

Response Summary:

Mine Worksheet

Goal: to identify patterns, extreme and subtle features about the data

Objectives: Students will identify basic descriptors for the data, and categorize the data according to the specifications from the Parse Worksheet

Outcomes: Three (3) specific questions to be answered using the data

1. Student Information *

First Name	Thomas
Last Name	Cluff
Course (e.g. CGT 270-001)	CGT 270-009
Term (e.g. F2019)	F2021

2. Email Address *

tcluff@purdue.edu

3. Visualization Assignment *

- Lab Assignment

Analyze

4. Basic Descriptors: for each data component from the Parse Worksheet, identify basic descriptors (basic statistics). Explain *

pokedex_number is a primary key (could be considered a continuous nominal integer). Mean, median, mode, max, and min can technically be calculated but it doesn't give any insight into the data except that there are no duplicate Pokémon listed.

name, german_name, and japanese_name are all nominal strings. you can calculate the length but they have no order to them besides the order given by their primary key.

generation is technically a discrete ordinal integer. While an integer is the data type, it represents which game in the series the Pokémon was introduced, so common math means nothing. It could just as easily be replaced with the string of the number ("two" instead of "2").

status is a nominal string that the length could be calculated. It is used to denote the rarity of the Pokémon (normal / legendary) which has gameplay and lore meaning.

species is a nominal string that denotes what type of animal in the real world this Pokémon is close to. Essentially it defines what a designer was inspired by to create this Pokémon.

type_number is a discrete ordinal integer that denotes how many elemental types a Pokémon has. Because there are only two options (1 and 2) this could also be represented with a string or even boolean.

type_1 and type_2 are both strings that say what elemental type a Pokémon has.

height_m and weight_kg are both continuous ratio floating point values that describe the body size of a Pokémon. These can be used to calculate min, max, average, etc.

abilities_number is a discrete ordinal integer that denotes how many abilities a Pokémon has. There are only 4 options (0-3). It could be replaced by a string so mathematic operations do not have value.

ability_1, ability_2, and ability_hidden are nominal strings that denote what abilities a Pokémon has.

total_points, hp, attack, defense, sp_attack, sp_defense, speed, catch_rate, base_friendship, and base_experience are continuous (for int) ratio integers that define the stats and characteristics of a Pokémon. Min, max, mean, median, and mode are all useful calculations.

growth_rate is a nominal string that categorizes how fast a Pokémon gains experience points to level up.

egg_type_number is a ordinal discrete integer that denotes how many egg types a Pokemon has

egg_type_1, egg_type_2 are both nominal strings to denote what egg type a Pokemon has.

percentage_male is a continuous floating point ratio data type that denotes the percentage chance for a pokemon's egg to hatch male vs. female (20 = 20% chance to be male).

egg_cycles is a discrete interval integer that denotes how long it takes an egg to hatch. Because values range from 5 to 120, there is no meaningful zero.

against_normal, against_fire, against_water, against_electric, against_grass, against_ice, against_fight, against_poison, against_ground, against_flying, against_psychic, against_bug, against_rock, against_ghost, against_dragon, against_dark, against_steel, and against_fairy are all discrete ratio floats that have values stepping in the 0.25 amounts from 0 to 2. These represent how effective a Pokémon is against that elemental type.

5. Categorize: consider what is similar and what is different? Categorize the data. Are the variables categorical (normal, ordinal, or rank). Are they quantitative (discrete or continuous)? Show categories. Explain. *

(I ACCIDENTALLY COVERED THIS QUESTION IN MY FIRST DESCRIPTOR ANSWER, SO I AM JUST GOING TO COPY PASTE THAT TEXT INTO THIS BOX INSTEAD OF DOING IT AGAIN)
pokedex_number is a primary key (could be considered a continuous nominal integer). Mean, median, mode, max, and min can technically be calculated but it doesn't give any insight into the data except that there are no duplicate Pokémon listed.

name, german_name, and japanese_name are all nominal strings. you can calculate the length but they have no order to them besides the order given by their primary key.

generation is technically a discrete ordinal integer. While an integer is the data type, it represents which game in the series the Pokémon was introduced, so common math means nothing. It could just as easily be replaced with the string of the number ("two" instead of "2").

status is a nominal string that the length could be calculated. It is used to denote the rarity of the Pokémon (normal / legendary) which has gameplay and lore meaning.

species is a nominal string that denotes what type of animal in the real world this Pokémon is close to. Essentially it defines what a designer was inspired by to create this Pokémon.

type_number is a discrete ordinal integer that denotes how many elemental types a Pokémon has. Because there are only two options (1 and 2) this could also be represented with a string or even boolean.

type_1 and type_2 are both strings that say what elemental type a Pokémon has.

height_m and weight_kg are both continuous ratio floating point values that describe the body size of a Pokémon. These can be used to calculate min, max, average, etc.

abilities_number is a discrete ordinal integer that denotes how many abilities a Pokémon has. There are only 4 options (0-3). It could be replaced by a string so mathematic operations do not have value.

ability_1, ability_2, and ability_hidden are nominal strings that denote what abilities a Pokémon has.

total_points, hp, attack, defense, sp_attack, sp_defense, speed, catch_rate, base_friendship, and base_experience are continuous (for int) ratio integers that define the stats and characteristics of a Pokémon. Min, max, mean, median, and mode are all useful calculations.

growth_rate is a nominal string that categorizes how fast a Pokémon gains experience points to level up.

egg_type_number is a ordinal discrete integer that denotes how many egg types a Pokemon has

egg_type_1, egg_type_2 are both nominal strings to denote what egg type a Pokemon has.

percentage_male is a continuous floating point ratio data type that denotes the percentage chance for a pokemon's egg to hatch male vs. female (20 = 20% chance to be male).

egg_cycles is a discrete interval integer that denotes how long it takes an egg to hatch. Because values range from 5 to 120, there is no meaningful zero.

against_normal, against_fire, against_water, against_electric, against_grass, against_ice, against_fight, against_poison, against_ground, against_flying, against_psychic, against_bug, against_rock, against_ghost, against_dragon, against_dark, against_steel, and against_fairy are all discrete ratio floats that have values stepping in the 0.25 amounts from 0 to 2. These represent how effective a Pokémon is against that elemental type.

6. Temporal: is the data streaming data? How is it stored (all at one time, over several years in years, days, minutes, seconds)? Explain. *

None of this data is dependent on time as it is hard coded into video games to support gameplay features and interactivity.

7. Range and Distribution: what is the distribution of the data? Few values, small size, evenly spread, sparse or dense? Explain. *

All most all of the data points exist for each row so I would say the data is dense. The columns that are missing data points are for categories where a Pokémon might have 1-2 types or 1-2 abilities. the column with missing data would be the type_2 column on a Pokémon row that only has one type. Looking at graphs of the Pokémon stats and effectiveness against elements (pretty much the only data that isnt nominal), the stats are evenly distributed. This makes sense because the data is designed for a video game, so balance in fighting stats is desired.

Evaluate

8. Questions and Assumptions: list at least 3 questions you plan to answer with the data or list the questions if they were provided. Must be complete sentences and end in a question mark. What assumptions are you making? *

Question 1	Which elemental type of Pokémon is on average the most effective against other elemental types?
Question 2	Which elemental type of Pokémon is on average the biggest (weight or height)?
Question 3	What element and egg type relation is the most prevalent amongst all the different Pokémon?
Assumptions	I am assuming that all data is up to date with the latest generation of games as data can change between games.
