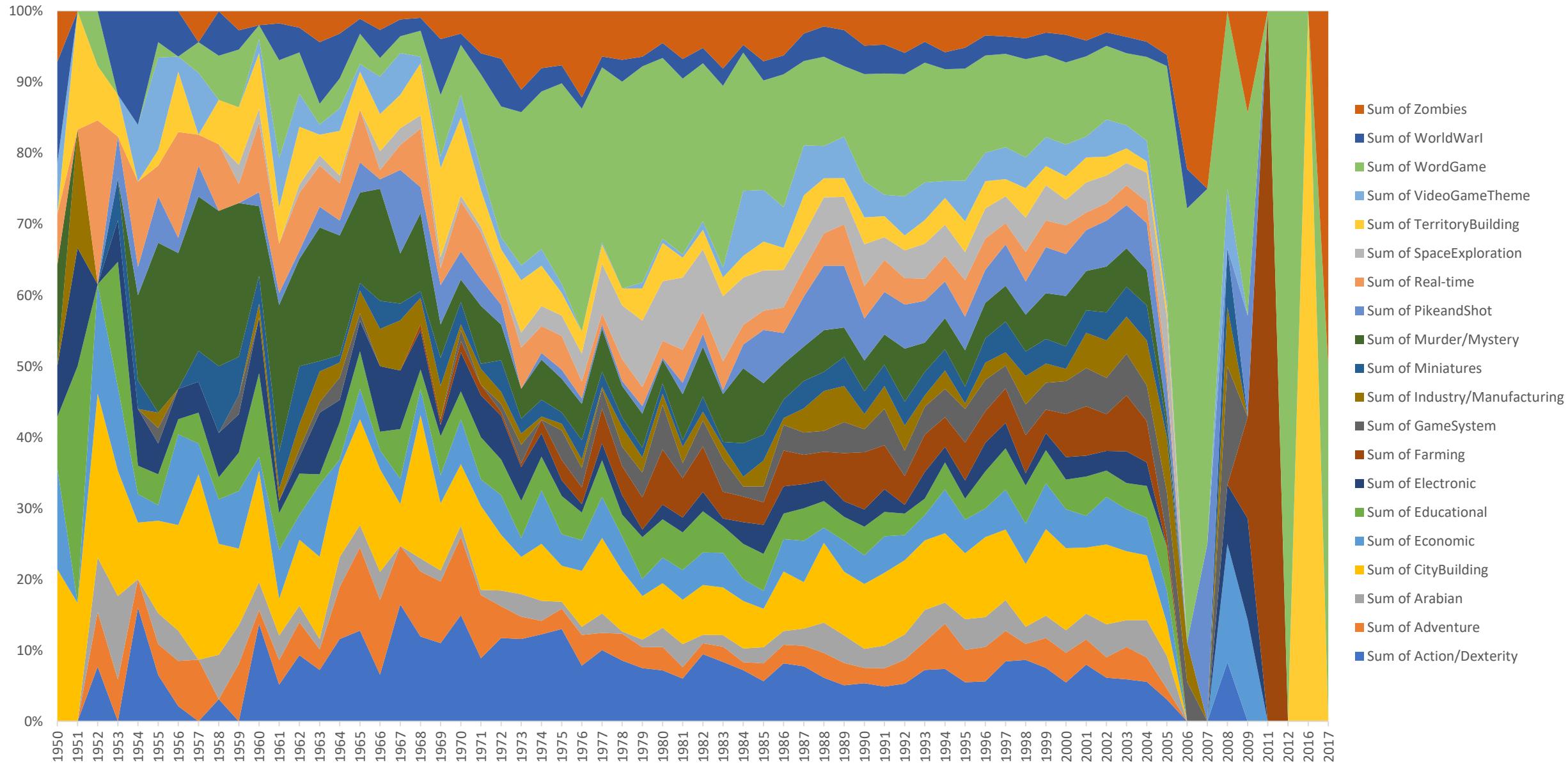
Game Types By All Years



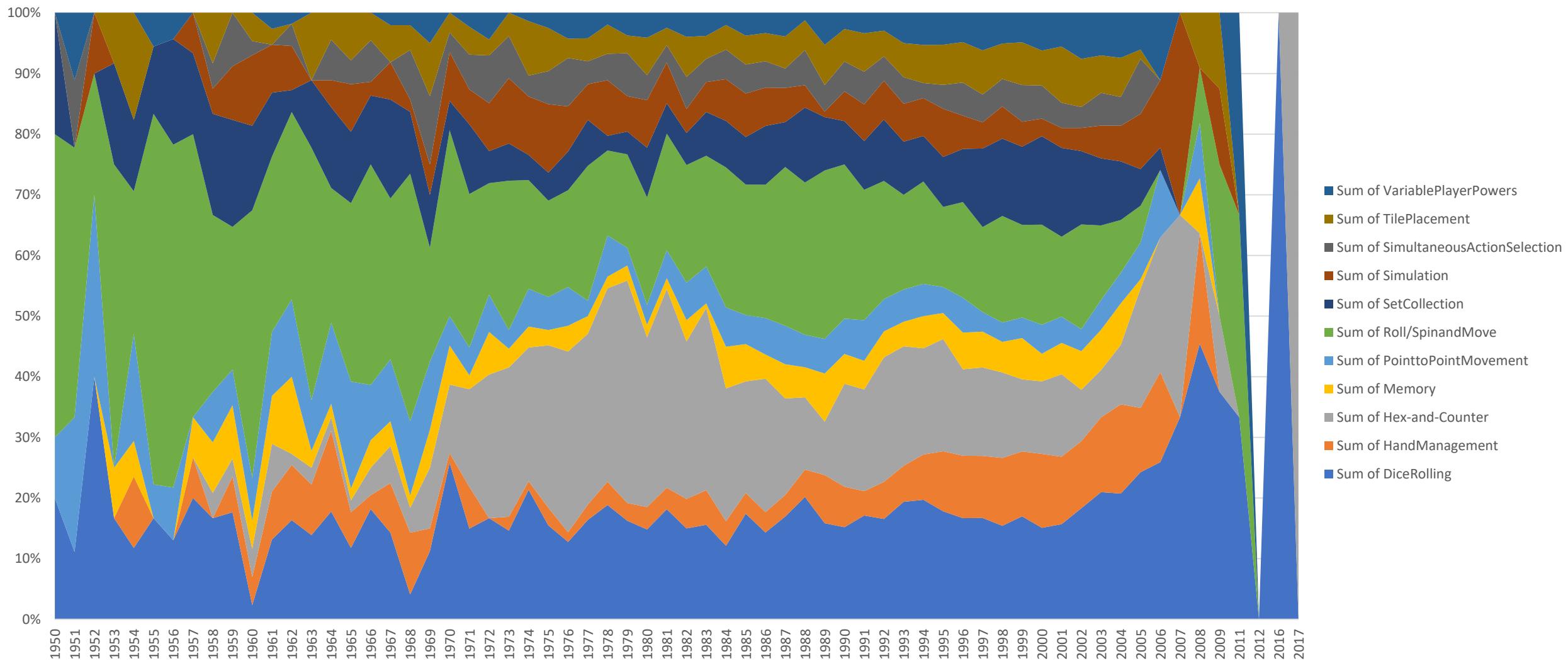
Game Types: 1950-2017 (stacked)



Top 20 Game Categories: 1950-2017 (stacked)



