

Introduction

IMGD 2905

What is data analysis for game development?

What is data analysis for game development?

- Using **game data** to inform the **game development** process
- Where does this data come from?
 - Users playing game
 - **Quantitative** (instrumented)
 - **Qualitative** (subjective evaluation)
 - (But often lots more of the former!)

What can game analysis do for game development?

What can game analysis do for game development?

- Improve level design – e.g., see where players are getting stuck
- Focus development on critical content – e.g., see what game modes or characters are not used
- Balance gameplay – e.g., tune parameters for more competitive and fun combat
- Broaden appeal – e.g., hear if content/story is engaging or repulsing

Why is data analysis for game development needed?

Why is data analysis for game development needed?

- Challenge
 - Games gotten larger and more complex
 - Number of reachable states, characters
 - Need for metrics to make sense of player behavior has increased
- Opportunity
 - New technologies enable aggregation, access and analysis

IMGD 2905 – Doing Data Analysis for Game Development

- **Data analysis pipeline** – get data from games, through analysis, to stakeholders
- **Summary statistics** – central tendencies of data
- **Visualization of data** – how to display analysis, illustrate messages
- **Statistical tests** – quantitatively determine relationships (e.g., correlation)
 - Probability needed as foundation
- **Regression** – model relationships
- More advanced topics (e.g., ML, Data management ...)

For this class:
Described in lecture
Read about in book
Applied in projects

Foundations for Data Analysis for Game Development at WPI

- Statistics classes
 - MA 2610 Applied Statistics for Life Sciences
 - MA 2611 Applied Statistics I
 - MA 2612 Applied Statistics II
- Probability classes
 - MA 2621 Probability for Applications
- Data Science minor
 - MA, CS, BUS
 - DS 3001 Foundations of Data Science
- Data Mining
 - CS 4445 Data Mining and Knowledge Discovery in Databases
- Other
 - CS 1004 Introduction to Programming for Non-Majors
 - CS 3431 Database Systems I

Note – other Stats and Probability classes are primarily geared for Math majors

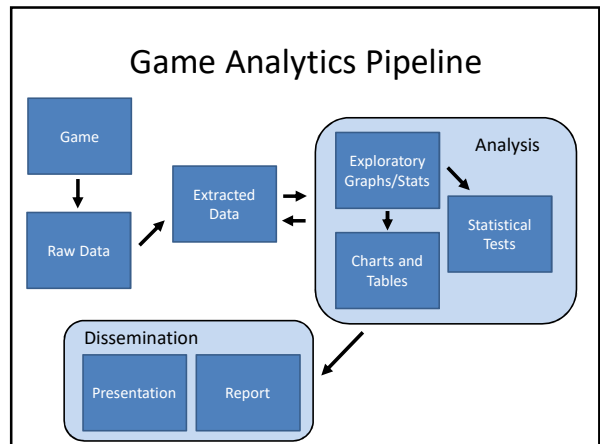
Outline

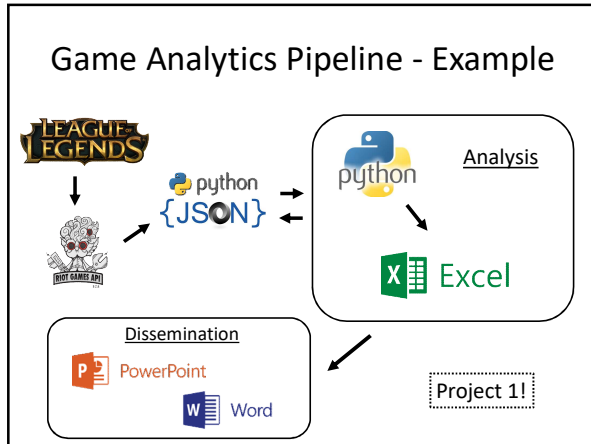
- Overview (done)
- Game Analytics Pipeline (next)
- Game Data Analysis Examples

Sources of Game Data

<p>Quantitative (Objective)</p> <ul style="list-style-type: none"> • Internal Testing <ul style="list-style-type: none"> – Developers – QA • External Testing <ul style="list-style-type: none"> – Usability testing – Beta tests – Long-term play data 	<p>Qualitative (Subjective)</p> <ul style="list-style-type: none"> • Surveys • Reviews • Online communities • Post mortems
---	---

