Introduction

IMGD 2905

Groupwork

- What is data analysis for game development?
- Where does this data come from?
- What can game analysis do for game development?

- Icebreaker, Groupwork, Questions

https://web.cs.wpi.edu/~imgd2905/d23/groupwork/1-introduction/handout.html

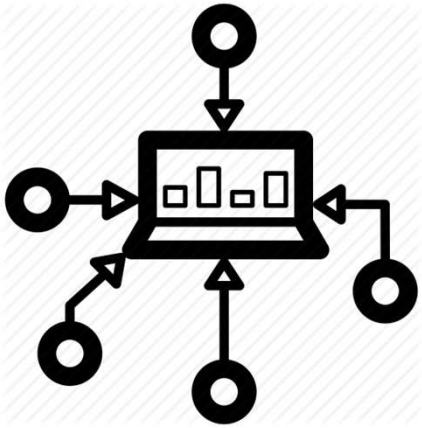
What is data analysis for game development?

What is data analysis for game development?

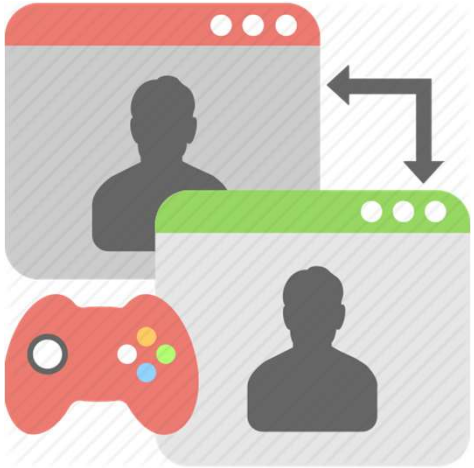
- Using game data to inform the game development process
- Where does this data come from?

→ *Players*, actually playing game

  - Quantitative (instrumented)
  - Qualitative (subjective evaluation)
  - (But often lots more of the former!)



https://cdn0.iconfinder.com/data/icons/big-data-2-1/128/Data-Integration-process-database-collection-512.png



https://cdn2.iconfinder.com/data/icons/sports-and-games-5-1/48/216-512.png

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
- Note: game data often informs *players*, too
  - Analytics not dissimilar

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

    → Game balance harder to achieve
  - Need for metrics to make sense of player behavior has increased
- Opportunity
  - New technologies enable collecting data, 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 (also used for game rules)
- Regression – model relationships
- Hint at more advanced topics
  - e.g., ML, Data management …

For this class:

Described in lecture
Read about in book
Applied in projects and homework

Foundations for Data Analysis @ 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 and major)
  - DS 1010 Introduction to Data Science
  - DS 2010 Modeling and Data Analysis
  - DS 3010 Computational Data Intelligence
  - DS 4433/CS4433 Big Data Management and Analytics
- 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

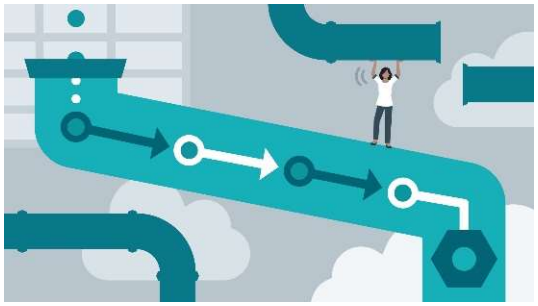
Quantitative (Objective)

- Internal Testing
  - Developers
  - QA
- External Testing
  - Usability testing
  - Beta tests
  - Long-term play data

Qualitative (Subjective)