Top 10 Game Mechanics: 1950-2017 (stacked)



Modelling Ratings and Ownership

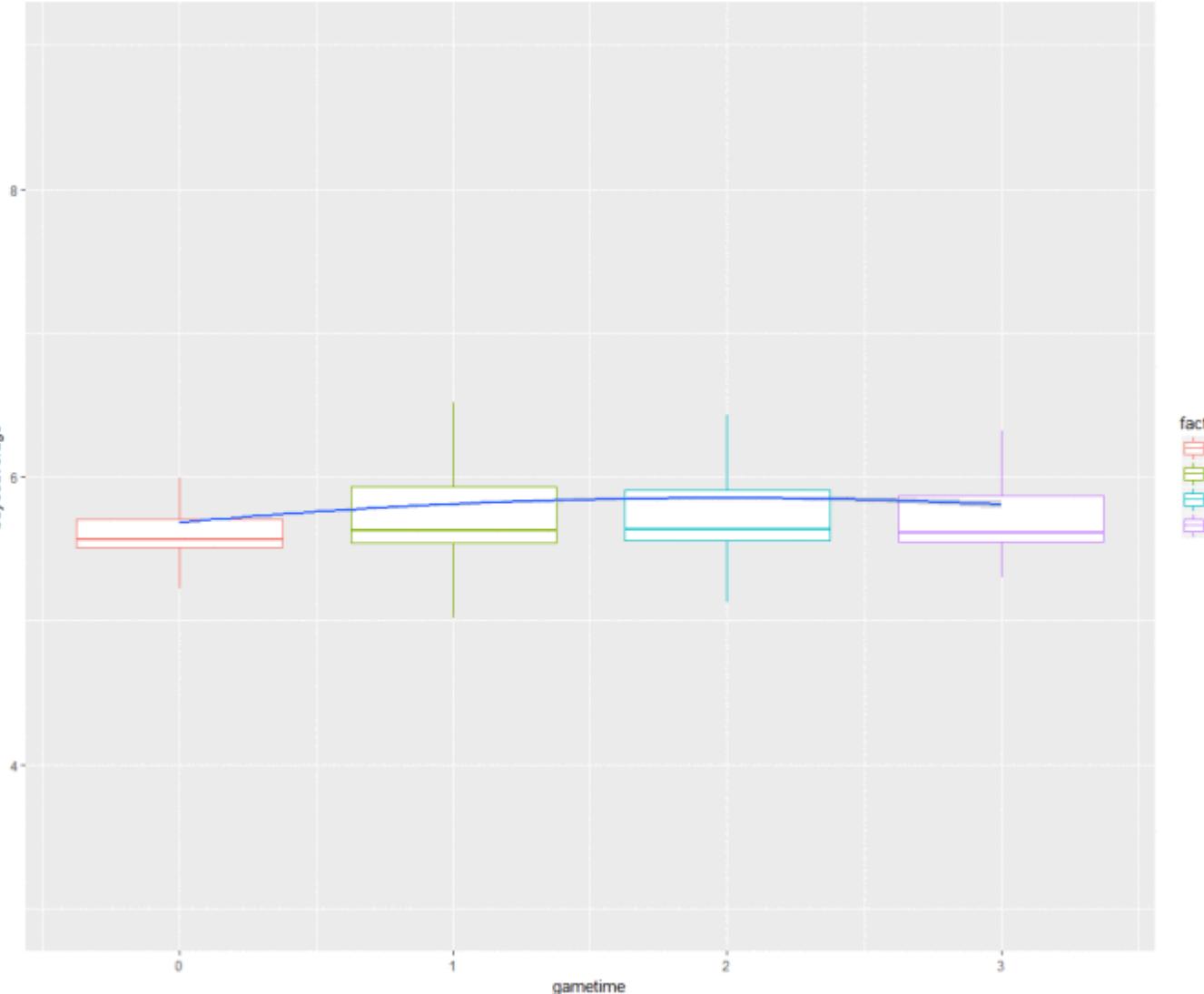
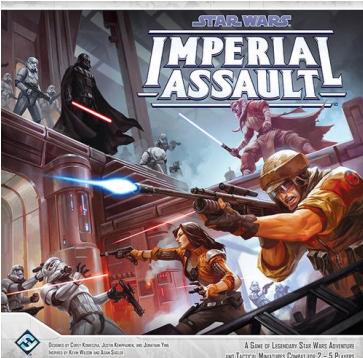
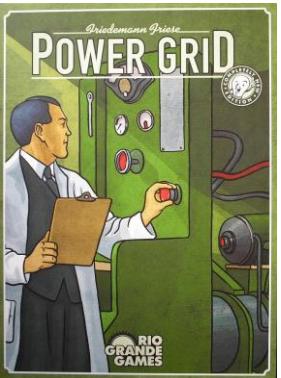