How to get from data to dissemination?
→ Game analytics pipeline







- ### Game Analytics Tools
- **Games** – breadth of experience with games, specific experience with game to be analyzed
 - **Statistics** – measures of central tendency, measures of spread, statistical tests
 - **Probability** – rules, distributions
 - **Data Visualization** – bar chart, scatter plot, histogram, error bars
 - **Technical Writing and Presentation** – white paper, technical talk; audience is peer group, developers, boss

- ### Outline
- Overview (done)
 - Game Analytics Pipeline (done)
 - Game Data Analysis Examples (next)


Example: Project Gotham Racing 4



K. Hullett, N. Nagappan, E. Schuh, and J. Hopson. "Data Analytics for Game Development", International Conference on Software Engineering (ICSE), May, 2011, Waikiki, Honolulu, HI, USA
<http://dl.acm.org/citation.cfm?id=1985952>



- **Publisher** – Microsoft 2007
 - 134 vehicles, 9 locations, 10 game modes
- **Analyzed data**
 - (Authors worked at Microsoft)
 - 3.1 million log entries, 1000s of users



Project Gotham Racing 4: Results

Game Mode	Races	% Total
OFFLINE_CAREER	1479586	47.63%
PGR_ARCADE	566705	18.24%
NETWORK_PLAY	584201	18.81%
SINGLE_PLAYER_PLAY	185415	5.97%
...		
NET_TOURNY_ELIM	2713	0.09%
...		
Group	Races	% Total
STREET_RACE	795334	25.60%
NET_STREET_RACE	543491	17.50%
ELIMINATION	216042	6.95%
HOTLAP	195949	6.31%
...		
TESTTRACK_TIME	7484	0.24%
CAT_N_MOUSE_FREE	3989	0.13%
CAT_N_MOUSE	53	0.00%

- Thoughts?
- What are some main messages?

Project Gotham Racing 4: Results


Game Mode	Races	% Total
OFFLINE_CAREER	1479586	47.63%
PGR_ARCADE	566705	18.24%
NETWORK_PLAY	584201	18.81%
SINGLE_PLAYER_PLAY	185415	5.97%
...		
NET_TOURNY_ELIM	2713	0.09%
...		
Group	Races	% Total
STREET_RACE	795334	25.60%
NET_STREET_RACE	543491	17.50%
ELIMINATION	216042	6.95%
HOTLAP	195949	6.31%
...		
TESTTRACK_TIME	7484	0.24%
CAT_N_MOUSE_FREE	3989	0.13%
CAT_N_MOUSE	53	0.00%

- **Mode**
 - Offline career dominates
 - Network tournament hardly used
- **Events**
 - Street race and network street race dominate
 - Cat and mouse never used
- **Vehicles** (not shown)
 - 1/3 used in less than 0.1% of races


Project Gotham Racing 4: Conclusion

- Content underused - 30-40% of content in less than 1% of races
- Use to shift emphases for DLC, next version
 - Asset creation costs significant, so even 25% reduction noticeable
- Other (not shown)
 - Encouraging new players to play career mode
 - Increasing likelihood of continuing play
 - Encouraging new players to stay with F Class longer
 - Rather than move to more difficult to control A Class


Example: Halo 3



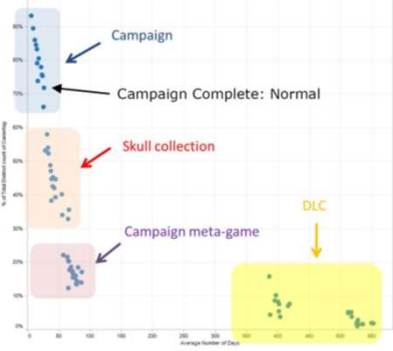
B. Phillips. "Peering into the Black Box of Player Behavior: The Player Experience Panel at Microsoft Game Studios". *Game Developers Conference (GDC)*, 2010. <http://www.gdcvault.com/play/1012387/Peering-into-the-Black-Box>



- Publisher – Microsoft 2007
 - Achievements: single player missions, challenges such as finding skulls, multiplayer accomplishments...
- Analyzed data
 - (Author worked at Microsoft)
 - 18,000 players

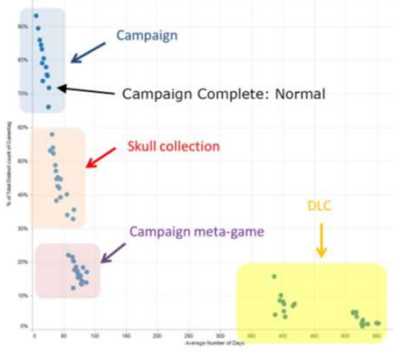


Halo 3: Results



- Thoughts?
- What are some main messages?