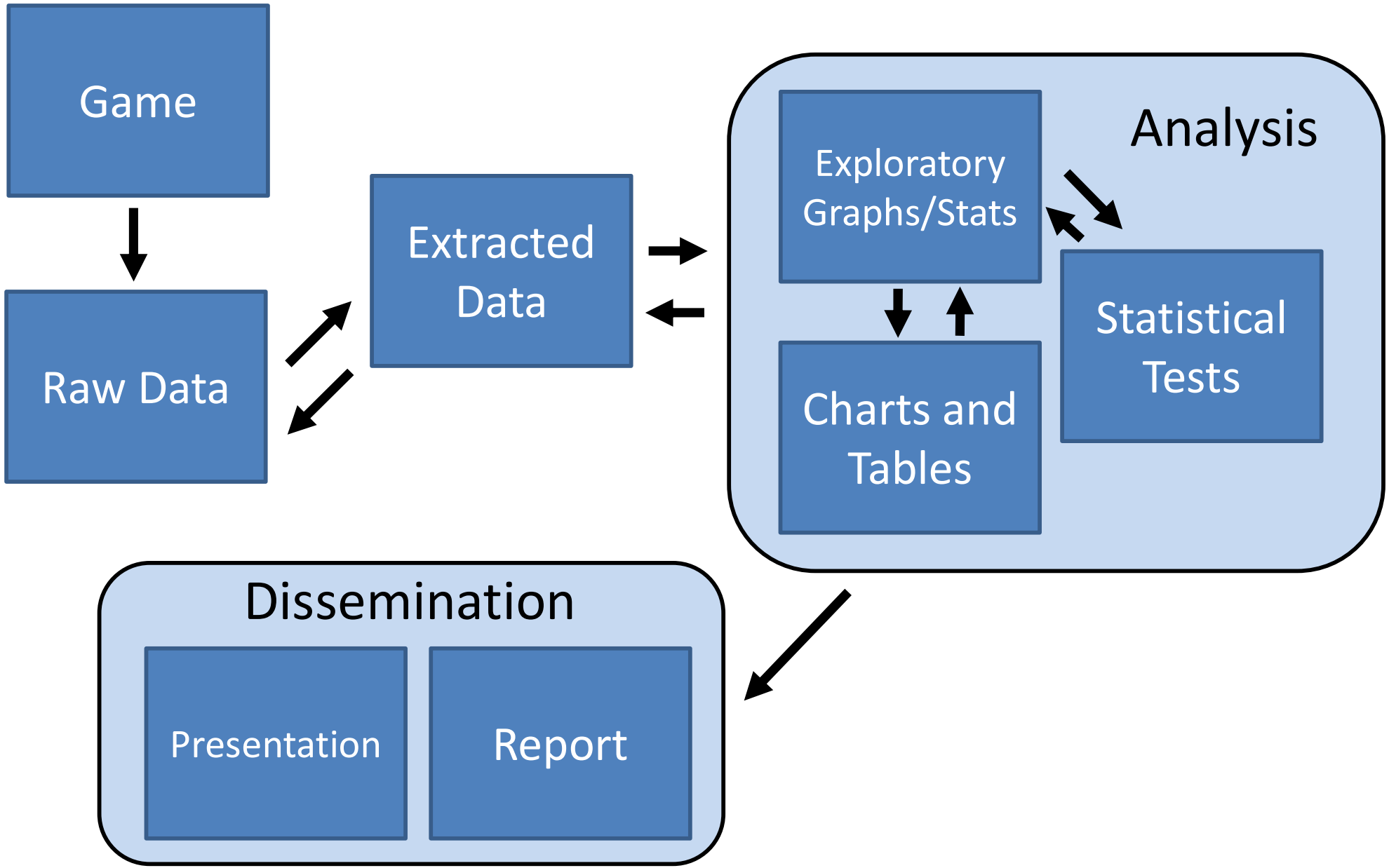
- Surveys
- Reviews
- Online communities
- Postmortems