Game Modelling

- How do we model the relationship between ratings and game length?
- Visual: box plot, binned by game length in hours
- Statistical: fit a curve and find the peak
- Done with R and *ggplot()* package

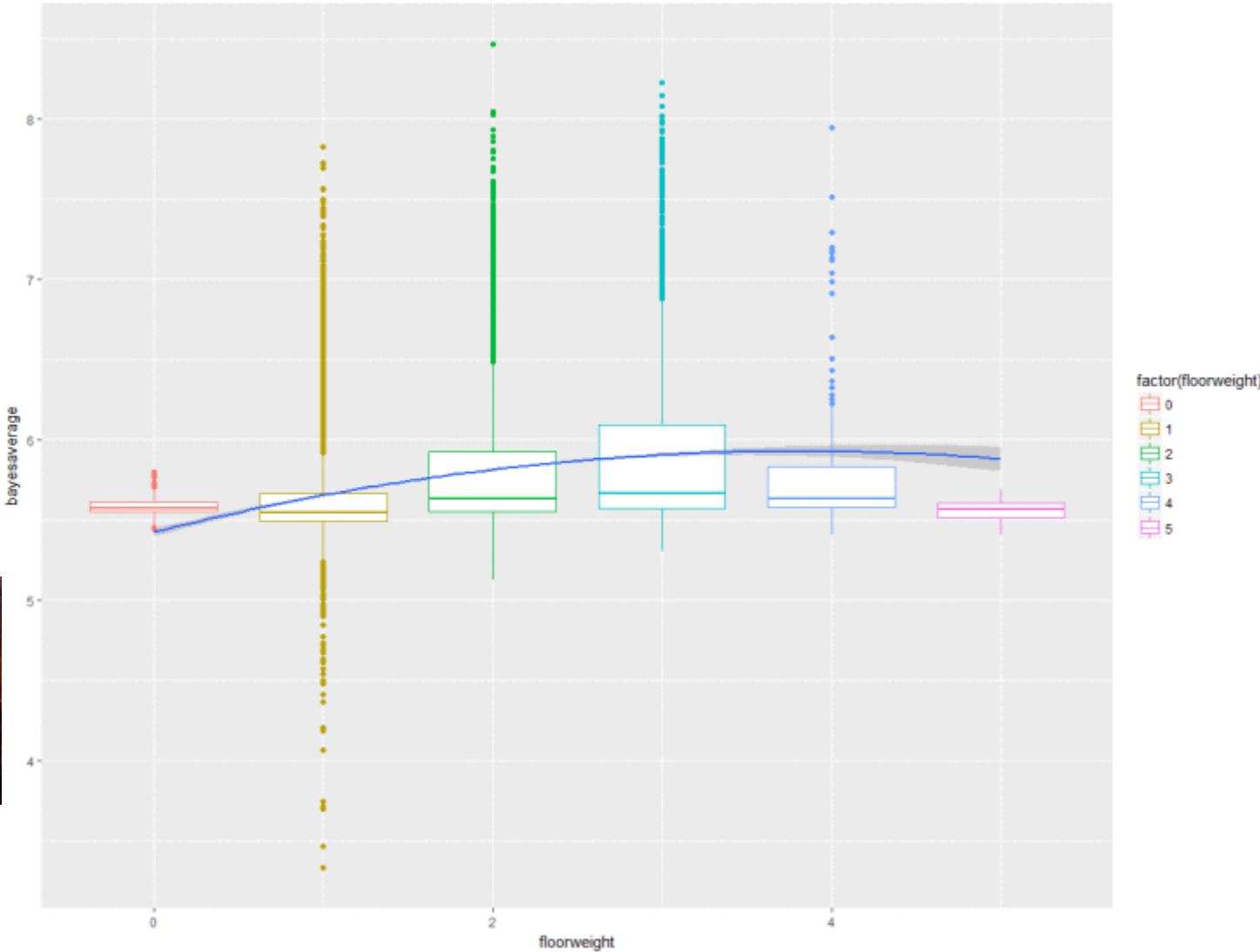
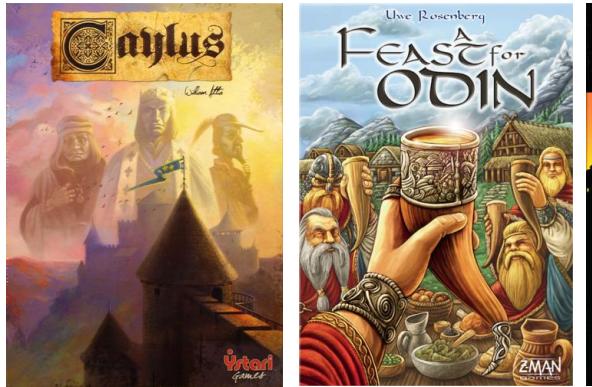
Rating as a function of playtime in hours

