Developing Data Science Skills Using Call of Duty® Data

Matt Slifko
Department of Statistics
The Pennsylvania State University

PennState
Eberly College
of Science

Motivation and Goals

- Interest in data, data science, and statistics is high

- GAISE Report[1] provides many useful recommendations for educators, for example:
  - Teach statistics as an investigative process of problem-solving and decision-making
  - Give students experience with multivariable thinking

- For teachers, there is always demand for interesting new data sets and examples

- Goals of this work:
  - Share resources
  - Share experiences

1 GAISE College Report: https://www.amstat.org/docs/default-source/amstat-documents/gaisecollege_full.pdf

PennState
Eberly College
of Science

Background

- Call of Duty® is popular video game franchise that began in 2003 and has sold over 400M units worldwide as of 2021
  - Guinness Book of World Records identified the franchise as the the best-selling first-person shooter series of all time[2] in 2021
  - Other sources identify the franchise among the most popular of all time along with Mario, Pokémon, and Tetris. [3-5]

- Our data focuses on a player's performance in an online multiplayer match

2 Guinness Book of World Records: https://www.guinnessworldrecords.com/world-records/87955-best-selling-first-person-shooter-fps-videogame-series
3 https://history-computer.com/discover-the-most-profitable-video-game-franchises-ever-made/
4 https://apptrigger.com/2023/01/16/highest-grossing-video-game-franchises/
5 https://www.titlemax.com/discovery-center/lifestyle/the-top-50-highest-grossing-video-game-franchises/

PennState
Eberly College
of Science

Data Sets Shared

- Gameplay Data: collect data from pregame lobby (map vote) and match results. 12 of the 25 variables include:

	Map1	Map2	Choice	MapVote	Result	Eliminations	Deaths	Score	Damage	TotalXP	XPType	GameType
1	Moscow	Miami Strike	Miami Strike	5 to 0	100–97	22	17	4070	634	11002	10% Boost	HC – TDM
2	Moscow	WMD	Moscow	2 to 0	76–89	20	15	5305	560	9451	10% Boost	HC – TDM
3	NA	NA	Yamantau	NA	100–92	18	11	3335	483	12948	10% Boost	HC – TDM
4	Drive–In	Jungle	Drive–In	2 to 0	80–100	10	19	2170	280	11502	Double XP + 10%	HC – TDM
5	Collateral Strike	Hijacked	Collateral Strike	3 to 3	71–100	11	19	2195	308	11133	Double XP + 10%	HC – TDM

- Along with a data dictionary, we will provide the following data sets:
  - 2 gameplay datasets (one semi-clean and one messy)
  - Maps dataset
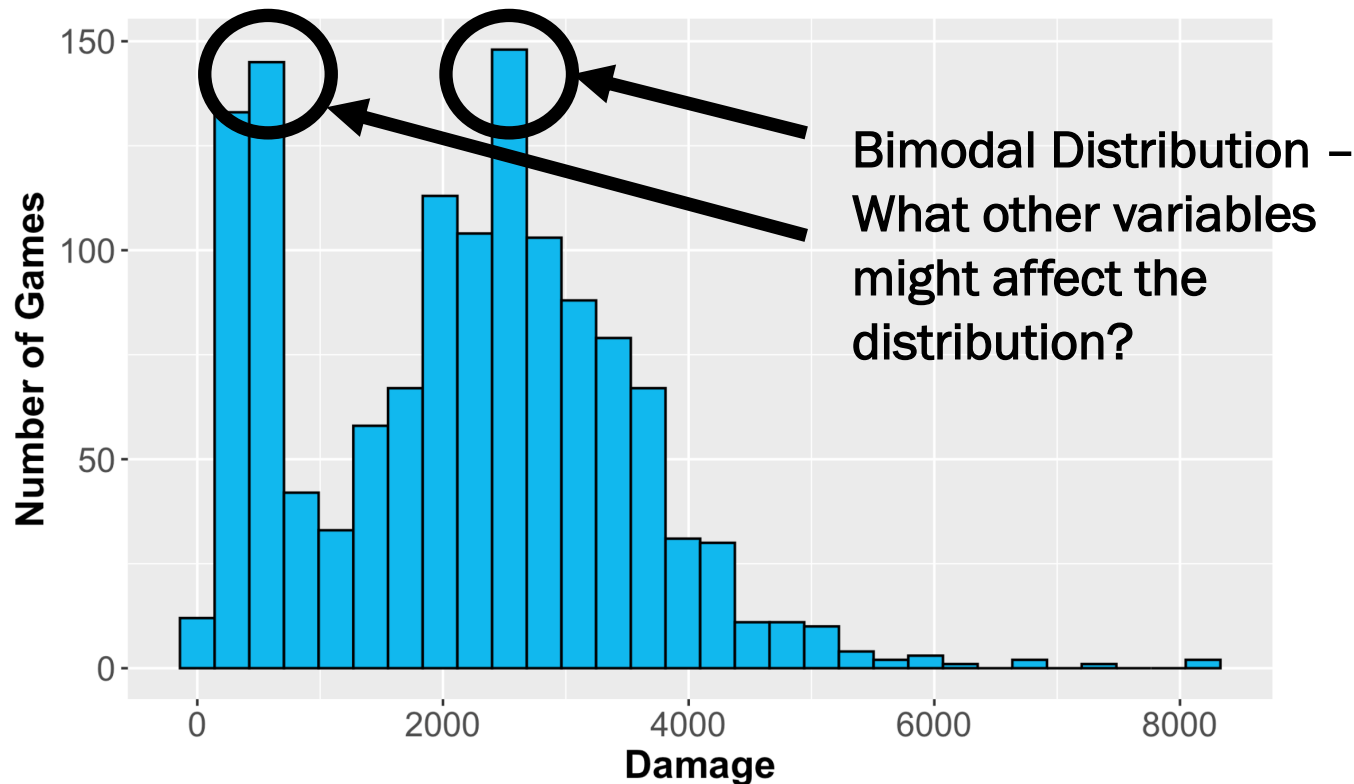 - Weapons dataset
  - Game modes dataset

