Data Science and Pokémon

Anthony Tan

McMaster University

tana4@mcmaster.ca

December 2018

Overview

Introduction

- Hand-held video game released in 1996.

- Currently there are 7 generations of Pokémon games.
- Each generation contains new species to be "discovered".
- Pokémon are trained for sport, or entertainment (generation 3 and onwards). The sport is battling, which will be our focus.

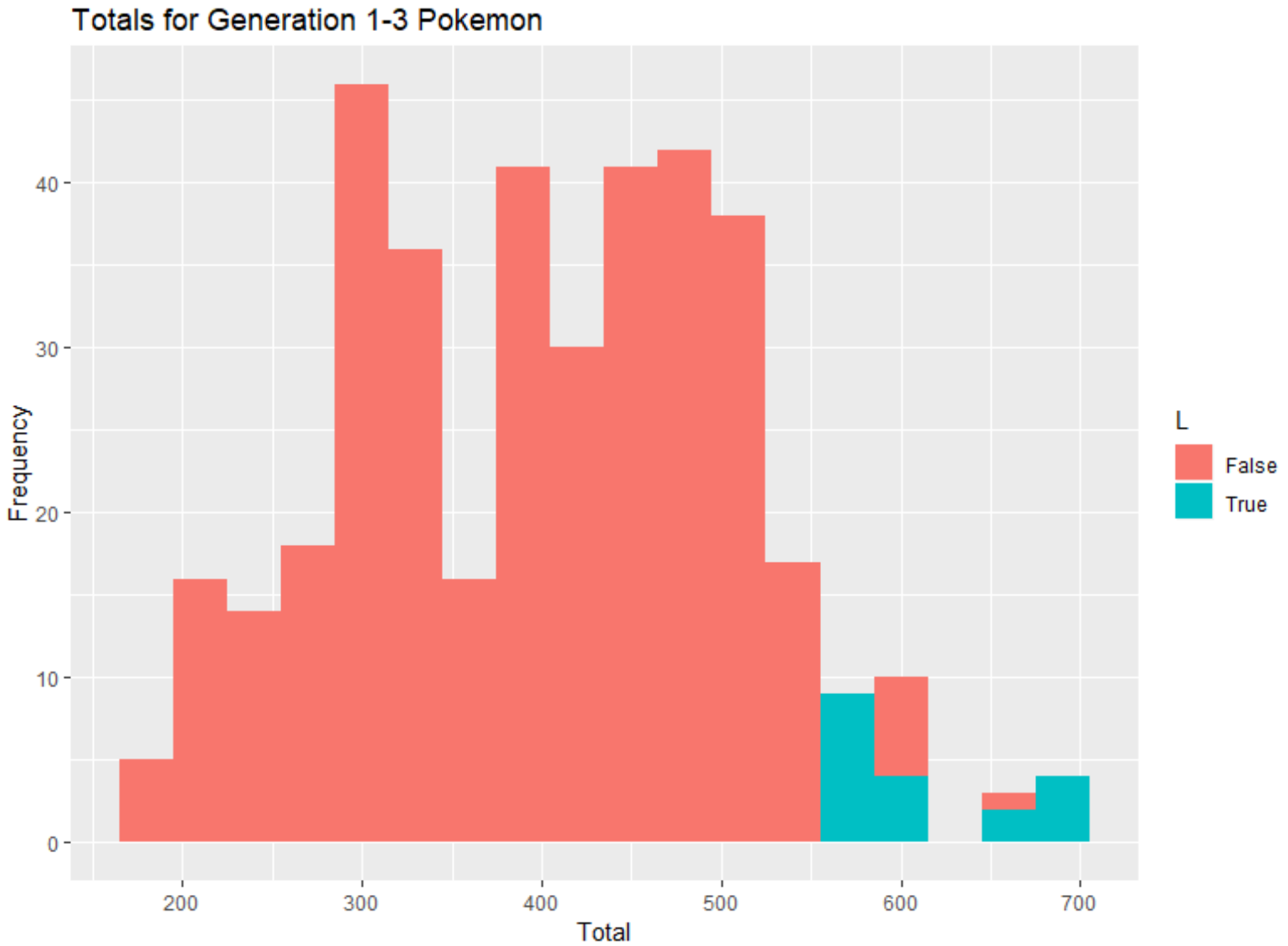
Data I: Organization

- For simplicity, we restrict ourselves to the first 3 generations. In particular, our population will be 386 out of 721.
- The population is divided into two categories: non-legendary (367) and legendary (19).
- Legendary Pokémon are mythological, and actually exist.
- Variables of interest (8):