- Games ratings peak at 1.987 hours



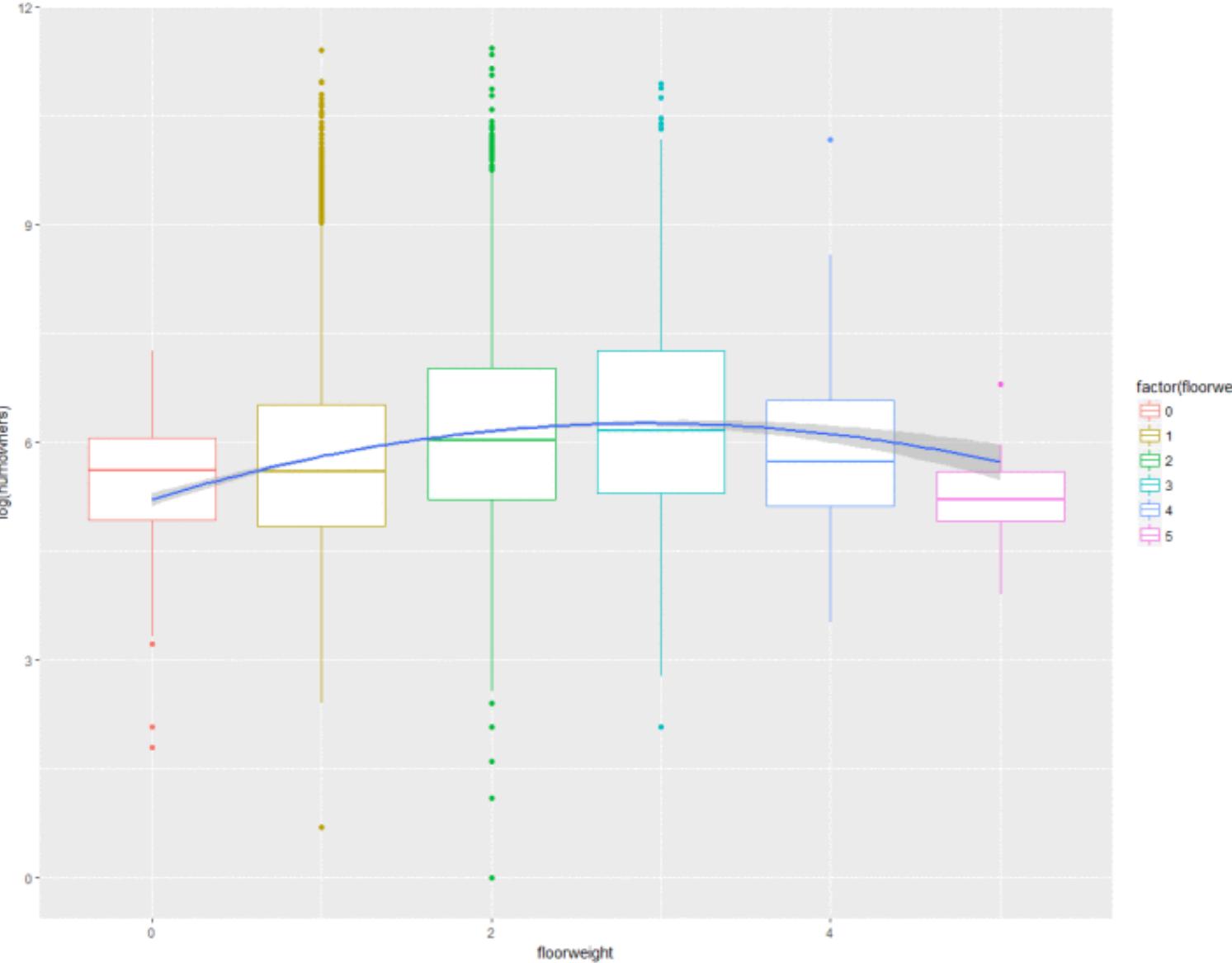
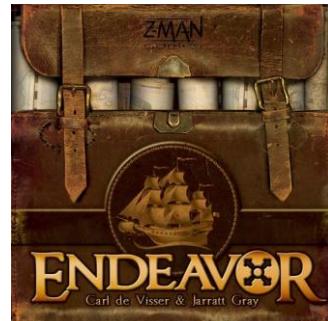
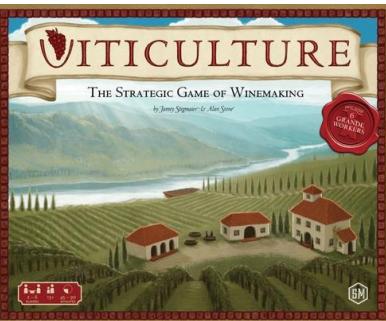
Rating as a function of game complexity

- Rating vs complexity peaks at 3.8233 / 5



Number of owners as a function of game complexity

- Peaks at 2.92 / 5



Games and Player Balance



Game balance

- “What’s a game that plays better with more people?”
 - Flawed statement – regression could start at 0 and go up to 50% recommended at 8 players and have a wild growth to it
- Reformulated question: “what’s a game that plays the best across all the players that its capable of supporting?”
- How do we define game balance?