- Please visit: https://github.com/matthewdslifko/CallOfDutyProject

PennState
Eberly College
of Science

Example – Exploratory Data Analysis 1

- Background: "Damage" represents the amount of damage issued by the player on the opposing team's players, weapons, vehicles

- Task: Explore the distribution of Damage.
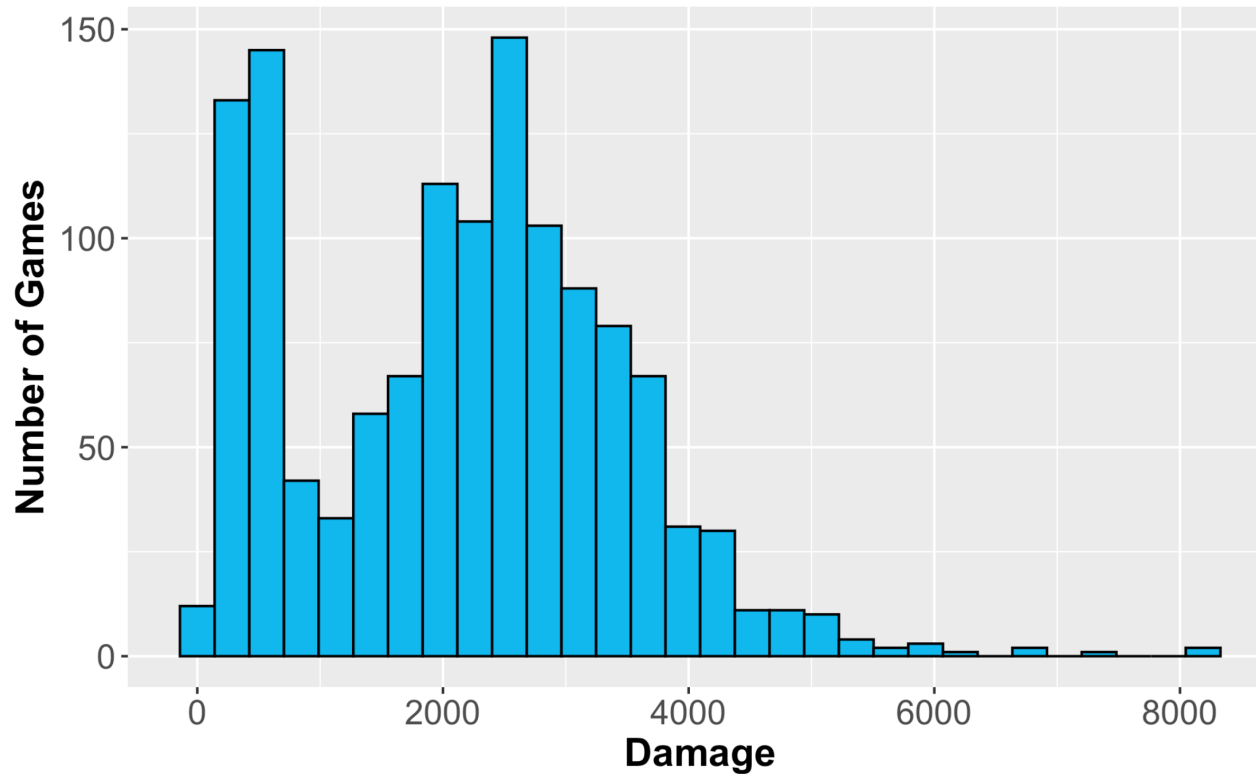
Example – Exploratory Data Analysis 2