https://tinyurl.com/y3gaja4j

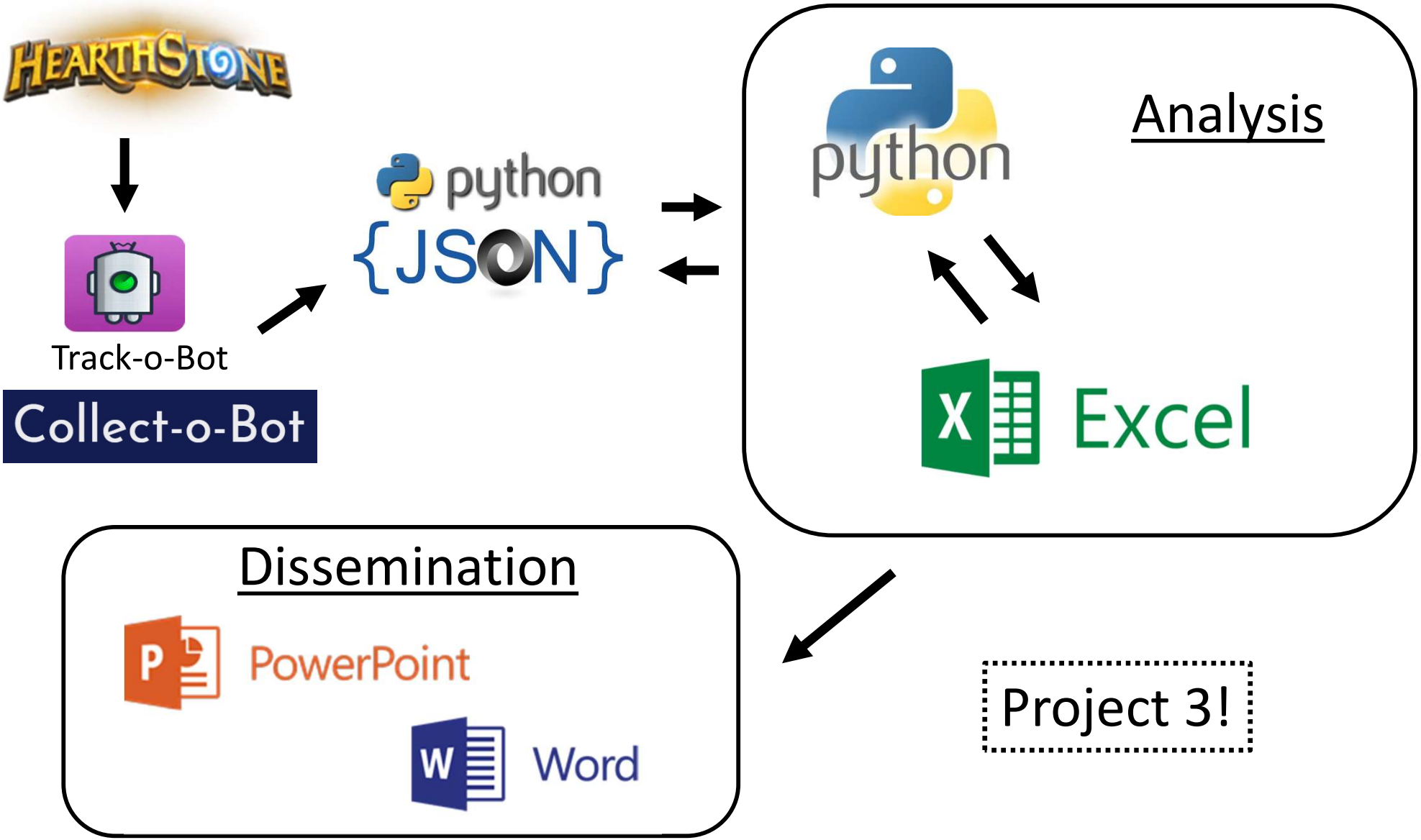

https://cdn.lynda.com/course/699344/699344-636703722500462287-16x9.jpg

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

Game Analytics Pipeline



Game Analytics Pipeline – Example



Game Analytics Components

- Games – breadth of experience with games, specific experience with game to be analyzed

- Tools – import, clean, filter, format data so can analyze
- Statistics – measures of central tendency, measures of spread, statistical tests
- Probability – rules, distributions
- Data Visualization – bar chart, scatter plot, histogram, error bars

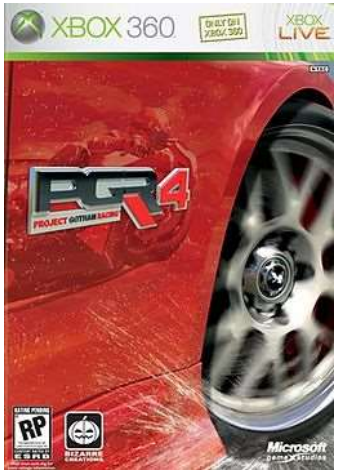
- Technical Writing and Presentation – white paper, technical talk; audience is peer group, developers, boss

Outline

- Overview (done)
- Game Analytics Pipeline (done)
- Game Data Analysis Examples (next)
  - Table
  - Scatter plot
  - Boxplot

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

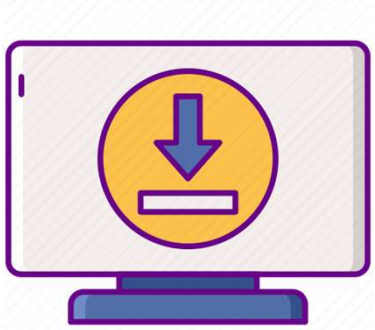
Game Mode	Races	% Total
OFFLINE_CAREER	1479586	47.63%
PGR_ARCADE	566705	18.24%
NETWORK_PLAY	584201	18.81%
SINGLE_PLAYER_PLAY	185415	5.97%
….		
NET_TOURNY_ELIM	2713	0.09%

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* play not well-used
- Events
  - *Street race* and *network street race* dominate
  - *Cat and mouse* almost never used
- Vehicles (not shown)
  - 1/3 used in less than 0.1% of races

Project Gotham Racing 4: Conclusion

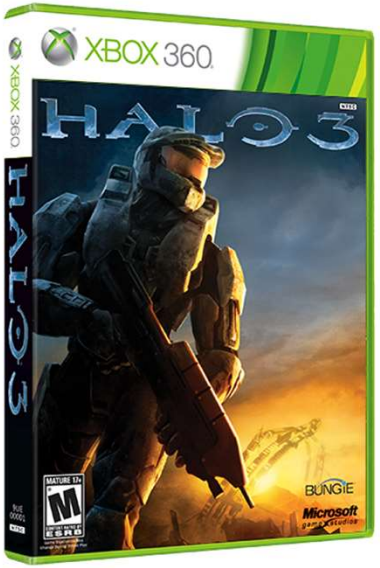
- Content underused - 30-40% of content in less than 1% of races
- Use to shift emphasis to DLC, next version
  - Asset creation costs significant, so even 25% reduction noticeable