Game balance

- Easier to discern between something that's recommended or not than between 'best' and 'recommended'
- Just take standard deviation of (1 – “not recommended %”) at each player count
- Most balanced games? Depends who you ask (or how many you ask)

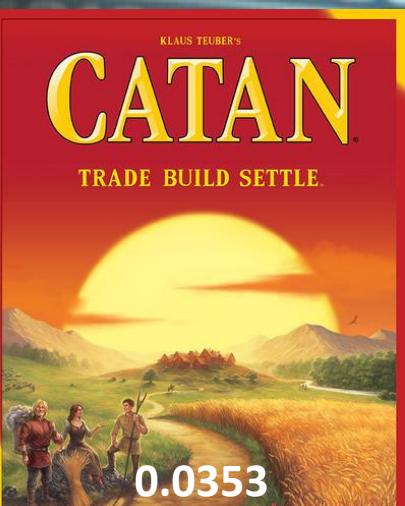
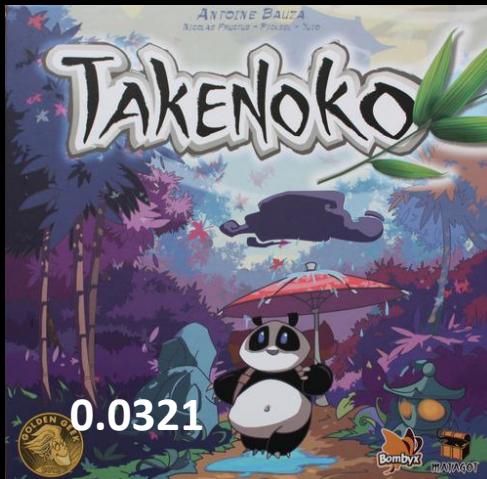
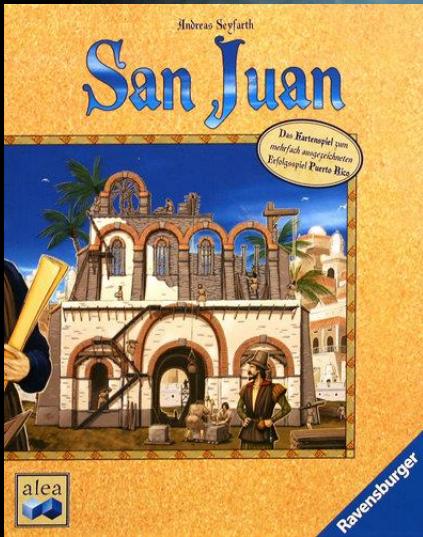
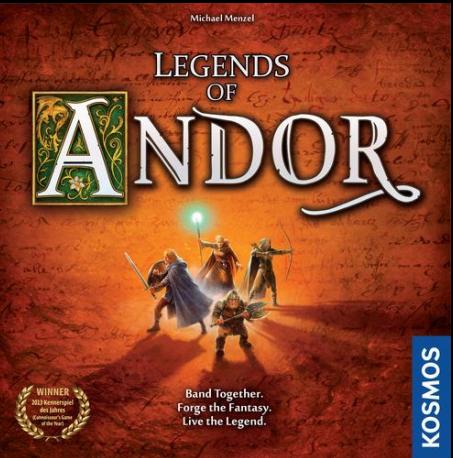
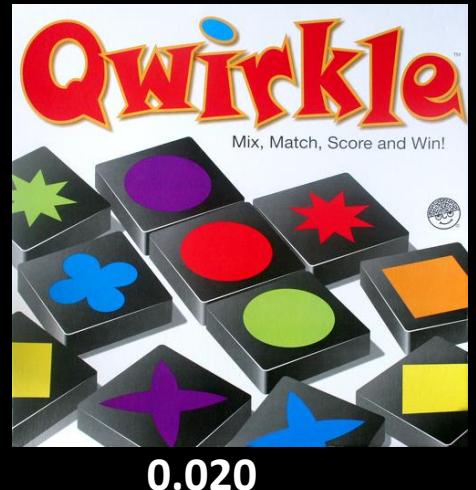
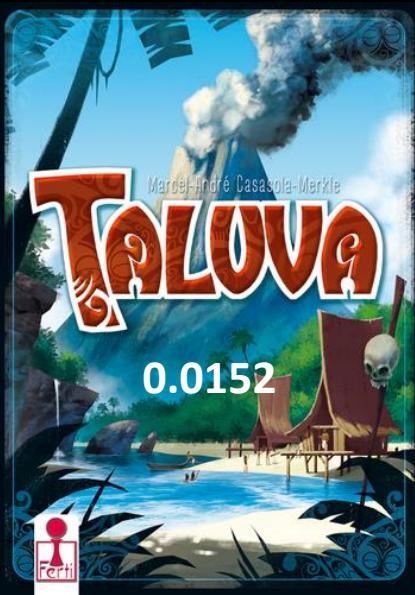
Example: 7 Wonders

- For each player count that the game supports (2-7)
- Recommended % = 1 – (Not Recommended %)
- $SD(0.39, 0.9, 0.98, 0.97, 0.9, 0.82) = 0.2216$