Halo 3: Results




- 73% of players completed campaign
 - Can compare to other Xbox games
- Took 26 days to accomplish
- Double that time for all original content
- DLC provides users up to 2 years of content

[Good Descriptive Example](#)

League of Legends Example: League of Legends

Mark Claypool, Jonathan Decelle, Gabriel Hall, and Lindsay O'Donnell. "Surrender at 20? Matchmaking in League of Legends," in *Proceedings of the IEEE Games, Entertainment, Media Conference (GEM)*, Toronto, Canada, October 2015. Online at: <http://www.cs.wpi.edu/~claypool/papers/lof-matchmaking/>

- Publisher – Riot Games 2009
 - Rank: ~5 Tiers, 5 divisions each → 25
- User study (52 players)
 - Play LoL in controlled environment
 - Record objective data
 - (e.g., player rank and game stats)
 - Provide survey for subjective data
 - (e.g., match balance and enjoyment)



Fun ↑

Game Balance

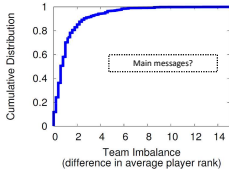
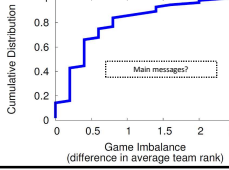
Too hard! ← Just right! → Too easy!

Sweet spot

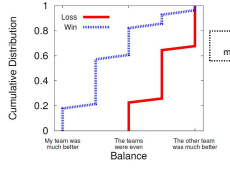
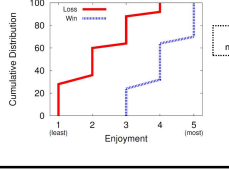
???

League of Legends: Results

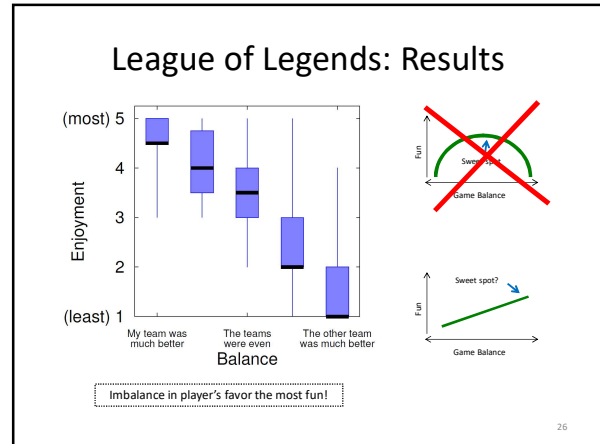
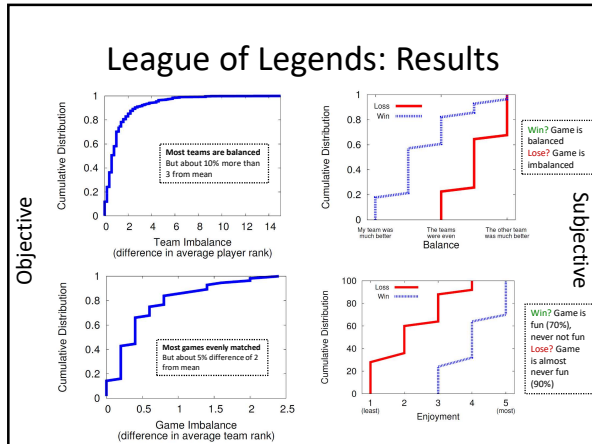
Objective

Subjective

Main messages?



League of Legends: Conclusion

Objective

- Teams are balanced
 - 50% players within 1 rank of each other
- Games are balanced
 - 80% teams within 1 average rank of each other

Subjective

- Games are *not* balanced
- When players win, perceive slight imbalance
- When players lose, perceive large imbalance
- Players enjoy winning more than losing (no surprise)
- (Surprise!)** Players most enjoy matches imbalanced in their favor!

Matchmaking systems may want to consider - e.g., balance not so important, as long as player not always on imbalanced side

27

Summary

- Data analysis for games increasingly important
 - Has potential to improve game development
- Knowledge and skills required
 - Scripting
 - Statistics
 - Data analysis
 - Writing and presentation