- 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*

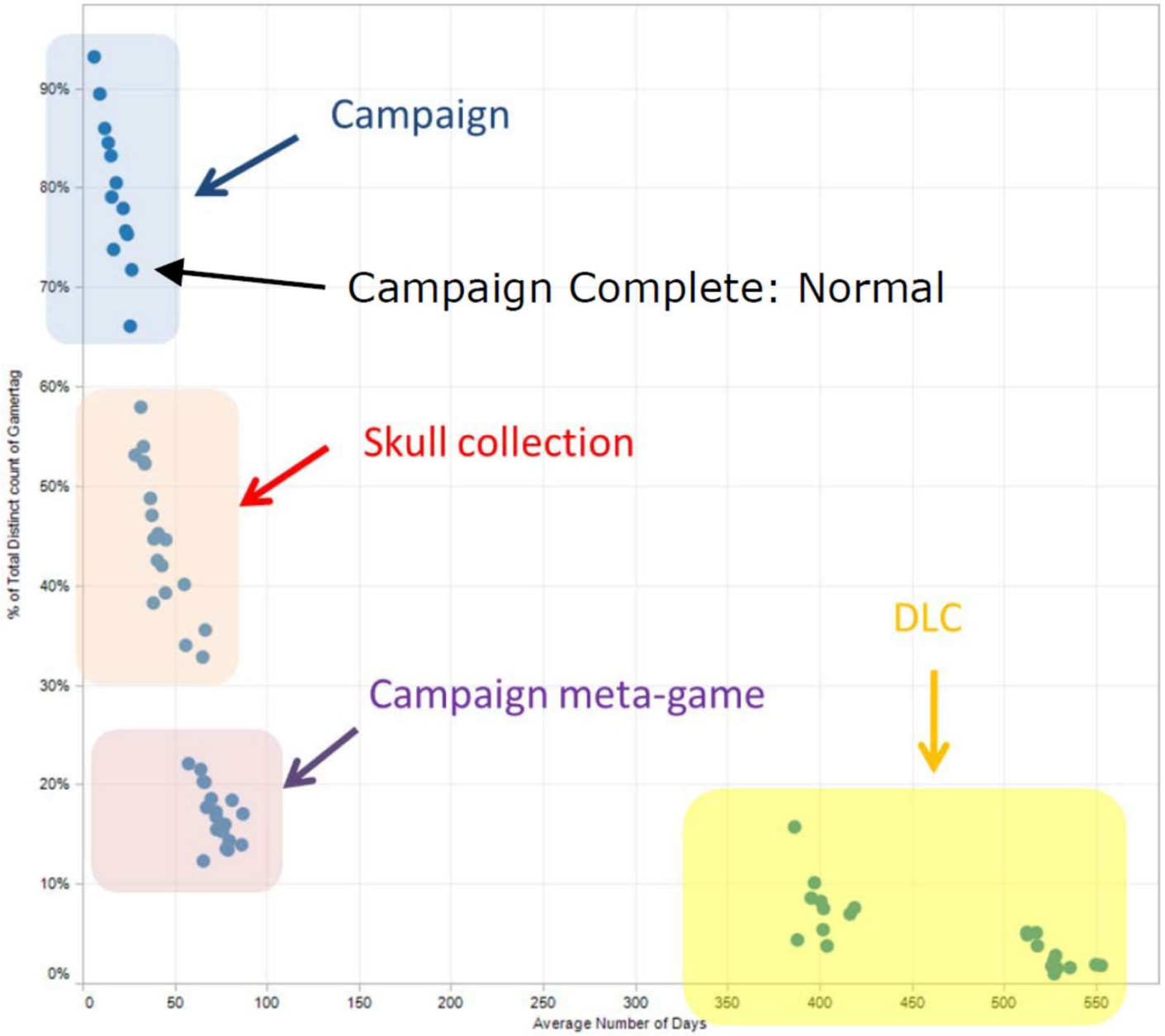
Example:
Halo 3

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



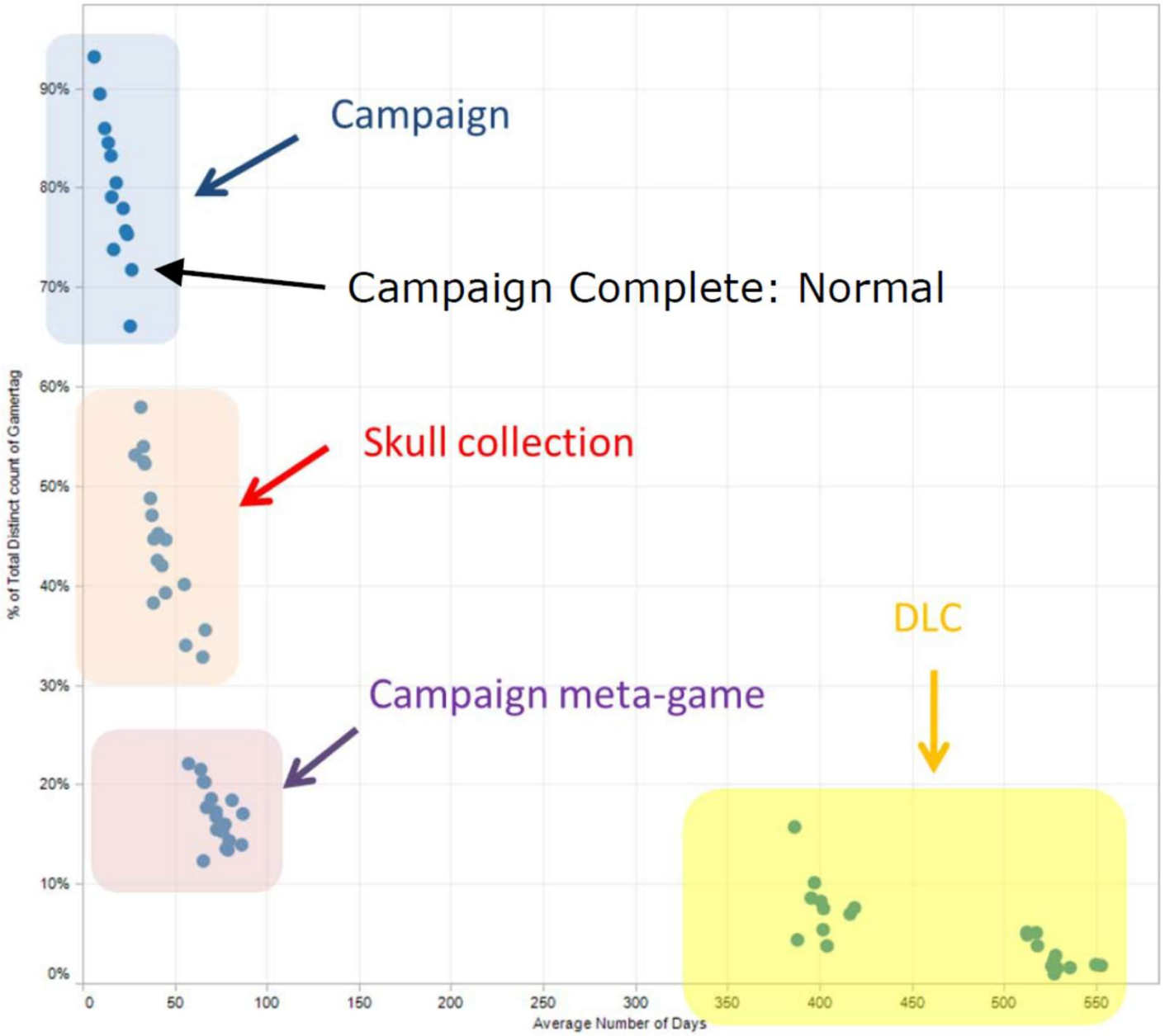
- Publisher – Microsoft 2007
  - Achievements: single player missions, challenges such as finding skulls, multiplayer accomplishments…
- Analyzed data
  - (Author worked at Microsoft)
  - 18,0000 players

Halo 3: Results



- Thoughts?
- What are some main messages?

Halo 3: Results