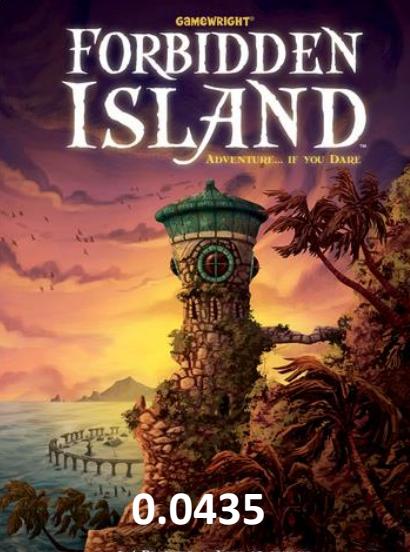
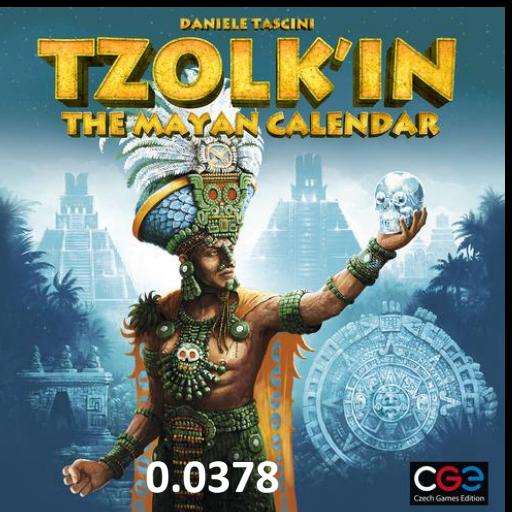
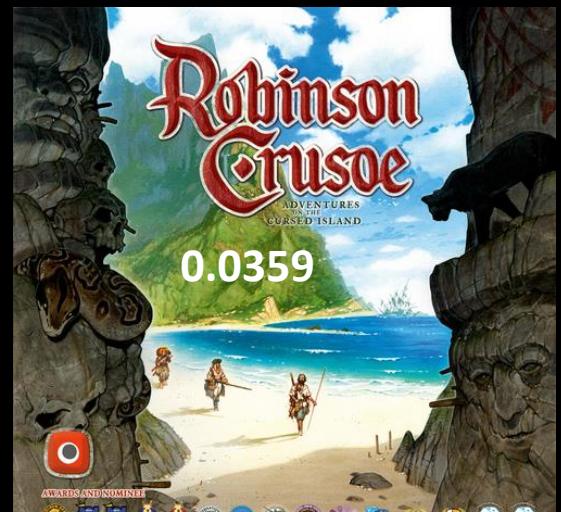
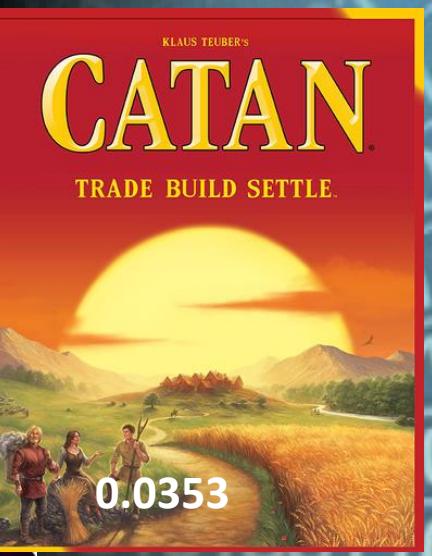
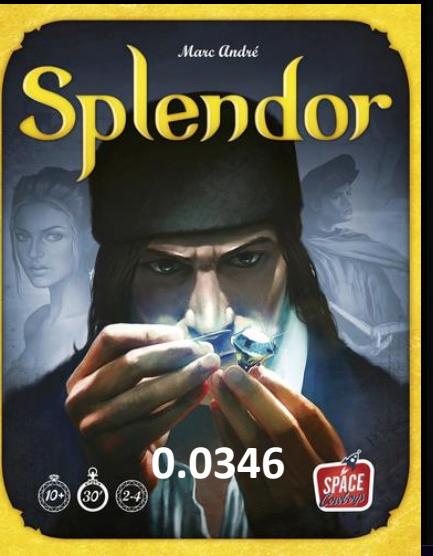
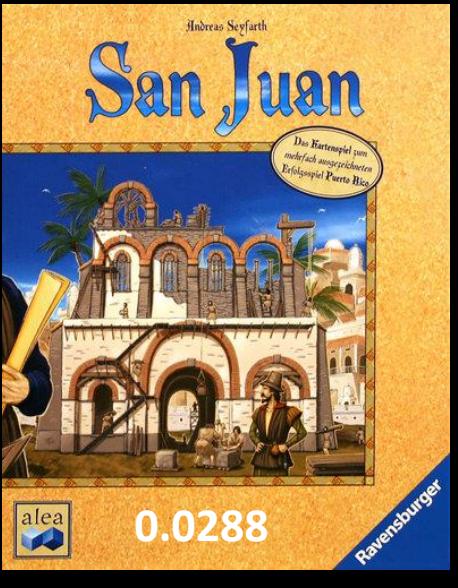
How many players do you recommend for this game? (OK to select more than one as "Best")