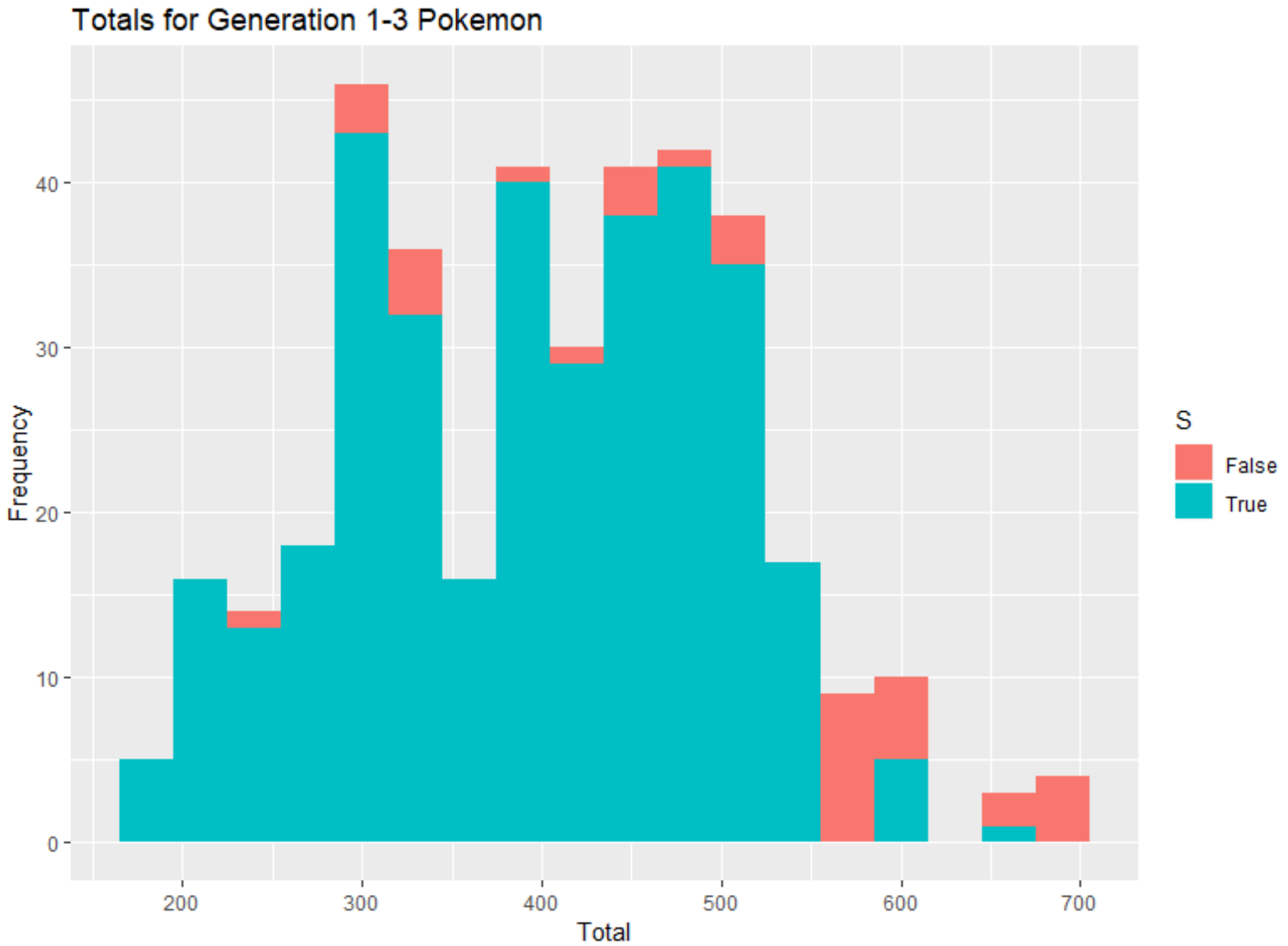
Variable	Type	Description
Combat Metrics	Discrete	Attack, special attack, defense, special defense, speed, and total.
hasGender	String/Discrete	A Pokémon may or may not have gender.
isLegendary	String/Discrete	A Pokémon is either legendary or not.

- The total variable in the table is the sum of attack, special attack, defense, special defense, and speed.

Data II: isLegendary speculation



Data III: hasGender speculation



Data IV: Observations

Observation 1

Legendary Pokémon have higher totals.

Observation 2

Legendary Pokémon do not have gender.

Question

Using Combat Metrics and hasGender, can we predict whether or not a Pokémon isLegendary?

Results

Random Forest		
	False	True
False	111	2
True	0	4
ARI = 0.7765		

Boosting		
	False	True
False	111	1
True	0	5
ARI = 0.8961		

Divisive Hierarchical		
	False	True
False	109	2
True	1	6
ARI = 0.8317		

KNN		
	False	True
False	109	0
True	2	5
ARI = 0.7338		

Conclusion I

Verdict

Performance-wise Ranking:

1. Boosting
2. Divisive Hierarchical
3. Random Forest
4. KNN

Reasons for Error

- Confusion among Pokémon that are legendary with gender and non-legendary without gender.
- Some Pokémon are mislabelled.

Conclusion II

Answer

Yes!

Suggestion(s) for Improvement

- Use catch rate as an additional predictor.

Questions?

Thank you for listening!