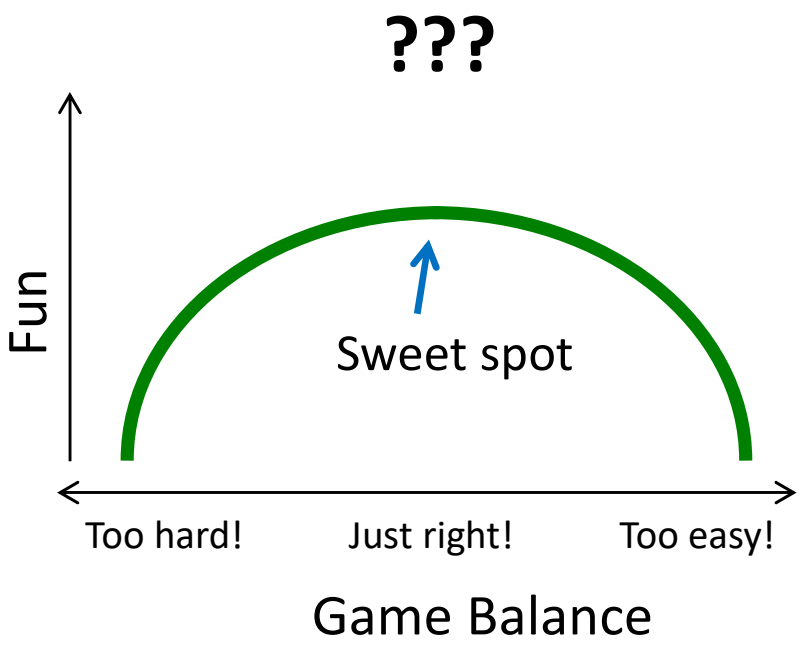
- 73% of players completed campaign
  - Can compare to other Xbox games
- Took 26 days to accomplish
- Double that time for all original content
- DLC provides users up to 2 years of content

Good Descriptive Example

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/lol-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)



League of Legends: Results

Objective

Team Imbalance

(difference in average player rank)

Min: 0

Median: 0.75

Max: 15

Main Messages?