	Best	Recommended	Not Recommended	Vote Count
1	0.3% (2)	1.3% (10)	98.5% (779)	791
2	10.4% (105)	28.8% (292)	60.8% (617)	1014
3	26.9% (314)	63.0% (737)	10.1% (118)	1169
4	59.1% (739)	38.8% (485)	2.2% (27)	1251
5	48.4% (589)	48.3% (588)	3.3% (40)	1217
6	23.7% (267)	66.3% (747)	10.0% (113)	1127
7	22.0% (245)	60.7% (676)	17.3% (192)	1113
more than 7	2.7% (16)	9.8% (59)	87.5% (526)	601
Total voters				1436

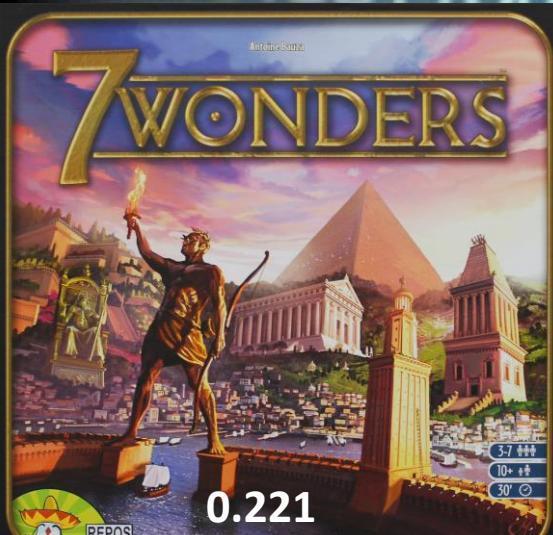
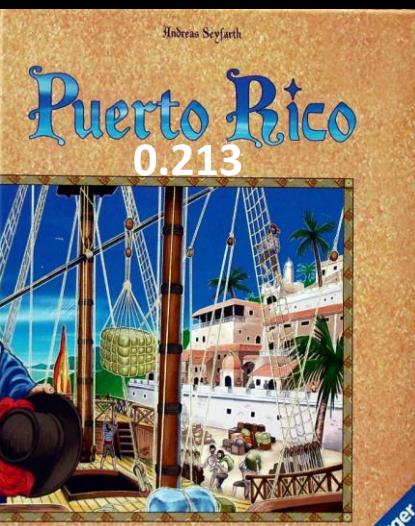
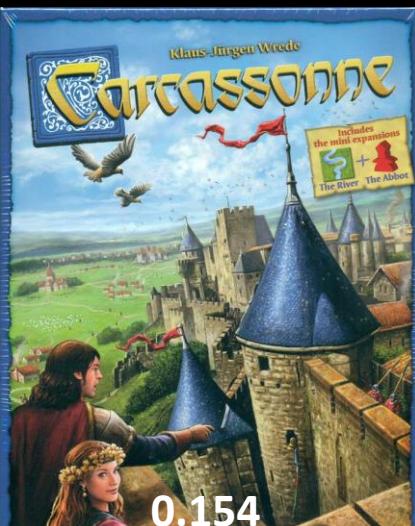
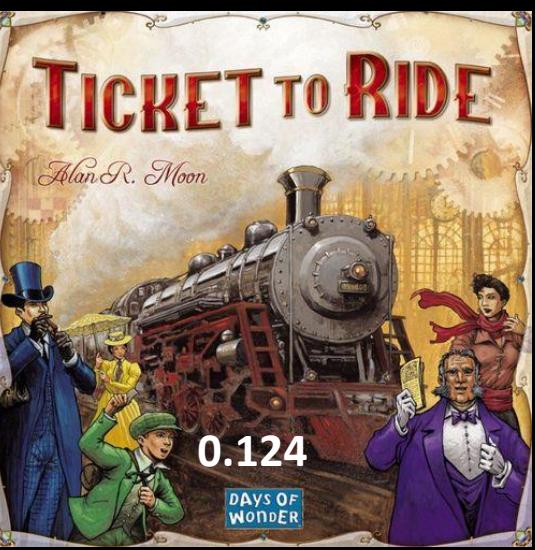
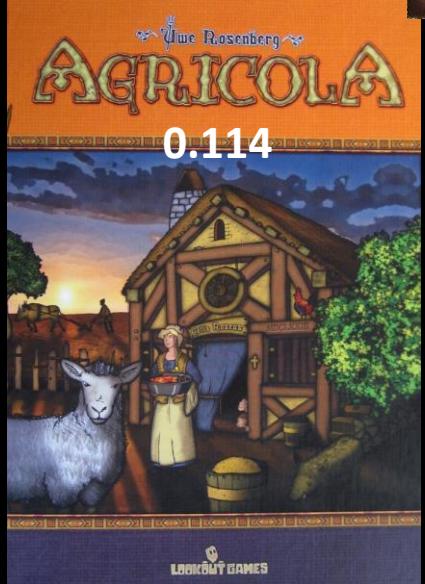
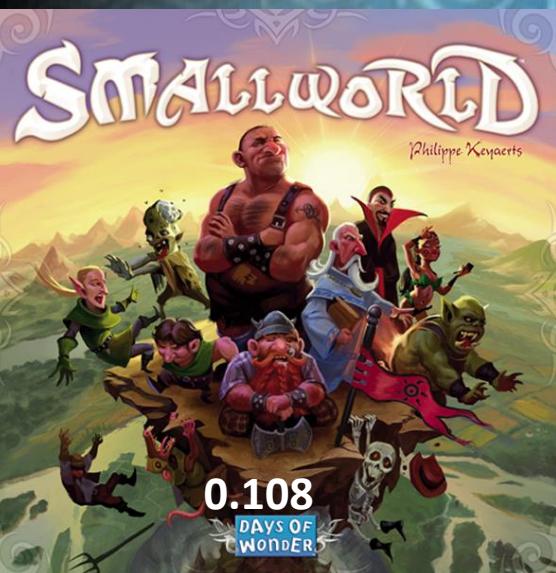
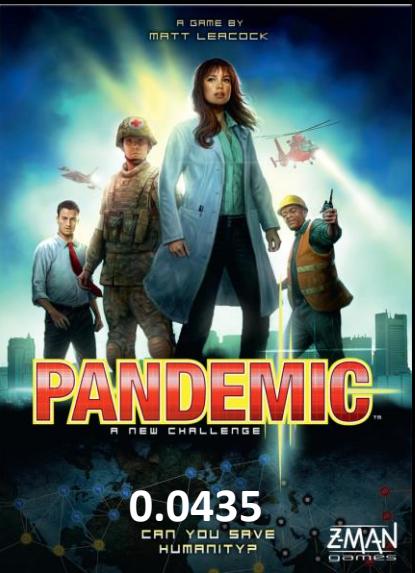
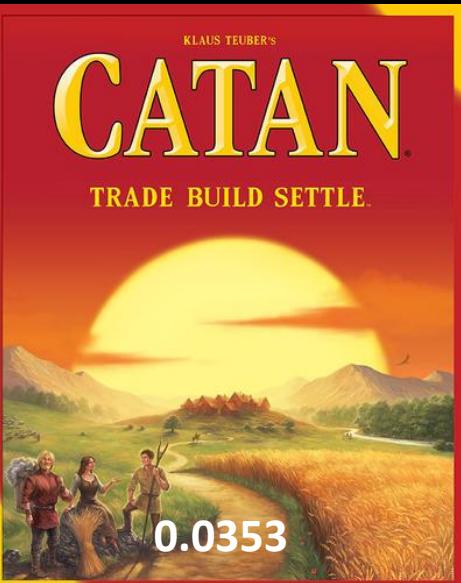
Most Balanced Games (> 500 votes)