- Background: "Damage" represents the amount of damage issued by the player on the opposing team's players, weapons, vehicles

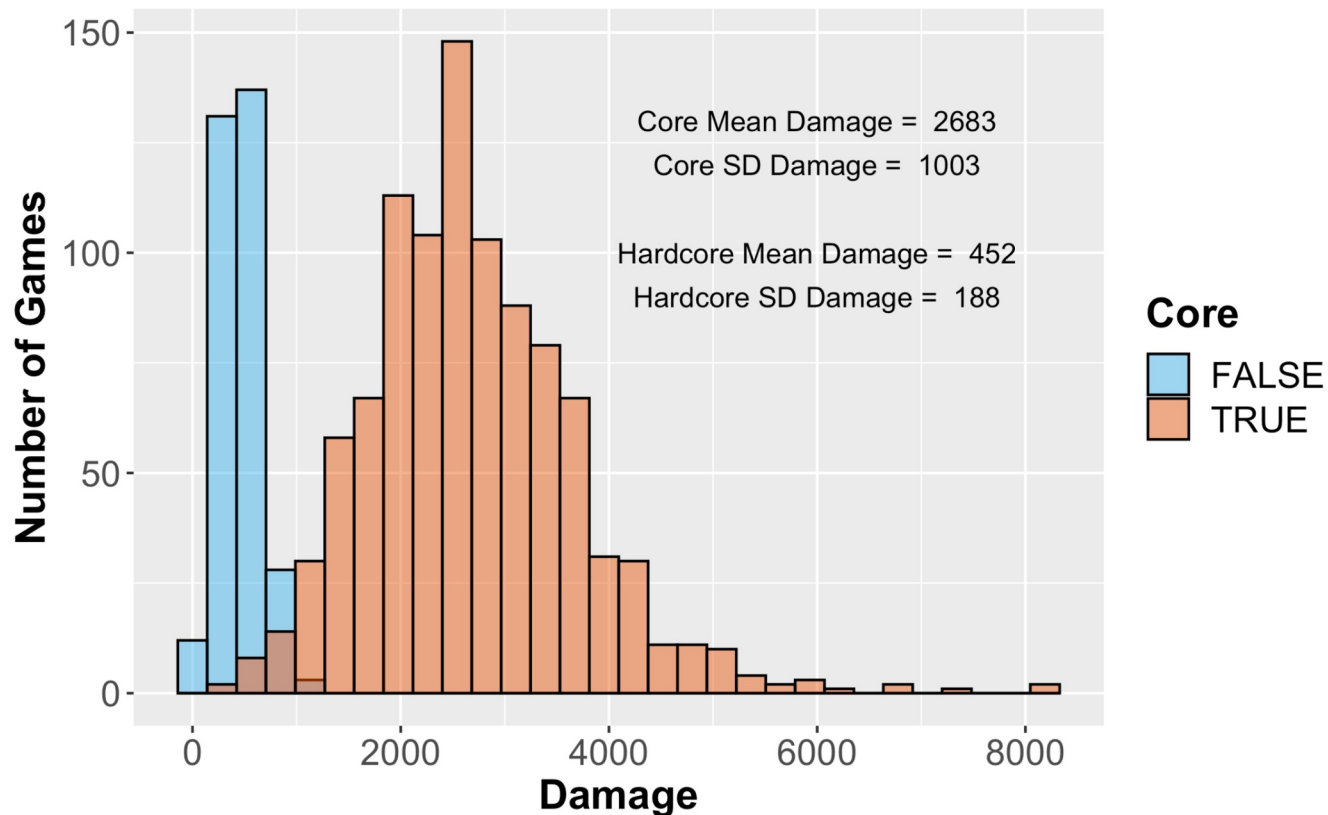- Task: Explore the distribution of Damage.

Bimodal Distribution – What other variables might affect the distribution?

PennState
Eberly College
of Science

Example – Exploratory Data Analysis 3

- Background: Some values in the "GameType" variable include the "HC" designation. Unlike Core games, in HardCore games, players begin with less health and health does not regenerate

- Task: Explore the distribution of Damage using this information.

Core Mean Damage = 2683
Core SD Damage = 1003

Hardcore Mean Damage = 452
Hardcore SD Damage = 188

Core
FALSE
TRUE

Thank you!

Come to the poster session to hear about my experiences using this data, see many more examples, and to discuss!!!

If you are interested in the data, please visit:
https://github.com/matthewdslifko/CallOfDutyProject

- Plan to update after SDSS and USCOTS

Thanks for your time!
Contact Information:

- Name: Matt Slifko

- Email: mds6457@psu.edu

PennState
Eberly College
of Science