Game Imbalance

(difference in average team rank)

Min: 0

Median: 0.5

Max: 2.5

Main Messages?

League of Legends: Results

Objective

Team Imbalance

(difference in average player rank)

Min: 0

Median: 0.75

Max: 15

Most teams are balanced
Over half less thant 1

Game Imbalance

(difference in average team rank)

Min: 0

Median: 0.5

Max: 2.5

Most games evenly matched
Over half less than 0.5

League of Legends: Results

Objective

Subjective

Team Imbalance

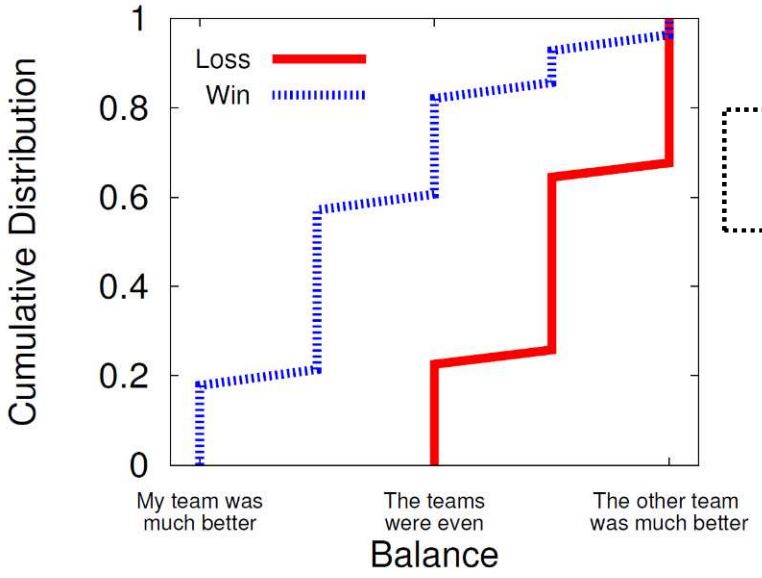
(difference in average player rank)

Min: 0

Median: 0.75

Max: 15

Most teams are balanced
Over half less thant 1



Main messages?

Game Imbalance

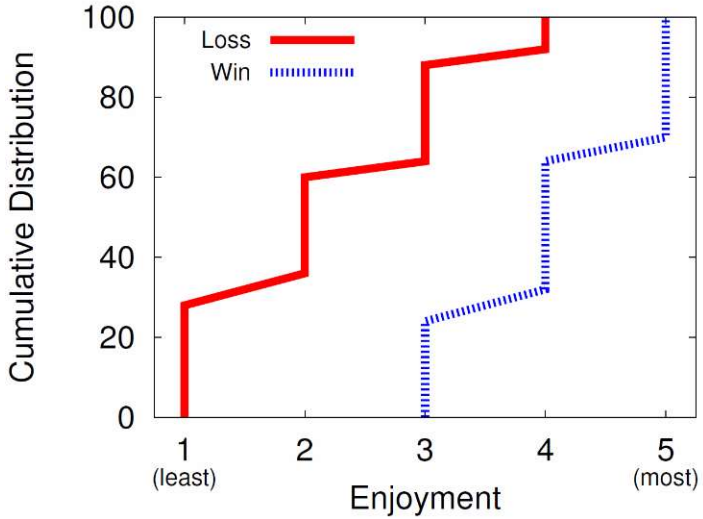
(difference in average team rank)

Min: 0

Median: 0.5

Max: 2.5

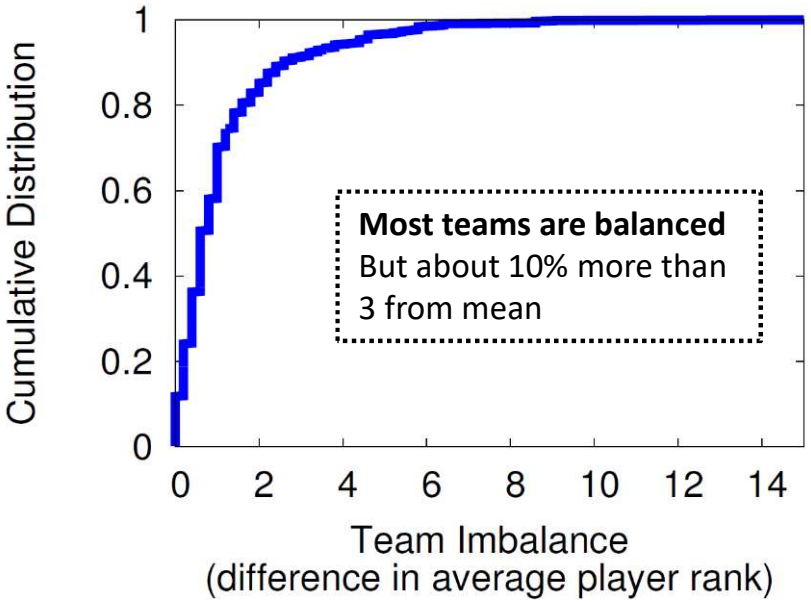
Most games evenly matched
Over half less than 0.5