Most Balanced Games (> 1000 votes)



Most Balanced Games (> 3000 votes)

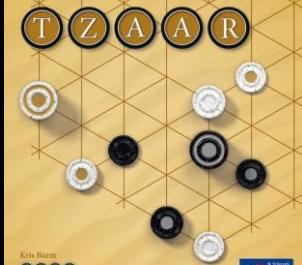
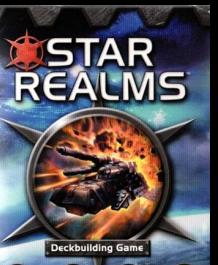
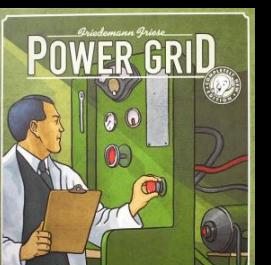
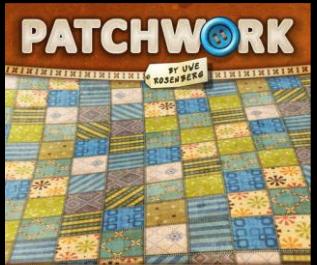
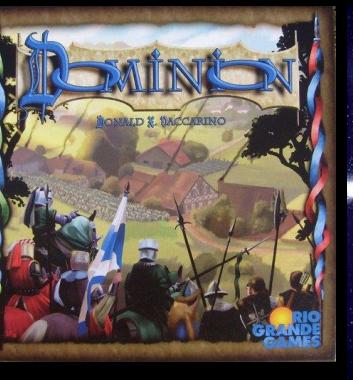
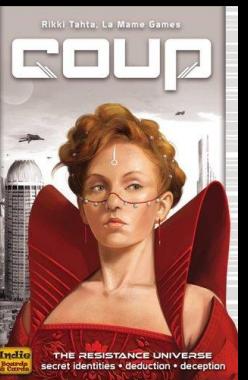
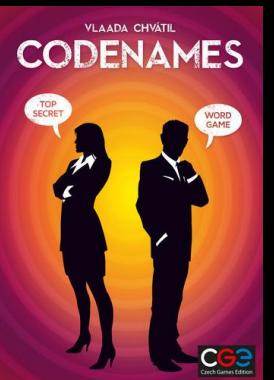
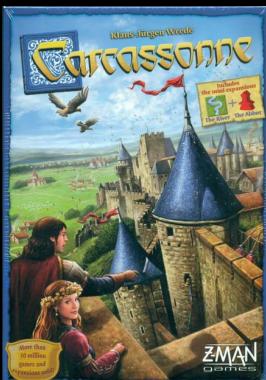


Final Thoughts

- Despite our childhood memories: there are no *bad* games
- Board games are booming
- Never been a better time to get into board games
- BoardGameGeek offers a lavish amount of data for investigation
 - Most games in the database are at a 5 or 6, so a 7 is actually pretty good!
 - BGG users like games that are about 2 hours long
 - BGG users like very complex games a lot
 - BGG users only really own games that are medium level complexity
 - Most balanced game – Race For the Galaxy

Find Some Fun!

- Not sure where to start? Find a game at PAX this weekend!
 - Tabletop Tournaments
 - Splendor at 4pm, Love Letter at 6pm – Olive 8
 - Games in the freeplay area – WSCC Level 2
 - My recommendations to try:



Resources for self-study

- source code, deck, raw data:

<https://github.com/scottburger/pax17>

- More in-depth code: svburger.com

<https://www.boardgamegeek.com/xmlapi2>