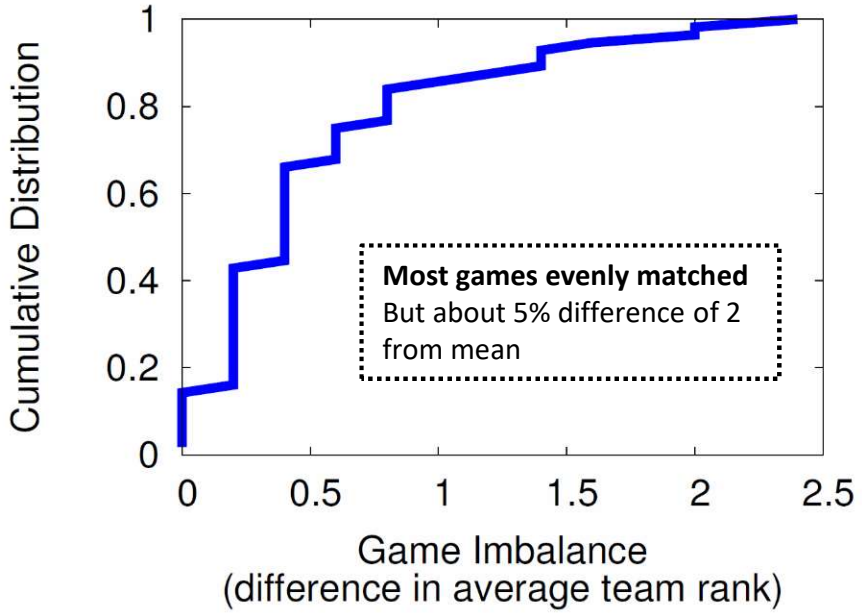


Main messages?

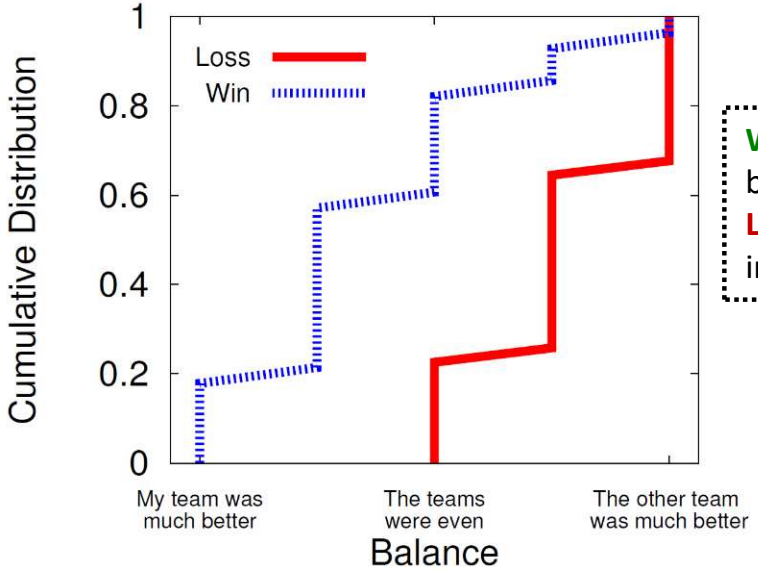
League of Legends: Results

Objective

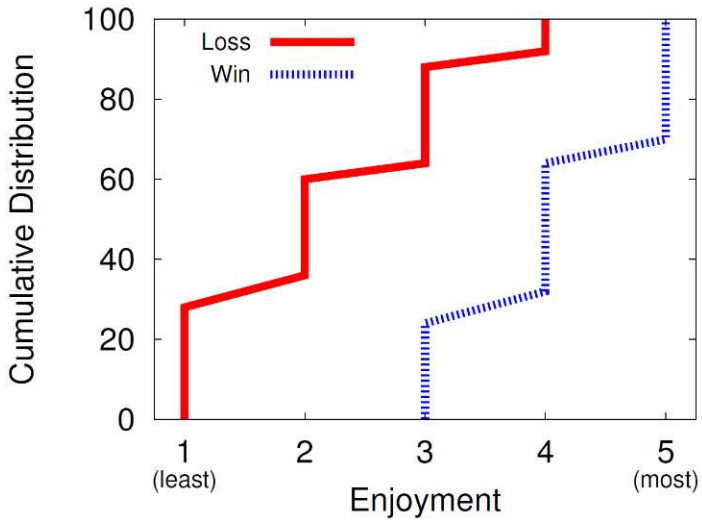




Subjective

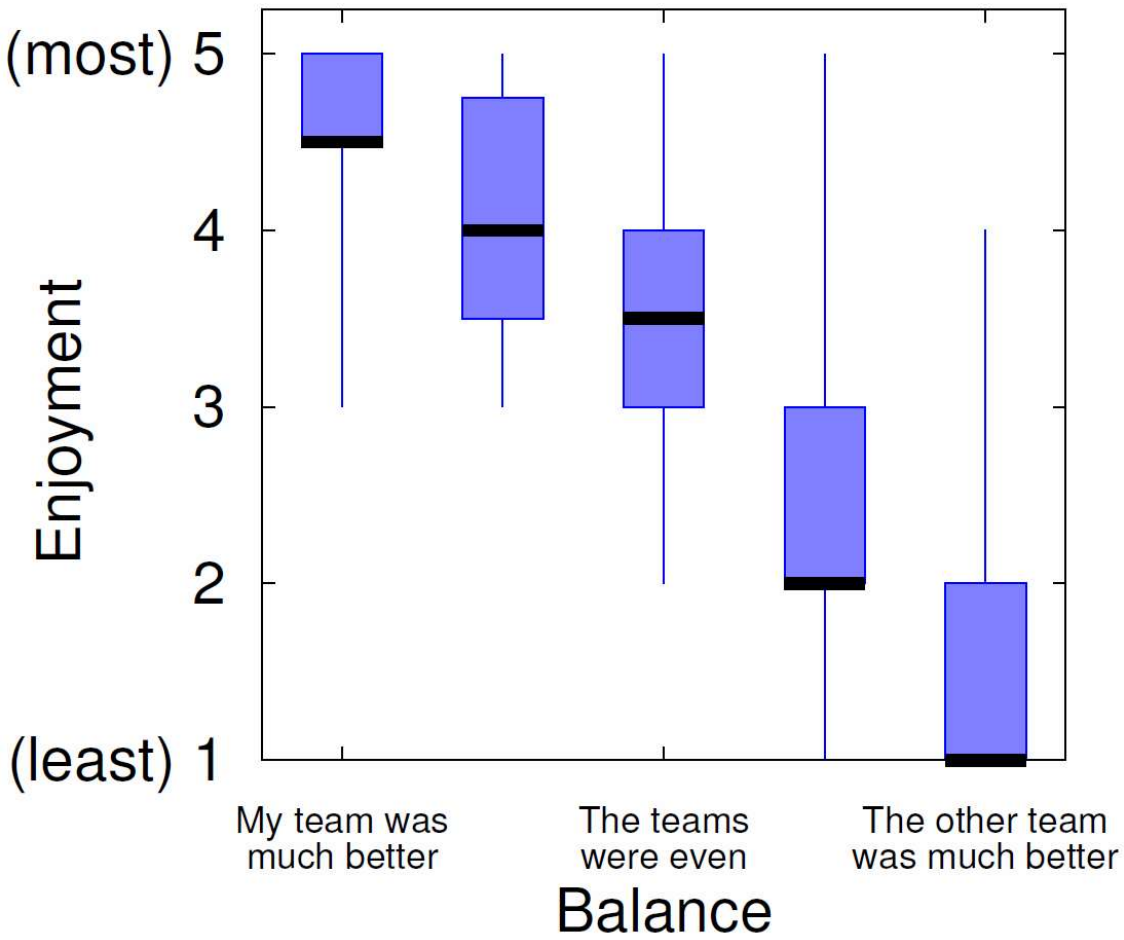


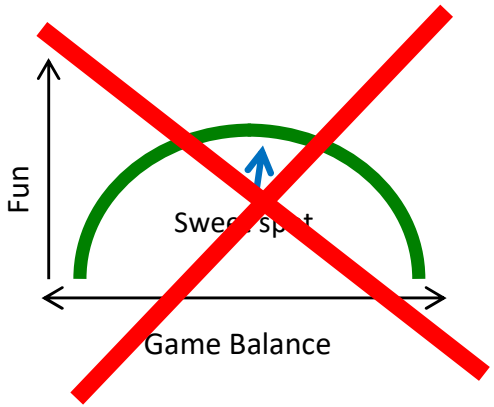
Win? Game is balanced
Lose? Game is imbalanced

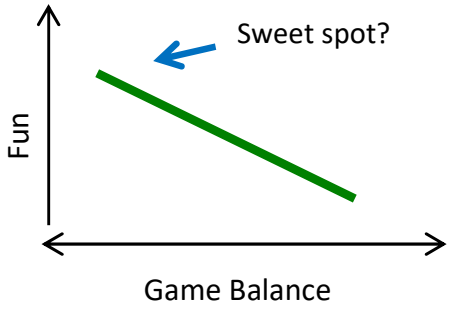


Win? Game is fun (70%), never not fun
Lose? Game is almost never fun (90%)

League of Legends: Results







Imbalance in player's favor the *most* fun!

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

Summary

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

https://1kabswnt2ua3ivl0cuqv2f17-wpengine.netdna-ssl.com/wp-content/uploads/2014/06/Skills.jpg

"Let's get to it, already!"
-- Tracer (Overwatch)